Personalised medicine approaches for investigation and application of new biomarkers

Vesna Gorenjak

▶ **To cite this version:**

Vesna Gorenjak. Personalised medicine approaches for investigation and application of new biomarkers. Biochemistry, Molecular Biology. Université de Lorraine, 2019. English. NNT : 2019LORR0106 . tel-02397858

HAL Id: tel-02397858

https://hal.univ-lorraine.fr/tel-02397858

Submitted on 6 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Ecole Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE DE LORRAINE

Mention: « Sciences de la Vie et de la Santé »

Par: **Vesna GORENJAK**

Stratégies de médecine personnalisée pour l'étude et l'utilisation de nouveaux biomarqueurs

Le 29 Octobre 2019

Membres du jury:

Rapporteurs:	**Mme. Janja MARC**	Professeur, University of Ljubljana, Faculty of Pharmacy, Ljubljana
	Mme. Csilla SIPEKY	Professeur, University of Turku, Institute of Biomedicine, Turku
Examinateurs:	**M. Bertrand RIHN**	Professeur, Université de Lorraine, Institut Jean Lamour, Nancy
	M. Georges DAGHER	DR1, INSERM US13, Paris
	Mme. Sophie VISVIKIS-SIEST	DR1, INSERM U1122; IGE-PCV, Université de Lorraine, Nancy – (Directeur de thèse)
	Mme. Maria STATHOPOULOU	IR2 (HDR), INSERM U1122; IGE-PCV, Université de Lorraine, Nancy – (Co-directeur de thèse)
Membre invité:	**M. Luc FERRARI**	Professeur, Université de Lorraine, Institut Jean Lamour, Nancy

Université de Lorraine - UMR INSERM U1122 ; IGE-PCV - Faculté de Pharmacié - 30, rue Lionnois - 54000 Nancy - France

Acknowledgements

*to **Dr Sophie Visvikis-Siest**,*

for your encouragement, when you proposed this three-year journey, and motivation all the way through. For all the exceptional opportunities I had! They helped to develop my expertise, meet remarkable scientists and gave me unforgettable work experiences. You trusted in my abilities and taught me the passion of a researcher, who always seeks for new adventures. Thank you!

*to **Dr Maria Stathopoulou,***

for always having an answer to a question and kindness for taking your time. Your expertise and knowledge helped me in reduction of my scientific papers. And your guidance and advice - they always made things easier.

*to **the members of my jury,***

for having accepted to evaluate my work, I express my sincere gratitude.

*to **my colleagues,***

*for always being around when needed. Thank you, **Alan**, for staying my friend. Thank you, **Alexandros**, for being a great fellow-student! Thank you, **Brigitte**, for your constant smile and support! Thank you, **Christine**, for the nice time we shared during the experiments! Thank you, **Georges**, for your kind support! Thank you, **Satish**, for all the positive energy! And thank you, **Sébastien**, for all interesting scientific discussions!*

***moji družini**,*

ki nikoli ni bila daleč stran …

Hvala za vašo vzpodbudo, ko je bilo hudo, in za skupno veselje, ko je bilo lepo. Rada vas imam.

*to **my friends**,*

from all around the globe! Without you, life would be so dull!

*et à **Morgan**,*

de m'avoir tenu la main en traversant ce chemin bouleversant que nous avons partagé. Je ne peux pas l'imaginer sans toi.

*A special thanks goes to the **LORRAINE Region** and **RHU,** who financed my PhD studies.*
Without their support, this work would not be possible.

*Also, I would like to gratefully acknowledge **GEENAGE** and **FEDER** projects,*
who financially supported our research work.

ABSTRACT

The fight against common chronic diseases, which are characterised by complex mechanisms of molecular regulation, requires the implementation of new risk prediction and prevention strategies. Personalised medicine offers sophisticated approaches for successful management of the morbidities of the ageing population. In this thesis, inspired by the principles of personalised medicine, we describe an integrative approach combining "-omics" methodologies. We use a model of a "common denominator" for cardiovascular disease (CVD) and other chronic diseases to identify biomarkers linked with common diseases risk factors and molecular pathways.

The results of this work are presented in five original publications, three review articles, and one publication under redaction. All publications aim to strengthen and facilitate the implementation of personalised medicine in healthcare. The three main biomarkers of our interest are triggering receptors expressed on myeloid cells (TREM2), vascular endothelial growth factor A (VEGF-A) and telomere length (TL) and are studied mainly in the supposed healthy population of the STANISLAS Family Study.

With the investigation of genetic variants, located in the region of the *TREM2* gene, we identified the association of the SNP rs6918289 with increased levels of TNF-α and intima-media thickness of the femoral artery (IMT-F). As inflammation and increased IMT-F are important indicators of the formation of the atherosclerotic plaque, the minor allele (T) of rs6918289 might be considered a risk allele for inflammatory diseases and atherosclerosis. With the use of an epigenome-wide association studies (EWAS), we identified novel epigenetic biomarkers, related to common diseases risk factors: central obesity and increased lipid levels. One methylation site (CpG) was associated with increased waist circumference (WC) (cg16170243), which could explain the epigenetic regulation of central obesity *via* increased insulin resistance. Moreover, an EWAS of the triglyceride (TG) levels identified two significant CpG sites, one of which was replicated in the adipose tissue (cg04580029), giving insights into epigenetic regulation of lipid levels. The results of these two last studies provided new epigenetic biomarkers and might contribute to future diagnostics and therapeutic interventions. An EWAS was also used to study the epigenetic regulation of VEGF-A concentrations; 20 CpG sites were identified, and their relations with VEGF-A were

analysed through detailed bioinformatics analysis. Methylation of genes, such as *TPX2* and *HAS,* could affect their activity and would in turn cause increased VEGF-A concentrations.

VEGF-A was further investigated for its relation with 11 cytokines, important mediators of common physiological pathways of the majority of chronic diseases. VEGF-A protein levels were associated with IL-4, MCP1 and EGF. Specific *VEGF-A* mRNA isoforms were also investigated for their association with cytokines; $VEGF_{165}$ showed significant associations with MCP1 and IL-1α and $VEGF_{189}$ with IL-4 and IL-6. Together with another important biomarker, TL, we studied the role of VEGF-A in atherosclerosis, and we identified one VEGF-A related genetic variant significantly associated with telomere attrition (calculated as a ratio between leukocyte TL and muscular TL). This genetic variant could present a common denominator of chronic diseases.

In conclusion, the employment of diverse methodologies for the investigation of common chronic diseases risk factors and pathways provided new diagnostic markers and generated results, which could help to improve the diseases risk prediction based on the individual genetic "make-up". New insights into associations between different biomarkers might help in understanding the (patho) physiological pathways common between CVDs and other chronic diseases. Finally, we hope that the results of this thesis through close collaboration with industry will facilitate the implementation of personalised healthcare.

RÉSUMÉ

La lutte contre les maladies chroniques courantes, qui sont caractérisées par des mécanismes complexes de régulation moléculaire, nécessite aujourd'hui la mise en œuvre de nouvelles stratégies de prédiction du risque et de prévention. La médecine personnalisée représente une approche sophistiquée pour réussir la prise en charge des morbidités de populations vieillissantes. Dans le cadre de ces travaux de thèse inspirés par les principes de la médecine personnalisée, nous décrivons une approche intégrative qui associe plusieurs méthodologies « -omiques ». Nous avons utilisé un modèle de « dénominateur commun » pour les maladies cardiovasculaires et d'autres maladies chroniques afin d'identifier des biomarqueurs associés aux facteurs de risque et aux voies biologiques de maladies courantes.

Les résultats de ces travaux sont présentés dans 5 publications originales, 3 articles de revue et 1 publication originale en cours de redaction. Ils sont tous destinés à renforcer et faciliter la mise en œuvre de la médecine personnalisée. Trois principaux biomarqueurs ont été étudiés, TREM2, VEGF-A et la longueur des télomères (LT), et ce dans la population d'individus en bonne santé apparente de l'étude familiale STANISLAS (SFS). Par l'étude de variants génétiques localisés dans la région comportant le gène *TREM2*, nous avons identifié une association entre le SNP rs6918289 et à la fois de taux élevés de TNF-α et une augmentation de l'épaisseur intima-média de l'artère fémorale (IMT-F). L'inflammation et l'IMT-F étant des indicateurs importants de la formation de plaque d'athérosclérose, l'allèle mineur (T) de rs6918289 pourrait être un allèle à risque pour les maladies inflammatoires et l'athérosclérose. Grâce à des études d'association panépigénomique (EWAS), nous avons identifié de nouveaux marqueurs épigénétiques liés à des facteurs de risque de maladies courantes, à l'obésité abdominale et à des taux élevés de lipides. Un site CpG de méthylation (cg16170243) était associé à une augmentation du tour de taille, contribuant à expliquer la régulation épigénétique de l'obésité abdominale par une résistance accrue à l'insuline. De plus, une étude EWAS des taux de triglycérides a permis d'identifier deux sites CpG significatifs. L'un de ces deux sites a pu être confirmé dans le tissu adipeux (cg04580029) et a permis de mieux comprendre la régulation épigénétique de l'augmentation du taux de triglycérides. Les résultats de ces deux dernières études ont donné lieu à de nouveaux biomarqueurs épigénétiques qui pourraient être utilisés dans l'avenir dans des applications diagnostiques et thérapeutiques. Une étude EWAS a également été réalisée pour décrire la régulation épigénétique des concentrations de VEGF-A. Vingt sites CpG ont pu être identifiés

ainsi et leurs relations avec la régulation du VEGF-A ont été examinées par analyse bioinformatique poussée. La méthylation de gènes tels que *TPX2* ou *HAS* pourrait avoir un effet sur leur activité et donc conduire à une augmentation de la concentration de VEGF-A. Les liens entre le VEGF-A et 11 cytokines, des médiateurs importants des voies physiopathologiques courantes impliquées dans la plupart des maladies chroniques, ont également été étudiés. Le taux de protéine VEGF-A était associé à IL-4, MCP1 et EGF. Les associations entre les cytokines et des isoformes spécifiques de l'ARNm du *VEGF-A* ont également été évaluées : le $VEGF_{165}$ était associé de manière significative à MCP1 et IL-1α, et le $VEGF_{189}$ à IL-4 et IL-6. Nous avons étudié le rôle du VEGF-A et d'un autre biomarqueur important, LT, dans l'athérosclérose. Cela a permis d'identifier un variant génétique lié au VEGF-A associé de manière significative à l'attrition des télomères (calculé sous la forme d'un rapport de LT dans les leucocytes et de LT dans les cellules musculaires). Ce variant génétique pourrait constituer un dénominateur commun pour les maladies chroniques.

En conclusion, l'utilisation de diverses méthodologies pour étudier les facteurs de risque et les voies impliquées dans des maladies chroniques courantes a permis d'identifier de nouveaux marqueurs diagnostiques et de générer des résultats qui pourraient améliorer la prédiction du risque de maladie basée sur le profil génétique de chaque individu. Les nouvelles données apportées sur les associations entre les différents biomarqueurs pourraient aider à mieux comprendre les voies physiopathologiques communes entre les maladies cardiovasculaires et d'autres maladies chroniques. Enfin, nous espérons que, grâce à une collaboration étroite avec des partenaires privés, les résultats de ces travaux de thèse faciliteront la mise en œuvre de la médecine personnalisée.

TABLE OF CONTENTS

ABSTRACT 1
RÉSUMÉ 3
TABLE OF CONTENTS 5
LIST OF ABBREVIATIONS 8
LIST OF FIGURES 11
LIST OF TABLES 12
LIST OF PUBLICATIONS 13
List of original publications 13
List of review publications 14
Other publications 14
Publication under redaction 15
PRÉSENTATION SYNTHÉTIQUE DE LA THÈSE 16
Situation du sujet 17
Hypothèses et objectifs 19
Discussion 21
Conclusion et perspectives 39
FOREWORD 44
Chapter I. INTRODUCTION 48
1.1 Chronic diseases 49
1.1.1 Cardiovascular diseases 50
1.1.2 Diabetes 52
1.1.3 Cancer 53
1.1.4 The major chronic disease risk factors 55
1.2 Common disease pathways 62
1.2.1 Inflammation 63
1.2.2 Angiogenesis 67
1.3 Biomarkers 69
1.3.1 VEGF-A 71
1.3.2 Telomere length 73
1.4 Methodologies of personalised medicine 75
1.4.1 Genetics 76
1.4.2 Epigenetics 82

1.4.3 Gene expression studies – transcriptomics 85
1.4.4 Challenges of personalised medicine 87
Chapter II. HYPOTHESIS AND OBJECTIVES 89
Chapter III. MATERIALS AND METHODS 93
3.1 Populations 94
3.1.1 The STANISLAS Family Study 94
3.1.2 Hypertensive population 94
3.1.3 The MuTHER study 95
3.1.4 The TELARTA study 95
3.2 DNA methylation assessment and genotyping 96
3.2.1 DNA methylation assessment 96
3.2.2 Genotyping 97
3.3 Gene expression analysis 97
3.3.1 PBMC isolation 97
3.3.2 mRNA quantification 98
3.4 Protein measurement 98
3.4.1 Total protein extraction from PBMCs 98
3.4.2 Protein quantification 98
3.5 Lipid measurement 99
3.6 Telomere length measurement 99
3.7 Statistical analysis 99
3.7.1 Methylation analysis 99
3.8 Characterisation: expression of recombinant antibody fragments 100
Chapter IV. PUBLICATIONS - RESULTS 101
Publication Nº 1 102
Publication Nº 2 130
Publication Nº 3 145
Publication Nº 4 156
Publication Nº 5 168
Publication Nº 6 194
Publication Nº 7 207
Publication Nº 8 235
Publication under redaction 253
Chapter V. GENERAL DISCUSSION 264
Chapter VI. CONCLUSION AND PERSPECTIVES 282
6.1 Perspectives of our results 286

BIBLIOGRAPHY 289
OTHER PUBLICATIONS 305
ANNEX 322

LIST OF ABBREVIATIONS

AD ….. Alzheimer's disease
AMPK ….. AMP/ATP-Binding subunit of AMP-activated protein kinase
ARL4A ….. ADP ribosylation factor like GTPase 4A
AUTS2 ….. Activator of transcription and developmental regulator AUTS2
BM ….. Basement membrane
BMI ….. Body mass index
BRC ….. Biological resources centre
C16orf95 ….. Chromosome 16 open reading frame 95
CAD ….. Coronary artery disease
CAPE ….. Combined analysis of pleiotropy and epistasis
CC ….. Chemotactic cytokines
CCPPRB ….. Comité consultatif de protection des personnes dans la recherche biomédicale
cDNA ….. Complementary DNA
CLHC1 ….. Clathrin heavy chain linker domain containing 1
CMP ….. Centre for preventive medicine
CNIL ….. Commission nationale de l'informatique et des libertés
CNVs ….. Copy-number variants
COPD ….. Chronic obstructive pulmonary disease
CpG ….. Cytosine-phosphate-guanine
CRAN ….. Comprehensive R archive network
CRP ….. C-reactive protein
CTNNB1 ….. Catenin beta 1
CVD ….. Cardiovascular disease
dbSNP ….. Single nucleotide polymorphism database
DNA ….. Deoxyribonucleic acid
DNAm ….. DNA methylation
EC ….. Endothelial cell
EDTA ….. Ethylenediaminetetraacetic acid
EGF ….. Epidermal growth factor
ELISA ….. Enzyme-linked immunosorbent assay
ESD ….. Esterase D
ETV1 ….. ETS Variant 1
EWAS ….. Epigenome-wide association studies
FDA ….. Food and drug administration
FDR ….. False discovery rate
FOG ….. Friend of GATA
FOXL1 ….. Forkhead box L1
GFP ….. Green fluorescent protein
GLIS3 ….. GLIS family zinc finger 3
GRS ….. Genetic risk score
GWAF ….. Genome-wide association/interaction analysis and rare variant analysis with family data
GWAS ….. Genome-wide association studies
HAS1 ….. Hyaluronan synthase 1
HDL ….. High-density lipoprotein
HGP ….. Human genome project

HWE ….. Hardy-Weinberg equilibrium
IFN ….. Interferon
IGE-PCV ….. Interactions gène-environnement en physiopathologie cardio-vasculaire
IL ….. Interleukin
IL17RD ….. Interleukin 17 receptor D
IMT-F ….. Intima–media thickness of femoral artery
IPTG ….. Isopropyl β-D-1-thiogalactopyranoside
KALRN ….. Kalirin rhoGEF kinase
KASP ….. Kompetitive allele specific PCR
KCNMB2 ….. Potassium calcium-activated channel subfamily m regulatory beta subunit 2
kDa ….. Kilodaltons
KDR ….. Kinase insert domain receptor
KREMEN2 ….. Kringle containing transmembrane protein 2
LDL ….. Low-density lipoprotein
LRP6 ….. Low-density lipoprotein receptor-related protein 6
LTL ….. Leukocyte telomere length
MAF ….. Minor allele frequency
MCP ….. Monocyte chemoattractant protein
miRNA ….. Micro RNA
MR ….. Mendelian randomisation
mRNA ….. Messenger RNA
MTL ….. Muscle telomere length
MuTHER ….. Multiple tissue human expression resource
NANOS3 ….. Nanos C2HC-type zinc finger 3
NHGRI-EBI ….. National human genome research institute - European bioinformatics institute
NK ….. Natural killer
NKT ….. Natural killer T
PBMC ….. Peripheral blood mononuclear cell
PIGB ….. Phosphatidylinositol glycan anchor biosynthesis Class B
PRKAG2 ….. Protein kinase AMP-activated non-catalytic subunit gamma 2
QTL ….. Quantitative trait locus
RNA ….. Ribonucleic acid
RPL31 ….. Ribosomal protein L31
RTKs ….. Receptors tyrosine kinases
RT-PCR ….. Reverse transcription - Polymerase chain reaction
SAA ….. Serum amyloid A
SBE ….. Single base extension
SEPSECS ….. Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase
SFS ….. STANISLAS Family Study
sICAM ….. Soluble intercellular adhesion molecule
SMIM21 ….. Small integral membrane protein 21
SNP ….. Single nucleotide polymorphism
SOX2OT ….. SOX2 overlapping transcript
ST8SIA5 ….. ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5
sVCAM ….. Soluble vascular cell adhesion molecule
TB ….. Terrific broth
TBL1XR1 ….. Transducin beta like 1 X-linked receptor 1
TC ….. Total cholesterol
TELARTA ….. Telomeres and arterial ageing

TET ….. Ten-eleven translocation
TFBS ….. Transcription factors binding sites
TG ….. Triglycerides
TL ….. Telomere length
TLR ….. Toll-like receptors
TNF-α ….. Tumour necrosis factor alpha
TPX2 ….. TPX2 microtubule nucleation factor
TREM ….. Triggering receptors expressed on myeloid cells
TREML ….. Triggering receptor expressed on myeloid cells like
TRL ….. Triglyceride-rich lipoprotein
TSHZ1 ….. Teashirt zinc finger homeobox 1
TSS ….. Transcription start sites
USP43 ….. Ubiquitin specific peptidase 43
VEGF ….. Vascular endothelial growth factor
VEGFR ….. Vascular endothelial growth factor receptor
VLDL ….. Very-low-density lipoprotein
VNTRs ….. Variable-number tandem repeats
WC ….. Waist circumference
WGS ….. Whole-genome sequencing
WHO ….. World health organisation
ZNF621 ….. Zinc finger protein 621
ZSWIM4 ….. Zinc finger SWIM-type containing 4

LIST OF FIGURES

Figure 1: Four major disease risk factors 49
Figure 2: Obesity-related comorbidities affect multiple organ systems. 50
Figure 3: Four main types of cardiovascular disease. 51
Figure 4: Cross-section of atherosclerotic plaque 52
Figure 5: Cancer cells form a tumour 54
Figure 6: Percentage of adults aged 18+ years old who are defined as obese based on their BMI 56
Figure 7: Obesity-induced inflammatory diseases 56
Figure 8: Pathophysiology of TG-rich lipoproteins in the progression of atherosclerosis 59
Figure 9: Sex difference in global mortality in the year 2017 61
Figure 10: Protein-protein interaction network 62
Figure 11: The inflammatory pathway 63
Figure 12: Production and action of cytokines 64
Figure 13: The cascade of events associated with sprouting angiogenesis 67
Figure 14: Two kinds of pathological situations caused by increased or reduced angiogenesis 68
Figure 15: VEGF-A role in different (patho)physiological mechanisms 71
Figure 16: Intracellular signalling pathways for VEGF-A 72
Figure 17: The structure of telomere 73
Figure 18: Shorter inherited telomeres have been related to advanced atherosclerotic lesions 74
Figure 19: The strategy of personalised medicine 75
Figure 20: The increasing number of GWAS studies and sample sizes in years 2007-2017 79
Figure 21: A schematic presentation of MR 82
Figure 22: Major epigenetic mechanisms 83
Figure 23: DNA cytosine ring 84
Figure 24: Transcription of DNA as a part of gene expression 86
Figure 25: Tools needed for prediction and personalised care 88
Figure 26: Graphical abstract of the published article 130
Figure 27: Methylated DNA 145
Figure 28: Structure of the VEGF-A protein 156
Figure 29: Manhattan plot of significant CpG associations with VEGF-A levels. 158
Figure 30: PBMCs are highly specialised immune cells 168
Figure 31: The relation between inflammatory cytokines and VEGF-A protein/mRNA levels. 171
Figure 32: Cytokine and macrophage infiltration in atherosclerotic lesions 194
Figure 33: Rs6993770 (red stripe) is located on the intron of the *ZFPM2* 259
Figure 34: Schematic presentation of applied methods 285

LIST OF TABLES

Table 1. Molecules from the class of cytokines and their role in the inflammation process. 65
Table 2. Association of methylation sites with TG levels. 132
Table 3. Replication of the significant results in adipose tissue. 132
Table 4. Association of methylation site with WC level, adjusted for BMI. 147
Table 5. Novel associations between DNA methylation and VEGF-A PBMCs levels. 159
Table 6. Significant associations between VEGF-A protein levels and inflammatory molecules. 170
Table 7. Significant associations between the expression of *VEGF-A* mRNA isoforms 171
Table 8. Association of the SNPs in *TREM2* region with TNF-α levels in children of the SFS. 196
Table 9. Association of rs6918289 with TNF-α levels 197
Table 10. Association of rs6918289 with IMT-F 198
Table 11. Results of MAF and HWE analysis 257
Table 12. Association between VEGF-A related genetic variants and LTL. 257
Table 13. Association between VEGF-A related genetic variants and MTL. 258
Table 14. Association between VEGF-A related genetic variants and LTL/MTL. 258
Table 15: Multiple regression analysis of VEGF-A 259

LIST OF PUBLICATIONS

List of original publications

V. Gorenjak*, A. A. Aldasoro Arguinano*, S. Dadé, M. G. Stathopoulou, D. R. Vance, C. Masson and S. Visvikis-Siest:
The polymorphism rs6918289 located in the downstream region of the TREM2 gene is associated with TNF-alpha levels and IMT-F.

Sci Rep, 8(1), 7160 (2018)

(Equal first author*)

T. Xie*, **V. Gorenjak***, M. G. Stathopoulou*, S. Dadé, E. Marouli, C. Masson, H. Murray, J. Lamont, P. Fitzgerald, P. Deloukas and S. Visvikis-Siest:
Epigenome-Wide Association Study (EWAS) of Blood Lipids in Healthy Population from STANISLAS Family Study (SFS).

Int J Mol Sci, 20(5) (2019)

(Equal first author*)

V. Gorenjak*, D. R. Vance*, A. M. Petrelis, M. G. Stathopoulou, S. Dadé, S. El Shamieh, H. Murray, C. Masson, J. Lamont, P. Fitzgerald and S. Visvikis-Siest:
Peripheral blood mononuclear cells extracts VEGF protein levels and VEGF mRNA: Associations with inflammatory molecules in a healthy population.

PLoS One, 14(8), e0220902 (2019)

(Equal first author*)

T. Xie*, **V. Gorenjak***, M. G. Stathopoulou, S. Dadé, E. Marouli, C. Masson, H. Murray, J. Lamont, P. Fitzgerald, P. Deloukas and S. Visvikis-Siest:
Epigenome-Wide Association Study Detects a Novel Loci Associated with Central Obesity in Healthy Subjects.

Under submission

(Equal first author*)

V. Gorenjak*, D. R. Vance*, S. Dadé, M. G. Stathopoulou, L. Doherty, T. Xie, H. Murray, C. Masson, J.Lamont, P. Fitzgerald and S. Visvikis-Siest:
Epigenome-wide association study in healthy individuals identifies significant associations with DNA methylation and PBMCs extracts VEGF-A concentration.

Under submission

(Equal first author*)

List of review publications

V. Gorenjak, S. Akbar, M. G. Stathopoulou and S. Visvikis-Siest:
The future of telomere length in personalized medicine.

Front Biosci (Landmark Ed), 23, 1628-1654 (2018)

S. Visvikis-Siest*, **V. Gorenjak*** and M. G. Stathopoulou*:
Personalised Medicine: The Odyssey from Hope to Practice.

J Pers Med, 8(4) (2018)

(Equal first author*)

V. Gorenjak, A. M. Petrelis, M. G. Stathopoulou S. Visvikis-Siest:
Telomere length determinants in childhood.

Clin Chem Lab Med (2019)

Other publications

M. G. Stathopoulou, T. Xie, D. Ruggiero, J. Chatelin, M. Rancier, G. Weryha, M. J. Kurth, A. A. Aldasoro Arguinano, **V. Gorenjak**, A. M. Petrelis, G. Dagher, G. Dedoussis, P. Deloukas, J. Lamont, J. Marc, M. Simmaco, R. Schaik, F. Innocenti, J. L. Merlin, J. Schneider, B. Z. Alizadeh, M. Ciullo, S. Seshadri and S. Visvikis-Siest:
A transnational collaborative network dedicated to the study and applications of the vascular endothelial growth factor-A in medical practice: the VEGF Consortium.

Clin Chem Lab Med (2017)

S. Visvikis-Siest, **V. Gorenjak**, M. G. Stathopoulou, A. M. Petrelis, G. Weryha, C. Masson, B. Hiegel, S. Kumar, R. Barouki, E. Boerwinkle, G. Dagher, P. Deloukas, F. Innocenti, J. Lamont, M. Marschler, H. Meyer, U. A. Meyer, C. Nofziger, M. Paulmichl, C. Vacher and L. Webster:

The 9th Santorini Conference: Systems Medicine, Personalised Health and Therapy. "The Odyssey from Hope to Practice", Santorini, Greece, 30 September(-)3 October 2018.

J Pers Med, 8(4) (2019)

Publication under redaction

Discovering the Relation between Telomere Length and VEGF-A.

PRÉSENTATION SYNTHÉTIQUE DE LA THÈSE

Situation du sujet

Au cours des dernières décennies, les systèmes de santé les plus avancés ont développé le concept de médecine personnalisée et ont adopté de nouvelles stratégies pour lutter contre les maladies chroniques courantes auxquelles sont confrontées les populations vieillissantes partout dans le monde. L'objectif est de formuler un diagnostic et un pronostic individualisés, et mettre en œuvre des traitements sur mesure d'après les données personnelles du patient et pas uniquement d'après la maladie, qui a abouti au développement de nouveaux outils diagnostiques et thérapeutiques. Mais surtout, ces approches prennent en compte les informations génétiques de l'individu qui, associées à d'autres données cliniques, permettent de poser un diagnostic plus précis, de mettre en œuvre un traitement plus efficace, et d'améliorer le pronostic et l'issue clinique.

Cela a marqué une nouvelle ère en médecine mais nécessitait d'énormes efforts en termes de développement de nouvelles technologies, méthodologies, bases de données (big data) et de cohortes à populations importantes afin de générer les données nécessaires pour expliquer les principes de la régulation génétique de maladies complexes. L'épidémiologie génétique a émergé en tant que nouvelle discipline à part entière consacrée à l'étude du rôle de facteurs génétiques et leurs interactions avec l'environnement. Les études d'association pangénomique (GWAS) à grande échelle qui analysent plusieurs milliers de polymorphismes mononucléotidiques (SNP) chez plusieurs milliers d'individus ont permis d'identifier de nombreux variants génétiques. Ces variants contribuent tous dans une certaine mesure à la survenue des maladies complexes polygéniques et multifactorielles. Pourtant, cela n'a pas suffi pour comprendre pleinement la régulation génétique des maladies chroniques. Le chaînon manquant entre l'exposition environnementale, la génétique et de nombreuses maladies courantes que les études d'association pangénomique n'arrivaient pas à expliquer a été attribué aux modifications épigénétiques de la séquence d'ADN. L'étude des modifications de la méthylation du génome est devenue ces derniers temps l'outil épigénomique le plus souvent utilisé pour étudier l'impact de l'environnement sur la régulation génique.

Mais de quelle façon toutes ces avancées ont-elles contribuées à changer la médecine ?

Nous nous trouvons encore aujourd'hui dans la période de transition : des tests génétiques sont maintenant utilisés pour dépister des maladies, les premières thérapies géniques sont en

voie d'approbation et de plus en plus d'outils de médecine personnalisée pénètrent le marché, mais les références en milieu hospitalier restent toujours les approches médicales classiques. Les laboratoires de recherche et l'industrie pharmaceutique ont la ferme intention d'améliorer le système actuel et de proposer de nouvelles stratégies basées sur le profil génétique de chaque individu.

Des liens doivent maintenant être établis entre les vastes quantités de données utiles générées au cours de la dernière décennie afin d'apporter de nouvelles connaissances en biologie et physiopathologie humaine. De tels efforts pourraient améliorer les stratégies actuelles de prévention, de prédiction et de traitement de maladies chroniques fréquentes et permettre de remplacer les approches existantes par des outils de médecine personnalisée. Cela implique le développement de biomarqueurs qui serviront à déterminer la susceptibilité par rapport à une maladie, le diagnostic, le pronostic et le traitement, et permettront de proposer de nouvelles cibles thérapeutiques basées sur la connaissance des voies biologiques et les liens entre les molécules impliquées.

L'unité de recherche UMR INSERM U1122 Interactions Gène-Environnement en Physiopathologie Cardiovasculaire (IGE-PCV) a déjà adopté les nouvelles méthodologies de la médecine personnalisée et s'attache à explorer les composants génétiques et environnementaux impliqués dans des maladies courantes liées à l'âge. Notre grand Centre de Ressources Biologiques (CRB IGE-PCV), constitué de populations saines et pathologiques permettant d'étudier les maladies cardiovasculaires, est au cœur des travaux de l'unité. Cette volonté de créer une grande base de données pour des études d'épidémiologie génétique a mené au lancement de l'étude familiale STANISLAS. Comprenant les données longitudinales de plus de 1000 familles, cette cohorte représente une des principales populations de la biobanque et a fourni une vaste quantité de données génétiques, épigénétiques et cliniques pour l'étude de la variabilité de phénotypes pathologiques intermédiaires et de déterminants environnementaux.

La disponibilité de données issues de notre biobanque et nos collaborations étroites avec d'autres cohortes nous ont permis d'étudier la régulation génétique et épigénétique de phénotypes intermédiaires impliqués dans plusieurs maladies liées à l'âge afin d'identifier des biomarqueurs communs.

Pour ce faire, nous avons mis en œuvre diverses approches, notamment les études d'association panépigénomique (EWAS) pour analyser les mécanismes de régulation

épigénétique de différentes molécules, et les études d'association entre des variants génétiques et différents phénotypes intermédiaires pour décrire l'architecture génétique de traits complexes.

À ce moment-là, nous concentrions tous nos efforts sur la connaissance des phénotypes intermédiaires impliqués dans les mécanismes physiopathologiques et des facteurs de risque présentant un intérêt spécial et potentiellement utilisables en pratique clinique. Comprendre les caractéristiques génétiques complexes de ces phénotypes permettrait de mieux -identifier les patients en prenant en compte leur profil génétique, et connaître les associations entre différentes molécules, ce qui permettrait de proposer des traitements complémentaires. Enfin, de nouveaux biomarqueurs de susceptibilité pourraient être identifiés grâce à la compréhension de la régulation épigénétique de facteurs de risque intermédiaires, ce qui faciliterait le pronostic précoce et la prévention de la survenue de maladies chroniques.

Hypothèses et objectifs

Des biomarqueurs génétiques et moléculaires impliqués dans plusieurs voies biologiques pourraient jouer un rôle important dans le développement de maladies chroniques. L'identification de voies biologiques communes et d'interactions partagées pourraient donc améliorer notre compréhension de la physiopathologie des maladies et nous aider à découvrir de nouveaux biomarqueurs, des « dénominateurs communs » pour les maladies et les comorbidités liées à l'âge.

Les études d'association pangénomique ont permis d'identifier un grand nombre de variants liés à des phénotypes spécifiques. Nous formulons l'hypothèse suivante : l'association des variants génétiques d'un biomarqueur d'intérêt à d'autres molécules impliquées dans la même maladie pourrait aider à identifier des voies moléculaires partagées entre les biomarqueurs et de nouveaux biomarqueurs génétiques qui serviraient à évaluer le risque de maladies apparentées.

Ces dernières années ont vu émerger les études épigénétiques qui augmentent grandement le potentiel d'approches de prédiction et de prévention personnalisées. Les études d'association panépigénomique (EWAS) permettent d'identifier des sites CpG méthylés qui jouent un rôle important dans la régulation des phénotypes intermédiaires. En présence de conditions

environnementales défavorables, la méthylation représente une des modifications précoces de l'ADN, les sites CpG liés à des facteurs de risque peuvent donc servir de biomarqueurs de susceptibilité pour toutes les maladies associées. En revanche, si le site CpG intervient dans la régulation d'un gène impliqué dans une maladie, cette méthylation pourrait éventuellement constituer une cible thérapeutique.

Lors de notre étude de biomarqueurs moléculaires, nous nous sommes concentrés sur différentes molécules impliquées dans des processus physiopathologiques courants, comme l'inflammation et l'angiogenèse. En plus des cytokines inflammatoires les plus courants, nous avons porté un intérêt particulier à TREM 2, au VEGF-A et à la longueur des télomères (LT). VEGF-A et LT sont des exemples parfaits de biomarqueurs impliqués dans différents mécanismes physiopathologiques. Nous avons donc supposé qu'analysés ensemble, ils pourraient constituer un outil puissant pour les applications de médecine personnalisée.

En outre, les collaborations entre les laboratoires de recherche et l'industrie représentent une part importante de la mise en œuvre progressive de la médecine personnalisée, la manière la plus rapide de transférer les connaissances fondamentales vers une application clinique. À cette fin, il est indispensable de participer activement aux projets des industriels cherchant à améliorer le diagnostic et le traitement de maladies chroniques courantes par des approches de médecine personnalisée.

Pour résumer, l'objectif de cette thèse était d'identifier des biomarqueurs communs à plusieurs maladies chroniques par l'étude des interactions entre des variants génétiques et des phénotypes intermédiaires au moyen de méthodologies « -omiques », afin de proposer un ensemble de biomarqueurs qui, lorsqu'analysés conjointement, pourraient améliorer la prédiction du risque de maladie et le traitement dans une stratégie globale de médecine personnalisée.

Les objectifs spécifiques établis pour le développement d'une approche de médecine personnalisée à plusieurs biomarqueurs sont :

- Résumer les avancées les plus marquantes dans le domaine de la médecine personnalisée.
- Étudier l'association entre les sites CpG de méthylation de l'ADN et les phénotypes intermédiaires tels que les taux de lipides sanguins et le tour de taille dans une

population d'individus sains, afin d'identifier les mécanismes sous-jacents qui relient la méthylation de l'ADN au métabolisme des lipides et l'obésité abdominale.

- Étudier l'association entre les sites CpG de méthylation de l'ADN et le taux de VEGF-A dans une population d'individus sains, afin d'identifier les mécanismes sous-jacents qui relient la méthylation de l'ADN à l'angiogenèse.
- Identifier des voies physiologiques communes entre le VEGF-A, TREM-2 et LT, et d'autres cytokines inflammatoires, au moyen de l'étude d'association.
- Évaluer le rôle de LT dans la médecine personnalisée, et son utilité sur les plans diagnostique, pronostique et thérapeutique.
- Se familiariser avec le milieu professionnel et la collaboration avec des partenaires de l'industrie du diagnostic clinique intéressés dans le transfert de résultats de recherche vers des outils de médecine personnalisée.

Discussion

Dans ce travail, nous avons exploré la méthylation de l'ADN et utilisé différentes méthodologies d'étude d'association afin d'identifier de nouveaux biomarqueurs et des voies moléculaires communes ayant un effet sur les phénotypes intermédiaires et les facteurs de risque de maladies chroniques. Nos résultats sont destinés à servir d'outils de prédiction pour contribuer à la découverte de nouvelles options thérapeutiques et à accroître les connaissances actuelles de la régulation moléculaire de processus biologiques communs. En cherchant à atteindre les objectifs définis, nous avons établi :

1. L'état des connaissances de la médecine personnalisée et la contribution importante apportée au système de santé actuel, ainsi que l'étendue de l'investissement scientifique nécessaire pour accompagner son développement et améliorer les programmes individualisés de soins.

2. Une association entre une augmentation du taux de TG (triglycérides) et les sites CpG cg08897188 et cg04580029. Le taux de TG est un facteur de risque intermédiaire important dans plusieurs maladies chroniques. Un site CpG lié au taux de TG pourrait être utilisé comme biomarqueur d'évaluation du risque ou intervenir dans l'élaboration de nouvelles stratégies thérapeutiques.

3. Une association entre un tour de taille élevé et le site CpG cg16170243, qui a permis de mieux comprendre la régulation complexe de l'obésité mais qui nécessite d'être étudiée davantage. Ce résultat pourrait se révéler utile lors de la recherche ultérieure de nouveaux facteurs de risque de l'obésité, ainsi que pour d'éventuelles applications thérapeutiques.

4. Une association entre le taux de VEGF-A et 20 sites CpG significatifs, ce qui souligne les mécanismes épigénétiques impliqués dans la régulation physiologique de ce biomarqueur angiogénique.

5. Des associations entre VEGF-A et IL-4, MCP-1 et EGF, entre l'isoforme $VEGF_{165}$ et MCP-1 et IL-1α, et entre l'isoforme $VEGF_{189}$ et IL-4 et IL-6. Ces associations fournissent un éclairage sur les liens entre les cytokines inflammatoires et le VEGF-A et permettent de mieux comprendre les interactions physiologiques entre l'angiogenèse et l'inflammation.

6. Une association entre rs6918289, un variant génétique lié à *TREM2*, et des valeurs élevées du taux de TNF-α et d'IMT-F, ce qui suggère que l'allèle mineur T de rs6918289 pourrait être considéré comme un facteur de risque de l'inflammation et de l'athérosclérose.

7. Une association entre rs6993770, un variant génétique lié au VEGF-A, et l'attrition des télomères (défini par LTL/LTM) dans une étude cas-contrôle sur l'athérosclérose, ce qui suggère une implication combinée de VEGF-A et LT dans l'étiologie des maladies cardiovasculaires.

8. L'état des connaissances actuelles portant sur l'utilité future des mesures de LT dans des applications cliniques de médecine personnalisée et le besoin d'études longitudinales débutant dès l'enfance. De telles études aideraient à comprendre la biologie des télomères au cours des premières années de vie, et permettraient d'évaluer l'impact de facteurs environnementaux pendant l'enfance sur LT chez les adultes et la susceptibilité par rapport à des maladies chroniques courantes.

Les maladies chroniques courantes, telles que les maladies cardiovasculaires, le cancer et le diabète de type 2, constituent la cause la plus fréquente de mortalité prématurée chez l'adulte (1). Elles ne se déclarent pas rapidement, mais surviennent petit à petit à force d'accumuler divers facteurs de risque environnementaux chez un individu au profil génétique défavorable (2). Nombreux de ces facteurs de risque pourraient être prévenus, et l'identification précoce des patients à risque, associée à des mesures de promotion de la santé appropriées, pourrait réduire la prévalence des maladies chroniques sans recours à des interventions thérapeutiques (3). La prévention des maladies chroniques ne doit pas débuter à la survenue de la maladie mais bien en amont en gérant et en luttant contre les expositions à effet délétère connu, d'où le besoin de nouveaux biomarqueurs de prédiction du risque (4).

Les phénotypes intermédiaires ou les traits quantitatifs correspondent aux manifestations causées par les mécanismes des maladies complexes et peuvent être utilisés lors d'études moléculaires et génétiques (5). Leur variation phénotypique est distribuée de manière continue dans les populations naturelles, en raison de la ségrégation des allèles au niveau de plusieurs loci et celle-ci est influencée par des effets génétiques et environnementaux complexes. La variation génétique quantitative peut expliquer la susceptibilité à des maladies complexes courantes et les variants génétiques impliqués peuvent être considérés comme des facteurs de risque potentiels (6, 7). De manière similaire, la variation épigénétique quantitative peut expliquer une partie de la variation des traits physiologiques et peut servir de biomarqueur de la susceptibilité (8).

Les phénotypes intermédiaires (appelés aussi facteurs de risque intermédiaires) sont l'expression de facteurs de risque modifiables, par exemple des taux élevés de cholestérol LDL et de TG pour l'évaluation d'une lipidémie anormale, un IMC et un tour de taille élevé pour l'évaluation de l'obésité, etc. Ces phénotypes sont d'un intérêt particulier pour la recherche dans le domaine de la médecine personnalisée car ils sont impliqués dans de nombreuses maladies chroniques et peuvent être évités en mettant en œuvre des mesures préventives appropriées (9, 10).

En médecine personnalisée, les biomarqueurs permettent d'évaluer les processus biologiques normaux, de détecter des processus anormaux et d'établir le pronostic d'une maladie (11, 12). À présent, l'identification de nouveaux biomarqueurs pour la médecine personnalisée nécessite non seulement une bonne compréhension de la maladie, mais aussi l'utilisation combinée d'approches génomiques et moléculaires afin de détecter des variants ayant une

valeur clinique prédictive. De nombreuses maladies partagent les mêmes mécanismes pathologiques, notamment l'inflammation et l'angiogenèse. Il est donc particulièrement important d'étudier ces processus. Les biomarqueurs identifiés dans de telles conditions sont susceptibles d'être impliqués dans plusieurs maladies et donc d'être plus performants pour le développement de stratégies thérapeutiques efficaces contre plusieurs comorbidités (13).

Dans ce travail de thèse, nous avons étudié les facteurs de risque intermédiaires et les voies biologiques impliquées dans un éventail de maladies chroniques. Nous avons mis en œuvre différentes méthodologies pour évaluer leur régulation et pour établir le lien entre les facteurs et la médecine personnalisée.

•

Il est aujourd'hui établi que l'exposition à des facteurs environnementaux indésirables peut induire des modifications de la méthylation de l'ADN de nombreuses années avant la survenue d'une maladie (14). Un biomarqueur de la méthylation de l'ADN démontré comme étant associé à un taux élevé de TG ou un tour de taille élevé pourrait, seul ou combiné à d'autres biomarqueurs, aider à identifier les personnes les plus susceptibles de présenter un risque accru de maladie chronique. Chez ces individus à risque, des mesures préventives précoces de gestion du mode de vie pourraient aider à limiter l'apparition de ces phénotypes intermédiaires (15).

De plus, contrairement à la séquence d'ADN, les modifications épigénétiques de l'ADN sont un des mécanismes clés de modulation de la transcription de l'ARNm et sont sensibles à des interventions environnementales et pharmacologiques. L'environnement peut modifier les profils épigénétiques qui jouent un rôle important dans la santé. Ces informations peuvent être transmises aux cellules filles lors de la réplication cellulaire (16). Des nombreuses substances thérapeutiques sont capables de modifier les profils de méthylation et de cibler des sites associés à un pronostic défavorable (17). Par conséquent, cette méthodologie ouvre la voie pour la découverte de nouvelles options thérapeutiques (18, 19).

Grâce à la méthodologie EWAS, les travaux menés au cours de cette thèse sur la population d'individus sains de la cohorte SFS ont permis d'identifier 2 nouveaux sites de méthylation associés au taux de TG, 1 nouveau site de méthylation associé au tour de taille, et 20 nouveaux sites qui expliquent la régulation épigénétique de l'expression du VEGF-A dans

des conditions physiologiques. Les résultats de ces trois études EWAS distinctes sont discutés ci-dessous.

Tout d'abord, nous avons étudié les profils de méthylation associés aux taux de cholestérol total, LDL, HDL et TG dans une population de 211 adultes et enfants sains de la cohorte SFS. Les résultats ont été publiés dans l'article intitulé **: « *Epigenome-Wide Association Study (EWAS) of Blood Lipids in Healthy Population from STANISLAS Family Study (SFS)* ».** La dérégulation de la lipidémie est un des facteurs de risque les plus importants pour les maladies cardiovasculaires (20). Les taux de lipides sont influencés par la génétique mais aussi par les expositions environnementales, qui peuvent modifier le profil de méthylation de l'ADN. La connaissance de ces profils épigénétiques pourrait expliquer la variabilité inter-individu des phénotypes lipidiques ainsi que les mécanismes de régulation sous-jacents (21). Différentes études ont été menées sur la régulation épigénétique des lipides chez des sujets présentant une maladie cardiovasculaire. La régulation de traits lipidiques peut varier de manière significative dans la population en bonne santé et il est vraisemblable que de nouveaux sites CpG impliqués dans la régulation des gènes pourront être identifiés et aideront à mieux comprendre les mécanismes complexes intervenant dans le métabolisme des lipides. Au moyen d'une approche sur l'ensemble de l'épigénome, nous avons identifié deux sites CpG associés de façon significative au taux de TG : cg08897188 (chromosome 7q36.1 ; β = -2,80 ; P = 1,39 × 10^{-8}) et cg04580029 (chromosome 16p13.3 ; β = 3,09 ; P = 5,75 × 10^{-9}).

Les modifications épigénétiques de l'ADN ont un effet régulateur important sur la différenciation cellulaire et il a été démontré que la méthylation de l'ADN pouvait être tissu-spécifique (22, 23). La méthylation de différents types de cellules pouvant varier de manière importante, cette étude a été préparée en prenant en considération le type de cellules utilisées, c'est-à-dire les cellules du sang, et d'autres ajustements techniques nécessaires (effets fixes, puce de méthylation).

Lors de la recherche de nouveaux biomarqueurs, la disponibilité du tissu est un point essentiel à prendre en considération. L'utilisation d'un tissu dont le prélèvement nécessite une intervention chirurgicale compliquerait la mise en place de l'approche sur le plan clinique. Il nous a donc semblé que les sites méthylés devaient être détectables dans l'ADN du sang et que l'étude EWAS de recherche de nouveaux biomarqueurs génétiques devait être effectuée sur l'ADN extrait du sang de la cohorte de découverte (24). Toutefois, pour mieux comprendre la régulation épigénétique d'un phénotype particulier, comme le taux de TG par

exemple, il a été convenu qu'un tissu lié au phénotype serait à utiliser également, si disponible (25), car cela pourrait apporter des données supplémentaires confirmant les mécanismes génétiques qui associent le site de méthylation au phénotype.

Le tissu adipeux étant étroitement lié à la biologie des TG (26), il a donc sélectionné comme source pour les analyses de méthylation de l'ADN dans la cohorte de réplication. En tout, des échantillons d'ADN provenant de tissu adipeux de 662 participants à l'étude MuTHER ont été testés pour l'association avec cg08897188 et cg04580029. L'association du site CpG 04580029 et du taux de TG a été confirmée, ce qui implique que la relation identifiée entre le site CpG et le taux de TG est également biologiquement plausible. En outre, la réplication dans le tissu adipeux du résultat démontré initialement dans le sang montre que l'étude de la méthylation de l'ADN dans le sang total est une approche pertinente pour l'identification de modifications épigénomiques, et confirme les conclusions d'études antérieures (27, 28).

Deux gènes ont été identifiés à proximité de la sonde cg08897188. *PRKAG2*, qui code pour la sous-unité fixant l'AMP/ATP de la protéine kinase activée par l'AMP (AMPK), et un transcrit non codant du gène *AC093583.1*, *LOC644090*. Le site Cg04580029 est localisé dans le promoteur du gène *KREMEN2* qui code pour dickkopf1 (Dkk1), un antagoniste sécrété de la voie de signalisation Wnt/β-caténine. Ces deux sites CpG associés au taux de TG étaient localisés dans un environnement génétique fortement régulé. L'identification de sites de méthylation significative à proximité d'une région de modification importante des histones est particulièrement intéressante car les modifications des histones sont un des composants clé de la régulation transcriptionnelle (29). Ces modifications sont à l'origine de l'organisation du génome en régions actives (l'ADN devient accessible à la transcription) et inactives (l'ADN est compact est moins accessible à la transcription ; elles participent donc à la régulation de l'expression génique (30). La méthylation à proximité d'une histone est donc susceptible d'être impliquée de manière active dans la régulation transcriptionnelle de la région.

Pour conclure, plusieurs études épidémiologiques, cliniques et génétiques ont démontré que des taux élevés de TG constituent un biomarqueur du risque de maladie cardiovasculaire (31). Les triglycérides contribuent au durcissement des artères et à l'épaississement de la paroi artérielle et donc augmentent le risque de coronaropathie (32), d'accident vasculaire cérébral (AVC), d'infarctus et de cardiopathie (33). De plus, des concentrations élevées en TG sont également impliquées dans d'autres maladies comme le diabète, l'insuffisance rénale chronique et le syndrome métabolique (34). Nos travaux de recherche de sites CpG associés

de manière significative au taux de TG a permis d'identifier deux biomarqueurs potentiels qui pourraient faciliter l'évaluation précoce du risque de maladie cardiovasculaire, ou qui pourrait servir de cible pour de nouvelles stratégies thérapeutiques basées sur la méthylation qui seront probablement largement développées dans l'avenir.

Une approche similaire a été employée pour étudier les profils de méthylation associés au tour de taille et au rapport taille/hanches (ajusté pour l'IMC). Les résultats de ces travaux ont été publiés dans l'article intitulé : **« *Epigenome-Wide Association Study Detects a Novel Loci Associated with Central Obesity in Healthy Subjects* ».** La population d'étude était composée de 211 adultes et enfants sains de la cohorte SFS. L'obésité est une accumulation excessive de graisses qui augmente le risque de plusieurs comorbidités, notamment le diabète de type 2, les maladies cardiovasculaires et le cancer (35). De nombreux effets néfastes sur la santé ont été associés à un excès de tissu adipeux autour de la partie centrale du corps, appelé obésité abdominale, et estimé habituellement en mesurant le tour de taille (36). L'obésité est un trait complexe dont la régulation n'est toujours pas pleinement établie. Malgré de nombreux travaux pour élucider les mécanismes de régulation épigénétique de l'obésité, de nouvelles cohortes doivent être constituées pour une entière compréhension de ce phénomène. L'utilisation de populations d'individus sains peut aider à éviter les divergences générées par les cas extrêmes d'obésité et de comorbidités dans les cohortes de populations spécifiques (37). Cette étude a permis d'identifier une association positive significative, jusqu'à-là inconnue, entre le tour de taille ajusté pour l'IMC et le site de méthylation cg16170243 (chromosome 18q21.1 ; $\beta = 2{,}32$; $P_{adj} = 0{,}048$).

Deux mécanismes peuvent expliquer l'association entre cg16170243 et le tour de taille. Premièrement, des facteurs environnementaux non spécifiques pourraient induire la méthylation du site cg16170243, ce qui entraînerait à son tour une augmentation du tour de taille et une susceptibilité accrue vis-à-vis de l'obésité abdominale. Deuxièmement, un environnement obésogène pourrait être à l'origine de la méthylation du site cg16170243, ce qui se traduirait par l'association de ce site de méthylation avec le tour de taille. Les deux mécanismes sont tout aussi possibles et sont discutés dans ce manuscrit, bien que des hypothèses plus détaillées aient été développées dans l'article pour expliquer le premier mécanisme.

Brièvement, le gène *ST8SIA5* dans lequel est localisée la sonde cg16170243 code pour une enzyme sialyltransférase impliquée dans la synthèse de gangliosides. Les gangliosides

interagissent avec des molécules des voies de transduction du signal, telles que les récepteurs à activité tyrosine kinase (RTK). Une modification de la composition en gangliosides induit la dissociation des RTK des microdomaines enrichis en glycolipides, ce qui pourrait aboutir à une phosphorylation réduite des récepteurs et une résistance à l'insuline. Ainsi, la méthylation de ce site CpG pourrait modifier le niveau d'expression du gène *ST8SIA5*, ce qui aurait une répercussion sur la synthèse des gangliosides, entraînant à son tour une résistance accrue à l'insuline qui est un facteur de risque connu pour le développement de l'obésité (38).

Le deuxième mécanisme suppose que l'adiposité aurait un effet sur la méthylation de l'ADN, comme discuté en détail par Wahl *et al.* (39). Dans leur étude, ces auteurs ont démontré que l'adiposité entraînait une modification de la méthylation de la plupart des sites CpG identifiés comme étant associés à l'IMC. La méthode employée pour évaluer la causalité de cette relation reposait sur un score de risque génétique pondéré qui combinait les effets connus pour influencer l'IMC sur l'ensemble des SNP. Dans notre étude, l'association entre les polymorphismes associés au tour de taille et la méthylation de cg16170243 a été testée, mais aucune association significative n'a été constatée. Toutefois, nous ne pouvons pas exclure l'hypothèse selon laquelle la méthylation du site cg16170243 soit due à une adiposité accrue. Des travaux supplémentaires sont nécessaires pour déterminer la nature du lien entre le site CpG identifié et le tour de taille.

Le rôle de la méthylation de l'ADN en tant que médiateur du lien de causalité entre des expositions variables et le risque de maladie peut être évalué par randomisation mendélienne à deux échantillons à l'aide des données statistiques résumées de l'EWAS pour le phénotype observé. Dans une première étape, l'effet causal du phénotype observé sur les marqueurs épigénétiques est établi. Lors d'une deuxième étape, la nature causale des marqueurs épigénétiques sur un état de santé est évaluée (40). Une telle approche pourrait permettre d'identifier les loci pertinents pour des études fonctionnelles (24).

Le troisième volet de ces travaux analysant la méthylation pangénomique s'est attaché à étudier la régulation épigénétique du VEGF-A dans une population de 211 adultes et enfants sains de la cohorte SFS. Les résultats ont été publiés dans l'article intitulé **« *Epigenome-wide association study in healthy individuals identifies significant associations with DNA methylation and PBMCs extracts VEGF-A concentration* »**. Le VEGF-A est un médiateur important de l'angiogenèse et de l'inflammation, ainsi qu'une cible biologique clé pour le traitement de plusieurs maladies où ces phénomènes sont impliqués dans le processus

pathologique (41-43). Le VEGF-A présente une héritabilité élevée (> 60 %, telle que déterminée lors de travaux antérieurs sur la cohorte SFS (44)), qui peut être expliquée en partie par 10 SNP associés de manière significative au VEGF-A (45, 46). Bien que sa régulation génétique ait déjà été étudiée au moyen d'études d'association pangénomique (GWAS), les résultats de la présente étude sont à notre connaissance les premiers à mettre en évidence la régulation épigénétique du VEGF-A par EWAS.

Nous avons identifié 20 sites CpG associés de manière significative au taux de VEGF-A dans des extraits de cellules mononuclées du sang périphérique (PBMCs) et nous avons discuté, dans l'article, des liens possibles avec les gènes à proximité des sites méthylés. De manière surprenante, aucun site CpG n'a été identifié dans un site de régulation d'un gène directement impliqué dans la biologie du VEGF-A, tel que *VEGF-A* (6p21.1), *FLT-1* (13q12.3), *KDR* (4q12), *HIF1A* (4q23.2) etc. Ceci pourrait être expliqué par une méthylation de sites non spécifiques induite par les taux élevés de VEGF-A. Cependant, nous avons identifié des sites de méthylation proches de gènes qui pourraient être associés indirectement à la régulation du VEGF-A. Les deux résultats les plus intéressants étaient les sites CpG cg21838233 (gène *TPX2*) et cg06785213 (gène *HAS1*). Le gène *TPX2* est impliqué dans la mise en place du fuseau mitotique, et est souvent surexprimé dans les cellules de cancer gastrique (47). Des études ont montré que l'inhibition de l'expression de *TPX2* peut empêcher la prolifération et l'invasion des cellules de cancer de côlon, et que ceci est peut-être lié à une baisse de l'expression du gène *VEGF-A* (48). Le gène *HAS1*, quant à lui, code pour la synthèse d'acide hyaluronique qui joue un rôle essentiel dans le développement tissulaire et l'homéostasie des protéines de la membrane plasmique. Il est également impliqué dans des arthropathies inflammatoires et dégénératives. Il a été démontré que des ARNsi ciblant le gène *HAS1* réduisent l'expression de *VEGF-A* (49).

En plus des deux gènes présentés ci-dessus, des liens avec le VEGF-A ont été trouvés pour 6 autres gènes à proximité des sites CpG identifiés. La méthylation pourrait expliquer certaines modifications de l'expression du *VEGF-A* observées dans des études antérieures mais à ce jour ells sont inexpliquées. Dans l'avenir, l'exploration de la méthylation de l'ADN associée au VEGF-A devrait cependant être axée sur des populations spécifiques, tels que les patients atteints de cancer ou de maladies cardiovasculaires. Chez les patients atteints de cancer, la production de VEGF-A est augmentée afin de déclencher le processus d'angiogenèse et permet de vasculariser le tumeur en développement (50). L'étude du rôle de la méthylation de l'ADN dans la régulation de la surexpression du *VEGF-A* pourrait offrir des

perspectives intéressantes. Associées aux résultats EWAS obtenus pour une population en bonne santé, les données d'une telle étude pourraient permettre d'identifier des sites de méthylation différentielle et peut-être contribuer au développement de stratégies alternatives aux traitements anti-angiogéniques actuels.

Dans cette étude, les taux de VEGF-A ont été mesurés dans des extraits cellulaires de PBMCs, et non dans le sérum ou le plasma, qui sont les sources biologiques les plus souvent utilisées pour la détermination du VEGF-A. Les PBMCs sont une source de VEGF-A et de nombreuses autres molécules de signalisation. Leur rôle important dans plusieurs voies biologiques a été établi dans de nombreuses études (51). La production du VEGF-A dans les cellules mononuclées est donc en lien direct avec les processus biologiques dans lesquels il est impliqué (angiogenèse et inflammation).

Comme dans toutes les discussions concernant les résultats d'études EWAS, il faut garder à l'esprit que la méthylation pourrait être la cause mais aussi la conséquence de taux élevés de VEGF-A. Cela pourrait être établi par randomisation mendélienne. Des avancées supplémentaires en épigénétique sont nécessaires pour le développement de nouvelles méthodologies permettant de déterminer de façon plus simple la causalité entre les sites CpG détectés et le phénotype. Cela améliorerait considérablement notre compréhension du rôle des marqueurs épigénétiques et accélérerait la mise en application des résultats sur le plan diagnostique et thérapeutique.

•

En raison de l'implication important du VEGF-A dans la physiopathologie des maladies au moyen de deux processus pathologiques majeurs (et interdépendants), l'angiogenèse et l'inflammation (52), l'association entre les taux de VEGF-A mesurés dans les extraits de PBMCs et 11 cytokines, des indicateurs importants de l'inflammation, a été étudiée. Les relations entre les deux processus physiologiques ont été décrites dans l'article intitulé **« *Peripheral blood mononuclear cell extract VEGF protein levels and VEGF mRNA. Associations with inflammatory molecules in a healthy population* ».** L'objectif premier de ces travaux était d'identifier les molécules inflammatoires qui pourraient être utilisées conjointement avec le taux de VEGF-A dans des applications prédictives ou thérapeutiques de médecine personnalisée.

Comme indiqué dans l'article suscité, une des originalités de l'étude était d'utiliser les extraits de PBMCs comme source pour toutes les mesures des molécules analysées et des isoformes du VEGF. Les PBMCs sont des cellules qui répondent de façon sélective à l'activation du système immunitaire et jouent un rôle majeur dans l'immunité chez l'homme en synthétisant et en libérant des cytokines inflammatoires et des facteurs de croissance (53, 54). La production simultanée de différents médiateurs par un même groupe de cellules semble suggérer une activité conjointe dans le même mécanisme physiopathologique.

L'association entre à la fois la concentration protéique de VEGF-A et le niveau d'expression des ARNm des quatre isoformes les plus abondantes du VEGF-A (*$VEGF_{121}$*, *$VEGF_{145}$*, *$VEGF_{165}$* et *$VEGF_{189}$*) et les molécules inflammatoires IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, INF-γ, TNF-α, MCP-1 et EGF a été évaluée. Les interactions avec l'expression protéique du VEGF-A ont été étudiées chez 285 individus de la cohorte SFS, alors qu'un sous-groupe de 110 individus a été utilisé pour l'étude d'association avec les ARNm des isoformes du gène *VEGF*. Ces travaux ont permis d'identifier des associations entre le VEGF-A et IL-4, MCP-1 et EGF, entre l'isoforme *$VEGF_{165}$* et MCP-1 et IL-1α, et entre l'isoforme *$VEGF_{189}$* et IL-4 et IL-6. Des recherches bibliographiques approfondies ont été menées pour expliquer les liens entre les cytokines identifiées et le VEGF-A.

Pour résumer, IL-4, une cytokine plurifonctionnelle ayant des activités anti-inflammatoires et anti-tumorales, était associée au taux de la protéine VEGF-A et à l'expression de l'ARNm de l'isoforme *$VEGF_{189}$*. Il a déjà été démontré que l'IL-4 est liée au VEGF-A en tant que facteur important de recrutement des macrophages associés aux tumeurs, connus pour promouvoir l'angiogenèse (55). En outre, une expression accrue de l'ARNm du gène *VEGF-A* a été mise en évidence dans des cellules musculaires lisses des voies respiratoires humaines lorsque stimulées par l'IL-4 en culture cellulaire (56). Dans cette étude, nous avons identifié l'isoforme *$VEGF_{189}$* comme l'isoforme principale responsable de l'association entre le VEGF-A et l'IL-4. Néanmoins, ce résultat doit être confirmé par d'autres études.

MCP-1, une chimiokine angiogénique impliquée dans la régulation, la migration et l'infiltration des monocytes et des macrophages, était associée aux niveaux d'expression de la protéine VEGF-A et de l'ARNm de l'isoforme *$VEGF_{165}$*. Ces résultats vont dans le même sens que ceux d'études antérieures qui ont montré que MCP-1 augmentait l'expression du gène *HIF1A* dans des cellules endothéliales d'aorte humaine, ce qui augmentait aussi l'expression du *$VEGF_{165}$* dans la paroi aortique (57).

Une autre molécule inflammatoire, l'EGF, était associée au taux de protéine VEGF-A dans notre population. L'EGF est un médiateur important de la croissance, de la prolifération et de la différenciation de nombreux types cellulaires. Il a été démontré dans différents types de cellules que ces deux facteurs de croissance, EGF et VEGF, jouent des rôles majeurs dans la tumorigénèse ainsi plusieurs mécanismes d'action conjointe ont été proposés. Le ciblage combiné des voies de signalisation du VEGF et de l'EGF a été étudié dans le but de développer de nouvelles approches thérapeutiques de lutte contre le cancer avec des résultats encourageants pour le développement de nouveaux médicaments (58-61).

Enfin, IL-6 et IL-1α étaient associées respectivement à l'expression des isoformes $VEGF_{189}$ et $VEGF_{165}$. Il a déjà été montré, lors d'études antérieures en culture cellulaire, que l'IL-6 et l'IL-1α augmentaient l'expression de l'ARNm du *VEGF-A*. IL-6 promeut l'angiogenèse médiée par le VEGF-A dans les tumeurs du col de l'utérus, les carcinomes gastriques (62) et les mésothéliomes malins (63). L'IL-1α stimule la synthèse de l'ARNm du *VEGF-A* dans les PBMCs de manière dose-dépendante. Dans des cellules de cancer du pancréas et dans des cellules d'adénome, les deux cytokines, IL-6 et IL-1α (64), participent ensemble à la régulation du VEGF-A (65).

Dans le corps humain, les voies physiologiques constituent des réseaux complexes régulés par de nombreuses interactions entre molécules. L'angiogenèse, qui permet le développement de nouveaux vaisseaux sanguins, est le principal processus médié par le VEGF-A. Toutefois, d'autres mécanismes, notamment l'inflammation, sont étroitement liés à sa régulation (66, 67). Les interactions entre ces deux mécanismes peuvent donner lieu à de nombreuses maladies chroniques courantes et doivent être considérées comme des cibles potentielles pour leur prévention.

•

La troisième approche méthodologique que nous avons mise en œuvre lors de cette thèse était l'étude des déterminants génétiques permettant d'évaluer les facteurs de risque intermédiaires et les voies biologiques impliquées dans différentes maladies chroniques. Les caractéristiques de la variation génétique quantitative ont été utilisées pour identifier des biomarqueurs génétiques destinés à prédire le risque d'athérosclérose, un processus pathologique commun à la plupart des maladies cardiovasculaires (68).

Dans l'article intitulé **« *The polymorphism rs6918289 located in the downstream region of the TREM2 gene is associated with TNF-alpha levels and IMT-F* »**, nous avons analysé deux traits quantitatifs, le taux de TNF-α et IMT-F (épaisseur intima-média fémorale), tous deux des indicateurs importants de l'athérosclérose. Puisqu'il a été récemment démontré que *TREM2* est exprimé dans les lésions d'athérosclérose chez l'homme (69), l'objectif de notre étude était d'étudier les variants génétiques localisés à proximité du gène *TREM2* et de rechercher une association avec le taux de TNF-α et l'IMT-F.

Tout d'abord, nous avons testé cinq SNP (rs7748777, rs6918289, rs7759295, rs9357347, rs6915083) localisés dans la région du gène *TREM2* (50 kb en amont et en aval du gène *TREM2*, chromosome 6p21) pour rechercher une association avec le taux de TNF-α chez 139 enfants de la cohorte SFS. Nous avons identifié un nouvel allèle à risque potentiel (T) pour le SNP rs6918289, qui était associé à des taux plasmatiques de TNF-α élevés ($\beta_{découverte} = 0,33$; $P_{découverte} = 0,00147$).

Pour augmenter la taille de la population de découverte, le SNP rs6918289 a été génotypé *de novo* chez 393 adultes et 277 enfants. Lorsque analysé dans ces deux sous-populations de la cohorte SFS, nous avons démontré une association significative entre l'allèle T du SNP rs6918289 et des taux plasmatiques élevés de TNF-α ($\beta_{SFS} = 0,49$; $P_{SFS} = 0,0017$). Ces résultats ont été confirmés dans une cohorte indépendante de 916 individus de nationalité française ($\beta = 0,202$; $P = 0,023$).

Une étude d'association a également été menée avec IMT-F dans un échantillon de 350 adultes issus de la cohorte SFS pour qui des mesures IMT-F étaient disponibles. Les résultats de l'étude ont montré une association significative avec l'allèle mineur (T) du SNP rs6918289 ($\beta = 0,024$; $P = 0,026$).

Bien que les premières études sur la protéine TREM2 portaient essentiellement sur les réponses inflammatoires médiées par les macrophages et les neutrophiles dans des maladies neurodégénératives comme la maladie d'Alzheimer, la sclérose latérale amyotrophique (maladie de Charcot) et la maladie de Parkinson (70), des résultats récents ont également démontré un rôle important pour la réponse des macrophages induite par TREM2 dans d'autres maladies (69). L'article suscité a apporté les premiers éléments éclaircissant l'association génétique de *TREM2* avec TNF-α et IMT-F, des indicateurs importants de la formation des plaques d'athérosclérose.

IMT-F est un marqueur précoce et non invasif de l'athérosclérose et aussi un facteur de prédiction indépendant des maladies cardiovasculaires qui pourrait faciliter la détection précoce de plaques d'athérosclérose par échographie (par mesure de la distance entre l'interface média-adventice et l'interface intima-lumière) (71-73). Le TNF-α intervient dans l'athérogenèse à travers différents mécanismes au moyen desquels cette cytokine inflammatoire agit sur l'endothélium (74). Le TNF-α peut augmenter la perméabilité vasculaire pour permettre l'infiltration de cellules inflammatoires (75) et promeut la migration de cellules immunitaires dans le tissu pathologique ou endommagé afin de le protéger de pathogènes potentiellement nocifs. Il peut inhiber la vasodilatation dépendante de l'endothélium en réduisant la formation et la biodisponibilité du NO, stimuler la production de l'anion superoxyde au niveau vasculaire et interférer avec les voies de signalisation de l'insuline dans les cellules endothéliales pour induire une résistance à l'insuline (76).

Compte-tenu de ce qui précède, les résultats de cette étude suggèrent que l'allèle mineur (T) de rs6918289, qui est associé à un taux de TNF-α et une valeur IMT-F élevés, pourrait constituer un allèle à risque pour les maladies inflammatoires et l'athérosclérose. Le mécanisme permettant d'expliquer ceci est encore inconnu. Toutefois, l'analyse bioinformatique de la région génétique a permis de formuler trois hypothèses concernant l'association entre rs6918289 et un taux de TNF-α élevé.

Premièrement, rs6918289 est impliqué dans le régulation post-transcriptionnelle et est associé à la protéine de liaison à l'ARN *PABPC1* (77), qui lie la région 3' poly(A) des ARNm. L'allèle mineur (T) de rs6918289 pourrait avoir un effet sur la protéine *PABPC1* qui à son tour affecterait la stabilité de l'ARNm de *TREM2*. Des études antérieures ont démontré que le niveau d'expression de *TREM2* peut avoir un effet sur les concentrations de TNF-α (78-80). L'effet de l'allèle mineur (T) conduirait à une production moindre de *TREM2*, ce qui entraînerait une augmentation du taux de TNF-α chez le même individu.

La deuxième hypothèse est en lien avec la localisation de rs6918289 entre deux séquences génomiques de liaison de CTCF (protéine à 11 motifs doigts de zinc) qui peuvent se lier ensemble et former une boucle de chromatine. Cette section de chromatine devient alors un domaine d'association topologique (TAD) qui peut avoir un effet sur l'expression génique et d'autres fonctions cellulaires. Nous formulons donc l'hypothèse que l'allèle mineur (T) du SNP rs6918289 promeut la fixation de CTCF et par conséquent déclenche une action insulatrice du gène *TREM2*.

Enfin, rs6918289 se trouve à proximité du gène Triggering Receptor Expressed On Myeloid Cells Like 1 (*TREML1*). Des analyses *in silico* ont mis en évidence une corrélation significative entre rs6918289 et sept SNP associés au gène *TREML1*. De ce fait, *TREML1* pourrait également être un médiateur de l'effet de rs6918289 sur la concentration de TNF-α.

Le remodelage d'une artère est étroitement dépendant de l'état inflammatoire. L'association entre rs6918289 et IMT-F est donc très probablement due à un taux élevé de TNF-α qui découle de l'effet de l'allèle mineur (T) du SNP rs6918289, ce qui suggère un rôle indirect de rs6918289 sur IMT-F.

Les cinq variants génétiques étudiés ont été sélectionnés sur la base de leur proximité avec le gène *TREM2* et non d'après une association identifiée auparavant lors d'études GWAS. Une telle décision se justifiait parce que tous les SNP n'étaient pas disponibles dans notre population génotypée. Nous avons choisi les variants génétiques d'après leur localisation qui présupposait qu'ils pourraient être impliqués dans la régulation du gène *TREM2*. D'autres travaux dans lesquels des variants *TREM2* identifiés lors d'études GWAS seraient utilisés pour explorer l'association entre TREM2, TNF-α et IMT-F sont nécessaires. Toutefois, une étude de réplication dans du tissu adipeux avec une population d'étude plus grande doit d'abord être effectuée pour confirmer nos résultats et évaluer leur éventuelle utilité clinique.

•

L'étude des variants génétiques a été poursuivie pour explorer l'association entre deux biomarqueurs, la longueur des télomères (LT) et le VEGF-A. L'objectif de ces travaux intitulés **« *Discovering the Relation between Telomere Length and VEGF-A* »** était d'évaluer l'association entre la longueur des télomères des leucocytes (LTL), la longueur des télomères des cellules musculaires (LTM) et l'attrition des télomères (calculée comme LTL/LTM) et 10 variants génétiques liés au VEGF-A (rs10761741, rs10738760, rs6921438, rs7043199, rs6993770, rs4416670, rs114694170, rs34528081, rs4782371 et rs2639990) (45, 46). LTL représente le longueur des télomères dans des cellules hématopoïétiques à prolifération rapide, alors que LTM correspond à la longueur dans un tissu musculaire squelettique à faible taux de réplication, et donc avec une LT comparable à la valeur initiale (81).

Le raccourcissement de la LT a été associé à l'athérosclérose et est considéré comme un facteur de risque possible pour les maladies cardiovasculaires (82). De manière similaire, il a

été démontré que le VEGF-A potentialisait le mécanisme physiopathologique de formation et de déstabilisation des plaques d'athérosclérose (83). Outre leur rôle important dans l'athérogenèse, ces deux biomarqueurs sont impliqués dans d'autres maladies chroniques courantes. L'étude TELARTA, actuellement toujours en cours, repose sur la création d'une cohorte cas-contrôle pour explorer les rôles du VEGF-A et de LT dans l'athérosclérose. Cependant, les résultats de cette étude pourront être extrapolés et appliqués à d'autres maladies où ces deux biomarqueurs sont impliqués, notamment le cancer.

Ces travaux ont permis d'identifier des liens entre l'allèle mineur (T) du SNP rs6993770 et le rapport LTL/LTM (P = 0,001143). Cela suggère l'existence d'un mécanisme de régulation génétique commune pour le VEGF-A et l'attrition des télomères, possiblement au moyen d'un processus qui agit sur les deux biomarqueurs. Le SNP rs6993770 est l'un des variants les plus significatifs associés à un taux élevé de VEGF-A circulant. Ensemble avec trois autres SNP (rs6921438, rs4416670 et rs10738760), ces associations permettent d'expliquer 48 % de l'héritabilité du taux sérique de VEGF (46). Il a également été établi que le SNP rs6993770 est associé à une variation du taux de cholestérol HDL (84), au nombre d'érythrocytes, de plaquettes, et à la concentration en IL-12 (85).

Dans le sang, le VEGF est présent dans le plasma, les plaquettes et les leucocytes (86). Plusieurs études ont signalé une corrélation entre la concentration en VEGF-A et les plaquettes. Ces dernières sont particulièrement importantes dans la cicatrisation et pourraient jouer un rôle de stimulation de la croissance tumorale dépendante de l'angiogenèse par l'intermédiaire de leur fonction de transporteurs du VEGF-A (87, 88).

Le SNP rs6993770 est localisé dans l'intron du gène *ZFPM2-201* (protéine à motifs doigt de zinc, membre 2 de la famille FOG). Les protéines FOG modulent l'action des protéines de la famille GATA qui sont des régulateurs importants de l'hématopoïèse et de la cardiogenèse (89). Le variant génétique rs6993770 pourrait avoir un effet sur l'activité du gène *ZFPM2-201* qui à son tour modifierait la régulation de l'hématopoïèse par une action sur les protéines GATA. Une augmentation de l'hématopoïèse pourrait entraîner une hausse du taux de renouvellement des cellules et donc à une accélération de l'attrition des télomères (90). De plus, une augmentation de l'hématopoïèse conduirait à une production accrue de plaquettes, ce qui expliquerait le taux plus élevé de VEGF-A. Ces hypothèses semblent vraisemblables car une accélération de l'attrition des télomères était le seul phénotype associé de manière

significative à ce variant génétique, alors qu'aucune association significative n'était détectée avec LTL et LTM.

La grande variabilité inter-individu de la longueur des télomères est connue, et est observable dès la naissance (91). Plus tard dans la vie de nombreux facteurs peuvent accélérer l'attrition (taux de renouvellement cellulaire élevé causé par l'inflammation, le stress oxydatif, *etc*. (92)), ou avoir un effet bénéfique sur LT (activité physique régulière ou alimentation saine (93)). Néanmoins, les effets de ces facteurs sur LT sont relativement faibles par rapport aux variations importantes entre les individus de la valeur initiale de LT (94). Pour ces raisons, LTL pourrait être utilisée comme biomarqueur sans comparaison à la valeur LT initiale à la naissance. L'analyse combinée de LT et de l'attrition des télomères avec le temps apporterait probablement plus d'informations que la mesure de LT seule (95). Il est assumé qu'une attrition accélérée de la longueur des télomères d'un individu représente un biomarqueur qui décrit plus précisément les processus pathologiques.

Les valeurs LTL réduites observées chez les patients atteints d'athérosclérose ont été attribuées à une hausse du renouvellement cellulaire causée par l'inflammation systémique chronique (96). Une étude récente a d'ailleurs développé cette hypothèse et a suggéré qu'une attrition accélérée des télomères est associée à l'athérosclérose, plutôt qu'à des valeurs LTL plus courtes. Cette étude a démontré que les différences entre les patients atteints d'athérosclérose et les contrôles étaient uniformes sur toute la plage d'âges de la population d'étude. De ce fait, les auteurs ont suggéré qu'une attrition accrue tôt dans la vie est susceptible d'être une des causes principales du raccourcissement de la LTL chez les patients atteints d'athérosclérose (81). Nos résultats semblent peut-être s'aligner avec ceux de l'étude précédente. Le SNP rs6993770 augmente l'hématopoïèse et donc également l'attrition des télomères dans les leucocytes, et ce dès l'enfance. Le raccourcissement des télomères est lié à une instabilité chromosomique qui entraîne l'apoptose. Un apoptose continu peut aboutir à une sénescence réplicative précoce du tissu vasculaire et donc aux premières étapes de formation d'une plaque d'athérosclérose.

Bien que cette hypothèse semble plausible, aucune association n'a été détectée entre la distribution de l'allèle à risque (T) de rs6993770 et le degré d'athérosclérose dans notre population d'étude. Nous pourrions donc supposer que rs6993770 représente un facteur de risque pour une attrition accélérée des télomères, qui à son tour entraîne l'athérosclérose. Toutefois, son rôle de facteur de prédiction des maladies cardiovasculaires doit être examiné

plus en détail. L'association de rs6993770 à l'attrition des télomères dans des études cas-contrôles d'autres maladies liées aux télomères, comme le cancer, doit également être évaluée.

•

La longueur des télomères est en effet un biomarqueur précieux lié à un grand nombre de maladies, allant de maladies monogéniques rares aux maladies chroniques courantes (97-99). Le raccourcissement continu de la longueur des télomères tout au long de la vie a lié LT à la longévité et a intrigué les scientifiques à la recherche d'une horloge biologique permettant de prédire l'espérance de vie (100). Son rôle dans la prédiction, le diagnostic et le traitement de maladies a été étudié au cours des dernières décennies, mais sans encore d'application clinique prometteuse (101).

Nous croyons que les télomères recèlent encore de nombreux secrets et nous avons consacré deux articles de revue à l'état actuel des connaissances sur la longueur des télomères et leur potentiel. Plus important, ces deux revues nous ont permis de souligner les nombreuses lacunes de connaissances qui nous empêchent de comprendre complètement la biologie des télomères et, si elles étaient comblées, pourraient contribuer à la réussite de la médecine personnalisée. Nous concluons que LT pourrait constituer un biomarqueur prospectif, mesuré systématiquement dès la naissance. En raison de la grande variabilité inter-individu, une mesure simple de LT ne serait pas représentative, alors que la détection d'une attrition accélérée pourrait apporter des informations précieuses et alerter sur d'éventuelles modifications pathologiques chez un individu. Des études longitudinales à long terme portant un intérêt particulier à la période de l'enfance seront nécessaires pour tester cette hypothèse.

•

En plus de tous les résultats décrits précédemment, nous avons participé activement, pendant toute la durée de ces travaux de thèse, au projet européen MAST4HEALTH, un programme d'échange de personnel de recherche et d'innovation faisant partie des actions Marie Skłodowska-Curie (MSCA RISE) du programme-cadre HORIZON 2020. Notre rôle dans ce projet était de déterminer les modifications de l'expression de biomarqueurs inflammatoires, tels que VEGF-A ($VEGF_{145}$, $VEGF_{121}$, $VEGF_{165}$, $VEGF_{165b}$, $VEGF_{189}$), TNF-α, et IL-6, chez des patients atteints de NAFLD traités par la plante grecque Pistacia lenticus (arbre à mastic), connue pour les propriétés antioxydantes, anti-inflammatoires et de réduction du taux de

lipides. Durant mon temps de travail chez Randox Laboratoires, une société qui propose des outils de diagnostic *in vitro*, nous avons développé et caractérisé des fragments d'anticorps monoclonaux qui serviront à mesurer de manière précise les biomarqueurs inflammatoires dans le sérum humain. Nous estimons que de telles collaborations de mise en application entre les laboratoires de recherche et l'industrie représentent la voie la plus rapide pour passer des découvertes en recherche à une application réussie en pratique clinique.

Conclusion et perspectives

Pendant ces trois années de travaux de recherche, différentes méthodologies ont été mises en œuvre pour explorer les facteurs de risque et voies moléculaires des maladies chroniques afin de répondre au besoin de nouveaux biomarqueurs utilisables en médecine personnalisée. Nous avons contribué aux connaissances actuelles qui viennent renforcer l'expérience déjà conséquente et le savoir de l'unité de recherche UMR INSERM U1122 IGE-PCV, tout en intégrant les approches les plus récentes d'étude de l'épidémiologie génétique de maladies chroniques courantes.

Le VEGF-A est un des biomarqueurs principaux ciblé par les travaux de recherche de notre unité qui a créé et dirige le Consortium VEGF international (http://www.vegfconsortium.org/) destiné à mettre en place un réseau de collaboration transnational dédié à de grandes études génomiques intégratives et pluridisciplinaires du VEGF. Parmi les projets collaboratifs nés de ce Consortium se trouvent deux grandes études GWAS qui ont permis d'identifier 10 variants génétiques associés au VEGF-A. Ces travaux de thèse ont approfondi les résultats de ces études et ont conduit à 1) l'identification de nouvelles associations avec des molécules inflammatoires, 2) la démonstration d'une association génétique avec l'attrition des télomères, et 3) les premières explications quant à sa régulation épigénétique.

Plus précisément, les travaux ont porté sur :

- L'étude des associations de plusieurs molécules qui interviennent avec le VEGF-A et ses différentes isoformes dans les processus d'inflammation et d'angiogenèse. Les résultats sont destinés à contribuer au développement de nouvelles stratégies thérapeutiques contre le VEGF-A, qui pourront être améliorées encore en ciblant des isoformes spécifiques et les cytokines qui leur sont associées.

- La recherche de liens entre la longueur des télomères (LT) et le VEGF-A. Un variant génétique a été identifié et offre de nouvelles hypothèses quant à l'étiologie de maladies chroniques liées aux télomères.
- L'identification de mécanismes de régulation épigénétique du VEGF-A. Il s'agissait de la première étude pangénomique cherchant à mettre en évidence des mécanismes de méthylation impliqués dans la biologie du VEGF-A.

Les études épigénétiques ouvrent une nouvelle ère dans le domaine de l'épidémiologie génétique en offrant un aperçu des mécanismes d'action des facteurs de risque génétiques et environnementaux dans la physiopathologie des malades, des biomarqueurs correspondant à différentes expositions et contribuant à la localisation de régions génomiques impliquées afin d'identifier des cibles potentielles.

Au cours des dernières années, nous avons travaillé sur le développement d'une stratégie de réalisation d'études EWAS et, en tout, nous avons réalisé trois études sur une population de découverte d'individus sains issus de la cohorte SFS. En plus du VEGF-A, nous avons cherché des marqueurs épigénétiques associés aux facteurs de risque intermédiaires communs à différentes maladies chroniques et nous avons mis en évidence des modifications du profil de méthylation associées au tour de taille et au taux de TG. Les possibilités d'applications cliniques sont nombreuses pour des sites de méthylation liés à une maladie ; par conséquent, il est hautement probable que ces résultats soient utilisés dans l'avenir en médecine personnalisée.

LT, le biomarqueur le plus étudié après le VEGF-A, est également une des cibles des travaux scientifiques de notre unité. Notre participation aux projets de Lorraine Université d'Excellence (LUE), soit les projets GEENAGE (*Functional Genomic, Epigenomic and Environment interplay to impact the understanding, diagnosis and management of healthy and pathological Ageing*) et TELARTA (*Telomeres and Arterial Ageing*) nous a encouragé à poursuivre nos explorations de la biologie des télomères et a donné lieu à des collaborations fructueuses. L'étude des associations génétiques de LT et de VEGF-A est l'une de ces collaborations.

Afin de surmonter les premiers obstacles sur la voie de l'utilisation de la longueur des télomères en tant que biomarqueur en médecine personnalisée, nous avons effectué des recherches bibliographiques exhaustives pour comprendre leur régulation génétique, leur rôle important dans certaines maladies et leurs associations avec d'autres molécules

inflammatoires. L'intérêt particulier que nous portons à LT chez les enfants nous a amené à concevoir une étude destinée à faire la lumière sur les déterminants génétiques et environnementaux de la LT pendant l'enfance et à combler les lacunes de connaissances dues au manque d'études longitudinales. Enfin, cette étude testera également l'hypothèse d'un lien entre une attrition rapide de LT pendant l'enfance et la survenue de maladies chroniques chez l'adulte. La publication d'un article de revue rassemblant les informations les plus pertinentes sur les télomères pendant l'enfance a constitué la première étape de cette démarche.

Enfin, les protéines de la famille TREM ont fait l'objet, ces dernières années, de recherches approfondies au sein de notre unité. Les résultats sur TREM2 présentés dans cette thèse font partie des travaux qui ont permis de démontrer un rôle important pour les variants génétiques liés à TREM dans les maladies cardiovasculaires. TREM2 est un des membres les moins connus de cette famille des protéines et représente un potentiel important de découvertes éventuellement applicables en médecine personnalisée.

Tous les travaux entrepris dans le cadre de cette thèse étaient destinés à contribuer à l'amélioration de la médecine personnalisée : compréhension des voies biologiques et de la régulation génétique des biomarqueurs, mécanismes épigénétiques des facteurs de risque et associations entre molécules. Un résumé des résultats obtenus dans les principaux projets de recherche est présenté sous forme graphique dans la figure 34 et montre les liens identifiés entre les biomarqueurs, les facteurs de risque et les voies moléculaires étudiés.

L'étude des associations entre le VEGF-A et les molécules inflammatoires a mis au jour des liens intéressants, la plupart desquels ont déjà été décrits dans d'autres études. Toutefois, des travaux supplémentaires sont nécessaires pour une recherche systématique de voies moléculaires communes entre le VEGF-A et les molécules inflammatoires. Tout d'abord, nos résultats doivent être confirmés sur un plus grand échantillon d'individus en bonne santé, avec en parallèle la mesure des valeurs plasmatiques pour les mêmes biomarqueurs. Cela permettrait de comparer la production de cytokines et de VEGF-A dans les PBMCs et les concentrations dans le plasma auxquelles contribuent tous les types cellulaires qui sécrètent des cytokines. Par la suite, des PBMCs devraient être mises en culture et stimulées par des facteurs angiogéniques ou inflammatoires pour mesurer la production de molécules et d'isoformes d'ARNm dans des conditions spécifiques. Les molécules associées pourraient ensuite être mesurées chez des patients spécifiques (cancer, maladies cardiovasculaires) pour évaluer leur utilité potentielle en tant que cibles thérapeutiques.

VEGF-A

Les résultats des recherches épigénomiques portant sur le VEGF-A décrits dans le cadre de cette thèse correspondent à une étude pilote des marques de méthylation du VEGF-A. Étant donné que des données d'expression des différentes isoformes du VEGF-A sont disponibles pour certains individus de la cohorte SFS, il était possible de réaliser une étude EWAS pour décrire la régulation épigénétique de l'expression du VEGF-A. Néanmoins, cette étude doit être répétée avec un échantillon d'étude plus grand provenant d'une population cas-contrôle comprenant des patients atteints de cancer qui ne reçoivent pas de traitement anti-angiogénique. Cela pourrait permettre l'identification de gènes présentant une méthylation différentielle impliqués dans la régulation de l'angiogenèse. De telles découvertes représenteraient des avancées très importantes pour le développement d'une nouvelle génération de traitements anti-angiogéniques.

Nous avons démontré une association entre le variant génétique rs6993770 lié au VEGF-A et l'attrition des télomères, ce qui suggère l'existence de mécanismes de régulation génétique communes agissant sur les deux biomarqueurs qui pourraient être intéressants en médecine personnalisée. Les associations entre des variants génétiques liés au VEGF-A et d'autres phénotypes et maladies, tels que les taux de lipides, la maladie d'Alzheimer, le diabète de type 2, le syndrome métabolique, ont déjà été étudiées. Une étude de randomisation mendélienne est nécessaire pour évaluer de manière systématique le rôle du VEGF-A dans différentes maladies, des facteurs de risque intermédiaires et la régulation d'autres molécules inflammatoires, en utilisant les variants génétiques liés au VEGF-A pour pallier au manque de données sur l'expression du VEGF-A.

TG et tour de taille

Il est également nécessaire de confirmer les sites CpG identifiés comme étant associés au taux de TG et au tour de taille à l'aide de cohortes plus importantes comprenant des populations d'individus obèses. Parce que le trait « tour de taille » est lié à une augmentation de l'accumulation de tissu adipeux autour de la taille, le site CpG associé à ce trait doit aussi être confirmé dans des échantillons d'ADN extraits de tissu adipeux. Des études de randomisation mendélienne seraient utiles pour déterminer le rôle causal des sites CpG identifiés dans des maladies chroniques courantes.

Longeur de télomères

Le variant génétique lié au VEGF-A, rs6993770, démontré comme étant associé à l'attrition des télomères pourrait expliquer les liens observés avec certains biomarqueurs et l'importance de LT dans l'étiologie des maladies chroniques. Une association entre LTL, LTM et l'attrition des télomères et le SNP rs6993770 est donc à rechercher dans des populations de patients atteints de différentes maladies associées à un raccourcissement de LTL (de nombreuses maladies métaboliques et inflammatoires, comme les maladies cardiovasculaires, le diabète, la rectocolite hémorragique, la cirrhose du foie et le lupus érythémateux systémique). Cela pourrait aider à déterminer si le variant identifié est associé à un raccourcissement accéléré dans toutes les maladies ou si celui-ci est spécifique à l'athérosclérose. Il conviendra de génotyper le SNP rs6993770 dans une population cas-contrôle plus importante de patients atteints d'athérosclérose afin d'évaluer son éventuelle utilité comme biomarqueur de la susceptibilité/du risque d'athérosclérose.

La réalisation d'une étude longitudinale de suivi d'individus dès la naissance et jusqu'à l'âge adulte est essentielle pour la mise en application de LT comme biomarqueur diagnostique/pronostique/de susceptibilité en médecine personnalisée. Grâce au développement rapide de technologies de séquençage du génome entier, LT pourrait être mesurée à l'aide de méthodes bioinformatiques. Des progrès sont donc attendre dans ce domaine.

TREM2

Le variant génétique rs6918289 localisé dans la région du gène *TREM2* doit être étudié par rapport aux taux de la protéine TREM2 afin de déterminer dans quelle mesure ce variant peut agir sur les concentrations plasmatiques de TREM2. Cela pourrait aider à comprendre les mécanismes qui relient rs6918289 à TNF-α et IMT-F. Des analyses transcriptomiques seront à réaliser ensuite pour expliquer la fonctionnalité du SNP et confirmer l'hypothèse proposée. Il conviendra également de tester ce variant génétique dans une population cas-contrôle plus importante de patients atteints d'athérosclérose afin d'évaluer son éventuelle utilité comme biomarqueur de la susceptibilité/du risque d'athérosclérose.

FOREWORD

During the last few decades, modern healthcare systems developed the model of personalised medicine, which adopted new strategies for the fight against common chronic diseases that are affecting the lives of ageing populations around the globe. The wish to create diagnostics, treatment and prognostics tailored to the individual's personal data, not only by disease, guided the development of new analytical tools and therapeutics. Most importantly, it included an individual's genetic information in combination with other clinical data to achieve more precise diagnostics, more efficient treatment and better prognostics of pathology outcomes.

The establishment of a whole new era in medicine demanded enormous efforts in the development of new technologies, methodologies, creating databases (big-data) and big-population cohorts to generate information, which would explain the basis of the genetic regulation of complex diseases. Genetic epidemiology emerged as a new discipline to study the role of genetic factors and their interplay with the environment. Large-scale genome-wide association studies (GWAS) that genotype many thousands of single nucleotide polymorphisms (SNPs) in thousands of individuals have identified variants, each one contributing to some extent to a polygenic, multifactorial complex disorder. Yet, this was not sufficient for the complete understanding of genetic regulation of chronic diseases. The missing link between environmental exposures, genetics and many common diseases that GWAS couldn't provide was partially attributed to epigenetic changes on DNA. The study of methylation changes on genome recently became the most frequently used epigenomic tool for studying the environmental impact on the regulation of genes. High-throughput technologies combined with computer science and informatics are now providing specialised tools which are used to view and analyse biological data from genomics, transcriptomics, proteomics and other related fields.

But how all the above contribute to change the medicine?

Today, we are still in the transition period, where genetic-based tests are being used for the screening of diseases, first gene therapies are being approved, and growing numbers of personalised medications are entering the market, whereas the classical medical approaches remain the gold standards in the hospitals. Research laboratories and the pharmaceutical industry are determined to improve the current system and offer new strategies based on the individual genetic "make-up".

Large quantities of useful information that were generated in the last decade now need to be linked together to provide new knowledge on human biology and pathology. Such an effort could improve the current strategies of prevention, prediction and treatment of common chronic diseases and enrich them with tools provided by personalised medicine. This includes identification of biomarkers that will allow determination of susceptibility for disease, diagnosis, prognosis and treatment and proposition of new therapeutic targets based on the profound knowledge of biological pathways and links between participating molecules.

The research Unit UMR INSERM U1122 Gene-Environment Interactions in Cardio-Vascular Physiopathology (IGE-PCV) has already embraced the new methodologies of personalised medicine and focused on the exploration of genetic and environmental components involved in common age-related diseases. The core of the Unit presents a large Biological Resources Centre (BRC) IGE-PCV, which consists of healthy and pathologic populations for the study of cardiovascular pathophysiology. The wish to create a big database that will serve for genetic epidemiology studies resulted in the initiation of the STANISLAS Family Study. With longitudinal data of more than 1000 families, it became one of the principal populations of the biobank and offered vast genetic, epigenetic and clinical data to study the variability of disease phenotypes and environmental determinants.

Large availability of information from our biobank together with tight collaborations with other cohorts have made it possible for us to explore the genetic and epigenetic regulation of intermediate phenotypes that are involved in diverse age-related pathologies, in order to identify biomarkers common to different diseases pathways.

To do so, we employed a variety of approaches, including epigenome-wide association studies (EWAS) for investigation of epigenetic regulatory mechanisms of different molecules, and studies of associations between genetic variants and various intermediate phenotypes to describe the genetic architecture of complex traits.

We were then continuously focused on building the knowledge around intermediate phenotypes involved in disease pathways and related to risk factors, which presented a special interest and potential for clinical application. Understanding the complex genetic background of these phenotypes would help to provide better stratification of patients by taking into account their genetic profile. The knowledge of associations among different molecules, on the other hand, would offer complementary treatments. Finally, understanding of epigenetic

regulation of intermediate risk factors would provide new susceptibility biomarkers, which could help in early prognostic and assist in the prevention of the onset of chronic pathologies.

Chapter I. INTRODUCTION

1.1 Chronic diseases

Chronic diseases are conditions that persist over a long period of time and can hinder independence and the quality of life of the individual. Generally, they cannot be prevented by vaccines or cured by medication. In the last decades, chronic diseases became a major cause of death and the main public health problem of occidental countries (102). Among them, most common types are cardiovascular diseases (CVD), cancer, chronic respiratory diseases and diabetes. Disease rates are increasing globally, currently accounting for almost 70% of all deaths worldwide (1).

Therefore, countries are investing in public health to make significant improvements in chronic disease prevention, control and management. Most chronic diseases emerge in middle age after a long exposure to an unhealthy lifestyle, caused primarily by four major risk factors: tobacco use, physical inactivity, harmful use of alcohol and unhealthy diets (Figure 1). As those factors may be modifiable, fighting against them is one of the essential strategies of chronic disease prevention. A proper understanding of their influence on health outcomes is the key to promote health and well-being (1).

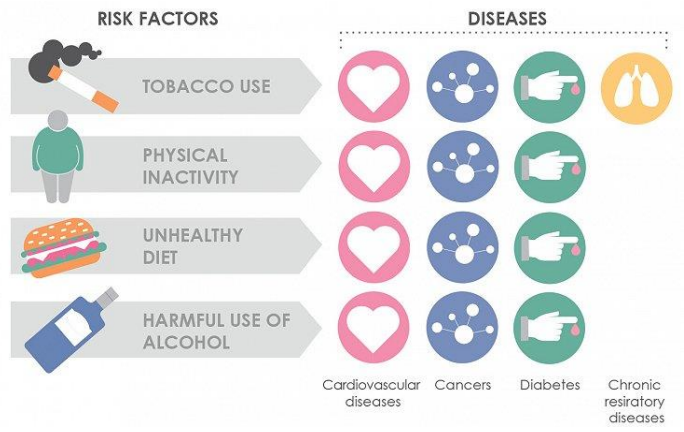


Figure 1: Four major disease risk factors: tobacco use, physical inactivity, unhealthy diet and harmful use of alcohol (1).

In reality, most common diseases are multi-factorial, caused by variants in numerous genes that interact with each other and with a wide range of environmental exposures. For this reason, public health genomics has evolved to integrate advancements in genomics into the fields of personalised medicine and public health (103).

Because of the combination of multiple genes and environmental factors involved in the aetiology of common chronic diseases, they are often referred to as "complex diseases". Their management is often further aggravated with the concurrence of multiple chronic conditions (comorbidities) within the same individual (Figure 2).

Figure 2: Obesity-related comorbidities affect multiple organ systems. (Source: Rethink obesity).

In occidental countries, about one in four adults have at least two chronic conditions and more than half of older adults suffer from three or more (104). Comorbidities are deteriorating the health outcomes, patient's functional status and are increasing the medication intake. The current health-care system is largely configured for individual diseases rather than multi-morbidity, which is also raising the risk of medical error. Complementary strategies are therefore needed to support clinicians and provide personalised and comprehensive patient care (3).

1.1.1 Cardiovascular diseases

CVDs are the most common cause of death worldwide, causing 31% of all global deaths. Heart attacks and strokes are the two prevailing cardiovascular events, representing together 85% of CVD mortality (105). There are different types of CVD, which often overlap by treatment, symptoms, and prevention strategies. They include coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, congenital heart disease and pulmonary embolism (Figure 3).

As for all chronic diseases, multiple factors are involved in the complex pathogenesis of CVD, from non-modifiable genetics, gender, age and ethnicity, to factors that can be modified by drugs or lifestyle change. The exposure to genetic and environmental risk factors can cause the manifestation of so-called "intermediate risk factors", which can be measured and can indicate an increased risk for cardiovascular complications. These are dyslipidaemia, hypertension, hyperglycaemia and obesity (105). Identification of intermediate risk factors can encourage lifestyle changes and initiate a drug treatment, which can normalise the condition and reduce the risk of CV events (4). Another level of complexity for CVD aetiology involves the interactions between risk factors, which are working in synergy with each other, and can considerably amplify each one's impact on a disease (68).

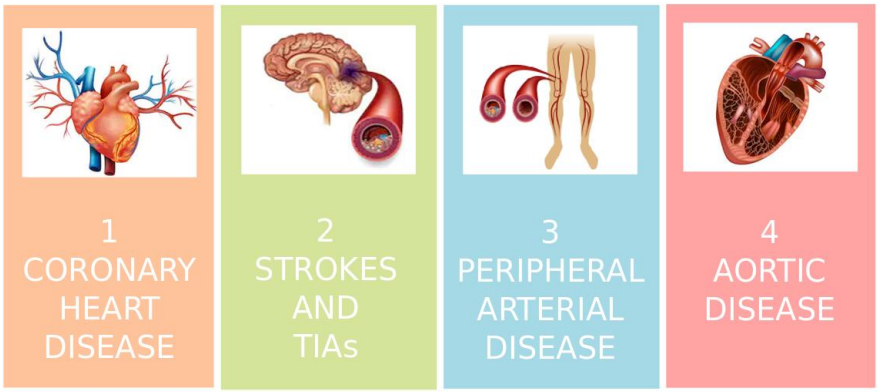


Figure 3: Four main types of cardiovascular disease.

Atherosclerosis is an arterial disease, which is believed to be a major cause of CVD (106). It is characterised with chronic inflammation, caused by interactions between modified lipoproteins, monocyte-derived macrophages, T cells and normal cellular elements of the arterial wall (Figure 4). Continuous inflammation can result in the development of complex lesions called atherosclerotic plaques, that decrease the arterial lumen (107).

The early-stage lesions of atherosclerosis consist of sub-endothelial accumulation of cholesterol-engorged macrophages termed 'foam cells'. Atherosclerosis develops gradually and does not cause any symptoms until the stiffed artery fails to supply an adequate quantity of blood to organs and tissue. The first appearances, named "fatty streaks", are not clinically significant, but are the precursors of more advanced lesions, which appear with the accumulation of lipid-rich necrotic debris and smooth muscle cells (68).

With time, a combination of calcification and ulceration at the luminal surface and haemorrhage from small vessels that grow into the lesion from the media of the blood vessel wall form increasingly complex plaques. These advanced lesions can disturb the blood flow and cause symptoms, which can finally result in serious conditions, *i.e.* coronary artery disease (CAD), cerebrovascular disease or peripheral arterial disease. However, the most important clinical complication is an acute occlusion of the vessel due to the formation of a thrombus or blood clot, resulting in myocardial infarction or stroke (68).

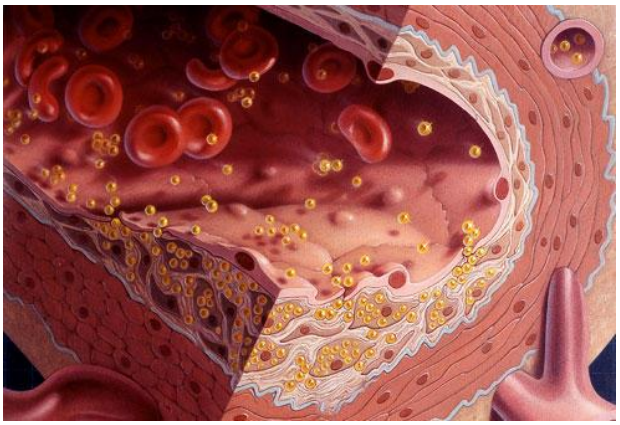


Figure 4: Cross-section of atherosclerotic plaque: disrupted endothelium, caused by increased inflammation in the arterial wall. (Source: Health Tap).

The most important risk factor for the development of atherosclerosis is elevated plasma cholesterol level, followed by hypertension, diabetes, smoking, male gender and inflammatory markers (108). Though some genetic risk factors have been identified, elevated cholesterol is so far the only identified factor, which is sufficient to drive the development of the disease (107). The overall genetic impact on disease susceptibility is still unknown and remains to be fully explored (108).

1.1.2 Diabetes

Diabetes is a group of metabolic diseases that impair the body's ability to process blood glucose, caused by defects in insulin secretion, insulin action, or both. The chronic hyperglycaemia can result in long-term damage, dysfunction and failure of different organs, especially the eyes, kidneys, nerves, heart and blood vessels. Therefore, diabetes is often related to cardiovascular pathologies, which are the most prevalent cause of morbidity and

mortality in diabetic patients (109, 110). According to the World Health Organisation (WHO), there are currently around 422 million people worldwide suffering from diabetes, which is equal to 1 out of 11 adults (111). There exist different types, which differ in pathogenic processes involved in the development of a disease. Most common conditions are type 1, type 2 and gestational diabetes.

Type 1 diabetes is autoimmune destruction of insulin-producing pancreatic β cells. Chronically inflamed pancreatic cell islets lose the function of insulin secretion, which results in deregulated blood glucose. Both, complex genetics of type 1 diabetes and the lack of knowledge regarding the immune response implicated in disease are hindering the progress in understanding and treatment of this disorder (112).

Type 2 diabetes is characterised by insulin resistance, a condition in which cells fail to respond to insulin properly. Insulin resistance demonstrates as decreased peripheral glucose uptake combined with increased endogenous glucose production, which can be further impaired with augmented lipolysis, elevated free fatty acid levels, along with the accumulation of intermediary lipid metabolites (113). Type 2 diabetes is a multifactorial disease with an important interaction between genetics and environmental causes (114).

Gestational diabetes is defined as any degree of glucose intolerance with onset or first recognition during pregnancy (115). It develops as a combination of insulin resistance, which is alone one of many physiological changes occurring during pregnancy, accompanied by pancreatic β-cell insufficiency. The genetic studies of gestational diabetes showed that susceptibility for a disorder is likely to be common to type 2 diabetes; however, it remains largely unknown (116).

1.1.3 Cancer

Cancer is the second largest cause of death globally and is estimated to account for almost 10 million deaths in 2018. It is a generic term for a group of diseases characterised by the growth of transformed cells that proliferate beyond their usual boundaries and tend to invade neighbouring tissues and spread to other organs. Among the most common types of cancers there are lung, female breast, colorectal and prostate cancer (117, 118).

The transformed cancer cells form tumours (Figure 5), which may be either benign or malignant. Only malignant tumours are properly referred to as cancers and differ from benign tumours in their capacity to invade surrounding normal tissue and spread throughout the body *via* the circulatory or lymphatic systems. Development of malignancy is a complex multistep process and diverse factors may affect the likelihood of cancer development (119).

Figure 5: Cancer cells form a tumour, which develops its own vascularisation to enable a continuous flow of oxygen and nutrients. (Source: Santé-Nutrition).

Transformations of cancer cells are driven by mutations that may be inherited, induced by environmental factors, or result from DNA replication errors (120). Therefore, many cancers are related to potentially modifiable risk factors. Cigarette smoking was associated with the most cancer deaths than any other risk factor, accounting for nearly 30% of deaths, followed by excess body weight (121). It has been estimated that between 30% and 50% of cancer deaths could be prevented by modifying or avoiding key risk factors, from where the need for early diagnosis and screening (117).

Identification of carriers of gene mutations, susceptible to common cancers has been shown as the most impactful method for cancer control. With screening, prevention and early detection, national population-testing programs to identify carriers of mutation will soon be integrated into healthcare systems. To fully deliver a precision prevention program, extended mutation studies are still required (122).

Epigenetic marks can also be considered as potential markers of cancer development and progression as they can define distinct cellular identities and specific cell-cell interactions

which are driving tumorigenesis. The reversibility of these alterations is bringing hope for future epigenetic therapies that could be used for clinical prevention and treatment (123).

1.1.4 The major chronic disease risk factors

The risk factors for chronic diseases are most often classified into two groups: modifiable and non-modifiable. Some of the most important modifiable risk factors are obesity, dyslipidaemia, physical inactivity and tobacco use, while age, sex and heritability are referred to as non-modifiable risk factors.

1.1.4.1 Obesity

Obesity is defined as abnormal or excessive fat accumulation that presents a risk for a number of comorbidities, including type 2 diabetes, CVD and cancer (35). A general population measure of obesity is the body mass index (BMI), *i.e.* a person's weight (kg) divided by the square of his or her height (m^2). A BMI $\geq$ 25 kg/m^2 is considered as overweight and BMI $\geq$ 30 kg/m^2 is defined as obesity. Many negative effects of obesity are linked to excessive fat tissue around the stomach, known as abdominal or central obesity. As tall people with abdominal obesity can have a normal BMI, waist circumference (WC) has been recommended as a measurement to assess this risk factor, rather than BMI (36).

Some countries have recognised obesity as a chronic disease, where foods with high energy density are the primary agent that adversely affects a host, alongside with other environmental factors, such as low physical activity (124). Other health professionals describe it rather as a behavioural problem that leads to disease (125). Nonetheless, obesity is a growing health issue (Figure 6), with nearly a third of the world's population classified as overweight or obese in 2019 (126). Understanding the underlying mechanisms leading to obesity and linking obesity to chronic diseases is necessary for the development of the proper therapeutic strategies against it (9).

Obesity and CVD are linked through high levels of fat mass, which worsen most of CVD risk factors, such as plasma lipids, blood pressure, glucose/insulin resistance and inflammation. Moreover, the body adaptation of carrying an extra load in daily activities has been shown to place an extra burden on the heart, resulting in left and right ventricular alterations that

ultimately lead to ventricular hypertrophy and enlargement, predisposing to heart failure (127).

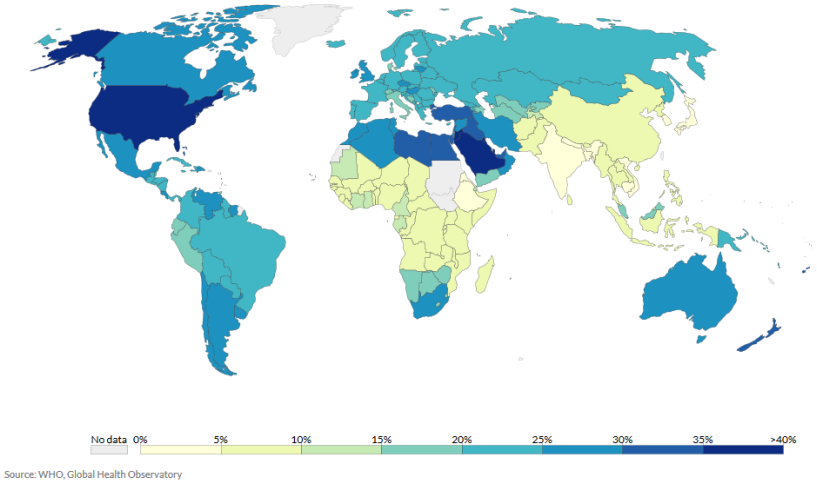


Figure 6: Percentage of adults aged 18+ years old who are defined as obese based on their BMI (35).

Adipose tissue has an important endocrine and metabolic role with extensive effect on the physiology of other tissues and processes, such as hormone metabolism. Moreover, proteins secreted by adipose tissue (*i.e.* adipokines) contribute to the regulation of immune response (leptin), inflammatory response (tumour necrosis factor-alpha (TNF-α), interleukin-6 (IL-6), SAA (serum amyloid A)), angiogenesis (vascular endothelial growth factor (VEGF)), and others (Figure 7).

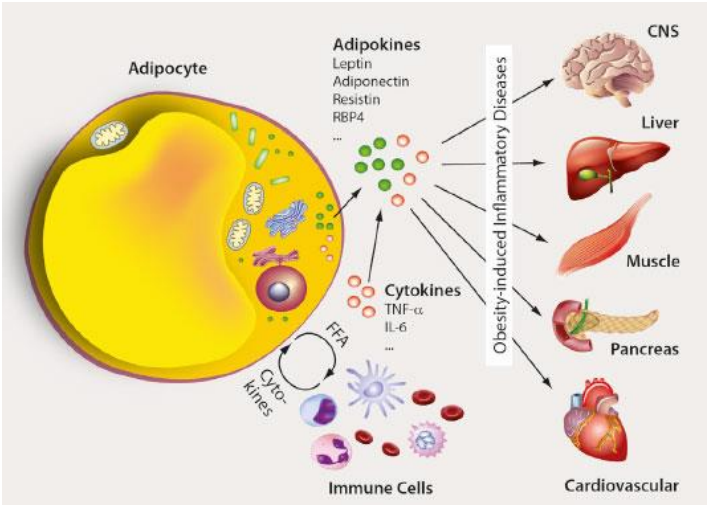


Figure 7: Obesity-induced inflammatory diseases: interaction between adipocytes and immune cells. (Source: Creative Diagnostics).

During obesity, adipokines are dysregulated and create a state of chronic low-grade inflammation responsible for different obesity-linked pathologies and the onset of insulin resistance. Obesity-associated dysregulation of adipokines is likely to contribute also to tumorigenesis and tumour progression, as well as metastatic potential (128).

1.1.4.2 Dyslipidaemia

Dyslipidaemia is a term which describes a deregulated concentration of lipids in the blood. It includes risk factors, such as a decreased concentration of high-density lipoprotein (HDL) cholesterol, qualitative changes in low-density lipoprotein (LDL) and raised triglycerides (TGs). In 46% to 60% of cases, elevated or decreased levels of these lipoproteins may be related to genetic alterations (129). Normally, however, dyslipidaemia reflects the interaction between both genetic influences and environmental factors, such as unhealthy diet and a sedentary lifestyle (130).

Regulation of these perturbations is essential in the prevention of coronary heart disease and other atherosclerotic and chronic diseases. Treatment of dyslipidaemia is normally straightforward; however, some of the disorders are complex and thus resistant to simple treatment strategies (20). The knowledge of the molecular basis of dyslipidaemias allows their correct diagnosis. Genetic diagnosis, on the other hand, can allow early detection, proper therapeutic guidance and as such contribute to the reduction of cardiovascular risk (129).

Hypercholesterolaemia

Hypercholesterolemia is the presence of high levels of cholesterol in the blood, resulting from an unhealthy diet, obesity, genetic factors or the presence of other diseases, such as type 2 diabetes. Most of the serum cholesterol is produced by the liver as triglyceride-rich particles in the form of very-low-density lipoprotein (VLDL), which transforms to LDL after removal of triglycerides. The formation of VLDL and LDL is increased in obesity and with a diet high in saturated fat. The role of LDL is to deliver cholesterol to tissues, as it is an essential component of cell membranes, hormones and vitamins. LDL is removed from the circulation by LDL receptors. Genetic defects in LDL receptor function are the usual cause of familial hypercholesterolaemia. HDL is involved in reverse cholesterol transport; it acquires excess

cholesterol from the tissues and returns it to the liver. It maintains the balance of cholesterol levels in the arteries and is considered as an atheroprotective molecule (131).

Atherosclerosis is the major human disease associated with cholesterol and lipid metabolism. The accumulation of plasma lipoproteins in the subendothelium is the earliest detectable event in atherogenesis (132). Serum levels of LDL play a key role in this process, whereas HDL has shown important protective effects by suppression of atherosclerosis progression (133). LDL can deliver huge amounts of cholesterol to macrophages and cause the formation of foam-cells, which secrete inflammatory cytokines and contribute to smooth-muscle-cell migration and T-cell-mediated inflammatory and immune response. With time, cholesterol-loaded foam cells in combination with smooth muscle cells, extracellular matrix material and smooth-muscle-cell-derived scar tissue lead to slowly progressive lumen occlusion– a formation of advanced atherosclerotic plaque (132).

Besides atherosclerosis, increased cholesterol levels were associated with several other CVDs (stroke, coronary heart disease, heart attacks), as well as cancer cell malignancies (134) and Alzheimer's disease (AD) (135).

Hypertriglyceridaemia

Hypertriglyceridaemia is defined as fasting plasma TG concentration ≥150 mg/dL (>1.7 mmol/l). It results from the interaction of genetic and non-genetic factors, specially poor glycaemic control and obesity (136). Plasma TG levels partially reflect the concentration of the triglyceride-rich lipoproteins (TRLs): VLDL, chylomicrons and their remnants. Hypertriglyceridaemia often causes the reduction of HDL levels and an increase in atherogenic small dense LDL levels and is an important risk factor for CVD, type 2 diabetes, chronic kidney disease and acute pancreatitis (137).

Macrophages incorporate cholesterol issued from TRLs, resulting in the formation of foam cells, which accumulate and contribute to the formation of atherosclerotic plaque (Figure 8). TRLs activate a number of proinflammatory, procoagulant and proapoptotic signalling pathways that play a fundamental role in the pathogenesis of atherosclerosis (137). Though LDL was considered as a primary risk factor for atherosclerosis, studies showed that elevated TG levels are independently associated with increased incidence of CV events, suggesting

that high TG levels should be individually recognised as an important target for therapy (137-139).

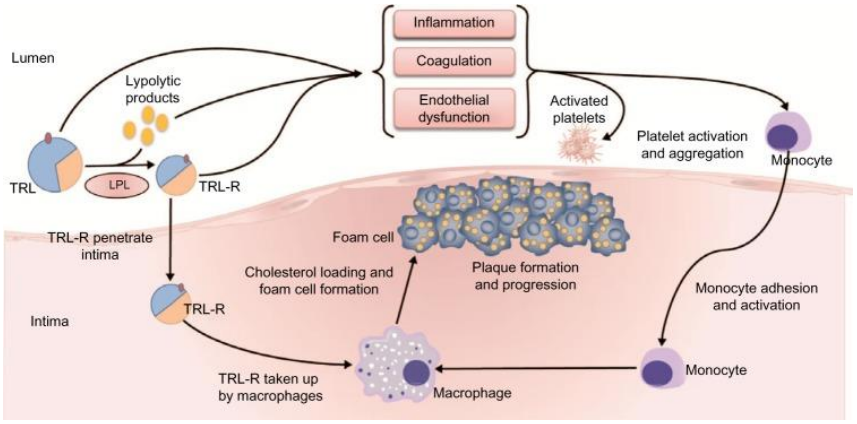


Figure 8: Pathophysiology of TG-rich lipoproteins in the progression of atherosclerosis (137).

1.1.4.3 Physical activity

Physical activity is considered as any movement produced by muscles, which results in energy expenditure. It is not limited to exercise but involves all kinds of daily activities. Insufficient physical activity is considered the fourth most important risk factor for global mortality (140).

Regular activity has been associated with a protective effect on CVD development and cancer, possibly through a beneficial impact on traditional risk factors, such as blood pressure, lipids, obesity and diabetes (141). It has been demonstrated that increase in physical activity can be beneficial also in middle-aged and older adults, including CVD and cancer patients, regardless of past activity levels and established risk factors, such as diet, BMI, blood pressure, TG and cholesterol (142).

1.1.4.4 Tobacco use

The use of tobacco has been known as the largest preventable cause of death and was linked with various cancerological, cardiovascular, neurological, and pulmonary diseases. In particular, it was estimated to cause over 90% of lung cancer in men, 70% of lung cancer among women and about 22% of all CVD (143). The adverse effect of smoking on health outcomes comes from toxic chemicals in tobacco smoke, of which nicotine, carbon monoxide

and oxidant chemicals are most commonly implicated in the pathogenesis of smoke-related morbidities. They affect systemic oxidants-antioxidants imbalance (causing lipid peroxidation and decreased levels of vitamins A and C) and lead to a low-grade systemic inflammatory response (causing elevated levels of C-reactive protein (CRP), fibrinogen, IL-6 and others). The effect of low-grade inflammation can persist even 10 to 20 years after quitting smoking.

In addition, cigarette smoke contains many carcerogenic substances, which are activated by metabolic processes and converted by cytochrome P450 enzymes to electrophilic entities. Electrophilic entities then covalently bind to DNA and form DNA adducts, which can initiate carcinogenesis (144-146).

1.1.4.5 Age and sex

In the occidental countries, age is the main independent risk factor for the development of most common chronic disorders, including cancer, CVD and neurodegenerative disorders, which are often referred to as "ageing diseases" (147). With advanced age, long-term accumulation of all kinds of risk factors causes alterations, altogether leading to a decline in function and an increase in the chance of disease or death (2). Age-related phenotypes are largely variable between individuals. They are regulated with complex genetic mechanisms that contribute to a large number of independent processes of damage accumulation with little or no common causality (148). Mechanisms implicated in ageing are related either to the maintenance of metabolic and energy homeostasis, or to damage repair, and show complex implication in aetiology of ageing diseases (2).

However, the above-discussed processes do not occur at the same time in men and women. Globally, under the age of 85 years, the number of deaths is greater for men than for women in all cause-of-mortality categories (149). Sex is, therefore, another important non-modifiable risk factor. Because it affects a wide range of physiological functions, it has an impact on susceptibility, age of onset and the progression of a variety of diseases. Sex differences are particularly well-investigated in CVDs because of strong epidemiological evidence that men and women face different risks and confront different outcomes (150). Often, women develop chronic pathologies later and are less likely to die from intentional and unintentional injuries (Figure 9).

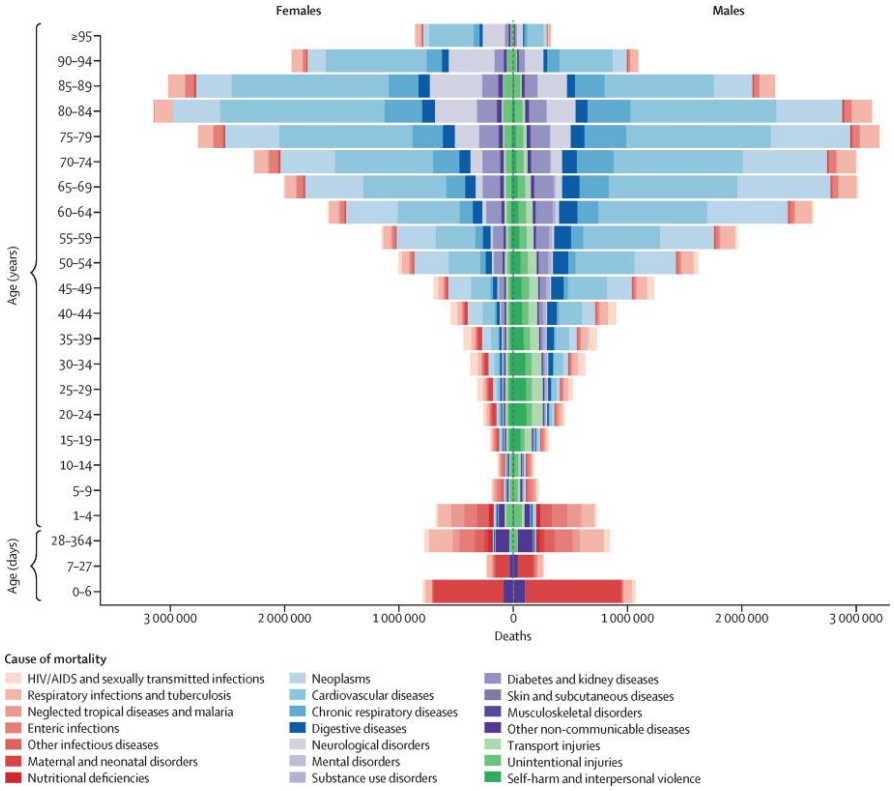


Figure 9: Sex difference in global mortality in the year 2017, presented by the age group (149).

The majority of the sex differences is caused by differences in endocrine and immune systems (151). Sex-specific diseases occur only in people of one sex, whereas sex-related diseases are more common to one sex. A classic example of the sex-specific difference is the occurrence of ovarian cancer only in women and prostate cancer only in men. The example of the sex-related difference, on the other hand, is multiple sclerosis, where women have a greater susceptibility to a disease, while men have worse disease progression (152). Finally, it is important to keep in mind that besides the impact of the immune and endocrine system, there are genetic influences on sex-specific differences in the epidemiology of common diseases, which are studied by gene-environment interactions (153).

1.1.4.6 Heritability

Family history of a disease is a piece of important information for understanding the risks of developing a chronic disease. It comes from common genetics, environment, lifestyle and habits shared among family members, which can significantly contribute to susceptibility for

disease. The main characteristics of inherited diseases are the following: they can occur earlier than expected (before the age of 55), are identified in more than one close relative, are not sex-specific (can appear in a sex which is normally less likely to have this disease, such as breast cancer for men) and can appear in specific combination among relatives (154). Unlike monogenic diseases, common chronic pathologies involve many genes with varied and complex inheritance patterns. The calculation of risk for such conditions is thus much more complicated. Therefore, observations of data collected from large families affected by a disease are compared to a family medical history as they can help to estimate the probability of developing a disease (155).

1.2 Common disease pathways

Complex diseases are mediated by different pathogenesis pathways, which can increase the incidence of comorbidities, suggesting that various human diseases could be biologically connected (156). The good understanding of such common pathways and interaction between them (Figure 10) is particularly interesting, as it might result in the development of therapeutic strategies effective across multiple diseases (13).

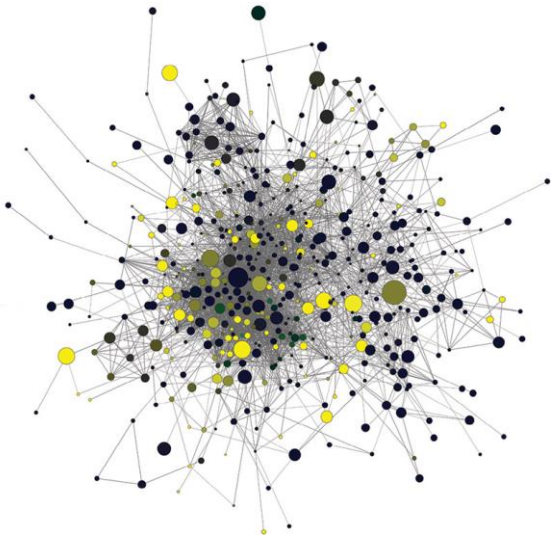


Figure 10: Protein-protein interaction network. (Source: Kettenbach Lab).

Two important processes, implicated in multifactorial diseases are inflammation and angiogenesis. They are mutually linked through common molecular mechanisms: immune

cells synthesise and secrete pro-angiogenic factors that promote neovascularisation during inflammatory reactions, whereas newly formed vessels maintain the inflammation by promoting the migration of inflammatory cells (157, 158).

1.2.1 Inflammation

Inflammation is an adaptive immune response activated by noxious stimuli and conditions, such as infection or injury. It enables survival by the maintenance of tissue homeostasis through a wide variety of physiological processes. Each inflammatory response is mediated by different mechanisms of induction, regulation, and resolution. Depending on the stressor that causes the induction, the sensors, mediators, and target tissues are regulated in the way that appropriate type of inflammatory response is induced (Figure 11) (159). Nevertheless, all of the inflammatory responses share the common mechanism, characterised by the four critical steps: 1) recognition of deleterious stimuli by cell surface pattern receptors; 2) activation of inflammatory pathways; 3) release of inflammatory mediators; and 4) recruitment of inflammatory cells (160).

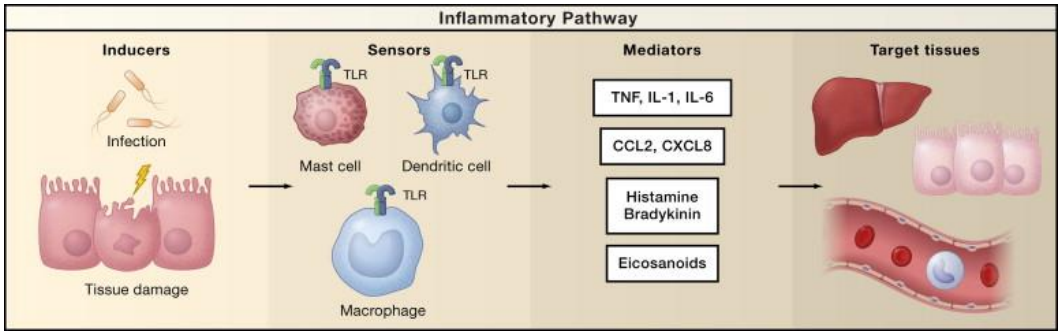


Figure 11: The inflammatory pathway consists of inducers, sensors, mediators, and target tissues (159).

The recognition of infection is mediated by tissue-resident macrophages and mast cells, which initiate the production of a variety of inflammatory mediators, including chemokines, cytokines, vasoactive amines, eicosanoids and products of proteolytic cascades. Mediators help in the recruitment of plasma proteins and leukocytes from the bloodstream to the affected tissue through the cytokine-activated endothelium. When leukocytes reach the affected tissue site they activate by direct contact with pathogens or through the actions of cytokines secreted by tissue-resident macrophages and act to remove the invading agent (161).

The acute inflammatory process is terminated when the inducers are removed, and the tissue is repaired. Inflammation is beneficial as long as it is well-controlled but can become harmful for the tissue if deregulated. In some cases, the process persists for a longer period, and a chronic inflammation appears. Normally, chronic inflammation is caused by continuous stimuli of allergens, unrepaired tissue damage or chronic infection. In some cases, the initiating trigger does not seem to be associated with any classical inducer (159, 161). Such chronic inflammatory response has been observed in a wide variety of conditions, including type 2 diabetes, cardiovascular and neurodegenerative diseases, obesity, cancer, asthma, chronic inflammatory rheumatic disorders and ageing (162). Inflammation and pathologic processes are often caught in a reciprocal circle of continuous positive feedback stimulation (159). The initiation of chronic inflammation is still not well understood, but it is believed to be related to the homeostatic imbalance caused by the malfunction of the tissue (159, 161).

Risk factors linked to chronic diseases have been shown to increase the inflammatory response. Therefore, downregulation of inflammation-associated risk factors could prevent or delay the onset of inflammation-induced diseases (163). A pathological inflammatory response can be controlled at multiple levels, especially through the production of inflammatory mediators. However, there are still substantial knowledge gaps in understanding of regulatory principles, due to the complexity of the inflammatory response and the multitude of components involved (159).

1.2.1.1 Cytokines

Cytokines are a category of small proteins (~5–20 kDa), important in cell signalling, and are common inflammation mediators (Figure 12).

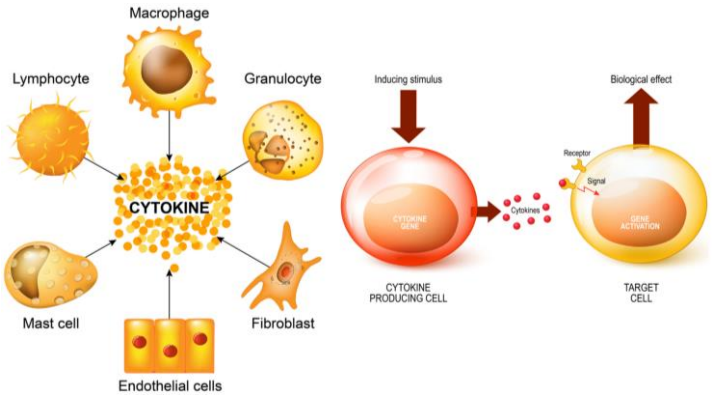


Figure 12: Production and action of cytokines. (Source: Shutterstock).

They consist of different protein families, including chemokines, interferons, interleukins, lymphokines, and tumour necrosis factors. Normally, cytokines are secreted from immune cells, such as monocytes, macrophages, and lymphocytes. They differ in pro- and anti-inflammatory functioning, through which they act to facilitate or inhibit inflammation (164). Most important cytokines and their function in inflammation are presented in Table 1.

Table 1. Molecules from the class of cytokines and their role in the inflammation process.

Cytokine	Family	Main sources	Function
IL-1α	IL-1	Macrophages, neutrophils, epithelial cells, endothelial cells	Pro-inflammation, proliferation, cytokine production, fever
IL-1β	IL-1	Macrophages, monocytes	Pro-inflammation, proliferation, apoptosis, differentiation
IL-2	IL-2	Activated CD4+ T cells, activated CD8+ T cells, NK cells, dendritic cells	T-cell differentiation, apoptosis
IL-4	IL-4	Th-cells	Anti-inflammation, T-cell and B-cell proliferation, B-cell differentiation
IL-6	IL-6	Macrophages, T-cells, adipocyte	Pro-inflammation, differentiation, cytokine production
IL-8	CXC-chemokine	Macrophages, epithelial cells, endothelial cells	Pro-inflammation, chemotaxis, angiogenesis
IL-10	IL-10	Monocytes, T-cells, B-cells	Anti-inflammation, inhibition of the pro-inflammatory cytokines
MCP-1	CC-chemokine	Monocytes, macrophages, dendritic cells	Pro-inflammation, migration and infiltration of monocytes/macrophages
TNF-α	TNF	Macrophages, NK cells, CD4+lymphocytes, adipocyte	Pro-inflammation, cytokine production, cell proliferation, apoptosis, anti-infection
IFN-γ	INF	T-cells, NK cells, NKT cells	Pro-inflammation, innate, adaptive immunity anti-viral

Cytokines modulate the inflammation through a complex network of interactions, by which they regulate the activation of the resident cells and newly recruited inflammatory cells. More

specifically, they are the major determinants of the make-up of the cellular infiltrate, state of cellular activation, and systemic responses to inflammation. Their functioning is often deregulated during chronic inflammation, which causes excessive inflammatory cytokine production that can lead to tissue damage, organ failure, and ultimately, death. A better knowledge of cytokine interaction in inflammatory response could help in the effort to regulate pathological processes with identification of the therapeutic agents, which could be used in the treatment of chronic diseases (160, 165).

1.2.1.2 TREM Family proteins

The triggering receptors expressed on myeloid cells (TREM) are a family of cell-surface molecules, responsible for the regulation of innate resistance, inflammation, bone homeostasis, neurological development and blood coagulation (166). At least one member of TREM family is expressed on the surface of each monocyte, macrophage, myeloid dendritic cell, plasmacytoid dendritic cell, neutrophil, microglia, osteoclast and platelet. They are essential for the defence against pathogens and participate in fine-tuning of immune reactivity by amplification or reducing of toll-like receptor (TLR)-induced signals, which activate innate immune cells to clear the pathogen. The process of fine-tuning is essential in the prevention of excessive inflammation that could result in tissue damage (167).

- TREM1

TREM1 is an amplifier of the systemic inflammatory response, commonly associated with sepsis. During infection, receptor expression is modulated, and soluble TREM-1 is released. This leads to a cascade of intracellular events that result in inflammatory effects, such as cytokine production, degranulation of neutrophils, and phagocytosis (168). TREM1 tunes the septic response to facilitate the efficient clearance of the pathogen without damaging the host (167).

- TREM2

TREM2 is an important negative regulator of autoimmunity. It is expressed on newly differentiated and alternatively activated macrophages, where it acts to restrain macrophage activation (78). TREM2 can suppress macrophage TNF-α and IL-6 production and promote an anti-inflammatory state. It was first detected on human monocyte-derived dendritic cells where it is also required for the differentiation of myelinating oligodendrocytes. Later, it was

also found on other cell types, including endothelial cells, but its function in this cell type is not yet fully characterised (166). Single nucleotide polymorphisms (SNPs) in the gene encoding TREM2 have been associated with AD, implying on a key role of TREM2 molecule in the pathology of disease (169). Moreover, model animal studies have shown its importance in other inflammatory diseases, such as chronic obstructive pulmonary disease (COPD), multiple sclerosis, and stroke (170). Its physiological role and mechanism in CVD remain to be fully investigated.

1.2.2 Angiogenesis

A proper vascularisation that enables a continuous flow of oxygen and nutrients is crucial for the normal functioning of every tissue. The construction of the vascular system starts in embryotic development with vasculogenesis, a process that allows the formation of the primary vascular plexus. Vasculogenesis is followed by angiogenesis, which enables a formation of new blood vessels from existing vasculature through the interaction of the endothelial and smooth muscle cells (171, 172).

Two different types of the angiogenic process occur in utero and in adults. Firstly, sprouting angiogenesis, which is characterised by the growth of sprouts from the pre-existing vascular network toward an angiogenic stimulus, such as VEGF-A (173). It is the most common and best-understood mechanism of angiogenesis (Figure 14).

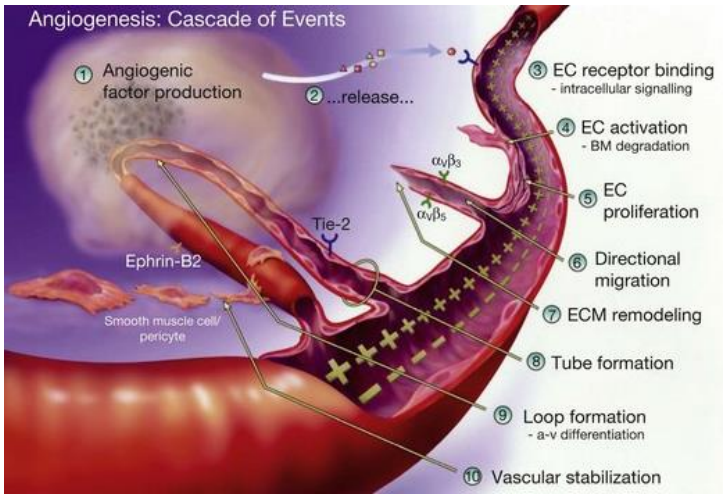


Figure 13: The cascade of events associated with sprouting angiogenesis. (EC: endothelial cell, BM: basement membrane). (Source: The Angiogenesis Foundation).

The second process is called intussusceptive angiogenesis, which occurs much more quickly than sprouting angiogenesis, as it allows the formation of new capillaries without the need of proliferation of endothelial cells. It is characterised by the formation of blood vessels by a splitting process, in which new blood vessel is created by dividing an existing vessel in two (173, 174).

Sprouting angiogenesis is initiated in the poorly perfused tissues, where sensors detect low oxygen levels and secrete pro-angiogenic stimuli to guide the proliferation and migration of endothelial cells. VEGF-A stimulates the formation of specialised endothelial cells and allows the release of large amounts of proteolytic enzymes to degrade the basement membrane and enable the migration of the endothelial cells. The endothelial cells proliferate in the surrounding matrix and migrate towards the source of the angiogenic stimuli, where they re-organise to form tubules with a central lumen and form connections with neighbouring vessels (173, 174).

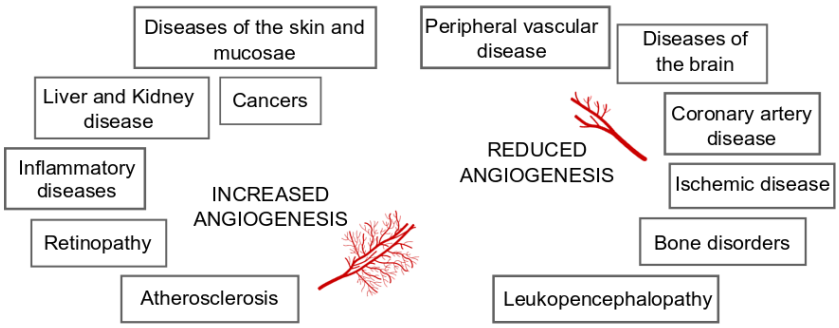


Figure 14: Two kinds of pathological situations caused by increased or reduced angiogenesis (175).

Angiogenesis is a physiological mechanism, which mostly occurs in embryogenesis. In adult tissues, it is regulated through tightly controlled processes and appears during the female reproductive cycle and in wound healing (176). However, unregulated angiogenesis is linked to a great number of pathologies, caused either by exaggerated or inadequate angiogenic activity (Figure 15). Stimulation and repression of these mechanisms have therefore been studied for application in many pathological conditions. Ischemic heart disease, peripheral arterial disease, and wound healing could benefit from induced neovascularisation. On the other hand, decreasing or inhibiting angiogenesis is already used in the treatment of several

types of cancer and ophthalmic conditions, and could be potentially used as therapeutic also in rheumatoid arthritis and other diseases (43, 174, 177).

VEGF-A signalling pathway is a principal target of anti-angiogenic drugs as it is increasingly expressed in most human cancers (178). The importance of VEGF-A in chronic diseases will be discussed in the next chapter.

Besides growth factors, cancer cells secrete a variety of cytokines that can also stimulate classical angiogenic signalling pathways, or they induce an inflammatory response that initiates angiogenesis and consequent vascularisation of the tumour (179). The common regulation of both processes could result in new agents, which could be used in the treatment of related disorders.

1.3 Biomarkers

A biomarker has been defined as "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (180). Generally speaking, biomarkers indicate a change in expression or state of measured characteristics that correlates with the risk or progression of a disease. They allow the examination of normal biological processes, diagnosis of pathogenic processes and prognosis of a disease. Moreover, they play a critical role in improving the drug development process as surrogate endpoints (11, 12).

Biomarkers have been classified in different categories based on their use in clinical practice (181):

- **Susceptibility/risk biomarkers** are used for indication of the potential for developing a disease or medical condition in an individual, who does not currently have a clinically apparent disease or medical condition.
- **Diagnostic biomarkers** are used for detection or confirmation of the presence of a disease or specific condition, or for identification of a subtype of the disease.
- **Monitoring biomarkers** are used for serial assessment of presence, status or extent of a disease or medical condition, or to provide evidence of exposure to (or an effect of) a medical product or environmental agent.

- **Prognostic biomarkers** are used to identify the likelihood of a clinical event, disease recurrence, or disease progression in patients with a disease or medical condition of interest.
- **Predictive biomarkers** are used to predict the likelihood to experience a favourable or unfavourable effect from the exposure to a medical product or environmental agent.
- **Pharmacodynamic/response biomarkers** are used for detection of changes in response to exposure to a medical product or an environmental agent.
- **Safety biomarkers** are used before or after an exposure to medical intervention or environmental agent to indicate the likelihood, presence, or extent of toxicity as an adverse event.

There is a special need for the use of biomarkers in the field of personalised medicine, which is directed towards precise diagnostics of disease and stratification of patients regarding their condition, in order to generate the optimal treatment and decrease the risk for side effects, caused by inappropriate therapy. The development of specific diagnostic biomarkers that would allow the identification of individuals with a disease and classification of a disease subtype are therefore particularly important (11). The need for the new biomarkers encouraged the scientists for investigation which resulted in an abundance of publications reporting clinically useful molecular biomarkers (182). Despite the enormous expansion in literature, there is still only a limited number of biomarkers currently used in discovery, treatment, or monitoring of common chronic diseases (183).

The problem stands in the complexity of chronic diseases, caused by the web of genetic variants and environmental risk factors that are each contributing to the development of pathological conditions (182). Fortunately, the identification of new biomarkers is now accelerated with fast advances in molecular biology, genetics, and computational biology, which enables the research of genomic biomarker as an alternative to traditionally measured characteristics. Consequentially, genetic tests for population screening, diagnostic, predictive, susceptibility testing and pharmacogenomics purposes are being developed (184).

Genomic biomarkers are defined as a "DNA or RNA characteristic that is an indicator of normal biologic processes, pathogenic processes, and response to therapeutic or other intervention." They reflect the expression, function or regulation of a gene. DNA characteristics include SNPs, DNA modification (*e.g.* methylation), insertions, deletions, copy

number variations and others. RNA characteristics include RNA sequence, RNA expression levels, micro RNA (miRNA) levels and others (184).

A good understanding of a disease, a combination of genomic and molecular features and identification of common pathological pathways in comorbidities are necessary for the design of biomarker research. Two important biomarkers, frequently involved in the pathology of chronic diseases are VEGF-A and telomere length.

1.3.1 VEGF-A

VEGF-A is a multifunctional signal protein, which works as a key regulator of both physiological and pathological angiogenesis, is involved in inflammation (41) and has been related to a variety of pathologies, such as cancer, CVDs and neurodegenerative disorders (185). It is known also as vascular permeability factor, because of its key role in inducing of vascular leakage, which is especially important in the process of inflammation (Figure 15).

VEGF-A works as a chemotactic agent for macrophages and granulocytes. Moreover, it induces the proliferation and migration of vascular endothelial cells and protects them against apoptosis to enable their survival in modified conditions (185, 186).

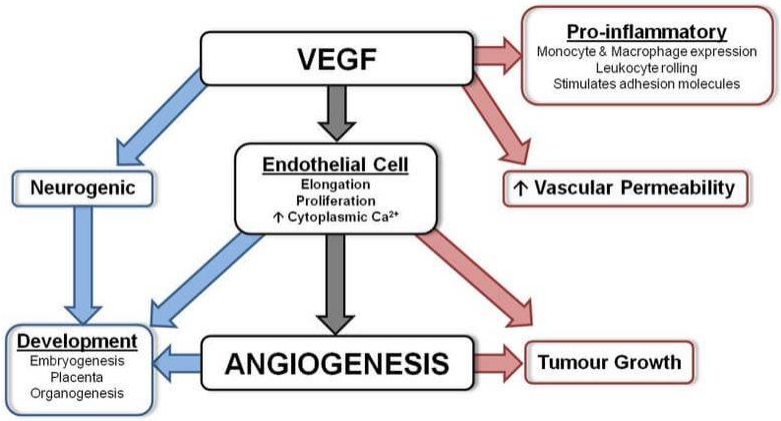


Figure 15: VEGF-A role in different (patho)physiological mechanisms. (Source: Selfhacked).

The action of VEGF-A signalling is exercised through binding to the tyrosine kinases Flt-1 (VEGFR-1) and Flk-1/KDR (VEGFR-2), high-affinity VEGF receptors (Figure 16), found on a wide variety of cell types (187). On ligand binding, VEGFR-2 dimerises and activates its

receptor tyrosine kinase activity, which results in autophosphorylation of the intracellular domains. This can then lead to the activation of a variety of signalling pathways (42).

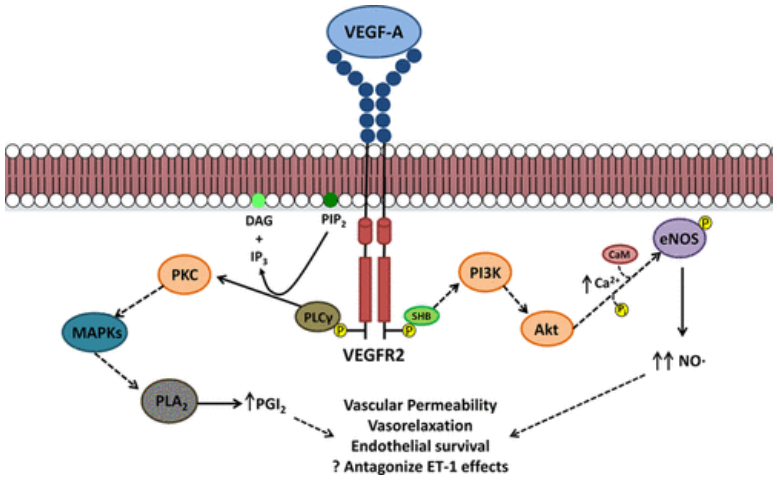


Figure 16: Intracellular signalling pathways for VEGF-A (42).

The VEGF-A plasmatic concentration is highly heritable phenotype, up to 60.5 % as demonstrated in the SFS (44). 50 % of this heritability was explained with four SNPs in GWAS (46). VEGF-A has 14 isoforms, classified in two main families: the *VEGFxxx* family and the *VEGFxxxb* family. The isoforms of the two families differ in the sequence of six amino acids, which is a result of alternative splicing of mRNA from a single, 8-exon of *VEGF-A* gene (188). The consequence of this difference is changed function of the molecule; *VEGFxxx* isoforms have pro-angiogenic properties, whereas the *VEGFxxxb* isoforms have anti-angiogenic properties (189, 190). The four most abundant *VEGF-A* isoforms are $VEGF_{165}$, $VEGF_{189}$, $VEGF_{145}$ and $VEGF_{121}$.

VEGF-A has become a prospective target for the design of anti-cancer treatment and bevacizumab was the first VEGF-A inhibitor approved for the treatment of cancer (191). Besides bevacizumab, few other anti-VEGF medications have already entered the market for the treatment of certain types of cancers. However, the trade-off for such therapy is often the occurrence of cardiovascular side effects (192, 193).

In normal physiological conditions in adults, anti-angiogenic isoforms are expressed in non-angiogenic tissues and represent more than 50% of the total amount of VEGF-A protein (194). In the occurrence of cancer, a reversal of the situation is observed, and pro-angiogenic isoforms are expressed in place of the anti-angiogenic isoforms, leading to neovascularisation

and development of tumour (188, 190). It has been suggested that the expression profile of isoforms of different types of tumours could affect the response of anti-VEGF therapies. The interesting fact is that alternative splicing, leading to the expression of different isoforms, may depend on environmental stimuli. It is expected that a better understanding of this mechanism could enable us to charge other molecules, not just VEGF-A, for tumour suppression activities (195). Because VEGF-A is a mediator of both, angiogenesis and inflammation, which are closely involved in many pathologies, inflammatory cells are likely to be the next potential targets for anti-angiogenic therapies (52).

1.3.2 Telomere length

A telomere is a repetitive sequence TTAGGG on the ends of the eukaryotic chromosomes (Figure 17) (196), which forms a protective cap that conserves the genetic material during cell divisions and protects it from constitutive exposure to the DNA damage response (197). It is shortening progressively with age (198), as a result of the end replication problem during cell division or oxidative stress. Chromosomes with critically short telomeres are recognized by DNA damage response proteins as a damaged DNA, which leads the cell to controlled telomere-initiated senescence (199). When cells become senescent, they undergo morphological and genetic changes that result in loss of tissue function.

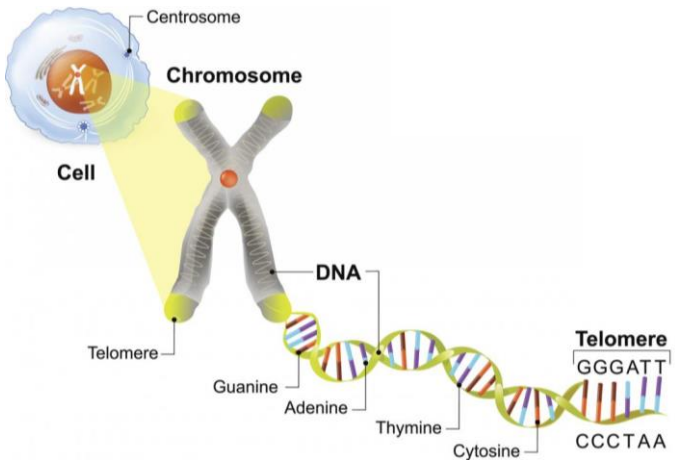


Figure 17: The structure of telomere. (Source: Medical news today).

A cell is capable of maintaining the length of telomeres with telomerase enzyme, which is extinguished during embryonic differentiation in most somatic cells but remains active in

germline cells, activated lymphocytes and certain types of stem cell populations. During normal human growth and development, telomerase activity is precisely regulated with a number of genes, in order to meet the proliferative demand of specific cellular functions (200). Cancer cells, on the other hand, can acquire a possibility of an infinitive number of divisions while allowing the telomerase enzyme active and thus restraining the shortening of telomeres (201).

Telomere length (TL) is a biomarker that could enable a good stratification of patients at risk of disease as it has high heritability, a strong relation to age and chronic diseases (97-99) and large inter-individual variability. The heritability of TL is estimated from 36 to 86 % from family and twin studies (202-205). So far, several GWAS (206-216) have been performed to identify common genetic variants associated with relative TL and 23 SNPs from different genes have been reported.

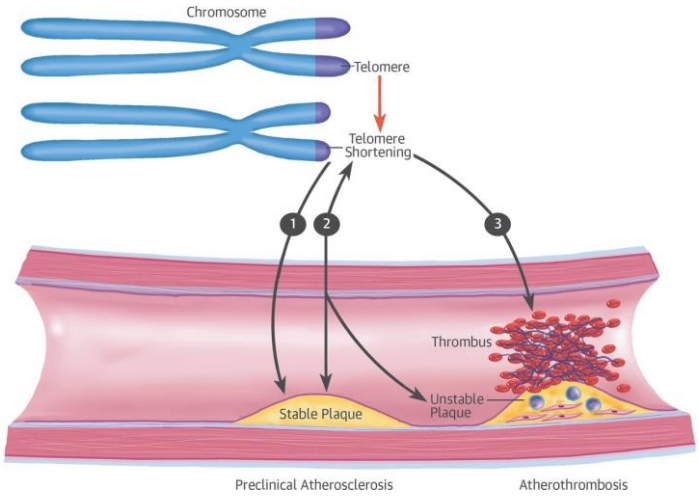


Figure 18: Shorter inherited telomeres have been related to advanced atherosclerotic lesions (217).

The alterations in genes involved in the regulation of telomere length can result in the occurrence of telomere syndromes, characterised by critically short telomeres that impact the fast-dividing tissues (218). Besides monogenic diseases, telomeres have been associated with many common age-related chronic diseases (219). Researches have shown an important role of telomeres also in atherosclerosis (Figure 18). Different hypotheses exist to explain this connection, whereas the exact mechanism of causality remains unknown. Firstly, shorter inherited telomeres could provoke early senescence of vascular tissue, leading to accelerated atherosclerosis development. Secondly, a combination of oxidative stress, ageing, and

inflammation could cause both telomere attrition and atherosclerosis development. And thirdly, short telomeres could induce haematopoietic stem cell senescence and inhibit production of endothelial cells, thus obstructing repair mechanisms and causing destabilisation of plaques (217). Recent studies showed the direction toward the first hypothesis, implying that shorter telomere detected in atherosclerotic patients compared to controls come from the increased attrition in early life (81).

Accumulating evidence also suggests a strong association of increased systemic inflammation with decreased TL (220, 221). Short TL has been associated with increased concentrations of inflammatory markers such as TNF-α, (222), CRP, sICAM-1, sVCAM-1 and SAA (221, 223). Other biomarkers include VEGF-A, EGF (224), E-selectin, L-selectin, P-selectin, IL-1β, IL-2, IL-6, IL-8, and IL-10 (221, 223) and show quite inconsistent results, pointing towards the complexity of these molecular interactions, which still need to be elucidated.

1.4 Methodologies of personalised medicine

Personalised medicine, commonly known as precision medicine (Figure 19), refers to a "medical model using characterisation of individuals' phenotypes and genotypes (*e.g.* molecular profiling, medical imaging and lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and to determine the predisposition to disease or to deliver timely and targeted prevention (225)."

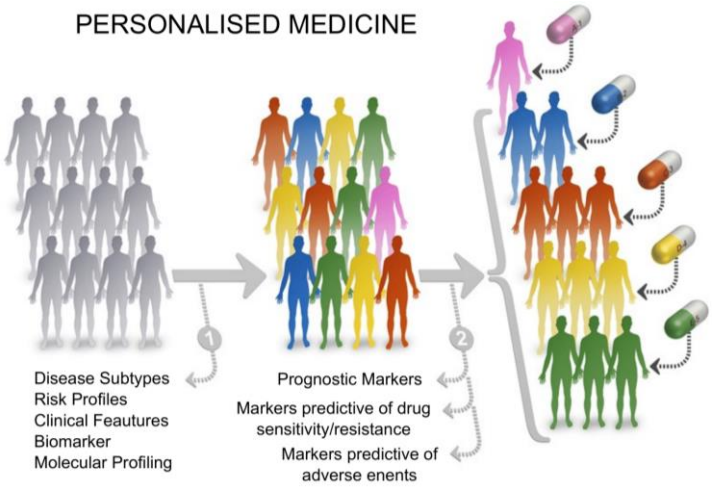


Figure 19: The strategy of personalised medicine – better diagnostic, prognostic and treatment. (Source: Frost & Sullivan).

In other words, personalised medicine changed the traditional medical approaches by adding a wide variety of data to the common clinical information of a patient, using a combined knowledge to enable better risk prediction, diagnosis, prognosis and treatment (226).

In the aspect of therapeutic development, personalised medicine can be divided into three different categories (227):

- **Individualised therapies**, where a single medicine is created for a single patient.
- **Precision medicines**, which are medicines accompanied by specific test to determine the overall likelihood of a beneficial effect.
- **Targeted medicines**, which are medicines specific to a single molecular target and can be robustly tested to identify patients who can receive the medicine.

The fast evolution of molecular diagnostics based on a more detailed understanding of a disease and its mechanisms has resulted in therapeutic alternatives in different clinical areas, in particular oncology. Today, there are more than 90 medicines on the market that require testing to determine accurate dosing (227).

Applications of personalised medicine are based on the fundamental discoveries, acquired from the use of the advanced technologies that enabled the insight into genetic code. At the beginning of the 21st century, the gold standard became the GWAS. Afterwards, high-throughput "-omics" technologies developed to understand genetic causality of complex traits of human diseases through parallel analysis of genes, transcripts, and proteins and to detect qualitative as well as quantitative levels of multiple genes, transcripts, and proteins simultaneously. They have been playing a key role in the identification of effective models that predict phenotypic traits and outcomes, in the discovery of biomarkers, and generation of important insights into the genetic foundations of the heritability of complex traits. "Omics" technologies helped to accelerate the medical-genetics analyses and contributed to the advances in molecular diagnosis and personalised medical treatment (228-230).

1.4.1 Genetics

The Human Genome Project (HGP), which was accomplished in 2003, provided a complete sequence of human DNA, which served as a foundation to gene catalogues, understanding of

structure and regulation of genes, determination of the variants in different populations and linking the variants to specific diseases. Understanding the genetic basis of disease was expected to lead to better targeted therapies. Therefore, the genetic diagnosis became an objective of medical scientists who were seeking an earlier identification of the affected patients and the improvement of patient's treatments (231).

The sequencing of amplified fragments corresponding to the particular gene regions is the conventional way of searching for mutations in a gene. It is mostly used for discoveries of genetic variants in genes related to simple monogenic diseases. Sophisticated molecular methods coupled with computational methods allowed the development of new analytical techniques which were designed to unravel the genetic basis of complex diseases (232). The innovation of genetic methodologies resulted in the production of large amounts of data that need the support of statistical and computational methods to be correctly processed. The main statistical and computational methods in genetic studies are presented below.

1.4.1.1 Association studies

Association studies use the genetic information to find a link between a complex disease or a particular condition and genetic or epigenetic variant. The genetic variants associated with a complex disease are often common polymorphisms that individually have little impact on the phenotype, but could be considered a cause of disease in combination with other risk variants (233).

The identification of variants associated with a specific disease gave the idea of a polygenic risk score calculation – a score reflecting the sum of all known risk alleles, weighted by how risky each variant is known to be to assess the genetic susceptibility for a particular disease (234). The polygenic risk score for determination of some complex cardiovascular pathologies and some types of cancers has already proved useful in the detection of individuals at risk. Such information can serve clinicians to take the same preventive measures as they do for patients meeting other high-risk criteria (235). It is expected that artificial intelligence-based algorithms for identification of patients with disease risk factors, patients with the need for a guideline-based screening or need for the application of pharmacogenetic guidelines to assist with drug selection and administration will soon be developed (236).

SNPs are the most common traits tested in genetic association studies. Simple sequence repeats (microsatellite markers), insertion/deletions, variable-number tandem repeats (VNTRs), and copy-number variants (CNVs) are other structural variants used for identification of genomic alterations involved in complex diseases (233).

Association studies are most often designed as a case-control study or quantitative trait approach.

- **Case-control study** is based on stratification of the population in two samples: "cases" and "controls", which are compared to detect variants with a different allelic frequency among two studied groups. It is probably one of the most common approaches used to study the genetic association (237). A Pearson test (2 df) and a Fisher exact test are the most common statistical methods to analyse the associations of a single SNP in case-control populations (238).

- **Quantitative trait approach** uses disease-associated quantitative traits (also known as intermediate phenotypes) as proxy disease phenotypes in molecular genetic investigations of complex diseases. Intermediate phenotypes are traits with normally distributed phenotypic variation in natural populations. Normally, they describe either morphology (height, weight), physiology (blood pressure), behaviour (aggression), or molecular phenotypes (TL, TNF-α levels, gene expression levels) (6, 7). Quantitative trait locus (QTL) analysis is a statistical method that links phenotypic data (trait measurements) and genotypic data (usually molecular markers) in an attempt to explain the genetic basis of variation in complex traits (239). The natural statistical tools for quantitative traits are linear regression and analysis of variance (ANOVA). ANOVA compares the null hypothesis of no association with a general alternative, whereas linear regression achieves a reduction in degrees of freedom from 2 to 1 by assuming a linear relationship between the mean value of the trait and genotype (238).

Most common types of association studies include candidate polymorphisms, candidate gene, fine mapping and genome-wide studies (238). Candidate polymorphism study focuses on one individual SNP that is suspected to be related to a disease. Linkage and candidate gene studies were mainly used to study the genetic basis of diseases before the mapping of SNPs across the

human genome and the realisation that a subset of these SNPs can capture ("tag") common genetic variation via linkage disequilibrium, which led to the development GWAS (233).

In GWAS, millions of commonly varying polymorphisms are genotyped to detect variants specific for the particular phenotype (such as a molecule or a disease). The variation of the entire genome is investigated to identify genetic variants that are related to the phenotype of interest. It is thus considered as a non-candidate gene driven approach (233). GWAS became a powerful tool for the investigation of common complex diseases in a population (Figure 20). In the years 2005-2018 there were 3639 GWAS reported, which investigated altogether 3508 traits (240).

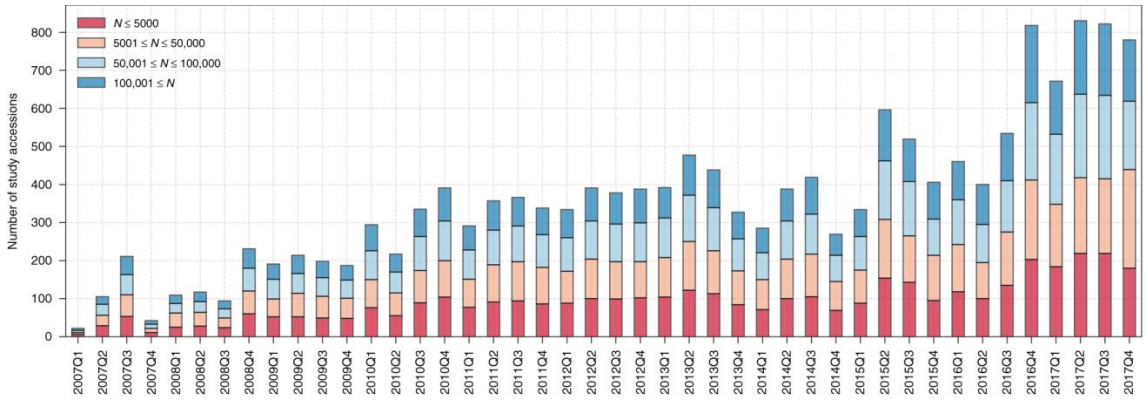


Figure 20: The increasing number of GWAS studies and sample sizes in years 2007-2017 (240).

The general statistical procedure in GWAS covers preliminary analyses, including 1) testing for **Hardy–Weinberg equilibrium** (HWE), to select a good subset of available SNPs. 2) **imputation of missing genotype data**, to replace missing genotypes with predictive values. 3) **inferring haplotype from genotype data** to generate more detailed information. 4) measures of **linkage disequilibrium** and **estimates of recombination rates**, to decrease the number of SNPs genotyped, and 5) choosing **tag SNPs** to retain the genetic information of the whole SNP set with the use of a minimal number of SNPs. The statistical analysis of the genetic data for a chosen phenotype in the study population is performed afterwards (238).

Epigenetic association studies are also becoming popular methods in the search for an association between epigenetic variation, such as DNA methylation (DNAm) or histone modifications, and disease risk. They examine the result of the environmental impact on

genetic changes without measuring individual exposures (241). The bases of association studies will be presented in a subchapter below.

1.4.1.2 Meta-Analysis

Meta-analysis is a set of statistical techniques that combines the results of multiple scientific studies in a mathematically appropriate way. By this, it creates a single, more precise estimate of an effect by increasing statistical power. Meta-analysis allows generation of useful results from many trials, which are individually too small to lead to consistent conclusions. Small trials are often underpowered, leading to larger standard deviations and standard errors. It is estimated that large trials and meta-analyses should give the same results in the absence of bias (242).

Meta-analysis can combine any quantitative data, including randomised controlled trials, observational studies, diagnostic test accuracy studies, and prevalence studies. A combination of data from participants across multiple data sets in the form of a meta-analysis is often used in contemporary GWAS (240, 243). It can increase the chances of finding true positives among the identified associations. However, some factors can influence the results of GWAS meta-analysis, such as heterogeneity of data, different sample size and design or utilisation of computed data. The integration of the data during the performance of meta-analysis should be therefore done systematically and rigorously, and the effect of heterogeneity should be questioned to avoid biased outcomes of the analysis (244).

1.4.1.3 The study of sequence variants

The significant variants identified in association studies have to be thoroughly investigated to determine 1) if the variants are causal, 2) what are the molecular functions of the causal variants, 3) which genes are affected by the causal variants, and 4) how changes in the function or regulation of the causal genes lead to altered disease risk (245).

The classical approach for identification of the genetic alteration of hereditary disease is the sequence of causative genes. The functional characterisation of the protein carrying the variant is the only direct criterion to demonstrate the pathogenicity of a variant (232). Nevertheless, bioinformatics analysis with the help of numerous databases (*p.ex.* Ensembl, dbSNP) can help to retrieve information about a variant and provide a list of previous studies

where the same variant was reported, if existing. Moreover, it can assist in the identification of the mutation type of the variant and in the understanding of its causative effect (232). Variants commonly identified variations include:

- **Synonymous mutations,** which affect only DNA or mRNA sequence, but do not cause an apparent change in the protein. They can modify the regulatory mechanisms at the basis of gene expression. Any change in the nucleotide sequence can lead to splicing alterations or to mRNA instability caused by alterations of secondary structure or by altered binding of miRNAs, resulting in decreased protein expression.
- **Splicing,** which affects regulatory mechanisms leading to intron retention or exon skipping. Each intronic, as well as an exonic variant, should be assessed for its potential effects on splicing.
- **Missense mutations**, which affect a change of a single amino acid in a protein. It is not deleterious if the affected amino acid is not included in the functional domains of the protein or if it is not essential in the protein folding, which is evaluated using the multiple alignment algorithms.
- **Deletion or insertion without reading frameshift,** which induces a deletion or an insertion of few amino acids and should be studied with respect to the conservation of involved region and possible alteration of protein structure.

1.4.1.4 Mendelian Randomisation

Epidemiological studies provide an abundance of information on associations between exposures and outcomes, but they cannot indicate the causal relations due to limitations introduced by confounding and reverse causality (246). Mendelian randomisation (MR) is an established genetic epidemiological approach that uses genetic variants (*e.g.* SNPs) as instrumental variables to assess causal relationships between exposures and outcomes (*e.g.* disease). It reports the association of the genetic variants directly associated with exposure, and the outcome (Figure 21). The genetic variant used in the study should only affect a single pathway on which the exposure of interest lies, and should not be associated with confounders, which impact on exposure and disease. As such, MR can provide unbiased evidence as to whether an exposure causes the investigated outcome (246).

MR is based on three key assumptions:

1. The genetic variant is reliably associated with the biomarker (exposure).
2. The genetic variant is not associated with any confounding factor (U).
3. The genetic variant only influences the risk of disease through the biomarker of the interest.

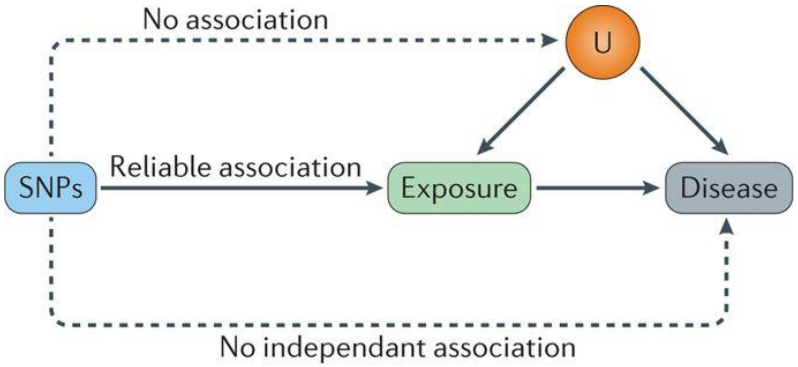


Figure 21: A schematic presentation of MR (247).

MR presents a valuable tool, especially when randomised controlled trials for the examination of causality are not feasible and observational studies provide biased associations because of confounding or reverse causality. It is widely used because it can overcome a major limitation of evidence from conventional observational studies: unmeasured confounding. Genetic variants are randomly allocated at conception and not subject to reverse causation. MR can thus support causal inferences about the effects of modifiable risk factors, which can overcome some types of confounding (248, 249). MR is now commonly carried out to identify potential targets for new drugs before performing a costly randomised controlled trial (250).

1.4.2 Epigenetics

Epigenetics is the study of heritable changes, which can modify the outcome of a locus or chromosome through modification of gene expression, without changing the underlying DNA sequence. Epigenetic programming is a dynamic process, influenced by environmental exposures during development, and includes several covalent or noncovalent modifications of DNA and histone proteins (Figure 22). The major epigenetic mechanisms, *i.e.* DNAm, histone modifications, chromatin remodelling, and small noncoding RNA (miRNA)-associated gene silencing are involved in complex interactions with each other and with classical transcription factor-based gene regulation (251-253).

Epigenetic modifications are critical for the regulation of genes and noncoding RNA expression. In turn, disruption of one of these systems can lead to inappropriate expression or silencing of genes, resulting in alterations in health and disease. Technological breakthroughs enabled large-scale epigenomic studies, which can help to learn how epigenetic marks and mutations are involved in pathophysiological mechanisms. Epigenetic alterations have been demonstrated in many diseases, primarily in cancer, but also in CVD, metabolic diseases and others. The reversibility of epigenetic changes and the discovery of numerous agents that alter methylation patterns or the modification of histone have encouraged the development of new therapeutic options for treatment of such diseases. This resulted in extensive studying in epigenetics for a large spectre of medical applications (254, 255).

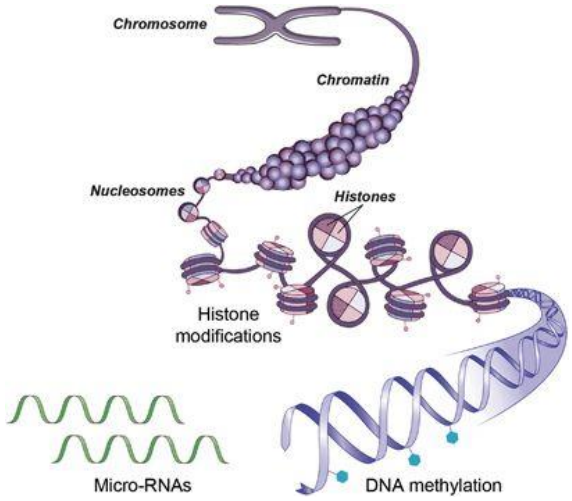


Figure 22: Major epigenetic mechanisms: chromatin changes, caused by DNA wrapping around a core octamer of histone proteins, DNAm at cytosine rings and miRNAs, involved in post-transcriptional control of gene expression (253).

1.4.2.1 DNA methylation

DNAm is the most widely studied epigenetic modification, which involves the covalent transfer of a methyl group to the carbon-5 position of the cytosine ring of DNA, catalysed by DNA methyltransferases (DNMTs) (Figure 23). It occurs almost exclusively, in more than 98% of cases, in the context of CpG dinucleotides, which tend to clusters in regions called CpG islands (256-258). CpGs are regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide in the linear sequence of bases.

DNAm is a key mechanism in epigenetic silencing of transcription and is essential for normal mammalian development, cellular differentiation and adaptation to environmental signals

(259). This process is not static but undergoes precise, highly coordinated changes that can be mediated both by environmental and genetic factors and inherited through mitotic cell divisions (17). DNA demethylation can appear actively or passively *via* oxidation of 5-methylcytosine by ten-eleven translocation (TET) family enzymes, which has been observed in different biological contexts. It is necessary for epigenetic reprogramming of genes and is directly involved in many important disease mechanisms (260).

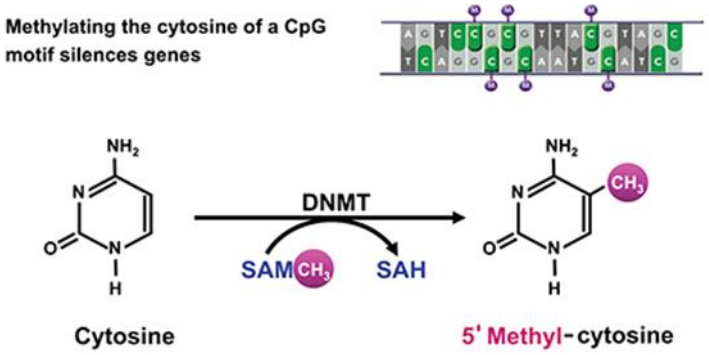


Figure 23: DNA cytosine ring is methylated by enzymes methyltransferases to form 5-methylcytosine at CpG sites. (Source: Dnaofbioscience).

DNAm has shown good potential for use as a biomarker of specific phenotype if it leads to a change of gene expression that induces that phenotype. Methylation biomarkers can be used at all stages of clinical disease management, as 1) **risk assessment biomarkers** (exposure to adverse environmental factors induces methylation changes long before disease appears), 2) **early disease detection biomarkers** (methylation changes occur early or they initiate a disease *e.g.* neoplastic transformation), 3) **clinical disease management biomarkers** (methylation changes can have prognostic significance for identification of individuals, likely to respond to intervention), 4) **post-treatment monitoring/surveillance biomarkers** (methylation changes can reflect a pathology, *e.g.* primary tumour in metastasis) (14). The use of epigenetic biomarkers has been clinically applied as a screening test for colorectal cancer in 2016 and is being tested for use in the diagnosis and management of ovarian cancer (261, 262).

As mentioned previously, epigenetic changes, especially DNAm, can be interesting also for therapeutic applications. DNAm inhibitors have become the mainstay for the treatment of certain haematological malignancies and show great potential for other diseases that have been caused by impaired epigenetic systems (18, 19). Currently, there are still no DNAm-based diagnostic biomarkers or treatments for CVDs, and further researches are warranted to

get a deeper understanding of the mechanisms underlying CVDs and enable the development of effective interventions for human cardiovascular disorders, based on DNAm (263).

1.4.2.2 Epigenome-wide association studies

EWAS are large-scale studies of human disease-associated DNAm variation, which enable identification of methylation markers that are specific for a disease or a particular disease risk factor (264). Genome-wide methylation profile of individuals is assessed through high throughput technology. Currently, Infinium HumanMethylation450 is the most commonly used array, which assesses the methylation status for >450 000 CpG sites located throughout the genome. Differential methylation analysis is performed using common statistical procedures after the required processing of methylation data (filtering of problematic probes and normalisation) (265).

EWAS studies present few challenges in the interpretation of the results, issued from the facts that DNAm patterns can be tissue-specific and are not static, but can change over time. Furthermore, EWAS associations can be causal as well as consequential for the phenotype of interest. Determination of causality in EWAS is critical as it helps elucidate the functional role of the phenotype-associated methylation site and potential utility in terms of diagnostics or therapeutics (264).

1.4.3 Gene expression studies – transcriptomics

Gene expression is a conversion of genetic information from DNA into a functional gene product (RNA or protein). It occurs in a strongly regulated two-step process: 1) **transcription of DNA to RNA**, catalysed by RNA polymerase enzymes, followed by processing of resulted mRNA molecule (Figure 24), and 2) **translation of RNA**, followed by subsequent post-translational processing of the protein molecule to provide a functional protein (266).

Gene regulation is a cellular process that controls the rate and manner of gene expression. It comprises a complex set of interactions between genes, RNA molecules, proteins (including transcription factors) and other components of the expression system. More particularly, it determines the rate of translation, regulates the processing of RNA molecules, including alternative splicing to produce more than one product from a single gene, regulates the

stability of mRNA molecules and the rate of translation. Constitutive gene expression is present in all cell types to enable the basic cellular function provided by the housekeeping genes. Other genes are expressed only in a certain cell type, which permits a cell to specialise and perform a particular function (267).

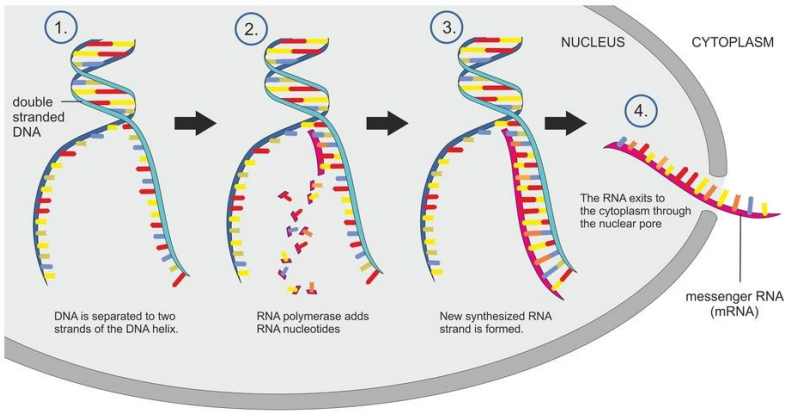


Figure 24: Transcription of DNA as a part of gene expression. (Source: Shutterstock).

Gene expression analysis enables to study the occurrence or activity of the formation of a gene product from its coding gene. It is a sensitive indicator of biological activity. A changing gene expression pattern reflects a change in biological process (268). Transcriptomics is the study of the transcriptome - the complete set of RNA transcripts that are produced by the genome. It is performed under specific circumstances or in a specific cell using high-throughput methods, such as RNA-Seq microarray analysis. Comparison of transcriptomes allows the identification of genes that are differentially expressed in distinct cell populations, or in response to different treatments (269).

Most common techniques used to measure and detect transcription activity are the following:

- **RT-PCR:** a sensitive *in vitro* method used for detecting and comparing the levels of mRNA and the surface proteins. In RT-PCR, the RNA molecule is first converted into a complementary DNA (cDNA) using reverse transcriptase. The cDNA is then used as a template for exponential amplification using PCR. It can be performed as real-time PCR for relative or absolute quantification, or as end-point PCR (270).
- **Northern blot:** a widely used procedure for analysing the molecular size and abundance of mRNA among a mixture of RNA. It can allow the analysis of a sample of RNA from a

particular tissue or cell type to measure the RNA expression of particular genes. The isolated RNA is denaturated and loaded on an agarose gel and separated by electrophoresis (271).

- **RNA-Seq:** a combination of next-generation sequencing techniques with computational methods to sequence whole transcriptomes. It allows the measurement of relative abundance of RNA, as well as the detection of additional variations such as fusion genes, post-transcriptional edits and novel splice sites (272).
- **DNA microarrays**: a technique used for gene expression profiling, which consists of short nucleotide oligomers, known as "probes," that are arrayed on a solid substrate. Transcript abundance is determined by hybridisation of fluorescently labelled transcripts to these probes (272).
- **In situ hybridisation**: a technique used for the high-resolution detection, quantification, and localisation of nucleic acid targets inside cells or tissues. It is based on the hybridisation of sequence-specific fluorescently labelled complementary probes (typically DNA sequences) to their target inside the cell (273).
- **Run-off transcription assay:** *in vitro* assay to identify transcription start sites (TSS) of a specific promoter. It can be used to quantitatively measure the effect of changing promoter regions on *in vitro* transcription levels (274).
- **MS2 tagging:** a technique used to monitor the appearance of RNA in living cells, at the site of transcription, or simply by observing the changes in RNA number in the cytoplasm The technique is based on the incorporation of a stem-loop structure into a gene, which then incorporates the structure also into newly synthesised RNA. The stem-loops can be detected using a fusion of GFP and the MS2 coat protein (275).

1.4.4 Challenges of personalised medicine

To summarise, personalised medicine is using complex approaches to form a medical model that combines characterisation of individuals' phenotypes and genotypes to provide a healthcare system which will better respond to patient needs.

To do so, it has to act in different time steps to provide appropriate biomarkers: to assess risk or clinical progression of a disease, to predict diagnose and support the therapeutic decision (Figure 25). Moreover, it assists in development of medicaments tailored by individual profile to provide appropriate medical intervention.

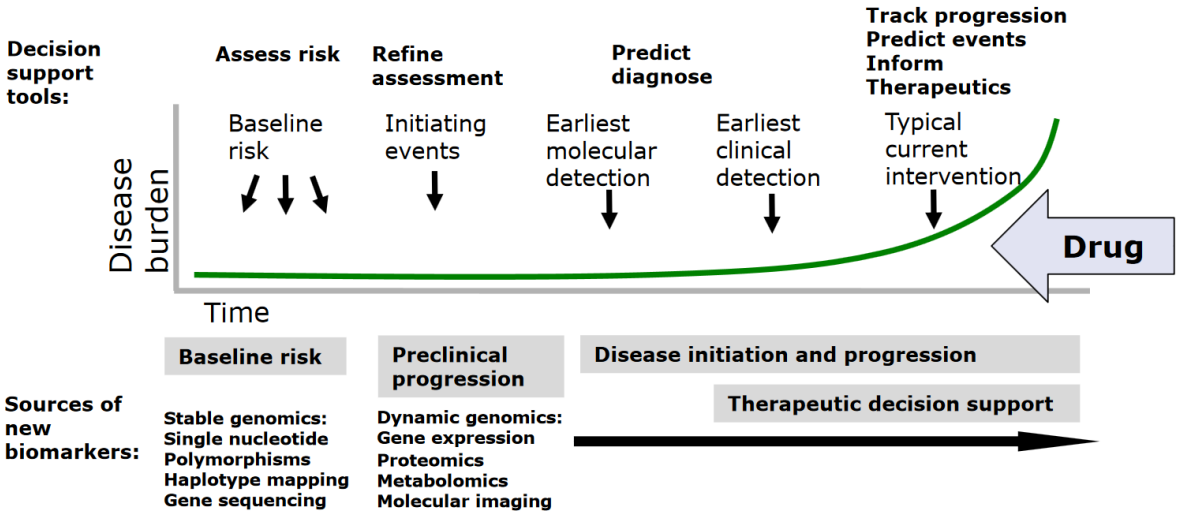


Figure 25: Tools needed for prediction and personalised care (276).

For the successful integration of new scientific results into public health, personalised medicine has to overcome five main challenges to which it is continuously confronted (277):

- Developing awareness and empowerment
- Integrating big data and ICT solutions.
- Translating basic to clinical research and beyond.
- Bringing innovation to the market.
- Shaping sustainable healthcare.

Chapter II. HYPOTHESIS AND OBJECTIVES

Common chronic diseases are complex traits, affected by the effects of multiple genes in combination with lifestyle and environmental risk factors. Thanks to the development of new technologies, personalised medicine enabled to improve their preventive, predictive and curative strategies. New scientific discoveries are now needed for further identification and characterisation of biomarkers, elucidation of molecular regulatory mechanisms and proposition of therapeutic targets that will serve in the application of personalised healthcare.

Genetic and molecular biomarkers involved in multiple pathways play a key role in the development of chronic diseases. The identification of common pathways and interactions could thus ameliorate our understanding of diseases' pathophysiology and would assist in discovering novel biomarkers, "common denominators" for ageing diseases and comorbidities.

GWAS studies identified a large number of variants related to specific phenotypes. We have hypothesised that a genetic variant associated with a biomarker of interest, which would show association with another biomarker involved in the same pathology may help in the identification of shared biological pathways between molecules. Moreover, such genetic variants could be considered as novel genetic biomarkers, useful in the risk assessment of comorbidities.

Studies of the epigenetics are emerging in the last few years, and are bringing a big potential for predictive and preventive personalised therapies. EWAS allow detection of methylated CpG sites that play an important role in the regulation of intermediate phenotypes. Because methylation can represent an early modifiable change of DNA, caused by unfavourable environmental conditions, CpG sites associated with a particular risk factor can be used as susceptibility biomarkers for all related pathologies. On the other hand, if the CpG site is involved in the regulation of gene involved in the aetiology of a disease, such methylation mark represents a potential therapeutic target.

In the investigation of molecular biomarkers, we primarily focused on various molecules implicated in common pathophysiological processes, such as inflammation and angiogenesis. Besides common inflammatory cytokines, we were particularly interested in TREM2, VEGF-A and TL. VEGF-A and TL are the perfect examples of biomarkers, involved in multiple pathophysiological mechanisms. We thus assumed that, combined, they could present a powerful tool for personalised medicine applications.

Furthermore, a collaboration between research and industry presents a key element for the progress in the implementation of personalised medicine, the fastest way to transfer the knowledge from a research laboratory to a clinical application. For this reason, we believe that active participation in industrial projects aiming to improve diagnostics and treatment of common chronic diseases, and serving to the field of personalised medicine is indispensable.

To summarise, the goal of this thesis was to identify biomarkers, common for different chronic diseases through investigation of association between genetic variants and intermediate phenotypes and application of "-omics" methodologies, to finally propose a combination of biomarkers, which would improve the disease risk prediction and treatment in a useful way for personalised medicine.

Three main objectives were pursued in this thesis work:

- To highlight new variants associated with the risk factors for common chronic diseases.
- To find molecular and genetic associations between biomarkers involved in shared biological pathways.
- To summarise the existing knowledge from our research field and offer it to the next generation of researchers.

The specific objectives set to develop a multiple-biomarker approach in personalised medicine are:

1. To summarise the greatest historical achievements in personalised medicine.

2. To investigate the association between DNAm (CpG sites) and intermediate phenotypes such as blood lipids levels and waist circumference, in a healthy population, thus to discover underlying mechanisms linking the DNA methylation with blood lipids metabolism and central obesity, respectively.

3. To investigate the association between DNAm (CpG sites) and VEGF-A levels in a healthy population, thus to discover underlying mechanisms linking the DNA methylation to angiogenesis.

4. To identify common physiological pathways among VEGF-A, TREM2, TL and other inflammatory cytokines with genetic association studies.

5. To evaluate the role of TL in personalised medicine and its potential in diagnostics, prognostics and treatment.

6. To be introduced in the working environment and collaboration with clinical diagnostics industry in the field of application of research results to personalised medicine.

Chapter III. MATERIALS AND METHODS

3.1 Populations

3.1.1 The STANISLAS Family Study

The STANISLAS Family Study (SFS) is a familial longitudinal cohort, aiming to study the role and the contribution of genetic and environmental factors to cardiovascular status (278, 279). Laboratory and demographic data were collected for 1006 families (two parents and at least two children aged 4 or more) from the East of France, in three recruitments during ten years. Participants were enrolled in the Center for Preventive Medicine (CMP) of Vandoeuvre-lès-Nancy, during a periodical health assessment. All the individuals included in the research were of European-Caucasian origin and supposed healthy, free from any declared acute or chronic disease. More than 600 data variables were collected on each individual with the help of trained professionals, including clinical data and lifestyle information, provided by questionnaires. Measurements were performed in the CMP and other collaborating research laboratories. Waist circumference was taken at the midpoint between the lower margin of the last palpable rib and the top of the iliac crest (hip bone). Measurements were recorded to the nearest 0.1 cm by trained professionals. BMI was calculated by the Quetelet index formula as weight (kilograms) divided by height² (meters)². Study protocols were approved by the institutional ethics committee CCPPRB de Lorraine (Comité consultatif de protection des personnes dans la recherche biomédicale) and CNIL (Commission Nationale de l'Informatique et des Libertés). All individuals gave written informed consent for their participation in the study. The SFS is part of the BRC IGE-PCV with identification number BB-0033-00051, which was initiated by the UMR INSERM U1122; IGE-PCV.

3.1.2 Hypertensive population

Hypertensive population available at the BRC IGE-PCV consists of hypertensive individuals (n=945) without any present diagnostics of CVD or other chronic diseases, and not treated with any antihypertensive therapy. This population has been enrolled using similar standardised protocols and methods for the collection of samples, data and biological measurements as those applied in the SFS cohort.

3.1.3 The MuTHER study

The "MuTHER" ("Multiple tissue human expression resource") study is a project under the Wellcome Trust Foundation, aiming to develop a resource of detailed genetic (genome-wide association, resequence) and genomic (expression, methylation) data from a range of tissues collected from UK twins, aiming to support efforts to understand mechanisms involved in common traits susceptibility (280). Lymphocytes, subcutaneous fat, muscle and skin biopsies have been obtained from 856 twins of European ancestry, from the well-characterised TwinsUK BioResource. All samples and information were collected with written and signed informed consent. The project has been approved by the local ethics committees of all institutions involved (280).

3.1.4 The TELARTA study

Telomeres and Arterial Ageing (TELARTA) is a case-control study, aiming to examine the links of arterial ageing and development of atherosclerosis with TL, as expressed in different tissues (leukocytes, muscle, skin, progenitor cells), and telomere length dynamics, expressed as a difference between TL of muscle and leukocytes. Muscle biopsies of patients with clinically significant atherosclerotic disease and controls without evidence of clinical CVD or history of previous CVD were collected during undergoing surgery or pacemaker/defibrillator implantation. Altogether, 131 atherosclerosis cases and 128 controls were enrolled in two French university hospitals, in Nancy and Marseille. All participants provided written informed consent approved by the Ethics Committee (Comité de Protection des Personnes) of Nancy, France. Subjects with BMI >40 kg/m^2, a glomerular filtration rate under 30 mL min^{-1} 1.73 m^{-2}, active malignancy, or history of chemotherapy/radiotherapy for cancer and subjects with aortic aneurysm were excluded from discovery cohort.

Replication cohort of the TELARTA study was composed of additional 91 individuals, enrolled in the university hospitals of Nancy and Marseille, and 52 individuals enrolled at 3 Athens hospitals (Onassis Cardiac Surgery Center, Iaso General Hospital, Hippokration Hospital). Moreover, the inclusion criteria for the replication cohort were less strict, and individuals that were excluded from the TELARTA discovery cohort based on BMI >40 kg/m^2, glomerular filtration rate under 30 mL min^{-1} 1.73 m^{-2}, malignancy or history of chemotherapy/radiotherapy were included in the replication cohort. Participants enrolled in France provided written informed consent approved by the Ethics Committee (Comité de

Protection des Personnes) of Nancy, France. Those enrolled in Athens provided written informed consent approved by the Ethics Committee of the University of Athens and Ethics Committee of each one of the three participating hospitals (81).

3.2 DNA methylation assessment and genotyping

Blood samples of all individuals from the SFS were taken after overnight fasting and were collected in EDTA vacuum tubes. Whole blood DNA was extracted by the Miller technique (281) and was stored on -80 °C in BRC IGE-PCV.

Punch biopsies of subcutaneous adipose tissue from a photo-protected area of the stomach adjacent and inferior to the umbilicus were obtained from 856 healthy female twins who participated in the MuTHER cohort. Skin from the punch biopsy was then dissected to separate it from adipose tissue, and both samples were weighed and immediately frozen with liquid nitrogen (282).

3.2.1 DNA methylation assessment

Blood DNA methylation assessment was performed by Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA). Illumina is using Infinium I and II arrays with probes for detection of methylated and unmethylated CpG sites. Methylation ratio referred to as beta value by Ilumina's software, is the proportion methylated / (methylated + unmethylated) for each CpG in the population of cells from which DNA was extracted (265, 283). A detection p-value was generated for every CpG with minfi R package.

Methylation levels were also estimated in subcutaneous adipose tissue-derived from the female twins enrolled in the MuTHER study. Methylation profiling was performed with Illumina's Infinium HumanMethylation450 BeadChip. Arrays were scanned with the IlluminaHiScan SQ scanner, and raw data were imported to the GenomeStudio v.2010.3 software with the methylation module 1.8.2 for the extraction of the image intensities.

3.2.2 Genotyping

Whole-genome genotyping of the SFS was performed using Illumina® human CNV370-Duo array following the manufacturer's protocols. Genotypes were called using Illumina® protocol for BeadStation genotyping solutions, based on the GenCall® software application to automatically cluster, call genotypes, and assign confidence scores using the GenCall® clustering algorithm GenTrain.

Genotyping of SNPs in the SFS, hypertensive population and TELARTA study was performed by the Laboratory of the Government Chemist (LGC Ltd., Teddington, UK), using a PCR-based KASP assay, which enables bi-allelic scoring of SNPs and insertions and deletions (Indels) at specific loci (284).

3.3 Gene expression analysis

3.3.1 PBMC isolation

Peripheral blood mononuclear cells (PBMCs) of individuals of the SFS were isolated using the density gradient centrifugation method, first developed by Bøyum in 1968 (285). Briefly, whole blood was collected in tubes containing sodium heparin. Hanks' Balanced Salt Solution (SIGMA Aldrich, reference: H6648) was added into 15 mL tubes with blood and poured gently into a 15 mL tube with Ficoll paque™ plus solution (Sigma Aldrich, reference: 17-1440-02). Ficoll paque medium is denser than lymphocytes, monocytes, and platelets (meaning these will remain above it), but less dense than granulocytes and erythrocytes, which will drop below it. The contents of the tube were centrifuged for 30 min at 300 g at room temperature. A PBMCs ring was retrieved and collected into a 15 mL tube, filled with Hanks' Balanced Salt Solution and centrifuged for 10 min at 1000 g at room temperature. The supernatant was aspirated and 2 mL of Hanks' Balanced Salt Solution was added. The solution was well suspended, filled up to 15 mL with Hanks' Balanced Salt Solution and centrifuged for 10 min at 1000 g at room temperature. The PBMC ring was collected into Eppendorf tube with 1 mL of Hanks' Balanced Salt Solution. PBMCs populations were evaluated by microscopic observation after May-Grunwald-Giemsa staining, and the PBMCs concentration was normalised to 10^6 cells/mL in Hank's Buffer. After final centrifugation for 5 min at 1000 g at room temperature, the supernatant was aspirated, and the pellet of PBMCs was processed immediately or stored at -80 °C.

3.3.2 mRNA quantification

Total mRNA was extracted and quantified from isolated PBMCs using the MagNA Pure LC RNA HP isolation kit and RNA HP Blood External lysis protocol (Roche Diagnostics, France) as previously described (286, 287). Briefly, 200 units of M-MuLV Reverse Transcriptase with 0.25 µg of oligos (dT) (Promega, France) were used to perform reverse transcription of total RNAs. Quantification of the transcripts coding for the VEGF-A isoforms ($VEGF_{121}$, $VEGF_{145}$, $VEGF_{165}$, $VEGF_{189}$) was performed using TaqMan® and LightCycler technologies (LC TaqMan Master Kit, Roche Diagnostics, France). All experiments were performed in duplicate. RT-PCR optimisation and specificity of Real Time-PCR products were conducted using SYBR® Green technology (LC FastStart DNA MasterPLUS SYBR Green I kit, Roche Diagnostic, France), melting curves analysis and agarose gel electrophoresis of the PCR amplicons. All mRNA levels were normalised to the mRNA levels of β2M gene.

3.4 Protein measurement

3.4.1 Total protein extraction from PBMCs

The lysis solution composed of cell lysis buffer (CelLyticTM-M, SIGMA Aldrich, reference: C2978) and protease inhibitor (0.5 %, Protease Inhibitor Cocktail, SIGMA Aldrich, reference: P8215) was added to the PBMC pellet, as recommended by the manufacturer (SIGMA Aldrich). It was stirred for 15 min at room temperature and centrifuged for 15 min at 12000 g at 4 °C. The supernatant was collected and was immediately used for further analysis or stored at -80 °C.

3.4.2 Protein quantification

Quantification of inflammatory molecules (IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, MCP-1, TNF-α, INF-γ, EGF and VEGF-A) from PBMCs extracts of individuals from the SFS and plasma samples of TELARTA population was performed by Randox high sensitivity multiplex cytokine and growth factor array (Evidence Investigator Analyzer, Randox Laboratories Ltd., Crumlin, United Kingdom). Evidence Investigator is a Randox patented 9 x 9 mm^2 activated biochip with spatially discrete test regions containing antibodies specific to each of the inflammatory molecules assessed.

3.5 Lipid measurement

For the SFS cohort, total serum cholesterol (TC) and serum TGs were measured using standard enzymatic methods (Merck, Darmstadt, Germany), automated on AU5021 (Olympus; Merck). Serum HDL was measured by phosphotungstate precipitation on a Cobas-Mira (Roche, Basel, Switzerland), while LDL levels were calculated using the Friedewald formula (288).

3.6 Telomere length measurement

TL in the samples of TELARTA study was measured in DNA extracted by the phenol/chloroform method from peripheral blood leukocytes and skeletal muscle in the surgical field. Measurements were performed in duplicate by Southern blots of the terminal restriction fragments, as previously described (81). The analysis was performed in the laboratory UMR INSERM U1116 DCAC.

3.7 Statistical analysis

Many statistical tests rely on the assumption that the residuals of a model are normally distributed. Therefore, values of studied phenotypes, such as lipid levels, waist circumference and inflammatory molecule levels were tested for normality. Log-transformation of the dependent variable was applied to satisfy the normality assumption if values were not normally distributed.

3.7.1 Methylation analysis

A linear mixed-effects model was used to analyse the association between methylation levels and phenotypes (lipid levels, waist circumference and VEGF-A levels) in the SFS population, accounting for relatedness between individuals. The model used was adjusted for gender, age, BMI, family structure, and other relevant covariates. Bonferroni correction and false discovery rate (FDR) methods (<0.05) were used for the correction of the results for multiple testing.

3.7.2 Correlation and association analyses

Linear regression models were applied to study associations between genes and intermediate phenotypes, using different inheritance models (additive, dominant and recessive model). Mixed models adjusted for the family structure were applied to the SFS population, to correct for possible issues of stratification due to familial resemblance. The models were adjusted for age and sex. Analyses were performed using the R packages GWAF and PLINK. Non-parametric correlation analyses (Spearman correlation) were performed to study links between VEGF-A and different inflammatory molecules.

3.8 Characterisation: expression of recombinant antibody fragments

Recombinant antibody fragments acquired using phage display technology were characterised for their specificity and stability, using a standardised protocol of Randox ltd., in Crumlin, UK. Briefly, extracted plasmid DNA (miniprep) of antibody fragment was mixed with WK6 cells (*E. coli*) and heat-shocked at 42 °C to allow the transfer of DNA into a bacterial cell. Transformed bacteria were spread over the agar plate containing ampicillin and streptomycin and incubated overnight. The contents of the grown agar plate were re-suspended into the 10 mL culture of Terrific broth (TB) medium heated on 37 °C. When the optimal growth was achieved, the culture was raised into 100 mL and 1 L, respectively, and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) for the overnight incubation. Cells were collected the next day and lysed to extract the synthesised protein. A sample of cell lysate was serially diluted and added on enzyme-linked immunosorbent assay (ELISA) plates coated with the antigen of interest and control antigens, to test the specificity of the isolated antibody. The rest of the collected protein was purified using Talon resin on gravity columns. The final concentration of the protein in the sample was measured using UV spectrophotometer. Samples of purified protein were stored on three different temperatures: 37 °C, 4 °C and -20 °C. Proteins that showed good specificity for the antigen were tested for stability after a week of storage. Samples from three different temperatures were prepared in 8 concentrations and were added on ELISA plates coated with the antigen of interest. Dilution curves of three temperatures were compared to assess the stability of the antibody.

Chapter IV. PUBLICATIONS - RESULTS

Publication Nº 1

Personalised Medicine: The Odyssey from Hope to Practice.

Sophie Visvikis-Siest[*], **Vesna Gorenjak[*]** and Maria G. Stathopoulou[*]

[*]equal contribution

J Pers Med, August 2018, 21;8(4)

The review article '*Personalised Medicine: The Odyssey from Hope to Practice'* was designed as an introduction to the most significant achievements in the field of personalised medicine, which lead the way to today's innovative and perspective healthcare. The review is an original mixture of the ancient Homer's poem '*The Odyssey'* together with the scientific facts, which are enrolling throughout the travelling of the poem's main hero Odysseus.

The article starts with the first and the most important achievement that enabled the expansion of genetic studies – sequencing of the whole genome within the HGP. Further on, it presents the technological and methodological development, discusses the practical application of the acquired knowledge in clinics and pharmacological studies, and summarises the ambitious international projects, which made a step forward in the understanding and use of genomics in everyday clinical and research practice. The article is addressed to the wide audience interested in learning about breakthrough discoveries, as well as to the health experts who seek an entertaining and educative reading of a general medical topic.

The review article was introduced during a biannual international conference on personalised medicine: *"Systems Medicine Personalised Health and Therapy – The Odyssey from Hope to Practice"*, which took place in Santorini, Greece, on 30th September - 3rd October 2018 and was organised by the Unit UMR INSERM U1122; IGE-PCV.

The review article *Personalised Medicine: The Odyssey from Hope to Practice* is a suitable introduction to the results of the thesis, as it shows the importance of investing to the research work in the emerging field of personalised medicine.

Journal of *Personalized Medicine*

MDPI

Review

Personalised Medicine: The Odyssey from Hope to Practice

Sophie Visvikis-Siest [1,2,*,†], Vesna Gorenjak [1,†] and Maria G. Stathopoulou [1,†]

[1] Inserm, IGE-PCV, Université de Lorraine, 54000 Nancy, France; gorenjak.vesna@gmail.com (V.G.); maria.stathopoulou@inserm.fr (M.G.S.)
[2] Department of Internal Medicine and Geriatrics, CHU Technopôle Nancy-Brabois, Rue du Morvan, F-54511 Vandoeuvre-lès-Nancy, France
* Correspondence: sophie.visvikis-siest@inserm.fr; Tel.: +33-6-0760-2569
† These authors contributed equally to this work.

Received: 4 September 2018; Accepted: 18 September 2018; Published: 21 September 2018

Abstract: In this endeavour, inspired by the Odyssey, we aim to embark with the reader on a journey on a ship from Troy to Ithaca, coursing through the history of the momentous events and achievements that paved the way for personalised medicine. We will set sail amidst important genetic discoveries, beginning with the discovery of the first human genome, and voyage through the projects that contributed to the progress of pharmacogenomic studies. Concurrently, we will propose methods to overcome the obstacles that are slowing the potential full implementation of accumulated knowledge into everyday practice. This journey aims to reflect on the frontiers of current genetic knowledge and the practical use of this knowledge in preventive, diagnostic and pharmacogenomic approaches to directly impact the socio-economic aspects of public health.

Keywords: personalised medicine; pharmacogenomics; ethics; omics; Odyssey

1. The Journey from DNA Structure to Personalised Medicine

> *"Tell me, O Muse, of that ingenious hero who travelled far and wide after he had sacked the famous town of Troy. Many cities did he visit, and many were the nations with whose manners and customs he was acquainted . . . "* [1]

Decades ago, the study of genetics was similar in ardour and endurance to the conquering of the famous city of Troy, described in Homer's epic poem, "The Iliad". Every story has its heroes, and at the end of the 1950s, the heroes in genetics were James D. Watson and Francis Crick [2], the most well-known scientists related to the discovery of DNA structure and function. *"The horse has entered the fortress and the battle began."* [1].

Since then, important findings followed one another: the DNA code was determined [3], and transcription [4], translation [5] and replication [6] were described. Laboratory techniques required for such research rapidly developed and computer science became more accessible. When Frederick Sanger presented a new method for DNA sequencing [7], for which he was awarded the Nobel prize in chemistry in 1980 [8], understanding the genetic code became even easier and, thus, expectations rose. The next goal for scientists was genotyping the entire human genome.

The ambitious Human Genome Project (HGP) was launched in 1990 under the direction of the U.S. National Center for Human Genome Research, which is currently known as the National Human Genome Research Institute (NHGRI) [9]. The sequencing of three billion base pairs was performed in collaboration with 20 groups from the USA, the UK, Japan, France, Germany and China and was completed 13 years later, in April 2003 [10]. The HGP provided vast and comprehensive catalogues

of genomic information on biological structure and function and allowed the investigation of basic genome physiology. Most importantly, it was a landmark in medical research.

Constituting a new focus of interest for scientists were the millions of single nucleotide polymorphisms (SNPs), which represent differences in single DNA building blocks [11]. Indeed, millions of small variations in the human genome lead to different disease and treatment susceptibilities. However, relating SNPs to a disease demands extensive methodologies that require high technological standards.

Despite the difficulty of categorising the complex web of environmental factors that also contribute to inter-individual variability, improvements in genomics research permitted the characterisation of genetic variants that account for this variability to a large extent. Due to the early initiation of the HGP, progress in genotyping and sequencing techniques was rapid. The facilitation of gene identification made exploring the genetic background of diseases a new common practice and provided a flood of new information that led to tremendous progress in understanding the pathophysiological processes of diseases [12], thereby ushering in personalised medicine.

The initial intent of personalised medicine was to collect "-omics" data to provide new information about the mechanisms underlying disease and to thus identify new strategies for prediction, prevention and treatment to ensure personalised and participatory healthcare [13]. Although personalised medicine first appeared as a scientific term in published studies in 1991 [14], the concept originated long ago in the history of medicine, when scientists realised that certain treatments did not apply to all patients and might be harmful in some cases. One of the best examples is blood transfusions. The ABO blood group system was discovered in 1900 [15] and such knowledge helped prevent complications during transfusion through a simple blood type test that categorised each individual into different groups based on the ABO genotypes. This simple illustration demonstrates well the mission of personalised medicine today, which is to classify patients into different subgroups based on their genetic predispositions to allow for early disease risk prediction, precise diagnosis, accurate prognosis and the use of the most efficient and least harmful treatments for each individual.

Because of the very broad and general meaning of the term "personalised" medicine, some scientists and clinicians refer to it as "precision medicine" to emphasise the primary aspect of the field, which concerns targeted treatments constructed on the basis of the genetic, phenotypic and psychosocial characteristics of an individual patient, through which such an individual differs from other patients with the same condition [16]. We prefer the term "personalised" medicine because of the broadness of the expression, which refers to all important domains of the field.

This article explores the research journey towards personalised medicine by connecting it to the adventurous trip of Odysseus, from Troy to Ithaca (Figure 1). His journey might have seemed sorrowful at the first glance, but above all it was rewarding and meaningful. Hence, it is in personalised medicine, where while fighting numerous obstacles and deceptions we finally end up realising remarkable discoveries and projects, which will greatly contribute to the next generation of health systems.

Figure 1. The journey of Odysseus. 0: Troy; 1: Ismarus; 2: The land of the Lotus-eaters; 3: The land of the Cyclopes; 4: The Aeolian island; 5: Telepylus, the city of the Laestrygonias; 6: Aeagean island (Circe's island); 7: The Underworld (The house of Hades); 8: The island of the two Sirens; 9: Cliffs of Scylla and Charybdis; 10: The island of Helios; 11: Ogygia (Calypso's island); 12: Scheria (island of the Phaeaians); 13: Ithaca.

2. First Stop: Ismarus, the City of the Cicons and the Difficulties of Genomic Research

"When I had set sail thence the wind took me first to Ismarus, which is the city of the Cicons. There I sacked the town and put the people to the sword. We took their wives and also much booty, which we divided equitably amongst us, so that none might have reason to complain." [1]

The journey of personalised medicine began with the launching of the HGP, which commenced at a time when the general public was beginning to become familiar with intriguing genetic discoveries, as governments invested in "prospective studies" within this new emerging field. Much optimism and excitement accompanied the beginning of the "-omics" era, and many scientists were eager to participate, just like Odysseus and his soldiers after the Trojan war were excited to embark on the journey that would take them home after a long and exhausting battle.

The initial idea of personalised medicine was extremely ambitious. However, progress was not as fast as expected. Most common chronic diseases appeared to be regulated by a complex mixture of genetic factors and an infinite number of environmental factors, which were more difficult to elucidate than initially expected. Expensive "-omics" studies showed some success, though the clinical applications were few. Hence, criticism began to ensue regarding genomic research [17]. Questions arose such as whether the associated thought processes were all in error from the beginning, whether the money spent on research and development might never pay off and whether greed for success had derailed some researchers from their goals.

After the great victory in Troy, Odysseus, the great hero of the war, stopped on his way home on the island of Ismarus and, drunk of success, made a fatal mistake of letting his men destroy a town. Thus, the angry gods called the winds that misled his ships and took him on a long and dangerous journey that would detain him from reaching his own goal, Ithaca.

3. Second Stop: The Lotus-Eaters and the International HapMap Project

"I was driven thence by foul winds for a space of nine days upon the sea, but on the tenth day we reached the land of the lotus-eaters, who live on a food that comes from a kind of flower."
" . . . so delicious that those who ate of it left off caring about home, and did not even want to go back and say what had happened to them, but were for staying and munching lotus with the lotus-eaters without thinking further of their return . . . " [1]

The International HapMap Project represented the first substantial achievement in the determination of common genetic patterns of DNA sequence variation, which allowed for the identification of the sequence variants that affect common diseases and their frequencies and correlations. The project began in October 2002 as an international collaboration and was a natural extension of the HGP [18]. Whereas the HGP focused on the genome, which is invariant across individuals, the project's haplotype mapping (HapMap) focused on DNA sequence differences among individuals.

Common genetic patterns of DNA sequence variation, called haplotypes, are specific combinations of alleles on a chromosome. This combination is predictable because individuals who carry a specific SNP allele at one site often carry specific alleles at other nearby variant sites. This observation indicates that neighbouring SNPs on the same chromosome are in so-called linkage disequilibrium (LD) [19].

The HapMap of the human genome mapped and correlated SNPs based on their LD, which thus provided the information needed to guide genetic studies on clinical phenotypes. Phase I of the project, completed in 2005, provided publicly available data of the human genome sequence, databases of common SNPs, insights into human LD, web-based tools for storing and sharing data and frameworks to accelerate the identification of the genetic factors that influence medical traits. More than 1.1 million SNPs of 270 individuals across different global populations were genotyped during the project, which had a central role in developing the methods for designing and analysing genome-wide association studies (GWAS). The complete characterisation of SNP variation and LD among common variants in the sampled populations led to an excellent understanding of genetic variation in the human population and allowed for the determination of the haplotypes that might be involved in specific diseases [19–22].

Phase II of the project characterised an additional 2.1 million SNPs in the same individuals, which thus achieved the goal of a SNP density of approximately one SNP per kilobase. Phases I and II of the International HapMap Project genotyped approximately 25–35% of the total 9–10 million common SNPs in the assembled human genome, with a minor allele frequency (MAF) of >0.05 [21]. The project was completed with the release of the Phase III dataset in spring 2009, for which 1.6 million SNPs were genotyped in 1184 reference individuals to generate an integrated dataset of common and rare alleles [23].

Combined with the HGP and the SNP consortium, the International HapMap Project contributed to the identification of approximately 10 million common DNA variants and the understanding of their LD patterns, which enabled GWAS that identified thousands of newly discovered genes related to diseases [23]. The International HapMap Project was the first considerable achievement after the HGP. It greatly contributed to the development of personalised medicine studies and provided tools that served scientists for a decade.

The achievements or feelings that we once found extraordinary and unforgettable sink into oblivion when our attention is drawn to new attractive things. In 2016, the National Centre for Biotechnology Information (NCBI) retired the HapMap dataset and website, which were overshadowed by new monumental research, including the 1000 Genome Project [24]. A similar shift occurred with some of Odysseus's soldiers after the second stop on the island of the Lotus-eaters. Seduced by the sweet elixir of the lotus flower, they forgot the victory they achieved in Troy and their craving to see the homeland. Their only interest became the lotus flower, so they decided to stay

on the island and never return to Ithaca. Thus, Odysseus chose to depart with the rest of his crew to continue towards new victories and left them behind on the island of the Lotus-eaters (Figure 2).

Figure 2. *Land of the Lotus Eaters*: Robert Duncanson, 1863.

4. Third Stop: The Land of Cyclopes and Pharmacogenomic Studies

"We sailed hence, always in much distress, till we came to the land of the lawless and inhuman Cyclopes. Now the Cyclopes neither plant nor plough, but trust in providence, and live on such wheat, barley, and grapes as grow wild without any kind of tillage, and their wild grapes yield them wine as the sun and the rain may grow them."
"When the child of morning, rosy-fingered Dawn appeared, we admired the island and wandered all over it . . . " [1]

Traditional medicine, along with the pharmaceutical industry, has been successfully saving lives for decades. However, with the sometimes surprising number of possible side effects listed on medication packages, we are left hoping to not be one of the "unlucky users". Individuals have different ways of responding to an administered drug, which is conditioned by their genetic predisposition. Indeed, small inherited variations in a single gene may affect a patient's susceptibility to specific drugs. These considerations were not accounted for during traditional drug design. Hence, currently, some of the highest-grossing drugs are effective for as few as 1 of 25 to 1 of 4 patients [25]. The latest meta-analysis on the prevalence of medication-related adverse events among inpatients in the Western world estimated that adverse drug events affect 19% of patients during hospitalisation [26]. Furthermore, more than 4% of urgent hospitalisations are estimated to be caused by adverse drug events, with age and polypharmacy among the primary risk factors [27], which signify a large burden for patients and health systems.

Pharmacogenomics studies represent the basis for the clinical application of personalised medicine. The study of the impact of genetic variations on drug development is leading to the development of therapies that might ensure the best response and highest safety for a patient, whilst preventing or minimising the risks of side effects. The goal of pharmacogenomic research is to provide effective therapies for smaller subgroups of patients who are stratified based on their genetic profile, despite having similar disease phenotypes. Moreover, these studies aim to prevent adverse effects using preliminary genetic tests to determine whether patients might benefit from the drug and will allow

dose adaptation for each patient based on their metabolic system characteristics. Finally, they will lead to the development of specific drugs for patients who are unresponsive to available treatments [28,29].

Pharmacogenetics is a subgroup of the pharmacogenomics field that focuses on examining genetic variations in drug metabolism [30]. Although bearing enormous potential, the translation from discoveries to clinical applications remains slower than desired [31]. In contrast, its successful applications increase hope and optimism and drive change for the future of practice in all aspects of personalised medicine.

Firstly, the case of abacavir demonstrated that through genetic examination, the response to a specific drug therapy might be predicted [32]. A severe adverse effect (abacavir hypersensitivity syndrome) manifested in 4–8% of patients treated with this anti-human immunodeficiency virus (HIV) drug. This reaction was later related to the major histocompatibility complex (MHC) class I allele *HLA-B**5701, with a negative predictive value of 100%. Currently, the HIV treatment guidelines strongly recommend *HLA-B**5701 testing prior to using abacavir [33,34]. Another example of drug effectiveness prediction using a personalised approach is gefitinib, a cancer drug prescribed for metastatic non-small-cell lung cancer, approved by the U.S. Food and Drug Administration (FDA) in 2003. Treatment with this therapy was only successful in 40% of patients. A year later, research groups demonstrated that these 40% of patients had mutated *EGFR* genes. This discovery allowed for the development of personalised lung cancer treatments using a new generation of targeted drugs [35–37].

Secondly, the testing of leukaemia patients for the thiopurine *S*-methyltransferase (*TPMT*) gene is a practical use of genetic testing to determine the ideal drug dose. *TPMT* is a gene responsible for producing an enzyme that metabolises thiopurines. A decreased ability to metabolise the drug was observed in 10% of the patients, due to a genetic change that reduces the number of high-activity copies. For these patients, a reduced thiopurine dose prevents undesirable effects and allows for successful therapy [38,39]. In fact, several known genes in addition to *TPMT* exist that are responsible for the variance in drug metabolism, specifically the cytochrome P450 gene family (*CYP*), vitamin K epoxide reductase complex subunit 1 (*VKORC1*), dihydropyrimidine dehydrogenase (*DPYD*) and uridine diphosphate glucuronosyltransferase (*UGT*) [40,41]. Whilst pharmacogenetic tests for these genes are indispensable for the correct pharmacokinetic calculations and drug dosages for patients, they are not fully implemented in daily clinical practice. Another example of drug dose adaptation aided by pharmacogenomics is the case of warfarin, a vitamin K antagonist used as an anticoagulant [42,43]. The FDA recommended genetic testing for *VKORC1* and *CYP2C9* polymorphisms, which account for 25–35% of the variance in warfarin dosing [44]. This testing is challenging because it is costly, inadequately available and time consuming [43]. However, additional studies are needed to incorporate genetic testing strategies into drug dose regulation in the future.

The first FDA-approved pharmacogenetic test was released onto the market in 2005. The AmpliChip™ CYP450 Test (Roche Molecular Systems, Inc.) is indicated to assess the metabolism rate of drugs metabolised by the CYP 2D6 and 2C19 isoenzyme, based on the genotyping of 27 alleles in *CYP2D6* and 3 alleles in *CYP2C19* genes associated with different metabolic phenotypes [45]. Currently, many more genetic tests are available to optimise drug therapy. Guidelines and tables of the available tests are presented on the Clinical Pharmacogenetics Implementation Consortium website [46].

We are still in the land of the Cyclopes, where many events have transpired since Odysseus left his ship and went exploring the unknown. Trapped in a cave with Polyphemus, son of Poseidon, who ate some of his men, Odysseus was like the researchers, trying to use all their creativity to escape the undesirable events. He blinded the eye of the giant Cyclope, hid himself and the rest of the men under the sheep (Figure 3), and thus provided himself a safe escape from the island. His group of soldiers survived, because blind Cyclope could not recognise them, as he would, if they would not be attached to the animals. And that is the simple principle of the pharmacogenomics treatment.

Figure 3. Odysseus in the Cave of Polyphemus: Jacob Jordaens, 1635.

5. Fourth Stop: Aeolian Island and Genetic Association Studies

"It is an island that floats (as it were) upon the sea, iron bound with a wall that girds it. Now, Aeolus has six daughters and six lusty sons, so he made the sons marry the daughters, and they all live with their dear father and mother, feasting and enjoying every conceivable kind of luxury."

"He [Aeolus] put the sack in the ship and bound the mouth so tightly with a silver thread that not even a breath of a side-wind could blow from any quarter. The West wind, which was fair for us, did he alone let blow as it chose; but it all came to nothing, for we were lost through our own folly." [1]

Genetic association studies are used to test the association between a specific disease (specific phenotype) and genetic variations to identify candidate genes or regions within the genome that might contribute to that disease [47]. The most widely tested genetic variations are SNPs, which are either assessed in candidate gene research [48] or through the more recently popularised GWAS [49]. Whilst candidate gene research confirms or refutes the correlation between a priori selected specific genetic variants and a disease, GWAS provides tests without a hypothesis for the specific regions, genes, or variants, albeit with a pre-selected design for genotyping platforms and analytical methodologies [50]. GWAS enabled the efficient and comprehensive assay of genetic variants that are common in a population and the identification of those that appear more commonly in patients with a given condition [51].

Although candidate gene studies brought remarkable success to the identification of high relative risk genes, they have not been as successful in the identification of genes involved in the complex forms of diseases [48]. Thus, the first GWAS, published in 2005 [52], generated enthusiasm among researchers.

Before leaving the Aeolian island and sailing towards his homeland, Odysseus received a special gift from the god of winds (Figure 4)—a tightly bounded sack. None of his soldiers was aware of its contents, and they became curious, believing that Odysseus was hiding treasures from them.

Figure 4. Aeolus Giving the Winds to Odysseus: Isaac Moillon, unknown date. Receiving a remarkable gift is of little value if it is not used appropriately.

"Nine days and nine nights did we sail, and on the tenth day, our native land showed on the horizon." [1]

Soon before reaching Ithaca, they opened the sack. The winds that were trapped inside started to blow again and took their ship back to the Aeolian island.

If GWAS at the beginning seemed like a gift, which will give the solutions to all problems, the high expectations were soon extinguished. These studies demanded a large number of samples for adequate statistical power [53]. Indeed, studies with hundreds of thousands of participants might only explain a small proportion of overall heritability. For example, significantly associated common variants in type 2 diabetes appear to merely explain approximately 6% of the increased risk of disease among relatives [54]. The number of genes related to complex diseases each accounted for only up to 10–30% of the genetic component of any given trait, thus indicating a new challenge, the so-called "hidden heritability" [55].

Nevertheless, GWAS connected thousands of novel genetic variants to many complex diseases, including diseases with no previously known genetic linkage. Their limitations, such as statistical strength and precision, were overcome using the statistical tool of meta-analysis, which combines the results of different genetic association studies to explore different sources of heterogeneity and to identify subgroups associated with the factor of interest [56]. Furthermore, the latest GWAS studies that are performed in very large populations, the improvements of computational tools, and the development of large datasets such as the UK Biobank provide new insights for the utility of GWAS studies. Recently, polygenic scores derived from GWAS studies have been shown to be as useful as monogenic mutations for the prediction of chronic diseases, thus opening a new direction for the application of GWAS results in clinical practice [57]. The next important step is to match these results to a larger biological construct to unravel the most complicated pathways related to a disease and to thereby contribute knowledge that might be translated into clinical and diagnostic tools [55].

Similarly, Odysseus and his crew had to take the next step and find their own way further on after the sorrowful event, for Aeolus, the god of winds, wished no more to help them.

6. Fifth Stop: The Laestrygones and Prediction Tools of Personalised Medicine

"Six days, night and day did we toil, and on the seventh day, we reached the rocky stronghold of Lamus—Telepylus, the city of the Laestrygonians, where the shepherd who is driving his sheep and goats [to be milked] salutes he who is driving out his flock [to feed] and answers the salute. In that country, a man who could do without sleep might earn double wages, one as a herdsman of cattle, and another as a shepherd, for they work much the same by night as they do by day." [1]

Another important aspect of personalised medicine, in addition to predicting the response to a treatment or adapting the drug dosage (discussed in the third stop of the journey), is predicting the likelihood of developing a specific disease. This ability may be indispensable in specific families with disease antecedents, wherein rapid discovery by genetic testing might prevent the onset of the disease. The famous case of actress Angelina Jolie raised awareness of these methods among the general population, when she announced she was undergoing a risk-reducing mastectomy after testing positive for the *BRCA1* gene mutation [58], a major risk factor for breast cancer. Six months after the announcement, the demand for genetic tests increased by more than 100% [59].

Currently, more than 2000 genetic tests are available for several different indications [60] to:

- identify genetic diseases in unborn babies
- determining whether people carry a disease gene and might pass it on to their children
- screen embryos for disease
- test for genetic diseases in adults before they are symptomatic
- diagnose a person who has disease symptoms
- determine the type or dosage of a medication that is optimal for a specific individual

The Genetic Testing Registry (GTR®), part of the National Center for Biotechnology Information (NCBI) (available at: https://www.ncbi.nlm.nih.gov/gtr/), provides detailed information on all available genetic tests for a condition or drug response.

Genetic testing is voluntary. A positive result may guide a person towards available prevention, monitoring, and treatment options. In some cases, although no solution might exist for the given condition, genetic tests may aid decision-making concerning family planning [60]. Science is diligently attempting to improve the potential for gene therapies that might prevent the diseases caused by gene mutations. In 2017, the first gene therapy was approved by FDA [61], raising a great optimism for future genetic disorder treatments.

When Odysseus reached the island of Lamus near the Telepylus, the city of the Laestrygonians, he was cautious. As one cannot surely know whether he has a genetic predisposition for a disease, so Odysseus could not predict whether some dangerous creatures were living in this unknown land. He decided to keep his ship outside the bay, before exploring the island, while other captains sailed their ships into the harbor and attached them close to one another. But the stay on this island was rather unfortunate. No people were living on it, except for the cannibal giants, who destroyed all the ships but one, hidden outside the bay. Thus, Odysseus lost the big part of his fellow companions, but kept himself and the men on his ship alive, so they continued alone on this long journey, sailing on the open sea.

7. Sixth Stop: Aeaean Island, where Circe lives, and Ethical Issues of Genomic Research

"Thence we sailed sadly on, glad to have escaped death, though we had lost our comrades, and came to the Aeaean island, where Circe lives—a great and cunning goddess who is sister to the magician Aeetes—for they are both children of the sun by Perse, who is daughter to Oceanus." [1]

Along with new, extensive information about individuals, genetics research raised ethical concerns about such experiments. Assuredly, sceptics first considered conspiracies of pharmaceutical and insurance companies, who might use this information for their own benefit. This scepticism is not surprising because genetics may not only predict susceptibility to a certain disease but also may interrogate the history of individual genes and loci under natural selection [62]. Hence, strong ethical and legal foundations were established to protect against gene-based discrimination to minimise the harms and maximise the benefits and confidentiality of genetic studies [63].

The United Nations Educational, Scientific and Cultural Organization (UNESCO) first established a Bioethics Programme in 1993, which was followed by three adaptations: the Universal Declaration on the Human Genome and Human Rights in 1997 [64], the International Declaration on Human Genetic Data in 2003, and the Universal Declaration on Bioethics and Human Rights in 2005 [65].

Protecting genetic information became one of the most important considerations in the development of personalised medicine. Nevertheless, only time will establish public readiness for the day when a genetic test will be as common as a simple blood analysis. Not all individuals will want to follow the new inroads led by medical progress. Indeed, freedom of choice remains one of the basic human rights to be structured into future health systems [29,66]. However, when choosing between life and death, principles are often forgotten in lieu of doing whatever is required to remain healthy and alive.

Ethical issues also troubled Odysseus when he met the goddess Circe, a beautiful witch, who was keen on using magic herbs on travellers that passed nearby, transforming them into wild beasts. He was advised by the god Mercury on how to resist her magic potions with a protective talisman and how to provide himself with her loyalty by going to bed with her. Odysseus was deeply in love with his faithful wife Penelope, but did not bother about keeping company with Circe for an entire year (Figure 5). Afterwards, the soldiers became anxious to return and the journey had to continue. Circe kept her promise to help and furthered them towards the House of Hades.

Figure 5. Odysseus and Circe: Angelica Kauffman, 1786. At what point might concerns and angst be disregarded to take advantage of what is offered, notwithstanding fear of the consequences?

8. Seventh Stop: The Underworld, Personal Genome Project and Cancer Genome Atlas

"Circe, that great and cunning goddess, sent us a fair wind that blew dead aft and stayed steadily with us, keeping our sails all the time well filled."
"All day long her sails were full as she held her course over the sea, but when the sun went down and darkness was over all the Earth, we got into the deep waters of the river Oceanus, where lie the land and city of the Cimmerians, who live enshrouded in mist and darkness, which the rays of the sun never pierce, neither at his rising nor as he goes down again out of the heavens, but the poor wretches live in one long melancholy night." [1]

In 2005, two important projects were launched that impacted personalised medicine. Firstly, the vision of the Personal Genome Project, as a natural successor of the HGP, was to publicly share genome, health and trait data for rapid scientific progress [67]. Initiated by the American geneticist, George McDonald Church, the project was a response to the demand for a highly integrated and comprehensive human genome and phenome datasets to aid research in human functional genomics and systems biology [68]. Currently, it is a network that connects researchers and institutions from the USA, Canada, the UK, Austria and China [67]. The project invites individuals to non-anonymously share their genomic data, traits and cells for free and open research [69,70]. To date, more than 5700 individuals have volunteered (https://my.pgp-hms.org/users) and established the foundation towards the desired goal of 100,000 participants.

Secondly, the Cancer Genome Atlas (TCGA), a collaboration between the National Cancer Institute (NCI) and NHGRI, aimed to catalogue and discover key genomic changes in large cohorts of human tumours [71] and demonstrated the importance of the information acquired through multidimensional genomic analysis for understanding the molecular basis of cancer [72]. The pilot project explored three types of cancers, which were glioblastoma multiforme, and lung and ovarian cancer [73]. Until 2017, it successfully generated comprehensive, multi-dimensional maps for 33 types of cancer and is now concluding [71]. Sharing the value placed on the importance of data sharing with the Personal Genome Project, diverse data from more than 11,000 patients, including clinical information about the patient, metadata about the samples, histopathology images and molecular information, such as gene expression, copy number, SNP genotyping, genome-wide DNA methylation, microRNA profiling and exon sequencing, are publicly available and have thus contributed to more than 1000 studies on cancer [71]. Identifying genes that contribute to functional changes in cells and oncogenic biomarkers allowed for clinical applications and improved the prevention, diagnosis, and treatment of cancer [74]. Finally, it became a model of a successful network and of collaboration for future projects in human health.

Vast innovative collaborations that create networks among researchers around the world are indispensable for generating representative amounts of samples for genomic research progress, which leads to personalised medicine. Integrated knowledge facilitates the path to important discoveries. The aforementioned successful projects provide outstanding examples for the future of research. Ultimately, the goals towards which we are striving are interchangeable worldwide.

In the Underworld, Odysseus was searching for the knowledge that would facilitate his way back home. Thus, he made a sacrifice under Circe's orders and offered blood to the ghosts with whom he desired to speak. The dead prophet Teiresias (Figure 6) and other ghosts shared important information, which allowed Odysseus to prepare himself for future adventures. He returned to Circe, buried the body of his dead soldier and departed towards the new challenges.

Figure 6. Teiresias and the Cancer Genome Atlas. (Adapted from: https://www.thoughtco.com/).

9. Eighth Stop: The Island of the Two Sirens and High-Throughput Technology

"I had hardly finished telling everything to the men before we reached the island of the two Sirens, for the wind had been very favourable. Then, all of a sudden, it fell dead calm; there was not a breath of wind nor a ripple upon the water, so the men furled the sails and stowed them; then taking to their oars they whitened the water with the foam they raised in rowing." [1]

At the beginning of our journey with Odysseus, we discussed the momentous discoveries that led to the rise of the genomics era, which must include one of the most valuable, that of Frederick Sanger. This British biochemist and two-time Nobel prize winner developed a method for DNA sequencing that became the gold standard for the "first generation" of DNA sequencing, and a key technology for the HGP [75].

This technique was used in daily research for more than 30 years, until the arrival of next-generation sequencing (NGS) in 2005 [76]. NGS differed from the Sanger method in its massively parallel analysis, high throughput, and reduced cost [77]. Sequencing methods are classified into three generations plus one future generation, which each have distinguishing characteristics (Figure 7).

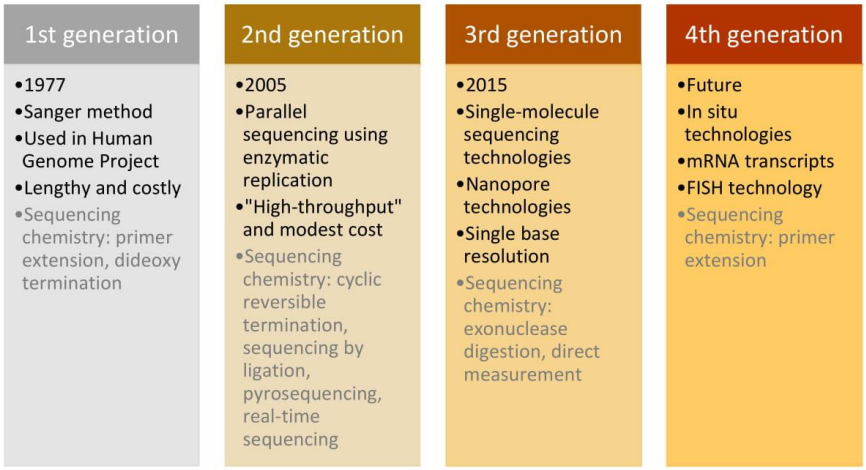


Figure 7. Schematic presentation of the four generations of DNA sequencing.

Firstly, the "second-generation" sequencing analyses clonally amplified templates that originated from a single DNA molecule, using microfluidic devices and the stepwise addition of four nucleobases. The product of this technology is a high-resolution image that provides accurate sequencing results [78,79]. The three most important technological systems were developed for "second-generation" sequencing, beginning with the 454 Sequencer (currently by Roche), which was the first commercially successful instrument released on the market. It was followed by the Genome Analyzer, developed by Solexa (later purchased by Illumina), which originally could generate 1 G per run. The technology improved and in 2010, Illumina released HiSeq 2000, with an output of 200 G per run. The third fundamental technology that marked the second generation of sequencing was the Sequencing by Oligo Ligation Detection (SOLid) system, made by Agencourt (purchased by Applied Biosystems in 2006) [80]. These technologies progressed over time to provide the most performative, accurate and efficient sequencing possible. They were the first throughput technologies with reduced costs and routine analytic methods for scientific research, which revolutionised genomic research.

Despite the success of second-generation methods, other innovative technologies emerged. The key characteristic of "third-generation" sequencing is the direct reading of a single DNA molecule template without prior processing using enzymatic replication. This feature decreased the bias of previous methods caused by polymerase chain reaction (PCR) amplification. Moreover, these methods increased the read length and decreased the processing time, which provided even higher overall accuracy and enabled rare variant detection at a lower cost [81].

Three categories of "third-generation" sequencing exist: "second-generation" technologies in which single molecules of DNA polymerase are studied, nanopore-sequencing technologies, and the direct imaging of individual DNA molecules using advanced microscopy techniques. The first nanopore DNA sequencer, MinION, commercially available since 2015, was developed by Oxford Nanopore Technologies and provides sequencing for as little as $1000/sample [82].

Finally, the "fourth generation" focuses on examining messenger RNA (mRNA) transcripts using in situ technology (fluorescence in situ hybridisation (FISH)), which allows the detection and genotyping of somatic point mutations with subcellular resolution by directly sequencing nucleic acids in cells and tissue. Furthermore, it employs the conversion of mRNA to localised complementary DNA (cDNA) and thus enables the genotyping of individual transcript molecules. Because expression in single cells may vary substantially from the mean expression detected in a heterogeneous cell population, this method has the potential to revolutionise the field of cancer biology because it enables

the mapping of the molecular and cellular tumour environment. Additionally, it is intended to detect defined DNA sequences (such as those that might have undergone a somatic mutation) rather than the complete genotypic sequence from the cell [77,83,84].

Although in situ sequencing has prodigious potential for future research, several technical improvements are needed before it might be broadly applied, specifically sample imaging, the low efficiency of molecular processes, and data handling and interpretation. Should these obstacles be overcome, the method might become a standard for sequencing tissue samples [83].

The "first-generation" sequencing method provided an opportunity for one of the greatest achievements in genomics—the decoding of the human chromosomal genome. This achievement boosted the development of high-throughput technologies that enabled the development of promising tools and led to large-scale studies, which are the engines that further power personalised medicine.

Passing by the island of Sirens, Odysseus knew that to hear their beautiful singing would lead to inevitable death, as Circe has warned him about the skeletons of unfortunate soldiers, misled by the beauty of the sound, which were lying all around the island, with their flesh still rotting. Thus, he took her advice and plugged the ears of his soldiers but attached himself to the mast, free to hear what no man alive had heard before.

With great amount of knowledge, we are free to do things that once seemed impossible. Sequencing that first took more than a decade to be accomplished is now an everyday practice. And Odysseus? His trick, simple and deft, enabled him to bypass the trap and continue on his way unharmed.

10. Ninth Stop: Cliffs of Scylla and Charybdis and Large Gains at a Small Cost

> *"Then, we entered the Straits in great fear of mind, for on the one hand was Scylla, and on the other dread Charybdis kept sucking up the salt water. As she vomited it up, it was like the water in a cauldron when it is boiling over upon a great fire, and the spray reached the top of the rocks on either side."* [1]

No discussion of genetics is complete without considering costs. The final cost of the HGP rose to $2.7 billion for the 13-year endeavour. The WGS of the first "reference" genome included approximately three billion bases, which represented a haploid copy of the human genome, and covered ≈99% of the euchromatic genome, accurate to an error rate of ≈1 event per 100,000 bases [85]. The cost of this project included a wide range of other expenses in addition to sequencing, such as technology development, physical and genetic mapping, model organism genome mapping and sequencing, bioethics research, and programme management. The NHGRI estimated the cost of the potential second "reference" human genome sequencing using the technology available in 2003 as up to $50 million. Though far lower than the initial cost, the use of such methods would still be too costly for routine studies and clinical applications [86]. Based on Sanger's technique, the diploid genome of the American geneticist Craig Venter was sequenced for approximately $100 million in 2007, and the results were deposited into the public GenBank database [87].

Whilst many ongoing parallel studies contributed to technological progress, the launching of a ground-breaking new project, called the "Advanced Sequencing Technology Awards" or the "$1000 Genome", set forth a new era in personalised medicine. The NHGRI began this project in 2004 to boost technological progress towards reducing the cost of DNA sequencing for an individual genome to $100,000 in 5 years and $1000 in 10 years. It committed more than $100 million to 50 academic and industrial teams of scientists, including major sequencing companies who, combined, were highly successful. Ultimately, the price of sequencing began to decrease at an accelerated rate after 2007 (Figure 8) [88,89].

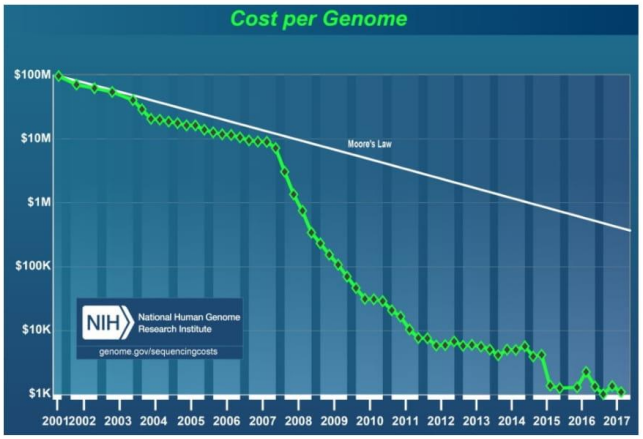


Figure 8. Costs for whole-genome sequencing (WGS). Years 2001–2007 reflect the costs of Sanger sequencing techniques, whilst the data after January 2008 reflect the costs of "second-generation" sequencing. Moore's Law (in white) is a comparison with the normal long-term trend of progress in the computer hardware industry.

Additional incentive was provided by Craig Venter's foundation, which offered a $500,000 award to stimulate DNA sequencing development with the same final goal of reducing the price of sequencing to $1000 per genome [90]. This award merged with The X Prize Foundation and, in October 2006, formed the Archon Genomics X Prize Foundation, which first announced a $10 million award for technology capable of successfully mapping 100 human genomes in 10 days [91]. The contest began in January 2013, and the prize was intended to be awarded to team(s) able to sequence 100 human genomes with 98% completeness, with an error rate ≈1 event per 1,000,000 bases and a total cost of $1000 per genome, within a time limit extended to 30 days [92]. However, the contest was cancelled in August 2013, due to an unexpectedly high rate of progress in sequencing technologies on the market that required no further encouragement [93].

Indeed, technology progressed much faster than expected. The first NGS of an entire human genome was performed using a Roche 454 Sequencer in 2008. The diploid human genome sequence of James D. Watson was completed in two months using massive parallel sequencing in picolitre-size reaction vessels and was also the "first individual genome to be sequenced for less than $1 million" [94,95]. This advance ushered in a massive reduction in price and the era of next-generation sequencing. Finally, the cost, though still high, was acceptable for routine research and thus permitted the use of a personalised medicine tool necessary for the research and development of new strategies for efficient patient diagnostics and treatment.

There is always a price to pay in science, as on the sea. The Sirens were not the only trap at sea that Circe warned Odysseus about. Two monsters, Scylla and Charybdis (Figure 9), were waiting for the sailors and were floating nearby. To avoid Charybdis, a whirlpool that could sink the whole ship, one would get too close to Scylla, a six-headed monster that would eat men alive. They would rather pay the price and continue on their chosen journey than sink after all the effort exerted during their travels.

Figure 9. Scylla and Charybdis: Salvador Dali, 1970. There is always a price to pay, but it rests with us to set the limits.

11. Tenth Stop: The Island of Helios, and the 1000 and 100,000 Genomes Projects

"When we had passed the rocks, with Scylla and terrible Charybdis, we reached the noble island of the sun-god, where were the goodly cattle and sheep belonging to the sun Hyperion." [1]

Many important projects contributed to personalised medicine, several of which were discussed during the recounting of the previous adventures of Odysseus. However, we cannot continue our journey without mentioning the last two, which might share similar titles, but should not be confused with each other.

The 1000 Genomes Project (1KGP) began as a collaboration among research groups from the USA, UK, China and Germany in 2008 and was an extension of the International HapMap Project, which aimed to establish the most extensive catalogue of human genetic variation. Although HapMap was a revolutionary project, it had several limitations, e.g., it only focused on SNPs and no other forms of genetic variations, with an MAF > 0.05. In contrast, the goal of the 1KGP was to establish a resource of at least 95% of the existing variants of human DNA and their haplotype contexts, with frequencies of at least 1% in multiple studied populations. This objective would be fulfilled by sequencing at least 1000 individuals from around the world using next-generation sequencing technologies [96–98].

The pilot project was designed to evaluate the strategies of genome-wide sequencing and determine the adequate coverage needed to detect the desired frequencies. It was divided into three steps: (i) WGS of 179 samples with low coverage (2–6×); (ii) deep sequencing of two mother–father–child trios (20–60× coverage); and (iii) exome targeting sequencing of 1000 gene regions in 697 samples (50× coverage). The pilot phase was completed in 2010 and produced robust protocols for whole-genome and targeted sequencing with validated algorithms. Low-coverage sequencing was efficient for detecting and genotyping rare variants and was selected for use in future phases [98,99].

The main project started with a phase two analysis, which generated low-coverage WGS data and exome sequencing for 1092 individuals from 14 populations. It provided a validated haplotype map of 38 million SNPs, 1.4 million insertions/deletions and 14,000 larger deletions, which aided the interpretation of the genetic variants that affect diseases, and provided a publicly available human genome map [100].

The final phase of the 1 KGP was completed in 2015 with the reconstruction of 2504 individual genomes from 26 populations with combinations of low-coverage WGS, deep exome sequencing and dense microarray genotyping [101,102]. Thus, the map included >99% of SNP variants with a frequency of >1% in multiple populations. The analysis of the results demonstrated that most common genetic variants are shared across populations, whilst rare variants are often restricted to closely related populations [103].

The 1KGP was the largest study to sequence the genetic information of such a large number of individuals and its ample variety of genetic information provided the foundation for future disease and population genetic studies. The International Gene Sample Resource is now charged with ensuring future access to and the expansion of the 1 KGP data [98].

In late 2012, the UK Prime Minister launched the 100,000 Genomes Project, which aims to sequence 100,000 whole genomes from approximately 70,000 people recorded in The National Health Service (NHS), England. Led by a public company, Genomics England, it established four main aims: (i) to create an ethical and transparent programme; (ii) to benefit patients and construct a genomic medicine service for the NHS; (iii) to enable new scientific discovery and medical insights; and (iv) to encourage the development of a UK genomics industry [104].

As of early December 2017, more than 41,500 genomes had been sequenced. The project is specifically focusing on rare diseases and cancer because these conditions are strongly linked to genetic mutations. Using longitudinal patient recording, the NHS might correlate a patient's genome data to phenotypes to thereby aid the understanding of the complex relationship between genes and diseases to not only facilitate the development of a personalised therapy for a patient but also to leave an important legacy for the future [104,105].

Furthermore, since the 1990s, the need for nationwide databases of genetic, clinical and environmental information has been recognised aiming to support the implementation of personalised medicine. The DeCODE project has resulted in the development of a complete database of all the Icelandic population starting at the end of 1990s. More recent efforts include the UK Biobank and the All of Us Research Program in USA.

When Odysseus arrived on the island with his crew, a huge storm commenced and lasted for a month. Although his crew began to run out of the food, he prohibited them from killing any of cattle grazing nearby because he was warned several times that this would signify the end for all his men. But they did not listen, and killed the biggest animal of the god of the sun, Helios, and all, except Odysseus, feasted. When the storm ended, they continued the journey, but Zeus destroyed their ship and everyone died but Odysseus, who continued floating on the open sea, all alone.

Great past and ongoing projects are the proof that in the battle to overcome obstacles in health, scientists, unlike Odysseus, are not alone, but are sailing together with the help of national government towards their Ithaca, better health care for the people.

12. Eleventh Stop: Ogygia, Calypso's Island, and '-omics' Development

> *"Hence, I was carried along for nine days till on the tenth night, the gods stranded me on the Ogygian island, where dwells the great and powerful goddess Calypso. She took me in and was kind to me, but I need say no more about this."* [1]

We have arrived at the point on our journey at which the foundation laid by all the important projects and discoveries thus far will permit researchers worldwide to perform independent analyses to unravel the complex mechanisms of various diseases. The era of "-omics" studies arose at the beginning of the 21th century on the wings of the technological revolution in rapid throughput technologies.

"-Omics" technologies are useful tools for detecting genes, mRNA, proteins, metabolites and epigenetic modifications. Genomics, with its well-known GWAS for detecting complex traits and WGS, allows for an understanding of human population and disease genetics, and was previously discussed in the fourth stop of our journey. However, "-omics" technologies provide a holistic view

of the molecules that constitute a cell, tissue or organism along with four other branches of this high-dimensional biology: transcriptomics, proteomics, metabolomics (Figure 10) [106].

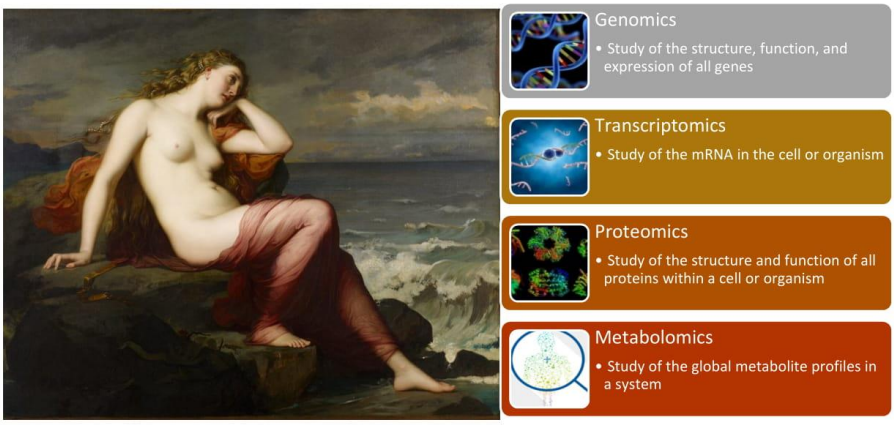


Figure 10. Calypso staring at the four most common "-omics" studies.

Transcriptomics emerged with the development of RNA sequencing technologies, which examine an organism's transcriptome, the sum of all RNA transcripts. Because the RNA profile directly reflects the actual gene activity, it measures the expression of an organism's genes and provides information regarding how genes are regulated [107]. In combination with genomics, RNA sequencing data provide a unique perspective on genetic functional variants and a deeper understanding of the cellular mechanisms of transcriptome variation [108].

Proteomics is the large-scale study of proteomes, which encompasses the aggregate of all expressed proteins within a cell, system or organism, and their structure and function. This field represents the next step after genomics and transcriptomics research for examining biological systems. For a long time, protein research was restricted due to technological limitations, which were overcome by the arrival of new mass spectrometry technologies. Proteomics aims to elucidate the functional role of proteins through their pathways and networks. It reflects gene and environmental dynamics and thus is closely related to functional genomics [107,109]. Proteomics remains a complex field because of the vast dynamics of proteins; however, it is a promising field for the discovery of new biomarkers because most commonly it is proteins that are affected during disease or a disease response. With future progress in technology, we will be able to achieve greater advances in the field and improve research on new therapeutic targets [110].

Finally, metabolomics is the last of the basic "-omics" technologies. It is the study of the chemical processes concerning metabolites, which thus reveals the energy status and metabolism of an organism. It reflects the changes in the transcriptome and proteome. Though it is the smallest of the four main "-omics" branches, it is the most physically and chemically complex [106,107].

Another emerging field is epigenetics modifications, which include heritable modifications of the gene expression regulation that are not due to nucleotide changes of DNA sequence. These modifications include the DNA methylation, histone modification and non-coding RNA (ncRNA)-associated gene silencing [111]. High throughput screening assays are available today for the study of epigenome, while large-scale projects aiming to the development of public datasets of epigenetic modifications exist. The Encyclopedia of DNA Elements (ENCODE) aims to identify all regions of transcription, transcription factor association, chromatin structure and histone modification in the human genome sequence [112]. The Natioanal Institutes of Health (NIH) Roadmap Epigenomics

Mapping Consortium is developing a public resource of human epigenomic data aiming to support basic and medical research related to the epigenetics field [113].

Systems biology is using "-omics" studies for detailed examinations of organisms and their molecular phenotypes. It facilitates the generation of complex molecular pathways that might only be revealed through a comprehensive approach. The accumulating knowledge from "-omics" research provides vast possibilities for the targeted treatment of diverse medical conditions, the discovery of novel possible biomarkers for diseases diagnostics and prognostication, and the transformation of conventional symptom-orientated medicine to preventive and personalised health care.

Such studies cannot bring immediate change but rather require years to transform discoveries into practical applications. Thus, the "-omics" era is passing through a long, calm period with much progress but no substantial perturbations thus far.

The calm period was there also for Odysseus, who settled in Ogygia, on the island of a nymph called Calypso. She fell in love with the hero and wanted to have him as her husband. For seven years, they lived together on the island, until the goddess Athena interrupts the story. Finally, Calypso had to let her lover go, and helped build a raft that would take him towards the final stop of his long journey, before reaching home.

13. Twelfth Stop: Scheria, the Island of the Phaeacians and Alliances of Personalised Medicine

> *"Thereon he floated about for two nights and two days in the water, with a heavy swell on the sea and death staring him in the face; but when the third day broke, the wind fell and there was a dead calm without so much as a breath of air stirring. As he rose on the swell he looked eagerly ahead, and could see land quite near. Then, as children rejoice when their dear father begins to get better after having for a long time borne sore affliction sent him by some angry spirit, but the gods deliver him from evil, so was Odysseus thankful when he again saw land and trees, and swam on with all his strength that he might once more set foot upon dry ground."* [1]

This account has consistently shown that only with mutual assistance, collaboration and scientific exchange can science achieve its goals of improving health systems through novel strategies of P4 medicine—predictive, preventive, personalised, and participatory. Currently, we are apparently wedged in a transitional period of personalised medicine, a time during which everyone is discussing it and researching its applications, but without a sense of substantial change during a simple visit to the doctor. This dissociation is why collaborations between physicians, public and private researchers and politicians are critical to allow the circulation of knowledge from research institutes supported by industry to hospitals and governments. This sequence will lead to more rapid advances and progressive changes for current health systems that will directly benefit patients.

Healthcare experts from across Europe are gathering at diverse international conferences, consortia and societies to accelerate the development and delivery of personalised medicine and diagnostics. Of these, firstly, the Santorini conference series, Systems Medicine and Personalised Health and Therapy (http://santoriniconference.org/), which was initiated in 2002 and aimed to assemble scientists from genetic and biochemical backgrounds, was dedicated to personalised medicine [114]. It is the oldest international conference in the field of pharmacogenomics and personalised medicine, and is held every two years [114–121]. One of the important outgrowths from the Santorini conference series was the formation of the European Society of Pharmacogenomics and Personalised Therapy (ESPT), founded in 2011 (https://esptnet.eu/). It soon became one of the leading organisations in the field of pharmacogenomics and personalised medicine, and aims to integrate multidisciplinary approaches with research and transform them into clinical benefits for professional training and education of the general public in all areas of human pharmacogenomics, clinical pharmacology, laboratory medicine and personalised medicine [122], extending the actions of the Santorini conference series. The organisation connects research groups with corporate members and national societies to enhance the scientific basis for the quality of patient diagnosis and therapy

through efficient communication with ESPT members, other healthcare providers, regulators and the public to disseminate information and educate them regarding outstanding scientific and educational achievements [122].

Secondly, the European Alliance for Personalised Medicine was launched in 2012 to change healthcare using personalised approaches towards each patient (https://www.euapm.eu/). By developing case studies, organising workshops, education, training and communication, they attempt to increase awareness and an understanding of personalised medicine and have an important impact on the progress of this field, through increased research and development [123].

Even this time, the journey did not go well, for the god, Poseidon, was still angry with Odysseus for blinding his son, so he sent the winds and storm over the sea. But Odysseus always had allies on his side. This time, the goddess Ino gave him her magic veil to help him reach Scheria, where Nausica, daughter of the King Alcinous, found him. Odysseus was invited to dinner, where he told his long story about the journey from Troy towards home. The Phaeacians, moved by his tale, offered him treasures and provided him safe passage back to Ithaca (Figure 11).

Figure 11. Departure of Ulysses from the Land of the Phaeacians: Claude Lorrain, 1646.

14. Thirteenth Stop: Ithaca and the Future of Personalised Medicine

> *"The ship bounded forward on her way as a four in hand chariot flies over the course when the horses feel the whip. Her prow curvetted as it were the neck of a stallion, and a great wave of dark blue water seethed in her wake. She held steadily on her course, and even a falcon, swiftest of all birds, could not have kept pace with her. Thus, then, she cut her way through the water, carrying one who was as cunning as the gods, but who was now sleeping peacefully, forgetful of all that he had suffered both on the field of battle and by the waves of the weary sea."* [1]

Odysseus managed to overcome all the obstacles to reach the Ithaca (Figure 12) he was longing for, yet the story did not conclude. Angry suitors were waiting, for faithful queen Penelope must finally choose to marry one of them. And Odysseus must fight them, to save his family, to make peace and order on his island again. He has not tired of the journey however, because he knows that a new, brighter future awaits.

A new, brighter future also awaits personalised medicine.

New extraordinary findings are reminding us every day that we are on the right path. With "-omics" information, we are illuminating new hypothesis on molecular mechanisms of disease, and translating them into advice for disease prevention and improved diagnostics and therapeutics. Moreover, researchers are now realising that genotyping information could give us many new answers on disease mechanisms, just by using the causal connections between identified variants and disease pathways [124]. With holistic "-omics" perspective, looking at genes, proteins, gut microbes and metabolic markers, practitioners will be able to see how these elements of health change as people progress through the course of their lives.

Pharmacogenomics studies using GWAS have discovered numerous rare and common variants in different populations that control individual drug responses. This led also to more precision trials, because drug candidates can be tested in more targeted subpopulations [125].

Finally, the new era in medicine started in 2017 with the first FDA-approved gene therapies for treating acute lymphoblastic leukemia (Kymriah), large B-cell lymphoma (Yescarta) and an inherited form of childhood blindness (Luxturna) [126]. Moreover, for the first time, scientists demonstrated the possibility of treating Duchenne muscular dystrophy in large animals, by delivering gene editing components that can restore affected protein [127]. All these new therapies are giving hope and posing questions. Will there be a day when we will no longer treat a disease, but our genome?

Until then, we must continue to confront the challenges posed by genomics; moreover, we must continue to compile genomic variants, each of which accounts for a small percentage of disease susceptibility. Each small discovery represents a step closer to a day when all the efforts of research and industry will be combined into diagnostic and therapeutic tools that will revolutionise medicine in favour of patients and healthcare providers and regulators.

Figure 12. The journey to Ithaca. (Adapted from: http://santoriniconference.org/). Ithaca–Konstantinos Kavafis (1863–1933).

When you set out on your journey to Ithaca,
pray that the road is . . .
full of adventure, full of knowledge . . .
That the summer mornings are many, when,
with such pleasure, with such joy you will enter ports seen for the first time . . .
Always keep Ithaca in your mind.
To arrive there is your ultimate goal . . .

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Homer. *The Odyssey*; Xist Publishing: Irvine, CA, USA, 2015.
2. Watson, J.D.; Crick, F.H. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **1953**, *4356*, 737. [CrossRef]
3. Crick, F.H.; Barnett, L.; Brenner, S.; Watts-Tobin, R.J. General nature of the genetic code for proteins. *Nature* **1961**, *192*, 1227–1232. [CrossRef] [PubMed]
4. Paigen, K. On the regulation of DNA transcription. *J. Theor. Biol.* **1962**, *3*, 268–282. [CrossRef]
5. McCarthy, B.; Holland, J. Denatured DNA as a direct template for in vitro protein synthesis. *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 880–886. [CrossRef] [PubMed]
6. Meselson, M.; Stahl, F.W. The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 671–682. [CrossRef] [PubMed]
7. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [CrossRef] [PubMed]
8. Sanger, F. Determination of nucleotide sequences in DNA. *Biosci. Rep.* **1981**, *1*, 3–18. [CrossRef] [PubMed]
9. Green, E.D.; Watson, J.D.; Collins, F.S. Human Genome Project: Twenty-five years of big biology. *Nature* **2015**, *526*, 29–31. [CrossRef] [PubMed]
10. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; Fitzhugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [PubMed]
11. Brookes, A.J. The essence of SNPs. *Gene* **1999**, *234*, 177–186. [CrossRef]
12. Lander, E.S. Initial impact of the sequencing of the human genome. *Nature* **2011**, *470*, 187–197. [CrossRef] [PubMed]
13. Hood, L.; Flores, M. A personal view on systems medicine and the emergence of proactive P4 medicine: Predictive, preventive, personalized and participatory. *New Biotechnol.* **2012**, *29*, 613–624. [CrossRef] [PubMed]
14. Jain, K.K. Personalized medicine. *Curr. Opin. Mol. Ther.* **2002**, *4*, 548–558. [PubMed]
15. Yamamoto, F.; Clausen, H.; White, T.; Marken, J.; Hakomori, S. Molecular genetic basis of the histo-blood group ABO system. *Nature* **1990**, *345*, 229–233. [CrossRef] [PubMed]
16. Jameson, J.L.; Longo, D.L. Precision medicine-personalized, problematic, and promising. *N. Engl. J. Med.* **2015**, *372*, 2229–2234. [CrossRef] [PubMed]
17. Ioannidis, J.P. Personalized genetic prediction: Too limited, too expensive, or too soon? *Ann. Intern. Med.* **2009**, *150*, 139–141. [CrossRef] [PubMed]
18. International HapMap Consortium. The International HapMap Project. *Nature* **2003**, *426*, 789–796. [CrossRef] [PubMed]
19. International HapMap Consortium. A haplotype map of the human genome. *Nature* **2005**, *437*, 1299–1320. [CrossRef] [PubMed]
20. Thorisson, G.A.; Smith, A.V.; Krishnan, L.; Stein, L.D. The International HapMap Project Web site. *Genome Res.* **2005**, *15*, 1592–1593. [CrossRef] [PubMed]
21. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **2007**, *449*, 851–861. [CrossRef] [PubMed]
22. Deloukas, P.; Bentley, D. The HapMap project and its application to genetic studies of drug response. *Pharmacogenomics J.* **2004**, *4*, 88–90. [CrossRef] [PubMed]
23. Altshuler, D.M.; Gibbs, R.A.; Peltonen, L.; Altshuler, D.M.; Gibbs, R.A.; Peltonen, L.; Dermitzakis, E.; Schaffner, S.F.; Yu, F.; Peltonen, L.; et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **2010**, *467*, 52–58. [PubMed]
24. Information TNCfB. NCBI Retiring HapMap Resource. 2017. Available online: https://www.ncbi.nlm.nih.gov/variation/news/NCBI_retiring_HapMap/ (accessed on 31 January 2018).
25. Schork, N.J. Personalized medicine: Time for one-person trials. *Nature* **2015**, *520*, 609–611. [CrossRef] [PubMed]
26. Laatikainen, O.; Miettunen, J.; Sneck, S.; Lehtiniemi, H.; Tenhunen, O.; Turpeinen, M. The prevalence of medication-related adverse events in inpatients-a systematic review and meta-analysis. *Eur. J. Clin. Pharmacol.* **2017**, *73*, 1539–1549. [CrossRef] [PubMed]

27. Pedros, C.; Quintana, B.; Rebolledo, M.; Porta, N.; Vallano, A.; Arnau, J.M. Prevalence, risk factors and main features of adverse drug reactions leading to hospital admission. *Eur. J. Clin. Pharmacol.* **2014**, *70*, 361–367. [CrossRef] [PubMed]
28. Vogenberg, F.R.; Isaacson Barash, C.; Pursel, M. Personalized medicine: Part 1: Evolution and development into theranostics. *Pharm. Ther.* **2010**, *35*, 560–576.
29. Vogenberg, F.R.; Barash, C.I.; Pursel, M. Personalized medicine: Part 2: Ethical, legal, and regulatory issues. *Pharm. Ther.* **2010**, *35*, 624–642.
30. Scott, S.A. Personalizing medicine with clinical pharmacogenetics. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2011**, *13*, 987–995. [CrossRef] [PubMed]
31. Huang, R.S.; Gamazon, E.R. Translating pharmacogenomics discoveries into the clinic: An implementation framework. *Genome Med.* **2013**, *5*, 94. [CrossRef] [PubMed]
32. Pucheril, D. The History and Future of Personalized Medicine. MANAGED CARE, 2011. Available online: https://www.managedcaremag.com/archives/2011/8/history-and-future-personalized-medicine (accessed on 31 January 2018).
33. Phillips, E.; Mallal, S. Successful translation of pharmacogenetics into the clinic. *Mol. Diagn. Ther.* **2009**, *13*, 1–9. [CrossRef] [PubMed]
34. Mallal, S.; Phillips, E.; Carosi, G.; Molina, J.M.; Workman, C.; Tomazic, J.; Jägel-Guedes, E.; Rugina, S.; Kozyrev, O.; Cid, J.F.; et al. *HLA-B*5701* screening for hypersensitivity to abacavir. *N. Engl. J. Med.* **2008**, *358*, 568–579. [CrossRef] [PubMed]
35. Eisenstein, M. Personalized medicine: Special treatment. *Nature* **2014**, *513*, S8–S9. [CrossRef] [PubMed]
36. Mok, T.; Wu, Y.L.; Zhang, L. A small step towards personalized medicine for non-small cell lung cancer. *Discov. Med.* **2009**, *8*, 227–231. [PubMed]
37. Lynch, T.J.; Bell, D.W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R.A.; Brannigan, B.W.; Harris, P.L.; Haserlat, S.M.; Supko, J.G.; Haluska, F.G.; et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **2004**, *350*, 2129–2139. [CrossRef] [PubMed]
38. Drew, L. Pharmacogenetics: The right drug for you. *Nature* **2016**, *537*, S60–S62. [CrossRef] [PubMed]
39. Lee, S.H.R.; Yang, J.J. Pharmacogenomics in acute lymphoblastic leukemia. *Best Pract. Res. Clin. Haematol.* **2017**, *30*, 229–236. [CrossRef] [PubMed]
40. Bosch, T.M.; Meijerman, I.; Beijnen, J.H.; Schellens, J.H. Genetic polymorphisms of drug-metabolising enzymes and drug transporters in the chemotherapeutic treatment of cancer. *Clin. Pharmacokinet.* **2006**, *45*, 253–285. [CrossRef] [PubMed]
41. Drogemoller, B.I.; Wright, G.E.; Warnich, L. Considerations for rare variants in drug metabolism genes and the clinical implications. *Expert Opin. Drug Metab. Toxicol.* **2014**, *10*, 873–884. [CrossRef] [PubMed]
42. Weeke, P.; Roden, D.M. Applied pharmacogenomics in cardiovascular medicine. *Annu. Rev. Med.* **2014**, *65*, 81–94. [CrossRef] [PubMed]
43. Alberts, M.J. Genetics of warfarin dosing—One polymorphism at a time. *Lancet (London, England)* **2013**, *382*, 749–751. [CrossRef]
44. Thompson, C.A. FDA encourages genetics-aided warfarin dosing. *Am. J. Health Syst. Pharm.* **2007**, *64*, 1994–1996. [CrossRef] [PubMed]
45. Squassina, A.; Manchia, M.; Manolopoulos, V.G.; Artac, M.; Lappa-Manakou, C.; Karkabouna, S.; Mitropoulos, K.; Del Zompo, M.; Patrinos, G.P. Realities and expectations of pharmacogenomics and personalized medicine: Impact of translating genetic knowledge into clinical practice. *Pharmacogenomics* **2010**, *11*, 1149–1167. [CrossRef] [PubMed]
46. CPIC. Clinical Pharmacogenetics Implementation Consortium. 2017. Available online: https://cpicpgx.org/ (accessed on 31 January 2018).
47. Lewis, C.M.; Knight, J. Introduction to genetic association studies. *Cold Spring Harbor Protoc.* **2012**, *2012*, 297–306. [CrossRef] [PubMed]
48. Tabor, H.K.; Risch, N.J.; Myers, R.M. Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nat. Rev. Genet.* **2002**, *3*, 391–397. [CrossRef] [PubMed]
49. Hirschhorn, J.N.; Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **2005**, *6*, 95–108. [CrossRef] [PubMed]

50. Kitsios, G.; Zintzaras, E. Genome-wide association studies: Hypothesis-"free" or "engaged"? *Transl. Res. J. Lab. Clin. Med.* **2009**, *154*, 161–164. [CrossRef] [PubMed]
51. Rodriguez-Fontenla, C.; Calaza, M.; Evangelou, E.; Valdes, A.M.; Arden, N.; Blanco, F.J.; Carr, A.; Chapman, K.; Deloukas, P.; Doherty, M.; et al. Assessment of osteoarthritis candidate genes in a meta-analysis of nine genome-wide association studies. *Arthrit. Rheumatol.* **2014**, *66*, 940–949. [CrossRef] [PubMed]
52. Haines, J.L.; Hauser, M.A.; Schmidt, S.; Scott, W.K.; Olson, L.M.; Gallins, P.; Spencer, K.L.; Kwan, S.Y.; Noureddine, M.; Gilbert, J.R.; et al. Complement factor H variant increases the risk of age-related macular degeneration. *Science (New York)* **2005**, *308*, 419–421. [CrossRef] [PubMed]
53. Hong, E.P.; Park, J.W. Sample size and statistical power calculation in genetic association studies. *Genom. Inform.* **2012**, *10*, 117–122. [CrossRef] [PubMed]
54. Cirulli, E.T.; Goldstein, D.B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **2010**, *11*, 415–425. [CrossRef] [PubMed]
55. Sainani, K. Unlocking the genetics of complex diseases. *Biomed. Comput. Rev.* **2015**, 22–28.
56. Lee, Y.H. Meta-analysis of genetic association studies. *Ann. Lab. Med.* **2015**, *35*, 283–287. [CrossRef] [PubMed]
57. Khera, A.V.; Chaffin, M. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **2018**, *50*, 1219–1224. [CrossRef] [PubMed]
58. Evans, D.G.R.; Barwell, J.; Eccles, D.M.; Collins, A.; Izatt, L.; Jacobs, C.; Donaldson, A.; Brady, A.F.; Cuthbert, A.; Harrison, R.; et al. The Angelina Jolie effect: How high celebrity profile can have a major impact on provision of cancer related services. *Breast Cancer Res.* **2014**, *16*, 442. [CrossRef] [PubMed]
59. Raphael, J.; Verma, S.; Hewitt, P.; Eisen, A. The impact of Angelina Jolie (AJ)'s story on genetic referral and testing at an Academic Cancer Centre in Canada. *J. Genet. Couns.* **2016**, *25*, 1309–1316. [CrossRef] [PubMed]
60. Medicine NUSNLo. MedlinePlus Trusted Health Information for You. 2017. Available online: https://medlineplus.gov/genetictesting.html (accessed on 31 January 2018).
61. FDA Approval Brings First Gene Therapy to the United States. 2017. Available online: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm574058.htm (accessed on 31 January 2018).
62. Vitti, J.J.; Cho, M.K.; Tishkoff, S.A.; Sabeti, P.C. Human evolutionary genomics: Ethical and interpretive issues. *Trends Genet.* **2012**, *28*, 137–145. [CrossRef] [PubMed]
63. Badzek, L.; Henaghan, M.; Turner, M.; Monsen, R. Ethical, legal, and social issues in the translation of genomics into health care. *J. Nurs. Scholarsh. Off. Publ. Sigma Theta Tau Int. Honor Soc. Nurs.* **2013**, *45*, 15–24.
64. Unesco. Universal declaration on the human genome and human rights (revised draft). *Bull. Med. Ethics.* **1997**, *126*, 9–11.
65. Langlois, A. The UNESCO Bioethics Programme: A review. *New Bioeth. Multidiscip. J. Biotechnol. Body* **2014**, *20*, 3–11. [CrossRef]
66. Samuel, G.N.; Farsides, B. Public trust and 'ethics review' as a commodity: The case of Genomics England Limited and the UK's 100,000 genomes project. *Med. Health Care Philos.* **2018**, *21*, 159–168. [CrossRef] [PubMed]
67. Genomes, P. The Personal Genome Project. 2017. Available online: http://www.personalgenomes.org/ (accessed on 31 January 2018).
68. Church, G.M. The personal genome project. *Mol. Syst. Biol.* **2005**, *1*, 2005.0030. [CrossRef] [PubMed]
69. Genome, P. The Harvard Personal Genome Project. 2017. Available online: https://pgp.med.harvard.edu/ (accessed on 31 January 2018).
70. Lunshof, J.E.; Bobe, J.; Aach, J.; Angrist, M.; Vorhaus, J.V.; Vorhaus, D.B.; Hoehe, M.R.; Church, G.M. Personal genomes in progress: From the Human Genome Project to the Personal Genome Project. *Dialogues Clin. Neurosci.* **2010**, *12*, 47–60. [PubMed]
71. Institute NHGR. The Cancer Genome Atlas. 2017. Available online: https://cancergenome.nih.gov/ (accessed on 31 January 2018).
72. De Mattos-Arruda, L.; Rodon, J. Pilot studies for personalized cancer medicine: Focusing on the patient for treatment aelection. *The Oncologist* **2013**, *18*, 1180–1188. [CrossRef] [PubMed]
73. The Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **2008**, *455*, 1061.

74. Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef] [PubMed]
75. Heather, J.M.; Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **2016**, *107*, 1–8. [CrossRef] [PubMed]
76. Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.J.; Chen, Z.; et al. Genome sequencing in open microfabricated high density picoliter reactors. *Nature* **2005**, *437*, 376–380. [CrossRef] [PubMed]
77. McGinn, S.; Gut, I.G. DNA sequencing—Spanning the generations. *New Biotechnol.* **2013**, *30*, 366–372. [CrossRef] [PubMed]
78. Gut, I.G. 2nd generation DNA sequencing meets functional genomics. *Aging* **2010**, *2*, 541. [CrossRef] [PubMed]
79. Metzker, M.L. Sequencing technologies—The next generation. *Nat. Rev. Genet.* **2010**, *11*, 31–46. [CrossRef] [PubMed]
80. Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; Law, M. Comparison of Next-Generation Sequencing Systems. *J. Biomed. Biotechnol.* **2012**, *2012*, 251364.
81. Schadt, E.E.; Turner, S.; Kasarskis, A. A window into third-generation sequencing. *Hum. Mol. Genet.* **2010**, *19*, R227–R240. [CrossRef] [PubMed]
82. Technologies, O.N. Nanopore Sequening. 2017. Available online: https://nanoporetech.com/ (accessed on 31 January 2018).
83. Ke, R.; Mignardi, M.; Hauling, T.; Nilsson, M. Fourth generation of next-generation sequencing technologies: Promise and consequences. *Hum. Mutat.* **2016**, *37*, 1363–1367. [CrossRef] [PubMed]
84. Larsson, C.; Grundberg, I.; Söderberg, O.; Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nat. Methods* **2010**, *7*, 395. [CrossRef] [PubMed]
85. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945.
86. Institute NHGR. The Cost of Sequencing a Human Genome. 2016. Available online: https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/ (accessed on 31 January 2018).
87. Levy, S.; Sutton, G.; Ng, P.C.; Feuk, L.; Halpern, A.L.; Walenz, B.P.; Axelrod, N.; Huang, J.; Kirkness, E.F.; Denisov, G.; et al. The diploid genome sequence of an individual human. *PLoS Biol.* **2007**, *5*, e254. [CrossRef] [PubMed]
88. Schloss, J.A. How to get genomes at one ten-thousandth the cost. *Nat. Biotechnol.* **2008**, *26*, 1113–1115. [CrossRef] [PubMed]
89. Hayden, E.C. Technology: The $1,000 genome. *Nature* **2014**, *507*, 294–295. [CrossRef] [PubMed]
90. World, B.I. Venter Raises Stakes for "$1000 Genome" Prize. 2005. Available online: http://www.bio-itworld.com/newsitems/2005/oct2005/10-19-05-news-genome-prize/ (accessed on 31 January 2018).
91. Ledford, H. Genome sequencing X Prize announced. *Nat. News* **2006**, *443*, 733. [CrossRef] [PubMed]
92. Kedes, L.; Campany, G. The new date, new format, new goals and new sponsor of the Archon Genomics X PRIZE competition. *Nat. Genet.* **2011**, *43*, 1055–1058. [CrossRef] [PubMed]
93. Diamandis, P. Cancellation of the Archon Genomics XPRIZE: A Public Debate. 2014. Available online: https://genomics.xprize.org/news/blog/cancellation-of-archon-genomics-xprize-public-debate (accessed on 31 January 2018).
94. Wheeler, D.A.; Srinivasan, M.; Egholm, M.; Shen, Y.; Chen, L.; McGuire, A.; He, W.; Chen, Y.-J.; Makhijani, V.; Roth, G.T.; et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **2008**, *452*, 872–876. [CrossRef] [PubMed]
95. Wadman, M. James Watson's genome sequenced at high speed. *Nature* **2008**, *452*, 788. [CrossRef] [PubMed]
96. Resource TIGS. The 1000 Genomes Project. 2017. Available online: http://www.internationalgenome.org/ (accessed on 31 January 2018).
97. Institute NHGR. 1000 Genomes Project. 2017. Available online: https://www.genome.gov/27528684/1000-genomes-project/ (accessed on 31 January 2018).
98. Resource TIGS. About IGSR and the 1000 Genomes Project. 2017. Available online: http://www.internationalgenome.org/about (accessed on 31 January 2018).

99. Abecasis, G.R.; Altshuler, D.; Auton, A.; Brooks, L.D.; Durbin, R.M.; Gibbs, R.A.; Hurles, M.E.; McVean, G.A. A map of human genome variation from population-scale sequencing. *Nature* **2010**, *467*, 1061–1073. [PubMed]
100. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1092 human genomes. *Nature* **2012**, *491*, 56–65. [CrossRef] [PubMed]
101. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **2015**, *526*, 68–74. [CrossRef] [PubMed]
102. Sudmant, P.H.; Rausch, T.; Gardner, E.J.; Handsaker, R.E.; Abyzov, A.; Huddleston, J.; Zhang, Y.; Ye, K.; Jun, G.; Fritz, M.H.-Y.; et al. An integrated map of structural variation in 2504 human genomes. *Nature* **2015**, *526*, 75–81. [CrossRef] [PubMed]
103. Variety of life. *Nature* **2015**, *526*, 5–6. [CrossRef] [PubMed]
104. England, G. The 100,000 Genomes Project. 2017. Available online: https://www.genomicsengland.co.uk/the-100000-genomes-project/ (accessed on 31 January 2018).
105. Rabesandratana, T. 100,000 Genomes Project Gets £300 Million to Finish the Job by 2017 2014. Available online: http://www.sciencemag.org/news/2014/08/uks-100000-genomes-project-gets-300-million-finish-job-2017 (accessed on 31 January 2018).
106. Horgan, R.P.; Kenny, L.C. 'Omic' technologies: Genomics, transcriptomics, proteomics and metabolomics. *Obstetrician Gynaecol.* **2011**, *13*, 189–195. [CrossRef]
107. Chen, R.; Snyder, M. Promise of personalized omics to precision medicine. *Wiley Interdisciplin. Rev. Syst. Biol. Med.* **2013**, *5*, 73–82. [CrossRef] [PubMed]
108. Lappalainen, T.; Sammeth, M.; Friedländer, M.R.; Hoen, P.A.C.; Monlong, J.; Rivas, M.A.; Gonzàlez-Porta, M.; Kurbatova, N.; Griebel, T.; Ferreira, P.G.; et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **2013**, *501*, 506–511. [CrossRef] [PubMed]
109. Wang, K.; Huang, C.; Nice, E. Recent advances in proteomics: Towards the human proteome. *Biomed. Chromatogr.* **2014**, *28*, 848–857. [CrossRef] [PubMed]
110. Rifai, N.; Gillette, M.A.; Carr, S.A. Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nat. Biotechnol.* **2006**, *24*, 971–983. [CrossRef] [PubMed]
111. Beyond the genome. *Nature* **2015**, *518*, 273. [CrossRef] [PubMed]
112. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [CrossRef] [PubMed]
113. Skipper, M.; Eccleston, A.; Gray, N.; Heemels, T.; Le Bot, N.; Marte, B.; Weiss, U. Presenting the epigenome roadmap. *Nature* **2015**, *518*, 313. [CrossRef] [PubMed]
114. From Genetic Variations to Risk Prediction and Pharmacogenomics. Proceedings of the First Santorini Conference, Santorini, Greece, 25–28 September 2002. *Clin. Chem. Lab. Med.* **2003**, *41*, 397–614.
115. From Systems Biology and Functional Genomics to Personalized Health. 5th Biologie Prospective Santorini Conference, Island of Santorini, Greece, 30 September–2 October 2010. *Drug Metab. Drug Interact.* **2011**, *26*, 41–42.
116. Gurwitz, D.; Lunshof, J.E.; Dedoussis, G.; Flordellis, C.S.; Fuhr, U.; Kirchheiner, J.; Licinio, J.; Llerena, A.; Manolopoulos, V.G.; Sheffield, L.J.; et al. Pharmacogenomics education: International Society of Pharmacogenomics recommendations for medical, pharmaceutical, and health schools deans of education. *Pharmacogenomics J.* **2005**, *5*, 221–225. [CrossRef] [PubMed]
117. Llerena, A.; Michel, G.; Jeannesson, E.; Wong, S.; Manolopoulos, V.G.; Hockett, R.D.; Boubekeur, K.; Siest, G.; Beaune, P.; Heafliger, C.; et al. Third Santorini conference pharmacogenomics workshop report: "Pharmacogenomics at the crossroads: What else than good science will be needed for the field to become part of Personalized Medicine?". *Clin. Chem. Lab. Med.* **2007**, *45*, 843–850. [CrossRef] [PubMed]
118. Manolopoulos, V.G.; Dechairo, B.; Huriez, A.; Kuhn, A.; Llerena, A.; van Schaik, R.H.; Yeo, K.T.; Ragia, G.; Siest, G. Pharmacogenomics and personalized medicine in clinical practice. *Pharmacogenomics* **2011**, *12*, 597–610. [CrossRef] [PubMed]
119. Siest, G.; Auffray, C.; Taniguchi, N.; Ingelman-Sundberg, M.; Murray, H.; Visvikis-Siest, S.; Ansari, M.; Marc, J.; Jacobs, P.; Meyer, U.; et al. Systems medicine, personalized health and therapy. *Pharmacogenomics* **2015**, *16*, 1527–1539. [CrossRef] [PubMed]

120. Siest, G.; Ndiaye, N.C.; El Shamieh, S.; Shahabi, P.; Stathopoulou, M.; Saleh, A.S.; Godjo, T.; Albertini, L.; Visvikis, V.-S. Conference scene: Systems biology and personalized health science and translation. *Pharmacogenomics* **2013**, *14*, 1953–1964. [CrossRef] [PubMed]
121. Visvikis-Siest, S.; Aldasoro Arguinano, A.A.; Stathopoulou, M.; Xie, T.; Petrelis, A.; Weryha, G.; Froguel, P.; Meier-Abt, P.; Meyer, U.A.; Mlakar, V.; et al. 8th Santorini Conference: Systems medicine and personalized health and therapy, Santorini, Greece, 3–5 October 2016. *Drug Metab. Pers. Ther.* **2017**, *32*, 119–127. [CrossRef] [PubMed]
122. ESPT. A Scientific Society for Individualised Medicine. 2017. Available online: https://esptnet.eu/S1/Home (accessed on 31 January 2018).
123. Medicine EAfP. The Right Pervention and Treatment for the Right Patient at the Right Time. 2017. Available online: https://www.euapm.eu/ (accessed on 31 January 2018).
124. Cranley, J.; MacRae, C.A. A new approach to an old problem: One brave idea. *Circ. Res.* **2018**, *122*, 1172–1175. [CrossRef] [PubMed]
125. Lau, E.; Wu, J.C. Omics, big data, and precision medicine in cardiovascular sciences. *Circ. Res.* **2018**, *122*, 1165–1168. [CrossRef] [PubMed]
126. Mullin, E. 2017 Was the Year of Gene-Therapy Breakthroughs: MIT Technology Review. 2018. Available online: https://www.technologyreview.com/ (accessed on 31 January 2018).
127. Amoasii, L.; Hildyard, J.C.W. Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science* **2018**. [CrossRef] [PubMed]

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Publication Nº 2

Epigenome-Wide Association Study (EWAS) of Blood Lipids in Healthy Population from STANISLAS Family Study (SFS).

Ting Xie*, **Vesna Gorenjak***, Maria G. Stathopoulou*, Sébastien Dadé, Eirini Marouli, Christine Masson, Helena Murray, John Lamont, Peter Fitzgerald, Panagiotis Deloukas and Sophie Visvikis-Siest

*equal contribution

Int J Mol Sci, February 2019, 26;20(5)

The prevention of CVDs requires a fast detection of dysregulation of cardiovascular risk factors. Blood lipids, *i.e.* HDL, LDL, TC and TGs, are strongly associated with adverse cardiovascular outcomes, as well as with other chronic diseases, such as diabetes (246, 289).

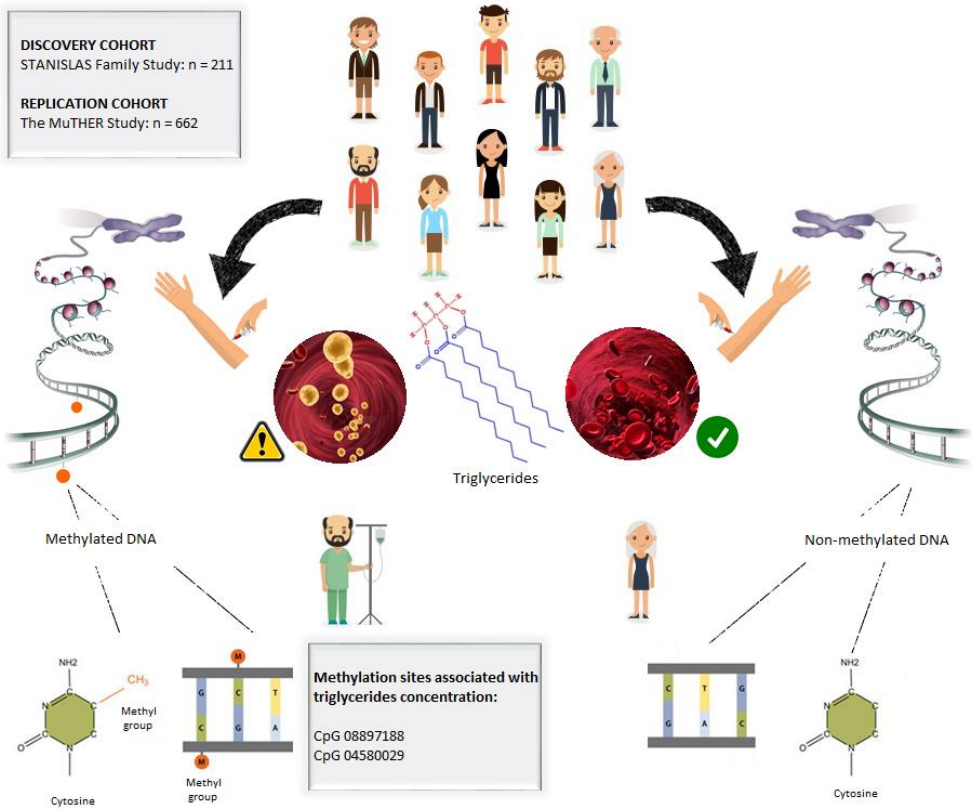


Figure 26: Graphical abstract of the published article.

Epigenetics mechanisms are implicated in the regulation of genes involved in metabolism and regulation of blood lipid levels (290). Studying the epigenetic regulation of blood lipids in an individual could enable the insight into the susceptibility for common chronic diseases even before the alert values of a particular phenotype (*e.g.* increased TG levels) are expressed (Figure 26).

Objective

The aim of the study was to perform an EWAS on blood lipid levels in a healthy population, in order to unravel the mechanisms of epigenetic regulation of blood lipids with the detection of CpG methylation sites associated with the levels of TG, TC, LDL and HDL.

Populations and analysis

Discovery cohort consisted of 211 healthy adults and children (75 families) from the first recruitment of the SFS. Blood DNA methylation assessment was performed by Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA). EWAS was carried out using the R packages minfi and CpGassoc in a linear mixed-effect model, accounting for age, gender, BMI, smoking and individual blood cell counts as fixed effect, and methylation chip and family structure as random effect.

Replication of the significant results from the discovery cohort was performed in the adipose tissue of 662 participants from the MuTHER study. Methylation profiling was performed with Illumina's Infinium HumanMethylation450 BeadChip. Raw data was imported to the GenomeStudio v.2010.3 software whereas the association analysis of the two significant probes (detected in the discovery cohort) with TG levels was performed by the CpGassoc package in R, in a linear mixed-effect model, accounted for age, gender, BMI and smoking as fixed effect, and methylation chip as random effect.

Genotyping was performed by the Infinium CoreExome Illumina assay. SNPs related to the phenotype of interest and located in the same chromosome as the significant methylation sites were selected from the NHGRI-EBI GWAS catalogue.

Results

From the four blood lipid phenotypes assessed in this study, significant results were observed only with TG levels (FDR < 0.05). More specifically, two probes demonstrated statistically significant associations (Table 2), namely cg08897188 located in the *PRKAG2* gene on chromosome 7 (7q36.1) and cg04580029 located in the *KREMEN2* gene on chromosome 16 (16p13.3).

Table 2. Association of methylation sites with TG levels.

CpG	Gene Name	Chromosome	Effect size	SE	P-Value	FDR
cg08897188	*PRKAG2*	7q36.1	-2.80	0.47	1.39×10^{-8}	0.049
cg04580029	*KREMEN2*	16p13.3	3.09	0.51	5.75×10^{-9}	0.049

SE: Standard error, **FDR:** False discovery rate (< 0.05)

In the replication cohort, the two significant probes cg08897188 and cg04580029 were determined in DNA acquired from the adipose tissue and were tested for the association with TG levels. An association at a nominal significance level was identified for cg04580029 (Table 3).

Table 3. Replication of the significant results in adipose tissue.

CpG	Gene Name	Effect size	SE	P-Value*
cg08897188	*PRKAG2*	-0.0008	0.0016	0.6265
cg04580029	*KREMEN2*	0.0084	0.0036	0.0196

SE: Standard error, *adjusted P-Value

Bioinformatics analysis showed that the replicated probe cg04580029 is located in the promoter region of the *KREMEN2* gene, three base pairs away of the transcription start site of the *KREMEN2* gene, overlapping the DNase hypersensitivity peak and histone peak. Probe cg08897188, on the other hand, is located between the *PRKAG2* gene and *AC093583.1* gene, in the proximity of transcription factor binding sites and histone peaks.

SNPs previously associated with TG levels in GWAS studies and located in the same chromosomes as the identified methylation sites were tested for their associations with the two methylation sites because DNA sequence variance can impact on methylation levels.

Altogether, 16 SNPs located on chromosomes 7 and 16 were identified. None of the polymorphisms showed the association with the identified CpG sites.

Conclusion

The study '*Epigenome-Wide Association Study (EWAS) of Blood Lipids in Healthy Population from STANISLAS Family Study (SFS)'* gave new insights into epigenetic regulation of TG levels, which could be useful in the future as a diagnostic tool for detecting early dysregulation of cardiovascular risk factors. Moreover, it revealed a link between DNA methylation and TG levels through possible metabolic pathways that could be associated with different diseases and would merit further investigation for future implication in personalised medicine.

A limitation of our study could be the use of the combined samples of related children (>6 years) and adults in the discovery population. The reason for such composition was the increased sample size in the combined population, which in turn increased the power of the analysis and helped us to identify the significant association. To avoid the spurious results, we performed the necessary adjustments for age, sex, BMI and family structure along with the analysis. Nevertheless, the replication of a result in the adult-only cohort confirmed that a combination of adult and children samples in the discovery cohort was not limiting for research.

International Journal of
Molecular Sciences

MDPI

Article

Epigenome-Wide Association Study (EWAS) of Blood Lipids in Healthy Population from STANISLAS Family Study (SFS)

Ting Xie [1,†], Vesna Gorenjak [1,†], Maria G. Stathopoulou [1,†], Sébastien Dadé [1], Eirini Marouli [2], Christine Masson [1], Helena Murray [3], John Lamont [3], Peter Fitzgerald [3], Panagiotis Deloukas [2] and Sophie Visvikis-Siest [1,4,*]

1 Université de Lorraine, Inserm, IGE-PCV, 54000 Nancy, France; ting.xie@inserm.fr (T.X.); gorenjak.vesna@gmail.com (V.G.); maria.stathopoulou@inserm.fr (M.G.S.); sebastien.dade@univ-lorraine.fr (S.D.); christine.masson@univ-lorraine.fr (C.M.)
2 Queen Mary University of London, E1 4NS London, UK; e.marouli@qmul.ac.uk (E.M.); p.deloukas@qmul.ac.uk (P.D.)
3 Randox Laboratories Ltd, Crumlin, BT29 4XL County Antrim, UK; helena.murray@randox.com (H.M.); john.lamont@randox.com (J.L.); peter.fitzgerald@randox.com (P.F.)
4 Department of Internal Medicine and Geriatrics, CHU Nancy-Brabois, 54000 Nancy, France
* Correspondence: sophie.visvikis-siest@inserm.fr; Tel.: +33-6-07-60-25-69
† These authors contributed equally to this work.

Received: 9 January 2019; Accepted: 21 February 2019; Published: 26 February 2019

check for updates

Abstract: Epigenome-Wide Association Studies (EWAS) are furthering our knowledge of epigenetic modifications involved in the regulation of lipids' metabolism. Furthermore, epigenetic patterns associated with lipid levels may play an important role in predicting the occurrence of cardiovascular events. To further investigate the relationship between methylation status and lipids, we performed an EWAS in 211 individuals from the STANISLAS Family study (SFS). Methylation at two CpG sites (*PRKAG2*; $p = 1.39 \times 10^{-8}$; *KREMEN2*; $p = 5.75 \times 10^{-9}$) were significantly associated with lipidomic profiles. Replication was sought in adipose tissue where one probe, cg08897188, was found to be nominally significant (*KREMEN2*; $p = 0.0196$). These results could provide new insight in the mechanisms underlying cardiovascular diseases and contribute to new therapeutic interventions.

Keywords: genetics; gene Expression/polymorphisms; lipids; adipose tissue; dyslipidemias; EWAS; DNA methylation; epigenetics

1. Introduction

Our health and well-being depend on diverse interactions between clinical, biological, environmental and lifestyle factors, which, all together, are involved in complex pathological mechanisms that may influence the disease susceptibility, development and prognosis of the treatment [1]. Dysregulated lipid levels, commonly referred to as dyslipidemia, are one of the leading risk factors for cardiovascular diseases (CVD), but also influence other chronic conditions such as type-2 diabetes mellitus [2]. Type-2 diabetes patients and individuals suffering from infectious diseases (e.g., AIDS and tuberculosis) were characterized by abnormal profiles of high triglycerides (TG), low levels of high density lipoprotein (HDL), high serum levels of very low density lipoproteins (VLDL) and increased LDL [2]. Cholesterol is one of the first biomarkers used in routine clinical diagnostics, and it can be well regulated with common treatment in patients with abnormal lipid profiles [3]. However, TG are now becoming the subject of new investigations in order to improve diagnosis and, potentially, in the prevention and treatment of CVD. Their predictive value for the onset of the

pathology is considered to be even higher than that of cholesterol [4,5]. With further research, potential new medications for the normalization of TG levels could be developed and used together with statins to help further reduce risk of chronic diseases. CVD is the leading cause of premature mortality and morbidity in occidental countries, and is an essential contributing factor in the rise of health care costs [6]. The study of factors affecting lipids' levels is an important field in the fight against CVD and other morbidities.

Lipid traits were found to be highly heritable, especially plasma cholesterol and TG concentrations (55–77%). However, changes in our environment can make modest, but clinically important changes regarding common risk factors [7]. The genetics of blood lipids have been assessed in a large genome-wide association study (GWAS), where as many as 157 lipid-associated loci were discovered (8), accounting for more than 12% of the total variance [8]. Strong associations were found between variants near *LIPC* gene with plasmalogen levels ($p < 10^{-40}$), and variants near *ABCA1* gene with sphingomyelin levels ($p < 10^{-5}$) [9]. These variants were previously associated with cardiovascular and metabolic traits including coronary artery disease, type 2 diabetes, blood pressure, waist/hip ratio, and body mass index, showing that even variants with small effect size can expose new pathways and therapeutic targets [9,10].

Epigenome-wide association studies (EWAS) are the epigenetic equivalent of GWAS, giving us information about associations between epigenomic perturbations and traits related to human diseases [11]. They allow us to assess the environmental impact on genetic regulation. With progress in epigenetics, new insights on disease mechanisms are possible through the exploration of the DNA accessibility and chromatin structure, and are enabling us to explain the regulatory mechanisms of gene expression. Epigenetic variations could explain parts of missing heritability of chronic diseases that have not yet been elucidated by GWAS [12]. The most common mechanism of epigenetic patterns is DNA methylation, which is the forming of 5-methylcytosine on the CpG (cytosine-phosphate-guanine) site of the genome; it normally results in silencing of the gene encoded in the sequence [13].

Recent EWAS studies are furthering the current knowledge of epigenetic modifications by investigating patterns associated with methylation at CpG sites in relation to the regulation of lipids. Thirty-three CpG sites were discovered and replicated in recent studies, located in genes associated with TG and HDL cholesterol, such as a reverse cholesterol transporter (*ABCG1*; $p = 7.2 \times 10^{-28}$). Pathway analysis showed the involvement of CpG sites in lipid, sterol, and cholesterol metabolic and biosynthetic processes [14]. Also, a differentially methylated locus was associated with CVD events (hazard ratio per SD increment, 1.38; 95% confidence interval, 1.15–1.66; $p = 0.0007$). These findings demonstrate an important role of epigenetic patterns in lipid metabolism and in predicting the occurrence of CVD events [14]. Several epigenetically regulated genes were shown to alter lipid levels, including carnitine palmitoyltransferase 1A *(CPT1A)*, ATP-binding cassette sub-family G member 1 *(ABCG1)*, sterol regulatory element binding transcription factor 1 *(SREBF1)*, troponin T1 (*TNNT1)*, microRNA 33b (*MIR33B)*, NFAIP3 interacting protein 3 *(TNIP)* and 24-dehydrocholesterol reductase (*DHCR24)* [14–18]. A differential methylation observed in EWAS was found to be a consequence, rather than a cause, of blood lipid abnormalities, since differential methylation was induced by TG, LDL and HDL, whereas there was no effect of DNA methylation on lipid levels in prime circulating immune cells [19].

All the aforementioned studies were performed on subjects suffering from cardiovascular condition. Regulation of lipid traits can differ in healthy population and new variants involved in gene regulation pathways can be discovered, which can contribute to a better understanding of the complex mechanisms involved in lipid metabolism. The aim of the current study was to fill this knowledge gap and analyze the association of epigenetic methylation patterns with lipid phenotypes (TG, cholesterol, LDL and HDL) in a healthy population and to expand the recent knowledge in the epigenetics of lipid profiles.

2. Results

2.1. Associations between Genome-Wide DNA Methylation and Lipid Levels in Whole Blood

Two probes demonstrated statistically significant levels of association with TG concentrations, including individual probes in *PRKAG2* on chromosome 7 (7q36.1), and *KREMEN2* on chromosome 16 (16p13.3) (Table 1). No significant associations were found between DNA methylation and other assessed lipid phenotypes ($p < 0.05$).

Table 1. Association of methylation values with TG level.

CpG	Gene Name	Chromosome	Effect Size	SE	Unadjusted *p*-Value	FDR
cg08897188	*PRKAG2*	7q36.1	−2.80	0.47	1.39×10^{-8}	0.049
cg04580029	*KREMEN2*	16p13.3	3.09	0.51	5.75×10^{-9}	0.049

SE: Standard error, FDR: False discovery rate (<0.05).

2.2. Bioinformatics Analysis

The cg08897188 probe maps on chromosome 7q36.1 and is located both within an intron of the *PRKAG2* gene and in the opposite strand within the *AC093583.1* non-coding transcript (Figure 1A). This gene has 2 splice variants: AC093583.1-201 (LOC644090) and AC093583.1-202. Cg08897188 is located in exon 3 of LOC644090 and in silico analysis showed the presence of transcription factor binding sites (Figure 1B).

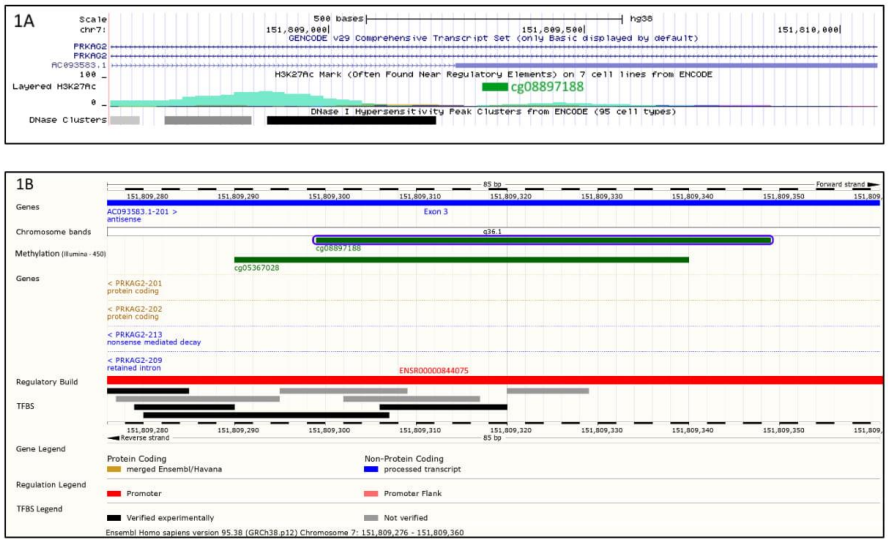


Figure 1. Environment of cg08897188 probe. As depicted by Figure 1A,B, cg08897188 (in green) is located on chromosome 7q36.1 within intron of *PRKAG2* gene of the forward strand and within exon 3 of the *AC093583.1* non-coding transcript in the opposite strand. (**A**) shows that cg08897188 is surrounded by regulatory elements indicated by acetylation of histone 3 on lysine 27 (H3K27Ac peaks, in turquoise). Moreover, as indicated by DNase I hypersensitivity peak clusters (black and grey rectangles), this region has also an accessible chromatin zone, indicating a transcriptional activity. (**B**) confirms that cg08897188 is located within a regulatory zone (promoter ENSR00000844075, in red) with clearly identified transcription factor binding sites (black and grey rectangles).

The cg04580029 on the chromosome 16p13.3 is located in the promoter flank of the *KREMEN2* gene (ENSR00000529174) (Figure 2). The cg04580029 probe fully overlaps a DNase I hypersensitivity site and a H3K27Ac histone peak (Figure 2A). Further analysis showed that cg04580029 is allocated 3 base pairs away of the transcription start site of *KREMEN2* gene, p2@KREMEN2 (Figure 2B).

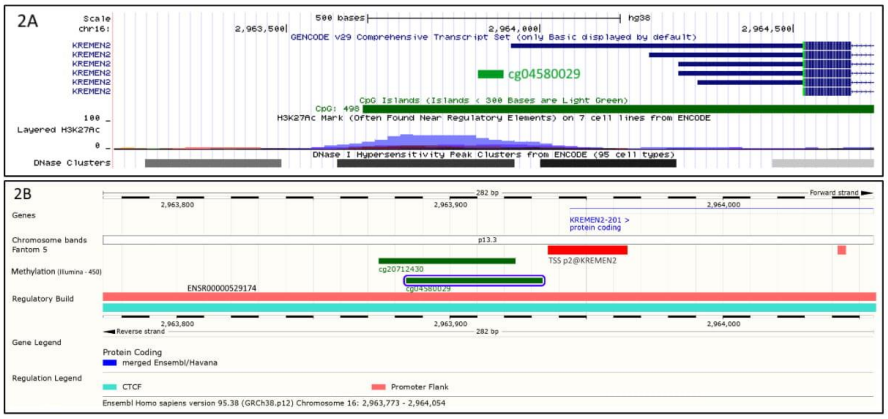


Figure 2. Environment of cg04580029 probe. As depicted by Figure 2A,B, cg04580029 (in green) is located on the chromosome 16p13.3 upstream of *KREMEN2* gene. (**A**) shows that cg04580029 is located in a regulatory element, indicated by acetylation of histone 3 on lysine 27 (H3K27Ac peak, in violet). Furthermore, DNase I hypersensitivity peak clusters (black and grey rectangles) reveal the presence of an accessible chromatin zone which is the sign of a transcriptional activity. (**B**) confirms that cg04580029 is located within a regulatory zone (promoter flank ENSR00000529174, in light red) and shows in addition that 3 base pairs away is located the transcription start site of *KREMEN2* gene, p2@KREMEN2.

Analysis of the regulatory activity of the ENSR00000529174 promoter, where the cg04580029 is involved, shows that, depending on the level of macrophage differentiation, the promoter is either in an active or inactive state (Figure 3).

	Promoter ENSR00000529174 of KREMEN2 gene
M0 macrophage (CB)	Active
M0 macrophage (VB)	Active
M1 macrophage (CB)	Inactive
M1 macrophage (VB)	Inactive
M2 macrophage (CB)	Active
M2 macrophage (VB)	Inactive
	Cg04580029

Figure 3. Regulatory status of the ENSR00000529174 promoter according to the macrophage cell type (CB: cord blood; VB: vein blood).

2.3. Associations between Genetic Variants and Methylation Sites

DNA sequence variation could influence methylation levels, thus we selected SNPs previously associated with TG levels in GWAS studies located in the same chromosomes of the identified methylation sites (Table 1) to test their associations with methylation levels. Sixteen polymorphisms

have been identified in the chromosomes 7 and 16, which have previously been associated with TG levels. Cut off point for significance was set to 0.003, but no significant *cis* methylation QTL was detected.

2.4. Replication of the Identified CpGs Associations with Lipids Levels in Adipose Tissue

The associations of the 2 probes with TG were tested in adipose tissue and the results are presented in Table 2. One of these associations was also identified, however, in nominal significance level (cg04580029).

Table 2. Replication of the associations of the 2 probes with TG levels in adipose tissue.

CpG	Gene Name	Effect Size	SE	*p*-Value
cg08897188	*PRKAG2*	−0.0008	0.0016	0.6265
cg04580029	*KREMEN2*	0.0084	0.0036	0.0196

SE: standard error.

3. Discussion

This study confirmed the involvement of DNA methylation in the regulation of TG levels through possible metabolic pathways that are associated with the development of chronic diseases. Knowing how the complex processes of methylation can impact disease development might highlight new targets for treatment and contribute to the improvement of personalized strategies of patient care. We reported two new CpG sites, cg08897188 and cg04580029 that are associated with TG levels in a healthy population and replicated the cg04580029 in the adipose tissue.

Adipose tissue is composed of adipocytes in its parenchymal part and also contains a stromal vascular fraction (STV) for its mesenchymal part. Thus, STV includes preadipocytes, fibroblasts, vascular endothelial cells and a variety of immune cells such as macrophages. It is therefore not surprising to note that the promoters of these cell-types are active and can be regulated by CpG methylation (Figure 3).

The cg08897188 probe is located closed to the distal end of a DNase I hypersensitive site and a H3K27Ac histone peak which is a mark for active transcription. As shown in Figure 1A, the proximity of cg08897188 to the H3K27Ac histone peak implies that it may be involved in the transcriptional regulation of this region. The region harbors the *PRKAG2* gene, coding for a protein involved in the regulation of fatty acid metabolic process [20] and fatty acid oxidation [21], as well as an EST-based antisense non-coding transcript LOC644090 (AC093583.1-201) on the opposite strand (Figure 1A,B). Therefore, the effect of cg08897188 could be exerted either via enhancing the transcription of LOC644090 which in turn down-regulates *PRKAG2* expression or directly modulating *PRKAG2* expression. PRKAG2 is an AMP/ATP-binding subunit of AMP-activated protein kinase (AMPK), an energy sensor protein kinase that plays a key role in regulating cellular energy metabolism and therefore may be implicated in fatty acid oxidation and modulation of TG level. The mutations in *PRKAG2* were first related to familial hypertrophic cardiomyopathy (HCM) [22], yet Arad et al. revealed that the *PRKAG2* mutations affect metabolic cardiomyopathy rather than HCM [23]. Finally, the mutation of *PRKAG2* is known to cause the glycogen-storage cardiomyopathy that resembles HCM but is distinguished by electrophysiological abnormalities [24]. Our results might support these findings, since the probe cg08897188 has a specific region that is able to bind MZF1 (Myeloid zinc finger 1 factors), which are associated with arrhythmogenic right ventricular dysplasia or cardiomyopathy (ARVD/C) with a similarity score of 0.992 [25]. There may be an association between the methylation profile that affects TG levels and a heritable heart muscle disease (ARVD) (Figure 4). However, further experiments are required to confirm this assumption and to make further progress in the investigation of the disease mechanisms.

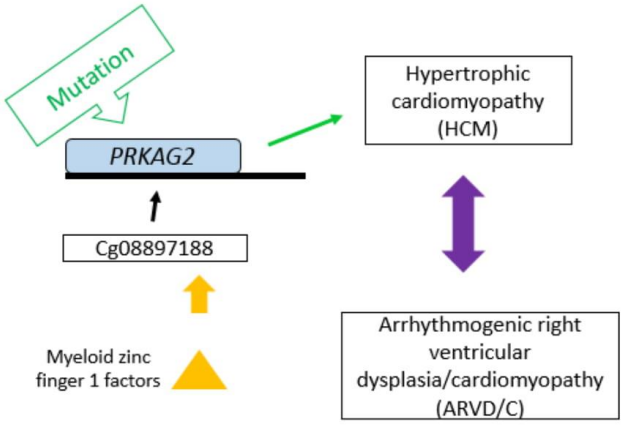


Figure 4. Possible mechanism of association of cg08897188 with CVD. Cg08897188 is involved in regulation of the *PRKAG2* gene. Mutation of *PRKAG2* can cause HCM or ARVD/C. Cg08897188 has a specific region able to bind myeloid zinc finger 1 factor, which is also associated with ARVD/C.

The cg04580029 is located in the promoter flank of *KREMEN2* gene, only 3 bp away of the TSS (p2@KREMEN2—Figure 2B). The observed overlap with a H3K27Ac histone peak which is a mark for active transcription, implies that modulation of methylation levels at cg04580029 will directly impact *KREMEN2* expression. It is well known that LRP6 (low-density lipoprotein-related receptors 6) is a co-receptor of WNT that transmits the canonical Wnt/β-catenin signaling cascade, which has been associated with many diseases including Alzheimer's disease [26] and coronary artery disease (CAD) [27,28]. In the canonical way, extracellular Wnt protein binds to a complex transmembrane receptor composed of Frizzled (Fz) and LRP (LDL-related protein) allowing to recruit intracellular Dishevelled (Dsh) protein. Dsh, in turn, recruit GSK3 (glycogen synthase kinase) protein that will not be able to phosphorylate β-catenin, which avoids to be degraded by proteasome. Thus, β-catenin can act as a transcription factor and stimulate gene expression. Due to its importance, this signaling pathway is closely controlled, in particular by an extracellular inhibitor called Dickkopf (DKK), which interacts with both the LRP5/6 co-receptor and the KREMEN1/2 proteins, and thus prevents Wnt dimerization of the LRP and Fz co-receptors (39). KREMEN2 protein prevents Wnt dimerization of the co-receptors LRP and Fz. Thus, a regulation of KREMEN2 via cg04580029 methylation near transcription start site might remove inhibition pathway of Wnt and in turn, β-catenin could stimulate specific genes expression leading to modulate TG level. Interestingly, our results found that the KREMEN2-related CpG site cg04580029 was significantly associated with TG levels in a healthy population. Therefore, our results suggest that KREMEN2 could be proposed as a link between TG and Alzheimer's disease and CAD through methylation profiles. Furthermore, this association was also present in the adipose tissue, however at a nominal level of significance.

TG-related differential methylation has previously been described in several genes [14,18]. Methylation on *ABCG1* locus, associated with expression of a gene involved in reverse cholesterol transport, was associated with a 38% higher risk of chronic heart disease per standard deviation increase in methylation [14,18]. CpG sites combined in a methylation risk score explained up to 9% of the variance in TG [18].

A limitation of our study is that although we have replicated the *KREMEN2* signal (cg04580029) in adipose tissue, we did not have access to an obese population to further validate this signal, which should be considered in a future study.

In conclusion, our findings demonstrate novel associations between DNA methylation and TG levels, and propose some mechanisms that relate these associations with pathways affecting different

diseases in humans that merit further investigation. These results could provide new insight in the mechanisms underlying CVD and contribute to new therapeutic interventions.

4. Materials and Methods

4.1. Participants

The discovery cohort of this study consists of 211 individuals from 73 families of the STANISLAS Family Study (SFS). The SFS is a 10-year longitudinal survey with 3 visits at 5-year intervals, involving 1,006 families from Vandoeuvre-lès-Nancy, France, first recruited between 1993 and 1995 [29]. It is a part of the Biological Resources Center "Interactions Gène-Environnement en Physiopathologie CardioVasculaire" (BRC-IGE-PCV, number BB-0033-00051). All subjects used for the study were of European-Caucasian origin, without the presence of serious and/or chronic disorders and personal history of CVD. The study protocols were approved by institutional ethics committees and all subjects gave written informed consent for their participation in the study. Population characteristics for the discovery cohort are compiled in Table 3.

Table 3. Populations characteristics of discovery cohort.

Population Characteristics	Total		Adults (116)		Children (95)		Male (105)		Female (106)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age (years)	28.17	14.83	40.48	7.53	13.15	2.58	27.09	15.14	29.24	14.37
Body Mass Index (kg/m^2)	21.52	4.00	24.06	3.22	18.43	2.31	21.77	4.17	21.27	3.79
Cholesterol (mmol/L)	5.24	1.00	5.70	0.96	4.68	0.74	5.12	1.04	5.37	0.95
Triglycerides (mmol/L)	0.86	0.49	0.98	0.56	0.72	0.32	0.91	0.55	0.82	0.41
High density lipoprotein (mmol/L)	1.44	0.41	1.45	0.42	1.43	0.39	1.34	0.38	1.54	0.42
Low density lipoprotein (mmol/L)	3.63	0.97	4.06	0.96	3.11	0.68	3.59	1.02	3.67	0.91

The replication cohort consists of 662 adipose tissue samples collected in the MuTHER study. The MuTHER study includes 856 female European-descent individuals aged 59.12±94.44 years, recruited from the TwinsUK Adult Twin Registry. All the procedures followed were in accordance with the ethical standards of the St. Thomas' Research Ethics Committee (REC reference 07/H0802/84) at St. Thomas' Hospital in London, and all study subjects provided written informed consent.

4.2. Data Collection

For the SFS cohort, participant information including personal health, lifestyle, biological and clinical data was collected using appropriate, validated questionnaires and procedures [29,30]. Body mass index (BMI) was expressed as weight (kilograms) divided by height^2·(meters)^2. Blood samples were collected between 8 and 9 a.m. after overnight fasting. DNA was extracted by the Miller technique [31] and was stored at –80 °C until further use.

For the MuTHER study, 8 mm punch biopsies were taken from a relatively photo-protected area adjacent and inferior to the umbilicus. Subcutaneous adipose tissue was carefully dissected from each biopsy, weighted and split into multiple pieces, and immediately stored in liquid nitrogen until analysis.

4.3. Biological Measurements

For SFS cohort, serum total cholesterol (TC) and serum TG levels were measured using standard enzymatic methods (Merck, Darmstadt, Germany), automated on AU5021 (Olympus; Merck). Serum HDL was measured by phosphotungstate precipitation on a Cobas-Mira (Roche, Basel, Switzerland), while LDL levels were calculated using the Friedewald formula [32]. Immunonephelometry was used for the detection of serum Apolipoprotein (Apo) B levels on Behring Nephelometer Analyser, with

Behring reagents (Behring Diagnostics, Rueil-Malmaison, France) and serum ApoE was measured by turbidimetry [33].

4.4. DNA Methylation Assessment

4.4.1. Discovery Cohort

Genome-wide DNA methylation profiling was performed by Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA) [34], which performs two measurements for each CpG: a methylated and unmethylated intensity. DNA methylation is described as a β value, a ratio between methylated and unmethylated intensities. A detection *p*-value was generated for every CpG with minfi R package [35]. Poor quality probes were excluded from the analysis using a detection *p*-value cutoff >0.05. Furthermore, probes, missing in >5% of the samples were excluded from all samples. Based on these two criteria, 764 probes were removed from analysis. To avoid spurious associations, the genomic information of probes already annotated in HumanMethlyation450 annotation files (probes containing single-nucleotide polymorphism (SNP), sex chromosomes, and a single base extension (SBE) site) was excluded. Finally, probes containing cross reactive and target polymorphic CpGs [36] were excluded, leaving 373626 probes for statistical analysis.

Before the association analysis, background correction and normalization were carried out with Illumina background correction and SWAN [37] in the R package minfi, respectively. One individual from our cohort was excluded after quality control checks of methylation array data (outlier of plotted median of the methylated against unmethylated samples, with the signal below 10.5), leaving 210 individuals for analysis (Table 3).

4.4.2. Replication Cohort

Methylation levels were estimated at 485,764 sites covering not only gene promoters but also several other genomic features in subcutaneous adipose tissue derived from the female twins as described in [38]. In brief, prior to bisulphite conversion, DNA samples were randomized, and exactly 700 ng of each DNA sample was taken for bisulphite conversion with the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, CA, USA) according to the supplier's protocol. Methylation profiling was performed with Illumina's Infinium HumanMethylation450 BeadChip. Arrays were scanned with the IlluminaHiScan SQ scanner, and raw data were imported to the GenomeStudio v.2010.3 software with the methylation module 1.8.2 for the extraction of the image intensities.

4.5. Genotyping and Selection of Single-Nucleotide Polymorphisms (SNPs)

Genotyping was performed by the Infinium CoreExome Illumina assay. Significant TG associated SNPs were selected from the NHGRI-EBI GWAS catalog [39], located in the same chromosome as the significant methylation sites.

4.6. Statistical Analysis

Levels of TC, TG, LDL and HDL for discovery cohort were not normally distributed and were therefore transformed on the e-log scale. A linear mixed effects model was used to analyze the association between methylation levels at each probe and log-transformed phenotypes accounting for relatedness between individuals. The model used was including gender, age, BMI, family structure, smoking and individual blood cell counts (neutrophils, lymphocytes, monocytes, eosinophils and basophils) as covariates, and chip array as random effect. Q values were estimated for false discovery rates [40] and *q* values (<0.05) were considered for defining the statistical significance. False discovery rate (FDR) method [40] was used for the correction of the results for multiple testing and Bonferroni correction method (cut-off 3.35×10^{-8}). The analysis of association of methylation values and the assessed phenotypes were performed by Package CpGassoc in R [41]. HumanMethlyation450 annotation files were used for annotating the probes and their corresponding genes.

Levels of TC in replication cohort followed the same statistical analysis procedure. The model used was adjusted for age, BMI, smoking status, family identification number, array batch and zygosity in a linear mixed effects model.

A linear mixed effects model with age, gender, BMI, and smoking as fixed effects, and methylation chip, and family structural as a random effect was carried out for the genetic association analysis using the statistical package lmerTest of R [42]. Significance cut off was set to 0.05/n where n is the number of significant TG associated SNPs in previous GWAS studies.

Deviation of Multiple Correlation Squared ρ2 From Constant (Random Model) of G*Power software was used in order to calculate statistical power [7], specifically the post hoc power analysis procedure. This procedure comprises a parameter analysis; it requires the test type (tails: one or two), effect size (H1 ρ2 and H0 ρ2), α error probability, total sample size, and number of predictors.

4.7. In Silico Analysis

The significant CpG sites were localized on the Human genome (GRCh38.p12) using Ensembl browser [43] and UCSC browser [44] as well as the establishment of regulatory feature activity of their respective promoters.

4.8. Data Access

Data is freely available upon request to the Biological Resources Centre IGE-PCV (http://www.u1122.inserm.fr/).

Author Contributions: V.G. and T.X. participated in study design, performed data analysis and interpretation and drafted the manuscript. S.D. participated in bioinformatics analysis and drafting of the manuscript. Epigenome-wide methylation analysis pipeline was designed by E.M. and P.D., C.M. prepared biological material and experiments. H.M., J.L. and P.F. contributed with biological measurements. M.G.S. and S.V.-S. were involved in study design and writing and obtaining grants for the funding of the projects. All authors participated in the interpretation of data and review of the manuscript. They all approved the final version of the manuscript.

Funding: The research was funded by the European Union within the frames of the Operational Programme FEDER-FSE Lorraine et Massif des Vosges 2014-2020 and by the Agence Nationale de la Recherche, programme d'Investissements d'avenir", grant number ANR-15RHU-0004. T.X. and V.G. were partially funded by the "Région Lorraine".

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vanuzzo, D.; Pilotto, L.; Mirolo, R.; Pirelli, S. Cardiovascular Risk and Cardiometabolic Risk: An Epidemiological Evaluation. *G. Ital. Cardiol. (Rome)* **2008**, *9* (Suppl. 1), 6s–17s.
2. Nayak, S.B.; Mohammed, S.B.; Nayak, A.S. Controlling Lipids Aids in the Prevention of Type 2 Diabetes, Hypertension, and Cardiovascular Diseases. *Int. J. Prev. Med.* **2017**, *8*, 39. [CrossRef] [PubMed]
3. Martin, M.J.; Hulley, S.B.; Browner, W.S.; Kuller, L.H.; Wentworth, D. Serum Cholesterol, Blood Pressure, and Mortality: Implications from a Cohort of 361,662 Men. *Lancet* **1986**, *2*, 933–936. [CrossRef]
4. Arsenault, B.J.; Boekholdt, S.M.; Kastelein, J.J. Lipid Parameters for Measuring Risk of Cardiovascular Disease. *Nat. Rev. Cardiol.* **2011**, *8*, 197–206. [CrossRef] [PubMed]
5. Hokanson, J.E.; Austin, M.A. Plasma Triglyceride Level Is a Risk Factor for Cardiovascular Disease Independent of High-Density Lipoprotein Cholesterol Level: A Meta-Analysis of Population-Based Prospective Studies. *J. Cardiovasc. Risk* **1996**, *3*, 213–219. [CrossRef] [PubMed]
6. Pirani, N.; Khiavi, F.F. Population Attributable Fraction for Cardiovascular Diseases Risk Factors in Selected Countries: A Comparative Study. *Mater. Sociomed.* **2017**, *29*, 35–39. [CrossRef] [PubMed]
7. Faul, F.; Erdfelder, E.; Buchner, A.; Lang, A.G. Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behav. Res. Methods* **2009**, *41*, 1149–1160. [CrossRef] [PubMed]
8. Teslovich, T.M.; Musunuru, K.; Smith, A.V.; Edmondson, A.C.; Stylianou, I.M.; Koseki, M.; Pirruccello, J.P.; Ripatti, S.; Chasman, D.I.; Willer, C.J.; et al. Biological, Clinical and Population Relevance of 95 Loci for Blood Lipids. *Nature* **2010**, *466*, 707–713. [CrossRef] [PubMed]

9. Willer, C.J.; Schmidt, E.M.; Sengupta, S.; Peloso, G.M.; Gustafsson, S.; Kanoni, S.; Ganna, A.; Chen, J.; Buchkovich, M.L.; Mora, S.; et al. Discovery and Refinement of Loci Associated with Lipid Levels. *Nat. Genet.* **2013**, *45*, 1274–1283. [CrossRef] [PubMed]
10. Musunuru, K.; Strong, A.; Frank-Kamenetsky, M.; Lee, N.E.; Ahfeldt, T.; Sachs, K.V.; Li, X.; Li, H.; Kuperwasser, N.; Ruda, V.M.; et al. From Noncoding Variant to Phenotype Via Sort1 at the 1p13 Cholesterol Locus. *Nature* **2010**, *466*, 714–719. [CrossRef] [PubMed]
11. Rakyan, V.K.; Down, T.A.; Balding, D.J.; Beck, S. Epigenome-Wide Association Studies for Common Human Diseases. *Nat. Rev. Genet.* **2011**, *12*, 529–541. [CrossRef] [PubMed]
12. Handy, D.E.; Castro, R.; Loscalzo, J. Epigenetic Modifications: Basic Mechanisms and Role in Cardiovascular Disease. *Circulation* **2011**, *123*, 2145–2156. [CrossRef] [PubMed]
13. Jones, P.A. Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond. *Nat. Rev. Genet.* **2012**, *13*, 484–492. [CrossRef] [PubMed]
14. Hedman, A.K.; Mendelson, M.M.; Marioni, R.E.; Gustafsson, S.; Joehanes, R.; Irvin, M.R.; Zhi, D.; Sandling, J.K.; Yao, C.; Liu, C.; et al. Epigenetic Patterns in Blood Associated with Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results from Genome-Wide Association Studies. *Circ. Cardiovasc. Genet.* **2017**, *10*, e001487. [CrossRef] [PubMed]
15. Sayols-Baixeras, S.; Irvin, M.R.; Arnett, D.K.; Elosua, R.; Aslibekyan, S.W. Epigenetics of Lipid Phenotypes. *Curr. Cardiovasc. Risk Rep.* **2016**, *10*, 31. [CrossRef] [PubMed]
16. Braun, K.V.; Voortman, T.; Dhana, K.; Troup, J.; Bramer, W.M.; Troup, J.; Chowdhury, R.; Dehghan, A.; Muka, T.; Franco, O.H. The Role of DNA Methylation in Dyslipidaemia: A Systematic Review. *Prog. Lipid Res.* **2016**, *64*, 78–91. [CrossRef] [PubMed]
17. Pfeiffer, L.; Wahl, S.; Pilling, L.C.; Reischl, E.; Sandling, J.K.; Kunze, S.; Holdt, L.M.; Kretschmer, A.; Schramm, K.; Adamski, J.; et al. DNA Methylation of Lipid-Related Genes Affects Blood Lipid Levels. *Circ. Cardiovasc. Genet.* **2015**, *8*, 334–342. [CrossRef] [PubMed]
18. Braun, K.V.E.; Dhana, K.; de Vries, P.S.; Voortman, T.; van Meurs, J.B.J.; Uitterlinden, A.G.; Hofman, A.; Hu, F.B.; Franco, O.H.; Dehghan, A. Epigenome-Wide Association Study (Ewas) on Lipids: The Rotterdam Study. *Clin. Epigenet.* **2017**, *9*, 15. [CrossRef] [PubMed]
19. Dekkers, K.F.; van Iterson, M.; Slieker, R.C.; Moed, M.H.; Bonder, M.J.; van Galen, M.; Mei, H.; Zhernakova, D.V.; van den Berg, L.H.; Deelen, J.; et al. Blood Lipids Influence DNA Methylation in Circulating Cells. *Genome Biol.* **2016**, *17*, 138. [CrossRef] [PubMed]
20. Burwinkel, B.; Scott, J.W.; Buhrer, C.; van Landeghem, F.K.; Cox, G.F.; Wilson, C.J.; Hardie, D.G.; Kilimann, M.W. Fatal Congenital Heart Glycogenosis Caused by a Recurrent Activating R531q Mutation in the Gamma 2-Subunit of Amp-Activated Protein Kinase (Prkag2), Not by Phosphorylase Kinase Deficiency. *Am. J. Hum. Genet.* **2005**, *76*, 1034–1049. [CrossRef] [PubMed]
21. Hardie, D.G.; Carling, D. The Amp-Activated Protein Kinase–Fuel Gauge of the Mammalian Cell? *Eur. J. Biochem.* **1997**, *246*, 259–273. [CrossRef] [PubMed]
22. Blair, E.; Redwood, C.; Ashrafian, H.; Oliveira, M.; Broxholme, J.; Kerr, B.; Salmon, A.; Ostman-Smith, I.; Watkins, H. Mutations in the Gamma(2) Subunit of Amp-Activated Protein Kinase Cause Familial Hypertrophic Cardiomyopathy: Evidence for the Central Role of Energy Compromise in Disease Pathogenesis. *Hum. Mol. Genet.* **2001**, *10*, 1215–1220. [CrossRef] [PubMed]
23. Arad, M.; Benson, D.W.; Perez-Atayde, A.R.; McKenna, W.J.; Sparks, E.A.; Kanter, R.J.; McGarry, K.; Seidman, J.G.; Seidman, C.E. Constitutively Active Amp Kinase Mutations Cause Glycogen Storage Disease Mimicking Hypertrophic Cardiomyopathy. *J. Clin. Investig.* **2002**, *109*, 357–362. [CrossRef] [PubMed]
24. Arad, M.; Maron, B.J.; Gorham, J.M.; Johnson, W.H., Jr.; Saul, J.P.; Perez-Atayde, A.R.; Spirito, P.; Wright, G.B.; Kanter, R.J.; Seidman, C.E.; et al. Glycogen Storage Diseases Presenting as Hypertrophic Cardiomyopathy. *N. Engl. J. Med.* **2005**, *352*, 362–372. [CrossRef] [PubMed]
25. GmbH, Genomatix. Genome Annotation and Regulation Analysis Tool. Available online: https://www.genomatix.de/ (accessed on 21 February 2019).
26. De Ferrari, G.V.; Papassotiropoulos, A.; Biechele, T.; De-Vrieze, F.W.; Avila, M.E.; Major, M.B.; Myers, A.; Saez, K.; Henriquez, J.P.; Zhao, A.; et al. Common Genetic Variation within the Low-Density Lipoprotein Receptor-Related Protein 6 and Late-Onset Alzheimer's Disease. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9434–9439. [CrossRef] [PubMed]

27. Mani, A.; Radhakrishnan, J.; Wang, H.; Mani, A.; Mani, M.A.; Nelson-Williams, C.; Carew, K.S.; Mane, S.; Najmabadi, H.; Wu, D.; et al. Lrp6 Mutation in a Family with Early Coronary Disease and Metabolic Risk Factors. *Science* **2007**, *315*, 1278–1282. [CrossRef] [PubMed]
28. Sarzani, R.; Salvi, F.; Bordicchia, M.; Guerra, F.; Battistoni, I.; Pagliariccio, G.; Carbonari, L.; Dessi-Fulgheri, P.; Rappelli, A. Carotid Artery Atherosclerosis in Hypertensive Patients with a Functional Ldl Receptor-Related Protein 6 Gene Variant. *Nutr. Metab. Cardiovasc. Dis.* **2011**, *21*, 150–156. [CrossRef] [PubMed]
29. Visvikis-Siest, S.; Siest, G. The Stanislas Cohort: A 10-Year Follow-up of Supposed Healthy Families. Gene-Environment Interactions, Reference Values and Evaluation of Biomarkers in Prevention of Cardiovascular Diseases. *Clin. Chem. Lab. Med.* **2008**, *46*, 733–747. [CrossRef] [PubMed]
30. Siest, G.; Visvikis, S.; Herbeth, B.; Gueguen, R.; Vincent-Viry, M.; Sass, C.; Beaud, B.; Lecomte, E.; Steinmetz, J.; Locuty, J.; et al. Objectives, Design and Recruitment of a Familial and Longitudinal Cohort for Studying Gene-Environment Interactions in the Field of Cardiovascular Risk: The Stanislas Cohort. *Clin. Chem. Lab. Med.* **1998**, *36*, 35–42. [CrossRef] [PubMed]
31. Miller, S.A.; Dykes, D.D.; Polesky, H.F. A Simple Salting out Procedure for Extracting DNA from Human Nucleated Cells. *Nucleic Acids Res.* **1988**, *16*, 1215. [CrossRef] [PubMed]
32. Friedewald, W.T.; Levy, R.I.; Fredrickson, D.S. Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, without Use of the Preparative Ultracentrifuge. *Clin. Chem.* **1972**, *18*, 499–502. [PubMed]
33. Noma, A.; Hata, Y.; Goto, Y. Quantitation of Serum Apolipoprotein a-I, a-Ii, B, C-Ii, C-Iii and E in Healthy Japanese by Turbidimetric Immunoassay: Reference Values, and Age- and Sex-Related Differences. *Clin. Chim. Acta* **1991**, *199*, 147–157. [CrossRef]
34. Bibikova, M.; Barnes, B.; Tsan, C.; Ho, V.; Klotzle, B.; Le, J.M.; Delano, D.; Zhang, L.; Schroth, G.P.; Gunderson, K.L.; et al. High Density DNA Methylation Array with Single Cpg Site Resolution. *Genomics* **2011**, *98*, 288–295. [CrossRef] [PubMed]
35. Aryee, M.J.; Jaffe, A.E.; Corrada-Bravo, H.; Ladd-Acosta, C.; Feinberg, A.P.; Hansen, K.D.; Irizarry, R.A. Minfi: A Flexible and Comprehensive Bioconductor Package for the Analysis of Infinium DNA Methylation Microarrays. *Bioinformatics* **2014**, *30*, 1363–1369. [CrossRef] [PubMed]
36. Chen, Y.; Lemire, M.; Choufani, S.; Butcher, D.T.; Grafodatskaya, D.; Zanke, B.W.; Gallinger, S.; Hudson, T.J.; Weksberg, R. Discovery of Cross-Reactive Probes and Polymorphic Cpgs in the Illumina Infinium Humanmethylation450 Microarray. *Epigenetics* **2013**, *8*, 203–209. [CrossRef] [PubMed]
37. Maksimovic, J.; Gordon, L.; Oshlack, A. Swan: Subset-Quantile within Array Normalization for Illumina Infinium Humanmethylation450 Beadchips. *Genome Biol.* **2012**, *13*, R44. [CrossRef] [PubMed]
38. Grundberg, E.; Meduri, E.; Sandling, J.K.; Hedman, A.K.; Keildson, S.; Buil, A.; Busche, S.; Yuan, W.; Nisbet, J.; Sekowska, M.; et al. Global Analysis of DNA Methylation Variation in Adipose Tissue from Twins Reveals Links to Disease-Associated Variants in Distal Regulatory Elements. *Am. J. Hum. Genet.* **2013**, *93*, 876–890. [CrossRef] [PubMed]
39. MacArthur, J.; Bowler, E.; Cerezo, M.; Gil, L.; Hall, P.; Hastings, E. The New Nhgri-Ebi Catalog of Published Genome-Wide Association Studies (Gwas Catalog). *Nucleic Acids Res.* **2017**, *45*, D896–D901. [CrossRef] [PubMed]
40. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [CrossRef]
41. Barfield, R.T.; Kilaru, V.; Smith, A.K.; Conneely, K.N. Cpgassoc: An R Function for Analysis of DNA Methylation Microarray Data. *Bioinformatics* **2012**, *28*, 1280–1281. [CrossRef] [PubMed]
42. Kuznetsova, A.; Brockhoff, P.; Christensen, R.H.B. Lmertest: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* **2017**, *2*. [CrossRef]
43. Zerbino, D.R.; Achuthan, P.; Akanni, W.; Amode, M.R.; Barrell, D.; Bhai, J.; Billis, K.; Cummins, C.; Gall, A.; Giron, C.G.; et al. Ensembl 2018. *Nucleic Acids Res.* **2018**, *46*, D754–D761. [CrossRef] [PubMed]
44. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The Human Genome Browser at Ucsc. *Genome Res.* **2002**, *12*, 996–1006. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Publication Nº 3

Epigenome-Wide Association Study Detects a Novel Loci Associated with Central Obesity in Healthy Subjects.

Ting Xie[*], **Vesna Gorenjak[*]**, Maria G. Stathopoulou[*], Sébastien Dadé, Eirini Marouli, Christine Masson, Helena Murray, John Lamont, Peter Fitzgerald, Panagiotis Deloukas and Sophie Visvikis-Siest

[*]equal contribution

Under submission

Obesity is a medical condition characterised by an excess amount of body fat, associated with a higher risk for a number of chronic diseases, including diabetes, CVD and cancer (291). BMI is the most widely used method for the definition of obesity, whereas indices of abdominal obesity are believed to be better discriminators of obesity-related health risks (292, 293). In most cases, obesity is a result of increased intake of hypercaloric food in combination with limited sports activity. Besides these environmental risk factors, an important genetic variability in susceptibility for obesity has been demonstrated in genetic studies, which identified a large number of common variants related to adiposity levels (294).

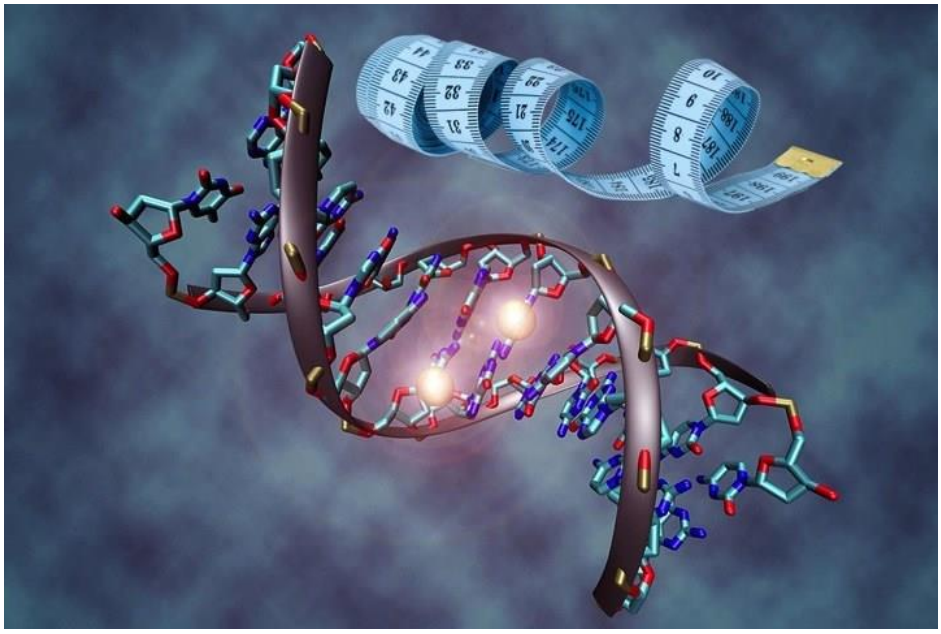


Figure 27: Methylated DNA - a link between environment and genetics (34).

In addition, an important role in the development and monitoring of obesity has been assigned to epigenetic mechanisms that are modulating the genetic expression and can thus affect a particular phenotype (Figure 27). Epigenetic markers that increase susceptibility for obesity can be inherited, as well as they can be modifiable and appear during the life course as a result of obesogenic environmental conditions (295).

Objective

The goal of this study was to perform an EWAS on waist circumference (WC) in a healthy population, to get an insight into epigenetic regulation of central obesity with detection of CpG methylation sites associated with WC.

Populations and analysis

Healthy related individuals of European-Caucasian origin from 73 families (n=211) from the SFS have been enrolled in this study, including 116 adults and 95 children. Blood DNA methylation assessment was performed by Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA). The R package minfi was used for the analysis of the methylation data and R package CpGassoc was used for the association analyses of methylation values and WC in a linear mixed-effects model adjusted for sex, age, BMI, family structure and individual blood cell counts (neutrophils, lymphocytes, monocytes, eosinophils and basophils) as covariates, and chip array as random effect. The relation between methylation and genetic variants associated with WC was investigated in a linear mixed-effects model using the R package nlme.

Genotyping was performed by the Infinium CoreExome Illumina assay. SNPs located in the same chromosome as the significant methylation site were selected from the NHGRI-EBI GWAS catalogue.

Results

The results of the EWAS (Table 4) showed a novel significant positive association of the probe cg16170243 with WC, adjusted for BMI (β=2.32, SE=0.41; P_{adj}=0.048).

Table 4. Association of methylation site with WC level, adjusted for BMI.

CpG	Gene Name	Chromosome	Effect size	SE	P-Value	FDR
cg16170243	AC090241.2	18q21.1	2.32	0.41	0.13×10^{-8}	0.048

SE: Standard error

Bioinformatics analysis showed that the probe cg16170243 corresponds to a 50 bp-length human methylation oligoprobe located on the chromosome 18q21.1, in the forward strand of the *ST8SIA5* (Alpha-2,8-sialyltransferase 8E) antisense gene (*AC090241.2*). On the opposite strand, the probe cg16170243 is located 245 base pairs upstream of the *ST8SIA5* gene, in the promoter flank, where it overlaps eight transcription factors binding sites (TFBS).

SNPs previously related to WC levels were tested for their associations with methylation. Altogether, three SNPs located on chromosome 18 were identified. No association was identified between these three polymorphisms and methylation at cg16170243, indicating that the association between CpG and WC does not originate from the relation between SNPs and methylated CpG site.

Conclusion

The study presented in the publication '*Epigenome-Wide Association Study Detects a Novel Loci Associated with Central Obesity in Healthy Subjects'* identified a novel CpG site associated with increased WC and gave new insights into epigenetic regulation of obesity. These findings improved the knowledge of epigenetic regulation of obesity and might be useful in a future determination of obesity risk factors.

A limitation of our study is combined population, small sample size and the lack of tissue-specific replication. Further analysis is warranted to overcome these issues.

Epigenome-Wide Association Study Detects a Novel Loci Associated with Central Obesity in Healthy Subjects

Xie T.[a,*], Gorenjak V.[a,*], Stathopoulou MG.[a,*], Dadé S.[a], Marouli E.[b], Masson C.[a], Murray H.[c], Lamont J.[c], Fitzgerald P.[c], Deloukas P.[b], Visvikis-Siest S.[a,d]

a. Université de Lorraine, Inserm, IGE-PCV, F-54000 Nancy, France
b. Randox Laboratories Ltd, Crumlin, UK
c. Queen Mary University of London, London, UK
d. Department of Internal Medicine and Geriatrics, CHU Nancy-Brabois, Nancy, France
** Equal first authors*

Abstract

Background and Aims:
Central obesity is a condition that poses a significant risk to global health and requires employment of novel scientific methods for exploration. The objective of this study is to use DNA methylation analysis to detect DNA methylation loci linked to obesity phenotypes.
Methods and Results:
Two-hundred and eleven healthy European participants from the STANISLAS Family Study (SFS), comprising 73 nuclear families, were comprehensively assessed for methylation status using Illumina Infinium HumanMethylation450 BeadChip. An epigenome-wide association study was performed, which identified a CpG site cg16170243 located on chromosome 18q21.2 significantly associated with waist circumference, after adjusting for BMI (β=2.32, SE=0.41, P_{adj}=0.048). No significant association was observed with waist-to-hip ratio adjusted for BMI (P_{adj}>0.05).
Conclusions:
CpG site cg16170243 corresponds to a 50 bp-length human methylation oligoprobe located within the *AC090241.2* gene, which exists in three transcription isoforms. The isoform 202 overlaps *ST8SIA5* gene, which codes for a sialyltransferase enzyme, involved in synthesis of gangliosides and contributes to the regulation of transmembrane proteins, such as insulin, epidermal growth factor and vascular endothelial growth factor receptors. Consequently, methylation on CpG site cg16170243 may result in insulin resistance, lipid accumulation and could account for increased susceptibility for central obesity.

Keywords: Central obesity, Methylation, EWAS, Epigenetics

1. INTRODUCTION

Epigenetic changes are covalent modifications of cytosine bases, histones and changes in nucleosome positioning that can modulate the effect of a genotype on particular phenotype and thus affect physiological mechanisms as well as the pathophysiology of many diseases (1). The most widely studied among them is cytosine methylation (2), a covalent attachment of methyl group to a DNA sequence that generally results in silencing of genes encoded in the methylated region (3). DNA methyltransferases are enzymes that enable this transformation, which normally takes place at the carbon-5 of cytosine ring in CpG dinucleotides. DNA methylation patterns are not static but undergo precise, highly coordinated changes that can be mediated both by environmental and genetic factors and inherited through mitotic cell divisions (4). This process occurs already during embryogenesis and is crucial for development, differentiation and cellular variability (3), as well as for the transcriptional regulation of genes and miRNA (5). On the other hand, atypical patterns of DNA methylation are associated with obesity, oxidative stress, hypertension, inflammation, angiogenesis and other pathological processes that are implicated in the development of chronic diseases (6).

Obesity is a systemic disease and a growing health problem that contributes to the increased risk of many common medical conditions (7). The excess of adipose tissue provokes chronic low-grade inflammation, associated with immunological activation and oxidative stress, insulin resistance, hypertension and dyslipidemia (8). This partly explains the association with many comorbidities, especially diabetes mellitus (type 2), cardiovascular disease (CVD) and cancer, which account for the greatest mortality risk linked with obesity (7). The most common approach to determine general obesity is body mass index (BMI), the ratio between the mass (weight) and height of an individual, expressed in kg/m^2. However, clinical evidence suggests that diseases, such as diabetes, are more associated with central obesity, where visceral adipose tissue is stored (9). Therefore, waist circumference (WC) and waist-to-hip ratio are

believed to better explain obesity-related health risk (10).

Increasing prevalence of obesity worldwide is mostly due to changes in the environment, whereas a person′s genetic profile is considered one of the main causes of individual difference in predisposition to weight gain. A high heritability of this phenotype has been confirmed in several epidemiological studies (11) and examined in GWAS, where 27-30 % of the total BMI variance was explained by common single nucleotide polymorphisms (SNPs) (12).

Furthermore, high nutritional input in early life was associated with an increased risk of a range of chronic diseases in adulthood, showing that early environmental influences can cause permanent effects (13). This interaction of biological and environmental factors is believed to be mediated by epigenetic mechanisms, by which the environmental factors could change gene expression and thus explain the increased prevalence of obesity in the last few decades (12-14).

Various approaches were therefore developed to study the associations between epigenetic mechanisms and particular phenotypes. The most recent among them, genome-wide analyses of DNA methylation, gave the possibility of in-depth insight into epigenetic changes (15). Several epigenome-wide association studies (EWAS) were performed to study DNA methylation to identify the common variation in the DNA methylome, related to obesity phenotypes.

One of the first large EWAS for obesity phenotypes identified an association between increased BMI and methylation at the *HIF3A* locus (involved in a variety of cellular and physiological responses to reduced oxygen concentrations), using large set of whole blood and adipose tissue samples from individuals of European ancestry (16). The study was followed by EWAS of BMI and WC as obesity phenotypes, performed in a cohort of African American adults (8) and in healthy adults of European American descent (17). In both studies, the epigenome-wide significant associations between the methylation status and both, BMI and WC, were confirmed and replicated. Results included methylation on *HIF3A, ABCG1* (8) and *CPT1A* loci (8, 17).

More recent was the EWAS of DNA methylation loci involved in adiposity, which was performed on sub-Saharan African population and identified three epigenome-wide significant loci associated with BMI and WC, including previously identified *CPT1A* (18). The REGICOR study replicated 16 CpGs previously reported and newly identified and validated 33 CpGs related to WC (19). Finally, large EWAS on obesity-related traits was conducted (n= 1450) and replicated (n=2097) in 2018, which identified novel differentially methylated genes for obesity-related traits and confirmed some of the findings of previous studies (20).

For successful prevention, prediction and treatment of chronic diseases, it is essential to obtain the fundamental knowledge of their regulatory mechanisms. Epigenetic studies have improved our understanding of obesity footprints, involved in changing of the process of gene regulation. Even though a lot of effort has been made to find epigenetic markers related to central obesity, new cohorts are continuously providing novel discoveries. In order to make further contribution to the comprehension of this disease trait, in this study, we are examining the methylation levels associated with central obesity, measured by WC and waist-to-hip ratio adjusted for BMI, in a healthy population. Using healthy individuals can help to avoid the discrepancies caused by extremes of obesity and comorbidity in population-specific cohorts. Our findings provided new insights into genetic regulation of visceral fat accumulation and are presenting new variant, which may increase the susceptibility to chronic diseases.

2. METHODS

2.1. Population

211 healthy related individuals from 73 families of the STANISLAS Family Study (SFS) have been enrolled in this study, including 116 adults and 95 children. The SFS is a 10-year longitudinal survey with three visits at 5-year intervals, involving 1,006 families from Vandoeuvre-ls-Nancy, France, first recruited between 1993-1995 (21). All subjects were of European-Caucasian origin, without the presence of chronic disorders (CVD or cancer). Descriptive characteristics are presented in Table 1.

The study protocols were approved by the institutional ethics committees and all subjects gave written informed consent for their participation in the study.

2.2. Data collection

Biological and clinical measurements for the SFS were collected using appropriate, validated procedures. All measurements of the clinical indicators were performed in the laboratory of the Centre for Preventive Medicine (CMP) in Vandoeuvre ls Nancy, France. WC was taken at the midpoint between the lower margin of the last palpable rib and the top of the iliac crest (hip bone). Trained professionals recorded measurements to the nearest 0.1 cm. Body mass index (BMI) was calculated by the Quetelet index formula as weight (kilograms) divided by height^2 (meters)^2. Data collection has been previously described (21, 22).

2.3. DNA methylation assay and quality control

Blood samples were taken between 8 and 9 a.m. after overnight fasting. Whole blood DNA was extracted by the Miller technique (23) and was stored at −80 °C. Genome-wide DNA methylation profiling was performed using the Infinium HumanMethylation450 BeadChip (Illumina). Methylation ratio, referred to as beta value by Illumina′s software, is the proportion methylated/(methylated + unmethylated) CpGs. Methylation arrays were analyzed and visualized using the R package minfi(version 1.16.1) (24). Detection p-value was generated for every CpG in all samples, indicating the quality of the signal. Poor quality probes were excluded from the analysis using a detection p-value cutoff (>0.05). Probes, missing in more than 5 % of samples were excluded from all samples. Background correction and normalisation were performed with Illumina background correction and SWAN (25) to all intensity values for a total of 485 512 probes. To avoid spurious

Table 1: Population characteristics

	Total (n=211)		Adults (n=116)		Children (n=95)		Males (n=105)		Females (n=106)	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Ages (years)	28.17	14.83	40.84	7.53	13.15	2.58	27.09	15.14	29.24	14.37
BMI (kg/m^2)	21.52	4.0	24.06	3.22	18.43	2.31	21.77	4.17	21.27	3.79
WC (cm)	72.92	11.82	80.11	10.28	64.15	6.29	76.62	13.08	69.26	8.96
HC (cm)	90.38	11.12	97.13	6.27	82.14	10.09	89.38	11.76	91.37	10.3
Waist-to-hip (cm)	0.81	0.08	0.82	0.08	0.79	0.06	0.86	0.07	0.79	0.05

associations, the probes already annotated in HumanMethlyation450 annotation files (probes containing SNPs, sex chromosomes, and a single base extension (SBE) sites) were excluded. Finally, the probes containing cross-reactive and target polymorphic CpGs (21) were excluded. All downstream analyses were carried using R programme.

2.4. Genotyping and selection of SNPs

Genotyping was performed by the Infinium CoreExome Illumina assay. Significant WC associated SNPs, located in the same chromosome as the significant methylation sites, were selected from the NHGRI-EBI GWAS catalogue.

2.5. Statistical analysis

Waist-to-hip ratio and WC were not normally distributed and were therefore transformed to the e-log scale. A linear mixed-effects model was used to analyse the association between methylation levels at each probe and log-transformed phenotypes. The model used included sex, age, BMI, family structure, and individual blood cell counts (neutrophils, lymphocytes, monocytes, eosinophils and basophils) as covariates and chip array as random effect. Bonferroni correction and false discovery rate (FDR) methods (>0.05) were used for the correction of the results for multiple testing. The association analyses of methylation values and the assessed phenotypes were performed by using the package CpGassoc in R (26). HumanMethlyation450 annotation file was used for annotating the probes and their corresponding genes. Deviation of Multiple Correlation Squared ρ^2 From Constant (Random Model) of G*Power software was used in order to calculate statistical power (27), specifically the *post hoc* power analysis procedure. This procedure is parameter analysis which requires the type of test (tails: one or two), effect size (H1 ρ^2 and H0 ρ^2), α error probability, total sample size, and a number of predictors. In order to investigate whether the observed association between WC and methylation level was related to genetic variants associated with WC, a linear mixed-effects model with age, sex, BMI, family structure, methylation batch array as fixed effects, and methylation chip as a random effect was carried out for association analysis using the R statistical package nlme (28).

2.6. In silico analysis

Cg16170243 was localised on the Human genome (GRCh38.p12) using the Ensembl browser. *In silico ST8SIA5* gene expression was obtained using the BLUEPRINT Data Analysis portal (29) and GTEx Portal (30).

3. RESULTS

Children included in the study had significantly lower WC (64.15 cm, SD=6.29) and BMI (18.43 kg/m^2, SD=2.31) than adult individuals (80.11 cm, SD = 10.28 and 24.06 kg/m^2, SD = 3.22, respectively). Mean age was 13.15 years for children and 40.48 years for adults (Table 1). 764 poor quality probes were excluded from the sample. In total, 77 % of probes passed quality control, excluding probes containing SNPs, sex chromosomes, SBE sites, cross-reactive and target polymorphic CpGs, thus leaving 373 626 probes for association analyses (Figure 1). The *post hoc* analysis of the statistical power of the result was calculated as 100 %.

One individual was excluded after quality control checks of the methylation array data (outlier of the plotted median of the methylated against unmethylated samples), thus 210 participants were included in the analyses. The results of the analysis showed a novel significant positive association of cg16170243 probe with WC adjusted for BMI (β=2.32, SE=0.41; P_{adj}=0.048) in the combined population. QQ plot of the genome-wide DNA methylation analysis is presented in Figure 2. CpG site cg16170243 (chr18:46759502-46759551) corresponds to a 50 bp-length human methylation oligoprobe located on chromosome 18q21.1 (Table 2). No significant associations were identified for waist-to-hip ratio.

Table 2: Association of methylation site with waist circumference level, adjusted for BMI

	Children	**Adults**	**All**
Effect size	0.01	0.59	2.32
SE	0.009	0.46	0.41
P-Value	0.01	0.041	0.13×10^{-3}
FDR	1	0.95	0.048

Three SNPs, previously associated with WC in GWAS (31) were identified on chromosome 18 (rs6567160, rs7239883, rs12970134). No significant associations were identified between these polymorphisms and methylation of the cg16170243.

The cg16170243 probe maps on chromosome 18q21.1 and is located in the forward strand of the *ST8SIA5* antisense gene (*AC090241.2*), whereas in the opposite strand the probe is located 245 base pairs upstream of the *ST8SIA5* gene (Figure 5). Cg16170243 is located both within an intron of *AC090241.2-202* and 173 base pairs upstream of *AC090241.2-203* (Figure 5). *In silico* analysis showed that cg16170243 is embedded

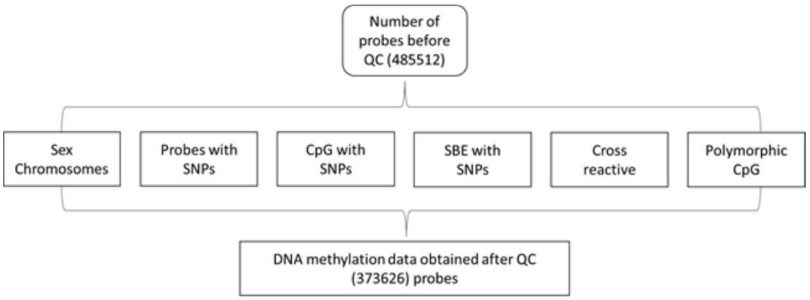


Figure 1: Quality control (QC) processing.

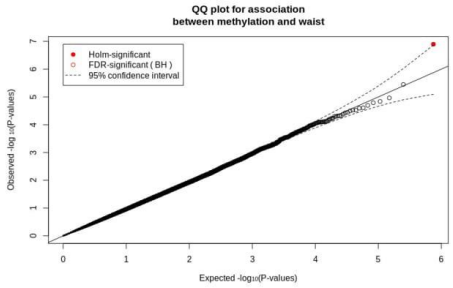


Figure 2: Quantile-quantile plot (QQ plot) of the distribution of observed log10 association P-Values against the expected null distribution.

in the promoter flank of *ST8SIA5* gene (ENSR00000575029) and contains transcription factor binding sites (TFBS) (Figure 5).

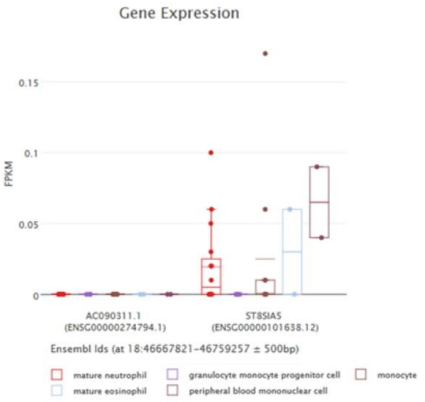


Figure 3: *In silico* expression of *ST8SIA5* gene and *AC090311.1* pseudogene (negative control).

Analysis of the expression data showed that *ST8SIA5* gene is expressed in several cell-types of the bloodline: neutrophils, peripheral blood mononuclear cells, eosinophils and monocytes (Figure 3). Furthermore, the additional data confirmed the expression of *ST8SIA5* in blood cells and indicated that the gene is to a greater extent expressed also in brain structures.

4. DISCUSSION

We have identified a novel significant positive association (β=2.32, SE=0.41; P_{adj}=0.048) between WC, adjusted for BMI, and methylation at the probe cg16170243. There was not association of cg16170243 with three candidate SNPs previously associated with WC. Therefore, the underlying mechanism of the CpG′s effect on WC is probably differentiated from the mechanisms linked to these genetic variants.

The *ST8SIA5* antisense gene has 3 splice variants: *AC090241.2-201*, *AC090241.2-202* and *AC090241.2-203*. Probe cg16170243 is located both in the intronic zone of *AC090241.2-202* variant and at the 5′ of *AC090241.2-203* variant (Figure 5). *AC090241.2-202* splice variant is an antisense biotype transcript, meaning that it overlaps the genomic span of a protein-coding locus on the opposite strand and can be an important regulator of gene expression (32). The overlapping gene on the opposite strand of *AC090241.2-202* is *ST8SIA5* (Alpha-2,8-sialyltransferase 8E) gene, coding for a sialyltransferase enzyme, involved in the synthesis of gangliosides GD1c, GT1a and GQ1b from GM1b, GD1a and GT1b, respectively (33). Gangliosides are glycosphingolipids with one or more sialic acids. They are a component of the bilayer lipid membrane on the cell surface, where they present the points of recognition for extracellular molecules on surfaces of neighbouring cells and serve for interaction between cells, adhesion, cell differentiation and transduction of signal (34).

Within the promoter flank of *ST8SIA5* gene, cg16170243 overlaps 8 TFBSs (Figure 5). Methylation of the CpG site could disrupt the transcription factor binding sites and thus affect the expression level of the *ST8SIA5* gene.

In view of the above, we could assume that the cg16170243 methylation can modify expression levels of *ST8SIA5* in two different ways: directly, by disrupting gene expression through

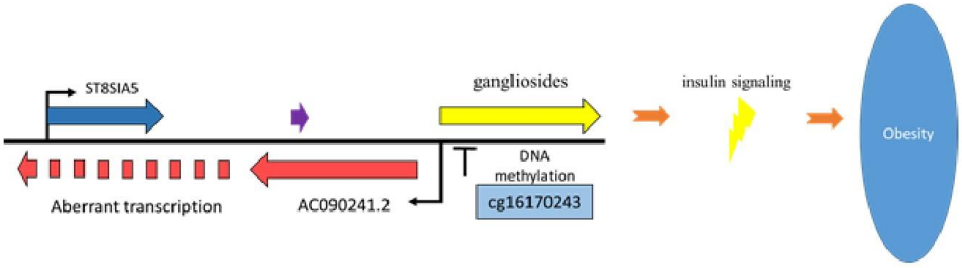


Figure 4: Possible mechanism explaining the relation between cg16170243 and visceral obesity.

methylation of TFBS, or indirectly, by acting on the expression of *AC090241.2* antisense gene. In both cases, this would lead to affected synthesis of gangliosides.

Expression data showed that *ST8SIA5* gene is active in several blood cells (Figure 3), such as neutrophil (in red), monocyte (brown), eosinophil (light blue), and PBMC (purple) [35] and also, to a greater extent, in many brain structures. Interestingly, studies showed that an inadequate ganglioside expression in mediobasal hypothalamic neurons deregulates neuronal leptin (35) and insulin signalling (36), which can affect body weight regulation and energy homeostasis. Gangliosides interact with molecules of signal transduction pathways, such as receptors tyrosine kinases (RTKs). Change in ganglioside composition induces the dissociation of RTKs from glycolipid-enriched microdomains, which results in reduced phosphorylation of the receptors and thus causes the insulin resistance (37). Accordingly, insulin is a critical regulator of adipocyte biology and resistance of insulin receptors is, on the one hand, one of the important causes of obesity, and on the other hand, one of the biggest contributors to the development of obesity (38).

It is plausible that the effect of cg16170243 on the *ST8SIA5* gene in blood cells reflects a process that also occurs on a larger scale in neuronal cells when methylation at cg16170243 site is present, causing a disruption of the insulin signalling pathway and contributing to accumulation of visceral fat (Figure 4). Further studies that would confirm such hypothesis are required.

Despite the above-proposed mechanisms of the methylation impact on the WC *via* insulin deregulation, we cannot exclude the possibility that methylation could be a consequence of modified WC, rather than a cause. Indeed, a meta-analysis from S. Wahl *et al.* has shown that variation in DNA methylation is most often a consequence of adiposity (39). Thus, molecular mechanisms involving the adipose tissue of visceral fat could enable the methylation of regulatory regions of genes involved in obesity, but exact mechanisms of this regulation remain to be elucidated.

Epigenetics studies of obese phenotypes on healthy individuals have been previously performed (40), however, this is in our knowledge the only EWAS study including only healthy subjects. The *post hoc* power analysis has shown that the statistical power of our result was very high (100 %). It should be mentioned, however, that analysis on separate children and adult populations was also performed and no significant associations were identified, probably due to decreased power of the smaller samples of the separate analyses. Similarly, we have performed separate analysis by sex and again no significant associations were identified.

A limitation of our study is a small sample size. Although we have identified a significant result with high statistical power we did not replicate it in a population-specific study. As epigenetic changes can be tissue-specific, the limitation of our study is also the use of blood samples without tissue-specific replications. Even though the within-subject correlation of CpG specific sites from blood and adipose tissue was previously confirmed, the use of methylation markers in blood to mirror the corresponding profile in the target tissue should be taken with caution (41).

In summary, we have identified a novel association between DNA methylation and WC. This association could be due to the modification of the regulatory region of *ST8SIA5* transcription, resulting in a perturbed synthesis of gangliosides. Eventually, methylation of the CpG site could be the consequence of increased WC, caused by the mechanisms that remain to be elucidated.

5. REFERENCES

1. Dupont C, Armant DR, Brenner CA. Epigenetics: definition, mechanisms and clinical perspective. Semin Reprod Med. 2009; 27: 351-7.
2. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010; 28: 1057-68.
3. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics. 2012; 13: 484-92.
4. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet. 2012; 13: 97-109.
5. Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. Human molecular genetics. 2007; 16 Spec No 1: R50-9.

6. Levy E, Spahis S, Bigras JL, Delvin E, Borys JM. The Epigenetic Machinery in Vascular Dysfunction and Hypertension. Curr Hypertens Rep. 2017; 19: 52.
7. Abdelaal M, le Roux CW, Docherty NG. Morbidity and mortality associated with obesity. Ann Transl Med. 2017; 5: 161.
8. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum Mol Genet. 2015; 24: 4464-79.
9. Vazquez G, Duval S, Jacobs DR, Jr., Silventoinen K. Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. Epidemiol Rev. 2007; 29: 115-28.
10. Janssen I, Katzmarzyk PT, Ross R. Waist circumference and not body mass index explains obesity-related health risk. Am J Clin Nutr. 2004; 79: 379-84.
11. Albuquerque D, Stice E, Rodriguez-Lopez R, Manco L, Nobrega C. Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective. Mol Genet Genomics. 2015; 290: 1191-221.
12. Pigeyre M, Yazdi FT, Kaur Y, Meyre D. Recent progress in genetics, epigenetics and metagenomics unveils the pathophysiology of human obesity. Clin Sci (Lond). 2016; 130: 943-86.
13. Bell CG. The Epigenomic Analysis of Human Obesity. Obesity (Silver Spring). 2017; 25: 1471-81.
14. Cordero P, Li J, Oben JA. Epigenetics of obesity: beyond the genome sequence. Curr Opin Clin Nutr Metab Care. 2015; 18: 361-6.
15. Flanagan JM. Epigenome-wide association studies (EWAS): past, present, and future. Methods Mol Biol. 2015; 1238: 51-63.
16. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. Lancet. 2014; 383: 1990-8.
17. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. Obesity (Silver Spring). 2015; 23: 1493-501.
18. Meeks KAC, Henneman P, Venema A, Burr T, Galbete C, Danquah I, et al. An epigenome-wide association study in whole blood of measures of adiposity among Ghanaians: the RODAM study. Clin Epigenetics. 2017; 9: 103.
19. Sayols-Baixeras S, Subirana I, Fernndez-Sanls A, Sent M, Llus-Ganella C, Marrugat J, et al. DNA methylation and obesity traits: An epigenome-wide association study. The REGICOR study. Epigenetics. 2017; 12: 909-16.
20. Dhana K, Braun KVE, Nano J, Voortman T, Demerath EW, Guan W, et al. An Epigenome-Wide Association Study (EWAS) of Obesity-Related Traits. Am J Epidemiol. 2018.
21. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med. 2008; 46: 733-47.
22. Siest G, Visvikis S, Herbeth B, Gueguen R, Vincent-Viry M, Sass C, et al. Objectives, design and recruitment of a familial and longitudinal cohort for studying gene-environment interactions in the field of cardiovascular risk: the Stanislas cohort. Clin Chem Lab Med. 1998; 36: 35-42.
23. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988; 16: 1215.
24. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014; 30: 1363-9.
25. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome biology. 2012; 13: R44.
26. Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. Bioinformatics. 2012; 28: 1280-1.
27. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behavior research methods. 2009; 41: 1149-60.
28. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. 2017.
29. Fernandez JM, de la Torre V, Richardson D, Royo R, Puiggros M, Moncunill V, et al. The BLUEPRINT Data Analysis Portal. Cell systems. 2016; 3: 491-5.e5.
30. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348: 648-60.
31. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). 2017; 45: D896-d901.
32. Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nature reviews Genetics. 2013; 14: 880-93.
33. Kim YJ, Kim KS, Do S, Kim CH, Kim SK, Lee YC. Molecular cloning and expression of human alpha2,8-sialyltransferase (hST8Sia V). Biochemical and biophysical research communications. 1997; 235: 327-30.
34. Yu RK, Tsai YT, Ariga T, Yanagisawa M. Structures, biosynthesis, and functions of gangliosides–an overview. J Oleo Sci. 2011; 60: 537-44.
35. Nordstrom V, Willershauser M, Herzer S, Rozman J, von Bohlen Und Halbach O, Meldner S, et al. Neuronal expression of glucosylceramide synthase in central nervous system regulates body weight and energy homeostasis. PLoS Biol. 2013; 11: e1001506.
36. Herzer S, Meldner S, Grone HJ, Nordstrom V. Fasting-Induced Lipolysis and Hypothalamic Insulin Signaling Are Regulated by Neuronal Glucosylceramide Synthase. Diabetes. 2015; 64: 3363-76.

37. Julien S, Bobowski M, Steenackers A, Le Bourhis X, Delannoy P. How Do Gangliosides Regulate RTKs Signaling? Cells. 2013; 2: 751-67.
38. Kahn BB, Flier JS. Obesity and insulin resistance. The Journal of clinical investigation. 2000; 106: 473-81.
39. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2016; 541: 81.
40. Boyne DJ, Friedenreich CM, McIntyre JB, Courneya KS, King WD. Associations between adiposity and repetitive element DNA methylation in healthy postmenopausal women. Epigenomics. 2017; 9: 1267-77.
41. Huang Y-T, Chu S, Loucks EB, Lin C-L, Eaton CB, Buka SL, et al. Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood. Epigenetics. 2016; 11: 227-36.

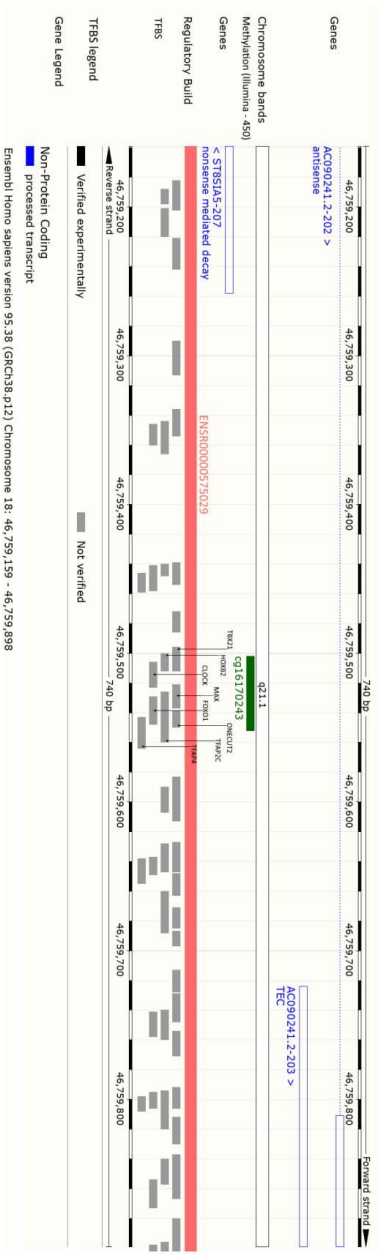


Figure 5: Environment of cg16170243 probe.

Publication Nº 4

Epigenome-wide association study in healthy individuals identifies significant associations with DNA methylation and PBMCs extracts VEGF-A concentration.

Vesna Gorenjak*, Dwaine R. Vance*, Sébastien Dadé, Maria G. Stathopoulou, Lauren Doherty, Helena Murray, Christine Masson, John Lamont, Peter Fitzgerald, Sophie Visvikis-Siest

Under submission

VEGF-A (Figure 29) is a mitogen that acts on endothelial cells and has various effects, including mediating increased vascular permeability, angiogenesis and endothelial cell growth, promoting cell migration, and inhibiting apoptosis (296). VEGF-A is involved in the pathogenesis of CVD (185), as well as other chronic diseases, such as cancer, type 2 diabetes, osteoporosis, osteoarthritis and COPD. It has been previously shown that blood levels of VEGF-A are highly heritable, with 60.5 % of variance attributable to genetic factors in the SFS (44). Ten polymorphisms were identified that explained more than 50% of the heritability of VEGF-A (45, 46) through GWAS (n > 16,000). Despite the well-established studies on VEGF-A genetic determinants, there is only little knowledge of VEGF-A epigenetic regulation. Knowing and understanding the mechanisms of epigenetic regulation of VEGF-A might allow further comprehension of the epigenetic regulation involved in chronic diseases, where VEGF-A plays a crucial role.

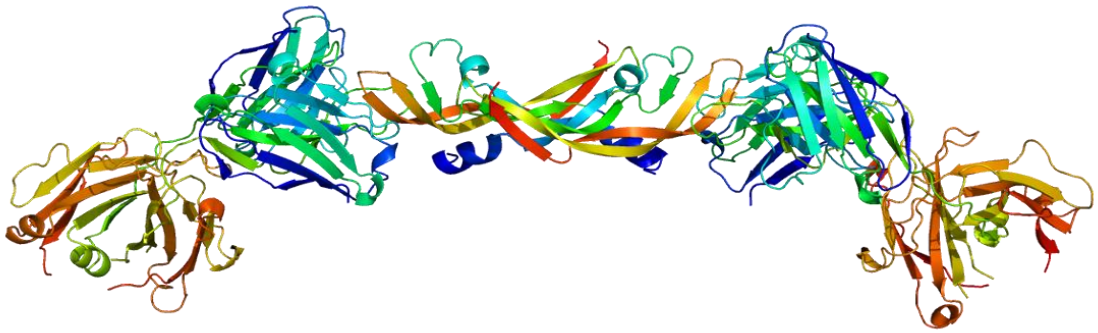


Figure 28: Structure of the VEGF-A protein: VEGF-A forms an antiparallel homodimer, with symmetrical binding sites for kinase domain receptors at each pole of the protein (297, 298).

Objective

The goal of this study was to perform an EWAS on VEGF-A levels in a healthy population, in order to get an insight into epigenetic regulation of VEGF-A levels from the PBMCs extracts with detection of CpG methylation sites associated with VEGF-A levels.

Populations and analysis

VEGF-A levels of 201 healthy participants from the SFS were tested for association with DNA methylation. DNA was extracted by the Miller technique (281) and was stored at – 80 °C. Whole-genome DNA methylation was investigated with Illumina HumanMethylation450 bead array (Illumina Inc., San Diego, CA, USA). Minfi R package (Bioconductor) was used to analyse and visualise Illumina Infinium methylation arrays, background correction, colour bias adjustment and Infinium I/II bias correction were carried out with Illumina background correction and SWAN. Finally, CpGassoc (CRAN) was used to test for association between methylation at CpG sites across the genome and VEGF-A concentration in PBMC extracts. The random mixed-effects model included gender, age, BMI, family structure and individual blood cell counts (neutrophils, lymphocytes, monocytes, eosinophils and basophils) as covariates and chip array as random effect.

VEGF-A was measured using the Randox high sensitivity multiplex cytokine and growth factor array (Evidence Investigator Analyzer, Randox Laboratories Ltd., Crumlin, United Kingdom). Measurements were performed in the protein extracts from PBMCs, isolated from the whole blood.

Bioinformatics analysis of results was performed using the Ensembl browser. Common biological pathways of annotated genes with VEGF-A were analysed using GeneMANIA app.

Results

EWAS identified 20 CpG sites significant after Holm-Bonferroni adjustment ($<1.6x10^{-7}$). Results are presented with the Manhattan plot (Figure 28), displaying adjusted P-values of the association between methylation probes and VEGF-A levels in PBMCs extracts.

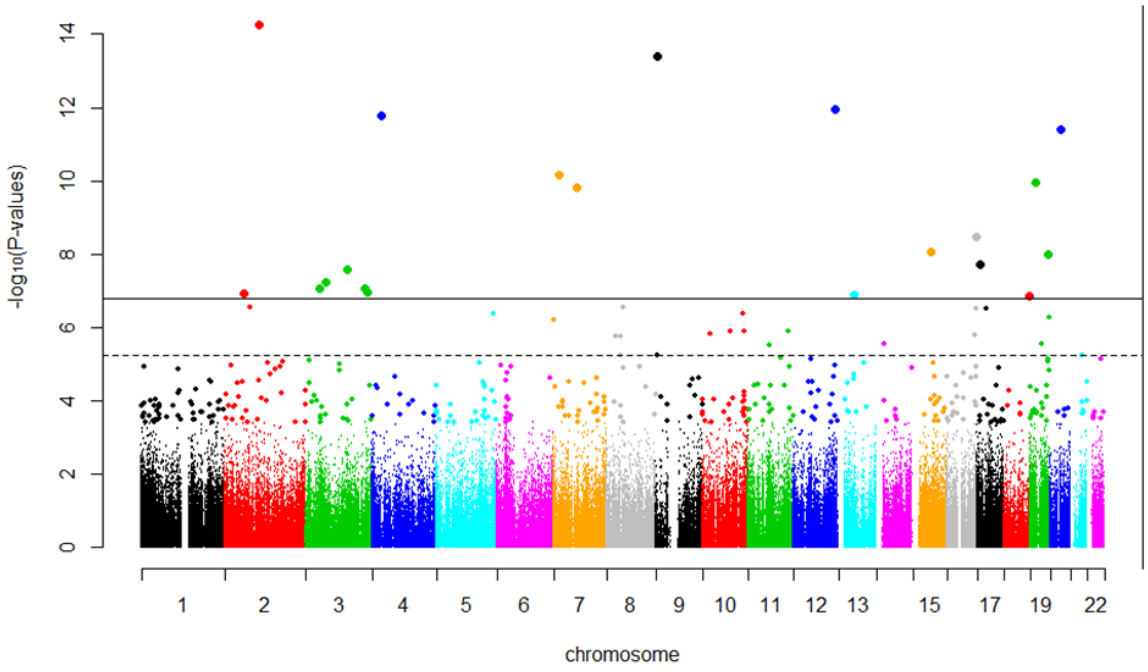


Figure 29: Manhattan plot of significant CpG associations with VEGF-A levels.

CpG sites were located on 13 different chromosomes and were found in the vicinity of different genes involved in diverse regulative processes. None of the CpG sites was detected on a gene with a direct implication in VEGF-A biology.

The annotation program did not identify the nearest gene for all significant CpGs. Therefore, we have conducted an *in silico* analysis to define the genetic environment in the vicinity of those CpG sites and detect the nearest regulatory features on which these methylation sites could act (nearest gene defined after bioinformatics analysis for each CpG site is listed in the legend of Table 5).

For all 20 CpG sites, we have performed detailed bioinformatics and bibliographical analysis to reveal possible genetic pathways common between VEGF-A and concerned genes. We have identified CpG sites, which have been previously implicated in VEGF-related biological processes, such as cell proliferation, cell growth, angiogenesis and related diseases. Such relations were found for *TPX2* gene, significantly associated with cg21838233 ($P= 4.01 \times 10^{-12}$) and *HAS1* gene found in the vicinity of cg06785213 ($P= 1.06 \times 10^{-08}$), as well as with *ARL4A*, *ETV1*, *CTNNB1*, *TBL1XR1* and *TSHZ1*.

Table 5. Novel associations between DNA methylation and VEGF-A PBMCs levels.

CpG	Gene Name	Chromosome	Effect size	P-Value
cg05739757	*RPL31*	2q11.2	0.00176	5.64×10^{-15}
cg23333878	*GLIS3*	9p24.2	0.00270	3.86×10^{-14}
cg21838233	*TPX2*	20q11.21	0.00191	4.01×10^{-12}
cg18815539	*SEPSECS*	4p15.2	-0.00197	1.70×10^{-12}
cg21968169	*LOC338799*	12q24.31	-0.00189	1.11×10^{-12}
cg16333561	*ND1*	7q11.23	-0.00130	7.02×10^{-11}
cg20547575	*AUTS2*	7q11.22	-0.00349	1.52×10^{-10}
cg15014826	*ND2*	19p13.12	-0.00157	1.12×10^{-10}
cg00117600	*PIGB*	15q21.3	0.00118	8.84×10^{-09}
cg08759276	*ND3*	16q24.1	-0.00153	3.31×10^{-09}
cg05275012	*ND4*	3p22.1	0.00176	8.81×10^{-08}
cg10517202	*ND5*	3q26.32	-0.00103	8.74×10^{-08}
cg09614565	*IL17RD*	3p14.3	-0.00283	5.69×10^{-08}
cg13689591	*KALRN*	3q21.1-q21.2	0.00199	2.58×10^{-08}
cg06934988	*USP43*	17p13.1	-0.00188	1.96×10^{-08}
cg06785213	*HAS1*	19q13.4	0.00087	1.06×10^{-08}
cg13332754	*ND6*	18q22.3	0.00164	1.39×10^{-07}
cg03551607	*ESD*	13q14.2	0.00115	1.34×10^{-07}
cg24364967	*CLHC1*	2p16.1	0.00122	1.22×10^{-07}
cg15057061	*SOX2OT*	3q26.33	-0.00151	1.08×10^{-07}

ND = No Data (ND_1 = *ARL4A* and *ETV1*, ND_2 = *ZSWIM4* and *NANOS3*, ND_3 = *FOXL1* and *C16orf95*, ND_4 = *ZNF621* and *CTNNB1*, ND_5 = *TBL1XR1* and *KCNMB2*, ND_6 = *TSHZ1* and *SMIM21*).

Conclusion

The study presented in the publication '*Epigenome-wide association study in healthy individuals identifies significant associations with DNA methylation and PBMCs extracts VEGF-A concentration'* detected 20 CpG sites associated with VEGF-A levels and gave new insights into epigenetic regulation of this important biomarker. This knowledge could be further useful in the understanding of genes that contribute to the regulation of VEGF-A levels *via* methylation-caused gene silencing.

A new study with a higher sample size is warranted to detect CpG sites related to the physiological regulation of VEGF-A. On the other hand, patient-populations are necessary to detect CpG sites involved in pathological regulation of VEGF-A level.

Epigenome-wide association study in healthy individuals identifies significant associations with DNA methylation and PBMCs extracts VEGF-A concentration

Gorenjak V[1], Vance RD[2], Dadé S[1], Stathopoulou MG[1], Doherty L[2], Xie T[1], Murray H[2], Masson C[1], Lamont J[2], Fitzgerald P[2], Visvikis-Siest S[1,3]

1. Université de Lorraine, Inserm, IGE-PCV, F-54000 Nancy, France
2. Randox Laboratories Ltd, Crumlin, UK
3. Department of Internal Medicine and Geriatrics, CHU Nancy-Brabois, Nancy, France

Abstract

Vascular endothelial growth factor A (VEGF-A) is a chemokine that induces proliferation and migration of vascular endothelial cells and is essential for both physiological and pathological angiogenesis. It is known for its high heritability (>60 %) and involvement in most common morbidities, which makes it a potentially interesting biomarker. Large GWAS have already identified polymorphisms related to VEGF-A. However, no previous research has provided epigenome-wide insight in regulation of VEGF-A. The following study aimed to comprehensively assess VEGF-A levels of 201 healthy participants from the STANISLAS Family Study (SFS) for association with DNA methylation. EWAS profile was determined using the 450K Infinium BeadChip Array (Illumina), VEGF-A concentration in PBMCs extracts was detected using a high-sensitivity multiplex Cytokine Array (Randox Laboratories, United Kingdom). EWAS analysis identified 41 methylation sites significantly associated with VEGF-A concentrations derived from PBMCs extracts. Twenty CpG sites within thirteen chromosomes reached Holm-Bonferroni significance. Significant values ranged from P=1.08 x 10^{-7} to P=5.64 x 10^{-15}. Methylation detected in promoter regions of *TPX2* and *HAS-1* could explain previously reported associations with *VEGFA* gene as well as the regulatory mechanisms of other genes located in the vicinity of detected CpG sites.

Keywords: VEGF-A, EWAS, Methylation, PBMC extracts

1. INTRODUCTION

Recent developments and discoveries in epigenetics provided new insights into disease regulation, among which exploration of DNA methylation has become the most intriguing (1). DNA methylation forms 5-methylcytosine on the CpG (cytosine-phosphate-guanine) site of a genome and normally results in silencing of the gene that is encoded in the sequence (2). This particularity was researched in various epigenome-wide methylation studies (EWAS), which managed to relate individual CpGs with cardiovascular diseases (3), cancer (4) and other pathologies (5, 6). In some cases, CpGs significantly associated with a certain disease are found on genes known to be involved with the aforementioned pathology or in promoter regions controlling gene expression (7). In many cases, associations with chromosomal positions of methylated sites and disease are not obvious. Intergenic regions with CpG islands are thus systematically studied to elucidate the role of methylation in genomic regions distant from protein-coding regions (8).

Vascular endothelial growth factor A (VEGF-A) is a myogenic protein that induces angiogenesis, endothelial cell proliferation and plays an important role in the regulation of vasculogenesis (9). VEGF-A is involved in the pathogenesis of cardiovascular disease (10), as well as other chronic diseases such as cancer (11), type 2 diabetes (12), osteoporosis, osteoarthritis (13) and chronic obstructive pulmonary disease (COPD) (14). Moreover, anti-VEGF medications containing a humanized antibody that blocks angiogenesis by inhibiting VEGF-A have already entered the market to treat a number of cancers, such as colon cancer, lung cancer, glioblastoma, and renal cell carcinoma, as well as age-related macular degeneration (15-17).

The involvement of VEGF-A in various diseases makes it a universal biomarker with great potential for patient stratification in personalised medicine. The precise understanding of its biological and genetic regulation is required to fully appreciate its clinical potential. In previous years, a major effort has resulted in the discovery of several strong genetic variants associated with growth factors, in particular VEGF-A concentration, using high powered genome-wide association studies (GWAS). Ten significant VEGF-A associated SNPs (19, 20) that explained more than 50 % of its variability has been identified. VEGF-A concentration is highly heritable, >60 % as demonstrated in the SFS (18). However, the genetic variants identified by GWAS explain only partially the variability of a phenotype. Previous research has not yet investigated the role of epigenetics or DNA methylation on VEGF-A concentration. Therefore epigenetic regulation could explain the missing heritability components (21).

Epigenetics is the study of gene transcription, regulation and expression that are not directly caused by the alteration of the genomic DNA sequence. Methylation (CpG) sites are often po-

sitioned in CpG islands (a densely populated area of CpG sites) within a promotor region, transcription start site (TSS), first or second exons of a gene, in an enhancer region, or upstream from genes with CpG island shores (2kb) or CpG shelves (2-4kb) (2). Previous studies have shown that epigenetics plays an important role in the regulation of promoter regions of *VEGFA* (22, 23) and *VEGFR* genes (24, 25), but no previous research studies have performed an EWAS of VEGF-A levels to determine the methylation sites responsible for regulation of *VEGFA*. Knowing and understanding the mechanisms of epigenetic regulation of *VEGFA* might allow further comprehension of the epigenetic regulation involved in chronic diseases, where VEGF-A plays a crucial role.

To this end, we comprehensively explored epigenome-wide associations of VEGF-A levels, measured from PBMCs extracts in a healthy population, in order to identify possible epigenetic mechanisms involved in *VEGFA* regulation, before the pathological onset of chronic disease. We performed a large *in silico* analysis to detect possible repeating patterns of CpG chromosomal positions that could explain the role of each CpG site in *VEGFA* regulation.

2. METHODS

2.1. Population

In this investigation we set out to explore links between genome-wide DNA methylation and PBMCs extracts VEGF-A concentration, in a population of 201 healthy individuals from the SFS. The SFS is a 10-year longitudinal survey with three visits at 5-year intervals, involving 1,006 families from Vandoeuvre-ls-Nancy, France, first recruited between 1993-1995 (26, 27). All subjects were of Caucasian origin, without the presence of chronic disorders, *e.g.* CVD or cancer, or previous personal history of such diseases. The study protocols were approved by the Comité Consultatif pour la Protection des Personnes dans la Recherche Biomédicale de Lorraine (Advisory Committee for the protection of people in biomedical research in Lorraine), and all subjects gave written informed consent for their participation in the study. All experiments were performed in accordance with relevant guidelines and regulations.

2.2. Data Collection

Biological and clinical measurements were determined using appropriate, validated procedures. Blood samples were collected between 8 and 9 a.m. after overnight fasting. DNA was extracted by the Miller technique (28) and was stored at −80 °C until further use. Body mass index (BMI) was calculated as weight (kg) divided by height^2 (m^2). All measurements were obtained by trained professionals.

2.3. Biological Measurements

2.3.1. Isolation of PBMCs

Full blood from healthy donors was collected into sodium heparin tubes. Samples were homogenized with Hanks Balanced Salt Solution (SIGMA Aldrich, reference: H6648) ($V_{Hanks} = V_{blood}$) and poured gently into a 15 mL tube with Ficoll^{TM} paque (Sigma Aldrich, reference: 17-1440-02) solution ($V_{Ficoll} = V_{Hanks} + V_{blood}$). The contents were centrifuged for 30 min at 300 g at room temperature. High density PBMC ring was retrieved and collected into a 15 mL tube, filled with Hanks Balanced Salt Solution and centrifuged for 10 min at 1000 g at room temperature (first washing). The supernatant was aspirated and 2 mL of Hanks Balanced Salt Solution was added. The tube was filled up to 15 mL with Hanks Balanced Salt Solution and centrifuged for a further 10 min at 1000 g at room temperature (second washing). The pellet was collected into an Eppendorf tube with 1 mL of Hanks Balanced Salt Solution. PBMCs populations were evaluated by microscopic observation after May-Grunwald-Giemsa staining and PBMCs concentration was normalised to 10^6 cells/mL in Hanks Buffer. After final centrifugation of 5 min at 1000 g at room temperature the supernatant was aspirated and the pellet of PBMCs was processed immediately or stored at −80 °C to maintain stability.

2.3.2. Total protein extraction

The lysis solution (lysate) was composed of 320 μL of cell lysis buffer (CelLytic^{TM}, SIGMA Aldrich, reference: C2978) and 1.6 μL of protease inhibitor (0.5%, Protease Inhibitor Cocktail, SIGMA Aldrich, reference: P8215) for the samples with counted cells (>10^6), and was added to the lymphocyte pellet. The mixture was stirred for 15 min at room temperature, and centrifuged for 15 min at 12000 g and 4 °C. The supernatant was collected and was immediately used for further analysis or stored at −80 °C to maintain stability.

2.3.3. VEGF-A measurement

PBMCs extracts concentrations of VEGF-A were estimated using the Randox high sensitivity multiplex cytokine and growth factor array (Evidence Investigator Analyzer, Randox Laboratories Ltd., Crumlin, United Kingdom). The descriptive statistics of studied population is presented in Table 4.

2.4. DNA methylation analysis

2.4.1. Methylation Assay

DNA methylation patterns were investigated using a method, previously described in details (29, 30). Briefly, genome-wide methylation profiling of bisulphite-converted genomic DNA was performed by Illumina HumanMethylation450 bead array (Illumina Inc., San Diego, CA, USA). Illumina is using Infinium I and II arrays with probes for detection of methylated and unmethylated CpG sites. Methylation ratio, referred to as beta value by Iluminas software, is the proportion methylated / (methylated + unmethylated) for each CpG in the population of cells from which we extracted DNA.

2.4.2. Quality control

Minfi R package (Bioconductor) was used to analyze and visualize Illumina Infinium methylation arrays (31). The first step in microarray data preprocessing consisted of removing all probes that can generate artifactual data. Firstly, a detection P-value was assigned to each probe. High detection P-value normally corresponds to a probe with a low quality sig-

nal, therefore probes with P>0.05 were removed from all samples. Furthermore, probes missing in >5 % of the samples were excluded. To avoid spurious associations, probes containing locations on the genome where variation is already annotated in HumanMethlyation450 annotation files (*i.e.* probes containing single-nucleotide polymorphism (SNP), sex chromosomes, and a single base extension (SBE) site) were excluded. Finally, probes containing cross-reactive and target polymorphic CpGs (32) were excluded, leaving 314 440 probes out of 484 777 for statistical analysis. In addition, one individual was excluded from our cohort after quality control checks of methylation array data (outlier of plotted median of the methylated against unmethylated intensity), leaving 200 individuals for the analysis.

2.4.3. Normalisation

The second step in microarray data preprocessing was removing sources of variation, related to technical limitations - data normalisation. Background correction, colour bias (dye bias) adjustment and Infinium I/II bias correction were carried out with Illumina background correction and SWAN (33) in the R package minfi.

2.4.4. Association study

CpGassoc (CRAN) was used to test for association between methylation at CpG sites across the genome and VEGF-A concentration in PBMCs extracts (34). The random mixed-effects model included gender, age, BMI, family structure and individual blood cell counts (neutrophils, lymphocytes, monocytes, eosinophils and basophils) as covariates and chip array as random effect. Holm-Bonferroni correction for multiple testing was applied to the result.

2.5. In silico analysis

Ensembl browser (35) was used for localisation of CpG sites on the Human Genome (GRCh38.p10), as well as for the establishment of regulatory features from their genomic environment. All annotated genes were investigated for interactions with VEGF-A gene using cytoscape app GeneMANIA (36).

3. RESULTS

The results of genome-wide methylation study pointed out forty-one probes that achieved statistical significance for methylated sites in relation to VEGF-A levels in cellular extracts. Twenty probes were significant after Holm-Bonferroni adjustment ($<1.6\text{x}10^{-7}$). The distribution of P-values for the level of methylation at each CpG site is shown in Figure1. Manhattan plot shows that methylation is spread across different chromosomes. Chromosome 19 and chromosome 3 showed more significantly associated methylation sites than other chromosomes.

Table 1 presents the list of twenty CpG sites that were significant after Holm-Bonferroni correction. Location and genes for each CpG site were retrieved from the annotation file. Six methylation sites did not have annotated genes. *In silico* analysis was thus conducted to explore the genetic environment of these methylation sites, results are presented in Figure3. Strong P-values have been detected, however, a small effect size was attributed to each significant CpG sites. Genes related to the CpG locations were not directly related to *VEGFA* gene. In order to reveal possible distant common genetic pathways, genomic environment of all CpG sites was studied in detail with Ensembl browser. Results are presented in the supplementary data (Sup. Figures 1 and 2). Moreover, common pathways were further analysed with GeneMANIA app.

GeneMANIA enables the construction of a composite gene-gene functional interaction network from a list of genes collected from many large, publicly available biological datasets (36). A list of 28 genes, retrieved from annotation results file and from *in silico* analysis, has been input into GeneMANIA to research their possible relation to *VEGFA*, namely *RPL31, GLIS3, TPX2, SEPSECS, LOC338799, SETD1B, AUTS2, PIGB, IL17RD, KALRN, USP43, HAS1, ESD, CLHC1, SOX2, ARL4A, ZSWIM4, FOXL1, ZNF621, TBL1XR1, TSHZ1, ZADH2, ETV1, NANOS3, C16orf95, CTNNB1, KCNMB2* and *SMIM21*. A gene network created as a result of this analysis is presented in Figure2. Some genes that were input were not found by the bioinformatics tool and are thus not presented in the figure. Some genes were not related to any of genes input into GeneMania and are also not presented in the figure (*TPX, C16orf95, KCNMB2, ZSWIM4, SETD1B, SMIM21, IL17RD* and *USP43*). Results revealed seven genes that were previously observed to have minor interactions with *VEGFA*, namely *ARL4A, ZADH2, SEPSECS, CTNNB1, TBL1XR1, GLIS3* and *ETV1* (37). *ZADH2, SEPSECS, CTNNB1, TBL1XR1, GLIS3* and *ETV1* had minor genetic interactions (presented with a green line in Figure2), *ARL4A* and *ZADH2* had similar expression levels with *VEGFA* in gene expression study (violate lines), *CTNNB1* and *VEGFA* were previously found to share a common molecular pathway (blue line). Most physical protein-protein interactions were detected within *SEPSECS, SOX* and *AUTS2* genes (rose lines).

4. DISCUSSION

We have comprehensively explored the DNA methylome in a population of healthy individuals and have identified 41 methylation sites significantly associated with VEGF-A concentrations derived from PBMCs extracts (see Sup. Table 2). Significance levels after Bonferroni correction ranged from P=1.08 x 10^{-7} (cg15057061) to P=5.64 x 10^{-15} (cg05739757) for 20 CpG sites. Ten CpGs produced a positive effect size (range: 0.00087 to 0.0027), whereas the remaining ten methylation biomarkers possessed a negative effect size based on VEGF-A concentrations (range: 0.00349 to 0.00103). This study is the first epigenome-wide association study investigating the links between DNA methylation and VEGF-A levels in a population of healthy individuals, and the importance of its findings will be discussed below.

We have analysed the Holm-Bonferroni significant CpGs using two different principles. Firstly, we have analysed the genes

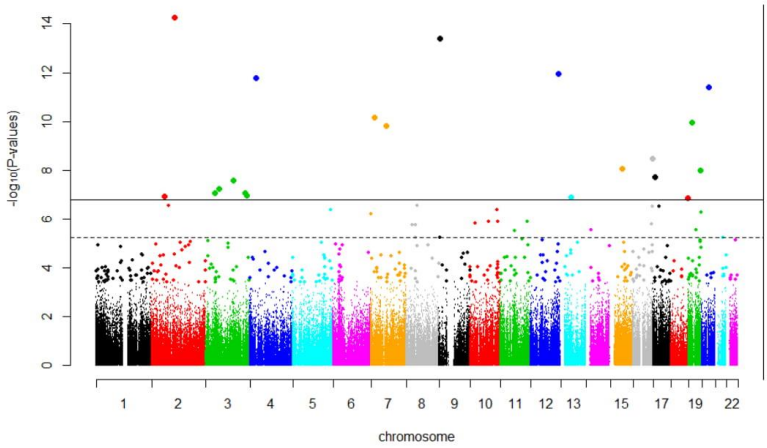


Figure 1: Manhattan plot displaying adjusted P-values of the association between methylation probes and VEGF-A levels in cell extracts. The dotted line represents FDR value, points above the full line indicate results that were found to be significant after Holm-Bonferroni testing.

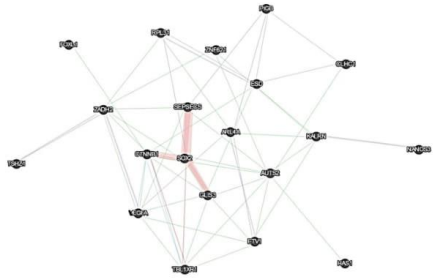


Figure 2: Network of genes related to CpG sites significantly associated with VEGF-A concentrations. Green lines present genetic associations, violet lines present co-expression, blue lines present common pathways and rose lines present physical interactions between connected genes.

encoded in the position of the CpG site, focusing on their function, relation to diseases and association with VEGF-A (Sup. Table 3). Fourteen genes were annotated in the result file obtained after performing EWAS (Table 1) using CpGassoc R package. To find the genomic features within the location of the rest six CpG sites, we have performed *in silico* analysis using Ensembl browser (Figure3). Secondly, we have studied the genetic environment of each CpG schematically, to detect possible common patterns related to the location of CpG on the genome (Sup. Figure 1 and 2). The overall aim of such analysis was to explore publically available databases in order to detect common biological pathways between VEGF-A and the concerned genes, which were previously associated, but not explained. DNA methylation could provide an answer to these interactions.

Furthermore, the patterns in which methylation occurs on the genome could provide us with new information on methylation function, and pave the way for novel hypotheses that could explain the function of CpGs in non-coding regions or locations without direct relation to a specific phenotype. All of the genes related to 20 Holm-Bonferroni significant CpGs were also input into GeneMania, to explore common genetic and physiological pathways. Seven genes were associated with *VEGFA* (***ARL4A***, *ZADH2, SEPSECS,* ***CTNNB1, TBL1XR1***, *GLIS3* and ***ETV1***). For the genes highlighted in bold the relation with VEGF-A has been further confirmed with bibliographical research (Sup. Table 3).

Methylation of CpGs, located on gene coding sections of DNA, has the potential to silence gene expression, which is especially important in disease development (38). Abnormal patterns of DNA methylation have been observed in cancer, most commonly in CpG islands in gene promoter regions (39). Methylation of the *VEGFA* gene has been observed in previous studies (40, 41). Moreover, research has shown that VEGF-A receptors *KDR* and *FLT4* are also silenced by DNA methylation. However, if the promoters are unmethylated, other factors (*e.g.* transactivation factors) determine the extent of *KDR* and *FLT4* expression (24, 42). Previous attempts of targeted methylation of the *VEGFA* promoter, provoked with generated fusion of de novo DNA methyltransferase and an artificial zinc-finger protein (23, 43) or lentivirus-mediated delivery of shRNA molecules (44), resulted in strong silencing of *VEGFA* expression and indicated that gene-specific methylation targeting can become a new strategy to achieve silencing of the gene in human cells.

CpG Site (strand)		cg16333561 (-)	cg15014826 (+)	cg08759276 (-)	cg05275012 (-)	cg10517202 (-)	cg13332754 (-)
Chromosome		7p21.2	19p13.12	16q24.1	3p22.1	3q26.32	18q22.3
Nearest genomic feature (product/strand)	Upstream	AC011287.1 (novel lincRNA/+)	ZSWIM4 (protein coding/+)	AC009108.4 (Unknown/+)	AC122683.1 (lincRNA/-)	LINC00578 (lincRNA/+)	TSHZ1 (protein coding/+)
	Upstream	AC005019.2 (novel lincRNA /-)	AC020916.1 (lincRNA/-)	AC009108.2 (lincRNA/-)	HMGN2P24 (pseudogene /-)		TSHZ1 (protein coding/+)
	Downstream					RN7SKP52 (misc RNA/-)	
Location		7:13803079-13803128	19:13833585-13833634	16:86610656-86610705	3:40619738-40619787	3:177469455-177469504	18:75290150-75290199
Regulatory features (T cells, natural killer cells or B cells)		1 Enhancer & 2 CTCF binding sites	8 Promoters	4 Promoters & 1 CTCF binding site	1 Promoter, 1 Enhancer & 3 CTCF binding sites	1 Promoter; 1 CTCF & 1 Enhancer	1 Promoter; 2 Enhancers & 1 CTCF binding site
Nearest coding gene (strand)	Upstream	ARL4A (+)	ZSWIM4 (+)	FOXL1 (+)	ZNF621 (+)	TBL1XR1 (-)	TSHZ1 (+)
	Downstream	ETV1 (-)	NANOS3 (+)	C16orf95 (-)	CTNNB1 (+)	KCNMB2 (+)	SMIM21 (-)

Figure 3: Supplementary information retrieved by *in silico* analysis for methylation sites without annotated gene and location.

None of the CpGs identified in this study was located within or near the *VEGF* gene family or its associated genes (see Table 3). However, some of the identified CpGs have been previously implicated in VEGF-related biological processes, such as cell proliferation, cell growth, angiogenesis and related diseases (see Sup. Table 3). One such relation was found with *TPX2* gene (probe cg21838233), (P= 4.01 x 10^{-12}). The *TPX2* gene is overexpressed in colon cancer, leading to vessel invasion and metastasis of colon cancer cells (45). *TPX2* gene silencing results in the inhibition of cell proliferation, and this effect has been linked to down-regulation of the expression of *VEGFA* (46). Cg21838233 is located within promoter region of *TPX2*, where methylation can play a crucial role in the control of gene expression (47). The degree of methylation and location of the methylation site may directly affect the transcription and subsequent expression of the gene. Therefore, it is biologically plausible to assume that the expression of *TPX2* is controlled by methylation of cg21838233, which in turn reflects further in the expression of *VEGFA*.

Another intriguing result was cg06785213 (P= 1.06 x 10^{-8}) which was found in the vicinity of the *HAS1* gene, 62 bp in 5 upstream region. The *HAS1* family of genes code for hyaluronic acid (HA), which has an essential role in tissue development and homeostasis, and directs the initiation and progression of various pathological conditions, including angiogenesis (48). Both proteins, HAS-1 and VEGF-A, have an important role in antigenic cascade. Moreover, it was suggested that engagement of the CD44 receptors for HA by depolymerised HA induces the endogenous release of VEGF-A (49, 50). Thus, methylation in the vicinity of the *HAS1* gene could play a distinct role in regulation of this process. Further research is required to confirm these estimations and elucidate new epigenetic pathways.

Five other genes in sup. Table 3 were related to antigenic processes, namely *ARL4A, ETV1, CTNNB1, TBL1XR1* and *TSHZ1*, located in the vicinity of CpGs detected in non-coding regions. In total, 6 out of 20 CpG sites from non-coding regions were significantly associated with VEGF-A. Little is known of whether such CpGs can have a real impact on genes in their proximity. However, it is known that it is not only the sequence in the immediate proximity of a region, such as a promoter, that can influence gene activity (51).

Moreover, DNA regions that were previously considered as junk DNA are now being considered as indispensable elements of regulation of gene expression (51). Looking upstream and downstream of 6 annotated CpG sites, we have discovered that the most common genetic features in their immediate proximity were long noncoding RNAs (lncRNAs), which are emerging as regulators of gene expression in pathogenesis (52). Cascade CpG-lincRNAs could take part in regulation of coding genes (*e.g. ARL4A, ETV1* or *CTNNB1*) and could thus impact on *VEGFA* regulation. Such estimation is very limited. However, it is important that we consider all of the options that might, in

Table 1: Novel associations between *VEGFA* gene expression and DNA methylation in a subset of the SFS, after Holm-Bonferroni correction (P<1.6 x 10^{-7}). Data retrieved from an annotation file.

CpG Site	Chromosone	Gene	Location	β	Effect Size	P-Value
cg05739757	2q11.2	RPL31	TSS200	0.07	0.00176	5.64 x 10-15
cg23333878	9p24.2	GLIS3	5'UTR; 1stExon	0.08	0.00270	3.86 x 10-14
cg21838233	20q11.21	TPX2	1stExon; 5'UTR	0.04	0.00191	4.01 x 10-12
cg18815539	4p15.2	SEPSECS	1stExon; 5'UTR	0.04	-0.00197	1.70 x 10-12
cg21968169	12q24.31	LOC338799, SETD1B	TSS1500	0.09	-0.00189	1.11 x 10-12
cg16333561	7q11.23	ND	ND	0.90	-0.00130	7.02 x 10-11
cg20547575	7q11.22	AUTS2	Body	0.03	-0.00349	1.52 x 10-10
cg15014826	19p13.12	ND	ND	0.03	-0.00157	1.12 x 10-10
cg00117600	15q21.3	PIGB	5'UTR;1stExon	0.04	0.00118	8.84 x 10-09
cg08759276	16q24.1	ND	ND	0.78	-0.00153	3.31 x 10-09
cg052750126	3p22.1	ND	ND	0.94	0.00176	8.81 x 10-08
cg10517202	3q26.32	ND	ND	0.88	-0.00103	8.74 x 10-08
cg09614565	3p14.3	IL17RD	TSS200	0.02	-0.00283	5.69 x 10-08
cg13689591	3q21.1-q21.2	KALRN	Body	0.88	0.00199	2.58 x 10-08
cg06934988	17p13.1	USP43	TSS200	0.02	-0.00188	1.96 x 10-08
cg06785213	19q13.4	HAS1	TSS200	0.20	0.00087	1.06 x 10-08
cg13332754	18q22.3	ND	ND	0.95	0.00164	1.39 x 10-07
cg03551607	13q14.2	ESD	5'UTR	0.04	0.00115	1.34 x 10-07
cg24364967	2p16.1	CLHC1	5'UTR	0.12	0.00122	1.22 x 10-07
cg15057061	3q26.33	SOX2OT	Body	0.04	-0.00151	1.08 x 10-07

Table 2: Population characteristics.

	Mean	SD
Age (years)	28.3	14.8
Sex (male %)	50.2	NA
VEGF-A	59.3	75.5
BMI	21.6	4.0
Neutrophils	53.77	9.14
Lymphocytes	36.01	8.46
Monocytes	6.22	2.45
Eosinophils	2.84	1.98
Basophils	0.64	0.39

Table 3: List of genes related to VEGF protein family and VEGF receptors.

Gene	Chr.	Band	Location
VEGFA	6	p21.1	43,770,183-43,786,487
VEGFB	11	q13.1	64,234,538-64,238,793
VEGFC	4	q34.3	176,683,538-176,792,727
VEGFD	X	p22.2	15,345,591-15,384,376
PGF	14	q24.3	74,941,834-74,955,784
FLT1	13	q12.3	28,300,344-28,495,145
KDR	4	q12	55,078,477-55,125,589
FLT4	5	q35.3	180,601,506-180,650,271
NRP1	10	p11.22	33,177,492-33,336,262
HIF1A	14	q23.2	61,695,401-61,748,259
VWF	12	p13.31	12:5,948,874-6,124,770
TGFB1	19	q13.2	41,301,587-41,353,933
SRC	20	q11.23	37,344,685-37,406,050
IGF1	12	q23.2	102,395,867-102,481,786
MMP9	20	q13.12	46,008,908-46,016,561

the future, elucidate important regulation pathways.

Schematic presentation of genetic regulatory elements in the vicinity of CpGs (Sup. Figure 1 and 2) demonstrated that most CpG sites were located within promoter regions, a regulatory region of DNA, where transcription is initiated. Normally, CpG islands within promoters are well characterised, but sometimes they are found in deserted areas. However, there is evidence that some orphan CpG islands may initiate transcription and are likely to represent either uncharacterised promoters of genes or promoters driving transcription of noncoding RNA (51). CTCF binding sites present the second most common element. It enables CTCF zinc-finger transcription factor to bind and thus activate or repress the activity of various genes. Moreover, it can act as enhancer-blocker (53). An enhancer is a third regulatory element found in the vicinity of CpGs. It enhances the transcription of genes by interactions with trans-acting factors, which allows specific control of gene activation, through chromatin looping of the intervening DNA (51).

We have noticed that regulatory elements are becoming less frequent with the distance from a CpG site. It means that identified CpG sites were located within regulatory vivid regions, indicating that CpGs could also be involved as an important element in regulation, without being located directly on the gene coding part.

5. CONCLUSION

We have found a significant association between DNA methylation and VEGF-A levels measured form the PBMCs cellular extracts. Epigenetic regulation of *TPX2* and *HAS* genes could affect their activity and thus explain the associations with VEGF-A levels, demonstrated in this study. Moreover, it could explain the regulation mechanisms of other genes located in vicinity of detected CpG sites.

6. REFERENCES

1. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell. 2007;128(4):635-8.
2. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484-92.
3. Handy DE, Castro R, Loscalzo J. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. Circulation. 2011;123(19):2145-56.
4. Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. Hum Mol Genet. 2007;16 Spec No 1:R50-9.
5. Portela A, Esteller M. Epigenetic modifications and human disease. Nature biotechnology. 2010;28(10):1057-68.
6. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nature reviews Genetics. 2011;12(8):529-41.
7. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465-76.
8. Medvedeva YA, Fridman MV, Oparina NJ, Malko DB, Ermakova EO, Kulakovskiy IV, et al. Intergenic, gene terminal, and intragenic CpG islands in the human genome. BMC Genomics. 2010;11(1):48.
9. Neufeld G, Cohen T, Gengrinovitch S, Poltorak Z. Vascular endothelial growth factor (VEGF) and its receptors. The FASEB journal. 1999;13(1):9-22.
10. Ferrara N, Gerber HP, LeCouter J. The biology of VEGF and its receptors. Nat Med. 2003;9(6):669-76.
11. Carmeliet P. VEGF as a key mediator of angiogenesis in cancer. Oncology. 2005;69(Suppl. 3):4-10.
12. Awata T, Inoue K, Kurihara S, Ohkubo T, Watanabe M, Inukai K, et al. A common polymorphism in the 5-untranslated region of the VEGF gene is associated with diabetic retinopathy in type 2 diabetes. Diabetes. 2002;51(5):1635-9.
13. Liu Y, Berendsen AD, Jia S, Lotinun S, Baron R, Ferrara N, et al. Intracellular VEGF regulates the balance between osteoblast and adipocyte differentiation. J Clin Invest. 2012;122(9):3101-13.
14. Ding Y, Niu H, Li Y, He P, Li Q, Ouyang Y, et al. Polymorphisms in VEGF-A are associated with COPD risk in the Chinese population from Hainan province. Journal of genetics. 2016;95(1):151-6.
15. Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. New England journal of medicine. 2004;350(23):2335-42.
16. Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, et al. Paclitaxelcarboplatin alone or with bevacizumab for nonsmall-cell lung cancer. New England Journal of Medicine. 2006;355(24):2542-50.
17. Escudier B, Pluzanska A, Koralewski P, Ravaud A, Bracarda S, Szczylik C, et al. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. The Lancet. 2007;370(9605):2103-11.
18. Berrahmoune H, Herbeth B, Lamont JV, Masson C, Fitzgerald PS, Visvikis-Siest S. Heritability for plasma VEGF concentration in the Stanislas family study. Annals of human genetics. 2007;71(Pt 1):54-63.
19. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, et al. Six Novel Loci Associated with Circulating VEGF Levels Identified by a Meta-analysis of Genome-Wide Association Studies. PLoS genetics. 2016;12(2):e1005874.
20. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, et al. Identification of cis- and trans-acting genetic variants explaining up to half the variation in circulating vascular endothelial growth factor levels. Circ Res. 2011;109(5):554-63.
21. Mayhew AJ, Meyre D. Assessing the Heritability of Complex Traits in Humans: Methodological Challenges and Opportunities. Current genomics. 2017;18(4):332-40.
22. Groh A, Jahn K, Burkert A, Neyazi A, Schares L, Janke E, et al. Epigenetic Regulation of the Promotor Region of Vascular Endothelial Growth Factor-A and Nerve Growth Factor in Opioid-Maintained Patients. European addiction research. 2017;23(5):249-59.
23. Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, Reinhardt R, et al. Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3aDnmt3L single-chain fusion protein with increased DNA methylation activity. Journal of molecular biology. 2013;425(3):479-91.
24. Kim JY, Whang JH, Zhou W, Shin J, Noh SM, Song IS, et al. The expression of VEGF receptor genes is concurrently influenced by epigenetic gene silencing of the genes and VEGF activation. Epigenetics. 2009;4(5):313-21.
25. Kim J, Hwang J, Jeong H, Song H-J, Shin J, Hur G, et al. Promoter methylation status of VEGF receptor genes: a possible epigenetic biomarker to anticipate the efficacy of intracellular-acting VEGF-targeted drugs in cancer cells. Epigenetics. 2012;7(2):191-200.
26. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clinical chemistry and laboratory medicine. 2008;46(6):733-47.

27. Siest G, Visvikis S, Herbeth B, Gueguen R, Vincent-Viry M, Sass C, et al. Objectives, design and recruitment of a familial and longitudinal cohort for studying gene-environment interactions in the field of cardiovascular risk: the Stanislas cohort. Clinical chemistry and laboratory medicine. 1998;36(1):35-42.
28. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988;16(3):1215.
29. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288-95.
30. Dedeurwaerder S, Defrance M, Bizet M, Calonne E, Bontempi G, Fuks F. A comprehensive overview of Infinium HumanMethylation450 data processing. Briefings in bioinformatics. 2014;15(6):929-41.
31. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.
32. Chen Y-a, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8(2):203-9.
33. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol. 2012;13(6):R44.
34. Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. Bioinformatics. 2012;28(9):1280-1.
35. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. Database : the journal of biological databases and curation. 2016;2016.
36. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server issue):W214-20.
37. Lin A, Wang RT, Ahn S, Park CC, Smith DJ. A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. Genome Research. 2010;20(8):1122-32.
38. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6-21.
39. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med. 2003;349(21):2042-54.
40. Pisani F, Cammalleri M, Dal Monte M, Locri F, Mola MG, Nicchia GP, et al. Potential role of the methylation of VEGF gene promoter in response to hypoxia in oxygen-induced retinopathy: beneficial effect of the absence of AQP4. Journal of cellular and molecular medicine. 2018;22(1):613-27.
41. Turunen MP, Yl-Herttuala S. Epigenetic regulation of key vascular genes and growth factors. Cardiovascular Research. 2011;90(3):441-6.
42. Quentmeier H, Eberth S, Romani J, Weich HA, Zaborski M, Drexler HG. DNA methylation regulates expression of VEGF-R2 (KDR) and VEGF-R3 (FLT4). BMC cancer. 2012;12(1):19.
43. Snowden AW, Gregory PD, Case CC, Pabo CO. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. Current biology. 2002;12(24):2159-66.
44. Turunen MP, Lehtola T, Heinonen SE, Assefa GS, Korpisalo P, Girnary R, et al. Efficient regulation of VEGF expression by promoter-targeted lentiviral shRNAs based on epigenetic mechanism: a novel example of epigenetherapy. Circulation research. 2009;105(6):604-9.
45. Wei P, Zhang N, Xu Y, Li X, Shi D, Wang Y, et al. TPX2 is a novel prognostic marker for the growth and metastasis of colon cancer. Journal of Translational Medicine. 2013;11:313.
46. JIAN J, HUANG Y, LIU L-z, LI S, DENG F. TPX2 gene-silencing inhibits the proliferation and invasion of human colon cancer SW480 cells. TUMOR. 2016;36(6):628-34.
47. Robertson KD, Jones PA. DNA methylation: past, present and future directions. Carcinogenesis. 2000;21(3):461-7.
48. Pardue EL, Ibrahim S, Ramamurthi A. Role of hyaluronan in angiogenesis and its utility to angiogenic tissue engineering. Organogenesis. 2008;4(4):203-14.
49. Murphy JF, Lennon F, Steele C, Kelleher D, Fitzgerald D, Long AC. Engagement of CD44 modulates cyclooxygenase induction, VEGF generation, and proliferation in human vascular endothelial cells. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2005;19(3):446-8.
50. Rodgers LS, Lalani S, Hardy KM, Xiang X, Broka D, Antin PB, et al. Depolymerized hyaluronan induces vascular endothelial growth factor, a negative regulator of developmental epithelial-to-mesenchymal transformation. Circ Res. 2006;99(6):583-9.
51. Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cellular and molecular life sciences : CMLS. 2012;69(21):3613-34.
52. SasChen A, Aure MR, Leibovich L, Carvalho S, Enuka Y, Krner C, et al. EMBO molecular medicine. 2016;8(9):1052-64.
53. Plasschaert RN, Vigneau S, Tempera I, Gupta R, Maksimoska J, Everett L, et al. CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. Nucleic Acids Res. 2014;42(2):774-89.

Publication Nº 5

Peripheral blood mononuclear cells extracts VEGF protein levels and VEGF mRNA: Associations with inflammatory molecules in a healthy population.

Vesna Gorenjak*, Dwaine R. Vance*, Alexandros M. Petrelis, Maria G. Stathopoulou, Sébastien Dadé, Said El Shamieh, Helena Murray, Christine Masson, John Lamont, Peter Fitzgerald, Sophie Visvikis-Siest

*equal contribution

PlosONE, August 2018

Angiogenesis is an important physiological mechanism that enables the formation of new blood vessels through VEGF-A signalling pathways and is implicated in the pathology of a wide range of diseases (43, 177). Besides VEGF-A-induced angiogenesis, other pathways are associated with progression of vessel-growth. For instance, inflammatory cells are likely to drive angiogenesis by contributing to tumour growth, progression, and immunosuppression (299). The elucidation of common pathways between VEGF-A isoforms and inflammatory molecules could serve to provide new biomarkers used for stratification of patients under an anti-angiogenic therapy, and in the development of novel targeted therapies for treatments of chronic pathologies related to these processes.

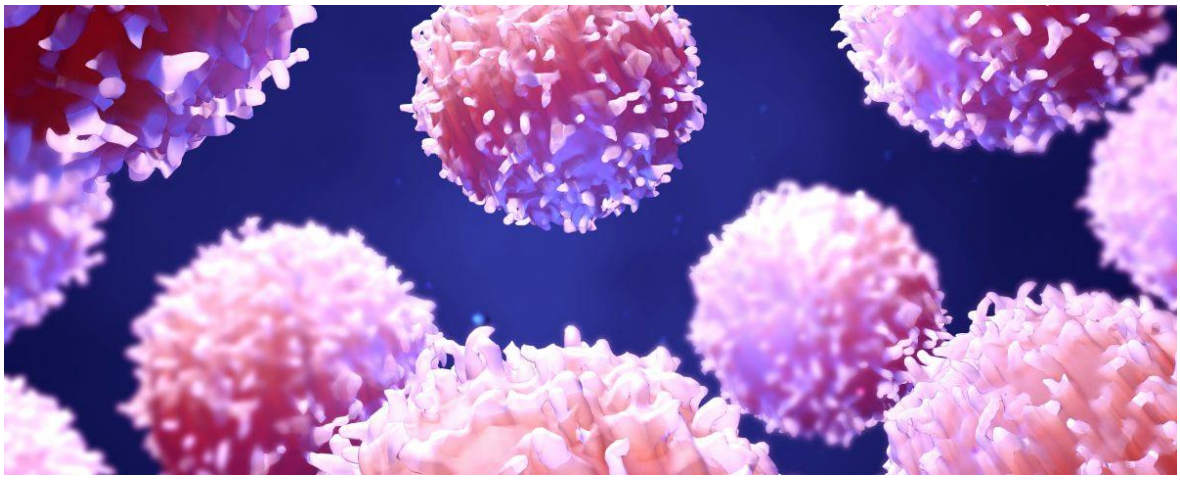


Figure 30: PBMCs are highly specialised immune cells, which are secreting critical components of the innate and adaptive immune system, among also inflammatory cytokines (300).

PBMCs are highly specialised immune cells (Figure 30), consisting of populations of B cells (~15 %), T cells (~70 %), monocytes (~5 %), and natural killer (NK) cells (~10 %) (301). PBMCs are secreting VEGF-A, as well as other important inflammatory proteins involved in complex molecular pathways that impact on physiological balance in the human organism (54). Therefore, PBMCs extracts might be a good source for studying the association of VEGF-A with other inflammatory markers. Moreover, the utilisation of the healthy population could help to elucidate common regulatory pathways in normal physiological mechanisms, which would, in turn, facilitate the understanding of pathophysiological processes.

Objective

The goal of this study was to investigate the associations of eleven molecules, commonly implicated in inflammation with *VEGF-A* mRNA levels of four most abundant VEGF-A isoforms ($VEGF_{165}$, $VEGF_{189}$, $VEGF_{145}$ and $VEGF_{121}$) and with VEGF-A protein levels in PBMCs, in order to determine common physiological pathways that could impact on the angiogenic process.

Populations and analysis

Participants from the SFS (n=285) were selected for the measurement of twelve molecules derived from PBMCs extracts (IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, MCP-1, TNF-α, INF-γ, EGF and VEGF-A).

Moreover, a subset of 110 participants was used for quantification of four specific *VEGF-A* mRNA isoforms from total RNAs derived from PBMCs (*i.e.* $VEGF_{121}$, $VEGF_{145}$, $VEGF_{165}$ and $VEGF_{189}$).

Non-parametric correlation analysis (Spearman correlation) and linear regression models were applied, adjusted for age and gender, to test for possible associations between VEGF protein levels and inflammatory molecules.

VEGF-A protein levels were pre-tested for normal distribution using the Kolmogorov-Smirnov test of normality. Consequently, VEGF-A protein levels were log-transformed to

normalise the distribution of data. Data was tested for outliers, which were removed before correlation and linear regression analysis.

Associations of specific *VEGF-A* mRNA isoforms and inflammatory molecules were assessed using identical statistical procedures.

Mixed models adjusted for the family structure were also applied to correct for possible issues of stratification due to familial resemblance.

Results

A correlation study has been performed between VEGF-A protein levels and inflammatory molecules. A significant trend was observed between VEGF-A and EGF, IL-1β, IL-8, MCP1 and TNF-α. Moreover, VEGF-A protein levels were associated with IL-4, MCP-1 and EGF levels in linear regression models adjusted for age and gender (Table 6).

Table 6. Significant associations between VEGF-A protein levels and inflammatory molecules.

	VEGF-A		
	β	**SE**	**P**
Interleukin 4	0.028	0.011	0.013
Monocyte Chemoattractant Protein 1	0.015	0.003	<0.0001
Epidermal Growth Factor	0.017	0.003	<0.0001

SE: Standard error, **β:** Effect size

A correlation study has been performed between *VEGF-A* mRNA levels and inflammatory molecules. A significant trend was observed between $VEGF_{121}$ and IL-4, between $VEGF_{165}$ and IL-4, IL-6, and between $VEGF_{189}$ and IFN-γ, IL-1β, IL-4, IL-6.

Two *VEGF-A* mRNA isoforms were significantly associated with four inflammatory molecules. $VEGF_{165}$ was associated with MCP-and IL-1α, whereas $VEGF_{189}$ was associated with IL-4 and IL-6 in a linear regression model adjusted for age and gender (Table 7).

Table 7. Significant associations between the expression of *VEGF-A* mRNA isoforms and inflammatory molecules.

	$VEGF_{165}$			$VEGF_{189}$		
	β	**SE**	**P**	**β**	**SE**	**P**
Monocyte Chemoattractant Protein 1	-0.319	0.006	0.002	-	-	-
Interleukin 1 alpha	-0.269	0.010	0.008	-	-	-
Interleukin 4	-	-	-	-0.290	0.017	0.019
Interleukin 6	-	-	-	-0.260	0.110	0.034

SE: Standard error, **β:** Effect size

The linear regression analyses for VEGF-A protein and mRNA levels were repeated using mixed models adjusted for family structure. The mixed model gave the same significant results as the linear regression model.

Altogether, five inflammatory molecules showed association with VEGF-A (protein or mRNA) levels (Figure 31). Detailed bibliographical research has been done to explain the relations of identified cytokines with VEGF-A.

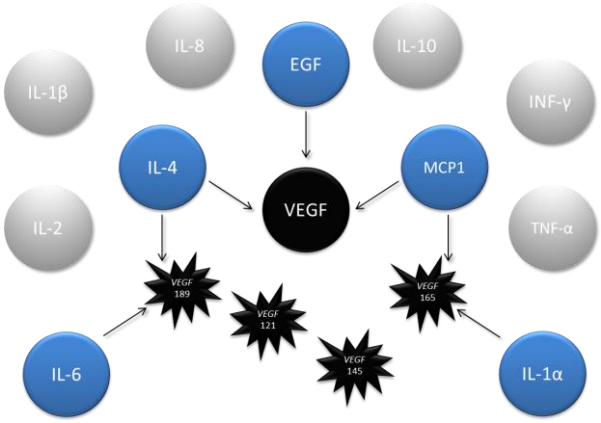


Figure 31: The relation between inflammatory cytokines and VEGF-A protein/mRNA levels.

Conclusion

The manuscript entitled '*Peripheral blood mononuclear cells extracts VEGF protein levels and VEGF mRNA: Associations with inflammatory molecules in a healthy population'* gave new insights in the association of inflammatory cytokines with VEGF-A, which could contribute to the understanding of the angiogenesis process and could ultimately offer novel targeted therapies for VEGF-A-related pathologies.

PLOS ONE

RESEARCH ARTICLE

Peripheral blood mononuclear cells extracts VEGF protein levels and VEGF mRNA: Associations with inflammatory molecules in a healthy population

Vesna Gorenjak[1☯], Dwaine R. Vance[2☯], Alexandros M. Petrelis[1], Maria G. Stathopoulou[1], Sébastien Dadé[1], Said El Shamieh[3], Helena Murray[2], Christine Masson[1], John Lamont[2], Peter Fitzgerald[2], Sophie Visvikis-Siest[1,4]*

1 Université de Lorraine, Inserm, IGE-PCV, Nancy, France, **2** Randox Laboratories Limited, Crumlin, Co. Antrim, Northern Ireland, United Kingdom, **3** Department of Medical Laboratory Technology, Faculty of Health Sciences, Beirut Arab University, Beirut, Lebanon, **4** Department of Internal Medicine and Geriatrics, CHU Technopôle Nancy-Brabois, Rue du Morvan, Vandoeuvre-lès-Nancy, France

☯ These authors contributed equally to this work.
* sophie.visvikis-siest@inserm.fr

Check for updates

OPEN ACCESS

Citation: Gorenjak V, Vance DR, Petrelis AM, Stathopoulou MG, Dadé S, Shamieh SE, et al. (2019) Peripheral blood mononuclear cells extracts VEGF protein levels and VEGF mRNA: Associations with inflammatory molecules in a healthy population. PLoS ONE 14(8): e0220902. https://doi.org/10.1371/journal.pone.0220902

Editor: Rudolf Kirchmair, Medical University Innsbruck, AUSTRIA

Received: February 6, 2019

Accepted: July 25, 2019

Published: August 16, 2019

Copyright: © 2019 Gorenjak et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data set was uploaded as supporting information.

Funding: Financial support was provided by the European Union within the frames of the Operational Programme FEDER-FSE Lorraine et Massif des Vosges 2014-2020, by Agence Nationale de la Recherche, programme d'Investissements d'avenir, grant number ANR-15RHU-0004 and GEENAGE project (Functional Genomic, Epigenomic and ENvironment interplay

Abstract

Background

Vascular endothelial growth factor (VEGF) is a signal protein, implicated in various physiological and pathophysiological processes together with other common inflammatory biomarkers. However, their associations have not yet been fully elucidated. In the present study, we investigated associations between VEGF and four specific *VEGF* mRNA isoforms with levels of 11 inflammation molecules, derived from peripheral blood mononuclear cells (PBMCs) extracts.

Methods

Healthy participants from the STANISLAS Family Study (n = 285) were included. Levels of VEGF (four mRNA isoforms and protein levels) and inflammatory molecules (IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, INF-γ, TNF-α, MCP-1, EGF) were measured in PBMCs extracts. Multiple regression analyses were performed, adjusted for age and gender.

Results

The analyses revealed significant associations between VEGF protein levels and levels of IL-4 (β = 0.028, P = 0.013), MCP-1 (β = 0.015, P<0.0001) and EGF (β = 0.017, P<0.0001). Furthermore, mRNA isoform $VEGF_{165}$ was associated with MCP-1 and IL-1α (P = 0.002 and P = 0.008, respectively); and mRNA isoform $VEGF_{189}$ was associated with IL-4 and IL-6 (P = 0.019 and P = 0.034, respectively).

to IMPACT management of healthy and pathological AGEing)" within the frames of the LUE project (Lorraine University of Excellence). VG was partially funded by the "Région Lorraine". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist. Some authors are employed by a commercial company: "Randox Laboratories Limited". This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

Conclusions

To our knowledge, the present study represents the first investigation that successfully demonstrates links between VEGF protein levels and inflammatory molecules levels derived from PBMCs extracts and identifies associations between specific *VEGF* mRNA isoforms and inflammatory molecules.

Impact

These findings provide novel insights that may assist in the development of new tissue and mRNA isoform specific measurements of VEGF levels, which may positively contribute to predicting the risk of common complex diseases and response of currently used anti-VEGF agents, and developing of novel targeted therapies for VEGF-related pathophysiology.

Introduction

Vascular endothelial growth factor A, VEGF-A (commonly referred as VEGF), is a multifunctional signal protein, which works as an important regulator of both physiological and pathological angiogenesis and has been related to a variety of pathologies, such as cancer and cardiovascular diseases (CVD) [1]. VEGF has become a perspective target for the design of anti-cancer treatments; anti-VEGF medications have already entered the clinical environment, however, the trade-off for the therapy is a common occurrence of cardiovascular side effects [2, 3].

VEGF is a prototype member of a cytokine family, which also includes placental growth factor (*PLGF*), VEGF-B, VEGF-C and VEGF-D, present in regulation of lymphangiogenesis, vasodilatation, chemotactic for different cells and vascular permeability [1]. Levels of VEGF represent a highly heritable phenotype (>60.5% as demonstrated in STANISLAS Family Study (SFS)) [4]. More than 50% of this variability is explained by ten single nucleotide polymorphisms identified through two genome-wide association studies (GWAS) [5, 6].

The *VEGF* gene produces more than 14 messenger RNA (mRNA) isoforms; the most predominant are $VEGF_{121}$, $VEGF_{145}$, $VEGF_{165}$ and $VEGF_{189}$, denoted by their length (number of amino acids). These isoforms differ in biochemical properties and receptor-binding characteristics that result in different effects on vessel growth [7]. Various tissues express different ratios of *VEGF* mRNA isoforms, including tumours, where growth appears to be most rapid when the isoform $VEGF_{164}$ is expressed [7]. Therefore, disease susceptibility may depend on transcription of specific *VEGF* mRNA isoforms, rather than the currently measured VEGF protein. As a result, studies exploring specific *VEGF* mRNA isoforms for association with inflammatory molecules, intermediate phenotypes and diseases are warranted.

Human VEGF isoforms are classified into two main families: the *VEGFxxx* family and the *VEGFxxxb* family (xxx denoting the number of amino acids) [8]. The isoforms of these families differ only in the sequence of carboxy-terminal six amino acids, as the result of alternative splicing of exon eight of the *VEGF* gene [9]. This difference leads to isoforms with opposite functions; VEGFxxx isoforms (e.g. $VEGF_{165}$) have pro-angiogenic properties whereas the VEGFxxxb isoforms (e.g. $VEGF_{165b}$) have anti-angiogenic properties [10, 11]. VEGFxxxb isoforms present more than half of the total VEGF expressed in vitreous fluid, circulating plasma, urine, renal cortex, colonic epithelium, bladder smooth muscle, lung and pancreatic islets, whereas in tissues with physiological angiogenesis (placenta) or pathological angiogenesis

(melanoma, colorectal or bladder cancer cells) VEGFxxx isoforms represent the majority of total VEGF [8].

VEGF is secreted from various cells: fibroblasts, tumour cells and inflammatory cells, such as lymphocytes [12, 13]. Lymphocytes are the most numerous components of peripheral blood cells (70–90% of PBMCs) and hold a central role in the regulation of the immune system [14]. Originating from a common lymphoid progenitor, there are three subgroups of lymphocytes, namely T-lymphocytes (70–85%), B-lymphocytes (5–20%) and natural killer cells (NK) (5–20%), each with their own function in immune response [15, 16]. Chronic inflammation often results in the infiltration of inflammatory cells at specific sites, among which the predominant are T-lymphocytes [17]. In addition to VEGF, other important signalling proteins are secreted from lymphocytes and are regulated by intracellular signalling control mechanisms or extracellular regulators [18]. Together, they are involved in complex molecular pathways that impact on physiological balance in human organism. The most important cytokine families involve hematopoietins, interferons, tumour necrosis factors-related molecules and chemokines [19]. VEGF levels have been previously related with cellular adhesion molecules (CAMs) [20], interleukin-1α (IL-1α)[21], interleukin-1β (IL-1β) [22], interleukin-4 (IL-4) [23–25], interleukin-6 (IL-6) [26, 27], nuclear factor-κB (NF-κB) transcription factor [28], endothelial growth factor (EGF) and others [29–31].

A detailed knowledge of biological pathways of VEGF is indispensable for further progress in pharmacological studies and detection of different isoforms of VEGF is of crucial importance for anticancer therapy with limited side effects. In addition, unravelling the relations between different biomarkers in physiological state of organisms can provide the basic knowledge that will help to understand these relations in pathological conditions.

The aim of this investigation is to assess the relationship between VEGF and inflammatory molecules, including cytokines and growth factors: IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, interferon-γ (INF-γ), tumour necrosis factor α (TNF-α), monocyte chemotactic protein 1 (MCP-1) and EGF, derived from PBMCs extracts in healthy individuals. PBMCs represent one of the most important sources of inflammatory molecules; therefore, we examined associations of cytokines and growth factors contained within PBMCs extracts utilizing a multiplex chemiluminescent biochip. Furthermore, because of the importance of VEGF isoforms, we examined the mRNA expression profiles of the four most biologically important VEGF isoforms ($VEGF_{121}$, $VEGF_{145}$, $VEGF_{165}$ and $VEGF_{189}$) and explored associations of each specific *VEGF* mRNA isoform with inflammatory molecules.

Materials and methods

Population

All subjects involved in this study make part of the STANISLAS Family Study (SFS). Information pertaining to this cohort has previously been described [32, 33]. Briefly, the SFS is a 10-year longitudinal survey with 3 visits at 5-year intervals, involving 1,006 families from Vandoeuvre-lès-Nancy, France, first recruited between 1993 and 1995. All subjects were of Northwest-European origin, without the presence of chronic disorders *e.g.* CVD or cancer, and without previous personal history of CVD. Study protocols were approved by the institutional ethics committees CCPPRB de Lorraine (Comité consultatif de protection des personnes dans la recherche biomédicale) and CNIL (Commission Nationale de l'Informatique et des Libertés). All individuals gave written informed consent for their participation in the study.

A subset of 285 participants (138 females and 147 males) from the SFS was selected for the measurement of twelve inflammatory molecules derived from PBMCs extracts.

A subset of 110 participants from the SFS was selected to quantify four specific *VEGF* mRNA isoforms from total RNA derived from PBMCs (*i.e.* $VEGF_{121}$, $VEGF_{145}$, $VEGF_{165}$ and $VEGF_{189}$), based on the sample availability.

Laboratory measurements

Isolation of PBMCs. Isolation of PBMCs was based on the method first described by Boyum in 1968 [34]. Briefly, whole blood was collected in tubes with sodium heparin and transported at room temperature. Hanks' Balanced Salt Solution (SIGMA Aldrich, reference: H6648) was added into 15 mL tubes with blood ($V_{Hanks} = V_{blood}$) and poured gently into a 15 mL tube with Ficoll paque plus (Sigma Aldrich, reference: 17-1440-02) solution ($V_{Ficoll} = V_{Hanks} + V_{blood}$). The contents were centrifuged for 30 min at 300 g at room temperature.

A PBMCs ring was retrieved and collected into a 15 mL tube, filled with Hanks' Balanced Salt Solution and centrifuged for 10 min at 1000 g at room temperature. The supernatant was aspirated and 2 mL of Hanks' Balanced Salt Solution were added. The solution was well suspended, filled up to 15 mL with Hanks' Balanced Salt Solution and centrifuged for 10 min at 1000 g at room temperature (second washing). The PBMC ring was collected into Eppendorf tube with 1 mL of Hanks' Balanced Salt Solution. PBMCs populations were evaluated by microscopic observation after May-Grunwald-Giemsa staining and the PBMCs concentration was normalized to 10^6 cells/mL in Hank's Buffer. After final centrifugation 5 min at 1000 g at room temperature the supernatant was aspirated and the pellet of PBMCs was processed immediately or stored at -80 ˚C to maintain stability.

Total protein extraction. The lysis solution (lysate) composed of cell lysis buffer (CelLytic-M, SIGMA Aldrich, reference: C2978) and protease inhibitor (0.5%, Protease Inhibitor Cocktail, SIGMA Aldrich, reference: P8215) was added to the PBMC pellet, as recommended by the manufacturer (SIGMA Aldrich). This was stirred for 15 min at room temperature, and centrifuged for 15 min at 12000 g at 4 ˚C. The supernatant was collected and was immediately used for further analysis or stored at -80 ˚C to maintain stability.

Protein measurement. Quantification of inflammatory molecules (IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, MCP-1, TNF-α, INF-γ, EGF and VEGF (combination of $VEGF_{121}$, $VEGF_{145}$, $VEGF_{165}$ and $VEGF_{189}$ isoforms protein levels)) from PBMCs extracts was performed by Randox high sensitivity multiplex cytokine and growth factor array (Evidence Investigator Analyzer, Randox Laboratories Ltd., Crumlin, United Kingdom), which is a Randox patented 9 x 9 mm^2 activated biochip with spatially discrete test regions containing antibodies specific to each of the inflammatory molecules assessed. In the present study, the combination of the four isoforms measured with the Cytokine array is referred to as "VEGF protein level".

Gene expression analysis. Total RNA was extracted and quantified from isolated PBMCs using the MagNA Pure LC RNA HP isolation kit and RNA HP Blood External lysis protocol (Roche Diagnostics, France) as previously described [35, 36]. In short, 200 units of M-MuLV Reverse Transcriptase with 0.25 μg of oligos (dT) (Promega, France) were used to perform reverse transcription of total RNAs. Quantitative real-time PCR was performed on LightCycler instrument (Roche Diagnostics, Mannheim, Germany) with TaqMan Master Kit for all *VEGF* transcripts. Specific primers were used to selectively quantify the transcripts coding for the *VEGF* isoforms (S1 Table). All quantifications were carried out in duplicates, the starting amount of cDNA template was the same in all samples (25 ng) to remove any bias resulting from difference in the initial RNA quantity. Positive controls with known concentration were included in every run to ensure reproducibility by comparing the expression of *VEGF* transcripts between different PCR runs. A standard curve was generated by plotting the whole

range of series dilution against the initial template quantity; this showed a linear standard curve and an efficiency of amplification near 99%. The absolute quantity of *VEGF* transcripts (number of copies/μL) in every sample was calculated by comparing the sample Ct value to the standard curve (Standard Curve Quantification method, The LightCycler 480 Software, Roche Diagnostic, France).

Statistical analysis

VEGF protein levels were tested for normal distribution using the Kolmogorov-Smirnov test of normality. Consequently, VEGF protein levels were log transformed (log_{10}) to normalize a distribution of data. Data was tested for outliers (1.5×IQR interquartile range), which were removed before further analysis. Non-parametric correlation analyses (Spearman correlation) were performed. Linear regression models were applied, adjusted for age and gender, to test for possible associations between VEGF protein levels and inflammatory molecules.

Associations of specific VEGF mRNA isoforms and inflammatory molecules were assessed using identical statistical procedures. Only VEGF isoform 145 values didn't follow the normal distribution and were log transformed before further analysis.

Mixed models adjusted for family structure were also applied to correct for possible issues of stratification due to familial resemblance.

All analyses were performed using SPSS 20.0 statistical software (SPSS, Armonk, New York: IBM Corp). Significance was determined at a two-tailed 0.05 level. Graphs for correlation and regression analyses were performed using R 3.5.2 statistical software with Jtools package.

Bioinformatics analysis

Alignment between different isoforms of VEGF was performed using Clustal-Omega program with the following parameters: default transition matrix Gonnet, gap opening penalty 6 bits and gap extension 1 bit. Clustal-Omega uses the HHalign algorithm and its default settings as its core alignment engine. The algorithm is described in details in Söding, J. [37].

Results

All twelve cytokines were detected in the studied population. All values represent the concentrations of proteins (pg/mL) measured in the cellular extracts. Concentrations are based on equal number of PBMCs (10^6) in all samples, counted before cell lysis. The characteristics of the study population are presented in Table 1.

Associations between VEGF levels and inflammatory molecules

VEGF protein levels were correlated with EGF, IL-1β, IL-8, MCP1 and TNF-α (Fig 1). Results are presented in Table 2, the significant P-values are presented in bold.

VEGF protein levels were associated with IL-4, MCP-1 and EGF levels in linear regression models adjusted for age and gender (Fig 2). Results are presented in Table 3.

Associations between VEGF mRNA isoforms and levels of inflammatory molecules extracted from PBMCs

The Randox cytokine array detects four most common VEGF isoforms, however it does not quantify them individually but displays a sum of all VEGF detected in the sample. It has been previously confirmed that various isoforms portray different roles in pathophysiological processes; this is why their quantification and association to other inflammatory molecules are of particular interest. As a result, we investigated associations between 11 inflammatory

Table 1. Characteristics of the study population (n = 285).

Characteristic	Mean	SD
Age (years)	39.13	14.57
Gender (%) male	51.58	-
Interleukin 1 alpha (pg/mL)	0.90	4.84
Interleukin 1 beta (pg/mL)	5.76	27.61
Interleukin 2 (pg/mL)	1.19	1.72
Interleukin 4 (pg/mL)	7.59	1.47
Interleukin 6 (pg/mL)	0.66	1.43
Interleukin 8 (pg/mL)	68.11	147.06
Interleukin 10 (pg/mL)	0.64	0.24
Interferon gamma (pg/mL)	0.55	3.06
Tumour Necrosis Factor alpha (pg/mL)	3.87	16.01
Monocyte Chemoattractant Protein 1 (pg/mL)	9.40	5.60
Epidermal Growth Factor (pg/mL)	6.80	6.50
Vascular Endothelial Growth Factor (pg/mL)	56.99	67.62

SD: Standard deviation

https://doi.org/10.1371/journal.pone.0220902.t001

molecules derived from PBMCs extracts and the expression of specific *VEGF* mRNA isoforms derived and quantified from PBMCs in a subset of individuals from the SFS cohort (Table 4).

Studies have established that $VEGF_{165}$ and $VEGF_{121}$ are the most abundantly expressed VEGF isoforms [38, 39]. Our detection of mRNA isoforms of VEGF in PBMCs has confirmed these findings.

$VEGF_{121}$ isoform was correlated with IL-4 (Fig 3), $VEGF_{165}$ isoform was correlated with IL4 and IL6 (Fig 4) and $VEGF_{189}$ isoform was correlated with IFN-γ, IL-1β, IL-4 and IL-6 (Fig 5). Results are presented in Table 5, the significant P-values are presented in bold.

Two *VEGF* mRNA isoforms were significantly associated with four inflammatory molecules in linear regression analysis. $VEGF_{165}$ was associated with MCP-1 ($P = 0.002$) and IL-1α ($P = 0.008$) (Fig 6), whereas $VEGF_{189}$ was associated with IL-4 ($P = 0.019$) and IL-6 ($P = 0.034$) (Fig 7). Results are presented in Table 6.

The STANISLAS cohort used in this study is composed of related individuals, therefore the analysis have been repeated using mixed models adjusted for family structure. The mixed model gave the same significant results as the linear regression model.

Multiple alignment of VEGF isoforms

Analysis of primary and secondary structures of VEGF isoforms provides elements of understanding that may explain the commitment to different signaling pathways. $VEGF_{189}$ and $VEGF_{165}$ represent the same protein and differ only by 25 amino acids (Fig 8).

P15692 corresponds to $VEGF_{206}$ considered as canonical sequence of VEGF. P15692-2 represents $VEGF_{189}$ and P15692-4 represents $VEGF_{165}$. Annotations are only depicted on the canonical sequence and stars indicate perfect alignment. The black frame shows the difference in amino acid composition (25) between $VEGF_{189}$ and $VEGF_{165}$.

Thus, $VEGF_{189}$ compared to $VEGF_{165}$ owns additional residues, allowing to interact with molecules, such as IL-4 and IL-6. Without them, the signaling pathway seems to implicate others actors, such as MCP-1 and IL-1α. A detailed analysis shows that these surnumerary amino acids harbor specific properties (Fig 9).

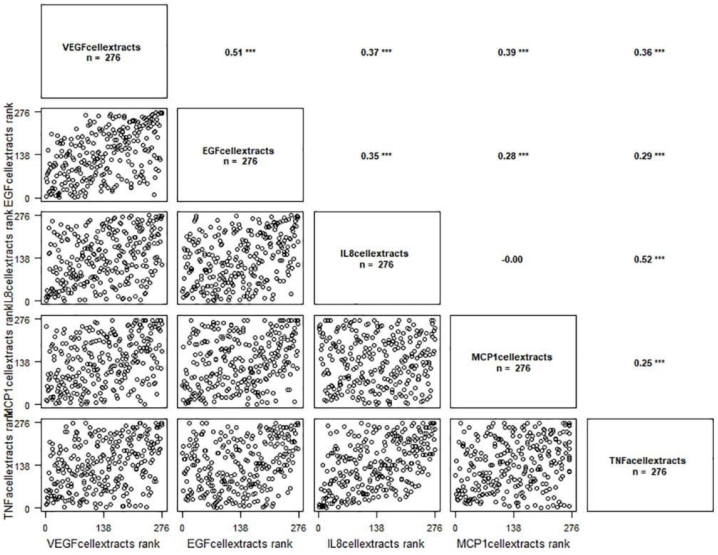


Fig 1. Correlation graph between VEGF protein levels (VEGF cellular extracts) and EGF, IL-8, MCP1 and TNF-α cellular extracts levels. '***' $p \leq 0.001$.

https://doi.org/10.1371/journal.pone.0220902.g001

P15692-2 corresponds to $VEGF_{189}$ and P15692-4 to $VEGF_{165}$. Analysis of amino acid properties revealed that $VEGF_{189}$ owns 25 amino acids with three features. First, 22 are polar (top alignment highlighted in purple), second, 16 belong to the category of bulky amino acids (highlighted in blue), and third, 13 have a positive charge (highlighted in green). The last one is a superposition of the three.

These additional residues may modify polarity, steric hindering and the electrostatic potential of $VEGF_{189}$ leading to change its counterparts and providing elements that could explain why these two close proteins are not associated with the same molecules.

Table 2. Correlation analysis between VEGF protein levels and inflammation molecules.

	VEGF	
	Correlation coefficient	**P-value**
EGF	0.5529	**< .0001**
IFN-γ	0.1677	0.2599
IL-1α	0.001	0.9945
IL-1β	0.2896	**0.0483**
IL-2	-0.0911	0.5427
IL-4	0.2256	0.1274
IL-6	-0.0077	0.959
IL-8	0.3393	**0.0197**
IL-10	-0.0571	0.7028
MCP1	0.3868	**0.0072**
TNF-α	0.394	**0.0061**

https://doi.org/10.1371/journal.pone.0220902.t002

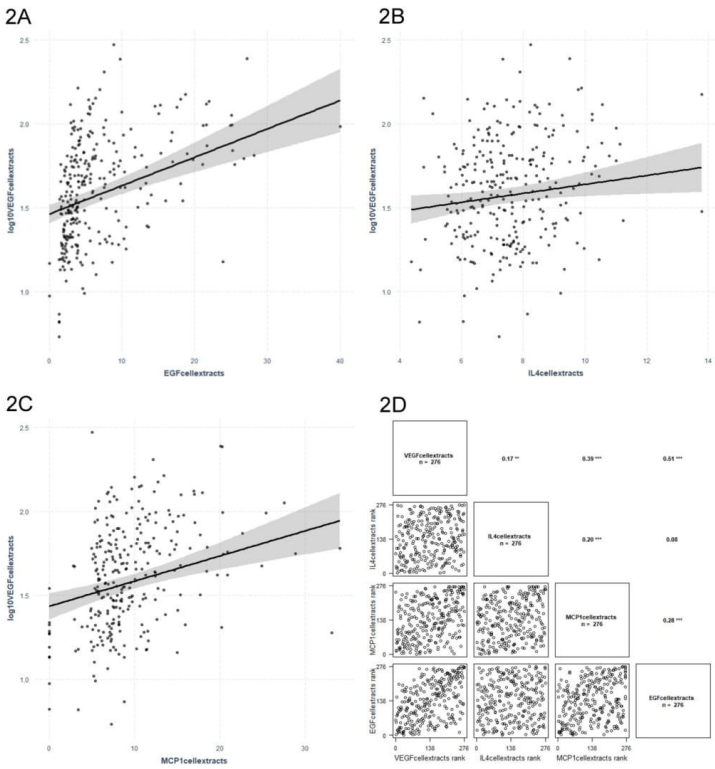


Fig 2. Regression graphs for VEGF protein levels (VEGF cellular extracts) as dependent variable and EGF (2A), IL-4 (2B) and MCP1 (2C) cellular extracts levels as independent variables. The related correlation graph is presented in (2D). '***' $p \leq 0.001$, '**' $p \leq 0.01$.

https://doi.org/10.1371/journal.pone.0220902.g002

Discussion

In the present study, we explored the relationship between VEGF protein levels and inflammatory molecules derived from PBMCs extracts from healthy individuals from the SFS cohort. In

Table 3. Significant determinants of VEGF protein levels extracted from PBMCs (linear regression analyses adjusted for age and gender).

		VEGF	
Inflammatory molecule	**β**	***SE***	***P***
Interleukin 4	0.028	0.011	0.013
Monocyte Chemoattractant Protein 1	0.015	0.003	<0.0001
Epidermal Growth Factor	0.017	0.003	<0.0001

SE: Standard error

https://doi.org/10.1371/journal.pone.0220902.t003

Table 4. Characteristics of the subsample of the population used for assessment of associations between inflammatory molecules levels from PBMCs extracts and *VEGF* mRNA isoforms (n = 110).

	Mean	SD
Age (years)	47.35	10.78
Gender (%) male	48.18	-
Interleukin 1 alpha (pg/mL)	0.81	2.75
Interleukin 1 beta (pg/mL)	4.88	17.42
Interleukin 2 (pg/mL)	0.90	1.3
Interleukin 4 (pg/mL)	7.75	1.55
Interleukin 6 (pg/mL)	0.614	1.22
Interleukin 8 (pg/mL)	67.65	135.91
Interleukin 10 (pg/mL)	0.65	0.24
Interferon gamma (pg/mL)	0.30	0.36
Tumour Necrosis Factor alpha (pg/mL)	3.85	16.55
Monocyte Chemoattractant Protein 1 (pg/mL)	9.58	5.59
Epidermal Growth Factor (pg/mL)	6.84	7.09
VEGF 121	47.96	19.95
VEGF 145	47.19	22.52
VEGF 165	250.52	111.77
VEGF 189	15.03	7.17

VEGF: Vascular endothelial growth factor, VEGF transcript copies number/25 ng cDNA

https://doi.org/10.1371/journal.pone.0220902.t004

addition, specific mRNA VEGF isoforms from PBMCs were assessed for association with inflammatory molecules. To the best of our knowledge, this is the first study to investigate levels of inflammatory molecules in PBMCs extracts and their relationship with VEGF protein levels and specific VEGF mRNA isoforms.

In summary, VEGF protein levels were correlated with EGF, IL1-β, IL-8, MCP1 and TNFα levels, while linear regression model adjusted for age and gender demonstrated that IL-4 ($P = 0.013$), MCP-1 ($P \leq 0.0001$) and EGF ($P \leq 0.0001$) levels were significant predictors of VEGF levels. The mRNA $VEGF_{121}$ isoform was correlated with IL-4 levels, while the $VEGF_{165}$ was correlated with IL-4 and IL-6 levels and the $VEGF_{189}$ was correlated with IFN-γ, IL-1β, IL-4 and IL6 levels. The regression models adjusted for age and sex demonstrated that for the expression of the mRNA $VEGF_{165}$ isoform the MCP-1 ($P = 0.002$) and IL-1α ($P = 0.008$) levels were significant predictors, while for the expression of the mRNA $VEGF_{189}$ isoform, the IL-4 ($P = 0.019$) and IL-6 ($P = 0.034$) levels were significant predictors.

PBMCs are particularly interesting cells, worthy of investigation because of their essential role in the synthesis and release of inflammatory cytokines and growth factors. Therefore, it is biologically plausible to explore associations of inflammatory molecules with VEGF protein levels and specific *VEGF* mRNA isoforms (*i.e.* $VEGF_{121}$, $VEGF_{145}$, $VEGF_{165}$ and $VEGF_{189}$) derived from PBMCs. PBMCs are a complement of cells derived from whole blood, composed of lymphocytes, monocytes, and dendritic cells. In humans, the frequencies of these cell populations vary across individuals. Lymphocytes are typically in the range of 70% to 90% of total PBMCs, monocytes range from 10% to 30% of total PBMCs, while dendritic cells are uncommon, estimated at only 1% to 2% of total PBMCs [16].

PBMCs are commonly used in immunology to examine cytokine secretion under modified condition [40, 41], in vaccine development [42, 43] and can serve as tissue for gene expression studies [44]. Gene expression of *VEGF* isoforms has been systematically measured in PBMCs for different purposes; to compare the expression of different isoforms in particular disease

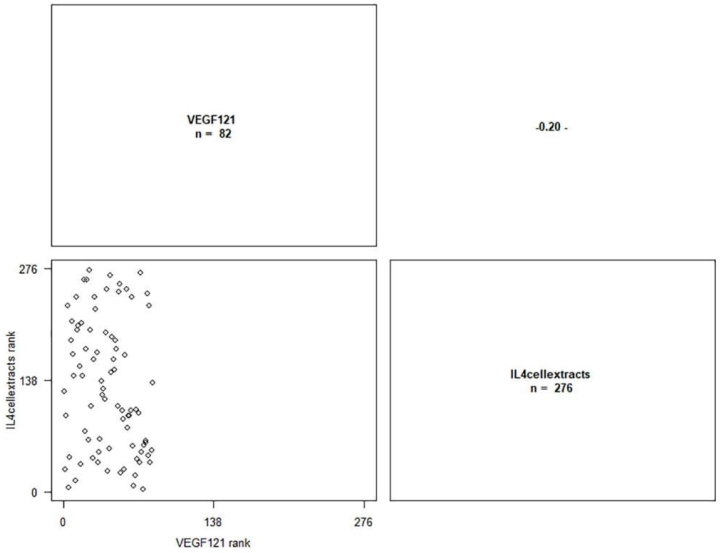


Fig 3. Correlation graph between $VEGF_{121}$ isoform mRNA levels and IL-4 cellular extracts levels. '.' $p \leq 0.1$.

https://doi.org/10.1371/journal.pone.0220902.g003

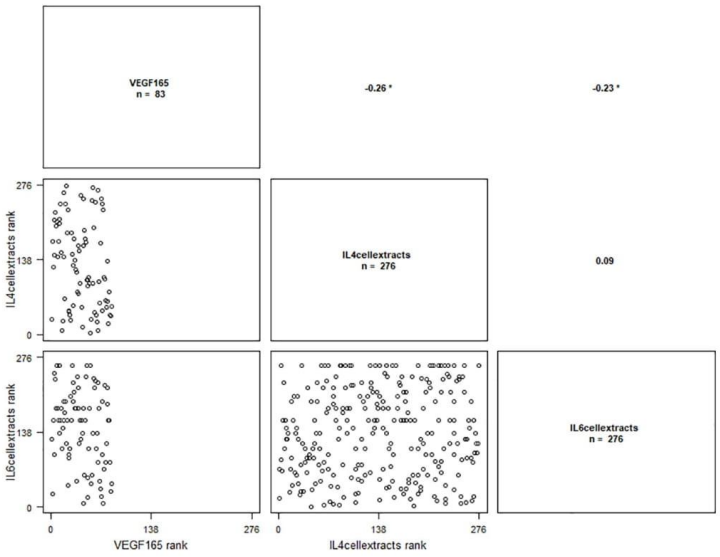


Fig 4. Correlation graph between $VEGF_{165}$ isoform mRNA levels and IL-4 and IL-6 cellular extracts levels. '*' $p \leq 0.05$.

https://doi.org/10.1371/journal.pone.0220902.g004

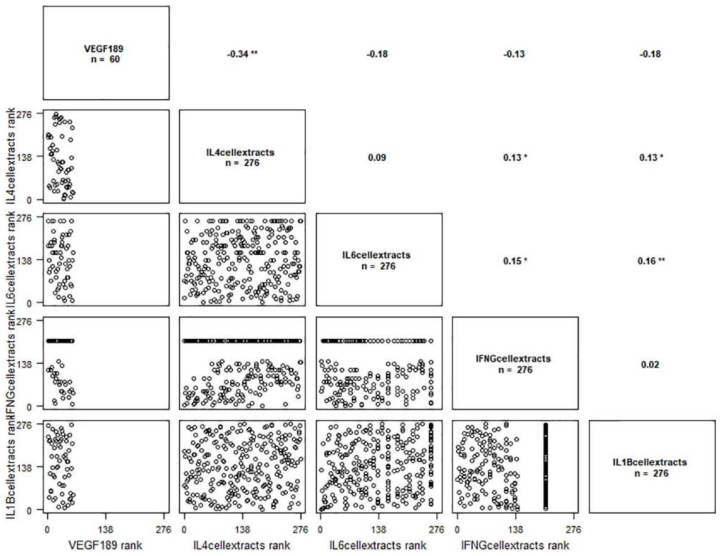


Fig 5. Correlation graph between $VEGF_{189}$ isoform mRNA levels and IL-4, IL-6, INFγ et IL-1β cellular extracts levels. '***' $p \leq 0.01$, '**' $p \leq 0.05$.

https://doi.org/10.1371/journal.pone.0220902.g005

[45, 46] or to understand the control of *VEGF* expression [47]. Besides *VEGF* expression profile, PBMCs have been used to detect RNAs for characterization of cardiovascular diseases [48, 49], diabetic nephropathy [50], rheumatic diseases [51], infectious diseases [52] and many other diseases. Protein expression of inflammatory cytokines from PBMCs has also been extensively studied. Increased expression of TNF-α and IL-6 along with increased secretion of leptin was found in obese cardiovascular patients [53]. Moreover, studies showed that protein expression or plasmatic levels of cytokines released by PBMCs can be indicators of pathologic state of organism [54]. Indeed, inflammation plays a central role in development of many

Table 5. Correlation analysis between VEGF isoforms and inflammation molecules.

	$VEGF_{121}$		$VEGF_{145}$		$VEGF_{165}$		$VEGF_{189}$	
	Correlation coefficient	**P-value**	**Correlation coefficient**	**P-value**	**Correlation coefficient**	**P-value**	**Correlation coefficient**	**P-value**
EGF	-0.1626	0.2748	-0.0607	0.6855	-0.1426	0.3388	-0.2128	0.1509
IFN-γ	-0.2412	0.1025	0.0034	0.9821	-0.2872	0.0503	-0.3415	**0.0188**
IL-1α	0.1276	0.3928	-0.236	0.1102	0.1256	0.4001	-0.0042	0.9778
IL-1β	-0.171	0.2504	-0.1735	0.2435	-0.1907	0.1992	-0.3088	**0.0347**
IL-2	-0.1759	0.237	0.129	0.3874	-0.1727	0.2457	-0.1155	0.4393
IL-4	-0.4345	**0.0023**	-0.1329	0.3731	-0.3643	**0.0118**	-0.3302	**0.0234**
IL-6	-0.2215	0.1346	-0.1797	0.2269	-0.3769	**0.009**	-0.3368	**0.0206**
IL-8	-0.0797	0.5942	0.0147	0.9216	-0.1354	0.3641	-0.1641	0.2702
IL-10	-0.0535	0.7208	-0.0787	0.5992	-0.0619	0.6794	-0.0676	0.6518
MCP1	-0.008	0.9577	-0.1209	0.4184	-0.0007	0.9963	-0.2326	0.1156
TNF-α	-0.0662	0.6584	0.0679	0.6503	-0.1271	0.3944	-0.2303	0.1194

https://doi.org/10.1371/journal.pone.0220902.t005

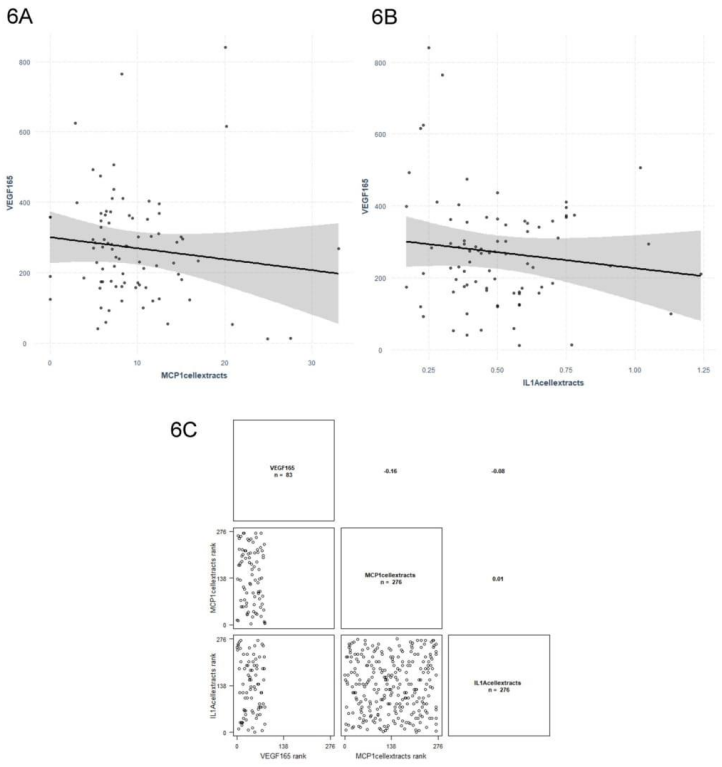


Fig 6. Regression graphs for *VEGF*$_{165}$ isoform mRNA levels as dependent variable and MCP1 (6A) and IL-1α (6B) cellular extracts levels as independent variables. The related correlation graph is presented in (6C). *** $p \leq 0.01$.

https://doi.org/10.1371/journal.pone.0220902.g006

chronic diseases [55] and PBMCs play an important role in inflammatory cytokine production.

Our investigation exploring the association between VEGF and inflammatory molecules followed a two-staged approach. Firstly, VEGF protein levels detected by microarray were analyzed. However, this method is not sensitive to specific VEGF isoforms; therefore, we used PBMCs also for mRNA isolation and quantification of *VEGF* expression isoforms. This enabled quantification of the expression of four isoforms in the second stage of the study. Two of the inflammatory molecules, IL-4 and MCP-1, were significantly associated to VEGF protein levels, and were also associated with the expression of *VEGF* specific isoforms 189 and 165, respectively.

Firstly, IL-4 was related with VEGF levels derived from PBMCs extracts *and VEGF*$_{189}$ mRNA levels detected from PBMCs. IL-4 is a multi-functional cytokine with anti-inflammatory and anti-tumor activity, produced by lymphocytes, basophils and mast cells [56]. Its anti-inflammatory properties are due to the ability of controlling the production of pro-inflammatory mediators by inhibiting their induction [57]. IL-4 has been previously related to VEGF as

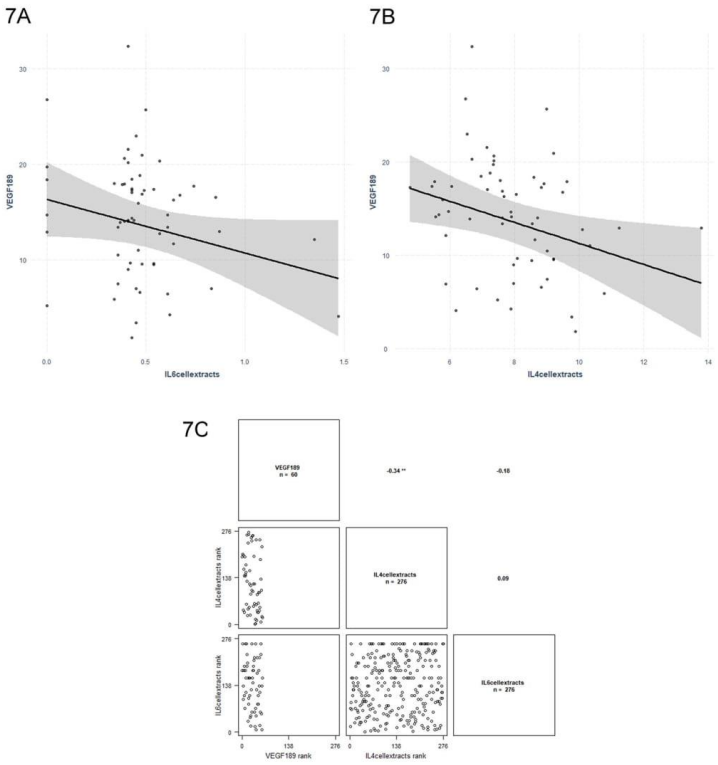


Fig 7. Regression graphs for $VEGF_{189}$ isoform mRNA levels as dependent variable and IL-6 (7A) and IL-4 (7B) cellular extracts levels as independent variables. The related correlation graph is presented in (7C).

https://doi.org/10.1371/journal.pone.0220902.g007

an important factor in the recruitment of tumor-associated macrophages (TAMs), which are known to promote angiogenesis, tissue remodeling and immunosuppression. As it was demonstrated in the study of Linde et. al [23, 58], VEGF is responsible for recruitment of monocytes from the peripheral circulation, whereas IL-4 induces their differentiation into tumor-

Table 6. Significant associations between the expression of *VEGF* mRNA isoforms and inflammatory molecules (linear regression models adjusted for age and gender).

	$VEGF_{165}$			$VEGF_{189}$		
Inflammatory molecule	**β**	***SE***	***P***	**β**	***SE***	***P***
Monocyte Chemoattractant Protein 1	-0.319	0.006	0.002	-	-	-
Interleukin 1 alpha	-0.269	0.010	0.008	-	-	-
Interleukin 4	-	-	-	-0.290	0.017	0.019
Interleukin 6	-	-	-	-0.260	0.110	0.034

VEGF: *Vascular endothelial growth factor*

https://doi.org/10.1371/journal.pone.0220902.t006

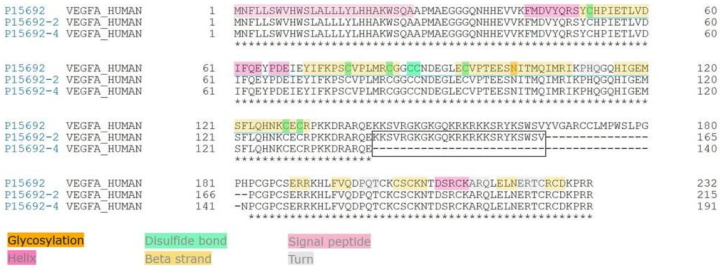


Fig 8. Alignment and elements of secondary structure of three VEGF isoforms (82% identity).

https://doi.org/10.1371/journal.pone.0220902.g008

promoting M2-like macrophages. The association of VEGF and IL-4 was also detected in human coronary artery endothelial cells [59] and in smooth muscle cells [25]; both studies noticed induced release of VEGF while treated with IL-4. Faffe et al. [25] also reported the increasing of VEGF-mRNA expression, caused by IL-4, but observed no alteration of VEGF promoter activity. Additionally, IL-4 has been shown to increase VEGF levels in fibroblast-like synoviocytes when applied alone, but portrayed an anti-angiogenic effect in the presence of transforming growth factor (TGF)-β, by inhibiting the VEGF production, in the study of patients with rheumatoid arthritis [56]. Thus, these results demonstrate the importance of understanding the synergy of different molecules, which can elucidate important biological pathways and thus open new possibilities for treatment of complex diseases. Previous studies have not proposed a specific VEGF isoform related to IL-4. Of note, our study is the first to

Fig 9. Focus on the 25 specific amino acids of $VEGF_{189}$ versus $VEGF_{165}$.

https://doi.org/10.1371/journal.pone.0220902.g009

propose a link between the isoform 189 and IL-4, but an additional research is warranted to confirm this finding.

Secondly, MCP-1 was another molecule statistically associated with both VEGF protein levels and *VEGF* mRNA. MCP-1 is an angiogenic chemokine, involved in the regulation, migration and infiltration of monocytes and macrophages [60]. Previous research found that MCP-1 induced angiogenesis is mediated by VEGF [61]. More specifically, the researchers identified that MCP-1 up-regulated hypoxia-inducible factor 1 alpha gene expression in human aortic endothelial cells, which as a result induced $VEGF_{165}$ expression in the aortic wall. The importance of MCP-1 in regulation of angiogenesis and immune system was discussed in the study of breast cancer tissue, where twelve inflammatory molecules were measured and evaluated for correlations [62]. This study confirmed that the expression of MCP-1 was correlated significantly with TAM accumulation in primary breast tumours, and with VEGF. Prognostic analysis revealed that high expression of MCP-1/VEGF was an independent indicator of early relapse of cancer. A study in vascular smooth muscle cells of rats [63] demonstrated that VEGF was responsible for the mitogenic activity of MCP-1, which leads to vessel wall remodeling. Moreover, severe hypoxia, previously known to induce VEGF release, potentiated the growth-promoting effect of MCP-1 [63]. This effect on vessel growth was studied also for therapeutic purposes. The therapy with dual delivery of VEGF and MCP-1 showed beneficial effects during transplantation of endothelial cells for therapeutic vascularisation [64], where VEGF improved survival of transplanted cells and MCP-1 induced mural cell recruitment. In the current study, MCP-1 was associated with VEGF protein levels and specific mRNA $VEGF_{165}$ isoform, which confirmed the results of previous studies.

Thirdly, our study identified association of VEGF isoform 165 also with protein IL-6. Several studies demonstrated IL-6 induction of *VEGF* expression. For example, Choen et al. demonstrated that treatment of cell lines with IL-6 results in significant increase of *VEGF* mRNA [65]. Moreover, IL-6 was shown to promote cervical tumorigenesis by activating VEGF-mediated angiogenesis via a STAT3 pathway [66]. Similarly, it has been shown to induce VEGF expression and as a result increase angiogenesis in gastric carcinoma patients [26] and in malignant mesothelioma [67]. These results suggest that treatment with IL-6R antibody might constitute a potential target therapy for certain cancer types. We would, again, reiterate the importance of testing for specific *VEGF* isoforms, which would, in combination with different targeted proteins, such as IL-6, contribute to specific and efficient therapies.

An association was observed also between IL-1α and $VEGF_{165}$ isoform. IL-1α is a prototypical proinflammatory cytokine, known also as hematopoietin 1, which is involved in various immune responses, inflammatory processes, and hematopoiesis [68]. It is produced by monocytes and macrophages and released in response to cell injury [69]. It has been previously shown that IL-1α stimulates VEGF secretion by PBMCs in a dose-depended manner, via induction of *VEGF* mRNA synthesis. Four VEGF isoforms were encoded with *de novo* synthetized mRNA, $VEGF_{121}$, $VEGF_{165}$, $VEGF_{189}$, and $VEGF_{206}$. Moreover, $VEGF_{121}$ and $VEGF_{165}$ gave the major signals both in unstimulated and IL-1α stimulated PBMCs, which is also in favour of our results [21]. Correlation of VEGF production with IL-1α and IL-6 in cell lines suggests that the expression of VEGF is regulated by IL-1α and IL-6 in pancreatic cancer [70] and in adenoma cells [71].

Finally, an association of VEGF protein levels and EGF has been observed in the current study. EGF plays an important role in the growth, proliferation and differentiation of numerous cell types, by binding with high affinity to the cell surface receptor [72]. Major role of both related growth factors in cancer development has been researched for various types of cancers and different mechanism of their common pathways have been proposed [73–75]. Combination of VEGF/EGF signaling pathways has been proposed for new therapeutic approaches of

anticancer therapy [76–79] and is showing positive predictions for future development of new medications.

VEGF isoforms are expressed in different quantities, based on the cell type that is producing them. Most cells express $VEGF_{121}$, $VEGF_{165}$ (the most abundant) and $VEGF_{189}$. Quantification of VEGF isoforms in our study confirmed the reports from literature, as the most expressed isoform from the PBMCs population was $VEGF_{165}$, followed by $VEGF_{121}$. $VEGF_{145}$ was also quantified in our research, although it is more commonly found in cells of placental origin [80, 81]. The notion that the identification of VEGF isoforms is important in fundamental biological and pharmacological studies has arisen the last few years. Specific VEGF isoforms were studied in different types of cancer [82], cardiovascular disease [83], kidney disease [84], autoimmune disease [85] and many others. Pro-angiogenic VEGFxxx and anti-angiogenic VEGFxxxb splicing variants are particularly interesting; it has been demonstrated that their ratio varies based on different pathology, therefore, determination of pro and anti-angiogenic isoforms could serve as diagnostic tools for such diseases [86–88]. Moreover, the quantification of pro and anti-angiogenic isoforms was shown to have a predictive value in response to a treatment with bevacizumab [89]. However, there is no available method that could separate pro and anti-angiogenic isoforms routinely, therefore, most of the studies quantify only pro-angiogenic isoforms. Likewise, in our research, only pro-angiogenic isoforms were detected and associated with specific protein levels. Still, we are aware of the importance of separation and quantification of all VEGF splice variants. We believe that in the future studies, anti-angiogenic isoforms should be routinely quantified and distinguished from pro-angiogenic isoforms for attempting diagnosis of different chronic diseases and prediction of response of anti-VEGF agents (pharmacogenomics).

With such progress, the discovery of anticancer treatments, which will specifically target VEGF isoforms and would decrease the risk of adverse effects of current anti-VEGF therapies will be only a matter of time. VEGF isoforms have already shown a good potential for significant and independent roles in the prediction of different types of cancers [90, 91] and differentiation of isoforms for therapeutic purposes is now being extensively studied [92, 93]. On the other hand, in the complex system of biological pathways, there is usually more than one factor impacting on a pathological development. Therefore, knowing the roles of various proteins related to VEGF is of special importance. It has been shown in previous studies that many cytokines and growth factors upregulate *VEGF* mRNA or induce VEGF release [31]. In support of this, the current investigation also identifies several significant associations.

$VEGF_{189}$ and $VEGF_{165}$ are two isoforms of VEGF protein sharing more than 88% sequence identity. Indeed, analysis of their multiple alignment shows that they differ by only 25 amino acids. These additional amino acids modifying polarity, steric hindering and basic properties confer a subtle difference, sufficient to significantly modify their features and involve them in different signaling pathways. Thus, according to our results, thanks to supplemental residues $VEGF_{189}$ could favor molecules interacting with IL-4 and IL-6 while $VEGF_{165}$ seems to be associated rather with MCP-1 and IL-1α.

One main limitation of our study is that we were unable to confirm the obtained results for VEGFmRNA/protein levels as the availability of biological materials was unfortunately limiting.

Conclusion

This investigation has revealed significant associations between VEGF levels and its specific mRNA isoforms with inflammatory molecules derived from PBMCs cellular extracts, specifically cytokines and growth factors IL-1α, IL4, IL-6, EGF, and MCP1 in a population of healthy

individuals from the SFS. This confirmed that cytokines and growth factors play an important role in VEGF regulation also at lymphocytes cellular extract level. Importantly, specific isoforms interact with different inflammatory molecules. Therefore, they should be identified in future investigations of molecular and biological pathways involving VEGF, which would contribute to an in-depth and specific understanding of the process and would ultimately offer novel targeted therapies for VEGF-related pathophysiology.

Supporting information

S1 Table. Primer sequences for quantification of VEGF transcripts.
(DOCX)

S1 Data. Supporting information—Data set.
(XLSX)

Author Contributions

Conceptualization: Dwaine R. Vance, Maria G. Stathopoulou, Sophie Visvikis-Siest.

Data curation: Helena Murray, John Lamont.

Formal analysis: Vesna Gorenjak, Dwaine R. Vance, Alexandros M. Petrelis, Maria G. Stathopoulou, Sébastien Dadé.

Funding acquisition: Vesna Gorenjak.

Investigation: Dwaine R. Vance.

Methodology: Maria G. Stathopoulou, Sébastien Dadé, Said El Shamieh.

Project administration: Vesna Gorenjak, Sophie Visvikis-Siest.

Resources: Helena Murray, Christine Masson, John Lamont, Peter Fitzgerald.

Supervision: Maria G. Stathopoulou, Sophie Visvikis-Siest.

Validation: Vesna Gorenjak, Dwaine R. Vance, Sophie Visvikis-Siest.

Visualization: Vesna Gorenjak.

Writing – original draft: Vesna Gorenjak, Dwaine R. Vance, Sébastien Dadé.

Writing – review & editing: Vesna Gorenjak, Alexandros M. Petrelis, Maria G. Stathopoulou, Said El Shamieh, Sophie Visvikis-Siest.

References

1. Ferrara N, Gerber HP, LeCouter J. The biology of VEGF and its receptors. Nat Med. 2003; 9(6):669–76. https://doi.org/10.1038/nm0603-669 PMID: 12778165
2. Zhang W, Shen Z, Luo H, Hu X, Zheng L, Zhu X. The Benefits and Side Effects of Bevacizumab for the Treatment of Recurrent Ovarian Cancer. Curr Drug Targets. 2017; 18(10):1125–31. https://doi.org/10.2174/1389450117666160502150237 PMID: 27138763
3. Riondino S, Monte GD, Fratangeli F, Guadagni F, Roselli M, Ferroni P. Anti-angiogenic drugs, vascular toxicity and thromboembolism in solid cancer. Cardiovasc Hematol Agents Med Chem. 2017.
4. Berrahmoune H, Herbeth B, Lamont JV, Masson C, Fitzgerald PS, Visvikis-Siest S. Heritability for plasma VEGF concentration in the Stanislas family study. Ann Hum Genet. 2007; 71(Pt 1):54–63. https://doi.org/10.1111/j.1469-1809.2006.00298.x PMID: 17227476
5. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, et al. Identification of cis- and trans-acting genetic variants explaining up to half the variation in circulating vascular endothelial growth

factor levels. Circ Res. 2011; 109(5):554–63. https://doi.org/10.1161/CIRCRESAHA.111.243790 PMID: 21757650

6. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, et al. Six Novel Loci Associated with Circulating VEGF Levels Identified by a Meta-analysis of Genome-Wide Association Studies. PLoS Genet. 2016; 12(2):e1005874. https://doi.org/10.1371/journal.pgen.1005874 PMID: 26910538
7. Vempati P, Popel AS, Mac Gabhann F. Extracellular regulation of VEGF: isoforms, proteolysis, and vascular patterning. Cytokine Growth Factor Rev. 2014; 25(1):1–19. https://doi.org/10.1016/j.cytogfr.2013.11.002 PMID: 24332926
8. Harper SJ, Bates DO. VEGF-A splicing: the key to anti-angiogenic therapeutics? Nature reviews Cancer. 2008; 8(11):880–7. https://doi.org/10.1038/nrc2505 PMID: 18923433
9. Woolard J, Bevan HS, Harper SJ, Bates DO. Molecular diversity of VEGF-A as a regulator of its biological activity. Microcirculation. 2009; 16(7):572–92. https://doi.org/10.1080/10739680902997333 PMID: 19521900
10. Nowak DG, Amin EM, Rennel ES, Hoareau-Aveilla C, Gammons M, Damodoran G, et al. Regulation of vascular endothelial growth factor (VEGF) splicing from pro-angiogenic to anti-angiogenic isoforms: a novel therapeutic strategy for angiogenesis. J Biol Chem. 2010; 285(8):5532–40. https://doi.org/10.1074/jbc.M109.074930 PMID: 19906640
11. Harper SJ, Bates DO. VEGF-A splicing: the key to anti-angiogenic therapeutics? Nat Rev Cancer. 2008; 8(11):880–7. https://doi.org/10.1038/nrc2505 PMID: 18923433
12. Mor F, Quintana FJ, Cohen IR. Angiogenesis-inflammation cross-talk: vascular endothelial growth factor is secreted by activated T cells and induces Th1 polarization. J Immunol. 2004; 172(7):4618–23. https://doi.org/10.4049/jimmunol.172.7.4618 PMID: 15034080
13. Kieran MW, Kalluri R, Cho YJ. The VEGF Pathway in Cancer and Disease: Responses, Resistance, and the Path Forward. Cold Spring Harb Perspect Med. 2012; 2(12).
14. Corkum CP, Ings DP, Burgess C, Karwowska S, Kroll W, Michalak TI. Immune cell subsets and their gene expression profiles from human PBMC isolated by Vacutainer Cell Preparation Tube (CPT()) and standard density gradient. BMC Immunol. 2015; 16.
15. Lackie JM. The Dictionary of Cell and Molecular Biology: Elsevier Science; 2012.
16. Miyahira A. Sanguine Bio Researcher Blog [Internet]. PBMC Basic: Sanguine Biosciences. 2012. [cited 2018]. https://technical.sanguinebio.com/types-of-immune-cells-present-in-human-pbmc/.
17. Angelo LS, Kurzrock R. Vascular endothelial growth factor and its relationship to inflammatory mediators. Clin Cancer Res. 2007; 13(10):2825–30. https://doi.org/10.1158/1078-0432.CCR-06-2416 PMID: 17504979
18. Turner MD, Nedjai B, Hurst T, Pennington DJ. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. Biochim Biophys Acta. 2014; 1843(11):2563–82. https://doi.org/10.1016/j.bbamcr.2014.05.014 PMID: 24892271
19. Paul WE, Seder RA. Lymphocyte responses and cytokines. Cell. 1994; 76(2):241–51. https://doi.org/10.1016/0092-8674(94)90332-8 PMID: 7904900
20. Krieglstein CF, Granger DN. Adhesion molecules and their role in vascular disease. Am J Hypertens. 2001; 14(6 Pt 2):44s–54s. https://doi.org/10.1016/s0895-7061(01)02069-6 PMID: 11411765
21. Salven P, Hattori K, Heissig B, Rafii S. Interleukin-1alpha promotes angiogenesis in vivo via VEGFR-2 pathway by inducing inflammatory cell VEGF synthesis and secretion. FASEB J. 2002; 16(11):1471–3. https://doi.org/10.1096/fj.02-0134fje PMID: 12205052
22. Repovic P, Fears CY, Gladson CL, Benveniste EN. Oncostatin-M induction of vascular endothelial growth factor expression in astroglioma cells. Oncogene. 2003; 22(50):8117–24. https://doi.org/10.1038/sj.onc.1206922 PMID: 14603252
23. De Palma M. Partners in crime: VEGF and IL-4 conscript tumour-promoting macrophages. J Pathol. 2012; 227(1):4–7. https://doi.org/10.1002/path.4008 PMID: 22367978
24. Hong K, Cho M, Min S, Shin Y, Yoo S, Choi J, et al. Effect of interleukin-4 on vascular endothelial growth factor production in rheumatoid synovial fibroblasts. Clin Exp Immunol. 2007; 147(3):573–9. https://doi.org/10.1111/j.1365-2249.2006.03295.x PMID: 17302909
25. Faffe DS, Flynt L, Bourgeois K, Panettieri RA, Shore SA. Interleukin-13 and Interleukin-4 Induce Vascular Endothelial Growth Factor Release from Airway Smooth Muscle Cells: Role of Vascular Endothelial Growth Factor Genotype. Am J Respir Cell Mol Biol. 2006; 34(2):213–8. https://doi.org/10.1165/rcmb.2005-0147OC PMID: 16210693
26. Huang SP, Wu MS, Shun CT, Wang HP, Lin MT, Kuo ML, et al. Interleukin-6 increases vascular endothelial growth factor and angiogenesis in gastric carcinoma. J Biomed Sci. 2004; 11(4):517–27. https://doi.org/10.1007/bf02256101 PMID: 15153787

27. Loeffler S, Fayard B, Weis J, Weissenberger J. Interleukin-6 induces transcriptional activation of vascular endothelial growth factor (VEGF) in astrocytes in vivo and regulates VEGF promoter activity in glioblastoma cells via direct interaction between STAT3 and Sp1. Int J Cancer. 2005; 115(2):202–13. https://doi.org/10.1002/ijc.20871 PMID: 15688401

28. Huang S, Robinson JB, DeGuzman A, Bucana CD, Fidler IJ. Blockade of nuclear factor-κB signaling inhibits angiogenesis and tumorigenicity of human ovarian cancer cells by suppressing expression of vascular endothelial growth factor and interleukin 8. Cancer Res. 2000; 60(19):5334–9. PMID: 11034066

29. Lau TP, Roslani AC, Lian LH, Lee PC, Hilmi I, Goh KL, et al. Association between EGF and VEGF functional polymorphisms and sporadic colorectal cancer in the Malaysian population. Genet Mol Res. 2014; 13(3):5555–61. https://doi.org/10.4238/2014.July.25.9 PMID: 25117311

30. Lequoy P, Murschel F, Liberelle B, Lerouge S, De Crescenzo G. Controlled co-immobilization of EGF and VEGF to optimize vascular cell survival. Acta Biomater. 2016; 29:239–47. https://doi.org/10.1016/j.actbio.2015.10.026 PMID: 26485166

31. Robinson CJ, Stringer SE. The splice variants of vascular endothelial growth factor (VEGF) and their receptors. J Cell Sci. 2001; 114(Pt 5):853–65. PMID: 11181169

32. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med. 2008; 46(6):733–47. https://doi.org/10.1515/CCLM.2008.178 PMID: 18601594

33. Siest G, Visvikis S, Herbeth B, Gueguen R, Vincent-Viry M, Sass C, et al. Objectives, design and recruitment of a familial and longitudinal cohort for studying gene-environment interactions in the field of cardiovascular risk: the Stanislas cohort. Clin Chem Lab Med. 1998; 36(1):35–42. https://doi.org/10.1515/CCLM.1998.007 PMID: 9594084

34. Boyum A. Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. Scand J Clin Lab Invest Suppl. 1968; 97:77–89. PMID: 4179068

35. Azimi-Nezhad M, Stathopoulou MG, Bonnefond A, Rancier M, Saleh A, Lamont J, et al. Associations of vascular endothelial growth factor (VEGF) with adhesion and inflammation molecules in a healthy population. Cytokine. 2013; 61(2):602–7. https://doi.org/10.1016/j.cyto.2012.10.024 PMID: 23201487

36. Marteau JB, Mohr S, Pfister M, Visvikis-Siest S. Collection and storage of human blood cells for mRNA expression profiling: a 15-month stability study. Clin Chem. 2005; 51(7):1250–2. https://doi.org/10.1373/clinchem.2005.048546 PMID: 15976103

37. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005; 21(7):951–60. https://doi.org/10.1093/bioinformatics/bti125 PMID: 15531603

38. Cai C, Bottcher MC, Werner JA, Mandic R. Differential expression of VEGF121, VEGF165 and VEGF189 in angiomas and squamous cell carcinoma cell lines of the head and neck. Anticancer Res. 2010; 30(3):805–10. PMID: 20393000

39. Keyt BA, Berleau LT, Nguyen HV, Chen H, Heinsohn H, Vandlen R, et al. The carboxyl-terminal domain (111–165) of vascular endothelial growth factor is critical for its mitogenic potency. J Biol Chem. 1996; 271(13):7788–95. https://doi.org/10.1074/jbc.271.13.7788 PMID: 8631822

40. Goodwin JS, Webb DR. Regulation of the immune response by prostaglandins. Clin Immunol Immunopathol. 1980; 15(1):106–22. PMID: 6987016

41. Aggarwal S, Pittenger MF. Human mesenchymal stem cells modulate allogeneic immune cell responses. Blood. 2005; 105(4):1815–22. https://doi.org/10.1182/blood-2004-04-1559 PMID: 15494428

42. Saurwein-Teissl M, Zisterer K, Schmitt T, Glück R, Cryz S, Grubeck-Loebenstein B. Whole virus influenza vaccine activates dendritic cells (DC) and stimulates cytokine production by peripheral blood mononuclear cells (PBMC) while subunit vaccines support T cell proliferation. Clin Exp Immunol. 1998; 114(2):271. https://doi.org/10.1046/j.1365-2249.1998.00714.x PMID: 9822287

43. McElhaney JE, Xie D, Hager WD, Barry MB, Wang Y, Kleppinger A, et al. T cell responses are better correlates of vaccine protection in the elderly. The Journal of Immunology. 2006; 176(10):6333–9. https://doi.org/10.4049/jimmunol.176.10.6333 PMID: 16670345

44. Langjahr M, Schubert AL, Sommer C, Uceyler N. Increased pro-inflammatory cytokine gene expression in peripheral blood mononuclear cells of patients with polyneuropathies. J Neurol. 2018; 265(3):618–27. https://doi.org/10.1007/s00415-018-8748-4 PMID: 29372388

45. Boiardi L, Casali B, Nicoli D, Farnetti E, Chen Q, Macchioni P, et al. Vascular endothelial growth factor gene polymorphisms in giant cell arteritis. J Rheumatol. 2003; 30(10):2160–4. PMID: 14528511

46. Iacobaeus E, Amoudruz P, Strom M, Khademi M, Brundin L, Hillert J, et al. The expression of VEGF-A is down regulated in peripheral blood mononuclear cells of patients with secondary progressive multiple sclerosis. PLoS One. 2011; 6(5):e19138. https://doi.org/10.1371/journal.pone.0019138 PMID: 21573104

47. Saito K, Nakaoka H, Takasaki I, Hirono K, Yamamoto S, Kinoshita K, et al. MicroRNA-93 may control vascular endothelial growth factor A in circulating peripheral blood mononuclear cells in acute Kawasaki disease. Pediatr Res. 2016; 80(3):425–32. https://doi.org/10.1038/pr.2016.93 PMID: 27089500

48. Hoekstra M, van der Lans CA, Halvorsen B, Gullestad L, Kuiper J, Aukrust P, et al. The peripheral blood mononuclear cell microRNA signature of coronary artery disease. Biochem Biophys Res Commun. 2010; 394(3):792–7. https://doi.org/10.1016/j.bbrc.2010.03.075 PMID: 20230787

49. Aziz H, Zaas A, Ginsburg GS. Peripheral blood gene expression profiling for cardiovascular disease assessment. Genomic Med. 2007; 1(3):105–12.

50. Shah VO, Dorin RI, Sun Y, Braun M, Zager PG. Aldose Reductase Gene Expression Is Increased in Diabetic Nephropathy1. The Journal of Clinical Endocrinology & Metabolism. 1997; 82(7):2294–8.

51. Giannopoulou EG, Elemento O, Ivashkiv LB. Use of RNA sequencing to evaluate rheumatic disease patients. Arthritis Res Ther. 2015; 17(1):167.

52. Nikolayeva I, Bost P, Casademont I, Duong V, Koeth F, Prot M, et al. A Blood RNA Signature Detecting Severe Disease in Young Dengue Patients at Hospital Arrival. J Infect Dis. 2018.

53. Erusan RR, Nalini D, Manohar G, Malathi R. Correlation between obesity and inflammation in cardiovascular diseases—evaluation of leptin and inflammatory cytokines. Open Journal of Endocrine and Metabolic Diseases. 2012; 2(02):7.

54. Nakagomi H, Pisa P, Pisa EK, Yamamoto Y, Halapi E, Backlin K, et al. Lack of interleukin-2 (IL-2) expression and selective expression of IL-10 mRNA in human renal cell carcinoma. Int J Cancer. 1995; 63(3):366–71. https://doi.org/10.1002/ijc.2910630311 PMID: 7591233

55. Hotamisligil GS. Inflammation and metabolic disorders. Nature. 2006; 444(7121):860. https://doi.org/10.1038/nature05485 PMID: 17167474

56. Hong KH, Cho ML, Min SY, Shin YJ, Yoo SA, Choi JJ, et al. Effect of interleukin-4 on vascular endothelial growth factor production in rheumatoid synovial fibroblasts. Clin Exp Immunol. 2007; 147(3):573–9. https://doi.org/10.1111/j.1365-2249.2006.03295.x PMID: 17302909

57. Hart PH, Vitti GF, Burgess DR, Whitty GA, Piccoli DS, Hamilton JA. Potential antiinflammatory effects of interleukin 4: suppression of human monocyte tumor necrosis factor alpha, interleukin 1, and prostaglandin E2. Proc Natl Acad Sci U S A. 1989; 86(10):3803–7. https://doi.org/10.1073/pnas.86.10.3803 PMID: 2786204

58. Linde N, Lederle W, Depner S, van Rooijen N, Gutschalk CM, Mueller MM. Vascular endothelial growth factor-induced skin carcinogenesis depends on recruitment and alternative activation of macrophages. J Pathol. 2012; 227(1):17–28. https://doi.org/10.1002/path.3989 PMID: 22262122

59. Skaria T, Burgener J, Bachli E, Schoedon G. IL-4 Causes Hyperpermeability of Vascular Endothelial Cells through Wnt5A Signaling. PLoS One. 2016; 11(5).

60. Deshmane SL, Kremlev S, Amini S, Sawaya BE. Monocyte Chemoattractant Protein-1 (MCP-1): An Overview. J Interferon Cytokine Res. 2009; 29(6):313–26. https://doi.org/10.1089/jir.2008.0027 PMID: 19441883

61. Hong KH, Ryu J, Han KH. Monocyte chemoattractant protein-1-induced angiogenesis is mediated by vascular endothelial growth factor-A. Blood. 2005; 105(4):1405–7. https://doi.org/10.1182/blood-2004-08-3178 PMID: 15498848

62. Ueno T, Toi M, Saji H, Muta M, Bando H, Kuroi K, et al. Significance of Macrophage Chemoattractant Protein-1 in Macrophage Recruitment, Angiogenesis, and Survival in Human Breast Cancer. Clin Cancer Res. 2000; 6(8):3282–9. PMID: 10955814

63. Parenti A, Bellik L, Brogelli L, Filippi S, Ledda F. Endogenous VEGF-A is responsible for mitogenic effects of MCP-1 on vascular smooth muscle cells. American Journal of Physiology-Heart and Circulatory Physiology. 2004; 286(5):H1978–H84. https://doi.org/10.1152/ajpheart.00414.2003 PMID: 14693680

64. Jay SM, Shepherd BR, Andrejecsk JW, Kyriakides TR, Pober JS, Saltzman WM. Dual delivery of VEGF and MCP-1 to support endothelial cell transplantation for therapeutic vascularization. Biomaterials. 2010; 31(11):3054–62. https://doi.org/10.1016/j.biomaterials.2010.01.014 PMID: 20110124

65. Cohen T, Nahari D, Cerem LW, Neufeld G, Levi BZ. Interleukin 6 induces the expression of vascular endothelial growth factor. J Biol Chem. 1996; 271(2):736–41. https://doi.org/10.1074/jbc.271.2.736 PMID: 8557680

66. Wei LH, Kuo ML, Chen CA, Chou CH, Lai KB, Lee CN, et al. Interleukin-6 promotes cervical tumor growth by VEGF-dependent angiogenesis via a STAT3 pathway. Oncogene. 2003; 22(10):1517–27. https://doi.org/10.1038/sj.onc.1206226 PMID: 12629515

67. Adachi Y, Aoki C, Yoshio-Hoshino N, Takayama K, Curiel DT, Nishimoto N. Interleukin-6 induces both cell growth and VEGF production in malignant mesotheliomas. Int J Cancer. 2006; 119(6):1303–11. https://doi.org/10.1002/ijc.22006 PMID: 16642474

68. Dembic Z. The Cytokines of the Immune System: The Role of Cytokines in Disease Related to Immune Response: Elsevier Science; 2015.

69. Dinarello CA. Biologic basis for interleukin-1 in disease. Blood. 1996; 87(6):2095–147. PMID: 8630372

70. Tang RF, Wang SX, Zhang FR, Peng L, Wang SX, Xiao Y, et al. Interleukin-1alpha, 6 regulate the secretion of vascular endothelial growth factor A, C in pancreatic cancer. Hepatobiliary Pancreat Dis Int. 2005; 4(3):460–3. PMID: 16109537

71. Borg SA, Kerry KE, Royds JA, Battersby RD, Jones TH. Correlation of VEGF production with IL1 alpha and IL6 secretion by human pituitary adenoma cells. Eur J Endocrinol. 2005; 152(2):293–300. https://doi.org/10.1530/eje.1.01843 PMID: 15745939

72. Carpenter G, Cohen S. Epidermal growth factor. J Biol Chem. 1990; 265(14):7709–12. PMID: 2186024

73. Saryeddine L, Zibara K, Kassem N, Badran B, El-Zein N. EGF-Induced VEGF Exerts a PI3K-Dependent Positive Feedback on ERK and AKT through VEGFR2 in Hematological In Vitro Models. PLoS One. 2016; 11(11):e0165876. https://doi.org/10.1371/journal.pone.0165876 PMID: 27806094

74. Witsch E, Sela M, Yarden Y. Roles for growth factors in cancer progression. Physiology (Bethesda). 2010; 25(2):85–101.

75. Citri A, Yarden Y. EGF–ERBB signalling: towards the systems level. Nature Reviews Molecular Cell Biology. 2006; 7:505. https://doi.org/10.1038/nrm1962 PMID: 16829981

76. Ciardiello F, Troiani T, Bianco R, Orditura M, Morgillo F, Martinelli E, et al. Interaction between the epidermal growth factor receptor (EGFR) and the vascular endothelial growth factor (VEGF) pathways: a rational approach for multi-target anticancer therapy. Ann Oncol. 2006; 17 Suppl 7:vii109–14.

77. Tortora G, Ciardiello F, Gasparini G. Combined targeting of EGFR-dependent and VEGF-dependent pathways: rationale, preclinical studies and clinical applications. Nat Clin Pract Oncol. 2008; 5(9):521–30. https://doi.org/10.1038/ncponc1161 PMID: 18594498

78. Tabernero J. The role of VEGF and EGFR inhibition: implications for combining anti-VEGF and anti-EGFR agents. Mol Cancer Res. 2007; 5(3):203–20. https://doi.org/10.1158/1541-7786.MCR-06-0404 PMID: 17374728

79. Winder T, Lenz HJ. Vascular endothelial growth factor and epidermal growth factor signaling pathways as therapeutic targets for colorectal cancer. Gastroenterology. 2010; 138(6):2163–76. https://doi.org/10.1053/j.gastro.2010.02.005 PMID: 20420953

80. Cao Y, Ji WR, Qi P, Rosin A, Cao Y. Placenta growth factor: identification and characterization of a novel isoform generated by RNA alternative splicing. Biochem Biophys Res Commun. 1997; 235 (3):493–8. https://doi.org/10.1006/bbrc.1997.6813 PMID: 9207183

81. Cressey R, Wattananupong O, Lertprasertsuke N, Vinitketkumnuen U. Alteration of protein expression pattern of vascular endothelial growth factor (VEGF) from soluble to cell-associated isoform during tumourigenesis. BMC Cancer. 2005; 5:128. https://doi.org/10.1186/1471-2407-5-128 PMID: 16202150

82. Canavese M, Ngo DT, Maddern GJ, Hardingham JE, Price TJ, Hauben E. Biology and therapeutic implications of VEGF-A splice isoforms and single-nucleotide polymorphisms in colorectal cancer. Int J Cancer. 2017; 140(10):2183–91. https://doi.org/10.1002/ijc.30567 PMID: 27943279

83. Kikuchi R. An anti-angiogenic isoform of VEGF-A contributes to impaired vascularization in peripheral artery disease. 2014; 20(12):1464–71.

84. Stevens M, Oltean S. Modulation of VEGF-A Alternative Splicing as a Novel Treatment in Chronic Kidney Disease. Genes (Basel). 2018; 9(2).

85. Manetti M, Guiducci S, Romano E, Ceccarelli C, Bellando-Randone S, Conforti ML, et al. Overexpression of VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, leads to insufficient angiogenesis in patients with systemic sclerosis. Circ Res. 2011; 109(3):e14–26. https://doi.org/10.1161/CIRCRESAHA.111.242057 PMID: 21636803

86. Barratt SL, Blythe T, Jarrett C, Ourradi K, Shelley-Fraser G, Day MJ, et al. Differential Expression of VEGF-Axxx Isoforms Is Critical for Development of Pulmonary Fibrosis. Am J Respir Crit Care Med. 2017; 196(4):479–93. https://doi.org/10.1164/rccm.201603-0568OC PMID: 28661183

87. Bates DO, Cui TG, Doughty JM, Winkler M, Sugiono M, Shields JD, et al. VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, is down-regulated in renal cell carcinoma. Cancer Res. 2002; 62(14):4123–31. PMID: 12124351

88. Bunni J, Shelley-Fraser G, Stevenson K, Oltean S, Salmon A, Harper SJ, et al. Circulating levels of anti-angiogenic VEGF-A isoform (VEGF-Axxxb) in colorectal cancer patients predicts tumour VEGF-A ratios. Am J Cancer Res. 2015; 5(6):2083–9. PMID: 26269767

89. Varey AH, Rennel ES, Qiu Y, Bevan HS, Perrin RM, Raffy S, et al. VEGF 165 b, an antiangiogenic VEGF-A isoform, binds and inhibits bevacizumab treatment in experimental colorectal carcinoma: balance of pro- and antiangiogenic VEGF-A isoforms has implications for therapy. Br J Cancer. 2008; 98 (8):1366–79. https://doi.org/10.1038/sj.bjc.6604308 PMID: 18349829

90. Jeng KS, Sheen IS, Wang YC, Gu SL, Chu CM, Shih SC, et al. Is the vascular endothelial growth factor messenger RNA expression in resectable hepatocellular carcinoma of prognostic value after resection? World J Gastroenterol. 2004; 10(5):676–81. https://doi.org/10.3748/wjg.v10.i5.676 PMID: 14991937

91. Paule B, Bastien L, Deslandes E, Cussenot O, Podgorniak MP, Allory Y, et al. Soluble isoforms of vascular endothelial growth factor are predictors of response to sunitinib in metastatic renal cell carcinomas. PLoS One. 2010; 5(5):e10715. https://doi.org/10.1371/journal.pone.0010715 PMID: 20502715

92. Clegg LE, Mac Gabhann F. A computational analysis of pro-angiogenic therapies for peripheral artery disease. Integr Biol (Camb). 2018; 10(1):18–33.

93. Guyot M, Hilmi C, Ambrosetti D, Merlano M, Lo Nigro C, Durivault J, et al. Targeting the pro-angiogenic forms of VEGF or inhibiting their expression as anti-cancer strategies. Oncotarget. 2017; 8(6):9174–88. https://doi.org/10.18632/oncotarget.13942 PMID: 27999187

Publication Nº 6

The polymorphism rs6918289 located in the downstream region of the *TREM2* gene is associated with TNF-alpha levels and IMT-F.

Vesna Gorenjak[*], Alex-Ander Aldasoro Arguinano[*], Sébastien Dade, Maria G. Stathopoulou, Dwaine R. Vance, Christine Masson and Sophie Visvikis-Siest

[*]equal contribution

Sci Rep, May 2018, 8;8(1):7160.

TREM2 is a cell surface receptor of the immunoglobulin superfamily, primarily expressed on myeloid cells and macrophages, which exposes important anti-inflammatory characteristics via inhibition of the cytokine production by macrophages (78). It is particularly known for its protective role in neurodegenerative diseases, where it prevents neuronal damages, caused by an inflammatory response *via* repression of microglia-mediated cytokine production (70). Its association with increased risk for AD was demonstrated in large GWAS, which identified heterozygous rare genetic variants of *TREM2* significantly associated with AD (302, 303).

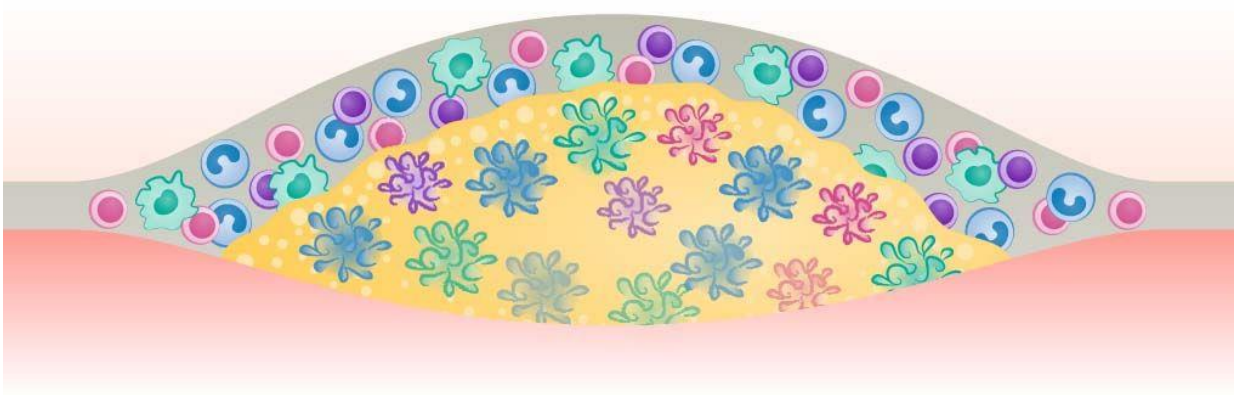


Figure 32: Cytokine and macrophage infiltration in atherosclerotic lesions: TREM2 high macrophages were detected in advanced atherosclerosis and *TREM2* expression was found in human atherosclerotic lesions (69).

TREM2 was previously associated with diverse inflammatory molecules which are known mediators of physiological and pathophysiological processes, thus with decreased infiltration of TNF-α and IFN-γ, but with increased secretion of IL-4 and IL-13 (45). Recent studies showed that *TREM2* expression was found in human atherosclerotic lesions (Figure 32), suggesting its involvement in the control of macrophage-mediated inflammation in atherosclerosis (46).

TNF-α is a key proinflammatory cytokine with an important role in CVD (74). Its role in the formation of atherosclerotic plaques has been previously demonstrated (304), however, no relation between TNF-α and TREM2 has been reported up to date.

Objective

The goal of this study was to investigate the association of genetic variants located in *TREM2* gene region with TNF-α and intima-media thickness of the femoral artery (IMT-F), in order to examine the role of TREM2 in the development of atherosclerotic plaques.

Populations and analysis

TNF-α

Discovery cohort consisted of 139 children, recruited in the SFS, where genotype information for the *TREM2* gene region was readily available through the Infinium CoreExome Illumina GWAS assay. Subsequently, 393 adults and 277 additional children were genotyped for the significant SNPs identified in the initial children sample and were included in the analysis. *De novo* genotyping was conducted by the Laboratory of the Government Chemist (LGC), using a PCR-based KASP assay. The association between the SNPs and TNF- α were tested through linear regression adjusted for age, gender and BMI. Analyses were performed using the R package GWAF (Genome-Wide Association/Interaction Analysis and Rare Variant analysis with Family Data), taking into account familial resemblance.

The replication cohort included 916 unrelated adults of French origin. *De novo* genotyping and analysis performed in the replication cohort corresponded to discovery cohort. A meta-

analysis of discovery and replication cohort was performed using the GWAMA software and a random-effect method.

TNF-α of discovery and replication cohort was measured in plasma by a commercially available ELISA (R&D Systems, UK), according to manufacturer instructions.

IMT-F

SNPs significantly associated with TNF-α in the analysis of the children cohort were tested also for association with IMT-F, in 350 adults from SFS, where measurements of IMT-F were available. IMT-F was measured in the right and left femoral arteries, examined with a 7.5 MHz probe, according to a protocol previously described (305). For each individual, two IMT-F measurements were obtained and right and left measurements were used to calculate the mean IMT-F (in mm).

Results

Five SNPs (rs7748777, rs6918289, rs7759295, rs9357347, rs6915083) located in the *TREM2* gene region (6p21) were tested for the association with TNF-α in 139 children from the SFS (Table 8).

Table 8. Association of the SNPs in *TREM2* region with TNF-α levels in children of the SFS.

Polymorphism	Model	P-Value	β	SE
rs7748777	Additive	0.04014	0.14	0.07
rs6918289	Additive	0.00147*	0.33	0.10
rs7759295	Additive	0.9966	-0.0005	0.11
rs9357347	Additive	0.02127	0.17	0.075
rs6915083	Additive	0.1634	0.1064	0.076

SE: Standard error, **β:** Effect size, P-value threshold is P<0.01, *significant P-value

The minor allele (T) of rs6918289 was significantly associated with increased concentration of TNF-α (p=0.0015). To increase the discovery population, rs6918289 was *de novo* genotyped in 393 adults and 277 additional children.

Because the underlying model of the genetic determinant of a trait was not known, in all analysis, three classical genetic models were used to test the association of the minor allele

(T) of rs6918289 with the phenotypes of interest. These models are grouping the alleles of the gene of interest in three groups: TT, GG (homozygote genotypes) and TG (heterozygote genotype). As T is the 'risk' allele, the genetic models can be presented as follows (306):

- Dominant: 'TT + TG' versus 'GG',
- Recessive: 'TT' versus 'TG + GG',
- Additive: 'TT' versus 'TG' versus 'GG'

Statistical analysis was performed in the combined SFS population, comprising 809 individuals (Table 10). Additive and recessive models showed a significant association between rs6918289 and TNF-α (p=0.0026 and p=0.0017, respectively).

Association analysis was repeated in the independent French population (n=916), where rs6918289 showed a positive association with TNF-α in the recessive model of the analysis (p=0.023) (Table 9).

Table 9. Association of rs6918289 with TNF-α levels in the discovery and replication populations.

Population	Model	N	β	SE	P-value
The SFS	Additive	809	0.13	0.04	0.0026*
	Dominant	809	0.11	0.05	0.022*
	Recessive	809	0.49	0.15	0.0017*
Replication	Additive	916	0.042	0.02	0.073
	Dominant	916	0.03	0.03	0.189
	Recessive	916	0.20	0.09	0.023*

SE: Standard error, **β:** Effect size, P-value threshold is P<0.01, *significant P-value

The meta-analysis of both cohorts showed a marginal association of rs6918289 with TNF-α using the additive model (p=0.072) and significant association using the recessive model (p=0.0003).

The minor allele (T) of rs6918289 was significantly associated with increased IMT-F in a population of adults from the SFS. Association was significant using the additive (p=0.026) and dominant (p=0.026) genetic model (Table 10).

Table 10. Association of rs6918289 with IMT-F in the discovery population.

Population	Model	N	β	SE	P-value
The SFS	Additive	350	0.02	0.009	0.026*
	Dominant	350	0.024	0.01	0.026*
	Recessive	350	0.024	0.025	0.34

SE: Standard error, **β:** Effect size, P-value threshold is P<0.01, *significant P-value

Conclusion

The study presented in the publication '*The polymorphism rs6918289 located in the downstream region of the TREM2 gene' is associated with TNF-α and IMT-F'* highlighted a new potential risk allele for inflammation and atherosclerosis, which affects the blood levels of TNF-α and is associated with increased IMT-F. These findings support the results of the previous studies linking variants within chromosomal 6p21.1 loci to atherosclerosis and are raising the awareness to consider genomic region 6p21.1 as a candidate susceptibility loci for atherosclerosis.

The five investigated genetic variants were chosen based on their proximity to the *TREM2* gene and were not based on the previous association identified through GWAS studies. The reason for such a decision is that not all SNPs were available in our genotyped population. We have chosen the genetic variants based on the predisposition that they are involved in the regulation of the *TREM2* gene. A new study, where genome-wide identified *TREM2* variants would be used for the association studies could be useful for the further exploration of the relations between TREM2, TNF-α and IMT-F. Nevertheless, a study with bigger sample size and replication of IMT-F results are needed to confirm our results and consider their further clinical utility.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

The polymorphism rs6918289 located in the downstream region of the *TREM2* gene is associated with TNF-α levels and IMT-F

Received: 25 October 2017
Accepted: 24 April 2018
Published online: 08 May 2018

Vesna Gorenjak[1], Alex-Ander Aldasoro Arguinano[1], Sébastien Dadé[1], Maria G. Stathopoulou[1], Dwaine R. Vance[2], Christine Masson[1] & Sophie Visvikis-Siest[1,3]

Triggering receptor expressed on myeloid cells 2 (TREM2) is known for its anti-inflammatory properties during the immune response, and influences negatively on TNF-α expression levels. Genetic epidemiology studies have identified polymorphisms located in the *TREM2* gene associated with neurodegenerative and chronic inflammatory diseases. TREM2 levels have been observed to affect plasma levels of TNF-α and plaque stability in symptomatic and asymptomatic patients with carotid stenosis. In this study, we investigated polymorphisms located in the *TREM2* gene region and association with TNF-α levels and the intima media thickness of the femoral artery. The discovery population from the STANISLAS Family Study comprised of 809 individuals, whereas the replication population utilized an independent cohort of French origin ($n = 916$). Our results suggest that the minor allele (T) of SNP rs6918289 is positively associated with elevated plasma levels of TNF-α in discovery and replication populations ($P = 0.0026$, $SE = 0.04$ and $P = 0.023$, $SE = 0.09$, respectively), including femoral artery thickness in the discovery cohort ($P = 0.026$, $SE = 0.009$). Results indicate that rs6918289 may be considered as a risk factor for inflammatory diseases and could be used in stratified medicine with patients diagnosed with chronic inflammatory-related conditions, such as atherosclerosis.

The triggering receptors expressed on myeloid cells (TREM) family molecules are members of the immunoglobulin superfamily of receptors. All five genes from the TREM family (Table 1) are situated in the 6p21.1 region of the chromosome[1] and mediate signaling in immune cells, thus playing critical roles in inflammatory responses[2]. The region 6p21.1 is in proximity to the MHC/HLA region of the genome, which is implicated in the immune response, autoimmunity and risk of autoimmune diseases[3,4]. Specifically, the TREM2 molecule is primarily expressed on the cell surface of macrophages and dendritic cells derived from monocytes, as well as in microglia and osteoclasts. TREM2 binds to the DAP12 trans-membrane molecule and is responsible for a series of tyrosine phosphorylation reactions that regulate various inflammatory responses[2,5]. TREM2 portrays anti-inflammatory properties during the immune response[6,7], including: stimulation of phagocytosis and suppression of cytokine production, *e.g.* TNF-α[8,9], one of the most important molecules for the regulation of inflammation, and reflects the degree of inflammatory response.

It is evident that TREM2 acts as a protective molecule in chronic inflammatory diseases. For example, studies in transgenic mice have demonstrated that deficiency of TREM2 protein may accelerate the aging process, reduce microglial activity and result in neuroinflammation, which plays a major role in all neurodegenerative diseases[10]. Furthermore, atherosclerosis is a chronic cardiovascular inflammatory disease, caused by activation of the immune system mediating the chronic inflammatory process of the arterial wall. Interestingly, TREM2 has also been shown to play an important role in the stability of atherosclerotic plaques[11].

Polymorphisms located in the *TREM2* gene have been correlated with neurodegenerative and chronic inflammatory diseases, such as Alzheimer's disease[12,13], frontotemporal dementia[14,15], Parkinson's disease[16,17],

[1]Université de Lorraine, Inserm, IGE-PCV, F-54000, Nancy, France. [2]Randox Laboratories Limited, Crumlin, Co. Antrim, Northern Ireland, United Kingdom. [3]Department of Internal Medicine and Geriatrics, CHU Technopôle Nancy-Brabois, Rue du Morvan, F-54511, Vandoeuvre-lès-Nancy, France. Vesna Gorenjak and Alex-Ander Aldasoro Arguinano contributed equally to this work. Correspondence and requests for materials should be addressed to S.V.-S. (email: sophie.visvikis-siest@inserm.fr)

Gene	Gene name	Chromosome
TREML1	triggering receptor expressed on myeloid cells like 1	6p21.1
TREML2	triggering receptor expressed on myeloid cells like 2	6p21.1
TREML4	triggering receptor expressed on myeloid cells like 4	6p21.1
TREM1	triggering receptor expressed on myeloid cells 1	6p21.1
TREM2	triggering receptor expressed on myeloid cells 2	6p21.1

Table 1. TREM family of genes.

Children STANISLAS Family Study (SFS)					
SNP	**Gene**	**Minor allele**	**MAF**	**Chromosome**	**HWE *P***
rs7748777	*LOC105375056*	A	0.44	6p21	1
rs6918289	*ADCY10P1*	T	0.13	6p21	0.46
rs7759295	*LOC105375056*	T	0.11	6p21	1
rs9357347	*LOC107986595*	C	0.29	6p21	1
rs6915083	*TREML2*	G	0.38	6p21	0.28
Total SFS					
rs6918289	*ADCY10P1*	T	0.13	6p21	0.73
Replication population (Adults)					
rs6918289	*ADCY10P1*	T	0.12	6p21	0.89

Table 2. Characteristics of the genotyped polymorphisms in *TREM 2* region. With the exception of the SNP rs6918289, the rest were only genotyped in children.

	SFS Children	SFS Total	SFS with IMT-F	Replication Population
Sample size [% female]	139 [52.5%]	809 [48.7%]	350 [53.1%]	916 [50.8%]
Age (years) [S.D]	15.4 [2.19]	30.3 [14.10]	31.9 [14.86]	55.5 [11.15]
BMI (kg/m^2) [S.D]	19.9 [2.22]	22.9 [3.99]	23.2 [4.19]	26.6 [3.47]
TNFa	3.57 [2.18]	3.26 [2.13]	—	2.68 [0.94]
IMT-F	—	—	0.50 [0.06]	—

Table 3. Demographic and clinical characteristics of studied populations.

inflammatory bowel disease[18] and stroke[19]. However, no genetic determinants have been identified in the *TREM2* locus affecting plasma levels of TNF-α and/or intima media thickness.

Due to the role of TREM2 in the inflammatory response and stability of atherosclerotic plaques, we hypothesize that polymorphisms in the *TREM2* gene region may influence the inflammatory process and subsequent atherosclerotic plaque formation. In this investigation, we have studied the association of variants located in the *TREM2* gene regionwith plasma levels of TNF-α and intima media thickness of the femoral artery (IMT-F).

Results

Information on the genotyped polymorphisms as well as the demographic and clinical characteristics of the studied populations is shown in Tables 2 and 3, respectively. All SNPs analyzed within this investigation were in agreement with Hardy-Weinberg equilibrium ($P > 0.001$).

Genetic association of SNPs in *TREM2* gene region with TNF-α concentration. Firstly, association analysis was performed in 139 children from the STANISLAS Family Study (SFS). SNPs that were located and previously genotyped in the *TREM2* gene region were tested for association with TNF-α concentration. Among the five SNPs studied, the minor allele (T) of rs6918289 was significantly associated with increased TNF-α concentrations ($P = 0.0015$, Table 4).

Secondly, rs6918289 was genotyped in 393 adult and additional 277 children relatives from the SFS. Analysis was performed in the combined population of SFS ($n_{total} = 809$) using three genetic models. The additive genetic model showed a positive association between polymorphism rs6918289, located in the *TREM2* gene downstream region and TNF-α plasma levels ($P = 0.0026$, $\beta = 0.13$, Table 5) (Fig. 1). The recessive model showed stronger genetic association with TNF-α plasma levels ($P = 0.0017$, $\beta = 0.498$, Table 5).

Further association analysis was performed in an independent French population of European ancestry, consisting of 916 individuals. However, only a marginal association for the additive model was observed ($P = 0.073$, $\beta = 0.042$, Table 5). A significant association was observed for the recessive model ($P = 0.023$, $\beta = 0.202$, Table 5). Interestingly, the minor allele (T) of SNP rs6918289 was associated with elevated levels of TNF−α in the discovery and replication populations.

Polymorphism	Model	P-value	Beta	S.E
rs7748777	Additive	0.04014	0.14	0.07
rs6918289	Additive	**0.00147**	0.33	0.10
rs7759295	Additive	0.9966	−0.0005	0.11
rs9357347	Additive	0.02127	0.17	0.075
rs6915083	Additive	0.1634	0.1064	0.076

Table 4. Genetic association of the SNPs in *TREM2* region with TNF-α levels in children of the SFS ($n = 139$). P-value threshold is $\boldsymbol{P < 0.01}$. Significant p-values are highlighted in bold.

Population	Genetic model	N	Beta	SE	P-value
SFS	**Additive**	809	0.13	0.04	**0.0026**
	Dominant	809	0.11	0.05	**0.022**
	Recessive	809	0.49	0.15	**0.0017**
Replication population	**Additive**	916	0.042	0.02	0.073
	Dominant	916	0.03	0.03	0.189
	Recessive	916	0.20	0.09	**0.023**

Table 5. Association analysis of the polymorphism rs6918289 with TNF-α levels in the discovery and replication populations. P-value threshold is $\boldsymbol{P < 0.05}$. Significant p-values are highlighted in bold.

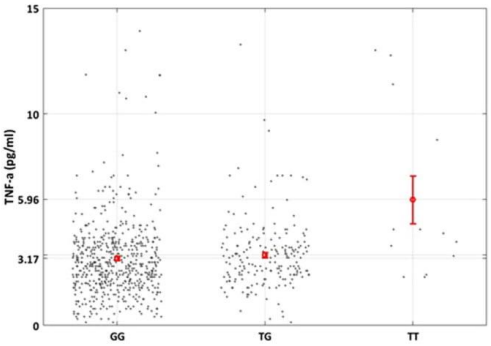


Figure 1. Mean values of TNF-α levels according to the different genotypes of rs6918289 (GG vs TG vs TT) in the SFS population. Thin bars show standard errors.

Meta-analysis of the SFS and the replication populations revealed similar results: A marginal association with TNF-α for the additive model results ($P = 0.072$) that becomes significant in the recessive model ($P = 0.0003$).

Genetic association of rs6918289 with IMT-F. A sub-group of 350 individuals, consisting of adults and children from the SFS population, where IMT-F measurements were available, was used to identify association with SNP rs6918289. The additive genetic model showed significant association ($P = 0.026$, $\beta = 0.02$, Table 6). The association was also significant for the dominant model ($P = 0.026$, $\beta = 0.024$, Table 6). Thus, the minor allele (T), of SNP rs6918289 was associated with increased thickness of the femoral artery (Fig. 2).

Bioinformatics analyses. *Polymorphism location.* SNP rs6918289 is located on chromosome 6 in p21.1 region. Located at 41 134 089 bp, it is an intron variant of adenylate cyclase 10 pseudogene 1 (ADC10P1) (Sup. Figure 1). The polymorphism on the forward strand is G > T with a minor allele frequency of 0.07 for thymine in 1000 Genome project[20,21].

Phylogenetic context. The guanine polymorphism of rs6918289 G > T is evolutionary and well-conserved in primates and mammals in general (Sup. Table 1). Further phylogenetic studies performed in 33 mammals and 46 vertebrates (Sup. Figure 2) showed that this polymorphism has a slower evolution rate than expected, having a PhyloP score of 0.061 and 0.056 respectively. This strengthens the idea that this variant is involved in important molecular mechanisms and that its preservation has been sustained throughout the natural selection processes.

Population	Genetic model	N	Beta	SE	P-value
SFS	**Additive**	350	0.02	0.009	**0.026**
	Dominant	350	0.024	0.01	**0.026**
	Recessive	350	0.024	0.025	0.34

Table 6. Association analysis of the polymorphism rs6918289 with IMT-F in the discovery population (not available in the replication population). P-value threshold is $P < 0.05$. Significant p-values are highlighted in bold.

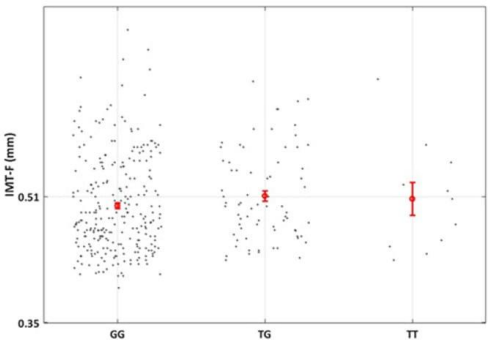


Figure 2. Mean values of intima media thickness of IMT-F according to the different genotypes of rs6918289 (GG vs TG vs TT) in the SFS population. Thin bars show standard errors.

Genomic context. SNP rs6918289 variant overlaps 2 transcripts and is located between two terminal exons of *ADCY10P1*: *ADCY10P1-202* (3785 bp - between exon 17 and 18) and *ADCY10P1-203* (4569 bp - between exon 18 and 19), both leading to transcripts that are not translated into proteins (Sup. Figure 2). Furthermore, according to the expressed sequence tags database (dbEST, https://www.ncbi.nlm.nih.gov/dbEST) available in Ensembl, the polymorphism rs6918289 is located within an intense transcriptionally active locus (Sup. Figure 3)[20,21].

Discussion

The anti-inflammatory effects of TREM2 have been described in several studies. Indeed, knockdown or silencing of *TREM2* gene results in increased levels of different pro-inflammatory molecules, among them TNF-α[6,7,22]. Despite the fact that several polymorphisms within the *TREM2* gene have been related with numerous neurodegenerative and inflammatory diseases[12,16,18,19], to date, no genetic determinants have been identified in the *TREM2* locus (6p21.1) affecting TNF-α concentrations.

In this study, we evaluated the effects of SNPs located in the *TREM2* gene region on TNF-α concentrations and IMT-F measurements. For 139 children from the SFS population, genotypes for five SNPs (rs7748777, rs6918289, rs7759295, rs9357347 and rs6915083) located in the TREM2 gene region were readily available. After conducting association analysis for the five aforementioned SNPs with TNF-α concentrations, we observed the minor allele (T) of SNP rs6918289, located in the downstream region of the *TREM2* gene, to be associated with elevated levels of TNF-α ($P = 0.0003$). Importantly, further analysis in a larger sample of SFS population, as well as replication in an independent population and a meta-analysis of discovery and replication populations confirmed this novel association. Subsequently, we also performed association analysis of rs6918289 with intima-media thickness of the femoral artery in 350 individuals of the discovery population. Results suggest that the minor allele (T) of SNP rs6918289 is associated with increasing intima-media thickness of the femoral artery. Therefore, this is evidence that the minor T allele may be considered as a risk allele for inflammatory diseases and atherosclerosis.

The role of both studied phenotypes in the development of atherosclerosis and the prognosis of atherosclerotic patients is well documented. TNF-α is a key regulator of immune response and alterations of its levels lead to elevated inflammation and a subsequent deterioration of the outcome of patients with cardiovascular diseases[23,24]. Also, one of the early processes that lead to atherosclerosis is the arterial remodeling and one effective way that provides information about this process is measuring the intima-media thickness. Indeed, IMT is predictive of atherosclerosis in asymptomatic individuals[25,26], and also provides information about the degree of atherosclerosis[27,28] as well as predicting the future risk of suffering a myocardial infarction event[29].

Chromosomal 6p21.1 genetic region, which is in proximity to the MHC and HLA regions, has been highly studied, and associated with autoimmune diseases[3,4]. Interestingly, previous GWAS studies have associated polymorphisms located within this genetic region with atherosclerotic stroke[30]. Although SNP rs6918289 is not in linkage disequilibrium with the variants associated with atherosclerotic stroke (rs556621 and rs556512), our results strengthen the idea that chromosomal 6p21.1 region could, indeed, be correlated with atherosclerosis risk.

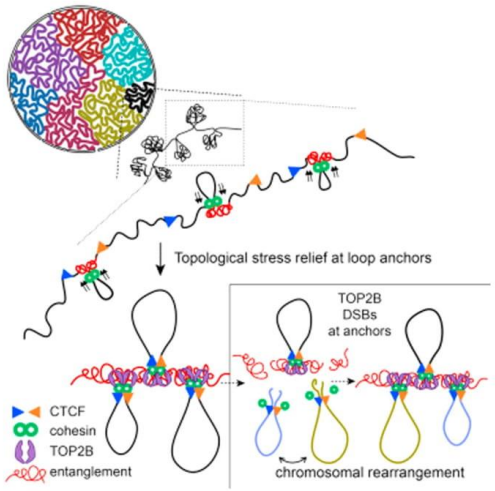


Figure 3. Organization of DNA forming topologically associating domain (TAD) within two CTCF sites in DNA sequence. (Retrieved from http://www.cell.com/cell/fulltext/S0092-8674(17)30718-3).

Our results are providing new insights on possible genetic regulations of pathological pathways that could lead to increased risk of atherosclerosis.

Despite the fact that this polymorphism contributes novel insights about its potential role in the development of cardiovascular diseases, the exact mechanisms by which this process is orchestrated is so far unknown and further studies are warranted. Indeed, there are no previous studies that associate this SNP or SNPs in linkage disequilibrium with TNF-α levels or IMT-F. In support of our findings, previous investigations suggest that TREM2 affects plasma TNF-α concentrations[7–9,22]. Also, our bioinformatic analyses indicate that rs6918289 is located in a transcriptional region of *ADC10P1* gene that could impact TREM2 transcription levels[31]. Indeed, rs6918289 is referred by rSNPBASE (http://rsnp.psych.ac.cn/) as a post-transcriptional regulatory element and, in lymphocyte B cells, rs6918289 is associated with the PABPC1 protein. The PABPC1 protein binds to the 3′ poly(A) region of the mRNAs. Although the binding of this protein is necessary for the translation initiation, it is also required for poly(A) shortening, which is the first step in mRNA decay[32]. Thus, the minor allele (T) of SNP rs6918289 could affect the PABPC1 protein, which would at the same time affect the stability of the TREM2 mRNA, consequently contributing to elevated plasma TNF-α concentrations.

An alternative or synergistic mechanism is also conceivable. As shown by several studies, many RNAs bind to transcriptional repressor CTCF (11-zinc finger protein) to modulate its regulatory functions[33]. SNP rs6918289 is located between two CTCF genomic sequences that are "together" (Sup. Figures 2 and 3), which means they are forming a chromatin loop leading to a topologically associating domain (TAD) (Fig. 3).

TAD is a large, megabase-sized local chromatin interaction domain that forms a self-interacting genomic region[34]. The structural and organizational changes of such region affect gene expression and other cellular functions, such as re-organization of local interactions between enhancers and promoters. Moreover, DNA sequences within a TAD physically interact with each other more frequently than with sequences outside the TAD[35]. Thus, we hypothesize that CTCF, in association with the non-coding RNAs of *ADCY10P1*, could affect TREM2 expression levels. The minor allele (T) of SNP rs6918289 could promote the binding of CTCF and consequently trigger an insulation mechanism for *TREM2* gene. These hypotheses should be tested in future works in order to conclude about the involved mechanisms.

Finally, the rs6918289 can be found in proximity to the *TREML1* gene. *In silico* analysis showed significant correlation of variant with seven SNPs related to TREML1 gene (Sup. Table 2). Therefore, *TREML1* could also be a possible mediator of the effect on plasma TNF-α concentration. *TREML1* gene (Triggering Receptor Expressed On Myeloid Cells Like 1) encodes a protein, involved in platelet aggregation, inflammation, and cellular activation. Together with TREM 2 they are involved in common pathways in the setup of Alzheimer's disease[36].

Concerning the IMT-F results, we think that the effect of SNP rs6918289 could be indirect. Indeed, the remodeling of the artery is highly dependent on the inflammatory state, thus, increased TNF-α levels produced by the minor allele (T) of SNP rs6918289 could be responsible for the increased thickness of the femoral artery. Further studies will be necessary in order to clarify if this polymorphism or polymorphisms in linkage disequilibrium are capable of modulating the TREM2 protein levels and to discover which are the mechanisms behind this observation. However, important applications could be identified as these findings may be used for personalized treatments in patients with chronic inflammatory diseases[37].

Limitation of our study is the unexplained mechanism of the rs6918289 effect on TNFα levels. Nevertheless, few pertinent hypotheses were proposed and future advances in genomic research with advanced bioinformatics tools will enable more precise explications of genetic interactions. Finally, genetic ancestry was not tested for populations used in this study; however strict inclusion criteria for origin were set aiming to collect homogeneous population, which are adequate for genetic association studies.

In summary, our study indicates variant rs6918289 located in the downstream region of the TREM2 gene as a candidate risk factor for inflammatory diseases because of its tight association with plasma TNF-α concentration. These findings support results of previous studies linking variants within chromosomal 6p21.1 loci to atherosclerosis and thus raising the awareness to consider genomic region 6p21.1 as candidate susceptibility loci for atherosclerosis.

Methods

Ethics statement. The samples are part of a human sample storage platform: BRC IGE-PCV number BB-0033–00051 located in Nancy, East France. All participants provided written informed consent. All populations involved in this study were recruited in accordance with the latest version of the Declaration of Helsinki for Ethical Principles for Medical Research Involving Human Subjects. All the protocols were approved by the local ethics committees for the protection of subjects for biomedical research: the Comité Consultatif de Protection des Persones dans la Recherche Biomédicale (CCPPRB).

Study participants. The discovery and replication populations are part of the Biological Resources Center 'Interactions Gène-Environnement en Physiopathologie CardioVasculaire' (BRC-IGE-PCV, number BB-0033–00051) in Nancy, France.

Individuals, comprising children and adults from the SFS[38] were used as the discovery population. The SFS include more than 1,000 nuclear families, each composed of at least four individuals (two parents and two children). All families are of French origin (parents and grandparents of French origin and residence in the Lorraine region) and were recruited at the Centre for Preventive Medicine of Vandoeuvre-lés-Nancy (East France). Even though strict criteria of French (European) origin were set for the inclusion of individuals in both populations, genetic ancestry was not tested. All individuals were free of chronic diseases. Firstly, we used samples from 139 children of the SFS, where genotype information was readily available for the *TREM2* gene region, to conduct association studies with plasma levels of TNF-α. Subsequently, 393 adult samples and 277 additional children samples from the corresponding families were included, reaching a total of 809 participants.

The replication cohort included 916 unrelated adults of French origin. Their inclusion criteria were the same as in the discovery cohort.

Blood samples and biological measurements. Blood samples were taken from individuals after an overnight fast (>8 hours). The plasma concentrations of TNF-α were measured by a commercially available enzyme-linked immunosorbent assay (ELISA) (R&D Systems, UK) according to manufacturer instructions. Body mass index (BMI) was calculated as weight divided by height squared (Kg/m^2). IMT-F was measured in 350 SFS individuals (including children and adults) using B-mode ultrasound methods[39]. The right and left femoral arteries were examined with a 7.5 MHz probe, according to a protocol already described[40,41]. For each individual, two IMT-F measurements were obtained and right and left measurements were used to calculate the mean IMT-F (in mm). IMT-F data were not available in the replication population.

Genotyping. Genome-wide genotypes were readily available for all children samples from the SFS ($n = 139$). Genotyping was performed using Illumina® human CNV370-Duo array[42]. The Illumina® protocol for the BeadStation genotyping platform was used, followed by GenCall® software analysis to automatically collect, call genotypes, and designate confidence scores using the GenTrain clustering algorithm. The selection of the SNPs was done by first extracting from the genome-wide assay all the SNPs available in the region around *TREM2* gene locus (50 kb upstream and downstream of the *TREM2* gene). We employed PLINK software[43] to conduct this analysis that resulted in the extraction and selection of five SNPs.

After performing initial association analysis in the child cohort of the SFS, polymorphism rs6918289 was *de novo* genotyped in the adult population ($n = 393$) and children population ($n = 277$) of SFS and the replication population ($n = 916$). Genotyping of rs6918289 in the replication population was conducted by Laboratory of the Government Chemist (LGC), using a PCR-based KASP assay[44].

Statistical analysis. Normal distribution was tested by Kolmogorov-Smirnov test. If phenotypes did not conform to normal distribution, data were log transformed in order to reach normality. The Hardy-Weinberg Equilibrium (HWE) was tested using the chi-square test ($P > 0.001$). The SNP effects on the studied phenotypes were tested through linear regression adjusted for age, gender and BMI under three inheritance models (additive, dominant and recessive) and using the minor allele as the reference allele. Analyses were performed using the R package GWAF (Genome-Wide Association/Interaction Analysis and Rare Variant analysis with Family Data)[45] taking into account familial resemblance. Alternatively, population characteristics were determined using SPSS statistical software version 20.0 (SPSS, Inc., Chicago, Illinois). Firstly, we tested the association of five SNPs located in the *TREM2* gene region that were extracted from the GWAS available in children of the SFS ($n = 139$) with plasma TNF-α concentrations. Secondly, we tested the significance of SNP rs6918289 with TNF-α plasma levels in all available individuals of SFS (children and adults combined; $n = 809$) and in the replication population ($n = 916$). Meta-analysis of the discovery and replication cohorts was performed using the GWAMA software and a random-effect method. Thirdly, in 350 individuals of the SFS, rs6918289 was tested for association with IMT-F. Bonferroni correction was applied in order to adjust for multiple testing. The P-value threshold was set at $P < 0.05/5 = 0.01$ for the first analysis made in children, $P < 0.05/1 = 0.05$ for the analysis made in all individuals of the discovery population (children and adults) and $P < 0.05/1 = 0.05$ in the replication population and the meta-analysis.

Bioinformatics analysis. Location, genomic and phylogenetic context of rs6918289 were determined on the Human genome (GRCh38.p10) using Ensembl browser[20]. The putative regulatory role of rs6918289 was established using rSNPBASE[21], and the PhyloP score was obtained using PhyloP software[46].

Data availability. Extensive data is provided with this article and further information is available from the authors on request.

References

1. Committee, H. H. G. N. *Gene names*, https://www.genenames.org/ (2017).
2. Ford, J. W. & McVicar, D. W. TREM and TREM-like receptors in inflammation and disease. *Curr Opin Immunol* **21**, 38–46, https://doi.org/10.1016/j.coi.2009.01.009 (2009).
3. Matzaraki, V., Kumar, V., Wijmenga, C. & Zhernakova, A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol* **18**, 76, https://doi.org/10.1186/s13059-017-1207-1 (2017).
4. Gough, S. C. & Simmonds, M. J. The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. *Curr Genomics* **8**, 453–465, https://doi.org/10.2174/138920207783591690 (2007).
5. Lanier, L. L. DAP10- and DAP12-associated receptors in innate immunity. *Immunol Rev* **227**, 150–160, https://doi.org/10.1111/j.1600-065X.2008.00720.x (2009).
6. Ito, H. & Hamerman, J. A. TREM-2, triggering receptor expressed on myeloid cell-2, negatively regulates TLR responses in dendritic cells. *Eur J Immunol* **42**, 176–185, https://doi.org/10.1002/eji.201141679 (2012).
7. Turnbull, I. R. *et al.* Cutting edge: TREM-2 attenuates macrophage activation. *J Immunol* **177**, 3520–3524 (2006).
8. Gao, X., Dong, Y., Liu, Z. & Niu, B. Silencing of triggering receptor expressed on myeloid cells-2 enhances the inflammatory responses of alveolar macrophages to lipopolysaccharide. *Mol Med Rep* **7**, 921–926, https://doi.org/10.3892/mmr.2013.1268 (2013).
9. Liu, D. *et al.* Impact of TREM-2 gene silencing on inflammatory response of endotoxin-induced acute lung injury in mice. *Mol Cell Biochem* **394**, 155–161, https://doi.org/10.1007/s11010-014-2091-6 (2014).
10. Raha, A. A. *et al.* Neuroprotective Effect of TREM-2 in Aging and Alzheimer's Disease Model. *Journal of Alzheimer's disease: JAD* **55**, 199–217, https://doi.org/10.3233/jad-160663 (2017).
11. Rai, V., Rao, V. H., Shao, Z. & Agrawal, D. K. Dendritic Cells Expressing Triggering Receptor Expressed on Myeloid Cells-1 Correlate with Plaque Stability in Symptomatic and Asymptomatic Patients with Carotid Stenosis. *PLoS One* **11**, e0154802, https://doi.org/10.1371/journal.pone.0154802 (2016).
12. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117–127, https://doi.org/10.1056/NEJMoa1211851 (2013).
13. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368**, 107–116, https://doi.org/10.1056/NEJMoa1211103 (2013).
14. Guerreiro, R. J. *et al.* Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. *JAMA Neurol* **70**, 78–84, https://doi.org/10.1001/jamaneurol.2013.579 (2013).
15. Cuyvers, E. *et al.* Investigating the role of rare heterozygous TREM2 variants in Alzheimer's disease and frontotemporal dementia. *Neurobiol Aging* **35**(726), e711–729, https://doi.org/10.1016/j.neurobiolaging.2013.09.009 (2014).
16. Benitez, B. A. & Cruchaga, C. & United States-Spain Parkinson's Disease Research, G. TREM2 and neurodegenerative disease. *N Engl J Med* **369**, 1567–1568, https://doi.org/10.1056/NEJMc1306509#SA4 (2013).
17. Rayaprolu, S. *et al.* TREM2 in neurodegeneration: evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease. *Mol Neurodegener* **8**, 19, https://doi.org/10.1186/1750-1326-8-19 (2013).
18. Correale, C. *et al.* Bacterial sensor triggering receptor expressed on myeloid cells-2 regulates the mucosal inflammatory response. *Gastroenterology* **144**, 346–356 e343, https://doi.org/10.1053/j.gastro.2012.10.040 (2013).
19. Sieber, M. W. *et al.* Attenuated inflammatory response in triggering receptor expressed on myeloid cells 2 (TREM2) knock-out mice following stroke. *PLoS One* **8**, e52982, https://doi.org/10.1371/journal.pone.0052982 (2013).
20. Aken, B. L. *et al.* Ensembl 2017. *Nucleic Acids Res* **45**, D635–D642, https://doi.org/10.1093/nar/gkw1104 (2017).
21. Guo, L., Du, Y., Chang, S., Zhang, K. & Wang, J. rSNPBase: a database for curated regulatory SNPs. *Nucleic Acids Res* **42**, D1033–1039, https://doi.org/10.1093/nar/gkt1167 (2014).
22. Takahashi, K., Prinz, M., Stagi, M., Chechneva, O. & Neumann, H. TREM2-transduced myeloid precursors mediate nervous tissue debris clearance and facilitate recovery in an animal model of multiple sclerosis. *PLoS Med* **4**, e124, https://doi.org/10.1371/journal.pmed.0040124 (2007).
23. Ridker, P. M. *et al.* Elevation of tumor necrosis factor-alpha and increased risk of recurrent coronary events after myocardial infarction. *Circulation* **101**, 2149–2153 (2000).
24. Koller-Strametz, J. *et al.* Circulating tumor necrosis factor-alpha levels in chronic heart failure: relation to its soluble receptor II, interleukin-6, and neurohumoral variables. *J Heart Lung Transplant* **17**, 356–362 (1998).
25. Rosvall, M., Janzon, L., Berglund, G., Engstrom, G. & Hedblad, B. Incidence of stroke is related to carotid IMT even in the absence of plaque. *Atherosclerosis* **179**, 325–331, https://doi.org/10.1016/j.atherosclerosis.2004.10.015 (2005).
26. Lorenz, M. W., von Kegler, S., Steinmetz, H., Markus, H. S. & Sitzer, M. Carotid intima-media thickening indicates a higher vascular risk across a wide age range: prospective data from the Carotid Atherosclerosis Progression Study (CAPS). *Stroke* **37**, 87–92, https://doi.org/10.1161/01.STR.0000196964.24024.ea (2006).
27. Burke, G. L. *et al.* Arterial wall thickness is associated with prevalent cardiovascular disease in middle-aged adults. The Atherosclerosis Risk in Communities (ARIC) Study. *Stroke* **26**, 386–391 (1995).
28. Nagai, Y. *et al.* Increased carotid artery intimal-medial thickness in asymptomatic older subjects with exercise-induced myocardial ischemia. *Circulation* **98**, 1504–1509 (1998).
29. Salonen, J. T. & Salonen, R. Ultrasound B-mode imaging in observational studies of atherosclerotic progression. *Circulation* **87**, II56–65 (1993).
30. Holliday, E. G. *et al.* Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet* **44**, 1147–1151, https://doi.org/10.1038/ng.2397 (2012).
31. Jones, B. L. & Swallow, D. M. The impact of cis-acting polymorphisms on the human phenotype. *Hugo J* **5**, 13–23, https://doi.org/10.1007/s11568-011-9155-4 (2011).
32. Jalkanen, A. L., Coleman, S. J. & Wilusz, J. Determinants and implications of mRNA poly(A) tail size–does this protein make my tail look big? *Semin Cell Dev Biol* **34**, 24–32, https://doi.org/10.1016/j.semcdb.2014.05.018 (2014).
33. Kung, J. T. *et al.* Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. *Mol Cell* **57**, 361–375, https://doi.org/10.1016/j.molcel.2014.12.006 (2015).
34. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376, https://doi.org/10.1038/nature11082, https://www.nature.com/articles/nature11082#supplementary-information (2012).
35. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* **16**, 245–257, https://doi.org/10.1038/nrm3965 (2015).

36. Carrasquillo, M. M. *et al.* A candidate regulatory variant at the TREM gene cluster associates with decreased Alzheimer's disease risk and increased TREML1 and TREM2 brain gene expression. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **13**, 663–673, https://doi.org/10.1016/j.jalz.2016.10.005 (2017).
37. Jiang, T., Yu, J. T., Zhu, X. C. & Tan, L. TREM2 in Alzheimer's disease. *Mol Neurobiol* **48**, 180–185, https://doi.org/10.1007/s12035-013-8424-8 (2013).
38. Visvikis-Siest, S. & Siest, G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. *Clinical chemistry and laboratory medicine* **46**, 733–747, https://doi.org/10.1515/cclm.2008.178 (2008).
39. Montauban van Swijndregt, A. D., De Lange, E. E., De Groot, E. & Ackerstaff, R. G. An *in vivo* evaluation of the reproducibility of intima-media thickness measurements of the carotid artery segments using B-mode ultrasound. *Ultrasound Med Biol* **25**, 323–330 (1999).
40. Gariepy, J., Massonneau, M., Levenson, J., Heudes, D. & Simon, A. Evidence for *in vivo* carotid and femoral wall thickening in human hypertension. Groupe de Prevention Cardio-vasculaire en Medecine du Travail. *Hypertension* **22**, 111–118 (1993).
41. Gariepy, J., Simon, A., Massonneau, M., Linhart, A. & Levenson, J. Wall thickening of carotid and femoral arteries in male subjects with isolated hypercholesterolemia. PCVMETRA Group. Prevention Cardio-Vasculaire en Medecine du Travail. *Atherosclerosis* **113**, 141–151 (1995).
42. Bouatia-Naji, N. *et al.* A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet* **41**, 89–94, https://doi.org/10.1038/ng.277 (2009).
43. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575, https://doi.org/10.1086/519795 (2007).
44. He, C., Holme, J. & Anthony, J. SNP genotyping: the KASP assay. *Methods Mol Biol* **1145**, 75–86, https://doi.org/10.1007/978-1-4939-0446-4_7 (2014).
45. Chen, M. H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580–581, https://doi.org/10.1093/bioinformatics/btp710 (2010).
46. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110–121, https://doi.org/10.1101/gr.097857.109 (2010).

Acknowledgements

This work was supported by the "Agence Nationale de la Recherche, programme d'Investissements d'avenir", grant number ANR-15RHU-0004. The funder provided support in the form of salaries for author Vesna Gorenjak, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. This work was supported by the University of Lorraine with a "Contrat de travail de droit public".

Author Contributions

Conceptualization: A.A.A. and S.V.S.; Formal analysis: A.A.A. and S.D.; Investigation: A.A.A., V.G., S.D., M.G.S., C.M.; Writing – original draft: A.A.A., V.G., S.D.; Writing – review & editing: A.A.A., V.G., S.D., S.V.S., M.G.S., D.R.V.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-25553-y.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

Publication Nº 7

The future of telomere length in personalised medicine.

Vesna Gorenjak, Samina Akbar, Maria G. Stathopoulou and Sophie Visvikis-Siest

Front Biosci (Landmark Ed), March 2018, 1;23:1628-1654

TL is recently being advertised by private companies as a biomarker of ageing and age-related conditions, yet, biochemical laboratories still haven't applied TL in routine clinical assessment (307). The abundance of information is making it difficult to understand the current role of TL in diagnostics and predict its future role in the field of personalised medicine.

In the review article entitled '*The future of telomere length in personalized medicine'* we gathered the most pertinent information to present the advances, the challenges and the perspectives of TL as a biomarker of ageing and a biomarker of diverse monogenic and chronic diseases. We discussed the role of TL in the setup of these diseases, as well as in treatment. Because we are aware of the complexity of pathological pathways and the importance of understanding and including multiple biomarkers approach in the next generation of TL research, we also discussed the association of TL with other biomarkers of inflammation, angiogenesis and oxidative stress. Finally, the heritability of TL was described and a large review of all GWAS studies of TL was included, with a systematic presentation of most significant genetic variants related to TL.

The review pointed out the perspective of novel techniques for TL measurement using whole-genome sequencing (WGS). WGS could permit a computational assessment of TL and allow large association studies using the existing whole-genome data (308). This would generate new information on genetic association of TL with a wide variety of diseases and phenotypes.

Because of the significant association of TL with many different pathologies (309) it is not likely that TL could ever be a unique biomarker for a specific age-related chronic disease. Alone or in combination with other molecules, TL could rather present a biomarker of the general health of an individual, where an accelerated rate of the telomere base-pair loss would exhibit an alarming signal, indicating an increased inflammation or cell senescence.

[Frontiers In Bioscience, Landmark, 23, 1628-1654, March 1, 2018]

The future of telomere length in personalized medicine

Vesna Gorenjak[1], Samina Akbar[1], Maria G. Stathopoulou[1], Sophie Visvikis-Siest[1, 2]

[1]INSERM UMR U1122, IGE-PCV, Gene-Environment Interactions in Cardio-Vascular Physiopathology, Faculty of Pharmacy – University of Lorraine, Nancy, 54000, France, [2]Department of Internal Medicine and Geriatrics, CHU Technopole Nancy-Brabois, Rue du Morvan, F-54511, Vandoeuvre-les-Nancy, France

TABLE OF CONTENTS

1. Abstract
2. Introduction
- *2.1. Telomere structure*
- *2.2. Telomere function*
- *2.3. Telomere-binding proteins*
- *2.4. Telomere lengthening*

3. The telomere syndromes and treatment
4. Leukocyte tl and methods of mesurements
5. Determinants of TL
- *5.1. Biomarkers related to TL*

6. Heritability and genetic variants associated with relative TL
7. TELOMEREs as a biomarker
- *7.1. Telomeres as aging biomarker*
- *7.2. Telomeres as a biomarker in age-related diseases*
- *7.3. Telomeres as a biomarker in cancer prognostic*
- *7.4. Telomeres as a biomarker in cardiovascular diseases (CVD)*
- *7.5. Telomeres as a biomarker in psychiatric disorders*
- *7.6. Telomeres as a biomarker in Alzheimer's disease*

8. Telomeres in anti-cancer treatment
- *8.1. Telomerase inhibitors*
  - *8.1.1. Telomerase inhibitor Imelstat (GRN163L)*
  - *8.1.2. G-quadruplex stabilisers*
  - *8.1.3. Natural products*
- *8.2. Future perspectives and other therapies*

9. Telomeres in treatment of other diseases
10. Commercial offer of telomere testing
11. Conclusions and new perspectives
12. References

1. ABSTRACT

Telomere length has been subject of studies for many decades, aiming to elucidate its role in physiological processes, in process of aging and in diverse pathologies. Yet today, there is still no "big title" discovery that would lead to a practical use of telomeres as a reliable biomarker or target for a new drug. However, therapies for chronic disease patients are being tested and companies are already offering commercial tests for telomere length measurement. The strong genetic heritability of telomeres is opening the place for pharmacogenomics researches that could promote the personalized treatment of diverse diseases. In this article, we present the recent knowledge of telomeres genetic determination obtained by genome-wide association studies (GWAS), important biomarkers related to telomere length and review the possibilities of telomere's practical implementation in the medical treatment of diverse

diseases and as a potential biomarker in personalized medicine. Furthermore, we summarise commercial offers of telomere length measurements available and we discuss the actions that should be taken to make steps forward into final application of the accumulated knowledge into practical use.

2. INTRODUCTION

2.1. Telomere structure

Telomeres are terminal nucleoprotein structures, composed of 5,000-12,000 base pairs of repetitive sequences of TTAGGG (3), playing a key role in the protection against genome instability-promoting events (4). During cell division, DNA polymerase cannot fully replicate the 3' end of linear DNA, resulting in the loss of 30-200 nucleotides from telomeric sequence by each cell division (5). Moreover, other mechanisms such as nuclease activation, oxidative damage and DNA replication stress may as well contribute to telomere shortening (6). After a critical degree of shortening, telomeres become dysfunctional and activate a DNA damage response, the p53 or the p16INK4a signaling pathway, which leads to senescence and finally to cell death, apoptosis (7). Because of their progressive shortening with age and sensitivity to oxidative stress, telomeres have been suggested as a biomarker of aging (6, 8) and age-related diseases (9).

2.2. Telomere function

Telomeres are preventing the end of the linear chromosomal DNA from being recognized by DNA reparative machinery as a broken end (6). They prevent distinct biochemical processes such as degradation of the terminal regions of chromosomes, fusion of a telomere, either with another telomere or with a broken DNA end and homologous recombination between the telomeric regions (4, 10). Apart from DNA end protection, the telomeres are also involved in the regulation of gene expression through transcriptional silencing of genes located close to the telomeres (11).

2.3. Telomere-binding proteins

Proteins binding to telomeres have a critical and complex role in regulating telomere function (12) and higher-order DNA conformations are necessary for normal telomere function (11). The structure of telomeres is determined by a complex of six telomere-specific proteins named shelterin (13). Telomeres have long G (guanine)-rich overhangs, called 3′-overhang or G-tail, overhanging 130-200 bases in length that result from degradation of both telomere ends (14). Shelterin affects the structure of G-tails by forming the so called "T-loop", which provides a general mechanism for protection of telomeres by masking the chromosome end from being recognized as DNA double-strand breaks (15). Additional protein-protein interactions mediate the communication of telomeres with structural components of the cell nucleus (16). Because of its crucial role in chromosome protection, it has been recently suggested that shorter G-tail length is more directly associated with age and higher disease risk factors than total telomere length (TL) (17), suggesting G-tail as a new potential biomarker, alternative to TL *per se* (18). Besides shelterin, G-quadruplexes are other higher-order structures, protecting the 3' overhang of telomeres from being accessed by the enzyme telomerase (19).

2.4. Telomere lengthening

Some normal human cells, immortalised cells and cancer cells are capable of preventing the shortening of telomeres with a specialised cellular ribonucleoprotein reverse transcriptase (10), which consists of the catalytic subunit *hTERT* (human Telomerase Reverse Transcriptase) and RNA component *hTERC* (human Telomerase RNA Component), a template for the elongation of the telomeric sequences by hTERT (20). Although *hTERT* and *hTERC* are sufficient to generate enzyme activity *in vitro*, telomerase relies on additional proteins to enable its complete functioning, within which, the best characterized is the dyskerin protein (*DKC1*) (21). By copying a template sequence, telomerase synthesizes one strand of the telomeric DNA running 5' to 3' towards the distal end of the chromosome and therefore extends the single-stranded overhang, G-tail (10). The telomerase activity in normal human cells appears during development of fetus and disappears in most somatic cells during embryonic differentiation, except in highly proliferative cells with a need for telomerase to maintain TL and genetic stability (22). This includes male germ cells, activated lymphocytes and certain types of stem cell populations (23).

Capped telomeres are inaccessible for telomerase; therefore, cells containing active telomerase can regulate the length of telomeres within well-defined limits by switching telomeres from capped to uncapped state with the higher-order telomeres DNA-protein complex (16). Capped state preserves the physical integrity of the telomere and enables cell division to proceed. However, whether telomere exists in capped or uncapped state depends also on its length; longer telomeres are more likely to form protective capped structure than shorter telomeres (24). Some immortalized human cells and some tumours maintain their telomeres in the absence of any detectable telomerase activity by a mechanism, named alternative lengthening of telomeres (ALT) (25). ALT is based on the homologous recombination, a common method of DNA repair, where a double-strand break of one chromosome gets repaired by DNA polymerase using template DNA, taken from a matching sister chromatid (26). During cancer development, cells mostly use the

© 1996-2018

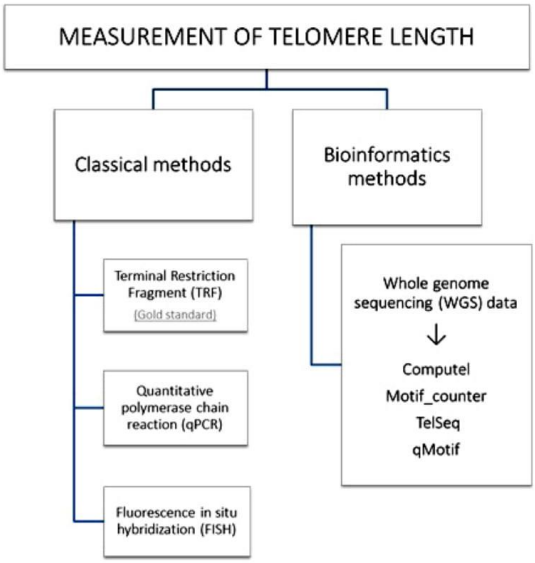


Figure 1. Presentation of methods used for TL measurement

telomerase-based telomere elongation, but also ALT process to acquire the ability to divide indefinitely (27).

3. THE TELOMERE SYNDROMES AND TREATMENT

Telomerase reconstruction for therapeutic use was firstly considered in diseases with distorted enzymatic activity of telomerase, the telomere syndromes (1). These are premature aging disorders with common defect of short telomeres, caused by the mutations in the genes that encode the essential telomerase components – *TERT, TERC* and *DKC1*. The most known and first discovered was dyskeratosis congenita (DC), a disease with childhood-onset, classically defined by mucocutaneous triad of skin hyperpigmentation, nail dystrophy and oral leukoplakia. All of the mutant genes identified in DC encode telomerase or telomere protein components. Mutations *DKC1* and *TIN2* gene are the most commonly identifiable (28). DC is associated with increased morbidity, often caused by bone-marrow failure, pulmonary fibrosis and/or severe immunodeficiency (21). Other disorders with onset in infancy are Hoyeraal–Hreidarsson syndrome and Revesz syndrome, both with severe telomere dysfunction.

More commonly, telomere disorders manifest in adulthood as a consequence of germline *hTERT* or *hTERC* mutations. The most common is idiopathic pulmonary fibrosis (IPF), followed by aplastic anemia. Among patients, 3-5% carry mutant *hTERT* and *hTERC* genes (28). Strategies of telomerase actie vation are raising safety concerns due to the close association with most cancers. Cycloastragenol, a molecule with a product name TA-65®, for example, has been shown to activate *hTERT* through activation of mitogenic pathways that lead to the activation of the oncogene c-myc and thus may drive cancer. For this reason, more targeted therapies are needed to ensure the safety. The most promising new strategy is hTERT therapy with gene therapy vectors (29), *demonstrated to improve blood counts and increase TL in the treatment of aplastic anemia, produced by short telomeres resulting from TERT deletion in mice (30),* that could be used particularly for the treatment of the human telomere syndromes (29). So far, no such study has been done in human population.

4. LEUKOCYTE TL AND METHODS OF MESUREMENTS

TL can be measured with different methods in different tissues. There are three methods that are currently used in laboratories. Telomere Restriction Fragment (TRF) (31) is a modification of Southern blot, considered as the “gold standard” of TL measurement. Quantitative real-time PCR (qPCR) (32) is the fastest and recently most used technique. The third method, fluorescence *in situ* hybridization (FISH) method (33), is using flow cytometry and defines the

© 1996-2018

length of the telomeres on individual chromosomes (Figure 1). Whole genome sequencing (WGS) - based TL measurement techniques (34) are new promising tools that use WGS data to calculate TL (Figure 1). Some studies have already calculated TL with bioinformatics tools such as Computel, Motif_counter, TelSeq and qMotif (34-40). These methods had a high level of concordance with q-PCR measurements (34). As a result, already existing WGS data could be used in the research of TL and could extend the range of possible studies with no further analytical cost.

Mean leukocyte TL (LTL) reflects systematic influence on telomere maintenance in other tissues and therefore, leukocytes are the most commonly used cell group for TL measurement (6). TL varies across somatic tissues in proportion to their replicative activity, but there is a strong correlation in TL across somatic tissues. Furthermore, it has been proven that within an adult, both highly and minimally proliferative tissues appear to display similar rate of age-dependent telomere shortening, suggesting that differences between TL in tissues were largely established at younger age (41).

5. DETERMINANTS OF TL

Since the discovery of DNA replication process, scientists pointed out the “end replication problem”(42, 43) and suggested that most cultured cells could survive only a limited number of cell divisions, commonly termed as the “Hayflick limit“, which might explain why physiologic function breaks down as organisms age (44). Thus, telomeres were since their discovery in 1970s an interesting subject of research because of their special nature, which is labelling them as the “aging biomarker“. Since then, hundreds of studies have been done, proving their inevitable correlation to age (8, 45-48). However, chronological age accounts for less than 10% of the variance in human TL (6). Besides age, gender (49) and ethnicity (50, 51) have also been considered as consistent predictors of TL, associating female gender and African and Hispanic origin with predisposition to longer telomeres in comparison to males and European origin. Human telomeres are relatively short in comparison to other mammals. From an evolutionary point of view, the cell division limit could be developed as a mechanism for tumor suppression (1). The presence of relatively shorter LTL in Europeans could attenuate the risk of melanoma in individuals with light skin pigmentation (50).

Environmental factors, such as smoking and physical activity, have been shown to be significantly correlated to shortening of TL (52-54), as well as traumatic events or certain psychiatric illnesses (55-58). In addition, studies on diseases’ risk factors, such as obesity, lipid profiles and hypertension have shown some inconsistent results. However, new evidences of correlation of these factors with TL are arising (59-61). The determination of epigenetic regulation of TL homeostasis with or without life style risk factors (eg obesity and smoking) remains an interesting area to be explored in telomeres biology (62-64), especially since genetic and epigenetic alterations seem important for a normal cell to suppress the telomere attrition and become malignant (65).

5.1. Biomarkers related to TL

A number of biomarkers related to different physiological processes are strongly associated with an acceleration of TL shortening, even though the mechanistic details and the molecular pathways of these connections have not been uncovered yet (66, 67).

Therefore, the blood or urine detection of protein biomarkers, produced as a consequence of telomere dysfunction in a disease state in any tissue, can provide a valuable clue towards the diagnosis of telomere-associated disorders (Table 1). Different physiological conditions, such as inflammation, oxidative stress and angiogenesis, play significant roles in determining an individual’s TL, both in health and disease (Figure 2) (61).

Chronic inflammation contributes significantly to the pathogenesis of multiple age-related diseases including cancer, atherosclerosis, autoimmune disorders, obesity, chronic obstructive pulmonary disease, diabetes, hematological disorders and neurodegenerative diseases (68-74), which in turn, are strongly associated with telomere shortening (66). Accumulating evidence from different epidemiological studies (Table 1) suggests a strong association of increased systemic inflammation with decreased TL (75, 76). The most significant inflammatory biomarkers include tumor necrosis factor alpha (77), C-reactive protein (76-82), serum amyloid A (76) and interleukin-6 (77, 78, 83). Recent studies have demonstrated that the inflammation is closely interconnected with angiogenesis, whereby the pathological angiogenesis exacerbates the chronic inflammation by stimulating the recruitment of inflammatory cells that release pro-angiogenic cytokines and growth factors (84). Various studies have shown an association between different angiogenic biomarkers with TL. Vascular endothelial growth factor-A (VEGF-A or VEGF) is a potent and cell-specific angiogenic factor (85), which has also been shown to be associated with various inflammatory markers (86). VEGF positively regulates telomerase activity in both *in vivo* and *in vitro* models (87, 88). However, some epidemiological studies have shown that plasma VEGF is negatively or not correlated with TL (76, 89). In addition, hepatocyte growth factor and granulocyte colony-stimulating factor, which contribute significantly both in physiological and pathological angiogenesis, were also found to be negatively correlated with TL in knee osteoarthritis patients (Table 1) (89).

© 1996-2018

Table 1. Association of TL with the biomarkers of inflammation, angiogenesis and oxidative stress in different population studies

Population	Study	Measurement method	Biomarkers	References
Inflammatory biomarkers				
1,517 Caucasian female twins, aged 18–79 years	TwinsUK	Southern blot	Higher CRP in postmenopausal than premenopausal women CRP levels inversely correlated with LTL in premenopausal but not postmenopausal women	(79)
419 men and women, aged 65–93 years	Cardiovascular Health Study (CHS)	Southern blot	Negative correlation of LTL with IL-6 and CRP levels	(78)
2,160 healthy Caucasian women twins, aged 18–79 years	TwinsUK	Southern blot	Positive correlation between high serum vitamin D levels and LTL Negative association of CRP levels with LTL and vitamin D concentrations	(81)
1,319 healthy Caucasian twins, aged 18–81 years	TwinsUK	Southern blot	Negative correlation between LTL and high CRP levels	(80)
136 patients with COPD, aged (mean±SD) 62.9±6.6, 113 age-and sex-matched smokers, aged (mean±SD) 62.2±7.7 and 42 nonsmokers, aged (mean±SD) 61.4±6.1 with normal lung function	Case-control study	Real time-qPCR	Negative correlation between IL-6 and LTL COPD patients	(83)
1,962 well-functioning men and women, aged 70–79 years	Health, Aging and Body Composition (Health ABC) Study	Real time-qPCR	High IL-6 and TNF-α levels associated with short LTL	(77)
87 male subjects over 18 years (mean±SD) 41.5±13 exposed to environments with high levels of occupational fine particulate matter (PM2.5)	The Harvard Boilermakers Longitudinal Study	Real time-qPCR	No change in TNF-α and IL-1β, IL-2, IL-6, IL-8, and IL-10 levels Negative correlation of plasma SAA and CRP levels with LTL	(76)
274 women volunteers: 150 with PCOS and 124 healthy women constituting the control group, aged 13-45 years	Case-control study	Real time-qPCR	High CRP levels negatively correlated with LTL in PCOS group	(82)
Angiogenic biomarkers				
87 male subjects over 18 years (mean±SD) 41.5±13, exposed to environments with high levels of occupational fine particulate matter (PM2.5)	The Harvard Boilermakers Longitudinal Study	Real time-qPCR	No change in VEGF levels Insignificant negative correlation with sICAM-1 and sVCAM-1	(76)
80 knee osteoarthritis patients (63 females and 17 males), aged 49–84 years and 60 healthy controls (34 females and 26 males); aged 50–80 years	Case-control study	Real time-qPCR	Negative correlation between RTL and plasma levels of VEGF, HGF, G-CSF	(89)
Oxidative stress biomarkers				
327 Caucasian men, aged 40–89 years	Offspring cohort of the Framingham Heart Study	Southern blot	Inverse correlation between LTL and 8-epi-PGF2α	(99)
34 T1D patients (21 men and 13 women), aged (mean±SD) 26.32±6.46; 62 T2D patients (35 men and 27 women), aged (mean±SD) 50.15±12.17; 40 non-diabetic controls (21 men and 19 women), aged (mean±SD) 32.25±9.74	Case-control study in Chinese Han participants	Real time-qPCR	Negative association of higher 8-OHdG levels and LTL in both T1D and T2D patients	(100)
88 (36 men and 52 women) metabolic syndrome subjects	LIPGENE cohort	Real time-qPCR	Elevated plasma protein carbonyl and urinary F2-isoprostanes levels in subjects with the shortest RTL	(98)
71 patients with newly diagnosed T2D (40 men and 31 women), aged (mean±SD) 54.55±8.37, with (n = 17) or without (n = 54) depression; 52 normal glycemic control subjects (30 men and 22 women), aged (mean±SD) 51.27±7.66; with (n = 6) or without (n = 46) depression	Case-control study	Real time-qPCR	8-OHdG levels contribute to TL shortening and depression development in newly diagnosed type 2 diabetic patients	(101)

© 1996-2018

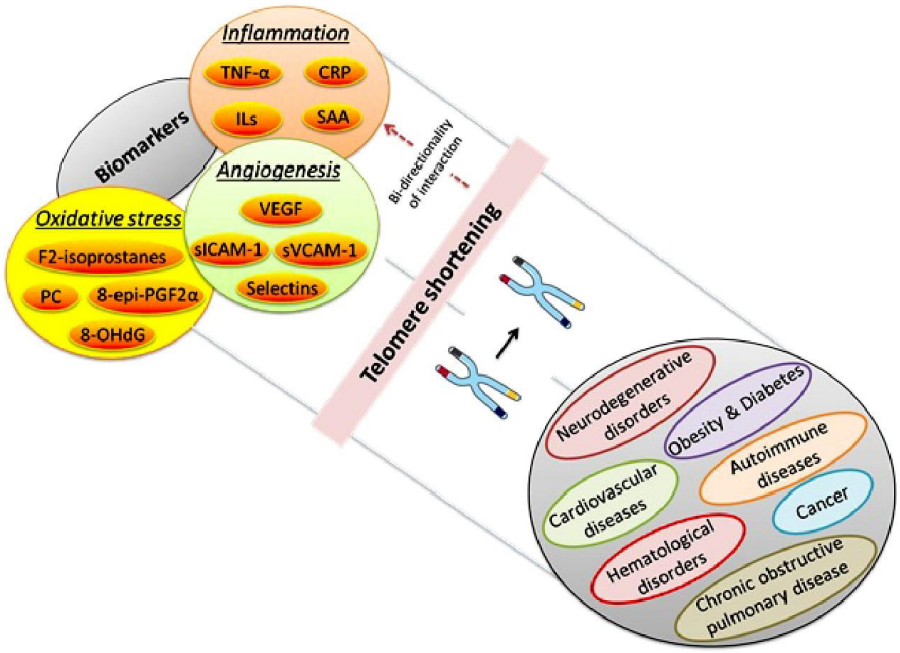


Figure 2. Inflammation, angiogenesis and oxidative stress biomarkers related to telomere shortening and increased risk of various age-related diseases.

Moreover, the relation between inflammation and angiogenesis is further bridged by the oxidative stress (84), which is implicated in aging and various age-related diseases including cancer, cardiovascular and neurodegenerative disorders (90-92). Oxidative stress plays an important role in telomere attrition (93), mainly because of the high content of guanines in telomeric sequences, which renders them highly prone to damage by free radicals (94-96). Another mechanism of telomere shortening by reactive oxygen species includes the single-strand breaks caused by hydroxyl radicals as well as the inefficient repair system for these lesions in the telomeric DNA (97). Elevated levels of urinary F2-isoprostanes and plasma protein carbonyl (produced through the oxidation of protein backbones) have been associated with the shortest relative TL in a population of 88 subjects with metabolic syndrome (98). Moreover, the LTL was also shown to be inversely correlated with another biomarker of oxidative stress, 8-epi-$PGF_{2\alpha}$, the product from isoprostane pathway (99). In addition, increased levels of 8-hydroxy-desoxyguanosine (8-OHdG), a marker of oxidative DNA damage, have also been negatively correlated with TL in type 1 and type 2 diabetes patients (Table 1) (100, 101).

To summarize, the cross-connections of these biomarkers with one another as well as with the telomerase and telomere-associated proteins could emerge as an excited area of research to develop the therapeutic approaches targeting specific pathological conditions. However, most of the studies linking TL to inflammatory biomarkers were based on cross-sectional design, and therefore, cannot explain the bi-directionality that exists between TL shortening and inflammation. It is, therefore, highly desirable to use data from longitudinal studies with large sample sizes in order to determine the evolution of various biomarkers at multiple time points, which will allow to better explore the telomere dynamics in relation with various chronic diseases (76, 89, 102, 103).

6 HERITABILITY AND GENETIC VARIANTS ASSOCIATED WITH RELATIVE TL

Previous studies have shown that TL is a highly heritable trait (36-86 % in different family and twin studies) (104-109). Whereas TL has a high heritability, the inheritance patterns require further investigations, as some studies suggested stronger paternal inheritance, indicating the possible modification by

© 1996-2018

parental age at conception and provided evidence that newborns with older fathers had significantly longer telomeres (109, 110). However, in some other studies, maternal TL inheritance was found to be more important (111).

Genome-wide association studies (GWAS) in large populations have been used to investigate associations between genetic variants and complex clinical conditions and phenotypic traits (112).

A search that we performed using the term "telomere" in GWAS Catalog (https://www.ebi.ac.uk/gwas/) and HuGE Navigator – GWAS Integrator (https://www.cdc.gov/genomics/hugenet/hugenavigator.htm) identified GWAS associated with TL. Selected articles concern those where only TL was reported as a trait and were excluded the diseases/traits with indirect connection to TL, such as sickle cell anaemia, testicular germ cell cancer or glioma. Eleven publications matched our searching criteria and are presented in Table 2. The presented studies identified common genetic variants associated with relative TL, where single nucleotide polymorphisms (SNPs) from different genes were proposed to explain the inter-individual variability of TL (113-123). In Table 2, we present only SNPs that have reached the threshold for genome-wide significance ($p \leq 5\times 10^{-8}$) or have shown supportive evidence for association ($p \leq 5x\ 10^{-5}$) (117). In case, where more SNPs had been identified for the same locus, the SNP with the strongest association is presented. Furthermore, the roles of the genes on loci with significantly associated variants are also described. The variance of TL explained by discovered SNPs is not higher than 1% for single identified locus, which indicates that until now, only a small part of genetic variability of TL has been identified (113-123) and further investigations are needed for the discovery of causal genetic biomarkers.

Most of SNPs that reached genome-wide significance threshold harbor genes that encode proteins with known functions in telomere biology (*i.e.* genes, directly involved in telomere maintenance). Among them, the most significantly associated locus is *TERC* on 3q26, reported and replicated in several studies, encoding telomerase RNA component. Furthermore, 5p15.3.3 locus harbors *TERT*, a gene encoding the reverse transcriptase subunit of the telomerase enzyme. The mutations of these genes are causing rare monogenic diseases that are associated with short TL (*e.g.* dysteratosis congenita) and are linked to several types of cancers (115). Gene *OBFC1* (oligonucleotide/oligosaccharide-binding fold containing 1) on 10q24.3.3 is specifically involved in the replication and capping of telomeres; along with *CTC1* and *TEN1* genes it encodes a component of the telomere-binding CST complex, which binds and protects telomeres *via* association with the shelterin complex. Shelterin complex consists of six other proteins (TRF1, TRF2, TIN2, TPP1, RAP1 and POT1). Zinc finger proteins can bind to G-quadruplex DNA and stabilize it; *ZNF676* and *ZNF208* are supposed to modify TL by direct binding to DNA and altering the expression of genes involved in telomere maintenance or through interaction with RNA to affect the post-translational signaling of these genes (113, 117). *MYNN* is located on 3q26.2. and encodes myoneurin, a member of the BTB/POZ and zinc finger domain-containing protein family, is controlling gene expression (124). Furthermore, gene *NAF1* (nuclear assembly factor 1) on 4q32.2. locus is involved in the formation of the telomerase enzyme and *RTEL1* on 20q13.3. is the regulator of telomere elongation helicase 1 involved in setting TL.

However, some significant loci do not harbor obvious genes related to telomere biology. In the second general category of genes associated with TL, we find genes that impact the turnover rate of hematopoietic stem cells (115). The 2p16.2. locus, for instance, contains the gene Acylphosphatase 2, muscle type (*ACYP2*) that has a specific role in muscle differentiation and stress induced apoptosis. Another gene, unlikely to be involved in telomere biology, is *PXK* gene, which codes for a serine/threonine kinase and is involved in regulation of electrical excitability and synaptic transmission. *DKK2* gene plays a role in embryonic development and may be involved in bone diseases, cancer and Alzheimer's disease in adults. The *PAPSS1* gene encodes a trypsinogen. Mutations on this gene are associated with hereditary pancreatitis. Finally, *CSNK2A2* encodes an enzyme, casein kinase II subunit alpha, which phosphorylates a large number of substrates and regulates numerous cellular processes, such as cell cycle progression, apoptosis and transcription. It is affiliated with the members of the shelterin complex involved in chromosome end protection, TL regulation and maintenance (113-123).

7 TELOMERES AS A BIOMARKER

By the definition of the Surrogate Endpoint Working Group, a biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (125). Short telomeres are associated with increased risk of early death and measurement of TL has been mentioned as a possible biomarker of aging and many age-related diseases, also because of (*i*) its tight correlation with oxidative stress and cellular senescence, two important regulators of human aging (9), (*ii*) its relation to basic biological mechanisms, (*iii*) record of past cell divisions, (*iv*) its high correlation across tissues and (*v*) association with lifestyle factors (126). A lot of effort has been done to evaluate whether clinical practice could benefit from TL measurement;

© 1996-2018

Table 2. Common genetic variants associated with relative TL

Number of individuals	Methods for TL analysis	SNP	p-value	Chromosomal Region	Reported gene(s)	Reference
1625 1165 (replication)	Southern blot	rs2162440	3×10^{-6}	18q12.2	BRUNOL4, PIKC3C	(118)
2917 9492 (replication)	Quantitative PCR	rs12696304	4×10^{-14}	3q26.2	TERC	(122)
3417 1893 (replication) + 2876 (in silico)	Southern blot	rs4387287 rs1975174 rs4452212 rs2736428	2×10^{-11} 2×10^{-6} 2×10^{-6} 3×10^{-6}	10q24.33 19p12 2q22.1 6p21.33	OBFC1 ZNF676 CXCR4 SLC44A4	(115)
459 890 (replication)	Quantitative PCR	rs6028466 rs621559 rs398652 rs654128	3×10^{-7} 2×10^{-6} 2×10^{-6} 3×10^{-6}	20q11.22 1p34.2 14q21 6q22.1	DHX35 WDR65 PELI2 KPNA5	(114)
3554 2460 (replication)	Quantitative PCR	rs12696304	2×10^{-14}	3q26.2	TERC	(120)
9190 2226 (replication)	Southern blot	rs9419958 rs1317082 rs412658 rs3027234	9×10^{-11} 1×10^{-8} 1×10^{-8} 2×10^{-8}	10q24.33 3q26.2 19p12 17p13.1	OBFC1 TERC ZNF676 CTC1	(117)
37 684 10 739 (replication)	Quantitative PCR	rs10936599 rs2736100 rs7675998 rs9420907 rs8105767 rs755017 rs11125529	3×10^{-31} 4×10^{-19} 4×10^{-16} 7×10^{-11} 1×10^{-9} 7×10^{-9} 7.5×10^{-10}	3q26.2 5p15.33 4q32.2 10q24.33 19p12 20q13.3 2p16.2	TERC TERT NAF1 OBFC1 ZNF208 RTEL1 ACYP2	(113)
2240 15 065 (replication) + 11 024 (replication)	Quantitative PCR	rs1317082 rs7726159 rs2487999 rs6772228 rs9257445 rs6060627	1×10^{-19} 5×10^{-17} 4×10^{-14} 4×10^{-10} 1×10^{-7} 6.5×10^{-7}	3q26.2 5p15.3 10q24.3 3p14.4 6p22.1 20q11.2	*TERC* *TERT* *OBFC1* *PXK* *ZNF311* *BCL2L1*	(119)
4289	Quantitative PCR	rs7680468 rs12638862 rs1317082 rs7100920 rs11787341 rs10904887 rs16859140 rs34596385 rs10466239 rs73394838 rs4902100	5×10^{-8} $<5 \times 10^{-8}$ $<5 \times 10^{-8}$ $<5 \times 10^{-8}$ 9×10^{-7} 4×10^{-6} 5×10^{-6} 6×10^{-6} 7×10^{-6} 9×10^{-6} 4×10^{-6}	4q25 3q26.2 3q26.2 10q24.33 8p21.3 10p13 3q13.2 6q24.1 10q11.21 22q12.2 14q23.2	DKK2, PAPSS1 TERC MYNN OBFC1 LOC100128993 TRDMT1 TMPRSS7 AK097143 FXYD4, RASGEF1A ASCC2 SYT16	(123)
2632 3917 (replication) + 696 (replication)	Quantitative PCR	rs2736100 rs17653722	2×10^{-5} 7×10^{-6}	5p15.33 12q13.13	TERT KRT80	(116)
1616 2397 (replication) + 2952 (replication)	Quantitative PCR	rs74019828 rs2098713 rs78869517	5×10^{-8} 3×10^{-6} 3×10^{-5}	16q21 5p13.2 5q21.3	CSNK2A2 C5orf42 FER	(121)

from determination of biological age of individual in order to predict mortality, to stratification of patients into different risk groups of precise medical conditions.

Recent discoveries showed that DNA can be damaged also with normal TL, if G-tail structure is disrupted, suggesting, that G-tail length contributes more to pathologies than TL itself (18). It has been associated with age and vascular risk factors (17) and methods for the examination of G-tail length with a high-throughput platform using genomic DNA or cell lysate are faster than classical methods for TL determination, giving G-tail priority as a biomarker choice in the future studies (127).

7.1. Telomeres as aging biomarker

Wide inter-individual differences among the subject of the same age-group suggest that chronological age is not a relative measure of person's health status, therefore, research of biomarkers has arisen to provide more information about an individual's biologi-

© 1996-2018

cal health, to point out those at high risk of age-related conditions, and to evaluate the progression of interventions, designed to delay the onset of age-related conditions (128). Aging Research proposed four criteria for aging biomarkers: 1) it must predict the rate of aging better than chronological age and tell exactly where a person is in their total life span, 2) it must monitor a basic process that underlies the aging process, not the effects of disease, 3) it must be able to be tested repeatedly without harming the person and 4) it must be something that works in humans and in laboratory animals in order to be tested in lab animals before being validated in humans (129).

Because it is difficult to separate the effects of chronic diseases from normal aging, there has been no such marker that would fulfill all the requested criteria. It has been suggested that, if TL is a biomarker of aging, it is a weak biomarker with poor predictive accuracy (9). However, looking wider at the perspective, TL can be considered as "a marker that predicts biological age of organism" and could be an ideal candidate for life course analysis, the study of long-term effects of physical and social exposures on chronic disease risk in different life stages, from childhood to late adulthood (9). Telomere shortening rates are not constant, but are influenced by positive or negative regulators (8). With systematic testing, factor such as TL can represent etiology of diseases more accurately and might thus be very advantageous.

Another perspective of epidemiological research of aging is studying TL in people who age slowly and exhibit "longevity potential" or "healthy aging". To do so, CHARGE Consortium (the Cohorts for Heart and Ageing Research in the Genomic Epidemiology) created longevity phenotype for GWAS (9). TL has been therefore investigated within the research of healthy aging and longevity biomarkers. However, it is hard to distinguish the changes causal to aging and longevity from those that are a consequence of normal aging and such marker, which could well explain the longevity potential of individual has not yet been discovered (130, 131).

7.2. Telomeres as a biomarker in age-related diseases

As telomeres are involved in the etiology of diverse diseases, new possible progress in diagnosis was seen in potential use of telomeres as a biomarker. Consequently, it was necessary first to elucidate their role in disease development, confirm the causality and separate their effect from any other disease risk factors. These tasks are complex and the conclusions obtained from the studies are not always clear. TL is therefore still not present in everyday clinical practice. Nevertheless, a lot of progress has been done and ongoing and prospective studies are getting closer to find the solutions for better and faster diagnosis and treatments.

7.3. Telomeres as a biomarker in cancer prognostic

Diverse studies have reported association of short TL in bladder (132), breast (133), colorectal (134), leukemia (135), lung (136), ovarian (137), pancreatic cancer (138) and others. Even though most of the studies reported higher risk of cancer in patients with shorter telomeres, these results are not consistent for all cancer types. Increasing evidence shows that mutations, which appear to lengthen telomeres, are linked to increased risk of cancer as well (139). Longer telomeres were associated with increased melanoma (140), glioma (141), endometrial cancer (142) and prostate cancer (143, 144). What are the reasons for the discrepancy of these results?

There are two different mechanisms in etiology of cancer, both related to telomeres. Short telomeres are promoting the events that lead cell to senescence or apoptosis, which causes accumulated mutations, genetic lesions and inactivated tumor suppressor checkpoints that can result in cancer. On the other hand, cancer cells are maintaining their capacity for infinitive division with the elongation of telomeres; elevated levels of various telomerase-stimulating factors may increase telomerase activity and contribute to longer TL (145). Second reason for discrepancy of results might be the study design. Different studies in TL differ in the use of measuring method, technique of sample collection, statistical analysis and other parameters (145). Only with additional studies, the hypothesis scientists try to confirm will be clarified and will open the way to new challenges. New insights into mechanisms of TL maintenance and telomerase expression and their transcriptional, post-transcriptional and epigenetic regulation can help to set telomeres as a new biomarker for early detection and prognosis of disease (146).

7.4. Telomeres as a biomarker in cardiovascular diseases (CVD)

As already discussed, reduced TL is associated with many intermediate phenotypes, namely chronic inflammation, hypertension, obesity and unhealthy lifestyle, all of which are risk factors for CVDs besides age, which is the major risk factor. Furthermore, several CVD manifestations such as coronary heart disease, carotid atherosclerosis, stroke, and abdominal aortic aneurysm have showed association with short telomeres, which might be shortened because of the interaction with common intermediate phenotypes of CVD (147). The mechanisms of causality still need to be revealed, however, growing understanding of cardiovascular

© 1996-2018

aging and telomere biology is enabling novel possible interventions and new therapies; maintenance of TL and telomerase activity regulation are already being tested for treatment of CVD (148). Moreover, some medications showed beneficial effects on TL maintenance. Statins and pioglitazone were associated with increased telomerase activity and protection of telomere by up-regulation of *TRF2,* and angiotensin II with reduction of oxidative stress (149-152).

In order to understand whether TL represents a unique biomarker for CVD, is it a marker of one primary risk factor such as inflammation, or even multiple risk factors, a study of TL in relation to 17 biomarkers of CVD risk was performed (153). The findings showed moderate association with multiple important CVD risk factors, such as BMI, waist circumference, percentage of body fat, HDL, TGs, pulse rate, CRP and cysteine C, probably by increased tissue inflammation and oxidative stress (148). However, there was no association with measures of overall CVD risk (the metabolic syndrome), suggesting that TL could be a biomarker of cardiovascular aging through established physiological mechanisms. TL is related to biomarkers of multiple regulatory systems that indicate risk for CVD; it could therefore serve as a cellular based indicator of systemic allostatic load, the "wear and tear on the body" (153).

Among traditional risk factors for CVD, decreased regenerative capacity, normally estimated by circulating levels of progenitor cells (PCs), was believed to be another major link to TL. However, a study of patients with coronary artery disease (CAD) showed that shorter TL is associated with lower levels of PCs, but they predicted adverse cardiovascular outcomes independently and additively one of another (154). The role of telomeres in CAD has been investigated in many studies (155-161), suggesting that TL is involved in the pathology of the disease, associating it with artery stiffness (156) and atherosclerosis (161). Furthermore, TL was found to be significantly shorter in young healthy adults from families with higher incidence of CAD (159). These results are suggesting the causal rather than consequential relation of telomeres and CAD. In favor of this hypothesis is also the association of genetic variants affecting TL with the risk of CAD (148); the polymorphism $^{-1327}$T/C of *hTERT* was found to be significantly associated with susceptibility to CAD among CAD patients in comparison to healthy controls. Moreover, in a large GWAS of TL in association with age-associated diseases, a sample of 22,233 coronary artery disease cases and 64,762 controls showed association of the alleles related to shorter TL with increased risk of CAD (113). It is therefore indisputable that TL has a big prognostic value in patients with CAD and could import additional support in preliminary CAD diagnostic.

Besides CAD, decreased TL was significantly correlated with heart failure (162) and peripheral arterial disease (163). Causality of TL in health outcomes was investigated with two-approach analysis also in other CVD and risk factors, using GWAS data and individual data of 3,734 individuals. However, examination of risk factors including diabetes, hypertension and obesity did not confirm any causal relationship with shorter telomeres, but showed decreased risk of stroke and increased risk of heart disease (164). Finally, studies of TL in stroke have shown very inconsistent results; another study investigating telomere-related SNPs in relation to predisposition for ischemic stroke in a case-control study suggested that variants of the *hTERC* and *hTERT* genes could have shortened LTL and lead to increased possibility of having ischemic stroke (165).

7.5. Telomeres as a biomarker in psychiatric disorders

Many psychiatric disorders, such as major depressive disorder (MDD), bipolar disorder (BD), post-traumatic stress disorder and schizophrenia have shown correlation to shorter leukocyte TL (40, 55-58, 166-168). A recent meta-analysis confirmed these findings in patients with MDD (169), a syndrome of "premature aging". Patients suffering from this disorder have an accelerated cellular aging that can reduce life expectancy up to 10 years (57). Telomere shortening may thus increase the vulnerability of psychiatric patients to premature death (167). Furthermore, rs10936599 SNP for T-carriers, located upstream of *hTERC* gene was found to predict a small, but significant increase in the risk of childhood-onset recurrent MDD (170). These findings suggest an important role of LTL in this disorder (169) and the telomerase enzyme might thus represent an important drug target for the prevention of early-onset MDD (170). Some pre-clinical studies suggested that the use of antidepressants, selective serotonin reuptake inhibitors (*SSRIs*) and lithium can increase telomerase activity; fluoxetine increased telomerase activity in hippocampus in mice (171) and continuous administration of lithium for 6 weeks significantly increased TERT expression and telomerase activity in hippocampus (172). Patients suffering from BD showed significantly increased LTL after receiving treatment with lithium. Thus, it might exert a protective effect against telomere shortening *via* telomerase activation (57).

However, it is still unknown which pathway underlies the association of shorter TL and psychiatric disorders; inflammation, metabolic alterations and smoking are important mediators in tight connection to psychiatric diseases. Therefore, additional studies should evaluate the role of lifestyle and telomerase in the evolution of mental illness in order to elucidate the complex mechanisms and to consider telomeres as a potential treatment target.

© 1996-2018

Table 3. Basic approaches of telomerase-based anti-cancer therapy

Direct enzyme inhibition (active site inhibitors, nucleotide substrates and allosteric inhibitors) by targeting TERT or TERC unit of enzyme
Direct enzyme inhibition (active site inhibitors, nucleotide substrates and allosteric inhibitors) by targeting TERT or TERC unit of enzyme
Active immunotherapy (antigen-presenting cells) by activation of native or memory TERT-specific T cells, which cooperate to kill tumour cells with displayed TERT peptides, through classical major histocompatibility complex (MHC) presentation
Telomere-disrupting agents by altering the structure of the telomere, leading to inability of telomerase to access the telomere, or to a TL-independent damage signal causing immediate cell arrest or death
TERT promoter-driven suicide gene therapy by delivering suicide genes with promoter regions (TERT and TERC) that tumour cells transcription factors activate
Blocking telomerase expression or biogenesis by down-regulating the amount of functionally active telomerase in a tumour cell, based on the growing understanding of how the telomerase holoenzyme is made, from transcription to post-translational modification, assembly and transport

7.6. Telomeres as a biomarker in Alzheimer's disease

Alzheimer's disease (AD) is considered an age-related pathology for which there is no causal therapy because of the still unknown mechanisms of neuronal cell death. However, it is known that telomere shortening plays an important role in cognitive function in AD. Evidences from studies suggest that AD patients have shorter LTL than age-matched controls (173-177) and that shortening is linked with the pathogenesis of AD *via* oxidative stress and inflammation (174, 178). A variant of the gene *APOE*, which encodes apolipoprotein E, is associated with an increased susceptibility to AD. One study (176) demonstrated the possible association between LTL and the *ApoE4/4* phenotype. Investigation of genetic variants of *hTERT* and *hTERC* genes in relation with AD susceptibility has shown that *hTERT* and *hTERC* genotypes which have been previously associated with reduced *hTERT* expression or shorter LTL, are implicated in AD development. This indicates that telomerase can be directly involved in the pathogenesis of AD (179). Therefore, it seems that TL therapy holds a big potential to become a next attractive drug target for the disease (174), but for this, fundamental knowledge of the role of telomeres in AD still has to be revealed.

8. TELOMERES IN ANTI-CANCER TREATMENT

Telomeres play an important role in cancer development since their length and telomerase activity are crucial for initiation and the survival of tumors, giving to cells the capability of faster and infinitive division. On the other hand, however, critically short telomeres can lead to chromosomal instability and as a result provoke a tumor growth in the affected cell. New insights into molecular regulation of tumorigenesis have significantly improved our understanding of the role of telomerase in cancer progression and provided the opportunities for the development of new diagnostic tools and effective anticancer molecules.

8.1. Telomerase inhibitors

Once the problem of "how the chromosomes can be copied in a complete way during cell divisions and how they are protected against degradation" was solved and awarded the 2009 medicine Nobel prize to three scientists for "adding a new dimension to our understanding of the cell, shedding light on disease mechanisms, and stimulating the development of potential new therapies" (180), extensive research was performed for the development of specific inhibitors of telomerase for targeted anti-cancer drugs. Studies demonstrated that 80-95% of all malignancies showed increased telomerase activity, enabling cancer cells a fast and continued replication and promoting cellular immortality, whereas in normal cells, a very low or almost undetectable telomerase activity was present (181). Therefore, telomerase upregulation is considered to be a critical step in cell tumorigenesis and a major hallmark of cancer (146). Up-regulation is achieved by activation of the normally silent *hTERT* gene, which later creates complexes with other proteins and *hTERC* component. The proposed mechanisms of this activation include mutations in *hTERT* promotor, *hTERT* alternative splicing, gene amplification and epigenetic changes (65). In 10-15% of tumors, the senescence is halted by the ALT mechanism, although telomerase maintains TL in the majority of cancer cells (11). Inhibition of telomerase may therefore lead to a decrease in TL, resulting in cell senescence and apoptosis in telomerase positive tumors, while having a minimal impact on normal cells (19). In this aspect, telomerase seems an excellent target for targeted anti-cancer treatment.

Currently, many anti-telomerase therapeutics are being evaluated for treatment of a variety of cancer types, using different approaches, including the nucleosides, oligonucleotides, small-molecule inhibitors, immunotherapeutic molecules, G-quadruplex stabilizers and gene therapy constructs to inhibit telomerase function (Table 3)(182).

© 1996-2018

8.1.1. Telomerase inhibitor Imelstat (GRN163L)

The most promising group of new agents seems to be the first category that led to the discovery of the so far most promising drug Imetelstat, a specific oligonucleotide telomerase inhibitor. GRN163L was taken into clinical trials after numerous preclinical studies provided evidence of its good pharmacokinetics and pharmacodynamics properties, sufficient potential, efficacy and safety (2). To date, Imetelstat has been the only telomerase-based inhibitor that was sufficiently developed to advance to late-phase clinical trials (183). Its activity was clinically demonstrated in patients with myeloproliferative neoplasms and with solid tumors but caused thrombocytopenia, which is considered as a common side-effect of the treatment (184). Imetelstat is now in the last phase of clinical trials for patients with lower risk of myelodysplastic syndromes and relapsed or refractory myelofibrosis (185).

8.1.2. G-quadruplex stabilisers

Another promising group of widely studied inhibitors are G-quadruplex stabilisers. Telomeres G-strand DNA can form a G4 DNA or G-quadruplex structures, a highly-order structure that is required for a proper regulation of TL *in vivo* by direct inhibition of telomerase binding to long telomere sequences (186). By stabilizing the G-quadruplex with small molecule ligands, the telomeres 3' overhang can be locked in place, thus blocking telomerase from accessing telomeres (19). G4s have become important drug targets that may regulate gene expression and telomere maintenance. Researchers are focusing on new methods to study the ligand binding affinities and selectivity (187) and to propose new bioactive chemotypes, identified after combined ligand-based virtual and experimental screening (188).

8.1.3. Natural products

Many studies have shown that natural phytochemicals can have inhibitory effects on telomerase activity through affecting its subunits and components (189). In traditional Chinese medicine, diverse ingredients with anti-aging effects, acting on different pathways including telomeres and telomerase, are being studied. Natural derivatives, such as astragalosides, polysacharides, flavonoids and others have shown to influence telomerase activity and represent a great potential for further research of anti-cancer drugs in the future (190). Chinese herbal medicine Tianshengyuan-1 (TSY-1) has been recently shown to increase telomerase activity in normal peripheral blood mononuclear cells and CD34+ hematopoietic stem cells with innately low telomerase activity but to decrease telomerase activity in human promyelocytic leukemia (HL60) cells with high intrinsic telomerase activity. Genome screening analysis identified TERT as potential target gene associated with the TSY-1, possibly via the methylation of TERT promoter (191).

8.2. Future perspectives and other therapies

Despite all the progress that has been achieved, the research has shown limited success with only one candidate drug waiting to enter the market (183). Telomerase targeting molecules present a long gap between the administration time and the appearance of a visible effect of producing critically short telomeres in cancer cells (182). This is why most tested therapeutics have shown to be more effective when combined with traditional therapies such as chemotherapy or radiotherapy, resulting in telomere shortening, tumor mass shrinkage and preventing resistance to single agent therapy (19). However, overall benefit is rather limited. One of the reasons might be the regulation of telomerase homeostasis, which is controlled by a complex network of genes having inter-connected signals (191). Nowadays, other targets than telomerase and its regulation are becoming the subject of research, such as ESCRT system, a multi-protein complex, required for the formation of transport vesicles within the cell, critical for safeguarding proper TL maintenance. A fully functional ESCRT system is required for proper telomere homeostasis; mutations in any of the genes encoding the complex could therefore cause telomere shortening. Deletion of ESCRT-0 gene seems to protect the cells from uncapped telomeres, making it suitable as a new anti-cancer drug target (192). Finally, even though an approved therapy based on telomerase inhibition has not yet been proposed, it is still one of the best targets to point out in oncology (182).

9. TELOMERES IN TREATMENT OF OTHER DISEASES

Telomere shortening has been associated with the occurrence of many psychiatric disorders, common aging morbidities, including diabetes, cancer, dys-regulated immune function and multiple aspects of cardiovascular diseases (CVD) as well as with rare monogenetic diseases. It is not surprising that fighting against telomere's shortening became a new considered strategy for prevention of these pathologies. In contrary to most cancers, the therapy that could serve for the treatment of these diseases relies on telomerase as a target for regenerative medicine. The week point, however, are complicated pathways of many diseases and the role of telomeres in it. "Are short telomeres the reason, or the consequence of aging and related diseases?". That is the question that still needs to be answered.

One of the isolated cases of medications in market is TA-65®, a small molecule activator of

© 1996-2018

Table 4. Basic approaches of telomerase-based anti-cancer therapy

Classical gene therapy (by creation of viral vectors) *can be used for tissue engineering, optimization of stem cell transplantation in donor cells with short telomeres and, in principle, also for the treatment of chronic diseases in the whole organism.*
Re-expression of silenced telomerase *(by various molecules and mechanisms such as histone deacetylase inhibitors and estrogen receptor agonists) can upregulate hTERT expression and/or activity.*
Activation of residual enzymatic activity *(by direct interaction with the telomerase holoenzyme or the telomerase activating signaling pathways) is an option for cells with residual telomerase activity to activate the telomerase activity itself.*
Modulation of the intracellular location *(by translocation of telomerase between the nucleus and the cytosol) can regulate telomerase activity.*

telomerase, discovered from screening of natural products from traditional Chinese medicine. Studies of the product showed improvements in biomarkers of aging, including immune, cardiovascular, metabolic, bone, and inflammatory markers and did not show any significant signs of toxicity. A purified compound from *Astragalus membranaceus* plant has been approved as a dietary supplement and is available commercially since 2007. Growing evidences are showing its beneficial role in diverse diseases. Further use of this molecule in disease treatment still calls for long-term prospective studies with larger samples that would reveal its positive effects or possible diverse side-effects (Table 4)(193-195).

10. COMMERCIAL OFFER OF TELOMERE TESTING

Several commercial laboratories over the world are nowadays promoting the TL testing; some of them in purely medical and diagnostic purposes (196), whereas others appeal on tests to reveal the biological aging of human body (197-202). The Canadian company RepeatDx, using Flow-FISH technology to assess TL of blood cells of different type, is offering licensed and validated clinical testing service since 2005 and promoting: *i)* check of possible telomere disorders, such as dyskeratosis congenita, *ii)* bone marrow donors screening to prevent the transplantation of hematopoietic cells with short telomeres, *iii)* testing the patients with pulmonary fibrosis to aid in disease course prediction and risk assessment and finally, *iv)* testing for its use as a biomarker for monitoring other diseases or family history risk. Commercial testing is possible only with the supervision of a practicing licensed physician (196); in the view of the founders of the company RepeatDx, testing outside the context of research studies and diseases related to telomeres is premature for wider public (203).

Other laboratories, such as Telomere Diagnostics, Inc., founded by a Nobel prize winner, Dr. Elizabeth Blackburn, are offering the assessment of age in "TeloYears", which is promoted as an indicator of a healthy lifestyle rather than a tool for screening, diagnosing, treating or preventing diseases or medical conditions (201, 204). Similar is the Spanish company Life Length; for them, telomere testing is a "valuable emerging diagnostic tool within the area of functional and preventive/personalized medicine with a number of clinical applications", such as being global biomarker of health, early detector of age-related diseases, determinant of prognosis and risk stratification (200). The mission of the company Titanovo is to promote longevity science and personalized medicine. Their aim is to "impact longevity research by making important correlations between TL, genotypes, and lifestyle choices". They are offering insight into key biomarkers in combination with longevity genotype (*ACE*, *APOC3*, *APOE*, *FOXO3A*). The aim of the company is also to encourage public to join a research, which "may impact the future of medicine" (202). Furthermore, Veritas, the genome company founded by Dr. George Church is offering telomere testing within the whole genome sequencing insight, where the traits related to disease risk or healthy lifestyle are investigated (197).

TL might easily be widely used as predictor of the biological age of an individual. But the important fact that there is still no reliable solution to repair or stop the process of telomere attrition is diminishing the utility of this testing. Diagnostic laboratories give some advices for individuals with short telomeres, such as proposing "healthy lifestyle and Mediterranean diet" (Cleveland HearthLab) (198) or appealing on the nutritional supplements (quality and balanced multivitamin) and pharmacologic treatments (ACEI, ARB, statins, Aspirin *etc.*) with positive effects on TL (SpectraCell Laboratories) (199). Combination of these advices may help to reduce risk for heart and other chronic diseases or for age-related ailments for which those individuals might be critically susceptible. With the strong scientific teams, which stand behind those companies and other laboratories around the world, it might not be too optimistic to expect that in some time, there will be a new value added to these tests - better understanding of the biomarker and a step towards the solution of premature aging.

11. CONCLUSIONS AND NEW PERSPECTIVES

Years of research have made a big improvement in our knowledge on telomeres biology. The influences of a complex mixture of genetics and environment are combined within the simple

© 1996-2018

The future of telomere length in personalized medicine

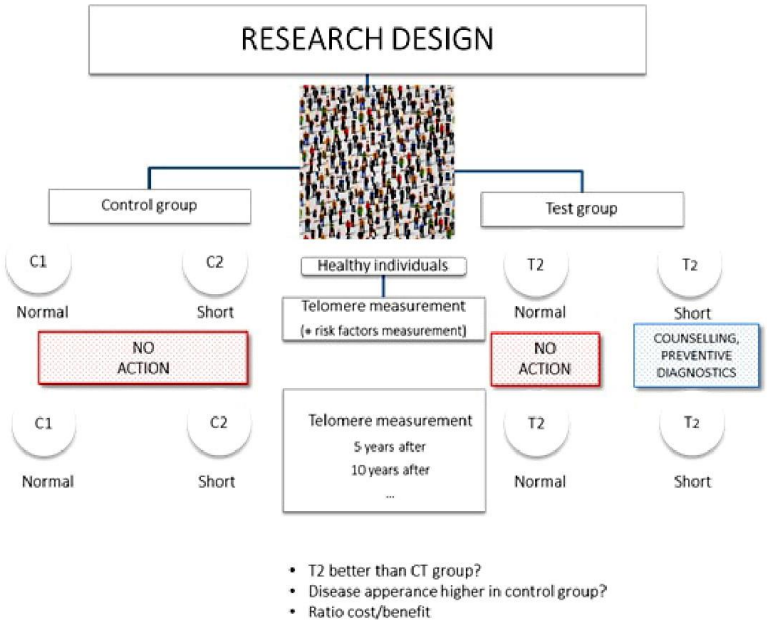


Figure 3 Presentation of the study design aiming to evaluate the potential benefit of TL measurement.

repeated structure TTAGGG, which is lively reacting to diverse factors of biology and society. Hundreds of studies have confirmed the correlation of telomeres with diseases, genetics and lifestyle. But how can we finally take the advantage of all this knowledge and put it into a successful clinical practice? The use of telomeres as a treatment target for diverse diseases is slowly making its progress. Even though there has not yet been an EMA or FDA approved medication that would act on telomeres, we might expect first results in following years, most probably in anticancer therapy, where tested drugs are approaching to the final steps of validation before their release in the market. Dietary supplements that are already being used as natural stimulators of the telomerase activity, such as TA-65®, still have to be further assessed for their beneficial effects and potential adverse effects. Moreover, gene therapy is showing a high potential; bigger progress in this field will open new possibilities also in treatment of telomere-related genes that are being involved in the onset of many chronological diseases.

GWAS studies gave a new insight in genetics and several SNPs were found to be correlated with TL, mostly within genes coding for telomere-related proteins. Rare variants of these genes can provoke diseases called “telomere syndrome”. Telomere measurement or genotyping are one of the most reliable methods of diagnosing these diseases. Furthermore, several candidate genes with function in telomere biology showed association with other age-related diseases. Further studies of ‘omics’ regulation of telomeres will contribute to better understanding of regulatory pathways which is crucial for successful drug design.

“Is TL a biomarker of … aging? Coronary artery disease? Breast cancer?” are the common questions that scientists tried to answer. Despite all hard work, so far, TL has not fulfilled all of the conditions for being considered as a “biomarker” for diagnostics of disease conditions or in predicting the “biological age” of individual. The reasons for this are many individual factors that are influencing TL and are preventing the identification of a clear casual relation between condition itself and telomeres. Next, there is an inconsistence of measuring methods of TL determination, used by different laboratories, another weak point of biomarker that we cannot neglect. However, the connection between TL and many diseases and risk factors has been proved in many different studies, suggesting a beneficial indicative ability of telomeres in disease prognostic.

© 1996-2018

Now it is time for new perspectives that will accelerate the development of diagnostic methods with TL measurement.

To summarize, TL should not be tested as a biomarker of the single disease condition, but as a general state of organism that reveals potential susceptibility for all-disease risk. Measuring of TL could present a follow-up of the patient's health and create an alert of possible disease setup, if patient would be classified in the lowest (or in some cases highest) quartile of TL of his age-group or would expose faster rate of telomere shorting than usual. The patients at risk would have further testing for particular diseases and additional counseling about the risk factors that might threaten their health from their personal doctor. As we know that changes in lifestyle may significantly decrease TL attrition and improve general health of individual, this might prevent or delay disease onset and ameliorate medical conditions. If the progress of telomere-related drugs will evaluate, a premature treatment will also be possible. In order to make a step further in application of telomeres into a personalized medicine, large prospective studies should be done, assessing the ratio between the cost-benefit of measuring TL in everyday clinical practice (Figure 3). Furthermore, with the advancements in the field of bioinformatics, WGS will become another future diagnostics tool. With WGS, TL determination would be possible without "classical methods" but with computational calculation only. This could become a new reference method for TL determination. Further research in this field, however, is essential and the establishment of unique, well performative technique is indispensable for faster progress in telomere research.

12. REFERENCES

1. K. Jager and M. Walter: Therapeutic Targeting of Telomerase. *Genes (Basel)*, 7(7) (2016)

2. C. B. Harley: Telomerase and cancer therapeutics. *Nat Rev Cancer*, 8(3), 167-79 (2008)

3. E. H. Blackburn: Structure and function of telomeres. *Nature*, 350, 569-573 (1991)

4. E. H. Blackburn: Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS letters*, 579, 859-862 (2005)

5. R. C. Allsopp, E. Chang, M. Kashefi-Aazam, E. I. Rogaev, M. A. Piatyszek, J. W. Shay and C. B. Harley: Telomere shortening is associated with cell division *in vitro* and *in vivo*. *Exp Cell Res*, 220(1), 194-200 (1995)

6. E. H. Blackburn, E. S. Epel and J. Lin: Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science (New York, N.Y.)*, 350, 1193-1198 (2015)

7. F. d'Adda di Fagagna, P. M. Reaper, L. Clay-Farrace, H. Fiegler, P. Carr, T. Von Zglinicki, G. Saretzki, N. P. Carter and S. P. Jackson: A DNA damage checkpoint response in telomere-initiated senescence. *Nature*, 426, 194-198 (2003)

8. S. Bekaert, T. D. Meyer and P. V. Oostveldt: Telomere Attrition as Ageing Biomarker. *Anticancer Research*, 25, 3011-3021 (2005)

9. J. L. Sanders and A. B. Newman: Telomere length in epidemiology: a biomarker of aging, age-related disease, both, or neither? *Epidemiologic Reviews*, 35, 112-131 (2013)

10. S. R. W. L. Chan and E. H. Blackburn: Telomeres and telomerase. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 109-121 (2004)

11. M. A. Jafri, S. A. Ansari, M. H. Alqahtani and J. W. Shay: Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med*, 8(1), 69 (2016)

12. K. A. Mattern, S. J. J. Swiggers, A. L. Nigg, B. Löwenberg, A. B. Houtsmuller and J. M. J. M. Zijlmans: Dynamics of Protein Binding to Telomeres in Living Cells: Implications for Telomere Structure and Function. *Molecular and Cellular Biology*, 24, 5587-5594 (2004)

13. T. d. Lange: Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & Development*, 19, 2100-2110 (2005)

14. V. L. Makarov, Y. Hirose and J. P. Langmore: Long G Tails at Both Ends of Human Chromosomes Suggest a C Strand Degradation Mechanism for Telomere Shortening. *Cell*, 88, 657-666 (1997)

15. J. D. Griffith, L. Comeau, S. Rosenfield, R. M. Stansel, A. Bianchi, H. Moss and T. de Lange: Mammalian telomeres end in a large duplex loop. *Cell*, 97, 503-514 (1999)

16. E. H. Blackburn: Switching and signaling at the telomere. *Cell*, 106, 661-673 (2001)

© 1996-2018

17. T. Nezu, N. Hosomi, T. Takahashi, K. Anno, S. Aoki, A. Shimamoto, H. Maruyama, T. Hayashi, M. Matsumoto and H. Tahara: Telomere G-tail Length is a Promising Biomarker Related to White Matter Lesions and Endothelial Dysfunction in Patients With Cardiovascular Risk: A Cross-sectional Study. *EBioMedicine*, 2, 960-967 (2015)

18. H. Seimiya: Predicting Risk at the End of the End: Telomere G-tail as a Biomarker. *EBioMedicine*, 2, 804-805 (2015)

19. M. Ruden and N. Puri: Novel anticancer therapeutics targeting telomerase. *Cancer Treat Rev*, 39(5), 444-56 (2013)

20. T. M. Nakamura and T. R. Cech: Reversing Time: Origin of Telomerase. *Cell*, 92, 587-590 (1998)

21. G. Sarek, P. Marzec, P. Margalef and S. J. Boulton: Molecular basis of telomere dysfunction in human genetic diseases. *Nat Struct Mol Biol*, 22(11), 867-74 (2015)

22. Y.-S. Cong, W. E. Wright and J. W. Shay: Human Telomerase and Its Regulation. *Microbiology and Molecular Biology Reviews*, 66, 407-425 (2002)

23. W. E. Wright, M. A. Piatyszek, W. E. Rainey, W. Byrd and J. W. Shay: Telomerase activity in human germline and embryonic tissues and cells. *Developmental Genetics*, 18, 173-179 (1996)

24. E. H. Blackburn: Telomere states and cell fates. *Nature*, 408(6808), 53-6 (2000)

25. M. A. Dunham, A. A. Neumann, C. L. Fasching and R. R. Reddel: Telomere maintenance by recombination in human cells. *Nature Genetics*, 26, 447-450 (2000)

26. C. M. Roake and S. E. Artandi: DNA repair: Telomere-lengthening mechanism revealed. *Nature*, 539, 35-36 (2016)

27. W. C. Hahn, S. A. Stewart, M. W. Brooks, S. G. York, E. Eaton, A. Kurachi, R. L. Beijersbergen, J. H. M. Knoll, M. Meyerson and R. A. Weinberg: Inhibition of telomerase limits the growth of human cancer cells. *Nature Medicine*, 5, 1164-1170 (1999)

28. M. Armanios and E. H. Blackburn: The telomere syndromes. *Nat Rev Genet*, 13(10), 693-704 (2012)

29. C. Bar and M. A. Blasco: Telomeres and telomerase as therapeutic targets to prevent and treat age-related diseases. *F1000Res*, 5 (2016)

30. C. Bar, J. M. Povedano, R. Serrano, C. Benitez-Buelga, M. Popkes, I. Formentini, M. Bobadilla, F. Bosch and M. A. Blasco: Telomerase gene therapy rescues telomere length, bone marrow aplasia, and survival in mice with aplastic anemia. *Blood*, 127(14), 1770-9 (2016)

31. I. Mender and J. W. Shay: Telomere Restriction Fragment (TRF) Analysis. *Bio Protoc*, 5(22) (2015)

32. R. M. Cawthon: Telomere measurement by quantitative PCR. *Nucleic Acids Res*, 30(10), e47 (2002)

33. G. M. Baerlocher, I. Vulto, G. de Jong and P. M. Lansdorp: Flow cytometry and FISH to measure the average length of telomeres (flow FISH). *Nat Protoc*, 1(5), 2365-76 (2006)

34. M. Lee, C. E. Napier, S. F. Yang, J. W. Arthur, R. R. Reddel and H. A. Pickett: Comparative analysis of whole genome sequencing-based telomere length measurement techniques. *Methods*, 114, 4-15 (2017)

35. Z. Ding, M. Mangino, A. Aviv, T. Spector, R. Durbin and U. K. Consortium: Estimating telomere length from whole genome sequence data. *Nucleic Acids Res*, 42(9), e75 (2014)

36. A. Hakobyan, L. Nersisyan and A. Arakelyan: Quantitative trait association study for mean telomere length in the South Asian genomes. *Bioinformatics*, 32(11), 1697-700 (2016)

37. M. Harland, M. Petljak, C. D. Robles-Espinoza, Z. Ding, N. A. Gruis, R. van Doorn, K. A. Pooley, A. M. Dunning, L. G. Aoude, K. A. Wadt, A. M. Gerdes, K. M. Brown, N. K. Hayward, J. A. Newton-Bishop, D. J. Adams and D. T. Bishop: Germline TERT promoter mutations are rare in familial melanoma. *Fam Cancer*, 15(1), 139-44 (2016)

38. M. Lee, M. Hills, D. Conomos, M. D. Stutz, R. A. Dagg, L. M. Lau, R. R. Reddel and H. A. Pickett: Telomere extension by telomerase and ALT generates variant repeats by

© 1996-2018

mechanistically distinct processes. *Nucleic Acids Res*, 42(3), 1733-46 (2014)

39. L. Nersisyan and A. Arakelyan: CompuTel: computation of mean telomere length from whole-genome next-generation sequencing data. *PLoS One*, 10(4), e0125201 (2015)

40. N. Cai, S. Chang, Y. Li, Q. Li, J. Hu, J. Liang, L. Song, W. Kretzschmar, X. Gan, J. Nicod, M. Rivera, H. Deng, B. Du, K. Li, W. Sang, J. Gao, S. Gao, B. Ha, H.-Y. Ho, C. Hu, J. Hu, Z. Hu, G. Huang, G. Jiang, T. Jiang, W. Jin, G. Li, K. Li, Y. Li, Y. Li, Y. Li, Y.-T. Lin, L. Liu, T. Liu, Y. Liu, Y. Liu, Y. Lu, L. Lv, H. Meng, P. Qian, H. Sang, J. Shen, J. Shi, J. Sun, M. Tao, G. Wang, G. Wang, J. Wang, L. Wang, X. Wang, X. Wang, H. Yang, L. Yang, Y. Yin, J. Zhang, K. Zhang, N. Sun, W. Zhang, X. Zhang, Z. Zhang, H. Zhong, G. Breen, J. Wang, J. Marchini, Y. Chen, Q. Xu, X. Xu, R. Mott, G.-J. Huang, K. Kendler and J. Flint: Molecular signatures of major depression. *Current biology: CB*, 25, 1146-1156 (2015)

41. L. Daniali, A. Benetos, E. Susser, J. D. Kark, C. Labat, M. Kimura, K. Desai, M. Granick and A. Aviv: Telomeres shorten at equivalent rates in somatic tissues of adults. *Nat Commun*, 4, 1597 (2013)

42. A. M. Olovnikov: A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J Theor Biol*, 41(1), 181-90 (1973)

43. J. D. Watson: Origin of concatemeric T7 DNA. *Nat New Biol*, 239(94), 197-201 (1972)

44. L. Hayflick and P. S. Moorhead: The serial cultivation of human diploid cell strains. *Exp Cell Res*, 25, 585-621 (1961)

45. P. Boukamp: Ageing mechanisms: the role of telomere loss. *Clin Exp Dermatol*, 26(7), 562-5 (2001)

46. H. Cherif, J. L. Tarry, S. E. Ozanne and C. N. Hales: Ageing and telomeres: a study into organ- and gender-specific telomere shortening. *Nucleic Acids Res*, 31(5), 1576-83 (2003)

47. H. Jiang, Z. Ju and K. L. Rudolph: Telomere shortening and ageing. *Z Gerontol Geriatr*, 40(5), 314-24 (2007)

48. J. Zhang, G. Rane, X. Dai, M. K. Shanmugam, F. Arfuso, R. P. Samy, M. K. Lai, D. Kappei, A. P. Kumar and G. Sethi: Ageing and the telomere connection: An intimate relationship with inflammation. *Ageing Res Rev*, 25, 55-69 (2016)

49. M. Gardner, D. Bann, L. Wiley, R. Cooper, R. Hardy, D. Nitsch, C. Martin-Ruiz, P. Shiels, A. A. Sayer, M. Barbieri, S. Bekaert, C. Bischoff, A. Brooks-Wilson, W. Chen, C. Cooper, K. Christensen, T. De Meyer, I. Deary, G. Der, A. Diez Roux, A. Fitzpatrick, A. Hajat, J. Halaschek-Wiener, S. Harris, S. C. Hunt, C. Jagger, H. S. Jeon, R. Kaplan, M. Kimura, P. Lansdorp, C. Li, T. Maeda, M. Mangino, T. S. Nawrot, P. Nilsson, K. Nordfjall, G. Paolisso, F. Ren, K. Riabowol, T. Robertson, G. Roos, J. A. Staessen, T. Spector, N. Tang, B. Unryn, P. van der Harst, J. Woo, C. Xing, M. E. Yadegarfar, J. Y. Park, N. Young, D. Kuh, T. von Zglinicki and Y. Ben-Shlomo: Gender and telomere length: systematic review and meta-analysis. *Exp Gerontol*, 51, 15-27 (2014)

50. M. E. B. Hansen, S. C. Hunt, R. C. Stone, K. Horvath, U. Herbig, A. Ranciaro, J. Hirbo, W. Beggs, A. P. Reiner, J. G. Wilson, M. Kimura, I. De Vivo, M. M. Chen, J. D. Kark, D. Levy, T. Nyambo, S. A. Tishkoff and A. Aviv: Shorter telomere length in Europeans than in Africans due to polygenetic adaptation. *Human Molecular Genetics*, 25, 2324-2330 (2016)

51. S. M. Lynch, M. K. Peek, N. Mitra, K. Ravichandran, C. Branas, E. Spangler, W. Zhou, E. D. Paskett, S. Gehlert, C. DeGraffinreid, T. R. Rebbeck and H. Riethman: Race, Ethnicity, Psychosocial Factors, and Telomere Length in a Multicenter Setting. *PLoS One*, 11(1), e0146723 (2016)

52. A. Muezzinler, U. Mons, A. K. Dieffenbach, K. Butterbach, K. U. Saum, M. Schick, H. Stammer, P. Boukamp, B. Holleczek, C. Stegmaier and H. Brenner: Smoking habits and leukocyte telomere length dynamics among older adults: Results from the ESTHER cohort. *Exp Gerontol*, 70, 18-25 (2015)

53. D. Sassenroth, A. Meyer, B. Salewsky, M. Kroh, K. Norman, E. Steinhagen-Thiessen and I. Demuth: Sports and Exercise at Different Ages and Leukocyte Telomere Length in Later Life--Data from the Berlin

© 1996-2018

Aging Study II (BASE-II). *PLoS One*, 10(12), e0142131 (2015)

54. A. M. Valdes, T. Andrew, J. P. Gardner, M. Kimura, E. Oelsner, L. F. Cherkas, A. Aviv and T. D. Spector: Obesity, cigarette smoking, and telomere length in women. *Lancet*, 366(9486), 662-4 (2005)

55. B. M. Coimbra, C. M. Carvalho, P. N. Moretti, M. F. Mello and S. I. Belangero: Stress-related telomere length in children: A systematic review. *J Psychiatr Res*, 92, 47-54 (2017)

56. G. W. Edmonds, S. E. Hampson, H. C. F. Côté, P. L. Hill and B. Klest: Childhood Personality, Betrayal Trauma, and Leukocyte Telomere Length in Adulthood: A Lifespan Perspective on Conscientiousness and Betrayal Traumas as Predictors of a Biomarker of Cellular Ageing. *European Journal of Personality*, 30, 426-437 (2016)

57. D. Lindqvist, E. S. Epel, S. H. Mellon, B. W. Penninx, D. Révész, J. E. Verhoeven, V. I. Reus, J. Lin, L. Mahan, C. M. Hough, R. Rosser, F. S. Bersani, E. H. Blackburn and O. M. Wolkowitz: Psychiatric disorders and leukocyte telomere length: Underlying mechanisms linking mental illness with cellular aging. *Neuroscience and Biobehavioral Reviews*, 55, 333-364 (2015)

58. L. H. Price, H.-T. Kao, D. E. Burgers, L. L. Carpenter and A. R. Tyrka: Telomeres and early-life stress: an overview. *Biological Psychiatry*, 73, 15-23 (2013)

59. J. L. Buxton, R. G. Walters, S. Visvikis-Siest, D. Meyre, P. Froguel and A. I. F. Blakemore: Childhood Obesity Is Associated with Shorter Leukocyte Telomere Length. *The Journal of Clinical Endocrinology and Metabolism*, 96, 1500-1505 (2011)

60. L. Ma, Y. Li and J. Wang: Telomeres and essential hypertension. *Clin Biochem*, 48(16-17), 1195-9 (2015)

61. S. Rizvi, S. T. Raza and F. Mahdi: Telomere length variations in aging and age-related diseases. *Current Aging Science*, 7, 161-167 (2014)

62. M. A. Blasco: The epigenetic regulation of mammalian telomeres. *Nature Reviews Genetics*, 8, 299-309 (2007)

63. M. García-Cao, R. O'Sullivan, A. H. F. M. Peters, T. Jenuwein and M. A. Blasco: Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nature Genetics*, 36, 94-99 (2004)

64. D. Melicher, E. I. Buzas and A. Falus: Genetic and epigenetic trends in telomere research: a novel way in immunoepigenetics. *Cellular and molecular life sciences: CMLS*, 72, 4095-4109 (2015)

65. J. W. Shay: Are short telomeres predictive of advanced cancer? *Cancer Discov*, 3(10), 1096-8 (2013)

66. V. Kordinas, A. Ioannidis and S. Chatzipanagiotou: The Telomere/Telomerase System in Chronic Inflammatory Diseases. Cause or Effect? *Genes*, 7 (2016)

67. T. von Zglinicki: Oxidative stress shortens telomeres. *Trends in Biochemical Sciences*, 27, 339-344 (2002)

68. P. T. King: Inflammation in chronic obstructive pulmonary disease and its role in cardiovascular disease and lung cancer. *Clin Transl Med*, 4(1), 68 (2015)

69. T. Kuwabara, F. Ishikawa, M. Kondo and T. Kakiuchi: The Role of IL-17 and Related Cytokines in Inflammatory Autoimmune Diseases. *Mediators Inflamm*, 2017, 3908061 (2017)

70. P. Libby: Inflammation in atherosclerosis. *Nature*, 420(6917), 868-74 (2002)

71. S. F. Moss and M. J. Blaser: Mechanisms of disease: Inflammation and the origins of cancer. *Nat Clin Pract Oncol*, 2(2), 90-7; quiz 1 p following 113 (2005)

72. A. D. Pradhan, J. E. Manson, N. Rifai, J. E. Buring and P. M. Ridker: C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. *JAMA*, 286(3), 327-34 (2001)

73. L. G. Schuettpelz and D. C. Link: Regulation of hematopoietic stem cell activity by inflammation. *Front Immunol*, 4, 204 (2013)

74. T. Wyss-Coray: Inflammation in Alzheimer disease: driving force, bystander or beneficial response? *Nat Med*, 12(9), 1005-15 (2006)

© 1996-2018

75. R. T. Calado and N. S. Young: Telomere Diseases. *The New England Journal of Medicine*, 361, 2353-2365 (2009)

76. J. Y. Y. Wong, I. De Vivo, X. Lin, S. C. Fang and D. C. Christiani: The relationship between inflammatory biomarkers and telomere length in an occupational prospective cohort study. *PloS One*, 9, e87348 (2014)

77. A. O'Donovan, M. S. Pantell, E. Puterman, F. S. Dhabhar, E. H. Blackburn, K. Yaffe, R. M. Cawthon, P. L. Opresko, W.-C. Hsueh, S. Satterfield, A. B. Newman, H. N. Ayonayon, S. M. Rubin, T. B. Harris and E. S. Epel: Cumulative Inflammatory Load Is Associated with Short Leukocyte Telomere Length in the Health, Aging and Body Composition Study. *PLoS ONE*, 6 (2011)

78. A. L. Fitzpatrick, R. A. Kronmal, J. P. Gardner, B. M. Psaty, N. S. Jenny, R. P. Tracy, J. Walston, M. Kimura and A. Aviv: Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am J Epidemiol*, 165(1), 14-21 (2007)

79. A. Aviv, A. Valdes, J. P. Gardner, R. Swaminathan, M. Kimura and T. D. Spector: Menopause modifies the association of leukocyte telomere length with insulin resistance and inflammation. *J Clin Endocrinol Metab*, 91(2), 635-40 (2006)

80. J. B. Richards, A. M. Valdes, J. P. Gardner, B. S. Kato, A. Siva, M. Kimura, X. Lu, M. J. Brown, A. Aviv and T. D. Spector: Homocysteine levels and leukocyte telomere length. *Atherosclerosis*, 200(2), 271-7 (2008)

81. J. B. Richards, A. M. Valdes, J. P. Gardner, D. Paximadas, M. Kimura, A. Nessa, X. Lu, G. L. Surdulescu, R. Swaminathan, T. D. Spector and A. Aviv: Higher serum vitamin D concentrations are associated with longer leukocyte telomere length in women. *Am J Clin Nutr*, 86(5), 1420-5 (2007)

82. D. C. Pedroso, C. L. Miranda-Furtado, G. S. Kogure, J. Meola, M. Okuka, C. Silva, R. T. Calado, R. A. Ferriani, D. L. Keefe and R. M. dos Reis: Inflammatory biomarkers and telomere length in women with polycystic ovary syndrome. *Fertil Steril*, 103(2), 542-7 e2 (2015)

83. L. Savale, A. Chaouat, S. Bastuji-Garin, E. Marcos, L. Boyer, B. Maitre, M. Sarni, B. Housset, E. Weitzenblum, M. Matrat, P. Le Corvoisier, D. Rideau, J. Boczkowski, J. L. Dubois-Rande, C. Chouaid and S. Adnot: Shortened telomeres in circulating leukocytes of patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 179(7), 566-71 (2009)

84. Y. W. Kim, X. Z. West and T. V. Byzova: Inflammation and oxidative stress in angiogenesis and vascular disease. *J Mol Med (Berl)*, 91(3), 323-8 (2013)

85. A. Hoeben, B. Landuyt, M. S. Highley, H. Wildiers, A. T. Van Oosterom and E. A. De Bruijn: Vascular endothelial growth factor and angiogenesis. *Pharmacol Rev*, 56(4), 549-80 (2004)

86. M. Azimi-Nezhad, M. G. Stathopoulou, A. Bonnefond, M. Rancier, A. Saleh, J. Lamont, P. Fitzgerald, N. C. Ndiaye and S. Visvikis-Siest: Associations of vascular endothelial growth factor (VEGF) with adhesion and inflammation molecules in a healthy population. *Cytokine*, 61(2), 602-7 (2013)

87. G. Zaccagnini, C. Gaetano, L. Della Pietra, S. Nanni, A. Grasselli, A. Mangoni, R. Benvenuto, M. Fabrizi, S. Truffa, A. Germani, F. Moretti, A. Pontecorvi, A. Sacchi, S. Bacchetti, M. C. Capogrossi and A. Farsetti: Telomerase mediates vascular endothelial growth factor-dependent responsiveness in a rat model of hind limb ischemia. *J Biol Chem*, 280(15), 14790-8 (2005)

88. N. Liu, D. Ding, W. Hao, F. Yang, X. Wu, M. Wang, X. Xu, Z. Ju, J. P. Liu, Z. Song, J. W. Shay, Y. Guo and Y. S. Cong: hTERT promotes tumor angiogenesis by activating VEGF via interactions with the Sp1 transcription factor. *Nucleic Acids Res*, 44(18), 8693-8703 (2016)

89. T. Poonpet, N. Saetan, A. Tanavalee, V. Wilairatana, P. Yuktanandana and S. Honsawek: Association between leukocyte telomere length and angiogenic cytokines in knee osteoarthritis. *Int J Rheum Dis* (2017)

90. I. Dalle-Donne, R. Rossi, R. Colombo, D. Giustarini and A. Milzani: Biomarkers of oxidative damage in human disease. *Clin Chem*, 52(4), 601-23 (2006)

91. M. Valko, D. Leibfritz, J. Moncol, M. T. Cronin, M. Mazur and J. Telser: Free

© 1996-2018

radicals and antioxidants in normal physiological functions and human disease. *Int J Biochem Cell Biol*, 39(1), 44-84 (2007)

92. L. A. Pham-Huy, H. He and C. Pham-Huy: Free radicals, antioxidants in disease and health. *Int J Biomed Sci*, 4(2), 89-96 (2008)

93. J. M. Houben, H. J. Moonen, F. J. van Schooten and G. J. Hageman: Telomere length assessment: biomarker of chronic oxidative stress? *Free Radic Biol Med*, 44(3), 235-46 (2008)

94. N. Sitte, G. Saretzki and T. von Zglinicki: Accelerated telomere shortening in fibroblasts after extended periods of confluency. *Free Radic Biol Med*, 24(6), 885-93 (1998)

95. S. Oikawa, S. Tada-Oikawa and S. Kawanishi: Site-specific DNA damage at the GGG sequence by UVA involves acceleration of telomere shortening. *Biochemistry*, 40(15), 4763-8 (2001)
DOI: 10.1021/bi002721g

96. S. Kawanishi and S. Oikawa: Mechanism of telomere shortening by oxidative stress. *Ann N Y Acad Sci*, 1019, 278-84 (2004)
DOI: 10.1196/annals.1297.047

97. S. Petersen, G. Saretzki and T. von Zglinicki: Preferential accumulation of single-stranded regions in telomeres of human fibroblasts. *Exp Cell Res*, 239(1), 152-60 (1998)
DOI: 10.1006/excr.1997.3893

98. L. Gonzalez-Guardia, E. M. Yubero-Serrano, O. Rangel-Zuniga, C. Marin, A. Camargo, P. Perez-Martinez, J. Delgado-Lista, F. Gomez-Delgado, A. Garcia-Rios, F. J. Tinahones, H. M. Roche, F. Perez-Jimenez and J. Lopez-Miranda: Influence of endothelial dysfunction on telomere length in subjects with metabolic syndrome: LIPGENE study. *Age (Dordr)*, 36(4), 9681 (2014)
DOI: 10.1007/s11357-014-9681-9

99. S. Demissie, D. Levy, E. J. Benjamin, L. A. Cupples, J. P. Gardner, A. Herbert, M. Kimura, M. G. Larson, J. B. Meigs, J. F. Keaney and A. Aviv: Insulin resistance, oxidative stress, hypertension, and leukocyte telomere length in men from the Framingham Heart Study. *Aging Cell*, 5(4), 325-30 (2006)
DOI: 10.1111/j.1474-9726.2006.00224.x

100. D. Ma, W. Zhu, S. Hu, X. Yu and Y. Yang: Association between oxidative stress and telomere length in Type 1 and Type 2 diabetic patients. *J Endocrinol Invest*, 36(11), 1032-7 (2013)

101. Z. Liu, J. Zhang, J. Yan, Y. Wang and Y. Li: Leucocyte telomere shortening in relation to newly diagnosed type 2 diabetic patients with depression. *Oxid Med Cell Longev*, 2014, 673959 (2014)
DOI: 10.1155/2014/673959

102. D. A. Glei, N. Goldman, M. Weinstein and R. A. Risques: Shorter Ends, Faster End? Leukocyte Telomere Length and Mortality Among Older Taiwanese. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 70, 1490-1498 (2015)
DOI: 10.1093/gerona/glu191

103. T. De Meyer, E. R. Rietzschel, M. L. De Buyzere, W. Van Criekinge and S. Bekaert: Studying telomeres in a longitudinal population based study. *Front Biosci*, 13, 2960-70 (2008)
DOI: 10.2741/2901

104. T. Andrew, A. Aviv, M. Falchi, G. L. Surdulescu, J. P. Gardner, X. Lu, M. Kimura, B. S. Kato, A. M. Valdes and T. D. Spector: Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. *American Journal of Human Genetics*, 78, 480-486 (2006)
DOI: 10.1086/500052

105. G. Atzmon, M. Cho, R. M. Cawthon, T. Budagov, M. Katz, X. Yang, G. Siegel, A. Bergman, D. M. Huffman, C. B. Schechter, W. E. Wright, J. W. Shay, N. Barzilai, D. R. Govindaraju and Y. Suh: Evolution in health and medicine Sackler colloquium: Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. *Proceedings of the National Academy of Sciences of the United States of America*, 107 Suppl 1, 1710-1717 (2010)
DOI: 10.1073/pnas.0906191106

106. S. L. Bakaysa, L. A. Mucci, P. E. Slagboom, D. I. Boomsma, G. E. McClearn, B. Johansson and N. L. Pedersen: Telomere length predicts survival independent of genetic influences. *Aging Cell*, 6, 769-774 (2007)
DOI: 10.1111/j.1474-9726.2007.00340.x

© 1996-2018

107. M. Vasa-Nicotera, S. Brouilette, M. Mangino, J. R. Thompson, P. Braund, J.-R. Clemitson, A. Mason, C. L. Bodycote, S. M. Raleigh, E. Louis and N. J. Samani: Mapping of a major locus that determines telomere length in humans. *American Journal of Human Genetics*, 76, 147-151 (2005)
DOI: 10.1086/426734

108. P. E. Slagboom, S. Droog and D. I. Boomsma: Genetic determination of telomere size in humans: a twin study of three age groups. *American Journal of Human Genetics*, 55, 876-882 (1994)

109. O. T. Njajou, R. M. Cawthon, C. M. Damcott, S.-H. Wu, S. Ott, M. J. Garant, E. H. Blackburn, B. D. Mitchell, A. R. Shuldiner and W.-C. Hsueh: Telomere length is paternally inherited and is associated with parental lifespan. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 12135-12139 (2007)
DOI: 10.1073/pnas.0702703104

110. M. Kimura, L. F. Cherkas, B. S. Kato, S. Demissie, J. B. Hjelmborg, M. Brimacombe, A. Cupples, J. L. Hunkin, J. P. Gardner, X. Lu, X. Cao, M. Sastrasinh, M. A. Province, S. C. Hunt, K. Christensen, D. Levy, T. D. Spector and A. Aviv: Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. *PLoS Genet*, 4(2), e37 (2008)
DOI: 10.1371/journal.pgen.0040037

111. T. De Meyer and D. T. Eisenberg: Possible technical and biological explanations for the 'parental telomere length inheritance discrepancy' enigma. *Eur J Hum Genet*, 23(1), 3-4 (2015)
DOI: 10.1038/ejhg.2014.65

112. D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff and H. Parkinson: The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*, 42(Database issue), D1001-6 (2014)
DOI: 10.1093/nar/gkt1229

113. V. Codd, C. P. Nelson, E. Albrecht, M. Mangino, J. Deelen, J. L. Buxton, J. J. Hottenga, K. Fischer, T. Esko, I. Surakka, L. Broer, D. R. Nyholt, I. Mateo Leach, P. Salo, S. Hägg, M. K. Matthews, J. Palmen, G. D. Norata, P. F. O'Reilly, D. Saleheen, N. Amin, A. J. Balmforth, M. Beekman, R. A. de Boer, S. Böhringer, P. S. Braund, P. R. Burton, A. J. M. de Craen, M. Denniff, Y. Dong, K. Douroudis, E. Dubinina, J. G. Eriksson, K. Garlaschelli, D. Guo, A.-L. Hartikainen, A. K. Henders, J. J. Houwing-Duistermaat, L. Kananen, L. C. Karssen, J. Kettunen, N. Klopp, V. Lagou, E. M. van Leeuwen, P. A. Madden, R. Mägi, P. K. E. Magnusson, S. Männistö, M. I. McCarthy, S. E. Medland, E. Mihailov, G. W. Montgomery, B. A. Oostra, A. Palotie, A. Peters, H. Pollard, A. Pouta, I. Prokopenko, S. Ripatti, V. Salomaa, H. E. D. Suchiman, A. M. Valdes, N. Verweij, A. Viñuela, X. Wang, H.-E. Wichmann, E. Widen, G. Willemsen, M. J. Wright, K. Xia, X. Xiao, D. J. van Veldhuisen, A. L. Catapano, M. D. Tobin, A. S. Hall, A. I. F. Blakemore, W. H. van Gilst, H. Zhu, C. consortium, J. Erdmann, M. P. Reilly, S. Kathiresan, H. Schunkert, P. J. Talmud, N. L. Pedersen, M. Perola, W. Ouwehand, J. Kaprio, N. G. Martin, C. M. van Duijn, I. Hovatta, C. Gieger, A. Metspalu, D. I. Boomsma, M.-R. Jarvelin, P. E. Slagboom, J. R. Thompson, T. D. Spector, P. van der Harst and N. J. Samani: Identification of seven loci affecting mean telomere length and their association with disease. *Nature Genetics*, 45, 422-427, 427e1-2 (2013)

114. J. Gu, M. Chen, S. Shete, C. I. Amos, A. Kamat, Y. Ye, J. Lin, C. P. Dinney and X. Wu: A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. *Cancer Prevention Research (Philadelphia, Pa.)*, 4, 514-521 (2011)
DOI: 10.1158/1940-6207.CAPR-11-0063

115. D. Levy, S. L. Neuhausen, S. C. Hunt, M. Kimura, S.-J. Hwang, W. Chen, J. C. Bis, A. L. Fitzpatrick, E. Smith, A. D. Johnson, J. P. Gardner, S. R. Srinivasan, N. Schork, J. I. Rotter, U. Herbig, B. M. Psaty, M. Sastrasinh, S. S. Murray, R. S. Vasan, M. A. Province, N. L. Glazer, X. Lu, X. Cao, R. Kronmal, M. Mangino, N. Soranzo, T. D. Spector, G. S. Berenson and A. Aviv: Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 9293-9298 (2010)
DOI: 10.1073/pnas.0911494107

116. Y. Liu, L. Cao, Z. Li, D. Zhou, W. Liu, Q. Shen, Y. Wu, D. Zhang, X. Hu, T. Wang, J. Ye, X. Weng, H. Zhang, D. Zhang, Z.

© 1996-2018

Zhang, F. Liu, L. He and Y. Shi: A genome-wide association study identifies a locus on TERT for mean telomere length in Han Chinese. *PloS One*, 9, e85043 (2014) DOI: 10.1371/journal.pone.0085043

117. M. Mangino, S.-J. Hwang, T. D. Spector, S. C. Hunt, M. Kimura, A. L. Fitzpatrick, L. Christiansen, I. Petersen, C. C. Elbers, T. Harris, W. Chen, S. R. Srinivasan, J. D. Kark, A. Benetos, S. El Shamieh, S. Visvikis-Siest, K. Christensen, G. S. Berenson, A. M. Valdes, A. Viñuela, M. Garcia, D. K. Arnett, U. Broeckel, M. A. Province, J. S. Pankow, C. Kammerer, Y. Liu, M. Nalls, S. Tishkoff, F. Thomas, E. Ziv, B. M. Psaty, J. C. Bis, J. I. Rotter, K. D. Taylor, E. Smith, N. J. Schork, D. Levy and A. Aviv: Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Human Molecular Genetics*, 21, 5385-5394 (2012) DOI: 10.1093/hmg/dds382

118. M. Mangino, J. B. Richards, N. Soranzo, G. Zhai, A. Aviv, A. M. Valdes, N. J. Samani, P. Deloukas and T. D. Spector: A genome-wide association study identifies a novel locus on chromosome 18q12.2. influencing white cell telomere length. *Journal of Medical Genetics*, 46, 451-454 (2009) DOI: 10.1136/jmg.2008.064956

119. K. A. Pooley, S. E. Bojesen, M. Weischer, S. F. Nielsen, D. Thompson, A. Amin Al Olama, K. Michailidou, J. P. Tyrer, S. Benlloch, J. Brown, T. Audley, R. Luben, K.-T. Khaw, D. E. Neal, F. C. Hamdy, J. L. Donovan, Z. Kote-Jarai, C. Baynes, M. Shah, M. K. Bolla, Q. Wang, J. Dennis, E. Dicks, R. Yang, A. Rudolph, J. Schildkraut, J. Chang-Claude, B. Burwinkel, G. Chenevix-Trench, P. D. P. Pharoah, A. Berchuck, R. A. Eeles, D. F. Easton, A. M. Dunning and B. G. Nordestgaard: A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Human Molecular Genetics*, 22, 5056-5064 (2013) DOI: 10.1093/hmg/ddt355

120. J. Prescott, P. Kraft, D. I. Chasman, S. A. Savage, L. Mirabello, S. I. Berndt, J. L. Weissfeld, J. Han, R. B. Hayes, S. J. Chanock, D. J. Hunter and I. De Vivo: Genome-wide association study of relative telomere length. *PloS One*, 6, e19635 (2011) DOI: 10.1371/journal.pone.0019635

121. R. Saxena, A. Bjonnes, J. Prescott, P. Dib, P. Natt, J. Lane, M. Lerner, J. A. Cooper, Y. Ye, K. W. Li, C. G. Maubaret, V. Codd, D. Brackett, L. Mirabello, P. Kraft, C. P. Dinney, D. Stowell, M. Peyton, S. Ralhan, G. S. Wander, N. K. Mehra, K. D. Salpea, J. Gu, X. Wu, M. Mangino, D. J. Hunter, I. De Vivo, S. E. Humphries, N. J. Samani, T. D. Spector, S. A. Savage and D. K. Sanghera: Genome-wide association study identifies variants in casein kinase II (CSNK2A2) to be associated with leukocyte telomere length in a Punjabi Sikh diabetic cohort. *Circulation. Cardiovascular Genetics*, 7, 287-295 (2014) DOI: 10.1161/CIRCGENETICS.113.000412

122. V. Codd, M. Mangino, P. van der Harst, P. S. Braund, M. Kaiser, A. J. Beveridge, S. Rafelt, J. Moore, C. Nelson, N. Soranzo, G. Zhai, A. M. Valdes, H. Blackburn, I. M. Leach, R. A. de Boer, A. H. Goodall, W. Ouwehand, D. J. van Veldhuisen, W. H. van Gilst, G. Navis, P. R. Burton, M. D. Tobin, A. S. Hall, J. R. Thompson, T. Spector and N. J. Samani: Variants near TERC are associated with mean telomere length. *Nature genetics*, 42, 197-199 (2010) DOI: 10.1038/ng.532

271. 123. J. H. Lee, R. Cheng, L. S. Honig, M. Feitosa, C. M. Kammerer, M. S. Kang, N. Schupf, S. J. Lin, J. L. Sanders, H. Bae, T. Druley, T. Perls, K. Christensen, M. Province and R. Mayeux: Genome wide association and linkage analyses identified three loci—4q25, 17q23.2., and 10q11.2.1—associated with variation in leukocyte telomere length: the Long Life Family Study. *Frontiers in Genetics*, 4 (2014)

273. 124. S. K. Do, S. S. Yoo, Y. Y. Choi, J. E. Choi, H. S. Jeon, W. K. Lee, S. Y. Lee, J. Lee, S. I. Cha, C. H. Kim and J. Y. Park: Replication of the results of genome-wide and candidate gene association studies on telomere length in a Korean population. *Korean J Intern Med*, 30(5), 719-26 (2015) DOI: 10.3904/kjim.2015.30.5.719

125. B. D. W. Group.: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*, 69, 89-95 (2001) DOI: 10.1067/mcp.2001.113989

126. S. E. Bojesen: Telomeres and human health. *Journal of Internal Medicine*, 274,

© 1996-2018

399-413 (2013)
DOI: 10.1111/joim.12083

127. H. Tahara: Telomere G-Overhang Length Measurement Method 2: G-Tail Telomere HPA. *Methods Mol Biol*, 1587, 63-69 (2017)
DOI: 10.1007/978-1-4939-6892-3_6

128. K. A. Mather, A. F. Jorm, R. A. Parslow and H. Christensen: Is telomere length a biomarker of aging? A review. *J Gerontol A Biol Sci Med Sci*, 66(2), 202-13 (2011)
DOI: 10.1093/gerona/glq180

129. T. E. Johnson: Recent results: biomarkers of aging. *Exp Gerontol*, 41(12), 1243-6 (2006)
DOI: 10.1016/j.exger.2006.09.006

130. J. Deelen, M. Beekman, M. Capri, C. Franceschi and P. E. Slagboom: Identifying the genomic determinants of aging and longevity in human population studies: progress and challenges. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 35, 386-396 (2013)
DOI: 10.1002/bies.201200148

131. J. Deelen, E. B. van den Akker, S. Trompet, D. van Heemst, S. P. Mooijaart, P. E. Slagboom and M. Beekman: Employing biomarkers of healthy ageing for leveraging genetic studies into human longevity. *Experimental Gerontology*, 82, 166-174 (2016)
DOI: 10.1016/j.exger.2016.06.013

132. A. Russo, F. Modica, S. Guarrera, G. Fiorito, B. Pardini, C. Viberti, A. Allione, R. Critelli, A. Bosio, G. Casetta, G. Cucchiarale, P. Destefanis, P. Gontero, L. Rolle, A. Zitella, D. Fontana, B. Frea, P. Vineis, C. Sacerdote and G. Matullo: Shorter leukocyte telomere length is independently associated with poor survival in patients with bladder cancer. *Cancer Epidemiol Biomarkers Prev*, 23(11), 2439-46 (2014)
DOI: 10.1158/1055-9965.EPI-14-0228

133. C. Duggan, R. Risques, C. Alfano, D. Prunkard, I. Imayama, S. Holte, K. Baumgartner, R. Baumgartner, L. Bernstein, R. Ballard-Barbash, P. Rabinovitch and A. McTiernan: Change in peripheral blood leukocyte telomere length and mortality in breast cancer survivors. *J Natl Cancer Inst*, 106(4), dju035 (2014)
DOI: 10.1093/jnci/dju035

134. H. Jia and Z. Wang: Telomere Length as a Prognostic Factor for Overall Survival in Colorectal Cancer Patients. *Cell Physiol Biochem*, 38(1), 122-8 (2016)
DOI: 10.1159/000438614

135. J. Williams, N. H. Heppel, B. Britt-Compton, J. W. Grimstead, R. E. Jones, S. Tauro, D. T. Bowen, S. Knapper, M. Groves, R. K. Hills, C. Pepper, D. M. Baird and C. Fegan: Telomere length is an independent prognostic marker in MDS but not in de novo AML. *Br J Haematol* (2017)
DOI: 10.1111/bjh.14666

136. B. Karimi, M. Yunesian, R. Nabizadeh, P. Mehdipour and A. Aghaie: Is Leukocyte Telomere Length Related with Lung Cancer Risk?: A Meta-Analysis. *Iran Biomed J*, 21(3), 142-153 (2017)
DOI: 10.18869/acadpub.ibj.21.3.142

137. M. Yang, J. Prescott, E. M. Poole, M. S. Rice, L. D. Kubzansky, A. Idahl, E. Lundin, I. De Vivo and S. S. Tworoger: Prediagnosis Leukocyte Telomere Length and Risk of Ovarian Cancer. *Cancer Epidemiol Biomarkers Prev*, 26(3), 339-345 (2017)
DOI: 10.1158/1055-9965.EPI-16-0466

138. E. J. Duell: Telomere length and pancreatic cancer risk: breaking down the evidence. *Gut*, 66(6), 1 (2017)
DOI: 10.1136/gutjnl-2016-313156

139. S. E. Stanley and M. Armanios: The short and long telomere syndromes: paired paradigms for molecular medicine. *Curr Opin Genet Dev*, 33, 1-9 (2015)
DOI: 10.1016/j.gde.2015.06.004

140. S. Ribero, M. Mangino and V. Bataille: Skin phenotypes can offer some insight about the association between telomere length and cancer susceptibility. *Med Hypotheses*, 97, 7-10 (2016)
DOI: 10.1016/j.mehy.2016.10.010

141. M. N. Bainbridge, G. N. Armstrong, M. M. Gramatges, A. A. Bertuch, S. N. Jhangiani, H. Doddapaneni, L. Lewis, J. Tombrello, S. Tsavachidis, Y. Liu, A. Jalali, S. E. Plon, C. C. Lau, D. W. Parsons, E. B. Claus, J. Barnholtz-Sloan, D. Il'yasova, J. Schildkraut, F. Ali-Osman, S. Sadetzki, C. Johansen, R. S. Houlston, R. B. Jenkins, D. Lachance, S. H. Olson, J. L. Bernstein, R. T. Merrell, M. R. Wrensch, K. M. Walsh, F. G. Davis, R. Lai, S. Shete, K. Aldape, C. I. Amos, P. A. Thompson, D. M. Muzny, R. A. Gibbs, B. S. Melin and M. L. Bondy: Germline mutations in shelterin complex

© 1996-2018

genes are associated with familial glioma. *J Natl Cancer Inst*, 107(1), 384 (2015) DOI: 10.1093/jnci/dju384

142. Y. Sun, L. Zhang, L. Zhao, X. Wu and J. Gu: Association of leukocyte telomere length in peripheral blood leukocytes with endometrial cancer risk in Caucasian Americans. *Carcinogenesis*, 36(11), 1327-32 (2015) DOI: 10.1093/carcin/bgv133

143. U. Svenson, G. Roos and P. Wikstrom: Long leukocyte telomere length in prostate cancer patients at diagnosis is associated with poor metastasis-free and cancer-specific survival. *Tumour Biol*, 39(2), 1010428317692236 (2017) DOI: 10.1177/1010428317692236

144. B. Julin, I. Shui, C. M. Heaphy, C. E. Joshu, A. K. Meeker, E. Giovannucci, I. De Vivo and E. A. Platz: Circulating leukocyte telomere length and risk of overall and aggressive prostate cancer. *Br J Cancer*, 112(4), 769-76 (2015) DOI: 10.1038/bjc.2014.640

145. L. Hou, X. Zhang, A. J. Gawron and J. Liu: Surrogate tissue telomere length and cancer risk: shorter or longer? *Cancer Lett*, 319(2), 130-135 (2012) DOI: 10.1016/j.canlet.2012.01.028

146. S. C. Akincilar, B. Unal and V. Tergaonkar: Reactivation of telomerase in cancer. *Cell Mol Life Sci*, 73(8), 1659-70 (2016) DOI: 10.1007/s00018-016-2146-9

147. P. M. Nilsson, H. Tufvesson, M. Leosdottir and O. Melander: Telomeres and cardiovascular disease risk: an update 2013. *Transl Res*, 162(6), 371-80 (2013) DOI: 10.1016/j.trsl.2013.05.004

148. J.-K. Yeh and C.-Y. Wang: Telomeres and Telomerase in Cardiovascular Diseases. *Genes*, 7 (2016) DOI: 10.3390/genes7090058

149. K. Bennaceur, M. Atwill, N. Al Zhrany, J. Hoffmann, B. Keavney, D. Breault, G. Richardson, T. von Zglinicki, G. Saretzki and I. Spyridopoulos: Atorvastatin induces T cell proliferation by a telomerase reverse transcriptase (TERT) mediated mechanism. *Atherosclerosis*, 236(2), 312-20 (2014) DOI: 10.1016/j.atherosclerosis.2014.07.020

150. K. E. Herbert, Y. Mistry, R. Hastings, T. Poolman, L. Niklason and B. Williams: Angiotensin II-mediated oxidative DNA damage accelerates cellular senescence in cultured human vascular smooth muscle cells via telomere-dependent and independent pathways. *Circ Res*, 102(2), 201-8 (2008) DOI: 10.1161/CIRCRESAHA.107.158626

151. I. Spyridopoulos, J. Haendeler, C. Urbich, T. H. Brummendorf, H. Oh, M. D. Schneider, A. M. Zeiher and S. Dimmeler: Statins enhance migratory capacity by upregulation of the telomere repeat-binding factor TRF2 in endothelial progenitor cells. *Circulation*, 110(19), 3136-42 (2004) DOI: 10.1161/01.CIR.0000142866.50300.EB

152. C. Werner, C. Gensch, J. Poss, J. Haendeler, M. Bohm and U. Laufs: Pioglitazone activates aortic telomerase and prevents stress-induced endothelial apoptosis. *Atherosclerosis*, 216(1), 23-34 (2011) DOI: 10.1016/j.atherosclerosis.2011.02.011

153. D. H. Rehkopf, B. L. Needham, J. Lin, E. H. Blackburn, A. R. Zota, J. M. Wojcicki and E. S. Epel: Leukocyte Telomere Length in Relation to 17 Biomarkers of Cardiovascular Disease Risk: A Cross-Sectional Study of US Adults. *PLoS Med*, 13(11), e1002188 (2016) DOI: 10.1371/journal.pmed.1002188

154. M. Hammadah, I. Al Mheid, K. Wilmot, R. Ramadan, N. Abdelhadi, A. Alkhoder, M. Obideen, P. M. Pimple, O. Levantsevych, H. M. Kelli, A. Shah, Y. V. Sun, B. Pearce, M. Kutner, Q. Long, L. Ward, Y. A. Ko, K. Hosny Mohammed, J. Lin, J. Zhao, J. D. Bremner, J. Kim, E. K. Waller, P. Raggi, D. Sheps, A. A. Quyyumi and V. Vaccarino: Telomere Shortening, Regenerative Capacity, and Cardiovascular Outcomes. *Circ Res*, 120(7), 1130-1138 (2017) DOI: 10.1161/CIRCRESAHA.116.309421

155. S. C. Hunt, M. Kimura, P. N. Hopkins, J. J. Carr, G. Heiss, M. A. Province and A. Aviv: Leukocyte telomere length and coronary artery calcium. *Am J Cardiol*, 116(2), 214-8 (2015) DOI: 10.1016/j.amjcard.2015.03.060

156. Y. Y. Wang, A. F. Chen, H. Z. Wang, L. Y. Xie, K. X. Sui and Q. Y. Zhang: Association of shorter mean telomere length with large artery stiffness in patients with coronary heart disease. *Aging Male*, 14(1), 27-32 (2011) DOI: 10.3109/13685538.2010.529196

© 1996-2018

157. A. G. Mainous, 3rd, V. Codd, V. A. Diaz, U. J. Schoepf, C. J. Everett, M. S. Player and N. J. Samani: Leukocyte telomere length and coronary artery calcification. *Atherosclerosis*, 210(1), 262-7 (2010) DOI: 10.1016/j.atherosclerosis.2009.10.047

158. J. Huzen, R. A. de Boer, D. J. van Veldhuisen, W. H. van Gilst and P. van der Harst: The emerging role of telomere biology in cardiovascular disease. *Front Biosci (Landmark Ed)*, 15, 35-45 (2010) DOI: 10.2741/3604

159. S. W. Brouilette, A. Whittaker, S. E. Stevens, P. van der Harst, A. H. Goodall and N. J. Samani: Telomere length is shorter in healthy offspring of subjects with coronary artery disease: support for the telomere hypothesis. *Heart*, 94(4), 422-5 (2008) DOI: 10.1136/hrt.2007.139675

160. Y. Matsubara, M. Murata, K. Watanabe, I. Saito, K. Miyaki, K. Omae, M. Ishikawa, K. Matsushita, S. Iwanaga, S. Ogawa and Y. Ikeda: Coronary artery disease and a functional polymorphism of hTERT. *Biochem Biophys Res Commun*, 348(2), 669-72 (2006) DOI: 10.1016/j.bbrc.2006.07.103

161. M. Ogami, Y. Ikura, M. Ohsawa, T. Matsuo, S. Kayo, N. Yoshimi, E. Hai, N. Shirai, S. Ehara, R. Komatsu, T. Naruko and M. Ueda: Telomere shortening in human coronary artery diseases. *Arterioscler Thromb Vasc Biol*, 24(3), 546-50 (2004) DOI: 10.1161/01.ATV.0000117200.46938.e7

162. V. G. Haver, I. Mateo Leach, J. Kjekshus, J. C. Fox, H. Wedel, J. Wikstrand, R. A. de Boer, W. H. van Gilst, J. J. McMurray, D. J. van Veldhuisen and P. van der Harst: Telomere length and outcomes in ischaemic heart failure: data from the COntrolled ROsuvastatin multiNAtional Trial in Heart Failure (CORONA). *Eur J Heart Fail*, 17(3), 313-9 (2015) DOI: 10.1002/ejhf.237

163. J. Raschenberger, B. Kollerits, A. Hammerer-Lercher, B. Rantner, M. Stadler, M. Haun, P. Klein-Weigel, G. Fraedrich and F. Kronenberg: The association of relative telomere length with symptomatic peripheral arterial disease: results from the CAVASIC study. *Atherosclerosis*, 229(2), 469-74 (2013) DOI: 10.1016/j.atherosclerosis.2013.05.027

164. R. Hamad, S. Walter and D. H. Rehkopf: Telomere length and health outcomes: A two-sample genetic instrumental variables analysis. *Exp Gerontol*, 82, 88-94 (2016) DOI: 10.1016/j.exger.2016.06.005

165. S. Zhang, G. Ji, Y. Liang, R. Zhang, P. Shi, D. Guo, C. Li, J. Feng, F. Liu, R. Peng and M. Chen: Polymorphisms in Telomere Length Associated TERC and TERT predispose for Ischemic Stroke in a Chinese Han population. *Sci Rep*, 7, 40151 (2017) DOI: 10.1038/srep40151

166. M. P. Boks, H. C. van Mierlo, B. P. F. Rutten, T. R. D. J. Radstake, L. De Witte, E. Geuze, S. Horvath, L. C. Schalkwyk, C. H. Vinkers, J. C. A. Broen and E. Vermetten: Longitudinal changes of telomere length and epigenetic age related to traumatic stress and post-traumatic stress disorder. *Psychoneuroendocrinology*, 51, 506-512 (2015) DOI: 10.1016/j.psyneuen.2014.07.011

167. J. K. Kiecolt-Glaser and S. J. Wilson: Psychiatric Disorders, Morbidity, and Mortality: Tracing Mechanistic Pathways to Accelerated Aging. *Psychosom Med*, 78(7), 772-5 (2016) DOI: 10.1097/PSY.0000000000000373

168. S. M. Darrow, J. E. Verhoeven, D. Revesz, D. Lindqvist, B. W. Penninx, K. L. Delucchi, O. M. Wolkowitz and C. A. Mathews: The Association Between Psychiatric Disorders and Telomere Length: A Meta-Analysis Involving 14,827 Persons. *Psychosom Med*, 78(7), 776-87 (2016) DOI: 10.1097/PSY.0000000000000356

169. P. Y. Lin, Y. C. Huang and C. F. Hung: Shortened telomere length in patients with depression: A meta-analytic study. *J Psychiatr Res*, 76, 84-93 (2016) DOI: 10.1016/j.jpsychires.2016.01.015

170. J. E. Michalek, A. Kepa, J. Vincent, S. Frissa, L. Goodwin, M. Hotopf, S. L. Hatch, G. Breen and T. R. Powell: Genetic predisposition to advanced biological ageing increases risk for childhood-onset recurrent major depressive disorder in a large UK sample. *J Affect Disord*, 213, 207-213 (2017) DOI: 10.1016/j.jad.2017.01.017

171. Q. G. Zhou, Y. Hu, D. L. Wu, L. J. Zhu, C. Chen, X. Jin, C. X. Luo, H. Y. Wu, J. Zhang

© 1996-2018

and D. Y. Zhu: Hippocampal telomerase is involved in the modulation of depressive behaviors. *J Neurosci*, 31(34), 12258-69 (2011)
DOI: 10.1523/JNEUROSCI.0805-11.2011

172. Y. B. Wei, L. Backlund, G. Wegener, A. A. Mathe and C. Lavebratt: Telomerase dysregulation in the hippocampus of a rat model of depression: normalization by lithium. *Int J Neuropsychopharmacol*, 18(7), pyv002 (2015)
DOI: 10.1093/ijnp/pyv002

173. V. Boccardi, L. Pelini, S. Ercolani, C. Ruggiero and P. Mecocci: From cellular senescence to Alzheimer's disease: The role of telomere shortening. *Ageing Res Rev*, 22, 1-8 (2015)
DOI: 10.1016/j.arr.2015.04.003

174. Z. Cai, L. J. Yan and A. Ratka: Telomere shortening and Alzheimer's disease. *Neuromolecular Med*, 15(1), 25-48 (2013)
DOI: 10.1007/s12017-012-8207-9

175. M. Liu, Y. R. Huo, J. Wang, C. Wang, S. Liu, S. Liu, J. Wang and Y. Ji: Telomere Shortening in Alzheimer's Disease Patients. *Ann Clin Lab Sci*, 46(3), 260-5 (2016)

176. Y. Takata, M. Kikukawa, H. Hanyu, S. Koyama, S. Shimizu, T. Umahara, H. Sakurai, T. Iwamoto, K. Ohyashiki and J. H. Ohyashiki: Association between ApoE phenotypes and telomere erosion in Alzheimer's disease. *J Gerontol A Biol Sci Med Sci*, 67(4), 330-5 (2012)
DOI: 10.1093/gerona/glr185

177. T. Hochstrasser, J. Marksteiner and C. Humpel: Telomere length is age-dependent and reduced in monocytes of Alzheimer patients. *Exp Gerontol*, 47(2), 160-3 (2012)
DOI: 10.1016/j.exger.2011.11.012

178. B. Bernardes de Jesus and M. A. Blasco: Assessing cell and organ senescence biomarkers. *Circ Res*, 111(1), 97-109 (2012)
DOI: 10.1161/CIRCRESAHA.111.247866

179. D. Scarabino, E. Broggio, G. Gambina, F. Pelliccia and R. M. Corbo: Common variants of human TERT and TERC genes and susceptibility to sporadic Alzheimers disease. *Exp Gerontol*, 88, 19-24 (2017)
DOI: 10.1016/j.exger.2016.12.017

180. T. O. W. S. f. N. Prize: The 2009 Nobel Prize in Physiology or Medicine. https://www.nobelprize.org/nobel_prizes/medicine/laureates/2009/press.html (2009)

181. J. W. Shay and S. Bacchetti: A survey of telomerase activity in human cancer. *Eur J Cancer*, 33(5), 787-91 (1997)
DOI: 10.1016/S0959-8049(97)00062-2

182. D. L. Gomez, R. G. Armando, C. S. Cerrudo, P. D. Ghiringhelli and D. E. Gomez: Telomerase as a Cancer Target. Development of New Molecules. *Curr Top Med Chem*, 16(22), 2432-40 (2016)
DOI: 10.2174/1568026616666160212122425

447. 183. G. M. Arndt and K. L. MacKenzie: New prospects for targeting telomerase beyond the telomere. *Nat Rev Cancer*, 16(8), 508-24 (2016)
DOI: 10.1038/nrc.2016.55

184. G. Mosoyan, T. Kraus, F. Ye, K. Eng, J. D. Crispino, R. Hoffman and C. Iancu-Rubin: Imetelstat, a telomerase inhibitor, differentially affects normal and malignant megakaryopoiesis. *Leukemia* (2017)
DOI: 10.1038/leu.2017.78

185. imetelstat.info: Information about imetelstat: the first clinical stage telomerase inhibitor. https://imetelstat.eu/ (2017)

186. A. M. Zahler, J. R. Williamson, T. R. Cech and D. M. Prescott: Inhibition of telomerase by G-quartet DNA structures. *Nature*, 350(6320), 718-20 (1991)
DOI: 10.1038/350718a0

187. M. J. Lecours, A. Marchand, A. Anwar, C. Guetta, W. S. Hopkins and V. Gabelica: What stoichiometries determined by mass spectrometry reveal about the ligand binding mode to G-quadruplex nucleic acids. *Biochim Biophys Acta* (2017)
DOI: 10.1016/j.bbagen.2017.01.010

188. D. Musumeci, J. Amato, P. Zizza, C. Platella, S. Cosconati, C. Cingolani, A. Biroccio, E. Novellino, A. Randazzo, C. Giancola, B. Pagano and D. Montesarchio: Tandem application of ligand-based virtual screening and G4-OAS assay to identify novel G-quadruplex-targeting chemotypes. *Biochim Biophys Acta* (2017)
DOI: 10.1016/j.bbagen.2017.01.024

189. A. Alibakhshi, J. Ranjbari, Y. Pilehvar-Soltanahmadi, M. Nasiri, M. Mollazade and N. Zarghami: An Update on Phytochemicals in Molecular Target Therapy of Cancer:

© 1996-2018

Potential Inhibitory Effect on Telomerase Activity. *Curr Med Chem*, 23(22), 2380-93 (2016)
DOI: 10.2174/0929867323666160425113705

190. C. Y. Shen, J. G. Jiang, L. Yang, D. W. Wang and W. Zhu: Anti-ageing active ingredients from herbs and nutraceuticals used in traditional Chinese medicine: pharmacological mechanisms and implications for drug discovery. *Br J Pharmacol* (2016)

191. W. Yu, X. Qin, Y. Jin, Y. Li, C. Santiskulvong, V. Vu, G. Zeng, Z. Zhang, M. Chow and J. Rao: Tianshengyuan-1 (TSY-1) regulates cellular Telomerase activity by methylation of TERT promoter. *Oncotarget*, 8(5), 7977-7988 (2017)
DOI: 10.18632/oncotarget.13939

192. A. K. Dieckmann, V. Babin, Y. Harari, R. Eils, R. Konig, B. Luke and M. Kupiec: Role of the ESCRT Complexes in Telomere Biology. *MBio*, 7(6) (2016)
DOI: 10.1128/mBio.01793-16

193. L. Salvador, G. Singaravelu, C. B. Harley, P. Flom, A. Suram and J. M. Raffaele: A Natural Product Telomerase Activator Lengthens Telomeres in Humans: A Randomized, Double Blind, and Placebo Controlled Study. *Rejuvenation Res*, 19(6), 478-484 (2016)
DOI: 10.1089/rej.2015.1793

194. C. T. Dow and C. B. Harley: Evaluation of an oral telomerase activator for early age-related macular degeneration - a pilot study. *Clin Ophthalmol*, 10, 243-9 (2016)
DOI: 10.2147/OPTH.S100042

195. C. B. Harley, W. Liu, M. Blasco, E. Vera, W. H. Andrews, L. A. Briggs and J. M. Raffaele: A natural product telomerase activator as part of a health maintenance program. *Rejuvenation Res*, 14(1), 45-56 (2011)
DOI: 10.1089/rej.2010.1085

196. RepeatDx: Telomeres| Telomere testing| Pulmonary Fibrosys| Repeat Diagnostics. http://www.repeatdx.com/

197. V. t. g. company: Veritas| Understand your genetics to live healthier, longer. https://www.veritasgenetics.com/

198. C. HearthLab: Telomere Testing May Predict Lifespan and Health Risks. http://www.clevelandheartlab.com/blog/horizons-telomere-testing-may-predict-lifespan-and-health-risks/

199. S. Laboratories: Telomere Testing. https://www.spectracell.com/clinicians/products/telomere-testing/

200. L. Length: Life Length Telomere Testing. http://lifelength.com/

201. TeloYears: Telomere Age Test - Testing Genetic Aging. https://www.teloyears.com/home/index.html

202. Titanovo: Telomere length testing and longevity genotyping. http://telomeres.titanovo.com/

203. H. Wolinsky: Testing time for telomeres. *EMBO Reports*, 12, 897-900 (2011)
DOI: 10.1038/embor.2011.166

204. T. Diagnostics: Company - Telomere Diagnostics. http://telomeredx.com/company/

Abbreviations: LTL: Leukocyte telomere length; TNF-α: Tumor necrosis factor α; CRP: C-reactive protein; COPD: Chronic obstructive pulmonary disease; sICAM-1: Serum intercellular adhesion molecule-1; sVCAM-1: Serum vascular cell adhesion molecule-1; SAA: Serum amyloid A; PCOS: Polycystic ovary syndrome; VEGF: Vascular endothelial growth factor; HGF: Hepatocyte Growth Factor; G-CSF: Granulocyte colony-stimulating factor; IL: Interleukin; T1D: type 1 diabetes; T2D: type 2 diabetes; 8-OHdG: 8-hydroxy-desoxyguanosine

Key Words: Telomere length, Biomarker, Personalized medicine, Treatment, Chronic diseases, Aging, Review

Send correspondence to: Sophie Visvikis-Siest, INSERM UMR U1122; IGE-PCV, Faculty of Pharmacy, University of Lorraine, 22 rue Lionnois, 54000 Nancy, France, Tel: 336-076-025-69, E-mail: sophie.visvikis-siest@inserm.fr

© 1996-2018

Publication Nº 8

Telomere length determinants in childhood.

Vesna Gorenjak, Alexandros M. Petrelis, Maria G. Stathopoulou and Sophie Visvikis-Siest

CCLM, September 2019

The review article entitled '*Telomere length determinants in childhood'* was a product of the preparation for a project in collaboration with the university hospital in Nancy, aiming to perform TL research in newborns. In order to understand the state-of-the-art of the subject, a detailed analysis of the literature, investigating the telomere attrition in early childhood and the determinants of newborn TL, was conducted.

The childhood seems particularly important period for the onset of the baseline TL (95), which also impacts on adulthood TL and eventually influences the pathological processes leading to common chronic diseases. Various environmental influences on newborn TL, especially during pregnancy, showed a significant impact and demonstrated the importance of a healthy pregnancy on the overall health of an individual (310-313).

Telomere dynamics has been described in several studies and despite the substantial limitations (low sample size, absence of longitudinal studies), we were able to draw general conclusions for telomere regulation (314-316). The main knowledge gap in the field is the understanding of genetic variants associated with childhood TL, as only three medium-sized genetic studies were conducted up to date. All of them agreed that genetic variants identified in adult studies do not reflect the TL in the children population. However, the studies did not replicate variants that could be significant for childhood TL (317, 318). Our review article pointed out the importance of new GWAS, performed in newborn and childhood populations, to obtain further information on genetic regulation of TL during childhood.

Unfortunately, the design of such a study would require a long preparation to acquire the appropriate population of sufficient sample size and all necessary ethical permissions. It would thus be impossible to finalise it during the time of this thesis. However, the article '*Telomere length determinants in childhood'* made a step further in the research of TL and

offered a general insight into childhood TL, which will be useful for anyone planning to study this field further.

Finally, we concluded that TL is a prospective biomarker that could be used in the future as an indicator of general health, especially in longitudinal patient follow-up during their lifetime, which was in accordance with our previous publication. Regular measurement of TL starting in the childhood could offer an insight into the rate of individual cellular turnover and could alert the physicians about the physiological changes in case of the accelerated telomere attrition.

DE GRUYTER Clin Chem Lab Med 2019; aop

Review

Vesna Gorenjak, Alexandros M. Petrelis, Maria G. Stathopoulou and Sophie Visvikis-Siest*

Telomere length determinants in childhood

https://doi.org/10.1515/cclm-2019-0235
Received March 1, 2019; accepted July 30, 2019

Abstract: Telomere length (TL) is a dynamic marker that reflects genetic predispositions together with the environmental conditions of an individual. It is closely related to longevity and a number of pathological conditions. Even though the extent of telomere research in children is limited compared to that of adults, there have been a substantial number of studies providing first insights into child telomere biology and determinants. Recent discoveries revealed evidence that TL is, to a great extent, determined already in childhood and that environmental conditions in adulthood have less impact than first believed. Studies have demonstrated that large inter-individual differences in TL are present among newborns and are determined by diverse factors that influence intrauterine development. The first years of child growth are associated with high cellular turnover, which results in fast shortening of telomeres. The rate of telomere loss becomes stable in early adulthood. In this review article we summarise the existing knowledge on telomere dynamics during the first years of childhood, highlighting the conditions that affect newborn TL. We also warn about the knowledge gaps that should be filled to fully understand the regulation of telomeres, in order to implement them as biomarkers for use in diagnostics or treatment.

Keywords: attrition; children; determinants; dynamics; genetics; telomere length.

***Corresponding author: Dr. Sophie Visvikis-Siest,** University of Lorraine, Inserm, IGE-PCV, 30 rue Lionnois, 54000 Nancy, France, Phone: +33 6 07 60 25 69, E-mail: sophie.visvikis-siest@inserm.fr; and Department of Internal Medicine and Geriatrics, CHU Technopôle Nancy-Brabois, Vandoeuvre-lès-Nancy, France
Vesna Gorenjak, Alexandros M. Petrelis and Maria G. Stathopoulou: University of Lorraine, Inserm, IGE-PCV, Nancy, France

Introduction to telomeres

Telomeres are non-coding repetitive sequences TTAGGG on the ends of the eukaryotic chromosomes [1]. They form a protective cap that conserves the genetic material during cell division and protect it from constitutive exposure to the DNA damage response [2]. In humans, telomeres consist of 4–15 kilobase pairs (kbp) [1]. From the early research it has become clear that they shorten progressively with age [3] as a result of the end replication problem during cell division and/or oxidative stress. Chromosomes with critically short telomeres are recognised by DNA damage response proteins as damaged DNA, which leads the cell to a controlled telomere-initiated senescence [4]. When cells become senescent, they undergo morphological and genetic changes that result in the loss of tissue function.

Cells are capable of maintaining the length of telomeres with the telomerase enzyme, which is extinguished during embryonic differentiation in most somatic cells, but remains active in germline cells, activated lymphocytes and certain types of stem cell populations [5]. During normal human growth and development, telomerase activity is precisely regulated by a number of genes in order to meet the proliferative demand of the specific cellular function [6]. Cancer cells, on the other hand, can acquire the possibility of an infinite number of divisions by enhancing the telomerase activity, thus restraining the shortening of telomeres [7], or with the mechanism of the alternative lengthening of telomeres (ALT) [8].

Telomere length (TL) has been an important topic of research because of its association with longevity, as well as with the occurrence and progression of common chronic diseases. An active role of TL in age-related human diseases arises because short telomeres increase the risk of diseases related to restricted cell proliferation and tissue degradation, such as cardio-vascular diseases (CVD); long telomeres increase the risk of diseases related to increased proliferative growth, such as major cancers [9]. Therefore, TL is considered as a potential biomarker for disease susceptibility or a possible target for a particular treatment. However, its complex biology, mixed with the substantial impact of genetic and environmental factors are making this task difficult [10].

Unauthenticated
Download Date | 8/30/19 1:36 PM

In the early 1990s, the interest in telomeres started to spread, and samples of individuals of different age ranges were used for the studies; among these were also the first samples of children. Knowing the biology behind telomeres in newborns and young children is crucial, as it provides the information about baseline TL [11]. The baseline TL represents a critically important characteristic of an individual's telomere biology, together with a second major determinant – TL attrition over time [9]. A reduction in newborn TL can result in a greater susceptibility for pathophysiological conditions in adulthood [11].

In the following article we summarise the current knowledge of telomeres in childhood. This important period of life is characterised by dynamic biology and strong dependence on genetic and environmental factors, and it plays a crucial role in telomere attrition [12]. We demonstrate why the precise understanding of telomere dynamics in childhood is indispensable for further consideration of telomeres as a biomarker or a treatment target.

Telomeres and diseases

Telomere syndromes

TL and telomerase activity are involved in the pathology of numerous diseases [13], and preterm telomere attrition is associated with reduced longevity [14]. The initial discoveries of the role of telomeres in human diseases primarily came from the studies of rare monogenic disorders of childhood. Monogenic, age-related dysfunctions (i.e. telomere syndromes) are associated with short telomeres, normally caused by genetic mutation in a gene associated with telomere maintenance [15]. Dyskeratosis congenita (DC) is a multisystem inherited syndrome which was first related to defective telomere maintenance [16]. Children inherit short TL from affected parents, together with a mutated copy of a gene encoding products involved in telomere maintenance, such as telomerase core components (*TERT* and *TERC*), factors required for telomerase biogenesis (*DKC1*, *NHP2*, *NOP2*, *NAF1* and *PARN*), telomerase trafficking (*WRAP53*), telomerase recruitment (*ACD*), telomere replication and end structure (*RTEL1*, *CTC1*, *STN1* and *POT1*), and other aspects of telomere biology (*TINF2*) [17]. The three most distinguished subtypes of DC are related to the mentioned genes: X-linked recessive DC with mutations in the *DKC1* gene, encoding a component of H/ACA small nucleolar ribonucleoprotein; autosomal dominant DC with heterozygous mutations in either *TERC* (RNA component of telomerase) or *TERT* (enzymatic component of telomerase); and autosomal recessive DC, for which the genes involved remain largely elusive [18]. The disease affects highly proliferative tissues, which require constant renewal and are normally highly regenerative, such as skin, gut and bone marrow [19, 20]. This leads to mucocutaneous abnormalities, abnormal nails, reticular skin pigmentation, oral leukoplakia and an increased predisposition to cancer. The principle cause of mortality of DC is bone marrow failure or aplastic anaemia; 40% of patients are affected by one of these conditions by the age of 40 [21].

Hoyeraal-Hreidarsson and Revesz syndromes are two very rare disorders with onset in infancy, are linked to very short telomeres, and represent a spectrum of the same disease. They are more severe than DC and are characterised by intrauterine growth retardation, microcephaly, cerebellar hypoplasia, progressive combined immune deficiency and aplastic anaemia. Revesz syndrome can be distinguished from Hoyeraal-Hreidarsson syndrome by the presence of bilateral exudative retinopathy [22].

Telomere disorders most commonly manifest as adult-onset diseases. Mutations in *TERT* and *TERC* can lead to diverse phenotypes in which severity depends on which family generation is affected. Idiopathic pulmonary fibrosis, for example, is normally diagnosed at the age of 50. However, later generations of a susceptible family will more often have children affected with aplastic anaemia along with classic features of DC [23].

Cancer

Telomere biology is very important in paediatric cancers, as most paediatric leukaemias and embryonal solid tumours activate the telomerase enzyme. Therefore, telomeres in paediatric cancer patients are often longer than in healthy subjects, but become shorter after chemotherapeutic treatment, which causes an excessive oxidative stress that harms telomeres. Telomerase was considered as a therapeutic target for paediatric cancers, as the enzyme plays a key role in conferring cellular immortality, is present in most tumours, and is relatively specific for cancer cells [24–26]. Despite all the efforts to develop a telomerase-targeting therapy, there is currently no approved treatment available on the market [27]. This is mainly due to the high number of cell doublings required to induce a tumour suppressive senescence after telomerase inhibition – a long period before a telomerase inhibitor becomes effective against a tumour [28]. In addition, such treatment could prompt a toxic effect on highly proliferative normal stem cells that can express a regulated telomerase activity [29]. A Mendelian randomisation study

Unauthenticated
Download Date | 8/30/19 1:36 PM

on the association between TL and risk of cancer and non-neoplastic diseases showed that increased TL due to germline genetic variation was generally associated with increased risk for site-specific cancers, but reduces the risk for some non-neoplastic diseases, including CVD [30]. Therefore, potential therapeutic applications based on TL should be carefully considered also for the trade-off of the risk for going from one disease to another.

Stress and other diseases

Many childhood clinical disorders can cause increased oxidative stress and chronic inflammation, including perinatal brain damage, asthma, cystic fibrosis, juvenile rheumatoid arthritis, cholestatic liver diseases and diarrhoeal diseases [31]. Chronic inflammation can increase the rate of cell proliferation and increase cellular turnover, therefore facilitating telomere erosion. Also, oxidative stress can increase the size of telomere repeats clipped from the ends of chromosomes, thus promoting telomere DNA double-strand breakage and telomere shortening [32]. Long-term childhood diseases that provoke an increased oxidative stress can therefore cause faster telomere shortening and result in accelerated cellular ageing that may manifest in adulthood as a common age-related pathology [33–36].

It was demonstrated that in childhood, stress originating from childhood maltreatment [37] or violence [38] can cause a premature shortening of telomeres [39]. Similarly, maternal stress during pregnancy can have an impact on TL in children. This subject will be further discussed in the section "Determinants of newborn TL".

Genetic determinants of TL

TL is a highly heritable trait (36%–86% in different family and twin studies) [40–45]. Several large genome-wide association studies (GWAS) were performed to investigate associations between genetic variants and relative TL in an adult population [46–56]. Most of the identified SNPs harbor genes that encode proteins with known functions in telomere biology (i.e. genes directly involved in telomere maintenance). Two GWAS studies were performed in paediatric populations to look for common SNPs related to relative TL (Table 1), and none replicated the results identified in adult cohorts. This suggests that possibly other genetic variants are involved in the regulation of TL in childhood and adolescence than in adulthood [57, 58]. Moreover, none of the identified SNPs was replicated among paediatric populations. One of the possible reasons for this might be the heterogeneity between populations, as studies were performed in healthy African-American children and adolescents [57] and in a healthy European population [58]. Genotyping of the SNPs identified in the European cohort was also performed in newborns in an Asian population, but no significant association for any individual variant was detected [59].

There is no doubt that genes related to telomere biology have a big impact on TL starting at birth; a pure example is the heritable childhood disease DC as discussed already [20]. However, individuals with such diseases are born with the very short telomeres, inherited by their parents, whereas the impact of affected genes can show only later in life. Accelerated loss of telomere base pairs after birth, caused by affected genes, might not be evident in the first decades of life, but would only become significant in adulthood.

The GWAS on paediatric populations are interesting for the identification of genes that reflect the baseline TL before the occurrence of the major attrition. However, paediatric studies to date have used small sample sizes compared to studies in adult cohorts. In order to confirm whether there is an age-specific variation of genetic factors contributing to TL regulation, and to identify novel variants that regulate TL in childhood, larger

Table 1: Genetic variants associated with relative TL in children population.

First author, year, reference	Sample size	SNP	p-Value	Chromosome	Reported gene(s)
Zeiger, 2018 [57]	492 Children (8–20 years)	rs1483898	7.86×10^{-8}	14	*LRFN5*
Stathopoulou, 2015 [58]	322 Children (6–17 years)	rs10496920[a]	5.00×10^{-5}	2	*LRP1B*
		rs528983[a]	2.93×10^{-5}	4	*NDST4*
		rs594119[a]	3.11×10^{-5}	6	*NKAIN2*
		rs12678295[a]	3.53×10^{-5}	8	*MYOM2*
		rs2300383[a]	1.88×10^{-5}	21	*ITSN1*
		rs11703393[a]	5.00×10^{-5}	22	*PARVB*

[a]SNPs did not reach genome-wide significance level, but reached the levels of suggestive association ($p \leq 5\times10^{-5}$).

Unauthenticated
Download Date | 8/30/19 1:36 PM

GWAS studies should be performed on the paediatric population.

Assessment of TL in healthy and diseased neonates and children

In the following section we will present the studies that focussed on determination of newborn TL and rate of TL attrition throughout childhood (Table 2) and summarise the important conclusions acquired from the researchers.

One of the first investigations that used a paediatric sample to study TL dates from 1993. The study compared lymphocyte TL of 21 individuals with Down syndrome (DS), aged 0–45 years, to 119 healthy controls aged 0–107 years. The rate of telomere loss was calculated as the decrease in mean telomere restriction fragment (TRF) length as a function of donor age. The study demonstrated that DS patients experience significantly higher rates of telomere loss (133 ± 15 bp/year) compared to healthy age-matched controls (41 ± 7 bp/year). Moreover, separation of samples according to gender showed faster rates of male telomere loss compared to female, but the difference was not statistically significant. The difference in the rate of TL shortening from birth until adulthood was not discussed [60].

The first study to focus on the rate of telomere sequence loss in children's leucocytes was published by Frenck et al. in 1998. The research was conducted on 75 individuals that comprised 12 unrelated healthy newborns and their relatives (parents, grandparents). Comparison among three generations showed a larger difference in TL between the newborns and their parents (4.8 kbp) than between parents and grandparents (2.0 kbp), even though the age interval was similar for both generations (25 years). Further research was conducted in order to observe the rate of telomere shortening during the first few years of life. Because of a disagreement of the Institutional Review board, the study could not provide the follow-up of the same children, but it compared 10 unrelated healthy children of age 5–48 months. Researchers discovered that the rate of telomere shortening is not stable, but appears in three characteristic phases: (1) from birth to the age of 4, characterised by a rapid decline in the average TL, (2) stable TL from the age of 4 until early adulthood, (3) gradual decline in mean TL associated with advancing age. The authors assumed that a high rate of proliferation appears in the most immature subset of haematopoietic progenitors in the first months of life. Haematopoiesis throughout life is later initiated from progenitors that have already undergone a substantial amount of telomeric loss [61]. The weak point of the study was a low number of the individuals (10 children) by which the whole theory was estimated. It is important to mention that variation in TL among newborns is as wide as the variation among adults [12]. Therefore, comparisons between non-related children of different ages might not perfectly describe the process of telomere shortening that is happening in the same individual, as inter-individual differences between the subjects of the same age can be significant and could cause biased results when using a small sample size.

This problem was overcome in the study published by Zeichner et al. in 1999, in which nine newborns of HIV-infected mothers were followed up for 3 years. Two adults with the risk of HIV infection with 8- and 10-year follow-up were used for comparison. The study detected a significant difference between TL measured in peripheral blood mononuclear cells (PBMCs) at 1 month and 36 months of age. The average rate for the TL shortening was 270 bp/year from 0 to 36 months, which was significantly higher than telomere loss in adult PBMCs (50 bp/year). Compared to Frenck et al., Zeichner et al. reported only two characteristic phases of telomere loss: (1) a phase of fast telomere sequence loss in the first 3 years, followed by (2) a relatively constant gradual loss of telomere base pairs [62]. The strong point of the study was the longitudinal follow-up of individuals, which enabled a much more refined view of telomere biology than cross-sectional studies. However, the follow-up of the children was conducted only until the age of 3, and the researchers estimated the telomere loss after the age of 3 by calculating the average age loss from the last two samples obtained from each child. The calculated value for the telomere loss was 50 bp/year, consistent with the results in adults; therefore, researchers assumed that from the age of 3 onward the telomere loss is constant [62].

The plateau of TL at the age of 4, first mentioned by Frenck et al., was further investigated in 70 children of Latino origin. Leucocyte telomere length (LTL) was assessed twice in a year, at the age of 4 and 5. The study reported LTL maintenance in most of the children during the period of 1 year, which is consistent with the results obtained by Frenck et al. [61].

In order to have an insight into foetal haematopoiesis and telomere attrition, changes in LTL were also investigated in preterm and full-term newborns. The findings of the cord blood TRF analysis of 15 preterm and 11 full-term newborns showed no significant differences between the groups, though a trend of shorter telomeres with increasing gestational age was observed. A rapid and significant decline between 27 and 32 weeks of gestation was noticed,

Unauthenticated
Download Date | 8/30/19 1:36 PM

Table 2: Studies assessing the TL dynamics in newborns and children.

First author, year, reference	Sample size	Age	TL measurement	Findings	Comments	Cells
Vaziri, 1993 [60]	140	0–107 years	Mean TRF length	DS patients shorter TL than controls; shortening of TL with age; men shorter TL than women	No detailed report in children	Lymphocytes
Frenck, 1998 [61]	75 Family members, 10 unrelated children	Three generations 5–48 months	Mean TRF length	Fast decrease in first years of childhood, plateau at age of 4, graduated attrition in the adulthood	Small sample size	Leukocytes
Zeichner, 1999 [62]	Nine newborns, two adults	Longitudinal study: 0 m (for 3 years) 28 and 30 years (for 8 and 10 years)	Mean TRF length	Faster shortening during first 3 years of life, then slower rate of shortening	Small sample size	PBMCs
Friedrich, 2001 [63]	15 Preterm, 11 full-term newborns	<37 weeks of gestation, >37 weeks of gestation	Mean TRF length	No big difference in groups of newborns, significant decline of TL with increasing gestational age between 27 and 32 weeks	Comparison of different children	Leukocytes
Okuda, 2002 [12]	168 Newborns	Newborns	Mean TRF length	No significant difference in male and female newborns in TL, synchrony among tissue	TL variation among newborns is as wide as among adults	Leukocytes, umbilical artery cells, foreskin cells
Akkad, 2006 [64]	34 SGA babies, 38 normal-grown newborns	35–42 weeks of gestation	Mean TRF length	Impaired fetal growth is not associated with TL	Small sample size	Leukocytes
Wojcicki, 2016 [65]	77 Children, 70 mothers	4 years old	qPCR	TL at the age of 4–5 maintains unchanged	Technique less precise than TRF	Leukocytes
Vasu, 2017 [66]	47 Preterm, 31 full-term newborns	<32 weeks of gestation, >37 weeks of gestation	qPCR	Preterm infants at term equivalent age have significantly *longer* telomere lengths than term born infants	Small sample size, less precise than TRF	Leukocytes

DS, Down syndrome; PBMCs, peripheral blood mononuclear cells; SGA, small-for gestational-age; TL, telomere length; TRF, terminal restriction fragment.

Unauthenticated
Download Date | 8/30/19 1:36 PM

pointing to an increased proliferation rate during this period of gestation. Moreover, after 32 weeks of gestation a high inter-individual variability of LTL was observed, possibly due to an important genetic influence on TL, as estimated by the researchers [63].

On the contrary, most other studies observed significant differences in preterm and full-term newborn TL. Vasu et al. confirmed that preterm infants had longer LTL than full-term newborns and that relative LTL is highly variable among newborn infants. In addition, they discovered that at birth, preterm infants had significantly longer LTL than full-term babies [66]. Holmes explained that the difference occurs because of the greater proliferative capacity of foetal haematopoietic stem cells in comparison to post-natal stem cells [67].

Finally, telomeres were also assessed in small-for-gestational-age (SGA) babies; however, there was no difference in comparison to normally grown controls [64].

All but two of the mentioned studies used TRF technique for measurement of TL (Table 2). A real-time PCR (qPCR) technique for TL measurement is a less precise method than TRF, and it might impact the obtained results [68]. The newborn's blood was in all cases collected from the umbilical cord, and TL was measured from blood leukocytes or PBMCs. All studies have substantial limitations due to the small sample size and/or lack of longitudinal observations. However, ethical considerations of such studies limit the research design, as it is unethical to follow-up healthy children for research that will not contribute to health improvement [61]. Hence, the follow-up was possible only in the cases of children with significant health risk (such as preterm babies and children of HIV-infected mothers), where closer medical follow-up was indispensable.

Nevertheless, some final conclusions can be drawn for TL dynamics in children:

- TL is highly variable among children of the same age.
- Preterm babies have longer telomeres than full-term babies.
- The rate of telomere loss is higher in the first 3 years of life (approximately 250 bp/year).
- The TL at the age of 3 until young adulthood might be stable, but it is more likely gradually decreasing (up to 50 bp/year).
- The rate of telomere loss becomes constant in young adulthood (approximately 50 bp/year).
- The differences in shortening rates at different periods of childhood are due to an increased cell replication rate in infants, especially during accelerated development of the immune system (expanding of haematopoietic stem cell [HSC] pool).

TL shortening in diverse tissues

In the following section we will discuss the difference in TL among various human tissues. Different views on the topic will be presented and the most pertinent conclusions will be pointed out.

Several studies have agreed that synchrony in TL exists among tissues of the human foetus, but this is lost during extrauterine life [69, 70]. Corresponding to this idea, TL in samples of three tested tissues in newborns (leucocytes, umbilical artery cells and skin cells) were found to be highly synchronised [12]. Indeed, telomerase is active during every stage of development of the embryo, and this enables the maintenance of equally long telomeres [71].

With ageing, cells in different tissues lose the synchrony in TL due to different proliferative activity, which mainly arises during the expanded growth and development in the first two decades of life. For a long time, there was a common opinion that TL in skeletal cells, which form a minimally proliferative tissue, represents the approximate size of TL at birth [72, 73]. Decary et al. demonstrated that skeletal satellite cells lose both proliferative capacity and TL during the first two decades of life. In adulthood, muscle development is complete and TL becomes stable. On the other hand, they realised that skeletal muscle cell myonuclei, generally presented as post-mitotic tissue, do not undergo any shortening of telomeres and stay consistent from birth to late age [73].

Scientists thus assumed that TL in skeletal muscle and the difference between LTL and skeletal muscle telomere length (MTL) could provide a broad account of LTL dynamics over the life course of the individual, as these two variables would represent birth LTL and its age-dependent attrition, respectively [74]. They concluded that the difference between TL in muscle and leucocytes provides a much better account for LTL dynamics than LTL alone (research was performed in dogs) [74].

The research in adult subjects, however, gave unexpected results. TL shortening in four types of somatic tissue cells – leucocytes, muscle, skin and fat – was examined in 87 individuals with an age range of 19–77 years. Despite different replicative status, all tissues displayed similar rates of age-dependent attrition. The difference in TL across the tissues is thus believed to be established during the first two decades of life. Later, stem cell division rates, including those of muscle, are synchronised across the examined somatic tissues. Muscle cells are therefore not an exact indicator of birth TL, which continuously shortens even in adulthood [75]. The authors explained this discovery, which is not consistent with the previous studies of TL in skeletal muscle, as the result of

Unauthenticated
Download Date | 8/30/19 1:36 PM

Table 3: Determinants of newborn TL.

First author, year, reference	Sample size	Age	TL measurement	Observed variable	Comments	Cells
Gender						
Okuda, 2002 [12]	168 Newborns	Newborns	Mean TRF length	Gender	No significant difference in male and female newborns in TL, synchrony among tissue	Leukocytes, umbilical artery cells, foreskin cells
Wojcicki, 2016 [102]	54 Newborns	Newborns	qPCR	Gender	Female gender associated with longer TL by ~350 base pairs	Cord blood cells
Ethnicity						
Drury, 2015 [103]	71 Newborns	Newborns	qPCR	Ethnicity	Black infants had significantly longer TL than white ($p = 0.0134$), strongest effect observed in black female	Blood spot cells
Vitamins						
Entringer, 2015 [100]	119 Mother-newborn dyads	Newborns	qPCR	Maternal folate concentration in early pregnancy	10 ng/mL increase in folate increased 5.8% in median TL ($p = 0.03$)	CBMCs
Kim, 2017 [99]	106 Mother-newborn dyads	Newborns	qPCR	Maternal vitamin D concentrations	Newborn TL were associated with maternal vitamin D concentrations ($\beta = 0.33$, $p < 0.01$)	Leukocytes
Maternal pre-pregnancy BMI						
Martens, 2016 [101]	$N_{cord} = 743$ $N_{plac} = 702$	Newborns	qPCR	Maternal pre-pregnancy BMI	Each kg/m^2 increase in pre-pregnancy BMI was associated with a −0.50% shorter cord blood TL and a −0.66% shorter placental TL	Cord blood cells, placental cells
Maternal stress						
Entringerer et al., 2013 [104]	27 Mother-newborn dyads	Newborns	Mean TRF length	Maternal psychosocial stress	Independent, linear effect of pregnancy-specific stress on newborn LTL accounted for 25% of the variance of LTL	Leukocytes
Marchetto et al., 2016 [94]	24 Mother-newborn dyads	Newborns	Mean TRF length	Maternal psychosocial stress	Significantly shorter TL in newborns whose mothers experienced a high level of stress during pregnancy	Cord blood cells
Send et al., 2017 [95]	319 Newborns and 318 mothers	Newborns	qPCR	Maternal psychosocial stress	Stress during pregnancy was associated with shorter telomeres in newborns but not with maternal TL	Cord blood cells
Maternal smoking						
Almanzar, 2013 [97]	$N_{Non\text{-}smoking\ mothers} = 111$ $N_{smoking\ mothers} = 58$	Newborns	qPCR	Tobacco exposure	Newborns exposed to tobacco had significantly longer telomeres than non-exposed newborns	Lymphocytes
Salihu et al., 2015 [105]	n = 86 Mother-newborn dyads	Newborns	qPCR	Tobacco exposure	Newborns exposed to tobacco had significantly shorter telomeres than non-exposed newborns	Leukocytes

Unauthenticated
Download Date | 8/30/19 1:36 PM

Table 3 (continued)

First author, year, reference	Sample size	Age	TL measurement	Observed variable	Comments	Cells
Hormones						
Entringer, 2015 [96]	n = 100 Infants	15 months old	qPCR	Maternal E_3 concentrations in early gestation	One-multiple-of-the-median increase in maternal E3 concentration during early pregnancy was associated with a 14.42% increase in infant TL	Buccal cells
Gestational and pre-gestational diabetes						
Cross et al., 2010 [106]	$N_{type1} = 26$ $N_{type2} = 20$ $N_{gestional} = 71$ $N_{control} = 202$	Newborns	Flow cytometry	GD and PGD	No difference in cord blood TL in pregnancies of women with diabetes compared with controls	CBMCs
Xu et al., 2014 [107]	$N_{gestional} = 82$ $N_{control} = 65$	Newborns	qPCR	GD and PGD	In the GD group, TL was significantly shorter than in non-GD pregnancy ($p = 0.028$)	Leukocytes
Gilfillan et al., 2016 [108]	Diabetes: $N_{gestational} = 20$ $N_{pre\text{-}gestational} = 14$ $N_{control} = 18$	Newborns	qPCR	GD and PGD	No significant telomere shortening in the offspring of mothers with PGD or GD	Cord blood cells
Parental age at conception						
Factor-Litvak, 2016 [109]	n = 490 Father-mother-newborn trios	Newborns	TRF	Parental age	1-year increase in father's age results in 0.016 kb increase in newborn LTL	Leukocytes

BMI, body mass index; CBMC, cord blood mononuclear cells; E_3, estriol; GD, gestational diabetes; PGD, pre-gestational diabetes, TL, telomere length; TRF, terminal restriction fragment.

Unauthenticated
Download Date | 8/30/19 1:36 PM

a more accurate method of TL measurement and a bigger sample size used for the investigation.

The difference between skeletal MTL and LTL is primarily an index of HSC telomere shortening due to the expansion of the HSC pool, which becomes stable in adulthood. Some differences in proliferative activities among two tissues can appear but are likely to be modest [76].

Finally, the difference between skeletal MTL and LTL was examined in the foetal and child samples. Surprisingly, the study demonstrated that MTL is longer than LTL already in foetal samples, assuming that TL is similar across tissues only until early embryonic development. Scientists concluded that variation in TL among individuals is wider than intra-individual variation of LTL and MTL, and that individuals with short/long MTL display short/long LTL, as was previously shown [77].

Different tissue cells display different proliferative activities. However, this difference is mainly established until adulthood, when cells require only a small number of divisions to maintain tissue integrity [75].

The main conclusions of the above discussed studies are the following:

- During embryonic development, cells of different tissues have similar TL due to the activity of telomerase in early intrauterine growth.
- The difference in TL among tissues (e.g. skeletal muscle and leucocytes) appears during late intrauterine development and is mainly established in the first two decades of life.
- TL of minimally proliferative muscle cells shortens much more slowly than TL of highly proliferative leucocyte cells during the first two decades of life.
- In adulthood, skin, fat, muscle and leucocyte cell TL shortening is synchronized.
- The difference in MTL and LTL represents the leucocyte telomere dynamics during early life and corresponds well to telomere dynamics of an individual.
- High LTL variability among individuals is expressed already *in utero*.
- Variation in TL between foetuses and children is as wide as variation among adults of the same age.
- In the same child or adult, a synchrony exists across somatic tissues (an individual with long LTL will have long MTL).

Determinants of newborn TL

As previously mentioned, since birth, a high variability exists in TL among children. One of the most important factors that determine this inter-individual difference is genetics, which was shown in several heritability studies [40–45]. Nevertheless, a substantial number of studies in adult cohorts showed that besides genetics, TL can be influenced by various other factors. Most commonly reported among those are:

- Gender (women have longer TL than men) [78]
- Race (African Americans have longer TL than Europeans) [79, 80]
- Paternal age (older age of the father is associated with longer TL) [81, 82]
- Smoking (smokers have shorter TL than non-smokers) [83, 84]
- Physical activity (non-active have shorter TL than active) [85]
- Traumatic events (cause shortening of TL) [39, 86–88]
- Obesity (causes shortening of TL) [84]
- Oxidative stress (causes shortening of TL) [89, 90].

Antenatal determinants were shown to be a particularly important factor of TL, as early developmental environment conditions the developmental process, which in turn influences the individual's predisposition to developing a complex common disorder [91]. TL in children can be affected by oxidative, immune, endocrine and metabolic pathways in a way that accelerates cellular dysfunction over the lifespan [11, 92], thus reflecting the effect of specific conditions on the later susceptibility to a disease. Maternal stress [93–95], hormones [96], smoking [97], race [98], vitamin concentrations [99, 100], pre-pregnancy body mass index (BMI) [101] and other factors showed correlation with TL and pointed out the importance of *in utero* exposures in the regulation of newborn TL. In this section we will focus on some of the most pertinent antenatal determinants. The summary of the newborn TL determinants is presented in Table 3.

Gender

The fact that women have a predisposition for longer telomeres than men have is well known and has been systematically confirmed in adult telomere studies [78, 110, 111]. At the beginning of the exploration, studies of telomeres in children did not take into account the gender of the individuals or any other covariate that could impact TL. The first study to focus on newborn gender in association with TL showed that at birth, male (10.95 ± 0.088 kb) and female (11.07 ± 0.077 kb, $p = 0.3$) newborns do not statistically differ in the length of their leucocyte telomeres, suggesting that sex difference in TL arises from different rates

Unauthenticated
Download Date | 8/30/19 1:36 PM

of attrition in extrauterine life [12]. This discovery was not in accordance with later studies [103, 112], and the reason for this discrepancy might be that no other potential TL determinant (race, socioeconomic status, etc.) was taken into account.

On the contrary, longer cord blood and placental telomeres were found in female newborns compared to male newborns in studies of Latino and mixed-population infants [102, 109], supposing that hormonal differences between male and female genders could play a role in telomere dynamics *in utero* [113].

It is possible that the impact of gender on TL is smaller in newborns than in adults of the same age. However, it remains an important determinant that should be systematically considered in all telomere studies. The degree of the gender impact on infant TL remains to be fully investigated.

Race

The explanation for differences in TL among different races lies in natural protection developed during the evolution process. The hypothesis is that a cell division limit was developed as a mechanism for tumour suppression [28]. Therefore, the presence of relatively shorter TL in Europeans is a result of polygenic adaptation of the northbound migration out of equatorial Africa that attenuated the risk of melanoma in individuals with light skin pigmentation [79]. TL was compared among 71 Black and White babies, measured in blood spots of newborns. The results showed that infant TL was significantly longer in Black compared to White newborns ($p = 0.0134$), in a model accounting for sex, birth date, birth weight, gestational age, parental age, maternal race and maternal highest level of education as a proxy for socioeconomic status [103]. This study demonstrated that racial differences in TL are significant from birth and highlighted origin as an important TL determinant.

Vitamins

Two studies give an example of the importance of appropriate maternal vitamin intake during pregnancy. First, maternal folate concentration was measured in the first trimester of pregnancy in serum samples of 119 mothers and was analysed in relation to TL measured from the cord blood mononuclear cells [100]. Adjusted covariates were specified *a priori*, including maternal socioeconomic status (annual family income), race/ethnicity, maternal pre-pregnancy BMI, the presence of obstetric complications, maternal age, infant sex, gestational age at birth and birth weight. The results demonstrated that each 10 ng/mL increase in maternal total folate concentration was associated with a 5.8% increase in median TL ($p = 0.03$).

Second, maternal vitamin D concentration was examined for associations with LTL in newborn cord blood samples in a model adjusted for the maternal age, BMI, LTL, white blood cell count, glycosylated haemoglobin level, health behaviours (smoking, exercise, body weight before pregnancy, medical history) and nutritional intake and for the newborn's sex and birthweight. The study concluded that maternal vitamin D levels were also positively associated with newborn LTL [99].

Maternal pre-pregnancy BMI

Maternal obesity during pregnancy can reflect an adverse nutritional status that affects an offspring and has a significant deleterious effect on the outcome of pregnancy [114]. This phenomenon was also observed in telomeres. Higher maternal pre-pregnancy BMI was associated with a decline in newborn TL in a study examining 743 samples from cord blood and 702 samples from placental tissue, independent of maternal and paternal age at birth, maternal education, gestational age, newborn's gender, ethnicity, birthweight, maternal smoking status, parity, caesarean section and pregnancy complications. Each kg/m^2 increase in pre-pregnancy BMI was associated with a −0.50% (95% CI, −0.83 to −0.17; $p = 0.003$) shorter cord blood TL and a −0.66% (95% CI, −1.06 to −0.25; $p = 0.002$) shorter placental TL [101]. Higher maternal pre-pregnancy BMI was also associated with increased newborn adiposity and inflammation [115]. Nonetheless, with adequate preventive measures maternal obesity could be reduced. This would result in decreased pregnancy complications and might also impact the overall quality of the child's life.

Maternal stress

Stress plays an important role in foetal development and can result in an altered endocrine and immune response, which serve in the progression of normal gestation. This results in increased oxidative stress that is particularly harmful for telomeres [91]. Psychological stress and psychiatric disorders were previously linked to telomere biology in adults [37, 39, 116, 117]. The significant impact of intrauterine stress exposure on LTL was first

Unauthenticated
Download Date | 8/30/19 1:36 PM

demonstrated in young adults [118]. The fact that exposure to maternal psychosocial stress may exert a "programming" effect on the newborn telomere biology was confirmed in 27 mother-newborn dyads, after accounting for the mother's age at birth, weight, sex and exposure to antepartum obstetric complications. The effect of the pregnancy-specific stress on newborn LTL accounted for 25% of the variance in LTL [104]. Likewise, a significant negative association between maternal stress and newborn TL was observed in a group of 24 mother-newborn dyads. Affected newborns had significantly shorter TL (6.98 ± 0.41 kb) measured in cord blood cells, compared to the newborns of mothers with low stress (8.74 ± 0.24 kb; $t = -3.99$, $p = 0.003$) [94]. Finally, it was demonstrated that maternal lifetime history of a psychiatric disorder causes shortening of maternal TL but does not affect newborn TL. Acute stress during pregnancy, on the other hand, is related to shorter TL of a newborn but does not affect the mother's TL [95].

Smoking

A recent systematic review and meta-analysis demonstrated that TL of smokers decreases significantly over time in contrast to non-smokers [119], which implies mechanisms linking tobacco smoke exposure to ageing-related diseases. The impact of intrauterine tobacco exposure on foetal TL was also investigated. First, a study of the effect of cigarette smoking during pregnancy on the lymphocyte subpopulations in newborns showed that newborns of smoking mothers had significantly *longer* telomeres compared to newborns of non-smoking mothers [97]. This unexpected discovery was disproved in a study by Salihu et al. in which evidence of a positive association between shortened LTL and smoking during pregnancy was demonstrated [105], suggesting the possibility of early intrauterine programming for accelerated ageing as a result of tobacco exposure. The reason for the discrepancy in the results of these two studies is not clear; however, there is undeniable proof that link cigarette smoke to telomere shortening caused by increased oxidation stress in cells [120]. We therefore postulate that smoking during pregnancy should be considered as an important environmental factor that causes accelerated chromosomal ageing via increased telomere loss.

Hormones

The hormones have a substantial effect on adult male and female TL, and the impact of hormonal therapy was a subject of numerous studies [121, 122]. Entringer et al. assessed the influence of estriol (E_3) concentration in early gestation on children's TL. TL was measured in buccal cells collected in children aged 14.6 months on average. After accounting for the effects of gestational age at maternal blood draw during pregnancy and the child's age and sex, there was a significant, independent effect of maternal E_3 concentrations on children's TL [96].

Gestational and pre-gestational diabetes

Association studies of a mother's gestational or pre-gestational diabetes with TL showed inconsistent results. First, Cross et al. compared TL of cord blood mononuclear cells of newborns, whose mothers suffered from pre-gestational type 1 diabetes ($n = 26$), type 2 diabetes ($n = 20$) or gestational diabetes ($n = 71$), with newborns of mothers without diabetes ($n = 45$, $n = 76$ and $n = 81$, respectively) [106]. Covariates included maternal smoking status, age of both parents, and offspring's birth weight and sex. The study found no difference in cord blood TL in pregnancies of women with diabetes compared to control subjects, but it identified higher cord blood telomerase activity in type 1 and gestational diabetes. Similarly, a study of 20 cases of mothers with gestational diabetes and 14 cases of pre-gestational diabetes ($n = 7$ type 1 and $n = 7$ type 2) compared to 18 control mother-baby pairs also showed no significant difference in the newborn's cord blood TL [108]. On the other hand, another study demonstrated that a group of babies born to mothers with gestational diabetes ($n = 82$) had significantly shorter LTL than babies born to healthy mothers ($p = 0.028$). The model was adjusted for maternal age, gestational age at delivery, neonatal birth weight and foetal gender [107]. In conclusion, further studies should be performed to confirm the impact of gestational or pre-gestational diabetes on newborn TL.

Paternal age at conception

Finally, one of the most intriguing observations regarding TL is the fact that children with older fathers have longer telomeres in comparison to their peers with younger fathers. The effect of paternal age at conception (PAC) was also demonstrated in newborns, where a 1-year increase in PAC corresponded to a 0.016 kb increase of newborn LTL [109]. This phenomenon was attributed to elongation of telomeres in the sperm of older men due to a higher telomerase activity [81, 123]. Studies remain inconclusive whether it is maternal or paternal inheritance that has

Unauthenticated
Download Date | 8/30/19 1:36 PM

a greater impact on TL; however, the effects of both are already evident in newborns [92].

To conclude, Entringer et al. understood the process of newborn TL development as a foetal programming of health and disease risk: '… intrauterine life represents a particularly sensitive time period when the effects of maternal states and conditions around conception and across pregnancy may be transmitted to the developing embryo/foetus.' TL is a key cellular target of these effects and carries important implications in long-term health and susceptibility to common age-related disorders [11].

Conclusions

During the last three decades, extensive studies were continually providing the evidences that linked shortened telomeres with common age-related diseases, disease risk factors and longevity. Telomeres were considered as a potential biomarker that could assess susceptibility to a specific pathology, or even as a mitotic clock that would limit cell life span and predict one's longevity. Even though there are reasonable explanations behind these theories, there is still missing knowledge in telomere biology that is restraining the implementation of telomeres as biomarkers in clinical practice.

Only recently, scientists agreed that the best model to describe telomere dynamics is not simply measuring TL *per se*, but determining baseline (newborn) TL and telomere attrition over time and how it is affected by the number of cell divisions, oxidative stress and other exposures that may reduce TL.

Baseline TL of newborns significantly differs among individuals, corresponding to the variation of TL found among adults of the same age. Twin studies, however, showed a synchrony among TL in siblings and therefore provided proof of strong genetic determination of telomere biology. Genetic influence on TL was confirmed with the identification of genetic variants that were significantly related to TL in adult cohorts. Surprisingly, GWAS in the child population did not identify the same significant variants, implying that genetic regulation of TL could be age dependent. Genetic loci detected in adult studies probably modify TL over the life span, while different genetic mechanisms affect the TL of newborns.

Baseline TL significantly differs also among foetuses of the same gestational age. During early intrauterine life, cells divide excessively but maintain their TL with the telomerase enzyme. This results in a synchrony of TL among cells of different tissues within the same individual. The telomerase activity is mainly extinguished after early embryonic development, and telomeres of cells with different proliferating activities start to shorten at different rates. Because leucocytes represent highly proliferative tissue with a high rate of turnover compared to minimally proliferative muscle cells, a shorter LTL of a full-term newborn compared to the MTL of the same individual can be detected.

Cumulative risk exposures, especially during early life, together with genetic make-up account for the likelihood of developing a common age-related pathology. During intrauterine development, the foetus is continuously exposed to numerous environmental factors which can transfer through distinct biological pathways and affect baseline TL in a process of so-called foetal programming. Baseline TL reflects the early environmental conditions and represents an important cellular marker for susceptibility to common age-related pathologies. TL attrition can accelerate under conditions that cause increased oxidative stress and chronic inflammation and can determine newborn TL during pregnancy. Some of the unfavourable conditions can be modified: avoiding smoking during pregnancy, maintaining appropriate BMI before and during pregnancy, and ensuring a proper vitamin intake. Therefore, it is important to be aware of the influence of such conditions on the child's health. Maternal stress and gestational diabetes are less easily avoided but can be managed with proper clinical help. Finally, some of the determinants of newborn TL, such as the child's sex, race or genetics cannot be modified, but it is important to understand and take into account their impact on TL, especially in designing the studies exploring new telomere determinants or association of telomeres with pathologies.

Besides baseline TL, telomere attrition over time represents another important element in telomere dynamics. However, telomeric sequence loss is not constant throughout life; the first 4 years of childhood are characterised by a rapid decline in TL due to a large turnover of highly proliferative cells. In this period, the main difference in TL of tissues with different proliferative capacities is established. Later, telomeres in most cells are gradually and synchronically shortened. The variance among cells in different tissues of an individual can be significant, but it remains smaller than the variance in TL among unrelated individuals.

The dynamics of TL between 4 years of age and early adulthood remain to be elucidated. Some discoveries suggest that telomeres during this life period reach a plateau and maintain their length, whereas other results point to a gradual shortening of approximately 50 bp/year

Unauthenticated
Download Date | 8/30/19 1:36 PM

that starts after the age of 4 and remains stable throughout life. Indeed, the biggest problem of the first studies of TL in children and newborns is the small sample size and the lack of longitudinal follow-up. Despite that, these studies made some consistent conclusions which provided the first insights into telomere dynamics from the first phases of gestation until adulthood. TL at birth reflects TL in adulthood: children with short telomeres will most likely grow into adults with short telomeres. Environmental factors in adulthood may affect TL, but this influence is probably smaller than in the dynamics at birth and in the first years of life.

Knowing the importance of TL in childhood, it is crucial to have a good understanding of the dynamics of telomeres in intrauterine life and in the first years of life. This could help to prevent early attrition and thus decrease the risk for later pathophysiological conditions. It is important to keep in mind that the spectrum of diseases related to relative TL in adulthood is vast, and though these pathologies manifest only in adulthood, early developmental environment plays a significant role in their generation.

In view of this, we can conclude that there are still some missing gaps that need further exploration for a general understanding of telomere biology. First, larger genetics studies in the child population are warranted to detect genetic variants that influence baseline TL. This may result in the discovery of novel genetic loci involved in the pathways of telomere biology and help understanding biological regulation of telomere length in childhood. Second, longitudinal studies on telomere dynamics from birth until adulthood are necessary. Follow-up of individuals through long periods would give answers to the missing information of telomere dynamics after the age of 4 and would confirm the effects of early environmental conditions on adult susceptibility to common diseases. Finally, new studies on newborn telomere determinants are required, to confirm previously discovered relations and to find new potential conditions that significantly influence newborn TL.

With precise knowledge of the biology and dynamics of telomeres, we might expect that one day TL will be used as a common biomarker in everyday clinical practice.

Acknowledgements: Many thanks to Prof. Dr. Georges Weryha for his motivation for this project and kind discussions around it.
Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: Financial support was provided by the European Union within the frames of the Operational Programme FEDER-FSE Lorraine et Massif des Vosges 2014–2020, by Agence Nationale de la Recherche, programme d'Investissements d'avenir, grant number ANR-15RHU-0004 and GEENAGE project "(Functional Genomic, Epigenomic and ENvironment interplay to IMPACT management of healthy and pathological AGEing)" within the frames of the LUE project (Lorraine University of Excellence). Additionally, this work was supported by the regional project CPER-ITM2P 2015–2020.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organisation(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, et al. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci USA 1988;85:6622–6.
2. O'Sullivan RJ, Karlseder J. Telomeres: protecting chromosomes against genome instability. Nat Rev Mol Cell Biol 2010;11: 171–81.
3. Lindsey J, McGill NI, Lindsey LA, Green DK, Cooke HJ. In vivo loss of telomeric repeats with age in humans. Mutat Res 1991;256:45–8.
4. Fagagna FD, Reaper PM, Clay-Farrace L, Fiegler H, Carr P, von Zglinicki T, et al. A DNA damage checkpoint response in telomere-initiated senescence. Nature 2003;426:194.
5. Hiyama E, Hiyama K. Telomere and telomerase in stem cells. Br J Cancer 2007;96:1020.
6. Cong Y-S, Wright WE, Shay JW. Human telomerase and its regulation. Microbiol Mol Biol Rev 2002;66:407–25.
7. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PD, et al. Specific association of human telomerase activity with immortal cells and cancer. Science 1994;266:2011–5.
8. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. Nat Rev Genet 2010;11:319.
9. Aviv A, Shay JW. Reflections on telomere dynamics and ageing-related diseases in humans. Proc R Soc Lond B Biol Sci 2018;373:20160436.
10. Gorenjak V, Akbar S, Stathopoulou MG, Visvikis-Siest S. The future of telomere length in personalized medicine. Front Biosci (Landmark Ed) 2018;23:1628–54.
11. Entringer S, de Punder K, Buss C, Wadhwa PD. The fetal programming of telomere biology hypothesis: an update. Proc R Soc Lond B Biol Sci 2018;373:20170151.
12. Okuda K, Bardeguez A, Gardner JP, Rodriguez P, Ganesh V, Kimura M, et al. Telomere length in the newborn. Pediatr Res 2002;52:377–81.
13. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet 2005;6:611–22.

Unauthenticated
Download Date | 8/30/19 1:36 PM

14. Tzanetakou IP, Nzietchueng R, Perrea DN, Benetos A. Telomeres and their role in aging and longevity. Curr Vasc Pharmacol 2014;12:726–34.
15. Armanios M, Blackburn EH. The telomere syndromes. Nat Rev Genet 2012;13:693–704.
16. Dokal I. Dyskeratosis congenita. Hematology Am Soc Hematol Educ Program 2011;2011:480–6.
17. Dodson L, Bertuch AA. Dyskeratosis congenita and the telomere biology disorders. In: Kupfer G, Reaman G, Smith F, editors. Bone marrow failure. Cham: Springer, 2018:111–35.
18. Kirwan M, Dokal I. Dyskeratosis congenita: a genetic disorder of many faces. Clin Genet 2008;73:103–12.
19. Shay JW, Wright WE. Telomeres in dyskeratosis congenita. Nat Genet 2004;36:437–8.
20. Mitchell JR, Wood E, Collins K. A telomerase component is defective in the human disease dyskeratosis congenita. Nature 1999;402:551–5.
21. Dokal I. A disease of premature ageing. Lancet 2001;358:S27.
22. Glousker G, Touzot F, Revy P, Tzfati Y, Savage SA. Unraveling the pathogenesis of Hoyeraal-Hreidarsson syndrome, a complex telomere biology disorder. Br J Haematol 2015;170:457–71.
23. Armanios M. Syndromes of telomere shortening. Annu Rev Genomics Hum Genet 2009;10:45–61.
24. Engelhardt M, Ozkaynak MF, Drullinsky P, Sandoval C, Tugal O, Jayabose S, et al. Telomerase activity and telomere length in pediatric patients with malignancies undergoing chemotherapy. Leukemia 1998;12:13–24.
25. Tabori U, Dome JS. Telomere biology of pediatric cancer. Cancer Invest 2007;25:197–208.
26. Beeharry N, Broccoli D. Telomere dynamics in response to chemotherapy. Curr Mol Med 2005;5:187–96.
27. Shay JW, Wright WE. Telomeres and telomerase: three decades of progress. Nat Rev Genet 2019;20:299–309.
28. Jager K, Walter M. Therapeutic targeting of telomerase. Genes (Basel) 2016;7:39.
29. Shay JW, Wright WE. Telomerase therapeutics for cancer: challenges and new directions. Nat Rev Drug Discov 2006;5:577–84.
30. Haycock PC, Burgess S, Nounu A, Zheng J, Okoli GN, Bowden J, et al. Association between telomere length and risk of cancer and non-neoplastic diseases: a Mendelian randomization study. JAMA Oncol 2017;3:636–51.
31. Granot E, Kohen R. Oxidative stress in childhood – in health and disease states. Clin Nutr 2004;23:3–11.
32. Poonpet T, Saetan N, Tanavalee A, Wilairatana V, Yuktanandana P, Honsawek S. Association between leukocyte telomere length and angiogenic cytokines in knee osteoarthritis. Int J Rheum Dis 2018;21:118–25.
33. Belsky DW, Shalev I, Sears MR, Hancox RJ, Lee Harrington H, Houts R, et al. Is chronic asthma associated with shorter leukocyte telomere length at midlife? Am J Respir Crit Care Med 2014;190:384–91.
34. Li J, Tang B, Qu Y, Mu D. Telomerase reverse transcriptase: a novel neuroprotective mechanism involved in neonatal hypoxic-ischemic brain injury. Int J Dev Neurosci 2011;29:867–72.
35. Eisenberg DT, Borja JB, Hayes MG, Kuzawa CW. Early life infection, but not breastfeeding, predicts adult blood telomere lengths in the Philippines. Am J Hum Biol 2017;29:e22962.
36. Picarelli MM, Danzmann LC, Grun LK, Junior NT, Lavandovsky P, Guma F, et al. Arterial stiffness by oscillometric device and telomere length in juvenile idiopathic artrhitis with no cardiovascular risk factors: a cross-sectional study. Pediatr Rheumatol Online J 2017;15:34.
37. Tyrka AR, Price LH, Kao H-T, Porton B, Marsella SA, Carpenter LL. Childhood maltreatment and telomere shortening: preliminary support for an effect of early stress on cellular aging. Biol Psychiatry 2010;67:531–4.
38. Moffitt TE. Childhood exposure to violence and lifelong health: clinical intervention science and stress-biology research join forces. Dev Psychopathol 2013;25:1619–34.
39. Price LH, Kao H-T, Burgers DE, Carpenter LL, Tyrka AR. Telomeres and early-life stress: an overview. Biol Psychiatry 2013;73:15–23.
40. Andrew T, Aviv A, Falchi M, Surdulescu GL, Gardner JP, Lu X, et al. Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. Am J Hum Genet 2006;78:480–6.
41. Atzmon G, Cho M, Cawthon RM, Budagov T, Katz M, Yang X, et al. Evolution in health and medicine Sackler colloquium: genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. Proc Natl Acad Sci USA 2010;107(Suppl 1):1710–7.
42. Bakaysa SL, Mucci LA, Slagboom PE, Boomsma DI, McClearn GE, Johansson B, et al. Telomere length predicts survival independent of genetic influences. Aging Cell 2007;6:769–74.
43. Vasa-Nicotera M, Brouilette S, Mangino M, Thompson JR, Braund P, Clemitson J-R, et al. Mapping of a major locus that determines telomere length in humans. Am J Hum Genet 2005;76:147–51.
44. Slagboom PE, Droog S, Boomsma DI. Genetic determination of telomere size in humans: a twin study of three age groups. Am J Hum Genet 1994;55:876–82.
45. Njajou OT, Cawthon RM, Damcott CM, Wu S-H, Ott S, Garant MJ, et al. Telomere length is paternally inherited and is associated with parental lifespan. Proc Natl Acad Sci USA 2007;104:12135–9.
46. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet 2013;45:422–7, 427e1–2.
47. Gu J, Chen M, Shete S, Amos CI, Kamat A, Ye Y, et al. A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. Cancer Prev Res (Phila) 2011;4:514–21.
48. Levy D, Neuhausen SL, Hunt SC, Kimura M, Hwang S-J, Chen W, et al. Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. Proc Natl Acad Sci USA 2010;107:9293–8.
49. Liu Y, Cao L, Li Z, Zhou D, Liu W, Shen Q, et al. A genome-wide association study identifies a locus on TERT for mean telomere length in Han Chinese. PLoS One 2014;9:e85043.
50. Mangino M, Hwang S-J, Spector TD, Hunt SC, Kimura M, Fitzpatrick AL, et al. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. Hum Mol Genet 2012;21:5385–94.
51. Mangino M, Richards JB, Soranzo N, Zhai G, Aviv A, Valdes AM, et al. A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length. J Med Genet 2009;46:451–4.
52. Pooley KA, Bojesen SE, Weischer M, Nielsen SF, Thompson D, Amin Al Olama A, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified

Unauthenticated
Download Date | 8/30/19 1:36 PM

loci show little association with hormone-related cancer risk. Hum Mol Genet 2013;22:5056–64.
53. Prescott J, Kraft P, Chasman DI, Savage SA, Mirabello L, Berndt SI, et al. Genome-wide association study of relative telomere length. PLoS One 2011;6:e19635.
54. Saxena R, Bjonnes A, Prescott J, Dib P, Natt P, Lane J, et al. Genome-wide association study identifies variants in casein kinase II (CSNK2A2) to be associated with leukocyte telomere length in a Punjabi Sikh diabetic cohort. Circ Cardiovasc Genet 2014;7:287–95.
55. Codd V, Mangino M, van der Harst P, Braund PS, Kaiser M, Beveridge AJ, et al. Variants near TERC are associated with mean telomere length. Nat Genet 2010;42:197–9.
56. Lee JH, Cheng R, Honig LS, Feitosa M, Kammerer CM, Kang MS, et al. Genome wide association and linkage analyses identified three loci – 4q25, 17q23.2, and 10q11.21 – associated with variation in leukocyte telomere length: the Long Life Family Study. Front Genet 2014;4:310.
57. Zeiger AM, White MJ, Eng C, Oh SS, Witonsky J, Goddard PC, et al. Genetic determinants of telomere length in African American youth. Sci Rep 2018;8:13265.
58. Stathopoulou MG, Petrelis AM, Buxton JL, Froguel P, Blakemore AI, Visvikis-Siest S. Genetic determinants of leucocyte telomere length in children: a neglected and challenging field. Paediatr Perinat Epidemiol 2015;29:146–50.
59. Weng Q, Du J, Yu F, Huang T, Chen M, Lv H, et al. The known genetic loci for telomere length may be involved in the modification of telomeres length after birth. Sci Rep 2016;6:38729.
60. Vaziri H, Schachter F, Uchida I, Wei L, Zhu X, Effros R, et al. Loss of telomeric DNA during aging of normal and trisomy 21 human lymphocytes. Am J Hum Genet 1993;52:661–7.
61. Frenck RW Jr., Blackburn EH, Shannon KM. The rate of telomere sequence loss in human leukocytes varies with age. Proc Natl Acad Sci USA 1998;95:5607–10.
62. Zeichner SL, Palumbo P, Feng Y, Xiao X, Gee D, Sleasman J, et al. Rapid telomere shortening in children. Blood 1999;93:2824–30.
63. Friedrich U, Schwab M, Griese EU, Fritz P, Klotz U. Telomeres in neonates: new insights in fetal hematopoiesis. Pediatr Res 2001;49:252–6.
64. Akkad A, Hastings R, Konje JC, Bell SC, Thurston H, Williams B. Telomere length in small-for-gestational-age babies. BJOG 2006;113:318–23.
65. Wojcicki JM, Shiboski S, Heyman MB, Elwan D, Lin J, Blackburn E, et al. Telomere length change plateaus at 4 years of age in Latino children: associations with baseline length and maternal change. Mol Genet Genomics 2016;291:1379–89.
66. Vasu V, Turner KJ, George S, Greenall J, Slijepcevic P, Griffin DK. Preterm infants have significantly longer telomeres than their term born counterparts. PLoS One 2017;12:e0180082.
67. Holmes DK, Bellantuono I, Walkinshaw SA, Alfirevic Z, Johnston TA, Subhedar NV, et al. Telomere length dynamics differ in foetal and early post-natal human leukocytes in a longitudinal study. Biogerontology 2009;10:279–84.
68. Eastwood JR, Mulder E, Verhulst S, Peters A. Increasing the accuracy and precision of relative telomere length estimates by RT qPCR. Mol Ecol Resour 2018;18:68–78.
69. Youngren K, Jeanclos E, Aviv H, Kimura M, Stock J, Hanna M, et al. Synchrony in telomere length of the human fetus. Hum Genet 1998;102:640–3.
70. Kimura M, Gazitt Y, Cao X, Zhao X, Lansdorp PM, Aviv A. Synchrony of telomere length among hematopoietic cells. Exp Hematol 2010;38:854–9.
71. Wright DL, Jones EL, Mayer JF, Oehninger S, Gibbons WE, Lanzendorf SE. Characterization of telomerase activity in the human oocyte and preimplantation embryo. Mol Hum Reprod 2001;7:947–55.
72. Ponsot E, Lexell J, Kadi F. Skeletal muscle telomere length is not impaired in healthy physically active old women and men. Muscle Nerve 2008;37:467–72.
73. Decary S, Mouly V, Hamida CB, Sautet A, Barbet JP, Butler-Browne GS. Replicative potential and telomere length in human skeletal muscle: implications for satellite cell-mediated gene therapy. Hum Gene Ther 1997;8:1429–38.
74. Benetos A, Kimura M, Labat C, Buchoff GM, Huber S, Labat L, et al. A model of canine leukocyte telomere dynamics. Aging Cell 2011;10:991–5.
75. Daniali L, Benetos A, Susser E, Kark JD, Labat C, Kimura M, et al. Telomeres shorten at equivalent rates in somatic tissues of adults. Nat Commun 2013;4:1597.
76. Aviv A, Kark JD, Susser E. Telomeres, atherosclerosis, and human longevity: a causal hypothesis. Epidemiology 2015;26:295.
77. Sabharwal S, Verhulst S, Guirguis G, Kark JD, Labat C, Roche NE, et al. Telomere length dynamics in early life: the blood-and-muscle model. FASEB J 2017;32:529–34.
78. Gardner M, Bann D, Wiley L, Cooper R, Hardy R, Nitsch D, et al. Gender and telomere length: systematic review and meta-analysis. Exp Gerontol 2014;51:15–27.
79. Hansen ME, Hunt SC, Stone RC, Horvath K, Herbig U, Ranciaro A, et al. Shorter telomere length in Europeans than in Africans due to polygenetic adaptation. Hum Mol Genet 2016;25:2324–30.
80. Lynch SM, Peek MK, Mitra N, Ravichandran K, Branas C, Spangler E, et al. Race, ethnicity, psychosocial factors, and telomere length in a multicenter setting. PLoS One 2016;11:e0146723.
81. Kimura M, Cherkas LF, Kato BS, Demissie S, Hjelmborg JB, Brimacombe M, et al. Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. PLoS Genet 2008;4:e37.
82. Prescott J, Du M, Wong JY, Han J, De Vivo I. Paternal age at birth is associated with offspring leukocyte telomere length in the Nurses' Health Study. Hum Reprod 2012;27:3622–31.
83. Muezzinler A, Mons U, Dieffenbach AK, Butterbach K, Saum KU, Schick M, et al. Smoking habits and leukocyte telomere length dynamics among older adults: results from the ESTHER cohort. Exp Gerontol 2015;70:18–25.
84. Valdes AM, Andrew T, Gardner JP, Kimura M, Oelsner E, Cherkas LF, et al. Obesity, cigarette smoking, and telomere length in women. Lancet 2005;366:662–4.
85. Sassenroth D, Meyer A, Salewsky B, Kroh M, Norman K, Steinhagen-Thiessen E, et al. Sports and exercise at different ages and leukocyte telomere length in later life – data from the Berlin Aging Study II (BASE-II). PLoS One 2015;10:e0142131.
86. Coimbra BM, Carvalho CM, Moretti PN, Mello MF, Belangero SI. Stress-related telomere length in children: a systematic review. J Psychiatr Res 2017;92:47–54.
87. Edmonds GW, Hampson SE, Côté HC, Hill PL, Klest B. Childhood personality, betrayal trauma, and leukocyte telomere length in adulthood: a lifespan perspective on conscientiousness and betrayal traumas as predictors of a biomarker of cellular ageing. Eur J Pers 2016;30:426–37.

Unauthenticated
Download Date | 8/30/19 1:36 PM

88. Lindqvist D, Epel ES, Mellon SH, Penninx BW, Révész D, Verhoeven JE, et al. Psychiatric disorders and leukocyte telomere length: underlying mechanisms linking mental illness with cellular aging. Neurosci Biobehav Rev 2015;55:333–64.
89. Ma D, Zhu W, Hu S, Yu X, Yang Y. Association between oxidative stress and telomere length in type 1 and type 2 diabetic patients. J Endocrinol Invest 2013;36:1032–7.
90. Babizhayev MA, Vishnyakova KS, Yegorov YE. Oxidative damage impact on aging and age-related diseases: drug targeting of telomere attrition and dynamic telomerase activity flirting with imidazole-containing dipeptides. Recent Pat Drug Deliv Formul 2014;8:163–92.
91. Entringer S, Buss C, Wadhwa PD. Prenatal stress, telomere biology, and fetal programming of health and disease risk. Sci Signal 2012;5:pt12.
92. Aviv A. The mitochondrial genome, paternal age and telomere length in humans. Philos Trans R Soc Lond B Biol Sci 2018;373.
93. Salihu HM, King LM, Nwoga C, Paothong A, Pradhan A, Marty PJ, et al. Association between maternal-perceived psychological stress and fetal telomere length. South Med J 2016;109:767–72.
94. Marchetto NM, Glynn RA, Ferry ML, Ostojic M, Wolff SM, Yao R, et al. Prenatal stress and newborn telomere length. Am J Obstet Gynecol 2016;215:94.e1–8.
95. Send TS, Gilles M, Codd V, Wolf I, Bardtke S, Streit F, et al. Telomere length in newborns is related to maternal stress during pregnancy. Neuropsychopharmacology 2017;42:2407–13.
96. Entringer S, Epel ES, Lin J, Blackburn EH, Buss C, Simhan HN, et al. Maternal estriol concentrations in early gestation predict infant telomere length. J Clin Endocrinol Metab 2015;100:267–73.
97. Almanzar G, Eberle G, Lassacher A, Specht C, Koppelstaetter C, Heinz-Erian P, et al. Maternal cigarette smoking and its effect on neonatal lymphocyte subpopulations and replication. BMC Pediatr 2013;13:57.
98. Jones CW, Gambala C, Esteves KC, Wallace M, Schlesinger R, O'Quinn M, et al. Differences in placental telomere length suggest a link between racial disparities in birth outcomes and cellular aging. Am J Obstet Gynecol 2017;216:294.e1–8.
99. Kim JH, Kim GJ, Lee D, Ko JH, Lim I, Bang H, et al. Higher maternal vitamin D concentrations are associated with longer leukocyte telomeres in newborns. Matern Child Nutr 2018;14:1.
100. Entringer S, Epel ES, Lin J, Blackburn EH, Buss C, Shahbaba B, et al. Maternal folate concentration in early pregnancy and newborn telomere length. Ann Nutr Metab 2015;66:202–8.
101. Martens DS, Plusquin M, Gyselaers W, De Vivo I, Nawrot TS. Maternal pre-pregnancy body mass index and newborn telomere length. BMC Med 2016;14:148.
102. Wojcicki JM, Olveda R, Heyman MB, Elwan D, Lin J, Blackburn E, et al. Cord blood telomere length in Latino infants: relation with maternal education and infant sex. J Perinatol 2016;36:235–41.
103. Drury SS, Esteves K, Hatch V, Woodbury M, Borne S, Adamski A, et al. Setting the trajectory: racial disparities in newborn telomere length. J Pediatr 2015;166:1181–6.
104. Entringer S, Epel ES, Lin J, Buss C, Shahbaba B, Blackburn EH, et al. Maternal psychosocial stress during pregnancy is associated with newborn leukocyte telomere length. Am J Obstet Gynecol 2013;208:134.e1–7.
105. Salihu HM, Pradhan A, King L, Paothong A, Nwoga C, Marty PJ, et al. Impact of intrauterine tobacco exposure on fetal telomere length. Am J Obstet Gynecol 2015;212:205.e1–8.
106. Cross JA, Temple RC, Hughes JC, Dozio NC, Brennan C, Stanley K, et al. Cord blood telomere length, telomerase activity and inflammatory markers in pregnancies in women with diabetes or gestational diabetes. Diabet Med 2010;27:1264–70.
107. Xu J, Ye J, Wu Y, Zhang H, Luo Q, Han C, et al. Reduced fetal telomere length in gestational diabetes. PLoS One 2014;9:e86161.
108. Gilfillan C, Naidu P, Gunawan F, Hassan F, Tian P, Elwood N. Leukocyte telomere length in the neonatal offspring of mothers with gestational and pre-gestational diabetes. PLoS One 2016;11:e0163824.
109. Factor-Litvak P, Susser E, Kezios K, McKeague I, Kark JD, Hoffman M, et al. Leukocyte telomere length in newborns: implications for the role of telomeres in human disease. Pediatrics 2016;137:e20153927.
110. Mayer S, Bruderlein S, Perner S, Waibel I, Holdenried A, Ciloglu N, et al. Sex-specific telomere length profiles and age-dependent erosion dynamics of individual chromosome arms in humans. Cytogenet Genome Res 2006;112:194–201.
111. Moller P, Mayer S, Mattfeldt T, Muller K, Wiegand P, Bruderlein S. Sex-related differences in length and erosion dynamics of human telomeres favor females. Aging (Albany NY) 2009;1:733–9.
112. Barrett EL, Richardson DS. Sex differences in telomeres and lifespan. Aging Cell 2011;10:913–21.
113. Benetos A, Dalgard C, Labat C, Kark JD, Verhulst S, Christensen K, et al. Sex difference in leukocyte telomere length is ablated in opposite-sex co-twins. Int J Epidemiol 2014;43:1799–805.
114. Galtier-Dereure F, Boegner C, Bringer J. Obesity and pregnancy: complications and cost. Am J Clin Nutr 2000;71:1242S–8S.
115. McCloskey K, Ponsonby AL, Collier F, Allen K, Tang ML, Carlin JB, et al. The association between higher maternal pre-pregnancy body mass index and increased birth weight, adiposity and inflammation in the newborn. Pediatr Obes 2018;13:46–53.
116. Humphreys J, Epel ES, Cooper BA, Lin J, Blackburn EH, Lee KA. Telomere shortening in formerly abused and never abused women. Biol Res Nurs 2012;14:115–23.
117. Parks CG, Miller DB, McCanlies EC, Cawthon RM, Andrew ME, DeRoo LA, et al. Telomere length, current perceived stress, and urinary stress hormones in women. Cancer Epidemiol Biomarkers Prev 2009;18:551–60.
118. Entringer S, Epel ES, Kumsta R, Lin J, Hellhammer DH, Blackburn EH, et al. Stress exposure in intrauterine life is associated with shorter telomere length in young adulthood. Proc Natl Acad Sci USA 2011;108:E513–8.
119. Astuti Y, Wardhana A, Watkins J, Wulaningsih W. Cigarette smoking and telomere length: a systematic review of 84 studies and meta-analysis. Environ Res 2017;158:480–9.
120. Babizhayev MA, Savel'yeva EL, Moskvina SN, Yegorov YE. Telomere length is a biomarker of cumulative oxidative stress, biologic age, and an independent predictor of survival and therapeutic treatment requirement associated with smoking behavior. Am J Ther 2011;18:e209–26.
121. Lee D-C, Im J-A, Kim J-H, Lee H-R, Shim J-Y. Effect of long-term hormone therapy on telomere length in postmenopausal women. Yonsei Med J 2005;46:471–9.
122. Townsley DM, Dumitriu B, Liu D, Biancotto A, Weinstein B, Chen C, et al. Danazol treatment for telomere diseases. N Engl J Med 2016;374:1922–31.
123. Unryn BM, Cook LS, Riabowol KT. Paternal age is positively linked to telomere length of children. Aging Cell 2005;4:97–101.

Unauthenticated
Download Date | 8/30/19 1:36 PM

Publication under redaction

Discovering the Relation between Telomere Length and VEGF-A.

Introduction

A telomere is a repeated sequence of TTAGGG nucleotides on the ends of eukaryotic chromosomes, serving in the maintenance of the chromosomal integrity during cell replication (319). As it is gradually shortening with each mitotic division, it has been considered as an indicator of accelerated cellular ageing in the pathogenesis of atherosclerosis (96). This process is characterised by increased inflammation, mediated by a production of cytokines mainly from blood-borne inflammatory and immune cells, but also from vascular endothelial and smooth muscle cells (320).

VEGF-A is a key regulator of physiological and pathological angiogenesis and a mediator of vascular permeability in inflammatory disorders (185). It plays a critical role in cancer, where several VEGF-A signalling inhibitors have already been approved by the Food and Drug Administration (FDA) for oncological treatment (321). VEGF-A also seems to be important in CVDs (322), including atherosclerosis, where it promotes neoangiogenesis in the early stages of the disease as well as in the advanced atherosclerotic plaques (83).

VEGF-A and TL are promising biomarkers for detection and monitoring of chronic pathologies, as they are both a) involved in disease pathways, such as inflammation and angiogenesis (41, 323), b) highly heritable (44, 324) and c) have important inter-individual variability (325, 326). In combination, these two biomarkers could become a new tool for calculation of risk factors in personalised diagnostics of chronic diseases. The association of plasma concentrations of VEGF-A with leukocyte telomere length (LTL) has been previously studied (327). To expand the knowledge of the common pathways of both biomarkers, the current study aimed to investigate the genetic associations between VEGF-A and TL.

Accordingly, ten VEGF-A genome-wide significant variants that together explain up to 52% of the VEGF-A phenotypic variance (45) were investigated for association with LTL and muscle TL (MTL), as well as telomere attrition (the ratio between LTL and MTL, calculated as LTL/MTL) in a *case-control study* of atherosclerotic patients. The study is using LTL as an

approximation of highly proliferative tissue (hematopoietic system), MTL as an approximation of a minimally proliferative tissue (skeletal muscle) and their ratio LTL/MTL as an approximation of telomere attrition throughout life.

Methods

Population

The population of TELARTA study was genotyped for ten VEGF-A related genetic variants (402 individuals with available measurements of LTL and MTL as previously described (81)).

Briefly, 259 French individuals recruited in the discovery cohort of the TELARTA study were included in this population. The aim of the TELARTA study was to measure the TL in skeletal muscle and leukocytes of atherosclerotic CVD (ASCVD) patients and controls, in order to examine the role of telomere dynamics in arterial ageing. All participants provided written informed consent approved by the Ethics Committee (Comité de Protection des Personnes) of Nancy, France.

The replication cohort of the TELARTA study included 91 French individuals, recruited under the same conditions as discovery cohort and 52 individuals from an independent Greek population. The samples from three Greek hospitals have been collected at the Laboratory for Experimental Surgery and Surgical Research of the University of Athens. All participants enrolled in Athens provided written informed consent approved by the Ethics Committee of the University of Athens and Ethics Committee of each one of the three participating hospitals.

The VEGF-A plasma levels were measured in 324 French individuals from the TELARTA study.

Telomere length measurement

TL was measured in skeletal muscle and blood leukocytes. Muscle biopsies were obtained by surgeons and cardiologists from six different sites in different patients undergoing surgery or pacemaker/defibrillator implantation, as previously described in (81, 328). Skeletal muscle biopsy (100 to 200 mg) was collected from each individual and stored on -80 °C. Blood

samples from each patient were collected in EDTA tubes prior to surgery, and stored on -80 °C until further use.

TL measurement was performed by the Southern blot analysis of terminal restriction fragments, as previously described (329). Briefly, DNA was extracted from the muscle tissue and leukocyte cells by the phenol/chloroform method, and treated overnight with restriction enzymes *Hinf*I and *Rsa*I (Roche Diagnostics GmbH, Germany). Digested DNA samples and DNA ladder were resolved on 0.5% (wt/vol) agarose gels for 23 hours. After the depurination, denaturation and neutralisation of agarose gel, DNA was transferred on a positively charged nylon membrane (Roche) using a vacuum blotter (Biorad, Hercules, CA). Membranes were hybridised at 42 °C with the digoxigenin-labelled telomeric probe to enable the luminescent detection and readings using the charge-coupled device camera (Las 4000, Fuji). Measurements were performed in duplicate on separate gels. The measurement repeatability, as determined by the intraclass correlation coefficient, was 0.99 (95% confidence interval, 0.817–1.0) and 0.98 (95% confidence interval, 0.81–1.0) for LTL and MTL, respectively. The repeatability of the means of two duplicates, known as the extrapolated repeatability, was 0.995 and 0.991 for LTL and MTL, respectively.

VEGF-A protein measurement

VEGF-A protein was measured in plasma samples, using Cytokine Array I on Randox semi-automated benchtop immunoanalyser (Evidence Investigator Analyzer, Randox Laboratories Ltd., Crumlin, United Kingdom). Cytokine Array I is a high sensitivity multiplex cytokine and growth factor array, which enables simultaneous detection of 12 cytokines and growth factors in a single sample.

Genotyping

Ten VEGF-A related genetic variants (rs10761741, rs10738760, rs6921438, rs7043199, rs6993770, rs4416670, rs114694170, rs34528081, rs4782371 and rs2639990), previously identified in GWAS, were genotyped in blood cell DNA samples using a PCR-based KASP assay (284). Genotyping was performed by the Laboratory of the Government Chemist (LGC Ltd., Teddington, UK).

Statistical Analysis

Minor allele frequencies (MAF) and Hardy-Weinberg equilibrium (HWE) were calculated for ten VEGF-A related genetic variants (rs10761741, rs10738760, rs6921438, rs7043199, rs6993770, rs4416670, rs114694170, rs34528081, rs4782371 and rs2639990). The SNP rs10761741 did not follow the HW equilibrium and was excluded from further analyses.

The direct effects of VEGF-A related genetic variants on three phenotypes of interest (LTL, MTL and LTL/MTL) were tested using the PLINK toolset in the additive genetic model. Epistatic interactions were tested using PLINK and R package CAPE. LTL was log-transformed to follow a normal distribution.

The significance level for the direct effects of the 9 SNPs and the three tested phenotypes (LTL, MTL and LTL/MTL) was calculated as 0.05/9/3= 0.0012. For epistatic interactions, the significance level for the nine SNPs and three phenotypes was calculated as 0.05/36/3= 0.00046.

Multiple regression analysis was performed to study the association of VEGF-A plasma concentration with LTL, MTL and LTL/MTL. As VEGF-A concentrations were not normally distributed, a log-transformation was performed. VEGF-A plasma concentrations were compared between atherosclerosis cardiovascular disease (ASCVD) case and control groups using a non-parametric Mann-Whitney test.

In silico Analysis

The genomic environment of the significant SNPs was explored using Ensemble browser of the human genome (GRCh38.p12) and NCBI dbSNP.

Results

Association between VEGF-A related genetic variants and TL

Ten VEGF-A related genetic variants were tested for association with LTL, MTL and telomere attrition (LTL/MTL). Values of MAF and results of HWE analysis are presented in Table 11.

Table 11. Results of MAF and HWE analysis for ten VEGF-A related genetic variants in the study population.

SNP	Chromosome	MAF	HWE (P-Value)
rs114694170	5	0.04602	0.5802
rs34528081	6	0.3505	0.3788
rs6921438	6	0.4325	1
rs4416670	6	0.4527	0.4209
rs6993770	8	0.2873	1
rs7043199	9	0.206	0.7591
rs10738760	9	0.4739	0.617
rs10761741	10	0.4439	0.01131*
rs4782371	16	0.3282	0.1659
rs2639990	18	0.1136	0.6178

MAF: Minor allele frequency, **HWE**: Hardy-Weinberg equilibrium, *significant P-Value

The SNP rs10761741 did not follow the HWE and was excluded from further analyses. HWE investigates the inheritance of genetic polymorphisms tested in the population-genetic studies. If the measured genotype frequencies of SNP deviate from the expected one, the results should be treated cautiously because the observed genotype distribution in tested population does not represent genotype distribution in the overall population (330).

Nine SNPs altogether were tested for direct effects with LTL, MTL and LTL/MTL in the additive genetic model. LTL-values were log-transformed to follow the normal distribution. None of the SNPs was significantly associated with LTL (Table 12) or MTL (Table 13).

Table 12. Association between VEGF-A related genetic variants and LTL.

SNP	Chromosome	β	SE	P-Value
rs114694170	5	0.002083	0.007452	0.7799
rs34528081	6	-0.001541	0.003261	0.6369
rs6921438	6	-0.0004508	0.003173	0.8871
rs4416670	6	-0.002907	0.00323	0.3687
rs6993770	8	0.007	0.003496	0.04594
rs7043199	9	-0.00173	0.003846	0.6531
rs10738760	9	0.002185	0.003119	0.4841
rs4782371	16	0.002397	0.003291	0.4669
rs2639990	18	-0.001899	0.004965	0.7023

SE: Standard error, **β:** Effect size, P-value threshold is P<0.0012

Table 13. Association between VEGF-A related genetic variants and MTL.

SNP	Chromosome	β	SE	P-Value
rs114694170	5	0.1713	0.1132	0.1312
rs34528081	6	-0.07143	0.04946	0.1495
rs6921438	6	-0.06758	0.04827	0.1623
rs4416670	6	0.007034	0.04927	0.8866
rs6993770	8	-0.02124	0.05354	0.6917
rs7043199	9	-0.08772	0.05874	0.1361
rs10738760	9	0.03656	0.04753	0.4422
rs4782371	16	0.02819	0.05035	0.5758
rs2639990	18	0.06804	0.07592	0.3707

SE: Standard error, **β:** Effect size, P-value threshold is P<0.0012

A significant association was identified between minor allele (T) of rs6993770 and LTL/MTL (P=0.001143, β=0.0148, SE=0.004516) (Table 14).

Table 14. Association between VEGF-A related genetic variants and LTL/MTL.

SNP	Chromosome	β	SE	P-Value
rs114694170	5	-0.01288	0.009716	0.1856
rs34528081	6	0.004154	0.004246	0.3285
rs6921438	6	0.004166	0.004146	0.3156
rs4416670	6	-0.005934	0.004207	0.1592
rs6993770	8	0.0148	0.004516	0.001143*
rs7043199	9	0.004878	0.005062	0.3358
rs10738760	9	0.0003862	0.004077	0.9246
rs4782371	16	0.002534	0.004284	0.5546
rs2639990	18	-0.008642	0.006481	0.1832

SE: Standard error, **β:** Effect size, P-value threshold is P<0.0012, *significant P-value

Bioinformatics analysis showed that rs6993770 is located on the intron region (Figure 33) of the *ZFPM2* (zinc finger protein, FOG family member 2) gene.

There was no significant epistatic interaction identified for three tested phenotypes (LTL, MTL and LTL/MTL) according to the selected threshold (p<0.00046).

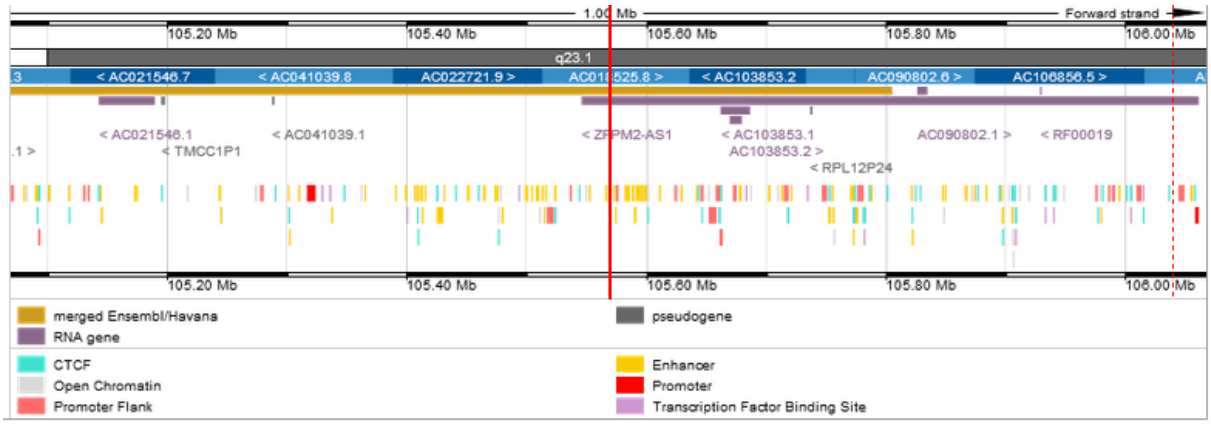


Figure 33: Rs6993770 (red stripe) is located on the intron of the *ZFPM2* gene (8q23.1).

Association between VEGF-A plasma concentrations and TL

VEGF-A plasma concentrations were tested for association with LTL, MTL and LTL/MTL using multiple regression analysis (Table 15). LTL and MTL values were adjusted for age and sex, whereas LTL/MTL values were adjusted only for age. There was no significant relation of VEGF-A plasma concentrations with investigated TL phenotypes.

Table 15: Multiple regression analysis of VEGF-A plasma concentration with LTL, MTL and LTL/MTL. Dependent variable: logVEGFA.

Independent Variable	Regression Coefficient ± SE	R^2	P-Value Level
LTL	-0.034 ± 0.029	0.4%	0.24
MTL	-0.045 ± 0.029	0.7%	0.12
LTL/MTL	0.000 ± 0.003	0.0%	0.99

SE: Standard error

Association between VEGF-A plasma concentrations and ASCVD

VEGF-A has been previously associated with atherosclerosis progression and lesion. To further explore its role in atherosclerosis, VEGF-A plasma concentrations of ASCVD patients (N =153) and controls (n=171) were compared. Results of the analysis showed that plasma concentrations of VEGF-A were significantly higher in the control group compared to the

group of ASCVD patients ($p<0.0001$). The difference remained significant after adjustment for age and sex.

Because treatment with lipid-lowering medication can significantly impact on VEGF-A plasma concentration, statin-treated individuals were excluded from the second analysis. The difference remained significant in VEGF-A plasma concentrations of the ASCVD patients (n=51) compared to controls (n=147) after the exclusion of the statin-treated individuals ($p=0.03$).

Discussion

The present study investigated genetic associations between VEGF-A and telomere length measured in leukocytes, skeletal muscle cells and their ratio (LTL/MTL). It identified a direct association between minor allele (T) of rs6993770 and LTL/MTL ($P=0.001143$). This result suggests a common genetic regulation between VEGF-A and telomere attrition, possibly through a molecular process that affects both biomarkers.

Rs6993770 is one of the most significant variants associated with circulating VEGF-A levels. The minor allele (T) of SNP has been previously related to decreased VEGF-A levels. Together with three other SNPs (rs6921438, rs4416670 and rs10738760), rs6993770 explained 48% of the heritability of serum VEGF-A levels (46). Besides VEGF-A, it has been related to variation in HDL cholesterol (84), erythrocyte count, IL-12 levels, and platelets (85).

In the blood, VEGF-A can be found in plasma, platelets and leukocytes (86). Several studies reported the correlation between the concentration of VEGF-A and platelets, which are particularly important in wound healing and may have a stimulating role in angiogenesis-dependent tumour growth through their function as transporters of VEGF-A (87, 88).

In the present study, minor allele (T) of rs6993770 was associated with increased LTL/MTL ratio, indicating the protective role of allele (T) in telomere attrition. Rs6993770 is located in the intron of the *ZFPM2* gene, coding for a FOG family member protein. The FOG (Friend of GATA) proteins can both activate and down-regulate expression of GATA-target genes, resulting in modulation of GATA family proteins activity. GATA proteins are crucial regulators of haematopoiesis and cardiogenesis *via* the control of haemoglobin synthesis (89).

The genetic variant rs6993770 could likely impact on the activity of the *ZFPM2* gene, which would, in turn, affect GATA protein regulation of haematopoiesis. The risk allele (A) of the identified SNP could lead to increased haematopoiesis, which may result in high cellular turnover and thus, faster telomere attrition (90). Moreover, increased haematopoiesis could lead to bigger production of platelets, which would explain the higher levels of VEGF-A in subjects with this risk variant and the previously identified association of rs6993770 with platelets (331). Such hypothesis seems plausible since telomere attrition was the only phenotype that was significantly related to the genetic variant, whereas LTL and MTL did not show significant association with rs6993770.

TL is known for its high interindividual variation, which is evident already at birth (91). Afterwards, numerous factors can accelerate the attrition, *e.g.* increased cellular turnover caused by increased inflammation, oxidative stress (92), or can have a protective effect on TL, *e.g.* regular sport activity, healthy diet (93). However, the effects of these factors on TL are rather small comparing to the large variations in the baseline TL among individuals (94). Therefore, the use of LTL as a biomarker could be limiting without comparison to the baseline TL. Determination of the LTL is not as informative as it would be in combination with telomere attrition over time. Accelerated attrition within an individual's TL is supposed to be a more accurate biomarker for describing pathological processes in the organism (95). As muscular cells present a minimally proliferative tissue, MTL can be considered as baseline TL, and overall LTL attrition can be calculated as LTL/MTL.

Shorter LTL in atherosclerotic disease has been previously explained as a consequence of the increased cellular turnover due to the chronic systemic inflammation (96). A recent study further developed this hypothesis and suggested that accelerated telomere attrition has been related to atherosclerosis, rather than shorter LTL. As the study demonstrated that differences in telomere attrition between atherosclerosis patients and controls were consistent across the age range of the studied population, it has been suggested that increased attrition in the early life is more likely be a major cause of shorter LTL in atherosclerosis patients (81). The finding of our study could be in agreement with this result. The risk allele (A) of SNP rs6993770 may cause increased haematopoiesis and thus increased leukocyte telomere attrition, especially in childhood, when cellular turnover is the highest. Shorter telomeres are linked to chromosomal instability, which causes apoptosis. Continuous apoptosis can lead to early replicative senescence of vascular tissue and thus the start of the early formation of atherosclerotic plaque.

Even though this hypothesis seems likely, no association was found between the distribution of the risk allele (A) of rs6993770 and atherosclerosis status in the studied population. We might, however, still suppose that allele (A) of rs6993770 presents a risk factor for faster telomere attrition, which can result in atherosclerosis. Nevertheless, its predictive role for the risk of the CVDs should be further investigated. Telomere attrition in the case-control studies of other telomere related pathologies, such as cancer, should also be investigated for an association with rs6993770.

Epistatic interactions between nine investigated genetic variants and three phenotypes have been analysed to detect a combined effect of genetic variants on investigated phenotypes. No significant epistatic interactions were observed at the established threshold level.

Besides the genetic association between TL and VEGF-A, this study also investigated the associations of VEGF-A plasma levels of with LTL, MTL and telomere attrition, and the differences in VEGF-A plasma levels of ASCVD patients and controls.

Linear regression analysis did not identify any association between VEGF-A plasma levels and LTL, MTL or telomere attrition. The previous studies which investigated the association between LTL and VEGF-A levels reported inconsistent findings. No statistically significant association was reported between LTL and plasma concentrations of VEGF-A in the longitudinal study population consisting of 87 subjects (332). On the other hand, a study of patients with knee osteoarthritis identified a negative correlation between VEGF-A plasma levels and LTL (327). Further studies are warranted to fully explore the association of both biomarkers.

Similarly, studies investigating the plasma levels of VEGF-A in groups of atherosclerotic patients and controls gave discordant results. VEGF-A levels were increased in patients with uncomplicated hyperlipidemia free of major underlying atherosclerosis and patients with hyperlipidemia plus established atherosclerosis (333), as well as in patients with peripheral artery disease (PAD) and CAD, compared to the controls (334). On the other hand, VEGF-A plasma concentration appeared to be similar in CAD patients and patients without diseased coronaries (335) in another study. Nevertheless, in all cases, statin-treated atherosclerosis patients had lower VEGF-A concentrations compared to untreated patients (333, 335).

In our study, ASCVD patients had significantly lower VEGF-A plasma levels compared to the controls before excluding the statin-treated individuals. When the analysis was repeated

only with individuals without prescribed lipid-lowering medication, this difference was smaller but remained significant.

To conclude, this is the first study investigating the association of VEGF-A related genetic variants with LTL, MTL and telomere attrition. The study identified a significant association between rs6993770 and telomere attrition, possibly through the modification of the expression of GATA proteins, which results in a direct impact on hematopoiesis and production of platelets. No association was observed between VEGF-A plasma levels and LTL, MTL and telomere attrition. The levels of VEGF-A did not significantly differ in groups or atherosclerosis patients and controls.

Chapter V. GENERAL DISCUSSION

In the present work, we explored DNA methylation and diverse association studies methodologies for the identification of novel biomarkers and common molecular pathways that affect intermediate phenotypes and risk factors for chronic diseases. Our findings aimed to serve as prediction tools, to contribute to new therapeutic possibilities and to expand the current knowledge of molecular regulation in common biological processes. Following the objectives pursued, we have demonstrated:

1. The state of the art of the large contribution that personalised medicine gave to the current health system, and the importance of further scientific investment to accompany it for the improvement of individualised patient care.

2. An association between increased TG levels and CpG sites cg08897188 and cg04580029. TG levels are an important intermediate risk factor for various chronic diseases. The identification of a CpG site associated with TG levels could serve as a risk assessment biomarker or could contribute to new treatment strategies.

3. An association between increased WC and CpG site cg16170243, which gave an insight into the complex regulation of obesity that merit further investigation. This discovery might be useful in the future determination of obesity risk factors, as well as in potential therapeutic applications.

4. An association between VEGF-A levels and 20 significant CpG sites, underlying the epigenetic mechanisms that contribute to the physiological regulation of this angiogenic biomarker.

5. Associations of VEGF-A with IL-4, MCP-1 and EGF, associations of $VEGF_{165}$ isoform with MCP-1 and IL-1α, and associations of $VEGF_{189}$ isoform with IL-4 and IL-6, which gave new insights into the links of inflammatory cytokines with VEGF-A and thus contributed to the understanding of the physiological interactions between angiogenesis and inflammation.

6. An association between *TREM2*-related genetic variant rs6918289 and increased TNF-α levels and IMT-F, suggesting that minor allele T of rs6918289 could be considered a risk factor for inflammation and atherosclerosis.

7. An association between VEGF-A related genetic variant rs6993770 and telomere attrition (defined as LTL/MTL) in an atherosclerosis case-control study, which suggested a common implication of VEGF-A and TL in the aetiology of CVDs.

8. A state of the art of the usefulness of TL measurements in future clinical applications in personalised medicine and the need for longitudinal research, starting from childhood. Such study will clarify the telomere biology in the first years of life and will assess the influence of childhood environmental factors on adult TL and susceptibility for common chronic diseases.

Personalised medicine has developed as an alternative to classical clinical approaches, based on the principle 'one size fits all' for treatment of patients with the same medical condition. It brought new strategies for better management of patients' health and therapies, combining and analysing individual's genomic information along with other clinical and diagnostic data. This enabled better identification of an individual's risk of developing a disease, faster detection of illnesses and determination of the most effective interventions (336). In this thesis work, we followed the principles of personalised medicine and used combined clinical and genetic information in research for new possibilities in the fight against common chronic diseases.

Common chronic diseases, such as CVD, cancer or type 2 diabetes, are the most common cause of premature death in adults (1). They do not occur at once but appear after a long-term accumulation of various environmental risk factors combined with an unfavourable genetic profile (2). Many of these factors are preventable, and early identification of patients at risk, combined with appropriate health promotion management could decrease the prevalence of chronic diseases even, in some cases, without the need of therapeutic interventions (3). Prevention of chronic diseases therefore does not start with the appearance of a disease itself, but with managing and fighting against the exposures known for deleterious impact on health, from where the need for new risk prediction biomarkers (4).

Intermediate phenotypes or quantitative traits are reflecting mechanism-related manifestations of complex diseases and can be used for molecular and genetic investigations (5). They have phenotypic variation continuously distributed in natural populations, which results from the segregation of alleles at multiple loci, and are influenced by complex genetic and

environmental factors. Quantitative genetic variation can underlie susceptibility to common complex diseases and genetic variants involved can be considered as potential risk factors (6, 7). Similarly, quantitative epigenetic variation can explain a portion of the variation of physiological traits and can serve as susceptibility biomarker (8).

Modifiable risk factors are expressed through intermediate phenotypes (also named intermediate risk factors), which include increased LDL cholesterol and increased TG levels for assessment of abnormal blood lipids, increased BMI and increased WC for assessment of obesity, and others. They are particularly interesting for research in personalised medicine because they are involved in multiple chronic diseases and can be avoided with the right preventive health measures (9, 10).

Biomarkers in personalised medicine allow the examination of normal biological processes, diagnosis of pathogenic processes and prognosis of the disease (11, 12). Nowadays, the identification of new diagnostic tools for the personalised medicine requires not only a good understanding of the pathology but also the use of a combination of genomic and molecular features, to detect variants with predictive value in clinics. As many diseases are linked through common pathological pathways, in particular inflammation or angiogenesis, those pathways are specifically important for investigation. Biomarkers identified under such conditions are likely to be involved in multiple diseases and have better potential to participate in the development of therapeutic strategies effective across comorbidities (13).

In this thesis work, we investigated intermediate risk factors and biological pathways that are involved in a variety of chronic diseases. We used different methodologies to assess their regulation and link them to personalised medicine.

First, we focused on epigenetic studies to investigate two risk factors involved in common chronic diseases, mentioned above: lipid levels and obesity. As previous EWAS of obesity focused mainly on BMI, we decided to investigate central obesity phenotypes, WC and waist-to-hip ratio, which we adjusted for BMI. These two studies identified altogether three risk variants – methylation sites, associated with the investigated phenotypes. Because VEGF-A presents an important biomarker in common chronic diseases, we decided to use the same methodology to investigate the epigenetic regulation of VEGF-A levels. This was in our knowledge the first study using EWAS for identification of methylation sites associated with VEGF-A. A particularity of the study worth mentioning was that VEGF-A levels were measured in PBMCs extracts, instead of using more common methods of determination in

plasma or serum samples. We identified many significant methylation sites which might help to explain VEGF-A epigenetic regulation, though none of them was identified on the gene which is directly involved in VEGF-A biology.

As one of the main hypothesis of the thesis was that a combination of biomarkers involved in the same pathology could better predict the susceptibility or outcomes of a disease, we continued our investigation of VEGF-A levels measured in PBMCs to unravel its links with cytokines and growth factors implicated in inflammation. We used VEGF-A protein levels, as well as mRNA levels of the four most abundant isoforms, which have different biological functions and should be systematically considered as separated biomarkers in biological studies. We identified interesting links with four cytokines and one growth factor, worthy of further attention.

TL is, as VEGF-A, involved in a variety of diseases and following the same hypothesis, we wished to unravel common links between both biomarkers to investigate their interplay in the pathogenesis of atherosclerosis. Relations between VEGF-A plasma concentrations and LTL have been previously studied. To expand the current knowledge we focused on their genetic links. We assessed direct associations between 10 GWAS identified SNPs of VEGF-A levels and TL measured in leukocytes, muscle cells and their ratio that reflects LTL attrition over time. We identified one genetic variant which was significantly associated with TL attrition and could present a risk factor for common chronic diseases, in which both, TL and VEGF-A, are involved. The study was performed on an atherosclerotic case-control population.

A similar approach for investigation of links between genetic variants and biomarkers was used in the research of associations between TREM2, TNF-α (biomarkers of inflammation) and IMT-F (indicator of atherosclerosis lesions). TREM2 is an attractive molecule which has been neglected in the frames of cardiovascular research and was more often related to neurodegenerative diseases. We identified one SNP, located in the region of the *TREM2* gene, which was significantly associated with TNF-α and IMT-F and could present a risk factor for atherosclerosis pathologies. This study was a part of the research project of our Unit, which dedicated itself to make the progress on TREM family molecule investigations.

The results of all publications form this thesis, which was dedicated to the discovery of new genetic and molecular biomarkers for identification of early risk factors for common chronic diseases, including lipids, obesity, inflammation, atherosclerosis and angiogenesis, are discussed below.

.

It is well-known that exposure to adverse environmental factors can induce DNA methylation changes long before disease appears (14). A DNA methylation biomarker that would show its relation with, *e.g.* increased TG levels or increased WC could alone, or in combination with other genetic biomarkers help in identification of individuals with high susceptibility for an increased risk for chronic diseases related to the observed phenotype. This could help in the early prevention of altered levels of intermediate risk factors with the right lifestyle management of individuals at risk (15).

Moreover, epigenetic modification of DNA is a key modulator of mRNA transcription, which is, unlike the DNA sequence, modifiable by environmental and pharmacological interventions. The environment can modify epigenetic patterns that play an important role in health. This information can be transmitted to daughter cells during cellular replication (16). Many therapeutic agents can alter methylation patterns and target sites related to unfavourable health outcomes (17). Subsequently, this methodology opened the gates for the discoveries of new therapeutic possibilities (18, 19).

Genome-wide DNA methylation studies are approaches for identification of novel biomarkers for risk assessment, diagnosis and prognosis, as well as for the potential therapeutic application. They are bringing new strategies for prevention and fight against diseases. Recent FDA approvals of DNA methylation-based diagnostic tools for colorectal cancer and DNA methylation-based therapeutics against haematological malignancies are the proof of the good direction and are giving the motivation for further exploration of the field (18, 19, 337).

Using the EWAS methodology, this thesis work reported two novel methylation sites linked to TG levels, one novel methylation site linked to WC, and 20 novel sites that explained epigenetic regulation of VEGF-A levels in physiological conditions, in the population of the healthy individuals from the SFS.

First, we investigated methylation patterns of TC, LDL, HDL and TG levels in a population of 211 healthy adults and children from the SFS, in research entitled "*Epigenome-Wide Association Study (EWAS) of Blood Lipids in Healthy Population from STANISLAS Family Study (SFS)*".

Deregulated blood lipid levels are among the most important risk factors of CVDs (20). They are influenced by genetics as well as environmental exposures, which can reflect on DNA

methylation patterns. Insight into these epigenetic patterns can explain the interindividual variability in lipid phenotypes as well as the underlying regulatory mechanisms (21). Previous studies that investigated the epigenetic regulation of lipids were performed on subjects suffering from a cardiovascular condition. Regulation of lipid traits can differ significantly in healthy population and novel CpG sites involved in gene regulation pathways can be discovered. Using the epigenome-wide approach, this study identified two CpG sites significantly related to TG levels, cg08897188 (chromosome 7q36.1, β=-2.80, $P=1.39 \times 10^{-8}$) and cg04580029 (chromosome 16p13.3, β=3.09, $P=5.75 \times 10^{-9}$).

Epigenetic DNA modifications are important regulators of tissue differentiation and it was previously demonstrated that DNA methylation could be tissue-specific (22, 23). Because methylation among different cell types can vary distinctively, accounting for blood cell type was used in the planning of this study, together with other required technical adjustments (fixed effects, methylation chip).

In research of new potential biomarkers, the availability of the tissue is an important point to consider. The collection of a tissue that demands a chirurgical intervention is less likely to be used in everyday clinical practice. Therefore, we believe that the methylated site should be detectable in the blood DNA and that EWAS of the blood DNA should be used as a discovery cohort in studies looking for new genetic biomarkers (24). However, we agree that for understanding the epigenetic regulation of a particular phenotype, such as TG, a phenotype-related tissue should also be considered for investigation, if available (25), as it can add an additional confirmation of the genetic mechanisms that are relating methylated site to the phenotype.

Adipose tissue is tightly linked to TG biology (26); therefore, it was used as a source for a DNA methylation assay in the replication cohort. Altogether, the DNA of the adipose tissue of 662 participants from the MuTHER study was tested for associations with cg08897188 and cg04580029. The association of CpG site cg04580029 and TG levels was replicated at a nominal significance level. The replication of the result in adipose tissue demonstrated that the whole blood DNA methylation is a relevant tissue for the identification of epigenomic changes, which is consistent with the conclusions of the previous studies (27, 28).

With bioinformatics analysis, we identified two genes in the vicinity of cg08897188 probe. *PRKAG2*, coding for AMP/ATP-binding subunit of AMP-activated protein kinase, and a non-coding transcript of *AC093583.1* gene, *LOC644090*. *PRKAG2* plays a key role in the lipid

metabolism *via* regulation of AMP/ATP-binding subunit of AMP-activated protein kinase (AMPK). AMPK maintains the energy balance within the cell and protects the body against metabolic disorders, such as obesity and type 2 diabetes (338). When activated, it switches off ATP-consuming pathways, *i.e.* glycogen, cholesterol and fatty acid synthesis and activates ATP-producing pathways, *i.e.* fatty acid oxidation (339). Mutations in the *PRKAG2* gene were previously associated with impaired glucose metabolism and excess glycogen storage in human cells (340) and can result in deformation of myocytes (341). Deformed myocytes can cause arrhythmias and hypertrophic cardiomyopathy-like manifestations (339). CpG cg08897188 could also affect a non-coding transcript of *AC093583.1* gene, *LOC644090*, located on the opposite strand. The effect of cg08897188 on TGs could be exerted *via* enhancing of the *LOC644090* transcription which would, in turn, down-regulate *PRKAG2*, or could directly affect *PRKAG2*.

Cg04580029 is located in the promoter of *KREMEN2* gene, coding for a secreted antagonist of the Wnt/β-catenin signalling pathway dickkopf1 (Dkk1) (342). Wnt signalling pathway is an important regulator of various cellular pathways, including cell migration, cell polarity, neural patterning and organogenesis during embryonic development (343). LRP6 (low-density lipoprotein receptor-related protein 6) is a co-receptor of Wnt signalling pathway, previously associated with AD and CAD (344, 345). Methylation at cg04580029 near transcription site could affect expression levels, which might result in modified TG concentrations. In addition, the methylation of the *KREMEN2* gene could also help in explaining the links of the Wnt/β-catenin signalling pathway with AD and CAD.

Both CpG sites related to TG levels were located in the regulatory rich genetic environment. Identification of significant methylation sites in the proximity of histone peak is especially important since histone modifications are the key component of transcriptional regulation (29). They organise the genome into active regions (DNA gets accessible for transcription) or inactive regions (DNA is compact and less accessible for transcription) and thus regulate gene expression (30). Methylation in the proximity of histone is therefore likely to be actively involved in the transcriptional regulation of the region.

To conclude, epidemiological, clinical and genetic studies demonstrated that elevated TG levels are a biomarker for CVD risk (31). As they contribute to the hardening of the arteries and thickening of the artery walls, they can increase the risk for CAD (32), stroke, heart attack and heart disease (33). With the identification of CpG sites significantly associated with TG

levels, we discovered two potential biomarkers that could help in the early assessment of the CVD risk or might be used in new methylation-based therapeutic strategies, which we can expect to be extensively developed in the future.

A similar approach was used for the investigation of the methylation patterns of WC and waist/hip ratio (adjusted for BMI) in research entitled "*Epigenome-Wide Association Study Detects a Novel Loci Associated with Central Obesity in Healthy Subjects*". The population consisted of 211 healthy adults and children from the SFS.

Obesity is an excessive fat accumulation that presents a risk for a number of comorbidities, including type 2 diabetes, CVD and cancer (35). Many negative effects are linked to the excessive fat tissue around the stomach, known as abdominal or central obesity, which is usually determined by WC (36). Obesity is a complex trait whose regulation is still not completely understood. Although big effort has been previously made to unravel the epigenetic regulatory mechanism, new cohorts are needed to contribute to the insights into epigenetic regulation of obesity phenotypes. Use of the populations of healthy individuals can help to avoid the discrepancies caused by extremes of obesity and comorbidity in population-specific cohorts (37). This study identified a novel significant positive association of WC adjusted for BMI with methylation site cg16170243 (chromosome 18q21.1, β=2.32, P_{adj}=0.048).

The association of cg16170243 and WC can be explained by two possible mechanisms. Firstly, non-specific environmental factors might cause the methylation at cg16170243, which would, in turn, lead to increased WC and greater susceptibility for central obesity. Secondly, the obesogenic environment could cause the methylation at cg16170243, which would result in the association of this methylation site with WC. Both mechanisms are equally plausible and are discussed in the manuscript, although more detailed hypotheses were developed in the article to explain the first hypothesis.

Briefly, the gene *ST8SIA5* overlapping the probe cg16170243 is coding for a sialyltransferase enzyme, involved in the synthesis of gangliosides. Gangliosides interact with molecules of signal transduction pathways, such as receptors tyrosine kinases (RTKs). Change in ganglioside composition induces the dissociation of RTKs from glycolipid-enriched microdomains, which may result in reduced phosphorylation of the receptors and cause insulin resistance. Thus, methylation of the CpG site could affect the expression level of the

ST8SIA5 gene, which would lead to the affected synthesis of gangliosides, followed by increased insulin resistance, which is a known risk factor for the development of obesity (38).

The second mechanism is suggesting that adiposity may influence DNA methylation, and this hypothesis was discussed in detail in the study of Wahl *et al.* (39). Their research demonstrated that adiposity caused the alterations in methylation at the majority of the identified CpG sites associated with BMI. The method used to investigate the causality of this relation was a GRS, which combines effects across SNPs known to influence BMI. In our study, the association between polymorphisms related to WC and methylation at cg16170243 has been tested, and no significant link was detected. However, we cannot exclude the possibility that increased adiposity determined the methylation at cg16170243. Further studies are warranted to determine the causal mechanism between the identified CpG site and WC.

The role of DNA methylation as a mediator on the causal pathway between modifiable exposures and disease risk can be assessed through two-sample Mendelian randomisation (MR), using EWAS summary statistics data of the observed phenotype. In the first step, the causal impact of the observed phenotype on epigenetic markers is established. In the second step, the casual nature of epigenetic markers on a health-related outcome is interrogated (40). Such an approach can help to identify relevant loci for future functional studies (24).

The third study using genome-wide methylation assay focused on the determination of the epigenetic regulation of VEGF-A levels in a population of 211 healthy adults and children from the SFS, and is presented in the publication entitled "*Epigenome-wide association study in healthy individuals identifies significant associations with DNA methylation and PBMCs extracts VEGF-A concentration*".

VEGF-A is an important mediator of angiogenesis and inflammation, as well as a key biological target for the treatment of several diseases related to these pathological pathways (41-43). The article identified 20 CpG sites (cg05739757, cg23333878, cg21838233, cg18815539, cg21968169, cg16333561, cg20547575, cg15014826, cg00117600, cg08759276, cg05275012, cg10517202, cg09614565, cg13689591, cg06934988, cg06785213, cg13332754, cg03551607, cg24364967, cg15057061) significantly associated with VEGF-A levels measured in PBMCs extracts, and discussed possible relations with genes, encoded in the vicinity of methylated sites.

Surprisingly, no identified CpG site was found in the regulatory region of a gene directly related to VEGF-A biology, such as *VEGF-A* (6p21.1), *FLT-1* (13q12.3), *KDR* (4q12), *HIF1A*

(4q23.2) *etc.* The reason for this might be the fact that increased levels of VEGF-A caused the methylation on unspecific sites. However, we identified methylation sites near genes that might be indirectly associated with VEGF-A regulation. Two most promising results were CpG cg21838233 (*TPX2* gene) and cg06785213 (*HAS1* gene). Firstly, *TPX2* is a gene involved in spindle fibre development, commonly overexpressed in gastric cancer cells (47). Studies showed that silencing of *TPX2* could inhibit proliferation and invasion of colon cancer cells, which is possibly related to the down-regulation of the *VEGF-A* gene (48). Secondly, *HAS1* encodes hyaluronic acid synthase, which has an essential role in tissue development and homeostasis of plasma membrane proteins and has been implicated in inflammatory and degenerative arthropathies. It has been previously shown that *HAS1*-siRNA can restrain the expression of *VEGF-A* (49).

Besides those two genes, links with VEGF-A related pathways were found for six other genes located in the vicinity of identified CpG sites. The methylation could explain the involvement in the regulation of *VEGF-A* expression that was previously identified (but not explained). The future exploration of DNA methylation related to VEGF-A levels should be, however, focused on specific populations, such as cancer or CVD patients. In cancer patients, VEGF-A production is up-regulated to provide enough signal for the angiogenic process, which enables vascularisation of a growing tumour (50). The role of DNA methylation in the regulation of VEGF-A overexpression could give interesting insights. In combination with the results of EWAS from the healthy population, such study could serve for the identification of differentially methylated sites and eventually contribute to the development of new alternatives to current antiangiogenic therapies.

VEGF-A levels in this study were measured in cellular extracts of PBMCs. PBMCs are a source of VEGF-A and many other signalling molecules. Their importance in different biological pathways has been previously reported (51). Determination of the VEGF-A production inside the mononuclear cells is therefore in straightforward link with related biological processes (angiogenesis and inflammation).

.

Because of the significant implication of VEGF-A in disease pathology *via* two major (and mutually dependent) pathological processes – angiogenesis and inflammation (52), VEGF-A levels detected in PBMCs extracts were further investigated for their association with 11 cytokines, important indicators of inflammation. The association of both physiological

pathways was demonstrated in the study entitled "*Peripheral blood mononuclear cells extracts VEGF protein levels and VEGF mRNA: Associations with inflammatory molecules in a healthy population*". The overall aim was to identify inflammatory molecules that could be used along with VEGF-A levels in predictive or therapeutic applications in personalised medicine.

As in the article above, particular originality of the study was the use of PBMCs extracts for measurement of all investigated molecules and determination of VEGF isoforms. PBMCs are cells that respond selectively to the immune system activation and are the major actors of the human body immunity, with an essential role in the synthesis and release of inflammatory cytokines and growth factors (53, 54). Simultaneous production of different mediators by the same cell group is likely to suggest their involvement in the same pathophysiological mechanism.

Both, molecular levels of VEGF-A as well as mRNA levels of four most abundant VEGF-A isoforms ($VEGF_{121}$, $VEGF_{145}$, $VEGF_{165}$ and $VEGF_{189}$) were investigated for their association with the inflammatory molecules IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, INF-γ, TNF-α, MCP-1 and EGF. Associations with VEGF-A protein levels were investigated in 285 individuals from the SFS, while a subset of 110 individuals was used for association study with *VEGF* mRNA isoforms. The study identified associations of VEGF-A with IL-4, MCP-1 and EGF, associations of $VEGF_{165}$ isoform with MCP-1 and IL-1α, and associations of $VEGF_{189}$ isoform with IL-4 and IL-6. Detailed bibliographical research has been done to explain the relations of identified cytokines with VEGF-A.

Briefly, IL-4, a multi-functional cytokine with anti-inflammatory and anti-tumour activity was associated with VEGF-A protein levels and $VEGF_{189}$ mRNA levels. IL-4 has been previously related to VEGF-A as an important factor in the recruitment of tumour-associated macrophages, which are known to promote angiogenesis (55). Moreover, increased *VEGF-A* mRNA expression was identified in IL-4 induced cell cultures of human airway smooth muscle cells (56). In this work, we identified $VEGF_{189}$ as the main isoform responsible for the association of IL-4 with VEGF-A levels.

MCP-1, an angiogenic chemokine involved in the regulation, migration and infiltration of monocytes and macrophages, was associated with VEGF-A protein levels and $VEGF_{165}$ mRNA levels. These findings were in accordance with previous studies, which showed that

MCP-1 up-regulated *HIF1A* gene expression in human aortic endothelial cells, which resulted in induced $VEGF_{165}$ expression in the aortic wall (57).

Another inflammatory molecule, EGF, was associated with VEGF-A protein levels in our population. EGF is an important mediator in growth, proliferation and differentiation of numerous cell types. The major role of both associated growth factors in tumorigenesis has been demonstrated in various types of cancers, and different mechanisms of their common pathways have been proposed. Combination of VEGF/EGF signalling pathways was studied for the development of new therapeutic approaches against cancer and is already showing positive predictions for the future development of new medications (58-61).

Finally, IL-6 and IL-1α were associated with $VEGF_{189}$ and $VEGF_{165}$, respectively. Previous studies on cell cultures detected induction of *VEGF-A* mRNA levels by both IL-6 and IL-1α. IL-6 was shown to induce VEGF-A-mediated angiogenesis in cervical tumorigenesis, gastric carcinoma patients (62) and in malignant mesothelioma (63). IL-1α stimulated the induction of *VEGF-A* mRNA synthesis in PBMCs in a dose-depended manner. Both IL-6 and IL-1α were linked together in the regulation of VEGF-A in pancreatic cancer (64) and in adenoma cells (65).

Physiological pathways in the human body are complex networks, regulated by various interacting molecules, which work together to build the physiological systems. Angiogenesis is the main process mediated by VEGF-A, which enables the growth of new blood vessels. However, other mechanisms, including inflammation, are tightly involved in its regulation (66, 67). The interplay of both mechanisms can result in many common pathologies and should be considered as a potential target for prevention of related diseases.

With this study, we successfully demonstrated links between VEGF-A and inflammation molecules derived from PBMCs extracts. This discovery gave insights into shared pathways between investigated molecules, which could be useful for applications in personalised medicine.

.

The third methodological approach used in this thesis was the study of genetic determinants for the investigation of the intermediate risk factors and biological pathways involved in different chronic diseases. The characteristics of the quantitative genetic variation were used

for identification of genetic biomarkers, which could serve in risk prediction of atherosclerosis, a common pathological process in the majority of CVDs (68).

In the study entitled "*The polymorphism rs6918289 located in the downstream region of the TREM2 gene is associated with TNF-alpha levels and IMT-F*", we focused on two quantitative traits, TNF-α levels and IMT-F, both important indicators of atherosclerosis. As recent discoveries showed that *TREM2* expression was found in human atherosclerotic lesions (69), the objective of our research was to study the genetic variants located in the vicinity of *TREM2* gene for an association with TNF-α levels and IMT-F.

First, we investigated five SNPs (rs7748777, rs6918289, rs7759295, rs9357347, rs6915083) located in the *TREM2* gene region (50 kb upstream and downstream of the *TREM2* gene, chromosome 6p21) for an association with TNF-α levels in 139 children from the SFS. We identified a new potential risk allele (T) of SNP rs6918289, which was associated with increased plasma levels of TNF-α ($\beta_{discovery}$=0.33, $P_{discovery}$=0.00147).

To increase the discovery population, SNP rs6918289 was *de novo* genotyped in 393 adults and 277 children. Analysis in combined SFS population reported a significant association between allele (T) of SNP rs6918289 and increased plasma levels of TNF-α (β_{SFS}=0.49, P_{SFS}=0.0017). These results were replicated in an independent cohort of 916 French individuals (β = 0.202, P = 0.023).

An association study was also performed with IMT-F in a sample of 350 adults from the SFS, where measurements of IMT-F were available. The results of the analysis showed significant association with minor allele (T) of the SNP rs6918289 (β = 0.024, P = 0.026).

Even though first researches related the TREM2 protein mostly with macrophage- and neutrophil-mediated inflammatory responses of neurodegenerative conditions, such as AD, amyotrophic lateral sclerosis, and Parkinson's disease (70), recent discoveries demonstrated the importance of TREM2-mediated macrophage response also in cardiovascular pathologies (69). The following article provided the first insights into the possible genetic association of TREM2 with TNF-α and IMT-F.

IMT-F is a non-invasive, early marker of atherosclerosis and an independent predictor of CVD that can help to detect an early atherosclerotic plaque formation by using ultrasound to calculate media-adventitia interface to the intima-lumen interface (71-73). The role of TNF-α in atherogenesis is explained through different mechanisms, by which this inflammatory

cytokine acts on endothelium (74). TNF-α can increase the vascular permeability to allow the infiltration of inflammatory cells (75) and promote migration of immune cells into the diseased or injured tissues, in order to protect the tissues from harmful pathogens. It can impair endothelium-dependent vasodilation by decreasing NO formation and bioavailability, stimulate vascular superoxide production and can interfere with endothelial insulin signalling to induce insulin resistance (76).

In the view of the above, the results of this study suggest that the minor allele (T) of the rs6918289, which was associated with increased TNF-α and IMT-F, might be considered a risk allele for inflammatory diseases and atherosclerosis. The mechanism of this effect is so far unknown; however, bioinformatics analysis of the genetic region indicated three possible hypotheses that could explain the association of rs6918289 and increased TNF-α levels.

Firstly, rs6918289 is involved in the post-transcriptional regulation and is associated with RNA binding protein PABPC1 (77), which binds to the 3' poly(A) region of the mRNAs. The minor allele (T) of rs6918289 could affect the PABPC1 protein, which would, in turn, disturb the stability of the *TREM2* mRNA. Previous studies have shown that TREM2 levels can affect TNF-α concentrations (78-80). The effect of the minor allele (T) would result in decreased TREM2 production, which would lead to elevated TNF-α levels in the same individual.

Secondly, rs6918289 is located between two CTCF (11-zinc finger protein) genomic sequences, which can bind together and form a chromatin loop, leading to a topologically associating domain (TAD) that can affect gene expression and other cellular functions. Thus, we hypothesise that the minor allele (T) of SNP rs6918289 could promote the binding of the CTCF and consequently trigger an insulation mechanism for the *TREM2* gene.

Finally, the rs6918289 can be found in proximity to the Triggering Receptor Expressed On Myeloid Cells Like 1 (*TREML1*) gene. *In silico* analysis showed a significant correlation of rs6918289 with seven SNPs related to *TREML1* gene. Therefore, *TREML1* should also be considered as a possible mediator of the rs6918289 effect on TNF-α concentration.

Remodelling of the artery is highly dependent on the inflammatory state; thus the association of the rs6918289 with IMT-F is most likely due to the increased TNF-α levels, produced by the minor allele (T) of SNP rs6918289, suggesting the indirect role of rs6918289 on IMT-F.

.

The study of the genetic variants was further utilised for investigation of the associations between two biomarkers, TL and VEGF-A. The aim of the research "*Discovering the Relation between Telomere Length and VEGF-A*" was to explore the association of LTL, MTL and telomere attrition (calculated as LTL/MTL) with 10 VEGF-A related genetic variants (rs10761741, rs10738760, rs6921438, rs7043199, rs6993770, rs4416670, rs114694170, rs34528081, rs4782371 and rs2639990) (45, 46). LTL represented TL in the highly proliferative hematopoietic cells, whereas MTL represented a minimally replicative skeletal muscle tissue, which can be comparable to the baseline TL (81).

Shorter TL has been previously associated with atherosclerosis and was considered as a possible risk factor for CVDs (82). Similarly, VEGF-A was demonstrated as a potential enhancer of the pathophysiologic mechanism of formation and destabilisation of atherosclerotic plaque (83). Besides their important role in atherogenesis, both biomarkers are implicated in other common chronic diseases. As the main population of this publication under redaction, the TELARTA study, is a case-control cohort of atherosclerosis patients, we investigated the role of VEGF-A and TL through this pathology. The results of the study, however, can be extrapolated to other diseases where both biomarkers play a distinct role, in particular cancer.

The present study identified links between the minor allele (T) of the SNP rs6993770 with LTL/MTL (P=0.001143). SNP rs6993770 is highly associated with VEGF-A plasma levels ($P = 2.44x10^{-60}$), as identified in a large GWAS (n>15,000) (45). We have conducted a bioinformatics analysis to investigate the genetic environment or the significant SNP and identify possible mechanisms that could help to explain the relation of VEGF-A and TL through significant genetic variant.

The bioinformatics analysis reviled that rs6993770, located on the *ZFPM2* gene, could impact on VEGF-A levels and telomere attrition through modification of *ZFPM2* activity. This would, in turn, affect the expression of GATA proteins, important regulators of haematopoiesis (346). Haematopoiesis is crucial for both investigated biomarkers. Increased cellular turnover causes faster telomere attrition, and also higher production of platelets, which are an important source of VEGF-A. The risk allele (A) of rs6993770 was associated with faster telomere attrition, platelet counts and higher concentrations of VEGF-A, which could all be a result of increased haematopoiesis.

Results of this study brought new insight into telomere regulation in diseases. The mechanism which relates shorter telomere to a variety of chronic diseases was investigated for a long time. One of the possible explanations for such an association was also faster LTL attrition, discussed in the recent studies (94). Our results suggested that faster telomere attrition could be dependent on genetic predispositions and possibly contribute to the pathogenesis of chronic diseases.

An interesting association between VEGF-A and telomeres has been identified from studies investigated the effects of genes involved in telomere regulation on VEGF-A. *hTERT* (human telomerase reverse transcriptase) was identified as an activator of *VEGF* gene expression through interactions with the *VEGF* promoter and the transcription factor Sp1 (347).

All these findings are suggesting a complex common regulation mechanism of both biomarkers, which is worthy of further investigation.

.

A telomere is indeed an important biomarker, related to a large variety of pathologies, starting from the rare monogenic disorders to common chronic diseases (97-99). Its special characteristics of continuous shortening throughout life have linked TL to longevity and intrigued the scientists looking for a biological clock that could predict a lifespan (100). Its role in prediction, diagnosis and treatment of diseases has been studied during the last few decades, yet it did not bring any promising clinical applications (101).

Because we believe that there is still much to be discovered, we have dedicated two review articles to summarise the current knowledge about TL, to assess its potential and most of all, to point out the knowledge gaps, which are obstructing the way to profound understanding of telomere biology and could lead to successful use in personalised medicine. We have concluded that TL could be a perspective biomarker, systematically measured from birth. Because of the high interindividual variability, a simple measurement of TL cannot be representative, whereas accelerated attrition could be extremely informative and could give the alert for the pathological changes in the individual. To assess this hypothesis, long-term longitudinal studies are warranted, with a special emphasis on the childhood period.

.

Besides all the above discoveries, during the preparation of this thesis work, we have actively participated in the European project MAST4HEALTH, HORIZON 2020 Marie Skłodowska-Curie actions - Research and Innovation Staff Exchange (MSCA RISE). Our role in this project was to determine the change in expression of inflammatory biomarkers, such as VEGF-A (*$VEGF_{145}$, $VEGF_{121}$, $VEGF_{165}$, $VEGF_{165b}$, $VEGF_{189}$*), TNF-α, and IL-6 after the treatment of NAFLD patients with a Greek plant mastiha, known for its antioxidant/anti-inflammatory and lipid-lowering properties. In the frames of the project, we have been also involved in 6-month secondment in *in vitro* diagnostics company Randox Laboratories, where we participated in developing and characterising of monoclonal antibody fragments, which will be used for a precise measurement of inflammatory biomarkers in human serum.

This experience was interesting from the scientific point of view; we autonomously prepared numerous experiments and thus contributed to the research of new diagnostic tools. Also, it was important from the collaborative perspective; the exchange of personnel created a tight link between the industrial and our research Unit. We believe that such applicative collaborations between research laboratories and industry can assure the fastest way from the intriguing discoveries towards successful application in patient care. The project MAST4HEALTH is still on course and the final results are expected in 2020.

Chapter VI. CONCLUSION AND PERSPECTIVES

During the three years of my research work, we used different methodologies to explore the risk factors and molecular pathways of chronic diseases in order to answer the need for biomarkers applicable in personalised medicine. We were building the knowledge based on the rich experiences and discoveries of the research Unit UMR INSERM U1122 IGE-PCV while integrating the most recent approaches for the investigation of the genetic epidemiology of common chronic diseases.

VEGF-A is one of the principal targeted biomarkers of our research Unit, which created and led the international **VEGF Consortium** (http://www.vegfconsortium.org/) that aims to develop a transnational collaborative network dedicated to large integrative and multidisciplinary genomic studies of VEGF. The collaborative projects of the VEGF Consortium, among others, consisted of two large GWAS that identified ten genetic variants related to VEGF-A.

The work of this thesis pushed the investigations of VEGF-A further and resulted in 1) identification of new associations with inflammatory molecules, 2) demonstration of genetic association with telomere attrition, and 3) first insights in epigenetic regulation.

More precisely, by:

- Studying the associations of several molecules that assist VEGF-A and its isoforms in the process of inflammation and angiogenesis. The findings should serve in the development of new therapeutic strategies against VEGF-A, which could be improved by acting on specific isoforms and their associated cytokines.
- Searching for the links between TL and VEGF-A. A genetic variant discovered in our study offered a new hypothesis for the explanation of the aetiology of telomere-related chronic diseases.
- Determining the epigenetic regulation of VEGF-A. This was the first essay to demonstrate genome-wide methylation mechanisms involved in VEGF-A biology.

Epigenetic studies are opening a new era in genetic epidemiology, enabling mechanistic insights into genetic and environmental risk factors for diseases, providing biomarkers of exposure and helping to localise disease-relevant genomic regions to identify potential intervention targets.

During the last few years, we were building a strategy to perform EWAS and carried out altogether three studies using the healthy SFS population as discovery cohort. Besides VEGF-

A, we focused on epigenetic markers related to intermediate risk factors common in chronic diseases and demonstrated the methylation changes related to WC and to TG levels. Given the large possibilities for clinical application of disease-relevant methylation sites, these results have a high potential for future use in personalised medicine.

TL is the second important biomarker highlighted in this thesis, also presenting a scientific target of our Unit. Our participation in the projects LUE: **GEENAGE** (Functional Genomic, Epigenomic and Environment interplay to impact the understanding, diagnosis and management of healthy and pathological Ageing) and **TELARTA** further encouraged our investment to explore telomere biology, which resulted in exciting collaborations. The investigation of the genetic associations with VEGF-A has been one of them.

As for the next challenges on the way towards the application of telomeres as a biomarker in personalised medicine, we performed bibliographic research that allowed us to understand their genetic regulation, their important role in diseases and their association with other inflammatory molecules. Our particular interest in TL in children motivated us to design a study, which would unravel genetic and environmental determinants of childhood TL, to fill in the knowledge gaps arising from the lack of longitudinal studies and to finally test the hypotheses that are linking the fast TL attrition in the childhood to chronic diseases in adulthood. A review article that gathered the most pertinent information on childhood telomere was the first step of this process.

Finally, the TREM family proteins have been extensively studied by the researches of our Unit in recent years. The research on TREM2 presented in this thesis was a part of the work that has demonstrated an important role of TREM-related genetic variants in cardiovascular pathologies. TREM2, as a less known member of its protein family, has a big potential for novel discoveries that may be useful for application in personalised healthcare.

All the efforts of this thesis were dedicated to the improvements of personalised medicine: understanding of biological pathways and genetic regulation of biomarkers, epigenetic mechanisms of risk factors and associations between molecules. The summary of the results acquired through the main research projects, showing the identified links between biomarkers, risk factors and molecular pathways investigated is schematically presented in Figure 34.

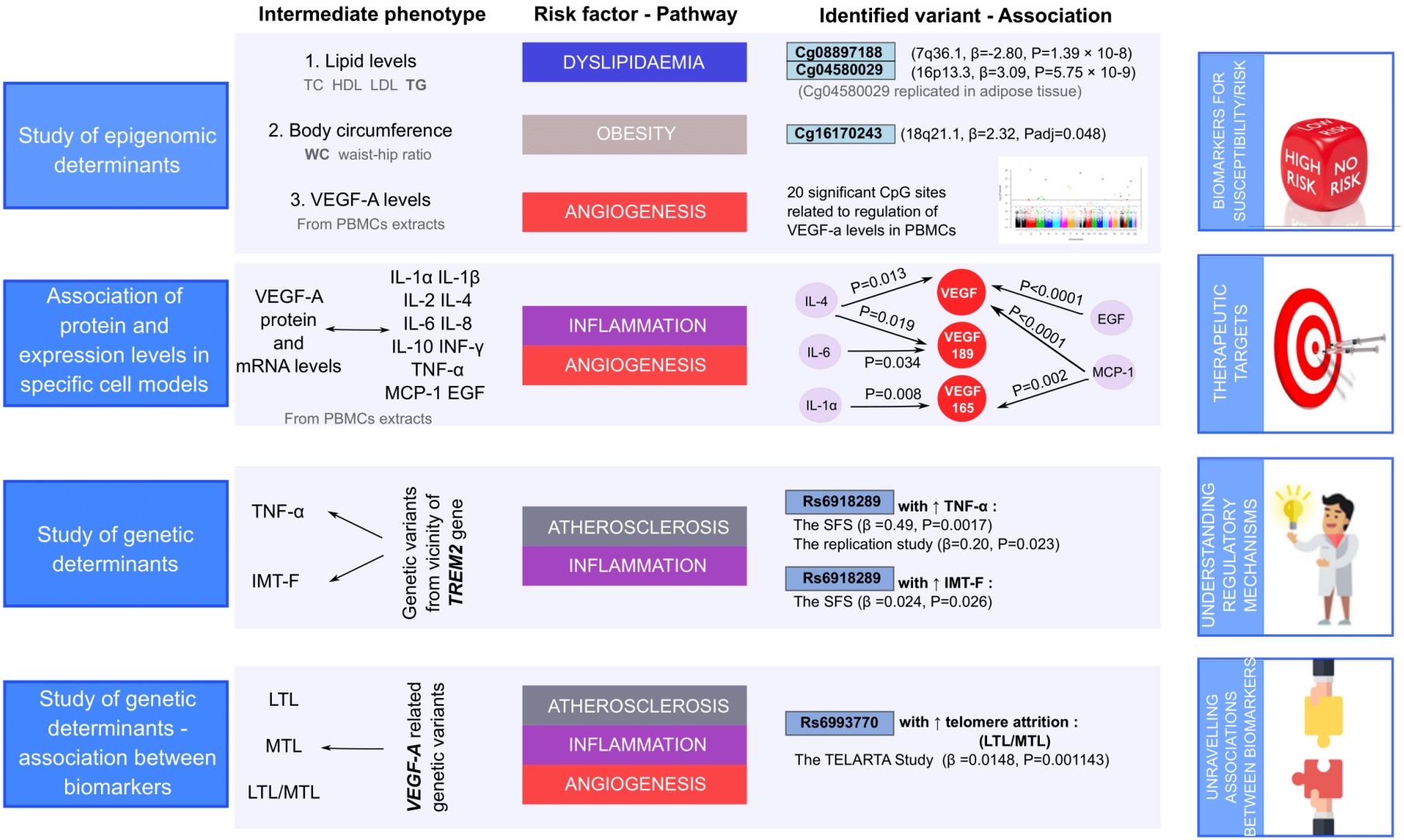


Figure 34: Schematic presentation of applied methods, investigated intermediate phenotypes, corresponding risk factors, and identified results.

Personalised medicine made an enormous breakthrough in the last decades, and our research Unit was contributing to its progress with rigorous research work, as well as with the preparation of the biannual congress on Systemes Medicine and Personalised Health and Therapy: **the Santorini Conferences series**. To honour the 9th edition, we have gathered all the most important achievement from the field of the personalised medicine in a review article, which was named after the conference: The Odyssey from hope to practice. We hope that this article will be an encouragement for the further efforts that pave the way of implementation of personalised medicine in everyday clinical practice.

6.1 Perspectives of our results

All genetic variants identified in our studies need a functional characterisation (validation) to tests the hypothesis generated by genomic and statistical approaches *in vitro* or *in vivo*. This can be done using molecular assays (*e.g.* mRNA expression) or by biological assays (*e.g.* human cell lines, animal models) (348).

More specifically, the perspectives of our results are the following:

VEGF-A

The studies of associations between VEGF-A and inflammatory molecules unravelled interesting links. However, systematic research of common molecular pathways of VEGF-A and inflammatory molecules needs further investigation. Firstly, the research should be repeated in a bigger sample of healthy individuals, and parallel measurement of the same biomarkers should be performed in plasma. This will enable a comparison of the cytokine and VEGF-A production by PBMCs to plasma concentrations released by all cytokine-producing cell types. Afterwards, cell cultures of PBMCs should be established and treated by angiogenic or inflammation stimuli to measure the production of molecules and mRNA isoforms under specific conditions. Associated molecules should then be measured in specific patients (cancer, cardiovascular pathologies) to consider their potential role as therapeutic targets.

The epigenomic research of VEGF-A reported in this thesis was the pilot study of VEGF-A methylation marks. As VEGF-A expression levels of different isoforms are also available for some of the individuals of the SFS, an EWAS could be performed to identify the epigenetic regulation of VEGF-A expression. Furthermore, the study should be repeated in a bigger sample, in a case-control population, where cancer patient without angiogenic therapy should be included. This could potentially result in the identification of differentially methylated genes involved in the regulation of angiogenesis. Such discovery would be extremely important for the novel generation of antiangiogenic treatments.

We have demonstrated an association of VEGF-A related genetic variant rs6993770 with telomere attrition, which could suggest common genetic regulatory mechanisms that impact both biomarkers and could thus be very interesting for application in personalised medicine. The associations of VEGF-A related genetic variants have been previously investigated with other phenotypes and diseases, such as lipids, AD, type 2 diabetes, metabolic syndrome *etc*. To systematically assess the role of VEGF-A in pathologies, intermediate risk factors and regulation of other inflammatory molecules, a Mendelian randomisation study should be performed, using the VEGF-A related genetic variants as instrumental variables for VEGF-A levels.

TG and WC

It would be necessary to replicate the identified CpG sites associated with TG and WC using bigger cohorts that would include obese populations. Because WC is related to increased deposit of adipose tissue on the waist, CpG site identified for its relation with WC should also be replicated using adipose DNA samples. Mendelian randomisation studies would be useful for identification of the causal role of identified CpGs in common chronic diseases.

Telomere length

VEGF-A related genetic variant rs6993770 that showed association with telomere attrition could explain interesting links between biomarkers and importance of TL in the aetiology of chronic diseases. LTL, MTL and telomere attrition should therefore be investigated for association with rs6993770 in populations of patients suffering from different pathologies previously associated with shorter LTL (numerous metabolic and inflammatory diseases, such

as CVD, diabetes mellitus, ulcerative colitis, liver cirrhosis and systemic lupus erythematous). This will help to determine whether the identified variant is associated with accelerated shortening in all diseases, or it is specific for atherosclerosis patients. SNP rs6993770 should be genotyped in bigger atherosclerosis case-control population, to investigate its potential for a susceptibility/risk biomarker in atherosclerotic diseases.

Longitudinal research that would follow individuals from birth to adulthood is essential in the efforts of implementation of TL as diagnostics/prognostic/susceptibility biomarker in personalised medicine. With the fast development of WGS technologies, TL could be measured using bioinformatics methods. Progress in this field is therefore necessary.

TREM-2

The genetic variant rs6918289 located in the *TREM2* gene region should be investigated in relation to TREM2 molecular levels, to determine to which extent it can affect TREM2 plasma concentrations. This would help to explain the mechanisms that are linking rs6918289 to TNF-α and IMT-F. Transcriptomic analyses should be performed afterwards to explain the functionality of the SNP and confirm the proposed hypothesis. The genetic variant should be tested in big atherosclerosis case-control population, to investigate its potential for a susceptibility/risk biomarker in atherosclerotic diseases.

BIBLIOGRAPHY

1. WHO. Chronic diseases [Available from: https://www.who.int/ncds/en/
2. Niccoli T, Partridge L. Ageing as a risk factor for disease. Curr Biol. 2012;22(17):R741-52.
3. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380(9836):37-43.
4. Helfand M, Buckley D, Fleming C, Fu R, Freeman M, Humphrey L, et al. Screening for intermediate risk factors for coronary heart disease. 2009.
5. Goldman D, Ducci F. Deconstruction of vulnerability to complex diseases: enhanced effect sizes and power of intermediate phenotypes. ScientificWorldJournal. 2007;7:124-30.
6. Silverman EK, Palmer LJ. Case-control association studies for the genetics of complex respiratory diseases. Am J Respir Cell Mol Biol. 2000;22(6):645-8.
7. Mackay TF. Q&A: Genetic analysis of quantitative traits. J Biol. 2009;8(3):23.
8. Jammes H, Junien C, Chavatte-Palmer P. Epigenetic control of development and expression of quantitative traits. Reprod Fertil Dev. 2011;23(1):64-74.
9. Marinou K, Tousoulis D, Antonopoulos AS, Stefanadi E, Stefanadis C. Obesity and cardiovascular disease: from pathophysiology to risk stratification. Int J Cardiol. 2010;138(1):3-8.
10. Hokanson JE, Austin MA. Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol level: a meta-analysis of population-based prospective studies. J Cardiovasc Risk. 1996;3(2):213-9.
11. Califf RM. Biomarker definitions and their applications. Exp Biol Med (Maywood). 2018;243(3):213-21.
12. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010;5(6):463-6.
13. Torkamani A, Topol EJ, Schork NJ. Pathway analysis of seven common diseases assessed by genome-wide association. Genomics. 2008;92(5):265-72.
14. Wojdacz TK. Methylation biomarker development in the context of the EU regulations for clinical use of in-vitro diagnostic devices. Expert review of molecular diagnostics. 2019;19(6):439-41.
15. Cordero P, Li J, Oben JA. Epigenetics of obesity: beyond the genome sequence. Curr Opin Clin Nutr Metab Care. 2015;18(4):361-6.
16. Mitchell C, Schneper LM, Notterman DA. DNA methylation, early life environment, and health outcomes. Pediatr Res. 2015;79:212.
17. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet. 2012;13(2):97-109.
18. Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD. Epigenetic therapy in immune-oncology. Nature Reviews Cancer. 2019;19(3):151-61.
19. Fardi M, Solali S, Farshdousti Hagh M. Epigenetic mechanisms as a new approach in cancer treatment: An updated review. Genes & diseases. 2018;5(4):304-11.
20. Durrington P. Dyslipidaemia. Lancet. 2003;362(9385):717-31.
21. Sayols-Baixeras S, Irvin MR, Arnett DK, Elosua R, Aslibekyan SW. Epigenetics of Lipid Phenotypes. Curr Cardiovasc Risk Rep. 2016;10(10).
22. Lokk K, Modhukur V, Rajashekar B, Märtens K, Mägi R, Kolde R, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biol. 2014;15(4):r54.
23. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, et al. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet. 2009;5(8):e1000602.
24. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman Å K, Aslibekyan S, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. PLoS Med. 2017;14(1):e1002215.
25. Huang Y-T, Chu S, Loucks EB, Lin C-L, Eaton CB, Buka SL, et al. Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood. Epigenetics. 2016;11(3):227-36.
26. Gong Y, Dou LJ, Liang J. Link between obesity and cancer: role of triglyceride/free fatty acid cycling. Eur Rev Med Pharmacol Sci. 2014;18(19):2808-20.

27. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. Lancet. 2014;383(9933):1990-8.
28. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum Mol Genet. 2015;24(15):4464-79.
29. Kouzarides T. Histone methylation in transcriptional control. Curr Opin Genet Dev. 2002;12(2):198-209.
30. Abcam. Histone modifications 2019 [Available from: https://www.abcam.com/epigenetics/histone-modifications-a-guide.
31. Budoff M. Triglycerides and Triglyceride-Rich Lipoproteins in the Causal Pathway of Cardiovascular Disease. Am J Cardiol. 2016;118(1):138-45.
32. Reiner Z. Hypertriglyceridaemia and risk of coronary artery disease. Nat Rev Cardiol. 2017;14(7):401-11.
33. Handelsman Y, Shapiro MD. TRIGLYCERIDES, ATHEROSCLEROSIS, AND CARDIOVASCULAR OUTCOME STUDIES: FOCUS ON OMEGA-3 FATTY ACIDS. Endocr Pract. 2017;23(1):100-12.
34. Miller M, Stone NJ, Ballantyne C, Bittner V, Criqui MH, Ginsberg HN, et al. Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. Circulation. 2011;123(20):2292-333.
35. WHO. Obesity 2019 [Available from: https://www.who.int/topics/obesity/en/.
36. Després J-P, Lemieux I. Abdominal obesity and metabolic syndrome. Nature. 2006;444(7121):881.
37. Dhana K, Braun KVE, Nano J, Voortman T, Demerath EW, Guan W, et al. An Epigenome-Wide Association Study (EWAS) of Obesity-Related Traits. Am J Epidemiol. 2018.
38. Kahn BB, Flier JS. Obesity and insulin resistance. J Clin Invest. 2000;106(4):473-81.
39. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2016;541:81.
40. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012;41(1):161-76.
41. Zittermann SI, Issekutz AC. Endothelial growth factors VEGF and bFGF differentially enhance monocyte and neutrophil recruitment to inflammation. J Leukoc Biol. 2006;80(2):247-57.
42. Pandey AK, Singhi EK, Arroyo JP, Ikizler TA, Gould ER, Brown J, et al. Mechanisms of VEGF (Vascular Endothelial Growth Factor) inhibitor–associated hypertension and vascular disease. Hypertension. 2018;71(2):e1-e8.
43. Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. Nature. 2000;407(6801):249.
44. Berrahmoune H, Herbeth B, Lamont JV, Masson C, Fitzgerald PS, Visvikis-Siest S. Heritability for plasma VEGF concentration in the Stanislas family study. Ann Hum Genet. 2007;71(Pt 1):54-63.
45. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, et al. Six Novel Loci Associated with Circulating VEGF Levels Identified by a Meta-analysis of Genome-Wide Association Studies. PLoS Genet. 2016;12(2):e1005874.
46. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, et al. Identification of cis- and trans-acting genetic variants explaining up to half the variation in circulating vascular endothelial growth factor levels. Circ Res. 2011;109(5):554-63.
47. Tomii C, Inokuchi M, Takagi Y, Ishikawa T, Otsuki S, Uetake H, et al. TPX2 expression is associated with poor survival in gastric cancer. World J Surg Oncol. 2017;15.
48. JIAN J, HUANG Y, LIU L-z, LI S, DENG F. TPX2 gene-silencing inhibits the proliferation and invasion of human colon cancer SW480 cells. TUMOR. 2016;36(6):628-34.
49. Wang G, Zhao W, Gao X, Zhang D, Li Y, Zhang Y, et al. HNF1AAS1 promotes growth and metastasis of esophageal squamous cell carcinoma by sponging miR214 to upregulate the expression of SOX-4. Int J Oncol. 2017;51(2):657-67.

50. Costache MI, Ioana M, Iordache S, Ene D, Costache CA, Saftoiu A. VEGF Expression in Pancreatic Cancer and Other Malignancies: A Review of the Literature. Rom J Intern Med. 2015;53(3):199-208.
51. Visvikis-Siest S, Marteau JB, Samara A, Berrahmoune H, Marie B, Pfister M. Peripheral blood mononuclear cells (PBMCs): a possible model for studying cardiovascular biology systems. Clin Chem Lab Med. 2007;45(9):1154-68.
52. Shaik-Dasthagirisaheb YB, Varvara G, Murmura G, Saggini A, Potalivo G, Caraffa A, et al. Vascular Endothelial Growth Factor (VEGF), Mast Cells and Inflammation. Int J Immunopathol Pharmacol. 2013;26(2):327-35.
53. De Jager W, te Velthuis H, Prakken BJ, Kuis W, Rijkers GT. Simultaneous detection of 15 human cytokines in a single sample of stimulated peripheral blood mononuclear cells. Clin Diagn Lab Immunol. 2003;10(1):133-9.
54. Turner MD, Nedjai B, Hurst T, Pennington DJ. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. Biochim Biophys Acta. 2014;1843(11):2563-82.
55. Linde N, Lederle W, Depner S, van Rooijen N, Gutschalk CM, Mueller MM. Vascular endothelial growth factor-induced skin carcinogenesis depends on recruitment and alternative activation of macrophages. J Pathol. 2012;227(1):17-28.
56. Faffe DS, Flynt L, Bourgeois K, Panettieri RA, Shore SA. Interleukin-13 and Interleukin-4 Induce Vascular Endothelial Growth Factor Release from Airway Smooth Muscle Cells: Role of Vascular Endothelial Growth Factor Genotype. Am J Respir Cell Mol Biol. 2006;34(2):213-8.
57. Hong KH, Ryu J, Han KH. Monocyte chemoattractant protein-1-induced angiogenesis is mediated by vascular endothelial growth factor-A. Blood. 2005;105(4):1405-7.
58. Ciardiello F, Troiani T, Bianco R, Orditura M, Morgillo F, Martinelli E, et al. Interaction between the epidermal growth factor receptor (EGFR) and the vascular endothelial growth factor (VEGF) pathways: a rational approach for multi-target anticancer therapy. Ann Oncol. 2006;17 Suppl 7:vii109-14.
59. Tortora G, Ciardiello F, Gasparini G. Combined targeting of EGFR-dependent and VEGF-dependent pathways: rationale, preclinical studies and clinical applications. Nat Clin Pract Oncol. 2008;5(9):521-30.
60. Tabernero J. The role of VEGF and EGFR inhibition: implications for combining anti-VEGF and anti-EGFR agents. Mol Cancer Res. 2007;5(3):203-20.
61. Winder T, Lenz HJ. Vascular endothelial growth factor and epidermal growth factor signaling pathways as therapeutic targets for colorectal cancer. Gastroenterology. 2010;138(6):2163-76.
62. Huang SP, Wu MS, Shun CT, Wang HP, Lin MT, Kuo ML, et al. Interleukin-6 increases vascular endothelial growth factor and angiogenesis in gastric carcinoma. J Biomed Sci. 2004;11(4):517-27.
63. Adachi Y, Aoki C, Yoshio-Hoshino N, Takayama K, Curiel DT, Nishimoto N. Interleukin-6 induces both cell growth and VEGF production in malignant mesotheliomas. Int J Cancer. 2006;119(6):1303-11.
64. Tang RF, Wang SX, Zhang FR, Peng L, Wang SX, Xiao Y, et al. Interleukin-1alpha, 6 regulate the secretion of vascular endothelial growth factor A, C in pancreatic cancer. Hepatobiliary Pancreat Dis Int. 2005;4(3):460-3.
65. Borg SA, Kerry KE, Royds JA, Battersby RD, Jones TH. Correlation of VEGF production with IL1 alpha and IL6 secretion by human pituitary adenoma cells. Eur J Endocrinol. 2005;152(2):293-300.
66. Loizzi V, Del Vecchio V, Gargano G, De Liso M, Kardashi A, Naglieri E, et al. Biological Pathways Involved in Tumor Angiogenesis and Bevacizumab Based Anti-Angiogenic Therapy with Special References to Ovarian Cancer. Int J Mol Sci. 2017;18(9).
67. Savant SS, Sriramkumar S, O'Hagan HM. The Role of Inflammation and Inflammatory Mediators in the Development, Progression, Metastasis, and Chemoresistance of Epithelial Ovarian Cancer. Cancers (Basel). 2018;10(8).
68. Lusis AJ. Atherosclerosis. Nature. 2000;407(6801):233-41.
69. Fernández-Ruiz I. Redefining leukocytes in atherosclerosis. Nature Reviews Cardiology. 2018;15(6):319-.

70. Jiang T, Yu J-T, Zhu X-C, Tan L. TREM2 in Alzheimer's disease. Mol Neurobiol. 2013;48(1):180-5.
71. Polak JF, Pencina MJ, Pencina KM, O'Donnell CJ, Wolf PA, D'Agostino RB, Sr. Carotid-wall intima-media thickness and cardiovascular events. N Engl J Med. 2011;365(3):213-21.
72. Giannoukas AD, Antoniou GA, Saleptsis V, Baros C, Griffin M, Nicolaides AN. Common femoral artery intima-media thickness as marker for cardiovascular disease in asymptomatic adults. Vasa. 2009;38(2):147-54.
73. Touboul PJ, Hennerici MG, Meairs S, Adams H, Amarenco P, Bornstein N, et al. Mannheim Carotid Intima-Media Thickness Consensus (2004–2006). Cerebrovasc Dis. 2007;23(1):75-80.
74. Urschel K, Cicha I. TNF-α in the cardiovascular system: from physiology to therapy. Internat J Interferon Cytokine Med Res. 2015;7:9-25.
75. Petrache I, Birukova A, Ramirez SI, Garcia JG, Verin AD. The role of the microtubules in tumor necrosis factor-alpha-induced endothelial cell permeability. Am J Respir Cell Mol Biol. 2003;28(5):574-81.
76. Steyers CM, 3rd, Miller FJ, Jr. Endothelial dysfunction in chronic inflammatory diseases. Int J Mol Sci. 2014;15(7):11324-49.
77. Guo L, Du Y, Chang S, Zhang K, Wang J. rSNPBase: a database for curated regulatory SNPs. Nucleic Acids Res. 2013;42(D1):D1033-D9.
78. Turnbull IR, Gilfillan S, Cella M, Aoshi T, Miller M, Piccio L, et al. Cutting edge: TREM-2 attenuates macrophage activation. J Immunol. 2006;177(6):3520-4.
79. Liu D, Dong Y, Liu Z, Niu B, Wang Y, Gao X. Impact of TREM-2 gene silencing on inflammatory response of endotoxin-induced acute lung injury in mice. Molecular and cellular biochemistry. 2014;394(1-2):155-61.
80. Gao X, Dong Y, Liu Z, Niu B. Silencing of triggering receptor expressed on myeloid cells-2 enhances the inflammatory responses of alveolar macrophages to lipopolysaccharide. Molecular medicine reports. 2013;7(3):921-6.
81. Benetos A, Toupance S, Gautier S, Labat C, Kimura M, Rossi PM, et al. Short leukocyte telomere length precedes clinical expression of atherosclerosis: the blood-and-muscle model. Circ Res. 2018;122(4):616-23.
82. Khan S, Chuturgoon AA, Naidoo DP. Telomeres and atherosclerosis. Cardiovasc J Afr. 2012;23(10):563-71.
83. Camaré C, Pucelle M, Nègre-Salvayre A, Salvayre R. Angiogenesis in the atherosclerotic plaque. Redox biology. 2017;12:18-34.
84. Stathopoulou MG, Bonnefond A, Ndiaye NC, Azimi-Nezhad M, El Shamieh S, Saleh A, et al. A common variant highly associated with plasma VEGFA levels also contributes to the variation of both LDL-C and HDL-C. J Lipid Res. 2013;54(2):535-41.
85. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46(D1):D754-d61.
86. Kut C, Mac Gabhann F, Popel AS. Where is VEGF in the body? A meta-analysis of VEGF distribution in cancer. Br J Cancer. 2007;97:978.
87. George ML, Eccles SA, Tutton MG, Abulafi AM, Swift RI. Correlation of plasma and serum vascular endothelial growth factor levels with platelet count in colorectal cancer: clinical evidence of platelet scavenging? Clin Cancer Res. 2000;6(8):3147-52.
88. Benoy I, Salgado R, Colpaert C, Weytjens R, Vermeulen PB, Dirix LY. Serum interleukin 6, plasma VEGF, serum VEGF, and VEGF platelet load in breast cancer patients. Clin Breast Cancer. 2002;2(4):311-5.
89. GeneCards. The human gene database 2019 [Available from: https://www.genecards.org/.
90. Rufer N, Brummendorf TH, Kolvraa S, Bischoff C, Christensen K, Wadsworth L, et al. Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. J Exp Med. 1999;190(2):157-67.
91. Youngren K, Jeanclos E, Aviv H, Kimura M, Stock J, Hanna M, et al. Synchrony in telomere length of the human fetus. Hum Genet. 1998;102(6):640-3.

92. Houben JM, Moonen HJ, van Schooten FJ, Hageman GJ. Telomere length assessment: biomarker of chronic oxidative stress? Free Radic Biol Med. 2008;44(3):235-46.
93. Ludlow AT, Zimmerman JB, Witkowski S, Hearn JW, Hatfield BD, Roth SM. Relationship between physical activity level, telomere length, and telomerase activity. Med Sci Sports Exerc. 2008;40(10):1764.
94. Aviv A, Shay JW. Reflections on telomere dynamics and ageing-related diseases in humans. Proc R Soc Lond B Biol Sci. 2018;373(1741):20160436.
95. Entringer S, de Punder K, Buss C, Wadhwa PD. The fetal programming of telomere biology hypothesis: an update. Proc R Soc Lond B Biol Sci. 2018;373(1741):20170151.
96. Samani NJ, Boultby R, Butler R, Thompson JR, Goodall AH. Telomere shortening in atherosclerosis. Lancet. 2001;358(9280):472-3.
97. Yang M, Prescott J, Poole EM, Rice MS, Kubzansky LD, Idahl A, et al. Prediagnosis Leukocyte Telomere Length and Risk of Ovarian Cancer. Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology. 2017;26:339-45.
98. Sanders JL, Newman AB. Telomere length in epidemiology: a biomarker of aging, age-related disease, both, or neither? Epidemiol Rev. 2013;35:112-31.
99. Falandry C, Horard B, Bruyas A, Legouffe E, Cretin J, Meunier J, et al. Telomere length is a prognostic biomarker in elderly advanced ovarian cancer patients: a multicenter GINECO study. Aging (Milano). 2015;7:1066-76.
100. Tzanetakou IP, Nzietchueng R, Perrea DN, Benetos A. Telomeres and their role in aging and longevity. Curr Vasc Pharmacol. 2014;12(5):726-34.
101. Gorenjak V, Akbar S, Stathopoulou MG, Visvikis-Siest S. The future of telomere length in personalized medicine. Front Biosci (Landmark Ed). 2018;23:1628-54.
102. Kuh D, Shlomo YB. A life course approach to chronic disease epidemiology: Oxford University Press; 2004.
103. Molster CM, Bowman FL, Bilkey GA, Cho AS, Burns BL, Nowak KJ, et al. The Evolution of Public Health Genomics: Exploring Its Past, Present, and Future. Frontiers in public health. 2018;6:247-.
104. Whitson HE, Boyd CM. Managing multiple comorbidities.
105. WHO. Cardiovascular diseases (CVDs) 2019 [Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
106. Scott J. Pathophysiology and biochemistry of cardiovascular disease. Curr Opin Genet Dev. 2004;14(3):271-9.
107. Glass CK, Witztum JL. Atherosclerosis: the road ahead. Cell. 2001;104(4):503-16.
108. Falk E. Pathogenesis of atherosclerosis. J Am Coll Cardiol. 2006;47(8 Supplement):C7-C12.
109. Leon BM, Maddox TM. Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment recommendations and future research. World J Diabetes. 2015;6(13):1246-58.
110. Association AD. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2010;33(Supplement 1):S62-S9.
111. WHO. WHO diabetes 2019 [Available from: https://www.who.int/health-topics/diabetes.
112. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. Lancet. 2014;383(9911):69-82.
113. Cersosimo E, Triplitt C, Solis-Herrera C, Mandarino LJ, DeFronzo RA. Pathogenesis of Type 2 Diabetes Mellitus. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, Dungan K, Grossman A, et al., editors. Endotext. South Dartmouth (MA): MDText.com, Inc.; 2000.
114. Szabo M, Máté B, Csép K, Benedek T. Genetic Approaches to the Study of Gene Variants and Their Impact on the Pathophysiology of Type 2 Diabetes. Biochem Genet. 2018;56(1):22-55.
115. Association AD. Gestational diabetes mellitus. Diabetes Care. 2004;27(suppl 1):s88-s90.
116. Robitaille J, Grant AM. The genetics of gestational diabetes mellitus: evidence for relationship with type 2 diabetes mellitus. Genet Med. 2008;10:240.
117. WHO. Cancer 2019 [Available from: https://www.who.int/cancer/en/.
118. UK Cr. Cancer statistics 2019 [Available from: https://www.cancerresearchuk.org/.

119. Cooper GM. The Cell, 2nd edition. Sunderland (MA): Sinauer Associates; 2000.
120. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science. 2017;355(6331):1330-4.
121. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA Cancer J Clin. 2018;68(1):31-54.
122. Turnbull C, Sud A, Houlston RS. Cancer genetics, precision prevention and a call to action. Nat Genet. 2018:1.
123. Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L. Cancer epigenetics: moving forward. PLoS Genet. 2018;14(6):e1007362.
124. The LDE. Should we officially recognise obesity as a disease? The lancet Diabetes & endocrinology. 2017;5(7):483.
125. Frellick M, Vega CP. Is Obesity a Disease? The Health Professionals Point of View: Medscape; 2018 [Available from: https://www.medscape.org/viewarticle/905706.
126. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. Metabolism. 2019;92:6-10.
127. Ortega FB, Lavie CJ, Blair SN. Obesity and cardiovascular disease. Circ Res. 2016;118(11):1752-70.
128. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nature Reviews Cancer. 2004;4(8):579-91.
129. García-Giustiniani D, Stein R. Genetics of Dyslipidemia. Arq Bras Cardiol. 2016;106(5):434-8.
130. Thompson GR. Management of dyslipidaemia. Heart. 2004;90(8):949-55.
131. Soran H, Adam S, Mohammad JB, Ho JH, Schofield JD, Kwok S, et al. Hypercholesterolaemia - practical information for non-specialists. Arch Med Sci. 2018;14(1):1-21.
132. Maxfield FR, Tabas I. Role of cholesterol and lipid organization in disease. Nature. 2005;438(7068):612.
133. Badimon L, Vilahur G. LDL-cholesterol versus HDL-cholesterol in the atherosclerotic plaque: inflammatory resolution versus thrombotic chaos. Ann N Y Acad Sci. 2012;1254:18-32.
134. Lin CJ, Lai CK, Kao MC, Wu LT, Lo UG, Lin LC, et al. Impact of cholesterol on disease progression. Biomedicine (Taipei). 2015;5(2):7.
135. Wingo TS, Cutler DJ, Wingo AP, Le N-A, Rabinovici GD, Miller BL, et al. Association of Early-Onset Alzheimer Disease With Elevated Low-Density Lipoprotein Cholesterol Levels and Rare Genetic Coding Variants of APOBAssociation of Early-Onset AD With Elevated LDL-C Levels and Rare APOB Coding VariantsAssociation of Early-Onset AD With Elevated LDL-C Levels and Rare APOB Coding Variants. JAMA Neurology. 2019;76(7):809-17.
136. Pejic RN, Lee DT. Hypertriglyceridemia. J Am Board Fam Med. 2006;19(3):310-6.
137. Toth PP. Triglyceride-rich lipoproteins as a causal factor for cardiovascular disease. Vasc Health Risk Manag. 2016;12:171-83.
138. Pang J, Chan DC, Watts GF. Origin and therapy for hypertriglyceridaemia in type 2 diabetes. World J Diabetes. 2014;5(2):165-75.
139. Tenenbaum A, Klempfner R, Fisman EZ. Hypertriglyceridemia: a too long unfairly neglected major cardiovascular risk factor. Cardiovasc Diabetol. 2014;13:159.
140. WHO. Physical activity 2019 [Available from: https://www.who.int/dietphysicalactivity/pa/en/.
141. Mora S, Cook N, Buring JE, Ridker PM, Lee IM. Physical activity and reduced risk of cardiovascular events: potential mediating mechanisms. Circulation. 2007;116(19):2110-8.
142. Mok A, Khaw KT, Luben R, Wareham N, Brage S. Physical activity trajectories and mortality: population based cohort study. BMJ. 2019;365:l2323.
143. WHO. Tobacco Free Initiative 2019 [Available from: https://www.who.int/tobacco/mpower/en/.
144. Yanbaeva DG, Dentener MA, Creutzberg EC, Wesseling G, Wouters EF. Systemic effects of smoking. Chest. 2007;131(5):1557-66.

145. Salahuddin S, Prabhakaran D, Roy A. Pathophysiological Mechanisms of Tobacco-Related CVD. Glob Heart. 2012;7(2):113-20.
146. Hecht SS. Cigarette smoking: cancer risks, carcinogens, and mechanisms. Langenbeck's archives of surgery. 2006;391(6):603-13.
147. Dhingra R, Vasan RS. Age as a risk factor. Med Clin North Am. 2012;96(1):87-91.
148. Partridge L. The new biology of ageing. Philos Trans R Soc Lond B Biol Sci. 2010;365(1537):147-54.
149. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1736-88.
150. Regitz-Zagrosek V. Sex and gender differences in health. Science & Society Series on Sex and Science. EMBO Rep. 2012;13(7):596-603.
151. Altevogt BM, Wizemann T, Pankevich DE. Sex differences and implications for translational neuroscience research: workshop summary: National Academies Press; 2011.
152. Golden LC, Voskuhl R. The importance of studying sex differences in disease: The example of multiple sclerosis. J Neurosci Res. 2017;95(1-2):633-43.
153. Ober C, Loisel DA, Gilad Y. Sex-specific genetic architecture of human disease. Nat Rev Genet. 2008;9(12):911-22.
154. Genetic NY-M-ACf, Services NS. Understanding genetics: a New York, mid-Atlantic guide for patients and health professionals: Lulu. com; 2009.
155. center Gsl. Learn.Genetics: Genetic Risk 2019 [Available from: https://learn.genetics.utah.edu/content/history/geneticrisk/.
156. Alderton GK. Common disease pathogenesis pathways. Nature Reviews Genetics. 2010;11:386.
157. Imhof BA, Aurrand-Lions M. Angiogenesis and inflammation face off. Nat Med. 2006;12(2):171-2.
158. Ribatti D, Crivellato E. Immune cells and angiogenesis. J Cell Mol Med. 2009;13(9a):2822-33.
159. Medzhitov R. Inflammation 2010: new adventures of an old flame. Cell. 2010;140(6):771-6.
160. Chen L, Deng H, Cui H, Fang J, Zuo Z, Deng J, et al. Inflammatory responses and inflammation-associated diseases in organs. Oncotarget. 2018;9(6):7204-18.
161. Medzhitov R. Origin and physiological roles of inflammation. Nature. 2008;454(7203):428-35.
162. Scrivo R, Vasile M, Bartosiewicz I, Valesini G. Inflammation as “common soil” of the multifactorial diseases. Autoimmun Rev. 2011;10(7):369-74.
163. Prasad S, Sung B, Aggarwal BB. Age-associated chronic diseases require age-old medicine: role of chronic inflammation. Prev Med. 2012;54 Suppl(Suppl):S29-37.
164. Zhang JM, An J. Cytokines, inflammation, and pain. Int Anesthesiol Clin. 2007;45(2):27-37.
165. Feghali CA, Wright TM. Cytokines in acute and chronic inflammation. Front Biosci. 1997;2(1):d12-d26.
166. Genua M, Rutella S, Correale C, Danese S. The triggering receptor expressed on myeloid cells (TREM) in inflammatory bowel disease pathogenesis. Journal of Translational Medicine. 2014;12(1):293.
167. Ford JW, McVicar DW. TREM and TREM-like receptors in inflammation and disease. Current opinion in immunology. 2009;21(1):38-46.
168. Arts RJ, Joosten LA, van der Meer JW, Netea MG. TREM-1: intracellular signaling pathways and interaction with pattern recognition receptors. Journal of leukocyte biology. 2013;93(2):209-15.
169. Reitz C, Mayeux R. TREM2 and neurodegenerative disease. The New England journal of medicine. 2013;369(16):1564-5.
170. Kober DL, Brett TJ. TREM2-ligand interactions in health and disease. Journal of molecular biology. 2017;429(11):1607-29.
171. Carmeliet P. Mechanisms of angiogenesis and arteriogenesis. Nat Med. 2000;6(4):389-95.
172. Risau W. Mechanisms of angiogenesis. Nature. 1997;386(6626):671.

173. Hofmann M, Heineke J. The Impact of Endothelial Transcription Factors in Sprouting Angiogenesis. Tumor Angiogenesis: A Key Target for Cancer Therapy. 2018:1-18.
174. Adair T, Montani J. Angiogenesis. 2010. Morgan & Claypool Life Sciences: San Rafael, CA. 2010.
175. Dehghanian F, Hojati Z, Kay M. New Insights into VEGF-A Alternative Splicing: Key Regulatory Switching in the Pathological Process. Avicenna J Med Biotechnol. 2014;6(4):192-9.
176. Plate KH, Breier G, Risau W. Molecular mechanisms of developmental and tumor angiogenesis. Brain Pathol. 1994;4(3):207-18.
177. Ellis LM, Hicklin DJ. VEGF-targeted therapy: mechanisms of anti-tumour activity. Nature reviews cancer. 2008;8(8):579.
178. Ferrara N, Kerbel RS. Angiogenesis as a therapeutic target. Nature. 2005;438(7070):967.
179. Sakurai T, Kudo M. Signaling pathways governing tumor angiogenesis. Oncology. 2011;81 Suppl 1:24-9.
180. Group. BDW. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69:89-95.
181. Group F-NBW. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring (MD): Food and Drug Administration (US); Bethesda (MD): National Institutes of Health (US). 2016.
182. Burke HB. Predicting Clinical Outcomes Using Molecular Biomarkers. Biomark Cancer. 2016;8:89-99.
183. Selleck MJ, Senthil M, Wall NR. Making Meaningful Clinical Use of Biomarkers. Biomark Insights. 2017;12:1177271917715236.
184. Novelli G, Ciccacci C, Borgiani P, Papaluca Amati M, Abadie E. Genetic tests and genomic biomarkers: regulation, qualification and validation. Clin Cases Miner Bone Metab. 2008;5(2):149-54.
185. Ferrara N, Gerber HP, LeCouter J. The biology of VEGF and its receptors. Nat Med. 2003;9(6):669-76.
186. Nör JE, Christensen J, Mooney DJ, Polverini PJ. Vascular endothelial growth factor (VEGF)-mediated angiogenesis is associated with enhanced endothelial cell survival and induction of Bcl-2 expression. Am J Pathol. 1999;154(2):375-84.
187. Ferrara N, Hillan KJ, Novotny W. Bevacizumab (Avastin), a humanized anti-VEGF monoclonal antibody for cancer therapy. Biochem Biophys Res Commun. 2005;333(2):328-35.
188. Woolard J, Bevan HS, Harper SJ, Bates DO. Molecular diversity of VEGF-A as a regulator of its biological activity. Microcirculation. 2009;16(7):572-92.
189. Nowak DG, Amin EM, Rennel ES, Hoareau-Aveilla C, Gammons M, Damodoran G, et al. Regulation of vascular endothelial growth factor (VEGF) splicing from pro-angiogenic to anti-angiogenic isoforms: a novel therapeutic strategy for angiogenesis. J Biol Chem. 2010;285(8):5532-40.
190. Harper SJ, Bates DO. VEGF-A splicing: the key to anti-angiogenic therapeutics? Nat Rev Cancer. 2008;8(11):880-7.
191. Meadows KL, Hurwitz HI. Anti-VEGF therapies in the clinic. Cold Spring Harb Perspect Med. 2012;2(10).
192. Zhang W, Shen Z, Luo H, Hu X, Zheng L, Zhu X. The Benefits and Side Effects of Bevacizumab for the Treatment of Recurrent Ovarian Cancer. Curr Drug Targets. 2017;18(10):1125-31.
193. Riondino S, Monte GD, Fratangeli F, Guadagni F, Roselli M, Ferroni P. Anti-angiogenic drugs, vascular toxicity and thromboembolism in solid cancer. Cardiovasc Hematol Agents Med Chem. 2017.
194. Ladomery MR, Harper SJ, Bates DO. Alternative splicing in angiogenesis: the vascular endothelial growth factor paradigm. Cancer Lett. 2007;249(2):133-42.
195. Nowak DG, Woolard J, Amin EM, Konopatskaya O, Saleem MA, Churchill AJ, et al. Expression of pro- and anti-angiogenic isoforms of VEGF is differentially regulated by splicing and growth factors. J Cell Sci. 2008;121(Pt 20):3487-95.
196. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, et al. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A. 1988;85(18):6622-6.

197. O'Sullivan RJ, Karlseder J. Telomeres: protecting chromosomes against genome instability. Nat Rev Mol Cell Biol. 2010;11(3):171-81.
198. Lindsey J, McGill NI, Lindsey LA, Green DK, Cooke HJ. In vivo loss of telomeric repeats with age in humans. Mutat Res. 1991;256(1):45-8.
199. Fagagna FdAd, Reaper PM, Clay-Farrace L, Fiegler H, Carr P, von Zglinicki T, et al. A DNA damage checkpoint response in telomere-initiated senescence. Nature. 2003;426:194.
200. Cong Y-S, Wright WE, Shay JW. Human Telomerase and Its Regulation. Microbiol Mol Biol Rev. 2002;66:407-25.
201. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PdL, et al. Specific association of human telomerase activity with immortal cells and cancer. Science. 1994;266(5193):2011-5.
202. Njajou OT, Cawthon RM, Damcott CM, Wu S-H, Ott S, Garant MJ, et al. Telomere length is paternally inherited and is associated with parental lifespan. Proc Natl Acad Sci U S A. 2007;104:12135-9.
203. Vasa-Nicotera M, Brouilette S, Mangino M, Thompson JR, Braund P, Clemitson J-R, et al. Mapping of a major locus that determines telomere length in humans. Am J Hum Genet. 2005;76:147-51.
204. Andrew T, Aviv A, Falchi M, Surdulescu GL, Gardner JP, Lu X, et al. Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. Am J Hum Genet. 2006;78:480-6.
205. Atzmon G, Cho M, Cawthon RM, Budagov T, Katz M, Yang X, et al. Evolution in health and medicine Sackler colloquium: Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. Proc Natl Acad Sci U S A. 2010;107 Suppl 1:1710-7.
206. Mangino M, Richards JB, Soranzo N, Zhai G, Aviv A, Valdes AM, et al. A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length. J Med Genet. 2009;46:451-4.
207. Codd V, Mangino M, van der Harst P, Braund PS, Kaiser M, Beveridge AJ, et al. Variants near TERC are associated with mean telomere length. Nat Genet. 2010;42:197-9.
208. Levy D, Neuhausen SL, Hunt SC, Kimura M, Hwang S-J, Chen W, et al. Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. Proc Natl Acad Sci U S A. 2010;107:9293-8.
209. Gu J, Chen M, Shete S, Amos CI, Kamat A, Ye Y, et al. A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. Cancer Prev Res (Phila). 2011;4:514-21.
210. Prescott J, Kraft P, Chasman DI, Savage SA, Mirabello L, Berndt SI, et al. Genome-wide association study of relative telomere length. PLoS One. 2011;6:e19635.
211. Mangino M, Hwang S-J, Spector TD, Hunt SC, Kimura M, Fitzpatrick AL, et al. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. Hum Mol Genet. 2012;21:5385-94.
212. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet. 2013;45:422-7, 7e1-2.
213. Pooley KA, Bojesen SE, Weischer M, Nielsen SF, Thompson D, Amin Al Olama A, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. Hum Mol Genet. 2013;22:5056-64.
214. Liu Y, Cao L, Li Z, Zhou D, Liu W, Shen Q, et al. A genome-wide association study identifies a locus on TERT for mean telomere length in Han Chinese. PLoS One. 2014;9:e85043.
215. Saxena R, Bjonnes A, Prescott J, Dib P, Natt P, Lane J, et al. Genome-wide association study identifies variants in casein kinase II (CSNK2A2) to be associated with leukocyte telomere length in a Punjabi Sikh diabetic cohort. Circ Cardiovasc Genet. 2014;7:287-95.
216. Walsh KM, Codd V, Smirnov IV, Rice T, Decker PA, Hansen HM, et al. Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk. Nat Genet. 2014;46:731-5.

217. Rietzschel ER, Bekaert S, De Meyer T. Telomeres and Atherosclerosis: The Attrition of an Attractive Hypothesis. J Am Coll Cardiol. 2016;67(21):2477-9.
218. Armanios M, Blackburn EH. The telomere syndromes. Nat Rev Genet. 2012;13:693-704.
219. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet. 2005;6:611-22.
220. Calado RT, Young NS. Telomere Diseases. The New England Journal of Medicine. 2009;361:2353-65.
221. Slattery ML, Curtin K, Baumgartner R, Sweeney C, Byers T, Giuliano AR, et al. IL6, aspirin, nonsteroidal anti-inflammatory drugs, and breast cancer risk in women living in the southwestern United States. Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology. 2007;16:747-55.
222. O'Donovan A, Pantell MS, Puterman E, Dhabhar FS, Blackburn EH, Yaffe K, et al. Cumulative inflammatory load is associated with short leukocyte telomere length in the Health, Aging and Body Composition Study. PLoS One. 2011;6(5):e19687.
223. Glei DA, Goldman N, Weinstein M, Risques RA. Shorter Ends, Faster End? Leukocyte Telomere Length and Mortality Among Older Taiwanese. The Journals of Gerontology Series A, Biological Sciences and Medical Sciences. 2015;70:1490-8.
224. Maida Y, Kyo S, Kanaya T, Wang Z, Yatabe N, Tanaka M, et al. Direct activation of telomerase by EGF through Ets-mediated transactivation of TERT via MAP kinase signaling pathway. Oncogene. 2002;21:4071-9.
225. Council E. Council conclusions on personalised medicine for patients. official Journal of the European Union. 2015;421.
226. Redekop WK, Mladsi D. The faces of personalized medicine: a framework for understanding its meaning and scope. Value Health. 2013;16(6 Suppl):S4-9.
227. Goodsaid F, Frueh F, Burczynski ME. Personalized Medicine. Drug Discovery and Evaluation: Methods in Clinical Pharmacology. 2018:1-14.
228. Kim D-H, Kim Y-S, Son N-I, Kang C-K, Kim A-R. Recent omics technologies and their emerging applications for personalised medicine. IET Syst Biol. 2017;11(3):87-98.
229. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype–phenotype interactions. Nature Reviews Genetics. 2015;16:85.
230. Thibaut F. From basic research to personalized medicine. Dialogues Clin Neurosci. 2016;18(3):231-2.
231. Nussbaum RL, McInnes RR, Willard HF. Thompson & Thompson genetics in medicine e-book: Elsevier Health Sciences; 2015.
232. Camastra F, Di Taranto MD, Staiano A. Statistical and computational methods for genetic diseases: An overview. Comput Math Methods Med. 2015;2015.
233. Witte JS. Genome-wide association studies and beyond. Annu Rev Public Health. 2010;31:9-20 4 p following
234. Zheutlin AB, Ross DA. Polygenic risk scores: What are they good for? Biological psychiatry. 2018;83(11):e51-e3.
235. Khera AV, Chaffin M, Aragam KG, Emdin CA, Klarin D, Haas ME, et al. Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. bioRxiv. 2017:218388.
236. Jameson JL, Longo DL. Precision medicine—personalized, problematic, and promising. Obstet Gynecol Surv. 2015;70(10):612-4.
237. Cardon LR, Bell JI. Association study designs for complex diseases. Nature Reviews Genetics. 2001;2(2):91-9.
238. Balding DJ. A tutorial on statistical methods for population association studies. Nature Reviews Genetics. 2006;7(10):781-91.
239. Miles CM, Wayne M. Quantitative Trait Locus (QTL) Analysis. Nature Education 2008;1(1).
240. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol. 2019;2:9.

241. Dorak MT. Genetic association studies: background, conduct, analysis, interpretation: Garland Science; 2016.
242. Heggenhougen K. International encyclopedia of public health: Elsevier Inc.; 2008.
243. Hoffman JI. Biostatistics for medical and biomedical practitioners: Academic Press; 2015.
244. Gupta PK, Kulwal PL, Jaiswal V. Chapter Two - Association Mapping in Crop Plants: Opportunities and Challenges. In: Friedmann T, Dunlap JC, Goodwin SF, editors. Advances in Genetics. 85: Academic Press; 2014. p. 109-47.
245. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. Am J Hum Genet. 2018;102(5):717-30.
246. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J. 2014;36(9):539-50.
247. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nature Reviews Cardiology. 2017;14:577.
248. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
249. Sekula P, Fabiola Del Greco M, Pattaro C, Köttgen A. Mendelian randomization as an approach to assess causality using observational data. J Am Soc Nephrol. 2016;27(11):3253-65.
250. Thanassoulis G, O'Donnell CJ, Raby BA, Sokol HN. Mendelian randomization. Topic 13514 Version 1202018.
251. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell. 2007;128(4):635-8.
252. Bemer M. Unraveling the Complex Epigenetic Mechanisms that Regulate Gene Activity. Methods in molecular biology (Clifton, NJ). 2018;1675:205-31.
253. Weber-Stadlbauer U. Epigenetic and transgenerational mechanisms in infection-mediated neurodevelopmental disorders. Transl Psychiatry. 2017;7(5):e1113.
254. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429(6990):457-63.
255. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010;28(10):1057-68.
256. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, et al. Editing DNA methylation in the mammalian genome. Cell. 2016;167(1):233-47. e17.
257. Schübeler D. Function and information content of DNA methylation. Nature. 2015;517(7534):321.
258. Robertson KD. DNA methylation and human disease. Nature reviews Genetics. 2005;6(8):597-610.
259. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes & cancer. 2011;2(6):607-17.
260. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472-9.
261. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool DNA testing for colorectal-cancer screening. The New England journal of medicine. 2014;370(14):1287-97.
262. Chang CC, Wang HC. The feasibility of detecting endometrial and ovarian cancer using DNA methylation biomarkers in cervical scrapings. 2018;29(1):e17.
263. Zhang W, Song M, Qu J, Liu GH. Epigenetic Modifications in Cardiovascular Aging and Diseases. Circulation research. 2018;123(7):773-86.
264. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12(8):529-41.
265. Dedeurwaerder S, Defrance M, Bizet M, Calonne E, Bontempi G, Fuks F. A comprehensive overview of Infinium HumanMethylation450 data processing. Brief Bioinform. 2014;15(6):929-41.
266. Alberts B, Bray D, Hopkin K, Johnson AD, Lewis J, Raff M, et al. Essential cell biology: Garland Science; 2013.

267. Leicester Uo. Gene Expression and Regulation 2019 [Available from: https://www2.le.ac.uk/projects/vgec/highereducation/topics/geneexpression-regulation.
268. Segundo-Val IS, Sanz-Lozano CS. Introduction to the Gene Expression Analysis. Methods Mol Biol. 2016;1434:29-43.
269. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009;10:57.
270. Tayebi L, Moharamzadeh K. Biomaterials for Oral and Dental Tissue Engineering: Woodhead Publishing; 2017.
271. Vidhyasekaran P. Handbook of molecular technologies in crop disease management: CRC Press; 2013.
272. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. PLoS Comput Biol. 2017;13(5):e1005457.
273. Huber D, von Voithenberg LV, Kaigala G. Fluorescence in situ hybridization (FISH): history, limitations and what to expect from micro-scale FISH? Micro and Nano Engineering. 2018.
274. Loewenstein PM, Song CZ, Green M. The use of in vitro transcription to probe regulatory functions of viral protein domains. Methods Mol Med. 2007;131:15-31.
275. George L, Indig FE, Abdelmohsen K, Gorospe M. Intracellular RNA-tracking methods. Royal Society Open Biology. 2018;8(10):180104.
276. Deverka P, Carlson RJ. Personalized Medicine: Current and Future Perspectives. 2019.
277. Vivot M. The 5 Challenges of Personalised Medicine: European Commission; 2016 [Available from: https://ec.europa.eu/digital-single-market/en/news/5-challenges-personalised-medicine.
278. Siest G, Visvikis S, Herbeth B, Gueguen R, Vincent-Viry M, Sass C, et al. Objectives, design and recruitment of a familial and longitudinal cohort for studying gene-environment interactions in the field of cardiovascular risk: the Stanislas cohort. Clin Chem Lab Med. 1998;36(1):35-42.
279. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med. 2008;46(6):733-47.
280. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet. 2011;7(2):e1002003.
281. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988;16(3):1215.
282. Glastonbury CA, Viñuela A, Buil A, Halldorsson GH, Thorleifsson G, Helgason H, et al. Adiposity-Dependent Regulatory Effects on Multi-tissue Transcriptomes. Am J Hum Genet. 2016;99(3):567-79.
283. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288-95.
284. He C, Holme J, Anthony J. SNP genotyping: the KASP assay. Methods Mol Biol. 2014;1145:75-86.
285. Boyum A. Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. Scand J Clin Lab Invest Suppl. 1968;97:77-89.
286. Azimi-Nezhad M, Stathopoulou MG, Bonnefond A, Rancier M, Saleh A, Lamont J, et al. Associations of vascular endothelial growth factor (VEGF) with adhesion and inflammation molecules in a healthy population. Cytokine. 2013;61(2):602-7.
287. Marteau JB, Mohr S, Pfister M, Visvikis-Siest S. Collection and storage of human blood cells for mRNA expression profiling: a 15-month stability study. Clin Chem. 2005;51(7):1250-2.
288. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18(6):499-502.

289. Schulze MB, Rimm EB, Shai I, Rifai N, Hu FB. Relationship between adiponectin and glycemic control, blood lipids, and inflammatory markers in men with type 2 diabetes. Diabetes Care. 2004;27(7):1680-7.
290. Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, et al. Epigenome-wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study. Circulation. 2014;130(7):565-72.
291. Kopelman PG. Obesity as a medical problem. Nature. 2000;404(6778):635-43.
292. Lee CMY, Huxley RR, Wildman RP, Woodward M. Indices of abdominal obesity are better discriminators of cardiovascular risk factors than BMI: a meta-analysis. J Clin Epidemiol. 2008;61(7):646-53.
293. Janssen I, Katzmarzyk PT, Ross R. Waist circumference and not body mass index explains obesity-related health risk. Am J Clin Nutr. 2004;79(3):379-84.
294. Moustafa JSE-S, Froguel P. From obesity genetics to the future of personalized obesity therapy. Nature Reviews Endocrinology. 2013;9(7):402.
295. van Dijk SJ, Molloy PL, Varinli H, Morrison JL, Muhlhausler BS. Epigenetics and human obesity. Int J Obes (Lond). 2015;39(1):85-97.
296. Byrne AM, Bouchier-Hayes DJ, Harmey JH. Angiogenic and cell survival functions of vascular endothelial growth factor (VEGF). J Cell Mol Med. 2005;9(4):777-94.
297. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.
298. Muller YA, Li B, Christinger HW, Wells JA, Cunningham BC, de Vos AM. Vascular endothelial growth factor: crystal structure and functional mapping of the kinase domain receptor binding site. Proc Natl Acad Sci U S A. 1997;94(14):7192-7.
299. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? The lancet. 2001;357(9255):539-45.
300. Muir T. Peripheral Blood Mononuclear Cells: A Brief Review: stemexpress; 2019 [Available from: https://www.stemexpress.com/blogs/peripheral-blood-mononuclear-cells/.
301. Corkum CP, Ings DP, Burgess C, Karwowska S, Kroll W, Michalak TI. Immune cell subsets and their gene expression profiles from human PBMC isolated by Vacutainer Cell Preparation Tube (CPT(™)) and standard density gradient. BMC Immunol. 2015;16.
302. Colonna M, Wang Y. TREM2 variants: new keys to decipher Alzheimer disease pathogenesis. Nat Rev Neurosci. 2016;17(4):201-7.
303. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 Variants in Alzheimer's Disease. N Engl J Med. 2013;368(2):117-27.
304. Boesten LS, Zadelaar AS, van Nieuwkoop A, Gijbels MJ, de Winther MP, Havekes LM, et al. Tumor necrosis factor-alpha promotes atherosclerotic lesion progression in APOE*3-Leiden transgenic mice. Cardiovasc Res. 2005;66(1):179-85.
305. Gariepy J, Simon A, Massonneau M, Linhart A, Levenson J. Wall thickening of carotid and femoral arteries in male subjects with isolated hypercholesterolemia. PCVMETRA Group. Prevention Cardio-Vasculaire en Medecine du Travail. Atherosclerosis. 1995;113(2):141-51.
306. Zhao F, Song M, Wang Y, Wang W. Genetic model. J Cell Mol Med. 2016;20(4):765.
307. Lee SH, Griffiths N, Shao L. Can telomere length be used as a biomarker for cardiovascular diseases?: insights from a large clinical study. J Physiol. 2017.
308. Lee M, Napier CE, Yang SF, Arthur JW, Reddel RR, Pickett HA. Comparative analysis of whole genome sequencing-based telomere length measurement techniques. Methods (San Diego, Calif). 2017;114:4-15.
309. Haycock PC, Burgess S, Nounu A, Zheng J, Okoli GN, Bowden J, et al. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. JAMA Oncol. 2017;3(5):636-51.
310. McCloskey K, Ponsonby AL, Collier F, Allen K, Tang MLK, Carlin JB, et al. The association between higher maternal pre-pregnancy body mass index and increased birth weight, adiposity and inflammation in the newborn. Pediatr Obes. 2018;13(1):46-53.

311. Salihu HM, Pradhan A, King L, Paothong A, Nwoga C, Marty PJ, et al. Impact of intrauterine tobacco exposure on fetal telomere length. Am J Obstet Gynecol. 2015;212(2):205.e1-8.
312. Send TS, Gilles M, Codd V, Wolf I, Bardtke S, Streit F, et al. Telomere Length in Newborns is Related to Maternal Stress During Pregnancy. Neuropsychopharmacology. 2017;42(12):2407-13.
313. Shalev I, Entringer S, Wadhwa PD, Wolkowitz OM, Puterman E, Lin J, et al. Stress and telomere biology: a lifespan perspective. Psychoneuroendocrinology. 2013;38(9):1835-42.
314. Martens DS, Plusquin M, Gyselaers W, De Vivo I, Nawrot TS. Maternal pre-pregnancy body mass index and newborn telomere length. BMC Med. 2016;14(1):148.
315. Friedrich U, Schwab M, Griese EU, Fritz P, Klotz U. Telomeres in neonates: new insights in fetal hematopoiesis. Pediatr Res. 2001;49(2):252-6.
316. Okuda K, Bardeguez A, Gardner JP, Rodriguez P, Ganesh V, Kimura M, et al. Telomere length in the newborn. Pediatr Res. 2002;52(3):377-81.
317. Stathopoulou MG, Petrelis AM, Buxton JL, Froguel P, Blakemore AI, Visvikis-Siest S. Genetic determinants of leucocyte telomere length in children: a neglected and challenging field. Paediatr Perinat Epidemiol. 2015;29(2):146-50.
318. Zeiger AM, White MJ, Eng C, Oh SS, Witonsky J, Goddard PC, et al. Genetic Determinants of Telomere Length in African American Youth. Sci Rep. 2018;8(1):13265.
319. Blackburn EH. Structure and function of telomeres. Nature. 1991;350:569-73.
320. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. N Engl J Med. 2005;352(16):1685-95.
321. Carmeliet P. VEGF as a key mediator of angiogenesis in cancer. Oncology. 2005;69(Suppl. 3):4-10.
322. Testa U, Pannitteri G, Condorelli GL. Vascular endothelial growth factors in cardiovascular medicine. J Cardiovasc Med (Hagerstown). 2008;9(12):1190-221.
323. Houben JM, Moonen HJ, van Schooten FJ, Hageman GJ. Telomere length assessment: biomarker of chronic oxidative stress? Free Radic Biol Med. 2008;44(3):235-46.
324. Honig LS, Kang MS, Cheng R, Eckfeldt JH, Thyagarajan B, Leiendecker-Foster C, et al. Heritability of telomere length in a study of long-lived families. Neurobiol Aging. 2015;36(10):2785-90.
325. Hetland ML, Christensen IJ, Lottenburger T, Johansen JS, Svendsen MN, Horslev-Petersen K, et al. Circulating VEGF as a biological marker in patients with rheumatoid arthritis? Preanalytical and biological variability in healthy persons and in patients. Dis Markers. 2008;24(1):1-10.
326. Toupance S, Villemonais D, Germain D, Gegout-Petit A, Albuisson E, Benetos A. The individual's signature of telomere length distribution. Sci Rep. 2019;9(1):685.
327. Poonpet T, Saetan N, Tanavalee A, Wilairatana V, Yuktanandana P, Honsawek S. Association between leukocyte telomere length and angiogenic cytokines in knee osteoarthritis. Int J Rheum Dis. 2018;21(1):118-25.
328. Chahine MN, Toupance S, El-Hakim S, Labat C, Gautier S, Moussallem T, et al. Telomere length and age-dependent telomere attrition: the blood-and-muscle model (1). Can J Physiol Pharmacol. 2019;97(4):328-34.
329. Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, et al. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. Nat Protoc. 2010;5(9):1596-607.
330. Kocsis I, Gyorffy B, Nemeth E, Vasarhelyi B. Examination of Hardy-Weinberg equilibrium in papers of Kidney International: an underused tool. Kidney Int. 2004;65(5):1956-8.
331. Schick UM, Jain D, Hodonsky CJ, Morrison JV, Davis JP, Brown L, et al. Genome-wide Association Study of Platelet Count Identifies Ancestry-Specific Loci in Hispanic/Latino Americans. Am J Hum Genet. 2016;98(2):229-42.
332. Wong JY, De Vivo I, Lin X, Fang SC, Christiani DC. The relationship between inflammatory biomarkers and telomere length in an occupational prospective cohort study. PLoS One. 2014;9(1):e87348.

333. Blann AD, Belgore FM, Constans J, Conri C, Lip GY. Plasma vascular endothelial growth factor and its receptor Flt-1 in patients with hyperlipidemia and atherosclerosis and the effects of fluvastatin or fenofibrate. The American journal of cardiology. 2001;87(10):1160-3.
334. BLANN AD, BELGORE FM, McCOLLUM CN, SILVERMAN S, LIP PL, Gregory Y. Vascular endothelial growth factor and its receptor, Flt-1, in the plasma of patients with coronary or peripheral atherosclerosis, or Type II diabetes. Clin Sci. 2002;102(2):187-94.
335. Alber HF, Frick M, Dulak J, Dörler J, Zwick RH, Dichtl W, et al. Vascular endothelial growth factor (VEGF) plasma concentrations in coronary artery disease. Heart. 2005;91(3):365-6.
336. Visvikis-Siest S, Gorenjak V, Stathopoulou MG. Personalised Medicine: The Odyssey from Hope to Practice. J Pers Med. 2018;8(4).
337. Souverijn JH. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med. 2014;371(2):187.
338. Carling D. The AMP-activated protein kinase cascade--a unifying system for energy control. Trends Biochem Sci. 2004;29(1):18-24.
339. Banankhah P, Fishbein GA, Dota A, Ardehali R. Cardiac manifestations of PRKAG2 mutation. BMC Med Genet. 2018;19(1):1.
340. Ha AC, Renaud JM, Dekemp RA, Thorn S, Dasilva J, Garrard L, et al. In vivo assessment of myocardial glucose uptake by positron emission tomography in adults with the PRKAG2 cardiac syndrome. Circ Cardiovasc Imaging. 2009;2(6):485-91.
341. Arad M, Benson DW, Perez-Atayde AR, McKenna WJ, Sparks EA, Kanter RJ, et al. Constitutively active AMP kinase mutations cause glycogen storage disease mimicking hypertrophic cardiomyopathy. J Clin Invest. 2002;109(3):357-62.
342. Mao B, Niehrs C. Kremen2 modulates Dickkopf2 activity during Wnt/LRP6 signaling. Gene. 2003;302(1-2):179-83.
343. Komiya Y, Habas R. Wnt signal transduction pathways. Organogenesis. 2008;4(2):68-75.
344. Foulquier S, Daskalopoulos EP, Lluri G, Hermans KC, Deb A, Blankesteijn WM. WNT signaling in cardiac and vascular disease. Pharmacol Rev. 2018;70(1):68-141.
345. De Ferrari GV, Inestrosa NC. Wnt signaling function in Alzheimer's disease. Brain Res Rev. 2000;33(1):1-12.
346. Tsai F-Y, Keller G, Kuo FC, Weiss M, Chen J, Rosenblatt M, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature. 1994;371(6494):221.
347. Liu N, Ding D, Hao W, Yang F, Wu X, Wang M, et al. hTERT promotes tumor angiogenesis by activating VEGF via interactions with the Sp1 transcription factor. Nucleic Acids Res. 2016;44(18):8693-703.
348. Koboldt D. Functional Validation of Genomic Discoveries 2013 [Available from: http://massgenomics.org/2013/07/functional-validation-genomic-discoveries.html.

OTHER PUBLICATIONS

Journal of *Personalized Medicine*

MDPI

Conference Report

The 9th Santorini Conference: Systems Medicine, Personalised Health and Therapy. "The Odyssey from Hope to Practice", Santorini, Greece, 30 September–3 October 2018

Sophie Visvikis-Siest [1,*], Vesna Gorenjak [1], Maria G. Stathopoulou [1], Alexandros M. Petrelis [1], Georges Weryha [1], Christine Masson [1], Brigitte Hiegel [1], Satish Kumar [1], Robert Barouki [2,3], Eric Boerwinkle [4], Georges Dagher [5], Panagiotis Deloukas [6], Federico Innocenti [7], John Lamont [8], Michael Marschler [9], Heiko Meyer [10], Urs A. Meyer [11], Charity Nofziger [12], Markus Paulmichl [13,14], Cora Vacher [15] and Lynn Webster [16]

1 Université de Lorraine, Inserm, IGE-PCV, 54000 Nancy, France; vesna.gorenjak@inserm.fr (V.G.); maria.stathopoulou@inserm.fr (M.G.S.); apetrelis@live.com (A.M.P.); g.weryha@chru-nancy.fr (G.W.); christine.masson@univ-lorraine.fr (C.M.); brigitte.hiegel@univ-lorraine.fr (B.H.); satish.kumar@univ-lorraine.fr (S.K.)
2 INSERM UMR-S 1124, Université Paris Descartes, Université Sorbonne Paris Cité, 75270 Paris, France; robert.barouki@parisdescartes.fr
3 Service de Biochimie Métabolomique et Protéomique, Hôpital Universitaire Necker Enfants Malades, AP-HP, 75015 Paris, France
4 Department of Epidemiology, Human Genetics and Environmental Sciences, UTHealth School of Public Health, Houston, TX 77030, USA; Eric.Boerwinkle@uth.tmc.edu
5 Inserm UMR-S 1124, Paris-Descartes University, 75270 Paris, France; georges.dagher@inserm.fr
6 Department of Clinical Pharmacology, Queen Mary University of London, London E1 4NS, UK; p.deloukas@qmul.ac.uk
7 Pharmacogenomics and Individualized Therapy, University of North Carolina Chapel Hill, Chapel Hill, NC 27599-2200, USA; innocent@email.unc.edu
8 Randox Laboratories Ltd., Crumlin, County Antrim BT29 4XL, UK; john.lamont@randox.com
9 PRAHealthSciences, 68165 Mannheim, Germany; marschlermichael@prahs.com
10 Agena Bioscience GmbH, 22761 Hamburg, Germany; heiko.meyer@agenabio.com
11 Biozentrum, University of Basel, 4056 Basel, Switzerland; urs-a.meyer@unibas.ch
12 PharmGenetix Gmbh, 1090 Vienna, Austria; charity.nofziger@pharmgenetix.com
13 Center for Health & Bioresources, Austrian Institute of Technology, 1210 Vienna, Austria; paulmichl@me.com
14 NESMOS Department, University of Rome Sapienza, 00185 Rome, Italy
15 Illumina, 81669 Munich, Germany; cvacher@illumina.com
16 PRAHealthSciences, Salt Lake City, UT 84124, USA; WebsterLynn@prahs.com
* Correspondence: sophie.visvikis-siest@inserm.fr; Tel.: +3-360-760-2569

Received: 3 December 2018; Accepted: 5 December 2018; Published: 12 December 2018

Abstract: The 9th traditional biannual conference on Systems Medicine, Personalised Health & Therapy—"The Odyssey from Hope to Practice", inspired by the Greek mythology, was a call to search for practical solutions in cardio-metabolic diseases and cancer, to resolve and overcome the obstacles in modern medicine by creating more interactions among disciplines, as well as between academic and industrial research, directed towards an effective 'roadmap' for personalised health and therapy. The 9th Santorini Conference, under the Presidency of Sofia Siest, the director of the INSERM U1122; IGE-PCV (www.u1122.inserm.fr), University of Lorraine, France, offered a rich and innovative scientific program. It gathered 34 worldwide distinguished speakers, who shared their passion for personalised medicine with 160 attendees in nine specific sessions on the following topics: First day: The Odyssey from hope to practice: Personalised medicine—landmarks and

challenges Second day: Diseases to therapeutics—genotype to phenotype an "-OMICS" approach: focus on personalised therapy and precision medicine Third day: Gene-environment interactions and pharmacovigilance: a pharmacogenetics approach for deciphering disease "bench to clinic to reality" Fourth day: Pharmacogenomics to drug discovery: a big data approach and focus on clinical data and clinical practice. In this article we present the topics shared among the participants of the conference and we highlight the key messages.

Keywords: systems medicine; personalised medicine; pharmacogenomics; clinical trials; cardio-metabolic diseases; cancer; genetic screening; "-OMICs" biomarkers; santorini conference

1. The 9th Santorini Conference

Figure 1. Participants of The 9th Santorini conference in Petros M. Nomikos Conference Centre. (Photo by George Mindrinos).

The 9th traditional biannual conference on Systems Medicine, Personalised Health and Therapy, under the auspices of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM), the Hellenic Society for Basic and Clinical Pharmacology (EEI), the Hellenic Society of Pharmacogenomics and personalised Diagnosis and Therapy (EEPHARM), and the European Society of Pharmacogenomics and Personalised Therapy (ESPT), took place in Santorini, Greece, on 30 September 2018.

It was organised with the help of different industrial companies: Randox (Crumlin, UK) (loyal Gold sponsor)—Siemens Healthineers (Erlangen, Germany), Illumina (Munich, Germany), Agena Bioscience (Hamburg, Germany), DiaSys (Connecticut, États-Unis) (Silver sponsors)—the Banque Populaire (Paris, France), Mastiha Growers Association (Chios, Greece), PharmGenetix (Vienna, Austria), and HMG systems Engineering (Fuerth, Germany) (Other sponsors).

The 9th Santorini Conference, under the Presidency of Sofia Siest, the director of the INSERM U1122; IGE-PCV (www.u1122.inserm.fr), University of Lorraine, France, offered a rich and innovative scientific program. It gathered 34 worldwide distinguished speakers, who shared their passion for personalised medicine with 160 attendees (Figure 1) in nine specific sessions on the following topics:

- First day: The Odyssey from hope to practice: Personalised medicine—landmarks and challenges
- Second day: Diseases to therapeutics—genotype to phenotype an "-OMICS" approach: Focus on personalised therapy and precision medicine
- Third day: Gene–environment interactions and pharmacovigilance: A pharmacogenetics approach for deciphering disease "bench to clinic to reality"
- Fourth day: Pharmacogenomics to drug discovery: A big data approach and focus on clinical data and clinical practice.

In this article, we present the topics shared among the participants of the conference and we highlight the key messages.

1.1. The Odyssey from Hope to Practice

The conference was officially open with the welcome by Sofia Siest, who made a brief summary on the evolution of this colloquium over the past 16 years, initiated in 2002, thus being the oldest International conference in the field of personalised Medicine medicine and pharmacogenomics and one of the most important conferences on genetic predisposition to health, diseases, and response to drugs and environment. After 16 years, the 9th Santorini Conference "Systems Medicine and Personalised Health and Therapy" aimed to reflect on the frontiers of current genomic knowledge in the fields of cardiometabolic diseases and cancer and to reveal the practical use of this knowledge in disease prevention, diagnosis, and pharmacogenomics aiming to directly impact the socioeconomic aspects of public health. Sophia Siest's presentation was inspired by the Odyssey, the Greek ancient epic poem of Homer, and she embarked with the attendees on a ship from Troy (Hope) to Ithaca (Practice), coursing through the history of the momentous events and achievements that paved the way for personalised medicine. She set sail amidst important genetic discoveries and obstacles that are slowing the full implementation of accumulated knowledge into everyday practice, beginning with the discovery of the first human genome and voyaging through the projects that contributed to the progress of personalised medicine. Her introduction was illustrated with an animated movie (available on https://www.youtube.com/watch?v=SReS8WmH2jU) presenting the journey of Odysseus on his way to Ithaca—from hope to practice—whilst reminding of great achievements that contributed enormously to the development of personalised medicine. The whole story describing the history of personalised medicine was carefully gathered in the eponymous article [1].

Following the opening words of Sofia Siest, the keynote lecture, presented by Eric Boerwinkle (Houston, TX, USA) carried on with the journey to personalised medicine and unravelled the landmarks and challenges of the field. He discussed the necessity of genome-wide association studies (GWAS) and their input in the discovery of new genes, but the inclusion of ethic variations is also necessary for taking knowledge further. He pointed out that apart from genes, their interactions with environment through time should be actively considered in genomic research. He also discussed pharmacogenomics and the importance of including genomic information in medical guidelines. The importance of technological advances was also mentioned as a lesson learned even from Homer's era. He concluded that Ithaca for personalised medicine should be based on relating clinicians and researchers and include improved sequencing to improve diagnostics and personalisation of therapy of severe and chronic diseases.

2. Conference Sessions

2.1. "-OMICs" Biomarkers Commonality in Cardiometabolic Diseases and Cancer—Can Sequencing Offer a Diagnosis?

Chairs: Panagiotis Deloukas, London, UK/Heiko Meyer, Hamburg, Germany

The session started with the morning lecture, given by Christopher B. Newgard (Durham, NC, USA), who discussed the application of metabolomics and other "-omics" tools for the understanding of mechanisms contributing to pandemic metabolic diseases of our era—diabetes, obesity, and cardiovascular disease [2]. With the practical examples from the investigation of the mechanistic and therapeutic significance of a metabolomics signature of perturbed branched chain amino acid (BCAA) catabolism in multiple cohorts of insulin resistant humans, compared to normally insulin sensitive controls [3], Dr Newgard demonstrated the potential of metabolic profiling for defining novel metabolic disease mechanisms and new therapeutic strategies.

Next, two remarkable projects that evolved the era of genomics were presented. Firstly, Mark Caulfield (London, UK) spoke about the 100,000 Genomes Project and its impact on transforming genomics in healthcare. In the frames of the project, thirteen National Health Service

(NHS) Genomic Medicine Centres of excellence across England were created in collaboration with the NHS to enable the generation of clinical data and sample flows. Moreover, in partnership with Illumina, one of the largest next generation sequencing centres in the world was created to generate the highest fidelity and most comprehensive whole genome DNA sequence produced from patients to date with high fidelity clinical data stored in a de-identified format within a multipetabyte data infrastructure.

Secondly, Peter Campbell (Hinxton, UK) from The Welcome Trust Sanger Institute introduced The International Cancer Genome project. The main aim is to provide important insights into the biology of cancer through discovering the genes that are frequently mutated in tumours, and with studying the patterns of mutations seen in cancer cells. Overall, the projects explore basic scientific questions on the role that somatic mutations play in clonal evolution, ageing, and development.

The first session of the conference was completed with the lecture of Eleftherios Diamandis (Toronto, ON, Canada) on personalised biomarkers in cancer. A slow progress in cancer biomarker discovery and general speculations of scientists, who believed it is unlikely to discover new serological biomarkers characterised by high sensitivity and specificity, encouraged researchers to propose a new way of improving the landscape of cancer biomarker research. Screening new patients for a large number of previously described biomarkers with high specificity (>90%) but low sensitivity (<30%) could identify new informative markers for clinical use, important for managing individual cancer patients. Moreover, this approach may explain the reasons for the increased value of some biomarkers that appear only in a small group of patients. These differences in expression are likely to be linked to specific genomic alterations, which could then be found with genomic sequencing [4].

2.2. Unmet Clinical Needs in the Prevention and Treatment of Cardiometabolic Diseases and Cancer-Comorbidities

Chairs: Georges Weryha, Nancy, France/John Lamont, Crumlin, County Antrim, UK

For the introduction to the new session, Manuel Rosa Garrido (Los Angeles, CA, USA) shared his findings in the epigenetics of cardiovascular diseases and comorbidities. Epigenetic mechanisms control gene expression at the individual locus scale as well as at the genomic scale via the formation of long-range regulatory interactions and the formation of chromosome territories. The presented case-control study on mice aimed to determine the impact of such mechanisms during the development of heart failure. Manuel Rosa Garrido proposed that global remodelling of chromatin after traverse aortic constriction and 11-zinc finger protein (CTCF) depletion drives heart failure. He concluded that heart failure involves conserved structural reprogramming of chromatin microenvironments with histone marks and DNA methylation, which play an important role in controlling expression.

In his speech, Abraham Aviv (New Jersey, USA) focused on the role of telomere length (TL) in the etymology of chronic diseases. He pointed out that short TLs have a protective role against cancer, while the trade-off for it might be susceptibility to cardiovascular diseases CVD, suggesting TL as a common factor of these diseases. Furthermore, associations manifest at the genomic level, as single nucleotide polymorphisms (SNPs) associated with short leucocyte telomere length (LTL) are also associated with CVD and SNPs associated with long leucocyte TL are also associated with several cancers. Therefore, he concluded that TL might play a causal role in CVD and some cancers. Finally, these findings also have considerable ramifications for the longevity of humans.

2.3. Pipeline Challenges in Precision Medicine—From Patient Sampling to a Clinical Report

Chairs: Georges Dagher, Paris, France/Charity Nofziger, Salzburg, Austria

Markus Paulmichl (Salzburg, Austria) presented the Austrian experience of a good practice in carrying patient data in the clinical report. He presented the national program of the Austrian government that aims for the increased implementation of pharmacogenomics into the daily clinical routine. He discussed the different steps for reporting pharmacogenomics data to clinicians, which can

be challenging, especially in cases of polytherapy. A schematic report warns clinicians about inappropriate medications and proposes possible alternative solutions.

The importance of standardised pre-analytical workflows that can ensure good quality samples for reliable analytical test results was recognised by EU. The result was the SPIDIA consortium that gathered academic institutions, international organisations, and private companies, in order to identify variables in the pre-analytical phase that can influence the quality of samples and should be controlled to guarantee a robust diagnosis. Uwe Oelmueller (Hilden, Germany) gave the overview on the new EU SPIDIA4P project and its potential for developing on these basic pre-analytical workflow standards that guide the laboratories workflow parameters and guarantees analytical test results. In relation to these findings, 22 standards are being drafted in the scope of SPIDIA4P. Ten of them are already published by the European Committee for Standardization as European CEN documents and now being further broadened to be published as International standards (ISO).

In relation to the previous speaker, Karl Friedrich Becker (Munich, Germany) gave the example of implementation of standardised pre-analytical CEN/TS and ISO/IS documents based on SPIDIA and SPIDIA4P projects in the current workflows, in order to provide a successful integration of proteomic studies in the clinic. He discussed the influences of sample processing, including sample collection, transport, stabilisation, and storage and analyte (e.g., protein) extraction on the final assay result, and highlighted the importance of standardisation of the entire workflow from test ordering to the report of the molecular assay, with special emphasis on the pre-analytical phase.

In addition to pre-analytical challenges in sample procession, there are other obstacles that can hinder the usage of patient samples for precision medicine. Circulating tumour cells firstly showed a potential utility as a surrogate biomarker of tumour biology via a liquid biopsy. Maria G. Daidone (Milan, Italy) presented the challenges in the use of such liquid biopsies for personalised medicine.

2.4. PGx-Regulatory Perspectives and Future Technical/Analytical Developments

Chairs: Markus Paulmichl, Salzburg, Austria/Lynn Webster, Salt Lake City, UT, USA

In personalised medicine, genomic data became an important source of information for the evaluation of efficacy and safety of medicinal products. Similar to standardisation of pre-analytical workflows to ensure the good quality of samples for reliable analysis, the use of genomic biomarkers in drug development should, in order to be of value, follow certain principles. Markus Paulmichl (Salzburg, Austria) presented the Good Pharmacogenomic Practice guideline from the pharmacogenomics working group within the European Medicines Agency (EMA), which provides recommendations for the realisation of genomic studies in relation to medical therapy.

To conclude the last daily session, Simone Vanoni (Salzburg, Austria), discussed how correct and reliable determination of *CYPD26* functional activity is extremely complex and needs careful evaluation. The in vitro assay developed at PharmGenetix, combined with extensive genotyping, will allow for a fast and precise definition of the activity of various *CYP2D6* alleles.

2.5. Gene-Environment Interactions in Cardio-Metabolic Diseases and Cancer

Chairs: Robert Barouki, Paris, France/Michael Marschler, Mannheim, Germany

Georges Dedoussis (Athens, Greece) inaugurated the third day of the conference with a presentation on nutrigenetics in non-alcoholic fatty liver disease (NAFLD). This disease is primarily activated by dietary factors, obesity, and insulin resistance. However, a polygenic background is believed to impact susceptibility to the onset and progression of NAFLD. Therefore, the study of gene–diet interactions that lead to NAFLD onset and progression became an important subject of research and key to enabling the application of personalised nutrition and thus providing patients with more efficient therapies. Georges illustrated his presentation with very interesting examples of diet experience: Diet with increased fish intake, which would normally be considered as healthy, resulted in increased intrahepatic

accumulation of triglycerides for the carriers of a TM6SF2 variant, one of the four genetic variants that have been identified as risk factors for NAFLD. Such findings demonstrate the importance of nutrigenetics research and application of the right diet for the patients with NAFLD.

Hugues Aschard (Paris, France) explained the theoretical aspects of interaction tests in regression models with the use of real data examples, including breast cancer, to illustrate the potential impact of having the right understanding of statistical analysis on detection and clinical utility of data. Hugues showed that the simplest biological interaction models—in which the magnitude of a genetic effect depends on a common exposure—are among the most difficult to identify. Moreover, he presented new insights on advantages and limitations of multivariate interaction models that can be leveraged for future method development and for the improvement of our understanding of the interplay between genetic variants and environmental exposures in multifactorial traits and diseases. The identification of precise and predictive biomarkers of health and disease is a critical objective of clinical biochemistry and biomedical research.

Robert Barouki (Paris, France) presented how recent developments in exposome studies and epigenomics could be used to support preventive action in medicine. Research in the exposome field allowed the development of sensors and biological biomarkers using "-omics" technologies that can support the prediction of the effect of those exposures on human health. Robert pointed out that precision medicine has primarily focused on adapting treatments to the genetic profiles of tumours, but in fact, it originally had a wider scope, including the use of robust biomarkers for disease prevention. Using different sensors, metabolomics and epigenomics, it now seems possible to generate precise observations that could be of value for prevention.

2.6. Latest Insights in Stroke: Clinical Trials and Applicability of '-OMICS' Data in Patients' Stratification

Chairs: John Lamont, Crumlin, County Antrim, UK/Federico Innocenti, Chapel Hill, NC, USA

As in many diseases, response to treatment in stroke differs from one individual to another. Pharmacogenetics aims to stratify patients based on their genetic information into groups that are more likely to benefit from a particular intervention, in order to select appropriate treatments. An increased understanding of stroke pharmacogenetics has been driven by advances in genotyping technology and increased interest in developing targeted pharmaceutical treatments. Guillaume Pare (Ontario, Canada) summarised the pharmacogenetics of stroke by providing an overview of the genetic variants that contribute to the individual responses to aspirin, clopidogrel, warfarin, and dabigatran and discussed considerations for evidence-based implementation. He suggested that the reasons for the limitations in implementation of pharmacogenetics in clinical settings can be found in the lack of awareness and need for evidence-based recommendations.

Alexander Haliassos (Athens, Greece) continued the discussion on stroke, mainly focusing on protein biomarkers in the diagnosis of ischemic stroke [5]. Fast and accurate diagnosis of patients is crucial but challenging, since ischemic stroke cannot be identified based only on clinical assessment. New non-invasive tests that could quickly distinguish stroke from stroke mimics and ischemic from haemorrhagic stroke would be a good alternative to CT or MRI imaging. Alexander Haliassos suggested that the identification of blood biomarkers of stroke is a prospective area of research, since their potential use is not limited to diagnosis and differentiation but can be applied to prognosis and patient treatment monitoring. However, he pointed out factors that are limiting such endeavours: The heterogeneity and complexity of stroke aetiology, analytical issues, and difficulties at the interpretation of laboratory measurements. To date, many biomarkers have been identified, but none so far have shown sufficient sensitivity and specificity to be used in the clinical setting.

2.7. Pain Management—A 'Journey' from 'Clinical Trial' via 'Post-Marketing Pharmacovigilance/Risk Management' to 'Success Story'

Chairs: Michael Marschler, Mannheim, Germany/Lynn Webster, Salt Lake City, UT, USA

The session started with the journey towards successful pain management using genetic biomarkers as common denominators for specific pain disorders. Lynn Webster (Salt Lake City, UT, USA) stated that every clinician knows patients may develop very different degrees of pain and suffering after the same standard procedures. This interindividual variability is the rule rather than the exception.

The clinical problem is that this huge variability is unpredictable. There are multiple reasons for the variability, but an individual's genotype is a major contributing factor. Genes have ethnicity-specific effects in response to pain. For example, the *GCH1* gene is protective for pain in a Caucasian population, and it is a risk factor for pain in African Americans with sickle cell pain. Then, there are genes that encode metabolic factors, and they have significant variation among ethnic populations. Thus, these ethnic-specific effects are common. Also very common are sex-specific effects of genes that are described for many pain genes. There are three scenarios of these sex-specific effects: The same genetic variations may have much stronger effects in one sex compared to another; they may have no effect in one sex compared to another; or they may have the opposite effect in females, for example, compared to males.

Genes also have combined effects on pain. For example, when analysing both the catechol-*O*-methyltransferase (*COMT*) and mu-opioid receptor (*OPRM1*) common-function SNPs, the morphine requirements can be much less than in patients without the common SNPs. In addition, the genetic biomarker, brain-derived neurotrophic factor gene (*BDNF*), has recently been shown to be associated with an increased risk of chronic postsurgical pain. The reason for this is unclear.

Dr Webster also described a large international study designed to see if genetic biomarkers for chronic low back pain (CLBP) could be identified. In a just published paper, the researchers reported that a meta-analysis of GWAS for CLBP identified and replicated genetic loci associated with CLBP (SOX5, CCDC26/GSDMC and DCC).

Laure Elens (Brussels, Belgium) continued the journey, presenting pharmacogenetics of opioid treatments. Interindividual differences in sensitivity to pain can be very problematic, especially during medical interventions and in managing discomfort after medical intervention. However, there is evidence that genetic variations and SNPs in genes, implicated in the opioid response pathway or the mechanisms of pain sensing, might explain the various phenotypes related to pain susceptibility and response to opioid therapy. Laure Elens summarised current knowledge in opioid pharmacogenetics and possible beneficial applications in clinical practice to allow better individualisation of pain discomfort when treated with opioids, and provided an overview of opioid pharmacogenetic studies and how it might be useful in understanding interindividual differences in pain perception and treatment effectiveness and/or drug-related toxicity, to determine whether genetic testing has any clinical significance in decision making for the management and control of pain, or not.

Noelia Martin Granado (Madrid, Spain) followed with a presentation about pharmacogenomics in post-marketing pharmacovigilance. Firstly, she reminded the audience that pharmacovigilance is the science and activities related to the detection, assessment, understanding, and prevention of adverse drug reactions (ADRs), since there is robust evidence that genomic factors may play a key role in the pathogenesis of both predictable and idiosyncratic ADRs. Over the past years, several studies have identified a number of common and rare variants that are associated with an increased risk of ADRs. In particular, polymorphisms in genes encoding drug transporters, drug-metabolising enzymes, and drug targets (e.g., enzymes, receptors) can lead to the occurrence of ADRs. As more reliable and affordable genetic testing tools become available, pharmacogenomics seems promising in terms of facilitating personalised drug therapy by maximising the therapeutic efficacy of drugs while minimising the occurrence of ADRs in patients. Due to the strong synergy existing between pharmacovigilance and pharmacogenomics disciplines, as both aim to understand the inter-relationships between drug therapeutic efficacy and safety, it has been proposed to put into practice pharmacogenovigilance,

a combination of pharmacogenomics and pharmacovigilance: Pharmacovigilance activities fed and guided by accompanying pharmacogenomics analyses.

Andrew Purchase (Swansea, UK) finished the journey and discussed how the adoption of pharmacogenomics leads to patient benefit. He provided an overview of the increasing acceptance and adoption of pharmacogenomics into clinical practice, and how this translates into benefits for the patient. The presentation also explored methodologies for raising patient awareness and understanding of the potential benefits linked to pharmacogenomics, as well as current/future technologies that support the use of genomic data. Genomic data have emerged as a key component of precision medicine, offering a way for providers to determine the most effective therapies for patients based on the unique makeup of their DNA. As an example, Andrew informed the audience that before prescribing the antiviral drug Abacavir (Ziagen), doctors now routinely test HIV-infected patients for a genetic variant that makes them more likely to have a negative reaction to the drug. The need to test patients for the *HLA-B* (type 5701) gene and to remind patients to contact their doctor immediately if they develop symptoms is suggestive of hypersensitivity. Andrew concluded that with reduced costs, PGx has the potential to further change the landscape of healthcare and improve the treatment options for patients, whilst reducing/removing risk.

2.8. Pharmacogenomics: Challenges of Clinical Translation

Chairs: Urs A. Meyer, Basel, Switzerland/Charity Nofziger, Salzburg, Austria

The first session of the day started with a lecture of Federico Innocenti (Chapel Hill, NC, USA), who discussed the role of genomics in safety of drug treatment in oncology [6]. He reminded the audience about the importance that adverse events have on drug treatment in cancer. Not only do they have a significant burden on the quality of life of the patient and their families, but they also reduce confidence in the treatment and often lead to permanent discontinuation of therapy. However, genetic analyses can increase patient safety and improve the safety of cancer drugs in several ways. These aspects are of particular importance for implementation, as access to genetic profiling is becoming more common for patients.

Commonly, drugs frequently associated with adverse events are known to be metabolised by enzymes with genetic polymorphisms. Accurate genetic analysis of patients using such treatment is therefore of crucial importance to prevent possible adverse events. *CYP2D6* is responsible for metabolising around 25% of clinically prescribed drugs, and is riddled with function altering genetic variations, including point mutations, insertions and deletions, as well as hybrid formation with its pseudogene, *CYP2D7*—all of which make accurate genotyping quite challenging [7]. In the following speech, Charity Nofziger (Salzburg, Austria) presented the technical challenges in genotyping *CYP2D6*, focusing on allele drop-out events that can severely alter the prediction of metabolising phenotypes within a particular patient.

Urs A. Meyer (Basel, Switzerland) continued the debate with an overview of present studies on the clinical implementation of digital signatures of the drug response profiles. In most therapies, efficacy and toxicity is determined by the combined action of multiple genetic, epigenetic, environmental, and host factors, and the genetic contribution may be too small for predicting drug response. Therefore, application of "-omics"-based data in combination with clinical and environmental factors to predict an individual's drug response is one of the visions of pharmacogenomics, personalised medicine, and precision medicine. For an increasing number of actionable gene–drug interactions, practice guidelines based on genetic and clinical information have been established (http://cpicpgx.org). The incorporation of pharmacogenetic test results, combined with clinical decision support (CDS) into machine-readable electronic medical records (EMRs) allows a clinical implementation of this information. In the view of the above, Urs discussed how to best apply the already existing digital genomic and other "-omics" information to optimise drug response.

Belgin Süsleyici (Istanbul, Turkey) presented an example of implementation of pharmacogenetics in Turkish clinical practice, focusing on the enzyme dihydropyrimidine dehydrogenase (DPD). DPD is an important catalyser of fluorouracil metabolism, a common drug for treatment of colorectal carcinoma. However, individuals carrying at least one copy of a loss of function DPYD variant may not be able to metabolise fluorouracil at normal rates and are at risk of life-threatening toxicity. In recent research, associations between fluorouracil treatment outcomes and germ line polymorphisms in DPD were analysed. Belgin demonstrated that the variances in the DPD gene can have no functional consequences on enzymatic activity, can decrease, or increase metabolism of the drug. She concluded that clinicians should be strongly encouraged to consider testing for DPD poor metaboliser variants as a rational pretreatment screening for patient candidates to a fluoropyrimidine-based therapy, in order to prevent toxicities.

For the closure of the session, Alexander Jetter (Zurich, Switzerland) shared his views on pharmacogenetics as the basis of individualised therapy decisions. He presented two clinical cases and discussed the choice of therapy for each presented case that was based on the genomic data of patients. For some medications, such as those containing carbamazepine, such a procedure is highly recommended. Moreover, different pharmacogenetic test approaches were presented. Alexander highlighted that pharmacogenetic tests should be used in combination with nongenetic information in order to treat "the right patient with the right drug in the right dose at the right time".

2.9. What Can We Learn from Electronic Health Records?

Chairs: Panagiotis Deloukas, London, UK/ Markus Paulmichl, Salzburg, Austria

The final session was introduced by Harry Hemingway (London, UK) continued the session with a presentation of a large-scale enquiry of longitudinal electronic health records (EHR) linked to genomic and "-omics" resources [8]. Such resources can offer discovery science with well-established designs, improve care through preventive genomics, and provide new kinds of scientific enquiry based on an agnostic understanding of longitudinal human phenome sequences. He presented the national platform CALIBER (www.ucl.ac.uk/health-informatics-caliber) that developed phenotyping algorithms for 72 diseases and risk factors and made them available through an open-access data portal, together with a set of open source tools. He concluded that in the era of precision medicine, robust approaches to creating, validating, and sharing EHR-derived phenotypes are critical to enable cross-source analyses of thousands of simple and complex traits in millions of individuals. The CALIBER approach provides a transparent methodology for transforming raw EHRs to reproducible phenotypes for use in such studies.

Richard Trembath (London, UK) introduced the East London Genes and Health cohort, containing primary health data of over 20,000 people, aiming to improve health among people of Pakistani and Bangladeshi heritage in East London by analysing the genes and health of local people.

The last presentation of the conference was given by Panagiotis Deloukas (London, UK), who shared his experiences of linking genetic data to risk of recurrent cardiovascular events through hospital records. So far, 70 loci were associated with coronary heart disease and myocardial infraction and are currently undertaking further analyses in UK Biobank, including gene–gene and gene–environment interactions, together with an integrated approach of using CHD risk factors, such as lipid levels, blood pressure, haematological traits, and height, through QTL analyses in large, well-phenotyped cohorts with an ultimate goal of developing a risk model for recurrent MI and testing its clinical utility.

3. Oral Communications Session

Chairs: Sofia Siest, Nancy, France/Eric Boerwinkle, Houston, USA

The conference invited young researchers and industrials to submit abstracts, to present their current work and win a reward, given by the organisation committee. Six participants were selected for 15 min oral presentations.

Said El Shamieh (Beirut, Lebanon): Genetic and protein profiling of cancer tumours, a first step towards personalised therapy.

Four cases of personalised therapy for cancer were presented. Molecular profiling has managed to provide better therapeutic options in 3 out of 4 of these cases. Furthermore, a guideline for a personalised treatment in endometrial cancer was presented.

Gilles Lunzenfichter (Luzern, Switzerland): Welcome to a new world of Intelligent Connected Care.

The presentation introduced us to M+ Intelligent Connected Care that enables synchronisation of patients' data with physician devices and enables the live sharing of patients' data and measurements. It is specialised in chronic conditions such as diabetes, hypertension, and cardiovascular and overweight indications.

Alexia Giannoula (Barcelona, Spain): Temporal comorbidity patterns in prostate cancer disease trajectories based on semantic, phenotypic, and genetic similarities.

Comorbodities and studies of molecular interactions were presented on prostate cancer patients. The aim was to unravel related pathogenetic processes and hidden diseases patterns.

Laurent Becquemont (Paris, France): Investigation of novel biomarkers of drug-induced kidney injury in renal transplant recipients undergoing graft biopsy.

New urinary biomarkers for drug-induced kidney injury were presented, in comparison to standard biomarkers. He concluded that new urinary biomarkers alone do not outperform the classical ones.

Bianca van den Bosch (Maastricht, The Netherlands): Evolution of Dihydropyrimidine Dehydrogenase (DPD) Diagnostics in a Single Center in a Time-Period of Eight Years.

The screening of the *DPYD* gene to detect variants related to change enzyme activity. She concluded that a combination of genetic screening and measurements of enzyme activity is better than these methods alone.

Carlos Malpica (Doha, Qatar): Bridging the Multi-Omics Precision Medicine Gap in the Middle East: The Valdia Health Experience.

Carlos discussed the implementation of precision medicine in Qatar and future challenges for establishing new networks of collaborations that will make this implementation possible.

The participants of the Santorini Conference voted for Bianca van den Bosch and Carlos Malpica to receive the Gerard SIEST Awards for the best oral presentations. Bianca van den Bosch obtained the award offered by the *Journal of Personalised Medicine* and Carlos Malpica the award offered by the University of Lorraine.

4. Posters

Sixty-three posters were presented that were classified in 2 groups (group 1: "-omics" biomarkers and group 2: Pharmacogenomics). Alzbeta Hlavackova received one "Gerard SIEST" award for the best poster in group 1 with the poster entitled "Changes in Plasma Metabolomic Profile during Acute Myeloid Leukemia Treatment" offered by De Gruyter publisher. Carolina Dagli-Hernandez won the second "Gerard SIEST" award for the best poster in group 2 with the poster entitled "Rare Variants in CYP2C9 and CYP3A5 Detected in Patients with Familial Hypercholesterolemia with Statin-Related Adverse Events" granted by the European Society of Pharmacogenomics and Personalised Therapy (ESPT).

5. Industrial Workshops

One of the important aims of the conference was also to establish strong relations among different sectors: Research–hospitals and industry. New improvements and techniques from innovative companies were presented during the conference.

5.1. Randox Workshop

Randox is a world leader in the in-vitro diagnostics industry, offering the most comprehensive insight into patient diagnosis, allowing for more effective disease management and treatment. The company developed biochip array technology (BAT), which is capable of simultaneous multianalyte diagnostic testing within the fields of clinical research and drugs of abuse testing. Cliona Johnston (Crumlin, County Antrim, UK) presented new possibilities for clinical evaluation of the type 1 diabetes based on genetic risk evaluation, using the innovative biochip array technology.

5.2. Illumina Workshop

Medical progress is vastly dependent upon tight collaboration between academy, hospitals, and industry. Illumina is an enterprise strongly involved in genetics and pharmacogenomics studies. The expert in pharmacogenomics Ron Van Schaik (Rotterdam, The Netherlands) discussed his experiences with pharmacogenetics using the Illumina GSA array. His speech was followed with a presentation by Lili Milani (Tartu, Estonia) on the Estonian personalised medicine initiative: A calculation of polygenic risk scores, pharmacogenetics, and rare mutations.

6. Closed Meetings

6.1. Closed Meeting—Mast4health Project

A closed meeting was reserved for the collaborators of the MAST4HEALTH project, a Marie Sklodowska-Curie Actions (MSCA) Research and Innovation Staff Exchange (RISE) program under the EU Horizon 2020 framework.

6.2. Closed Meeting—VEGF Consortium

The members of the VEGF Consortium (www.vegfconsortium.org) gathered together in a closed meeting. They hdiscussed the past work of the consortium and future projects, aiming to maintain and further develop the transnational collaborative network, dedicated to large integrative and multidisciplinary genomic studies of the Vascular Endothelial Growth Factor, in order to generate applicable knowledge for medical practice.

7. Conclusions and Perspective Remarks

The Conference officially finished with conclusions and perspective remarks given by Sofia Siest (Nancy, France) and Urs A. Meyer (Basel, Switzerland). They underlined the success of the conference, which, with pertinent illustrations and examples, successfully answered the questions that had been set:

➢ Can genetic screening help to identify individuals at greatest risk for cardiometabolic diseases and cancer?
➢ What is the greatest clinical need with regard to diagnosis, prediction, and patient stratification for these pathologies and how is this/can this be addressed?
➢ Does a genetic risk score identify patients at the highest risk and is its use justified in clinical practice?
➢What are the challenges of the current clinical trials?
➢ What about comorbidities with ageing?
➢ What is the impact of pharmacogenomics on:

- The deliverance of more predictable responses to drug therapy?
- The minimisation of the occurrence and severity of adverse drug reactions?
- The conduction of more cost-effective clinical trials?
- Drug discovery and the drug development process?

➢ Do the existing diagnostic tools answer the needs of pharmacogenomics?

Finally, Sofia Siest announced the next event—The 10th Santorini Conference, in 2020 (28 September–1 October).

Author Contributions: S.V.S. was responsible for the preparation of scientific programme and overall organisation of the conference, V.G. and MGS were responsible for drafting the manuscript and participated in the organisation of the conference, A.M.P., G.W., C.M., B.H. and S.K. participated at the conference organisation, E.B., G.D., P.D., F.I., J.L., M.M., H.M., U.A.M., C.N., M.P., C.V. and L.W. were the chairs of the sessions and were responsible for the scientific content of this article.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Visvikis-Siest, S.; Gorenjak, V.; Stathopoulou, M.G. Personalised Medicine: The Odyssey from Hope to Practice. *J. Pers. Med.* **2018**, *8*, 31. [CrossRef] [PubMed]
2. Newgard, C.B. Metabolomics and Metabolic Diseases: Where Do We Stand? *Cell Metab.* **2017**, *25*, 43–56. [CrossRef] [PubMed]
3. White, P.J.; McGarrah, R.W.; Grimsrud, P.A.; Tso, S.C.; Yang, W.H.; Haldeman, J.M.; Grenier-Larouche, T.; An, J.; Lapworth, A.L.; Astapova, I.; et al. The BCKDH Kinase and Phosphatase Integrate BCAA and Lipid Metabolism via Regulation of ATP-Citrate Lyase. *Cell Metab.* **2018**, *27*, 1281–1293. [CrossRef] [PubMed]
4. Kulasingam, V.; Prassas, I.; Diamandis, E.P. Towards personalized tumor markers. *NPJ Precis. Oncol.* **2017**, *1*, 17. [CrossRef] [PubMed]
5. Makris, K.; Haliassos, A.; Chondrogianni, M.; Tsivgoulis, G. Blood biomarkers in ischemic stroke: Potential role and challenges in clinical practice and research. *Crit. Rev. Clin. Lab. Sci.* **2018**, *55*, 294–328. [CrossRef] [PubMed]
6. Innocenti, F. Bringing a genomic perspective to the safety of drug treatment in oncology. *F1000 Research* **2017**, *6*. [CrossRef] [PubMed]
7. Nofziger, C.; Paulmichl, M. Accurately genotyping CYP2D6: Not for the faint of heart. *Pharmacogenomics* **2018**, *19*. [CrossRef] [PubMed]
8. Rapsomaniki, E.; Timmis, A.; George, J.; Pujades-Rodriguez, M.; Shah, A.D.; Denaxas, S.; White, I.R.; Caulfield, M.J.; Deanfield, J.E.; Smeeth, L.; et al. Blood pressure and incidence of twelve cardiovascular diseases: Lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people. *Lancet* **2014**, *383*, 1899–1911. [CrossRef]

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

DE GRUYTER Clin Chem Lab Med 2017; x(x): xxx–xxx

Letter to the Editor

Maria G. Stathopoulou, Ting Xie, Daniela Ruggiero, Jerome Chatelin, Marc Rancier, George Weryha, Mary Jo Kurth, Alex-Ander Aldasoro Arguinano, Vesna Gorenjak, Alexandros M. Petrelis, Georges Dagher, George Dedoussis, Panagiotis Deloukas, John Lamont, Janja Marc, Maurizio Simmaco, Ron H.N. van Schaik, Federico Innocenti, Jean-Louis Merlin, Jochen Schneider, Behrooz Ziad Alizadeh, Marina Ciullo, Sudha Seshadri and Sophie Visvikis-Siest*, The VEGF Consortium

A transnational collaborative network dedicated to the study and applications of the vascular endothelial growth factor-A in medical practice: the VEGF Consortium

https://doi.org/10.1515/cclm-2017-0838
Received September 18, 2017; accepted September 19, 2017

Keywords: collaborative network; multidisciplinary genomic studies; personalized medicine; VEGF-A.

To the Editor,

The Vascular Endothelial Growth Factor European Genomic Federation (VEGF) Consortium (www.vegfconsortium.org) was founded in June 2014 by Sophie Visvikis-Siest (chair of the consortium) and an international group of researchers with an interest focused on VEGF-A and its implications in personalized medicine.

Here, we present the VEGF Consortium and describe its objectives and ambitions, its structure and its components together with the methodologies used in projects and preliminary results.

The VEGF Consortium aims to develop a transnational collaborative network dedicated to large integrative and multidisciplinary genomic studies of the VEGF-A in order to generate applicable knowledge for medical practice thanks to the following specific objectives:

- to combine data from multiple cohorts in order to identify VEGF-A '-omics' profiling in health and non-communicable diseases
- to elucidate the pivotal role of VEGF-A in the pathophysiology of non-communicable diseases
- to demonstrate the patients' stratification potential of VEGF-A '-omics' profiling
- to implement the research results into clinical practice and establish the role of VEGF-A as a predictive, preventive, diagnostic and prognostic biomarker
- to provide information on the effect of VEGF-A '-omics' profiling in side effects and response to therapy through pharmacogenomics studies

***Corresponding author: Dr. Sophie Visvikis-Siest,** UMR INSERM U1122; IGE-PCV "Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire", Université de Lorraine, 30 Rue Lionnois, 54000 Nancy, France, Phone: +33 (0)6 07 60 25 69, Fax: +33 (0)3 83 32 13 22, E-mail: sophie.visvikis-siest@inserm.fr
Maria G. Stathopoulou, Ting Xie, Jerome Chatelin, Marc Rancier, George Weryha, Alex-Ander Aldasoro Arguinano, Vesna Gorenjak and Alexandros M. Petrelis: UMR INSERM U1122; IGE-PCV "Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire", Université de Lorraine, Nancy, France
Daniela Ruggiero and Marina Ciullo: Institute of Genetics and Biophysics, National Research Council of Italy, Naples, Italy
Mary Jo Kurth and John Lamont: Randox Laboratories Ltd., Crumlin, UK
Georges Dagher: Biobanking and Biomolecular Resources Research Infrastructure (BBMRI)/INSERM US 13, BIOBANQUES, Paris, France
George Dedoussis: Department of Nutrition Dietetics, Harokopio University of Athens, Athens, Greece
Panagiotis Deloukas: Queen Mary University of London, London, UK
Janja Marc: Faculty of Pharmacy, University of Ljubljana, Ljubljana, Slovenia
Maurizio Simmaco: Sant'Andrea Hospital – Sapienza University of Rome, Rome, Italy
Ron H.N. van Schaik: European Society of Pharmacogenomics and Personalised Therapy (ESPT), Nancy, France
Federico Innocenti: University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
Jean-Louis Merlin: Institut de Cancérologie de Lorraine et Université de Lorraine, Vandoeuvre les Nancy, France
Jochen Schneider: University of Luxembourg, Luxembourg, Europe
Behrooz Ziad Alizadeh: Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
Sudha Seshadri: Department of Neurology, Boston University School of Medicine, Boston, MA, USA

- to propose implementation strategies and European guidelines involving VEGF-A '-omics' profiling for the management of non-communicable diseases
- to share methodologies, data and knowledge in the field of '-omics' management and innovative statistics
- to develop standardized teaching and evaluation methods practiced and validated by the consortium.

It comprises 11 working groups, where the partners of the consortium are participating based on their expertise:

1. VEGF-A '-omics' profiling in health
2. VEGF-A '-omics' profiling in diseases
3. '-Omics' technologies
4. Methodological aspects
5. VEGF-A clinical implementation
6. Pharmagenomics
7. Endothelins and endothelial factors
8. VEGF-A basic research (cancer cell lines, animal models)
9. VEGF-A and inflammation
10. Communication and scientific/educational meetings
11. Raising awareness of populations

Prospective, longitudinal, family-based or population-based cohorts of healthy individuals are the core of the consortium. These are the following:

1. The STANISLAS family study: This is a longitudinal family structure cohort of community-based population of French origin recruited in 1993–1995, followed up for 15 years. A number of 1006 nuclear families comprising two parents and at least two biological children over 6 years old are included in the study [1].
2. The Framingham Heart Study: It is an ongoing, longitudinal, community-based, observational cohort study that was initiated in 1948 to prospectively investigate the risk factors for cardiovascular disease (CVD). The original cohort enrolled 5209 men and women, the offspring cohort enrolled 5124 participants (including 3514 biological offspring) and Gen 3 included 4095 individuals [2–4].
3. Cilento study: This is a population-based study that aims at identifying genetic risk factors for common diseases and traits. The sample includes isolated populations from three villages (Campora, Gioi and Cardile). The overall sample size of individuals participating to the study is 2100 [5].
4. The LifeLines Cohort Study and Biobank (LLs): LLs is a multidisciplinary prospective population-based cohort study examining the health and health-related behaviors of 167,000 persons living in the north east region of The Netherlands in a three-generation design [6].

Furthermore, cohorts of patients and case-control studies are also included in the consortium:

5. The Hellenic Study of Interactions between SNPs and Eating in Atherosclerosis Susceptibility: case-control study of coronary artery disease.
6. Ljubljana patients: case-control studies recruited from Ljubljana, Slovenia that include cases of osteoporosis, osteoarthritis, CVD and diabetes [7–9].

In addition, partners of the consortium have access to large biobanks such as the UK Biobank (Panagiotis Deloukas), Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium BBMRI-ERIC (including Sophie Visvikis-Siest with the Biological Resources Center IGE-PCV – BB-0033-00051) and the Alliance for Clinical Trials in Oncology (Federico Innocenti).

Given the wide range of the research field and the need for application of different methodologies in order to successfully achieve its objectives, the consortium is composed of scientists with different and complementary expertise and with a large range of resources such as large study populations, research materials and harmonized data.

Although the VEGF Consortium was officially founded on 2014, it was based on the long-term collaboration between some of its founding partners. Therefore, a number of significant results have been published already, and many projects are ongoing, with promising preliminary results.

Among the most basic steps was the identification through two GWAS of 10 genetic variants that explain >50% of the circulating VEGF-A levels variability [10, 11]. This is a unique finding among GWASs. In most GWASs, the identified variants do not explain >10% of the individual variability of the assessed traits. This finding has strengthened our belief that VEGF-A will indeed be used as strong biomarker for personalized medicine. Significant associations between some of these polymorphisms and intermediate phenotypes of chronic diseases have been identified since then: high-density and low-density lipoprotein [12], L-selectin gene expression [13] and free tri-iodothyronin (FT3) levels [14]. Significant epistatic interactions between these variants were observed for intercellular adhesion molecule 1 (ICAM-1), E-selectin, interleukin 6 and tumor necrosis factor α (TNF-α) plasma levels [13]. Concerning specific disease risk, we have shown that these polymorphisms and/or

their epistatic interactions can affect the risk for depression [15], and for autoimmune thyroid diseases [14], whereas no associations were found for diabetes type 2 [16].

Concerning the expression isoforms of *VEGF-A* gene, we have shown that these are significantly associated with ICAM-1, L-selectin and TNF-α expression [13] and with specific autoimmune thyroid diseases [17].

Furthermore, we have also identified associations between VEGF-A circulating levels and thyroid hormones levels [17] and with ICAM-1 and E-selectin levels [13].

Through a candidate gene approach of polymorphisms in genes involved in angiogenesis, we have identified direct and epistatic effects of variants on nitric oxide synthase 3 (*NOS3*), *CD14*$^{+}$ monocytes, matrix metalloproteinases (*MMPs*) and interleukin 4 receptor (*ILR4*) genes with levels of VEGF-A and VEGF-A expression isoforms, but also gene × environment interactions [18].

An important result is also the production of two patents based on the results of studies performed by partners of the consortium [19, 20].

Several projects are ongoing and focused on CVD intermediate phenotypes, thyroid diseases, cancer and stroke.

Three face-to-face meetings have been organized to date: Paris (2014), Budapest (2015) and Santorini (2016).

The VEGF Consortium is an ambitious international collaboration that aims to pave the way for the implementation of VEGF-A in personalized medicine and routine clinical practice. The designed projects take advantage of the wide expertise of its partners, the large infrastructures of cohorts and biobanks and a combination of the most up-to-date "-omics" approaches for generating multidimensional data, as well as 'systems medicine' approaches, network analysis and computational modeling methodologies.

Among its major originalities is that it integrates commercialization and communication platforms, targeting the valid and easy measurement of the identified biomarkers in large-scale settings but also the education of the general population, patients, scientists and health practitioners.

The VEGF Consortium is a novel consortium with an innovating structure and original goals. The ultimate goal is to ameliorate life expectancy, quality of life for individuals and financial benefits for health systems.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med 2008;46:733–47.
2. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol 2007;165:1328–35.
3. Dawber T, Meadors G, Moore F. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 1951;41:279–81.
4. Feinleib M, Kannel W, Garrison R, McNamara P, Castelli W. The Framingham Offspring Study. Design and preliminary data. Prev Med 1975;4:518–25.
5. Colonna V, Nutile T, Ferrucci RR, Fardella G, Aversano M, Barbujani G, et al. Comparing population structure as inferred from genealogical versus genetic information. Eur J Hum Genet 2009;17:1635–41.
6. Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al. Cohort profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2015;44:1172–80.
7. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet 2012;44:491–501.
8. Mirjanic-Azaric B, Vekic J, Zeljkovic A, Jelic-Ivanovic Z, Djeric M, Milivojac T, et al. Interrelated cathepsin S-lowering and LDL subclass profile improvements induced by atorvastatin in the plasma of stable angina patients. J Atheroscler Thromb 2014;21:868–77.
9. Zupan J, van't Hof RJ, Vindisar F, Haring G, Trebse R, Komadina R, et al. Osteoarthritic versus osteoporotic bone and intra-skeletal variations in normal bone: evaluation with microCT and bone histomorphometry. J Orthop Res 2013;31:1059–66.
10. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, et al. Six novel loci associated with circulating VEGF levels identified by a meta-analysis of genome-wide association studies. PLoS Genet 2016;12:e1005874.
11. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, et al. Identification of cis- and trans-acting genetic variants explaining up to half the variation in circulating vascular endothelial growth factor levels. Circ Res 2011;109:554–63.
12. Stathopoulou MG, Bonnefond A, Ndiaye NC, Azimi-Nezhad M, El Shamieh S, Saleh A, et al. A common variant highly associated

with plasma VEGFA levels also contributes to the variation of both LDL-C and HDL-C. J Lipid Res 2012;54:535–41.
13. Azimi-Nezhad M, Stathopoulou MG, Bonnefond A, Rancier M, Saleh A, Lamont J, et al. Associations of vascular endothelial growth factor (VEGF) with adhesion and inflammation molecules in a healthy population. Cytokine 2013;61:602–7.
14. Zaaber I, Rancier M, Stathopoulou MG, Saleh A, Marmouch H, Masson C, et al. Plasma VEGF-related polymorphisms are implied in autoimmune thyroid diseases. Autoimmunity 2016;49:229–35.
15. Xie T, Stathopoulou MG, de Andres F, Siest G, Murray H, Martin M, et al. VEGF-related polymorphisms identified by GWAS and risk for major depression. Transl Psychiatry 2017;7:e1055.
16. Bonnefond A, Saulnier PJ, Stathopoulou MG, Grarup N, Ndiaye NC, Roussel R, et al. What is the contribution of two genetic variants regulating VEGF levels to type 2 diabetes risk and to microvascular complications? PLoS One 2013;8:e55921.
17. Rancier M, Zaaber I, Stathopoulou MG, Chatelin J, Saleh A, Marmouch H, et al. Pro- and anti-angiogenic VEGF mRNAs in autoimmune thyroid diseases. Autoimmunity 2016;49:366–72.
18. Saleh A, Stathopoulou MG, Dade S, Ndiaye NC, Azimi-Nezhad M, Murray H, et al. Angiogenesis related genes NOS3, CD14, MMP3 and IL4R are associated to VEGF gene expression and circulating levels in healthy adults. BMC Med Genet 2015;16:90.
19. Visvikis-Siest S. Association of vascular endothelial growth factor genetic variant with metabolic syndrome, 2014.
20. Randox. Identification of Genetic Variants – VEGF, 2012.

ANNEX

SUPPLEMENTARY DATA

Epigenome-wide association study in healthy individuals identifies significant associations with DNA methylation and PBMCs extracts VEGF-A concentration

Vesna Gorenjak[1], Dwaine R.Vance[2], Sébastien Dade[1], Maria G. Stathopoulou[1], Lauren Doherty[2], Ting Xie[1], Helena Murray[2], Christine Masson[1], John Lamont[2], Peter Fitzgerald[2], Sophie Visvikis-Siest[¥, 1, 3]

[1] Université de Lorraine, Inserm, IGE-PCV, F-54000 Nancy, France

[2] Randox Laboratories Limited, Crumlin, Co. Antrim, Northern Ireland, United Kingdom

[3] Department of Internal Medicine and Geriatrics, CHU Technopôle Nancy-Brabois, Rue du Morvan, F-54511, Vandoeuvre-lès-Nancy, France

[¥]Corresponding author:

Dr. Sophie VISVIKIS-SIEST

INSERM UMR U1122; IGE-PCV, Faculté de Pharmacie – Université de Lorraine

30 rue Lionnois

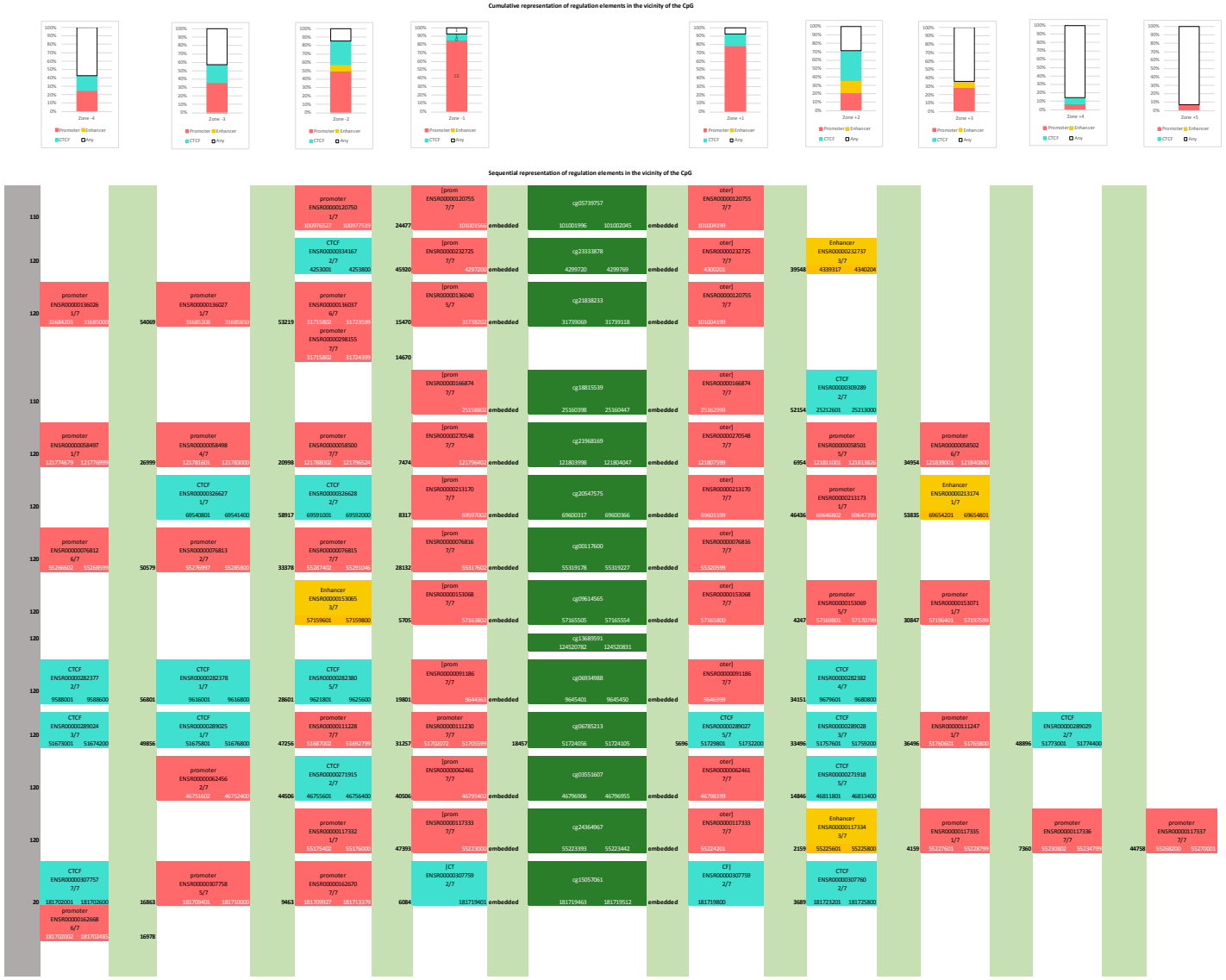
54000 Nancy, France

Tel: +33.6.07.60.25.69

E-mail address: sophie.visvikis-siest@inserm.fr

Sup. Figure 1: Genomic environment of six CpG sites. *Dark green regions present a CpG site, numbers on left and right of the box indicate a location, within which CpG can be found. Nearest genomic features upstream (left) or downstream (right) are presented for each CpG. Distance (bp) between each CpG and genomic feature is indicated in light green regions. Turquoise squares present CTCF regions, red promoter region and yellow enhancers. Square brackets [ ] indicate that CpG is located within genomic feature. Seven different PBMC cell types were looked up (Sup. Table 1). Number 1-7/7 in each box of particular genomic region is indicating in what extent this genomic feature is presented in PBMCs. Diagrams on the top are presenting patterns of genomic features that can be found in the genomic environment of CpGs. We can see that in the immediate proximity of CpG enhancers are the most common and that with distance, genetic features become less common (regions of non-coding DNA).*

Sup. Figure 2: *Genetic environment of fourteen CpG sites.*



Sup. Table 1: *PBMC cell types investigated in in silico analysis of CpG genomic environment.*

Cell type	Description
B cells (PB) Roadmap	REMC Epigenome (Class 5) for Primary B cells from Peripheral Blood
CD4+ ab T cell (VB)	CD4+ ab T cell from Venous Blood
CD4+ ab T cell (VB)	CD4+ ab T cell from Venous Blood
CM CD4+ ab T cell (VB)	CM CD4+ ab T cell from Venous Blood
Natural Killer cells (PB)	REMC Epigenome (Class 5) for Primary Natural Killer Cells from Peripheral Blood
T cells (PB) Roadmap	REMC Epigenome (Class 5) for Primary T Cells from Peripheral Blood
Naive B cell (VB)	Naive B cell from Venous Blood

Sup. Table 2: *Forty-one significant CpG sites related to VEGF concentration derived from PBMCs extracts.*

CpG Labels	P-value	Holm. Sig.	FDR	Effect size	Chromosome	Strand
cg15144153	2.80 x 10^ -07	FALSE	0.00400	0.00200	chr2	-
cg24364967	1.22 x 10^ -07	TRUE	0.00214	0.00122	chr2	-
cg15057061	1.08 x 10^ -07	TRUE	0.00200	-0.00151	chr3	-
cg23012579	4.29 x 10^ -07	FALSE	0.00519	-0.00178	chr5	-
cg18262852	1.25 x 10^ -06	FALSE	0.01312	0.00081	chr11	+
cg03551607	1.34 x 10^ -07	TRUE	0.00219	0.00115	chr13	+
cg06934988	1.96 x 10^ -08	TRUE	0.00051	-0.00188	chr17	-
cg17949256	3.20 x 10^ -07	FALSE	0.00422	-0.00123	chr17	+
cg03521258	2.87 x 10^ -06	FALSE	0.02492	-0.00183	chr19	+
cg05739757	5.64 x 10^ -15	TRUE	1.77 x 10^ -09	0.00177	chr2	-
cg05275012	8.81 x 10^ -08	TRUE	0.00173	0.00176	chr3	+
cg09614565	5.69 x 10^ -08	TRUE	0.00128	-0.00283	chr3	+
cg10517202	8.74 x 10^ -08	TRUE	0.00173	-0.00104	chr3	+
cg13689591	2.58 x 10^ -08	TRUE	0.00062	0.00200	chr3	+
cg18815539	1.70 x 10^ -12	TRUE	1.347 x 10^ -07	-0.00197	chr4	+
cg03663456	6.32 x 10^ -07	FALSE	0.00709	0.00144	chr7	-
cg16333561	7.02 x 10^ -11	TRUE	3.68 x 10^ -06	-0.00130	chr7	-
cg20547575	1.52 x 10^ -10	TRUE	5.96 x 10^ -06	-0.00350	chr7	-
cg12141623	1.83 x 10^ -06	FALSE	0.01658	-0.00116	chr8	-
cg19497501	6.06 x 10^ -06	FALSE	0.04651	0.00068	chr8	-
cg20085411	1.84 x 10^ -06	FALSE	0.01658	0.00099	chr8	+
cg22735324	2.76 x 10^ -07	FALSE	0.00400	0.00155	chr8	-
cg13496481	6.02 x 10^ -06	FALSE	0.04651	0.00321	chr9	-
cg23333878	3.86 x 10^ -14	TRUE	6.06 x 10^ -09	0.00270	chr9	-
cg04843252	4.20 x 10^ -07	FALSE	0.00519	0.00104	chr10	-
cg10928925	1.33 x 10^ -06	FALSE	0.01348	-0.00091	chr10	+
cg19004007	1.25 x 10^ -06	FALSE	0.01312	0.00134	chr10	-
cg19864972	1.46 x 10^ -06	FALSE	0.01438	0.00126	chr10	+
cg14682277	3.09 x 10^ -06	FALSE	0.02555	-0.00099	chr11	-
cg21968169	1.11 x 10^ -12	TRUE	1.16 x 10^ -07	-0.00189	chr12	-
cg18073151	2.93x 10^ -06	FALSE	0.02492	-0.00220	chr14	-
cg00117600	8.84 x 10^ -09	TRUE	0.00028	0.00118	chr15	-
cg03790192	3.22 x 10^ -07	FALSE	0.00422	-0.00131	chr16	+
cg08759276	3.31 x 10^ -09	TRUE	0.00012	-0.00153	chr16	+
cg16698748	1.67 x 10^ -06	FALSE	0.01593	0.00072	chr16	-
cg13332754	1.39 x 10^ -07	TRUE	0.00219	0.00164	chr18	+
cg06785213	1.06 x 10^ -08	TRUE	0.00030	0.00087	chr19	-
cg15014826	1.12 x 10^ -10	TRUE	5.04 x 10^ -06	-0.00157	chr19	-
cg25492620	5.26 x 10^ -07	FALSE	0.00612	-0.00154	chr19	-
cg21838233	4.01 x 10^ -12	TRUE	2.52 x 10^ -07	0.00191	chr20	-
cg06295071	5.99 x 10^ -06	FALSE	0.04651	0.00241	chr21	-

Sup. Table 3: *Summary table explaining the potential functionality and biological plausibility of each of the 20 significant CpGs and their nearby genes.*

CpG	Nearest Coding Gene(s)	Function	Disease	Interactions with VEGF
cg05739757	*RPL31*	Encodes a ribosomal protein	Increased expression in colorectal cancer (1) and in prostate cancer (2)	None identified
cg23333878	*GLIS3*	Acts as both a repressor and activator of transcription	Neonatal diabetes mellitus, with congenital hypothyroidism (3)	None identified
cg21838233	*TPX2*	Required for spindle fibre development	Overexpressed in gastric cancer cells (4)	*TPX2* silencing inhibit proliferation and invasion of colon cancer cells, possible related to down-regulation of *VEGF* (5)
cg18815539	*SEPSECS*	Implicated in the catalysation of O-phosphoseryl-tRNA (Sec) to selenocysteinyl-tRNA (Sec)	Pontocerebellar hypoplasia type 2D (6)	None identified
cg21968169	*LOC338799*	Long intergenic non-protein coding RNA 1089	Hormone-independent breast cancer,serves as a driver of mammary cell migration (7)	None identified
	SETD1B	A component of a histone methyltransferase complex that produces trimethylated H3 at Lys4	Gastric and colorectal cancers (8)	None identified
cg16333561 **(Located in enhancer region)**	*ARL4A*	A member of the ADP-ribosylation factor family of GTP-binding proteins	None identified	*ARL4A* has been differentially regulated in LV165 and ML20 in response to bevacizumab treatment (10)
	ETV1	ETS proteins regulate many target genes that modulate biological processes like cell growth, angiogenesis, migration, proliferation and differentiation.	Overexpressed in prostate cancer, melanoma and gastrointestinal stromal tumor (9)	EWS-ETS (ETS-family of ETV1 protein) fusion proteins upregulate the transcription of *VEGF*, overexpressed in Ewing tumors (11)
cg20547575	*AUTS2*	Activator of transcription and development regulator	Neurological disorders, autism, overexpressed in liver metastases of pancreatic cancer (12)	None identified
cg15014826 **(Located in enhancer region)**	*ZSWIM4*	Protein coding gene for zinc finger SWIM-Type Containing 4	None identified	None identified
	NANOS3	Protein coding gene for Nanos C2HC-Type zinc finger 3, maintenance of the undifferentiated state of germ cells, affects cell proliferation	Ovarian failure (13), male sterility (14), lung tumor (15)	None identified
cg00117600	*PIGB*	Required for the development of a glycosyl-phosphatidylinositol anchor, a glycolipid serving to anchor proteins to the cell surface	Hyperphosphatasia mental retardation syndrome (16)	None identified
cg08759276 **(Located in enhancer region)**	*FOXL1*	Regulation of metabolism, cell proliferation and gene expression during ontogenesis	Overexpression of *FOXL1* has protective role in breast (17), gastric (18), bladder (19), renal (20) cancer	None identified

	C16orf95	Protein coding gene for chromosome 16 open reading frame 95	Bilateral vesicoureteral Reflux (21)	None identified
cg05275012	*ZNF621*	Zinc finger protein is involved in transcriptional regulation	None identified	None identified
	CTNNB1	Part of a complex of proteins that constitute adherens junctions, regulators of epithelia cell growth and adhesion between cells	Colorectal cancer, pilomatrixoma, medulloblastoma, and ovarian cancer (22)	*VEGF* expression is regulated by beta-catenin (CTNNB1) in colon cancer (23, 24); *VEGF-D* mRNA is down-modulated by beta-catenin (25)
cg10517202 **(Located in enhancer region)**	*TBL1XR1*	Required for transcriptional activation by a variety of transcription factors mediated by nuclear receptors	Ovarian cancer (26), autism Spectrum Disorder, Intellectual disability, tumours, Pierpont syndrome (27)	Silencing of *TBL1XR1* decreased *VEGF-C* expression - *TBL1XR1* may function as an upstream regulator of *VEGF-C* (26)
	KCNMB2	Involved in the control of smooth muscle tone and neuronal excitability		None identified
cg09614565	*IL17RD*	Encodes a component of the interleukin 17 receptor signalling complex	Hypogonadotropic Hypogonadism (28)	None identified
cg13689591 **(Located in enhancer region)**	*KALRN*	Involved in the trafficking of vesicles. Its protein, Kalirin has been shown to be atheroprotective	Early onset coronary heart disease (29), Huntington's disease	None identified
cg06934988	*USP43*	Involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins	None identified	None identified
cg06785213	*HAS1*	Encodes predicted plasma membrane proteins with multiple transmembrane domains	Inflammatory and degenerative arthropathies such as rheumatoid arthritis, esophageal squamous cell carcinoma	HAS1-siRNA restrains the expression of *VEGF* (30)
cg13332754	*TSHZ1*	2H2-type zinc-finger protein family, involved in transcriptional regulation of developmental processes	Congenital aural atresia syndrome (31)	Circulating VEG-A levels in GWAS were related to variant in chromosome 18, located in an intergenic region downstream of the *ZADH2* gene and upstream of the *TSHZ1* gene (32)
	SMIM21	Protein coding gene for small integral membrane protein	None	None
cg03551607	*ESD*	Enzyme involved in the recycling of sialic acids	Retinoblastoma (33) and Wilson's disease	None
cg24364967	*C2orf63 (CLHC1)*	Unknown	None	None
cg15057061	*SOX2OT*	Transcription factor involved in embryonic development and cell fate determination	Optic nerve hypoplasia, breast cancer (34)	None

Most information gathered in the supplementary Table 3 were retrieved from public databases such as NCBI (https://www.ncbi.nlm.nih.gov/gene/), WikiGenes (https://www.wikigenes.org/) and GeneCards (http://www.genecards.org/).

REFERENCES

1. K. A. Chester, L. Robson, R. H. Begent, I. C. Talbot, J. H. Pringle, L. Primrose, A. J. Macpherson, G. Boxer, P. Southall and A. D. Malcolm: Identification of a human ribosomal protein mRNA with increased expression in colorectal tumours. Biochim Biophys Acta, 1009(3), 297-300 (1989)

2. Y. Maruyama, T. Miyazaki, K. Ikeda, T. Okumura, W. Sato, K. Horie-Inoue, K. Okamoto, S. Takeda and S. Inoue: Short hairpin RNA library-based functional screening identified ribosomal protein L31 that modulates prostate cancer cell growth via p53 pathway. PLoS One, 9(10), e108743 (2014) doi:10.1371/journal.pone.0108743

3. V. Senee, C. Chelala, S. Duchatelet, D. Feng, H. Blanc, J. C. Cossec, C. Charon, M. Nicolino, P. Boileau, D. R. Cavener, P. Bougneres, D. Taha and C. Julier: Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat Genet, 38(6), 682-7 (2006) doi:10.1038/ng1802

4. C. Tomii, M. Inokuchi, Y. Takagi, T. Ishikawa, S. Otsuki, H. Uetake, K. Kojima and T. Kawano: TPX2 expression is associated with poor survival in gastric cancer. World J Surg Oncol, 15 (2017) doi:10.1186/s12957-016-1095-y

5. J. JIAN, Y. HUANG, L.-z. LIU, S. LI and F. DENG: TPX2 gene-silencing inhibits the proliferation and invasion of human colon cancer SW480 cells. TUMOR, 36(6), 628-634 (2016)

6. K. Iwama, M. Sasaki, S. Hirabayashi, C. Ohba, E. Iwabuchi, S. Miyatake, M. Nakashima, N. Miyake, S. Ito, H. Saitsu and N. Matsumoto: Milder progressive cerebellar atrophy caused by biallelic SEPSECS mutations. J Hum Genet, 61(6), 527-31 (2016) doi:10.1038/jhg.2016.9

7. A. Sas-Chen, M. R. Aure, L. Leibovich, S. Carvalho, Y. Enuka, C. Körner, M. Polycarpou-Schwarz, S. Lavi, N. Nevo, Y. Kuznetsov, J. Yuan, F. Azuaje, I. Ulitsky, S. Diederichs, S. Wiemann, Z. Yakhini, V. N. Kristensen, L. Børresen-Dale A-, Y. Yarden, T. Sauer, J. Geisler, S. Hofvind, T. F. Bathen, E. Borgen, O. Engebråten, Ø. Fodstad, Ø. Garred, G. A. Geitvik, R. Kåresen, B. Naume, G. M. Mælandsmo, H. G. Russnes, E. Schlichting, T. Sørlie, O. C. Lingjærde, K. K. Sahlberg, H. K. Skjerven and B. Fritzman. EMBO Mol Med, 8(9), 1052-64 (2016) doi:10.15252/emmm.201606198

8. Y. J. Choi, H. R. Oh, M. R. Choi, M. Gwak, C. H. An, Y. J. Chung, N. J. Yoo and S. H. Lee: Frameshift mutation of a histone methylation-related gene SETD1B and its regional heterogeneity in gastric and colorectal cancers with high microsatellite instability. Hum Pathol, 45(8), 1674-81 (2014) doi:10.1016/j.humpath.2014.04.013

9. B. G. Jang, H. E. Lee and W. H. Kim: ETV1 mRNA is specifically expressed in gastrointestinal stromal tumors. Virchows Arch, 467(4), 393-403 (2015) doi:10.1007/s00428-015-1813-9

10. Y. Gokmen-Polar, C. P. Goswami, R. A. Toroni, K. L. Sanders, R. Mehta, U. Sirimalle, B. Tanasa, C. Shen, L. Li, M. Ivan, S. Badve and G. W. Sledge, Jr.: Gene Expression Analysis Reveals Distinct Pathways of Resistance to Bevacizumab in Xenograft Models of Human ER-Positive Breast Cancer. J Cancer, 5(8), 633-45 (2014) doi:10.7150/jca.8466

11. B. Fuchs, C. Y. Inwards and R. Janknecht: Vascular endothelial growth factor expression is up-regulated by EWS-ETS oncoproteins and Sp1 and may represent an independent predictor of survival in Ewing's sarcoma. Clin Cancer Res, 10(4), 1344-53 (2004)

12. Y. Han, G. Q. Ru, X. Mou, H. J. Wang, Y. Ma, X. L. He, Z. Yan and D. Huang: AUTS2 is a potential therapeutic target for pancreatic cancer patients with liver metastases. Med Hypotheses, 85(2), 203-6 (2015) doi:10.1016/j.mehy.2015.04.029

13. M. G. Santos, A. Z. Machado, C. N. Martins, S. Domenice, E. M. Costa, M. Y. Nishi, B. Ferraz-de-Souza, S. A. Jorge, C. A. Pereira, F. C. Soardi, M. P. de Mello, A. T. Maciel-Guerra, G. Guerra-Junior and B. B. Mendonca: Homozygous inactivating mutation in NANOS3 in two sisters with primary ovarian insufficiency. Biomed Res Int, 2014, 787465 (2014) doi:10.1155/2014/787465

14. K. Kusz, L. Tomczyk, A. Spik, A. Latos-Bielenska, P. Jedrzejczak, L. Pawelczyk and J. Jaruzelska: NANOS3 gene mutations in men with isolated sterility phenotype. Mol Reprod Dev, 76(9), 804 (2009) doi:10.1002/mrd.21070
15. S. Grelet, V. Andries, M. Polette, C. Gilles, K. Staes, A. P. Martin, C. Kileztky, C. Terryn, V. Dalstein, C. W. Cheng, C. Y. Shen, P. Birembaut, F. Van Roy and B. Nawrocki-Raby: The human NANOS3 gene contributes to lung tumour invasion by inducing epithelial-mesenchymal transition. J Pathol, 237(1), 25-37 (2015) doi:10.1002/path.4549
16. Y. Murakami, N. Kanzawa, K. Saito, P. M. Krawitz, S. Mundlos, P. N. Robinson, A. Karadimitris, Y. Maeda and T. Kinoshita: Mechanism for release of alkaline phosphatase caused by glycosylphosphatidylinositol deficiency in patients with hyperphosphatasia mental retardation syndrome. J Biol Chem, 287(9), 6318-25 (2012) doi:10.1074/jbc.M111.331090
17. J. Zhong, H. Wang, J. Yu, J. Zhang and H. Wang: Overexpression of Forkhead Box L1 (FOXL1) Inhibits the Proliferation and Invasion of Breast Cancer Cells. Oncol Res, 25(6), 959-965 (2017) doi:10.3727/096504016x14803482769179
18. Z. Ertao, C. Jianhui, C. Chuangqi, Q. Changjiang, C. Sile, H. Yulong, C. Shirong and W. Hui: Low level of FOXL1 indicates a worse prognosis for gastric cancer patients. Tumour Biol, 37(8), 11331-7 (2016) doi:10.1007/s13277-016-4890-8
19. Y. Qin, W. Gong, M. Zhang, J. Wang, Z. Tang and Z. Quan: Forkhead box L1 is frequently downregulated in gallbladder cancer and inhibits cell growth through apoptosis induction by mitochondrial dysfunction. PLoS One, 9(7), e102084 (2014) doi:10.1371/journal.pone.0102084
20. F. Q. Yang, F. P. Yang, W. Li, M. Liu, G. C. Wang, J. P. Che, J. H. Huang and J. H. Zheng: Foxl1 inhibits tumor invasion and predicts outcome in human renal cancer. Int J Clin Exp Pathol, 7(1), 110-22 (2014)
21. M. G. Butler, S. L. Dagenais, J. L. Garcia-Perez, P. Brouillard, M. Vikkula, P. Strouse, J. W. Innis and T. W. Glover: Microcephaly, intellectual impairment, bilateral vesicoureteral reflux, distichiasis, and glomuvenous malformations associated with a 16q24.3 contiguous gene deletion and a Glomulin mutation. Am J Med Genet A, 158a(4), 839-49 (2012) doi:10.1002/ajmg.a.35229
22. K. Na, E. K. Kim, W. Jang and H. S. Kim: CTNNB1 Mutations in Ovarian Microcystic Stromal Tumors: Identification of a Novel Deletion Mutation and the Use of Pyrosequencing to Identify Reported Point Mutation. Anticancer Res, 37(6), 3249-3258 (2017) doi:10.21873/anticanres.11688
23. V. Easwaran, S. H. Lee, L. Inge, L. Guo, C. Goldbeck, E. Garrett, M. Wiesmann, P. D. Garcia, J. H. Fuller, V. Chan, F. Randazzo, R. Gundel, R. S. Warren, J. Escobedo, S. L. Aukerman, R. N. Taylor and W. J. Fantl: beta-Catenin regulates vascular endothelial growth factor expression in colon cancer. Cancer Res, 63(12), 3145-53 (2003)
24. M. Grazia Lampugnani, A. Zanetti, M. Corada, T. Takahashi, G. Balconi, F. Breviario, F. Orsenigo, A. Cattelino, R. Kemler, T. O. Daniel and E. Dejana: Contact inhibition of VEGF-induced proliferation requires vascular endothelial cadherin, beta-catenin, and the phosphatase DEP-1/CD148. J Cell Biol, 161(4), 793-804 (2003) doi:10.1083/jcb.200209019
25. M. Orlandini, S. Semboloni and S. Oliviero: Beta-catenin inversely regulates vascular endothelial growth factor-D mRNA stability. J Biol Chem, 278(45), 44650-6 (2003) doi:10.1074/jbc.M304255200
26. M. Ma and N. Yu: Over-Expression of TBL1XR1 Indicates Poor Prognosis of Serous Epithelial Ovarian Cancer. Tohoku J Exp Med, 241(3), 239-247 (2017) doi:10.1620/tjem.241.239
27. A. Slavotinek, H. Pua, U. Hodoglugil, J. Abadie, J. Shieh, J. Van Ziffle, M. Kvale, H. Lee, P. Y. Kwok, N. Risch and M. Sabbadini: Pierpont syndrome associated with the p.Tyr446Cys missense mutation in TBL1XR1. Eur J Med Genet, 60(10), 504-508 (2017) doi:10.1016/j.ejmg.2017.07.003
28. H. Miraoui, A. A. Dwyer, G. P. Sykiotis, L. Plummer, W. Chung, B. Feng, A. Beenken, J. Clarke, T. H. Pers, P. Dworzynski, K. Keefe, M. Niedziela, T. Raivio, W. F. Crowley, Jr., S. B. Seminara, R. Quinton, V. A. Hughes, P. Kumanov, J. Young, M. A. Yialamas, J. E. Hall, G. Van Vliet, J. P. Chanoine, J. Rubenstein, M. Mohammadi, P. S. Tsai, Y. Sidis, K. Lage and N. Pitteloud: Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism. Am J Hum Genet, 92(5), 725-43 (2013) doi:10.1016/j.ajhg.2013.04.008

29. L. Wang, E. R. Hauser, S. H. Shah, M. A. Pericak-Vance, C. Haynes, D. Crosslin, M. Harris, S. Nelson, A. B. Hale, C. B. Granger, J. L. Haines, C. J. Jones, D. Crossman, D. Seo, S. G. Gregory, W. E. Kraus, P. J. Goldschmidt-Clermont and J. M. Vance: Peakwide mapping on chromosome 3q13 identifies the kalirin gene as a novel candidate gene for coronary artery disease. Am J Hum Genet, 80(4), 650-63 (2007) doi:10.1086/512981

30. G. Wang, W. Zhao, X. Gao, D. Zhang, Y. Li, Y. Zhang and W. Li: HNF1AAS1 promotes growth and metastasis of esophageal squamous cell carcinoma by sponging miR214 to upregulate the expression of SOX-4. Int J Oncol, 51(2), 657-667 (2017) doi:10.3892/ijo.2017.4034

31. I. Feenstra, L. E. Vissers, R. J. Pennings, W. Nillessen, R. Pfundt, H. P. Kunst, R. J. Admiraal, J. A. Veltman, C. M. van Ravenswaaij-Arts, H. G. Brunner and C. W. Cremers: Disruption of teashirt zinc finger homeobox 1 is associated with congenital aural atresia in humans. Am J Hum Genet, 89(6), 813-9 (2011) doi:10.1016/j.ajhg.2011.11.008

32. S. H. Choi, D. Ruggiero, R. Sorice, C. Song, T. Nutile, A. Vernon Smith, M. P. Concas, M. Traglia, C. Barbieri, N. C. Ndiaye, M. G. Stathopoulou, V. Lagou, G. B. Maestrale, C. Sala, S. Debette, P. Kovacs, L. Lind, J. Lamont, P. Fitzgerald, A. Tönjes, V. Gudnason, D. Toniolo, M. Pirastu, C. Bellenguez, R. S. Vasan, E. Ingelsson, A. L. Leutenegger, A. D. Johnson, A. L. DeStefano, S. Visvikis-Siest, S. Seshadri and M. Ciullo: Six Novel Loci Associated with Circulating VEGF Levels Identified by a Meta-analysis of Genome-Wide Association Studies. PLoS Genet, 12(2) (2016) doi:10.1371/journal.pgen.1005874

33. R. S. Sparkes, M. C. Sparkes, M. G. Wilson, J. W. Towner, W. Benedict, A. L. Murphree and J. J. Yunis: Regional assignment of genes for human esterase D and retinoblastoma to chromosome band 13q14. Science, 208(4447), 1042-4 (1980)

34. M. E. Askarian-Amiri, V. Seyfoddin, C. E. Smart, J. Wang, J. E. Kim, H. Hansji, B. C. Baguley, G. J. Finlay and E. Y. Leung: Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. PLoS One, 9(7), e102140 (2014) doi:10.1371/journal.pone.0102140

ABSTRACT

The fight against common chronic diseases, characterised by complex mechanisms of molecular regulation, requires the implementation of new risk prediction and prevention strategies. Personalised medicine offers sophisticated approaches for successful management of the morbidities of the ageing population. In this thesis, inspired by the principles of personalised medicine, we describe an integrative approach combining "-omics" methodologies. We use a model of a "common denominator" for cardiovascular disease (CVD) and other chronic diseases to identify biomarkers linked with common diseases risk factors and molecular pathways.

The results of this work are presented in five original publications, three review articles, and one publication under redaction. All aim to strengthen and facilitate the implementation of personalised medicine in healthcare. The three main biomarkers of our interest are TREM2, VEGF-A and telomere length (TL) and are studied mainly in the supposed healthy population of The STANISLAS family study.

With the investigation of genetic variants, located in the region of the TREM2 gene, we identified the association of SNP rs6918289 with increased levels of TNF-α and intima-media thickness of the femoral artery (IMT-F). As inflammation and increased IMT-F are important indicators of the formation of the atherosclerotic plaque, the minor allele (T) of rs6918289 might be considered a risk allele for inflammatory diseases and atherosclerosis. With the use of epigenome-wide association studies (EWAS), we identified novel epigenetic biomarkers, related to common diseases risk factors: central obesity and increased lipid levels. One methylation site (CpG) was associated with increased waist circumference (cg16170243), which could explain the epigenetic regulation of central obesity via increased insulin resistance. Moreover, an EWAS of the triglyceride (TG) levels identified two significant CpG sites, one of which was replicated in the adipose tissue (cg04580029), giving insights into epigenetic regulation of lipid levels. The results of these two last studies provided new epigenetic biomarkers and might contribute to future diagnostics and therapeutic interventions. An EWAS was also used to study the epigenetics of VEGF-A concentrations; 20 CpG sites were identified, and their relations with VEGF-A regulation were analysed through detailed bioinformatics analysis. Methylation of genes, such as TPX2 and HAS, could affect their activity and would in turn cause increased VEGF-A concentrations.

VEGF-A was further investigated for its relation with 11 cytokines, important mediators of common physiological pathways of the majority of chronic diseases. VEGF-A protein levels were associated with IL-4, MCP1 and EGF. Specific VEGF-A mRNA isoforms were also investigated for their association with cytokines; VEGF165 showed significant associations with MCP1 and IL-1α and VEGF189 with IL-4 and IL-6. Together with another important biomarker, TL, we studied the role of VEGF-A in atherosclerosis, and we identified one VEGF-A related genetic variant significantly associated with telomere attrition (calculated as a ratio between leukocyte TL and muscular TL). This genetic variant could present a common denominator of chronic diseases.

In conclusion, the employment of diverse methodologies for the investigation of common chronic diseases risk factors and pathways provided new diagnostic markers and generated results, which could help to improve the diseases risk prediction based on the individual genetic "make-up". New insights into associations between different biomarkers might help in understanding the (patho) physiological pathways common between CVDs and other chronic diseases. Finally, we hope that the results of this thesis through close collaboration with industry will facilitate the implementation of personalised healthcare.

Keywords: Biomarkers, VEGF-A, Telomeres, Personalised medicine, Epigenetics, Genetic associations

---

RÉSUMÉ

La lutte contre les maladies chroniques courantes nécessite aujourd'hui la mise en œuvre de nouvelles stratégies de prédiction du risque et de prévention. La médecine personnalisée représente une approche sophistiquée pour réussir la prise en charge des morbidités de populations vieillissantes. Dans le cadre de ces travaux de thèse inspirés par les principes de la médecine personnalisée, nous décrivons une approche intégrative qui associe plusieurs méthodologies « -omiques ». Nous avons utilisé un modèle de « dénominateur commun » pour les maladies cardiovasculaires et d'autres maladies chroniques afin d'identifier des biomarqueurs associés aux facteurs de risque et aux voies biologiques de maladies courantes.

Les résultats de ces travaux sont présentés dans 5 publications originales, 3 articles de revue et 1 publication originale en cours de redaction. Trois principaux biomarqueurs ont été étudiés, TREM2, VEGF-A et la longueur des télomères (LT), et ce dans la population d'individus en bonne santé apparente de l'étude familiale STANISLAS (SFS). Par l'étude de variants génétiques localisés dans la région comportant le gène TREM2, nous avons identifié une association entre le SNP rs6918289 et à la fois de taux élevés de TNF-α et une augmentation de l'épaisseur intima-média de l'artère fémorale (IMT-F). L'inflammation et l'IMT-F étant des indicateurs importants de la formation de plaque d'athérosclérose, l'allèle mineur (T) de rs6918289 pourrait être un allèle à risque pour les maladies inflammatoires et l'athérosclérose. Grâce à des études d'association panépigénomique (EWAS), nous avons identifié de nouveaux marqueurs épigénétiques liés à des facteurs de risque de maladies courantes, à l'obésité abdominale et à des taux élevés de lipides. Un site CpG de méthylation (cg16170243) était associé à une augmentation du tour de taille, contribuant à expliquer la régulation épigénétique de l'obésité abdominale par une résistance accrue à l'insuline. De plus, une étude EWAS des taux de triglycérides a permis d'identifier deux sites CpG significatifs. L'un de ces deux sites a pu être confirmé dans le tissu adipeux (cg04580029) et a permis de mieux comprendre la régulation épigénétique de l'augmentation du taux de triglycérides. Les résultats de ces deux dernières études ont donné lieu à de nouveaux biomarqueurs épigénétiques qui pourraient être utilisés dans l'avenir dans des applications diagnostiques et thérapeutiques. Une étude EWAS a également été réalisée pour décrire la régulation épigénétique des concentrations de VEGF-A. Vingt sites CpG ont pu être identifiés ainsi et leurs relations avec la régulation du VEGF-A ont été examinées par analyse bioinformatique poussée. La méthylation de gènes tels que TPX2 ou HAS pourrait avoir un effet sur leur activité et donc conduire à une augmentation de la concentration de VEGF-A. Les liens entre le VEGF-A et 11 cytokines, des médiateurs importants des voies physiopathologiques courantes impliquées dans la plupart des maladies chroniques, ont également été étudiés. Le taux de protéine VEGF-A était associé à IL-4, MCP1 et EGF. Les associations entre les cytokines et des isoformes spécifiques de l'ARNm du VEGF-A ont également été évaluées : le VEGF165 était associé de manière significative à MCP1 et IL-1α, et le VEGF189 à IL-4 et IL-6. Nous avons étudié le rôle du VEGF-A et d'un autre biomarqueur important, LT, dans l'athérosclérose. Cela a permis d'identifier un variant génétique lié au VEGF-A associé de manière significative à l'attrition des télomères (calculé sous la forme d'un rapport de LT dans les leucocytes et de LT dans les cellules musculaires). Ce variant génétique pourrait constituer un dénominateur commun pour les maladies chroniques.

En conclusion, l'utilisation de diverses méthodologies pour étudier les facteurs de risque et les voies impliquées dans des maladies chroniques courantes a permis d'identifier de nouveaux marqueurs diagnostiques et de générer des résultats qui pourraient améliorer la prédiction du risque de maladie basée sur le profil génétique de chaque individu. Les nouvelles données apportées sur les associations entre les différents biomarqueurs pourraient aider à mieux comprendre les voies physiopathologiques communes entre les maladies cardiovasculaires et d'autres maladies chroniques. Enfin, nous espérons que, grâce à une collaboration étroite avec des partenaires privés, les résultats de ces travaux de thèse faciliteront la mise en œuvre de la médecine personnalisée.

Mots clés : Biomarqueurs, VEGF-A, Télomères, Médecine personnalisée, Epigénétique, Associations génétiques