Development of risk prediction algorithms for cardiovascular diseases and Alzheimer's disease using key biomarkers : telomere length and vascular endothelial growth factor-A

Alexandros Petrelis

▶ **To cite this version:**

Alexandros Petrelis. Development of risk prediction algorithms for cardiovascular diseases and Alzheimer's disease using key biomarkers : telomere length and vascular endothelial growth factor-A. Human health and pathology. Université de Lorraine, 2022. English. NNT : 2022LORR0097 . tel-03817885

HAL Id: tel-03817885

https://hal.univ-lorraine.fr/tel-03817885v1

Submitted on 17 Oct 2022

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.

UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

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 bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

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 **Alexandros PETRELIS**

Développement d'algorithmes de prédiction de risque pour les maladies cardiovasculaires et la maladie d'Alzheimer à l'aide de biomarqueurs-clés: Longueur des Télomères et Facteur de Croissance Endothélial Vasculaire-A

29 avril 2022

Membres du jury :

Rapporteurs :	**Mme. Belgin SUSLEYICI**	**Professeur, Marmara University, Istanbul, Turkey**
	M. Nikolaos DRAKOULIS	**Professeur, National and Kapodistrian University of Athens, Athens, Greece**
Examinateurs :	**M. Bertrand RIHN**	**Professeur, Université de Lorraine, Nancy, France**
	Mme Marie-Dominique DEVIGNES	**Directrice de Recherche CNRS, LORIA, Vandœuvre lès Nancy, France**
	M. Maurizio SIMMACO	**Professeur, Sapienza University of Rome, Rome, Italy**
	Mme. Sofia SIEST	**Directrice de Recherche à l'INSERM - DR1, Université de Lorraine, Nancy, France, directeur de thèse**

EA_1122; IGE-PCV, Université de Lorraine, 30 Rue Lionnois, 54000 Nancy, France

Acknowledgements

This journey would not be possible without the multi-factor experience that I have taken from the support and guidance of many valuable people next to me.
This encouragement came from my personal environment, my family and friends, who supported me professionally and from my professional environment that I had the chance to be treated as family.

I was very lucky to have and I would like to thank:

- My supervisor and mentor Dr. Sofia Siest for her trust and guidance.
- The colleagues-members of EA_1122; IGE-PCV laboratory for their support and friendship: Maria, Brigitte, Christine, Said, Vesna, Alan, Ting, Abdel and Satish.
- The industrial partners that I have worked with: The Dassault Systèmes for being the first people from industry accepting me aboard on BioIntelligence project, PMI and the RANDOX Laboratories for the support and recognition in my thesis projects.
- Karim Oumnia and Sandrine Oumnia, as this thesis wouldn't have been completed without the understanding, support and agility from Epsilon group and the companies Digitsole and Zhortech.
- The PhD committee and the defense jury for accepting to follow-up and review my work and for their valuable advices.

My soft skills were very critical to successfully achieve my professional and my personal development until today. These skills were further developed by the love and support of my dear family and friends. This journey wouldn't have been so amazing without you!

Therefore, I would like to thank my friends and family:

- Greece
  - Athens: Maria, Maria, Marilia, Emmanuel, Peny, Dimitris, Maira, Giorgos Andreas, Michalis, Kiriakos, Donata, Giorgos, Kelly, Danai, Irene, David
  - Zakinthos: Katerina, Sakis, Filippos, Andreas, Markos
  - Kalamata: Dimitra, Eleni, Roubini

- France
    - Nancy: Lydia, Nikos, Eleni, Costantinos, Chrisa, Lazaros, Georgia, Maria, Dimitris, Marina, Antonis, Annabelle, Nico, Claudy, Marwen
    - All the members of the association Maison Francohellenic Lorraine, especially Katerina
    - Nice: Tatiana, Pantelis, Amaya, Marios, Elissavet, Alexandros, Pinelopi, Christoforos, Dimitris, Sofia, Valia, Thanos

However, as not all the important people of my life can be listed here,

I have gathered a collection of photos of them.

Travelling is feeding my soul but landscapes are empty without my favorite people who make my world shine!

to my most loved

my Mother Maria & my Wife Maria & my Sister Peny

my Children Marilia & Emmanuel

Abstract

Cardiovascular diseases (CVD) and Alzheimer's disease (AD) are common chronic diseases multifactorial and polygenic and their management could significantly benefit from personalised medicine. The vascular endothelial growth factor A (VEGF-A) and the telomere length (TL) are optimal biomarkers for use in personalised risk prediction as they are involved in common diseases pathways and have an important heritable background.

The aim of this thesis was to identify "-omics" determinants of VEGF-A and TL, to investigate their relationships with CVD and AD and to examine their usefulness in disease risk prediction. For valuably achieving the aim, an emphasis was given in the standardisation of the analytical process and the visualisation of the datasets used in the thesis projects.
A review article presented the knowledge on telomere dynamics during childhood including the factors that affect TL in embryonic life and in newborns. We emphasise the role of TL as an index of health and we propose its regular measurement.

In a next publication, we found significant associations between VEGF-A levels and its mRNA isoforms with IL-1α, IL-4, IL-6, EGF, and MCP1 in PBMCs of healthy individuals. Our results further support the important role of cytokines and growth factors in VEGF-A regulation at lymphocytes cellular extracts.

Then, we demonstrated that two polymorphisms, the rs12696304 and rs10936599 near the *TERC* gene, do not affect TL at birth (estimated by muscle TL - MTL), but are associated with higher telomere attrition (estimated by leucocytes telomere length - LTL/MTL) during lifetime. Thus, polymorphisms located on genes coding for telomere maintenance proteins may contribute to higher TL attrition and could impact the development of chronic diseases like atherosclerotic CVD.

A significant association between the VEGF-A related polymorphism rs6993770 and TL attrition has been found. The risk allele A could lead to increased haematopoiesis, thus higher cellular turnover and faster TL attrition. This polymorphism is a common determinant of VEGF-A and TL

and could be an important marker in the development of prediction tools for chronic diseases where both biomarkers are involved, like CVD.

We identified novel interactions between *APOE, LSR* and VEGF-A related polymorphisms that are predicting the risk of AD in a machine learning model with high accuracy. The interaction rs7043199*rs6993770 (VEGF-A related polymorphisms) was the strongest predictive factor, with a protective role higher even than that of the absence of *ε4 APOE* allele. This interaction was followed by the absence of the *ε4 APOE* allele and the interaction rs2375981*rs34528081(VEGF-A related polymorphisms). The results provide new insights about molecular mechanisms involved in AD development and could have an impact as biomarkers and/or treatment targets.

We have also developed an automated pipeline of standardised procedures for statistical analyses and the development of prediction binary models by using the elastic net method and a related interactive interface (R Shiny Genetic Elastic Net, RSGEN). A desktop application was developed in Power BI software (Power Medical Intelligence, PMI) to allow visualisation of the data of the populations studied in the thesis as well as preliminary analytical steps.

In conclusion, this thesis is proposing determinants of VEGF-A and TL that are linked through known and novel mechanisms with CVD and AD: IL-4, IL-6, MCP-1, EGF, IL-1α, *TERG* variants, VEGF-A-related variants (especially the rs6993770), *APOE* and *LSR* polymorphisms. A risk prediction model for AD with a satisfactory accuracy has also been developed, which includes mostly genetic variants and their interactions, thus it is a highly personalised tool. Furthermore, we have developed two informatic and statistical tools that can be used in other populations and contribute to the development of robust personalised models for chronic diseases prediction.

Résumé

Les maladies cardiovasculaires (MCV) et la maladie d'Alzheimer (MA) sont des maladies chroniques, multifactorielles et polygéniques et leur gestion pourrait fortement bénéficier de la médecine personnalisée. Le facteur de croissance endothélial vasculaire A (VEGF-A) et la longueur des télomères (TL) sont deux biomarqueurs optimaux pour la prédiction du risque personnalisée car ils sont impliqués dans les voies métaboliques des maladies courantes et ont un contenu héréditaire important.

Le but de cette thèse était d'identifier les déterminants « -omiques » du VEGF-A et de la TL, d'étudier leurs relations avec les MCV et la MA et d'étudier leurs utilisations dans la prédiction du risque pour ces maladies.

Un article de revue présente la dynamique des télomères pendant l'enfance. Nous soulignons l'utilité du TL comme indice de santé et nous proposons sa quantification régulière.

Nous avons trouvé des associations significatives entre les taux de VEGF-A et ses isoformes d'ARNm avec IL-1α, IL-4, IL-6, EGF et MCP1 dans les PBMCs d'individus sains. Nos résultats soutiennent le rôle des cytokines et des facteurs de croissance dans la régulation du VEGF-A.
Ensuite, nous avons démontré que les variants rs12696304 et rs10936599, proche du gène *TERC*, n'affectent pas la TL à la naissance (estimée par la TL des muscles squelettiques - MTL), mais sont associés à une plus grande attrition des télomères (estimée par la longueur des télomères des leucocytes - LTL/ MTL) au cours de la vie. Ainsi, ces variants pourraient avoir un impact sur le développement de maladies chroniques telles que les MCV.

Dans l'étude suivante, une association significative entre le variant lié au VEGF-A rs6993770 et l'attrition des télomères a été trouvée. L'allèle à risque A pourrait entraîner une augmentation de l'hématopoïèse, donc un renouvellement cellulaire plus élevé et une attrition plus rapide des télomères. Ce polymorphisme est un déterminant commun du VEGF-A et de la TL et pourrait être un marqueur important dans le développement d'outils de prédiction des maladies chroniques.
De plus, nous avons identifié de nouvelles interactions entre les polymorphismes liés au VEGF-A

et ceux des gènes d'*APOE* et *LSR* qui prédisent le risque de MA dans un modèle d'apprentissage automatique avec une grande précision. L'interaction rs7043199*rs6993770 (polymorphismes liés au VEGF-A) est le facteur prédictif le plus fort du modèle, avec un rôle protecteur contre la MA plus élevé même que celui de l'absence d'allèle *ε4 APOE*, suivie par l'absence de l'allèle *APOE ε4* et de l'interaction rs2375981*rs34528081(polymorphismes liés au VEGF-A). Les résultats donnent de nouvelles informations sur les mécanismes moléculaires impliqués dans la MA et pourraient avoir un impact en tant que biomarqueurs et/ou cibles thérapeutiques.

Nous avons également développé un pipeline automatisé de procédures standardisées pour les analyses statistiques et le développement de modèles binaires de prédiction avec la méthodologie Elastic Net et une interface interactive (R Shiny Genetic Elastic Net, RSGEN). Une application a été développée dans le logiciel Power BI (Power Medical Intelligence, PMI) pour permettre la visualisation des données des populations étudiées dans la thèse.

En conclusion, cette thèse propose des déterminants de VEGF-A et TL liés par des mécanismes connus et nouveaux avec les MCV et la MA: IL-4, IL-6, MCP-1, EGF, IL-1α, des variants du *TERG*, des polymorphismes liés au VEGF-A, en particulier le rs6993770, et des variants d'*APOE* et *LSR*. Un modèle de prédiction du risque de MA avec une précision satisfaisante a été développé, qui inclut en majorité des variants génétiques et leurs interactions, il s'agit donc d'un outil hautement personnalisé. De plus, nous avons développé deux outils informatiques et statistiques qui peuvent être utilisés dans d'autres populations et contribuer au développement de modèles personnalisés robustes pour la prédiction des maladies chroniques.

Contents

Acknowledgements *2*

Abstract *5*

Résumé *7*

List of abbreviations *11*

List of figures *14*

List of tables *16*

List of publications *17*

Présentation synthétique de la thèse *20*

Introduction *47*

Chronic diseases **48**

1. Cardiovascular diseases **49**

1.1 Definition and types 49

1.2 Epidemiology and cost 50

1.3 Pathophysiology 50

1.4 Major risk factors 51

2. Alzheimer's disease **51**

2.1 Definition 51

2.2 Epidemiology and costs 52

2.3 Major risk factors 52

2.4 Pathophysiology 54

3. Common diseases' pathways **55**

3.1 Inflammation 55

3.2 Angiogenesis 57

4. Biomarkers **60**

4.1 Presentation of biomarkers of interest 60

4.2 Role of selected biomarkers in chronic diseases 65

5. Personalised medicine **70**

5.1 Definition of PM 70

5.2 Challenges of PM 70

5.3 The role of "-omics" 71

5.4 Genetics of chronic diseases 72

6. Diseases risk prediction **76**

6.1. CVD risk prediction tools 77

6.2. AD risk prediction tools 78

6.3. Artificial intelligence and diseases risk prediction 78

Hypothesis and Objectives *85*

Methods and Materials *87*

Populations **88**
The STANISLAS Family Study 88
The TELARTA study 88
Case-control AD population 89
DNA extraction **89**
Genotyping **90**
Peripheral blood mononuclear cells (PBMCs) isolation **90**
Gene expression / mRNA quantification **91**
Total protein extraction from PBMCs **91**
Protein quantification **91**
Telomere length measurement **92**
Statistical analysis **92**
Development of an analytical pipeline for the diseases risk prediction model **95**
Development of a user-friendly application for data visualisation **95**
Results ***97***
Publication 1 ..... **98**
Publication 2 ..... **115**
Publication 3 ..... **140**
Publication 4 ..... **151**
Publication 5 ..... **165**
Publication 6 ..... **182**
General discussion, Conclusion and Perspectives ***196***
General discussion **197**
Perspectives **200**
References ***202***
Other publications ***214***
Annex1 ..... ***248***
RSGEN script **249**
PMI script **272**

List of abbreviations

AA	Apprentissage automatique
AD	Alzheimer's disease
ABCA7	ATP Binding Cassette Subfamily A Member 7
Aβ	Amyloid β
AI	Artificial intelligence
AMI	Acute myocardial infarction
AP-1	Activator protein-1
APP	Amyloid precursor protein
ApoE	Apolipoprotein E
α- syn	α- synuclein ()
ASCVD	Atherosclerotic cardiovascular disease
AUC	Area Under the Curve
B2M	Beta-2-Microglobulin
BACE1	β-secretase
BBB	Blood–brain barrier
BI	Business intelligence
BIN1	Bridging Integrator 1
BMI	Body mass index
BRC	Biological resources centre
BTL	Blood telomere length
C6orf223	Chromosome 6 Open Reading Frame 223
CABG	Coronary artery bypass graft
CAD	Coronary artery disease
CAM	Cell adhesion molecules
CAPE	Combined analysis of pleiotropy and epistasis
CBF	Cerebral blood flow
CCPPRB	Comité consultatif de protection des personnes dans la recherche biomédicale
CD2AP	CD2 Associated Protein
CDK5	Cyclin dependent kinase 5
CD33	Sialic acid binding Ig-like lectin 3
CD36	Scavenger receptor class B, member 3
CHD	Coronary heart disease
CHF	Chronic heart failure
CHIP	Clonal hematopoiesis of indeterminate potential
CLU	Clusterin
CMP	Centre for preventive medicine
CNIL	Commission nationale de l'informatique et des libertés
COX-2	Cyclooxygenase 2
CR1	Complement C3b/C4b Receptor 1 (Knops Blood Group)
CSF	Cerebrospinal fluid
CVD	Cardiovascular diseases
CXCR4	C-X-C Motif Chemokine Receptor 4
DNA	Deoxyribonucleic acid

EC	Endothelial cell
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EN	Elastic Net
EPHA1	EPH Receptor A1
ET-1	Endothelin-1
FOG	Friend of GATA
GWAS	Genome-wide association studies
HDL	High-density lipoprotein
HF	Heart failure
HIF	Hypoxia-inducible factor
HWE	Hardy-Weinberg equilibrium
IA	Intelligence artificielle
ICD	International Classification of Diseases
IFN	Interferon
IGE-PCV	Interactions gène-environnement en physiopathologie cardio-vasculaire
IL	Interleukin
IPA	Ingenuity Pathway Analysis
KASP	Competitive allele specific PCR
KCNV2	Potassium Voltage-Gated Channel Modifier Subfamily V Member 2
kDa	Kilodaltons
LDL	Low-density lipoprotein
LRP1	Lipoprotein receptor-related protein 1
LTL	Leukocyte telomere length
LSR	Lipolysis stimulated lipoprotein receptor
MA	Maladie d'Alzheimer
MAF	Minor allele frequency
MCP-1	Monocyte chemoattractant protein-1
MCV	Maladies cardiovasculaires
MI	Myocardial infraction
ML	Machine learning
MMP-9	Matrix metalloproteinase
mRNA	Messenger RNA
MS4A	Membrane Spanning 4-Domains A
MTL	Muscle telomere length
NFT	Neurofibrillary tangles
NK cells	Natural killer cells
NGS	Next generation sequencing
NF-Kb	Nuclear factor kappa
OBFC1	STN1 Subunit Of CST Complex
PBMCs	Peripheral blood mononuclear cell
PICALM	Phosphatidylinositol Binding Clathrin Assembly Protein
PLGF	Placental growth factor
PMI	Power Medical Intelligence
PM	Personalised medicine

PSEN	Presenilin
RAGE	Receptor for advanced glycation products
ROC	Receiver Characteristic Operator
ROS	Reactive oxygen species
RNA	Ribonucleic acid
RSGEN	R Shiny Genetic Elastic Net
RT-PCR	Reverse transcription - Polymerase chain reaction
SCORE	Systemic Coronary Risk Evaluation
SE	Standard error
SFS	STANISLAS Family Study
SNP	Single nucleotide polymorphism
SORL1	Sortilin Related Receptor 1
STAT3	Single transducer and activator of transcription 3
TDP-43	TAR DNA-binding protein-43
TELARTA	Telomeres and arterial ageing
TERC	Telomerase RNA Component
TET2	Tet methylcytosine dioxygenase 2
TJ	Tight junctions
TL	Telomere length
TNF-α	Tumour necrosis factor alpha
TLR4	Toll like receptor 4
ui	Interface utilisateur
VEGF Consortium	Vascular endothelial growth factor European Genomic Federation consortium
VEGF-A	Vascular endothelial growth factor-A
VEGFR	Vascular endothelial growth factor receptor
VLDL	Very-low-density lipoprotein
VLDLR-AS1	Very-low-density lipoprotein receptor antisense RNA 1
VSMC	Vascular smooth muscle cell
VTL	Vascular telomere length
WHO	World health organization
ZADH2	Zinc Binding Alcohol Dehydrogenase Domain Containing 2
ZFPM2	Zinc finger protein, FOG family member 2

List of figures

Figure synthétique. Représentation synthétique des résultats 40
Figure 1. A critical appraisal of amyloid-β-targeting therapies for Alzheimer disease [119] 54
Figure 2. Inflammatory-induced pathways in course of chronic disorders 55
Figure 3. The involvement of inflammation as a key inducer of CVD 56
Figure 4. The process underlying angiogenesis in atherosclerotic arterial wall 58
Figure 5. A scheme revealing the involvement of angiogenesis in AD pathology 59
Figure 6. Members of VEGF family and their receptors 61
Figure 7. The contribution of different factors to TL as a clinical biomarker..... 63
Figure 8. The architecture of *APOE* gene 64
Figure 9. A clarified diagram of the dynamic interplay among oxidative stress, TL, CAD 66
Figure 10. The major functional pathways of AD affected by APOE..... 69
Figure 11. Summary of genes that have been found to be involved in AD 73
Figure 12. Privilege of AI in PM compared to traditional care data analysis..... 79
Figure 13. Contribution of ML algorithms for the advance of medicine through integration and interpretation of data from different sources 80
Figure 14. The protocol of supervised ML algorithm generation from genotype input data 83
Figure 15. Architect of Power BI tool 96
Figure 16. *TERC* variants association with telomere attrition (LTL/MTL)..... 141
Figure 17. The coefficient plot of the EN model with the highest accuracy 167
Figure 18. Presentation of the first dashboard of the analytical interface with the criteria for customised analyses 185
Figure 19. The second dashboard of the analytical interface with the presentation of the dataset 185
Figure 20. Descriptive analysis of the dataset in the third dashboard of the analytical interface 186
Figure 21. Missing values quantification and qualification in the forth dashboard of the analytical interface..... 186

Figure 22. Genetic quality control by summarising HW testing in the fifth dashboard of the analytical interface 187
Figure 23. The histograms of the genetic data of the sixth dashboard of the analytical interface 187
Figure 24. The violin plots of the numeric data of the seventh dashboard of the analytical interface 188
Figure 25. The histograms of the categorical data of the eighth dashboard of the analytical interface 188
Figure 26. The table of coefficients from the predictive model of the ninth dashboard of the analytical interface 189
Figure 27. The table of results and the characteristics of the predictive model of the tenth dashboard of the analytical interface 189
Figure 28. Reporting generator of the eleventh dashboard of the analytical interface 190
Figure 29. Prediction application of the twelfth dashboard of the analytical interface 190
Figure 30. The first dashboard of the PMI descriptive interface with the map of available data 191
Figure 31. The second dashboard of the PMI descriptive interface with the selection of databases 192
Figure 32. The third dashboard of the PMI descriptive interface with description of cases and controls 193
Figure 33. The forth dashboard of the PMI descriptive interface with the selection of a continuous variable 193
Figure 34. The fifth dashboard of the PMI descriptive interface with the descritpive characteristics of SNPs 194
Figure 35. The sixth dashboard of the PMI descriptive interface with the correlation analyses 194
Figure of synthesis. Synthetic representation of results 198

List of tables

Table 1. Distinct etiological hypothesis of AD 53
Table 2. Specific functions of VEGF-A isoforms 62
Table 3. Expression of VEGF-A in different CVD 67
Table 4. Findings of different studies showing the association between different *VEGF-A* polymorphisms and CVDs. 76
Table 5. Different risk anticipation models' performance 81
Table 6. Significant associations between VEGF-A levels and inflammatory molecules... 116
Table 7. Significant associations between *VEGF-A* mRNA isoforms and inflammatory molecules 116
Table 8. Associations between VEGF-A related genetic variants and LTL/MTL* 152
Table 9. Univariate and multiple regression analysis of VEGF-A plasma concentration with LTL, MTL and LTL/MTL. 153

List of publications

List of publications related to the thesis subject

1. V. Gorenjak, **A. M. Petrelis**, M. G. Stathopoulou S. Visvikis-Siest. Telomere length determinants in childhood. Clin Chem Lab Med. 2020 Jan 28;58(2):162-177 (Review)

2. Gorenjak V*, Vance DR*, **Petrelis AM**, Stathopoulou MG, Dadé S, Shamieh SE, Murray H, Masson C, Lamont J, Fitzgerald P, Visvikis-Siest S. Peripheral blood mononuclear cells extracts VEGF protein levels and VEGF mRNA: Associations with inflammatory molecules in a healthy population. PLoS One. 2019 Aug 16;14(8):e0220902. * equal first authors

3. Toupance S, Stathopoulou MGS, **Petrelis AM**, Gorenjak V, Labat C, Lai TP, Visvikis-Siest S, Benetos A. TERC Variants Associated with Short Leukocyte Telomeres: Implication of Higher Early Life Leukocyte Telomere Attrition as Assessed by the Blood-and-Muscle Model. Cells. 2020 May 9(6), 1360

4. Gorenjak V *, **Petrelis AM***, Stathopoulou MG**, Toupance S**, Labat C, Masson C, Murray H, Lamont J, Fitzgerald P, Benetos A***, Visvikis-SiestS***. A genetic determinant of VEGF-A levels is associated with telomere attrition. Aging (Albany NY). 2021 Oct 18;13(20):23517-23526. * equal first authors, **equal second authors, ***equal last authors

5. **Petrelis AM**, Stathopoulou MG, Kafyra M, Murray H, Masson C, Lamont J, Fitzgerald P, Dedoussis G, Yen FT, Visvikis-Siest S. VEGF-A related genetic variants protect against Alzheimer's disease. *Under revision in Aging.*

6. **Petrelis AM**, …., Visvikis-Siest S. Development of an automated analysis pipeline for diseases risk prediction models and an application for populations data visualisation. *Under preparation.*

Other publications

1. 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. 2018 Mar 28;56(4):83-86

2. 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. 2018 Dec 12;8(4):43

3. Ahluwalia TS, Prins BP, Abdollahi M, Armstrong NJ, Aslibekyan S, Bain L, Jefferis B, Baumert J, Beekman M, Ben-Shlomo Y, Bis JC, Mitchell BD, de Geus E, Delgado GE, Marek D, Eriksson J, Kajantie E, Kanoni S, Kemp JP, Lu C, Marioni RE, McLachlan S, Milaneschi Y, Nolte IM, **Petrelis AM**, Porcu E, Sabater-Lleal M, Naderi E, Seppälä I, Shah T, Singhal G, Standl M, Teumer A, Thalamuthu A, Thiering E, Trompet S, Ballantyne CM, Benjamin EJ, Casas JP, Toben C, Dedoussis G, Deelen J, Durda P, Engmann J, Feitosa MF, Grallert H, Hammarstedt A, Harris SE, Homuth G, Hottenga JJ, Jalkanen S, Jamshidi Y, Jawahar MC, Jess T, Kivimaki M, Kleber ME, Lahti J, Liu Y, Marques-Vidal P, Mellström D, Mooijaart SP, Müller-Nurasyid M, Penninx B, Revez JA, Rossing P, Räikkönen K, Sattar N, Scharnagl H, Sennblad B, Silveira A, Pourcain BS, Timpson NJ, Trollor J; CHARGE Inflammation Working Group, van Dongen J, Van Heemst D, Visvikis-Siest S, Vollenweider P, Völker U, Waldenberger M, Willemsen G, Zabaneh D, Morris RW, Arnett DK, Baune BT, Boomsma DI, Chang YC, Deary IJ, Deloukas P, Eriksson JG, Evans DM, Ferreira MA, Gaunt T, Gudnason V, Hamsten A, Heinrich J, Hingorani A, Humphries SE,

Jukema JW, Koenig W, Kumari M, Kutalik Z, Lawlor DA, Lehtimäki T, März W, Mather KA, Naitza S, Nauck M, Ohlsson C, Price JF, Raitakari O, Rice K, Sachdev PS, Slagboom E, Sørensen TIA, Spector T, Stacey D, Stathopoulou MG, Tanaka T, Wannamethee SG, Whincup P, Rotter JI, Dehghan A, Boerwinkle E, Psaty BM, Snieder H, Alizadeh BZ. Genome-wide association study of circulating interleukin 6 levels identifies novel loci. Hum Mol Genet. 2021 Apr 27;30(5):393-409.

Présentation synthétique de la thèse

Introduction

Les maladies cardiovasculaires (MCV) sont des maladies complexes et multifactorielles dont la régulation est polygénique. L'identification des variants génétiques qui expliquent la variabilité interindividuelle des MCV est importante pour la prévention, la prévision et le diagnostic personnalisés et devrait améliorer notre compréhension des processus biologiques et moléculaires impliqués dans ces maladies pour le développement de thérapies spécifiques personnalisées. En d'autres termes, l'application translationnelle de ces connaissances dans la pratique clinique conduirait à une meilleure personnalisation de la prédiction et de la prévision du risque et du traitement, ce qui aurait un grand impact sur la santé publique. Cependant, la contribution des facteurs génétiques à la prédiction du risque de MCV est encore limitée.

La sélection de biomarqueurs est un domaine très important et difficile. De multiples voies biologiques, dont l'angiogenèse, l'inflammation et la longueur des télomères sont impliquées dans le développement des maladies cardiovasculaires.

La longueur des télomères des leucocytes (LTL) est liée à une variété de maladies humaines, telles que le cancer et les MCV, y compris les facteurs de risque pour ces situations de santé comme l'obésité, l'hypertension et le diabète [1-3]. L'âge est un déterminant fondamental de la longueur des télomères, tout comme le sexe et l'origine ethnique [4]. L'accélération du raccourcissement de la LTL a été associée à l'inflammation chronique et au stress oxydatif [2]. En outre, des études antérieures ont montré que le LTL est un trait fortement héréditaire (70 %) [5]. L'étude de la LTL pourrait fournir des informations importantes sur les MCV qui pourraient être utilisées pour la prédiction du risque de MCV et donc la prévention précoce.

L'angiogenèse physiologique se produit au cours du développement du foetus et, chez l'adulte, pendant la croissance osseuse, la grossesse et la cicatrisation, ainsi que dans des contextes inflammatoires, de greffe tissulaire ou d'ischémie. Une angiogenèse insuffisante ou excessive est liée à des maladies complexes, dont les MCV et le cancer [6]. Parmi les régulateurs de l'angiogenèse, le facteur de croissance endothélial vasculaire A (VEGF-A) est le facteur angiogénique le plus puissant et est considéré comme une molécule clé pour l'angiogenèse

physiologique et pathologique [7]. Le VEGF-A est un biomarqueur unique pour le système cardiovasculaire. Il joue un rôle important dans l'athérogenèse [8]. Plusieurs études, dont celles de EA_1122 ; IGE-PCV, ont montré des associations entre les taux de VEGF-A, les isoformes transcriptomiques du VEGF-A et les polymorphismes liés au VEGF-A avec les facteurs de risque et les MCV [9-12]. Le VEGF-A est également fortement héréditaire, comme l'a démontré une étude de l'EA_1122 ; IGE-PCV [13]. Dix variants génétiques ont été identifiés dans deux études d'association pangénomique (GWAS) qui expliquent plus de 50 % de la variabilité individuelle des niveaux de VEGF-A [14, 15]. Ces déterminants génétiques ont été associés à des phénotypes intermédiaires de MCV et d'autres maladies chroniques telles que les maladies thyroïdiennes auto-immunes et la dépression [9, 12, 16] [17, 18]. Le VEGF-A est donc un biomarqueur idéal pour la prédiction du risque de MCV.

L'EA_1122 ; IGE-PCV a créé en 2015 le consortium VEGF - "Vascular endothelial growth factor European Genomic Federation"- (santoriniconference.org), réunissant, 20 partenaires internationaux visant à améliorer l'étude du VEGF-A et de ses déterminants qui affectent le risque de maladies [19].

Le VEGF-A est également considéré comme un facteur de risque pour d'autres maladies chroniques, notamment la maladie d'Alzheimer (MA). La famille des VEGF est fortement impliquée dans la régulation angiogénique, la neurogenèse et la survie neuronale [20]. La diminution des niveaux de VEGF-A dans le sérum et le liquide céphalo-rachidien est liée au risque de la MA [21, 22] mais des résultats contradictoires ont également été démontrés [23]. La MA est la forme la plus courante de démence et c'est une maladie dévastatrice pour laquelle il n'existe aucun traitement permettant de contrôler sa progression [24]. Elle est associée à l'accumulation de plaques amyloïdes neuronales et à des lésions précoces principalement dans l'hippocampe. Les facteurs de risque des MCV jouent un rôle important dans le développement de la MA et le déclin accéléré de la cognition.

Le gène de l'apolipoprotéine E (*APOE*) (isoforme *ε4*) est le facteur de risque le plus important de la MA à apparition tardive, en raison de son association avec un dépôt amyloïde accru [25] et un facteur de risque connu de MCV (isoforme *ε4*) [26], tandis que l'allèle *ε2* est le facteur de risque

d'hyperlipidémie de type III [27]. L'unité de recherche EA_1122 ; IGE-PCV a précédemment démontré des interactions épistatiques significatives des variants génétiques de *l'APOE* et du récepteur des lipoprotéines stimulé par la lipolyse (*LSR*) associées à la MA [28].

Le rôle des variants génétiques du VEGF-A et de leurs interactions épistatiques avec les polymorphismes de l'*APOE* et du récepteur des lipoprotéines stimulé par la lipolyse (*LSR*) dans le risque de maladie d'Alzheimer n'a pas encore été évalué.

Hypothèses et objectifs

Les MCV et la MA sont deux des principales maladies chroniques, qui pèsent lourdement sur la santé et la vie des individus et sur les systèmes de santé. Elles partagent certaines voies physiopathologiques communes comme l'inflammation, l'angiogenèse et les anomalies du métabolisme des lipides. Ces deux maladies ont une forte composante héréditaire et de nombreux efforts ont été déployés pour identifier les variants qui définissent génétiquement le risque de ces maladies. Bien que les études d'association pangénomiques aient permis de réaliser des progrès considérables dans la détermination de certains de ces variants, une part importante de l'héritabilité de ces maladies reste à découvrir. Par conséquent, la personnalisation du risque de maladie est encore limitée. Il est donc de plus en plus nécessaire d'identifier de nouveaux déterminants du risque de maladie qui permettront à terme d'adapter et de personnaliser la prédiction du risque. Une approche plus ciblée, comme celle des gènes candidats des biomarqueurs clés, pourrait améliorer cet objectif.

La LT et le VEGF-A sont des biomarqueurs clés qui relient différentes maladies chroniques entre elles [29-31]. Ces biomarqueurs ont une importante composante génétique et constituent des biomarqueurs optimaux pour la médecine personnalisée [13, 32]. Il est important de noter que pour le VEGF-A, nous avons identifié une grande partie de sa variabilité génétique [14, 15] et les variants qui expliquent celle-ci peuvent servir comme nouveaux déterminants pour l'évaluation du risque de maladies chroniques. Les déterminants moléculaires de ces biomarqueurs pourraient également être utilisés à cette fin. Nous avons donc émis l'hypothèse que les déterminants de ces biomarqueurs pourraient être associés aux MCV, à la MA et à leurs phénotypes intermédiaires et

pourraient être d'une grande valeur pour la prédiction du risque de et la prévention précoce des maladies.

Précisément, l'objectif de cette thèse était d'identifier les déterminants " -omiques " du VEGF-A et de la TL, d'étudier leurs relations avec les MCV et la MA et d'étudier leur utilité dans la prédiction du risque. En outre, afin de permettre la réplication, la généralisation et l'extension de nos résultats, l'accent a été mis sur la standardisation du processus analytique et la visualisation des ensembles de données qui ont été utilisés dans les projets de thèse.

Quatre objectifs ont été évalués :
Objectif 1 : Identifier les déterminants de la LTL et du VEGF-A dans la santé, les MCV et la MA
Objectif 2 : Développer des modèles prédictifs en utilisant les déterminants identifiés ci-dessus
Objectif 3 : Développer des algorithmes pour l'utilisation des modèles prédictifs précédemment développés
Objectif 4 : Développer une application pour la présentation descriptive des populations et des analyses primaires

Résultats

Publication 1

Telomere length determinants in childhood

Gorenjak V, **Petrelis AM**, Stathopoulou MG, Visvikis-Siest S.

Clin Chem Lab Med. 2020 Jan 28;58(2):162-177.

Ma contribution spécifique à cette étude a été : Recherche bibliographique et participation à la rédaction du manuscrit

Cet article de revue a pour but de présenter l'état actuel des connaissances sur la dynamique des télomères pendant l'enfance, y compris les facteurs qui affectent la TL pendant la vie embryonnaire et chez les nouveau-nés.

L'enfance est une période particulièrement importante pour la TL [33] qui détermine celle-ci tout au long de la vie et sa relation avec les maladies chroniques courantes. L'environnement peut jouer un rôle néfaste pendant la grossesse pour la TL des nouveau-nés [34-37]. Un domaine très défiant est la génétique de la TL dans l'enfance qui semble être régulée de manière différentielle par rapport aux adultes [38]. Nous soulignons l'utilité de la TL comme indice de santé et nous proposons sa quantification régulière qui pourrait avoir une importance clinique dès l'enfance.

Cet article de revue a servi comme introduction à la thèse, se concentrant sur l'un des deux biomarqueurs d'intérêt et établissant la base de connaissances pour les publications originales sur les déterminants génétiques de la TL et ses liens avec les biomarqueurs des maladies chroniques.

Publication 2

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

Gorenjak V*, Vance DR*, **Petrelis AM**, Stathopoulou MG, Dadé S, Shamieh SE, Murray H, Masson C, Lamont J, Fitzgerald P, Visvikis-Siest S.

PLoS One. 2019 Aug 16;14(8):e0220902.

* equal first authors

Ma contribution spécifique à cette étude a été : Participation à l'analyse, à la rédaction et à la révision du manuscrit

Objectif

Cette étude a été réalisée afin d'évaluer les déterminants du VEGF-A dans les cellules mononucléaires du sang périphérique (PBMCs). Les PBMCs sont des cellules immunitaires qui comprennent des populations de cellules B (~15 %), de cellules T (~70 %), de monocytes (~5 %) et de cellules tueuses naturelles (NK) (~10 %) [39]. Elles sécrètent du VEGF-A et des molécules inflammatoires [40]. Les extraits de PBMCs sont un bon modèle pour étudier l'association du VEGF-A avec d'autres molécules inflammatoires.

L'objectif spécifique de cette étude était d'examiner les associations de 11 molécules inflammatoires avec les taux d'ARNm de quatre isoformes du VEGF-A (VEGF-A_{165}, VEGF-A_{189},

VEGF-A$_{145}$ et VEGF-A$_{121}$) et avec les taux de protéine du VEGF-A dans les PBMCs, dans une population saine, permettant ainsi d'identifier les déterminants physiologiques du VEGF-A.

Populations et analyse

Des participants de l'étude STANISLAS Family Study (SFS) (n=285) ont été inclus dans l'étude. Les taux d'interleukine (IL) 1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, protéine chimioattractante monocytaire 1 (MCP-1), facteur de nécrose tumorale α (TNF-α), interféron γ (INF-γ), facteur de croissance épidermique (EGF) et VEGF-A ont été mesurés dans des extraits de PBMCs. Quatre isoformes d'ARNm du VEGF-A provenant de l'ARN total dérivé des PBMCs (VEGF-A$_{121}$, VEGF-A$_{145}$, VEGF-A$_{165}$ et VEGF-A$_{189}$) ont été quantifiées chez 110 participants.

Les taux de VEGF-A ont été transformés en logarithme pour normaliser la distribution des données. Les données avec des valeurs aberrantes ont été éliminées avant l'analyse de corrélation et de régression linéaire. Une analyse de corrélation non paramétrique (corrélation de Spearman) et des modèles de régression linéaire ont été appliqués, ajustés pour l'âge et le sexe, afin de tester les associations possibles entre les taux de protéine VEGF-A et les molécules inflammatoires. Des modèles mixtes ajustés pour la structure familiale ont également été appliqués pour corriger les éventuels biais de stratification dus à la ressemblance familiale.

Résultats

Des corrélations significatives ont été observées entre le VEGF-A et l'EGF, l'IL-1β, l'IL-8, le MCP-1 et le TNF-α. Les taux de VEGF-A étaient significativement associés à l'IL-4, au MCP-1 et à l'EGF dans les modèles de régression linéaire ajustés pour l'âge et le sexe.
Des corrélations significatives ont été observées entre le VEGF-A$_{121}$ et l'IL-4, entre le VEGF-A$_{165}$ et l'IL-4, l'IL-6, et entre le VEGF-A$_{189}$ et l'IFN-γ, l'IL-1β, l'IL-4, l'IL-6.

Le VEGF-A$_{165}$ était associé au MCP et à l'IL-1α, tandis que le VEGF-A$_{189}$ était associé à l'IL-4 et à l'IL-6 dans les modèles de régression linéaire ajustés pour l'âge et le sexe.
Les analyses des associations ont été répétées en utilisant des modèles mixtes ajustés pour la structure familiale, ce qui a donné les mêmes résultats.

Conclusions

Ces travaux ont permis de mieux comprendre les liens entre les molécules inflammatoires et le VEGF-A, contribuant ainsi à accroître nos connaissances sur la régulation de l'angiogenèse.

L'étude a révélé des associations significatives entre les taux de VEGF-A et ses isoformes d'ARNm avec IL-1α, IL-4, IL-6, EGF, et MCP1 dans les PBMCs d'individus sains. Ces résultats confirment le rôle important des cytokines et des facteurs de croissance dans la régulation du VEGF-A au niveau des extraits cellulaires des lymphocytes. Les interactions des isoformes spécifiques avec différentes molécules inflammatoires devraient être prises en compte dans les recherches sur les voies moléculaires et biologiques impliquant le VEGF-A, l'angiogenèse et le développement de maladies chroniques. Elles pourraient également être utiles pour la prédiction des maladies chroniques où l'inflammation et l'angiogenèse sont impliquées, comme les MCV, le cancer et la MA.

Publication 3

TERC variants associated with short leukocyte telomeres: implication of higher early life leukocyte telomere attrition as assessed by the blood-and-muscle model.

Toupance S, Stathopoulou MGS, **Petrelis AM**, Gorenjak V, Labat C, Lai TP, Visvikis-Siest S, Benetos A.

Cells. 2020 May 9(6), 1360

Ma contribution spécifique à cette étude a été : Participation aux analyses, à la préparation et à la révision du manuscrit

Objectif

La LT a un contenu génétique important et de multiples études GWAS ont identifié des variants fortement associés à la LTL [41]. Comme les leucocytes sont un tissu hautement prolifératif, la LTL représente la TL à un moment donné de la vie, tandis que la TL dans les tissus peu prolifératifs, comme la longueur des télomères des muscles squelettiques (MTL), représente une estimation de la TL au début de la vie. En outre, le rapport LTL/MTL représente une estimation de l'attrition de la LT tout au long de la vie. Le contexte génétique de ces deux indices de TL n'a pas été étudié auparavant. Par conséquent, cette étude visait à évaluer les associations entre les

gènes candidats pour la LTL et la MTL et le ratio LTL/MTL dans une population cas-témoin de maladie cardiovasculaire athérosclérotique (ASCVD).

Population et analyse

La population fait partie de la cohorte TELARTA (Telomere in Arterial Aging) et se composait d'une cohorte de découverte de 259 individus avec (n=131) ou sans (n=128) manifestation clinique de MCV athérosclérotique, d'une cohorte de réplication indépendante de 91 individus, recrutés dans les mêmes conditions et de 52 individus provenant d'une population grecque indépendante [42]. Au total, 402 individus ont été inclus et ont été génotypés pour quinze polymorphismes (rs11125529, rs6772228, rs12696304, rs10936599, rs7675998, rs7726159, rs2736100, rs9419958, rs9420907, rs4387287, rs3027234, rs8105767, rs412658, rs6028466 et rs755017), dont on a observé précédemment qu'ils étaient associés à la LTL dans plusieurs études GWAS [43]. Les mesures de la TL ont été effectuées par analyse de Southern plot des fragments de restriction terminaux, comme décrit précédemment [44].

Les effets directs des variants sur la LTL, la MTL et le LTL/MTL ont été testés à l'aide du logiciel PLINK dans l'hypothèse d'un modèle génétique additif. Les interactions épistatiques ont été testées à l'aide de PLINK et du package R CAPE.

Résultats

Deux polymorphismes liés au LTL (rs12696304 et rs10936599), situés sur le gène *TERC* (composant ARN de la télomérase), étaient associés au ratio LTL/MTL ($p = 0.007$ et $p = 0.037$, respectivement) mais pas au MTL ($p = 0.78$ et $p = 0.32$, respectivement). Les allèles associés à un LTL plus court étaient associés à un ratio LTL/MTL plus faible, signalant ainsi une attrition LTL plus élevée. Aucune interaction épistatique significative n'a été identifiée.

Conclusions

Nous avons démontré que les variants rs12696304 et rs10936599, qui étaient précédemment associés à un LTL court, n'affectent pas le TL à la naissance (estimé par le MTL), mais sont associés à une attrition plus élevée des télomères (estimée par le LTL/MTL) au cours de la vie. Ainsi, les polymorphismes situés sur les gènes codant pour les protéines de maintien des télomères

peuvent contribuer à une attrition plus élevée des TL et pourraient avoir un impact sur le développement de maladies chroniques comme les ASCVD. Par conséquent, cette étude propose de nouvelles perspectives dans l'association entre le LTL court et les ASCVD et les variants identifiés pourraient être testés pour leur valeur prédictive dans le développement de l'athérosclérose.

Publication 4

A genetic determinant of VEGF-A levels is associated with telomere attrition

Gorenjak V *, **Petrelis AM***, Stathopoulou MG**, Toupance S**, Labat C, Masson C, Murray H, Lamont J, Fitzgerald P, Benetos A***, Visvikis-Siest S***.

Aging (Albany NY). 2021 Oct 18;13(20):23517-23526

*premiers auteurs égaux, **deuxièmes auteurs égaux, ***derniers auteurs égaux

Ma contribution spécifique dans cette étude a été : Participation à la conception de l'idée de recherche, soutient à la rédaction du manuscrit, à la conception et à la réalisation des analyses, à l'interprétation des résultats, et à la révision du manuscrit.

Objectif

Le VEGF-A et le TL sont des biomarqueurs qui sont impliqués dans les voies pathophysiologiques des maladies chroniques, notamment l'inflammation et l'angiogenèse. Ils ont tous les deux un fort caractère héréditaire [13, 45] et une variabilité interindividuelle significative [46, 47]. Ainsi, l'identification de déterminants génétiques communs pour ces deux biomarqueurs pourrait faciliter leur utilisation dans la prédiction personnalisée des maladies chroniques. Cette étude visait à élargir les connaissances sur les voies communes qui régulent ces deux biomarqueurs en étudiant les associations génétiques entre le VEGF-A et la TL.

Population et analyse

La population de l'étude précédente (Publication 3) a été utilisée dans cette étude. Elle faisait partie de la cohorte TELARTA et était composée d'une cohorte de découverte de 259 individus avec (n=131) ou sans (n=128) manifestation clinique de MCV athérosclérotique, d'une cohorte de réplication indépendante de 91 individus, recrutés dans les mêmes conditions et de 52 individus

provenant d'une population grecque indépendante [42]. Au total, 402 individus ont été inclus et ont été génotypés pour dix polymorphismes liés au VEGF-A : rs10761741, rs10738760, rs6921438, rs7043199, rs6993770, rs4416670, rs114694170, rs34528081, rs4782371 et rs2639990.

La protéine VEGF-A a été mesurée dans des échantillons de plasma, en utilisant le Cytokine Array I sur un immunoanalyseur semi-automatique de Randox (Evidence Investigator Analyzer, Randox Laboratories Ltd., Crumlin, Royaume-Uni) et le dosage de TL a été effectué par l'analyse de Southern plot des fragments de restriction terminaux, comme décrit précédemment [44].
Les effets directs des variants génétiques liés au VEGF-A sur trois phénotypes d'intérêt (LTL, MTL et LTL/MTL) ont été testés à l'aide de modèles de régression linéaire ajustés pour l'âge et le sexe avec le logiciel PLINK et les interactions épistatiques ont été testées à l'aide de PLINK et du package R CAPE.

Résultats

Nous n'avons trouvé aucune association significative entre les variants liés au VEGF-A et les LTL et MTL. Cependant, l'allèle mineur T du polymorphisme rs6993770 était significativement associé à une augmentation du ratio LTL/MTL ($P=0.001143$, $\beta=0.0148$, $SE=0.004516$). Aucune interaction épistatique significative n'a été identifiée.
De même, il n'y avait pas d'association significative entre les concentrations plasmatiques de VEGF-A et les phénotypes de TL étudiés.

Conclusion

C'est la première étude sur les associations entre les variants génétiques liés au VEGF-A et la LTL, la MTL et l'attrition télomérique. Une association significative entre rs6993770 et l'attrition télomérique a été trouvée.
Le rs6993770 est situé dans la région intronique du gène *ZFPM2* (zinc finger protein, FOG family member 2). L'allèle mineur T de ce polymorphisme a été précédemment associé à une diminution des taux de VEGF-A [14] et également aux taux de HDL [12], la numération érythrocytaire, les taux d'IL-12 et les plaquettes [48]. Cette étude a démontré que l'allèle T avait un rôle protecteur dans l'attrition des télomères. Cette association pourrait s'expliquer par la modification de

l'expression des protéines GATA, ce qui pourrait avoir un impact direct sur l'hématopoïèse et la production de plaquettes. Le rs6993770 pourrait affecter l'activité du gène *ZFPM2*, qui affecte la régulation de l'hématopoïèse par les protéines GATA. L'allèle à risque A pourrait entraîner une augmentation de l'hématopoïèse, donc un renouvellement cellulaire plus important et une attrition plus rapide des télomères [49].

L'absence d'association entre les taux plasmatiques de VEGF-A et la LTL, la MTL ou l'attrition des télomères permet de conclure que l'effet du rs6993770 sur l'attrition des télomères n'est pas directement lié aux taux de VEGF-A et est probablement médié par d'autres mécanismes, tels que l'augmentation de l'hématopoïèse que nous proposons ici. Les rapports scientifiques précédents sur l'association entre la LTL et les niveaux de VEGF-A ont rapporté des résultats incohérents, donc d'autres études sont nécessaires pour étudier pleinement l'association de ces biomarqueurs.

En conclusion, le polymorphisme rs6993770 pourrait être un marqueur important dans le développement d'outils de prédiction pour les maladies chroniques où la TL et le VEGF-A sont tous les deux impliqués, comme les MCV.

Publication 5

VEGF-A related genetic variants protect against Alzheimer's disease

Petrelis AM, Stathopoulou MG, Kafyra M, Murray H, Masson C, Lamont J, Fitzgerald P, Dedoussis G, Yen FT, Visvikis-Siest S.

En cours de révision à Aging

Ma contribution spécifique à cette étude a été : Participation à l'élaboration de l'étude, conception et réalisation des analyses, participation à l'interprétation des données et rédaction de la première version du manuscrit.

Objectif

La MA à apparition tardive est une maladie multifactorielle avec un fort contenu génétique. Cependant, l'utilisation des polymorphismes génétiques identifiés dans les études GWAS dans la modélisation de la prédiction de la maladie est très limitée. Cette constatation est commune à la plupart des maladies chroniques et s'explique par le fait que la majorité des polymorphismes

identifiés n'expliquent qu'un faible risque relatif et une petite proportion de la contribution génétique sous-jacente. La capacité prédictive des modèles peut être améliorée par l'inclusion de variants fonctionnels, l'incorporation d'effets épistatiques et la combinaison avec des biomarqueurs non-génétiques [50]. Dans cette étude, nous avons cherché à développer un modèle de prédiction du risque de la maladie d'Alzheimer en utilisant une méthodologie d'apprentissage automatique dans une population cas-témoins de la MA, qui combinerait des nouveaux gènes candidats pour la MA et leurs interactions épistatiques avec les polymorphismes fonctionnels connus des gènes *APOE* et *LSR*.

Population et analyse

Cette étude a porté sur une population cas-témoins de la MA à apparition tardive, composée de 1078 personnes non-apparentées d'origine européenne (602 témoins et 476 cas). Le recrutement et la collecte des données de cette population ont été décrits précédemment [25]. Les données de génotypage étaient disponibles dans un sous-échantillon de la population. La taille finale de l'échantillon après contrôle de qualité pour le génotypage et les valeurs aberrantes était de 323 (180 témoins et 143 cas).

Deux polymorphismes du gène *LSR* (rs34259399 et rs916147) précédemment identifiés comme gènes candidats pour la MA ont été génotypés, ainsi que les 10 polymorphismes liés au VEGF-A (rs6921438, rs7043199, rs6993770, rs2375981, rs34528081, rs4782371, rs2639990, rs10761741, rs114694170, rs1740073) et les variants *APOE* rs429358 (Cys112Arg) et rs7412 (Arg158Cys).
Nous avons appliqué une méthode d'apprentissage automatique appelée Elastic Net (EN) [51] visant à identifier les prédicteurs les plus forts du risque de MA en combinant toutes les données de génotypage ainsi que l'âge et le sexe. Cette méthode a déjà été utilisée pour prédire le risque de la MA à partir de facteurs cliniques, de l'imagerie et des données "-omiques" [52-54] mais pas pour étudier l'effet des gènes candidats comme facteurs prédictifs.

L'outil QIAGEN Ingenuity Pathway Analysis (IPA) a été utilisé pour identifier les liens de causalité et les relations entre les gènes significatifs identifiés pour prédire le risque de MA.

Résultats

Après avoir appliqué EN aux données, nous avons développé un modèle prédictif du risque de MA avec une précision de 72% (intervalle de confiance CI= (0.6, 0.8) pour 95%). La précision ci-dessus a été calculée dans le cadre d'un modèle dont la valeur P est de 0.008922. L'aire sous la courbe (AUC) du modèle est de 81 %, ce qui indique la détection de vrais positifs (vrais cas prédits comme des cas) par rapport aux faux positifs (vrais contrôles prédits comme des cas).

La présence de 2 allèles *ε4* (homozygotes *ε4/ε4*) et l'âge était liée à une augmentation du risque de MA, comme attendu, tandis que l'absence d'allèles ε4 était lié à une diminution du risque. Deux interactions précédemment [55] identifiées entre les polymorphismes *APOE* et *LSR* sont également incluses dans le modèle.

Le modèle comprenait cinq interactions épistatiques entre les variants liés au VEGF-A, une interaction *APOE*LSR* (rs34259399) et une interaction *APOE**polymorphisme lié au VEGF-A (rs6993770) associées à une diminution du risque de MA. L'interaction rs7043199*rs6993770 a présenté le coefficient le plus élevé, il s'agit donc du facteur prédictif le plus fort du modèle, avec un rôle protecteur contre la MA supérieur à celui de l'absence de l'allèle ε4 APOE, suivi par l'absence de l'allèle *ε4 APOE* et l'interaction rs2375981*rs34528081.

Les gènes identifiés dans le modèle, *VLDLR-AS1, APOE, KCNV2, ZADH2, C6orf223, LSR, ZFPM2*, et *VEGF-A*, ont été analysés dans l'outil IPA qui a montré que ces gènes étaient couramment impliqués dans différentes maladies, dont les MCV, les maladies ophtalmiques, les lésions et anomalies de l'organisme, les maladies neurologiques et le cancer figurant dans les cinq premières positions.

Conclusion

Cette étude a identifié de nouvelles interactions épistatiques entre les polymorphismes liés au VEGF-A et les variants de l'*APOE* et du *LSR* qui permettent de prédire le risque de la MA. La précision du modèle de prédiction est de 72 % et est comparable à celle d'autres outils [56-58]. Il s'agit d'une précision très élevée pour un modèle qui utilise principalement des facteurs génétiques, car d'autres études ont montré que l'ajout de variables génétiques entraîne généralement une faible

amélioration de la prédiction des modèles de variables classiques [59]. Une analyse récente de 61 articles décrivant des modèles de risque de démence a montré que la plupart d'entre eux avaient une capacité de prédiction modérée à élevée (AUC>0.70). La valeur la plus élevée de l'AUC était de 0.932 [60] et notre AUC était de 81 %. Ainsi, ce modèle peut être très utile et pourrait être testé sur de grandes populations. Les résultats fournissent également de nouvelles informations sur les mécanismes moléculaires impliqués dans le développement de la MA qui pourraient avoir un impact en tant que biomarqueurs et/ou cibles thérapeutiques pour cette maladie. Nous proposons deux interactions épistatiques (rs7043199*rs6993770 et rs2375981*rs34528081) entre les polymorphismes liés au VEGF-A comme facteurs de protection contre la MA. Nos résultats confirment l'hypothèse selon laquelle les déterminants génétiques des biomarqueurs sélectionnés (VEGF-A, *LSR* et *APOE*) peuvent agir comme des liens communs entre les maladies chroniques les plus importantes tels que les MCV, le cancer et les maladies neurologiques comme la MA et peuvent jouer un rôle important dans la prédiction de ces maladies chroniques.

Publication 6

Development of an automated analysis pipeline for diseases risk prediction models and an application for populations data visualisation.

Petrelis AM, …., Visvikis-Siest S.

En préparation

Ma contribution spécifique à cette étude a été : Participation à l'élaboration de l'étude, conception et réalisation des analyses, participation à l'interprétation des données, rédaction de la première version du manuscrit.

Introduction

La prédiction de risque de maladies chroniques est un objectif important pour permettre la prévention précoce des maladies et leur gestion efficace [61]. Le développement des modèles de prédiction de risque et d'algorithmes facilite cette tâche et ces modèles peuvent être des outils essentiels pour les décisions médicales [62]. Au cours des dernières décennies, les progrès de la génétique humaine et de la génomique permettent également la personnalisation des modèles de prédiction des risques avec l'inclusion de variables génétiques ainsi que des données cliniques

dans le cadre du concept de la médecine génomique [63]. Ces modèles peuvent aider les cliniciens à prendre des décisions médicales et préventives bien informées et à personnaliser et adapter les traitements [64], mais ils doivent être compréhensibles et faciles à utiliser afin d'être applicables dans la pratique clinique.

Une amélioration majeure dans le développement des modèles de prédiction de risque est l'application de l'intelligence artificielle (IA), en particulier des méthodologies d'apprentissage automatique (AA), car elles comprennent un pouvoir évolutif pour personnaliser la prédiction de risque avec de multiples avantages [65, 66].

De plus, la quantité croissante de données utilisées dans la médecine personnalisée et en particulier les études génétiques et génomiques, nécessite la mise en œuvre la visualisation des données et l'application des méthodes AA [67]. La visualisation interactive des données peut conduire à une interprétation réussie des données en abordant les connaissances connues et en identifiant de nouvelles connaissances grâce à la formation de nouvelles hypothèses et permettre en outre une communication large des découvertes [68, 69]. Les outils de visualisation doivent être simples à utiliser, gérer de grandes quantités de données et fournir des images très complètes et de haute qualité pour une meilleure communication [70].

Objectif

Deux outils ont été créés :

1. Un pipeline automatisé pour les analyses pour le développement d'algorithmes de prédiction de risque de maladies.
2. Une application pour la visualisation des données des populations

Le pipeline automatisé qui a été développé sur la base de la méthodologie utilisée pour la création du modèle de prédiction du risque de la AD décrit précédemment avait un double objectif :

1. Faciliter les analyses pour le développement de nouveaux modèles de prédiction dans de nouvelles populations avec des variables identiques ou différentes et pour différentes maladies
2. Permettre l'application rapide des modèles prédictifs chez de nouveaux individus afin de définir leur risque pour des maladies chroniques

Le premier objectif conduit à la conception d'un algorithme automatisé d'analyses standardisées pour les chercheurs, qui veulent développer de nouveaux modèles de prédiction pour les maladies en utilisant la méthodologie « EN-Net ». Le deuxième objectif est attendu de transformer le modèle de prédiction en un outil essentiel pour la prédiction du risque de maladies des nouveaux patients par les chercheurs et les prestataires de santé.

Le but de l'application de visualisation est de permettre une visualisation rapide des caractéristiques des différentes populations et des analyses préliminaires qui peuvent faciliter le développement de nouvelles hypothèses et études.

Population et analyse

Pour le développement du pipeline automatisé, la population de cas-témoins AD a été utilisée et les étapes de l'analyse pour le développement du modèle de prédiction de la publication 5 ont été standardisées et combinées dans une interface utilisant le paquet R Shiny du logiciel R.
L'algorithme d'analyse comprend les étapes suivantes :

1. Importation des données
2. Contrôle de qualité
3. Analyse EN

L'application de visualisation a été développée à l'aide de Microsoft Power BI. Power BI permet la connexion de différents bases de données, la visualisation et/ou la découverte de données et le partage de données [71].

Il peut être installé sur n'importe quel bureau tandis que son utilisation est très simple grâce à son format qui permet aux utilisateurs non spécialisés, comme les chercheurs et le personnel médical, de profiter de ses services. La population des publications 3 et 4 a été utilisée pour présenter l'interface.

Résultats

L'algorithme a été réalisé en langage R. Le script R de cet algorithme nécessite l'installation de R et du studio R. Ensuite, tous les packages requis seront automatiquement installés.

Les scripts ont été integrés dans une interface qui a été développée avec le paquet R Shiny, nommé RSGEN (R Shiny Genetic Elastic Net). Cette interface permet l'automatisation du modèle de prédiction identifié pour AD et un pipeline standardisé avec les étapes nécessaires au développement de nouveaux modèles de prédiction pour toute condition binaire.

Tous les scripts nécessaires de l'interface ont été combinés dans un seul fichier de script appelé app.R.
App.R a trois composants :

- Un objet d'interface utilisateur (ui72.R)
- Une fonction serveur (serv72.R)
- Un appel à la fonction shinyApp

L'objet interface utilisateur (ui) contrôle la mise en page et l'apparence de l'application. La fonction serveur contient les instructions dont l'ordinateur a besoin pour construire l'application. Enfin, la fonction shinyApp crée des objets d'application Shiny à partir d'une paire ui/serveur explicite.
Les scripts détaillés pour le développement de l'interface sont présentés en **annexe**.

L'utilisateur doit avoir R et le R Studio installés sur son ordinateur et en ouvrant le fichier app.R et en exécutant les commandes, l'interface apparaît automatiquement. Le premier tableau de bord de l'interface RSGEN permet l'importation des données et la sélection de valeurs personnalisées pour les paramètres de contrôle qualité qui ont été pris en compte dans les analyses du modèle de prédiction du risque de l'AD : Hardy Weinberg p-value, pourcentage de données manquantes par individu, pourcentage de génotypes manquants par individu et pourcentage de valeurs manquantes par variable. En outre, il permet la sélection du pourcentage pour la partition de la population pour la création aléatoire du test et former des bases de données pour l'algorithme d'apprentissage EN. Il présente un résumé de toutes les variables : catégorielles, numériques et génétiques. C'est l'onglet où la variable à prédire est spécifiée, par ex. Cas AD ou témoin. Le deuxième tableau de bord présente l'ensemble de données qui a été téléchargé. Le troisième tableau de bord génère des données descriptives pour l'ensemble de données. Le quatrième tableau de bord présente les pourcentages des valeurs manquantes et l'investigation d'un modèle potentiel de valeurs manquantes. Le cinquième tableau de bord présente un contrôle de qualité génétique en résumant les tests HW. Le sixième tableau de bord présente les caractéristiques descriptives des données

génétiques. Le septième tableau de bord présente les caractéristiques descriptives des métriques continues. Le huitième tableau de bord présente les caractéristiques descriptives des métriques catégorielles. Le neuvième tableau de bord présente un tableau des coefficients du modèle prédictif. Le dixième tableau de bord présente un tableau des résultats et des caractéristiques du modèle prédictif et une visualisation graphique des résultats du modèle prédictif. Le onzième tableau de bord présente le rapport de des résultats de tous les tableaux de bord. Ils seront automatiquement extraits et sauvegardés dans un format doc sur le dossier où les scripts sont stockés. Le douzième tableau de bord présente une application pour la prédiction de nouveaux individus basée sur les coefficients qui ont été trouvés à partir du modèle de prédiction. Un fichier d'entrée avec les données demandées de la nouvelle population sera téléchargé et les prédictions sont immédiatement présentées.

L'application de visualisation nommé PMI (Power Medical Intelligence) se compose de différents tableaux de bord qui ont été créés pour donner un aperçu des données des populations, la description des caractéristiques des populations et des variants génotypés et d'effectuer des étapes analytiques initiales comprenant le contrôle de la normalité des distributions de variables et des analyses de corrélation. Les scripts détaillés sont présentés en **annexe**.

Le premier tableau de bord présente quelques principales caractéristiques descriptives projetées sur une carte du monde. Le deuxième tableau de bord présente les bases de données téléchargées et permet la sélection de chacune d'elles. L'interface peut être utilisée pour plusieurs bases de données.

Le tableau de bord suivant s'ouvre après la sélection d'une base de données et comprend un module de sélection à gauche de l'écran et le module de visualisation avec les tracés relatifs à droite de l'écran. Les modules de sélection incluent la description des populations et des cas-témoins séparément pour les populations cas-témoins. Ensuite, il y a un module de test de normalité pour toutes les variables continues (numériques) et des modules pour les variables-clés, les caractéristiques descriptives des SNPs et enfin un module d'analyses de corrélation pour toutes les variables-clés.

Conclusion

Deux outils ont été développés pour la construction de modèles de prédiction des risques de maladies en utilisant la méthodologie AA et la visualisation des données de populations. Ces outils ne s'adressent pas nécessairement à des analystes spécialisés mais à tous les chercheurs de différents domaines et cliniciens ayant des compétences analytiques limitées, en raison de leur automatisation complète. On s'attend à ce qu'ils soient utilisés pour l'application de la personnalisation dans la prédiction des risques de maladies et la promotion de la médecine personnalisée.

Discussion générale

Dans cette thèse, nous nous sommes concentrés sur deux biomarqueurs-clés, le VEGF-A et la TL, qui sont impliqués dans de nombreuses maladies chroniques, y compris les MCV et la MA, et nous avons cherché à étudier profondément leur implication dans la variabilité des phénotypes intermédiaires et des conditions qui mènent à ces pathologies. Nous avons identifié de nouveaux déterminants du VEGF-A et du TL, révélé des liens entre ces molécules et des conditions telles que l'inflammation et l'attrition de la TL qui sont impliquées dans les MCV et la MA. En outre, nous avons démontré leur valeur dans la prédiction du risque de maladie. Nous avons utilisé différentes approches, en commençant par les gènes candidats pour étudier les effets directs et les interactions épistatiques entre les polymorphismes et la méthodologie d'apprentissage automatique pour développer un algorithme de prédiction personnalisé de la MA basé sur les variants génétiques. Nos résultats proposent de nouveaux mécanismes qui relient ces deux biomarqueurs aux maladies chroniques et fournissent des outils originaux pour la personnalisation de la prédiction du risque.

Plus précisément, nous avons démontré :

1. Par une revue des connaissances actuelles sur la dynamique de la longueur des télomères dans l'enfance, que la LT doit être utilisée comme un indice de santé régulièrement mesuré tout au long de la vie, en commençant par l'enfance.
2. Des associations entre les taux de VEGF-A et les taux d'IL-4, MCP-1 et EGF dans les extraits de PBMCs et également des associations entre les niveaux de l'ARNm VEGF-A_{165} avec

MCP-1 et IL-1α et entre les taux de VEGF-A_{189} et IL-4 et IL-6 en condition saine.

3. Des associations entre les variants génétiques liés aux LTL et l'attrition des télomères dans une population cas-témoins atteinte d'athérosclérose.
4. Des associations entre le variant rs6993770 lié au VEGF-A et l'attrition télomérique dans la même population.
5. Des interactions épistatiques entre les variants liés au VEGF-A, les polymorphismes de *l'APOE* et du *LSR* et la MA dans une population cas-témoins, en particulier les interactions prédictives protectrices rs7043199*rs6993770 et rs2375981*rs34528081 et un modèle de prédiction du risque génétique d'une précision satisfaisante.
6. Par le développement d'un algorithme, la standardisation et l'automation des analyses pour le développement des modèles de prédiction du risque des maladies.
7. Par le développement d'une application, la visualisation des données des populations et la réalisation des analyses préliminaires.

Une représentation synthétique de tous nos résultats est présentée dans la Figure synthétique suivante.

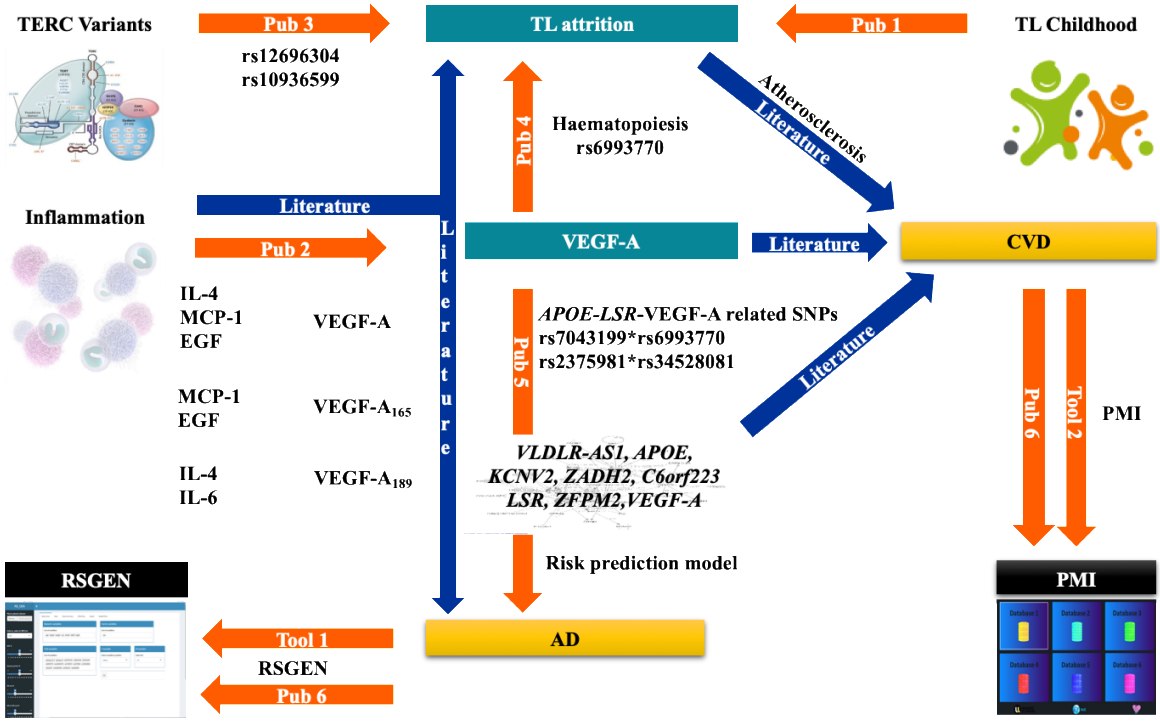

Figure synthétique. Représentation synthétique des résultats

Plus précisément, dans notre publication de revue (publication 1), nous avons présenté la dynamique de la TL dans l'enfance et proposé la TL comme un indice de santé. Dans la première publication originale (publication 2), nous avons montré des liens spécifiques entre l'inflammation, une condition connue pour être impliquée dans les MCV [72, 73] et de la MA [74], et le VEGF-A et l'attrition des télomères [32]. Nous avons utilisé les taux de protéines et les isoformes d'expression génique du VEGF-A dans les PBMCs, un type de cellules qui sécrètent les molécules de l'inflammation et jouent un rôle essentiel dans l'inflammation chronique. Les taux de l'anti-inflammatoire IL-4 et des molécules de croissance MCP-1 et EGF sont associés à une augmentation des taux de VEGF-A, tandis que l'expression de l'isoforme de l'ARNm pro-angiogénique du VEGF-A le plus abondant, le VEGF-A_{165}, est régulée par MCP-1 et EGF. L'expression du VEGF-A_{189} est régulée par l'anti-inflammatoire IL-4 et la pro-inflammatoire IL-6. Ainsi, nous avons identifié des molécules d'inflammation spécifiques comme déterminants du VEGF-A. Dans la deuxième publication originale (publication 3), nous avons montré que deux variants de *TERC*, un composant de la télomérase connu pour diminuer la LTL, sont également des déterminants d'une attrition accélérée des télomères, qui est liée à l'athérosclérose [75], aux MCV mais aussi à la MA [76, 77]. Un autre déterminant de l'attrition des télomères identifié dans notre troisième publication originale (publication 4) est le variant rs6993770, qui est un déterminant des taux de VEGF-A. Ce variant est associé à une diminution de l'attrition des télomères, probablement par une modification de l'hématopoïèse, et pourrait donc être un facteur protecteur de l'athérosclérose et des MCV mais aussi de la MA comme l'a montré notre quatrième publication originale (publication 5). Dans ce travail, nous avons démontré plusieurs interactions épistatiques entre les variants liés au VEGF-A et les gènes *APOE* et *LSR*, des gènes candidats connus pour les MCV et la MA, qui agissent comme des déterminants de protection ou de risque de la MA. Parmi elles, l'interaction la plus forte est la rs6993770* rs7043199, qui a un rôle protecteur contre la MA encore plus élevé que celui de l'absence de l'allèle *ε4 APOE*, et le rs2375981*rs34528081 avec un effet similaire à l'absence de l'allèle *ε4 APOE*. Il convient de noter que l'EA_1122 ; IGE-PCV a été parmi les premiers à identifier l'allèle 4 du gène *APOE* comme le facteur de risque génétique le plus fort pour la MA [78, 79] et nous proposons actuellement les interactions rs7043199*rs6993770 et rs2375981*rs34528081 comme étant de puissants facteurs de protection contre la MA. La première interaction implique le rs6993770, dont nous avons montré précédemment qu'il est également un déterminant du VEGF-A et de l'attrition des

télomères. Il s'agit donc d'un variant-clé dans nos recherches reliant ces deux biomarqueurs aux deux maladies chroniques étudiées. L'ApoE et le LSR sont impliqués dans le métabolisme des lipides et sont donc liés aux MCV, mais aussi à la MA [80] et nos résultats confirment ces liens. Dans la même publication, les gènes des variants impliqués dans le modèle de prédiction du risque identifié, se retrouvent couramment dans des maladies comme les MCV, les maladies neurologiques et le cancer. Ils participent à un réseau avec d'autres molécules et modulateurs qui méritent d'être étudiés pour être utilisés comme déterminants potentiels du risque des MCV et de la MA. En outre, un modèle de prédiction du risque de la MA a été proposé. Il comprend principalement des variants génétiques et leurs interactions et constitue donc un outil personnalisé, développé par une méthodologie d'apprentissage automatique. Un algorithme a été développé pour l'automatisation et la standardisation des analyses requises pour le développement d'outils de prédiction du risque par la méthodologie d'apprentissage automatique EN par le biais d'une interface automatisée appelée RSGEN (publication 6). Enfin, une application appelée PMI a été développée en Microsoft Power BI pour la visualisation des caractéristiques des populations et des données disponibles et la réalisation d'analyses préliminaires (publication 6)

Conclusion

En conclusion, cette thèse propose plusieurs déterminants du VEGF-A et de la TL qui sont liés par des mécanismes connus et nouveaux aux MCV et à la MA : IL-4, IL-6, MCP-1, EGF, IL-1α, variants de *TERG*, variants liés au VEGF-A, notamment le rs6993770, polymorphismes d'*APOE* et de *LSR*. Certains d'entre eux ont été utilisés avec succès pour développer un modèle de prédiction du risque de MA avec une précision satisfaisante. Il s'agit d'un modèle original qui inclut principalement des variants génétiques et leurs interactions pour prédire le risque de MA, ce qui en fait un outil hautement personnalisé. Il souligne également l'importance de l'étude de l'épistasie entre les variants génétiques, qui pourrait mieux rendre compte de la variabilité génétique des maladies multifactorielles et fournir une méthodologie puissante pour la prédiction du risque de maladies. Il est développé par le biais d'une méthodologie d'apprentissage automatique, qui est très puissante pour le développement de modèles de prédiction du risque, libre de restrictions méthodologiques des méthodes statistiques classiques et est actuellement la méthodologie de choix en médecine personnalisée. De plus, nous avons développé deux outils informatiques et

statistiques qui ont servi cette thèse : une application pour l'analyse descriptive des données des populations et un algorithme pour la modélisation du risque des maladies par la méthodologie EN. Ces outils peuvent être utilisés dans d'autres populations afin de permettre la visualisation et l'analyse de données similaires, de faciliter la validation et la réplication de nos résultats et de contribuer au développement de modèles personnalisés robustes pour la prédiction du risque des maladies chroniques.

Perspectives

La prochaine étape la plus importante est la validation de nos résultats dans des populations plus grandes, puis la combinaison des résultats afin de développer des outils de prédiction du risque étendus dans des populations où la disponibilité des déterminants identifiés et des données relatifs est assurée. Ces outils personnalisés de prédiction du risque devraient être testés dans de grandes cohortes cas-témoins et prospectives de patients atteints de MCV et de MA et comparés aux outils de prédiction existants afin de démontrer leur valeur ajoutée en termes d'utilité et de rentabilité. Ils devraient également être testés dans des populations d'origine différente afin de généraliser leur utilisation.

Ces ressources pourraient être trouvées dans les cadres du consortium VEGF (consortium "Vascular endothelial growth factor European Genomic Federation"-www.santoriniconference.org), qui a été créé en 2015 et coordonné par l'EA_1122 ; IGE-PCV. Il comprend 20 partenaires internationaux qui s'intéressent au VEGF-A, à la médecine personnalisée et qui ont d'accès à des populations importantes disposantes d'une grande quantité de données "-omiques" [19]. En outre, des ressources importantes peuvent également être trouvées dans les collaborations de l'EA_1122; IGE-PCV dans le cadre de la Conférence de Santorini (www.santoriniconference.org), qui est organisée par l'équipe [81].

Une contribution significative de cette thèse est le développement d'une application pour la visualisation et l'analyse descriptive des données des populations et un algorithme pour l'élaboration du modèle de prédiction du risque. L'application peut être facilement alimentée avec les nouvelles populations et peut permettre une visualisation rapide des caractéristiques spécifiques de chaque cohorte basée sur l'analyse descriptive et les graphiques associés et la sélection

d'individus selon des critères précis. Un contrôle de qualité rapide et l'exclusion d'individus et de variants sur la base de critères d'exclusion sont également effectués par l'application et les populations sélectionnées peuvent être extraites vers des bases de données. Ces bases de données peuvent ensuite être utilisées pour l'application de l'algorithme d'analyse. Cet algorithme applique la méthode du EN pour fournir le modèle de prédiction le plus précis en fonction des déterminants de chaque population. Ainsi, ces deux outils, associés à la disponibilité de grandes populations qui peuvent être fournies par le réseau scientifique de l'EA_1122 ; IGE-PCV, peuvent garantir que les perspectives ambitieuses de ce travail peuvent être réalisées et que des modèles de prédiction du risque personnalisés, validés et reproduits, basés sur les déterminants du risque identifiés dans cette thèse, seront produits et utilisés par les professionnels de la santé.

Foreword

Chronic diseases represent today one of the biggest challenges for human health and well-being. For these multifactorial diseases that have also a polygenic heritable background, personalised medicine (PM) has been proposed as a promising strategy against them. Furthermore, early prediction constitutes a major goal as it allows earlier interventions for prevention and/or therapy and thus slow development of the disease and better prognosis.

Traditionally, risk prediction tools have been developed for some chronic diseases, including cardiovascular diseases (CVD), which are based on clinical and lifestyle risk factors. The involvement of "-omics" variables in these tools has been limited and their value in increasing the accuracy of the model has been discouraging. Therefore, these tools remain highly non-tailored to the individual and for this reason a significant amount of diseases risk is still not predictable.
The reason that the application of "-omics" has been limited even though the advances in genetics and genomics of chronic diseases are important is because a large part of the heritability of the chronic diseases is still unknown.

The EA_1122; IGE-PCV research unit has a very long experience in the study of genetic epidemiology of chronic diseases, especially CVD and Alzheimer's disease (AD). Concerning the issue of missing heritability, the unit has proposed and mastered several approaches: the parallel use of genome-wide association studies (GWAS) approaches and candidate-genes, the study of epistatic and gene × environment interactions, the use of supposed healthy and case-control populations, the study of intermediate phenotypes and the research of specific biomarkers implicated in multiple diseases pathways and having a significant genetic component. The application of these approaches has led to the identification of a number of valuable determinants of risk factors and diseases phenotypes.

The unit has extensively studied the vascular endothelial growth factor A (VEGF-A) as a challenging biomarker for PM for chronic diseases. The researchers of the unit have showed that the heritable component of is more that 60%. Ten polymorphisms have been identified in two large

GWAS that explain more than half of VEGF-A variability. This is an exceptional finding, unique in the research of genetics of polygenetic biomarkers, as for most markers, the variability explained by genetic factors is around 5-10%. Numerous publications of the team have focused on these variants and have demonstrated significant associations with intermediate phenotypes of chronic diseases (cholesterol levels, blood pressure and inflammation molecules), with angiogenesis molecules, with risk for depression, and with thyroid hormones levels in patients with autoimmune thyroid diseases. The unit has also a special interest in the transcriptomic isoforms of VEG-A that are numerous and with opposite actions: anti-angiogenic and pro-angiogenic. The unit has developed and is coordinating a consortium specialised in the study of VEGF-A as a biomarker in PM, the VEGF Consortium (Vascular endothelial growth factor European Genomic Federation consortium) (santoriniconference.org) that gathers 20 international high-level scientists and institutions.

Another biomarker of interest for the EA_1122; IGE-PCV unit is the telomere length (TL). They have participated in a GWAS on TL, studied the associations of obesity with TL and focused on the genetic determinants of TL and its dynamic in childhood and in atherosclerosis. They have proposed TL as a significant biomarker for PM.

Furthermore, the unit has developed methods and infrastructures to cope with the challenges of the research field of PM, especially the need for populations and advanced analytical methods. They coordinate the Biological Resources Centre "Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire" (BRC IGE-PCV, BB-0033-00051), which is linked with a powerful analytical platform that manages the available data and can accommodate analytical pipelines to facilitate and standardise the analyses.

Given the above, the present thesis was designed to promote the knowledge on personalised disease risk prediction by using and extending the expertise of the EA_1122; IGE-PCV research team. It is proposing the use of populations of the BRC IGE-PCV and the application of modern analytical methods like machine-learning to study the "-omics" determinants of VEGF-A and TL and their role in the risk prediction of CVD and AD.

Introduction

Chronic diseases

Definition

The World Health Organization (WHO) describes chronic diseases as non-communicable conditions of long duration and usually with slow progression [82, 83]. Chronic diseases include a spectrum of physical and mental abnormalities that last over a year and result in health impairment and need constant supervision and remedy [84]. The advent of chronic diseases as a global leading issue is un-controvertible, as non-communicable conditions stand for almost two thirds of global mortality [85]. Chronic disorders include mainly CVD, stroke, cancer, chronic obstructive pulmonary disease, diabetes mellitus type 2, neurodegenerative diseases and renal failure [86].

Epidemiology and costs

The increased longevity has revealed the growing burden of chronic diseases that escalates as the population ages [87]. They are responsible for the death of 41 million people each year, equivalent to 71% of all deaths globally. Fifteen million people die from chronic diseases at ages 30-69 years and > 85% of these "premature" deaths occur in low- and middle-income countries [83]. In the latest decades, chronic diseases dominantly occur in low- and middle-income countries [88]. This can be attributed to the considerable rise in the chronic diseases risk factors in developing countries such as smoking and obesity [89] in addition to population demographics like aging and health disparities that also escalate the prevalence of chronic conditions [85]. It is note-worthy that the rising economic costs of chronic diseases impacts all wide-world countries [89]. However, developed countries have addressed chronic diseases burden exerting enormous efforts, in contrast to developing countries where the growing burden has earned insufficient consideration [89]. Evidently, the cost of chronic disorders possesses a global effect on health platforms in terms of resources and finance [90, 91]. Chronic disorders represent a huge burden in health care systems and need intuitive interference [91]. Several concerns have been raised including the enlargement of the primary-care cost, the demand for greater funding, shortage of financial resources, and the top priority for reforming the strategies of resource distribution [91]. Based on the reports of the Centers for Disease Control, chronic disorders contribute for about 75% of the gross health care costs that corresponds for approximately 5300$ per individual yearly, in the U.S. population alone [84]. Globally, chronic disorders have negatively influenced individuals' wellbeing and quality of

life in addition to magnifying healthcare expenditure at different levels [84]. Data shows that people having one chronic condition are more vulnerable for encountering other diseases [92]. Diabetes may be the perfect example as it is involved in elevating the risk of developing other chronic complications as CVD and renal failure [93, 94]. Patients with comorbidities usually suffer from medication disparity, increased need and insufficient access to health services and multi-drug regimen, that may raise the likelihood of reduced effectiveness and considerable drug interactions. These issues can reduce the quality of people's life and increase the death risk. Furthermore, it is challenging for medical stuff and the health care systems to deal with multi-morbid patients. Thus, the European health policy schedule has prioritised conducting chronic diseases research in aging population, resulting in founding of the "European Innovation Partnership on Active and Healthy Ageing" and collaborative measures on chronic disorders, in which the coexistence of chronic diseases is addressed [87].

1. Cardiovascular diseases

1.1 Definition and types

CVD that constitute the world's leading health concern affecting the heart, the brain and the vasculature, are complex multifactorial group of diseases having a broad spectrum of both environmental and genetic factors [95]. CVD encompass an extensive range of distinct conditions commonly including coronary heart disease (diseases of the blood vessels supplying the heart muscle), cerebrovascular disease (diseases of the blood vessels supplying the brain), peripheral arterial disease (disease of blood vessels supplying the arms and legs), rheumatic heart disease (damage to the heart muscle and heart valves from rheumatic fever, caused by streptococcal bacteria), congenital heart disease (malformations of heart structure existing at birth), deep vein thrombosis and pulmonary embolism (blood clots in the leg veins, which can dislodge and move to the heart and lungs) [96]. It also includes other diseases like hypertension and hypotension, heart failure, cardiac arrhythmias and others as mentioned in the group 11 “Diseases of the circulatory system” of the World Health Organisation (WHO) International Classification of Diseases ICD11 [97]. The majority of these diseases have a high heritable component [98].

1.2 Epidemiology and cost

CVD are the leading cause of global morbidity [99]. According to WHO, over 17 million people die from CVD annually, particularly, about 85% of these deaths, are due to heart attacks and stroke [96]. CVD incidents are keeping on rising steadily in low- and middle-income countries, while their prevalence in high income countries were decreasing dramatically over the last years [100]. The financial burden of CVD is also very high [101]. Based on the World Economic Forum, it is estimated that CVD costed about 863 billion dollars in 2010 at the global level and this expense was expected to escalate 22 percent by the year of 2030 [101]. The risk of developing CVD is increasing with an estimated 50% risk of encountering their complication by the age of 45 [102]. CVD incidence and age are directly proportional, with few differences among males and females such that the prevalence is greater in males at younger stages of life, however the prevalence disparities diminish considerably after menopause [102]. It is important to know that with distinct ethnic variations, there are disparities in prevalence and types of CVD [103].

1.3 Pathophysiology

A variety of etiologies may underlie CVD, entailing emboli in individuals having atrial fibrillation that leads to ischemic stroke through blocking in an artery that supplies blood to the brain and, rheumatic fever which is an inflammatory disease that triggers valvular cardiac disorder, however, the development of atherosclerosis is the leading cause of CVD pathophysiology [102]. Almost all CVD are originated from atherosclerotic basis that results in the pathogenesis of coronary artery disease (CAD), cerebrovascular disease, venous thromboembolism and, peripheral vascular disorders, with eventual development of myocardial infraction (MI), heart arrhythmias and stroke [104]. Atherosclerosis is a pathological phenomenon occurring in the intima of various arteries of medium and large sizes including the coronary arteries, cerebral arteries, iliac and femoral arteries, and aorta, particularly where the vessels ramify [105, 106]. Atherosclerosis decreases the blood flow through narrowing blood vessels and thus triggering distinct CVD [102]. Different processes are associated with atherosclerosis including dyslipidemia, immunologic incidents, inflammation, and endothelial dysfunction. Plaque formation consisting of fat, cholesterol, calcium and other substances that build up in arterial walls, represent the hallmark of atherosclerosis, that often starts in childhood and progresses with age. Eventually with time, the plaques harden, reduce the diameter of the arteries and limit the blood flow. At further stages, plaques can break and generate

thrombus or blood clot that can further block the perfusion of body organs by oxygenated blood [102]. If this occurs in one of the two main coronary arteries that nourish the heart, heart attack will result and when thrombosis occurs in one of the arteries to the brain, a stroke will happen, while if it happens in arteries of the limbs, it can lead to peripheral artery diseases. It is noteworthy that atherosclerosis itself is rarely lethal, in contrast to mobile superimposed thrombus or ruptured plaque at advanced stages that represents the direct cause of CVD incidents [107]. Strategies with ultimate goal to stop the development of atherosclerosis represents potential features for CVD prevention [102].

1.4 Major risk factors

CVD possess a wide spectrum of eminent risk factors such as hypertension, hyper-lipidemia, diabetes, obesity, smoking and sedentary lifestyle. In about all epidemiological surveys, these factors together are responsible for > 90 % of the CVD incident [104]. In a recent study, hypertension was found to be significantly associated with CVD among individuals with high body mass index (BMI). Hypertensive people have 2-3 times increased hazard of developing CVD compared to normotensive individuals. Hence, hypertension seems to be an earlier prognostic indicator for CVD in overweight subjects [108]. In addition to the previously mentioned risk factors, the World Heart Federation highlighted additional CVD risk factors like ethnicity, based on the study that shows that African or Asian ancestry are at greater risks of encountering CVD compared to other races and family history that also impose a risk for CVD where several researches have shown that major factors contributing to CVD like hypertension and obesity may have genetic basis [109].

2. Alzheimer's disease

2.1 Definition

AD, the most common cause of dementia, corresponds to a disorder of the brain characterised by mental and functional impairment along with marked neuro-pathologies. AD accounts for 50 to 75% of dementia incidents, and it is an age-related disorder whose prevalence double each 5 years in individuals over the age of 65 [110]. It was first described in 1906, by a German psychiatrist named Alois Alzheimer where he presented the case of a 51-year-old female patient, whom he

first met in 1901, suffering from mental impairment, mental confusion, delusions and other behavioral alterations [111]. There are two basic types of AD: familial or early onset AD that contribute for lower than 5% of disease incidence and sporadic or late onset AD that accounts for 79% of disease burden [112].

2.2 Epidemiology and costs

AD is considered by the WHO as a growing universal public health issue with enormous consequences on individuals and communities [110]. AD, being a major form of dementia characterised by mental decline and diminished ability to generate recent memories, is sufficient to affect everyday activities of the patient, causing dependency, disability and subsequent fatality [110, 113]. It is estimated that around 5.5 million individuals in US are impacted with AD, and that the global incidence may be over 24 million [113]. Combined datasets derived from European populations stated that the prevalence of AD in individuals > 65 years is 4.4% (6.4% for dementia) [114]. Until 2040, the incidence rate of AD is predicted to multiply by two each 20 years, considering the increased longevity in developed and developing countries [113] In addition to causing extensive drawbacks for important number of patients and families in the world's population, AD costs over 100 billion USD annually as health care expenses [115]. Such growing and demanding disorder is expected to raise tremendous obstacles to public health and geriatric care services in all nations around the globe with time [114].

2.3 Major risk factors

AD as a complex multifactorial disorder it has a broad range of risk factors with old age standing as the major risk predictor, implicating that aging-based physiological events may be involved in AD pathogenesis [114]. Moreover, the solid relation between AD and aging may partly show the aggregate impact of distinct risk and protective markers throughout life, entailing the complicated interplay among genetic, psycho-social, biological and environmental factors encountered throughout lifetime. The genetic risk hypothesis highlights that the majority of AD cases can be explained by genetic factors along with environmental factors. The apolipoprotein E (*APOE*) ε4 allele represents the strongest anticipating gene for AD [114]. People with one ε4 allele have 2-3 fold elevated chance of developing AD while those with two ε4 alleles have about 12-fold increased risk relative to people with no ε4 alleles [116]. It is noteworthy that the ε2 allele of *APOE*

has a shielding role being linked with a reduced risk for AD [116].

Furthermore, the risk of AD dementia was found to be highly elevated by CVD incidents such as strokes. Information derived from the Cardiovascular Health Study revealed that CVD were linked with elevated risk of AD, mainly in cases of peripheral arterial diseases, indicating that atherosclerosis may constitute a risk for encountering AD. The principal vulnerability and preventive factors for AD are outlined in Table 1 [114].

Table 1. Distinct etiological hypothesis of AD [114]

Etiologic hypothesis	**Contents: risk and protective factors**	**Epidemiologic evidence**
Genetic susceptibility	Risk factors: *APOE* ε4 allele and familial aggregation	Strong
Vascular pathway hypothesis	Risk factors: midlife high blood pressure and high BMI, diabetes, cerebrovascular disease, and smoking; Protective factors: light-to-moderate alcohol consumption, and antihypertensive therapy	Moderate or sufficient
Psychosocial hypothesis	Protective factors: high educational attainment, mentally stimulating activities, social activity and enriched social network, and physical activity	Moderate or sufficient
Nutritional and dietary hypothesis	Risk factor deficiency in folate, vitamin B12, and antioxidants (vitamins A, E, and C); Protective factors: Fish (omege-3 fatty acids) and vegetable consumption.	Insufficient or limited/mixed
Others (eg, toxic or inflammatory factors)	Risk factors: traumatic head injuries, occupational exposure to toxins and electromagnetic fields, depression, and hormone replacement therapy; Protective factors: nonsteroidal anti-inflammatory drugs	Insufficient or limited/mixed

2.4 Pathophysiology

Several potential pathways tend to illustrate AD pathogenesis, in which neuro-degeneration and atrophy are the major hallmarks notably in the brain's temporal and parietal lobes [117]. AD pathogenesis has been extensively linked with extra-cellular oligomers of amyloid β (Aβ) plaques and intracellular neurofibrillary tangles (NFT) that consists of hyper-phosphorylated tau τ-protein (Figure 1) [118]. Amyloid pathological process begins with modification in the cleavage pattern of the amyloid precursor protein (APP) by β-secretase (BACE1) and γ-secretase enzymes generating insoluble Aβ fibrils. This is followed by Aβ oligomerisation and intervention in signaling at synaptic gaps. The insoluble amyloid fibrils cluster to form plaques. This incidence activates kinases that results in τ protein hyper-phosphorylation and NFT formation. The deposition of plaques and tangles produces elevated oxidative stress and synaptic damage inducing the microglia to infiltrate and become activated, triggering inflammation and toxicity and neuronal cell death. Hence, the amyloidogenic pathway results in increased inflammation, oxidative stress, synaptic dysfunction, neurotransmitter deficiency and ultimately neuronal cell death contributing to AD pathology [118].

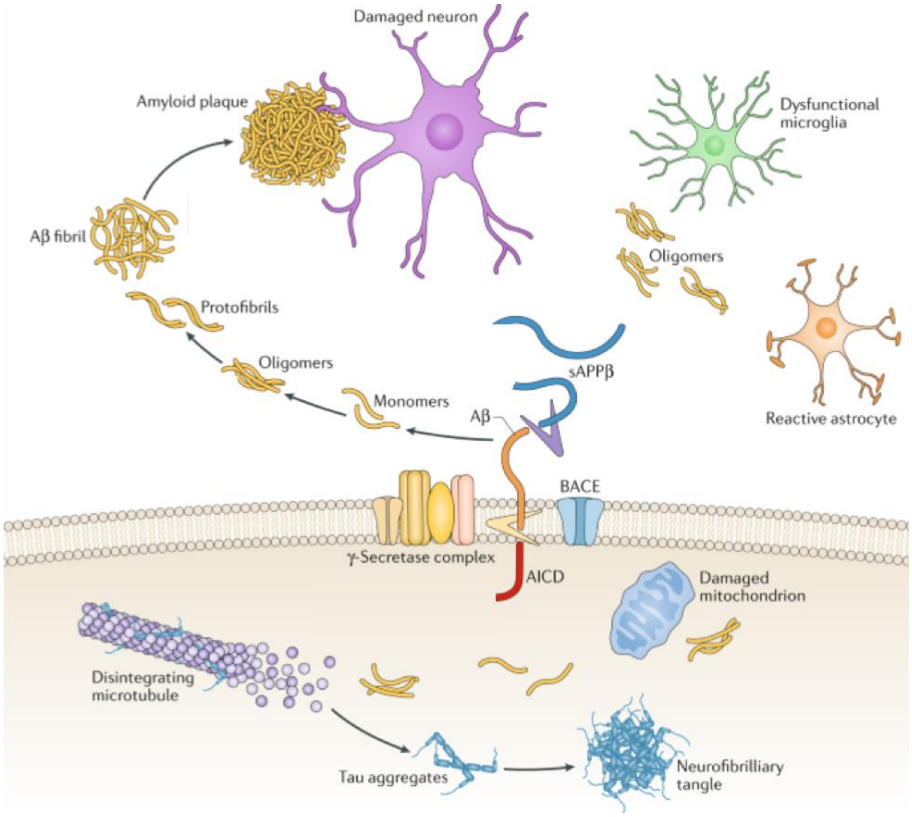

Figure 1. A critical appraisal of amyloid-β-targeting therapies for Alzheimer disease [119]

3. Common diseases' pathways

Distinct but interconnected disorders often coincide in overlapping basic pathological pathways. Detecting and analyzing the common mechanisms and factors among diseases can enhance our knowledge of such pathologies and assist in the generation of novel potential therapeutic methods [120].

3.1 Inflammation

The dominant risk factors for chronic disorders induce the disease phenotype through triggering inflammation [121]. Inflammation, the body's defense mechanism against foreign particles or stimuli aiming to maintain homeostasis, can be acute or chronic. Acute inflammatory response occurs for a brief amount of time, while chronic inflammation is long-lasting contributing to chronic diseases such as CVD and neurodegenerative disorders through interfering with many signaling pathways. In chronic diseases, multiple inflammatory biological markers are modified such as transcription factors, including the nuclear factor kappa (NF-Kb) light chain enhancer of activated B cells, the single transducer and activator of transcription 3 (STAT3) and the subsequent by-products like inflammatory cytokines tumor necrosis factor α (TNF-α), interleukins (IL-1, IL-6, IL-8) and pro-inflammatory enzymes like cyclooxygenase 2 (COX-2), matrix metalloproteinase 9 (MMP-9), cell adhesion molecules (CAM), VEGF-A and others. NF-κB is the major mediator in the course of inflammation. This transcription factor presents in a trimeric assembly of subunits p50, p65 and IκBα inhibitory subunit. After being activated by inflammatory inducements, over 400 genes implicated in inflammatory responses are activated by p50 and p65 that attach to genes' promoters (Figure 2) [121].

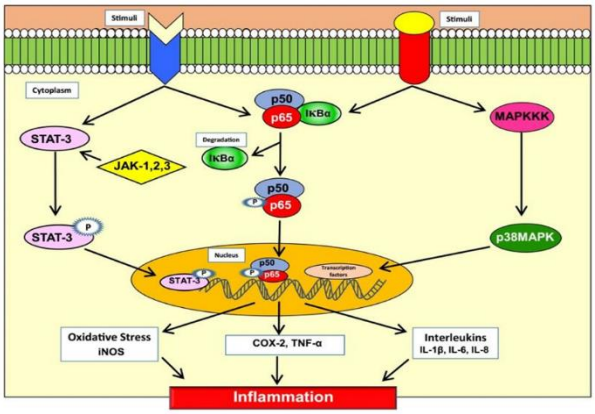

Figure 2. Inflammatory-induced pathways in course of chronic disorders [121]

3.1.1 Inflammation in CVD

Moreover, chronic inflammation, for example C-reactive protein highly increased levels, have been reported as potential hazard predictor for CVD incidences [122]. Inflammation mediates a key role in atherosclerosis and subsequent CVD through many processes (Figure 3). Foam macrophages containing oxidised low-density lipoprotein (LDL) are implicated in atherosclerosis pathogenicity via numerous inflammatory reactions. Cholesterol within macrophages trigers scavenger receptor class B, member 3 (CD36) that induces toll like receptor 4 (TLR4) and 6 in addition to inflammasome activation resulting in synthesis and release of IL-1 and IL-18. Furthermore, specific immune reactions participate in atherosclerosis pathogenesis. Th1 and Th17 pro-inflammatory T lymphocytes subtypes secretes interferon (IFN) -γ, TNF-α and IL-17 contributing to the prompt advance of atherosclerosis. Interestingly, clonal hematopoiesis of indeterminate potential (CHIP) has been linked with elevated CVD likelihood based on the evidence that the possibility of developing myocardial infarction (MI) and strokes is 2 times elevated and that the chance of developing premature atherosclerosis in young individuals escalated 4 times in comparison to controls with no CHIP. Mutations in several genes have been linked with CVD. Loss of function mutation in Tet methylcytosine dioxygenase 2 (*TET2)* gene, an example of these genes, speeds up atherosclerosis and enhances foam cell formation and secretion of inflammatory mediators [122].

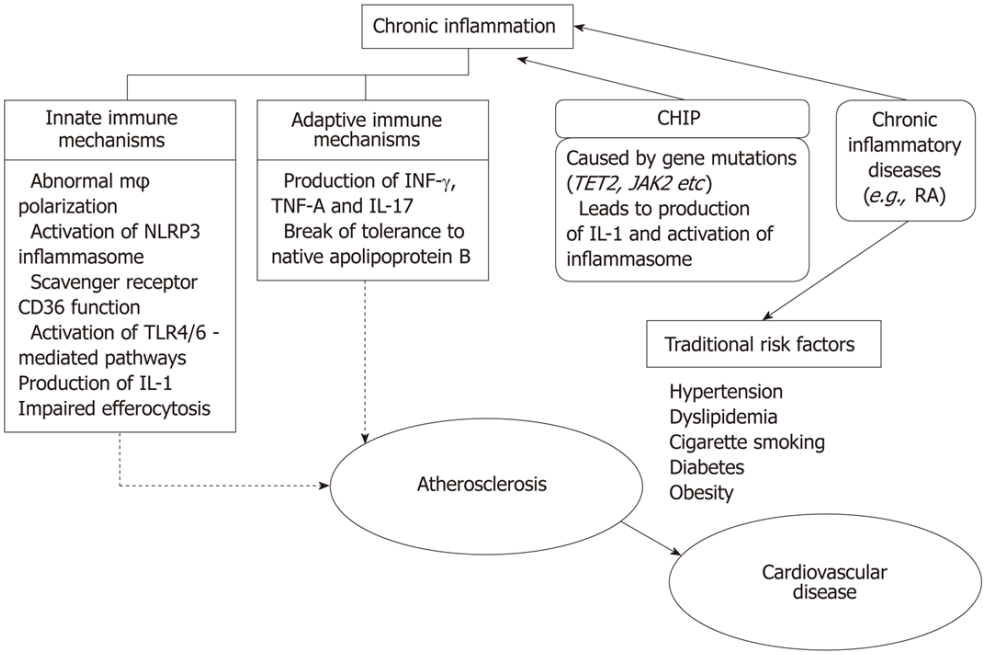

Figure 3. The involvement of inflammation as a key inducer of CVD [122]

3.1.2 Inflammation in AD

Inflammation in the brain possesses a neuroprotective function in case of acute response, that turns damaging once the chronic response is established [123]. In AD pathology, neuronal inflammation is a key contributor, being an inducer and an outcome at the same time. The formation of AD causative plaques and NFT is leastwise partially induced by aging-triggered inflammation. The occurrence of these pathological events yields further inflammation which speeds up the synthesis of more plaques and tangles and thus worsening AD hallmarks [123]. A large number of data has shown that the extracellular accumulation of Aβ in AD results in increased expression of inflammatory factors such as IL-1β, IL-6 and TNF-α via functional microglia cells, that stimulate further inflammation-related cascades through COX-2 and NF-κB [121]. It is important to note that increased amount of IL-1 increases APP synthesis and therefore consequent Aβ amount production. On the other hand, IL-6 induces the cyclin dependent kinase 5 (CDK5) activation which is a kinase that hyperphosphorylates τ-protein. Hence, inflammation seems to play a critical function in triggering AD causative pathologies [121].

3.2 Angiogenesis

The process of angiogenesis stands for the formation of new blood vessels from existing vasculature. Angiogenesis takes place in both physiological and pathological states throughout life. Given the role of angiogenesis and its contributions in distinct pathologies, enormous efforts have been exerted over the last decades to generate therapies that can induce revascularisation in disorders like ischemia or impede angiogenesis in diseases like cancer [124].

3.2.1 Angiogenesis in CVD

Additionally, in the course of CVD, neovascularisation interferes in atherosclerotic plaque development and destabilisation leading to its rupture and thus contributing to further pathogenicity [125]. Neovascularisation predominates in more inflammatory and vulnerable plaques [126]. In atherosclerosis, the related pathological hallmarks as inflammation, oxidative stress and hypoxia, trigger the production of distinct angiogenic mediators that induce formation of blood vessels from existing vasa vasorum. This process nourishes the affected area locally with oxygen and nutrients, increases macrophage infiltration, wall thickening, lipid aggregation and inflammation and ultimately induces the plaque to progress and remodel. The un-matured and frail

structure of sprouted capillaries trigger the plaque to hemorrhage which is coupled with thin cap atheroma, further foam cells infiltration and big necrotic core formation contributing to plaque vulnerability and subsequent break off [125]. The process by which angiogenesis takes place may initiate with hypoxia that elevate the level of hypoxia-inducible factor (HIF) and endothelin-1 (ET-1) and the subsequent production of VEGF-A[126]. Oxidative stress yields the same alterations in addition to the induction of MMP-2 and MMP-9 through triggering NF-κB and activator protein-1 (AP-1) expression (Figure 4) [126].

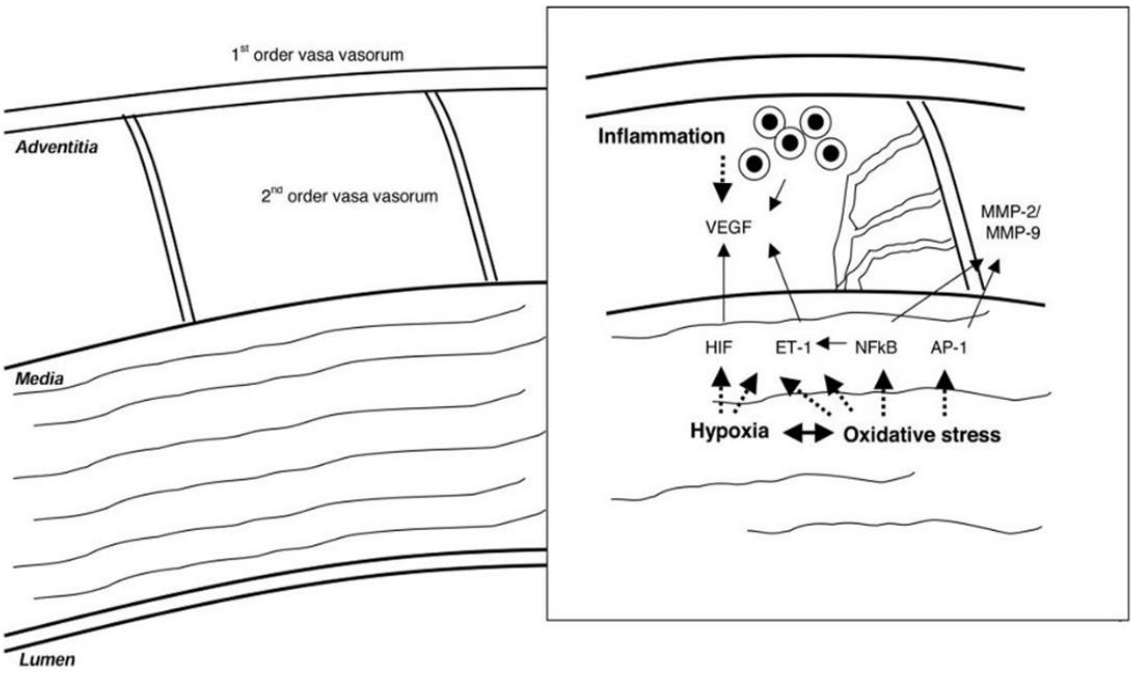

Figure 4. The process underlying angiogenesis in atherosclerotic arterial wall [126]

3.2.2 Angiogenesis in AD

Extensive researches demonstrate that, in addition to Aβ and tangles, vascular disruption has a hand in AD pathogenesis [127]. It is evident that alterations in central and peripheral vasculature affect brain perfusion contributing to AD development. This alternative hypothesis is driven by findings showing impairment in blood–brain barrier (BBB) and cerebral blood flow (CBF) (hypoperfusion) implicating that the BBB's weakened clearance pathways lead to Aβ pathogenesis in AD brains. Furthermore, it has been reported that chronic cerebral angiogenesis furnishes as an outcome of Aβ aggregation and produces BBB malfunction in AD cases. The BBB, is the brain's microvasculature which impedes influx of most substances from blood to brain, consists of

endothelial cells attached by tight junctions (TJ), astrocytes and pericytes. BBB damage may take place in response to impairment of its components. For example, the reduction of pericytes that occurs with aging can diminish the TJ protein-binding modules production and hence induce breakdown of BBB. Disruption of CBF, BBB breakdown due to TJ disturbance and impairment in Aβ regulatory peptides in the brain including receptor for advanced glycation products (RAGE) that controls the entrance of peripheral Aβ to the brain and lipoprotein receptor-related protein 1 (LRP1) that is involved in clearing Aβ from the brain-facing side of the BBB are all components that induce increased Aβ production and consecutive angiogenesis. On the other hand, the chronic inflammation triggers the production of angiogenic cytokines including VEGF-A and thrombin that additionally induce remodeling of vasculature in AD patients' brain. Moreover, angiogenesis may take place in response to hypoperfusion for compensation. Collectively, these events trigger more Aβ production by endothelial cells that intensify angiogenesis and reactive oxygen species (ROS) release, endothelium disruption and neurotoxicity (Figure 5) [127].

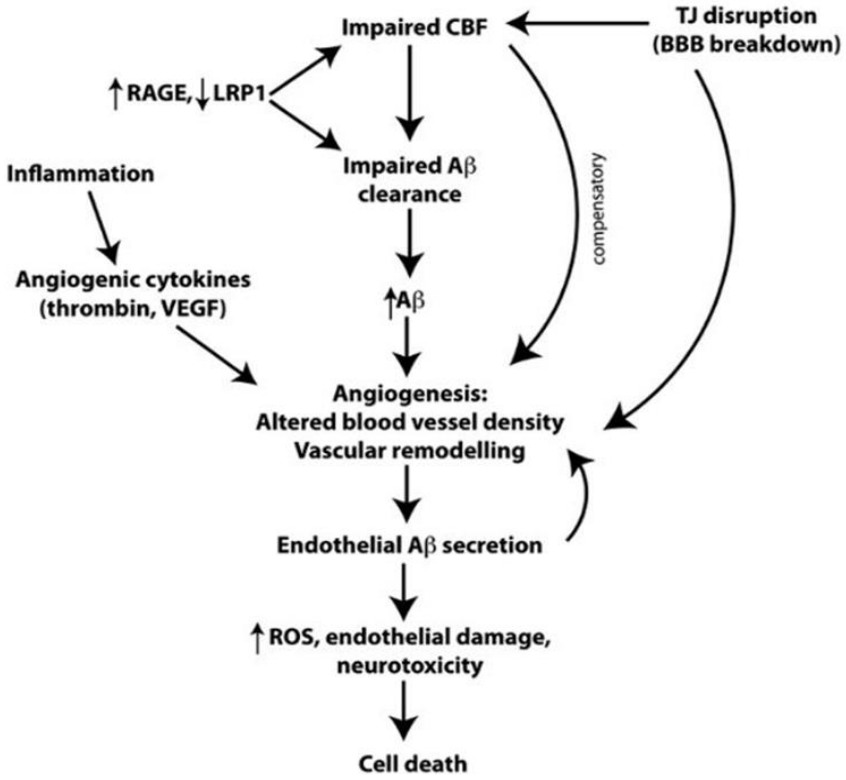

Figure 5. A scheme revealing the involvement of angiogenesis in AD pathology [127]

4. Biomarkers

The implication of biological markers or biomarkers in medical research and applications has become increasingly broadened and their existence as primary endpoints in clinical studies is approved nowadays. The National Institutes of Health Biomarkers Definitions Working Group described a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [128]. Hence, biomarkers of pathological conditions are highly critical, not only for prognosis and diagnosis but also as indicators of therapeutic effectiveness [129].

4.1 Presentation of biomarkers of interest

4.1.1 Vascular endothelial growth factor A

In humans, VEGF family, the major modulators of vascular growth, includes five distinct gene products: VEGF-A, VEGF-B and placental growth factor (PLGF), which are the main angiogenesis modulators; and VEGF-C and VEGF-D, which regulate lymphogenesis. These factors possess distinct binding compatibility for three tyrosine kinase receptors (VEGFR1, 2 and 3). Amongst these factors, the function of VEGF-A and its receptors in angiogenesis and vasculogenesis is the best comprehended [129]. VEGF-A is a dimer whose two monomers are joined via disulfide bridges [130]. In humans, *VEGF-A* gene is situated at 6p21.3 and consists of eight exons with seven introns. Like other genes involved in angiogenesis that have been shown to be alternatively spliced, VEGF-A protein possesses different subtypes due to alternative splicing of *VEGF-A* gene. The major isoforms of *VEGF-A* are: *VEGF-A*$_{111}$, *VEGF-A*$_{121}$, *VEGF-A*$_{145}$, *VEGF-A*$_{165}$, *VEGF-A*$_{183}$, *VEGF-A*$_{189}$ and *VEGF-A*$_{206}$ (the number indicates the number of amino acids of each isoform). As exons 6a and 7 determine the binding capacity of the isoforms to heparin and heparin sulphate proteoglycans on the cell surface and extracellular matrix (ECM), the splicing in these areas and the related isoforms have different binding properties. The VEGF-A$_{145}$, VEGF-A$_{189}$ and VEGF-A$_{206}$ bind strongly to both the cell surface and ECM, while VEGF-A$_{111}$ and VEGF-A$_{121,}$ (isoforms lacking exons 6 and 7) are diffusible. VEGF-A$_{165}$, which is the most abundantly expressed isoform and most potent initiator of angiogenesis, is moderately diffusible (50–70% being bound to the cell surface or ECM). The alternative splicing of exon 8 produces another family of isoforms with the same number of amino acids, but alternative sequences in the

C terminus. These isoforms are noted as *VEGF-A$_{XXX}$b* and 4 alternative isoforms have been identified: *VEGF-A$_{165}$b, VEGF-A$_{121}$b, VEGF-A$_{189}$b*, and *VEGF-A$_{145}$b*. It is generally accepted that the VEGF-A$_{xxx}$b family of isoforms are inhibitors of angiogenesis, and the VEGF-A$_{xxx}$a family promotes angiogenesis. Some of the known functions of VEGF-A isoforms are summarised in Table 2 [131]. VEGF-A is a key player in the control of blood vessels formation in physiologic and pathologic conditions particularly through its binding affinity with the tyrosine kinase receptors VEGFR-1 and VEGFR-2 [130]. VEGF-A axis is highly involved in embryonic growth via angiogenesis modulation [129]. Moreover, through activation of signal transduction pathways, it controls the majority of endothelial cellular events including proliferation, survival and migration, and microvascular permeability (Figure 6). Furthermore, VEGF-A signaling controls organs and tissues maintenance in adulthood. It triggers endothelial cells to secrete paracrine factors that facilitate tissue restoration in case of subjection to serious damage. On the other hand, VEGF-A signaling contributes greatly to the pathogenicity of disorders linked with angiogenesis. It acts as an extensive biomarker in course of diseases [129]. VEGF-A expression is affected by several inducers including hypoxia, growth factors, p53 mutation, estrogen, thyroid-stimulating hormone (TSH), cancer inducers and nitric oxide (NO) [130]. Such pathologic stimuli increase VEGF-A concentration and result in biological and chemical alterations that may activate membrane receptors, alter signaling within cells, metabolic and genetic transcription [129].

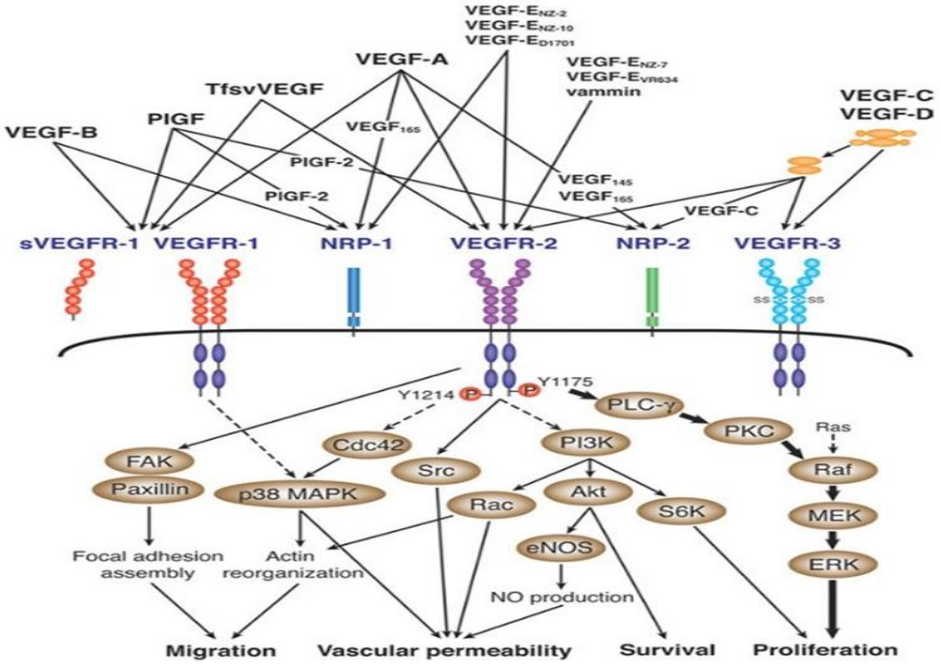

Figure 6. Members of VEGF family and their receptors [130]

Table 2. Specific functions of VEGF-A isoforms [131]

Splice Variants	Function
VEGF-A_{111}	Proangiogenic Diffusible
VEGF-A_{121}	Proangiogenic Diffusible Implicated in tumorigenesis
VEGF-A_{145}	Proangiogenic Bind to cell surface and ECM
VEGF-A_{165}	Proangiogenic Moderately diffusible Implicated in tumorigenesis
VEGF-A_{189}	Proangiogenic Implicated in tumorigenesis
VEGF-A_{206}	Proangiogenic Strongly bind to cell surface and ECM
VEGF-A_{183}	Proangiogenic
VEGF-A_{xxx}b	Anti-angiogenic Downregulated in cancer, diabetic retinopathy, Denys Drash syndrome, retinal vein occlusion Upregulated in systemic sclerosis and asthma

4.1.1. Telomere length

Telomere length (TL) represents an efficient and functional biomarker for basic aging processes, diagnosis of disorders associated with aging, and multiple pathologies' progression [29]. Telomeres, the genetic fragments of tandem repeat DNA sequences situated at the chromosomes' ends, have been known since 1930s to sustain the genomic make-up coherence [132, 133]. These structures display a mechanistic process for aging cells to point their longevity beside inducing the activation of the repair system in case of DNA damage or the senescence axis for great damage [134]. It is known that the TL reduces with each cell division and eventually, regular shortening of TL will result in genome exposure triggering cellular senescence [135]. TL is likely to be regulated by substantial synergisms between genetic and non-genetic agents such as oxidative stress and inflammatory reactions [136]. Commonly, leukocytes telomere length (LTL) is selected for conducting TL studies because of the key function of leukocytes in immune response, inflammatory reactions and cell clearance in addition to being subjected to metabolic substances as ROS [134]. Such pathological incidences like chronic inflammation and oxidative stress are known to reduce LTL rendering them as biomarkers of chronic diseases. Clinical studies on adaptive LTL offer definitive findings for LTL assessment as valid biomarker in chronic disorders (Figure 7) [134].

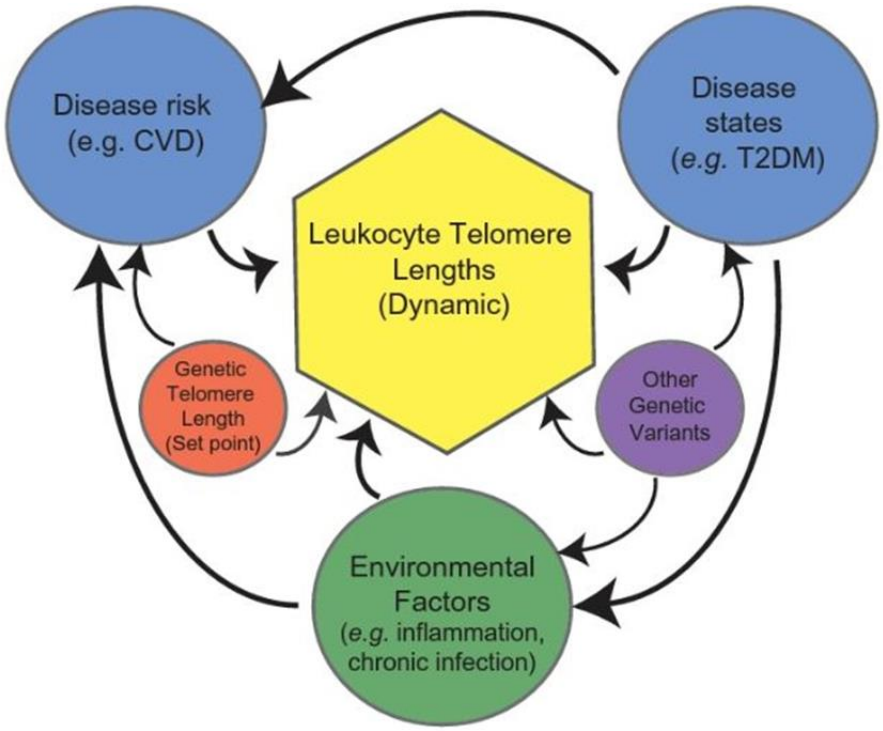

Figure 7. The contribution of different factors to TL as a clinical biomarker [134]

4.1.2 Apolipoprotein E

ApoE represents a versatile protein functioning mainly in lipid metabolism, neurology, and neurodegenerative pathologies [137]. ApoE, a 35 kDa glycoprotein of 299 amino acids, widely expressed in the liver and the brain, is encoded by *APOE* gene [138]. The ApoE domain structure is made up of two major domains, the N- and C-terminus flanking a hinge zone (Figure 8). Residues 1 to 167 form the N-terminal region including a four alpha-helix bundle with an arginine-lysine enriched portion spanning residues 135 to 150 and has the role of LDL-receptor binding domain [138]. Residues 206 to 299 form the C-terminal region with both hydrophobic and hydrophilic alpha helices characterising the family of apolipoprotein and having residues 244 to 272 making the lipid binding domain [138].

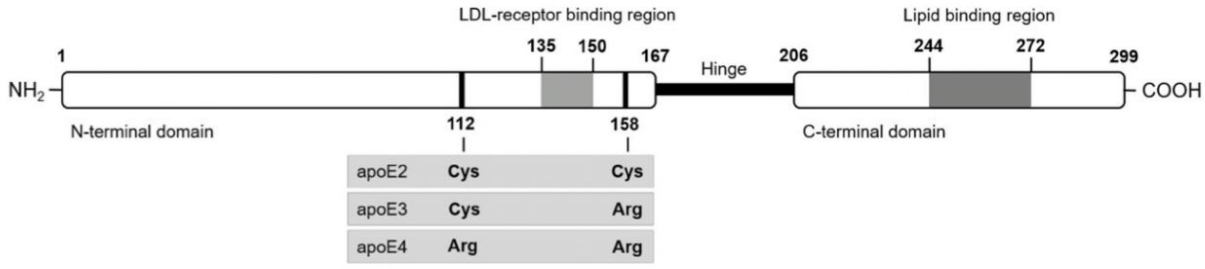

Figure 8. The architecture of *APOE* gene [138]

APOE possesses three typical isoforms *APOE2*, *APOE3*, and *APOE4* that differ only at residue 112 or 158 by one or two amino acids, however, these slight differences significantly modify its function [137]. Its main role is the clearance of remnants of triglyceride-rich lipoproteins from the circulation into the liver as it is the ligand for the LDL receptor family of proteins and heparan sulphate proteoglycans. In the brain, it is associated with assimilation and transfer of lipids [139]. The levels of ApoE are increased in conditions of axonal injuries aiming to transport lipids towards reforming nerve fibers and Schwann cells for myelin sheath formation. Additionally, within cells, ApoE can regulate distinct physiologic or pathophysiologic mechanisms, entailing alignment and cohesion of cytoskeleton, stability and activity of mitochondria, structure and role of dendrites. ApoE is broadly implicated in a spectrum of diseases that have to do with neurons repairing, remodeling and death via its interplay with distinct components in different pathways. In fact, APOE4 has been found to be related with the development or adverse clinical results of neurodegenerative disorders [137][126][126][126][126][126][126][126][126][126][126][126][126][126][126][126][126][126][126][124][124][124]

4.1.3 Lipolysis-stimulated lipoprotein receptor (LSR)

Upstream of *APOE* gene on the long arm of chromosome 19q13.1, lies the lipolysis-stimulated lipoprotein receptor (*LSR*) gene, which encrypts the LSR receptor that recognises ApoE as ligand [55]. The loci on which *LSR* and *APOE* are found have been associated with CVD intermediate phenotypes including plasma lipid concentration, dyslipidemia, adiposity and other pathologies according to genome wide association studies (GWAS) [140]. Being an ApoE receptor, LSR is engaged in the process of managing and maintaining lipid balance in the peripheral and central nervous system [141, 142]. LSR is involved in the clearance of triglyceride-rich lipoprotein from blood through binding chylomicrons, LDL and VLDL in presence of free fatty acids and allows their subsequent uptake in the cells (www.uniprot.org/uniprot/Q86X29).

4.2 Role of selected biomarkers in chronic diseases

4.2.1 Links between TL and CVD

TL has been considered a putative biomarker for CVD [143] and is associated with CVD complications [2]. LTL is linked with aging-related abnormalities and environmental factor exposures that are associated mainly to atherosclerosis and other aspects of heart and blood vessels aging [144]. A report from the Cardiovascular Health Study stated that every decreased kilo base pair of TL stands for a 3-fold elevated risk for developing MI and stroke [133]. Recently, meta-interpretive review studies have manifested that shortened LTL and cardio-metabolic complications' susceptibility are inversely proportional, where the attenuation of one standard deviation (SD) in LTL was substantially linked with a 21% escalated risk of stroke and 24% incremented risk of MI [133]. A link between family history of coronary atherosclerotic disease (CAD) and TL reduction was established in a paired case-control analysis, suggesting that the existence of shortened TL is a key defect characterising CAD [135]. Moreover, shorter TL was observed in hypertensive individuals in comparison with healthy people [135]. Chronic or ischemic cardiac failure represents additional domain of concern for TL examination in CAD [135]. Patients suffering from chronic cardiac failure possess diminished LTL compared to healthy individuals [135].

Multiple potential pathways that underlie the causative relation between shortening of TL and CVD have been suggested [143]. Firstly, TL reduction induces cell senescence and aggregation of

apoptotic cells along with lipid-laden macrophages resulting in inflammatory response and ultimately tissue remodeling, that are critical hallmarks of atherosclerosis progression. Second, TL shortening is accompanied with reduced endothelial cells proliferation ability and cell apoptosis resulting in endothelial dysfunction that furnishes as a common feature in atherosclerotic plaque formation and therefore CVD manifestation. Thirdly, important reduction in TL trigger p53-dependent mechanistic axis that triggers mitochondrial impairment and cardiac cell death. Furthermore, it was revealed recently that reduced TL can alter β-cell metabolism and insulin production, resulting in escalated oxidative stress and consequently CVD development [136, 143]. Figure 9 presents diagrammatically some links between CAD and TL.

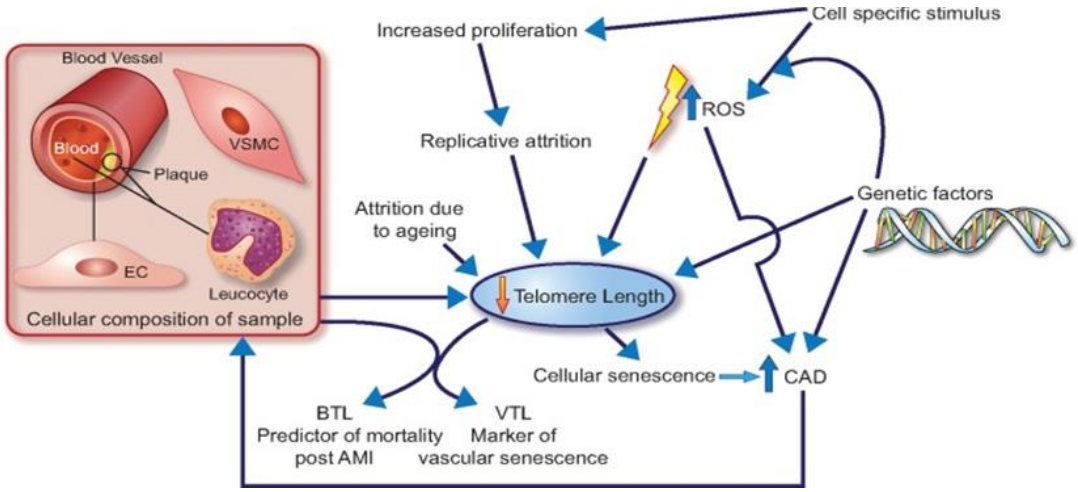

Figure 9. A clarified diagram of the dynamic interplay among oxidative stress, TL, and CAD

AMI, acute myocardial infarction; BTL, blood telomere length; EC, endothelial cell; VSMC, vascular smooth muscle cell; VTL, vascular telomere length [136].

4.2.2 Links between VEGF-A and CVD

VEGF-A is positively related with CVD [145]. Cardiomyocytes releases VEGF-A during inflammatory reactions, once subjected to mechanical stress and stimulated by cytokines [30]. Furthermore, individuals manifesting diverse CVD possess elevated levels of VEGF-A which is commonly linked with adverse complications and pathological conditions' gravity [30]. VEGF-A can trigger lipoprotein pro-atherogenic alterations [125]. In coronary arteries, VEGF-A and VEGFR are normally absent, however, they are highly detected in atherosclerotic regions, and in

particular in small capillaries endothelium, in activated monocytes and slightly differentiated smooth muscle cells [125]. VEGF-A is considered a biomarker of atherosclerosis [30]. It plays two roles in atherosclerosis. It triggers the elevation of anti-apoptotic markers expression and NO production mediating a protective role for endothelial cells. Contrastingly, VEGF-A causes damaging effects functioning as a mitotic-triggering agent through mediating endothelial cells regrowth and restoration of endothelial damage that may result in atherosclerosis development. Besides, VEGF-A enhances the permeability of endothelial, increases the expression of adhesion molecules and macrophages chemokine attractants and therefore increases monocytes infiltration through enhancing their adhesion, diapedesis and functionalisation. The impact of VEGF-A in different CVD is summarised in Table 3 [30].

Table 3. Expression of VEGF-A in different CVD [30]

Cardiovascular Diseases	**VEGF-A**	**Effects**
Ischemic heart disease: →Myocardial infarction	Increased	Microvessel density in the infarcted area, myocardial remodeling angiogenesis, neovascularisation, destabilisation of inflamed plaques
→Stroke	Increased	Vasodilation, vascular permeability, endothelial cells migration, dissociation of smooth muscle cells, loosening of the extracellular matrix.
Atherosclerosis: →Atherothrombosis	Decreased	Vascular toxicity, arterial thrombosis.
→Coronary artery disease	Increased	Angiogenesis, growth and vulnerability of plaques, intraplaque hemorrhage, inflammatory cell and erythrocyte recruitment, coronary artery revascularisation.

4.2.3 Links between VEGF-A and AD

VEGF-A is also identified as a potential biomarker for AD [21]. VEGF-A seems to have a protective role against intellectual disability, mainly in the course of AD [21, 22], however, controversial findings concerning the up- or down-regulation of *VEGF-A* gene and protein in the brain, blood, and cerebrospinal fluid (CSF) in dementia exist. VEGF-A concentration in CSF [23], plasma [31] and the medial parietal cortex [146] have been found to be increased in patients with AD in comparison to control subjects. Additionally, elevated VEGF-A concentration in the brains of AD patients was shown to be linked with pericytes reduction, BBB permeability elevation [147] and much chronic pathological tangles [148]. However, some studies have revealed that VEGF-A is downregulated in AD. Probably, the accurate link seems to rely on spatial and temporal acquiring of the sample, the tissue type and the diseases phase [149].

4.2.4 Links between ApoE and AD

It has been documented for decades that the most important predictors for AD is the *ε4* allele of *APOE* gene [78, 79, 138]. Extensive studies have stated that a key process by which *APOE ε4* increases the AD susceptibility is via triggering a greater bulk of Aβ pathology in the brains of individuals carrying *APOE ε4* [150]. ApoE may be associated with AD through its direct synergism with Aβ proteins. The existence of ApoE in Aβ plaques of AD brains has been reported so far [151]. ApoE and Aβ interaction takes place through binding of the Aβ protein residues 13 to17 with the ApoE N-terminal domain residues 144–148 and the C-terminal domain residues 244–248 [138]. Moreover, knock-out studies of *apoe* gene in mice have revealed that ApoE is critical for the initiation and retention of Aβ plaques. Furthermore, the direct synergistic link of ApoE and tau protein furnishes another theory behind the association of ApoE and AD. Previously, the existence of ApoE has been detected in tau-NFT mediated deposits [152]. Up-regulation of *APOE ε4* in neuronal cells of genetically engineered mice triggered a rise in tau hyper-phosphorylation induced by extracellular signal regulated kinase stimulation [138]. Additionally, *APOE ε4* can serve as a causative factor and has mitigated efficacy in several axes involved in synaptic integrity, brain homeostasis, involving lipid translocation, glucose catabolism and brain vasculature [150]. APOE has also been shown to be involved in TAR DNA-binding protein-43 (TDP-43) pathological hazard in the AD brains and in protein pathologies involving tau or α-synuclein (α- syn) (Figure 10) [150].

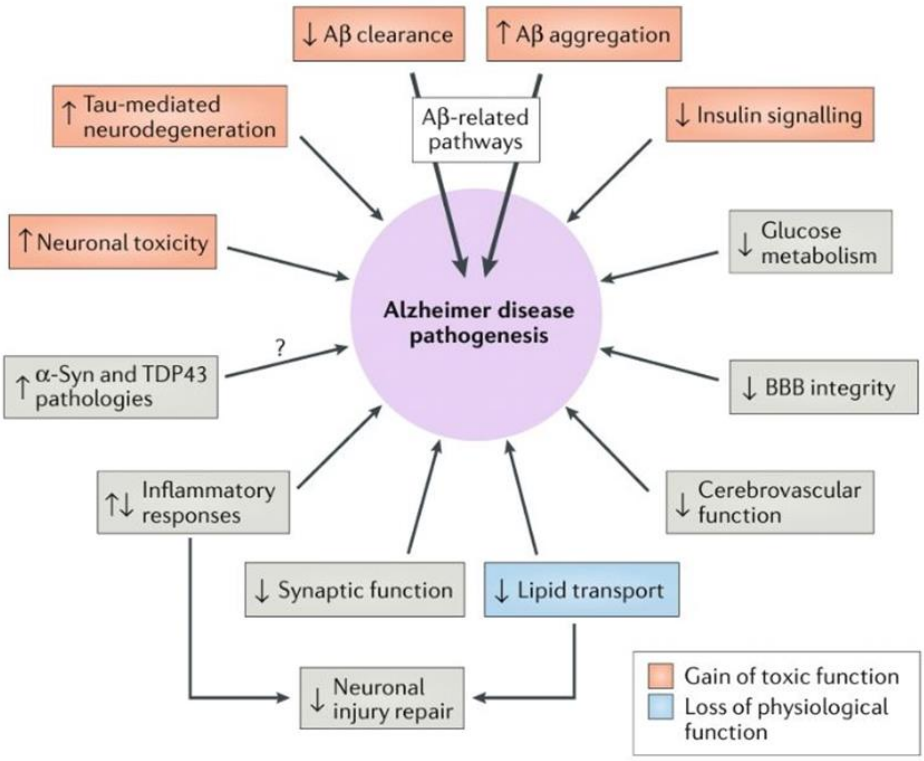

Figure 10. The major functional pathways of AD affected by APOE [150]

4.2.5 Links between LSR and AD

Many studies have highlighted the link between elevated cholesterol amount in plasma with higher AD susceptibility. Cholesterol modifies the cleavage of APP by enzymes bounded to the membrane leading to the production of Aβ. In the brain, cholesterol allocation and homeostasis maintenance are achieved by ApoE through binding to distinct lipoprotein receptors [55]. Diminished expression of LSR exhibit altered cholesterol homeostasis in the brain of aged mice coupled with mental impairment that are increased when subjected to Aβ stress [153]. Interestingly, *LSR* gene shows an epistatic interaction with *APOE* resulting in elevated AD risk [55]. The epistasis between *LSR* and *APOE* polymorphisms provides a novel genetic relation through which individuals at increased risk of developing AD can be identified in addition to paving the way of novel therapeutic strategies for AD.

5. Personalised medicine

5.1 Definition of PM

PM is an emergent medical field that invests the knowledge about an individual's own genome and its derivatives (RNA, proteins and metabolites) to prevent, predict, , diagnose, or treat a disease [154]. The concept of PM is heavily related to "-omics" such as genetics, genomics and others [154]. This approach utilises scientific insights brought on by human genome sequencing into the genetic and molecular basis of health and diseases with a general purpose of offering "the right treatment for the right patient at the right time" and thereby improving that person's health [155, 156]. In contrast to "one-size-fits-all", this tailored medical approach concentrates on the genetic make-up of the patient, inheritance, behavioral profile and environmental exposures rather than focusing on the disease phenotype [157] and is expected to play a significant role in the management of chronic diseases.

5.2 Challenges of PM

Even though PM demonstrates an extensive and expanding progress in research and innovation, it raises unique challenges [158, 159]. Ethical, social, and legal issues are raised by the initiation of PM [160, 161]. It will be crucial to figure strategies to protect participants' privacy and the confidentiality of their medical data [161]. The risks and benefits of PM research will need to be understood by participants, which means that researchers will have to develop a rigorous informed consent procedure [154]. Education in PM is crucial for its transition [154]. The shortage of the expertise of health professionals in interpreting and operating on pharmacogenetic data as well as patient perception and autonomy is one of the main challenges for PM to be broadly approved and accepted as a routine approach used by different regulatory agencies and health care stakeholders [160].

Furthermore, significant contributions need to be done by the scientific community in bioinformatics, biomathematics and biostatistics in order to speed up the shift towards PM and to improve translational interpretation of "-omics" data and make the best use of high-throughput technologies [162]. Computational and storage capacity today are limited and do not fully allow the process of "-omics" data in ordinary research laboratories [162]. Investments in informatics

support for software customisation and utilisation are advisable for laboratory-hosted servers [162]. Additionally, it is demanding and labor-intensive, but vitally crucial task, to create secure and error-free software [162]. Thus, before admitting the therapeutic utility of "-omics" information, it is necessary to confirm the accuracy of research platforms [162].

5.3 The role of "-omics"

The sequencing of the human genome together with the continuous development and implementation of novel high throughput "-omics" technologies as genomics, transcriptomics, proteomics, metabolomics and interactomics are improving our knowledge on human health and disease and have already, and will continue, to furnish a sound foundation for PM [154, 163]. In the last decade, genetics and molecular sciences have made several contributions for the advance of medicine, including the Human Genome Project, the International HapMap Project and the GWAS project (International HapMap Consortium, 2005) [159]. Single nucleotide polymorphisms (SNPs) are presently identified as the primary source of human genetic diversity and they represent an important reference for mapping complex genetic phenotypes [159]. Integrating genetic effects on phenotypes and medications enable PM to tailor therapies to specific genotype groups of patients [159]. Presently, there are various types of molecular tests that interpret the human DNA, RNA and proteins and can detect mutations, variations, metabolic profiles and genetic abnormalities that are linked with diseases and that can impact greatly the therapeutic response [154]. Thus, they can intervene in all disease stages: vulnerability, preclinical progression, assessment and diagnosis, disease progression and treatment response [154]. After all, "-omics" technologies, in particular genetics and genomics, represent the pillars of the personalised approach of medicine [155].

In the "-omics" realm, however, there is plenty to be acquired from its constant progression [155]. The Human Genome Project has started a new era of attempts to prevent and cure diseases [155]. When projects such as the Human Genome Project have emerged, the notion of PM has been offered a new definition. As we start to understand more about the heterogeneity amongst humans and existing diseases, our capacity to personalise and tailor medicine will be enhanced considerably [155]. For this cause, as new developments are introduced, it is likely that the concept of PM will evolve [155].

5.4 Genetics of chronic diseases

Over the past century, there has been a steady rise in the average lifespan of people [164]. In view of the fact that aging population amplification is associated with a growing load of chronic diseases, this has significantly impacted healthcare costs [164]. This poses the critical dispute of how to approach people with chronic health conditions for the sake of enhancing medical expenses [164]. CVD, diabetes, neurodegenerative diseases and the majority of cancers are the utmost predominant chronic diseases attributed with age and most of them have a significant genetic background [164]. Thus, PM will provide significant advantages in the fight against these diseases. The study of diseases, intermediate phenotypes and biomarkers genetics is a significant approach in the PM research. Mapping gene loci linked with complex traits needs solid levels of information and analysis, but, since 2007, approaches to achieve this aim have grown, and genetic mapping for complex phenotypes in humans has become a truth [98]. Basic approaches include linkage analysis and genetic association study through candidate genes and GWAS methodologies [98].

5.4.1 Genetics of CVD

CVD stand as the world's leading health concern encompassing an extensive range of health conditions commonly including coronary heart disease (CHD), stroke, hypertension, arrhythmias, atherosclerosis, MI and congestive cardiac failure [95]. The majority of these diseases are heritable [98]. The genetics of CVD widely involve the genetics of exposure for atherosclerosis, cardiomyopathy, arrhythmia, congenital heart disease and other related conditions [103]. With distinct ethnic variations, there are disparities in prevalence and types of CVDs, which confirms the significance of recognising potential genetic factors that could affect susceptibility [103]. Certain forms of CVD manifest a simple pattern of inheritance indicative of a single causal gene that results in a significant effect on phenotype [98]. However, most CVD phenotypes, such as MI or LDL plasma concentration, possess multifactorial inheritance, implicative of interaction of both genetic and environmental factors [98]. The GWAS have bring significant progress in CVD through novel insights into their genetic architecture [164]. These studies resulted in the screening of multiple loci linked with CVD and their risk factors [164]. About 1,500 CVD genes from approximately 24,000 research articles are found in CardioGenBase online resource that includes GWAS, genetic, epigenetic and pharmacological studies [165].

5.4.2. Genetics of AD

There are two basic types of AD: familial or early onset AD and sporadic or late onset AD [112]. Familial AD is rare autosomal dominant Mendelian disorder, contributing for lower than 5% of disease incidence [112]. This form of AD is induced by variations in three genes: *APP* coding for amyloid precursor protein, *PSEN1* coding for presenilin-1, and *PSEN2* coding for presenilin-22 [112]. On the other hand, late onset AD commonly affects people over the age of 65 years [112]. This form of AD is highly heritable accounting for 79% of disease burden, and it is etiologically heterogeneous originating from a mix of multiple genetic and non-genetic environmental contributors [112]. The sporadic AD involves significant number of genes that have been attributed with its appearance [112]. *ABCA7* (ATP Binding Cassette Subfamily A Member 7), *APOE*, *BIN1* (Bridging Integrator 1), *CD2AP* (CD2 Associated Protein), *CD33* (sialic acid binding Ig-like lectin 3), *CLU* (Clusterin), *CR1* (Complement C3b/C4b Receptor 1 (Knops Blood Group)), *EPHA1* (EPH Receptor A1), Membrane Spanning 4-Domains A4A/A4E/A6E *(MS4A4A/MS4A4E/MS4A6E)*, *PICALM* (Phosphatidylinositol Binding Clathrin Assembly Protein), and *SORL1* (Sortilin Related Receptor 1) are for instance some of these genes that have been verified by several researches [112]. Over 40 genes/loci have been associated with the risk of AD through the last 10 years, thanks for the routine utilisation of high-throughput approaches (Figure 11) [166].

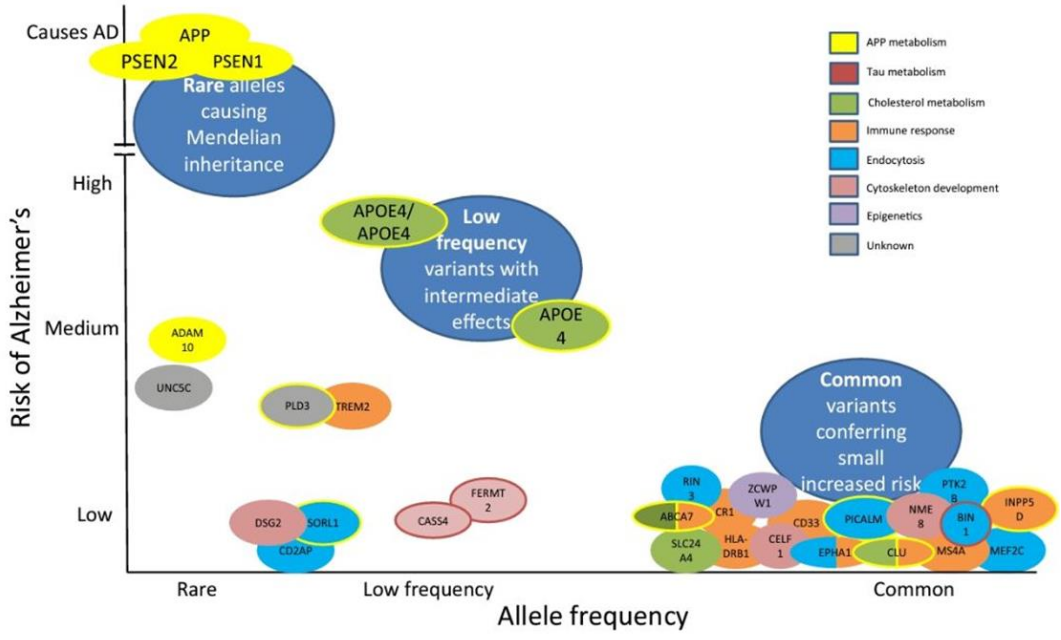

Figure 11. Summary of genes that have been found to be involved in AD [110]

It has been documented for decades that the most important genetic prognosticators for AD is the *ε4* allele of *APOE* gene which is strongly expressed in the brain [138]. Among the several SNPs of the *APOE* gene, there are three most common allelic variants that produce alterations in the protein coding sequence leading to the formation of three different *APOE* isoforms: *APOE2* (cys112, cys158), *APOE3* (cys112, arg158), and *APOE4* (arg112, arg158) [116, 138]. These three isoforms differ only at residue 112 or 158 by one or two amino acids, however, these slight differences significantly modify the formation and role of ApoE. People with one *ε4* allele have 2-3fold elevated chance of developing AD while those with two ε4 alleles have about 12-fold increased risk relative to people with no *ε4* alleles [116]. Furthermore, in the *APOE ε4* AD population the age of onset is lower. It is noteworthy that the *ε2* allele of *APOE* has a shielding role being linked with a reduced risk for AD [116].

Human *APOE* gene is situated on the long arm of chromosome 19q13.1, an AD-associated zone as reported by GWAS [55]. This region also holds, upstream of *APOE*, the *LSR* gene. Recently, a study was performed comparing AD cases with controls to examine the effects of *LSR* SNPs and their interplay with *APOE* gene allelic variants [55]. This study was the first to report the existence of epistatic interaction between *LSR* SNPs and *APOE* in AD patients [55] and thus supporting the notion that *LSR* is a candidate gene for AD.

5.4.3 Genetics of TL

More than 100 genetic variants have been identified as determinants of TL in GWAS of different populations [41]. These studies have revealed principal regulating mechanisms of LTL by identifying highly replicated variants harboring loci of known genes responsible for TL maintenance, such as *TERC* (Telomerase RNA Component), *OBFC1* (STN1 Subunit Of CST Complex) and *CXCR4* (C-X-C Motif Chemokine Receptor 4), that integrate into the wider image of the LTL-dependent atherosclerosis [144]. Moreover, TL related genetic variations can be used as further indicators of tailored individual's susceptibility to disorder risks.

5.4.4 Genetics of VEGF-A

VEGF-A gene is highly polymorphic and exhibits multiple polymorphisms in its promoter and 5′ untranslated region [167]. The *VEGF-A* gene is positioned on the short arm of chromosome 6p21.3 and bears a 14–kb coding region including eight exons and seven introns [168]. *VEGF-A* gene polymorphisms are of marked attentiveness for several angiogenesis–associated diseases [168]. Many studies have published the linkage between distinct *VEGF-A* SNPs and CVD like CAD, MI, and other related anomalies (Table 4) [167]. A meta-analysis study has shown that other polymorphisms, encompassing *VEGF-A* –2578A/C (rs699947), −1154G/A (rs1570360), +405C/G (rs2010963), and + 936C/T (rs3025039) are correlated with CAD susceptibility [168]. Moreover, VEGF-A polymorphisms were also found to be partnered with atherosclerosis manifestation, presumably via manipulating the expression of VEGF-A [169]. Furthermore, it was demonstrated that premenopausal patients can be at risk of developing hypertension coupled with overweight if they carry the homozygous GG genotype of the polymorphism -634 G/C (rs2010963) of the *VEGF-A* gene that stands as prior indication of CVD [170]. The results of a pooled study suggested that the possession of *VEGF-A* rs699947 C > A, rs3025039 C > T or rs2010963 G > C polymorphisms may constitute potential predictors for developing CHD [171].

Two large GWAS have focused on the identification of genetic determinants of VEGF-A levels and have demonstrated that 10 polymorphisms explain more than 50% of the variability of VEGF-A levels in plasma [14, 15]. This is an extraordinary finding in the genetics of multigenetic phenotypes, as most variables usually explain very low variability of these traits. Thus, these polymorphisms have a high potential for their use in PM and risk prediction of chronic diseases.

Table 4. Findings of different studies showing the association between different *VEGF-A* polymorphisms and CVDs [167].

Clinical Event	Author	Journal	Polymorphism	Risk genotype
CHF	Van der meer *et al.*	*J Card Fail* 2005	−634 C/G	CC
CAD	Howell *et al.*	*J Med Genet* 2005	−2578 C/A	AA
MI	Petrovic *et al.*	*Cardiology* 2007	−634 C/G	CC
CAD	Biselli *et al.*	*Heart Vessels* 2008	−2578 C/A	AA
HF after MI	Douvaras *et al.*	*Cardiology* 2009	−634 C/G	CC
Death after CABG	Pastuszczak *et al.*	*Pol Arch Med Wew* 2009	−634 C/G	GG
MI	Licastro *et al.*	*Curr Pharm Des* 2010	−2578 C/A	CC
CAD	Lin *et al.*	*Am J Hypertens* 2010	−634 C/G	CC

CABG, coronary artery bypass graft; CAD, coronary artery disease; CHF, chronic heart failure; HF, heart failure; MI, myocardial infarction; VEGF-A, vascular endothelial growth factor.

6. Diseases risk prediction

An essential public health requirement arising nowadays is to pinpoint people at high risk for a particular disease to allow for early prevention [61]. Models for risk prediction have been a major component of medical decisions and conclusions [62]. They offer fast and convenient approaches to determine the person's vulnerability for a particular disorder, which will then guide medication decisions [62]. Each individual is subjected to various risk factors and at different extents of exposure; consequently, personalised tailored medical decisions are urgently needed to specify an individual's risk to develop a disease and therefore to deliver the corresponding personalised care [172]. The concept of "genomic medicine" approach to health care has been widely used lately [63]. This approach employ the information derived from the genetic make-up of an individual to detect those who are at a greater risk to get affected by a given disease and to intervene to prevent this disease at an earlier stage [63]. The era of "big data" in medicine and health has paved the way for risk prediction tools vowing to foresee human health incidents [64]. Multiple benefits result

from the use of disease risk anticipating approaches that reliably forecast the prevalence and prognosis of a disease, or a person's response to a given medication [64]. Such tools allow people to make well-informed medical and preventive decisions; assist clinicians to accurately personalise and customise treatments via organising medical care and prevention; and help medical management systems in distributing assets to risky patients [64]. Risk prediction tools are not created to take the place of doctors, but to aid health professionals in their interpretations by offering objective risk predictions [173].

In chronic diseases prediction domain, creating a validated prediction model is critical, yet other features are demanded before considering the model clinically functional. Foremost, the employment of a prediction model in wide scale of populations with thorough model authentication and continual updates is needed. Besides, it is necessary for the model to be comprehensible and easy to use. The model should be easily accessed by people and clinicians to figure out effortlessly the disease risk and to elucidate the results to promote general use. Moreover, the prediction model outcomes should be coupled with clinical application [172].

6.1. CVD risk prediction tools

Based on applicable worldwide and local recommendations for CVD prevention, the efficacy of early interventions relies on the estimation of the CVD risk of an individual [174]. Through years, several risk algorithmic approaches have been developed to anticipate CVD morbidity over 10-years or lifelong vulnerability in various demographic groups, in apparently healthy people, patients with previous CVD and diabetic patients [173]. Risk algorithms are set at disposal for utility in clinical practice via tools, typically interactive and with online available access [173]. Distinct CVD risk algorithms are required for distinct groups of patients [173]. There are numerous risk algorithms available, including the Systemic Coronary Risk Evaluation (SCORE) for the prognostication of the 10-year vulnerability to CVDs mortality in Europe, the QRISK for forecasting of the composite consequence of coronary cardiac anomalies, ischemic stroke or acute ischemic attack in the UK, and the pooled cohort equations in North America for the prediction of the possibility of developing atherosclerosis incident over 10-years [173]. The Framingham Risk Score is the most familiar health risk prediction model that evaluates an individuals' CVD risk and recommends lifestyle alterations and early pharmacologic interventions for those who are likely to develop CVD in a particular timeframe [172].

6.2. AD risk prediction tools

Likewise, precise detection of people at high risk for AD is very critical for early diagnosis and intervention, such as close monitoring, enhanced care and risk factors targeted intervention. Early detection is important for the prosperous investment of disease modifying treatments during pre-symptomatic phase of the disease before the pathological amyloid and tau accumulation leading to extensive brain damage [175]. Multiple approaches have been used for forecasting AD-linked parameters like medical assessment, cognitive and imaging assays that exploit multimodal biological molecules features present in AD. Moreover, conventional panel studies relying on regression analysis present the relation between dependent variables and other familiar variables, like clinical examination and cognitive test scores [175]. A further approach includes machine learning methodologies and supervised predictive analytics tools like support vector machines, random forests, and artificial neural networks, that employ automated data regularities identification to study the association between a set of indicators or biomarkers and their tags or corresponding analysis [175]. These technologies are employed to distinguish AD cases and for identifying people having possibilities of developing AD in a given period of time [175].

6.3. Artificial intelligence and diseases risk prediction

Based on how essential data-intensive assays are to uncover relevant intervention targets and strategies for PM, implementation of artificial intelligence (AI) in healthcare can play an extensive role in the growth of PM at all stages of the clinical development and implementation of novel tailored health products, from identifying suitable intervention targets to examining their usefulness [65]. AI is the part of information technology that focuses on the ability of machinery to mimic or boost brilliant human intelligent behaviors [66]. AI contributes to the advance of PM through refining relevant assays and methods of storage, aggregation, accession and ultimately integration of the produced data [65]. Through intelligently incorporating panomics with medical, imaging, epidemics and population-based analysis data, PM has the ability to refine and enhance the conventional symptom-driven use of medicine and enable a broad scale of emergency prompt responses to optimise and customise better tailored care [66]. Moreover, healthcare problems including misdiagnosis, overtreatment, medical errors, reduced performance, poor handling of clinical data, high cost and spending can be mitigated considerably by the employment of the

evolution in informational technology, that is AI, at every level of care (Figure 12) [66].

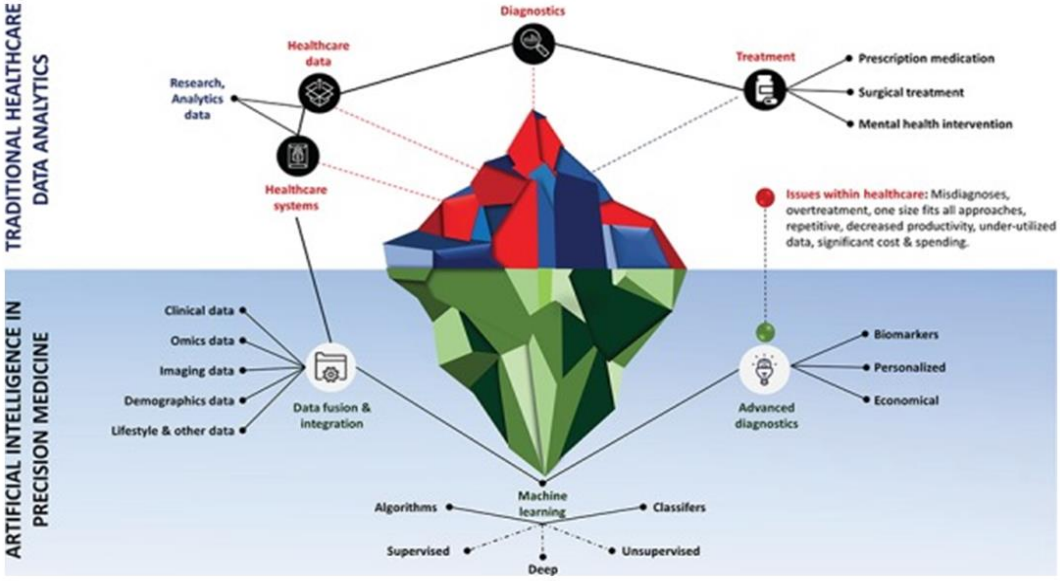

Figure 12. Privilege of AI in PM compared to traditional care data analysis [66]

6.3.1. Machine learning in diseases risk prediction

Machine learning (ML), a subset of AI, is now comprising the evolutionary power for guiding public health management with multiple computational advantages, which involve instantaneous patient tracking, disorders profiles examination, abnormalities identification and recommendation of medical remedy, personalised care delivery with improved medication, medical inaccuracy mitigation, anticipating scoring, remedial accountability, spotting of elevated disease probability for rapid intervention, phenotypic data recognition, scanning hypothesis, specifying hallmarks of chronic maladies, risk forecasting, strategic imperative and resources distribution [66]. Support vector machine, deep learning, logistic regression, discriminant analysis, decision tree, random forest, linear regression, Naïve Bayes, K-nearest neighbor, and hidden Markov model are the most extensively implemented ML platforms in medicine (Figure 13) [66].

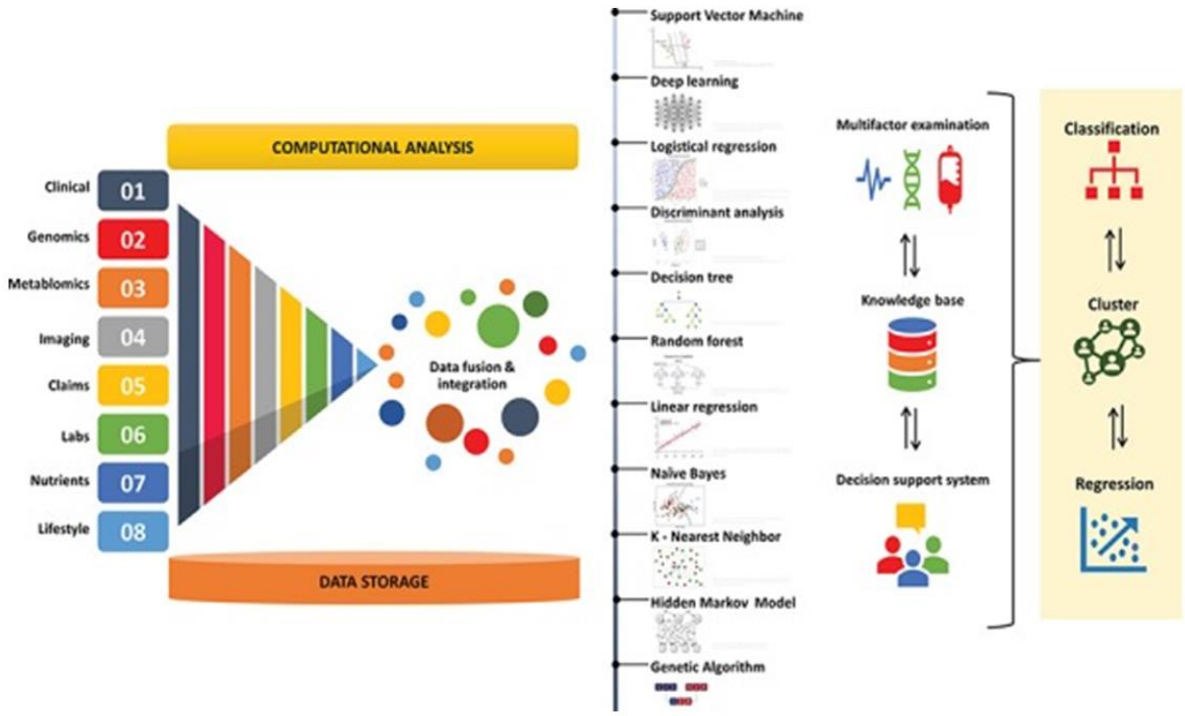

Figure 13. Contribution of ML algorithms for the advance of medicine through integration and interpretation of data from different sources [66]

6.3.2. Machine Learning risk prediction of CVD

Most medical scientists are attracted by the state-of-art ML-oriented techniques for data analysis and interpretation that can enhance the ability of detecting people at risk of developing chronic diseases through broad databases to point new risk predictors and their complicated interactions. Lately, a modern automated ML method based on AutoPrognosis algorithmic framework was developed to assess CVD risk prediction in symptomless subjects [176]. This approach was established using records of about 400,000 UK Biobank subjects, including more than 450 variants per subject. Compared to popular scoring systems that are established on standard disease predictors and presently endorsed by preventive recommendations as Framingham score, AutoPrognosis greatly increased the accuracy of CVDs risk prediction (Table 5). Besides, this approach has permitted the finding of new predictors for CVD. Notably, AutoPrognosis revealed interactions among distinct elements of an individual, leading to the detection of hazard prognosticators of subgroups for whom the present tools were giving inaccurate disease susceptibility forecasts [176].

Table 5. Different risk anticipation models' performance [176]

Model	AUC-ROC	Absolute AUC-ROC Change
Framingham Score	0.724 ± 0.004	Baseline model
Cox PH Model (7 core variables)	0.734 ± 0.005	+ 1.0%
Cox PH Model (all variables)	0.758 ± 0.005	+ 3.4%
Support Vector Machines	0.709 ± 0.061	- 1.5%
Random Forest	0.730 ± 0.004	+ 0.6%
Neural Networks	0.755 ± 0.005	+ 3.1%
AdaBoost	0.759 ± 0.004	+ 3.5%
Gradient Boosting	0.769 ± 0.005	+ 4.5%
AutoPrognosis (7 core variables)	0.744 ± 0.005	+ 2.0%
AutoPrognosis (369 non-lab. variables)	0.761 ± 0.005	+ 3.7%
AutoPrognosis (104 lab. variables)	0.735 ± 0.008	+ 1.1%
AutoPrognosis (all variables)	0.774 ± 0.005	+ 5.0%

AUC, Area Under the Curve; ROC, Receiver Characteristic Operator

6.3.3. Machine Learning risk prediction of AD

Another state of art ML model was developed for the prediction of AD future incidence exploiting broad-range administrative health information [177]. Knowing that conventional prior risk prediction tools possess reliability fluctuating between 0.5 to 0.78 in AUC, this ML-based algorithmic framework was able to predict AD potential occurrence with 0.781 reasonable exactness in terms of AUC in a year forecast. The accurate performance of this model established on huge national-wide contexts can promote the future usefulness of AD predictive tools build on executive information [177]. A systematic review revealed that multifaceted approaches better anticipate the likelihood for neurodegenerative complications, while approaches based on one factors perform weakly, proposing that effective prediction of AD prospect needs broad plethora

of data, and thus, supporting data-driven AI-ML approaches that gathers important data records from massive medical databases incorporating thousands of information of individuals' clinical records and deliver well-tailored AD risk prognostication [177].

6.3.4. Machine Learning risk prediction with genetic data and visualisation

Genetic technologies based on GWAS and next generation sequencing (NGS) are empowering the exploration of multifactorial disorders' genetic constitution to construct more exact disease prognosticative tools that will potentially promote successful methodologies to tailored prevention and medication options for different disorders. Computational ML approaches can make use of masked linkages among genetic and other predisposing parameters, anticipating the disease probability through engaging genetic aspects and network forecasting tools [178]. ML algorithms place marked weight on optimising the performance of tailored anticipating approaches. Although ML-dependent data processing methodologies may furnish suitable basis for using vast array of genetic records to prognosticate the likelihood of encountering a disease, their progress seems to be in preliminary level [178]. Implementing extremely adaptive data mining and processing algorithms for collecting genetic features is essential for rendering these frameworks quite efficient for extracting data from existing GWAS and candidate genes and analysing them in multitude of subjects together with the evolving information from high throughput sequencing fieldwork, like the 1000 Genomes project [178]. ML approaches conform a panel of complicated analytical and computer based programs like support vector machine or random forest to prognosticate through establishing a mathematic map displaying the sophisticated relations between a batch of candidate allelic variants and complex disorders' traits [179]. Supervised ML prognosticating tools are formed by training the machine using well-labeled datasets to outline the link between the genotype information and the related disorder, that is to predict outcomes for hidden data [179]. Following the training phase, algorithms with optimal prognosticating ability are sorted out for verification [179]. The workflow for generating a ML approach from a genotype input data is clarified in Figure 14 [179].

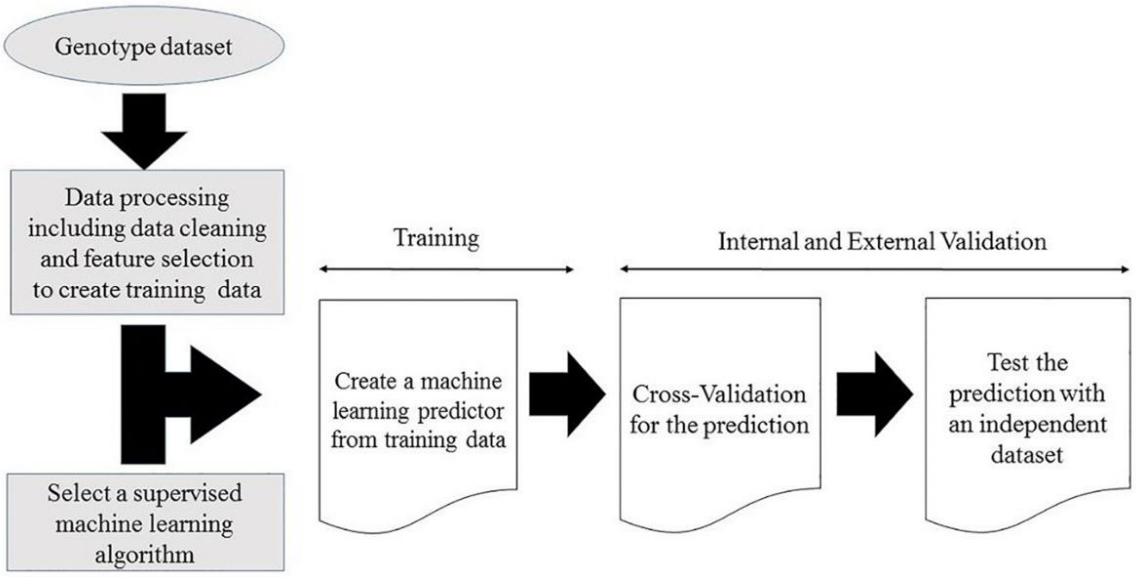

Figure 14. The protocol of supervised ML algorithm generation from genotype input data [179]

Actually, the solidity of ML data manipulation in multifactorial disorders incidence prognostication resides in its processing of synergetic extensive data [179].

The immensity of data explosion in the biomedical research in contemporary "big data" era is evident [68]. The growing amount of data employed in GWAS, and the complication of the methods utilised to assess the data, require a new integrative approach to genetics discovery [180]. We must depend on the powers of machines and humans to thoroughly capture the potentiality of today's genetic data [67]. In the genetics realm, where the quantity of achievable data keeps growing, the implementation of data visualisation and ML techniques has the ability to boost the progress in genetics discovery [67]. Data interpretation is important to transform the increasing volume of data into knowledge and this interpretation is facilitated vastly by interactive and user-friendly data visualisation [69]. Application of effective visualisation technologies is critical not only to address the known but also to uncover the hidden, enabling assumptions, insights and principles to be inferred [68]. The visualisation of genomic information is important for elucidation and postulate formulation, in addition to practically assisting communicating discoveries. Interactive visual tools bridge the distance between algorithms and investigators' mental abilities [181]. Multiple strategies for visualising genomic-derived data in an interactive manner were developed in response to the massive information basis of biomedicine and the absence of clear-cut postulations in several disciplines [181]. Such tools should be simple to use and manage big

quantities of input data in a way that users can retrieve and access essential information and import the high resolution image for their publications [70]. One of this tools that is recently developed for multiple applications going from science into business, is the Microsoft Power BI (Business intelligence). Power BI consists of software services, applications, and connectors, which transform database into coherent and interactive insights using a large variety of plots and graphs. Power BI allows the connection of different datasets, the data visualisation and/or discovery, and the data sharing [71].

Hypothesis and Objectives

CVD and AD are two of the major chronic diseases that have a vast burden in the individual's health and life and the healthcare systems. They share some common pathophysiological pathways like inflammation, angiogenesis and lipid metabolism abnormalities. The two diseases have a strong heritable component and many efforts have been done to identify the variants that genetically define the risk for these diseases. Although with the GWAS a large progress has been made in the determination of some of these variants, an important part of the heritability of these diseases remains to be uncovered. Consequently, the personalisation of disease risk is still limited. Therefore, there is an increased need for identification of novel determinants of disease risk that will eventually allow tailoring and personalisation of risk prediction. A more focused approach, like the candidate gene approach of key biomarkers, could improve this desired objective.

TL and VEGF-A are key biomarkers that link different chronic diseases together [29-31]. These biomarkers have a large genetic background and are optimal biomarkers for use in PM [13, 32]. Importantly, for VEGF-A, we have identified a large part of its genetic background [14, 15] and the variants that explain its variability can serve as novel determinants for chronic diseases risk evaluation. Molecular determinants of these biomarkers could also be used for this purpose. We thus hypothesised that the determinants of these biomarkers could be associated with CVD, AD and their intermediate phenotypes and could be of important value for the prediction of disease risk and the earliest disease prevention.

Precisely, the aim of this thesis was to identify the "-omics" determinants of VEGF-A and TL, to investigate their relationships with CVD and AD and to study their usefulness in disease risk prediction. Furthermore, in order to allow the replication, generalisation and extension of our results, an emphasis was given in the standardisation of the analytical process and the visualisation of the datasets that were used in the thesis projects.

Four objectives have been evaluated:

***Objective:* 1.** Identify determinants of LTL and VEGF-A in health, CVD and AD

Objective: 2. Develop predictive models with the use of the above identified determinants

Objective: 3. Develop algorithms for the use of the previously developed predictive models

Objective: 4. Develop a user-friendly application for the descriptive presentation of populations and primary analyses

Methods and Materials

Populations

The STANISLAS Family Study

The STANISLAS Family Study (SFS) is a familial longitudinal cohort aimed to study the genetic and environmental effects on cardiovascular intermediate phenotypes using the model of nuclear families [182, 183]. Laboratory, clinical, lifestyle and demographic data were initially collected on 1006 families (two parents and at least two children aged 4 or more) from the East of France by trained professionals. Follow-up visits were performed 2 more times in an interval of 5 years (total follow-up period of 10 years). Participants were recruited in the Center for Preventive Medicine (CMP) of Vandoeuvre-lès-Nancy, during a periodical health assessment. Inclusion criteria included an European-Caucasian origin and lack of any declared acute and/or chronic disease. There are > 600 variables collected for each individual. Measurements were performed in the CMP and in other collaborating research laboratories. 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 Biological Ressources Centre 'Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire' (BRC IGE-PCV) with identification number BB-0033-00051.

The TELARTA study

Telomeres and Arterial Aging (TELARTA) study aimed to investigate the links between arterial aging and development of atherosclerosis with TL in different tissues and LTL dynamics. It is involving the study of the difference between TL of muscle (estimate of TL at birth) and leukocytes (current estimate of TL) that is an index of telomere attrition through life. The main population of the study (discovery cohort) is a case-control study of patients with clinically significant atherosclerotic disease and controls without evidence of clinical CVD or history of previous CVD. Muscle biopsies were collected during undergoing surgery or pacemaker/defibrillator implantation. The university hospitals of Nancy and Marseille (France) recruited 131 atherosclerosis cases and 128 controls. Exclusion criteria included: body mass index (BMI) >40 kg/m^2, glomerular filtration rate < 30 mL min^{-1} 1.73 m^{-2}, active malignancy, or history of chemotherapy/radiotherapy for cancer, and presence of aortic aneurism.

A replication population (replication cohort) was also enrolled and consisted of additional 91 individuals from the university hospitals of Nancy and Marseille, and 52 individuals recruited at 3 located in Athens (Greece) hospitals (Onassis Cardiac Surgery Center, Iaso General Hospital, Hippokration Hospital). The exclusion criteria for the replication cohort were less strict and individuals excluded from the TELARTA discovery cohort based on BMI >40 kg/m^2, glomerular filtration rate values, malignancy or history of chemotherapy/radiotherapy were included in replication cohort. All participants gave their written informed consent approved either by the Ethics Committee (Comité de Protection des Personnes) of Nancy, France or by the Ethics Committee of the University of Athens and Ethics Committee of each one of the 3 participating hospitals of Athens [42].

Case-control AD population

The case-control of late-onset AD population consisted of 1078 (602 controls and 476 cases) unrelated individuals of European origin (9 European countries) and its recruitment aimed to study the effect of ApoE levels and *APOE* polymorphism on AD risk. The recruitment and data collection of this population was previously described [25]. The clinical diagnosis of AD was based on 5 criteria: the National Institute of Neurological and Communicative Disorders and Stroke and the AD and Related Disorders Association classification of probable AD; the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; the International Classification of Diseases and Related Health Problems, 10th Edition; the Mini Mental State Examination score equal to or less than 23 (Mini Mental State Examination >23 for controls); as well as the modified ischemic scale less than three. Secondary causes of dementia were excluded by computerised tomography scan of the brain. Exclusion criteria included other chronic diseases like cancer, CVD, cerebrovascular diseases and neurological diseases like Parkinson's diseases. Controls followed similar assessment and interviews as cases, except from the computerised tomography scan. They had a Mini Mental State Examination >23 and were free of dementia. The project was approved by the related ethics committee, and all subjects gave their informed consent to participate in the study. Genotyping data were available in a subsample of the population.

DNA extraction

Blood samples of individuals were taken after overnight fasting and were collected to EDTA

(ethylenediaminetetraacetic acid) vacuum tubes. For individuals from SFS and the AD case-control population, whole blood DNA was extracted by Miller technique [184] and was stored on -80 °C in BRC IGE-PCV, while in TELARTA cohort the DNA extraction was performed by the phenol/chloroform/isoamyl alcohol method.

Genotyping

Genotyping of SNPs in the SFS, the AD case-control population and the TELARTA cohort was performed by the Laboratory of the Government Chemist (LGC Ltd., Teddington, UK), using the competitive allele-specific PCR (KASP) chemistry coupled with a Förster resonance energy transfer-based genotyping system (http://www.kbioscience.co.uk/reagents/KASP/KASP.html), which enables bi-allelic scoring of SNPs and insertions and deletions (Indels) at specific loci [185], and by Randox genotyping VEGF-A assay.

Peripheral blood mononuclear cells (PBMCs) isolation

PBMCs of individuals of the SFS were isolated using the density gradient centrifugation method [186]. Whole blood was collected in tubes with 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^{TM} 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 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 PBMCs 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 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.

Gene expression / mRNA quantification

In SFS, 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 [9, 187]. 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-A_{121}, VEGF-A_{145}, VEGF-A_{165}, VEGF-A_{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 *B2M* (Beta-2-Microglobulin) gene.

Total protein extraction from PBMCs

In SFS total protein from PBMCs was extracted using 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) added to the PBMCs 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 stored at -80 °C.

Protein quantification

Quantification of inflammatory molecules (IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, MCP-1 (monocyte chemoattractant protein), TNF-α, INF-γ, EGF (epidermal growth factor) and VEGF-A) from PBMCs extracts of individuals from the SFS and from plasma samples of TELARTA cohort was performed by Randox high sensitivity multiplex cytokine and growth factor array (Evidence Investigator Analyzer, Randox Laboratories Ltd., Crumlin, United Kingdom) Cytokine Array I. This array enables simultaneous detection of 12 cytokines and growth factors in a single sample. It is a Randox patented array of 9 x 9 mm^2 activated biochip with spatially discrete test regions containing antibodies specific to each of the inflammatory molecules assessed.

Telomere length measurement

MTL (muscle telomere length) and LTL were measured in DNA extracted from muscle biopsies and peripheral blood leukocytes respectively [42, 188]. Skeletal muscle biopsies (~100 to 200 mg in the surgical field) were collected from individual during surgery, flash frozen in liquid nitrogen and stored on -80 °C until DNA extraction.

DNA samples passed an integrity testing using a 1% (wt/vol) agarose gel before TL measurement, which was performed by the Southern blot analysis of terminal restriction fragments, as described previously [44]. Briefly, DNA samples were 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 depurination, denaturation and neutralisation, DNA was transferred on a positively charged nylon membrane (Roche) using a vacuum blotter (Biorad, Hercules, CA). Membranes were hybridised at 42°C with a digoxigenin-labelled telomeric probe. The probe was later detected by the DIG luminescent detection procedure (Roche) and exposed on a 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. The analysis was performed in the laboratory UMR INSERM U1116 DCAC.

Statistical analysis

Normality testing was performed for all dependent variables using the Kolmogorov-Smirnov test and transformations were applied to achieve normal distribution if possible. For normally distributed variables (transformed or non-transformed) parametric statistical tests were performed, while for variables not normally distributed (even after transformations) non-parametric tests were used.

Non-parametric Spearman test was performed to study the correlation between VEGF-A and inflammatory molecules levels in PBMCs extracts and linear regression models adjusted for age and gender were used to test the associations of these variables. Associations between VEGF-A

mRNA isoforms and inflammatory molecules levels in PBMCs were tested using linear regression models adjusted for age and gender. Mixed models adjusted for family structure were also applied to correct for possible issues of stratification due to familial resemblance in SFS. Analyses were performed using the SPSS 20.0 software.

The association of VEGF-A plasma levels with LTL, MTL and LTL/MTL were tested using multiple regression analyses adjusted for age, sex and statin use, because statins were significantly associated with VEGF-A levels in the TELARTA population. Non-genetic analyses were performed using the SPSS 20.0 software. The direct effects of TL-related and VEGF-A-related genetic variants on LTL, MTL and LTL/MTL were tested using linear regression models adjusted for age and sex using PLINK under the assumption of an additive genetic model. Reference allele for all variants was the minor allele. Epistatic interactions were tested using R and the R package CAPE (combined analysis of pleiotropy and epistasis).

In the AD case-control population quality control was performed before the analysis and individuals with missing data and outliers were excluded from further analysis. The identification of outliers was performed using the principal component analysis. Then, quality control was performed on genotyping data: exclusion of individuals that genotyping has failed for >10% of the polymorphisms, exclusion of polymorphisms that failed in >20% of the individuals, and exclusion of polymorphisms that are not in agreement with Hardy-Weinberg equilibrium tested using the χ^2 test. Missing values for the remaining individuals were imputed using the missMDA R package.
The prediction model for the AD risk was developed using a machine-learning approach called elastic net (EN) [51]. It is a method extensively used in the development of disease prediction models, especially when these integrate "-omics" data [189-192]. It is also being used with genetic data and GWAS studies [193, 194] and it is considered to function better in cases where gene * environment interactions are involved in the prediction of a disease [195]. Comparison studies have shown that EN is a powerful tool, especially when additive gene effect is expected [196]. It has also been previously used in AD risk prediction with clinical factors, imaging and "-omics" [52-54] but not to study the effect of candidate genes as predicting factors. It is testing hundreds of logistic regression models and penalising them until it finds the optimal one. It is a regularised regression methodology that combines Lasso and Ridge regression. Logistic regression model

indicates the conditional probabilities through a linear function of the predictors:

$$\Pr(G = 1/x) = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\boldsymbol{\beta_0} + \boldsymbol{x}^T\boldsymbol{\beta})}}$$

$$\Pr(G = 2/x) = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{+(\boldsymbol{\beta_0} + \boldsymbol{x}^T\boldsymbol{\beta})}} = 1 - \Pr(G = 1/x)$$

Alternatively, this implies that

$$log\frac{Pr(G\ =\ 1/x)}{Pr(G\ =\ 2/x)}\ =\ \boldsymbol{\beta_0} + \boldsymbol{x}^T\boldsymbol{\beta}$$

In the regularised maximum binomial likelihood, $p(x_i) = \Pr(G = 1/x_i)$ is the probability for the i observation at a specific value for the parameters (β_0,β). To maximise the penalised log-likelihood

$$max_{\,(\,\beta_0,\beta)\in\mathbb{R}^{p+1}}\ [\ \frac{1}{N}\sum_{i=1}^{N}\quad \{\ I(g_i=1)*\log p(x_i) + I(g_i=2)*\log(1-p(x_i))\ \}\ -\ \lambda P_a(\beta)]$$

In the final step of the EN, the use of coordinate descent for solving the penalised weighted least-squares problem has as an objective function, the penalised negative binomial log-likelihood, and it is:

$$min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}}\quad -[\frac{1}{N}\sum_{i=1}^{N} y_i\ (\beta_0\ +\ x_i^T\beta)\ -\ log(1\ +\ e^{(\beta_0\ +\ x_i^T\beta)})]\qquad +\qquad \lambda[(1-\alpha)||\beta||_2^2/\ 2+\alpha||\beta||_1]$$

where λ is a parameter that controls shrinkage, 0 is no penalty and ∞ is entirely penalty, and α regulates how much of the ridge vs. lasso, 0 is the ridge, and 1 is the lasso. In our process, we run multiple cross-validations for different values inside this interval [0,1] to define the optimal α for our model. To apply this methodology in R software, we used the glmnet package (version 4.1-1).

Development of an analytical pipeline for the diseases risk prediction model

The R scripts that were used to perform the analyses of the AD cases-control population were combined into a standardised automated algorithm with a user-friendly interface (RSGEN) using the R Shiny package. Shiny is an R package that makes it easy to build interactive web applications straight from R. It was selected because it allows the development of customised tools that can be modified and extended based on the needs of the research teams.

Development of a user-friendly application for data visualisation

As already mentioned, the big data era is facing the challenge of managing and visualising big data. The large volume of information can be practically used and fully exploited by user-friendly ways of presentation. In this thesis, a user-friendly application for the data visualisation has been developed with the use of a novel visualisation tool, the PMI is based on Power BI (Microsoft) which allows for interactive visualisation [71]. It is a tool that consists of software services, cloud-based applications, and connectors, which synergistically turn the data of different populations into visually immersive, and interactive insights. It allows connection of different data sources, data visualisation and discovery, data querying and data sharing. It is based on a self-service environment where users can create and share reports and dashboards. It is designed to make data analysis quick and easy and it provides connectivity to different datasets with a simple interface that makes dashboard creation fast and accessible. It was selected for the populations data visualisation task because it is a perfect tool for reporting and produces great-looking visualisations while it can be easily used and provide meaningful outputs for scientists of different disciplines and not only specialists in informatics. Also, the fact that it is developed by Microsoft increases the reliability of the tool.

Input data included the population of the publications 3 and 4 datasets in Excel spreadsheet. These data were stored locally in the EA_1122; IGE-PCV desktops. The Power BI tool was connected with R locally and the customised application was developed as desktop application. However, Power BI allows the input of data from cloud for example a university cloud, the use of cloud applications and the development of a customised web application easily shared and accessible.

However, the use of web tools requires specific authorisations and safety procedures for data security and due to the sensitive nature of the medical data included in this project the local procedures were followed. This architecture is presented in Figure 15.

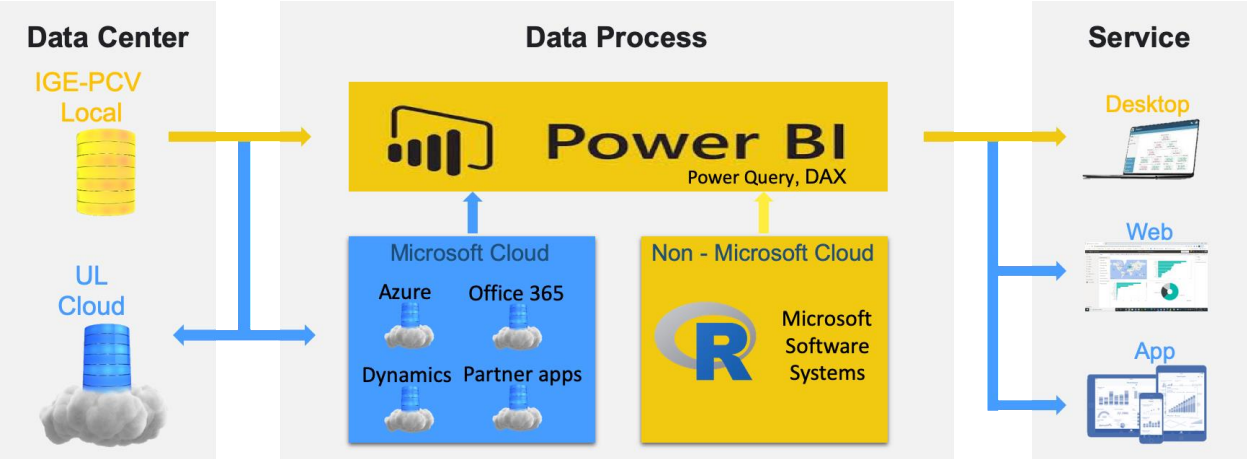

Figure 15. Architect of Power BI tool

Yellow color indicates the procedures that were followed in the thesis. The blue color indicates possible extensions of the application depending on authorisations to share data online.

Results

Publication 1

Telomere length determinants in childhood [32]

Gorenjak V, **Petrelis AM**, Stathopoulou MG, Visvikis-Siest S.

Clin Chem Lab Med. 2020 Jan 28;58(2):162-177.

My specific contribution in this study was: Bibliography search and participation in the drafting of the manuscript

This review article aimed to present the current situation of knowledge on telomere dynamics during childhood including also the factors that affect TL in embryonic life and in newborns.

The childhood is a particularly important period for TL [33] that determines TL throughout life and its relation with common chronic diseases. The environment plays a detrimental role during pregnancy for the TL of neonates [34-37]. A very challenging field is the genetics of TL in childhood that seems to be differentially regulated compared to adults [38]. We emphasise the usefulness of TL as an index of health and we propose its regular measurement that could be of clinical importance starting from childhood.

This review article served as an introduction to the thesis, focusing on one of the two biomarkers of interests and setting the knowledge base for the original publications on TL genetic determination and its links with biomarkers of chronic.

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; previously published online August 29, 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].

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

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 ($p \leq 5\times10^{-5}$).

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,

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.

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

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/m² 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

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.

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

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

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

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

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.

Acknowledgments: 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.

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

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.

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.

Publication 2

Peripheral blood mononuclear cells extracts VEGF-A protein levels and VEGF-A mRNA: Associations with inflammatory molecules in a healthy population [197]

Gorenjak V*, Vance DR*, **Petrelis AM**, Stathopoulou MG, Dadé S, Shamieh SE, Murray H, Masson C, Lamont J, Fitzgerald P, Visvikis-Siest S.

PLoS One. 2019 Aug 16;14(8):e0220902.

* equal first authors

My specific contribution in this study was: Participation in the analysis, in the drafting and the revision of the manuscript

Objective

This investigation has been performed in order to assess the determinants of VEGF-A in PBMCs. PBMCs are immune cells that include populations of B cells (~15 %), T cells (~70 %), monocytes (~5 %), and natural killer (NK) cells (~10 %) [39]. They are secreting VEGF-A and inflammatory molecules [40]. PBMCs extracts is a good model for investigating the association of VEGF-A with other inflammatory molecules.

The specific objective of this study was to investigate the associations of 11 inflammatory molecules with *VEGF-A* mRNA levels of four VEGF-A isoforms (*VEGF-A*$_{165}$, *VEGF-A*$_{189}$, *VEGF-A*$_{145}$ and *VEGF-A*$_{121}$) and with VEGF-A protein levels in PBMCs in a healthy population, allowing to identify physiological determinants of VEGF-A.

Populations and analysis

Participants from the STANISLAS Family Study (SFS) (n=285) were included in the study. The levels of IL1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, MCP-1, TNF-α, INF-γ, EGF and VEGF-A were measured in PBMCs extracts. Four *VEGF-A* mRNA isoforms from total RNA derived from PBMCs (*VEGF-A*$_{121}$, *VEGF-A*$_{145}$, *VEGF-A*$_{165}$ and *VEGF-A*$_{189}$) were quantified in 110 participants. VEGF-A levels were log-transformed to normalise the distribution of data. Data were tested for outliers, which were removed before correlation and linear regression analysis. Non-parametric correlation analysis (Spearman correlation) and linear regression models were applied, adjusted for age and gender, to test for possible associations between VEGF-A protein levels and

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

Results

Significant correlations were observed between VEGF-A and EGF, IL-1β, IL-8, MCP-1 and TNF-α. VEGF-A levels were significantly associated with IL-4, MCP-1 and EGF in linear regression models adjusted for age and gender (Table 6).

Table 6. Significant associations between VEGF-A 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

Significant correlations were observed between *VEGF-A*$_{121}$ and IL-4, between *VEGF-A*$_{165}$ and IL-4, IL-6, and between *VEGF-A*$_{189}$ and IFN-γ, IL-1β, IL-4, IL-6.

VEGF-A$_{165}$ was associated with MCP-and IL-1α, whereas *VEGF-A*$_{189}$ was associated with IL-4 and IL-6 in linear regression models adjusted for age and gender (Table 7).

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

	***VEGF-A*$_{165}$**			***VEGF-A*$_{189}$**		
	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 analyses for associations were repeated using mixed models adjusted for family structure, which gave the same significant results.

Conclusions

This work gave new insights into the links between inflammatory molecules and VEGF-A, thus contributing to increase our knowledge on the regulation of angiogenesis.

It has revealed significant associations between VEGF-A levels and its mRNA isoforms with IL-1α, IL-4, IL-6, EGF, and MCP1 in PBMCs of healthy individuals. These results further support the important role of cytokines and growth factors in VEGF-A regulation also at lymphocytes cellular extract level. The interactions of specific isoforms with different inflammatory molecules should be considered in the investigations of molecular and biological pathways involving VEGF-A, angiogenesis and chronic diseases development. They could also be of value for prediction of chronic diseases where inflammation and angiogenesis are involved like CVD, cancer and AD.

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).

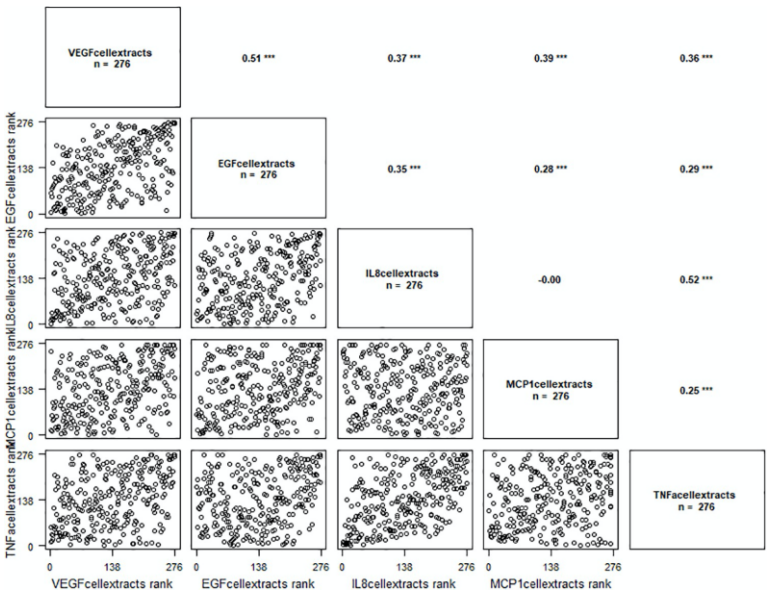

Fig 1. Correlation graph between VEGF protein levels (VEGF cellular extracts) and EGF, IL-8, MCP1 and TNF-α cellular extracts levels. '****' $p \leq 0.001$.

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

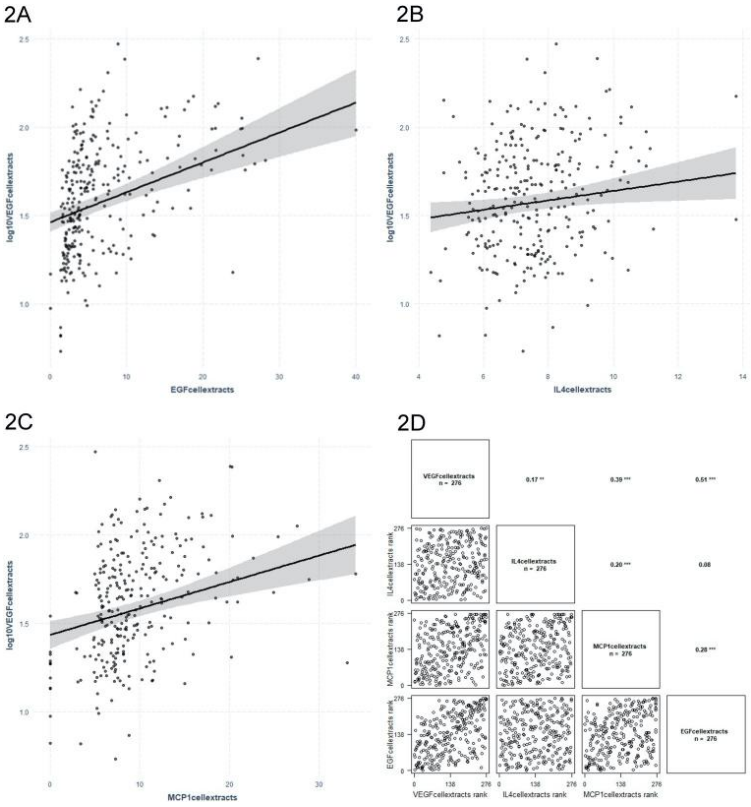

Fig 2. Regression graphs for VEGF protein levels (VEGF cellular extracts) as dependent variable and EGF (2A), IL-4 (2B) and MCP1 (2C) cellular extracts levels as independent variables. The related correlation graph is presented in (2D). '***' $p \leq 0.001$, '**' $p \leq 0.01$.

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

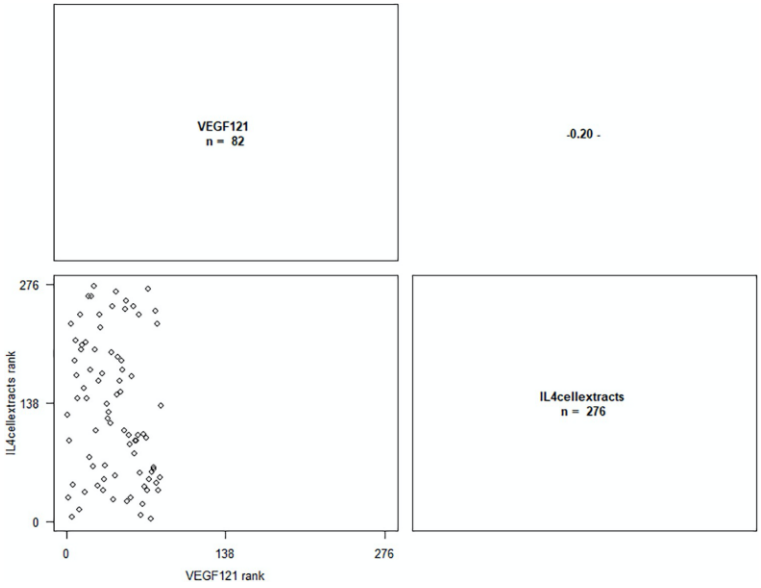

Fig 3. Correlation graph between *VEGF*$_{121}$ isoform mRNA levels and IL-4 cellular extracts levels. '-' $p \leq 0.1$.

https://doi.org/10.1371/journal.pone.0220902.g003

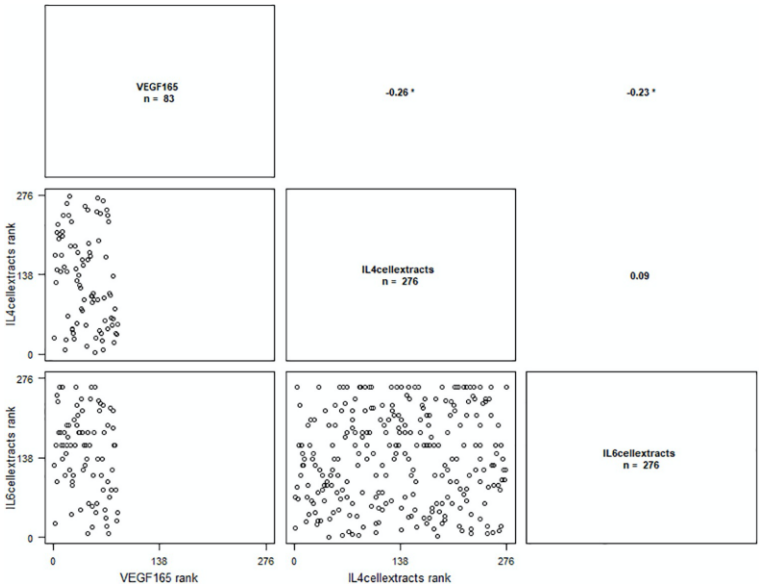

Fig 4. Correlation graph between *VEGF*$_{165}$ isoform mRNA levels and IL-4 and IL-6 cellular extracts levels. '*' $p \leq 0.05$.

https://doi.org/10.1371/journal.pone.0220902.g004

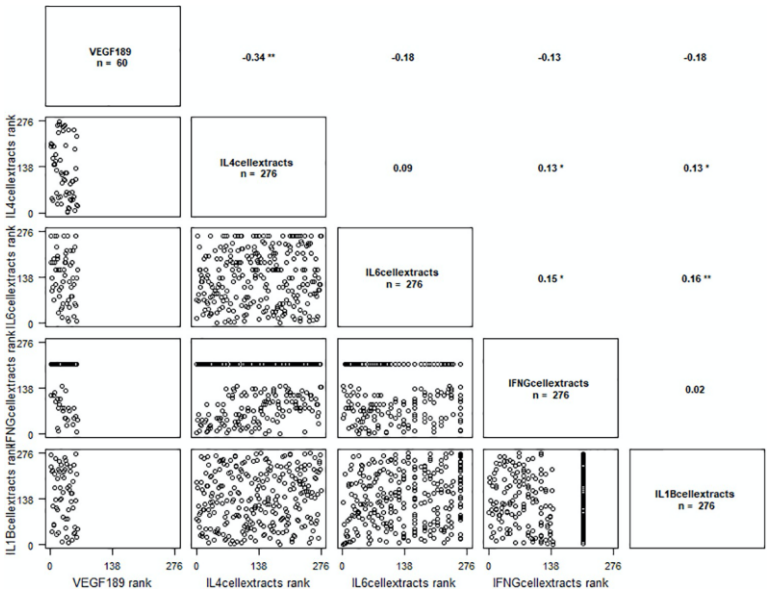

Fig 5. Correlation graph between $VEGF_{189}$ isoform mRNA levels and IL-4, IL-6, INFγ et IL-1β cellular extracts levels. "**" $p \leq 0.01$, "*" $p \leq 0.05$.

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

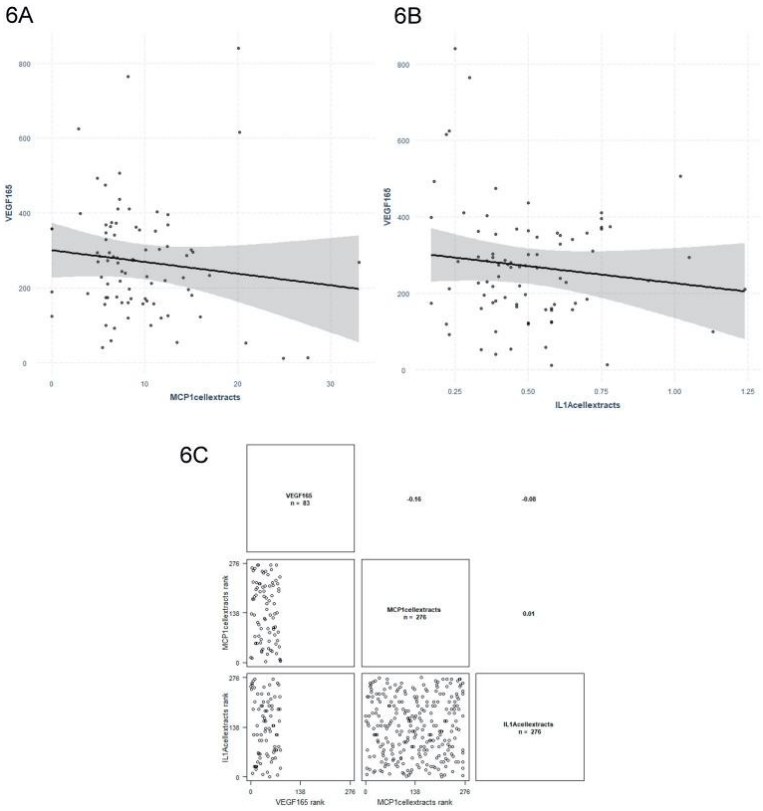

Fig 6. Regression graphs for $VEGF_{165}$ isoform mRNA levels as dependent variable and MCP1 (6A) and IL-1α (6B) cellular extracts levels as independent variables. The related correlation graph is presented in (6C). "**" $p \leq 0.01$.

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

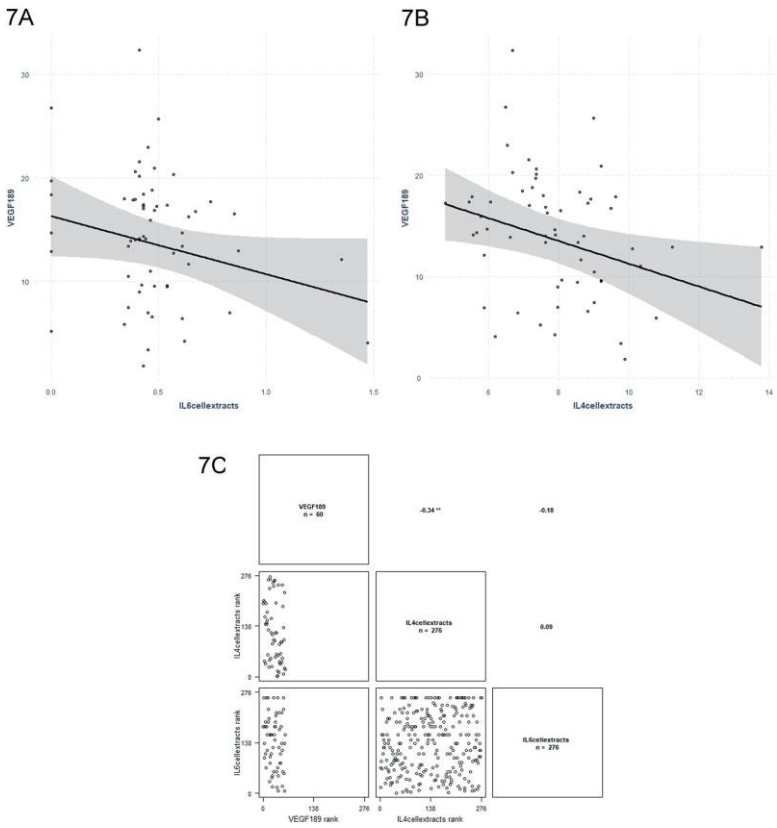

Fig 7. Regression graphs for $VEGF_{189}$ isoform mRNA levels as dependent variable and IL-6 (7A) and IL-4 (7B) cellular extracts levels as independent variables. The related correlation graph is presented in (7C).

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

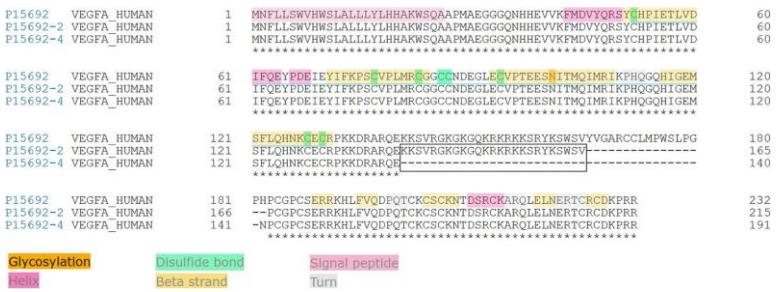

Fig 8. Alignment and elements of secondary structure of three VEGF isoforms (82% identity).

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 3

TERC variants associated with short leukocyte telomeres: implication of higher early life leukocyte telomere attrition as assessed by the blood-and-muscle model. [198]

Toupance S, Stathopoulou MGS, **Petrelis AM**, Gorenjak V, Labat C, Lai TP, Visvikis-Siest S, Benetos A.

Cells. 2020 May 9(6), 1360

My specific contribution in this study was: Participation in the analyses, in the preparation and the revision of the manuscript

Objective

Telomere length has an important genetic background and multiple GWAS studies have identified variants that are strongly associated with TL in leucocytes [41]. As leucocytes are highly proliferative tissue, LTL represents TL at any given moment of life, while TL in minimally proliferative tissues, such as skeletal muscle telomere length (MTL), represent an estimation of the TL at the beginning of life. Furthermore, the ratio LTL/MTL provides an estimation of TL attrition throughout life. The genetic background of these two TL indexes has not been studied before. Therefore, this investigation aimed to assess the associations between candidate genes for LTL with LTL, MTL, and LTL/MTL in a case-control population of atherosclerotic cardiovascular disease (ASCVD).

Population and analysis

The population was part of the TELARTA cohort and consisted of one discovery cohort of 259 individuals with (n=131) or without (n=128) clinical manifestation of atherosclerotic CVD, an independent replication cohort of 91 individuals, recruited under the same conditions and 52 individuals from an independent Greek population [42]. In total, 402 individuals were included and were genotyped for fifteen SNPs (rs11125529, rs6772228, rs12696304, rs10936599, rs7675998, rs7726159, rs2736100, rs9419958, rs9420907, rs4387287, rs3027234, rs8105767, rs412658, rs6028466 and rs755017), observed previously to be associated with LTL in several GWAS [43]. TL measurements was done by Southern blot analysis of terminal restriction fragments, as previously described [44].

The direct effects of variants on 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.

Results

Two LTL-related SNPs (rs12696304 and rs10936599), located on the *TERC* gene, were associated with the LTL/MTL ratio (p = 0.007 and p = 0.037, respectively) but not with MTL (p = 0.78 and p = 0.32 respectively). The alleles associated with shorter LTL were associated with a lower LTL/MTL ratio, thus signaling higher LTL attrition (Figure 16). No significant epistatic interaction was identified.

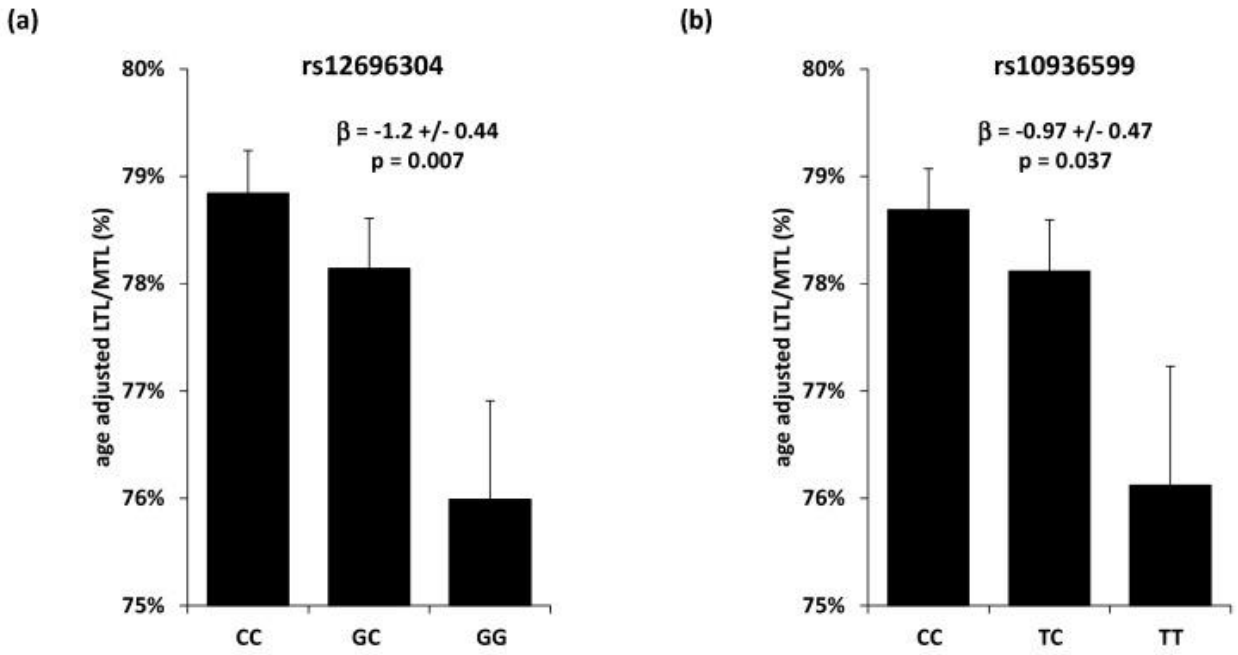

Figure 16. *TERC* variants association with telomere attrition (LTL/MTL)

(a) LTL/MTL ratio (expressed in %) according to rs12696304 genotype; (b) LTL/MTL ratio (expressed in %) according to rs10936599 genotype.

β: change in ratio per copy of the effect allele

Values are presented as mean +/− standard error of the mean

Analyses are age-adjusted.

Conclusions

We have demonstrated that the rs12696304 and rs10936599, which were previously associated with short LTL, do not affect TL at birth (estimated by MTL), but are associated with higher telomere attrition (estimated by LTL/MTL) during lifetime. Thus, polymorphisms located on genes coding for telomere maintenance proteins may contribute to higher TL attrition and could have an impact on the development of chronic diseases like ASCVD. Therefore, this study proposes novel insights in the association between short LTL and ASCVD and the identified variants could be tested for their predictive value for the development of atherosclerosis.

cells

MDPI

Communication

TERC Variants Associated with Short Leukocyte Telomeres: Implication of Higher Early Life Leukocyte Telomere Attrition as Assessed by the Blood-and-Muscle Model

Simon Toupance [1,2], Maria G. Stathopoulou [3], Alexandros M. Petrelis [3], Vesna Gorenjak [3], Carlos Labat [1], Tsung-Po Lai [4], Sophie Visvikis-Siest [3] and Athanase Benetos [1,2,*]

1 Université de Lorraine, Inserm, DCAC, F-54000 Nancy, France; s.toupance@chru-nancy.fr (S.T.); carlos.labat@inserm.fr (C.L.)
2 Université de Lorraine, CHRU-Nancy, Pôle "Maladies du Vieillissement, Gérontologie et Soins Palliatifs", F-54000 Nancy, France
3 Université de Lorraine, IGE-PCV, F-54000 Nancy, France; maria.stathopoulou@inserm.fr (M.G.S.); apetrelis@live.com (A.M.P.); gorenjak.vesna@gmail.com (V.G.); sophie.visvikis-siest@inserm.fr (S.V.-S.)
4 Center of Human Development and Aging, Rutgers, New Jersey Medical School, The State University of New Jersey, Newark, NJ 07103, USA; tl599@njms.rutgers.edu
* Correspondence: a.benetos@chru-nancy.fr; Tel.: +33-383-15-33-22; Fax: +33-383-15-76-68

Received: 29 April 2020; Accepted: 27 May 2020; Published: 31 May 2020

Abstract: Short leukocyte telomere length (LTL) is associated with atherosclerotic cardiovascular disease (ASCVD). Mendelian randomisation studies, using single nucleotide polymorphisms (SNPs) associated with short LTL, infer a causal role of LTL in ASCVD. Recent results, using the blood-and-muscle model, indicate that higher early life LTL attrition, as estimated by the ratio between LTL and skeletal muscle telomere length (MTL), rather than short LTL at conception, as estimated by MTL, should be responsible of the ASCVD-LTL connection. We combined LTL and MTL measurements and SNPs profiling in 402 individuals to determine if 15 SNPs classically described as associated with short LTL at adult age were rather responsible for higher LTL attrition during early life than for shorter LTL at birth. Two of these SNPs (rs12696304 and rs10936599) were associated with LTL in our cohort ($p = 0.027$ and $p = 0.025$, respectively). These SNPs, both located on the *TERC* gene, were associated with the LTL/MTL ratio ($p = 0.007$ and $p = 0.037$, respectively), but not with MTL ($p = 0.78$ and $p = 0.32$ respectively). These results suggest that SNPs located on genes coding for telomere maintenance proteins may contribute to a higher LTL attrition during the highly replicative first years of life and have an impact later on the development of ASCVD.

Keywords: telomere length maintenance; SNPs; TERC; atherosclerosis; cohort studies

1. Introduction

Short leukocyte telomere length (LTL) is associated with atherosclerotic cardiovascular disease (ASCVD) and its complications [1,2]. Genome-wide association studies (GWAS) have found several single nucleotide polymorphisms (SNPs) associated with short LTL over the years [3,4] and recently Mendelian randomisation studies showed that these genetic variants are associated with ASCVD [5–7] inferring a causal role of short LTL in atherosclerosis. The hypothesis is that inherited short LTL is associated with ASCVD. However, we recently showed, using the blood and muscle model, that the association between short LTL and ASCVD was not related to shorter telomere length (TL) at the beginning of life—as estimated by minimally proliferative skeletal muscle telomere length (MTL)—but

to higher LTL attrition, especially during the first years of life, as estimated by the LTL/MTL ratio [8]. This lead us to the hypothesis that genetic variants associated with short LTL measured at adult age identified by GWAS could be related to higher LTL attrition during early life rather than to shorter LTL at conception. In this genetic study, we tested this hypothesis by combining SNPs profiling of 15 variants associated with short LTL and telomere length measurements in leukocytes and skeletal muscles. We found that SNPs associated with short LTL in our cohort, *TERC* variants rs12696304 and rs10936599, are linked with higher estimated early life LTL attrition rather than with shorter estimated LTL at birth.

2. Materials and Methods

2.1. The Cohort

The aim of the TELARTA (Telomere in Arterial Aging) study is to examine the implication of telomere dynamics in arterial aging and the development of atherosclerosis using the blood and muscle model. The study and its goals have been described previously [8]. Briefly, 259 men and women (older than 20 years), who were admitted for various surgical procedures, were enrolled in university hospitals in Nancy (n = 215) and Marseille (n = 44), France. A replication cohort of 143 individuals was enrolled in 3 sites: 91 individuals in original sites (Nancy and Marseille) and 52 individuals in 3 Athens hospitals (Onassis Cardiac Surgery Center, Surgeon KP; Iaso General Hospital Surgeon MVG; Hippokration Hospital, Surgeon EM). All 402 subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by Ethics Committee (Comité de Protection des Personnes) of Nancy, France, Ethics Committee of the University of Athens and Ethics Committee of each one of the 3 participating hospitals. This study is registered on http://www.clinicaltrials.gov under unique identifier: NCT02176941.

2.2. Telomere Length Measurements

Leukocyte TL (LTL) and Muscle TL (MTL) were measured in DNA extracted by the phenol/chloroform/isoamyl alcohol method from peripheral blood leukocytes and skeletal muscle biopsies (~ 100 mg in the surgical field), respectively. DNA samples passed an integrity testing using a 1% (wt/vol) agarose gel before TL measurements performed by Southern blotting of the terminal restriction fragments (TRFs), as described previously [9]. Briefly, DNA samples were digested (37 °C) overnight with the restriction enzymes Hinf I and Rsa I (Roche Diagnostics GmbH, Mannheim, Germany). Digested DNA samples and DNA ladders were resolved on 0.5% (wt/vol) agarose gels during 23 h. DNA was depurinated, denatured, neutralised and transferred onto a positively charged nylon membrane (Roche Diagnostics GmbH) using a vacuum blotter (Biorad, Hercules, CA, USA). Membranes were hybridised at 65 °C with a DIG-labelled telomeric probe after which the probe was detected by the DIG luminescent detection procedure (Roche Diagnostics GmbH) and exposed on CCD camera (Las 4000, Fujifilm Life Sciences, Cambridge, MA, USA). Optical density values (OD) versus DNA migration distances (in pixels) were obtained from pictures with image processing software (Multigauge, Fujifilm Life Sciences). Distances were converted to molecular weight (MW) in kilobases (kb) using a power function transformation owing to DNA ladders. Mean TRF length was calculated in the 3–20 kb range using the formula mean TRF (kb) = Σ (ODi)/Σ (ODi/MWi), where ODi is the OD at position i and MWi is the MW (i.e., TRF length) at that position. Leukocyte and muscle DNA from one patient were always measured simultaneously on the same membrane, as shown in the representative blot (Figure 1), and measurements were performed in duplicate on separate membranes. 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 2 duplicates (used in the analysis), known as the

extrapolated repeatability, was 0.995 and 0.991 for LTL and MTL, respectively. The LTL/MTL ratio was calculated.

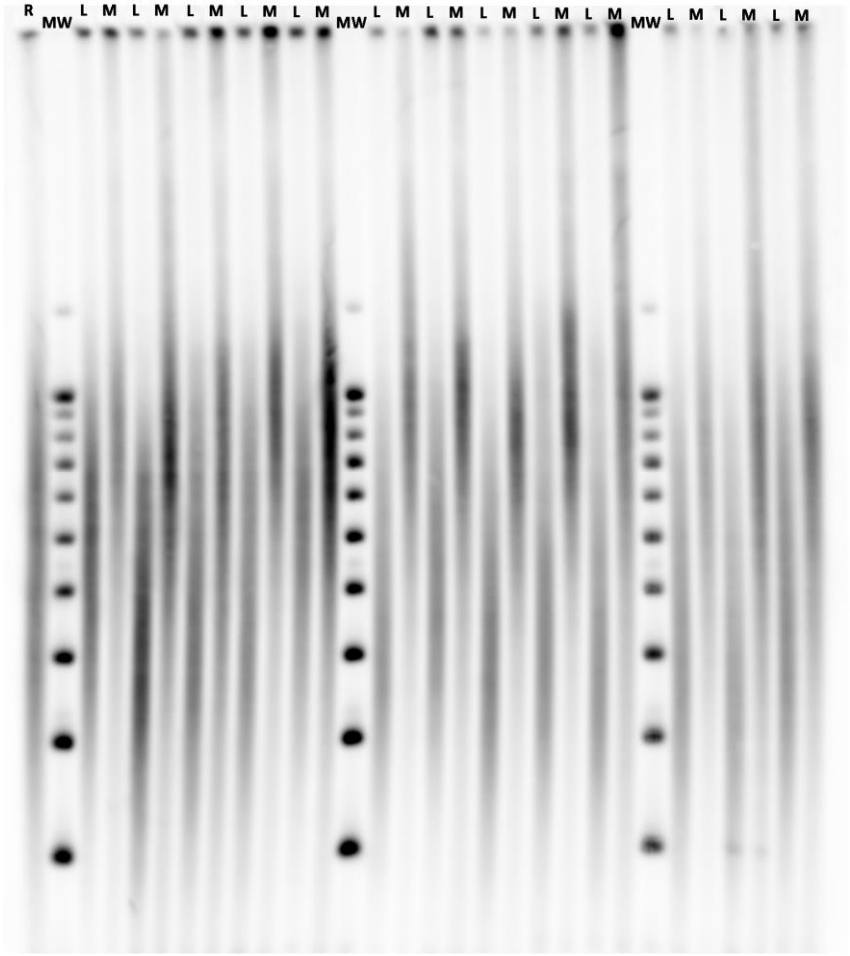

Figure 1. Illustrative membrane of telomere length measurements performed by Southern blotting of the terminal restriction fragments (TRFs). A sample of known telomere length serves as an internal reference (**R**). Three molecular weight ladders (**MW**) are resolved across the gel, and the one that is closest to a given sample is used for the computation of its mean TRF. Leukocyte (**L**) and Muscle (**M**) DNA samples of one participant are resolved on adjacent lanes.

2.3. Single Nucleotide Polymorphism Genotyping

Fifteen SNPs (rs11125529, rs6772228, rs12696304, rs10936599, rs7675998, rs7726159, rs2736100, rs9419958, rs9420907, rs4387287, rs3027234, rs8105767, rs412658, rs6028466 and rs755017), observed previously to be associated with LTL in several GWAS [5], were genotyped in DNA samples using a PCR-based KASP assay [10]. Minor allele frequencies (MAF) were calculated and the Hardy-Weinberg equilibrium was tested using the χ^2 method. SNPs which did not follow Hardy-Weinberg equilibrium were excluded in the subsequent analyses.

2.4. Statistical Analyses

Continuous variables are presented as means ± SD or mean ± SE and discrete variables as frequencies or percentages. Pairwise comparisons were performed using the Mann–Whitney and χ^2 tests, as appropriate. Telomere length values are presented and compared with or without adjustment to age and sex. Bivariate relations between continuous variables were determined using Pearson correlation coefficients. Statistical analyses were performed using the Number Crunching Statistical System (NCSS) 9 statistical software package (NCSS, Kaysville, UT, USA).

For SNP analyses, LTL was transformed into logLTL to normalize its distribution. Statistical analyses were performed using linear regression models adjusted for age and sex for LTL and MTL, and age only for the LTL/MTL ratio (MTL and LTL but not LTL/MTL are sex-dependant). Direct SNP associations were examined with PLINK software in an additive model (using minor alleles as reference alleles) and epistatic interactions with PLINK and R softwares and the CAPE R package. Taking into account the limited sample size, and in order to limit multiple testing, the LTL phenotype was initially tested and only SNPs associated with LTL in our cohort were tested against MTL and LTL/MTL phenotypes.

3. Results

3.1. Population Characteristics

Participants were of European ancestry, aged 60 ± 15 years (with an age range of 20–94 years) and 32% were females. Among them, 48% displayed clinical manifestations of ASCVD in either coronary, carotid, cerebral, iliac, femoral or popliteal arteries.

3.2. Telomere Length Dynamics

LTL (mean ± SD; 6.71 ± 0.84 kb) was shorter than MTL (8.57 ± 0.72 kb; $p < 0.0001$) in all individuals. Mean LTL/MTL ratio was 0.78 ± 0.07. Both LTL and MTL shortened with age with a slope of the effect of age steeper for LTL than MTL (−31 bp/year and −15 bp/year, respectively). The LTL/MTL ratio decreased with age (−0.0021 ± 0.0002/year). Age-adjusted LTL was longer in women than men (6.86 ± 0.06 kb and 6.64 ± 0.04 kb, respectively; $p < 0.005$). Similarly, age-adjusted MTL was longer in women than men (8.74 ± 0.06 kb and 8.49 ± 0.04 kb, respectively; $p < 0.001$), whereas age-adjusted LTL/MTL was not influenced by sex (0.78 ± 0.005 in women vs. 0.78 ± 0.004 in men; $p = 0.69$).

3.3. Single Nucleotide Polymorphisms Association with Telomere Length

The 13 SNPs in accordance with Hardy–Weinberg equilibrium (rs11125529, rs6772228, rs12696304, rs10936599, rs7675998, rs2736100, rs9419958, rs9420907, rs4387287, rs3027234, rs8105767, rs412658 and rs6028466) were included in the analyses. Two of these SNPs, rs12696304 and rs10936599, were significantly associated with LTL ($p = 0.027$ and $p = 0.025$, respectively, Table 1). These SNPs map to a locus harbouring *TERC*, the gene encoding the RNA subunit of telomerase. The minor alleles, allelic frequencies, and directions of the associations were similar to previous findings [5]. These two SNPs showed no association with MTL ($p = 0.78$ and $p = 0.32$, respectively, Table 2) but were associated with the LTL/MTL ratio ($p = 0.007$ and $p = 0.037$, respectively), in a way that the allele associated with shorter LTL was associated with a lower LTL/MTL ratio, thus signalling higher LTL attrition (Figure 2). No significant epistatic interaction was identified.

Table 1. Single nucleotide polymorphisms association with leukocyte telomere length (LTL).

							Log LTL		
SNP	**CHR**	**Position**	**Gene**	**Effect Allele**	**Other Allele**	**MAF**	**BETA**	**SE**	**P**
rs11125529	2	54248729	*ACYP2*	A	C	0.12	0.000024	0.0050	0.99
rs6772228	3	58390292	*PXK*	A	T	0.045	−0.0094	0.0073	0.20
rs12696304	3	169763483	*TERC*	G	C	0.29	−0.0074	0.0033	0.027
rs10936599	3	169774313	*TERC*	T	C	0.25	−0.0080	0.0036	0.025
rs7675998	4	163086668	*NAF1*	A	G	0.24	−0.0012	0.0037	0.74
rs2736100	5	1286401	*TERT*	A	C	0.47	−0.0038	0.0032	0.24
rs9419958	10	103916188	*OBFC1*	T	C	0.16	−0.00093	0.0044	0.84
rs9420907	10	103916707	*OBFC1*	C	A	0.17	0.00032	0.0042	0.94
rs4387287	10	103918139	*OBFC1*	A	C	0.21	0.0038	0.0039	0.32
rs3027234	17	8232774	*CTC1*	T	C	0.21	0.0017	0.0038	0.66
rs8105767	19	22032639	*ZNF208*	G	A	0.30	−0.0027	0.0034	0.42
rs412658	19	22176638	*ZNF676*	T	C	0.39	−0.0016	0.0032	0.63
rs6028466	20	39500359	*DHX35*	A	G	0.070	−0.0022	0.0062	0.72

LTL = Leukocyte telomere length, SNP = Single nucleotide polymorphism, CHR = Chromosome, Position = base-pair position (GRCh38.p3), MAF = minor allelic frequency, BETA = regression coefficient (change in telomere length per copy of the effect allele), SE = standard error. LTL values are transformed in Log LTL to normalize their distribution. Analyses are age- and sex-adjusted.

Table 2. TERC single nucleotide polymorphisms (SNPs) association with muscle telomere length (MTL).

						MTL		
SNP	**CHR**	**Position**	**Gene**	**Effect Allele**	**Other Allele**	**BETA**	**SE**	**P**
rs12696304	3	169763483	*TERC*	G	C	−0.014	0.051	0.78
rs10936599	3	169774313	*TERC*	T	C	−0.054	0.054	0.32

MTL = Muscle telomere length, SNP = Single nucleotide polymorphism, CHR = Chromosome, Position = base-pair position (GRCh38.p3), BETA = regression coefficient (change in telomere length per copy of the effect allele), SE = standard error. Analyses are age- and sex-adjusted.

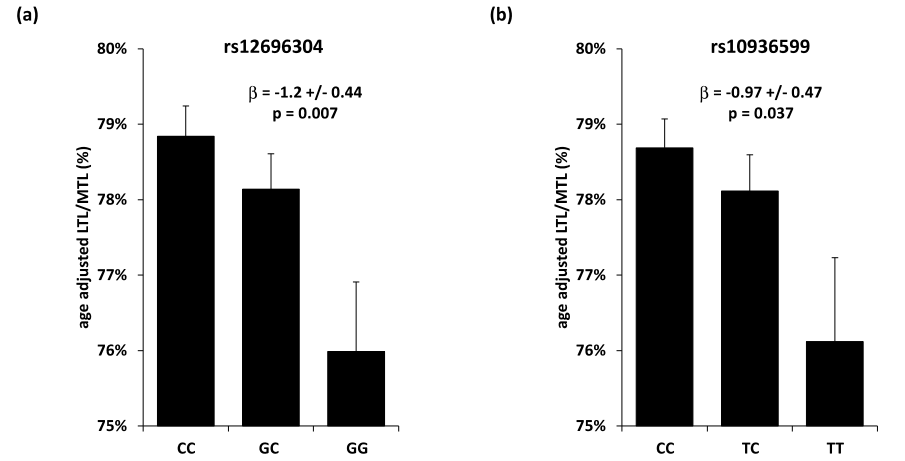

Figure 2. *TERC* SNPs association with LTL/MTL (**a**) LTL/MTL ratio (expressed in %) according to Table 12696304. (*TERC*) genotype; (**b**) LTL/MTL ratio (expressed in %) according to the rs10936599 (*TERC*) genotype. LTL = Leukocyte telomere length, MTL = Muscle telomere length, β = change in ratio per copy of the effect allele Values are presented as mean +/− SEM. Analyses are age-adjusted.

4. Discussion

In our cohort of patients, SNPs that were found to be associated with short LTL are associated with a lower LTL/MTL ratio but are not associated with MTL. The present results indicate that rs12696304 and rs10936599, two *TERC* SNPs previously described as associated with short LTL, do not influence TL at conception as reflected by MTL, but rather higher LTL attrition from conception onwards as expressed in the LTL/MTL ratio. We have reported previously that the gap between LTL and MTL is essentially constituted during the first decade of life [11]. SNPs located on genes coding for telomere maintenance proteins may thus contribute to higher TL attrition during this critical highly replicative period and have an impact later on the development of ASCVD [12–14].

The observed associations of these SNPs with LTL but not MTL may also suggest that person-to-person variation in TL dynamics specifically in the hematopoietic system but not in skeletal muscle, and presumably not in most other tissues, might contribute to the pathogenesis of atherosclerosis. This contribution may possibly be mediated through inflammation regulation [15]. Indeed, short telomeres have been shown to upregulate the NLRP3 (NOD (nucleotide oligomerization domain)-, LRR (leucine-rich repeat)-, and PYD (pyrin domain)-containing protein 3) inflammasome of macrophages in a murine model [16]. NLRP3, an innate immune signalling complex, is known to activate the proinflammatory IL-1 pathway, which contribute to atherogenesis [17]. Moreover, clonal haematopoiesis of indeterminate potential (CHIP), a phenomenon which is associated with a doubling of the risk of ASCVD [18,19], has been recently linked to short LTL [20,21] and the upregulation of the macrophage NLRP3 inflammasome in human [22,23]. Another possible explanation of the link observed between short LTL and ASCVD is through the hemothelium paradigm. As the hematopoietic system and vascular endothelium share a common precursor, the hemogenic endothelium or hemothelium [24], short LTL might reflect short TL in progenitor cells involved in endothelial repair. Compromised proliferation capacity of endothelial progenitor cells with short TL may influence endothelium repair capacity and impact ASCVD progression [25].

This study is limited by the small sample size in regard to SNPs analyses, but it is difficult to obtain very large number of participants due to the heavy procedure needed to obtain results for MTL. However, this drawback was offset at least for TL measurements by the use of high reproducibility TRF method [26,27].

In conclusion, taking advantage from the data obtained by the blood and muscle telomere model and those obtained from Mendelian randomisation studies, the present study proposes a new mechanical insight in the association between short LTL and ASCVD. Our findings suggest that *TERC* variants might impact early life TL attrition in hematopoietic cells and contribute later to the pathogenesis of atherosclerosis.

Author Contributions: S.T. and A.B. conceived and designed the study. S.T. and T.-P.L. performed the T.-P.L. measurements. M.G.S. and V.G. performed the SNP profiling. S.T., M.G.S., A.M.P. and C.L. performed the analyses. M.G.S., A.M.P., V.G. and C.L. contributed to the manuscript preparation. S.T. and A.B. wrote the manuscript with input from all co-authors. A.B. and S.V.-S. oversaw all experimental design, data analysis, and manuscript preparation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the French National Research Agency (ANR), Translationnelle: N°ID RCB: 2014-A00298-39: 2014-2017, the French regional project CPER-ITM2P 2015-2020, the French PIA project « Lorraine Université d'Excellence », reference ANR-15-IDEX-04-LUE and the Investments for the Future program under grant agreement No. ANR-15-RHU-0004.

Acknowledgments: Members of the TELARTA consortium: S. Gautier, G. Watfa, P. Lacolley, V. Regnault, H. Louis, C. Lakomy, M. Folio, M. Schmitt, A. Didier, S. Vauthier, N. François, A. Aviv, M. Kimura, P.M. Rossi, F. Dignat George, F. Sabatier, S. Simoncini, C. Dumoulin, P. Béranger, S. Malikov, N. Settembre, J. Hubert, L. Frimat, B. Bertrand, M. Boufi, X. Flecher, N. Sadoul, P. Eschwege, M. Kessler, I.P. Tzanetakou, I.P. Doulamis, P. Konstantopoulos, A. Tzani, M. Korou, A. Gkogkos, K. Perreas, E. Menenakos, G. Samanidis, M. Vasiloglou-Gkanis, J.D. Kark, S. Verhulst.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. D'Mello, M.J.J.; Ross, S.A.; Briel, M.; Anand, S.S.; Gerstein, H.; Paré, G. Association between shortened leukocyte telomere length and cardiometabolic outcomes: Systematic review and meta-analysis. *Circ. Cardiovasc. Genet.* **2015**, *8*, 82–90. [CrossRef] [PubMed]
2. Haycock, P.C.; Heydon, E.E.; Kaptoge, S.; Butterworth, A.S.; Thompson, A.; Willeit, P. Leucocyte telomere length and risk of cardiovascular disease: Systematic review and meta-analysis. *BMJ* **2014**, *349*, g4227. [CrossRef] [PubMed]
3. Codd, V.; Nelson, C.P.; Albrecht, E.; Mangino, M.; Deelen, J.; Buxton, J.L.; Hottenga, J.J.; Fischer, K.; Esko, T.; Surakka, I.; et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* **2013**, *45*, 422–427. [CrossRef] [PubMed]
4. Mangino, M.; Hwang, S.-J.; Spector, T.D.; Hunt, S.C.; Kimura, M.; Fitzpatrick, A.L.; Christiansen, L.; Petersen, I.; Elbers, C.C.; Harris, T.; et al. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum. Mol. Genet.* **2012**, *21*, 5385–5394. [CrossRef] [PubMed]
5. Haycock, P.C.; Burgess, S.; Nounu, A.; Zheng, J.; Okoli, G.N.; Bowden, J.; Wade, K.H.; Timpson, N.J.; Evans, D.M.; Willeit, P.; et al. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA Oncol.* **2017**, *3*, 636–651. [CrossRef] [PubMed]
6. Scheller Madrid, A.; Rode, L.; Nordestgaard, B.G.; Bojesen, S.E. Short Telomere Length and Ischemic Heart Disease: Observational and Genetic Studies in 290 022 Individuals. *Clin. Chem.* **2016**, *62*, 1140–1149. [CrossRef] [PubMed]
7. Zhan, Y.; Karlsson, I.K.; Karlsson, R.; Tillander, A.; Reynolds, C.A.; Pedersen, N.L.; Hägg, S. Exploring the Causal Pathway From Telomere Length to Coronary Heart Disease: A Network Mendelian Randomization Study. *Circ. Res.* **2017**, *121*, 214–219. [CrossRef]
8. Benetos, A.; Toupance, S.; Gautier, S.; Labat, C.; Kimura, M.; Rossi, P.M.; Settembre, N.; Hubert, J.; Frimat, L.; Bertrand, B.; et al. Short Leukocyte Telomere Length Precedes Clinical Expression of Atherosclerosis: The Blood-and-Muscle Model. *Circ. Res.* **2018**, *122*, 616–623. [CrossRef]
9. Kimura, M.; Stone, R.C.; Hunt, S.C.; Skurnick, J.; Lu, X.; Cao, X.; Harley, C.B.; Aviv, A. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nat. Protoc.* **2010**, *5*, 1596–1607. [CrossRef]
10. He, C.; Holme, J.; Anthony, J. SNP genotyping: The KASP assay. *Methods Mol. Biol.* **2014**, *1145*, 75–86. [CrossRef]
11. Sabharwal, S.; Verhulst, S.; Guirguis, G.; Kark, J.D.; Labat, C.; Roche, N.E.; Martimucci, K.; Patel, K.; Heller, D.S.; Kimura, M.; et al. Telomere length dynamics in early life: The blood-and-muscle model. *FASEB J.* **2018**, *32*, 529–534. [CrossRef] [PubMed]
12. Barraclough, J.Y.; Skilton, M.R.; Garden, F.L.; Toelle, B.G.; Marks, G.B.; Celermajer, D.S. Early and late childhood telomere length predict subclinical atherosclerosis at age 14 yrs.-The CardioCAPS study. *Int. J. Cardiol.* **2019**, *278*, 250–253. [CrossRef] [PubMed]
13. Skilton, M.R.; Nakhla, S.; Ayer, J.G.; Harmer, J.A.; Toelle, B.G.; Leeder, S.R.; Jones, G.; Marks, G.B.; Celermajer, D.S.; Childhood Asthma Prevention Study Group. Telomere length in early childhood: Early life risk factors and association with carotid intima-media thickness in later childhood. *Eur. J. Prev. Cardiol.* **2016**, *23*, 1086–1092. [CrossRef] [PubMed]
14. Toupance, S.; Labat, C.; Temmar, M.; Rossignol, P.; Kimura, M.; Aviv, A.; Benetos, A. Short Telomeres, but Not Telomere Attrition Rates, Are Associated With Carotid Atherosclerosis. *Hypertension* **2017**, *70*, 420–425. [CrossRef] [PubMed]
15. Chiriacò, M.; Georgiopoulos, G.; Duranti, E.; Antonioli, L.; Puxeddu, I.; Nannipieri, M.; Rosada, J.; Blandizzi, C.; Taddei, S.; Virdis, A.; et al. Inflammation and Vascular Ageing: From Telomeres to Novel Emerging Mechanisms. *High Blood Press. Cardiovasc. Prev. Off. J. Ital. Soc. Hypertens.* **2019**, *26*, 321–329. [CrossRef] [PubMed]
16. Kang, Y.; Zhang, H.; Zhao, Y.; Wang, Y.; Wang, W.; He, Y.; Zhang, W.; Zhang, W.; Zhu, X.; Zhou, Y.; et al. Telomere Dysfunction Disturbs Macrophage Mitochondrial Metabolism and the NLRP3 Inflammasome through the PGC-1α/TNFAIP3 Axis. *Cell Rep.* **2018**, *22*, 3493–3506. [CrossRef]
17. Grebe, A.; Hoss, F.; Latz, E. NLRP3 Inflammasome and the IL-1 Pathway in Atherosclerosis. *Circ. Res.* **2018**, *122*, 1722–1740. [CrossRef]

18. Amorós-Pérez, M.; Fuster, J.J. Clonal hematopoiesis driven by somatic mutations: A new player in atherosclerotic cardiovascular disease. *Atherosclerosis* **2020**, *297*, 120–126. [CrossRef]
19. Jaiswal, S.; Natarajan, P.; Silver, A.J.; Gibson, C.J.; Bick, A.G.; Shvartz, E.; McConkey, M.; Gupta, N.; Gabriel, S.; Ardissino, D.; et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N. Engl. J. Med.* **2017**, *377*, 111–121. [CrossRef]
20. Zink, F.; Stacey, S.N.; Norddahl, G.L.; Frigge, M.L.; Magnusson, O.T.; Jonsdottir, I.; Thorgeirsson, T.E.; Sigurdsson, A.; Gudjonsson, S.A.; Gudmundsson, J.; et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **2017**, *130*, 742–752. [CrossRef]
21. Schratz, K.E.; Haley, L.; Danoff, S.K.; Blackford, A.; DeZern, A.; Gocke, C.D.; Duffield, A.S.; Armanios, M. Cancer spectrum and outcomes in the Mendelian short telomere syndromes. *Blood* **2020**. [CrossRef] [PubMed]
22. Aviv, A.; Levy, D. Hemothelium, Clonal Hematopoiesis of Indeterminate Potential, and Atherosclerosis. *Circulation* **2019**, *139*, 7–9. [CrossRef] [PubMed]
23. Steensma, D.P. Clinical Implications of Clonal Hematopoiesis. *Mayo Clin. Proc.* **2018**, *93*, 1122–1130. [CrossRef] [PubMed]
24. Lacaud, G.; Kouskoff, V. Hemangioblast, hemogenic endothelium, and primitive versus definitive hematopoiesis. *Exp. Hematol.* **2017**, *49*, 19–24. [CrossRef]
25. Aviv, A.; Levy, D. Telomeres, atherosclerosis, and the hemothelium: The longer view. *Annu. Rev. Med.* **2012**, *63*, 293–301. [CrossRef]
26. Aviv, A.; Hunt, S.C.; Lin, J.; Cao, X.; Kimura, M.; Blackburn, E. Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR. *Nucleic Acids Res.* **2011**, *39*, e134. [CrossRef]
27. Nettle, D.; Seeker, L.; Nussey, D.; Froy, H.; Bateson, M. Consequences of measurement error in qPCR telomere data: A simulation study. *PLoS ONE* **2019**, *14*, e0216118. [CrossRef]

© 2020 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 4

A genetic determinant of VEGF-A levels is associated with telomere attrition. [199]

Gorenjak V *, **Petrelis AM***, Stathopoulou MG**, Toupance S**, Labat C, Masson C, Murray H, Lamont J, Fitzgerald P, Benetos A***, Visvikis-SiestS***.

Aging (Albany NY). 2021 Oct 18;13(20):23517-23526

* equal first authors, **equal second authors, ***equal last authors

My specific contribution in this study was: Participation in the conception of the research idea, support the writing of the manuscript, the conception and the realisation of analyses, the interpretation of the results, and the revision of the manuscript.

Objective

VEGF-A and TL are biomarkers that are implicated in chronic diseases pathways, including inflammation and angiogenesis. They both have a strong heritable background [13, 45] and significant inter-individual variability [46, 47], thus, the identification of common genetic determinants for these two biomarkers could facilitate their use in personalised prediction of chronic diseases. This study aimed to expand the knowledge of the common pathways that regulate both biomarkers through the investigation of the genetic associations between VEGF-A and TL.

Population and analysis

The population of the previous study (Publication 3) was used in this investigation. It was part of the TELARTA cohort and consisted of one discovery cohort of 259 individuals with (n=131) or without (n=128) clinical manifestation of atherosclerotic CVD, an independent replication cohort of 91 individuals, recruited under the same conditions and 52 individuals from an independent Greek population [42]. In total, 402 individuals were included and were genotyped for ten VEGF-A-related polymorphisms: rs10761741, rs10738760, rs6921438, rs7043199, rs6993770, rs4416670, rs114694170, rs34528081, rs4782371 and rs2639990.

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) and TL measurements was done by Southern blot analysis of terminal

restriction fragments, as previously described [44].

The direct effects of VEGF-A related genetic variants on three phenotypes of interest (LTL, MTL and LTL/MTL) were tested using linear regression models adjusted for age and sex with the PLINK toolset, while epistatic interactions were tested using R package CAPE.

Results

We found no significant association between the VEGF-A-related variants and LTL and MTL. However, the minor allele T of the polymorphism rs6993770 was significantly associated with increased LTL/MTL (P=0.001143, β=0.0148, SE=0.004516) (Table 8). No significant epistatic interaction was identified.

Also, there was no significant association between VEGF-A plasma concentrations and the investigated TL phenotypes (Table 9).

Table 8. Associations 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,

* models were adjusted for age and sex

**significant P-value

Table 9. Univariate and multiple regression analysis of VEGF-A plasma concentration with LTL, MTL and LTL/MTL.

Dependent variable: logVEGF-A.

Independent Variable*	Regression Coefficient (SE)	R^2	P-Value
Non adjusted			
LTL	0.019 (0.160)	0.19%	0.42
MTL	-0.009 (0.028)	0.03%	0.74
LTL/MTL	0.471 (0.300)	0.73%	0.12
Adjusted*			
LTL	-0.075 (0.107)	0.15%	0.49
MTL	-0.170 (0.105)	0.81%	0.36
LTL/MTL	0.756 (0.919)	0.21%	0.41

SE: Standard error

* each TL variable was tested as independent variable in separate models. Thus, the table presents the results of 3 different models

** Models were adjusted for age, sex and statin use.

Conclusion

This is the first study of the associations of VEGF-A related genetic variants with LTL, MTL and telomere attrition. A significant association between rs6993770 and telomere attrition has been found.

The rs6993770 is located on the intronic region of the *ZFPM2* (zinc finger protein, FOG family member 2) gene. The minor allele T of this polymorphism has been previously associated with decreased VEGF-A levels [14] and also with HDL levels [12], erythrocyte count, IL-12 levels, and platelets [48]. This investigation demonstrated that the allele T had a protective role in telomere attrition. This association could be explained through the modification of the expression of GATA proteins, which could result in a direct impact on hematopoiesis and production of platelets. The rs6993770 could affect the activity of the *ZFPM2* gene, which affects GATA protein regulation of haematopoiesis. The risk allele A could lead to increased haematopoiesis, thus higher cellular turnover and faster telomere attrition [49].

The lack of association between VEGF-A plasma levels and LTL, MTL or estimated telomere attrition leads to the conclusion that the effect of rs6993770 on telomere attrition is not directly linked to VEGF-A levels and is probably mediated by other mechanisms, such as the increased haematopoiesis that we propose here. Previous scientific reports on the association between LTL and VEGF-A levels have reported inconsistent findings, thus more studies are required to fully investigate the association of these biomarkers.

In conclusion, this polymorphism could be an important marker in the development of prediction tools for chronic diseases where both TL and VEGF-A are involved, like CVD.

www.aging-us.com AGING 2021, Vol. 13, No. 20

Research Paper

A genetic determinant of VEGF-A levels is associated with telomere attrition

Vesna Gorenjak[1,*,#], Alexandros M. Petrelis[1,*,#], Maria G. Stathopoulou[1,*], Simon Toupance[2,*], Satish Kumar[1], Carlos Labat[2], Christine Masson[1], Helena Murray[3], John Lamont[3], Peter Fitzgerald[3], Athanase Benetos[2,4,*], Sophie Visvikis-Siest[1,*], TELARTA Consortium

[1]Université de Lorraine, IGE-PCV, Nancy F-54000, France
[2]Université de Lorraine, Inserm, DCAC, Nancy F-54000, France
[3]Randox Laboratories Limited, Crumlin, Co. Antrim BT29 4QY, Northern Ireland, United Kingdom
[4]Université de Lorraine, CHRU-Nancy, Pôle "Maladies du Vieillissement, Gérontologie et Soins Palliatifs", Nancy F-54000, France
*Equal contribution
[#]Co-first author

Correspondence to: Sophie Visvikis-Siest; **email:** sophie.visvikis-siest@inserm.fr
Keywords: VEGF-A, telomere length, telomere attrition, accelerated cellular aging
Received: March 18, 2021 **Accepted:** October 3, 2021 **Published:** October 18, 2021

Copyright: © 2021 Gorenjak et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Telomere length (TL) is a hallmark of cellular aging and is associated with chronic diseases development. The vascular endothelial growth factor A (VEGF-A), a potent angiogenesis factor, is implicated in the pathophysiology of many chronic diseases. The aim of the present study was to investigate the associations between VEGF-A and TL.

TL in leukocytes (LTL) and skeletal muscle (MTL) were measured, 10 VEGF-related polymorphisms genotyped, and VEGF-A plasma concentrations determined in 402 individuals from the TELARTA cohort. LTL/MTL ratio was calculated as an estimate of lifelong TL attrition. Associations between VEGF-A variants and levels, and TL parameters were investigated.

We identified one significant association between the minor allele (T) of rs6993770 variant and LTL/MTL ratio (P=0.001143, β=0.0148, SE=0.004516). The rs6993770 is an intronic variant of the *ZFPM2* gene, which is involved in haematopoiesis and the identified association with increased telomere attrition could be due to increased haematopoiesis. No significant epistatic interaction was identified, and no association was found between levels of VEGF-A and any of assessed phenotypes.

We identified a potential common genetic regulation between VEGF-A and telomere length attrition that could be explained by mechanisms of increased hematopoiesis and production of platelets. VEGF-A and TL could play an important role in personalized medicine of chronic diseases and identification of molecular links between them can promote the understanding of their complex implications.

INTRODUCTION

Telomeres are repeated sequences of TTAGGG nucleotides at the end of eukaryotic chromosomes, participating in the maintenance of chromosomal integrity during cell replication [1]. As telomere length (TL) is gradually shortening with each mitotic division it has been considered as an indicator of accelerated

cellular aging. Inflammation and oxidative stress are increasing the shortening rate of TL and further affect the aging process and are considered as a basic link between TL and many chronic diseases [2]. Short leukocyte TL (LTL) is associated with the pathogenesis of atherosclerosis [3]. 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 [4]. TL has also been associated with other chronic diseases such as cancer [5], osteoporosis [6], chronic kidney diseases [7], neurodegenerative diseases [8] and others.

LTL is known for its high interindividual variation, which is evident already at birth [9]. Afterwards, numerous factors may accelerate LTL attrition, e.g. increased cellular turnover caused by increased inflammation, oxidative stress [10], or may have a protective effect on LTL shortening, e.g. regular sport activity, healthy diet [11]. Although the effects of these factors on LTL are rather small compared to the large variation observed in the baseline TL among individuals [12], the use of LTL alone as a biomarker could be not as informative as it would be in combination with an estimate of lifelong telomere attrition [13]. As skeletal muscle is a minimally proliferative tissue, skeletal muscle TL (MTL) can be considered as a baseline TL proxy, and estimated lifelong TL attrition can be calculated by the LTL/MTL ratio [14].

The vascular endothelial growth factor A (VEGF-A) is a key regulator of physiological and pathological angiogenesis and a mediator of vascular permeability in inflammatory disorders [15]. It plays a critical role in cancer and several VEGF-A signaling inhibitors have already been approved by the Food and Drug Administration (FDA) for oncological treatment [16]. VEGF-A also seems to be important in cardiovascular diseases (CVD) [17], including atherosclerosis, where it promotes neoangiogenesis in the early stages of the disease as well as in the advanced atherosclerotic plaques [18].

VEGF-A and TL are both involved in chronic diseases pathways, such as inflammation and angiogenesis. They have been commonly assessed in studies of different conditions [19–22], however, their associations have not been extensively studied [10, 23]. Furthermore, these biomarkers are highly heritable [24, 25] and have an important inter-individual variability [26, 27] and could have a great value in personalized medicine as they could become a new tool for calculation of risk factors in personalized diagnostics of chronic diseases. The identification of common genetic determinants for both markers could play an important role towards this direction. Therefore, in order to expand the knowledge of the common pathways that regulate both biomarkers, the current study aimed to investigate the genetic associations between VEGF-A and TL.

Accordingly, ten variants that were associated in a GWAS study with VEGF-A levels and together explain up to 52% of the VEGF-A phenotypic variance [28] were investigated for association with LTL and MTL, as well as estimated lifelong telomere attrition (LTL/MTL ratio) in a cohort of adult subjects.

RESULTS

The descriptive characteristics of the population are presented in Table 1.

Association between VEGF-A related genetic variants and TL

Values of MAF and results of HWE analysis are presented in Table 2.

The SNP rs10761741 did not follow the HWE and was excluded from further analyses.

None of the 9 assessed SNPs was significantly associated with LTL (Table 3) or MTL (Table 4).

A significant association was identified between the minor allele (T) of rs6993770 and LTL/MTL ($P=0.001143$, $\beta=0.0148$, $SE=0.004516$) (Table 5). There was no significant epistatic interaction identified for three tested phenotypes (LTL, MTL and LTL/MTL) according to the selected threshold ($p<0.00046$).

A bioinformatics analysis showed that rs6993770 is located on the intron region (Figure 1) of the *ZFPM2* (zinc finger protein, FOG family member 2) gene.

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 6) with VEGF-A levels used as the dependent variable. There was no significant relation of VEGF-A plasma concentrations with the investigated TL phenotypes.

DISCUSSION

The present study investigated associations between VEGF-A genetic determinants and telomere length dynamics. It identified a direct association between minor allele (T) of rs6993770 and LTL/MTL ratio

Table 1. Characteristics of the population.

Variable	n	mean	SD
Age (years)	402	60	15
Sex (F/M %)	402	32/68	
LTL (kb)	402	6.71	0.84
MTL (kb)	402	8.57	0.72
LTL/MTL	402	0.78	0.07
VEGF-A (pg/ml)	324	25.55	16.97
Statin use (%)	324	38.88	

SD, standard deviation.
LTL and VEGF are not normally distributed so median values and interquartile ranges are also given.
LTL: 6.62 (6.11 – 7.26) kb.
VEGF-A: 15.79 (10.07 - 29.38) pg/ml.

Table 2. 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.

Table 3. Association between VEGF-A related genetic variants and LTL (log transformed)*.

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

B, Effect size; SE, Standard error; P-value threshold is P<0.0012.
*models are adjusted for age and sex.

Table 4. 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

B, Effect size; SE, Standard error; P-value threshold is P<0.0012.
*models are adjusted for age and sex.

Table 5. 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
*models are adjusted for age and sex.

(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 [29]. Besides VEGF-A, it has been related to variation in HDL cholesterol [30], erythrocyte count, IL-12 levels, and platelets count [31].

In the blood, VEGF-A can be found in plasma, platelets and leukocytes [32]. 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 tumor growth through their function as transporters of VEGF-A [33, 34].

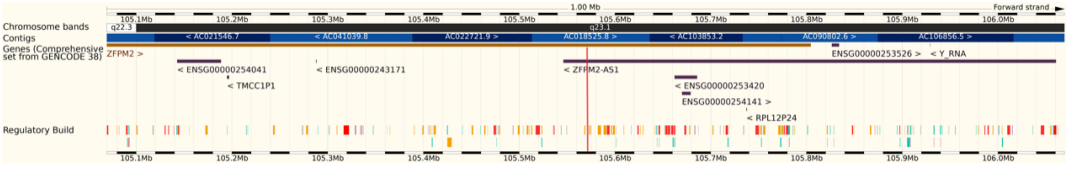

Figure 1. Rs6993770 (red stripe) is located on the intron 4 of the *ZFPM2* gene (8q23.1).

Table 6. Univariate and multiple regression analysis of VEGF-A plasma concentration with LTL, MTL and LTL/MTL.

Independent variable*	Regression coefficient (SE)	R^2	P-value
Non adjusted			
LTL	0.019 (0.160)	0.19%	0.42
MTL	-0.009 (0.028)	0.03%	0.74
LTL/MTL	0.471 (0.300)	0.73%	0.12
Adjusted*			
LTL	-0.075 (0.107)	0.15%	0.49
MTL	-0.170 (0.105)	0.81%	0.36
LTL/MTL	0.756 (0.919)	0.21%	0.41

SE, Standard error.
*each TL variable was tested as independent variable in separate models. Thus, the table presents the results of 3 different models. * Models were adjusted for age, sex and statin use.
Dependent variable: logVEGF-A.

In the present study, the 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 (Friend of GATA) family member protein. The FOG proteins can both activate and down-regulate expression of GATA-target genes, resulting in modulation of GATA family proteins activity. The *ZFPM2* gene codes for the FOG family member 2 that has been linked with repression of GATA mediated transcriptional activation [35, 36]. GATA proteins are crucial regulators of haematopoiesis and cardiogenesis *via* the control of haemoglobin synthesis [37].

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 [38]. 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 [39]. Such a hypothesis seems plausible since telomere attrition was the only TL phenotype that was significantly related to the genetic variant, whereas LTL and MTL did not show significant association with rs6993770. Estimated lifelong attrition as expressed by the LTL/MTL ratio has been suggested to be impacted mainly by TL attrition during early life [40]. The finding of our study could be in agreement with this statement. 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.

Besides the genetic association between TL and VEGF-A, this study did not identify any direct association between VEGF-A plasma levels and LTL, MTL or estimated 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 [41]. On the other hand, a study of patients with knee osteoarthritis identified a negative correlation between VEGF-A plasma levels and LTL [42]. Further studies are warranted to fully explore the association of these biomarkers. Thus, the genetic association of rs6993770 with telomere attrition seems to not be directly linked with VEGF-A levels and is probably mediated by other mechanisms, such as the increased haematopoiesis that we propose here.

We would like to acknowledge the limitations of the study. The power is limited due to small sample size, so false negative results cannot be excluded. Therefore, we propose the replication of the results in bigger populations.

To conclude, this study is the first to investigate 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. We propose that this association could be explained possibly through the modification of the expression of GATA proteins, which could result in a direct impact on hematopoiesis and production of platelets. This hypothesis remains to be replicated and verified in future studies.

MATERIALS AND METHODS

Population

This analysis included 402 individuals with measurements of LTL and MTL from the TELARTA (TELomere in ARTerial Aging) cohort. The aim of the TELARTA study was to examine the role of telomere length dynamics in arterial aging using the "Blood-and-Muscle model" [43, 44]. Briefly, 259 French individuals, who were admitted for various surgical procedures, were recruited in two centers (Nancy and Marseille) and constituted the discovery cohort of the TELARTA study. The replication cohort of the study included 91 French individuals, recruited under the same conditions as in the discovery cohort and 52 individuals from an independent Greek population enrolled in Athens. All French participants provided written informed consent approved by the Ethics Committee (Comité de Protection des Personnes) of Nancy, France. All Greek participants 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 study was conducted in accordance with the Declaration of Helsinki and is registered on http://www.clinicaltrials.gov under unique identifier: NCT02176941.

Telomere length measurement

TL in skeletal muscle (MTL) and in leukocytes (LTL) were measured in DNA extracted from muscle biopsies and peripheral blood leukocytes respectively [14, 43]. Skeletal muscle biopsies (~100 to 200 mg in the surgical field) were collected from individual during surgery, flash frozen in liquid nitrogen and stored on -80° C until DNA extraction. Whole blood samples were collected in EDTA tubes prior to surgery and stored on -80° C until DNA extraction.

DNA was extracted from the muscle tissue and leukocyte by the phenol/chloroform/isoamyl alcohol method. DNA samples passed an integrity testing using a 1% (wt/vol) agarose gel before TL measurement was performed by the Southern blot analysis of terminal restriction fragments, as described previously [45]. Briefly, DNA samples were 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 depurination, denaturation and neutralization, DNA was transferred on a positively charged nylon membrane (Roche) using a vacuum blotter (Biorad, Hercules, CA, USA). Membranes were hybridized at 42° C with a digoxigenin-labelled telomeric probe. The probe was later detected by the DIG luminescent detection procedure (Roche) and exposed on a 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. The LTL/MTL ratio was calculated for each individual by dividing the LTL by the MTL value. As MTL can be considered as a proxy of TL at birth and LTL represents the current status of TL, a ratio of 1 indicates no telomere attrition throughout life, while a smaller ratio indicates lower LTL values compared to MTL and thus greater telomere attrition.

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, UK). 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. VEGF-A plasma levels were measured in 324 French individuals from the TELARTA study.

Genotyping

Ten VEGF-A related genetic variants (rs10761741, rs10738760, rs6921438, rs7043199, rs6993770, rs4416670, rs114694170, rs34528081, rs4782371 and rs2639990), previously identified by a GWAS, were genotyped in leukocyte DNA samples using a PCR-based KASP assay [46]. Genotyping was performed by the Laboratory of the Government Chemist (LGC Ltd., Teddington, UK) using the competitive allele-specific PCR (KASP) chemistry coupled with a Förster resonance energy transfer-based genotyping system (http://www.kbioscience.co.uk/reagents/KASP/KASP.html) and by Randox genotyping VEGF-A assay.

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 (χ^2 test) and was excluded from further

analyses due to suspected technical error in the genotyping.

LTL was log-transformed to follow a normal distribution. The direct effects of VEGF-A related genetic variants on three phenotypes of interest (LTL, MTL and LTL/MTL) were tested using linear regression models adjusted for age and sex using the PLINK toolset under the assumption of an additive genetic model. Reference allele for all variants was the minor allele. Epistatic interactions were tested using R package CAPE.

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 analyses were performed to study the association of VEGF-A plasma concentration with LTL, MTL and LTL/MTL. As VEGF-A levels were not normally distributed, a log-transformation was performed. In the models, VEGF-A levels were used as the dependent variable and they were adjusted for age, sex and statin use, because statins were significantly associated with VEGF-A levels in our population (data not presented).

***In silico* analysis**

The genomic environment of the significant SNPs was explored using Ensembl browser of the human genome (GRCh38.p12) and NCBI dbSNP.

Abbreviations

CVD: Cardiovascular disease; CVD: Cardiovascular diseases; FDA: Food and Drug Administration; FOG: Friend of GATA; GWAS: Genome-wide association study; HDL: High-density lipoproteins; HWE: Hardy-Weinberg equilibrium; IL: Interleukin; LTL: Leukocyte telomere length; MAF: Minor allele frequency; MTL: Muscle telomere length; SE: Standard error; SNP: Single nucleotide polymorphism; TL: Telomere length; VEGF-A: Vascular endothelial growth factor A; *ZFP*: Zinc finger protein.

AUTHOR CONTRIBUTIONS

VG, AP, MGS and ST designed the present idea. VG wrote the manuscript with support of AP. AP and MGS conceived and performed the statistical analysis with support of CL. VG, AP, MGS and ST performed the analysis of the results. SK and CM contributed to sample preparation, JL, HM and PF carried out experiments. AB and SVS conceived the present idea and supervised the realization of this project. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGMENTS

We would like to thank the Operational Programme FEDER-FSE Lorraine et Massif des Vosges 2014-2020, the "Région Lorraine", the Agence Nationale de la Recherche "programme d'Investissements d'avenir" (PIA) and the French PIA project "Lorraine Université d'Excellence"- LUE to support our project and making therefore possible to realize our research work.

We also thank all the members of the TELARTA consortium:

Sylvie Gautier, Ghassan Watfa, Véronique Regnault, Patrick Lacolley, Huguette Louis, Oualid Ayad, Cécile Lakomy, Mélanie Folio, Margaux Schmitt, Agnes Didier, Sandrine Vauthier, Nelly François, Masayuki Kimura, Tsung-Po Lai, Abraham Aviv, Pascal M. Rossi, Françoise Dignat George, Florence Sabatier, Stéphanie Simoncini, Chloé Dumoulin, Patricia Béranger, Serguei Malikov, Nicla Settembre, Jacques Hubert, Luc Frimat, Baptiste Bertrand, Mourad Boufi, Xavier Flecher, Nicolas Sadoul, Pascal Eschwege, Michèle Kessler, Irene P. Tzanetakou, Ilias P. Doulamis, Panagiotis Konstantopoulos, Aspasia Tzani, Marilina Korou, Anastasios Gkogkos, Konstantinos Perreas, Evangelos Menenakos, Georgios Samanidis, Michail Vasiloglou-Gkanis, Jeremy D. Kark and Simon Verhulst.

CONFLICTS OF INTEREST

Authors declare no conflicts of interest.

FUNDING

The TELARTA cohort was supported by the French National Research Agency (ANR), Translationnelle: N°ID RCB: 2014-A00298-39: 2014-2017. 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 the French regional project CPER-ITM2P 2015-2020, by Agence Nationale de la Recherche, programme d'Investissements d'avenir (PIA), grant number ANR-15RHU-0004 and GEENAGE project (Functional Genomic, Epigenomic and Evironment interplay to IMPACT management of healthy and pathological AGEing)" within the frames of the French PIA project "Lorraine Université d'Excellence" - LUE, reference

ANR-15-IDEX-04-LUE. VG was partially funded by the “Région Lorraine”.

REFERENCES

1. Blackburn EH. Structure and function of telomeres. Nature. 1991; 350:569–73. https://doi.org/10.1038/350569a0 PMID:1708110

2. Zhang J, Rane G, Dai X, Shanmugam MK, Arfuso F, Samy RP, Lai MK, Kappei D, Kumar AP, Sethi G. Ageing and the telomere connection: An intimate relationship with inflammation. Ageing Res Rev. 2016; 25:55–69. https://doi.org/10.1016/j.arr.2015.11.006 PMID:26616852

3. Samani NJ, Boultby R, Butler R, Thompson JR, Goodall AH. Telomere shortening in atherosclerosis. Lancet. 2001; 358:472–73. https://doi.org/10.1016/S0140-6736(01)05633-1 PMID:11513915

4. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. N Engl J Med. 2005; 352:1685–95. https://doi.org/10.1056/NEJMra043430 PMID:15843671

5. McNally EJ, Luncsford PJ, Armanios M. Long telomeres and cancer risk: the price of cellular immortality. J Clin Invest. 2019; 129:3474–81. https://doi.org/10.1172/JCI120851 PMID:31380804

6. Fragkiadaki P, Nikitovic D, Kalliantasi K, Sarandi E, Thanasoula M, Stivaktakis PD, Nepka C, Spandidos DA, Tosounidis T, Tsatsakis A. Telomere length and telomerase activity in osteoporosis and osteoarthritis. Exp Ther Med. 2020; 19:1626–32. https://doi.org/10.3892/etm.2019.8370 PMID:32104213

7. Ameh OI, Okpechi IG, Dandara C, Kengne AP. Association Between Telomere Length, Chronic Kidney Disease, and Renal Traits: A Systematic Review. OMICS. 2017; 21:143–55. https://doi.org/10.1089/omi.2016.0180 PMID:28253088

8. Wang J, Liu Y, Xia Q, Xia Q, Wang B, Yang C, Liang J, Liu X. Potential roles of telomeres and telomerase in neurodegenerative diseases. Int J Biol Macromol. 2020; 163:1060–78. https://doi.org/10.1016/j.ijbiomac.2020.07.046 PMID:32673712

9. Youngren K, Jeanclos E, Aviv H, Kimura M, Stock J, Hanna M, Skurnick J, Bardeguez A, Aviv A. Synchrony in telomere length of the human fetus. Hum Genet. 1998; 102:640–43. https://doi.org/10.1007/s004390050755 PMID:9703424

10. Houben JM, Moonen HJ, van Schooten FJ, Hageman GJ. Telomere length assessment: biomarker of chronic oxidative stress? Free Radic Biol Med. 2008; 44:235–46. https://doi.org/10.1016/j.freeradbiomed.2007.10.001 PMID:18021748

11. 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; 40:1764–71. https://doi.org/10.1249/MSS.0b013e31817c92aa PMID:18799986

12. Aviv A, Shay JW. Reflections on telomere dynamics and ageing-related diseases in humans. Philos Trans R Soc Lond B Biol Sci. 2018; 373:20160436. https://doi.org/10.1098/rstb.2016.0436 PMID:29335375

13. Entringer S, de Punder K, Buss C, Wadhwa PD. The fetal programming of telomere biology hypothesis: an update. Philos Trans R Soc Lond B Biol Sci. 2018; 373:20170151. https://doi.org/10.1098/rstb.2017.0151 PMID:29335381

14. Chahine MN, Toupance S, El-Hakim S, Labat C, Gautier S, Moussallem T, Yared P, Asmar R, Benetos A. Telomere length and age-dependent telomere attrition: the blood-and-muscle model[1]. Can J Physiol Pharmacol. 2019; 97:328–34. https://doi.org/10.1139/cjpp-2018-0582 PMID:30874445

15. Ferrara N, Gerber HP, LeCouter J. The biology of VEGF and its receptors. Nat Med. 2003; 9:669–76. https://doi.org/10.1038/nm0603-669 PMID:12778165

16. Carmeliet P. VEGF as a key mediator of angiogenesis in cancer. Oncology. 2005 (Suppl 3); 69:4–10. https://doi.org/10.1159/000088478 PMID:16301830

17. Testa U, Pannitteri G, Condorelli GL. Vascular endothelial growth factors in cardiovascular medicine. J Cardiovasc Med (Hagerstown). 2008; 9:1190–221. https://doi.org/10.2459/JCM.0b013e3283117d37 PMID:19001927

18. Camaré C, Pucelle M, Nègre-Salvayre A, Salvayre R. Angiogenesis in the atherosclerotic plaque. Redox Biol. 2017; 12:18–34. https://doi.org/10.1016/j.redox.2017.01.007 PMID:28212521

19. MacEneaney OJ, Kushner EJ, Westby CM, Cech JN,

Greiner JJ, Stauffer BL, DeSouza CA. Endothelial progenitor cell function, apoptosis, and telomere length in overweight/obese humans. Obesity (Silver Spring). 2010; 18:1677–82. https://doi.org/10.1038/oby.2009.494 PMID:20057362

20. Farhat N, Thorin-Trescases N, Voghel G, Villeneuve L, Mamarbachi M, Perrault LP, Carrier M, Thorin E. Stress-induced senescence predominates in endothelial cells isolated from atherosclerotic chronic smokers. Can J Physiol Pharmacol. 2008; 86:761–69. https://doi.org/10.1139/Y08-082 PMID:19011671

21. Iberl S, Meyer AL, Müller G, Peters S, Johannesen S, Kobor I, Beier F, Brümmendorf TH, Hart C, Schelker R, Herr W, Bogdahn U, Grassinger J. Effects of continuous high-dose G-CSF administration on hematopoietic stem cell mobilization and telomere length in patients with amyotrophic lateral sclerosis - a pilot study. Cytokine. 2019; 120:192–201. https://doi.org/10.1016/j.cyto.2019.05.003 PMID:31100684

22. Chen L, Xia W, Hou M. Mesenchymal stem cells attenuate doxorubicin-induced cellular senescence through the VEGF/Notch/TGF-β signaling pathway in H9c2 cardiomyocytes. Int J Mol Med. 2018; 42:674–84. https://doi.org/10.3892/ijmm.2018.3635 PMID:29693137

23. Zittermann SI, Issekutz AC. Endothelial growth factors VEGF and bFGF differentially enhance monocyte and neutrophil recruitment to inflammation. J Leukoc Biol. 2006; 80:247–57. https://doi.org/10.1189/jlb.1205718 PMID:16818728

24. 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; 71:54–63. https://doi.org/10.1111/j.1469-1809.2006.00298.x PMID:17227476

25. Honig LS, Kang MS, Cheng R, Eckfeldt JH, Thyagarajan B, Leiendecker-Foster C, Province MA, Sanders JL, Perls T, Christensen K, Lee JH, Mayeux R, Schupf N. Heritability of telomere length in a study of long-lived families. Neurobiol Aging. 2015; 36:2785–90. https://doi.org/10.1016/j.neurobiolaging.2015.06.017 PMID:26239175

26. Hetland ML, Christensen IJ, Lottenburger T, Johansen JS, Svendsen MN, Hørslev-Petersen K, Nielsen L, Nielsen HJ. Circulating VEGF as a biological marker in patients with rheumatoid arthritis? Preanalytical and biological variability in healthy persons and in patients. Dis Markers. 2008; 24:1–10. https://doi.org/10.1155/2008/707864 PMID:18057530

27. Toupance S, Villemonais D, Germain D, Gegout-Petit A, Albuisson E, Benetos A. The individual's signature of telomere length distribution. Sci Rep. 2019; 9:685. https://doi.org/10.1038/s41598-018-36756-8 PMID:30679552

28. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, Concas MP, Traglia M, Barbieri C, Ndiaye NC, Stathopoulou MG, Lagou V, Maestrale GB, 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. https://doi.org/10.1371/journal.pgen.1005874 PMID:26910538

29. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, Safa R, Azimi Nezhad M, Sawyer D, Marteau JB, Xanthakis V, Siest G, Sullivan L, 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. https://doi.org/10.1161/CIRCRESAHA.111.243790 PMID:21757650

30. Stathopoulou MG, Bonnefond A, Ndiaye NC, Azimi-Nezhad M, El Shamieh S, Saleh A, Rancier M, Siest G, Lamont J, Fitzgerald P, Visvikis-Siest S. 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; 54:535–41. https://doi.org/10.1194/jlr.P030551 PMID:23204297

31. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, et al. Ensembl 2018. Nucleic Acids Res. 2018; 46:D754–61. https://doi.org/10.1093/nar/gkx1098 PMID:29155950

32. 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; 97:978–85. https://doi.org/10.1038/sj.bjc.6603923 PMID:17912242

33. 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; 6:3147–52. PMID:10955796

34. 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; 2:311–15.
https://doi.org/10.3816/cbc.2002.n.008
PMID:11899364

35. Svensson EC, Tufts RL, Polk CE, Leiden JM. Molecular cloning of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. Proc Natl Acad Sci USA. 1999; 96:956–61.
https://doi.org/10.1073/pnas.96.3.956
PMID:9927675

36. Lin AC, Roche AE, Wilk J, Svensson EC. The N termini of Friend of GATA (FOG) proteins define a novel transcriptional repression motif and a superfamily of transcriptional repressors. J Biol Chem. 2004; 279:55017–23.
https://doi.org/10.1074/jbc.M411240200
PMID:15507435

37. GeneCards. The human gene database. 2019. https://www.genecards.org/

38. Rufer N, Brümmendorf TH, Kolvraa S, Bischoff C, Christensen K, Wadsworth L, Schulzer M, Lansdorp PM. 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; 190:157–67.
https://doi.org/10.1084/jem.190.2.157
PMID:10432279

39. Schick UM, Jain D, Hodonsky CJ, Morrison JV, Davis JP, Brown L, Sofer T, Conomos MP, Schurmann C, McHugh CP, Nelson SC, Vadlamudi S, Stilp A, et al. Genome-wide Association Study of Platelet Count Identifies Ancestry-Specific Loci in Hispanic/Latino Americans. Am J Hum Genet. 2016; 98:229–42.
https://doi.org/10.1016/j.ajhg.2015.12.003
PMID:26805783

40. Sabharwal S, Verhulst S, Guirguis G, Kark JD, Labat C, Roche NE, Martimucci K, Patel K, Heller DS, Kimura M, Chuang D, Chuang A, Benetos A, Aviv A. Telomere length dynamics in early life: the blood-and-muscle model. FASEB J. 2018; 32:529–34.
https://doi.org/10.1096/fj.201700630R
PMID:28855279

41. 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:e87348.
https://doi.org/10.1371/journal.pone.0087348
PMID:24475279

42. 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.
https://doi.org/10.1111/1756-185X.12988
PMID:28205383

43. Benetos A, Toupance S, Gautier S, Labat C, Kimura M, Rossi PM, Settembre N, Hubert J, Frimat L, Bertrand B, Boufi M, Flecher X, Sadoul N, et al. Short Leukocyte Telomere Length Precedes Clinical Expression of Atherosclerosis: The Blood-and-Muscle Model. Circ Res. 2018; 122:616–23.
https://doi.org/10.1161/CIRCRESAHA.117.311751
PMID:29242238

44. Toupance S, Stathopoulou MG, Petrelis AM, Gorenjak V, Labat C, Lai TP, Visvikis-Siest S, Benetos A. TERC Variants Associated with Short Leukocyte Telomeres: Implication of Higher Early Life Leukocyte Telomere Attrition as Assessed by the Blood-and-Muscle Model. Cells. 2020; 9:1360.
https://doi.org/10.3390/cells9061360 PMID:32486379

45. Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, Harley CB, Aviv A. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. Nat Protoc. 2010; 5:1596–607.
https://doi.org/10.1038/nprot.2010.124
PMID:21085125

46. He C, Holme J, Anthony J. SNP genotyping: the KASP assay. Methods Mol Biol. 2014; 1145:75–86.
https://doi.org/10.1007/978-1-4939-0446-4_7
PMID:24816661

Publication 5

Alzheimer's Disease Risk is associated with VEGF-A-related polymorphisms.

Petrelis AM, Stathopoulou MG, Kafyra M, Masson C, Murray H, Lamont J, Fitzgerald P, Dedoussis G, Yen FT, Visvikis-Siest S.

Under revision to Aging

My specific contribution in this study was: Participation in the design of the study, conception and realisation of analyses, participation in data interpretation and drafting of the first version of the manuscript.

Objective

Late-onset AD is a multifactorial disease with a strong genetic background. However, the use of common genetic polymorphisms identified in GWAS in disease prediction modelling is very limited. This is a common finding in most chronic diseases and it is due to the fact that the majority of identified polymorphisms explain small relative risk and a small proportion of the underlying genetic contribution. The predictive ability of models can be improved with the inclusion of functional variants, the incorporation of epistatic effects and the combination with nongenetic biomarkers [50]. In this study we aimed to develop a model for risk prediction for AD using a machine-learning methodology in a case-control population of late-onset AD patients that would combine novel candidate genes for AD and their epistatic interactions with known functional polymorphisms of *APOE* and *LSR* genes.

Population and analysis

A case-control population of late-onset AD was involved in this study that consisted of 1078 (602 controls and 476 cases) unrelated individuals of European origin. The recruitment and data collection of this population was previously described [25]. Genotyping data were available in a subsample of the population. The final sample size after quality control for genotyping and outliers was 323 (180 controls and 143 cases).

Two polymorphisms of *LSR* gene (rs34259399 and rs916147) previously identified as candidate genes for AD were genotyped, along with the 10 VEGF-A-related polymorphisms (rs6921438,

rs7043199, rs6993770, rs2375981, rs34528081, rs4782371, rs2639990, rs10761741, rs114694170, rs1740073) and the common *APOE* variants rs429358 (Cys112Arg) and rs7412 (Arg158Cys).

We have applied a machine learning method called EN [51] aiming to identify the strongest predictors for the risk of AD combining all genotyping data as well as age and sex. This method has been previously used in AD risk prediction with clinical factors, imaging and "-omics" [52-54] but not to study the effect of candidate genes as predicting factors.

The QIAGEN Ingenuity Pathway Analysis (IPA) tool was used to identify causal links and relationships between the significant genes identified to predict the AD risk.

Results

After applying EN in the data, we developed a predictive model for risk of AD with an accuracy of 72% (confidence interval CI= (0.6, 0.8) for 95%). The above accuracy has been calculated under a model with P-value: 0.008922. The AUC of the model is 81%, which indicates the detection of true positives (true cases predicted as cases) versus false positives (true controls predictive as cases).

This model included 11 factors demonstrated in Figure 17.

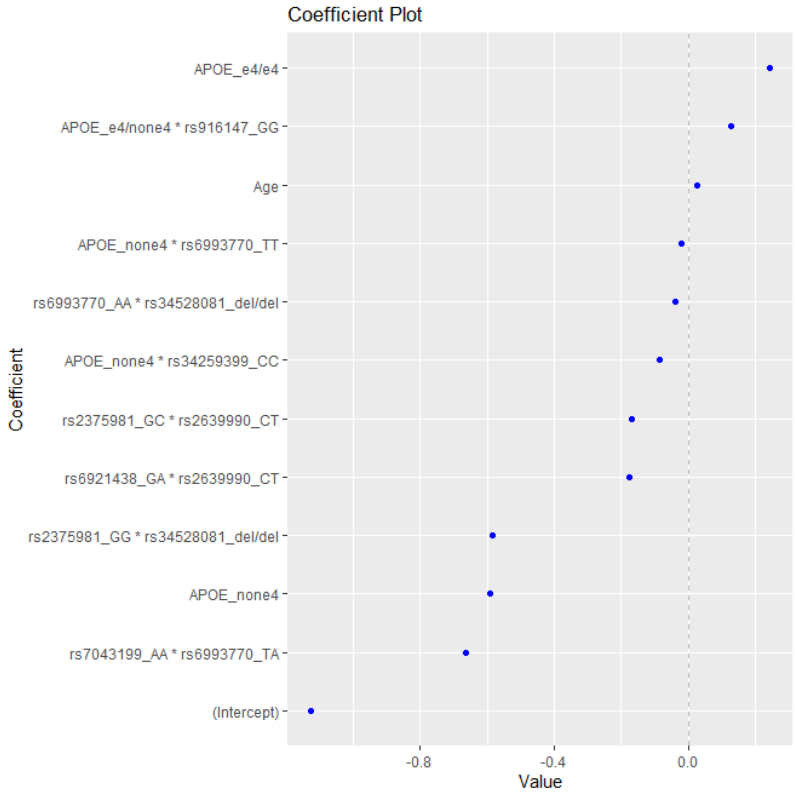

Figure 17. The coefficient plot of the EN model with the highest accuracy

The presence of 2 *ε4* alleles (homozygotes *ε4*/*ε4*) and age were linked with increased risk for AD as expected, while the absence of *ε4* alleles was linked with decreased risk. Two interactions previously [55] identified between *APOE* and *LSR* polymorphisms are also included in the model. The model included five epistatic interactions between VEGF-A-related variants, one *APOE***LSR* (rs34259399) interaction and one *APOE**VEGF-A-related polymorphism interaction (rs6993770) associated with decreased risk for AD. The interaction rs7043199*rs6993770 presented the highest coefficient, thus is the strongest predictive factor of the model, with a protective role against AD higher than that of the absence of *ε4 APOE* allele, followed by the absence of the *ε4 APOE* allele and the interaction rs2375981*rs34528081.

The identified genes in the model, *VLDLR-AS1*, *APOE*, *KCNV2*, *ZADH2*, *C6orf223*, *LSR*, *ZFPM2*, and *VEGF-A*, were analysed in the IPA tool and they were found to be commonly involved in different diseases with CVD, ophthalmic disease, organismal injury and abnormalities, neurological disease, and cancer ranging in the top 5 position.

Conclusion

This study identified novel epistatic interactions between *APOE*, *LSR* and VEGF-A related polymorphisms that are predicting the risk of AD. The accuracy of the prediction model is 72% and is comparable to other tools [56-58]. This is very high accuracy for a model that uses mostly genetic factors as other studies have shown that the addition of genetic variables leads usually to a small improvement of prediction of the classical variables' models [59]. A recent review of 61 papers describing dementia risk models, most of them had moderate-to-high predictive ability (AUC>0.70). The highest AUC value was 0.932 [60], with our AUC being 81%. Thus, this model can be very useful and could be tested in bigger populations. The results also provide new insights about molecular mechanisms involved in AD development and could have an impact as biomarkers and/or treatment targets for this disease. We are proposing two epistatic interactions (rs7043199*rs6993770 and rs2375981*rs34528081) between VEGF-A-related polymorphisms as strong protective factors against AD. Our results are further supporting the hypothesis that the genetic determinants of the selected biomarkers (VEGF-A, LSR and APOE) can act as common links between the most important chronic diseases such as CVD, cancer and neurological diseases like AD and can play an important general role in the prediction of these chronic diseases.

www.aging-us.com AGING 2022, Vol. 14, No. 6

Research Paper

VEGF-A-related genetic variants protect against Alzheimer's disease

Alexandros M. Petrelis[1], Maria G. Stathopoulou[2], Maria Kafyra[1,3], Helena Murray[4], Christine Masson[1], John Lamont[4], Peter Fitzgerald[4], George Dedoussis[1,3], Frances T. Yen[5], Sophie Visvikis-Siest[1]

[1]IGE-PCV, Université de Lorraine, Nancy 54000, France
[2]Inserm, C3M, Team Control of Gene Expression (10), Université Cote d'Azur, Nice, France
[3]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens 17671, Greece
[4]Randox Laboratories Limited, Crumlin, County Antrim BT29 4QY, United Kingdom
[5]Qualivie, UR AFPA laboratory, University of Lorraine, Vandoeuvre-les-Nancy, Lorraine, France

Correspondence to: Sophie Visvikis-Siest; **email:** sophie.visvikis-siest@inserm.fr
Keywords: Apolipoprotein E, LSR, VEGF-A, epistasis, Alzheimer's disease, elastic net, machine learning, prediction
Received: December 31, 2021 **Accepted:** March 14, 2022 **Published:** March 28, 2022

Copyright: © 2022 Petrelis et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The Apolipoprotein E (*APOE*) genotype has been shown to be the strongest genetic risk factor for Alzheimer's disease (AD). Moreover, both the lipolysis-stimulated lipoprotein receptor (LSR) and the vascular endothelial growth factor A (VEGF-A) are involved in the development of AD. The aim of the study was to develop a prediction model for AD including single nucleotide polymorphisms (SNP) of *APOE*, *LSR* and VEGF-A-related variants.

The population consisted of 323 individuals (143 AD cases and 180 controls). Genotyping was performed for: the *APOE* common polymorphism (rs429358 and rs7412), two *LSR* variants (rs34259399 and rs916147) and 10 VEGF-A-related SNPs (rs6921438, rs7043199, rs6993770, rs2375981, rs34528081, rs4782371, rs2639990, rs10761741, rs114694170, rs1740073), previously identified as genetic determinants of VEGF-A levels in GWAS studies. The prediction model included direct and epistatic interaction effects, age and sex and was developed using the elastic net machine learning methodology.

An optimal model including the direct effect of the *APOE ε4* allele, age and eight epistatic interactions between *APOE* and *LSR*, *APOE* and VEGF-A-related variants was developed with an accuracy of 72%. Two epistatic interactions (rs7043199*rs6993770 and rs2375981*rs34528081) were the strongest protective factors against AD together with the absence of *ε4 APOE* allele. Based on pathway analysis, the involved variants and related genes are implicated in neurological diseases.

In conclusion, this study demonstrated links between *APOE*, *LSR* and VEGF-A-related variants and the development of AD and proposed a model of nine genetic variants which appears to strongly influence the risk for AD.

INTRODUCTION

Alzheimer's disease (AD) is a chronic, neuro-degenerative disorder and the most common cause of dementia, characterized by mental and functional impairment [1]. It is associated with accumulation of neuronal amyloid plaques and early lesions primarily in hippocampus. It is estimated that AD prevalence doubles every 5 years in individuals over the age of 65. The World Health Organization has stated that AD constitutes a growing universal public health issue with enormous consequences on both individuals and communities [2].

AD severely affects the life of the patient, causing dependency, disability and subsequent fatality [2, 3].

There are two basic types of AD: a) familial or early onset AD which is responsible for more than 5% of the disease incidence and; b) sporadic or late onset AD that accounts for 79% of the disease burden. Late onset AD is highly heritable and it is etiologically heterogeneous originating from a mixture of multiple genetic and environmental risk factors.

Some of the genes that have been associated with the risk of sporadic AD include the *ABCA7*, *APOE*, *BIN1*, *CD2AP*, *CD33*, *CLU*, *CR1*, *EPHA1*, *MS4A4A/ MS4A4E/MS4A6E*, *PICALM*, and *SORL1* genes [4]. More than 40 genes/loci have been associated with the risk of AD through the last 10 years, based on genome-wide association studies (GWAS) [5].

The most important genetic factor for AD is the Apolipoprotein E gene (*APOE*). The *APOE* gene codes for a 35 kDa glycoprotein, the apolipoprotein E (ApoE), which is strongly expressed in the brain [6]. There are three most common allelic variants in the *APOE* gene that alter the protein sequence leading to the formation of three different *APOE* isoforms: *APOE2* (cys112, cys158), *APOE3* (cys112, arg158), and *APOE4* (arg112, arg158) [6, 7] arising from 3 alleles, respectively, *ε2*, *ε3* and *ε4*. These alleles are associated with different ApoE roles [7]. The *ε4* allele is the strongest risk factor for late-onset AD [8, 9], due to its association with increased amyloid deposition and is a known risk factor for cardiovascular disease (CVD) [10]. Individuals with one *ε4* allele have a 2 to 3-fold elevated risk of developing AD, while those with two *ε4* alleles have about 12-fold increased risk compared to individuals who do not have the *ε4* allele. On the other hand, the *ε2* allele of the *APOE* gene appears to display a protective role, as it is associated with reduced risk for AD [7], but remains a risk factor for Type III hyperlipidemia [11]. Thus, this common polymorphism is an excellent candidate to study in the development of genetic risk prediction models for AD.

The human *APOE* gene is situated on the long arm of chromosome 19q13.1, an AD-associated zone as reported by GWAS [5]. The lipolysis-stimulated lipoprotein receptor (*LSR*) gene is also located in the same region and encodes the lipolysis-stimulated lipoprotein receptor (LSR) which recognizes ApoE as ligand [12]. As an ApoE receptor, LSR is involved in the process of managing and maintaining lipid balance in the peripheral and central nervous system [13, 14] Recently, our team identified significant epistatic interactions between two *LSR* gene single nucleotide polymorphisms (SNPs) and *APOE* in AD patients [12], namely the rs34259399 and the rs916147 SNPs. The former is located on exon 6 and the latter is located in a splicing junction between intron 5 and exon 6. Both these SNPs have been studied by our group and preliminary results indicate a functionality in terms of gene expression modification, while the rs916147 was also associated with lipids in a population of obese individuals (Yen et al., unpublished results).

The vascular endothelial growth factor A (VEGF-A) is also considered as a risk factor for chronic diseases, including AD. The VEGF family plays important roles in angiogenic regulation, neurogenesis and neuronal survival [15]. Although an inverse relationship has also been demonstrated [16], decreased levels of VEGF-A in serum and cerebrospinal fluid have been linked with increased risk for AD and cognitive impairment [17, 18].

Our group has focused on the study of VEGF-A for many years and has been involved in two GWAS by which ten genetic variants have been identified explaining more than 50% of the individual variability of VEGF-A levels [19, 20]. This exceptionally high percentage of variability explained by these variants makes them optimal target SNPs to be used in the candidate genes association studies as determinants of VEGF-A levels. In several previous studies we have demonstrated associations of these polymorphisms with intermediate phenotypes of CVD and other chronic diseases, such as autoimmune thyroid disease and depression [21–25], where VEGF-A is involved in several of their pathophysiology pathways. We have initiated and we are coordinating the Vascular Endothelial growth factor European Genomic Federation Consortium - VEGF Consortium (Sophie Visvikis-Siest coordinator, http://www.vegfconsortium.org) for the study of VEGF-A in chronic diseases and personalized medicine [26].

The precise detection of individuals at high risk for AD is very critical for early diagnosis and appropriate management enabling closer monitoring, enhanced care, as well as closer supervision of targeted risk factors-based interventions [27]. Even though late-onset AD is known to be a multifactorial disease with a strong genetic component, the use of common genetic variations identified in GWAS in disease prediction modelling has been of limited value so far, given that such polymorphisms explain a small relative risk and proportion of the underlying genetic contribution. Thus, it has been proposed that the predictive ability of models would be improved with the inclusion of true functional variants, the incorporation of epistatic effects and the combination with nongenetic biomarkers [28].

In the present study, a machine-learning predictive model for AD risk was developed, using a case-control population of late-onset AD patients and combining novel variants on candidate molecules for AD that have

Table 1. Populations' characteristics.

	Age		Sex (male)	
	Mean	**SD**	***n***	**%**
Controls (*n* = 182)	74.59	8.39	68	37.36
Patients (*n* = 145)	69.94	8.66	69	47.58

Abbreviation: SD: standard deviation.

Table 2. Characteristics of the genotyped polymorphisms.

Variants	**Genes**	**All**		**Controls**		**Patients**	
		MAF	**HW *p*-value**	**MAF**	**HW *p*-value**	**MAF**	**HW *p*-value**
rs10761741	*JMJD1C*	0.45	0.77	0.47	0.77	0.44	0.77
rs6921438	*LOC100132354 and C6orf223*	0.42	0.43	0.41	0.43	0.42	0.43
rs7043199	*VLDLR-AS1*	0.24	0.90	0.26	0.90	0.23	0.90
rs6993770	*ZFPM2*	0.33	0.71	0.33	0.71	0.34	0.71
rs114694170	*MEF2C*	0.45	**0.00**	0.47	**0.00**	0.44	**0.00**
rs1740073	*POLR1C*	0.38	0.33	0.38	0.33	0.38	0.33
rs2375981	*KCNV2 and VLDLR*	0.47	0.26	0.44	0.26	0.49	0.26
rs34528081	*VEGF-A*	0.39	0.93	0.41	0.93	0.36	0.93
rs916147	*LSR*	0.37	0.29	0.38	0.29	0.36	0.29
rs34259399	*LSR*	0.14	0.76	0.12	0.76	0.15	0.76
rs4782371	*ZFPM1*	0.30	0.96	0.34	0.96	0.26	0.96
rs2639990	*ZADH2*	0.09	0.94	0.12	0.94	0.07	0.94
rs429358	*APOE*	0.22	0.77	0.22	0.77	0.22	0.77
rs7412	*APOE*	0.06	0.43	0.06	0.43	0.06	0.43

Abbreviations: MAF: Minor allele frequency; HW: Hardy-Weinberg.

not been assessed before (VEGF-A related variants), as well as their epistatic interactions with known functional polymorphisms (*APOE* and *LSR*).

RESULTS

The characteristics of the final population used in the analysis are presented in Table 1. The mean age of participants was 74.59 and 69.94 years old in patients and non-patients, respectively. Men constituted 37.36% of the control population, whereas they contributed to almost half in the cases group (i.e., 47.58%).

The SNPs included in the analysis, their minor allele frequencies and their annotation on the genome are presented in Table 2. Since the rs114694170 SNP did not abide by the Hardy-Weinberg equilibrium law, it was removed from the analysis.

After applying the EN method to the data, the model with the highest accuracy included the variables presented in Figure 1. As expected, the presence of 2 *ε4* alleles (homozygotes *e4*/*e4*) was the first variable to be linked with increased risk for AD, while the absence of *ε4* allele was shown to be associated with decreased risk.

The epistatic interaction of *ε4* allele heterozygotes with the rs916147 variant of *LSR* gene was also associated with higher risk for AD, followed by age as the third risk factor for AD (Table 3).

The model included five epistatic interactions between VEGF-A-related variants, one *APOE***LSR* (rs34259399) interaction and one *APOE**VEGF-A-related polymorphism interaction (for the rs6993770 SNP), associated with decreased risk for AD (Table 3). The strongest predictive factor of the model appears to be the interaction rs7043199*rs6993770, which decreases the risk of AD, thus playing a protective role, followed by the absence of the *e4 APOE* allele and the interaction rs2375981*rs34528081.

Table 3. Coefficients of the prediction model.

		Coefficient
Risk factor variants	*APOE_ε4/ε4*	0.24
	APOE_ε4/non *ε4**rs916147	0.13
	Age	0.03
Protective factor variants	rs7043199*rs6993770	−0.66
	*APOE*_non *ε4*	−0.59
	rs2375981*rs34528081	−0.58
	rs6921438*rs2639990	−0.18
	rs2375981*rs2639990	−0.17
	*APOE*_non *ε4**rs34259399	−0.09
	rs6993770*rs34528081	−0.04
	*APOE*_non *ε4**rs6993770	−0.02

The accuracy of the model is 72% with a confidence interval CI = (0.6, 0.8) for 95%. The above accuracy has been calculated under a model with *P*-value: 0.009. The area under the curve (AUC) of the model is 81%, which indicates the detection of true positives (true cases predicted as cases) versus false positives (true controls predictive as cases).

Based on the annotation of the identified polymorphisms of the prediction model, we identified the following list of genes involved in AD risk: *VLDLR-AS1, APOE, KCNV2, ZADH2, C6orf223, LSR, ZFPM2,* and *VEGF-A*. After uploading the list to the IPA tool, the top five diseases found to be associated with the aforementioned were CVD, ophthalmic diseases,

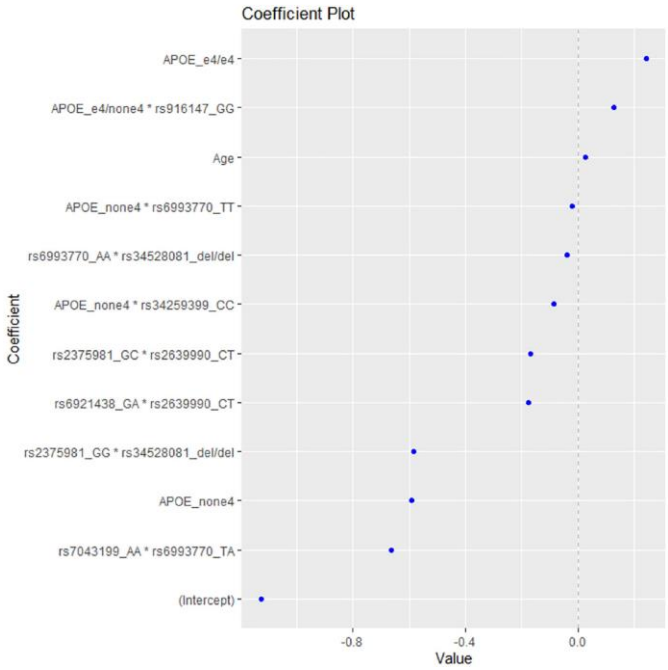

Figure 1. The coefficient plot of the EN model with the highest accuracy.

organismal injury and abnormalities, neurological diseases, and cancer.

Subsequently, a network linking most of these genes and other mediators was developed using IPA tool (Figure 2). The genes and mediators of this network have functions that correspond to cancer, dermatological diseases and conditions, organismal injury and abnormalities.

The network displayed in Figure 2 shows the multiple associations identified from the analyses, with the *VEGF-A* gene located at the center of many of the observed relationships involved in several metabolic pathways connected to cancer and inflammation. In this context, *VEGF-A* appears to interact with the *EPO, AKT, ERK, TGFB1* genes and the PI3K complex, indicating a potential effect in cellular activities, such as proliferation and apoptosis, where the *AKT, PI3K, ERK* and *TGFB1* cascades and pathways play a significant role. The identified network further enhances the aforementioned notions, by highlighting respective relationships between the *ERK, AKT, PI3K, TGFB1, APP* and the *APOE* gene. In addition, IPA analysis shows direct association between *APP* and the VEGF-A gene, among other associations including the *TGFB1, MPV17L, ZADH2, NPL, VASH2, GTDC1, EPB41L4A, LRRTM3, TRIM25, and ARHGAP27* genes.

DISCUSSION

Early detection of AD is important for the prognosis of patients, especially to initiate treatments during its pre-symptomatic phase and before the pathological amyloid and tau protein accumulation and the extensive brain damage, which can significantly decrease patient autonomy. Multiple approaches have been used for the early identification of AD-linked parameters including

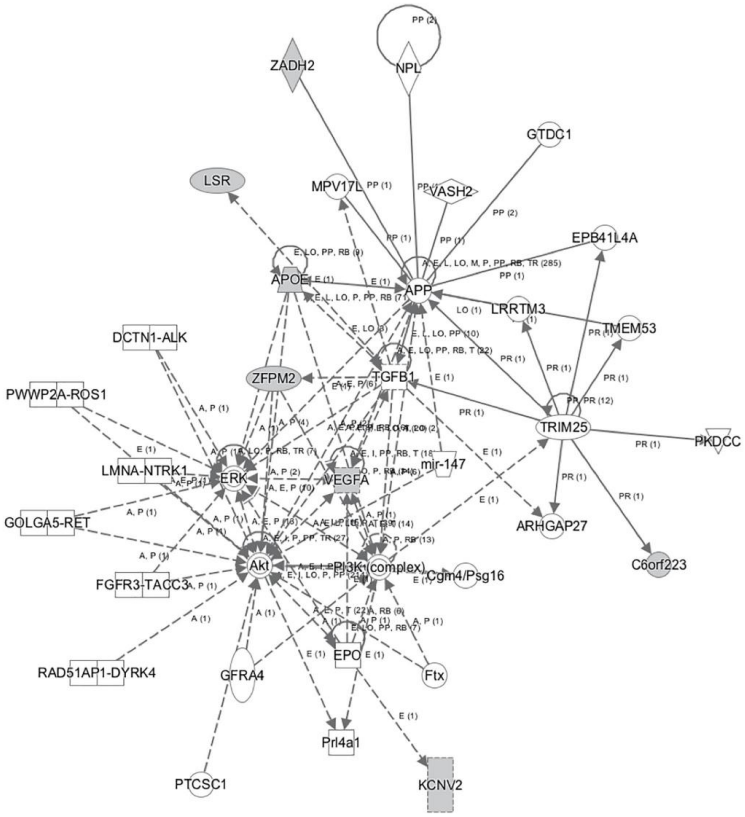

Figure 2. The network that links most of the identified genes that predict AD as generated by IPA tool.

medical assessment, as well as cognitive evaluations and imaging analyses that exploit multimodal biological and molecular features present in AD. Moreover, studies that use regression analyses have demonstrated the relation between AD and variables including clinical examination and cognitive test scores [27]. More recently, artificial intelligence approaches have been used, including supervised predictive analytics tools such as support vector machines, random forests, and artificial neural networks to distinguish AD cases from controls, and for identifying individuals with higher risk of AD in a given period of time [27]. In the present study, we used EN [29], a supervised machine learning method, to assess the role of genetic variants of common AD biomarkers on the prediction of AD risk.

The EN revealed one model with an adequate accuracy of 72% that mostly included epistatic interactions between the assessed variants as predictors of AD risk. The first and strongest genetic predictor was the *APOE e4* allele, which is in agreement with a large volume of scientific results that support this finding [9, 28]. Despite this, the precise processes by which ApoE affects AD remains unclear. AD is characterized by two main features encompassing the existence of extracellular deposition of Aβ generating amyloid plaques, and intracellular occurrence of neurofibrillary tangles (NFT) consisting of clusters of hyper-phosphorylated tau protein. ApoE may be associated with AD through its direct synergism with Aβ proteins. Indeed, the ApoE has been found in Aβ plaques of AD brains [30], and knock-out studies of *APOE* gene in mice have revealed that ApoE is critical for the initiation and retention of Aβ plaques. Furthermore, the direct synergistic link between ApoE and tau protein may also contribute to the association of ApoE and AD. Previously, the existence of ApoE was detected in Tau-NFT mediated deposits [31]. In addition, up-regulation of *APOE4* in neuronal cells of genetically engineered mice triggered a rise in tau hyper-phosphorylation induced by Erk stimulation [6].

Using the EN model, we observed the significant role of the interactions between *APOE* e4 allele polymorphism and the 2 SNPs of the *LSR* gene which we have identified in a previous study [12]. LSR is present in the central nervous system [13] and the *lsr*+/− mice display increased memory deficits following intracerebroventricular injection of the oligomeric soluble form of the β-amyloid peptide [32]. Animal studies have shown the presence of *LSR* gene transcripts in endothelial cells (ECs) of the blood-brain barrier (BBB) [33]. In *lsr*−/− knockout mice, the BBB does not appear to seal during embryogenesis [34], highlighting LSR's critical role in maintaining BBB integrity. Furthermore, given its role as an ApoE receptor, LSR is involved in lipid metabolism in the brain [13]; Herzine et al, manuscript in preparation) and could, therefore, play a role in AD development, with the present genetic associations further supporting this hypothesis.

The novel results of the present model include five epistatic interactions between VEGF-A-related variants and one *APOE**VEGF-A-related polymorphism interaction (rs6993770) associated with decreased risk for AD.

VEGF-A has been proposed as a promising novel therapeutic approach for AD [35]. Transplantation of mesenchymal stem cells into the double transgenic AD mouse model (APPswe/PS1dE9 mutations) leads to improvement of cognitive function [36]. Furthermore, higher VEGF-A concentration in the cerebrospinal fluid has been associated with slower cognitive decline in patients with AD risk [17]. Thus, VEGF-A could be considered as a protective factor for individuals having high risk of AD.

Concerning the interactions between *APOE* and VEGF-A, this is consistent with studies showing that VEGF-A exerts a neuroprotective effect in humanized *APOE ε4* mice, where treatment with VEGF-A leads to improvements of behavioral deficits [37]. A recent study also demonstrated that *APOE ε4* interacts with *VEGF-A* gene expression in the brain to affect cognitive performance [38]. Therefore, VEGF-A alone or in interaction with ApoE, seems to play an important role in AD risk, consistent with the results of this study.

The polymorphism rs6993770 is involved in 2 epistatic interactions with other VEGF-A-related SNPs and one interaction with *APOE ε4* allele. This constitutes an intronic variant of the *ZFPM2* (zinc finger protein, FOG family member 2) gene. The latter codes for the FOG family member 2, which is linked with repression of GATA mediated transcriptional activation [39, 40] and thus with hematopoiesis. The T minor allele has been associated with decreased VEGF-A levels [19, 20]. In the present model, this allele interacted with the non-*ε4* alleles of *APOE* gene to decrease the risk for AD. Also, the TA genotype of rs6993770 interacted with the AA genotype of rs7043199 to decrease the risk for AD. The rs7043199 SNP in an intronic variant of *VLDLR-AS1* gene and is located close to the *VLDLR* gene and its A allele has been associated with decreased VEGF-A levels [19, 20]. The VLDL receptor is a member of the low-density lipoprotein receptor family and binds ApoE. It is involved in pathways essential for the development of laminated structures and for the synaptic plasticity of the brain and is, thus, considered as a receptor that could be involved in the development

of AD [41]. It is important to note that this interaction presented the highest coefficient, thus is the strongest predictive factor of the model, with a protective role against AD higher than that of the absence of *ε4 APOE* allele. Our team was among the first to identify the allele 4 of the *APOE* gene as the strongest genetic risk factor for AD [8] and in this investigation we are proposing the rs7043199*rs6993770 interaction as a strong protective factor against AD. Finally, the A major allele of the rs6993770 (AA genotype) interacted with the deletion genotype of the rs34528081, which is an intergenic SNP close to *VEGF-A* gene, with the overall result of decreasing the risk for AD. Both alleles of these 2 SNPs have been associated with increased levels of VEGF-A [19, 20]. Thus, this interaction could be linked with further increase in VEGF-A levels that have been shown to have a protective effect against AD [17, 18]. The same deletion genotype of rs34528081 also interacted with the GG genotype of rs2375981, which is an intergenic SNP between *KCNV2* and *VLDLR* gene and whose G allele has been associated with decreased levels of VEGF-A [19, 20]. This interaction that is associated with decreased risk of AD could either be explained by a modification of VEGF-A levels or through an effect on the *VLDLR* gene. It has a very high coefficient in the model, similar to the absence of *APOE ε4* allele, and thus has a strong protective effect. Furthermore, the GC genotype of rs2375981 interacted with the CT genotype of rs2639990 to decrease the risk of AD. This is an intronic variant of *ZADH2* gene and the T allele has been associated with increased levels of VEGF-A [19, 20]. Hypomethylation of differentially methylated positions located on *ZADH2* has been observed in AD patients and have been associated with memory performance and cerebrospinal fluid levels of Aβ and tau [42], thus indicating a role of this gene on AD. Also, the same genotype of this SNP interacted with the GA genotype of rs6921438 to decrease the risk of AD. The rs6921438 is an intergenic SNP located between *LOC100132354* (lnc-RNA) and the *C6orf223* gene (encoding an uncharacterized protein) and is near the *VEGF-A* gene. The A allele is associated with decreased levels of VEGF-A [19, 20] and explains the highest percentage of VEGF-A levels variability (41.19%). This SNP (A allele) has also been associated with decreased HDL and increased LDL [22, 43]. It is thus a marker linked with both VEGF-A levels and lipid metabolism, which could mediate its relationship with AD.

The strengths of the present study include the use of the EN method in the development of the prediction model, which is a machine learning approach more powerful than classical statistics methodologies. Furthermore, the identified model showed a sufficient accuracy of 72% and an AUC of 81%. The accuracy of the model is comparable to other tools [44–46] and this is very important as our model uses genetic factors. In fact, in other studies, the addition of a genetic score led to a small improvement of prediction of the classical variables model [47]. In a recent review, 61 papers describing dementia risk models were identified and most of them had moderate-to-high predictive ability (AUC > 0.70). The highest AUC value was 0.932 [48] and our AUC is 81%. All these data highlight the satisfactory accuracy of the identified prediction model.

A few limitations of the study, however, include the relatively small sample size and the lack of VEGF-A levels' measurements in the studied populations.

In summary, the prediction model proposed in the present study consists of 8 epistatic interactions that, in combination with the *APOE ε4* allele, directly affect the risk for AD. These interactions involve 9 polymorphisms in 8 genes: *VLDLR-AS1, APOE, KCNV2, ZADH2, C6orf223, LSR, ZFPM2,* and *VEGF-A*.

IPA analysis highlighted relationships between the identified genes and neurological diseases, within the first top five disorders associated with said genes, including CVD and cancer. This finding indicates that the genetic determinants of the selected biomarkers (VEGF-A, LSR and APOE) could act as common links between important chronic diseases. In fact, most of these genes are shown to be linked in the context of an enlarged common network, the functions of which correspond to cancer, dermatological diseases and conditions, organismal injury and abnormalities.

In conclusion, these novel epistatic interactions between *APOE*, *LSR* and VEGF-A related polymorphisms allow for prediction of AD risk, constituting not only a useful prediction model, but also providing new insights about molecular mechanisms that can be involved in AD development which could be useful as biomarkers and/or treatment targets. We are also proposing two epistatic interactions (rs7043199*rs6993770 and the rs2375981*rs34528081) between VEGF-A-related polymorphisms as strong protective factors against AD.

METHODS

Population and data collection

The study population consisted of 1078 (602 controls and 476 cases) unrelated adults of European origin, recruited during the period 1996–1998. The study was approved by the related ethics committees and all participants provided written consent prior to their enrollment in the study.

The recruitment and data collection procedures of the present population have been previously extensively described [9]. The clinical diagnosis of AD was based on 5 criteria: the National Institute of Neurological and Communicative Disorders and Stroke and the AD and Related Disorders Association classification of probable AD [28]; the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [28]; the International Classification of Diseases and Related Health Problems, 10th Edition [28]; the Mini Mental State Examination score equal to or less than 23 (Mini Mental State Examination >23 for controls) [28]; and, lastly the modified ischemic scale less than three [28]. Secondary causes of dementia were excluded by computerized tomography scan of the brain. Individuals with other chronic or neurological diseases, such as cancer and Parkinson's diseases, respectively, were excluded from the study. Participants serving as controls followed similar assessment and interviews as cases, except for the computerized tomography scan. All controls presented a Mini Mental State Examination >23 and were free of dementia.

Genotyping data were available in a subsample of the population. The final sample size after quality control for genotyping and outliers exclusion was 323 individuals (143 AD cases and 180 controls).

Genotyping analyses

DNA was extracted from peripheral blood [49] and all samples were stored in biobanks of the BRC IGE-PCV (Biological Resources Center 'Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire' BB-0033-00051). Two SNPs of the *LSR* gene (rs34259399 and rs916147) previously identified as candidate variants for AD were genotyped, along with the 10 VEGF-A-related polymorphisms (rs6921438, rs7043199, rs6993770, rs2375981, rs34528081, rs4782371, rs2639990, rs10761741, rs114694170, rs1740073) and the common *APOE* variants. The genotyping analyses were performed in LGC genomics (http://www.lgcgroup.com) using the competitive allele-specific PCR (KASP) chemistry coupled with a Förster resonance energy transfer-based genotyping system (http://www.kbioscience.co.uk/reagents/KASP/KASP.html). Two of the *APOE* common polymorphisms rs429358 (Cys112Arg) and rs7412 (Arg158Cys) were genotyped as previously described [50].

Statistical analyses

The agreement of the frequencies of genotypes with the Hardy-Weinberg equilibrium was tested using the chi-squared test.

For the purposes of developing the prediction model, a machine learning method was applied, entitled "Elastic Net" (EN) [29], with the aim of identifying the strongest predictors for the risk of AD combining all genotyping data (direct effects and epistatic interactions), as well as age and sex. In short, the method tests hundreds of logistic regression models and penalizes each of them, in order to reach the final, optimal one. The former constitutes an extensively used method, especially in the development of disease prediction models with special attention to the integration of omics data [51–54]. Furthermore, the method can allow accurate predictions with smaller sample sizes as it tolerates a big number of predictors [29] and it is also used in analyses entailing genetic data, as well as GWAS studies [55, 56] and it is considered to function better in cases where gene * environment interactions are involved in the prediction of a disease [57]. Comparison studies have shown that EN is a powerful tool, especially when additive gene effect is expected [58]. It has also been previously used in AD risk prediction with clinical factors, imaging and omics [59–61] but not to study the effect of candidate genes as predicting factors.

In the context of the present analyses, *APOE* genotypes were divided into 3 groups according to the potential presence of *ε4* allele (coded as 2 = *ε4/4*; 1 = *ε2/4*, *ε3/4*; and 0 = *ε2/2, ε3/2, ε3/3)*. The complete dataset was split into two separate datasets, the "train" dataset that included the 80% of the whole population (AD patients + controls) and the "test" dataset that included the remaining 20% of the whole population (AD patients + controls). This selection was random and was performed by the R command 'createDataPartition' that is specific to this. The EN analysis was performed using the R software.

Pathway analysis

All genes identified to participate in AD risk prediction were further assessed using the QIAGEN Ingenuity Pathway Analysis (IPA) tool, in order to examine potential relationships and identify causal links, with the overall aim of proposing mechanisms to explain the results deriving from the prediction model.

Abbreviations

AD: Alzheimer's disease; APOE/ApoE: Apolipoprotein E; APP: Amyloid Beta Precursor; ARHGAP27: Rho GTPase Activating Protein 27; BBB: Blood-brain barrier; ECs: Endothelial Cells; EN: Elastic Net; EPB41L4A: Erythrocyte Membrane Protein Band 4.1 Like 4A; EPO: Erythropoietin; GTDC1: Glycosyltransferase Like Domain Containing 1;

LRRTM3: Leucine Rich Repeat Transmembrane Neuronal 3; LSR: Lipolysis Stimulated Lipoprotein Receptor; MPV17L: Mitochondrial Inner Membrane Protein Like; NPL: N-Acetylneuraminate Pyruvate Lyase; PI3K: Phospoinositide 3-kinase; SNP(s): Single Nucleotide Polymorphism(s); TRIM25: Tripartite Motif Containing 25; VASH2: Vasohibin 2; VLDL: Very low-density lipoprotein; ZADH2: Zinc Binding Alcohol Dehydrogenase Domain Containing 2; SD: Standard deviation; MAF: Minor allele frequency; HW: Hardy-Weinberg; ZFPM2: Zinc finger protein, FOG family member 2; NFT: Neurofibrillary tangles; IPA: Ingenuity Pathway Analysis; VEGF-A: Vascular endothelial growth factor A; CVD: Cardiovascular disease.

AUTHOR CONTRIBUTIONS

VSS conceived hypothesis and with PAM the design of the study. PAM designed and performed the statistical analysis. MC, MH, LJ and FP contributed to data collection and genotyping. PAM, SMG, KM, DG, FTY and VSS contributed to data interpretation. PAM drafted the first version of the manuscript. All authors critically contributed to the final version of the manuscript. VSS supervised and assured project implementation.

ACKNOWLEDGMENTS

We would like to thank the “Agence Nationale de la Recherche - Programme d’Investissements d’Avenir” - PIA (grant number ANR-15RHU-0004), the “Fondation Plan Alzheimer”- DYSLIP-AD project – 2019, the “Region Grand Est” - project “Gutenberg Chair” (Georges DEDOUSSIS) – 2018 and the VEGF Consortium (http://vegfconsortium.org).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

The present investigation was funded in part by the “Agence Nationale de la Recherche - Programme d’Investissements d’Avenir” - PIA (grant number ANR-15RHU-0004), the “Fondation Plan Alzheimer”- DYSLIP-AD project - 2019 and the “Region Grand Est_ Gutenberg Chair” project (Georges DEDOUSSIS) – 2018.

REFERENCES

1. Soria Lopez JA, González HM, Léger GC. Alzheimer's disease. Handb Clin Neurol. 2019; 167:231–55. https://doi.org/10.1016/B978-0-12-804766-8.00013-3 PMID:31753135

2. Lane CA, Hardy J, Schott JM. Alzheimer's disease. Eur J Neurol. 2018; 25:59–70. https://doi.org/10.1111/ene.13439 PMID:28872215

3. Mayeux R, Stern Y. Epidemiology of Alzheimer disease. Cold Spring Harb Perspect Med. 2012; 2:a006239. https://doi.org/10.1101/cshperspect.a006239 PMID:22908189

4. Tanzi RE. The genetics of Alzheimer disease. Cold Spring Harb Perspect Med. 2012; 2:a006296. https://doi.org/10.1101/cshperspect.a006296 PMID:23028126

5. Institute NHGR. GWAS Catalog. 2021. https://www.ebi.ac.uk/gwas/.

6. Muñoz SS, Garner B, Ooi L. Understanding the Role of ApoE Fragments in Alzheimer's Disease. Neurochem Res. 2019; 44:1297–305. https://doi.org/10.1007/s11064-018-2629-1 PMID:30225748

7. Kim J, Basak JM, Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. Neuron. 2009; 63:287–303. https://doi.org/10.1016/j.neuron.2009.06.026 PMID:19679070

8. Lucotte G, David F, Visvikis S, Leininger-Müller B, Siest G, Babron MC, Couderc R. Apolipoprotein E-epsilon 4 allele and Alzheimer's disease. Lancet. 1993; 342:1309. PMID:7901625

9. Siest G, Bertrand P, Qin B, Herbeth B, Serot JM, Masana L, Ribalta J, Passmore AP, Evans A, Ferrari M, Franceschi M, Shepherd J, Cuchel M, et al. Apolipoprotein E polymorphism and serum concentration in Alzheimer's disease in nine European centres: the ApoEurope study. ApoEurope group. Clin Chem Lab Med. 2000; 38:721–30. https://doi.org/10.1515/CCLM.2000.102 PMID:11071064

10. Liu S, Liu J, Weng R, Gu X, Zhong Z. Apolipoprotein E gene polymorphism and the risk of cardiovascular disease and type 2 diabetes. BMC Cardiovasc Disord. 2019; 19:213. https://doi.org/10.1186/s12872-019-1194-0 PMID:31521122

11. Geisel J, Bunte T, Bodis M, Oette K, Herrmann W. Apolipoprotein E2/E2 genotype in combination with mutations in the LDL receptor gene causes type III hyperlipoproteinemia. Clin Chem Lab Med. 2002; 40:475–9.

https://doi.org/10.1515/CCLM.2002.08
PMID:12113291

12. Xie T, Akbar S, Stathopoulou MG, Oster T, Masson C, Yen FT, Visvikis-Siest S. Epistatic interaction of apolipoprotein E and lipolysis-stimulated lipoprotein receptor genetic variants is associated with Alzheimer's disease. Neurobiol Aging. 2018; 69:292.e1–e5. https://doi.org/10.1016/j.neurobiolaging.2018.04.013 PMID:29858039

13. Stenger C, Pinçon A, Hanse M, Royer L, Comte A, Koziel V, Olivier JL, Pillot T, Yen FT. Brain region-specific immunolocalization of the lipolysis-stimulated lipoprotein receptor (LSR) and altered cholesterol distribution in aged LSR+/- mice. J Neurochem. 2012; 123:467–76. https://doi.org/10.1111/j.1471-4159.2012.07922.x PMID:22909011

14. Yen FT, Roitel O, Bonnard L, Notet V, Pratte D, Stenger C, Magueur E, Bihain BE. Lipolysis stimulated lipoprotein receptor: a novel molecular link between hyperlipidemia, weight gain, and atherosclerosis in mice. J Biol Chem. 2008; 283:25650–9. https://doi.org/10.1074/jbc.M801027200 PMID:18644789

15. Beazley-Long N, Hua J, Jehle T, Hulse RP, Dersch R, Lehrling C, Bevan H, Qiu Y, Lagrèze WA, Wynick D, Churchill AJ, Kehoe P, Harper SJ, et al. VEGF-A165b is an endogenous neuroprotective splice isoform of vascular endothelial growth factor A in vivo and in vitro. Am J Pathol. 2013; 183:918–29. https://doi.org/10.1016/j.ajpath.2013.05.031 PMID:23838428

16. Tarkowski E, Issa R, Sjögren M, Wallin A, Blennow K, Tarkowski A, Kumar P. Increased intrathecal levels of the angiogenic factors VEGF and TGF-beta in Alzheimer's disease and vascular dementia. Neurobiol Aging. 2002; 23:237–43. PMID:11804709

17. Hohman TJ, Bell SP, Jefferson AL, and Alzheimer's Disease Neuroimaging Initiative. The role of vascular endothelial growth factor in neurodegeneration and cognitive decline: exploring interactions with biomarkers of Alzheimer disease. JAMA Neurol. 2015; 72:520–9. https://doi.org/10.1001/jamaneurol.2014.4761 PMID:25751166

18. Huang L, Jia J, Liu R. Decreased serum levels of the angiogenic factors VEGF and TGF-β1 in Alzheimer's disease and amnestic mild cognitive impairment. Neurosci Lett. 2013; 550:60–3. https://doi.org/10.1016/j.neulet.2013.06.031 PMID:23827227

19. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, Safa R, Azimi Nezhad M, Sawyer D, Marteau JB, Xanthakis V, Siest G, Sullivan L, 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. https://doi.org/10.1161/CIRCRESAHA.111.243790 PMID:21757650

20. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, Concas MP, Traglia M, Barbieri C, Ndiaye NC, Stathopoulou MG, Lagou V, Maestrale GB, 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. https://doi.org/10.1371/journal.pgen.1005874 PMID:26910538

21. Azimi-Nezhad M, Stathopoulou MG, Bonnefond A, Rancier M, Saleh A, Lamont J, Fitzgerald P, Ndiaye NC, Visvikis-Siest S. Associations of vascular endothelial growth factor (VEGF) with adhesion and inflammation molecules in a healthy population. Cytokine. 2013; 61:602–7. https://doi.org/10.1016/j.cyto.2012.10.024 PMID:23201487

22. Stathopoulou MG, Bonnefond A, Ndiaye NC, Azimi-Nezhad M, El Shamieh S, Saleh A, Rancier M, Siest G, Lamont J, Fitzgerald P, Visvikis-Siest S. 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; 54:535–41. https://doi.org/10.1194/jlr.P030551 PMID:23204297

23. Xie T, Stathopoulou MG, de Andrés F, Siest G, Murray H, Martin M, Cobaleda J, Delgado A, Lamont J, Peñas-LIedó E, LLerena A, Visvikis-Siest S. VEGF-related polymorphisms identified by GWAS and risk for major depression. Transl Psychiatry. 2017; 7:e1055. https://doi.org/10.1038/tp.2017.36 PMID:28267147

24. Rancier M, Zaaber I, Stathopoulou MG, Chatelin J, Saleh A, Marmouch H, El Shamieh S, Masson C, Murray H, Lamont J, Fitzgerald P, Mahjoub S, Said K, et al. Pro- and anti-angiogenic VEGF mRNAs in autoimmune thyroid diseases. Autoimmunity. 2016; 49:366–72. https://doi.org/10.1080/08916934.2016.1199019 PMID:27494076

25. Zaaber I, Rancier M, Stathopoulou MG, Saleh A, Marmouch H, Masson C, Murray H, Kurth MJ, Lamont J, Fitzgerald P, Mahjoub S, Said K, Bel Hadj Jrad

Tensaout B, et al. Plasma VEGF-related polymorphisms are implied in autoimmune thyroid diseases. Autoimmunity. 2016; 49:229–35. https://doi.org/10.3109/08916934.2016.1151005 PMID:26955881

26. Stathopoulou MG, Xie T, Ruggiero D, Chatelin J, Rancier M, Weryha G, Kurth MJ, Aldasoro Arguinano AA, Gorenjak V, Petrelis AM, Dagher G, Dedoussis G, Deloukas P, et al, and 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. Clin Chem Lab Med. 2018; 56:83–6. https://doi.org/10.1515/cclm-2017-0838 PMID:29087954

27. Marinescu RV, Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW, Barkhof F, Fox NC, Golland P, Klein S, Alexander DC. TADPOLE Challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data. Predict Intell Med. 2019; 11843:1–10. https://doi.org/10.1007/978-3-030-32281-6_1 PMID:32587957

28. Van Cauwenberghe C, Van Broeckhoven C, Sleegers K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med. 2016; 18:421–30. https://doi.org/10.1038/gim.2015.117 PMID:26312828

29. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B (Stat Methodol). 2005; 67:301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x

30. Burns MP, Noble WJ, Olm V, Gaynor K, Casey E, LaFrancois J, Wang L, Duff K. Co-localization of cholesterol, apolipoprotein E and fibrillar Abeta in amyloid plaques. Brain Res Mol Brain Res. 2003; 110:119–25. https://doi.org/10.1016/s0169-328x(02)00647-2 PMID:12573540

31. Yamaguchi H, Ishiguro K, Sugihara S, Nakazato Y, Kawarabayashi T, Sun X, Hirai S. Presence of apolipoprotein E on extracellular neurofibrillary tangles and on meningeal blood vessels precedes the Alzheimer beta-amyloid deposition. Acta Neuropathol. 1994; 88:413–9. https://doi.org/10.1007/BF00389492 PMID:7847069

32. Pinçon A, Thomas MH, Huguet M, Allouche A, Colin JC, Georges A, Derrien A, Lanhers MC, Malaplate-Armand C, Oster T, Corbier C, Pillot T, Olivier JL, Yen FT. Increased susceptibility of dyslipidemic LSR+/- mice to amyloid stress is associated with changes in cortical cholesterol levels. J Alzheimers Dis. 2015; 45:195–204. https://doi.org/10.3233/JAD-142127 PMID:25690661

33. Daneman R, Zhou L, Agalliu D, Cahoy JD, Kaushal A, Barres BA. The mouse blood-brain barrier transcriptome: a new resource for understanding the development and function of brain endothelial cells. PLoS One. 2010; 5:e13741. https://doi.org/10.1371/journal.pone.0013741 PMID:21060791

34. Sohet F, Lin C, Munji RN, Lee SY, Ruderisch N, Soung A, Arnold TD, Derugin N, Vexler ZS, Yen FT, Daneman R. LSR/angulin-1 is a tricellular tight junction protein involved in blood-brain barrier formation. J Cell Biol. 2015; 208:703–11. https://doi.org/10.1083/jcb.201410131 PMID:25753034

35. Shim JW, Madsen JR. VEGF Signaling in Neurological Disorders. Int J Mol Sci. 2018; 19:275. https://doi.org/10.3390/ijms19010275 PMID:29342116

36. Garcia KO, Ornellas FL, Martin PK, Patti CL, Mello LE, Frussa-Filho R, Han SW, Longo BM. Therapeutic effects of the transplantation of VEGF overexpressing bone marrow mesenchymal stem cells in the hippocampus of murine model of Alzheimer's disease. Front Aging Neurosci. 2014; 6:30. https://doi.org/10.3389/fnagi.2014.00030 PMID:24639647

37. Salomon-Zimri S, Glat MJ, Barhum Y, Luz I, Boehm-Cagan A, Liraz O, Ben-Zur T, Offen D, Michaelson DM. Reversal of ApoE4-Driven Brain Pathology by Vascular Endothelial Growth Factor Treatment. J Alzheimers Dis. 2016; 53:1443–58. https://doi.org/10.3233/JAD-160182 PMID:27372644

38. Moore AM, Mahoney E, Dumitrescu L, De Jager PL, Koran MEI, Petyuk VA, Robinson RA, Ruderfer DM, Cox NJ, Schneider JA, Bennett DA, Jefferson AL, Hohman TJ. APOE ε4-specific associations of VEGF gene family expression with cognitive aging and Alzheimer's disease. Neurobiol Aging. 2020; 87:18–25. https://doi.org/10.1016/j.neurobiolaging.2019.10.021 PMID:31791659

39. Svensson EC, Tufts RL, Polk CE, Leiden JM. Molecular cloning of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. Proc Natl Acad Sci U S A. 1999; 96:956–61. https://doi.org/10.1073/pnas.96.3.956 PMID:9927675

40. Lin AC, Roche AE, Wilk J, Svensson EC. The N termini of Friend of GATA (FOG) proteins define a novel transcriptional repression motif and a superfamily of transcriptional repressors. J Biol Chem. 2004; 279:55017–23. https://doi.org/10.1074/jbc.M411240200 PMID:15507435

41. Dlugosz P, Nimpf J. The Reelin Receptors Apolipoprotein E receptor 2 (ApoER2) and VLDL Receptor. Int J Mol Sci. 2018; 19:3090. https://doi.org/10.3390/ijms19103090 PMID:30304853

42. Madrid A, Hogan KJ, Papale LA, Clark LR, Asthana S, Johnson SC, Alisch RS. DNA Hypomethylation in Blood Links B3GALT4 and ZADH2 to Alzheimer's Disease. J Alzheimers Dis. 2018; 66:927–34. https://doi.org/10.3233/JAD-180592 PMID:30372681

43. Belhassan K, Ouldim K, Sefiani AA. Genetics and genomic medicine in Morocco: the present hope can make the future bright. Mol Genet Genomic Med. 2016; 4:588–98. https://doi.org/10.1002/mgg3.255 PMID:27896281

44. Danso SO, Zeng Z, Muniz-Terrera G, Ritchie CW. Developing an Explainable Machine Learning-Based Personalised Dementia Risk Prediction Model: A Transfer Learning Approach With Ensemble Learning Algorithms. Front Big Data. 2021; 4:613047. https://doi.org/10.3389/fdata.2021.613047 PMID:34124650

45. Qiu S, Joshi PS, Miller MI, Xue C, Zhou X, Karjadi C, Chang GH, Joshi AS, Dwyer B, Zhu S, Kaku M, Zhou Y, Alderazi YJ, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain. 2020; 143:1920–33. https://doi.org/10.1093/brain/awaa137 PMID:32357201

46. Wang L, Li P, Hou M, Zhang X, Cao X, Li H. Construction of a risk prediction model for Alzheimer's disease in the elderly population. BMC Neurol. 2021; 21:271. https://doi.org/10.1186/s12883-021-02276-8 PMID:34233656

47. Chouraki V, Reitz C, Maury F, Bis JC, Bellenguez C, Yu L, Jakobsdottir J, Mukherjee S, Adams HH, Choi SH, Larson EB, Fitzpatrick A, Uitterlinden AG, et al, and International Genomics of Alzheimer's Project. Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. J Alzheimers Dis. 2016; 53:921–32. https://doi.org/10.3233/JAD-150749 PMID:27340842

48. Hou XH, Feng L, Zhang C, Cao XP, Tan L, Yu JT. Models for predicting risk of dementia: a systematic review. J Neurol Neurosurg Psychiatry. 2019; 90:373–9. https://doi.org/10.1136/jnnp-2018-318212 PMID:29954871

49. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988; 16:1215. https://doi.org/10.1093/nar/16.3.1215 PMID:3344216

50. Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. J Lipid Res. 1990; 31:545–8. PMID:2341813

51. Li J, Guasch-Ferré M, Chung W, Ruiz-Canela M, Toledo E, Corella D, Bhupathiraju SN, Tobias DK, Tabung FK, Hu J, Zhao T, Turman C, Feng YA, et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. Eur Heart J. 2020; 41:2645–56. https://doi.org/10.1093/eurheartj/ehaa209 PMID:32406924

52. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, Zhu S, Ye Z. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med. 2019; 17:119. https://doi.org/10.1186/s12967-019-1860-0 PMID:30971285

53. Das J, Gayvert KM, Bunea F, Wegkamp MH, Yu H. ENCAPP: elastic-net-based prognosis prediction and biomarker discovery for human cancers. BMC Genomics. 2015; 16:263. https://doi.org/10.1186/s12864-015-1465-9 PMID:25887568

54. Millstein J, Budden T, Goode EL, Anglesio MS, Talhouk A, Intermaggio MP, Leong HS, Chen S, Elatre W, Gilks B, Nazeran T, Volchek M, Bentley RC, et al, and AOCS Group. Prognostic gene expression signature for high-grade serous ovarian cancer. Ann Oncol. 2020; 31:1240–50. https://doi.org/10.1016/j.annonc.2020.05.019 PMID:32473302

55. Cho S, Kim H, Oh S, Kim K, Park T. Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. BMC Proc. 2009 (Suppl 7); 3:S25. https://doi.org/10.1186/1753-6561-3-s7-s25 PMID:20018015

56. Waldmann P, Mészáros G, Gredler B, Fuerst C, Sölkner J. Evaluation of the lasso and the elastic net in genome-wide association studies. Front Genet. 2013; 4:270.

https://doi.org/10.3389/fgene.2013.00270
PMID:24363662

57. Frost HR, Shen L, Saykin AJ, Williams SM, Moore JH, and Alzheimers Disease Neuroimaging Initiative. Identifying significant gene-environment interactions using a combination of screening testing and hierarchical false discovery rate control. Genet Epidemiol. 2016; 40:544–57.
https://doi.org/10.1002/gepi.21997
PMID:27578615

58. Momen M, Mehrgardi AA, Sheikhi A, Kranis A, Tusell L, Morota G, Rosa GJM, Gianola D. Predictive ability of genome-assisted statistical models under various forms of gene action. Sci Rep. 2018; 8:12309.
https://doi.org/10.1038/s41598-018-30089-2
PMID:30120288

59. Bouts MJR, Möller C, Hafkemeijer A, van Swieten JC, Dopper E, van der Flier WM, Vrenken H, Wink AM, Pijnenburg YAL, Scheltens P, Barkhof F, Schouten TM, de Vos F, et al. Single Subject Classification of Alzheimer's Disease and Behavioral Variant Frontotemporal Dementia Using Anatomical, Diffusion Tensor, and Resting-State Functional Magnetic Resonance Imaging. J Alzheimers Dis. 2018; 62:1827–39.
https://doi.org/10.3233/JAD-170893
PMID:29614652

60. Liu Y, Thalamuthu A, Mather KA, Crawford J, Ulanova M, Wong MWK, Pickford R, Sachdev PS, Braidy N. Plasma lipidome is dysregulated in Alzheimer's disease and is associated with disease risk genes. Transl Psychiatry. 2021; 11:344.
https://doi.org/10.1038/s41398-021-01362-2
PMID:34092785

61. de Vos F, Schouten TM, Hafkemeijer A, Dopper EG, van Swieten JC, de Rooij M, van der Grond J, Rombouts SA. Combining multiple anatomical MRI measures improves Alzheimer's disease classification. Hum Brain Mapp. 2016; 37:1920–9.
https://doi.org/10.1002/hbm.23147
PMID:26915458

Publication 6

Development of an automated analysis pipeline for diseases risk prediction models and an application for populations data visualisation.

Petrelis AM, …., Visvikis-Siest S.

Under preparation

My specific contribution in this study was: Participation in the design of the study, conception and realisation of analyses, participation in the data interpretation, and drafting of the first version of the manuscript.

Introduction

Chronic diseases risk prediction is an important target to allow early diseases prevention and management [61]. The development of risk prediction models and algorithms is mediating this task and these models can be useful tools for medical decisions [62]. During the last decades, the advances of human genetics and genomics allow also the personalisation of risk prediction models with the inclusion of genetic variables along with clinical data as part of the concept of "genomic medicine" [63]. These models can assist clinicians to make well-informed medical and preventive decisions and to personalise and customise treatments [64], but should be comprehensible and easy to use in order to be applicable in clinical practice.

A major improvement in the development of risk prediction models is the application of AI, especially ML methodologies, as they comprise an evolutionally power for personalising risk prediction with multiple computational advantages [65, 66].

Furthermore, the growing amount of data employed in PM and especially the genetics and genomics studies, require the implementation of data visualisation and ML techniques [67]. Interactive and user-friendly data visualisation can lead to successful data interpretation by addressing known and identifying novel knowledge through new assumptions and hypotheses formation and further allow an extended communication of the discoveries [68, 69]. Visualisation tools should be simple to use and manage large amounts of data and provide highly comprehensive and quality images for better communication [70].

Objective

Two tools have been created:

1) An automated pipeline for the analyses for the development of diseases risk prediction algorithms
2) A desktop application for the visualisation of populations' data

The automated pipeline that was developed based on the methodology used for the creation of the previously described AD risk prediction model had a double aim:

1) To facilitate the analyses for the development of new prediction models in new populations with the same or different variables and for different diseases
2) To enable the fast application of the predictive models in new individuals in order to define their risk for chronic diseases

The first aim leads to the design of an automated algorithm for standardised analyses addressed to researchers, who want to develop novel prediction models for diseases using the EN methodology. The second aim is expected to transform the prediction model into a useful tool for the prediction of diseases risk of new patients by researchers and health providers.
The aim of the visualisation application is to allow fast visualisation of the characteristics of different populations and preliminary analyses that can facilitate the development of new hypotheses and studies.

Population and analysis

For the development of the automated pipeline, the AD cases-control population was used and the steps of the analysis for the development of the prediction model of the Publication 5 were standardized and combined in an interface using R Shiny package of R software.
The algorithm for analyses consists of the following steps:

1. Data importation
2. Quality control
3. EN analysis

The visualisation application was developed using the Microsoft Power BI. Power BI allows the connection of different datasets, the data visualisation and/or discovery, and the data sharing [71]. It can be installed at any desktop while its use is very easy due to its user-friendly format that allows non-specialised users, like researchers and medical stuff, to profit from its services. The

populations of publications 3 and 4 was used to present the interface.

Results

The algorithm has been made in R language. The R script of this algorithm requires the installation of R and the R studio. Then all requiered packages will automatically be installed.

The scripts have been integrated and trasformed in a user-friendly interface that was developed with R Shiny package, named RSGEN. This interface allows the automation of the identified prediction model for AD and a standardised pipeline with necessary steps for the development of new prediction models for any binary condition.

All the necessary scripts of the interface have been combined in a single script file called app.R. app.R has three components:

- a user interface function (ui72.R)
- a server function (serv72.R)
- a call to the shinyApp function

The user interface (ui) object controls the layout and appearance of the application. The server function contains the instructions that the computer needs to build the application. Finally, the shinyApp function creates Shiny app objects from an explicit ui/server pair.

The detailed scripts for the development of the interface are presented in the **annex1**.

By opening R Studio and the app.R file and running it's commands, the interface appears automatically.

The RSGEN interface is presented in the Figures 18-29.

The first dashboard allows the importation of the data and the selection of customised values for quality control parameters that were considered in the analyses for the AD risk prediction model: Hardy Weinberg p-value, percentage of missing data per individual, percentage of missing genotypes per individuals and percentage of missing values per variable. Also, it allows the selection of the percentage for the partition of the population for the random creation of the test and train datasets for the machine learning EN algorithm (Figure 18). It presents a summary of all variables: categorical, numeric and genetic. It is the tab where the variable to predict is specified e.g. AD case or control.

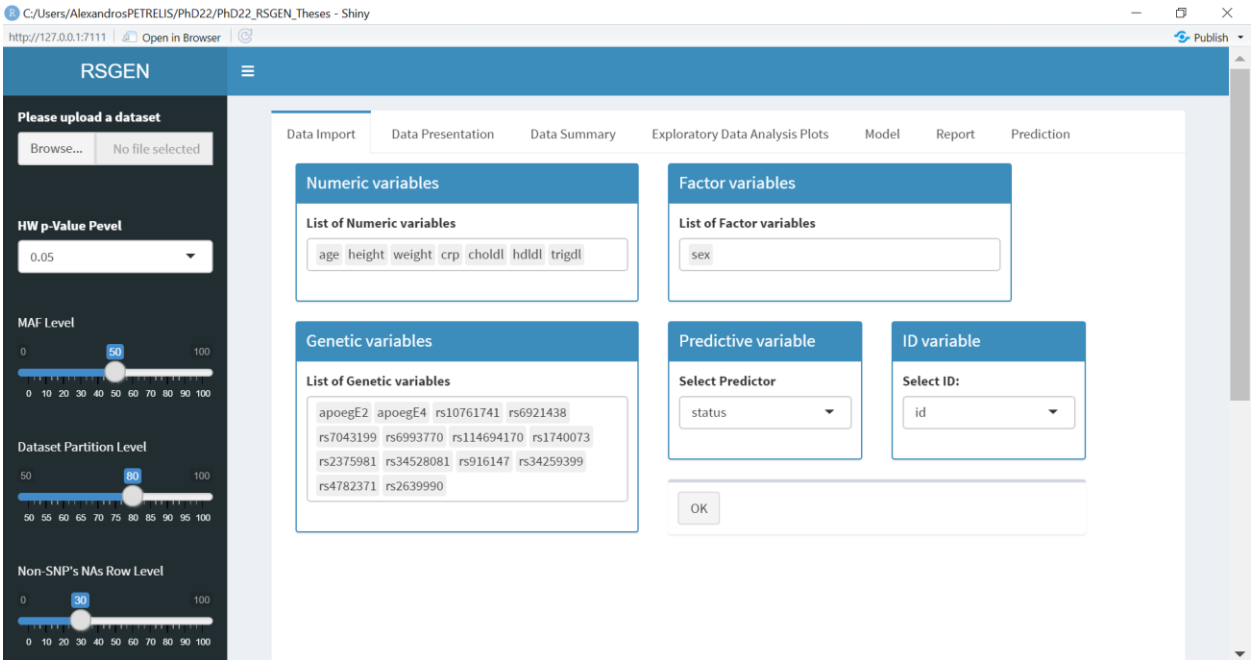

Figure 18. Presentation of the first dashboard of the analytical interface with the criteria for customised analyses

The second dashboard presents the dataset that was uploaded (Figure 19)

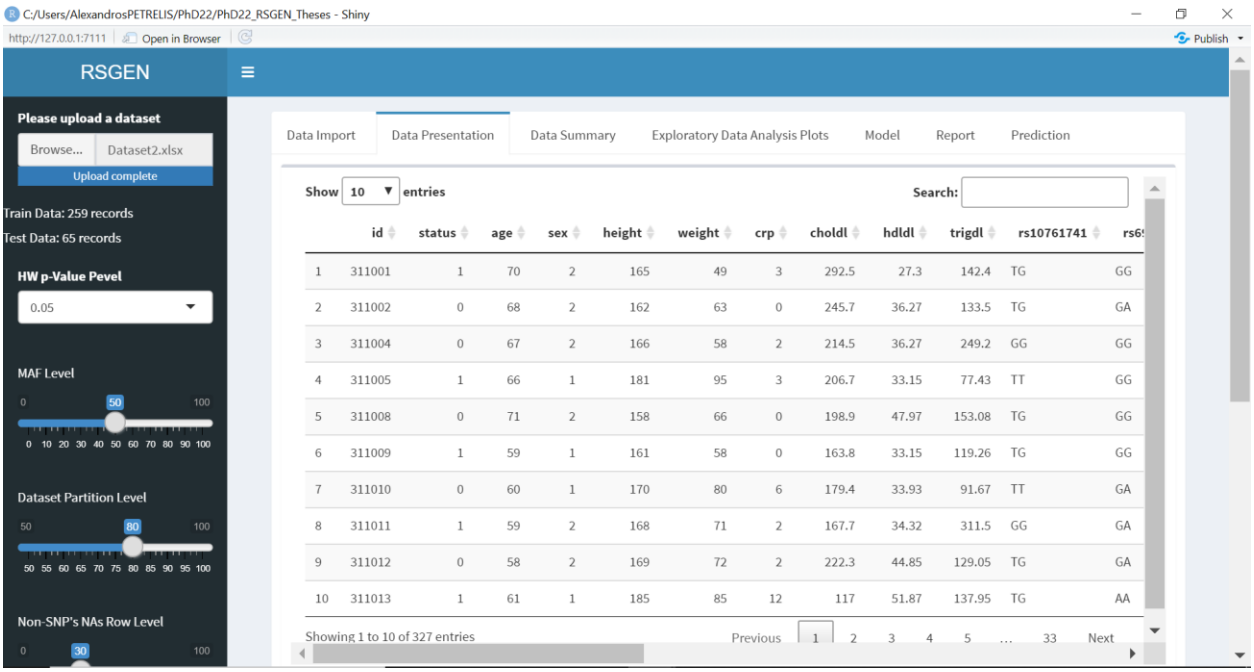

Figure 19. The second dashboard of the analytical interface with the presentation of the dataset

The third dashboard is generating descriptive data for the dataset (Figure 20).

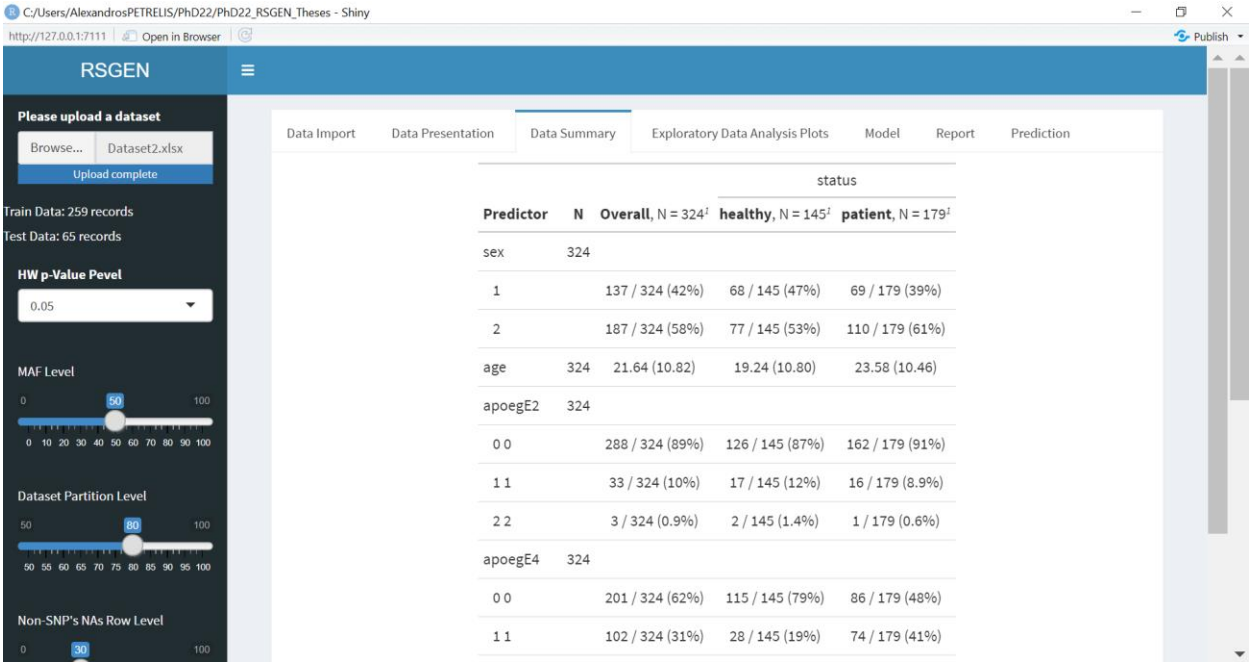

Figure 20. Descriptive analysis of the dataset in the third dashboard of the analytical interface

The fourth dashboard presents the percentages of the missing values and the investigation of a potential missing values pattern (Figures 21).

Figure 21. Missing values quantification and qualification in the forth dashboard of the analytical interface

The fifth dashboard presents a genetic quality control by summarising HW testing (Figure 22).

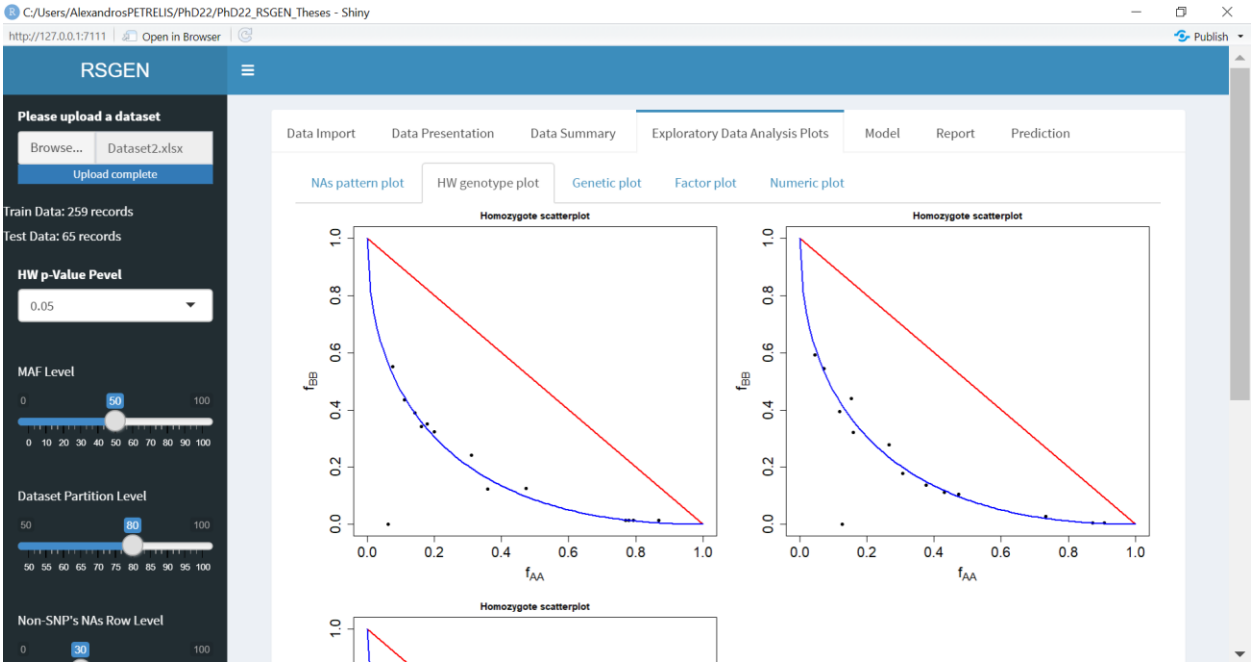

Figure 22. Genetic quality control by summarising HW testing in the fifth dashboard of the analytical interface

The sixth dashboard presents the descriptive characteristics of genetic metrics (Figure 23).

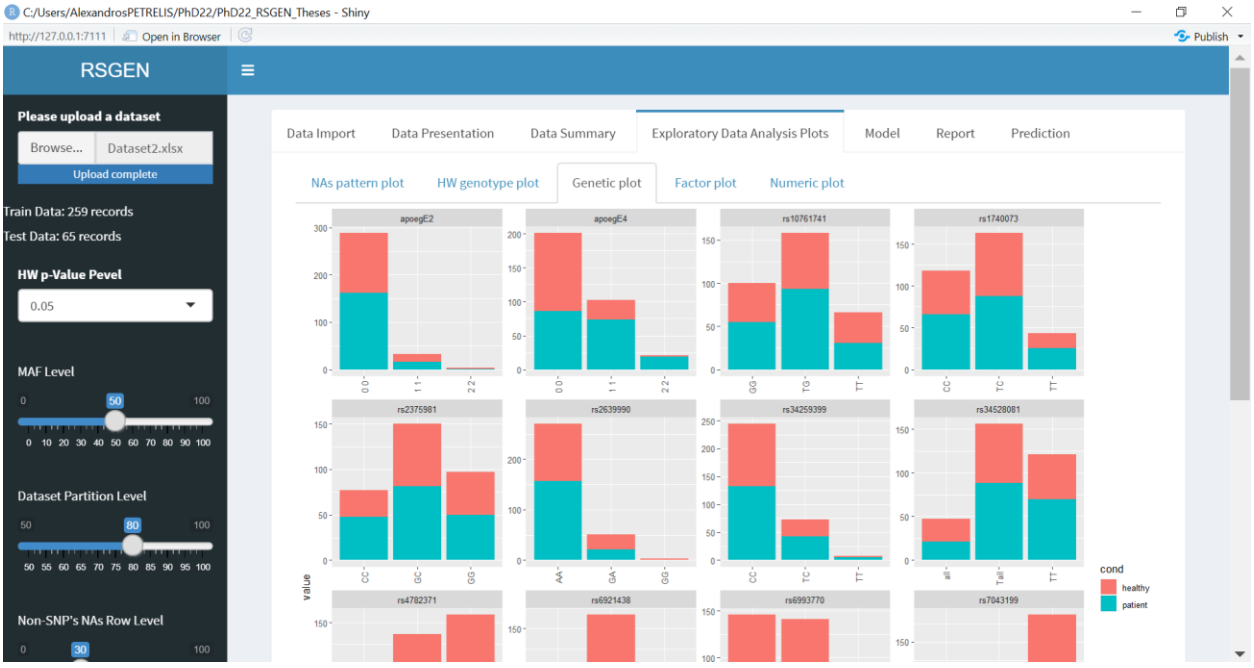

Figure 23. The histograms of the genetic data of the sixth dashboard of the analytical interface

The seventh dashboard presents the descriptive characteristics of continues metrics (Figure 24).

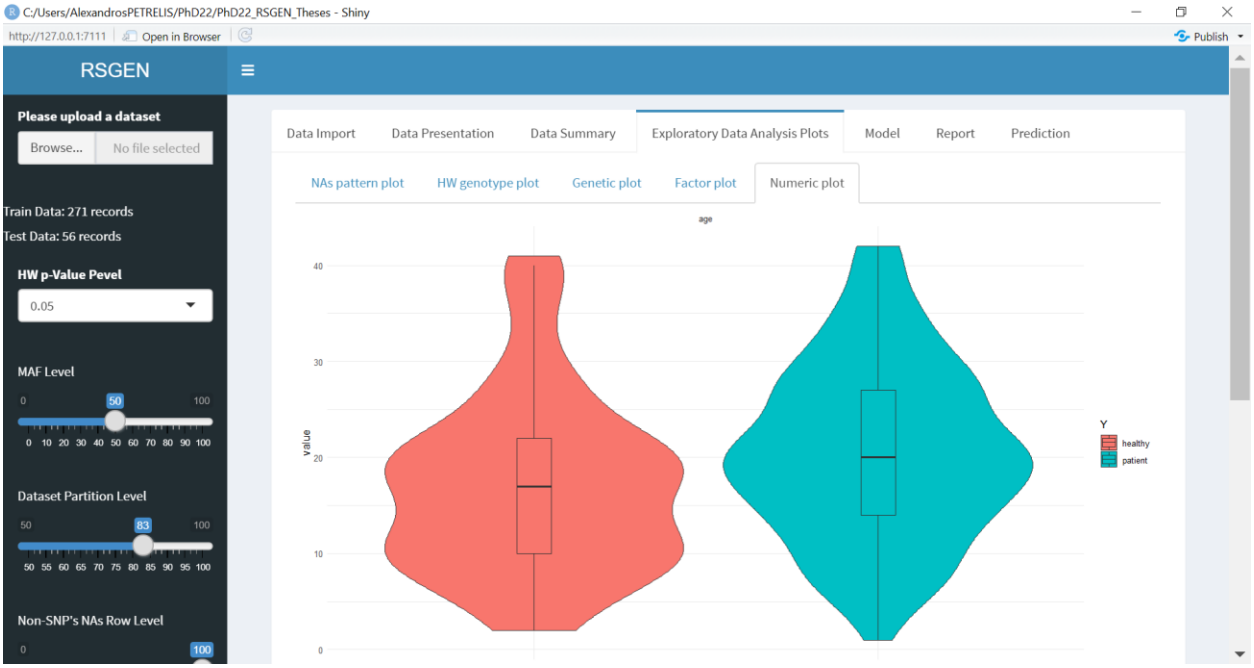

Figure 24. The violin plots of the numeric data of the seventh dashboard of the analytical interface

The eighth dashboard presents the descriptive characteristics of categorical metrics (Figure 25).

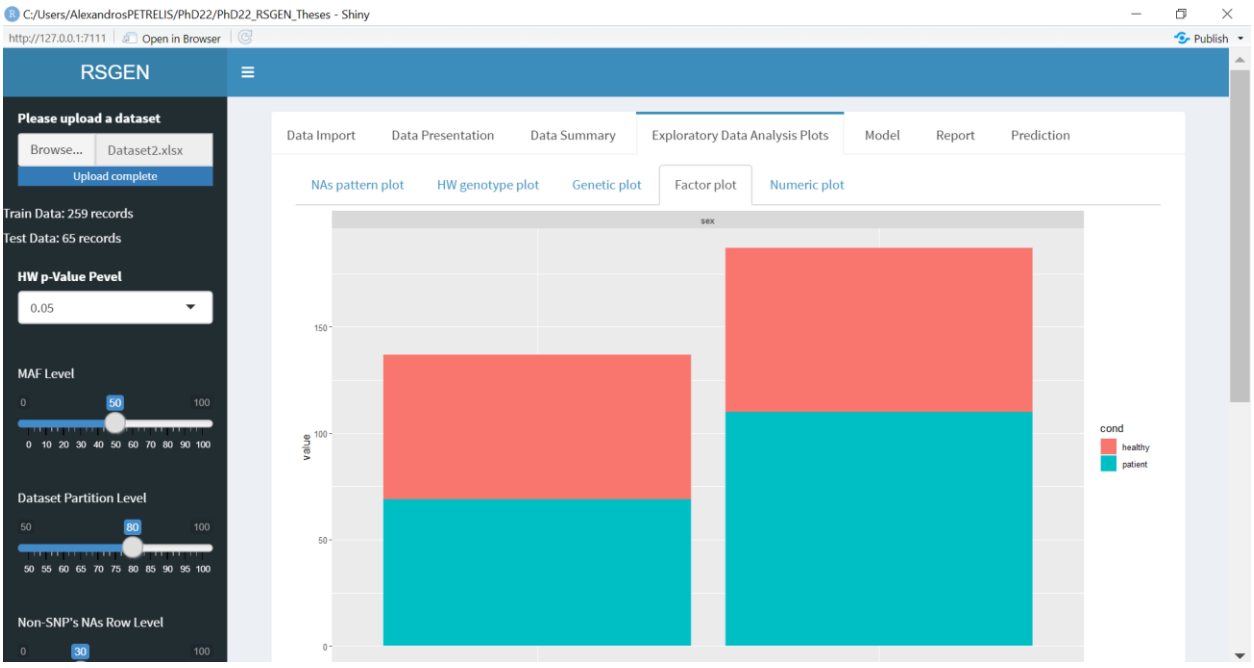

Figure 25. The histograms of the categorical data of the eighth dashboard of the analytical interface

The ninth dashboard presents a table of the coefficients from the predictive model (Figure 26).

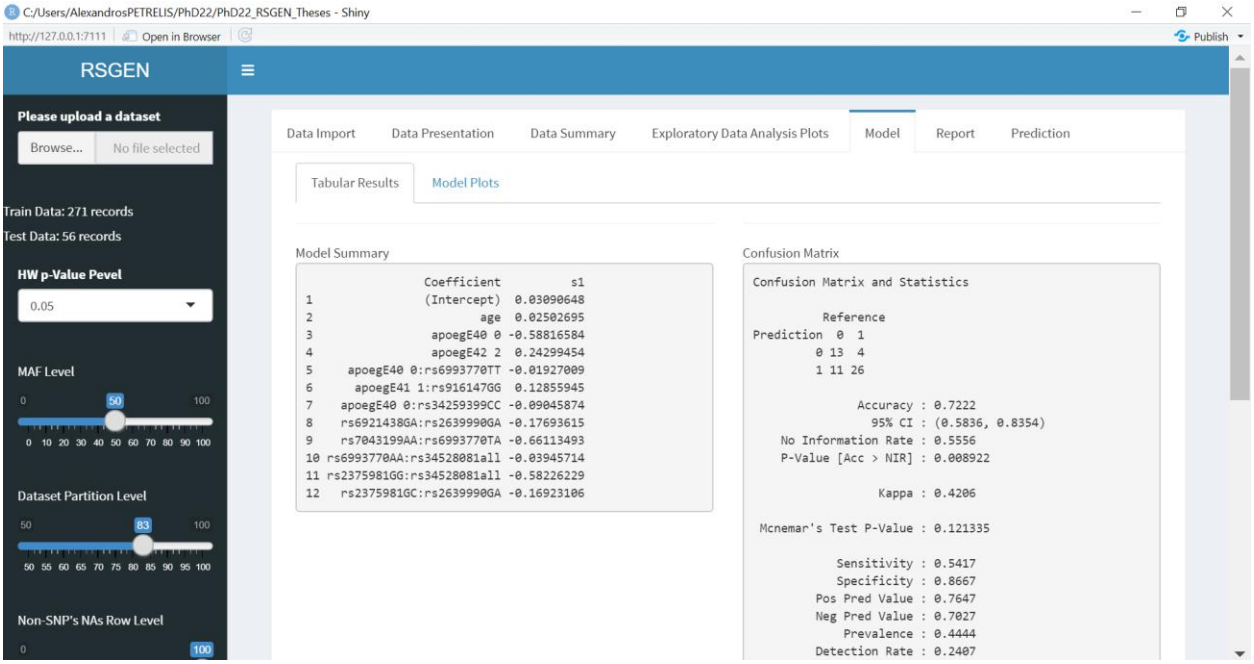

Figure 26. The table of coefficients from the predictive model of the ninth dashboard of the analytical interface

The tenth dashboard presents a graphical visualisation of the results of the predictive model (Figure 27).

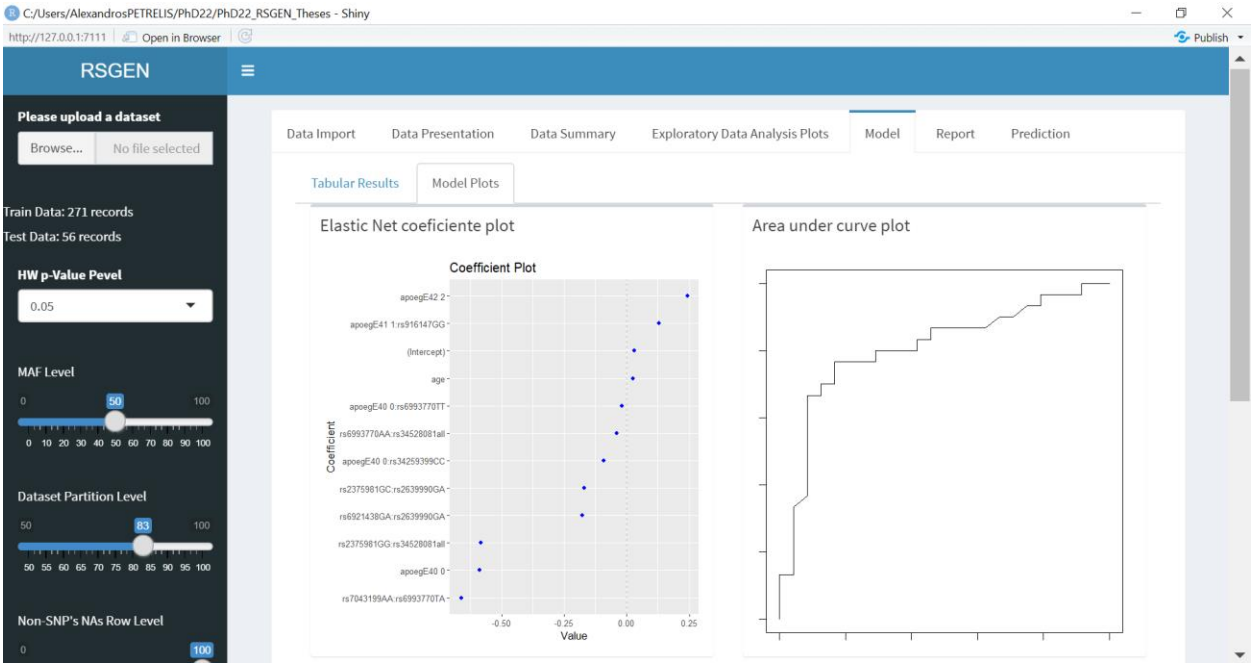

Figure 27. The table of results and the characteristics of the predictive model of the tenth dashboard of the analytical interface

The eleventh dashboard presents the generation report of results from all dashboards. They will be automatically extracted and stored in a doc format on the folder where the scripts have been stored (Figure 28)

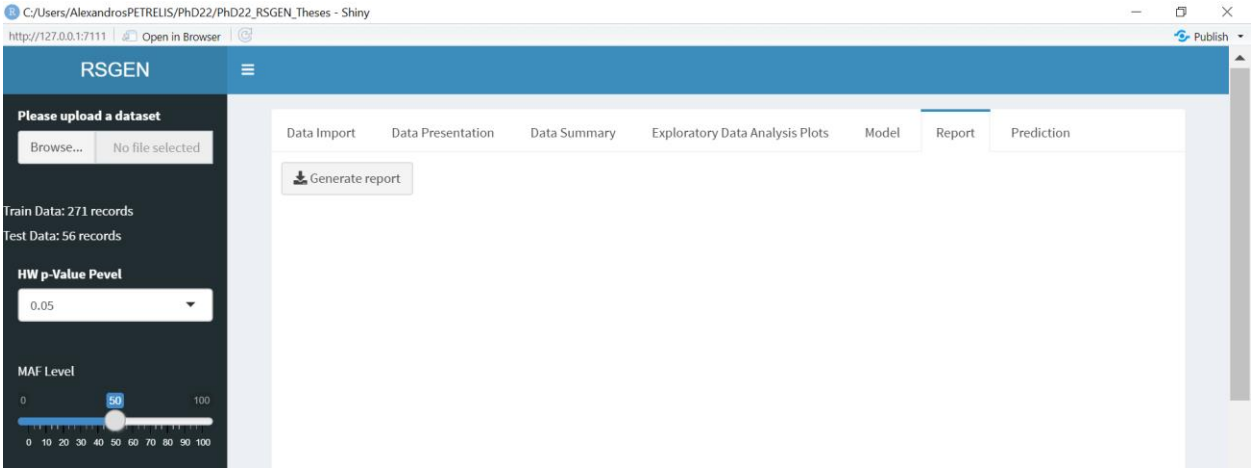

Figure 28. Reporting generator of the eleventh dashboard of the analytical interface

The twelfth dashboard presents an application for the prediction of novel individuals based on the coefficients that have been found from the prediction model. An input file with the requested data of the new population will be uploaded and the predictions are immediately presented (Figure 29)

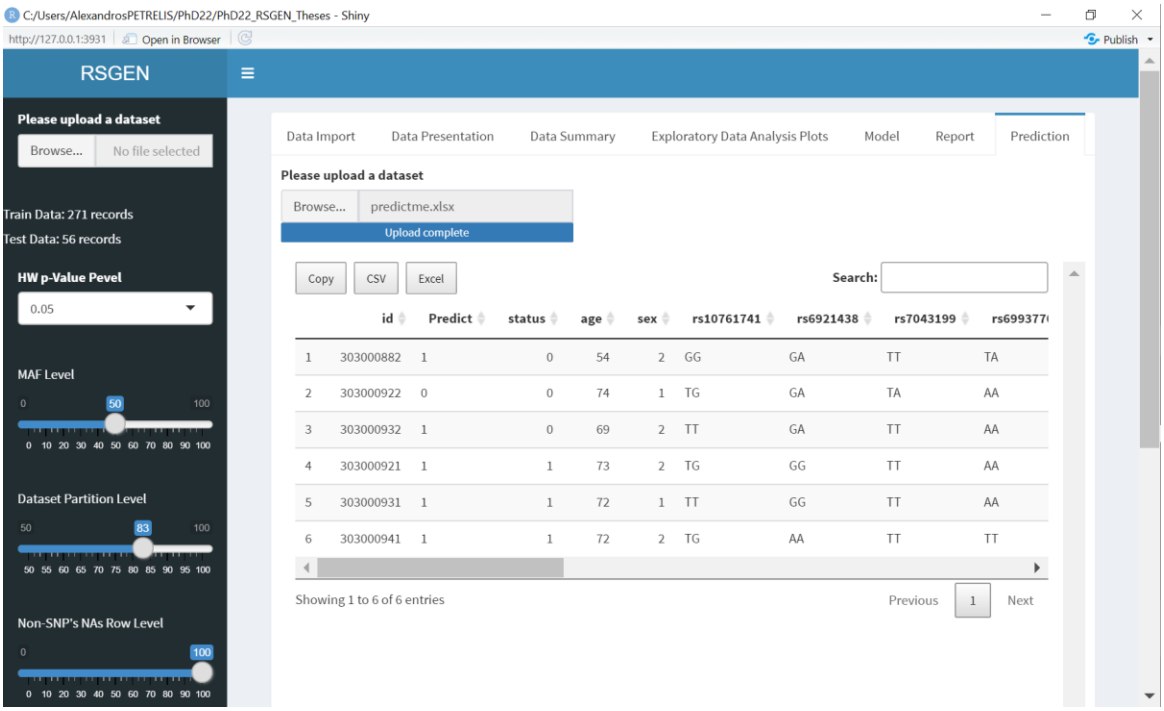

Figure 29. Prediction application of the twelfth dashboard of the analytical interface

The visualisation interface PMI consists of different dashboards that have been created to provide an overview of the data of the populations, the description of the populations' characteristics and of the genotyped variants and to perform initial analytical steps including control of the normality of the distributions of the variables and correlation analyses.

The detailed scripts are presented in the **annex1**.
The interface is presented in Figures 30-34

The first dashboard presents some main descriptive characteristics projected on a world map. (Figure 30). The interface can be used for more datasets.

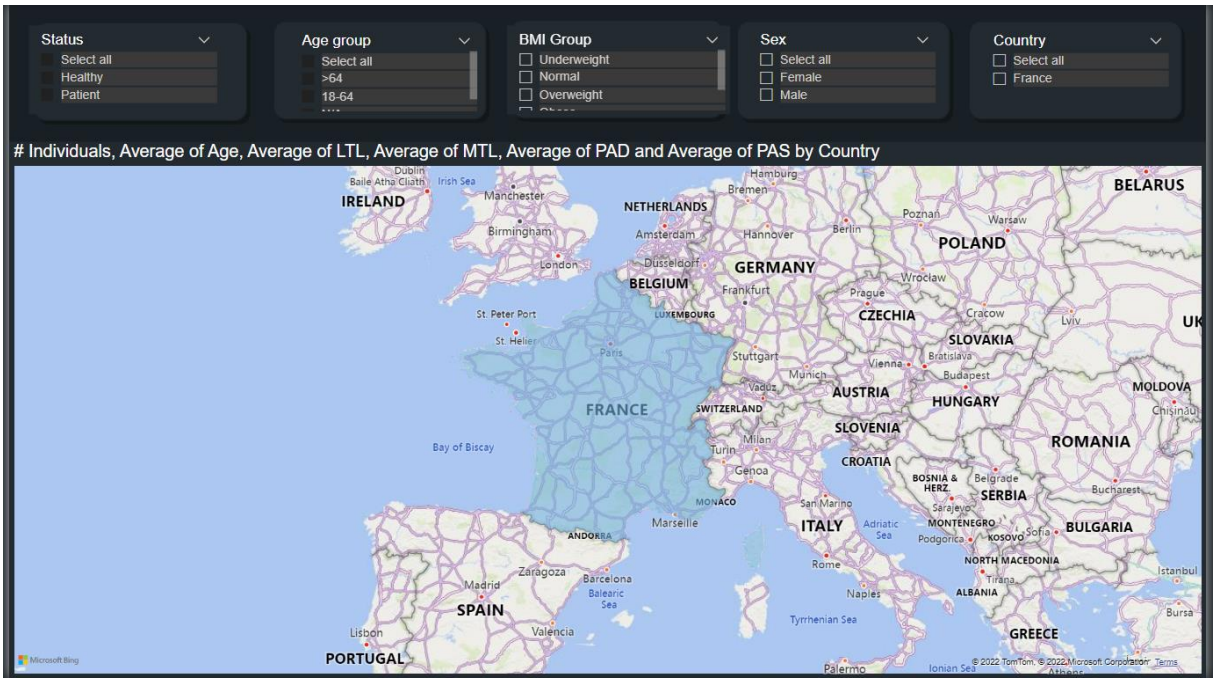

Figure 30. The first dashboard of the PMI descriptive interface with the map of available data

The second dashboard presents the uploaded databases and allows the selection of each of them (Figure 31).

Figure 31. The second dashboard of the PMI descriptive interface with the selection of databases

The next dashboard opens after the selection of a database and includes a selection module at the left of the screen and the visualisation module with the relative plots at the right of the screen. The selection modules include the description of the populations and the cases-controls (Figure 32) separately for the cases-controls populations. Then, there is a normality test module for all continuous (numeric) variables and modules for the key variables (Figure 33), the descriptive characteristics of the available SNPs (Figure 34) and finally a correlation analyses module for all key variables (Figure 35).

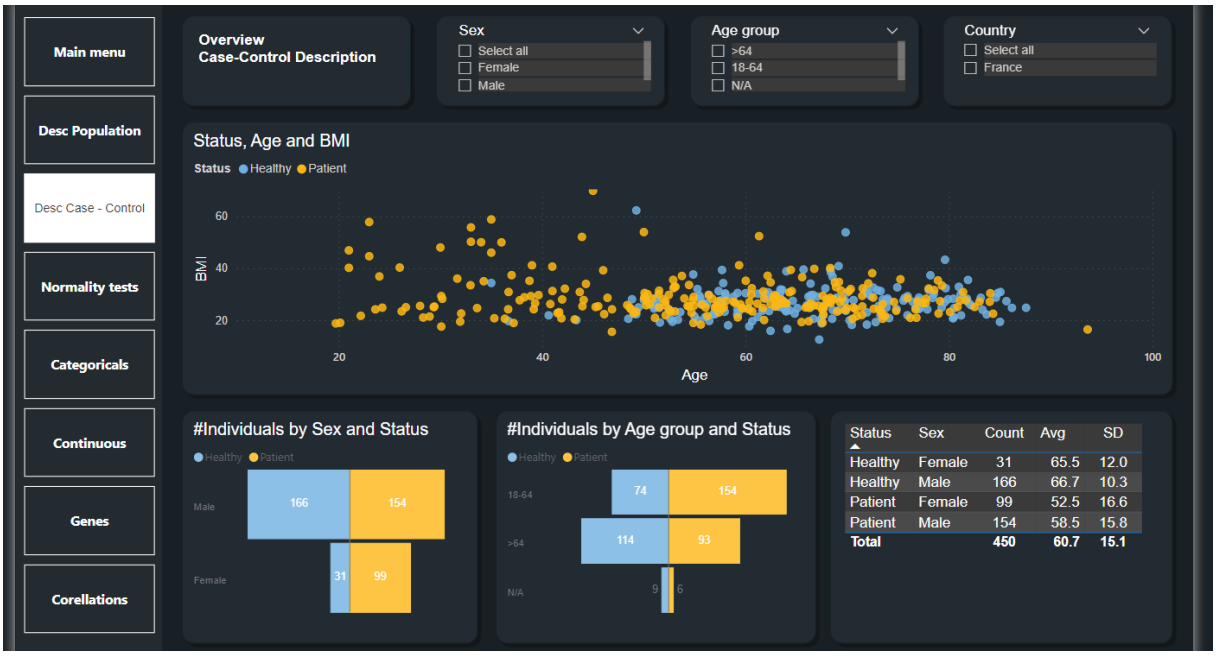

Figure 32. The third dashboard of the PMI descriptive interface with description of cases and controls

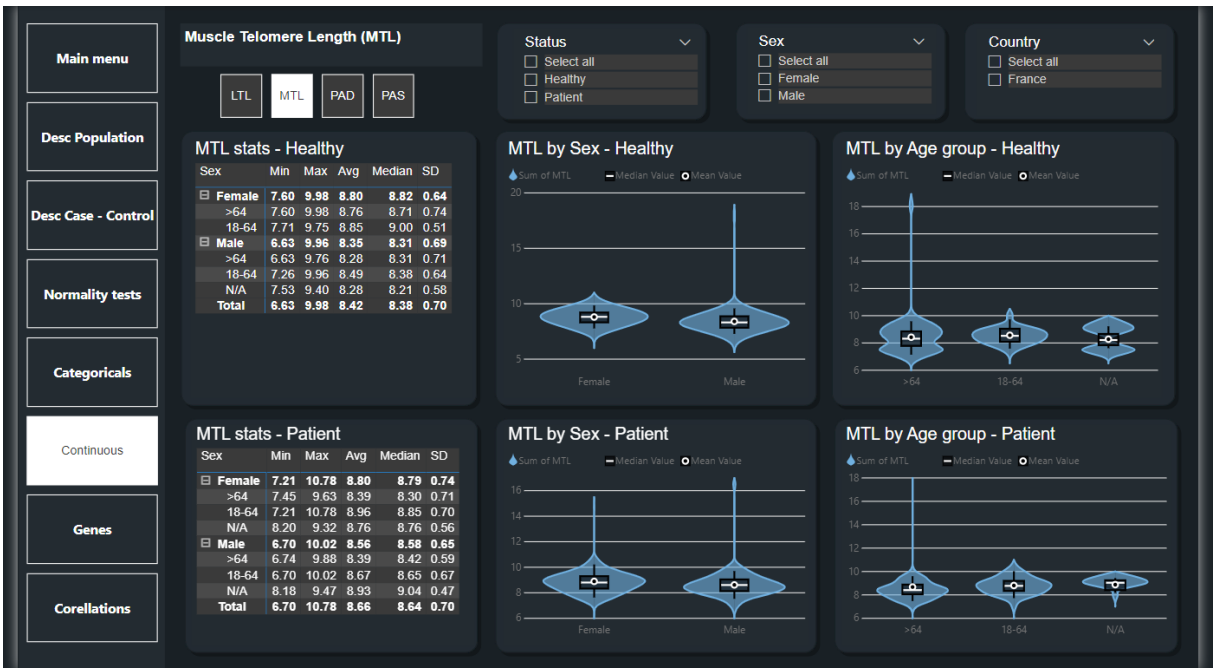

Figure 33. The forth dashboard of the PMI descriptive interface with the selection of a continuous variable

Figure 34. The fifth dashboard of the PMI descriptive interface with the descritpive characteristics of SNPs

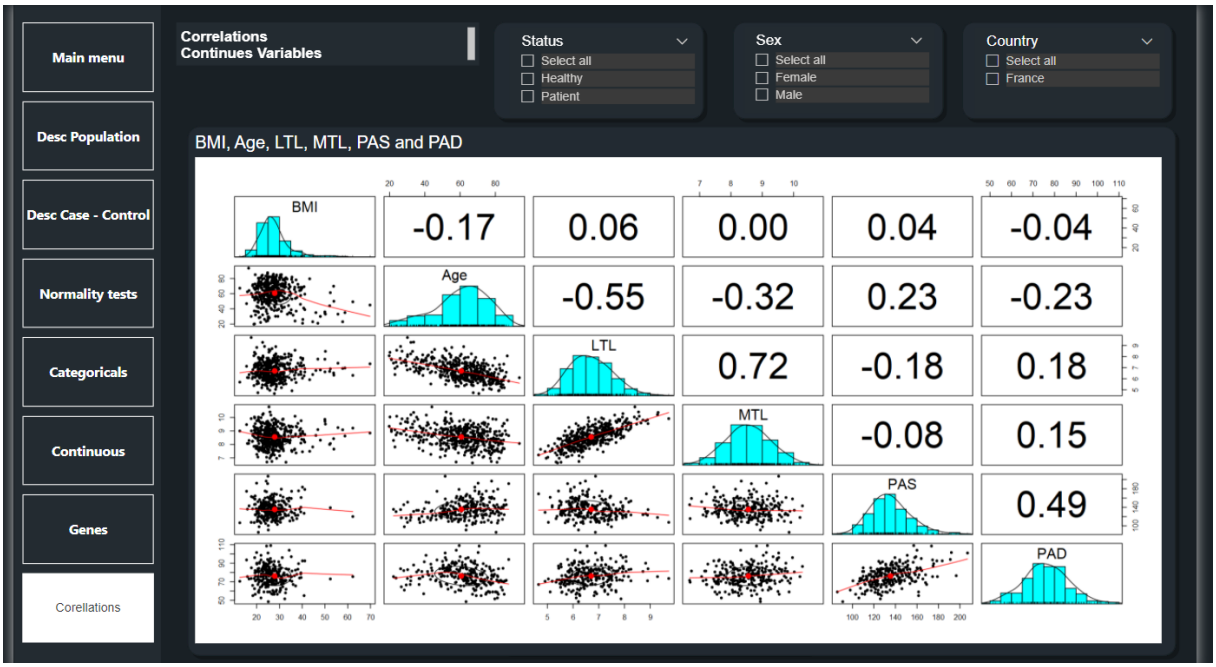

Figure 35. The sixth dashboard of the PMI descriptive interface with the correlation analyses

Conclusion

Two user-friendly tools have been developed for the construction of diseases risk prediction models using ML methodology and the data visualisation of large populations. These tools are not necessarily addressed to specialised analysts but to all researchers of different fields and clinicians with limited analytical skills due to their comprehensive automation. They are expected to be used for the application of personalisation in diseases risk prediction and the promotion of PM.

General discussion, Conclusion and Perspectives

General discussion

In this thesis we have focused on two key biomarkers, VEGF-A and TL, that are implicated in many chronic diseases, including CVD and AD and we aimed to further study their involvement in variability of intermediate phenotypes and conditions that lead to these pathologies. We have identified novel determinants of VEGF-A and TL, revealed links between these molecules and conditions like inflammation and telomere length attrition that are involved in both CVD and AD. Furthermore, we demonstrate their value in the disease risk prediction. We have used different approaches starting from candidate genes to investigate direct effects and epistatic interactions between polymorphisms and machine learning methodology to develop a personalised prediction algorithm for AD based on genetic variants. Our results are proposing new mechanisms that connect these two biomarkers with chronic diseases and provide original tools for the personalisation of risk prediction. More specifically we have demonstrated:

1. A review of the current knowledge on telomere length dynamics in childhood, where we propose that TL should be used as an index of health regularly measured throughout life starting from childhood.
2. Associations between VEGF-A levels and levels of IL-4, MCP-1 and EGF in PBMCs extracts and also associations between VEGF-A_{165} mRNA levels with MCP-1 and IL-1α and between VEGF-A_{189} mRNA levels and IL-4 and IL-6 in healthy condition.
3. Associations between LTL-related genetic variants with telomere attrition in a case-control population with atherosclerosis
4. Associations between the VEGF-A-related variant rs6993770 with telomere attrition in the same population
5. Epistatic interactions between VEGF-A-related variants, *APOE* and *LSR* polymorphisms and AD in a case-control population, especially the protective predictive interactions rs7043199*rs6993770 and rs2375981*rs34528081 and a genetic risk prediction model with a satisfactory accuracy
6. Development of an algorithm for the standardisation and automation of the analyses for the development of risk prediction models
7. Development of a user-friendly application for the visualisation of data of the populations and preliminary analyses

A synthetic representation of all our results is shown in Figure of synthesis below.

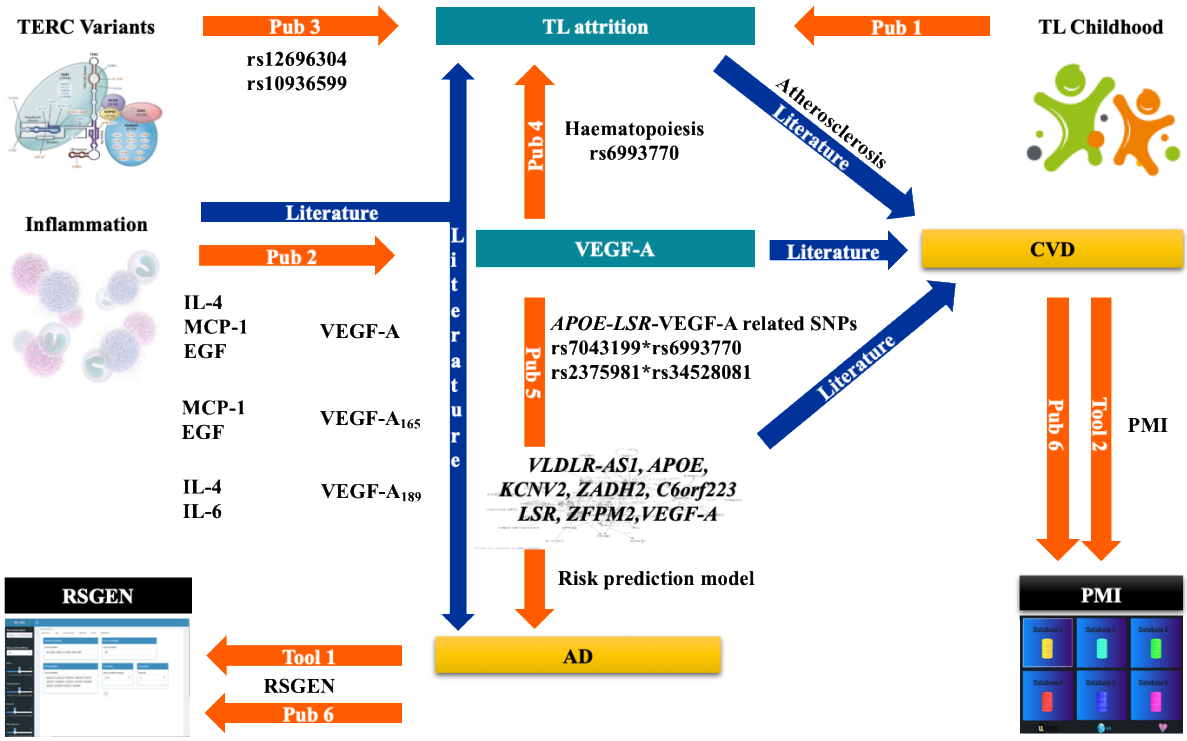

Figure of synthesis. Synthetic representation of results

More precisely, in our review publication (publication 1) we have presented the dynamics of TL in childhood and proposed TL as an index of health. In the first original publication (publication 2), we have shown specific links between inflammation, a condition known to be involved within CVD [72, 73] and AD [74] development, VEGF-A and with telomere attrition [32]. We have used protein levels and gene expression isoforms of VEGF-A in PBMCs, a type of cells that secrete inflammation molecules and play an essential role in chronic inflammation. The levels of anti-inflammatory IL-4 and the growth molecules MCP-1 and EGF are associated with increased VEGF-A levels, while the expression of the most abundant pro-angiogenic VEGF-A mRNA isoform, the VEGF-$A_{165,}$ is downregulated by MCP-1 and EGF. The expression of the VEGF-$A_{189,}$ is downregulated by anti-inflammatory IL-4 and pro-inflammatory IL-6. Thus, we have identified specific inflammation molecules as determinants of VEGF-A. In the second original publication (publication 3), we have shown that two *TERC* variants, a component of telomerase known to

decrease LTL, are also determinants of higher telomere attrition, which is linked with atherosclerosis [75] and CVD, but also with AD [76, 77]. Another determinant of telomere attrition identified in our third original publication (publication 4) is the variant rs6993770, which is a determinant of VEGF-A levels. This variant is associated with decreased telomere attrition, probably through modification on haematopoiesis, and could therefore be a protective factor for atherosclerosis and CVD but also for AD as found in our last original publication (publication 5). In this last work, we demonstrated several epistatic interactions between VEGF-A-related variants and the *APOE* and *LSR* genes, known candidate genes for both CVD and AD, that act as protective or risk determinants of AD. Among them the strongest interaction is the rs6993770* rs7043199, which has a protective role against AD even higher than that of the absence of *ε4 APOE* allele and the rs2375981*rs34528081 with an effect similar to the absence of *ε4 APOE* allele. It is noteworthy that the EA_1122; IGE-PCV was among the first to identify the allele 4 of the *APOE* gene as the strongest genetic risk factor for AD [78, 79] and currently we are proposing the rs7043199*rs6993770 and the rs2375981*rs34528081 interactions as strong protective factors against AD. This first interaction involves the rs6993770, which we previously showed that it is also a determinant for both VEGF-A and telomere attrition, thus is a key variant in our investigations linking these two assessed biomarkers with the two chronic diseases of interest. ApoE and LSR are involved in lipid metabolism and thus connected with CVD, but also with AD [80] and our results further support this connexion. In the same publication, the genes of the implicated variants in the identified prediction model are commonly found in diseases like CVD, neurological diseases and cancer. They participate in one network along with other molecules and modulators that deserve further investigation for their use as potential determinants of CVD and AD risk. Furthermore, an accurate risk prediction model for AD has been proposed that includes mainly genetic variants and their interactions and is thus a personalised tool, developed though machine learning methodology. An algorithm has been developed for the automation and the standardisation of the required analyses for the development of risk prediction tools by EN machine learning methodology through an automated interface called RSGEN (publication 6). Finally, a user-friendly application named PMI has been developed in Power BI for the visualisation of the populations' characteristics and available data and the realisation of preliminary analyses (publication 6).

Conclusion

In conclusion, this thesis is proposing several determinants of VEGF-A and TL that are linked through known and novel mechanisms with CVD and AD: IL-4, IL-6, MCP-1, EGF, IL-1α, *TERG* variants, VEGF-A-related variants, especially the rs6993770, *APOE* and *LSR* polymorphisms. Some of these have been successfully used to develop a risk prediction model for AD with a satisfactory accuracy. This is an original model that includes mainly genetic variants and their interactions to predict the risk for AD, thus it is a highly personalised tool. It is also highlighting the importance of the study of epistasis between genetic variants, that could better capture the genetic variability of the multifactorial diseases and provide a powerful methodology for diseases risk prediction. It is developed through a machine-learning methodology, which is very powerful for the development of risk prediction models, has less methodological restrictions compared to classical statistical methods and is currently the methodology of choice in PM. Furthermore, we have developed two informatic and statistical tools that served this thesis: an application for descriptive analysis of populations' data and an algorithm for the EN modelling. These tools can be used in other populations so as to enable visualisation and analysis of similar data, facilitate the validation and replication of our results and contribute in the development of robust personalised models for chronic diseases prediction.

Perspectives

The most important next step is the validation of our results in larger populations and then combine results in order to develop extended prediction risk tools in populations where availability of the identified determinants and related data is assured. These personalised risk prediction tools should be tested in large case-control and prospective cohorts of CVD and AD patients and compared with existing prediction tools to demonstrate their added value in utility and cost-effectiveness. They should also be tested in populations of different origin so as to generalise their use.

These resources could be found within the frames of the VEGF Consortium (www.santoriniconference.org), which was funded in 2015 and coordinated by the EA_1122 ; IGE-PCV. It includes 20 international partners with interests on VEGF-A and PM and access to important populations with large availability of "-omics" data [19]. Furthermore, significant

resources can be also found within collaborations of EA_1122; IGE-PCV within the frames of the Santorini Conference (www.santoriniconference.org), which is organised by the team [81].

A significant contribution of this thesis is the development of an application for the visualisation and descriptive analysis of populations' data and the algorithm for the elaboration of the risk prediction model. The application can be easily fed with the new populations and can allow a fast overview of the specific characteristics of each cohort based on descriptive analysis and related plots and graphs and selection of individuals based on precise criteria. Fast quality control and exclusion of individuals and variants based on exclusion criteria is also performed through the application and the selected populations can be extracted to datasets. These datasets can then be used for the application of the analysis algorithm. This algorithm is applying the EN method to provide the most accurate prediction model based on the determinants of each population. Thus, these two tools together with the availability of large populations that can be provided by the scientific network of EA_1122; IGE-PCV can ensure that the challenging perspectives of this work can be realised and that validated and replicated personalised risk prediction models based on the identified risk determinants will be produced and delivered for use to health professionals.

References

1. Armanios M and Blackburn EH. The telomere syndromes. Nat Rev Genet. 2012; 13(10):693-704.
2. Calado RT and Young NS. Telomere diseases. N Engl J Med. 2009; 361(24):2353-2365.
3. Blackburn EH, Epel ES and Lin J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. Science (New York, NY). 2015; 350(6265):1193-1198.
4. Muezzinler A, Zaineddin AK and Brenner H. A systematic review of leukocyte telomere length and age in adults. Ageing Res Rev. 2013; 12(2):509-519.
5. Broer L, Codd V, Nyholt DR, Deelen J, Mangino M, Willemsen G, Albrecht E, Amin N, Beekman M, de Geus EJ, Henders A, Nelson CP, Steves CJ, et al. Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. Eur J Hum Genet. 2013; 21(10):1163-1168.
6. Carmeliet P and Jain RK. Molecular mechanisms and clinical applications of angiogenesis. Nature. 2011; 473(7347):298-307.
7. Ferrara N. VEGF-A: a critical regulator of blood vessel growth. European cytokine network. 2009; 20(4):158-163.
8. Ribatti D, Levi-Schaffer F and Kovanen PT. Inflammatory angiogenesis in atherogenesis--a double-edged sword. Ann Med. 2008; 40(8):606-621.
9. Azimi-Nezhad M, Stathopoulou MG, Bonnefond A, Rancier M, Saleh A, Lamont J, Fitzgerald P, Ndiaye NC and Visvikis-Siest S. Associations of vascular endothelial growth factor (VEGF) with adhesion and inflammation molecules in a healthy population. Cytokine. 2013; 61(2):602-607.
10. Bonnefond A, Saulnier PJ, Stathopoulou MG, Grarup N, Ndiaye NC, Roussel R, Nezhad MA, Dechaume A, Lantieri O, Hercberg S, Lauritzen T, Balkau B, El-Sayed Moustafa JS, 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(2):e55921.
11. Saleh A, Stathopoulou MG, Dade S, Ndiaye NC, Azimi-Nezhad M, Murray H, Masson C, Lamont J, Fitzgerald P and Visvikis-Siest S. Angiogenesis related genes NOS3, CD14, MMP3 and IL4R are associated to VEGF gene expression and circulating levels in healthy adults. BMC medical genetics. 2015; 16:90.
12. Stathopoulou MG, Bonnefond A, Ndiaye NC, Azimi-Nezhad M, El Shamieh S, Saleh A, Rancier M, Siest G, Lamont J, Fitzgerald P and Visvikis-Siest S. 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):535-541.
13. Berrahmoune H, Herbeth B, Lamont JV, Masson C, Fitzgerald PS and Visvikis-Siest S. Heritability for plasma VEGF concentration in the Stanislas family study. Annals of human genetics. 2007; 71(Pt 1):54-63.
14. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, Safa R, Azimi Nezhad M, Sawyer D, Marteau JB, Xanthakis V, Siest G, Sullivan L, 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):554-563.
15. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, Concas MP, Traglia M, Barbieri C, Ndiaye NC, Stathopoulou MG, Lagou V, Maestrale GB, 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.
16. Xie T, Stathopoulou MG, de Andres F, Siest G, Murray H, Martin M, Cobaleda J, Delgado A, Lamont J, Penas LE, A LL and Visvikis-Siest S. VEGF-related polymorphisms identified by GWAS and risk for major depression. Transl Psychiatry. 2017; 7(3):e1055.
17. Rancier M, Zaaber I, Stathopoulou MG, Saleh A, Marmouch H, Masson C, Murray H, Kurth MJ, Lamont J, Fitzgerald P, Mahjoub S, Said K, Bel Hadj Jrad Tensaout B, et al. Pro- and anti-angiogenic VEGF mRNAs in autoimmune thyroid diseases. Autoimmunity. 2016; In press.

18. Zaaber I, Rancier M, Stathopoulou MG, Saleh A, Marmouch H, Masson C, Murray H, Kurth MJ, Lamont J, Fitzgerald P, Mahjoub S, Said K, Bel Hadj Jrad Tensaout B, et al. Plasma VEGF-related polymorphisms are implied in autoimmune thyroid diseases. Autoimmunity. 2016; 49(4):229-235.
19. Stathopoulou MG, Xie T, Ruggiero D, Chatelin J, Rancier M, Weryha G, Kurth MJ, Aldasoro Arguinano AA, Gorenjak V, Petrelis AM, Dagher G, Dedoussis G, Deloukas P, et al. 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. 2018; 56(4):83-86.
20. Beazley-Long N, Hua J, Jehle T, Hulse RP, Dersch R, Lehrling C, Bevan H, Qiu Y, Lagrèze WA, Wynick D, Churchill AJ, Kehoe P, Harper SJ, et al. VEGF-A165b is an endogenous neuroprotective splice isoform of vascular endothelial growth factor A in vivo and in vitro. Am J Pathol. 2013; 183(3):918-929.
21. Hohman TJ, Bell SP and Jefferson AL. The role of vascular endothelial growth factor in neurodegeneration and cognitive decline: exploring interactions with biomarkers of Alzheimer disease. JAMA Neurol. 2015; 72(5):520-529.
22. Huang L, Jia J and Liu R. Decreased serum levels of the angiogenic factors VEGF and TGF-β1 in Alzheimer's disease and amnestic mild cognitive impairment. Neurosci Lett. 2013; 550:60-63.
23. Tarkowski E, Issa R, Sjögren M, Wallin A, Blennow K, Tarkowski A and Kumar P. Increased intrathecal levels of the angiogenic factors VEGF and TGF-beta in Alzheimer's disease and vascular dementia. Neurobiol Aging. 2002; 23(2):237-243.
24. World Health Organisation W. (2020). Fact sheets on dementia.
25. Siest G, Bertrand P, Qin B, Herbeth B, Serot JM, Masana L, Ribalta J, Passmore AP, Evans A, Ferrari M, Franceschi M, Shepherd J, Cuchel M, et al. Apolipoprotein E polymorphism and serum concentration in Alzheimer's disease in nine European centres: the ApoEurope study. ApoEurope group. Clin Chem Lab Med. 2000; 38(8):721-730.
26. Liu S, Liu J, Weng R, Gu X and Zhong Z. Apolipoprotein E gene polymorphism and the risk of cardiovascular disease and type 2 diabetes. BMC Cardiovasc Disord. 2019; 19(1):213.
27. Geisel J, Bunte T, Bodis M, Oette K and Herrmann W. Apolipoprotein E2/E2 genotype in combination with mutations in the LDL receptor gene causes type III hyperlipoproteinemia. Clin Chem Lab Med. 2002; 40(5):475-479.
28. Xie T, Akbar S, Stathopoulou MG, Oster T, Masson C, Yen FT and Visvikis-Siest S. Epistatic interaction of apolipoprotein E and lipolysis-stimulated lipoprotein receptor genetic variants is associated with Alzheimer's disease. Neurobiol Aging. 2018; 69:292 e291-292 e295.
29. Fossel M. Use of Telomere Length as a Biomarker for Aging and Age-Related Disease. Current Translational Geriatrics and Experimental Gerontology Reports. 2012; 1(2):121-127.
30. Braile M, Marcella S, Cristinziano L, Galdiero MR, Modestino L, Ferrara AL, Varricchi G, Marone G and Loffredo S. VEGF-A in Cardiomyocytes and Heart Diseases. Int J Mol Sci. 2020; 21(15).
31. Chiappelli M, Borroni B, Archetti S, Calabrese E, Corsi MM, Franceschi M, Padovani A and Licastro F. VEGF gene and phenotype relation with Alzheimer's disease and mild cognitive impairment. Rejuvenation Res. 2006; 9(4):485-493.
32. Gorenjak V, Petrelis AM, Stathopoulou MG and Visvikis-Siest S. Telomere length determinants in childhood. Clin Chem Lab Med. 2020; 58(2):162-177.
33. Entringer S, de Punder K, Buss C and Wadhwa PD. The fetal programming of telomere biology hypothesis: an update. Proc R Soc Lond B Biol Sci. 2018; 373(1741):20170151.
34. McCloskey K, Ponsonby AL, Collier F, Allen K, Tang MLK, Carlin JB, Saffery R, Skilton MR, Cheung M, Ranganathan S, Dwyer T, Burgner D and Vuillermin P. The association between higher maternal pre-pregnancy body mass index and increased birth weight, adiposity and inflammation in the newborn. Pediatric obesity. 2018; 13(1):46-53.

35. Salihu HM, Pradhan A, King L, Paothong A, Nwoga C, Marty PJ and Whiteman V. Impact of intrauterine tobacco exposure on fetal telomere length. American journal of obstetrics and gynecology. 2015; 212(2):205.e201-208.
36. Send TS, Gilles M, Codd V, Wolf I, Bardtke S, Streit F, Strohmaier J, Frank J, Schendel D, Sutterlin MW, Denniff M, Laucht M, Samani NJ, et al. Telomere Length in Newborns is Related to Maternal Stress During Pregnancy. Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology. 2017; 42(12):2407-2413.
37. Shalev I, Entringer S, Wadhwa PD, Wolkowitz OM, Puterman E, Lin J and Epel ES. Stress and telomere biology: a lifespan perspective. Psychoneuroendocrinology. 2013; 38(9):1835-1842.
38. Stathopoulou MG, Petrelis AM, Buxton JL, Froguel P, Blakemore AI and Visvikis-Siest S. Genetic determinants of leucocyte telomere length in children: a neglected and challenging field. Paediatr Perinat Epidemiol. 2015; 29(2):146-150.
39. Corkum CP, Ings DP, Burgess C, Karwowska S, Kroll W and 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 Immunology. 2015; 16.
40. Turner MD, Nedjai B, Hurst T and Pennington DJ. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2014; 1843(11):2563-2582.
41. Institute NHGR. (2021). GWAS Catalog. In: Institute NHGR, ed.
42. Benetos A, Toupance S, Gautier S, Labat C, Kimura M, Rossi PM, Settembre N, Hubert J, Frimat L and Bertrand B. Short leukocyte telomere length precedes clinical expression of atherosclerosis: the blood-and-muscle model. Circulation research. 2018; 122(4):616-623.
43. Haycock PC, Burgess S, Nounu A, Zheng J, Okoli GN, Bowden J, Wade KH, Timpson NJ, Evans DM, Willeit P, Aviv A, Gaunt TR, Hemani G, et al. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. JAMA Oncol. 2017; 3(5):636-651.
44. Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, Harley CB and Aviv A. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. Nature protocols. 2010; 5(9):1596-1607.
45. Honig LS, Kang MS, Cheng R, Eckfeldt JH, Thyagarajan B, Leiendecker-Foster C, Province MA, Sanders JL, Perls T, Christensen K, Lee JH, Mayeux R and Schupf N. Heritability of telomere length in a study of long-lived families. Neurobiology of aging. 2015; 36(10):2785-2790.
46. Hetland ML, Christensen IJ, Lottenburger T, Johansen JS, Svendsen MN, Horslev-Petersen K, Nielsen L and Nielsen HJ. Circulating VEGF as a biological marker in patients with rheumatoid arthritis? Preanalytical and biological variability in healthy persons and in patients. Disease markers. 2008; 24(1):1-10.
47. Toupance S, Villemonais D, Germain D, Gegout-Petit A, Albuisson E and Benetos A. The individual's signature of telomere length distribution. Scientific Reports. 2019; 9(1):685.
48. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG, Gil L, Gordon L, Haggerty L, et al. Ensembl 2018. Nucleic acids research. 2018; 46(D1):D754-d761.
49. Rufer N, Brummendorf TH, Kolvraa S, Bischoff C, Christensen K, Wadsworth L, Schulzer M and Lansdorp PM. 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. The Journal of experimental medicine. 1999; 190(2):157-167.
50. Van Cauwenberghe C, Van Broeckhoven C and Sleegers K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med. 2016; 18(5):421-430.
51. Zou H and Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B (Stat Methodol). 2005; 67:301-320.

52. Bouts M, Möller C, Hafkemeijer A, van Swieten JC, Dopper E, van der Flier WM, Vrenken H, Wink AM, Pijnenburg YAL, Scheltens P, Barkhof F, Schouten TM, de Vos F, et al. Single Subject Classification of Alzheimer's Disease and Behavioral Variant Frontotemporal Dementia Using Anatomical, Diffusion Tensor, and Resting-State Functional Magnetic Resonance Imaging. J Alzheimers Dis. 2018; 62(4):1827-1839.
53. Liu Y, Thalamuthu A, Mather KA, Crawford J, Ulanova M, Wong MWK, Pickford R, Sachdev PS and Braidy N. Plasma lipidome is dysregulated in Alzheimer's disease and is associated with disease risk genes. Transl Psychiatry. 2021; 11(1):344.
54. de Vos F, Schouten TM, Hafkemeijer A, Dopper EG, van Swieten JC, de Rooij M, van der Grond J and Rombouts SA. Combining multiple anatomical MRI measures improves Alzheimer's disease classification. Hum Brain Mapp. 2016; 37(5):1920-1929.
55. Xie T, Akbar S, Stathopoulou MG, Oster T, Masson C, Yen FT and Visvikis-Siest S. Epistatic interaction of apolipoprotein E and lipolysis-stimulated lipoprotein receptor genetic variants is associated with Alzheimer's disease. Neurobiol Aging. 2018; 69:292.e291-292.e295.
56. Danso SO, Zeng Z, Muniz-Terrera G and Ritchie CW. Developing an Explainable Machine Learning-Based Personalised Dementia Risk Prediction Model: A Transfer Learning Approach With Ensemble Learning Algorithms. Front Big Data. 2021; 4:613047.
57. Qiu S, Joshi PS, Miller MI, Xue C, Zhou X, Karjadi C, Chang GH, Joshi AS, Dwyer B, Zhu S, Kaku M, Zhou Y, Alderazi YJ, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain. 2020; 143(6):1920-1933.
58. Wang L, Li P, Hou M, Zhang X, Cao X and Li H. Construction of a risk prediction model for Alzheimer's disease in the elderly population. BMC Neurol. 2021; 21(1):271.
59. Chouraki V, Reitz C, Maury F, Bis JC, Bellenguez C, Yu L, Jakobsdottir J, Mukherjee S, Adams HH, Choi SH, Larson EB, Fitzpatrick A, Uitterlinden AG, et al. Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. J Alzheimers Dis. 2016; 53(3):921-932.
60. Hou XH, Feng L, Zhang C, Cao XP, Tan L and Yu JT. Models for predicting risk of dementia: a systematic review. J Neurol Neurosurg Psychiatry. 2019; 90(4):373-379.
61. Khera AV, Chaffin M, Zekavat SM, Collins RL, Roselli C, Natarajan P, Lichtman JH, D'Onofrio G, Mattera J, Dreyer R, Spertus JA, Taylor KD, Psaty BM, et al. Whole-Genome Sequencing to Characterize Monogenic and Polygenic Contributions in Patients Hospitalized With Early-Onset Myocardial Infarction. Circulation. 2019; 139(13):1593-1602.
62. Pate A, Emsley R, Ashcroft DM, Brown B and van Staa T. The uncertainty with using risk prediction models for individual decision making: an exemplar cohort study examining the prediction of cardiovascular disease in English primary care. BMC Med. 2019; 17(1):134.
63. Hall WD, Morley KI and Lucke JC. The prediction of disease risk in genomic medicine. EMBO Rep. 2004; 5 Spec No(Suppl 1):S22-26.
64. Chapman BP, Lin F, Roy S, Benedict RHB and Lyness JM. Health risk prediction models incorporating personality data: Motivation, challenges, and illustration. Personal Disord. 2019; 10(1):46-58.
65. Schork NJ. Artificial Intelligence and Personalized Medicine. Cancer Treat Res. 2019; 178:265-283.
66. Ahmed Z, Mohamed K, Zeeshan S and Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford). 2020; 2020.
67. O'Donoghue SI, Baldi BF, Clark SJ, Darling AE, Hogan JM, Kaur S, Maier-Hein L, McCarthy DJ, Moore WJ, Stenau E, Swedlow JR, Vuong J and Procter JB. Visualization of Biomedical Data. Annual Review of Biomedical Data Science. 2018; 1(1):275-304.

68. Pavlopoulos GA, Malliarakis D, Papanikolaou N, Theodosiou T, Enright AJ and Iliopoulos I. Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future. Gigascience. 2015; 4:38.
69. Sarris K, Komianou A, Patrinos GP and Katsila T. Application of the DruGeVar Database in Cancer Genomics and Pharmacogenomics. Public Health Genomics. 2017; 20(2):142-147.
70. Sun L, Dong S, Ge Y, Fonseca JP, Robinson ZT, Mysore KS and Mehta P. DiVenn: An Interactive and Integrated Web-Based Visualization Tool for Comparing Gene Lists. Front Genet. 2019; 10:421.
71. Microsoft. (Assessed 12/2021). Power BI - basic concepts
72. Soysal P, Arik F, Smith L, Jackson SE and Isik AT. Inflammation, Frailty and Cardiovascular Disease. Adv Exp Med Biol. 2020; 1216:55-64.
73. Poznyak AV, Bezsonov EE, Popkova TV, Starodubova AV and Orekhov AN. Immunity in Atherosclerosis: Focusing on T and B Cells. Int J Mol Sci. 2021; 22(16).
74. de Oliveira J, Kucharska E, Garcez ML, Rodrigues MS, Quevedo J, Moreno-Gonzalez I and Budni J. Inflammatory Cascade in Alzheimer's Disease Pathogenesis: A Review of Experimental Findings. Cells. 2021; 10(10).
75. Benetos A, Toupance S, Gautier S, Labat C, Kimura M, Rossi PM, Settembre N, Hubert J, Frimat L, Bertrand B, Boufi M, Flecher X, Sadoul N, et al. Short Leukocyte Telomere Length Precedes Clinical Expression of Atherosclerosis: The Blood-and-Muscle Model. Circ Res. 2018; 122(4):616-623.
76. Rizvi S, Raza ST and Mahdi F. Telomere length variations in aging and age-related diseases. Curr Aging Sci. 2014; 7(3):161-167.
77. Thanseem I, Viswambharan V, Poovathinal SA and Anitha A. Is telomere length a biomarker of neurological disorders? Biomark Med. 2017; 11(9):799-810.
78. Schmechel DE, Saunders AM, Strittmatter WJ, Crain BJ, Hulette CM, Joo SH, Pericak-Vance MA, Goldgaber D and Roses AD. Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. Proc Natl Acad Sci U S A. 1993; 90(20):9649-9653.
79. Lucotte G, David F, Visvikis S, Leininger-Müller B, Siest G, Babron MC and Couderc R. Apolipoprotein E-epsilon 4 allele and Alzheimer's disease. Lancet. 1993; 342(8882):1309.
80. Mallick R, Basak S and Duttaroy AK. Fatty acids and evolving roles of their proteins in neurological, cardiovascular disorders and cancers. Prog Lipid Res. 2021; 83:101116.
81. Visvikis-Siest S, Gorenjak V, Stathopoulou MG, Petrelis AM, Weryha G, Masson C, Hiegel B, Kumar S, Barouki R, Boerwinkle E, Dagher G, Deloukas P, Innocenti F, et al. 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. 2018; 8(4).
82. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, Bobba S and Zwar N. A systematic review of chronic disease management interventions in primary care. BMC Family Practice. 2018; 19(1):11.
83. WHO. (Assessed 2021). Fact sheets: Noncommunicable diseases.
84. Raghupathi W and Raghupathi V. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach. Int J Environ Res Public Health. 2018; 15(3).
85. Bauer UE, Briss PA, Goodman RA and Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet. 2014; 384(9937):45-52.
86. Bernell S and Howard SW. Use Your Words Carefully: What Is a Chronic Disease? Front Public Health. 2016; 4:159.
87. Prados-Torres A, Poblador-Plou B, Gimeno-Miguel A, Calderón-Larrañaga A, Poncel-Falcó A, Gimeno-Feliú LA, González-Rubio F, Laguna-Berna C, Marta-Moreno J, Clerencia-Sierra M, Aza-Pascual-Salcedo M, Bandrés-Liso AC, Coscollar-Santaliestra C, et al. Cohort Profile: The Epidemiology of Chronic Diseases and Multimorbidity. The EpiChron Cohort Study. Int J Epidemiol. 2018; 47(2):382-384f.

88. Gruskin S, Ferguson L, Tarantola D and Beaglehole R. Noncommunicable diseases and human rights: a promising synergy. Am J Public Health. 2014; 104(5):773-775.
89. Yach D, Hawkes C, Gould CL and Hofman KJ. The global burden of chronic diseases: overcoming impediments to prevention and control. Jama. 2004; 291(21):2616-2622.
90. Abegunde DO, Mathers CD, Adam T, Ortegon M and Strong K. The burden and costs of chronic diseases in low-income and middle-income countries. Lancet. 2007; 370(9603):1929-1938.
91. Arredondo A and Aviles R. Costs and epidemiological changes of chronic diseases: implications and challenges for health systems. PLoS One. 2015; 10(3):e0118611.
92. Wolff JL, Starfield B and Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Arch Intern Med. 2002; 162(20):2269-2276.
93. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2009; 32 Suppl 1(Suppl 1):S62-67.
94. Dokken BB. The Pathophysiology of Cardiovascular Disease and Diabetes: Beyond Blood Pressure and Lipids. Diabetes Spectrum. 2008; 21(3):160-165.
95. Abbate R, Sticchi E and Fatini C. Genetics of cardiovascular disease. Clin Cases Miner Bone Metab. 2008; 5(1):63-66.
96. WHO. Fact sheets: Cardiovascular Diseases Assessed 2021.
97. WHO. (Assessed 2021). ICD-11 for Mortality and Morbidity Statistics (ICD-11 MMS)
98. Kathiresan S and Srivastava D. Genetics of human cardiovascular disease. Cell. 2012; 148(6):1242-1257.
99. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, Ahmed M, Aksut B, Alam T, Alam K, Alla F, Alvis-Guzman N, Amrock S, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol. 2017; 70(1):1-25.
100. Yan R, Li W, Yin L, Wang Y and Bo J. Cardiovascular Diseases and Risk-Factor Burden in Urban and Rural Communities in High-, Middle-, and Low-Income Regions of China: A Large Community-Based Epidemiological Study. J Am Heart Assoc. 2017; 6(2).
101. Karmali KN, Persell SD, Perel P, Lloyd-Jones DM, Berendsen MA and Huffman MD. Risk scoring for the primary prevention of cardiovascular disease. Cochrane Database Syst Rev. 2017; 3(3):Cd006887.
102. Olvera Lopez E, Ballard BD and Jan A. (2021). Cardiovascular Disease. StatPearls. (Treasure Island (FL): StatPearls Publishing
Copyright © 2021, StatPearls Publishing LLC.).
103. Arbour L, Asuri S, Whittome B, Polanco F and Hegele RA. The Genetics of Cardiovascular Disease in Canadian and International Aboriginal Populations. Can J Cardiol. 2015; 31(9):1094-1115.
104. Flora GD and Nayak MK. A Brief Review of Cardiovascular Diseases, Associated Risk Factors and Current Treatment Regimes. Curr Pharm Des. 2019; 25(38):4063-4084.
105. National Research Council Committee on D and Health. (1989). Diet and Health: Implications for Reducing Chronic Disease Risk. (Washington (DC): National Academies Press (US)
Copyright © 1989 by the National Academy of Sciences.).
106. Frostegård J. Immunity, atherosclerosis and cardiovascular disease. BMC Med. 2013; 11:117.
107. Falk E. Pathogenesis of atherosclerosis. J Am Coll Cardiol. 2006; 47(8 Suppl):C7-12.
108. Qiao W, Zhang X, Kan B, Vuong AM, Xue S, Zhang Y, Li B, Zhao Q, Guo D, Shen X and Yang S. Hypertension, BMI, and cardiovascular and cerebrovascular diseases. Open Med (Wars). 2021; 16(1):149-155.
109. World Heart Federation. (Assessed November 2021).
110. Lane CA, Hardy J and Schott JM. Alzheimer's disease. Eur J Neurol. 2018; 25(1):59-70.
111. Serý O, Povová J, Míšek I, Pešák L and Janout V. Molecular mechanisms of neuropathological changes in Alzheimer's disease: a review. Folia Neuropathol. 2013; 51(1):1-9.
112. Tanzi RE. The genetics of Alzheimer disease. Cold Spring Harb Perspect Med. 2012; 2(10).

113. Mayeux R and Stern Y. Epidemiology of Alzheimer disease. Cold Spring Harb Perspect Med. 2012; 2(8).
114. Qiu C, Kivipelto M and von Strauss E. Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. Dialogues Clin Neurosci. 2009; 11(2):111-128.
115. Vagnucci AH, Jr. and Li WW. Alzheimer's disease and angiogenesis. Lancet. 2003; 361(9357):605-608.
116. Kim J, Basak JM and Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. Neuron. 2009; 63(3):287-303.
117. Oboudiyat C, Glazer H, Seifan A, Greer C and Isaacson RS. Alzheimer's disease. Semin Neurol. 2013; 33(4):313-329.
118. Tiwari S, Atluri V, Kaushik A, Yndart A and Nair M. Alzheimer's disease: pathogenesis, diagnostics, and therapeutics. Int J Nanomedicine. 2019; 14:5541-5554.
119. Panza F, Lozupone M, Logroscino G and Imbimbo BP. A critical appraisal of amyloid-β-targeting therapies for Alzheimer disease. Nat Rev Neurol. 2019; 15(2):73-88.
120. Li P, Nie Y and Yu J. An Effective Method to Identify Shared Pathways and Common Factors among Neurodegenerative Diseases. PLoS One. 2015; 10(11):e0143045.
121. Kunnumakkara AB, Sailo BL, Banik K, Harsha C, Prasad S, Gupta SC, Bharti AC and Aggarwal BB. Chronic diseases, inflammation, and spices: how are they linked? J Transl Med. 2018; 16(1):14.
122. Katsiari C, Bogdanos D and Sakkas L. WJTM. World. 2019; 8(1):1-8.
123. Kinney JW, Bemiller SM, Murtishaw AS, Leisgang AM, Salazar AM and Lamb BT. Inflammation as a central mechanism in Alzheimer's disease. Alzheimers Dement (N Y). 2018; 4:575-590.
124. Carmeliet P. Angiogenesis in life, disease and medicine. Nature. 2005; 438(7070):932-936.
125. Camaré C, Pucelle M, Nègre-Salvayre A and Salvayre R. Angiogenesis in the atherosclerotic plaque. Redox Biol. 2017; 12:18-34.
126. Herrmann J, Lerman LO, Mukhopadhyay D, Napoli C and Lerman A. Angiogenesis in atherogenesis. Arterioscler Thromb Vasc Biol. 2006; 26(9):1948-1957.
127. Jefferies WA, Price KA, Biron KE, Fenninger F, Pfeifer CG and Dickstein DL. Adjusting the compass: new insights into the role of angiogenesis in Alzheimer's disease. Alzheimers Res Ther. 2013; 5(6):64.
128. Strimbu K and Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010; 5(6):463-466.
129. Latham AM, Molina-París C, Homer-Vanniasinkam S and Ponnambalam S. An integrative model for vascular endothelial growth factor A as a tumour biomarker. Integr Biol (Camb). 2010; 2(9):397-407.
130. Takahashi H and Shibuya M. The vascular endothelial growth factor (VEGF)/VEGF receptor system and its role under physiological and pathological conditions. Clin Sci (Lond). 2005; 109(3):227-241.
131. Bowler E and Oltean S. Alternative Splicing in Angiogenesis. Int J Mol Sci. 2019; 20(9).
132. Artandi SE. Telomeres, telomerase, and human disease. N Engl J Med. 2006; 355(12):1195-1197.
133. Yeh JK and Wang CY. Telomeres and Telomerase in Cardiovascular Diseases. Genes (Basel). 2016; 7(9).
134. Fasching CL. Telomere length measurement as a clinical biomarker of aging and disease. Crit Rev Clin Lab Sci. 2018; 55(7):443-465.
135. Saliques S, Zeller M, Lorin J, Lorgis L, Teyssier JR, Cottin Y, Rochette L and Vergely C. Telomere length and cardiovascular disease. Arch Cardiovasc Dis. 2010; 103(8-9):454-459.
136. Rezvan A. Telomeres, oxidative stress, and myocardial infarction. Eur Heart J. 2017; 38(41):3105-3107.
137. Huang Y and Mahley RW. Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. Neurobiol Dis. 2014; 72 Pt A:3-12.
138. Muñoz SS, Garner B and Ooi L. Understanding the Role of ApoE Fragments in Alzheimer's Disease. Neurochem Res. 2019; 44(6):1297-1305.

139. Marais AD. Apolipoprotein E in lipoprotein metabolism, health and cardiovascular disease. Pathology. 2019; 51(2):165-176.
140. Stenger C, Corbier C and Yen FT. (2012). Structure and Function of the Lipolysis Stimulated Lipoprotein Receptor. Chemical Biology.
141. Stenger C, Pinçon A, Hanse M, Royer L, Comte A, Koziel V, Olivier JL, Pillot T and Yen FT. Brain region-specific immunolocalization of the lipolysis-stimulated lipoprotein receptor (LSR) and altered cholesterol distribution in aged LSR+/- mice. J Neurochem. 2012; 123(4):467-476.
142. Yen FT, Roitel O, Bonnard L, Notet V, Pratte D, Stenger C, Magueur E and Bihain BE. Lipolysis stimulated lipoprotein receptor: a novel molecular link between hyperlipidemia, weight gain, and atherosclerosis in mice. J Biol Chem. 2008; 283(37):25650-25659.
143. Zhan Y and Hägg S. Telomere length and cardiovascular disease risk. Curr Opin Cardiol. 2019; 34(3):270-274.
144. Aviv A. Genetics of leukocyte telomere length and its role in atherosclerosis. Mutat Res. 2012; 730(1-2):68-74.
145. Au Yeung SL, Lam H and Schooling CM. Vascular Endothelial Growth Factor and Ischemic Heart Disease Risk: A Mendelian Randomization Study. J Am Heart Assoc. 2017; 6(8).
146. Miners JS, Palmer JC and Love S. Pathophysiology of Hypoperfusion of the Precuneus in Early Alzheimer's Disease. Brain Pathol. 2016; 26(4):533-541.
147. Miners JS, Schulz I and Love S. Differing associations between Aβ accumulation, hypoperfusion, blood-brain barrier dysfunction and loss of PDGFRB pericyte marker in the precuneus and parietal white matter in Alzheimer's disease. J Cereb Blood Flow Metab. 2018; 38(1):103-115.
148. Thomas T, Miners S and Love S. Post-mortem assessment of hypoperfusion of cerebral cortex in Alzheimer's disease and vascular dementia. Brain. 2015; 138(Pt 4):1059-1069.
149. Mahoney ER, Dumitrescu L, Moore AM, Cambronero FE, De Jager PL, Koran MEI, Petyuk VA, Robinson RAS, Goyal S, Schneider JA, Bennett DA, Jefferson AL and Hohman TJ. Brain expression of the vascular endothelial growth factor gene family in cognitive aging and alzheimer's disease. Mol Psychiatry. 2021; 26(3):888-896.
150. Yamazaki Y, Zhao N, Caulfield TR, Liu CC and Bu G. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. Nat Rev Neurol. 2019; 15(9):501-518.
151. Burns MP, Noble WJ, Olm V, Gaynor K, Casey E, LaFrancois J, Wang L and Duff K. Co-localization of cholesterol, apolipoprotein E and fibrillar Abeta in amyloid plaques. Brain Res Mol Brain Res. 2003; 110(1):119-125.
152. Yamaguchi H, Ishiguro K, Sugihara S, Nakazato Y, Kawarabayashi T, Sun X and Hirai S. Presence of apolipoprotein E on extracellular neurofibrillary tangles and on meningeal blood vessels precedes the Alzheimer beta-amyloid deposition. Acta Neuropathol. 1994; 88(5):413-419.
153. Pinçon A, Thomas MH, Huguet M, Allouche A, Colin JC, Georges A, Derrien A, Lanhers MC, Malaplate-Armand C, Oster T, Corbier C, Pillot T, Olivier JL, et al. Increased susceptibility of dyslipidemic LSR+/- mice to amyloid stress is associated with changes in cortical cholesterol levels. J Alzheimers Dis. 2015; 45(1):195-204.
154. Sanzo MD, Cipolloni L, Borro M, Russa RL, Santurro A, Scopetti M, Simmaco M and Frati P. Clinical Applications of Personalized Medicine: A New Paradigm and Challenge. Current Pharmaceutical Biotechnology. 2017; 18(3).
155. Redekop WK and Mladsi D. The faces of personalized medicine: a framework for understanding its meaning and scope. Value Health. 2013; 16(6 Suppl):S4-9.
156. Currie G and Delles C. Precision Medicine and Personalized Medicine in Cardiovascular Disease. Adv Exp Med Biol. 2018; 1065:589-605.
157. Zhang S, Bamakan SMH, Qu Q and Li S. Learning for Personalized Medicine: A Comprehensive Review From a Deep Learning Perspective. IEEE Rev Biomed Eng. 2019; 12:194-208.

158. Sigman M. Introduction: Personalized medicine: what is it and what are the challenges? Fertil Steril. 2018; 109(6):944-945.
159. Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ and Altman RB. Bioinformatics challenges for personalized medicine. Bioinformatics. 2011; 27(16):2323-2323.
160. Goetz LH and Schork NJ. Personalized medicine: motivation, challenges, and progress. Fertil Steril. 2018; 109(6):952-963.
161. Overby CL and Tarczy-Hornoch P. Personalized medicine: challenges and opportunities for translational bioinformatics. Per Med. 2013; 10(5):453-462.
162. Alyass A, Turcotte M and Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics. 2015; 8:33.
163. Liang Y and Kelemen A. Dynamic modeling and network approaches for omics time course data: overview of computational approaches and applications. Brief Bioinform. 2018; 19(5):1051-1068.
164. Strianese O, Rizzo F, Ciccarelli M, Galasso G, D'Agostino Y, Salvati A, Del Giudice C, Tesorio P and Rusciano MR. Precision and Personalized Medicine: How Genomic Approach Improves the Management of Cardiovascular and Neurodegenerative Disease. Genes (Basel). 2020; 11(7).
165. V A, Nayar PG, Murugesan R, Mary B, P D and Ahmed SS. CardioGenBase: A Literature Based Multi-Omics Database for Major Cardiovascular Diseases. PLoS One. 2015; 10(12):e0143188.
166. Bellenguez C, Grenier-Boley B and Lambert JC. Genetics of Alzheimer's disease: where we are, and where we are going. Curr Opin Neurobiol. 2020; 61:40-48.
167. Awata T. Vascular endothelial growth factor gene polymorphisms in susceptibility to coronary artery disease. Am J Hypertens. 2010; 23(9):938-939.
168. Ma WQ, Wang Y, Han XQ, Zhu Y and Liu NF. Association of genetic polymorphisms in vascular endothelial growth factor with susceptibility to coronary artery disease: a meta-analysis. BMC Med Genet. 2018; 19(1):108.
169. Howell WM, Ali S, Rose-Zerilli MJ and Ye S. VEGF polymorphisms and severity of atherosclerosis. J Med Genet. 2005; 42(6):485-490.
170. Iaresko M and Kolesnikova E. THE ROLE OF POLYMORPHISM - 634 G/C (RS 2010963) OF VEGF-A GENE IN THE DEVELOPMENT OF HYPERTENSION AND OBESITY IN PREMENOPAUSAL WOMEN. Georgian Med News. 2016; (256-257):33-37.
171. Wang Y, Huang Q, Liu J, Wang Y, Zheng G, Lin L, Yu H, Tang W and Huang Z. Vascular endothelial growth factor A polymorphisms are associated with increased risk of coronary heart disease: a meta-analysis. Oncotarget. 2017; 8(18):30539-30551.
172. Oh SM, Stefani KM and Kim HC. Development and application of chronic disease risk prediction models. Yonsei Med J. 2014; 55(4):853-860.
173. Rossello X, Dorresteijn JA, Janssen A, Lambrinou E, Scherrenberg M, Bonnefoy-Cudraz E, Cobain M, Piepoli MF, Visseren FL, Dendale P, This Paper Is AC-PBEJOPCEHJACC and European Journal Of Cardiovascular N. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardiovascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). Eur J Prev Cardiol. 2019; 26(14):1534-1544.
174. Grammer TB, Dressel A, Gergei I, Kleber ME, Laufs U, Scharnagl H, Nixdorff U, Klotsche J, Pieper L, Pittrow D, Silber S, Wittchen HU and März W. Cardiovascular risk algorithms in primary care: Results from the DETECT study. Sci Rep. 2019; 9(1):1101.
175. Marinescu RV, Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW, Barkhof F, Fox NC, Golland P, Klein S and Alexander DC. TADPOLE Challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data. Predict Intell Med. 2019; 11843:1-10.

176. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF and van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. PLoS One. 2019; 14(5):e0213653.
177. Park JH, Cho HE, Kim JH, Wall MM, Stern Y, Lim H, Yoo S, Kim HS and Cha J. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. NPJ Digit Med. 2020; 3:46.
178. Okser S, Pahikkala T and Aittokallio T. Genetic variants and their interactions in disease risk prediction - machine learning and network perspectives. BioData Min. 2013; 6(1):5.
179. Ho DSW, Schierding W, Wake M, Saffery R and O'Sullivan J. Machine Learning SNP Based Prediction for Precision Medicine. Front Genet. 2019; 10:267.
180. Curtis RE, Yin J, Kinnaird P and Xing EP. Finding genome-transcriptome-phenome association with structured association mapping and visualization in GenAMap. Pac Symp Biocomput. 2012:327-338.
181. Nusrat S, Harbig T and Gehlenborg N. Tasks, Techniques, and Tools for Genomic Data Visualization. Comput Graph Forum. 2019; 38(3):781-805.
182. Siest G, Visvikis S, Herbeth B, Gueguen R, Vincent-Viry M, Sass C, Beaud B, Lecomte E, Steinmetz J, Locuty J and Chevrier P. 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.
183. Visvikis-Siest S and 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):733-747.
184. Miller SA, Dykes DD and Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic acids research. 1988; 16(3):1215.
185. He C, Holme J and Anthony J. SNP genotyping: the KASP assay. Methods in molecular biology (Clifton, NJ). 2014; 1145:75-86.
186. 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. Scandinavian journal of clinical and laboratory investigation Supplementum. 1968; 97:77-89.
187. Marteau JB, Mohr S, Pfister M and Visvikis-Siest S. Collection and storage of human blood cells for mRNA expression profiling: a 15-month stability study. Clinical chemistry. 2005; 51(7):1250-1252.
188. Chahine MN, Toupance S, El-Hakim S, Labat C, Gautier S, Moussallem T, Yared P, Asmar R and Benetos A. Telomere length and age-dependent telomere attrition: the blood-and-muscle model (1). Canadian journal of physiology and pharmacology. 2019; 97(4):328-334.
189. Li J, Guasch-Ferré M, Chung W, Ruiz-Canela M, Toledo E, Corella D, Bhupathiraju SN, Tobias DK, Tabung FK, Hu J, Zhao T, Turman C, Feng YA, et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. Eur Heart J. 2020; 41(28):2645-2656.
190. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, Zhu S and Ye Z. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med. 2019; 17(1):119.
191. Das J, Gayvert KM, Bunea F, Wegkamp MH and Yu H. ENCAPP: elastic-net-based prognosis prediction and biomarker discovery for human cancers. BMC Genomics. 2015; 16(1):263.
192. Millstein J, Budden T, Goode EL, Anglesio MS, Talhouk A, Intermaggio MP, Leong HS, Chen S, Elatre W, Gilks B, Nazeran T, Volchek M, Bentley RC, et al. Prognostic gene expression signature for high-grade serous ovarian cancer. Ann Oncol. 2020; 31(9):1240-1250.
193. Cho S, Kim H, Oh S, Kim K and Park T. Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. BMC Proc. 2009; 3 Suppl 7(Suppl 7):S25.

194. Waldmann P, Mészáros G, Gredler B, Fuerst C and Sölkner J. Evaluation of the lasso and the elastic net in genome-wide association studies. Front Genet. 2013; 4:270.
195. Frost HR, Shen L, Saykin AJ, Williams SM and Moore JH. Identifying significant gene-environment interactions using a combination of screening testing and hierarchical false discovery rate control. Genet Epidemiol. 2016; 40(7):544-557.
196. Momen M, Mehrgardi AA, Sheikhi A, Kranis A, Tusell L, Morota G, Rosa GJM and Gianola D. Predictive ability of genome-assisted statistical models under various forms of gene action. Sci Rep. 2018; 8(1):12309.
197. Gorenjak V, Vance DR, Petrelis AM, Stathopoulou MG, Dadé S, El Shamieh S, Murray H, Masson C, Lamont J, Fitzgerald P and Visvikis-Siest S. Peripheral blood mononuclear cells extracts VEGF protein levels and VEGF mRNA: Associations with inflammatory molecules in a healthy population. PLoS One. 2019; 14(8):e0220902.
198. Toupance S, Stathopoulou MG, Petrelis AM, Gorenjak V, Labat C, Lai TP, Visvikis-Siest S and Benetos A. TERC Variants Associated with Short Leukocyte Telomeres: Implication of Higher Early Life Leukocyte Telomere Attrition as Assessed by the Blood-and-Muscle Model. Cells. 2020; 9(6).
199. Gorenjak V, Petrelis AM, Stathopoulou MG, Toupance S, Kumar S, Labat C, Masson C, Murray H, Lamont J, Fitzgerald P, Benetos A and Visvikis-Siest S. A genetic determinant of VEGF-A levels is associated with telomere attrition. Aging (Albany NY). 2021; 13(20):23517-23526.

Other publications

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; previously published online October 31, 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 "Gene-Environment Interactions in Cardio-Vascular Physiopathology", University of 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 "Gene-Environment Interactions in Cardio-Vascular Physiopathology", University of Lorraine, Nancy, France
Daniela Ruggiero: 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
Marina Ciullo: Institute of Genetics and Biophysics, National Research Council of Italy, Naples, Italy; and IRCCS Neuromed, Pozzilli, Isernia, Italy
Sudha Seshadri: Department of Neurology, Boston University School of Medicine, Boston, MA, USA

Brought to you by | University of Groningen
Authenticated
Download Date | 4/18/18 2:11 PM

– 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 genome-wide association studies (GWAS) of 10 genetic variants that explain >50% of the circulating VEGF-A levels variability [10, 11]. This is a unique finding among GWAS. In most GWAS, 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

Brought to you by | University of Groningen
Authenticated
Download Date | 4/18/18 2:11 PM

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

Brought to you by | University of Groningen
Authenticated
Download Date | 4/18/18 2:11 PM

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. Patent number: WO2014006231 A1. Available at: http://www.google.com.pg/patents/WO2014006231A1?cl=en&hl=fr. Accessed: 1 June 2017.
20. Randox. Identification of genetic variants – VEGF, 2012. Patent number: US20130202587 A1. Available at: https://www.google.com/patents/US20130202587. Accessed: 1 June 2017.

Brought to you by | University of Groningen
Authenticated
Download Date | 4/18/18 2:11 PM

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/).

Human Molecular Genetics, 2021, Vol. 30, No. 5 393–409

doi: 10.1093/hmg/ddab023
Advance Access Publication Date: 30 January 2021
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Genome-wide association study of circulating interleukin 6 levels identifies novel loci

Tarunveer S. Ahluwalia[1,2,*,†], Bram P. Prins[3], Mohammadreza Abdollahi[3], Nicola J. Armstrong[4], Stella Aslibekyan[5], Lisa Bain[6], Barbara Jefferis[7], Jens Baumert[8], Marian Beekman[9], Yoav Ben-Shlomo[10], Joshua C. Bis[11], Braxton D. Mitchell[12], Eco de Geus[13,14], Graciela E. Delgado[15], Diana Marek[16], Joel Eriksson[17], Eero Kajantie[18,19], Stavroula Kanoni[20], John P. Kemp[21,22], Chen Lu[23], Riccardo E. Marioni[24,25], Stela McLachlan[26], Yuri Milaneschi[27], Ilja M. Nolte[3], Alexandros M. Petrelis[28], Eleonora Porcu[29], Maria Sabater-Lleal[30,31], Elnaz Naderi[3], Ilkka Seppälä[32], Tina Shah[33], Gaurav Singhal[34], Marie Standl[8], Alexander Teumer[35,‡], Anbupalam Thalamuthu[36], Elisabeth Thiering[8,37], Stella Trompet[38,39], Christie M. Ballantyne[40], Emelia J. Benjamin[41,42], Juan P. Casas[43], Catherine Toben[34], George Dedoussis[44], Joris Deelen[9,45], Peter Durda[46], Jorgen Engmann[33], Mary F. Feitosa[47], Harald Grallert[8,48], Ann Hammarstedt[49], Sarah E. Harris[25,50], Georg Homuth[51], Jouke-Jan Hottenga[13,14], Sirpa Jalkanen[52,53], Yalda Jamshidi[54], Magdalene C. Jawahar[34], Tine Jess[55], Mika Kivimaki[56], Marcus E. Kleber[15], Jari Lahti[57,58], Yongmei Liu[59], Pedro Marques-Vidal[60,61], Dan Mellström[17], Simon P. Mooijaart[39], Martina Müller-Nurasyid[62,63,¶], Brenda Penninx[27], Joana A. Revez[6,‖], Peter Rossing[1,64], Katri Räikkönen[58], Naveed Sattar[65], Hubert Scharnagl[66], Bengt Sennblad[30,67], Angela Silveira[30], Beate St Pourcain[22,68,69], Nicholas J. Timpson[22], Julian Trollor[36,70], CHARGE Inflammation Working Group[3], Jenny van Dongen[13,14], Diana Van Heemst[40], Sophie Visvikis-Siest[28], Peter Vollenweider[60,61], Uwe Völker[52], Melanie Waldenberger[8], Gonneke Willemsen[13,14], Delilah Zabaneh[71], Richard W. Morris[72], Donna K. Arnett[73], Bernhard T. Baune[74,75,76],

[†]Tarunveer S. Ahluwalia, http://orcid.org/0000-0002-7464-3354
[‡]Alexander Teumer, http://orcid.org/0000-0002-8309-094X
[¶]Martina Müller-Nurasyid, http://orcid.org/0000-0003-3793-5910
[‖]Joana A. Revez, http://orcid.org/0000-0003-3204-5396
Received: December 2, 2020. **Revised:** December 2, 2020. **Accepted:** January 4, 2021

© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Dorret I. Boomsma[13,14], Yen-Pei C. Chang[12], Ian J. Deary[25,50], Panos Deloukas[20,77], Johan G. Eriksson[78,79], David M. Evans[21,22], Manuel A. Ferreira[6], Tom Gaunt[80,81], Vilmundur Gudnason[82,83], Anders Hamsten[30], Joachim Heinrich[8,84,85], Aroon Hingorani[33], Steve E. Humphries[33], J. Wouter Jukema[39,86], Wolfgang Koenig[87,88,89], Meena Kumari[56,90], Zoltan Kutalik[16,91], Deborah A. Lawlor[80,81], Terho Lehtimäki[32], Winfried März[15,66,92], Karen A. Mather[36,93], Silvia Naitza[29], Matthias Nauck[94,95], Claes Ohlsson[17], Jackie F. Price[26], Olli Raitakari[96,97,98], Ken Rice[99], Perminder S. Sachdev[36,100], Eline Slagboom[9,45], Thorkild I.A. Sørensen[101,102], Tim Spector[103], David Stacey[104], Maria G. Stathopoulou[28], Toshiko Tanaka[105], S. Goya Wannamethee[7], Peter Whincup[106], Jerome I. Rotter[107], Abbas Dehghan[108], Eric Boerwinkle[109,110], Bruce M. Psaty[11,111], Harold Snieder[3] and Behrooz Z. Alizadeh[3,*]

[1]Steno Diabetes Center Copenhagen, Gentofte DK2820, Denmark, [2]Department of Biology, The Bioinformatics Center, University of Copenhagen, Copenhagen DK2200, Denmark, [3]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands, [4]Mathematics and Statistics, Murdoch University, Perth 6150, Australia, [5]Department of Epidemiology, University of Alabama at Birmingham School of Public Health, Birmingham, Alabama 35233, USA, [6]QIMR Berghofer Medical Research Institute, Brisbane 4006, Australia, [7]Department of Primary Care & Population Health, UCL Institute of Epidemiology & Health Care, University College London, London NW3 2PF, UK, [8]Institute of Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg 85764, Germany, [9]Department of Biomedical Data Sciences, Section of Molecular Epidemiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands, [10]Population Health Sciences, University of Bristol, Bristol BS8 2PS, UK, [11]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA, [12]Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21202, USA, [13]Department of Biological Psychology, Behaviour and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam 1081 BT, The Netherlands, [14]Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Amsterdam 1105 AZ, The Netherlands, [15]Vth Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim 68167, Germany, [16]SIB Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland, [17]Department of Internal Medicine and Clinical Nutrition, Sahlgrenska Academy, Centre for Bone and Arthritis Research (CBAR), University of Gothenburg, Gothenburg 41345, Sweden, [18]Chronic Disease Prevention Unit, National Institute for Health and Welfare, PO Box 30, Helsinki 00271, Finland, [19]Hospital for Children and Adolescents, Helsinki University Central Hospital and University of Helsinki, Helsinki 00014, Finland, [20]William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London, London EC1M 6BQ, UK, [21]The University of Queensland Diamantina Institute, The University of Queensland, Woolloongabba, Queensland 4102, Australia, [22]MRC Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol BS8 2BN, UK, [23]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA, [24]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK, [25]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK, [26]Usher Institute, University of Edinburgh, Edinburgh EH8 9AG, UK, [27]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit, Amsterdam 1081 HJ, The Netherlands, [28]Université de Lorraine, Inserm, IGE-PCV, Nancy 54000, France, [29]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato (CA) 09042, Italy, [30]Cardiovascular Medicine, Department of Medicine Solna, Center for Molecular Medicine, Karolinska Institutet, Stockholm 17176, Sweden, [31]Unit of Genomics of Complex Diseases, Institut d'Investigació Biomèdica Sant Pau (IIB-Sant Pau), Barcelona 08041, Spain, [32]Department of Clinical Chemistry, Fimlab

Laboratories, and Finnish Cardiovascular Research Center—Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland, [33]Institute of Cardiovascular Science, University College London, London WC1E 6BT, UK, [34]Discipline of Psychiatry, Adelaide Medical School, University of Adelaide, Adelaide 5005, Australia, [35]Institute for Community Medicine, University Medicine Greifswald, Greifswald 17475, Germany, [36]Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney 2052, Australia, [37]Division of Metabolic Diseases and Nutritional Medicine, Ludwig-Maximilians-University of Munich, Dr. von Hauner Children's Hospital, Munich 80337, Germany, [38]Department of Cardiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands, [39]Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands, [40]Baylor College of Medicine, Houston, TX 77030, USA, [41]National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA 01702, USA, [42]Section of Cardiovascular Medicine and Preventive Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA, [43]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA 02130, USA, [44]Department of Nutrition-Dietetics, Harokopio University, Athens 17671, Greece, [45]Max Planck Institute for Biology of Ageing, Cologne 50931, Germany, [46]Department of Pathology and Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, VT 05405, USA, [47]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110-1093, USA, [48]German Center for Diabetes Research (DZD), Neuherberg 85764, Germany, [49]The Lundberg Laboratory for Diabetes Research, Department of Molecular and Clinical Medicine, Sahlgrenska Academy at the University of Gothenburg, Gothenburg SE-41345, Sweden, [50]Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK, [51]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald 17475, Germany, [52]MediCity Research Laboratory, University of Turku, Turku 20520, Finland, [53]Department of Medical Microbiology and Immunology, University of Turku, Turku 20520, Finland, [54]Genetics Research Centre, Molecular and Clinical Sciences Institute, St George's University of London, London SW17 0RE, UK, [55]Department of Epidemiology Research, Statens Serum Institute, Copenhagen DK2300, Denmark, [56]Department of Epidemiology & Public Health, UCL Institute of Epidemiology & Health Care, University College London, London WC1E 7HB, UK, [57]Turku Institute for Advanced Studies, University of Turku, Turku 20014, Finland, [58]Department of Psychology and Logopedics, University of Helsinki, Helsinki 00014, Finland, [59]Department of Epidemiology and Prevention, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA, [60]Department of Internal Medicine, Lausanne University Hospital (CHUV), Lausanne 1011, Switzerland, [61]University of Lausanne, Lausanne 1011, Switzerland, [62]IBE, Faculty of Medicine, Ludwig Maximilians University (LMU) Munich, Munich 81377, Germany, [63]Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johhanes Gutenberg University, Mainz 55101, Germany, [64]Department of Clinical Medicine, University of Copenhagen, Copenhagen DK2200, Denmark, [65]BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow G12 8TA, UK, [66]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8036, Austria, [67]Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala 75124, Sweden, [68]Max Planck Institute for Psycholinguistics, Nijmegen XD 6525, The Netherlands, [69]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen 6525 AJ, The Netherlands, [70]Department of Developmental Disability Neuropsychiatry, School of Psychiatry, University of New South Wales, Sydney 2031, Australia, [71]Department of Genetics, Environment and Evolution, University College London Genetics Institute, London WC1E 6BT, UK, [72]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 1UD, UK, [73]Dean's Office, College of Public Health, University of Kentucky, Lexington, KY 40536, USA, [74]Department of Psychiatry, Melbourne Medical School, The University of Melbourne, Parkville 3000, Australia, [75]Department of Psychiatry and Psychotherapy, University of Muenster, Muenster 48149, Germany, [76]The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville 3000, Australia, [77]Centre for Genomic Health, Queen Mary University of London, London EC1M 6BQ, UK, [78]National Institute for Health and Welfare, University of Helsinki, Helsinki 00014, Finland, [79]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki 00014, Finland, [80]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol BS6 2BN, UK, [81]Population Health Science, Bristol

Medical School, University of Bristol, Bristol BS8 2BN, UK, [82]Icelandic Heart Association, Kópavogur 201, Iceland, [83]Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland, [84]Institute and Clinic for Occupational, Social and Environmental Medicine, University Hospital, LMU Munich, Munich 81377, Germany, [85]Allergy and Lung Health Unit, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne 3010, Australia, [86]Durrer Center for Cardiogenetic Research, Amsterdam 1105 AZ, The Netherlands, [87]Deutsches Herzzentrum München, Technische Universität München, Munich 80636, Germany, [88]DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich 80336, Germany, [89]Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm 89081, Germany, [90]Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ, Germany, [91]University Center for Primary Care and Public Health, University of Lausanne, Lausanne 1010, Switzerland, [92]SYNLAB Academy, SYNALB Holding Deutschland GmbH, Mannheim 68163, Germany, [93]Neuroscience Research Australia, Sydney 2031, Australia, [94]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald 17475, Germany, [95]DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald 17475, Germany, [96]Centre for Population Health Research, University of Turku, Turku University Hospital, Turku 20520, Finland, [97]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland, [98]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20014, Finland, [99]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA, [100]Neuropsychiatric Institute, Prince of Wales Hospital, Sydney 2031, Australia, [101]Novo Nordisk Foundation Center For Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen DK2200, Denmark, [102]Department of Public Health, Section on Epidemiology, University of Copenhagen, Copenhagen DK1014, Denmark, [103]Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK, [104]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK, [105]Longitudinal Study Section, Translational Gerontology Branch, National Institute on Aging, Baltimore, MD 21224, USA, [106]Population Health Research Institute, St George's, University of London, London SW17 0RE, UK, [107]Department of Pediatrics, The Institute for Translational Genomics and Population Sciences, The Lundquist Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA, [108]Department of Epidemiology, Erasmus MC, Rotterdam 3000 CA, The Netherlands, [109]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030, USA, [110]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA and [111]Departments of Epidemiology and Health Services, University of Washington, Seattle, WA 98101, USA

*To whom correspondence should be addressed at: Steno Diabetes Center Copenhagen, Gentofte, and Department of Biology, The Bioinformatics Center, University of Copenhagen, Copenhagen, Denmark. Email: tarun.veer.singh.ahluwalia@regionh.dk (Tarunveer S. Ahluwalia); Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. Email: b.z.alizadeh@umcg.nl (Behrooz Z. Alizadeh)

Abstract

Interleukin 6 (IL-6) is a multifunctional cytokine with both pro- and anti-inflammatory properties with a heritability estimate of up to 61%. The circulating levels of IL-6 in blood have been associated with an increased risk of complex disease pathogenesis. We conducted a two-staged, discovery and replication meta genome-wide association study (GWAS) of circulating serum IL-6 levels comprising up to 67 428 ($n_{discovery} = 52\,654$ and $n_{replication} = 14\,774$) individuals of European ancestry. The inverse variance fixed effects based discovery meta-analysis, followed by replication led to the identification of two independent loci, *IL1F10/IL1RN* rs6734238 on chromosome (Chr) 2q14, ($P_{combined} = 1.8 \times 10^{-11}$), *HLA-DRB1/DRB5* rs660895 on Chr6p21 ($P_{combined} = 1.5 \times 10^{-10}$) in the combined meta-analyses of all samples. We also replicated the *IL6R* rs4537545 locus on Chr1q21 ($P_{combined} = 1.2 \times 10^{-122}$). Our study identifies novel loci for circulating IL-6 levels uncovering new immunological and inflammatory pathways that may influence IL-6 pathobiology.

Introduction

Interleukin 6 (IL-6) is a multifunctional cytokine, which is involved in a wide range of immunomodulatory processes, from cellular migration and adhesion to proliferation and maturation (1,2). Interleukins are involved in immune cell differentiation and activation (3). IL-6 is synthesized by a variety of different immune cells such as monocytes (4), B cells (5) and T cells (6) and also non-immune cells such as epithelial and smooth muscle cells (7), adipocytes (8), endothelial cells (9) and osteoblasts (10).

Several factors have been implicated in circulating IL-6 levels. We have previously demonstrated that IL-6 levels decrease with age in children and increase with age in adults (11). Also, increased levels of IL-6 have been observed in various diseases, not surprisingly in autoimmune diseases such as rheumatoid arthritis (12) and systemic juvenile idiopathic arthritis (13), but also cardio-metabolic diseases like type 2 diabetes (14), heart failure, coronary heart disease (15) and atherosclerosis (16), as well in cancers (17), atopic dermatitis (18) and psychological disorders like depression (19). Due to its implications in the pathogenesis of different disorders, Il-6 has been used as an appropriate choice for drug targeting and used as a monitoring biomarker of disease progression and response to treatments (20). The most illustrious IL-6 inhibitor is tocilizumab (21), a monoclonal antibody binding the IL-6 receptor, which is already in use for treating patients with allergic asthma (22), and immune system disorders like rheumatoid arthritis (23) and systemic juvenile idiopathic arthritis (24), with high efficacy with some initial benefits towards respiratory illnesses like COVID-19 (25).

IL-6 baseline levels are heritable with estimates from twin studies ranging between 15 and 61% (26–29). However, efforts to identify genetic variants associated with levels of IL-6 constituted relatively small-scale GWAS (30–33) or sequencing-based candidate gene association studies (34). To date, variants in the IL-6 receptor gene (*IL-6R*) and the gene encoding histo-blood group ABO system transferase (*ABO*) have been identified as statistically significant for an association to IL6-levels. Also, the genetic risk score constructed of *IL-6* variants identified in the study by Shah and colleagues explained up to 2% of the variation in IL-6 levels (33), leaving a substantial part of its heritability unexplained. These seemingly sparse results and limited findings could be due to limitations in the study power caused by low sample size or a great inter-individual variability of IL-6 levels. One may speculate a substantial increase in the study size by increasing the number of participants, which would very likely lead to the identification of additional variants explaining IL-6 levels (35–37).

The current study is the (till date) largest meta GWAS study including 67 428 individuals of European ancestry to identify genetic variants explaining the levels of circulating IL-6 and to understand underlying genetic mechanisms implicated in the pathophysiology of this cytokine.

Results

A total of 52 654 individuals of European descent from 26 cohorts were included in the discovery GWAS meta-analysis with up to 2 454 025 autosomal SNPs passing quality control. Four cohorts (ALSPAC, MONICA/KORA, NTR and SardiNIA) identified genome-wide significant associations in the *ABO* region, whereas none of the other 22 cohorts did, either individually or combined. These cohorts conditioned their results on their relevant top-SNP in *ABO*, the results of which were included in the discovery meta-analyses. The overall genomic control inflation factor (λ_{GC} after correction) at the discovery stage meta-analysis was 1.0.

We identified 94 variants that were genome-wide significantly ($P_{discovery} < 5.0 \times 10^{-8}$; Supplementary Material, Table S1) associated with IL-6 levels, representing two independent genetic loci on chromosomes 1q21 and 6p21. Two common SNPs (rs4537545 and rs660895), one per locus, Chr. 1q21 (*IL6R*), and Chr. 6p21 (*HLA-DRB1/HLA-DRB5*), showed the most significant association with IL-6 levels (index SNPs) and the third SNP (rs6734238) mapped on Chr. 2q14 (*IL1F10/IL1RN*) locus showed suggestive ($5.0 \times 10^{-8} < P_{discovery} < 1.0 \times 10^{-5}$) association in addition to 5 other loci (*LHFPL3*, *LZTS1*, *GPC5/GPC6*, *USP32/APPBP2*, *STAU1*; Supplementary Material, Table S2). Manhattan and QQ plots have been depicted in Figures 1A and 1B.

The minor alleles of *IL6R* rs4537545*T ($\beta = 0.091$; $P_{discovery} = 8.39 \times 10^{-85}$), *IL1F10/IL1RN* rs6734238*G ($\beta = 0.025$; $P_{discovery} = 1.45 \times 10^{-7}$) and *HLA-DRB1/5* rs660895*G ($\beta = 0.036$; $P_{discovery} = 1.80 \times 10^{-9}$) associated with increased circulating IL-6 levels (Table 1). Two additional genome-wide significant SNPs in the *IL1R* locus, rs11265618 ($\beta = 0.047$; $P_{discovery} = 1.21 \times 10^{-15}$) and rs10796927 ($\beta = 0.034$; $P_{discovery} = 1.24 \times 10^{-11}$), in low LD ($r^2 < 0.25$) with the lead SNP rs4537545 were carried forward for replication, and later conditional analysis as they seemed potential candidates as independent signals.

Overall, 12 SNPs spanning over 9 loci at a $P_{discovery} < 1 \times 10^{-5}$ in the discovery GWAS meta-analyses were selected for the replication stage (Supplementary Material, Table S2). This included the three index SNPs, two additional SNPs from the 1q21 locus (GWS but in low LD, $r^2 < 0.25$ with index SNP) plus an additional set of seven statistically suggestive SNPs with a *P*-value of $5 \times 10^{-8} < P < 1 \times 10^{-5}$ in the discovery meta-analyses (either in low LD, $r^2 < 0.25$ with the index SNP or independent loci). Additionally, 3 SNPs as negative controls and 3 SNPs in LD ($r^2 > 0.25$) with the Chr.1 index SNP, to control for possible genotyping errors of index SNP across replication cohorts, were also added to the replication list, yielding 18 SNPs for replication stage (Supplementary Material, Table S3).

Three loci including Chr.1q21 *IL6R*, Chr.6p21 *HLA-DRB1/5* and Chr.2q14 *ILF10/IL1RN* replicated at $P_{replication} < 0.05$, reaching GWS; 1q21 rs4537545, $P_{combined} = 1.20 \times 10^{-122}$; 6p21 rs660895, $P_{combined} = 1.55 \times 10^{-10}$; and 2q14 rs6734238, $P_{combined} = 1.84 \times 10^{-11}$ in the combined meta-analyses (Table 1 and Supplementary Material, Table S3). Locus zoom plots available in Figures 1C, 1D, and 1E. The two additional signals at Chr.1q21 *IL6R* locus were replicated at $P_{replication} = 1.7 \times 10^{-4}$ for rs11265618 and $P = 0.03$ for rs10796927, reaching $P_{combined} = 2.5 \times 10^{-9}$ and $P_{combined} = 4.1 \times 10^{-13}$, respectively (Supplementary Material, Table S3). The conditional analysis confirmed that rs11265618 and rs10796927 SNPs were not independent from (Supplementary Material, Table S4) but were driven by the index rs4537545 SNP.

In both, discovery and replication association analyses, the effect directionality was generally consistent across individual studies for GWS variants, while there was some evidence of borderline heterogeneity in one of the two novel loci (I^2 (*P* value) < 0.05) during the discovery and combined meta-analysis (Table 1, Fig. 2). The imputation quality scores (r^2) for the GWS (index) SNPs for each cohort (discovery and replication) are available in Supplementary Material, Table S5. The other seven SNPs that showed suggestive association in the discovery stage, and expectedly the negative control SNPs did not reach GWS in the combined meta-analyses (Supplementary Material, Table S3).

Table 1. Novel and replicated loci associated with circulating IL-6 levels at $P < 5.0 \times 10^{-8}$ in the combined GWAS meta analyses

Chr	Lead SNP (rsID)	BP (Hg19)	Effect/Other allele	EAF	Beta (SE)	$P_{discovery}$	$P_{replication}$	$P_{combined}$	Annotation	Nearest Genes	P_{het} (I^2)
Novel Loci											
2q14	rs6734238	113 841 030	G/A	0.42	0.025 (0.005)	1.45×10^{-7}	3.24×10^{-5}	1.84×10^{-11}	Intergenic	*IL1F10/IL1RN*	0.03 (32)
6p21	rs660895	32 577 380	G/A	0.19	0.036 (0.006)	1.80×10^{-9}	3.38×10^{-2}	1.55×10^{-10}	Intergenic	*HLA-DRB5/DRB1*	0.08 (26)
Replicated Known Locus											
1q21	rs4537545	154 418 879	T/C	0.39	0.091 (0.005)	8.39×10^{-85}	7.88×10^{-37}	1.20×10^{-122}	Intronic	*IL6R*	<0.01 (72)

Index SNPs that reached $P < 5 \times 10^{-8}$ in the combined analysis from each locus are reported here. Sample sizes: discovery cohorts, $n = 52\,654$; replication cohorts, $n = 14\,774$; combined, $n = 67\,428$. The effect sizes (β) in the discovery phase, given for the effect allele. EAF: effect allele frequency; effect sizes and standard error (SE) values are based on natural log transformed IL6 (pg/ml) levels. Phet: combined meta-analysis heterogeneity P value; I^2: heterogeneity measure.

The three GWAS index SNPs when combined explained approximately 1.06% of the variance in circulating levels of IL-6 using data from the NESDA cohort. The phenotypic variance explained by all the common variants was estimated to be 4.45% using the SumVg method (38).

Replication of other known/suggestive loci for IL-6

IL6R was the only IL6 known locus that we replicate at GWS. *IL1RN* and *HLA-DRB1*, our primary findings have been reported as suggestive loci ($1 \times 10^{-6} < P < 1 \times 10^{-4}$) by Shah *et al.* while some known/suggestive IL6 loci (*ABO*, *BUD13*, *TRIB3* and *SEZ6L*) did not replicate ($P_{discovery} > 0.05$) in the current study.

SNP functionality

We looked up SNPs in LD with the index SNPs from the immunologically associated loci including *IL-6R*, rs4537545, 1q21; *IL1F10*, *IL1RN*, rs6734238, 2q14, intergenic; and *HLA-DRB1/DRB5*, rs660895, 6p21, intergenic. The search for functional/missense variants in high LD ($r^2 > 0.8$) with the lead SNPs led to the identification of only one nonsynonymous rs2228145 SNP in LD ($r^2 = 0.95$) with the rs4537545 index SNP from the *IL6R* locus. We used the Combined Annotation-Dependent Depletion (CADD) database to identify the functionality, i.e. deleterious, disease causal, pathogenicity, of rs2228145 in *IL6R*. CADD is an integrative annotation based on multiple genomic features scored into a single metric (39). *IL6R* missense the rs2228145 variant has a CADD score of 15.98 (https://cadd.gs.washington.edu).

Associations with other traits and gene expression data

Genome-wide significant associations between IL6-associated top SNPs and other traits, and gene expression, data were mined using the Pheno Scanner v2 database (accessed, October 2020).

GWAS-based *IL1F10/IL1RN* rs6734238*G allele has been associated with increased levels of serum C-reactive protein (CRP) and decreased fibrinogen levels, and blood cell traits in recent GWAS reports (40,41) (PMID:27863252; Supplementary Material, Table S6).

HLA-DRB1/DRB5 rs660895*G allele is associated with increased risk of rheumatoid arthritis (RA) in Europeans and Asians (42), IgA nephropathy in Asians (43), while the decreased risk of ulcerative colitis and inflammatory bowel disease (IBD) (44). *IL-6R* rs4537545*T allele has been associated with increased circulating CRP levels (45), a decreased risk of RA (42) in mixed ancestries, while an increased risk of diabetes and asthma from the UK Biobank Neale's lab rapid GWAS (See Web Resources; Supplementary Material, Table S6). *IL6R* rs4537545T* allele is also associated with C-reactive protein, allergic disease, rheumatoid arthritis and coronary artery disease (Supplementary Material, Table S6).

Gene expression. *IL1F10/IL1RN* rs6734238 is associated with IL1F10/IL1RN expression levels in the skin, peripheral blood and whole blood ($P < 5.0 \times 10^{-8}$; Supplementary Material, Table S7). *HLA-DRB1/DRB5* rs660895 has been associated with HLA-DRB1/DRB5/DRB6/DQB1/DQB2 expression levels in multiple tissues including peripheral blood, whole blood, monocytes, adipose tissue, thyroid, tibial artery, coronary artery, heart, lung, brain, colon, skeletal muscle, tibial nerve, skin and lymphoblastoid cell lines ($P < 5.0 \times 10^{-8}$; Supplementary Material, Table S7). *IL6R* rs4537545 SNP is also associated with IL6R expression levels

in peripheral and whole blood ($P < 5.0 \times 10^{-8}$; Supplementary Material, Table S7).

Power estimates

Based on power calculator and assumptions mentioned under methods section, the estimated power for the 2 novel index SNPs was 98.3% rs6734238 (effect allele frequency, EAF: 0.42), and 76.9% rs660895 (EAF: 0.19), respectively.

Discussion

We performed the largest (to date) GWAS meta-analysis for circulating IL-6 levels, which includes 66 341 individuals of European ancestry. We identified three loci associated with levels of circulating of IL-6 in the general population amongst which two are novel (Chr6p21, and Chr2q14), located in/nearby genes (*HLA-DRB1* and *IL1RN*/*IL-38*) with inflammatory roles explaining up to 1.06% variance.

The strongest associated SNP, interleukin 6 receptor (*IL-6R*) rs4537545 at the 1q21 locus, is in high LD ($r^2 = 0.95$) with a missense *IL-6R* SNP rs2228145 (D358A) that results in an amino acid substitution at position 358 (Asp → Ala) on the extracellular domain of IL-6R and a high CADD score suggesting that the variant is pathogenic or functional or deleterious (among top 10% variants of the genome). The missense SNP is known to impair the responsiveness of cells targeted by IL-6 (46) by reducing IL-6R expression on cell surfaces (47), and increasing levels of soluble IL-6R in individuals homozygous for this mutation (48,49). Recently it has been demonstrated that increased levels of sIL-6R induced by this variant can be explained by ectodomain shedding off IL-6R, a mechanism in which membrane-associated proteins are rapidly converted into soluble effectors whereby simultaneously cell surface expression of the same protein is reduced (50). Increased levels of sIL-6R may act as a counterbalance to limit exaggerated IL-6 signaling and may explain the protective effect of the 358A allele for various cardiovascular diseases including coronary artery disease (CAD) (51–53), atrial fibrillation (54), lung function in asthmatics (55) and abdominal aortic aneurysm (56) as well as RA (57). However, in contrast with this finding, the IL-6-sIL-6R complex itself is capable of transducing IL-6 signaling to non-IL-6R expressing cells, known as trans-signaling (58), and it is this mechanism, as opposed to classic signaling, that is linked to chronic inflammatory disorders including IBD and RA (59). Blocking IL-6 signaling cascades can be achieved by using an IL-6R specific inhibitor in the form of a monoclonal antibody, tocilizumab, which is a widely used therapy in the treatment of RA. Several variants in IL-6R, including rs2228145, may assist in the prediction of patient response to tocilizumab in RA (60). The rs4537545*T allele that is associated with IL6 levels is known to associate with increased circulating CRP levels (61) and a decreased risk of RA (42) in studies comprising mixed ancestries. Moreover, this SNP has been associated with IL6R expression in peripheral blood, skin, brain and adipose tissue (Supplementary Material, Table S7). The causal involvement of IL-6 levels in disease remains to be elucidated, but a recent study using a Mendelian randomisation (MR) approach did demonstrate that by using this SNP as instrumental variable, modelling the effects of tocilizumab, that IL-6R signalling has a causal effect on CAD (52). On the other hand, pleiotropic nature of the IL-6R locus, influencing IL-6, CRP and fibrinogen levels, prohibits instrumental variable analysis and attribution of causality to one particular intermediate. Finally, several other genes encompass the 1q21 locus, including Src homology 2 domain containing E (*SHE*), and Tudor domain containing 10 (*TDRD10*), but have been ruled out to play a role at this locus (33).

At the identified chromosome 2 locus the lead SNP, rs6734238, is intergenic and has also been associated with circulating CRP and fibrinogen levels (40,41,62). The nearest genes to this locus are the interleukin 1 family member 10 (*IL1F10*, distance = 7.6 kb, currently known as *IL-38*) and interleukin 1 receptor antagonist (*IL1RN*, distance = 34.4 kb). *IL1F10*/*IL-38* and *IL1RN* variants (rs6759676 and rs4251961) in partial LD with the lead SNP (r^2_{LD}:0.10 and 0.61) have been recently reported to be protective against the development of insulin resistance (63). This further supports the molecular mechanisms behind IL-6-mediated insulin secretion via glucagon-like peptide 1 (GLP-1) (64) contributing to type 2 diabetes (T2D) pathophysiology. For IL-6 specifically, it has been found that synthesis increases when dendritic cells are stimulated by bacterial lipopolysaccharides (LPS) in the presence of *IL1F10* (65). *IL-1RN* is another member of the interleukin 1 cytokine family, with suggestive evidence for involvement in determining IL-6 levels in the blood. One study found significant associations of *IL-1RN* rs4251961 with plasma CRP and IL-6 levels, albeit not independently replicated and not genome-wide significant ($P = 1 \times 10^{-4}$ and $P = 0.004$) (66). Our lead SNP was not in high LD ($r^2 < 0.8$) with variants in either neighboring genes and therefore in conjunction with its intergenic position, identifying a causal variant in this locus remains non-trivial.

The 6p21 rs660895, which was identified, resides within the HLA region, which forms one of the most complex genomic regions to study due to its large LD blocks and sequence diversity. This region has some population substructure in Europeans, which may have influenced the results; however, (1) each cohort population substructure adjustment was applied, followed by genomic correction for overall discovery stage meta-analyses. Thus, we reduced the chances that the population substructure may have had on this locus. The nearest genes to the index SNP, *HLA-DRB1* (distance = 19.8 kb) and *HLA-DQA1* (distance = 27.8 kb) are both histocompatibility complex genes encoding proteins that form cell surface complexes for certain immune system cells helping in antigen presentation to trigger an immune response. It is noteworthy that variations at this locus code for antigen-presenting complexes (APCs), which have been previously associated with diseases having a dysfunctional immune system; while we report for the first time that there exists also a strong association of this locus with circulating cytokine levels. Therefore, the association of this locus with the disease may corroborate through its effect through IL6 levels. One high-LD SNP (rs9272422, $r^2 = 0.82$ with our index SNP, rs660895) residing in the promoter region of *HLA-DQA1* support this hypothesis and has been identified previously for systemic lupus erythematosus (67) and ulcerative colitis (68). rs660895*G allele is associated with increased risk of RA in Europeans and Asians (42), IgA nephropathy in Asians (43), while the decreased risk of ulcerative colitis and IBD. (44)

Various studies aimed to identify genetic variation underlying levels of IL-6 (22–26) have found genome-wide significant associations in the *IL-6R* and *ABO* genes. The study performed by Shah and colleagues (33) found suggestive evidence (non-GWS; $P = 3.8 \times 10^{-6}$, respectively) for additional loci, including *ABO*, *BUD13*, *TRIB3* and *SEZ6L*, none of which replicate in the current study ($P_{discovery} > 0.05$) indicating that these might be false-positive findings due to low sample size (~7800) or loci with sex-specific effects (associations based on women dominant population) or due to technical shortcomings with measurement assay (*ABO* locus).

It is surprising that even with increased statistical power ($n_{discovery} = 52\,654$; $n_{replication} = 14\,774$) in the current study (compared to the previous IL6 GWAS) (33), we could identify three genetic loci (1q21, 2q14 and 6p21) accounting for ~1% of the genetic variance for circulating IL-6 levels. According to the current estimates, the heritability levels for IL6 levels range between 15 and 61%, suggesting that an enormous increase in sample sizes would be required to identify additional variants explaining this remaining heritability. Multiple explanations for this so-called missing heritability phenomenon have been proposed in the past, which can be sought in rare or low frequency coding variants as observed for a similar metabolic quantitative trait by us (69) or can be explained by non-additive effects, which may cause inflated estimates of heritability. Plausible evidence for other sources of unexplained heritability that have been found are epigenetic changes, and haplotypes of common SNPs.

Collectively, our results provided additional insights into the biology of circulating IL-6. We identify new loci, limited by common variants in the Hap Map Reference panel. Albeit this is comparable to the 1000 genomes reference panel (70) but narrower compared to some newly available panels that show greater variant coverage in numbers and frequency range. Future studies are recommended to aim for identification of additional common but also rare variants, by firstly using richer imputation panels, such as UK10K project or the Haplotype Reference Consortium, a strategy that holds great promise, and secondly by making use of genetically isolated populations. Thirdly, we would like to stress the importance of phenotype harmonization. As we identified genome-wide variants in the *ABO* locus, in four studies participating in the discovery, but not in the remaining 22 cohorts, there is a strong indication that this locus may be assay specific. However, a proper demonstration of this hypothesis would require further testing, including repeating the GWAS in *ABO*-positive cohorts using a different IL-6 assay. Indeed it is emerging that the *ABO* locus has pleiotropic effects on many different traits and diseases (71), which would suggest a more thorough analysis before disregarding this signal. Also, conventionally increasing sample sizes without correction for population substructures may raise heterogeneity within populations (72), likely concealing the SNPs that affect particular subgroups. Future specific studies should counter the widely held assumption of unconditional risk alleles of complex traits and focus on the importance of studying more homogenous subgroups to, for example, investigate the age-dependent effect of genetic variants (73,74). Here, while further exploring the pleiotropic effect of IL-6-related variants, we identified phenotypes differentially regulated by diverse variants in the 1q21 locus. Biologic systems are dynamic complex networks and are evolving through lifespan and investigating the interrelationships existing between phenotypes as well as between genetic variations and phenotypic variations has the potential for uncovering the complex mechanisms. This is the case here for IL-6 and tailored methodologies should be devoted to the study of such traits, hopefully resulting in clinically significant breakthroughs. Future collaborative efforts therefore should strive to use well-calibrated assays, z-standardized protocols for sample handling, and processing (75), though this will be difficult to achieve in practice. Lastly, we have attempted to perform formal association-based causal analysis to identify the likely causal loci, using the DEPICT approach; unfortunately, instead with only 2 novel GWS findings, our analyses were underpowered and thus not included. We also mine the gene expression and eQTL data for the identified SNPs using established databases; however, we were unable control for random co-localization signals or other confounders as we had limited access to summary level data. In conclusion, we identify two novel common genetic variants associated with circulating IL-6 levels that may influence the pathophysiology of complex cardio-metabolic, psychiatric and immunological traits, among individuals of European ancestry. This is a step further towards unravelling new biological pathways and potential therapeutic targets that can be developed for the IL-6-related disorders, while suggesting looking deeper into the genome for coding variants (rare and common) having larger individual effects (Figs 1 and 2).

Material and Methods

Discovery stage

Study populations. The overall study design (Supplementary Material, Fig. S1) involved the discovery cohorts with 53 893 individuals. After overlapping individuals with available genotype and phenotype data, the discovery stage included 52 654 individuals from 26 cohorts of European ancestry listed under Supplementary Material, Table S8, described in Supplementary Text S1 and study summary characteristics in Supplementary Material, Table S9. Only population-based samples or healthy controls from case–control studies were included in the final analyses.

Serum IL-6 measurements. Each study typically collected venous blood samples stored below −80°C until the time of measurement using various types of immunoassays and expressed as pg/ml as presented in Supplementary Material, Table S10. The trait transformation and phenotype data quality control (QC) were presented by Supplementary Material, Text S3 (Supplementary Material, Text S3.1 and S3.2). In brief, participating cohorts have checked for the percentage of missingness in IL6 measurements and evaluated for indices of QC (Supplementary Material, Text S3.2), yielded the final number of participants with available validated IL6 levels, of whom those with available genotype data were included in the study as characterized in Supplementary Material, Table S9 and in Supplementary Material, Text S1 and S2.

Genotyping and imputation. Each participating cohort performed genome-wide genotyping using a variety of genotyping platforms and applied a predefined quality control (QC) of genotype data (Supplementary Material, Table S11) followed by performing imputation of non-genotyped genetic variants, on the backbone of haplotypes inferred from the Hapmap Phase II reference panel (NCBI Build 36), and using statistical software such as IMPUTE (75), MACH, Minimac (76) or BIMBAM (77) (Supplementary Material, Table S11). Each cohort was recommended a set of general SNP quality filters including MAF < 0.01; Hardy Wienberg Equilibrium (HWE) $P \leq 10^{-6}$; imputation quality $r^2 \leq 0.3$; and genotyping call rate < 0.95 (Supplementary Material, Fig. S1). Once we received summary results from each participating study, we ran a series of QC checks. Firstly, these included a set standard checks, including the imputation quality filters (basis the imputation program used and/or $r^2_{imputation} < 0.3$ were excluded), and then checks for genomic inflation (quantile–quantile or QQ plots). We adapted filter thresholds per cohort to reduce any observed deviation from the null while missing SNP loss due to the QC process. Finally, ~2.45 million (2 454 025) common SNPs were part of the discovery meta-analysis.

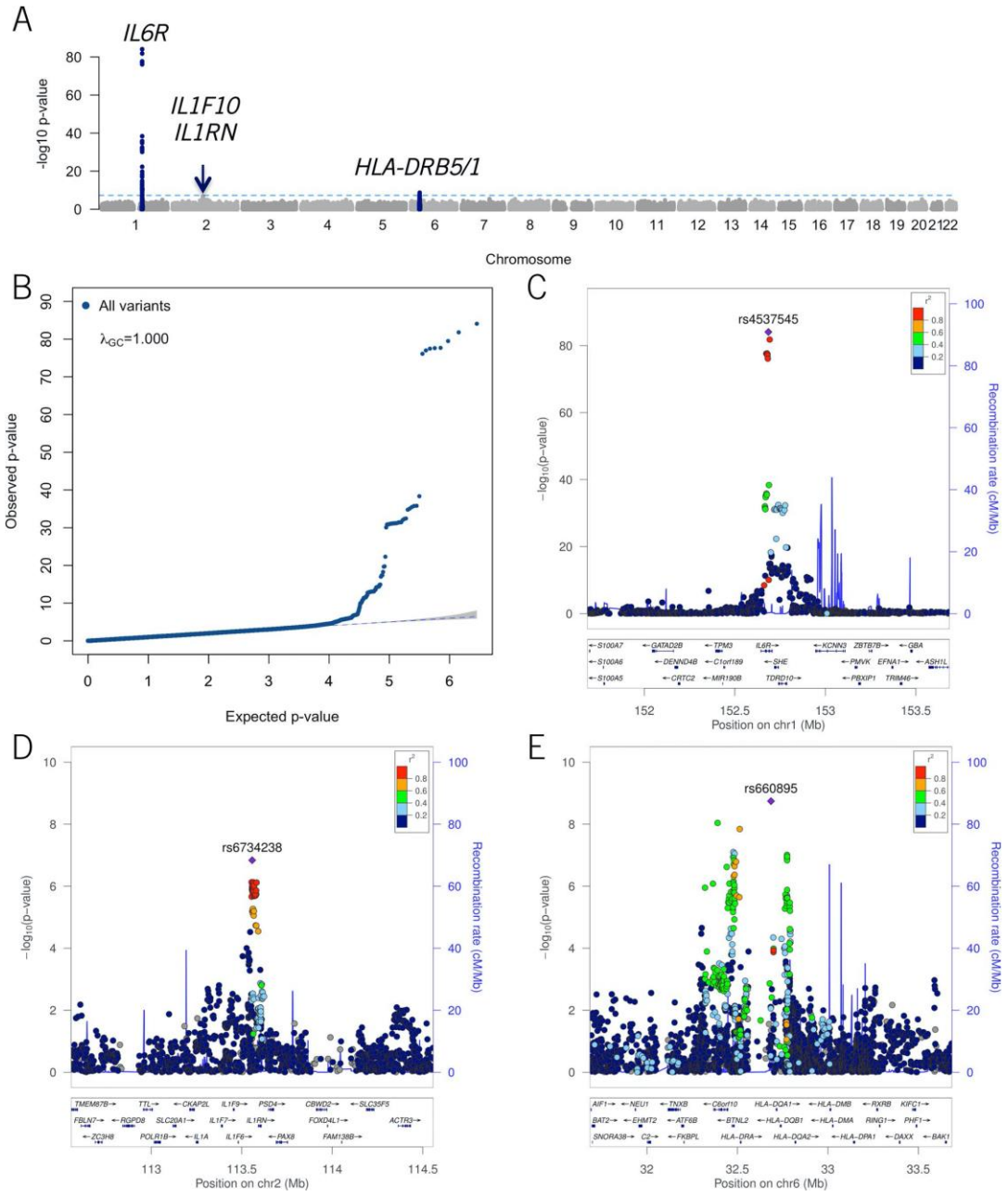

Figure 1. Manhattan, QQ and LocusZoom plots of the discovery GWAS meta-analyses. (**A**) Manhattan plot showing the association of SNPs with IL-6. Loci coloured in red or blue, three in total, represent those for which the lead SNPs reached genome-wide significance ($P = 5 \times 10^{-8}$). Horizontal axis: relative genomic position of variants on the genome, vertical axis: −log10 *P*-value of each SNP; (**B**) Quantile–quantile plot for *P*-values obtained from the meta-analysis. The horizontal and vertical axes represents the expected distribution of −log10 (*P*-values) under the null hypothesis of no association, whereas the vertical axis shows the observed −log10 (*P*-values). The blue dashed line represents the null, and λ_{GC} value represents the genomic inflation factor lambda. Each data point represents the observed versus the expected *P*-value of a variant included in the association analyses; (**C–E**) Regional association plots for each of the three genome-wide significant loci, 1q21, 2q14 and 6p21, respectively. Pairwise LD (r^2) with the lead SNP is indicated following a color-coded scale. Horizontal axis: relative genomic position of variants within the locus, vertical axis: −log10 *P*-value of each SNP.

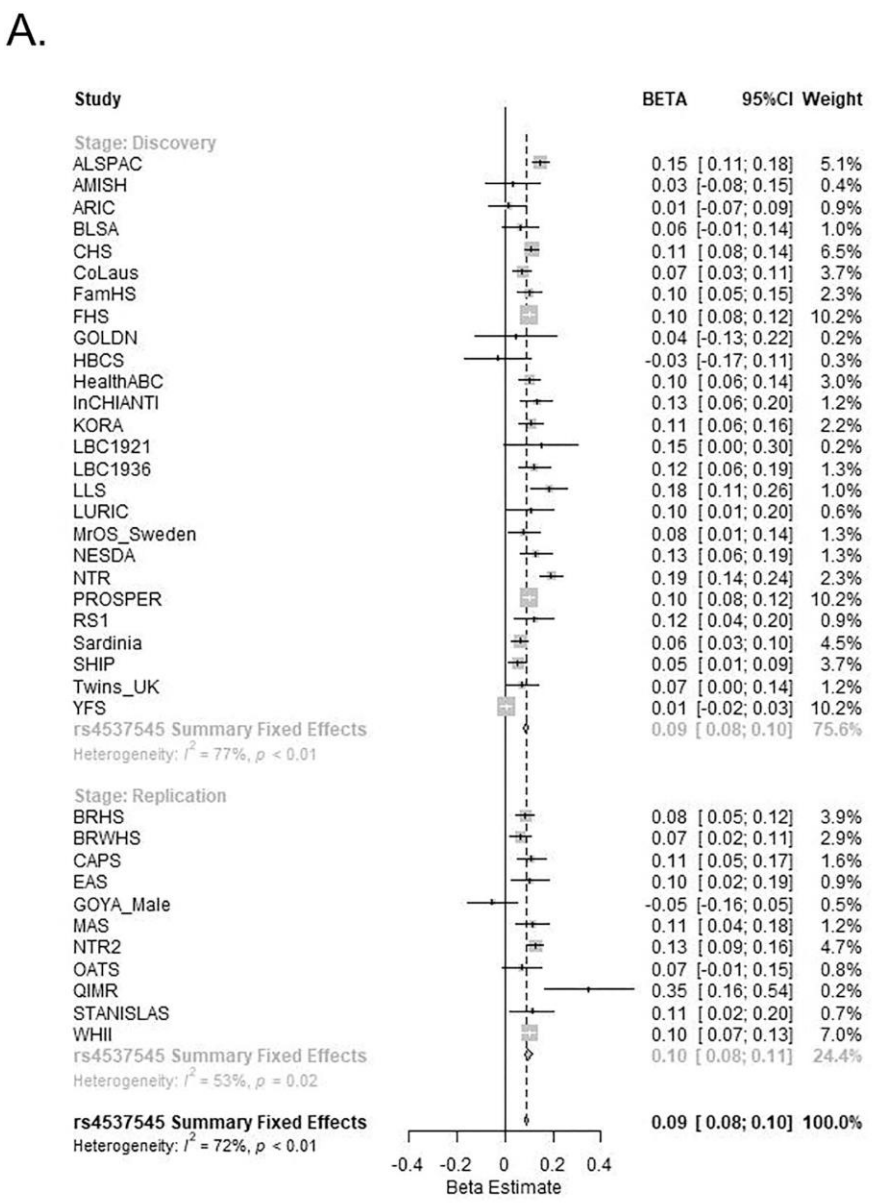

Figure 2. Combined discovery and replication forest plots for the GWAS Index SNPs. Forrest plots for (**A**) *IL6R* rs4537545 (chr. 1q21), (**B**) *IL1RN* rs6734238 (chr. 2q14), (**C**) *HLA-DRB5* rs660895 (chr. 6p21) with discovery, replication and combined effect estimates, 95% CI and weights based on the fixed effects inverse variance meta-analyses.

Statistical methods

GWAS *analysis*. Each study conducted an independent GWAS analysis between SNPs and natural log-transformed values of serum IL-6 levels following a predefined analysis plan (Supplementary Material, Methods S4). Association analyses were conducted using linear regression model, or linear mixed effect models to account for familial correlation when warranted, with additive genetic effects, accounting for imputation uncertainty while adjusting for age, sex, population substructure (through study-specific principal components) and/or study-specific site, when necessary. GWAS summary result obtained from each cohort underwent a series of QC checks using the QCGWAS package in R (78) (Supplementary Material, Text S3 and Supplementary Material, Fig. S1). Being aware of the potential false-positive

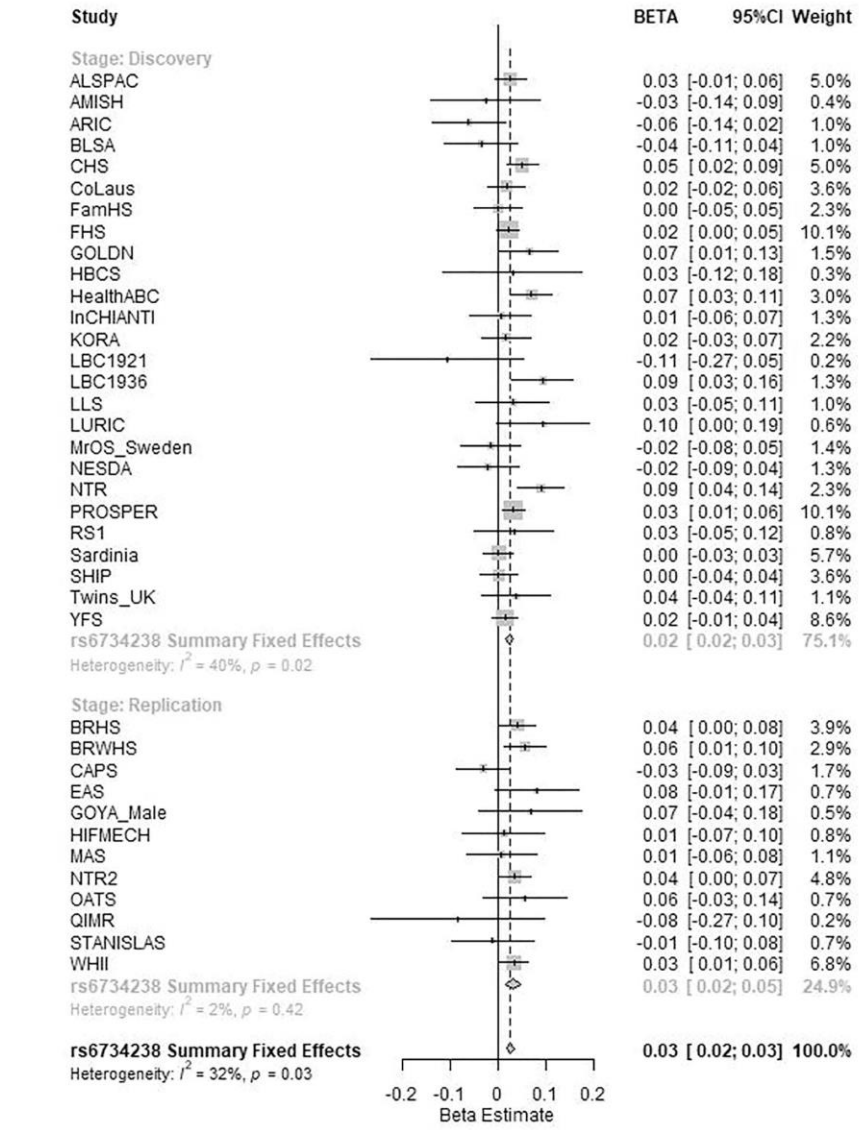

Figure 2. Continued.

association in the *ABO* region on chromosome 9 (28,30), while using an R&D systems high-sensitivity assay kit to measure IL6 levels (R&D systems, Minneapolis, MN, USA), four (out of 22) discovery cohorts that observed genome-wide significant results in the *ABO* locus were asked to rerun the GWAS analysis conditional on the top *ABO* SNP (i.e. rs8176704) before including them in the final discovery meta-analysis (Supplementary Material, Text S3.3).

Discovery GWAS meta-analyses. Individual GWAS results from 26 European studies were meta-analyzed using the inverse variance weighted, fixed-effects method as implemented in GWAMA

C.

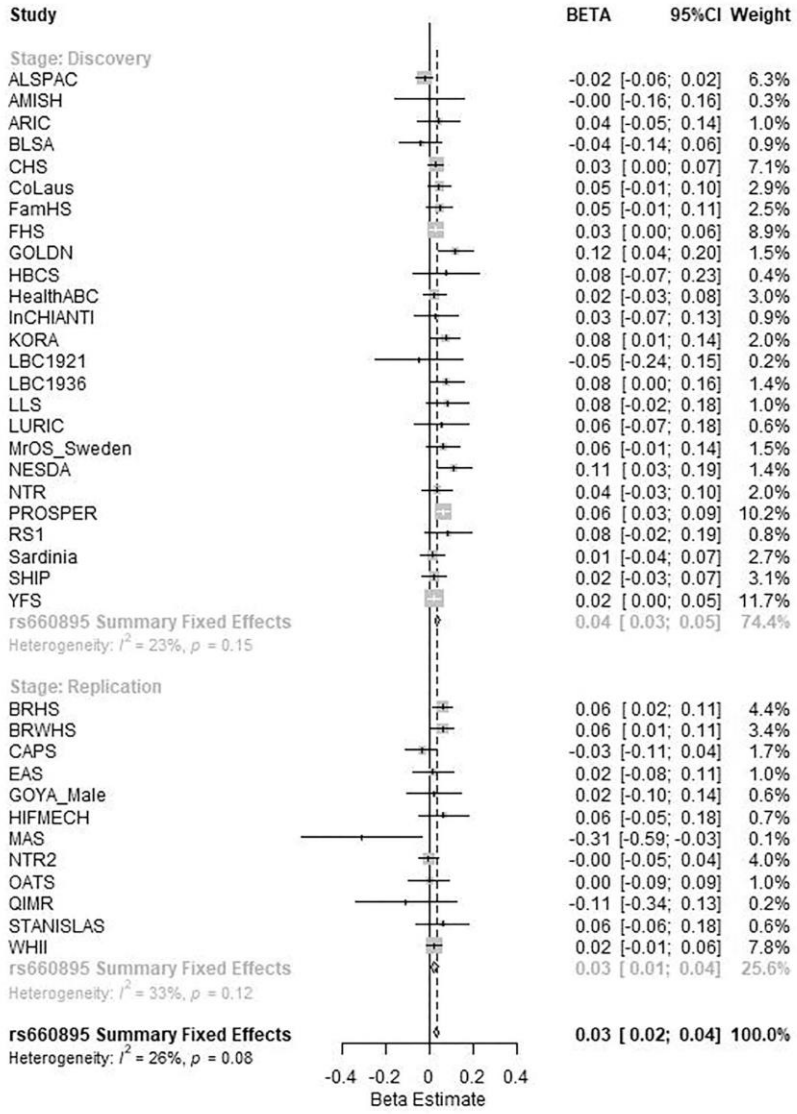

Figure 2. Continued.

while applying the double genomic control (GC) correction for population stratification, i.e. first to each study individually and subsequently also to the pooled results after meta-analysis (79).

Regional association plots for the discovered loci were generated through the LocusZoom (78) tool. We used the SNAP tool (80) to perform the pairwise LD checks (HapMap release 22 data) and to verify low LD with secondary signals. All SNPs selected for the replication stage had to fulfill the following criteria: (1) having an association $P_{discovery} \leq 1 \times 10^{-5}$ and being in very low LD with the index SNP ($r^2 < 0.2$) and (2) available in at least 50% of study cohorts.

Replication and combined meta-analysis

Study population, phenotyping and QC. The overall study (Supplementary Material, Fig. S1) comprised 15 785 individuals for replication. After removing individuals with missing data, the replication analyses were performed using a combination

of *in silico* and *de novo* genotyping in 14 774 individuals from 12 cohorts of European ancestry as described in Supplementary Material, Text S2. Similar QC (Supplementary Material, Text S3 and Supplementary Material, Table S11) and statistical checks were made as in the discovery stage.

Venous blood samples (serum or plasma) were collected and stored at −80°C. Serum/plasma IL-6 levels (pg/ml) were measured using various immunoassay methods described in Supplementary Material, Table S10. Each cohort tested the selected SNPs using the same statistical model as for the discovery association analyses (Supplementary Material, Fig. S1). Effect size estimates of all replication SNPs from each replication study were compared with the effect size estimates from the discovery meta-analyses. When effect sizes from individual cohorts did not align, we excluded these cohorts from the replication meta-analyses ($n_{\text{cohorts}} = 3$). To account for the inter-study assay differences insensitivity, we combined results across the replication studies using a fixed effect sample size weighted Z-score meta-analysis as implemented in the METAL package (https://genome.sph.umich.edu/wiki/METAL) (81).

The summary GWAS meta-analyses result from the discovery and replication stages were then used to perform a combined (discovery + replication) GWAS meta-analysis using a sample size weighted Z-score method. Test for heterogeneity was also performed as part of the meta-analysis package using METAL where I^2 statistic denoting the percentage of variation across studies was estimated ($I^2 = 100\% \times (Q - df)/Q$) where Q is the Chi-Square statistic. Significance for heterogeneity was denoted by the heterogeneity (or Het_P) P values. Variants that were significant in the replication meta-analysis at $P < 0.05$ and had an overall $P_{\text{combined}} < 5 \times 10^{-8}$ in the combined meta-analysis were considered statistically GWS. SNPs within the range of 1 Mb (or 10^6 bases) on either side of the most significant (i.e. index) SNP (with LD, $r^2 > 0.25$) were considered part of the same locus, whereas those in low LD ($r^2 < 0.25$) were tested if they were independent using conditional analysis.

Conditional analysis. We performed an approximate joint conditional analysis to identify distinct signals in a specific chromosomal region as implemented in GCTA (82) using high-quality genome-wide genotyped/imputed reference data from two studies (NEtherlands Study of Depression and Anxiety (NESDA) from the Netherlands and/or Genetics of Obesity in Young Adults (GOYA) from Denmark) to estimate linkage disequilibrium (LD) (83) between SNPs.

Conditional analysis for identification of independent signals was performed on GWS SNPs (±1 Mb to the index SNP and having low LD, $r^2 < 0.2$ with the index SNP) using summary statistics from the discovery GWAS meta-analysis data (—COJO option in GCTA) after confirming the GWS loci from the combined meta-analysis.

Heritability estimates. We approximated the variance explained by all distinct lead SNPs from the meta-analysis using the following formula:

$$\sum_{i=1}^{n} \frac{\beta_i^2 \cdot 2 \cdot \text{EAF}_i \cdot (1 - \text{EAF}_i)}{\sigma^2\,(\text{residuals}\,(\ln(\text{IL6})))}$$

where EAF is the effect allele frequency, β_i the effect size of the individual variants, and n is the total number of lead variants. The current formula may overestimate the variance to a small extent as some level of SNP correlation was existent (LD $r^2 < 0.25$). The variance of the residuals of ln (IL-6) was calculated using data from the NESDA cohort ($n = 2517$). The total common SNP heritability of serum IL-6 levels explained by all GWAS variants was estimated using the observed Z-statistics from the discovery analyses for a subset of pruned SNPs. Following the original method (SumVg) (38), we pruned the imputed (based on the 1000G *Phase1 Integrated Release, Version 3, 2012.04.30* reference panel) genotypes of the NESDA cohort using PLINK v1.07 (84), by removing correlated SNPs at $r^2 > 0.25$ within a 100-SNP sliding window, and with a step size of 25 SNPs per forwarding slide. This resulted in a pruned set of 163 459 SNPs.

SNP mapping and functionality. We searched for variants in high LD ($r^2 > 0.8$) within a 1 Mb region on either side of the lead SNPs using 1000 Genomes sequence data (Phase1 Integrated Release, Version 3, 2012.04.30), utilizing tools available in Liftover (85), VCFtools (86) and clumping in PLINK (84). We subsequently annotated these variants using ANNOVAR (87) with the RefSeq (88) database for variant function and genic residence or distance. We used the Combined Annotation-Dependent Depletion (CADD) database to identify the functionality, i.e. deleterious, disease causal, pathogenicity, for the index SNPs.

Associations with other traits and gene expression data. Pheno Scanner v2 (89) data mining tool was used (Access date October 2020) to identify existing GWS (at $P < 5 \times 10^{-8}$) associations between IL6 identified SNPs and other traits, and gene expression/eQTLs data.

Power calculation

We used GWAs power estimator (see Web Resources) by assuming a relative risk of 1.10 (or an effect estimate of 0.10), given $N = 66\,000$, alpha (P-value) $= 5 \times 10^{-8}$ (also GCTA power calculator, Supplementary Text S3.5).

Supplementary Material

Supplementary Material is available at *HMG* online. GWAS summary statistics data is available upon request to corresponding authors T.S.A. and/or B.Z.A.

Web Resources

QCGWAS, https://cran.r-project.org/web/packages/QCGWAS/index.html

GWAMA, http://www.well.ox.ac.uk/gwama/

METAL, http://csg.sph.umich.edu//abecasis/metal/

GCTA, http://www.complextraitgenomics.com/software/gcta/

LocusZoom, http://csg.sph.umich.edu/locuszoom/

1000 Genomes, ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/

PLINK, http://pngu.mgh.harvard.edu/purcell/plink/

VCFtools, http://vcftools.sourceforge.net/

ANNOVAR, http://www.openbioinformatics.org/annovar/

PhenoScanner, http://www.phenoscanner.medschl.cam.ac.uk/phenoscanner

Power calculations: http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html

The UK Biobank Neale's lab rapid GWAS: (http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank).

Authors Contributions

The guarantors of the study are T.S.A., B.P. and B.Z.A., from conception and design to conduct of the study and acquisition of

data, data analysis and interpretation of data. T.S.A. and B.P. have written the first draft of the manuscript. All co-authors have provided important intellectual input and contributed considerably to the analyses and interpretation of the data. All authors guarantee that the accuracy and integrity of any part of the work have been appropriately investigated and resolved and all have approved the final version of the manuscript. BP had full access to the data and the corresponding authors T.S.A. and B.Z.A. had the final responsibility for the decision to submit for publication.

Acknowledgments

We acknowledge the participants of this study and the funding resources that have contributed to achieving this effort. Detailed acknowledgments can be found in Supplementary Material, Table S12.

Conflict of Interest. There is no conflict of interest from any of the participating co-authors. T.S.A. is a shareholder with Novo Nordisk A/S and Zealand Pharma A/S.

Funding

Novo Nordisk Foundation (NNF18OC0052457 to T.S.A.). Detailed Funding statements from participating study cohorts can be found in Supplementary Material, Table S12.

References

1. Kishimoto, T. (2010) IL-6: from its discovery to clinical applications. *Int. Immunol.*, **22**, 347–352.
2. Nishimoto, N. and Kishimoto, T. (2006) Interleukin 6: from bench to bedside. *Nat. Clin. Pract. Rheumatol.*, **2**, 619–626.
3. Brocker, C., Thompson, D., Matsumoto, A., Nebert, D.W. and Vasiliou, V. (2010) Evolutionary divergence and functions of the human interleukin (IL) gene family. *Hum. Genomics*, **5**, 30–55.
4. Gelinas, L., Falkenham, A., Oxner, A., Sopel, M. and Legare, J.F. (2011) Highly purified human peripheral blood monocytes produce IL-6 but not TNFalpha in response to angiotensin II. *J. Renin-Angiotensin-Aldosterone Syst.*, **12**, 295–303.
5. Kitani, A., Hara, M., Hirose, T., Harigai, M., Suzuki, K., Kawakami, M., Kawaguchi, Y., Hidaka, T., Kawagoe, M. and Nakamura, H. (1992) Autostimulatory effects of IL-6 on excessive B cell differentiation in patients with systemic lupus erythematosus: analysis of IL-6 production and IL-6R expression. *Clin. Exp. Immunol.*, **88**, 75–83.
6. Li, T. and He, S. (2006) Induction of IL-6 release from human T cells by PAR-1 and PAR-2 agonists. *Immunol. Cell Biol.*, **84**, 461–466.
7. Ng, E.K., Panesar, N., Longo, W.E., Shapiro, M.J., Kaminski, D.L., Tolman, K.C. and Mazuski, J.E. (2003) Human intestinal epithelial and smooth muscle cells are potent producers of IL-6. *Mediat. Inflamm.*, **12**, 3–8.
8. Fain, J.N. (2006) Release of interleukins and other inflammatory cytokines by human adipose tissue is enhanced in obesity and primarily due to the nonfat cells. *Vitam. Horm.*, **74**, 443–477.
9. Podor, T.J., Jirik, F.R., Loskutoff, D.J., Carson, D.A. and Lotz, M. (1989) Human endothelial cells produce IL-6. Lack of responses to exogenous IL-6. *Ann. N. Y. Acad. Sci.*, **557**, 374–385 discussion 386-377.
10. Sanchez, C., Gabay, O., Salvat, C., Henrotin, Y.E. and Berenbaum, F. (2009) Mechanical loading highly increases IL-6 production and decreases OPG expression by osteoblasts. *Osteoarthr. Cartil.*, **17**, 473–481.
11. Haddy, N., Sass, C., Maumus, S., Marie, B., Droesch, S., Siest, G., Lambert, D. and Visvikis, S. (2005) Biological variations, genetic polymorphisms and familial resemblance of TNF-alpha and IL-6 concentrations: STANISLAS cohort. *Eur. J. Hum. Genet.*, **13**, 109–117.
12. Madhok, R., Crilly, A., Watson, J. and Capell, H.A. (1993) Serum interleukin 6 levels in rheumatoid arthritis: correlations with clinical and laboratory indices of disease activity. *Ann. Rheum. Dis.*, **52**, 232–234.
13. de Benedetti, F., Massa, M., Robbioni, P., Ravelli, A., Burgio, G.R. and Martini, A. (1991) Correlation of serum interleukin-6 levels with joint involvement and thrombocytosis in systemic juvenile rheumatoid arthritis. *Arthritis Rheum.*, **34**, 1158–1163.
14. Ahluwalia, T.S., Kilpelainen, T.O., Singh, S. and Rossing, P. (2019) Editorial: novel biomarkers for type 2 diabetes. *Front Endocrinol (Lausanne)*, **10**, 649.
15. Cesari, M., Penninx, B.W., Newman, A.B., Kritchevsky, S.B., Nicklas, B.J., Sutton-Tyrrell, K., Rubin, S.M., Ding, J., Simonsick, E.M., Harris, T.B. and Pahor, M. (2003) Inflammatory markers and onset of cardiovascular events: results from the health ABC study. *Circulation*, **108**, 2317–2322.
16. Qu, D., Liu, J., Lau, C.W. and Huang, Y. (2014) IL-6 in diabetes and cardiovascular complications. *Br. J. Pharmacol.*, **171**, 3595–3603.
17. Mauer, J., Denson, J.L. and Bruning, J.C. (2015) Versatile functions for IL-6 in metabolism and cancer. *Trends Immunol.*, **36**, 92–101.
18. Mucha, S., Baurecht, H., Novak, N., Rodríguez, E., Bej, S., Mayr, G., Emmert, H., Stölzl, D., Gerdes, S., Jung, E.S. *et al.* (2020) Protein-coding variants contribute to the risk of atopic dermatitis and skin-specific gene expression. *J. Allergy Clin. Immunol.*, **145**, 1208–1218.
19. Zhang, C., Wu, Z., Zhao, G., Wang, F. and Fang, Y. (2016) Identification of IL6 as a susceptibility gene for major depressive disorder. *Sci. Rep.*, **6**, 31264.
20. Calabrese, L.H. and Rose-John, S. (2014) IL-6 biology: implications for clinical targeting in rheumatic disease. *Nat. Rev. Rheumatol.*, **10**, 720–727.
21. Scheinecker, C., Smolen, J., Yasothan, U., Stoll, J. and Kirkpatrick, P. (2009) Tocilizumab. *Nat. Rev. Drug Discov.*, **8**, 273–274.
22. Revez, J.A., Bain, L.M., Watson, R.M., Towers, M., Collins, T., Killian, K.J., O'Byrne, P.M., Gauvreau, G.M., Upham, J.W. and Ferreira, M.A. (2019) Effects of interleukin-6 receptor blockade on allergen-induced airway responses in mild asthmatics. *Clin. Transl. Immunol.*, **8**, e1044.
23. Yazici, Y., Curtis, J.R., Ince, A., Baraf, H., Malamet, R.L., Teng, L.L. and Kavanaugh, A. (2012) Efficacy of tocilizumab in patients with moderate to severe active rheumatoid arthritis and a previous inadequate response to disease-modifying antirheumatic drugs: the ROSE study. *Ann. Rheum. Dis.*, **71**, 198–205.
24. Yokota, S., Imagawa, T., Mori, M., Miyamae, T., Aihara, Y., Takei, S., Iwata, N., Umebayashi, H., Murata, T., Miyoshi, M. *et al.* (2008) Efficacy and safety of tocilizumab in patients with systemic-onset juvenile idiopathic arthritis: a randomised, double-blind, placebo-controlled, withdrawal phase III trial. *Lancet*, **371**, 998–1006.
25. Gupta, S., Wang, W., Hayek, S.S., Chan, L., Mathews, K.S., Melamed, M.L., Brenner, S.K., Leonberg-Yoo, A., Schenck, E.J.,

Radbel, J. *et al.* (2021) Association between early treatment with tocilizumab and mortality among critically ill patients with COVID-19. *JAMA Intern. Med.*, **181**, 41–51.
26. Worns, M.A., Victor, A., Galle, P.R. and Hohler, T. (2006) Genetic and environmental contributions to plasma C-reactive protein and interleukin-6 levels–a study in twins. *Genes Immun.*, **7**, 600–605.
27. Sas, A.A., Jamshidi, Y., Zheng, D., Wu, T., Korf, J., Alizadeh, B.Z., Spector, T.D. and Snieder, H. (2012) The age-dependency of genetic and environmental influences on serum cytokine levels: a twin study. *Cytokine*, **60**, 108–113.
28. Neijts, M., van Dongen, J., Kluft, C., Boomsma, D.I., Willemsen, G. and de Geus, E.J. (2013) Genetic architecture of the pro-inflammatory state in an extended twin-family design. *Twin Res. Hum. Genet.*, **16**, 931–940.
29. Amaral, W.Z., Krueger, R.F., Ryff, C.D. and Coe, C.L. (2015) Genetic and environmental determinants of population variation in interleukin-6, its soluble receptor and C-reactive protein: insights from identical and fraternal twins. *Brain Behav. Immun.*, **49**, 171–181.
30. Melzer, D., Perry, J.R., Hernandez, D., Corsi, A.M., Stevens, K., Rafferty, I., Lauretani, F., Murray, A., Gibbs, J.R., Paolisso, G. *et al.* (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet.*, **4**, e1000072.
31. Naitza, S., Porcu, E., Steri, M., Taub, D.D., Mulas, A., Xiao, X., Strait, J., Dei, M., Lai, S., Busonero, F. *et al.* (2012) A genome-wide association scan on the levels of markers of inflammation in Sardinians reveals associations that underpin its complex regulation. *PLoS Genet.*, **8**, e1002480.
32. Comuzzie, A.G., Cole, S.A., Laston, S.L., Voruganti, V.S., Haack, K., Gibbs, R.A. and Butte, N.F. (2012) Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One*, **7**, e51954.
33. Shah, T., Zabaneh, D., Gaunt, T., Swerdlow, D.I., Shah, S., Talmud, P.J., Day, I.N., Whittaker, J., Holmes, M.V., Sofat, R. *et al.* (2013) Gene-centric analysis identifies variants associated with interleukin-6 levels and shared pathways with other inflammation markers. *Circ. Cardiovasc. Genet.*, **6**, 163–170.
34. Sidore, C., Busonero, F., Maschio, A., Porcu, E., Naitza, S., Zoledziewska, M., Mulas, A., Pistis, G., Steri, M., Danjou, F. *et al.* (2015) Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat. Genet.*, **47**, 1272–1281.
35. Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M. and Crawford, D.C. (2010) PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*, **26**, 1205–1210.
36. Pendergrass, S.A., Brown-Gentry, K., Dudek, S.M., Torstenson, E.S., Ambite, J.L., Avery, C.L., Buyske, S., Cai, C., Fesinmeyer, M.D., Haiman, C. *et al.* (2011) The use of phenome-wide association studies (Phe WAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genet. Epidemiol.*, **35**, 410–422.
37. Pendergrass, S.A., Brown-Gentry, K., Dudek, S., Frase, A., Torstenson, E.S., Goodloe, R., Ambite, J.L., Avery, C.L., Buyske, S., Bůžková, P. *et al.* (2013) Phenome-wide association study (Phe WAS) for detection of pleiotropy within the population architecture using genomics and epidemiology (PAGE) network. *PLoS Genet.*, **9**, e1003087.
38. So, H.C., Li, M. and Sham, P.C. (2011) Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet. Epidemiol.*, **35**, 447–456.
39. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. and Kircher, M. (2019) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.*, **47**, D886–D894.
40. Sabater-Lleal, M., Huang, J., Chasman, D., Naitza, S., Dehghan, A., Johnson, A.D., Teumer, A., Reiner, A.P., Folkersen, L., Basu, S. *et al.* (2013) Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation*, **128**, 1310–1324.
41. Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P. *et al.* (2011) Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*, **123**, 731–738.
42. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. *et al.* (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376–381.
43. Yu, X.Q., Li, M., Zhang, H., Low, H.Q., Wei, X., Wang, J.Q., Sun, L.D., Sim, K.S., Li, Y., Foo, J.N. *et al.* (2011) A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat. Genet.*, **44**, 178–182.
44. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T. *et al.* (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.*, **47**, 979–986.
45. Elliott, P., Chambers, J.C., Zhang, W., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund, P., Engert, J.C., Bennett, D. *et al.* (2009) Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA*, **302**, 37–48.
46. Ferreira, R.C., Freitag, D.F., Cutler, A.J., Howson, J.M., Rainbow, D.B., Smyth, D.J., Kaptoge, S., Clarke, P., Boreham, C., Coulson, R.M. *et al.* (2013) Functional IL6R 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genet.*, **9**, e1003444.
47. Stone, K., Woods, E., Szmania, S.M., Stephens, O.W., Garg, T.K., Barlogie, B., Shaughnessy, J.D., Jr., Hall, B., Reddy, M., Hoering, A. *et al.* (2013) Interleukin-6 receptor polymorphism is prevalent in HIV-negative Castleman disease and is associated with increased soluble interleukin-6 receptor levels. *PLoS One*, **8**, e54610.
48. Rafiq, S., Frayling, T.M., Murray, A., Hurst, A., Stevens, K., Weedon, M.N., Henley, W., Ferrucci, L., Bandinelli, S., Corsi, A.M. *et al.* (2007) A common variant of the interleukin 6 receptor (IL-6r) gene increases IL-6r and IL-6 levels, without other inflammatory effects. *Genes Immun.*, **8**, 552–559.
49. Galicia, J.C., Tai, H., Komatsu, Y., Shimada, Y., Akazawa, K. and Yoshie, H. (2004) Polymorphisms in the IL-6 receptor (IL-6R) gene: strong evidence that serum levels of soluble IL-6R are genetically influenced. *Genes Immun.*, **5**, 513–516.
50. Hayashida, K., Bartlett, A.H., Chen, Y. and Park, P.W. (2010) Molecular and cellular mechanisms of ectodomain shedding. *Anat Rec (Hoboken)*, **293**, 925–937.
51. Collaboration, I.R.G.C.E.R.F, Sarwar, N., Butterworth, A.S., Freitag, D.F., Gregson, J., Willeit, P., Gorman, D.N., Gao, P., Saleheen, D., Rendon, A. *et al.* (2012) Interleukin-6 receptor

pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. *Lancet*, **379**, 1205–1213.
52. Interleukin-6 Receptor Mendelian Randomisation Analysis, C, Swerdlow, D.I., Holmes, M.V., Kuchenbaecker, K.B., Engmann, J.E., Shah, T., Sofat, R., Guo, Y., Chung, C., Peasey, A. *et al.* (2012) The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet*, **379**, 1214–1224.
53. Consortium, C.A.D., Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T.L., Thompson, J.R., Ingelsson, E., Saleheen, D., Erdmann, J. *et al.* (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.*, **45**, 25–33.
54. Schnabel, R.B., Kerr, K.F., Lubitz, S.A., Alkylbekova, E.L., Marcus, G.M., Sinner, M.F., Magnani, J.W., Wolf, P.A., Deo, R., Lloyd-Jones, D.M. *et al.* (2011) Large-scale candidate gene analysis in whites and African Americans identifies IL6R polymorphism in relation to atrial fibrillation: the National Heart, Lung, and Blood Institute's candidate Gene Association resource (CARe) project. *Circ. Cardiovasc. Genet.*, **4**, 557–564.
55. Hawkins, G.A., Robinson, M.B., Hastie, A.T., Li, X., Li, H., Moore, W.C., Howard, T.D., Busse, W.W., Erzurum, S.C., Wenzel, S.E. *et al.* (2012) The IL6R variation asp (358) ala is a potential modifier of lung function in subjects with asthma. *J. Allergy Clin. Immunol.*, **130**, 510, e511–515.
56. Harrison, S.C., Smith, A.J., Jones, G.T., Swerdlow, D.I., Rampuri, R., Bown, M.J., Aneurysm, C., Folkersen, L., Baas, A.F., de Borst, G.J. *et al.* (2013) Interleukin-6 receptor pathways in abdominal aortic aneurysm. *Eur. Heart J.*, **34**, 3707–3716.
57. Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K. *et al.* (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.*, **44**, 1336–1340.
58. Scheller, J., Ohnesorge, N. and Rose-John, S. (2006) Interleukin-6 trans-signalling in chronic inflammation and cancer. *Scand. J. Immunol.*, **63**, 321–329.
59. Chalaris, A., Schmidt-Arras, D., Yamamoto, K. and Rose-John, S. (2012) Interleukin-6 trans-signaling and colonic cancer associated with inflammatory bowel disease. *Dig. Dis.*, **30**, 492–499.
60. Enevold, C., Baslund, B., Linde, L., Josephsen, N.L., Tarp, U., Lindegaard, H., Jacobsen, S. and Nielsen, C.H. (2014) Interleukin-6-receptor polymorphisms rs12083537, rs2228145, and rs4329505 as predictors of response to tocilizumab in rheumatoid arthritis. *Pharmacogenet. Genomics*, **24**, 401–405.
61. Ligthart, S., Vaez, A., Vosa, U., Stathopoulou, M.G., de Vries, P.S., Prins, B.P., Van der Most, P.J., Tanaka, T., Naderi, E., Rose, L.M. *et al.* (2018) Genome analyses of >200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. *Am. J. Hum. Genet.*, **103**, 691–706.
62. Reiner, A.P., Beleza, S., Franceschini, N., Auer, P.L., Robinson, J.G., Kooperberg, C., Peters, U. and Tang, H. (2012) Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. *Am. J. Hum. Genet.*, **91**, 502–512.
63. Herder, C., Nuotio, M.L., Shah, S., Blankenberg, S., Brunner, E.J., Carstensen, M., Gieger, C., Grallert, H., Jula, A., Kahonen, M. *et al.* (2014) Genetic determinants of circulating interleukin-1 receptor antagonist levels and their association with glycemic traits. *Diabetes*, **63**, 4343–4359.
64. Ellingsgaard, H., Hauselmann, I., Schuler, B., Habib, A.M., Baggio, L.L., Meier, D.T., Eppler, E., Bouzakri, K., Wueest, S., Muller, Y.D. *et al.* (2011) Interleukin-6 enhances insulin secretion by increasing glucagon-like peptide-1 secretion from L cells and alpha cells. *Nat. Med.*, **17**, 1481–1489.
65. van de Veerdonk, F.L., Stoeckman, A.K., Wu, G., Boeckermann, A.N., Azam, T., Netea, M.G., Joosten, L.A., van der Meer, J.W., Hao, R., Kalabokis, V. *et al.* (2012) IL-38 binds to the IL-36 receptor and has biological effects on immune cells similar to IL-36 receptor antagonist. *Proc. Natl. Acad. Sci. USA.*, **109**, 3001–3005.
66. Reiner, A.P., Wurfel, M.M., Lange, L.A., Carlson, C.S., Nord, A.S., Carty, C.L., Rieder, M.J., Desmarais, C., Jenny, N.S., Iribarren, C. *et al.* (2008) Polymorphisms of the IL1-receptor antagonist gene (IL1RN) are associated with multiple markers of systemic inflammation. *Arterioscler. Thromb. Vasc. Biol.*, **28**, 1407–1412.
67. Hom, G., Graham, R.R., Modrek, B., Taylor, K.E., Ortmann, W., Garnier, S., Lee, A.T., Chung, S.A., Ferreira, R.C., Pant, P.V. *et al.* (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.*, **358**, 900–909.
68. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A. *et al.* (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, **491**, 119–124.
69. Ahluwalia, T.S., Schulz, C.A., Waage, J., Skaaby, T., Sandholm, N., van Zuydam, N., Charmet, R., Bork-Jensen, J., Almgren, P., Thuesen, B.H. *et al.* (2019) A novel rare CUBN variant and three additional genes identified in Europeans with and without diabetes: results from an exome-wide association study of albuminuria. *Diabetologia*, **62**, 292–305.
70. Apelqvist, J. (2012) Diagnostics and treatment of the diabetic foot. *Endocrine*, **41**, 384–397.
71. Pickrell, J.K., Berisa, T., Liu, J.Z., Segurel, L., Tung, J.Y. and Hinds, D.A. (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.*, **48**, 709–717.
72. Mac Rae, C.A. and Vasan, R.S. (2011) Next-generation genome-wide association studies: time to focus on phenotype? *Circ. Cardiovasc. Genet.*, **4**, 334–336.
73. Dumitrescu, L., Carty, C.L., Franceschini, N., Hindorff, L.A., Cole, S.A., Buzkova, P., Schumacher, F.R., Eaton, C.B., Goodloe, R.J., Duggan, D.J. *et al.* (2013) Post-genome-wide association study challenges for lipid traits: describing age as a modifier of gene-lipid associations in the population architecture using genomics and epidemiology (PAGE) study. *Ann. Hum. Genet.*, **77**, 416–425.
74. Kulminski, A.M., Culminskaya, I., Arbeev, K.G., Arbeeva, L., Ukraintseva, S.V., Stallard, E., Wu, D. and Yashin, A.I. (2015) Birth cohort, age, and sex strongly modulate effects of lipid risk alleles identified in genome-wide association studies. *PLoS One*, **10**, e0136319.
75. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
76. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. and Abecasis, G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955–959.
77. Servin, B. and Stephens, M. (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.*, **3**, e114.

78. van der Most, P.J., Vaez, A., Prins, B.P., Munoz, M.L., Snieder, H., Alizadeh, B.Z. and Nolte, I.M. (2014) QCGWAS: a flexible R package for automated quality control of genome-wide association results. *Bioinformatics*, **30**, 1185–1186.
79. Magi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, **11**, 288.
80. Pei, Y.F., Zhang, L., Li, J. and Deng, H.W. (2010) Analyses and comparison of imputation-based association methods. *PLoS One*, **5**, e10827.
81. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
82. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, **42**, 565–569.
83. Penninx, B.W., Beekman, A.T., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W., Assendelft, W.J. *et al.* (2008) The Netherlands study of depression and anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.*, **17**, 121–140.
84. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
85. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F. *et al.* (2006) The UCSC genome browser database: update 2006. *Nucleic Acids Res.*, **34**, D590–D598.
86. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
87. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
88. Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2005) NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.
89. Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J. *et al.* (2016) Pheno scanner: a database of human genotype-phenotype associations. *Bioinformatics*, **32**, 3207–3209.

Annex1

Detailed script for the analytical algorithm

RSGEN script

App

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
rm(list                                  =                                  ls())

source("serv72.R")
source("ui72.R")

shinyApp(ui,server)
#shinyApp(ui,server,options = list(display.mode='showcase'))
```

Server

```
if                                                          (!require("pacman"))
  install.packages("pacman")
pacman::p_load(
  "shinyjs",
  "shinythemes",
  "shinycssloaders",
  "conflicted",
  "summarytools",
  "Hmisc",
  "Formula",
  "survival",
  "rmarkdown",
  "ROCR",
  "devtools",
  "usethis",
  "VIM",
  "colorspace",
  "HardyWeinberg",
  "nnet",
  "Rsolnp",
  "mice",
  "psych",
  "missMDA",
  "factoextra",
  "FactoMineR",
  "reshape2",
  "doParallel",
  "iterators",
  "foreach",
```

```
  "caret",
  "lattice",
  "coefplot",
  "glmnet",
  "Matrix",
  "broom",
  "forcats",
  "stringr",
  "dplyr",
  "purrr",
  "readr",
  "tidyr",
  "tibble",
  "tidyverse",
  "useful",
  "gt",
  "gtsummary",
  "plotly",
  "DT",
  "ggplot2",
  "rio",
  "janitor",
  "shinydashboard",
  "shiny",
  data.table
)
conflict_prefer("select",                                        "dplyr")
conflict_prefer("filter",                                        "dplyr")
conflict_prefer("box",                                  "shinydashboard")
conflict_prefer("melt",                                        "reshape2")
conflict_prefer("dcast",                                       "reshape2")
#                              source                              functions
list.files("funs", full.names = TRUE, pattern = ".R$") %>% walk(source)
#                              default                              dataset
dataset                    <-                  (import("Dataset2.xlsx"))
##                                      dataset$id=as.character(dataset$id)
server       =       function(input,       output,       session)       {
  #                              input                                data
  shinyjs::disable("report")
  shinyjs::disable("file_pred")
  InputDataset                        <-                        reactive({
    #          read          default          or          user          data
    if                    (is.null(input$file_mtb))                      {
      dataset
    }                                else                                 {
      #                              get                              input
```

```
    infile                      <-                      input$file_mtb
    #                 read                  and                  return
    #                         clean_names(rio::import(infile$datapath))
    rio::import(infile$datapath)
    #        remove_outliers_df(clean_names(rio::import(infile$datapath)))
    #
  }
})
#                  DT                  original                  data
output$Data                <-                renderDT(InputDataset())
#        update        options        for        Y        and        ID
observe({
  lstname                  <-                  names(InputDataset())
  updateSelectInput(
    session                         =                         session,
    inputId                        =                       "SelectY",
    choices                         =                         lstname,
    selected         =         str_subset(lstname,         "stat")[1]
  )
  updateSelectInput(
    session                         =                         session,
    inputId                       =                       "SelectID",
    choices                         =                         lstname,
    selected         =         str_subset(lstname,         "id")[1]
  )
})
  #  identify  factors-  numerics-  genetics  (after  setting  ID,  Y)
toListen                        <-                        reactive({
  list(input$SelectY                ,                input$SelectID)
})
vars            <-            eventReactive(toListen(),            {
  form_fun(InputDataset(),
             resp                   =                  input$SelectY,
             id                    =                  input$SelectID)
})
observeEvent(toListen()                        ,                    {
  #                                                  req(input$SelectY)
  #                                                 req(input$SelectID)
  updateSelectInput(
    session                         =                         session,
    inputId                       =                      "SelectNUM",
    choices                       =                      vars()$nums,
    selected                       =                      vars()$nums
  )
  updateSelectInput(
    session                         =                         session,
```

```
      inputId                          =                          "SelectFAC",
      choices                          =                         vars()$facts,
      selected                          =                         vars()$facts
    )
    updateSelectInput(
      session                            =                            session,
      inputId                          =                          "SelectGEN",
      choices                         =                        vars()$genetix,
      selected                         =                        vars()$genetix
    )
  })
 #                              remove                              outliers
  InputDataset_clean          <-          eventReactive(input$go,          {
     dd=remove_outliers_df(InputDataset())
     })
  #                              fun2                                HW_MAF
  list_HW_MAF                 <-                reactive(                 {
    print("list_HW_MAF")
req(InputDataset_clean())
    list_HW_MAF                            =                          list()
    i                                   =                                  0
    for          (k          in          c(0,          1,          3))          {
      i                 =                 i                 +                 1
      list_HW_MAF[[i]]          =          fun2_HW_MAF(InputDataset_clean(),
vars()$genetix,  status=input$SelectY,  k  =  k,  HWpval=input$p_val  ,
MAF=input$MAF/100)
names(list_HW_MAF)[i]              =              paste0("k",              i)
    }
    list_HW_MAF
  })
 #                      HW                      MF                      plots
  HW_MAF_plist                          <-                        reactive({
    req(list_HW_MAF())
    par(mar=c(1,1,1,1))
    plot.new()
    par(mfrow=c(2,2))
    i                                   =                                  0
    for          (k          in          c(0,          1,          3))          {
      i                 =                 i                 +                 1
HWGenotypePlot(list_HW_MAF()[[i]][[3]],plottype=2,cex=0.7,cex.lab=1.5,
cex.axis=1.5    ,cex.main=1,    main="Homozygote    scatterplot",asp=0.5)
    }
    plist=recordPlot()
  })
  output$HW_MAF_plist                     <-                     renderPlot({
   replayPlot(                                              HW_MAF_plist())
```

```
  })
  dat3                          <-                          reactive({
    req(list_HW_MAF())
dat2=    InputDataset_clean()    %>%    select(one_of(c(input$SelectY,
        vars()$facts,        vars()$nums,        list_HW_MAF()$k3[[1]]
      )))
      #                                                          aggr
      #     Remove     variables     based     on     missing     values
      par(mar=c(1,1,1,1))
      plot.new()
    aggr_dat=    aggr(dat2,    col=c('navyblue','red'),    numbers=TRUE,
sortVars=TRUE,       labels=names(dat2),       cex.axis=.7,       gap=3,
ylab=c("Histogram         of         missing         data","Pattern"))
      aggr_plot<-                                          recordPlot()
      aggr_plot
      MVV<-as.data.frame(aggr_dat$missings)
      MVV[2]<-round(MVV[2]/dim(dat2)[1],       digits       =       2)
      MVV     <-     subset(MVV,     MVV[2]     <     input$MVV_prop/100)
      dd3    <-       dat2    %>%    select(    one_of(MVV$Variable)    )
      #         delete         rows         with         NAs>         tol
        dat22    =      dd3    %>%    rowwise()    %>%    mutate(nmis_oth    =
sum(is.na(c_across(one_of(
          c(vars()$facts,                                  vars()$nums)
        ))))       /       length(c(vars()$facts,       vars()$nums)),
        nmis_gen              =              sum(is.na(c_across(one_of(
          c(list_HW_MAF()$k3[[1]])
        ))))       /       length(c(list_HW_MAF()$k3[[1]])))       %>%
        filter(nmis_oth<=input$nmis_oth_tol/100                      &
nmis_gen<=input$nmis_gen_tol/100)    %>%    select(-c(nmis_oth,nmis_gen))
       return(list(                    dat=dat22,aggr_plot=aggr_plot))
  })
 #                              plot                              aggr
  output$NA_plots                    <-                    renderPlot({
    replayPlot(dat3()$aggr_plot)
  })
  #                                  fun3::                  imputation
 dat_model            <-            eventReactive(dat3(),            {
   req(dat3())
fun3_imp(dat3()$dat,1,                                input$SelectNUM)
  })
  #       Prepare       dataset       for       summary       and       Plots
  InputDataset_summ        <-        eventReactive(dat_model(),        {
req(dat_model)
    dt                                                                <-
      dat_model()      %>%      select(-one_of(input$SelectID))      %>%
select(one_of(
```

```
        c(
          input$SelectY,
          input$SelectFAC,
          input$SelectNUM,
          input$SelectGEN
        )
      ))       %>%       rename(Y       =       input$SelectY)       %>%
      mutate(
        across(one_of(input$SelectNUM),                    as.numeric),
        across(one_of(input$SelectFAC),                     as.factor),
        across(Y,                                           as.factor),
        across(one_of(input$SelectGEN),                   as.character)
      )                                                              %>%
      mutate(Y = recode_factor(Y, "0" = "healthy",   "1" = "patient"))
  })
  #                              Summary                             Table
  summ_tab=reactive({
   req(                                InputDataset_summ()           )
    InputDataset_summ()                                              %>%
      tbl_summary(
        by                              =                             Y,
        statistic                         =                        list(
          all_continuous()          ~          "{mean}          ({sd})",
          all_categorical()      ~      "{n}      /      {N}      ({p}%)"
        ),
        digits          =          all_continuous()          ~          2,
        missing_text                    =                    "(Missing)"
      )                                                              %>%
      add_overall()                                                  %>%
      add_n()                                                        %>%
      modify_header(label         ~         "**Predictor**")         %>%
      modify_spanning_header(c("stat_1",   "stat_2")   ~   input$SelectY)
  })
  output$Summ                                                         <-
    render_gt(
      summ_tab()%>%
        as_gt()
    )
  #                              EDA                                 Plots
  plots                        <-                       reactiveValues()
  observeEvent(InputDataset_summ(),                                    {
    req(InputDataset_summ)
    #                            genetic                             plot
    bb = InputDataset_summ() %>% select(Y, one_of(input$SelectGEN)) %>%
      pivot_longer(cols = -Y) %>% group_by(name, Y, value) %>% count %>%
mutate(Y                                                              =
```

```
as.factor(Y))
    names(bb)      =      c("var",      "cond",      "cat",      "value")
    plots$plotGEN      <-                 bb      %>%      na.omit      %>%
      ggplot()                                                            +
      geom_bar(aes(y    =    value,    x    =    cat,    fill    =    cond),
               stat                    =                    "identity",
               position              =              'stack')              +
      facet_wrap(      ~      var,      scales      =      'free')      +
      theme(axis.text.x                  =                 element_text(
        angle                            =                            90,
        vjust                           =                           0.5,
        hjust                             =                            1
      ))
    #                              factor                             plot
    bb2 = InputDataset_summ() %>% select(Y, one_of(input$SelectFAC)) %>%
      pivot_longer(cols = -Y) %>% group_by(name, Y, value) %>% count %>%
mutate(Y                                                                  =

as.factor(Y))
    names(bb2)      =      c("var",      "cond",      "cat",      "value")
    plots$plotFAC      <-                 bb2      %>%      na.omit      %>%
      ggplot()                                                            +
      geom_bar(aes(y    =    value,    x    =    cat,    fill    =    cond),
               stat                    =                    "identity",
               position              =              'stack')              +
      facet_wrap(      ~      var,      scales      =      'free')      +
      theme(axis.text.x                  =                 element_text(
        angle                            =                            90,
        vjust                           =                           0.5,
        hjust                             =                            1
      ))
    #                              numeric                            plot-
    dat3  =  InputDataset_summ()  %>%  select(Y,  where(is.numeric))  %>%
mutate(Y                                                                  =
as.factor(Y))                    %>%                    pivot_longer(-Y)
    dodge          <-          position_dodge(width          =          0.4)
    plots$plotNUM           <-                    ggplot(dat3)           +
      aes(x     =     Y,     y     =     value,     fill     =     Y)     +
      geom_violin(position              =              dodge)             +
      geom_boxplot(width                        =                       .1,
                   outlier.colour                    =                  NA,
                   position              =              dodge)             +
      theme_minimal()                                                     +
      facet_wrap(~          name,          scales          =          'free')
  })
  output$plotGEN                      <-                      renderPlot({
```

```
    plots$plotGEN
  })
  output$plotFAC                    <-                    renderPlot({
    plots$plotFAC
  })
  output$plotNUM                    <-                    renderPlot({
    plots$plotNUM
  })
  #                                                          Partition
  splitSlider                       <-                     reactive({
    input$partition                        /                       100
  })
  set.seed(2022)   # setting seed to reproduce results of random sampling
  trainingRowIndex                                                  <-
    reactive({
      req(InputDataset_summ)
      sample(1:nrow(InputDataset_summ()),
             splitSlider()         *         nrow(InputDataset_summ()))
    })#        row        indices        for        training        data
  trainingData                      <-                     reactive({
    tmptraindt                  <-                 InputDataset_summ()
    tmptraindt[trainingRowIndex(),]
  })
  testData                        <-                       reactive({
    tmptestdt                   <-                 InputDataset_summ()
    tmptestdt[-trainingRowIndex(),                                  ]
  })
  output$cntTrain                                                   <-
    renderText(paste("Train  Data:",  NROW(trainingData()),  "records"))
  output$cntTest                                                    <-
    renderText(paste("Test   Data:",   NROW(testData()),   "records"))
#                              end                              partition
  #MODEL
#                                                              formula
  formula=reactive({
    req(dat_model())
    #       exclude       vars       with       only       1       value
    datmod=dat_model()    %>%    select(where(~n_distinct(.)    >    1))
    EN_import<-fun4_EN_formula(datmod,       resp=       input$SelectY,
id=input$SelectID)
   EN_import$form
  })
#                                fun                                4
  model_out                <-                reactive(                {
req(dat_model())
      datmod=datmod1  %>%  mutate(across(everything(),~as.factor(.)))%>%
```

```
mutate(across(any_of(input$SelectNUM),~as.numeric(.x)))
    fun4_EN(datmod,  resp=  input$SelectY,  p=splitSlider(),  formula  =
formula())
  })
 #                    model                    results                    -
   output$confmat          <-          renderPrint(model_out()$confmat)
   output$coefs            <-            renderPrint(model_out()$coefs)
   output$acc              <-              renderPrint(model_out()$acc)
   output$auc              <-              renderPrint(model_out()$auc)
   output$mod_pl1=renderPlot({                        model_out()$pl1})
   output$mod_pl2=renderPlot({                        model_out()$pl2})
   #                                                          PREDICTION
   output$panelstatus                    <-                   reactive({
     req(dat_model())
TRUE
   })
   outputOptions(output,  "panelstatus",  suspendWhenHidden  =  FALSE)
 predDataset                        <-                        reactive({
     #          read          default          or          user          data
     if                   (is.null(input$file_pred))                   {
      datmod1     %>%     mutate(across(everything(),~as.factor(.)))%>%
mutate(across(any_of(input$SelectNUM),~as.numeric(.x)))
     }                                 else                            {
       #                               get                          input
        infile                      <-                    input$file_pred
       #                  read                  and                  return
          rio::import(infile$datapath)
     }
 })
   pred_rst=reactive({
     req(dat_model())
     req(predDataset())
    #     bind     to     dat_model     toget     correct     col     classes
zz    =    rbindlist(list(dat_model(),    predDataset()),    fill    =    T)
     preda=     zz%>%     mutate(across(everything(),~as.factor(.)))%>%
mutate(across(one_of(input$SelectNUM),~as.numeric(.x)))   %>%   select(-
one_of(input$SelectID))  %>%  select(where(~n_distinct(.)  >  1))  %>%
select(where(~nlevels(.)       >       1       |       is.numeric(.)))
     test_x     <-     build.x(formula(),preda,contrasts     =     FALSE)
     preds=                                                     predict(
       model_out()$model,
       s                          =                        "lambda.min",
       newx                            =                          test_x,
       type                            =                          "class"
     )
       predDat       =       predDataset()       %>%       mutate(Predict
```

```
=tail(preds,nrow(predDataset()))  %>%  select(input$SelectID,Predict,
everything())
    predDat
  })
      #                          DT                          prediction
 output$DataPred               <-               DT::renderDataTable({
  DT::datatable(
    pred_rst(),
    extensions                    =                    "Buttons",
    options                        =                        list(
      paging                       =                        TRUE,
      scrollX                      =                        TRUE,
      searching                     =                       TRUE,
      ordering                     =                        TRUE,
      dom                          =                     'Bfrtip',
      buttons        =        c('copy',        'csv',        'excel'),
      pageLength                    =                        15,
      lengthMenu          =          c(3,          5,          10)
    )
  )
 })
  #                                                          REPORT
  observeEvent(model_out(),{
    shinyjs::enable("report")
    shinyjs::enable("file_pred")
  })
  output$report                  <-                  downloadHandler(
    filename                     =                     "report.doc",
    content              =              function(file)              {
      tempReport      <-      file.path(tempdir(),      "report.Rmd")
      file.copy("report.Rmd",    tempReport,    overwrite    =    TRUE)
      #    Set    up    parameters    to    pass    to    Rmd    document
      params                                                      <-
        list(
          file=input$file_mtb$name,
          summ_tab=summ_tab(),
          plotGEN=plots$plotGEN,
          plotFAC=plots$plotFAC,
          plotNUM=                                  plots$plotNUM,
          HW_MAF_p=                                HW_MAF_plist(),
          aggrPlot=dat3()$aggr_plot,
          p1                      =                   model_out()$pl1,
          p2                      =                   model_out()$pl2,
          coefs                    =                model_out()$coefs,
          confmat                  =               model_out()$confmat,
          auc                      =                    model_out()$auc
```

```
        )
      rmarkdown::render(tempReport,      output_file      =      file,
                        params                =                 params,
                        envir    =    new.env(parent    =    globalenv())
      )
    }
  )
}
```

UI

```
if                                                    (!require("pacman"))
  install.packages("pacman")
pacman::p_load(
  "shinyjs",
  "shinythemes",
  "shinycssloaders",
  "shinydashboard",
  "shiny",
  data.table
)
ui<-dashboardPage(
  dashboardHeader(title  =  "RSGEN",  dropdownMenuOutput("msgOutput")),
  #                                                        dashboardSidebar
  dashboardSidebar(
    useShinyjs(),
    fileInput(inputId = "file_mtb", label = "Please upload a  dataset",
              multiple  =  FALSE,  accept  =  c('.xlsx',  '.xls',  '.csv'),
placeholder       =       "              No       file       selected"
    ),
    h5(textOutput("cntTrain")),
    h5(textOutput("cntTest")),
    selectInput(
      "p_val",
      #                          fun2-->row                          51
      label          =          "HW          p-Value          Pevel",
      choices                   =                   c(0.05,0.001,0.1),
      multiple                          =                          F,
      selected                         =                        0.05
    ),
    sliderInput(
      "MAF",
      #                          fun2-->                          /100
      label               =               h5("MAF               Level"),
      min                              =                            0,
```

```
      max                                 =                                 100,
      value                                 =                                 50
    ),
    sliderInput(
      "partition",
      #                                                                     fun4
      label         =         h5("Dataset         Partition         Level"),
      min                                 =                                 50,
      max                                 =                                 100,
      step=1,
      value                                 =                                 80
    ),
    sliderInput(
      #nmis
      "nmis_oth_tol",
      label       =       h5("Non-SNP's       NAs       Row       Level"),
      min                                 =                                 0,
      max                                 =                                 100,
      step=5,
      value                                 =                                 30
    ),
    sliderInput(
      #nmis_gen
      "nmis_gen_tol",
       label       =       h5("SNP's       NAs       Row       Level"),
      min                                 =                                 0,
      max                                 =                                 100,
      step=5,
      value                                 =                                 30
    ),
    sliderInput(
      #MVV
      "MVV_prop",
      label          =          h5("NAs          Column          Level"),
      min                                 =                                 0,
      max                                 =                                 100,
      step=5,
      value                                 =                                 5
    ),
    br()
  ),
  #                                                                 dashboardBody
  dashboardBody(
  fluidPage(
      tabBox(
      id                                 =                                 "tabset1",
```

```
      height                          =                          "1000px",
      width                             =                              12,
  #                              Select                              Vars
      tabPanel("Data                                              Import",
               box(
                 selectInput(
                   "SelectNUM",
                   label     =     "List     of     Numeric     variables",
                   choices                      =                      "",
                   multiple                     =                    TRUE,
                   selected                       =                      ""
                 ),
                 solidHeader                    =                    TRUE,
                 width                        =                       "5",
                 status                     =                   "primary",
                 title           =           "Numeric           variables"
               ),
               box(
                 selectInput(
                   "SelectFAC",
                   label     =     "List     of     Factor     variables",
                   choices                      =                      "",
                   multiple                     =                    TRUE,
                   selected                       =                      ""
                 ),
                 solidHeader                    =                    TRUE,
                 width                        =                       "5",
                 status                     =                   "primary",
                 title           =           "Factor            variables"
               ),
               box(
                 selectInput(
                   "SelectGEN",
                   label     =     "List     of     Genetic     variables",
                   choices                      =                      "",
                   multiple                     =                    TRUE,
                   selected                       =                      ""
                 ),
                 solidHeader                    =                    TRUE,
                 width                        =                       "5",
                 status                     =                   "primary",
                 title           =           "Genetic           variables"
               ),
               box(
                 selectInput("SelectY",  label  =  "Select  Predictor",
choices                          =                                  ""),
```

```
                  solidHeader                    =                    TRUE,
                  width                        =                     "3",
                  status                     =                   "primary",
                  title          =          "Predictive          variable"
                ),
                box(
                  selectInput("SelectID", label = "Select ID:", choices
=                                                                      ""),
                  solidHeader                    =                    TRUE,
                  width                        =                     "3",
                  status                     =                   "primary",
                  title             =             "ID             variable"
                ),
                box(               actionButton("go",               "OK")
                    )
                ,
                width=12
                ),
      #                                                                Data
      tabPanel("Data                                          Presentation",
                box(withSpinner(
                  column(width                        =                  12,
                         DT::dataTableOutput("Data"),style             =
"height:500px;       overflow-y:       scroll;overflow-x:       scroll;"
                  )
                ),              width               =               18)),
      #                              Data                               summ
      tabPanel(
        "Data                                                  Summary",
      withSpinner(                                      gt_output("Summ"))
      ),
      #                                EDA                                Plot
            tabPanel("Exploratory      Data      Analysis      Plots",
               fluidPage(
          tabsetPanel(
                     tabPanel("NAs              pattern              plot",
                box(withSpinner(plotOutput(
                  "NA_plots"
                )),                width               =               12)),
                tabPanel("HW                genotype                plot",
                         withSpinner(plotOutput(
                           "HW_MAF_plist",                    height=800
                         ))),
           tabPanel("Genetic                                          plot",
                withSpinner(plotOutput(
                  "plotGEN",                                  height=800
```

```
              ))),
          tabPanel("Factor                                          plot",
                   withSpinner(plotOutput(
                     "plotFAC",                                 height=500
                   ))),
          tabPanel("Numeric                                         plot",
                   withSpinner(plotOutput(
                     "plotNUM",                                 height=500
                   )))
      )                                                                 )),
  #                                                                   Model
        tabPanel(
       "Model",
       fluidPage(
         tabsetPanel(
          #                          Tabular                          Results
           tabPanel("Tabular                                      Results",
         fluidRow(column(6,                                   align="left",
      hr("Model                                                  Summary"),
         withSpinner(verbatimTextOutput("coefs"))
       ),
       column(6,
          hr("Confusion                                          Matrix"),
                withSpinner(verbatimTextOutput("confmat")),
                     hr("AreaUunder            the             Curve"),
           withSpinner(verbatimTextOutput("auc"))
       )
      )),

     #                              Model                              Plots
      tabPanel(                                "Model                Plots",
        box(withSpinner(plotOutput("mod_pl1")),  width  =  6,  title  =
"Elastic              Net              coeficiente              plot"),
        box(withSpinner(plotOutput("mod_pl2")), width = 6, title = "Area
under                             curve                             plot")
      )
         )
       )
        )
           ,
  #                                                                   Report
  tabPanel(
    "Report",
    conditionalPanel(
      condition              =              'output.panelstatus==true',
          downloadButton("report",          "Generate          report")
```

```
    )
  ),
  #                                                          Prediction
      tabPanel(
        "Prediction",
      conditionalPanel(
        condition              =              'output.panelstatus==true',
        fileInput(inputId  =  "file_pred",  label  =  "Please  upload  a
dataset",
                  multiple = FALSE, accept = c('.xlsx', '.xls', '.csv'),
placeholder          =          "          No          file          selected"
        ),
        withSpinner(
          column(width                          =                          12,
                 DT::dataTableOutput("DataPred"),style = "height:500px;
overflow-y:                  scroll;overflow-x:                  scroll;"
          )
        )
        )
      ,                   width                    =                    18)
    )
  )
  )
)
```

Import

```
form_fun=function(  dat,  resp=  "statut_atherome",           id="id"){
  xx=crossing(var1 = c("A","T","G","C","1 ","1 ","2 ", "0 ", "1 ", "2 "
), var2 = c("A","T","G","C","1","1","2","0", "1", "2")) %>% unite("z",
var1:var2,       remove       =       T,sep="")       %>%       pull(z)
  genetix=dat                                                        %>%
    select(where(~  any(.  %in%  xx,  na.rm  =  TRUE)))  %>%  names()
  #      identify      integers-      >      convert      to      factors
  is_whole   <-   function(x)   all(floor(x)   ==   x,   na.rm   =   T)
  unixx=function(x)                                  length(unique(x))<=5
  integs = dat %>% select_if(is.numeric) %>%  select_if( is_whole )%>%
select_if(            unixx            )            %>%            names
  #       convert        integers-       characters       to       factor
  dat2=dat    %>%    mutate_if(sapply(.,    class)    %in%    c("factor",
"character","integer"),                  as.factor)                  %>%
mutate(across(integs,~as.factor(.x)))
  #                  factor                    col                    names
  factors=dat2      %>%      select_if(is.factor)      %>%      names()
  facts=factors[!factors            %in%            c(genetix,resp,id)]
  #                  numeric                   col                    names
```

```
  nums=dat2 %>% select_if(is.numeric) %>% select(-one_of(id))%>% names()
  #                                                                formula
  form=as.formula(paste(resp,"~",                          paste(facts,
collapse="+"),"+",paste(nums,         collapse="+"),"+(",paste(genetix,
collapse="+"),")^2"))
  return(list(form=form,        genetix=genetix,facts=facts,nums=nums))
}
```

HW MAF

```
fun2_HW_MAF=function(dataset,col_names_genetic,status,k,  HWpval=0.05  ,
MAF=0.5){
  #                                  dataset=janitor::clean_names(dataset)
  if                (k                 ==                3)                {
    dataset<-dataset
  }                                 else                                  {
    dataset<-dataset             %>%            filter(!!sym(status)==k)
  }
  dataset  <-    dataset  %>%  dplyr::select(  c(col_names_genetic  )  )
  n_row<-dim(dataset)[2]
    x<-matrix(nrow        =        n_row,        ncol        =        3)
  HWC<-matrix(nrow        =        n_row,        ncol        =        3)
  HWE<-matrix(nrow        =        n_row,        ncol        =        3)
  dataset_n_temp<-matrix(nrow      =      n_row,      ncol      =      3)
  HW_MAF<-matrix(nrow       =       n_row,       ncol       =       4)
    for                 (i                 in                 1:n_row){
    dataset_n_temp[i]<-list(table(dataset[i]))
      for                 (j                 in                 1:3){
      x[i,j]<-sapply(dataset_n_temp[i],  function(x){as.numeric(x[j])})
    }
    x[is.na(x)]                          =                          0
    HWC[i]<-
list(HWChisq(c(MM=x[i,1],MN=x[i,2],NN=x[i,3]),verbose=TRUE))
    HWE[i]<-
list(HWExact(c(MM=x[i,1],MN=x[i,2],NN=x[i,3]),verbose=TRUE))
  }
  #maf
  maf<-matrix(nrow        =        n_row,        ncol        =        4)
  for                 (i                 in                 1:n_row){
    m                              <-                              x[i,]
    m                            <-                            m/sum(m)
    HW_MAF[i,1]<-                                 col_names_genetic[i]
    HW_MAF[i,2]<-        round(maf(m),        digits        =        2)
    HW_MAF[i,3]<-round(HWC[[i]]$chisq,        digits        =        2)
    HW_MAF[i,4]<-round(HWC[[i]]$pval,        digits        =        2)
    #remove       snps       that       HWpval<0.05       or       MAF>0.5
```

```
    if      ((HW_MAF[i,2]>MAF)      |      (HW_MAF[i,4]<HWpval))      {
      dataset<-dataset[,-c(i)]
    }
  }
  HW_MAF<-as.data.frame(HW_MAF)
  names(HW_MAF)[1]                    <-                    paste("SNP")
  names(HW_MAF)[2]                    <-                    paste("MAF")
  names(HW_MAF)[3]                  <-                  paste("HWchisq")
  names(HW_MAF)[4]                   <-                   paste("HWpval")
  n_ind<-dim(dataset)[1]
  return(list(names(dataset),HW_MAF,             x,             n_ind))
}
```

Outlier

```
#     convert     ALL     outliers     to     NA     in     each     col
remove_outliers    <-    function(x,    na.rm    =    TRUE,    ...)    {
  qnt   <-   quantile(x,   probs=c(.25,   .75),   na.rm   =   na.rm,   ...)
  H      <-      1.5      *      IQR(x,      na.rm      =      na.rm)
  y                                  <-                                  x
  y[x        <        (qnt[1]        -        H)]        <-        NA
  y[x        >        (qnt[2]        +        H)]        <-        NA
  y
}
remove_outliers_df=function(df){
  df2=df                                                              %>%
    mutate(across(where(is.numeric),             ~remove_outliers(.x)))
  return(df2)
}
```

Imputation

```
fun3_imp    =    function(dat,    ncpMax    =    5,    col_names_numeric)    {
  dataset      =      dat      %>%                ungroup()            %>%
    mutate(across(everything(),              ~      as.factor(.x)))      %>%
    mutate(across(any_of(col_names_numeric),      ~    as.numeric(.x)))    %>%
select(where(~                        n_distinct(.)           >           1))
  res.ncp   <-   try(estim_ncpFAMD(dataset,   ncp.max   =   ncpMax),silent=T)
  if             (!inherits(res.ncp,             "try-error"))             {
    print("IMP1")
    res.famd   <-try(   imputeFAMD(dataset,   ncp   =   res.ncp$ncp),silent=T)
    if             (!inherits(res.famd,             "try-error"))             {
      dataset_imp1                  <-                  res.famd$completeObs
      dataset_imp2                                                        <-
```

```
        as.data.frame(lapply(dataset_imp1,                          function(x)
          sub(".*_",                          "",                          x)))
      dataset_imp              =              dataset_imp2              %>%
        mutate(across(everything(),            ~      as.factor(.)))      %>%
        mutate(across(any_of(col_names_numeric),                          ~
as.numeric(.x)))
    }                                                                  else{
      print("ERR")
      dataset_imp                          =                          dataset
    }
  }else{
    print("ERR")
    dataset_imp                           =                          dataset
  }
    return(dataset_imp)
      }
```

Elastic Net formula

```
fun4_EN_formula=function(    dataset,    resp=    "status",    id="id"){
  #        exclude        vars        with        only        1        value
  dataset=dataset     %>%     select(where(~n_distinct(.)     >     1))
  #                  identify                  genetic                  cols
  #            all            combinations            of            A-T-C-G
  xx=crossing(var1 = c("A","T","G","C","1 ","1 ","2 ", "0 ", "1 ", "2 "
), var2 = c("A","T","G","C","1","1","2","0", "1", "2")) %>% unite("z",
var1:var2,       remove       =       T,sep="")       %>%       pull(z)
  genetic=dataset                                                    %>%
    select(where(~  any(.  %in%  xx,  na.rm  =  TRUE)))  %>%  names()
  #       identify       integers-       >       convert       to       factors
  is_whole   <-   function(x)   all(floor(x)   ==   x,   na.rm   =   T)
  unixx=function(x)                                length(unique(x))<=8
  integs = dataset %>% select_if(is.numeric) %>%  select_if( is_whole
)%>%              select_if(       unixx       )       %>%       names
  #        convert        integers-        characters        to        factor
  dataset2=dataset  %>%  mutate_if(sapply(.,  class)  %in%  c("factor",
"character","integer"),                  as.factor)                  %>%
mutate(across(integs,~as.factor(.x)))
  #                   factor                   col                   names
  factors=dataset2     %>%     select_if(is.factor)     %>%     names()
  factor=factors[!factors        %in%        c(genetic,resp,        id)]
  #                   numeric                   col                   names
  numeric=dataset2     %>%     select_if(is.numeric)     %>%     names()
  #                                                                formula
  formula=as.formula(paste(resp,"~",    paste(factor,    collapse="+"),
                          "+",paste(numeric,            collapse="+"),
```

```
                              "+(",paste(genetic,    collapse="+"),")^2"))
  return(list(formula=formula,
genetic=genetic,factor=factor,numeric=numeric))
}
```

Elastic Net

```
extractGlmnetInfo                  <-                  function(object)
{
  #                                find                                lambdas
  lambdaMin                        <-                    object$lambda.min
  lambda1se                        <-                    object$lambda.1se
  #    figure    out    where    those    lambdas    fall    in    the    path
  whichMin       <-       which(object$lambda       ==       lambdaMin)
  which1se       <-       which(object$lambda       ==       lambda1se)
  # build a one line data.frame with each of the selected lambdas and
  # its corresponding error figures338 Chapter 22 Regularization and Shrinkage
  data.frame(lambda.min=lambdaMin,      error.min=object$cvm[whichMin],
             lambda.1se=lambda1se,      error.1se=object$cvm[which1se])
}
nonzeroCoef      =      function      (beta,      bystep      =      FALSE)
{
#   bystep   =   FALSE   means   which   variables   were   ever   nonzero
#  bystep  =  TRUE  means  which  variables  are  nonzero  for  each  step
  nr=nrow(beta)
  if          (nr          ==          1)          {#degenerate          case
    if                                                          (bystep)
      apply(beta,     2,     function(x)     if     (abs(x)     >     0)
        1
        else                                                          NULL)
    else                                                              {
      if                (any(abs(beta)                >                0))
        1
      else                                                          NULL
    }
  }
  else                                                                {
    beta=abs(beta)>0          #          this          is          sparse
    which=seq(nr)
    ones=rep(1,ncol(beta))
    nz=as.vector((beta%*%ones)>0)
    which=which[nz]
    if                                (bystep)                          {
      if(length(which)>0){
        beta=as.matrix(beta[which,,drop=FALSE])
```

```
        nzel      =      function(x,      which)      if      (any(x))
          which[x]
        else                                                    NULL
        which=apply(beta,           2,           nzel,           which)
        if(!is.list(which))which=data.frame(which)# apply can return a
matrix!!
        which
      }
      else{
        dn=dimnames(beta)[[2]]
        which=vector("list",length(dn))
        names(which)=dn
        which
      }
    }
    else                                                        which
  }
}
fun4_EN=function(dat,  resp=  "status",  p=0.83,  formula="formula"){
  set.seed(1564641)
  dat=   dat   %>%   mutate(across(where(is.character),   as.factor))
  #       exclude       vars       with       only       1       value
  dat=dat       %>%       select(where(~n_distinct(.)       >       1))
  intrain<-createDataPartition(y=dat[[resp]],p=p,list=FALSE)
  training<-dat[intrain,]
  testing<-dat[-intrain,]
  #build                  x                  for                  training
  train_x<-build.x(formula      ,training,contrasts      =      FALSE)
  #build             y             for             training             set
  train_y<-build.y(formula,training)
  #apoeg
  test_x     <-     build.x(formula,testing,contrasts     =     FALSE)
  test_y                  <-                  build.y(formula,testing)
  set.seed(1863561)
  #          run          the          cross-validated          glmnet
  fit1 <- cv.glmnet(x=train_x, y=train_y, family="binomial",weights =
NULL, offset = NULL, standardize = TRUE, intercept = TRUE, nfold=5, alpha
=                                                                   0)
  #                             start                             point
  set.seed(2834673)
  theFolds     <-     sample(rep(x=1:5,     length.out=nrow(train_x)))
  alphas         <-         seq(from=0,         to=1,         by=.05)
  #set    the    seed    for    repeatability    of    random    results
  set.seed(5127151)
  #       start       a       cluster       with       two       workers
  cl                          <-                          makeCluster(8)
```

```
  # register the workers
  registerDoParallel(cl)
  # keep track of timing
  before <- Sys.time()
  # build foreach loop to run in parallel
 # several arguments
  fit_all2 <- foreach(i=1:length(alphas), .errorhandling="remove",
                      .inorder=FALSE, .multicombine=TRUE,
                      .export=c("train_x", "train_y", "alphas",
"theFolds"),
                      .packages="glmnet") %dopar%
    {print(alphas[i])
      cv.glmnet(x=train_x, y=train_y, family="binomial", standardize =
TRUE, intercept = TRUE, nfolds=20,
                foldid=theFolds, alpha=alphas[i])
    }
  # stop timing
  after <- Sys.time()
    # make sure to stop the cluster when done
  stopCluster(cl)
    # time difference
  # this will depend on speed, memory & number of cores of the machine
  sapply(fit_all2, class)
  alphaInfo <- Reduce(rbind, lapply(fit_all2, extractGlmnetInfo))
  # make a column listing the alphas
  alphaInfo$Alpha <- alphas
  # melt the data into long format
  alphaMelt <- melt(alphaInfo, id.vars="Alpha", value.name="Value",
                    variable.name="Measure")
  alphaMelt$Type <- str_extract(string=alphaMelt$Measure,
                                pattern="(min)|(1se)")
  # some housekeeping
  alphaMelt$Measure <- str_replace(string=alphaMelt$Measure,
                                   pattern="\\.(min|1se)",
                                   replacement="")
  alphaCast <- dcast(alphaMelt, Alpha + Type ~ Measure,
                     value.var="Value")
  alpha1000=alphaInfo$Alpha[which.min(alphaInfo$error.min)]
  set.seed(5127151)
  acsCV3 <- cv.glmnet(x=train_x, y=train_y, family="binomial",
standardize = TRUE, intercept = TRUE, nfold=20,
                      alpha=alphaInfo$Alpha[which.min(
                        alphaInfo$error.min
                      )])
  x <- coef(acsCV3, s="lambda.min")
  theCoef <- as.matrix(coef(acsCV3, s="lambda.min"))
```

```
  coefDF <- data.frame(Value=theCoef, Coefficient=rownames(theCoef))
  coefDF <- coefDF[nonzeroCoef(coef(acsCV3, s="lambda.min")), ]
  pred1 <- predict(acsCV3,s = "lambda.min", newx = test_x, type =
"response")
  pred2 <- predict(acsCV3,s = "lambda.min", newx = test_x, type =
"class")
  #take test_y from first time
  conf <- data.frame(pred2, test_y)
  aa=confusionMatrix(as.factor(conf[,1]), as.factor(conf[,2]))
  # c = confusionchart(yTest,pred2);
  p <- predict(acsCV3,s = "lambda.min", newx = test_x)
  pr <- prediction(p, as.data.frame(test_y))
  prf <- performance(pr, measure = "tpr", x.measure = "fpr")
  # plot(prf)
    auc <- performance(pr, measure = "auc")
  auc <- auc@y.values[[1]]
  #values to be presented
  auc
  plot<- ggplot(coefDF, aes(x=s1, y=reorder(Coefficient, s1))) +
    geom_vline(xintercept=0, color="grey", linetype=2) +
    geom_point(color="blue") +
    labs(x="Value", y="Coefficient", title="Coefficient Plot")
  # coefDF[,c(1,2)]
  rownames(coefDF) <- NULL
  plot.new()
   plot(prf)
   pl2=recordPlot()
return(list(coefs=coefDF[,c(2,1)],pl1=plot, pl2=pl2,auc=auc,acc=
aa[["overall"]][["Accuracy"]],confmat=aa,model=acsCV3))
}
```

PMI script

Tables

Table	QueryDefinition
TL	let
	Source = Excel.Workbook(File.Contents(FolderPath & Filename), null, true),
	TL_Sheet = Source{[Item="TL",Kind="Sheet"]}[Data],
	#"Promoted Headers" = Table.PromoteHeaders(TL_Sheet, [PromoteAllScalars=true]),
	#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"NUMERO", type text}, {"Age", type number}, {"rs2736100", type text}, {"rs10936599", type text}, {"rs755017", type text}, {"rs9420907", type text}, {"rs7675998", type text}, {"rs11125529", type text}, {"rs8105767", type text}, {"rs2162440", type text}, {"rs12696304", type text}, {"rs7726159", type text}, {"rs2487999", type text}, {"rs74019828", type text}, {"rs7680468", type text}, {"rs6028466", type text}, {"rs621559", type text}, {"rs398652", type text}, {"rs1975174", type text}, {"rs4452212", type text}, {"rs654128", type text}, {"rs2736428", type text}, {"rs2098713", type text}, {"rs17653722", type text}, {"rs412658", type text}, {"rs6772228", type text}, {"rs3027234", type text}, {"rs4387287", type text}, {"rs9419958", type text}, {"rs10761741", type text}, {"rs10738760", type text}, {"rs6921438", type text}, {"rs7043199", type text}, {"rs6993770", type text}, {"rs4416670", type text}, {"rs114694170", type text}, {"rs34528081", type text}, {"rs4782371", type text}, {"rs2639990", type text}, {"LTL", type number}, {"MTL", type number}, {"BMI", type number}, {"PAS", type number}, {"PAD", type number}}),
	#"Removed Other Columns" = Table.SelectColumns(#"Changed Type",{"Country","NUMERO", "SEXE", "Age", "Statut Athérome", "BMI", "MTL", "LTL", "PAD", "PAS"}),
	#"Renamed Columns" = Table.RenameColumns(#"Removed Other Columns",{{"NUMERO", "MeasurementID"}, {"SEXE", "Sex"}, {"Statut Athérome", "isHealthy"}}),
	#"Filtered Rows" = Table.SelectRows(#"Renamed Columns", each ([Sex] <> null)),
	#"Renamed Columns1" = Table.RenameColumns(#"Filtered Rows",{{"MeasurementID", "ID"}}),
	#"Changed Type1" = Table.TransformColumnTypes(#"Renamed Columns1",{{"Sex", type text}}),
	#"Replaced Value" = Table.ReplaceValue(#"Changed

	Type1","2","Female",Replacer.ReplaceText,{"Sex"}),
	#"Replaced Value1" = Table.ReplaceValue(#"Replaced Value","1","Male",Replacer.ReplaceText,{"Sex"}),
	#"Changed Type2" = Table.TransformColumnTypes(#"Replaced Value1",{{"isHealthy", type text}}),
	#"Replaced Value2" = Table.ReplaceValue(#"Changed Type2","1","Healthy",Replacer.ReplaceText,{"isHealthy"}),
	#"Replaced Value3" = Table.ReplaceValue(#"Replaced Value2","0","Patient",Replacer.ReplaceText,{"isHealthy"}),
	#"Renamed Columns2" = Table.RenameColumns(#"Replaced Value3",{{"ID", "Individuals"}, {"isHealthy", "Status"}})
	in
	#"Renamed Columns2"
SNPs	let
	TL_Sheet = Source{[Item="TL",Kind="Sheet"]}[Data],
	#"Promoted Headers" = Table.PromoteHeaders(TL_Sheet, [PromoteAllScalars=true]),
	#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"NUMERO", type text}, {"SEXE", Int64.Type}, {"Age", type number}, {"Statut Athérome", Int64.Type}, {"rs2736100", type text}, {"rs10936599", type text}, {"rs755017", type text}, {"rs9420907", type text}, {"rs7675998", type text}, {"rs11125529", type text}, {"rs8105767", type text}, {"rs2162440", type text}, {"rs12696304", type text}, {"rs7726159", type text}, {"rs2487999", type text}, {"rs74019828", type text}, {"rs7680468", type text}, {"rs6028466", type text}, {"rs621559", type text}, {"rs398652", type text}, {"rs1975174", type text}, {"rs4452212", type text}, {"rs654128", type text}, {"rs2736428", type text}, {"rs2098713", type text}, {"rs17653722", type text}, {"rs412658", type text}, {"rs6772228", type text}, {"rs3027234", type text}, {"rs4387287", type text}, {"rs9419958", type text}, {"rs10761741", type text}, {"rs10738760", type text}, {"rs6921438", type text}, {"rs7043199", type text}, {"rs6993770", type text}, {"rs4416670", type text}, {"rs114694170", type text}, {"rs34528081", type text}, {"rs4782371", type text}, {"rs2639990", type text}, {"LTL", type number}, {"MTL", type number}, {"BMI", type number}, {"PAS", type number}, {"PAD", type number}}),
	#"Removed Columns" = Table.RemoveColumns(#"Changed Type",{"SEXE", "Age", "Statut Athérome", "LTL", "MTL", "BMI", "PAS", "PAD", "Country"}),
	#"Unpivoted Other Columns" = Table.UnpivotOtherColumns(#"Removed Columns", {"NUMERO"}, "Attribute", "Value"),

	#"Renamed Columns" = Table.RenameColumns(#"Unpivoted Other Columns",{{"Attribute", "Gene"}, {"NUMERO", "Individuals"}}),
	#"Filtered Rows" = Table.SelectRows(#"Renamed Columns", each ([Value] <> "Bad" and [Value] <> "Uncallable")),
	#"Replaced Value" = Table.ReplaceValue(#"Filtered Rows","?","NA",Replacer.ReplaceText,{"Value"}),
	#"Renamed Columns1" = Table.RenameColumns(#"Replaced Value",{{"Gene", "SNP"}})
	in
	#"Renamed Columns1"
Measures	let
	Source = Table.FromRows(Json.Document(Binary.Decompress(Binary.FromText("i45WSklJUYqNBQA=", BinaryEncoding.Base64), Compression.Deflate)), let _t = ((type nullable text) meta [Serialized.Text = true]) in type table [Column1 = _t]),
	#"Changed Type" = Table.TransformColumnTypes(Source,{{"Column1", type text}})
	in
	#"Changed Type"
Selections	let
	Source = Table.FromRows(Json.Document(Binary.Decompress(Binary.FromText("i45WcvL1VIrViVbyCfEB075QOsDRBUoHK8XGAgA=", BinaryEncoding.Base64), Compression.Deflate)), let _t = ((type nullable text) meta [Serialized.Text = true]) in type table [Selection = _t]),
	#"Changed Type" = Table.TransformColumnTypes(Source,{{"Selection", type text}})
	in
	#"Changed Type"
Genes Descriptions	let
	Source = Excel.Workbook(File.Contents(FolderPath & Filename), null, true),
	GeneDescriptions_Table = Source{[Item="GeneDescriptions",Kind="Table"]}[Data],
	#"Changed Type" = Table.TransformColumnTypes(GeneDescriptions_Table,{{"Gene", type text}, {"Description", type text}})
	in
	#"Changed Type"

Metrics

Column Name	Expression
Age group	switch(true(), [Age] >= 0 && [Age]<=17,"0-17", [Age] >= 18 && [Age]<=63,"18-64", [Age] >= 64,">64", "N/A")
BMI Group	switch(true(), [BMI] <18.5,"Underweight", [BMI] >= 18.5 && [bmi]<25,"Normal", [BMI] >= 25 && [bmi]<30,"Overweight", [BMI] >= 30 && [bmi]<35,"Obese", [BMI] >35,"Extremely Obese")
BMI Group Sort	switch(true(), [BMI] <18.5,"1", [BMI] >= 18.5 && [bmi]<25,"2", [BMI] >= 25 && [bmi]<30,"3", [BMI] >= 30 && [bmi]<35,"4", [BMI] >35,"5")
Age (bins)	IF(ISBLANK('TL'[Age]), BLANK(), IF('TL'[Age] >= 0, ROUNDDOWN('TL'[Age] / 5, 0) * 5, ROUNDUP('TL'[Age] / 5, 0) * 5))
Description	related(GeneDescriptions[Description])

Meausres

Measure Name	Expression
Average of Age average per Country	AVERAGEX (KEEPFILTERS (VALUES ('TL'[Country])), CALCULATE (AVERAGE ('TL'[Age])))
BMI Percentage to total	VAR Total = CALCULATE (COUNT (TL[Individuals]), ALL (tl)) VAR ThisOne = COUNT (TL[Individuals]) VAR pct = thisone / Total RETURN pct
Minimum of MTL LTL	VAR choice = IF (HASONEVALUE (Selections[selection]), VALUES (selections[selection]), "BMI") VAR LTL_min = MIN (TL[LTL]) VAR MTL_min = MAX (TL[MTL]) VAR result = SWITCH (choice, "MTL", MTL_min, "LTL", LTL_min, "MTL" RETURN

	RESULT
Count GA	COUNTROWS (FILTER (Genes, Genes[Value] = "GA"))
Count AA	COUNTROWS (FILTER (Genes, Genes[Value] = "AA"))
Count GG	COUNTROWS (FILTER (Genes, Genes[Value] = "GG"))
#Individuals	COUNT (TL[Individuals])
#Healthy	CALCULATE (DISTINCTCOUNT (TL[Individuals]), FILTER ('TL', TL[Status] = "1"))
# Individuals	COUNT (Genes[Individuals])

R code

Normal distribution

```
y<-dataset$PAD
name<-names(dataset)
st<-shapiro.test(y)
stp<-round(st$p.value,3)
hist(y, prob=TRUE, main = paste("Histogram of",names(dataset),", p-Value =" , stp))
curve(dnorm(x, mean(y), sd(y)), add=TRUE, col="darkblue", lwd=2)
```

HW pval

```
library(HardyWeinberg)
data <- `dataset`
HW.test <- HWChisq(table(data$Value),verbose=TRUE)
x<-round(HW.test$pval,2)
```

MAF

```
library(HardyWeinberg)
data <- `dataset`
HW.test <- HWChisq(table(data$Value),verbose=TRUE)
m<-as.matrix(table(data$Value)/(sum(table(data$Value))))
x<- round(maf(m[,1]),2)
bb = barplot(x, width = 6)
text(bb,x,labels=x,cex=4, adj = c(0.5, 1.5))
```

HW Genotype Plot

```
library(HardyWeinberg)
data <- `dataset`
m<-as.matrix(table(data$Value)/(sum(table(data$Value))))
HWGenotypePlot(m[,],plottype=1)
```

Correlations

```
library(psych)
pairs.panels(dataset)
```

Abstract

Cardiovascular diseases (CVD) and Alzheimer's disease (AD) are common chronic diseases multifactorial and polygenic and their management could significantly benefit from personalised medicine. The vascular endothelial growth factor A (VEGF-A) and the telomere length (TL) are optimal biomarkers for use in personalised risk prediction as they are involved in common diseases pathways and have an important heritable background.

The aim of this thesis was to identify "-omics" determinants of VEGF-A and TL, to investigate their relationships with CVD and AD and to examine their usefulness in disease risk prediction. For valuably achieving the aim, an emphasis was given in the standardisation of the analytical process and the visualisation of the datasets used in the thesis projects.

A review article presented the knowledge on telomere dynamics during childhood including the factors that affect TL in embryonic life and in newborns. We emphasise the role of TL as an index of health and we propose its regular measurement. In a next publication, we found significant associations between VEGF-A levels and its mRNA isoforms with IL-1α, IL-4, IL-6, EGF, and MCP1 in PBMCs of healthy individuals. Our results further support the important role of cytokines and growth factors in VEGF-A regulation at lymphocytes cellular extracts. Then, we demonstrated that two polymorphisms, the rs12696304 and rs10936599 near the TERC gene, do not affect TL at birth (estimated by muscle TL - MTL), but are associated with higher telomere attrition (estimated by leucocytes telomere length - LTL/MTL) during lifetime. Thus, polymorphisms located on genes coding for telomere maintenance proteins may contribute to higher TL attrition and could impact the development of chronic diseases like atherosclerotic CVD. A significant association between the VEGF-A related polymorphism rs6993770 and TL attrition has been found. The risk allele A could lead to increased haematopoiesis, thus higher cellular turnover and faster TL attrition. This polymorphism is a common determinant of VEGF-A and TL and could be an important marker in the development of prediction tools for chronic diseases where both biomarkers are involved, like CVD. We identified novel interactions between APOE, LSR and VEGF-A related polymorphisms that are predicting the risk of AD in a machine learning model with high accuracy. The interaction rs7043199*rs6993770 (VEGF-A related polymorphisms) was the strongest predictive factor, with a protective role higher even than that of the absence of ε4 APOE allele. This interaction was followed by the absence of the ε4 APOE allele and the interaction rs2375981*rs34528081(VEGF-A related polymorphisms). The results provide new insights about molecular mechanisms involved in AD development and could have an impact as biomarkers and/or treatment targets. We have also developed an automated pipeline of standardised procedures for statistical analyses and the development of prediction binary models by using the elastic net method and a related interactive interface (R Shiny Genetic Elastic Net, RSGEN). A desktop application was developed in Power BI software (Power Medical Intelligence, PMI) to allow visualisation of the data of the populations studied in the thesis as well as preliminary analytical steps.

In conclusion, this thesis is proposing determinants of VEGF-A and TL that are linked through known and novel mechanisms with CVD and AD: IL-4, IL-6, MCP-1, EGF, IL-1α, TERG variants, VEGF-A-related variants (especially the rs6993770), APOE and LSR polymorphisms. A risk prediction model for AD with a satisfactory accuracy has also been developed, which includes mostly genetic variants and their interactions, thus it is a highly personalised tool. Furthermore, we have developed two informatic and statistical tools that can be used in other populations and contribute to the development of robust personalised models for chronic diseases prediction.

Key-words: VEGF-A, telomere length, CVD, Alzheimer's disease, risk prediction, personalised medicine

Résumé

Les maladies cardiovasculaires (MCV) et la maladie d'Alzheimer (MA) sont des maladies chroniques, multifactorielles et polygéniques et leur gestion pourrait fortement bénéficier de la médecine personnalisée. Le facteur de croissance endothélial vasculaire A (VEGF-A) et la longueur des télomères (TL) sont deux biomarqueurs optimaux pour la prédiction du risque personnalisée car ils sont impliqués dans les voies métaboliques des maladies courantes et ont un contenu héréditaire important.

Le but de cette thèse était d'identifier les déterminants « -omiques » du VEGF-A et de la TL, d'étudier leurs relations avec les MCV et la MA et d'étudier leurs utilisations dans la prédiction du risque pour ces maladies.

Un article de revue présente la dynamique des télomères pendant l'enfance. Nous soulignons l'utilité du TL comme indice de santé et nous proposons sa quantification régulière. Nous avons trouvé des associations significatives entre les taux de VEGF-A et ses isoformes d'ARNm avec IL-1α, IL-4, IL-6, EGF et MCP1 dans les PBMCs d'individus sains. Nos résultats soutiennent le rôle des cytokines et des facteurs de croissance dans la régulation du VEGF-A. Ensuite, nous avons démontré que les variants rs12696304 et rs10936599, proche du gène TERC, n'affectent pas la TL à la naissance (estimée par la TL des muscles squelettiques - MTL), mais sont associés à une plus grande attrition des télomères (estimée par la longueur des télomères des leucocytes - LTL/ MTL) au cours de la vie. Ainsi, ces variants pourraient avoir un impact sur le développement de maladies chroniques telles que les MCV. Dans l'étude suivante, une association significative entre le variant lié au VEGF-A rs6993770 et l'attrition des télomères a été trouvée. L'allèle à risque A pourrait entraîner une augmentation de l'hématopoïèse, donc un renouvellement cellulaire plus élevé et une attrition plus rapide des télomères. Ce polymorphisme est un déterminant commun du VEGF-A et de la TL et pourrait être un marqueur important dans le développement d'outils de prédiction des maladies chroniques. De plus, nous avons identifié de nouvelles interactions entre les polymorphismes liés au VEGF-A et ceux des gènes d'APOE et LSR qui prédisent le risque de MA dans un modèle d'apprentissage automatique avec une grande précision. L'interaction rs7043199*rs6993770 (polymorphismes liés au VEGF-A) est le facteur prédictif le plus fort du modèle, avec un rôle protecteur contre la MA plus élevé même que celui de l'absence d'allèle ε4 APOE, suivie par l'absence de l'allèle APOE ε4 et de l'interaction rs2375981*rs34528081(polymorphismes liés au VEGF-A). Les résultats donnent de nouvelles informations sur les mécanismes moléculaires impliqués dans la MA et pourraient avoir un impact en tant que biomarqueurs et/ou cibles thérapeutiques. Nous avons également développé un pipeline automatisé de procédures standardisées pour les analyses statistiques et le développement de modèles binaires de prédiction avec la méthodologie Elastic Net et une interface interactive (R Shiny Genetic Elastic Net, RSGEN). Une application a été développée dans le logiciel Power BI (Power Medical Intelligence, PMI) pour permettre la visualisation des données des populations étudiées dans la thèse.

En conclusion, cette thèse propose des déterminants de VEGF-A et TL liés par des mécanismes connus et nouveaux avec les MCV et la MA: IL-4, IL-6, MCP-1, EGF, IL-1α, des variants du TERG, des polymorphismes liés au VEGF-A, en particulier le rs6993770, et des variants d'APOE et LSR. Un modèle de prédiction du risque de MA avec une précision satisfaisante a été développé, qui inclut en majorité des variants génétiques et leurs interactions, il s'agit donc d'un outil hautement personnalisé. De plus, nous avons développé deux outils informatiques et statistiques qui peuvent être utilisés dans d'autres populations et contribuer au développement de modèles personnalisés robustes pour la prédiction des maladies chroniques.

Mots-clés: VEGF-A, longueur des télomères, MCV, maladie d'Alzheimer, prédiction du risque, médecine personnalisée