Genomics of rare and complex/common metabolic diseases and their comorbidities

Sébastien Hergalant

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 **Sébastien HERGALANT**

Génomique des maladies rares et complexes/communes du métabolisme et de ses comorbidités

Le 15 décembre 2023

Membres du jury :

Rapporteurs :	**Mme Karin TARTE**	**PU/PH, UMR1236 INSERM, Université de Rennes**
	M. Hugues ASCHARD	**DR, USR3756 CNRS - Institut Pasteur, Paris**
Examinateurs :	**Mme Anne FRIEDRICH**	**MCF, UMR7156 CNRS, Université de Strasbourg**
	M. Olivier SAND	**IR, UAR3601 CNRS – IFB Core, Lille**
	M. Guillaume GAUCHOTTE	**PU/PH, Université de Lorraine, invité**
	M. Julien BROSÉUS	**PU/PH, Université de Lorraine, président**
	M. David MEYRE	**PU/PH, Université de Lorraine, directeur de thèse**

UMR1256 INSERM - NGERE Nutrition, Génétique et Exposition aux Risques Environnementaux – Campus Biologie Santé - 54500 Vandœuvre-lès-Nancy

REMERCIEMENTS

A mon directeur de thèse, futur directeur de labo, collègue et ami **David Meyre** : l'ordre est protocolaire et ne reflète pas vraiment celui qui m'importe. Merci, David, d'avoir accepté de m'encadrer et de m'avoir suivi dans cette aventure qui n'était pas tout à fait la tienne. Merci d'avoir fait front aux mêmes difficultés et d'avoir tenu bon. Merci pour ton excellence scientifique, ta rigueur millimétrique et ta vision. C'est rafraîchissant ! Merci pour ton incroyable dévouement aux autres, à ta fonction, à la Science, et au laboratoire. Puissent les années à venir et nos futures collaborations abonder dans le sens qui t'es cher. Ceci est une prière, il n'y a pas de raison qu'elle ne soit pas entendue, et donc exaucée. Car Dieu aide ceux qui s'aident eux-mêmes.

A mon collègue et cher ami **Julien Broséus**. Que dire ici que nous ne nous soyons jamais dit ? Au-delà de tout ce que l'on a pu vivre et traverser ces douze dernières années, et des ascenseurs émotionnels qui nous ont pris, tantôt à tenter de survivre dans ce labo, tantôt à nous investir passionnément dans nos projets, ici à jouer de diplomatie, là de politique et de calculs, tout cela pour une chose : essayer d'avancer comme nous le méritions indubitablement ! Et c'est ce que nous avons fait, par conviction, par intégrité, par amour de ce qui est juste et par acharnement au travail. Merci pour ton répondant dans toute situation, ton amitié, ton incomparable loyauté et ton soutien indéfectible. Puisses-tu avoir la plus longue carrière possible, comme tu le souhaites. Et je l'espère de tout cœur, trouver un peu de réconfort et surmonter tes épreuves actuelles.

A mon collègue et ami **Guillaume Gauchotte**. Merci de m'avoir accordé l'opportunité de m'appliquer sur ces cohortes de tumeurs du SNC. Puis de m'avoir laissé dérouler les travaux dans mon style. La génomique neuro-onco est un dada qui me gratouille, puissions-nous poursuivre avec des projets ambitieux. Merci pour ta gentillesse.

Chers membres du jury, rapporteurs et examinateurs : Madame le Professeur **Karin Tarte**, Madame le Docteur **Anne Friedrich**, Monsieur le Docteur **Hugues Aschard** et Monsieur le Docteur **Olivier Sand**. Merci pour votre temps et votre intérêt pour ce travail. Puissent nos futurs échanges planer plus haut que la sombre nature, dans les stratosphères de la Science.

A mon collègue et ami **Pierre Rouyer**. Merci pour tout ce que tu fais au quotidien et qui soulage le mien. La reconnaissance que tu méritais s'est longuement fait attendre, espérons que la prochaine viendra plus vite !

A mon ex-collègue, rêvons futur collègue, et ami **Jean-Claude Chèvre**. Merci pour ces moments et conversations passionnées et passionnantes qui m'ont élevé. Pour cet instinct scientifique qui te donne une longueur d'avance. Pour ta veille bibliographique qui m'a parfois illuminé. Merci pour ta loyauté et ton amitié. Puissent ces occasions partagées, ainsi que les bons plats et franches rigolades, revenir.

A mes collègues de bureau, étudiants et amis **Romain Piucco**, **Ghislain Fiévet** et **Marion Divoux**. Merci de me supporter au quotidien. Cette situation n'est pas idéale mais elle a le mérite de nous faire partager un petit peu de tout. Comme ce serait égoïste de vous souhaiter le meilleur, je m'abstiens... Godspeed !

Aux **membres de NGERE**, de peur d'en oublier et parce que le va et vient est continuel, je préfère rester vague, mais bien sûr je pense à vous tous. Cela dit, je citerai tout de même quelques piliers solides de mon environnement, de tous bords et de tous temps, sur qui il est (ou était) toujours possible de compter : **Jean-Marc Alberto**, **Djésia Arnone**, **Shyue-Fang Battaglia-Hsu**, **Aline Cazé**, **Pierre Feugier**, **Dominique Guillaume**, **Almut Heinken**, **Isabelle Koscinski**, **Laurent Peyrin-Biroulet**, **Caroline Reppert**, **Catherine Tavera** et **Rémy Umoret**. Merci, collègues, ex-collègues et amis, de contribuer à faire de cet endroit ce qu'il est.

Merci à tous les **scientifiques** et **médecins** avec qui j'ai pu collaborer par le passé, avec qui je collabore aujourd'hui et avec qui collaborerai dans le futur. C'est le pied !

A ma famille,

Rue de la Chèvre, à METZ

Gentile or Jew
O you who turn the wheel and look to windward,
Consider Phlebas, who was once handsome and tall as you.
Thomas Stearns ELIOT

Souviens-toi que le Temps est un joueur avide
Qui gagne sans tricher, à tout coup ! C'est la loi.
Le jour décroît : la nuit augmente, souviens-toi !
Le gouffre a toujours soif ; la clepsydre se vide.
Charles BAUDELAIRE

PUBLICATIONS ET COMMUNICATIONS (PRÉSENTÉES DANS CE MANUSCRIT)

Articles dans des revues

1. **Sébastien Hergalant**, Alicia Grima and David Meyre. Comprehensive identification of pleiotropic associations for serum adiponectin levels using the NHGRI-EBI catalog of published genome-wide association studies. *In preparation*.
2. **Sébastien Hergalant**, Jean-Matthieu Casse, Abderrahim Oussalah, Rémi Houlgatte, Déborah Helle et al. MicroRNAs miR-16 and miR-519 control meningioma cell proliferation via overlapping transcriptomic programs shared with the RNA-binding protein HuR. *Frontiers in Oncology*, 2023, 13. ⟨10.3389/fonc.2023.1158773⟩
3. Julien Broséus*, **Sébastien Hergalant***, Julia Vogt, Eugen Tausch, Markus Kreuz et al. Molecular characterization of Richter syndrome identifies de novo diffuse large B-cell lymphomas with poor prognosis. *Nature Communications*, 2023, 14 (1), pp.309. ⟨10.1038/s41467-022-34642-6⟩
4. **Sébastien Hergalant***, Chloé Saurel*, Marion Divoux, Fabien Rech, Celso Pouget et al. Correlation between DNA Methylation and cell proliferation identifies new candidate predictive markers in meningioma. *Cancers*, 2022, 14 (24), pp.6227. ⟨10.3390/cancers14246227⟩
5. Celso Pouget*, **Sébastien Hergalant***, Emilie Lardenois, Stéphanie Lacomme, Rémi Houlgatte et al. Ki-67 and MCM6 labelling indices are correlated with overall survival in anaplastic oligodendroglioma, IDH1 -mutant and 1p/19q-codeleted: a multicenter study from the French POLA network. *Brain Pathology*, 2020, 30 (3), pp.465-478. ⟨10.1111/bpa.12788⟩

Communications dans des congrès

6. **Sébastien Hergalant**, Romain Piucco, Ghislain Fiévet, Emil Chteinberg, Stephan Stilgenbauer et al. Multi-omics characterization of Richter syndrome unlocks classifiers and predictors of outcome into broader and heterogeneous lymphoma datasets. *JOBIM*, Multisites & Remote (Plouzané, Pointe-à-Pitre, Tours, Nancy, Nice, France). 27-30 Jun 2023, pp.39. ⟨https://jobim2023.sciencesconf.org/data/pages/proceedings.pdf⟩
7. Romain Morizot*, **Sébastien Hergalant***, Romain Piucco, Florian Bouclet, Anne Quinquenel et al. Large-Scale Proteomics Identifies Distinct Signatures for Richter Syndrome and De Novo Diffuse Large B-Cell Lymphoma: A French Study from the Filo Group. *Blood*, 2020, 136 (Supplement 1), pp.29-30. ⟨10.1182/blood-2020-137061⟩

* Contributions égales en tant que 1ers auteurs

TABLE DES MATIÈRES

LISTE DES FIGURES ET DES TABLES

Liste des Annexes

LISTE DES ABRÉVIATIONS

1p19qcodel	codélétion 1p + 19q
ABC	*Activated B-Cell like*
ADN	Acide désoxyribonucléique
AMP	Adénosine monophosphate
AMPK	*AMP-activated Protein Kinase*
AO	*Anaplastic Oligodendroglioma*
ARN	Acide ribonucléique
ASH	*American Society of Hematology*
BCR	*B-Cell Receptor*
BTK	*Bruton Tyrosine Kinase*
CAR-T	*Chimeric Antigenic Receptor - T*
ChIP	*Chromatin ImmunoPrecipitation*
CIRC	Centre International de Recherche sur le Cancer
CNV	*Copy Number Variations*
CLL	*Chronic Lymphocytic Leukemia*
DLBCL	*Diffuse Large B-Cell Lymphoma*
DT2	Diabète de Type 2
EBI	*European Bioinformatics Institute*
eQTL	*expression Quantitative Trait Locus*
FGF	*Fibroblast Growth Factor*
FISH	*Fluorescence* In Situ *Hybridization*
GCB	*Germinal Center B-cell like*
GWAS	*Genome-Wide Association Studies*
HCL	Hôpital Civil de Lyon
HDL	*High-Density Lipoprotein*
HMW	*High Molecular Weight*
HuR	*Human antigen R*
IARC	*International Agency for Research on Cancer*
IDH	Isocitrate déshydrogénase
IDHmt	muté pour IDH
IGF	*Insulin Growth Factor*
IGHV	*Immunoglobulin Heavy chain Variable region*
IL	Interleukine
IMC	Indice de Masse Corporelle
INCa	Institut National du Cancer
JOBIM	Journées Ouvertes de Biologie, Informatique et Mathématiques

Ki-67	Antigène marqueur de prolifération (lymphome de Hodg**ki**n clone **67**)
LBDGC	Lymphome B Diffus à Grandes Cellules
LDL	*Low-Density Lipoprotein*
LLC	Leucémie Lymphoïde Chronique
LMW	*Low Molecular Weight*
MAP	*Mitogen-Activated Protein*
MAPK	*MAP Kinase*
MCM6	*Mini-Chromosome Maintenance 6*
MMW	*Medium Molecular Weight*
mTOR	*mammalian Target Of Rapamycin*
NHGRI	*National Human Genome Research Institute*
OA	Oligodendrogliome anaplasique
OMS	Organisation Mondiale de la Santé
OWAS	*Omics-Wide Association studies*
POLA	Prise en charge des Oligodendrogliomes Anaplasiques
PRC	*Polycomb Repressive Complex*
R-CHOP	Rituximab + Cyclophosphamide, Adriamycine, Vincristine, Prednisone
RNA	*Ribonucleic Acid*
RNA-seq	*RNA Sequencing*
RS	*Richter Syndrome*
SNP	*Single Nucleotide Polymorphism*
SR	Syndrome de Richter
TCGA	*The Cancer Genome Atlas*
TNF	*Tumor Necrosis Factor*
UTR	*Untranslated Region*
VEGF	*Vascular Endothelial Growth Factor*
WCRF	*World Cancer Research Fund*
WHO	*World Health Organization*

INTRODUCTION

Avant-propos : Les maladies du métabolisme

Ce manuscrit traite des « maladies du métabolisme cellulaire » à prendre au sens large, c'est-à-dire toute maladie affectant les métabolismes dans la cellule, héréditaire ou acquise, congénitale ou environnementale. Ces métabolismes touchent les voies biologiques de l'énergie, du contrôle hormonal, de la transduction du signal, du cycle cellulaire, de la maintenance des programmes moléculaires lors du développement et de la différentiation des cellules. Nous aborderons en particulier les aspects génétiques et génomiques des pathologies comme l'obésité et le cancer, et explorerons le lien qui les unit entre elles et aux autres maladies métaboliques d'origines multifactorielles (**Figure 1**).

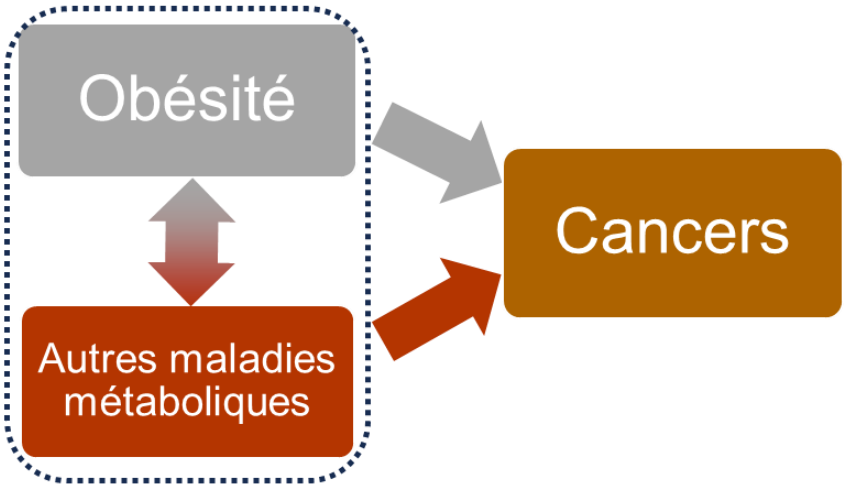


Figure 1. *Représentation schématique du triptyque relationnel « maladies métaboliques / obésité / cancers » tel qu'abordé dans ce document.*

Les omiques et la biomédecine des systèmes

Les maladies du métabolisme, qu'elles soient polygéniques, monogéniques, rares ou communes, sont toutes la cause, d'un point de vue génomique, d'une dérégulation des niveaux moléculaires et voies cellulaires normalement à l'équilibre chez les sujets sains (*1*, *2*). Ces variations d'expression, mesurées à l'échelle du génome et intégrées avec les données phénotypiques des patients, sont les ensembles d'observations permettant d'expliquer, au moins en partie, si l'on considère aussi l'effet de l'environnement, les manifestations de ces maladies. Leur caractérisation passe, entre autres, par l'identification des reprogrammations

épigénétiques, des voies de signalisations impliquées, des carrefours métaboliques, des biomarqueurs et/ou des cibles thérapeutiques potentielles.

Les systèmes biologiques que représentent les multiples couches omiques (par exemple l'exome, le génome, le transcriptome, le méthylome, l'épigénome, le métabolome, le protéome) dressent ainsi la carte d'identité des pathologies, indispensables en médecine de précision (*3*), et apportent une vue holistique du dysfonctionnement cellulaire associé (*4*). Au prix d'une grande complexité dans les données générées : quantité, dimensionalité conséquente, bruit, biais technologiques, multiplication et superposition des signaux, cofacteurs indissociables. Les enjeux sont à la hauteur des défis en fouille de données. Dans cette quête d'extraction de la connaissance, le bioinformaticien s'appuie en premier lieu sur celles établies indubitablement et répertoriées dans les bases de données structurées, contrôlées et hiérarchisées (ontologies) sur les gènes, leurs fonctions, les associations connues avec les maladies, leur régulation et les voies dans lesquels ils interviennent. En génétique des populations, les approches pangénomiques GWAS (*Genome-Wide Association Studies*) ont par exemple permis d'établir un grand nombre d'associations avérées entre polymorphisme génétique, fonction du gène, trait phénotypique et pathologie, posant la première pierre à l'édifice (*2*). Les bases de données se sont considérablement étoffées et démocratisées durant les deux dernières décennies. Leur interopérabilité favorise aujourd'hui une multitude d'interactions complexes, le flux de données phénotypiques (phénomes) et environnementales (exposomes) s'ajustant aux désormais très abondants flux de données biologiques (**Figure 2**).

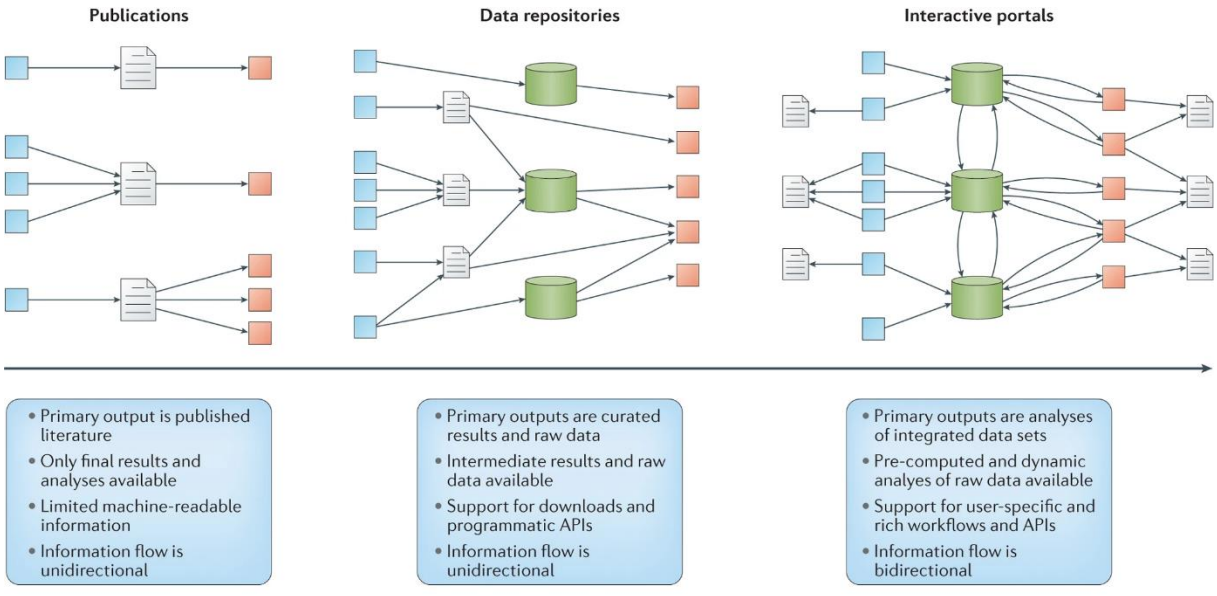


Figure 2*. Evolution des bases de connaissance sur les gènes, et nécessité d'interconnecter les données à tous les niveaux de représentation pour interpréter les maladies et les traits complexes (5).*

Le terme « omique » fait référence à la caractérisation et la quantification de ces flux de données constitués d'un grand nombre de variables :

- Molécules regroupées selon des similitudes biologiques structurelles ou fonctionnelles fondamentales. Il s'agit par exemple des ARN messagers, des ARN non codants, des ARNs dégradés, des protéines, des peptides, des modifications post-traductionnelles, des gènes, des exons, des métabolites comme les lipides, les acides aminés ou les glucides, des génomes, des populations bactériennes représentées par leurs génomes, des flux moléculaires. Dans ces conditions, les protéomes représenteront par exemple toutes les protéines et/ou peptides d'une ou plusieurs cellules, techniquement capturables et mesurables à un instant donné dans des conditions expérimentales données. Les exomes identifieront, eux, tous les variants génétiques associés à un trait ou une maladie et touchant les régions codantes sur le génome. Ces flux biologiques donnent par conséquent une vue, que l'on espère représentative, du fonctionnement cellulaire/tissulaire à chaque niveau moléculaire (*6*, *7*).
- Phénotypes cliniques, biologiques et sémiologiques (par exemple l'âge, le sexe, l'ethnicité, la pathologie étudiée, les antécédents cliniques possibles). Indispensables pour la compréhension des mécanismes sous-jacents des maladies du métabolisme considérées.
- Données relatives à l'environnement (ou d'exposition, comme les polluants, les pesticides, les zones à risque, le mode de vie, l'activité physique, les addictions) et à la socio-démographie (comme le niveau d'éducation, la profession, la situation familiale, les revenus). Pour élucider l'étiologie et pour les études d'interaction gène x environnement.

Les aspects techniques et méthodologiques concernant les différentes analyses omiques sont pléthore. Ceux-ci étant bien détaillés dans les matériels et méthodes, résultats, et parties introductives des différents articles présentés dans ce manuscrit, nous ne les aborderons pas dans ce chapitre.

L'intégration des données omiques entre elles est nécessaire pour mieux comprendre la complexité des mécanismes moléculaires et leurs dérégulations dans les pathologies. Prises isolément, les analyses mono-omiques, bien qu'indispensables en premier lieu, ne fournissent pas suffisamment d'informations pour répondre pleinement à ces enjeux. Ce que proposent la biologie et la médecine des systèmes, c'est d'interpréter les variations et les différents changements observés à chaque niveau étudié comme faisant partie d'un ensemble interactif et

fonctionnel (*4*). L'intégration « multi-omique » des données (**Figure 3**) peut se faire en *cis*, en traversant verticalement les couches moléculaires (qui concerne les mêmes variables sur les différents niveaux, comme un gène, l'ARN messager et la protéine correspondante), ou en *trans*, en considérant les interactions horizontales et transversales (en suivant des voies cellulaires, des molécules cibles, des mécanismes de régulations, des complexes moléculaires), et permet principalement :

i. De réduire le bruit. Nous parlons ici du bruit de fond, des signaux secondaires ou indirects, des artefacts et des faux positifs. Le signal final est concentré sur ce qui se répercute d'un niveau à l'autre.
ii. D'améliorer la précision dans les données complexes. Les omiques sont majoritairement des « *Big Data* » de haute complexité (*8*).
iii. De donner un aperçu des réseaux cellulaires. Comme les voies métaboliques, la transduction du signal et les réseaux de régulation ou d'interaction.
iv. De mettre en évidence les mécanismes fonctionnels centraux potentiellement impliqués dans le destin cellulaire.

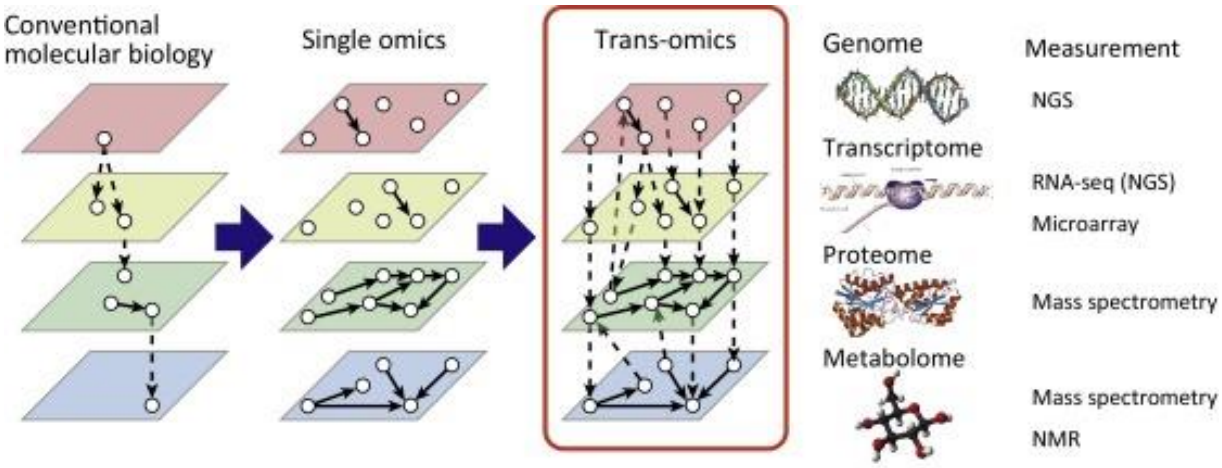


***Figure 3**. Réseau trans-omique s'appuyant sur de multiples couches omiques : génome, transcriptome, protéome, métabolome. Le même concept s'applique en cis (corrélomiques) et à n'importe quel autre omique : épigénome, méthylome, phénome, exposome, régulome, miRnome (9).*

Les cancers

Le cancer est une maladie de la division cellulaire, qui intervient après l'acquisition d'anomalies moléculaires provoquant une transformation de cellules anormales et déclenchant leur multiplication excessive. La cancérogénèse résulte d'une accumulation de plusieurs altérations irréversibles selon 3 grandes phases successives : initiation, promotion et prolifération. Les cancers rassemblent des pathologies très diverses de formes et de

conséquences, mais partagent cependant un ensemble fonctionnel de caractéristiques acquises (attributs ou « *hallmarks* », ou encore endophénotypes) leur permettant de se développer, de s'adapter et de survivre dans leur environnement aux dépends des mécanismes naturels de l'organisme (*10*).

La prévalence des cancers varie fortement dans la population, ce sont des maladies typiquement multifactorielles alliant terrain individuel et histoire personnelle (compte-rendu de l'Académie des Sciences (*11*)). Mutations monogéniques héréditaires ou sporadiques touchant des localisations chromosomiques préférentielles, vieillissement, style de vie, maladies préexistantes, facteurs hormonaux et susceptibilités polygéniques sont tous en cause à divers degrés dans la cancérogénèse. Nous ne sommes pas tous égaux face au cancer, mais notre exposition aux facteurs de risques biologiques et environnementaux jouerait au premier plan : 5-10% des cas sont attribués à des facteurs génétiques uniquement contre 25-30% au tabagisme, 30-35% aux régimes alimentaires et aux comportements (alcoolisme, dépendances, obésité, consommation et cuisson des viandes), 15-20% aux infections, et 10-25% à d'autres facteurs environnementaux (stress, sédentarité, pollutions, rayons ionisants) (*12*).

Plusieurs types de cancers sont en augmentation à travers le monde, avec de fortes disparités géographiques (*13*). Celle-ci serait surtout la résultante d'un accroissement et du vieillissement de la population, tous facteurs confondus (*14*). Alors, que l'incidence de certains cancers augmente, notamment du poumon, de l'ovaire, des testicules, les tumeurs cérébrales et les hémopathies malignes, elle diminue pour d'autres, par exemple le cancer du col de l'utérus, de l'œsophage et de l'estomac (incidence des principaux cancers en France métropolitaine en 2023 et tendances depuis 1990 ; sources : Institut National du Cancer (INCa) / Santé publique France / réseau FRANCIM / Hospices Civils de Lyon (HCL), bulletin épidémiologique hebdomadaire, numéro 12-13). Les répercussions des modifications de l'environnement et des modes de vie sur l'augmentation de l'incidence de certains cancers sont devenues une préoccupation majeure de santé publique à l'échelle mondiale.

Selon l'OMS (Organisation Mondiale de la Santé), le cancer est la deuxième cause de décès dans le monde entier et a fait 9,6 millions de morts en 2018, soit un décès sur six (www.who.int). En France, les cancers représentaient en 2005 la première cause de mortalité chez les hommes et la deuxième chez les femmes, avec des estimations à 320 000 nouveaux cas de cancers par an, 180 000 chez les hommes et 140 000 chez les femmes. Si l'on tient compte des évolutions démographiques, l'incidence des cancers a progressé de 48 % chez l'homme et de 46 % chez la femme entre 1980 et 2005 (*15*). En 2023, c'est 433 000 nouveaux cas (57% d'hommes) qui ont été enregistrés, toutes localisations confondues, avec un taux d'incidence stabilisé chez l'homme depuis 2012, mais qui progresse continuellement chez la femme (+ 0,9% par an ; sources : INCa / Santé publique France / FRANCIM / HCL).

La démocratisation des dépistages et l'exploitation des connaissances sur les marqueurs sériques ou sanguins a conséquemment provoqué l'augmentation de l'incidence du risque dans les populations où ces pratiques sont rendues possibles pour tous (*13*).

Génomique des cancers

Outre l'indentification de biomarqueurs qui soient facilement réutilisables en clinique pour le diagnostic et le pronostic, les grands défis de la génomique en cancérologie concernent :

i. La dissection des endophénotypes (**Figure 4**). Les voies cellulaires associées à la pathogénèse des cancers sont d'une importance capitale à déchiffrer bioinformatiquement et à valider mécanistiquement.
ii. La découverte de sous-types génétiques / épigénétiques. La meilleure approche est multi-omique, avec des méthodes non-supervisées (sans *a priori* sur les données, par exemple des classifications hiérarchiques, des k-moyennes ou des réductions de dimensions) sur des cohortes suffisamment grandes pour confirmer l'existence d'un sous-groupe, même minoritaire. C'est le principe même de la médecine de précision qui est ciblé ici.
iii. La (re-)classification des tumeurs selon des critères moléculaires avancés. Les profils histologiques à la détermination subjective, les marqueurs de surface inadaptés ou les signatures géniques datées et appauvries constituent des limitations au diagnostic et à la prise en charge thérapeutique.
iv. L'identification de la cellule d'origine du cancer. Il est nécessaire pour cela de s'appuyer sur de multiples groupes de références (tissus / cellules à plusieurs stades d'un développement normal ou même pathologiques) pour lesquels les cellules d'origines sont bien identifiées, ou de faire des explorations en « *single-cell* », bien que cette dernière technique ne garantisse pas la réussite à elle-seule.
v. La détection des vulnérabilités thérapeutiques. L'identification de médicaments existants qui soient recyclables dans la pathologie, ou l'identification de carrefours moléculaires aisément ciblables, vulnérables et potentiellement saufs pour les tissus normaux, sont des résultats particulièrement scrutés car à retombées immédiates. Disposer du protéome ou du métabolome est d'un intérêt tout particulier pour y parvenir.
vi. Le déchiffrage de l'agressivité et de la prolifération cellulaire. Il s'agit ici de comprendre comment la cellule débloque et détourne le cycle cellulaire, non pas comme un des endophénotypes du cancer, mais pour identifier des biomarqueurs prédictifs de la survie des patients.

vii. La compréhension des mécanismes physiopathologiques sous-jacents à la progression et/ou la transformation de la maladie. Il s'agit ici aussi d'identifier des marqueurs précoces de mauvais pronostic afin d'anticiper et d'améliorer la prise en charge des patients.

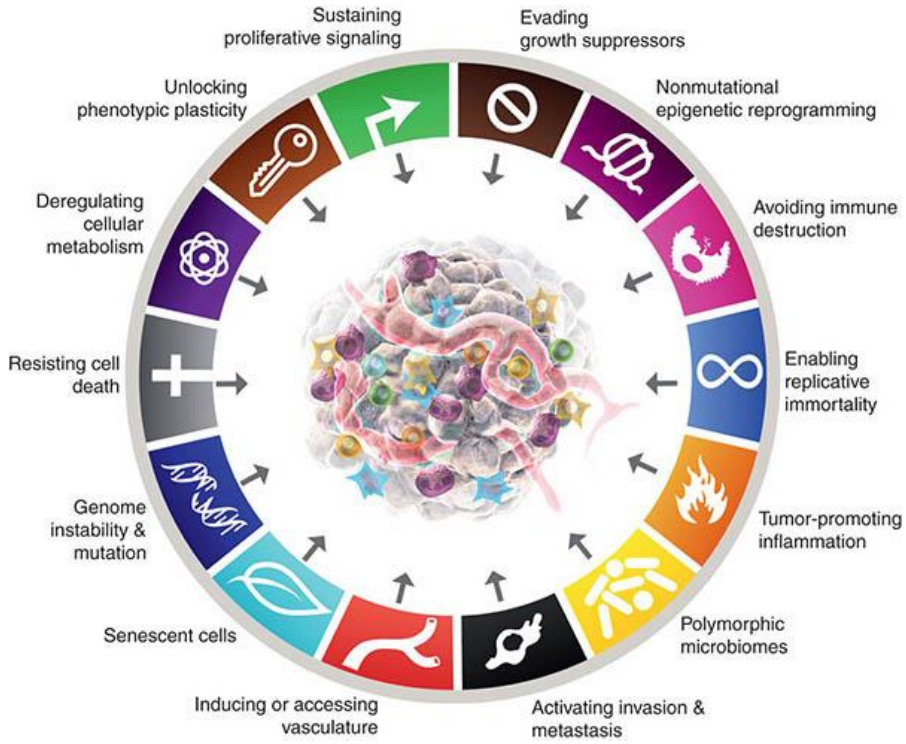


Figure 4. *Les 14 caractéristiques (endophénotypes) des cancers, selon la dernière révision de 2022 (10).*

La cellule cancéreuse reprogramme son épigénome comme un mécanisme rapide de modification de l'expression des gènes qui va lui permettre de survivre et de proliférer, d'échapper au système immunitaire et de s'adapter à son environnement : des endophénotypes de la cancérogénèse (*10*). Une empreinte plus durable est ensuite visible sur les niveaux de méthylation de régions fonctionnelles sur l'ADN : le méthylome. En règle générale, durant le développement normal des tissus, les niveaux de méthylation et les marques de la chromatine changent, et sont très souvent corrélés entre eux et à l'action de facteurs de transcription spécifiques (*16*). En fonction du stade de développement cellulaire, des tumeurs très différentes peuvent émerger, mais qui présentent toutes un lien avec leur programme épigénétique d'origine, même si celui-ci est largement modifié (différentiel) par rapport à la cellule normale (**Figure 5**) (*17*). Dans cette mise en scène particulière, la compréhension physiologique et mécanistique de la reprogrammation épigénétique est rendue possible en intégrant différents omiques successifs : le méthylome, l'épigénome et le transcriptome.

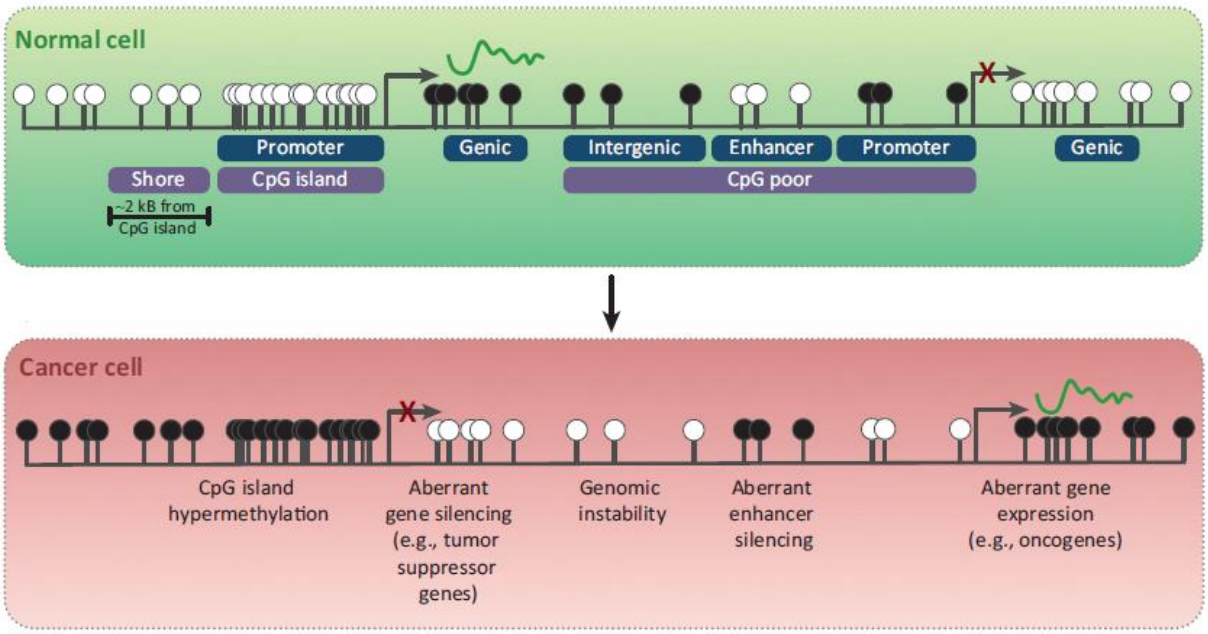


Figure 5. *Vue simplifiée de la méthylation (telle qu'on peut la mesurer par puces à ADN), et de sa reprogrammation dans les cellules cancéreuses (17).*

Traverser les couches omiques facilement corrélables entre elles, car liées directement, est une technique largement utilisée pour comprendre les dérégulations moléculaires qui se répercutent d'un niveau à l'autre dans les cancers, où les reprogrammations sont généralement franches et agressives (**Figure 6**). La corrélomique est une méthode multi-omique corrélative fonctionnant en *cis*, où chaque gène est corrélé sur ses différentes couches moléculaires (par exemple son niveau d'expression génique *versus* protéique), ce qui permet de se concentrer sur les effets différentiels conservés – donc sur une forme de régulation linéaire – et d'éliminer le bruit de fond inhérent aux modes mono-omiques, et réduit sensiblement la complexité des données (*8*). Les corrélomiques cherchent à calquer au plus proche les cas d'école des dynamiques homo-directionnelles ou inversées entre les niveaux moléculaires (par exemple : une hypométhylation dans une région régulatrice mène à une surexpression du gène sous le contrôle de cette région, qui mène à son tour à une surexpression de son produit. La corrélation est respectivement négative puis positive) (*18*).

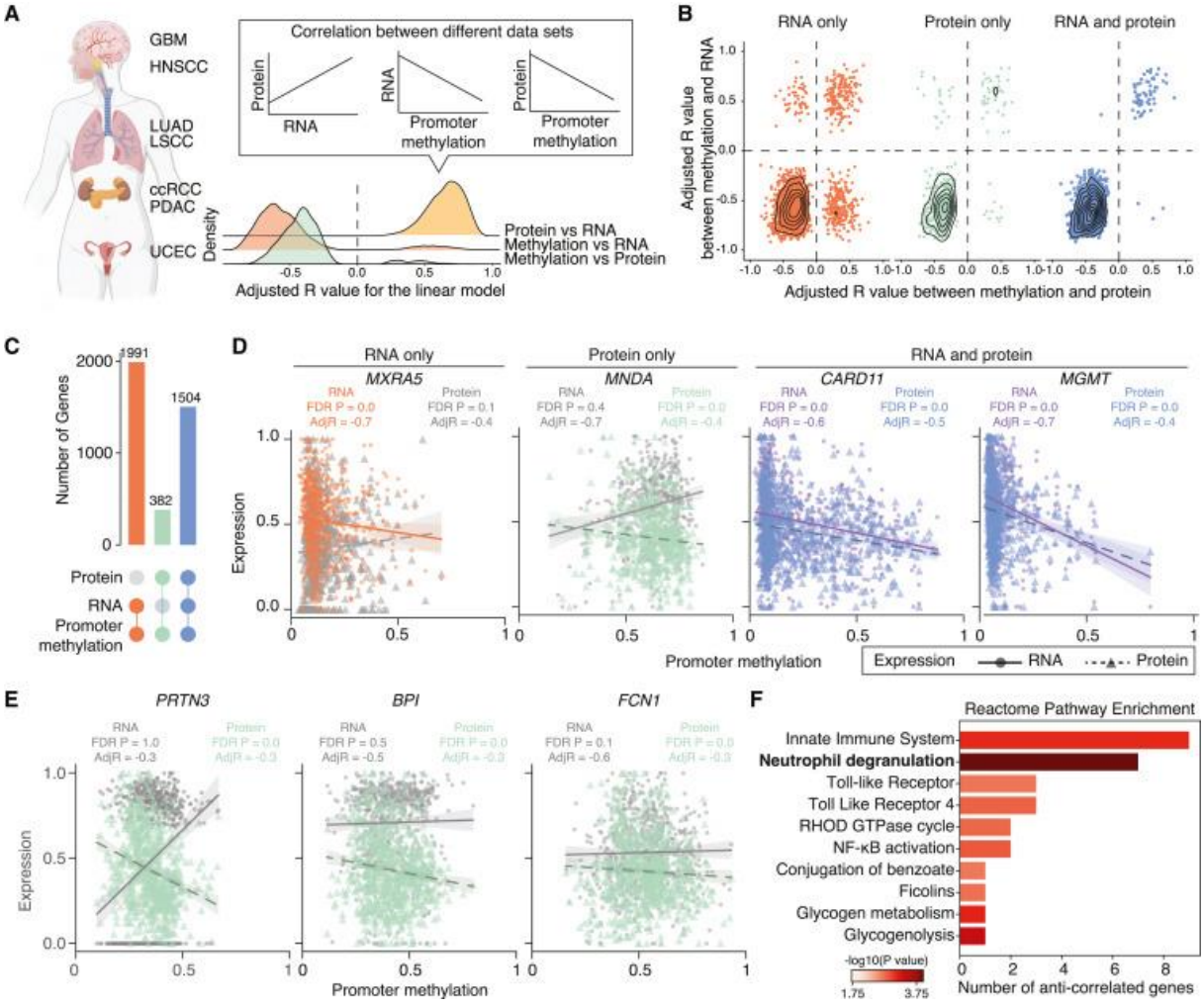


Figure 6. *Etude pan-cancer par la méthode multi-omique corrélative (corrélomique) sur le méthylome, le transcriptome et le protéome de patients souffrant de cancers solides. De la présentation générale de la cohorte et des données (A, B et C), en zoomant sur les gènes les plus remarquables (D et E), aux annotations fonctionnelles des voies issues de la base de données Reactome (F) (18).*

Les hémopathies malignes

Les hémopathies malignes regroupent un ensemble hétérogène de cancers des cellules sanguines et de leurs précurseurs. Parmi cet ensemble, on distingue les leucémies, les syndromes myélodysplasiques et les lymphomes. Dans ce manuscrit, nous nous focaliserons sur les lymphomes non hodgkiniens – les plus fréquents – de la lignée B, notamment la leucémie lymphoïde chronique (LLC), les lymphomes B diffus à grandes cellules (LBDGC), et le syndrome de Richter (SR), une transformation dramatique de la LLC en LBDGC secondaire.

Les lymphomes B

Les lymphomes sont des cancers se développant aux dépends des structures lymphoïdes secondaires, notamment les ganglions lymphatiques. Leur incidence est en hausse constante depuis des années (estimations nationales de l'incidence et de la mortalité par cancer en France

métropolitaine entre 1990 et 2018 ; source : INCa / Santé publique France / FRANCIM / HCL, hémopathies malignes, volume 2). Il s'agit en réalité d'un ensemble très hétérogène avec de très nombreux types histologiques, une présentation clinique très variable avec des formes de pronostic très différent et une prise en charge médicale allant de l'abstention thérapeutique et la surveillance à l'hospitalisation en urgence.

Dans les lymphomes B, la reprogrammation épigénétique touche presque toutes les régions du génome et l'on y observe une hypométhylation globale des régions intergéniques par rapport aux lymphocytes B normaux, ainsi qu'une hyperméthylation des régions correspondant à la chromatine bivalente, cibles des répresseurs transcriptionnels polycomb (PRC pour *Polycomb Repressive Complex*), impliqués dans le développement et la différenciation cellulaire (**Figure 7**). Cette mécanistique est liée à l'histoire proliférative de l'établissement des lymphomes, et des cancers en général (elle en constitue un des endophénotypes). Les promoteurs sont très majoritairement hypométhylés, et les activateurs sont hypo- ou hyperméthylés spécifiquement en fonction des facteurs de transcription qui les ciblent (*19*).

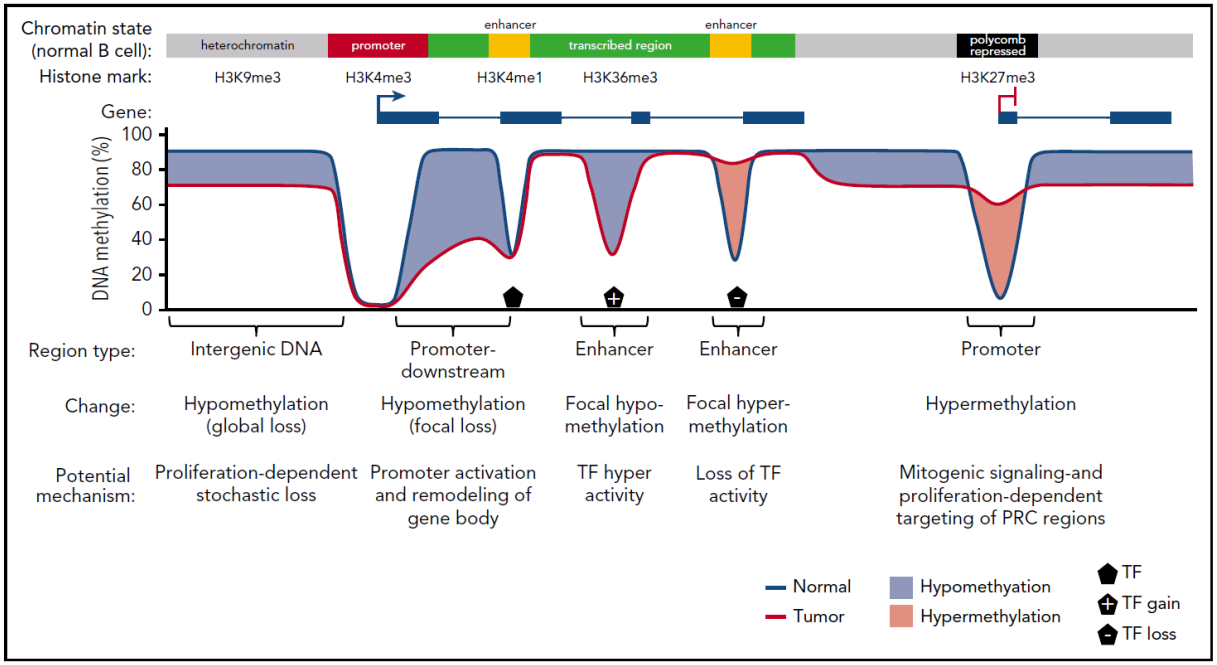


Figure 7. *Altérations de l'épigénome dans les lymphomes B : méthylation de l'ADN, états de la chromatine et mécanistique proposée (19).*

Développement des lymphomes B

Ce sont les lymphomes les plus fréquents. On oppose les lymphomes B à petites cellules, d'évolution clinique souvent indolente, aux lymphomes B à grandes cellules, qui ont systématiquement une présentation agressive, nécessitant une prise en charge thérapeutique

rapide. Devant le nombre important de sous-types différents de lymphomes B, les premières classifications reposaient tout d'abord sur l'identification de la cellule d'origine au sein de la lignée lymphocytaire B normale et sur sa localisation géographique au sein du ganglion. Parmi les lymphomes B indolents, la LLC présente une des incidences les plus élevées (5000 nouveaux cas par an en France), et se singularise par la présence constante d'une phase circulante sanguine, c'est-à-dire de la circulation sanguine de petits lymphocytes clonaux. Le LBDGC constitue la forme de lymphome agressif la plus fréquente et prend naissance au niveau du centre germinatif (**Figure 8**) (*20*).

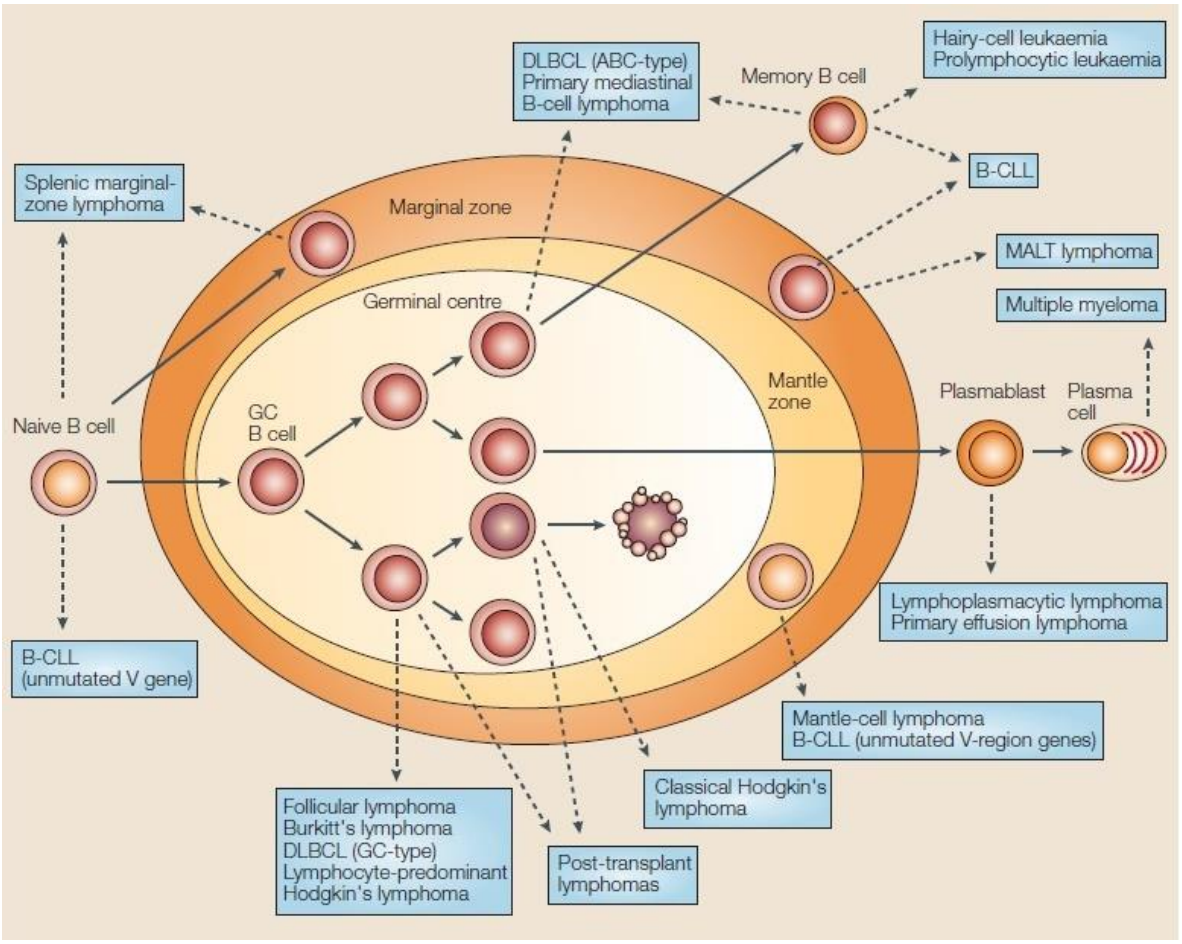


Figure 8. *Aperçu des différents sous-types de lymphomes B, distribués selon leur zone anatomique d'origine (20).*

La lymphopoïèse B, le processus physiologique de fabrication des lymphocytes B, comporte deux étapes de maturation, la première au sein de la moelle osseuse, la deuxième au sein des structures lymphoïdes secondaires. Le schéma de maturation des lymphocytes B est intimement lié à l'évolution des gènes codant pour l'immunoglobuline de surface, notamment le gène *IGHV*, encodant la partie variable de sa chaîne lourde. La genèse d'un large panel de lymphocytes B, destiné à couvrir la diversité antigénique de l'environnement, est due à un mécanisme de réarrangement du gène *IGHV* à l'origine d'une grande diversité combinatoire. L'étape médullaire permet la synthèse d'un vaste répertoire de lymphocytes B naïfs, tous

uniques car présentant un récepteur de surface BCR (*B-Cell Receptor*) spécifique à chacun, notamment par la partie hypervariable CDR3 de l'immunoglobuline de surface. Ce BCR est constitué par un anticorps de surface combiné à une machinerie de transduction du signal. En cas de réponse immunitaire, les lymphocytes B présentent une affinité pour l'antigène, et le gène *IGHV* subit alors une étape additionnelle d'hypermutations somatiques pour augmenter encore cette affinité. Ces hypermutations sont le reflet du passage ganglionnaire du lymphocyte B, et donc de son niveau de maturité, opposant les lymphocytes B naïfs préganglionnaires et donc « non-mutés pour *IGHV* », aux lymphocytes B mémoires, dont le gène *IGHV* a subi l'étape d'hypermutations somatiques.

Au cours de la lymphopoïèse B, la méthylation globale de l'ADN diminue, un processus corrélé à la méthylation des régions régulatrices cibles importantes dans la maturation des cellules, et aux marques de la chromatine (modifications post-traductionnelles des histones) associées. Comme vu plus haut, les régions réprimées par le PRC sont, elles, négativement corrélées à ce phénomène. Les étapes incrémentales de ces changements sont placées sous le contrôle de facteurs de transcription spécifiques du développement précoce, médian, ou tardif des lymphocytes B (**Figure 9**) (*16*).

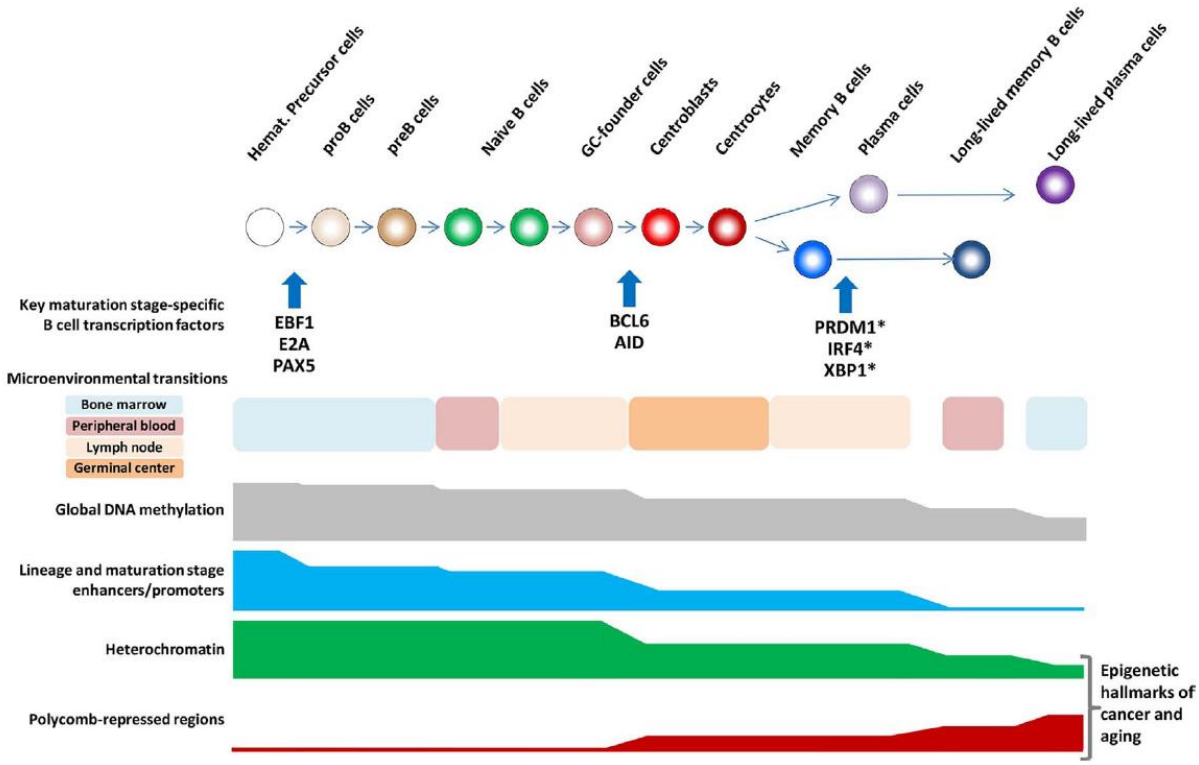


Figure 9. *Schéma intégratif de la différentiation des lymphocytes B, incluant les facteurs de transcription majeurs, les modifications du microenvironnement et les dynamiques épigénétiques (16).*

La leucémie lymphoïde chronique (LLC)

Définition et épidémiologie

La LLC est une pathologie du sujet âgé, avec une médiane d'âge de 72 ans au diagnostic (*21*). C'est la forme de leucémie la plus fréquente dans les pays occidentaux, avec une incidence annuelle de près de 5000 cas par an en France (estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018 ; source : Santé publique France / Francim / HCL / INCa, hémopathies malignes, volume 2). La LLC est un lymphome B à petites cellules comportant par définition une phase circulante sanguine (hyperlymphocytose) justifiant l'appellation de « leucémie lymphoïde ». Le point de départ de la maladie est donc ganglionnaire. Il est suivi du passage sanguin d'une population homogène de petits lymphocytes matures clonaux avec un faible potentiel de prolifération et donc une évolution clinique lente, d'où le terme de leucémie lymphoïde « chronique ». Le lymphome lymphocytique est une entité partageant l'essentiel des caractéristiques biologiques et cliniques de la LLC, à l'exception de l'absence de phase circulante sanguine. Il s'agit d'une forme purement ganglionnaire de la maladie.

Diagnostic et évolution clinique

Le diagnostic repose sur l'objectivation d'une hyperlymphocytose sanguine et l'identification morphologique et cytométrique de caractères de malignité. Les lymphocytes B malins circulants constituent une population monomorphe de petits lymphocytes matures (**Figure 10**). A l'opposé des autres sous-types de lymphomes dont le diagnostic repose sur une biopsie ganglionnaire, le diagnostic de la LLC peut être réalisé avec un simple prélèvement sanguin puisque les lymphocytes malins circulants présentent un phénotype caractéristique, identifiable par cytométrie en flux. Le diagnostic de LLC repose sur l'identification d'une expansion clonale de lymphocyte B avec expression des marqueurs CD5 et CD23, une faible expression de l'immunoglobuline de surface κ ou λ, la faible expression de FMC7 et de la chaîne accessoire du récepteur pour l'antigène CD79b, selon un score d'au moins 4 pour cet ensemble de 5 critères (score de Matutes) (*22*).

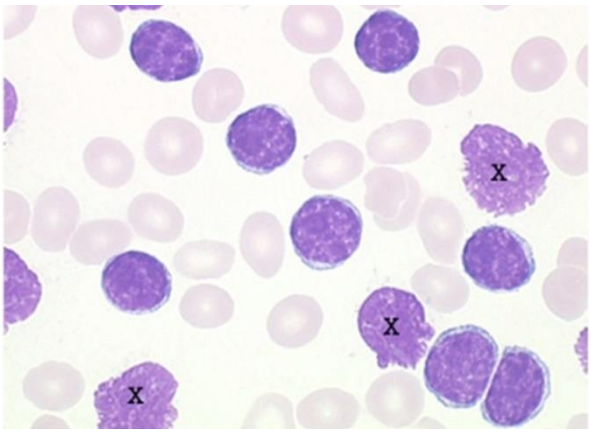


Figure 10. *Aperçu d'un frottis sanguin coloré au May-Grünwald-Giemsa et observé au microscope (objectif x100).*

La LLC constitue en réalité un ensemble hétérogène au regard de l'évolution clinique, avec des formes stables durant des années, à l'opposé de formes très rapidement évolutives et agressives. Cette hétérogénéité est sous-tendue par la diversité des caractéristiques biologiques sous-jacentes.

Hétérogénéité cytogénétique et moléculaire

Les LLC sont réparties entre 2 sous-types moléculaires majeurs qui diffèrent par leur degré d'hypermutations somatiques du gène *IGHV*. Comme dit plus haut, les hypermutations somatiques constituent un processus physiologique de maturation des lymphocytes B prenant place dans le centre germinatif des structures lymphoïdes secondaires. Les LLC *IGHV* non mutées, dont la séquence du gène *IGHV* partage plus de 98% d'homologie avec la séquence germinale, dérivent d'un lymphocyte B naïf pré-ganglionnaire et sont associées à un mauvais pronostic comparativement aux LLC *IGHV* mutées, issues de la transformation néoplasique d'un lymphocyte B mémoire. Certains BCR sont dits « stéréotypiques », c'est-à-dire qu'ils sont présents dans les lymphocytes B malins chez une proportion de patients beaucoup plus élevée que les statistiques le prévoient, indiquant une distribution non aléatoire, probablement due à une sélection de certaines sous-populations lymphocytaires par une stimulation antigénique chronique. Environ 1/3 des LLC présentent un BCR stéréotypique, dont certains sont associés à un mauvais pronostic (*23*).

La FISH (*Fluorescence In Situ Hybridization*) a permis d'identifier les principales anomalies cytogénétiques associées à la LLC. 80% d'entre elles sont associées à au moins une des 4 principales anomalies décrites : (i) la délétion 13q, qui emporte le cluster miRNA 15a/16-1 ; (ii) la délétion 11q qui englobe les gènes *ATM* et *BIRC3*, (iii) la délétion 17p, qui emporte notamment le gène *TP53* et (iv) la trisomie 12. Ces anomalies définissent différents sous-groupes pronostiques (*24*).

Le paysage mutationnel de la LLC est lui-aussi très hétérogène, avec un large spectre de mutations affectant différentes fonctions cellulaires, telles que: (i) la réponse aux dommages à l'ADN et le contrôle du cycle cellulaire (*TP53*, *ATM*, *POT1*, *ATRX*), (ii) la maturation et l'exportation des ARN (*SF3B1*, *XPO1*, *RPS15*, *DDX3X*, *ZNF292*, *MED12*, *NXF1*), (iii) la voie NOTCH (*NOTCH1*, *FBXW7*), (iv) la voie du BCR (*EGR2*, *KLHL6*, *BCOR*, *IRF4*, *IKZF3*, *ITKB*, *CARD11*), (v) la physiologie chromatinienne (*CHD2*, *BAZ2A*, *SETD2*, *ASXL1*, *ZMYM3*, *HIST1H1E*, *ARID1A*), (vi) la voie NFκB (*BIRC3*, *MYD88*, *TRAF3*, *NFKB1E*), (vii) la réponse inflammatoire (*SAMHD1*, *RIPK1*), (viii) la développement lymphocytaire B (*IKZF3*, *PAX5*), (ix) la voie MAPK-ERK (*MAPK*, *MAP2K1*, *ERK*, *BRAF*, *KRAS*) et (x) la signalisation associée à MYC (*MGA*, *PTPN11*) (**Figure 11**) (*25*). La distribution et la fréquence de ces anomalies sont différentes entre les LLC non-mutées pour *IGHV* et les LLC mutées pour *IGHV* (*23*).

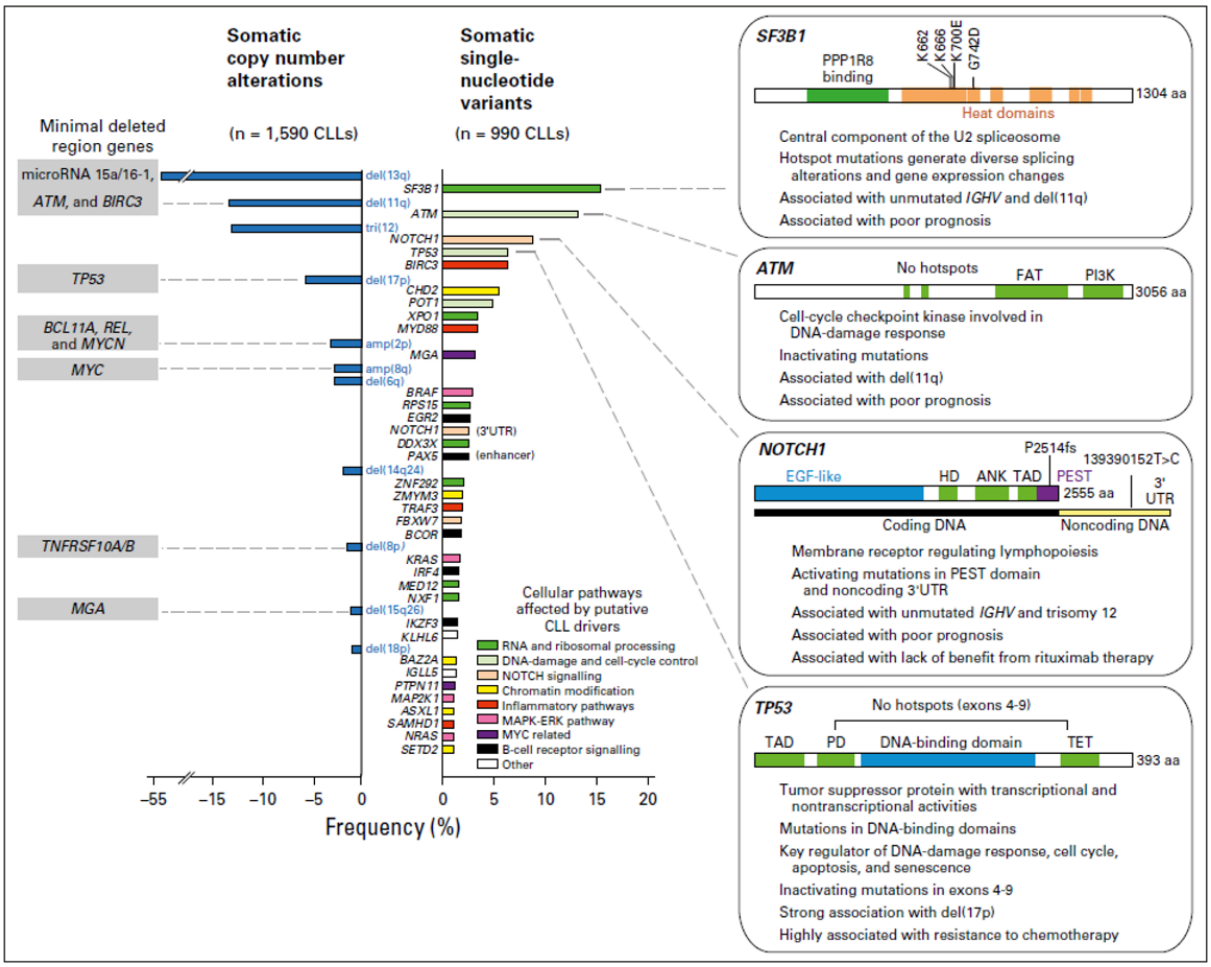


Figure 11. *Panorama mutationnel de la LLC, avec fréquence des principales anomalies cytogénétiques et moléculaires (25).*

Prise en charge thérapeutique

L'immunochimiothérapie par Rituximab-Fludarabine-Cyclophosphamide est devenu en 2010 le traitement de référence en première intention chez les patients de moins de 65 ans et/ou en bon état général et atteints d'une LLC sans délétion 17p ni mutation de *TP53* (*26*). Les progrès scientifiques récents ont permis d'identifier les bases physiopathologiques de la LLC

et ont permis l'avènement récent de traitements sans chimiothérapie, ciblant les voies de signalisation qui sont à l'origine de la prolifération tumorale et de la résistance à l'apoptose.

Les inhibiteurs de BTK (*Bruton Tyrosine Kinase*), qui bloquent le signal prolifératif du BCR, ou les inhibiteurs de la protéine anti-apoptotique BCL-2, ont considérablement amélioré la prise en charge. Les inhibiteurs de BTK permettent de restaurer l'apoptose dans les cellules malignes, et de les forcer à quitter leur environnement ganglionnaire protecteur d'origine. Les essais cliniques montrent d'excellents résultats dans le traitement des LLC réfractaires ou en rechute après l'immunochimiothérapie conventionnelle (*27*), mais aussi en traitement de première ligne (*28*). L'expression élevée de la protéine anti-apoptotique BCL-2 est un mécanisme de résistance tumoral fréquent dans la LLC. Le Venetoclax est un inhibiteur oral hautement spécifique de la protéine anti-apoptotique BCL-2. Il induit la mort cellulaire en reproduisant l'effet du domaine BH3. L'efficacité de ce médicament a été démontrée en traitement des LLC de haut risque et chimiorésistantes (*29*). Contrairement aux traitements conventionnels par chimiothérapie, ces nouveaux traitements ciblés présentent moins de toxicité et ont l'intérêt d'être très bien tolérés par les malades âgés et/ou présentant des comorbidités.

Les lymphomes B diffus à grandes cellules (LBDGC)

Les LBDGC sont les lymphomes agressifs les plus fréquents, et représentent 30 à 40% des lymphomes malins non Hodgkiniens. Sur le plan histologique, les LBDGC se présentent sous la forme de grandes plages continues de centroblastes et/ou d'immunoblastes (**Figure 12**).

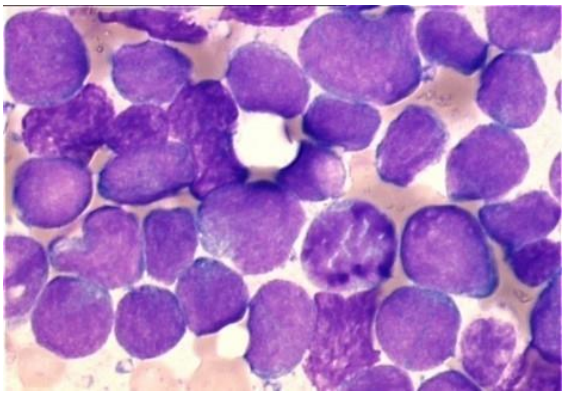
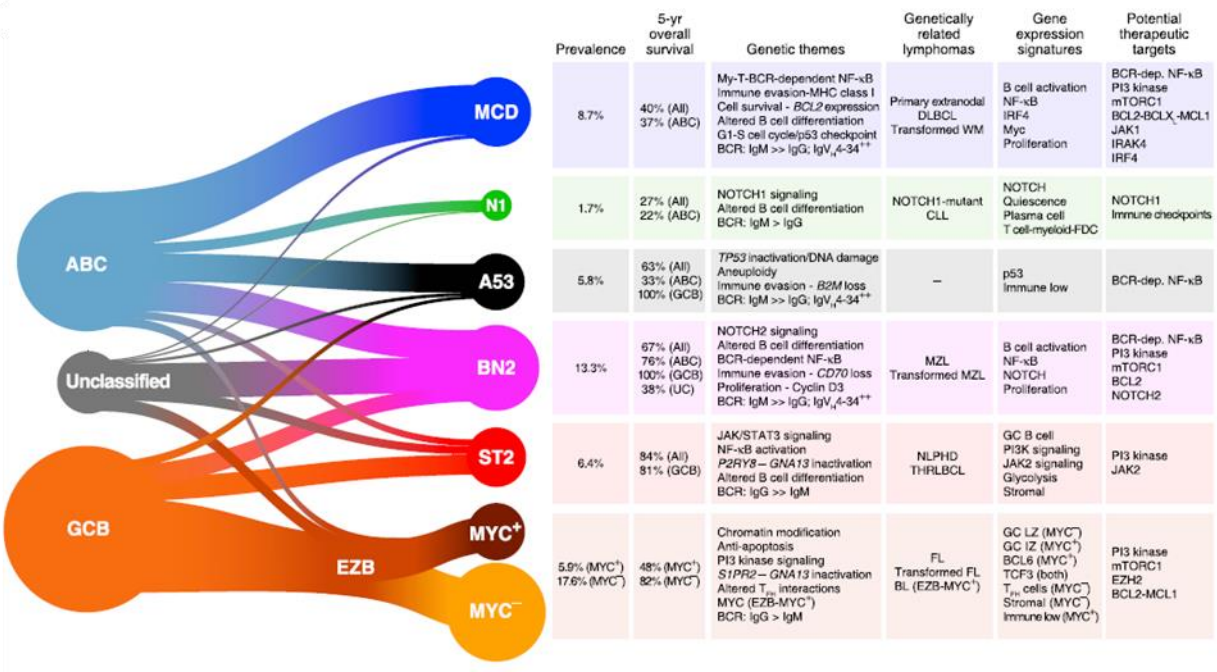


Figure 12. *Présentation morphologique d'un LBDGC sur une coupe de tissu ganglionnaire.*

Bien que non différentiable morphologiquement, il est essentiel de faire la distinction entre (i) les LBDGC primitifs, constituant une forme native de lymphome B agressif survenant sur un terrain vierge de toute pathologie lymphocytaire B clonale antérieure et (ii) les LBDGC secondaires, qui constituent une évolution tumorale péjorative d'un lymphome B à petites cellules préexistant.

Les LBDGC primitifs (*de novo*)

Ils constituent un ensemble très hétérogène sur le plan clinique, moléculaire et pronostique. Les techniques récentes de génomique ont permis de mieux comprendre l'hétérogénéité biologique et la physiopathologie de la maladie. Les 2 sous-types majeurs de LBDGC primitifs décrits se distinguent par leur cellule d'origine, qui est attestée par leurs signatures transcriptomiques. Ainsi, les LBDGC de type GCB (*Germinal Center B-cell like*) présentent un profil d'expression génique proche de celui des lymphocytes B normaux du centre germinatif et les LBDGC de type ABC (*Activated B-Cell like*) ont un profil dérivant de lymphocytes B activés après rencontre avec l'antigène (*30*). Le panorama mutationnel de ces deux sous-types de LBDGC a été caractérisé par la suite sur de grandes cohortes de patients. Les LBDGC de type GCB sont liés à des anomalies mutationnelles affectant la voie PI3K/Akt/mTOR, la régulation épigénétique, BCL2, BCL6, MYC, TP53 et GNA13. Les LBDGC de type ABC présentent un panorama mutationnel à l'origine d'une activation constitutionnelle des voies du BCR et des toll-like receptors, et *in fine* du facteur de transcription NF-κB (*31, 32*). Des travaux récents consacrés à la génomique des LBDGC primitifs ont permis d'affiner leur classification au-delà du statut GCB/ABC, sur la base des profils mutationnels et de l'évolution clinique (**Figure 13**) (*33-36*).



	Prevalence	5-yr overall survival	Genetic themes	Genetically related lymphomas	Gene expression signatures	Potential therapeutic targets
MCD	8.7%	40% (All) 37% (ABC)	My-T-BCR-dependent NF-κB Immune evasion-MHC class I Cell survival - *BCL2* expression Altered B cell differentiation G1-S cell cycle/p53 checkpoint BCR: IgM >> IgG; IgV$_H$4-34^{++}	Primary extranodal DLBCL Transformed WM	B cell activation NF-κB IRF4 Myc Proliferation	BCR-dep. NF-κB PI3 kinase mTORC1 BCL2-BCLX$_L$-MCL1 JAK1 IRAK4 IRF4
N1	1.7%	27% (All) 22% (ABC)	NOTCH1 signaling Altered B cell differentiation BCR: IgM > IgG	NOTCH1-mutant CLL	NOTCH Quiescence Plasma cell T cell-myeloid-FDC	NOTCH1 Immune checkpoints
A53	5.8%	63% (All) 33% (ABC) 100% (GCB)	*TP53* inactivation/DNA damage Aneuploidy Immune evasion - *B2M* loss BCR: IgM >> IgG; IgV$_H$4-34^{++}	–	p53 Immune low	BCR-dep. NF-κB
BN2	13.3%	67% (All) 76% (ABC) 100% (GCB) 38% (UC)	NOTCH2 signaling Altered B cell differentiation BCR-dependent NF-κB Immune evasion - *CD70* loss Proliferation - Cyclin D3 BCR: IgM >> IgG; IgV$_H$4-34^{++}	MZL Transformed MZL	B cell activation NF-κB NOTCH Proliferation	BCR-dep. NF-κB PI3 kinase mTORC1 BCL2 NOTCH2
ST2	6.4%	84% (All) 81% (GCB)	JAK/STAT3 signaling NF-κB activation *P2RY8 – GNA13* inactivation Altered B cell differentiation BCR: IgG >> IgM	NLPHD THRLBCL	GC B cell PI3K signaling JAK2 signaling Glycolysis Stromal	PI3 kinase JAK2
MYC+ MYC−	5.9% (MYC$^+$) 17.6% (MYC$^-$)	48% (MYC$^+$) 82% (MYC$^-$)	Chromatin modification Anti-apoptosis PI3 kinase signaling *S1PR2 – GNA13* inactivation Altered T$_{FH}$ interactions MYC (EZB-MYC$^+$) BCR: IgG > IgM	FL Transformed FL BL (EZB-MYC$^+$)	GC LZ (MYC$^-$) GC IZ (MYC$^+$) BCL6 (MYC$^+$) TCF3 (both) T$_{FH}$ cells (MYC$^-$) Stromal (MYC$^-$) Immune low (MYC$^+$)	PI3 kinase mTORC1 EZH2 BCL2-MCL1

Figure 13. *Fréquence des différents sous-types de LBDGC primitifs selon leurs profils mutationnels et leurs signatures transcriptomiques (36).*

Sur le plan thérapeutique, l'immunochimiothérapie R-CHOP (Rituximab, Cyclophosphamide, Adriamycine, Vincristine et Prednisone) a considérablement amélioré la survie des patients atteints de LBDGC (*37*). Malgré cette avancée majeure, une rechute de la maladie est observée chez environ 30% des patients. Dans ce contexte, l'application des

traitements de deuxième ligne est conditionnée à l'éligibilité ou non des patients à des chimiothérapies lourdes, et à la greffe de cellules souches hématopoïétiques. Seuls 10 à 15% des patients en rechute après R-CHOP sont guéris au terme du traitement de deuxième ligne (*38*). La disponibilité de nouveaux traitements ciblés et de nouvelles immunothérapies comme les lymphocytes T recombinants (CAR-T *cells*) a permis d'améliorer la survie des patients, y compris dans les situations de LBDGC réfractaires ou en rechute.

Les LBDGC secondaires

Ce sont des transformations histologiques en LBDGC se développant le plus souvent dans le contexte de lymphomes à petites cellules préexistants, et dont elles constituent l'évolution clonale agressive. Les LBDGC secondaires n'ont pas le même panorama mutationnel que les primitifs (*39*). Les plus fréquents sont les lymphomes folliculaires transformés en LBDGC, observés dans l'évolution de près de 30% des lymphomes folliculaires. La forme de LBDGC secondaire la plus grave est cependant le Syndrome de Richter, transformation très agressive se déclarant au cours de l'évolution de 5 à 10% des LLC (**Figure 14**). Toutefois, la filiation entre le clone LLC et le composant à grandes cellules B n'est démontrée que dans 80% des cas. Dans les 20% restant, le LBDGC émergeant dans un contexte de LLC constitue une maladie nouvelle et indépendante (ou *de novo*).

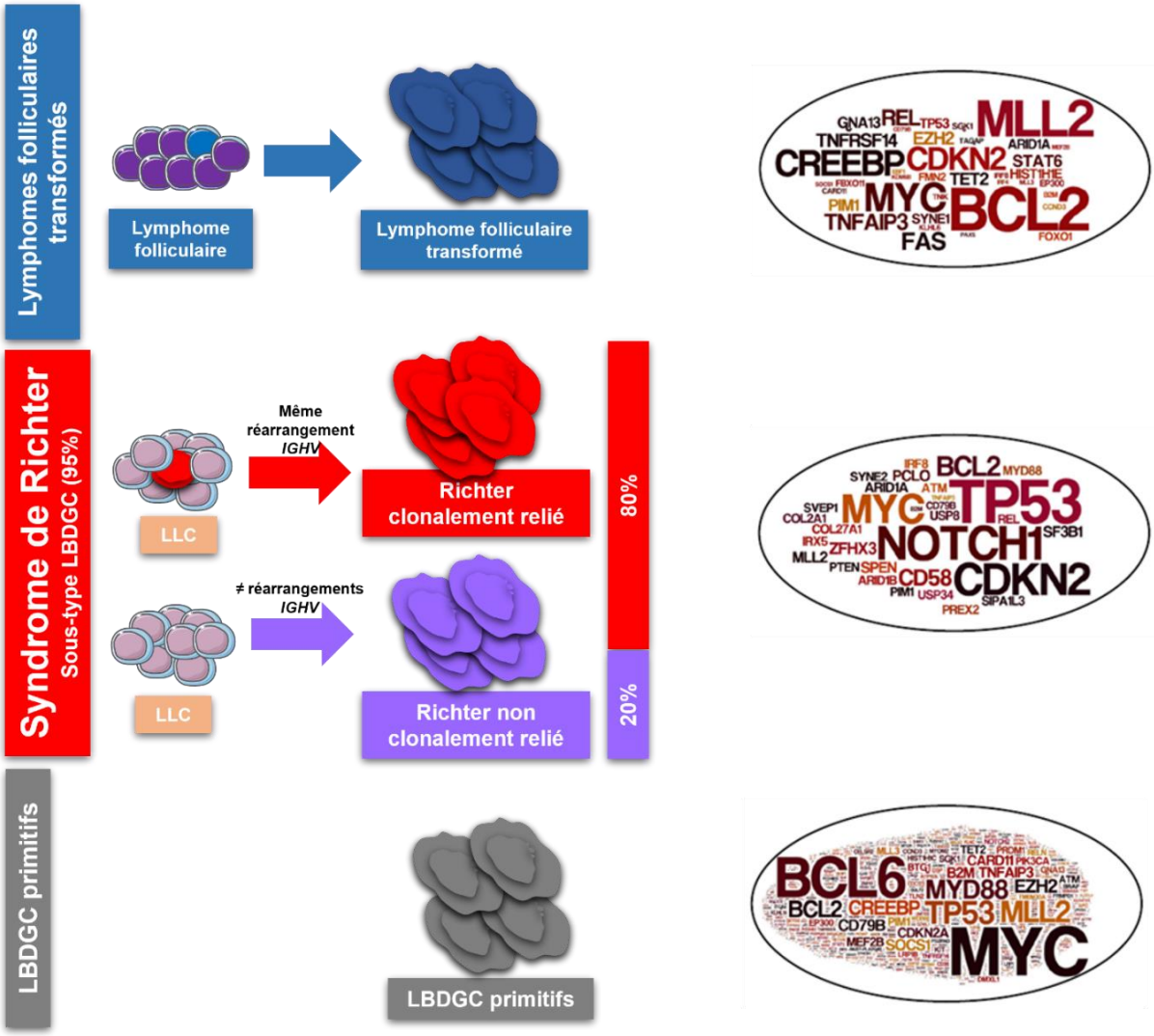


Figure 14. *Panorama mutationnel différentiel entre LBDGC primitifs et LBDGC secondaires de types lymphomes folliculaires transformés et syndrome de Richter, LBDGC secondaire à la LLC.*

Le syndrome de Richter (SR)

Le SR, ou transformation de Richter, correspond à l'émergence d'un lymphome agressif dans le contexte d'une LLC, le plus souvent un LBDGC (> 95% des cas) (*22*). Le variant type lymphome de Hodgkin (5% restant), induit le plus souvent par le virus d'Epstein-Barr, est pris en charge comme les lymphomes de Hodgkin primitifs, avec un bon pronostic. Ce chapitre se focalisera sur le variant le plus fréquent et le plus agressif : le sous-type LBDGC.

D'un point de vue morphologique, il est très difficile de faire la distinction entre un LBDGC et un SR si l'on n'a pas connaissance de la LLC antérieure. Néanmoins, d'un point de vue phénotypique, on notera que les cellules de SR sont caractérisées par l'expression des marqueurs CD20, quand seulement 15 à 70 % des cellules expriment CD23 et seulement 30 % expriment CD5. On notera également la présence de PD-1 dans 80 % des cas et de MYC dans 30 à 40 % des cas (*40*).

Transformation, chimiorésistance, relation clonale

Globalement, le SR se déclare au cours de l'évolution de 5 à 10% des LLC, mais en réalité, cette incidence est variable selon les séries, allant de 2 à 9% sur les LLC non sélectionnées, à plus de 20% dans les séries de LLC réfractaires aux traitements et/ou présentant une délétion 17p (*40*). Le SR associe une forte agressivité tumorale à un profil étendu de résistance aux chimiothérapies antimitotiques, mais aussi aux traitements ciblés efficaces dans la LLC. Son émergence constitue un tournant évolutif majeur chez les patients suivis pour une LLC, la médiane de survie se réduisant alors à 12 mois (*40, 41*). De nombreux facteurs augmentent statistiquement le risque de transformation d'une LLC en SR, mais ne sont pas utilisés en pratique courante dans le suivi des LLC (*42*).

Le facteur pronostique ayant le plus fort impact est la relation de clonalité existant entre le clone LLC et le clone Richter (*41*). La majorité des SR (80 à 90%) dérivent réellement de la LLC, ainsi que le démontre l'homologie de la séquence $IGHV_{CDR3}$ (partie hypervariable, la plus spécifique) entre le clone LLC et le clone SR. Ceux-ci sont associés à une survie de quelques mois. En revanche, les SR sans lien clonal avec la LLC (10% à 20% des cas) ont une survie médiane comparable à celle des LBDGC primitifs (62,5 mois) et sont considérés par la plupart des auteurs non pas comme de vrais LBDGC secondaires mais comme des LBDGC primitifs indépendants (*de novo*) se développant à la faveur du contexte d'immunodépression de la LLC.

Aspects moléculaires

Le panorama mutationnel du SR est très différent de celui des LBDGC *de novo*, avec une prévalence beaucoup plus faible des mutations concourant à l'activation constitutive de NF-κB et l'absence de translocations chromosomiques impliquant BCL2 ou BCL6. Le SR se caractérise également par une majorité de cas (64%) présentant un gène *IGHV* non muté et une forte prévalence de BCR stéréotypiques du groupe 8 (*41*). Dans le modèle moléculaire actuel, l'évolution du clone LLC en SR est associée à des dérégulations de la prolifération cellulaire, de l'apoptose et du cycle cellulaire, principalement en raison d'anomalies de *TP53* (60%), *NOTCH1* (30%), *MYC* (30%) et *CDKN2A/B* (30%) (**Figure 15**). Seuls 20% des SR ne présentent aucune de ces 4 anomalies (*43, 44*). Le séquençage du génome entier a permis d'étoffer le panorama mutationnel du SR, notamment par la description de mutations de *TRAF3* et *DUSP2*. Ce travail a aussi apporté des données de transcriptomique permettant d'explorer l'évolution du profil d'expression génique entre le stade LLC et le stade SR. L'étude du transcriptome était limitée à un panel de 800 gènes et a permis de confirmer le rôle central des anomalies touchant les mécanismes cellulaires de contrôle des dommages à l'ADN, et de montrer le rôle de la dérégulation de la voie des MAP kinases (*45*). Enfin, le SR possède un

profil de méthylation de l'ADN distinct des LBDGC primitifs, impliquant l'hyperméthylation de promoteurs de gènes impliqués dans la régulation du cycle cellulaire dont *TP53*, *RB1* et *CDKN2A* (*46*).

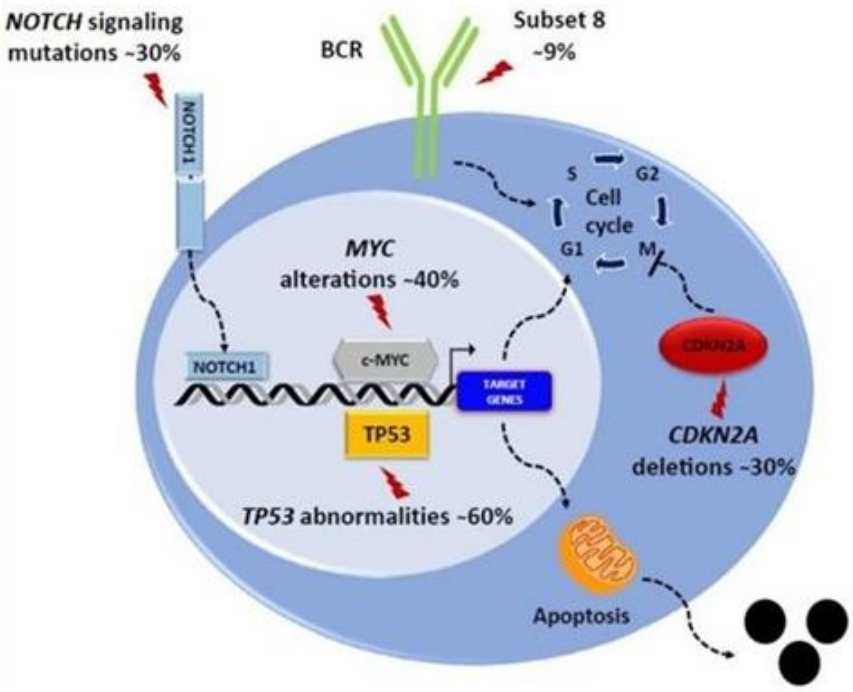


***Figure 15**. Représentation schématique des principaux acteurs moléculaires dérégulés impliqués dans le développement du SR (40).*

Options thérapeutiques

La faible médiane de survie dans le SR est due à l'inefficacité des protocoles de chimiothérapies classiques, le R-CHOP restant aujourd'hui la meilleure option et permettant un taux de réponse globale de 67%, mais seulement 7% de rémissions complètes. Des protocoles de chimiothérapie plus intensifs ont permis d'obtenir un meilleur taux de réponse complète mais sans nette amélioration de la survie globale, pour une toxicité souvent importante. L'apport des nouvelles thérapies ciblées ne semble pas comparable à ce qui a été observé pour la LLC. Quelques réponses, modestes et/ou très courtes, ont été rapportées avec les inhibiteurs de BTK, de BCL2 ou de XPO1. D'autres pistes sont actuellement à l'étude, parmi lesquelles les inhibiteurs du checkpoint immunitaire, d'autant que PD-1 est surexprimé dans le SR. Utilisés seuls ou en combinaison, ces traitements montrent des taux de réponse encourageants et prolongés (*40*, *47*). Une récente perspective thérapeutique consisterait à exploiter *in vivo* la relative domination épigénomique et transcriptomique des voies prolifératives, de la phosphorylation oxydative et du BCR observées et inhibées *in vitro* dans le SR (**Figure 16**) (*48*).

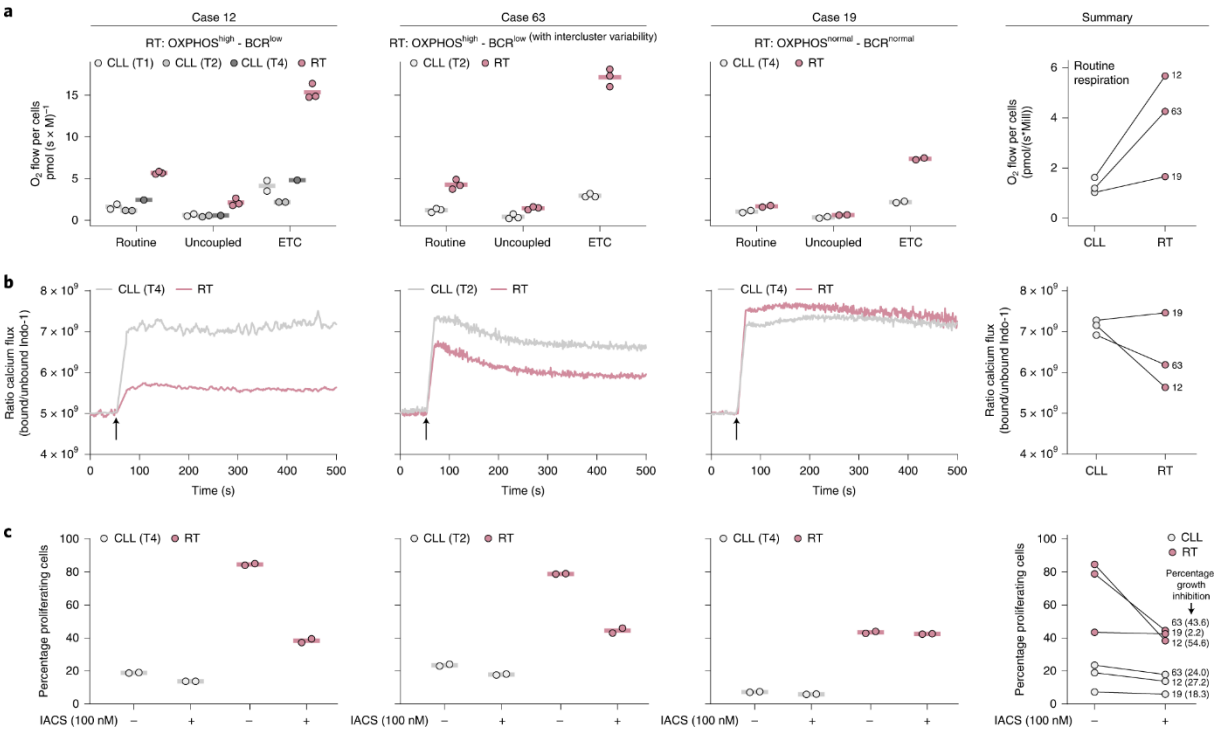


Figure 16*. Potentiel inhibiteur* in vitro *de la phosphorylation oxydative (a), de la voie du BCR (b), et de la respiration cellulaire (c) dans le SR (48).*

En conclusion, l'avènement des traitements ciblés dans la LLC ne préserve pas de la transformation en SR, qui représente 30 à 50% des cas d'échappement thérapeutique de la LLC. L'évolution de la LLC en SR constitue donc à la fois un modèle biologique de transformation agressive d'un lymphome à petites cellules en un lymphome à grandes cellules, mais aussi le principal obstacle au contrôle à long terme de la LLC.

Les cancers cérébraux

Ce chapitre traite plus généralement des tumeurs du système nerveux central, ici vulgarisé en « cancers cérébraux » car nous nous intéresserons principalement aux tumeurs intracrâniennes et leur progression vers l'agressivité et la malignité. Toutefois, qu'elles soient malignes ou non, ces tumeurs sont dangereuses car cloisonnées dans un espace restreint, provoquant des compressions pathologiques des tissus cérébraux et nécessitant souvent des interventions chirurgicales délicates.

Les origines des tumeurs du cerveau sont multiples et complexes : il en existe plusieurs centaines (**Figure 17**), classées selon leur localisation, leur taille et leur agressivité (*49*). Celles-ci sont dites bénignes lorsqu'elles sont non cancéreuses, et l'on parle dans ce cas de méningiomes ou épendymomes (tumeurs du système nerveux central qui peuvent se développer à la fois dans le cerveau et dans la moelle épinière). Lorsqu'elles sont malignes, ou cancéreuses, elles peuvent être de deux types selon leur provenance :

- Un cancer du cerveau primitif, se caractérisant par l'apparition d'une ou plusieurs tumeurs cancéreuses ou anaplasiques touchant des tissus intra-crâniaux victimes d'un dysfonctionnement du développement cellulaire. Il porte généralement le nom des cellules à partir desquelles il se développe. Il peut provenir d'une évolution dramatique de la tumeur bénigne.

- Un cancer métastatique, provenant d'une migration de cellules cancéreuses à partir d'un autre tissu tumoral distant (poumon, prostate, sein, colon), et qui ont atteint le cerveau. Il s'agit d'une tumeur cérébrale secondaire. A l'inverse, un cancer du cerveau primitif ne provoque jamais de métastases sur d'autres localisations. Les métastases cérébrales sont beaucoup plus fréquentes que les tumeurs primitives.

Parmi les cancers cérébraux primitifs, les plus courants sont les gliomes, les méningiomes, et les médulloblastomes. En France, ces tumeurs se trouvent au seizième rang des cancers en termes d'incidence. Chez l'enfant, c'est le deuxième cancer le plus fréquent après les leucémies (source : INCa).

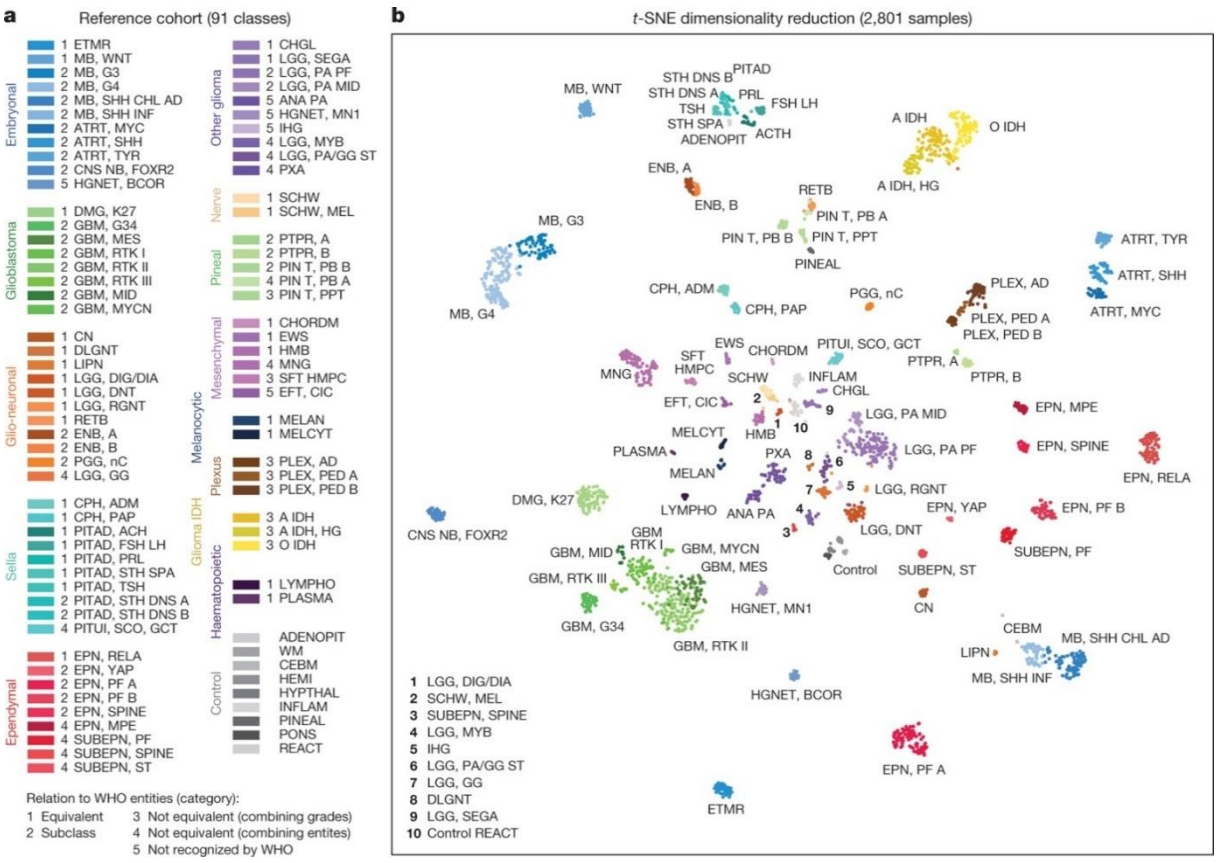


***Figure 17**. Partitionnement par réduction de dimension des tumeurs du système nerveux central à partir des données de méthylation de l'ADN (cohorte de référence de 2801 échantillons) (50).*

Ici encore, les profils épigénétiques ont tout d'abord permis une caractérisation fine de la plupart de ces cancers, en identifiant leurs cellules d'origines par rapport aux tissus cérébraux

sains (**Figure 17**). Des signatures spécifiques et réduites, extraites pour chaque type et chaque groupe de tumeur, permettent également de les reclassifier indépendamment du grade OMS, et ainsi d'améliorer drastiquement le diagnostic et le pronostic, tout en restant fidèle aux critères histologiques (**Figure 18**) (*50*). L'apport du méthylome seul permet ici de mieux refléter l'agressivité et le potentiel prolifératif et évolutif des cancers cérébraux.

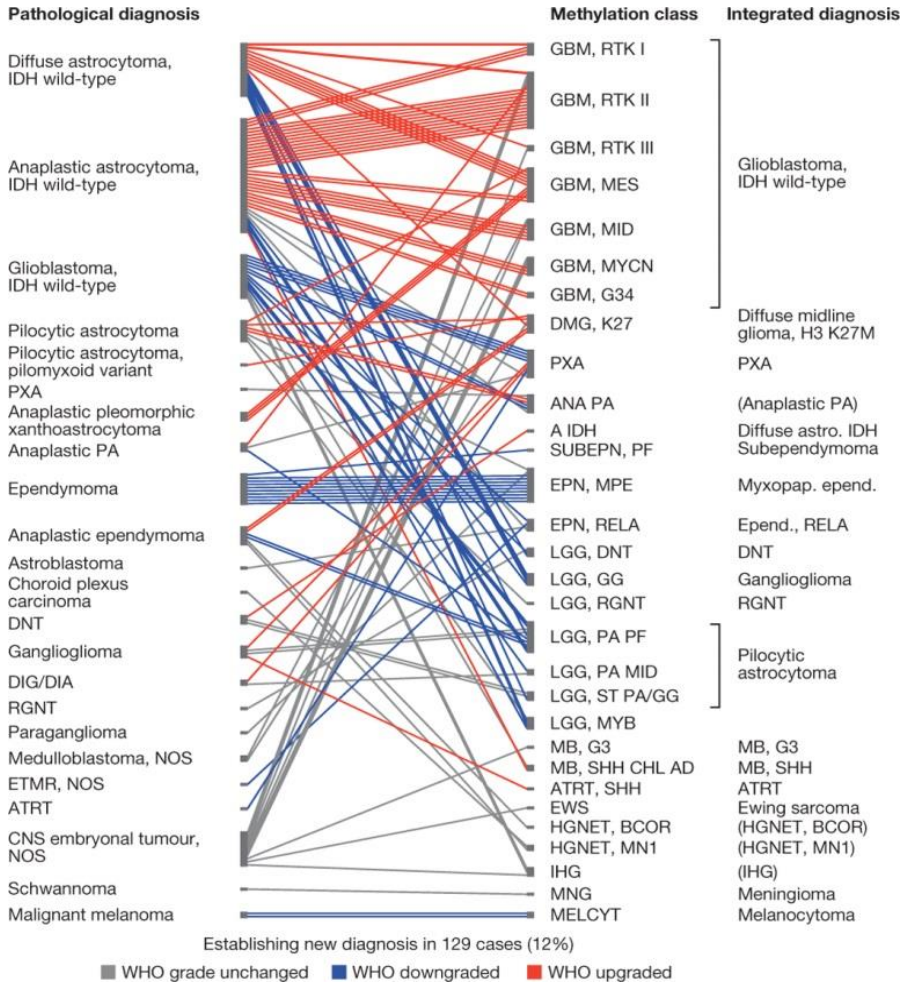


***Figure 18**. Reclassification de ces mêmes tumeurs selon les signatures moléculaires et établissement de nouveaux diagnostics /amélioration des diagnostics existants par rapport à ceux calibrés sur les critères de l'OMS (50).*

Les méningiomes

Définition, épidémiologie, étiologie

Les méningiomes sont des tumeurs développées en dehors du système nerveux, au dépend des méninges, fine membrane transparente recouvrant le cerveau, le cervelet et la moëlle épinière, vraisemblablement au niveau de la couche arachnoïde comprise entre la dure-mère et la pie-mère (*51*). Leur évolution est très variable : la plupart des méningiomes sont bénins et à croissance lente (75-80% des cas) mais certains peuvent être beaucoup plus agressifs, avec une forte tendance à récidiver et à résister aux traitements.

Il s'agit de la tumeur intracrânienne de l'adulte la plus fréquente, touchant plus couramment les femmes que les hommes (*sex-ratio* : 2/1), et qui représentait le tiers des tumeurs du système nerveux central aux États-Unis entre 2007 et 2011 (*52*). L'incidence des méningiomes augmente avec l'âge, avec un âge moyen au diagnostic autour de 65 ans.

L'influence des hormones sexuelles (œstrogènes, progestérone, anti-androgènes) et les radiations ionisantes sont des facteurs associés à l'apparition de méningiomes. L'impact des traumatismes crâniens n'est pas tout à fait établi, mais serait également en cause dans certains méningiomes agressifs (*53*). Certaines prédispositions génétiques sont également en cause, avec des syndromes de susceptibilité comme la neurofibromatose de type 2 (touchant le gène *NF2*), les schwannomatoses et des formes familiales de méningiomatoses (sources : référentiel INCa téléchargeable sur e-cancer.fr).

Les méningiomes cérébraux se manifestent principalement par des crises d'épilepsie, des signes focaux de compression cérébrale (atteintes neurologiques et cognitives, troubles sensitivo-moteurs), et des signes d'augmentation de pression intracrânienne (céphalées, nausées, troubles de la vision) (*51*). Certains méningiomes sont asymptomatiques et décelés fortuitement au cours d'un examen d'imagerie cérébrale.

Dans la majorité des cas, la chirurgie est le principal traitement des méningiomes. Le degré de résection chirurgicale de la tumeur est même un facteur indépendant de récurrence prématurée, la chirurgie agressive étant à bannir, sans impact sur la morbidité (*54*).

Classification OMS, récurrence et survie

En pratique courante, c'est la classification OMS en grades de gravité et d'agressivité croissantes qui est utilisée. Celle-ci repose sur des observations essentiellement histologiques, et distingue 15 sous-types de méningiomes distribués sur 3 grades (**Figure 19**) :

- Les méningiomes de grade 1, dits bénins, d'évolution lente. Ce sont les plus fréquents, avec 70% des cas. Il en existe 9 sous-types histologiques (méningothélial, transitionnel, fibroblastique, angiomateux, psammomateux, microkystique, sécrétoire, métaplasique, riche en lymphocytes et plasmocytes) et même un type mixte. Le taux de récurrence varie de 7 à 25% selon les cohortes. La survie à 5 ans est de 92%.

- Les méningiomes de grade 2, avec trois sous-types histologiques (atypique, à cellules claires, chordoïde), et qui se caractérisent entre autres par une prolifération accrue, avec un indice mitotique ≥ 4 et < 20 mitoses pour 1,6 mm², une infiltration inflammatoire chronique, un envahissement du parenchyme cérébral, une hypercellularité et/ou une

nécrose tumorale. Ils représentent 20 à 30% des cas, avec un taux de récurrence de 29 à 59%.

- Les méningiomes de grade 3 sont de 3 sous-types (rhabdoïde, papillaire et anaplasique (malin)), et prolifèrent agressivement, avec un indice mitotique ≥ 20 mitoses pour 1,6 mm². Ils peuvent porter une mutation du promoteur de *TERT* et/ou une délétion homozygote de *CDKN2A/B*. Ces critères moléculaires sont les seuls pris en compte dans la classification OMS en date (2021). Ces méningiomes représentent 1 à 3% des cas, avec un taux de récurrence de 50 à 94%.

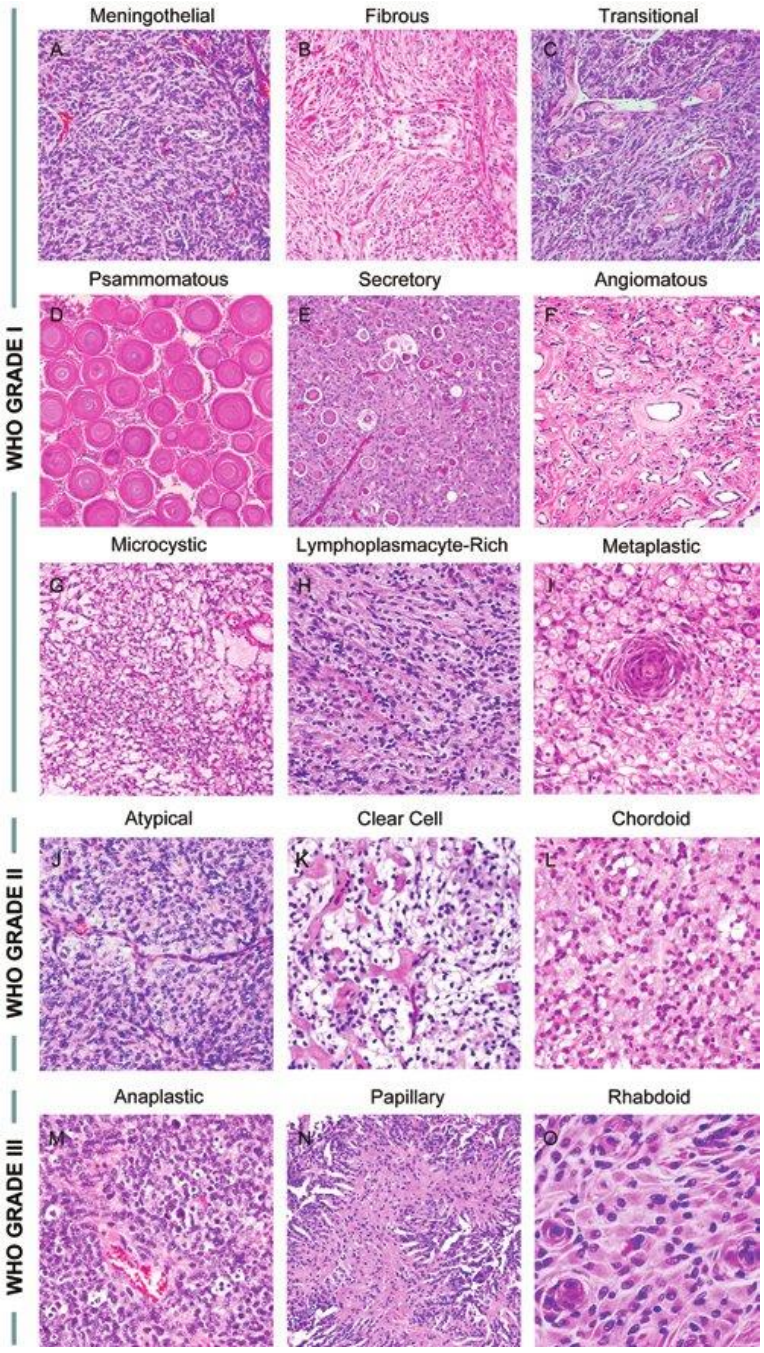


***Figure 19**. Photomicrographies montrant les sous-types histologiques de méningiomes, selon leur grade OMS (55).*

Le taux de récurrence augmente drastiquement avec le grade (*56*), accompagné par une survie globale de plus en plus faible (*57*, *58*). Cependant, cette classification est par nature subjective car dépendante des observations anatomopathologiques (**Figure 20**), et demeure imparfaite, avec par exemple des méningiomes de grade 1 qui pourtant sont d'évolution défavorable, et inversement des tumeurs de grade 2 et même parfois de grade 3 qui montrent une évolution favorable.

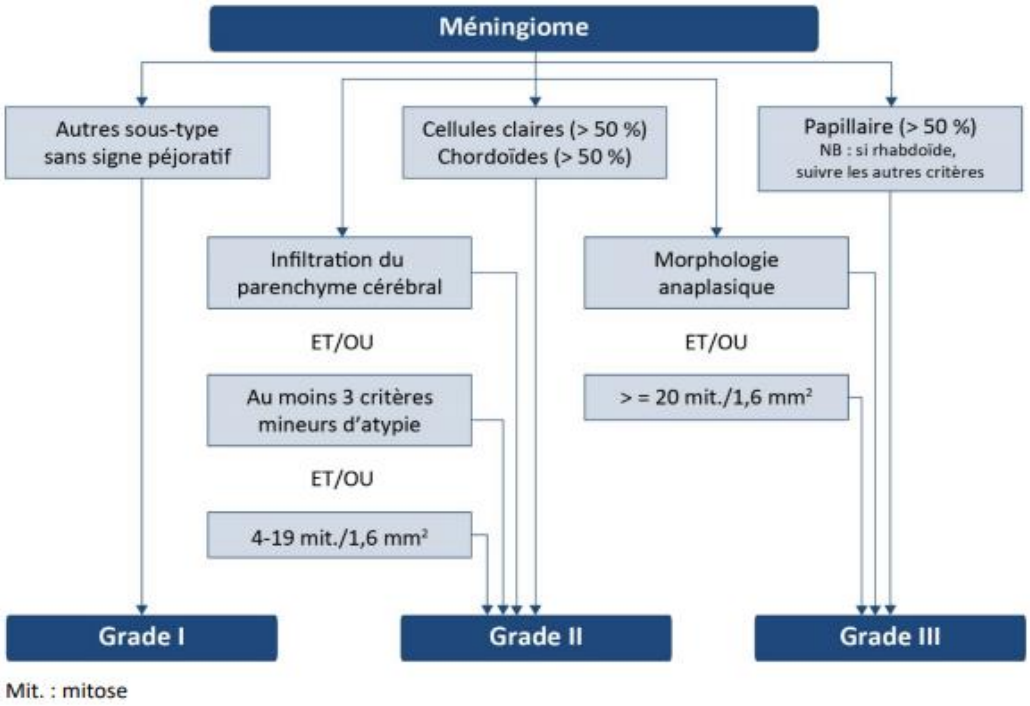


Figure 20. *Arbre décisionnel anatomopathologique, d'après le référentiel INCa 2020 basé sur les critères OMS 2016 (59).*

De nouvelles classifications à valeur pronostique supérieure à celle des grades OMS ont récemment été proposées à partir des profils moléculaires dans diverses cohortes de méningiomes.

Aspects moléculaires

La première altération génétique décrite dans les méningiomes est la délétion du chromosome 22q. Elle implique l'inactivation du gène suppresseur de tumeur *NF2* codant pour la merline (**Figure 21**) (*60*). Elle est parfois associée à des mutations de *SMARCB1* et/ou *SMARCE1*, qui interviennent dans le remodelage de la chromatine. L'inactivation bi-allélique de *NF2* concerne 80% des méningiomes de grade 2 et 3, et 43% des tumeurs de grade 1. Les méningiomes sans inactivation de *NF2* comportent principalement des altérations situées sur 4 voies de signalisation quasi mutuellement exclusives : i) la voie Hedgehog, ii) les mutations de *TRAF7, iii)* les mutations de *POLR2A* dans les méningiomes de grade 1, plus souvent de type

méningothéliaux, et iv) les mutations plus rares et de faibles fréquences (*BRCA1*, *BAP1*) (*61-63*).

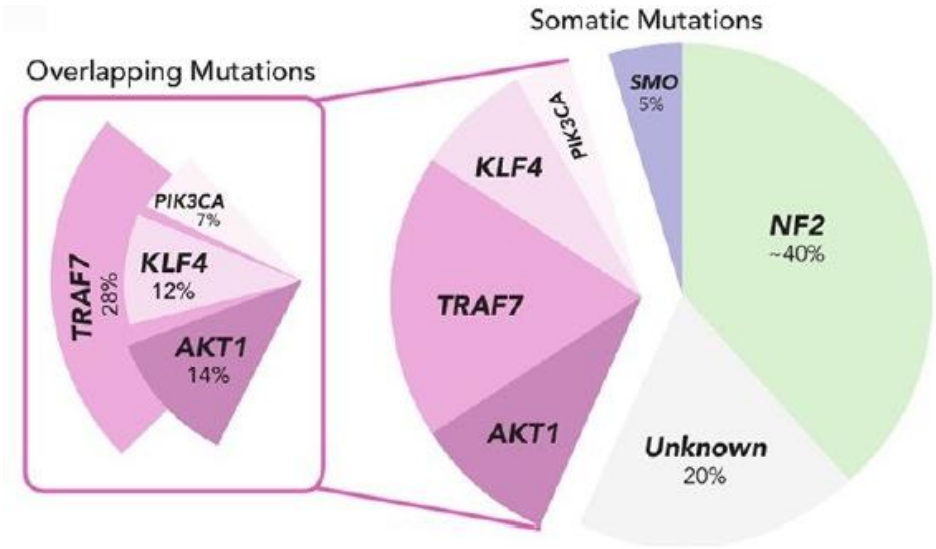


Figure 21. *Fréquence des mutations somatiques dans les méningiomes (64).*

Les méningiomes de grade 2 et 3 présentent une plus grande instabilité génomique, une charge mutationnelle plus élevée, et plus d'évènements altérant les CNVs (C*opy Number Variations*, ou variations du nombre de copie des gènes) (*62*). Les pertes chromosomiques sont plus fréquentes que les gains dans les 3 grades de méningiomes, avec une monosomie 22 en tête des anomalies.

Des mutations touchant le promoteur de *TERT* sont présentes dans 6,5 à 11% des méningiomes. *TERT* étant normalement réprimé dans les cellules somatiques, ces altérations dans sa région promotrice provoquent *in fine* l'expression de TERT et une stabilisation des télomères, et par conséquent promeuvent l'immortalité et la prolifération cellulaire. En effet, les tumeurs portant ces mutations sont plus agressives et présentent un taux de récidive plus élevé (*65*).

Plus récemment, Sahm *et al*. ont défini des signatures moléculaires permettant de classifier les méningiomes selon leur profil de méthylation de l'ADN (*66*). Au total, ce sont 497 échantillons répartis dans les 3 grades OMS qui ont été reclassifiés en 2 grands groupes épigénétiques puis en 6 classes distinctes (**Figure 22**), chacune démontrant une spécificité clinique et biologique : survie, localisation, histologie prédominante, expression des gènes, aberrations chromosomiques et mutations ponctuelles (**Figure 23**). Surtout, ces classes de méthylation ont un potentiel prédictif et une meilleure valeur pronostique que les grades OMS actuellement utilisés (*66*).

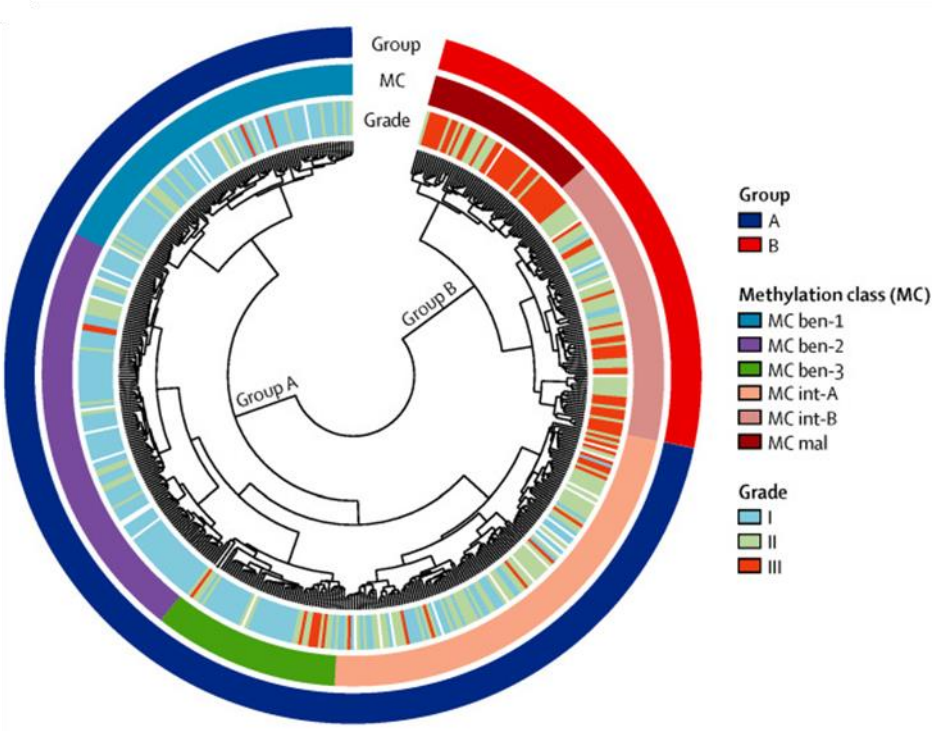


***Figure 22**. Classification des méningiomes selon leur paysage épigénétique (profils de méthylation de l'ADN) illustrant les manquements et la subjectivité du grade histologique OMS (66).*

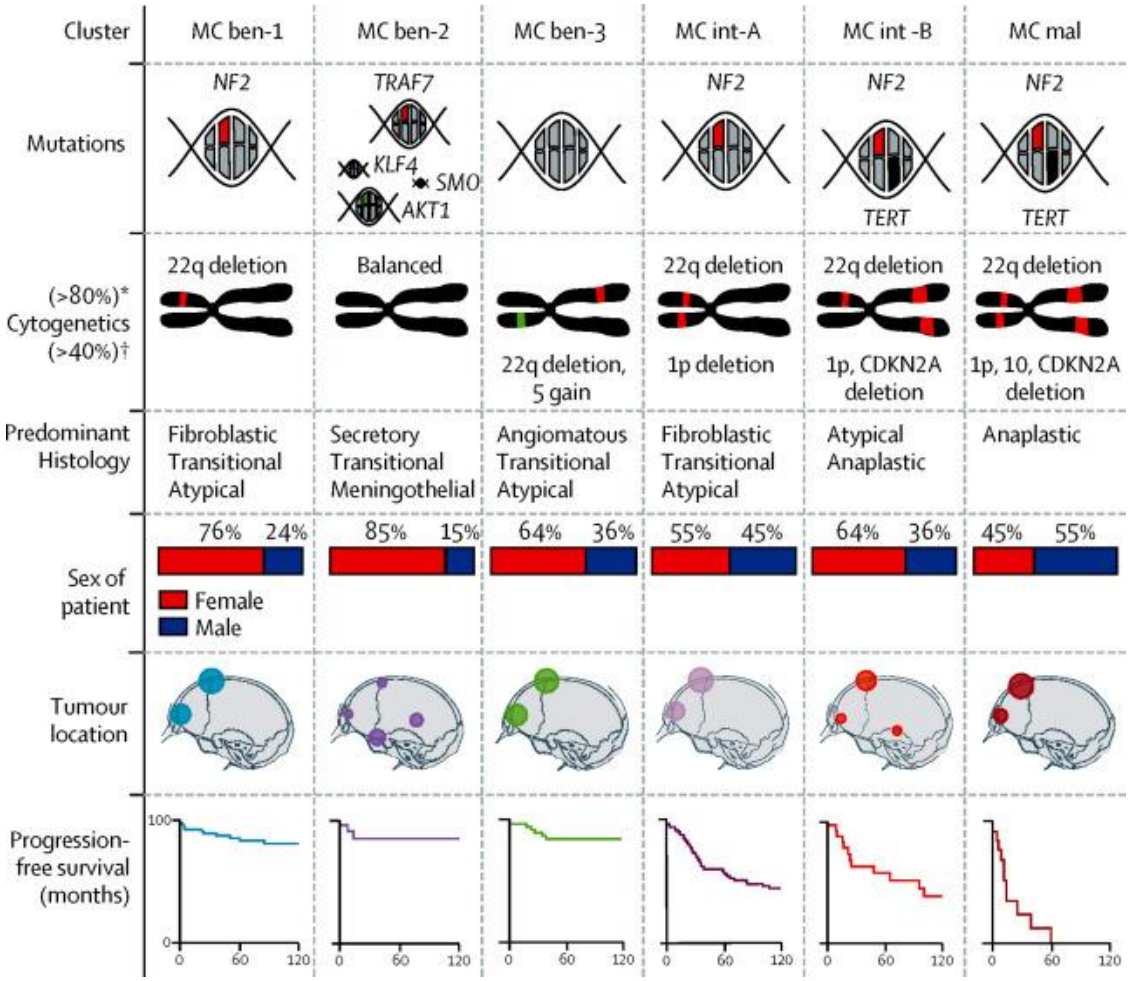


***Figure 23**. Vue schématique des 6 classes de méthylation (MC) des méningiomes et de leurs caractéristiques moléculaires et cliniques (66).*

Une autre étude majeure sur 121 méningiomes caractérisés en multi-omique avec intégration à grande échelle des exomes, génomes, épigénomes, transcriptomes (en *bulk* et en *single-cell*) et protéomes a permis de distinguer 4 catégories de tumeurs : le groupe « immunogénique », le groupe « *NF2* non muté », le groupe « hypermétabolique » et le groupe « prolifératif » (*67*). Les noms des 4 groupes sont bien sûr dérivés des voies moléculaires caractéristiques et sur-représentées dans chacun d'entre eux. L'apport de cette classification, en plus de sa valeur prédictive sur la survie sans récurrence, informe sur les orientations thérapeutiques les plus adaptées. De même que ci-dessus, chaque groupe présente des spécificités biologiques au-delà des considérations cliniques (**Figure 24)**. A titre d'exemple, le niveau de corrélation entre expressions transcriptomiques, protéogénomiques et immunohistochimiques est sans appel : ce sont bien les fonctions biologiques et les voies moléculaires qui décrivent le mieux la pathologie, avec une précision millimétrée et en définitive, transposable à l'individu.

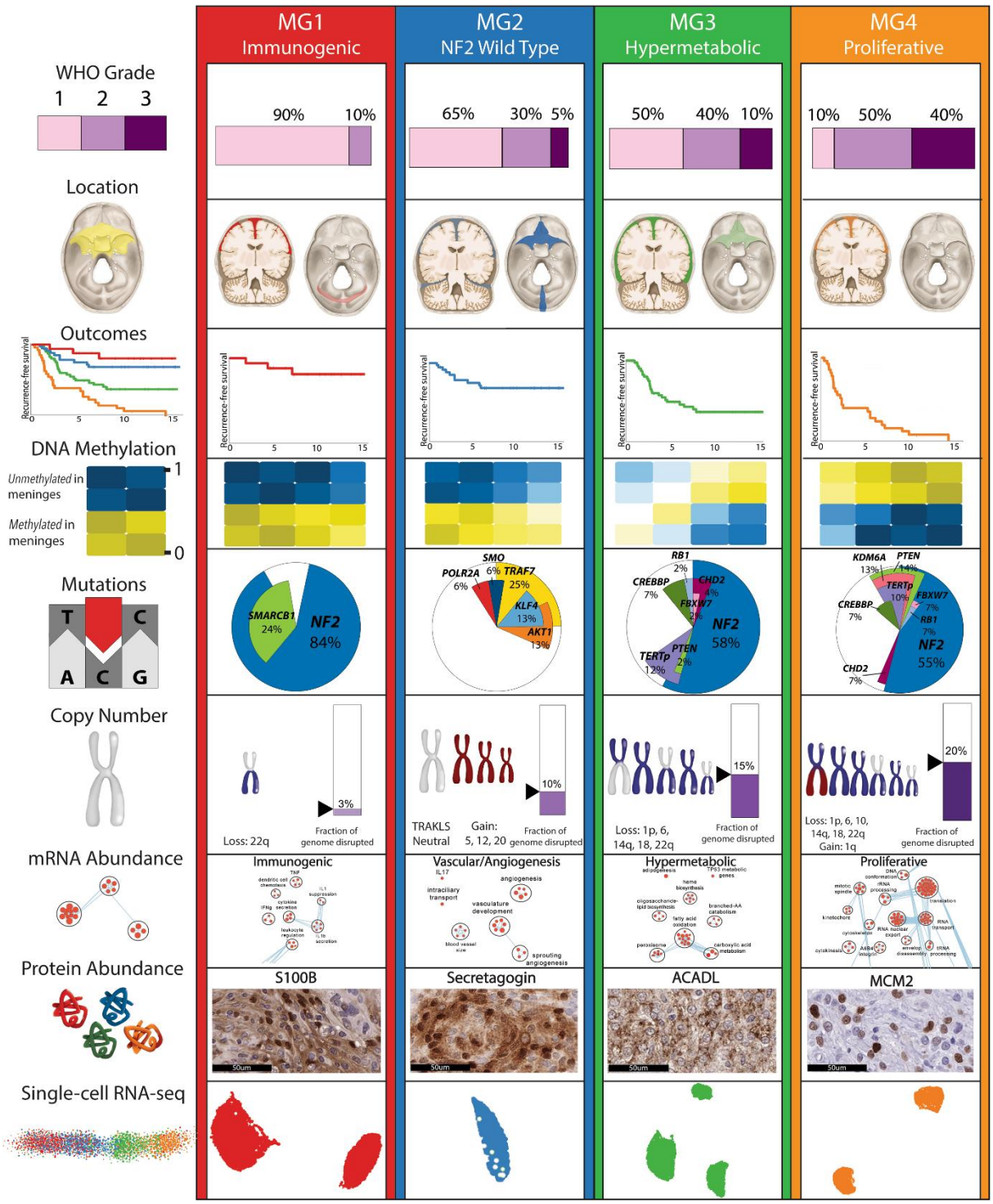


***Figure 24**. Vue schématique des 4 groupes de méningiomes (MG1, MG2, MG3 et MG4) selon leurs profils moléculaires, avec leurs caractéristiques biologiques prototypiques (67).*

Les oligodendrogliomes

Définition et généralités

Les oligodendrogliomes sont des tumeurs cérébrales dont on pense qu'elles se développent à partir des cellules souches précurseurs des oligodendrocytes, des cellules de la névroglie interstitielle dont la principale fonction est la formation de la gaine de myéline entourant et protégeant les axones du système nerveux central. Leur origine ne fait cependant

pas consensus et reste sujet à débat : ils pourraient provenir de cellules souches cancéreuses (*68*).

Ces tumeurs rares représentent environ 2 à 5% des tumeurs cérébrales primitives et sont le troisième type le plus commun des tumeurs gliales de l'adulte, avec des pics diagnostics vers 40 et 50 ans, selon leur agressivité, bien qu'ils puissent toucher les enfants et les adolescents (*69, 70*). L'étiologie de ces cancers n'est pas avérée, la grande majorité intervenant de manière sporadique, sans cause héréditaire directe connue. Les oligodendrogliomes se révèlent le plus souvent par une longue histoire clinique typiquement marquée par des crises d'épilepsies, des symptômes de pressions intracrâniennes comme des céphalées et des vomissements, des troubles visuels ou de la personnalité. Ces troubles dépendent de la localisation cérébrale. Dans les formes malignes, le tableau clinique est marqué par une progression rapide des signes neurologiques.

Bien que moins agressifs que les autres tumeurs du système nerveux central, les oligodendrogliomes restent incurables, malgré un traitement par radiothérapie et chimiothérapie retardant les rechutes (*68*).

Classification OMS

Les oligodendrogliomes sont des tumeurs cérébrales individualisées parmi les autres tumeurs gliales en raison d'anomalies génétiques et de profils moléculaires particuliers (**Figure 25**), ainsi qu'une plus grande sensibilité à la chimiothérapie. On les distingue selon leur classification OMS, qui depuis 2016 les différencie définitivement des astrocytomes sur des critères mutationnels, bannissant les entités mixtes ou intermédiaires (*59, 71*).

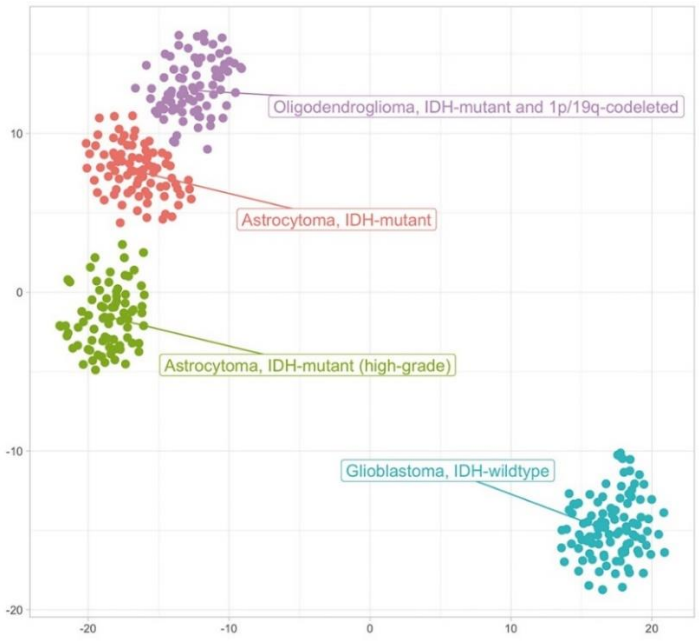


***Figure 25**. Profils de méthylation de l'ADN de gliomes diffus de l'adulte par réduction de dimension t-SNE (t-distributed Stochastic Neighbor Embedding) (50).*

Les oligodendrogliomes sont dits bénins ou de bas grade (OMS grade 2), ou malins ou de haut grade, ou encore anaplasiques (OMS grade 3), ces derniers se caractérisant histologiquement par une cellularité et une activité mitotique accrues, une prolifération microvasculaire, un polymorphisme nucléaire et/ou une nécrose (**Figure 26**). Les formes malignes peuvent provenir de la transformation d'une tumeur bénigne plusieurs années après le diagnostic initial. Au cours de la maladie, la tumeur peut éventuellement dégénérer en oligodendrogliome hautement malin, de grade 4.

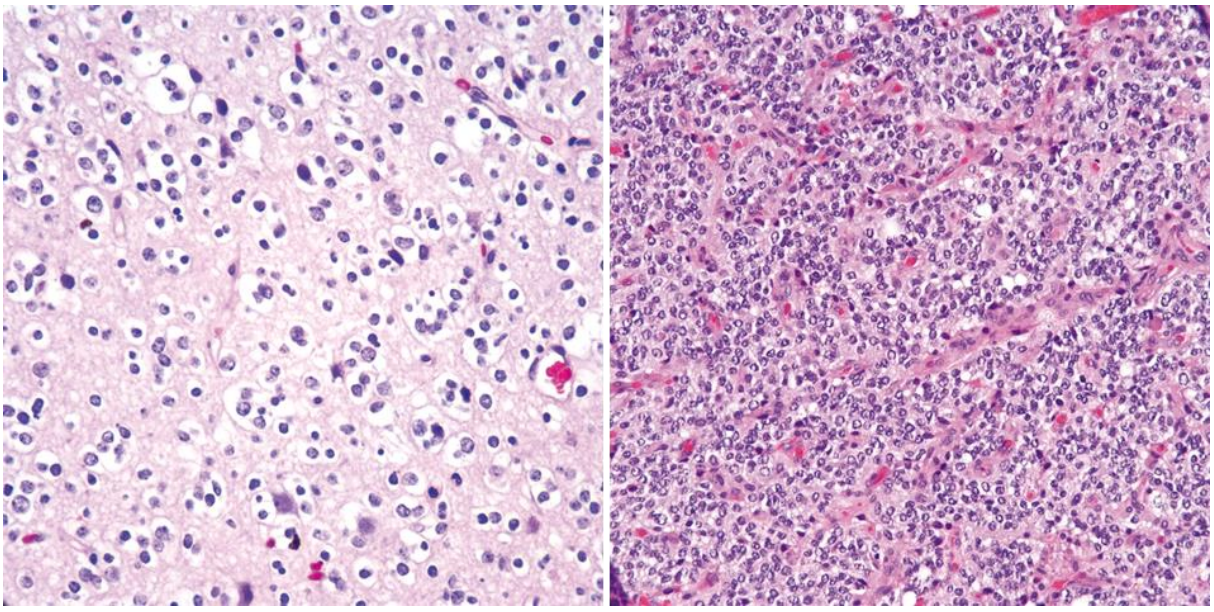


***Figure 26**. Coupes histologiques d'oligodendrogliomes de grade 2 montrant une infiltration corticale en regroupement autour des corps neuronaux (à gauche), et de grade 3 illustrant hypercellularité et hypervascularité dans ces tumeurs (à droite) (source : WHO classification of tumors, online, Daniel J. Brat).*

Les oligodendrogliomes anaplasiques (OA) sont des gliomes de haut grade, correspondant histologiquement aux tumeurs cérébrales primitives de grade 3 (*59*). Par définition, et selon la classification OMS 2016 qui intègre désormais des critères génétiques et phénotypiques, les OA portent à la fois des mutations *IDH1/2* et une codélétion 1p/19q (IDHmt/1p19qcodel) (*72, 73*). Ils sont rares, représentant 0,5 % de toutes les tumeurs cérébrales primitives et un tiers des tumeurs oligodendrogliales (*74*).

Aspects moléculaires et groupes pronostics

Des altérations génétiques caractéristiques telles que des mutations dans un des gènes de l'isocitrate déshydrogénase (mutations faux-sens : pour *IDH1*, p.R132H dans 90% des cas ; pour *IDH2*, au niveau du codon 172) et des délétions du bras p du chromosome 1 et du bras q du chromosome 19 sont de bons prédicteurs de la survie dans les oligodendrogliomes. La survie sans récidive est, par rapport à d'autres tumeurs gliales, relativement bonne, même dans les formes anaplasiques, grâce au traitement combinant la chirurgie, la chimiothérapie et la

radiothérapie (*75*). Des délétions concomitantes 1p et 19q sont observées dans 50 à 70% des tumeurs bénignes et des OA (*76*), et proviennent de translocations déséquilibrées entre les chromosomes 1 et 19 suivies de pertes alléliques des bras complets. Cette codélétion peut être considérée comme favorable pour le patient et rend plus probable une réponse à la radiothérapie ou à la chimiothérapie (**Figure 27**). La grande majorité des oligodendrogliomes portent des mutations dans la région promotrice de *TERT*, mais jamais dans les cas touchant les adolescents (*77*)**.** Ces mutations provoquent une stabilisation des télomères, et par conséquent promeuvent l'immortalité et la prolifération cellulaire.

Comme pour d'autres cancers cérébraux, le pronostic et la survie restent mieux associés au profil moléculaire qu'au diagnostic morphologique / histologique (*78*). C'est particulièrement vrai dans le cas des gliomes mutants pour IDH, dont les oligodendrogliomes, pour lesquels le grade OMS reste discutable (*73*).

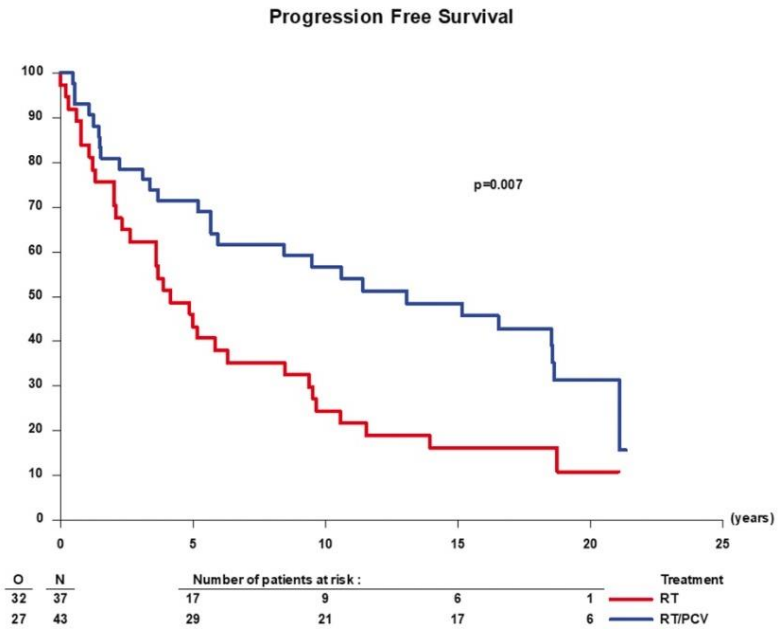


Figure 27*. Survie sans progression de patients atteints d'oligodendrogliomes et traités par radiothérapie seule (en bleu) ou radiothérapie + chimiothérapie (en rouge). Etude prospective Européenne sur 80 patients (79).*

Des caractéristiques histologiques telles que la prolifération microvasculaire, l'indice mitotique et la nécrose ont été associées à trois sous-groupes pronostiques d'OA IDH-mt/1p19qcodel présentant des altérations génomiques distinctes (*80*). Celles-ci incluent des mutations de CIC et TCF12, toutes deux liées à des OA plus agressifs (*81-83*), et une perte allélique de 9p21.3 associée à une survie plus courte dans les OA IDHmt/1p19qcodel (*84*).

Une approche multi-omique utilisant le transcriptome, le génome et le méthylome a également permis l'identification de facteurs pronostiques potentiels, et la classification des

tumeurs oligodendrogliales en trois sous-groupes moléculaires ayant des comportements cliniques distincts (*85*). Une technique d'apprentissage automatique permettrait de plus une stratification pronostique en identifiant les CNV somatiques dans les OA (*86*).

Toutes ces méthodes nécessitent des analyses moléculaires approfondies qui ne peuvent pas être généralisées facilement en clinique. En revanche, et en ce qui nous concerne, les marqueurs de prolifération tels que Ki-67, dont l'indice immunohistochimique est fortement corrélé au pronostic de nombreux types de tumeurs, constituent des moyens simples, efficaces et économiques pour évaluer l'agressivité tumorale. Ki-67, et MCM6 (un autre marqueur de progression du cycle cellulaire, au stade réplicatif), constituent effectivement de bons prédicteurs de la survie dans les OA (*87*).

L'obésité

L'OMS définit l'obésité comme une accumulation anormale et/ou excessive de graisse qui présente un risque pour la santé (http://www.who.int). En augmentation continuelle dans les pays en développement et développés, cette maladie chronique est un problème de santé mondial désormais vu comme « épidémique » (*88*). En 2015, la prévalence chez l'adulte était de 12% dans le monde et de 17,2% en France. L'obésité est un important facteur de risque de multiples comorbidités (*89*), qui entraine une stigmatisation sociale et une faible estime de soi, et qui a été associée à une moins bonne qualité de vie ainsi qu'à une réduction de l'espérance de vie pouvant aller jusqu'à 20 ans dans ses formes les plus graves (*90, 91*). Les complications que l'obésité entraine (diabète de type 2, maladies cardiovasculaires, maladies inflammatoires, cancers) concernent au moins 2,8 millions de personnes chaque année dans le monde, toujours selon l'OMS. Il existe différentes approches de traitement : modification du style de vie, modification du comportement, pharmacothérapie et chirurgie bariatrique. Cependant, les programmes actuels ont en grande partie échoué à faire ralentir la progression de la maladie, malgré les énormes investissements financiers représentant jusqu'à 10% des dépenses nationales de santé (*92*). Dans ce contexte, approfondir les recherches sur les causes de l'obésité pourrait améliorer la prévision, la prévention et le soin sur le long terme (*93*).

L'indice de masse corporelle (IMC)

L'IMC est le paramètre le plus fréquemment utilisé pour diagnostiquer l'obésité en population. D'après l'OMS, les sujets adultes sont classés selon différentes catégories :

insuffisance pondérale (IMC<18.5 kg/m^2), corpulence normale (18.5≤IMC<25 kg/m^2), surpoids (25≤IMC<30 kg/m^2), obésité modérée (35≤IMC<40 kg/m^2), sévère (35≤IMC<40 kg/m^2) ou morbide (IMC≥40 kg/m^2) (**Figure 28**).

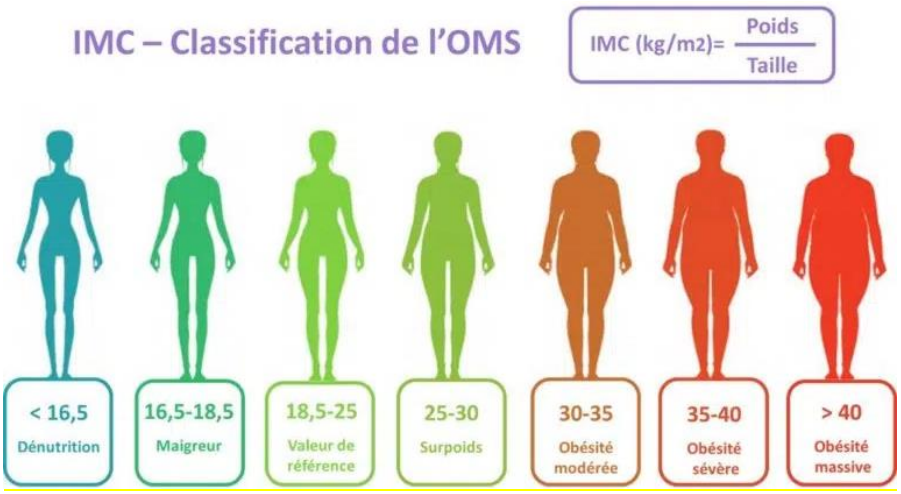


***Figure 28**. Catégories OMS des individus adultes selon leur Indice de Masse Corporelle (source : Département Prévention Cancer Environnement, Centre Léon Bérard, 2020, https://www.cancer-environnement.fr).*

Analysé comme un trait continu ou binaire / catégoriel, l'IMC est aussi le phénotype le plus couramment utilisé dans la recherche sur l'obésité. Bien que l'IMC soit un trait hautement héritable, facile et peu coûteux à mesurer, sa pertinence pour mesurer le degré d'adiposité a tout de même été remise en question car l'IMC ne fait aucune distinction entre la masse maigre et la masse grasse (*94*). Bien que fortement corrélé à la masse grasse chez les personnes obèses, la corrélation avec l'IMC reste toutefois plus faible pour les personnes de poids normal et en insuffisance pondérale (*95*, *96*). Une variation substantielle de la masse grasse a été décrite dans la population générale à un IMC donné (*95*). Bien que l'étude du pourcentage de masse grasse soit l'approche de référence, accumuler ces données dans de larges populations peut être difficile en pratique (*94*). Sur la base de ces considérations, l'IMC est une mesure pertinente dans la recherche sur l'obésité si elle est utilisée avec circonspection (*94*).

Facteurs environnementaux et biologiques, génétique

Alors que la proportion d'adultes obèses a plus que doublé en France entre 1980 et 2015, les origines de cette épidémie sont multifactorielles. Le bouleversement récent des habitudes alimentaires contribue de façon très significative à ce problème. Au-delà des aspects nutritionnels et génétiques, des facteurs environnementaux ont été impliqués dans le développement et l'installation de l'obésité, en interaction avec des facteurs biologiques **(Figure 29)** (*93*). Il existe des facteurs de risques environnementaux bien établis, tels qu'une mauvaise alimentation, l'inactivité physique, le stress psychosocial et le manque de sommeil.

Parmi les individus présentant un mode de vie obésogène, tous ne deviendront pas obèses, ce qui suggère l'implication de nombreux contributeurs biologiques dans la survenue de la pathologie. Parmi ces facteurs, on retrouve essentiellement l'âge, le sexe, l'origine ethnique, les facteurs *in utero*, les antécédents médicaux, le microbiome intestinal, l'épigénétique et la génétique (*93, 97*).

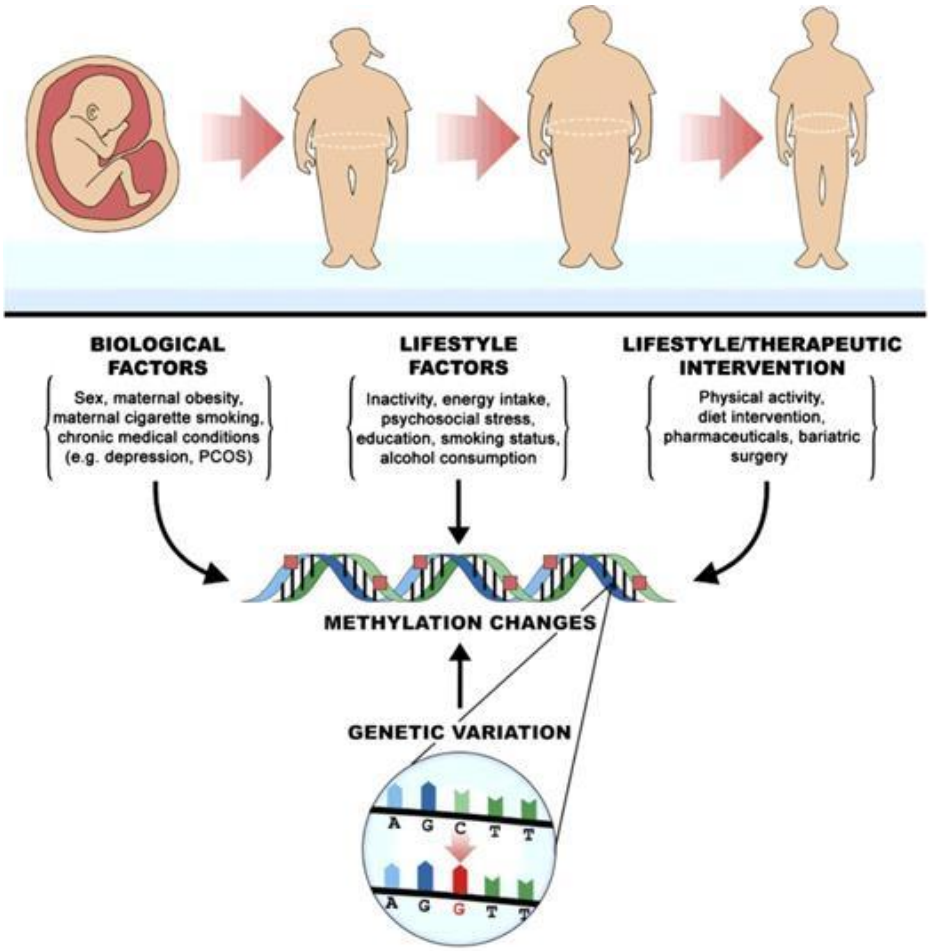


Figure 29. *Interactions gènes/environnement dans l'obésité (93).*

Les antécédents familiaux d'obésité sont également un facteur connu de risque d'obésité infantile, pour lesquels une multiplication du risque par 2 a été observée (*98*). Des études de jumeaux, de familles et dans la population générale ont estimé une héritabilité de 40 à 75% pour l'IMC et le risque d'obésité (*99, 100*). De nombreuses anomalies génétiques responsables d'obésités monogéniques syndromiques, monogéniques non syndromiques, et polygéniques ont été identifiées au cours de ces 25 dernières années (*93, 101*).

Les approches pangénomiques GWAS, permettant d'analyser des millions de variants génétiques (SNP pour *Single Nucleotide Polymorphism*) dans de larges cohortes, ont identifié près de 1000 locus associés à l'IMC. Toutefois ces SNPs n'ont collectivement qu'un effet modeste sur le risque d'obésité et contribuent pour seulement 6% de la variance de l'IMC (*102*). On citera néanmoins la découverte en 2007 d'un cluster de variants fréquents dans le premier

intron du locus *FTO*, associé de façon indubitable avec l'IMC (*103*, *104*). Un de ces variants est retrouvé dans 16 % de la population sous forme homozygote et est corrélé avec un risque majoré d'obésité commune (augmenté d'environ 40%), débutant dès l'enfance (*103*). Si leur contribution à la physiopathologie de la maladie reste un mystère, ces gènes situés à proximité de ce millier de SNPs sont presque tous exprimés dans le système nerveux central et particulièrement dans les régions cérébrales intervenant dans les mécanismes de récompense et d'addiction (*105*). Les voies moléculaires impliquées touchent à la fonction synaptique, à la signalisation par le glutamate, à la sécrétion et l'action de l'insuline, au métabolisme énergétique, à la biologie des lipides et à l'adipogénèse (*106*).

Il existe également des formes plus rares, dites monogéniques ou monofactorielles, au-delà de ces formes communes d'obésité, et qui touchent 5-10% des sujets obèses dans les pays occidentaux (hors populations consanguines). Un seul évènement génétique rare, une mutation ponctuelle ou une anomalie chromosomique, est suffisant pour provoquer la maladie. Celle-ci intervient souvent à un âge précoce, est sévère, et est généralement associée à une hyperphagie (déséquilibre de la prise alimentaire par perte de la satiété, faim incontrôlée). Certaines formes syndromiques exceptionnellement rares sont parfois associées à une déficience intellectuelle et une dysmorphie en même temps que l'obésité (par exemple les syndromes de Prader-Willi et de Bardet-Biedl). Dans les formes monogéniques non syndromiques, une quinzaine de gènes ont pour le moment été identifiés ou sont fortement suspectés d'être responsables des obésités monogéniques : *LEP* (codant pour la leptine), *LEPR* (récepteur de la leptine), *POMC* (proopiomélanocortine), *PCSK1* (proprotéine convertase 1), *MC4R* (récepteur de type 4 des mélanocortines), *MC3R* (récepteur de type 3 des mélanocortines), *MRAP2* (protéine accessoire 2 du récepteur de la mélanocortine 2), *SIM1* (s*ingle-minded homolog* 1), *BDNF* (*brain-derived neurotrophic factor*), *NTRK2* (récepteur à tyrosine kinase de type 2 neurotrophique), *SH2B1* (*SH2B adaptor protein* 1), *ADCY3* (adénylate cyclase 3), *AGRP* (*agouti-related protein*), *PHIP* (*pleckstrin homology domain interacting protein*) et *KSR2* (*kinase suppressor of Ras2*) (*107*). Tous ces gènes codent des protéines impliquées dans la voie leptine / mélanocortines, régulant l'appétit, et principalement active au niveau de l'hypothalamus (**Figure 30**).

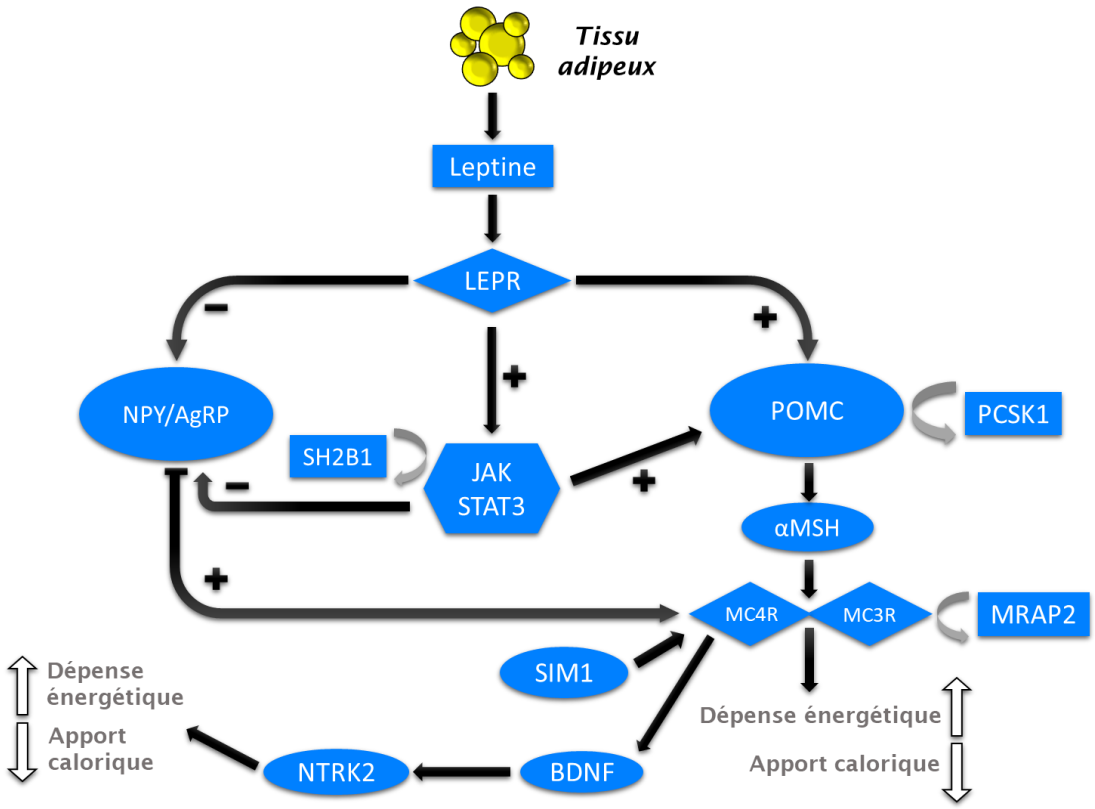


Figure 30. *Produits des gènes de la voie leptine / mélanocortines impliqués dans les formes monogéniques de l'obésité via la régulation de l'apport alimentaire et de la dépense énergétique. Une insuffisance dans n'importe quelle protéine peut déclencher une obésité sévère.*

Les adipokines

Les adipokines sont des cytokines produites principalement par le tissu adipeux. Elles jouent un rôle régulateur important dans le métabolisme des glucides et des lipides, et sont associées, entre autres, à divers traits phénotypiques en lien avec l'obésité. Par exemple, la leptine et l'adiponectine améliorent la sensibilité à l'insuline, alors que la visfatine, la fétuine-A, la résistine et l'inhibiteur de l'activateur du plasminogène-1 (PAI-1) contribuent au développement de l'intolérance au glucose. Leptine et adiponectine augmentent également l'oxydation des acides gras, préviennent la formation de cellules spumeuses (cellules gonflées de gouttelettes de lipoprotéines de basse densité (LDL pour *low-density lipoprotein*) promotrices du processus d'athérosclérose) et améliorent le métabolisme des lipides, tandis que la visfatine, la fétuine-A, la PAI-1 et la résistine ont des propriétés pro-athérogènes (**Figure 31**) (*108*).

Nous nous intéresserons ici à deux hormones majeures de l'homéostasie énergétique : la leptine et l'adiponectine.

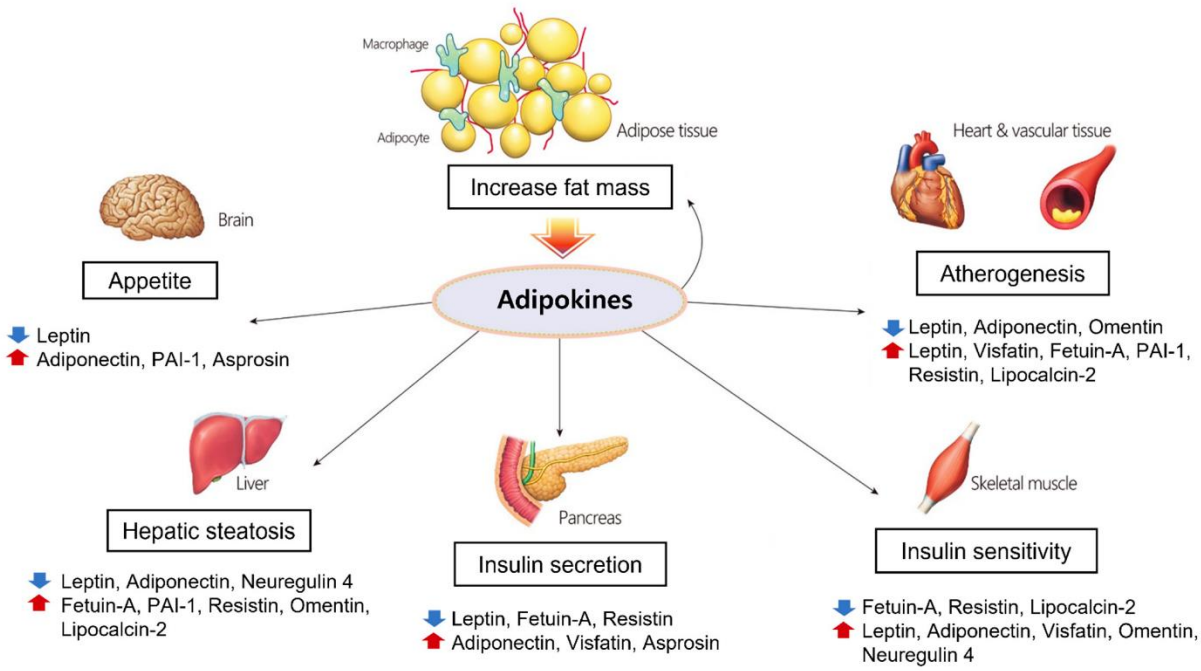


***Figure 31**. Processus physiologiques régulés par les adipokines dans les organes cibles (108).*

Leptine

La leptine est secrétée au niveau des adipocytes et se lie à son récepteur au niveau de l'hypothalamus. Elle contribue principalement à l'équilibre homéostasique, influence la prise alimentaire et régule la dépense énergétique, fonctions contrôlées selon la voie leptine / mélanocortines abordée précédemment (**Figure 30**), La liaison de la leptine inhibe la production du neuropeptide Y / *agouti-related protein* (NPY/AgRP) et stimule la production de POMC, qui après modifications post-traductionnelles permet la production de peptides (α-MSH et β-MSH notamment) sous l'action de la prohormone convertase 1 (PC1/3). α-MSH et β-MSH se lient aux récepteurs MCR3 et MCR4 pour permettre leur activation, conduisant à une baisse de l'apport alimentaire et une augmentation de la dépense énergétique. MRAP2 est capable de diminuer l'activité des récepteurs MC3R et MC4R en réponse à α-MSH et β-MSH, et de conduire à l'obésité. La protéine SIM1, un facteur de transcription, favorise l'activité du récepteur MC4R. La stimulation de MC4R déclenche la production de BDNF, qui se lie au récepteur neurotrophique TrkB (*NTRK2*) et influence ainsi l'apport calorique et la dépense énergétique. La fixation de la leptine à son récepteur active également la voie de signalisation JAK/STAT via le facteur de transcription Stat3 qui, une fois présent dans le noyau de la cellule, permet l'expression de gènes cibles spécifiques conduisant aux mêmes effets anorexigènes.

Outre la régulation de l'homéostasie de l'énergie, la leptine a été corrélée des actions pléiotropiques pour d'autres fonctions physiologiques telles que la communication neuroendocrinienne, la reproduction, l'angiogenèse, la formation osseuse, la réponse immunitaire et l'inflammation (*109-112*).

Adiponectine

L'adiponectine est une autre hormone majeure de l'homéostasie énergétique. Elle est sous-exprimée dans les états de résistance à l'insuline et ses niveaux sont corrélés négativement à plusieurs situations physiopathologiques dont l'obésité abdominale, le syndrome métabolique, le diabète de type 2, la stéatose et la stéato-hépatite non alcoolique (*113*). On lui prête également d'autres effets pléiotropiques, anti-inflammatoires, anti-athérogènes, cardioprotecteurs, et des actions sur le développement de certains cancers, des pathologies intestinales et rénales, la polyarthrite rhumatoïde, la démence liée à l'âge et d'autres maladies cognitives (*113, 114*).

L'adiponectine, connue sous les appellations ACRP30, apM1, AdipoQ (*115, 116*) et GBP28 (*116*), est une hormone excrétée majoritairement par les tissus adipeux sous-cutanés et viscéraux (*117*). Elle cible des cellules spécifiques dans des tissus variés, comme les cellules bêta-pancréatiques, les hépatocytes et les cardiomyocytes (*118*). La protéine est codée par le gène *ADIPOQ* qui comporte 3 exons et 2 introns sur 16 kb, mais seuls 2 exons sont traduits chez l'homme. Le produit du gène est structuré sur 244 acides aminés en forme monomérique de 28 kDa qui subira des modifications post-traductionnelles importantes pour la formation d'oligomères d'adiponectine, fondamentales pour l'inhibition de la néoglucogénèse hépatique et la liaison du ligand à ses récepteurs membranaires (*119*). L'adiponectine circule sous différentes formes moléculaires : trimérique (LMW ou *low molecular weight*), hexamèrique (MMW ou *medium molecular weight*), et une forme multimérique de haute masse moléculaire (HMW ou *high molecular weight*) agrégeant au moins 18 protomères, qui est majoritaire à plus de 80% et serait la plus active métaboliquement, notamment pour les effets insulino-sensibilisateurs (**Figure 32**) (*119*).

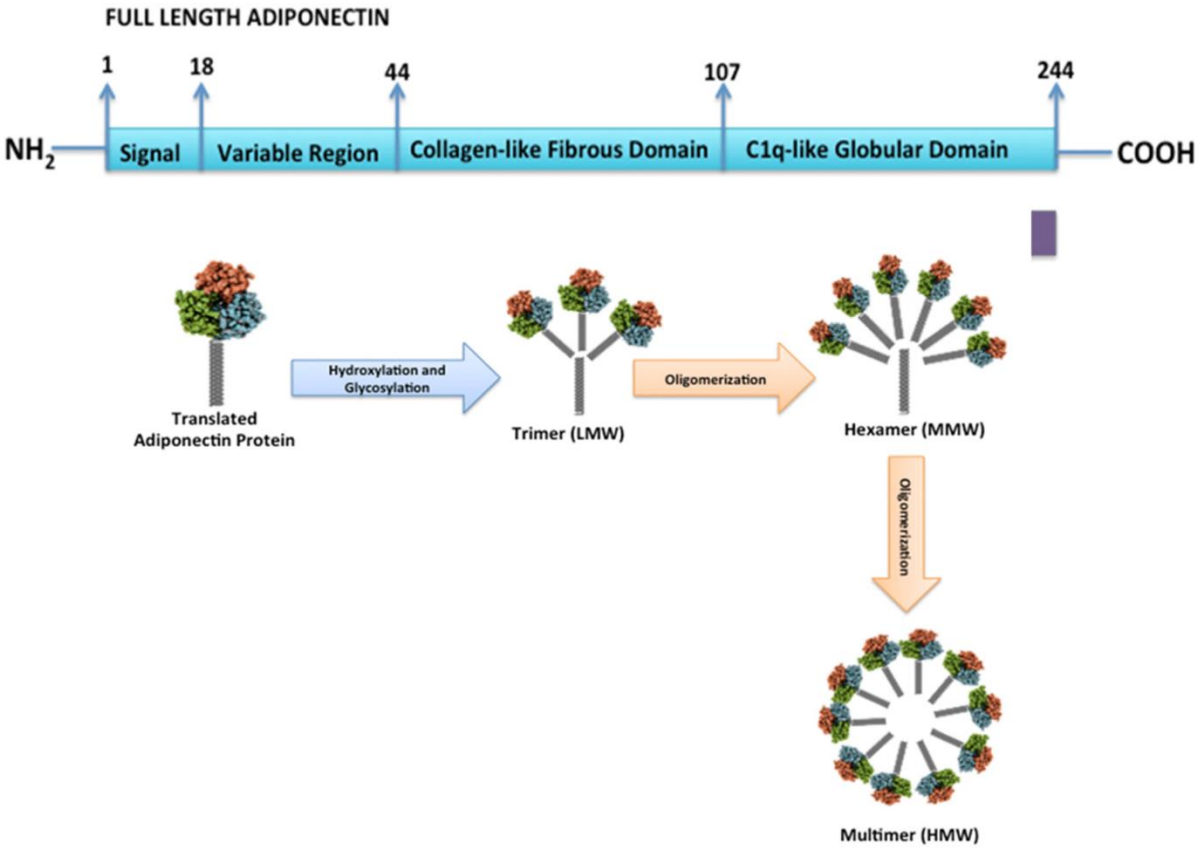


Figure 32. *Domaines et structures de l'adiponectine (114).*

Les niveaux sériques d'adiponectine sont très riches, jusqu'à 1000 plus élevés que pour d'autres hormones comme l'insuline, et que les autres adipokines, comme la leptine (*118*). Ils sont modulés par le régime alimentaire, l'activité physique, le tabagisme, l'exposition au froid, la prise médicamenteuse de sibutramine (traitement d'appoint contre l'obésité nutritionnelle, et parfois en cas de diabète de type 2 ou de dyslipidémie), et les fluctuations du poids (*116, 120-125*). Les niveaux sériques d'adiponectine sont aussi influencés par des facteurs biologiques : le sexe, l'âge, l'ethnicité, l'épigénétique et la génétique (*118, 126-128*). Les estimations d'héritabilité pour les niveaux d'adiponectine sérique s'échelonnent de 33 à 93% pour les enfants et adultes de populations diverses (*129-131*). Des études de gène candidat, de liaison, d'associations pangénomiques et pan-exomiques ont associé de nombreux variants fréquents et rares aux niveaux d'adiponectine (*132-136*).

Il existe 2 récepteurs à l'adiponectine, AdipoR1 et AdipoR2 (*118, 137*). AdipoR1 est exprimé de façon ubiquitaire alors qu'AdipoR2 est surtout exprimé dans le foie (*118*). La liaison des oligomères d'adiponectine à ses récepteurs activerait la protéine adaptatrice APPL1, qui initierait alors une transduction de signal complexe en activant PPAR-α, et en phosphorylant l'AMPK et p38-MAPK (**Figure 33**) (*114*). La signalisation vers l'AMPK puis vers mTOR joue un rôle primordial sur la régulation de l'homéostasie énergétique (*138*). L'adiponectine a des effets anti-apoptotiques, anti-inflammatoires, antiathérogènes, est insulino-sensibilisante et agit

comme un messager pour la communication entre le tissu adipeux et les organes métaboliques (*118, 139, 140*). Chez les souris obèses, les modèles ob/ob soumis à un régime riche en graisses présentent des niveaux diminués d'adiponectine plasmatique (*118*). Les niveaux d'adiponectine sérique sont également inversement associés à divers traits cardiométaboliques chez l'homme, comme le diabète de type 2 (*141, 142*), les maladies cardiovasculaires et l'obésité (*142*). Cependant, toutes les études épidémiologiques conduites à ce stade ont des limitations observationnelles et ne permettent pas d'établir clairement les relations causales entre les traits mentionnés. Des travaux supplémentaires sont donc nécessaires pour associer définitivement l'adiponectine avec les complications cardiométaboliques (*143*).

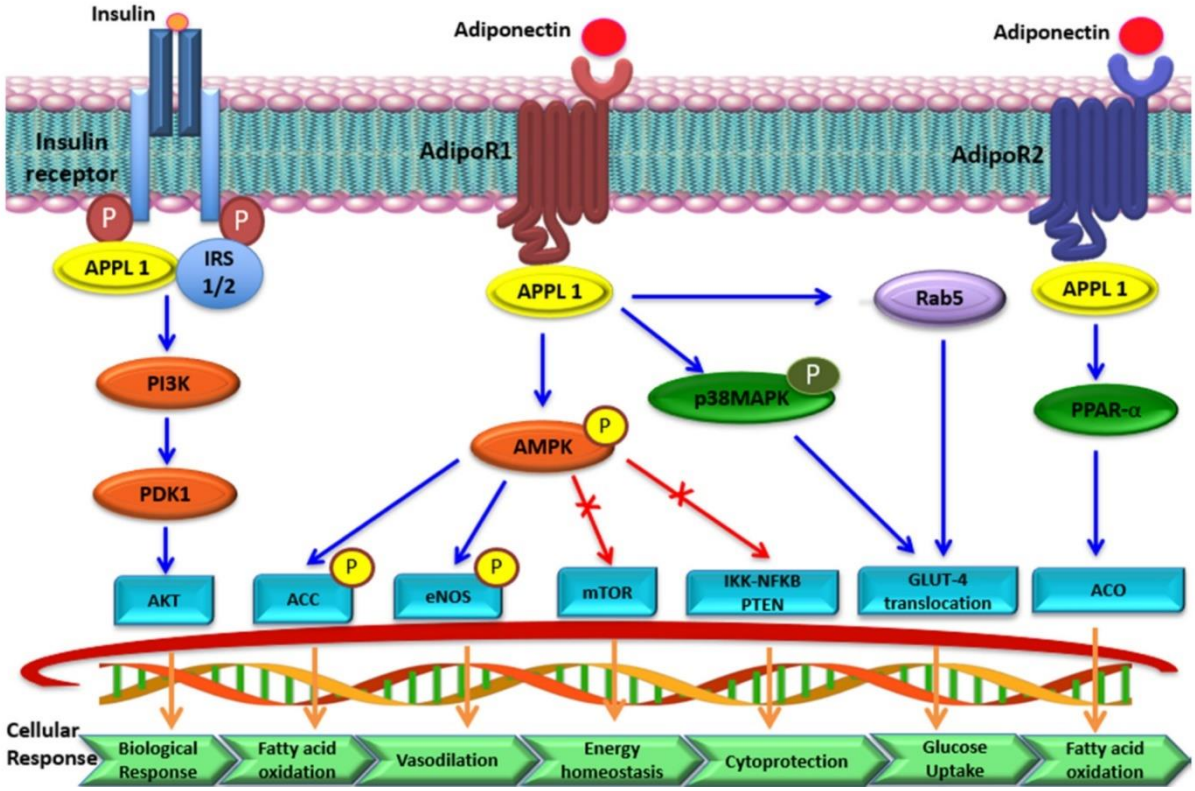


***Figure 33**. Principales voies de signalisation de l'adiponectine (selon l'état des connaissances actuelles) (114).*

Insulinorésistance et diabète de type 2 (DT2)

Bien que leur origine génétique soit différente, l'obésité et le diabète sont deux maladies intrinsèquement liées. En effet, l'obésité est le premier facteur de risque du diabète et 80-90% des diabétiques sont obèses (*144*). Si le diabète, dont le DT2 est la forme largement majoritaire (> 90%), se développe dans le monde, le nombre de diabétiques est passé de 170 millions au début du siècle à presque 400 millions prévus en 2030 (*145*), c'est principalement parce que de plus en plus de personnes sont obèses ou en surpoids. Tout comme l'obésité, le diabète est une maladie qui empêche de vieillir en bonne santé et réduit l'espérance de vie. Une étude très récente sur 23 millions de participants, et couvrant une période allant de 1960 à 2020, démontre

que le diagnostic de plus en plus précoce pour le DT2 voit l'espérance de vie des patients se réduire d'une moyenne de 3-4 ans pour chaque décade de précocité (*146*). Celui-ci peut d'ailleurs être diagnostiqué des années après l'apparition des premiers symptômes.

Dans le DT2, la glycémie ne réagit plus à l'insuline sécrétée par le pancréas. Les cellules bêta des îlots Langerhans y sont défaillantes, entrainant une perte de contrôle de la glycémie et la pathogénèse du DT2, qui devient inévitable si cette altération est accompagnée d'une résistance à l'insuline par les cellules périphériques (*147, 148*). L'insulinorésistance provoque une élévation des acides gras plasmatiques, une diminution du transport du glucose dans les cellules du muscle squelettique et une augmentation de la dégradation des graisses conduisant à une production élevée de glucose hépatique. Le glucose est, de plus, mal capté par les cellules, expliquant l'hyperglycémie observée. Toute personne en surpoids ou obèse présente une forme d'insulinorésistance ou une autre, bien que le diabète ne se développe que chez ceux qui présentent un ratio sécrétion / résistance insuffisant pour l'insuline, son taux pouvant tout de même être élevé dans le DT2 (*149*).

En plus de la prévalence croissante de l'obésité dans tous les groupes d'âges, d'autres facteurs environnementaux et biologiques comme le sexe, l'activité physique, une mauvaise alimentation et l'urbanisation (l'occidentalisation des modes de vie), sont associés à une augmentation des diagnostics de DT2 (*150*). Le tabagisme est également un facteur de risque, probablement en augmentant l'insulinorésistance (*151*). Le syndrome métabolique, déclenché lui-aussi par l'installation d'une obésité androïde qui témoigne d'une insulinorésistance, se situe au confluent de tous ces facteurs et aux avant-postes du DT2 (*152, 153*).

Le risque de développer le diabète au cours de la vie augmente de 7 à 70% lorsque l'IMC passe de moins de 18,5 kg/m à plus de 35 kg/m chez les hommes de plus de 18 ans. Chez les femmes, ce risque augmente de 12 à 74% pour les mêmes valeurs d'IMC (*154*). Au contraire, une perte de poids entraîne une réduction significative de l'incidence du diabète dans les populations à risque. Des modifications du mode de vie telles qu'une légère perte de poids (5 à 10%) et une intensification de l'activité physique (pour atteindre au moins 150 minutes par semaine) réduisent l'incidence du diabète de plus de 50% (*155*). La chirurgie bariatrique, elle, réduit jusqu'à 5 fois l'incidence du DT2 sur une période post-opératoire de 7 ans (*156*). La perte de poids s'accompagne d'un retour proportionnel du contrôle glycémique, et mène parfois à la rémission complète (*157*).

Du point de vue de l'architecture génétique, les mécanismes qui sous-tendent les différences individuelles dans la prédisposition au DT2 restent relativement obscurs. Depuis 2001, les études pangénomiques sur des cohortes de plus en plus grandes et incluant de

nouvelles ethnies ont permis d'identifier plus d'une centaine de marqueurs génétiques du DT2 (**Figure 34**) (*5*). La plupart de ces locus ont été associés à une dysfonction de la cellule bêta pancréatique. Néanmoins, ces marqueurs génétiques permettent d'expliquer moins de 10% de l'héritabilité du DT2.

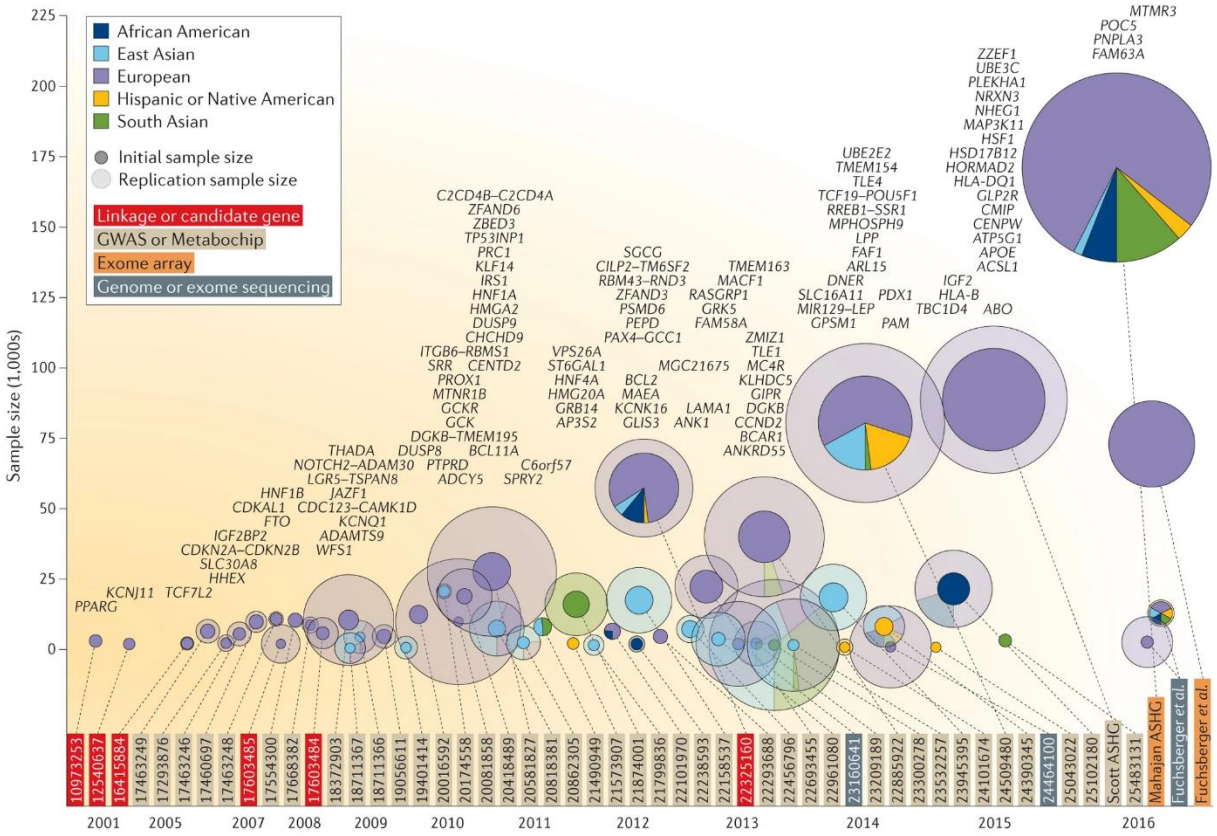


***Figure 34**. Progression des connaissances génétiques grâce aux associations pangénomiques pour le diabète de type 2 (de 2000 à 2016) (5).*

Récemment, une équipe a dévoilé 42 nouveaux variants à risques pour le DT2, dont 3 variants rares, en méta-analysant 63 000 patients *versus* 600 000 contrôles dans les populations européennes. En intégrant leurs résultats avec des données d'expression (eQTL pour *expression Quantitative Trait Locus*) dans le sang pour 17 000 échantillons, ils priorisent 33 gènes potentiellement fonctionnels dans la maladie. Une dernière intégration avec le méthylome de 2000 patients et des annotations de référence pour l'épigénome dévoile les mécanismes régulateurs probables pour au moins 3 gènes (*CAMK1D*, *TP53INP1*, et *ATP5G1*) pour lesquels les variants influencent directement la méthylation de l'ADN puis l'expression des transcrits (*158*). Bien que révélatrices de ce que les études d'associations à grande échelle permettent encore d'obtenir aujourd'hui si l'on augmente à la fois la taille des cohortes, les dimensions des méta-analyses, et si l'on additionne d'autres feuillets omiques au génome, ces analyses n'ont permis d'explorer qu'une petite partie du potentiel de découverte qu'elles recèlent. A quelque exception près, les auteurs ont effectivement limité cette étude multi-omique aux interactions en *cis*, en comparant tous les niveaux omiques gène à gène, c'est-à dire pour un même gène

donné à chaque fois, en descendant verticalement à travers les couches moléculaires du génome au transcriptome. On pourra regretter que l'aspect *trans*, qui prend en compte les interactions horizontales et/ou diagonales entre les molécules, y ait été négligé. C'est ce que proposent les approches trans-omiques, qui pourraient constituer une augmentation non négligeable du potentiel GWAS sur des études de ce type, pour aboutir à une association fine des traits complexes du phénome avec les variants du génome (**Figure 35**) (*9*).

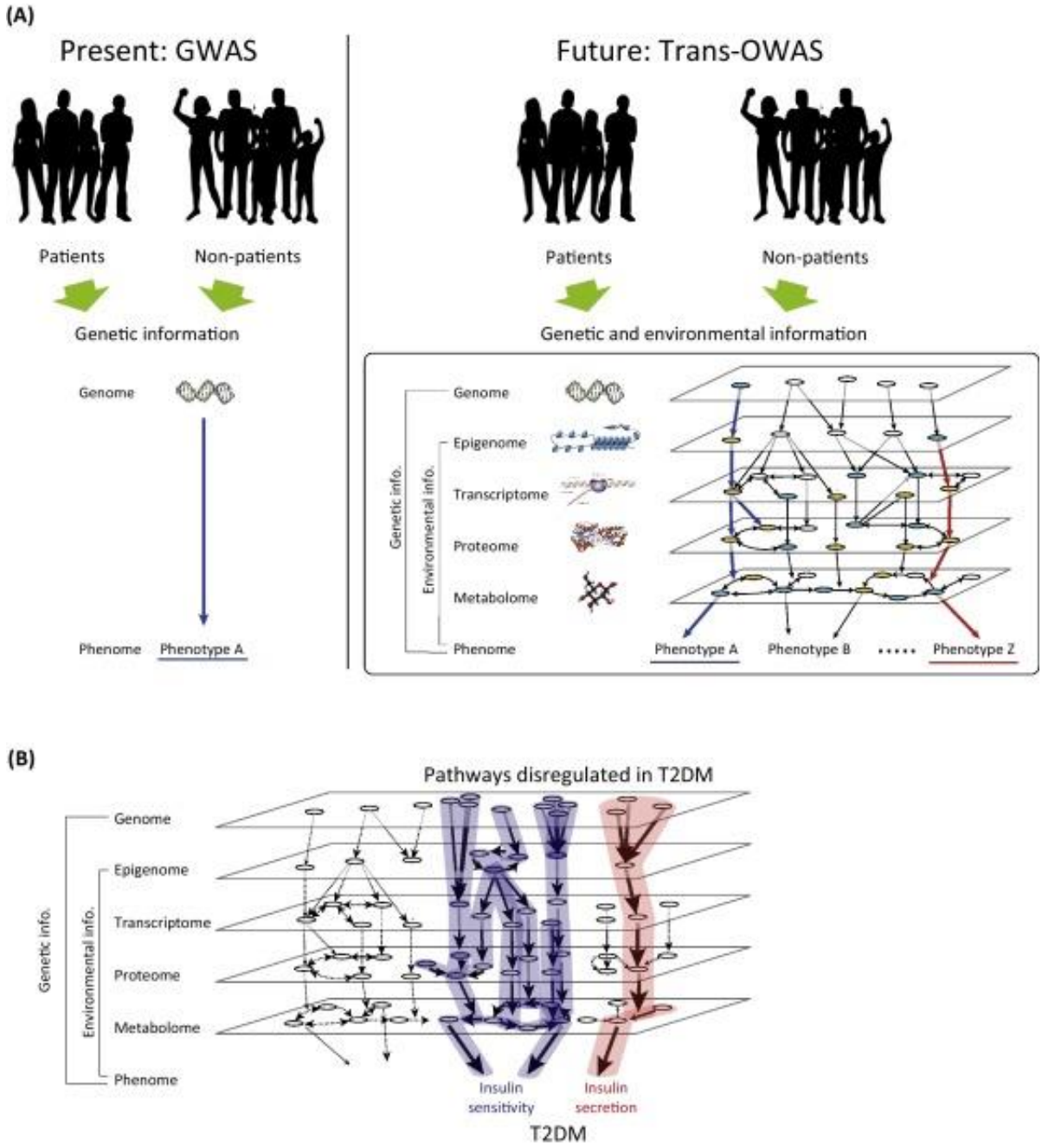


***Figure 35**. Une augmentation possible du GWAS : les trans-omic-wide association studies. (A) A partir des réseaux moléculaires fluctuant entre les couches omiques, il est théoriquement possible d'associer les traits complexes du phénome aux variants du génome. (B) A titre d'exemple, application de la méthode au DT2, permettant de focaliser le signal sur les voies essentielles dans la pathologie (9).*

Lien avec les autres maladies métaboliques

Pléïotropie

L'obésité étant une maladie systémique, complexe et multifactorielle, il est peu étonnant que des dizaines de gènes à fonction pléïotrope puissent contribuer cumulativement et de plusieurs manières à sa pathogénèse. Ce paragraphe n'a pas pour but d'être exhaustif mais de livrer quelques exemples représentatifs supplémentaires des relations obésité / maladies métaboliques, notamment dans les maladies cardiovasculaires, les dysfonctions hépatiques, et l'inflammation.

Outre les adipokines exposées plus haut, aux effets antiathérogènes et hypotenseurs, nous pouvons citer une autre hormone intervenant dans l'obésité par le biais de la stimulation de l'appétit : la ghréline (*159*), un facteur orexigène et adipogène influençant de nombreuses autres voies, gastro-intestinales, neuroendocriniennes, métaboliques et extra endocriniennes, notamment au sein du système cardiovasculaire (*160*). En plus d'avoir des effets sur la performance cardiaque, la ghréline agit comme un facteur de survie en protégeant les cardiomyocytes et les cellules endothéliales contre l'apoptose induite par la doxorubicine *in vitro*. Elle est elle-même synthétisée et sécrétée par les cardiomyocytes (*161*). De fait, la régulation de l'appétit est lui-aussi un phénomène très complexe.

Dans le foie des patients obèses atteints de stéatose, la concentrations des acides gras polyinsaturés oméga-3, aux effets anti-inflammatoires, est diminuée, mais pas des oméga-6, précurseurs de la prostaglandine, impliquée entre autres dans les manifestations inflammatoires, la vasodilatation et la douleur (*162*). L'équilibre entre les deux classes de lipides détermine *in fine* la tendance de la cellule à générer et entretenir l'inflammation par la production de prostaglandines et de leucotriènes. Les acides gras polyinsaturés sont donc eux aussi qualifiés de pléïotropes, parce qu'ils agissent sur une grande variété de mécanismes physiologiques. Outre leur rôle structurel et protecteur comme constituants de la membrane cytoplasmique, et leur fonction dans la réponse immunitaire, ils interviennent aux niveaux de l'expression génique et de la transduction des signaux. Ainsi les oméga-6 stimuleraient, et les oméga-3 inhiberaient, la formation de tissus adipeux pendant la période périnatale (*163*).

Chez les patients obèses, le tissu adipeux, hypertrophié, est une source de médiateurs de l'inflammation tels que le TNFα et l'interleukine-6 (IL-6). De surcroit, ces molécules entravent la lipolyse et la perte de poids (*164*). Un autre médiateur intracellulaire, NF-κB, se déplace vers le noyau une fois activé pour réguler l'expression de médiateurs inflammatoires, tels que l'IL-6 et l'IL-18. Sa protéine accessoire, IKK-β, est sensibles aux niveaux d'acides gras cytoplasmiques et active NF-κB (*165*), ce qui conduit à une élévation conséquente de la

production des cytokines inflammatoires et des effets immunosuppresseurs dans l'obésité (**Figure 36**) (*166*). L'inflammation et l'immunosuppression systémiques contribueraient en partie à expliquer l'association entre l'obésité et le diabète (*167*), l'asthme (*168*), la dépression et le trouble bipolaire (*169*), et le cancer (*170*), entre autres comorbidités.

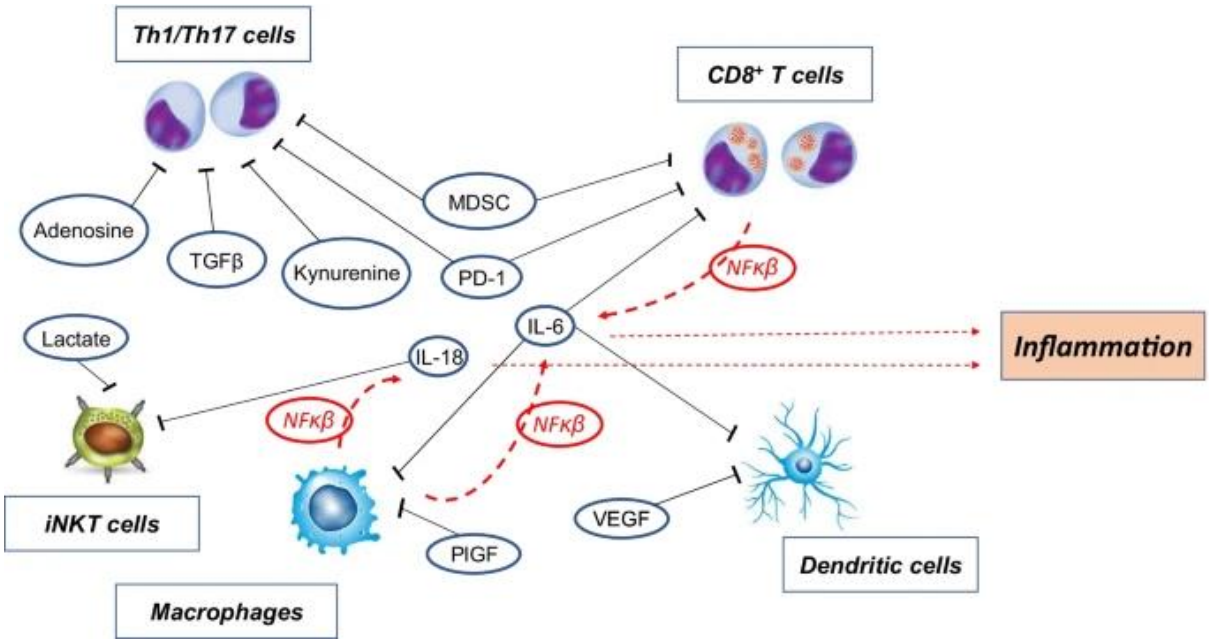


***Figure 36**. Effets de l'obésité sur l'immunosuppression et l'inflammation (166).*

Réseaux pléiotropiques

Chez l'homme, des études génétiques concernant un grand nombre de traits ont maintenant été réalisées. Cette accumulation de connaissances ouvre la voie à l'étude de la pléïotropie, lorsqu'un seul changement génétique affecte plusieurs traits. Cette exploration porte le potentiel de découverte thérapeutique, soit en augmentant le nombre d'indications potentielles d'un médicament, soit en évitant leurs effets secondaires indésirables.

Les protéines qui interagissent entre elles ont tendance à participer aux mêmes fonctions cellulaires et contribuent aux mêmes effets sur l'organisme (*171*). Dans ce sens, les réseaux d'interaction protéine / protéine peuvent être utilisés pour prédire la fonction d'autres gènes dans une maladie (*172*). Ces réseaux augmentent donc le potentiel des études pangénomiques : en utilisant des gènes associés par GWAS comme points de départ dans un réseau d'interactions protéiques construit pour identifier des gènes supplémentaires ou manquant, ces derniers seront eux aussi associés aux mêmes traits (*173, 174*). Les modules d'interaction obtenus partagent des traits phénotypiques : c'est sur la base de ce raisonnement qu'a été établi le concept de réseaux d'interactions pléiotropiques (**Figure 37**) (*175*).

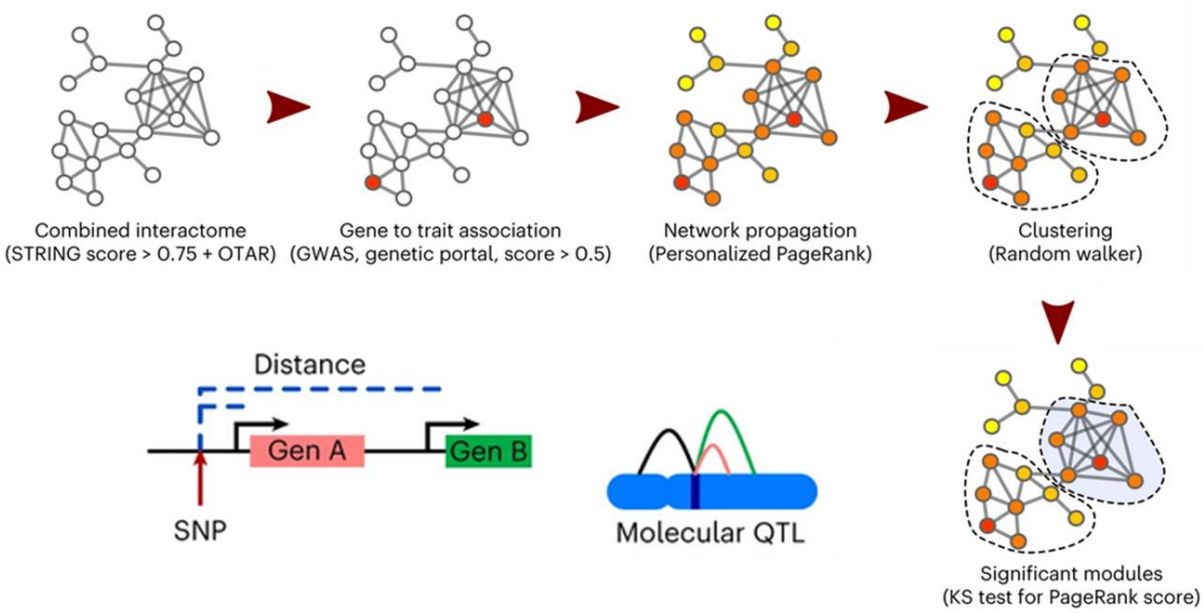


Figure 37*. Une seconde augmentation du GWAS : les cheminements pas à pas dans les réseaux d'interactions protéine / protéine à partir des scores d'associations permettent de singulariser des modules pléiotropiques (175).*

Les auteurs de cette excellente approche montrent que l'ensemble des locus GWAS significatifs à l'échelle du génome est enrichi en gènes codant pour des cibles médicamenteuses bien établies. Les gènes liés à un ou plusieurs traits par expansion du réseau sont enrichis de la même manière, même en excluant les gènes ayant un support génétique direct. La méthode permet également d'augmenter les données GWAS pour plus de 1000 traits par expansion du réseau d'interactions, et à identifier un grand nombre de processus cellulaires pléïotropes chez l'humain. En conclusion, cette expansion du réseau permet de récupérer des gènes impliqués dans les maladies, connus mais non associés par GWAS, d'identifier des groupes de traits sous l'influence des mêmes processus cellulaires, et de définir une carte pléiotropique pour la biologie cellulaire (*175*). A titre d'exemple, les traits cardiovasculaires y sont représentés en relation étroite avec certains biomarqueurs sanguins connus, avec les lipoprotéines, les maladies endocrines, et les cancers.

Obésité et cancers

Environ 30 à 50% des cancers parmi les plus communs pourraient être évités par un poids optimal, entretenu par une alimentation saine et équilibrée et une activité physique suffisante, selon les recommandations du *World Cancer Research Fund* (WCRF) (*176*). Bien que le message soit clair et répandu, seule une faible fraction des gens sondés (moins de 10%) sur tous les continents connait la relation entre obésité et cancer et suit ces recommandations (**Figure 38**) (*176, 177*).

Figure 38. *Affiche grand public présentant les conséquences d'un surpoids ou de l'obésité sur le cancer (source : Département Prévention Cancer Environnement, Centre Léon Bérard, 2020, https://www.cancer-environnement.fr).*

En France, en 2015, 18 600 nouveaux cas de cancers (5,4%) étaient imputés à une surcharge pondérale (**Figure 39**) (*178*). Selon l'IARC (*International Agency for Research on Cancer* ou Centre International de Recherche sur le Cancer - CIRC), une entité dépendant de l'OMS, 481 000 nouveaux cas de cancer par an (environ 4%) peuvent être attribués au surpoids et à l'obésité dans le monde (*179*). Cette proportion est plus importante chez les femmes (5,4%) que chez les hommes (1,9%). En cause : le nombre élevé de cancers spécifiquement féminins qui ont une origine hormonale (sein post-ménopause, endomètre, ovaires), en lien avec la surcharge pondérale (*180*). En effet, une production œstrogénique anormale et accrue par un complexe d'aromatases à partir d'hormones des graisses corporelles est observée dans l'obésité (*180*). L'obésité de la mère aggrave aussi le risque de cancer du testicule chez le futur enfant, probablement pour les mêmes raisons hormonales (*181*).

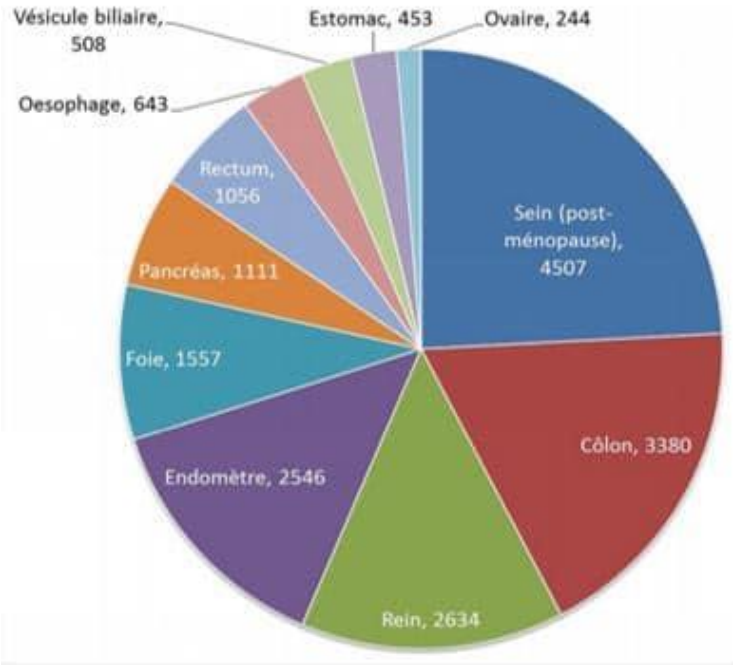


***Figure 39**. Nombre de cancers attribuables au surpoids et à l'obésité en France en 2015 chez les hommes et les femmes (178).*

Plus largement, les cancers colorectaux, de la vésicule biliaire, de la prostate, du pancréas, de la vessie, de l'œsophage, de l'estomac, de la bouche et des reins seraient favorisés par la surcharge pondérale : de nombreuses études associent obésité, surpoids, et prédispositions aux cancers de façon convaincante (**Table 1**) (*176, 179, 182*). Un IMC de 30-35 augmenterait même le risque de mourir du cancer de plus de 30% (sources : Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. WCRF / *American Institute for Cancer Research*, rapport 2018). Les hémopathies malignes comme les leucémies, lymphomes et myélomes (*183*), et les tumeurs du système nerveux central comme les méningiomes (*184*) semblent également favorisées par la surcharge pondérale, avec un niveau de preuve jugé satisfaisant par le WCRF et/ou l'IARC. Pour certains cancers, l'obésité au moment du diagnostic, ainsi que la prise de poids au cours du traitement, sont associées à un risque accru de récidive, de développement d'un second cancer et de mortalité liée au cancer initial (**Figure 38**) (*185, 186*). Les chances de survie sont moindres chez l'obèse dans la plupart des cancers, sauf semble-t-il pour le cancer du poumon, du rein et le mélanome (méta-analyse de 203 études et 6,3 millions de participants) (*187*). En outre, les cancers des obèses seraient détectés plus tardivement et traités moins efficacement (*188*). Enfin, l'obésité au cours de l'enfance et/ou l'adolescence augmenterait le risque de développer un cancer à l'âge adulte (*189*). Des travaux récents ont conclu que le risque de cancer du pancréas augmentait progressivement avec l'IMC pour ceux qui étaient obèses ou en surpoids dès l'adolescence, par rapport à ceux qui n'ont jamais été obèses ou en surpoids à cet âge (*190, 191*).

Localisation de cancer ou type	Indications chez l'homme	Risque relatif de la catégorie d'IMC la plus haute par rapport à un IMC normal (IC à 95%)
Œsophage : adénocarcinome	Suffisantes	4,8 (3,0–7,7)
Cancer du cardia	Suffisantes	1,8 (1,3–2,5)
Côlon et rectum	Suffisantes	1,3 (1,3–1,4)
Foie	Suffisantes	1,8 (1,6–2,1)
Vésicule biliaire	Suffisantes	1,3 (1,2–1,4)
Pancréas	Suffisantes	1,5 (1,2–1,8)
Sein : postménopause	Suffisantes	1,1 (1,1–1,2)
Corps de l'utérus	Suffisantes	7,1 (6,3–8,1)
Ovaire	Suffisantes	1,1 (1,1–1,2)
Rein (cellules rénales)	Suffisantes	1,8 (1,7–1,9)
Méningiome	Suffisantes	1,5 (1,3–1,8)
Thyroïde	Suffisantes	1,1 (1,0–1,1)
Myélome multiple	Suffisantes	1,5 (1,2–2,0)
Cancer du sein chez l'homme	Limitée	SO
Cancer fatal de la prostate	Limitée	SO
Lymphome diffus à grandes cellules B	Limitée	SO
Œsophage : carcinome épidermoïde	Insuffisantes	SO
Cancer de l'estomac non cardial	Insuffisantes	SO
Voies biliaires extrahépatiques	Insuffisantes	SO
Poumon	Insuffisantes	SO
Peau : mélanome cutané	Insuffisantes	SO
Testicule	Insuffisantes	SO
Vessie	Insuffisantes	SO
Cerveau ou moelle épinière : gliome	Insuffisantes	SO

***Table 1**. Localisations de cancers associés à l'obésité et niveaux de preuve correspondants (source : IARC Lyon). IC : intervalle de confiance, SO : sans objet.*

L'obésité ne change pas que la physiologie corporelle et le métabolisme, elle modifie profondément les programmes protéomiques, transcriptomiques et épigénétiques dans de multiples tissus, et ces dérégulations peuvent être liées à de nombreux attributs des cancers (*10*). En favorisant un état inflammatoire chronique qui se traduit par une augmentation des taux sanguins de facteurs pro-inflammatoires tels que le TNFα, l'interleukine 6, et la protéine C-réactive, l'obésité a pour conséquence de rompre l'équilibre entre prolifération, différentiation cellulaire et apoptose et ainsi de contribuer au processus d'initiation de la cancérogenèse (*182*). En supplément des cytokines et hormones anaboliques citées précédemment (œstrogènes), et chez les sujets ayant un IMC élevé à très élevé (> 30), on observe également une dérégulation des taux endogènes de plusieurs facteurs de croissance, enzymes du métabolisme énergétique, adipokines et autres hormones : insuline, IGF-1, AMPK, leptine, ghréline, résistine, adiponectine (*192*). Certains mécanismes sont communs à toutes les localisations de cancers, tels que l'augmentation de la résistance à l'insuline due à l'excès de tissus adipeux, ou l'augmentation des niveaux de leptine, et contribueraient à leur promotion par des effets mitogènes, pro-angiogènes et anti-apoptotiques (**Figure 40**).

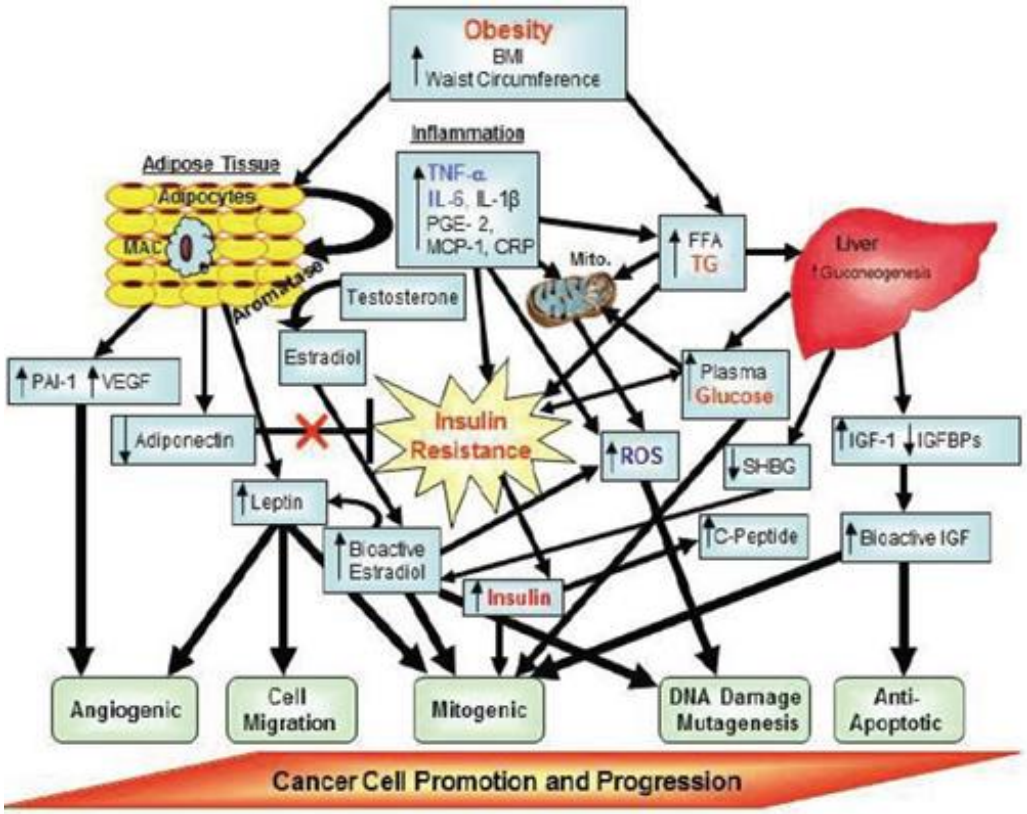


Figure 40 *: Facteurs potentiellement impliqués dans la carcinogénèse en lien avec l'obésité (source :* The Role of Obesity in Cancer Survival and Recurrence: Workshop Summary, Nathan Berger, 2012).

Dans une revue récente de la littérature, Harris *et al.* discutent de la manière dont les endophénotypes des cancers pourraient être affectés par l'obésité (*166*), en mettant en évidence des points communs majeurs et émergents, tels que le lien entre l'inflammation favorisant la croissance tumorale et l'état inflammatoire chronique observé dans l'obésité (*193*). Leur message principal se résume aux interactions avérées ou probables entre une soixantaine de biomarqueurs, moléculaires pour la plupart, et 10 caractéristiques incontournables de l'initiation, de la promotion et de la prolifération des cancers (**Figure 41**). Ces biomarqueurs sont regroupés en une dizaine de facteurs systémiques affectés par le terreau pondéral : les adipokines (leptine, adiponectine, résistine), les facteurs alimentaires (acides gras, dérivés du cholestérol, glucose, PCSK9, kinurénine, chymotrypsine), les onco-métabolites (lipides peroxydés, dérivés réactifs de l'oxygène, lipopolysaccharides, acide désoxycholique), les cytokines et apparentées (interleukines, chimiokines, interféron γ, TNFα, TGFβ1, MCP-1), les cellules immunitaires et leurs marqueurs de surface (lymphocytes B épuisés, lymphocytes T CD8+, lymphocytes T régulateurs, lymphocytes T *helpers*, lymphocytes T *natural killers* invariants, lymphocytes T *mucosal-associated* invariants, cellules souches myéloïdes, éosinophiles, macrophages pro-inflammatoires, macrophages du tissu adipeux et métaboliquement actifs, pièges extracellulaires des neutrophiles), les protéines fibrinolytiques (PAI-1, facteur tissulaire), les hormones (insuline, œstrogènes, cholécystokinines), les facteurs de croissances et leurs interacteurs (des familles VEGF, IGF et FGF principalement), les

composants de la matrice extracellulaire, et quelques autres molécules comme des microARNs ou la céruloplasmine. On notera que les 4 attributs des cancers « signalisation proliférative continue », « activation de l'invasion tissulaire et métastases », « inflammation pro-tumorale » et « échappement au système immunitaire » sont hautement pléiotropiques sur la base de ces marqueurs, leurs dérégulations chroniques, même faibles, s'accumulant probablement entre elles jusqu'à provoquer le basculement irréversible. Dans ce tableau, le rôle joué par le système immunitaire serait de premier plan. Toutefois, la majorité des études décrites dans cette revue ont été réalisées sur des modèles animaux et des lignées cellulaires (*166*). Des travaux supplémentaires et innovants du point de vue moléculaire sont nécessaires pour désambiguïser la relation complexe et cumulative entre obésité et carcinogénèse chez le patient.

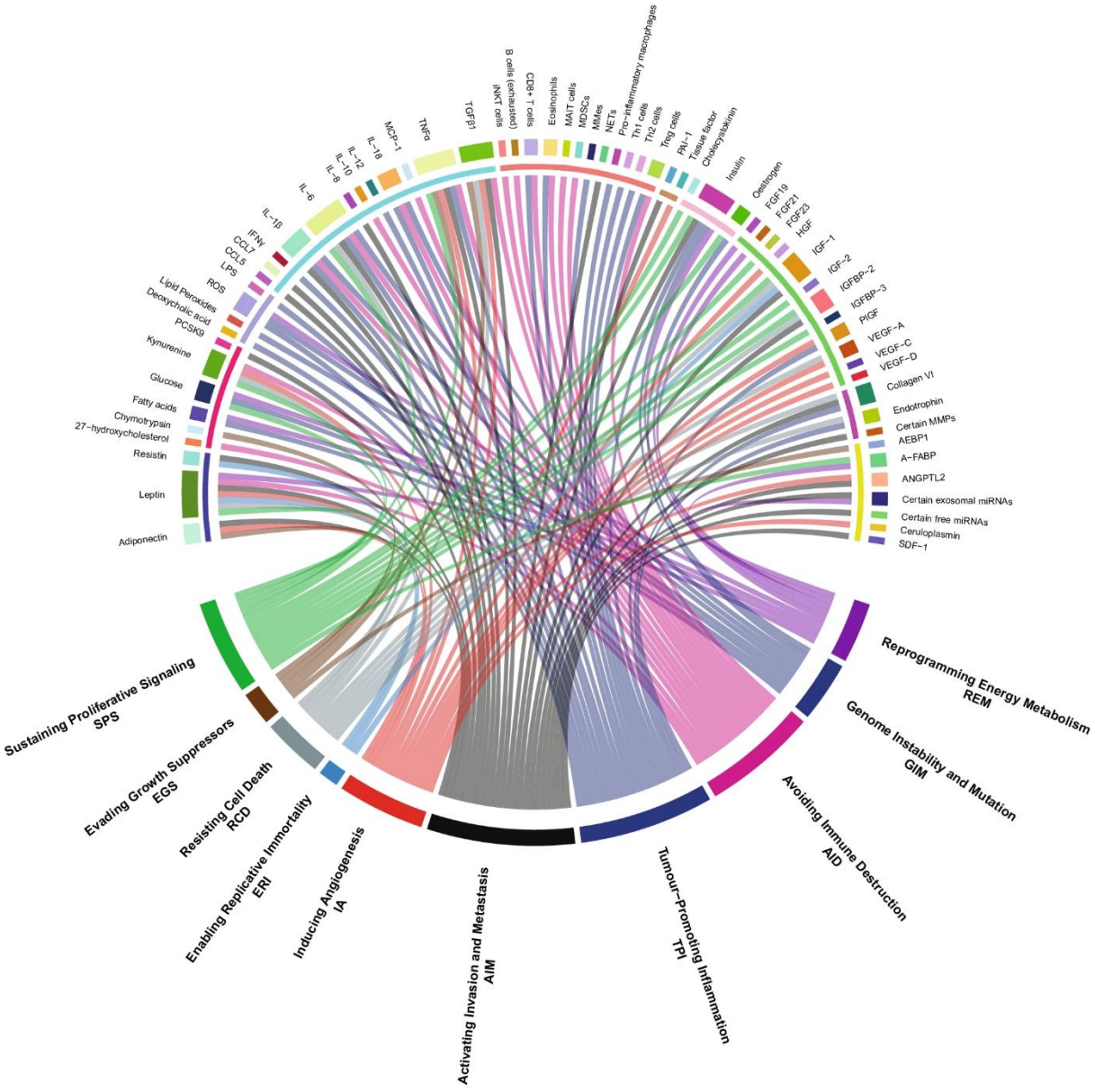


Figure 41. *Influences potentielles de l'obésité sur les 10 principaux attributs des cancers, d'un point de vue moléculaire. Circos plot avec biomarqueurs de l'obésité en haut et caractéristiques des cancers en bas (166).*

RÉSULTATS

PARTIE 1 : GÉNOMIQUE DES CANCERS CÉRÉBRAUX

1. KI-67 AND MCM6 LABELING INDICES ARE CORRELATED WITH OVERALL SURVIVAL IN ANAPLASTIC OLIGODENDROGLIOMA, *IDH1*-MUTANT AND 1P/19Q-CODELETED: A MULTICENTER STUDY FROM THE FRENCH POLA NETWORK

Contexte

Le premier article de cette série « Cancers Cérébraux », indexé officiellement en mai 2020 (prépublication en ligne en septembre 2019), donne la mesure des travaux effectués dans l'équipe « Génomique des Cancers » du laboratoire et permet de retracer l'évolution du développement d'approches bioinformatiques en fonction des besoins des projets. On peut voir ici le point culminant de pipelines d'analyses du transcriptome bien rodés, et de l'exome dans une moindre mesure, avec des méthodes concurrentielles d'inter-validations supervisées (essentiellement biostatistiques) / non supervisées (classifications, partitionnement et réductions de dimensions). Pour ce travail, une innovation est particulièrement intéressante : la classification par gradient de k-moyennes, qui permet de découvrir des clusters différentiels selon un poids d'ordonnancement des échantillons : ici ce sont des expressions incrémentales de protéines ou de leurs ARN messagers.

Points clés

- Découvrir de nouveaux biomarqueurs à valeurs pronostiques dans les oligodendrogliomes anaplasiques, mutés pour *IDH1* et codélétés 1p + 19q. Ce sont des gliomes de haut grade au statut mutationnel bien établi, agressifs, traitables mais incurables sur le long terme.
- Évaluer et décrire l'expression de deux marqueurs de prolifération, Ki-67 et MCM6, et l'influence de la surexpression de *MKI67* et *MCM6* sur les transcriptomes d'une large cohorte de patients du réseau multicentrique français POLA (Prise en charge des Oligodendrogliomes Anaplasiques).
- Comprendre l'association de ces marqueurs avec la survie (générale et progression) et valider les résultats de transcriptomique sur des jeux de données du TCGA (*The Cancer Genome Atlas*).

RESEARCH ARTICLE

Ki-67 and MCM6 labeling indices are correlated with overall survival in anaplastic oligodendroglioma, *IDH1*-mutant and 1p/19q-codeleted: a multicenter study from the French POLA network

Celso Pouget[1,2,*]; Sébastien Hergalant[2,*]; Emilie Lardenois[1,2]; Stéphanie Lacomme[3]; Rémi Houlgatte[2]; Catherine Carpentier[4]; Caroline Dehais[5]; Fabien Rech[6,7]; Luc Taillandier[8]; Marc Sanson[4,5,9]; Romain Appay[10,11]; Carole Colin[10]; Dominique Figarella-Branger[10,11]; Shyue-Fang Battaglia-Hsu[2]; Guillaume Gauchotte[1,2,3,*] Investigators: The POLA Network

[1] Department of Pathology, CHRU, Nancy, France.
[2] INSERM U1256, NGERE, Faculté de Médecine de Nancy, Université de Lorraine, Vandoeuvre-lès-Nancy, France.
[3] Centre de Ressources Biologiques, CHRU, BB-0033-00035, Nancy, France.
[4] Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Inserm U 1127, CNRS UMR 7225, ICM, F-75013, Paris, France.
[5] AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de Neurologie 2-Mazarin, 75013, Paris, France.
[6] Department of Neurosurgery, CHRU, Nancy, France.
[7] Institut des Neurosciences, INSERM U1051, Montpellier, France.
[8] Department of Neurology, CHRU, Nancy, France.
[9] Onconeurotek, Groupe Hospitalier Pitié-Salpêtrière, Paris, France.
[10] Aix-Marseille Univ, CNRS, INP, Inst. Neurophysiopathol, Marseille, France.
[11] AP-HM, Hôpital de la Timone, Service d'Anatomie Pathologique et de Neuropathologie and Centre de Ressources Biologiques CRB-TBM, BB-0033-00097, Marseille, France.

Keywords
1p/19q codeletion, anaplastic oligodendroglioma, glioma, immune response, immunohistochemistry, Ki-67, MCM6, proliferation, pro-neural, transcriptomics.

Corresponding author:
Guillaume Gauchotte, MD, PhD, Service d'Anatomie et Cytologie Pathologiques, Hôpital Central, CHRU de Nancy, 29, avenue du Maréchal De Lattre de Tassigny, Nancy 54000, France
(E-mail: *g.gauchotte@chru-nancy.fr*)

Received 23 May 2019
Accepted 19 September 2019
Published Online Article Accepted 27 September 2019

*These two authors have contributed equally to this work.

doi:10.1111/bpa.12788

Abstract

Anaplastic oligodendroglioma (AO), IDH-mutant and 1p/19q codeleted (IDHmut+/1p19qcodel), is a high-grade glioma with only limited prognostic markers. The primary objective of this study was to evaluate, by immunohistochemistry, the prognostic value of two proliferation markers, MCM6 and Ki-67, in a large series of IDHmut+/1p19qcodel AO included in the POLA ("Prise en charge des Oligodendrogliomes Anaplasiques") French national multicenter network. We additionally examined the transcriptome obtained from this series to understand the functional pathways dysregulated with the mRNA overexpression of these two markers. The labeling indices (LI) of MCM6 and Ki-67 were obtained via computer-assisted color image analyses on immunostained AO tissues of the cohort (n = 220). Furthermore, a subgroup of AO (n = 68/220) was used to perform transcriptomic analyses. A high LI of either MCM6 (≥50%) or Ki-67 (≥15%) correlated with shorter overall survival, both in univariate (P = 0.013 and P = 0.004, respectively) and multivariate analyses (P = 0.027; multivariate Cox model including age, mitotic index, MCM6 and Ki-67). MCM6 and Ki-67 LI also correlated with overall survival in an additional retrospective cohort of 30 grade II IDHmut+/1p19qcodel oligodendrogliomas. The prognostic value of *MCM6* mRNA level was confirmed in The Cancer Genome Atlas (TCGA) IDHmut+/1p19qcodel gliomas. The transcriptomic approach revealed that high transcriptional expressions of *MCM6* and *MKI67* were both linked positively with cell cycle progression, DNA replication, mitosis, pro-neural phenotype as well as neurogenesis, and negatively with microglial cell activation, immune response, positive regulation of myelination, oligodendrocyte development, beta-amyloid binding and postsynaptic specialization. In conclusion, the overexpression of MCM6 and/or Ki-67 is independently associated to shorter overall survival in IDHmut+/1p19qcodel AO. These two easy-to-use and cost-effective markers could thus be used concurrently in routine pathology practice. Additionally, the transcriptomic analyses showed that AO with high proliferation index have down-regulated immune response and lower microglial cells activation, and bears pro-neural phenotype.

INTRODUCTION

According to the World Health Organization (WHO), *IDH*-mutant and 1p/19q-codeleted (IDHmut+/1p19qcodel) anaplastic oligodendrogliomas (AOs) are high-grade gliomas, corresponding histologically to WHO grade III primary brain tumors (35). AO are rare, accounting for 0.5% of all primary brain tumors and one-third of oligodendroglial tumors (38). They are histologically defined as tumors of the oligodendroglial cells, with features of anaplasia, including increased cellularity and mitotic activity, microvascular proliferation and/or necrosis. Although Ki-67 labeling index (LI) is usually used to help differentiate low-grade from high-grade oligodendroglioma, no clear cut-off point has been established (12). By definition, and according to the 2016 update of the WHO Classification of Central Nervous System, AO carry both *IDH1/2* mutations and 1p/19q codeletion (10, 43). The standard of care for these tumors includes surgical resection in conjunction with adjuvant therapies like combined adjuvant radiochemotherapy with procarbazine, CCNU (lomustine), and vincristine (RT-PCV) to improve the overall survival (OS) (46).

The prognostic role of the WHO grading in IDH-mutant gliomas remains debatable, and this is particularly true in the case of oligodendrogliomas (43). In AO patients, young age, high Karnofsky index/clinical performance status, complete surgical resection and adjuvant therapy such as RT-PCV have been shown to associate with longer OS (45, 46). Histological features such as microvascular proliferation, mitotic index (MI) and necrosis have been found associated with three distinct prognostic subgroups of IDH-mut+/1p19qcodel AO with other genomic alterations (19). These include mutations of *CIC* and *TCF12,* both linked to more aggressive AO (4, 22, 32), and allelic loss of 9p21.3, associated with shorter progression-free survival (PFS) and shorter OS in IDH-mut+/1p19qcodel AO (2). Omics approaches using transcriptome, genome and methylome have also been used to identify potential prognostic factors. One such study has successfully classified oligodendroglial tumors into three molecular subgroups with distinct clinical behaviors (29). Machine learning algorithms aiming to identify copy number variations have also been applied recently to AO samples; such technique seems promising in helping with prognostic stratification (41). Despite of these advances, however, it is worth noting that all the above-mentioned markers of AO need expansive and time-consuming molecular analyses, and thus cannot be widely applied to all samples in all laboratories.

In contrast, immunohistochemical proliferation markers such as Ki-67, whose LI correlates strongly with prognosis of many tumor types, are simple, effective and economical means to assess tumor aggressivity. In the case of AO, past few evidence obtained using univariate—but not multivariate—analysis appeared to support the prognostic value of Ki-67 LI (40). However, as less than 50% of the AO tissues in this latter study (performed prior to the 2016 revision of the WHO classification) actually harbored *IDH* mutations and 1p/19q codeletion, its validity shall be questioned. In a more recent study, Zeng *et al* evaluated the prognostic value of Ki-67 upon classifying gliomas into subgroups with either the presence or the absence of *IDH1/2*; however, data concerning 1p/19q values codeletion were lacking in the studied gliomas (49).

Several proliferation markers other than Ki-67 have also been studied in brain tumors. Notably, we previously reported that Minichromosome Maintenance Complex component 6 (MCM6) was overexpressed in meningioma, and was associated with higher histological grade and risk of recurrence (21). In adamantinomatous craniopharyngiomas, MCM6 correlated with a higher risk of long-term recurrence (48). In gliomas, especially in glioblastomas of the Chinese Cancer Genome Atlas (CCGA), *MCM6* mRNA overexpression was also reported to be correlated with poor overall survival (9).

Minichromosome Maintenance proteins (MCMs) play a key role in DNA synthesis and replication, forming a hexameric helicase complex around the DNA (31). All MCMs (MCM2-7) are detectable during the different phases of the cell cycle, including G1, S, G2 and M, but are absent in G0 (33). These proteins are also expressed earlier during G1, in comparison with Ki-67. In other solid tumors, like non-small cell lung carcinomas (47), hepatocellular carcinomas (34), endometrial carcinomas (26), low-grade chondrosarcomas (24) and mantle cell lymphomas (42), a high MCM6 LI correlated with a worse prognosis. To our knowledge since the revision of the 4th WHO classification of 2016, MCMs, and notably MCM6, have never been specifically studied in IDHmut+/1p19qcodel AO.

The primary goal of this study was therefore to evaluate and compare, by immunohistochemistry, the prognostic value of MCM6 and Ki-67 in a large series of IDHmut+/1p19qcodel AO obtained from the POLA ("Prise en charge des Oligodendrogliomes Anaplasiques") French national multicenter network. We additionally examined the transcriptomes obtained from part of this series to understand the functional pathways dysregulated with the mRNA overexpression of these two markers.

MATERIAL AND METHODS

Population and clinicopathological data

Two hundred and thirty-one cases of IDHmut+/1p19qcodel AO were retrieved from the French national multicenter POLA cohort. Clinical data, such as age, sex, extent of surgical removal, type of adjuvant treatment and OS are available in the database, as well as molecular data, including *IDH* mutation status and presence of 1p/19q codeletion. All cases included in this cohort have been centrally reviewed by the neuropathologists of the national board of French national POLA network and were classified according to the 2016 4th WHO classification update. The MI and Ki-67 LI were evaluated in whole tissue sections, as previously described (18). In 220 out of 231 patients, tissue microarray (TMA) blocks were designed with representative samples of tumors, in order to perform immunohistochemical analyses. One to three spots were available for each case.

Additionally, in order to evaluate the prognostic value of MCM6 and Ki-67 LI in low-grade tumors, 30 cases of grade II IDHmut+/1p19qcodel oligodendrogliomas (IDHmut+/1p19qcodel OII) were retrieved from the files of the Department of Neuropathology at Pitié-Salpétrière Hospital (AP-HP, Paris; OncoNeuroTek database; Pitié-Salpétrière) and the Department of Pathology at Nancy University Hospital (CHRU, Nancy; Centre de Ressources Biologiques, BB-0033-00035).

Ethics

Anonymity was strictly respected, according to the principles of the declaration of Helsinki and national ethical guidelines. Patients consent for clinical data collection and genetic analyses have been obtained prospectively, according to POLA network policies. The study was approved by the ethics committee of Hôpital Universitaire la Pitié-Salpêtrière.

Immunohistochemistry

Paraffin sections of 5 μm thickness were immersed in a 10mM sodium citrate buffer (pH 6) for 20 minutes at 97°C for dewaxing and antigen retrieval. The following primary antibodies were used: MCM6 (1/400; goat polyclonal, Santa Cruz Biotechnology, Heidelberg, Germany), Ki-67 (1/200; mouse monoclonal, MIB-1, Dako Cytomation, Glostrup, Denmark).

Immunohistochemistry was performed with Dako Autostainer Plus (Dako) and the Flex + Envision revelation system (Dako).

Ki-67 and MCM6 evaluation

Ki-67 and MCM6 labeling indices (LI) were defined as the percentage of cells with positive nuclear stain of the two individual markers, independently from the signal intensity. For each analyzed TMA spot, the field with the strongest immunostaining was selected by the observer, blinded to the clinical data and outcome of the respective patients. To limit inter-observer variability, cell counting was performed with a computerized color image analyzer (Olympus Cellsens Dimension, Olympus Medical System and Micro-Imaging Group, Hamburg, Germany). A x20 objective was used to take one microphotograph of the area of interest. A minimal object size of 50 pixels was required to count positive nuclei. Vessels and microcalcifications were excluded from the analysis. The percentage of positive and negative cells were automatically computed by the color image analyzer. A mean LI was calculated for each case. Additionally, in order to validate this computer-based LI, manual Ki-67 and MCM6 LI were evaluated in 50 randomly selected cases, by counting 1000 cells in TMA spots.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 23.0 (IBM Corp., Armonk, NY, USA).

Because the quantitative variables did not pass the Kolmogorov–Smirnov normality test, non-parametric tests were used. Spearman correlation test was used to explore the correlation between Ki-67 LI, MCM6 LI and the MI. The concordance between image analysis and manual LI was evaluated with the intraclass correlation coefficient. The correlation between MCM6 and Ki-67 LI and WHO grade was evaluated with the Mann–Whitney Wilcoxon test. Overall survival (OS) analyses were performed with Kaplan–Meier estimation (log-rank test) and the Cox model using univariate and multivariate analyses. For Ki-67 and MCM6 LI, the optimal threshold was computed using the Cutoff Finder online tool (8), in order to separate patients in two groups. For Cox multivariate analyses, only variables which were significant on univariate analyses and WHO grade were integrated into the model. Survival analyses were performed in the POLA cohort of IDHmut+/1p19qcodel AO, then in the IDHmut+/1p19qcodel OII cohort. A *P*-value (*P*) of less than 0.05 was considered statistically significant.

Transcriptomic analyses

POLA cohort: Gene expression data from POLA (29) were downloaded from http://gliovis.bioinfo.cnio.es/ under the "Kamoun" and "POLA Network" tags. From these publicly available data, we identified 68 samples satisfying the following phenotypic criteria: histology = oligodendroglioma, grade = III (anaplastic), co-deletion 1p-19q status = yes, IDH1 or IDH2 status = mutant. These data were normalized (lowess algorithm) against a median profile of all 68 samples and further underwent descriptive statistics, which helped define groups of samples before proceeding with downstream analyses. Hence, all samples were individually labeled MCM6- and MKI67-up or -down based on their relative MCM6 and MKI67 expression (Supplementary Figure S1).

K-means clustering, functional annotations, enrichments computations, differential expression statistics and renderings were achieved using protocols and tools as described before (20). Further hierarchical clustering was applied for each significant k-means cluster to evaluate the level of correlation and quickly delineate gene or sample outliers in each gene signature.

Differential expression *P*-values were computed on normalized data using a two-way t-test between both MCM6 and MKI67 up and down groups. These were adjusted to allow for false discovery rate (FDR) using the Benjamini–Hochberg procedure and FDR < 0.01 was considered to indicate statistical significance.

Linear correlations between MCM6 and MKI67 expression levels were performed with Pearson's method, while Spearman's rank correlation coefficients were used to explore the dependence between both MCM6 and MKI67 gene expressions with their respective MCM6 and Ki-67 immunohistochemical LI.

TCGA cohort: Additionally, in order to evaluate the results obtained on the POLA cohort, further analyses were run on a subset of The Cancer Genome Atlas Lower Grade Glioma cohort (TCGA-LGG). From the available RNA-Seq processed data (FPKM-UQ normalized counts), 98 samples corresponded to unique cases with the following criteria:

IDH1 and/or IDH2 mutated, presence of the 1p-19q codeletion, and clearly clinically labeled as grade III and/or Anaplastic OD (35 samples) or grade II OD (63 samples). The original 60483 gene Ensembl ID were re-annotated (using https://biotools.fr) and their expression values were aggregated to obtain a set of unique official gene symbols. Data were then filtered to remove genes with very low expression values across samples (FPKM-UQ < 10 in more than 9/10th of the samples), which left 16345 genes. Values were re-normalized according to library size and log2 transformed. Two datasets were then created, one set containing the 35 grade III samples, with the aim to conduct validation analyses of the POLA high-grade OD transcriptomics, and another set with the entire selected TCGA cohort of 98 grade II + grade III samples for extending the results to lower grade OD. These two data sets finally underwent the same unsupervised workflow as described above (sample ordering according to *MCM6* mRNA expression levels, k-means clustering, functional annotations, enrichment analyses).

RESULTS

Clinicopathological data

The mean age of the patients of the POLA cohort was 49 years (range 19–80) (Table 1), with a male-to-female ratio of 1.22:1. A large majority of patients had surgery first, with 33% (72/220) of them showing macroscopic total resection, 31% (68/220) subtotal resection and 21% (46/220) partial resection. As adjuvant therapy, 32% (71/220) received RT-PCV therapy and 21% (46/220) combined radiotherapy with Temozolomide (Stupp Protocol). For 29% (64/220) of them, adjuvant radiotherapy was chosen. Fourteen patients did not receive any treatment (6%; 14/220). Follow-up data were available for 220 patients and were collected over a median follow-up of 40.9 (0.3–92) months. Disease-related death occurred in 9% (19/220) of patients. Using univariate analyses, age was the only variable significantly associated with survival (P = 0.001; HR: 1.055 [95%CI: 1.021–1.091]) (Table 2). A high MI (at least eight mitoses per 1.6 mm^2) was significantly associated with a shorter OS (P = 0.018; HR: 2.587 [95%CI: 1.177–5.686]). No significant correlation between OS and microvascular proliferation (P = 0.511; HR: 1.493 [95%CI: 0.451–4.939] or necrosis (P = 0.192; HR: 1.648 [95%CI: 0.778–3.492] was found.

In the grade II series, mean age was 44 years (range 18–81), with a male-to-female ratio of 1.21:1. Disease-related death occurred in 13% (4/30) of patients, with a median follow-up of 38.7 (2.4–79.9) months. When combining the two cohorts, WHO grade was not correlated to OS (P = 0.642).

MCM6 labeling index and correlation with survival

MCM6 staining was interpretable in 94% (206/220) of the cases with an average count of 1383 cells per case. Image analysis-based LI was strongly correlated to manual counting (ICC: 0.864; 95%CI [0.772–0.920]). Mean MCM6 LI was 24% (range 0.1–87%; median 21.4; standard deviation 18.8) (Figure 1A,B). MCM6 LI was significantly correlated to MI (ρ = 0.253; P < 0.0001).

In the POLA cohort, log-rank survival analyses showed that patients with high MCM6 expression (upper than 50%) had significantly shorter overall survival (OS) (P = 0.013) (Figure 2A). Using univariate Cox model, MCM6 expression correlated inversely with OS (P = 0.018; HR: 3.28; 95%CI [1.22–8.83]) (Table 2). In the group of patients with grade II oligodendroglioma, a high MCM6 LI was also correlated to a shorter OS (P = 0.001) (Figure 3A). MCM6 LI was not significantly higher in grade III than in grade II oligodendrogliomas (23% vs. 18%; P = 0.232).

Table 1. Clinico-pathological characteristics and molecular data.

Variable	Results
Age (mean; min-max)	49; 19–80 years
Sex	Male-to-female ratio: 1.22:1 (121/99)
Surgery	Biopsy: 4.1% (9/220)
	Total resection: 32.7% (72/220)
	Subtotal resection 30.9% (68/220)
	Partial resection 20.9% (46/220)
	Missing data 11.4% (25/220)
Type of treatment	RT-PCV: 32.3% (71/220)
	Radiotherapy: 29.1% (64/220)
	PCV: 2.3% (5/220)
	Stupp protocol: 20.9% (46/220)
	Temozolomide: 4.1% (9/220)
	Other: 0.5% (1/220)
	No treatment: 6.4% (14/220)
	Missing data: 3.6% (8/220)
Survival	Progression: 30.7% (71/220)
	Death: 8.2% (19/220)
Molecular data	*TERT* promoter mutation: 98.3% (216/220)
	CIC loss: 61% (141/220)

RT-PCV = radiation therapy – procarbazine, CCNU (lomustine), and vincristine.

Ki-67 labeling index and correlation with survival

For Ki-67 LI, 91% (199/220) of the cases were interpretable. We found a high reproducibility between image analysis and manual counting (ICC: 0.900; 95%CI [0.827–0.943]). Ki-67 LI evaluated by image analysis was also strongly correlated to the Ki-67 LI previously obtained from whole slides (ρ = 0.553; P < 0.0001). The mean number of cells counted for each TMA was 1121. The mean Ki-67 LI was 6.3% (range 0.1–36.9; median 3.7; standard deviation 6.7) (Figure 1C,D). Ki-67 LI was significantly correlated to MI (ρ = 0.275; P < 0.0001) and MCM6 LI (ρ = 0.449; P < 0.0001).

The log-rank test and the univariate regression Cox model revealed that a Ki-67 LI equal to or greater than 15% was

Brain Pathology **30** (2020) 465–478

Table 2. Univariate and multivariate Cox analyses for overall survival.

	Cox univariate (OS)		Cox multivariate (OS) *Model 1*		Cox multivariate (OS) *Model 2*	
Variable	HR [95%CI]	*P*-value	HR [95%CI]	*P*-value	HR [95%CI]	*P*-value
Age	1.055 [1.021–1.091]	0.001*	1.060 [1.020–1.103]	0.003*	1.051 [1.011–1.092]	**0.012***
Mitotic index ≥ 8/1.6 mm^2	2.587 [1.177–5.686]	0.018*	1.439 [0.538–3.851]	0.469	1.588 [0.606–4.162]	0.347
MCM6 LI ≥ 50%	3.283 [1.221–8.826]	0.018*	2.896 [0.964–8.702]	0.058		
Ki-67 LI ≥ 15%	3.948 [1.442–10.41]	0.008*	2.713 [0.935–7.875]	0.066		
MCM6 LI ≥ 50% and/or Ki-67 LI ≥ 15%	3.875 [1.603–9.370]	0.003*			2.872 [1.125–7.328]	**0.027***

HR = hazard ratio; LI = labeling index; OS = overall survival.
*Statistically significant ($P < 0.05$).

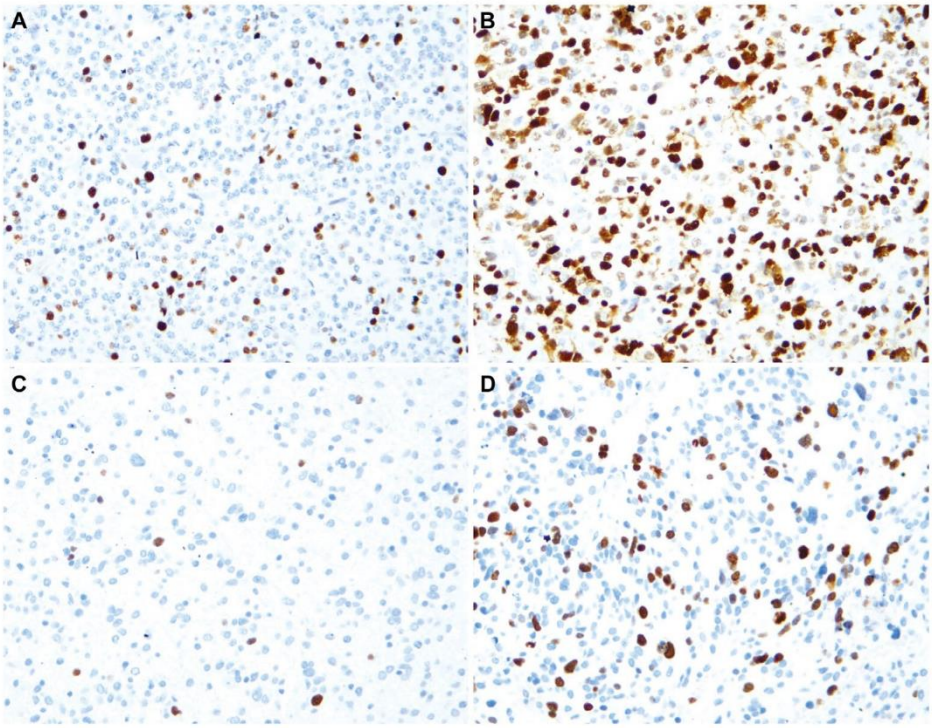


Figure 1. *Immunolabeling for MCM6 and Ki-67 (immunohistochemistry, x200).* **A**. Low MCM6 labeling index (LI). **B**. High MCM6 LI. **C**. Low Ki-67 LI. **D**. High Ki-67 LI (same case as B).

correlated with a shorter OS (log-rank: $P = 0.004$; Cox: $P = 0.008$; HR = 3.948; 95%CI [1.442–10.41]) (Figure 2B; Table 2). Similarly, in the group of patients with grade II oligodendroglioma, a high Ki-67 LI was correlated to shorter OS ($P = 0.027$; log-rank test) (Figure 3B). Ki-67 LI was significantly higher in grade III than in grade II tumors (6% vs. 3%; $P = 0.001$).

Multivariate survival analyses

The variables that were significant in univariate analyses (age, MI ≥ 8/1.6 m^2, MCM6 LI ≥ 50%, Ki-67 LI ≥ 15%) were included in the first multivariate Cox model. Only age remained significantly correlated to OS ($P = 0.003$), whereas MCM6 and Ki-67 LI were close to significance ($P = 0.058$ and $P = 0.066$, respectively) (Table 2).

In order to evaluate if MCM6 could be an interesting marker to use in complement to Ki-67, we designed a second Cox multivariate model, with the variable "MCM6 LI ≥ 50% and/or Ki-67 LI ≥ 15%", that was strongly significant in univariate analysis (log-rank: $P = 0.001$; Cox: $P = 0.003$; HR: 3.875; 95%CI [1.603–9.370]). With this model, age and "MCM6 LI ≥ 50% and/or Ki-67 LI ≥ 15%" were significantly correlated to OS ($P = 0.012$ and $P = 0.027$, respectively), but not MI ($P = 0.347$) (Figure 2C; Table 2).

Additionally, we evaluated the prognostic value of MCM6 and Ki-67 LI in all cases of IDHmut+/1p19qcodel oligodendrogliomas, including grade II and III tumors. In a

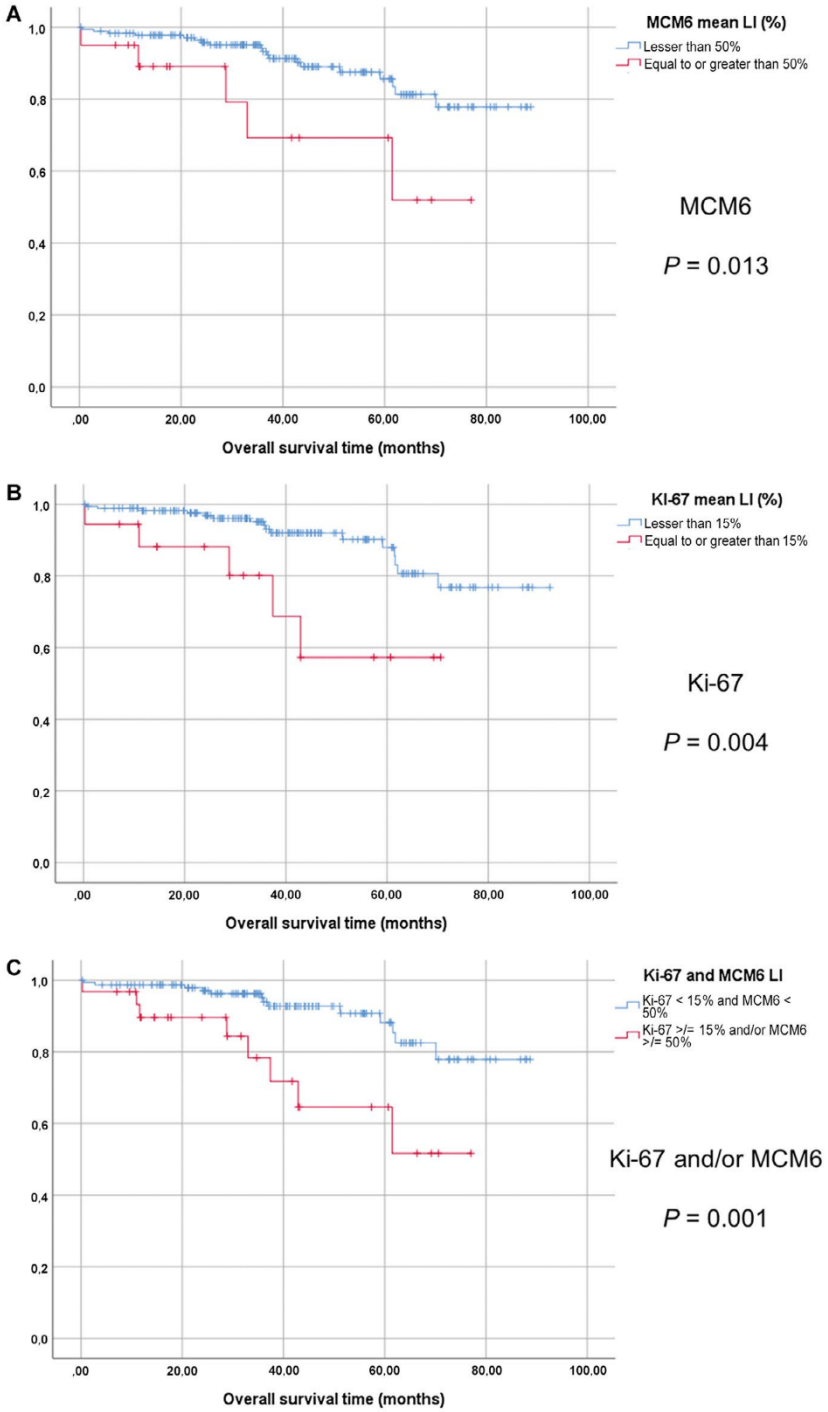


Figure 2. *Survival analyses in the POLA cohort of anaplastic oligodendroglimas. Kaplan-Meier curves with log-rank tests (overall survival).* **A**. MCM6 labeling index, 50% threshold (LI). **B**. Ki-67 LI, 15% threshold. **C**. MCM6 LI ≥ 50% and/or Ki-67 LI ≥ 15% vs. MCM6 < 50% and Ki-67 LI < 15%.

multivariate model including age, MI, WHO grade and the two proliferation markers, we found that age (P = 0.022) and MCM6 LI ≥ 50% and/or Ki-67 LI ≥ 15% (P = 0.001; HR: 4.148; 95%CI [1.735–9.920] were significantly associated with OS, but not WHO grade (P = 0.108) or MI (P = 0.637).

Brain Pathology **30** (2020) 465–478

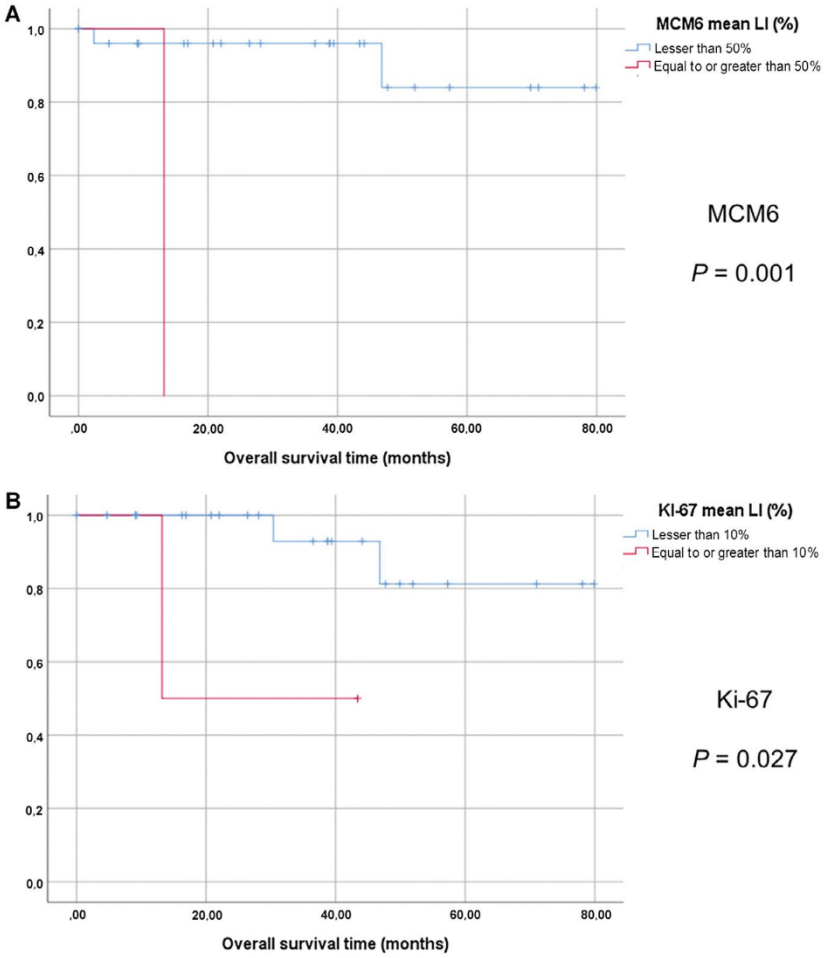


Figure 3. *Survival analyses the cohort of grade II oligodendrogliomas.* Kaplan-Meier curves with log-rank tests (overall survival). **A**. MCM6 labeling index, 50% threshold (LI). **B**. Ki-67 LI, 10% threshold.

Transcriptomics in the POLA cohort

For both *MCM6* and *MKI67*, we expected positive correlations between their LI and mRNA level in AO cases for which both transcriptomic data and LI were available. The Spearman tests indeed confirmed such correlations (ρ = 0.43, P = 0.003, n = 46; and ρ = 0.63, $P = 4.7 \times 10^{-5}$, n = 38, respectively) (Supplementary Figure S2). A highly positive correlation was also found between the mRNA level of *MCM6* and that of the *MKI67* in the AO transcriptomes of the POLA network (n = 68; Pearson's test: 0.69; $P = 5.6 \times 10^{-11}$) (Supplementary Figure S3).

Molecular events associated with differential expressions of either *MCM6* or *MKI67* in these AO samples were subsequently analyzed (n = 68). This was achieved using the median expression levels for *MCM6* and *MKI67* as "up" and "down" subgroup delimiters (Supplementary Figure S1) in two separate analyses based on the relative expressions of *MCM6* and *MKI67*. When the median *MCM6* mRNA level was used to organize the 68 AO samples, 34 were classified and labeled as *MCM6*-down (sample expression < median expression), and the remaining 34 as MCM6-up (expression > median). The same was applied for *MKI67*. In these two analyses, k-means clustering coupled with a distance computation based on the weighted correlation method, permitted the identification of clusters with distinct molecular functions differentiating the "up" from the "down" *MCM6* or *MKI67*-expresser AO (Figure 4). Remarkably, these *MCM6*- and *MKI67* gradient-driven clusterings resulted in individual clusters sharing high level of identity, as the clusters produced between the *MCM6* and *MKI67* series of experiments overlapped with each other (ranging from 85% to 99%, the most significant clusters presenting with the greatest identity) (Figure 4). Moreover, among the 10 clusters identified in each series (from c0 to c9), 2 clusters of upregulated genes (c3 and c7) and 3 clusters of downregulated genes (c1, c2 and c6) were highly conserved between the two series (98%, 99%, 97%, 97%, 96%, respectively; Figure 4). Further functional annotation of the up- and downregulated clusters revealed that tumors expressing higher levels of *MCM6* and *MKI67* were

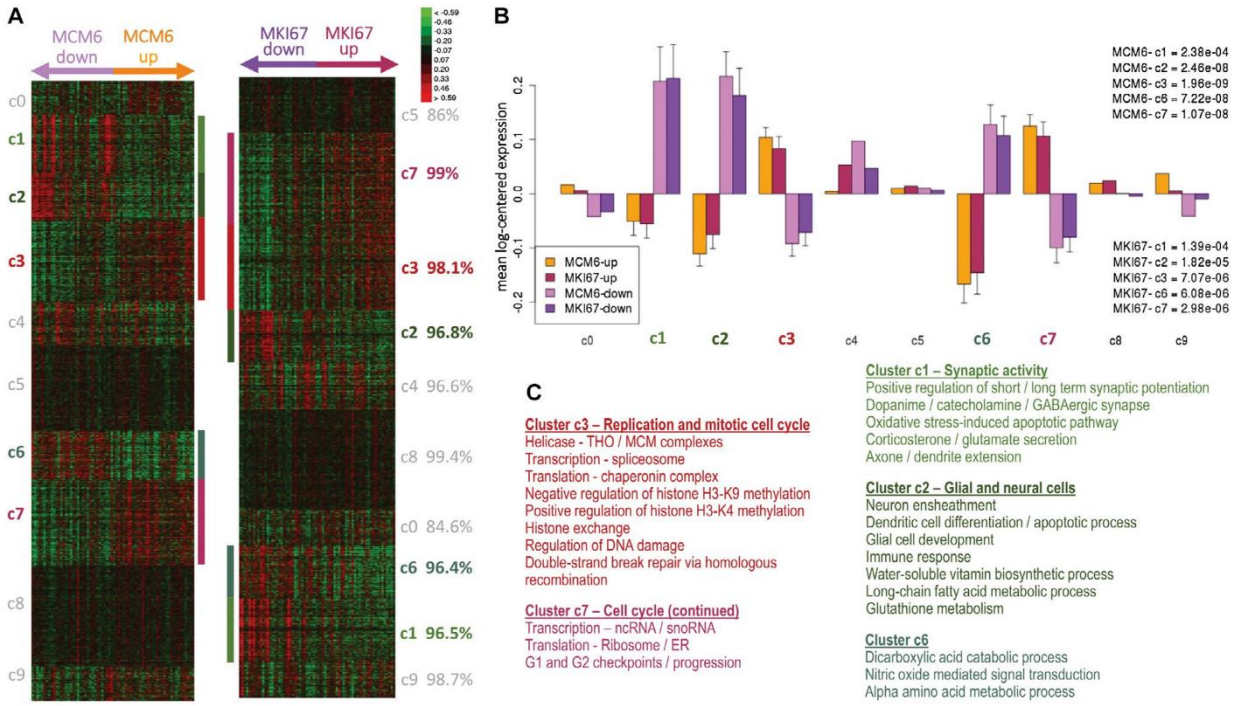


Figure 4. *K-means clustering from* MCM6 *and* MKI67 *expression gradients (from lowest to highest) of 68 transcriptomes of anaplastic oligodendrogliomas from the POLA network.* The two independent K-means clustering each delineated five significant groups of co-expressed genes ($P < 0.01$) showing highly correlated expression profiles. **A.** Heatmap of the ten identified k-means clusters in the two clustering results, and lowest identity (in %) found for each cluster (green, red and black indicate downregulated, upregulated, and median genes, respectively). **B.** Mean gene expression (overall with standard error at the mean) for each cluster in *MCM6* and *MKI67* analyses where both upregulated samples for *MCM6* or *MKI67* are compared to their downregulated counterparts (AU – arbitrary units). Upregulated and downregulated samples are defined relatively to the median gene expression for either *MCM6* or *MKI67*. **C.** Overview of the most relevant functional annotations for the significantly over- and under-expressed clusters (in red and green hues, respectively, all $p(FDR) \leq 0.01$).

also enriched with genes controlling replication and cell cycle (c3 and c7), but downregulated with genes associated with synaptic activity (c1), neuron ensheathment, glial cell development, immune and inflammatory response, vitamin biosynthetic process (c2), and alpha amino acid metabolic process (c6) (FDR < 0.0001) (Supplementary Table S1).

Additionally, we explored differential expression statistics (two-sided t-test) between *MCM6*-up and *MCM6*-down, and between *MKI67*-up and *MKI67*-down samples. Four significant gene lists were produced (all FDR < 0.01), one list for over- and one for under-expressed genes in *MCM6*-up samples (871 and 461 genes; fold changes > 4/3 and < 3/4, respectively), and likewise for *MKI67*-up samples (352 and 237 genes, respectively) (Supplementary Table S2). *MCM6* overexpressed samples presented with a significant enrichment in genes involved in cell cycle functions (Table 3), including most notably DNA replication, mitotic centrosome separation and mitotic chromosome condensation and genes involved in internal ribosome entry site (IRES) dependent translational initiation. Tumors harboring lower *MCM6* expression showed significant upregulation of pathways involved in glial differentiation and immune response, such as microglial cell activation, myelination, oligodendrocyte development and oligodendrocyte differentiation. Similarly, tumors with higher expression of *MKI67* (Table 4) were enriched in genes involved with cell cycle functions such as DNA strand elongation in DNA replication during cell cycle, pre-replicative complex assembly, mitotic chromosome condensation and MCM complex. Tumors with lower *MKI67* expression were significantly enriched in genes involved in beta-amyloid binding, synaptic vesicle, axon terminus, trans-synaptic signaling, neuronal cell body and cell projection.

When focusing on genes associated with a pro-neural signature, such as *SOX2, SOX4, SOX11, OLIG1, OLIG2, INSM1, FERMT1, DCX, CTTNBP2, ATOH8* and *ASCL1*, we found a significant overexpression of these genes in tumors showing higher levels of *MCM6* and *MKI67* (FDR < 0.01) (Supplementary Figure S4A). Surprisingly, a number of genes involved in immunological responses were significantly downregulated in these tumors (*CX3CR1, TLR4, SYK, FAS, HSPA2, CEBPB, TRIM59, ITGB5, PLP1, C3AR1, PREX1, MNDA, GAB2, VAMP3*; FDR < 0.01) (Supplementary Figure S4B). Genes controlling the production of molecular mediators involved in inflammatory response were more significantly

Table 3. Top 12 Gene Ontology functional annotations associated with the expression of MCM6 (sorted by fold enrichment).

GO biological process	Fold enrichment	FDR
Enrichment in tumors showing MCM6 higher expression		
Cell cycle DNA replication initiation (GO:1902292)	22.1	<0.0001
DNA strand elongation involved in cell cycle DNA replication (GO:1902296)	21.7	0.0001
Ndc80 complex (GO:31262)	21.7	0.0001
Pre-replicative complex assembly (GO:36388)	21.3	<0.0001
Condensed nuclear chromosome kinetochore (GO:778)	15.2	<0.0001
Mitotic centrosome separation (GO:7100)	15.2	0.001
Mitotic chromosome condensation (GO:7076)	14.8	<0.0001
IRES-dependent translational initiation (GO:2192)	14.6	0.0001
Aster (GO:5818)	13.8	0.0001
Exodeoxyribonuclease activity (GO:4529)	12.6	0.003
Mitotic prophase (GO:88)	12.1	<0.0001
Regulation of transcription involved in G1/S transition (GO:83)	11.9	<0.0001
Enrichment in tumors showing MCM6 lower expression		
Myelin sheath adaxonal region (GO:35749)	11.2	0.009
Negative regulation of inclusion body assembly (GO:90084)	8.7	0.02
Sodium ion export from cell (GO:36376)	8.7	0.02
Positive regulation of organic acid transport (GO:32892)	8.2	0.02
Microglial cell activation (GO:1774)	8.4	<0.0001
Macrophage activation involved in immune response (GO:2281)	8.2	0.009
Central nervous system myelination (GO:22010)	7.6	0.002
Positive regulation of myelination (GO:31643)	6.5	0.02
Main axon (GO:44304)	4.6	<0.0001
Proteoglycan binding (GO:43394)	4.5	0.01
Oligodendrocyte development (GO:14003)	4.2	0.005
Positive regulation of glial cell differentiation (GO:45687)	4.0	0.02

FDR = false discovery rate; GO = gene ontology.

downregulated with *MCM6*-up than *MKI67*-up (*EPHX2, APOD, VAMP3, SYK, ALOX5, DUSP10, SNX6, ADORA3, ALOX5AP, TNF, LYN, BTK, FCER1G, IL17RA* and the pro-inflammatory mediator *IL17D*; FDR < 0.01) (Supplementary Table S2).

Other gene signatures were also found to overexpress with *MCM6* and *MKI67* high transcription, such as epigenetic markers (*DNMT1A, EZH2, TYMS, DHFR*; FDR < 0.01) and cell cycle progression key protagonists (all *MCMs, PCNA, CCNB1, CCND1, CDK1, CDK4*; FDR < 0.01), whereas myelination and glial cell differentiation genes were under-expressed with *MCM6* (FDR < 0.01) but less clearly with *MKI67* high levels, which is in accordance with the functional annotations reported above (Supplementary Figures S4B,C,E,F). Strongly upregulated KEGG pathways were all associated with an *MCM6* or *MKI67* overexpression (ribosome, RNA transport, splicing and degradation, DNA replication and repair, P53 signaling; q-value < 0.05). Both *MCM6*-up and *MKI67*-up sets of significantly disturbed pathways were nearly identical, with cell cycle controlling genes showing the same overexpression patterns (Supplementary Figure S5).

Confirmation analyses based on TCGA data

To validate the observations made on the POLA cohort, we ran a validation analysis with 98 similar cases identified in TCGA (The Cancer Genome Atlas) database. These were all identified as IDH1/IDH2 codeleted OD cases and included 35 fully annotated grade III and 63 grade II samples. Our survival analyses of the 98 TCGA cases showed that similar to that observed in our own grade II/III cohort, high levels of *MCM6* mRNA (equal to or greater than the median expression; arbitrary unit) indeed correlated to a shorter OS (P = 0.027) (Supplementary Figure S6A). However, the mRNA level of *MKI67* in the TCGA samples did not correlate with OS (P = 0.398) (Supplementary Figure S6B). WHO grade did not correlate to survival (P = 0.617) (Supplementary Figure S6C). Only age was independently correlated to survival (Cox multivariate analysis, P = 0.007). *MCM6* and *MKI67* mRNA levels were strongly correlated (ρ = 0.623; P < 0.0001). No significant correlation existed between the WHO grade and the mRNA level of either *MCM6* or *MKI67* (P > 0.50).

We also compared the transcriptomic signatures obtained with anaplastic POLA samples with the 35 grade III independent cases retrieved from TCGA (Supplementary Figure S7). We did so by conducting the analyses with the same unsupervised protocols on the TCGA data set as were applied to the POLA study. As sample grouping according to *MCM6* and *MKI67* median expression levels led to the same configuration, results for both comparisons were fused as one. Similar molecular signatures were indeed uncovered in TCGA samples, each signature presenting with an enrichment with its POLA counterpart ranging from 3.3 to 7.6 times higher than expected (all $P < 1 \times 10^{-100}$, Fisher's exact test). These showed a downregulation of inflammation, glial differentiation,

Table 4. Top 12 Gene Ontology functional annotations associated with the expression of MKI67 (sorted by fold enrichment).

GO biological process	Fold enrichment	FDR
Enrichment in tumors showing MKI67 higher expression		
Condensin complex (GO:796)	50.2	<0.0001
Replication fork protection complex (GO:31298)	42.1	<0.0001
DNA strand elongation involved in cell cycle DNA replication (GO:1902296)	40.1	<0.0001
Ndc80 complex (GO:31262)	40.1	<0.0001
Pre-replicative complex assembly (GO:36388)	35.1	<0.0001
Leading strand elongation (GO:6272)	35.1	<0.0001
Condensed nuclear chromosome kinetochore (GO:778)	31.6	<0.0001
MCM complex (GO:42555)	28.1	0.0001
Mitotic centrosome separation (GO:7100)	28.1	0.0001
Mitotic chromosome condensation (GO:7076)	27.5	<0.0001
Kinetochore microtubule (GO:5828)	25.5	0.0001
Aster (GO:5818)	25.5	0.0001
Enrichment in tumors showing MKI67 lower expression		
Regulation of inclusion body assembly (GO:90083)	12.3	0.03
Beta-amyloid binding (GO:1540)	10.5	0.005
Neurotrophin TRK receptor signaling pathway (GO:48011)	7.9	0.04
Import into cell (GO:98657)	6.6	0.01
Sodium ion transmembrane transport (GO:35725)	4.7	0.03
Synaptic vesicle (GO:8021)	4.1	0.007
Postsynaptic specialization (GO:99572)	3.7	0.001
Axon part (GO:33267)	3.7	<0.0001
Excitatory synapse (GO:60076)	3.7	0.001
Myelin sheath (GO:43209)	3.4	0.01
Dendrite (GO:30425)	2.6	0.001
Trans-synaptic signaling (GO:99537)	2.5	0.001

FDR = false discovery rate; GO = gene ontology.

myelin sheath and synaptic activity in tumors that overexpressed *MCM6* and/or *MKI67*, and an upregulation of mitotic cell cycle and DNA replication (all FDR < 0.01).

We further asked if these results extended to grade II cases and applied the same experimental protocol on TCGA combined grades II/III cohort (n = 98; Supplementary Figure S8). Clusters of genes involved in mitotic cell cycle/replication, myelin sheath/glial cell differentiation and axoneme/oxacid metabolic process were indeed strongly differential between *MCM6*-down and *MCM6*-up groups (all FDR < 0.01), and presented with high identities with their POLA counterparts (73.6%, 60.6% and 48.2%, respectively). However, we did not find well-defined signatures such as those reported for grade III alone of both POLA and TCGA cohorts for the other clusters. The immune response/inflammation signature, in particular, was lost when including grade II samples to the mixture.

DISCUSSION

The aim of our study was to evaluate the prognostic value of MCM6 and Ki-67 LI in AOs. Previously, other members of MCM family, such as MCM2, MCM3 and MCM7, have indeed been studied in gliomas, and their overexpression was associated with a poor overall prognosis; however, their applicability in clinical practice seems to be unclear (14, 16, 27). The prognostic value of MCM6 has been confirmed in various types of brain tumors like meningiomas, adamantinomatous craniopharyngiomas and gliomas (9, 21, 48). In the present study, we specifically focused on AOs with the presence of defined mutations of IDH1/2 mutations and 1p/19q codeletion to avoid confusing factors due to molecular alterations. Indeed, *IDH1/2* mutations and 1p/19q codeletion are known to correlate with better outcome, due notably to a better response to adjuvant chemotherapy. Only few prognostic markers have been reported in the specific group of IDHmut+/1p19qcodel AO. In this work, PFS was not assessed, because a consensual definition of progression in glioma is still lacking (1, 7, 28).

In this study, we found a good correlation between the mean LIs of MCM6 and Ki-67. MCM6 LI was always higher than Ki-67 LI in a given tumor (Figure 1). This is consistent with data of the literature, and is due to the fact that MCMs are expressed earlier in the cell cycle (15, 30). Furthermore, we have identified a cutoff point at 50% MCM6 LI as the optimal criteria for survival prediction: at greater than a LI of 50%, MCM6 overexpression correlated with a shorter overall survival. Based on the results we obtained earlier in meningioma, in which high MCM6 LI were found correlated with survival (21), this marker may even be applicable to other brain tumor types. Consequently, this high labeling index of MCM6 at 50% threshold can be considered as an easy-to-use tool for routing practice of the pathology laboratories, in complement to Ki-67.

We also found that high mean LIs of Ki-67 were associated with poor outcome, with a threshold of 15%. This

result is consistent with previous studies using the EORTC (European Organization for Research and Treatment of Cancer) Brain Tumor Group (40).

Given the results reported above, it shall be noted that although Ki-67 LI had good inter-observer agreement in previous studies (11, 39), there is no consensual method for counting Ki-67 LI in AO. Usually, most of the neuropathologists select areas showing the highest proliferation. Here, in order to have a more reproducible counting method for both Ki-67 and MCM6, we adapted computerized color image analyses based on previous results for Ki-67 cell scoring (6, 36). Consequently, a wide area was analyzed for each TMA spot. However, as we were not sure whether the selected spots corresponded to areas with the highest cell proliferation (despite the fact that individual spots were made with histologically representative samples of the tumors), we determined in our series the correlation between the Ki-67 LI obtained from TMA and the one obtained from whole slides; we found that these values correlated significantly ($\rho = 0.553$; $P < 0.0001$; data not shown), confirming the adequacy of our prepared TMA spots.

Multivariate analyses showed that MCM6 may be an interesting marker to use in association to Ki-67. Indeed, tumors harboring MCM6 LI ≥ 50% and/or Ki-67 LI ≥ 15% were significantly correlated to shorter survival, both in univariate and multivariate analyses. High MCM6 and Ki-67 LI were also correlated to shorter survival in grade II tumors, while in our study WHO grade was not significantly correlated to survival. However, these results are limited by a relatively short median follow-up time and a low number of OS events, thus the observed survival differences were based on few patients. Another limitation is that we used a statistical method that specifically seeks for the best cutoff in a given data set, which may not be applicable in another cohort. Further studies are needed in order to confirm the validity of these thresholds.

Analyses of the transcriptome data of the POLA grade III cohort showed that *MCM6* and/or *MKI67* overexpressing tumors share near identical whole transcriptional profiles, in spite of a correlation between *MCM6* and *MKI67* mRNA levels not being perfect. These tumors upregulated genes linked to DNA replication during the cell cycle (including all *MCM*s), DNA strand elongation, pre-replicative complex assembly, mitotic centrosome condensation and separation or regulation of helicase activity. This is consistent with the highly proliferative activity of these aggressive cancers. In addition, the *MCM6*-up and *MKI67*-up tumors of the POLA AO cohort overexpressed pro-neuronal differentiation genes such as *SOX4*, *SOX11*, *INSM1* and *ASCL1*. These findings are consistent with the pro-neural signature of the AO previously highlighted by Bielle *et al* These authors identified a subgroup of AO overexpressing neuronal intermediate progenitor genes, associated with immunohistochemical similarities to embryonic subventricular zone, expression of *INSM1* and no expression of *SOX9* (5).

On the contrary, in AO cases showing an overall lower expression of *MCM6* and/or *MKI67*, transcriptomic analyses revealed enrichments of genes involved in myelin sheath production, myelinization or oligodendroglial differentiation. These results suggest that both *MCM6*-down and *MKI67*-down tumors lack a pro-neural signature, and unlike their *MCM6*-*MKI67*-up counterparts, are well-differentiated oligodendroglial-like tumors.

In the *MCM6*-down subgroup, we also found genes like *CX3CR1*, *TLR4* and *SYK*, involved in immune response and microglial cells activation. Such downregulations in immune responsive genes have been observed in other types of glioma. For example, in murine models, loss of *CX3CR1* expression promoted gliomagenesis in glioblastoma with pro-neural signature through larger inflammatory monocytes accumulation (17). Lower levels of *TLR4*, a member of the toll-like receptors involved in macrophagic immune response, have been implicated in the immune evasion of glioblastoma stem cells (3). The immune evasion in high-grade gliomas is well documented. The non-neoplastic cells of the glioma microenvironment, such as fibroblasts, endothelial cells or microglial cells, produce survival and growth factors, helping tumors cells to grow and infiltrate the brain parenchyma (23). The stimulatory role of spleen tyrosine kinase receptors (SYK) in gliomagenesis, brain invasion, and recruitment of immune cells has recently been studied in glioblastoma stem cells by Moncayo *et al* using SYK-inhibitors, but their involvement has never been specifically studied in AO (37). In another study, SYK appeared as a tumor suppressor gene in breast carcinomas (13). First-generation SYK-inhibitors have been successfully tested in some hematologic malignancies, as B-acute lymphoid leukemia and chronic lymphocytic lymphoma (25, 44). The assessment of their efficiency in high-grade gliomas needs further investigations.

Finally, the predictive prognostic power of *MCM6* and/or *MKI67* was tested using a TCGA cohort of IDH1-IDH2-codel glioma composing of both grade II (n = 68) and III (n = 35) tumors. We found that high *MCM6* expression indeed correlated with OS, but not *MKI67*.

CONCLUSION

In this study, we assessed the prognostic value of MCM6 and Ki-67 LI in *IDH*-mutant and 1p/19q codeleted AOs of the POLA cohort. Our multivariate analyses showed that the overexpression of MCM6 and/or Ki-67 was independently correlated to shorter survival. The prognostic value of MCM6 was confirmed in TCGA grade II–III *IDH*-mutant and 1p/19q codeleted gliomas. These two easy-to-use and cost-effective markers could thus be used concurrently in routine pathology practice. Potentially, these markers could also be integrated into therapeutic as well as clinicoradiological monitoring strategies to improve disease outcomes. We also would like to underline the important new insight gained from the transcriptomic analyses of the AO, which is that AO with high proliferation have downregulated immune response and lower microglial cells activation.

ACKNOWLEDGMENTS

The authors wish to thank all the team of the CHRU Nancy Pathology Department for technical support, the members of

the POLA network for contributing case material and Marion Divoux for her help in manuscript preparation. The results shown here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

REFERENCES

1. Abdulla S, Saada J, Johnson G, Jefferies S, Ajithkumar T (2015) Tumour progression or pseudoprogression? A review of post-treatment radiological appearances of glioblastoma. *Clin Radiol* **70**:1299–1312.
2. Alentorn A, Dehais C, Ducray F, Carpentier C, Mokhtari K, Figarella-Branger D *et al* (2015) Allelic loss of 9p21.3 is a prognostic factor in 1p/19q codeleted anaplastic gliomas. *Neurology* **85**:1325–1331.
3. Alvarado AG, Thiagarajan PS, Mulkearns-Hubert EE, Silver DJ, Hale JS, Alban TJ *et al* (2017) Glioblastoma cancer stem cells evade innate immune suppression of self-renewal through reduced TLR4 expression. *Cell Stem Cell* **20**:450–461.
4. Bettegowda C, Agrawal N, Jiao Y, Sausen M, Wood LD, Hruban RH *et al* (2011) Mutations in CIC and FUBP1 contribute to human oligodendroglioma. *Science* **333**:1453–1455.
5. Bielle F, Ducray F, Mokhtari K, Dehais C, Adle-Biassette H, Carpentier C *et al* (2017) Tumor cells with neuronal intermediate progenitor features define a subgroup of 1p/19q co-deleted anaplastic gliomas. *Brain Pathol* **27**:567–579.
6. Blaker YN, Brodtkorb M, Maddison J, Hveem TS, Nesheim JA, Mohn HM *et al* (2015) Computerized image analysis of the Ki-67 proliferation index in mantle cell lymphoma. *Histopathology* **67**:62–69.
7. Brandsma D, Stalpers L, Taal W, Sminia P, Van den Bent MJ (2008) Clinical features, mechanisms, and management of pseudoprogression in malignant gliomas. *Lancet Oncol* **9**:453–461.
8. Budczies J, Klauschen F, Sinn BV, Győrffy B, Schmitt WD, Darb-Esfahani S, Denkert C (2012) Cutoff finder: A comprehensive and straightforward web application enabling rapid biomarker cutoff optimization. *PLoS ONE* **7**:e51862.
9. Cai HQ, Chen ZJ, Zhang HP, Wang PF, Zhang Y, Hao JJ *et al* (2018) Overexpression of MCM6 predicts poor survival in patients with glioma. *Hum Pathol* **78**:182–187.
10. Cancer Genome Atlas Research Network, Brat DJ, Verhaak RGW, Aldape KD, Yung WKA, Salama SR *et al* (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med* **372**:2481–2498.
11. Coleman KE, Brat DJ, Cotsonis GA, Lawson D, Cohen C (2006) Proliferation (MIB-1 expression) in oligodendrogliomas: assessment of quantitative methods and prognostic significance. *Appl Immunohistochem Mol Morphol* **14**:109–114.
12. Coons SW, Johnson PC, Pearl DK (1997) The prognostic significance of Ki-67 labeling indices for oligodendrogliomas. *Neurosurgery* **41**:878–884; discussion 884–885.
13. Coopman PJ, Do MT, Barth M, Bowden ET, Hayes AJ, Basyuk E *et al* (2000) The Syk tyrosine kinase suppresses malignant growth of human breast cancer cells. *Nature* **406**:742–747.
14. Erkan EP, Ströbel T, Lewandrowski G, Tannous B, Madlener S, Czech T *et al* (2014) Depletion of minichromosome maintenance protein 7 inhibits glioblastoma multiforme tumor growth in vivo. *Oncogene* **33**:4778–4785.
15. Eward KL, Obermann EC, Shreeram S, Loddo M, Fanshawe T, Williams C *et al* (2004) DNA replication licensing in somatic and germ cells. *J Cell Sci* **117**:5875–5886.
16. Facoetti A, Ranza E, Benericetti E, Ceroni M, Tedeschi F, Nano R (2006) Minichromosome maintenance protein 7: a reliable tool for glioblastoma proliferation index. *Anticancer Res* **26**:1071–1075.
17. Feng X, Szulzewsky F, Yerevanian A, Chen Z, Heinzmann D, Rasmussen RD *et al* (2015) Loss of CX3CR1 increases accumulation of inflammatory monocytes and promotes gliomagenesis. *Oncotarget* **6**:15077–15094.
18. Figarella-Branger D, Mokhtari K, Dehais C, Carpentier C, Colin C, Jouvet A *et al* (2016) Mitotic index, microvascular proliferation, and necrosis define 3 pathological subgroups of prognostic relevance among 1p/19q co-deleted anaplastic oligodendrogliomas. *Neuro Oncol* **18**:888–890.
19. Figarella-Branger D, Mokhtari K, Dehais C, Jouvet A, Uro-Coste E, Colin C *et al* (2014) Mitotic index, microvascular proliferation, and necrosis define 3 groups of 1p/19q codeleted anaplastic oligodendrogliomas associated with different genomic alterations. *Neuro Oncol* **16**:1244–1254.
20. Gauchotte G, Hergalant S, Vigouroux C, Casse JM, Houlgatte R, Kaoma T *et al* (2017) Cytoplasmic overexpression of RNA-binding protein HuR is a marker of poor prognosis in meningioma, and HuR knockdown decreases meningioma cell growth and resistance to hypoxia. *J Pathol* **242**:421–434.
21. Gauchotte G, Vigouroux C, Rech F, Battaglia-Hsu SF, Soudant M, Pinelli C *et al* (2012) Expression of minichromosome maintenance MCM6 protein in meningiomas is strongly correlated with histologic grade and clinical outcome. *Am J Surg Pathol* **36**:283–291.
22. Gleize V, Alentorn A, Connen de Kérillis L, Labussière M, Nadaradjane AA, Mundwiller E *et al* (2015) CIC inactivating mutations identify aggressive subset of 1p19q codeleted gliomas. *Ann Neurol* **78**:355–374.
23. Hambardzumyan D, Gutmann DH, Kettenmann H (2016) The role of microglia and macrophages in glioma maintenance and progression. *Nat Neurosci* **19**:20–27.
24. Helfenstein A, Frahm SO, Krams M, Drescher W, Parwaresch R, Hassenpflug J (2004) Minichromosome maintenance protein (MCM6) in low-grade chondrosarcoma: distinction from enchondroma and identification of progressive tumors. *Am J Clin Pathol* **122**:912–918.
25. Hoellenriegel J, Coffey GP, Sinha U, Pandey A, Sivina M, Ferrajoli A *et al* (2012) Selective, novel Spleen tyrosine kinase (Syk) inhibitors suppress chronic lymphocytic leukemia B cell activation and migration. *Leukemia* **26**:1576–1583.
26. Hotton J, Agopiantz M, Leroux A, Charra-Brunaud C, Marie B, Busby-Venner H *et al* (2018) Minichromosome maintenance complex component 6 (MCM6) expression correlates with histological grade and survival in

endometrioid endometrial adenocarcinoma. *Virchows Arch* **472**:623–633.
27. Hua C, Zhao G, Li Y, Bie L (2014) Minichromosome Maintenance (MCM) Family as potential diagnostic and prognostic tumor markers for human gliomas. *BMC Cancer* **14**:526.
28. Jain R, Narang J, Sundgren PM, Hearshen D, Saksena S, Rock JP *et al* (2010) Treatment induced necrosis versus recurrent/progressing brain tumor: going beyond the boundaries of conventional morphologic imaging. *J Neurooncol* **100**:17–29.
29. Kamoun A, Idbaih A, Dehais C, Elarouci N, Carpentier C, Letouzé E *et al* (2016) Integrated multi-omics analysis of oligodendroglial tumours identifies three subgroups of 1p/19q co-deleted gliomas. *Nat Commun* **7**.
30. Kingsbury SR, Loddo M, Fanshawe T, Obermann EC, Prevost AT, Stoeber K, Williams GH (2005) Repression of DNA replication licensing in quiescence is independent of geminin and may define the cell cycle state of progenitor cells. *Exp Cell Res* **309**:56–67.
31. Labib K, Kearsey SE, Diffley JF (2001) MCM2-7 proteins are essential components of prereplicative complexes that accumulate cooperatively in the nucleus during G1-phase and are required to establish, but not maintain, the S-phase checkpoint. *Mol Biol Cell* **12**:3658–3667.
32. Labreche K, Simeonova I, Kamoun A, Gleize V, Chubb D, Letouzé E *et al* (2015) TCF12 is mutated in anaplastic oligodendroglioma. *Nat Commun* **6**.
33. Lindner K, Gregán J, Montgomery S, Kearsey SE (2002) Essential role of MCM proteins in premeiotic DNA replication. *Mol Biol Cell* **13**:435–444.
34. Liu Z, Li J, Chen J, Shan Q, Dai H, Xie H *et al* (2018) MCM family in HCC: MCM6 indicates adverse tumor features and poor outcomes and promotes S/G2 cell cycle progression. *BMC Cancer* **18**:200.
35. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK *et al* (2016) The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* **131**:803–820.
36. Markiewicz T, Grala B, Kozlowski W, Osowski S (2010) Computer system for cell counting in selected brain tumors at Ki-67 immunohistochemical staining. *Anal Quant Cytol Histol* **32**:323–332.
37. Moncayo G, Grzmil M, Smirnova T, Zmarz P, Huber RM, Hynx D *et al* (2018) SYK inhibition blocks proliferation and migration of glioma cells and modifies the tumor microenvironment. *Neuro-Oncology* **20**:621–631.
38. Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, Barnholtz-Sloan JS (2017) CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro-Oncology* **19**:v1–v88.
39. Prayson RA, Castilla EA, Hembury TA, Liu W, Noga CM, Prok AL (2003) Interobserver variability in determining MIB-1 labeling indices in oligodendrogliomas. *Ann Diagn Pathol* **7**:9–13.
40. Preusser M, Hoeftberger R, Woehrer A, Gelpi E, Kouwenhoven M, Kros JM *et al* (2012) Prognostic value of Ki67 index in anaplastic oligodendroglial tumours–a translational study of the European Organization for Research and Treatment of Cancer Brain Tumor Group. *Histopathology* **60**:885–894.
41. Rosenberg S, Ducray F, Alentorn A, Dehais C, Elarouci N, Kamoun A *et al* (2018) Machine learning for better prognostic stratification and driver gene identification using somatic copy number variations in anaplastic oligodendroglioma. *Oncologist* **23**(12):1500–1510.
42. Schrader C, Janssen D, Klapper W, Siebmann JU, Meusers P, Brittinger G *et al* (2005) Minichromosome maintenance protein 6, a proliferation marker superior to Ki-67 and independent predictor of survival in patients with mantle cell lymphoma. *Br J Cancer* **93**:939–945.
43. Suzuki H, Aoki K, Chiba K, Sato Y, Shiozawa Y, Shiraishi Y *et al* (2015) Mutational landscape and clonal architecture in grade II and III gliomas. *Nat Genet* **47**:458–468.
44. Uckun FM, Qazi S (2014) SYK as a new therapeutic target in B-cell precursor acute lymphoblastic leukemia. *J Cancer Ther* **5**:124–131.
45. Van den Bent MJ, Brandes AA, Taphoorn MJB, Kros JM, Kouwenhoven MCM, Delattre JY *et al* (2013) Adjuvant procarbazine, lomustine, and vincristine chemotherapy in newly diagnosed anaplastic oligodendroglioma: long-term follow-up of EORTC brain tumor group study 26951. *J Clin Oncol* **31**:344–350.
46. Van den Bent MJ, Carpentier AF, Brandes AA, Sanson M, Taphoorn MJB, Bernsen HJJA *et al* (2006) Adjuvant procarbazine, lomustine, and vincristine improves progression-free survival but not overall survival in newly diagnosed anaplastic oligodendrogliomas and oligoastrocytomas: a randomized European Organisation for Research and Treatment of Cancer phase III trial. *J Clin Oncol* **24**:2715–2722.
47. Vigouroux C, Casse JM, Battaglia-Hsu SF, Brochin L, Luc A, Paris C *et al* (2015) Methyl(R217)HuR and MCM6 are inversely correlated and are prognostic markers in non small cell lung carcinoma. *Lung Cancer* **89**:189–196.
48. Xu J, Zhang S, You C, Huang S, Cai B, Wang X (2007) Expression of human MCM6 and DNA Topo II alpha in craniopharyngiomas and its correlation with recurrence of the tumor. *J Neurooncol* **83**:183–189.
49. Zeng A, Hu Q, Liu Y, Wang Z, Cui X, Li R *et al* (2015) IDH1/2 mutation status combined with Ki-67 labeling index defines distinct prognostic groups in glioma. *Oncotarget* **6**:30232–30238.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site:

Figure S1. *MCM6* and *MKI67* mRNA level distributions into two subgroups of higher ("-up") and lower ("-down") expressions. Transcriptomic samples (n = 68) are separated at the median expression for each gene and ordered according to their expression gradient for either *MCM6* or *MKI67*.

Figure S2. Correlations between immunohistochemistry (IHC) and mRNA level for Ki-67 (*MKI67*) and MCM6.

Figure S3. Correlation between mRNA expression levels of *MKI67* and *MCM6*.

Figure S4. Log2-normalized gene expression levels for the upregulated pro-neural **(A)**, epigenetic **(B)**, and cell cycle **(C)** signatures, as well as the downregulated immune response **(D)**, myelination **(E)** and glial cells differentiation **(F)** signatures for both *MCM6* and *MKI67* "-down" and "-up" samples (transcriptomics).

Figure S5. Significantly dysregulated KEGG pathways with an overall upregulation (Table as obtained with Gage) and map

of the cell cycle (GraphViz rendered with Pathview) in both *MCM6* **(A)** and *MKI67* **(B)** upregulated samples. Significantly (FDR < 0.01) down- and upregulated genes are colored from green to red, with respect to their fold change of expression (log2).

Figure S6. Survival analyses in grades II/III IDH-mutant/1p19q codeleted oligodendrogliomas selected from TCGA database. Kaplan–Meier curves with log-rank tests (overall survival). **A.** *MCM6* mRNA level, median threshold (arbitrary units). **B.** *MKI67* mRNA level, median threshold (arbitrary units). **C.** WHO 2016 grade.

Figure S7. K-means clustering based on *MCM6* and/or *MKI67* mRNA expression gradient (from lowest to highest) of the 35 transcriptomes of anaplastic IDH-mutant/1p19q codeleted oligodendrogliomas selected from TCGA bank. The K-means clustering delineated five significant clusters of co-expressed genes ($P < 0.05$) with highly correlated molecular functions. **A.** Heatmap of the 10 k-means clusters organized based on the extent of *MCM6* and *MKI67* expression (green, red and black indicate downregulated, upregulated and median genes, respectively). **B.** Overview of the most relevant functional annotations for the significantly over- and under-expressed clusters (in red and green hues, respectively, all $P(FDR) \leq 0.01$). Each of these clusters was compared to the clusters found of the POLA cohort, and their identities calculated based on the corresponding POLA clusters. **C.** Mean gene expression (overall with standard error at the mean) for each cluster where both upregulated samples for *MCM6* and/or *MKI67* are compared to their downregulated counterparts (AU—arbitrary units). Upregulated and downregulated samples are defined relatively to *MCM6* median gene expression.

Figure S8. K-means clustering based on *MCM6* and/or *MKI67* mRNA expression gradient (from lowest to highest) of the 98 transcriptomes of grades II and III IDH-mutant/1p19q codeleted oligodendrogliomas selected from TCGA bank. The K-means clustering delineated five significant clusters of co-expressed genes ($P < 0.05$) with highly correlated molecular functions. **A.** Heatmap of the ten k-means clusters organized based on the extent of *MCM6* and *MKI67* expression (green, red and black indicate downregulated, upregulated and median genes, respectively). **B.** Overview of the most relevant functional annotations for the significantly over- and under-expressed clusters (in red and green hues, respectively, all $P(FDR) \leq 0.01$). Each of these clusters was compared to the clusters found of the POLA cohort, and their identities calculated based on the corresponding POLA clusters. **C.** Mean gene expression (overall with standard error at the mean) for each cluster where both upregulated samples for *MCM6* and/or *MKI67* are compared to their downregulated counterparts (AU—arbitrary units). Upregulated and downregulated samples are defined relatively to *MCM6* median gene expression.

Table S1. K-means clustering from *MCM6* and *MKI67* expression gradients of 68 transcriptomes of anaplastic oligodendrogliomas from the POLA network. Functional annotations of the genes present in the five significant groups of co-expressed genes ($P < 0.01$), sorted by gene ontology enrichment (FDR, false discovery rate; GO, gene ontology).

Table S2. Differential gene expression statistics (t-test) between *MCM6*-up and *MCM6*-down, and between *MKI67*-up and *MKI67*-down samples, sorted by *P*-value (BH, Benjamini–Hochberg; FDR, false discovery rate; GO, gene ontology; SD, standard deviation).

Table S3. K-means clustering from *MCM6* (or *MKI67*) expression gradients of 35 transcriptomes of grade III oligodendrogliomas from the TCGA bank. Functional annotations of the genes present in the five significant groups of co-expressed genes ($P < 0.01$), sorted by gene ontology enrichment (FDR, false discovery rate; GO, gene ontology).

Table S4. K-means clustering from *MCM6* expression gradients of 98 transcriptomes of grade II and grade III oligodendrogliomas from the TCGA bank. Functional annotations of the genes present in the six significant groups of co-expressed genes ($P < 0.01$), sorted by gene ontology enrichment (FDR, false discovery rate; GO, gene ontology).

APPENDIX

List of investigators

POLA Network: Amiens (Christine Desenclos, H. Sevestre), Angers (Philippe Menei, A. Rousseau), Annecy (T. Cruel, S. Lopez), Besançon (M.I. Mihai, A. Petit), Brest (R. Seizeur, I. Quintin-Roué), Bicêtre (C. Adam, F. Parker), Bordeaux (S. Eimer, H. Loiseau), Caen (L. Bekaert, F. Chapon), Clamart (D. Ricard), Clermont-Ferrand (C. Godfraind, T. Khallil), Clichy (D. Cazals-Hatem, T. Faillot), Colmar (C. Gaultier, M. C. Tortel), Cornebarrieu (I. Carpiuc, P. Richard), Créteil (W. Lahiani), Dijon (H. Aubriot-Lorton, F. Ghiringhelli), Lille (C-A. Maurage, E. Le Rhun), Limoges (E. M. Gueye, F. Labrousse), Lyon (F. Ducray, D. Meyronet), Marseille (O. Chinot), Montpellier (L. Bauchet, V. Rigau), Nancy (P. Beauchesne), Nantes (M. Campone, D. Loussouarn), Nice (D. Fontaine, F. Vandenbos-Burel), Nîmes (A. Le. Floch, P. Roger), Orléans (C. Blechet, M. Fesneau), Paris (A. Carpentier, A. Idbaih, J. Y. Delattre [POLA Network National coordinator], K. Mokhtari, F. Bielle, S. Hamdi, M. Polivka), Poitiers (S. Milin), Reims (P. Colin, M. D. Diebold), Rennes (D. Chiforeanu, E. Vauleon), Rouen (O. Langlois, A. Laquerriere), Saint-Etienne (F. Forest, M. J. Motso-Fotso), Saint-Pierre de la Réunion (M. Andraud, G. Runavot), Strasbourg (B. Lhermitte, G. Noel), Suresnes (S. Gaillard, C. Villa), Toulon (N. Desse), Tours (C. Rousselot-Denis, I. Zemmoura), Toulouse (E. Cohen-Moyal, E. Uro-Coste), Villejuif (F. Dhermain).

2. CORRELATION BETWEEN DNA METHYLATION AND CELL PROLIFERATION IDENTIFIES NEW CANDIDATE PREDICTIVE MARKERS IN MENINGIOMA

Contexte

Cet article plus récent explore les marqueurs de l'épigénome, ici le méthylome de l'ADN, dans les méningiomes, et leur lien avec la progression dans cette tumeur cérébrale répandue, généralement indolente et à évolution lente. Une autre dimension omique de notre travail est présentée sous les traits de la méthylation de l'ADN par puces Illumina®, en lien avec la plateforme de génomique fonctionnelle du laboratoire. Cette étude met en avant l'aspect biostatistiques, prépondérant dans ce type d'approche contrainte : modèles linéaires versatiles (Bayes empirique) pour les analyses différentielles / catégorielles et continues, chevauchement et réduction des signatures selon de multiples variables, corrélations exhaustives à l'échelle du génome, et associations multivariées et pan-épigénomique à la survie avec modèles de régression de Cox (*Cox Proportional Hazards*). Notons le développement et l'application de protocoles dédiés aux annotations fonctionnelles des régions régulatrices, par essence complexes, biaisées et difficiles à interpréter avec les données du méthylome.

Points clés

- Associations avec / corrélations entre différents marqueurs / indices de prolifération et méthylation de l'ADN dans une cohorte de 48 patients atteints de méningiomes répartis sur 3 grades d'agressivité, tels que définis par l'OMS en 2021.
- Identification de régions régulatrices récurrentes, et par recoupement, d'une signature épigénétique de prolifération dans les méningiomes associée à la progression de la maladie et à la survie des patients.
- Découverte de biomarqueurs liés à la croissance cellulaire et au neurodéveloppement pouvant constituer des cibles pronostiques et thérapeutiques dans ces tumeurs.
- Confirmation en profondeur, avec une nouvelle exploration omique, que les grades OMS, basés sur l'histologie, ne reflètent pas aussi bien les classifications bâties sur les mesures moléculaires à grande échelle au regard de l'agressivité et des risques de récurrence après chirurgie.

cancers

MDPI

Article

Correlation between DNA Methylation and Cell Proliferation Identifies New Candidate Predictive Markers in Meningioma

Sébastien Hergalant [1,†], Chloé Saurel [2,†], Marion Divoux [1], Fabien Rech [3,4], Celso Pouget [2], Catherine Godfraind [5], Pierre Rouyer [1], Stéphanie Lacomme [6], Shyue-Fang Battaglia-Hsu [4] and Guillaume Gauchotte [1,2,6,*]

1 UMR Inserm 1256 NGERE (Nutrition, Génétique et Exposition aux Risques Environnementaux), Université de Lorraine, 54000 Nancy, France
2 Department of Biopathology CHRU-ICL, CHRU, 54500 Nancy, France
3 Department of Neurosurgery, CHRU, 54500 Nancy, France
4 CRAN, CNRS, Université de Lorraine, 54000 Nancy, France
5 Neuropathology, CHU of Clermont-Ferrand, UMR INSERM/Université d'Auvergne U1071, 63000 Clermont-Ferrand, France
6 Centre de Ressources Biologiques, BB-0033-00035, CHRU, 54500 Nancy, France
* Correspondence: g.gauchotte@chru-nancy.fr; Tel.: +33-3-83-65-60-17
† These authors contributed equally to this work.

Simple Summary: In adults, meningioma is the most common primary tumor of the brain. It is classified into three clinical grades of aggressiveness. Whereas disease recurrence after surgery and survival are associated with grade, it is worth investigating proliferation at a molecular level to identify markers capable of improving the clinical management of meningioma. In this study, we explore the DNA methylation profiles of 48 tumors of various grades and conduct statistical analyses on several proliferation indices and markers, such as mitotic index, grade, and Ki-67 or MCM6 expression levels. We identify differential methylation profiles between grades, loci highly correlated with cell growth and division, and a specific methylation signature of regulatory regions persistently associated with proliferation indices, grade, and survival. Finally, we report candidate genes under the control of these regions with potential prognostic and therapeutic value and deserving clinical evaluation.

Abstract: Meningiomas are the most common primary tumors of the central nervous system. Based on the 2021 WHO classification, they are classified into three grades reflecting recurrence risk and aggressiveness. However, the WHO's histopathological criteria defining these grades are somewhat subjective. Together with reliable immunohistochemical proliferation indices, other molecular markers such as those studied with genome-wide epigenetics promise to revamp the current prognostic classification. In this study, 48 meningiomas of various grades were randomly included and explored for DNA methylation with the Infinium MethylationEPIC microarray over 850k CpG sites. We conducted differential and correlative analyses on grade and several proliferation indices and markers, such as mitotic index and Ki-67 or MCM6 immunohistochemistry. We also set up Cox proportional hazard models for extensive associations between CpG methylation and survival. We identified loci highly correlated with cell growth and a targeted methylation signature of regulatory regions persistently associated with proliferation, grade, and survival. Candidate genes under the control of these regions include *SMC4*, *ESRRG*, *PAX6*, *DOK7*, *VAV2*, *OTX1*, and *PCDHA-PCDHB-PCDHG*, i.e., the protocadherin gene clusters. This study highlights the crucial role played by epigenetic mechanisms in shaping dysregulated cellular proliferation and provides potential biomarkers bearing prognostic and therapeutic value for the clinical management of meningioma.

Keywords: genome-wide DNA methylation; meningioma; methylome; proliferation signature; biomarkers; survival; Ki-67; MCM6

check for updates

Citation: Hergalant, S.; Saurel, C.; Divoux, M.; Rech, F.; Pouget, C.; Godfraind, C.; Rouyer, P.; Lacomme, S.; Battaglia-Hsu, S.-F.; Gauchotte, G. Correlation between DNA Methylation and Cell Proliferation Identifies New Candidate Predictive Markers in Meningioma. *Cancers* **2022**, *14*, 6227. https://doi.org/10.3390/cancers14246227

Academic Editor: David Wong

Received: 2 November 2022
Accepted: 14 December 2022
Published: 17 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Meningiomas are the most common primary tumors of the central nervous system in adults. The annual incidence rate ranges from 1.3/100,000 to 7.8/100,000 for cerebral meningiomas, with a tendency towards constant augmentation over the past few years [1]. The widely adopted WHO (World Health Organization) classification divides meningiomas into 15 subtypes and 3 grades of malignancy, mainly based upon histology [2]. Grade 1, 2, and 3 meningiomas represent about 70%, 20–30%, and 1–3% of reported cases, respectively. They correlate with recurrence risk (7–25% for grade 1, 29–59% for grade 2, and 50–94% for grade 3 [3]), as well as with 5-year and 10-year overall survival [4,5]. However, the histological criteria are rather subjective and are often associated with significant interobserver bias [6,7]; more reliable markers can thus improve the adequacy of treatments based on tumor grade. The last WHO classification (5th Edition, 2021) integrated a *TERT* promoter mutation or a homozygous deletion of *CDKN2A* and/or *CDKN2B* as new criteria for the recognition of grade 3 meningiomas. The standard of care is as follows: when deemed adequate, most patients undergo surgery, whereas adjuvant therapy is not systematic; for grade 2 and grade 3 meningioma, after surgery, conformational radiotherapy is recommended, particularly in cases of incomplete resection in grade 2 and in all grade 3 cases [8]. To date, no drug therapy has been validated for meningioma treatment.

The discovery of new molecular targets may present new therapeutic options in meningioma management. The study of the molecular landscape in meningioma is thus an important issue. However, until recently, few genetic variations have been described. These included the earliest finding of chromosome 22q deletion, which causes the loss of the tumor suppressor gene *NF2* [9], the inactivation of which was observed in about half of the meningiomas studied [10]. More recently, several genes with recurrent mutations were identified in meningioma, including proapoptotic E3 ubiquitin ligase TNF receptor-associated factor 7 (*TRAF7*), pluripotency transcription factor Kruppel-like factor 4 (*KLF4*), proto-oncogene v-Akt murine thymoma viral oncogene homolog 1 (*AKT1*), Hedgehog pathway-signaling member "smoothened" (*SMO*), and phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit A (*PIK3CA*). Approximately 40% of sporadic meningiomas harbor at least one of these variations [11,12]. Inhibitors of *SMO*, *AKT1*, and *PIK3CA* are therefore of therapeutic interest [8,10]. Other recently identified mutations are linked to a phosphatase tensin homolog on chromosome 10 (*PTEN*), as well as cyclin kinases *CDKN2A*/*CDKN2B*, main tumor suppressor genes predominantly implicated in meningioma progression [13–15]. However, approximately 20% of meningiomas present no known oncogenic mutation [10]. Other genetic rearrangements may be implicated; these include copy-number alteration and chromosomal abnormalities, both of which are associated with higher grade and poor tumor prognosis [16–19]. These unbalanced profiles can impact genes involved in cell cycle maintenance and progression, dysregulate major functional pathways, activate oncogenes, and inactivate tumor suppressor genes [13,20].

Recently, DNA methylation (DNAm) profiles have been studied to elaborate new prognostic classifications [21–24]. Two were landmark studies that greatly advanced our understanding of the role of DNAm in meningioma. One study investigated meningioma genome-wide DNAm patterns and classified them into three distinct and clinically relevant methylation classes (benign, intermediate, and malignant) and six methylation subclasses (benign-1, benign-2, benign-3, intermediate-A, intermediate-B, and malignant). This approach more efficiently predicted tumor recurrence than the WHO classification [23]. The other study identified four key molecular/phenotypic features associated with meningioma malignancy based on an integrative analysis of multi-omic data gathered from DNAm, somatic point mutations, copy-number aberrations, and mRNA abundance. These features presented immunogenic, benign *NF2* wild-type, hypermetabolic, and proliferative characteristics of the tumor tissue and could be used to determine the most appropriate therapeutic strategy. This study further associated independent immunohistochemical markers with each molecular group. For instance, a high expression of MCM proteins—from the helicase complex involved in DNA replication, which can be used as a proliferation marker—was

discovered to be characteristic of the proliferative group, with levels correlating with poor prognosis. This finding is consistent with our own observation that a high MCM6 index correlates with shorter progression-free survival [25], underscoring the crucial role of cell cycle progression and proliferation in meningioma progression.

Taken together, these insights prompted us to investigate the correlation between DNAm and meningioma cell proliferation and evaluate their levels of dependency. To achieve this, we studied tumor cell proliferation based on the immunohistochemical markers Ki-67 and MCM6, which are associated with histopathological factors such as WHO grade and mitotic index on one hand and genome-wide DNAm profiles of a meningioma cohort composed of 48 tumor tissues of various grades on the other hand. Our goal was to identify specific genes correlating with/presenting differentially methylated regions as a function of tumor proliferation. To conclude the potential clinical application of our results, we evaluated the association between DNAm and survival in an attempt to facilitate the discovery of potential new prognostic markers and/or therapeutic targets.

2. Materials and Methods

2.1. Population and Clinical Data

Forty-eight samples from surgical meningioma resections analyzed in the Department of Pathology of Nancy University Hospital (CHRU Nancy, France) between 2006 and 2018, with available frozen tissue, were randomly included, focusing on high-grade meningioma associated with a grade 1 control group. All samples were anonymized. The study was conducted in accordance with local ethical guidelines.

For each sample, we collected the WHO grade and other histopathological data from the pathology reports and clinical data from the medical records. Main clinical variables of interest were age, gender, localization of the tumor, quality of the excision (complete or not), treatment by chemotherapy or/and radiotherapy, recurrence of the meningioma and vital status. Because the date of first symptoms or diagnostic imaging was not available for all the patients, the diagnostic date was considered as the date of surgery. The progression date was considered as the day of the radiological exam during which progression was noted. The last consultation date and the last news date were also collected from the medical records. They were used to calculate overall survival (OS) and progression-free survival (PFS).

2.2. Histopathology

The WHO grade was established by a neuropathologist based on the WHO 2021 classification. The mitotic count per 10 high power fields (HPF; 1.6 mm^2) was assessed independently by two pathologists, and the mean value was calculated. The level of agreement between the two raters was measured with the intraclass correlation coefficient. Both analyses were performed on hematoxylin, eosin, and saffron (HES) slides, blinded to the clinical and molecular data.

2.3. Immunohistochemistry

For each case, all immunohistochemical (IHC) staining was performed from the same block of formalin-fixed, paraffin-embedded (FFPE) tissue, which was selected after reviewing all the HES slides. Slides were manually prepared by paraffin sections of 4 µm, followed by deparaffinization, rehydration, and antigen retrieval. IHC staining was performed using the following antibodies: MCM6 (mouse monoclonal antibody, clone H-8, sc-3936-16, 1/2000 dilution, Santa Cruz) and Ki-67 (mouse monoclonal antibody, clone MIB-1, GA62661, prediluted, Dako, Agilent). Immunochemistry was performed with a Dako Omnis (Dako Agilent) automate using an Envision Flex revelation system (Dako Agilent). The labeling index (LI) for MCM6 and Ki-67 was calculated as the percentage of tumor cells with nuclear staining counted among a total of 1000 tumoral cells.

2.4. DNA Methylation Analysis

DNA was extracted with a Macherey Nagel DNA extraction kit (Macherey-466 Nagel, Düren, Germany). After qualitative control, 900 ng of the extracted DNA was used to perform analyses with an Illumina MethylationEPIC BeadChip following the manufacturer's instructions, as previously described [24,25]. Raw data files (IDAT) were generated and used for downstream bioinformatics with the minfi package in R v4.1. The EPIC microarray interrogates 850K CpG sites, enabling a genome-wide methylome study including proximal promoters, distal regulatory regions, gene bodies, and intergenic features. Numerous preparation and filtering steps for quality control (as described in [26]), including technical checks and CpG removal from X and Y chromosomes, led to a 787,087 CpG working dataset. Normalization was carried out following the FunNorm procedure [27].

Supervised analyses included statistical modeling with empirical Bayes for case/control studies (R limma package) and reported the differentially methylated CpGs, with *p*-values for each comparison adjusted for false discovery rate (FDR) following the Benjamini–Hochberg procedure. Differentially methylated regions (DMRs) between groups were determined using the same linear models with the dmrcate function of the R package DMRcate (parameters: lambda = 1000, C = 3, less than a 1000 bp gap, and at least 3 CpGs). Unbiased functional annotations on ontological terms (gene ontology, GO) were achieved at CpG and DMR levels with the R package missMethyl [28]. Spearman's correlation tests were used for correlation analyses between CpG methylation levels and quantitative variables (mitotic index: number of mitoses, Ki-67 and MCM6 labeling indices: proliferative marker expressions). For each 787087 CpG, correlative metrics (Spearman's Rho, *p*-value, and FDR by Benjamini–Hochberg) and beta-value statistics (average, median, standard deviation, methylation range max/min, and Q3/Q1) were computed and reported, along with CpG island (CGI) and gene contexts. By default, beta value is expressed from 0 (demethylated) to 1 (methylated), whereas methylation differentials, changes, and ranges are expressed in % and represent the real proportion of beta-value change: a 20% change signifies a 0.2 increase/decrease in beta value.

Cox proportional hazard models were used to evaluate the association between CpG methylation levels and survival, either for PFS or OS analyses (R survival and survminer packages). After selection of the top CpGs with univariate setups in PFS in OS, multivariate analyses were conducted using age and grade as covariates. For each 787,087 CpG, associative metrics (Wald test *p*-values and effect size) were computed and reported, along with CGI and gene annotations. The proportional hazard assumption was tested on each fitted univariate model. Two statistics were considered: the *p*-value of non-random distribution for the methylation variable (CpG) and the global *p*-value for the model. When both *p*-values > 0.05, the CpG was retained for further multivariate evaluation. Proportional hazard assumptions were not rechecked for age and sex, as we only considered the statistical significance of the methylation covariate.

In every statistical approach, an FDR < 0.05 was considered significant. Supervised analyses (statistical modeling and DMR search) were conducted on methylation M values. Unsupervised analyses (hierarchical clustering) and correlative studies were performed with methylation B values (beta values).

Samples also underwent brain tumor classification and copy-number variation (CNV) estimation according to Capper et al. [29] and Sahm et al. [23]. Raw methylation files (IDAT) were uploaded to the MolecularNeuropathology.org server using the v11b4 classifier for brain tumor classification and MNGv2.4 for meningioma subtype classification. The evaluations of copy-number aberrations (along with CNV plots), *CDKN2A/B* loss, and *PTEN* loss were extracted from the CNV profile section of the generated report.

3. Results

3.1. Clinicopathological Data

Our cohort was composed of 48 patients with tumor samples of diverse WHO 2021 grades, including grade 1 (21%; 10/48), grade 2 (atypical; 69%; 33/48), and grade 3

(anaplastic; 10%; 5/48) meningiomas. In comparison with the original pathology reports, the grade changed in one case due to the detection of a homozygous *CDKN2A/B* deletion, upgrading the case from grade 2 to 3. The median age was 57 years, with a male-to-female ratio of 0.5. Meningioma tissues were mainly localized in the convexity (69%) and skull base (25%). No neoadjuvant treatment was delivered. Twenty-two patients (46%) underwent radiotherapy (postoperative and/or after progression; Table S1), on patient (2%) received post-progression chemotherapy, and four patients received antiangiogenic (bevacizumab) therapy. Disease progression occurred in 19 cases (40%; median PFS time: 39 months (16;55)). Nine patients died during the follow-up, with a median OS time of 52 months (31;95). The mean mitotic index was 5.4 mitoses/mm^2 (grade 1: 0.5/1.6 mm^2; grade 2: 4.9/1.6 mm^2; grade 3: 23/1.6 mm^2), with good inter-observer agreement (intraclass correlation coefficient: 0.75). Average Ki-67 and MCM6 labeling indices were 21% and 51%, respectively (Table 1).

Table 1. Clinical and pathological features.

WHO Grade (2021 Classification)	***n* = 48**
Grade 1	10 (21%)
Grade 2	33 (69%)
Grade 3	5 (10%)
Age	57 (48; 67)
Sex	
Female	32 (67%)
Male	16 (33%)
Localization	
Skull base	12 (25%)
Convexity	33 (69%)
Ventricular	2 (4%)
Spinal	1 (2%)
Complete resection	
Yes	26 (54%)
No	16 (33%)
Unknown	6 (13%)
Adjuvant chemotherapy	1 (2%)
Adjuvant radiotherapy	22 (46%)
Progression	19 (40%)
Median progression-free survival (months)	39 (16; 55)
Death	9 (19%)
Median overall survival (months)	52 (31; 95)
Ki67 (%)	21 (9; 38)
MCM6 (%)	51 (29; 73)
Mitoses/1.6 mm^2	2 (1; 6)

Results are presented as absolute values (*n*, %) or median (Q1; Q3). WHO: World Health Organization.

3.2. Molecular Data Based on Molecular Neuropathology Classifiers and Copy-Number Variations

On MolecularNeuropathology.org, the brain tumor classification was successfully run on all 48 samples, reporting the methylation class "meningioma" with a valid score (calibrated score > 0.9) for all samples (100%). The additional algorithm for meningioma

subtype identification classified 31 meningiomas (65%) with a calibrated score > 0.9: 13 in the benign class, 14 in the intermediate class, and 4 in the malignant class (Table 2). Grade 1 meningiomas were classified in benign and intermediate classes (80% and 20%, respectively). Among the 19 grade 2 cases with a calibrated score > 0.9, 5 were classified in benign (15%), 12 in intermediate (36%), and 2 in malignant (6%) classes.

Table 2. Molecular features from Sahm et al.'s methylation classifier reports according to the WHO grade.

WHO Grade	Grade 1 (n = 10)	Grade 2 (n = 33)	Grade 3 (n = 5)	Total (n = 48)
Methylation class				
Benign	8 (80%)	5 (15%)	0 (0%)	13 (27%)
Intermediate	2 (20%)	12 (36%)	0 (0%)	14 (29%)
Malignant	0 (0%)	2 (6%)	2 (40%)	4 (8%)
No match (calibrated score < 0.9)	0 (0%)	14 (42%)	3 (60%)	17 (35%)
Methylation sub-class				
1	2 (20%)	1 (3%)	0 (0%)	3 (6%)
2	4 (40%)	3 (9%)	0 (0%)	7 (15%)
3	0 (0%)	0 (0%)	0 (0%)	0 (0%)
4	1 (10%)	9 (27%)	0 (0%)	10 (21%)
5	0 (0%)	2 (6%)	0 (0%)	2 (4%)
6	0 (0%)	2 (6%)	2 (40%)	4 (8%)
No match (calibrated score < 0.9)	3 (30%)	16 (48%)	3 (60%)	22 (46%)
CDKN2A/B homozygous loss	0 (0%)	0 (0%)	2 (40%)	2 (4%)
PTEN loss	0 (0%)	14 (42%)	4 (100%)	18 (38%)
NF2 loss	5 (50%)	30 (91%)	4 (100%)	39 (81%)

Results are presented as absolute values (n, %). WHO: World Health Organization.

Copy-number variations (CNVs) identified by DNA methylation (DNAm) in the cohort of 48 meningiomas are represented in Figure 1. The total number of CNVs (complete or partial gain/deletion) averaged at 7.2 ± 5.7. The CNV profiles showed a homozygous loss of *CDKN2A*/B, loss of *PTEN*, and/or *NF2* genes in 0%, 0%, and 50% of grade 1 meningiomas; 0%, 42%, and 91% of grade 2 meningiomas; and 40%,100%, and 100% of grade 3 meningiomas, respectively (Table 2).

3.3. DNA Methylation and WHO Grade

At CpG resolution, differential analyses yielded 2099 significant hits between high-grade (grades 2 and 3; n = 38) and low-grade (grade 1; n = 10) meningiomas (Figure 2), with 1158 hypomethylated and 941 hypermethylated CpGs (moderated t-test; FDR < 0.05) (Figure S1). Of the total 2099 CpGs with FDR < 0.05, 2064 (98.3%) presented with a beta-value change > 1%, 1752 (83.5%) had a methylation differential > 5%, and 1641 (78.2%) showed > 10% change. Overall, hypomethylation occurred at already unmethylated and low methylation loci (beta value < 0.3 on a scale ranging from 0 to 1), with very low methylation dynamics (<10% change, a metric corresponding to 0.1 in beta-value change). Hypermethylation occurred to a greater extent (>50% change, i.e., >0.5 beta change) and mostly impacted medium-/high- methylation sites (beta value > 0.5) (Figure 3). Further analysis of areas spanning multiple CpGs areas consolidated these into 222 DMRs (with methylation differential > 10%), and in high grades, a majority of these were hypermethylated (200 DMRs) with only few hypomethylations (22 DMRs) (Tables 3 and S2). Indeed, hypermethylated loci had rather restricted distribution, as 33.8% of these CpGs hit the same

Cancers **2022**, *14*, 6227

gene at least twice. Furthermore, they were distributed preferentially within CpG islands (CGI context) in promoter and intergenic regions (gene context), hinting at enhancer-linked functions. The hypomethylated loci were spread more evenly, with 88.5% hypomethylations located within the gene body of exclusive genes and predominantly in regions outside the CGI (open-sea probes) (Figure 3).

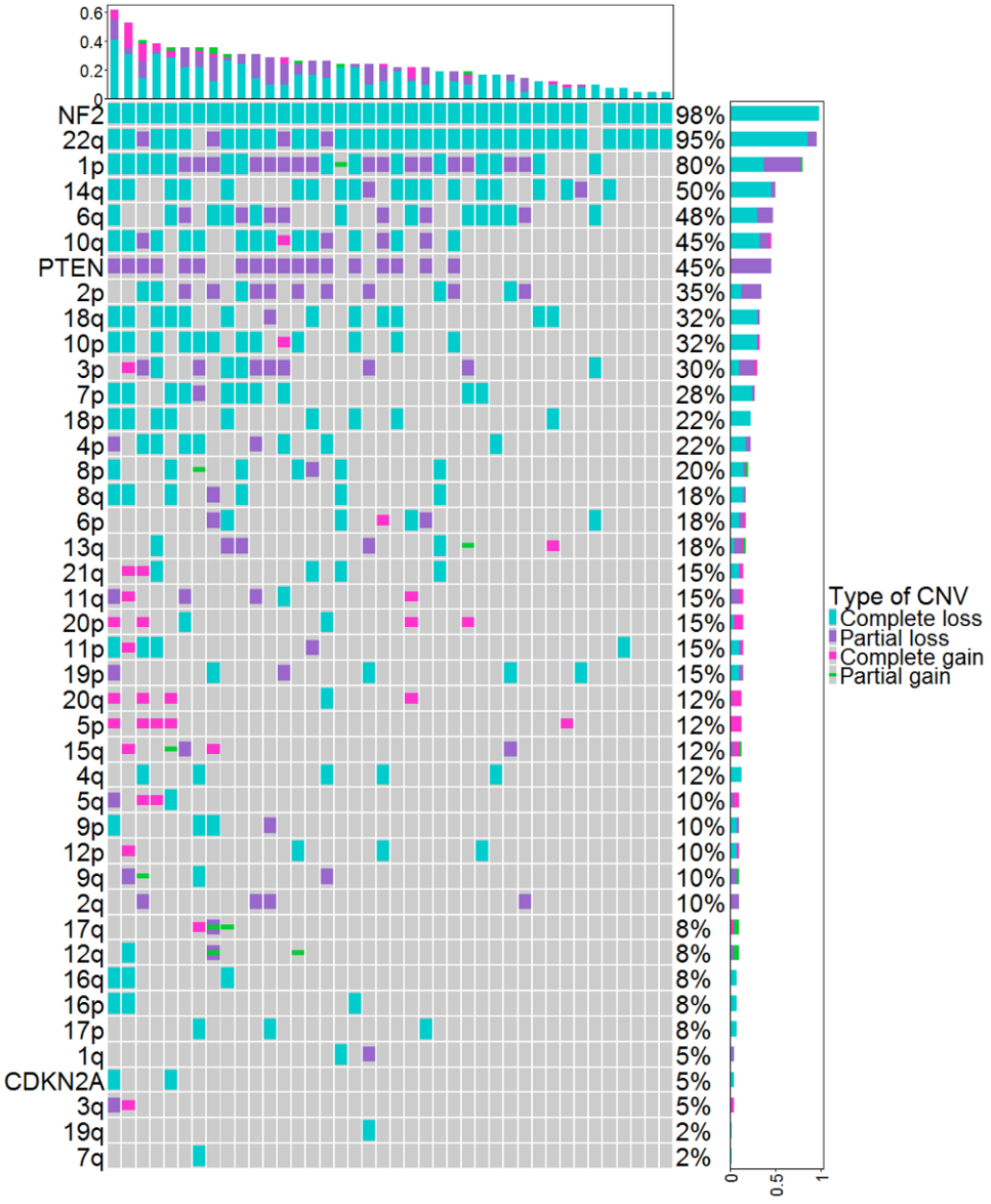


Figure 1. CNVs identified by DNAm in the cohort of 48 meningiomas. Each column represents individual patients ordered from left to right by increasing frequency of genetic alterations. Genetic alterations are ordered on the y axis from top to bottom by decreasing frequency of genetic alterations. CNV: copy-number variation, DNAm: DNA methylation.

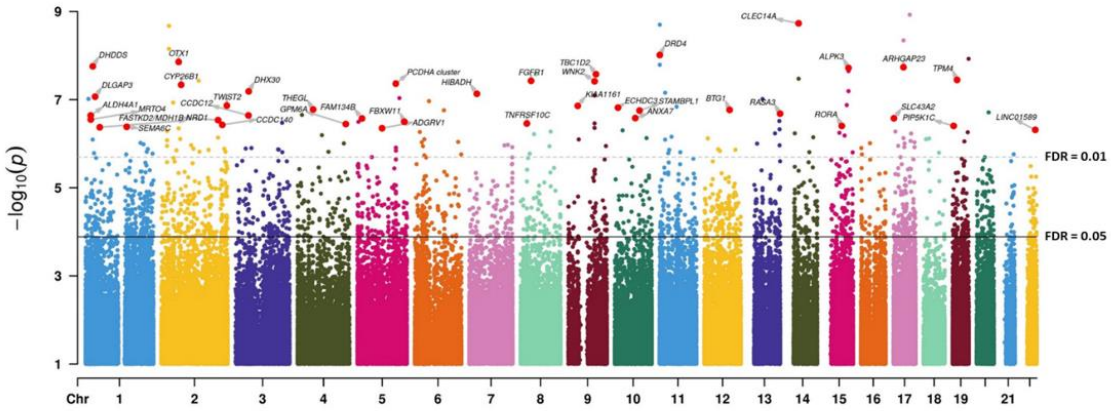


Figure 2. Manhattan plot representing the association between DNAm at the CpG level and high-grade (>1) vs. low-grade meningiomas. The dashed and solid lines indicate FDR thresholds after *p*-value correction for genome-wide multitesting (Benjamini–Hochberg). FDR: false discovery rate.

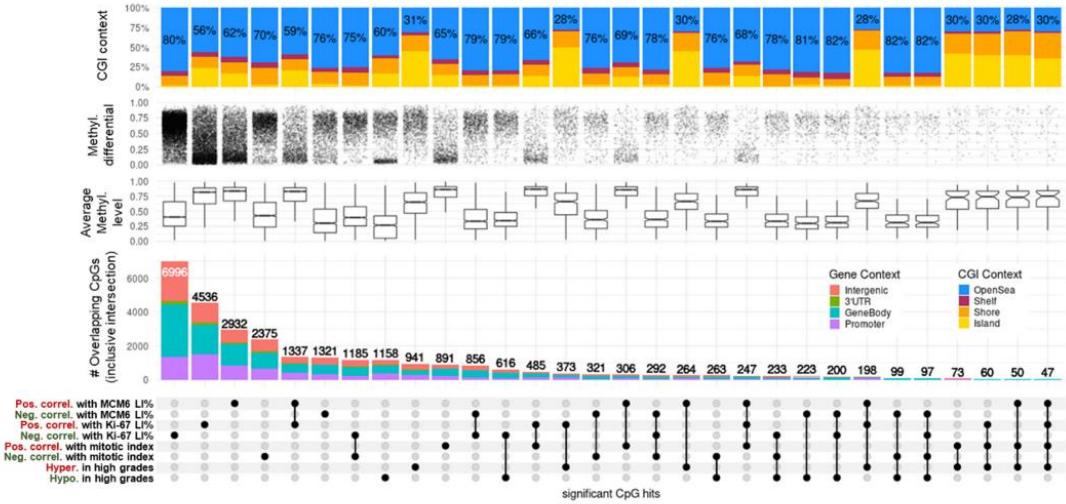


Figure 3. UpSet plot summarizing the overlap between CpGs highly correlated with proliferation indices and grades in meningiomas. Every possible intersection is displayed (30 in total, including the 8 original hit lists shown on the bottom-left side). All intersections are inclusive. CpG regulatory contexts (gene and CGI annotations) are represented as proportions of barplot heights. Proportions of open-sea CpGs are displayed on CGI context bar plots. Average methylation levels are represented as boxplots of means (beta value) for each CpG of the intersected lists. Methylation differentials represent the range of each CpG beta value (max–min) and the densities of changes. DNAm: DNA methylation; methyl: methylation; CGI: CpG island; pos. correl: positive correlation; neg. correl: negative correlation; hypo: hypomethylation; hyper: hypermethylation; LI: labeling index.

We next conducted unbiased gene ontology analysis on the set of genes associated with the 2099 differentially methylated CpGs. The most enriched biological process ontologies in high-grade versus low-grade meningiomas concerned main functions such as morphogenesis, neurogenesis, and cell differentiation (all FDR $< 1 \times 10^{-13}$) (Table 4) and included growth (FDR = 2.80×10^{-3}), cell leading edge (FDR = 2.89×10^{-3}), cell cycle (FDR = 2.95×10^{-3}), and apoptotic process (FDR = 4.97×10^{-6}). Highly enriched GO terms comprised more specific processes, such as cell morphogenesis involved in neuron differentiation (FDR = 2.81×10^{-11}, 464/555 genes), positive regulation of transcription by RNA polymerase II (FDR = 9.16×10^{-10}, 893/1151 genes), regulation of nervous

system development (FDR = 3.41×10^{-9}, 712/894 genes), Rho protein signal transduction (FDR = 3.31×10^{-5}, 159/191 genes), mitotic cell cycle checkpoint (FDR = 3.31×10^{-4}, 132/164 genes), regulation of the extrinsic apoptotic pathway (FDR = 4.79×10^{-4}, 121/149 genes), DNA damage checkpoint (FDR = 9.34×10^{-3}, 117/142 genes), and cellular response to prostaglandin stimulus (FDR = 0.02, 21/21 genes) (Table S3).

Table 3. Top differentially methylated regions (DMRs; 15 hypermethylated, 10 hypomethylated) ordered by average methylation differential (Mean Diff.) between high-grade (2, 3) and low-grade (1) meningiomas.

	Linked Gene	Chromosome	# CpGs	FDR	Mean. Diff.	Max. Diff.
Hypermethylated in high grades	CYP26B1	chr2	3	5.21×10^{-15}	+29%	+42%
	REC8	chr14	13	1.09×10^{-38}	+29%	+37%
	C2CD4D	chr1	5	6.10×10^{-13}	+28%	+35%
	KIFC2	chr8	4	2.45×10^{-14}	+28%	+35%
	CALCB	chr11	5	4.77×10^{-14}	+28%	+38%
	HEPACAM	chr11	4	1.02×10^{-16}	+28%	+38%
	DCDC2C	chr2	5	7.55×10^{-14}	+27%	+41%
	PAX6	chr11	15	7.63×10^{-23}	+27%	+42%
	SPEG	chr2	6	8.06×10^{-22}	+27%	+35%
	LTBP4	chr19	5	2.84×10^{-11}	+26%	+32%
	WNK2	chr9	8	1.11×10^{-35}	+25%	+43%
	PITX1	chr5	6	5.17×10^{-14}	+25%	+34%
	KLB	chr4	4	2.85×10^{-11}	+25%	+33%
	B4GALNT1	chr12	5	1.41×10^{-14}	+24%	+36%
	IRX1	chr5	8	2.14×10^{-15}	+24%	+37%
Hypomethylated in high grades	SMC4/miR16	chr3	4	2.30×10^{-16}	−41%	−46%
	ARHGAP23	chr17	3	6.68×10^{-24}	−36%	−44%
	PATJ	chr1	3	6.07×10^{-12}	−31%	−33%
	CACNA1H	chr16	3	3.94×10^{-14}	−28%	−35%
	THSD4	chr15	3	4.52×10^{-12}	−25%	−27%
	DNAJB6	chr7	7	2.61×10^{-13}	−23%	−36%
	TP63	chr3	9	2.38×10^{-19}	−23%	−42%
	LINC01589	chr22	4	1.30×10^{-16}	−23%	−32%
	DHX30	chr3	3	1.14×10^{-13}	−20%	−28%
	RBM47	chr4	11	6.62×10^{-19}	−19%	−28%

CpGs: number of CpG probes, FDR: false discovery rate, Mean. Diff. = mean difference (average methylation differential throughout every CpG of the DMR), Max. Diff = maximum difference (single CpG within the DMR showing the maximum methylation differential). Linked genes obtained from overlapping annotations (see Methods). DMRs without linked genes are omitted (full list in Supplementary Table S2).

Among the DMRs and multiple gene hits (Tables 3 and S2, Figure 2), notable hypomethylated regions in high-grade meningiomas were linked to cell cycle, cell differentiation, and cell fate genes, such as *SMC4/miR16* (four CpGs spanning 1036 bp, 41% average methylation differential (AMD)), *PATJ* (three CpGs spanning 155 bp, 31% AMD), and *TP63* (nine CpGs spanning 650 bp, 23% AMD; coding for the tumor protein p63). DMR hypermethylation in high-grade meningiomas concerned important genes of neural development, such as *CALCB* (CpGs spanning 770 bp, 28% AMD), *PAX6* (DMRs interspersed by 4.6 Kb and associated with alternative transcription start sites: (i) 15 CpGs, 2.76 kb, 27% AMD; and (ii) 9 CpGs, 1.21 kb, 17% AMD), the *PCDH* gene clusters (3 CpGs, 374 bp, 18% AMD), and *WNK2* (8 CpGs, 1169 bp, 25% AMD). There was also a strong enrichment in *SUZ12* (4.8-fold, q-value = 1.33×10^{-14}) and EZH2 (5.9-fold, q-value = 2.87×10^{-4}) target regions, both determinants of the polycomb repressor complex 2, which is involved in gene-silencing processes.

Table 4. Top 20 gene ontologies (biological processes) based on the highest significance (FDR) between high-grade (>1) and low-grade meningiomas.

Reference	Ontology Term	N	DE (%)	P.DE	FDR
GO:0009653	anatomical structure morphogenesis	2629	77.7%	1.44×10^{-27}	3.27×10^{-23}
GO:0007399	nervous system development	2264	79.1%	6.68×10^{-27}	7.60×10^{-23}
GO:0048856	anatomical structure development	5697	73.0%	3.36×10^{-26}	2.55×10^{-22}
GO:0032502	developmental process	6073	72.6%	1.65×10^{-25}	9.36×10^{-22}
GO:0016043	cellular component organization	6081	73.2%	5.20×10^{-25}	2.37×10^{-21}
GO:0007275	multicellular organism development	5227	73.1%	2.03×10^{-24}	7.71×10^{-21}
GO:0071840	cellular component organization or biogenesis	6261	72.9%	3.91×10^{-24}	1.11×10^{-20}
GO:0048518	positive regulation of biological process	5830	72.6%	2.35×10^{-23}	5.94×10^{-20}
GO:0048522	positive regulation of cellular process	5143	73.1%	5.74×10^{-23}	1.31×10^{-19}
GO:0048731	system development	4689	73.2%	3.42×10^{-22}	7.08×10^{-19}
GO:0048468	cell development	2096	77.8%	2.10×10^{-20}	3.97×10^{-17}
GO:0048869	cellular developmental process	4222	73.0%	1.74×10^{-19}	3.05×10^{-16}
GO:0022008	neurogenesis	1563	79.5%	3.76×10^{-19}	6.11×10^{-16}
GO:0009893	positive regulation of metabolic process	3450	73.9%	4.98×10^{-19}	7.56×10^{-16}
GO:0048699	generation of neurons	1465	79.7%	1.95×10^{-18}	2.77×10^{-15}
GO:0010604	positive regulation of macromolecular metabolic process	3193	74.1%	3.98×10^{-18}	5.33×10^{-15}
GO:0030182	neuron differentiation	1310	80.4%	5.16×10^{-18}	6.52×10^{-15}
GO:0030154	cell differentiation	4040	72.7%	1.28×10^{-17}	1.53×10^{-14}
GO:0031325	positive regulation of cellular metabolic process	3166	74.0%	1.79×10^{-17}	2.04×10^{-14}
GO:0010646	regulation of cell communication	3381	73.9%	2.44×10^{-17}	2.64×10^{-14}

GO: gene ontology; N: number of genes in the GO category; DE: percentage of differentially methylated genes (corrected for representation bias and normalized against background) among the GO category; P.DE: over-representation *p*-value; FDR: false discovery rate.

3.4. Correlation between DNAm and Mitotic Index

CpG methylation levels were significantly correlated with mitotic index (MI) in 891 occurrences for positive correlations (PC) and 2,375 (77%) for negative correlations (NC) (FDR < 0.05, Spearman test; Table S4, Figure 3). Moreover, with 94.2% very high correlations (|rho| > 2/3, 32/34 CpGs) being negative, DNAm was predominantly negatively correlated with MI. Most NC-linked CpGs were in the medium-methylation category (0.3 > beta value < 0.7) and displayed high methylation dynamics (60–90% variation). Conversely, PC with MI mostly covered methylated CpGs (beta-value > 0.7) with feeble methylation changes (<10%). Overall, these 3,266 correlated loci represented open-sea probes (70% in NC and 65% in PC) and were distributed evenly in proximal promoters, gene bodies, and intergenic contexts (Figure 3).

We obtained further molecular insights after cross-examination of significant loci reported between grade and MI. Intersection of grade-related CpGs with MI-related CpGs resulted in 73 hypermethylations/PC and 263 hypomethylations/NC. No other overlap existed between the four lists, implying the presence of a bidirectional DNAm regulatory

mechanism for PC CpGs within CGI and regulatory regions and for NC CpGs outside CGIs in gene bodies (Figure 3). A list of the top correlated hits/genes with MI highlights the biologically relevant methylation candidate markers in meningiomas (Table 5, upper part). These hits were selected based not only on their very high methylation ranges (max–min > 50% and Q3–Q1 > 10%) but also on their gene-wise aggregation in the regulatory elements of the genes (annotations as islands, shores, and shelves within the CGI context, as well as in proximal or distal promoters in the gene context). These included CpGs linked to *SMC4/miR16*, p53 effector *CD82*, and probable methyltransferase *METLL24* (seven, two, and two hits, respectively) for NC and CpGs linked to transcription factors *ESRRG*, *PAX9*, and *OTX1* for PC (six, two, and two hits, respectively). Together with other dynamic and island-restricted CpGs, such as *ARHGDIA* (3 hits, NC), *DOK7* (2 hits, NC), *CAPN2* (2 hits, NC), the *PCDH* clusters (10 hits, PC) and *PAX6* (10 hits, PC), these candidates constitute potential proliferative biomarkers associated with disease progression (Table S4).

Table 5. Top correlated CpGs between DNAm levels (beta values) and mitotic indices, Ki-67 labeling indices, and MCM6 labeling indices in regulatory region contexts.

		CpG	Gene (# Hits)	Chr.	Rho	*p*-Value	DNAm Median Level (Beta-Value)	DNAm Diff. Range (%)
Mitotic index	negative	cg21942721	ASB4 (1)	chr7	−0.71	1.27×10^{-8}	High (0.84)	60 (26)
		cg18568061	PTRF (1)	chr17	−0.69	4.94×10^{-8}	Medium (0.37)	63 (21)
		cg01764105	SMC4/miR16 (7)	chr3	−0.69	7.46×10^{-8}	Low (0.08)	76 (21)
		cg17605814	CD82 (2)	chr11	−0.68	1.22×10^{-7}	Medium (0.57)	75 (25)
		cg22624818	SDPR (2)	chr2	−0.68	1.34×10^{-7}	High (0.80)	66 (17)
		cg16166651	DEPDC1 (3)	chr1	−0.67	1.53×10^{-7}	Low (0.16)	62 (12)
		cg06003566	METTL24 (2)	chr6	−0.67	2.19×10^{-7}	High (0.86)	79 (14)
	positive	cg25588576	MIR7641-2 (1)	chr14	0.66	3.52×10^{-7}	High (0.79)	69 (31)
		cg21784383	ESRRG (6)	chr1	0.65	5.83×10^{-7}	Low (0.18)	71 (27)
		cg18361098	PAX9 (2)	chr14	0.62	2.80×10^{-6}	Low (0.10)	80 (43)
		cg10640333	OTX1 (2)	chr2	0.61	4.33×10^{-6}	High (0.76)	90 (20)
		cg13244312	TTC9 (1)	chr14	0.61	4.81×10^{-6}	Medium (0.58)	55 (19)
Ki-67 LI%	negative	cg01464849	SMC4/miR16 (9)	chr3	−0.73	3.33×10^{-9}	Low (0.26)	84 (58)
		cg18943088	IQCJ-SCHIP1 (16)	chr3	−0.71	1.78×10^{-8}	Low (0.30)	83 (42)
		cg22800629	RAB33B (2)	chr4	−0.69	7.13×10^{-8}	Low (0.09)	53 (15)
		cg17253087	HIPK3 (1)	chr11	−0.68	7.97×10^{-8}	Medium (0.55)	78 (45)
		cg11629830	IQGAP2 (3)	chr5	−0.68	1.04×10^{-7}	Medium (0.69)	68 (36)
		cg13944632	VAV2 (4)	chr9	−0.66	2.97×10^{-7}	Medium (0.37)	75 (15)
	positive	cg03126579	ZFR2 (1)	chr19	0.71	1.27×10^{-8}	Medium (0.65)	93 (37)
		cg10269365	CCDC140 (10)	chr2	0.69	6.92×10^{-8}	Medium (0.55)	84 (37)
		cg08139247	CLEC14A (5)	chr14	0.68	9.42×10^{-8}	Medium (0.53)	83 (48)
		cg21784383	ESRRG (7)	chr1	0.66	3.34×10^{-7}	Low (0.18)	71 (27)
		cg26418900	NPY (4)	chr7	0.66	3.34×10^{-7}	Medium (0.40)	87 (32)
		cg10640333	OTX1 (4)	chr2	0.65	5.83×10^{-7}	High (0.76)	90 (20)
MCM6 LI%	negative	cg04570316	GMNN (1)	chr6	−0.62	2.45×10^{-6}	High (0.89)	71 (11)
		cg09130952	CCDC108 (1)	chr2	−0.62	2.52×10^{-6}	High (0.71)	68 (25)
		cg16959792	SLC50A1/EFNA1 (1)	chr1	−0.62	2.57×10^{-6}	Low (0.23)	39 (18)
		cg24310126	FLJ46361 (1)	chr10	−0.62	2.99×10^{-6}	High (0.85)	78 (16)
		cg03805253	CACNA1G (1)	chr17	−0.62	3.05×10^{-6}	Medium (0.68)	61 (19)
		cg10422455	MRGPRX2 (1)	chr11	−0.61	3.69×10^{-6}	Medium (0.66)	82 (33)
	positive	cg03126579	ZFR2 (1)	chr19	0.68	8.32×10^{-8}	Medium (0.65)	93 (37)
		cg03552103	SEPT10/ANKRD57 (1)	chr2	0.67	1.70×10^{-7}	Medium (0.69)	50 (13)
		cg15415136	ZNF540 (2)	chr19	0.66	3.60×10^{-7}	Low (0.15)	56 (10)
		cg06488443	TBR1 (2)	chr2	0.66	3.93×10^{-7}	Low (0.29)	61 (22)
		cg05008496	SSPN (1)	chr12	0.64	1.20×10^{-6}	Medium (0.56)	52 (22)
		cg22151446	PCDHabg clusters (27)	chr5	0.63	1.61×10^{-6}	Medium (0.31)	66 (38)
		cg12052661	CACNA1B (2)	chr9	0.63	1.72×10^{-6}	Medium (0.39)	52 (23)

DNAm: DNA methylation; LI: labeling index; # hits: total number of significant CpGs in the same regulatory region; Chr: chromosome; diff: differential. Low (methylation beta value $\leq$ 0.3), medium (beta value > 0.3 and <0.7), and high (beta value $\geq$ 0.7) median levels. Differential ranges in max–min % (Q3–Q1%). Spearman's Rho correlations. Full lists and metrics in Tables S3–S5.

3.5. DNAm and Ki-67 Labeling Index

CpG methylation levels and Ki-67 labeling index (LI) correlated significantly in 11,532 occurrences, with 4536 PC and 6996 (61%) NC (FDR < 0.05; Table S5, Figure 3). DNAm was also largely negatively correlated with Ki-67 levels, with 88.4% NC at a very

high cutoff (|rho| > 2/3, 38/43 CpGs). Similar to what was observed with MI, the majority of NC-linked CpGs were in the medium (0.3 > beta value < 0.7) methylation category and displayed high (50–90%) methylation dynamics, whereas DNAm PC with Ki-67 expression covered methylated CpGs (beta value > 0.7), mostly displaying moderate methylation variations (<25%). The proportion of CGI and known promoters was also higher in PC than in NC CpGs (Figure 3).

Ki-67 expression was highly correlated with MI (Spearman's Rho = 0.71, p-value = 1.22×10^{-8}). Surprisingly, less than half of the CpGs correlating with MI also correlated with Ki-67 levels (1185 for NC and 485 for PC). None of the loci in NC with Ki-67 overlapped with those in PC with MI and vice-versa. As with MI, we found no CpG overlap between NC with Ki-67 and hypermethylated sites in high grades, nor for CpGs in PC with Ki-67 and hypomethylated sites in high grades (Figure 3).

A list of the top correlated hits/genes relevant to Ki-67 LI is presented in Table 5 (middle part; same criteria as above with MI), including CpGs linked to *SMC4/miR16* and the GTPases *RAB33B* and *VAV2* (9, 2, and 4 hits, respectively) for NC and CpGs linked to long non-coding RNA *CCDC140* and *ESRRG* and neuropeptide *NPY* for PC (10, 7, and 4 hits, respectively). Candidates such as *DOK7* (3 hits, NC), the *PCDH* clusters (32 hits, PC), *PAX6* (15 hits, PC), and *TBR1* (6 hits, PC), among others with dynamic and island-restricted CpGs, constitute potential proliferative biomarkers in meningiomas (Table S5).

3.6. DNAm and MCM6 Labeling Index

CpG methylation levels and MCM6 LI correlated significantly in 4253 occurrences, with 2932 (69%) PC and 1321 NC (FDR < 0.05; Table S6, Figure 3). In contrast to what was observed with MI and Ki-67 LI, DNAm was generally positively correlated with MCM6 levels, with 70% PC at a very high cutoff (|rho| > 2/3, 7/10 CpGs). NC-linked CpGs were in the low/medium (beta-value < 0.4) methylation category and displayed high (50–90%) dynamics, whereas PC with MCM6 expression covered highly methylated CpGs (beta value > 0.8), most of which displayed moderate methylation variations (<25%). The proportion of island CpGs was also much higher in PC than in NC (Figure 3).

Nevertheless, MCM6 expression correlated positively with MI (Rho = 0.6, p-value = 5.31×10^{-6}) and Ki-67 LI (Rho = 0.69, p-value = 7.77×10^{-8}). Less than one-fifth of the CpGs correlating with MCM6 levels also correlated with MI (321 for NC and 306 for PC), with still no overlap between PC in MI and NC in MCM6 and vice-versa. Conversely, more than a half of the CpGs correlating with MCM6 levels also correlated with Ki-67 levels (856 for NC and 1,337 for PC). Again, no overlap was found between PC in MCM6 and NC in Ki-67 and vice-versa. Finally, as with MI and Ki-67, we found no CpG overlap between those in NC with MCM6 and hypermethylated sites in high grades, nor for CpGs in PC with MCM6 and hypomethylated sites in high grades (Figure 3). Thus, the controlled correlation structure observed between top loci associated with MI and WHO grades held and propagated with Ki-67 and MCM6 markers. The biologically relevant methylation markers based on the top correlated hits with MCM6 LI included CpGs linked with DNA replication inhibitor *GMNN* and ephrin *EFNA1* (a single hit for both) for NC CpGs linked with brain transcription repressor *TBR1* and *PCDH* gene clusters for PC (2 and 27 hits, respectively) (Table 5, lower part). Along with other candidates also associated with WHO grades, such as *SMC4* (three hits, NC) and *PAX6* (four hits, PC), these genes may constitute other interesting progression biomarkers in meningiomas (Table S6).

3.7. DNAm Proliferative Signature in Meningiomas

We did not observe a tight relationship between MI and Ki-67 DNAm markers and even less so with MCM6 levels, as only a small proportion of hits intersected (Figure 3). However, the overlapping structures were conservative between the three correlation experiments. Hence, we next evaluated the dependencies and the complementarities of these measurements (MI, Ki-67, and MCM6 LI) and derived a DNAm proliferation signature in meningiomas. This 310-CpG signature was obtained by cumulating top hits

Cancers **2022**, *14*, 6227

from the three positive (247 intersected CpGs) and three negative (292 intersected CpGs) correlation analyses (Figure 3) and by limiting the results to highly dynamic variables only (standard deviation > 10%; Table S7). Hierarchical clustering revealed a strongly correlated CpG structure, with an overall progressive sample demethylation according to both proliferation—with a very good adequation with MI—and disease progression, as the signature was also associated with WHO grade (Figure 4). Indeed, benign meningiomas (left side of the heat map) were methylated for a large open-sea CpG set and unmethylated at island CpGs, which constituted one-fifth of the signature. Samples with very low MCM6 LI (<30%) aggregated within this cluster. Grade 3 samples belonged to another cluster (right side of the heat map) of extreme methylation values for these CpGs. This cluster also aggregated samples with very high Ki-67 LI (>70%).

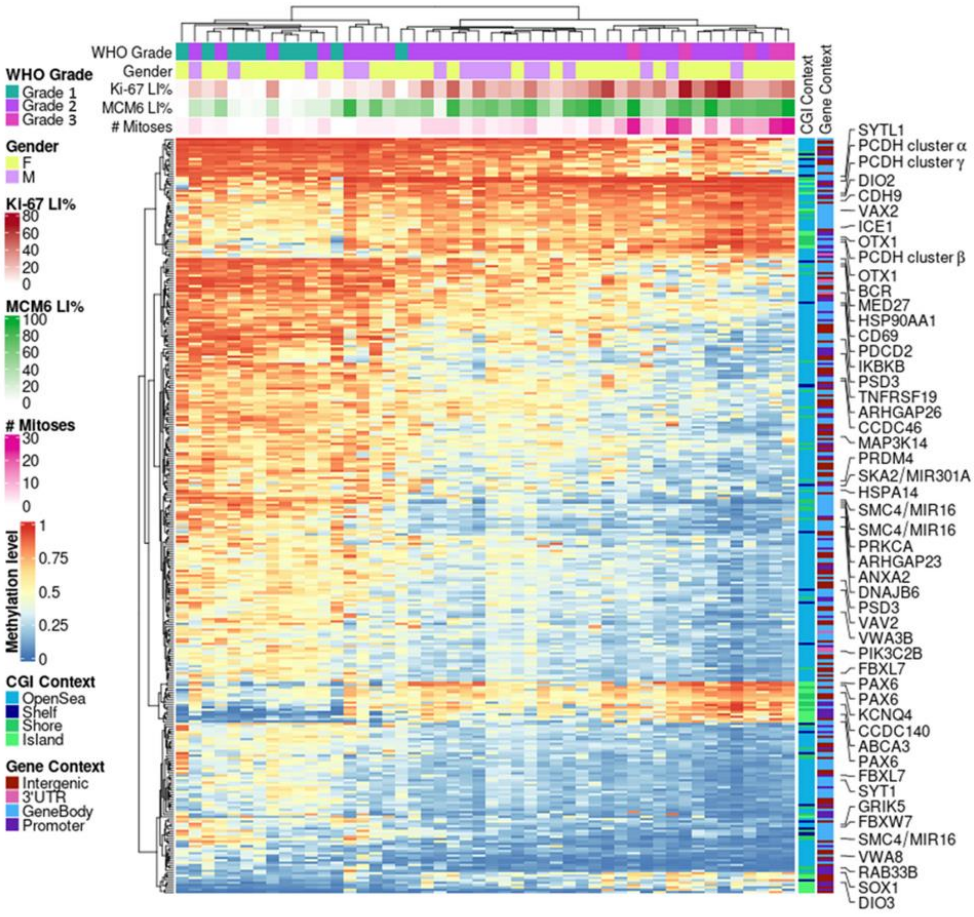


Figure 4. DNAm proliferation signature in meningiomas. Hierarchical clustering heat map depicting 310 CpGs highly correlated between DNAm and three proliferative markers: mitotic index, Ki-67, and MCM6 expressions (metrics: Euclidean distance and complete linkage). CGI: CpG island; DNAm: DNA methylation; LI: labeling index; WHO: World Health Organization.

This DNAm signature was validated on an external meningioma dataset (GSE200321) of 60 QC-passed samples of various grades and histological subtypes, which also included invasiveness information (Figure S2). Hierarchical clustering again revealed the same patterns of progressive demethylation at open-sea CpGs across samples with increasing invasive and aggressive properties, accompanied by a strong methylation of island CpGs. Notably, chordoid meningioma is a morphological subtype designated as WHO grade 2 but

mostly present with a low proliferation profile comparable to that of grade 1 meningiomas. These results are in line with recent findings suggesting that histology alone may not justify a grade 2 designation for chordoid meningiomas [30].

3.8. Associations between DNAm and Survival

We next conducted survival analyses with Cox proportional hazard models to identify specific loci associated with PFS and OS. We first examined the DNAm proliferative signature described above. As expected and independent of WHO grade and age, these 310 CpGs were all associated with survival, albeit better for PFS, with *p*-values in the range of [1.34×10^{-5}–1.47×10^{-3}], than for OS, with *p*-values in the range of [1.69×10^{-3}–2.8×10^{-2}] (Table S7; Wald test of multivariate settings).

We then examined genome-wide univariate associations with survival and observed a stronger DNAm association with PFS (69 hits, $p < 1 \times 10^{-5}$) than with OS (no hits, $p < 1 \times 10^{-5}$; 54 hits with $p < 1 \times 10^{-4}$) (Figure 5). Furthermore, PFS was associated massively with chr1q and chr20 loci, with 9.5- and 4.7-fold enrichments against background CpGs, respectively (both $p < 2.2 \times 10^{-16}$, chi-squared test; hits with $p < 1 \times 10^{-4}$). No such associations were found with OS. From these top univariate hits, we derived multivariate associations with PFS (Table S8) and OS (Table S9). One key conclusion is that for PFS, most loci (48/69) had a positive effect size, meaning that a methylation gain was linked to improved survival. This result echoes our findings based on grade/proliferation: the higher the proliferation (and the higher the grade), the lower the extent of DNAm (Figure 4). This trend was not observed for OS.

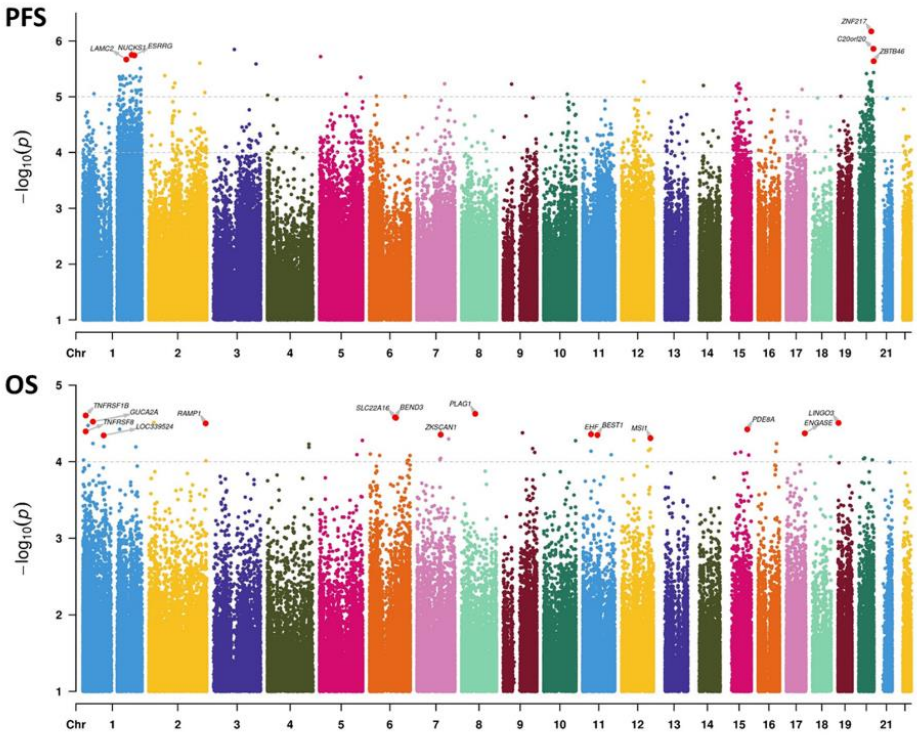


Figure 5. Manhattan plots representing the association between DNAm at the CpG level and survival. Upper panel: associations with progression-free survival (PFS). Lower panel: associations with overall survival (OS). Univariate Cox statistics (Wald test). *p*-value cutoffs running multivariate analyses on selected CpGs are indicated by the topmost dashed lines in each panel. For the PFS panel, the second dashed line represents the threshold for computing hit enrichments over chromosomic regions.

4. Discussion

DNAm is an innovative tool that is increasingly used for the classification of brain tumors and useful to predict their clinical outcome [29]. In our setup of 48 samples with varied disease progression, all samples were recognized as meningiomas by the main brain tumor classification tool. However, regarding the specific meningioma algorithm, MNGv2.4, and despite our choice of using only freshly frozen tissues, we were still confronted with a high rate of non-classifiable meningiomas (35%). This rate was even higher for meningioma subclassification (46%), which limited its use. Our study unveiled the association between indices of proliferation with disease progression and DNAm and consequently helped to identify a targeted proliferation-relevant DNAm signature comprising only a few hundred CpGs. As this signature could easily be used to score or classify meningioma samples, we propose that such a methylation-based classifier could ameliorate their categorization.

In our attempt to link proliferation/grade/disease outcomes and DNAm, we observed two main dynamics associated with cell proliferation in meningiomas. First, there were few but strong methylations at CGIs located in proximal promoters and enhancers mostly affiliated with neural transcription factors and tumor suppressor genes. One such hit is the developmentally targeted epigenetic silencing by polycomb repressor complex 2; although it is difficult to extrapolate these previous results to meningiomas, this silencing is well-documented in brain cancers [31,32]. Second, there was an extended and progressive demethylation at open sea, mostly within gene bodies. This observation has also been reported in cancers [33,34] although in pediatric brain cancers, the enriched hits were found to be intergenic, with no gene affiliation [6]. This mechanism has been associated with chromosomal instability [35] and malignant progression of lower-grade glioma [36] and was predictive of response to standard chemotherapy in osteosarcoma [37].

Among the epigenetic changes between low- and high-grade meningiomas, we found expected pathways driving unchecked cell division, such as the hypomethylation of genes promoting cell cycle, growth, differentiation, and fate. We also reported dysregulation of regulatory pathways of DNA damage checkpoints and extrinsic apoptosis, along with genes with full driving potential, such as *TP63* [38,39] and *SMC4*. *SMC4* was found to be involved in tumor cell growth, migration, and invasion. It is also correlated with poor prognosis in some cancers. Indeed, its overexpression is suspected to play a role in numerous cancers, such as hepatocellular, colorectal, breast, and endometrial cancer [40–44]. In glioma, *SMC4* overexpression promotes aggressive phenotypes by TGFβ/Smad signaling [45–47]. Hence, the hypomethylation of *SMC4* in high-grade meningiomas could likely lead to overexpression, which is consistent with previous data indicating poor prognosis in other tumors. It would therefore be interesting to evaluate its predictive role in meningioma, as it was also highly correlated with all proliferation indices. Additionally, hypermethylated markers in high-grade meningiomas included proximal promoters of *PAX6* and *PCDH* genes, in addition to correlation with proliferation markers. PAX6, a transcription factor playing an important role in the development of the central nervous system, constitutes a good candidate in our study, as in many regulatory loci, it was positively correlated with all proliferative markers. In glioblastoma, a few studies have suggested that a low *PAX6* expression level should be considered prognostically pejorative [48–50]. The protocadherin (PCDH) family of proteins plays an important role in neural cell aggregation, cell recognition, and neural development. PCDHs are the "barcodes" of the cell, generating single-cell diversity in the brain [51]. They are encoded by combinations obtained by alternative splicing from the three mapped-in-tandem, multigene *PCDH* α, β, and γ clusters. This splicing is controlled through DNAm states of alternative promoters [52], *PCDH* cluster expression involving CTCF interaction, which can be affected by methylation alterations, leading to long-range epigenetic silencing in cancers [53]. Here, we highlight strong and spread-out PCDH promoter hypermethylations linked to disease progression and cell proliferation markers. We suggest that these aberrant methylations may be of diagnostic value in meningioma. Finally, our findings with respect to the hypermethylation of the tumor suppressor and familial meningioma *WNK2* in grade 2 and 3 meningioma are consistent

with previously published data, which suggested epigenetic alterations to be the dominant, grade-specific mechanism of gene inactivation [54,55].

When considering mitotic activity in meningiomas, we found negative correlations with DNAm for *ARHGDIA* and *DOK7* and positive correlations for *ESRRG* and *OTX1* (both correlating with all proliferation indices). *ARHGDIA* codes for a Rho GDP dissociation inhibitor protein with antiapoptotic activity and is involved in cellular processes, such as cell proliferation, cell cycle progression, and cell migration [56]. Its expression is altered in various cancer, such as breast, prostate, and hepatocellular cancer [57–59]. In glioma, its level was reported to be correlated with positive prognosis and was considered both an independent predictive marker for OS [56] and an actionable therapeutic target [60]. *DOK7* was also reported to be correlated with the Ki-67 index and may function as a tumor suppressor gene, which is consistent with our findings. However, in breast carcinoma, significant *DOK7* promoter hypomethylation implicated its role in early tumorigenesis [61,62]. In lung cancer, lower *DOK7* expression was associated with lower survival. In in vitro studies, *DOK7* was reported to inhibit proliferation and migration by downregulating the PI3K/AKT/mTOR pathway [63,64]. Similarly, *DOK7* was found to be downregulated in glioma, and its overexpression inhibited both in vitro and in vivo proliferation of glioma cells [65]. ESRRG is a transcriptional activator of *DNMT1* (DNA methyltransferase 1). Its downregulation correlates with poor clinical outcome in gastric cancer, where it acts by suppressing cell growth and tumorigenesis. It also antagonizes Wnt signaling [66]. *ESRRG* promoter hypermethylation was used as a diagnostic and prognostic biomarker in laryngeal squamous cell carcinoma [67]. OTX1 is a transcription factor that is required for proper brain and sensory organ development in mice [68], where it regulates cell cycle progression, with its knockdown leading to diminished neurons but increased astrocytes [69]. Few studies have linked *OTX1* overexpression with tumorigenesis and growth in cancers [70,71], and to the best of our knowledge, it has not been yet reported as an epigenetic marker in brain cancers.

Between DNAm and Ki-67 levels, we report negative correlations for VAV2, a guanine exchange factor playing an important role in angiogenesis. Its recruitment by phosphorylated EPHA2 is critical for EFNA1-induced RAC1 GTPase activation, as well as vascular endothelial cell migration and assembly. Interestingly, we also found EFNA1 DNAm to be correlated with MCM6 expression. VAV2 is overexpressed in numerous cancers [72–74], but its dysregulation has not yet been reported in brain cancers. It is implicated in both cutaneous and head and neck squamous cell carcinomas, where it promotes regenerative proliferation [75]. In esophageal squamous cell carcinoma, it is required for DNA repair and radiotherapy resistance [76]. Because it is a Rho guanine exchange factor, VAV2 is an attractive pharmacological target, with druggable catalytic sites and a more restricted tissue distribution pattern than other Rho proteins. *TBR1* DNAm is positively correlated with the Ki-67 and MCM6 indices. TBR1 is a transcriptional repressor involved in multiple aspects of cortical development, including neuronal migration and axonal projection. Recurrent variations of *TRB1* have been described in medulloblastoma, and a high frequency of its copy-number loss has been detected in glioblastoma, suggesting a possible involvement in tumorigenesis or progression [77,78]. To the best of our knowledge, no data are available in the literature with respect to any implication of these genes in meningioma.

With respect to MCM6 levels, we found negative correlations in *GMNN*- (geminin) and *EFNA1*- linked DNAm. GMNN was shown to inhibit DNA replication by preventing the incorporation of MCM complexes into prereplication complexes (pre-RC) [79–83]. Further evidence implies its participation in the inhibition of the transcriptional activity of a subset of Hox proteins, linking GMNN to proliferative cell cycle control [83]. *EFNA1* codes for the receptor protein-tyrosine kinase Ephrin A1, which is known to mediate developmental events and is involved in migration, repulsion, and adhesion during neuronal, vascular, and epithelial development in the nervous system and in erythropoiesis. In the context of tumor biology, it was shown to be regulated by hypoxia, and its main function involves angiogenesis and tumor neovascularization [84]. The recruitment of VAV2 is criti-

cal for EFNA1-induced RAC1 GTPase activation and vascular endothelial cell migration and assembly.

When correlating proliferative indices (mitotic, Ki-67, and MCM6) with DNAm levels, we found that the positive and negative correlations exclusively occurred in the hyper- and hypomethylated DMRs differentiating WHO grades. This interrelation hints at a tight epigenetic control of cell proliferation. Thus, either on islands or in open-sea regions, non-concerted modification of methylation does not seem to be the key determinant driving cell division in meningioma. Moreover, fewer overlaps than expected were found between hits with our three chosen markers of proliferation, suggesting their complementing diagnostic value. As expected, because these markers are expressed during different stages of the cell cycle with different magnitudes of expression level, Ki-67 was found to perform well with high-grade and highly proliferative meningiomas, whereas MCM6 expression better delineated low-grade, low-proliferation tumors. Overall, the mitotic index was progressively correlated with demethylation in our proliferation signature.

Due to the limited number of samples in our dataset, the survival statistics were capped to a threshold of associations incompatible with genome-wide multitesting corrections. To overcome this, we first focused on the subset of 310 CpGs with proliferation-dependent DNAm and found that all CpGs were associated with PFS or OS. We next looked at genome-wide results and found that PFS associations with DNAm were the strongest, with unexpectedly numerous hits of high significance in broad genomic regions such a chromosomes 1q and 20. Results with PFS also confirmed the directionality of our DNAm proliferation signature, with progressive demethylation associated with worse progression. The lack of such an observation in OS indicates that no high adequation exists between these proliferation markers and occurrence of death; however, this could also be explained by a smaller number of events and the inclusion of deaths due to another pathology (i.e., not specific to OS). OS may be more strongly associated with independent or indirectly linked loci, such as the *MSI1* regulatory region. This gene encodes an RNA binding protein that could be involved in the maintenance of stem cells in the central nervous system and in cell proliferation. In glioblastoma, it was found to promote the expression of stem cell marker CD44 by impairing miRNA function [85]. In these cells, it was shown to be regulated and stabilized by HuR [86], another RNA binding protein and biomarker of interest in meningiomas [87].

In the present study, given the importance of proliferation in the prognosis of meningiomas, we designed an original approach focusing on the correlation between DNAm and cell proliferation. Proliferation is one of the main mechanisms of tumorigenesis and involves a considerable number of pathways shared by multiple types of tumors. However, epigenetic regulation is only a part of the mechanisms by which cells alter their gene expression. Although we do not have evidence of how DNAm changes affect gene expression, DNAm could potentially regulate downstream genes expression, which may contribute to tumorigenesis or progression. To better dissect the underlying molecular pathways of aberrant cellular growth in meningiomas, integrative and multi-omics studies are needed. As a first step, in this pilot mono-omic study integrating molecular and histological indices, we identified a proliferation signature encompassing hundreds of regulated genes, with several candidates serving as potential predictive and prognostic biomarkers or new therapeutic targets. Most notably, *SMC4, DOK7, PAX6, ARHGDIA, ESRRG, VAV2*, and *OTX1*, as well as the three protocadherin gene clusters PCDHα, β, and γ, may be particularly relevant and deserve further preclinical and clinical investigations.

5. Conclusions

In conclusion, this study highlights the crucial role played by epigenetic mechanisms in shaping dysregulated cellular proliferation in meningioma. It provides molecular biomarkers with potential to revamp the current prognostic classification, as well as new druggable targets, adding therapeutic value to clinical management. The reported findings are novel

and show the additional value of DNAm evaluation in diagnosis and prognostication for patients with meningioma.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers14246227/s1, Table S1. Clinical treatments: preoperative or post-progression radiotherapy, post-progression chemotherapy, post-progression chemotherapy (no neoadjuvant treatment). Table S2. Differentially methylated regions in high grade (>1) vs. low-grade meningiomas. Table S3. GO functional annotations of differential CpGs between high-grade (2 and 3) and low-grade (1) meningiomas. Table S4. Correlation statistics between DNA methylation and mitotic index. Table S5. Correlation statistics between DNA methylation and Ki-67 labeling index. Table S6. Correlation statistics between DNA methylation and MCM6 labeling index. Table S7. DNA methylation proliferation signature in meningiomas. Table S8. Top CpGs associated with progression-free survival. Table S9. Top CpGs associated with overall survival. Figure S1. Principal component analysis of the top 2099 CpGs from the differential analysis between grade 1 and grade 2 + 3 meningiomas. Figure S2. Validation of the proliferative signature on an external methylation dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE200321 (accessed on 23 November 2022); from Daoud et al. [30]). Hierarchical clustering against the 310 CpGs and 60 samples of various meningioma grades and histological types (6 grade 1: 5 meningothelial and 1 transitional; 50 grade 2: 10 chordoid, 27 atypical, and 13 atypical and invasive; 4 grade 3: anaplastic). Raw data from this dataset were processed with the same analytical pipeline as the main dataset presented in this work.

Author Contributions: Conceptualization, S.H. and G.G.; Data curation, C.G., P.R., C.S., S.L., C.G. and G.G.; Formal analysis, S.H. and G.G.; Data analysis, C.G., M.D., S.H., C.S., M.D. and S.L.; Investigation, C.S., M.D., S.L., F.R., C.G., C.P. and G.G.; Methodology, S.H., S.-F.B.-H. and G.G.; Project administration, G.G.; Supervision, G.G.; Writing—original draft, C.S. and S.H.; Writing—review and editing, S.-F.B.-H., S.H. and G.G. All authors have read and agreed to the published version of the manuscript.

Funding: La Région Lorraine, Grand Est—projet de recherche d'intérêt régional 2016; INSERM U1256 NGERE.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the CHRU Nancy Hospital (protocol code 2021PI0154-127; date of approval: 3 November 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE200321. The data generated and presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: The authors thank the team of the Pathology Department of Nancy (CHRU-ICL) for technical support.

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

References

1. Baldi, I.; Engelhardt, J.; Bonnet, C.; Bauchet, L.; Berteaud, E.; Grüber, A.; Loiseau, H. Epidemiology of Meningiomas. *Neurochirurgie* **2018**, *64*, 5–14. [CrossRef] [PubMed]
2. International Agency for Research on Cancer. *WHO Classification of Tumours Editorial Board. Central Nervous System Tumours*, 5th ed.; WHO Classification of Tumours Series; WHO: Geneva, Switzerland, 2021; Volume 6.
3. Wang, N.; Osswald, M. Meningiomas: Overview and New Directions in Therapy. *Semin. Neurol.* **2018**, *38*, 112–120. [CrossRef] [PubMed]
4. Aizer, A.A.; Bi, W.L.; Kandola, M.S.; Lee, E.Q.; Nayak, L.; Rinne, M.L.; Norden, A.D.; Beroukhim, R.; Reardon, D.A.; Wen, P.Y.; et al. Extent of Resection and Overall Survival for Patients with Atypical and Malignant Meningioma: Extent of Resection and Recurrence in Meningioma. *Cancer* **2015**, *121*, 4376–4381. [CrossRef] [PubMed]
5. van Alkemade, H.; de Leau, M.; Dieleman, E.M.T.; Kardaun, J.W.P.F.; van Os, R.; Vandertop, W.P.; van Furth, W.R.; Stalpers, L.J.A. Impaired Survival and Long-Term Neurological Problems in Benign Meningioma. *Neuro-Oncology* **2012**, *14*, 658–666. [CrossRef]

6. Pathology Concordance Levels for Meningioma Classification and Grading in NRG Oncology RTOG Trial 0539. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4799683/ (accessed on 20 February 2022).
7. Goldbrunner, R.; Minniti, G.; Preusser, M.; Jenkinson, M.D.; Sallabanda, K.; Houdart, E.; von Deimling, A.; Stavrinou, P.; Lefranc, F.; Lund-Johansen, M.; et al. EANO Guidelines for the Diagnosis and Treatment of Meningiomas. *Lancet Oncol.* **2016**, *17*, e383–e391. [CrossRef]
8. Institut National du Cancer. Conduite a Tenir Devant des Patients Atteints de Méningiomes de Grade II et III/Synthèse, septembre 2020. *e-cancer.fr*. 2020. Available online: https://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Conduites-a-tenir-devant-des-patients-atteints-d-un-meningiome-de-grade-II-et-III-Synthese (accessed on 1 November 2022).
9. Rouleau, G.A.; Merel, P.; Lutchman, M.; Sanson, M.; Zucman, J.; Marineau, C.; Hoang-Xuan, K.; Demczuk, S.; Desmaze, C.; Plougastel, B. Alteration in a New Gene Encoding a Putative Membrane-Organizing Protein Causes Neuro-Fibromatosis Type 2. *Nature* **1993**, *363*, 515–521. [CrossRef]
10. Bi, W.L.; Zhang, M.; Wu, W.W.; Mei, Y.; Dunn, I.F. Meningioma Genomics: Diagnostic, Prognostic, and Therapeutic Applications. *Front. Surg.* **2016**, *3*, 40. [CrossRef]
11. Bi, W.L.; Mei, Y.; Agarwalla, P.K.; Beroukhim, R.; Dunn, I.F. Genomic and Epigenomic Landscape in Meningioma. *Neurosurg. Clin. N. Am.* **2016**, *27*, 167–179. [CrossRef]
12. Proctor, D.T.; Ramachandran, S.; Lama, S.; Sutherland, G.R. Towards Molecular Classification of Meningioma: Evolving Treatment and Diagnostic Paradigms. *World Neurosurg.* **2018**, *119*, 366–373. [CrossRef]
13. Galani, V.; Lampri, E.; Varouktsi, A.; Alexiou, G.; Mitselou, A.; Kyritsis, A.P. Genetic and Epigenetic Alterations in Meningiomas. *Clin. Neurol. Neurosurg.* **2017**, *158*, 119–125. [CrossRef]
14. Yakubov, E.; Ghoochani, A.; Buslei, R.; Buchfelder, M.; Eyüpoglu, I.Y.; Savaskan, N. Hidden Association of Cowden Syndrome, PTEN Mutation and Meningioma Frequency. *Oncoscience* **2016**, *3*, 149–155. [CrossRef] [PubMed]
15. Alexiou, G.A.; Voulgaris, S. The Role of the PTEN Gene in Malignant Gliomas. *Neurol. Neurochir. Pol.* **2010**, *44*, 80–86. [CrossRef]
16. Bi, W.L.; Greenwald, N.F.; Abedalthagafi, M.; Wala, J.; Gibson, W.J.; Agarwalla, P.K.; Horowitz, P.; Schumacher, S.E.; Esaulova, E.; Mei, Y.; et al. Genomic Landscape of High-Grade Meningiomas. *NPJ Genomic Med.* **2017**, *2*, 15. [CrossRef] [PubMed]
17. Brastianos, P.K.; Horowitz, P.M.; Santagata, S.; Jones, R.T.; McKenna, A.; Getz, G.; Ligon, K.L.; Palescandolo, E.; Van Hummelen, P.; Ducar, M.D.; et al. Genomic Sequencing of Meningiomas Identifies Oncogenic SMO and AKT1 Mutations. *Nat. Genet.* **2013**, *45*, 285–289. [CrossRef] [PubMed]
18. Lekanne Deprez, R.H.; Riegman, P.H.; van Drunen, E.; Warringa, U.L.; Groen, N.A.; Stefanko, S.Z.; Koper, J.W.; Avezaat, C.J.; Mulder, P.G.; Zwarthoff, E.C. Cytogenetic, Molecular Genetic and Pathological Analyses in 126 Meningiomas. *J. Neuropathol. Exp. Neurol.* **1995**, *54*, 224–235. [CrossRef] [PubMed]
19. Cordova, C.; Kurz, S.C. Advances in Molecular Classification and Therapeutic Opportunities in Meningiomas. *Curr. Oncol. Rep.* **2020**, *22*, 84. [CrossRef] [PubMed]
20. Suvà, M.L.; Louis, D.N. Next-Generation Molecular Genetics of Brain Tumours. *Curr. Opin. Neurol.* **2013**, *26*, 681–687. [CrossRef]
21. Katz, L.M.; Hielscher, T.; Liechty, B.; Silverman, J.; Zagzag, D.; Sen, R.; Wu, P.; Golfinos, J.G.; Reuss, D.; Neidert, M.C.; et al. Loss of Histone H3K27me3 Identifies a Subset of Meningiomas with Increased Risk of Recurrence. *Acta Neuropathol. (Berl.)* **2018**, *135*, 955–963. [CrossRef]
22. Paramasivam, N.; Hübschmann, D.; Toprak, U.H.; Ishaque, N.; Neidert, M.; Schrimpf, D.; Stichel, D.; Reuss, D.; Sievers, P.; Reinhardt, A.; et al. Mutational Patterns and Regulatory Networks in Epigenetic Subgroups of Meningioma. *Acta Neuropathol. (Berl.)* **2019**, *138*, 295–308. [CrossRef]
23. Sahm, F.; Schrimpf, D.; Stichel, D.; Jones, D.T.W.; Hielscher, T.; Schefzyk, S.; Okonechnikov, K.; Koelsche, C.; Reuss, D.E.; Capper, D.; et al. DNA Methylation-Based Classification and Grading System for Meningioma: A Multicentre, Retrospective Analysis. *Lancet Oncol.* **2017**, *18*, 682–694. [CrossRef]
24. Nassiri, F.; Liu, J.; Patil, V.; Mamatjan, Y.; Wang, J.Z.; Hugh-White, R.; Macklin, A.M.; Khan, S.; Singh, O.; Karimi, S.; et al. A Clinically Applicable Integrative Molecular Classification of Meningiomas. *Nature* **2021**, *597*, 119–125. [CrossRef]
25. Gauchotte, G.; Vigouroux, C.; Rech, F.; Battaglia-Hsu, S.-F.; Soudant, M.; Pinelli, C.; Civit, T.; Taillandier, L.; Vignaud, J.-M.; Bressenot, A. Expression of Minichromosome Maintenance MCM6 Protein in Meningiomas Is Strongly Correlated with Histologic Grade and Clinical Outcome. *Am. J. Surg. Pathol.* **2012**, *36*, 283–291. [CrossRef]
26. Zgheib, R.; Battaglia-Hsu, S.-F.; Hergalant, S.; Quéré, M.; Alberto, J.-M.; Chéry, C.; Rouyer, P.; Gauchotte, G.; Guéant, J.-L.; Namour, F. Folate Can Promote the Methionine-Dependent Reprogramming of Glioblastoma Cells towards Pluripotency. *Cell Death Dis.* **2019**, *10*, 596. [CrossRef]
27. Fortin, J.-P.; Labbe, A.; Lemire, M.; Zanke, B.W.; Hudson, T.J.; Fertig, E.J.; Greenwood, C.M.; Hansen, K.D. Functional Normalization of 450k Methylation Array Data Improves Replication in Large Cancer Studies. *Genome Biol.* **2014**, *15*, 503. [CrossRef]
28. Phipson, B.; Maksimovic, J.; Oshlack, A. MissMethyl: An R Package for Analyzing Data from Illumina's HumanMethylation450 Platform. *Bioinforma. Oxf. Engl.* **2016**, *32*, 286–288. [CrossRef]
29. Capper, D.; Jones, D.T.W.; Sill, M.; Hovestadt, V.; Schrimpf, D.; Sturm, D.; Koelsche, C.; Sahm, F.; Chavez, L.; Reuss, D.E.; et al. DNA Methylation-Based Classification of Central Nervous System Tumours. *Nature* **2018**, *555*, 469–474. [CrossRef]
30. Daoud, E.V.; Zhu, K.; Mickey, B.; Mohamed, H.; Wen, M.; Delorenzo, M.; Tran, I.; Serrano, J.; Hatanpaa, K.J.; Raisanen, J.M.; et al. Epigenetic and Genomic Profiling of Chordoid Meningioma: Implications for Clinical Management. *Acta Neuropathol. Commun.* **2022**, *10*, 56. [CrossRef]

31. Esteller, M. Epigenetic Gene Silencing in Cancer: The DNA Hypermethylome. *Hum. Mol. Genet.* **2007**, *16*, R50–R59. [CrossRef]
32. Mack, S.C.; Hubert, C.G.; Miller, T.E.; Taylor, M.D.; Rich, J.N. An Epigenetic Gateway to Brain Tumor Cell Identity. *Nat. Neurosci.* **2016**, *19*, 10–19. [CrossRef]
33. Rechache, N.S.; Wang, Y.; Stevenson, H.S.; Killian, J.K.; Edelman, D.C.; Merino, M.; Zhang, L.; Nilubol, N.; Stratakis, C.A.; Meltzer, P.S.; et al. DNA Methylation Profiling Identifies Global Methylation Differences and Markers of Adrenocortical Tumors. *J. Clin. Endocrinol. Metab.* **2012**, *97*, E1004–E1013. [CrossRef]
34. Song, M.-A.; Tiirikainen, M.; Kwee, S.; Okimoto, G.; Yu, H.; Wong, L.L. Elucidating the Landscape of Aberrant DNA Methylation in Hepatocellular Carcinoma. *PLoS ONE* **2013**, *8*, e55761. [CrossRef]
35. Rodriguez, J.; Frigola, J.; Vendrell, E.; Risques, R.-A.; Fraga, M.F.; Morales, C.; Moreno, V.; Esteller, M.; Capellà, G.; Ribas, M.; et al. Chromosomal Instability Correlates with Genome-Wide DNA Demethylation in Human Primary Colorectal Cancers. *Cancer Res.* **2006**, *66*, 8462–9468. [CrossRef]
36. Nomura, M.; Saito, K.; Aihara, K.; Nagae, G.; Yamamoto, S.; Tatsuno, K.; Ueda, H.; Fukuda, S.; Umeda, T.; Tanaka, S.; et al. DNA Demethylation Is Associated with Malignant Progression of Lower-Grade Gliomas. *Sci. Rep.* **2019**, *9*, 1903. [CrossRef]
37. Lietz, C.E.; Newman, E.T.; Kelly, A.D.; Xiang, D.H.; Zhang, Z.; Luscko, C.A.; Lozano-Calderon, S.A.; Ebb, D.H.; Raskin, K.A.; Cote, G.M.; et al. Genome-Wide DNA Methylation Patterns Reveal Clinically Relevant Predictive and Prognostic Subtypes in Human Osteosarcoma. *Commun. Biol.* **2022**, *5*, 213. [CrossRef]
38. Cancino, G.I.; Yiu, A.P.; Fatt, M.P.; Dugani, C.B.; Flores, E.R.; Frankland, P.W.; Josselyn, S.A.; Miller, F.D.; Kaplan, D.R. P63 Regulates Adult Neural Precursor and Newly Born Neuron Survival to Control Hippocampal-Dependent Behavior. *J. Neurosci. Off. J. Soc. Neurosci.* **2013**, *33*, 12569–12585. [CrossRef]
39. Dugani, C.B.; Paquin, A.; Fujitani, M.; Kaplan, D.R.; Miller, F.D. P63 Antagonizes P53 to Promote the Survival of Embryonic Neural Precursor Cells. *J. Neurosci. Off. J. Soc. Neurosci.* **2009**, *29*, 6710–6721. [CrossRef]
40. Feng, X.-D.; Song, Q.; Li, C.-W.; Chen, J.; Tang, H.-M.; Peng, Z.-H.; Wang, X.-C. Structural Maintenance of Chromosomes 4 Is a Predictor of Survival and a Novel Therapeutic Target in Colorectal Cancer. *Asian Pac. J. Cancer Prev. APJCP* **2014**, *15*, 9459–9465. [CrossRef]
41. Chen, Y.; Huang, F.; Deng, L.; Yuan, X.; Tao, Q.; Wang, T.; Li, D.; Fan, Y.; Peng, Q.; Tang, D. HIF-1-MiR-219-SMC4 Regulatory Pathway Promoting Proliferation and Migration of HCC under Hypoxic Condition. *BioMed Res. Int.* **2019**, *2019*, 8983704. [CrossRef]
42. Yan, Y.; Liu, C.; Zhang, J.; Li, W.; Yin, X.; Dong, L.; Pang, S.; Li, X. SMC4 Knockdown Inhibits Malignant Biological Behaviors of Endometrial Cancer Cells by Regulation of FoxO1 Activity. *Arch. Biochem. Biophys.* **2021**, *712*, 109026. [CrossRef]
43. Ma, R.-M.; Yang, F.; Huang, D.-P.; Zheng, M.; Wang, Y.-L. The Prognostic Value of the Expression of SMC4 MRNA in Breast Cancer. *Dis. Markers* **2019**, *2019*, 2183057. [CrossRef]
44. Zhou, B.; Chen, H.; Wei, D.; Kuang, Y.; Zhao, X.; Li, G.; Xie, J.; Chen, P. A Novel MiR-219-SMC4-JAK2/Stat3 Regulatory Pathway in Human Hepatocellular Carcinoma. *J. Exp. Clin. Cancer Res. CR* **2014**, *33*, 55. [CrossRef] [PubMed]
45. Jiang, L.; Zhou, J.; Zhong, D.; Zhou, Y.; Zhang, W.; Wu, W.; Zhao, Z.; Wang, W.; Xu, W.; He, L.; et al. Overexpression of SMC4 Activates TGFβ/Smad Signaling and Promotes Aggressive Phenotype in Glioma Cells. *Oncogenesis* **2017**, *6*, e301. [CrossRef] [PubMed]
46. You, A.; Rao, G.; Wang, J.; Li, J.; Zhang, Y.; Gu, J.; Ge, X.; Zhang, K.; Gao, X.; Wu, X.; et al. MiR-433-3p Restrains the Proliferation, Migration and Invasion of Glioma Cells via Targeting SMC4. *Brain Res.* **2021**, *1767*, 147563. [CrossRef]
47. Wang, Y.; Wu, Z. The Clinical Significance and Transcription Regulation of a DNA Damage Repair Gene, SMC4, in Low-Grade Glioma via Integrated Bioinformatic Analysis. *Front. Oncol.* **2021**, *11*, 761693. [CrossRef] [PubMed]
48. Chang, J.Y.; Hu, Y.; Siegel, E.; Stanley, L.; Zhou, Y.-H. PAX6 Increases Glioma Cell Susceptibility to Detachment and Oxidative Stress. *J. Neurooncol.* **2007**, *84*, 9–19. [CrossRef] [PubMed]
49. Pavlakis, E.; Tonchev, A.B.; Kaprelyan, A.; Enchev, Y.; Stoykova, A. Interaction between Transcription Factors PAX6/PAX6-5a and Specific Members of MiR-183-96-182 Cluster, May Contribute to Glioma Progression in Glioblastoma Cell Lines. *Oncol. Rep.* **2017**, *37*, 1579–1592. [CrossRef]
50. Zhou, Y.-H.; Wu, X.; Tan, F.; Shi, Y.-X.; Glass, T.; Liu, T.J.; Wathen, K.; Hess, K.R.; Gumin, J.; Lang, F.; et al. PAX6 Suppresses Growth of Human Glioblastoma Cells. *J. Neurooncol.* **2005**, *71*, 223–229. [CrossRef] [PubMed]
51. Chen, W.V.; Maniatis, T. Clustered Protocadherins. *Dev. Camb. Engl.* **2013**, *140*, 3297–3302. [CrossRef]
52. Tasic, B.; Nabholz, C.E.; Baldwin, K.K.; Kim, Y.; Rueckert, E.H.; Ribich, S.A.; Cramer, P.; Wu, Q.; Axel, R.; Maniatis, T. Promoter Choice Determines Splice Site Selection in Protocadherin Alpha and Gamma Pre-MRNA Splicing. *Mol. Cell* **2002**, *10*, 21–33. [CrossRef]
53. Vega-Benedetti, A.F.; Loi, E.; Moi, L.; Blois, S.; Fadda, A.; Antonelli, M.; Arcella, A.; Badiali, M.; Giangaspero, F.; Morra, I.; et al. Clustered Protocadherins Methylation Alterations in Cancer. *Clin. Epigenet.* **2019**, *11*, 100. [CrossRef]
54. Jun, P.; Hong, C.; Lal, A.; Wong, J.M.; McDermott, M.W.; Bollen, A.W.; Plass, C.; Held, W.A.; Smiraglia, D.J.; Costello, J.F. Epigenetic Silencing of the Kinase Tumor Suppressor WNK2 Is Tumor-Type and Tumor-Grade Specific. *Neuro-Oncology* **2009**, *11*, 414–422. [CrossRef]
55. He, S.; Pham, M.H.; Pease, M.; Zada, G.; Giannotta, S.L.; Wang, K.; Mack, W.J. A Review of Epigenetic and Gene Expression Alterations Associated with Intracranial Meningiomas. *Neurosurg. Focus* **2013**, *35*, E5. [CrossRef] [PubMed]

56. Lu, W.; Wang, X.; Liu, J.; He, Y.; Liang, Z.; Xia, Z.; Cai, Y.; Zhou, L.; Zhu, H.; Liang, S. Downregulation of ARHGDIA Contributes to Human Glioma Progression through Activation of Rho GTPase Signaling Pathway. *Tumour Biol.* **2016**, *37*, 15783–15793. [CrossRef]
57. Wang, H.; Wang, B.; Liao, Q.; An, H.; Li, W.; Jin, X.; Cui, S.; Zhao, L. Overexpression of RhoGDI, a Novel Predictor of Distant Metastasis, Promotes Cell Proliferation and Migration in Hepatocellular Carcinoma. *FEBS Lett.* **2014**, *588*, 503–508. [CrossRef]
58. Yamashita, T.; Okamura, T.; Nagano, K.; Imai, S.; Abe, Y.; Nabeshi, H.; Yoshikawa, T.; Yoshioka, Y.; Kamada, H.; Tsutsumi, Y.; et al. Rho GDP-Dissociation Inhibitor Alpha Is Associated with Cancer Metastasis in Colon and Prostate Cancer. *Pharm.-Int. J. Pharm. Sci.* **2012**, *67*, 253–255.
59. Bozza, W.P.; Zhang, Y.; Hallett, K.; Rosado, L.A.R.; Zhang, B. RhoGDI Deficiency Induces Constitutive Activation of Rho GTPases and COX-2 Pathways in Association with Breast Cancer Progression. *Oncotarget* **2015**, *6*, 32723–32736. [CrossRef]
60. Harding, M.A.; Theodorescu, D. RhoGDI Signaling Provides Targets for Cancer Therapy. *Eur. J. Cancer Oxf. Engl. 1990* **2010**, *46*, 1252–1259. [CrossRef]
61. Shirkavand, A.; Boroujeni, Z.N.; Aleyasin, S.A. Examination of Methylation Changes of VIM, CXCR4, DOK7, and SPDEF Genes in Peripheral Blood DNA in Breast Cancer Patients. *Indian J. Cancer* **2018**, *55*, 366–371. [CrossRef]
62. Heyn, H.; Carmona, F.J.; Gomez, A.; Ferreira, H.J.; Bell, J.T.; Sayols, S.; Ward, K.; Stefansson, O.A.; Moran, S.; Sandoval, J.; et al. DNA Methylation Profiling in Breast Cancer Discordant Identical Twins Identifies DOK7 as Novel Epigenetic Biomarker. *Carcinogenesis* **2013**, *34*, 102–108. [CrossRef]
63. Zhao, H.; Chen, G.; Ye, L.; Yu, H.; Li, S.; Jiang, W.G. DOK7V1 Influences the Malignant Phenotype of Lung Cancer Cells through PI3K/AKT/MTOR and FAK/Paxillin Signaling Pathways. *Int. J. Oncol.* **2019**, *54*, 381–389. [CrossRef]
64. Chen, G.; Yu, H.; Satherley, L.; Zabkiewicz, C.; Resaul, J.; Zhao, H.; Mu, H.; Zhi, X.; He, J.; Ye, L.; et al. The Downstream of Tyrosine Kinase 7 Is Reduced in Lung Cancer and Is Associated with Poor Survival of Patients with Lung Cancer. *Oncol. Rep.* **2017**, *37*, 2695–2701. [CrossRef] [PubMed]
65. Hua, C.-D.; Bian, E.-B.; Chen, E.-F.; Yang, Z.-H.; Tang, F.; Wang, H.-L.; Zhao, B. Repression of Dok7 Expression Mediated by DNMT1 Promotes Glioma Cells Proliferation. *Biomed. Pharmacother. Biomedecine Pharmacother.* **2018**, *106*, 678–685. [CrossRef] [PubMed]
66. Kang, M.-H.; Choi, H.; Oshima, M.; Cheong, J.-H.; Kim, S.; Lee, J.H.; Park, Y.S.; Choi, H.-S.; Kweon, M.-N.; Pack, C.-G.; et al. Estrogen-Related Receptor Gamma Functions as a Tumor Suppressor in Gastric Cancer. *Nat. Commun.* **2018**, *9*, 1920. [CrossRef] [PubMed]
67. Shen, Z.; Hu, Y.; Zhou, C.; Yuan, J.; Xu, J.; Hao, W.; Deng, H.; Ye, D. ESRRG Promoter Hypermethylation as a Diagnostic and Prognostic Biomarker in Laryngeal Squamous Cell Carcinoma. *J. Clin. Lab. Anal.* **2019**, *33*, e22899. [CrossRef]
68. Acampora, D.; Mazan, S.; Avantaggiato, V.; Barone, P.; Tuorto, F.; Lallemand, Y.; Brûlet, P.; Simeone, A. Epilepsy and Brain Abnormalities in Mice Lacking the Otx1 Gene. *Nat. Genet.* **1996**, *14*, 218–222. [CrossRef]
69. Huang, B.; Li, X.; Tu, X.; Zhao, W.; Zhu, D.; Feng, Y.; Si, X.; Chen, J.-G. OTX1 Regulates Cell Cycle Progression of Neural Progenitors in the Developing Cerebral Cortex. *J. Biol. Chem.* **2018**, *293*, 2137–2148. [CrossRef]
70. Jiang, L.; Zuo, Z.; Lin, J.; Yang, C. Orthodenticle Homeobox OTX1 Is a Potential Prognostic Biomarker for Bladder Cancer. *Bioengineered* **2021**, *12*, 6559–6571. [CrossRef]
71. Tu, X.-P.; Li, H.; Chen, L.-S.; Luo, X.-N.; Lu, Z.-M.; Zhang, S.-Y.; Chen, S.-H. OTX1 Exerts an Oncogenic Role and Is Negatively Regulated by MiR129-5p in Laryngeal Squamous Cell Carcinoma. *BMC Cancer* **2020**, *20*, 794. [CrossRef]
72. Jiang, Y.; Prabakaran, I.; Wan, F.; Mitra, N.; Furstenau, D.K.; Hung, R.K.; Cao, S.; Zhang, P.J.; Fraker, D.L.; Guvakova, M.A. Vav2 Protein Overexpression Marks and May Predict the Aggressive Subtype of Ductal Carcinoma in Situ. *Biomark. Res.* **2014**, *2*, 22. [CrossRef]
73. Citterio, C.; Menacho-Márquez, M.; García-Escudero, R.; Larive, R.M.; Barreiro, O.; Sánchez-Madrid, F.; Paramio, J.M.; Bustelo, X.R. The Rho Exchange Factors Vav2 and Vav3 Control a Lung Metastasis-Specific Transcriptional Program in Breast Cancer Cells. *Sci. Signal.* **2012**, *5*, ra71. [CrossRef]
74. Tan, B.-B.; Li, Y.; Fan, L.-Q.; Zhao, Q.; Liu, Q.-W.; Liu, Y.; Wang, D.; Jia, N. Upregulated Vav2 in Gastric Cancer Tissues Promotes Tumor Invasion and Metastasis. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* **2017**, *39*, 1010428317698392. [CrossRef]
75. Lorenzo-Martín, L.F.; Fernández-Parejo, N.; Menacho-Márquez, M.; Rodríguez-Fdez, S.; Robles-Valero, J.; Zumalave, S.; Fabbiano, S.; Pascual, G.; García-Pedrero, J.M.; Abad, A.; et al. VAV2 Signaling Promotes Regenerative Proliferation in Both Cutaneous and Head and Neck Squamous Cell Carcinoma. *Nat. Commun.* **2020**, *11*, 4788. [CrossRef]
76. Liu, W.; Miao, C.; Zhang, S.; Liu, Y.; Niu, X.; Xi, Y.; Guo, W.; Chu, J.; Lin, A.; Liu, H.; et al. VAV2 Is Required for DNA Repair and Implicated in Cancer Radiotherapy Resistance. *Signal Transduct. Target. Ther.* **2021**, *6*, 322. [CrossRef]
77. Jones, D.T.W.; Jäger, N.; Kool, M.; Zichner, T.; Hutter, B.; Sultan, M.; Cho, Y.-J.; Pugh, T.J.; Hovestadt, V.; Stütz, A.M.; et al. Dissecting the Genomic Complexity Underlying Medulloblastoma. *Nature* **2012**, *488*, 100–105. [CrossRef]
78. Nakahara, Y.; Shiraishi, T.; Okamoto, H.; Mineta, T.; Oishi, T.; Sasaki, K.; Tabuchi, K. Detrended Fluctuation Analysis of Genome-Wide Copy Number Profiles of Glioblastomas Using Array-Based Comparative Genomic Hybridization. *Neuro-Oncol.* **2004**, *6*, 281–289. [CrossRef]
79. McGarry, T.J.; Kirschner, M.W. Geminin, an Inhibitor of DNA Replication, Is Degraded during Mitosis. *Cell* **1998**, *93*, 1043–1053. [CrossRef]
80. Miotto, B.; Struhl, K. HBO1 Histone Acetylase Activity Is Essential for DNA Replication Licensing and Inhibited by Geminin. *Mol. Cell* **2010**, *37*, 57–66. [CrossRef]

81. Sugimoto, N.; Tatsumi, Y.; Tsurumi, T.; Matsukage, A.; Kiyono, T.; Nishitani, H.; Fujita, M. Cdt1 Phosphorylation by Cyclin A-Dependent Kinases Negatively Regulates Its Function without Affecting Geminin Binding. *J. Biol. Chem.* **2004**, *279*, 19691–19697. [CrossRef]
82. Caillat, C.; Pefani, D.-E.; Gillespie, P.J.; Taraviras, S.; Blow, J.J.; Lygerou, Z.; Perrakis, A. The Geminin and Idas Coiled Coils Preferentially Form a Heterodimer That Inhibits Geminin Function in DNA Replication Licensing. *J. Biol. Chem.* **2013**, *288*, 31624–31634. [CrossRef]
83. Zhou, B.; Liu, C.; Xu, Z.; Zhu, G. Structural Basis for Homeodomain Recognition by the Cell-Cycle Regulator Geminin. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8931–8936. [CrossRef]
84. Hao, Y.; Li, G. Role of EFNA1 in Tumorigenesis and Prospects for Cancer Therapy. *Biomed. Pharmacother.* **2020**, *130*, 110567. [CrossRef]
85. Pötschke, R.; Haase, J.; Glaß, M.; Simmermacher, S.; Misiak, C.; Penalva, L.O.F.; Kühnöl, C.D.; Hüttelmaier, S. MSI1 Promotes the Expression of the GBM Stem Cell Marker CD44 by Impairing MiRNA-Dependent Degradation. *Cancers* **2020**, *12*, 3654. [CrossRef]
86. Vo, D.T.; Abdelmohsen, K.; Martindale, J.L.; Qiao, M.; Tominaga, K.; Burton, T.L.; Gelfond, J.A.L.; Brenner, A.J.; Patel, V.; Trageser, D.; et al. The Oncogenic RNA-Binding Protein Musashi1 Is Regulated by HuR via MRNA Translation and Stability in Glioblastoma Cells. *Mol. Cancer Res. MCR* **2012**, *10*, 143–155. [CrossRef]
87. Gauchotte, G.; Hergalant, S.; Vigouroux, C.; Casse, J.-M.; Houlgatte, R.; Kaoma, T.; Helle, D.; Brochin, L.; Rech, F.; Peyre, M.; et al. Cytoplasmic Overexpression of RNA-Binding Protein HuR Is a Marker of Poor Prognosis in Meningioma, and HuR Knockdown Decreases Meningioma Cell Growth and Resistance to Hypoxia. *J. Pathol.* **2017**, *242*, 421–434. [CrossRef]

3. MicroRNAs miR-16 and miR-519 Control Meningioma Cell Proliferation via Overlapping Transcriptomic Programs Shared with the RNA-Binding Protein HuR

Contexte

Nous abordons une étude récemment publiée mais concluant un projet engagé en 2014 et portant sur le rôle de HuR (*Human antigen R*) dans la prolifération cellulaire des méningiomes (*194*). HuR est une protéine qui se lie aux ARNs en 3'UTR pour en assurer le transport et la stabilité, et qui donc les régule post-transcriptionnellement. Ses ARNs cibles sont pléthore, et elle et son gène (*ELAVL1*) sont exprimés dans une majorité de tissus et de conditions. L'implication de HuR était décrite dans de nombreux carcinomes, y compris les cancers cérébraux de haut grade, mais pas les méningiomes et les tumeurs non cancéreuses avant ces travaux. Ici nous nous intéressons plus précisément à deux microARNs, miR-16 et miR-519, et examinons leur rôle dans la progression et la prolifération cellulaire dans des méningiomes de grade 1 (bénins) et 2 (atypiques) ainsi que dans des lignées cellulaires de méningiomes, agressifs ou non. Dans ces systèmes, nous élucidons notamment le programme transcriptionnel sous le contrôle de ces microARNs et le relions à celui observé chez HuR. Pourquoi miR-16 et miR-519 ? Le premier est bien connu pour réguler, entre autres, des gènes suppresseurs de tumeurs, la balance apoptotique, ou *TP53*, et le second cible *ELAVL1*, de façon clairement démontrée, dans plusieurs tissus et tumeurs (*195, 196*).

Points clés

- Mise en évidence du rôle suppresseur de tumeur des deux miRs chez 80 patients (60 méningiomes et 20 tissus sains), via un contrôle transcriptionnel du cycle cellulaire, notamment du complexe pré-réplicatif, et du neurodéveloppement. *A la mode* de HuR.
- Description d'une signature transcriptomique robuste et reproductible, partagée entre cibles de miR-16, de miR-519 et de HuR, proposant des biomarqueurs originaux de progression et de récurrence dans ces tumeurs, avec potentiel prédictif dans les bégnines.
- Validation, à chaque étape du travail, des résultats avec des données publiques de transcriptome, de miRnome (l'ensemble des microARNs), en mono-omique ou multi-omique, et couvrant un large spectre de conditions (lignées, patients, grades, contrôles, récurrence, statut mutationnel, progression).

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 02 August 2023
DOI 10.3389/fonc.2023.1158773

OPEN ACCESS

EDITED BY
Yuriy Gusev,
Georgetown University, United States

REVIEWED BY
Sabrina Battista,
National Research Council (CNR), Italy
Sharon K. Michelhaugh,
Virginia Tech, United States

*CORRESPONDENCE
Sébastien Hergalant
sebastien.hergalant@univ-lorraine.fr
Guillaume Gauchotte
g.gauchotte@chru-nancy.fr

†These authors have contributed equally to this work and share last authorship

RECEIVED 04 February 2023
ACCEPTED 14 July 2023
PUBLISHED 02 August 2023

CITATION
Hergalant S, Casse J-M, Oussalah A, Houlgatte R, Helle D, Rech F, Vallar L, Guéant J-L, Vignaud J-M, Battaglia-Hsu S-F and Gauchotte G (2023) MicroRNAs miR-16 and miR-519 control meningioma cell proliferation *via* overlapping transcriptomic programs shared with the RNA-binding protein HuR. *Front. Oncol.* 13:1158773. doi: 10.3389/fonc.2023.1158773

MicroRNAs miR-16 and miR-519 control meningioma cell proliferation *via* overlapping transcriptomic programs shared with the RNA-binding protein HuR

Sébastien Hergalant[1*], Jean-Matthieu Casse[1], Abderrahim Oussalah[1,2,3], Rémi Houlgatte[1], Déborah Helle[1], Fabien Rech[4,5], Laurent Vallar[6], Jean-Louis Guéant[1,2,3], Jean-Michel Vignaud[1,7,8], Shyue-Fang Battaglia-Hsu[2,3,5†] and Guillaume Gauchotte[1,7,8*†]

[1]INSERM, U1256, NGERE – Nutrition, Genetics, and Environmental Risk Exposure, Faculty of Medicine of Nancy, University of Lorraine, Vandoeuvre-lès-Nancy, France, [2]Department of Molecular Medicine and Personalized Therapeutics, University Hospital of Nancy (CHRU), Vandoeuvre-lès-Nancy, France, [3]Department of Biochemistry, Molecular Biology, Nutrition, and Metabolism, University Hospital of Nancy (CHRU), Vandoeuvre-lès-Nancy, France, [4]Department of Neurosurgery, University Hospital of Nancy (CHRU), Nancy, France, [5]CNRS, UMR7039, CRAN - Centre de Recherche en Automatique de Nancy, Université de Lorraine, Vandoeuvre-lès-Nancy, France, [6]Genomics and Proteomics, Department of Oncology, Luxembourg Institute of Health, Luxembourg, Luxembourg, [7]Department of Biopathology Institut De Cancérologie de Lorraine (CHRU-ICL), University Hospital of Nancy (CHRU), Nancy, France, [8]Centre de Ressources Biologiques BB-0033-00035, University Hospital of Nancy (CHRU), Nancy, France

Introduction: Meningiomas are the most common type of primary central nervous system tumors. In about 80% cases, these tumors are benign and grow very slowly, but the remainder 20% can unlock higher proliferation rates and become malignant. In this study we examined two miRs, miR-16 and miR-519, and evaluated their role in tumorigenesis and cell growth in human meningioma.

Methods: A cohort of 60 intracranial grade 1 and grade 2 human meningioma plus 20 healthy meningeal tissues was used to quantify miR-16 and miR-519 expressions. Cell growth and dose-response assays were performed in two human meningioma cell lines, Ben-Men-1 (benign) and IOMM-Lee (aggressive). Transcriptomes of IOMM-lee cells were measured after both miR-mimics transfection, followed by integrative bioinformatics to expand on available data.

Results: In tumoral tissues, we detected decreased levels of miR-16 and miR-519 when compared with arachnoid cells of healthy patients (miR-16: P=8.7e-04; miR-519: P=3.5e-07). When individually overexpressing these miRs in Ben-Men-1 and IOMM-Lee, we observed that each showed reduced growth (P<0.001). In IOMM-Lee cell transcriptomes, downregulated genes, among which ELAVL1/HuR (miR-16: P=6.1e-06; miR-519:P=9.38e-03), were linked to biological

processes such as mitotic cell cycle regulation, pre-replicative complex, and brain development (FDR<1e-05). Additionally, we uncovered a specific transcriptomic signature of miR-16/miR-519-dysregulated genes which was highly enriched in HuR targets (>6-fold; 79.6% of target genes).

Discussion: These results were confirmed on several public transcriptomic and microRNA datasets of human meningiomas, hinting that the putative tumor suppressor effect of these miRs is mediated, at least in part, via HuR direct or indirect inhibition.

KEYWORDS

miR-16, miR-519, microRNA, HuR, proliferation, tumorigenesis, meningioma, transcriptomics

1 Introduction

Meningiomas are the most common type of primary tumors of the central nervous system in adult. For brain meningioma alone, the annual incidence rate ranges from 1.3/100 000 to 7.8/100 000, a trend now under constant acceleration (1). WHO (World Health Organization) stratifies meningiomas into 3 grades of malignancy and 15 subtypes. These tumors originate from arachnoid cap cells forming one of the layers of the protective meninges, along with the dura and the pia mater, a membrane covering the brain and spinal cord. Regardless of grade, most patients undergo surgery if deemed adequate, but adjuvant therapy is not systematic because, to date, there is none validated for meningioma treatment (2). Plus, for grades 2 and 3, conformational radiotherapy is recommended after surgery (3). Thus, investigating the disease at a molecular level is an important issue as it may unlock new diagnostic and therapeutic options. However, it is only recently that meningioma genomic and epigenomic landscapes were described with enough accuracy to be helpful in precision medicine (2, 4, 5). By paving the way toward refined and clinically relevant classification systems (6–8), by fueling biomarker and drug target discoveries (9–12), these omic studies opened new areas of exploration to decipher the molecular characteristics of various meningioma subgroups (13–15).

MicroRNAs (miRs) are a class of 21–23 nucleotide-long non-coding RNA molecules involved in gene silencing and can modify gene expression at post-transcriptional level. They are of vital importance for the maintenance of balanced biological processes like cell proliferation and differentiation, metabolism, signaling, and death (16). Indeed, tissue-specific dysregulation of these miRs can trigger pathological consequences, and cancer. Interventions targeting abnormal miR expression account for effective treatment strategies for diverse diseases (17), offering alternatives with improved clinical outcomes (18). MiR-16, for example, constitutes a potentially useful biomarker for early detection in cancer diagnosis (19, 20) and an attractive therapeutic target (21, 22). Both miR-16 and miR-519 are dysregulated in several types of tumors (19, 20, 23–29), including glioma (30–32) and glioblastoma (33, 34). In human meningioma, however, neither the *in vivo* expression of these two miRs, nor their *in vitro* use as potential tumor suppressors have been evaluated, and to this day, few works have examined miR expression profiling in tumor tissue or serum of meningioma patients (35–40).

HuR (*ELAVL1*), a ubiquitously expressed RNA-binding protein involved in mRNA processing, stability, and transport, accounts for another promising drug target in anticancer treatment (41, 42). In meningioma, we previously described HuR overexpression as a marker of poor prognosis (43). Because miR-16 and miR-519 may negatively regulate HuR directly or indirectly (23, 25, 26, 44–46), we asked if restoration of these miRs in meningioma cells might reduce HuR and have anti-proliferative consequence. Therefore, the aims of this study were to determine whether miR-16 and miR-519 are differentially expressed in human meningioma relative to normal meningeal tissues, and to evaluate the effects of their overexpression on cell proliferation in human meningioma cell lines. Relative to healthy arachnoid tissues, we report miR-16 and miR-519 reduced levels in human meningiomas. Additionally, we explored the transcriptome-wide effects of miR-16 and miR-519 overexpression in high-grade meningioma IOMM-Lee cells and investigated the way these two miRs altered the expression of HuR and its target genes. Compared with our previous results on HuR transcriptomics and other human meningioma datasets of available miR profiling and transcriptome studies, these findings suggest that the putative tumor suppressor effect of miR-16 and miR-519 is mediated, at least in part, *via* HuR.

2 Materials and methods

2.1 Population and clinicopathological data

Sixty consecutive cases of intracranial grade 1 and grade 2 meningioma tissues were retrospectively retrieved from the Department of Pathology of the University Hospital of Nancy (institutional review board DC2008-459), and reviewed to confirm their initial diagnosis and grading according to the 2016 WHO classification criteria (47). Twenty samples of normal meningeal tissue were studied, including 10 samples of arachnoid membrane collected during autopsies and 10 surgical samples of

non-neoplastic dura mater. Both meningioma and control tissues were fixed in formalin for 24 h.

2.2 Quantification of the relative expression levels of miR-16 and miR-519

In all tissue samples, the relative expression levels of miR-16 and miR-519 were determined *via* quantitative reverse transcription-polymerase chain reaction (qRT-PCR). Paraffin-embedded tumors and normal tissues were dissected from tissue blocks. Total RNA extraction was performed using TRIzol (Invitrogen, Life Technologies, Carlsbad, CA, USA). TaqMan MicroRNA Assays (Applied Biosystems, Foster City, CA, USA) were used for the quantification of miR-16 (hsa-miR-16, miRBase ID hsa-miR-16-5p, Applied Biosystems) and miR-519a (hsa-miR-519a, miRBase ID hsa-miR-519a-3p, Applied Biosystems), as previously described (24) and normalized against MiR-191 (hsa-miR-191-5p, Applied Biosystems), the reference microRNA in all experiments.

2.3 Cell lines

We used two cell lines, i) a human malignant meningioma cell line, IOMM-Lee cells (intraosseous malignant meningioma; a generous gift from Dr Gillespie and Dr Jensen, University of Utah, USA) (48), and ii) a benign grade 1 meningioma cell line, Ben-Men-1 cells, which were immortalized by retroviral transduction with human telomerase reverse transcriptase (Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Germany) (49). The cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM, Life Technologies, Carlsbad, California, USA) supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 mg/ml streptomycin at 37°C in 5% CO2.

2.4 MiR mimics transfection, cell growth, dose-response assays, and Ki-67 labeling index

The overexpression of miR-16 and miR-519 was achieved by transfection of mirVana miR Mimic hsa-miR-16-5p (Ambion, Life Technologies) and mirVana miR Mimic hsa-miR-519a-3p (Ambion, Life Technologies), respectively. MirVana miR Mimic Negative Control (Ambion, Life Technologies) was used as a control and referred throughout the work as miR-mimic negative control. Cell transfections, using Lipofectamine RNAiMAX Transfection Reagent, were performed following the manufacturer's instructions. Transfection efficacy was verified 48h later using qRT-PCR technique.

For performing anchorage-dependent cell growth assay, 25,000 cells were incubated per well in 24-well plates. Cells were transfected with 17 nM of miR mimics. The number of cells per microliter was counted at 2, 4, and 6 days after transfection using LUNA Automated Cell Counter (Logos Biosystems, Annandale, USA). For performing dose-response assay, 10,000 cells were incubated per well in 48-well plates, 24 h before transfection. Cells were transfected with 0 nM, 0.17 nM, 1.7 nM, 3.4 nM, 17 nM and 170 nM of miRs. The number of cells per microliter was counted 96 h after transfection. Each measurement was performed three times after three independent transfections (n=9).

Additionally, cell proliferation was evaluated in IOMM-Lee cells based on the expression of Ki-67. The Ki-67 labeling index (LI) was evaluated 72 h after transfection, using anti-Ki-67 primary antibody (1/500; mouse monoclonal, MIB-1, Dako Cytomation), and fluorescent FITC anti-mouse Alexa Fluor (1/1000; Life Technologies) secondary antibody. A total of 500 cells in areas showing maximal nuclear intensity were used to compute the LI. Each measurement was performed three times after three independent transfections (n=9 in total).

2.5 Transcriptomics

Seventy-two hours after independent transfection with miR-16 (n=6, miR Mimic hsa-miR-16-5p), miR-519 (n=6, miR Mimic hsa-miR-519a-3p), and miR-mimic negative control (n=6, mirVana miR Mimic Negative Control), total RNA of the transfected IOMM-Lee cells was extracted using the TRIzol protocol (Invitrogen, Life Technologies, Carlsbad, CA, USA). Gene expression experiments were performed using the Affymetrix Human Gene v.2.0 ST Arrays according to GeneChip® WT PLUS Reagent Kit, Manual Target Preparation for GeneChip® Whole Transcript Expression Arrays P/N 703174 Rev.2 protocol; 100 ng of Total RNA were used as a starting amount for microarrays experiments; 3.5 µg of labeled DNA were injected into the Affymetrix cartridge. The arrays were hybridized with rotation at 60 rpm for 16 hours at 45°C. The arrays were washed and scanned according to the protocol GeneChip® Expression Wash, Stain and Scan For Cartridge Arrays P/N 702731 Rev. 4.

Fluorescence values corresponding to raw expression data for each sample were extracted from each Affymetrix CEL files (one file per sample) using the R (v3.6) oligo package with the corresponding microarray platform definitions (pd.hugene.2.0.st). The extraction method included no normalization or background correction with the RMA algorithm. Positive and negative control probes were removed. The remaining 44.629 probes were annotated with up-to-date gene symbols using our local Ensembl database (version 83_38), allowing for accurate miR precursors and other ncRNA determination. Non-linear effects such as background or saturation were corrected by LOWESS normalization against a median profile of all samples (50). Data were then subjected to hierarchical clustering, which delineated clusters of co-expressed genes on one dimension and classified samples according to their expression profiles on another dimension. The method was applied on log2-transformed and gene-median-centered data, using uncentered Pearson's correlation as similarity metric and average linkage to reconstruct the gene and sample dendrograms. Gene clusters were delimited by applying a distance threshold of 1/5 on the gene tree. Gene clusters separating control, miR-16 and miR-519 samples

were then extracted, and a collective p-value (Student t-test) was computed between each group. For each sample, a mean expression value of all genes from the initial gene cluster was calculated and these values were compared between the selected groups. This strategy, based on strong correlation of gene expression, allowed us to avoid multi-testing as a means of p-value correction for the unsupervised analyses.

Differential gene expression analyses and statistics were achieved with moderated t-tests [linear modeling with empirical Bayes (51)] and corrected for the false discovery rate (FDR) with the Benjamini-Hochberg procedure. All clusterings were performed with Cluster 3.0 (52). For each identified gene list (gene cluster, gene signature, differential genes), functional annotations were performed using enrichR on multiple databases and gene sets (53). In-house enrichment analyses were conducted by calculating the ratio of frequencies Observed/Expected, where Observed was the frequency of the GO-Pathway-Disease term in the cluster or list, and Expected was the background frequency on the whole chip. Fisher's exact tests were used to statistically validate the results.

MiR profiling and bulk-transcriptome public data were downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database as raw gene datasets when possible, or processed datasets otherwise, and underwent the same quality control, preparation and annotation steps as described above.

2.6 Cell protein extraction and western blot

Seventy-two hours after IOMM-Lee cells transfection with miR-16, miR-519 and negative control miR mimics, total cellular proteins were extracted with RIPA buffer. Expression levels of HuR and GAPDH were then analyzed by Western blotting (n=9). The following primary antibodies were used: HuR (1/1000; rabbit polyclonal, Millipore), GAPDH (1/2000; chicken polyclonal, Millipore). Densitometry of all samples of Western blots were measured with Image J 1.42u (Wayne Rasband, National Institutes of Health, USA).

2.7 Statistical analyses

All quantitative variables are described as medians and percentiles [Interquartile range (IQR), 25–75th percentile]. All proportions are expressed as percentages with 95% confidence intervals (95% CI). Comparisons of miR-16 and miR-519 expression levels across the three tissue groups were performed using the Kruskall-Wallis test. Comparisons of miR-16 and miR-519 expression level of normal tissue and meningioma (both grade 1 and grade 2 subtypes) were carried out using the Mann-Whitney U test. When a statistically significant difference was found, the effect size estimated (r) for the difference between the two groups was calculated and interpreted according to Cohen's method using z value. Cumulative probabilities of relapse-free survival were estimated by the Kaplan-Meier method. To evaluate the potential association between miR-16 and miR-519 and time to relapse, univariate analyses, using log-rank test were carried out on measures of miR-16, miR-519 and Ki-67 divided in medians. Log-rank tests were also performed using the online Cutoff Finder tool to screen for significant cutoff values (54). Cox proportional-hazard regression analysis was performed to identify independent variables predictive of relapse, using the following covariates: meningioma grade (1 or 2), miR-16, and miR-519 as continuous variables or as quartilized variables. Results were shown as hazard ratios (HRs) with 95% confidence intervals.

For serial measurements of cellular growth in the three experimental groups, we tested the change over time in cell viability and the difference between the three experimental groups over time using repeated measures analysis of variance (ANOVA) of log-transformed data. Two summary measures of interest were considered in serial analyses, namely: i) the area under curve considering the first value as the baseline value and ii) the percentage of the difference between the first and the last values. *Post-hoc* analysis for pairwise group comparisons was performed using the Student-Newman-Keuls test to avoid multiple testing issues. The measurement of progression of cell viability in the three experimental groups at successive times was carried out using the Friedman test for testing the difference between several related samples, as the same parameter was measured under different conditions in the same group. In the dose-effect study, three cell line groups were compared according to the type of miR transfected (miR-mimics for miR-16, miR-519, or negative control), and their concentration used which varied from 0 to 170 nM. The absolute number of viable cells at 96 h after the initiation of transfection was compared across the three groups by repeated measures ANOVA of log-transformed data. At each dose point, one-way Student t-tests were used to compare the number of viable cells between controls and miR-16 or miR-519.

3 Results

3.1 MiR-16 and miR-519 were underexpressed in human meningioma samples

In patient meningioma samples (clinical data detailed in Table 1), lower levels of miR-16 were found in tumoral relative to control tissues (P = 1.23e-04; Mann-Whitney U-test), whether healthy arachnoid (P = 8.72e-04) or dura mater (P = 8.66e-03) (Figure 1A). MiR-519 expression was also lower in meningioma vs. control (P = 2.31e-03), with a clear differential against arachnoids (P = 3.52e-07) and no difference against dura mater (P = 0.76) (Figure 1B). In these samples, we found no significant difference in miR-16 and miR-519 levels between grade 1 and grade 2 tumors (Figure S1). Among this cohort, 22/56 (39%) patients were recurrence-free with the actuarial survival probabilities at their last known follow-up; neither the level of miR-16 nor that of miR-519 (expressed in medians) associated with post-surgical recurrence (Figures 2A, B). Positive associations were found between tumor grade, Ki-67 labeling index and higher risk of recurrence (P = 0.011 and P = 0.012, respectively) (Figures 2C, D), tumor grade and Ki-67 LI being highly correlated (rho = 0.74,

TABLE 1 Summary of the demographic and clinical features of the patients with meningioma.

Features	WHO grade 1 (n = 32) %		WHO grade 2 (n = 28) %	
Gender				
Male	34%		61%	
Female	66%		39%	
Patient outcomes				
Death	0%		8%	
Recurrence	25%		60%	
Simpson's grade				
Grade 1	44%		55%	
Grade 2	16%		17%	
Grade 3	40%		28%	
Treatment (excluding surgery)				
Radiation therapy	4%		44%	
Pre-operative embolization	8%		11%	
	Median	(IQR, 25^{th} – 75^{th})	**Median**	(IQR, 25^{th} – 75^{th})
Age, years (IQR, 25^{th} – 75^{th})	59	(54–70)	68	(57–72)
Follow-up duration, months	36	(12–55)	22	(11–46)

QR, interquartile range. Simpson's grade (55): 1, macroscopically complete removal, including dura and bones; 2, macroscopically complete removal, dural coagulation; 3, complete removal, dura not coagulated. Grade: 2021 World Health Organization grading.

P = 1.1e-10; Spearman's test). Further analyses by the Cutoff Finder online tool (54) revealed that miR-16 and miR-519 showed no significant threshold (P = 0.07 and P = 0.20, respectively). Multivariate analysis using Cox proportional-hazards regression model consistently indicated that meningioma tumor grade was the only independent predictor of disease recurrence after adjusting for age, sex, miR-16, and miR-519 (P = 0.03).

3.2 *In vitro* investigation of miR-16 and miR-519 in IOMM-Lee and Ben-Men-1 cells

3.2.1 MiR-16 and miR-519 level in control and transfected cells

Given the results of decreased miR-16 and miR-519 in human meningioma against healthy tissues, we next examined the level of these miRs in two cultured meningioma cells lines, IOMM-Lee and Ben-Men-1 cells. We found that while miR-16 and miR-519 were nearly undetectable in IOMM-Lee cells, their levels were significantly higher in Ben-Men-1 cells (P = 2e-03 and P = 5e-03, respectively; Figure S2), with equivalent miR-16 and 5-times lower miR-519 levels in Ben-Men-1 as compared with healthy arachnoids. We proceeded to overexpress these miRs into both cell lines, noting that the transfection with miR-16 mimic did not significantly alter the expression of miR-519 (P = 1 in both IOMM-Lee and Ben-Men-1 cells), and vice versa (P = 0.13 in IOMM-Lee cells; P = 0.81 in Ben-Men-1 cells). The efficiency of transfection with miR-16 and miR-519 mimics was confirmed by qRT-PCR (Figure S3).

3.2.2 Effect of miR-16 and miR-519 transfection on cell proliferation

Cell growth of transfected IOMM-Lee cells was compared among three groups of cells, namely control (miR-mimic negative control), miR-16 and miR-519 mimics, on day 0, 2, 4, and 6 after transfection. Significant differences for both miRs were noted 2 to 4 days after miR transfection, with cells transfected with miR-16 showing the lowest growth (ANOVA of log-transformed data, P < 1e-04; pairwise comparisons, P < 0.05) (Figure 3A). Similar effects of the two miR-mimics were observed in Ben-Men-1 cells (both P < 1e-04) (Figure 3B). Ki-67 LI was lower in miR-16 (median, 81.4%; IQR, 72.9–85.2%) and in miR-519 (median, 96.4%; IQR, 95.8%–97.3%) miR-mimics transfected cells than in miR-mimics negative control (median, 99.2%; IQR, 98.8%–99.4%) transfected cells (P < 0.0001; *post-hoc*, P < 0.05 for all pairwise comparisons) (Figure 3E).

3.2.3 Dose-response study of the effects of miR-16 and miR-519 on cell growth

In IOMM-Lee cell line, the number of viable cells was significantly lower in miR-16 transfected group than in both

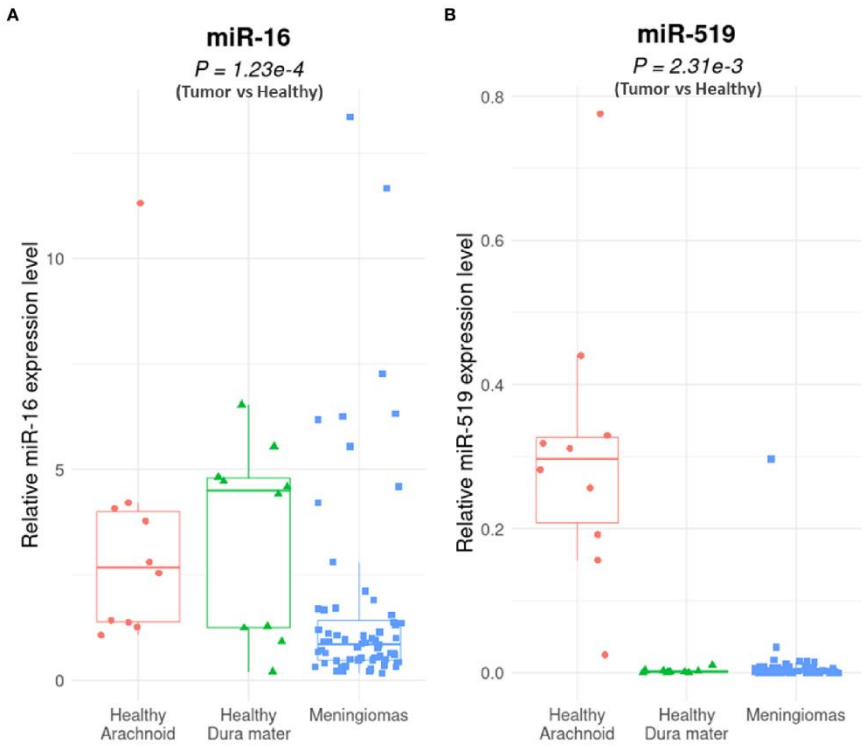


FIGURE 1
MiR-16 and miR-519 expression levels in human meningioma vs. healthy tissues. Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analysis of miR-16 **(A)** and miR-519 **(B)** expression in human meningiomas (n = 64: 30 grade-1 and 34 grade-2) vs. non-tumoral tissue (n = 20: 10 arachnoid and 10 dura mater) (P = 1.23e-4 for miR-16 and P = 2.31e-3 for miR-519). Boxplot (minimum, first quartile, median, third quartile, and maximum) with in-dividual scatterplots.

control and miR-519 transfected groups in all concentrations tested (0.17, 1.7, 17 and 170 nM) (all P = 1e-03) (Figure 3C), and the number of viable cells was significantly correlated with miR-16 transfection concentration (rho = -0.950; P < 1e-04). The number of viable cells did not differ between miR-519 and negative control-transfected groups, except at the highest dose (170 nM) where it was significantly reduced (P < 1e-04). Similar results were found in Ben-Men-1 cell line with the number of viable cells correlating significantly with the concentration of miR-16 (P < 1e-04) (Figure 3D). No significant inhibitory doses were found for miR-519 on this cell line.

3.3 MiR-16 and mir-519 transcriptomics

3.3.1 Impact of the transfection with miR-16 and miR-519 mimics on the transcriptome of IOMM-Lee cells

The transcriptome-wide effects of miR-16 and miR-519 transfection were analyzed in human anaplastic meningioma IOMM-Lee cell line (Figure 4). Hierarchical clustering was performed on three groups of samples: miR-16-mimics (n=6), miR-519-mimics (n=6) and miR-mimics negative controls (n=6). Each group was clearly separated from the other while retaining a high individual correlative structure (Figure 4A). These intergroup difference and intragroup cohesion were confirmed with principal component analysis, which illustrated the relative equidistance from controls and both miR-16 and miR-519 profiles, albeit on different axes of variances (Figure 4B). Shared features were also observed for the two miR-mimic groups. This segregation was driven by 5 clusters of strongly correlated genes (C1 to C5; Figure 4A), each of them differential vs. controls (Figure 4C), two of which displaying similar expression profiles for both miR-mimics (C3 and C4, down- and upregulated in both miR-16 and miR-519, respectively), the three others functioning in opposite directions (C1, downregulated in miR-16, C2, upregulated in miR-519, and C5, upregulated in miR-16 but downregulated in miR-519). These 5 clusters were functionally annotated (Figure 4D, Tables S1–S5) and associated with significant processes and pathways such as mitotic cell cycle *via* TP53, replication complexes (including MCMs) and gene expression *via* the DREAM complex (C1), TNFα signaling and immune response (C2), brain and cilium development (C3), regulation of apoptosis, hypoxia, cell migration, and the modulation of the extracellular matrix (C4), or NF-KB signaling and macroautophagy (C5) (all adjusted-P < 0.05). As expected, they were also highly enriched in miR-16 (mainly C1 with 34%) and miR-519 (mainly C5 with 26.5%) mRNA targets (Figures 4D, E), and in meningioma signature genes (C1 and C4). Multivariate

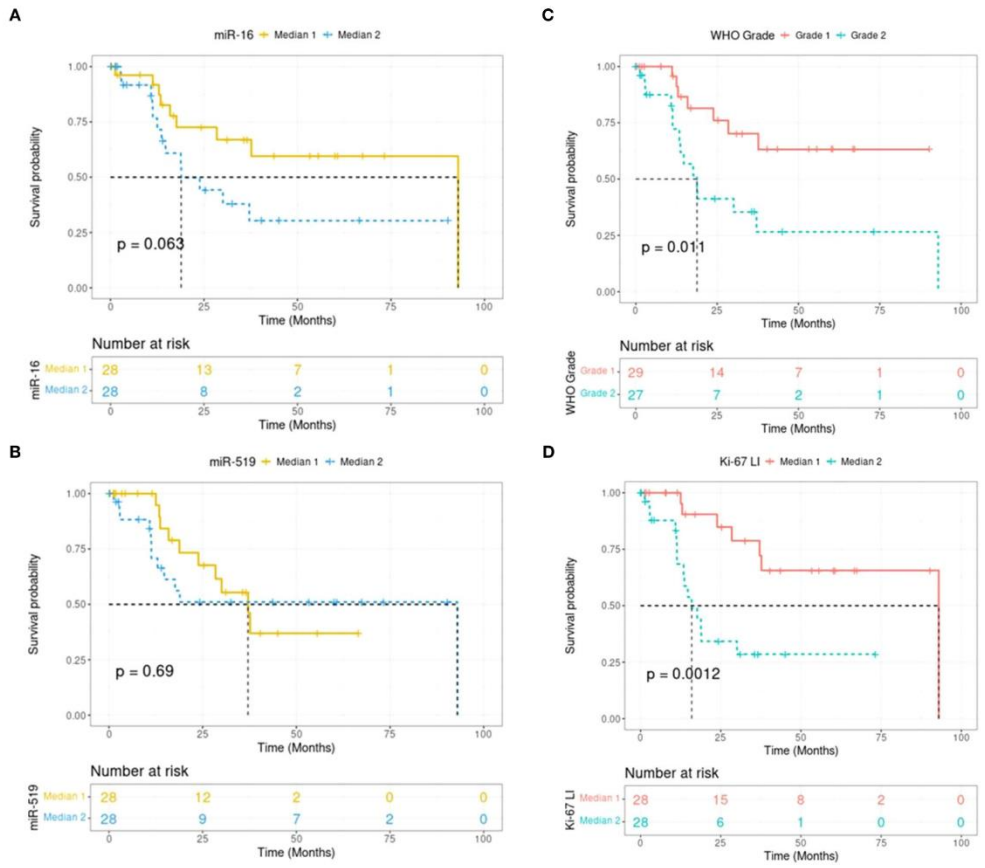


FIGURE 2
Kaplan-Meier plots showing recurrence-free survival in human meningiomas (n = 56). Survival curves are split in 2 groups for miR-16 **(A)**, miR-519 **(B)**, WHO grade **(C)** and Ki-67 **(D)**. Expression levels are grouped into medians, with Median 1 being the lower expression group. WHO, world health organization; LI, labeling index.

statistics confirmed these results and yielded 510 differential genes between miR-16 and control transcriptomes, and 152 genes between miR-519 and controls (FDR < 0.05, differential > 2-fold-change; moderated t-tests on linear modeling with empirical Bayes; Tables S6, S7). Few of these top genes overlapped between miR-16 and miR-519 (35 unique genes up in both miR-mimics, 2 unique genes down in both miR-mimics, 2 unique genes down in miR-16 while up in miR-519).

3.3.2 Transcriptome-wide effect of miR-mimics on HuR (*ELAVL1*)

Next, we asked if and how these five differential and functional clusters obtained with miR transcriptomics (C1 to C5) were enriched in HuR targets and in the HuR transcriptomic signature previously identified by our group in a similar setup on meningioma samples with HuR knockdown (GSE95212, 43). Compared to cells expressing miR-mimic negative control, both miR-16 and miR-519 transfected cells had significantly lowered level of *ELAVL1* (HuR) mRNA (FDR = 6.1e-06 and FDR = 9.38e-03, respectively; moderated t-tests; Figure 4C). Western blot showed that the expression level of HuR was significantly lower (2.2-fold decrease) following the transfection with miR-16, but not miR-519 mimics (P < 1e-04 and P = 0.8, respectively). Remarkably, *ELAVL1* was one of the downregulated genes of cluster C1 with functions in pre-replicative complex and cell cycle. Further gene enrichment analysis showed that 79.6% of known HuR targets [2802-gene list obtained from Starbase v2 https://starbase.sysu.edu.cn/starbase2/ for HuR transcriptomic compatibility (43)] were distributed in the five clusters, the most represented being C1 (26.1%; 2.6-fold enrichment), C4 (15.9%; 2.7-fold enrichment) and C5 (21.4%; 2-fold enrichment) (all P < 2.2e-16; Fisher's exact test; Figure 4E). Furthermore, genes that were dysregulated in our HuR knockdown transcriptomic experiment were also found enriched in the present miR-mimics signature. Downregulated genes in HuR knockdown mainly distributed in clusters C3, C4 and C5 (5.5%, 8.7% and 22.1%; 2.2, 2.2 and 3-fold enrichment, respectively). Conversely, genes that

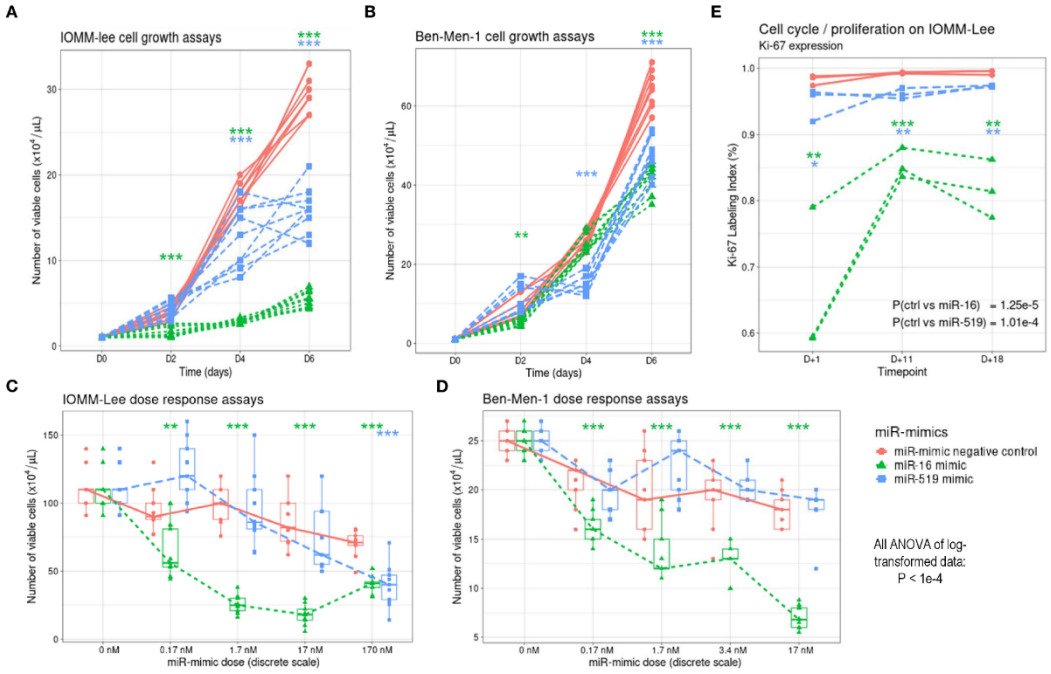


FIGURE 3
Effects of miR-16 and miR-519 overexpression in IOMM-Lee and Ben-Men-1 cells. **(A)** Meningioma cell viability measured in IOMM-Lee cells transfected with 17 nM of miR-16 and miR-519 mimics, and miR-mimic negative control (n = 9 in each group). **(B)** Similarly, meningioma cell growth measured in Ben-Men-1 cells. **(C)** Dose-response curve constructed by determining the number of viable IOMM-Lee cells 96 h after transfection with 0 nM, 0.17 nM, 1.7 nM, 17 nM and 170 nm of miR-16 and miR-519 (n = 9 in each group). **(D)** Similarly, dose-response curve obtained in Ben-Men-1 cells after transfection with 0 nM, 0.17 nM, 1.7 nM, 3.4 nM and 17 nm of miR-16 and miR-519. **(E)** Ki-67 expression at 3 different time points after transfection with 17 nM of miR-16, miR-519 and miR-mimic negative control (n = 3 in each group). *P < 0.05, **P < 0.01, ***P < 0.001 (one-way Student t test).

were upregulated by HuR knockdown distributed in C1, C2 and C4 (15.8%, 8.9% and 11.2%; 2.2, 2.4 and 2.8-fold enrichment, respectively). Genes outside the miR-mimics signature were 3-times depleted in HuR knockdown genes (all P < 2.2e-16; Figure 4E).

3.3.3 MiR-mimics signature of shared miR-16, miR-519 and HuR targets

Considering the high levels of enrichment of the five transcriptomic clusters in miR-16 or miR-519 target mRNAs (Figure 4E), we consolidated this large 5-cluster signature into a more informative feature containing only 208 shared targets of both miRs. Nearly all genes (202/208) were also HuR targets (Figure 4F). Given their differential expressions and their combined functions, the redistribution of these genes back into C1 to C5 can be considered highly representative of the larger signature (Figure S4, Table S8). Indeed, the 208-gene signature contained both transcriptional regulators involved in cell cycle (such as E2F genes) and protagonists of proliferation (*MKI67*, *CCND2*, *CDKN1A*). The signaling network reconstructed with these 208 genes linked them functionally around meningioma relevant hubs such as AKT3, CDKN1A, PAK2 and PRKAA1 and revealed functions and pathways including cell cycle progress, differentiation, DNA damage response, mRNA nucleus export, growth factors and fibrosis (all FDR < 0.05; Figure S4).

3.3.4 The miR-mimics signature differentiates between subgroups of meningiomas with distinct proliferative features

Finally, we evaluated our findings on relevant public datasets of human meningiomas. On a miR profiling dataset [GSE126563 (40)] of primary (n=44) and secondary (n=15) tumors we found diminished miR-16 levels when compared to controls (n=5; P = 0.047 vs. all tumors, and P = 0.035 vs. primary meningiomas alone). On another dataset of various meningioma grades [GSE50641 (36)], we found lower miR-16 expression in grade 2 (n=11) as compared to grade 1 (n=33), which already displayed very low levels (P = 0.037). In either set, miR-519 was below detection and could not be tested.

We also investigated a landmark transcriptome of 121 meningiomas [GSE85135 (14)], where the miR-mimics signature highlighted two dominant clusters with different expression profiles, named left and right branch (LB and RB, respectively) of the clustering tree (Figure 5A). According to this 208-gene hierarchical clustering, healthy controls expectedly displayed a correlated pattern for both embryologic and adult tissues, distinct from LB and RB profiles. Interestingly, the dura mater control samples clustered preferentially with tumoral samples on the LB, instead of regrouping with the rest of the controls. Investigating on these profiles from the whole transcriptome led to the same

Hergalant et al.
10.3389/fonc.2023.1158773

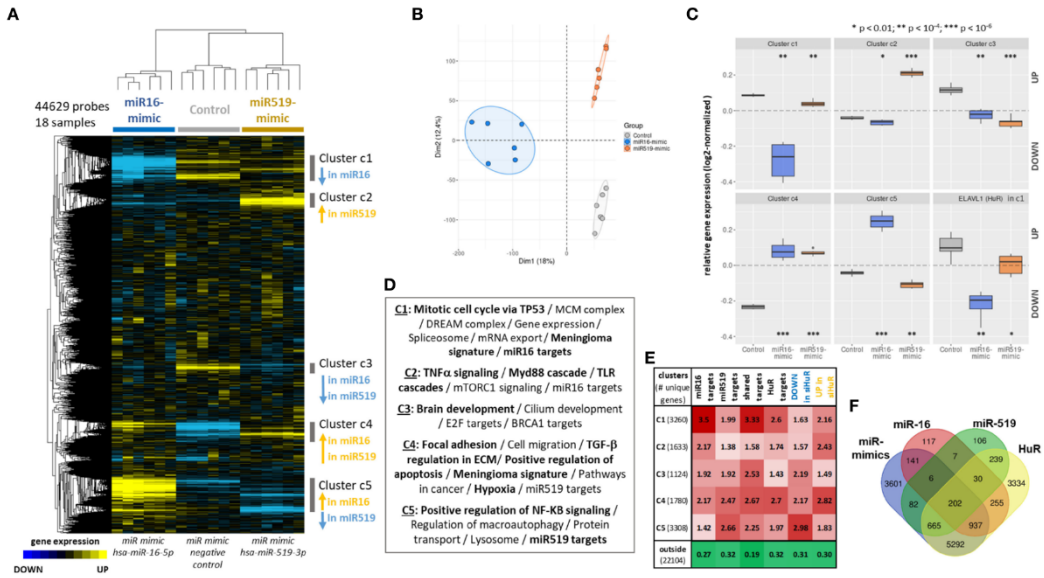


FIGURE 4
Effects of miR-16 and miR-519 transfection on the transcriptome of IOMM-Lee cells. A differential transcriptomic signature is extracted, dissected and compared for each miRNA (miR16: n = 6; miR-519: n = 6; negative controls: n = 6). **(A)** Hierarchical clustering heat map of the whole transcriptome delineating five significant clusters of co-expressed genes. Median-centered log2-transformed gene expression (arbitrary unit). Blue, yellow and black indicate downregulated, upregulated, and median genes, respectively. **(B)** Principal Component Analysis confirms the correlative structure within each group and the directional scattering between groups. **(C)** Boxplot comparing overall gene expression for the 5 significant identified gene clusters (c1 to c5) and for ELAVL1 (HuR). Student's t-tests. **(D)** Top functional annotations associated with each gene clusters (with EnrichR, all q-values < 0.05). **(E)** Relative fold-enrichments computed for each five clusters (c1 to c5) and the rest of the microarray (outside) in target lists (from ENCORI https://starbase.sysu.edu.cn/ under high stringency): miR16 targets (1,695), miR519 targets (1,337), miR16 and miR519 overlapping targets (245), HuR targets (2,802; from Starbase v2), DOWN-regulated signature in siHuR transcriptome (4,725 genes), and UP-regulated signature in siHuR transcriptome (4,787 genes) (43). Fisher's exact tests, all FDR < 0.01. **(F)** Venn diagram showing the overlaps between the miR-mimics signature comprising the five differential clusters and miR-16, miR519 and HuR targets (10,954; from ENCORI under high stringency). FDR: false discovery rate. * p < 0.01; ** p < 1e-04; *** p < 1e-06.

conclusion, with little to no overlap with the adult arachnoids. The overlap with tumoral and embryologic tissues was enriched in miR-519 targets (>1.8-fold; P < 2.2e-16), confirming our findings in human tissues (Figure 1B). After removing controls, transcriptome-wide differential statistics between LB and RB respectively yielded 2,279 and 1,816 up- and downregulated genes in LB (FDR < 0.01, Figure 5B), which amounted to 18% of the measured genes. Remarkably, these genes were extremely enriched in HuR targets, 56.2% of which were UP in LB (DOWN in RB) and 34.3% UP in RB (or DOWN in LB). Genes UP in LB revealed functions linked to mitotic cell cycle, DNA repair and immune response and were also enriched in miR-16 and HIF1A targets. Genes UP in RB were associated to mRNA processing and transport (all FDR < 0.05). Moreover, we checked *EP300* level, which was reported to be a solid marker of meningioma recurrence, independently of WHO grade (11). Consistent differences were observed between the two subgroups (P = 2.75e-10; Figure 5C) for this mRNA, along with other markers associated with aggressiveness/proliferation, also outside the 208-gene list (*HIF1A*, *RB1*, *BAD* or *IDH1*; all FDR < 1e-14). Furthermore, *ELAVL1* (HuR) showed reduced levels in the subgroup with diminished gene expression and mRNA transport processes and was overexpressed in the subgroup with increased cell cycle and proliferation features (FDR = 3.76e-04). On another transcriptomic dataset [GSE74385 (9)], the miR-mimics signature again classified samples in two groups. WHO grade 1 (13/16) and non-recurrent (16/20) meningioma were over-represented in the first, the second regrouping higher grades (29/37) and recurrent tumors (14/16), with decreased *EP300* (P = 1.13e-04) and increased *FOXM1* (P = 6.62e-07) levels, previously reported in recurrent meningiomas, as down- and upregulated, respectively (11).

Finally, to validate our main transcriptomic findings, we performed an integrative analysis of a multi-omic meningioma dataset with bulk transcriptome coupled with miR measurements for the same patients (GSE88721), including samples of heterogeneous histological origin and grades (1 and 2) (56). Focusing on miR-16 and miR-519 mature products, along with their premature forms, we associated the miR expression levels with that of hallmark target genes representing the whole correlated panorama (Figure S5). Expression of *ELAVL1* was found negatively correlated with miR-519 (r2 = -0.69; Pearson's correlation) but not with its premature forms (r2 = 0.38), potentially indicating a preferential miR-519 inhibition of *ELAVL1* transcripts in meningiomas, rather than the inhibition of miR-519 by HuR and/or as a consequence of other transcriptomic changes. The hypoxia factor gene *HIF1A* correlated with both miR-16 and miR-519 levels (r2 < -0.4, respectively). G1/S Cyclin D3 (*CCND3*), from the pre-replicative complex and cell cycle progression marker, was found significant with both miRs, along with *RB1* disease progression marker (and directly inhibited by Cyclin D3 upon phosphorylation), strongly

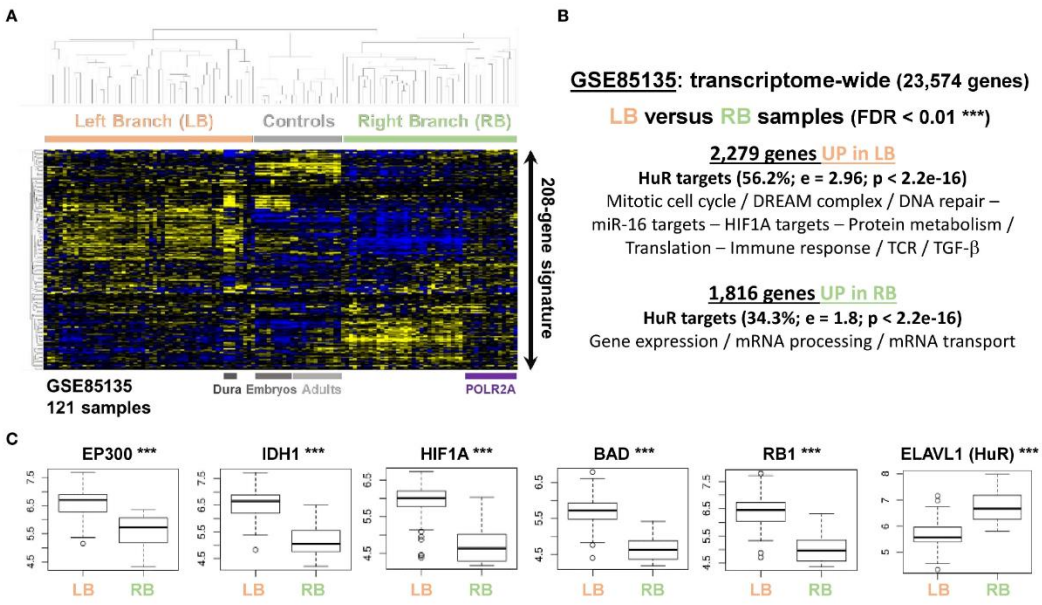


FIGURE 5
The miR transcriptomic signature distinguishes two subgroups of samples with different proliferative features in a landmark dataset of grade 1 meningiomas (GSE85135 (14)). **(A)** Unsupervised hierarchical clustering according to the reduced 208-gene miR-16/miR-519 signature. **(B)** Transcriptome-wide differential statistics between the 2 groups of samples identified with the reduced signature (LB, RB). Up- and downregulated gene lists were functionally annotated with EnrichR (adjusted p-values < 0.05). Control samples were not included in the statistical analyses. **(C)** Boxplots representing the differential expressions between LB and RB for a select list of genes reported to be consistent markers of meningioma recurrence and aggressiveness, and HuR. These genes do not belong to the 208-gene signature. Y-axis: normalized gene expression (log2). Enrichment in HuR targets: Fisher's exact test; e, relative fold-enrichment. LB, left branch of the dendrogram; RB, right branch of the dendrogram. *** FDR < 0.01.

associated with miR-519a (r2 = -0.62). In line with our transcriptomic clusters, we also report *BCL2L1* (potent apoptosis regulator and caspase inhibitor) and *NFKB1* associations with miR-519, and *MTOR* and *TNF* associations with miR-16.

4 Discussion

MicroRNA expression is an open subject in human meningioma where limited number of studies have explored tissue and serum samples (2). For example, Zhi et al. found increased serum content of miR-106a-5p, miR-219-5p, miR-375 and miR-409-3p, and decreased content of miR-197 and miR-224 in meningioma patients (37). The putative functions of these miRs seem to depend on their relative expression level in tumor vs. normal tissue. For instance, the overexpression of miR-335 in meningioma samples was used to evidence its role as "oncomiR" (57), while the downregulations of miR-200a and miR-145 were used to support their tumor suppression function (58). Likewise, in a retrospective study, higher miR-190a expression level was reported to be an independent prognostic factor of meningioma recurrence rates and lower miR-29c-3p and miR-219-5p were found to be associated with advanced clinical stages of meningioma (35). Ludwig et al. identified several dysregulated miRs between different subtypes of benign meningiomas, and in anaplastic vs. benign tumors. They further marked a 4-miR signature, miR-222, miR-34a, miR-136, and miR-497, as differentiating WHO grade 2 from grade 1 meningiomas in a setup of 55 samples of various histological types (36). MiR-34a was next confirmed as differential between grade 2 and grade 1 meningiomas in a study led by another group (38). However, no other reported miR otherwise overlapped with the previous works, suggesting that larger cohorts of patients might be needed to overcome the heterogeneity of meningioma subtypes. More recently, miR-15a, miR-146a, and miR-331 were identified as good prognosticators of relapse (40), in a design of paired primary vs. recurrent tumors and a large validation cohort. However, they did not validate the previously reported miR-190a as differentially expressed. Conversely, Negroni et al. confirmed miR-497 as a circulating biomarker for high-grade meningiomas, with lower levels in serum exosome samples as compared with benign meningiomas (39). Here we found both miR-16 and miR-519 downregulated in benign and atypical meningioma vs. normal arachnoid tissues.

MiR-16 and miR-519 have previously been shown to be tumor suppressor miRs in several types of tumors, including laryngeal squamous cells, non-small cell lung carcinoma, breast carcinoma, nasopharyngeal carcinoma, prostate cancer, glioma and glioblastoma (23–34, 59), but were not investigated as such in meningioma. We found them both downregulated in human samples. This result was replicated for miR-16 in a miR dataset of healthy tissues and primary + secondary tumors (40), albeit only as

a trend probably because of the limited number of controls samples ($P < 0.05$). On another meningioma dataset (36), we found diminished miR-16 levels in grade 2 versus grade 1 tumors. This trend ($P < 0.05$) should be validated in larger and more homogeneous cohorts given that benign tumors were subdivided into meningothelial, fibroblastic and transitional meningioma variants. In our data we did not find significant change of miR-16 levels between grades. Moreover, in all available data miR-16 expression was already very low in grade 1, making any comparison between grades hazardous. For miR-519, no exploitable data was available as expression reached background levels in every dataset. In-depth sequencing could solve these problems and thus appears to be a much-needed endeavor in meningioma miR profiling. Especially because very few miR datasets are publicly available and/or computationally exploitable.

In addition to their downregulation in human tissues, here we studied the consequences of miR-16 and miR-519 overexpression in benign as well as malignant cell lines. We found both miRs tumor-suppressive. In addition, we show that the repressive effect of miR-16 is dose-dependent, like what was reported by Reid et al. in malignant pleural mesothelioma cells (60). Interestingly, this cell growth inhibition was much less pronounced in the benign Ben-Men-1 cell line than in the anaplastic meningioma IOMM-Lee cell line, probably since IOMM-Lee cells have lower basal level of miR-16 and miR-519 than Ben-Men-1 cells, resulting in their higher sensitivity to the overexpression of these miRs. Confirming the growth inhibition on another cell line is therefore necessary before drawing conclusions. Nevertheless, our findings suggest that miR-16 and miR-519 mediate anti-tumorigenic processes *via* inhibition of cell proliferation. A fact corroborated by lower Ki-67 labeling index in IOMM-Lee cells overexpressing either miR-16 or miR-519.

The results of the transcriptomic study also suggest that the cell growth inhibition by either miR is mediated by downregulations of both the pre-replicative complex and cell cycle *via* p53, part of the transcriptome-wide consequences of their overexpression. In fact, the common dysregulated clusters we extracted as a result were far more informative as they were composed of meningioma-associated genes and pathways of regulation of apoptosis and of brain development. These processes all agree with the tumor suppressor potential of these miRs. Further in line with previous results, gene expression *via* the DREAM complex was again pointed out as a culprit in meningioma progression (12), which reinforces the usefulness of our miR-mimics signature as it was able to segregate samples according partly to this feature in external human datasets. Of note, in the landmark cohort from Clark et al. (14), POLR2A-mutated samples clustered outside the main subgroups and were mostly unresponsive to the signature, which delineates its direct implication in cell cycle progression and gene expression. In the larger 5-clusters signature, target genes of miR-16 and miR-519 were expectedly enriched, but were also targets of HuR, which our group reported as upregulated and as a poor prognosis factor in meningioma progression and recurrence (43). Here, HuR was strongly under-expressed following overexpression of both miRs, and part of the gene expression/cell cycle cluster. We propose thus that the interplay of the three markers, namely HuR, miR-16, and miR-519 is of importance in meningioma development and progression. In this regard, the restricted 208-gene signature deserves attention as many of these genes (Table S8) may emerge as precious additions to the meningioma biomarker repertoire. Moreover, many targets genes and signatures proposed by the transcriptomics (HuR, HIF1A and hypoxia, EP300) have already been functionally validated or meta-analytically cross-validated in previous works (11, 43, 61). In a previous study on HuR in meningioma patients and following knockdown in the same meningioma cell lines, we already correlated mRNA and protein levels, as well as HuR cellular localization and post-translational modifications. We also extensively studied HuR mRNA targets *via* transcriptomics (43). In the present work, we overlap the signatures obtained previously and the miR-mimic signatures. By doing so we recover the hypoxia signature that was functionally validated in HuR work, including its action on cell growth. Hypoxia being under the tight control of HIF1A, we demonstrated it to be a hallmark of meningioma progression. Concerning the transcription factor EP300, we previously correlated a methylation signature with grade, progression, and proliferation markers such as Ki-67 and MCM6 and showed that the regulatory regions associated with meningioma growth are highly enriched in CpG islands located in enhancers in distal regions (61). This methylation signature is known to be a mark of tissue-specific EP300 activity, and involved in cell growth and division in cancers (62, 63). In our final integrative experiment on human tissues, we provide the hint of a mechanistic link between the two miR expressions and that of hallmark target genes we report from our own transcriptomic findings: markers from each of the five clusters, miR-16 and/or miR-519 targets, and progression markers associated with meningioma aggressiveness. These results translate directly *in vivo*, onto meningioma samples of various histological subtypes and grades, therefore we believe our proposed biomarkers and signatures could have wide biological and clinical meaning.

In healthy tissues, little is known on the molecular differences between the arachnoid and the dura mater. Contrary to miR-16, here we observed similarly diminished miR-519 levels in the dura mater and in tumoral tissues. Further investigating the transcriptomic profiles available for healthy controls with and without the 208-gene signature uncovered a dura mater pattern overlapping with other embryologic tissues and an enrichment in miR-519 mRNA targets, and no overlap with the adult arachnoids, confirming our first observation on human meningiomas.

Decreased miR-16 levels have been observed in colorectal cancer (64), non-small cell lung carcinoma (24, 59, 65), chronic lymphocytic leukemia (66), pituitary adenomas (67), and gliomas (30). Our findings are consistent with these previous observations and suggest the involvement of miR-16 in tumor suppression. The molecular basis for the suppressive action in meningioma growth, however, is not clear. Yang et al. reported that miR-16 inhibits cell growth and reduces invasive properties in a glioma cell line through the suppression of BCL2 and NF-kappaB1/MMP-9 signaling pathway (30). Alternatively, miR-16 may mediate its action through the inhibition of FGF receptors or SMAD3 (68–70). MiR-16 may interact directly with HuR mRNA at its 3'UTR or with HuR protein itself. Indeed, Xu et al. showed that miR-16 decreases the expression of the pro-oncogenic HuR protein in

breast cancer by inhibiting the translatability of its mRNA *via* direct interaction within the 3'UTR of HuR transcript (25). It is also compatible with an evidence in colon cancer cells indicating disrupted miR-16 binding to its cytoplasmic targets due to miR-16/HuR interaction. Incidentally, these competitive interactions are thought to occur in the cytoplasm as weaker association between HuR and miR-16 was noted when nucleocytoplasmic trafficking of HuR was inhibited (64).

Like miR-16, miR-519 has been linked to tumor suppression. Its downregulation has been reported in laryngeal squamous cell carcinoma (23), as well as ovary (45), lung (24), and kidney cancers (44). In several cancer cell lines (ovarian, colon, and laryngeal), miR-519 was shown to inhibit cell growth and proliferation, and, in animal model, the anti-tumorigenic properties of miR-519 were demonstrated in cultured HeLa cells xenografted in athymic mice (23, 44–46). Its mechanism of action may also be mediated through HuR as two miR-519 interaction sites have been evidenced within HuR mRNA: one within the coding region, and the other in the 3'UTR (45). Possibly, much like miR-16, miR-519 may alter HuR expression by inhibiting the translation of HuR mRNA (23, 45). MiR-519 may also exert its action *via* other signaling molecules independent of HuR. Abdelmohsen et al. identified numerous miR-519 targets in addition to HuR, *via* a combination of proteome, microarray, and miR-519-mRNA interaction analyses (71). They found that miR-519 inhibits the growth and survival of tumor cells *via* repressing the expression of proteins involved in DNA maintenance (including DUT1, EXO1, RPA2, and POLE4) and intracellular calcium homeostasis (ATP2C1 and ORAI1). In this work we report that miR-519 effects on cell growth are linked with transcriptomic programs related to cancer hallmarks such as the regulation of apoptosis and hypoxia pathways, in the fashion of what we observed with HuR activity (43).

In meningioma, one additional question concerns the upstream mechanisms leading to the downregulation of both miR-16 and miR-519. Chromosomal deletions at 13q14 have been linked to miR-16 downregulation in several hematological malignancies (66, 72). DNA methylation may also participate in the inhibition of certain miR-related gene transcription (73). For example, our group recently described methylation of miR-16-linked regulatory regions as being strongly correlated with proliferation markers and indices (61). It remains unclear, however, what causes the downregulation of miR-519 in meningiomas. Despite lacking a complete understanding of the transcriptional regulation of miR-16 and miR-519 and of their downstream effects, we investigated possible prognostic values of these miRs in meningiomas by searching for correlations between miR expression, WHO histological grade, and progression-free survival. We found no such correlation. Possible explanations include that these miRs participate in tumor formation during early stages, and that their expression levels, albeit high in normal tissue, decrease in tumors to levels close to the detection limit of the chosen assay method and are therefore difficult to quantify reliably. In other types of cancers, evidence suggests that miR-16 may be of prognostic value. For example, in colorectal cancer, the 5-year overall survival rate was significantly reduced for patients with lower miR-16 expression (67, 74). Also, in investigated T lymphoblastic lymphoma/acute lymphoblastic leukemia (T-LBL/ALL) lymph node samples, authors found evidence of improved overall 1-year survival rate for patients with higher miR-16 expression levels (75).

The use of chemotherapy as an additional treatment for patients with recurrent meningioma was considered by Balik et al. The authors showed that *in vitro* chemosensitivity was most effectively obtained with cisplatin (76), which was shown to inhibit cell proliferation *via* upregulation of miR-16 in neuroblastoma both *in vivo* and *in vitro* (77). Additionally, therapeutic applications of miRs represent a novel strategy to influence clinical outcomes in cancer patients. Fujita et al. reviewed the recent trials on small RNAs, focusing on the modulation of miR levels (18). Reid et al. demonstrated that the restoration of miR-16 levels results in inhibition of growth in malignant pleural mesothelioma *in vitro* and improves antimetabolite drug sensitivity, justifying the onset of phase I clinical trials ("MesomiR-1", ClinicalTrials.gov identifier: NCT02369198) (60, 78). Here, we showed that miR-16 is an interesting candidate for miR replacement therapy in meningioma. Indeed, miR-16 experimental overexpression resulted in a significant decrease of cell growth, both in the anaplastic IOMM-Lee cell line and in the benign meningioma Ben-Men-1 cell line, showing significant effects on the cell cycle. We found that the inhibition of cell growth by miR-16 mimic is dose-dependent, this important pharmacologic property reinforcing its attractivity for therapeutic purpose. These first pre-clinical results need now to be validated with *in vivo* experimental studies. Similar to "TargomiRs", which showed interesting preliminary results in mesothelioma (78), this miR could be specifically addressed to meningioma tumor cells through vectors loaded with miR-16 mimics and targeted to receptors specifically expressed by meningioma cells (e.g., SSTR2).

Some authors have argued about the meningothelial origin of IOMM-Lee cells and whether it can be considered a realistic model of meningioma (49, 79). While these cells demonstrate specific features of malignancy, here we make extensive use of this cell line and by doing so reviewed its molecular relevance as a high-grade anaplastic cell line. Apart for the fact that we needed a fully compatible model to link and overlap our results with what we previously validated on HuR and hypoxia, we persistently uncovered dysregulated meningioma signature genes, such as *CDKN1A*, *HIF1A*, *EGFR*, *MUC1* (EMA), *NRAS*, *MMP2*, *STAT3*, *ETV6*, *MN1*, *ERCC2*, *MDM2*, *NF2* and *TP53*, some of them well known to be often associated together in meningioma. In our opinion, the upregulation of many of these genes with the miR-mimics or HuR knockdown demonstrate a profile correlating with that of well-differentiated meningiomas. Furthermore, expression of proteins like SSTR2A (somatostatin receptor 2A), EMA (epithelial membrane antigen, *MUC1* gene) and PR (progesterone receptor) is known to fit with meningioma diagnosis. Indeed, the transcriptomes of IOMM-Lee cell line, whether in control, miR-mimic or siHuR, showed expression of respective corresponding genes *SSTR2*, *MUC1* and *PGR* above the median level, with levels more than 2-fold higher basal expression in every samples.

In conclusion, the present study provides the first evidence for the downregulation of both miR-16 and miR-519 in human

meningioma. We show that the overexpression of these two miRs can independently inhibit meningioma cell growth. The data from the dose-response experiments reported here indicate that miR-16 exerts strong inhibitory effects against cell growth. We also uncover a highly specific transcriptomic signature of miR-16/miR-519-dysregulated genes, enriched in cell cycle genes and HuR targets, and confirmed on external datasets of human meningiomas, suggesting that the putative tumor suppressor effect of these miRs is mediated, at least in part, *via* HuR direct or indirect inhibition.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE98848 (produced in this work), GSE95212, GSE85135, GSE126563, GSE50641, GSE74385, GSE88721.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board DC2008-459. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization, SH, J-LG, J-MV, S-FB-H and GG; Data curation, J-MC; Formal analysis, SH, J-MC and GG; Funding acquisition, J-LG and GG; Investigation, DH, FR, LV, SH and GG; Methodology, SH, AO, DH, FR, LV, S-FB-H and GG; Project administration, J-LG and GG; Software, SH, AO and RH; Supervision, S-FB-H and GG; Writing – original draft, SH, J-MC, AO and GG; Writing – review & editing, SH, RH, J-MV, S-FB-H and GG. All authors contributed to the article and approved the submitted version.

Funding

GIRCI-EST and La Région Lorraine. 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, and in the decision to publish the results.

Acknowledgments

The authors wish to thank Drs. GILLESPIE and JENSEN (University of Utah, USA) for their generous gift of IOMM-Lee cells; GIRCI Est and La Région Lorraine for providing financial support; Mrs. Sophie LORENTZ, Mrs Justine FLAYAC (INSERM U1256), Mrs. Lydia BROCHIN and all members of the Pathology Department of Nancy (CHU) for providing technical support; Dr. Natacha DREUMONT (INSERM U1256), Dr. Rose GHEMRAWI (INSERM U1256), Dr. Jean-Marc ALBERTO (INSERM U1256), Dr. Nicolas GAMBIER and Dr. Julien SCALA-BERTOLA (Department of Clinical Pharmacology and Toxicology, University Hospital of Nancy; UMR, CNRS 7365 IMoPA) for their advices; Mrs. Nathalie NICOT and M. Tony KAOMA for transcriptomics analyses (Genomics and Proteomics, Department of Oncology, Luxembourg Institute of Health, L-1526 Luxembourg, Luxembourg).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1158773/full#supplementary-material

References

1. Baldi I, Engelhardt J, Bonnet C, Bauchet L, Berteaud E, Grüber A, et al. Epidemiology of meningiomas. *Neurochirurgie.* (2018) 64(1):5–14. doi: 10.1016/j.neuchi.2014.05.006

2. Ogasawara C, Philbrick BD, Adamson DC. Meningioma: A review of epidemiology, pathology, diagnosis, treatment, and future directions. *Biomedicines.* (2021) 9(3):319. doi: 10.3390/biomedicines9030319

3. Goldbrunner R, Stavrinou P, Jenkinson MD, Sahm F, Mawrin C, Weber DC, et al. EANO guideline on the diagnosis and management of meningiomas. *Neuro Oncol* (2021) 23(11):1821–34. doi: 10.1093/neuonc/noab150

4. Bi WL, Mei Y, Agarwalla PK, Beroukhim R, Dunn IF. Genomic and epigenomic landscape in meningioma. *Neurosurg Clin N Am* (2016) 27(2):167–79. doi: 10.1016/j.nec.2015.11.009

5. Bi WL, Zhang M, Wu WW, Mei Y, Dunn IF. Meningioma genomics: diagnostic, prognostic, and therapeutic applications. *Front Surg* (2016) 3:40. doi: 10.3389/fsurg.2016.00040

6. Nassiri F, Liu J, Patil V, Mamatjan Y, Wang JZ, Hugh-White R, et al. A clinically applicable integrative molecular classification of meningiomas. *Nature.* (2021) 597 (7874):119–25. doi: 10.1038/s41586-021-03850-3

7. Sahm F, Schrimpf D, Stichel D, Jones DTW, Hielscher T, Schefzyk S, et al. DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. *Lancet Oncol* (2017) 18(5):682–94. doi: 10.1016/S1470-2045(17)30155-9

8. Zador Z, Landry AP, Saha A, Cusimano MD. Gene expression signatures identify biologically homogenous subgroups of grade 2 meningiomas. *Front Oncol* (2020) 10:541928. doi: 10.3389/fonc.2020.541928

9. Dunn J, Lenis VP, Hilton DA, Warta R, Herold-Mende C, Hanemann CO, et al. Integration and comparison of transcriptomic and proteomic data for meningioma. *Cancers (Basel)* (2020) 12(11):3270. doi: 10.3390/cancers12113270

10. Dunn J, Ferluga S, Sharma V, Futschik M, Hilton DA, Adams CL, et al. Proteomic analysis discovers the differential expression of novel proteins and phosphoproteins in meningioma including NEK9, HK2 and SET and deregulation of RNA metabolism. *EBioMedicine.* (2019) 40:77–91. doi: 10.1016/j.ebiom.2018.12.048

11. Zador Z, Landry AP, Haibe-Kains B, Cusimano MD. Meta-gene markers predict meningioma recurrence with high accuracy. *Sci Rep* (2020) 10(1):18028. doi: 10.1038/s41598-020-74482-2

12. Patel AJ, Wan YW, Al-Ouran R, Revelli JP, Cardenas MF, Oneissi M, et al. Molecular profiling predicts meningioma recurrence and reveals loss of DREAM complex repression in aggressive tumors. *Proc Natl Acad Sci U S A.* (2019) 116 (43):21715–26. doi: 10.1073/pnas.1912858116

13. Harmanci AS, Youngblood MW, Clark VE, Coskun S, Henegariu O, Duran D, et al. Integrated genomic analyses of *de novo* pathways underlying atypical meningiomas. *Nat Commun* (2017) 8:14433. doi: 10.1038/ncomms14433

14. Clark VE, Harmancı AS, Bai H, Youngblood MW, Lee TI, Baranoski JF, et al. Recurrent somatic mutations in POLR2A define a distinct subset of meningiomas. *Nat Genet* (2016) 48(10):1253–9. doi: 10.1038/ng.3651

15. Youngblood MW, Miyagishima DF, Jin L, Gupte T, Li C, Duran D, et al. Associations of meningioma molecular subgroup and tumor recurrence. *Neuro Oncol* (2021) 23(5):783–94. doi: 10.1093/neuonc/noaa226

16. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature.* (2005) 435(7043):834–8. doi: 10.1038/nature03702

17. Chakraborty C, Sharma AR, Sharma G, Doss CGP, Lee SS. Therapeutic miRNA and siRNA: Moving from Bench to Clinic as Next Generation Medicine. *Mol Ther Nucleic Acids* (2017) 8:132–43. doi: 10.1016/j.omtn.2017.06.005

18. Fujita Y, Kuwano K, Ochiya T. Development of small RNA delivery systems for lung cancer therapy. *Int J Mol Sci* (2015) 16(3):5254–70. doi: 10.3390/ijms16035254

19. Huang Z, Chen W, Du Y, Guo Q, Mao Y, Zhou X, et al. Serum miR-16 as a potential biomarker for human cancer diagnosis: results from a large-scale population. *J Cancer Res Clin Oncol* (2019) 145(3):787–96. doi: 10.1007/s00432-019-02849-8

20. Reis PP, Drigo SA, Carvalho RF, Lopez Lapa RM, Felix TF, Patel D, et al. Circulating miR-16-5p, miR-92a-3p, and miR-451a in Plasma from Lung Cancer Patients: Potential Application in Early Detection and a Regulatory Role in Tumorigenesis Pathways. *Cancers (Basel)* (2020) 12(8):2071. doi: 10.3390/cancers12082071

21. Bonafé GA, Dos Santos JS, Ziegler JV, Umezawa K, Ribeiro ML, Rocha T, et al. Growth inhibitory effects of dipotassium glycyrrhizinate in glioblastoma cell lines by targeting microRNAs through the NF-κB signaling pathway. *Front Cell Neurosci* (2019) 13:216. doi: 10.3389/fncel.2019.00216

22. Huang X, Hou Y, Weng X, Pang W, Hou L, Liang Y, et al. Diethyldithiocarbamate-copper complex (CuET) inhibits colorectal cancer progression *via* miR-16-5p and 15b-5p/ALDH1A3/PKM2 axis-mediated aerobic glycolysis pathway. *Oncogenesis.* (2021) 10(1):4. doi: 10.1038/s41389-020-00295-7

23. Shen Z, Zhan G, Deng H, Ren Y, Ye D, Xiao B, et al. MicroRNA-519a demonstrates significant tumour suppressive activity in laryngeal squamous cells by targeting anti-carcinoma HuR gene. *J Laryngol Otol* (2013) 127(12):1194–202. doi: 10.1017/S0022215113003174

24. Vigouroux C, Casse JM, Battaglia-Hsu SF, Brochin L, Luc A, Paris C, et al. Methyl(R217)HuR and MCM6 are inversely correlated and are prognostic markers in non small cell lung carcinoma. *Lung Cancer.* (2015) 89(2):189–96. doi: 10.1016/j.lungcan.2015.05.008

25. Xu F, Zhang X, Lei Y, Liu X, Liu Z, Tong T, et al. Loss of repression of HuR translation by miR-16 may be responsible for the elevation of HuR in human breast carcinoma. *J Cell Biochem* (2010) 111(3):727–34. doi: 10.1002/jcb.22762

26. Ren L, Li Y, Zhao Q, Fan L, Tan B, Zang A, et al. miR-519 regulates the proliferation of breast cancer cells *via* targeting human antigen R. *Oncol Lett* (2020) 19 (2):1567–76. doi: 10.3892/ol.2019.11230

27. Yu G, Zhang T, Jing Y, Bao Q, Tang Q, Zhang Y. miR-519 suppresses nasopharyngeal carcinoma cell proliferation by targeting oncogene URG4/URGCP. *Life Sci* (2017) 175:47–51. doi: 10.1016/j.lfs.2017.03.010

28. Jin W, Chen F, Wang K, Song Y, Fei X, Wu B. miR-15a/miR-16 cluster inhibits invasion of prostate cancer cells by suppressing TGF-β signaling pathway. *BioMed Pharmacother.* (2018) 104:637–44. doi: 10.1016/j.biopha.2018.05.041

29. Yang R, Xu J, Hua X, Tian Z, Xie Q, Li J, et al. Overexpressed miR-200a promotes bladder cancer invasion through direct regulating Dicer/miR-16/JNK2/MMP-2 axis. *Oncogene.* (2020) 39(9):1983–96. doi: 10.1038/s41388-019-1120-z

30. Yang TQ, Lu XJ, Wu TF, Ding DD, Zhao ZH, Chen GL, et al. MicroRNA-16 inhibits glioma cell growth and invasion through suppression of BCL2 and the nuclear factor-kappaB1/MMP9 signaling pathway. *Cancer science.* (2014) 105(3):265–71. doi: 10.1111/cas.12351

31. Tian R, Wang J, Yan H, Wu J, Xu Q, Zhan X, et al. Differential expression of miR16 in glioblastoma and glioblastoma stem cells: their correlation with proliferation, differentiation, metastasis and prognosis. *Oncogene.* (2017) 36(42):5861–73. doi: 10.1038/onc.2017.182

32. Krell A, Wolter M, Stojcheva N, Hertler C, Liesenberg F, Zapatka M, et al. MiR-16-5p is frequently down-regulated in astrocytic gliomas and modulates glioma cell proliferation, apoptosis and response to cytotoxic therapy. *Neuropathol Appl Neurobiol* (2019) 45(5):441–58. doi: 10.1111/nan.12532

33. Roth P, Wischhusen J, Happold C, Chandran PA, Hofer S, Eisele G, et al. A specific miRNA signature in the peripheral blood of glioblastoma patients. *J neurochemistry.* (2011) 118(3):449–57. doi: 10.1111/j.1471-4159.2011.07307.x

34. Wang H, Pan J, Yu L, Meng L, Liu Y, Chen X. MicroRNA-16 inhibits glioblastoma growth in orthotopic model by targeting cyclin D1 and WIP1. *Onco Targets Ther* (2020) 13:10807–16. doi: 10.2147/OTT.S250369

35. Zhi F, Zhou G, Wang S, Shi Y, Peng Y, Shao N, et al. A microRNA expression signature predicts meningioma recurrence. *Int J cancer.* (2013) 132(1):128–36. doi: 10.1002/ijc.27658

36. Ludwig N, Kim YJ, Mueller SC, Backes C, Werner TV, Galata V, et al. Posttranscriptional deregulation of signaling pathways in meningioma subtypes by differential expression of miRNAs. *Neuro Oncol* (2015) 17(9):1250–60. doi: 10.1093/neuonc/nov014

37. Zhi F, Shao N, Li B, Xue L, Deng D, Xu Y, et al. A serum 6-miRNA panel as a novel non-invasive biomarker for meningioma. *Sci Rep* (2016) 6:32067. doi: 10.1038/srep32067

38. El-Gewely MR, Andreassen M, Walquist M, Ursvik A, Knutsen E, Nystad M, et al. Differentially expressed microRNAs in meningiomas grades I and II suggest shared biomarkers with malignant tumors. *Cancers* (2016) 8(3):31. doi: 10.3390/cancers8030031

39. Negroni C, Hilton DA, Ercolano E, Adams CL, Kurian KM, Baiz D, et al. GATA-4, a potential novel therapeutic target for high-grade meningioma, regulates miR-497, a potential novel circulating biomarker for high-grade meningioma. *EBioMedicine.* (2020) 59:102941. doi: 10.1016/j.ebiom.2020.102941

40. Slavik H, Balik V, Vrbkova J, Rehulkova A, Vaverka M, Hrabalek L, et al. Identification of meningioma patients at high risk of tumor recurrence using microRNA profiling. *Neurosurgery.* (2020) 87(5):1055–63. doi: 10.1093/neuros/nyaa009

41. Schultz CW, Preet R, Dhir T, Dixon DA, Brody JR. Understanding and targeting the disease-related RNA binding protein human antigen R (HuR). *Wiley Interdiscip Rev RNA.* (2020) 11(3):e1581. doi: 10.1002/wrna.1581

42. Majumder M, Chakraborty P, Mohan S, Mehrotra S, Palanisamy V. HuR as a molecular target for cancer therapeutics and immune-related disorders. *Adv Drug Delivery Rev* (2022) 188:114442. doi: 10.1016/j.addr.2022.114442

43. Gauchotte G, Hergalant S, Vigouroux C, Casse JM, Houlgatte R, Kaoma T, et al. Cytoplasmic overexpression of RNA-binding protein HuR is a marker of poor prognosis in meningioma, and HuR knockdown decreases meningioma cell growth and resistance to hypoxia. *J Pathol* (2017) 242(4):421–34. doi: 10.1002/path.4916

44. Abdelmohsen K, Kim MM, Srikantan S, Mercken EM, Brennan SE, Wilson GM, et al. miR-519 suppresses tumor growth by reducing HuR levels. *Cell Cycle* (2010) 9 (7):1354–9. doi: 10.4161/cc.9.7.11164

45. Abdelmohsen K, Srikantan S, Kuwano Y, Gorospe M. miR-519 reduces cell proliferation by lowering RNA-binding protein HuR levels. *Proc Natl Acad Sci U S A.* (2008) 105(51):20297–302. doi: 10.1073/pnas.0809376106

46. Ristimaki A. Tumor suppressor effect of the microRNA miR-519 is mediated *via* the mRNA-binding protein HuR. *Cell Cycle* (2010) 9(7):1234. doi: 10.4161/cc.9.7.11322

47. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Ellison DWWHO. *WHO Classification of Tumours of the Central Nervous System, Revised. 4th ed.* Lyon: IARC Press (2016).

48. Lee WH. Characterization of a newly established malignant meningioma cell line of the human brain: IOMM-Lee. *Neurosurgery.* (1990) 27(3):389–95. doi: 10.1097/00006123-199009000-00008

49. Puttmann S, Senner V, Braune S, Hillmann B, Exeler R, Rickert CH, et al. Establishment of a benign meningioma cell line by hTERT-mediated immortalization. *Lab Invest.* (2005) 85(9):1163–71. doi: 10.1038/labinvest.3700307

50. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* (2002) 30(4):e15. doi: 10.1093/nar/30.4.e15

51. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007

52. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics.* (2004) 20(9):1453–4. doi: 10.1093/bioinformatics/bth078

53. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. *Curr Protoc* (2021) 1(3):e90. doi: 10.1002/cpz1.90

54. Budczies J, Klauschen F, Sinn BV, Gyorffy B, Schmitt WD, Darb-Esfahani S, et al. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PloS One* (2012) 7(12):e51862. doi: 10.1371/journal.pone.0051862

55. Simpson D. The recurrence of intracranial meningiomas after surgical treatment. *J Neurol Neurosurg Psychiatry* (1957) 20(1):22–39. doi: 10.1136/jnnp.20.1.22

56. Dalan AB, Gulluoglu S, Tuysuz EC, Kuskucu A, Yaltirik CK, Ozturk O, et al. Simultaneous analysis of miRNA-mRNA in human meningiomas by integrating transcriptome: A relationship between PTX3 and miR-29c. *BMC Cancer.* (2017) 17 (1):207. doi: 10.1186/s12885-017-3198-4

57. Shi L, Jiang D, Sun G, Wan Y, Zhang S, Zeng Y, et al. miR-335 promotes cell proliferation by directly targeting Rb1 in meningiomas. *J Neurooncol.* (2012) 110 (2):155–62. doi: 10.1007/s11060-012-0951-z

58. Saydam O, Shen Y, Wurdinger T, Senol O, Boke E, James MF, et al. Downregulated microRNA-200a in meningiomas promotes tumor growth by reducing E-cadherin and activating the Wnt/beta-catenin signaling pathway. *Mol Cell Biol* (2009) 29(21):5923–40. doi: 10.1128/MCB.00332-09

59. Bandi N, Zbinden S, Gugger M, Arnold M, Kocher V, Hasan L, et al. miR-15a and miR-16 are implicated in cell cycle regulation in a Rb-dependent manner and are frequently deleted or down-regulated in non-small cell lung cancer. *Cancer Res* (2009) 69(13):5553–9. doi: 10.1158/0008-5472.CAN-08-4277

60. Reid G, Pel ME, Kirschner MB, Cheng YY, Mugridge N, Weiss J, et al. Restoring expression of miR-16: a novel approach to therapy for malignant pleural mesothelioma. *Ann Oncol Off J Eur Soc Med Oncol* (2013) 24(12):3128–35. doi: 10.1093/annonc/mdt412

61. Hergalant S, Saurel C, Divoux M, Rech F, Pouget C, Godfraind C, et al. Correlation between DNA methylation and cell proliferation identifies new candidate predictive markers in meningioma. *Cancers (Basel)* (2022) 14(24):6227. doi: 10.3390/cancers14246227

62. Zhou P, Gu F, Zhang L, Akerberg BN, Ma Q, Li K, et al. Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *Elife* (2017) 6:e22039. doi: 10.7554/eLife.22039

63. Durbin AD, Wang T, Wimalasena VK, Zimmerman MW, Li D, Dharia NV, et al. EP300 selectively controls the enhancer landscape of MYCN-amplified neuroblastoma. *Cancer Discovery* (2022) 12(3):730–51. doi: 10.1158/2159-8290.CD-21-0385

64. Young LE, Moore AE, Sokol L, Meisner-Kober N, Dixon DA. The mRNA stability factor HuR inhibits microRNA-16 targeting of COX-2. *Mol Cancer Res MCR.* (2012) 10(1):167–80. doi: 10.1158/1541-7786.MCR-11-0337

65. Navarro A, Diaz T, Gallardo E, Vinolas N, Marrades RM, Gel B, et al. Prognostic implications of miR-16 expression levels in resected non-small-cell lung cancer. *J Surg Oncol* (2011) 103(5):411–5. doi: 10.1002/jso.21847

66. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci United States America.* (2002) 99 (24):15524–9. doi: 10.1073/pnas.242606799

67. Bottoni A, Piccin D, Tagliati F, Luchin A, Zatelli MC, degli Uberti EC. miR-15a and miR-16-1 down-regulation in pituitary adenomas. *J Cell Physiol* (2005) 204 (1):280–5. doi: 10.1002/jcp.20282

68. Andriani F, Majorini MT, Mano M, Landoni E, Miceli R, Facchinetti F, et al. MiR-16 regulates the pro-tumorigenic potential of lung fibroblasts through the inhibition of HGF production in an FGFR-1- and MEK1-dependent manner. *J Hematol Oncol* (2018) 11(1):45. doi: 10.1186/s13045-018-0594-4

69. Schelch K, Kirschner MB, Williams M, Cheng YY, van Zandwijk N, Grusch M, et al. A link between the fibroblast growth factor axis and the miR-16 family reveals potential new treatment combinations in mesothelioma. *Mol Oncol* (2018) 12(1):58–73. doi: 10.1002/1878-0261.12150

70. Zhu C, Huang Q, Zhu H. Melatonin inhibits the proliferation of gastric cancer cells through regulating the miR-16-5p-smad3 pathway. *DNA Cell Biol* (2018) 37 (3):244–52. doi: 10.1089/dna.2017.4040

71. Abdelmohsen K, Srikantan S, Tominaga K, Kang MJ, Yaniv Y, Martindale JL, et al. Growth inhibition by miR-519 *via* multiple p21-inducing pathways. *Mol Cell Biol* (2012) 32(13):2530–48. doi: 10.1128/MCB.00510-12

72. Lovat F, Fassan M, Gasparini P, Rizzotto L, Cascione L, Pizzi M, et al. miR-15b/16-2 deletion promotes B-cell malignancies. *Proc Natl Acad Sci United States America.* (2015) 112(37):11636–41. doi: 10.1073/pnas.1514954112

73. Suzuki H, Maruyama R, Yamamoto E, Kai M. DNA methylation and microRNA dysregulation in cancer. *Mol Oncol* (2012) 6(6):567–78. doi: 10.1016/j.molonc.2012.07.007

74. Qian J, Jiang B, Li M, Chen J, Fang M. Prognostic significance of microRNA-16 expression in human colorectal cancer. *World J Surg* (2013) 37(12):2944–9. doi: 10.1007/s00268-013-2205-4

75. Xi Y, Li J, Zan L, Wang J, Wang G, Ning Y. Micro-RNA-16 expression in paraffin-embedded specimen correlates with overall survival of T-lymphoblastic lymphoma/leukemia. *Hum Pathol* (2013) 44(6):1011–6. doi: 10.1016/j.humpath.2012.08.023

76. Balik V, Sulla I, Park HH, Sarissky M. *In vitro* testing to a panel of potential chemotherapeutics and current concepts of chemotherapy in benign meningiomas. *Surg Oncol* (2015) 24(3):292–9. doi: 10.1016/j.suronc.2015.06.004

77. Sun YX, Yang J, Wang PY, Li YJ, Xie SY, Sun RP. Cisplatin regulates SH-SY5Y cell growth through downregulation of BDNF *via* miR-16. *Oncol Rep* (2013) 30 (5):2343–9. doi: 10.3892/or.2013.2731

78. van Zandwijk N, Pavlakis N, Kao SC, Linton A, Boyer MJ, Clarke S, et al. Safety and activity of microRNA-loaded minicells in patients with recurrent malignant pleural mesothelioma: a first-in-man, phase 1, open-label, dose-escalation study. *Lancet Oncol* (2017) 18(10):1386–96. doi: 10.1016/S1470-2045(17)30621-6

79. Mei Y, Bi WL, Greenwald NF, Agar NY, Beroukhim R, Dunn GP, et al. Genomic profile of human meningioma cell lines. *PloS One* (2017) 12(5):e0178322. doi: 10.1371/journal.pone.0178322

PARTIE 2 : GÉNOMIQUE DES LYMPHOMES

4. MOLECULAR CHARACTERIZATION OF RICHTER SYNDROME IDENTIFIES DE NOVO DIFFUSE LARGE B-CELL LYMPHOMAS WITH POOR PROGNOSIS

Contexte

Ce manuscrit sur la caractérisation épigénomique très exhaustive du syndrome de Richter, de la méthylation de l'ADN à l'expression des gènes, est l'aboutissement d'un effort considérable de plusieurs années. Il synthétise en le diversifiant un savoir-faire en génomique et bioinformatique, avec contrôles qualité et méthodologies rigoureuses, approches exploratoires non supervisées et observations empiriques, représentation des données complexes à l'échelle du génome, et vision biostatistique support qui valide les résultats tout en les formalisant. Le travail débouche sur l'élaboration de classifieurs généralisables à de vastes panels de lymphomes agressifs et la réanalyse de jeux de données publiques repères, reconnus pour leur apport, leur qualité, et leur taille conséquente.

Etat de l'art

Les lymphomes regroupent l'ensemble des transformations cancéreuses des lymphocytes ; ils représentent aujourd'hui la 6ème cause de cancer en France en termes d'incidence. Ils se manifestent le plus souvent par l'apparition de ganglions pathologiques (de taille augmentée), infiltrés par des lymphocytes anormaux. Ici nous travaillons plus spécifiquement sur le syndrome de Richter (SR), qui constitue la transformation d'une leucémie lymphoïde chronique (LLC) en un lymphome B diffus à grandes cellules (LBDGC).

Le SR est la forme la plus grave d'évolution d'un lymphome à petites cellules en un lymphome agressif à grandes cellules, avec une survie médiane des patients inférieure à 12 mois en raison de sa forte agressivité et d'un profil étendu de chimiorésistance. Les analyses cytogénétiques et moléculaires de l'ADN ont permis de dresser avec précision le profil de cette évolution dramatique. Celui-ci regroupe les anomalies parmi les plus graves décrites dans les lymphomes agressifs. En revanche, peu d'études ont exploré les aspects épigénétiques du SR, en particulier les mécanismes de dérégulation des programmes cellulaires.

En 2023, le traitement du SR repose toujours sur des polychimiothérapies, le plus souvent peu efficaces. De nouveaux moyens thérapeutiques sont indispensables, notamment des immunothérapies ou des traitements ciblant plus spécifiquement des récepteurs ou des voies de signalisation intracellulaire impliquées dans le développement de la maladie.

Méthodologies

Notre cohorte est constituée de 64 SR dont 25 sont appariés par suivi à la LLC, de 240 LLC, de 96 LBDGC et de 102 échantillons sains couvrant le développement normal des lymphocytes B. Plusieurs aspects omiques permettent de caractériser finement les mécanismes de la maladie : mutations de l'ADN par exome ciblé, méthylation de l'ADN par puces de dernière génération (Illumina® 850k), états de la chromatine et modifications post-traductionnelles des histones par ChIP-seq pour 7 épigénomes de référence de LLC, et profils d'expression génique par RNA-seq.

Une déconvolution cellulaire est nécessaire pour focaliser le signal de méthylation sur les lymphocytes B car la LLC est circulante et les échantillons sont prélevés dans le sang, alors que les lymphomes agressifs sont obtenus à partir de biopsies de ganglions lymphatiques. Les effets *batch* sont corrigés en observant la dérive des échantillons réplicats (par réduction de dimension) d'une technologie de puce à une autre et d'un laboratoire à un autre. La pureté tumorale et la qualité des échantillons sont scrutées de bout en bout pour nous assurer qu'elles ne sont pas en lien avec nos observations ou le fonctionnement des modèles statistiques.

Nous construisons des scores cumulatifs, avec cohortes d'apprentissage et cohorte de validation, en nous appuyant sur des signatures et ensembles de variables informatives extraites depuis le méthylome, le transcriptome, et l'intégrome entre régions régulatrices et expression des gènes sous le contrôle de ces régions.

Résultats

Ces travaux permettent d'identifier, dans des cohortes publiées de LBDGC, des lymphomes agressifs présentant un profil moléculaire et un tableau clinique comparable à celui du SR. Ils expliquent en outre jusqu'à 50% de cas supplémentaires de lymphomes qui étaient jusqu'à présent inclassifiables génomiquement, et parmi les pires en termes de pronostic.

Nous établissons une signature transcriptomique spécifique et lui associons un score statistique permettant de classifier linéairement les LBDGC en fonction de la survie et de la progression de la maladie. Ceci indépendamment des profils moléculaires et variables cliniques connus pour être en lien avec la survie dans les LBDGC.

Nous proposons également une meilleure caractérisation des mécanismes biologiques du développement de la maladie. Consécutivement, nous mettons au point un classifieur utile en clinique pour le diagnostic et une prise de décision thérapeutique rapide, en établissant clairement la relation clonale du SR à la LLC sous-jacente sans pour cela avoir besoin d'échantillons au stade LLC.

Conclusion

Ces travaux permettent de retracer la trajectoire de l'évolution tumorale et d'anticiper le profil de chimiorésistance des SR. Nous proposons des thérapies ciblées et adaptées si elles existent, et orientons les recherches vers de nouveaux traitements en proposant une panoplie de biomarqueurs et de voies dérégulés dans cette maladie.

Nous construisons des scores facilement réutilisables et identifiant spécifiquement les profils de pronostic sombre au sein du vaste panorama des lymphomes agressifs, proposons un classifieur substitutif permettant de déterminer la relation clonale entre SR et LLC sans connaissance préalable du stade LLC, et reconstruisons bioinformatiquement les réseaux de dérégulations épigénétiques pour désigner les principaux déterminants de la maladie et de potentielles cibles thérapeutiques.

nature communications

Article

https://doi.org/10.1038/s41467-022-34642-6

Molecular characterization of Richter syndrome identifies de novo diffuse large B-cell lymphomas with poor prognosis

Received: 4 February 2022

Accepted: 1 November 2022

Published online: 19 January 2023

Check for updates

Julien Broséus[1,2,3,32] ✉, Sébastien Hergalant[2,32], Julia Vogt[4], Eugen Tausch[1], Markus Kreuz[5], Anja Mottok[4], Christof Schneider[1], Caroline Dartigeas[6], Damien Roos-Weil[7], Anne Quinquenel[8], Charline Moulin[9,10], German Ott[11], Odile Blanchet[12], Cécile Tomowiak[13,14], Grégory Lazarian[15], Pierre Rouyer[2], Emil Chteinberg[4], Stephan H. Bernhart[16], Olivier Tournilhac[17], Guillaume Gauchotte[2,18], Sandra Lomazzi[19], Elise Chapiro[20,21], Florence Nguyen-Khac[20,21], Céline Chery[2,22], Frédéric Davi[21,23], Mathilde Hunault[24], Rémi Houlgatte[2], Andreas Rosenwald[25], Alain Delmer[8], David Meyre[2], Marie-Christine Béné[26,27], Catherine Thieblemont[28], Peter Lichter[29], Ole Ammerpohl[4], Jean-Louis Guéant[2,22], ICGC MMML-Seq Consortium*, Romain Guièze[17], José Ignacio Martin-Subero[30,31], Florence Cymbalista[15], Pierre Feugier[2,9,33], Reiner Siebert[4,33] & Stephan Stilgenbauer[1,33] ✉

Richter syndrome (RS) is the transformation of chronic lymphocytic leukemia (CLL) into aggressive lymphoma, most commonly diffuse large B-cell lymphoma (DLBCL). We characterize 58 primary human RS samples by genome-wide DNA methylation and whole-transcriptome profiling. Our comprehensive approach determines RS DNA methylation profile and unravels a CLL epigenetic imprint, allowing CLL-RS clonal relationship assessment without the need of the initial CLL tumor DNA. DNA methylation- and transcriptomic-based classifiers were developed, and testing on landmark DLBCL datasets identifies a poor-prognosis, activated B-cell-like DLBCL subset in 111/1772 samples. The classification robustly identifies phenotypes very similar to RS with a specific genomic profile, accounting for 4.3-8.3% of de novo DLBCLs. In this work, RS multi-omics characterization determines oncogenic mechanisms, establishes a surrogate marker for CLL-RS clonal relationship, and provides a clinically relevant classifier for a subset of primary "RS-type DLBCL" with unfavorable prognosis.

Chronic lymphocytic leukemia (CLL) is the most frequent leukemia in Western countries[1]. While generally considered an indolent B cell disease, CLL is in fact associated with a highly heterogeneous clinical course. CLLs are classified into two major molecular subtypes that differ in their degree of somatic hypermutations in the immunoglobulin heavy chain variable (IGHV) domains. IGHV-unmutated CLLs (U-CLL) are associated with an inferior prognosis than IGHV-mutated CLLs (M-CLL)[2]. CLL transformation into a more aggressive histology is

A full list of affiliations appears at the end of the paper. *A list of authors and their affiliations appears at the end of the paper.
✉e-mail: julien.broseus@univ-lorraine.fr; Stephan.Stilgenbauer@uniklinik-ulm.de

Article
https://doi.org/10.1038/s41467-022-34642-6

termed Richter syndrome (RS)[3]. Diffuse large B cell lymphoma (DLBCL) subtype accounts for 90–95% of RS cases. Around 80% of RS cases are IGHV-unmutated while the remainder are IGHV-mutated[4]. In contrast, most de novo DLBCL (from now on called DLBCL) are IGHV-mutated as they originate from germinal center (GC) or post-GC B cells. Based on gene expression patterns, different cell-of-origin (COO) derivations of DLBCL include GC B cell like (GCB) and activated B cell-like (ABC) DLBCL[5]. Recent genomic studies combining DNA and RNA sequencing extended DLBCL subtyping beyond COO[6–10], identifying DLBCL subgroups defined by their genomic alteration patterns and associated clinical courses, but a notable proportion remains unclassified[7,8,10]. Moreover, although studies have shown some extent of association of genetically defined groups with transcriptionally defined COO signatures, the transcriptome in its entirety is not fully used in current classifications.

As compared to other lymphoid malignancies, the availability of in vitro or in vivo models to study RS is limited[11–15], and therefore our current knowledge on RS biology remains incomplete. The few genomic studies attempting to decipher oncogenic mechanisms underlying RS described disabled DNA damage response and cell cycle control through *TP53* abnormalities and *CDKN2A* deletions, chronic B cell receptor (BCR) signaling, and NOTCH, MYC, and MAPK pathway deregulations[16–20]. A recent report using multiome and single cell approaches in sequential CLL-RS samples describes that the increased molecular complexity of RS does not seem to be the consequence of clonal evolution over time but rather the selection of minute subclones present at CLL diagnosis and years before overt transformation[21]. Additionally, recent studies focusing on DNA methylation (DNAm) further captured the genomic complexity of CLL[22–26], RS[27], de novo DLBCL[28–30], and other B cell neoplasms[31–33]. A better understanding of epigenetic signatures is needed, whether related to B cell development or tumor transformation mechanisms.

Distinguishing between CLL-derived RS and de novo DLBCL in a diagnostic setting based on histology and immunochemistry alone is challenging. Around 80% of RS cases are clonally related to the CLL disease stage while the remainder are unrelated (i.e. independent de novo DLBCL). This dichotomy is of importance for treatment decisions. De novo DLBCLs are chemosensitive in most patients, whereas CLL-derived RS is mainly characterized by chemoresistance and poor outcome, with a median overall survival (OS) of around 12 months.

In this study, we perform genome-wide DNAm analysis and whole-transcriptome profiling for a large series of primary human RS samples, and comprehensively compare our findings to those in CLL and DLBCL. We extensively characterize the epigenetic architecture of the RS samples and find the majority retain a CLL imprint. Remarkably, applying DNAm- and gene expression-based classifiers to datasets from landmark studies identifies a subset of "RS-type" DLBCL that is not previously described at the genomic level, is enriched in cases with an ABC-like COO signature, and has an unfavorable prognosis [7,8,10].

Results

Data quality controls

The study workflow is described in Fig. 1. We investigated DNAm using array-based technologies, exploring a total of 433 samples, including 58 RS samples, 25 CLLs paired with RS (i.e. tumor samples were available at both CLL and RS stages; hereafter "paired-CLLs"), 68 DLBCLs, and additional published methylomes from 190 other CLLs, and 92 samples representing normal B cell subpopulations (Supplementary Fig. 1)[22,25,26,34,35]. Limiting the batch effect is critical for comparing large cohorts explored with different platforms in different facilities. In this regard, we used EPIC and 450K Illumina microarray platforms, as these provide accurate, robust, and reproducible genome-wide coverage of CpG sites[36,37]. We extensively explored potential batch effects and showed it was completely removed after applying strict quality

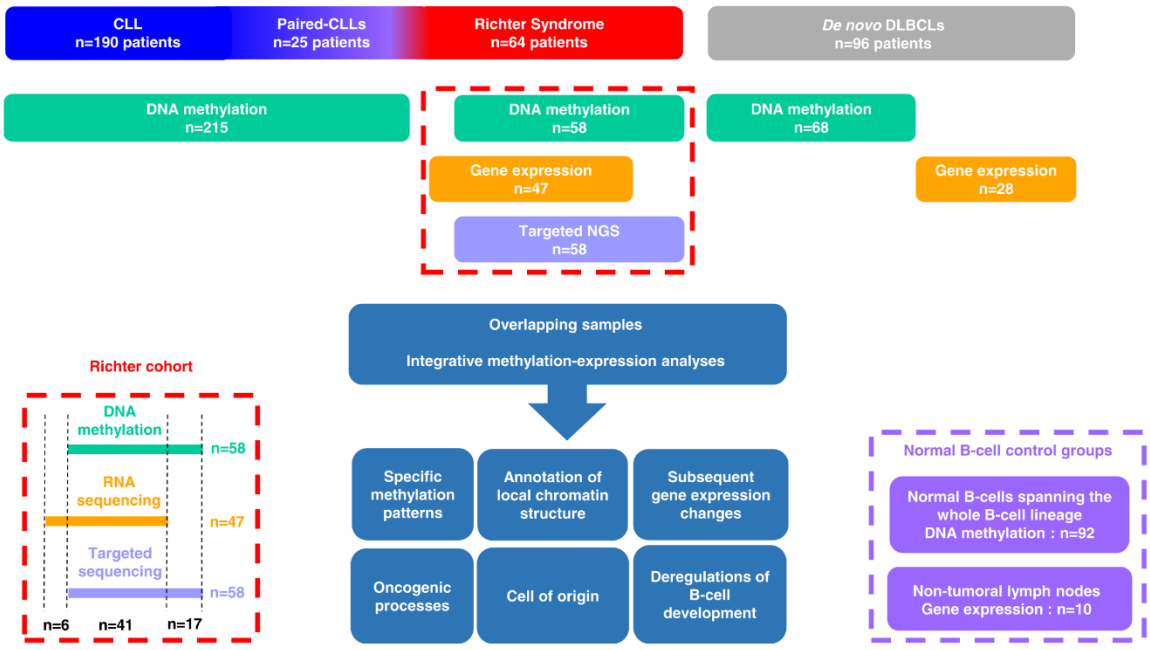
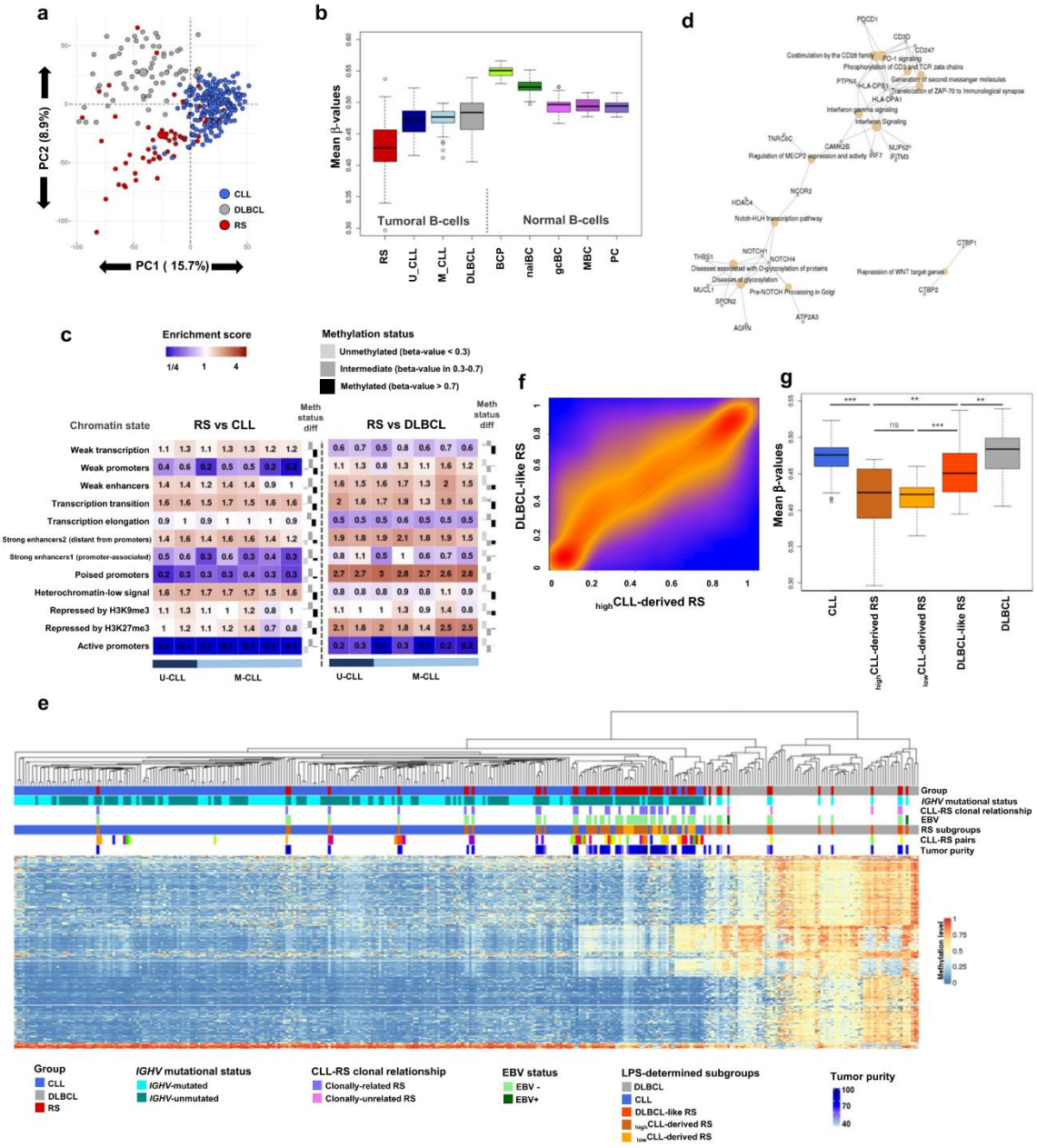


Fig. 1 | Study workflow. Genome-wide DNA methylation data were available for 58 RS, 25 CLLs paired with RS (tumor DNA samples were available at both CLL and RS stages), 190 other CLLs, 68 de novo DLBCLs, and 92 samples from normal B cells spanning the entire B lineage. All 58 RS samples were also documented for mutations in a custom panel of 13 CLL driver genes, and RNA-sequencing data were concomitantly available for 41 RS samples, allowing integrative analyses and detailed exploration of oncogenic processes and epigenetic network deregulations. RNA sequencing data were obtained for another 6 RS, 28 de novo DLBCLs, and 10 non-tumoral lymph nodes. Data acquired from normal B cell control groups were used for methodologic purposes only (see "Methods"). CLL chronic lymphocytic leukemia, DLBCL de novo diffuse large B cell lymphoma, NGS next-generation sequencing, RS Richter syndrome.

Article
https://doi.org/10.1038/s41467-022-34642-6



controls (see "Methods"). In addition, we applied a bioinformatic deconvolution method to separate methylation data attributable to five subtypes of normal white blood cells (CD4+ T-lymphocytes, CD8+ T-lymphocytes, neutrophils, monocytes, B cells). Use of respective cell composition data as covariates in supervised analyses limited the influence of tumor cell content of our samples.

RS is a DNA hypomethylated entity versus CLL and de novo DLBCL

Unsupervised principal component analysis (PCA) showed a clear partitioning between RS, CLL, and DLBCL samples in the most variable components, highlighting different DNAm patterns in each group (Fig. 2a and Supplementary Figs. 2 and 3). Principal component 1 separated CLL from RS and DLBCL, while principal component 2 separated DLBCL from RS. However, some RS clustered within the DLBCLs or the CLLs. Decreased DNAm was observed in RS compared to CLL, DLBCL, and normal B cells (Fig. 2b and Supplementary Figs. 4 and 5). DNAm levels of the paired-CLLs were intermediate between RS and the other CLLs (Fig. S6). Hypomethylated and hypermethylated CpGs in RS were differentially distributed regarding CpG islands but similarly distributed regarding genomic context (Supplementary Figs. 7 and 8).

Next, we annotated CpGs differentially methylated between RS, CLL, and DLBCL according to 12 chromatin states reported in 7 CLL reference epigenomes[26]. The 102,614 CpGs differentially methylated between RS and CLL (two-way moderated *t* test adjusted for a false

Article https://doi.org/10.1038/s41467-022-34642-6

Fig. 2 | DNA methylation comparative analysis with CLL and de novo DLBCL shows that RS is a heterogeneous and hypomethylated entity. a Unsupervised principal component analysis of the adjusted DNAm values of RS, CLL, and DLBCL. Geometrical centers are represented by bigger circles of the same color. **b** Boxplots of sample-averaged methylation levels with all 397,769 CpGs. RS ($n = 58$) versus U-CLL ($n = 112$): $p = 7.74e{-}11$; RS versus M-CLL ($n = 103$): $p = 4.46e{-}12$); RS versus DLBCL ($n = 68$): $p = 6.07e{-}12$. **c** Distribution of differential CpGs (FDR < 0.01; methylation differential >10%) according to the reported chromatin states in 7 CLL reference epigenomes[26]. Enrichments are shown as a heatmap and were calculated from the position of the selected CpGs. Their distribution was reported among 12 different chromatin state categories. Barplots in the right part of each panel show the methylation status difference in RS versus CLL or DLBCL. Differentially methylated CpGs are distributed among 3 methylation level categories. Upward bars indicate a comparative gain of CpGs in RS for the corresponding category, while downward bars indicate a comparative loss in RS. **d** RS versus CLL top annotations network (ReactomePA) from 238 differential DMRs computed with DMRcate (Fisher's multiple comparison statistics: min_smoothed_FDR and HMFDR both <0.01; max beta-value differential >30%; at least 3 CpGs in the DMR with no gap >1 kb between CpGs). **e** DNAm-based linear predictor score (LPS) CpG architecture. Hierarchical clustering of 4863 CpGs differential between CLL and DLBCL (FDR < 0.01; beta-value differential >30%; moderated *t* test). **f** Density map of DNAm between $_{high}$CLL-derived and DLBCL-like RS. Smoothed beta-value densities from the EPIC dataset. Scale from blue (no density) to yellow (medium density) and red (high density). **g** Boxplots showing general methylation levels for $_{high}$CLL-derived ($n = 33$), $_{low}$CLL-derived ($n = 12$), and DLBCL-like RS ($n = 13$), de novo DLBCLs ($n = 68$), and CLLs ($n = 215$). CLL versus $_{high}$CLL-derived RS: $p = 2.2e{-}16$; $_{high}$CLL-derived RS versus DLBCL-like RS: $p = 5e{-}3$; $_{low}$CLL-derived RS versus DLBCL-like RS: $p = 9.9e{-}3$; DLBCL-like RS versus DLBCL: $p = 3.5e{-}2$. BCP B cell precursors, CLL chronic lymphocytic leukemia, DLBCL de novo diffuse large B cell lymphoma, DNAm DNA methylation, EBV Epstein–Barr virus, FDR false discovery rate, gcBC germinal center B cells, $_{high}$CLL-derived RS CLL-derived RS with a high LPS, HMFDR harmonic mean of the individual components FDR, MBC memory B cells, M-CLL IGHV-mutated CLL, $_{low}$CLL-derived RS CLL-derived RS with a LPS score below threshold, LPS linear predictor score, naiBC naive B cells, PC plasma cells, PC1/2 principal component 1/2, RS Richter syndrome, U-CLL IGHV-unmutated CLL. *p* values were derived from two-sided *t* tests. $^{**}p < 0.01$; $^{***}p < 0.001$; ns not significant. For all box plots, center line indicates median; box limits indicate upper and lower quartiles; whiskers indicate 1.5× interquartile range; points indicate outliers. Source data are provided as a Source data file.

discovery rate (FDR) < 0.01; 90.8% hypomethylations in RS) were: depleted (ratio < 0.75) in active promoters, poised promoters, promoter-associated strong enhancers, and weak promoters; and enriched (ratio > 1.5) in transcription transition regions and heterochromatin (Fig. 2c). The 82,940 CpGs differentially methylated between RS and DLBCL (96.4% hypomethylations in RS) were: depleted in active promoters; and enriched in poised promoters and regions repressed by H3K27me3. Differentially methylated regions (DMRs; see "Methods") between RS and DLBCL were strongly enriched in targets of polycomb complex components SUZ12 ($p = 1.2e{-}121$) and EZH2 ($p = 1.5e{-}30$), which likely corresponds to the derivation of DLBCL from GC or post-GC B cells. Notably, genes associated with the extracellular matrix were overrepresented in this subset (Supplementary Fig. 9 and Supplementary Data 1). DMRs between RS and CLL were linked to NOTCH and Wnt pathways, and to the adaptive immune system, with PD-1 signaling and T cell/B cell co-stimulations (Fig. 2d and Supplementary Data 2), which likely corresponds to the driver role of NOTCH and PD-1 signaling in RS onset.

DNA methylation separates CLL-derived and DLBCL-like RS subgroups

The PCA principal component 2 split the RS samples into two subgroups, one with a profile similar to CLL, the other closer to DLBCL (Fig. 2a). We postulated that "CLL-derived RS" (maintaining a CLL imprint) could be separated from "DLBCL-like RS" (distinct from the preceding CLL and closer to DLBCL). To test this, we modeled a linear predictor score (LPS)[38], computing two underlying probabilities (p): one to label samples according to their CLL-derived RS profile ($p_{CLL\text{-}derived}$), one for DLBCL-like RS ($p_{DLBCL\text{-}like}$), defining $p_{CLL\text{-}derived} \geq 98\%$ and $p_{DLBCL\text{-}like} \geq 98\%$ to obtain highly specific and homogeneous groups (see "Methods"; Supplementary Fig. 10). The statistical model devised to compute LPS was constructed with 4863 CpGs robust in separating CLL from DLBCL. Since de novo DLBCLs are usually IGHV-mutated whereas CLL may be IGHV-mutated or -unmutated, we excluded CpGs highly differential according to IGHV status[22,24] from the LPS calculation to focus on other distinctive features between CLL and DLBCL. The LPS scoring system was confirmed with hierarchical clustering (Fig. 2e), non-negative matrix factorization (NMF), PCA (Supplementary Figs. 11 and 12), and displayed differential patterns on normal cells spanning the B cell lineage (Supplementary Fig. 13). The scoring system identified 33 CLL-derived RS (57%) and 13 DLBCL-like RS (22%), leaving 12 intermediate samples (21%). This latter subgroup clustered within the CLL and CLL-derived RS branch, albeit marginally (Fig. 2e). The subgroup was then referred to as "low-LPS score CLL-derived RS" ($_{low}$CLL-derived RS), in contrast to the "high-LPS score CLL-derived RS" ($_{high}$CLL-derived RS). Comparing $_{high}$CLL-derived RS and DLBCL-like RS confirmed global hypomethylation of $_{high}$CLL-derived RS. In addition, DLBCL-like RS genomic distribution of DNAm did not coincide with that of DLBCL, with most locations hypomethylated in DLBCL-like RS (Fig. 2f, g and Supplementary Fig. 14). This subgrouping was not influenced by the tumor cell content (Supplementary Fig. 10).

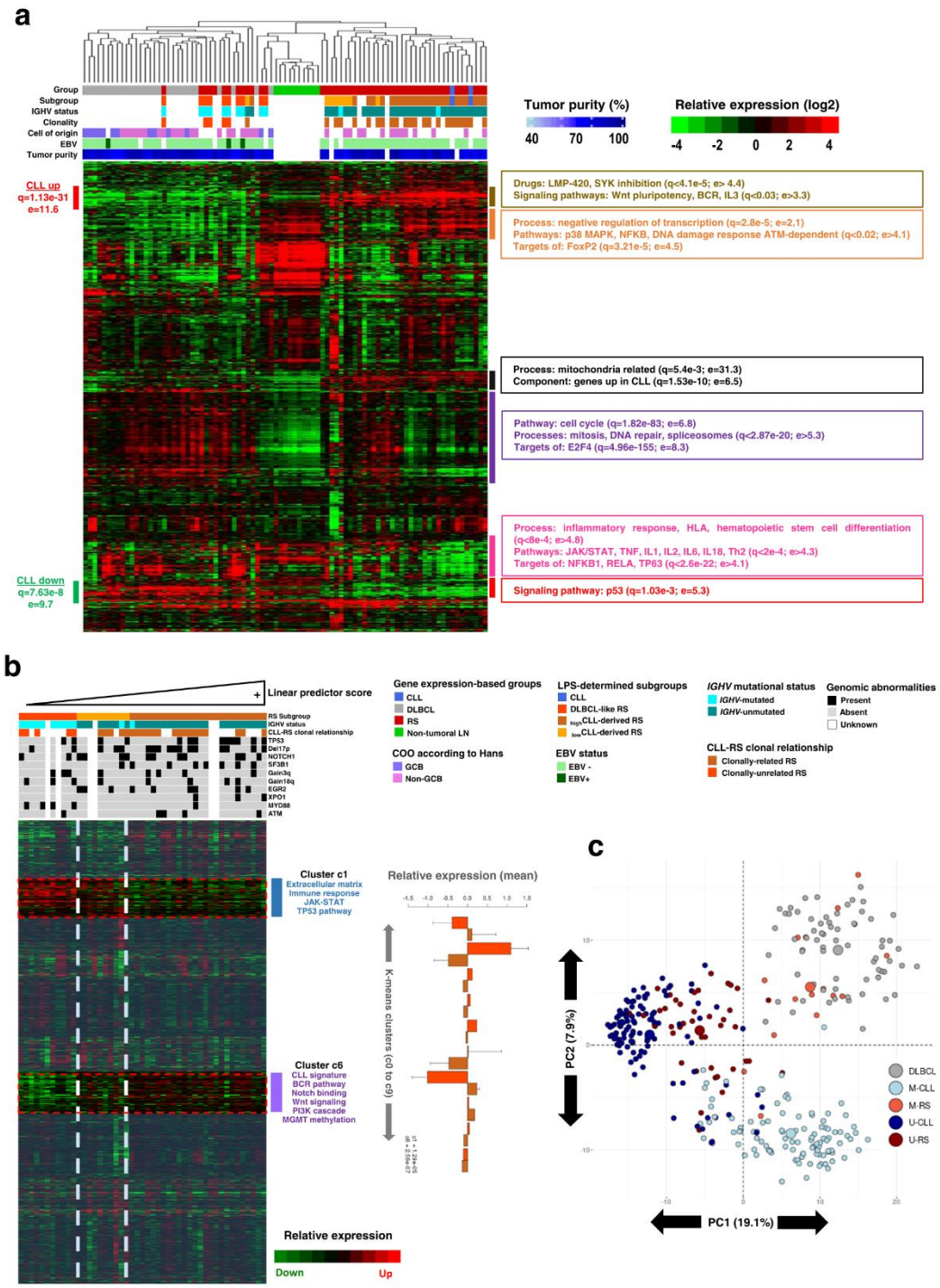
RS homogeneous subgrouping corroborates with gene expression

Among the 58 RS samples investigated for DNAm, 41 also underwent whole-transcriptome profiling. RNA samples from 6 independent RS cases were also sequenced. In total, the RNA-sequencing experiment included lymph node samples of 47 RS, 2 paired CLLs, and 28 DLBCLs, plus 10 non-tumoral samples for methodologic validation purposes (see "Methods"). Hierarchical clustering of the 23,508 identified genes confirmed clear subgrouping among RS samples (Fig. 3a). All RS classified as DLBCL-like RS by DNAm clustered with DLBCL (predominantly with non-GCB subtype) and separated from CLL-derived RS. This supports the existence of CLL-derived RS and DLBCL-like RS, through cross-validation using an orthogonal technique (>95% concordance). Annotations of gene clusters showed that CLL-derived RS shared a solid CLL gene expression signature, with upregulated genes involved in the BCR pathway and downregulated genes involved in the immune response, p53-signaling, and JAK-STAT pathways. Furthermore, *K*-means gene clustering of the 47 RS samples ranked according to LPS gradient revealed two main clusters of differentially expressed genes between $_{high}$CLL-derived and DLBCL-like RS (Fig. 3b). One cluster is downregulated in $_{high}$CLL-derived RS, is related to the extracellular matrix and TLR signaling, and included methylation-regulated p53 activity as an interesting feature (Supplementary Data 3). The other cluster is reminiscent of a CLL signature, overexpressed in $_{high}$CLL-derived RS, and linked with NOTCH, PI3K signaling, and DNAm metabolism (Supplementary Data 4).

RS subgroups correlate with IGHV mutational status and CLL-RS clonal relationship

To reduce the influence of IGHV mutational status on LPS, CpGs highly differential between U-CLL and M-CLL were filtered from the scoring CpGs. However, IGHV mutational status is associated with major DNAm changes in CLL[22,24,34]. Therefore, we next performed PCA on the 10,000 most variable CpGs, whether associated or not with GC reaction, tagging samples with IGHV annotations (Fig. 3c). CLL-derived RS accounted for nearly 80% of our RS samples and displayed a high

Article
https://doi.org/10.1038/s41467-022-34642-6



prevalence of IGHV-unmutated samples. In contrast, 12/13 (93%) DLBCL-like RS were IGHV-mutated. RS subgrouping was thus highly associated with IGHV mutational status ($p = 6.3e{-9}$). This raises the possibility that RS subgroup partitioning simply reflects DNAm patterns of U-CLL and M-CLL. However, while most CLL-derived RS samples gathered among U-CLL, DLBCL-like RS samples regrouped with DLBCL, well separated from M-CLL (Fig. 3c).

Moreover, none of the DLBCL-like RS were clonally related to their respective CLL component ($n = 5$ pairs), confirming that DLBCL-like RS were not M-CLL-derived RS but rather de novo DLBCLs. In contrast, CLL epigenetic imprint is a feature of CLL-derived RS, likely an entity arising from CLL cells (Supplementary Fig. 15). This CLL-RS clonal relationship was further confirmed by identical IGHV-CDR3 sequences found in paired CLL and RS samples ($n = 26$ pairs; $p = 5.8e{-6}$). To

Article

https://doi.org/10.1038/s41467-022-34642-6

Fig. 3 | RS gene expression profiles corroborate DNA methylation subgrouping. **a** Unsupervised hierarchical clustering of RS and de novo DLBCL transcriptomes (RNA-Seq; 23,508 genes). **b** *K*-means consensus clustering of RS transcriptomes according to DNA methylation-based LPS gradient. Expression level statistics for each cluster are displayed as barplots. Barplot: data are presented as mean values +/− standard deviation from the mean. Cluster 1: $n = 1657$ genes; $p = 1.29e{-}5$. Cluster 6: $n = 2203$ genes; $p = 2.56e{-}7$. *p* values were derived from two-sided *t* tests. Source data are provided as a Source data file. Differential clusters are functionally annotated to the right. Mutational statuses as reported with NGS, or abnormalities determined with CNV analysis on DNAm data, are added below sample annotation for a selected panel frequently described in CLL and RS. **c** Sample partitioning according to IGHV mutational status. Unsupervised PCA clustering of U-RS, M-RS, U-CLL, M-CLL, and DLBCL according to the 10,000 most variable CpGs in the dataset. The focus is made on the most variable CpGs because these are highly representative of the IGHV signature in CLL (59% of these CpGs are strongly differential between U-CLL and M-CLL). Indeed, PC1 separates IGHV-unmutated from IGHV-mutated B cell malignancies, with U-CLLs and U-RS segregating in the same area. Conversely, M-RS partition with DLBCLs, clearly separated from M-CLLs on PC2. CLL chronic lymphocytic leukemia, COO cell of origin, DLBCL de novo diffuse large B cell lymphoma, DLBCL-like RS DLBCL-like Richter syndrome, e enrichment, EBV Epstein–Barr virus, GCB germinal center B cell, $_{high}$CLL-derived RS CLL-derived RS with a high LPS, LN lymph node, $_{low}$CLL-derived RS CLL-derived RS with an LPS score below threshold, LPS linear predictor score, M-CLL IGHV-mutated CLL, M-RS IGHV-mutated Richter syndrome, PC1/2 principal component 1/2, q *q*-value (corrected *p* value), RS Richter syndrome, U-CLL IGHV-unmutated CLL, U-RS IGHV-unmutated Richter syndrome.

confirm the ability of the LPS to identify CLL-derived RS, we set up an independent validation EPIC 850 K experiment, investigating 52 samples (see "Methods" and Supplementary Fig. 16): (i) 44 new samples, including 18 new RS, the CLL component of 14 of these, 6 new DLBCLs, and 6 new CLLs; (ii) 8 samples from the first series: 4 RS samples (3 clonally related and 1 clonally unrelated), with the 4 respective CLL components. LPS classified 5/22 RS samples (22.7%; including the clonallyunrelated RS from the first series) as DLBCL-like RS. Absence of clonal relationship with preceding CLL was confirmed by IGHV sequencing for 3 of these (data unavailable for the 2 other cases). The other 17 RS samples were identified as CLL-derived RS, with IGHV-assessed clonal relationship for 15/15 samples with concomitant CLL (Supplementary Fig. 16). These findings clearly indicate that DNAm is a powerful tool to determine the cellular origin in cases diagnosed as RS, as it differentiates DLBCL arising in a patient with CLL from true morphological transformations of CLL.

To further characterize our RS samples, we sequenced a panel of 13 CLL driver genes. Data integrated with copy number variations obtained from DNAm showed a high prevalence of CLL-driver mutations in RS samples harboring a CLL methylation signature (Supplementary Fig. 17). CLL-derived RS and DLBCL-like RS clinical features are displayed in Table 1. Both RS groups were uniformly treated with rituximab-based chemotherapy regimens, yet with inferior outcome for CLL-derived RS ($p = 1.7e{-}3$). This was further confirmed with gene-expression profiling, where RS samples aggregating in the CLL-derived branch of the dendrogram (Fig. 3a) were associated with a median OS of only 8 months. In contrast, RS samples clustering with the DLBCLs were associated with a longer median OS (35.5 months; $p = 0.018$) (Supplementary Fig. 18).

CLL-derived and DLBCL-like RS feature different epigenetic networks

To better understand the epigenetic architecture of RS subgroups, we performed an integrative analysis based on correlations between DNAm and gene expression data (see "Methods"). The resulting integrome associated 674,567 transcripts with methylation loci. From these, 63,305 (9.4%) significant correlations ($p < 0.01$, Spearman's rho <−0.33 and >0.33) were first selected. Compared with DLBCL-like RS, $_{high}$CLL-derived RS were mainly hypomethylated, which transcribed into a dominant direction of overexpression (Fig. 4a). Matching density maps were observed for $_{high}$CLL-derived and $_{low}$CLL-derived RS, with only slight differences. In contrast, DLBCL-like epigenomic programs largely differed (Supplementary Fig. 19), so we undertook an in-depth comparison of their integrome against that of $_{high}$CLL-derived RS. Significant correlations between the two RS groups accumulated at regulatory locations and were mostly negative (77.3%; Fig. 4b). Genes under the control of these regions were related to cell proliferation (cell cycle, NOTCH pathway, PLCγ-mediated BCR signaling), epigenetic regulation and RNA processing, immune response (T- and B-lymphocyte activation and differentiation), and transcriptional regulation, including STAT family transcription factors (TF). Negative correlations between promoter methylation levels and gene expression (rho < −0.33; at least three hits in the same regulatory region; Supplementary Data 5) led to a list of 666 unique associations showing enrichment in TF binding sites of SUZ12, TP63, TP53, and target genes of early B cell development TFs. Conversely, 234 regions correlated positively between DNAm and gene expression levels (22.7%; 3 hits with rho >0.33; Supplementary Data 6 and Fig. 4b). These were involved in controlling cellular proliferation and differentiation, regulation of transcription, protein metabolism, and immune response. Taken together, positively and negatively correlated locations amounted to 861 unique genes summarizing the most prominent features of $_{high}$CLL-derived compared to DLBCL-like RS in terms of transcriptional mechanisms. Substantial differences in B cell development programs were highlighted, including the lower expression of B-lymphocyte-associated TFs *EBF1* and E2F partner *MSC/ABF1*, and the higher expression of *CD5*, *CCND1*, *ZAP70*, *ID3*, *BLK*, *WNT3*, *PRKCZ*, and *MGMT* in $_{high}$CLL-derived RS (Fig. 4c, Supplementary Fig. 20, and Supplementary Data 7).

Methylome and transcriptome integration provide insights into RS regulatory features

Key players of RS epigenetic deregulations were further identified in $_{high}$CLL-derived RS, using DLBCL-like RS as a reference, and the 861 genes transcriptionally controlled through methylation. Among these, 156 were identified as TFs (18.1%; 2.3-fold enrichment; $p < 1e{-}16$)[39]. The regulatory network reconstructed in silico from these genes showed a central role of p53-like TFs and STAT proteins, an extensive control emanating from master regulators such as TP53, NF-KB1, and FOXC1, an essential developmental TF in many tissues which may have a role as a tumor suppressor. Over-represented target genes included those of the transcriptional repressors ZNF418 (6.1-fold; FDR = 1.87e−21) and ZNF217 (2.1-fold; FDR = 1.56e−8), involved in differentiation and antagonizing cell death, respectively. On the network, downstream effectors were mainly involved in epigenetic repression via the polycomb complex Prc2 (Supplementary Fig. 21 and Supplementary Data 8), for which we noted a SUZ12 signature (FDR = 5.68e−4) and an EZH2 target enrichment (FDR = 2.84e−4) in B cells, also linked with H3K9me3, H3K27me, and H3K27me3 epigenetic marks (FDR < 3.85e−6 in GM12878 cell line). The 156 TFs were strongly enriched in KRAB domain/C2H2-ZF-type TFs defining homeobox developmental proteins[40]. We observed P300 favored interactions (4.2-fold increase; FDR = 3.1e−3), denoting enhancers as enriched targets[41]. These results support our previous findings and highlight critical pathway reprogramming through selected epigenetic control of key TFs as an important mechanism in RS.

RS-based classifiers uncover "RS-type" DLBCLs with poor outcome

DLBCL histological presentation of RS is essential to be distinguished from de novo DLBCL because they differ greatly in terms of prognosis. We thus developed a gene expression based linear classifier score

Article https://doi.org/10.1038/s41467-022-34642-6

Table 1 | Biological characteristics of the different RS subgroups, according to DNA methylation profiling

Characteristic	Full cohort		CLL-derived RS		DLBCL-like RS		CLL-derived versus DLBCL-like RS
	n/N	%	*n/N*	%	*n/N*	%	
Clinical features at CLL diagnosis							
Age at diagnosis (years)							
Median (range)	60 (35–82)		59 (35–80)		64 (52–82)		$p = 0.1$ (NS)
Number of CLL treatment lines before RS transformation							
0	18/56	32	10/44	23	8/12	66	$p = 0.02$
1	14/56	25	12/44	27	2/12	17	
≥2	24/56	43	22/44	50	2/12	17	
Clinical and biologic features at RS diagnosis							
Male (%)	39/58	67	31/45	69	8/13	62	$p = 0.73$ (NS)
Age at diagnosis (y)							
Median (range)	66 (42–88)		65 (42–83)		69 (59–88)		$p = 0.12$ (NS)
Time to RS transformation (y)							
Time <2 y	15/56	27	10/44	23	5/12	42	$p = 0.44$ (NS)
2 y ≤ time ≤5 y	10/56	18	8/44	18	2/12	16	
Time >5 y	31/56	55	26/44	59	5/12	42	
CLL status at RS diagnosis							
Binet A	34/50	68	27/40	68	7/10	70	$p = 0.41$ (NS)
Binet B	10/50	20	7/40	17	3/10	30	
Binet C	6/50	12	6/40	15	0/10	0	
Response	13/52	25	12/43	28	1/9	11	$p = 0.42$ (NS)
Progression	39/52	75	31/43	72	8/9	89	
ECOG PS > 1	28/52	54	21/42	50	7/10	70	$p = 0.30$ (NS)
Ann Arbor stage I–II	8/55	15	7/43	16	1/12	8	$p = 0.67$ (NS)
Ann Arbor stage III–IV	47/55	85	36/43	84	11/12	92	
RS score							
0–1	30/49	61	21/39	54	9/10	90	$p = 0.07$ (NS)
2–3	19/49	39	18/39	46	1/10	10	
Rossi score[17]							
High risk	28/50	56	21/40	52	7/10	70	$p = 0.67$ (NS)
Intermediate risk	17/50	34	15/40	38	2/10	20	
Low risk	5/50	10	4/40	10	1/10	10	
First-line RS treatment							
R-CHOP/R-ACVBP	46/53	87	37/43	86	9/10	90	$p = 1$ (NS)
Platinum-based immuno-chemotherapies	7/53	13	6/43	14	1/10	10	
Response to RS first-line treatment							
Complete remission	15/53	28	10/42	24	5/11	45	$p = 0.35$ (NS)
Partial remission	2/53	4	2/42	5	0/11	0	
Stable disease progression	36/53	68	30/42	71	6/11	55	
OS < 12 months	42/56	75	35/44	80	7/12	58	$p = 1.7 \times 10^{-3}$
12 ≤ OS ≤ 48 months	8/56	14	8/44	18	0/12	0	
OS > 48 months	6/56	11	1/44	2	5/12	42	
EBV positive	3/21	14	1/16	6	2/5	40	$p = 0.12$ (NS)
IGHV unmutated	43/58	74	42/45	93	1/13	7	$p = 6.3 \times 10^{-9}$
Stereotyped IGHV	12/58	21	10/45	22	2/13	15	$p = 0.71$ (NS)
CLL clonally related	26/31	84	26/26	100	0/5	0	$p = 5.8 \times 10^{-6}$
Large cell component (%), median [range]	80 [50–95]		80 [50–95]		80 [50–90]		$p = 0.44$ (NS)
Del 17p (13.1)	26/58	45	23/45	51	3/13	23	$p = 0.11$ (NS)
Del 11q (22.3)	6/58	10	6/45	13	0/13	0	$p = 0.32$ (NS)
Trisomy 12	11/58	19	9/45	20	2/13	15	$p = 1$ (NS)
Del 13q (14.3)	10/58	17	10/45	22	0/13	0	$p = 0.09$ (NS)
TP53	21/58	36	17/45	38	4/13	31	$p = 0.75$ (NS)
NOTCH1	21/58	36	18/45	40	3/13	23	$p = 0.33$ (NS)
SF3B1	12/58	22	12/45	27	0/13	0	$p = 0.05$ (NS)
EGR2	11/58	19	11/45	24	0/13	0	$p = 0.055$ (NS)

Article

https://doi.org/10.1038/s41467-022-34642-6

Table 1 (continued) | Biological characteristics of the different RS subgroups, according to DNA methylation profiling

Characteristic	Full cohort		CLL-derived RS		DLBCL-like RS		CLL-derived versus DLBCL-like RS
	n/N	%	*n/N*	%	*n/N*	%	
XPO1	7/58	12	7/45	16	0/13	0	$p = 0.33$ (NS)
MYD88	5/58	8	1/45	2	4/13	31	$p = 7 \times 10^{-3}$
ATM	4/58	7	4/45	11	0/13	0	$p = 1$ (NS)
POT1	3/58	5	3/45	7	0/13	0	$p = 0.1$ (NS)
RPS15	2/58	3.5	2/45	4	0/13	0	$p = 1$ (NS)
FBXW7	1/58	2	0/45	0	1/13	8	$p = 0.22$ (NS)
BIRC3	1/58	2	1/45	2	0/13	0	$p = 0.4$ (NS)
BRAF	1/58	2	1/45	2	0/13	0	$p = 0.4$ (NS)

Two-sided Student's *t* tests.
CLL chronic lymphocytic leukemia, *DLBCL* diffuse large B cell lymphoma, *EBV* Epstein–Barr virus, *ECOG PS* Eastern Cooperative Oncology Group performance status, *NS* non-significant, *OS* overall survival, *RS* Richter syndrome.

(LCS) to discriminate CLL-derived RS cases among DLBCL samples. We used a set of 215 genes selected from the transcriptomic CLL-derived RS signature (Supplementary Data 9; see "Methods") to screen external datasets of supposedly de novo DLBCL for the CLL-derived RS imprint. We first explored an independent gene expression dataset containing RS samples, untransformed CLLs, and EBV-positive DLBCL cell lines (GSE103265). The 215-gene set allowed unequivocal clustering of RS and CLL samples, well separated from DLBCLs (Supplementary Fig. 22). To cross-validate the previously described DNAm-based classifier (LPS) with the gene expression-derived classifier (LCS), we explored array-based DNAm and transcriptome-sequencing data of the ICGC MMML-Seq consortium (both classifiers can be used independently). Four (6.2%) DLBCL samples with classical DLBCL morphology showed extreme DNAm and gene expression scores, suggesting a CLL-like RS profile (Supplementary Fig. 23). Applying the gene expression-based classifier to array-based gene expression data of 430 DLBCL from the MMML-network identified 31 samples (7.2%) with a statistically significant score (see "Methods"). Next, we mined four large external cohorts of de novo DLBCL, including 1342 samples[8–10,42]. As with previous datasets, gene expression-based LCS distributions were biased toward overrepresenting extreme positive values (Supplementary Fig. 24). Our transcriptomic classifier identified 35/420 (8.3%; series from Lenz and colleagues)[42], 8/137 (5.8%; series from Chapuy and colleagues)[8], 13/223 (5.8%; series from Dubois and colleagues)[9], and 24/562 (4.3%; series from Wright and colleagues)[10] samples harboring the CLL-derived RS signature with a score above the threshold, for a total of 80/1342 (5.9%) samples. In the four datasets, 91.6% to 100% of these samples were of ABC-like subtype. We cross-compared this 215-gene signature and a discriminant 44-gene signature of ABC-type DLBCL[38], identifying *LMO2* as the only common gene. *LMO2* is an important gene of the ABC signature but holds no more weight in our classifier than the other 214 genes. Indeed, instead of just outlining every ABC-subtype DLBCL, our classifier extracted DLBCL with outstanding features, enriched in, but not exclusively, ABC-subtype DLBCL, with an overlap between ABC and GCB DLBCL and a subset of ABC-subtype DLBCL associated with a low LCS (Supplementary Fig. 25). DLBCL sharing the extreme score values with CLL-like RS showed a shorter progression-free survival (PFS) and/or OS (*p* values ranging from $<10^{-3}$ to 0.02 depending on the cohort) compared to all other samples, and compared to other ABC-subtype DLBCL (*p* values ranging from $<10^{-3}$ to 0.07) (Fig. 5a, b and Supplementary Figs. 26–28). We next conducted a multivariate analysis with Cox Proportional Hazards models, including all available covariates to evaluate the association of gene expression-based LCS with survival (OS and PFS). This association was set up in binary (top 25% versus the rest) as well as linear (as a continuous variable) models and provided estimates and effect size for each covariate (IPI, *TP53* and *MYC/BCL2* double hit status; Supplementary Fig. 29). This systematic analysis confirmed strong associations with survival, independently from other covariates. In particular, this shorter survival was unrelated to international prognostic index distribution (Supplementary Fig. 30).

We next explored whether this effect might be due to the enrichment of a previously described genomic subgroup of ABC-like DLBCL associated with unfavorable prognosis[10]. In the 562-sample dataset from Wright and colleagues, the 25 cases with top LCS scores were enriched in formerly unassigned (1.5-fold relative enrichment; $p = 0.04$) and N1 subgroups (6-fold; $p = 6e{-}3$) while depleted in EZB subtype ($p = 0.03$). These cases were also strongly enriched (6.74-fold; $p = 4.1e{-}4$) in samples collected at relapse, raising the hypothesis of the ability of our classifier to identify DLBCL prone to relapse. Thus, the extreme LCS values seemed to characterize a distinct subset of ABC-type DLBCL, accounting for 4.3–8.3% of de novo DLBCL, with poor prognosis. The highest 25% scores in the series from Wright and colleagues showed biased distributions in genomic subgroups, dominated by unclassified cases, and associated with shorter PFS and OS (Fig. 6). These findings suggest an ability of the LCS classifier to: (i) identify high-scoring DLBCL samples as a separate DLBCL entity within de novo DLBCL, associated with ABC phenotypes and other features comparable to RS; and (ii) linearly classify other samples according to survival and overall prognosis (Supplementary Figs. 29 and 31). Interestingly, while absent from the 215-gene list, the CLL-associated marker *CD5* was overexpressed in RS versus DLBCL (2.4-fold; FDR = 2.13e−3) and $_{high}$CLL-derived versus DLBCL-like RS (2.3-fold; FDR = 0.01). In the dataset from Wright and colleagues[10], *CD5* expression was higher in samples within the top 25% LCS than in other samples ($p = 5.8e{-}7$), corroborating our results. Last, in a dataset with concomitant transcriptome and CD5 immunochemistry staining GSE66770, the majority (17/22; 77.2%) of the top 25% samples were CD5+ DLBCL (2.1-fold enrichment) while this proportion was significantly lower (16/68; 23.5%) in the rest of the cohort ($p = 4.73e{-}3$).

Discussion

In this study, by using genome-wide DNAm analysis and whole-transcriptome gene expression profiling, we extensively characterized the epigenetic architecture of primary human RS samples. We identified a CLL epigenetic imprint that can act as a surrogate for identifying whether an RS is clonally related to CLL or has arisen de novo. Discovery of the CLL imprint in an RS sample avoids reliance on obtaining tumor DNA at the CLL stage. Considering de novo DLBCL, DNAm- and gene expression-based classifiers delineated an RS-like subset in datasets from several landmark studies that was not previously described at the genomic level, was enriched in cases with an ABC-like COO signature, and had an unfavorable prognosis[7,8,10].

Previous extensive explorations with exome or full genome-sequencing had found differences in genomic landscapes between DLBCL-subtype RS and de novo DLBCL[16–20]. Here we used a different

Article
https://doi.org/10.1038/s41467-022-34642-6

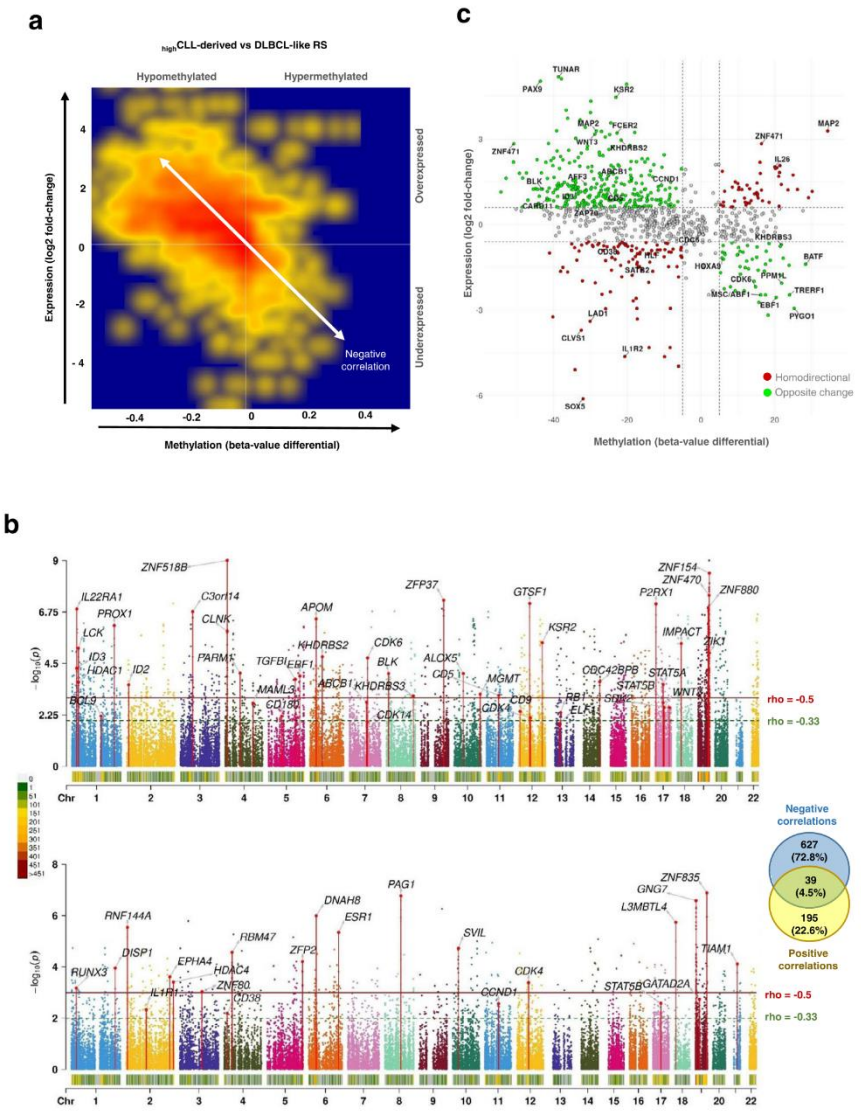


Fig. 4 | Integrative analysis of DNA methylation and transcriptome data highlights different epigenetic programs in $_{high}$CLL-derived and DLBCL-like RS. **a** Density map (smoothed density scatterplot) representing overall DNA methylation versus gene expression changes between $_{high}$CLL-derived RS and DLBCL-like RS. Scale ranges from blue (no density), to yellow (medium density) and red (high density). Only genes with at least one significant correlation (Spearman's test; p value <0.01) were retained. Locations of the corresponding CpGs were mainly distributed in proximal and distal regulatory regions, with specific enrichments in TSS features for negative (TSS200: 2.6-fold, TSS1500: 2.2-fold) and positive (TSS200: 1.2-fold, TSS1500: 1.6-fold) correlations. Hypo/hyper-methylations and under/over-expressions are indicated relatively to the $_{high}$CLL-derived RS subgroup. **b** Manhattan plots of negatively and positively correlated regulatory regions and associated transcript expressions. Chromosomes are displayed at the bottom of each plot, with a color code (from green to red) indicating the density of correlations over sliding windows of 1 Mb. Series of vertically aligned dots indicate DMRs (of at least 3 CpGs with a hit in TSS-associated location) significantly correlated with gene expression. Upper part: negative correlations, amounting to 666 unique genes; bottom part: positive correlations, amounting to 234 unique genes; a VENN diagram indicates the overlap between negative and positive correlations. **c** Quadrant scatterplot displaying methylation levels of regulatory sequences and corresponding expression levels for the 861 selected genes (overall absolute correlations: rho = 0.72; p < 2.2e−16; Spearman's tests). The upper left and lower right quadrants show genes with a negative correlation between methylation and expression. Lower left and upper right areas: genes with positive correlations. CLL chronic lymphocytic leukemia, DLBCL de novo diffuse large B cell lymphoma, DLBCL-like RS DLBCL-like Richter syndrome, DMR differentially methylated region, $_{high}$CLL-derived RS CLL-derived RS with a high linear predictor score, RS Richter syndrome, TSS transcription start site.

study design and methodological approach to expand this knowledge. Firstly, we studied epigenetic deregulations using robust and proven methods, and profiled the RS molecular landscape beyond gene mutations and copy number variations. Secondly, we conducted a comprehensive analysis of RS pathophysiology which combined the analysis of genome-wide DNAm and whole transcriptome profiling, rather than pinpointing a limited number of specific targets. Thirdly, we compared the RS epigenetic profile to that of large cohorts of diverse CLL and de novo DLBCL, which contrasts with previous work mostly focusing on the RS transformation process.

Human-derived xenograft mouse models and cell lines were recently reported to study RS biology and test drug response[11–15].

Article
https://doi.org/10.1038/s41467-022-34642-6

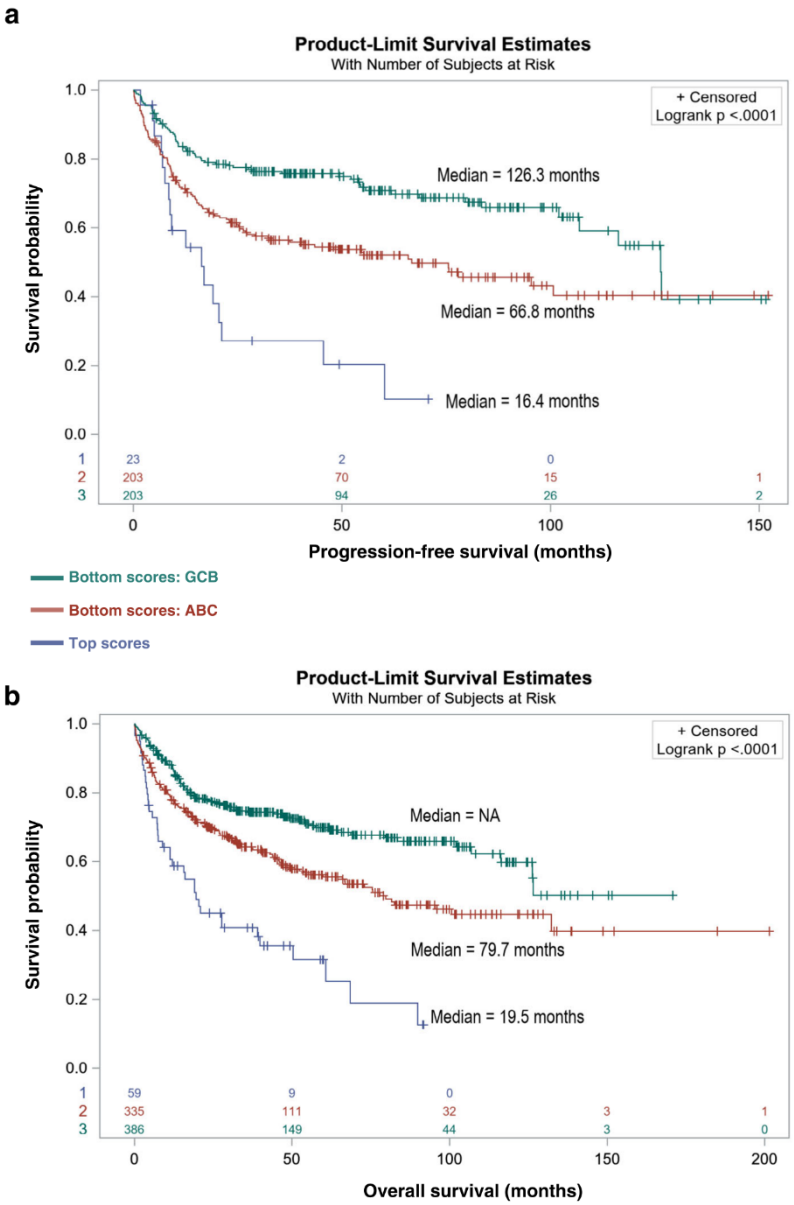
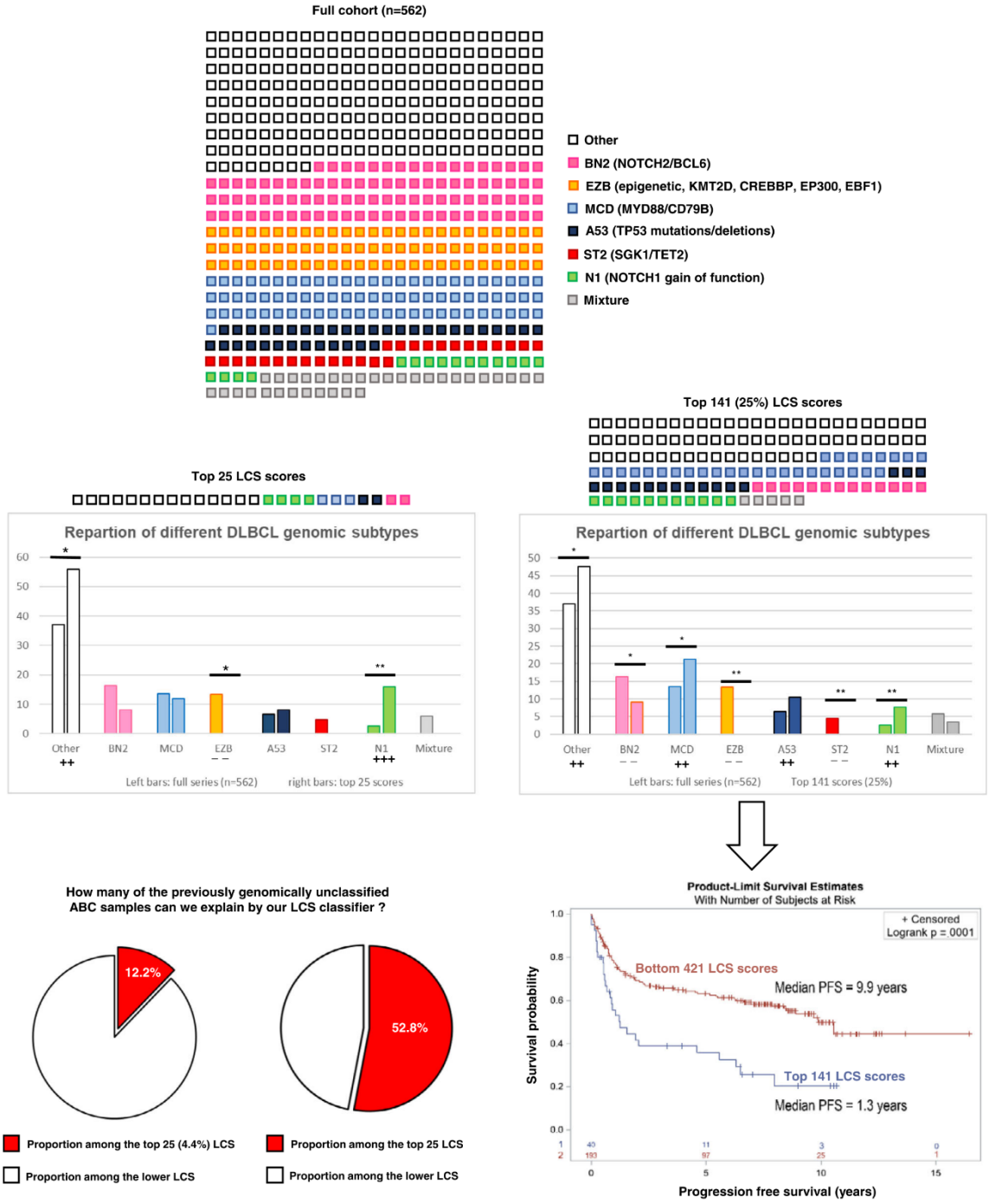


Fig. 5 | DLBCLs harboring the CLL-derived RS epigenetic signature are associated with ABC phenotype and worse outcome. a Kaplan–Meier estimates of progression-free survival for $n = 429$ patients from three combined and clinically annotated public DLBCL datasets[8–10]. Comparative PFS between patients with top LCS and the rest of the cohorts, according to COO ($p = 8.4e{-}8$). **b** Kaplan–Meier estimates of overall survival for $n = 780$ patients from four combined and clinically annotated DLBCL public datasets[8–10,42]. Comparative OS between patients with top LCS and the rest of the cohorts, according to COO ($p = 1.1e{-}11$). Statistical comparisons were performed with the log-rank test. Bonferroni method was used for multitesting adjustments. Datasets: from Lenz et al. ($n = 420$; microarray, accession under GSE10846; PMID: 21546504); from Chapuy et al. ($n = 137$; microarray, accession under GSE98588; PMID: 29713087); from Dubois et al. ($n = 223$; microarray, accession under GSE87371; PMID: 31648986); from Wright et al. ($n = 562$; RNA-Seq; PMID: 32289277). ABC activated B cell, CLL chronic lymphocytic leukemia, COO cell of origin, DLBCL de novo diffuse large B cell lymphoma, GCB germinal center B cell, LCS linear classifier score, OS overall survival, PFS progression-free survival, RS Richter syndrome.

However, the availability of these models is limited and they cannot recapitulate the full heterogeneity of RS, as they were generated from a limited number of tumor samples. Our approach using large cohorts of primary human RS samples and comparative tumor material also holds promise for discoveries and better characterize the wide RS epigenetic complexity. We cross-validated our epigenetic findings using DNAm patterns, that were largely corroborated by transcriptome data, in an independent manner.

Our genome-wide DNAm data provide a more complete RS hypomethylation profile description. The DNAm patterns confirm previous findings that RS is a DNA-hypomethylated entity as compared with CLL and de novo DLBCL[27]. Such global hypomethylation may in

Article https://doi.org/10.1038/s41467-022-34642-6



part reflect a more extensive proliferative history of the RS subclone[21], as measured by the epiCMIT mitotic clock[33]. Using a reproducible DNAm microarray uniformly spanning the vast majority of regulatory regions at a whole-genome scale[37], we characterized the epigenetic architecture underlying the commonly accepted dichotomic heterogeneity with regard to whether a primary RS is clonally related to CLL or has arisen de novo[17]. As expected, around 80% of our RS samples harbored a CLL epigenetic imprint (likely derived from a pre-existing CLL clone). This was confirmed by identical IGHV-CDR3 sequences for all CLL-RS follow-ups. As nearly all de novo DLBCLs harbor a mutated IGHV, we propose that RS clonally related to the underlying CLL clone are: (i) IGHV-unmutated DLBCL; and (ii) IGHV-mutated DLBCL with a CLL imprint. Determining CLL history using DNAm and gene expression by identifying a CLL imprint independently from matched-CLL availability is a step forward, and is essential for clinical and therapeutic management. Interestingly, DLBCL-like RS would conversely be DLBCL without clonal relationship with the CLL counterpart. However, DNAm of DLBCL-like RS differed from that of de novo DLBCL in terms of

Article
https://doi.org/10.1038/s41467-022-34642-6

Fig. 6 | The gene expression-based LCS linearly classifies de novo DLBCL samples, with high scores enriched in N1, unclassified genomic profiles[10], and shorter progression-free survival. Dataset from Wright et al. ($n = 562$; RNA-Seq; PMID: 32289277). Two-sided *t* tests were used to assess statistical significance. Top 25 LCS scores: enrichment in "other" subtype ($e = 1.51$; $p = 4.6e{-}2$); depletion in EZB subtype ($e = 0$; $p = 3.0e{-}2$); enrichment in N1 subtype ($e = 5.99$; $p = 6.4e{-}3$). Top 141 (25%) LCS scores: enrichment in "other" subtype ($e = 1.28$; $p = 1.4e{-}2$); depletion in BN2 subtype ($e = 0.56$; $p = 1.9e{-}2$); enrichment in MCD subtype ($e = 1.57$; $p = 1.7e{-}2$); depletion in EZB subtype ($e = 0$; $p = 1.7e{-}8$); enrichment in A53 subtype ($e = 1.62$; $p = 7.5e{-}2$); depletion in ST2 subtype ($e = 0$; $p = 2.6e{-}3$); enrichment in N1 subtype ($e = 2.92$; $p = 7.0e{-}3$). Survival curves: Kaplan-Meier estimates of progression-free survival for $n = 233$ patients from a clinically and genomically annotated dataset from Wright and colleagues. Comparative PFS between patients with top 25% LCS and the rest of the cohort. Statistical comparisons were performed with the log-rank test ($p = 1e{-}4$). Source data are provided as a Source Data file. ABC activated B cell like, A53 *TP53* mutations/deletions-associated DLBCL subgroup, BN2 DLBCL subgroup associated with lesions of *BCL6* and/or *NOTCH2*, COO cell of origin, DLBCL de novo diffuse large B cell lymphoma, EZB DLBCL subgroup associated with abnormalities of epigenetic regulators *KMT2D*, *CREBBP*, *EP300*, and/or *EBF1*, GCB germinal center B cell, LCS linear classifier score, MCD DLBCL subgroup associated with lesions of *MYD88* and/or *CD79B*, N1 DLBCL subgroup associated with *NOTCH1* gain of function, PFS progression-free survival, RS Richter syndrome, ST2 DLBCL subgroup associated with lesions of *SGK1* and/or *TET2*. *p value <0.05; **p value <0.01; ++: enrichment >1.2; +++: enrichment >5; −: depletion <0.6.

increased cell cycle activity and IGF1, ERK/MAPK, PI3K/AKT, and PD-1 signaling pathways. These differences suggest influences of the CLL-invaded microenvironment for the development of a specific DLBCL pathogenesis[43].

Moreover, by integrating the DNAm and transcriptomic data, we evidenced different epigenetic networks in CLL-derived and DLBCL-like RS. Epigenetic architecture remodeling and subsequent deregulation of EZH2 and Wnt pathways, as well as PI3kinase/AKT and IGFR1 signaling cascades, unravel CLL-derived RS underlying mechanisms potentially responsible for chemotherapy resistance. These mechanisms are potentially druggable through EZH2, PI3K/AKT, or IGFR1 inhibitors. IGFR1 pathway triggering was recently described as a resistance mechanism to targeted therapy in CLL[44]. Interestingly, O6-methylguanine-DNA methyltransferase *MGMT* regulatory sequences are hypomethylated and *MGMT* is consequently overexpressed in CLL-derived RS. *MGMT* promoter hypomethylation status is a known negative prognostic marker in glioblastoma[45], de novo DLBCL[46], and an actionable target. This marker is easily assessable in the context of DLBCL diagnosis and routinely used to guide therapeutic decisions.

Our results show B cell-specific TF implication and epigenetic imprint in CLL-derived RS, and emphasize the previously described important role of TP53, FOXC1, NF-KB, and epigenetic regulators in oncogenic mechanisms. Strikingly, genes involved in the regulation of TP53 activity through methylation were overexpressed in CLL-derived RS, confirming the central role of TP53 in clonally-related RS and the primary importance of epigenetic deregulation in the transformation process. An interesting finding of this study is the putative role of the FOXC1 TF in the RS regulatory network. FOXC1 has previously been described as cooperating with HOX family members for orchestrating mesenchymal tissue development, through NF-KB signaling[47]. FOXC1 is PRC2 repressed during hematopoietic development, but frequently derepressed in hematopoietic progenitors in acute myeloid leukemia[48]. Our data identified FOXC1 derepression as a hallmark of CLL-derived RS, likely associated with the blockade of B cell development and proliferation due to NF-KB signaling unleashing. We also observed hypomethylation of DMRs regulating the expression of genes involved in the extracellular matrix organization, and in the immune system. These observations suggest a strong influence of the microenvironment in RS development.

Notably, our findings directly translate into classification and prognostication of de novo DLBCL, the most common human B cell lymphoma. We provide a gene expression-based, stable, reproducible, and potentially widely applicable classifier, on the basis of a CLL-derived RS epigenetic imprint. The classifier differentiates a particular DLBCL subgroup from supposedly de novo DLBCL datasets. Of clinical importance, cases assigned to this subgroup are frequently not detected by recently described genomic and gene expression classifiers of DLBCL, and they are associated with an unfavorable prognosis. These cases were ABC-like DLBCL, enriched in unclassified or N1 DLBCL genomic subtypes[7,10]. This is in line with the association of RS with a particular gene expression profile and with *NOTCH1* mutations and NOTCH pathway activation. Given the efficacy of ibrutinib plus R-CHOP chemotherapy in N1 subtype DLBCL[49], the enrichment in N1 profile within RS samples supports research into whether these patients may also benefit from BTK inhibition combined with R-CHOP chemotherapy. However, a recent single cell transcriptome analysis of sequential CLL-RS samples revealed that, as compared to the CLL cells, RS cells downregulate genes related to BCR signaling and upregulate those involved in oxidative phosphorylation[21], and therefore RS may be less sensitive to Ibrutinib. By applying a stringent cut-off to our transcriptomic score, generalized to all studied DLBCL datasets[8–10,42], we identified a separate de novo DLBCL subset associated with a median PFS comparable to that of clonally-related RS. Based on our observations, 4-8% of DLBCL diagnosed as de novo DLBCL, non-otherwise specified, may in fact be a subgroup of DLBCLs sharing common epigenetic and transcriptional features with clonally related RS, and with a similar unfavorable outcome. We propose a stable and reproducible expression-based classifier widely applicable to transcriptomic data, enabling the identification of this specific entity within supposedly de novo DLBCL, termed "RS-type DLBCL." Limitations of the transcriptome scoring method are dataset size and composition (DLBCL features associated with outcome), which by design prevent the exploration of single samples independently and may exert biases. However, the method also demonstrated the linear association of DLBCL scores with poor outcome and clinical variables of cancer aggressiveness, and so constitutes a means for improving the current DLBCL classification system.

In conclusion, our study has revealed several relevant aspects of RS biology, including the complete RS hypomethylation profile and differentiation of clonal versus non-clonal RS according to DNAm patterns and gene expression profiles. The discovery of a CLL imprint allows clonal relationship assessment without the need for tumor DNA at the CLL stage. Subgrouping of primary RS samples according to extensive characterization of the epigenetic architecture has provided information underlying oncogenic processes, with clear clinical implications. In particular, identification of RS-type DLBCL cases helps to advance the current DLBCL classification system and could be incorporated in treatment decisions, potentially improving disease management. Our findings also enable the evaluation of larger cohorts recruited in clinical trials and the development of novel treatment approaches, which are urgently needed in RS.

Methods

Our methods and results made extensive use of data from previous landmark studies[26,31]. Care was taken to follow good practices in the analyses of methylome and transcriptome data, employing widely approved procedures previously used in other high-standard studies. Regarding the handling of large cohorts, we used sample correlations, performed genotype checks between omics data, and added technical and biologic replicates wherever possible.

Ethics statement

This study complies with all relevant ethical regulations and we have obtained written informed consent for all participants. No

Article
https://doi.org/10.1038/s41467-022-34642-6

compensation was provided. We obtained consent to use and publish information that identifies individuals, including indirect identifiers such as gender and age. Individuals recruited for this study can no longer be identified by the information provided, due to sample anonymization and processing of the genomic data. All procedures were in accordance with Helsinki declaration. Study protocol was approved by the Institutional Review Boards and Ethics Committees of Nancy, Kiel (#A150/10), Ulm (#349/11; #459/19 and #96/08) and Barcelona university hospitals, and by the French national ethics committee (Comité de Protection des Personnes Ouest IV 09/05/2017).

Patients and materials

A multicenter registry of RS accrual was established across nine centers affiliated to the French Innovative Leukemia Organization (ClinicalTrials.gov Identifier: NCT03619512). Sixty-four patients diagnosed with DLBCL-subtype RS were enrolled. Fresh frozen biopsies were gathered at RS diagnosis and met the criteria for DLBCL, including diffuse patterns of large B cells with the same size as macrophages or twice the size of normal lymphocytes[3,50]. For all patients, diagnoses were reviewed and confirmed by two independent pathologists. Only RS samples with at least 50% (median 80%, range 50-95%) high-grade component assessed by pathology review were selected for analysis. The same process was applied for assembling a validation cohort of 58 samples, further reduced to 52 QC-passed samples, which we processed to an independent EPIC 850K experiment. This 52-sample validation cohort included 44 new samples: 18 new RS samples, the CLL component of 14 of these, 6 new DLBCL samples, and 6 additional CLLs. In addition, 8 samples from the training series were used as controls: 4 RS samples (3 clonally related and 1 clonally unrelated), with the 4 respective CLL components (Supplementary Fig. 16). Thus, this EPIC 850K experiment investigated 22 RS, 6 new DLBCLs, and 24 CLLs, including 18 paired-CLLs.

Fifty-eight of the 64 enrolled patients with RS were from a previously described cohort, and both targeted NGS sequencing and DNAm exploration were performed; 56 of these 58 patients with RS underwent 18F-fluorodeoxyglucose positron emission tomography/computed tomography for initial diagnosis[51]. For the other six patients with RS, the fresh frozen biopsy was too small for extracting both DNA and RNA, and due to the large cellular component (>70%), we prioritized gene expression data and only RNA sequencing was performed. The minimal tumor purity was raised to 70% for RNA analysis, as contamination by signal from residual normal cells strongly influences global gene expression, especially for a subset of transcripts with very low expression in tumor cells but high expression in residual normal cells.

Additional data for CLL ($n = 215$), and 92 normal B cells spanning the entire B lineage development were obtained as part of previously published studies[22,25,26,34,35]. DNA methylation from 68 de novo DLBCL cases were also used as a reference. These DLBCLs originate from a larger lymphoma cohort gathered by the ICGC MMML-seq consortium[52]. Finally, 10 lymph nodes from healthy subjects were analyzed as a control group for transcriptome sequencing.

Methylome data analyses

EPIC microarray. DNAm status of 866,562 CpG sites was interrogated on the *Infinium Methylation EPIC array* (Illumina, San Diego, CA, USA; see Supplementals), later referred to as the EPIC 850K platform.

Dataset generation. Datasets were created using the minfi package[53]. The EPIC set comprises 90 distinct samples (58 RS, 25 CLL, plus a subset of 7 DLBCL replicates also available on 450K), interrogated on EPIC 850K. DNA methylation data from the control groups (215 CLL, 68 DLBCL, 92 normal B cells spanning the entire B-lineage) were acquired with the Illumina Infinium® HumanMethylation450 BeadChip (later referred to as the 450K platform)[22,25,26,34]. These and the EPIC 850K data were processed from IDAT files. Analyzes were run under R 3.6 with Bioconductor 3.10 and later versions. The FULL dataset comprises 433 distinct samples (92 benign B cells, 215 CLLs, 68 DLBCLs and 58 RS), combined into a single 450K object containing probes shared by 850K and 450K microarrays: (i) raw IDAT files corresponding to 96 and 377 samples for the 850K (866,091 CpGs) and 450K (485,512 CpGs) platforms, respectively, and included technical replicates; (ii) each subset was loaded independently, stored into a dedicated *RGChannelSet* minfi object, along with full sample annotations, then both were combined into a third subset containing 473 samples × 452,567 CpGs using the *combineArrays* function with output type as "IlluminaHumanMethylation450k"; (iii) the EPIC dataset stems from the first (850K) subset alone, the FULL dataset is obtained from the combined subsets.

To reduce technological issues and biases, the same preparation protocol was applied to both EPIC (850K) and FULL (combined) subsets. The main stages of the filtering and quality control pipeline are as follows: (i) technical checks, filtering, and evaluations (ii) data normalization with SWAN[54]; (iii) probes located on X and Y chromosomes, flagged as cross-hybridization probes, or located near known SNPs were further removed with the *rmSNPandCH* function (with parameters dist = 2 and mafcut = 0.05) available from the DMRcate package[55]; (iv) imputation of the remaining failed β-value positions with *imputePCA* of the R missMDA package[56,57], (v) 2 × 2 sample correlation checks (Supplementary Figs. 32 and 33). Correlation heat maps were rendered with the R corrplot package; (vi) extended quality control step to remove sample outliers and check for residual post-normalization batch effects (Supplementary Fig. 34); (vii) ultimately, technical replicates were averaged into unique samples as all replicates were found comparable (Supplementary Fig. 35). These filtering steps led to the final EPIC (90 samples × 794,927 CpGs) and FULL (433 samples × 397,769 CpGs) datasets.

Technical checks, filtering, evaluations, and quality control. These steps included failed CpGs removal (>10% samples with a detection p value >0.01), gender check between clinical data and gender returned by the *getSex* function, and genotype checks (Supplementary Data 10) between RNA-seq data (see Supplementals) and genotypes inferred with the *beta2genotype* function available from the R OmicsPrint package[58].

Cell composition deconvolution. Cell type composition was estimated for each sample with the *estimateCellCounts* function against a library of 6 normal white blood cells (CD8 T cells, CD4 T cells, NK, B cells, monocytes, and granulocytes) (Supplementary Data 11). The proportions of each explored cell type were reported and later used as covariates in statistical models to adjust for B cell representation in the mixes. Blood samples deprived in B cells (<30%) were thus discarded from further analyses.

Downstream bioinformatics

Supervised analyses. As a rule, β-values were used for direct interpretation and graphical representation, while M-values were favored for statistics and computations. Linear modeling based on empirical Bayesian methods was used to assess for CpG differential methylation. When applicable, these models included cell deconvolution results as added covariates to correct for B cell content. Additionally, at this point, any unwanted methylation variation such as residual batch effects were removed by using the *RUVm* function from package missMethyl[59]. The overall dispersion was calculated on the entire dataset, then p values for each comparison were obtained with a two-way moderated t test and adjusted for FDR following the Benjamini–Hochberg procedure. At probe level, an FDR < 0.01 indicated statistical significance. Differentially methylated region (DMR) determination was performed on the same linear models with *dmrcate* (package DMRcate), with lambda = 1000 and C = 3. FDR cut-off for first allowing a CpG to initiate a DMR was set to FDR = 0.01, and DMRs were

Article https://doi.org/10.1038/s41467-022-34642-6

considered statistically significant if both min_smoothed_fdr and HMFDR output probabilities were <0.01.

Unsupervised analyses. Explorations were conducted on β-values, and all methods used Euclidean distances as (dis)similarity metrics. PCAs were performed with R packages FactoMineR and factoextra, on the entire datasets or a subset of the top variant CpGs across all considered samples. Hierarchical clusterings included complete and average linkage criteria, and resulting heat maps and dendrograms were rendered with the R package ComplexHeatmap. Non-Negative Matrix Factorizations were performed with the R package NMF, either on all CpGs or a subset of the most variant ones according to the context, with method *lee*, and parameters ranking = 3 and iterations = 50.

Feature annotations. Methylome data were analyzed using the available Illumina 450K and EPIC platform annotations, which strongly rely upon the hg19 assembly. As several tools like minfi and DMRcate still use those by default, CpG and DMR locations/annotations were lifted to hg38 coordinates as an after-computation-process when required, especially when dealing with integrations with transcriptomic and epigenomic data. Additional CpG annotations included B cell development modules[34], UCSC tracks for the EBV-transformed GM12878 cell line, such as DNaseI and chromatin marks from ENCODE and transcription factor ChIP-Seq peaks from ENCODE3, histone modifications, and chromHMM chromatin states for 7 reference CLL epigenomes (2 U-CLLs and 5 M-CLLs)[26]. Any *liftOver* of coordinates between hg19- and hg38-annotated data was achieved with the UCSC table browser or with R packages liftOver and XGR.

Gene set and pathway analyses. Unbiased functional annotations on ontological terms (GO) and KEGG pathways were achieved at CpG and DMR levels with the R package missMethyl. Additional enrichment analyses were conducted on curated gene lists with Enrichr[60]. Reactome pathway overrepresentations and enrichment analyses were performed with ReactomePA[61] on curated sets of unique genes associated with identified DMRs. Enrichments in sets of CpGs, DMRs, target genes, GO terms, or pathways were calculated as the occurrences of the selection against a background representing the entire dataset (enrichment = observed frequency/expected frequency). p values associated with enrichment analyses were obtained with (i) an overrepresentation test, (ii) Fisher's exact test, or (iii) a Chi-square test, depending on the context and group size.

Linear predictor score (LPS). To formally distribute RS samples into subgroups, we developed a scoring predictor inspired by the work of Wright and colleagues on transcriptome data, that successfully separated GCB from ABC DLBCL[38]. Here, we applied LPS on methylome data, with a cohort composed of all 58 RS samples, 215 CLLs, and 68 DLBCLs. As we aimed to best discriminate between CLL and DLBCL profiles, only the highly differential CpGs between the two groups were considered in the analysis (261,085; FDR < 0.01; moderated t-statistics were retained for further use in the score computation). To lessen the impact of B cell IGHV maturity on the scoring model, we next subtracted CpGs that were also differential between U-CLL and M-CLL (128,408; FDR < 0.01). The 181,231 remaining CpGs were then filtered into 4863 CpGs with high methylation differential (beta-value differential or β-Fold-Change), that is, >30%. This amount was considered appropriate as: (i) statistical power to discriminate such a methylation differential was reached; (ii) probe composition was balanced between regulatory region/gene body/intergenic location as compared to the background; (iii) it provided sufficient number to expect a normal distribution of LPS within subgroups; and (iv) those CpGs demonstrated a strong correlation structure among the groups of samples (Fig. 2e).

Finally, from each of these 4863 CpGs and for each sample S of the cohort, the score

$$\mathrm{LPS}(S) = \sum_{i=1}^{n} t_i.S_i \quad (1)$$

was calculated, with t_i representing the moderated t-statistic for CpG i and S_i the corresponding methylation β-value. Known score distribution of CLL and DLBCL samples within their respective subgroup $G \in$ [CLL,DLBCL] allowed the Bayesian likelihood approximation for RS samples S to belong in each one of them, with probability

$$P(S \text{ in } G = \mathrm{CLL}) = \frac{\Phi\left(\mathrm{LPS}(S), \hat{\mu}_{\mathrm{CLL}}, \hat{\sigma}^2{}_{\mathrm{CLL}}\right)}{\Phi\left(\mathrm{LPS}(S), \hat{\mu}_{\mathrm{CLL}}, \hat{\sigma}^2{}_{\mathrm{CLL}}\right) + \Phi\left(\mathrm{LPS}(S), \hat{\mu}_{\mathrm{DLBCL}}, \hat{\sigma}^2{}_{\mathrm{DLBCL}}\right)} \quad (2)$$

and $P(S \text{ in } G = \mathrm{CLL}) \simeq 1 - P(S \text{ in } G = \mathrm{DLBCL})$ where Φ computes the normal density function with the estimated means $\hat{\mu}$ and variances $\hat{\sigma}^2$ of LPS within either subgroup G. To finally obtain highly specific and homogeneous subgroups, thresholds were defined as follows: (i) $p(S \text{ in } G = \mathrm{CLL}) \geq 0.98$ for CLL-derived (namely $p_{\mathrm{CLL\text{-}derived}}$), and (ii) $p(S \text{ in } G = \mathrm{DLBCL}) \geq 0.98$ for DLBCL-like ($p_{\mathrm{DLBCL\text{-}like}}$) labeling, since the gray zone between the two main groups is centered on scores for which the probability density functions overlap at values >0.02 either way (Supplementary Fig. 10).

Transcriptomics

All samples were processed within the same batch. Demultiplexed single-end sequencing data corresponding to 50-nucleotide-long reads were available in FASTQ files, one for each of the 87 samples, and used in the next processing steps. The cohort was composed of 47 RS samples, 2 paired-CLLs, 28 DLBCLs, and 10 controls from normal lymph nodes. The 10 normal controls were added for methodologic purposes: (i) normalization; (ii) checking benign profiles against B cell malignancies; (iii) checking the feeble amplitude of the transcriptomic component separating inflammatory ($n = 3$) from non-inflammatory lymph nodes ($n = 7$); and (iv) validation of the efficiency of the developed scoring methods.

Transcriptome reconstruction pipeline. First quality controls were conducted using FastQC v0.11.5 [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/] results as a guideline. No adapter content or known overrepresented sequence needed to be removed at this step. Read mapping and the main filtering were performed using HISAT2 v2.0.4[62] against a reference index built to account for human population SNPs as well as known transcripts (this index can be obtained from ftp://ftp.ccb.jhu.edu/pub/infphilo/hisat2/data/grch38_snp_tran.tar.gz). The following scoring constraints were applied during alignment:–score-min L, 0,−0.2–sp 10.3–dta. Samtools v1.3.1 [http://github.com/samtools/samtools] was used for manipulating the alignment files throughout the downstream analysis. PCR duplicates were flagged with Picard v1.13 MarkDuplicates [https://broadinstitute.github.io/picard/]. Differentially spliced transcripts were assembled from the obtained alignments with StringTie v2.1.0[63]. The present protocol took advantage of the proposed workflow for identifying known as well as novel isoforms, using an annotation file for hg38 in gtf format as a guide [ftp://ftp.ensembl.org/pub/release88/gtf/homo_sapiens/Homo_sapiens.GRCh38.90.gtf.gz]. The following parameters were used: (i) first step is applied for each sample -f 0.2 -j 3 -c 10 -M 0.5, (ii) second step merges all transcripts of all samples–merge -m 200 and (iii) the last step estimates abundances and read coverage for all the merged transcripts, for each sample -A -C -f 0.2 -j 3 -c 10 -M 0.5. Two tables were generated from these results, one compiling raw read count at the gene level, and another at the transcript level.

Article
https://doi.org/10.1038/s41467-022-34642-6

Raw abundance filtering and normalization. Raw counts were filtered by applying a minimum expression threshold for a gene or transcript. Those had to be expressed (non-zero value) in at least two samples and present an average expression value across all samples higher than 1/5,000,000 of the average library size (64 ± 3 million reads per sample), that is, at least 20 reads per feature. Data was further adjusted with the TMM normalization method[64], and finally was log2 and cpm (count per million) transformed[65]. A total of 23,508 genes and 77,491 transcripts were identified and reported at the end of the process. Pearson's correlations for gene expression levels averaged at 0.92 for genes and 0.75 for transcripts and were very stable across samples (data not shown).

Gene and transcript annotations. All transcriptomic analyses were performed using the hg38 reference assembly of the human genome. Results were fully annotated with known symbols corresponding to gene and transcript genomic locations whenever possible. Upon completion of the transcript assembly, gene symbols were assigned Ensembl IDs based on overlapping positions with known transcripts (90% overlap minimum). In case of failed overlap, custom and unique IDs were used. Therefore, gene and transcript assignments were based on the Ensembl[66] GRCh38 annotations available in both *core* and *funcgene* databases, version 90. These were downloaded from ftp://ftp.ensembl.org/pub/release-90/mysql/ for local installation and query with in-house custom tools).

Transcriptome explorations. Unsupervised analyses were all carried out with hierarchical and *K*-means clustering techniques, as previously described[67]. Expression values were median-centered, and uncentered Pearson's correlation was used as distance metrics. Supervised analyses were performed through linear modeling (empirical Bayes), and differential expression *p* values were obtained using a two-way moderated *t* test then adjusted for FDR following the Benjamini–Hochberg procedure. An FDR < 0.01 indicated statistical significance. Cluster dissection was achieved with functional annotation tools for target gene associations, such as the Open Targets platform[68], and gene signature correlation with public datasets from multiple databases, such as GEO (Gene Expression Omnibus), with Enrichr [https://maayanlab.cloud/Enrichr].

Methylome and transcriptome data integrations

Here we focused on the RS cohort, for which 41 RS samples overlapped between methylome and transcriptome experiments. A subset of the methylome EPIC dataset (M-values, normalized and curated) and part of the transcriptome dataset (gene and transcript CPMs – also normalized and filtered) were integrated to eliminate unwanted signals and pinpoint the functional mechanisms linking DNA methylation of regulatory regions with gene expression in RS.

Both datasets were re-annotated with biomaRt[69] and linked using two methods: (i) with shared Ensembl identifiers; and (ii) by genomic coordinates for refined feature overlap when the first method failed. We used "TSS200," "TSS1500," and "first exon" CpG information to define associations with promoter regions in the next analysis steps, and overlap was considered successful within 2 kb between CpG and gene transcription start sites (TSS). The integromes generated at this step represented 475,148 and 674,567 associations at the gene and transcript levels, respectively. As described in a similar setup[70], Spearman's correlations were calculated for each association. Correlations at the gene level were used for generating density plots and presenting a general view, whereas transcripts were used for precise analyses and final results. These were filtered into candidate transcriptional effector locations, by selecting "promoter regions" containing at least three negatively-correlated CpGs (rho < −1/3; *p* value <0.001) or three positively correlated CpGs (rho > 1/3; *p* value <0.001) with features corresponding to "TSS200," "TSS1500," "first exon," or "TSSoverlap2kb" (each linked to the same transcript identifier).

Manhattan representations were plotted against the background with the R CMplot package. Gene set enrichment and pathway analyses of selected candidate lists were carried out as described in *Methylome data analyses*. Interaction networks of putative TFs encoded by candidate genes, protein domain enrichments, and effector functions were performed with STRING tools [https://string-db.org/][71]. A curated database of 1639 human TFs with DNA-binding domain information was obtained from http://humantfs.ccbr.utoronto.ca/. Regulatory networks were built with NetworkAnalyst [www.networkanalyst.ca][72].

Methodology for building the gene expression-based scoring system

CLL-derived RS signature. A 215-gene set was obtained by extracting two clusters of strongly correlated up- and down-regulated profiles from the transcriptome hierarchical clustering tree (Fig. 3a, Supplementary Fig. 36, and Supplementary Data 9). The two initial clusters displayed a very high enrichment in CLL genes and mainly drove the whole sample aggregation process. These were further reduced to protein-coding genes, to avoid biases when applying the signature to transcriptomes of different origins, which may not contain ncRNAs or genes of undefined biotype. The reduced set was then overlapped with genes integrating significantly between transcriptome and methylome. The resulting 215-gene signature contained 93 protein-coding genes underexpressed in CLL-derived RS and 122 protein-coding genes overexpressed in CLL-derived RS.

Linear classifier score (LCS). For each analyzed dataset, scores were obtained according to the following procedure, to render the process as reproducible as possible. (i) When applicable, raw expression data with relevant sample annotations were retrieved from the Gene Expression Omnibus curated database (https://www.ncbi.nlm.nih.gov/geo/) with GEOquery[73]. Expression matrices were then prepared, described statistically, and normalized according to a well-established protocol[74]. Otherwise, already normalized expression data were used "as is". (ii) Whole transcriptomes were reduced to their features (genes, transcripts, probes) corresponding to matches with the 215-gene signature. (iii) Expression values were summed up over genes to obtain an aggregated and unique expression for each gene. (iv) Data were scaled, i.e., mean-centered and standard-deviation-reduced. (v) Positive outliers were trimmed at the last permille (99.9%) to reduce the impact of extreme gene expression values on the score but preserve high enough values as essential markers. Trimmed values were replaced with the last permille value. After a distribution check, no negative outliers were found in any dataset. (vi) For each of the 215 genes, weights were assigned: those originating from the upregulated cluster were weighted +1 and those originating from the downregulated cluster were weighted −1. (vii) Finally, LCS scores were computed as the mean of weighted gene expressions for each sample *S* of the dataset:

$$\mathrm{LCS}(S) = \frac{1}{n}\sum_{i=1}^{n} G_i.W_i \qquad (3)$$

with *n* the number of genes in the signature, and G_i representing the gene *i* weighted by W_i. LCS scores were then standardized (mean-centering to 0 and standard-deviation-reducing to obtain scores fully comparable between datasets). The obtained *Z*-scores were compared to a normal distribution in a one-way test to calculate a *p* value, used to define the initial LCS cutoff ($p < 0.05$) in each dataset.

Statistics and reproducibility

No statistical method was used to predetermine sample size. Data exclusion criteria according to quality controls are explained in the

"Methods" section. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Raw DNA methylation, gene expression and targeted NGS data generated in this study from RS samples have been deposited in the European Genome-Phenome Archive (study EGAS00001005495) under accession number EGAD00010002194 for DNA methylation data; accession number EGAD00001007922 for transcriptomic data, and accession number EGAD00001009509 for targeted NGS data. The raw data are protected and available under restricted access. Clinical and genomic data can be obtained by contacting the data access committee, according to the European Genome-Phenome Archive's procedure. Data access will be granted if their use complies with the data use conditions, including a commitment to strictly use these data for a clearly identified academic research programs and according to good practice recommendations. The Data Access Committee will respond to requests within 2 weeks. Once access to the data is granted, these are available until the end of the research program they support. Previously published DNA methylation datasets from the ICGC MMML-seq consortium that were used in this study are available upon request from the data access committee at the ICGC consortium data portal [https://dcc.icgc.org/]. Published datasets can be found under the following accession codes: GSE103265; GSE66770; GSE10846; GSE98588; GSE87371. All other data supporting the findings of this study are available from the corresponding authors upon request. Source data are provided with this paper.

Code availability
The source code developed for this study for designing the DNam and gene expression classifiers and the methylome–transcriptome integrative analyses is available on the GitHub platform, [https://github.com/zetcheuv/RichterOmicsCode]. All other source data supporting the findings of this study are available from the corresponding authors.

References
1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33 (2021).
2. Kipps, T. J. et al. Chronic lymphocytic leukaemia. *Nat. Rev. Dis. Prim.* **3**, 17008 (2017).
3. Hallek, M. et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood* **131**, 2745–2760 (2018).
4. Mao, Z. et al. IgVH mutational status and clonality analysis of Richter's transformation: diffuse large B-cell lymphoma and Hodgkin lymphoma in association with B-cell chronic lymphocytic leukemia (B-CLL) represent 2 different pathways of disease evolution. *Am. J. Surg. Pathol.* **31**, 1605–1614 (2007).
5. Alizadeh, A. A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
6. Reddy, A. et al. Genetic and functional drivers of diffuse large B cell lymphoma. *Cell* **171**, 481.e15–494.e15 (2017).
7. Schmitz, R. et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N. Engl. J. Med.* **378**, 1396–1407 (2018).
8. Chapuy, B. et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat. Med.* **24**, 679–690 (2018).
9. Dubois, S. et al. Refining diffuse large B-cell lymphoma subgroups using integrated analysis of molecular profiles. *EBioMedicine* **48**, 58–69 (2019).
10. Wright, G. W. et al. A probabilistic classification tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications. *Cancer Cell* **37**, 551–568.e14 (2020).
11. Vaisitti, T. et al. Novel Richter syndrome xenograft models to study genetic architecture, biology, and therapy responses. *Cancer Res.* **78**, 3413–3420 (2018).
12. Chakraborty, S. et al. B-cell receptor signaling and genetic lesions in TP53 and CDKN2A/CDKN2B cooperate in Richter transformation. *Blood* **138**, 1053–1066 (2021).
13. Iannello, A. et al. Synergistic efficacy of the dual PI3K-δ/γ inhibitor duvelisib with the Bcl-2 inhibitor venetoclax in Richter syndrome PDX models. *Blood* **137**, 3378–3389 (2021).
14. Vaisitti, T. et al. ROR1 targeting with the antibody-drug conjugate VLS-101 is effective in Richter syndrome patient-derived xenograft mouse models. *Blood* **137**, 3365–3377 (2021).
15. Schmid, T. et al. U-RT1 - a new model for Richter transformation. *Neoplasia* **23**, 140–148 (2021).
16. Scandurra, M. et al. Genomic profiling of Richter's syndrome: recurrent lesions and differences with de novo diffuse large B-cell lymphomas. *Hematol. Oncol.* **28**, 62–67 (2010).
17. Rossi, D. et al. The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation. *Blood* **117**, 3391–3401 (2011).
18. Fabbri, G. et al. Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. *J. Exp. Med.* **210**, 2273–2288 (2013).
19. Chigrinova, E. et al. Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. *Blood* **122**, 2673–2682 (2013).
20. Klintman, J. et al. Genomic and transcriptomic correlates of Richter's transformation in chronic lymphocytic leukemia. *Blood* **122**, 2800–2816 (2021).
21. Nadeu, F. et al. Detection of early seeding of Richter transformation in chronic lymphocytic leukemia. *Nat. Med.* **28**, 1662–1671 (2022).
22. Kulis, M. et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 1236–1242 (2012).
23. Oakes, C. C. et al. Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov.* **4**, 348–361 (2014).
24. Queirós, A. C. et al. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* **29**, 598–605 (2015).
25. Oakes, C. C. et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.* **48**, 253–264 (2016).
26. Beekman, R. et al. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.* **24**, 868–880 (2018).
27. Rinaldi, A. et al. Promoter methylation patterns in Richter syndrome affect stem-cell maintenance and cell cycle regulation and differ from de novo diffuse large B-cell lymphoma. *Br. J. Haematol.* **163**, 194–204 (2013).
28. Shaknovich, R. et al. DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood* **116**, e81–e89 (2010).
29. Chambwe, N. et al. Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes. *Blood* **123**, 1699–1708 (2014).
30. Pan, H. et al. Epigenomic evolution in diffuse large B-cell lymphomas. *Nat. Commun.* **6**, 6921 (2015).

31. Kretzmer, H. et al. DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. *Nat. Genet.* **47**, 1316–1325 (2015).
32. Queirós, A. C. et al. Decoding the DNA methylome of mantle cell lymphoma in the light of the entire B cell lineage. *Cancer Cell* **30**, 806–821 (2016).
33. Duran-Ferrer, M. et al. The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. *Nat. Cancer* **1**, 1066–1081 (2020).
34. Kulis, M. et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.* **47**, 746–756 (2015).
35. Lee, S. T. et al. A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res.* **40**, 11339–11351 (2012).
36. Bibikova, M. et al. High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–295 (2011).
37. Pidsley, R. et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* **17**, 208 (2016).
38. Wright, G. et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl Acad. Sci. USA* **100**, 9991–9996 (2003).
39. Lambert, S. A. et al. The human transcription factors. *Cell* **175**, 598–599 (2018).
40. Ecco, G., Imbeault, M. & Trono, D. KRAB zinc finger proteins. *Development* **144**, 2719–2729 (2017).
41. Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).
42. Lenz, G. et al. Stromal gene signatures in large-B-cell lymphomas. *N. Engl. J. Med.* **359**, 2313–2323 (2008).
43. Augé, H. et al. Microenvironment remodeling and subsequent clinical implications in diffuse large B-cell histologic variant of Richter syndrome. *Front. Immunol.* **11**, 594841 (2020).
44. Scheffold, A. et al. IGF1R as druggable target mediating PI3K-δ inhibitor resistance in a murine model of chronic lymphocytic leukemia. *Blood* **134**, 534–547 (2019).
45. Kitange, G. J. et al. Evaluation of MGMT promoter methylation status and correlation with temozolomide response in orthotopic glioblastoma xenograft model. *J. Neurooncol.* **92**, 23–31 (2009).
46. Esteller, M. et al. Hypermethylation of the DNA repair gene O(6)-methylguanine DNA methyltransferase and survival of patients with diffuse large B-cell lymphoma. *J. Natl Cancer Inst.* **94**, 26–32 (2002).
47. Wang, J. et al. FOXC1 regulates the functions of human basal-like breast cancer cells by activating NF-κB signaling. *Oncogene* **31**, 4798–4802 (2012).
48. Somerville, T. D. et al. Frequent derepression of the mesenchymal transcription factor gene FOXC1 in acute myeloid leukemia. *Cancer Cell* **28**, 329–342 (2015).
49. Wilson, W. H. et al. Effect of ibrutinib with R-CHOP chemotherapy in genetic subtypes of DLBCL. *Cancer Cell* **39**, 1643–1653.e3 (2021).
50. Soilleux, E. J. et al. Diagnostic dilemmas of high-grade transformation (Richter's syndrome) of chronic lymphocytic leukaemia: results of the phase II National Cancer Research Institute CHOP-OR clinical trial specialist haemato-pathology central review. *Histopathology* **69**, 1066–1076 (2016).
51. Moulin, C. et al. Clinical, biological, and molecular genetic features of Richter syndrome and prognostic significance: a study of the French Innovative Leukemia Organization. *Am. J. Hematol.* **96**, E311–E314 (2021).
52. Hübschmann, D. et al. Mutational mechanisms shaping the coding and noncoding genome of germinal center derived B-cell lymphomas. *Leukemia* **35**, 2002–2016 (2021).
53. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
54. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* **13**, R44 (2012).
55. Peters, T. J. et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 6 (2015).
56. Josse, J. & François, H. missMDA: a package for handling missing values in multivariate data analysis. *J. Stat. Softw.* **70**, 1–31 (2016).
57. Lena, P. D., Sala, C., Prodi, A. & Nardini, C. Methylation data imputation performances under different representations and missingness patterns. *BMC Bioinformatics* **21**, 268 (2020).
58. Van Iterson, M., Cats, D., Hop, P., Heijmans, B. T. & Consortium, B. omicsPrint: detection of data linkage errors in multiple omics studies. *Bioinformatics* **34**, 2142–2143 (2018).
59. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286–288 (2016).
60. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, 90–97 (2016).
61. Yu, G. & He, Q. Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
62. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
63. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
64. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
65. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
66. Cunningham, F. et al. Ensembl 2019. *Nucleic Acids Res.* **47**, D745–D751 (2019).
67. Pouget, C. et al. Ki-67 and MCM6 labeling indices are correlated with overall survival in anaplastic oligodendroglioma, IDH1-mutant and 1p/19q-codeleted: a multicenter study from the French POLA network. *Brain Pathol.* **30**, 465–478 (2020).
68. Ochoa, D. et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
69. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
70. Zgheib, R. et al. Folate can promote the methionine-dependent reprogramming of glioblastoma cells towards pluripotency. *Cell Death Dis.* **10**, 596 (2019).
71. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
72. Zhou, G. et al. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234–W241 (2019).
73. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).
74. Willekens, J. et al. Wnt signaling pathways are dysregulated in rat female cerebellum following early methyl donor deficiency. *Mol. Neurobiol.* **56**, 892–906 (2019).

Article https://doi.org/10.1038/s41467-022-34642-6

Acknowledgements
The authors would like to thank the divisions of clinical hematology, hematology laboratory and pathology of Nancy (Dr Hélène Busby, Pr Hervé Sartelet, Dr Ludovic Dubouis), Poitiers, Angers, Reims (Dr Pascale Cornillet-Lefebvre), Clermont-Ferrand (Dr Lauren Veronèse and Dr Albane Ledoux-Pilon), Tours (Dr Flavie Arbion), Avicenne, Saint-Louis (Dr Véronique Meignin) and Pitié-Salpêtrière (Dr Frédéric Charlotte and Pr Isabelle Brocheriou). The authors would like to thank the tumor libraries biological resource centers of Nancy (BB-0033-00035), Poitiers (BB-0033-00068), Caen (Pr Xavier Troussard), Tours, Clermont-Ferrand, Angers (BB -0033-00038), Reims-Champagne-Ardenne, Besançon (Franck Monnien, Dr Etienne Daguindau) and Bordeaux (Marie-Pierre Fort, Dr Fontanet Bijou) who provided us with the biological material. The authors would like to thank Véronique Saunier (direction of research at University Hospital of Nancy) for supporting the project. RNA sequencing was performed by the GenomEast platform, a member of the "France Génomique" consortium (ANR–10-INBS-0009). The authors would like to thank Louis Staudt (Center for Cancer Genomics, National Cancer Institute, Bethesda, MD 20892, USA) for giving access to the data published in Schmitz and colleagues (2018) and Wright and colleagues (2020). The authors would like to thank the members of ICGC the MMML-seq consortium for contribution to the generation of the DLBCL omics datasets and the MMML-seq consortium for data access. The authors would like to thank Pr Catherine Wu and Dr Erin Parry for shared expertise and language editing, and Dr Cath Carsberg for English language editing. This work was supported in part by the Cancéropôle Est (J.B., S.H., P.F.), the Ligue contre le Cancer (J.B., P.F.), the University Hospital of Nancy (J.B., P.F.), the association of SILLC patients (J.B., P.F.), and the Association des Chefs de Services of the University Hospital of Nancy (J.B.). E.T., S.S., and R.S. were supported by the DFG (SFB1074 projects B1, B9, and B10). The ICGC MMML-Seq consortium has been supported by the German Ministry of Science and Education in the framework of the ICGC MMML-Seq consortium (01KU1002) and ICGC DE-Mining (01KU1505).

Author contributions
Conception and design: J.B., S.H., P.F., R.S., S.S. Development and methodology: S.H., J.B., E.T., M.K., P.F., J.I.M.-S., R.S., S.S. Sample and clinical data providing, acquired and managed patients: C.D., D.R.W., A.Q., O.B., C. Tomowiak, G.L., G.G., S.L., E.C., F.N.K., F.D., A.R., M.-C.B., A.D., O.T., G.O., M.H., C. Thieblemont, R.G., J.I.M.-S. and F.C. Acquisition of data (data production and techniques, provided facilities): J.B., S.H., J.V., M.K., R.H., P.R., C.C., D.M., E.C., C.S., S.L., E.T., S.B., G.O., J.-L.G., ICGC MMML-seq consortium, MMML consortium, P.F., R.S., S.S. Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.H., J.B., C.M., C.S., A.M., M.K., R.S., S.S. Writing, review and/or revision of the manuscript: J.B., S.H., P.F., R.S. and S.S. wrote the first and the revised version of the paper. All authors critically reviewed and agreed on the final version of the manuscript. Administrative, technical and material support (i.e., reporting or organizing data, constructing databases): S.H., J.B., J.V., C.M., E.C., S.L., E.T., P.L., O.A., ICGC MMML-seq consortium, MMML consortium, P.F., R.S., S.S.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-34642-6.

Correspondence and requests for materials should be addressed to Julien Broséus or Stephan Stilgenbauer.

Peer review information *Nature Communications* thanks Silvia Deaglio and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

[1]Division of CLL. Department of Internal Medicine III, Ulm University, Ulm, Germany. [2]Inserm UMRS1256 Nutrition-Génétique et Exposition aux Risques Environnementaux (N-GERE), Université de Lorraine, Nancy, France. [3]Université de Lorraine, CHRU-Nancy, Service d'Hématologie Biologique, Pôle Laboratoires, F54000 Nancy, France. [4]Institute of Human Genetics, Ulm University & Ulm University Medical Center, Ulm, Germany. [5]Fraunhofer Institute for Cell Therapy and Immunology IZI, Leipzig, Germany. [6]Department of Haematology, University Hospital of Tours, Tours, France. [7]Department of Hematology, Hôpital de la Pitié-Salpêtrière, AP-HP, Paris, France. [8]Université de Reims Champagne-Ardenne, IRMAIC, Centre Hospitalier Universitaire de Reims, Hématologie Clinique, Reims, France. [9]Department of Hematology, University Hospital of Nancy, Vandoeuvre-lès-Nancy, France. [10]Inserm, CHRU, University of Lorraine, CIC Clinical Epidemiology, Nancy, France. [11]Department of Clinical Pathology, Robert-Bosch-Krankenhaus, and Dr. Margarete Fischer-Bosch Institute for Clinical Pharmacology, Stuttgart, Germany. [12]CHU Angers, Biological Resource Center of Angers (CRB-CHU Angers), BB-0033-00038, Laboratoire d'Hématologie, Angers, France. [13]Department of Hematology, CHU Poitiers, Poitiers, France. [14]CIC1402 Inserm Poitiers, Poitiers, France. [15]Hematology Laboratory, Avicenne Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France. [16]Bioinformatics Group, Department of Computer Science and Interdisciplinary Center for Bioinformatics, Leipzig University, Leipzig, Germany. [17]Hematology department, Clermont-Ferrand University Hospital, Clermont-Ferrand, France. [18]Department of Biopathology CHRU-ICL, BBB, CHRU Nancy, Vandoeuvre-lès-Nancy, France. [19]Biological Resource Center of Nancy, BB-0033-00035, CHRU de Nancy, Nancy, France. [20]Sorbonne Université, Cytogénétique Hématologique, Hôpital Pitié-Salpêtrière, AP-HP, Paris, France. [21]Centre de Recherche des Cordeliers, INSERM, Université Sorbonne Paris Cite, Université Paris Descartes, Université Paris Diderot, F-75006 Paris, France. [22]CHRU of Nancy, Service de Biochimie-Biologie Moléculaire-Nutrition, Pôle Laboratoires, F54000 Nancy, France. [23]Hematology Department, Hôpital Pitié-Salpêtrière,

Article https://doi.org/10.1038/s41467-022-34642-6

AP-HP, Sorbonne University, Paris, France. [24]Department of Hematology, University Hospital of Angers, Angers, France. [25]Institute of Pathology, University Hospital of Würzburg, Bavaria, Germany. [26]Hematology Biology, University Hospital of Nantes, Hôtel-Dieu, France. [27]Inserm 1232 Centre de Recherche en Cancérologie et Immunologie Nantes Angers (CRCINA), Nantes, France. [28]Department of Hematology, Hôpital Saint-Louis, Paris, France. [29]Division of Molecular Genetics, German Cancer Consortium (DKTK) and National Center for Tumor Diseases (NCT) Heidelberg, German Cancer Research Center (DKFZ), Heidelberg, Germany. [30]Biomedical Epigenomics Group, Institut d'investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain. [31]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [32]These authors contributed equally: Julien Broséus, Sébastien Hergalant. [33]These authors jointly supervised this work: Pierre Feugier, Reiner Siebert, Stephan Stilgenbauer.
✉e-mail: julien.broseus@univ-lorraine.fr; Stephan.Stilgenbauer@uniklinik-ulm.de

ICGC MMML-Seq Consortium

Ole Ammerpohl[4], Stephan Bernhart[16], Markus Kreuz[5], Peter Lichter[29], German Ott[11], Andreas Rosenwald[25], Reiner Siebert[4,33] & Stephan Stilgenbauer[1,33]

A full list of members and their affiliations appears in the Supplementary Information.

5. MULTI-OMICS CHARACTERIZATION OF RICHTER SYNDROME UNLOCKS CLASSIFIERS AND PREDICTORS OF OUTCOME INTO BROADER AND HETEROGENEOUS LYMPHOMA DATASETS

Synthèse de la communication

Les Journées Ouvertes de Biologie, Informatique et Mathématiques (JOBIM) sont une conférence avec comité de relecture, en anglais, internationalement reconnue pour sa qualité, son accessibilité pour les jeunes chercheurs et ingénieurs, et son intérêt participatif bien au-delà de la communauté bioinformatique francophone. Il y avait plus de 350 participants et une trentaine d'intervenants en 2023, avec une organisation un peu spéciale puisque le congrès était réparti sur cinq sites et en distanciel.

J'y ai détaillé une rétrospective et les perspectives multi-omiques de notre travail sur le syndrome de Richter, en mettant l'accent sur les aspects bioinformatiques, biostatistiques et méthodologiques présentés dans l'article *Nature Communications* précédent. En partant du design expérimental, qui adresse aussi bien les questions biologiques posées qu'il anticipe les problématiques techniques. Pour arriver sur l'élaboration des classifieurs à partir des observations empiriques, en montrant le fil du raisonnement et la prise de recul critique nécessaire pour évaluer la fiabilité et la validité des résultats.

Programme détaillé :

https://jobim2023.sciencesconf.org/resource/page?id=21&forward-action=page&forward-controller=resource&lang=en

Actes du congrès :

https://jobim2023.sciencesconf.org/data/pages/proceedings.pdf

Multi-omics characterization of Richter syndrome unlocks classifiers and predictors of outcome into broader and heterogeneous lymphoma datasets

Sébastien HERGALANT[1], Romain PIUCCO[1], Ghislain FIÉVET[1], Emil CHTEINBERG[2], Stephan STILGENBAUER[3], Reiner SIEBERT[2], David MEYRE[1], Pierre FEUGIER[1] and Julien BROSÉUS[1]

[1] Nutrition, Genetics and Environmental Risk Exposure (NGERE) - Inserm U1256, Campus Biologie Santé, Nancy, France

[2] Institute of Human Genetics, Ulm University Medical Center, Ulm, Germany

[3] Department of Internal Medicine III, Ulm University, Ulm, Germany

Corresponding Author: sebastien.hergalant@inserm.fr

Reference paper: **Broséus & Hergalant *et al.* (2023) Molecular characterization of Richter syndrome identifies de novo diffuse large B-cell lymphomas with poor prognosis, Nature Communications, 2023, 14, 309.**
https://doi.org/10.1038/s41467-022-34642-6

Richter syndrome (RS) is the onset of a dismal diffuse large B-cell lymphoma (DLBCL) subtype that exemplifies aggressiveness and chemoresistance occurring in the context of indolent chronic lymphocytic leukemia (CLL). In this study we characterize a large series of human RS samples [1] by genome-wide DNA methylation and whole-transcriptome profiling from a multi-omics setup including additional copy number alterations, exome and proteome data. We comprehensively compare them to i) the paired CLL component, i.e. samples acquired before the aggressive transformation in the same patients, ii) a broad CLL reference group [2] and iii) *de novo* DLBCLs of different cell-of-origin (COO) and molecular classification.

Recent genomic studies combining DNA and RNA sequencing extended DLBCL subtyping beyond COO, identifying subgroups defined by their genomic alteration patterns and associated clinical courses, but a notable proportion remains unclassified [3]. Distinguishing between CLL-derived RS and *de novo* DLBCL in a diagnostic setting based on histology and immunochemistry alone is challenging. Most RS cases arise from the preceding CLL clone, while the remainder are in fact independent *de novo* DLBCLs, a dichotomy of importance for treatment decisions. Indeed, *de novo* DLBCLs are chemosensitive in most patients, whereas CLL-derived RS are chemoresistant, with a median overall survival of around 12 months.

Adjusting for the lack of appropriate human or animal models to study RS, our integrative approach provides insights into its epigenomic architecture, corroborates two evolutionary groups of RS [4] and unravels a CLL epigenetic imprint in clonally related samples. Thus removing the need for the initial CLL tumor DNA, a significant improvement since CLL stage is often undiagnosed and/or corresponding samples unavailable. We also define two novel classifiers: a methylation-based predictor to detect a stable CLL "memory" over disease evolution, and a gene-expression-based scoring method outlining a novel DLBCL subgroup from public datasets harboring this CLL-derived RS epigenetic imprint. Applying both classifiers to omics data from landmark studies uncovers a subset of "RS-type" DLBCL enriched in cases with a specific COO signature, unclassified or undetected by other genomic classifiers, and with the same unfavorable prognosis as RS. These findings directly translate prognostication of *de novo* DLBLCs, the most common human B-cell lymphoma, and associate them linearly with overall and progression-free survival, independently of known clinical factors and biological covariates.

References

1. Charline Moulin, Francis Guillemin, Thomas Remen, Florian Bouclet *et al*. Clinical, biological, and molecular genetic features of Richter syndrome and prognostic significance: A study of the French Innovative Leukemia Organization. *American Journal of Hematology*, *96*(9), 311-314, 2021.
2. Renée Beekman, Vincente Chapaprieta, Núria Russiñol, Roser Vilarrasa-Blasi *et al*. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nature medicine*, *24*(6), 868-880, 2018.
3. George Wright, Da Wei Huang, James Phelan, Zana Coulibaly *et al*. A probabilistic classification tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications. *Cancer cell*, *37*(4), 551-568. 2020.
4. Erin Parry, Ignaty Leshchiner, Romain Guièze, Connor Johnson *et al*. Evolutionary history of transformation from chronic lymphocytic leukemia to Richter syndrome. *Nature medicine*, 29, 158–169, 2023.

6. LARGE-SCALE PROTEOMICS IDENTIFIES DISTINCT SIGNATURES FOR RICHTER SYNDROME AND *DE NOVO* DIFFUSE LARGE B-CELL LYMPHOMA: A FRENCH STUDY FROM THE FILO GROUP

Synthèse de la communication

L'ASH (*American Society of Hematology*) organise chaque année une conférence très prisée des hématologues, et très regardante sur la qualité des travaux sélectionnés pour une présentation orale. Nous avons eu la chance d'en faire partie, en co-présentant fin 2020 avec Julien Broséus nos résultats préliminaires sur le profil protéomique du syndrome de Richter, mis en concurrence avec celui d'autres lymphomes agressifs (LBDGC) et lymphomes folliculaires transformés. J'ai choisi d'intégrer ce résumé long car, sans rentrer dans les détails, il est représentatif du manuscrit principal correspondant à ce travail et qui est en cours de finition actuellement.

Nous nous intéressons ici particulièrement aux profils de chimiorésistance dans le SR et à l'identification de biomarqueurs et de voies cibles pour des médicaments / thérapies préexistantes. En plus de ce qui est présenté ici, nous intégrons désormais cette couche omique avec les autres déjà disponibles (exome, méthylome, transcriptome), et identifions des axes associés à la relation clonale LLC-RS et à la survie.

blood

622.LYMPHOMA BIOLOGY-NON-GENETIC STUDIES | NOVEMBER 5, 2020

Blood (2020) 136 (Supplement 1) : 29.

http://doi.org/10.1182/blood-2020-137061

Large-Scale Proteomics Identifies Distinct Signatures for Richter Syndrome and *De Novo* Diffuse Large B-Cell Lymphoma: A French Study from the Filo Group

Romain Morizot[1*], **Sébastien Hergalant**[2*], Romain Piucco[3], Florian Bouclet[4], Anne Quinquenel[5], Caroline Dartigeas[6], Hélène Augé[1], Eugen Tausch[7], Sandra Lomazzi[8], Hélène Busby[9], Cécile Tomowiak[10], Veronique Leblond[11], Catherine Thiéblemont[12], Florence Cymbalista[13], Marie C Bene[14], Stephan Stilgenbauer[15], Romain Guièze[16], Christine Carapito[17], Aurore Perrot[18], Luc-Mathieu Fornecker[19], Pierre Feugier[20] and **Julien Broséus**[21]

***Equal contribution**
Speaker: julien.broseus@univ-lorraine.fr ; sebastien.hergalant@inserm.fr

[1]Haematology Department, Nancy University Hospital, Vandoeuvre-lès-Nancy, France; [2]U1256, Inserm, Vandoeuvre-lès-Nancy, France; [3]U1256, Inserm, VANDOEUVRE, France; [4]CHU Estaing, Unit of adult cell therapy and clinical hematology, University Hospital of Clermont-Ferrand, Clermont-Ferrand, Auvergne, France; [5]Haematology Department, Robert Debré University Hospital, Reims, France; [6]CHU Tours, Tours, France; [7]Department of Internal Medicine III, Ulm University, Ulm, Germany; [8]Biocollection repository, Nancy University Hospital, Vandoeuvre-lès-Nancy, France; [9]Pathology Department, Nancy University Hospital, Vandoeuvre-lès-Nancy, France; [10]Department of Oncology-Haematology and Cell Therapy, University Hospital of Poitiers, Poitiers, France; [11]Département d' Hématologie Hôpital Pitié-Salpêtrière APHP, UPMC Université Paris, Paris, France; [12]Hôpital Saint-Louis, Paris, France; [13]Laboratoire d'Hématologie, APHP Hôpital Avicenne, Bobigny, France; [14]Hematology Biology, Nantes University Hospital, Nantes, France; [15]Department of Internal Medicine III, University Hospital Ulm, Ulm, Germany; [16]CHU Estaing, Unit of adult cell therapy and clinical hematology, University Hospital of Clermont-Ferrand, Brookline, MA; [17]Laboratoire de Spectrométrie de Masse BioOrganique, Université de Strasbourg, CNRS, IPHC, UMR 7178, Strasbourg, France; [18]Institut Universitaire du Cancer de Toulouse-Oncopole, Nancy, France; [19]Department of Hematology, Strasbourg University Hospital, Strasbourg, France; [20]Hematology Department, Nancy University Hospital, Vandoeuvre-lès-Nancy, France; [21]Laboratory Department, CHRU Nancy, Vandoeuvre-lès-Nancy, France

Richter syndrome (RS) occurs during the disease course of 2 to 10% of Chronic Lymphocytic Leukemia (CLL). Diffuse Large B-Cell Lymphoma (DLBCL) subtype accounts for 90-95% of RS cases. While presenting with the same morphology as *de novo* DLBCLs, DLBCL subtype of RS is associated with a very poor outcome.

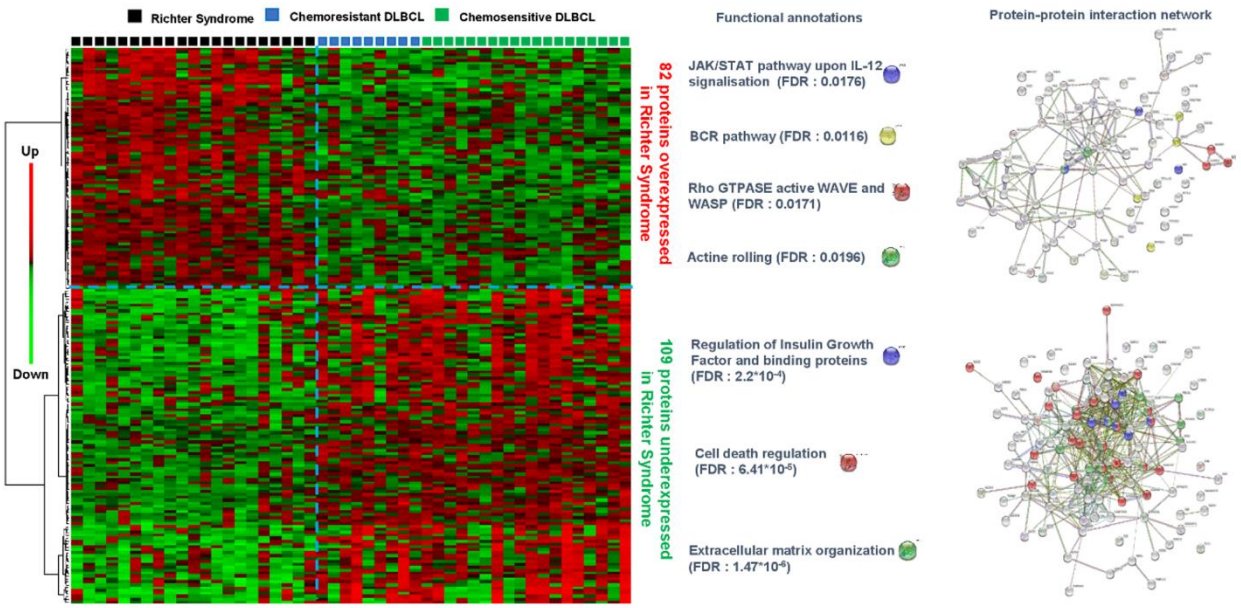
Proteins are the primary cellular biological effectors. Proteome composition is highly dependent on regulatory mechanisms located both upstream and downstream translation (transcriptional regulation, post-translational modifications, protein metabolism). Thus, the analysis of the genome and the transcriptome only allows a putative extrapolation of the expressed proteome. Proteomic studies have been performed in the context of de novo DLBCLs (Fornecker et al. Sci Rep. 2019), unravelling a set of proteins associated with refractoriness. In

CLL, it showed different profiles according to IGHV mutational status after B-Cell Receptor activation (Perrot et al. Blood 2011). No proteomic study of RS has been carried out to date.

RS sample selection was performed across 7 French institutions affiliated to the FILO (French Innovative Leukemia Organization). A total of 49 fresh frozen biopsies were collected, including 28 de novo DLBCLs and 21 RS, mostly treated with first line R-CHOP. All biopsies were centrally reviewed. RS samples were characterized, with data on CLL-RS clonal relationship and mutational status for a 13-gene panel representing the most frequently mutated genes in CLL. Only DLBCL subtype RS samples with at least 50% tumor purity (range 50-95%) and a minimum 10 mg weight were selected. Peptide measurements were performed using liquid chromatography coupled with tandem mass spectrometry, according to published methods (Muller et al. Sci Rep. 2018). Stringent quality controls were applied to ensure sample integrity, abundance accuracy and overall reproducibility. Proteome reconstruction at peptide and at protein level was achieved with a specifically devised pipeline involving conditional filtering, full normalization, categorization and imputation of missing values. These tools made use of the R/Bioconductor DEP package (Zhang et al. Nat Protoc. 2018). Supervised (Bayesian linear models) and unsupervised (hierarchical clustering, K-means, PCA) analyses were further applied to identify differential protein signatures. These were functionally annotated with ReactomePA (Yu et al. Mol Biosyst. 2016) for pathways and STRING (Szklarczyk et al. Nucleic Acids Res. 2019) for association networks.

Extended proteomics analysis identified 1,772 proteins, among which 191 were differentially expressed (False Discovery Rate/FDR < 0.05) in RS samples compared to *de novo* DLBCLs, with 82 increased and 109 decreased proteins. Hierarchical clustering revealed a highly correlated expression profile of these top candidates and clearly separated the 21 RS and the 28 *de novo* DLBCL samples (Figure 1). Sample distribution was independent from chemosensitivity/resistance, for DLBCL samples, and also unrelated to GCB/Non-GCB phenotype according to Hans algorithm, Epstein-Barr virus positivity or tumor purity. Functional interactome is an *in silico* protein-protein interaction network built on published

Figure 1: Heatmap of 191 differentially expressed proteins between Richter Syndrome (n=21) and de novo DLBCLs (n=28) (False Discovery Rate < 0.05) with main functional annotations and protein-protein interaction networks.



data from the literature and available in public databases. The functional interactome computed

from the 82 proteins overexpressed in RS showed a strongly enriched association network (protein-protein interactions; p-value < 1e-16), with an over-representation in BCR pathway, VEGF signaling, JAK-STAT pathway and Interleukin-12, Rho GTPase, and actin coiling (FDR < 0.05; Figure 1). Proteins underexpressed in RS (109) also displayed highly associated interactions (p-value < 1e-16) with a main node including proteins involved in cell death regulation, extracellular matrix organization, regulation of Insulin Growth Factor, and signaling by receptor tyrosine kinase (FDR < 0.05; Figure 1).

Here we performed proteomics on a 49-sample cohort of 28 *de novo* DLBCLs and 21 RS, which revealed a specific and differential signature in RS. This includes increased expression of targets within the druggable signaling pathways BCR and JAK-STAT. Furthermore the decrease in proteins involved in cell death regulation and extracellular matrix organization suggests resistance mechanisms to apoptosis and immune system in RS.

PARTIE 3 : GÉNÉTIQUE DE L'OBÉSITÉ

7. COMPREHENSIVE IDENTIFICATION OF PLEIOTROPIC ASSOCIATIONS FOR SERUM ADIPONECTIN LEVELS USING THE NHGRI-EBI CATALOG OF PUBLISHED GENOME-WIDE ASSOCIATION STUDIES

Contexte

L'article présenté ci-après n'est pas encore soumis à un journal scientifique, et constitue donc une communication confidentielle. Il a cependant atteint un stade très avancé, presque mature, bien que les résultats obtenus nous fassent désormais nourrir l'ambition de le conclure par une dernière expérience (randomisation mendélienne) qui changerait la donne car apportant une vue mécanistique et fonctionnelle sur ce que nous y présentons. Ce travail représentait l'opportunité d'une acquisition de nouvelles compétences en génétique des populations, et une plongée dans les problématiques autour de l'obésité, et du lien qui unit cette pathologie, par effets pléiotropiques, à l'inflammation et aux cancers par exemple, qui seront les 3 thématiques centrales de notre future équipe lors du prochain contrat quinquennal du laboratoire (2024-2028).

Nous travaillons ici sur les locus génétiques associés aux niveaux sériques de l'adiponectine, une adipokine importante pour l'homéostasie énergétique, et explorons leur potentiel pléiotropique, c'est-à-dire leurs associations et leurs effets sur d'autres traits phénotypiques et maladies. En d'autres termes, nous cherchons à savoir si les variants génétiques conduisant à des niveaux plus élevés d'adiponectine circulante sont aussi des variants prédisposant à manifester d'autres phénotypes, et si ceux-ci sont augmentés, ou diminués. Nous reconstruisons également un réseau d'interaction pléiotropique autour des gènes de l'adiponectine, ceux mis en cause par les locus génétiques identifiés plus haut, afin de proposer des mécanismes moléculaires pour les différents résultats observés. Enfin, nous comparons et interprétons les résultats obtenus avec les connaissances actuelles sur les mécanismes en jeu et les états physiologiques connus.

Une analyse de causalité inter-traits par randomisation mendélienne, permettant de reconstruire le réseau des relations causales autour de l'adiponectine, constituera la dernière étape de cette étude. Celle-ci est rendue possible par l'accession à la base de données *UK Biobank* contenant les génotypes et informations riches de plus d'un demi-million d'individus.

Points clés

- Association de 26 blocs génétiques indépendants avec les niveaux plasmatiques d'adiponectine. Ces blocs sont définis par déséquilibre de liaison avec les 56 SNPs de l'adiponectine identifiés par criblage exhaustif du *GWAS Catalog* et des données de la littérature, à la recherche de toutes les associations pangénomiques significatives sur des cohortes importantes, avec prise en compte de l'ethnicité lors de la découverte ($P < 5 \times 10^{-8}$; $N > 1000$ individus).
- Catalogage algorithmique de tous les autres traits du *GWAS Catalog* significativement associés aux 767 *proxy SNPs* et détermination de la direction de l'effet par rapport à la direction de l'adiponectine pour chacun de ses blocs génétiques.
- Développement d'une approche permettant de construire un réseau d'interaction protéine / protéine autour des gènes touchés par les SNPs de l'adiponectine. Identification de sous-réseaux pléiotropiques par propagation pas à pas des annotations fonctionnelles (ontologies sur les gènes, associations avec les maladies, les phénotypes, les types cellulaires et les autres génotypes) et par ajout d'interacteurs manquant, au fur et à mesure de la progression dans les branches du réseau.
- Les résultats obtenus par ces 2 techniques (sur les traits et sur les gènes) corrèlent très majoritairement avec les connaissances rapportées dans les modèles animaux et cellulaires, ou les mécanismes pressentis *in vivo* chez l'homme. Il y a un recoupement entre la génétique et la physiologie.
- Nous proposons pour finir un modèle d'action de l'adiponectine sur les traits cognitifs, hormonaux, cardiométaboliques, anthropométriques, comportementaux, psychologiques, pathogéniques, des biomarqueurs, et sur différentes fonctions tissulaires, que nous espérons pouvoir approfondir avec des analyses de causalité.

Comprehensive identification of pleiotropic associations for serum adiponectin levels using the NHGRI-EBI catalog of published genome-wide association studies

Sébastien Hergalant[1], Alicia A. Grima[2], David Meyre[1,2,3*]

[1]INSERM UMR_S 1256, Nutrition, Genetics, and Environmental Risk Exposure (NGERE), Faculty of Medicine of Nancy, University of Lorraine, Nancy, France; [2]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada; [3]Department of Molecular Medicine, Division of Biochemistry, Molecular Biology, and Nutrition, University Hospital of Nancy, Nancy, France.

[*]**Corresponding author**: Pr David Meyre, Inserm UMRS 1256 N-GERE (Nutrition-Genetics-Environmental Risks) - University of Lorraine, Faculty of Medicine of Nancy, Bâtiment C, 2ème étage, 9 Avenue de la Forêt de Haye, 54500, Vandoeuvre les Nancy, France, Tel: (33)-372746132, Email: david.meyre@univ-lorraine.fr; Department of Health Research Methods, Evidence, and Impact, McMaster University, Michael DeGroote Centre for Learning & Discovery, Room 3205, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada. Tel: 905.525.9140 Ext. 26802. Fax: 905.528.2814. Email: meyred@mcmaster.ca.

Key words: Serum adiponectin level, Genome-wide association study, Single nucleotide polymorphism, Pleiotropy, Psychologic and cognitive traits, Metabolic and physiologic traits, Tissue-specific functions, Diseases, Protein-protein interaction networks, Gene modules.

Abstract

Background

Adiponectin, an adipocyte secreted hormone, targets a variety of cell types, protecting against apoptosis and inflammation, and exhibiting anti-atherogenic and insulin-sensitising properties. Adiponectin levels are negatively associated with adiposity and several cardiometabolic conditions. Several rare and common genetic variants have been associated with serum adiponectin levels. Pleiotropic effects associated with circulating forms of adiponectin have been identified in patients. However, systematic investigation of gene pleiotropy between adiponectin genetic variants and other phenotypic traits is lacking. In this work we catalogue single nucleotide polymorphisms (SNPs) associated with serum adiponectin levels and explore their potential in the context of gene pleiotropy.

Methods

Using the genome-wide association studies (GWAS) Catalog, we inventoried SNPs reaching genome-wide significance for serum adiponectin levels. Next, we sought proxy SNPs (pSNPs) that were in strong to perfect linkage disequilibrium (LD, $r^2 > 0.9$) with the lead adiponectin SNPs (lSNPs). Adiponectin genetic blocks were defined as an agglomeration of SNPs in LD with the lSNPs ($r^2 \geq 0.1$) and were assigned ethnicities according to the reported lSNP discovery cohorts. We then investigated all other genome-wide significant associations between human traits and lSNPs + pSNPs with an ethnicity match. We computed and reported the direction of effects for each cross-traits relatively to an adiponectin level increase. We functionally annotated protein-coding genes mapped around the genetic blocks using an augmented protein/protein interaction network approach to identify pleiotropic modules highly associated with specific biological functions and pathways.

Results

We report 46 serum adiponectin level SNPs into 26 genetic blocks aggregating 721 pSNPs. After filtering, 1123 pleiotropic traits were summarized into 161 unique cross-trait associations which were included into 17 trait categories. The most enriched trait categories included leptin (a closely related adipokine), addiction-related behaviors, body fat and lipid traits, while insulin-related, inflammation (galectin-3), fat mass and executive function (cognitive) were among the top over-represented cross-traits. Direction of effect correlations with adiponectin levels largely confirmed physiological and mechanistical scientific literature in patients and animal models: negative with type 2 diabetes, cardiovascular and most anthropometric traits, positive by BMI, lipids, body fat and kidney activity. We further document new negative relations with bone mineral density, stress, anxiety and smoking, and positive ones with alcohol consumption, intelligence, and hematocrit, along with the molecular pathways leading to inflammation, hypoxia, fibrinolysis, vasodilation, angiogenesis and thermoregulation, and linked to brain development, chromatin remodeling, insulin signaling, oxidative phosphorylation and mTOR regulation.

Conclusions

Here we provide a comprehensive network of pleiotropic effects surrounding the adiponectin genetic architecture. Using integrative bioinformatics, we revisit the accumulated knowledge obtained with population genetics through association studies, reconstruct the pleiotropic networks around the adiponectin genes and highlight the pathways leading to glucose and fatty acid metabolism, angio- and atherogenesis, insulin signaling, liver and kidney functions, cognition and psychological-specific phenotypes, psychiatric disorders, inflammation, and cancers.

Abbreviations

AFR = Africans
ALAT = Alanine aminotransferase
AMR = Americans
ASAT = Aspartate aminotransferase
BMI = Body mass index
EAS = East Asians
EBI = European Bioinformatics Institute
EUR = Europeans
FDR = False discovery rate
gamma-GT = Gamma-glutamyl transferase
GO = Gene ontology
GWAS = Genome-wide association study
HbA1c = Glycated haemoglobin
HDL = High-density lipoprotein
LD = linkage disequilibrium
LDL = low-density lipoprotein
lSNP = Lead single nucleotide polymorphism
N/A = Not applicable
NAFLD = Non-alcoholic fatty liver disease
NCI = National Cancer Institute
NHGRI = National Human Genome Research Institute
NR = Not reported
NT5DC2 = 5' nucleotidase domain containing 2
OR = Odds ratio
PPI = Protein-protein interaction
pSNP = Proxy single nucleotide polymorphism
QC = Quality control
RBC = Red blood cell
SAS = South Asians
SNP = Single nucleotide polymorphism
T2D = Type 2 diabetes

TG = Triglyceride
UCSC = University of California Santa Cruz
VEGFA = Vascular endothelial growth factor A
vLDL = Very-low-density lipoprotein
WBC = White blood cell

1. Introduction

Adiponectin, commonly known as ACRP30, apM1, AdipoQ [1, 2], and GBP28 [2], is a hormone secreted predominantly by subcutaneous but also by visceral adipose tissue [3]. Adiponectin targets cell types in various tissues including pancreatic beta cells, liver and heart cells [4]. Its structure consists of 244 amino acids that form the initial protomer, an adiponectin monomer that is next combined into 3 main oligomeric forms of low (trimer), medium (hexamer) and high (at least 18 monomers) molecular weights, the latter representing 80% of all circulating adiponectin and being the most metabolically active [4, 5].The circulating plasma levels of adiponectin are up to 1000-fold higher than other hormones, such as leptin and insulin [4]. There are two adiponectin receptors, AdipoR1, and AdipoR2 [4, 6]. AdipoR1 has an ubiquitous expression, whereas AdipoR2 is expressed mainly in the liver [4]. Adiponectin protects from apoptosis, reduces inflammation, has anti-atherogenic and insulin-sensitizing properties, and acts as a messenger for communication between adipose tissue and metabolic organs [4, 7, 8]. Studies have shown that obese mice, both ob/ob or high-fat-diet models, have decreased adiponectin levels [4]. Adiponectin serum levels have also been inversely associated with various cardiometabolic traits in humans such as type 2 diabetes [9, 10], cardiovascular disease [10], and obesity [10]. However, because of the limitations of observational epidemiology to establish causal relationships between traits, more research is needed at this stage to understand the links between adiponectin and cardio-metabolic complications [11].

Serum adiponectin levels are modulated by diet, physical activity, smoking status, exposure to cold, sibutramine medication, and weight fluctuations [2, 12-17]. Adiponectin levels are also influenced by biological factors: sex, age, ethnicity, epigenetics, and genetics [4, 18-20]. Heritability estimates for serum adiponectin levels range from 33 to 93% in children and adults of diverse ethnicities [21-23]. Candidate gene, linkage, genome-wide, and exome-wide association studies have identified multiple low-frequency and common variants associated with serum adiponectin level [24-28].

Gene pleiotropy is the influence of one gene on more than one unrelated phenotypic trait, and can be biological/horizontal or mediated/vertical [29-31]. When a genetic variant influences more than one phenotype independently it is referred to as biological / horizontal pleiotropy. Mediated / vertical pleiotropy refers to the situation where a genetic variant influences one phenotypic trait which in turn influences a second phenotypic trait. The variant is then indirectly associated with the second trait [29, 32, 33]. There has been growing interest in using adiponectin-associated genetic variants to explore pleiotropic associations with other common traits. As an illustration, pleiotropic associations between adiponectin and traits such as insulin resistance [11, 34, 35], type 2 diabetes (T2D) [35, 36], obesity [37, 38], and coronary artery disease [39-41] have been identified in some but not all studies using Mendelian randomization. However, all these studies are driven by prior hypotheses, and no systematic / hypothesis-

free investigation of gene pleiotropy for adiponectin genetic variants has been reported to date [32]. Recently, Kaur et al. proposed an original methodology to perform such an agnostic investigation for waist to hip ratio [32]. This prompted us to analyze single nucleotide polymorphisms (SNPs) associated with serum adiponectin level in the context of gene pleiotropy in the NHGRI-EBI catalog of published genome-wide association studies.

2. Material and Methods

2.1 Participants

We used summary statistics from previously published literature. All corresponding studies were approved by local ethics committees, and were performed in accordance with relevant guidelines/regulations [42]. Written informed consent was obtained from each subject, their parents or legal guardians prior to participation, in accordance with the Declaration of Helsinki [42].

2.2 NHGRI-EBI Catalog of Published Genome-Wide Association Studies

The present work makes extensive use of the NHGRI-EBI Catalog of Published Genome-Wide Association Studies [43] (later referred to as GWAS Catalog) up to date from 9 July 2022 and annotated for the GRCh38/hg38 human reference genome. GWAS Catalog databases were downloaded from https://www.ebi.ac.uk/gwas/docs/file-downloads for local use, with versions tagged as 1.0.2 for GWAS associations and 1.0.3 for GWAS studies. Files were loaded into then interrogated under R 4.1 [44]. We restricted our analysis on the SNP associated trait only ('Mapped Trait') and did not focus on the background and disease trait information because of their limited value outside the original context. For each association, we retrieved the SNP strongest risk allele, P-value, effect size (odds ratio and/or beta) and direction of effect. Genomic positions, associated genes ('Mapped Genes'), genotyping technologies, related publications and study information were also stored for downstream analyses and manual checks. Discovery ('Initial Sample Size') and replication ('Replication Sample Size') cohorts were decomposed into 5 main populations along with the cumulated sample size in each: EUR (Europeans), EAS (East Asians), SAS (South Asians), AFR (Africans) and AMR (Americans). These were double-checked manually and further processed for use in ethnicity analyses.

2.3 Selection of SNPs associated with serum adiponectin level

We selected SNPs that reached genome-wide significant levels of association ($P < 5e\text{-}8$) for serum adiponectin in at least one study. A genome-wide significant $P < 5e\text{-}8$ cut-off was used in order to restrict our data to conclusive associations and avoid reporting on possible false positive results [45]. SNPs were extracted through an independent search from GWAS Catalog [43] using "adiponectin measurement" and "BMI-adjusted adiponectin measurement" as mapped trait, and by a comprehensive manual literature check following each publication IDs to ascertain reported information about risk alleles, P-values and effect sizes. Associations concerning small GWAS cohorts ($N < 1{,}000$), interaction studies, copy-number

variations or that did not use genome/exome-wide arrays as genotyping technology were excluded. We then used the 1000 Genome Project reference panel to identify whether the adiponectin SNPs located in or near the same gene were independent. SNPs were considered independent if $r^2 \leq 0.1$. Those in linkage disequilibrium (LD) were grouped as one into genetic blocks. For all adiponectin SNPs (lead SNPs (lSNP)), we identified proxy SNPs (pSNPs) that were in strong to perfect LD with the GWAS lSNPs ($r^2 \geq 0.9$). This information was retrieved from the LDlink tool website (https://ldlink.nci.nih.gov) [46] for the five ethnic groups from the 1000 Genome Project reference panel described above. In each discovery and replication cohort, multi-ethnic populations that displayed a predominant ethnic group (> 90%) were considered as a single ethnic group for the lSNPs. We used the R package LDlinkR [47] for custom batch queries, following the procedure for programmatic access to extensively compute all r^2 for each lSNP according to their determined ethnicities.

2.4 Cross-trait GWAS associations

For each adiponectin lSNP and pSNPs, we scanned GWAS Catalog for the presence of genome-wide significant (P < 5e-8) associations with other traits. Cross-trait associations concerning small cohorts (N < 1000), genotype by environment interaction studies, correlated traits, pairwise, multivariate or pleiotropic analyses, copy-number variations or that did not use genome/exome-wide arrays as genotyping technology were discarded for consistency in our inclusion criteria. Cases of missing data such as unknown risk alleles and unreported odds ratio or beta values, were first checked against the original publications and considered as not reported if still undetermined. Each individual remaining cross-trait study was codified for ethnicity as described above. In multi-ethnic cohorts, ethnic groups were retained when their sample size was > 1,000. We then added information on original lSNPs, genetic blocks and the previously determined block ethnicity and proceeded to check for ethnicity mismatches for each cross-trait. The cross-trait association was discarded as a mismatch if the discovery ethnicity for the adiponectin trait and the cross-trait (lSNP or pSNP) were different. For the remaining traits, direction of effect was encoded as positive or negative according to their increase or decrease with the reported risk allele and after investigating for the effect size variable type (beta or odds ratio). Then the direction of effect with the adiponectin risk alleles for each association was determined as a combination of 3 parameters: 1) the adiponectin SNP in strongest LD with the cross-trait SNP was selected as representative of the latter within the genetic block; 2) the adiponectin direction of effect was linked to the representing adiponectin risk allele; 3) cross-trait risk allele LD was checked against the adiponectin risk allele and if reversed the effect association was reversed. In case of unfound risk allele LD the process was reiterated with the next best representing adiponectin SNP. If still unfound, the direction of effect with adiponectin was considered undefined. Cross-trait direction of effects were reported as positive (homodirectional) or negative (opposite) as compared with serum adiponectin level traits. Finally, similar traits were hierarchized and catalogued into phenotypic categories to consolidate the data.

2.5 Gene-based functional annotations and pleiotropic networks

Mapped genes associated with each adiponectin genetic block were combined into 32 unique genes including 23 protein coding, 7 non-coding RNAs and 2 pseudogene biotypes. Functional annotations were conducted using the EnrichR resource (https://maayanlab.cloud/Enrichr/) [48] which was accessed programmatically and queried on multiple databases: 1) pathways: Reactome, KEGG, BioCarta and BioPlanet; 2) gene ontologies: GO molecular function and GO biological process; 3) phenotype-related: GWAS Catalog, PhenGenI, dbGAP and UK biobank; and 4) diseases: DisGenet and Jensen diseases. Functional terms and enrichments were reported if the over-representation test corrected q-value were < 0.05. We used multiple databases in each search category to further validate the results for consistency, as every ontology database undergoes varying curation processes and provides hierarchized structures focusing differently on depth, level of details and term labelling. The 23 protein-coding genes were used to infer protein/protein interaction (PPI) networks for the adiponectin in human using String (https://string-db.org/) v11.5 [49]: 1) first-rank interactors (from 20 to 30) were added on a median to high confidence basis (scores ranking from 0.4 to 0.7); 2) the obtained networks were clustered with kmeans to separate main interacting subgroups; 3) each cluster was enriched with 10 more high-confidence (score 0.7) first-rank interactors; 4) at each step, functional annotations were retrieved from either the provided String tool, or with EnrichR as described above; 5) steps 2, 3 and 4 were reiterated in each branch until the loss of initial adiponectin annotation enrichment in GWAS Catalog (q-value > 0.05). This protocol allowed us to move forward into the mechanistic network associated with adiponectin genes, evaluate their relationships and their relevance into its pleiotropic functions, and identify high-enrichment points associated to specific interactions. Each part of the PPI network could be analysed for their contributions to the reported cross-traits.

2.6 Statistics

The GWAS Catalog was used to extract SNPs associated with each trait. Pairwise r^2 values between the SNPs were calculated based on the 1000 Genome populations and were obtained through LDlink [46]. The SNPs were subsequently grouped into unique loci (genetic blocks) with $r^2 \leq 0.1$ as the threshold for independence [50]. For cross-trait GWAS associations, enrichments in cross-traits and trait categories were calculated following the relation: frequency of the observed trait / expected frequency, where the observed frequency represents the number of reported occurrences for the trait in our adiponectin selection among the total number of reported adiponectin cross-trait occurrences, and the expected frequency represents the overall significant occurrences for the same trait among the total number of significant associations in the complete GWAS Catalog (P < 5e-8). Two-way Fisher's exact test was used to evaluate the enrichment statistical significance. All statistical analyses were conducted using R V4.1 [44].

3. Results

3.1 GWAS-identified adiponectin SNPs and proxy SNP search

A search of GWAS studies for adiponectin and BMI-adjusted adiponectin serum level SNPs resulted in 58 associations and 46 unique lSNPs in 26 different LD blocks identified in 24 independent studies (Figure 1, Table 1, Table S1). Following ethnicity encoding, 721 pSNPs were identified in a search around the 46 adiponectin lSNPs (Table S1).

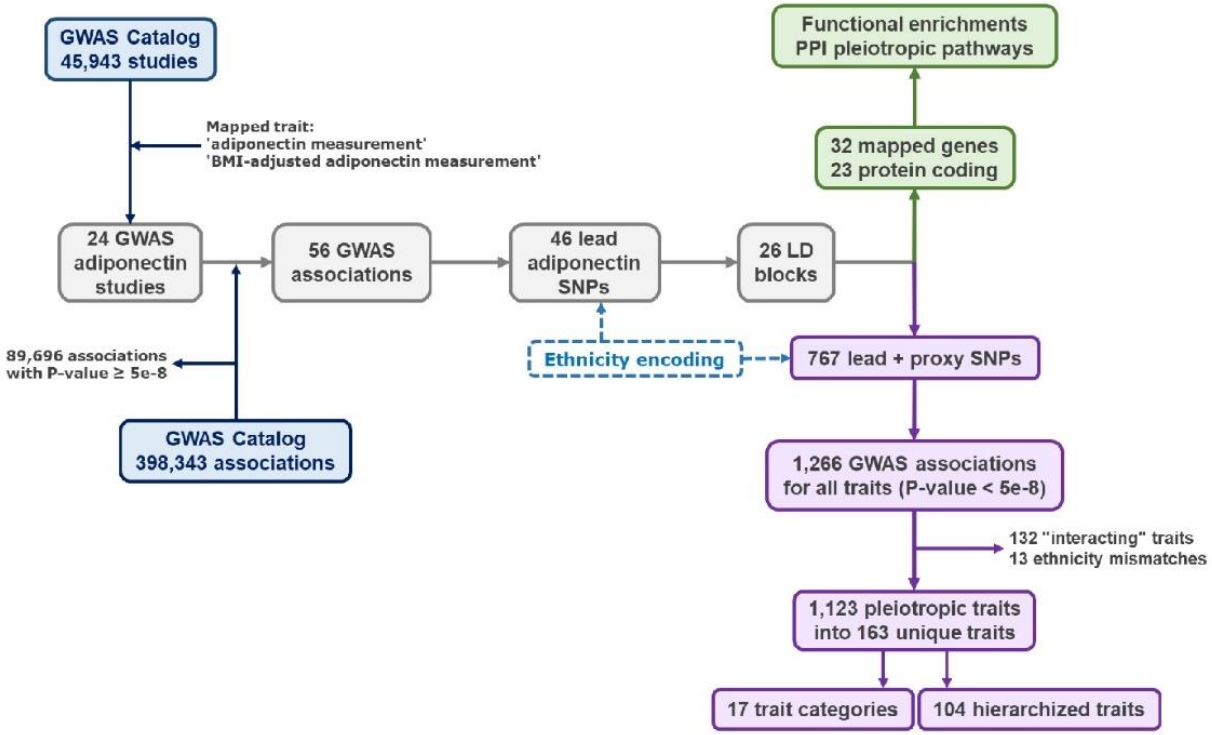


Figure 1. Flowchart illustrating the step-by-step extraction of adiponectin SNPs and cross-trait associations.

Abbreviations: BMI = body mass index; GWAS = genome-wide association study; LD = linkage disequilibrium; PPI = protein-protein interaction; SNP = single-nucleotide polymorphism.

Table 1. GWAS studies for serum adiponectin levels, with discovery cohort sizes, ethnicities and number of significant associations.

PMID	Study title	Study accession	First author	Initial sample size	Replication sample size	Ethnicity	Mapped trait	Disease / Trait	Significant SNPs	Platform [SNPs passing QC]	Genotyping
22065538	CDH13 gene coding T-cadherin influences variations in plasma adiponectin levels in the Japanese population.	GCST001314	Morisaki H	3,310 Japanese ancestry individuals		EAS	adiponectin measurement	Adiponectin levels	1	Illumina [348622]	Genome-wide genotyping array
21771975	A genome-wide association study reveals a quantitative trait locus of adiponectin on CDH13 that predicts cardiometabolic outcomes.	GCST001165	Chung CM	382 Han Chinese ancestry young-onset-hypertensive cases	559 Han Chinese ancestry young-onset-hypertensive cases	EAS	adiponectin measurement	Adiponectin levels	1	Illumina [509174]	Genome-wide genotyping array
28317342	Genetic determinants of adiponectin regulation revealed by pregnancy.	GCST004289	Hivert MF	1,322 European ancestry individuals	522 European ancestry, 174 French Canadian founder individuals	EUR	adiponectin measurement	Adiponectin levels in pregnancy	3	Illumina [NR]	Genome-wide genotyping array
26299439	Pathway-Based Genome-wide Association Studies Reveal That the Rac1 Pathway Is Associated with Plasma Adiponectin Levels.	GCST003096	Li WD	737 European ancestry females		EUR	adiponectin measurement	Adiponectin levels	5	Illumina [up to 550000]	Genome-wide genotyping array
26299439	Pathway-Based Genome-wide Association Studies Reveal That the Rac1 Pathway Is Associated with Plasma Adiponectin Levels.	GCST003095	Li WD	737 European ancestry females		EUR	BMI-adjusted adiponectin measurement	Adiponectin levels (BMI-adjusted)	4	Illumina [up to 550000]	Genome-wide genotyping array
31178129	Exome-Derived Adiponectin-Associated Variants Implicate Obesity and Lipid Biology.	GCST010050	Spracklen CN	up to 60,465 European ancestry, up to 2,568 East Asian ancestry, up to 3,271 African American, up to 1,435 Hispanic individuals		EUR, EAS, AFR, AMR	adiponectin measurement	Adiponectin levels	20	Illumina [265780]	Exome genotyping array [Exome array]
19165155	Genome-wide linkage and association analyses to identify genes influencing adiponectin levels: the GEMS Study.	GCST000319	Ling H	997 European ancestry cases, 989 European ancestry controls		EUR	adiponectin measurement	Adiponectin levels	2	Affymetrix [398625]	Genome-wide genotyping array
20018283	Clear detection of ADIPOQ locus as the major gene for plasma adiponectin: results of genome-wide association analyses including 4659 European individuals.	GCST000537	Heid IM	4,659 European ancestry individuals	13,795 European ancestry individuals	EUR	adiponectin measurement	Adiponectin levels	1	Affymetrix, Illumina [2585854]	Genome-wide genotyping array
21700879	Novel locus FER is associated with serum HMW adiponectin levels.	GCST001122	Qi L	1,591 European ancestry female individuals	626 European ancestry individuals from 235 families	EUR	adiponectin measurement	Adiponectin levels	5	Affymetrix [2543887]	Genome-wide genotyping array
24105470	A meta-analysis of genome-wide association studies for adiponectin levels in East Asians identifies a novel locus near WDR11-FGFR2.	GCST002233	Wu Y	7,827 East Asian ancestry individuals	10,252 East Asian ancestry individuals	EAS	adiponectin measurement	Adiponectin levels	10	Affymetrix, Illumina [~2500000]	Genome-wide genotyping array

PMID	Study title	Study accession	First author	Initial sample size	Replication sample size	Ethnicity	Mapped trait	Disease / Trait	Significant SNPs	[SNPs passing QC]	Genotyping
20011104	A genome-wide association study reveals variants in ARL15 that influence adiponectin levels.	GCST000538	Richards JB	8,531 European ancestry individuals	6,202 European ancestry individuals	EUR	adiponectin measurement	Adiponectin levels	3	Affymetrix, Illumina [~2200000]	Genome-wide genotyping array
20876611	Genome-wide association study for adiponectin levels in Filipino women identifies CDH13 and a novel uncommon haplotype at KNG1-ADIPOQ.	GCST000812	Wu Y	1,776 Filipino ancestry female individuals	1,774 Filipino ancestry offspring	EAS	adiponectin measurement	Adiponectin levels	3	Affymetrix [2073674]	Genome-wide genotyping array
22479202	Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals.	GCST001463	Dastani Z	29,347 European ancestry, 4,232 African American, 1,776 Filipino ancestry individuals		EUR, AFR, EAS	adiponectin measurement	Adiponectin levels	10	Affymetrix, Illumina [NR]	Genome-wide genotyping array
22479202	Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals.	GCST001465	Dastani Z	29,347 European ancestry individuals	10,536 European ancestry individuals	EUR	adiponectin measurement	Adiponectin levels	21	Affymetrix, Illumina [NR]	Genome-wide genotyping array
20887962	Adiponectin concentrations: a genome-wide association study.	GCST000828	Jee SH	4,001 Korean ancestry individuals	2,304 Korean ancestry individuals	EAS	adiponectin measurement	Adiponectin levels	1	Affymetrix [up to 354357]	Genome-wide genotyping array
28240269	Connecting genetic risk to disease end points through the human blood plasma proteome.	GCST901006210	Suhre K	997 European ancestry individuals	338 Greater Middle Eastern (Middle Eastern, North African or Persian), South Asian ancestry, Asian ancestry individuals	EUR	adiponectin measurement	Adiponectin levels	1	Affymetrix [509946]	Genome-wide genotyping array
33562295	Circulating Adiponectin and Its Association with Metabolic Traits and Type 2 Diabetes: Gene-Diet Interactions Focusing on Selected Gene Variants and at the Genome-Wide Level in High-Cardiovascular Risk Mediterranean Subjects.	GCST012167	Coltell O	348 European ancestry high cardiovascular risk elderly men		EUR	adiponectin measurement	Adiponectin levels	9	Illumina [625127]	Genome-wide genotyping array
33562295	Circulating Adiponectin and Its Association with Metabolic Traits and Type 2 Diabetes: Gene-Diet Interactions Focusing on Selected Gene Variants and at the Genome-Wide Level in High-Cardiovascular Risk Mediterranean Subjects.	GCST012166	Coltell O	606 European ancestry high cardiovascular risk elderly women		EUR	adiponectin measurement	Adiponectin levels	8	Illumina [625127]	Genome-wide genotyping array

PMID	Study title	Study accession	First author	Initial sample size	Replication sample size	Ethnicity	Mapped trait	Disease / Trait	Significant SNPs	Platform [SNPs passing QC]	Genotyping
33562295	Circulating Adiponectin and Its Association with Metabolic Traits and Type 2 Diabetes: Gene-Diet Interactions Focusing on Selected Gene Variants and at the Genome-Wide Level in High-Cardiovascular Risk Mediterranean Subjects.	GCST012165	Coltell O	450 European ancestry high cardiovascular risk elderly individuals		EUR	adiponectin measurement	Adiponectin levels in type 2 diabetes	1	Illumina [625127]	Genome-wide genotyping array
33562295	Circulating Adiponectin and Its Association with Metabolic Traits and Type 2 Diabetes: Gene-Diet Interactions Focusing on Selected Gene Variants and at the Genome-Wide Level in High-Cardiovascular Risk Mediterranean Subjects.	GCST012164	Coltell O	504 European ancestry high cardiovascular risk non-diabetic elderly individuals		EUR	adiponectin measurement	Adiponectin levels	4	Illumina [625127]	Genome-wide genotyping array
33562295	Circulating Adiponectin and Its Association with Metabolic Traits and Type 2 Diabetes: Gene-Diet Interactions Focusing on Selected Gene Variants and at the Genome-Wide Level in High-Cardiovascular Risk Mediterranean Subjects.	GCST012168	Coltell O	954 European ancestry high cardiovascular risk elderly individuals		EUR	adiponectin measurement	Adiponectin levels	8	Illumina [625127]	Genome-wide genotyping array
32915782	Adiponectin GWAS loci harboring extensive allelic heterogeneity exhibit distinct molecular consequences.	GCST90011881	Spracklen CN	9,262 European ancestry individuals		EUR	BMI-adjusted adiponectin measurement	Adiponectin levels (BMI-adjusted)	12	Illumina [16607452]	Genome-wide genotyping array
32915782	Adiponectin GWAS loci harboring extensive allelic heterogeneity exhibit distinct molecular consequences.	GCST90011882	Spracklen CN	9,262 European ancestry individuals		EUR	adiponectin measurement	Adiponectin levels (fat percentage adjusted)	0	Illumina [16607452]	Genome-wide genotyping array
32915782	Adiponectin GWAS loci harboring extensive allelic heterogeneity exhibit distinct molecular consequences.	GCST90011883	Spracklen CN	9,262 European ancestry individuals		EUR	BMI-adjusted adiponectin measurement	Adiponectin levels (BMI-adjusted)	0	Illumina [39120605]	Genome-wide genotyping array

Abbreviations: AFR = African; AMR = American; BMI = Body mass index; BP = Blood pressure; EAS = East Asian; EUR = European; GWAS = Genome-wide association studies; NR = Not reported; QC = Quality control; SAS = South Asian; SNP = Single nucleotide polymorphism.

3.2 Cross-trait associations

First, we extracted 308,647 significant (P < 5e-8) associations from the GWAS Catalog database. In total, 767 lSNPs and pSNPs were searched for cross-trait associations in the GWAS Catalog and as a result 1,266 GWAS associations were found. Following the removal of associations with ethnicity mismatch (n=13) or gene x environment interactions (n=132), 1,123 pleiotropic associations for 163 unique traits remained (Table S2), including 56 associations and 2 unique traits for adiponectin and BMI-adjusted adiponectin levels that were removed from further analysis, 436 lSNP cross-traits and 631 pSNP cross-traits. Figure 1 provides an overview of the search and extraction process.

The 161 unique traits were classified into 104 hierarchized traits and 17 trait categories (Table S3). The 1,123 cross-trait associations for each trait category were as followed: adipokines (n=7 association signals), inflammation (n=4), branched amino acids (n=3), lipids (n=472), sexual hormones (n=49), blood formula (n=87), biomarkers (n=36), anthropometric (n=231), body fat (n=13), cognitive function (n=16), neuropsychiatric and neurologic (n=11), addiction-related (n=10), blood pressure (n=15), cardiovascular (n=12), kidney function (n=16), hepatic function (n=36), and type 2 diabetes (T2D) (n=49) related traits. These observed trait category occurrences were normalized against background for all 308,647 significant traits in the GWAS Catalog and reported as enrichments as compared to the expected values (Table S3) and visually summarized for all 17 trait categories and 104 hierarchized traits (Figure 2).

Overall, the adiponectin pleiotropic panorama displayed significantly higher than expected enrichments with adipokines/leptin (29.8-fold; p=6.5e-9; two-way Fisher's exact test), addiction-related (8.25-fold; p=3.5e-7), body fat (5.3-fold; p=9.5e-7), lipids (4.2-fold; p=2.3e-170), branched amino-acids (3.2-fold; p=0.03), type 2 diabetes (3.1-fold; p=6.5e-12), hepatic function (1.8-fold; p=4.6e-4) and cardiovascular (1.7-fold; p=0.04) traits, indicating preferential SNP loci within adiponectin blocks for these associations, and lower than expected enrichments with cognitive (0.4-fold; p=6.9e-6) and blood formula (0.6-fold; p=1.1e-10) traits, indicating a more dispersed SNP distribution for these associations (Figure 2A). A more detailed view let appear high enrichments (>2-fold; n≥2; p<0.05) in SNPs associated with circulating insulin, leptin, smoking behavior, alcohol use disorder, anxiety, high-density lipoproteins (HDL), very-low-density lipoproteins (vLDL), triglycerides, metabolic syndrome, birth weight, familial hyperlipidemia and waist-hip ratio, while lowest enrichments (>0.5-fold; n≥2; p<0.05) concerned SNPs involved in red blood cell (RBC) distribution, mean corpuscular hemoglobin, platelet count, mean corpuscular volume and monocyte count (Figure 2B). A few categories regrouped a majority of enriched traits (body fat, lipids, addictions, sexual hormones, branched amino acids, cardiovascular and T2D), but for others, trait enrichments mixed highs and lows, according to their nature, and the number and distribution of SNPs involved (cognition, neuropsychiatric, inflammation, anthropometric, blood formula, biomarkers, hepatic and kidney function).

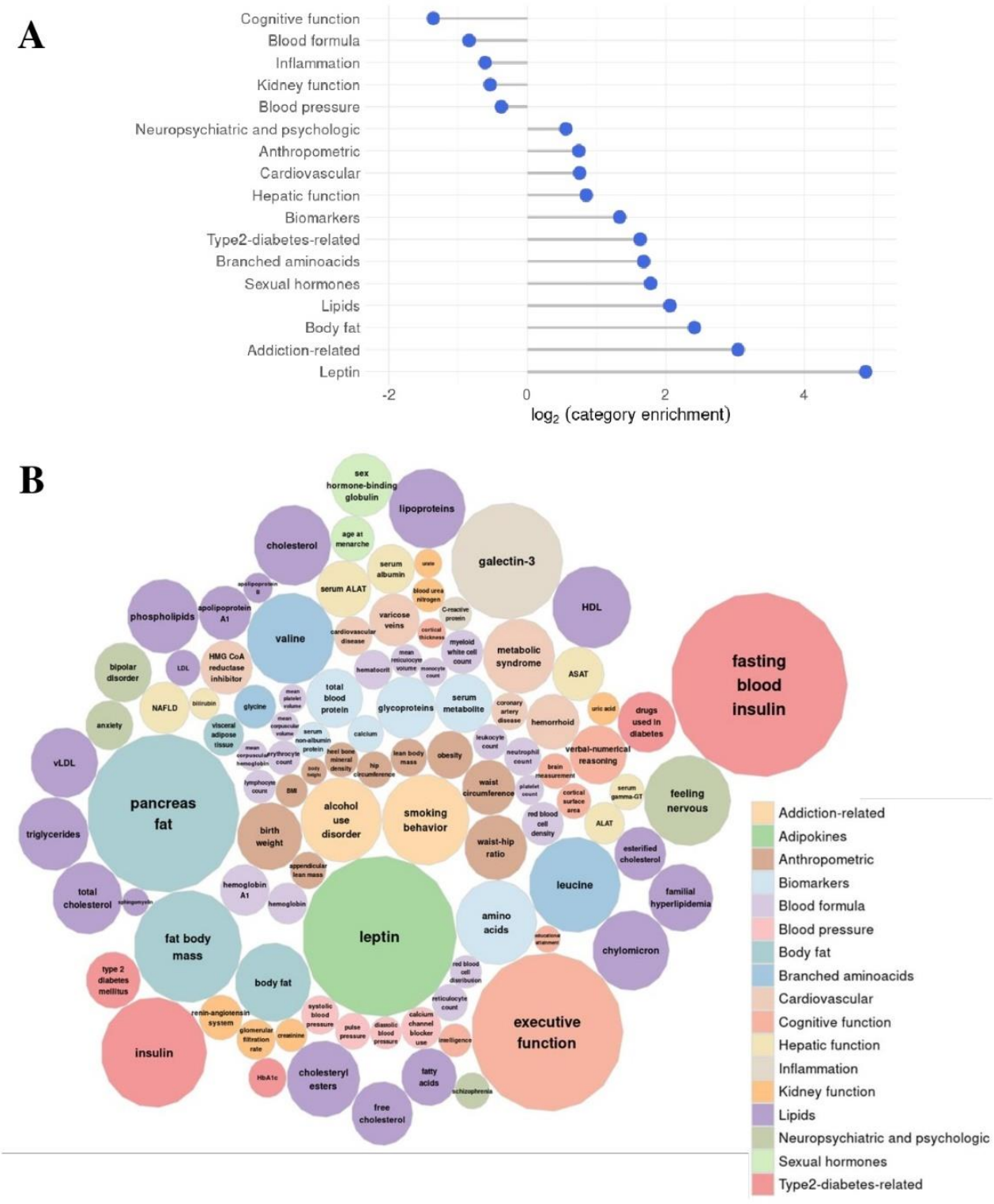


Figure 2. Cross-trait associations with adiponectin. **(A)** Normalized enrichments for each of the 17 adiponectin pleiotropic trait categories as compared to the GWAS Catalog background. **(B)** Visual representation of the 104 classified traits where the size of each circle is proportional to the normalized enrichment of pleiotropic associations found for each as compared to background.

Abbreviations: ALAT = alanine aminotransferase; ASAT = aspartate aminotransferase; HbA1c = glycated haemoglobin; HDL = high-density lipoprotein; LDL = low-density lipoprotein; NAFLD = non-alcoholic fatty liver disease; gamma-GT = gamma-glutamyl transferase; VLDL = very-low-density lipoprotein.

3.3 Cross-trait direction of effect with adiponectin

All 56 adiponectin lSNPs reporting publications were verified thoroughly for determination of effect allele, effect size (beta value) and direction of effect was encoded as positive or negative for an increase or a decrease in serum adiponectin levels, respectively (Table S2). Cross-trait lSNPs and pSNPs underwent the same codification process, then were each assigned a representative adiponectin lSNP within their genetic block (highest LD, see Methods) and each cross-trait effect allele was tested against its assigned lSNP effect allele to determine the final direction of effect with adiponectin. We tagged positive (+) for homodirectional effects, negative (-) for opposite directions, undefined if the effect allele and/or beta/odds ratio could not be determined, and aggregated the results for each 161 cross-trait and each 17 categories (Table 2, Table S2).

Homodirectional effects with adiponectin were observed with alcohol consumption (n=2/2 defined SNPs, located in the same genetic block); BMI (n=5/6) and hip circumference (n=11/13) but no other anthropometric traits; calcium measurement (n=2/2); mean corpuscular volume (n=3/3), neutrophil percentage (n=2/2), platelet count (n=2/2) and RBC distribution (n=4/4) for blood formula traits; body fat percentage (n=6/6); intelligence (n=3/3 SNP clustering in one genetic block around *GLN3* and *PBRM1*); glomerular filtration rate (n=6/8) as unique positively associated kidney function trait; and total cholesterol (n=14/14), total cholesteryl esters (n=11/11), free cholesterols (n=11/11), fatty acids (n=12/12) and apolipoprotein A1 (n=9/10), among others included in lipid traits.

Opposite directions were found with smoking behavior traits (n=7/7); leptin (n=7/7), the only other pleiotropic adipokine association, with all SNPs clustering on a single genetic block near *ADIPOQ*; birth weight (n=7/7 SNPs also in the same single block near *ADIPOQ*), BMI-adjusted waist-hip ratio (n=83/92) and heel bone mineral density (n=5/5) for anthropometric traits; hematocrit (n=3/3), hemoglobin-related (n=11/12), lymphocyte-related (n=5/5), neutrophil count (n=2) and reticulocyte traits (n=11/11) for blood formula; all blood pressure traits (n=11/12); branched amino acids (n=2/2); all cardiovascular traits (n=7/7); brain (n=2/2) and cortical surface (n=4/4) measurements; serum gamma-glutamyl transferase (n=3) as the only hepatic function trait; creatinine measurements (n=2/2) and urea-related traits (n=3/3) for kidney function; familial hyperlipidemia (n=3/3), triglyceride measurements (n=106/125) and all vLDL measurements (n=52/54) for lipids; stress and anxiety (n=3/3 SNPs in a single genetic block including *GLN3* and *PBRM1*) for psychologic traits; and all insulino-resistance and T2D (n=23/23).

Mixed effects, with positive and negative pleiotropic associations with adiponectin levels, were reported for inflammation, with an expected negative effect with a galectin-3 SNP and surprisingly consistent positive correlations with C-reactive protein measurements (n=3/3 SNPs clustering on the same genetic block around *GIMAP7*); unadjusted waist-hip ratio (7 positive and 10 negative counts) as anthropometric trait; aspartate aminotransferase (3+ and 5-) and serum alanine aminotransferase (5+ and

5-) measurements as hepatic function traits; and sexual hormones with a negative correlation for age at menarche and 23+/17- associations for sex hormone-binding globulin measurements.

Table 2. Pleiotropic variants within adiponectin genetic blocks, 161 unique cross-trait associations with adiponectin SNPs in strong LD and direction of effects relative to adiponectin traits.

GWAS mapped trait (original name)	Direction of effect with adiponectin level traits			Significant associations (P < 5e-8)	Number of adiponectin blocks	Genetic block IDs
	+	-	undef			
Addiction-related	**2**	**7**	**1**	**10**	**6**	**B3, B5, B9, B25, B27, B37**
alcohol use disorder measurement	2	0	0	2	1	
smoking behavior, BMI-adjusted waist circumference	0	2	1	3	2	
smoking behavior, BMI-adjusted waist-hip ratio	0	5	0	5	4	
Adipokines	**0**	**7**	**0**	**7**	**1**	**B9**
BMI-adjusted leptin measurement	0	4	0	4	1	
leptin measurement	0	3	0	3	1	
Anthropometric	**45**	**145**	**41**	**231**	**14**	**B1, B3, B5, B9, B25, B26, B27, B32, B34, B35, B36, B37, B38, B44**
appendicular lean mass	3	1	1	5	4	
birth weight	0	6	1	7	1	
birth weight, parental genotype effect measurement	0	1	0	1	1	
BMI-adjusted hip circumference	9	2	0	11	7	
BMI-adjusted waist circumference	8	36	12	56	9	
BMI-adjusted waist-hip ratio	8	76	8	92	9	
BMI-adjusted waist-hip ratio, physical activity measurement	1	7	0	8	5	
body height	0	0	1	1	1	
body mass index	5	1	5	11	8	
heel bone mineral density	0	5	5	10	2	
hip circumference	2	0	0	2	1	
lean body mass	0	0	1	1	1	
obesity	1	0	0	1	1	
waist circumference	1	0	0	1	1	
waist-hip ratio	7	10	7	24	9	
Biomarkers	**4**	**5**	**27**	**36**	**8**	**B3, B5, B9, B28, B29, B30, B37, B44**
amino acid measurement	0	1	0	1	1	
calcium measurement	2	0	0	2	2	
glycoprotein measurement	0	1	0	1	1	
serum metabolite measurement	0	0	27	27	1	
serum non-albumin protein measurement	1	0	0	1	1	
total blood protein measurement	1	3	0	4	4	

GWAS mapped trait (original name)	Direction of effect with adiponectin level traits			Significant associations (P < 5e-8)	Number of adiponectin blocks	Genetic block IDs
	+	-	undef			
Blood formula	**23**	**42**	**22**	**87**	**9**	**B3, B5, B9, B25, B27, B28, B29, B37, B44**
erythrocyte count	1	1	3	5	3	
hematocrit	0	3	4	7	2	
hemoglobin A1 measurement	0	3	0	3	3	
hemoglobin measurement	1	8	1	10	3	
leukocyte count	2	3	1	6	3	
lymphocyte count	0	3	1	4	2	
lymphocyte percentage of leukocytes	0	2	0	2	1	
mean corpuscular hemoglobin	1	0	2	3	2	
mean corpuscular hemoglobin concentration	0	2	0	2	1	
mean corpuscular volume	3	0	0	3	3	
mean platelet volume	0	0	1	1	1	
mean reticulocyte volume	1	0	0	1	1	
monocyte count	1	0	0	1	1	
monocyte percentage of leukocytes	1	0	0	1	1	
myeloid white cell count	2	3	1	6	3	
neutrophil count	0	2	4	6	2	
neutrophil percentage of leukocytes	2	0	0	2	1	
neutrophil-to-lymphocyte ratio	0	0	1	1	1	
platelet count	2	0	1	3	1	
red blood cell density measurement	2	1	2	5	4	
red blood cell distribution width	4	0	0	4	2	
reticulocyte count	0	9	0	9	4	
reticulocyte measurement	0	2	0	2	1	
Blood pressure	**1**	**11**	**3**	**15**	**7**	**B1, B3, B5, B9, B27, B29, B44**
Calcium channel blocker use measurement	0	1	0	1	1	
diastolic blood pressure	0	1	1	2	2	
pulse pressure measurement	1	3	0	4	4	
systolic blood pressure	0	6	2	8	6	
Body fat	**7**	**3**	**3**	**13**	**6**	**B1, B3, B9, B25, B27, B44**
body fat percentage	6	0	2	8	4	
fat body mass	0	1	0	1	1	
lean mass-adjusted fat body mass	1	0	0	1	1	
pancreas fat measurement	0	0	1	1	1	
sum of skinfolds	0	1	0	1	1	
visceral adipose tissue measurement	0	1	0	1	1	
Branched amino acids	**0**	**2**	**1**	**3**	**2**	**B3, B30**
glycine measurement	0	0	1	1	1	
leucine measurement	0	1	0	1	1	

GWAS mapped trait (original name)	Direction of effect with adiponectin level traits			Significant associations (P < 5e-8)	Number of adiponectin blocks	Genetic block IDs
	+	-	undef			
valine measurement	0	1	0	1	1	
Cardiovascular	**0**	**7**	**5**	**12**	**5**	**B3, B9, B27, B30, B37**
cardiovascular disease	0	0	1	1	1	
coronary artery disease	0	0	3	3	3	
hemorrhoid	0	0	1	1	1	
HMG CoA reductase inhibitor use measurement	0	2	0	2	2	
metabolic syndrome	0	4	0	4	4	
Varicose veins	0	1	0	1	1	
Cognitive function	**3**	**7**	**6**	**16**	**3**	**B5, B28, B38**
brain measurement	0	2	1	3	2	
brain measurement, neuroimaging measurement	0	0	1	1	1	
cortical surface area measurement	0	4	0	4	2	
cortical thickness	0	0	1	1	1	
educational attainment	0	1	0	1	1	
executive function measurement	0	0	1	1	1	
intelligence	3	0	0	3	1	
neuroimaging measurement, brain volume measurement	0	0	1	1	1	
verbal-numerical reasoning measurement	0	0	1	1	1	
Hepatic function	**10**	**17**	**9**	**36**	**9**	**B1, B3, B5, B9, B26, B27, B28, B29, B44**
aspartate aminotransferase measurement	3	5	4	12	5	
aspartate aminotransferase to alanine aminotransferase ratio	1	0	0	1	1	
bilirubin measurement	0	1	0	1	1	
non-alcoholic fatty liver disease	0	0	1	1	1	
serum alanine aminotransferase measurement	5	5	4	14	4	
serum albumin measurement	1	3	0	4	4	
serum gamma-glutamyl transferase measurement	0	3	0	3	2	
Inflammation	**3**	**1**	**0**	**4**	**1**	**B29**
C-reactive protein measurement	3	0	0	3	1	
galectin-3 measurement	0	1	0	1	1	
Kidney function	**7**	**8**	**1**	**16**	**3**	**B3, B11, B26**
Agents acting on the renin-angiotensin system use measurement	1	1	0	2	1	
blood urea nitrogen measurement	0	1	0	1	1	
creatinine measurement	0	2	0	2	1	
glomerular filtration rate	6	2	1	9	2	
urate measurement	0	1	0	1	1	
uric acid measurement	0	1	0	1	1	

GWAS mapped trait (original name)	Direction of effect with adiponectin level traits			Significant associations (P < 5e-8)	Number of adiponectin blocks	Genetic block IDs
	+	-	undef			
Lipids	**227**	**224**	**21**	**472**	**15**	**B1, B3, B5, B9, B25, B26, B27, B28, B29, B30, B32, B34, B36, B37, B44**
apolipoprotein A 1 measurement	9	1	0	10	6	
apolipoprotein B measurement	0	1	0	1	1	
cholesterol:total lipids ratio, high density lipoprotein cholesterol measurement	8	0	0	8	4	
cholesterol:total lipids ratio, intermediate density lipoprotein measurement	2	0	0	2	2	
cholesterol:total lipids ratio, low density lipoprotein cholesterol measurement	4	0	0	4	3	
cholesteryl ester measurement, chylomicron measurement, very low density lipoprotein cholesterol measurement	0	2	0	2	2	
cholesteryl ester measurement, high density lipoprotein cholesterol measurement	12	0	0	12	4	
cholesteryl ester measurement, intermediate density lipoprotein measurement	1	0	0	1	1	
cholesteryl ester measurement, very low density lipoprotein cholesterol measurement	0	5	0	5	3	
cholesteryl esters:total lipids ratio, high density lipoprotein cholesterol measurement	9	0	0	9	4	
cholesteryl esters:total lipids ratio, intermediate density lipoprotein measurement	2	0	0	2	2	
chylomicron measurement, total cholesterol measurement, very low density lipoprotein cholesterol measurement	0	2	0	2	2	
chylomicron measurement, triglyceride measurement, very low density lipoprotein cholesterol measurement	0	2	0	2	2	
chylomicron measurement, very low density lipoprotein cholesterol measurement	0	2	0	2	2	
chylomicron measurement, very low density lipoprotein cholesterol measurement, lipid measurement	0	2	0	2	2	
chylomicron measurement, very low density lipoprotein cholesterol measurement, phospholipid measurement	0	2	0	2	2	
esterified cholesterol measurement	1	0	0	1	1	
familial hyperlipidemia	0	3	0	3	2	
fatty acid measurement	11	0	0	11	4	
fatty acid measurement, linoleic acid measurement	1	0	0	1	1	
free cholesterol measurement, chylomicron measurement, very low density lipoprotein cholesterol measurement	0	2	0	2	2	
free cholesterol measurement, high density lipoprotein cholesterol measurement	9	0	0	9	3	
free cholesterol measurement, intermediate density lipoprotein measurement	1	0	0	1	1	

GWAS mapped trait (original name)	Direction of effect with adiponectin level traits			Significant associations (P < 5e-8)	Number of adiponectin blocks	Genetic block IDs
	+	-	undef			
free cholesterol measurement, low density lipoprotein cholesterol measurement	1	0	0	1	1	
free cholesterol measurement, very low density lipoprotein cholesterol measurement	0	6	0	6	3	
free cholesterol:total lipids ratio, high density lipoprotein cholesterol measurement	5	2	0	7	4	
high density lipoprotein cholesterol measurement	64	36	6	106	11	
high density lipoprotein particle size measurement	3	0	0	3	3	
lipid measurement, high density lipoprotein cholesterol measurement	10	0	0	10	3	
lipoprotein measurement	2	0	0	2	2	
low density lipoprotein cholesterol measurement	0	0	3	3	2	
low density lipoprotein cholesterol measurement, free cholesterol:total lipids ratio	6	0	0	6	3	
low density lipoprotein particle size measurement	1	0	0	1	1	
phospholipid measurement, high density lipoprotein cholesterol measurement	9	0	0	9	3	
phospholipids:total lipids ratio, high density lipoprotein cholesterol measurement	2	6	0	8	5	
sphingomyelin measurement	1	0	0	1	1	
total cholesterol measurement, high density lipoprotein cholesterol measurement	9	0	0	9	4	
total cholesterol measurement, intermediate density lipoprotein measurement	1	0	0	1	1	
total cholesterol measurement, very low density lipoprotein cholesterol measurement	0	6	0	6	3	
triglyceride measurement	19	44	12	75	11	
triglyceride measurement, high density lipoprotein cholesterol measurement	0	6	0	6	3	
triglyceride measurement, intermediate density lipoprotein measurement	0	1	0	1	1	
triglyceride measurement, low density lipoprotein cholesterol measurement	0	8	0	8	3	
triglyceride measurement, phospholipid measurement	0	4	0	4	4	
triglyceride measurement, very low density lipoprotein cholesterol measurement	0	13	0	13	3	
triglycerides:total lipids ratio, high density lipoprotein cholesterol measurement	0	15	0	15	4	
triglycerides:total lipids ratio, intermediate density lipoprotein measurement	0	2	0	2	2	
triglycerides:total lipids ratio, low density lipoprotein cholesterol measurement	0	6	0	6	2	
triglycerides:total lipids ratio, very low density lipoprotein cholesterol measurement	0	7	0	7	2	
very low density lipoprotein cholesterol measurement	0	9	0	9	3	

GWAS mapped trait (original name)	Direction of effect with adiponectin level traits			Significant associations (P < 5e-8)	Number of adiponectin blocks	Genetic block IDs
	+	-	undef			
very low density lipoprotein cholesterol measurement, cholesterol:total lipids ratio	6	0	0	6	2	
very low density lipoprotein cholesterol measurement, cholesteryl esters:total lipids ratio	7	0	0	7	3	
very low density lipoprotein cholesterol measurement, free cholesterol:total lipids ratio	9	0	0	9	4	
very low density lipoprotein cholesterol measurement, lipid measurement	0	10	0	10	3	
very low density lipoprotein cholesterol measurement, phospholipid measurement	0	8	0	8	3	
very low density lipoprotein cholesterol measurement, phospholipids:total lipids ratio	2	8	0	10	4	
very low density lipoprotein particle size measurement	0	3	0	3	3	
Neuropsychiatric and psychologic	**0**	**3**	**8**	**11**	**1**	**B5**
anxiety	0	1	0	1	1	
bipolar disorder	0	0	4	4	1	
feeling nervous measurement	0	2	0	2	1	
schizophrenia	0	0	4	4	1	
Sexual hormones	**23**	**18**	**8**	**49**	**8**	**B1, B3, B5, B9, B26, B27, B37, B44**
age at menarche	0	1	2	3	2	
sex hormone-binding globulin measurement	23	17	6	46	6	
Type2-diabetes-related	**0**	**23**	**26**	**49**	**7**	**B1, B3, B25, B26, B27, B37, B44**
BMI-adjusted fasting blood insulin measurement	0	3	0	3	3	
Drugs used in diabetes use measurement	0	3	0	3	2	
fasting blood insulin measurement	0	3	0	3	3	
HbA1c measurement	0	2	1	3	3	
insulin measurement	0	4	0	4	4	
type 2 diabetes mellitus	0	8	25	33	5	

Abbreviations: + = risk allele homodirectional effect; - = risk allele reverse effect; GWAS = Genome-wide association studies; undef = undefined risk allele.

3.4 Molecular pathways and PPI network augmentation

Using phenotype/genotype, gene ontology (GO), pathway and disease databases, we functionally annotated the molecular network encompassing the 23 protein-coding genes which overlapped the 26 adiponectin genetic blocks and that were reported as "mapped genes" in the GWAS Catalog (Table S4).

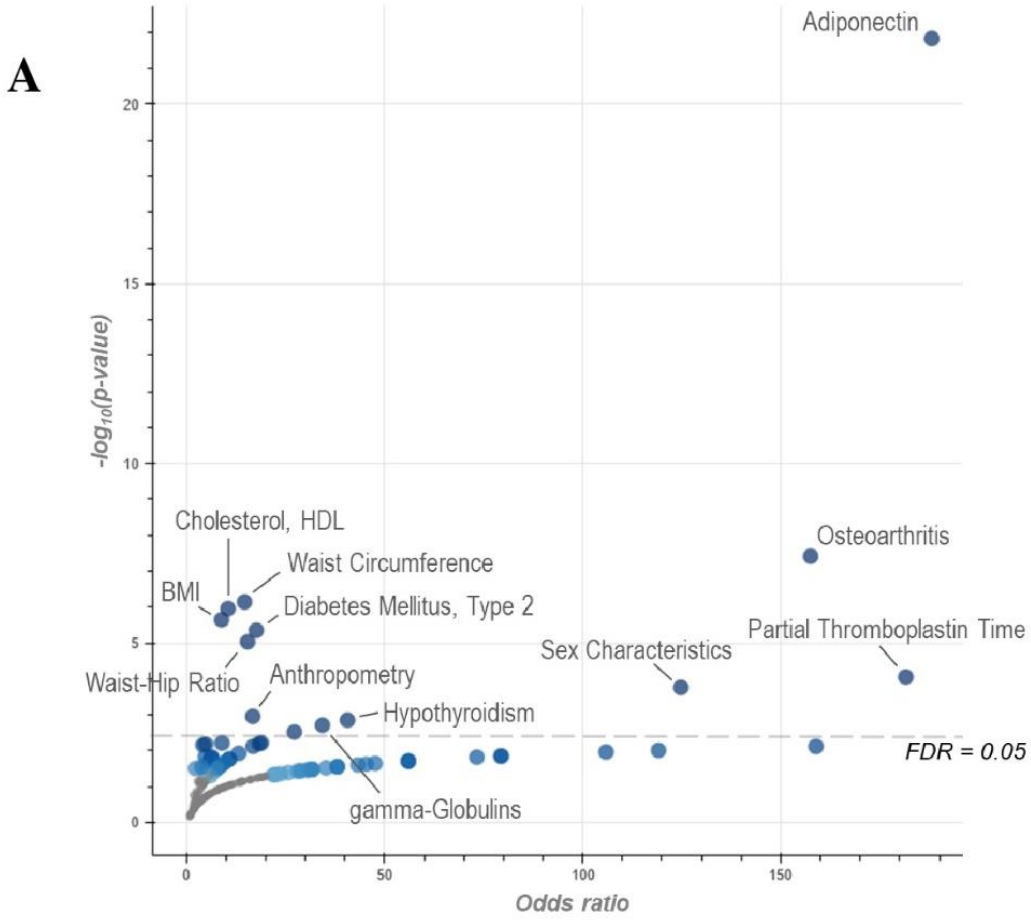
Among the most enriched terms uncovered "as is" were adiponectin, osteoarthritis, partial thromboplastin time, sex characteristics, hypothyroidism, gamma-globulins, T2D, HDL and

anthropometric traits (Figure 3A; FDR<0.05; Phenotype/Genotype Integrator (PheGenI)). These terms overlapped with findings from other databases, which also included tunica media, respiratory function tests and myocardial infarctions (database of genotype and phenotype (dbGAP)), coronary atherosclerosis, whole or part of body fat mass, ischaemic heart disease, diastolic blood pressure, RBC distribution, birth weight, neutrophil and lymphocyte percentage (UK Biobank GWAS v1) (Table S4; all FDR<0.05). These findings are consistent with the genetic pleiotropic associations found with previous cross-trait analyses. Enriched pathways included actions of nitric oxide in the heart, VEGF, hypoxia and angiogenesis, and visceral fat deposits and the metabolic syndrome (Figure 2B, left panel; BioCarta), response to elevated platelet cytosolic Ca2+ (Reactome), chromatin remodelling and control of gene expression, interferon-beta enhancer, validated nuclear estrogen receptor beta network, complement and coagulation cascades, dissolution of fibrin clot, adipocytokine signaling pathway and AMPK signaling (BioPlanet) (Table S4; all FDR<0.05). GO terms pointed out low-density lipoprotein (LDL) particle binding as molecular functions, and positive regulations of several biological processes: positive chemotaxis, vascular permeability, MAP kinase activity and reproduction (Table S4; all FDR<0.05).

However, these pathways and GO results remain incomplete and too imprecise, even after filtering (FDR<0.05) the overlaps between the 23-gene list and the over-represented terms was often too low to allow for the identification of gene modules with high enrichments in informative cellular and molecular functions. To circumvent this, we constructed PPI networks around the adiponectin-associated genes and added first-rank robust interactors to increase the number of possible connections and clustered this augmented network into pleiotropic modules (see Methods). Indeed, interacting proteins often participate in the same biological functions. As demonstrated in Figure 3B, augmented network annotations uncover new pathways and highlight or conceal others by modifying their relative enrichments and over-representation metrics. For example, neuronal migration and development, Lis1 pathway increases in significance and reveal new potential causal genes with network augmentation (FDR from 0.03 to 3.6e-05; enrichment from 48 to 90), whereas visceral far deposit and the metabolic syndrome decreases (FDR=0.02; enrichment from 130 to 70), indicating rippling and center effects for the culprit genes, respectively.

After two steps of network augmentation (+30 first-rank medium-to-high-confidence interactors, k-means clustering into 5 modules, +10 first-rank high-confidence interactors in every cluster), we functionally analyzed each of the highly pleiotropic modules obtained (Figure 3C, Table S5, with 5 modules depicted in different colors). The "red" module clustered proteins around initial ADIPOQ, ARL15, CDH13, CLIP1, CMIP, DNAH10, FAM13A and PEPD. This module over-represented functions such as leptin and adiponectin, dynein and dynactin complexes, microtubule motor activity, microtubule plus-end, centrosome, neuronal migration, cerebral cortex development, retrograde axonal transport, fatty acid oxidation, glucose transport, mitosis, lissencephaly and hypoplasia of the brainstem. The "yellow" module started from PBRM1 and annotated with neurogenesis, mental retardation, chromatin organization, chromatin binding, regulation of transcription, thermogenesis, and positive regulation of glucose mediated signaling. The "green" module included ATP6V0A2 and GNL3 and extended functions like ribosome biogenesis, insulin receptor signaling, insulin receptor recycling, transferrin transport, transferrin recycling, oxidative phosphorylation, synaptic vesicle and amino acid regulation of mTOR. The "cyan" module was the central hub of the complete PPI network and rejected

most starting genes apart from VEGFA. Functions associated with this module included VEGF signaling, HIF1A signaling and hypoxia, axon guidance, axon extension, neuron projection, autonomic nervous system development, innervation, vasculature, generation of neurons and regulation of cell death. Last, the "purple" module linked HRG and KNG1 and presented with functions including positive regulation of fibrinolysis, fibrin clot formation (intrinsic pathway and clotting cascade), vasodilation, plasminogen activation, cell adhesion mediated by integrins, inflammatory response, blood coagulation, heparin binding, extracellular matrix disassembly, glycosaminoglycan binding and blood vessel development (Figure 3C; Table S5; all FDR<0.05). Of note, proteins unable to interact with the extended network most likely represented genes incorrectly mapped with the discovery SNPs, or with still unknown functions.



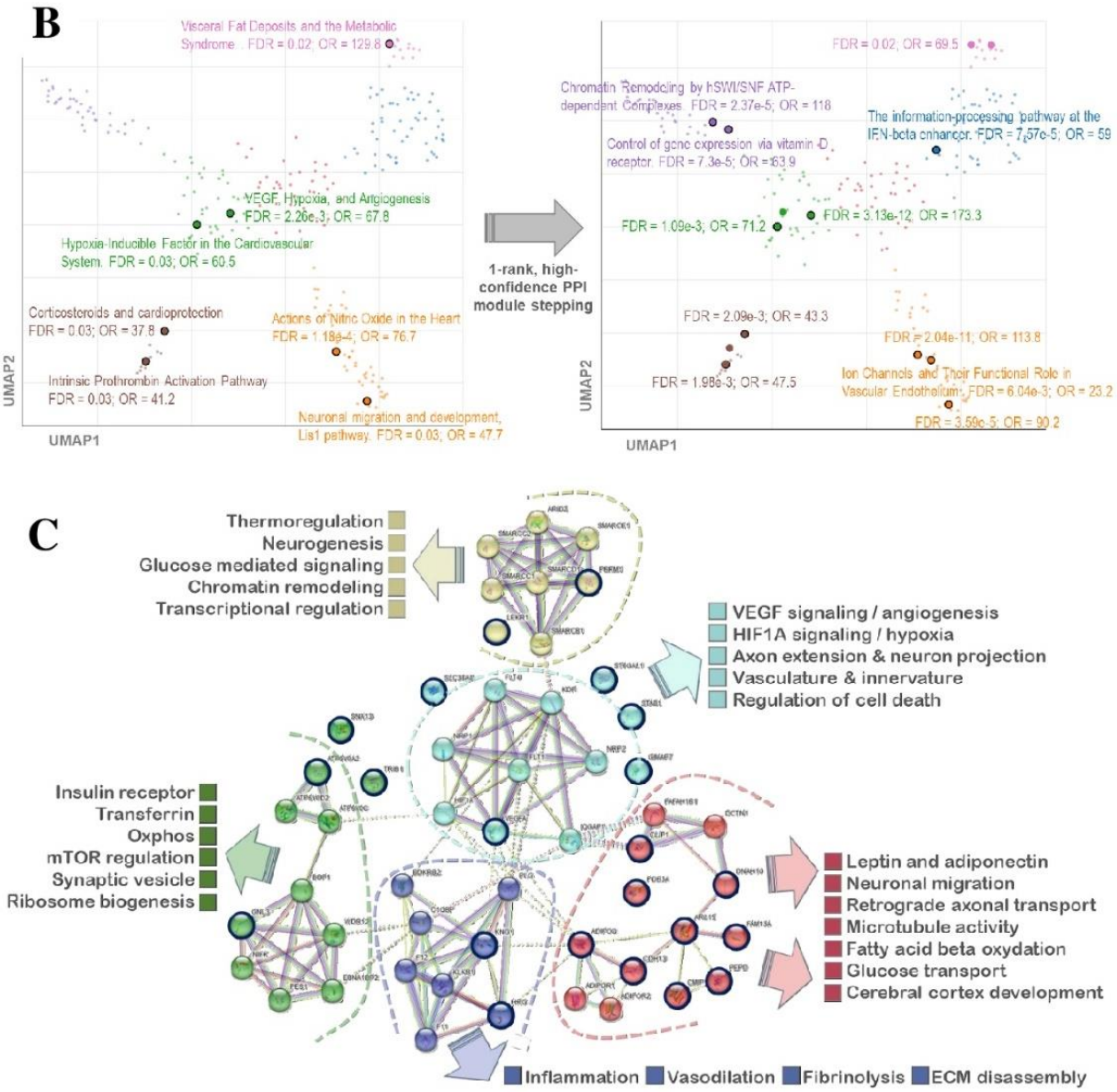


Figure 3. Molecular pathways linked to the adiponectin gene network. (A) Functional annotations of 23 protein-coding genes overlapping the 26 adiponectin genetic blocks using the Phenotype-Genotype Integrator (PheGenI 2021) merging most trait association databases, including GWAS Catalog. **(B)** First step augmentation of the adiponectin PPI network with 20 high-confidence (> 0.7) interactors pinpoint new pathways and modifies relative enrichments. Term frequency-inverse document frequency values are calculated for each gene sets (here, from terms in the Biocarta 2016 database). Obtained values are projected in UMAPs. **(C)** After k-means clustering of the adiponectin network, each cluster is augmented as exemplified in (B) and functionally annotated against molecular function databases. The 23 starting genes are circled in dark blue.

Abbreviations: BMI = Body mass index; ECM = extracellular matrix; FDR = False Discovery Rate; GWAS = Genome-wide association studies; HDL = High-density lipoprotein; OR = Odds ratio;

Oxphos = Oxidative phosphorylation; PPI = Protein-protein interactions; UMAP = Unified Manifold Approximation and Projection.

3.4 Diseases and cancers

Association with diseases were conducted on the initial 23 adiponectin genes (Table S4) and on the augmented adiponectin gene network (Table S6; +30 first-rank medium-to-high-confidence interactors). Many databases catalogue disease terms and pathways in diseases. Here we mainly focused on KEGG (Kyoto Encyclopedia of Genes and Genomes) disease pathways, DisGenet (gene variants-disease associations) and Jensen Diseases (gene-disease associations mined from literature) databases.

KEGG most enriched terms included pathways in cancer (n=5 genes; 4-fold enrichment), hepatocellular carcinoma (n=7; 20-fold), renal cell carcinoma (n=2; 12-fold) and rheumatoid arthritis (n=5; 24-fold) (Table S6; KEGG; FDR<0.05). With Jensen diseases, most representative terms covered carcinoma (n=40; 2.6-fold), kidney cancer (n=14; 2.5-fold), lung cancer (n=5; 5.3-fold), melanoma (n=5; 4-fold), cerebrovascular disease (n=6; 11.8-fold), coronary disease (n=6; 12.9-fold), vascular disease (n=5; 29-fold), angioedema (n=5; 151.5-fold), hypertension (n=8; 12.7-fold), T2D (n=5; 7.9-fold) and factor XII deficiency (n=5; 530.4-fold) (Table S6; Jensen Diseases; FDR<0.05; n≥5). The top 50 DisGenet annotations encompassed thrombopathies, kidney diseases, retinopathies and macular degenerations, and up to 13/50 cardiovascular and 17/50 cancer terms. Among these cancers and tumors were lymphatic metastasis (n=6; 19-fold), neoplasms, metastasis (n=27; 4.4-fold), myeloid leukemia (n=10; 9.4-fold), brain neoplasms (n=11; 8.2-fold), epithelial ovarian cancer (n=15; 5.7-fold), stomach neoplasms (n=12; 7-fold) and malignant neoplasms of lung (n=20; 4.5-fold). Other relevant diseases included obesity (n=13; 3-fold), hyperlipidemia (n=6; 11.4-fold), atherosclerosis (n=13;5.6-fold), cancer relapse (n=3; 111-fold), smoking behaviors (n=6;12.7-fold), chronic obstructive airway diseases (n=11;7.9-fold), ischemic colitis (n=3; 152.6-fold) and inflammation (n=10; 11.1-fold) (Table S6; DisGenet; FDR < 0.001). These diseases and cancers unfold on functions and pathways consistent with our previous molecular findings, and mostly coincide with the proposed pleiotropic modules.

4. Discussion

The cross-trait analysis conducted in this study identified traits sharing unique genetic associations with serum adiponectin levels. These traits could be organized into the following 17 categories: addictions, adipokines, anthropometric traits, biomarkers, blood formula, blood pressure, body fat, branched amino acids, cardiovascular traits, cognitive function, hepatic function, inflammation, kidney function, lipids, neuropsychiatric and psychologic traits, sexual hormones and T2D-related traits. The most enriched trait categories included leptin (a closely related adipokine), addiction-related behaviors, body fat and lipid traits, while insulin-related, inflammation (galectin-3), fat mass and executive function (cognitive) were among the top over-represented cross-traits. Direction of effect correlations with adiponectin levels largely confirmed physiological and mechanistical scientific literature in patients and animal models: negative with T2D, cardiovascular and most anthropometric

traits, positive by BMI, lipids, body fat and kidney activity. We further document new negative relations with bone mineral density, stress, anxiety and smoking, and positive ones with alcohol consumption, intelligence, and hematocrit, along with the molecular pathways leading to inflammation, hypoxia, fibrinolysis, vasodilation, angiogenesis and thermoregulation, and linked to brain development, chromatin remodeling, insulin signaling, oxidative phosphorylation and mTOR regulation.

We further provide a comprehensive network of pleiotropic effects surrounding the adiponectin genetic architecture. Using integrative bioinformatics, we revisit the accumulated knowledge obtained with population genetics through association studies, reconstruct the pleiotropic networks around the adiponectin genes and highlight the pathways leading to glucose and fatty acid metabolism, angio- and atherogenesis, insulin signaling, liver and kidney functions, cognition and psychological-specific phenotypes, psychiatric disorders, inflammation, and cancers.

Many of the associations noted in our study were consistent with the epidemiological literature. For example, adiponectin promotes insulin sensitivity [4], where insulin resistance is an important aspect of T2D [51], and has been shown to be inversely associated with risk of T2D [9, 10]. Likewise, our study found decreased T2D risk with serum adiponectin risk alleles.

There is a well established association between decreased HDL cholesterol levels and an increased risk of heart disease [52]. Although not as well established, there is some evidence that increased TG levels are associated with increased coronary disease risk [53]. Adiponectin has been shown to be inversely related to cardiovascular disease [10]. In agreement with this, our study found a majority of associations showing increased HDL cholesterol levels and decreased TG levels with serum adiponectin level risk alleles.

Adiponectin has been associated with anti-inflammatory properties [4, 7] and one marker of inflammation is elevation of WBC counts [54]. Consistent with this, our results found a decrease in traits categorized under WBC counts with serum adiponectin level risk alleles.

Our study also found negative associations with the use of two medications and serum adiponectin level risk alleles. The first medication was calcium channel blockers, which are used as antihypertensive drugs, and the second was HMG-CoA reductase inhibitors (i.e. statins), which are lipid lowering medications used in those at risk of cardiovascular disease. This is consistent with the known associations between elevated adiponectin levels and a decreased risk of hypertension [55], and elevated adiponectin and decreased cardiovascular disease [10]. The association between adiponectin and statins is also consistent with the associations with lipids observed in this study.

A genetic study using Mendelian randomization investigated the association between body fat distribution and adiponectin levels [37]. Specifically, they looked at abdominal (proxied by waist circumference) and gluteofemoral fat (proxied by hip circumference). They found a negative association between abdominal fat and adiponectin concentrations and a positive association between gluteofemoral fat and adiponectin concentrations [37]. Likewise, our study found a majority negative association between serum adiponectin level and waist circumference and a positive direction of effect with hip circumference. The association between waist-to-hip ratio and serum adiponectin level risk alleles was also negative.

A number of associations in our study are not well investigated. Our study found two negative associations between heel bone mineral density and serum adiponectin level risk alleles. To our knowledge, little epidemiological research has been done on this specific association, although some supporting research has been conducted on bone density. One such study found that adiponectin was negatively associated with bone mineral density in children [56] and another study found a negative association in post-menopausal women [57].

Furthermore, there is also very little research on the association between adiponectin level and the age of onset of menarche. One study did not find any significant association between the two [58]. Our study found a negative direction of effect for the association between the age of onset of menarche and serum adiponectin level risk alleles.

We also showed decreased reticulocyte count with serum adiponectin risk alleles. Epidemiologically, there is a limited amount of research on the association between adiponectin and reticulocyte count. A study conducted in patients with chronic kidney disease sought to determine whether high adiponectin levels was associated with low hemoglobin and could predict anemia in these patients. The researchers found a negative association between adiponectin level and reticulocyte production index, which is supportive of our findings [59].

Although our study focused on cross-trait associations that had two or more independent signals found, this does not mean that the associations with only one signal are unimportant. For example, a negative direction of effect between adiponectin and circulating leptin levels were found. This association reflects the biological relationship between leptin and adiponectin seen in obesity; while adiponectin decreases with an increase in adiposity, leptin has the opposite response and increases in such a scenario [1].

Our study found several associations with body characteristics that were in contrast to epidemiologic literature. A positive association was found with birthweight, which is reflected in a study by Tsai et al., which found positive associations between cord plasma adiponectin levels and birthweight [60]. However, this study also found a positive association between adiponectin and adiposity, determined through the measurement of skinfold thickness [60]. This is in contrast to the negative association found between adiponectin risk alleles, and fat mass in newborns and sum of skinfolds in newborns in our study. The positive direction of effect found between BMI and adiponectin risk alleles also contradicts the associations found in epidemiological studies, which finds negative correlations between adiponectin levels and BMI [61].

These multiple discrepancies between observational and genetic epidemiologic data in the direction of effect are not unique to our study. Other studies have also called attention to the discrepancies seen between these types of epidemiological studies. A study by Dastani et al. (2012) found a positive direction of effect, similar to our study, between adiponectin level and BMI [24]. This was posited to be due to SNPs at two loci in their study [24]. Mente et al. (2013) conducted a multi-ethnic observational study and found signification associations between adiponectin level and several metabolic traits, including HDL cholesterol, BMI, and fasting insulin level. However, the researchers only found one convincing genetic association [11]. One possible suggestion for this was that the associations between adiponectin and these traits may be influenced by confounding factors and that not all associations may

be causal [11]. As such, the discrepancies observed in our study are not fully unexpected. A possible explanation for these observations in our study may be the reduced availability of data in the literature restricting the number of associations that were found in the study, as a limited number of associations restricts the ability to make conclusions from the data [32].

There are several limitations to this study. By investigating associations and SNPs identified through GWAS studies, the methodology of this research therefore relies on the current discovery status of traits. A similar study by Kaur et al. (2018) that investigated waist-to-hip ratio (adjusted for BMI) as opposed to adiponectin yielded an initial 160 SNPs and 3675 unique pSNPs [32] compared to our study's results of 56 and 721, respectively. This difference illustrates how the extent of available data across different traits varies and may be influenced by factors such as, research conducted on the trait or disease area, the significance of findings, or the size of the studies [32].

Bias may also exist in GWAS due to many causes, including population selection. A recent paper identified how GWAS of a disease where participants would have survived other more common diseases are subject to type II errors [62], or false negatives. As our study relies on data from GWAS, any inherent bias that may influence GWAS results would be carried over to our experiment. Additionally, data in GWAS are subject to bias within the ethnic groups studied. For example, 78% of individuals included in GWAS are of European decent and 52% and 21% of studies in the Catalog were conducted in European and Asian populations, respectively [63]. This is reflected in our study as the majority of associations included European individuals and some associations were removed for ethnicity mismatch prior to analysis.

5. Conclusion

Our study sought to analyze SNPs associated with serum adiponectin level in the context of gene pleiotropy in the GWAS Catalog. Forty-six unique association signals representing 1123 cross-traits in 17 trait categories were found, with many of the uncovered cross-trait associations matching physiological mechanisms proposed in the available literature for circulating adiponectin, and others linking original phenotypes and molecular functions. Therefore, this work provides a comprehensive network of pleiotropic effects surrounding the adiponectin genetic architecture. Using integrative bioinformatics, we revisit the accumulated knowledge obtained with population genetics through association studies, reconstruct the pleiotropic networks around the adiponectin genes and highlight the pathways leading to glucose and fatty acid metabolism, angio- and atherogenesis, insulin signaling, liver and kidney functions, cognition and psychological-specific phenotypes, psychiatric disorders, inflammation and cancers.

AUTHOR'S CONTRIBUTION

D.M. designed the study; S.H., A.A.G. and D.M. conducted research; S.H. and D.M. analyzed data; S.H. performed bioinformatics, developed tools and programs; S.H. and D.M. wrote the manuscript; S.H. and D.M. designed the tables and figures; A.A.G. critically reviewed the manuscript for important intellectual

content; S.H. and D.M. had primary responsibility for final content. All authors read and approved the final manuscript.

ACKNOWLEDGMENT

We are grateful to all participants involved in this study. S.H. was supported by the following grants: APOGEE 2023 (Université de Lorraine - pôle Biologie Médecine Santé), D.M. was supported by the following grants: Chaire Gutenberg 2020.

CONFLICT OF INTEREST

No potential conflicts of interest relevant to this article were reported.

References

1. Stanley, S., et al., *Hormonal regulation of food intake.* Physiological reviews, 2005. **85**(4): p. 1131-1158.
2. Imbeault, P., *Environmental influences on adiponectin levels in humans.* Applied Physiology, Nutrition, and Metabolism, 2007. **32**(3): p. 505.
3. Meyer, L.K., et al., *Adipose tissue depot and cell size dependency of adiponectin synthesis and secretion in human obesity.* Adipocyte, 2013. **2**(4): p. 217-26.
4. Wang, Z.V. and P.E. Scherer, *Adiponectin, the past two decades.* J Mol Cell Biol, 2016. **8**(2): p. 93-100.
5. Wang, Y., et al., *Post-translational modifications of adiponectin: mechanisms and functional implications.* Biochem J, 2008. **409**(3): p. 623-33.
6. Yamauchi, T., et al., *Adiponectin receptors: a review of their structure, function and how they work.* Best Practice and Research Clinical Endocrinology and Metabolism, 2014. **28**(1): p. 15-23.
7. Goldstein, B.J., R.G. Scalia, and X.L. Ma, *Protective vascular and myocardial effects of adiponectin.* Nature Clinical Practice Cardiovascular Medicine, 2009. **6**(1): p. 27.
8. Combs, T.P., et al., *Endogenous glucose production is inhibited by the adipose-derived protein Acrp30.* J Clin Invest, 2001. **108**(12): p. 1875-81.
9. Fall, T., et al., *The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis.* PLoS medicine, 2013. **10**(6): p. e1001474.
10. Pischon, T. and E.B. Rimm, *Adiponectin: a promising marker for cardiovascular disease.* Clinical Chemistry, 2006.
11. Mente, A., et al., *Causal relationship between adiponectin and metabolic traits: a Mendelian randomization study in a multiethnic population.* PLoS one, 2013. **8**(6): p. e66808.
12. Vella, C.A., et al., *Physical Activity and Adiposity-related Inflammation: The MESA.* Med Sci Sports Exerc, 2017. **49**(5): p. 915-921.
13. Kotani, K., et al., *Adiponectin and smoking status: a systematic review.* J Atheroscler Thromb, 2012. **19**(9): p. 787-94.
14. Brien, S.E., et al., *Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies.* BMJ, 2011. **342**: p. d636.
15. Salehi-Abargouei, A., V. Izadi, and L. Azadbakht, *The effect of low calorie diet on adiponectin concentration: a systematic review and meta-analysis.* Horm Metab Res, 2015. **47**(8): p. 549-55.
16. De Vincentis, A., et al., *Effect of Sibutramine on Plasma C-Reactive Protein, Leptin and Adipon ectin Concentrations: A Systematic Review and Meta-Analysis of Randomized Contr olled Trials.* Curr Pharm Des, 2017. **23**(6): p. 870-878.
17. Butner, K.L., et al., *A review of weight loss following Roux-en-Y gastric bypass vs restrictive bariatric surgery: impact on adiponectin and insulin.* Obes Surg, 2010. **20**(5): p. 559-68.

18. Ohman-Hanson, R.A., et al., *Ethnic and Sex Differences in Adiponectin: From Childhood to Adulthood.* J Clin Endocrinol Metab, 2016. **101**(12): p. 4808-4815.
19. Kim, A.Y., et al., *Obesity-induced DNA hypermethylation of the adiponectin gene mediates insulin resistance.* Nat Commun, 2015. **6**: p. 7585.
20. Vasseur, F., D. Meyre, and P. Froguel, *Adiponectin, type 2 diabetes and the metabolic syndrome: lessons from human genetic studies.* Expert Rev Mol Med, 2006. **8**(27): p. 1-12.
21. Miljkovic-Gacic, I., et al., *Genetic determination of adiponectin and its relationship with body fat topography in multigenerational families of African heritage.* Metabolism, 2007. **56**(2): p. 234-238.
22. Butte, N.F., et al., *Genetic and environmental factors influencing fasting serum adiponectin in Hispanic children.* J Clin Endocrinol Metab, 2005. **90**(7): p. 4170-6.
23. Menzaghi, C., et al., *Circulating high molecular weight adiponectin isoform is heritable and shares a common genetic background with insulin resistance in nondiabetic White Caucasians from Italy: evidence from a family-based study.* J Intern Med, 2010. **267**(3): p. 287-94.
24. Dastani, Z., et al., *Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals.* PLoS Genet, 2012. **8**(3): p. e1002607.
25. Bowden, D.W., et al., *Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study.* Hum Mol Genet, 2010. **19**(20): p. 4112-20.
26. Bouatia-Naji, N., et al., *ACDC/adiponectin polymorphisms are associated with severe childhood and adult obesity.* Diabetes, 2006. **55**(2): p. 545-50.
27. Spracklen, C.N., et al., *Exome-Derived Adiponectin-Associated Variants Implicate Obesity and Lipid Biology.* Am J Hum Genet, 2019. **105**(1): p. 15-28.
28. Spracklen, C.N., et al., *Adiponectin GWAS loci harboring extensive allelic heterogeneity exhibit distinct molecular consequences.* PLoS Genet, 2020. **16**(9): p. e1009019.
29. Solovieff, N., et al., *Pleiotropy in complex traits: challenges and strategies.* Nature Reviews Genetics, 2013. **14**(7): p. 483.
30. Stearns, F.W., *One hundred years of pleiotropy: a retrospective.* Genetics, 2010. **186**(3): p. 767-73.
31. Kaur, Y., et al., *Comprehensive identification of pleiotropic loci for body fat distribution using the NHGRI-EBI Catalog of published genome-wide association studies.* Obes Rev, 2019. **20**(3): p. 385-406.
32. Kaur, Y., et al., *Comprehensive identification of pleiotropic loci for body fat distribution using the NHGRI-EBI Catalog of published genome-wide association studies.* Obesity Reviews, 2019. **20**(3): p. 385-406.
33. Hackinger, S. and E. Zeggini, *Statistical methods to detect pleiotropy in human complex traits.* Open Biology, 2017. **7**(11): p. 170125.
34. Gao, H., et al., *Evidence of a causal relationship between adiponectin levels and insulin sensitivity: a Mendelian randomization study.* Diabetes, 2013. **62**(4): p. 1338-1344.
35. Yaghootkar, H., et al., *Mendelian randomization studies do not support a causal role for reduced circulating adiponectin levels in insulin resistance and type 2 diabetes.* Diabetes, 2013. **62**(10): p. 3589-98.
36. Nielsen, M.B., et al., *Low Plasma Adiponectin in Risk of Type 2 Diabetes: Observational Analysis and One- and Two-Sample Mendelian Randomization Analyses in 756,219 Individuals.* Diabetes, 2021. **70**(11): p. 2694-2705.
37. Borges, M., et al., *Obesity-induced hypoadiponectinaemia: the opposite influences of central and peripheral fat compartments.* International journal of epidemiology, 2017. **46**(6): p. 2044-2055.
38. Nielsen, M.B., et al., *Causal Relationship between Plasma Adiponectin and Body Mass Index: One- and Two-Sample Bidirectional Mendelian Randomization Analyses in 460 397 Individuals.* Clin Chem, 2020. **66**(12): p. 1548-1557.
39. Dastani, Z., et al., *The shared allelic architecture of adiponectin levels and coronary artery disease.* Atherosclerosis, 2013. **229**(1): p. 145-8.

40. Borges, M.C., et al., *Role of Adiponectin in Coronary Heart Disease Risk: A Mendelian Randomization Study.* Circ Res, 2016. **119**(3): p. 491-9.
41. Chen, D., et al., *Causal associations between circulating adipokines and cardiovascular disease: A Mendelian randomization study.* J Clin Endocrinol Metab, 2022.
42. *World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.* JAMA, 2013. **310**(20): p. 2191-4.
43. MacArthur, J., et al., *The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog).* Nucleic Acids Res, 2017. **45**(D1): p. D896-D901.
44. Team, R.C., *R: A language and environment for statistical computing.* . R Found. Stat. Comput., 2014.
45. Dudbridge, F. and A. Gusnanto, *Estimation of significance thresholds for genomewide association scans.* Genet Epidemiol, 2008. **32**(3): p. 227-34.
46. Machiela, M.J. and S.J. Chanock, *LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants.* Bioinformatics, 2015. **31**(21): p. 3555-7.
47. Myers, T.A., S.J. Chanock, and M.J. Machiela, *LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations.* Front Genet, 2020. **11**: p. 157.
48. Xie, Z., et al., *Gene Set Knowledge Discovery with Enrichr.* Curr Protoc, 2021. **1**(3): p. e90.
49. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.
50. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
51. Kahn, B.B.J.C., *Type 2 diabetes: when insulin secretion fails to compensate for insulin resistance.* 1998. **92**(5): p. 593-596.
52. Goldbourt, U., et al., *Isolated low HDL cholesterol as a risk factor for coronary heart disease mortality: a 21-year follow-up of 8000 men.* 1997. **17**(1): p. 107-113.
53. Bitzur, R., et al., *Triglycerides and HDL cholesterol: stars or second leads in diabetes?* 2009. **32**(suppl 2): p. S373-S377.
54. So, W.-Y., et al., *Low plasma adiponectin level, white blood cell count and Helicobacter pylori titre independently predict abnormal pancreatic β-cell function.* 2009. **86**(2): p. 89-95.
55. Kim, D.H., et al., *Adiponectin levels and the risk of hypertension: a systematic review and meta-analysis.* Hypertension, 2013. **62**(1): p. 27-32.
56. Tubic, B., et al., *Relation between bone mineral density, biological markers and anthropometric measures in 4-year-old children: a pilot study within the IDEFICS study.* Int J Obes (Lond), 2011. **35 Suppl 1**: p. S119-24.
57. Tanna, N., et al., *The relationship between circulating adiponectin, leptin and vaspin with bone mineral density (BMD), arterial calcification and stiffness: a cross-sectional study in post-menopausal women.* J Endocrinol Invest, 2017. **40**(12): p. 1345-1353.
58. Gavela-Pérez, T., et al., *High prepubertal leptin levels are associated with earlier menarcheal age.* 2016. **59**(2): p. 177-181.
59. Kim, H., et al., *High serum adiponectin is associated with anemia development in chronic kidney disease: The results from the KNOW-CKD study.* 2018. **103**: p. 1-9.
60. Tsai, P.J., et al., *Cord plasma concentrations of adiponectin and leptin in healthy term neonates: positive correlation with birthweight and neonatal adiposity.* 2004. **61**(1): p. 88-93.
61. Arita, Y., et al., *Paradoxical decrease of an adipose-specific protein, adiponectin, in obesity.* 1999. **257**(1): p. 79-83.
62. Schooling, C.J.b., *Biases in GWAS–the dog that did not bark.* 2019: p. 709063.
63. Sirugo, G., S.M. Williams, and S.A.J.C. Tishkoff, *The missing diversity in human genetic studies.* 2019. **177**(1): p. 26-31.

DISCUSSION GÉNÉRALE DES ARTICLES

Les articles présentés en résultats sont délibérément choisis et organisés de façon à retracer la progression de mes travaux, mêmes si les dates de publication ne reflètent pas exclusivement cet agencement, principalement pour des raisons techniques de finition et de soumission dans les différents journaux. Par exemple l'article sur les microARNs dans les méningiomes est issu d'une étude entamée de longue date et qui a évolué dans le temps pour intégrer un nombre croissant d'analyses en bioinformatique et génomique, jusqu'à en faire les axes principaux de l'étude. Le travail introductif de 2020 sur le protéome de Richter est lui présenté comme une perspective d'une étude plus complète en cours de finition, multi-omique en mode protéome-centrique, et qui est planifiée pour soumission à un journal début 2024.

Cet axe chronologique de présentation des articles reflète également une progression technique, du savoir-faire professionnel, et une acquisition de compétences dans tous les domaines abordés. Evolution accompagnée d'une vision périphérique, interdisciplinaire, et d'une approche globale des problèmes. La sélection des publications comprend l'assortiment suivant :

- Quatre articles de recherche en génomique des cancers, tous acceptés dans des journaux avec évaluation par les pairs et de facteur d'impact supérieur à 5 (*Brain Pathology*, *Cancers*, *Frontiers in Oncology* et *Nature Communications*) (*87, 197-199*),
- Une communication orale longue de type *highlight* publiée dans des actes de congrès international (*JOBIM 2023*), sélectionnée après évaluation par un minimum de 2 *peer-reviewers*, et dont la présentation devrait être accessible à tous et mise en ligne prochainement par le comité d'organisation et la Société Française de Bioinformatique (SFBI, https://www.sfbi.fr/) (*200*),
- Une autre communication dans une conférence internationale avec comité de relecture, sélectionnée pour une présentation longue et qui a fait l'objet d'une publication dans un numéro spécial de *Blood (201)*,
- Un dernier article de recherche en génétique de l'obésité, non soumis mais à un stade avancé d'écriture, et qui mérite sa place dans ce manuscrit pour son importance vis-à-vis de la thématique abordée, et de son intérêt qui pourrait justifier une publication en l'état. Toutefois, au vu des résultats obtenus, nous jugeons propice d'ajouter des analyses de randomisation mendélienne à ces travaux, afin de terminer l'histoire avec des preuves mécanistiques validant notre modèle fonctionnel.

CONCLUSION ET PERSPECTIVES

Ce manuscrit de thèse présente une compilation étendue de thématiques reliées autour des maladies du métabolisme et des cancers, des approches actuelles de génomique, les « omiques » et le développement bioinformatique qui les accompagne. Son articulation cherche une cohérence dans l'évolution de mes travaux, appliqués à un vaste panel de sujets, qui finalement se fondent tous dans le projet de l'équipe « obésité, cancers et maladies inflammatoires » du prochain contrat quinquennal de notre unité de recherche. Equipe pour laquelle le rôle de la bioinformatique et des biostatistiques sera central, et pour le futur positionnement de l'unité, et pour la progression de ces thématiques localement, nationalement et internationalement, puisque de nombreuses collaborations régionales, nationales et internationales sont concernées par ces projets. Mon implication ne s'arrête pas là puisque l'équipe de bioinformatique a vocation à poursuivre son travail en participant aux projets transversaux de la communauté scientifique nancéenne, et au-delà, par le biais de la création de consortium internationaux, par exemple en génétique de l'obésité, avec un projet ambitieux et fédérateur pour élucider l'architecture polygénique de l'obésité morbide de l'adulte dans les populations européennes. En hématologie, nous poursuivons nos riches collaborations avec les équipes allemandes des Pr Siebert (*Institute of Human Genetics*) et Stilgenbauer (*Comprehensive Cancer Center Ulm* & *Division of CLL*), l'équipe espagnole du Dr Martín-Subero (*University of Barcelona*) et l'équipe américaine du Pr Wu (*Broad Institute* & *Dana-Farber Cancer Institute*) (*202-204*).

La vision générale et élargie de la biologie intégrative abordée ici donne une dimension originale aux méthodes que je mets en place car elles sont applicables à tous les domaines et apportent un souffle bienvenu à certaines approches qui pourraient être qualifiés de statiques, par exemple pour les analyses post-GWAS en génétique (augmentations trans-omiques et pléiotropiques), ou pour les annotations fonctionnelles des signatures génomiques, réputées difficiles ou trop permissives avec certains omiques ou dans le cas de données manquantes (en utilisant les réseaux d'interaction protéine / protéine multi-tissus, dont on peut voir un aperçu dans l'article sur l'adiponectine). Les travaux à venir comprennent par conséquent une partie méthodologie avec développement d'approches et d'outils bioinformatiques, et une partie applicative, en reproduisant par exemple ce que nous avons fait sur l'adiponectine à d'autres traits supposés hautement pléiotropiques et pour lesquels un certain nombre de locus GWAS sont bien décrits (par exemple : niveaux de leptine sérique, pourcentage de masse grasse). Concernant les réseaux d'interaction pléiotropiques, la mise au point de scores basés sur l'expression des gènes et des protéines dans les tissus, plutôt que l'utilisation des scores

d'association, permettrait de reconstruire des réseaux tissu-spécifiques et de s'affranchir définitivement des faux positifs dus aux types cellulaires non désirés dans les annotations fonctionnelles.

En génomique des cancers cérébraux, nous avons désormais bouclé une étude débutée il y a une dizaine d'années sur le rôle de HuR dans les méningiomes, mais poursuivons le travail sur la valeur prédictive et l'utilité clinique des marqueurs de prolifération comme Ki-67 et MCM6 (*205*). Nous complétons également les expériences sur le méthylome par le séquençage des ARNs couvrant les mêmes échantillons de méningiomes atypiques (projet TransMenAtyp), afin d'identifier sur le transcriptome des biomarqueurs pronostiques de la survie sans progression et des marqueurs prédictifs de la réponse locale à la radiothérapie, pour sélectionner des marqueurs immunohistochimiques fiables en routine clinique. L'intégration corrélative méthylome / transcriptome interviendra en premier plan dans les critères de sélection.

Enfin, en génomique des lymphomes les perspectives sont nombreuses et répondent principalement aux enjeux du projet « caractérisation génomique et protéomique du Syndrome de Richter », d'envergure nationale multi-centrique, et internationale avec les équipes de Ulm (Pr Reiner Siebert et Pr Stephan Stilgenbauer), de Barcelone (Dr Jose-Ignacio Martín-Subero) et de Boston (Pr Catherine Wu et Dr Erin Parry), pour ne citer qu'elles. Trois grandes questions restent en suspens à l'issue du travail présenté ici :

i) Nous ne disposons pas aujourd'hui de marqueurs biologiques prédictifs de la transformation. Existe-t-il dès le stade LLC des anomalies moléculaires prédictives de la transformation agressive, permettant son anticipation ? Ceci constitue la suite directe de l'article *Nature Communications*, et nous répondrons à cette question avec les mêmes cohortes et données omiques, en nous focalisant cette fois sur les paires LLC-RS afin de déterminer l'histoire évolutive propre à chaque patient (évolution clonale). Nous nous appuierons également sur les méthylomes et épigénomes couvrant tous les stades de développement des lymphocytes B sains. Ces travaux sont déjà bien avancés.

ii) Quelles sont les anomalies non présentes au stade LLC, qui sont donc des anomalies additionnelles qui accompagnent la transformation vers le stade Richter, et qui déclenchent la bascule phénotypique d'une maladie d'évolution lente vers une maladie agressive, létale en quelques mois ? Autrement dit, il s'agit d'élucider les mécanismes et les déterminants de l'évolution agressive, ce que nous proposons de faire avec l'étude de l'évolution clonale, et l'étude du

miRnome selon un design expérimental assez proche de celui présenté précédemment (un des axes du travail de thèse de Romain Piucco).

iii) Au stade SR, quels sont les mécanismes oncogéniques à l'origine de l'inefficacité de l'immunochimiothérapie et des traitements ciblés ? Il s'agit ici de disséquer la physiopathologie du SR pour identifier de nouvelles cibles thérapeutiques, ce que nous sommes en train de faire avec l'étude multi-omique axée sur le protéome de SR, dont une communication orale est présentée dans ce document.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. J. A. Timmons *et al.*, A coding and non-coding transcriptomic perspective on the genomics of human metabolic disease. *Nucleic Acids Res* **46**, 7772-7792 (2018).
2. S. O'Rahilly, Human genetics illuminates the paths to metabolic disease. *Nature* **462**, 307-314 (2009).
3. J. Nielsen, Systems Biology of Metabolism: A Driver for Developing Personalized and Precision Medicine. *Cell Metab* **25**, 572-579 (2017).
4. J. Rendleman, H. Choi, C. Vogel, Integration of large-scale multi-omic datasets: a protein-centric view. *Curr Opin Syst Biol* **11**, 74-81 (2018).
5. J. Flannick, J. C. Florez, Type 2 diabetes: genetic data sharing to advance complex disease research. *Nat Rev Genet* **17**, 535-549 (2016).
6. N. Perakakis, A. Yazdani, G. E. Karniadakis, C. Mantzoros, Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. *Metabolism* **87**, A1-A9 (2018).
7. Y. Hasin, M. Seldin, A. Lusis, Multi-omics approaches to disease. *Genome Biol* **18**, 83 (2017).
8. G. Tini, L. Marchetti, C. Priami, M. P. Scott-Boyer, Multi-omics integration-a comparison of unsupervised clustering methodologies. *Brief Bioinform* **20**, 1269-1279 (2019).
9. K. Yugi, H. Kubota, A. Hatano, S. Kuroda, Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple 'Omic' Layers. *Trends Biotechnol* **34**, 276-290 (2016).
10. D. Hanahan, Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31-46 (2022).
11. J. Clavel, Progress in the epidemiological understanding of gene-environment interactions in major diseases: cancer. *C R Biol* **330**, 306-317 (2007).
12. P. Anand *et al.*, Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res* **25**, 2097-2116 (2008).
13. H. Sung *et al.*, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249 (2021).
14. H. K. Weir, T. D. Thompson, S. L. Stewart, M. C. White, Cancer Incidence Projections in the United States Between 2015 and 2050. *Prev Chronic Dis* **18**, E59 (2021).
15. A. Belot *et al.*, Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique* **56**, 159-175 (2008).
16. J. I. Martin-Subero, C. C. Oakes, Charting the dynamic epigenome during B-cell development. *Semin Cancer Biol* **51**, 139-148 (2018).
17. C. Stirzaker, P. C. Taberlay, A. L. Statham, S. J. Clark, Mining cancer methylomes: prospects and challenges. *Trends Genet* **30**, 75-84 (2014).
18. W. W. Liang *et al.*, Integrative multi-omic cancer profiling reveals DNA methylation patterns associated with therapeutic vulnerability and cell-of-origin. *Cancer Cell* **41**, 1567-1585.e1567 (2023).
19. C. C. Oakes, J. I. Martin-Subero, Insight into origins, mechanisms, and utility of DNA methylation in B-cell malignancies. *Blood* **132**, 999-1006 (2018).
20. R. Küppers, Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* **5**, 251-262 (2005).
21. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer Statistics, 2021. *CA Cancer J Clin* **71**, 7-33 (2021).
22. M. Hallek *et al.*, iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood* **131**, 2745-2760 (2018).
23. T. J. Kipps *et al.*, Chronic lymphocytic leukaemia. *Nat Rev Dis Primers* **3**, 16096 (2017).
24. H. Döhner *et al.*, Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* **343**, 1910-1916 (2000).
25. G. Lazarian, R. Guièze, C. J. Wu, Clinical Implications of Novel Genomic Discoveries in Chronic Lymphocytic Leukemia. *J Clin Oncol* **35**, 984-993 (2017).
26. M. Hallek *et al.*, Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* **376**, 1164-1174 (2010).

27. R. R. Furman *et al.*, Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N Engl J Med* **370**, 997-1007 (2014).
28. T. Robak, S. Stilgenbauer, A. Tedeschi, Front-line treatment of CLL in the era of novel agents. *Cancer Treat Rev* **53**, 70-78 (2017).
29. S. Stilgenbauer *et al.*, Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study. *Lancet Oncol* **17**, 768-778 (2016).
30. A. A. Alizadeh *et al.*, Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511 (2000).
31. M. Roschewski, L. M. Staudt, W. H. Wilson, Diffuse large B-cell lymphoma-treatment approaches in the molecular era. *Nat Rev Clin Oncol* **11**, 12-23 (2014).
32. L. H. Sehn, R. D. Gascoyne, Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood* **125**, 22-32 (2015).
33. A. Reddy *et al.*, Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* **171**, 481-494.e415 (2017).
34. R. Schmitz *et al.*, Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N Engl J Med* **378**, 1396-1407 (2018).
35. B. Chapuy *et al.*, Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* **24**, 679-690 (2018).
36. G. W. Wright *et al.*, A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. *Cancer Cell* **37**, 551-568.e514 (2020).
37. B. Coiffier *et al.*, CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med* **346**, 235-242 (2002).
38. C. Gisselbrecht *et al.*, Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol* **28**, 4184-4190 (2010).
39. D. Rossi *et al.*, Molecular prediction of durable remission after first-line fludarabine-cyclophosphamide-rituximab in chronic lymphocytic leukemia. *Blood* **126**, 1921-1924 (2015).
40. D. Rossi, V. Spina, G. Gaidano, Biology and treatment of Richter syndrome. *Blood* **131**, 2761-2772 (2018).
41. D. Rossi *et al.*, The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation. *Blood* **117**, 3391-3401 (2011).
42. S. A. Parikh, N. E. Kay, T. D. Shanafelt, How we treat Richter syndrome. *Blood* **123**, 1647-1657 (2014).
43. E. Chigrinova *et al.*, Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. *Blood* **122**, 2673-2682 (2013).
44. G. Fabbri *et al.*, Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. *J Exp Med* **210**, 2273-2288 (2013).
45. J. Klintman *et al.*, Genomic and transcriptomic correlates of Richter transformation in chronic lymphocytic leukemia. *Blood* **137**, 2800-2816 (2021).
46. A. Rinaldi *et al.*, Promoter methylation patterns in Richter syndrome affect stem-cell maintenance and cell cycle regulation and differ from de novo diffuse large B-cell lymphoma. *Br J Haematol* **163**, 194-204 (2013).
47. H. Augé *et al.*, Microenvironment Remodeling and Subsequent Clinical Implications in Diffuse Large B-Cell Histologic Variant of Richter Syndrome. *Front Immunol* **11**, 594841 (2020).
48. F. Nadeu *et al.*, Detection of early seeding of Richter transformation in chronic lymphocytic leukemia. *Nat Med* **28**, 1662-1671 (2022).
49. D. N. Louis *et al.*, The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* **23**, 1231-1251 (2021).
50. D. Capper *et al.*, DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469-474 (2018).
51. A. R. Fathi, U. Roelcke, Meningioma. *Curr Neurol Neurosci Rep* **13**, 337 (2013).
52. Q. T. Ostrom *et al.*, CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007-2011. *Neuro Oncol* **16 Suppl 4**, iv1-63 (2014).
53. P. François *et al.*, Post-traumatic meningioma: three case reports of this rare condition and a review of the literature. *Acta Neurochir (Wien)* **152**, 1755-1760 (2010).
54. K. Gousias, J. Schramm, M. Simon, The Simpson grading revisited: aggressive surgery and its place in modern meningioma management. *J Neurosurg* **125**, 551-560 (2016).

55. W. L. Bi *et al.*, Genomic landscape of intracranial meningiomas. *J Neurosurg* **125**, 525-535 (2016).
56. N. Wang, M. Osswald, Meningiomas: Overview and New Directions in Therapy. *Semin Neurol* **38**, 112-120 (2018).
57. H. van Alkemade *et al.*, Impaired survival and long-term neurological problems in benign meningioma. *Neuro Oncol* **14**, 658-666 (2012).
58. A. A. Aizer *et al.*, Extent of resection and overall survival for patients with atypical and malignant meningioma. *Cancer* **121**, 4376-4381 (2015).
59. D. N. Louis *et al.*, The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* **131**, 803-820 (2016).
60. G. A. Rouleau *et al.*, Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature* **363**, 515-521 (1993).
61. V. E. Clark *et al.*, Recurrent somatic mutations in POLR2A define a distinct subset of meningiomas. *Nat Genet* **48**, 1253-1259 (2016).
62. W. L. Bi *et al.*, Genomic landscape of high-grade meningiomas. *NPJ Genom Med* **2**, (2017).
63. R. A. Buerki *et al.*, An overview of meningiomas. *Future Oncol* **14**, 2161-2177 (2018).
64. W. L. Bi, M. Zhang, W. W. Wu, Y. Mei, I. F. Dunn, Meningioma Genomics: Diagnostic, Prognostic, and Therapeutic Applications. *Front Surg* **3**, 40 (2016).
65. S. Suppiah *et al.*, Molecular and translational advances in meningiomas. *Neuro Oncol* **21**, i4-i17 (2019).
66. F. Sahm *et al.*, DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. *Lancet Oncol* **18**, 682-694 (2017).
67. F. Nassiri *et al.*, A clinically applicable integrative molecular classification of meningiomas. *Nature* **597**, 119-125 (2021).
68. N. A. Paleologos, H. B. Newton, *Oligodendroglioma: Clinical Presentation, Pathology, Molecular Biology, Imaging, and Treatment*. (Elsevier Science, 2019).
69. Q. T. Ostrom *et al.*, CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016. *Neuro Oncol* **21**, v1-v100 (2019).
70. C. Hartmann *et al.*, Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas. *Acta Neuropathol* **118**, 469-474 (2009).
71. F. Sahm *et al.*, Farewell to oligoastrocytoma: in situ molecular genetics favor classification as either oligodendroglioma or astrocytoma. *Acta Neuropathol* **128**, 551-559 (2014).
72. D. J. Brat *et al.*, Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* **372**, 2481-2498 (2015).
73. H. Suzuki *et al.*, Mutational landscape and clonal architecture in grade II and III gliomas. *Nat Genet* **47**, 458-468 (2015).
74. Q. T. Ostrom *et al.*, CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro Oncol* **19**, v1-v88 (2017).
75. M. J. van den Bent, Thirty years of progress in the management of low-grade gliomas. *Rev Neurol (Paris)* **179**, 425-429 (2023).
76. G. Simonetti *et al.*, Clinical management of grade III oligodendroglioma. *Cancer Manag Res* **7**, 213-223 (2015).
77. J. Lee *et al.*, Oligodendrogliomas, IDH-mutant and 1p/19q-codeleted, arising during teenage years often lack TERT promoter mutation that is typical of their adult counterparts. *Acta Neuropathol Commun* **6**, 95 (2018).
78. J. Buckner *et al.*, Management of diffuse low-grade gliomas in adults - use of molecular diagnostics. *Nat Rev Neurol* **13**, 340-351 (2017).
79. M. J. van den Bent *et al.*, Adjuvant procarbazine, lomustine, and vincristine chemotherapy in newly diagnosed anaplastic oligodendroglioma: long-term follow-up of EORTC brain tumor group study 26951. *J Clin Oncol* **31**, 344-350 (2013).
80. D. Figarella-Branger *et al.*, Mitotic index, microvascular proliferation, and necrosis define 3 groups of 1p/19q codeleted anaplastic oligodendrogliomas associated with different genomic alterations. *Neuro Oncol* **16**, 1244-1254 (2014).
81. C. Bettegowda *et al.*, Mutations in CIC and FUBP1 contribute to human oligodendroglioma. *Science* **333**, 1453-1455 (2011).

82. V. Gleize *et al.*, CIC inactivating mutations identify aggressive subset of 1p19q codeleted gliomas. *Ann Neurol* **78**, 355-374 (2015).
83. K. Labreche *et al.*, TCF12 is mutated in anaplastic oligodendroglioma. *Nat Commun* **6**, 7207 (2015).
84. A. Alentorn *et al.*, Allelic loss of 9p21.3 is a prognostic factor in 1p/19q codeleted anaplastic gliomas. *Neurology* **85**, 1325-1331 (2015).
85. A. Kamoun *et al.*, Integrated multi-omics analysis of oligodendroglial tumours identifies three subgroups of 1p/19q co-deleted gliomas. *Nat Commun* **7**, 11263 (2016).
86. S. Rosenberg *et al.*, Machine Learning for Better Prognostic Stratification and Driver Gene Identification Using Somatic Copy Number Variations in Anaplastic Oligodendroglioma. *Oncologist* **23**, 1500-1510 (2018).
87. C. Pouget *et al.*, Ki-67 and MCM6 labeling indices are correlated with overall survival in anaplastic oligodendroglioma, IDH1-mutant and 1p/19q-codeleted: a multicenter study from the French POLA network. *Brain Pathol* **30**, 465-478 (2020).
88. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet* **390**, 2627-2642 (2017).
89. D. P. Guh *et al.*, The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health* **9**, 88 (2009).
90. K. R. Fontaine, D. T. Redden, C. Wang, A. O. Westfall, D. B. Allison, Years of life lost due to obesity. *JAMA* **289**, 187-193 (2003).
91. K. E. Friedman *et al.*, Weight stigmatization and ideological beliefs: relation to psychological functioning in obese adults. *Obes Res* **13**, 907-916 (2005).
92. D. Withrow, D. A. Alter, The economic burden of obesity worldwide: a systematic review of the direct costs of obesity. *Obes Rev* **12**, 131-141 (2011).
93. H. Reddon, J. L. Gueant, D. Meyre, The importance of gene-environment interactions in human obesity. *Clin Sci (Lond)* **130**, 1571-1597 (2016).
94. A. Li, D. Meyre, Challenges in reproducibility of genetic association studies: lessons learned from the obesity field. *Int J Obes (Lond)* **37**, 559-567 (2013).
95. M. J. Muller, A. Bosy-Westphal, M. Krawczak, Genetic studies of common types of obesity: a critique of the current use of phenotypes. *Obes Rev* **11**, 612-618 (2010).
96. A. M. Gradmark *et al.*, Computed tomography-based validation of abdominal adiposity measurements from ultrasonography, dual-energy X-ray absorptiometry and anthropometry. *Br J Nutr* **104**, 582-588 (2010).
97. M. Pigeyre, F. T. Yazdi, Y. Kaur, D. Meyre, Recent progress in genetics, epigenetics and metagenomics unveils the pathophysiology of human obesity. *Clin Sci (Lond)* **130**, 943-986 (2016).
98. B. Patro *et al.*, Maternal and paternal body mass index and offspring obesity: a systematic review. *Ann Nutr Metab* **63**, 32-41 (2013).
99. C. Stryjecki, A. Alyass, D. Meyre, Ethnic and population differences in the genetic predisposition to human obesity. *Obes Rev* **19**, 62-80 (2018).
100. J. Yang *et al.*, Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* **47**, 1114-1120 (2015).
101. I. O. Lunsky, D. Meyre, Decoding Mendelian obesity. *Curr Opin Endocr Metab Res* **4**, 21-28 (2019).
102. L. Yengo *et al.*, Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649 (2018).
103. T. M. Frayling *et al.*, A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-894 (2007).
104. A. Scuteri *et al.*, Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* **3**, e115 (2007).
105. F. K. Ndiaye *et al.*, The expression of genes in top obesity-associated loci is enriched in insula and substantia nigra brain regions involved in addiction and reward. *Int J Obes (Lond)* **44**, 539-543 (2020).
106. A. E. Locke *et al.*, Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).

107. R. J. F. Loos, G. S. H. Yeo, The genetics of obesity: from discovery to biology. *Nat Rev Genet* **23**, 120-133 (2022).
108. J. E. Kim *et al.*, The Roles and Associated Mechanisms of Adipokines in Development of Metabolic Syndrome. *Molecules* **27**, (2022).
109. R. Coppari, C. Bjørbæk, Leptin revisited: its mechanism of action and potential for treating diabetes. *Nat Rev Drug Discov* **11**, 692-708 (2012).
110. O. Gualillo, Further evidence for leptin involvement in cartilage homeostases. *Osteoarthritis Cartilage* **15**, 857-860 (2007).
111. N. Ouchi, J. L. Parker, J. J. Lugus, K. Walsh, Adipokines in inflammation and metabolic disease. *Nat Rev Immunol* **11**, 85-97 (2011).
112. M. Scotece *et al.*, Adipokines as drug targets in joint and bone disease. *Drug Discov Today* **19**, 241-258 (2014).
113. C. Cuerq *et al.*, [Mutltifaceted biological roles of adiponectin]. *Ann Biol Clin (Paris)* **78**, 243-252 (2020).
114. H. M. Choi, H. M. Doss, K. S. Kim, Multifaceted Physiological Roles of Adiponectin in Inflammation and Diseases. *Int J Mol Sci* **21**, (2020).
115. S. Stanley, K. Wynne, B. McGowan, S. Bloom, Hormonal regulation of food intake. *Physiological reviews* **85**, 1131-1158 (2005).
116. P. Imbeault, Environmental influences on adiponectin levels in humans. *Appl Physiol Nutr Metab* **32**, 505-511 (2007).
117. L. K. Meyer, T. P. Ciaraldi, R. R. Henry, A. C. Wittgrove, S. A. Phillips, Adipose tissue depot and cell size dependency of adiponectin synthesis and secretion in human obesity. *Adipocyte* **2**, 217-226 (2013).
118. Z. V. Wang, P. E. Scherer, Adiponectin, the past two decades. *J Mol Cell Biol* **8**, 93-100 (2016).
119. Y. Wang, K. S. Lam, M. H. Yau, A. Xu, Post-translational modifications of adiponectin: mechanisms and functional implications. *Biochem J* **409**, 623-633 (2008).
120. C. A. Vella *et al.*, Physical Activity and Adiposity-related Inflammation: The MESA. *Med Sci Sports Exerc* **49**, 915-921 (2017).
121. K. Kotani *et al.*, Adiponectin and smoking status: a systematic review. *J Atheroscler Thromb* **19**, 787-794 (2012).
122. S. E. Brien, P. E. Ronksley, B. J. Turner, K. J. Mukamal, W. A. Ghali, Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *BMJ* **342**, d636 (2011).
123. A. Salehi-Abargouei, V. Izadi, L. Azadbakht, The effect of low calorie diet on adiponectin concentration: a systematic review and meta-analysis. *Horm Metab Res* **47**, 549-555 (2015).
124. A. De Vincentis *et al.*, Effect of Sibutramine on Plasma C-Reactive Protein, Leptin and Adipon ectin Concentrations: A Systematic Review and Meta-Analysis of Randomized Contr olled Trials. *Curr Pharm Des* **23**, 870-878 (2017).
125. K. L. Butner, S. M. Nickols-Richardson, S. F. Clark, W. K. Ramp, W. G. Herbert, A review of weight loss following Roux-en-Y gastric bypass vs restrictive bariatric surgery: impact on adiponectin and insulin. *Obes Surg* **20**, 559-568 (2010).
126. R. A. Ohman-Hanson *et al.*, Ethnic and Sex Differences in Adiponectin: From Childhood to Adulthood. *J Clin Endocrinol Metab* **101**, 4808-4815 (2016).
127. A. Y. Kim *et al.*, Obesity-induced DNA hypermethylation of the adiponectin gene mediates insulin resistance. *Nat Commun* **6**, 7585 (2015).
128. F. Vasseur, D. Meyre, P. Froguel, Adiponectin, type 2 diabetes and the metabolic syndrome: lessons from human genetic studies. *Expert Rev Mol Med* **8**, 1-12 (2006).
129. I. Miljkovic-Gacic *et al.*, Genetic determination of adiponectin and its relationship with body fat topography in multigenerational families of African heritage. *Metabolism* **56**, 234-238 (2007).
130. N. F. Butte *et al.*, Genetic and environmental factors influencing fasting serum adiponectin in Hispanic children. *J Clin Endocrinol Metab* **90**, 4170-4176 (2005).
131. C. Menzaghi *et al.*, Circulating high molecular weight adiponectin isoform is heritable and shares a common genetic background with insulin resistance in nondiabetic White Caucasians from Italy: evidence from a family-based study. *J Intern Med* **267**, 287-294 (2010).

132. Z. Dastani *et al.*, Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* **8**, e1002607 (2012).
133. D. W. Bowden *et al.*, Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. *Hum Mol Genet* **19**, 4112-4120 (2010).
134. N. Bouatia-Naji *et al.*, ACDC/adiponectin polymorphisms are associated with severe childhood and adult obesity. *Diabetes* **55**, 545-550 (2006).
135. C. N. Spracklen *et al.*, Exome-Derived Adiponectin-Associated Variants Implicate Obesity and Lipid Biology. *Am J Hum Genet* **105**, 15-28 (2019).
136. C. N. Spracklen *et al.*, Adiponectin GWAS loci harboring extensive allelic heterogeneity exhibit distinct molecular consequences. *PLoS Genet* **16**, e1009019 (2020).
137. T. Yamauchi, M. Iwabu, M. Okada-Iwabu, T. Kadowaki, Adiponectin receptors: a review of their structure, function and how they work. *Best Practice and Research Clinical Endocrinology and Metabolism* **28**, 15-23 (2014).
138. N. Kubota *et al.*, Adiponectin stimulates AMP-activated protein kinase in the hypothalamus and increases food intake. *Cell Metab* **6**, 55-68 (2007).
139. B. J. Goldstein, R. G. Scalia, X. L. Ma, Protective vascular and myocardial effects of adiponectin. *Nature Clinical Practice Cardiovascular Medicine* **6**, 27 (2009).
140. T. P. Combs, A. H. Berg, S. Obici, P. E. Scherer, L. Rossetti, Endogenous glucose production is inhibited by the adipose-derived protein Acrp30. *J Clin Invest* **108**, 1875-1881 (2001).
141. T. Fall *et al.*, The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis. *PLoS medicine* **10**, e1001474 (2013).
142. T. Pischon, E. B. Rimm, Adiponectin: a promising marker for cardiovascular disease. *Clinical Chemistry*, (2006).
143. A. Mente *et al.*, Causal relationship between adiponectin and metabolic traits: a Mendelian randomization study in a multiethnic population. *PLoS one* **8**, e66808 (2013).
144. S. N. Bhupathiraju, F. B. Hu, Epidemiology of Obesity and Diabetes and Their Cardiovascular Complications. *Circ Res* **118**, 1723-1735 (2016).
145. P. M. McKeigue, B. Shah, M. G. Marmot, Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet* **337**, 382-386 (1991).
146. E. R. F. Collaboration, Life expectancy associated with different ages at diagnosis of type 2 diabetes in high-income countries: 23 million person-years of observation. *Lancet Diabetes Endocrinol* **11**, 731-742 (2023).
147. M. Kasuga, Insulin resistance and pancreatic beta cell failure. *J Clin Invest* **116**, 1756-1760 (2006).
148. A. S. Al-Goblan, M. A. Al-Alfi, M. Z. Khan, Mechanism linking diabetes mellitus and obesity. *Diabetes Metab Syndr Obes* **7**, 587-591 (2014).
149. M. E. Røder, D. Porte, R. S. Schwartz, S. E. Kahn, Disproportionately elevated proinsulin levels reflect the degree of impaired B cell secretory capacity in patients with noninsulin-dependent diabetes mellitus. *J Clin Endocrinol Metab* **83**, 604-608 (1998).
150. A. G. Ershow, Environmental influences on development of type 2 diabetes and obesity: challenges in personalizing prevention and management. *J Diabetes Sci Technol* **3**, 727-734 (2009).
151. J. Maddatu, E. Anderson-Baucum, C. Evans-Molina, Smoking and the risk of type 2 diabetes. *Transl Res* **184**, 101-107 (2017).
152. J. P. Després, Intra-abdominal obesity: an untreated risk factor for Type 2 diabetes and cardiovascular disease. *J Endocrinol Invest* **29**, 77-82 (2006).
153. S. M. Haffner, M. P. Stern, H. P. Hazuda, B. D. Mitchell, J. K. Patterson, Cardiovascular risk factors in confirmed prediabetic individuals. Does the clock for coronary heart disease start ticking before the onset of clinical diabetes? *JAMA* **263**, 2893-2898 (1990).
154. K. M. Narayan, J. P. Boyle, T. J. Thompson, E. W. Gregg, D. F. Williamson, Effect of BMI on lifetime risk for diabetes in the U.S. *Diabetes Care* **30**, 1562-1566 (2007).
155. W. C. Knowler *et al.*, Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* **346**, 393-403 (2002).
156. H. Booth *et al.*, Incidence of type 2 diabetes after bariatric surgery: population-based matched cohort study. *Lancet Diabetes Endocrinol* **2**, 963-968 (2014).

157. J. R. Harkess *et al.*, Human ehrlichiosis in Oklahoma. *J Infect Dis* **159**, 576-579 (1989).
158. A. Xue *et al.*, Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* **9**, 2941 (2018).
159. M. Tschöp, D. L. Smiley, M. L. Heiman, Ghrelin induces adiposity in rodents. *Nature* **407**, 908-913 (2000).
160. M. T. Bluet-Pajot *et al.*, [Ghrelin: a striking example of neuroendocrine peptide pleiotropy]. *Med Sci (Paris)* **21**, 715-721 (2005).
161. M. J. Iglesias *et al.*, Growth hormone releasing peptide (ghrelin) is synthesized and secreted by cardiomyocytes. *Cardiovasc Res* **62**, 481-488 (2004).
162. A. Elizondo *et al.*, Polyunsaturated fatty acid pattern in liver and erythrocyte phospholipids from obese patients. *Obesity (Silver Spring)* **15**, 24-31 (2007).
163. F. Massiera *et al.*, A Western-like fat diet is sufficient to induce a gradual enhancement in fat mass over generations. *J Lipid Res* **51**, 2352-2361 (2010).
164. R. Stienstra, C. Duval, M. Müller, S. Kersten, PPARs, Obesity, and Inflammation. *PPAR Res* **2007**, 95974 (2007).
165. G. S. Hotamisligil, E. Erbay, Nutrient sensing and inflammation in metabolic diseases. *Nat Rev Immunol* **8**, 923-934 (2008).
166. B. H. L. Harris *et al.*, Obesity: a perfect storm for carcinogenesis. *Cancer Metastasis Rev* **41**, 491-515 (2022).
167. K. E. Wellen, G. S. Hotamisligil, Inflammation, stress, and diabetes. *J Clin Invest* **115**, 1111-1119 (2005).
168. D. M. Mannino *et al.*, Boys with high body masses have an increased risk of developing asthma: findings from the National Longitudinal Survey of Youth (NLSY). *Int J Obes (Lond)* **30**, 6-13 (2006).
169. B. I. Goldstein, D. E. Kemp, J. K. Soczynska, R. S. McIntyre, Inflammation and the phenomenology, pathophysiology, comorbidity, and treatment of bipolar disorder: a systematic review of the literature. *J Clin Psychiatry* **70**, 1078-1090 (2009).
170. E. E. Calle, R. Kaaks, Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer* **4**, 579-591 (2004).
171. H. Carter, M. Hofree, T. Ideker, Genotype to phenotype via network analysis. *Curr Opin Genet Dev* **23**, 611-621 (2013).
172. O. Vanunu, O. Magger, E. Ruppin, T. Shlomi, R. Sharan, Associating genes and protein complexes with disease via network propagation. *PLoS Comput Biol* **6**, e1000641 (2010).
173. H. Fang *et al.*, A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat Genet* **51**, 1082-1091 (2019).
174. J. K. Huang *et al.*, Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Syst* **6**, 484-495.e485 (2018).
175. I. Barrio-Hernandez *et al.*, Network expansion of genetic associations defines a pleiotropy map of human cell biology. *Nat Genet* **55**, 389-398 (2023).
176. E. Kerschbaum, V. Nüssler, Cancer Prevention with Nutrition and Lifestyle. *Visc Med* **35**, 204-209 (2019).
177. S. J. Nomura *et al.*, Adherence to diet, physical activity and body weight recommendations and breast cancer incidence in the Black Women's Health Study. *Int J Cancer* **139**, 2738-2752 (2016).
178. B. Lauby-Secretan, L. Dossus, C. Marant-Micallef, M. His, [Obesity and Cancer]. *Bull Cancer* **106**, 635-646 (2019).
179. S. K. Clinton, E. L. Giovannucci, S. D. Hursting, The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. *J Nutr* **150**, 663-671 (2020).
180. P. Bhardwaj *et al.*, Estrogens and breast cancer: Mechanisms involved in obesity-related development, growth and progression. *J Steroid Biochem Mol Biol* **189**, 161-170 (2019).
181. R. Dobson, in *Bmj*. (Copyright © 2005, BMJ Publishing Group Ltd., 2005), vol. 331, pp. 368.
182. F. Bianchini, R. Kaaks, H. Vainio, Overweight, obesity, and cancer risk. *Lancet Oncol* **3**, 565-574 (2002).
183. M. A. Lichtman, Obesity and the risk for a hematological malignancy: leukemia, lymphoma, or myeloma. *Oncologist* **15**, 1083-1101 (2010).
184. K. I. Avgerinos, N. Spyrou, C. S. Mantzoros, M. Dalamaga, Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism* **92**, 121-135 (2019).

185. E. K. Amankwah, A. M. Saenz, G. A. Hale, P. A. Brown, Association between body mass index at diagnosis and pediatric leukemia mortality and relapse: a systematic review and meta-analysis. *Leuk Lymphoma* **57**, 1140-1148 (2016).
186. B. L. Ecker *et al.*, Impact of obesity on breast cancer recurrence and minimal residual disease. *Breast Cancer Res* **21**, 41 (2019).
187. F. Petrelli *et al.*, Association of Obesity With Survival Outcomes in Patients With Cancer: A Systematic Review and Meta-analysis. *JAMA Netw Open* **4**, e213520 (2021).
188. E. H. Allott, S. D. Hursting, Obesity and cancer: mechanistic insights from transdisciplinary studies. *Endocr Relat Cancer* **22**, R365-386 (2015).
189. X. Fang *et al.*, Causal association of childhood obesity with cancer risk in adulthood: A Mendelian randomization study. *Int J Cancer* **149**, 1421-1425 (2021).
190. L. Nogueira, R. Stolzenberg-Solomon, M. Gamborg, T. I. A. Sørensen, J. L. Baker, Childhood body mass index and risk of adult pancreatic cancer. *Curr Dev Nutr* **1**, (2017).
191. L. Zohar *et al.*, Adolescent overweight and obesity and the risk for pancreatic cancer among men and women: a nationwide study of 1.79 million Israeli adolescents. *Cancer* **125**, 118-126 (2019).
192. J. Vendrell *et al.*, Resistin, adiponectin, ghrelin, leptin, and proinflammatory cytokines: relationships in obesity. *Obes Res* **12**, 962-971 (2004).
193. M. F. Gregor, G. S. Hotamisligil, Inflammatory mechanisms in obesity. *Annu Rev Immunol* **29**, 415-445 (2011).
194. G. Gauchotte *et al.*, Cytoplasmic overexpression of RNA-binding protein HuR is a marker of poor prognosis in meningioma, and HuR knockdown decreases meningioma cell growth and resistance to hypoxia. *J Pathol* **242**, 421-434 (2017).
195. Z. Shen *et al.*, MicroRNA-519a demonstrates significant tumour suppressive activity in laryngeal squamous cells by targeting anti-carcinoma HuR gene. *The Journal of laryngology and otology* **127**, 1194-1202 (2013).
196. L. Ren *et al.*, miR-519 regulates the proliferation of breast cancer cells via targeting human antigen R. *Oncol Lett* **19**, 1567-1576 (2020).
197. J. Broséus *et al.*, Molecular characterization of Richter syndrome identifies de novo diffuse large B-cell lymphomas with poor prognosis. *Nat Commun* **14**, 309 (2023).
198. S. Hergalant *et al.*, MicroRNAs miR-16 and miR-519 control meningioma cell proliferation. *Front Oncol* **13**, 1158773 (2023).
199. S. Hergalant *et al.*, Correlation between DNA Methylation and Cell Proliferation Identifies New Candidate Predictive Markers in Meningioma. *Cancers (Basel)* **14**, (2022).
200. S. Hergalant *et al.*, paper presented at the JOBIM2023, Plouzané (Brest), France, 2023-06-27 2023.
201. R. Morizot *et al.*, Large-Scale Proteomics Identifies Distinct Signatures for Richter Syndrome and De Novo Diffuse Large B-Cell Lymphoma: A French Study from the Filo Group. *Blood* **136**, 29-30 (2020).
202. E. M. Parry *et al.*, ZNF683 marks a CD8. *Cancer Cell*, (2023).
203. E. M. Parry *et al.*, Evolutionary history of transformation from chronic lymphocytic leukemia to Richter syndrome. *Nat Med* **29**, 158-169 (2023).
204. E. Ten Hacken *et al.*, In Vivo Modeling of CLL Transformation to Richter Syndrome Reveals Convergent Evolutionary Paths and Therapeutic Vulnerabilities. *Blood Cancer Discov* **4**, 150-169 (2023).
205. G. Gauchotte *et al.*, A High MCM6 Proliferative Index in Atypical Meningioma Is Associated with Shorter Progression Free and Overall Survivals. *Cancers (Basel)* **15**, (2023).

ANNEXES

Annexe 1. Tables et figures supplémentaires de l'article « *Ki-67 and MCM6 labeling indices are correlated with overall survival in anaplastic oligodendroglioma, IDH1-mutant and 1p/19q-codeleted: a multicenter study from the French POLA network* »

Tables supplémentaires

Table S1. https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fbpa.12788&file=bpa12788-sup-0009-TableS1.xlsx
Table S2. https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fbpa.12788&file=bpa12788-sup-0010-TableS2.xlsx
Table S3. https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fbpa.12788&file=bpa12788-sup-0011-TableS3.xlsx
Table S4. https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fbpa.12788&file=bpa12788-sup-0012-TableS4.xlsx

Figures supplémentaires

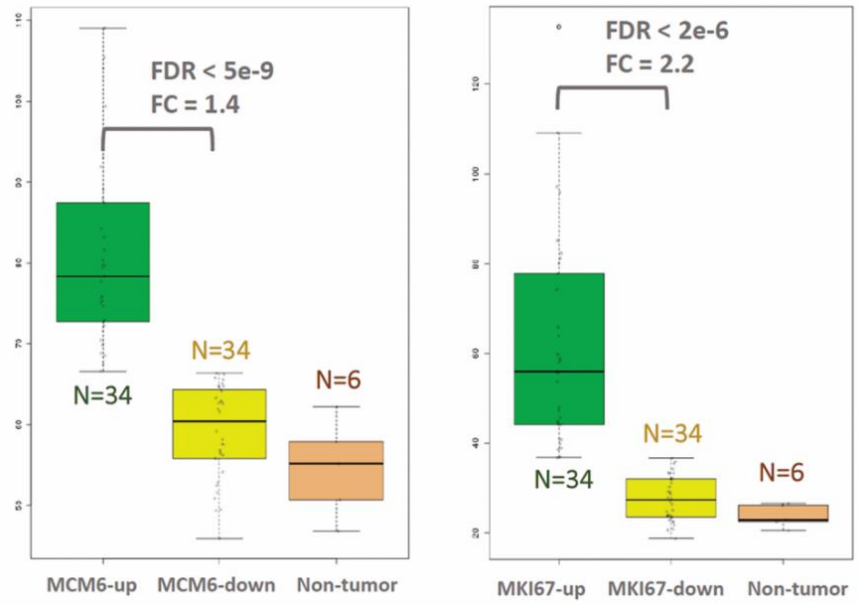


Figure S1. *MCM6* and *MKI67* mRNA level distributions into two subgroups of higher ("-up") and lower ("-down") expressions. Transcriptomic samples (n = 68) are separated at the median expression for each gene and ordered according to their expression gradient for either *MCM6* or *MKI67*.

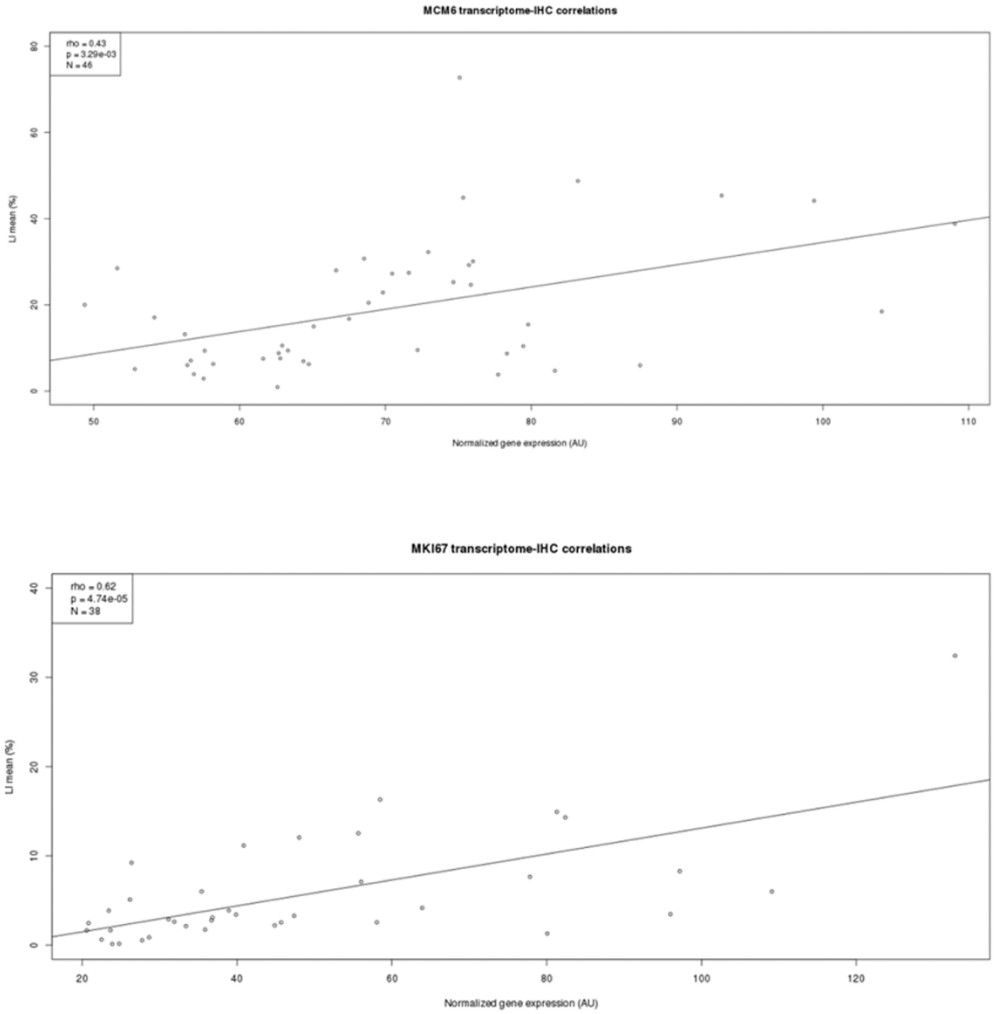


Figure S2. Correlations between immunohistochemistry (IHC) and mRNA level for Ki-67 (*MKI67*) and MCM6.

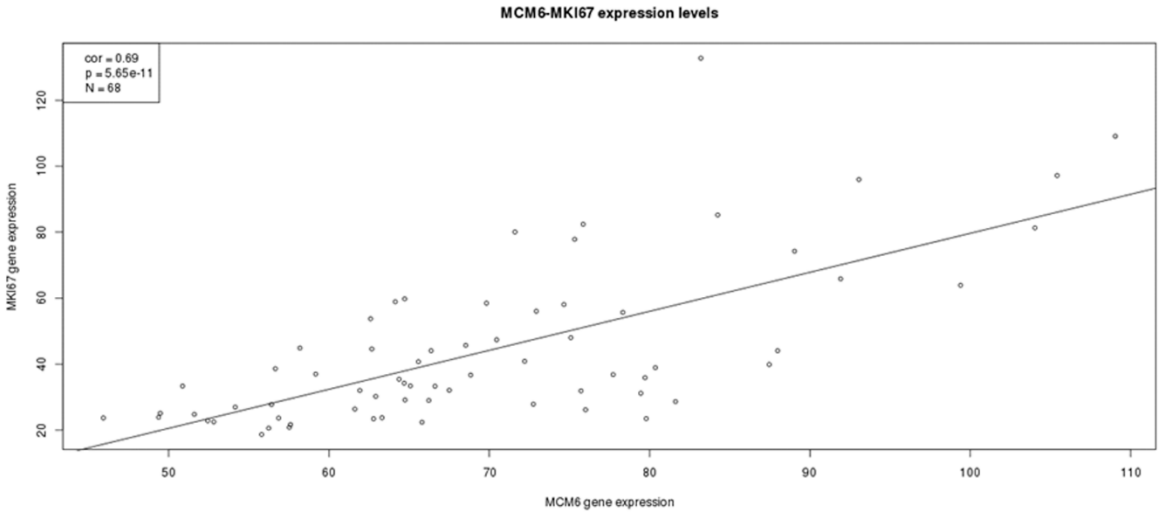


Figure S3. Correlation between mRNA expression levels of *MKI67* and *MCM6*.

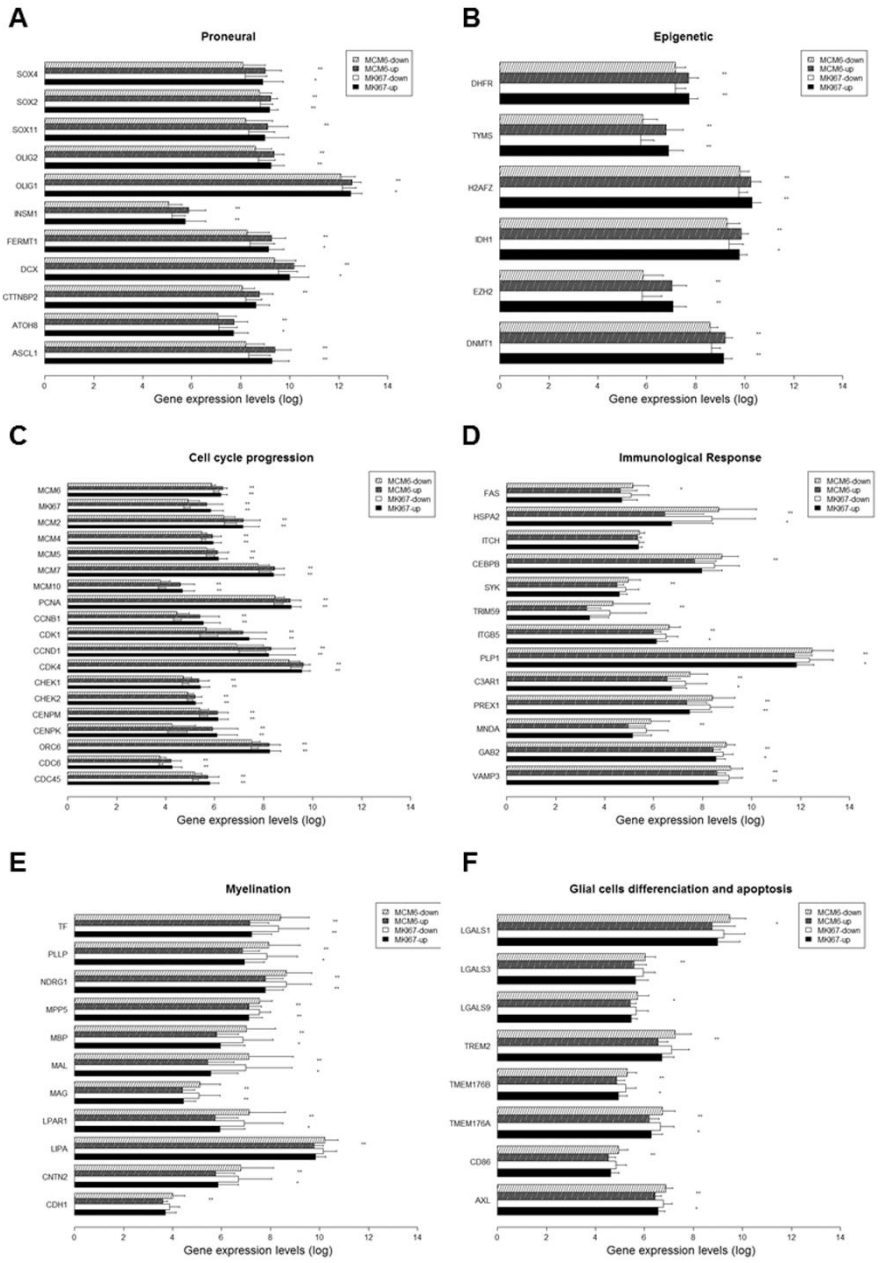


Figure S4. Log2-normalized gene expression levels for the upregulated pro-neural **(A)**, epigenetic **(B)**, and cell cycle **(C)** signatures, as well as the downregulated immune response **(D)**, myelination
(E) and glial cells differentiation **(F)** signatures for both *MCM6* and *MKI67* "-down" and "-up" samples (transcriptomics).

A

	p.val	q.val
hsa03010 Ribosome	1.09E-10	1.77E-08
hsa04110 Cell cycle	4.64E-09	3.78E-07
hsa03013 RNA transport	8.82E-08	4.79E-06
hsa03030 DNA replication	5.18E-07	1.73E-05
hsa03040 Spliceosome	5.30E-07	1.73E-05
hsa03008 Ribosome biogenesis in eukaryotes	3.98E-06	0.00010803
hsa03018 RNA degradation	2.16E-05	0.00050352
hsa03430 Mismatch repair	0.00065689	0.01338404
hsa03410 Base excision repair	0.00081396	0.01474171
hsa00240 Pyrimidine metabolism	0.00134891	0.02165669
hsa03440 Homologous recombination	0.00151398	0.02165669
hsa04115 p53 signaling pathway	0.00159436	0.02165669
hsa03420 Nucleotide excision repair	0.00178626	0.0223969

B

	p.val	q.val
hsa04110 Cell cycle	3.18E-09	5.18E-07
hsa03030 DNA replication	1.15E-06	9.34E-05
hsa03010 Ribosome	2.16E-06	0.00011713
hsa03013 RNA transport	3.98E-05	0.00162063
hsa03008 Ribosome biogenesis in eukaryotes	0.00032958	0.00915883
hsa03040 Spliceosome	0.00033713	0.00915883
hsa04115 p53 signaling pathway	0.00040669	0.0094701
hsa03018 RNA degradation	0.00087649	0.01785845
hsa03440 Homologous recombination	0.00125725	0.02277023
hsa03430 Mismatch repair	0.00168658	0.02749125
hsa00240 Pyrimidine metabolism	0.00348181	0.04781854
hsa03410 Base excision repair	0.00352038	0.04781854
hsa03420 Nucleotide excision repair	0.00382388	0.0479456

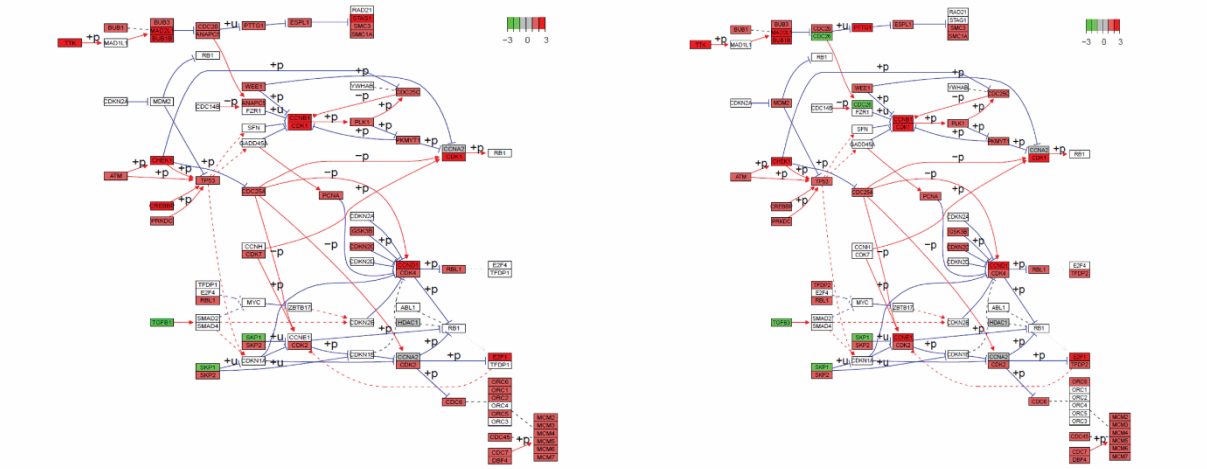


Figure S5. Significantly dysregulated KEGG pathways with an overall upregulation (Table as obtained with Gage) and map of the cell cycle (GraphViz rendered with Pathview) in both *MCM6* **(A)** and *MKI67* **(B)** upregulated samples. Significantly (FDR < 0.01) down- and upregulated genes are colored from green to red, with respect to their fold change of expression (log2).

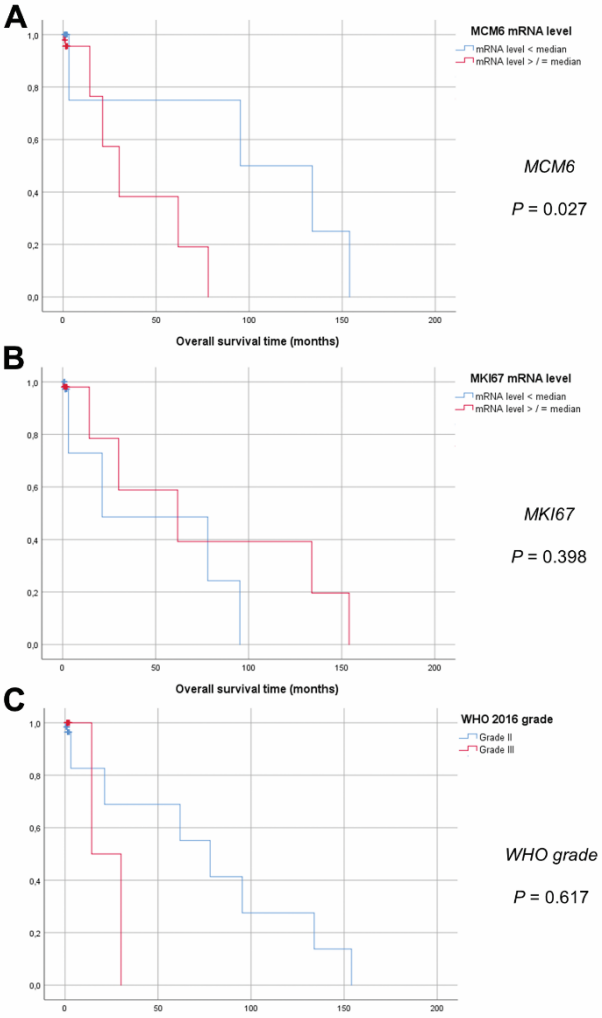


Figure S6. Survival analyses in grades II/III IDH-mutant/1p19q codeleted oligodendrogliomas selected from TCGA database. Kaplan–Meier curves with log-rank tests (overall survival). **A**. *MCM6* mRNA level, median threshold (arbitrary units). **B**. *MKI67* mRNA level, median threshold (arbitrary units). **C.** WHO 2016 grade.

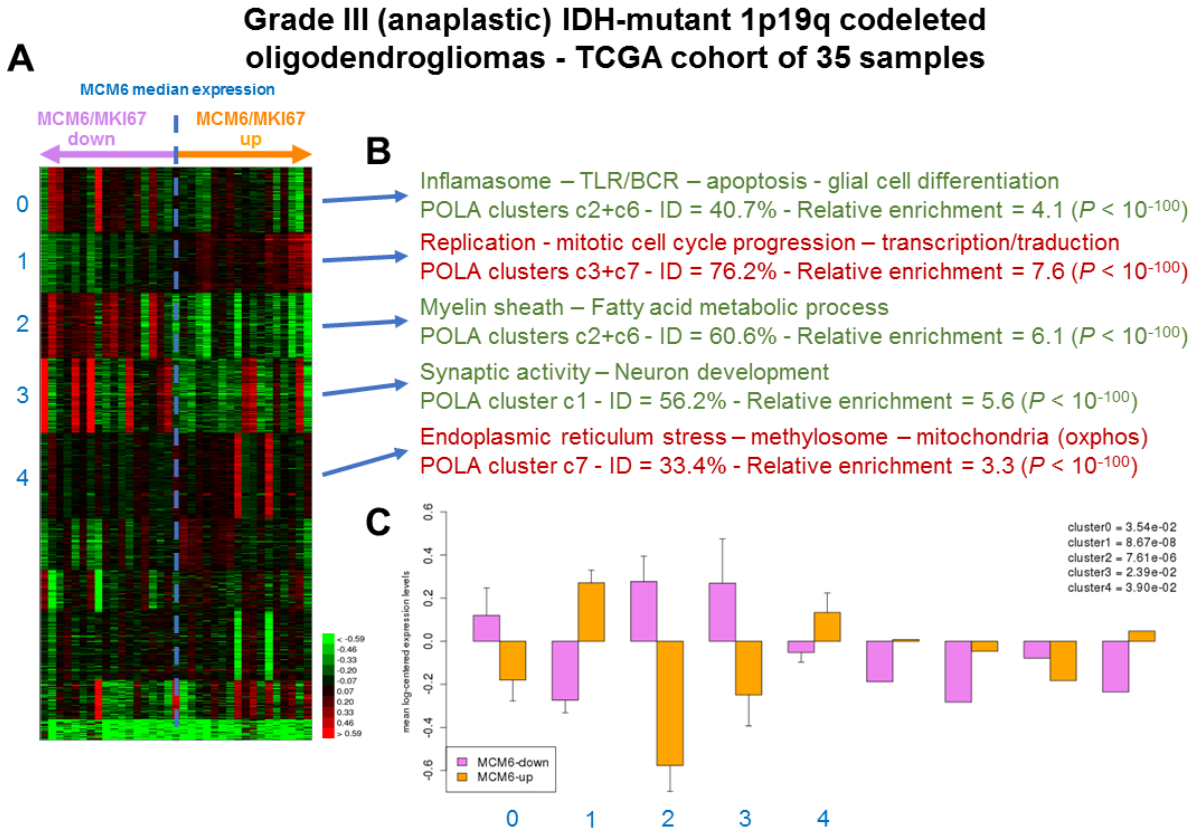


Figure S7. K-means clustering based on *MCM6* and/or *MKI67* mRNA expression gradient (from lowest to highest) of the 35 transcriptomes of anaplastic IDH-mutant/1p19q codeleted oligodendrogliomas selected from TCGA bank. The K-means clustering delineated five significant clusters of co-expressed genes ($P < 0.05$) with highly correlated molecular functions. **A.** Heatmap of the 10 k-means clusters organized based on the extent of *MCM6* and *MKI67* expression (green, red and black indicate downregulated, upregulated and median genes, respectively). **B.** Overview of the most relevant functional annotations for the significantly over- and under-expressed clusters
(in red and green hues, respectively, all *P(FDR)* ≤ 0.01). Each of these clusters was compared to the clusters found of the POLA cohort, and their identities calculated based on the corresponding POLA clusters. **C.** Mean gene expression (overall with standard error at the mean) for each cluster where both upregulated samples for *MCM6* and/or *MKI67* are compared to their downregulated counterparts (AU—arbitrary units). Upregulated and downregulated samples are defined relatively to *MCM6* median gene expression.

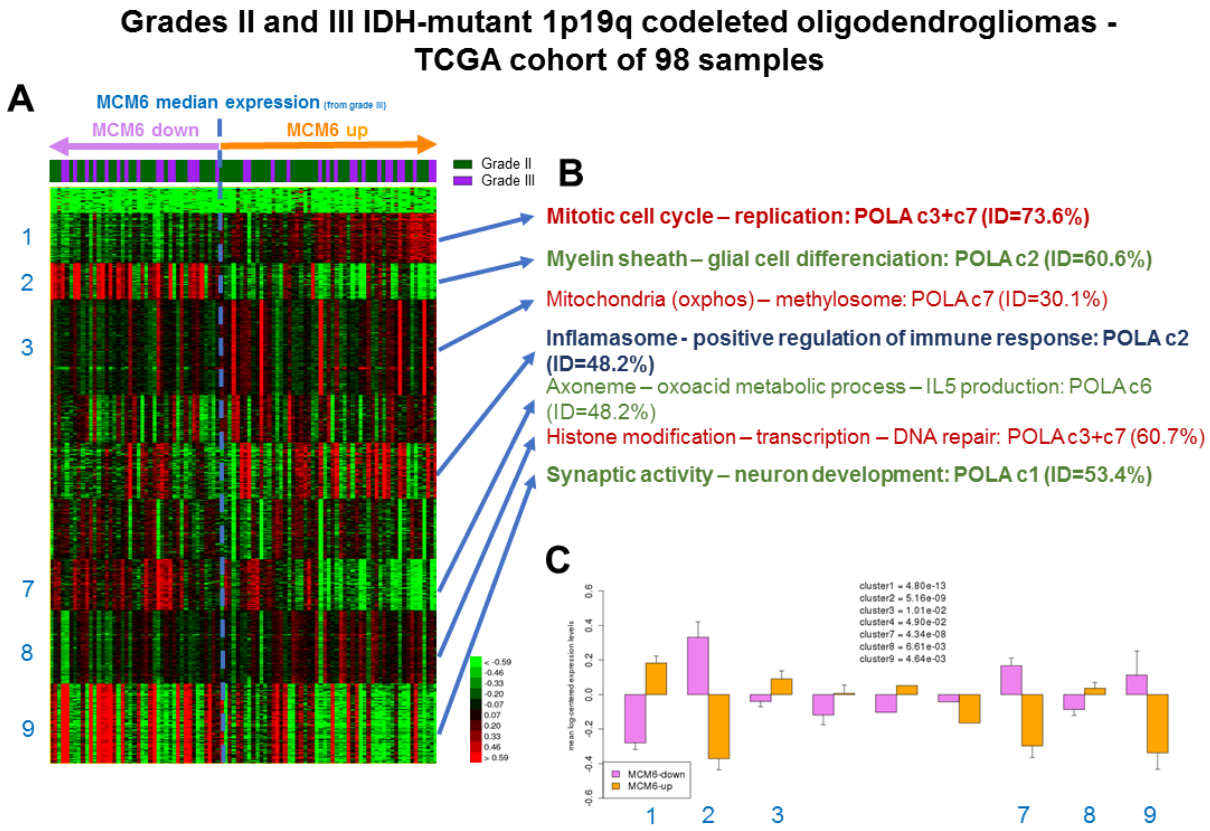


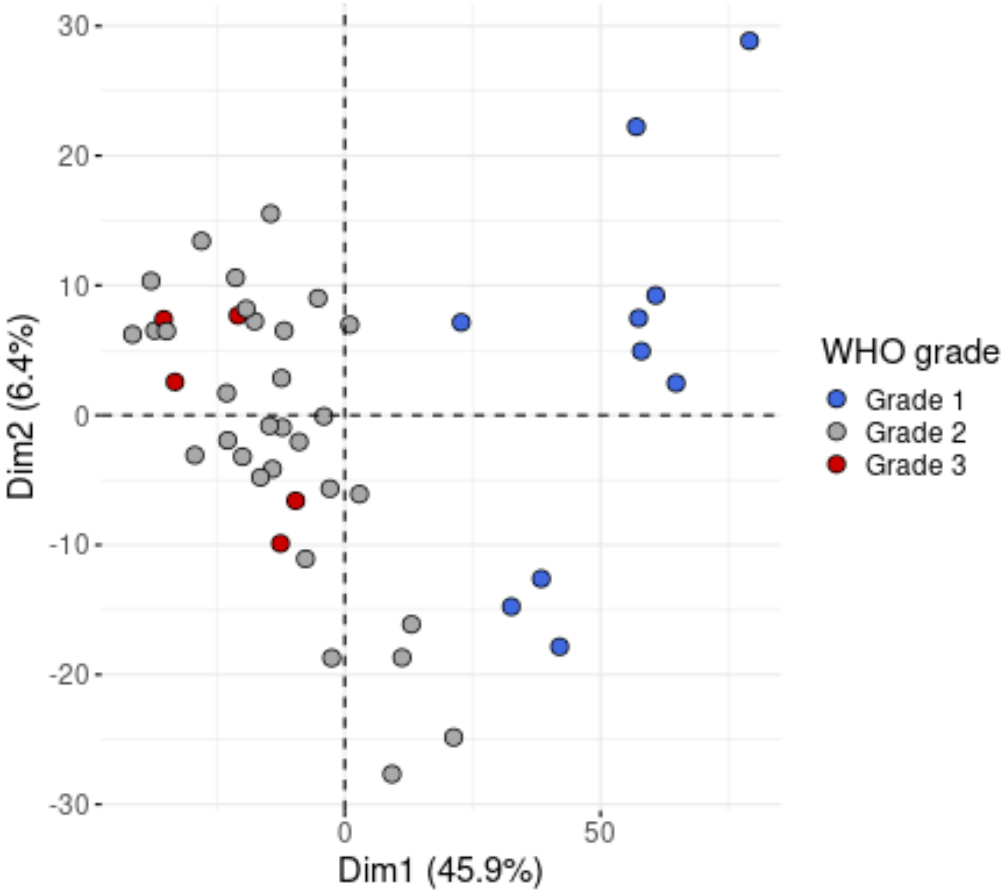
Figure S8. K-means clustering based on *MCM6* and/or *MKI67* mRNA expression gradient (from lowest to highest) of the 98 transcriptomes of grades II and III IDHmutant/1p19q codeleted oligodendrogliomas selected from TCGA bank. The K-means clustering delineated five significant
clusters of co-expressed genes ($P < 0.05$) with highly correlated molecular functions. **A.** Heatmap of the ten k-means clusters organized based on the extent of *MCM6* and *MKI67* expression (green, red and black indicate downregulated, upregulated and median genes, respectively). **B.** Overview of the most relevant functional annotations for the significantly over- and under-expressed clusters (in red and green hues, respectively, all *P(FDR)* ≤ 0.01). Each of these clusters was compared to the clusters found of the POLA cohort, and their identities calculated based on the corresponding
POLA clusters. **C.** Mean gene expression (overall with standard error at the mean) for each cluster where both upregulated samples for *MCM6* and/or *MKI67* are compared to their downregulated counterparts (AU—arbitrary units). Upregulated and downregulated samples are defined relatively to *MCM6* median gene expression.

Annexe 2. Tables et figures supplémentaires de l'article « *Correlation between DNA methylation and cell proliferation identifies new candidate predictive markers in meningioma* »

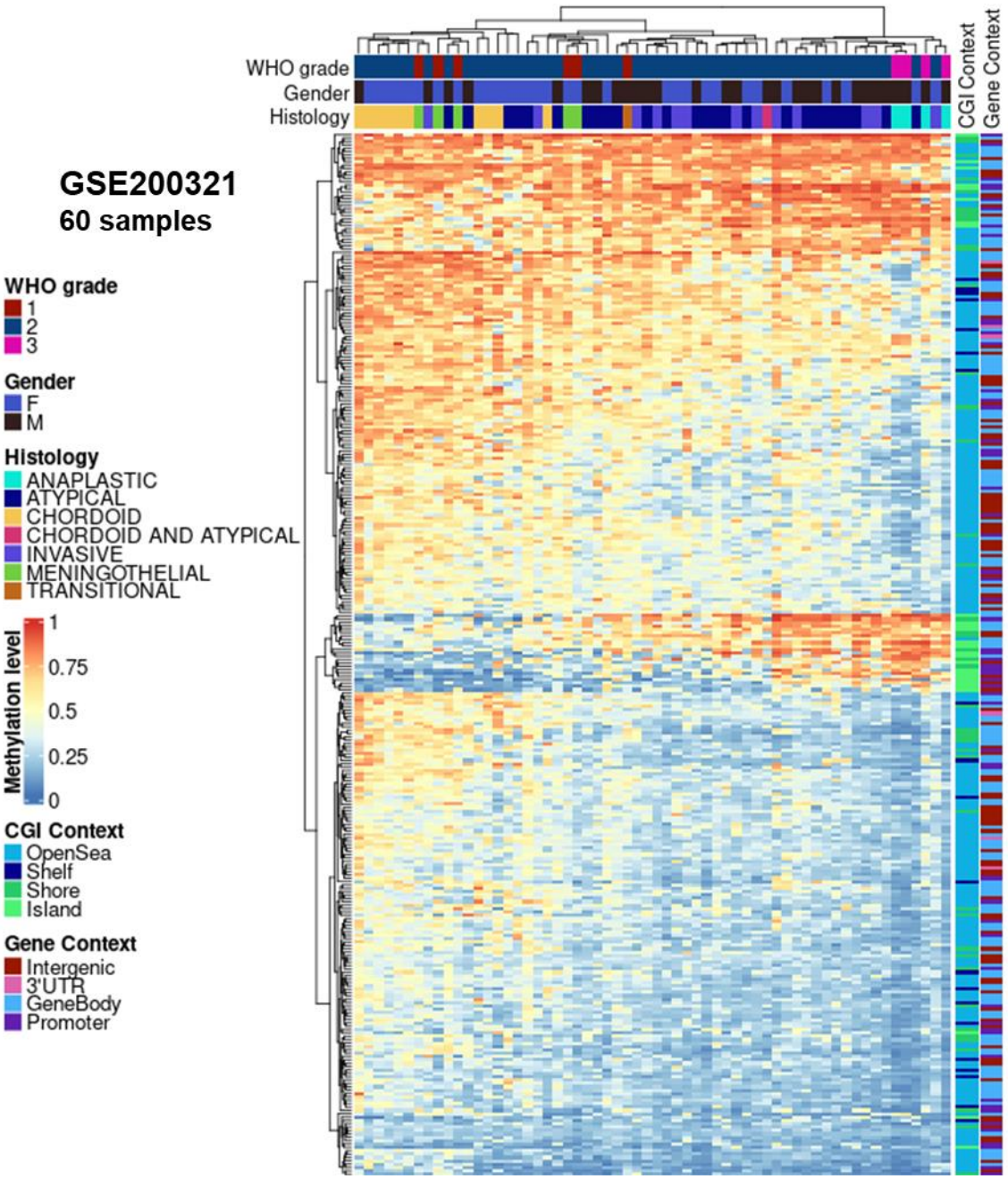
Tables supplémentaires

Tables S1 à S9. https://www.mdpi.com/article/10.3390/cancers14246227/s1

Figures supplémentaires



Supplementary Figure S1. Principal component analysis of the top 2099 CpGs from the differential analysis between grade 1 and grade 2+3 meningiomas.



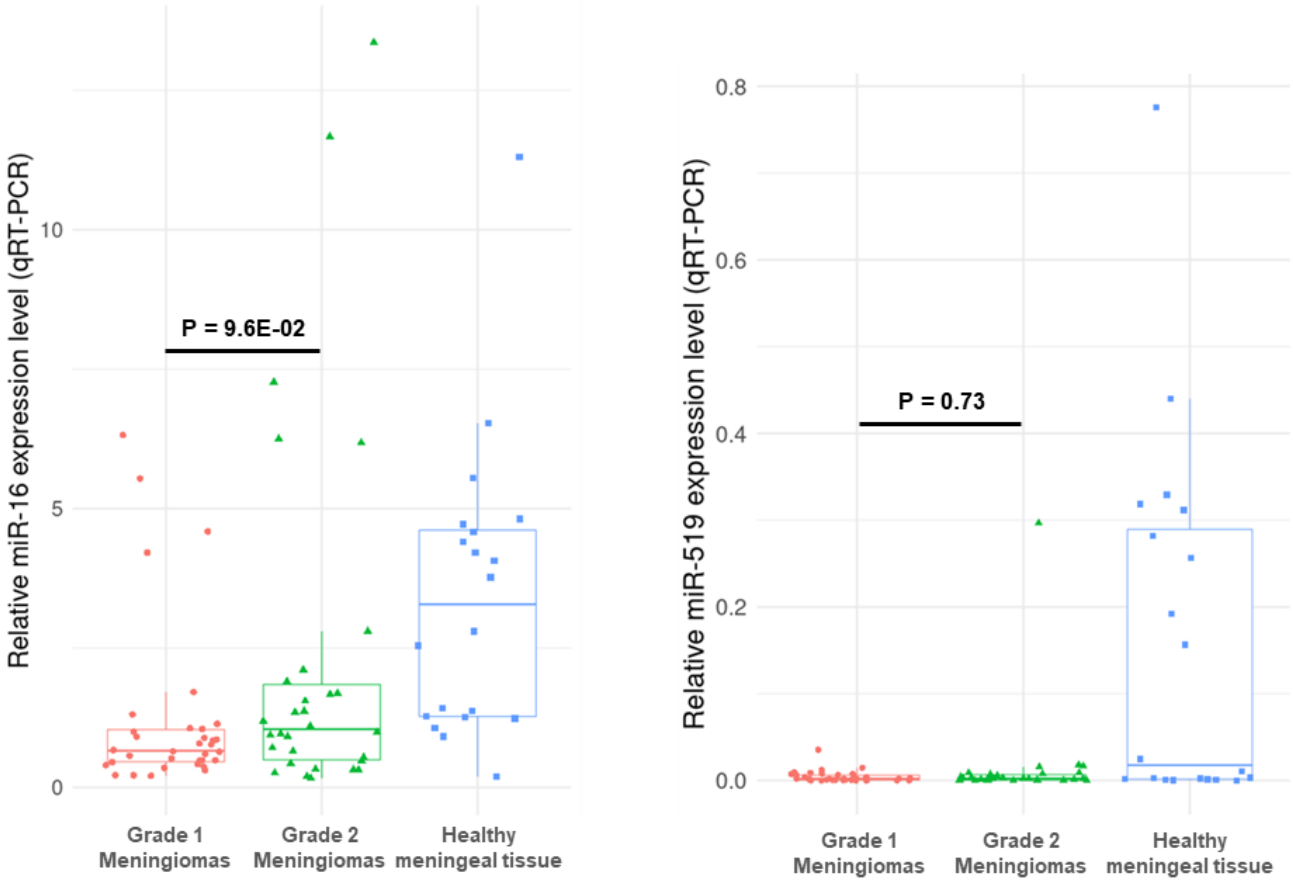
Supplementary Figure S2. Validation of the proliferative signature on an external methylation dataset (GSE200321). Hierachical clustering against the 310 CpGs and 60 samples of various meningioma grades and histological types (6 grade 1: 5 meningothelial and 1 transitional; 50 grade 2: 10 chordoid, 27 atypical and 13 atypical and invasive; 4 grade 3: anaplastic). Raw data from this dataset was processed with the same analytical pipeline as the main dataset presented in this work. CGI: CpG island, WHO: World Health Organization.

Annexe 3. Tables et figures supplémentaires de l'article « *MicroRNAs miR-16 and miR-519 control meningioma cell proliferation via overlapping transcriptomic programs shared with the RNA-binding protein HuR* »

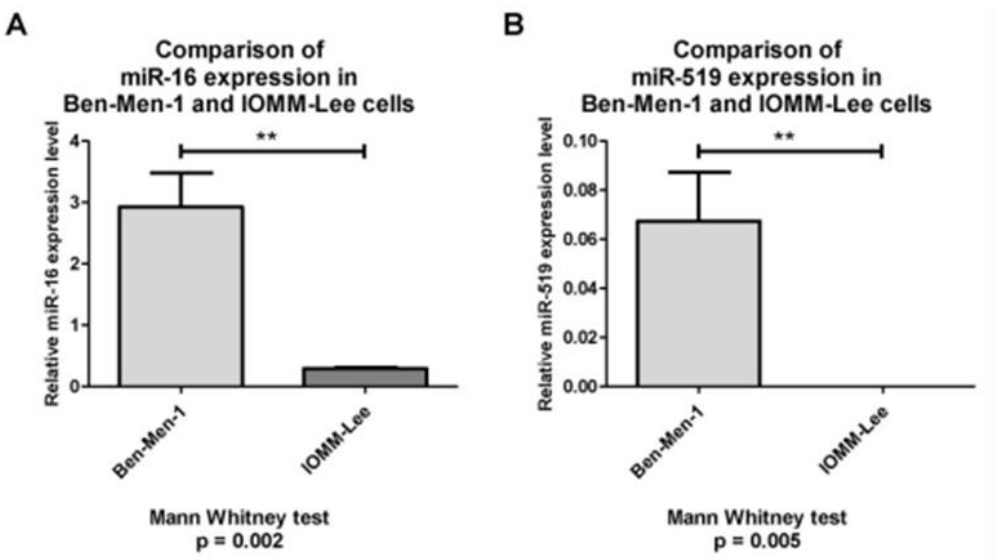
Tables supplémentaires

Tables S1 à S8. https://www.frontiersin.org/articles/file/downloadfile/1158773_supplementary-materials_tables_1_xlsx/octet-stream/Table%201.XLSX/1/1158773

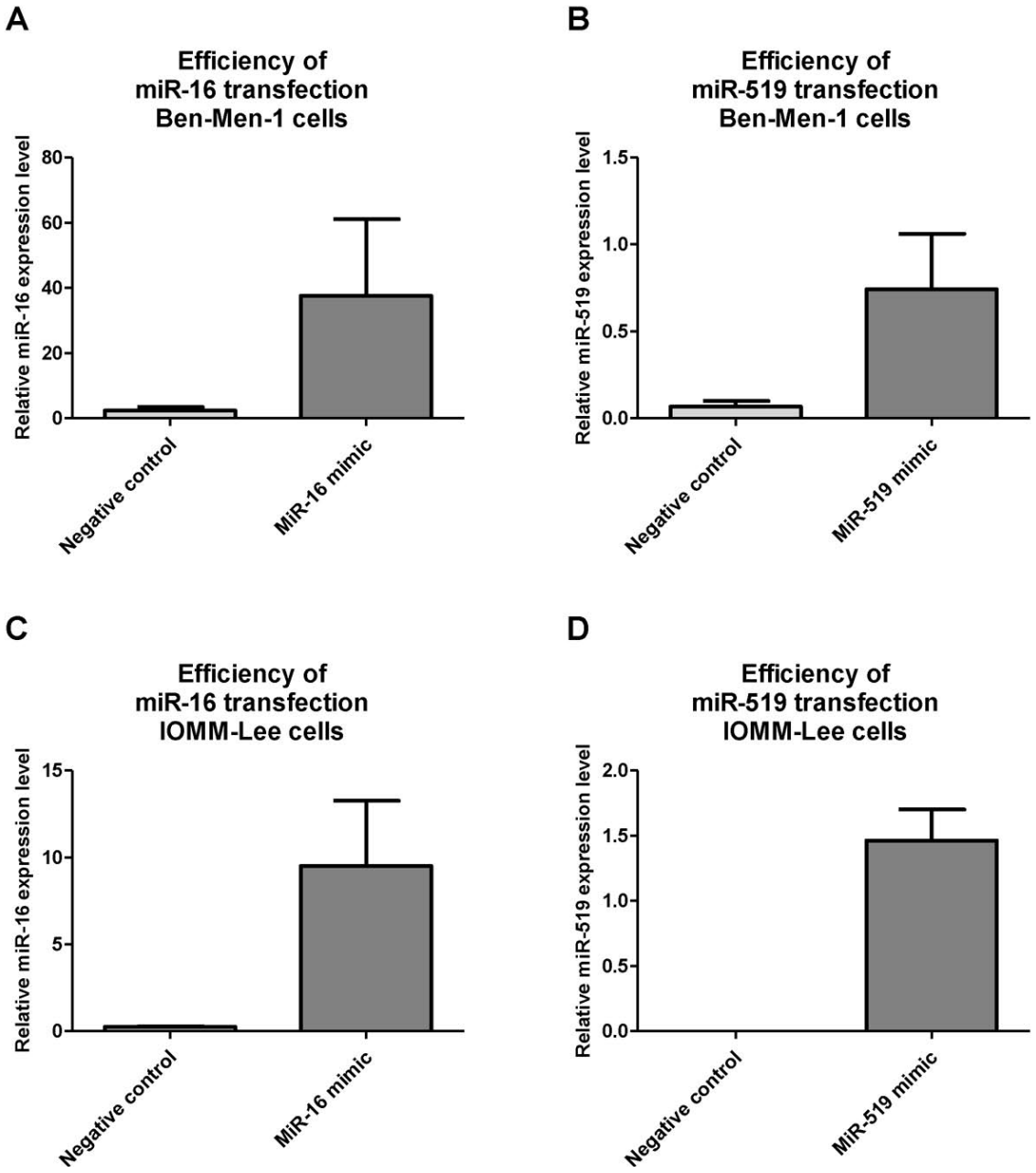
Figures supplémentaires



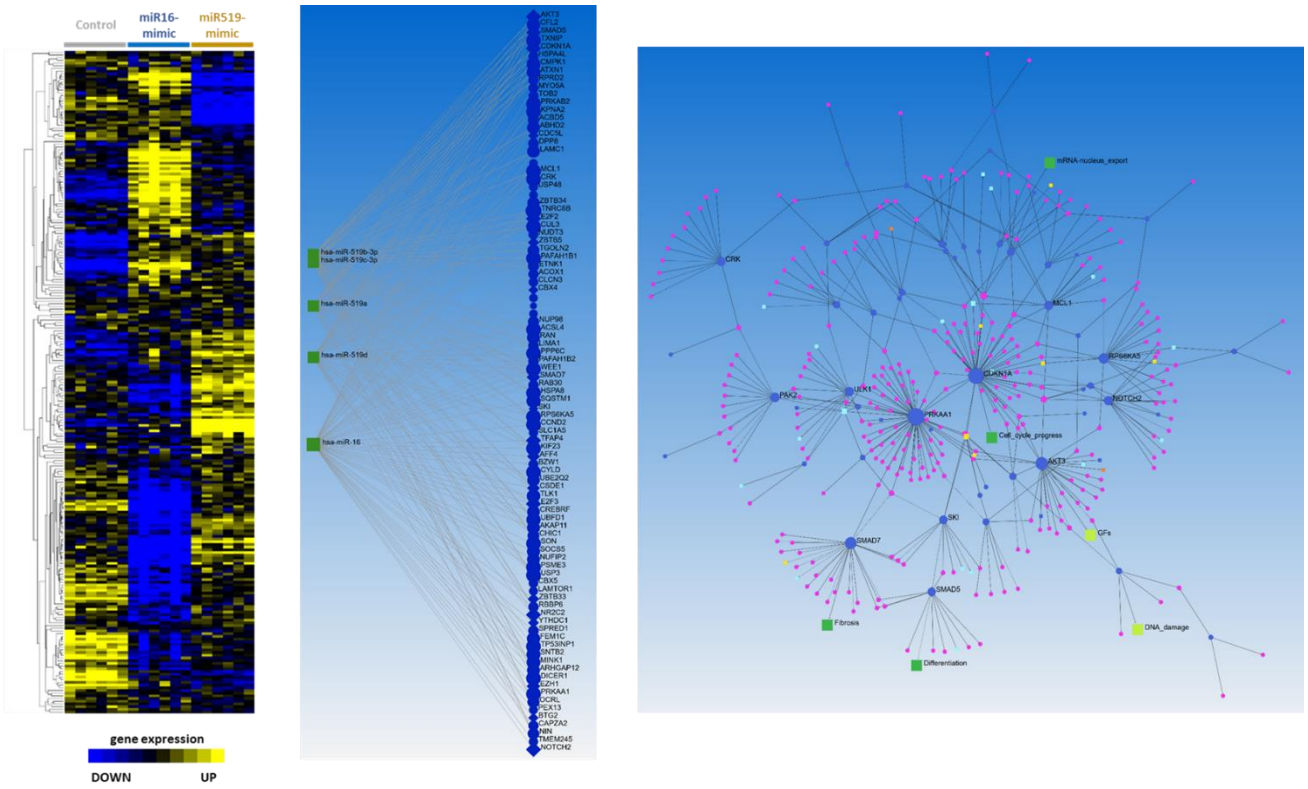
Supplementary Figure S1. MiR-16 and miR-519 expression levels in human meningioma and healthy meningeal tissues. Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analysis of miR-16 (A) and miR-519 (B) expression in human meningiomas (grade 1: n = 30; grade 2: n = 34) and non-tumoral tissue (n = 20). Boxplot (minimum, first quartile, median, third quartile, and maximum) with individual scatterplots. Wilcoxon Mann-Whitney U-tests.



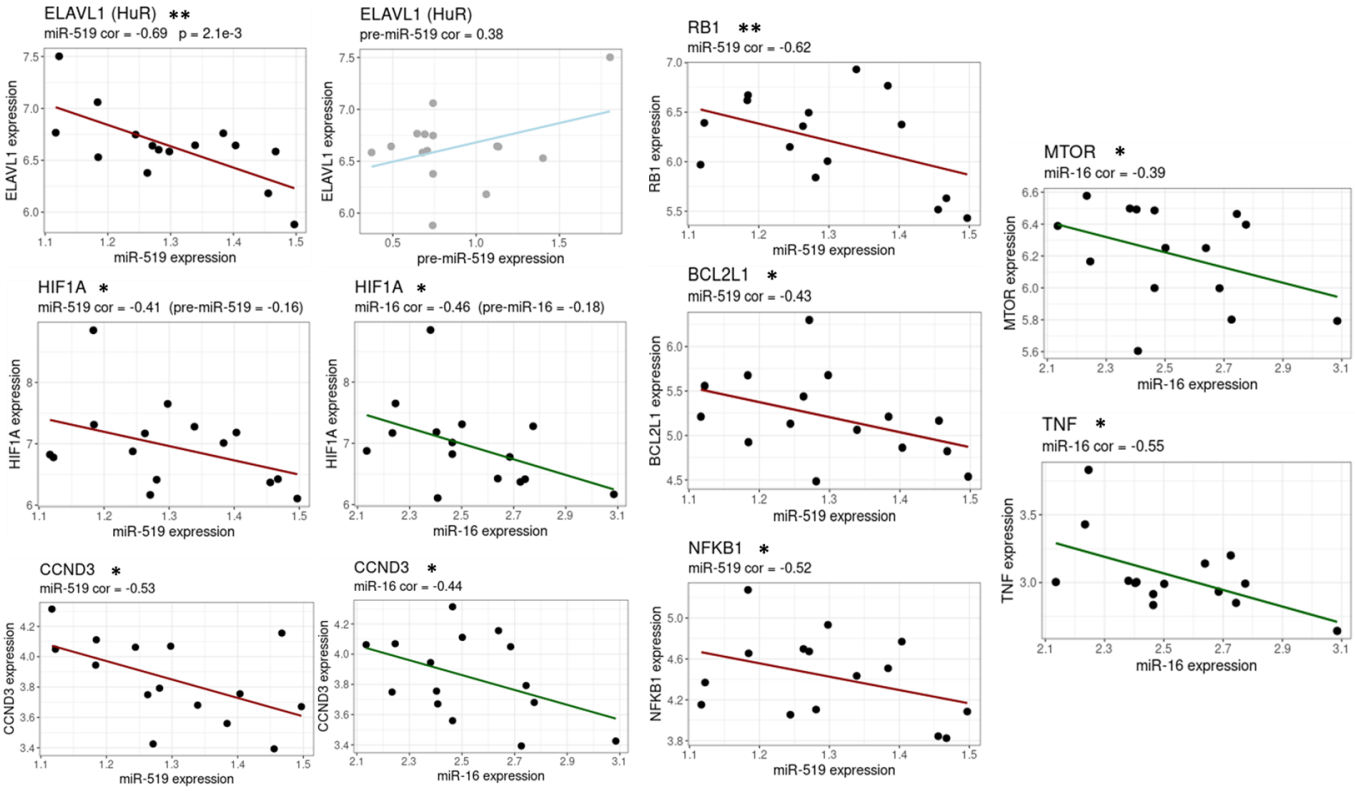
Supplementary Figure S2. Comparison of basal expression level of miR-16 and miR-519 between the human benign meningioma immortalized Ben-Men-1 cell line and the human anaplastic meningioma IOMM-Lee cell line, showing a significantly lower expression of miR-16 (A) and miR-519 (B) in the IOMM-Lee cells (qRT-PCR; n=6).



Supplementary Figure S3. Efficiency of miR mimics transfection evaluated by qRT-PCR against miR-mimic negative controls. (A) Transfection of hsa-miR16 mimic in Ben-Men-1 cell line. (B) Transfection of hsa-miR519 mimic in Ben-Men-1 cell line. (C) Transfection of hsa-miR16 mimic, IOMM-Lee cell line. (D) Transfection of hsa-miR519 mimic in IOMM-Lee cell line (n=3).



Supplementary Figure S4. Gene signature identified with miR-16 and miR-519 transcriptomics. Left panel: Hierarchical clustering heat map of the 208 genes from the five differential clusters (c1 to c5) further reduced to mir-16, miR-519 overlapping targets. The highly differential and correlative structure is representative of the five clusters. Middle panel: miR network of shared targets of miR-16 and miR-519. Squares: miRs, losanges: transcription factors / transcriptional regulators; circle: other targets). Shape size is proportionally linked with the number of shared miRs. Right panel: Signaling network reconstructed with the 208-gene signature identified with miR-16 and miR-519 transcriptomics. Circles (pink and blue): gene targets, blue squares: chemicals, green squares (light and dark): stimuli and phenotypes (ontologies), turquoise and yellow squares: protein complexes and families, orange squares: other small molecules. Networks were constructed with NetworkAnalyst (https://www.networkanalyst.ca/).



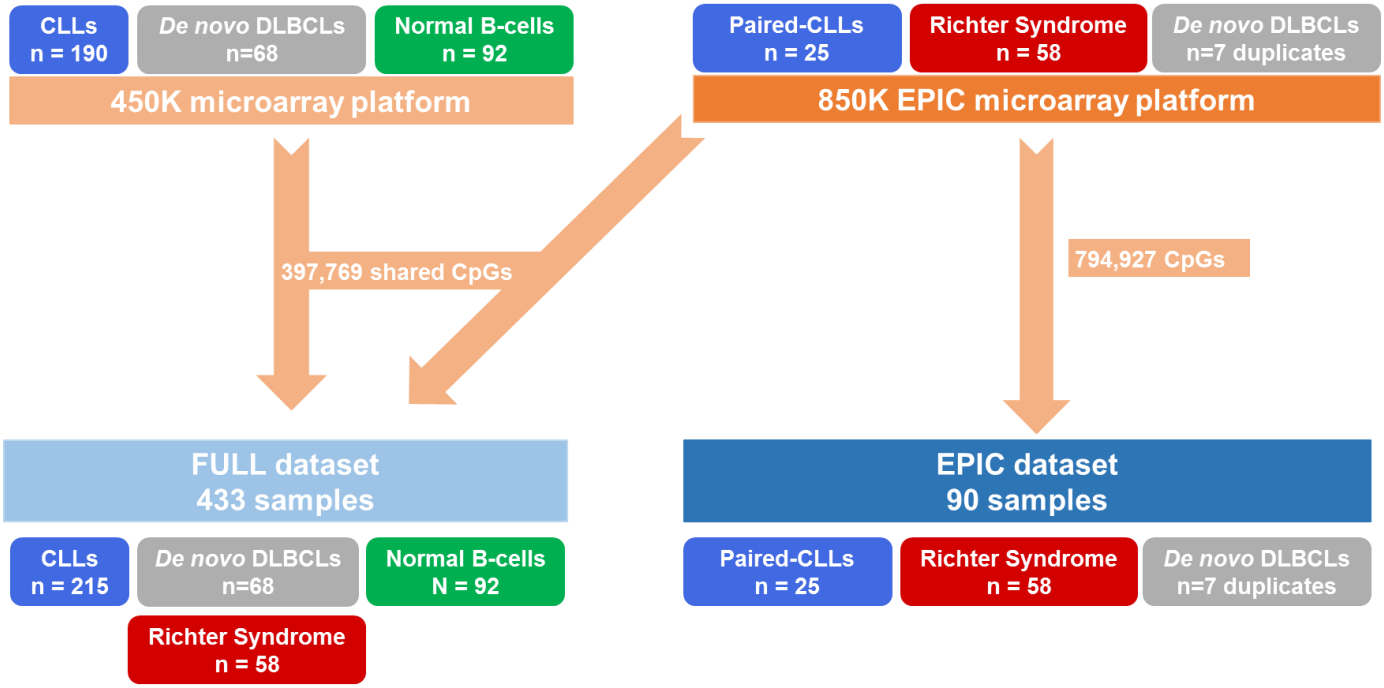
Supplementary Figure S5. Validation of main transcriptomic results through the integration of meningioma patient transcriptomes and corresponding microRNA microarrays (public dataset GSE88721 from Dalan et al, 2017; PMID: 28327132; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE88721). Hallmark genes from the original 5-cluster transcriptomic signature, the meningioma signature and proposed progression markers are tested against candidate miR-16 and miR-519 and preprocessed pre-miR-16 and pre-miR519 (Pearson's correlations; * $p < 0.05$; ** $p < 0.01$), including direct miR-16 and/or miR-519 target RNAs. Patient samples include diverse histological subtypes (including one healthy meninges) and meningioma grades 1 and 2. Expression levels are normalized against background and log2-transformed.

Annexe 4. Tables, figures, méthodes et références supplémentaires de l'article « *Molecular characterization of Richter syndrome identifies de novo diffuse large B-cell lymphomas with poor prognosis* »

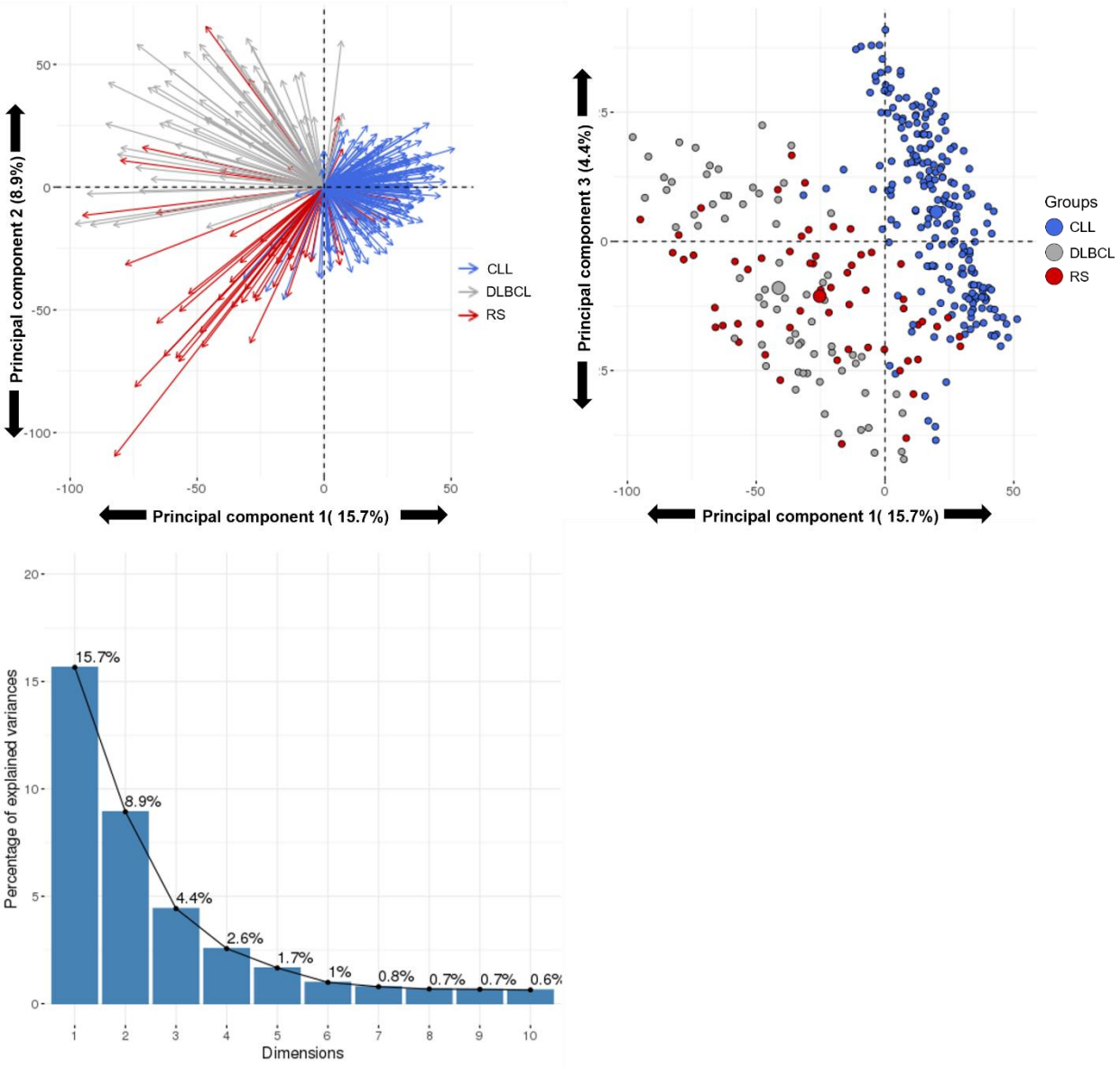
Tables supplémentaires

Tables S1 à S11. https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-022-34642-6/MediaObjects/41467_2022_34642_MOESM4_ESM.xlsx

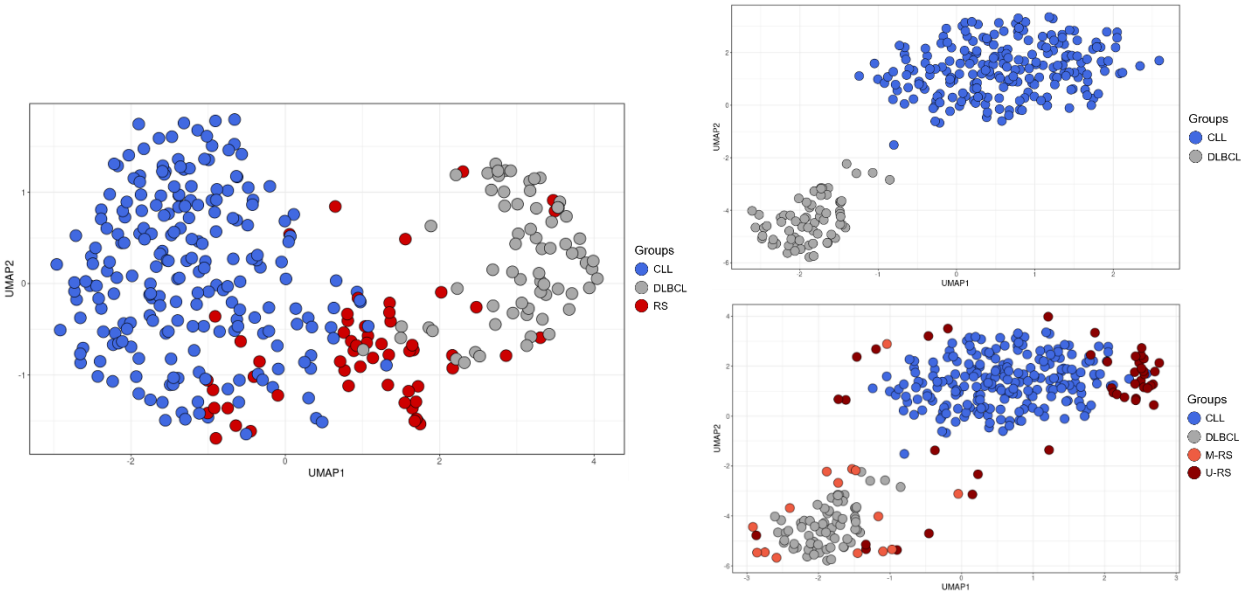
Figures supplémentaires



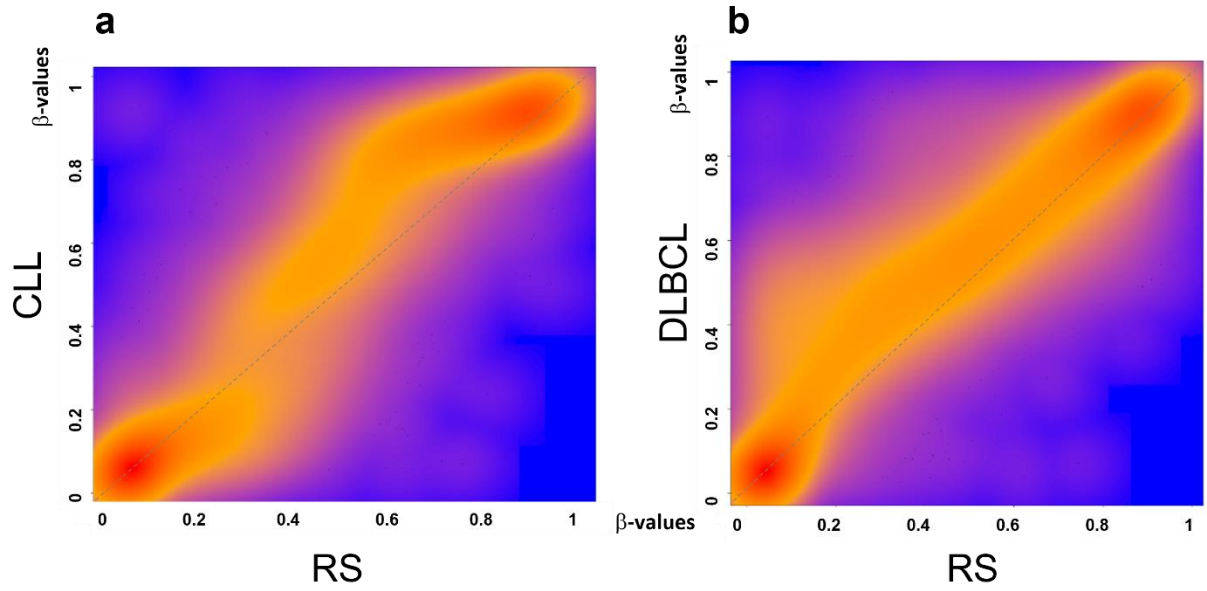
Supplementary Fig. 1. DNA methylation data analysis workflow. CLL: chronic lymphocytic leukemia; DLBCL: diffuse large B-cell lymphoma.



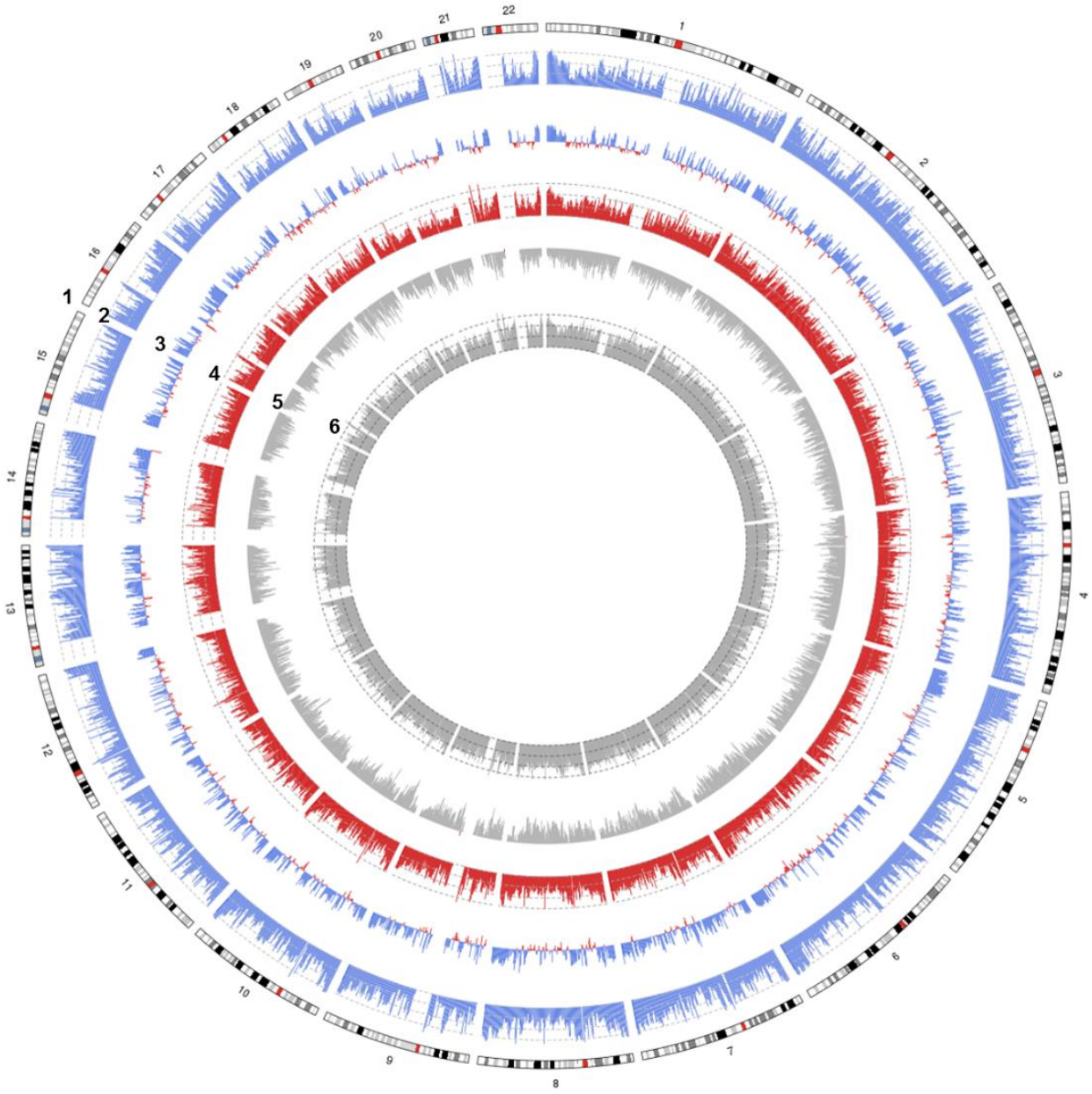
Supplementary Fig. 2. RS is hypomethylated as compared with *de novo* DLBCL or CLL. Unsupervised PCA of 58 RS, 215 CLLs and 68 DLBCLs. On PC1: RS is hypomethylated against CLL. On PC2: RS is hypomethylated against DLBCL. On PC3: CLL progression. Other dimensions each explain a small amount of variance (as seen on the scree-plot). For each group, geometrical centers are represented by bigger circles of the same color. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; PCA: principal component analysis; RS: Richter syndrome.



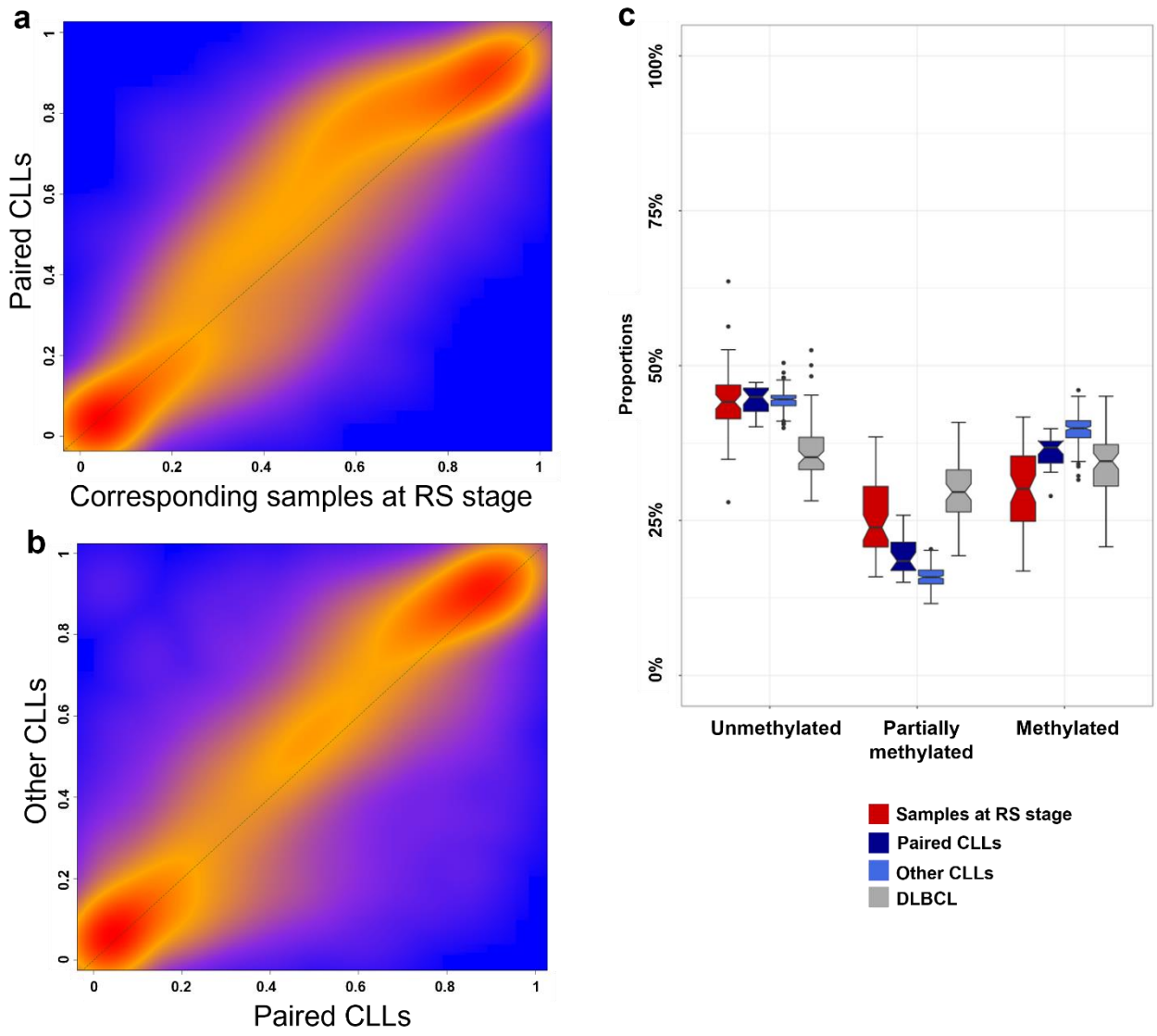
Supplementary Fig. 3. UMAP of all CpGs from FULL dataset.
Left part: UMAP applied to the 58 RS, 215 CLL and 68 DLBCL cases, on 397,769 QC-selected CpGs shared by both 450K and EPIC microarray platforms. Right part: UMAP of CLL and *de novo* DLBCLs (top), with predicted scattering of RS cases (bottom), demonstrating that sample distribution is not driven by *IGHV* mutational status. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; M-RS: IGHV-mutated Richter syndrome; RS: Richter syndrome; U-RS: IGHV-unmutated Richter syndrome.



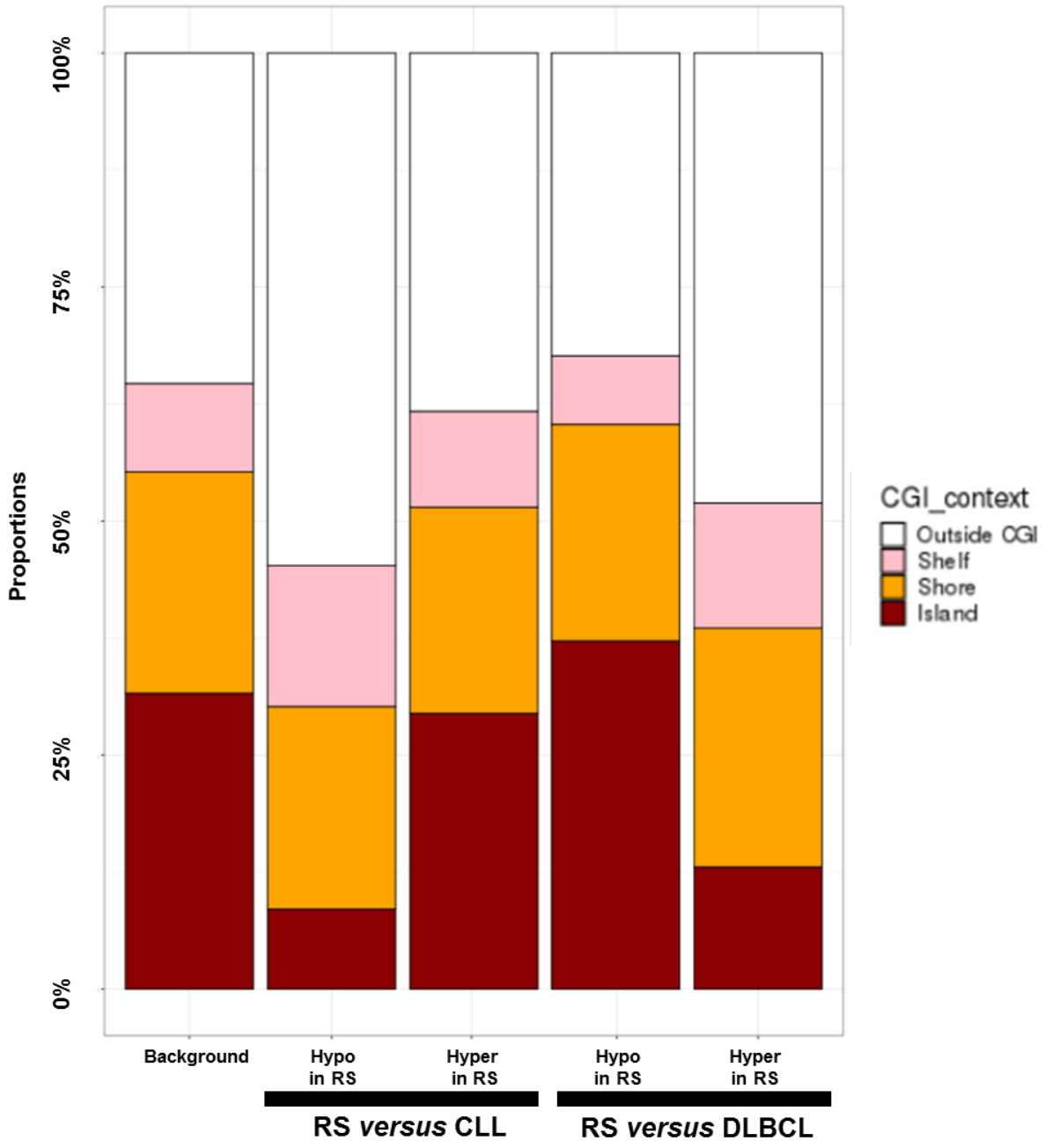
Supplementary Fig. 4. Comparative smoothed methylation scatterplots between RS and CLL or DLBCL. Each dot represents one of the 397,769 CpGs in the FULL dataset. For each CpG, coordinates on the x-axis correspond to the mean beta-value for the RS group and coordinates on the y-axis correspond to the mean beta-value for **a** the CLL group or **b** the DLBCL group. Scale from blue (no density) to yellow (medium density) and red (high density), purple and orange are medium-low and medium-high densities, respectively. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; RS: Richter syndrome.



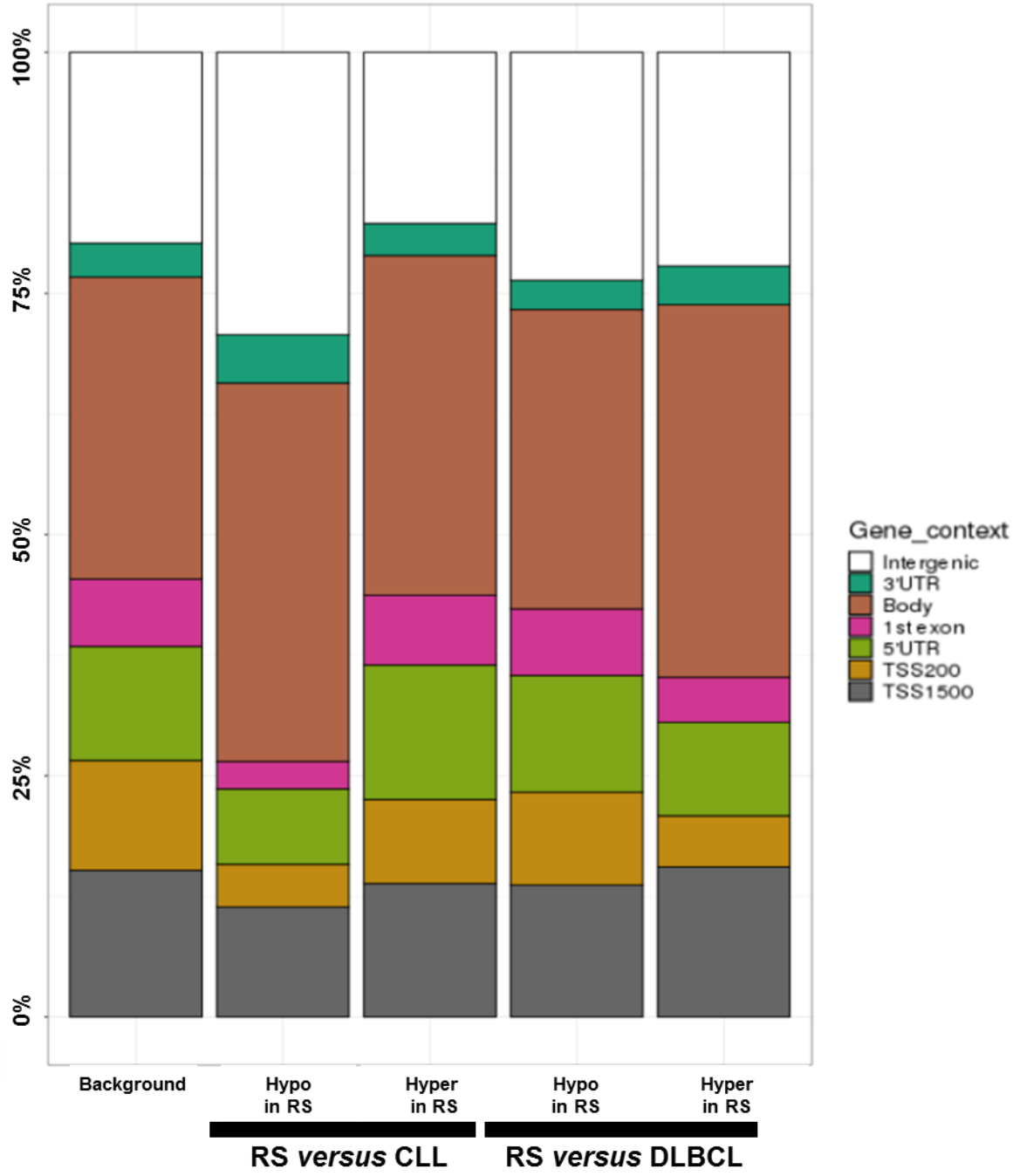
Supplementary Fig. 5. Circular plot of median CpG methylation levels and differences in CLL (blue), RS (red) and DLBCL (grey) over sliding windows of 500 kb. RS is hypomethylated as compared with CLL and DLBCLs, but CLL relative hypermethylation is due to localized and highly methylated areas while DLBCL relative hypermethylation is evenly distributed throughout the genome. From outer to inner track: 1) Chromosome number and ideograms with cytobands; 2) CLL methylation levels; 3) CLL minus RS methylation differences. Range from -10% to +30%. Areas hypermethylated in CLL are represented with blue peaks pointing towards the CLL track while areas hypermethylated in RS are represented with red peaks pointing towards the RS track; 4) RS methylation levels; 5) RS minus DLBCL methylation differences (range from -30% to +10%). Areas hypermethylated in RS are represented with red peaks pointing towards the RS track while areas hypermethylated in DLBCLs are represented with grey peaks pointing towards the DLBCL track; 6) DLBCL methylation levels. On tracks 2, 4 and 6, dashed lines represent 0%, 25%, 50%, and 75% average methylation. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; RS: Richter syndrome.



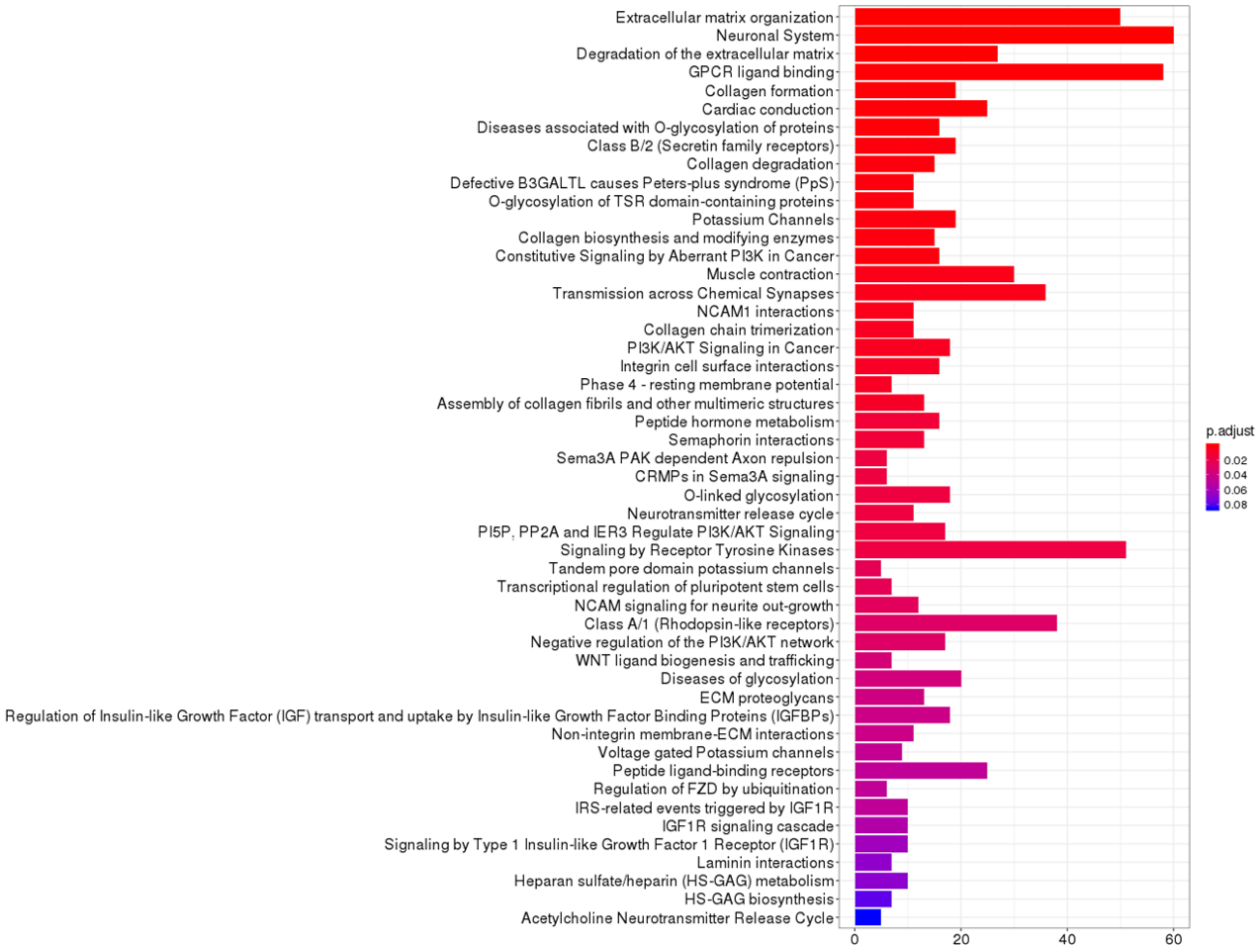
Supplementary Fig. 6. Comparative smoothed methylation scatterplots between: a paired CLLs and corresponding samples at RS stage or **b** paired CLLs and other CLLs. Scale from blue (no density) to yellow (medium density) and red (high density). **c** Distribution of unmethylated (beta-value < 0.3), partially methylated (0.3 < beta-value < 0.7) and methylated (beta-value >0.7) CpGs across RS, DLBCL, paired-CLLs and other CLLs, with sample variability. Box plot: the center line, box limits, whiskers and points represent the mean, 25th and 75th percentile, 1.5x interquartile range and individual samples beyond the 1.5x interquartile, respectively. RS: n=58 biologically independent samples; paired-CLLs: n=25 biologically independent samples; other CLLs: n=190 biologically independent samples; DLBCL: n= 68 biologically independent samples. Source data are provided as a Source Data file. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; RS: Richter syndrome.



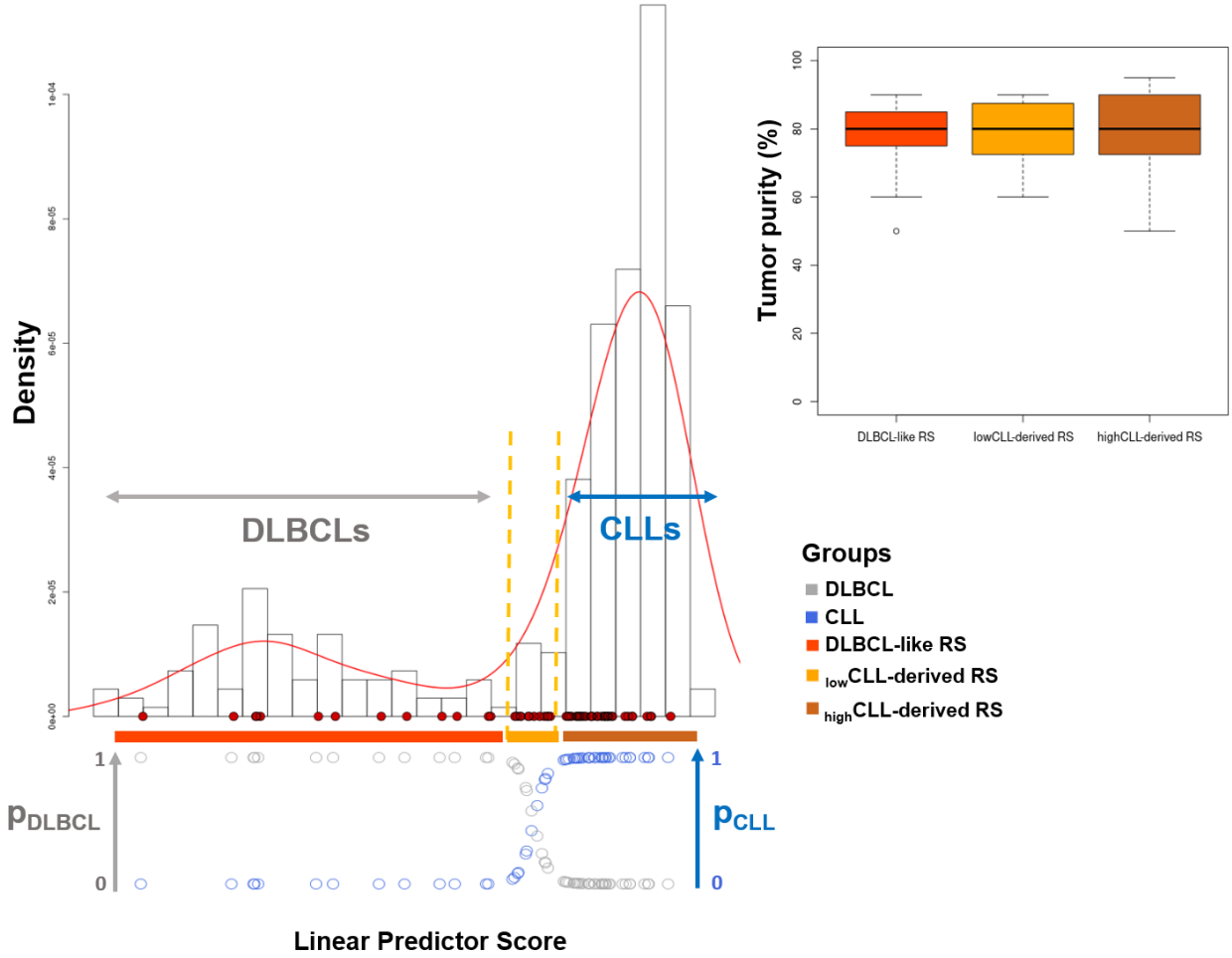
Supplementary Fig. 7. Location of RS vs CLL and DLBCL differential CpGs regarding regulatory region context. Distribution of significant and differential CpGs (FDR < 0.01; beta-value differential > 10%) outside CGI and inside CGI (Shelf, Shore, Island). A methylation hallmark of RS is the gain and loss of differential CpGs outside and inside CpG islands, respectively. Distribution bias against background for CGI: CpGs hypomethylated in RS versus CLL were enriched outside CGI and impoverished in Island (p=2.98e-4). CpGs hypermethylated in RS versus DLBCLs were enriched outside CGI and impoverished in Island (p=1.54e-2). CGI: CpG island; CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; FDR: false discovery rate; Hyper: hypermethylated; Hypo: hypomethylated; RS: Richter syndrome.



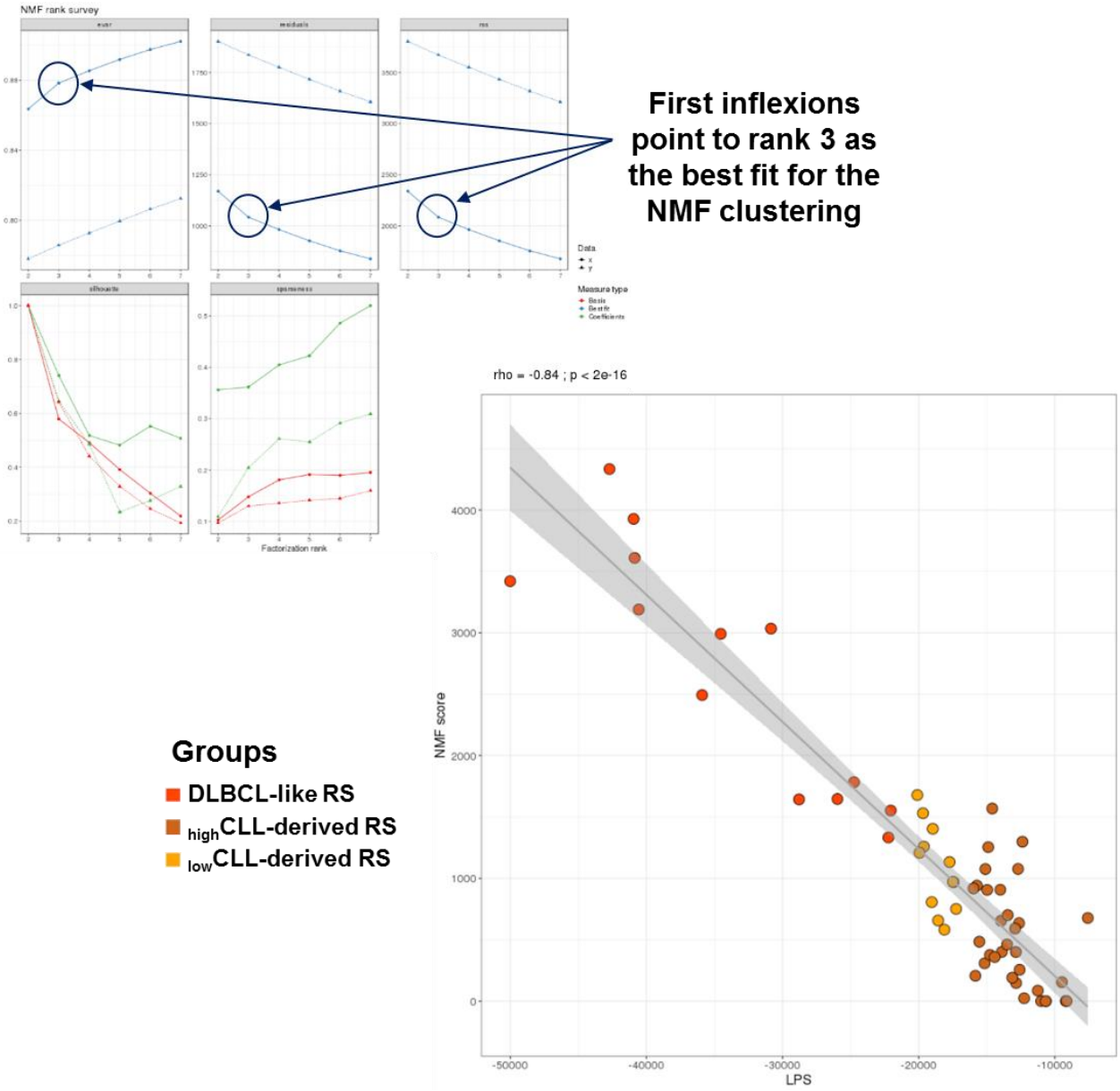
Supplementary Fig. 8. Location of RS vs CLL and DLBCL differential CpGs regarding gene context. Distribution of differential CpGs (FDR < 0.01; beta-value differential > 10%) among the different genomic locations around genes (3'UTR, body, first exon, 5'UTR, TSS200, TSS1500) and outside genes (intergenic). TSS200 and TSS1500 represent areas located at 200 bp and 1500 bp of transcription start sites, respectively. There is no significant bias of RS differential CpGs regarding gene context, whether against CLL or DLBCL. CpGs hypomethylated in RS vs CLL were enriched 1.5-fold in intergenic CpGs. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; FDR: false discovery rate; Hyper: hypermethylated; Hypo: hypomethylated; RS: Richter syndrome; TSS: transcription start site; UTR: untranslated region.



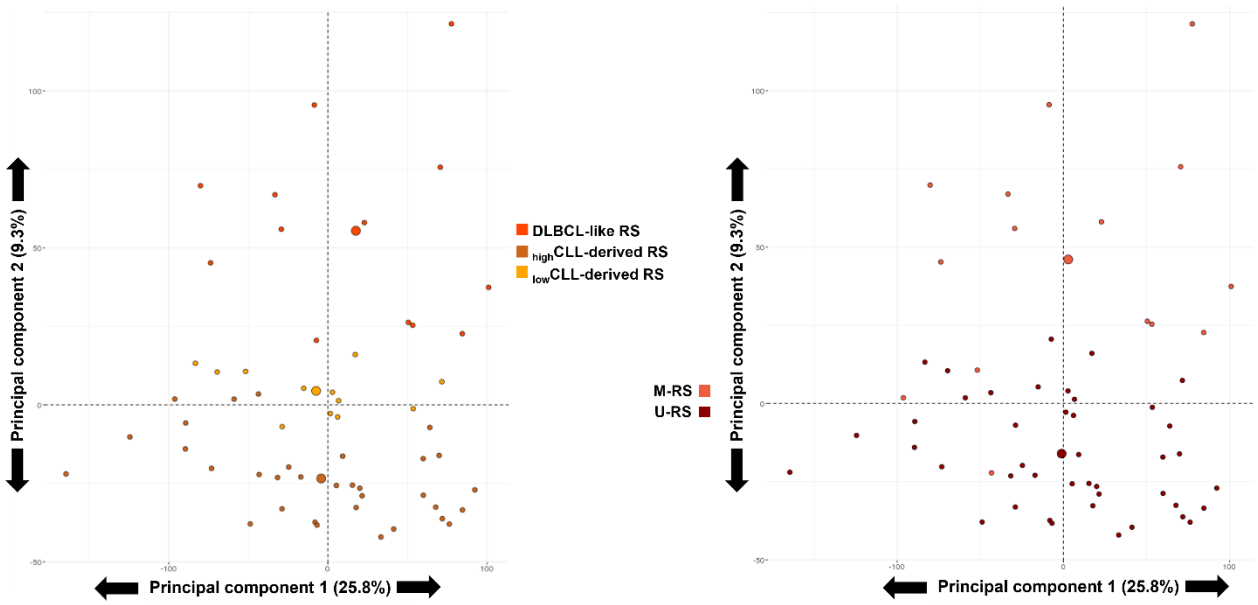
Supplementary Fig. 9. RS vs DLBCL top functional annotations, as returned by ReactomePA. Gene Set Enrichment Analysis permutation tests adjusted for the FDR. From 1,615 differential DMRs associated with 1,768 genes (min_smoothed_FDR, HMFDR and Fisher all < 0.01; max(beta-value differential) > 30%; at least 3 CpGs with gap < 1 kb). 98.9% of these DMRs were hypomethylated in RS. DLBCL: *de novo* diffuse large B-cell lymphoma; FDR: false discovery rate; Fisher: Fisher's multiple comparison statistic; HMFDR: harmonic mean of the individual components FDR; RS: Richter syndrome.



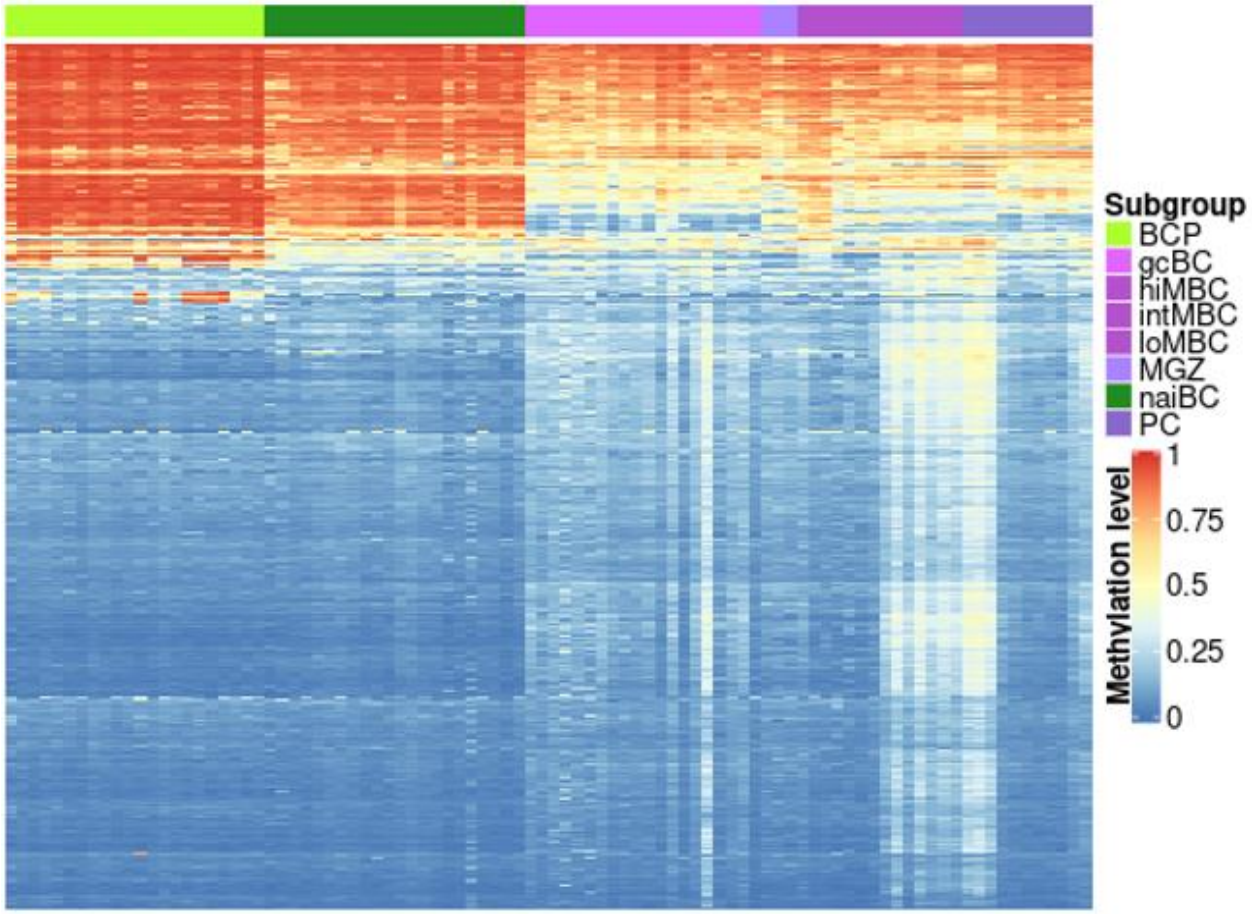
Supplementary Fig. 10. Demonstration of normal distribution of linear predictor scores (LPS) within subgroups. RS samples are displayed as red dots on the x-axis. Rationale for the 0.98 probability cut-off is found in the « unclassified zone » where the two normal distributions cross, between the dashed lines. Construction of the LPS is based on CLL vs DLBCL most differential CpGs, from which the IGHV signature has been subtracted. In grey and blue dots: application of the LPS on RS samples. Additionally, sample distribution is not a consequence of tumor purity (boxplots at the top-right side). Box plot: the center line, box limits, whiskers and points represent the mean, 25th and 75th percentile, 1.5x interquartile range and individual samples beyond the 1.5x interquartile, respectively. DLBCL-like RS group: n=13 independent samples; $_{low}$CLL-derived RS group: n=11 independent samples; $_{high}$CLL-derived RS group: n=32 independent samples. Source data are provided as a Source Data file. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; RS: Richter syndrome.



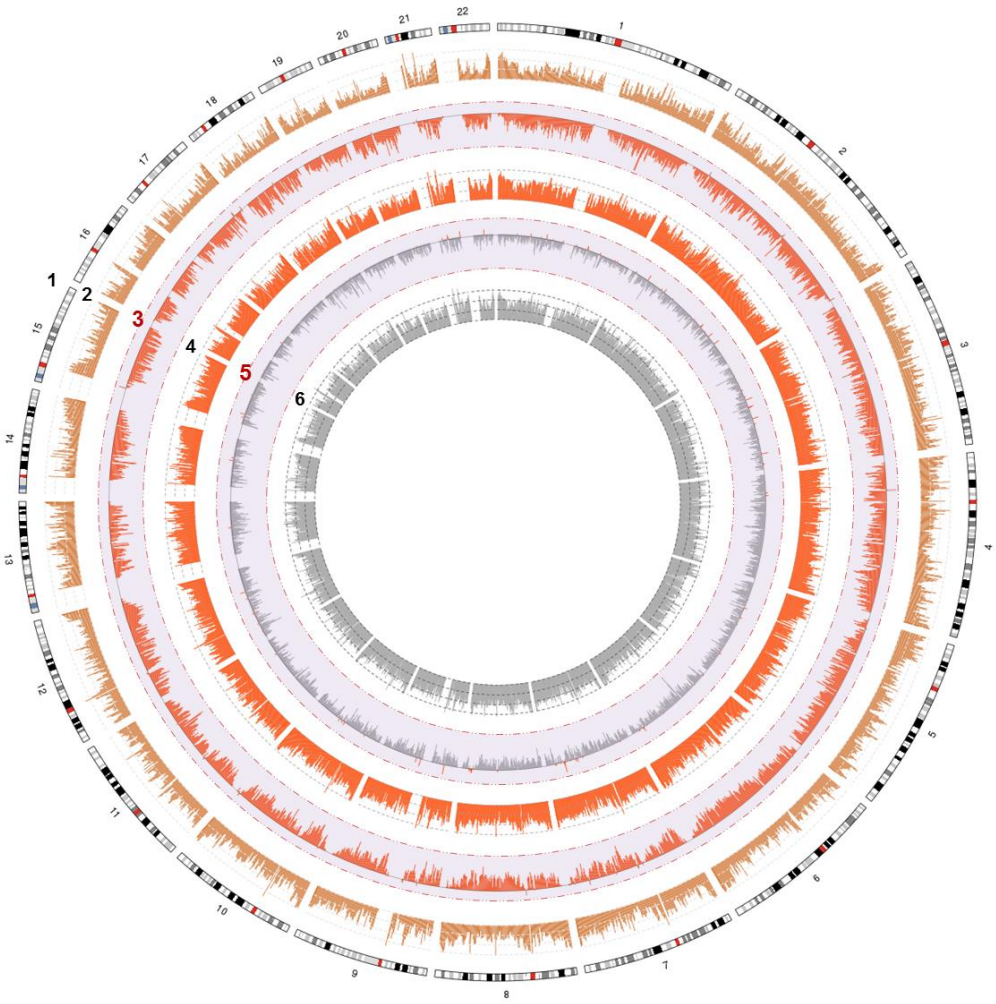
Supplementary Fig. 11. Non-negative matrix factorization (NMF) classification of RS samples correlates with grouping by DNA methylome LPS. Best sample classification is in 3 subgroups: rank 3 is the best fit as shown on the NMF rank survey (circles: test data; squares: best fit with test data; triangles: randomized data). On the figure, NMF clustering was performed with the 10,000 most differential CpGs from the EPIC dataset. Similar results have independently been obtained on every dataset considered (whether EPIC with 794,927 CpGs or FULL with 397,769 CpGs), with every CpGs, with the 10,000 most variant CpGs, or with the 4,863 LPS scoring CpGs (construction = CLLs + DLBCLs, test = RS). DLBCL-like RS: n=13; $_{low}$CLL-derived RS: n=12; $_{high}$CLL-derived RS: n=33. rho=-0.84; p<2e-16 (Spearman's correlation test). CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; LPS: linear predictor score; NMF: non-negative matrix factorization; RS: Richter syndrome.



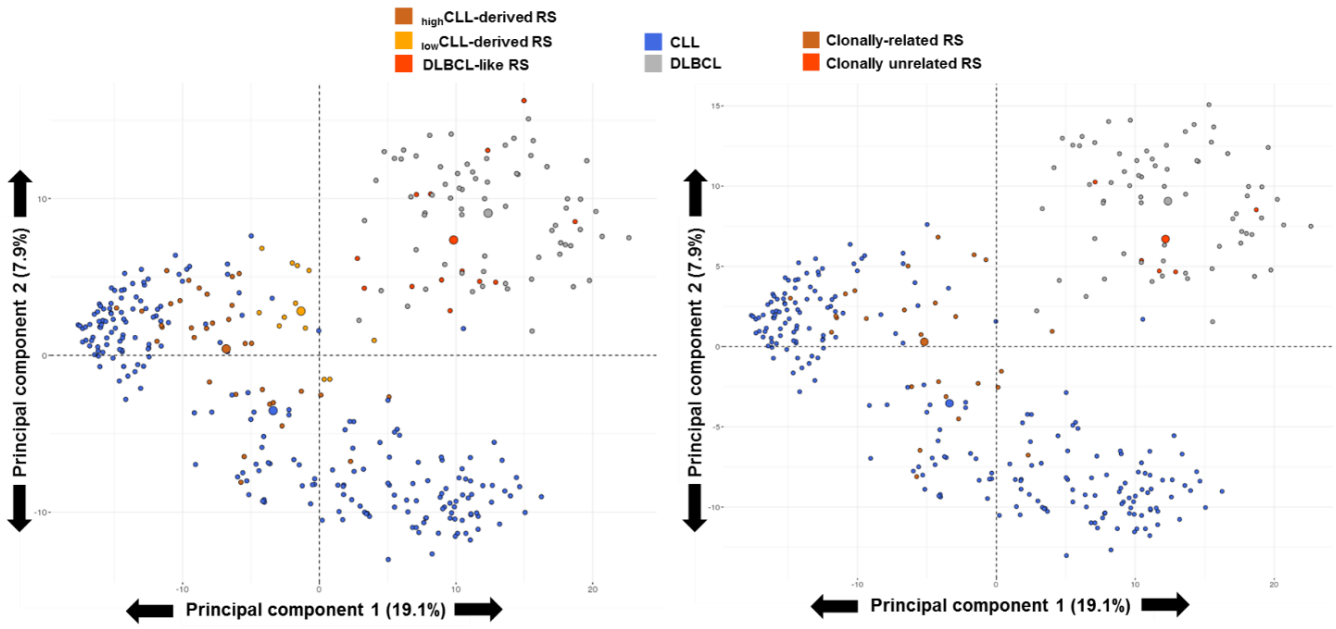
Supplementary Fig. 12. Unsupervised PCA of all RS samples using the 794,927 CpGs from the EPIC dataset. For each group, geometrical centers are displayed in bigger dots. Left panel: with RS subgroup annotations; Right panel: with *IGHV* mutational status. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; M-RS: IGHV-mutated RS; PCA: principal component analysis; RS: Richter syndrome; U-RS: IGHV-unmutated RS.



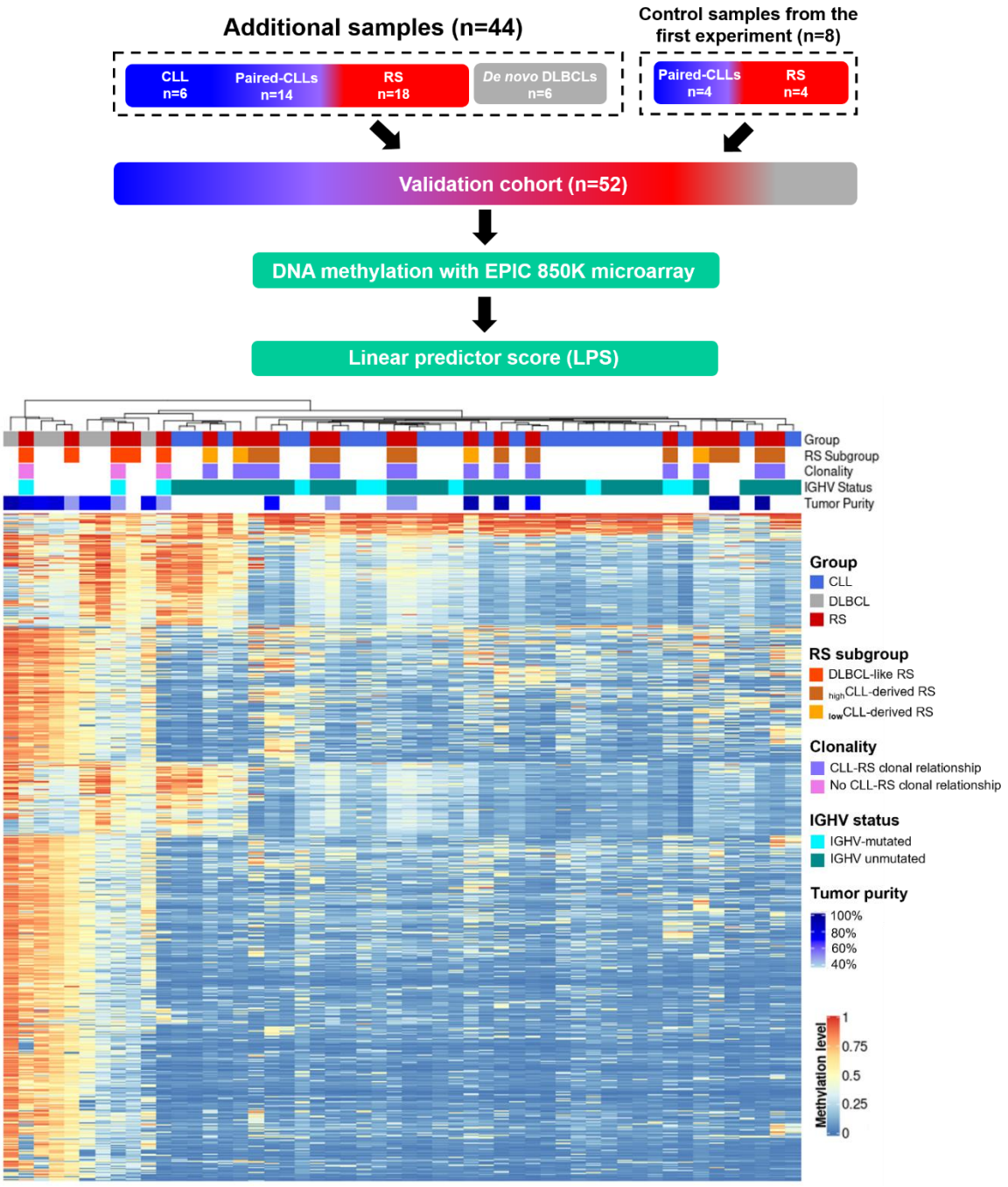
Supplementary Fig. 13. Methylation status of the 4,863 CpGs used by the linear classifier score (LCS) in normal B-cells. BCP: B-Cell precursors; gcBC: germinal center B-cells; MBC: memory B-cells; MGZ: marginal zone; naiBC: naive B-cells; PC: plasma cells.



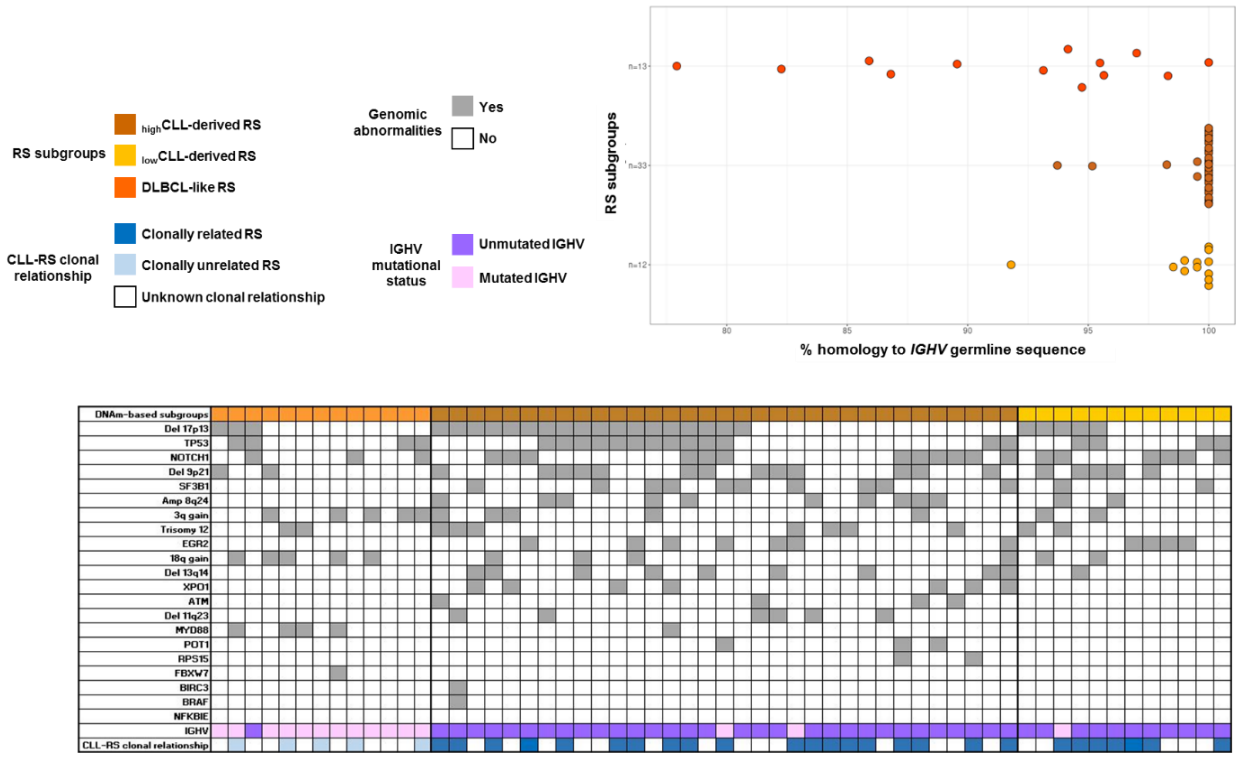
Supplementary Fig. 14. Genomic distribution of DNA methylation in $_{high}$CLL-derived RS (brown), DLBCL-like RS (orange) and DLBCLs (grey). Circular plot of median methylation levels and differences over sliding windows of 500 kb. Highlighted tracks 3 and 5 show the DNAm changes between $_{high}$CLL-derived, DLBCL-like RS and DLBCL, which are different in both extent and locations. From outer to inner track: 1) Chromosome number and ideograms with cytobands (X and Y chromosomes were removed from the analysis); 2) $_{high}$CLL-derived RS methylation levels; 3) $_{high}$CLL-derived minus DLBCL-like RS methylation differences (range -10 to +30%). Areas hypermethylated in DLBCL-like RS are represented with orange peaks pointing towards the DLBCL-like RS track; 4) DLBCL-like RS methylation levels; 5) DLBCL-like RS minus DLBCL methylation differences (range -19 to +7%). Areas hypermethylated in DLBCL-like RS are represented with orange peaks pointing toward the DLBCL-like RS track while areas hypermethylated in DLBCLs are represented with grey peaks pointing towards the DLBCL track; 6) DLBCL methylation levels. Dashed lines represent 0%, 25%, 50%, and 75% average methylation. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; RS: Richter syndrome.



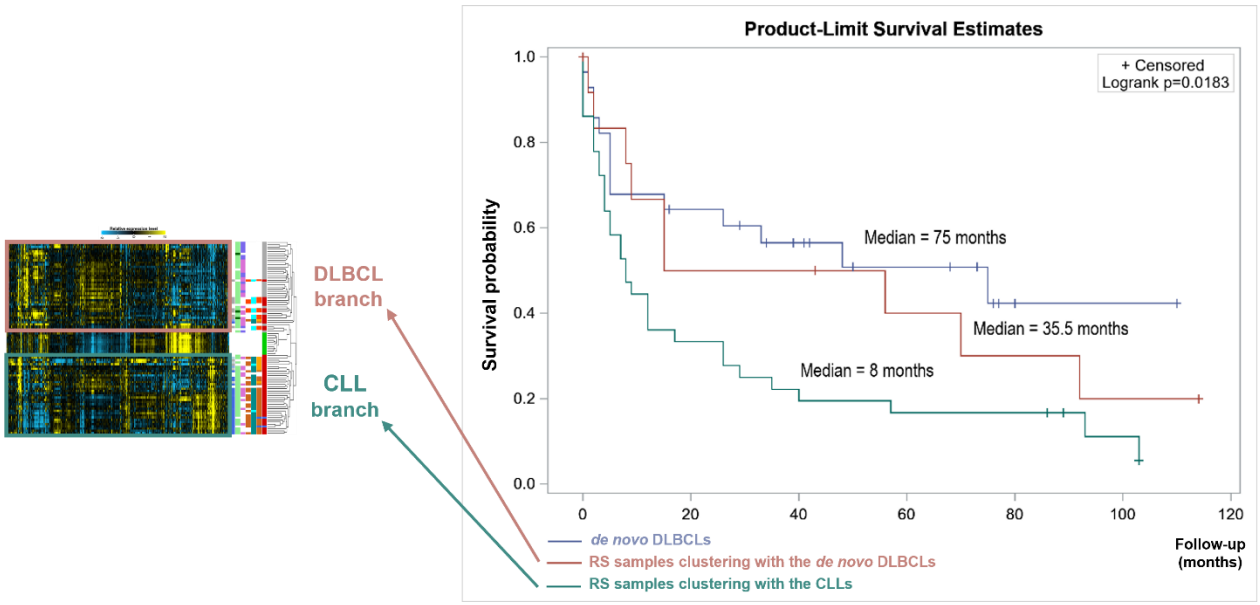
Supplementary Fig. 15. Sample partitioning according to IGHV mutational status. Additional data for Figure 3C, PCA on the 10,000 most variable CpGs of the dataset. Left panel: samples are tagged according to DNA methylation profile; Right panel: samples are tagged according to CLL-RS clonal relationship. RS samples with unknown clonality status are hidden. For each group, geometrical centers are displayed in bigger dots. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; PCA: principal component analysis; RS: Richter syndrome.



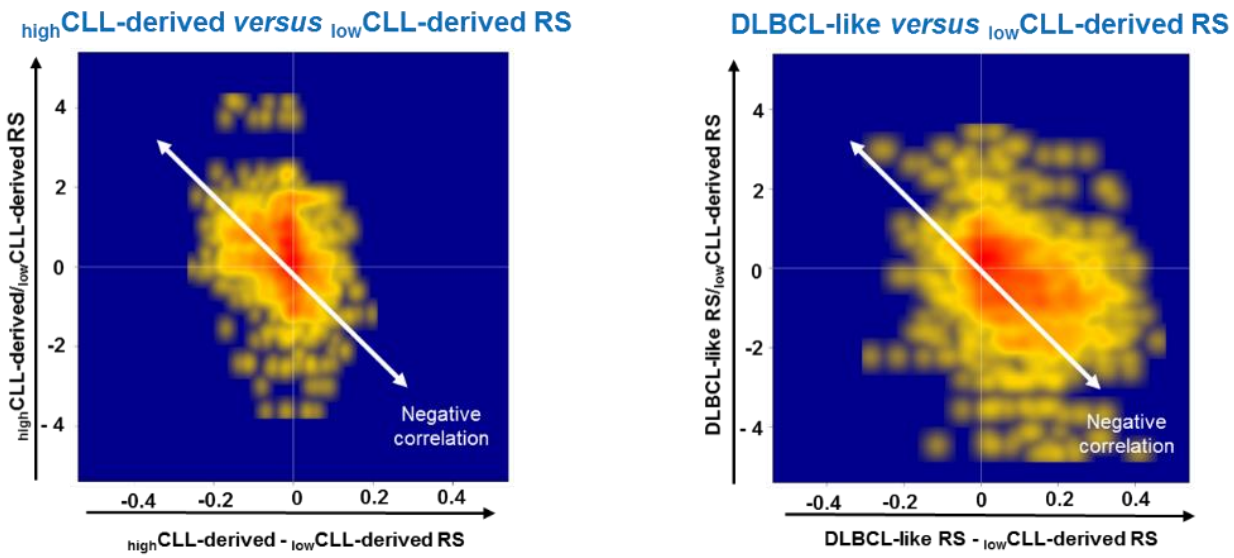
Supplementary Fig. 16. Hierarchical clustering of DNAm data retrieved from the validation cohort. This 52-sample validation cohort included 44 new samples: 18 new RS samples, the CLL component of 14 of these, 6 new DLBCL samples, and 6 additional CLLs. In addition, 8 samples from the training series were used as controls: 4 RS samples (3 clonally related and 1 clonally unrelated), with the 4 respective CLL components. RS sample classification according to LPS is displayed in the "RS Subgroup" annotations. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; LPS: linear predictor score; RS: Richter syndrome.



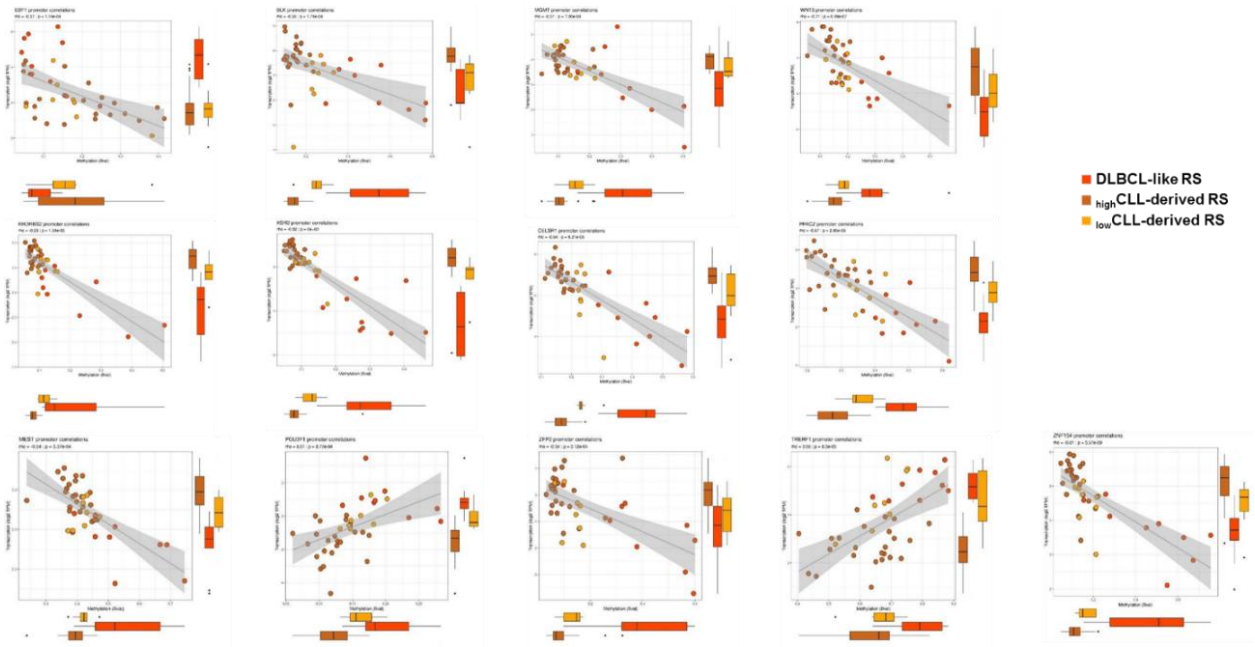
Supplementary Fig. 17. Annotation of RS DNA methylation subgrouping with a panel of genomic abnormalities frequently described in CLL and RS. CLL: chronic lymphocytic leukemia; RS: Richter syndrome.



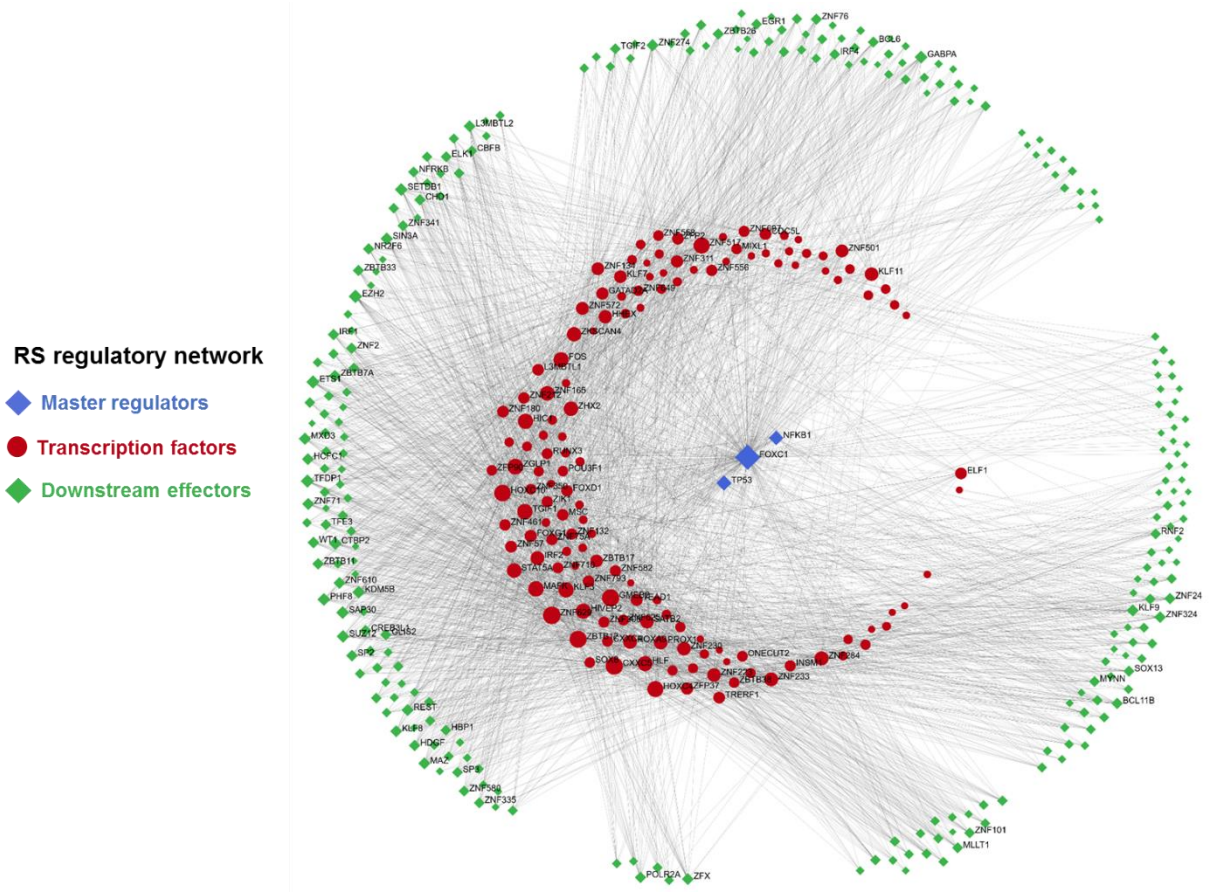
Supplementary Fig. 18. CLL-derived RS are associated with a worse prognosis than DLBCL-like RS and *de novo* DLBCL. Kaplan-Meier estimates of overall survival for 77 patients. Statistical comparisons were performed with Log-rank test. Bonferroni method was used for multitesting adjustment. CLL: chronic lymphocytic leukemia; DLBCL: diffuse large B-cell lymphoma; RS: Richter syndrome.



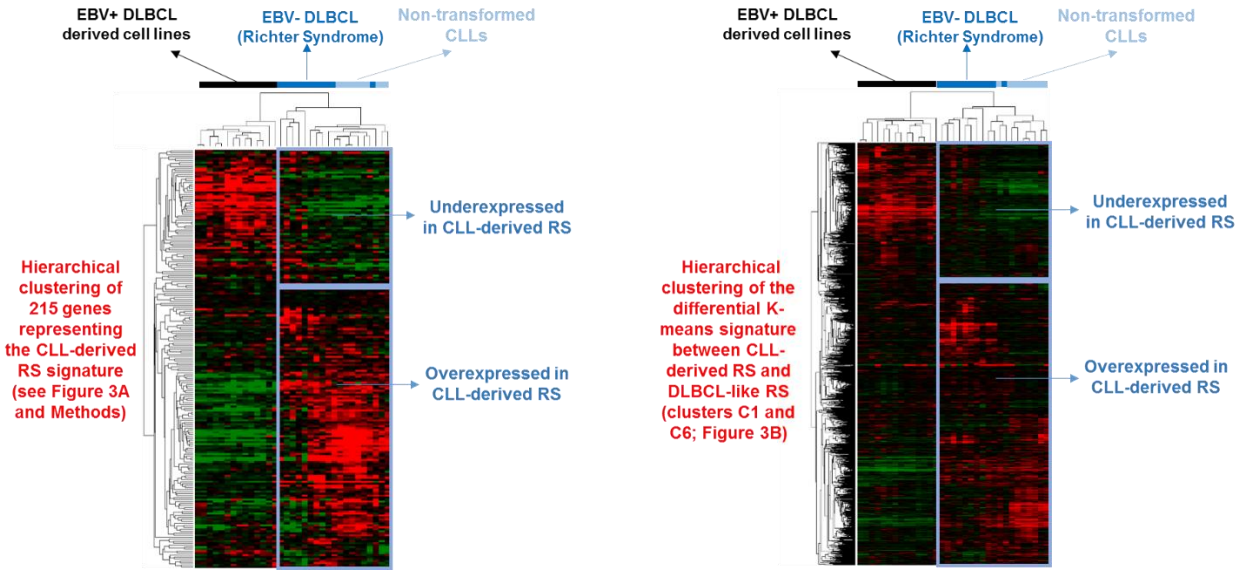
Supplementary Fig. 19. Density maps (smoothed density scatterplots) representing overall DNA methylation vs gene expression changes. Scale ranges from blue (no density), to yellow (medium density) and red (high density). Only genes with at least 1 significant correlation (cutoff p-values < 0.01) were retained. The exact same limits for x and y axes are chosen for each plot. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; RS: Richter syndrome.



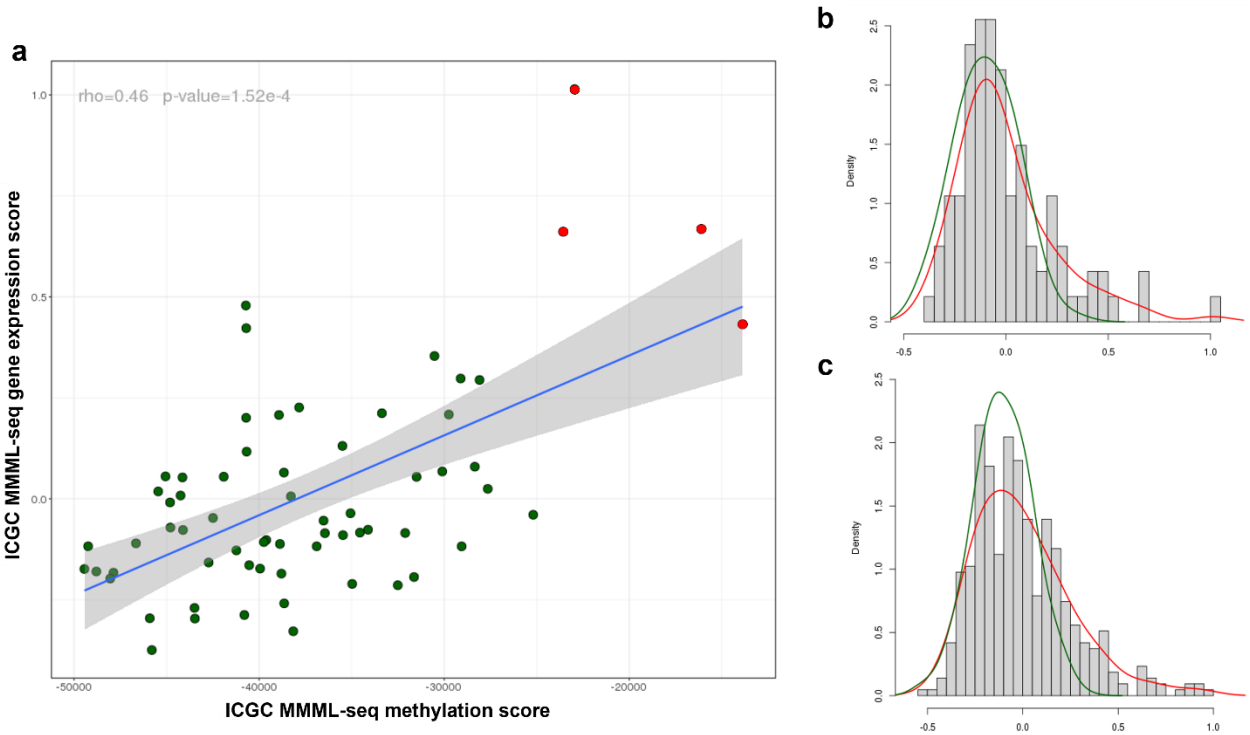
Supplementary Fig. 20. Correlation scatterplots between DNA methylation and gene expression values, viewed for a selection of genes. X-axis: DNA methylation average beta-value for negatively or positively correlating regulatory regions (see main text and Methods); y-axis: transcriptional activity of the corresponding gene, in transcript per million (TPM, normalized expression). Boxplots are displayed at the bottom and on the right hand side for methylation and transcription summaries, respectively. DLBCL-like RS group: n=9 independent samples; $_{low}$CLL-derived RS group: n=8 independent samples; $_{high}$CLL-derived RS group: n=24 independent samples. For all box plots, the center line, box limits, whiskers and points represent the median, 25th and 75th percentile, 1.5x interquartile range and individual samples beyond the 1.5x interquartile, respectively. Bval: Beta-value; CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; RS: Richter syndrome; TPM: transcript per million.



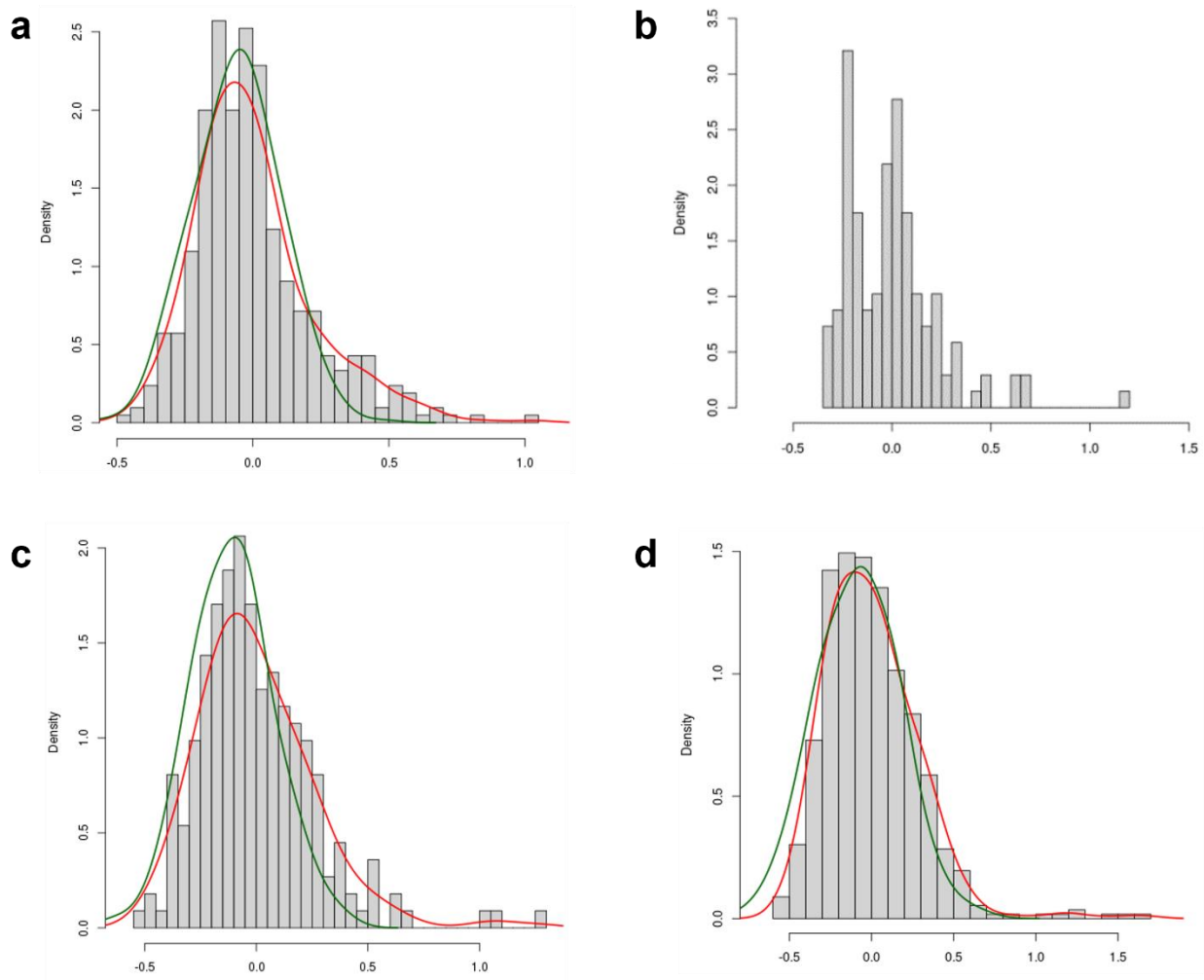
Supplementary Fig. 21. Regulatory network inferred from the set of 156 transcription factors correlating between methylome and transcriptome. These TFs are used as seeds for reconstructing the network (red dots, 1st line). Upstream, they are controlled by three master regulators (TP53, NFKB1 and FOXC1, blue dots). In turn, they regulate a large set of downstream TF effectors (green dots; 2nd line). TF interactions were computed from ENCODE and JASPAR databases. TF: transcription factor.



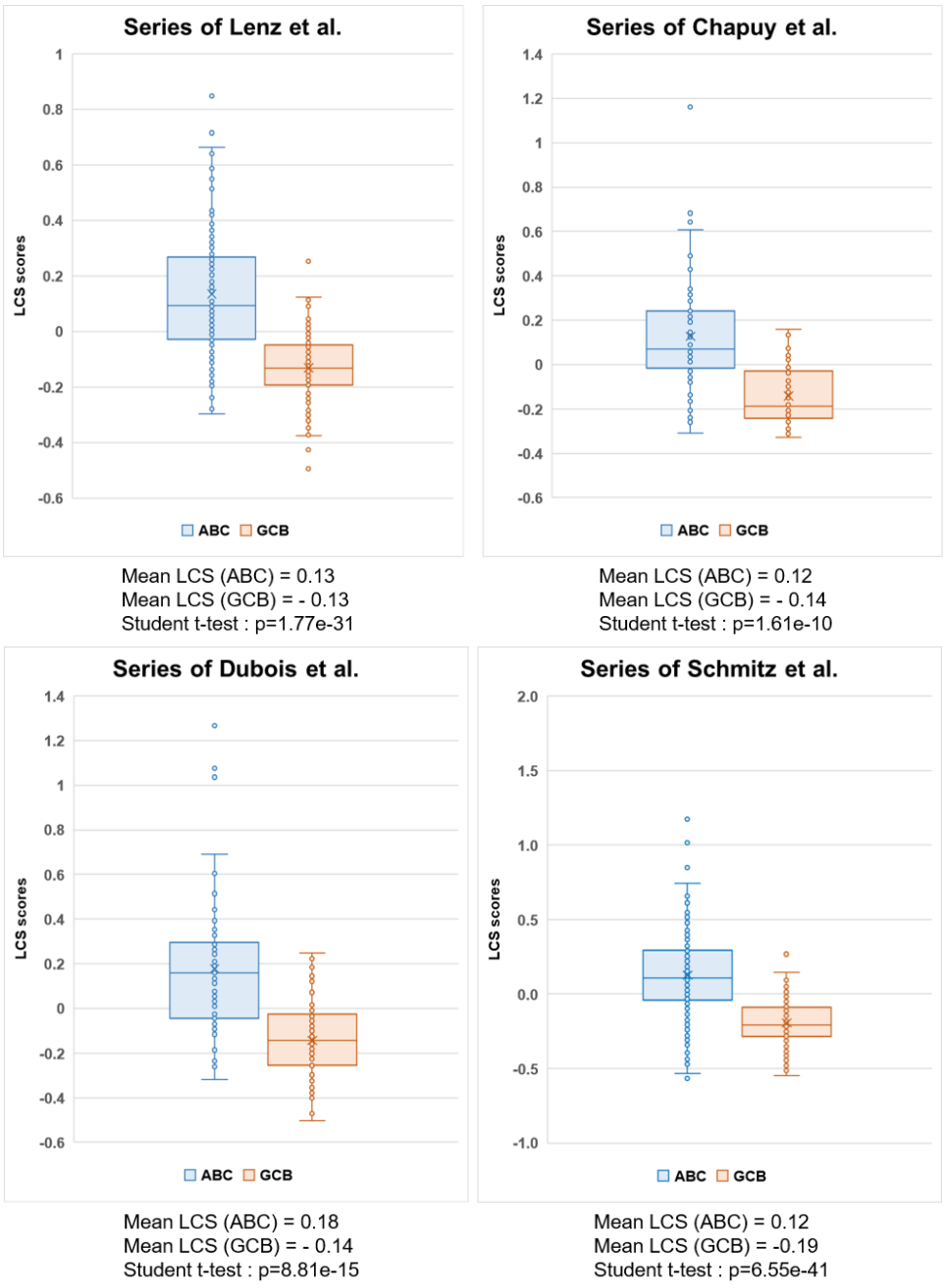
Supplementary Fig. 22. Validation of the CLL-derived RS gene signature on a DLBCL cohort including RS samples. From the public dataset GSE103265 available from the Gene Expression Omnibus database. CLL: chronic lymphocytic leukemia; DLBCL: diffuse large B-cell lymphoma; EBV: Epstein-Barr virus; RS: Richter syndrome.



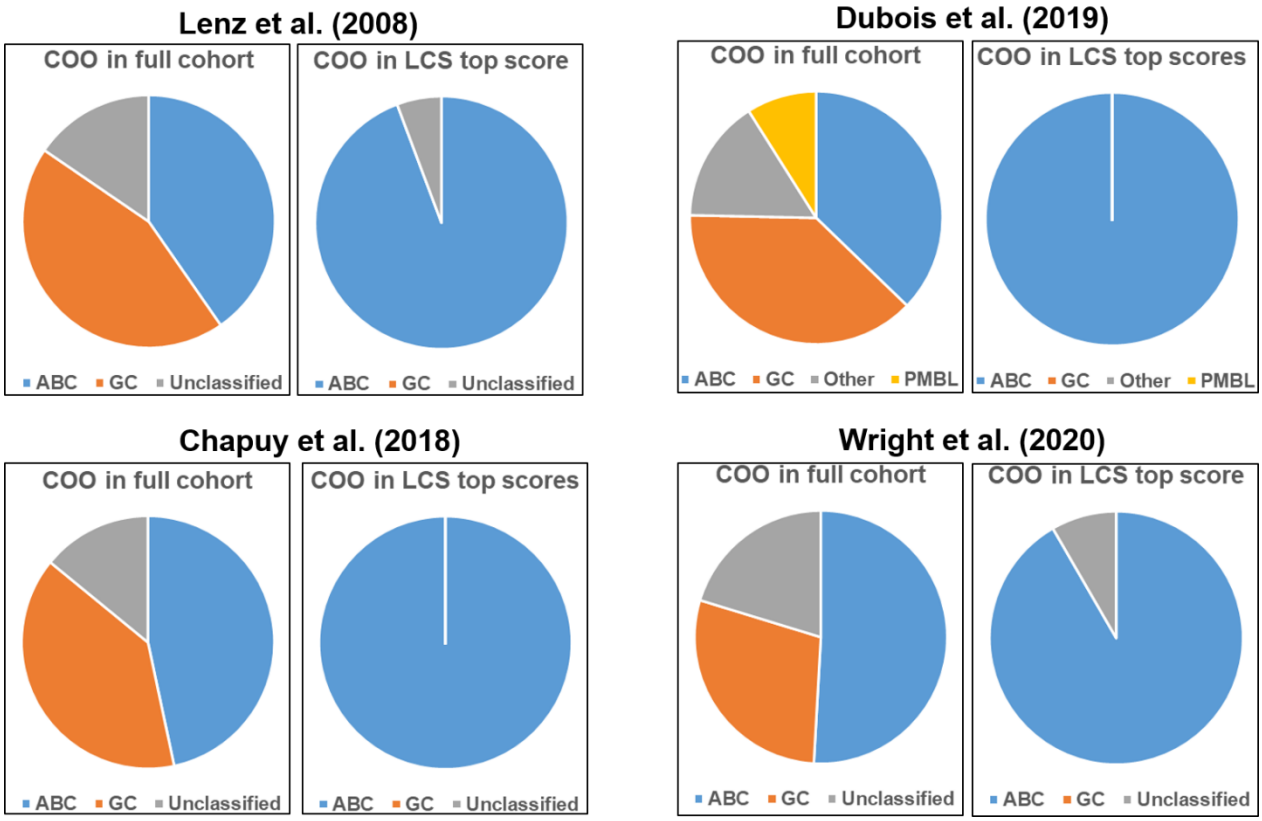
Supplementary Fig. 23. a Correlation between methylome and transcriptome scores (LPS and LCS, respectively; Spearman's test) in the ICGC MMML-seq dataset (scatterplot, linear modelling fit with 5% confidence interval). Samples with extreme scores in both DNAm and gene expression are labeled in red. rho=0.46, p=1.52e-4. **b** LCS densities for the ICGC MMML-seq dataset (n=94); **c)** LCS densities for MMML dataset (n=430). **b** and **c** Red curve: observed distribution (lowess-smoothed); Green curve: simulated/expected distribution from the real observations (lowess-smoothed). ICGC: International Cancer Genome Consortium; LCS: linear classifier score (from gene expression); LPS: linear predictor score (from DNA methylation); MMML: Molecular Mechanisms in Malignant Lymphoma.



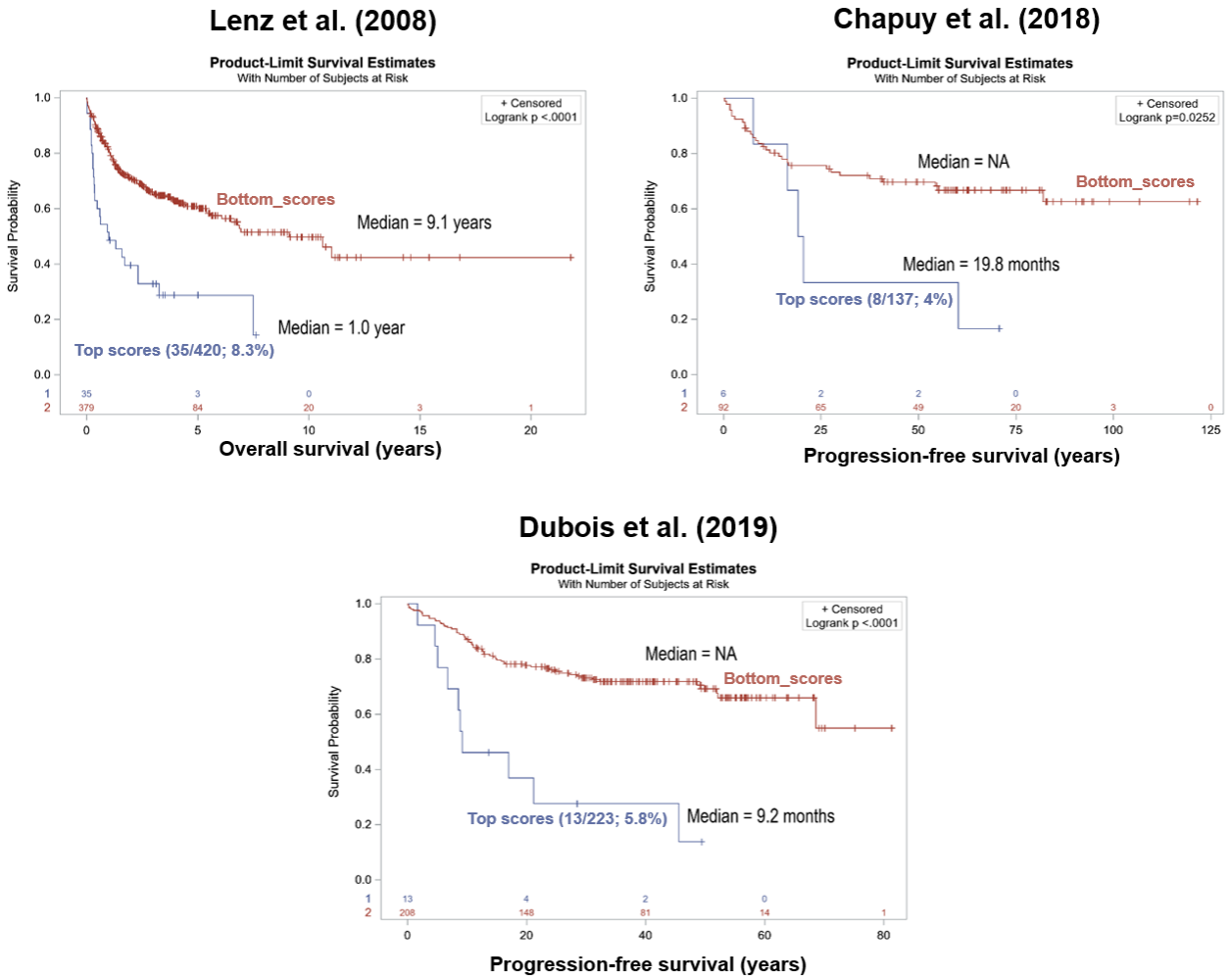
Supplementary Fig. 24. Linear classifier score distributions for all other studied datasets. Histograms of LCS from: **a** Lenz dataset (n=420; microarray, accession under GSE10846; PMID: 21546504); **b** Chapuy dataset (n=137; microarray, accession under GSE98588; PMID: 29713087); **c** Dubois dataset (n=223; microarray, accession under GSE87371; PMID: 31648986); **d** Wright dataset (n=562; RNA-Seq; PMID: 32289277). Red curve: observed distribution (lowess-smoothed); Green curve: simulated/expected distribution from the real observations (lowess-smoothed). LCS: linear classifier score.



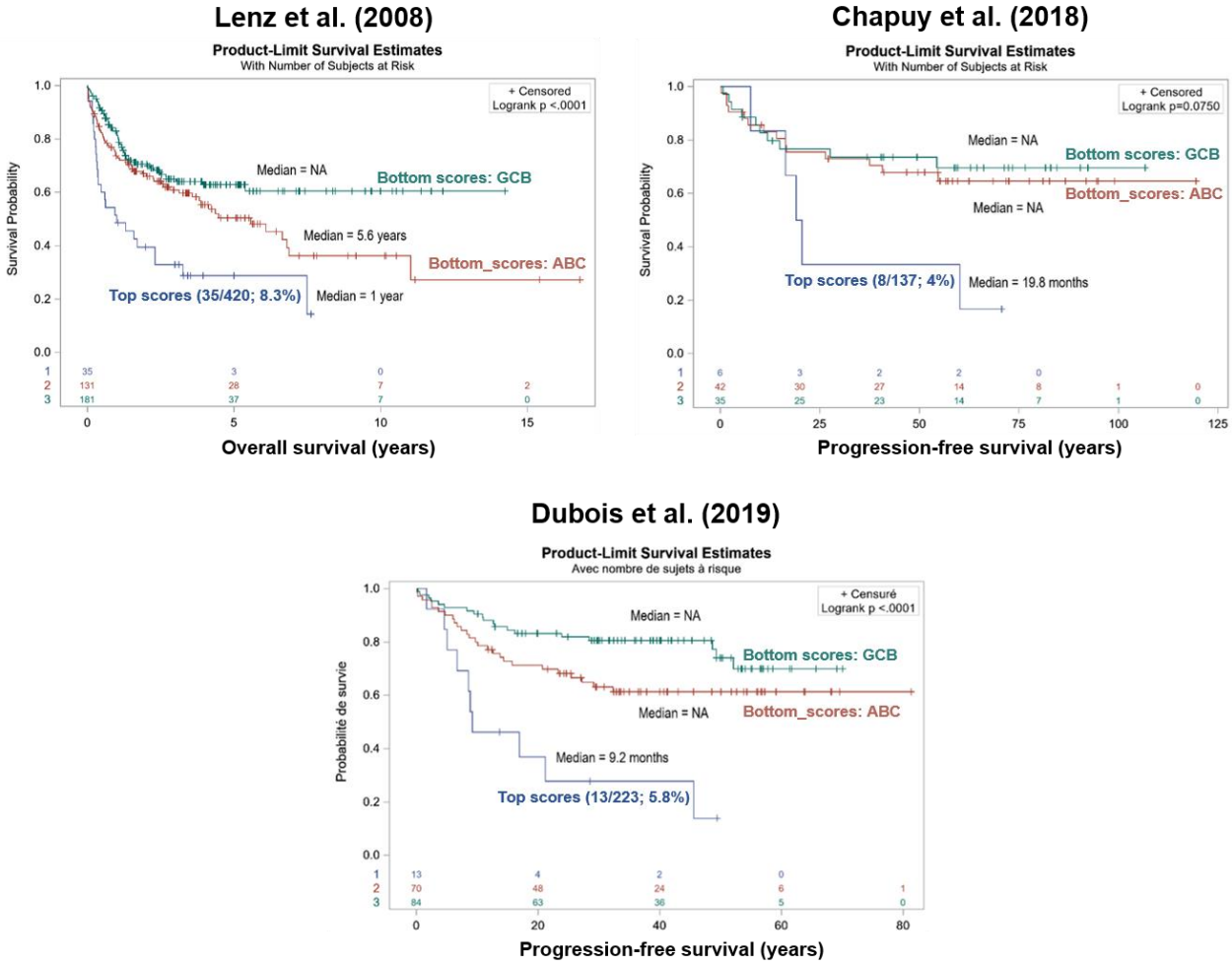
Supplementary Fig. 25. Distribution of LCS scores among ABC- and GCB-subtype DLBCLs in 4 large series. Series from Lenz et al: n=167 independent ABC subtype DLBCL cases and n= 183 independent GCB subtype DLBCL cases. Series from Chapuy et al: n=63 independent ABC subtype DLBCL cases and n=53 independent GCB subtype DLBCL cases. Series from Dubois et al: n=83 independent ABC subtype DLBCL cases and n=85 independent GCB subtype DLBCL cases. Series from Schmidt et al: n=286 independent ABC subtype DLBCL cases and n=162 independent GCB subtype DLBCL cases. Box plots: the center line, box limits, whiskers and points represent the median, 25th and 75th percentile, 1.5x interquartile range and individual samples, respectively. Two-sided Student t-test. Source data are provided as a Source Data file. ABC: activated B-cell like; GCB: germinal center B-cell like; LCS: linear classifier score.



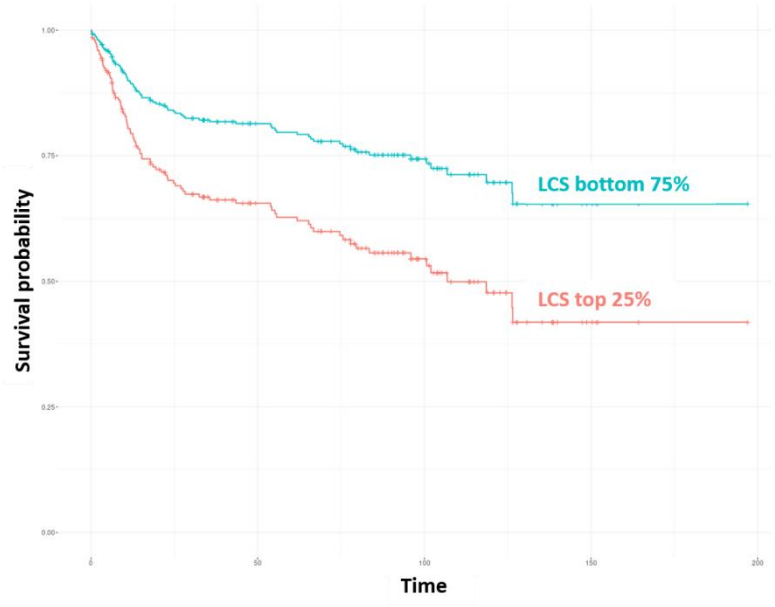
Supplementary Fig. 26. Distribution of COO in the entire cohorts and among the top-LCS samples in 4 datasets. ABC: activated B-cell like; COO: cell-of-origin; GCB: germinal center B-cell like; LCS: linear classifier score.



Supplementary Fig. 27. Kaplan-Meier estimates of overall survival or progression-free survival for 733 patients from three DLBCL datasets comparing patients with top-LCS and the rest of the cohorts. Series from Lenz and colleagues: n=420 patients; p=3.5e-7; series from Chapuy and colleagues: n=137 patients; p=0.0252; series from Dubois and colleagues: n=223 patients; p=3.2e-6. Statistical comparisons were performed with the log-rank test.



Supplementary Fig. 28. Kaplan-Meier estimates of overall survival or progression-free survival for 733 patients from three DLBCL datasets. Comparison between patients with top LCS and the rest of the cohorts, according to COO. Series from Lenz and colleagues: n=420 patients; p=5.2e-6; series from Chapuy and colleagues: n=137 patients; p=0.075; series from Dubois and colleagues: n=223 patients; p=4.2e-6. Statistical comparisons were performed with the log-rank test. Bonferroni method was used for multitesting adjustments. COO: cell-of-origin; DLBCL: *de novo* diffuse large B-cell lymphoma.



Cox PH multivariate statistics for PFS, variable = LCS, covariates = IPI + TP53 + DoubleHit (LR, Wald & logrank tests all < 1e-7)

	Beta	HR	P-Value	95% CI [LL, UL]	
LCS	0,4421	1,556	8,16E-05	[1.2488,1.939]	***
IPIcategory	0,7628	2,1442	2,11E-06	[1.5645,2.939]	***
TP53	0,41	1,5068	0,275	[0.7218,3.146]	
DoubleHit	-0,19	0,8269	0,569	[0.4296,1.592]	

Cox PH multivariate statistics for OS, variable = LCS, covariates = IPI + TP53 + DoubleHit (LR, Wald & logrank tests all < 9e-7)

	Beta	HR	P-Value	95% CI [LL, UL]	
LCS	0,36538	1,44106	0,00338	[1.1287,1.84]	**
IPIcategory	0,8426	2,32241	9,26E-07	[1.6587,3.252]	***
TP53	0,01682	1,01696	0,96862	[0.44,2.351]	
DoubleHit	-0,23064	0,79402	0,52805	[0.3879,1.625]	

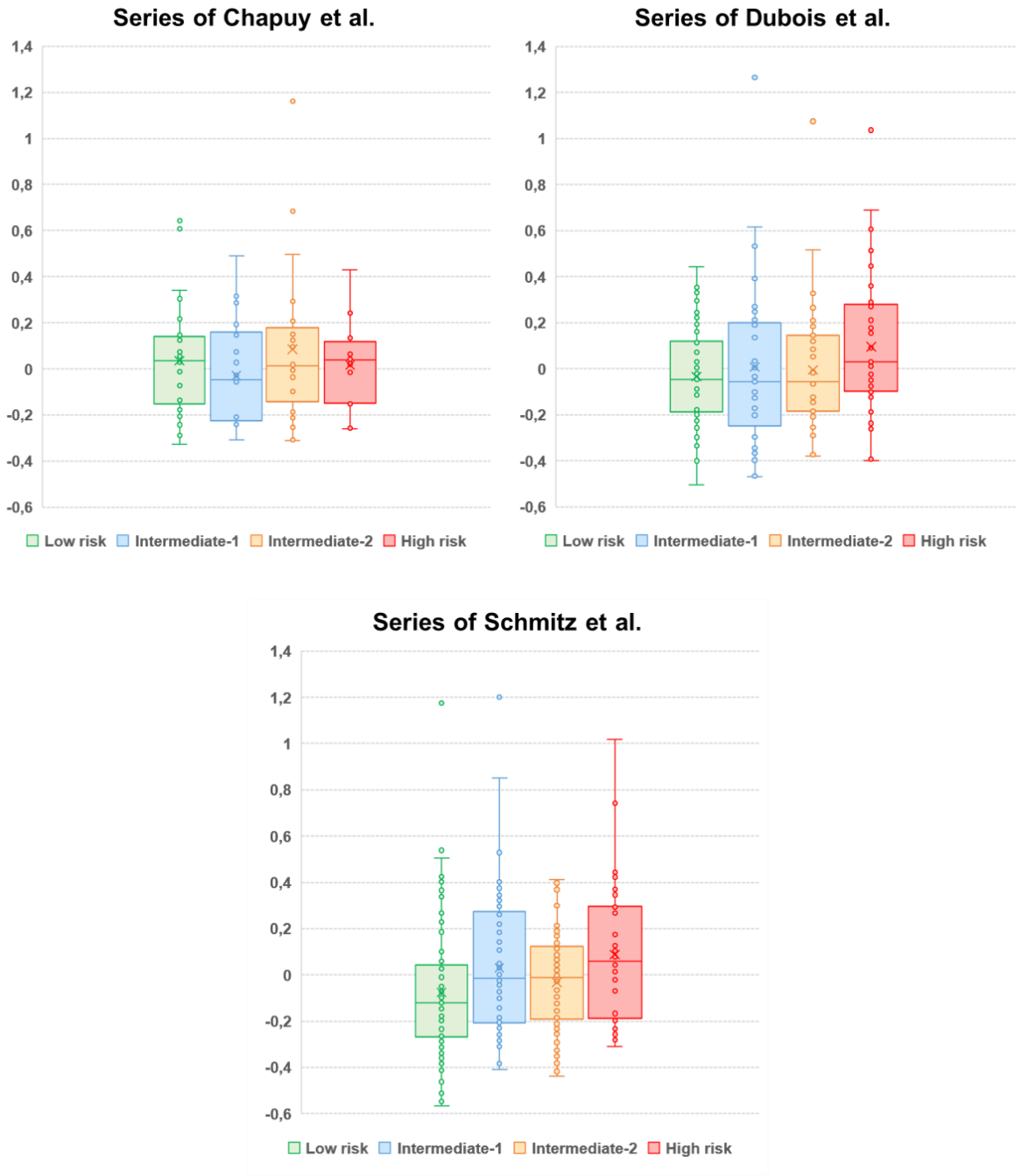
Cox PH multivariate statistics for PFS, variable = top25, covariates = IPI + TP53 + DoubleHit (LR, Wald & logrank tests all < 2e-6)

	Beta	HR	P-Value	95% CI [LL, UL]	
top25	0,7127	2,0395	0,00327	[1.2683,3.279]	**
IPIcategory	0,7805	2,1826	8,68E-07	[1.5993,2.979]	***
TP53	0,3648	1,4402	0,33169	[0.6895,3.008]	
DoubleHit	-0,1739	0,8404	0,59978	[0.4389,1.609]	

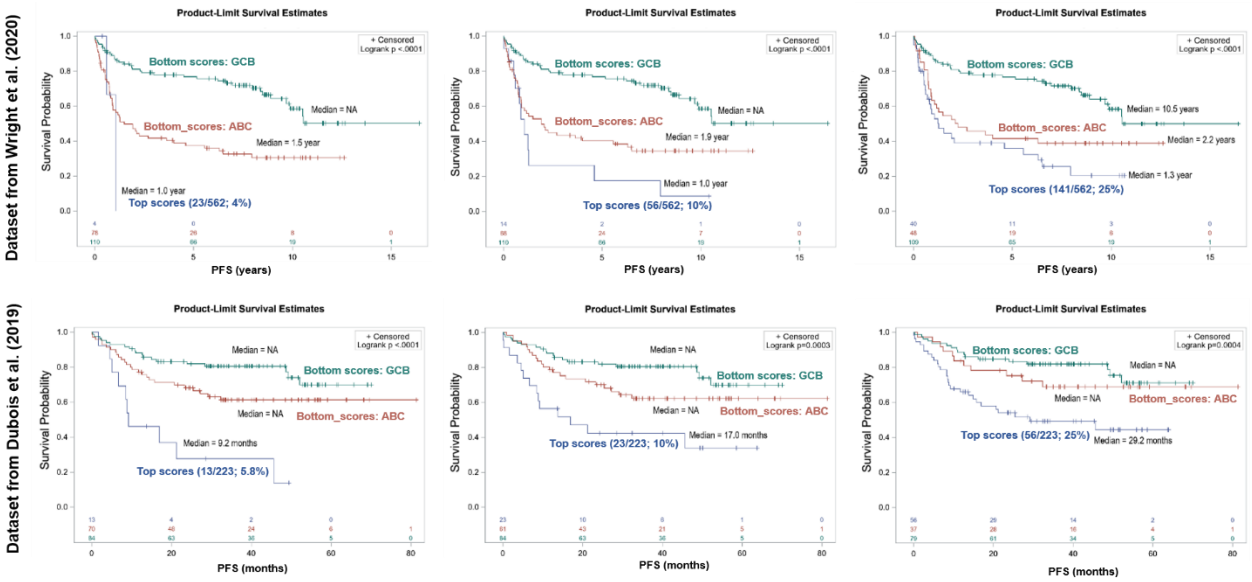
Cox PH multivariate statistics for OS, variable = top25, covariates = IPI + TP53 + DoubleHit (LR, Wald & logrank tests all < 5e-6)

	Beta	HR	P-Value	95% CI [LL, UL]	
top25	0,587363	1,799238	0,0265	[1.0708,3.023]	*
IPIcategory	0,861689	2,367155	4,48E-07	[1.694,3.308]	***
TP53	0,007843	1,007874	0,9854	[0.4353,2.334]	
DoubleHit	-0,212922	0,808219	0,5566	[0.3974,1.644]	

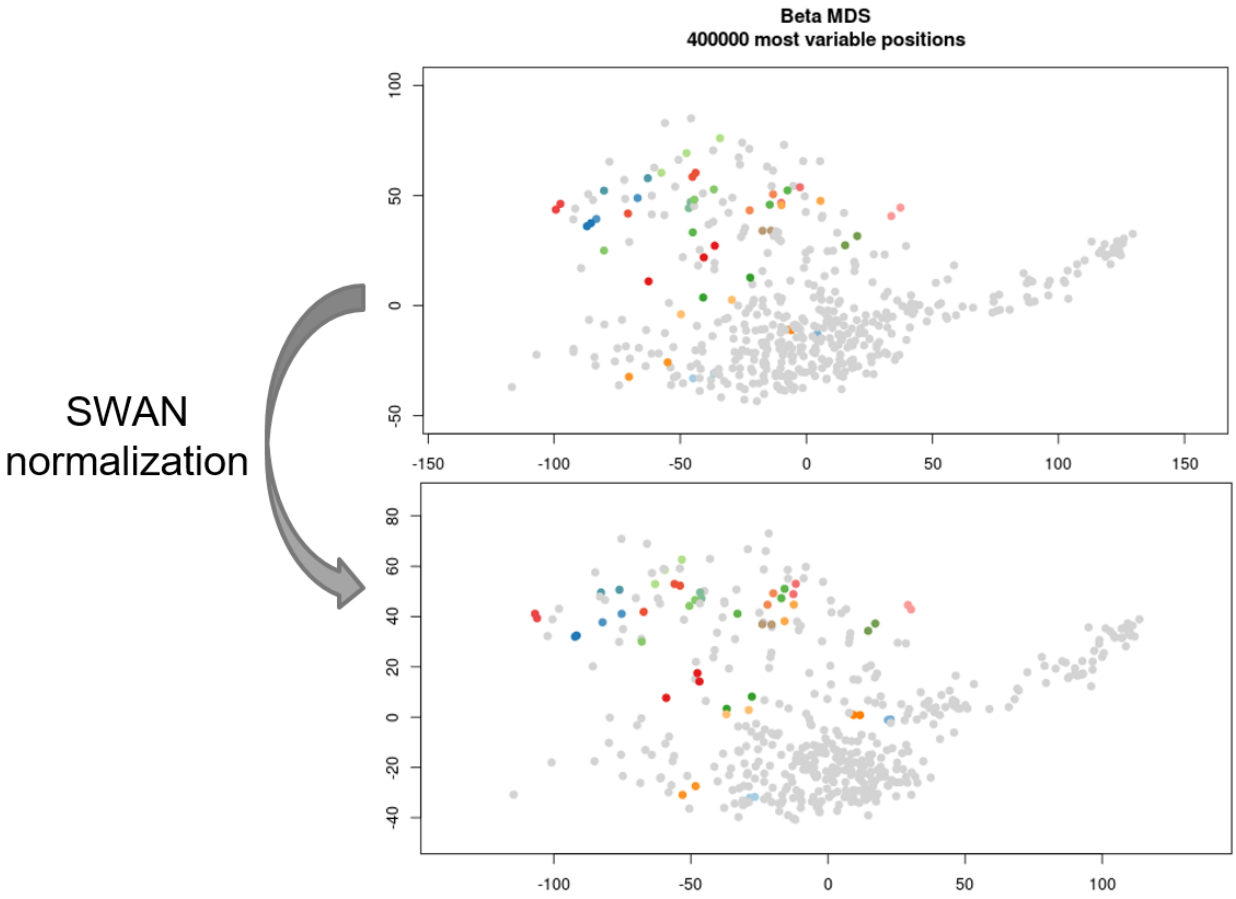
Supplementary Fig. 29. Multivariate analysis with a Cox proportional hazards model, including all available informative covariates to evaluate the association of LCS with survival (OS and PFS). This association was calculated in binary (top 25% LCS versus the rest, tagged LCS in the tables) as well as linear (LCS as a continuous variable, tagged Zscore in the table) modes. Cox proportional-hazards multivariate models: (i) PFS model includes LCS, IPI, *TP53* abnormalities, and *MYC/BCL2* double hit (Wald = 1e-8); (ii) OS model includes LCS, IPI, *TP53* abnormalities and *MYC/BCL2* double hit (Wald = 5e-7). Statistical tests (Wald, LR) are two-tailed. CI: confidence interval, HR: hazard ratio; LCS: linear classifier score; LL: lower limit; LR: likelihood ratio; UL: upper limit. *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.



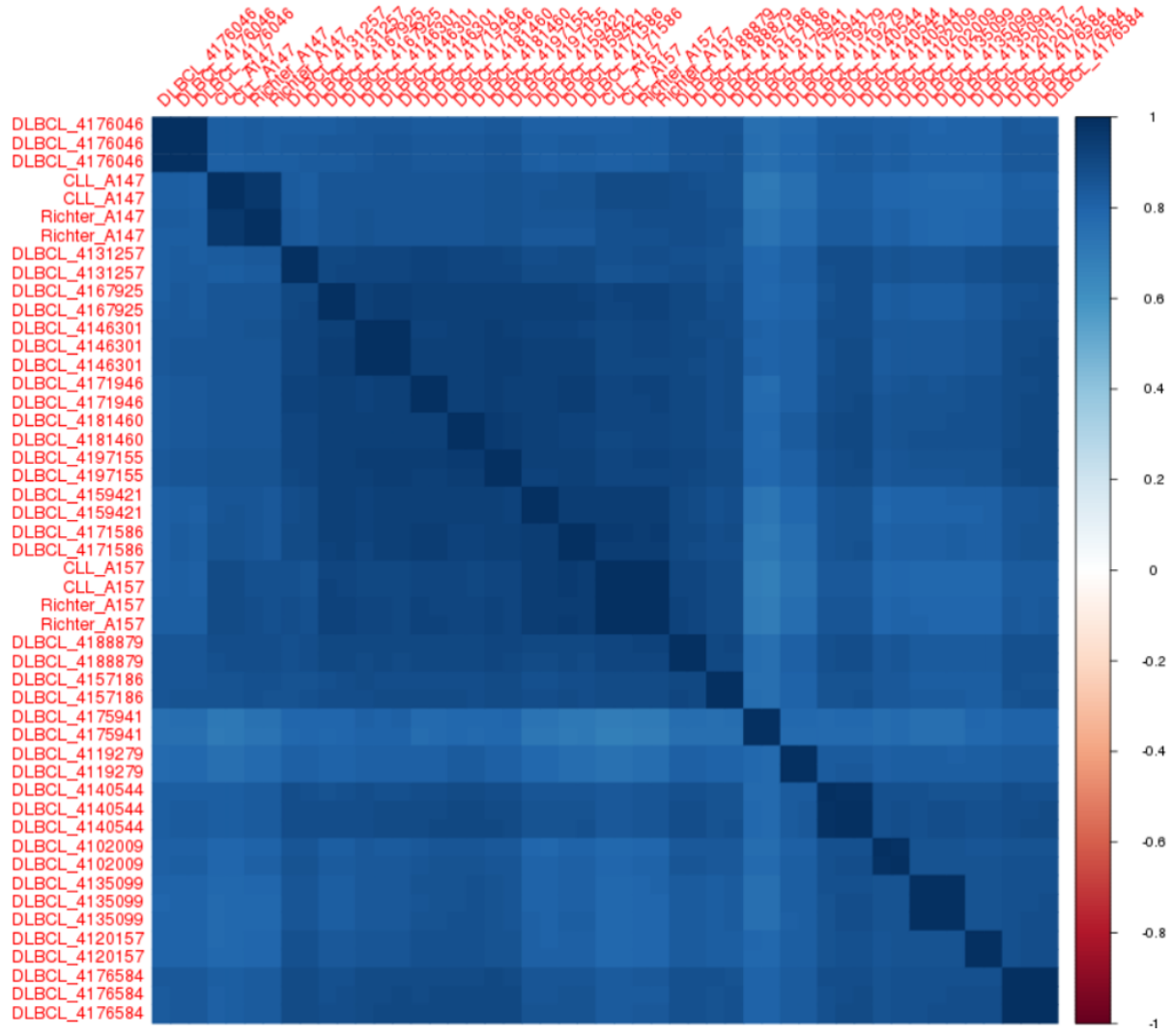
Supplementary Fig. 30. Distribution of LCS scores among prognostic groups according to international prognostic index in three large DLBCL series. Series from Chapuy et al: n=29 independent low-risk DLBCLs, n=22 independent intermediate-1 DLBCLs, n=21 independent intermediate-2 DLBCLs, and n=16 independent high-risk DLBCLs. Series from Dubois et al: n=51 independent low-risk DLBCLs, n=29 independent intermediate-1 DLBCLs, n=43 independent intermediate-2 DLBCLs, and n=45 high-risk DLBCLs. Series from Schmitz et al: n=85 independent low-risk DLBCLs, n=72 independent intermediate-1 DLBCLs, n=66 independent intermediate-2 DLBCLs, and n=43 independent high-risk DLBCLs. Box plots: the center line, box limits, whiskers and points represent the median, 25th and 75th percentile, 1.5x interquartile range and individual samples, respectively. Two-sided Student t-test. Source data are provided as a Source Data file.



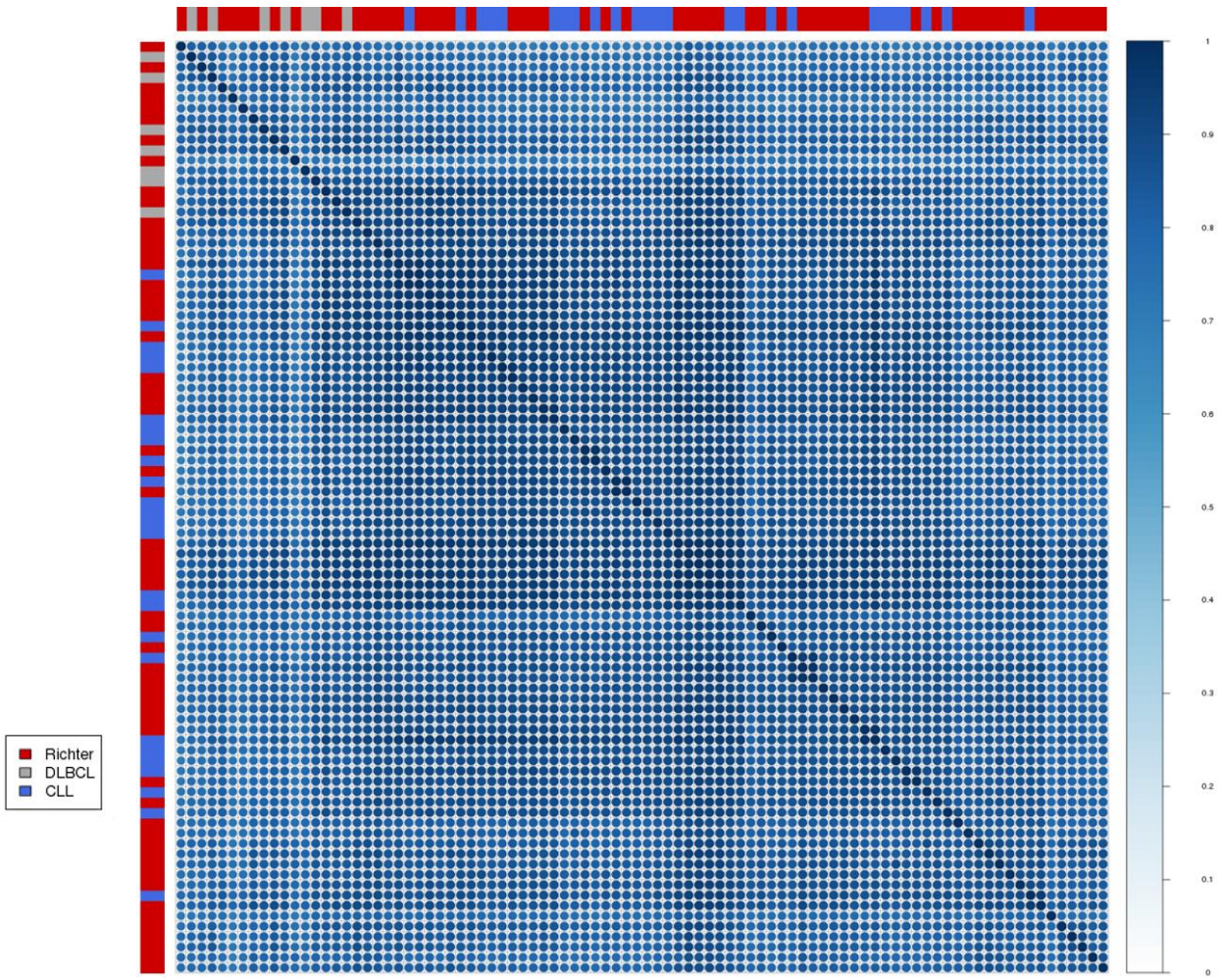
Supplementary Fig. 31. Kaplan-Meier estimates of PFS for 785 patients from two combined DLBCL datasets with clinical annotations. Comparative PFS between groups of patients determined according to progressive thresholds with top-LCS RS samples and the rest of the cohorts, according to COO. Upper panel: dataset from Wright and colleagues (PMID: 32289277): n=562 patients; p=1.2e-6 (top 25 LCS); p=1.9e-7 (top 10% LCS); p=8.5e-7 (top 25% LCS). Lower panel: dataset from Dubois and colleagues (PMID: 31648986): n=223 patients; p=1.6e-5 (top 25 LCS); p=3e-4 (top 10% LCS); p=4e-4 (top 25% LCS). Statistical comparisons were performed with Log-rank test. Bonferroni method was used for multitesting adjustments. COO: cell-of-origin; DLBCL: *de novo* diffuse large B-cell lymphoma. ABC: activated B-cell; COO: cell of origin; GCB: germinal center B-cell; LCS: linear classifier score; PFS: progression-free survival; RS: Richter syndrome.



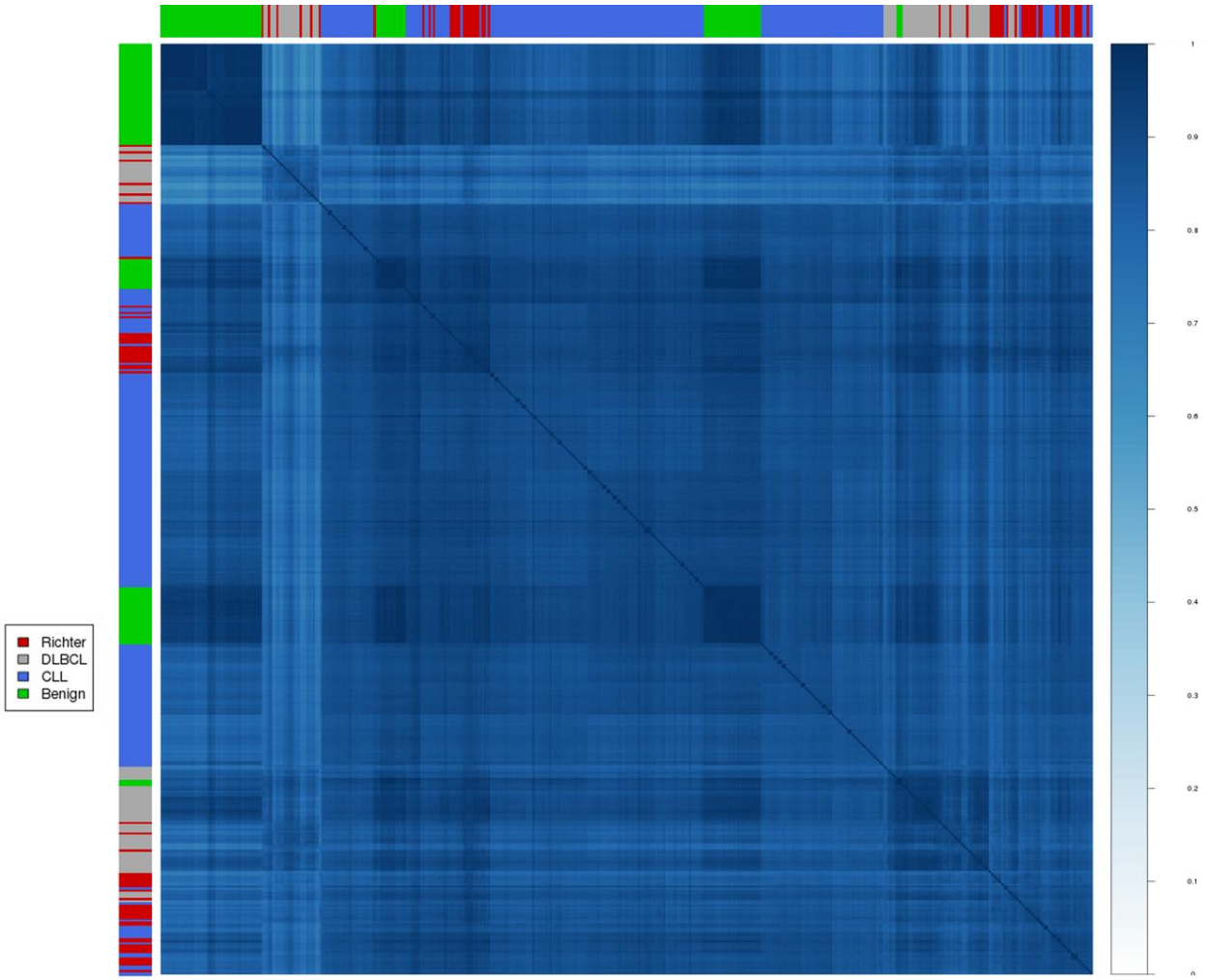
Supplementary Fig. 32. Multi-dimensional scaling before and after normalization of the FULL dataset. Color emphasis on the 7 *de novo* DLBCL replicates between French and German facilities, and 2 RS and 2 CLL replicates between French batches. Upper panel: before normalization, showing a small shift between many replicates, always in the same direction. Lower panel: after normalization, showing the correction of the shift, as seen here between replicate samples.



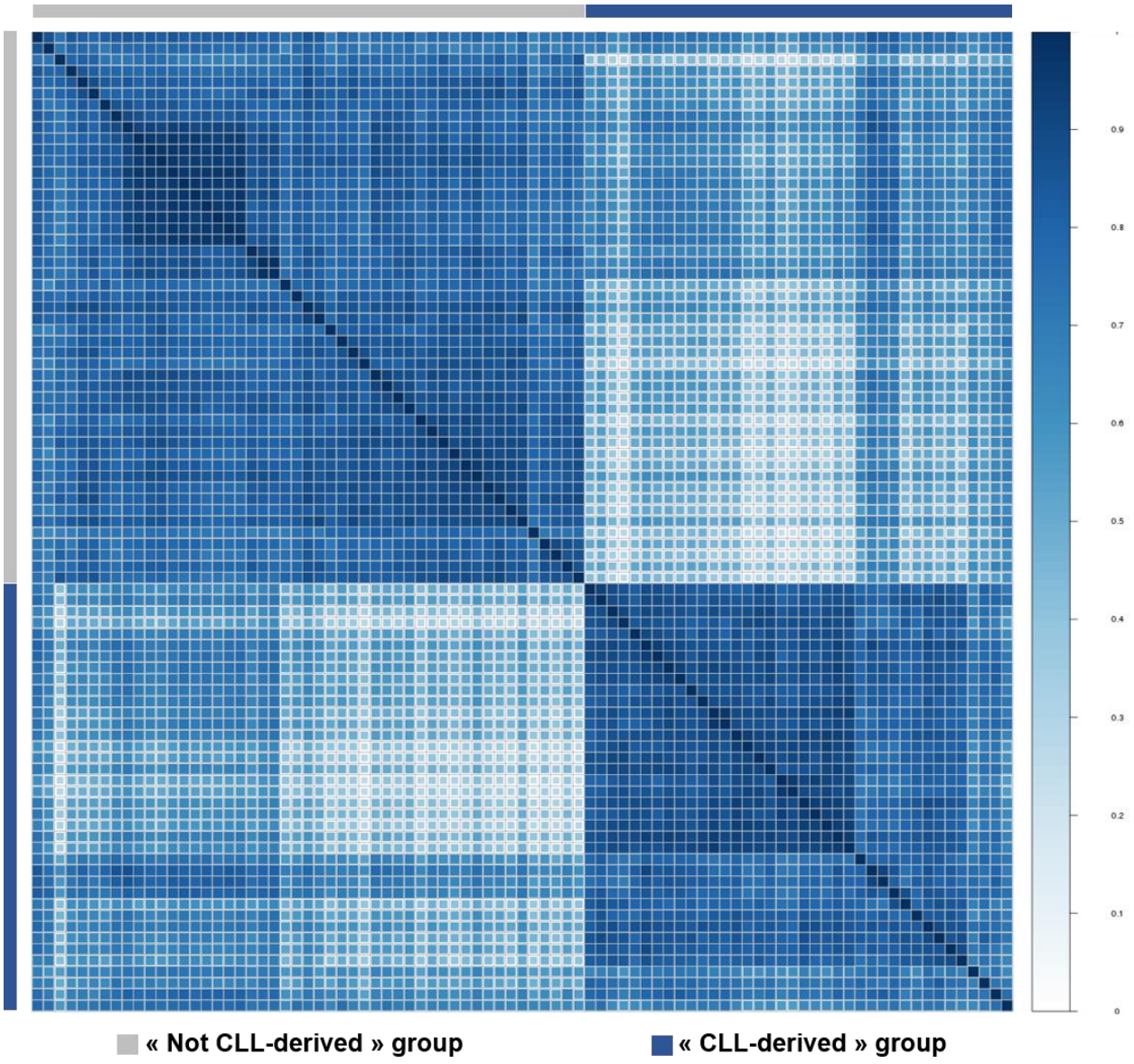
Supplementary Fig. 33. Correlation heat map of all replicate samples in the FULL dataset after complete QC, SWAN normalization and extended checks. SWAN: subset-quantile within array normalization; QC: quality control.



Supplementary Fig. 34. Correlation heat map for the entire 90-sample EPIC dataset after SWAN normalization, extended QC and checks. Pearson's correlation values: min = 0.64; median = 0.86; mean = 0.85. CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; QC: quality control; RS: Richter syndrome; SWAN: Subset-quantile within array normalization.



Supplementary Fig. 35. Correlation heat map for the entire 433-sample FULL dataset after SWAN normalization, extended QC and checks. Pearson's correlation values: min = 0.58; median = 0.88; mean = 0.87. Benign: normal B-cells spanning different stages of the B-lymphopoiesis; CLL: chronic lymphocytic leukemia; DLBCL: *de novo* diffuse large B-cell lymphoma; QC: quality control; RS: Richter syndrome; SWAN: subset-quantile within array normalization.



Supplementary Fig. 36. Correlation matrix of the ranked 215-gene signature. There is a clear discrimination between two groups according to this gene signature: i) the « CLL-derived group » is the CLL-derived RS signature-sharing samples, composed of CLLs + $_{low}$CLL-derived RS + $_{high}$CLL-derived RS, and ii) the « not CLL-derived group », composed of all other samples, naming de novo DLBCLs, tMZL (1 sample), tWM (1 sample), DLBCL-like RS and benigns. Benign: normal lymph node; CLL: chronic lymphocytic leukemia; DLBCL: diffuse large B-cell lymphoma; tMZL: transformed marginal zone lymphoma; tWM: transformed Waldenström Macroglobulinemia; RS: Richter syndrome.

Méthodes supplémentaires

Targeted Next-Generation Sequencing

Targeted NGS was performed using Ampli-Seq (Illumina; cat. no. 20020495 and cat. no. 20019102) method according to the manufacturer's protocol. A 13-gene panel was used, including *TP53* (exons 2-10), *SF3B1* (exons 12-16), *NOTCH1* (exon 34), *ATM* (exons 2-63), *BIRC3* (exons 6 and 7), *MYD88* (exons 3 and 5), *FBXW7* (exons 1-11), *POT1* (exons 4-7 and 9), *XPO1* (exons 15 and 16), *RPS15* (exons 3 and 4), *NFKBIE* (exons 1-6), *EGR2* (exons 1 to 3), *BRAF* (exon 15). Sequences were aligned, and mutation calling was performed using a custom bioinformatics pipeline. (1) A threshold of 5% variant allelic fraction was chosen to retain mutations.

Methylome data analyses

Dataset generation. First datasets were created using functions, objects and protocols mainly derived from the minfi package. (2) Several steps led to EPIC and FULL final datasets, which were the starting point of all downstream analyses. The EPIC set comprises 90 distinct samples (58 RS, 25 CLL, plus a subset of 7 DLBCL replicates also available on 450K), interrogated on EPIC 850K. The FULL dataset comprises 433 distinct samples from all cohorts (92 benign B-cells, 215 CLLs, 68 DLBCLs and 58 RS), combined into a single 450K object containing probes shared by 850K and 450K microarrays. Technically, it transcribes into (i) raw IDAT files corresponding to 96 and 377 samples for the 850K (866,091 CpGs) and 450K (485,512 CpGs) platforms, respectively, and included technical replicates; (ii) each subset was loaded independently, stored into a dedicated *RGChannelSet* minfi object, along with full sample annotations, then both were combined into a third subset containing 473 samples x 452,567 CpGs using the *combineArrays* function with output type as "IlluminaHumanMethylation450k"; (iii) the EPIC dataset stems from the first (850K) subset alone, the FULL dataset is obtained from the third (combined) subset. The global project workflow is displayed in Figure 1.

At this point, to reduce technological issues and biases, the same preparation protocol was applied to both 850K and combined subsets. The main stages of the filtering and quality control pipeline are as follows: (i) technical checks, filtering, and evaluations (ii) data normalization with SWAN; (3) (iii) probes located on X and Y chromosomes, flagged as cross-hybridization probes, or located near known SNPs were further removed with the rmSNPandCH function (with parameters dist=2 and mafcut=0.05) available from the DMRcate package; (4) (iv) imputation of the remaining failed β-value positions with *imputePCA* of the R missMDA package, (5) as is considered good practice when dealing with methylation data; (6) (v) 2x2 sample correlation checks (Supplementary Fig. 33 and 34). Correlation heat maps were rendered with the R corrplot package; (vi) extended quality control step to remove sample outliers and check for residual post-normalization batch effects (Supplementary Fig. 32); (vii)

ultimately, technical replicates were averaged into unique samples as all replicates were found comparable (Supplementary Fig. 35). These filtering steps led to the final EPIC (90 samples x 794,927 CpGs) and FULL (433 samples x 397,769 CpGs) datasets.

Copy number variations and detection of CpG deletions. Copy number variation (CNV) analyses, gene and promoter regions probe retrieval, β-value visualization and CpG deletion detections were performed on the EPIC dataset with the R package SeSAMe. [7] The *openSesame* pipeline was used with default parameters for background subtraction with *noobsb*, dye bias correction with *dyeBiasCorr*, probe masking on quality and detection. β-values were finally obtained by using *getBetas* with sum.TypeI = TRUE as a parameter to rescue probes with SNP hitting the extension base and hence switching color channel. [8] Promoter regions were plotted using β-values with *visualizeGene* and *visualizeRegion*, with deleted CpGs marked as masked or removed. CNVs were normalized against a copy-number-normal data set (EPIC.5. normal; available in the package), computed with the *cnSegmentation* function, and visualized with *visualizeSegments*. DNAcopy segmentation files were generated for all Richter samples, with image files created for each chromosome.

Downstream bioinformatics

Supervised analyses. As a rule, β-values were used for direct interpretation and graphical representation, while M-values were favored for statistics and computations. Linear modeling based on empirical Bayesian methods was used to assess for CpG differential methylation. When applicable, these models included cell deconvolution results as added covariates to correct for B-cell content. Additionally, at this point, any unwanted methylation variation such as residual batch effects were removed by using the *RUVm* function from package missMethyl. [9] The overall dispersion was calculated on the entire dataset, then p-values for each comparison were obtained with a two-way moderated t-test and adjusted for False Discovery Rate (FDR) following the Benjamini–Hochberg procedure. At probe level, an FDR < 0.01 indicated statistical significance. Differentially methylated region (DMR) determination was performed on the same linear models with *dmrcate* (package DMRcate), with lambda = 1000 and C = 3. FDR cut-off for first allowing a CpG to initiate a DMR was set to FDR=0.01, and DMRs were considered statistically significant if both min_smoothed_fdr and HMFDR output probabilities were < 0.01.

Unsupervised analyses. Explorations were conducted on β-values, and all methods used Euclidean distances as (dis)similarity metrics. PCAs were performed with R packages FactoMineR and factoextra, on the entire datasets or a subset of the top variant CpGs across all considered samples. Hierarchical clusterings included complete and average linkage criteria, and resulting heat maps and dendrograms were rendered with the R package ComplexHeatmap. Non-Negative Matrix Factorizations were performed with the R package

NMF, either on all CpGs or a subset of the most variant ones according to the context, with method *lee*, and parameters ranking = 3 and iterations = 50.

Feature annotations. Methylome data were analyzed using the available Illumina 450K and EPIC platform annotations, which strongly rely upon the hg19 assembly. As several tools like minfi and DMRcate still use those platform manifests by default, CpG and DMR locations/annotations were lifted to hg38 coordinates as an after-computation-process when required, especially when dealing with integrations with transcriptomic and epigenomic data. Additional CpG annotations included B-cell development modules, [10] UCSC tracks for the EBV-transformed GM12878 cell line, such as DNaseI and chromatin marks from ENCODE and transcription factor ChIP-Seq peaks from ENCODE3, histone modifications, and chromHMM chromatin states for 7 reference CLL epigenomes (2 U-CLLs and 5 M-CLLs). [11] Any *liftOver* of coordinates between hg19- and hg38-annotated data was achieved with the UCSC table browser or with R packages liftOver and XGR.

Gene set and pathway analyses. Unbiased functional annotations on ontological terms (GO) and KEGG pathways were achieved at CpG and DMR levels with the R package missMethyl. Additional enrichment analyses were conducted on curated gene lists with Enrichr. [12] Reactome pathway over-representations and enrichment analyses were performed with ReactomePA [13] on curated sets of unique genes associated with identified DMRs. Enrichments in sets of CpGs, DMRs, target genes, GO terms, or pathways were calculated as the occurrences of the selection against a background representing the entire dataset, which is equivalent to the formulation: Enrichment = Observed frequency / Expected frequency. P-values associated with enrichment analyses were obtained with either i) an overrepresentation test, ii) Fisher's exact test, or iii) a Chi-squared test, depending on the context and group size.

Transcriptomics

Transcriptome reconstruction pipeline. First quality controls were conducted using FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) results as a guideline. No adapter content or known overrepresented sequence needed to be removed at this step. Read mapping and the main filtering were performed using HISAT2 v2.0.4 [14] against a reference index built to account for human population SNPs as well as known transcripts (this index can be obtained from ftp://ftp.ccb.jhu.edu/pub/infphilo/hisat2/data/grch38_snp_tran.tar.gz). The following scoring constraints were applied during alignment: --score-min L, 0,-0.2 --sp 10,3 --dta. Samtools v1.3.1 (http://github.com/samtools/samtools) was used for manipulating the alignment files throughout the downstream analysis. PCR duplicates were flagged with Picard v1.13 MarkDuplicates (https://broadinstitute.github.io/picard/). Differentially spliced transcripts were assembled from the obtained alignments with StringTie v2.1.0. [15] The present protocol took advantage of the proposed workflow for identifying known as well as novel isoforms (https://ccb.jhu.edu/software/stringtie/index.shtml?t=manual), using an annotation file for hg38

in gtf format as a guide (downloaded from ftp://ftp.ensembl.org/pub/release88/gtf/homo_sapiens/Homo_sapiens.GRCh38.90.gtf.gz. The following parameters were used: i) first step is applied for each sample -f 0.2 -j 3 -c 10 -M 0.5, ii) second step merges all transcripts of all samples --merge -m 200 and iii) the last step estimates abundances and read coverage for all the merged transcripts, for each sample -A -C -f 0.2 -j 3 -c 10 -M 0.5. Two tables were generated from these results, one compiling raw read count at the gene level, and another at the transcript level.

Raw abundance filtering and normalization. Raw counts were filtered by applying a minimum expression threshold for a gene or transcript. Those had to be expressed (non-zero value) in at least two samples and present an average expression value across all samples higher than 1/5,000,000 of the average library size (64 ± 3 million reads per sample), that is, at least 20 reads per feature. Data was further adjusted with the TMM normalization method, [16] and finally was log2- and cpm- (count per million) transformed -- following good-practice measures for assessing differential expression - as previously described. [17] A total of 23,508 genes and 77,491 transcripts were identified and reported at the end of the process. Pearson's correlations for gene expression levels averaged at 0.92 for genes and 0.75 for transcripts and were very stable across samples (data not shown).

Gene and transcript annotations. All transcriptomic analyses were performed using the hg38 reference assembly of the human genome. Results were fully annotated with known symbols corresponding to gene and transcript genomic locations whenever possible. Upon completion of the transcript assembly, gene symbols were assigned Ensembl IDs based on overlapping positions with known transcripts (90% overlap minimum). In case of failed overlap, custom and unique IDs were used. Therefore, gene and transcript assignments were based on the Ensembl [18] GRCh38 annotations available in both *core* and *funcgene* databases, version 90. These were downloaded from ftp://ftp.ensembl.org/pub/release-90/mysql/ for local installation and query with in-house custom tools).

Variant calling from expressed transcripts. SNPs and small indels for each sample were called against reference hg38 from pileups generated with samtools with parameters -A -C 50 -q 10 -Q 30 -m 2 to discard bad nucleotide or sequence alignment quality. Multiallelic variants were called with bcftools v1.3.1 (http://github.com/samtools/bcftools) with mutation rate set at 1/1000 and default ploidy for GRCh38 and were then filtered to exclude low depth of coverage (DP < 10) and quality biased (QUAL < 10) calls. Samples were merged, indels normalized to the reference sequence, and multiallelic sites split into multiple biallelic entries, then genotypes and call statistics were re-evaluated for every locus. Low-quality positions (based on reported statistical tests) were finally discarded and replaced with missing genotypes. Variants were annotated with Annovar version 2019-10-24 [19] with databases among avsnp147, cosmic70, clinvar_20160302, dbscsnv11 and dbnsfp30a, Ensembl and refSeq annotations.

Transcriptome explorations. Unsupervised analyses were all carried out with hierarchical and K-means clustering techniques, as previously described. [20] Expression values were median-centered, and uncentered Pearson's correlation was used as distance metrics. Supervised analyses were performed through linear modeling (empirical Bayes), and differential expression p-values were obtained using a two-way moderated t-test then adjusted for false discovery rate (FDR) following the Benjamini–Hochberg procedure. An FDR < 0.01 indicated statistical significance. Cluster dissection was achieved with functional annotation tools for target gene associations, such as the Open Targets platform, [21] and gene signature correlation with public datasets from multiple databases, such as GEO (Gene Expression Omnibus), with Enrichr (https://maayanlab.cloud/Enrichr). Additional gene set, pathway, and enrichment analyses were carried out as described in the Methylome methods section.

Références supplémentaires

1. Tausch E, Schneider C, Robrecht S, Zhang C, Dolnik A, Bloehdorn J, et al. Prognostic and predictive impact of genetic markers in patients with CLL treated with obinutuzumab and venetoclax. Blood. 2020;135(26):2402-12.
2. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.
3. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol. 2012;13(6):R44.
4. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin. 2015;8:6.
5. Josse J, François H. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. Journal of Statistical Software; 2016. p. 1-31.
6. Lena PD, Sala C, Prodi A, Nardini C. Methylation data imputation performances under different representations and missingness patterns. BMC Bioinformatics. 2020;21(1):268.
7. Zhou W, Triche TJ, Laird PW, Shen H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. Nucleic Acids Res. 2018;46(20):e123.
8. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res. 2017;45(4):e22.
9. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32(2):286-8.
10. Kulis M, Merkel A, Heath S, Queirós AC, Schuyler RP, Castellano G, et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nat Genet. 2015;47(7):746-56.
11. Beekman R, Chapaprieta V, Russiñol N, Vilarrasa-Blasi R, Verdaguer-Dot N, Martens JHA, et al. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat Med. 2018;24(6):868-80.
12. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7.
13. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016;12(2):477-9.
14. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357-60.
15. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290-5.
16. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.
17. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.
18. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, et al. Ensembl 2019. Nucleic Acids Res. 2019;47(D1):D745-D51.
19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

20. Pouget C, Hergalant S, Lardenois E, Lacomme S, Houlgatte R, Carpentier C, et al. Ki-67 and MCM6 labeling indices are correlated with overall survival in anaplastic oligodendroglioma, IDH1-mutant and 1p/19q-codeleted: a multicenter study from the French POLA network. Brain Pathol. 2020;30(3):465-78.
21. Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. Nucleic Acids Res. 2021;49(D1):D1302-D10.

Résumé :

Les maladies du métabolisme, qu'elles soient complexes, héritées, rares ou communes, sont toutes la cause, d'un point de vue génomique, d'une dérégulation des niveaux moléculaires et voies cellulaires normalement à l'équilibre chez les sujets sains. Les systèmes biologiques que représentent les multiples couches omiques (transcriptome, méthylome, épigénome, protéome, métabolome, exome, génome) permettent d'en dresser la carte d'identité, caractérisations indispensables en médecine de précision, en déchiffrant par biologie intégrative les voies du dysfonctionnement cellulaire associé. L'obésité, qui correspond à un excès de masse grasse entrainant des conséquences néfastes sur la santé, est fortement liée aux maladies métaboliques, et prédispose à certains types de cancers, notamment via des mécanismes inflammatoires. Ce travail se propose de mieux comprendre les multiples liens associant obésité, maladies métaboliques, inflammation et cancers, pour déchiffrer leur origine environnementale et multifactorielle. Il s'intéresse particulièrement à la génomique des cancers cérébraux (méningiomes et oligodendrogliomes) et des hémopathies malignes de la lignée des lymphocytes B (syndrome de Richter, leucémie lymphoïde chronique, lymphomes B diffus à grandes cellules) par des approches mono et multi-omiques permettant leur caractérisation détaillée. Y sont identifiés des biomarqueurs et des signatures moléculaires utilisables pour le pronostic, la thérapeutique et les (re-)classifications en entités aux physiopathologies bien distinctes. Nous étudions également la pléïotropie des gènes de l'adiponectine, une hormone majeure de l'homéostasie énergétique, liée à l'obésité et au diabète de type 2. En revisitant bioinformatiquement les connaissances accumulées en génétique des populations par des études d'associations, nous reconstruisons les réseaux pléiotropiques autour de l'adiponectine et élucidons les voies aboutissant aux traits phénotypiques associés, du métabolisme des acides gras aux fonctions hépatiques, de la cognition aux désordres psychologiques et psychiatriques, de l'inflammation aux cancers.

Mots-clés : bioinformatique, génétique, génomique, omique, obésité, cancer, pléïotropie.

Abstract:

From a genomic point of view, metabolic diseases – whether complex, inherited, rare or common – consistently induce molecular dysregulations in cellular pathways normally maintained at equilibrium in healthy subjects. Such biological systems as those represented by the multiple omic layers (transcriptome, methylome, epigenome, proteome, metabolome, exome, genome) allow the elaboration of patient's identity maps which are critical in precision medicine. Obesity corresponds to the accumulation of excess fat mass leading to harmful physical and psychological consequences. It is strongly linked to metabolic diseases and predisposes to certain types of cancer, particularly through inflammatory mechanisms. This work aims to better understand the multiple links associating obesity, metabolic diseases, inflammation and cancers, and to decipher their environmental and multifactorial origin. We first focus on cancer genomics, with brain cancers (meningiomas and oligodendrogliomas) and haematological malignancies of the B lineage (Richter syndrome, chronic lymphocytic leukemia and diffuse large B cell lymphomas) using single and multi-omics approaches allowing their detailed characterization. We identify biomarkers and molecular signatures that can be used for prognosis, therapy and (re-)classification into entities with very distinct physiopathologies. We also decipher the pleiotropy surrounding the adiponectin, an energy homeostasis key hormone linked to obesity and type 2 diabetes. Using integrative bioinformatics, we revisit the accumulated knowledge obtained with population genetics through association studies, reconstruct the pleiotropic networks around the adiponectin genes and highlight the pathways leading to correlated phenotypic traits, from fatty acid metabolism to liver functions, from cognition to psychological and psychiatric disorders, from inflammation to cancers.

Keywords: bioinformatics, genetics, genomics, omics, obesity, cancer, pleiotropy.