Epigenetic and integrative cross-omics analyses of cerebral white matter hyperintensities on MRI

Yunju Yang,[1,†] Maria J. Knol,[2,†] Ruiqi Wang,[3,†] Aniket Mishra,[4] Dan Liu,[5] Michelle Luciano,[6] Alexander Teumer,[7,8,9] Nicola Armstrong,[10] Joshua C. Bis,[11] Min A Jhun,[12] Shuo Li,[3] Hieab H. H. Adams,[2,13] Nasir Ahmad Aziz,[5,14] Mark E. Bastin,[15] Mathieu Bourgey,[16,17] Jennifer A. Brody,[11] Stefan Frenzel,[18] Rebecca F. Gottesman,[19] Norbert Hosten,[20] Lifang Hou,[21] Sharon L. R. Kardia,[12] Valerie Lohner,[5] Pascale Marquis,[16,17] Susana Muñoz Maniega,[15] Claudia L. Satizabal,[22,23,24] Farzaneh A. Sorond,[25] Maria C. Valdés Hernández,[15] Cornelia M. van Duijn,[2,26] Meike W. Vernooij,[2,13] Katharina Wittfeld,[18,27] Qiong Yang,[3,23] Wei Zhao,[12] Eric Boerwinkle,[28,29] Daniel Levy,[23,30] Ian J. Deary,[6] Jiyang Jiang,[31] Karen A. Mather,[31,32] Thomas H. Mosley,[33] Bruce M. Psaty,[11,34] Perminder S. Sachdev,[31,35] Jennifer A. Smith,[12] Nona Sotoodehnia,[11] Charles S. DeCarli,[36] Monique M. B. Breteler,[5,37] M. Arfan Ikram,[2] Hans J. Grabe,[18,27] Joanna Wardlaw,[15] W. T. Longstreth Jr,[34,38] Lenore J. Launer,[39] Sudha Seshadri,[22,23,24] Stephanie Debette[4,24,40] and Myriam Fornage[1,28]

[†]**These authors contributed equally to this work.**

1 Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science at Houston, Houston, Texas, USA

2 Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

3 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA

4 University of Bordeaux, Inserm, Bordeaux Population Health Research Center, team VINTAGE, UMR 1219, Bordeaux, France

© The Author(s) 2022. Published by Oxford University Press on behalf of the Guarantors of Brain. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

5 Population Health Sciences, German Centre for Neurodegenerative Diseases (DZNE), Bonn, Germany

6 Department of Psychology, University of Edinburgh, Edinburgh, UK

7 Institute for Community Medicine, University Medicine Greifswald, Greifswald 17475, Germany

8 German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, Germany

9 Department of Population Medicine and Lifestyle Diseases Prevention, Medical University of Bialystok, Bialystok, Poland

10 Mathematics and Statistics, Curtin University, Perth, Australia

11 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA

12 Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA

13 Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands

14 Department of Neurology, Faculty of Medicine, University of Bonn, Bonn, Germany

15 Centre for Clinical Brain Sciences, Department of Neuroimaging Sciences, University of Edinburgh, Edinburgh, UK

16 Canadian Centre for Computational Genomics, McGill University, Montréal, Quebec, Canada

17 Department for Human Genetics, McGill University Genome Centre, McGill University, Montréal, Quebec, Canada

18 Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany

19 Stroke Branch, National Institutes of Neurological Disorders and Stroke, Bethesda, Maryland, USA

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

20 Department of Radiology and Neuroradiology, University Medicine Greifswald, Germany

21 Department of Preventive Medicine, Northwestern University, Chicago, Illinois, USA.

22 Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases and Department of Population Health Sciences, UT Health San Antonio, San Antonio, Texas, USA

23 The Framingham Heart Study, Framingham, Massachusetts, USA

24 Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA

25 Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

26 Nuffield Department of Population Health, Oxford University, Oxford, UK

27 German Center for Neurodegenerative Diseases (DZNE), Site Rostock/ Greifswald, Rostock, Germany

28 Human Genetics Center, School of Public Health, University of Texas Health Science at Houston, Houston, Texas, USA

29 Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA

30 Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA

31 Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, New South Wales, Australia

32 Neuroscience Research Australia, Sydney, Australia

33 The Memory Impairment Neurodegenerative Dementia (MIND) Research Center, University of Mississippi Medical Center, Jackson, Mississippi, USA

34 Department of Epidemiology, University of Washington, Seattle, Washington, USA

35 Neuropsychiatric Institute, The Prince of Wales Hospital, University of New South Wales, Randwick, New South Wales, Australia

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

36 Department of Neurology and Center for Neuroscience, University of California at Davis, Sacramento, California, USA

37 Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Bonn, Germany

38 Department of Neurology, University of Washington, Seattle, Washington, USA

39 Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA

40 CHU de Bordeaux, Department of Neurology, Bordeaux, France

Correspondence to: Myriam Fornage, PhD

The Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston

1825 Pressler street, Suite 530F

Houston, Texas, USA 77030

E-mail: Myriam.Fornage@uth.tmc.edu

Running title: Epigenetics of white matter hyperintensities

Keywords: epigenome-wide association study; white matter hyperintensities; cerebral small vessel disease; integrative cross-omics analysis; blood-brain barrier dysfunction

Abbreviations: BBB = blood-brain barrier; CADASIL = Cerebral Autosomal Dominant Arteriopathy with Sub-cortical Infarcts and Leukoencephalopathy; cSVD = cerebral small vessel disease; DNAm = DNA methylation; WMH = cerebral white matter hyperintensities

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

Abstract

Cerebral white matter hyperintensities on MRI are markers of cerebral small vessel disease, a major risk factor for dementia and stroke. Despite the successful identification of multiple genetic variants associated with this highly heritable condition, its genetic architecture remains incompletely understood. More specifically, the role of DNA methylation has received little attention.

We investigated the association between white matter hyperintensity burden and DNA methylation in blood at approximately 450,000 CpG sites in 9,732 middle-aged to older adults from 14 community-based studies. Single-CpG and region-based association analyses were carried out. Functional annotation and integrative cross-omics analyses were performed to identify novel genes underlying the relationship between DNA methylation and white matter hyperintensities.

We identified 12 single-CpG and 46 region-based DNA methylation associations with white matter hyperintensity burden. Our top discovery single CpG, cg24202936 (P=7.6x10^{-8}), was associated with *F2* expression in blood (P=6.4x10^{-5}), and colocalized with *FOLH1* expression in brain (posterior probability =0.75). Our top differentially methylated regions were in *PRMT1* and in *CCDC144NL-AS1*, which were also represented in single-CpG associations (cg17417856 and cg06809326, respectively). Through Mendelian randomization analyses cg06809326 was putatively associated with white matter hyperintensity burden (P=0.03) and expression of *CCDC144NL-AS1* possibly mediated this association. Differentially methylated region analysis, joint epigenetic association analysis, and multi-omics colocalization analysis consistently identified a role of DNA methylation near *SH3PXD2A*, a locus previously identified in genome-wide association studies of white matter hyperintensities. Gene set enrichment analyses revealed

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

functions of the identified DNA methylation loci in the blood-brain barrier and in the immune response. Integrative cross-omics analysis identified 19 key regulatory genes in two networks related to extracellular matrix organization, and lipid and lipoprotein metabolism. A drug repositioning analysis indicated antihyperlipidemic agents, more specifically peroxisome proliferator-activated receptor alpha, as possible target drugs for white matter hyperintensities.

Our epigenome-wide association study and integrative cross-omics analyses implicate novel genes influencing white matter hyperintensity burden, which converged on pathways related to the immune response and to a compromised blood brain barrier possibly due to disrupted cell-cell and cell-extracellular matrix interactions. The results also suggest that antihyperlipidemic therapy may contribute to lowering risk for white matter hyperintensities possibly through protection against blood brain barrier disruption.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

Introduction

Cerebral white matter hyperintensities (WMH) on MRI are indicative of cerebral small vessel disease (cSVD) and are part of the spectrum of brain vascular injury that impacts cognitive function, also known as vascular contributions to cognitive impairment and dementia (VCID).[1,2] While the pathophysiology of WMH is little understood and likely heterogeneous, it likely has ischemic and neurodegenerative origins.[1] Historical pathology data suggested chronic ischemia resulting in demyelination and axonal loss as an underlying mechanism; however, neuroimaging data point to blood-brain barrier dysfunction, dysfunctional blood flow linked with impaired cerebrovascular autoregulation, vascular stiffness, periarteriolar inflammation, and more recently protein deposition (i.e. amyloid angiopathy).[2] Genetics plays a significant role in WMH with a heritability estimated from 54 to 80%[3–7]; however, the genetic variants identified in association studies explain only ~29% of WMH variance.[8,9] Epigenetic changes such as DNA methylation (DNAm), which regulate gene expression, have emerged as another key component of the genetic architecture of complex traits.[10] Unlike DNA sequence variation, which remains unchanged throughout life, DNAm is plastic and highly sensitive to changes in the environment and aging.[10,11] To date, its role in cSVD has received little attention. We hypothesized that there may be patterns of DNAm associated with WMH that are common across all populations. We also hypothesized that the interplay between genotype, epigenotype, and risk factor exposure underlies cSVD etiology and used an integrated analytic framework to identify such relationships.

Materials and methods

Overview

This study comprises five analytic parts to implicate novel genes and gene networks in WMH etiology (Figure 1). First, we performed an epigenome-wide association analysis to identify DNAm loci, both cytosine-phosphate-guanine (CpG) sites and differentially methylated regions (DMRs), associated with WMH burden. The identified DNAm loci were then annotated for regulatory features, pathways, and association with other traits. Second, we investigated the contribution of genetic variation to variation in DNAm at the identified CpGs and used

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

Mendelian randomization (MR) techniques to test for causal association with WMH burden and for the mediating role of expression of nearby genes. Third, we examined the role of DNAm at established WMH genome-wide association study (GWAS) loci. Fourth, we integrated gene expression and expression quantitative trait loci (eQTL) data to prioritize candidate genes associated with the identified CpGs. Lastly, we performed integrative cross-omics analyses to derive WMH-associated genes networks and their key drivers and to reposition drug targets.

Study subjects

The sample included 9,732 middle-aged to older adults of European (EA) and African ancestry (AA) from 14 community-based studies. Our discovery sample includes 5,715 subjects of European ancestry (EA, *n*=4,610) and of African ancestry (AA, *n*=1,105) from Atherosclerosis Risk in Communities (ARIC),[12] Biobanking and BioMolecular resources Research Infrastructure (BBMRI),[13] Cardiovascular Health Study (CHS),[14] Coronary Artery Risk Development in Young Adults (CARDIA),[15] Framingham Heart Study (FHS) offspring study,[16,17] Genetic Epidemiology Network of Arteriopathy (GENOA) study,[18] Lothian Birth Cohort (LBC) 1936,[19,20] Rotterdam Study (RS),[21,22] and Study of Health in Pomerania (SHIP) [23]. To replicate our findings, we accessed data on 3,398 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI),[24,25] FHS 3rd generation study,[26] the Older Australian Twin Study (OATS),[27,28] and the Rhineland study[29]. Additionally, we included a secondary replication sample (*n*=619) from the BRIDGET consortium.[30] Subjects with history of stroke or dementia were excluded. Details about participating studies and study-specific ethics statements are provided in Supplementary Data I. Each study obtained written informed consent from all participants and approval from the appropriate institutional review boards.

WMH burden measurements

Brain MRI was taken in the same or the closest subsequent visit to the visit in which DNAm was measured. In each study, MRI scans were performed and interpreted using standardized procedures without reference to demographic or clinical information. The field strength of the scanners used ranged mostly from 1.5 to 3.0 Tesla. T1-, T2-, and/or proton-density-weighted scans were obtained for all participants. The majority of the studies used fully automated segmentation method to quantify WMH burden. MRI procedures and WMH

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

quantification in each study are detailed in Supplementary Data II.

DNAm profiling

DNAm levels were measured at ~450K CpGs from whole blood samples with the Illumina Infinium Human450 Methylation BeadChip in most participating cohorts. GENOA study measured methylation levels at ~27K CpGs with the Illumina Infinium HumanMethylation 27 BeadChip, entirely covered by the Human450 BeadChip. CARDIA, SHIP-TREND, ADNI, and Rhineland Study used the Illumina MethylationEPIC BeadChip with a denser coverage of CpGs (~850K). Each study independently performed quality control (QC) for DNAm data, complying with the agreed minimum QC guidelines; CpGs with more than 95% of samples with a detection P<0.01 and samples with more than 95% of CpGs with a detection P <0.01 were selected. DNAm values were then standardized using an intra-array normalization method. The BRIDGET Consortium measured DNAm levels using Hi-seq bisulfate sequencing, and DNAm sites with sample coverage less than 95% were excluded. Details of DNAm data collection and processing in participating studies are presented in Supplementary Data III.

Cohort-level epigenome-wide association analyses

We tested association between DNAm level (untransformed beta values) and WMH burden (ln(WMH+1)) using a linear mixed regression model by ancestry group adjusted for age, sex, study site if applicable, total (intra)cranial volume (cm^3), white blood cell proportion (%),[31] and within-ancestry principal components (PCs) as fixed effects and technical covariates (i.e. plate, chip-position, row, and column) as random effects. In FHS, family structure is also adjusted as a random effect. Multi-ancestry studies with a small number of subjects in each ancestry, namely CHS and CARDIA, performed a pooled-ancestry analysis that also adjusted for ancestry group as fixed effects. Additionally, subgroup analyses by hypertension status were conducted. Hypertension was defined if either systolic or diastolic blood pressure (SBP or DBP) is greater than 140 mmHg or 90 mmHg, respectively, or if a subject was taking any anti-hypertensive medication at the time of MRI measurement. In the BRIDGET study, we tested the association of DNAm with an extreme-SVD phenotype defined as excessive WMH volume with or without brain infarcts accounting for age, sex, country, the sequencing read counts, and sample relatedness.[32] DNAm measurements and statistical models used in participating studies are

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

described in Supplementary Data III.

Epigenome-wide meta-analysis and replication analysis

We combined EWAS results based on sample-size-weighted z-score-based fixed-effect method in METAL[33] because WMH was measured on different scales in the various cohorts and because our primary aim was to identify novel DNA methylation loci for WMH burden rather than estimate effect sizes of methylation probes.[34] Hypertensive and normotensive subgroup meta-analyses and ancestry-specific meta-analyses (excluding CHS and CARDIA) were also performed. Study-specific results were corrected for inflation during meta-analysis if inflation was detected (genomic inflation factor (λ)>1.0). An association was considered as significant if *P* is smaller than Bonferroni threshold (approximately 1.2×10^{-7}). A less stringent threshold was also set as 1.0×10^{-5} to detect suggestive associations. CpGs on sex chromosomes were not considered because our analytic plan did not account for hemi-methylation on the X chromosome due to chromosome X inactivation in women. Cross-reactive CpGs reported by Chen *et al.*[35] and those showing evidence of heterogeneity (Cochran Q P-value <0.05) were removed from the results *post hoc*. In the replication samples, associations for the identified CpGs were tested. CpGs were considered replicated if they were significant at the Bonferroni threshold (0.05/the number of the CpGs). We plotted epigenetic associations *in cis* (±50 kb) using R 'coMET' package.[36]

Annotation of regulatory features and traits

We scored genomic positions of the identified CpGs according to RegulomeDB's [37] ranking criteria ranging from one (likely to affect binding and linked to expression of a gene target) to five (minimal binding evidence) and also computed a probability score within a range from zero to one (the most likely to be a regulatory variant). CpGs at the locations with significant regulatory features (rank category one or two, and probability score $\geq$ 0.9) are discussed. We also identified enhancers or promoters mapped to CpGs using the database of genome-wide enhancer-to-gene or promoter-to-gene associations computed based on five elements: eQTLs, eRNA co-expression, transcription factor co-expression, capture Hi-C and gene target distance (GeneHancer DB).[38] Identified CpGs were also searched in EWAS catalog[39] and EWAS atlas[40] to identify associated traits reported in previous EWAS. Lastly, to examine possible correlations among the CpGs, Spearman correlations were calculated in 906 EA and 639

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

AA subjects from the ARIC study.

DMR analysis

We performed a DMR analysis to identify a group of CpGs that collectively influence WMH burden using two specific methods, Comb-p[41] and DMRcate[42], accounting for their spatial correlations. Briefly, Comb-p detects regional enrichment of low *P*s at varying distance using the Stouffer-Liptak-Kechris correction for adjacent *P*s.[41] DMRcate models Gaussian kernel smoothing within pre-defined distance (1Kbp in this study) and collapses contiguous significant CpGs (P <0.05) after multiple testing correction. DMR identified by both Comb-p (Šidák P <0.05) and DMRCate (FDR <0.05) was considered significant. To replicate, individual association *Ps* were pooled at each identified DMRs using DMRCate in the replication samples.

Gene set enrichment analysis of WMH-associated epigenetic Loci

Identified CpGs and DMRs were tested for enrichment in gene sets from MSigDB c5 gene ontology database[43,44] and KEGG pathway database[45], using '*gsameth'* and '*gsaregion'* functions built in R 'missMethyl' package[46].

Shared epigenetics with BP

BP is an influential risk factor for WMH.[47–49] To investigate the shared epigenetics between WMH burden and BP, we performed a pairwise multivariate association test using summary statistics from a previous EWAS of SBP and DBP[50]. CpGs associated with both traits were tested against the null hypothesis H_0: $\beta_{WMH}=\beta_{SBP\ or\ DBP}=0$. The test uses Z-scores for each trait and estimates multivariate test statistics accounting for the trait correlation calculated based on the null associations (trait-specific $P>1\times10^{-5}$). This method is implemented in the '*metaUSAT*' software.[51] To avoid false positive associations driven entirely by one trait, we included CpGs showing significance ($P<0.001$) for both traits. Bonferroni threshold was set at 8.33×10^{-3} (=0.05/6) based on the number of associations tested.

Heritability analysis and GWAS of WMH-associated CpGs

Inter-individual variation in DNAm may result from differences in environmental exposures, stochastic variation, or genetic influences. To examine the contribution of genetic

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

variation to variation in DNAm at the identified CpGs, we estimated the narrow sense heritability (h^2_{meth}) in the FHS Offspring Cohort subjects (n=2,377) adjusting for age, sex, blood cell counts, PCs and technical covariates. Body mass index (BMI) and smoking status were additionally corrected in sensitivity analyses.

To further identify genetic variants associated with DNAm levels at the WMH-associated CpGs, we performed GWAS in ARIC EA subjects (n=984). Genotypes were measured with Affymetrix 6.0 array and imputed from 1000 Genome phase one version three reference using MaCH v1.0.16. Variants were excluded if minor allele frequency (MAF)<0.01, sample call rate<95%, or imputation quality<0.3. The untransformed methylation beta value was tested for genetic association adjusting for age, sex, top 42 methylation PCs, and blood cell components.

Bi-directional MR analysis of the identified CpGs and WMH burden

To determine if the WMH-associated DNAm level is a causal factor for WMH burden (Forward-MR) or a secondary outcome of WMH burden (Reverse-MR), we performed a bi-directional two-sample MR analysis[52] for the identified CpGs with at least three instrumental variables (IVs). We identified methylation quantitative trait loci (mQTL) associations in-cis (±1Mb) from the FHS study (n=4,170) that had been validated using ARIC data (n=963).[53] Those mQTLs were clumped at linkage disequilibrium (LD) r^2<0.05 for independence. For WMH, the UK biobank (UKBB) GWAS summary statistics (n=11,226)[54] was downloaded from the Cerebrovascular Disease Knowledge Portal (http://www.cerebrovascularportal.org/) on 01/09/2019. Reverse-MR analysis was performed using eight clumped genome-wide significant associations (LD r^2<0.05). Since the FHS mQTL study shares only significant associations *in cis*,[53] we used the mQTL association statistics from ARIC EA subjects for reverse-MR analysis.

For each CpG in both directions, causal association was tested based on the IVW method in the R package '*TwoSampleMR*'.[55] To validate the MR result, sensitivity analyses based on weighted median and MR Egger methods, and built-in tests for pleiotropy and heterogeneity were also performed. For existence of pleiotropy (MR egger intercept test P<0.05), the Egger regression estimate was assessed instead of the IVW estimate.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

Mediating effect of *in cis* genes between CpGs and WMH burden

To investigate if expression of nearby genes mediates the relationship between the identified CpG and WMH burden, a two-step MR analysis was performed. We tested the directional relationships 1) from "*the exposure* (CpGs)" to "*the mediator* (gene expression)" (step 1) and 2) from "*the mediator* (gene expression)" to "the outcome (WMH burden) (step 2) using the identified mQTL IVs, the WMH GWAS associations,[54] and eQTL associations from the GTEx version eight brain eQTL data accessed via eQTL Catalogue (https://www.ebi.ac.uk/eqtl/) on 2020/11/12. Among available GTEx brain tissues, cortex (n=205), frontal cortex (n=175), cerebellum (n=209), cerebellar hemisphere (n=175), and caudate basal ganglia (n=194) were selected. MR association based on the IVW method was again tested and sensitivity analysis was also performed. Gene expression with IVW P<0.05 at both steps was considered as a potential mediator in the association between the identified CpG and WMH burden.

Cis-acting genes associated with the identified CpGs in blood

To functionally annotate the identified CpGs, we tested associations with gene expression in blood in long-range [53]cis-regions (±5Mb) in 1,966 and 728 EA subjects from FHS and RS, respectively. Expression of the nearest gene/ mRNA was regressed on DNAm β score at the CpG adjusting for age, sex, population structure, and family structure (FHS only), blood cell counts and technical covariates. Technical covariates and family structure were modeled as random effects. In sensitivity analyses, smoking status and BMI were added to the model. Estimates from two studies were then combined for each gene using the sample-sized based meta-analysis method in METAL[33].

Genes colocalizing with the identified CpGs in Brain

To investigate cis-acting genes colocalized with the identified CpGs in the brain, we performed a multiple-trait colocalization (moloc) analysis using brain QTL data. Prior to this analysis, we examined the inter-individual correlations between DNAm levels in whole blood and in prefrontal cortex at the identified CpGs, using publicly available data.[56] For CpGs with significant correlation (P <0.05) between blood and prefrontal cortex, we tested the posterior probabilities for full-colocalization (PPFC) that multiple traits (DNAm, gene expression, and

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

WMH burden) share causal variants at each locus, given the data. We used coloc priors of $1x10^{-5}$. We identified EA-specific GWAS associations[8] and brain mQTL (n=543) and eQTL (n=534) associations accessed via http://mostafavilab.stat.ubc.ca/xqtl/.[57] If PPFC is greater than 0.7, we considered the gene is significantly colocalized with CpG and WMH burden. Moloc analysis was performed using the R package "*moloc*".[58]

Epigenetic regulation of known GWAS loci

We next investigated the role of DNAm at established WMH GWAS loci, which may not have been detectable at the genome-wide significance threshold. Among 26 loci reported in the latest WMH GWAS,[8] we mapped 450K-array CpGs to 21 loci. EWAS associations at each of these 21 loci were pooled using the Brown's method (implemented in the package "poolr") adjusting for dependence among CpGs.[59] For dependency information, we calculated correlation among CpGs in the GWAS loci using ARIC methylation data (906 EA and 639 AA subjects). A GWAS locus with combined P was considered significant if P is smaller than Bonferroni-adjusted threshold (0.05/number of loci tested).

Alternatively, we performed a moloc analysis at the 21 GWAS loci, again using the GWAS and brain QTL data.[8,57] With the priors of $1x10^{-5}$, we considered genes with a PPFC greater than 70% as convincingly colocalized with DNAm and WMH burden.

Identification of biological pathways using multi-dimensional data integration

Integrating multi-omics associations for WMH may boost power to identify novel genes influencing WMH burden. We integrated genetic[8], transcriptomic[60], and epigenetic genome-wide association studies of WMH using the R package '*mergeomics'* (version 1.2). [61] To reduce noise in the GWAS data, the top 50% of genetic associations[8] were included and pruned at $r^2<0.5$ based on HapMap3 LD information as recommended.[61] For transcriptomic associations, we used the recent WMH transcriptome-wide association study (TWAS) results.[60] For epigenetic associations, we used our discovery EWAS. Markers were primarily mapped to the nearest genes. For CpGs, cis-acting genes reported in the MesaEpiGenomics study[62] were additionally annotated. For each GWAS, EWAS and TWAS, we tested marker-level enrichment with

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

hierarchical permutation size of 20,000 based on biological pathways from pre-defined public databases: KEGG[45], REACTOME[63], Biocarta[64], and the gene ontology knowledgebase[65,66]. Then, we meta-analyzed the enriched gene sets from association studies and identified the WMH-associated gene sets (FDR-adjusted P <0.05).

To describe the regulatory network of the identified gene sets and identify its local hub genes, we performed a weighted key driver analysis (wKDA) using the web-based software *Mergeomics version 2.0.*[67] Gene regulatory network was constructed using in-house brain-specific Bayesian network (minimum hub overlap 0.33 and directed edge type) [68] and visualized via Cytoscape version 3.8.2.[69]

We also conducted an overlap-based drug repositioning analysis "*PharmOmics"* based on the identified key driver genes (FDR <0.05) to predict potential drugs or small molecules targeting WMH.[70] *PharmOmics* comprises a curated drug signature database covering 941 drugs, constructed from transcriptomic data across >20 tissues from rat, human, and mouse. For our analysis, we selected drug signatures from relevant tissues (*in vivo* human transcriptome data in cardiovascular and nervous system, and *in vitro* transcriptome data from murine oligodendroglial precursor cells), and examined the overlap between these drug signature genes and key-driver genes from our identified WMH-associated gene sets.

Data availability

The data that support the findings of this study are included in this manuscript. Full EWAS summary statistics are available in dbGaP at phs000930.v9.p1.

Results

Identification of epigenetic changes associated with WMH burden

Study sample characteristics

In the discovery sample, the mean age ranges from 49.7 years in SHIP to 74.6 in CHS. Sex ratios are balanced in all studies except for GENOA study, which has 72.8% female. ARIC, CARDIA, and CHS have both EA and AA subjects, other studies consist of single ancestry

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

subjects (AA or EA). In the primary replication sample, subjects from FHS 3rd generation and Rhineland Study (mean age 47.1 and 54.1 years, respectively), which compose 86.2% of the replication study, are younger than most discovery studies and show relatively smaller median WMH burden (0.34 in FHS 3rd generation study and 0.40 in Rhineland Study). All subjects in the replication studies are of EA. Demographic characteristics of participating cohorts are shown in Supplementary Table 1.

Novel DNAm loci are associated with WMH burden

In the discovery sample, we identified a novel epigenome-wide significant association between WMH burden and level of DNAm at cg24202936 (Z=5.38, P=7.58x10^{-8}) in *SEPTIN7P11*. Associations at cg24202936 in each study are presented in a forest plot (Supplementary Figure 1) and regional associations within 50 kb are presented with annotations (Supplementary Figure 2). At the suggestive significance threshold of 1x10^{-5}, we identified 11 additional loci (**Table 1**). The associations remained significant (P <0.05/12=4.17x10^{-3}) after adjusting for BMI, smoking status, and SBP and DBP. Quantile-quantile (QQ) and Miami plots are presented in Supplementary Figure 3 and 4. All subsequent analyses focus on these 12 CpGs, which are referred to as "target CpGs". None of the target CpGs associations were replicated in independent samples and a meta-analysis of the discovery and replication samples showed significant heterogeneity in many of the resulting associations, which was not present in the discovery cohorts (Supplementary Table 2). Target CpGs showed consistent associations with WMH in subgroup analyses by ancestry and hypertension status (Supplementary Table 3 and 4). Cg06450373 in *CDH18* (P=6.48x10^{-8}) was identified in normotensive subjects (Supplementary Table 5); but not replicated. In a gene set enrichment analysis on discovered CpGs (P<1.0x10^{-5}), "cell-cell junction organization" was identified as the top pathway (P=1.32x10^{-3}, false discovery rate (FDR)=0.32)**.**

Annotated regulatory functions of target CpGs

We found significant regulatory features from RegulomeDB at the genomic positions of cg24202936 (rank 2b and score 0.93), and cg06809326 (rank 2b and score 0.91) (Supplementary Table 6). Cg24202936 resides near a transcriptional starting site (0.2 Kb upstream), and identified as a transcriptional factor binding site computationally annotated with 20 genes

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

(Supplementary Table 6). Previously reported EWAS traits associated with target CpGs are presented in Supplementary Table 7. In particular, cg24202936 was previously reported associated with HIV infection.[71] Cg06450373, cg031161214, cg01506471, and cg14547240 were correlated each other in both ancestries with weak to moderate r (0.23 to 0.55) (Supplementary Figure 5). In AA, cg23586595 showed weak but significant correlations with cg13476133 (r=0.32), cg03116124 (r=-0.42), and cg14547240 (r=-0.36). No correlated CpG (|r|>0.3) was identified for our top CpG, cg24202936, in both ancestries**.**

WMH-associated DMRs are enriched in immune response-related pathways

We identified 46 DMRs in associations with WMH burden (Supplementary Table 8). Notably, one DMR was in *SH3PXD2A*, previously identified in genome-wide association studies (GWAS).[8,72–74] Identified DMRs were enriched in several gene ontologies, including STAT (signal transducer and activator of transcription) family protein binding (FDR=4.91×10^{-3}) and defense response to virus (FDR=5.68×10^{-3}), which are related to the immune response (Supplementary Table 9). Of the 46 identified DMRs, *PRMT1, ABAT, BHMT2, C11orf21, IZUMO1, C5orf66, ENPEP, SLC35F3, FBXO47, SLC45A4, KCTD16, KITLG,* and *UCN3* were replicated (Supplementary Table 8). Of note, *ENPEP, SLC35F3,* and *SLC45A4* were previously reported in BP GWAS.[75–81]

Shared epigenetic loci between WMH and BP

At the Bonferroni-corrected threshold (P <8.33×10^{-3}), we identified six CpGs associated with both WMH burden and BP (Supplementary Table 10). For WMH-DBP, cg23291754 in *MOBKL1A* (P=2.38×10^{-7}) and cg24372586 in *GNL1* (P=7.84×10^{-7}) were identified. For WMH-SBP, cg00711496 in *CDC42BPB* (P=1.99×10^{-7}), cg04987734 in *C19orf76;PRMT1* (P=3.09×10^{-7}), cg00934987 in *SEPT4* (P=1.07×10^{-6}), and cg18770635 in *KLHDC7B* (P=1.68×10^{-6}) were identified.

Heritability of the WMH-associated CpGs

Significant h^2_{meth} was estimated for cg17417856 (40.4%, P=1.37×10^{-8}), cg06809326 (26.5%, P=1.03×10^{-4}), cg23586595 (24.2%, P=1.47×10^{-3}), cg17577122 (14.3%, P=2.80×10^{-2}), and cg24202936 (15.5%, P=1.34×10^{-2}) (Table 2). Additional adjustment for BMI and smoking

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

status did not significantly modify these estimates. In GWAS of the target CpGs in the ARIC EA sample, we observed significant *cis* genetic influence on cg06809326, cg13476133, and cg24202936 (Supplementary Figure 6). This result agrees with a previous publication that included the same dataset.[53]

Mendelian randomization analyses between target CpGs and WMH burden

Forward two-sample multiple IV MR analysis was performed for two target CpGs, cg06809326 and cg24202936, which have at least three independent cis-mQTL IVs in Huan T *et al*[53] (Supplementary Table 11). We found a marginally significant causal relationship from cg06809326 to WMH burden (P=2.91x10^{-2}). Higher methylation level at the locus is associated with greater WMH burden (odds ratio (OR) [95% confidence interval (CI)]=1.39 [1.03, 1.87]). Evidence was lacking for horizontal pleiotropy (P=0.41) or heterogeneity (P=0.42) (Supplementary Table 12). In reverse-MR analysis, evidence that WMH causally influence methylation levels at any of the target CpGs was lacking (Supplementary Table 12 and 13).

Using the same three IVs, we also investigated whether cg6809326 is causally associated with expression of nearby genes (step 1). Two *cis* transcripts were annotated to this CpG in GTEx version eight data.[82] They both encode a long noncoding RNA designated as *CCDC144NL*, and *CCDC144NL-AS1* and we identified one IV for both transcripts. In all five brain tissues, we found evidence of causal association between cg06809326 and both *CCDC144NL* and *CCDC144NL-AS1* (Supplementary Table 14). In step two, a marginal association between *CCDC144NL* and WMH burden was observed in caudate basal ganglia and cortex (step one P=1.11x10^{-3} and step two P=3.94x10^{-2} in caudate basal ganglia; step one P=1.21x10^{-3} and step two P=4.28x10^{-2} in cortex).

DNAm at established GWAS loci and WMH burden

[8]We estimated the combined effect of DNAm at each locus from our EWAS results at the 21 established GWAS loci.[8] Consistent with our DMR results, CpGs at the GWAS locus *SH3PXD2A* were jointly associated with WMH (P=8.48x10^{-3}), but evidence of DNAm effects on WMH at other loci was lacking (Supplementary Table 15). We also conducted a multiple trait colocalization analysis (moloc)[58] of brain mQTL and expression QTL (eQTL)[57], and WMH-associated single nucleotide polymorphisms (SNPs). At 17 out of the 21 GWAS loci, we

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

identified significant colocalization evidence (PPFC >0.7) (Supplementary Table 16 and Supplementary Figure 7). At eight loci, the SNPs with the highest PPFC were the sentinel SNPs in the GWAS.

Candidate genes implicated by gene-expression associations with the target CpGs

At the Bonferroni threshold ($6.93x10^{-5}$=0.05/722 cis-genes in ±5Mb of the target CpGs), we identified significant associations between cg23586595 and *PLAC8* (P=$2.98x10^{-7}$), and between cg24202936 and *F2* (P=$6.39\ x10^{-5}$) (Table 3). Adjusting for additional covariates (smoking status and BMI) did not change these associations.

Cg24202936, cg01506471, and cg06809326 showed significant correlation estimates (|r|>0.3) between blood and brain (r=0.33, 0.87, and 0.57, respectively) (Supplementary Table 17) and, thus, were tested for colocalization. We found that mQTLs for cg24202936 and WMH GWAS SNPs colocalize with *FOLH1* expression in dorsolateral prefrontal cortex (DLPFC) (PPFC=0.75) (Supplementary Table 18). Also, suggestive evidence existed for colocalization of cg06809326 mQTLs, *CCDC144NL-AS1* eQTLs, and WMH SNPs (PPFC=0.69).

Integrative cross-omics analysis

Integrative cross-omics analysis identifies novel gene regulatory networks

At FDR <0.05, we identified 576 WMH-associated gene sets enriched from the integrated data of GWAS, EWAS, and TWAS out of 12,303 gene sets from curated databases.[45,63–66] Top associated gene-sets includes "regulation of actin cytoskeleton" (P=$1.14x10^{-45}$, 211 genes), "telomeres, telomerase, cellular aging, and immortality" (P=$1.10x10^{-35}$, 18 genes), "integrin-mediated cell surface interactions" (P=$3.17x10^{-34}$, 84 genes), "thrombin signaling through proteinase activated receptors" (P=$1.41x10^{-33}$, 32 genes), and "Nef protein mediated CD4 down-regulation" (P=$4.70x10^{-32}$, nine genes). All enriched pathways with FDR P <0.05 are listed in Supplementary Table 19.

We derived two WMH burden-associated gene networks in brain. The first network is comprised of four sub-networks. Five key driver genes (*FMOD, COL3A1, SERPING1, SLC13A4,*

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

and *ISLR*) represent a sub-network of "extracellular matrix (ECM) organization, ECM-receptor interaction, focal adhesion, and collagen formation". Additionally, three related sub-networks, "smooth muscle contraction" with key driver *TAGLN;* "G-protein-coupled receptor (GPCR) ligand binding" with key drivers *GAL*, *ECEL1*, *ESR1*, and *NTS;* and "cytokine signaling in immune system" with key drivers *IFIT1* and *RTP4*, make up the network (Figure 2 and Supplementary Table 20). We also identified an independent second network associated with "lipid and lipoprotein metabolism", with key driver gene *KNG1*. Genes included in each subnetwork are presented in Supplementary Table 21.

Overlap-based drug repositioning analysis of WMH-associated genes

Using drug signatures derived from *in vivo* cardiovascular and nervous system data, we predicted antihyperlipidemic drugs, including PPAR-α (peroxisome proliferator-activated receptor-alpha) agonist "fenofibrate", as the top therapeutic target. Using drug signatures derived from murine oligodendroglial precursor cells data, we predicted several small molecules, including a glycogen synthase kinase inhibitor and a phenylalanyl tRNA synthetase inhibitor that may have therapeutic potential for Alzheimer's disease[83] and autoimmune diseases[84], respectively (Supplementary Table 22).

Discussion

This first EWAS of WMH burden in 9,732 middle-aged to older adults from 14 community-based cohorts identified several novel epigenetic loci. Although we could not independently replicate the association of single CpGs with WMH, likely due to a limited sample size and differences between the discovery and replication sample, functional annotation and bioinformatic analyses provided strong supportive evidence. Moreover, powerful DMR analyses identified 46 DMRs of which 13 were replicated. Integrative analyses of multi-omics information also suggested novel gene networks with key drivers and potential drug targets for WMH.

We identified a novel epigenetic locus, cg24204936, mapping to a pseudogene *SEPTIN7P11*. Functional integration revealed two candidate genes whose expression may be influenced by variation in DNAm at this locus: *F2* in blood and *FOLH1* in DLPFC. Prothrombin encoded by *F2* plays an essential role in blood clot formation, angiogenesis, tissue repair and

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

vascular integrity. A prothrombotic state or circulating prothrombin has been reported for symptomatic cSVD,[85,86] WMH and stroke[87,88]. However, it remains unclear whether coagulation plays a major role in the etiology of WMH or is secondary to injury to the cerebral small vessels and white matter.[89] *FOLH1* encodes glutamate carboxypeptidase II that catalyzes the hydrolysis of N-acetylaspartylglutamate (NAAG). An elevated level of NAAG in the cerebrospinal fluid has been reported in two patients with almost complete absence of myelin in the central nervous system[90] and has been proposed as a diagnostic biomarker for rare diseases of the white matter.[91]

An epigenetic locus mapping to *PRMT1*, which encodes a protein arginine N-methylase, was identified in single-CpG and DMR analyses and also as a shared epigenetics locus with BP. The biological link between DNAm at *PRMT1* and WMH burden may involve pathways related to endothelial dysfunction, which have previously been implicated in WMH etiology.[92] *PRMT1*, a predominant member of the PRMT family, methylates histone and non-histone proteins to regulate various cellular functions.[93] *PRMT1* is essential for the development of neurons, astrocytes, and oligodendrocytes and is critical for myelin formation.[94] PRMTs also catalyze the formation of ADMA (asymmetric dimethylarginine), which reduces nitric oxide production, promotes endothelial dysfunction in the blood-brain barrier (BBB), and triggers the immune response in atherosclerosis.[92,95,96] Higher ADMA levels have been repeatedly associated with cSVD and its monogenic form, CADASIL (Cerebral Autosomal Dominant Arteriopathy with Sub-cortical Infarcts and Leukoencephalopathy).[97–103]

Single-CpG association combined with functional genomic analyses and DMR analyses identified a novel epigenetic locus near *CCDC144NL;CCDC144NL-AS1* (coiled-coil domain containing 144 family and its antisense RNA1). Cg06809326 is under strong cis-genetic control and brain expression of its nearest gene,*CCDC144NL;CCDC144NL-AS1*, may mediate the association between DNAm and WMH burden (Supplementary Figure 8). A TWAS of WMH using blood gene expression data[60] did not report a significant association for *CCDC144NL;CCDC144NL-AS1* expression, possibly due to its low expression in blood. *CCDC144NL-AS1* encodes a long non-coding mRNA transcript that controls expression of target genes by acting as a molecular sponge for various regulatory miRNAs.[104–109] *In vitro* studies have uncovered several of its target genes with potentially relevant function to cSVD. These include matrix metalloproteinases MMP2 and MMP9[108]; F-actin and vimentin[110]; and transforming growth factor beta (TGF-β)-activated kinase 1 (TAK1).[106] MMP2 and MMP9can damage the

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

BBB by triggering recruitment of immune cells[111] and have been implicated in white matter injury and cSVD.[112] F-actin plays an important role in maintaining the shape of endothelial cells and the integrity of the BBB.[113] Disturbed TGF-β signaling has been implicated in the pathogenesis of several monogenic forms of cSVD.[114–117] Deficiency of TAK1 in mouse brain endothelial cells resulted in endothelial cell death, small vessel rarefaction, and disruption of the BBB.[118]

A central role of endothelial dysfunction, possibly resulting in a compromised BBB, in WMH burden[119] is further suggested by identified DNAm associations in genes involved in cell junctions. Claudins are integral membrane proteins that comprise tight junctions specifically in brain microvascular endothelial[120] cells and that regulate BBB permeability.[121] Claudin-5 mapped to cg17577122 is the most enriched tight junction protein in the BBB, and its dysfunction has been implicated in neurodegenerative and neuroinflammatory diseases, and cSVD.[122–126] A recent DMR analysis using DLPFC DNAm levels also identified *CLDN5* to be associated with cognitive decline.[127] In normotensive subjects, we identified a CpG in cadherin 18 (*CDH18*) that encodes an adherens junction protein, which mediates calcium-dependent cell-cell adhesion. *CDH18* is also involved in cell junction organization process and in cell signaling pathways including G-proteins signaling together with *F2*.[128]

Our DMR analysis, aggregated epigenetic associations using Brown's method, [59] and *moloc* analysis using brain QTL data consistently identified an epigenetic association at a known WMH GWAS locus, *SH3PXD2A* (SH3 and PX-domain-containing protein 2A).[8,72–74] Several genome-wide associations with WMH-related traits have also been reported at *SH3PXD2A*, including white matter microstructure[129] and stroke[130,131]. *SH3PXD2A* encodes an adaptor protein (TKS5) involved in the formation of podosomes that act as sites of close contact to as well as degradation of ECM.[132] Gene set enrichment of identified DNAm loci and integrative cross-omics analyses collectively point to a central role of the ECM in WMH burden. One of the two WMH burden networks identified through the Mergeomics approach centered around key driver genes involved in ECM organization and function and the top associated module was "regulation of actin cytoskeleton". Notably, actin polymerization and disassembly of junctional proteins within microvascular endothelial cells were shown to play a key role in early BBB disruption in a murine model.[133]

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

Another network includes genes that function in lipid and lipoprotein metabolism and our overlap-based drug repositioning analyses suggested antihyperlipidemic drugs as potential drug targets. A recent MR analysis showed that genetically increased high-density lipoprotein cholesterol (HDL-C) level was associated with lower WMH volume and lower risk of small vessel stroke.[134] Statin therapy for cSVD has also been regarded as promising since individuals with high WMH burden typically carry higher vascular risk factors. Few randomized clinical trials assessing the effect of lipid lowering on WMH progression have been conducted and they have generally provided mixed results.[135–137] While they suggest a possible role of statins, in particular rosuvastatin, in preventing WMH progression, the lack of high-quality data prevents strong evidence-based recommendation at this time.[138] It has been postulated that statins improve endothelial function and stabilize the BBB in cSVD.[139,140] Studies that investigated membrane proteins including phospholipid flippase (ATP11B) and aquaporin-4 showed that the loss of these proteins cause pathological features of cSVD including endothelial cell dysfunction with reduced tight junctions, nitric oxide, oligodendrocyte progenitor cell maturation block and microglial activation.[126,141]

Finally, this study provides further emphasis concerning the long-observed perivascular inflammation as an additional crucial player in cSVD pathology and provides a possible explanation. Interestingly, gene set enrichment analyses identified a possible role of the defense response to viral infection with several DMR-associated genes related to interferon gamma signaling and the innate immune response (*DTX3L-PARP9, BNIP3,* and *IFITM1)*. Our top associated CpG has been previously reported in an EWAS of chronic HIV infection[71] and our drug-repositioning analysis also identified a HIV antiviral as a possible drug target. Several studies have reported that people with HIV are at higher risk of an increased burden of WMH compared to uninfected controls.[142,143]

Several limitations of our study must be acknowledged. First, many of our EWAS discoveries were not independently confirmed. Since a series of functional analyses showed biological relevance, we suspect that the lack of replication may stem from the limited size of the replication sample and from differences between the discovery and replication samples as hinted by the increased heterogeneity in the DNAm association observed in the meta-analysis (Supplementary Table 2). Indeed, variation in WMH burden was smaller in the replication studies

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

than in the discovery studies perhaps due to the younger age of the participants. The younger cohorts, CARDIA (n=277) with a mean age 53.9 years and SHIP (n=214) with a mean age 49.7 years, make up only 8.59% of the discovery sample; whereas the Rhineland Study with a mean age 54.1 years and FHS 3rd generation cohort with a mean age of 47.1 years, make up over 86% of the replication sample. Replication of several WMH-associated loci identified through more powerful DMR analyses further underscore an underpowered replication study for single CpG associations. Additional studies are needed to confirm the findings presented here. Second, we conducted a subgroup analysis stratified by hypertension status, but statistical power in each stratum was limited. A more ideal design to study this and other modifiable risk factors of cSVD will be a longitudinal study or a stratified association study with a larger sample size. Similarly, our study was not sufficiently powered to examine ancestry-specific associations of DNAm with WMH and possible ancestry difference in epigenetic patterns could not be investigated. Third, we did not adjust for additional lifestyle factors or comorbidities to maximize our sample size by minimizing the number of covariates in the models. Our primary goal was to identify novel DNAm loci associated with WMH burden and we cannot exclude the possibility that the identified loci may reflect, in part, variation in those risk factors. Fourth, the currently publicly available brain QTL data are limited to cis-regions of omics markers and, thus, our *in-silico* bioinformatics analyses were restricted only to the CpGs with substantial cis-acting genetic influence. For example, cg17417856 in *PRMT1* had a strong heritability estimate (h^2=0.40, P=1.37x10^{-8}) but was not followed-up because it was under polygenic control. Lastly, the study was conducted in blood and cell type-specific associations, most notably in brain, may have been missed. To extrapolate the findings in blood to brain, we assessed the correlation with DNAm in brain, and utilized available brain QTL data. Due to the difficulties of getting both brain DNAm and MRI data from a large population-based sample, an EWAS of WMH burden using brain DNAm may not be easy to achieve. However, findings from this large blood-based study may provide a basis for an epigenetic candidate gene study in the brain.

Acknowledgements

The authors thank the staff and participants of ARIC study, BBMRI, CARDIA, CHS, FHS, GENOA, LBC1936, RS, SHIP, ADNI, Rhineland Study, OATS, and BRIDGET for their pivotal contributions. We acknowledge collaborative contributions from the neuroCHARGE

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

working group. We also acknowledge Dr. Dan Levy's team (Drs. Roby Joehanes and Tianxiao Huan) for the mQTL and eQTL data used in this study. A full set of study-specific acknowledgments is provided in the Supplementary Data I**.**

Funding

We thank all study participants for contributing to this work. This project was largely supported by grant R01NS087541 from US National Institute of Neurological Disorders and Stroke with additional support from U01AG052409 and R01HL105756. A full set of study-specific funding sources is provided in the Supplementary Data I**.**

Competing interests

Dr. DeCarli is a consultant of Novartis pharmaceuticals on a safety study for heart failure. Dr. Wardlaw is supported by the UK Dementia Research Institute which is funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK. Dr. Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Dr. Grabe has received travel grants and speaker's honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag as well as research funding from Fresenius Medical Care.

Supplementary material

Supplementary material is available at *Brain* online.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

References

1. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol*. 2010;9(7):689-701.

2. Wardlaw JM, Benveniste H, Williams A. Cerebral Vascular Dysfunctions Detected in Human Small Vessel Disease and Implications for Preclinical Studies. *Annu Rev Physiol*. 2022;84(1):409-434.

3. Carmelli D, DeCarli C, Swan GE, *et al*. Evidence for genetic variance in white matter hyperintensity volume in normal elderly male twins. *Stroke*. 1998;29(6):1177-1181.

4. Atwood LD, Wolf PA, Heard-Costa NL, *et al*. Genetic variation in white matter hyperintensity volume in the Framingham Study. *Stroke*. 2004;35(7):1609-1613.

5. Kochunov P, Glahn D, Winkler A, *et al.* Analysis of genetic variability and whole genome linkage of whole-brain, subcortical, and ependymal hyperintense white matter volume. *Stroke*. 2009;40(12):3685-3690.

6. Turner ST, Jack CR, Fornage M, Mosley TH, Boerwinkle E, de Andrade M. Heritability of leukoaraiosis in hypertensive sibships. *Hypertension*. 2004;43(2):483-487.

7. Duperron MG, Tzourio C, Sargurupremraj M, *et al*. Burden of Dilated Perivascular Spaces, an Emerging Marker of Cerebral Small Vessel Disease, Is Highly Heritable. *Stroke*. 2018;49(2):282-287.

8. Sargurupremraj M, Suzuki H, Jian X, *et al*. Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun*. 2020;11(1):6285.

9. Jian X, Satizabal CL, Smith A v, *et al*. Exome Chip Analysis Identifies Low-Frequency and Rare Variants in MRPL38 for White Matter Hyperintensities on Brain Magnetic Resonance Imaging. *Stroke*. 2018;49(8):1812-1819.

10. Handel AE, Ebers GC, Ramagopalan S v. Epigenetics: molecular mechanisms and implications for disease. *Trends Mol Med*. 2010;16(1):7-16.

11. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33(3S):245-254.

12. Wright JD, Folsom AR, Coresh J, *et al.* The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. *J Am Coll Cardiol*. 2021;77(23):2939-2959.

13. BBMRI Stakeholder's Forum. BBMRI: A Step Closer - Stakeholder's Forum Report.; 2010. Accessed February 14, 2022. https://www.bbmri-eric.eu/wp-content/uploads/2016/07/stakeholders-forum-report-a-step-closer-a4.pdf

14. Fried LP, Borhani NO, Enright P, *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1(3):263-276.

15. Friedman GD, Cutter GR, Donahue RP, *et al.* CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol*. 1988;41(11):1105-1116.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

16. Dawber TR, Kannel WB. The Framingham study. An epidemiological approach to coronary heart disease. *Circulation.* 1966;34(4):553-555. doi:10.1161/01.cir.34.4.553

17. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* 1975;4(4):518-525.

18. Daniels PR, Kardia SLR, Hanis CL, *et al.* Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am J Med*. 2004;116(10):676-681.

19. Wardlaw JM, Bastin ME, Valdés Hernández MC, *et al.* Brain aging, cognition in youth and old age and vascular disease in the Lothian Birth Cohort 1936: rationale, design and methodology of the imaging protocol. *Int J Stroke*. 2011;6(6):547-559.

20. Deary IJ, Gow AJ, Taylor MD, *et al*. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr*. 2007;7(1):1-12.

21. Ikram MA, Brusselle G, Ghanbari M, *et al*. Objectives, design and main findings until 2020 from the Rotterdam Study. *Eur J Epidemiol*. 2020;35(5):483-517.

22. Ikram MA, van der Lugt A, Niessen WJ, *et al*. The Rotterdam Scan Study: design update 2016 and main findings. *Eur J Epidemiol*. 2015;30(12):1299-1315. doi:10.1007/s10654-015-0105-7

23. Volzke H, Alte D, Schmidt CO, *et al.* Cohort profile: the study of health in Pomerania. *Int J Epidemiol.* 2011;40(2):294-307.

24. Khachaturian ZS. Perspective on the Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement*. 2010;6(3):199-201.

25. Petersen RC, Aisen PS, Beckett LA, *et al.* Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology*. 2010;74(3):201-209.

26. Splansky GL, Corey D, Yang Q, *et al*. The third generation cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*. 2007;165(11):1328-1335.

27. Sachdev PS, Lee T, Lammel A, *et al*. Cognitive functioning in older twins: The Older Australian Twins Study. *Australas J Ageing.* 2011;30(s2):17-23.

28. Sachdev PS, Lammel A, Trollor JN, *et al.* A comprehensive neuropsychiatric study of elderly twins: the Older Australian Twins Study. *Twin Res Hum Genet*. 2009;12(6):573-582.

29. Breteler MMB, Wolf H. P2-135: The Rhineland study: a novel platform for epidemiologic research into Alzheimer disease and related disorders. *Alzheimers Dement*. 2014;10:P520.

30. University of Bordeaux. BRIDGET: BRain Imaging, cognition, Dementia and next generation GEnomics. Accessed February 14, 2022. https://bridget.u-bordeaux.fr/Project

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

31. Houseman EA, Accomando WP, Koestler DC, *et al*. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.

32. Mishra A, Chauhan G, Violleau MH, *et al.* Association of variants in HTRA1 and NOTCH3 with MRI-defined extremes of cerebral small vessel disease in older subjects. *Brain.* 2019;142(4):1009-1023.

33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.

34. Pereira T v, Patsopoulos NA, Salanti G, Ioannidis JPA. Discovery properties of genome-wide association signals from cumulatively combined data sets. *Am J Epidemiol.* 2009;170(10):1197-1206.

35. Chen Y an, Lemire M, Choufani S, *et al*. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* 2013;8(2):203-209.

36. Martin TC, Yet I, Tsai PC, Bell JT. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinformatics.* 2015;16(1):131.

37. Boyle AP, Hong EL, Hariharan M, *et al*. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22(9):1790-1797.

38. Fishilevich S, Nudel R, Rappaport N, *et al*. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database.* 2017:bax028.

39. Battram T, Yousefi P, Crawford G, *et al*. The EWAS Catalog: A Database of Epigenome-wide Association Studies. *OSF Preprints*. [Preprint] doi:10.31219/osf.io/837wn

40. Li M, Zou D, Li Z, *et al.* EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res*. 2019;47(D1):D983-D988. doi:10.1093/nar/gky1027

41. Pedersen BS, Schwartz DA, Yang I v, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics.* 2012;28(22):2986-2988.

42. Peters TJ, Buckley MJ, Statham AL, *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin.* 2015;8(1):6.

43. Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102(43):15545.

44. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417-425.

45. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30.

46. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics.* 2016;32(2):286-288.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

47. Debette S, Schilling S, Duperron MG, Larsson SC, Markus HS. Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. *JAMA Neurol*. 2019;76(1):81-94.

48. Prins ND, Scheltens P. White matter hyperintensities, cognitive impairment and dementia: an update. *Nat Rev Neurol.* 2015;11:157.

49. Prins ND, van Dijk EJ, den Heijer T, *et al*. Cerebral small-vessel disease and decline in information processing speed, executive function and memory. *Brain*. 2005;128(9):2034-2041.

50. Richard MA, Huan T, Ligthart S, *et al.* DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. *Am J Hum Genet*. 2017;101(6):888-902.

51. Ray D, Boehnke M. Methods for meta-analysis of multiple traits using GWAS summary statistics. *Genet Epidemiol.* 2018;42(2):134-145.

52. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* 2013;178(7):1177-1184.

53. Huan T, Joehanes R, Song C, *et al*. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat Commun*. 2019;10(1):4267.

54. Traylor M, Tozer DJ, Croall ID, *et al*. Genetic variation in PLEKHG1 is associated with white matter hyperintensities (n = 11,226). *Neurology*. 2019;92(8):e749-e757.

55. Hemani G, Zheng J, Elsworth B, *et al*. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7.

56. Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics.* 2015;10(11):1024-1032.

57. Ng B, White CC, Klein HU, *et al*. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci.* 2017;20(10):1418-1426.

58. Giambartolomei C, Zhenli Liu J, Zhang W, *et al.* A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics.* 2018;34(15):2538-2545.

59. Brown MB. 400: A method for combining non-independent, one-sided tests of significance. *Biometrics.* 1975;31:987-992.

60. Lin H, Satizabal C, Xie Z, *et al.* Whole blood gene expression and white matter Hyperintensities. *Mol Neurodegener.* 2017;12(1):67.

61. Shu L, Zhao Y, Kurt Z, *et al*. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC Genomics*. 2016;17(1):874.

62. Liu Y, Ding J, Reynolds LM, *et al*. Methylomics of gene expression in human monocytes. *Hum Mol Genet*. 2013;22(24):5065-5074.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

63. Jassal B, Matthews L, Viteri G, *et al*. The reactome pathway knowledgebase. *Nucleic Acids Res.* 2020;48(D1):D498-D503.

64. Nishimura D. BioCarta. *Biotech Software & Internet Report: The Computer Software Journal for Scient.* 2001;2(3):117-120.

65. Carbon S, Douglass E, Good BM, *et al*. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 2021;49(D1):D325-D334.

66. Ashburner M, Ball CA, Blake JA, *et al.* Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25(1):25-29.

67. Arneson D, Bhattacharya A, Shu L, Makinen VP, Yang X. Mergeomics: a web server for identifying pathological pathways, networks, and key regulators via multidimensional data integration. *BMC Genomics.* 2016;17(1):722.

68. Zhu J, Lum PY, Lamb J, *et al*. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet Genome Res.* 2004;105(2-4):363-374.

69. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504.

70. Chen YW, Diamante G, Ding J, *et al*. PharmOmics: A Species- and Tissue-specific Drug Signature Database and Online Tool for Drug Repurposing. *bioRxiv.* [Preprint] doi:10.1101/837773

71. Gross AM, Jaeger PA, Kreisberg JF, *et al*. Methylome-wide Analysis of Chronic HIV Infection Reveals Five-Year Increase in Biological Age and Epigenetic Targeting of HLA. *Mol Cell.* 2016;62(2):157-168.

72. Persyn E, Hanscombe KB, Howson JMM, Lewis CM, Traylor M, Markus HS. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat Commun*. 2020;11(1):2175.

73. Verhaaren BFJ, Debette S, Bis JC, *et al*. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ Cardiovasc Genet*. 2015;8(2):398-409.

74. Armstrong NJ, Mather KA, Sargurupremraj M, *et al*. Common Genetic Variation Indicates Separate Causes for Periventricular and Deep White Matter Hyperintensities. *Stroke*. 2020;51(7):2111-2121.

75. Giri A, Hellwege JN, Keaton JM, *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet*. 2019;51(1):51-62. doi:10.1038/s41588-018-0303-9

76. Kichaev G, Bhatia G, Loh PR, *et al*. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet*. 2019;104(1):65-75.

77. Sakaue S, Kanai M, Tanigawa Y, *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53(10):1415-1424.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

78. Warren HR, Evangelou E, Cabrera CP, *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet.* 2017;49(3):403-415.

79. Takeuchi F, Akiyama M, Matoba N, *et al.* Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nat Commun.* 2018;9(1):5052.

80. Kato N, Takeuchi F, Tabara Y, *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet.* 2011;43(6):531-538.

81. Zhu Z, Wang X, Li X, *et al.* Genetic overlap of chronic obstructive pulmonary disease and cardiovascular disease-related traits: a large-scale genome-wide cross-trait analysis. *Respir Res.* 2019;20(1):64.

82. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318-1330.

83. Lauretti E, Dincer O, Praticò D. Glycogen synthase kinase-3 signaling in Alzheimer's disease. Biochim Biophys *Acta Mol Cell Res.* 2020;1867(5):118664.

84. Kwon NH, Fox PL, Kim S. Aminoacyl-tRNA synthetases as therapeutic targets. *Nat Rev Drug Discov.* 2019;18(8):629-650.

85. Tomimoto H, Akiguchi I, Ohtani R, *et al.* The Coagulation-Fibrinolysis System in Patients With Leukoaraiosis and Binswanger Disease. *Arch Neurol.* 2001;58(10):1620-1625.

86. Wiseman S, Marlborough F, Doubal F, Webb DJ, Wardlaw J. Blood Markers of Coagulation, Fibrinolysis, Endothelial Dysfunction and Inflammation in Lacunar Stroke versus Non-Lacunar Stroke and Non-Stroke: Systematic Review and Meta-Analysis. *Cerebrovasc Dis.* 2014;37(1):64-75.

87. Nagai M, Hoshide S, Kario K. Association of Prothrombotic Status With Markers of Cerebral Small Vessel Disease in Elderly Hypertensive Patients. *Am J Hypertens.* 2012;25(10):1088-1094.

88. Kario K, Yano Y, Matsuo T, Hoshide S, Eguchi K, Shimada K. Additional impact of morning haemostatic risk factors and morning blood pressure surge on stroke risk in older Japanese hypertensive patients. *Eur Heart J.* 2011;32(5):574-580.

89. Markus HS, Hunt B, Palmer K, Enzinger C, Schmidt H, Schmidt R. Markers of Endothelial and Hemostatic Activation and Progression of Cerebral White Matter Hyperintensities. *Stroke.* 2005;36(7):1410-1414.

90. Wolf NI, Willemsen MAAP, Engelke UF, *et al.* Severe hypomyelination associated with increased levels of N-acetylaspartylglutamate in CSF. *Neurology.* 2004;62(9):1503.

91. Mochel F, Boildieu N, Barritault J, *et al.* Elevated CSF N-acetylaspartylglutamate suggests specific molecular diagnostic abnormalities in patients with white matter diseases. *Biochim Biophys Acta*. 2010;1802(11):1112-1117.

92. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. *Lancet Neurol.* 2019;18(7):684-696.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

93. Wei H, Mundade R, Lange KC, Lu T. Protein arginine methylation of non-histone proteins and its role in diseases. *Cell Cycle.* 2014;13(1):32-41.

94. Hashimoto M, Murata K, Ishida J, Kanou A, Kasuya Y, Fukamizu A. Severe Hypomyelination and Developmental Defects Are Caused in Mice Lacking Protein Arginine Methyltransferase 1 (PRMT1) in the Central Nervous System. *J Biol Chem.* 2016;291(5):2237-2245.

95. Dowsett L, Higgins E, Alanazi S, Alshuwayer NA, Leiper FC, Leiper J. ADMA: A Key Player in the Relationship between Vascular Dysfunction and Inflammation in Atherosclerosis. *J Clin Med.* 2020;9(9):3026.

96. Watson CP, Pazarentzos E, Fidanboylu M, Padilla B, Brown R, Thomas SA. The transporter and permeability interactions of asymmetric dimethylarginine (ADMA) and L-arginine with the human blood-brain barrier in vitro. *Brain Res*. 2016;1648(Pt A):232-242.

97. Janes F, Cifù A, Pessa ME, *et al.* ADMA as a possible marker of endothelial damage. A study in young asymptomatic patients with cerebral small vessel disease. *Sci Rep*. 2019;9(1):14207.

98. Pikula A, Böger RH, Beiser AS, *et al.* Association of plasma ADMA levels with MRI markers of vascular brain injury: Framingham offspring study. *Stroke.* 2009;40(9):2959-2964.

99. Guan J, Yan C, Gao Q, *et al.* Analysis of risk factors in patients with leukoaraiosis. *Medicine.* 2017;96(8):e6153-e6153.

100. Khan U, Hassan A, Vallance P, Markus HS. Asymmetric Dimethylarginine in Cerebral Small Vessel Disease. *Stroke.* 2007;38(2):411-413.

101. Notsu Y, Nabika T, Bokura H, *et al*. Evaluation of Asymmetric Dimethylarginine and Homocysteine in Microangiopathy-Related Cerebral Damage. *Am J Hypertens*. 2009;22(3):257-262.

102. Gao Q, Fan Y, Mu LY, Ma L, Song ZQ, Zhang YN. S100B and ADMA in cerebral small vessel disease and cognitive dysfunction. *J Neurol Sci.* 2015;354(1):27-32.

103. Rufa A, Blardi P, de Lalla A, *et al.* Plasma Levels of Asymmetric Dimethylarginine in Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarct and Leukoencephalopathy. *Cerebrovasc Dis.* 2008;26(6):636-640.

104. Guo L, Peng Y, Meng Y, *et al.* Expression profiles analysis reveals an integrated miRNA-lncRNA signature to predict survival in ovarian cancer patients with wild-type BRCA1/2. *Oncotarget.* 2017;8(40):68483-68492.

105. He J, Guan J, Liao S, *et al*. Long Noncoding RNA CCDC144NL-AS1 Promotes the Oncogenicity of Osteosarcoma by Acting as a Molecular Sponge for microRNA-490-3p and Thereby Increasing HMGA2 Expression. *Onco Targets Ther*. 2021;14:1-13.

106. Fan H, Ge Y, Ma X, *et al.* Long non-coding RNA CCDC144NL-AS1 sponges miR-143-3p and regulates MAP3K7 by acting as a competing endogenous RNA in gastric cancer. *Cell Death Dis.* 2020;11(7):521.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

107. Zhang L, Chi B, Chai J, *et al*. LncRNA CCDC144NL-AS1 Serves as a Prognosis Biomarker for Non-small Cell Lung Cancer and Promotes Cellular Function by Targeting miR-490-3p. *Mol Biotechnol.* 2021;63(10):933-940.

108. Zhang Y, Zhang H, Wu S. LncRNA-CCDC144NL-AS1 Promotes the Development of Hepatocellular Carcinoma by Inducing WDR5 Expression via Sponging miR-940. *J Hepatocell Carcinoma.* 2021;8:333-348.

109. Niu P, Yao B, Wei L, Zhu H, Fang C, Zhao Y. Construction of prognostic risk prediction model based on high-throughput sequencing expression profile data in childhood acute myeloid leukemia. *Blood Cells Mol Dis.* 2019;77:43-50.

110. Zhang C, Wu W, Zhu H, *et al*. Knockdown of long noncoding RNA CCDC144NL-AS1 attenuates migration and invasion phenotypes in endometrial stromal cells from endometriosis. *Biol Reprod.* 2019;100(4):939-949.

111. Anwar MM, Özkan E, Gürsoy-Özdemir Y. The role of extracellular matrix alterations in mediating astrocyte damage and pericyte dysfunction in Alzheimer's disease: A comprehensive review. *Eur J Neurosci*. 2021:1-23.

112. Montaner J, Ramiro L, Simats A, *et al.* Matrix metalloproteinases and ADAMs in stroke. *Cell Mol Life Sci.* 2019;76(16):3117-3140.

113. Prasain N, Stevens T. The actin cytoskeleton in endothelial cell phenotypes. *Microvasc Res.* 2009;77(1):53-63.

114. Hara K, Shiga A, Fukutake T, *et al.* Association of HTRA1 Mutations and Familial Ischemic Cerebral Small-Vessel Disease. *New Engl J Med.* 2009;360(17):1729-1739.

115. Kast J, Hanecker P, Beaufort N, *et al.* Sequestration of latent TGF-β binding protein 1 into CADASIL-related Notch3-ECD deposits. *Acta Neuropathol Commun*. 2014;2:96.

116. Hamel E. Cerebral Circulation: Function and Dysfunction in Alzheimer's Disease. *J Cardiovasc Pharm.* 2015;65(4).

117. Smahi A, Courtois G, Vabres P, *et al.* Genomic rearrangement in NEMO impairs NF-κB activation and is a cause of incontinentia pigmenti. *Nature.* 2000;405(6785):466-472.

118. Ridder DA, Wenzel J, Müller K, *et al*. Brain endothelial TAK1 and NEMO safeguard the neurovascular unit. *J Exp Med.* 2015;212(10):1529-1549.

119. Wardlaw JM, Makin SJ, Valdés Hernández MC, *et al.* Blood-brain barrier failure as a core mechanism in cerebral small vessel disease and dementia: evidence from a cohort study. *Alzheimers Dement.* 2017;13(6):634-643.

120. Sukriti S, Tauseef M, Yazbeck P, Mehta D. Mechanisms regulating endothelial permeability. *Pulm Circ.* 2014;4(4):535-551.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

121. Lippmann ES, Azarin SM, Kay JE, *et al*. Derivation of blood-brain barrier endothelial cells from human pluripotent stem cells. *Nat Biotechnol*. 2012;30(8):783-791.

122. Greene C, Hanley N, Campbell M. Claudin-5: gatekeeper of neurological function. *Fluids Barriers CNS*. 2019;16(1):3.

123. Yang Y, Kimura-Ohba S, Thompson JF, *et al*. Vascular tight junction disruption and angiogenesis in spontaneously hypertensive rat with neuroinflammatory white matter injury. *Neurobiol Dis*. 2018;114:95-110.

124. Berndt P, Winkler L, Cording J, *et al*. Tight junction proteins at the blood-brain barrier: far more than claudin-5. *Cell Mol Life Sci*. 2019;76(10):1987-2002.

125. Bailey EL, Wardlaw JM, Graham D, Dominiczak AF, Sudlow CLM, Smith C. Cerebral small vessel endothelial structural changes predate hypertension in stroke-prone spontaneously hypertensive rats: a blinded, controlled immunohistochemical study of 5- to 21-week-old rats. *Neuropathol Appl Neurobiol*. 2011;37(7):711-726.

126. Rikesh RM, Sophie Q, Silvie RR, *et al*. Reversal of endothelial dysfunction reduces white matter vulnerability in cerebral small vessel disease in rats. *Sci Transl Med*. 2018;10(448):eaam9507.

127. Hüls A, Robins C, Conneely KN, *et al*. Brain DNA Methylation Patterns in CLDN5 Associated With Cognitive Decline. *Biol Psychiat*. 2021;91(4):389-398.

128. Belinky F, Nativ N, Stelzer G, *et al*. PathCards: multi-source consolidation of human biological pathways. *Database*. 2015:bav006.

129. Zhao B, Zhang J, Ibrahim JG, *et al*. Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Mol Psychiatr*. 2021;26(8):3943-3955.

130. Malik R, Chauhan G, Traylor M, *et al*. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50(4):524-537.

131. Ishigaki K, Akiyama M, Kanai M, *et al*. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet*. 2020;52(7):669-679.

132. Murphy DA, Courtneidge SA. The "ins" and "outs" of podosomes and invadopodia: characteristics, formation and function. *Nat Rev Mol Cell Bio*. 2011;12(7):413-426.

133. Shi Y, Zhang L, Pu H, *et al*. Rapid endothelial cytoskeletal reorganization enables early blood-brain barrier disruption and long-term ischaemic reperfusion brain injury. *Nat Commun*. 2016;7:10523.

134. Georgakis MK, Malik R, Anderson CD, Parhofer KG, Hopewell JC, Dichgans M. Genetic determinants of blood lipids and cerebral small vessel disease: role of high-density lipoprotein cholesterol. *Brain*. 2020;143(2):597-610.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

135. Zhang H, Cui Y, Zhao Y, *et al.* Effects of sartans and low-dose statins on cerebral white matter hyperintensities and cognitive function in older patients with hypertension: a randomized, double-blind and placebo-controlled clinical trial. *Hypertens res.* 2019;42(5):717-729.

136. Guo Y, Li Y, Liu X, *et al*. Assessing the effectiveness of statin therapy for alleviating cerebral small vessel disease progression in people ≥75 years of age. *BMC Geriatr.* 2020;20(1):292.

137. ten Dam VH, van den Heuvel DMJ, van Buchem MA, *et al.* Effect of pravastatin on cerebral infarcts and white matter lesions. *Neurology.* 2005;64(10):1807-1809.

138. Wardlaw JM, Debette S, Jokinen H, *et al.* ESO Guideline on covert cerebral small vessel disease. *Eur Stroke J.* 2021;6(2):CXI-CLXII.

139. Zhang CE, Wong SM, van de Haar HJ, *et al.* Blood–brain barrier leakage is more widespread in patients with cerebral small vessel disease. *Neurology.* 2017;88(5):426.

140. Giannopoulos S, Katsanos AH, Tsivgoulis G, Marshall RS. Statins and cerebral hemodynamics. *J Cereb Blood Flow Metab.* 2012;32(11):1973-1976.

141. Holland PR, Searcy JL, Salvadores N, *et al.* Gliovascular disruption and cognitive deficits in a mouse model with features of small vessel disease. *J Cereb Blood Flow Metab.* 2015;35(6):1005-1014.

142. Moulignier A, Savatovsky J, Assoumou L, *et al.* Silent Cerebral Small-Vessel Disease Is Twice as Prevalent in Middle-Aged Individuals With Well-Controlled, Combination Antiretroviral Therapy–Treated Human Immunodeficiency Virus (HIV) Than in HIV-Uninfected Individuals. *Clin Infect Dis.* 2018;66(11):1762-1769.

143. Mina Y, Wu T, Hsieh HC, *et al.* Association of White Matter Hyperintensities With HIV Status and Vascular Risk Factors. *Neurology.* 2021;96(14):e1823-e1834.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

Figure 1 Overview of the study analytic scheme.

Figure 2 WMH-associated Gene Networks. WMH-associated genes based on multi-molecular evidence are organized around the 19 key driver genes. a. WMH-associated network consisting of four sub-networks - extracellular matrix (ECM) organization (*FMOD*, *COL3A1*, *SEPING1*, *SLC13A4*, and *ISLR*); smooth muscle contraction (*TAGLN*); G-protein-coupled receptor (GPCR) ligand binding (*GAL*, *ECEL1*, *ESR1*, and *NTS*) and cytokine signaling in immune system (*IFIT1* and *RTP4*) b. WMH-associated network of lipid and lipoprotein metabolism (*KNG1*). Key drivers and associated gene networks identified in the Mergeomics analysis are colored in orange (RGB: 255,166,127). Neighboring genes are grouped into networks and labelled in random colors.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

Table 1 Single CpG associations with white matter hyperintensities burden in the discovery sample ($P < 1 \times 10^{-5}$)

CpG	Chr:Position (hg19)	Nearest Gene	Reduced model					Full model		
			N	Z	P	Q	FDR	N	Z	P
cg24202936	11:50257256	*SEPTIN7P11*	5359	5.38	7.58×10^{-8}	0.03	0.04	4930	5.28	1.30×10^{-7}
cg17417856	19:50191637	*PRMT1;ADM5*	4917	−4.95	7.42×10^{-7}	0.15	0.28	4526	−4.40	1.11×10^{-5}
cg01506471	7:3990479	*SDK1*	5359	−4.81	1.52×10^{-6}	0.21	0.3	4930	−4.00	6.41×10^{-5}
cg14547240	4:15428750	*C1QTNF7*	5359	−4.71	2.48×10^{-6}	0.25	0.3	4930	−4.17	3.10×10^{-5}
cg21547371	3:52869521	*MUSTN1*	5359	−4.65	3.30×10^{-6}	0.25	0.3	4930	−4.06	4.95×10^{-5}
cg03116124	1:231293208	*TRIM67*	5129	−4.64	3.54×10^{-6}	0.25	0.31	4700	−4.58	4.63×10^{-6}
cg06809326	17:20799526	*CCDC144NL-AS1*	5359	4.57	4.80×10^{-6}	0.28	0.34	4930	3.44	5.88×10^{-4}
cg13476133	7:44185646	*GCK*	5359	4.55	5.46×10^{-6}	0.28	0.36	4930	4.03	5.65×10^{-5}
cg14133539	9:104568	*FOXD4*	4917	−4.53	5.98×10^{-6}	0.28	0.38	4526	−4.45	8.41×10^{-6}
cg17577122	22:19511967	*CLDN5*	5359	4.50	6.88×10^{-6}	0.29	0.4	4930	4.79	1.68×10^{-6}
cg23586595	4:84034390	*PLAC8*	5359	4.45	8.45×10^{-6}	0.32	0.43	4930	3.93	8.36×10^{-5}
cg23054394	3:140784675	*SPSB4*	5359	−4.42	9.88×10^{-6}	0.34	0.45	4930	−4.01	6.07×10^{-5}

The reduced model is adjusted for age, sex, study site (if applicable), total (intra)cranial volume (cm^3), white blood cell proportion (%), technical covariates and genetic principal components. The full model is additionally adjusted for body mass index, smoking status, and systolic and diastolic blood pressure measures. Chr: chromosome, EA: European ancestry, FDR: local false discovery rate value, N: number of subjects tested for the CpG, P: *P* value, Q: *Q* value, SE: standard error, and Z: Z-score.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

Table 2 Heritability estimates of WMH-associated CpGs

CpG	Nearest Gene	Reduced model			Full Model		
		h^2_{meth}	S.E.	*P*	h^2_{meth}	S.E.	*P*
cg0150647	*SDK1*	0.02	0.07	0.38	0.01	0.07	0.42
cg03116124	*TRIM67*	0.01	0.07	0.45	0.01	0.07	0.47
cg06809326	*CCDC144NL*	0.26	0.07	$1.03 \times 10^{-4*}$	0.27	0.07	$9.51 \times 10^{-5*}$
cg13476133	*GCK*	0.09	0.07	0.11	0.09	0.07	0.12
cg14133539	*FO × D4*	0.08	0.07	0.14	0.07	0.07	0.17
cg14547240	*C1QTNF7*	0.06	0.07	0.20	0.06	0.07	0.18
cg17417856	*PRMT1;ADM5*	0.40	0.08	$1.37 \times 10^{-8*}$	0.40	0.08	$3.06 \times 10^{-8*}$
cg17577122	*CLDN5*	0.14	0.08	2.80×10^{-2}	0.15	0.08	2.27×10^{-2}
cg21547371	*MUSTN1*	0.00	–	0.50	0.00		0.50
cg23054394	*SPSB4*	0.00	–	0.50	0.00		0.50
cg23586595	*PLAC8*	0.24	0.08	$1.47 \times 10^{-3*}$	0.23	0.08	$2.51 \times 10^{-3*}$
cg24202936	*LOC441601*	0.15	0.07	1.34×10^{-2}	0.16	0.07	1.17×10^{-2}

h^2_{meth}: the narrow-sense heritability in an additive genetic model. S.E.: standard error

Reduced model is adjusted for age, sex, blood cell counts, principal components of the ancestry and technical covariates. Full model is additionally adjusted for body mass index and smoking.

*Significant after adjustment for multiple testing burden.

Table 3 Cis-Genes (±5Mb) whose expression is significantly associated with identified CpGs

CpG	Gene	Region (hg19)	n	$Z_{reduced}$	$P_{reduced}$	Z_{full}	P_{full}
cg23586595	*PLAC8*	4:84011211–84138405	2687	−5.13	2.98×10^{-7}	−5.11	3.27×10^{-7}
cg24202936	*F2*	11:46740749–46761054	1963	−4.00	6.39×10^{-5}	−4.01	6.04×10^{-5}

Z scores and P values from the reduced and full model are presented. Reduced model is adjusted for age, sex, blood cell counts, principal components of the ancestry and technical covariates. Full model is additionally adjusted for body mass index and smoking. *PLAC8*: placenta-associated 8 and *F2*: coagulation factor II.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

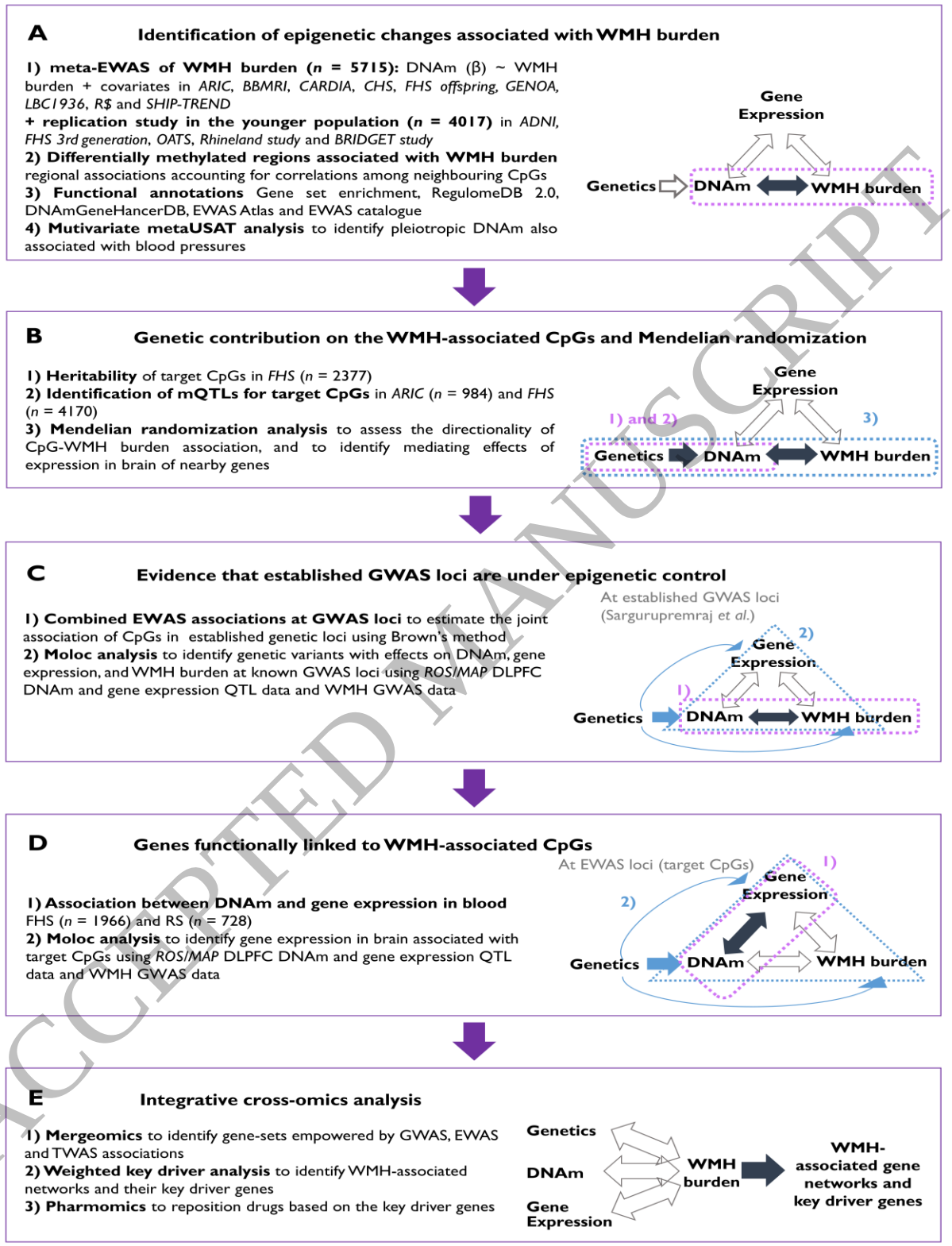

Figure 1

160x241 mm (.46 x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022

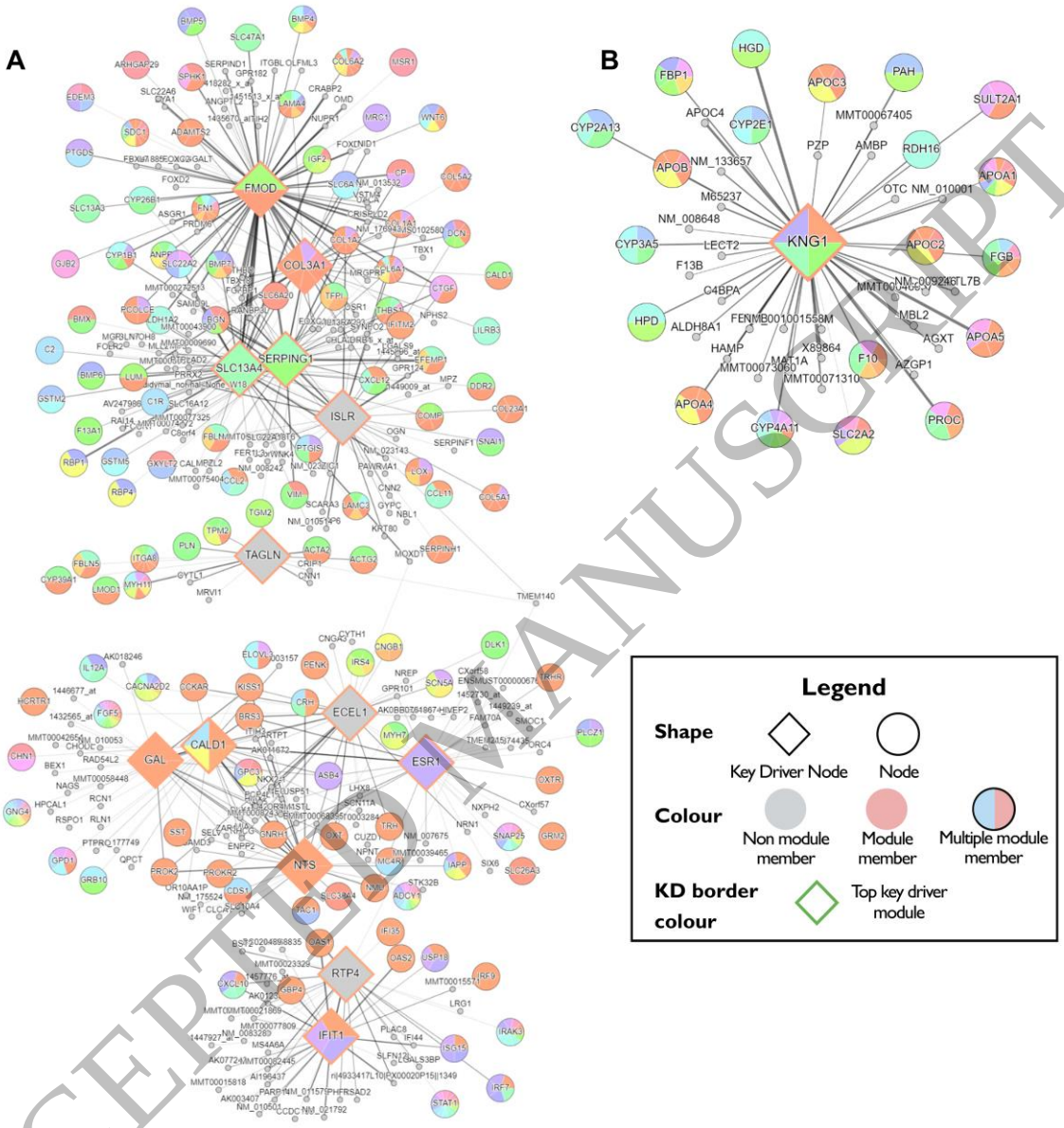

Figure 2
160x158 mm (.46 x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awac290/6659012 by Universite de Bordeaux user on 29 October 2022