

Spring 5-2020

THE ROLE OF DNA METHYLATION IN WHITE MATTER HYPERINTENSITY BURDEN: AN INTEGRATIVE APPROACH

YUNJU YANG

UTHealth School of Public Health

Follow this and additional works at: https://digitalcommons.library.tmc.edu/uthsph_dissertsopen



Part of the [Community Psychology Commons](#), [Health Psychology Commons](#), and the [Public Health Commons](#)

Recommended Citation

YANG, YUNJU, "THE ROLE OF DNA METHYLATION IN WHITE MATTER HYPERINTENSITY BURDEN: AN INTEGRATIVE APPROACH" (2020). *UT School of Public Health Dissertations (Open Access)*. 114.
https://digitalcommons.library.tmc.edu/uthsph_dissertsopen/114

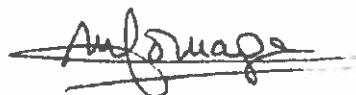
This is brought to you for free and open access by the School of Public Health at DigitalCommons@TMC. It has been accepted for inclusion in UT School of Public Health Dissertations (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

THE ROLE OF DNA METHYLATION IN WHITE MATTER HYPERINTENSITY
BURDEN: AN INTEGRATIVE APPROACH

by

YUNJU YANG, MPH

APPROVED:



MYRIAM FORNAGE, PHD



JAN BRESSLER, PHD



MICHAEL SWARTZ, PHD



DEAN, THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH

Copyright
by
Yunju Yang, MPH, PhD
2020

THE ROLE OF DNA METHYLATION IN WHITE MATTER HYPERINTENSITY

BURDEN: AN INTEGRATIVE APPROACH

by

YUNJU YANG

BPH, KOREA UNIVERSITY, 2011

MPH, SEOUL NATIONAL UNIVERSITY, 2013

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH

Houston, Texas

May, 2020

ACKNOWLEDGEMENTS

I would like to thank the committee members - Drs. Myriam Fornage, Jan Bressler, and Michael Swartz-, and the external reviewer Dr. Han Chen for offering their time, supports and advices. Especially, I would like to express my deep and sincere gratitude to my extraordinary advisor as well as the committee chair, Dr. Fornage, for providing invaluable research opportunities and guidance on this research. Her passion, vision, and motivation has inspired me to pursue her footage. I was privileged to work with her and study under supervision.

Also, I wish to thank the below projects that enabled my research.

The CHARGE consortium: We acknowledge the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium for encouraging CHARGE studies to participate in this effort and for the contributions of CHARGE members to the analyses conducted for this research. Study-specific acknowledgements are included in the Supplemental Material.

The GTEx Project: The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

THE ROLE OF DNA METHYLATION IN WHITE MATTER HYPERINTENSITY
BURDEN: AN INTEGRATIVE APPROACH

Yunju Yang, MPH, PhD
The University of Texas
School of Public Health, 2020

Dissertation Chair: Myriam Fornage, PhD

ABSTRACT

Cerebral white matter hyperintensities (WMH) on MRI are an indicator of cerebral small vessel disease, a major risk factor for vascular dementia and stroke. DNA methylation may contribute to the molecular underpinnings of WMH, which are highly heritable. We performed a meta-analysis of 11 epigenome-wide association studies in 6,019 middle-aged to elderly subjects, who were free of dementia and stroke and were of African (AA) or European (EA) descent. In each study, association between WMH volume and each CpG was tested within ancestry using a linear mixed model, adjusted for age, sex, total intracranial volume, white blood cell count, technical covariates, BMI, smoking and blood pressure (BP). To detect differentially methylated regions (DMRs), we also calculated region-based p-values accounting for spatial correlations among CpGs. No individual CpG reached epigenome-wide significance, but suggestive novel associations were identified with cg17577122 (*CLDN5*, $P=2.39E-7$), cg24202936 (*LOC441601*, $P=3.78E-7$), cg03116124 (*TRIM67*, $P=6.55E-7$), cg04245766 (*BMP4*, $P=3.78E-7$) and cg06809326 (*CCDC144NL*, $P=6.14E-7$). Gene enrichment analyses implicated pathways involved in regulation of cell

development and differentiation, especially of endothelial cells. We identified 11 DMRs ($P_{\text{Sidak}} < 0.05$) and two were mapped to BP-related genes (*HIVEP3*, *TCEA2*). The most significant DMRs were mapped to *PRMT1*, a protein arginine methyltransferase involved in glioblastomagenesis ($P = 7.9\text{E-}12$), and mapped to *CCDC144NL-AS1*, an antisense transcript of *CCDC144NL* ($P = 1.6\text{E-}11$). Genes mapping to DMRs were enriched in biological processes related to lipoprotein metabolism and transport. Bi-directional Mendelian randomization analysis showed that DNA methylation level at cg06809326 influenced WMH burden (OR [95% CI] = 1.7[1.2-2.5], $P = 0.001$) but not the reverse ($P = 0.89$). Additionally, increased methylation at cg06809326 was associated with lower expression of *CCDC144NL* ($P = 3.3\text{E-}2$), and two-step Mendelian randomization analysis supported its mediating role in the association of cg06809326 and WMH burden. *CCDC144NL* is known to be associated with diabetic retinopathy and the coiled coil proteins in general promotes integrin-dependent cell adhesion. Integrin-related pathway was further supported by integrative genetic analyses. In conclusion, we identified novel epigenetic loci associated with WMH burden, and further supported the role of cg06809326 in the WMH etiology implicating integrin-mediated pathology.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT.....	v
Table of Contents	vii
List of Tables	viii
List of Figures	viii
List of Appendices	x
BACKGROUND	1
Literature Review.....	1
Public Health Significance.....	25
Specific Aims and Hypotheses	28
METHODS	30
Aim 1: Meta-analysis of epigenome-wide association studies (EWASs) of WMH burden	30
Aim 2: Genetic control and transcriptional impact of WMH epigenetics	38
Aim 3: Integrative genetic analyses of WMH burden	41
RESULTS	45
Aim 1: Meta-analysis of epigenome-wide association studies (EWASs) of WMH burden	45
Aim 2: Genetic control and transcriptional impact of WMH epigenetics	56
Aim 3: Integrative genetic analyses of WMH burden	63
DISCUSSION	72
CONCLUSIONS.....	79
APPENDICES	82
Appendix A. Supplemental Material	82
Appendix B. Supplemental Figures	103
Appendix C. Supplemental Tables	113
REFERENCES	130

LIST OF TABLES

Table 1 Genome-wide significant loci associated with WMH	20
Table 2 Genomic inflation in meta-analyses	46
Table 3 Top associations from a meta-analysis	47
Table 4 Top associations from a stratified meta-analysis	48
Table 5 DMR analysis result.....	50
Table 6 FUMA gene-enrichment results.....	54
Table 7 Multivariate joint analysis of CA and WMH epigenetic associations.....	55
Table 8 Heritability estimates of WMH-associated methylation probes	57
Table 9 Bi-directional MR analysis result	59
Table 10 Associations between target probes and the nearest genes	60
Table 11 TWAS results in target probes for WMH-associated genes	62
Table 12 Two-step MR analysis result	63
Table 13 Meta-analysis of marker set enrichment.....	64
Table 14 MOLOC result in whole blood	67
Table 15 MOLOC result in brain tissues	70

LIST OF FIGURES

Figure 1 Signs of cerebral small vessel disease Adapted from Inzitari et al, BMJ. 2009 Jul 6;339:b2477. doi: 10.1136/bmj.b2477	4
Figure 2 Blood-brain Barrier (BBB) and the neurovascular unit (NVU) Adapted from Yu Yamazaki et al. 2017.....	9
Figure 3 The schema of overall study designs.....	30

LIST OF APPENDICES

Appendix A. Supplemental Material

Appendix B. Supplemental Figures

Appendix C. Supplemental Tables

BACKGROUND

Literature Review

Cerebral Small Vessel Disease: Definition and classifications

Small vessel disease is a general term for various abnormalities related to small blood vessels often referred to in the cardiovascular system.¹ Blood vessels play a crucial role in the cardiovascular system, and any abnormalities can lead to severe sequelae. Small vessel disease in the heart involves coronary arterioles and the sub-endocardial plexus of microvessels.² Abnormalities developed in the heart vessels can lead to acute or stable coronary syndromes, as well as heart failure in the long term.^{3,4} Cerebral small vessel disease is an umbrella term for a group of abnormalities of the small arteries, arterioles, venules, and capillaries in the brain.⁵ Most often, cerebral small vessel disease studies paid attentions to the arterial part.⁶ A single occlusion of these vessels may result in lacunar stroke syndrome. The main focus of this dissertation is on cerebral small vessel disease (CSVD).

Blood vessels are widely and densely dispersed in the brain. These perforating vessels play a crucial role of carrying essential amount of oxygen and nutrients, thus maintaining optimal functioning of the brain's metabolically active nuclei and complex white matter networks.⁷ The brain greatly depends on the circulatory system in that it receives up to 20% of cardiac output and consumes up to 20% of the total body's oxygen, though it only constitutes ~2% of total body mass.⁸ Arterial blood is supplied to the brain by two major sets of vessels: carotid and vertebral arteries.⁹ Large arteries - anterior cerebral artery (ACA) and middle cerebral artery (MCA) - which supply blood to most parts of the cerebral hemispheres arise from carotid arteries; while posterior cerebral arteries (PCA) are formed from vertebral

arteries. The two major circulation arteries divide into smaller branch vessels to reach out to every corner of the brain. Small vessels in the brain originate from two sources: superficial and deep.^{10–13} Superficial small vessels, stemming from the subarachnoid circulation, pass the cortical layers and meet the deep originated small vessels which have passed the deep grey matter in the deepest areas of the subcortical white matter.

Since abnormalities of brain vessels can be characterized clinically differently by size, they are conveniently called either as a “large” with a diameter larger than 0.5 to 1.0 millimeter or a “small” vessel with a diameter less than 0.5 millimeter. In CSVD, “small” vessels generally means those with a diameter less than 0.2 millimeter.¹⁴ Clinical trials performed to develop the ischemic stroke classification system, also named as the TOAST (Trial of Org 10172 in Acute Stroke Treatment), claimed that stroke caused from large artery atherosclerosis should be differentiated from that caused from small-vessel occlusion.¹⁵ Additionally, two other widely used classification systems distinguish these two origins of ischemic stroke.^{16,17}

Small vessel disease is the most common pathological neurological process;¹⁸ its consequence can be pathologically discerned either as hemorrhagic or ischemic.⁵ A disease with ischemic origin occurs when blood flow is blocked by a blood clot; while that with hemorrhagic origin happens when a weakened blood vessel wall bursts and bleeds into the lesion. Ischemic small vessel disease accounts for a third of acute cerebral ischemic events.¹⁹ In terms of stroke, a recent study on behalf of the American Heart Association reported that 87% of all stroke is ischemic.²⁰ Hemorrhagic stroke accounts for a relatively small portion of all strokes, but the consequence is more acute and devastating.²¹ Ischemic stroke subtypes

can be further divided into large artery atherosclerotic stroke (LAS), cardioembolic stroke (CES), stroke due to small vessel disease (SVS), stroke of other known etiologies, and stroke of undetermined etiologies according to the TOAST criteria. Small vessel disease can explain about a quarter of ischemic strokes.²²⁻²⁴

Small vessel disease systematically affects multiple organs and areas of the body and can be classified based on its etiopathogenesis. There has been proposed six subtypes of small vessel disease; the most prevalent forms are type 1: arteriolosclerosis and type 2: sporadic and hereditary cerebral amyloid angiopathy.⁵ A group of inherited CSVDs such as CADASIL (cerebral autosomal dominant arteriopathy with subcortical ischemic strokes and leukoencephalopathy) and Fabry's disease have been recently reported,²⁵ and these forms are crucial to understand the pathogenesis of sporadic CSVDs.

White matter hyperintensity (WMH)

White matter lesions are confluent areas on brain magnetic resonance image (MRI) that are bilaterally and symmetrically sited in the periventricular and deep white matter of the cerebral hemispheres, the basal ganglia which is deep grey matter, the pons, and occasionally in the other parts of the brainstem and cerebellar white matter.^{5,26} It appears hyperintense on T2-weighted and fluid-attenuated inversion recovery (FLAIR) images (**Figure 1**²⁷) or appears with decreased signal on T1-weighted MR imaging.^{5,26} It is unclear whether differential locations of white matters indicate different disease stages or distinct mechanisms, but white matter tends to be numerous distributed in the cerebral hemispheres first and then to appear in the brainstem.

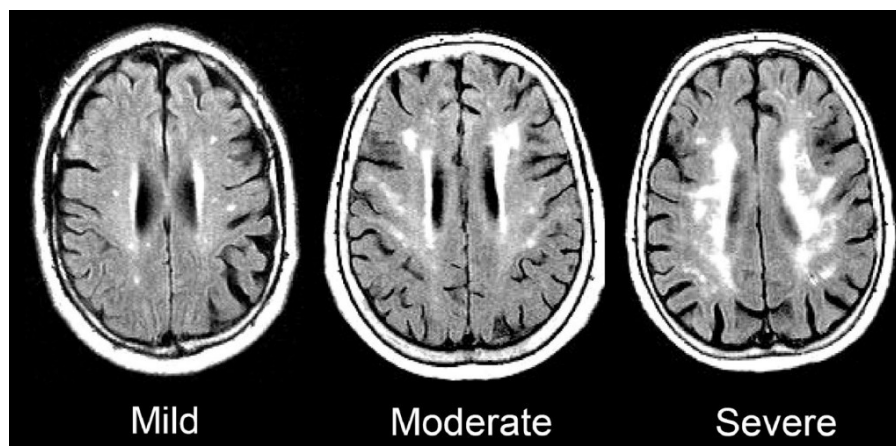


Figure 1 Signs of cerebral small vessel disease Adapted from Inzitari et al, BMJ. 2009 Jul 6;339:b2477. doi: 10.1136/bmj.b2477

The confluent areas in the brain first drew scientific attention more than 30 years ago with the term leukoaraiosis.²⁸ At the time that the term “leukoaraiosis” was introduced, most

studies measured the lesion with computed tomography (CT) imaging, while the most updated white matter lesion studies do so with MRI. The lesions are developed by chronic hypoperfusion of the white matter and disruption of the blood-brain barrier which cause chronic leakage of plasma into the white matter. Pathological studies on the lesions included myelin pallor, tissue rarefaction associated with loss of myelin and axons, and mild gliosis.^{13,29} A further study with consideration of the lesion's location and risk factors found that leukoaraiosis is likely to be another expression of CSVD.³⁰ As MRI has become widely available in various clinical settings, the data on white matter lesions becomes tremendously abundant and white matter lesions are now generally considered a consequence of CSVD.

WMH is widely used as a surrogate of CSVD.^{31–33} Symptomatic presentations can be used as markers of CSVD since it is difficult to visualize small vessels in vivo.³⁴ As mentioned above, CSVD affects the brain vessels which include the perforating cerebral arterioles, capillaries, and venules resulting in brain damage in the cerebral white and deep grey matter.⁵ On conventional brain MRI taken at 1.5 or 3T, CSVDs manifest in several forms such as lacunar infarcts or hemorrhages, lacunes,³⁵ WMH (mostly clinically silent),³⁶ visible perivascular spaces,³⁷ and microbleeds³⁸. As one of the key markers of CSVD, WMH is more common in lacunar stroke, which is a clinical manifestation of CSVD, than other subtypes of stroke.³⁶ Additionally, WMH is correlated with other CSVD correlates *i.e.* lacunes,³⁹ perivascular spaces,³⁷ microbleeds³⁸ and brain atrophy.⁴⁰ With the advance of imaging techniques, large population studies and clinical trials use white matter lesion traits as a surrogate marker of CSVD in that WMH can be measured quantitatively via non-invasively means.³² Epidemiologic studies employ several different endpoints of white

matter. Frequently used terms are white matter hyperintensities (WMH), white matter lesions, white matter progression, and white matter changes. WMH is a term for white matter lesions appearing hyperintense on brain MRI. This dissertation will focus on WMH, which is the most commonly studied marker of CSVD, to provide an additional body of evidence to the current knowledge in CSVD with comparability to other studies.

Epidemiology of WMH

As WMH is highly prevalent in the aging population and associated with age-related neurological changes;^{41,42} it has been hypothesized that WMH is a neuroimaging correlate of age-associated disabilities.⁴³ Even though white matter lesions may develop in young age,^{44,45} prevalence increases with age in a “dose-response relationship”. WMH becomes universal in the older ranging from 11-21% in adults aged around 64 to 94% at age 82.^{32,35,46–50} Furthermore, white matter lesions are associated with many geriatric disabilities which include cognitive disorders, gait and mood disturbances, and urinary problems.⁵¹

Risk factors for WMH are a similar set to those for vascular diseases;⁵² hypertension plays a crucial role in WMH.^{53,54} Large population-based studies have shown that advancing age, female gender, higher blood pressure or hypertension, African American ethnicity, smoking, alcohol use, lower education attainment, low income, silent infarct, diabetes, and lower forced expiratory volume in 1 second (FEV1) are associated with WMH as a marker of CSVD.⁵⁵ Hypertension is considered as the most important modifiable risk factor for CSVD and, particularly, WMH and lesions.^{48,49,56–63} Some of those studies have also shown that

duration of hypertension is associated with WMH.^{60,61} In this perspective, the pharmacological treatment standard of CSVD has been designed to manage blood pressure.⁶⁴

Despite that the pathophysiology of WMH is little known and that various pathologies can lead to WMH,^{33,65} it is likely to have ischemic or degenerative origin that damages the small vessels of the brain, leading to chronic cerebral hypoperfusion and myelin rarefaction.⁶⁶ Large-scale brain imaging studies have shown that WMH is associated with incident stroke, vascular cognitive impairment, dementia, Alzheimer's disease or death.^{42,67} Specifically, the association of cognitive decline with WMH has been reported for more than 20 years.^{31,32,68} A meta-analysis of 23 cross-sectional and 14 longitudinal studies reported that presence of WMH was associated with concurrent cognitive deficits in every test domain, and the effect was constant over time.⁶⁹ Studies in older subjects without neurological diseases reported mixed results on individual domains, but WMH is considered to be particularly correlated with decline in information-processing speed and executive function.⁷⁰

Physiological characteristics of cerebral small vessels

As mentioned earlier, blood flow to the brain is crucial. A few minutes interruption of blood flow can cause severe brain damage and even the slightest insufficient blood flow can lead to chronic brain injury.⁷¹ The brain exhibits unique functional properties to ensure this crucial blood flow.⁷² First, since the brain has little means to store energy, the brain works to regulate vessels to sustain an adequate supply of blood in cerebral blood flow (CBF) under a variety of conditions. CBF is primarily determined by mechanisms including perfusion pressure, autoregulatory mechanisms, and vascular reactivity to the partial pressure of arterial CO₂.⁷³ Second, the brain has an auto-regulation mechanism to maintain the constant transmural pressure against variations in arterial pressure during daily activity. Myogenic reactivity is known to play a major role in auto-regulation, and there are other candidate mechanisms to be studied.^{73,74} Third, the brain has a unique anatomical structure called the blood-brain-barrier (BBB), which is an important mechanism to regulate the entry and exit of blood composites including ions, nutrients, macro molecules and energy metabolites.⁷⁵ Specifically, BBB protects the brain from physiological fluctuation in plasma composites, and from harmful agents circulating in the blood supply.⁷⁶ In essence, the homeostasis in the brain is precisely controlled by an elaborate system of unique physiological structures and mechanisms.

Physiological studies found that the unique feature of the cerebral vasculature, BBB does not function independently, but as a module called the neurovascular unit (NVU),⁷⁵ which enables tight regulation of blood flow to the brain. The NVU consists of endothelial cells and smooth muscle cells, and are surrounded by a basement membrane, the perivascular

(Virchow-Robin) space, pia mater, and astrocyte endfeet.⁷⁷ Pial arteries running on the brain surface in the subarachnoid space branch out into the smaller, penetrating arteries. Likewise, penetrating arteries further branch out into parenchymal arteriole and further into capillaries. As shown in **Figure 2**,⁷⁷ pial arteries and penetrating arteries are covered by smooth muscle cells and are separated from brain tissues by parenchymal basement membrane. Branching out into parenchymal arterioles, both parenchymal basement membrane and pia matter disappears; further branching out into the capillaries, eventually smooth muscle cells disappear. Without smooth muscle cells and pia mater, capillaries gain pericytes between endothelial cells and astrocytes; thus, neurons and astrocytes are in direct contact with them.

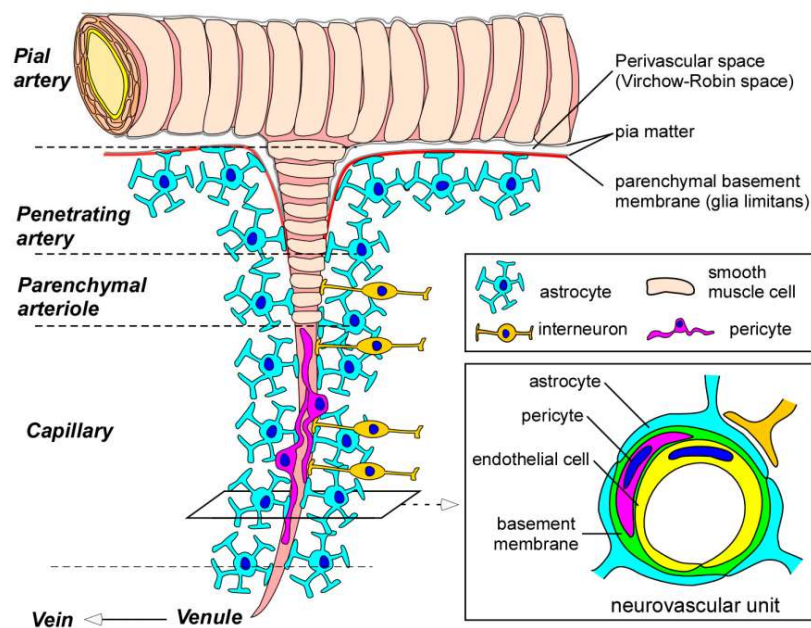


Figure 2 Blood-brain Barrier (BBB) and the neurovascular unit (NVU) Adapted from Yu Yamazaki et al. 2017

Mechanisms underlying CSVD: Current understanding

Animal models are useful to simulate ischemic changes in the brain; however, due to the lack of pathological understanding the optimal animal models in CSVD or WMH have not yet been developed. Hypertension, as an important risk factor in CSVD, has been thoroughly studied in rodents and other model animals. Studies confirmed deleterious effects of hypertension on the cerebral vascular structure and functions; those changes are thought to be associated with cerebral blood flow (CBF) reduction, impaired vasodilation and vasodilator reserve, and shifts in the autoregulatory curve.^{78,79} The studies suggested several underlying mechanisms, and the effect of angiotensin II mediated by oxidant-dependent and immune-dependent mechanisms also seems to play a crucial role in the pathology.^{78,79} Angiotensin II mechanism is often a therapeutic target in hypertension and other cardiovascular disease treatments. Although animal models showed the effect of hypertension on vasculature, the causal relationship with brain lesions is lacking.⁷² Also, a common animal model known as “sporadic small vessel disease model” typically induces malignant severe hypertension which is beyond the range of autoregulation;^{72,80} thus, the sporadic model cannot show the specific influence of modest blood pressure variability. A recent animal study for dementia implemented a device called an ameroid constrictor that precisely replicates chronic cerebral hypoperfusion by gradually narrowing cerebral vessels.⁸¹ A study with a novel animal model using this device showed the role of reduced CBF in white matter damage.⁸²

Modern concepts in CSVD etiopathology originate in the post-mortem work of C. Miller Fisher during 1955-1973, which took clinicopathological-vascular examinations in 20

patients with lacunes, lacunar infarcts, and perforating arterioles.^{83–87} Lacunar ischemic stroke is a type of stroke attributable to a recent small infarct in the white matter, basal ganglia, pons, or brainstem. Lacunar stroke patients provided a clue about the CSVD etiology, as a clinically overt stage of CSVD. Historically, it was believed that symptomatic lacunar infarcts are due to atherosclerosis or embolism in larger arterioles and the silent ones are due to lipohyalinosis or fibrinoid necrosis in the smaller arterioles.⁸⁷

More recently, three possible etiologies for lacunar ischemic stroke are suggested:^{26,88} atheroma, embolism, and intrinsic small vessel disease. Atherosclerosis in the middle cerebral artery (MCA), a type of parent artery, has been suspected as a common cause of lacunar stroke in Fisher's studies, but stenosis does not seem to address lacunar strokes. In the Warfarin Aspirin Symptomatic Intracranial Disease (WASID) trial in subjects with stroke history and current stenosis, only 11% (38/347) of randomized subjects had lacunar stroke as an index.⁸⁹ Also, intracranial stenosis is rare in the Caucasian population at less than a 1 to 2 % prevalence, but still a quarter of ischemic stroke is lacunar in the population.⁹⁰ To address this discrepancy it was hypothesized that there are two distinct subtypes of lacunar stroke: those due to proximal perforating arteriolar atheroma and the others due to lipohyalinosis/arteriolosclerosis.⁹¹ However, it is unclear how to differentiate them reliably.

The next possible mechanism is embolism, but it also does not appear to be a common cause of CSVD. Although embolism can cause an acute lacunar infarct, cohort studies and meta-analyses showed that only 10-15% of lacunar infarcts and few WMH are explained by embolism.^{92–94} Also, a systematic review of experimental studies that induced stroke in animals concluded that emboli mostly caused cortical, but not small subcortical, lesions.⁹⁵

One of the studies reported that only less than 6% of emboli injected into the carotid arteries entered the perforating arterioles.⁹⁶ Thus, lacunar stroke is less likely to be associated with overt emboli compared to non-lacunar stroke,^{92,93} and, thusly, the stenosis is unlikely to cause the infarct⁹³ or WMH⁹⁴.

In his post-mortem studies Fisher claimed blood pressure is the cause of pathologies in CSVD *i.e.* arteriolosclerosis, lipohyalinosis or fibrinoid necrosis; however, the relationship between blood pressure and CSVD or WMH is not clearly understood yet. In a study of 70 brain autopsies, many CSVD patients did not have classic vascular risk factors including hypertension.⁹⁷ Blood pressure is still considered as an important risk factor for WMH,^{57,60,61,63,98} but the relationship between blood pressure and WMH is complex. It is still unclear whether systolic⁶³ or diastolic^{62,99} BP is important in WMH; it is not consistent if BP does predict WMH progression⁵⁷ or not¹⁰⁰.

The emerging hypothesis in CSVD pathology is that the loss of normal endothelial function in the blood-brain barrier (BBB) causes pathological changes in the brain. BBB consists of endothelial cells tightly anchored to each other. The failure of the cerebral arteriolar and capillary endothelium could explain the arteriolar wall infiltration and thickening via perivascular tissue changes (*e.g.* damaged arteriolar smooth muscle and fibrin deposition known as lipohyalinosis and fibrinoid necrosis).¹⁰¹ These changes were referred to as “microvascular pathology” in Fisher’s original paper.¹⁰² Cumulative microvascular pathology in perivascular tissue can lead to rarefaction and demyelination, which are pathologically observed in WMH carriers.²⁹ More severe consequences of endothelium failure include vessel lumen dilation and narrowing and stiffened vessels with loss of normal

autoregulation; as scientists long suspected for CSVD etiology.^{103,104} As the mechanism involving endothelium dysfunction has been hypothesized relatively recently, our understanding of the role of cerebrovascular endothelium in CSVD pathology is lacking.¹⁰⁵

Previous Genetic Studies of CSVD and WMH

Even though a myriad of studies put intensive efforts to decipher molecular mechanisms underlying CSVD, the precise mechanism is not fully understood. Poor understanding of biological mechanisms behind the disease continues to impede the development of effective and precise treatments for CSVD. Studying microvessels in the brain has been a struggle due to the limited accessibility to them, the difficulties of in-vivo-visualization, and the lack of adequate animal models.¹⁰⁶ However, genetic studies have been expanding our understanding of CSVD mechanisms; monogenic forms of CSVD were discovered and results from large-scale studies are promising to decode the genetic basis of sporadic CSVDs. From the current understanding it is hypothesized that CSVDs are likely to result from various pathogeneses such as atheroma (decreased blood flow), emboli (blockage-causing piece of material inside a blood vessel), inflammation with BBB disruption, cell degeneration, and myelin breakdown. Genetic studies are designed both to investigate whether or not WMH is developed by these mechanisms and to identify genetic markers for prevention or intervention purposes.

Animal model

Transgenic mouse models have been numerously employed to study mechanisms of known causal genes in monogenic CSVDs such as CADASIL. Vascular protein deposits (granular osmiophilic material) containing the *Notch3* extracellular domain have been identified in transgenic mouse models expressing a pathogenic mutant,^{107–111} but not in knockout mice.^{112–114} Through these animal studies *NOTCH3* is now strongly claimed to have a gain-of-toxic-function rather than a loss-of-function mechanism.^{114,115} Proteomic

profiles of *HTRA1* gene knockout mouse was shown to overlap with the protein profiles of CADASIL patients.¹¹⁴ Unlike monogenic CSVD, animal studies in common and complex form of CSVD are lacking. Since large-scale epidemiologic studies have recently identified genetic variants associated with common and complex form of CSVDs presented as WMH, brain infarcts, and other clinical manifestations; it is necessary to confirm the mechanistic function of those genes in transgenic animal studies. A recent study using CRISPR/Cas9 models examined the role of a SNP (rs9349379) located in *PHACTR1* gene on five vascular diseases including hypertension.¹¹⁶ This suggests a potential utility of the modern gene-editing technology with WMH and CSVD causal variants.

Compared to non-cases a family history of stroke is more likely to be found in stroke patients, and a positive family history of stroke is strongly associated with CSVDs [odds ratio of 1.93 to 2.76].^{117,118} Specifically for WMH, twin studies showed that correlations of WMH volume in pairs are greater in monozygotic twins than dizygotic twins, and the heritability estimate of 71% adjusted for age and head size was very high.¹¹⁹ High heritability in WMH volume was also estimated in family-based studies ranging from 52 to 78%.^{120–122} The Framingham Heart Study suggested that the heritability estimate may depend on gender and age.¹²⁰ However, heritability estimates based on common genetic variants, known as narrow-sense heritability, is generally lower; the estimate in small vessel-related stroke and deep intracerebral hemorrhage ranged from 15 to 30%,^{123,124} and was only 6.4 % for WMH specifically.¹²⁵ This discrepancy between the estimates can be explained by other genomic variants or mechanisms not included in GWASs (*i.e.* rare variants, DNA methylation, gene expression, gene-gene interaction, and gene-environment interaction).

Genetic studies have identified nine genes for mendelian forms of CSVD including *NOTCH3*, *COL4A1*, *COL4A2*, *TREX1*, *HTRA1*, *FOXC1*, *PITX2*, *CECRI*, and *CTSA*.¹²⁶ *NOTCH3*, known to be associated with CADASIL,^{127–130} encodes the *NOTCH3* transmembrane receptor involved in arterial differentiation and vascular smooth muscle cell remodeling.¹³¹ *HTRA1* (high temperature requirement protein A1) is involved in TGF- β signaling, playing an important role in vessel development and maintenance.¹³² Also, *HTRA1* is associated with cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL).^{25,133} Studies also identified that some genes cause specific CSVD in a monogenic manner; basement membrane abnormality related to *COL4A1* and *COL4A2*, which are associated with COL4-related angiopathies,^{134,135} changes in endothelial and pericyte proliferation as well as impairment of BBB integrity were found to be associated with *FOXC1/PITX2* in animal models,^{136–138} and *CECRI* is associated with *ADA2* deficiency leading to early-onset strokes and inflammations.^{139,140} Additionally, *TREX1* is associated with retinal vasculopathy with cerebral leukodystrophy and systemic manifestations (RVCL-S) and the gene plays a role in basement membrane defects in capillaries.^{141,142} Recently, investigators found a novel association of *CTSA* with a subtype named cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *CTSA* encodes cathepsin A which is involved in the inactivation of neuropeptides and regulation of a lysosomal pathway of protein degradation and pathological change leads to fibrous thickening of the small vessels.^{143,144} Mutations in *NOTCH3*, *COL4A1* and *COL4A2* are thought to explain a large portion of familial CSVDs, and penetrance of these genes are high so that individuals with even a single mutation are at almost 100% risk of the disease.¹²⁶ Even though monogenic

CSVDs are considered genetically heterogeneous and rare, the shared clinical and radiological features between monogenic and sporadic CSVD implies that the genetic underpinning of monogenic CSVDs provides important information to understand the genetics of the common and prevalent forms of CSVD.¹⁰⁶

Knowing the presence of genetic components in WMH, scientific efforts to map the trait-associated locus were made with linkage studies. Linkage studies test a genetic locus transmitted with the disease in larger regions (tens of centiMorgans). The co-transmission of two close-enough genetic loci is called “linkage disequilibrium (LD)” and the degree of LD is assessed with a logarithm of the odds (LOD) score, the odds of co-transmission compared to expected independent inheritance. Linkage studies of WMH identified the locus at 4 cM on chromosome 4 in healthy elderly white subjects.^{145,146} Also suggestive LOD scores have been reported at 95cM on chromosome 17 (LOD=1.78)¹⁴⁵, at 118 cM on chromosome 11 (LOD=2.21), and at 13 cM on chromosome 21 (LOD=1.75) in white subjects, and at 36 cM on chromosome 22 (LOD=2.02) as well as at 58 cM on chromosome 21 (LOD=1.99) in black subjects.¹⁴⁷ Additionally, bivariate linkage analysis suggested that 1q24, 1q42, 10q22-q26 and 15q26 loci are in LD both in WMH volumes and BP measurements, indicating pleiotropic effects in the regions.¹⁴⁸

Moreover, candidate gene studies investigated dozens of biologically plausible genes in the association with WMH. A recent systematic review described 46 candidate gene studies having investigated the association between WMH and 19 genes which are mostly involved in lipid metabolism, control of vascular tone, or blood pressure regulation.¹⁴⁹ The study also performed a meta-analysis for four genes (*APOE*, *ACE*, *MTHFR*, and *AGT*), and none of

them were significantly associated with the WMH grade.¹⁴⁹ However, the *AGT* gene has been repeatedly implicated in WMH linkage studies and the most studied SNP in the gene,^{147,148} M235T has been reported in a genome-wide association study¹⁵⁰ suggesting that M235T is likely to be in LD with functional causal variants for WMH.¹⁵¹

The advance of genotyping and computational technologies enabled finer mapping of causal loci associated with WMH. In addition to the linkage findings, genome-wide association studies (GWAS) explored which single nucleotide polymorphism (SNP) is associated with WMH (**Table 1**). The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium has published GWAS associations of WMH in disease-free populations. These studies excluded subjects with stroke history due to WMH being prevalent in stroke patients and thus discounting the cases of WMH formed as a stroke comorbidity. The first large-scale association study by CHARGE identified SNPs in six genes (*TRIM65*, *TRIM47*, *WBP2*, *MRPL38*, *FBF1*, and *ACOX1*) at 17q25;¹⁵⁰ a follow-up study with a larger number of subjects in a multiethnic population additionally mapped *SH3PXD2A* at 10q24, *HAAO* at 2p21, *PMF1-BGLAP* at 1q22, and *EFEMP1* at 2p16.¹⁵² Under the hypothesis of shared genetics among stroke and stroke-free populations, a meta-analysis of the latter CHARGE GWAS and a GWAS in 3,670 ischemic stroke patients was performed for the previously reported 15 SNPs.¹⁵³ This candidate-gene meta-analysis identified four significant loci, *NBEAL1* at 2q33, *EVL* at 14q32, *CIQL1* at 17q21, and *COL4A2* at 13q34.¹⁵³ The most recent and largest meta-analysis of GWASs, including 2,797 stroke subjects and 8,429 stroke-free individuals from UK Biobank, identified a novel locus in *PLEKHG1* at 6q25 and confirmed two previously reported loci *i.e.* 17q25 and 2p16.¹²⁵ The

novel gene *PLEKHG1* was also suggestively significant in a GWAS of stroke-free population (P value= 7.9×10^{-7}).¹²⁵ There was also an effort to identify rare variants associated with WMH. Jian et al. performed an exome-wide association analysis including up to 250,000 exonic variants in 20,719 individuals.¹⁵⁴ In this analysis the known loci 17q25 and 2q33 were confirmed and a novel gene in the locus (*MRPL38*) was identified at the genome-wide significance level. Most recently, a meta-analysis of two of the exclusively largest GWASs from CHARGE and UK Biobank confirmed previous findings and additionally identified ten novel loci for WMH. This study is currently under review for publication, and the identified genome-wide significant loci were informed during personal communications with the authors.

Functions of some genome-wide significant SNPs can be plausibly explained in relation to the CSVD pathology. *HAAO* at 2p21 encodes 3-hydroxyanthranilate 3,4-dioxygenase, which catalyzes the synthesis of quinolinic acid (*QUIN*). *QUIN* has been identified in studies of Alzheimer's disease and Huntington's disease. *PMF1-BGLAP* at 1q22 encodes polyamine-modulated factor 1, a nuclear protein regulated by polyamines. This protein is required in chromosome alignment and segregation during mitosis. Genes including *NEURL* (neuralized E3 ubiquitin protein ligase 1), *TRIM47* (tripartite motif containing 47), *SH3PXD2A* (tyrosine kinase substrate with five SH3 domains), *EEMPI* (endothelial cell membrane glycoprotein fibulin-3) and *NBEAL1* (neurobeachin like 1) were implicated in white matter-related malignant brain tumors involving glial cells.¹⁵² *COL4A2*, aforementioned in relation to the monogenic form of CSVD, encodes one of the six subunits of type IV collagen, the major structural components of the basement membrane. Variants in

this gene were reported to be associated with intracerebral hemorrhage and ischemic small-vessel stroke.^{155,156} Lastly, *PLEKHG1* encodes pleckstrin homology and RhoGEF domain containing G1 which is one of the Rho guanine nucleotide exchange factors (Rho-GEFs). Rho-GEFs are involved in cyclic stretch-induced cell reorientation in the vascular endothelium.¹⁵⁷

Table 1 Genome-wide significant loci associated with WMH

Locus	Gene	Population	Analysis Method	Ref
1q22	<i>PMF1, BGLAP</i>	Pooled	GWAS	152
1q24	200cM	Hispanic	Bivariate Linkage ^a	148
1q31.3	<i>DENND1B</i>	EA	GWAS	PC
1q42 ^s	272cM	Hispanic	Bivariate Linkage ^a	148
2p16	<i>EFEMP1</i>	Pooled, EA	GWAS	125,152
2p21	<i>HAAO</i>	EA	GWAS	152
2q32.1	<i>AC007319.1</i>	EA	GWAS	PC
2q33	<i>NBEAL1</i>	Pooled, EA	GWAS	152–154
	<i>WDR12</i>	Pooled, EA	GWAS	154
	<i>CARF</i>	Pooled, EA	GWAS	154
4p	4cM	EA	Linkage	145,146
5q23.2	CTC-441N14.4	EA	GWAS	PC
6q25.1	<i>PLEKHG1</i>	EA	GWAS	125
7q22.1	<i>ZAN</i>	EA	GWAS	PC
10q24-26 ^s	15cM	Hispanic	Bivariate Linkage ^a	148
10q24	<i>PDCD11</i>	EA	GWAS	152
	<i>NEURL</i>	Pooled, EA	GWAS	152
	<i>SH3PXD2A</i>	Pooled, EA	GWAS	152
	<i>PDCD11</i>	Pooled	GWAS	152
				147
11 ^s	118cM	EA	Linkage	147
13q34	<i>COL4A2</i>	EA	GWAS	PC
14q32.2	<i>EVL, DEGS2</i>	EA	GWAS	PC
15q22.31	<i>MTFMT, SLC51B</i>	EA	GWAS	PC
15q26 ^s	94cM	Hispanic	Bivariate Linkage ^a	148
16q12.1	<i>RP11-437L7.1</i>	EA	GWAS	PC
16q24.2	<i>C16orf95</i>	EA	GWAS	PC
17 ^s	95CM	EA	Linkage	145
17p11.2	<i>EPN2</i>	EA	GWAS	PC
17q21.31	<i>DKAKD, NMT1</i>	EA	GWAS	PC

17q25	<i>UNC13D</i>	Pooled, EA	GWAS	152
	<i>TRIM65</i>	Pooled, EA	GWAS	125,150,152,154
	<i>TRIM47</i>	EA	GWAS	150,152
	<i>WBP2</i>	EA	GWAS	150
	<i>FBF1</i>	Pooled, EA	GWAS	152,154
	<i>ACOX1</i>	Pooled	GWAS	152,154
	<i>MRPL38</i>	EA	GWAS	154
19q13.32	<i>APOE</i>	EA	GWAS	PC
21^s	13cM, 58cM	AA, EA	Linkage	147
22^s	36cM	AA	Linkage	147

^a Bivariate analysis of WMH and blood pressure measures (pulse pressure and arterial blood pressure). ^s Linkage result with suggestive LOD score. EA: Subjects of European ancestry, AA: Subjects of African ancestry. PC: based on personal communication

Epigenetic studies and integrative analysis

In genetic epidemiology “missing heritability” has been a mystery to unravel. Although GWASs found a myriad of common variants associated with complex diseases, a large portion of heritability is not completely accounted for yet. This unexplained portion of heritability (“missing heritability”) can be due to overestimated heritability or underestimated SNP effects in GWASs. In the post-genome era, several additional explanations are suggested to fill this gap. Rare alleles with bigger effects are being studied actively as the first alternative explanation. Tagger SNPs in GWASs do not cover rare alleles enough plus the ability to impute those is limited;¹⁵⁸ exome-wide association studies covering rare alleles (*i.e.* minor allele frequency <0.01 or 0.05) may find extra genetic variants. Some rare variants may have greater effects than common variants which may indicate a single-gene disorder.¹⁵⁹ Secondly, it is shown that genetic interactions or epistasis account for a significant portion of missing heritability. Estimation of heritability requires an assumption that there is no genetic interactions or epistasis in the heritability estimation; thus, existence of epistasis can lead to over-estimation. This overestimated gap is referred as “phantom heritability”. Thirdly, epigenetics is another explanatory factor for the missing heritability. The genome is wrapped around units of histones, which are collectively called chromatin. Chromatin can be modified through DNA methylation, modifications of histone proteins, or post-transcriptional silencing mediated by micro-RNAs (miRNAs).^{160,161} Epigenetic changes are mostly transient, but some epigenetic changes can be inherited in the short term.¹⁶² Epigenetics is dynamically influenced by the environmental factors such as smoking, stress, and nutrients^{161,163} and

regulates gene expression; thus being likely to contribute to phenotypic variance and explain some extent of missing heritability.

Epigenetics receives considerable interest among many aspects of genomics as it involves the regulatory process of translating encoded DNA into proteins. Since epigenetic phenomenon is reversible, it provides information on the complex dynamics between genes and environment. DNA methylation has been most studied of the several epigenetic changes since it is more feasible to measure in an epidemiological setting.¹⁶⁴ DNA methylation is an epigenetic phenomenon that cytosine in the genome is attached with a methyl group onto its C5 position by DNA methyltransferases (DNMTs). It is well shown in cancer cells that changes including chromosomal instability, activation of oncogenes, silencing of tumor suppressor genes and inactivation of DNA repair systems are not caused only by genetic but also by epigenetic abnormalities such as genome-wide hypomethylation and regional hypermethylation of CpG islands.^{165–169} The role of epigenetics has also been studied in the epidemiologic settings and studies have identified DNA methylation sites associated with complex diseases such as cardiovascular diseases and stroke.^{170–173} In the search for DNA methylation sites involved in disease etiology, the analytic strategy in GWAS is used in epigenetic studies. A genome-wide investigation of associations between DNA methylation levels and a complex trait is now coined as an epigenome-wide association study (EWAS).^{174–176}

As our understanding broadens with data at different molecular levels (*i.e.* genome, epigenome, and transcriptome), an integrative analysis of the multi-omics data became important. Although GWAS approaches identified genetic loci related to WMH burden, the

precise mechanistic pathway involving molecular events (*e.g.* DNA methylation, regulation of gene transcription, translation, post-translation modification, etc.) is not fully addressed. Integrative genomics has been introduced to combine genomics and emerging functional omics (*i.e.* epigenomics, transcriptomics, proteomics, and metabolomics) and derive a meaningful causal etiological pathway utilizing multi-omics data. An integrative approach has been successful in many complex traits including immunology,¹⁷⁷ oncology,^{178,179} cardiovascular diseases,¹⁸⁰ and some neurological diseases such as narcolepsy,¹⁸¹ Alzheimer's disease,¹⁸² schizophrenia,¹⁸³ and autism¹⁸⁴.

Public Health Significance

The use of WMH burden as an endpoint may improve CSVD prevention, thusly facilitating general brain health. Since brain imaging is non-invasive and became accessible, prevention strategies based on WMH burden is widely applicable to the public. Large-scale brain imaging studies have shown that WMH is associated with public health burdens of extreme significance such as incident stroke, vascular cognitive impairment, dementia, Alzheimer's disease, and death.^{42,67} Early detection of relevant brain diseases is more feasible with utilizing WMH burden as the accessible endpoint rather than diagnosing based on symptoms. For instance, using MRI-defined infarcts, silent ischemic stroke was found to be significantly more prevalent than symptomatic events.¹⁸⁵ Furthermore, genetic epidemiology may enable us to earlier identify individuals at higher risk of WMH burden^{125,150,152} and prevent brain diseases before the damage accumulates and the changes become irreversible.

Understanding WMH etiology can help to establish an effective and safe CSVD treatment strategy. Precision medicine, the medical care designed to optimize efficiency or therapeutic benefit, can be achieved first by subtyping CSVD.¹⁸⁶ Although the pathophysiology is not fully addressed (and various pathologies can lead to WMH),^{33,65} WMH is likely to have an ischemic or degenerative origin that damages the small vessels of the brain, leading to chronic cerebral hypoperfusion and myelin rarefaction.⁶⁶ Due to the limited understanding of the etiology, current practice of ischemic vascular disease prevention is barely limited to management of modifiable risk factors such as high blood pressure and compositions of lipids;¹⁸⁷ moreover, these treatments have not shown agreed upon and consistent effectiveness for CSVD. The current treatments of CSVDs are listed as

thrombolysis, antihypertensive, statins and antiplatelets. Observational studies - the Rotterdam Scan Study and the Epidemiology of Vascular Ageing study - showed that effective hypertension treatment was associated with lower risk of CSVD manifested as WMH,^{60,61} while it has shown inconsistent effects in randomized controlled trials.^{188–190} Likewise, lipid-lowering treatments also did not show optimal effectiveness in preventing WMH,^{191,192} even worse, safety issues were elevated with these medications. Excessive blood pressure decrease may even induce cognitive decline in older patients with extensive WMH as shown in the Secondary Prevention of Small Subcortical Strokes trial.¹⁹³ A recent large statin study in a stroke-free population showed that statin use can even increase the risk of ischemic stroke.¹⁹⁴ Also, antiplatelet therapy should be avoided due to the excessive risk of intracerebral hemorrhage in CSVDs.¹⁹⁵ Furthermore, treatments for strokes are not effective in CSVDs. The gold-standard treatment of acute ischemic stroke, tissue plasminogen activator (t-PA) showed an increased risk of hemorrhage in patients with WMH.^{196–199} Failure of traditional treatment options leads us to the conclusion that CSVD is etiologically dynamic and heterogeneous; therefore, the treatment strategy needs to be designed based on its etiology and progression. Subtyping WMH etiology and developing a precision medicine strategy based on epigenetics can be the new way to achieve the second prevention of CSVD.

The main goal of this dissertation is to identify the role of DNA methylation in WMH burden in the reasonable hope that it may provide a key to understanding the CSVD etiology in relation to aging. As life expectancy increases, disabilities in the older population are even more subject to public health interest. One of the important risk factors in many neurological

diseases is advancing age, since many neurological changes develop as an individual ages.²⁰⁰ Senescence, the condition or process of deterioration with age, is composed of many neurological changes⁵³ - including cerebral white matter development.²⁰¹ WMH is very prevalent in the aging population; 95% of older adults aged 60-90 have these signs on brain MRI. Age-related changes in WM are also associated with increased risk of stroke and stroke mortality.²⁰² DNA methylation is known to be affected by aging,^{201,203} and it is suggested as a biomarker for cognitive dysfunction and brain aging.²⁰⁴ Epigenetics is a composite factor of genetic composition nurtured by environmental exposures such as aging, making it a useful tool to study age-related genetics in CSVD.

To summarize, results of this dissertation might bring a few positive impacts to public health. First, the use of WMH burden as the endpoint would facilitate CSVD prevention. Since WMH burden can be measured using MRI techniques, it is accessible, non-invasive, quantitative, and a representative marker of CSVD. Therefore, a prevention strategy based on WMH burden is more widely applicable to the public. Second, the discovery of epigenetic risk factors in addition to genetic risk factors may improve personalized medicine for CSVD. CSVD is a complex disease leading to heterogeneous medical conditions; understanding the genetic architecture behind WMH burden will help us classify individuals by their potential risk genotypes. Furthermore, we can develop a more effective treatment strategy that is customized for distinct etiological pathways associated with each genotype. Thirdly, because epigenetic factors are influenced by both genetic and environmental exposures, the identification of CSVD-associated epigenetic changes may provide an avenue to intervene on “modifiable” environmental risk factors to prevent CSVD.

Specific Aims and Hypotheses

Emerging evidence suggests that a complex disorder can be understood better with epigenetics. WMH on magnetic resonance imaging (MRI) are a subclinical marker of vascular brain injury, therefore a clinical biomarker of CSVD.⁴³ Twin and family studies have shown that WMH is highly heritable with an estimated heritability ranging from 52 to 78%^{120–122}. However, the underlying genetic architecture is still unresolved in that all identified risk loci explain only a small portion of the heritability (6.4%).^{125,150,152–154} To fill this gap, research has drawn attention to new aspects of genomic variation. In this dissertation, we aim to investigate one aspect called “epigenetics” in association with WMH burden. Epigenetics receives increasing attention since it is involved in the regulatory process of gene expression. Also, epigenetics reflects the complex interactions of gene and environment, and thus, it is expected to provide novel findings and explain an additional layer of genetic architecture in WMH, which is previously unrealized.

The objective of this dissertation is to identify novel genes associated with WMH burden, implicate through variation in DNA methylation - the most commonly studied epigenetic phenomenon. First, we will complete an epigenome-wide association study (EWAS) of WMH, where DNA methylation sites associated with WMH will be identified. Second, we will examine the causal direction between the identified DNA methylation and WMH using genetic instruments in order both to confirm if the epigenetic association is under genetic control and to establish directionality of the EWAS associations. Lastly, we will incorporate multi-omics information (*i.e.* genetic, epigenetic and transcriptional

information) to aid in interpretation, and characterize functional relationships that will help infer etiology of WMH.

Below are the study questions in more scientific forms.

Aim 1: To identify DNA methylation (DNAm) loci associated with WMH burden

(Epigenome-wide association study)

Q1. Are there DNAm loci associated with WMH burden?

Q2. Are the associated DNAm loci implicated in biological pathways or etiology of relevant diseases or traits?

Aim 2: To identify the causal pathway in DNAm-WMH associations using genetic instruments

Q1. Are there single nucleotide polymorphisms (SNPs) that influence DNA methylation levels at loci associated with WMH?

Q2. How is the causal direction between DNAm and WMH burden?

Q3. Is gene expression associated with the WMH-associated DNAm?

Aim 3: To investigate the role of DNAm in WMH etiology using multi-dimensional (genetic, epigenetic and transcriptional) data

Q1. Are there etiological pathways identifiable by the integrated multi-dimensional WMH burden associations?

Q2. Does any gene play a mediator role in the association between DNA methylation and WMH burden?

Q3. Do the genetic associations of multi-dimensions (*i.e.* DNAm, gene expression and WMH burden) co-localize at the known WMH loci?

METHODS

Aim 1: Meta-analysis of epigenome-wide association studies (EWASs) of WMH burden

The overall study design is shown in **Figure 3**. In this aim (Aim 1), we performed a meta-analysis of EWASs in WMH burden. Accounting for spatial correlation among DNAm, P values were collated to identify a group of DNAm associated with WMH burden. Results were interpreted in the context of biological functions using two strategies based on: 1) the genes mapped to the identified DNAm and 2) the shared epigenetics.

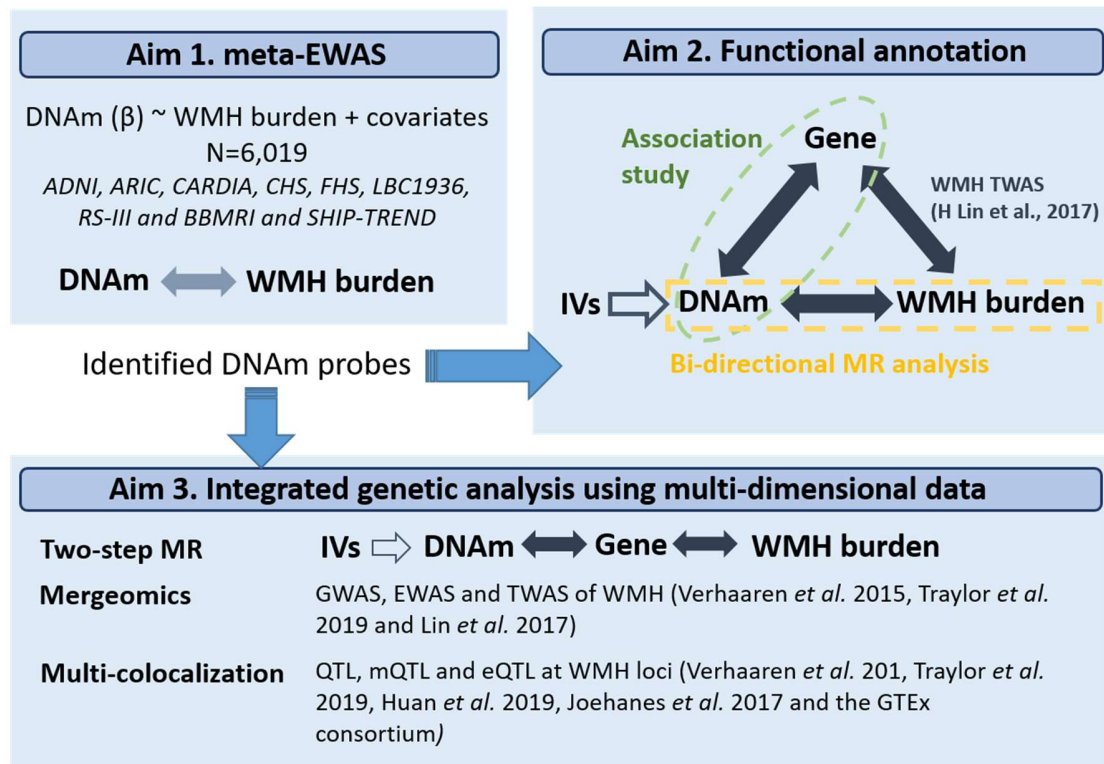


Figure 3 The schema of overall study designs

Study subjects

We included ten population- or family-based cohort studies - Alzheimer's Disease Neuroimaging Initiative (ADNI), Atherosclerosis Risk in Communities (ARIC) study, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI), Coronary Artery Risk Development in Young Adults (CARDIA), Cardiovascular Health Study (CHS), Framingham Heart Study (FHS) Offspring Cohort, Genetic Epidemiology Network of Arteriopathy (GENOA), Lothian Birth Cohort 1936 (LBC1936), Rotterdam Study III (RS3) and The Study of Health in Pomerania TREND (SHIP-TREND). Eligible subjects were middle-age to elderly, free of stroke or dementia, and of African or European ancestry (AA or EA). Description of participating cohorts and study-specific methods are described in the ***Supplemental Material***. Descriptive statistics in participating cohorts are presented in the ***Table S1***.

Protection of human subjects

All participating studies had received institutional review board approval for its informed consent process, clinical examinations (brain imaging, anthropometric measures, blood pressure measures, etc.), methylation sample collection and its use, and data access and distribution. Approvals were granted to individual studies and the study-specific ethic statements are presented in ***Appendix A. Supplemental Material***. Subjects who provided written informed consent for study participation, MRI scanning, and use of methylation data were entered in this study. Study protocol for this dissertation was also reviewed by the University of Texas Health Science Center Committee for the Protection of Human Subjects, the approval letter is attached in ***Appendix A. Supplemental Material***.

White matter hyperintensity (WMH) measurements

WMH on brain MRI in this study indicates the high-intensity areas within cerebral white matter on T2-weighted images. MRI was taken in the same or the closest subsequent visit to the visit where DNA methylation sample was taken. Diagnosis of clinical dementia or stroke for eligibility were made at MRI scans. MRI was scanned and interpreted according to the standardized method in each study, with demographic or clinical information of subjects blinded. Participating studies obtained T2-weighted scans using either spin-echo pulse sequences or fluid-attenuated inversion recovery (FLAIR) sequences from all subjects. MRI scan of WMH is measured either on a quantitative volume (mL) or qualitative scale (grade). Since WMH distribution is typically right-skewed, WMH burden is quantified as the log-transformation of a validated WMH value +1. Addition of 1 is to make the WMH burden be always a positive value for convenience in later analyses. MRI procedures and WMH quantification in each study are detailed in *Appendix A. Supplemental Material*.

DNA methylation profiling

DNA methylation levels have been measured in whole blood samples with the Illumina Infinium Human450 Methylation BeadChip in most participating cohorts. The GENOA study measured methylation levels at 26,449 probes with the Illumina Infinium HumanMethylation 27 BeadChip, which are covered by the Illumina Human450 Methylation BeadChip. ADNI, CARDIA and SHIP-TREND measured methylation with the Illumina MethylationEPIC BeadChip with a denser coverage of probes (~850K). Each study independently performed the quality control (QC) for DNA methylation data, complying the agreed minimum QC guideline; methylation probes with detection P value < 0.01 were

selected, and then probes having more than 5% of samples with a detection P value of >0.01 were excluded; a sample was removed if the total proportion of probes with a detection P value <0.01 is less than 95%. DNA methylation values were then standardized using an intra-array normalization method. Beta values from the output were used as a dependent variable in an individual EWAS. The details of DNA methylation measurements are described in *Supplemental Material*.

Cohort-level epigenome-wide association analyses

Individual CpG probes were tested for the association with the WMH burden in two linear mixed effect regression models *i.e.* a reduced- and a full-adjustment model according to the following formulae. In two models, the untransformed methylation beta value for CpG sites is the dependent variable and WMH burden is the independent variable of interest. In a reduced-adjustment model, age, sex, study site if applicable, total cranial volume (cm^3), white blood cell proportion (%), technical covariates (*i.e.* plate, chip-position, row, and column) and race-specific genetic principal components of ancestry (PCs) were included as covariates. In a full-adjustment model, smoking status (pack-year), body mass index (BMI, kg/m^2), systolic blood pressure and diastolic blood pressure (SBP and DBP, mmHg) were added as covariates. White blood cell proportion was estimated with the Houseman method.²⁰⁵ In both models, technical covariates were treated as random effects and the remaining covariates were treated as fixed effects. Family structure was also adjusted according to cohort-specific strategies. In ARIC, EWASs were performed separately by ancestry. Multi-ethnic studies with a small number of subjects in each ancestry, namely CARDIA and CHS, performed the EWAS in a pooled population. Additionally, subgroup

analyses in hypertensive and normotensive groups were conducted to investigate distinct etiology depending on the existence of the most important risk factor, hypertension. Hypertension was defined if either systolic or diastolic blood pressure 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.

Epigenome-wide meta-analysis

The meta-analysis combined EWAS results based on sample-size weighted z-score-based fixed-effect method in METAL.²⁰⁶ This study primarily aims to identify novel DNA methylation loci for WMH burden rather than estimate accurate effect sizes of methylation probes. For this reason, a fixed-effect model-based meta-analysis was chosen to maximize statistical power in this study.²⁰⁷ EWASs were first combined altogether for a pooled analysis, then race-specific EWASs were meta-analyzed within each ancestry. Subgroup meta-analyses were also performed within either hypertensive or normotensive groups, in a pooled population and in each ancestry. An association was considered as significant if the P value is below 0.05 either of Q value (Benjamini-Hochberg false discovery rate²⁰⁸) or of Bonferroni threshold corrected for the number of methylation probes (approximately 1.0E-7). The less stringent threshold was also set as 1.0E-06 to detect putative associations. Probes on sex chromosomes and cross-reactive probes reported by Chen et al.²⁰⁹ were removed from the meta-analyses *post hoc*.

Differentially methylated region analysis

We performed a differentially methylated region (DMR) analysis to identify a group of DNA methylation sites that collaboratively influence WMH burden. Two specific

methods, Comb-P²¹⁰ and DMRcate²¹¹ compute a region-based P value using summary statistics of individual DNA methylation sites accounting for their spatial correlations. Comb-P computes correlations at varying distance lags and uses them to compute correlated P values at each CpG site. DMRcate collapses contiguous significant CpGs that are within a certain number of nucleotides from each other. Multiple testing error was corrected with a one-step Sidak method.²¹² We considered a DMR as valid if it was identified by both Comb-P and DMRcate.

Biological functions associated with EWAS findings

To explore population structures and inter-relationships in DNAm, Spearman correlations among 11 EWAS probes (“target probes”) were calculated in 906 EA and 639 AA subjects in the ARIC study. DNAm may regulate functional DNA elements in specific tissues. The RegulomeDB²¹³ annotates DNA features and regulatory elements such as transcription factor binding sites, position-weight matrix for transcription factors (TF) binding (PWM), DNase footprinting, DNase sensitivity, chromatin states, eQTLs, differentially methylated regions, manually curated regions and validated functional SNPs to genomic positions of interest. We scored genomic positions of target probes according to the RegulomeDB’s criteria ranging from 5 (minimal binding evidence) to 1 (likely to affect binding and linked to expression of a gene target). Probes at the locations with significant DNA features or regulatory functions (score 1 or 2) were discussed. Furthermore, we investigated tissue- and cell type-specific regulation by our target probes using the experimentally derived database, the Functional element Overlap analysis of ReGions from EWAS (eFORGE ver. 2.0).^{214,215} Probes were pruned to be at least 1 kb apart then tested for

overlap with DNase I hypersensitive sites against 1,000 annotation-matched background probe sets from the 850K array, ENCODE, and Roadmap Epigenomics data. Hypergeometric P values were generated from counts of overlaps between a TF and all target probes, and counts of overlaps between a TF and all probes.^{214,215} Multiple-testing corrected Q values were calculated using the Benjamini-Yekutieli method.²¹⁶ TF binding sites were considered “significantly enriched” where Q value<0.05.

To explore the shared genetics between WMH burden and its comorbidities, genes mapped to significant DNAm probes and DMRs were searched within publicly available GWAS results of relevant traits *i.e.* blood-pressure, dementia, Alzheimer’s disease (AD), stroke and white matter abnormalities.²¹⁷ The search terms for each trait are as below.

Relevant trait	Search term
Blood Pressure (BP)	BP, Systolic BP, Systolic BP change measurement, Diastolic BP, Diastolic BP change measurement, hypertension, mean arterial pressure, pulse pressure measurement
Dementia	dementia, vascular dementia, lewy-body dementia measurement and lewy-body dementia
Stroke	stroke, ischemic stroke, cardioembolic stroke, small vessel stroke and small artery occlusion in brain
Alzheimer’s disease (AD)	AD and late-onset AD

In addition, genes mapped to significant probes and DMRs were tested for gene-enrichment. We investigated tissue-specific expression patterns of the genes and biological pathways that the genes are enriched in via *GENE2FUNC* analysis using the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA)²¹⁸ platform. Gene

enrichment was evaluated with a hypergeometric test against the gene sets obtained from two large databases, namely MsigDB²¹⁹ and WikiPathways²²⁰..

Shared epigenetics with cognitive ability (CA)

Cognitive decline is one comorbidity of CSVD, in particular of WMH.^{31,32,68,221} To investigate the shared epigenetics between WMH burden and CA, we compared our WMH burden EWAS with our recent EWAS of seven cognitive ability tests:²²² animal naming (Semantic Verbal Fluency abbreviated as “*Animal*”), letter fluency (Phonemic Verbal Fluency abbreviated as “*Verbal Fluency*”), executive function (Trail Making Test Part B abbreviated as “*Trails*”), verbal declarative memory (Wechsler Logical Memory abbreviated as “*Logical Memory*”), processing speed (Wechsler Digit Symbol Test/Symbol digit Modalities Test/Letter Digit Substitution Test abbreviated as “*Digit Symbol*”), vocabulary (Boston Naming Test or National Adult Reading Test abbreviated as “*Vocabulary*”) and global cognitive function (Mini-Mental State Examination abbreviated as “*MMSE*”). Also, a DMR analysis was performed using summary statistics and regions identified in both WMH burden and CA. In the CA EWAS, the reduced- and full-adjustment models included similar sets of covariates as in our WMH EWAS; thus, comparisons were made in both models.

Furthermore, we performed a multivariate association test to identify pleiotropic DNA methylation which influences both WMH burden and CA. DNA methylation probes associated with both traits were tested against the null hypothesis $H_0: \beta_{WMH} = \beta_{CA} = 0$. The test uses Z-scores for each trait and estimates multivariate test statistics accounting for the trait correlation. This method is implemented in the ‘*metaUSAT*’ software.²²³ The trait correlation matrix R , 2×2 correlation matrix of two traits, is estimated using Z-scores of the

null SNPs which are not associated with any trait (P_{CA} and $P_{WMH} > 1E-5$). P values were Bonferroni-adjusted with the number of DNA methylation probes included in each analysis.

Aim 2: Genetic control and transcriptional impact of WMH epigenetics

To achieve interventions based on WMH burden epigenetics, it is crucial to understand the causal relationship with genetics and transcriptome. This aim is designed to investigate 1) genetic control behind the target probes and 2) the probes' biological functions (represented by gene expression). A recent study showed that DNAm is heritable and 50% of CpG sites had evidence of a significant genetic component.²²⁴ To examine genetic control behind the WMH epigenetics, we estimated heritability of WMH-associated probes ("the target probes"), and performed a methylation quantitative locus (mQTL) analysis. Based on the mQTL associations, we also tested the causal direction between each target probe and WMH burden. Secondly, we investigated gene expressions associated with the target probes to imply biological functions in action.

Heritability of WMH-associated DNAm

For the target probes, the narrow sense heritability (h^2_{meth}) was estimated in the FHS Offspring Cohort subjects (N= 2,377). The narrow sense heritability is defined as the proportion of the total phenotypic variance of a DNA methylation level explained by the additive polygenic variance: $h^2_{meth} = \sigma^2_A / \sigma^2_{meth}$, where σ^2_A denotes the additive polygenic genetic variance and σ^2_{meth} denotes the total methylation variance at the probe. Age, sex, blood cell counts, principal components and technical covariates were included for

adjustment in the reduced model. BMI and smoking status were additionally corrected in the full model.

Identification of methylation quantitative loci (mQTLs)

DNAm levels at the target probes may be influenced by nearby (cis-effect) or distant (trans-effect) DNA sequence. Methylation QTL analysis was performed for the target probes in EA (n=984) and AA (n=2,225) individuals from ARIC. Our target CpGs are from WMH EWAS (Aim 1). Inverse-normal transformed methylation beta values were regressed for age, sex, and significant population structure factors (up to the first 10 genetic principal components (PCs) and the first 10 methylation PCs in each ancestry). A genome-wide association study was then performed on 1000 Genomes Phase I imputed SNPs (GRCh37/hg19). SNPs with low imputation quality ($r^2 < 0.3$) and with low minor allele frequency (MAF < 0.05) were excluded from analyses. Association from mQTL analysis was regarded as genome-wide significant if the P value is below $5E-8$. Since cis-effects tend to be significantly larger than trans-effects²²⁵, our primary mQTL for later analyses was limited to cis-effect mQTLs. The cis-significance was set to 0.05 divided by the number of mQTLs in the region ($CpG \pm 1Mb$). We validated significant mQTL associations with mQTL associations in the FHS mQTL study of 4,170 EA subjects²²⁶.

Bi-directional Mendelian Randomization (MR) analysis

To determine if the DNAm level is a causal factor for WMH burden (Forward MR) or a secondary outcome of WMH burden (Backward MR), we performed a bidirectional two-sample MR analysis.²²⁷ Two sample MR analysis requires summary statistics including the effect size (β) and its standard error for exposure and outcome traits from independent

samples. For DNAm associations, we used our mQTL analysis result, and for WMH associations, a previous GWAS in UK biobank (UKBB) participants (n=11,226) was used.¹²⁵ Summary statistics were downloaded from the GWAS catalog website (Accession ID: GCST007305) accessed on 01/09/2019. Forward MR analysis tested if any of the target probes influences WMH. Instrumental variables (IVs) for forward MR analysis are cis-effect mQTLs identified in ARIC EA participants. To meet the fundamental assumption of independence in MR analysis, IVs were pruned at $r^2 < 0.2$. For target probes having at least three IVs, we performed a multi-instrument MR analysis. The backward MR analysis was used to test if target probes are influenced by WMH. Twenty three sentinel SNPs were reported by previous GWASs (including the recent meta-analysis under review) and an exome-wide association study.^{125,152,154} The SNP rs12120143 was excluded from analysis due to low frequency in the UKBB GWAS (minor allele frequency (MAF) =0.02). Sentinel SNPs that did not exist in the UKBB GWAS were replaced with a SNP in high linkage disequilibrium (LD) in the HapMap European population - rs7894407 replaced with rs2176448 ($r^2=0.94$), rs17576323 with rs13417165, rs17148926 with rs11957829 and rs3215395 with rs3891038 ($r^2=1.0$).

For each target probe in both MR directions, causation was tested based on the inverse-variance weighted (IVW) effects across all IV-DNAm and IV-WMH estimates in the R package *TwoSampleMR*.²²⁸ Causation with P value <0.05 was considered significant. To validate the MR result, built-in tests for pleiotropy, heterogeneity and directionality were also performed. For existence of pleiotropy (MR egger intercept test P value <0.05), Egger regression estimates was assessed instead of the IVW method.

Gene expression associated with WMH epigenetics

Gene expression associated with target probes was investigated to functionally annotate target probes. First, we tested associations between a DNAm level at each probe and an expression level at the nearest gene in 1,966 EA subjects from FHS. The DNAm β score at each target probe was regressed on the expression level adjusting for age, sex, population and family structure, blood cell counts and technical covariates. Technical covariates and family structure were modeled as random effects. Secondly, to explore the genome-wide effects of target probes, transcriptome-wide association studies (TWASs) were performed in the FHS (n=1,966) and RS (n=728) subjects. In this analysis, we included the expression level as a dependent variable and the DNAm level as an independent variable with the same set of covariates. In RS, mRNA expression levels were dependent variables. Estimates from two studies were combined for each gene using the sample-sized based meta-analysis method in METAL. We corrected the associations for multiple-testing burden based on the Bonferroni method. As a secondary threshold, Q value using the FDR method was calculated, and associations with $Q < 0.05$ were considered significant.

Aim 3: Integrative genetic analyses of WMH burden

In the last aim, we used multi-dimensional data to get holistic evidence to support WMH burden etiology. Firstly, we performed gene-based pathway analyses on WMH association studies at genome-wide, epigenome-wide and transcriptome-wide levels and merged them to identify the common biological function shared by multiple layers. Second, we tested co-localization of quantitative trait loci (QTLs) associated with multi-dimensional

traits (*i.e.* DNAm, gene expression and WMH burden) at the WMH loci to show how the genetics influences WMH burden involving different molecular levels. Lastly, causation from DNAm to WMH mediated by gene expression was tested using two-step MR analysis.

Two-step Mendelian randomization

To confirm if the functionally annotated gene expression mediates the relationship between the target probe and WMH burden, a two-step Mendelian randomization (2SMR) analysis was performed.²²⁹ In the first step, we tested the directional relationship from “*the exposure*” target probes to “*the mediator*” gene expression. The ARIC mQTL associations for target probes were selected as IVs after pruning to be independent ($r^2 < 0.2$). For the IVs, we confirmed associations with gene expression in the FHS whole blood eQTL data ($n=5,257$) and in the GTEx ver.7 brain eQTL data. For brain tissues, cortex, frontal cortex, anterior cingulate cortex, and caudate basal ganglia were selected. The second step was to establish the causation from the implicated genes to “*the outcome*” WMH. For the genes confirmed significant in the first step, eQTLs were taken as IVs and pruned to be independent ($r^2 < 0.2$). Then we evaluated associations with WMH in the UKBB GWAS.¹²⁵ Gene expression validated in both steps with P value < 0.05 was considered as a mediator in the directional relationship of the target DNAm and WMH burden.

Identification of biological pathways with multi-dimensional data

Based on hierarchical gene-based permutation, we performed marker set enrichment analyses (MSEAs) on genetic, epigenetic and transcriptional associations of WMH in the R package *Mergeomics*.^{230,231} For the genetic association, we meta-analyzed two WMH GWASs (CHARGE¹⁵² and UK Biobank¹²⁵) based on the sample-size weighted z-score based

fixed-effect. Markers were then pruned to be independent ($r^2 < 0.5$) based on HapMap3 LD information. For epigenetic and transcriptional associations, we used our EWAS result and a recent TWAS of WMH,²³² respectively. Markers were primarily mapped to the nearest genes. For epigenetic markers, cis-acting genes reported in the MesaEpiGenomics study²³³ were additionally annotated. We tested marker-level enrichment with permutation size 20,000 based on biological pathways from pre-defined public databases: KEGG, REACTOME, Biocarta and the Gene Ontology (GO) knowledgebase.^{234,235} Then, enriched results from individual studies were meta-analyzed using the Meta MSEA function in the package. Pathways with a permutation-based P value smaller than 0.05 were considered significant.

Co-localization of multi-dimensional QTLs

The P value based Bayesian co-localization test was conducted using QTLs associated with multi-dimensions – genotype associated with WMH (QTL), with DNAm (mQTL) and with gene expression (eQTL) - at the WMH loci using the R package “*coloc*”.²³⁶ In this analysis, “the WMH locus” indicates the cytoband where a previously reported sentinel variant is located. We have selected QTLs ± 500 kb of 23 sentinel SNPs from WMH GWASs including the recent meta-analysis, 11 sentinel DNAm probes from our EWAS and 13 genes from WMH TWAS.^{125,152,153,232} For QTL, we used the UKBB GWAS associations (n=11,226). For mQTL and eQTL, we identified associations at the target loci from publicly available data. In FHS, mQTLs and eQTLs were identified using whole blood samples from 4,170 and 5,257 EA subjects.^{226,237} To explore co-localization in brain tissues, we also used the GTEx Consortium data (version 7) in four brain tissues – cortex (n=175),

frontal cortex (n=143), caudate basal ganglia (n=174) and anterior cingulate cortex (n=129).²³⁸ Due to limited number of subjects in the GTEx brain eQTL data, we took the FHS whole blood as a primary endpoint.²³⁹ The GTEx version 7 data was obtained from the GTEx Portal on 12/19/19. Out of 13 TWAS loci, the *ARL17A* locus did not have mQTL associations, therefore it was excluded from the analysis. We evaluated the posterior probability of full co-localization (PPFC), that multiple traits (WMH, gene expression and DNAm) share a causal variant at each locus, given the data, using the R *moloc* package. Co-localization was tested for a unique combination of DNAm probe and a gene with at least 5 QTLs in common. We considered the WMH locus with a PPFC greater than 80% as a convincingly co-localized locus.

RESULTS

Aim 1: Meta-analysis of epigenome-wide association studies (EWASs) of WMH burden

Study sample characteristics

The number of subjects in each participating study ranged from 194 to 1,545. Mean age ranged from 49.7 to 74.6 years, and proportion of female participants ranged from 40.5 to 72.8%. Mean BMI was similar among studies ranging from 26.4 to 31.0 kg/m². Smoking prevalence was the highest in the ADNI study (40.8%), but similarly low among other studies (from 5.6% to 20.7%). Prevalence of hypertension ranged from 21.6 to 85.4 %. Two (ARIC and GENOA) studies included subjects of African ancestry (AA) and seven other cohorts and ARIC included subjects of European ancestry (EA). CARDIA and CHS included subjects of multiple ancestries, but they contributed only to the trans-ethnic analyses due to small sample sizes. Median WMH burden (log transformed WMH grade or volume) ranged from 0.18 to 2.1 (Q1: 0.10 to 1.81; Q3: 0.41 to 2.56). Numbers of markers and subjects, and demographic descriptions in participating studies are presented in **Table S1**.

Epigenome-wide association study (EWAS) model diagnostics

The genomic inflation statistics (λ s) were heterogeneous in studies ranging from 0.71 to 1.33. Individual studies corrected the statistics according to their protocols where significant inflation exists. The lambdas in the trans-ethnic meta-analyses were 1.03 and 1.01 for the reduced and full-adjustment model, respectively (**Table 2 and Figure S2**). In

all meta-analyses including race-specific and blood-pressure subgroup analyses, we did not observe significant inflation evidence, with the lambdas from 0.93 to 1.02.

Table 2 Genomic inflation in meta-analyses

	Pooled			Hypertensive			Normotensive		
Model	All	EA	AA	All	EA	AA	All	EA	AA
Reduced	1.03	1.00	0.97	1.02	0.99	0.97	1.01	1.01	0.95
Full	1.01	0.99	0.99	1.01	0.97	0.98	0.99	0.99	0.93

Meta-analysis of EWASs

We included 6,019 (EA: 4,452 and AA: 995) and 5,589 (EA: 4,023, AA: 994) subjects in reduced- and full-adjustment models, respectively. There was no genome-wide significant methylation probe either at Bonferroni corrected P value threshold (P value $<1.12 \times 10^{-7}$) or a Benjamini-Hochberg false discovery rate (Q <0.05). However, with suggestive significance (P value $<1.0 \times 10^{-6}$), three probes, cg24202936 mapped to septin 7 pseudogene (*LOC441601*; P value= 3.78×10^{-7}), cg06809326 mapped to coiled-coil domain containing 144 A N-terminal pseudogene (*CCDC144NL*; P value= 6.14×10^{-7}) and cg03116124 mapped to tripartite motif containing 67 gene (*TRIM67*; P value= 9.32×10^{-7}) were identified in a reduced model. Adjusting for BMI and smoking, we identified cg17577122 mapped to claudin 5 gene (*CLDN5*; P value= 2.39×10^{-7}) in addition to cg24202936 (P value= 5.20×10^{-7}) and cg03116124 (P value= 6.55×10^{-7}). In a race-specific analysis, a suggestive association at cg04245766 mapped to bone morphogenetic protein 4 gene (*BMP4*; P value= 3.78×10^{-7}) was identified from the full-adjustment analysis in EA

subjects. Top associations from trans-ethnic analyses are shown in **Table 3** and Manhattan plots of epigenome-wide associations are presented in **Figure S1**.

Table 3 Top associations from a meta-analysis

Probe	Gene	Regulatory feature	Chr	Position	Z score	P value	Q value	Weight
Reduced								
cg24202936	<i>LOC441601</i>	Body	11	50257256	5.08	3.78E-7	0.122	5,663
cg06809326	<i>CCDC144NL</i>	TSS200	17	20799526	4.99	6.14E-7	0.122	5,663
cg03116124	<i>TRIM67</i>		1	231293208	-4.91	9.32E-7	0.122	5,433
Full								
cg17577122	<i>CLDN5</i>	1 st Exon; Body	22	19511967	5.12	2.39E-7	0.106	5,234
cg24202936	<i>LOC441601</i>	Body	11	50257256	5.02	5.20E-7	0.106	5,234
cg03116124	<i>TRIM67</i>		1	231293208	-4.97	6.55E-7	0.106	5,004

Stratified meta-analyses by hypertensive status

At the genome-wide significance level, no probe was further identified in the stratified analysis (**Table 4**). However, different sets of putative associations were identified in hypertensive and normotensive groups. In hypertensive subjects, cg25317585 mapped to *FGF14* (fibroblast growth factor 14) showed robust significance in both reduced- and a full-adjustment models (P value= 2.05E-7 and 2.76E-7). In normotensive subjects, two probes, namely cg06450373 mapped to *CDH18* (cadherin 18) (P value=3.36E-7) and cg18950108 mapped to *DPCR1* (diffuse panbronchiolitis critical region protein 1) (P value=4.22E-7) were identified as suggestive associations from the reduced model. We also performed race-specific stratified analyses. In hypertensive AA subjects, a putative association at cg02741985 mapped to coiled-coil domain containing 57 gene (*CCDC57*) was identified (P value=3.93E-7) from the reduced-model. More suggestive associations were found in normotensive EA subjects; cg07675682 at the *SLC26A11* (solute

carrier family 26 member 11) locus (P value =6.30E-7) and cg10051414 at the *MAFF* (MAF bZIP transcription factor F) locus (P value =9.37E-7) in the full model. So far, from the pooled and stratified analyses, we identified 11 probes in the association with WMH burden. To evaluate whether the effects of those 11 probes differ by hypertensive status, we showed the association statistics in hypertensive and normotensive groups (**Table S2**).

Table 4 Top associations from a stratified meta-analysis

Probe	Gene	Regulatory feature	Chr	Position	Z score	P value	Q value	Weight
Hypertensive subgroup								
Reduced								
cg25317585	<i>FGF14</i>	Body;1 st exon	13	102568886	5.19	2.05E-07	0.098	3,006
cg19059707			12	64080622	-4.47	7.66E-06	0.594	1,377
cg16790423	<i>HSPAIL</i>	5'UTR	6	31781328	-4.45	8.61E-06	0.594	1,377
Full								
cg25317585	<i>FGF14</i>	Body;1 st exon	13	102568886	5.14	2.76E-07	0.121	2,858
cg27501293	<i>SFTPD</i>	3'UTR	10	81697500	-4.86	1.16E-06	0.187	2,858
cg08106319	<i>ZFAND3</i>	Body	6	37849427	-4.61	4.05E-06	0.470	2421
Normotensive subgroup								
Reduced								
cg06450373	<i>CDH18</i>	Body	5	19766434	-5.10	3.36E-07	0.101	2,437
cg18950108	<i>DPCR1</i> ; <i>MUCL3</i>	Body	6	30920171	-5.06	4.22E-07	0.101	2,599
cg20443254	<i>ASCL4</i>	1 st Exon	12	108169020	4.37	1.23E-05	0.797	2,599
Full								
cg25313122			7	6121416	4.81	1.53E-06	0.371	1,926
cg12176783	<i>TCEA2</i>	TSS200;5'UTR	20	62694000	-4.69	2.74E-06	0.384	2,321
cg06450373	<i>CDH18</i>	Body	5	19766434	-4.62	3.78E-06	0.384	2,186

Differentially methylated regions (DMRs)

Fourteen DMRs were identified from the reduced model (Šidák P value <0.05) and regions near *PRMT1*, *CCDC144NL-AS1*, *IFITM10*, *VARS2*, *SLC12A4*, *SLC45A4*, *KITLG* and *ZNF311* loci were overlapped with intron or exon of genes (**Table 5**). Five regions near *PRMT1*, *ZNF311*, *SLC45A4*, *LCP1* and *KAT6B* were also identified from the full-

adjustment model as well as six additional regions; DMRs at the *PRMT1*, *ZNF311*, *HIVEP3*, *SLC45A4* and *TCEA2* loci were within intron or exon of genes. In a DMR analysis of EA population, 13 and 15 DMRs were identified from the reduced- and the full-adjustment models, respectively (**Table S3**). Among DMRs identified from the pooled analysis, the *PRMT1*, *CCDC144NL-AS1* and *LCPI* loci in the reduced model and the *PRMT1*, *ALLC*, *ZNF311* and *LCPI* loci in the full model were replicated in the EA population. DMRs at the *PRMT1*, *KDM2B*, *ATP13A4*; *ATP13A4-AS1*, *SPIRE2*, *SH3PXD2A*, *LCPI*, *RHPN1* and *LOC101930114* loci were identified regardless of adjustment models in EA. In the AA population, 15 and 17 DMRs were identified in the reduced- and the full-adjustment model, respectively, and none of pooled analysis results were replicated. Nine loci (*MIR596*, *LOC100294145*, *LAMB2*, *C3orf56*, *FAM120B*, *WASH3P*, *EDARADD*, *EBF3*, and *TM9SF2*) overlapped in both models. None of the DMRs overlapped between two races.

Table 5 DMR analysis result

Chr	start	end	#probe	P _{Comb-P}	P _{DMRCate}	Gene	Distance from Gene
Reduced-adjustment Model							
19	50191341	50191882	6/8	7.97E-12	3.25E-11	<i>PRMT1</i>	0
17	20799407	20799769	10/13	1.63E-11	5.57E-14	<i>CCDC144NL</i>	0
						<i>-AS1</i>	
11	1769288	1769645	8/10	5.92E-07	2.12E-06	<i>IFITM10</i>	0
6	30882707	30883203	6/13	4.31E-05	1.11E-06	<i>VAR2S2</i>	0
6	28973317	28973497	9/17	1.86E-04	9.61E-06	<i>ZNF311</i>	0
11	130185383	130185670	3/3	1.23E-03	7.96E-04	<i>ZBTB44</i>	-776
7	157092681	157092982	2/2	1.35E-03	2.64E-03	<i>UBE3C</i>	30,615
8	142233381	142233705	5/5	1.35E-03	3.50E-03	<i>SLC45A4</i>	0
12	88974242	88974665	8/11	3.88E-03	1.13E-02	<i>KITLG</i>	0
13	46757317	46757415	3/3	4.52E-03	5.25E-03	<i>LCPI</i>	-858
16	67997857	67998164	4/5	8.70E-03	2.03E-02	<i>SLC12A4</i>	0
10	76573395	76573506	2/4	1.31E-02	4.87E-04	<i>KAT6B</i>	-12,664
22	50984367	50984597	3/3	2.37E-02	2.74E-02	<i>KLHDC7B</i>	-1,864
2	233284401	233284661	2/3	2.81E-02	5.05E-03	<i>ALPPL2</i>	8,977
Full-adjustment Model							
19	50191341	50191683	5/8	1.54E-07	4.58E-07	<i>PRMT1</i>	0
6	28973317	28973497	9/17	1.52E-04	2.59E-05	<i>ZNF311</i>	0
1	42384389	42384564	3/10	2.79E-04	1.56E-05	<i>HIVEP3</i>	0
8	142233426	142233642	3/7	5.72E-04	5.96E-06	<i>SLC45A4</i>	0
10	76573395	76573506	2/5	1.87E-03	8.35E-05	<i>KAT6B</i>	-12,664
2	3704500	3704793	6/7	3.07E-03	1.55E-02	<i>ALLC</i>	-992
13	46757317	46757415	3/3	4.39E-03	4.97E-03	<i>LCPI</i>	-858
20	62693942	62694005	6/9	8.72E-03	6.92E-03	<i>TCEA2</i>	0
3	145879429	145879710	6/2	1.78E-02	4.44E-02	<i>PLD2</i>	-147
11	89956572	89956777	6/8	4.82E-02	1.42E-02	<i>CHORDC1</i>	-40
6	31148369	31148666	14/16	4.57E-02	8.50E-03	<i>PSORS1C3</i>	-2,693

P_{Comb-P}: combined p value after Sidak multiple testing adjustment, P_{DMRCate}: minimum adjusted p-value from the CpGs constituting the significant region

DMR analysis in hypertensive subgroup

In the hypertensive subgroup, five DMRs were identified from the reduced model and seven from the full model, and *OSR2* and *ATP13A4*; *ATP13A4-AS1* loci were identified

in both (**Table S4**). In the EA specific analysis, nine and twelve associations were identified from the reduced and full models (**Table S5**). Irrespective of adjustment models, DMRs at the *PON3*, *RGMA* and *ATP13A4*; *ATP13A4-AS1* loci were identified. Among DMRs identified in the pooled hypertensive population, DMRs at the *ATP13A4/ATP13A4-AS1* and *MIR4291* were replicated in the EA analyses. In the AA specific analysis, ten and eleven DMRs were identified from the reduced and full models (**Table S5**). Among them, DMRs at *LDLRAD4*, *RABL6*, *SLC36A3*, *FAM222A*, *RASIP1* and *MUC4* were found regardless of models. One of ten DMRs at the *CCDC57* locus involved the cg02741985 probe, which was a suggestive signal from our meta-analysis in the AA hypertensive subgroup. The *LDLRAD4* locus was one of the loci found in the hypertensive pooled analysis. None of the DMRs overlapped between two races.

DMR analysis in normotensive subgroup

In the pooled population, 13 and 16 DMRs were identified to be associated with WMH in the reduced- and the full-adjustment model, respectively (**Table S6**). Five regions at the *CHORDC1*, *FBXO47*, *LINC01983*, *NFAM1* and *TCEA2* loci were found in both models. Five (at the *DRC7*, *FBXO47*, *KDM2B*, *NRBP1*; *KRTCAP3* and *TCEA2* loci) and six (at the *DRC7*, *FBXO47*, *LINC00702*, *NFAM1*, *TCEA2* and *THBS1* loci) were replicated in the respective models in the EA population. In the EA population, 10 and 12 DMRs were found to be associated with WMH in two models, and five loci of *CD248*, *DRC7*, *FBXO47*, *KHDC3L* and *TCEA2* were identified regardless of adjustment models (**Table S7**). Among the ten DMRs identified in the reduced model, a region at the *CCDC144NL* locus involved the cg06809326 probe which was identified in a meta-analysis of the pooled population. In

the AA population, a different set of DMRs were identified from both models (**Table S7**). DMRs at the *DNHDI*, *F10*, *LAMB2*, *LOC151174*; *LOC643387*, *PAPD7*, and *VTRNA2-1* loci showed significance irrespective of models. Only one DMR at the *LINC01983* locus from the pooled analysis was replicated in AA.

Functional genomics of target probes

Significant correlations ($r > 0.3$ with $p\text{-value} < 0.05$) among cg03116124, cg06450373, and cg18951018 were observed in both ancestries. Correlations among 11 EWAS probes (“target probes”) are presented in **Figure S3**. Genomic positions of 11 EWAS probes (“target probes”) were searched in the RegulomeDB. We identified that cg17577122, cg06809326 and cg10051414 are situated in the genomic position likely to affect protein bindings. In umbilical vein endothelial cells, RNA Polymerase II Subunit A (*POLR2A*) was found to be likely bound to the genomic positions of three probes: cg17577122 (chr22:19511967), cg04245766 (chr14:54421051) and cg10051414 (chr22:38598733). Interestingly, the positions of cg06809326 (chr17:20799526), cg10051414 (chr22:38598733) and cg25317585 (chr13:102568886) were likely to be bound by the CCCTC-binding transcriptional regulator protein (*CTCF*) in the brain microvascular endothelial cells. Proteins binding to target probes in brain tissues and endothelial cells are shown in **Table S8**. The eFORGE tested enrichment of transcription factor (TF) binding sites for 11 target probes based on common TF databases. DNase I hypersensitivity signals surrounding target probes are shown in **Figure S4**. We observed significant overlap with transcription factor binding sites at 5 target CpGs: cg06450373, cg06809326, cg17577122, cg24202936, and cg25317585 (**Table S9**).

Genetic sharing with blood pressure, stroke, dementia and white matter-related traits

From the GWAS catalog, we identified 1,990 BP genes, 59 white matter trait genes, 317 stroke genes, 67 dementia genes and 913 Alzheimer's disease genes as of 03/06/2019. Among 11 genes identified with suggestive evidence from our EWAS, *CDH18* was reported in the association with blood-pressure. From 159 genes annotated to significant DMRs, we found 15 BP-related genes (*CETP*, *NRBP1*, *PRDM8*, *TM9SF2*, *WBP1L*, *UBE3C*, *TNXB*, *TCEA2*, *MCF2L*, *HIVEP3*, *PLEC*, *IRF6*, *GRAMD1B*, *RPTOR* and *LOC102546294*), four stroke-related genes (*SLC44A2*, *CETP*, *NRBP1* and *SH3PXD2A*) and one Alzheimer's disease gene (*IRF6*).

Genetic function enrichment in FUMA

We input four genes, *CLDN5*, *CCDC144NL*, *TRIM67* and *BMP4*, mapped to WMH-associated methylation probes for the GENE2FUNC program implemented in the FUMA platform. Three genes, *BMP4*, *CLDN5* and *TRIM67*, were enriched in the regulation of cell development, especially the first two genes were enriched in the regulation of endothelial cell differentiation (**Table 6**). We also performed a DMR analysis in CA using summary statistics. Based on genes mapped to all DMRs we identified, we found enriched pathways related to lipoprotein metabolism and transport. Genes mapped to the probes found in the pooled and race-specific analyses, were significantly enriched with previous GWAS findings in HIV-1 control and squamous cell lung carcinoma.

Table 6 FUMA gene-enrichment results

Gene Set	Size	P value	P _{FDR}	genes
Single				
Positive regulation of endothelial cell differentiation	24	2.68E-6	0.020	<i>BMP4, CLDN5</i>
Regulation of endothelial cell differentiation	42	8.35E-6	0.031	<i>BMP4, CLDN5</i>
Positive regulation of cell development	538	1.41E-5	0.035	<i>TRIM67, BMP4, CLDN5</i>
Positive regulation of epithelial cell differentiation	65	2.02E-5	0.037	<i>BMP4, CLDN5</i>
Outflow tract morphogenesis	77	2.84E-5	0.042	<i>BMP4, CLDN5</i>
DMR – all genes				
High density lipoprotein particle remodeling	15	1.80E-5	0.040	<i>APOC3, SCARB1, CETP</i>
Regulation of plasma lipoprotein particle levels	45	1.85E-5	0.040	<i>APOC3, SCARB1, CETP</i>
Reverse cholesterol transport	17	2.68E-5	0.040	<i>APOC3, SCARB1, LMF1, RP11-161M6.2, CETP</i>
DMR – excluding stratified analyses				
HIV-1 control	22	4.35E-6	0.008	<i>RNF39, VARS2, PSORS1C3</i>
Squamous cell lung carcinoma	132	4.15E-5	0.038	<i>HIVEP3, PLOD2, ZNF311, VARS2</i>

Shared epigenetics with cognitive ability

In seven CA tests, 61 methylation probes were reported with putative associations (P value <10E-5).²²² For 51 probes also included in our WMH EWAS, we confirmed the epigenetic associations in both CA and WMH EWASs at the Bonferroni significance threshold of 9.8E-4 (0.05/51). We identified three methylation probes significantly shared with both CA and WMH burden. In the reduced model, cg17417856 mapped to *PRMT1* ($P_{WMH}=2.40E-06$ and $P_{CA}=1.82E-4$ in Digital Symbol). In the full model, cg18676110 mapped to *RAFI* ($P_{WMH}=1.12E-06$ and $P_{CA}=3.67E-4$ in Vocabulary). In both models, cg05704155 mapped to *PGD* was identified for the MMSE test ($P_{WMH}=1.07E-06$ and $P_{CA}=8.65E-4$ in the reduced model; $P_{WMH}=1.07E-06$ and $P_{CA}=4.11E-4$ in the full model). A

multivariate joint analysis of two EWASs identified significant probes shared by both WMH burden and CA: 115 probes for Digit Symbol, 10 for MMSE, 12 for Verbal Fluency and 1 for Vocabulary. Weak associations in one trait were strengthened by the other trait based on the trait correlation (**Table S10**). In **Table 7**, we showed the significant shared epigenetic association which were significant for individual traits (P_{WMH} and $P_{\text{CA}} < 1\text{E-}3$). Two methylation probes, cg17417856 (*PRMT1*) and cg05704155 (*PGD*), that we identified in the comparisons were confirmed with the same CA domains. Especially, *PGD* was consistently identified regardless of statistical models. *PGD* encodes a 6-phosphogluconate dehydrogenase enzyme which has previously been reported to be involved in hyperglycemia-induced oxidative stress in the neonatal diabetic brain.²⁴⁰

Table 7 Multivariate joint analysis of CA and WMH epigenetic associations

Test	CpG	Gene	Z _{WMH} , P _{WMH}	Z _{CA} , P _{CA}	P _{joint}	Q _{joint}
Reduced model						
Verbal Fluency	cg12582616	<i>JARID2</i>	-3.54, 3.96E-04	-4.15, 3.38E-05	4.66E-08	9.83E-03
	cg12025310	<i>TMPRSS6</i>	3.43, 6.00E-04	-4.05, 5.16E-05	1.36E-06	4.77E-02
MMSE	cg05951221		-3.61, 3.12E-04	4.84, 1.29E-06	1.45E-08	6.10E-03
	cg25923609		3.59, 3.32E-04	-4.46, 8.17E-06	8.71E-08	7.35E-03
	cg05704155	<i>PGD</i>	4.55, 5.26E-06	-3.33, 8.65E-04	1.39E-07	9.75E-03
	cg01127300		-3.97, 7.21E-05	3.74, 1.86E-04	3.93E-07	2.37E-02
	cg19859270	<i>GPR15</i>	-3.87, 1.09E-04	3.60, 3.18E-04	9.48E-07	4.44E-02
	cg05951221		-3.61, 3.12E-04	5.75, 9.12E-09	4.55E-10	4.79E-05
Digit Symbol	cg17417856	<i>PRMT1; C19orf76</i>	-4.72, 2.40E-06	3.74, 1.82E-04	3.77E-08	1.22E-03
	cg19859270	<i>GPR15</i>	-3.87, 1.09E-04	4.26, 2.00E-05	1.54E-07	3.41E-03
	cg04885881		-3.34, 8.48E-04	4.25, 2.17E-05	9.43E-07	9.02E-03
	cg01127300		-3.97, 7.21E-05	3.42, 6.33E-04	2.09E-06	1.40E-02
	cg04994217	<i>C9orf170</i>	-3.32, 9.12E-04	3.64, 2.69E-04	9.07E-06	3.38E-02
	cg05951221		-3.61, 3.12E-04	5.75, 9.12E-09	4.55E-10	4.79E-05
Full model						
MMSE	cg05704155	<i>PGD</i>	4.88, 1.07E-06	-3.53, 4.11E-04	1.42E-08	6.00E-03

Test: A cognitive ability test domain, P_{WMH}: P value estimated in the reduced-adjusted trans-ethnic WMH burden EWAS, P_{CA}: P value estimated in the reduced-adjusted trans-ethnic cognitive abilities EWAS, P_{joint} and Q_{joint}: P value and local FDR-adjusted P value estimated in the joint analysis

In a reduced-adjusted DMR analysis for CA, we identified 130 DMRs associated with 7 CA tests: 9 for Animal naming, 35 for Digit Symbol, 14 for Logical Memory, 15 for MMSE, 15 for Trails, 26 for Verbal Fluency, and 16 for Vocabulary. In a fully-adjusted DMR analysis for CA, we identified 98 DMRs associated with 7 CA tests: 10 for Animal naming, 21 for Digit Symbol, 12 for Logical Memory, 15 for MMSE, 15 for Trails, 15 for Verbal Fluency, and 10 for Vocabulary. We identified a DMR within *ALPP2* (2q37.1) shared by Digit Symbol test and WMH burden with reduced-adjustment (Sidak $P_{WMH}=0.028$ and Sidak $P_{CA}=9.10E-11$). In EA subjects, a DMR near *SH3PXD2A* was identified in both adjustment models (Sidak $P_{WMH}=0.028$ and Sidak $P_{CA}=3.82E-4$ in the reduced model; Sidak $P_{WMH}=0.028$ and Sidak $P_{CA}=7.55E-4$ in the full model).

Aim 2: Genetic control and transcriptional impact of WMH epigenetics

Heritability of the target probes

For 11 methylation probes which we identified from trans-ethnic, race-specific and stratified meta-analyses, heritability was estimated in two linear models. We found that a big proportion (48.2%) of DNAm variance at cg06450373 was explained by the additive genetic variance (**Table 8**). Moderate heritability was estimated for cg02741985 (36.3%) and cg06809326 (26.5%). Small, but significant heritability estimates were calculated for cg24202936 (15.5%), cg25317585 (12.0%) and cg17577122 (14.3%). For cg03116124, cg07675682, cg10051414 and cg18950108, we could not identify meaningful heritability estimates. Estimates additionally adjusted for BMI and smoking were not significantly different from the reduced model estimates.

Table 8 Heritability estimates of WMH-associated methylation probes

Probe ID	Reduced model			Full model		
	h^2	S.E.	P value	h^2	S.E.	P value
cg02741985	0.363	0.071	2.95E-08*	0.367	0.071	2.06E-08*
cg03116124	0.009	0.069	4.46E-01	0.006	0.069	4.67E-01
cg06450373	0.482	0.075	1.92E-11*	0.480	0.075	2.35E-11*
cg06809326	0.265	0.075	1.03E-04*	0.266	0.075	9.51E-05*
cg07675682	0.000	-	5.00E-01	0.000	-	5.00E-01
cg10051414	0.009	0.073	4.51E-01	0.007	0.073	4.64E-01
cg18950108	0.000	-	5.00E-01	0.000	-	5.00E-01
cg24202936	0.155	0.073	1.34E-02*	0.158	0.073	1.17E-02*
cg25317585	0.120	0.072	4.31E-02*	0.116	0.072	5.07E-02
cg17577122	0.143	0.078	2.80E-02*	0.150	0.078	2.27E-02*

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

Methylation QTLs for target probes

To show genome-wide patterns of genetic control, we performed GWASs of 11 target probes in ARIC study (984 EA and 2,225 AA subjects). Cg04245766 was only included in the analysis of AA subjects since the probe did not pass the QC criteria in the EA population. The genome-wide pattern of genetic control in target probes are shown in Manhattan plots (**Figure S5**). As shown in the plots, we observed the genetic influence in *cis* on cg02741985, cg06450373, cg06809326, cg07675682 and cg24202936 regardless of ancestry. Cg02741985 showed more amplified *cis*-genetic influence in AA subjects and cg03116124 appeared to have *cis*-influence only in AA subjects. *Cis*-mQTLs were defined in EA subjects for a Mendelian randomization analysis. In EA, numbers of markers within *cis*-regions of target probes were 5,902 for cg02741985, 5,207 for cg03116124, 5,118 for cg06450373, 3,070 for cg06809326, 6,538 for cg07675682, 4,346 for cg10051414, 4,475 for cg17577122, 15,062 for cg18950108, 5,903 for cg24202936 and 5,317 for cg2531758.

The mQTL associations significant at Bonferroni threshold ($0.05/\text{number of markers in cis-region}$) were identified for four target probes: 57 mQTLs for cg02741985, 74 for cg06450373, 162 for cg06809326 and 2,981 for cg24202936. We validated the same four probes with mQTLs in the FHS mQTL study,²²⁶ the same four target probes were also identified with significant mQTLs.

Causal relationship from target probes to WMH burden

The forward MR analysis was performed for four target probes having mQTLs. We found a significant causal relationship from cg06809326 to WMH burden using three independent genetic proxies. The odds ratio (OR) was estimated as 1.78, indicating that a 1% increase in methylation level at cg06809326 is associated with a 78% increment in WMH burden unit. Three IVs for cg06809326 were rs116963987, rs73298018 and rs7501645. In the backward MR analysis, we confirmed that WMH burden does not significantly influence the methylation level at cg06809326. In addition, a more causal relationship that a standard deviation increase in WMH burden leads to 14 % higher methylation level was observed at cg03116124 ($P \text{ value}=0.03$). Pleiotropy tests for MR associations in both directions showed non-significant P values except for two backward MR analyses in cg03116124 and cg06450373. For the analyses with pleiotropy evidence, we took MR Egger-based estimates as the primary result instead of inverse variance-weighted estimates. Results from all MR analyses are presented in **Table 9**.

Table 9 Bi-directional MR analysis result

Probe ID	No.IV	Beta (S.E.)	OR[95%CI]	P _{MR}	P _{pleiotropy}
Forward					
cg02741985	2	-1.56E-2(0.428)	0.98[0.43,2.28]	0.97	NA
cg06450373	2	3.89E-2(0.280)	1.04[0.60,1.81]	0.89	NA
cg06809326	3	0.5751(0.178)	1.78[1.25,2.52]	1.21E-3*	-0.80
cg24202936	3	0.193(0.214)	1.21[0.80,1.85]	0.37	0.56
Backward					
cg02741985	18	3.50E-3(1.65E-2)	1.00[0.97,1.04]	0.83	-0.57
cg03116124	18	0.134(5.65E-2)	1.14[1.02,1.28]	3.07E-2*	-1.80E-2*
cg06450373	18	0.100(5.45E-2)	1.11[0.99,1.23]	0.08	-4.88E-2*
cg06809326	18	-2.39E-3(2.36E-2)	1.00[0.95,1.04]	0.92	-0.53
cg07675682	18	1.60E-3(1.77E-2)	1.00[0.97,1.04]	0.93	-0.19
cg10051414	18	2.59E-5(4.38E-3)	1.00[0.99,1.01]	1.00	-0.72
cg17577122	18	3.53E-3(2.81E-3)	1.00[1.00,1.01]	0.21	-0.43
cg18950108	18	-4.84E-3(1.61E-2)	1.00[0.96,1.03]	0.76	-0.12
cg24202936	18	-2.13E-2(1.92E-2)	0.98[0.94,1.02]	0.27	-0.25
cg25317585	18	5.81E-4(1.25E-2)	1.00[0.98,1.03]	0.96	0.40

No.IV: Number of instrumental variables included. OR: Odds ratio, P_{MR} and Beta (S.E.): P value, effect size and its standard error for the causal direction, primarily based on the inverse variance-weighted method, or in existence of pleiotropy (P_{pleiotropy} < 0.05) based on MR Egger method. P_{pleiotropy}: P value estimated from MR Egger intercept pleiotropy test

Gene expression associated with target probes

Excluding missing cg04245766, the association between each of 10 target probes and its nearest gene was tested. As shown in **Table 10**, we found significant associations for relationships between cg03116124 and *TRIM67* (P value=0.02) and between cg06809326 with *CCDC144NL* (P value=0.03).

Table 10 Associations between target probes and the nearest genes

CpG	Gene	Chr	Start	Stop	Coefficient	SE	P-value
cg03116124	<i>TRIM67</i>	1	231244657	231357304	2.0E-2	8.9.E-3	2.4.E-2*
cg06809326	<i>CCDC144NL</i>	17	20739760	20844644	-9.7E-3	4.5.E-3	3.3.E-2*
cg06450373	<i>CDHI8</i>	5	19473080	19988319	-9.1E-3	9.6.E-3	0.35
cg25317585	<i>FGF14</i>	13	102344869	103054799	-1.6E-3	1.9.E-3	0.40
cg17577122	<i>CLDN5</i>	22	19510550	19515068	-6.1E-4	1.1.E-3	0.59
cg24202936	<i>LOC441601</i>	11	50205720	50280104	-3.1E-3	6.6.E-3	0.65
cg07675682	<i>SLC26A11</i>	17	78194232	78227297	1.7E-3	4.5.E-3	0.71
cg02741985	<i>CCDC57</i>	17	80081038	80191192	-1.1E-3	3.1.E-3	0.72
			80059350	80111346	8.8E-4	3.2.E-3	0.78
cg18950108	<i>DPCR1</i>	6	30902400	30921998	-8.0E-4	2.3.E-3	0.72
cg10051414	<i>MAFF</i>	22	38576894	38612509	-1.1E-4	9.3.E-4	0.91

Next, we performed TWASs of ten target probes to examine both cis- and trans-effects. In FHS (n=1,966), 17,873 transcripts were included, and a study-level Bonferroni threshold was set to 2.80E-6. No gene-DNA probe association was identified at the transcriptome-wide significance level in FHS. In RS (n=728), TWASs were performed with 21,238 mRNA transcripts, 1 gene (*CHKB*; choline kinase-like protein) for cg06809326, 7 genes for cg02741985, 38 genes for cg03116124, 61 genes for cg06450373, 296 genes for cg18950108 and 4 genes for cg253147585 were identified as significant at the Bonferroni threshold of 2.35E-6. Among significant associations, three WMH-associated genes²³² - *FCRL6*, *TGFBR3* and *IL4R* – were found to be associated with our target probes

(cg03116124, cg06450373 and cg08950108) (**Table 11**). However, a meta-analysis did not find any significant association due to lack of significance in FHS.

Table 11 TWAS results in target probes for WMH-associated genes

CpG Site	Gene	CHR	FHS (n=1,966)		RS (n=728)		Meta-analysis				WMH TWAS	
			T	Q value	T	Q value	N	Z-score	P value	HetP	T	Q value
cg03116124	<i>TGFBR3</i>	1	0.56	0.858	-5.00	0.002	2,683	0	1.00	0.31	4.27	0.036
cg06450373	<i>TGFBR3</i>	1	-0.43	0.976	-5.89	2.16E-05	2,686	-0.99	0.32	0.41	4.27	0.036
cg18950108	<i>TGFBR3</i>	1	-0.22	0.997	-11.40	2.58E-06	2,691	0.24	0.81	0.41	4.27	0.036
cg18950108	<i>FCRL6</i>	1	-1.85	0.250	-7.38	2.58E-06					5.23	0.003
cg18950108	<i>IL4R</i>	16	0.97	0.997	4.72	0.002	2,691	1.53	0.13	0.52	-5.50	2.61E-04

CHR: chromosome. HetP: P-value for heterogeneity test ($H_1: Z_{FHS} \neq Z_{RS}$). FDR Q: false-discovery rate adjusted P-value

Aim 3: Integrative genetic analyses of WMH burden

Gene expression as a mediator in DNAm-WMH causal relationship

Previously we confirmed that four of our target probes (*i.e.* cg02741985, cg06450373, cg06809326 and cg24202936) are under *in cis* genetic control. Also, we found that cg06809326 was associated with the nearest gene *CCDC144NL*. In the two-step Mendelian randomization analysis, we confirmed that *CCDC144NL* mediates the relationship between cg06809326 and WMH burden in caudate basal ganglia, but not in the other brain tissues or whole blood (**Table 12**). The selected instrumental variable for step 1, rs7501645 was in high linkage disequilibrium (LD) with the instrumental variable for step 2, rs11653586 ($r^2=0.98$).

Table 12 Two-step MR analysis result

Step 1 (CpG -> Gene)			Step 2 (Gene->WMH burden)			
Beta±S.E.	IV	P _{IV}	Beta±S.E.	IV	P _{IV}	Tissue
11.52±2.43	rs7501645	2.13E-6	0.05±0.03	rs11653586	4.80E-2	Caudate basal ganglia

Meta-analysis of enriched pathways from multi-dimensional associations

Based on independent SNPs, we observed three significantly enriched gene ontologies: extracellular matrix constituent conferring elasticity, microfibril and camera-type eye development (**Table 13**). From epigenetic markers, copulation, insulin secretion involved in cellular response to glucose stimulus, positive regulation of heat generation, pre-synapse assembly and negative regulation of dopamine secretion were significantly enriched. Lastly, from transcriptional markers, we found two ontologies including adaptive immune response and co-receptor activity enriched. Pathway-based meta-analysis of the three MSEA outputs identified that multi-omics associations in WMH burden were enriched in extracellular matrix structural constituent, camera-type eye development, alphaV-beta3 integrin-*IGF-I-IGFIR* complex, retinol dehydrogenase activity and phosphatidylinositol-4-phosphate binding.

Table 13 Meta-analysis of marker set enrichment

MODULE	P	FDR	DESCR
<i>GWAS</i>			
GO:0030023	2.38E-05	4.35E-02	Extracellular matrix constituent conferring elasticity
GO:0001527	3.03E-05	4.58E-02	Microfibril
GO:0043010	4.70E-05	4.89E-02	Camera-type eye development
GO:0035583	4.98E-05	5.01E-02	Sequestering of TGF-beta in extracellular matrix
GO:0006066	1.53E-04	9.15E-02	Alcohol metabolic process
GO:0008090	2.04E-04	0.103	Retrograde axonal transport
GO:0060346	2.31E-04	0.108	Bone trabecula formation
GO:0004022	2.98E-04	0.118	Alcohol dehydrogenase (NAD) activity
GO:0035276	3.14E-04	0.120	Ethanol binding
rctm0368	3.28E-04	0.120	Elastic fiber formation
<i>EWAS</i>			
GO:0007620	7.60E-06	0.026	Copulation

MODULE	P	FDR	DESCR
GO:0035773	7.86E-06	0.026	Insulin secretion involved in cellular response to glucose stimulus
GO:0031652	1.34E-05	0.036	Positive regulation of heat generation
GO:0099054	1.76E-05	0.039	Presynapse assembly
GO:0033602	3.48E-05	0.049	Negative regulation of dopamine secretion
GO:0033685	3.89E-05	0.051	Negative regulation of luteinizing hormone secretion
GO:0034260	4.33E-05	0.052	Negative regulation of GTPase activity
rctm0313	4.64E-05	0.053	Degradation of GABA
GO:0007269	4.66E-05	0.053	Neurotransmitter secretion
GO:0009450	4.71E-05	0.053	Gamma-aminobutyric acid catabolic process
<i>TWAS</i>			
GO:0002250	4.27E-06	0.022	Adaptive immune response
GO:0015026	1.10E-05	0.037	Coreceptor activity
M4085	3.00E-05	0.068	Primary immunodeficiency
GO:0006955	3.07E-05	0.069	Immune response
M6856	7.03E-05	0.134	Hematopoietic cell lineage
GO:0009897	9.84E-05	0.148	External side of plasma membrane
GO:0030020	1.02E-04	0.150	Extracellular matrix structural constituent conferring tensile strength
rctm0115	1.13E-04	0.152	Antigen activates B cell receptor leading to generation of second messengers
GO:0000145	1.28E-04	0.155	Exocyst
GO:0070371	1.48E-04	0.157	ERK1 and ERK2 cascade
<i>Combined</i>			
GO:0005201	4.23E-06	0.019	Extracellular matrix structural constituent
GO:0043010	8.26E-06	0.021	Camera-type eye development
GO:0035867	8.47E-06	0.022	AlphaV-beta3 integrin-IGF-1-IGF1R complex
GO:0004745	9.52E-06	0.023	Retinol dehydrogenase activity
GO:0070273	1.99E-05	0.039	Phosphatidylinositol-4-phosphate binding
GO:0006955	4.03E-05	0.069	Immune response
rctm1075	6.36E-05	0.096	SHC-related events triggered by IGF1R
M6856	7.41E-05	0.101	Hematopoietic cell lineage
GO:0019841	8.74E-05	0.108	Retinol binding
GO:0032024	1.44E-04	0.163	Positive regulation of insulin secretion
rctm0115	1.77E-04	0.169	Antigen activates B cell receptor leading to generation of second messengers

Genetic loci associated with DNAm, gene expression and WMH burden

At the WMH loci identified previously - 23 GWAS, 11 EWAS and 13 TWAS loci- we tested co-localization for each combination of DNAm probe and gene expression. The numbers of combinations were 182,499 in whole blood and 472 (77 in the same cytoband) combinations were identified as significant (PPFC>0.8). Significant co-localization combinations were located for 4 GWAS loci including rs3215395 at 7q22.1, rs6587216 at 17p11.2, rs8071429 at 17q21.31 and rs7214628 at 17q25.1. (**Table 14**) At 7q22.1, we observed multi-level QTLs co-localized at *MAP11* (microtubule associated protein 11), *MOSPD3* (motile sperm domain containing 3) and *SLC12A9* (potassium-chloride transporter 9). At 17p11.2, 10 probes were co-localized with *CCDC144A* (coiled-coil domain containing 144A), 15 probes with *FAM106A* (family with sequence similarity 106 member A) and 6 probes with *GRAPL* (growth factor receptor-bound protein 2-related adaptor protein). At 17q21.3, *KANSL1* (lysine acetyltransferase 8 regulatory nonspecific lethal complex subunit 1) and *PLEKHMI* (pleckstrin homology domain containing protein family member 1) were identified as co-localized. At 17q25.1, we identified *EXOC7* (exocyst complex component 7); *ZACN* (zinc activated ion channel), *GALK1* (galactokinase 1); *ITGB4* (integrin subunit beta 4), *H3F3B;H3F3A;H3F3C* (H3 histone family member 3B;3A;3C), *TEN1*(telomere length regulation protein TEN1 homolog);*TEN1-CDK3* (cell division protein kinase 3);*ACOX1* (acyl-CoA oxidase 1), *UNC13D* (unc-13 homolog D) and *WBP2* (WW domain binding protein 2) co-localized with WMH QTLs and mQTLs. Three genes with full co-localization – *UNC13D*, *ACOX1* and *WBP2*- were previously reported;^{150,152} however, other

reported genes including *TRIM47*, *TRIM67*, and *FBF1* did not show the full co-localization. One of the EWAS loci, cg06809326 is also located at 17p11.2, and one of the TWAS loci, *ARL17A* is also located at 17q21.31. Significant co-localization at cg06809326 was not identified; however, 10 methylation probes were co-localized with *CCDC144A*, which is a paralog of the gene mapped to the probe, *CCDC144NL*. We did not conduct the analysis at *ARL17A*, since the cis region for this gene does not have any mQTLs.

Table 14 MOLOC result in whole blood

Gene	CpG	nsnp	PPFC	PPFC_SNP
7q22.1				
<i>MAP11</i>	cg21406967	14	0.87	7:100285888
<i>MOSPD3</i>	cg04742719	23	0.85	7:100359454
<i>SLC12A9</i>	cg17194182	21	0.89	7:100300898
17p11.2				
<i>CCDC144A; CCDC144B; CCDC144C</i>	cg01534423	14	0.89	17:19165179
	cg15971010	6	0.89	17:19253613
	cg21495715	16	0.89	17:19165179
	cg13305951	14	0.87	17:19172196
	cg16674433	16	0.86	17:19172196
	cg04279801	17	0.85	17:19204863
	cg13832201	23	0.85	17:19196440
	cg27589594	25	0.85	17:19165179
	cg20538986	25	0.84	17:19194812
	cg11465639	29	0.84	17:19165179
<i>FAM106A; FAM106CP; FAM106B</i>	cg15971010	9	0.92	17:19253613
	cg21158317	17	0.91	17:19204863
	cg13305951	17	0.91	17:19224397
	cg21495715	17	0.91	17:19224397
	cg01627492	10	0.91	17:19165179
	cg16674433	17	0.9	17:19247075
	cg04279801	17	0.9	17:19239432
	cg01534423	17	0.89	17:19213335
	cg22671072	17	0.89	17:19194812
	cg27589594	17	0.89	17:19165179
	cg13832201	17	0.88	17:19196440
	cg14817867	17	0.87	17:19165179

Gene	CpG	nsnp	PPFC	PPFC_SNP
	cg18633784	17	0.85	17:19165179
	cg01735398	17	0.84	17:19165179
	cg12137220	17	0.81	17:19217261
<i>GRAPL;GRAP</i>	cg13305951	19	0.88	17:19196440
	cg01534423	20	0.88	17:19196440
	cg16674433	23	0.83	17:19196440
<i>GRAPL;LOC79999;LOC100132472</i>	cg21495715	19	0.89	17:19224397
	cg13305951	18	0.89	17:19224397
	cg01534423	20	0.84	17:19224397
<i>17q21.31</i>				
<i>KANSL1</i>	cg17123029	21	0.89	17:43153006
	cg14750215	46	0.85	17:43161353
<i>PLEKHM1;PLEKHM1P</i>	cg16826777	11	0.80	17:43376447
<i>17q25.1</i>				
<i>EXOC7;ZACN</i>	cg16649771	8	0.90	17:73926121
	cg02497559	10	0.90	17:73904155
	cg04225089	8	0.90	17:73896299
	cg19804488	16	0.87	17:73896299
	cg05611719	18	0.86	17:73904155
	cg10935138	19	0.83	17:73896299
	cg08125733	19	0.83	17:73896299
<i>GALK1;ITGB4</i>	cg10138630	9	0.94	17:73881056
	cg09795027	11	0.89	17:73881056
	cg03700492	9	0.89	17:73850819
	cg18082885	36	0.87	17:73881056
<i>H3F3B;H3F3A;H3F3C</i>	cg00759043	12	0.86	17:73904155
<i>TEN1;TEN1-CDK3;ACOX1</i>	cg18082885	6	0.94	17:73899533
<i>UNC13D</i>	cg26780705	14	0.97	17:73876473
	cg00759043	12	0.97	17:73931787
	cg23806715	19	0.94	17:73764500
	cg00943312	13	0.94	17:73874684
	cg19026231	23	0.92	17:73764500
	cg10295360	23	0.90	17:73874684
	cg20772300	38	0.84	17:73876473
	cg24166112	38	0.82	17:73874684
<i>WBP2</i>	cg26780705	14	0.98	17:73876473
	cg11015353	6	0.97	17:73884948
	cg15177359	26	0.97	17:73839366
	cg00943312	13	0.96	17:73874684
	cg23806715	19	0.94	17:73796617

Gene	CpG	nsnp	PPFC	PPFC_SNP
	cg10295360	23	0.92	17:73874684
	cg24563967	55	0.92	17:73874684
	cg00759043	12	0.91	17:73931787
	cg19026231	23	0.91	17:73775825
	cg06407111	91	0.87	17:73874684
	cg04442638	65	0.87	17:73874684
	cg24166112	42	0.87	17:73874684
	cg14955151	75	0.86	17:73874684
	cg20772300	42	0.86	17:73874684
	cg09253696	77	0.86	17:73874684
	cg03700492	73	0.85	17:73874684
	cg04225089	91	0.83	17:73874684
	cg10644544	65	0.82	17:73874684
	cg18082885	90	0.81	17:73874770

nsnp: number of SNPs included for co-localization test. PPABC_SNP: the SNP with the most evidence of colocalization

Out of 234,336 gene-probe combinations in cortex, significant co-localization evidence was observed in 22, in 9 of 232358 combinations in frontal cortex, 12 of 230,792 in anterior cingulate cortex, and 9 of 234,905 in caudate basal ganglia. In cortex, co-localizations at 7q22.1, 17q21.31 and 17q25.1 were observed, some of genes at 17q25.1 were replicated in other brain tissues – frontal and anterior cingulate cortex tissues and caudate basal ganglia tissue (**Table 15**). At 7q22.1, co-localization was identified for non-coding genes only. Many genes were co-localized at 17q25.1 including *TRIM47* in all cortex tissues and *FBF1* in all brain tissues. *TRIM47* was not included in the whole blood analysis since eQTLs at the locus did not have any *in cis* associations with the gene. There were also tissue-specific co-localizations. In only cortex tissue, we observed co-localization of QTLs, mQTLs for cg03706109 and eQTLs for *DKAKD* at 17q21.31. Co-localization at *TEN1* was confirmed in frontal cortex, *TRIM65* and *CDK3* in anterior cingulate cortex, and *WBP2* in caudate basal

ganglia. We confirmed the genes reported by previous GWASs with full-co-localization - *DCAKD* at 17q21.31, *FBF1*, *TRIM47*, *TRIM65* and *WBP2* at 17q25.1. Except for *WBP2*, these genes were not co-localized with whole blood eQTLs.

Table 15 MOLOC result in brain tissues

Symbol	Tissue	DNAm	nsnp	PPABC	PPABC_SNP
7q22.1					
<i>AC069294.1</i>	Caudate basal ganglia	cg04120578	66	0.91	7:100264893
		cg14016363	20	0.86	7:100359797
		cg00621899	53	0.86	7:100359797
		cg25923345	75	0.85	7:100264893
		cg25224090	71	0.84	7:100359797
		cg05392513	81	0.81	7:100359797
		cg22813798	148	0.81	7:100359797
<i>RP11-395B7.2</i>	Cortex	cg04120578	68	0.90	7:100264893
		cg25923345	79	0.82	7:100264893
17q21.31					
<i>DCAKD</i>	Cortex	cg03706109	148	0.87	17:43128906
17q25.1					
<i>CDK3</i>	Anterior cingulate cortex	cg00943312	13	0.82	17:73874684
<i>FBF1</i>	Anterior cingulate cortex	cg00759043	12	0.82	17:73943942
	Caudate basal ganglia	cg00759043	12	0.92	17:73909948
	Cortex	cg00943312	13	0.93	17:73890363
		cg00759043	12	0.92	17:73931787
		cg26780705	19	0.92	17:73876473
		cg10295360	23	0.84	17:73874684
		cg24166112	57	0.81	17:73874684
	Frontal cortex	cg00759043	12	0.91	17:73931787
<i>TEN1</i>	Frontal cortex	cg00943312	13	0.83	17:73874684
<i>TRIM47</i>	Anterior cingulate cortex	cg00943312	13	0.96	17:73874684
		cg26780705	19	0.96	17:73873656
		cg08125733	286	0.86	17:73850940
		cg10935138	278	0.85	17:73862646
		cg04442638	180	0.85	17:73874684
		cg09253696	217	0.85	17:73874684
		cg06407111	246	0.85	17:73874684
		cg04225089	279	0.83	17:73874684
		cg10295360	23	0.83	17:73874684
	Cortex	cg26780705	19	0.97	17:73876473

		cg00943312	13	0.94	17:73874684
		cg10295360	23	0.88	17:73876473
		cg20772300	57	0.84	17:73876473
		cg04442638	180	0.84	17:73874684
		cg06407111	246	0.83	17:73874684
		cg09253696	217	0.83	17:73874684
		cg24166112	57	0.83	17:73876473
		cg12714757	188	0.82	17:73876473
		cg04225089	279	0.82	17:73874684
		cg10935138	278	0.81	17:73862646
		cg14955151	136	0.81	17:73874684
		cg03700492	98	0.81	17:73876473
		cg10644544	79	0.80	17:73876473
	Frontal cortex	cg26780705	19	0.96	17:73873656
		cg00943312	13	0.95	17:73873656
		cg04442638	180	0.87	17:73874684
		cg09253696	217	0.86	17:73874684
		cg06407111	246	0.86	17:73874684
		cg10295360	23	0.85	17:73874684
		cg04225089	279	0.83	17:73874684
<i>TRIM65</i>	Anterior cingulate cortex	cg00759043	12	0.84	17:73943942
<i>WBP2</i>	Caudate basal ganglia	cg26780705	19	0.89	17:73876473

nsnp: number of SNPs included for co-localization test. PPABC_SNP: the SNP with the most evidence of colocalization

DISCUSSION

Our study identified 11 methylation probes associated with WMH burden. These probes mapped to genes that may have an implication in CSVD etiology. *CLDN5* (*claudin 5*) mapped to cg17577122 encodes integral membrane proteins that comprise tight junctions specifically in brain microvascular endothelial cells that regulate blood-brain barrier (BBB) permeability.²⁴¹ Claudin-5 is the most enriched tight junction protein at BBB, and its dysfunction has been implicated in neurodegenerative and neuroinflammatory diseases.²⁴² In a rat model of WM injury, a decrease in endothelial *CLDN5* gene expression was observed in regions of BBB damage.²⁴³ The profile of claudins and tight junction-associated marvel proteins was studied in rats; up-regulation of *Cldn5* after middle cerebral artery occlusion was paralleled with down-regulated *Cldn-1*, *3*, and *12*, and occludin in cerebral capillaries in pathological ischemia.²⁴⁴ Supporting this, *CLDN5* was enriched in the regulation of endothelial cell differentiation with another EWAS gene, *BMP4* (bone morphogenetic protein 4), in the FUMA gene enrichment analysis. Transforming growth factor beta 1 (*TGFB1*) is dysregulated in the heredity form of CSVD, CARASIL. BMPs are the largest group of the TGFB superfamily, and a study using post mortem samples from SVD patients showed that *BMP4* was highly expressed in white matter pericytes and suggested that ischemic white matter damage evolves in parallel with *BMP4* upregulation in pericytes,²⁴⁵ that inhibits maturation of oligodendrocytes and blocks the expression of myelin proteins.²⁴⁶

Accounting for the spatial correlation among DNAm probes, we could enhance DNAm associations for WMH burden. Among genes mapped to the DMRs, we could implicate several genes to the WMH pathology. Protein arginine N-methylase 1 (*PRMT1*)

transfers methyl group to arginine residues on histones; methylation mediated by *PRMT1* is promoted in genes involved in glioblastomagenesis.²⁴⁷ Generally, arginine methyltransferase enzymes (*PRMT1*, *PRMT4*, and *PRMT5*) regulate myeloid differentiation.¹³² Mice lacking *PRMT1* were characterized with severe central nervous system defects, particularly defects in oligodendrocyte maturation processes.²⁴⁸ *PRMT1* is a predominant type I protein arginine methyltransferase,²⁴⁹ which is mainly expressed in the endothelium and smooth muscle cells of the cardiovascular system.²⁵⁰ Under oxidative stress, type 1 PRMT enzymes enhanced by low-density lipoproteins methylate arginine competing against asymmetric dimethylarginine (ADMA); thus leading to increased level of ADMA that inhibits synthesizing nitric oxide resulting in endothelial dysfunction.^{251,252} DNAm at *PRMT1* was also identified as the shared epigenetics between cognitive ability (processing speed in specific) and WMH burden. The link between DNAm at *PRMT1* and cognition-related white matter changes needs to be further studied; however, the mechanism is likely to be related to ischemia in endothelial cells. Another DMR was mapped to the mast cell growth factor gene (*KITLG*, KIT ligand gene), encoding the ligand of the tyrosine-kinase receptor which is a pleiotropic factor in neural cell development. The gene is overexpressed by neurons to recruit neural stem cells to the cerebral injury site for angiogenic response.²⁵³

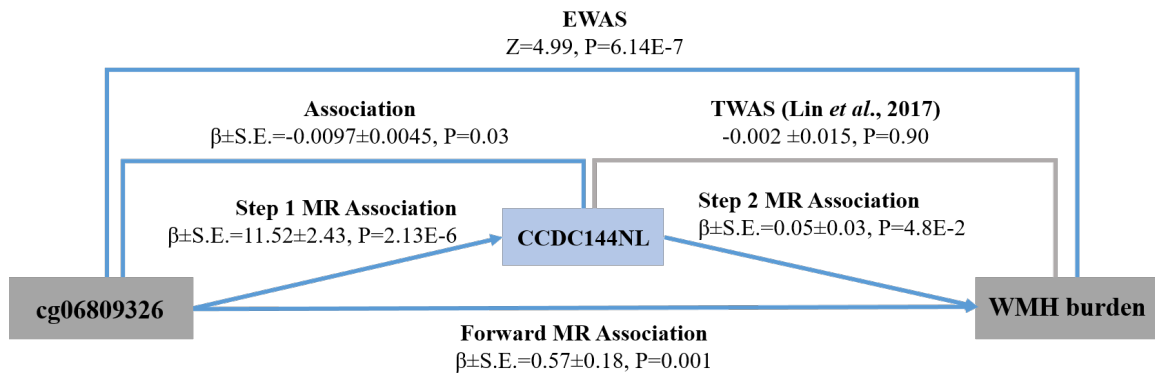
WMH burden-associated DMRs and the probes shared with cognitive abilities suggested that immunity- or infection- related functions are closely related to the CSVD etiology. Innate and adaptive immune responses coincide with leukocyte adhesion to the blood-vessel wall; therefore, the influx of leukocytes through endothelial cells influences the immune reactions in the tissues.^{254,255} In this study, genes mapped to DMRs were enriched in

HIV-1 control (enrichment FDR-adjusted P value =0.008) and squamous cell lung carcinoma²⁵⁶ (FDR-adjusted P value=0.038) (**Table 5**). The arginine methyltransferase enzyme encoded by *PRMT1* binds to the intracytoplasmic domain of the type I interferon complex; reduced *PRMT1* expression was found with the reduced antiproliferative effect of interferon-beta.²⁵⁷ More DMRs were mapped to genes in infection-related or immune system-related pathways. Zinc finger protein 311 (*ZNF311*) is known to be included in the herpes simplex virus 1 infection superpathway with many other zinc finger proteins.²⁵⁸ In addition, two genes, *UBE3C* and *KITLG*, involved in the class I MHC mediated antigen processing,²⁵⁹ lymphocyte cytosolic protein 1 (*LCPI*) associated with B-cell non-Hodgkin lymphoma,²⁶⁰ human immunodeficiency virus (HIV) type I enhancer binding protein 3 (*HIVEP3*) and mitochondrial valyl-tRNA synthetase 2 (*VAR2*) implicated in vaccinia virus infection,^{261,262} procollagen-lysine 5-dioxygenase 2 (*PLD2*) implicated in decreased HIV-1 infection,²⁷⁶ psoriasis susceptibility 1 candidate 3 (*PSORS1C3*),²⁶³ and solute blood cell counts (*SLC45A4*) associated with blood cell counts (plateletcrit, leukocyte and reticulocyte counts) were mapped to DMRs.^{264,265} *HIVEP3* has been also reported in blood pressure and eosinophil GWASs.^{265–268}

Cg06809326 mapped to *CCDC144NL* was identified in the trans-ethnic analysis when BMI, smoking status and blood pressure measures were not adjusted. Blood pressure is an important risk factor for WMH burden, and BMI and smoking are known to influence DNAm levels. To remove false positive signals, the risk factors should be corrected for in the models; however, the unadjusted association is still meaningful as a crude measure of DNAm-WMH burden relationship. And we confirmed that cg06809326 or *CCDC144NL*

have not been reported in the epigenetic association with BMI²⁶⁹, smoking²⁷⁰, or blood pressure²⁷¹. A regional plot at cg06809326 shows that the epigenetic locus is likely to associate with WMH burden with multiple signals (**Figure S5**). Correlations among methylation probes in 1,597 ARIC subjects were shown below the plot, probes making the association signal at the locus were significantly correlated. The heritability estimate for cg06809326 was 26.5% and the genetic control indicated by mQTL associations was present *in cis* in both EA and AA subjects (**Figure S6**). The directionality from the DNAm level at cg06809326 to WMH burden was established by the bi-directional MR analysis. To further elaborate the cg06809326-linked etiology, we confirmed the direct association with the gene expression of *CCDC144NL* as well as the gene's mediation in the relationship between the DNAm level at cg06809326 and WMH burden. However, a TWAS in FHS subjects did not report a significant association between *CCDC144NL* and WMH burden.²³² The *CCDC144NL* gene (coiled-coil domain containing 144 family (*CCDC144A*) N-terminal pseudogene) is a pseudogene associated with diabetic retinopathies.²⁷² Diabetic retinopathy is blood vessel damage in the retina, and it shares risk factors with CSVD such as high blood pressure and cholesterol levels. The coiled-coil protein, vimentin, assembles into filaments potentially involved in integrin-mediated cell adhesion and angiogenesis.²⁷³ In the backward MR analysis, we could find WMH burden influences the methylation level at cg03116124. Absence of genetic influence on the probe but association with WMH burden might be explained by this backward association. The DNAm level at cg03116124 can be the secondary outcome of WMH burden; and might have a potential as a serum biomarker. The cg03116124 probe was found to be associated with expression of *TRIM67*, a member of the

tripartite motif containing family; previous large-scale GWASs identified other members of the gene family - *TRIM47* and *TRIM65* - in the association with WMH burden or volume.^{125,150,152,274}



Only five of our target probes showed significant genetic components (Table 7), lack of genetic components may implicate environmental influences. Also, polygenic control was inferred by that we could not identify the genetic control *in cis* for heritable probes.

Cg17577122 was mapped to *CLDN5*, implicated in the CSVD etiology via endothelial cell-related mechanisms. The heritability estimate for cg17577122 was 14.3%; however, no mQTLs *in cis* were identified. Recently, a region-based EWAS reported a DMR mapped to *CLDN5* in the association with cognitive trajectory.²⁷⁵ In the study, genetic variations within *CLDN5* were associated with DNAm, but not with cognitive trajectory. Considering the correlation between WMH and cognitive function, the result supports our conclusion that WMH-associated DNAm at *CLDN5* is not under *in cis* genetic control. In our study, genetic controls over this probe appears to be scattered throughout the genome, especially in AA subjects multiple mQTL loci were shown (**Figure S6**). Generally, trans-effects of genetics

explain only a small portion of genetic control over DNA methylation;^{225,226} however, it is still possible that some DNAm is controlled by multiple genetic loci in a polygenic manner.

We studied DNAm levels in whole blood samples and the utility of DNAm probes as preventive biomarkers can be secured by checking the correlation with the levels in brain tissues. For 450,000 probes on the Illumina Human450 Methylation BeadChip, correlations between DNAm level in whole blood and the levels in four brain tissues (*i.e.* prefrontal cortex, entorhinal cortex, superior temporal gyrus and cerebellum) were reported on the Blood Brain DNA Methylation Comparison Tool website

(<https://epigenetics.essex.ac.uk/bloodbrain/>).²⁷⁶ We searched 11 target probes in the tool, and marginal to high correlations of DNAm levels in blood and brain tissues were observed at several target probes (**Table S11**). The correlation coefficients at cg06809326 were similarly significant in four tissues from 0.566 to 0.711. At cg04245766, DNAm levels in blood and the brain were highly correlated ($r=0.923$ to 0.954). Marginal correlation was confirmed at cg24202936 in prefrontal and entorhinal cortex and superior temporal gyrus ($r=0.324$ to 0.398). At some probes, tissue-specific correlation was found - $r=0.0494$ in entorhinal cortex (cg02741985), $r=0.242$ in cerebellum (cg03116124), $r=-0.312$ in cerebellum (cg10051414), and $r=0.255$ in superior temporal gyrus (cg18950108). Probes with high correlated blood and brain methylation levels are more likely to be eligible as an epigenetic biomarker for WMH burden.

Integrative genetic analyses provided us novel and additional information about the CSVD etiology. Using multi-level associations (GWAS, EWAS and TWAS), we meta-analyzed gene-based associations and identified biological functions associated with WMH

burden. A significant association was identified in the alphaV-beta3 integrin-*IGF-1-IGF1R* complex (FDR adjusted P value=0.022) ontology. *IGFBP-2*, which binds *IGF-1* to the integrin complex, was recently suggested as a biomarker for dementia and Alzheimer's disease.²⁷⁷ By co-localizing QTL, mQTL, and eQTL associations at WMH loci, our analysis linked several GWAS loci with DNAm and gene expression possibly involved in the pathology. Especially, the well-established WMH locus, 17q25.1, showed co-localization evidence for multiple genes including *GALK1*; *ITGB4*. Co-localization at the integrin subunit beta 4 gene integrin-related gene (*ITGB4*) again suggests integrin-mediated cell adhesion and angiogenesis as a possible etiological mechanism.

CONCLUSIONS

This dissertation study was designed to investigate the role of DNA methylation, a kind of epigenetic variations, in WMH burden, a shared pathology among critical public health burdens including stroke and dementia. As the first step, we have identified DNAm levels at 11 novel loci in association with WMH burden and implicated biological functions related to the genes mapped to those loci. In the next step, we characterized the relationship between DNAm at the identified probes and the genetics or gene expression. We confirmed significant genetic components in six DNAm probes and identified genetic variants (mQTLs) associated with four of them. Based on mQTLs, we tested a causal direction between DNAm and WMH burden in either way. DNAm at cg06809326 was found to influence WMH burden, while WMH burden influences DNAm at cg03116124. Additionally, expression levels of the closest genes to these two probes were confirmed to be associated with the DNAm at the probes (*i.e.* *CCDC144NL* for cg06809326, and *TRIM67* for cg03116124). The causal association from the genetics through DNAm and then gene expression towards WMH burden was also confirmed at cg06809326. We also investigated the WMH etiology utilizing multi-dimensional data available. We integrated genome-, transcriptome-, and our epigenome-wide associations and identified a novel ontology related to integrin complex associated with WMH burden. Had tested co-localization of genetic associations at multi-dimensions at the reported WMH burden loci, we found significant co-occurrences of mQTL, eQTL, and QTLs at four GWAS reported loci including *ITGB4* (integrin subunit beta 4). This indicates that some genetic influences lead to WMH burden through related epigenetics and gene expressions. In conclusion, we found novel loci associated with WMH burden and

functionally annotated these loci in relation to the etiology, suggesting involvement of the integrin-mediated pathway.

This is the first epigenome-wide association study of WMH in a large-scale. Aging influences DNA methylation,²⁷⁸ and WMH burden is known to be associated with older ages^{32,35,46–50}. Thus, epigenetics might unravel the unknown etiology of WMH burden. In this study, we identified novel epigenetic loci associated with WMH burden, and these loci might provide a novel explanation to the aging-related WMH pathology. The main finding of this study was the integrin-associated pathway. Integrins play a crucial role in cell-to-cell and cell-to-extra-cellular matrix interactions, and these interactions are essential for effective T-cell-mediated immune response.²⁷⁹ The ageing process often leads to reduced efficiency of adaptive immunity, while B-cell functions and innate immune system tend to maintain.²⁸⁰ Integrin-mediated functions might be able to explain the shared pathology of ageing and WMH burden. This study exhibits several limitations as well; however, those can be overcome with future studies. MetaUSAT multivariate analysis computes a joint statistic using two sets of summary statistics. In a multi-trait analysis of GWASs, the joint association becomes stronger when correcting for the correlation structure of SNPs. Likewise, the joint analysis can improve with a future study in the correlation structure among methylation probes. Heritability estimates showed that only five of our EWAS probes have genetic components. The target probes with small or non-significant genetic components need to be studied for their association with environmental factors including blood pressure. We observed that mQTLs showed underpowered significance in UKBB GWAS, this might improve with greater sample size. We used the summary statistics from a previous GWAS to establish directional associations for target probes. Mendelian randomization analysis

requires the effect size and its standard error for exposure and outcome associations to compute the statistics for directional association. The largest CHARGE GWAS including 17,936 EA subjects provides Z-scores only, so it was not applicable for MR analyses. We used the summary statistics from the UKBB GWAS; however, it covers fewer SNPs. Many mQTLs and eQTLs are not included in the UKBB dataset, and the CHARGE GWAS does not provide effect size. One of the target probes, cg04245766 mapped to *BMP4*, was not included in the causal association analyses since it did not pass QC criteria in EA subjects of ARIC, FHS, and RS. The EWAS association for cg04245766 was driven by ADNI, ARIC AA subjects, CARDIA, and SHIP-TREND. When mQTL associations for cg04245766 are available, we can test the directional association toward WMH burden with or without mediation of gene expression. In the functional analysis, the *BMP4* gene was enriched in endothelial cell differentiation and regulation; thus, confirming the causal pathway related to cg04245766 might be a valuable addition to the understanding of the WMH burden-associated epigenetics. Lastly, the Bayesian method-based co-localization test simply confirms the co-occurrence of independent associations, not providing the evidence on the causal pathway. Based on the co-localization results, the etiology of WMH burden involving genetics, DNAm and gene expression can be further investigated.

APPENDICES

Appendix A. Supplemental Material

IRB approval from the UTHSC committee



Dr. Myriam Fornage
UT-H - IMM - IMM - Genetics

Committee for the Protection of Human Subjects

6410 Fannin Street, Suite 1100
Houston, Texas 77030

NOTICE OF CONTINUING REVIEW APPROVAL

January 22, 2019

HSC-IMM-14-0008 - *An Integrated Genetic and Epigenetic Approach to Cerebral Small Vessel Disease*

PI: Myriam Fornage

PROVISOS: Unless otherwise noted, this approval relates to the research to be conducted under the above referenced title and/or to any associated materials considered at this meeting, e.g. study documents, informed consents, etc.

APPROVED: By Expedited Review and Approval

REVIEW DATE: 01/21/2019

APPROVAL DATE: 01/22/2019

CHAIRPERSON: Rebecca Lunstroth, JD

Two handwritten signatures in black ink, one appearing to be "RL" and the other a more complex signature, likely belonging to Rebecca Lunstroth and another committee member.

Upon review, the CPHS finds that this research is being conducted in accord with its guidelines and with the methods agreed upon by the principal investigator (PI) and approved by the Committee.

PLEASE NOTE: Due to revisions to the common rule that went into effect July 19, 2018, this study that was approved under expedited approval no longer needs to submit for continuing review. Changes to the study, adverse events, protocol deviations, personnel changes, and all other types of reporting must still be submitted to CPHS for review and approval. When this study is complete, the PI must submit a study closure report to CPHS.

CHANGES: The PI must receive approval from the CPHS before initiating any changes, including those required by the sponsor, which would affect human subjects, e.g. changes in methods or procedures, numbers or kinds of human subjects, or revisions to the informed consent document or procedures. The addition of co-investigators must also receive approval from the CPHS. ALL PROTOCOL REVISIONS MUST BE SUBMITTED TO THE SPONSOR OF THE RESEARCH.

UNANTICIPATED RISK OR HARM, OR ADVERSE DRUG REACTIONS: The PI will immediately inform the CPHS of any unanticipated problems involving risks to subjects or others, of any serious harm to subjects, and of any adverse drug reactions.

RECORDS: The PI will maintain adequate records, including signed consent documents if required, in a manner which ensures subject confidentiality.

Alzheimer's Disease Neuroimaging Initiative (ADNI)

Study Description

ADNI is a longitudinal, multi-site observational study including Alzheimer's disease (AD), mild cognitive impairment (MCI), and elderly individuals with normal cognition assessing clinical and cognitive measures, MRI and PET scans (FDG and 11C PIB) and blood and CNS biomarkers.^{281,282} In 2004, the initial phase, ADNI-1, recruited 400 subjects with MCI, 200 subjects with early AD and 200 cognitively normal elderly from multiple site in North America. The study was extended with ADNI-GO in 2009 including 500 ADNI-1 subjects (normal controls and MCI cases) and 200 new early MCI cases. In 2011, one more extension, ADNI-2, started to study approximately 450-500 ADNI-1 subjects, 200 ADNI-GO subjects and 650 new cases (150 controls, 150 early MCI cases, 150 late MCI cases and 200 mild AD cases). In this study, 387 normal or early MCI subjects who provided brain MRI and methylation data were included. Brain MRI for WMH burden assessment was taken at ADNI-1 in 73 subjects and at ADNI-2 in 314 subjects. Samples for DNA methylation were drawn in ADNI-GO and ADNI-2 studies. Days between blood sampling and brain MRI were within 365 days in 85% of subjects (n=329), between a year and two years in 8% (n=31), between two and three years in 16 subjects and more than 3 years in 11 subjects.

Ethic Statement

The study protocols for the ADNI studies were approved by the Institutional Review Boards of all the participating institutions.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and

DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Atherosclerosis Risk in Communities (ARIC)

Study Description

The ARIC Study is a prospective longitudinal investigation of the development of atherosclerosis and its clinical sequelae in which 15,792 individuals aged 45 to 64 years were

enrolled at baseline. A detailed description of the ARIC study has been reported previously.²⁸³ At the inception of the study in 1987-1989, participants were selected by probability sampling from four communities in the United States: Forsyth County, North Carolina; Jackson, Mississippi (African-Americans only); the suburbs of Minneapolis, Minnesota; and Washington County, Maryland. Four examinations were carried out at three-year intervals (exam 1, 1987-1989; exam 2, 1990-1992; exam 3, 1993-1995; exam 4, 1996-1998). A fifth clinical examination was completed in 2011-2013. Subjects were contacted annually to update their medical histories between examinations. After QC, 906 European-American and 639 African-American subjects with methylation data and brain MRI were included in this study.

Ethic Statement

Written informed consent was provided by all study participants, and the study design and methods were approved by institutional review boards at the collaborating medical institutions: University of Mississippi Medical Center Institutional Review Board (Jackson Field Center); Wake Forest University Health Sciences Institutional Review Board (Forsyth County Field Center); University of Minnesota Institutional Review Board (Minnesota Field Center); and the Johns Hopkins School of Public Health Institutional Review Board (Washington County Field Center).

Acknowledgments

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and

HHSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions. Funding was also supported by RC2HL102419 and R01NS087541. This study was funded by the NIH contract (5R01NS087541-04) granted to Myriam Fornage and Eric Boerwinkle.

Coronary Artery Risk Development in Young Adults (CARDIA)

Study Description

Coronary Artery Disease Risk Development in Young Adults (CARDIA) is a longitudinal study designed to trace the development of risk factors for coronary heart disease in 5,115 relatively young individuals aged 18-30 (52% African American and 55% women). The study collected information on anthropometric, lifestyle factors as well as biomarkers including brain MRI, genetic and DNA methylation data. Baseline measurements were repeated, and additional measurements performed, at Years 2, 5, 7, 10, 15, 20, and 25.²⁸⁴ DNA methylation level was measured in 1,089 subjects at the examination year 15 and 1,092 subjects at the year 20. In this study, 126 subjects of European and 68 subjects of African-ancestry with the brain MRI image were contributed from CARDIA.

Ethic Statement

The CARDIA study is multicenter with recruitment in Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA. The study protocol was approved by the institutional review boards of the coordinating center and the 4 participating field centers, and written informed consent was obtained from participants at all examinations.

Acknowledgments

The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005).

Cardiovascular Health Study (CHS)

Study Description

The CHS is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥ 65 years conducted across four field centers in the United States:

Sacramento County, California; Washington County, Maryland; Forsyth County, North Carolina; and Pittsburgh, Allegheny County, Pennsylvania. The original cohort of 5,201 persons, predominantly of European ancestry, was recruited in 1989-1990 from random samples of the Medicare eligibility lists. Later, 687 African-Americans were added to the cohort, making the total cohort of 5,888.²⁸⁵ DNA methylation was measured on a randomly selected subjects without presence of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack at the

exam year 5. Post QC, 378 European ancestry and African-American participants with methylation data and brain MRI were included in this study.

Ethic Statement

Written informed consent to genetic research has been obtained on all individuals included in this study. CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease. The DNA methylation dataset from CHS can be requested at <https://chs-nhlbi.org/node/6222>.

Acknowledgements

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. The CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, N01HC15103; and NHLBI grants U01HL080295, R01HL087652, R01HL092111, R01HL105756, R01HL103612, R01HL111089, R01HL116747 and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 and R01AG033193 from the National Institute on Aging (NIA) as well as Laughlin Family, Alpha Phi Foundation, and Locke Charitable Foundation. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Framingham Heart Study (FHS) offspring cohort

Study Description

The FHS is a single-site, community-based, ongoing cohort study that was initiated in 1948 to investigate prospectively the risk factors for CVD.^{286–288} The original cohort (n=5,209) was randomly recruited in the town of Framingham, MA, USA, and examined every 2 year since 1948; their children and spouses of the children (n=5,124), the offspring cohort, was enrolled in 1971 and examined approximately every 4 year since 1971. DNA methylation was measured on 2,846 offspring participants around 8th examination cycle (2005-2008). This study includes 403 and 1,323 offspring subjects with dse and flair MRI scans, respectively, as well as DNA methylation data.

Ethic Statement

Written informed consent to genetic research has been obtained on all individuals included in this study. Ethics permission for FHS and genetic research in FHS was obtained from the Institutional Review Board of Boston University Medical Campus (IRB number H-32132, H-26671)

Acknowledgements

This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study Contract No. N01-HC-25195 and No. HHSN268201500001I, and by grants from the National Institute of Aging (R01s AG033193, AG008122, AG054076, AG033040, AG049607, AG05U01-AG049505), and the National Heart, Lung and Blood

Institute (R01 HL093029, HL096917). The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, and by NIH contract N01-HC-25195. The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health.

Genetic Epidemiology Network of Arteriopathy (GENOA)

Study Description

The Genetic Epidemiology Network of Arteriopathy (GENOA) study was designed to investigate the genetics of hypertension and target organ damage in hypertensive sibships.²⁰ The GENOA study is a community-based study including African Americans from Jackson, Mississippi and non-Hispanic whites from Rochester, Minnesota.²⁸⁹ The initial GENOA study, from 1996 to 2001, recruited any members in sibships having more than two individuals with clinically diagnosed hypertension before age 60. Exclusion criteria of the GENOA study were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. Eighty percent of African Americans (1,482 subjects) and 75% of non-Hispanic whites (1,213 subjects) from the initial study population returned for the second examination (Phase II: 2001-2005). Study visits were made in the morning after an overnight fast of at least eight hours. Demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples were collected in each phase. DNA methylation levels were measured only in African Americans participants. After quality control, DNA methylation and brain MRI data were available for 356 African Americans.

Ethic Statement

Written informed consent was obtained from all subjects, and approval was granted by participating institutional review boards at the University of Mississippi Medical Center, the Mayo Clinic, and the University of Michigan.

Acknowledgements

Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the NHLBI (HL054457, HL100185, HL119443, and HL133221) and the NINDS (NS041558) of the NIH.

Lothian Birth Cohort 1936 (LBC1936)

Study Description

The Lothian Birth Cohorts of 1936 (LBC1936) is a longitudinal study of ageing. Participants were born in 1936 and mostly live in the Lothain region of Scotland.^{290,291} The study was derived of the Scottish Mental Surveys of 1947 that conducted a general cognitive ability test on nearly all 11 year old children in Scotland. Residents in the Lothian area of Scotland (n=1,091) were recruited in late-life at mean age 70 and followed-up at their ages 70, 73, and 76. In this study, 230 subjects who provided epigenetic information and brain imaging, blood biomarkers and anthropomorphic and lifestyle measures were included.

Ethic Statement

Following written informed consent, venesected whole blood was collected for DNA extraction. Ethics permission for the LBC1936 was obtained from the Multi-Centre Research Ethics Committee for Scotland (Wave 1: MREC/01/0/56), the Lothian Research Ethics

Committee (Wave 1: LREC/2003/2/29), and the Scotland A Research Ethics Committee (Waves 2 and 3: 07/MRE00/58).

Acknowledgements

We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by the Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. Analysis of these data were carried out within the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE). CCACE is supported by funding from the BBSRC, the Medical Research Council (MRC), and the University of Edinburgh as part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1). Funding from the Australian NHMRC (613608) supported quality control analysis of the methylation data.

Rotterdam Study III (RS3)

Study Description

The Rotterdam Study is a prospective, population-based cohort study that includes inhabitants of the well-defined district Ommoord, located in the city of Rotterdam in the Netherlands.²⁹² The cohort was initiated in 1990 among approximately 7,900 persons aged 55 years and older, who underwent a home interview and extensive physical examination at the baseline and during the follow-up rounds every 3-4 years (RS-I). The cohort was extended in 2000-2001 (RS-II, 3,011 individuals, aged 55 years and older) and 2006-2008 (RS-III, 3,932

individuals aged 45 and older). Collected data include genetic information, epigenetic information, imaging (of heart, blood vessels, eyes, skeleton and brain) and numerous blood biomarkers, anthropomorphic and lifestyle measures.^{293,294} Post QC, DNA methylation and brain MRI used for the analyses were available in 547 subjects from the first visit of RS-III and 442 subjects from the fifth visit of RS-I, the third visit of RS-II and the second visit of RS-III. The latter dataset is part of the Biobanking and Biomolecular Resources Research Infrastructure for The Netherlands (BBMRI-NL), BIOS (Biobank-based Integrative Omics Studies) project [<http://www.bbmri.nl/?p=259>]. Whole blood DNA methylation was quantified in a random subset of 750 individuals with genotyping and RNA expression data available in the first cohort of RS-III.

Ethic Statement

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

Acknowledgements

The generation and management of the Illumina 450K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands.

The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk, and Lisette Stolk PhD for their help in creating the methylation database.

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. HHHA is supported by ZonMW grant number 916.19.151.

The Study of Health in Pomerania TREND (SHIP-TREND)

Study Description

The Study of Health in Pomerania (SHIP-Trend) is a longitudinal population-based cohort study in West Pomerania, a region in the northeast of Germany, assessing the prevalence and incidence of common population-relevant diseases and their risk factors. Baseline examinations for SHIP-Trend were carried out between 2008 and 2012, comprising 4,420 participants aged 20 to 81 years. Study design and sampling methods were previously described.²⁹⁵ The medical ethics committee of the University of Greifswald approved the study protocol, and oral and written informed consents were obtained from each of the study

participants. DNA was extracted from blood samples of n=256 SHIP-Trend participants to assess DNA methylation using the Illumina HumanMethylationEPIC BeadChip array. Samples were randomly selected based on availability of multiple OMICS data, excluding type II diabetes, and enriched for prevalent MI. The samples were taken between 07:00 AM and 04:00 PM, and serum aliquots were prepared for immediate analysis and for storage at -80 °C in the Integrated Research Biobank (Liconic, Liechtenstein). Processing of the DNA samples was performed at the Helmholtz Zentrum München. After exclusion of subjects who refused participation or who fulfilled exclusion criteria for MRI (e.g. cardiac pacemaker), 2189 subjects from SHIP-Trend-0 underwent the MRI scanning. After exclusion of scans with technical artifacts, major structural abnormalities and stroke, full data sets with EWAS data and MRI scans were available in 214 subjects in SHIP-TREND.

Ethic Statement

SHIP and SHIP TREND were approved by the local ethics committee. After complete description of the study to the subjects, written informed consent was obtained. The medical ethics committee of the University of Greifswald approved the study protocol.

Acknowledgements

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network ‘Greifswald Approach to Individualized Medicine (GANI_MED)’ funded by the Federal Ministry of Education and Research (grant 03IS2061A). DNA methylation data have been supported by the DZHK (grant 81X3400104). The University of Greifswald is a member of

the Caché Campus program of the InterSystems GmbH. The SHIP authors are grateful to Paul S. DeVries for his support with the EWAS pipeline.

WMH measurements in each cohort

Study	MRI technology	Scanning Protocol	Image Processing	WMH quantification
ADNI ²⁹⁶⁻²⁹⁹	1.5T (ADNI-1) and 3T (ADNI-2) MRI scanners from General Electric Medical Systems, Philips Medical Systems, or Siemens Medical Solutions	A series of sagittal T1-weighted scans and axial proton-density/ T2- weighted fast spin echo scans. T1-, T2- and PD-weighted MRI scans were co-registered, and skull stripped, then corrected for bias field.	ADNI-1: Brain images were reviewed by a neuroradiologist to exclude infarcts and other abnormalities and graded by a trained individual. ADNI-2: the FreeSurfer software package version 4.3 and 5.1 (http://surfer.nmr.mgh.harvard.edu/) by the Schuff and Tosun laboratory at the University of California-San Francisco	ADNI-1: WMH was quantified based on T1-, T2- and PD-intensities, the prior probability of WMH, and the probability of WMH conditioned on the presence of WMH at neighboring voxels. ADNI-2: WMH was measured with a Bayesian approach using segmentation of high resolution 3D T1 and FLAIR sequences.
ARIC ³⁰⁰	1.5-Tesla scanners (General Electric Medical Systems or Picker Medical Systems)	A series of sagittal T1-weighted scans and axial proton-density, T2- weighted and T1-weighted scans with 5 mm thickness and no interslice gaps.	Images were interpreted directly from a PDS-4 digital workstation consisting of four 1024 X 1024-pixel monitors capable of displaying all 96 images simultaneously.	WMHs were estimated as the relative total volume of periventricular and subcortical WM signal abnormality on proton density– weighted axial images by visual comparison with eight templates that successively increased from barely detectable WM changes (Grade 1) to extensive, confluent changes (Grade 8). Individuals with no WM changes received Grade 0, and those with changes worse than Grade 8 received Grade 9.
CARDIA ³⁰¹	3T scanners (Siemens Medical Solutions and Philips Medical Systems)	A sequence of sagittal 3DT2 with slice thickness 1 mm; sagittal 3D FLAIR with slice thickness 1 mm; and sagittal 3D MPRAGE with slice thickness 1mm.	Structural MR images were processed using previously described methods that were based on an automated multispectral computer algorithm.	We classified all supratentorial brain tissue into GM, WM, and CSF. GM and WM were characterized as normal and abnormal (ischemic).
CHS ³⁰⁰	General Electric or Picker 1.5-Tesla scanners and Toshiba 0.35-Tesla scanner	A series of sagittal T1-weighted scans and axial PD, T2- weighted and T1-weighted scans with 5 mm thickness and no interslice gaps.	Images were interpreted directly from a PDS-4 digital workstation consisting of four 1024 X 1024-pixel monitors capable of displaying all 96 images simultaneously.	WMHs were estimated as the relative total volume of periventricular and subcortical WM signal abnormality on proton density– weighted axial images by visual comparison with eight templates that successively increased from barely detectable WM changes (Grade 1) to extensive, confluent changes (Grade 8). Individuals with no WM changes received Grade 0, and those with changes worse than Grade 8 received Grade 9.
FHS	1 or 1.5 T Siemens Magnetom scanner	Offspring cohort: 3D T1 and double echo proton density (PD) and T2 DSE coronal images were acquired in 4-mm contiguous slices 3rd Generation cohort: FLAIR	Analyzed with QUANTA 6.2 at the University of California–Davis Medical Center. Images were analyzed and interpreted blind to subject data and in random order. Semi-automated analysis of pixel distributions, based on	For segmentation of WMH from other brain tissues the first and second echo images from T2 sequences were summed and a lognormal distribution was fitted to the summed data (after removal of CSF and correction of image intensity non-uniformities). A segmentation threshold for WMH was

			mathematical modeling of MRI pixel intensity histograms for CSF, WM, and GM, were used to determine the optimal threshold of pixel intensity to best distinguish CSF from brain matter based on previously published methods. The intracranial vault above the tentorium was outlined manually to determine the TIV.	determined as 3.5 standard deviations in pixel intensity above the mean of the fitted distribution of brain parenchyma.
GENOA ³⁰²	1.5 T MRI scanner from General Electric Medical Systems	FLAIR images consisting of 48 contiguous 3-mm interleaved slices with no interslice gap. Total intracranial volume (head size) was measured from T1-weighted spin echo sagittal images, each set consisting of 32 contiguous 5 mm thick slices with no interslice gap.	A fully automated algorithm was used to segment each slice of the edited multi-slice FLAIR sequence into voxels assigned to one of three categories: brain, CSF, or leukoaraiosis.	Interactive imaging processing steps were performed by a research associate who had no knowledge of the subjects' personal or medical histories or biological relationships. Total leukoaraiosis volumes were determined from axial FLAIR images.
LBC1936 ²⁸⁸	1.5 T HDx MRI scanner from General Electric Medical Systems	T1-w coronal and T2-W, FLAIR, and T2*-weighted axial whole brain images	A semi-automatic computational program written specifically for the project, MCMxxxVI, a multispectral color fusion method that combines different pairs of sequences in red-green color space and performs minimum variance quantization to highlight different tissues.	WMH were measured in the cerebral hemispheres, cerebellum and brainstem. Intracranial volume, brain and WMH volume were extracted and manually corrected as necessary to remove false positive lesions in the insular cortex, cingulate gyrus, anterior temporal cortex and around the floor of the third ventricle, and correct false negatives (http://www.bric.ed.ac.uk/research/imageanalysis.html). All focal stroke lesions were manually removed.
RSIII and RSI-III (BBMRI) ^{293,303}	1.5 T scanner (GE Signa Excite) using an 8-channel head coil	Structural imaging is performed with T1-weighted (T1w), proton density-weighted (PDw) and fluid-attenuated inversion recovery (FLAIR) sequences.	The combination of different MR contrasts provided by acquired sequences can be used for automated brain tissue and white matter lesion segmentation. For this purpose, the T1w scan is acquired in 3D at high in-plane resolution and with thin slices (voxel size <1 mm ³).	WMH volume was quantified using two fully automated methods, which was described previously in more detail. Either the HASTE, PD and T2 sequences were used, or the FLAIR, T1 and PD sequences. Briefly, cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM) are segmented by an atlas-based k-nearest neighbor classifier on multi-modal magnetic resonance imaging data. This classifier is trained by registering brain atlases to the subject. The resulting GM segmentation is used to automatically find a WMH threshold in a fluid-attenuated inversion recovery scan.
SHIP-TREND ³⁰⁴	1.5-T MR imager from Siemens Medical systems	A series of T1, MP-RAGE/ axial plane with the resolution of 1.0 x 1.0 x 1.0mm ³ and T2 FLAIR / axial plane with the resolution of 0.9 x 0.9 x 3.0mm ³ .	The FreeSurfer 5.1.0 software	WMH quantification was based on the following sequences: T1, MP-RAGE/ axial plane, TR=1900 ms, TE=3.4 ms, Flip angle=15°, resolution of 1.0 x 1.0 x 1.0mm ³ and T2 FLAIR / axial plane, TR= 5000, TE= 325, voxel= 0.9 x 0.9 x 3.0.
Abbreviations: Double spin echo (DSE). Fluid-attenuated inversion recovery (FLAIR). Proton-density (PD). Cerebrospinal fluid (CSF). White matter (WM). Grey matter (GM). Total (intra)cranial volume (TIV/TCV)				

DNA methylation data collection and association study in each cohort

	Technology	Tissue source	Normalization	QC methods	Experimental/ Biological covariates	Covariate Assessment	Technical Covariates	Methods of Adjustment for Covariates
ADNI	Illumina 850K	Whole blood	BMIQ	removed probes with - a low detection rate ($P > 0.01$ for >5% of samples) removed samples with - low call rate ($P < 0.01$ for <95% of probes)	Model 1: ICV, age, sex, WBC, study site, visit, time between MRI and DNAm visits, early MCI status, population structure (surrogates); Model 2: + Smoking, BMI SBP, DBP	- Smoking status was coded as “Yes” or “No”. - Systolic and diastolic blood pressure was measured at the MRI visit.	chip ID, chip Position	Linear mixed models of DNAm~WMH burden adjusted for biological and technical covariates. Technical covariates as random effects.
ARIC ³⁰⁵	Illumina 450K	Blood leukocytes	BMIQ	removed probes with - a low detection rate ($P > 0.01$ for >5% of samples) removed samples with - low call rate ($P < 0.01$ for <99% of probes)	In each ancestry, Model 1: age, sex, WBC, technical covariates, population structure (surrogates); Model 2: + Smoking, BMI, SBP, DBP	- Smoking was obtained at either exam 2 or 3 using an interviewer- administered questionnaire and was classified as current, former, or never. - Blood pressure was assessed at exam 2 and 3, the average value of two visit was used in this study.	Chip ID, Chip Row, plate	Linear mixed models of DNAm~WMH burden adjusted for biological and technical covariates. Technical covariates as random effects.
CARDIA	Illumina 850K	Whole blood	RCP	- excluded 6,209 CpGs with a detection rate <95% - excluded 87 samples with the percentage of low-quality methylation measurements >5% - extremely low intensity of bisulfite conversion probes (less than mean intensity – 3 × standard deviation of the intensity across samples).	Model 1: age, sex, race, WBC, technical covariates, population structure (surrogates); Model 2: + Smoking, BMI, SBP, DBP	- Smoking status was coded as “never”, “ex-smoker”, and “current smoker” based on a self- administered questionnaire. - Systolic and diastolic blood pressure in mmHg measured at the same examination as the DNA sampling.	chip ID, chip position	Linear mixed models of DNAm~WMH burden adjusted for biological and technical covariates. Technical covariates as random effects.
CHS	Illumina 450K	Blood leukocytes	SWAN	excluded samples with - a proportion of probes falling detection of greater than 0.5% - low median intensities of below 10.5 (log2) across the methylated and unmethylated channels - QC probes falling greater than 3 standard deviation from the	Model 1: age, sex, 4 genetic PCs, batch, WBC estimates, race, technical covariates Model 2: Model 1 + Smoking, BMI, SBP, DBP	- Smoking was assessed with a standard questionnaire during in- person interview at the time of blood assessment for methylation. Smoking status was categorized into never, former or current smoker. - Systolic and diastolic blood	chip ID, chip position	Linear mixed models of DNAm~WMH burden adjusted for biological and technical covariates. Technical covariates as random effects.

				mean - with sex-check mismatches - failed concordance with prior genotyping or > 0.5% of probes with a detection p-value > 0.01		pressure in mmHg measured at the same examination as the DNA sampling.		
FHS	Illumina 450K	Whole blood	DASEN (waterRmel on)	Samples were excluded if - outliers in multidimensional scaling (MDS) analysis, high missing rate (>1%), poor matching to SNP genotype Probes were excluded if - high missing rate (>20%), mapped to multiple locations, had SNP (MAF>5% in EUR 1000G) at CpG site or ≤10 bp of Single Base Extension.	Model 1: age, sex, WBC, technical covariates, population structure (surrogates); Model 2: + Smoking, BMI	- Current smoking is defined as regularly smoking in the past year which was self-reported at each examination. Using examination 8 and previous examination data, participants were classified as current smoker, ex-smoker, never smoker.	chip ID, chip position	Linear mixed models of DNAm~WMH burden adjusted for biological and technical covariates. Family structure is also adjusted.
GENOA	Illumina 27K	Blood leukocytes	SWAN	- Remove probe if >5% of samples have a detection p-value of >0.01; - Remove sample if >5% of probes have a detection p-value of >0.01;	Model 1: age, sex, total intracranial volume, 4 genetic PCs, WBC, technical covariates, familial relationship; Model 2: + smoking, BMI, SBP, DBP	- Smoking status: current smoker (smokes currently or quit smoking within the past year), ex-smoker (formerly smoked, but quit more than 1 year ago), and never-smoker (smoked less than 100 cigarettes in his/her entire life). - Blood Pressure: sitting systolic and diastolic blood pressure in mmHg were measured three times with a random zero sphygmomanometer. The average of the last two measurement was used.	plate, row, column	Linear mixed models of DNAm~WMH burden adjusted for biological and technical covariates. Technical covariates and familial relationship were treated as random effects.
LBC1936 ³⁰⁶	Illumina 450K	Whole blood	minfi	Remove probes with - a low detection rate (P>0.01 for >5% of samples) - low quality (manual inspection) Remove samples with - low call rate (P<0.01 for <95% of probes) - a poor match between genotypes and SNP control probes - incorrect predicted sex.	Model 1: age + Sex + ICV+ 4 genetic PCs + WBC + technical covariates, Model 2: BMI + current smoking + dbp3sit + sbp3sit + age + Sex + ICV+ 4 genetic PCs + WBC + technical covariates	- Smoking status was self-reported in three categories: current smoker, ex-smoker, never smoker. - Blood Pressure: dbp3sit + sbp3sit (3rd reading)	Chip, position, sample plate, hybridization date	Linear mixed models of DNAm~WMH burden adjusted for biological and technical covariates.

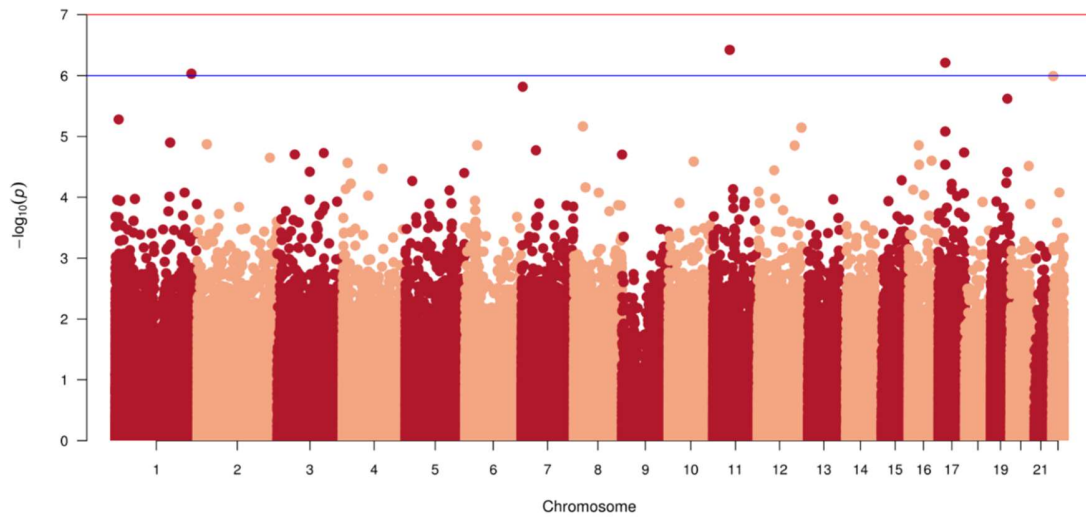
RSIII and RSI-III (BBMRI)	Illumina 450K	Whole blood	DASEN (watermelon)	<ul style="list-style-type: none"> - Excluded samples with incomplete bisulfite treatment, a low detection rate (<99%), or gender mismatches. - Excluded probes with a detection p-value>0.01 in >1% samples 	Model 1: ICV, age, sex, array number, array position, WBC counts. Model 2: model 1 + smoking, BMI, SBP, DBP	<ul style="list-style-type: none"> - Smoking status: Smoking status was coded as self-reported “never smoker”, “ever smoker”, and “current smoker” based on a home interview. - Blood Pressure: Blood pressure was measured twice at the center visit, and the average of the two measurements was used. 	Array number, array position.	Linear mixed models of DNAm~WMH burden adjusted for biological and technical covariates. Technical covariates as random effects.
SHIP- TREND	Illumina 850K	Whole blood	CPACOR	<ul style="list-style-type: none"> - probes with detection p-value$\geq 1E-16$ were set to missing - arrays with call rates ≤ 0.95 were excluded - arrays with observed technical problems during steps like bisulfite conversion, hybridization or extension, as well as arrays with mismatch between sex of the proband and sex determined by the chr X and Y probe intensities were excluded. 	Model 1: ICV, age, sex, WBC subtypes, 6 genetic PCs, technical covariates; Model 2: + smoking, BMI, SBP, DBP	<ul style="list-style-type: none"> - Smoking status: current smoker determined by questionnaire - Blood Pressure: measured (mean of 2nd and 3rd measurement) 	control probe	linear regression adjusted for all covariates listed

Gene expression data description

Study	Technology	Tissue source	Normalization	GEO accession ID
FHS ³⁰⁷	Affymetrix Human Exon Array ST 1.0	Whole blood	The raw data were quantile-normalized and log2 transformed, followed by summarization using Robust Multi-array Average. The percentages of each imputed cell type were then normalized, where the negative predicted values were set to 0 and the sum of the percentages for all cell types were set 100%.	
RS ^{308,309}	Illumina HumanHT12v4 Expression Beadchip	Whole blood	Quantile-normalized to the median distribution and subsequently log2-transformed. The probe and sample means were centered to zero and Z-transformed.	GSE33828

Appendix B. Supplemental Figures

Reduced Adjustment Model



Full Adjustment Model

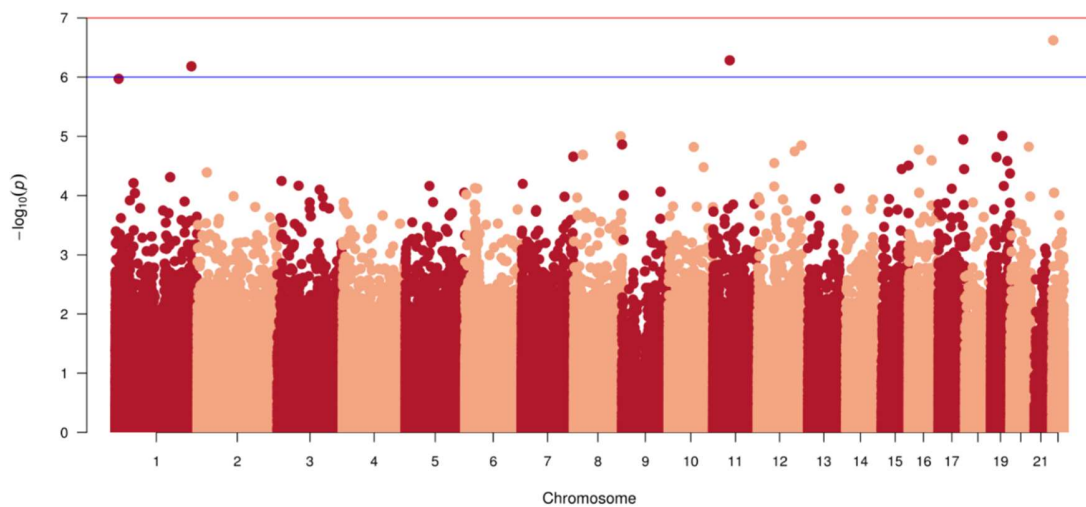


Figure S 1 Manhattan Plots of genome-wide WMH burden associations at each DNA methylation (DNAm) site in the trans-ethnic meta-analyses. The DNAm locations are plotted against the $-\log_{10}(p)$ value. The red line indicates the Bonferroni threshold (1×10^{-7}) for epigenome-wide significance. The blue line marks the putative threshold (1×10^{-6}).

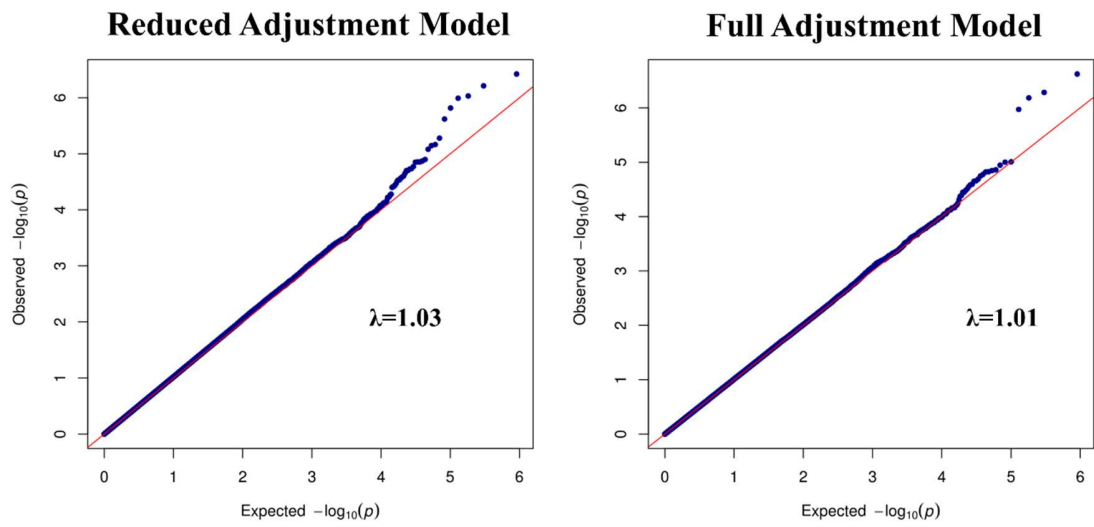


Figure S 2 QQ plots for observed vs expected $-\log_{10}$ (P value) at each DNAm site in the trans-ethnic meta-analyses.

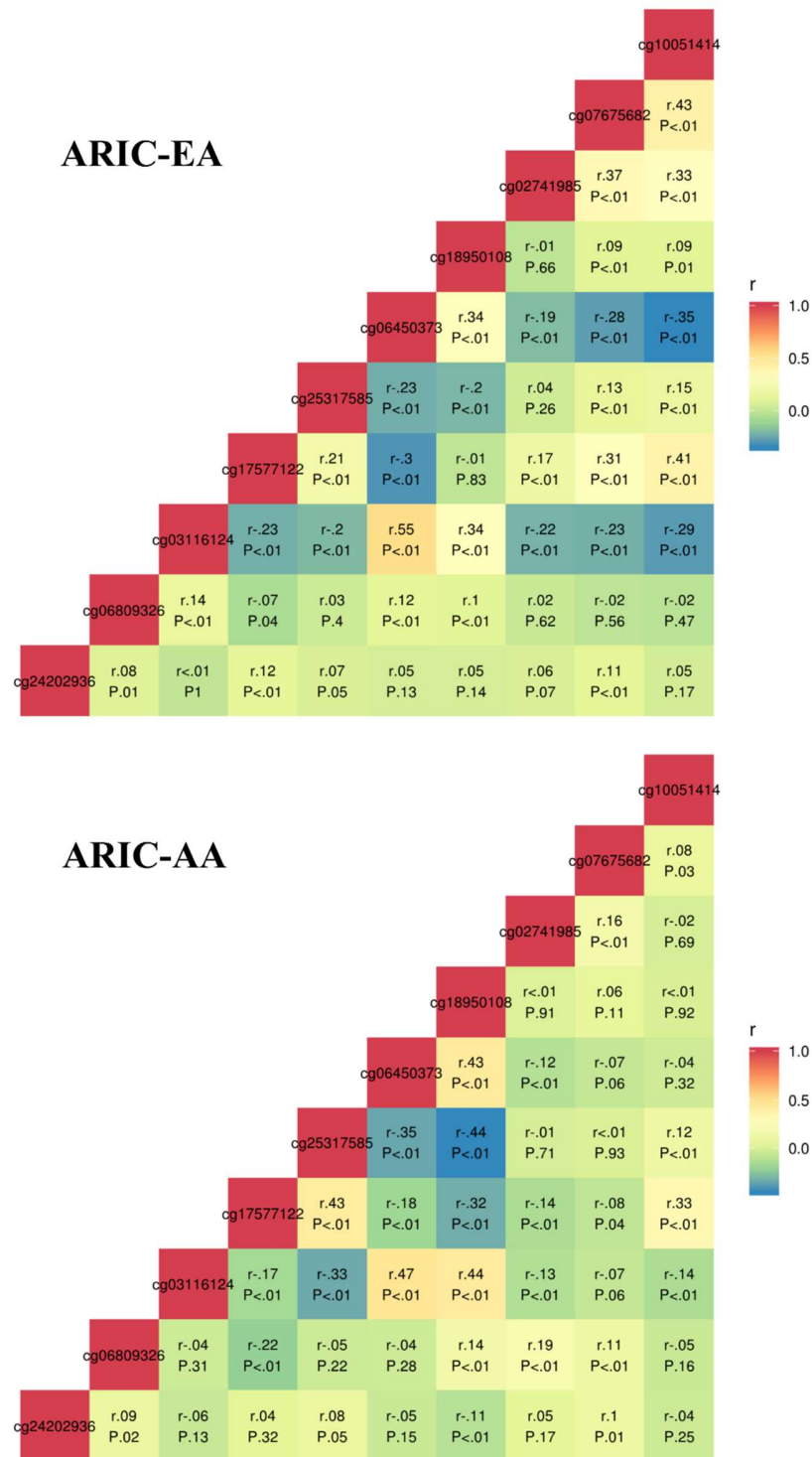
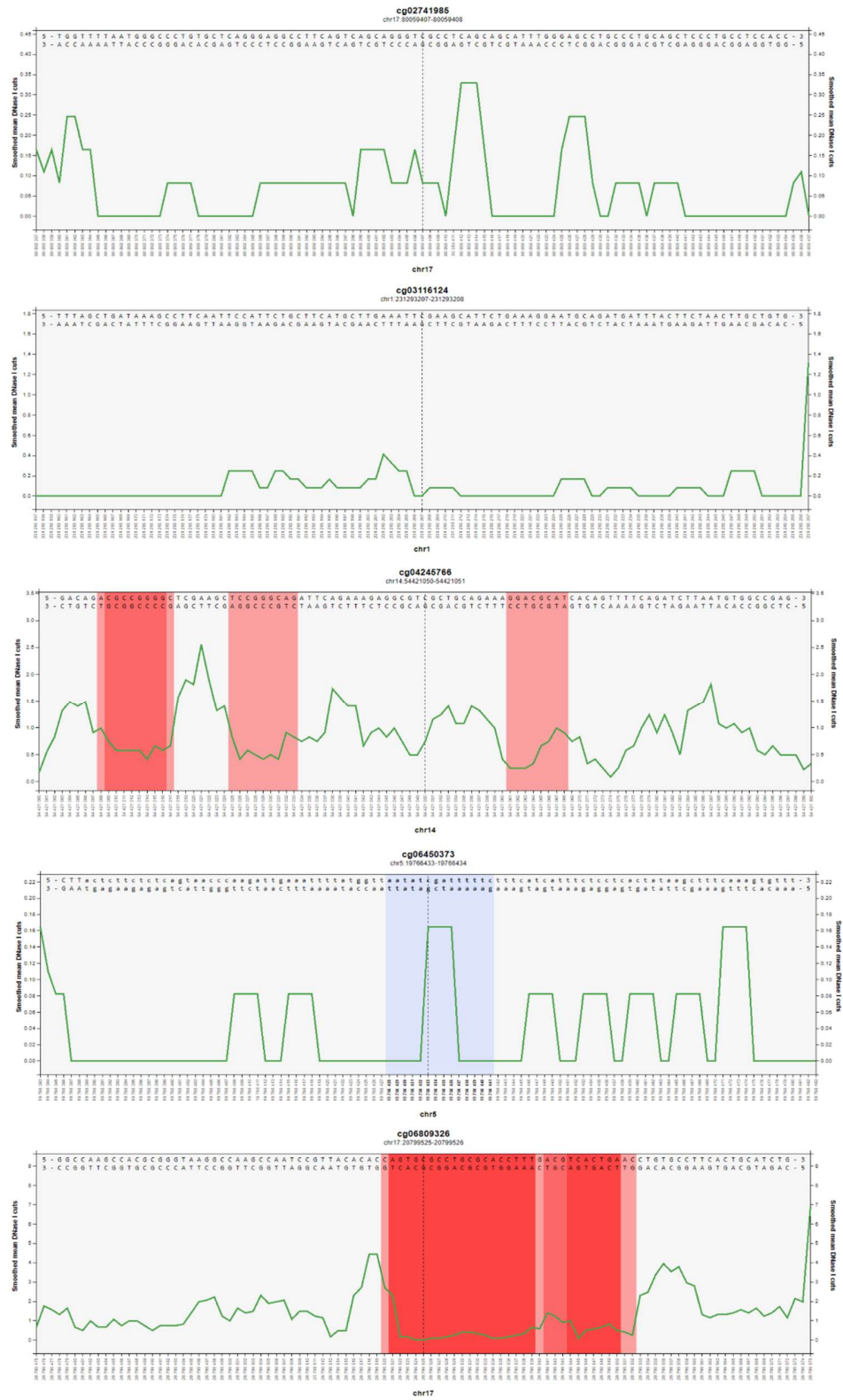


Figure S 3 Spearman correlation among DNA methylation probes associated with WMH burden in ARIC subjects of European ancestry (ARIC-EA) and of African ancestry (ARIC-AA). The color scale corresponds to strength of correlation, where inverse correlations are blue and direct correlations are red.



To be continued

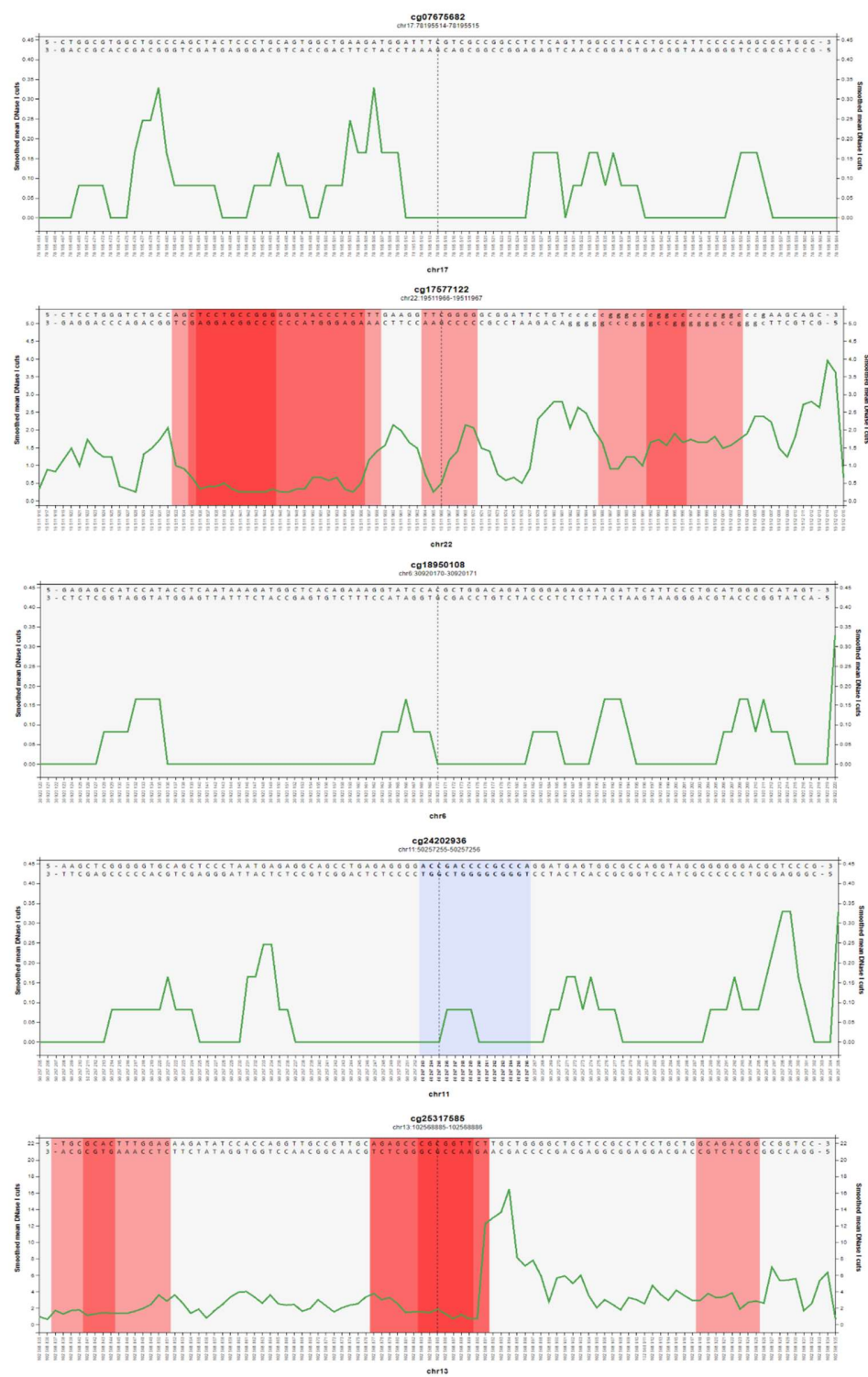


Figure S 4 DNase I hypersensitivity signal around EWAS probes

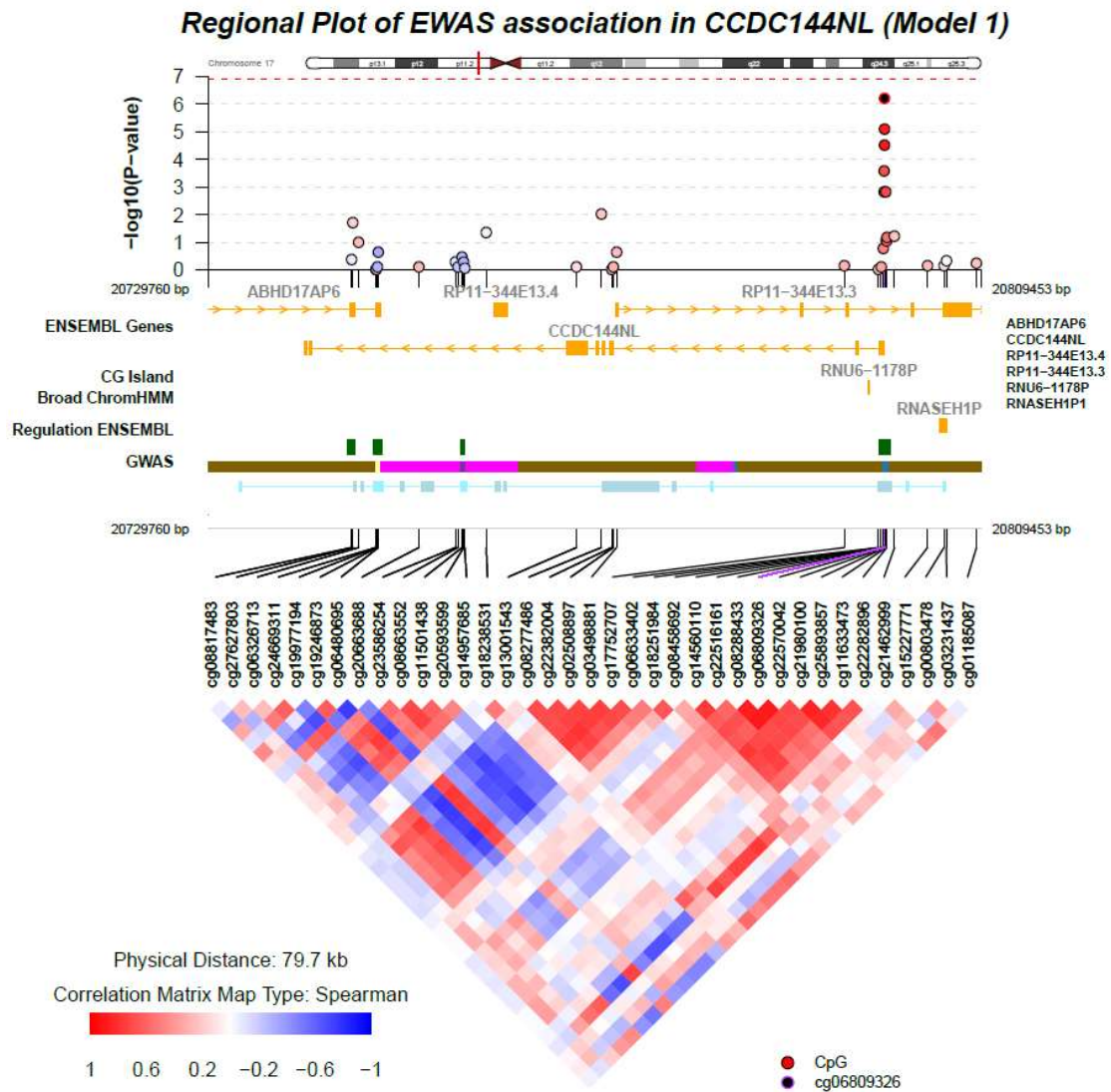
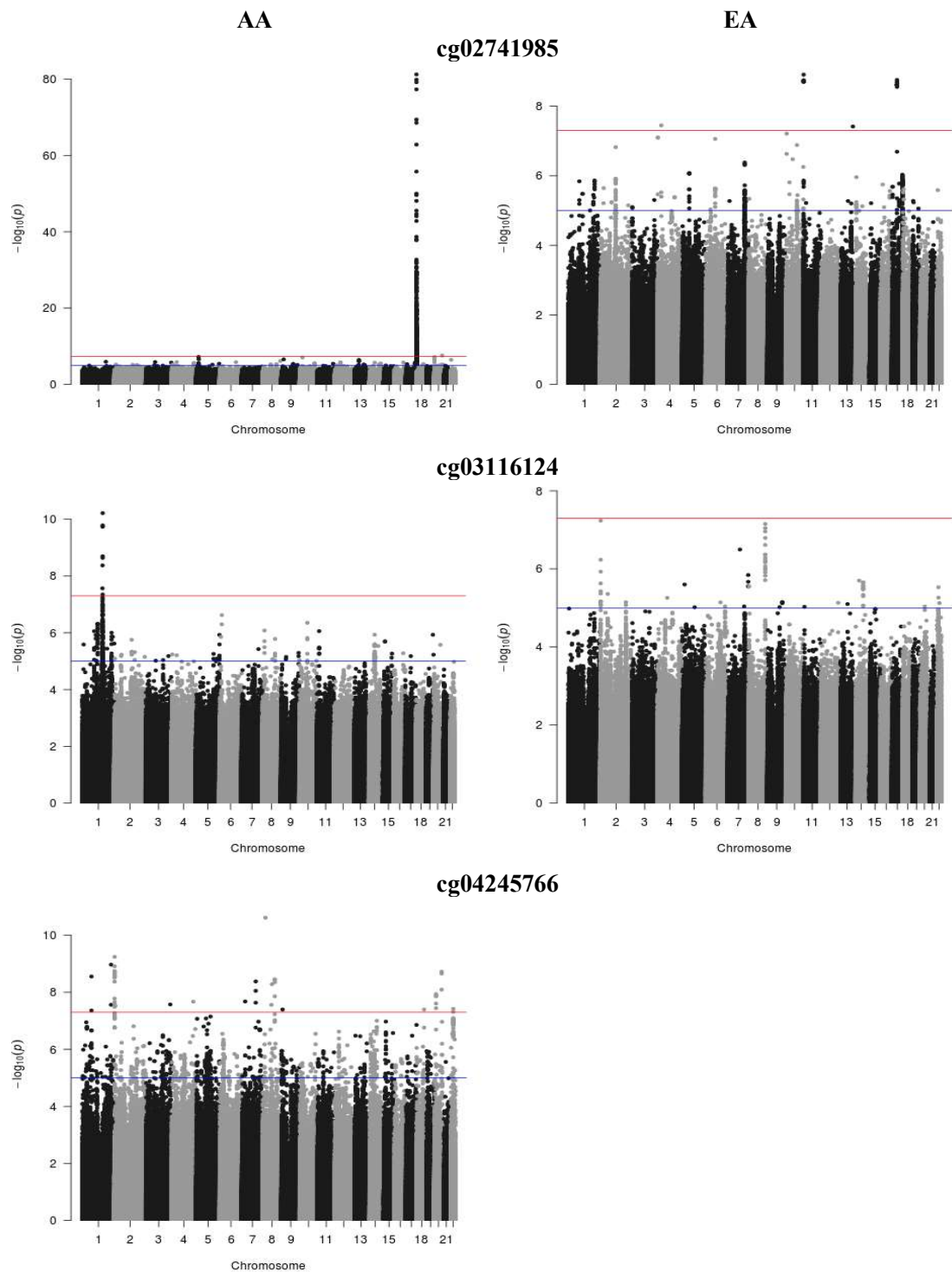


Figure S 5 Regional Plot of cg06809326

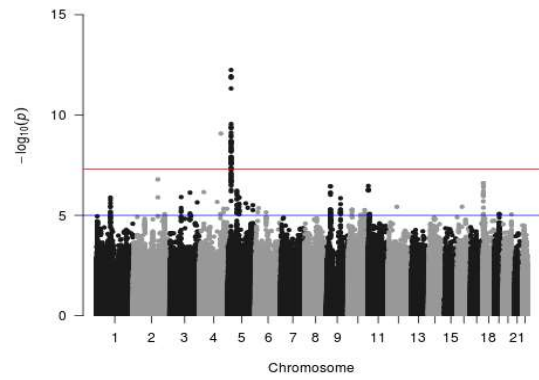
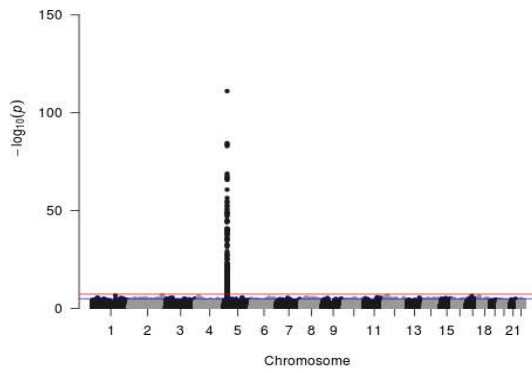


To be continued

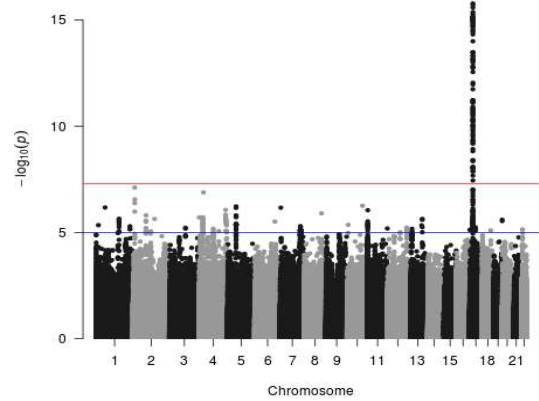
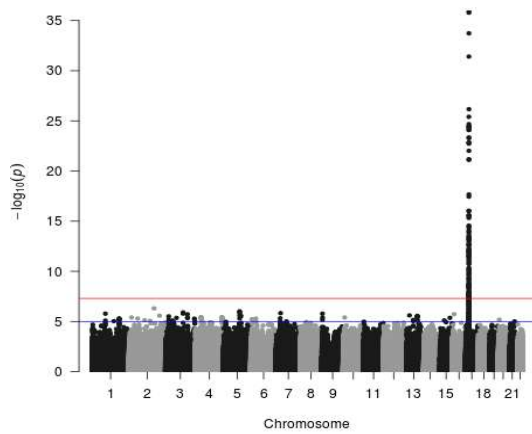
AA

EA

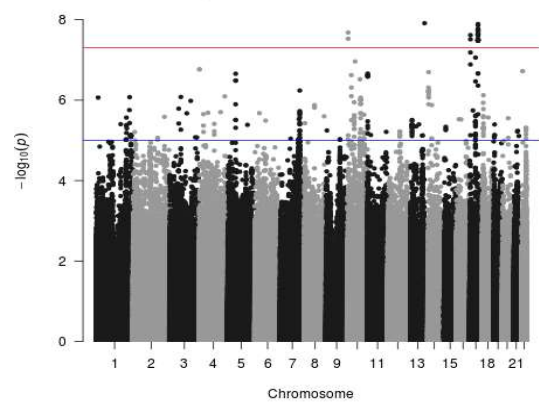
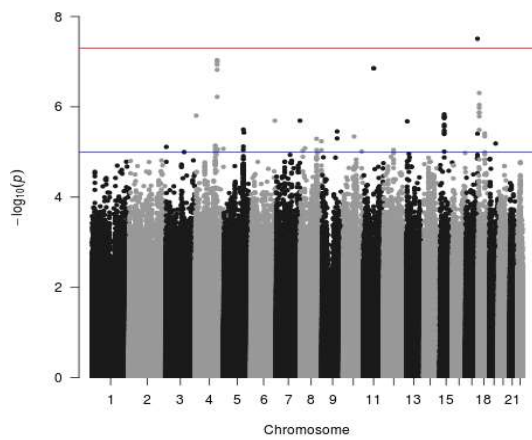
cg06450373



cg06809326



cg07675682

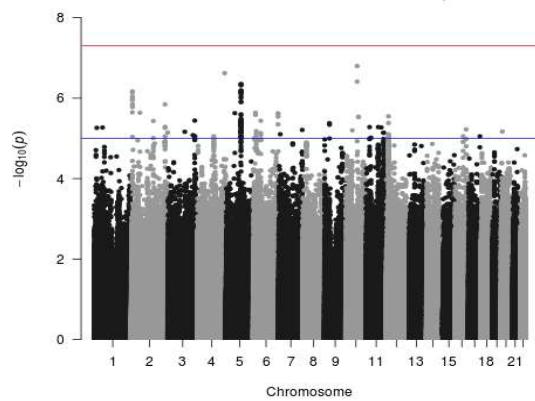
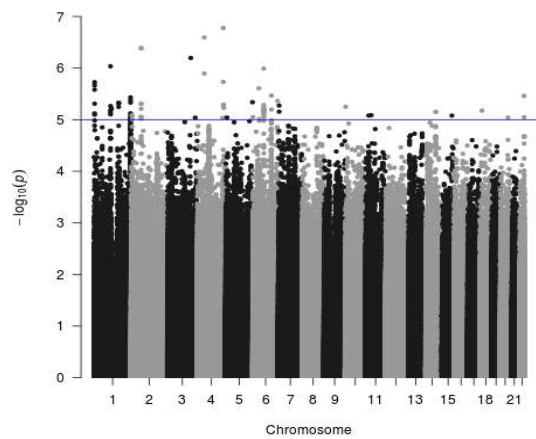


To be continued

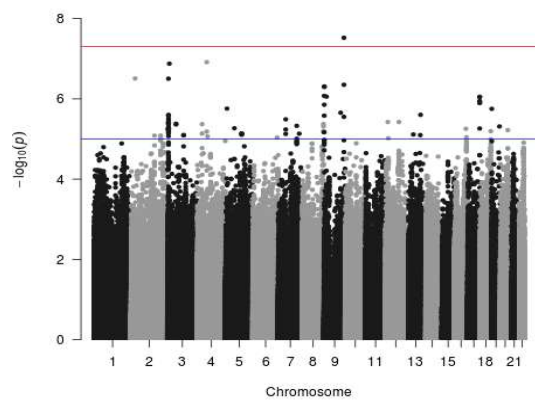
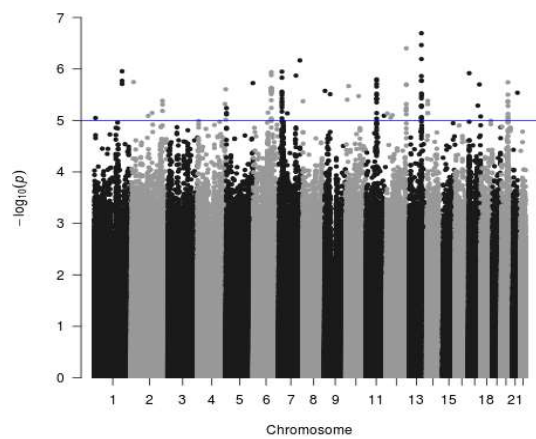
AA

EA

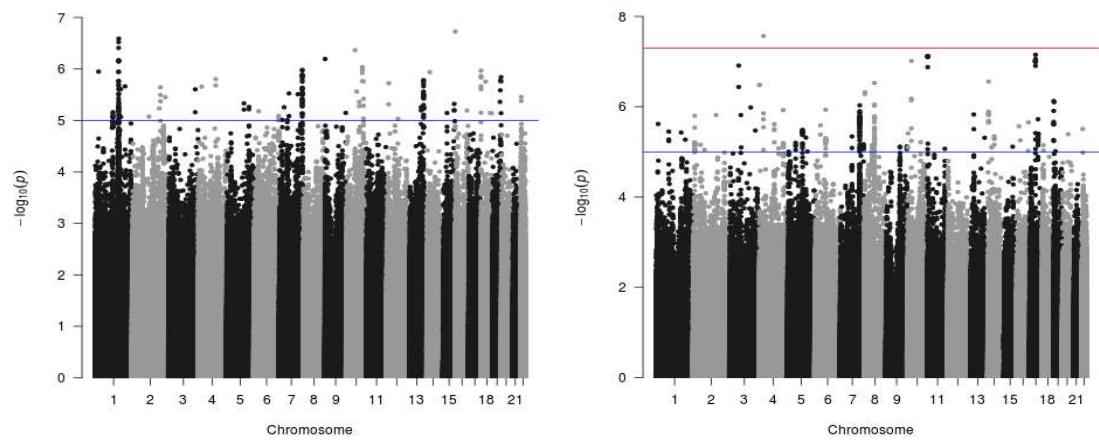
cg10051414



cg17577122



cg18950108

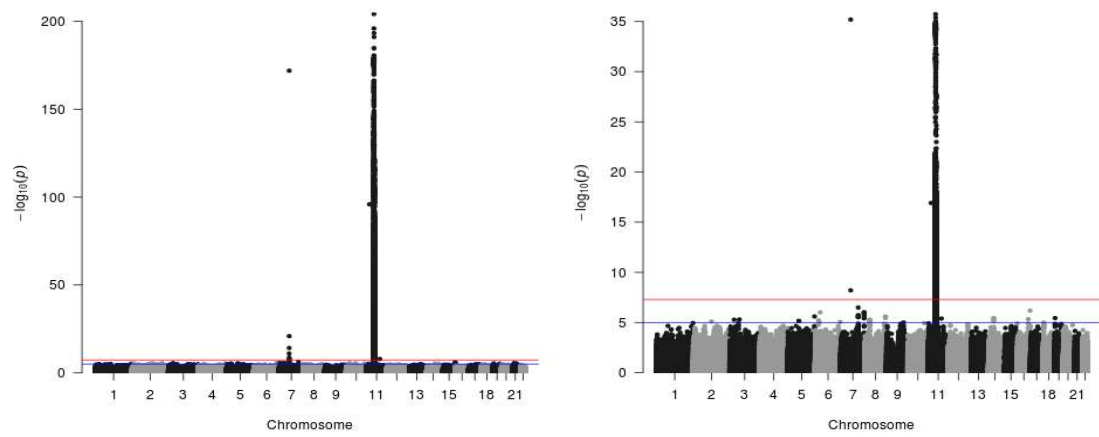


To be continued

AA

EA

cg24202936



cg25317585

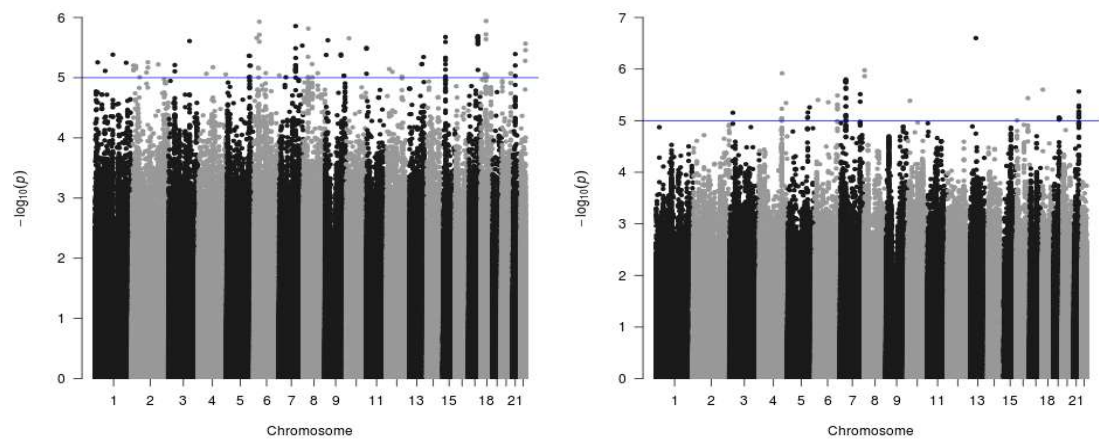


Figure S 6 Manhattan plots of mQTLs for 11 target probes

Appendix C. Supplemental Tables

Table S 1 Descriptive statistics of participating cohorts

Study		#Markers	Ancestry	No.	GI (λ)	Female	Age	BMI	HT	Smoking	WMH burden
ADNI		864,460	EA	387	0.98/ 0.99	187 (48.3%)	74.1±7.0	27.2±4.7	274 (70.8%)	158 (40.8%)	1.29 [0.58,2.01]
ARIC		483,526	EA	906	1.00/ 1.00	519 (57.3%)	63.1±4.4	26.4±4.6	306 (33.8%)	162 (17.9%)	0.69 [0.69,1.10]
			AA	639	0.97/ 0.99	413 (64.8%)	61.8±4.5	29.6±5.4	412 (64.5%)	127 (19.9%)	0.69 [0.69,1.10]
BBMRI		419,938	EA	442/ 404	1.04/ 1.06	255 (57.7%)	64.7±5.8	27.4±3.9	279 (63.1%)/ 269 (66.6%)	63 (14.2%)	1.40 [1.03,1.92]
CARDIA		860,628	Pooled	194	0.71/ 0.73	88 (45.4%)	50.2±3.5	28.8±5.3	42 (21.6%)	27 (13.9%)	0.18 [0.10,0.41]
CHS		483,624	Pooled	378	1.33/ 1.30	153 (40.5%)	74.6±5.4	27.5±5.0	240 (63.5%)	49 (13.0%)	1.10 [0.69- 1.38]
FHS Offspring	FLAIR	443,314	EA	1,323/ 1,320	0.96/ 0.99	711 (53.7%)	66.9±8.7	28.0±5.0	764 (57.7%)/ 761 (57.7%)	101 (7.6%)	1.18 [0.79,1.69]
	DSE			403/ 399	0.98/ 1.01	230 (57.1%)	60.6±9.6	28.1±5.7	186 (46.2%)/ 185 (46.4%)	61 (15.1%)	0.46 [0.28,0.79]
GENOA		26,449	AA	356/ 355	0.85/ 0.83	259 (72.8%)	66.3±7.19	31.0±5.98	291 (81.7%)/ 290 (85.4%)	50 (14.0%)	2.10 [1.81, 2.56]
LBC1936		450,727	EA	230	1.23/ 1.23	107 (46.5%)	72.1±5.2	27.7±4.51	190 (82.6%)	13 (5.6%)	2.04 [1.33, 2.76]
RS3		474,529	EA	547/ 163	1.31/ 1.19	291 (53.2%)	62.6±7.3	27.2±4.1	229 (41.9%)/ 95 (58.3%)	113 (20.7%)	1.34 [0.97,1.77]
SHIP-TREND		865,860	EA	214	1.23/ 1.20	117 (54.7%)	49.7±13.3	27.2±4.1	84 (39.3%)	42 (19.6%)	0.74 [0.59,0.93]
#Markers: number of markers. No.: Number of subjects in model1 (reduced) /model2 (full). GI: genomic inflation factor in model1 (reduced) /model2 (full). Female: number of female subjects, %. Age: mean± standard deviation (s.d.) in years. BMI: mean and s.d. of body mass index, kg/m ² . HT: hypertensive subjects, number (%) in model1 (reduced) /model2 (full). Smoking: current smoker, number (%). WMH burden: median and the IQR (interquartile range; Q1-Q3) of log(WMH+1).											

Table S 2 Coefficient point estimates and P values for 11 target probes in the meta-analyses among hypertensive and normotensive individuals

	HT		NT	
	Z score (S.E.)	N	Z score (S.E.)	N
<i>Reduced model</i>				
cg03116124	-2.83(0.005)	2,816	-3.009(0.003)	2,599
cg24202936	3.18(0.001)	3,006	2.866(0.004)	2,599
cg17577122	3.46(0.0005)	3,006	3.319(0.0009)	2,599
cg06809326	3.34(0.0009)	3,006	3.289(0.001)	2,599
cg04245766	1.09(0.276)	798	-2.21(0.027)	587
cg06450373	0.61(0.543)	2,727	-5.102(3.36E-07)	2,437
cg18950108	-1.66(0.097)	3,006	-5.059(4.22E-07)	2,599
cg07675682	-0.49(0.627)	3,296	-3.664(0.0002)	2,664
cg10051414	-0.27(0.789)	2,606	3.648(0.0003)	2,204
cg25317585	5.20(2.05E-07)	3,006	1.061(0.289)	2,599
cg02741985	2.09(0.037)	3,006	0.019(0.985)	2,599
<i>Full model</i>				
cg03116124	-2.71(0.007)	2,668	-3.141(0.002)	2,321
cg24202936	2.96(0.003)	2,858	1.971(0.049)	2,321
cg17577122	3.64(0.0003)	2,858	3.967(7.28E-05)	2,321
cg06809326	3.31(0.0009)	2,858	2.615(0.009)	2,321
cg04245766	0.82(0.410)	798	-2.319(0.020)	587
cg06450373	0.53(0.598)	2,589	-4.623(3.78E-06)	2,186
cg18950108	-2.11(0.035)	2,858	-4.421(9.81E-06)	2,321
cg07675682	-0.49(0.624)	3,147	-4.095(4.22E-05)	2,386
cg10051414	-0.56(0.574)	2,458	3.853(0.0001)	1,926
cg25317585	5.14(2.76E-07)	2,858	0.185(0.853)	2,321
cg02741985	2.13(0.034)	2,858	-0.186(0.852)	2,321

Table S 3 Top DMRs identified from race-specific analysis

Chr	start	end	P _{min}	Size	P _{Comb-p}	P _{Sidak}	Gene
EA							
Model 1							
19	50191341	50191882	2.68E-08	6	9.46E-14	8.29E-11	<i>PRMT1</i>
17	20799407	20799769	7.60E-08	10	1.68E-12	2.19E-09	<i>CCDC144NL-AS1</i>
12	122018896	122019117	4.47E-05	7	3.12E-09	6.68E-06	<i>KDM2B</i>
3							<i>ATP13A4</i>
	193272560	193272869	2.11E-03	4	2.95E-07	4.53E-04	<i>;ATP13A4-AS1</i>
5	140515674	140515961	0.006	3	3.99E-07	6.59E-04	<i>PCDHB5</i>
16	89894097	89894329	0.003	3	5.54E-07	0.001	<i>SPIRE2</i>
10	105428384	105428818	0.008	5	1.94E-06	0.002	<i>SH3PXD2A</i>
13	46757317	46757415	0.008	3	2.14E-06	0.010	<i>LCP1</i>
8	144451642	144451924	0.013	3	6.14E-06	0.010	<i>RHPNI</i>
11	130185383	130185670	0.019	3	6.34E-06	0.010	<i>ZBTB44</i>
1	209921095	209921370	1.26E-02	4	7.48E-06	1.28E-02	<i>LOC101930114</i>
9	116861287	116861532	3.15E-02	4	1.20E-05	0.023	<i>KIF12</i>
17	78800573	78800806	0.032	4	1.25E-05	0.025	<i>RPTOR</i>
Model 2							
6	30039129	30039600	3.03E-07	21	1.39E-11	1.40E-08	<i>RNF39</i>
19	50191341	50191882	1.44E-06	6	5.42E-11	4.75E-08	<i>PRMT1</i>
2	3704500	3704793	3.25E-06	6	1.65E-10	2.66E-07	<i>ALLC</i>
12	122018896	122019117	4.10E-05	7	6.45E-09	1.38E-05	<i>KDM2B</i>
6	28973317	28973497	4.23E-05	9	1.09E-07	0.0003	<i>ZNF311</i>
18	23713728	23714084	0.001	6	7.31E-07	0.001	<i>PSMA8</i>
10	105428384	105428818	3.61E-03	5	1.04E-06	0.001	<i>SH3PXD2A</i>
16	89894097	89894329	6.69E-03	3	2.08E-06	0.004	<i>SPIRE2</i>
1	209921030	209921370	6.69E-03	5	2.16E-06	3.01E-03	<i>LOC101930114</i>
13	46757317	46757415	0.010	3	3.22E-06	0.015	<i>LCP1</i>
							<i>PCDHA6;PCDHA12;</i>
							<i>PCDHA9;PCDHA1;</i>
							<i>PCDHA2;PCDHA3;</i>
							<i>PCDHA4;PCDHA5;</i>
							<i>PCDHA10;PCDHA7;</i>
							<i>PCDHA8;PCDHA11;</i>
5	140305946	140306458	0.014	10	7.57E-06	6.98E-03	<i>PCDHAC1;PCDHA13</i>
							<i>ATP13A4;ATP13A4-</i>
3	193272560	193272869	1.28E-02	4	9.52E-06	1.45E-02	<i>AS1</i>
14	88621423	88621721	2.66E-02	5	1.42E-05	2.24E-02	<i>KCNK10</i>
15	101093777	101093900	2.95E-02	3	1.67E-05	6.24E-02	<i>PRKXPI</i>
8	144451642	144451924	1.01E-02	3	2.00E-05	3.31E-02	<i>RHPNI</i>
AA							
Model 1							

8	1765065	1765820	1.71E-07	13	4.11E-13	2.57E-10	<i>MIR596</i>
6	32847197	32847845	1.71E-07	24	7.24E-11	5.27E-08	<i>LOC100294145</i>
3	49170495	49171185	1.39E-06	9	9.19E-11	6.29E-08	<i>LAMB2</i>
3	126911726	126911953	6.74E-05	4	7.56E-09	1.57E-05	<i>C3orf56</i>
11	58869662	58870494	0.0002	6	3.49E-08	1.98E-05	<i>FAM111B</i>
16	89118837	89119709	5.96E-04	7	1.14E-07	6.19E-05	<i>ACSF3</i>
6	170732119	170732353	8.19E-04	3	1.54E-07	3.12E-04	<i>FAM120B</i>
15	102501639	102501654	2.25E-03	3	6.09E-07	1.90E-02	<i>WASH3P</i>
1	236557181	236557758	2.35E-03	5	6.75E-07	0.001	<i>EDARADD</i>
10	131669405	131669630	0.003	3	8.08E-07	0.002	<i>EBF3</i>
4	20985622	20985984	0.006	4	2.59E-06	0.003	<i>KCNIP4</i>
13	100217961	100218552	3.63E-03	4	5.79E-06	4.62E-03	<i>TM9SF2</i>
7	42533000	42533414	0.013	5	6.30E-06	0.007	<i>LINC01448</i>
21	15352607	15352983	1.80E-03	7	7.39E-06	0.009	<i>ANKRD20A11P</i>
19	49699495	49700427	0.029	2	9.05E-05	0.045	<i>TRPM4</i>

Model 2

6	32847209	32847845	3.50E-07	23	8.30E-12	6.16E-09	<i>LOC100294145</i>
3	49170495	49171185	4.44E-07	9	1.32E-11	9.01E-09	<i>LAMB2</i>
3	126911726	126911953	4.49E-06	4	3.23E-10	6.73E-07	<i>C3orf56</i>
10	29697904	29698685	9.44E-07	10	8.62E-10	5.21E-07	<i>SVIL-AS1</i>
8	1765295	1765477	7.26E-06	8	3.51E-09	9.11E-06	<i>MIR596</i>
13	100217961	100218552	1.68E-06	4	4.85E-09	3.87E-06	<i>TM9SF2</i>
6	170732119	170732353	1.75E-03	3	4.60E-07	9.28E-04	<i>FAM120B</i>
10	131669405	131669630	0.003	3	9.25E-07	0.002	<i>EBF3</i>
15	102501639	102501654	0.005	3	1.45E-06	0.045	<i>WASH3P</i>
1	236557181	236557758	1.02E-02	5	3.65E-06	2.99E-03	<i>EDARADD</i>
8	145008109	145008397	0.002	3	8.38E-06	0.01365	<i>PLEC</i>
19	49223833	49224454	0.002	5	1.15E-05	0.009	<i>RASIP1</i>
7	116786596	116786870	2.76E-02	3	1.57E-05	0.027	<i>ST7</i>
5	150678161	150678334	2.77E-02	2	1.67E-05	4.45E-02	<i>SLC36A3</i>
13	113633378	113634042	0.014	8	2.00E-05	0.014	<i>MCF2L</i>
2	114341630	114342108	1.23E-02	3	2.31E-05	2.25E-02	<i>WASH2P</i>
17	37893802	37894636	0.028	6	5.74E-05	0.032	<i>GRB7</i>

P_{min}: the minimum P value in the DMR. Size: Number of methylation probes in the DMR.

P_{Comb-p}: combined Z-score based on Comb_P. P_{Sidak}: Sidak multiple-testing corrected P value

Table S 4 Top DMRs from gene-based analysis in hypertensive subjects

Chr	start	end	P _{min}	Size	P _{Comb-p}	P _{Sidak}	Gene
Model 1							
8	99984349	99985049	6.11E-3	5	1.16E-06	7.85E-04	<i>OSR2</i> <i>ATP13A4;ATP13A4-</i>
3	193272560	193273057	1.23E-2	5	2.57E-06	2.45E-03	<i>AS1</i>
22	45680497	45681036	1.86E-2	12	6.30E-06	0.00552	<i>UPK3A</i>
2	210074019	210074359	3.67E-2	3	1.70E-05	2.34E-02	<i>MAP2</i>
12	125225315	125225808	0.048	3	2.74E-05	0.02598	<i>SCARB1</i>
Model 2							
18	13611369	13611824	2.94E-5	7	5.57E-10	5.81E-07	<i>LDLRAD4</i>
19	50191341	50191882	1.03E-3	6	1.25E-09	1.09E-06	<i>PRMT1</i>
9	96623031	96623674	0.00136	4	7.50E-08	5.53E-05	<i>MIR4291</i>
8	99984349	99985049	1.12E-2	5	4.64E-06	3.14E-03	<i>OSR2</i>
12	4918390	4918848	1.71E-5	2	9.46E-06	9.74E-03	<i>KCNA6</i> <i>ATP13A4;ATP13A4-</i>
3	193272560	193273057	3.29E-2	5	1.15E-05	1.09E-02	<i>AS1</i>
20	13200684	13201551	1.80E-2	13	4.97E-05	2.68E-02	<i>ISMI</i>

P_{min}: the minimum P value in the DMR. Size: Number of methylation probes in the DMR.

P_{Comb-p}: combined Z-score based on Comb_P. P_{Sidak}: Sidak multiple-testing corrected P value

Table S 5 Top DMRs from race-specific analysis in hypertensive subjects

Chr	start	end	P _{min}	Size	P _{Comb-p}	P _{Sidak}	Gene
EA							
Model 1							
6	31691226	31692375	3.55E-8	29	1.04E-10	4.28E-8	<i>MPIG6B</i>
7	95025406	95026248	2.44E-4	21	1.50E-8	8.42E-6	<i>PON3</i>
19	843540	843995	2.13E-5	4	5.24E-8	5.45E-5	<i>PRTN3</i>
3	193272560	193273057	2.47E-3	5	2.89E-7	2.76E-4	<i>ATP13A4;ATP13A4-AS1</i>
15	93580021	93580327	0.002	5	3.65E-7	5.64E-4	<i>RGMA</i>
18	23713594	23714084	1.38E-2	9	4.15E-6	4.01E-3	<i>PSMA8</i>
1	209979282	209979779	2.94E-2	12	1.38E-5	1.31E-2	
5	176755198	176755805	0.040	5	2.26E-5	1.75E-2	<i>LMAN2</i>
9	96623031	96623480	0.014	3	2.91E-5	3.03E-2	<i>MIR4291</i>
17	41924042	41924577	3.59E-2	11	2.97E-5	2.60E-2	<i>CD300LG</i>
Model 2							
1	42384389	42384564	1.28E-5	3	3.39E-8	9.17E-5	<i>HIVEP3</i>
20	36148619	36149271	0.001	29	9.79E-8	7.11E-5	<i>BLCAP</i>
7	95025735	95026211	8.18E-4	18	2.48E-7	2.47E-4	<i>PON3</i>
16	979487	979662	0.002	3	3.60E-7	9.75E-4	<i>LMF1;LMF1-AS1</i>
18	77545027	77545223	0.005	3	5.97E-7	1.44E-3	<i>CTDP1</i>
15	93580021	93580327	5.09E-3	5	7.73E-7	1.20E-3	<i>RGMA</i>
3	193272560	193272869	9.77E-3	4	1.69E-6	2.59E-3	<i>ATP13A4;ATP13A4-AS1</i>
10	104196242	104196541	0.015	3	3.29E-6	5.20E-3	<i>MIR146B</i>
2	210074019	210074276	3.76E-3	2	4.08E-6	7.50E-3	<i>MAP2</i>
1	209979469	209979779	1.38E-2	9	5.70E-6	8.67E-3	<i>IRF6</i>
6	168491617	168491686	2.30E-2	3	5.91E-6	3.98E-2	<i>FRMD1</i>
5	147699717	147699892	0.031	3	9.37E-6	2.50E-2	<i>LOC102546294</i>
AA							
Model 1							
6	32118203	32119041	1.53E-6	17	7.51E-11	4.23E-8	<i>PRRT1</i>
17	80059407	80059540	0.0003	2	1.22E-8	4.33E-5	<i>CCDC57</i>
7	27183368	27184853	7.11E-4	43	1.93E-8	6.15E-6	<i>HOXA-AS3</i>
18	13611369	13611824	0.001	7	9.85E-8	1.02E-4	<i>LDLRAD4</i>
9	139715700	139716599	0.002	4	3.57E-7	1.88E-4	<i>RABL6</i>
5	150678161	150678334	0.004	2	9.82E-7	2.68E-3	<i>SLC36A3</i>
12	110156244	110156459	1.04E-2	5	3.21E-6	7.03E-3	<i>FAM222A</i>
19	49223833	49224454	1.62E-3	5	4.86E-6	3.69E-3	<i>RASIP1</i>
3	195489707	195490309	0.010	8	6.19E-6	4.84E-3	<i>MUC4</i>
8	145730817	145731409	7.40E-3	3	1.39E-5	1.11E-2	<i>GPT</i>
Model 2							
5	23507029	23507752	8.01E-7	12	1.61E-11	1.05E-8	<i>PRDM9</i>

7	1609340	1609908	1.76E-6	11	5.10E-10	4.24E-7	<i>PSMG3;PSMG3-AS1</i>
3	195489305	195490309	2.42E-5	9	1.79E-9	8.40E-7	<i>MUC4</i>
19	49223813	49224454	8.04E-5	6	6.19E-9	4.56E-6	<i>RASIP1</i>
18	13611369	13611824	8.04E-	7	7.48E-9	7.76E-6	<i>LDLRAD4</i>
5	150678161	150678334	0.002	2	3.37E-7	9.19E-4	<i>SLC36A3</i>
14	100203941	100204528	0.005	7	8.87E-7	7.14E-4	<i>CYP46A1</i>
11	123430574	123431162	2.03E-2	5	6.69E-6	5.36E-3	<i>GRAMD1B</i>
12	110156244	110156459	0.030	5	1.23E-5	2.67E-2	<i>FAM222A</i>
9	139716025	139716599	0.016	3	1.40E-5	1.14E-2	<i>RABL6</i>
16	87736669	87737091	0.015	4	3.43E-5	3.76E-2	<i>KLHDC4</i>

P_{min}: the minimum P value in the DMR. Size: Number of methylation probes in the DMR.

P_{Comb-p}: combined Z-score based on Comb_P. P_{Sidak}: Sidak multiple-testing corrected P value

Table S 6 Top DMRs from gene-based analysis in normotensive subjects

Chr	start	end	P _{min}	Size	P _{Comb-p}	P _{Sidak}	Gene
Model 1							
20	62693942	62694005	1.39E-08	6	4.93E-13	3.71E-09	<i>TCEA2</i>
2	27665016	27665306	3.04E-09	7	4.73E-12	7.73E-09	<i>NRBP1;KRTCAP3</i>
17	37123637	37123949	2.10E-05	9	9.75E-10	1.48E-06	<i>FBXO47</i>
3	195576389	195578280	9.28E-03	8	1.06E-09	2.66E-07	<i>LINC01983</i>
12	122018896	122019117	0.000243	7	1.44E-08	3.08E-05	<i>KDM2B</i>
1	84326461	84326856	2.80E-04	8	5.04E-08	6.04E-05	<i>LINC01725</i>
22	42828124	42828418	0.006238	5	1.61E-06	0.002586	<i>NFAM1</i>
11	89956516	89956777	9.21E-04	7	1.97E-06	3.57E-03	<i>CHORDC1</i>
16	57728631	57728700	0.008	2	2.33E-06	0.01584	<i>DRC7</i>
2	198650008	198651076	9.65E-04	2	2.59E-06	1.15E-03	<i>BOLL</i>
10	133946673	133946881	0.031	3	1.35E-05	0.0303	<i>JAKMIP3</i>
7	150498593	150498843	0.002	2	1.40E-05	2.62E-02	<i>TMEM176A</i>
19	57742216	57742444	0.014	8	1.43E-05	0.02918	<i>AURKC</i>
Model 2							
20	62693656	62694005	4.63E-13	7	2.82E-18	3.83E-15	<i>TCEA2</i>
3	195576389	195578280	3.92E-03	8	9.27E-11	2.32E-08	<i>LINC01983</i>
15	39871807	39872071	1.13E-06	6	3.34E-10	6.00E-07	<i>THBS1</i>
6	74072241	74072447	1.43E-05	7	6.03E-10	1.39E-06	<i>KHDC3L</i>
17	37123637	37123949	2.15E-05	9	1.40E-09	2.13E-06	<i>FBXO47</i>
16	8806530	8807043	8.69E-05	12	1.11E-08	1.02E-05	<i>ABAT</i>
8	142233381	142233705	1.94E-03	5	5.82E-07	8.51E-04	<i>SLC45A4</i>
22	42828124	42828516	6.57E-05	7	6.17E-07	0.000746	<i>NFAM1</i>
19	10736005	10736355	0.005097	7	9.67E-07	0.001308	<i>SLC44A2</i>
10	4230412	4230752	0.006747	6	1.40E-06	0.001948	<i>LINC00702</i>
11	89956572	89956777	4.49E-03	6	2.32E-06	5.34E-03	<i>CHORDC1</i>
4	165877874	165878317	2.26E-03	8	3.13E-06	3.34E-03	<i>FAM218A;TRIM61</i>
7	158280409	158280480	0.0143	2	4.23E-06	0.02781	<i>PTPRN2</i>
1	206785999	206786222	1.77E-02	6	5.58E-06	1.18E-02	<i>EIF2D</i> <i>LINC00390;SMIM2- AS1;</i>
13	44716512	44716778	0.01952	2	6.36E-06	1.13E-02	<i>SMIM2-AS1</i>
12	108168823	108169020	0.01409	3	9.70E-06	0.02305	<i>ASCL4</i>

P_{min}: the minimum P value in the DMR. Size: Number of methylation probes in the DMR.

P_{Comb-p}: combined Z-score based on Comb_P. P_{Sidak}: Sidak multiple-testing corrected P value

Table S 7 Top DMRs from race-specific gene-based analysis in normotensive subjects

Chr	start	end	P _{min}	Size	P _{Comb-p}	P _{Sidak}	Gene
EA							
Model 1							
							<i>CCDC144NL-AS1</i> ;
17	20799407	20799769	1.24E-09	10	3.01E-14	3.94E-11	<i>CCDC144NL</i>
2	27665016	27665306	1.28E-09	7	4.31E-12	7.05E-09	<i>NRBP1</i> ; <i>KRTCAP3</i>
17	37123637	37123949	1.53E-06	9	8.41E-11	1.28E-07	<i>FBXO47</i>
20	62693942	62694005	2.49E-06	6	2.19E-09	1.65E-05	<i>TCEA2</i>
12	122018769	122019117	2.31E-05	8	2.25E-09	3.06E-06	<i>KDM2B</i>
6	74072241	74072447	4.99E-05	7	5.47E-09	1.26E-05	<i>KHDC3L</i>
16	57728631	57728700	3.59E-04	2	4.77E-08	3.27E-04	<i>DRC7</i>
7	24323674	24323840	2.09E-05	6	3.68E-07	1.05E-03	<i>NPY</i>
11	66085249	66085299	5.39E-03	2	1.07E-06	1.01E-02	<i>CD248</i>
10	104535791	104536052	0.02529	8	1.04E-05	1.87E-02	<i>WBP1L</i>
Model 2							
20	62693942	62694005	4.26E-09	6	4.13E-13	3.11E-09	<i>TCEA2</i>
17	37123637	37123949	1.04E-06	9	4.40E-11	6.69E-08	<i>FBXO47</i>
15	39871807	39872071	1.04E-06	6	2.15E-10	3.86E-07	<i>THBS1</i>
6	74072241	74072447	4.39E-06	7	2.87E-10	6.61E-07	<i>KHDC3L</i>
17	17110119	17110353	2.35E-06	5	1.64E-09	3.31E-06	<i>PLD6</i>
8	144120334	144120706	5.88E-04	7	6.10E-07	7.77E-04	<i>C8orf31</i>
2	241497411	241497663	5.32E-03	7	1.10E-06	2.07E-03	<i>ANKMY1</i>
22	42828124	42828418	0.007968	5	2.12E-06	0.00341	<i>NFAM1</i>
7	158280409	158280480	9.85E-03	2	2.83E-06	1.87E-02	<i>PTPRN2</i>
11	66085249	66085299	1.10E-02	2	3.23E-06	3.02E-02	<i>CD248</i>
16	57728631	57728700	0.01098	2	3.45E-06	2.34E-02	<i>DRC7</i>
10	4230481	4230752	4.00E-02	5	2.84E-05	4.84E-02	<i>LINC00702</i>
AA							
Model 1							
5	135414857	135416613	4.67E-25	20	1.15E-34	3.10E-32	<i>VTRNA2-1</i>
5	6755025	6755843	9.57E-21	8	4.39E-25	2.54E-22	<i>PAPD7</i>
11	6518107	6518907	1.03E-11	11	7.47E-16	4.59E-13	<i>DNHDI</i>
19	37825008	37825679	5.20E-08	8	8.18E-11	5.76E-08	<i>HKR1</i>
3	49170495	49171185	1.36E-06	9	3.00E-10	2.05E-07	<i>LAMB2</i>
16	56995200	56995856	7.71E-06	6	4.87E-09	3.51E-06	<i>CETP</i>
3	195576389	195578280	2.47E-02	8	7.79E-09	1.95E-06	<i>LINC01983</i>
7	24323260	24323799	3.27E-06	7	6.64E-08	5.82E-05	<i>NPY</i>
2	239139910	239140733	2.00E-04	15	9.66E-08	5.54E-05	<i>LOC151174</i> ; <i>LOC643387</i>
19	58595667	58596261	0.001005	3	2.49E-07	0.000198	<i>ZSCAN18</i>
4	81118793	81119473	1.27E-03	7	7.53E-07	5.23E-04	<i>PRDM8</i>
17	79225572	79225974	0.00173	4	2.03E-06	0.002379	<i>SLC38A10</i>

11	65315029	65315466	7.45E-06	4	2.13E-06	2.30E-03	<i>LTBP3</i>
13	113776872	113777160	2.52E-02	9	1.39E-05	2.25E-02	<i>F10</i>
12	132859440	132859951	0.04213	5	2.77E-05	2.53E-02	<i>GALNT9</i>
Model 2							
5	135415128	135416613	2.54E-12	19	1.25E-18	3.98E-16	<i>VTRNA2-1</i>
6	32063393	32064212	5.43E-10	24	1.79E-14	1.04E-11	<i>TNXB</i>
5	6755025	6755439	4.95E-12	6	1.22E-13	1.39E-10	<i>PAPD7</i>
11	6518107	6518907	1.00E-08	11	9.34E-13	5.52E-10	<i>DNHD1</i>
1	45965448	45966115	9.51E-08	11	1.02E-11	7.22E-09	<i>MMACHC; CCDC163</i>
13	113776872	113777160	1.78E-07	9	2.34E-11	3.84E-08	<i>F10</i>
2							<i>STON1-GTF2A1L;</i>
	48844727	48845068	6.80E-07	8	1.07E-10	1.48E-07	<i>GTF2A1L</i>
3	49170495	49171343	1.50E-04	10	4.17E-08	2.32E-05	<i>LAMB2</i>
2	239139517	239140733	3.21E-08	16	6.01E-08	2.33E-05	<i>LOC151174; LOC643387</i>
21	47717405	47718080	0.000997	5	3.69E-07	0.000258	<i>YBEY</i>
20	259897	259925	0.005748	2	2.85E-06	0.04691	<i>C20orf96</i>
4	74718977	74719597	6.25E-03	6	3.28E-06	2.50E-03	<i>PF4V1</i>
10	135341869	135342936	0.01292	6	1.03E-05	0.004552	<i>CYP2E1</i>
6	149777805	149778058	0.01522	3	1.09E-05	0.02007	<i>ZC3H12D</i>
11	116699588	116700049	1.25E-02	3	1.87E-05	1.90E-02	<i>APOC3</i>

Table S 8 DNA features and regulatory elements at target DNAm loci

ProbeID	Chr	Pos	Score	BP	Relevant Tissue*
cg03116124	1	231293208	5		
cg24202936	11	50257256	5		
cg17577122	22	19511967	2b	<i>POLR2A</i>	umbilical vein endothelial cells
cg06809326	17	20799526	2b	<i>CTCF</i>	umbilical vein endothelial cells, brain microvascular endothelial cells, glioblastoma cells
cg04245766	14	54421051	4	<i>POLR2A</i>	umbilical vein endothelial cells
cg06450373	5	19766434	6		
cg18950108	6	30920171	5		
cg07675682	17	78195515	5		
cg10051414	22	38598733	2b	<i>POLR2A</i>	umbilical vein endothelial cells
				<i>MAX</i>	umbilical vein endothelial cells
				<i>CTCF</i>	umbilical vein endothelial cells, brain microvascular endothelial cells
				<i>REST</i>	glioblastoma/ astrocytoma
cg25317585	13	102568886	4	<i>EZH2</i>	astrocytes
				<i>CTCF</i>	umbilical vein endothelial cells, brain microvascular endothelial cells, glioblastoma cells, neuroblastoma cells
				<i>RAD21</i>	neuroblastoma cells
				<i>REST</i>	glioblastoma/ astrocytoma cells
				<i>YY1</i>	neuroblastoma cells
				<i>REST</i>	neuroblastoma cells
cg02741985	17	80059408	4	<i>REST</i>	neuroblastoma cells

Chr: Chromosome. Pos: Position. BP: Binding Protein. *Endothelial cells and brain tissues were reviewed.

Table S 9 Transcription factors enriched in the fetal brain tissue - eFORGE

ProbeID	Motif Name	Binding Site	Database	P value	Q value
cg06450373	ONECUT1 CUT 1	5:19766428-19766442:-	Taipale/SELEX	1.75E-07	4.65E-06
	ONECUT1 CUT 2	5:19766428-19766442:-	Taipale/SELEX	2.05E-07	4.65E-06
	ONECUT2 CUT 1	5:19766428-19766442:-	Taipale/SELEX	2.05E-07	4.65E-06
	ONECUT3 CUT 1	5:19766428-19766442:-	Taipale/SELEX	1.65E-07	4.65E-06
cg24202936	V MTERF 01	11:50257253-50257267:-	TRANSFAC	3.61E-06	6.55E-05
	Rara secondary	11:50257251-50257267:-	UniProbe	1.92E-04	2.18E-03
	Klf7 primary	11:50257254-50257270:+	UniProbe	7.55E-03	3.42E-02
	Ascl2 secondary	11:50257254-50257270:+	UniProbe	9.50E-03	3.91E-02
cg25317585	V_NRSF_Q4	13:102568875-102568894:+	TRANSFAC	1.17E-03	8.16E-03
cg06809326	PAX2 PAX 1	17:20799510-20799528:+	Taipale/SELEX	5.46E-05	8.25E-04
	PAX5 PAX 1	17:20799510-20799528:+	Taipale/SELEX	6.69E-05	8.67E-04
	NRF1_NRF_1	17:20799523-20799535:-/+	Taipale/SELEX	4.64E-04	4.21E-03
	V_NRF1_Q6	17:20799525-20799535:+17:20799523-20799533:-	TRANSFAC	4.63E-04	4.21E-03
	V_ZFP206_01	17:20799524-20799535:-17:20799523-20799534:+	TRANSFAC	1.04E-03	7.85E-03
	MA0014.1-Pax5	17:20799509-20799529:-	JASPAR	1.66E-03	9.50E-03
	V_PAX5_01	17:20799505-20799533:-	TRANSFAC	3.61E-03	1.72E-02
cg17577122	Egr3.mouse_C2H2_1	22:19511961-19511976:-	Taipale/SELEX	8.72E-04	7.18E-03
	Egr1.mouse_C2H2_1	22:19511961-19511977:-	Taipale/SELEX	1.40E-03	9.08E-03
	EGR1_C2H2_1	22:19511962-19511976:-	Taipale/SELEX	1.68E-03	9.50E-03
	EGR3_C2H2_1	22:19511962-19511977:-	Taipale/SELEX	2.49E-03	1.33E-02
	Egr1_primary	22:19511962-19511976:-	UniProbe	3.31E-03	1.67E-02
	Sp4_primary	22:19511962-19511979:-	UniProbe	8.92E-03	3.85E-02
	V_KROX_Q6	22:19511962-19511976:-	TRANSFAC	1.18E-02	4.67E-02
	V_SP1_Q6	22:19511966-19511979:+	TRANSFAC	1.72E-02	6.50E-02

TF binding site: Transcription factor binding site. P value: p-value calculated for sequence overlaps with each of the motifs. Q-value: Multiple-testing corrected p-value with the Benjamini-Yekutieli method (FDR 0.1%)

Table S 10 Shared epigenetic associations between white matter hyperintensities (WMH) burden and cognitive abilities (CA)

CpG	Gene	Z _{WMH}	P _{WMH}	Z _{CA}	P _{CA}	P _{joint}	Q _{joint}
Reduced model							
<i>Digital Symbol</i>							
cg05575921	AHRR	-2.79	5.30E-03	6.62	3.59E-11	3.19E-11	1.35E-05
cg18181703	SOCS3	-1.59	1.13E-01	6.62	3.53E-11	2.11E-10	4.29E-05
cg21566642		-2.81	4.99E-03	6.24	4.49E-10	3.06E-10	4.29E-05
cg05951221		-3.61	3.12E-04	5.75	9.12E-09	4.55E-10	4.79E-05
cg12593793		-1.99	4.70E-02	6.20	5.82E-10	1.86E-09	1.57E-04
cg03366574		1.83	6.68E-02	-6.15	8.00E-10	3.09E-09	2.17E-04
cg03636183	F2RL3	-2.92	3.54E-03	5.67	1.47E-08	5.24E-09	3.16E-04
cg06126421		-2.82	4.87E-03	5.60	2.17E-08	9.61E-09	5.06E-04
cg26470501	BCL3	-2.98	2.90E-03	5.34	9.49E-08	2.29E-08	1.07E-03
cg06946797		-1.82	6.95E-02	5.73	1.02E-08	3.43E-08	1.22E-03
cg17417856	PRMT1;C19orf76	-4.72	2.40E-06	3.74	1.82E-04	3.77E-08	1.22E-03
cg17759224		0.79	4.28E-01	5.67	1.47E-08	3.26E-08	1.22E-03
cg22953759	ANKRD33B	-0.73	4.63E-01	5.85	4.88E-09	3.71E-08	1.22E-03
cg00134210	FAM107B	-1.63	1.04E-01	5.65	1.58E-08	6.53E-08	1.73E-03
cg13092901	TYMP;SCO2	-2.90	3.72E-03	5.16	2.50E-07	6.57E-08	1.73E-03
cg17386185	GLYCTK	-2.21	2.71E-02	5.49	4.00E-08	6.31E-08	1.73E-03
cg14943908	BAT2	-1.66	9.70E-02	5.51	3.59E-08	1.35E-07	3.33E-03
cg19405484		-1.26	2.10E-01	5.57	2.57E-08	1.47E-07	3.41E-03
cg19859270	GPR15	-3.87	1.09E-04	4.26	2.00E-05	1.54E-07	3.41E-03
cg01940273		-1.94	5.30E-02	5.39	6.92E-08	1.67E-07	3.47E-03
cg03183540		-1.43	1.54E-01	5.49	4.12E-08	1.97E-07	3.47E-03
cg04987734	CDC42BPB	3.24	1.21E-03	-4.68	2.87E-06	2.19E-07	3.47E-03
cg07195224	AIM2	-1.67	9.57E-02	5.41	6.20E-08	2.21E-07	3.47E-03
cg16594502		-2.64	8.29E-03	5.07	4.08E-07	1.95E-07	3.47E-03
cg21108085	CD82;CD82	-1.61	1.08E-01	5.46	4.90E-08	1.91E-07	3.47E-03
cg22331294	LIMD1	0.99	3.25E-01	5.27	1.35E-07	2.21E-07	3.47E-03
cg25376310	ZDHHC14	-2.36	1.83E-02	5.17	2.29E-07	2.22E-07	3.47E-03
cg01657995	C6orf48;SNORD5						
cg07660627	2	-1.05	2.95E-01	5.50	3.92E-08	2.49E-07	3.75E-03
cg07660627	ACACA	-0.56	5.74E-01	-5.36	8.55E-08	2.66E-07	3.86E-03
cg00252813	GAPDH	-1.61	1.08E-01	5.38	7.41E-08	2.82E-07	3.96E-03
cg12142865	ALDH2	-0.82	4.14E-01	5.46	4.80E-08	3.32E-07	4.51E-03
cg01406381	SLC1A5	-1.25	2.12E-01	5.37	7.79E-08	4.21E-07	5.49E-03
cg05673882	POLK	-2.64	8.35E-03	4.89	1.03E-06	4.42E-07	5.49E-03
cg12473916	SHC1	-2.09	3.67E-02	5.14	2.71E-07	4.43E-07	5.49E-03
cg02003183	CDC42BPB	2.41	1.58E-02	-4.99	6.03E-07	4.64E-07	5.58E-03
cg14602222	RAD52	2.59	9.73E-03	-4.89	1.02E-06	4.97E-07	5.82E-03
cg00780520	PVT1	-1.40	1.61E-01	5.30	1.14E-07	5.26E-07	5.99E-03
cg03881294		-0.50	6.18E-01	5.33	1.00E-07	6.95E-07	7.71E-03

cg02657160	<i>CPOX</i>	-2.59	9.65E-03	4.79	1.68E-06	7.64E-07	7.86E-03
cg10475172	<i>CCR9</i>	0.91	3.63E-01	-5.30	1.17E-07	7.57E-07	7.86E-03
cg21171320	<i>TRPM3</i>	-0.35	7.24E-01	5.30	1.13E-07	7.57E-07	7.86E-03
cg01955153		0.89	3.74E-01	5.06	4.26E-07	8.25E-07	8.28E-03
cg27050612	<i>NFE2L1</i>	-2.28	2.27E-02	4.92	8.67E-07	8.52E-07	8.35E-03
cg04885881		-3.34	8.48E-04	4.25	2.17E-05	9.43E-07	9.02E-03
cg15095917	<i>CD68</i>	-0.38	7.05E-01	5.26	1.44E-07	9.63E-07	9.02E-03
cg11231349	<i>NOS1AP</i>	-2.02	4.31E-02	4.99	6.00E-07	1.01E-06	9.10E-03
cg23842572	<i>MPRIIP</i>	0.66	5.06E-01	-5.25	1.49E-07	1.02E-06	9.10E-03
cg00310412	<i>SEMA7A</i>	-1.65	9.91E-02	5.11	3.30E-07	1.06E-06	9.30E-03
cg16736826	<i>EDN2</i>	-2.54	1.11E-02	4.68	2.90E-06	1.38E-06	1.19E-02
cg04158018	<i>NFE2</i>	-1.20	2.32E-01	5.14	2.80E-07	1.46E-06	1.21E-02
cg06809326	<i>CCDC144NL</i>	4.99	6.14E-07	0.75	4.52E-01	1.47E-06	1.21E-02
cg04973995		0.82	4.13E-01	4.96	7.16E-07	1.55E-06	1.26E-02
cg10721220		-0.91	3.64E-01	5.15	2.56E-07	1.60E-06	1.27E-02
cg24202936	<i>LOC441601</i>	5.08	3.78E-07	0.19	8.47E-01	1.74E-06	1.36E-02
cg26140475		0.54	5.91E-01	5.00	5.66E-07	1.80E-06	1.38E-02
cg02068690	<i>DTNB</i>	0.72	4.75E-01	-5.12	2.99E-07	1.97E-06	1.39E-02
cg02507296	<i>C10orf95</i>	-0.23	8.16E-01	5.11	3.20E-07	1.98E-06	1.39E-02
cg02519286	<i>GAPDH</i>	-0.43	6.65E-01	5.13	2.85E-07	1.89E-06	1.39E-02
cg04424621	<i>HIST1H2BJ</i>	-1.80	7.16E-02	4.92	8.67E-07	2.04E-06	1.39E-02
cg11146034	<i>ELK3</i>	-2.10	3.56E-02	4.80	1.57E-06	2.05E-06	1.39E-02
cg21139312	<i>MSI2</i>	0.84	4.03E-01	-5.12	3.02E-07	1.92E-06	1.39E-02
cg26856257		0.41	6.79E-01	5.01	5.35E-07	1.98E-06	1.39E-02
cg01127300		-3.97	7.21E-05	3.42	6.33E-04	2.09E-06	1.40E-02
cg07960624	<i>SAMD12</i>	-1.18	2.38E-01	5.06	4.23E-07	2.18E-06	1.44E-02
cg09586583	<i>SSBP3</i>	-1.66	9.72E-02	-4.56	5.18E-06	2.30E-06	1.49E-02
cg00611192	<i>ZNF692</i>	-2.41	1.58E-02	4.60	4.26E-06	2.57E-06	1.53E-02
cg02010481	<i>JAZF1</i>	-1.80	7.19E-02	4.89	1.04E-06	2.40E-06	1.53E-02
cg09182189	<i>NADK</i>	-2.19	2.89E-02	4.71	2.45E-06	2.57E-06	1.53E-02
cg09676630	<i>FBRSL1</i>	-4.49	7.16E-06	2.63	8.54E-03	2.47E-06	1.53E-02
cg14277403	<i>ANP32B</i>	-2.95	3.14E-03	4.28	1.83E-05	2.46E-06	1.53E-02
cg18006990	<i>GTPBP2</i>	-2.22	2.65E-02	4.70	2.61E-06	2.55E-06	1.53E-02
cg14753356		-1.40	1.61E-01	4.97	6.73E-07	2.74E-06	1.59E-02
cg19572487	<i>RARA</i>	-1.91	5.66E-02	4.82	1.47E-06	2.75E-06	1.59E-02
cg16878214	<i>PLEKHF2</i>	-1.79	7.36E-02	-4.45	8.54E-06	2.82E-06	1.60E-02
cg02767093	<i>STK24</i>	-3.19	1.40E-03	4.04	5.34E-05	3.17E-06	1.77E-02
cg03707168	<i>PPP1R15A</i>	-0.12	9.08E-01	5.01	5.56E-07	3.20E-06	1.77E-02
cg05438378	<i>SMAD3</i>	-2.43	1.53E-02	4.53	5.97E-06	3.38E-06	1.83E-02
cg26729380	<i>TNF</i>	0.70	4.87E-01	4.84	1.31E-06	3.40E-06	1.83E-02
cg16618104	<i>CHST11</i>	0.85	3.96E-01	4.79	1.68E-06	3.47E-06	1.85E-02
cg24296397	<i>BSN</i>	1.31	1.91E-01	-4.93	8.08E-07	3.57E-06	1.88E-02
cg26695387	<i>MAP2K3</i>	-1.14	2.54E-01	4.95	7.28E-07	3.73E-06	1.94E-02
cg11344352	<i>ERCC1</i>	-1.59	1.12E-01	4.85	1.22E-06	3.78E-06	1.94E-02
cg04949225		-0.95	3.41E-01	4.97	6.76E-07	3.93E-06	2.00E-02

cg00415665	<i>ZHX2</i>	0.55	5.85E-01	-4.98	6.43E-07	4.18E-06	2.03E-02
cg17577122	<i>CLDN5</i>	4.89	1.02E-06	-1.37	1.70E-01	4.14E-06	2.03E-02
cg19940644		-2.80	5.13E-03	4.26	2.05E-05	4.06E-06	2.03E-02
cg21241410		0.89	3.71E-01	4.74	2.16E-06	4.17E-06	2.03E-02
cg03116124		-4.91	9.32E-07	1.22	2.22E-01	4.39E-06	2.03E-02
cg05168229		-0.41	6.84E-01	4.96	6.91E-07	4.44E-06	2.03E-02
cg06285727	<i>ATG16L2</i>	0.46	6.47E-01	4.85	1.23E-06	4.28E-06	2.03E-02
cg15705273	<i>NCLN</i>	2.53	1.16E-02	-4.41	1.01E-05	4.30E-06	2.03E-02
cg16346032		-0.02	9.88E-01	4.93	8.22E-07	4.41E-06	2.03E-02
cg08620426		-0.31	7.61E-01	-4.87	1.12E-06	4.58E-06	2.07E-02
cg10717214	<i>TNF</i>	0.83	4.05E-01	4.73	2.20E-06	5.04E-06	2.25E-02
cg25375916	<i>SLC33A1</i>	0.51	6.08E-01	4.82	1.47E-06	5.08E-06	2.25E-02
cg13854219		1.28	2.02E-01	4.53	5.86E-06	5.40E-06	2.35E-02
cg27115863		-1.41	1.59E-01	4.82	1.42E-06	5.40E-06	2.35E-02
cg07092212	<i>DGKZZ</i>	-0.61	5.44E-01	4.92	8.89E-07	5.49E-06	2.36E-02
cg12966875	<i>SLPI</i>	0.71	4.78E-01	4.74	2.12E-06	5.76E-06	2.45E-02
cg00144180	<i>HDAC4</i>	0.60	5.47E-01	-4.90	9.42E-07	6.04E-06	2.54E-02
cg13709639	<i>TUBA1B</i>	-1.80	7.16E-02	4.67	2.99E-06	6.15E-06	2.57E-02
cg24837149	<i>CCNL2;LOC148413</i>	-2.41	1.58E-02	4.38	1.21E-05	6.49E-06	2.65E-02
cg26079320	<i>POGK</i>	-0.52	6.03E-01	4.89	1.02E-06	6.48E-06	2.65E-02
cg15030712	<i>CHN2</i>	0.06	9.55E-01	4.84	1.28E-06	6.66E-06	2.70E-02
cg02610723	<i>FAM38A</i>	0.98	3.27E-01	4.62	3.80E-06	6.98E-06	2.78E-02
cg05492306	<i>ERCC1</i>	-1.43	1.52E-01	4.76	1.93E-06	6.94E-06	2.78E-02
cg25648203	<i>AHRR</i>	-0.64	5.20E-01	4.87	1.15E-06	7.24E-06	2.85E-02
cg10021364	<i>MYADM</i>	-2.49	1.29E-02	4.28	1.84E-05	7.89E-06	3.07E-02
cg25607249	<i>SLC1A5</i>	-0.46	6.43E-01	4.85	1.21E-06	7.95E-06	3.07E-02
cg03535253	<i>BTBD7</i>	0.01	9.96E-01	-4.83	1.38E-06	8.28E-06	3.17E-02
cg16253157	<i>CYB561D1</i>	-0.22	8.25E-01	-4.80	1.58E-06	8.50E-06	3.23E-02
cg27067618	<i>CYP4F3</i>	-0.09	9.30E-01	4.83	1.39E-06	8.65E-06	3.25E-02
cg04994217	<i>C9orf170</i>	-3.32	9.12E-04	3.64	2.69E-04	9.07E-06	3.38E-02
cg02656560		-2.39	1.69E-02	4.30	1.68E-05	9.65E-06	3.57E-02
cg00711496	<i>C19orf76;PRMT1</i>	-4.12	3.86E-05	2.69	7.11E-03	9.98E-06	3.66E-02
<i>MMSE</i>							
cg01127300		-3.97	7.21E-05	3.74	1.86E-04	3.93E-07	2.37E-02
cg03636183	<i>F2RL3</i>	-2.92	3.54E-03	4.97	6.78E-07	7.13E-08	7.35E-03
cg05575921	<i>AHRR</i>	-2.79	5.30E-03	5.14	2.77E-07	4.38E-08	7.35E-03
cg05704155	<i>PGD;PGD</i>	4.55	5.26E-06	-3.33	8.65E-04	1.39E-07	9.75E-03
cg05951221		-3.61	3.12E-04	4.84	1.29E-06	1.45E-08	6.10E-03
cg06126421		-2.82	4.87E-03	4.60	4.31E-06	5.47E-07	2.88E-02
cg19859270	<i>GPR15</i>	-3.87	1.09E-04	3.60	3.18E-04	9.48E-07	4.44E-02
cg21450381		-0.13	8.98E-01	-5.72	1.04E-08	7.45E-08	7.35E-03
cg22529900	<i>PARK7</i>	-4.47	7.87E-06	-2.53	1.15E-02	1.12E-06	4.74E-02
cg25923609		3.59	3.32E-04	-4.46	8.17E-06	8.71E-08	7.35E-03
<i>Verbal Fluency</i>							

cg04987734	<i>CDC42BPB</i>	3.24	1.21E-03	-4.47	7.89E-06	4.75E-07	2.86E-02
cg07102705	<i>HTR4</i>	3.08	2.07E-03	-4.48	7.56E-06	7.23E-07	3.54E-02
cg10313673	<i>CILP2</i>	1.75	7.99E-02	-5.07	3.90E-07	9.85E-07	3.78E-02
cg10922280	<i>DPEP2</i>	1.23	2.20E-01	-5.59	2.33E-08	1.31E-07	1.38E-02
cg12025310	<i>TMPRSS6</i>	3.43	6.00E-04	-4.05	5.16E-05	1.36E-06	4.77E-02
cg12507869	<i>INPP5A</i>	-0.04	9.69E-01	-5.81	6.44E-09	3.78E-08	9.83E-03
cg12582616	<i>JARID2</i>	-3.54	3.96E-04	-4.15	3.38E-05	4.66E-08	9.83E-03
cg13696609	<i>FOXE1</i>	-1.21	2.27E-01	-4.97	6.66E-07	8.40E-07	3.54E-02
cg14945937	<i>PLEKHF1</i>	-1.95	5.16E-02	-4.79	1.68E-06	4.19E-07	2.86E-02
cg15909981	<i>TMPRSS6</i>	-1.49	1.36E-01	-4.88	1.06E-06	7.64E-07	3.54E-02
cg16201957	<i>MAML3</i>	1.17	2.44E-01	-5.67	1.41E-08	8.45E-08	1.19E-02
cg23722790	<i>SLC35D1</i>	0.47	6.39E-01	-5.56	2.71E-08	1.97E-07	1.66E-02
<i>Vocabulary</i>							
cg04513006	<i>ESRP2</i>	0.41	6.85E-01	-5.65	1.60E-08	1.18E-07	4.96E-02
Full model							
<i>MMSE</i>							
cg05704155	<i>PGD</i>	4.88	1.07E-06	-3.53	4.11E-04	1.42E-08	6.00E-03
cg21450381		-0.05	9.57E-01	-5.66	1.55E-08	1.15E-07	2.43E-02
<i>Verbal Fluency</i>							
cg12507869	<i>INPP5A</i>	-0.82	4.11E-01	-5.96	2.47E-09	8.79E-09	3.71E-03
cg16201957	<i>MAML3</i>	1.35	1.78E-01	-5.46	4.71E-08	1.80E-07	3.79E-02

Table S 11 Correlation between Methylation Levels in Blood and Brain

Brain tissues	cg02741985	cg03116124	cg04245766	cg06450373	cg06809326	cg07675682
PFC (n=74)	-0.053, 0.652	0.292, 1.16E-02	0.923, 1.38E-31	0.227, 0.052	0.566, 1.49E-07	-0.114, 0.333
EC (n=71)	0.248, 3.68E-02	0.177, 0.141	0.954, 6.89E-38	-0.037, 0.763	0.711, 3.83E-12	-0.078, 0.518
STG (n=75)	0.049, 0.674	0.218, 6.00E-02	0.924, 4.17E-32	-0.036, 0.763	0.601, 1.15E-08	-0.048, 0.686
CER (n=71)	0.202, 9.17E-02	0.242, 4.2E-02	0.939, 1.16E-33	0.003, 0.982	0.582, 1.02E-07	-0.070, 0.562
Brain tissues	cg10051414	cg17577122	cg18950108	cg24202936	cg25317585	
PFC (n=74)	-0.102, 0.388	-0.158, 0.178	0.122, 0.302	0.327, 0.005	0.102, 0.389	
EC (n=71)	-0.158, 0.189	0.122, 0.311	-0.010, 0.937	0.398, 5.81E-04	0.043, 0.724	
STG (n=75)	0.066, 0.575	0.051, 0.666	0.255, 2.73E-02	0.324, 4.60E-03	-0.123, 0.295	
CER (n=71)	-0.312, 8.18E-03	0.122, 0.311	0.106, 0.380	0.158, 0.189	0.020, 0.871	

Correlation r and P value. PFC: prefrontal cortex, EC: entorhinal cortex, STG: superior temporal gyrus and CER: cerebellum

REFERENCES

1. Colin B, Novalia S, C. PA, et al. Small-Vessel Disease in the Heart and Brain: Current Knowledge, Unmet Therapeutic Need, and Future Directions. *J Am Heart Assoc.* 2019;8(3):e011104. doi:10.1161/JAHA.118.011104
2. Fulton WFM. Arterial Anastomoses in the Coronary Circulation: I. Anatomical Features in Normal and Diseased Hearts Demonstrated by Stereoarteriography. *Scott Med J.* 1963;8(11):420-434. doi:10.1177/003693306300801102
3. Noel BMC, J. PC, Norine WM, et al. Ischemia and No Obstructive Coronary Artery Disease (INOCA). *Circulation.* 2017;135(11):1075-1092. doi:10.1161/CIRCULATIONAHA.116.024534
4. Ford TJ, Corcoran D, Berry C. Stable coronary syndromes: pathophysiology, diagnostic advances and therapeutic need. *Heart.* 2018;104(4):284-292. doi:10.1136/heartjnl-2017-311446
5. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol.* 2010;9(7):689-701. doi:10.1016/S1474-4422(10)70104-6
6. Pantoni L, Sarti C, Alafuzoff I, et al. Postmortem examination of vascular lesions in cognitive impairment: a survey among neuropathological services. *Stroke.* 2006;37(4):1005-1009. doi:10.1161/01.STR.0000206445.97511.ae
7. Bullmore E, Sporns O. The economy of brain network organization. *Nat Rev Neurosci.* 2012;13(5):336-349. doi:10.1038/nrn3214

8. Zlokovic B V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nat Rev Neurosci.* 2011;12(12):723-738. doi:10.1038/nrn3114
9. Blumenfeld H. CHAPTER 10 Cerebral Hemispheres and Vascular Supply. In: *Neuroanatomy Through Clinical Cases. Second.* Sunderland, Massachusetts: Sinauer Associates, Inc.; 2010:391-399.
10. Van Den Bergh R. Centrifugal Elements in the Vascular Pattern of the Deep Intracerebral Blood Supply. *Angiology.* 1969;20(2):88-94.
doi:10.1177/000331976902000205
11. Rowbotham GF, Little E. Circulations of the cerebral hemispheres. *Br J Surg.* 1965;52:8-21.
12. de Reuck J. The Human Periventricular Arterial Blood Supply and the Anatomy of Cerebral Infarctions. *Eur Neurol.* 1971;5(6):321-334. doi:10.1159/000114088
13. Pantoni L, Garcia JH. Pathogenesis of leukoaraiosis: a review. *Stroke.* 1997;28(3):652-659.
14. Hainsworth AH. White matter lesions in cerebral small vessel disease: Underperfusion or leaky vessels? *Neurology.* 2019;92(15):687-688.
doi:10.1212/WNL.00000000000007258
15. Adams HPJ, Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke.* 1993;24(1):35-41.

16. Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG. New approach to stroke subtyping: the A-S-C-O (phenotypic) classification of stroke. *Cerebrovasc Dis*. 2009;27(5):502-508. doi:10.1159/000210433
17. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke*. 2007;38(11):2979-2984. doi:10.1161/STROKEAHA.107.490896
18. Thompson CS, Hakim AM. Living beyond our physiological means: small vessel disease of the brain is an expression of a systemic failure in arteriolar function: a unifying hypothesis. *Stroke*. 2009;40(5):e322-30. doi:10.1161/STROKEAHA.108.542266
19. Grau-Olivares M, Arboix A. Mild cognitive impairment in stroke patients with ischemic cerebral small-vessel disease: a forerunner of vascular dementia? *Expert Rev Neurother*. 2009;9(8):1201-1217. doi:10.1586/ern.09.73
20. J. BE, Paul M, Alvaro A, et al. Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. *Circulation*. 2019;139(10):e56-e528. doi:10.1161/CIR.0000000000000659
21. Gonzalez-Perez A, Gaist D, Wallander M-A, McFeat G, Garcia-Rodriguez LA. Mortality after hemorrhagic stroke: data from general practice (The Health Improvement Network). *Neurology*. 2013;81(6):559-565. doi:10.1212/WNL.0b013e31829e6eff
22. Warlow C, Sudlow C, Dennis M, Wardlaw J, Sandercock P. Stroke. *Lancet* (London, England). 2003;362(9391):1211-1224. doi:10.1016/S0140-6736(03)14544-8

23. Petty GW, Brown RDJ, Whisnant JP, Sicks JD, O'Fallon WM, Wiebers DO. Ischemic stroke subtypes : a population-based study of functional outcome, survival, and recurrence. *Stroke*. 2000;31(5):1062-1068.
24. Bamford J, Sandercock P, Jones L, Warlow C. The natural history of lacunar infarction: the Oxfordshire Community Stroke Project. *Stroke*. 1987;18(3):545-551.
25. Hara K, Shiga A, Fukutake T, et al. Association of HTRA1 Mutations and Familial Ischemic Cerebral Small-Vessel Disease. *N Engl J Med*. 2009;360(17):1729-1739.
doi:10.1056/NEJMoa0801560
26. Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging. *Lancet Neurol*. 2013;12(5):483-497.
doi:10.1016/S1474-4422(13)70060-7
27. Inzitari D, Pracucci G, Poggesi A, et al. Changes in white matter as determinant of global functional decline in older independent outpatients: three year follow-up of LADIS (leukoaraiosis and disability) study cohort. *BMJ*. 2009;339:b2477.
doi:10.1136/bmj.b2477
28. Hachinski VC, Potter P, Merskey H. Leuko-araiosis. *Arch Neurol*. 1987;44(1):21-23.
29. Fazekas F, Kleinert R, Offenbacher H, et al. Pathologic correlates of incidental MRI white matter signal hyperintensities. *Neurology*. 1993;43(9):1683-1689.
30. Kwa VI, Stam J, Blok LM, Verbeeten BJ. T2-weighted hyperintense MRI lesions in the pons in patients with atherosclerosis. Amsterdam Vascular Medicine Group. *Stroke*. 1997;28(7):1357-1360.

31. Prins ND, Scheltens P. White matter hyperintensities, cognitive impairment and dementia: an update. *Nat Rev Neurol*. 2015;11:157.
<https://doi.org/10.1038/nrneurol.2015.10>.
32. Debette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ*. 2010;341:c3666. doi:10.1136/bmj.c3666
33. Gouw AA, Seewann A, van der Flier WM, et al. Heterogeneity of small vessel disease: a systematic review of MRI and histopathology correlations. *J Neurol Neurosurg Psychiatry*. 2011;82(2):126-135. doi:10.1136/jnnp.2009.204685
34. Wardlaw JM, Dennis MS, Warlow CP, Sandercock PA. Imaging appearance of the symptomatic perforating artery in patients with lacunar infarction: occlusion or other vascular pathology? *Ann Neurol*. 2001;50(2):208-215.
35. Vermeer SE, Longstreth WTJ, Koudstaal PJ. Silent brain infarcts: a systematic review. *Lancet Neurol*. 2007;6(7):611-619. doi:10.1016/S1474-4422(07)70170-9
36. Rost NS, Rahman RM, Biffi A, et al. White matter hyperintensity volume is increased in small vessel stroke subtypes. *Neurology*. 2010;75(19):1670-1677.
doi:10.1212/WNL.0b013e3181fc279a
37. Doubal FN, MacLulich AMJ, Ferguson KJ, Dennis MS, Wardlaw JM. Enlarged perivascular spaces on MRI are a feature of cerebral small vessel disease. *Stroke*. 2010;41(3):450-454. doi:10.1161/STROKEAHA.109.564914

38. Cordonnier C, Al-Shahi Salman R, Wardlaw J. Spontaneous brain microbleeds: systematic review, subgroup analyses and standards for study design and reporting. *Brain*. 2007;130(Pt 8):1988-2003. doi:10.1093/brain/awl387
39. Vermeer SE, Hollander M, van Dijk EJ, Hofman A, Koudstaal PJ, Breteler MMB. Silent brain infarcts and white matter lesions increase stroke risk in the general population: the Rotterdam Scan Study. *Stroke*. 2003;34(5):1126-1129. doi:10.1161/01.STR.0000068408.82115.D2
40. Aribisala BS, Valdes Hernandez MC, Royle NA, et al. Brain atrophy associations with white matter lesions in the ageing brain: the Lothian Birth Cohort 1936. *Eur Radiol*. 2013;23(4):1084-1092. doi:10.1007/s00330-012-2677-x
41. Fornage M, Beecham AH. The emerging genetic landscape of cerebral white matter hyperintensities. *Neurology*. 2019;92(8):355-356. doi:10.1212/WNL.0000000000006936
42. Debette S, Schilling S, Duperron M-G, 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. doi:10.1001/jamaneurol.2018.3122
43. Wardlaw JM, Smith EE, Biessels GJ, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol*. 2013;12(8):822-838. doi:10.1016/S1474-4422(13)70124-8

44. Wadhwa R, Wen W, Frankland A, et al. White matter hyperintensities in young individuals with bipolar disorder or at high genetic risk. *J Affect Disord.* 2019;245:228-236. doi:10.1016/j.jad.2018.10.368
45. Williamson W, Lewandowski AJ, Forkert ND, et al. Association of Cardiovascular Risk Factors With MRI Indices of Cerebrovascular Structure and Function and White Matter Hyperintensities in Young Adults. *JAMA.* 2018;320(7):665-673. doi:10.1001/jama.2018.11498
46. Ylikoski A, Erkinjuntti T, Raininko R, Sarna S, Sulkava R, Tilvis R. White matter hyperintensities on MRI in the neurologically nondiseased elderly. Analysis of cohorts of consecutive subjects aged 55 to 85 years living at home. *Stroke.* 1995;26(7):1171-1177.
47. Poels MMF, Ikram MA, van der Lugt A, et al. Incidence of cerebral microbleeds in the general population: the Rotterdam Scan Study. *Stroke.* 2011;42(3):656-661. doi:10.1161/STROKEAHA.110.607184
48. Longstreth WTJ, Manolio TA, Arnold A, et al. Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301 elderly people. The Cardiovascular Health Study. *Stroke.* 1996;27(8):1274-1282.
49. Liao D, Cooper L, Cai J, et al. The prevalence and severity of white matter lesions, their relationship with age, ethnicity, gender, and cardiovascular disease risk factors: the ARIC Study. *Neuroepidemiology.* 1997;16(3):149-162. doi:10.1159/000368814

50. Breteler MM, van Swieten JC, Bots ML, et al. Cerebral white matter lesions, vascular risk factors, and cognitive function in a population-based study: the Rotterdam Study. *Neurology*. 1994;44(7):1246-1252.
51. Kuo H-K, Lipsitz LA. Cerebral White Matter Changes and Geriatric Syndromes: Is There a Link? *Journals Gerontol Ser A*. 2004;59(8):M818-M826.
doi:10.1093/gerona/59.8.M818
52. Moran C, Phan TG, Srikanth VK. Cerebral small vessel disease: a review of clinical, radiological, and histopathological phenotypes. *Int J Stroke*. 2012;7(1):36-46.
doi:10.1111/j.1747-4949.2011.00725.x
53. Christophe T, Stéphane L, Stéphanie D. Is Hypertension Associated With an Accelerated Aging of the Brain? *Hypertension*. 2014;63(5):894-903.
doi:10.1161/HYPERTENSIONAHA.113.00147
54. Launer LJ. Epidemiology of white matter lesions. *Top Magn Reson Imaging*. 2004;15(6):365-367.
55. Das AS, Regenhardt RW, Vernooij MW, Blacker D, Charidimou A, Viswanathan A. Asymptomatic Cerebral Small Vessel Disease: Insights from Population-Based Studies. *J stroke*. April 2019. doi:10.5853/jos.2018.03608
56. Verhaaren BFJ, Vernooij MW, de Boer R, et al. High blood pressure and cerebral white matter lesion progression in the general population. *Hypertens (Dallas, Tex 1979)*. 2013;61(6):1354-1359. doi:10.1161/HYPERTENSIONAHA.111.00430
57. van Dijk EJ, Breteler MMB, Schmidt R, et al. The association between blood pressure, hypertension, and cerebral white matter lesions: cardiovascular determinants of

- dementia study. *Hypertens (Dallas, Tex 1979)*. 2004;44(5):625-630.
doi:10.1161/01.HYP.0000145857.98904.20
58. Longstreth WTJ. Brain vascular disease overt and covert. *Stroke*. 2005;36(10):2062-2063.
 59. Jeerakathil T, Wolf PA, Beiser A, et al. Stroke risk profile predicts white matter hyperintensity volume: the Framingham Study. *Stroke*. 2004;35(8):1857-1861.
doi:10.1161/01.STR.0000135226.53499.85
 60. de Leeuw F-E, de Groot JC, Oudkerk M, et al. Hypertension and cerebral white matter lesions in a prospective cohort study. *Brain*. 2002;125(Pt 4):765-772.
 61. Dufouil C, de Kersaint-Gilly A, Besancon V, et al. Longitudinal study of blood pressure and white matter hyperintensities: the EVA MRI Cohort. *Neurology*. 2001;56(7):921-926.
 62. Guo X, Pantoni L, Simoni M, et al. Blood pressure components and changes in relation to white matter lesions: a 32-year prospective population study. *Hypertens (Dallas, Tex 1979)*. 2009;54(1):57-62.
doi:10.1161/HYPERTENSIONAHA.109.129700
 63. Gottesman RF, Coresh J, Catellier DJ, et al. Blood pressure and white-matter disease progression in a biethnic cohort: Atherosclerosis Risk in Communities (ARIC) study. *Stroke*. 2010;41(1):3-8. doi:10.1161/STROKEAHA.109.566992
 64. Mok V, Kim JS. Prevention and Management of Cerebral Small Vessel Disease. *J stroke*. 2015;17(2):111-122. doi:10.5853/jos.2015.17.2.111

65. Schmidt R, Schmidt H, Haybaeck J, et al. Heterogeneity in age-related white matter changes. *Acta Neuropathol.* 2011;122(2):171-185. doi:10.1007/s00401-011-0851-x
66. Pantoni L. Pathophysiology of age-related cerebral white matter changes. *Cerebrovasc Dis.* 2002;13 Suppl 2:7-10. doi:10.1159/000049143
67. Debette S, Seshadri S, Beiser A, et al. Midlife vascular risk factor exposure accelerates structural brain aging and cognitive decline. *Neurology.* 2011;77(5):461-468. doi:10.1212/WNL.0b013e318227b227
68. Pantoni L, Garcia JH. The significance of cerebral white matter abnormalities 100 years after Binswanger's report. A review. *Stroke.* 1995;26(7):1293-1301.
69. Kloppenborg RP, Nederkoorn PJ, Geerlings MI, van den Berg E. Presence and progression of white matter hyperintensities and cognition: a meta-analysis. *Neurology.* 2014;82(23):2127-2138. doi:10.1212/WNL.0000000000000505
70. Arvanitakis Z, Fleischman DA, Arfanakis K, Leurgans SE, Barnes LL, Bennett DA. Association of white matter hyperintensities and gray matter volume with cognition in older individuals without cognitive impairment. *Brain Struct Funct.* 2016;221(4):2135-2146. doi:10.1007/s00429-015-1034-7
71. Iadecola C. The Neurovascular Unit Coming of Age: A Journey through Neurovascular Coupling in Health and Disease. *Neuron.* 2017;96(1):17-42. doi:10.1016/j.neuron.2017.07.030
72. Joutel A, Faraci FM. Cerebral small vessel disease: insights and opportunities from mouse models of collagen IV-related small vessel disease and cerebral autosomal

- dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Stroke*. 2014;45(4):1215-1221. doi:10.1161/STROKEAHA.113.002878
73. Cipolla MJ. Chapter 2: Anatomy and Ultrastructure. In: *The Cerebral Circulation*. San Rafael, CA: Morgan & Claypool Life Sciences; 2009.
 74. Wei EP, Kontos HA. Increased venous pressure causes myogenic constriction of cerebral arterioles during local hyperoxia. *Circ Res*. 1984;55(2):249-252.
 75. McConnell HL, Kersch CN, Woltjer RL, Neuwelt EA. The Translational Significance of the Neurovascular Unit. *J Biol Chem*. 2017;292(3):762-770.
doi:10.1074/jbc.R116.760215
 76. Redzic Z. Molecular biology of the blood-brain and the blood-cerebrospinal fluid barriers: similarities and differences. *Fluids Barriers CNS*. 2011;8(1):3.
doi:10.1186/2045-8118-8-3
 77. Yamazaki Y, Kanekiyo T. Blood-Brain Barrier Dysfunction and the Pathogenesis of Alzheimer's Disease. *Int J Mol Sci*. 2017;18(9). doi:10.3390/ijms18091965
 78. Faraci FM. Protecting against vascular disease in brain. *Am J Physiol Heart Circ Physiol*. 2011;300(5):H1566-82. doi:10.1152/ajpheart.01310.2010
 79. Faraco G, Iadecola C. Hypertension: a harbinger of stroke and dementia. *Hypertens (Dallas, Tex 1979)*. 2013;62(5):810-817.
doi:10.1161/HYPERTENSIONAHA.113.01063
 80. Sommer CJ. Ischemic stroke: experimental models and reality. *Acta Neuropathol*. 2017;133(2):245-261. doi:10.1007/s00401-017-1667-0

81. Kitamura A, Fujita Y, Oishi N, et al. Selective white matter abnormalities in a novel rat model of vascular dementia. *Neurobiol Aging*. 2012;33(5):1012.e25-1012.e35. doi:<https://doi.org/10.1016/j.neurobiolaging.2011.10.033>
82. Xuefang R, Dominic Q, Heng H, et al. Abstract WP417: Novel Animal Model for Cerebral Small Vessel Disease. *Stroke*. 2019;49(Suppl_1):AWP417-AWP417. doi:10.1161/str.49.suppl_1.WP417
83. Fisher CM. Lacunes: Small, deep cerebral infarcts. *Neurology*. 1965;15:774-784. doi:10.1212/wnl.15.8.774
84. Fisher CM. The arterial lesions underlying lacunes. *Acta Neuropathol*. 1969;12(1):1-15. doi:10.1007/BF00685305
85. Fisher CM. Capsular infarcts: the underlying vascular lesions. *Arch Neurol*. 1979;36(2):65-73.
86. Fisher CM. Lacunar strokes and infarcts: a review. *Neurology*. 1982;32(8):871-876. doi:10.1212/wnl.32.8.871
87. Fisher CM. Lacunar infarcts—a review. *Cerebrovasc Dis*. 1991;1(6):311-320.
88. Shi Y, Wardlaw JM. Update on cerebral small vessel disease: a dynamic whole-brain disease. *Stroke Vasc Neurol*. 2016;1(3):83-92. doi:10.1136/svn-2016-000035
89. Khan A, Kasner SE, Lynn MJ, Chimowitz MI. Risk factors and outcome of patients with symptomatic intracranial stenosis presenting with lacunar stroke. *Stroke*. 2012;43(5):1230-1233. doi:10.1161/STROKEAHA.111.641696

90. Wardlaw JM, Doubal FN, Eadie E, Chappell F, Shuler K, Cvorovic V. Little association between intracranial arterial stenosis and lacunar stroke. *Cerebrovasc Dis*. 2011;31(1):12-18. doi:10.1159/000319773
91. Del Bene A, Palumbo V, Lamassa M, Saia V, Piccardi B, Inzitari D. Progressive lacunar stroke: review of mechanisms, prognostic features, and putative treatments. *Int J Stroke*. 2012;7(4):321-329. doi:10.1111/j.1747-4949.2012.00789.x
92. Jackson CA, Hutchison A, Dennis MS, et al. Differing risk factor profiles of ischemic stroke subtypes: evidence for a distinct lacunar arteriopathy? *Stroke*. 2010;41(4):624-629. doi:10.1161/STROKEAHA.109.558809
93. Mead GE, Lewis SC, Wardlaw JM, Dennis MS, Warlow CP. Severe ipsilateral carotid stenosis and middle cerebral artery disease in lacunar ischaemic stroke: innocent bystanders? *J Neurol*. 2002;249(3):266-271.
94. Potter GM, Doubal FN, Jackson CA, Sudlow CLM, Dennis MS, Wardlaw JM. Lack of association of white matter lesions with ipsilateral carotid artery stenosis. *Cerebrovasc Dis*. 2012;33(4):378-384. doi:10.1159/000336762
95. Bailey EL, McCulloch J, Sudlow C, Wardlaw JM. Potential animal models of lacunar stroke: a systematic review. *Stroke*. 2009;40(6):e451-8. doi:10.1161/STROKEAHA.108.528430
96. Macdonald RL, Kowalczyk A, Johns L. Emboli enter penetrating arteries of monkey brain in relation to their size. *Stroke*. 1995;26(7):1241-1247.
97. Lammie GA, Brannan F, Slattery J, Warlow C. Nonhypertensive cerebral small-vessel disease. An autopsy study. *Stroke*. 1997;28(11):2222-2229.

98. Godin O, Tzourio C, Maillard P, Mazoyer B, Dufouil C. Antihypertensive treatment and change in blood pressure are associated with the progression of white matter lesion volumes: the Three-City (3C)-Dijon Magnetic Resonance Imaging Study. *Circulation*. 2011;123(3):266-273. doi:10.1161/CIRCULATIONAHA.110.961052
99. Marcus J, Gardener H, Rundek T, et al. Baseline and longitudinal increases in diastolic blood pressure are associated with greater white matter hyperintensity volume: the northern Manhattan study. *Stroke*. 2011;42(9):2639-2641. doi:10.1161/STROKEAHA.111.617571
100. Gouw AA, van der Flier WM, Fazekas F, et al. Progression of white matter hyperintensities and incidence of new lacunes over a 3-year period: the Leukoaraiosis and Disability study. *Stroke*. 2008;39(5):1414-1420. doi:10.1161/STROKEAHA.107.498535
101. Lammie GA, Brannan F, Wardlaw JM. Incomplete lacunar infarction (Type Ib lacunes). *Acta Neuropathol*. 1998;96(2):163-171.
102. Wardlaw JM, Sandercock PAG, Dennis MS, Starr J. Is breakdown of the blood-brain barrier responsible for lacunar stroke, leukoaraiosis, and dementia? *Stroke*. 2003;34(3):806-812. doi:10.1161/01.STR.0000058480.77236.B3
103. Stevenson SF, Doubal FN, Shuler K, Wardlaw JM. A systematic review of dynamic cerebral and peripheral endothelial function in lacunar stroke versus controls. *Stroke*. 2010;41(6):e434-42. doi:10.1161/STROKEAHA.109.569855

104. Knottnerus ILH, Ten Cate H, Lodder J, Kessels F, van Oostenbrugge RJ. Endothelial dysfunction in lacunar stroke: a systematic review. *Cerebrovasc Dis.* 2009;27(5):519-526. doi:10.1159/000212672
105. Neuwelt E, Abbott NJ, Abrey L, et al. Strategies to advance translational research into brain barriers. *Lancet Neurol.* 2008;7(1):84-96. doi:10.1016/S1474-4422(07)70326-5
106. Rutten-Jacobs LCA, Rost NS. Emerging insights from the genetics of cerebral small-vessel disease. *Ann N Y Acad Sci.* January 2019. doi:10.1111/nyas.13998
107. Emmanuel C, Céline B-M, Valérie D-D, et al. Archetypal Arg169Cys Mutation in NOTCH3 Does Not Drive the Pathogenesis in Cerebral Autosomal Dominant Arteriopathy With Subcortical Infarcts and Leukoencephalopathy via a Loss-of-Function Mechanism. *Stroke.* 2014;45(3):842-849. doi:10.1161/STROKEAHA.113.003339
108. Ehret F, Vogler S, Pojar S, et al. Mouse model of CADASIL reveals novel insights into Notch3 function in adult hippocampal neurogenesis. *Neurobiol Dis.* 2015;75:131-141. doi:https://doi.org/10.1016/j.nbd.2014.12.018
109. Ghosh M, Balbi M, Hellal F, Dichgans M, Lindauer U, Plesnila N. Pericytes are involved in the pathogenesis of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Ann Neurol.* 2015;78(6):887-900. doi:10.1002/ana.24512
110. Joutel A, Monet-Lepretre M, Gosele C, et al. Cerebrovascular dysfunction and microcirculation rarefaction precede white matter lesions in a mouse genetic model of

cerebral ischemic small vessel disease. *J Clin Invest*. 2010;120(2):433-445.

doi:10.1172/JCI39733

111. Ruchoux MM, Domenga V, Brulin P, et al. Transgenic Mice Expressing Mutant Notch3 Develop Vascular Alterations Characteristic of Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy. *Am J Pathol*. 2003;162(1):329-342. doi:10.1016/S0002-9440(10)63824-2
112. Hua L, Wenbo Z, Simone K, B. CR, Brenda L. Notch3 Is Critical for Proper Angiogenesis and Mural Cell Investment. *Circ Res*. 2010;107(7):860-870. doi:10.1161/CIRCRESAHA.110.218271
113. L. HT, Annika K, Liqun H, et al. Notch3 Is Necessary for Blood Vessel Integrity in the Central Nervous System. *Arterioscler Thromb Vasc Biol*. 2015;35(2):409-420. doi:10.1161/ATVBAHA.114.304849
114. Zellner A, Scharrer E, Arzberger T, et al. CADASIL brain vessels show a HTRA1 loss-of-function profile. *Acta Neuropathol*. 2018;136(1):111-125. doi:10.1007/s00401-018-1853-8
115. Donahue CP, Kosik KS. Distribution pattern of Notch3 mutations suggests a gain-of-function mechanism for CADASIL. *Genomics*. 2004;83(1):59-65. doi:10.1016/s0888-7543(03)00206-4
116. Gupta RM, Hadaya J, Trehan A, et al. A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell*. 2017;170(3):522-533.e15. doi:10.1016/j.cell.2017.06.049

117. Polychronopoulos P, Gioldasis G, Ellul J, et al. Family history of stroke in stroke types and subtypes. *J Neurol Sci.* 2002;195(2):117-122.
118. Jerrard-Dunne P, Cloud G, Hassan A, Markus HS. Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke.* 2003;34(6):1364-1369. doi:10.1161/01.STR.0000069723.17984.FD
119. 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.
120. 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. doi:10.1161/01.STR.0000129643.77045.10
121. 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. doi:10.1161/STROKEAHA.109.565390
122. Turner ST, Jack CR, Fornage M, Mosley TH, Boerwinkle E, de Andrade M. Heritability of leukoaraiosis in hypertensive sibships. *Hypertens (Dallas, Tex 1979).* 2004;43(2):483-487. doi:10.1161/01.HYP.0000112303.26158.92
123. Bevan S, Traylor M, Adib-Samii P, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke.* 2012;43(12):3161-3167. doi:10.1161/STROKEAHA.112.665760
124. Devan WJ, Falcone GJ, Anderson CD, et al. Heritability estimates identify a substantial genetic contribution to risk and outcome of intracerebral hemorrhage. *Stroke.* 2013;44(6):1578-1583. doi:10.1161/STROKEAHA.111.000089

125. Traylor M, Tozer DJ, Croall ID, et al. Genetic variation in PLEKHG1 is associated with white matter hyperintensities (n = 11,226). *Neurology*. January 2019;10.1212/WNL.0000000000006952. doi:10.1212/WNL.0000000000006952
126. Yamamoto Y, Craggs L, Baumann M, Kalimo H, Kalaria RN. Review: molecular genetics and pathology of hereditary small vessel diseases of the brain. *Neuropathol Appl Neurobiol*. 2011;37(1):94-113. doi:10.1111/j.1365-2990.2010.01147.x
127. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserre E, Boussier M-G. Cadasil. *Lancet Neurol*. 2009;8(7):643-653. doi:10.1016/S1474-4422(09)70127-9
128. Domenga V, Fardoux P, Lacombe P, et al. Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. *Genes Dev*. 2004;18(22):2730-2735. doi:10.1101/gad.308904
129. Joutel A, Andreux F, Gaulis S, et al. The ectodomain of the Notch3 receptor accumulates within the cerebrovasculature of CADASIL patients. *J Clin Invest*. 2000;105(5):597-605. doi:10.1172/JCI8047
130. Rutten JW, Haan J, Terwindt GM, van Duinen SG, Boon EMJ, Lesnik Oberstein SAJ. Interpretation of NOTCH3 mutations in the diagnosis of CADASIL. *Expert Rev Mol Diagn*. 2014;14(5):593-603. doi:10.1586/14737159.2014.922880
131. Joutel A, Corpechot C, Ducros A, et al. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature*. 1996;383(6602):707-710. doi:10.1038/383707a0
132. Beaufort N, Scharrer E, Kremmer E, et al. Cerebral small vessel disease-related protease HtrA1 processes latent TGF-beta binding protein 1 and facilitates TGF-beta

- signaling. *Proc Natl Acad Sci U S A*. 2014;111(46):16496-16501.
doi:10.1073/pnas.1418087111
133. Oka C, Tsujimoto R, Kajikawa M, et al. HtrA1 serine protease inhibits signaling mediated by Tgfbeta family proteins. *Development*. 2004;131(5):1041-1053.
doi:10.1242/dev.00999
 134. Joutel A, Haddad I, Ratelade J, Nelson MT. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *J Cereb Blood Flow Metab*. 2016;36(1):143-157. doi:10.1038/jcbfm.2015.62
 135. Kuo DS, Labelle-Dumais C, Gould DB. COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets. *Hum Mol Genet*. 2012;21(R1):R97-110. doi:10.1093/hmg/ddc346
 136. Ryan AK, Blumberg B, Rodriguez-Esteban C, et al. Pitx2 determines left-right asymmetry of internal organs in vertebrates. *Nature*. 1998;394(6693):545-551.
doi:10.1038/29004
 137. Seo S, Singh HP, Lacal PM, et al. Forkhead box transcription factor FoxC1 preserves corneal transparency by regulating vascular growth. *Proc Natl Acad Sci U S A*. 2012;109(6):2015-2020. doi:10.1073/pnas.1109540109
 138. Siegenthaler JA, Choe Y, Patterson KP, et al. Foxc1 is required by pericytes during fetal brain angiogenesis. *Biol Open*. 2013;2(7):647-659. doi:10.1242/bio.20135009
 139. Westendorp WF, Nederkoorn PJ, Aksentijevich I, Hak AE, Lichtenbelt KD, Braun KPJ. Unexplained early-onset lacunar stroke and inflammatory skin lesions: Consider

ADA2 deficiency. *Neurology*. 2015;84(20):2092-2093.

doi:10.1212/WNL.0000000000001581

140. Zhou Q, Yang D, Ombrello AK, et al. Early-onset stroke and vasculopathy associated with mutations in ADA2. *N Engl J Med*. 2014;370(10):911-920.

doi:10.1056/NEJMoa1307361

141. Kavanagh D, Spitzer D, Kothari PH, et al. New roles for the major human 3'-5' exonuclease TREX1 in human disease. *Cell Cycle*. 2008;7(12):1718-1725.

doi:10.4161/cc.7.12.6162

142. Ophoff RA, DeYoung J, Service SK, et al. Hereditary vascular retinopathy, cerebretinal vasculopathy, and hereditary endotheliopathy with retinopathy, nephropathy, and stroke map to a single locus on chromosome 3p21.1-p21.3. *Am J Hum Genet*. 2001;69(2):447-453. doi:10.1086/321975

143. Bugiani M, Kevelam SH, Bakels HS, et al. Cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *Neurology*. 2016;87(17):1777-1786.

doi:10.1212/WNL.0000000000003251

144. Caciotti A, Catarzi S, Tonin R, et al. Galactosialidosis: review and analysis of CTSA gene mutations. *Orphanet J Rare Dis*. 2013;8:114. doi:10.1186/1750-1172-8-114

145. DeStefano AL, Atwood LD, Massaro JM, et al. Genome-wide scan for white matter hyperintensity: the Framingham Heart Study. *Stroke*. 2006;37(1):77-81.

doi:10.1161/01.STR.0000196987.68770.b3

146. Seshadri S, DeStefano AL, Au R, et al. Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the

- Framingham Study. BMC Med Genet. 2007;8 Suppl 1:S15. doi:10.1186/1471-2350-8-S1-S15
147. Turner ST, Fornage M, Jack CRJ, et al. Genomic susceptibility Loci for brain atrophy, ventricular volume, and leukoaraiosis in hypertensive sibships. Arch Neurol. 2009;66(7):847-857. doi:10.1001/archneurol.2009.110
148. Kochunov P, Glahn D, Lancaster J, et al. Whole brain and regional hyperintense white matter volume and blood pressure: overlap of genetic loci produced by bivariate, whole-genome linkage analyses. Stroke. 2010;41(10):2137-2142. doi:10.1161/STROKEAHA.110.590943
149. Paternoster L, Chen W, Sudlow CLM. Genetic determinants of white matter hyperintensities on brain scans: a systematic assessment of 19 candidate gene polymorphisms in 46 studies in 19,000 subjects. Stroke. 2009;40(6):2020-2026. doi:10.1161/STROKEAHA.108.542050
150. Fornage M, DeBette S, Bis JC, et al. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. Ann Neurol. 2011;69(6):928-939. doi:10.1002/ana.22403
151. Freudenberger P, Schmidt R, Schmidt H. Genetics of age-related white matter lesions from linkage to genome wide association studies. J Neurol Sci. 2012;322(1-2):82-86. doi:10.1016/j.jns.2012.06.016
152. 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. doi:10.1161/CIRCGENETICS.114.000858

153. Traylor M, Zhang CR, Adib-Samii P, et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology*. 2016;86(2):146-153. doi:10.1212/WNL.0000000000002263
154. 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. doi:10.1161/STROKEAHA.118.020689
155. Rannikmae K, Davies G, Thomson PA, et al. Common variation in COL4A1/COL4A2 is associated with sporadic cerebral small vessel disease. *Neurology*. 2015;84(9):918-926. doi:10.1212/WNL.0000000000001309
156. Rannikmae K, Sivakumaran V, Millar H, et al. COL4A2 is associated with lacunar ischemic stroke and deep ICH: Meta-analyses among 21,500 cases and 40,600 controls. *Neurology*. 2017;89(17):1829-1839. doi:10.1212/WNL.0000000000004560
157. Abiko H, Fujiwara S, Ohashi K, et al. Rho guanine nucleotide exchange factors involved in cyclic-stretch-induced reorientation of vascular endothelial cells. *J Cell Sci*. 2015;128(9):1683-1695. doi:10.1242/jcs.157503
158. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-1073. doi:10.1038/nature09534
159. Marth GT, Yu F, Indap AR, et al. The functional spectrum of low-frequency coding variation. *Genome Biol*. 2011;12(9):R84. doi:10.1186/gb-2011-12-9-r84

160. Marian AJ. Elements of “missing heritability”. *Curr Opin Cardiol*. 2012;27(3):197-201. doi:10.1097/HCO.0b013e328352707d
161. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33(3S):245-254. doi:10.1038/ng1089
162. Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, Fire A. Amplification of siRNA in *Caenorhabditis elegans* generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet*. 2012;44(2):157-164. doi:10.1038/ng.1039
163. Handel AE, Ebers GC, Ramagopalan S V. Epigenetics: molecular mechanisms and implications for disease. *Trends Mol Med*. 2010;16(1):7-16. doi:10.1016/j.molmed.2009.11.003
164. Zhong J, Agha G, Baccarelli AA. The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies. *Circ Res*. 2016;118(1):119-131. doi:10.1161/CIRCRESAHA.115.305206
165. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*. 1983;301(5895):89-92.
166. Gama-Sosa MA, Slagel VA, Trewyn RW, et al. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res*. 1983;11(19):6883-6894. doi:10.1093/nar/11.19.6883
167. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 2002;3(6):415-428. doi:10.1038/nrg816

168. Esteller M, Herman JG. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *J Pathol.* 2002;196(1):1-7.
doi:10.1002/path.1024
169. Gokul G, Khosla S. DNA methylation and cancer. *Subcell Biochem.* 2013;61(35):597-625. doi:10.1007/978-94-007-4525-4_26
170. Baccarelli A, Wright R, Bollati V, et al. Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology.* 2010;21(6):819-828.
doi:10.1097/EDE.0b013e3181f20457
171. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. *PLoS One.* 2010;5(3):e9692.
doi:10.1371/journal.pone.0009692
172. Movassagh M, Vujic A, Foo R. Genome-wide DNA methylation in human heart failure. *Epigenomics.* 2011;3(1):103-109. doi:10.2217/epi.10.70
173. Talens RP, Jukema JW, Trompet S, et al. Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction. *Int J Epidemiol.* 2012;41(1):106-115. doi:10.1093/ije/dyr153
174. Flanagan JM. Epigenome-wide association studies (EWAS): past, present, and future. *Methods Mol Biol.* 2015;1238:51-63. doi:10.1007/978-1-4939-1804-1_3
175. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet.* 2013;14(8):585-594. doi:10.1038/nrg3405

176. Relton CL, Smith GD. Epigenetic epidemiology of common complex disease: Prospects for prediction, prevention, and treatment. *PLoS Med.* 2010;7(10):e1000356. doi:10.1371/journal.pmed.1000356
177. Mitchell CJ, Getnet D, Kim MS, et al. A multi-omic analysis of human naïve CD4⁺ T cells. *BMC Syst Biol.* 2015;9(1):75. doi:10.1186/s12918-015-0225-4
178. He DX, Gu F, Gao F, et al. Genome-wide profiles of methylation, microRNAs, and gene expression in chemoresistant breast cancer. *Sci Rep.* 2016;6:24706. doi:10.1038/srep24706
179. Lin S, Yin YA, Jiang X, Sahni N, Yi S. Multi-OMICs and Genome Editing Perspectives on Liver Cancer Signaling Networks. *Biomed Res Int.* 2016;2016:1-14. doi:10.1155/2016/6186281
180. Arneson D, Shu L, Tsai B, Barrere-Cain R, Sun C, Yang X. Multidimensional Integrative Genomics Approaches to Dissecting Cardiovascular Disease. *Front Cardiovasc Med.* 2017;4:8. doi:10.3389/fcvm.2017.00008
181. Shimada M, Miyagawa T, Toyoda H, Tokunaga K, Honda M. Epigenome-wide association study of DNA methylation in narcolepsy: an integrated genetic and epigenetic approach. *Sleep.* 2018;41(4). doi:10.1093/sleep/zsy019
182. Humphries CE, Kohli MA, Nathanson L, et al. Integrated whole transcriptome and DNA methylation analysis identifies gene networks specific to late-onset Alzheimer's disease. *J Alzheimers Dis.* 2015;44(3):977-987. doi:10.3233/JAD-141989

183. Hannon E, Dempster E, Viana J, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol.* 2016;17(1):176. doi:10.1186/s13059-016-1041-x
184. Grayson DR, Guidotti A. Merging data from genetic and epigenetic approaches to better understand autistic spectrum disorder. *Epigenomics.* 2016;8(1):85-104. doi:10.2217/epi.15.92
185. Leary MC, Saver JL. Annual incidence of first silent stroke in the United States: a preliminary estimate. *Cerebrovasc Dis.* 2003;16(3):280-285. doi:10.1159/000071128
186. Konig IR, Fuchs O, Hansen G, von Mutius E, Kopp M V. What is precision medicine? *Eur Respir J.* 2017;50(4). doi:10.1183/13993003.00391-2017
187. B. GL, Robert A, J. AM, et al. Primary Prevention of Ischemic Stroke. *Stroke.* 2006;37(6):1583-1633. doi:10.1161/01.STR.0000223048.70103.F1
188. Dufouil C, Chalmers J, Coskun O, et al. Effects of blood pressure lowering on cerebral white matter hyperintensities in patients with stroke: the PROGRESS (Perindopril Protection Against Recurrent Stroke Study) Magnetic Resonance Imaging Substudy. *Circulation.* 2005;112(11):1644-1650. doi:10.1161/CIRCULATIONAHA.104.501163
189. Weber R, Weimar C, Blatchford J, et al. Telmisartan on top of antihypertensive treatment does not prevent progression of cerebral white matter lesions in the prevention regimen for effectively avoiding second strokes (PRoFESS) MRI substudy. *Stroke.* 2012;43(9):2336-2342. doi:10.1161/STROKEAHA.111.648576
190. Kjeldsen SE, Narkiewicz K, Burnier M, Oparil S. Intensive blood pressure lowering prevents mild cognitive impairment and possible dementia and slows development of

- white matter lesions in brain: the SPRINT Memory and Cognition IN Decreased Hypertension (SPRINT MIND) study. *Blood Press.* 2018;27(5):247-248.
doi:10.1080/08037051.2018.1507621
191. 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.
doi:10.1212/01.WNL.0000161844.00797.73
 192. Amarenco P, Benavente O, Goldstein LB, et al. Results of the Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) trial by stroke subtypes. *Stroke.* 2009;40(4):1405-1409. doi:10.1161/STROKEAHA.108.534107
 193. Muller M, Sigurdsson S, Kjartansson O, et al. Joint effect of mid- and late-life blood pressure on the brain: the AGES-Reykjavik study. *Neurology.* 2014;82(24):2187-2195. doi:10.1212/WNL.0000000000000517
 194. Ribe AR, Vestergaard CH, Vestergaard M, et al. Statins and Risk of Intracerebral Haemorrhage in a Stroke-Free Population: A Nationwide Danish Propensity Score Matched Cohort Study. *EClinicalMedicine.* 2019.
 195. Benavente OR, Hart RG, McClure LA, Szychowski JM, Coffey CS, Pearce LA. Effects of clopidogrel added to aspirin in patients with recent lacunar stroke. *N Engl J Med.* 2012;367(9):817-825. doi:10.1056/NEJMoa1204133
 196. Neumann-Haefelin T, Hoelig S, Berkefeld J, et al. Leukoaraiosis is a risk factor for symptomatic intracerebral hemorrhage after thrombolysis for acute stroke. *Stroke.* 2006;37(10):2463-2466. doi:10.1161/01.STR.0000239321.53203.ea

197. Palumbo V, Boulanger JM, Hill MD, Inzitari D, Buchan AM. Leukoaraiosis and intracerebral hemorrhage after thrombolysis in acute stroke. *Neurology*. 2007;68(13):1020-1024. doi:10.1212/01.wnl.0000257817.29883.48
198. Demchuk AM, Khan F, Hill MD, et al. Importance of leukoaraiosis on CT for tissue plasminogen activator decision making: evaluation of the NINDS rt-PA Stroke Study. *Cerebrovasc Dis*. 2008;26(2):120-125. doi:10.1159/000139658
199. Aries MJH, Uyttenboogaart M, Vroomen PC, De Keyser J, Luijckx GJ. tPA treatment for acute ischaemic stroke in patients with leukoaraiosis. *Eur J Neurol*. 2010;17(6):866-870. doi:10.1111/j.1468-1331.2010.02963.x
200. Cai W, Zhang K, Li P, et al. Dysfunction of the neurovascular unit in ischemic stroke and neurodegenerative diseases: An aging effect. *Ageing Res Rev*. 2017;34:77-87. doi:10.1016/j.arr.2016.09.006
201. Berger SL, Kouzarides T, Shiekhhattar R, Shilatifard A. An operational definition of epigenetics. *Genes Dev*. 2009;23(7):781-783. doi:10.1101/gad.1787609
202. Oksala NKJ, Oksala A, Pohjasvaara T, et al. Age related white matter changes predict stroke death in long term follow-up. *J Neurol Neurosurg Psychiatry*. 2009;80(7):762-766. doi:10.1136/jnnp.2008.154104
203. Marttila S, Kananen L, Hayrynen S, et al. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*. 2015;16:179. doi:10.1186/s12864-015-1381-z

204. Chouliaras L, Pishva E, Haapakoski R, et al. Peripheral DNA methylation, cognitive decline and brain aging: pilot findings from the Whitehall II imaging study. *Epigenomics*. 2018;10(5):585-595. doi:10.2217/epi-2017-0132
205. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86. doi:10.1186/1471-2105-13-86
206. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340
207. 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. doi:10.1093/aje/kwp262
208. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57(1):289-300.
209. Chen Y, 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. doi:10.4161/epi.23470
210. 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. doi:10.1093/bioinformatics/bts545

211. 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. doi:10.1186/1756-8935-8-6
212. Šidák Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc*. 1967;62(318):626-633.
213. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797. doi:10.1101/gr.137323.112
214. Breeze CE, Paul DS, van Dongen J, et al. eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. *Cell Rep*. 2016;17(8):2137-2150. doi:10.1016/j.celrep.2016.10.059
215. Breeze CE, Reynolds AP, van Dongen J, et al. eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics*. 2019;35(22):4767-4769. doi:10.1093/bioinformatics/btz456
216. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29(4):1165-1188.
217. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120
218. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5

219. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260
220. Kutmon M, Riutta A, Nunes N, et al. WikiPathways: Capturing the full diversity of pathway knowledge. *Nucleic Acids Res*. 2016;44(D1):D488-D494. doi:10.1093/nar/gkv1024
221. 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. doi:10.1093/brain/awh553
222. Marioni RE, McRae AF, Bressler J, et al. Meta-analysis of epigenome-wide association studies of cognitive abilities. *Mol Psychiatry*. 2018;23(11):2133-2144. doi:10.1038/s41380-017-0008-y
223. Ray D, Boehnke M. Methods for meta-analysis of multiple traits using GWAS summary statistics. *Genet Epidemiol*. 2018;42(2):134-145. doi:10.1002/gepi.22105
224. McRae AF, Powell JE, Henders AK, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol*. 2014;15(5):R73. doi:10.1186/gb-2014-15-5-r73
225. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol*. 2016;17:61. doi:10.1186/s13059-016-0926-z

226. 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. doi:10.1038/s41467-019-12228-z
227. Timpson NJ, Nordestgaard BG, Harbord RM, et al. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. *Int J Obes (Lond).* 2011;35(2):300-308. doi:10.1038/ijo.2010.137
228. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* 2018;7. doi:10.7554/eLife.34408
229. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol.* 2012;41(1):161-176. doi:10.1093/ije/dyr233
230. Arneson D, Bhattacharya A, Shu L, Makinen V-P, Yang X. Mergeomics: a web server for identifying pathological pathways, networks, and key regulators via multidimensional data integration. *BMC Genomics.* 2016;17(1):722. doi:10.1186/s12864-016-3057-8
231. 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. doi:10.1186/s12864-016-3198-9
232. Lin H, Satizabal C, Xie Z, et al. Whole blood gene expression and white matter Hyperintensities. *Mol Neurodegener.* 2017;12(1):67. doi:10.1186/s13024-017-0209-5
233. Liu Y, Ding J, Reynolds LM, et al. Methylomics of gene expression in human monocytes. *Hum Mol Genet.* 2013;22(24):5065-5074. doi:10.1093/hmg/ddt356

234. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 1999;27(1):29-34. doi:10.1093/nar/27.1.29
235. Croft D, Mundo AF, Haw R, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* 2014;42(Database issue):D472-7. doi:10.1093/nar/gkt1102
236. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. Williams SM, ed. *PLoS Genet.* 2014;10(5):e1004383. doi:10.1371/journal.pgen.1004383
237. Joehanes R, Zhang X, Huan T, et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol.* 2017;18(1):16. doi:10.1186/s13059-016-1142-6
238. Gibson G. Human genetics. GTEx detects genetic effects. *Science.* 2015;348(6235):640-641. doi:10.1126/science.aab3002
239. Giambartolomei C, Zhenli Liu J, Zhang W, et al. A Bayesian Framework for Multiple Trait Colocalization from Summary Association Statistics. *bioRxiv.* January 2018:155481. doi:10.1101/155481
240. Rosa AP, Jacques CED, de Souza LO, et al. Neonatal hyperglycemia induces oxidative stress in the rat brain: the role of pentose phosphate pathway enzymes and NADPH oxidase. *Mol Cell Biochem.* 2015;403(1-2):159-167. doi:10.1007/s11010-015-2346-x

241. 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. doi:10.1038/nbt.2247
242. Greene C, Hanley N, Campbell M. Claudin-5: gatekeeper of neurological function. *Fluids Barriers CNS.* 2019;16(1):3. doi:10.1186/s12987-019-0123-z
243. 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. doi:10.1016/j.nbd.2018.02.012
244. 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. doi:10.1007/s00018-019-03030-7
245. Uemura MT, Ihara M, Maki T, et al. Pericyte-derived bone morphogenetic protein 4 underlies white matter damage after chronic hypoperfusion. *Brain Pathol.* 2018;28(4):521-535. doi:10.1111/bpa.12523
246. Harnisch K, Teuber-Hanselmann S, Macha N, et al. Myelination in Multiple Sclerosis Lesions Is Associated with Regulation of Bone Morphogenetic Protein 4 and Its Antagonist Noggin. *Int J Mol Sci.* 2019;20(1). doi:10.3390/ijms20010154
247. Takai H, Masuda K, Sato T, et al. 5-Hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-methylosome complex. *Cell Rep.* 2014;9(1):48-60. doi:10.1016/j.celrep.2014.08.071
248. 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. doi:10.1074/jbc.M115.684514
249. Tang J, Frankel A, Cook RJ, et al. PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. *J Biol Chem.* 2000;275(11):7723-7730. doi:10.1074/jbc.275.11.7723
 250. Vallance P, Leiper J. Cardiovascular biology of the asymmetric dimethylarginine:dimethylarginine dimethylaminohydrolase pathway. *Arterioscler Thromb Vasc Biol.* 2004;24(6):1023-1030. doi:10.1161/01.ATV.0000128897.54893.26
 251. Boger RH, Sydow K, Borlak J, et al. LDL cholesterol upregulates synthesis of asymmetrical dimethylarginine in human endothelial cells: involvement of S-adenosylmethionine-dependent methyltransferases. *Circ Res.* 2000;87(2):99-105. doi:10.1161/01.res.87.2.99
 252. Cardounel AJ, Cui H, Samouilov A, et al. Evidence for the pathophysiological role of endogenous methylarginines in regulation of endothelial NO production and vascular function. *J Biol Chem.* 2007;282(2):879-887. doi:10.1074/jbc.M603606200
 253. Sun L, Hui A-M, Su Q, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell.* 2006;9(4):287-300. doi:10.1016/j.ccr.2006.03.003
 254. Dejana E. Endothelial cell-cell junctions: happy together. *Nat Rev Mol Cell Biol.* 2004;5(4):261-270. doi:10.1038/nrm1357

255. Feng D, Nagy JA, Pyne K, Dvorak HF, Dvorak AM. Neutrophils emigrate from venules by a transendothelial cell pathway in response to FMLP. *J Exp Med*. 1998;187(6):903-915. doi:10.1084/jem.187.6.903
256. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*. 2017;49(7):1126-1132. doi:10.1038/ng.3892
257. Altschuler L, Wook J-O, Gurari D, Chebath J, Revel M. Involvement of Receptor-Bound Protein Methyltransferase PRMT1 in Antiviral and Antiproliferative Effects of Type I Interferons. *J Interf Cytokine Res*. 1999;19(2):189-195. doi:10.1089/107999099314333
258. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27
259. Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48(D1):D498-D503. doi:10.1093/nar/gkz1031
260. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. 2019;47(D1):D941-D947. doi:10.1093/nar/gky1015
261. Sivan G, Martin SE, Myers TG, et al. Human genome-wide RNAi screen reveals a role for nuclear pore proteins in poxvirus morphogenesis. *Proc Natl Acad Sci U S A*. 2013;110(9):3519-3524. doi:10.1073/pnas.1300708110
262. Schmidt EE, Pelz O, Buhlmann S, Kerr G, Horn T, Boutros M. GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update. *Nucleic Acids Res*. 2013;41(Database issue):D1021-6. doi:10.1093/nar/gks1170

263. Pletscher-Frankild S, Pallejà A, Tsafou K, Binder JX, Jensen LJ. DISEASES: Text mining and data integration of disease–gene associations. *Methods*. 2015;74:83-89. doi:<https://doi.org/10.1016/j.ymeth.2014.11.020>
264. Astle WJ, Elding H, Jiang T, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*. 2016;167(5):1415-1429.e19. doi:[10.1016/j.cell.2016.10.042](https://doi.org/10.1016/j.cell.2016.10.042)
265. Kichaev G, Bhatia G, Loh P-R, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet*. 2019;104(1):65-75. doi:[10.1016/j.ajhg.2018.11.008](https://doi.org/10.1016/j.ajhg.2018.11.008)
266. Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet*. 2016;48(10):1171-1184. doi:[10.1038/ng.3667](https://doi.org/10.1038/ng.3667)
267. Hoffmann TJ, Ehret GB, Nandakumar P, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet*. 2017;49(1):54-64. doi:[10.1038/ng.3715](https://doi.org/10.1038/ng.3715)
268. 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](https://doi.org/10.1038/s41588-018-0303-9)
269. Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*. 2017;541(7635):81-86. doi:[10.1038/nature20784](https://doi.org/10.1038/nature20784)

270. Elliott HR, Tillin T, McArdle WL, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics*. 2014;6(1):4. doi:10.1186/1868-7083-6-4
271. 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. doi:10.1016/j.ajhg.2017.09.028
272. Pan J, Liu S, Farkas M, et al. Serum molecular signature for proliferative diabetic retinopathy in Saudi patients with type 2 diabetes. *Mol Vis*. 2016;22:636-645. <https://www.ncbi.nlm.nih.gov/pubmed/27307695>.
273. Dave JM, Bayless KJ. Vimentin as an integral regulator of cell adhesion and endothelial sprouting. *Microcirculation*. 2014;21(4):333-344. doi:10.1111/micc.12111
274. Rutten-Jacobs LCA, Tozer DJ, Duering M, et al. Genetic Study of White Matter Integrity in UK Biobank (N=8448) and the Overlap With Stroke, Depression, and Dementia. *Stroke*. 2018;49(6):1340-1347. doi:10.1161/STROKEAHA.118.020811
275. Hüls A, Robins C, Conneely KN, et al. Brain DNA Methylation Patterns in CLDN5 Associated With Cognitive Decline. *bioRxiv*. January 2019:857953. doi:10.1101/857953
276. 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. doi:10.1080/15592294.2015.1100786

277. McGrath E, Himali J, Levy D, et al. Circulating IGFBP-2: a novel biomarker for incident dementia. *Ann Clin Transl Neurol.* 2019;6. doi:10.1002/acn3.50854
278. Pal S, Tyler JK. Epigenetics and aging. *Sci Adv.* 2016;2(7):e1600584. doi:10.1126/sciadv.1600584
279. Burbach BJ, Medeiros RB, Mueller KL, Shimizu Y. T-cell receptor signaling to integrins. *Immunol Rev.* 2007;218:65-81. doi:10.1111/j.1600-065X.2007.00527.x
280. Crooks C V, Cross ML, Wall CR. Age-related differences in integrin expression in peripheral blood lymphocytes. *Immun Ageing.* 2010;7:5. doi:10.1186/1742-4933-7-5
281. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology.* 2010;74(3):201-209. doi:10.1212/WNL.0b013e3181cb3e25
282. Khachaturian ZS. Perspective on the Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement.* 2010;6(3):199-201. doi:10.1016/j.jalz.2010.04.002
283. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* 1989;129(4):687-702.
284. 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. doi:10.1016/0895-4356(88)90080-7
285. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991;1(3):263-276. doi:10.1016/1047-2797(91)90005-w

286. 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
287. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med (Baltim)*. 1975;4(4):518-525. doi:10.1016/0091-7435(75)90037-7
288. 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.
289. 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. doi:10.1016/j.amjmed.2003.12.032
290. 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.
291. 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):28.
292. Ikram MA, van der Lugt A, Niessen WJ, et al. The Rotterdam Scan Study: design and update up to 2012. *Eur J Epidemiol*. 2011;26(10):811-824.

293. 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
294. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol.* 2017;32(9):807-850.
doi:10.1007/s10654-017-0321-4
295. Volzke H, Alte D, Schmidt CO, et al. Cohort profile: the study of health in Pomerania. *Int J Epidemiol.* 2011;40(2):294-307. doi:10.1093/ije/dyp394
296. Jack Jr. CR, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging.* 2008;27(4):685-691.
doi:10.1002/jmri.21049
297. Wolz R, Julkunen V, Koikkalainen J, et al. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. *PLoS One.* 2011;6(10):e25446.
doi:10.1371/journal.pone.0025446
298. DeCarli C, Murphy DG, Teichberg D, Campbell G, Sobering GS. Local histogram correction of MRI spatially dependent image pixel intensity nonuniformity. *J Magn Reson Imaging.* 1996;6(3):519-528. doi:10.1002/jmri.1880060316
299. Tosto G, Zimmerman ME, Hamilton JL, Carmichael OT, Brickman AM. The effect of white matter hyperintensities on neurodegeneration in mild cognitive impairment. *Alzheimers Dement.* 2015;11(12):1510-1519. doi:10.1016/j.jalz.2015.05.014

- 300. Bryan RN, Manolio TA, Schertz LD, et al. A method for using MR to evaluate the effects of cardiovascular disease on the brain: the cardiovascular health study. *AJNR Am J Neuroradiol.* 1994;15(9):1625-1633.
- 301. Lao Z, Shen D, Liu D, et al. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. *Acad Radiol.* 2008;15(3):300-313. doi:10.1016/j.acra.2007.10.012
- 302. Jack CRJ, O'Brien PC, Rettman DW, et al. FLAIR histogram segmentation for measurement of leukoaraiosis volume. *J Magn Reson Imaging.* 2001;14(6):668-676. doi:10.1002/jmri.10011
- 303. Vrooman HA, Cocosco CA, van der Lijn F, et al. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage.* 2007;37(1):71-81. doi:10.1016/j.neuroimage.2007.05.018
- 304. Hegenscheid K, Kuhn JP, Volzke H, Biffar R, Hosten N, Puls R. Whole-body magnetic resonance imaging of healthy volunteers: pilot study results from the population-based SHIP study. *Rofo.* 2009;181(8):748-759. doi:10.1055/s-0028-1109510
- 305. Demerath EW, Guan W, Grove ML, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet.* 2015;24(15):4464-4479. doi:10.1093/hmg/ddv161

- 306. Shah S, McRae AF, Marioni RE, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* 2014;24(11):1725-1733.
doi:10.1101/gr.176933.114
- 307. Joehanes R, Ying S, Huan T, et al. Gene expression signatures of coronary heart disease. *Arterioscler Thromb Vasc Biol.* 2013;33(6):1418-1426.
doi:10.1161/ATVBAHA.112.301169
- 308. Westra H-J, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013;45(10):1238-1243.
doi:10.1038/ng.2756
- 309. Schurmann C, Heim K, Schillert A, et al. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. *PLoS One.* 2012;7(12):e50938.
doi:10.1371/journal.pone.0050938