Exome chip analysis identifies low-frequency and rare variants in MRPL38 for white matter hyperintensities on brain MRI

Exome chip analysis identifies low-frequency and rare variants in *MRPL38* for white matter hyperintensities on brain MRI

Xueqiu Jian, PhD[1]; Claudia L. Satizabal, PhD[2]; Albert V. Smith, PhD[3]; Katharina Wittfeld, PhD[4]; Joshua C. Bis, PhD[5]; Jennifer A. Smith, PhD[6]; Fang-Chi Hsu, PhD[7]; Kwangsik Nho, PhD[8]; Edith Hofer, PhD[9]; Saskia P. Hagenaars, PhD[10]; Paul A. Nyquist, MD[11]; Aniket Mishra, PhD[12]; Hieab H.H. Adams, PhD[13]; Shuo Li, MS[14]; Alexander Teumer, PhD[15]; Wei Zhao, PhD[6]; Barry I. Freedman, MD[16]; Yasaman Saba, MS[17]; Lisa R. Yanek, MPH[18]; Ganesh Chauhan, PhD[12]; Mark A. van Buchem, MD, PhD[19]; Mary Cushman, MD[20]; Natalie A. Royle, PhD[10]; R. Nick Bryan, MD, PhD[21]; Wiro J. Niessen, PhD[22]; Beverly G. Windham, MD[23]; Anita L. DeStefano, PhD[14]; Mohamad Habes, PhD[24]; Susan R. Heckbert, MD, PhD[5]; Nicholette D. Palmer, PhD[25]; Cora E. Lewis, MD[26]; Gudny Eiriksdottir, MS[3]; Pauline Maillard, PhD[27]; Rasika A. Mathias, ScD[18]; Georg Homuth, PhD[28]; Maria del C. Valdés-Hernández, PhD[10]; Jasmin Divers, PhD[7]; Alexa S. Beiser, PhD[14]; Sönke Langner, MD[29]; Kenneth M. Rice, PhD[30]; Mark E. Bastin, PhD[10]; Qiong Yang, PhD[14]; Joseph A. Maldjian, MD[31]; John M. Starr, PhD[10]; Stephen Sidney, MD[32]; Shannon L. Risacher, PhD[8]; André G. Uitterlinden, PhD[33]; Vilmundur G. Gudnason, MD, PhD[3]; Matthias Nauck, MD[34]; Jerome I. Rotter, MD[35]; Pamela J. Schreiner, PhD[36]; Eric Boerwinkle, PhD[37]; Cornelia M. van Duijn, PhD[13]; Bernard Mazoyer, MD, PhD[38]; Bettina von Sarnowski, MD[39]; Rebecca F. Gottesman, MD, PhD[11]; Daniel Levy, MD[40]; Sigurdur Sigurdsson, MS[3]; Meike W. Vernooij, MD, PhD[41]; Stephen T. Turner, MD[42]; Reinhold Schmidt, MD[9]; Joanna M. Wardlaw, PhD[10]; Bruce M. Psaty, MD, PhD[5]; Thomas H. Mosley, PhD[23]; Charles S. DeCarli, MD[27]; Andrew J. Saykin, PsyD[8]; Donald W. Bowden, PhD[25]; Diane M. Becker, ScD[18]; Ian J. Deary, PhD[10]; Helena Schmidt, MD, PhD[17]; Sharon L.R. Kardia, PhD[6]; M. Arfan Ikram, MD, PhD[43]; Stéphanie Debette, MD, PhD[12]; Hans J. Grabe, MD[44]; W.T. Longstreth, Jr., MD[45]; Sudha Seshadri, MD[2]; Lenore J. Launer, PhD[46]; and Myriam Fornage, PhD[1] for the neuroCHARGE Working Group

1. Institute of Molecular Medicine, McGovern Medical School at The University of Texas Health Science Center at Houston, Houston, TX
2. Department of Neurology, Boston University School of Medicine, Boston, MA
3. Icelandic Heart Association, Kópavogur, Iceland
4. German Center for Neurodegenerative Diseases, Site Rostock/Greifswald, Greifswald, Germany
5. Cardiovascular Health Research Unit, University of Washington, Seattle, WA
6. Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI
7. Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC
8. Center for Neuroimaging, Indiana University School of Medicine, Indianapolis, IN
9. Department of Neurology, Medical University of Graz, Graz, Austria
10. Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, United Kingdom
11. Department of Neurology and Neurosurgery, Johns Hopkins School of Medicine, Baltimore, MD

12. Bordeaux Population Health Research Centre U1219, Inserm-University of Bordeaux, Bordeaux, France
13. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, Netherlands
14. Department of Biostatistics, Boston University School of Public Health, Boston, MA
15. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
16. Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC
17. Institute of Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria
18. Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD
19. Department of Radiology, Leiden University Medical Center, Leiden, Netherlands
20. Department of Medicine, The University of Vermont Larner College of Medicine, Burlington, VT
21. Department of Diagnostic Medicine, Dell Medical School at The University of Texas at Austin, Austin, TX
22. Departments of Radiology and Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands
23. Department of Medicine, The University of Mississippi School of Medicine, Jackson, MS
24. Department of Radiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
25. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC
26. Department of Epidemiology, The University of Alabama at Birmingham School of Public Health, Birmingham, AL
27. Department of Neurology, UC Davis School of Medicine, Davis, CA
28. Institute of Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany
29. Institute for Diagnostic Radiology and Neuroradiology, University Medicine Greifswald, Greifswald, Germany
30. Department of Biostatistics, University of Washington School of Public Health, Seattle, WA
31. Department of Radiology, The University of Texas Southwestern Medical Center, Dallas, TX
32. Division of Research, Kaiser Permanente Northern California, Oakland, CA
33. Departments of Internal Medicine, Epidemiology and Clinical Chemistry, Erasmus University Medical Center, Rotterdam, Netherlands
34. Institute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
35. Institute for Translational Genomics and Population Sciences, Harbor-UCLA Medical Center, Torrance, CA

36. Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN
37. Human Genetics Center, The University of Texas Health Science Center at Houston School of Public Health, Houston, TX
38. Neurodegeneratives Diseases Institute-CNRS UMR 5293, University of Bordeaux, Bordeaux, France
39. Department of Neurology, University Medicine Greifswald, Greifswald, Germany
40. Population Sciences Branch, National Heart, Lung, and Blood Institute, Bethesda, MD
41. Departments of Radiology and Epidemiology, Erasmus University Medical Center, Rotterdam, Netherlands
42. Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN
43. Departments of Epidemiology, Radiology and Neurology, Erasmus University Medical Center, Rotterdam, Netherlands
44. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany
45. Departments of Neurology and Epidemiology, University of Washington, Seattle, WA
46. Laboratory of Epidemiology and Population Science, National Institute on Aging, Bethesda, MD

Correspondence to:
Myriam Fornage, PhD, FAHA
The University of Texas Health Science Center at Houston
1825 Pressler St, SRB 530F
Houston, TX 77030
Tel: 713-500-2463
Email: Myriam.Fornage@uth.tmc.edu

Correspondence may also be addressed to:
Xueqiu Jian, PhD
The University of Texas Health Science Center at Houston
1825 Pressler St, SRB 530.13
Houston, TX 77030
Tel: 713-500-2476
Email: Xueqiu.Jian@uth.tmc.edu

Cover title: Genetic risk of white matter hyperintensities

Table/Figure count: Tables 1; Figures 2

Key words: white matter hyperintensities; brain MRI; cerebral small vessel disease; exome chip; low-frequency/rare coding variant; meta-analysis

Abstract

Background and Purpose White matter hyperintensities (WMH) on brain magnetic resonance imaging are typical signs of cerebral small vessel disease and may indicate various pre-clinical, age-related neurological disorders such as stroke. Though WMH are highly heritable, known common variants explain a small proportion of the WMH variance. The contribution of low-frequency/rare coding variants to WMH burden has not been explored.

Methods In the discovery sample we recruited 20,719 stroke/dementia-free adults from 13 population-based cohort studies within the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium, among which 17,790 were of European ancestry (EA) and 2,929 of African ancestry (AA). We genotyped these participants at ~250,000 mostly exonic variants with Illumina HumanExome BeadChip arrays. We performed ethnicity-specific linear regression on rank-normalized WMH in each study separately, which were then combined in meta-analyses to test for association with single variants and genes aggregating the effects of putatively functional low-frequency/rare variants. We then sought replication of the top findings in 1,192 adults (EA) with whole exome/genome sequencing data from two independent studies.

Results At 17q25, we confirmed the association of multiple common variants in *TRIM65*, *FBF1*, and *ACOX1* ($p<6\times10^{-7}$). We also identified a novel association with two low-frequency non-synonymous variants in *MRPL38* (lead: rs34136221, $p_{EA}=4.5\times10^{-8}$) partially independent of known common signal ($p_{EA(conditional)}=1.4\times10^{-3}$). We further identified a locus at 2q33 containing common variants in *NBEAL1*, *CARF*, and *WDR12* (lead: rs2351524, $p_{all}=1.9\times10^{-10}$). Although our novel findings were not replicated due to limited power and possible differences in study design, meta-analysis of the discovery and replication samples yielded stronger association for the two low-frequency *MRPL38* variants ($p_{rs34136221}=2.8\times10^{-8}$).

Conclusions Both common and low-frequency/rare functional variants influence WMH. Larger replication and experimental follow-up are essential to confirm our findings and uncover the biological causal mechanisms of age-related WMH.

Introduction

White matter hyperintensities (WMH) on brain magnetic resonance imaging (MRI) refer to areas of high intensity signal within cerebral white matter on T2-weighted images. These abnormalities are believed to reflect demyelination and axonal loss as a result of chronic ischemia and blood-brain barrier dysfunction caused by cerebral small vessel disease (SVD)[1] and are commonly observed in the aging population.[2,3] Substantial evidence supports an association between a high WMH burden and an increased risk of stroke, dementia, and death.[4]

The prevalence and severity of WMH increase with advancing age and presence of cardiovascular risk factors, notably hypertension.[5–7] Besides, susceptibility to WMH has a large genetic component, with heritability estimates ranging from 55% to 80%.[8–10] Genome-wide association studies (GWASs) have identified common genetic variants on chromosome 17q25 associated with WMH.[11,12] These findings were consistently confirmed in independent studies.[13–16] Other loci were also identified genome-wide significant in a multi-ethnic GWAS, including common variants at 10q24, 2p21, 1q22, and 2p16.[12] Though additional associations are yet to be discovered, common variants are estimated to account for at most a quarter of the WMH phenotypic variance.[17] The remaining heritability is still unexplained.

Putatively functional, low-frequency (minor allele frequency (MAF) 1~5%) and rare variants (MAF≤1%) within the protein-coding region of the genome (exome), which are not well captured by GWAS arrays, have been proposed to play an important role in complex traits.[18,19] Yet their effect on WMH has not been explored. Taking advantage of an international collaboration of 13 population-based cohort studies within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium,[20] we conducted exome-wide association analysis in over 20,000 participants of European or African ancestry genotyped with

Illumina HumanExome BeadChip (exome chip) to identify novel coding variants influencing WMH burden.

Methods

Summary data for this meta-analysis will be available through the database of Genotypes and Phenotypes (dbGaP) CHARGE Summary Results site,[21] which can be downloaded via authorized access.

Study sample

Thirteen cohort studies within CHARGE consortium were included in the discovery sample (Supplemental Methods). They followed standardized procedure for subject inclusion, genotype calling, phenotype harmonization, covariate selection, and study-level analysis. In addition, two samples of European ancestry from the Three-City Dijon (3C-Dijon) Study and the Alzheimer's Disease Neuroimaging Initiative (ADNI) were used for replication (Supplemental Methods). Study participants were included in the analyses if they had phenotype, genotype and covariate data available and did not have stroke/dementia at the time of MRI scan. Institutional review boards approved all participating studies, and study participants provided written informed consent.

MRI scan and WMH measurement

MRI scans were performed in each study separately, following a standard procedure (Supplemental Methods). In brief, the magnetic field strength of the scanners used in different studies ranged from 1.5 to 3.0 tesla, except for a single site in the Cardiovascular Health Study (CHS) where the strength was 0.35 tesla. T2-weighted spin-echo pulse sequence were used and complemented by either proton density or fluid-attenuated inversion recovery sequence to contrast WMH against cerebrospinal fluid signal. Axial images were acquired and WMH were

estimated on a semi-quantitative visual rating scale (Atherosclerosis Risk in Communities (ARIC) Study Visit 3 (V3) and CHS) or using a quantitative volumetric method (other studies). The two methods have been compared within ARIC and CHS, and showed high agreement with each other.[22,23]

Genotyping and quality control

Exome chip is a genotyping array focusing on ~250,000 mostly coding variants discovered through exome sequencing in ~12,000 individuals and observed at least three times across at least two existing sequence datasets, including non-synonymous, splicing, stop-altering variants, most of which are rare (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Samples from all discovery cohorts were genotyped with Illumina HumanExome BeadChip. Variant calling and quality control in these studies were performed either jointly[24] or individually following the same protocol (Supplemental Methods). For the replication samples, whole exome sequencing (WES) was performed only on 3C-Dijon individuals with the extremes of the WMH distribution, according to its study design; Whole genome sequencing was performed on ADNI subjects (Supplemental Methods).

Statistical analysis

Figure 1 shows a flow chart illustrating the analytical approach implemented in this study. Detailed description of the analytical methods can be found in Supplemental Methods. In brief, at the individual study level, we performed ethnicity-specific linear regression on rank-normalized WMH in each study separately, which were then combined in meta-analyses to test for association with single variants and genes aggregating the effects of putatively functional low-frequency/rare variants. Analyses conditioning on hypertension status and lead single nucleotide polymorphisms (SNPs) in previous GWAS were also performed, respectively

(Supplemental Methods). The statistical method utilized in the replication analysis for exome-wide significant variants identified is described in Supplemental Methods.

Annotation and functional interpretation

The predicted consequences of all single variants were obtained via dbNSFP,[25] which were then used to aggregate putatively functional variants in the gene-based analysis. Functional prediction scores for the exome-wide significant variants from SIFT,[26] PolyPhen,[27] and CADD[28] were obtained via Ensembl Variant Effect Predictor.[29] We also used the Genotype-Tissue Expression (GTEx) database to investigate whether these variants affect gene expression in multiple brain tissues.[30]

Results

After exclusion and quality control, the discovery sample included 17,790 participants of European ancestry (EA, mean age = 66 years) and 2,929 of African ancestry (AA, mean age = 62 years). Descriptive statistics of the participants from each discovery cohort are summarized in Supplemental Table I. A summary of variants included in the meta-analysis by chromosome is shown in Supplemental Table II.

Our primary analysis identified several common and low-frequency/rare variants associated with WMH ($p<6\times10^{-7}$) in the EA and combined samples but not in the AA sample. They include six variants at 17q25 (four common variants are known and two low-frequency variants in *MRPL38* at 17q25 are novel) and three common variants at 2q33 (Table 1, Figure 2, Supplemental Figure I). No other GWAS loci were significant in our analysis (Supplemental Results). Gene-based analysis showed that *MRPL38* was associated with WMH ($p<3.6\times10^{-6}$) in the EA and combined samples but not in the AA sample (Supplemental Table III, Supplemental Figure II). The association was consistent across studies (Supplemental Table IV). However,

among all variants aggregated in this gene, the two significant variants in the single variant analysis contributed most (Supplemental Table V).

Adjustment for hypertension status did not meaningfully affect the signals in both single variant and gene-based analysis (Table 1, Supplemental Table III). When adjusting for rs7214628, the lead known GWAS SNP, 17q25 associations were no longer significant ($p>0.05$), except those in *MRPL38*. The two low-frequency variants in *MRPL38* remained nominally significant, with p-value ~0.001 in the EA sample and 0.006 in the combined sample (Supplemental Table VI). The same trend was observed in the gene-based analysis (Supplemental Table III).

The replication samples included WES data from 498 individuals with extremes on WMH scale and WGS data from 694 individuals (Supplemental Table VII). We tested nine exome-wide significant variants identified in the discovery sample for replication, which represented three independent signals. Thus the significance threshold was set to $p<0.017$. The results are shown in Supplemental Table VIII. Four known variants in *TRIM65*, *FBF1*, and *ACOX1* at 17q25 were significant in the 3C-Dijon sample. Two low-frequency variants in *MRPL38* at 17q25 and three common variants at 2q33 were not significant. However, the direction of the association for these variants was the same between the discovery and replication samples, and meta-analysis of the p-values for the discovery and replication samples yielded more significant results for the two variants in *MRPL38*. Test of heterogeneity between discovery and replication samples yielded negative results.

Functional predictions for the nine exome-wide significant variants by SIFT, PolyPhen, and CADD are summarized in Supplemental Table IX. Of note, most 17q25 variants have a scaled CADD score >10, indicating they are predicted to be among the 10% most deleterious variants in

the genome. Interestingly, the two low-frequency *MRPL38* variants have the highest scaled CADD score among our top variants (23.9 and 32 for rs34136221 and rs9191, respectively), meaning that they are among the 1% and 0.1% most deleterious variants, respectively. Moreover, both variants are predicted deleterious by SIFT, and rs9191 is predicted probably damaging by PolyPhen.

The expression quantitative trait locus (eQTL) results for these top variants from GTEx database are summarized in Supplemental Table X. In brief, common variants at 17q25 are significant eQTLs for their nearby genes (*FBF1*, *MRPL38*, *TRIM47*, and *TRIM65*) in multiple brain tissues. One of the two low-frequency variants in *MRPL38*, rs34136221, acts as a cis-eQTL for *TRIM47* in cerebellum ($p=1.8\times10^{-6}$) (Supplemental Figure III (A)). At 2q33, two common variants, rs72932557 and rs35212307, are associated with the expression level of *ICA1L* in frontal cortex ($p=2.6\times10^{-5}$) (Supplemental Figure III (B)).

Discussion

In this meta-analysis of association studies between WMH burden and exome chip genotypes in 13 community-based cohorts of stroke/dementia-free adults of European and African ancestry, we identified both common and low-frequency/rare variants significantly associated with WMH. At the known 17q25 locus, four exome-wide significant common variants were identified in our previous GWAS because they were true signals and the majority of the participants in our exome chip analysis were also included in the previous GWAS. Meanwhile, we showed that the association between the low-frequency non-synonymous variants in *MRPL38* and WMH was partially independent of the known GWAS signal. Although this gene was reported previously for WMH,[12,16] the variants identified were mostly common and within the linkage disequilibrium

(LD) block of the lead GWAS SNP at 17q25. According to the design of exome chip, which focuses on different genomic regions and allele frequency spectrum from those for GWAS, it is unlikely that the overlapping sample used in the current study and previous GWAS contributed to the novel low-frequency variants identified here. We also identified significant association at 2q33. In our previous GWAS, this locus (rs72934505, r^2>0.96) was only suggestively associated with WMH,[12] though it reached genome-wide significance when combining evidence from an additional GWAS of WMH in stroke patients.[31] We failed to replicate the novel low-frequency variants in *MRPL38* at 17q25 or in those genes at 2q33, possibly due to the low frequency of the variants (*MRPL38*), the small effect size (2q33), the limited sample size, and the difference in sample selection strategy, phenotypic transformation, or regression model used in the replication samples. Using the 3C-Dijon data, we calculated that we had less than 15% power to detect association for the two low-frequency variants in *MRPL38*, less than 7% for the three variants at 2q33, comparing to over 70% power to detect association for those known variants at 17q25 (*TRIM65*, *FBF1*, and *ACOX1*). However, we confirmed the direction of their effect, and indeed, the improved p-values by meta-analyzing the discovery and replication samples as well as no evidence of heterogeneity provided additional support for the novel association of *MRPL38* variants with WMH (Supplemental Table VIII).

The two low-frequency variants in *MRPL38*, rs34136221 and rs9191, are in complete LD with each other. Both have MAF of ~1.6% in our sample, which are consistent with allele frequencies observed in the 1000 Genomes Project. In this study, these low-frequency variants contributed primarily to the significance of the gene-based association of *MRPL38*, which encodes the mitochondrial 39S ribosomal protein L38 containing 380 amino acids. Both variants are non-synonymous. The "A" allele of rs34136221 leads to a change from arginine to

tryptophan at position 99 (R99W), and the "G" allele of rs9191 changes aspartic acid at position 371 to histidine (D371H). These changes are predicted to be deleterious to the protein function by SIFT and CADD (Supplemental Table IX). This protein is an important constituent of the large subunit of mitochondrial ribosomes responsible for assembling mitochondrial DNA-coded proteins essential for oxidative phosphorylation.[32] The L38 protein maintains the core architecture of the large subunit's central protuberance, which interacts with the small subunit and with mitochondrial transfer RNAs bound to the ribosome, and is hence critical for mitochondrial translation.[32] The exact function of *MRPL38* is still not clear, but interestingly, the structure of the L38 protein is similar to the phosphatidylethanolamine-binding proteins (PEBPs). PEBPs are identified in numerous tissues and have various functions, including a role in neural development and differentiation, which have been implicated in Alzheimer's disease and gliomas.[33] Evidence is also accumulating that mutated mitochondrial ribosomal protein genes are involved in impaired mitochondrial translation leading to several neurological diseases.[34] For instance, *MRPL18*, another mitochondrial ribosomal protein gene that was found to be upregulated in active multiple sclerosis lesions,[35] was also differentially expressed in brains of young spontaneously hypertensive stroke-prone rats and associated with WMH.[16] In addition to the deleterious effect on its encoding protein, rs34136221 also acted as a cis-eQTL for *TRIM47* in cerebellum (Supplemental Figure III (A)), a gene in which common variants have been associated with WMH in previous GWASs,[11,12] indicating that this variant may mediate the common effect of *TRIM47* on WMH. Taken together, these data provide biological plausibility supporting the role of deleterious non-synonymous variants in *MRPL38* in the pathophysiologic changes in the white matter.

Three significant common variants at 2q33 are in high LD with each other (r^2>0.98) with a MAF of ~12% in our sample. rs2351524 is located at the 5' untranslated region of *NBEAL1*. This variant was associated with coronary artery disease[36] and is likely to affect transcription factor binding.[37] *NBEAL1* encodes the neurobeachin-like 1 protein, one of nine Beige and Chédiak-Higashi domain-containing proteins (BDCPs).[38] The functions of BDCPs remain largely unknown, but BDCP gene mutations may affect lysosome size, apoptosis, autophagy, granule size, or synapse formation.[39] Though NBEAL1 is the least studied and understood BDCP, its coding gene is highly expressed in brain and upregulated in glioma.[38] The association of an *NBEAL1* variant reported here is in line with our previous observation that many loci significantly associated with WMH contain variants in genes implicated in malignant brain tumors of the white matter that involve glial cells and further suggests that the genetic pathway may be shared between WMH in aging and glioma, perhaps, through glial cell activation or oxidative DNA damage.[40] rs72932557 is a non-synonymous SNP in *CARF*. This gene encodes a calcium responsive transcription factor, a DNA-binding protein that modulates the transcription of brain-derived neurotrophic factor (BDNF).[41] BDNF is a nerve growth factor that promotes neuronal survival in the adult brain, and a BDNF polymorphism has been associated with WMH in healthy elderly population.[42] Animal studies have shown that knockdown of Carf reduced cortical BDNF expression and impaired memory.[43] Whether the observed association of the *CARF* variant with WMH is direct or through its effect on BDNF remains to be tested. The recent finding that astrocyte-derived BDNF may promote oligodendrogenesis and recovery from white matter damage is consistent with such a hypothesis.[44] rs35212307 is a non-synonymous SNP in *WDR12*. This gene encodes the WD repeat domain 12, a component of the PeBoW complex required for ribosome biogenesis.[45] Variants in *WDR12* have been identified in GWAS

of early myocardial infarction[46] and coronary artery disease[47] and have been associated with carotid intima-media thickness.[48] The relationship between carotid atherosclerosis and WMH has long been reported.[49] In a recent meta-analysis of 10 population-based studies, carotid atherosclerosis was strongly associated with the presence of WMH (odds ratio =1.42, $p<0.0001$),[50] suggesting a possible role of *WDR12* on WMH through atherosclerosis. However, this hypothesis was not supported by a recent large meta-analysis of more than 50,000 participants with carotid intima-media thickness, in which exome chip variants in *WDR12* were not significant.[51] In the eQTL analysis by GTEx, both rs72932557 and rs35212307 were associated with the expression level of *ICA1L* in frontal cortex (Supplemental Figure III (B)), a gene suggestively associated with neuroticism.[52]

Strengths of our study include the large community-based sample of middle-aged to older adults free from stroke/dementia and the first investigation of low-frequency/rare putatively functional variants in the coding regions of the genome in relation to WMH. Several limitations should be noted. In particular, the use of a rank-based inverse normal transformation of WMH did not allow for meaningful estimates of effect size, while having the advantages of harmonizing measures of WMH taken on different scale among individual studies, as well as maintaining desirable statistical properties of the analytical models for low-frequency/rare variant analysis. It should also be noted that, despite close agreement between visual rating and volumetric measurement of WMH, there are still known discrepancies in certain circumstances.[53] Possible inconsistent definitions of dementia used to filter out participants across studies as well as presence of correlated SVD traits such as lacunes and microbleeds may represent sources of heterogeneity in our sample, which may have diminished our power to detect genetic effects. In addition, variants captured by exome chip represent only a fraction of the genetic diversity of the

genome; we were unable to examine even rarer variants or those that lie outside the coding regions. Whole genome sequencing can provide a more comprehensive view of the genome. Recently, an initial WGS effort for brain imaging traits was made in the Framingham Heart Study, in which the investigators identified an independent common signal at 17q25 (rs9889965) associated with WMH volume.[54] Intriguingly, consistent with our novel finding of low-frequency *MRPL38* variant (rs34136221), this variant is also an eQTL for *TRIM47* in brain tissues, which provides additional evidence for the role of eQTLs in the development of WMH through the expression of *TRIM47* involved in glial cell activation. The WGS study also identified two novel loci. While confirmation of these loci in additional samples is needed, they were not observed in our data (not shown).

In conclusion, in this exome chip analysis of over 20,000 stroke/dementia-free individuals of European or African ancestry, we identified low-frequency non-synonymous variants in *MRPL38* at 17q25 associated with WMH partially independently of known common signals. In addition, we showed that common variants at 2q33, a previous suggestive locus in stroke-free populations, reached exome-wide significance in our analysis. Functional annotation provides further support for a role of pathways involved in glial cell activation and oxidative damage, possibly common to WMH and glioma. Larger replication and experimental follow-up are essential to confirm our findings and uncover the biological causal mechanisms of age-related WMH, and thus to provide a foundation for identifying targets for diagnosis, treatment and prevention.

Sources of Funding

This study was supported in part through the National Institute of Neurological Disorders and Stroke grant R01NS087541 and the National Institute on Aging (NIA) grant U01AG049506. Infrastructure for the CHARGE consortium was supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756 and for the neuroCHARGE working group through the NIA grant R01AG033193. Ian J. Deary and Mark E. Bastin are receiving funding from Age UK (Disconnected Mind programme) and the Medical Research Council (MR/K026992/1 and MR/M01311/1). Hans J. Grabe has funding from the Federal Ministry of Education and Research, the German Research Foundation, the DAMP Foundation, and Fresenius Medical Care. Wiro J. Niessen received funding from the Netherlands Organisation for Scientific Research. Helena Schmidt has funding from the Austrian Science Fund. Funding for each participating cohort is reported in the Online Supplement.

Disclosures

Rebecca F. Gottesman is an Associated Editor of *Neurology*. Hans J. Grabe has received speaker honoraria and travel support from Fresenius Medical Care and Janssen. Daniel Levy is Editor-in-Chief of the *Journal of the American Society of Hypertension*. Wiro J. Niessen is founder, stockholder and scientific lead of Quantib BV. Bruce M. Psaty services on the Data and Safety Monitoring Board for a clinical trial of a device funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson.

References

1. Pantoni L, Simoni M. Pathophysiology of cerebral small vessels in vascular cognitive impairment. *Int. Psychogeriatrics*. 2003;15:59–65.
2. Longstreth WT Jr, Manolio TA, Arnold A, Burke GL, Bryan N, Jungreis CA, et al. Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301 elderly people. The Cardiovascular Health Study. *Stroke*. 1996;27:1274–1282.
3. van Dijk EJ, Prins ND, Vermeer SE, Koudstaal PJ, Breteler MMB. Frequency of white matter lesions and silent lacunar infarcts. *J. Neural Transm. Suppl.* 2002;25–39.
4. 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.
5. Pantoni L, Garcia JH. Pathogenesis of leukoaraiosis: a review. *Stroke*. 1997;28:652–659.
6. Launer LJ. Epidemiology of white matter lesions. *Top. Magn. Reson. Imaging*. 2004;15:365–367.
7. Habes M, Erus G, Toledo JB, Zhang T, Bryan N, Launer LJ, et al. White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain*. 2016;139:1164–1179.
8. Atwood LD, Wolf PA, Heard-Costa NL, Massaro JM, Beiser A, D'Agostino RB, et al. Genetic variation in white matter hyperintensity volume in the Framingham Study. *Stroke*. 2004;35:1609–1613.
9. Carmelli D, DeCarli C, Swan GE, Jack LM, Reed T, Wolf PA, et al. Evidence for genetic variance in white matter hyperintensity volume in normal elderly male twins. *Stroke*. 1998;29:1177–1181.
10. Turner ST, Jack CR, Fornage M, Mosley TH, Boerwinkle E, De Andrade M. Heritability

of leukoaraiosis in hypertensive sibships. *Hypertension*. 2004;43:483–487.

11. Fornage M, Debette S, Bis JC, Schmidt H, Ikram MA, Dufouil C, et al. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Ann. Neurol.* 2011;69:928–939.
12. Verhaaren BF, Debette S, Bis JC, Smith JA, Ikram MK, Adams HH, et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ. Cardiovasc. Genet.* 2015;8:398–409.
13. Verhaaren B, de Boer R, Vernooij M, Rivadeneira F, Uitterlinden A, Hofman A, et al. Replication study of chr17q25 with cerebral white matter lesion volume. *Stroke*. 2011;42:3297–3299.
14. Adib-Samii P, Rost N, Traylor M, Devan W, Biffi A, Lanfranconi S, et al. 17q25 Locus is associated with white matter hyperintensity volume in ischemic stroke, but not with lacunar stroke status. *Stroke*. 2013;44:1609–1615.
15. Tabara Y, Igase M, Okada Y, Nagai T, Uetani E, Kido T, et al. Association of Chr17q25 with cerebral white matter hyperintensities and cognitive impairment: the J-SHIPP study. *Eur. J. Neurol.* 2013;20:860–862.
16. Lopez LM, Hill WD, Harris SE, Valdes Hernandez M, Munoz Maniega S, Bastin ME, et al. Genes from a translational analysis support a multifactorial nature of white matter hyperintensities. *Stroke*. 2015;46:341–347.
17. Adib-Samii P, Devan W, Traylor M, Lanfranconi S, Zhang CR, Cloonan L, et al. Genetic architecture of white matter hyperintensities differs in hypertensive and nonhypertensive ischemic stroke. *Stroke*. 2015;46:348–353.
18. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to

common diseases. *Nat. Genet.* 2008;40:695–701.

19. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* 2010;11:415–425.

20. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* 2009;2:73–80.

21. Rich SS, Wang ZY, Sturcke A, Ziyabari L, Feolo M, O'Donnell CJ, et al. Rapid evaluation of phenotypes, SNPs and results through the dbGaP CHARGE Summary Results site. *Nat. Genet.* 2016;48:702–703.

22. Kuller LH, Longstreth WT Jr, Arnold AM, Bernick C, Bryan RN, Beauchamp NJ. White matter hyperintensity on cranial magnetic resonance imaging: a predictor of stroke. *Stroke*. 2004;35:1821–1825.

23. Gottesman RF, Coresh J, Catellier DJ, Sharrett AR, Rose KM, Coker LH, et al. Blood pressure and white-matter disease progression in a biethnic cohort: Atherosclerosis Risk in Communities (ARIC) study. *Stroke*. 2010;41:3–8.

24. Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One*. 2013;8:e68095.

25. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* 2013;34:E2393–E2402.

26. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4:1073–1081.

27. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat. Methods*. 2010;7:248–249.

28. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 2014;46:310–315.

29. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17:122.

30. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 2013;45:580–585.

31. Traylor M, Zhang CR, Adib-Samii P, Devan WJ, Parsons OE, Lanfranconi S, et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology*. 2016;86:146–153.

32. Brown A, Amunts A, Bai XC, Sugimoto Y, Edwards PC, Murshudov G, et al. Structure of the large ribosomal subunit from human mitochondria. *Science (80-.).* 2014;346:718–722.

33. Ling HH, Mendoza-Viveros L, Mehta N, Cheng H-YM. Raf kinase inhibitory protein (RKIP): functional pleiotropy in the mammalian brain. *Crit. Rev. Oncog.* 2014;19:505–516.

34. Boczonadi V, Horvath R. Mitochondria: impaired mitochondrial translation in human disease. *Int. J. Biochem. Cell Biol.* 2014;48:77–84.

35. Fischer MT, Sharma R, Lim JL, Haider L, Frischer JM, Drexhage J, et al. NADPH oxidase expression in active multiple sclerosis lesions in relation to oxidative tissue damage and mitochondrial injury. *Brain*. 2012;135:886–899.

36. The CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* 2013;45:25–33.

37. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22:1790–1797.

38. Chen J, Lu Y, Xu J, Huang Y, Cheng H, Hu G, et al. Identification and characterization of NBEAL1, a novel human neurobeachin-like 1 protein gene from fetal brain, which is up regulated in glioma. *Mol. Brain Res.* 2004;125:147–155.

39. Cullinane AR, Schäffer AA, Huizing M. The BEACH is hot: a LYST of emerging roles for BEACH-domain containing proteins in human disease. *Traffic*. 2013;14:749–766.

40. Al-Mashhadi S, Simpson JE, Heath PR, Dickman M, Forster G, Matthews FE, et al. Oxidative glial cell damage associated with white matter lesions in the aging human brain. *Brain Pathol.* 2015;25:565–574.

41. Tao X, West AE, Chen WG, Corfas G, Greenberg ME. A calcium-responsive transcription factor, CaRF, that regulates neuronal activity-dependent expression of BDNF. *Neuron*. 2002;33:383–795.

42. Huang CC, Liu ME, Chou KH, Yang AC, Hung CC, Hong CJ, et al. Effect of BDNF Val66Met polymorphism on regional white matter hyperintensities and cognitive function in elderly males without dementia. *Psychoneuroendocrinology*. 2014;39:94–103.

43. McDowell KA, Hutchinson AN, Wong-Goodrich SJE, Presby MM, Su D, Rodriguiz RM, et al. Reduced cortical BDNF expression and aberrant memory in Carf knock-out mice. *J. Neurosci.* 2010;30:7453–7465.

44. Miyamoto N, Maki T, Shindo A, Liang AC, Maeda M, Egawa N, et al. Astrocytes

promote oligodendrogenesis after white matter damage via brain-derived neurotrophic factor. *J. Neurosci.* 2015;35:14002–14008.

45. Hölzel M, Rohrmoser M, Schlee M, Grimm T, Harasim T, Malamoussi A, et al. Mammalian WDR12 is a novel member of the Pes1–Bop1 complex and is required for ribosome biogenesis and cell proliferation. *J. Cell Biol.* 2005;170:367–378.

46. Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 2009;41:334–341.

47. Saade S, Cazier J-B, Ghassibe-Sabbagh M, Youhanna S, Badro DA, Kamatani Y, et al. Large scale association analysis identifies three susceptibility loci for coronary artery disease. *PLoS One*. 2011;6:e29427.

48. Zabalza M, Subirana I, Lluis-Ganella C, Sayols-Baixeras S, de Groot E, Arnold R, et al. Association between coronary artery disease genetic variants and subclinical atherosclerosis: an association study and meta-analysis. *Rev. española Cardiol. (English ed.)*. 2015;68:869–877.

49. Bots ML, van Swieten JC, Breteler MM, de Jong PT, van Gijn J, Hofman A, et al. Cerebral white matter lesions and atherosclerosis in the Rotterdam Study. *Lancet*. 1993;341:1232–1237.

50. Moroni F, Ammirati E, Magnoni M, D'Ascenzo F, Anselmino M, Anzalone N, et al. Carotid atherosclerosis, silent ischemic brain damage and brain atrophy: A systematic review and meta-analysis. *Int. J. Cardiol.* 2016;223:681–687.

51. Natarajan P, Bis JC, Bielak LF, Cox AJ, Dörr M, Feitosa MF, et al. Multiethnic exome-wide association study of subclinical atherosclerosis. *Circ. Cardiovasc. Genet.*

2016;9:511–520.

52. Okbay A, Baselmans BML, De Neve J-E, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses. *Nat. Genet.* 2016;48:624–633.

53. Valdés Hernández MC, Morris Z, Dickie DA, Royle NA, Muñoz Maniega S, Aribisala BS, et al. Close correlation between quantitative and qualitative assessments of white matter lesions. *Neuroepidemiology*. 2013;40:13–22.

54. Sarnowski C, Satizabal CL, DeCarli C, Pitsillides AN, Cupples LA, Vasan RS, et al. Whole genome sequence analyses of brain imaging measures in the Framingham Study. *Neurology*. 2018;90:e188–e196.

Figure Legends

Figure 1. A flow chart illustrating the analytical approach implemented in this study. MRI: magnetic resonance imaging; WMH: white matter hyperintensities; ICV: intracranial volume; HT: hypertension; SNP: single nucleotide polymorphism; PC: principal component; GWAS: genome-wide association study; EA: European ancestry; AA: African ancestry. *Field center was adjusted if available; Total ICV was adjusted in cohorts with volumetric measure of WMH; Family relationship was adjusted in cohorts with family data; HT was adjusted in the conditional analysis only. †GWAS lead SNP was adjusted in the conditional analysis assessing independence of significant findings at known loci only.

Figure 2. Manhattan plots for single variant analysis without adjustment for hypertension status in the sample of **(A)** European ancestry (EA), **(B)** African ancestry (AA), and **(C)** their combination (EA + AA), respectively. The minor allele frequency threshold was >0.1%. The significance threshold was $p<6\times10^{-7}$ (grey horizontal line). Two significant loci (17q25 and 2q33) in the EA and combined sample were highlighted in green.

Tables

Table 1. Summary of exome-wide significant variants in either EA, AA, or combined sample.

rsID	Locus	chr:pos	Gene	Function	RA	Sample	N	RAF	p (Unadjusted)	p (HT-adjusted)
rs3760128	17q25	17:73886888	*TRIM65*	NS	G	EA	17,790	0.324	+8.18E−10	+3.50E−10
						AA	2,747	0.711	+7.06E−02	+7.65E−02
						All	20,537	0.375	+1.54E−10	+7.14E−11
rs2351524	2q33	2:203880992	*NBEAL1*	UTR5	T	EA	17,103	0.123	−5.67E−09	−1.83E−08
						AA	2,323	0.062	−4.25E−03	−5.28E−03
						All	19,426	0.115	−1.92E−10	−7.43E−10
rs2305913	17q25	17:73922941	*FBF1*	NS	C	EA	17,790	0.327	+6.47E−09	+3.42E−09
						AA	2,747	0.679	+3.92E−01	+4.19E−01
						All	20,537	0.374	+8.27E−09	+5.03E−09
rs1135640	17q25	17:73949540	*ACOX1*	NS	G	EA	17,422	0.332	+2.21E−08	+1.03E−08
						AA	2,929	0.687	+3.97E−01	+4.32E−01
						All	20,351	0.383	+2.83E−08	+1.57E−08
rs72932557	2q33	2:203846817	*CARF*	NS	T	EA	17,790	0.123	−3.49E−08	−6.53E−08
						AA	2,929	0.034	−1.32E−01	−2.65E−01
						All	20,719	0.11	−1.13E−08	−3.44E−08
rs35212307	2q33	2:203765756	*WDR12*	NS	C	EA	17,790	0.123	−3.85E−08	−1.11E−07
						AA	2,929	0.033	−1.05E−01	−1.46E−01
						All	20,719	0.11	−1.09E−08	−3.89E−08
rs34136221	17q25	17:73898188	*MRPL38*	NS	A	EA	17,790	0.016	+4.47E−08	+4.15E−08
						AA	2,929	0.004	−7.69E−02	−1.06E−01
						All	20,719	0.014	+6.11E−07	+4.93E−07
rs9191	17q25	17:73894963	*MRPL38*	NS	G	EA	17,790	0.016	+1.76E−07	+1.71E−07
						AA	2,929	0.004	−1.25E−01	−1.71E−01
						All	20,719	0.014	+1.62E−06	+1.36E−06
rs1135889	17q25	17:73926121	*FBF1*	NS	A	EA	17,790	0.221	+4.72E−07	+2.08E−07
						AA	2,929	0.194	−7.73E−01	−8.32E−01
						All	20,719	0.217	+4.03E−06	+1.75E−06

EA: European ancestry; AA: African ancestry; RA: risk allele; RAF: risk allele frequency; HT: hypertension; NS: non-synonymous; UTR5: 5' untranslated region. The sign of the p-values indicates the direction of the risk allele.