doi:10.1093/brain/awab253 BRAIN 2021: 144; 2670–2682 | 2670

BRAIN
ORIGINAL ARTICLE

Whole-exome sequencing reveals a role of HTRA1 and EGFL8 in brain white matter hyperintensities

Rainer Malik,[1,†] Nathalie Beaufort,[1,†] Simon Frerich,[1] Benno Gesierich,[1] Marios K. Georgakis,[1] Kristiina Rannikmäe,[2] Amy C. Ferguson,[2] Christof Haffner,[1] Matthew Traylor,[3,4] Michael Ehrmann,[5,6] Cathie L. M. Sudlow[2,7,8] and Martin Dichgans[1,9,10]

[†]These authors contributed equally to this work.

White matter hyperintensities (WMH) are among the most common radiological abnormalities in the ageing population and an established risk factor for stroke and dementia. While common variant association studies have revealed multiple genetic loci with an influence on their volume, the contribution of rare variants to the WMH burden in the general population remains largely unexplored. We conducted a comprehensive analysis of this burden in the UK Biobank using publicly available whole-exome sequencing data (*n* up to 17 830) and found a splice-site variant in *GBE1*, encoding 1,4-alpha-glucan branching enzyme 1, to be associated with lower white matter burden on an exome-wide level [c.691+2T > C, β = –0.74, standard error (SE) = 0.13, $P = 9.7 \times 10^{-9}$].

Applying whole-exome gene-based burden tests, we found damaging missense and loss-of-function variants in *HTRA1* (frequency of 1 in 275 in the UK Biobank population) to associate with an increased WMH volume ($P = 5.5 \times 10^{-6}$, false discovery rate = 0.04). *HTRA1* encodes a secreted serine protease implicated in familial forms of small vessel disease. Domain-specific burden tests revealed that the association with WMH volume was restricted to rare variants in the protease domain (amino acids 204–364; β = 0.79, SE = 0.14, $P = 9.4 \times 10^{-8}$). The frequency of such variants in the UK Biobank population was 1 in 450. The WMH volume was brought forward by ∼11 years in carriers of a rare protease domain variant.

A comparison with the effect size of established risk factors for WMH burden revealed that the presence of a rare variant in the HTRA1 protease domain corresponded to a larger effect than meeting the criteria for hypertension (β = 0.26, SE = 0.02, $P = 2.9 \times 10^{-59}$) or being in the upper 99.8% percentile of the distribution of a polygenic risk score based on common genetic variants (β = 0.44, SE = 0.14, $P = 0.002$). In biochemical experiments, most (6/9) of the identified protease domain variants resulted in markedly reduced protease activity. We further found *EGFL8*, which showed suggestive evidence for association with WMH volume ($P = 1.5 \times 10^{-4}$, false discovery rate = 0.22) in gene burden tests, to be a direct substrate of HTRA1 and to be preferentially expressed in cerebral arterioles and arteries.

In a phenome-wide association study mapping ICD-10 diagnoses to 741 standardized Phecodes, rare variants in the HTRA1 protease domain were associated with multiple neurological and non-neurological conditions including migraine with aura (odds ratio = 12.24, 95%CI: 2.54–35.25; $P = 8.3 \times 10^{-5}$]. Collectively, these findings highlight an important role of rare genetic variation and the HTRA1 protease in determining WMH burden in the general population.

Received March 30, 2021. Revised June 01, 2021. Accepted June 19, 2021
© The Author(s) (2021). Published by Oxford University Press on behalf of the Guarantors of Brain. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

1 Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, 81377 Munich, Germany
2 Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh EH16 4TL, UK
3 Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London EC1M 6BQ, UK
4 The Barts Heart Centre and NIHR Barts Biomedical Research Centre - Barts Health NHS Trust, The William Harvey Research Institute, Queen Mary University of London, London, UK
5 Center of Medical Biotechnology, Faculty of Biology, University Duisburg-Essen, Essen 45141, Germany
6 School of Biosciences, Cardiff University, Cardiff CF10 3AX, UK
7 Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH16 4TL, UK
8 Health Data Research UK Scotland, University of Edinburgh, Edinburgh EH16 4TL, UK
9 Munich Cluster for Systems Neurology, Munich 81377, Germany
10 German Center for Neurodegenerative Diseases (DZNE), Munich 81377, Germany

Correspondence to: Martin Dichgans
Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich
Feodor-Lynen-Str. 17, 81377 Munich, Germany
E-mail: martin.dichgans@med.uni-muenchen.de

Keywords: white matter hyperintensities; whole-exome sequencing; UK Biobank; HTRA1; burden test

Abbreviations: CARASIL = cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy; GWAS = genome-wide association study; PheWAS = phenome-wide association study; PRS = polygenic risk score; SVD = small vessel disease; WES = whole-exome sequencing; WMH = white matter hyperintensities

Introduction

White matter hyperintensities (WMH) are common in elderly individuals[1,2] and are an increasingly recognized risk factor for stroke,[3,4] dementia[3–5] and functional decline in older age.[6] They are further associated with poor long-term outcomes after ischaemic stroke, including an increased risk of stroke recurrence, dementia and mortality.[7] While not specific to any particular aetiology, WMH are considered to be markers of cerebral small vessel disease (SVD).[5,8] Hypertension is the single strongest treatable risk factor for WMH.[4,7,9] However, the precise mechanisms underlying WMH remain largely elusive.

WMH are highly heritable with estimates ranging from 18% to 54%.[10,11] Accordingly, recent genome-wide association studies (GWAS) found common genetic variants at multiple loci to be associated with WMH burden.[11–14] In addition to confirming a causal link with hypertension,[13] these studies also pinpoint specific molecular pathways and biological mechanisms. Among the most prominent themes are perturbations of the extracellular matrix as evidenced by associations at loci that encode matrisome proteins. These loci include *COL4A2* (encoding collagen type IV alpha 2 chain), *EFEMP1* (encoding EGF containing fibulin extracellular matrix protein 1), *VCAN* (encoding versican) and *NID2* (encoding nidogen). Additional themes that have emerged from recent GWAS include dysfunction of vascular endothelial and mural cells, the blood–brain barrier and inflammatory mechanisms.[11,13,15] Results from recent GWAS on WMH further highlight links with small vessel stroke, ischaemic stroke, intracerebral haemorrhage and neurodegenerative disease.[13,16,17] As such, large scale genetic studies have been instrumental in uncovering core pathways and mechanisms underlying WMH and defining relationships with related phenotypes, SVD in particular.

A close link between WMH and SVD is further supported by observations in familial forms of the disease.[18] WMH are a regular feature in carriers of a pathogenic mutation in *NOTCH3* (encoding neurogenic locus notch homolog protein 3), the gene implicated in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL).[19,20] They are further seen in patients with cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL), a severe type of SVD caused by homozygous or bi-allelic *HTRA1* mutations.[21] *HTRA1* encodes a secreted serine protease that associates with the extracellular matrix and has been shown to process various molecular constituents of it proteolytically, including latent TGF-β binding protein, vitronectin and elastin.[22] The outstanding importance of perturbations of the extracellular matrix in the pathophysiology of WMH is further illustrated by highly penetrant mutations in *COL4A1* and *COL4A2*.[23–25] These mutations cause a broad range of phenotypes, including WMH, small vessel stroke, intracerebral haemorrhage, porencephalopathy and extracerebral manifestations.[24,25] Notably, extracerebral manifestations such as ocular and renal manifestations sometimes occur in rare forms of hereditary SVD.[26]

While the above studies have contributed to the understanding of common genetic variation in WMH and the role of highly penetrant mutations in familial SVD, the significance of rare genetic variations for WMH in the general population remains largely unexplored. To the best of our knowledge, previous studies focused on variants included on HumanExome BeadChip arrays[27] and on genes implicated in familial SVD.[28–30] Of note, some of these included samples that had been selected on the basis of extreme phenotypes.[28]

The UK Biobank (UKB)[31] is a large-scale (~500 000 participants) prospective community-based study that recruits from the general midlife population aged between 40 and 69 years and offers phenotypic information on multiple traits including brain imaging, with quantitative data on WMH volumes available for about 44 000 individuals.[32] The UKB further offers detailed clinical information mapped to international classification of diseases (10th revision of the International Statistical Classification of Diseases and Related Health Problems, ICD-10) codes, genome-wide genotyping of single nucleotide polymorphisms (SNPs) and, as of recently, whole-exome sequencing (WES) data. In contrast to genome-wide genotyping, WES enables the comprehensive analysis of all genetic variation in coding regions, including rare variants that have previously been inaccessible. In many complex traits and diseases, studying this variation has contributed to understanding their genetic basis.[33–35]

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

Here, we leveraged data from the UKB to systemically investigate associations between rare variation and WMH burden in a population-based setting. Specifically, we set out to: (i) identify single variants associated with WMH burden at an exome-wide level; (ii) identify genes associated with WMH load in burden tests; (iii) experimentally determine the functional consequences of selected variants; (iv) identify genes implicated in pathways and interaction networks relevant to SVD; and (v) explore the phenotypic spectrum of alternative allele carriers in a phenome-wide association study (PheWAS).

Materials and methods

UK Biobank population

The UKB received ethics approval from the National Health Service Research Ethics Service (London, UK; reference 11/NW/0382). This analysis was carried out under UKB project 2532. Our resource was the UK Biobank Exome 200k release from October 2020. Primary and secondary analyses were performed with an updated Functional Equivalence protocol that retains original quality scores in the CRAM files (referred to as the OQFE protocol).[36] We included only variants that met the following criteria: individual and variant missingness $<10\%$, Hardy-Weinberg equilibrium P-value $> 10^{-15}$, minimum read coverage depth of 7 for SNPs and 10 for indels, at least one sample per site passed the allele balance threshold >0.15 for SNPs and 0.20 for indels.

The 200k release encompasses 1135 parent-offspring pairs, 3855 full-sibling pairs, including 101 trios, 27 monozygotic twin pairs and 7461 second degree genetically determined relationships. To avoid bias due to relatedness, we selected an unrelated set of individuals up to the second degree (KING cut-off 0.0877).[37] For the WMH volume analysis, we preferentially retained individuals with more extreme trait values (i.e. further away from the mean) in the analysis using PRIMUS.[38] For PheWAS analyses, we selected a standard unrelated set (also up to the second degree), which was used for all subsequent analyses ($n = 166\,897$). We further excluded individuals showing an excess of heterozygosity from the UKB genotyping analysis and individuals of non-white British ancestry.

For the WMH analyses, we used UKB field ID 25781 ($n = 38\,347$; n with WES = 17 830; n unrelated = 16 511). This study used the January 2020 release of UKB imaging data on ~44 000 individuals. MRI was performed on two identical Siemens Skyra 3.0 T scanners (Siemens Medical Solutions), running VD13A SP4, with a standard Siemens 32-channel RF receiver head coil. Identical acquisition parameters and detailed quality control was used for all scans. Using the T_2-FLAIR sequence, WMH volumes were generated by an image-processing pipeline developed and run on behalf of UKB and were available as part of the UKB central analysis (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf, accessed 13 September 2021).[39–41] We excluded three extreme volume outliers [>6 standard deviations (SD)] and used the log-transformed WMH (logWMH) volume to ensure normal distribution.

Single variant analysis

Single variant analysis of rare variants in the WES data was performed using REGENIE.[42] REGENIE accounts for relatedness and subtle population stratification through a mixed-model approach. The mixed model parameters were estimated using 200 000 genotyped common variants. Saddle point approximation regression was used in favour of Firth's correction due to performance issues. We set a conservative exome-wide threshold of 5×10^{-8} as the significance cut-off in our analysis.

Whole-exome burden test on white matter hyperintensities

We determined the functional consequences of exome-wide variants using the variant effect predictor (VEP) tool v101.[43] For the whole-exome burden test on WMH analysis, we selected rare variants [minor allele frequency (MAF) < 0.01] that are either predicted to be damaging by REVEL[44] (REVEL score > 0.5) or predicted to exert a high-confidence loss-of-function effect using the LoFTEE[45] plugin in VEP. We used the GrCh38 refFlat definition of genes as provided by the UCSC genome annotation database.

For the domain specific analyses of *HTRA1*, we extracted all missense variants according to their position in the domain structure as defined by Uniprot identifier Q92743[46]: signal peptide domain [amino acids (AA) 1–22], IGFBP domain (AA 33–100), Kazal-like domain (AA 100–157), protease domain (AA 204–364) and PDZ domain (AA 365–467).

Whole-exome burden tests were carried out using rvtests[47] and by calculating multivariable burden tests using the combined multivariate and collapsing burden schema.[48] Sex, age at imaging and 10 genomic principal components were used as covariates in all analyses. For the primary logWMH analysis, we additionally corrected for vascular risk factors that were shown to associate with WMH volume[49] (smoking, hypertension, pulse pressure, diabetes and waist-to-hip ratio) in a sensitivity analysis. To test the stability of our results, we also performed the burden test using a weighting scheme suggested by Morris and Zeggini,[50] a SKAT test[51] and a SKAT-O test.[52] Results were deemed significant at a false discovery rate (FDR) of 5% after FDR correction. To estimate effect sizes and standard errors, we used linear regression with minor allele carrier status as a dependent variable and age at imaging, sex and 10 genomic principal components as covariates.

To compare effect sizes to dichotomized risk factors for WMH, we extracted the following information from the UKB: hypertension status (yes/no), defined as either (i) diagnosis of hypertension through self-report or ICD-10 codes (I10–I16); (ii) use of antihypertensive drugs; or (iii) systolic blood pressure >140 or diastolic blood pressure >90; diabetes mellitus (yes/no, UKB field 2443); body mass index (above mean versus below mean; UKB field 21001); and smoking (ever smoked versus never smoked; UKB field 20116). We further calculated a polygenic risk score (PRS) based on common genetic variants from the most recent WMH GWAS.[13] Specifically, we selected the 27 genome-wide significant lead SNPs and constructed a weighted allelic risk score for each individual in our dataset. Variants were weighted by their respective effect sizes in their association with WMH. We compared individuals above the mean versus below the mean with a PRS > 95% percentile versus PRS < 95% percentile and a PRS > 99.8% percentile versus PRS < 99.8% percentile. The cut-off of 99.8% was chosen to achieve the same case number as for the HTRA1 protease domain variant carrier status. Thus, the standard error (SE) is identical.

Recombinant protein expression

Expression vectors

The cDNA encoding human full-length HTRA1 (AA 1–480) or human full-length EGFL8 (AA 1–293, Origene) was cloned into a pcDNA4/TO/myc-His expression plasmid (Invitrogen). The cDNA encoding the N-terminal region of human LTBP1 (latent-transforming growth

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

factor β-binding protein 1, AA 1–689) fused to a C-terminal V5-His tag was cloned into the pTT5 plasmid. Mutagenesis was conducted using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies).

Cell culture and transfection

Human embryonic kidney (HEK) 293E cells were grown in Dulbecco's modified Eagle medium (DMEM) containing GlutaMAX, 10% (v/v) foetal calf serum (FCS), 100 U/ml penicillin and 100 μg/ml streptomycin (all from Invitrogen). Cells were transfected with Lipofectamine™ 2000 (Invitrogen) and maintained for 48 h in FCS-free DMEM before the culture medium was collected and centrifuged for 10 min at 400*g* to remove debris.

EGFL8 enrichment

Stably transfected HEK cells were selected using Zeocin (100 μg/ml, Invitrogen), then maintained in FCS-free DMEM for 4–5 days. The culture medium was collected and centrifuged at 1000*g* for 15 min, dialysed at room temperature against 0.5× phosphate-buffered saline (PBS) for 2 h and overnight at 4°C against 0.5× PBS containing 200 mM NaCl. The medium was gently agitated for 1 h at room temperature in the presence of Talon resin (5 μl solution per ml medium, Clontech) and the resin subsequently transferred to a gravity flow column. After washing, EGFL8 was eluted with 100 mM EDTA in PBS and the eluate dialysed overnight at 4°C against 1000 volumes of Tris 50 mM, NaCl 150 mM, pH 8.0.

HTRA1 purification

Human HTRA1 wild-type lacking the N-terminal Mac domain (AA 158–480) was produced as described previously[53] with minor modifications. N-terminally StrepII-tagged wild-type HTRA1 was affinity-purified with a strep-tactin resin material (IBA Lifesciences) and subsequently subjected to size exclusion chromatography using a Superdex™ 200 preparation grade column (GE Healthcare) in 20 mM HEPES and 50 mM NaCl, pH 7.5.

Protease activity assays

Medium from HEK cells overexpressing LTBP1 was treated with medium from cells overexpressing HTRA1 for 24 h at 37°C. LTBP1 proteolysis was evaluated by anti-V5 immunoblot. Relative protease activity was calculated as the ratio of cleaved to intact LTBP1 and normalized to HTRA1 expression, assessed by anti-Myc immunoblot.

EGFL8 was treated with 100 or 500 nM purified wild-type HTRA1 for 24 h at 37°C. Where indicated, 5 μM of the HTRA1 inhibitor NVP-LBG976 (obtained from Novartis Pharmaceuticals) were added to the reaction.[54] Substrate proteolysis was assessed by anti-Myc (EGFL8) immunoblot.

SDS-PAGE and immunoblot

Proteins were separated by SDS-PAGE and transferred onto polyvinylidene difluoride membranes (Immobilon-P, Millipore). Membranes were blocked in Tris-buffered saline (TBS) containing 0.2% (v/v) Tween 20 and 4% (w/v) skimmed milk and probed with anti-Myc (Santa Cruz Biotechnology, #sc-40, 1:5000), anti-V5 (Invitrogen, #R960-25, 1:10 000) or anti-HTRA1 (R&D Systems, #MAB2916, 1:5000) primary antibodies prepared in TBS/Tween/milk. Detection was performed with horseradish peroxidase-coupled secondary antibodies (Dako), Immobilon Western kit reagents (Millipore) and the Fusion FX7 documentation system (Vilbert Lourmat). Signal intensity was quantified with ImageJ.

Immunohistochemistry

Frozen human brain (frontal subcortex) from a 55-year-old individual with no known cerebrovascular disorder was provided by the Brain-Net Europe Biobank (Ludwig-Maximilian University, Munich) and analysed following ethical approval by the local ethics committee (reference number 17–140). The participant gave written informed consent according to the Declaration of Helsinki. For EGFL8 detection, 10 μm tissue sections were thawed to room temperature and fixed for 20 min in 4% (w/v) paraformaldehyde then permeabilized and blocked in PBS added with 0.1% (v/v) Triton X-100 and 5% (w/v) bovine serum albumin (BSA) for 1 h at room temperature. Primary antibodies were diluted in PBS/Triton/BSA as follows: anti-EGFL8 Ab (Sigma-Aldrich, #HPA061173, 1:50), anti-collagen IV Ab (SouthernBiotech, #1340–01, 1:200), Cy3-coupled anti-α smooth muscle actin (SMA) Ab (Merck, #C6198, 1:100), and incubation was performed overnight at 4°C. After washing with PBS, sections were incubated with Alexa Fluor 488- or 647-coupled secondary antibodies (Abcam, 1:100) and 4′,6-diamidino-2-phenylindol (DAPI, ThermoFisher Scientific) diluted in PBS for 1 h at room temperature. Tissue was washed, mounted with Fluoromount (Sigma-Aldrich) and images captured by confocal microscopy (LSM800, Zeiss).

PheWAS

To explore the association of specific regions and variants in the full range of phenotypes encoded within the UKB, we used the Phecode Map 1.2 to map UKB ICD-10 codes to Phecodes.[55] We used all ICD10 codes (main position, secondary position, death records) from the UKB. We excluded Phecodes with <100 cases and those that were male or female-specific. Individuals were assigned a case status if >1 ICD-10 code mapped to the respective Phecode. Individuals meeting the prespecified exclusion criteria were removed from the analysis; otherwise, the individual was assigned a control status. Given the known association between WMH and stroke, we further analysed any stroke (4143 cases/162 811 controls), any ischaemic stroke (1948 cases/165 006 controls) and intracerebral haemorrhage (413 cases/166 541 controls) as defined by the algorithmically-defined stroke outcomes (UKB fields 42006–42013), parental history of stroke (42 875 with family history/124 022 without family history) and family history of dementia (23 730 with family history/143 167 without family history) (UKB fields 20107 and 20110). The results were deemed significant at an FDR level of 5% after FDR correction. To approximate the effect size in a logistic regression framework, we used minor allele carrier status as an independent variable and age at baseline, sex and 10 genomic principal components as covariates. We used Firth's correction for unbalanced case/control ratios in our logistic regression analysis. Additionally, we analysed serum biomarker levels available in the UKB [UKB category 17518; diastolic blood pressure, mean arterial pressure, total cholesterol, urate, systolic blood pressure, LDL cholesterol, apolipoprotein B, total protein, HDL cholesterol, glucose, lipoprotein A, glycated haemoglobin (HbA1c), apolipoprotein A, triglycerides, pulse pressure, cystatin C, urea, creatinine, C-reactive protein, body mass index and vitamin D].

Data availability

UKB data are available through a procedure described at http://www.ukbiobank.ac.uk/using-the-resource/ (accessed 13 September 2021).

Results

We analysed up to 17 830 individuals from the UKB for whom both WES data and quantitative data on WMH volume were available.

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

Their demographic characteristics are presented in Supplementary Table 1.

A predicted splice donor loss-of-function variant in *GBE1* associates with WMH volume

In the single variant analysis using a mixed model approach and correcting for sex, age at imaging and 10 principal components, we found one rare variant (MAF < 0.01) to be associated at the genome-wide significance level ($P < 5 \times 10^{-8}$) with a lower logWMH volume (β = –0.74, SE = 0.13) in 17 830 individuals. This variant is a predicted splice donor loss-of-function variant in *GBE1* (rs192044702, c.691+2T>C, MAF = 0.00131) encoding 1,4-alpha-glucan branching enzyme 1. We further found three single variants in *KLHL21* (rs891958177), *PARP2* (rs750147320) and *ABCC1* (rs748043154) to reach a suggestive threshold of $P < 5 \times 10^{-7}$ for association with logWMH volume (Supplementary Table 2).

Whole-exome burden test reveals *HTRA1* as a risk gene for white matter hyperintensity burden

We next performed a whole-exome burden test using the combined multivariate and collapsing method implemented in rvtests software. Analyses were corrected for sex, age at imaging and 10 genomic principal components as covariates with the logWMH volume as the dependent variable ($n = 16\,511$ unrelated individuals). The results were subsequently filtered to retain genes with >20 alternative alleles in the analysed dataset. *HTRA1* was the only gene significantly associated at an FDR level of 5% ($P = 5.6 \times 10^{-6}$), based on 59 alternative *HTRA1* allele carriers (prevalence of ~3.6/1000 in the UKB) at 18 genomic positions. Four additional genes including *GBE1*, *DCAKD*, *EGFL8* and *RGS12* showed suggestive associations at an FDR level of 25% (Table 1, Fig. 1 and Supplementary Table 3). To estimate the effects of carrier status on the logWMH volume, we performed a linear regression analysis on the log transformed WMH volume and minor allele carrier status, corrected for sex, age at imaging and 10 genomic principal components. Carriers of rare variants in either *HTRA1*, *EGFL8* or *RGS12* showed increased WMH volumes compared with non-carriers (*HTRA1*: β = 0.36, SE = 0.09, $P = 0.0003$; *EGFL8*: β = 0.22, SE = 0.06, $P = 9.6 \times 10^{-5}$; *RGS12*: β = 0.29, SE = 0.10, $P = 0.003$, respectively), whereas carriers of rare variants in *GBE1* (β = –0.20, SE = 0.05, $P = 6.7 \times 10^{-5}$) and *DCAKD* (β = –0.21, SE = 0.05, $P = 0.0002$) showed lower WMH volumes. The results did not materially change when further correcting for vascular risk factors (smoking, hypertension, pulse pressure, diabetes and waist-to-hip ratio) (Supplementary Table 4). In an additional sensitivity analysis, we only considered loss-of-function mutations in the exome-wide burden test. The associations of *GBE1*, *DCAKD* and *EGFL8* with WMH volume were of similar magnitude as in the combined damaging missense and loss-of-function analysis. (Supplementary Table 5) In contrast, the association of *HTRA1*, which was based on only nine loss-of-function variants, was greatly diminished ($P = 0.97$), and therefore, *HTRA1* loss-of-function variants were not considered in further analyses. There was no association with loss-of-function alleles in *RGS12* ($P = 0.69$). We did not identify additional FDR-significant genes when considering different weighting schemes in the burden test (Supplementary Tables 6–8).

Domain-specific burden test of *HTRA1* shows specific association with the protease domain

Given the association between rare variants in *HTRA1* and WMH volume in the burden tests and the overall structure of the HTRA1 protease, which is organized into functionally distinct domains (Fig. 1), we next repeated the analyses focusing on rare missense variants in individual domains. Specifically, we performed a separate combined multivariate and collapsing burden test for the signal peptide (AA 1–22, $n = 2$ variants, prevalence 7.8/1000), IGFBP domain (AA 33–100, $n = 6$ variants, prevalence 0.6/1000), Kazal-like domain (AA 100–157, $n = 4$ variants, prevalence 0.9/1000), protease domain (AA 204–364, $n = 9$ variants, prevalence 2.2/1000) and PDZ domain (AA 365–467, $n = 11$ variants, prevalence 1.6/1000). Rare missense variants in the HTRA1 protease domain were significantly associated with altered WMH volumes ($P = 9.5 \times 10^{-8}$), whereas there was no signal for the other domains (Fig. 1 and Table 2) and the results remained stable after exclusion of one compound heterozygote individual (D320N/R403W). When including loss-of-function mutations in the model, the association was attenuated ($P = 1.6 \times 10^{-6}$). Baseline characteristics of protease domain variant carriers did not differ from those of non-carriers (Supplementary Table 9). We also calculated results for the combined protease and linker domain as the linker domain (AA 158–204; two variants) serves as an activator of protease function and found the association to be slightly attenuated ($P = 5.5 \times 10^{-7}$).

To approximate the effects of carrier status on WMH volume, we performed linear regression analysis correcting for sex, age at imaging and 10 genomic principal components. The carrier status of an alternative allele in the HTRA1 protease domain was associated with an increase in the logWMH volume (β = 0.79, SE = 0.14, $P = 9.4 \times 10^{-8}$; FDR < 5%), whereas we found no effect of carrier status on WMH volume in other domains (Table 2). In an age-stratified analysis, carriers exhibited logWMH volumes comparable to non-carriers that were 11.4 years older (Fig. 1B). Compared with other dichotomized risk factors for WMH volume, protease domain mutation carrier status showed a larger effect than hypertension (β = 0.28, SE = 0.02), diabetes (β = 0.25, SE = 0.04), body mass index (β = 0.18, SE = 0.01) and smoking (β = 0.10, SE = 0.01). Also, comparison with individuals on a PRS distribution based on 27 common genetic variants[13] showed a larger effect size for the carrier status of an alternative allele in the HTRA1 protease domain than for individuals with a PRS above versus below the 99.8% percentile of the PRS distribution (β = 0.44, SE = 0.14) (Fig. 2). Because of sample overlap in the discovery and replication sets of the PRS, we performed a sensitivity analysis using a PRS based on six variants associated with WMH volume published by the CHARGE consortium.[56] The results were slightly attenuated compared with the full PRS from the CHARGE + UKB sample (PRS above mean versus PRS below mean: β = 0.18, SE = 0.02; PRS > 95th percentile versus PRS < 95th percentile: β = 0.26, SE = 0.04; PRS > 98.8th percentile versus PRS <98.8th percentile: β = 0.29, SE = 0.14).

Mutations in the HTRA1 protease domain are associated with a loss of protease activity

To elucidate the functional consequence of the identified *HTRA1* variants, we performed activity assays using human cell-derived recombinant proteins (Fig. 3). Specifically, we evaluated HTRA1 protease activity towards its known physiological substrate LTBP1. Wild-type HTRA1 and an active site mutant obtained by replacing catalytic serine with an alanine (S328A) served as positive and negative controls, respectively. We investigated the enzymatic activity of two Kazal-like domain, two PDZ domain and nine protease domain-mutants. Three known CARASIL loss-of-function mutants were also included in the assays.[22,57] We found the protease activity of Kazal-like and PDZ domain mutants to be comparable to that of wild-type HTRA1, in accordance with the fact that these domains are dispensable for protease activity.[58,59] In contrast, six of nine protease domain mutants (V221L, V221M, R227W, P275L, M314V and D320N) exhibited a marked reduction of protease activity.

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

Table 1 Whole-exome burden test on white matter hyperintensities

Gene	Genomic position (hg38)	No. of rare variants	Minor allele count	P-value	FDR-adjusted P-value
HTRA1	10:122461552–122514907	18	59	**5.5×10^{-6}**	**0.044**
GBE1	3:81489704–81761645	48	281	3.2×10^{-5}	0.131
DCAKD	17:45023337–45061129	9	225	7.1×10^{-5}	0.192
EGFL8	6:32164594–32168285	15	237	1.5×10^{-4}	0.229
RGS12	4: 3314146–3439913	17	72	1.7×10^{-4}	0.229

Genes reaching exome wide significance (FDR-adjusted $P < 0.05$; in bold) or suggestive evidence (FDR-adjusted $P < 0.25$) for an association with logWMH volume are listed. The results were derived from a unidirectional mutivariable burden test on logWMH and adjusted for age at imaging, sex and the first 10 genomic principal components. Only the variants that were predicted to be damaging (REVEL score > 0.5) or predicted to exert a loss-of-function effect by LoFTEE were considered in the analysis. Genes with a minor allele count <20 were filtered out.

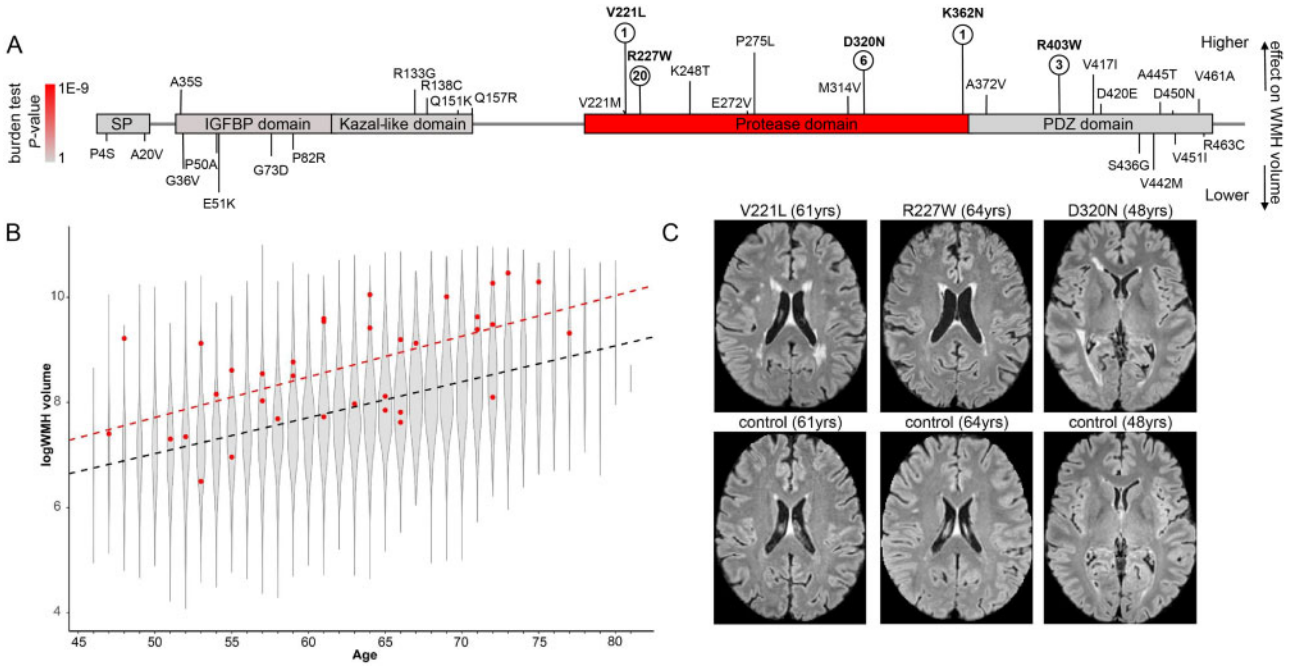

Figure 1 Overview of rare *HTRA1* variants and the association with WMH burden. (**A**) The domain structure of the HTRA1 protein and the position of rare missense variants in the 16 511 UKB participants included in the current analysis. The vertical lines reflect effect sizes (line length) and directionality of effect (increasing versus decreasing) of the respective variant on WMH volume. Variants reaching statistical significance in single variant tests ($P < 0.05$) are shown in bold with the number of minor allele carriers indicated below. (**B**) A violin plot of the age-stratified WMH burden in the UKB. Carriers of a rare variant in the HTRA1 protease domain are depicted in red. The dashed lines represent fitted linear regression lines (WMH∼age) for all participants (black) and protease mutations carriers (red). (**C**) Example brain MRI images of HTRA1 protease domain variant carriers and age-matched non-carriers. Non-carriers were chosen to represent individuals with a mean logWMH value near the mean of the age group.

EGFL8 is proteolytically processed by HTRA1 and expressed in the human brain vasculature

Given the prominence of variants in the HTRA1 protease domain in the association analyses and the results of the protease assays, we next performed a targeted analysis of genes encoding potential substrates of HTRA1. Potential substrates were identified from a recent proteomic study in HTRA1-deficient mice[60] that found 38 proteins to be enriched in the brain vasculature of *Htra1*$^{-/-}$ compared with wild-type mice, among them several verified HTRA1 substrates including EFEMP1, VTN and LTBP1 (Supplementary Table 10).[60,61] Notably, one of the most strongly enriched proteins in that dataset was EGFL8. Focusing on the corresponding genes encoding these 38 candidate substrates in our burden test, *EGFL8* was the only gene that reached experiment-wide significance at an FDR level of 5% ($P = 1.54 \times 10^{-4}$).

We next set out to determine experimentally whether the human EGFL8 protein is a substrate of HTRA1. Treatment of EGFL8 with HTRA1 resulted in dose-dependent degradation of EGFL8, which was prevented in the presence of a selective HTRA1 inhibitor (Fig. 4A). These observations thus identified EGFL8 as a novel HTRA1 substrate. We further examined the cerebral expression pattern of EGFL8 using immunohistochemistry. In the human cortex, EGFL8 was predominantly localized to the αSMA-positive vasculature, thus corresponding to arteries and arterioles (Fig. 4B).

Rare variants in the HTRA1 protease domain associate with multiple neurological and non-neurological phenotypes

HTRA1 mutation carriers exhibit a wide range of cerebral and extracerebral phenotypes.[26] To comprehensively assess disease

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

Table 2 *HTRA1* domain-specific burden test on white matter hyperintensities

Domain	Variants	MAC	β	SE	P-value
Signal peptide		129	–0.02	0.07	7.80×10^{-1}
	P4S	2	–0.21	0.61	7.34×10^{-1}
	A20V	127	–0.07	0.08	3.79×10^{-1}
IGFBP domain		11	0.18	0.23	4.30×10^{-1}
	A35S	1	0.66	0.87	4.45×10^{-1}
	G36V	3	–0.92	0.50	6.54×10^{-2}
	P50A	2	–0.52	0.61	4.01×10^{-1}
	E51K	1	–1.53	0.87	7.77×10^{-2}
	G73D	1	–0.59	0.87	4.98×10^{-1}
	P82R	3	–0.42	0.50	4.06×10^{-1}
Kazal domain		15	0.12	0.22	5.20×10^{-1}
	R133G	1	0.34	0.56	7.80×10^{-1}
	R138C	1	0.23	0.66	4.50×10^{-1}
	Q151K	12	0.02	0.25	9.34×10^{-1}
	Q157R	1	0.32	0.87	7.17×10^{-1}
Protease domain		35	0.79	0.14	**9.50×10^{-8}**
	V221L	1	1.87	0.87	**3.10×10^{-2}**
	V221M	1	0.08	0.87	9.26×10^{-1}
	R227W	20	0.73	0.19	**1.71×10^{-4}**
	K248T	1	0.85	0.87	3.26×10^{-1}
	E272V	3	0.03	0.50	9.59×10^{-1}
	P275L	1	1.61	0.87	6.40×10^{-2}
	M314V	1	0.48	0.87	5.83×10^{-1}
	D320N	6	1.17	0.35	**9.77×10^{-4}**
	K362N	1	1.82	0.87	**3.60×10^{-2}**
PDZ domain		26	0.06	0.17	4.30×10^{-1}
	A372V	1	0.53	0.61	3.89×10^{-1}
	R403W	3	1.01	0.50	**4.51×10^{-2}**
	V417I	1	1.06	0.87	2.24×10^{-1}
	D420E	4	0.26	0.43	5.49×10^{-1}
	S436G	1	–0.54	0.87	5.35×10^{-1}
	V442M	1	–0.96	0.87	2.71×10^{-1}
	A445T	4	0.30	0.43	4.95×10^{-1}
	D450N	2	0.01	0.61	9.83×10^{-1}
	V451I	6	–0.30	0.35	4.05×10^{-1}
	V461A	1	0.36	0.87	6.75×10^{-1}
	R463C	2	–0.02	0.61	9.72×10^{-1}

The results were derived from a multivariable burden test (combined multivariate and collapsing, for domains) or from a Wald association statistic (for single variants). All analyses were adjusted for age at imaging, sex and the first 10 genomic principal components. All missense variants mapping to the respective domains were considered. Results are displayed for the specific domains or the single variants. Significant *P*-values ($P < 0.05$) are displayed in bold.

outcomes associated with rare variants in *HTRA1* or *EGFL8*, we performed a rare-variant burden PheWAS on biomarkers and standardized Phecodes in the UKB. Specifically, we matched patients' ICD-10 codes (both primary, secondary and death-record based) to PheCodes. After excluding Phecodes that were sex-specific or had <100 cases in the analysis, 741 different Phecodes remained that were analysed in burden tests. We further performed analyses for any stroke, any ischaemic stroke and intracerebral haemorrhage as defined by the algorithmically defined stroke outcomes (UKB fields 42006–42013) and for parental history of stroke and family history of dementia (UKB fields 20107 and 20110).

Considering all identified damaging and loss-of-function variants in *HTRA1* (Supplementary Tables 11 and 12), we found conductive hearing loss (Phecode 389.2) to be associated at an FDR level of 5% (Supplementary Table 13 and Supplementary Fig. 2) with minor allele carriers having a higher risk in Firth logistic regression [OR: 5.37, CI95 (2.39–9.49)]. Focusing on damaging variants in the protease domain, we identified 12 Phecodes at an FDR level of 5% (Fig. 5, Table 3 and Supplementary Table 14) including migraine with aura (Phecode 340.1), encephalitis (Phecode 323), aneurysms and dissection of heart (Phecode 411.41) and heart valve replaced (Phecode 395.6). Notably, for all significant Phecodes, minor allele carrier status was associated with higher risk of the respective phenotype. We did not find any FDR-significant associations with biomarkers available in the UKB (Supplementary Table 15). For *EGFL8*, we identified two Phecodes at an FDR level of 5%: 578.8 (haemorrhage of rectum and anus) and 696.41 (psoriasis vulgaris), the latter confirming prior work in the UKB (Supplementary Table 16 and Supplementary Fig. 3).[62]

Discussion

Using WES data from the UKB, we showed that rare variants in *HTRA1* are associated with a higher burden of radiological WMH in the general population. We further showed that the association with WMH burden is largely restricted to variants affecting the HTRA1 protease domain and provided experimental evidence that the variants giving the strongest association signal result in a loss of proteolytic activity of HTRA1. Our PheWAS results revealed that heterozygous variants in the HTRA1 protease domain, which are present at a frequency of 1 in 450 in the UKB population, are associated with a wide range of phenotypes including migraine with aura. We further demonstrated that the gene product of *EGFL8*, which shows suggestive association with WMH burden in the UKB, is a direct substrate of HTRA1 and is expressed in the brain vasculature. Collectively, these findings highlight an important role of the HTRA1 protease in maintaining white matter integrity and provide insights into disease mechanisms.

HTRA1 has been implicated in familial SVD. Biallelic mutations in *HTRA1* cause CARASIL, an early-onset condition that is characterized by stroke, dementia, spondylosis deformans and alopecia.[21,22] In contrast, heterozygous mutations in *HTRA1* cause an autosomal dominantly inherited SVD that displays a more restricted phenotype of stroke, cognitive decline and gait disturbance and manifests at a later age of onset.[63] Our results extend these findings by showing that rare variants in *HTRA1*, and more specifically variants in the HTRA1 protease domain, are associated with the WMH burden on a population-wide level. Notably, we found no association of such variants with stroke, dementia or a family history of stroke or dementia in the UKB. Instead, we found an association with both neurological and non-neurological phenotypes that have not previously been associated with *HTRA1* mutations.

The domain-specific analyses for HTRA1 in combination with the protease assays suggest that the mechanism underlying the observed association between rare variants in *HTRA1* and WMH burden is a loss of proteolytic function of the HTRA1 protease. This would be in line with the proposed mechanisms for mutations implicated in *HTRA1*-related familial SVD.[22,63] However, our experiments focused on a set of variants identified in the UKB and were limited to biochemical assays using recombinant proteins. Other putative loss-of-function mechanisms related to e.g. gene expression, mRNA or protein stability[21,64] were not evaluated. Moreover, we did not go further in examining possible dominant negative effects, as recently reported for a subset of *HTRA1* variants linked to hereditary SVD.[57,58] Of interest, one of the variants (D320N) detected here has previously been reported in an Asian CARASIL patient[65] who was compound heterozygous for D320N. In contrast, R227W, which had an allele frequency of 0.04% in the UKB population with available WES data (0.06% in those with imaging), was associated with WMH burden and has so far not been linked to neurological disease. More work is needed to delineate genotype-phenotype correlations at this locus.

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

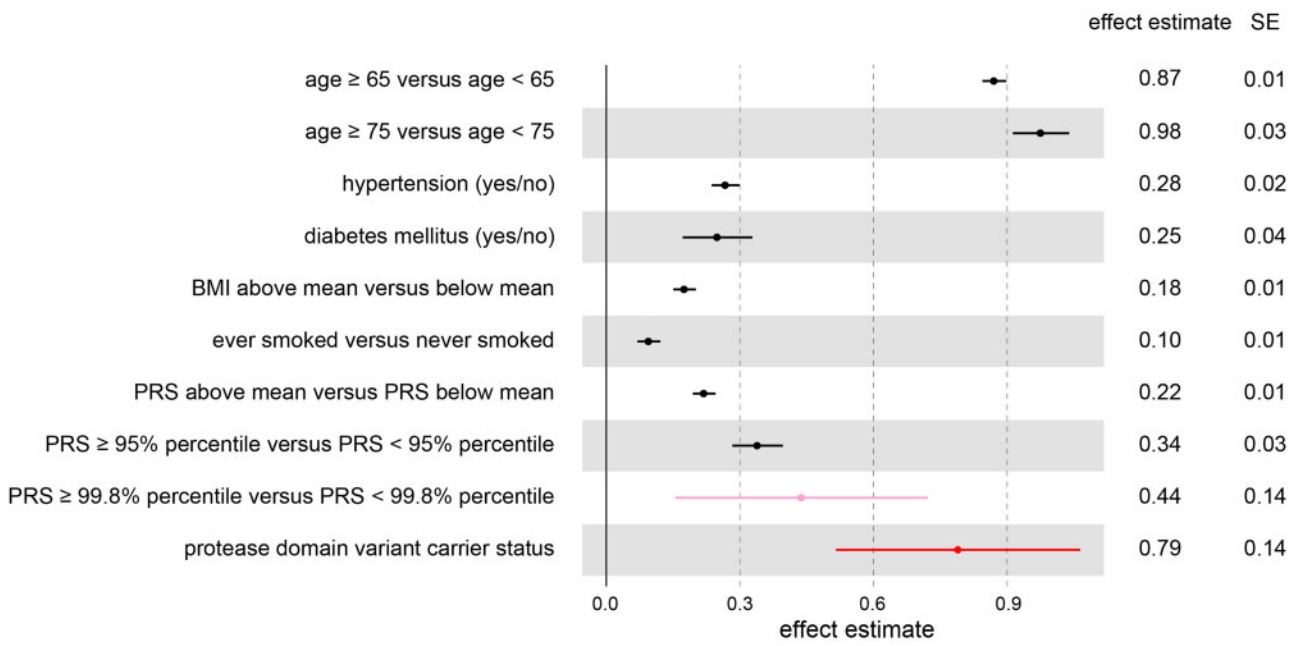

Figure 2 Effect size of protease domain variant carrier status on logWMH volume in comparison with established risk factors. The effect sizes and standard errors of dichotomized risk factors (age, hypertension, diabetes, smoking and body mass index), a dichotomized PRS based on common genetic variants (pink) and the HTRA1 protease domain variant carrier status (red) are shown. Effect sizes and standard errors were derived for the logWMH volume change using a linear regression model adjusting for age (except for the age categories), sex and 10 genetic principal components.

While *DCAKD* and *EGFL8* did not reach exome-wide significance for association with WMH burden, we consider these genes to exhibit strong evidence for a causal involvement in WMH. Recent GWAS found common variants at *DCAKD* to reach genome-wide significance for association with WMH, while leaving the responsible gene unidentified.[11,13] The current study pinpoints *DCAKD* as a likely causal gene in this region and further demonstrates a role of rare variants in *DCAKD*. A causal involvement of *EGFL8* is supported by the following observations. First, we found the gene product of *EGFL8* to be efficiently cleaved by HTRA1 in our protease assays, thus demonstrating a functional interaction between EGFL8 and a protease with an established role in SVD and cerebral WMH. Notably, common variants in *EFEMP1* encoding another substrate of HTRA1,[11–13,61] have consistently been associated with WMH in recent GWAS.[11–13,61] Second, we found human EGFL8 to be prominently expressed in the brain vasculature, which would be in line with the role of vascular mechanisms in WMH. The biological function of *EGFL8*, encoding EGF like domain multiple 8, remains largely unexplored. A recent study[66] found EGFL8 to act as neuritogen and rewire cellular signalling pathways by activating kinases involved in neurogenesis. Notably, *EGFL7*, a close paralogue of *EGFL8*, which is likewise expressed in brain vessels, has been implicated in various vascular functions including angiogenesis and elastogenesis.[67–69] *DCAKD*, encoding dephospho-CoA kinase domain containing protein, is ubiquitously expressed in brain, has a putative role in neurodevelopment and has previously been shown to be implicated in psychiatric disease.[70,71] Moreover, studies on expression quantitative loci have linked expression levels of *DCAKD* to white matter microstructure and WMH volume.[11,70]

Among the phenotypes shown here to be associated with rare variants in the HTRA1 protease domain is migraine with aura, thus pointing to a potential mechanistic link between WMH and migraine with aura. Previous population-based studies found a higher frequency of radiological WMH in migraineurs.[72–75] However, this was not consistently observed in other studies[76,77] and there was no clear pattern in terms of a preferential association with migraine with aura compared to migraine without aura.[72–75] Of further interest, common variant association studies reveal that several of the risk loci for migraine overlap with known risk loci for WMH including *COL4A1*, *COL4A2*, *NBEAL1-CARF*, and *ADAMTSL4*.[11,13,78] The association between rare variants in *HTRA1* and migraine with aura observed here further adds to recent GWAS that found common variants at *HTRA1* to be associated with migraine.[78,79] Studies in additional datasets are needed to better delineate the relationship between rare variants in *HTRA1* and migraine and a potential mechanistic link with WMH.

Another interesting finding is the observed association of HTRA1 protease domain carrier status and 'aneurysm and dissection of heart'. *LRP1*, a major risk gene for coronary artery dissection,[80,81] is a direct regulator of HTRA1 in the vasculature, and has been shown to play an important role in maintaining vessel integrity.[82] Interestingly, patients suffering from coronary artery dissections display a higher prevalence of migraine[83] and *LRP1* itself is a risk gene for migraine,[78] adding to the complex relationship between genetic variation in these genes and vascular disease.

The setting of the UKB enabled us to provide an estimate of the frequency of rare variants and their impact on WMH on a population level. Altogether, we found rare variants in the HTRA1 protease domain to occur at a frequency of ~1 in 400. As expected, some variants, such as R227W, were found to be more prevalent than others and effect sizes varied between variants, although these results need to be interpreted with caution given the low number of risk allele carriers. Contrasting with the current gene-burden tests that considered rare variants in *HTRA1*, common variants at *HTRA1* did not reach genome-wide significance for association with WMH in two recent GWAS that mostly included individuals

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

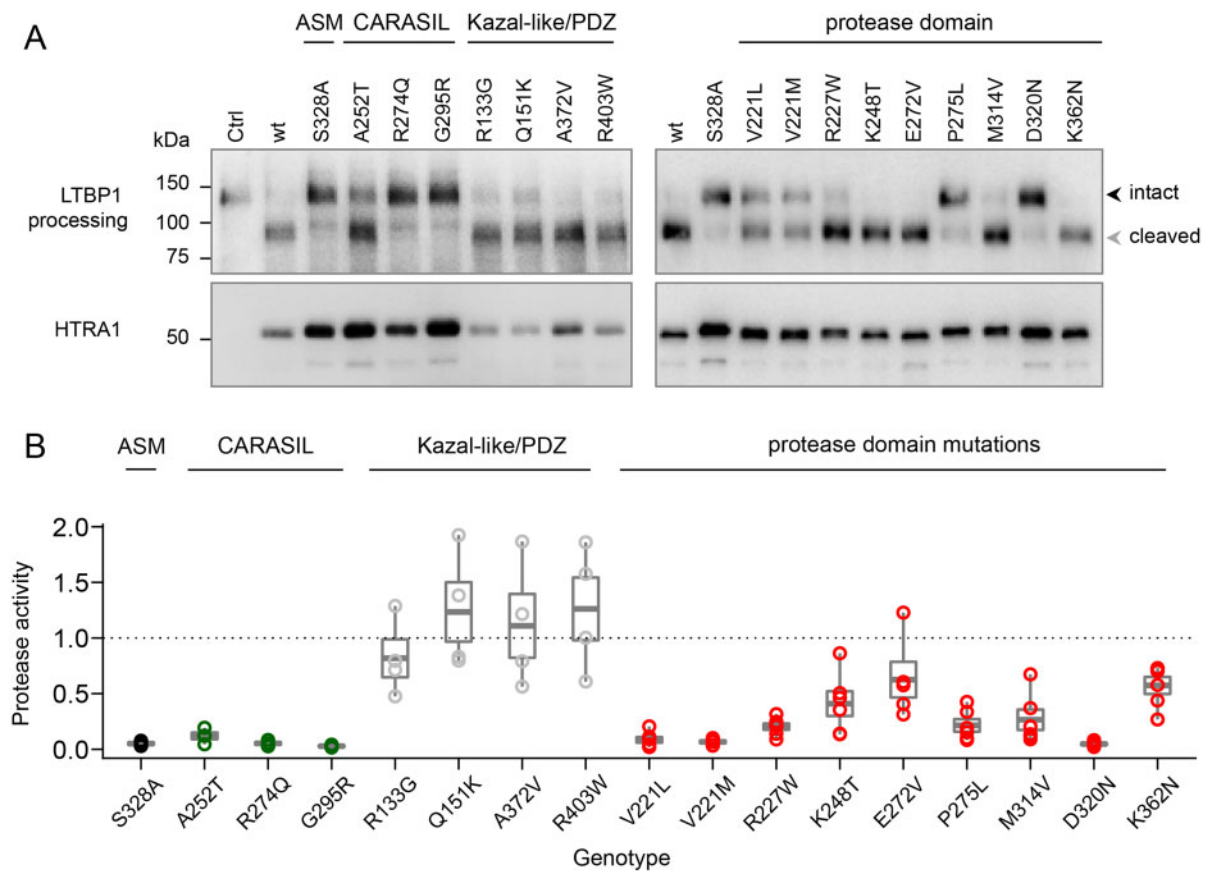

Figure 3 Functional consequences of rare *HTRA1* variants. (**A**) Culture medium from transfected cells was collected to assess HTRA1 expression (anti-Myc immunoblots) and protease activity towards LTBP1 (anti-V5 immunoblots). Black arrowhead = intact LTBP1; grey arrowhead = cleaved LTBP1. (**B**) Protease activity was determined as the ratio between cleaved and intact LTBP1 and was normalized to HTRA1 levels. Data are presented as box-and-whisker plots [median ± standard error of the mean (SEM); the activity of wild-type HTRA1 was set to 1]. Data-points represent independent measurements. At least three independent biological replicates were analysed. ASM = active site mutant.

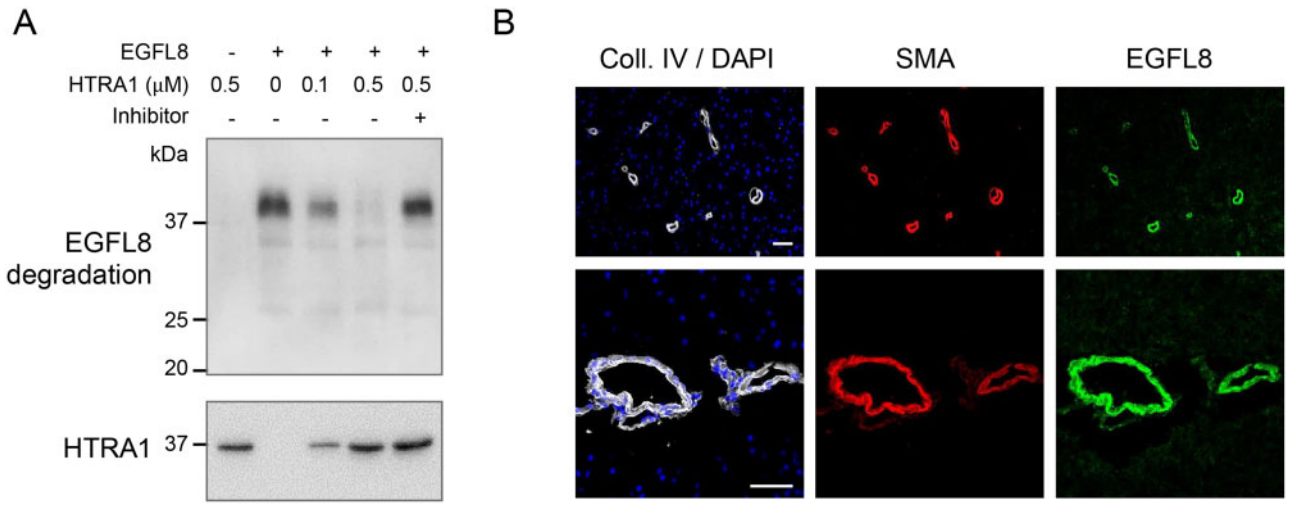

Figure 4 EGFL8 is a direct substrate of HTRA1 and expressed in the human brain vasculature. (**A**) EGFL8 was exposed to increasing concentrations of purified HTRA1 in the absence or presence of an HTRA1 inhibitor. EGFL8 and HTRA1 were detected by anti-Myc and anti-HTRA1 immunoblots, respectively. (**B**) Immunohistochemical detection of EGFL8 in human brain sections. The vascular basement membrane component collagen IV (Coll. IV) and the smooth muscle cell marker αSMA were detected as controls. Representative images are shown. Scale bar = 50 μm.

from the general population.[11,13] However, a recent candidate gene association study in population-based subjects with extreme WMH volumes found a common intronic variant in *HTRA1* (rs2293871; allele frequency: 19%) to be associated with WMH volume.[28] Also, *HTRA1* reached genome-wide significance in a multi-trait analysis that combined GWAS data on WMH and lacunar

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

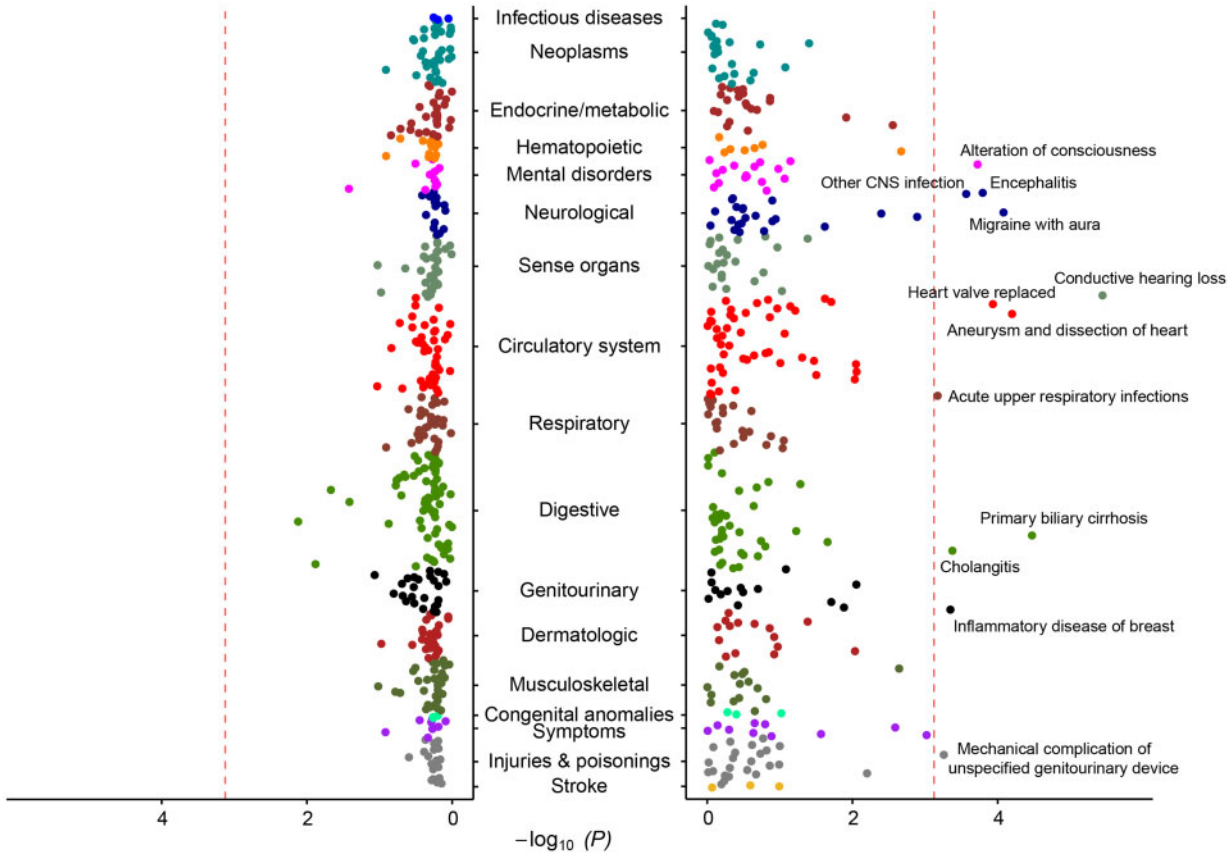

Figure 5 Rare variants in the HTRA1 protease domain associate with multiple neurological and non-neurological phenotypes. A PheWAS Manhattan plot for carriers of rare variants in the HTRA1 protease domain. All of the standardized Phecodes along with stroke-related phenotypes are shown. For phenotypes to the *right*, HTRA1 protease domain variant carriers show higher risk, for phenotypes to the *left*, carriers are at lower risk. The red dashed line represents the FDR 5% cut-off.

stroke.[17] Overall, these findings emphasize the relevance of *HTRA1* for vascular brain injury beyond rare hereditary arteriopathies.

Our single variant analysis revealed one variant that reached exome-wide significance and three additional variants that reached suggestive evidence for association with WMH. rs192044702 in *GBE1* has previously been reported in compound heterozygous individuals with glycogen storage disease IV,[84,85] which is allelic to adult polyglucosan body disease (APBD, OMIM #263570).[86] The majority of patients with this disease have WMH on MRI,[86,87] which seems at odds with our observation of a protective effect of rs192044702 on WMH burden. However, carriers of a single APBD mutation are generally asymptomatic and the effects by which variants in GBE1 might influence white matter integrity in one or the other direction remain poorly understood.

Our study has limitations. First, while the analysis of a quantitative trait offers higher statistical power than a binary phenotype our study still had limited power to detect small to moderate effect sizes. Also, the low number of alternative allele carriers may have resulted in inaccurate effect estimates with wide confidence intervals and potentially also false-positive findings, especially in our PheWAS analysis. Second, as with any WES approach, there remains a possibility of errors in calling very rare variants. However, this would be expected to bias the results from gene burden tests towards the null hypothesis whereas we found a high consistency in the directionality of effects for associated variants in *HTRA1*. Also, despite rigorous quality control of WES data in the UKB,[36] we cannot exclude the possibility of miscalled ultra-rare variants. Third, the results from our PheWAS might be more favorably powered for phenotypes with an early age-of-onset. Given the relatively young age of UKB participants at study inclusion (mean age 56.8 years) and the low number of participants with long-term follow up (median 8.98 years), associations with late-onset phenotypes such as stroke and dementia might have been missed. As a possible indication, we found rare variants in the HTRA1 protease domain to reach only nominal significance for Phecodes representing 'transient cerebral ischaemia' (Phecode 433.21) and 'occlusion of cerebral arteries' (Phecode 433.31). Fourth, the level of phenotyping for stroke in the UKB precluded analyses restricted to lacunar stroke which would be of considerable interest. Fifth, our PheWAS analysis was limited to hospital episode stay and death record ICD-10 codes mapped to Phecodes. Several symptoms and phenotypes of interest (e.g. alopecia) are not sufficiently reflected by these codes. Sixth, we did not assess longitudinal changes in WMH burden, as the number of individuals with available follow-up data on WMH in the UKB is still very low (~1500 individuals). Seventh, we focused on non-synonymous variants. Synonymous variants have been shown to impact transcription, splicing, folding or mRNA stability.[88] While these variants certainly exist, including all synonymous variants in our burden test would add noise to the results. Furthermore, we chose REVEL to provide pathogenicity predictions of single variants. While REVEL outperforms other methods,[89] there remains a risk of misclassification of non-synonymous variants. Eighth, our results cannot be extrapolated to other ethnicities and populations, which differ in terms of genetic architecture, vascular risk factors and environmental factors.[13,90,91] Ninth, in constructing the common variant PRS, we were only able to include genome-wide significant variants, as we did not have access to the full summary statistics. Using different

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

Table 3 *HTRA1* protease domain PheWAS

Phecode	Phenotype	P-value	FDR-adjusted P-value	Odds ratio	95% confidence interval
389.2	Conductive hearing loss	3.60×10^{-6}	0.0025	10.75	2.99–26.90
571.6	Primary biliary cirrhosis	3.37×10^{-5}	0.012	13.37	2.77–38.77
411.41	Aneurysm and dissection of heart	6.34×10^{-5}	0.015	6.68	2.22–15.20
340.1	Migraine with aura	8.31×10^{-5}	0.015	12.24	2.54–35.25
395.6	Heart valve replaced	0.00011	0.016	4.70	1.92–9.49
323	Encephalitis	0.00016	0.019	11.43	2.38–32.92
291.8	Alteration of consciousness	0.00018	0.019	7.68	2.14–19.13
324	Other CNS infection and poliomyelitis	0.00027	0.024	10.80	2.24–31.13
575.1	Cholangitis	0.00042	0.032	7.08	1.97–17.68
613.1	Inflammatory disease of breast	0.00044	0.032	7.04	1.96–17.64
857	Mechanical complication of unspecified genitourinary device, implant, graft	0.00055	0.036	4.62	1.74–9.78
465	Acute upper respiratory infections of multiple or unspecified sites	0.00067	0.040	4.53	1.70–9.58
	Any stroke	0.255	0.833	1.43	0.75–2.47
	Any ischaemic stroke	0.102	0.833	1.99	0.88–3.83
	Intracerebral haemorrhage	0.857	0.957	1.79	0.20–6.42
	Family history of stroke	0.145	0.833	1.85	0.76–2.93
	Family history of dementia	0.74	0.957	1.56	0.32–4.54

PheWAS analysis using Phecodes as the phenotypes of interest. Shown is the association of variants in the protease domain of HTRA1 with the individual Phecodes from a combined multivariate and collapsing burden test. Odds ratios were calculated using Firth's regression and rare allele carrier status as an independent variable.

variant selection criteria and algorithms might optimize the model and lead to higher power in the analysis. Finally, in the absence of available large datasets offering both WES and quantitative WMH assessment, we were unable to perform independent replication of rare variant or burden results. The upcoming release of the full UKB WES dataset will provide an excellent opportunity to replicate our findings in an identical setting.

Funding

This project received funding from the European Union's Horizon 2020 research and innovation programme (666881), SVDs@target (to M.D.; 667375), CoSTREAM (to M.D.); the DFG as part of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy: ID 390857198), the CRC 1123 (B3; to M.D.), DI 722/16-1 (project ID: 428668490), DI 722/13-1 und BE 6169/1-–1; and the Fondation Leducq (Transatlantic Network of Excellence on the Pathogenesis of Small Vessel Disease of the Brain; to M.D.). M.E. is funded by Deutsche Forschungsgemeinschaft (EH 100/18-1). K.R. is funded by a Health data Research UK Rutherford fellowship (MR/S004130/1). A.F. is funded by a BHF award RE/18/5/34216.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Garde E, Mortensen EL, Krabbe K, Rostrup E, Larsson HB. Relation between age-related decline in intelligence and cerebral white-matter hyperintensities in healthy octogenarians: A longitudinal study. *Lancet*. 2000;356(9230):628–634.
2. 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.
3. Debette S, Schilling S, Duperron MG, Larsson SC, Markus HS. Clinical significance of magnetic resonance imaging markers of vascular brain injury: A systematic review and meta-analysis. *JAMA Neurol*. 2019;76(1):81–94.
4. Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: Insights from neuroimaging. *Lancet Neurol*. 2013;12(5):483–497.
5. Dichgans M, Leys D. Vascular cognitive impairment. *Circ Res*. 2017;120(3):573–591.
6. Ladis Study Group, Poggesi A, Pantoni L, et al. 2001-2011: A decade of the LADIS (Leukoaraiosis And DISability) study: What have we learned about white matter changes and small-vessel disease? *Cerebrovasc Dis*. 2011;32(6):577–588.
7. Georgakis MK, Duering M, Wardlaw JM, Dichgans M. WMH and long-term outcomes in ischemic stroke: A systematic review and meta-analysis. *Neurology*. 2019;92(12):e1298–e1308.
8. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: Mechanisms and clinical implications. *Lancet Neurol*. 2019;18(7):684–696.
9. 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.
10. Duperron MG, Tzourio C, Sargurupremraj M, et al. Burden of dilated perivascular spaces, an emerging marker of cerebral small vessel disease, is highly heritable. *Stroke*. 2018;49(2):282–287.
11. Persyn E, Hanscombe KB, Howson JMM, Lewis CM, Traylor M, Markus HS. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat Commun*. 2020;11(1):2175.
12. Armstrong NJ, Mather KA, Sargurupremraj M, et al. Common genetic variation indicates separate causes for periventricular and deep white matter hyperintensities. *Stroke*. 2020;51(7):2111–2121.
13. Sargurupremraj M, Suzuki H, Jian X, et al.; International Headache Genomics Consortium (IHGC). Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun*. 2020;11(1):6285.

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

14. 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.
15. Neurology Working Group of the Cohorts for Heart Aging Research in Genomic Epidemiology Consortium, the Stroke Genetics Network, the International Stroke Genetics Consortium. Identification of additional risk loci for stroke and small vessel disease: A meta-analysis of genome-wide association studies. *Lancet Neurol*. 2016;15(7):695–707.
16. Malik R, Chauhan G, Traylor M, et al.; MEGASTROKE Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50(4):524–537.
17. Traylor M, Persyn E, Tomppo L, et al. Genetic basis of lacunar stroke: A pooled analysis of individual patient data and genome-wide association studies. *Lancet Neurol*. 2021;20(5):351–361.
18. Dichgans M, Pulit SL, Rosand J. Stroke genetics: Discovery, biology, and clinical applications. *Lancet Neurol*. 2019;18(6):587–599.
19. Dichgans M, Mayer M, Uttner I, et al. The phenotypic spectrum of CADASIL: Clinical findings in 102 cases. *Ann Neurol*. 1998;44(5):731–739.
20. 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.
21. 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.
22. Beaufort N, Scharrer E, Kremmer E, et al. Cerebral small vessel disease-related protease HtrA1 processes latent TGF-β binding protein 1 and facilitates TGF-β signaling. *Proc Natl Acad Sci U S A*. 2014;111(46):16496–16501.
23. Gould DB, Phalan FC, van Mil SE, et al. Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N Engl J Med*. 2006;354(14):1489–1496.
24. Verdura E, Herve D, Bergametti F, et al. Disruption of a miR-29 binding site leading to COL4A1 upregulation causes pontine autosomal dominant microangiopathy with leukoencephalopathy. *Ann Neurol*. 2016;80(5):741–753.
25. Jeanne M, Gould DB. Genotype-phenotype correlations in pathology caused by collagen type IV alpha 1 and 2 mutations. *Matrix Biol*. 2017;57-58:29–44.
26. Rannikmae K, Henshall DE, Thrippleton S, et al. Beyond the brain: Systematic review of extracerebral phenotypes associated with monogenic cerebral small vessel disease. *Stroke*. 2020;51(10):3007–3017.
27. Jian X, Satizabal CL, Smith AV, et al.; for the neuroCHARGE Working Group. 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.
28. Mishra A, Chauhan G, Violleau MH, et al. Association of variants in HTRA1 and NOTCH3 with MRI-defined extremes of cerebral small vessel disease in older subjects. *Brain*. 2019;142(4):1009–1023.
29. Rutten JW, Hack RJ, Duering M, et al. Broad phenotype of cysteine-altering NOTCH3 variants in UK Biobank: CADASIL to non-penetrance. *Neurology*. 2020;95(13):e1835–e1843.
30. Cho BPH, Nannoni S, Harshfield EL, et al. NOTCH3 variants are more common than expected in the general population and associated with stroke and vascular dementia: An analysis of 200 000 participants. *J Neurol Neurosurg Psychiatry*. 2021;92(7):694–701.
31. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–209.
32. Littlejohns TJ, Holliday J, Gibson LM, et al. The UK Biobank imaging enhancement of 100,000 participants: Rationale, data collection, management and future directions. *Nat Commun*. 2020;11(1):2624.
33. Flannick J, Mercader JM, Fuchsberger C, et al.; AMP-T2D-GENES. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature*. 2019;570(7759):71–76.
34. Wang Q, Dhindsa RS, Carss K, et al. Surveying the contribution of rare variants to the genetic architecture of human disease through exome sequencing of 177,882 UK Biobank participants. *bioRxiv*. [Preprint] doi:10.1101/2020.12.13.422582
35. Jurgens SJ, Choi SH, Morrill VN, et al. Rare Genetic Variation Underlying Human Diseases and Traits: Results from 200,000 Individuals in the UK Biobank. *bioRxiv*. [Preprint] doi:10.1101/2020.11.29.402495
36. Szustakowski JD, Balasubramanian S, Kvikstad E, et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. Nat Genet. 2021;53(7):942–948.
37. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867–2873.
38. Staples J, Qiao D, Cho MH, et al.; University of Washington Center for Mendelian Genomics. PRIMUS: Rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am J Hum Genet*. 2014;95(5):553–564.
39. Griffanti L, Zamboni G, Khan A, et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. *Neuroimage*. 2016;141:191–205.
40. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19(11):1523–1536.
41. Alfaro-Almagro F, Jenkinson M, Bangerter NK, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage*. 2018;166:400–424.
42. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021;53(7):1097–1103.
43. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122.
44. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877–885.
45. Karczewski KJ, Francioli LC, Tiao G, et al.; Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.
46. Consortium U. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506–D515.
47. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics*. 2016;32(9):1423–1426.
48. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: Study designs and statistical tests. *Am J Hum Genet*. 2014;95(1):5–23.
49. Cox SR, Lyall DM, Ritchie SJ, et al. Associations between vascular risk factors and brain MRI indices in UK Biobank. *Eur Heart J*. 2019;40(28):2290–2300.
50. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol*. 2010;34(2):188–193.
51. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011;89(1):82–93.

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024

52. Lee S, Emond MJ, Bamshad MJ, et al.; NHLBI GO Exome Sequencing Project—ESP Lung Project Team. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*. 2012;91(2):224–237.
53. Poepsel S, Sprengel A, Sacca B, et al. Determinants of amyloid fibril degradation by the PDZ protease HTRA1. *Nat Chem Biol*. 2015;11(11):862–869.
54. Grau S, Baldi A, Bussani R, et al. Implications of the serine protease HtrA1 in amyloid precursor protein processing. *Proc Natl Acad Sci U S A*. 2005;102(17):6021–6026.
55. Wu P, Gifford A, Meng X, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform*. 2019;7(4):e14325.
56. Verhaaren BF, 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.
57. Nozaki H, Kato T, Nihonmatsu M, et al. Distinct molecular mechanisms of HTRA1 mutants in manifesting heterozygotes with CARASIL. *Neurology*. 2016;86(21):1964–1974.
58. Uemura M, Nozaki H, Kato T, et al. HTRA1-related cerebral small vessel disease: A review of the literature. *Front Neurol*. 2020;11:545.
59. Eigenbrot C, Ultsch M, Lipari MT, et al. Structural and functional analysis of HtrA1 and its subdomains. *Structure*. 2012;20(6):1040–1050.
60. 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.
61. Lin MK, Yang J, Hsu CW, et al. HTRA1, an age-related macular degeneration protease, processes extracellular matrix proteins EFEMP1 and TSP1. *Aging Cell*. 2018;17(4):e12710.
62. Emdin CA, Khera AV, Chaffin M, et al. Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat Commun*. 2018;9(1):1613.
63. Verdura E, Herve D, Scharrer E, et al. Heterozygous HTRA1 mutations are associated with autosomal dominant cerebral small vessel disease. *Brain*. 2015;138(Pt 8):2347–2358.
64. Fasano A, Formichi P, Taglia I, et al. HTRA1 expression profile and activity on TGF-β signaling in HTRA1 mutation carriers. *J Cell Physiol*. 2020;235(10):7120–7127.
65. Xie F, Zhang LS. A Chinese CARASIL patient caused by novel compound heterozygous mutations in HTRA1. *J Stroke Cerebrovasc Dis*. 2018;27(10):2840–2842.
66. Weiss T, Taschner-Mandl S, Janker L, et al. Schwann cell plasticity regulates neuroblastic tumor cell differentiation via epidermal growth factor-like protein 8. *Nat Commun*. 2021;12(1):1624.
67. Parker LH, Schmidt M, Jin SW, et al. The endothelial-cell-derived secreted factor Egfl7 regulates vascular tube formation. *Nature*. 2004;428(6984):754–758.
68. Nichol D, Stuhlmann H. EGFL7: A unique angiogenic signaling factor in vascular development and disease. *Blood*. 2012;119(6):1345–1352.
69. Lelievre E, Hinek A, Lupu F, Buquet C, Soncin F, Mattot V. VE-statin/egfl7 regulates vascular elastogenesis by interacting with lysyl oxidases. *EMBO J*. 2008;27(12):1658–1670.
70. Barbu MC, Spiliopoulou A, Colombo M, et al. Expression quantitative trait loci-derived scores and white matter microstructure in UK Biobank: A novel approach to integrating genetics and neuroimaging. *Transl Psychiatry*. 2020;10(1):55.
71. Bipolar Disorder Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell*. 2018;173(7):1705–1715.e16.
72. Kruit MC, van Buchem MA, Hofman PA, et al. Migraine as a risk factor for subclinical brain lesions. *JAMA*. 2004;291(4):427–434.
73. Kurth T, Mohamed S, Maillard P, et al. Headache, migraine, and structural brain lesions and function: Population based epidemiology of vascular ageing-MRI study. *BMJ*. 2011;342:c7357.
74. Hamedani AG, Rose KM, Peterlin BL, et al. Migraine and white matter hyperintensities: The ARIC MRI study. *Neurology*. 2013;81(15):1308–1313.
75. Palm-Meinders IH, Koppen H, Terwindt GM, et al. Structural brain changes in migraine. *JAMA*. 2012;308(18):1889–1897.
76. Monteith T, Gardener H, Rundek T, et al. Migraine, white matter hyperintensities, and subclinical brain infarction in a diverse community: The northern Manhattan study. *Stroke*. 2014;45(6):1830–1832.
77. Gaist D, Garde E, Blaabjerg M, et al. Migraine with aura and risk of silent brain infarcts and white matter hyperintensities: An MRI study. *Brain*. 2016;139(Pt 7):2015–2023.
78. Hautakangas H, Winsvold BS, Ruotsalainen SE, et al. Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. *medRxiv*. [Preprint] doi:10.1101/2021.01.20.21249647
79. Gormley P, Anttila V, Winsvold BS, et al.; International Headache Genetics Consortium. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet*. 2016;48(8):856–866.
80. Turley TN, O'Byrne MM, Kosel ML, et al. Identification of susceptibility loci for spontaneous coronary artery dissection. *JAMA Cardiol*. 2020;5(8):929–938.
81. Saw J, Yang ML, Trinder M, et al.; Million Veteran Program. Chromosome 1q21.2 and additional loci influence risk of spontaneous coronary artery dissection and myocardial infarction. *Nat Commun*. 2020;11(1):4432.
82. Muratoglu SC, Belgrave S, Hampton B, et al. LRP1 protects the vasculature by regulating levels of connective tissue growth factor and HtrA1. *Arterioscler Thromb Vasc Biol*. 2013;33(9):2137–2146.
83. Kok SN, Hayes SN, Cutrer FM, et al. Prevalence and clinical factors of migraine in patients with spontaneous coronary artery dissection. *J Am Heart Assoc*. 2018;7(24):e010140.
84. Fernandez C, Halbert C, De Paula AM, et al. Non-lethal neonatal neuromuscular variant of glycogenosis type IV with novel GBE1 mutations. *Muscle Nerve*. 2010;41(2):269–271.
85. Ravenscroft G, Thompson EM, Todd EJ, et al. Whole exome sequencing in foetal akinesia expands the genotype-phenotype spectrum of GBE1 glycogen storage disease mutations. *Neuromuscul Disord*. 2013;23(2):165–169.
86. Mochel F, Schiffmann R, Steenweg ME, et al. Adult polyglucosan body disease: Natural history and key magnetic resonance imaging findings. *Ann Neurol*. 2012;72(3):433–441.
87. Lopez Chiriboga AS. Teaching NeuroImages: Prominent spinal cord atrophy and white matter changes in adult polyglucosan body disease. *Neurology*. 2017;88(20):e194–e195.
88. Rauscher R, Ignatova Z. Timing during translation matters: Synonymous mutations in human pathologies influence protein folding and function. *Biochem Soc Trans*. 2018;46(4):937–944.
89. Gunning AC, Fryer V, Fasham J, et al. Assessing performance of pathogenicity predictors using clinically relevant variant datasets. *J Med Genet*. 2021;58(8):547–555.
90. Brickman AM, Schupf N, Manly JJ, et al. Brain morphology in older African Americans, Caribbean Hispanics, and whites from northern Manhattan. *Arch Neurol*. 2008;65(8):1053–1061.
91. Zahodne LB, Manly JJ, Narkhede A, et al. Structural MRI predictors of late-life cognition differ across African Americans, Hispanics, and Whites. *Curr Alzheimer Res*. 2015;12(7):632–639.

Downloaded from https://academic.oup.com/brain/article/144/9/2670/6385778 by guest on 07 June 2024