

Whole-exome sequencing in 16,511 individuals reveals a role of the HTRA1 protease and its substrate 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 LM Sudlow ^{2,7,8}, Martin Dichgans ^{1,9,10}

¹ Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany

² Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK

³ Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London, 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, Germany

⁶ School of Biosciences, Cardiff University, Cardiff, UK

⁷ Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

⁸ Health Data Research UK Scotland, University of Edinburgh, Edinburgh, UK

⁹ Munich Cluster for Systems Neurology, Munich, Germany

¹⁰ German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

* These authors contributed equally

Correspondence should be addressed to:

Prof. Martin Dichgans

Institute for Stroke and Dementia Research (ISD)

Feodor-Lynen-Str. 17

81377 München

Germany

M: Martin.dichgans@med.uni-muenchen.de

P: +49-89-4400-46018

Word count: 5,888

Figures: 5

Tables: 3

ABSTRACT

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 WMH volume, the contribution of rare variants to WMH burden in the general population remains largely unexplored. We conducted a comprehensive analysis of WMH burden in the UK Biobank using publicly available whole-exome sequencing data (N=16,511) 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, beta=-0.74, se=0.13, p=9.7E-9). Applying whole-exome gene-based burden tests, we found damaging missense and loss-of-function variants in *HTRA1* to associate with increased WMH volume (p=5.5E-6, FDR=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; beta=0.79, se=0.14, p=9.4E-8). The frequency of such variants in the UK Biobank population was 1 in 450. WMH volume was brought forward by approximately 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 (beta=0.26, se=0.02, p=2.9E-59) or being in the upper 99.8% percentile of the distribution of a polygenic risk score based on common genetic variants (beta=0.44, se=0.14, p=0.002). In biochemical experiments, most (6/9) of the identified protease domain variants resulted in a markedly reduced protease activity. We further found *EGFL8*, which showed suggestive evidence for association with WMH volume (p=1.5E-4, FDR=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 (PheWAS) 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 (OR=12.24, 95%CI [2.54-35.25], p=8.3E-5). Collectively, these findings highlight an important role of rare genetic variation and of the HTRA1 protease in determining WMH burden in the general population.

INTRODUCTION

White matter hyperintensities (WMH) are common in elderly individuals ([Garde *et al.*, 2000](#); [Debette and Markus, 2010](#)) and an increasingly recognized risk factor for stroke ([Wardlaw *et al.*, 2013](#); [Debette *et al.*, 2019](#)), dementia ([Wardlaw *et al.*, 2013](#); [Dichgans and Leys, 2017](#); [Debette *et al.*, 2019](#)), and functional decline in the older age ([Ladis Study Group, 2011](#)). They are further associated with poor long-term outcomes after ischemic stroke including an increased risk of stroke recurrence, dementia, and mortality ([Georgakis *et al.*, 2019](#)). While not specific for any particular etiology, WMH are considered a marker of cerebral small vessel disease (SVD) ([Dichgans and Leys, 2017](#); [Wardlaw *et al.*, 2019](#)). Hypertension is the single strongest treatable risk factor for WMH ([Dufouil *et al.*, 2001](#); [Wardlaw *et al.*, 2013](#); [Georgakis *et al.*, 2019](#)). However, the precise mechanisms underlying WMH remain largely elusive.

WMH are highly heritable with estimates ranging from 18-54% ([Duperron *et al.*, 2018](#); [Persyn *et al.*, 2020](#)). Accordingly, recent genome-wide association studies (GWAS) found common genetic variants at multiple loci to be associated with WMH burden ([Fornage *et al.*, 2011](#); [Armstrong *et al.*, 2020](#); [Persyn *et al.*, 2020](#); [Sargurupremraj *et al.*, 2020](#)). Besides confirming a causal link with hypertension ([Sargurupremraj *et al.*, 2020](#)) these studies also pinpoint specific molecular pathways and biological mechanisms. Among the most prominent themes are perturbations of the extracellular matrix (ECM) 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 ([Neurology Working Group of the Cohorts for Heart Aging Research in Genomic Epidemiology Consortium *et al.*, 2016](#); [Persyn *et al.*, 2020](#); [Sargurupremraj *et al.*, 2020](#)). The results from recent GWAS on WMH further highlight links with small vessel stroke, ischemic stroke, intracerebral hemorrhage, and neurodegenerative disease ([Malik *et al.*, 2018](#); [Sargurupremraj *et al.*, 2020](#); [Traylor *et al.*, 2021](#)). As such, large scale genetic studies have been instrumental in uncovering core pathways and mechanisms

underlying WMH and defining relationships with related phenotypes, in particular SVD.

A close link between WMH and SVD is further supported by observations in familial forms of SVD (Dichgans *et al.*, 2019). 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) (Joutel *et al.*, 1996; Dichgans *et al.*, 1998). 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 (Hara *et al.*, 2009). *HTRA1* encodes a secreted serine protease that associates with the ECM and has been shown to proteolytically process various molecular constituents of the ECM including latent TGF- β binding protein, vitronectin, and elastin (Beaufort *et al.*, 2014). The outstanding importance of perturbations of the ECM in the pathophysiology of WMH is further illustrated by highly penetrant mutations in *COL4A1* and *COL4A2* (Gould *et al.*, 2006; Verdura *et al.*, 2016; Jeanne and Gould, 2017). These mutations cause a broad range of phenotypes including WMH, small vessel stroke, intracerebral hemorrhage, porencephalopathy, and extracerebral manifestations (Verdura *et al.*, 2016; Jeanne and Gould, 2017). Notably, extracerebral manifestations such as ocular and renal manifestations are a common feature in various forms of hereditary SVD (Rannikmae *et al.*, 2020).

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 variation 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 (Jian *et al.*, 2018) and on genes implicated in familial SVD (Mishra *et al.*, 2019; Rutten *et al.*, 2020; Cho *et al.*, 2021). Of note, some of them included samples that had been selected on the basis of extreme phenotypes (Mishra *et al.*, 2019).

The UK Biobank (UKB) (Bycroft *et al.*, 2018) is a large-scale (~500.000 participants) prospective community-based study that recruited 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 in about

44,000 individuals ([Littlejohns et al., 2020](#)). The UKB further offers detailed clinical information mapped to international classification of diseases (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 comprehensive analysis of all genetic variation in coding regions, including rare variants that have previously been inaccessible. In many complex traits and diseases, studying to this variation has contributed to understanding their genetic basis ([Flannick et al., 2019](#); [Jurgens et al., 2020](#); [Wang et al., 2020](#)).

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

This analysis was carried out under the UKB project 2532.

UK Biobank population

Our resource was the UK Biobank Exome 200k release from October 2020. Primary and secondary analyses were performed with an updated Functional Equivalence (FE) protocol that retains original quality scores in the CRAM files (referred to as the OQFE protocol) ([Szustakowski et al., 2020](#)).

The 200k release encompasses 1,135 parent-offspring pairs, 3,855 full-sibling pairs, including 101 trios, 27 monozygotic twin pair and 7,461 second degree genetically determined relationships. To avoid bias due to relatedness, we selected an unrelated set of individuals up to a 2nd degree (KING cut-off 0.0877) ([Manichaikul et al., 2010](#)). 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 ([Staples et al., 2014](#)). For PheWAS analyses, we selected a standard unrelated set (also up to a 2nd degree) which was used for all subsequent analyses (N=166,897). We further excluded individuals showing an excess of heterozygosity from the UK Biobank 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 UK Biobank imaging data on approximately 44,000 individuals. MRI was performed on two identical Siemens Skyra 3.0 T scanners (Siemens Medical Solutions, Germany), 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 T2-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 UK Biobank central analysis (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf) ([Griffanti et al., 2016](#); [Miller et al., 2016](#); [Alfaro-Almagro et al., 2018](#)). We excluded three extreme volume outliers (>6SD) and used log-transformed WMH volume to ensure normal distribution.

Single Variant analysis

Single variant analysis of rare variants in the WES data was done using REGENIE (Mbatchou *et al.*, 2020). 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 (SPA) regression was used in favor of Firth's correction due to performance issues. We set a conservative exome-wide threshold of $5E-8$ as significance cut-off in our analysis.

Whole-exome burden test on white matter hyperintensities (WMH)

We determined the functional consequences of exome-wide variants using the Variant Effect Predictor (VEP) tool v101 (McLaren *et al.*, 2016). For the whole-exome burden test on WMH analysis, we selected rare variants (MAF < 0.01) that are either predicted to be damaging by REVEL (Ioannidis *et al.*, 2016) (REVEL score > 0.5) or predicted to exert a high-confidence loss-of-function effect using the LoFTEE (Karczewski *et al.*, 2020) plugin in VEP.

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 (UniProt Consortium, 2019): 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 test were done using *rvtests* (Zhan *et al.*, 2016) and by calculating multivariable burden tests using the CMC burden schema (Lee *et al.*, 2014). Sex, age at imaging and 10 genomic principal components (PCs) were used as covariates in all analyses. For the primary logWMH analysis, we additionally corrected for vascular risk factors (SBP, LDL, BMI, Hba1C) in a sensitivity analysis. Results were deemed significant at an FDR level 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 PCs as covariates.

To compare effect sizes to dichotomised risk factors for WMH, we extracted the following information from the UKB: hypertension status (yes/no), defined as a systolic blood pressure (SBP) >140 mmHg or a diastolic blood pressure (DBP) >90mmHg (UKB fields 4080 and 4079) taking the mean of available measurements

and correcting SBP and DBP for antihypertensive drug use (UKB field 2003) as appropriate; diabetes mellitus (yes/no, UKB field 2443); BMI (above mean vs. below mean; UKB field 21001); smoking (ever smoked vs never smoked; UKB field 20116). We further calculated a polygenic risk score (PRS) based on common genetic variants from the most recent WMH GWAS ([Sargurupremraj et al., 2020](#)). 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 vs below the mean, with a PRS >95% percentile vs PRS <95% percentile, and a PRS > 99.8% percentile vs 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 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) were cloned into a pcDNA4/TO/myc-His expression plasmid (Invitrogen). The cDNA encoding the N-terminal region of human LTBP1 (Latent-transforming growth factor beta-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's medium (DMEM) containing GlutaMAX, 10% (v/v) fetal 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 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. Culture medium was collected and centrifuged at 1,000 g for 15 min, dialyzed at room temperature against 0.5 x phosphate-buffered saline (PBS) for 2 h and overnight at 4°C against 0.5 x PBS containing 200 mM NaCl. 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 washes,

EGFL8 was eluted with 100 mM EDTA in PBS and the eluate dialyzed overnight at 4°C against 1,000 volume of Tris 50 mM, NaCl 150 mM, pH 8.0.

HTRA1 purification. Human HTRA1 wild type (wt) lacking the N-terminal Mac domain (AA 158-480) was produced as described previously ([Poepsel et al., 2015](#)) with small modifications. N-terminally StrepII-tagged wt HTRA1 was affinity-purified with a streptactin resin material (IBA lifesciences) and subsequently subjected to size exclusion chromatography using a Superdex 200 preparation grade column (GE Healthcare) in 20 mM HEPES, 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 was normalized to HTRA1 expression, assessed by anti-Myc immunoblot.

EGFL8 was treated with 100 or 500 nM purified wt 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 ([Grau et al., 2005](#)). 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) skim milk and probed with anti-Myc (Santa Cruz Biotechnology, #sc-40, 1:5,000), anti-V5 (Invitrogen, #R960-25, 1:10,000) or anti-HTRA1 (R&D Systems, #MAB2916, 1:5,000) primary antibodies prepared in TBS/Tween/milk. Detection was performed with Horseradish peroxidase (HRP)-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-

Maximilians-University Munich) and analyzed according to the guidelines of the local ethical committees. For EGFL8 detection, ten μm tissue sections were thawed to room temperature (RT) and fixed for 20 min in 4% (w/v) paraformaldehyde then permeabilized and blocked in PBS added with 0.1% (v/v) Triton-X100 and 5% (w/v) bovine serum albumin (BSA) for 1 h at RT. Primary antibodies were diluted in PBS/Triton/BSA or PBS/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. Following washing with PBS, sections were incubated with Alexa Fluor 488- or 647-coupled secondary antibodies (Abcam, 1:100) and 4',6-Diamidin-2-phenylindol (DAPI, ThermoFisher Scientific) diluted in PBS for 1 h at RT. 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 UK Biobank, we used the Phecode Map 1.2 to map UKB ICD10-codes to Phecodes ([Wu et al., 2019](#)). We used all ICD10 codes (main position, secondary position, death records) from the UKB. We excluded Phecodes with <100 cases and Phecodes that are male- or female-specific. Individuals were assigned a case status if >1 ICD10 code mapped to the respective Phecode. Individuals meeting the pre-specified exclusion criteria were removed from the analysis, otherwise the individual was assigned a control status. Given the known association between white matter hyperintensities and stroke, we further analyzed any stroke (AS), any ischemic stroke (AIS) and intracerebral hemorrhage (ICH) as defined by the algorithmically defined stroke outcomes (UKB fields 42006-42013). Results were deemed significant at an FDR level of 5% after FDR correction. To approximate 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 PCs as covariates. We used Firth's correction for unbalanced case/control ratios in our logistic regression analysis.

RESULTS

We analyzed up to 17,830 individuals from the UK Biobank with both WES data and quantitative data on WMH volume available. Their demographic characteristics are presented in **Supp 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 (PCs), we found one rare variant (MAF < 0.01) to be associated at genome-wide significance level ($p < 5 \times 10^{-8}$) with lower logWMH volume ($\beta = -0.74$, $se = 0.13$) in 17,830 individuals. This variant is a predicted splice donor loss-of-function (LoF) 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 (**Supp Table 2**).

Whole-exome burden test reveals *HTRA1* as a risk gene for WMH burden

We next performed a whole-exome burden test using the Combined Multivariate and Collapsing (CMC) method implemented in the rvtsts software.—Analyses were corrected for sex, age at imaging, and 10 genomic PCs as covariates with 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 analyzed 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 association at an FDR level of 25% (**Table 1, Supp Table 3, Supp Fig 1**). To estimate the effects of carrier status on logWMH volume we performed linear regression analysis on log transformed WMH volume and minor allele carrier status, corrected for sex, age at imaging, and 10 genomic PCs. Carriers of rare variants in either *HTRA1*, *EGFL8* or *RGS12* showed increased WMH volumes compared to non-carriers (*HTRA1*: $\beta = 0.36$, $se = 0.09$, $p = 0.0003$; *EGFL8*: $\beta = 0.22$, $se = 0.06$, $p = 9.6 \times 10^{-5}$; *RGS12*: $\beta = 0.29$, $se = 0.10$, $p = 0.003$,

respectively), whereas carriers of rare variants in *GBE1* (beta=-0.20, se=0.05, p=6.7E-5) and *DCAKD* (beta=-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 (SBP, LDL, Hba1C, BMI and smoking) (**Supp Table 4**) In an additional sensitivity analysis, we only considered LoF 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 missense and LoF analysis. (**Supp Table 5**) In contrast, the association of *HTRA1*, which was based on only 9 LoF variants, was greatly diminished (p=0.97), and therefore, *HTRA1* LoF variants were not considered in further analyses. There was no association with LoF alleles in *RGS12* (p=0.69).

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 CMC 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=9.5E-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). Baseline characteristics of protease domain variant carriers did not differ from those of non-carriers (**Supp Table 6**). We also calculated results for the combined protease and linker domain as the linker domain (AA 158-204; 2 variants) serves as an activator of protease function and found the association to be slightly attenuated (p=5.5E-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 PCs. Carrier status of an alternative allele in the HTRA1 protease domain was associated with an increase in logWMH volume (beta=0.79, se= 0.14, p=9.4E-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**). When compared to other dichotomized risk factors for WMH volume, protease domain mutation carrier status showed a larger effect than hypertension (beta=0.26, se=0.02), diabetes (beta=0.25, se=0.04), BMI (beta=0.18, se=0.01) and smoking (beta=0.10, se=0.01). Also, comparison to individuals on a polygenic risk score (PRS) distribution based on 27 common genetic variants ([Sargurupremraj et al., 2020](#)) showed a larger effect size for carrier status of an alternative allele in the HTRA1 protease domain than for individuals with a PRS above vs below the 99.8% percentile of the PRS distribution (beta=0.44, se=0.14) (**Fig 2**).

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

To further 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 (wt) HTRA1 and an active site mutant (ASM) obtained by replacement of the catalytic Ser by an Ala (S328A) served as positive and negative controls, respectively. We investigated the enzymatic activity of 2 Kazal-like domain, 2 PDZ domain and 9 protease domain mutants. Three known CARASIL loss-of-function mutants were also included in the assays ([Beaufort et al., 2014](#); [Nozaki et al., 2016](#)). We found the protease activity of Kazal-like and PDZ domain mutants to be comparable to that of wt HTRA1, in accord with the fact that these domains are dispensable for protease activity ([Eigenbrot et al., 2012](#); [Uemura et al., 2020](#)). In contrast, 6 out of 9 protease domain mutants exhibited a marked reduction of protease activity (V221L, V221M, R227W, P275L, M314V, D320N).

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 result 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 (Zellner *et al.*, 2018) that found 38 proteins to be enriched in the brain vasculature of *Htra1*^{-/-} compared to wild-type mice, among them several verified HTRA1 substrates including Efemp1, Vtn, and Ltbp1 (Lin *et al.*, 2018; Zellner *et al.*, 2018) (Supp Table 7). 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% ($p=1.54E-4$).

We next set out to experimentally determine whether the human EGFL8 protein is a substrate of HTRA1. Treatment of EGFL8 with HTRA1 resulted in a dose-dependent degradation of EGFL8, which was prevented in the presence of a selective HTRA1 inhibitor (Fig 4A). These observations thus identify EGFL8 as a novel HTRA1 substrate. We further examined the cerebral expression pattern of EGFL8 by immunohistochemistry. In the human cortex, EGFL8 was predominantly localized to the alpha smooth muscle actin 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 (Rannikmae *et al.*, 2020). To comprehensively assess disease outcomes associated with rare variants in *HTRA1* or *EGFL8*, we performed a rare-variant burden PheWAS on standardized Phecodes in the UK Biobank. Specifically, we matched patients' ICD10 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 analyzed in burden tests. We further performed analyses for any stroke (AS), any ischemic stroke (AIS) and intracerebral hemorrhage (ICH) as defined by the algorithmically defined stroke outcomes (UKB fields 42006-42013) and for parental history of stroke and family history of dementia (UKB field 20107 and 20110).

Considering all identified damaging and LoF variants in *HTRA1* (**Supp Tables 8 and 9**), we found conductive hearing loss (Phecode 389.2) to be associated at an FDR level of 5% (**Supp Table 10, Supp Fig 2**) with minor allele carriers having a higher risk in Firth logistic regression (OR: 5.37, CI95 [2.39-9.49]). Focusing on missense variants in the protease domain we identified 12 Phecodes at an FDR level of 5% (**Fig 5, Table 3 and Supp Table 11**) 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. For *EGFL8*, we identified two Phecodes at an FDR level of 5%: 578.8 (Hemorrhage of rectum and anus) and 696.41 (Psoriasis vulgaris), the latter confirming prior work in the UKB ([Emdin et al., 2018](#)) (**Supp Table 12, Supp Fig 3**).

DISCUSSION

Using whole-exome sequencing data from the UK Biobank, we show that rare variants in *HTRA1* are associated with a higher burden of radiological white matter hyperintensities in the general population. We further show that the association with WMH burden is largely restricted to variants affecting the HTRA1 protease domain and provide experimental evidence that variants giving the strongest association signal result in a loss of proteolytic activity of HTRA1. Our PheWAS results reveal 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 demonstrate 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 in disease mechanisms.

HTRA1 has been implicated in familial small vessel disease (SVD). Biallelic mutations in *HTRA1* cause cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL), an early-onset condition that is characterized by stroke, dementia, spondylosis deformans, and alopecia ([Hara et al., 2009](#); [Beaufort et al., 2014](#)). 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 ([Verdura et al., 2015](#)). Our results extend these findings by showing that rare variants in *HTRA1*, and more specifically variants in the HTRA1 protease domain, are associated with 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 ([Beaufort et al., 2014](#); [Verdura et al., 2015](#)).

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 (Hara *et al.*, 2009; Fasano *et al.*, 2020) were not evaluated. Also, we did not go further in examining possible dominant negative effects, as recently reported for a subset of *HTRA1* variants linked to hereditary SVD (Nozaki *et al.*, 2016; Uemura *et al.*, 2020). Of interest, one of the variants (D320N) detected here has previously been reported in an Asian CARASIL patient (Xie and Zhang, 2018) 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.

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 (Persyn *et al.*, 2020; Sargurupremraj *et al.*, 2020). 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 (Lin *et al.*, 2018; Armstrong *et al.*, 2020; Persyn *et al.*, 2020; Sargurupremraj *et al.*, 2020), have consistently been associated with WMH in recent GWAS (Lin *et al.*, 2018; Armstrong *et al.*, 2020; Persyn *et al.*, 2020; Sargurupremraj *et al.*, 2020). 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 (Weiss *et al.*, 2021) found EGFL8 to act as neuritogen and to rewire cellular signaling 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 ([Parker et al., 2004](#); [Lelievre et al., 2008](#); [Nichol and Stuhlmann, 2012](#)). *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 ([Bipolar Disorder Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2018](#); [Barbu et al., 2020](#)). Moreover, studies on expression quantitative loci (eQTL) have linked expression levels of *DCAKD* to white matter microstructure and WMH volume ([Barbu et al., 2020](#); [Persyn et al., 2020](#)).

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 ([Kruit et al., 2004](#); [Kurth et al., 2011](#); [Palm-Meinders et al., 2012](#); [Hamedani et al., 2013](#)). However, this was not consistently observed in other studies ([Monteith et al., 2014](#); [Gaist et al., 2016](#)) and there was no clear pattern in terms of a preferential association with migraine with aura compared to migraine without aura ([Kruit et al., 2004](#); [Kurth et al., 2011](#); [Palm-Meinders et al., 2012](#); [Hamedani et al., 2013](#)). 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* ([Persyn et al., 2020](#); [Sargurupremraj et al., 2020](#); [Hautakangas et al., 2021](#)). 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 ([Gormley et al., 2016](#); [Hautakangas et al., 2021](#)). 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.

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 about 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 from the general population

(Persyn *et al.*, 2020; Sargurupremraj *et al.*, 2020). 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 (Mishra *et al.*, 2019). Also, *HTRA1* reached genome-wide significance in a multi-trait analysis that combined GWAS data on WMH and lacunar stroke (Traylor *et al.*, 2021). 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 (Fernandez *et al.*, 2010; Ravenscroft *et al.*, 2013), which is allelic to adult polyglucosan body disease (APBD, OMIM #263570) (Mochel *et al.*, 2012). The majority of patients with APBD have WMH on MRI (Mochel *et al.*, 2012; Lopez Chiriboga, 2017), 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. 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 QC of WES data in the UKB (Szustakowski *et al.*, 2020), we cannot exclude the possibility of mis-called 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 56.8yrs) and the low number of participants with long-term follow up (median 8.98yrs) 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 only reach nominal significance for Phecodes

representing “transient cerebral ischemia” (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. Fifth, our PheWAS analysis was limited to hospital episode stay and death record ICD10 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 (approx. 1,500 individuals). Finally, our results cannot be extrapolated to other ethnicities and populations, which differ in terms of genetic architecture, vascular risk factors, and environmental factors ([Brickman *et al.*, 2008](#); [Zahodne *et al.*, 2015](#); [Sargurupremraj *et al.*, 2020](#)).

FUNDING

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

COMPETING INTERESTS

The authors declare no competing interests.

TABLES

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
<i>HTRA1</i>	10:122461552-122514907	18	59	5.5E-6	0.044
<i>GBE1</i>	3:81489704-81761645	48	281	3.2E-5	0.131
<i>DCAKD</i>	17:45023337- 45061129	9	225	7.1E-5	0.192
<i>EGFL8</i>	6:32164594-32168285	15	237	1.5E-4	0.229
<i>RGS12</i>	4: 3314146-3439913	17	72	1.7E-4	0.229

Shown are genes reaching exome wide significance (FDR-adjusted $p < 0.05$; in bold) or suggestive evidence (FDR-adjusted $p < 0.25$) for association with logWMH volume. The results are derived from a unidirectional multivariable burden test (CMC) on logWMH and are adjusted for age at imaging, sex and the first 10 genomic principal components. Only variants that are 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 Minor allele count < 20 were filtered out.

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

Domain	Variants	MAC	beta	se	p-value
Signal peptide		129	-0.02	0.07	7.80E-01
	P4S	2	-0.21	0.61	7.34E-01
	A20V	127	-0.07	0.08	3.79E-01
IGFBP domain		11	0.18	0.23	4.30E-01
	A35S	1	0.66	0.87	4.45E-01
	G36V	3	-0.92	0.50	6.54E-02
	P50A	2	-0.52	0.61	4.01E-01
	E51K	1	-1.53	0.87	7.77E-02
	G73D	1	-0.59	0.87	4.98E-01
	P82R	3	-0.42	0.50	4.06E-01
Kazal domain		15	0.12	0.22	5.20E-01
	R133G	1	0.34	0.56	7.80E-01
	R138C	1	0.23	0.66	4.50E-01
	Q151K	12	0.02	0.25	9.34E-01
	Q157R	1	0.32	0.87	7.17E-01
Protease domain		35	0.79	0.14	9.50E-08
	V221L	1	1.87	0.87	3.10E-02
	V221M	1	0.08	0.87	9.26E-01
	R227W	20	0.73	0.19	1.71E-04
	K248T	1	0.85	0.87	3.26E-01
	E272V	3	0.03	0.50	9.59E-01
	P275L	1	1.61	0.87	6.40E-02
	M314V	1	0.48	0.87	5.83E-01
	D320N	6	1.17	0.35	9.77E-04
	K362N	1	1.82	0.87	3.60E-02
PDZ domain		26	0.06	0.17	4.30E-01
	A372V	1	0.53	0.61	3.89E-01
	R403W	3	1.01	0.50	4.51E-02
	V417I	1	1.06	0.87	2.24E-01
	D420E	4	0.26	0.43	5.49E-01
	S436G	1	-0.54	0.87	5.35E-01
	V442M	1	-0.96	0.87	2.71E-01
	A445T	4	0.30	0.43	4.95E-01
	D450N	2	0.01	0.61	9.83E-01
	V451I	6	-0.30	0.35	4.05E-01
	V461A	1	0.36	0.87	6.75E-01
	R463C	2	-0.02	0.61	9.72E-01

The results are derived from a multivariable burden test (CMC, 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 (grey) or the single variants.

Table 3. *HTRA1* protease domain PheWAS

Phecode	Phenotype	P-value	FDR-Adjusted P-value	Odds Ratio	95% Confidence Interval
389.2	Conductive hearing loss	3.60E-06	0.0025	10.75	2.99-26.90
571.6	Primary biliary cirrhosis	3.37E-05	0.012	13.37	2.77-38.77
411.41	Aneurysm and dissection of heart	6.34E-05	0.015	6.68	2.22-15.20
340.1	Migraine with aura	8.31E-05	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

Results from a 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 in a CMC burden test. Odds Ratios were calculated using Firth's regression and rare allele carrier status as an independent variable. FDR: false discovery rate.

FIGURE LEGENDS

Figure 1. Overview of rare *HTRA1* variants and the association with WMH burden.

(A) shown is the domain structure of the HTRA1 protein and the position of rare missense variants in the 16,511 UK Biobank participants included in the current analysis. Vertical lines reflect effect sizes (line length) and directionality of effect (increasing vs. 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) violin plot of age-stratified WMH burden in the UK Biobank. 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.

Figure 2. Effect size of protease domain variant carrier status on logWMH volume in comparison with established risk factors.

Shown are the effect sizes and standard errors of dichotomized risk factors (age, hypertension, diabetes, smoking, BMI), a dichotomized polygenic risk score based on common genetic variants, and HTRA1 protease domain variant carrier status. Effect sizes and standard errors were derived for logWMH volume change using a linear regression model adjusting for age, sex and 10 genetic PCs.

Figure 3. Functional consequences of rare *HTRA1* variants.

(A) Culture medium from transfected cells was collected to assess HTRA1 expression (anti-Myc immunoblot) and protease activity towards LTBP1 (anti-V5 immunoblots). ASM: active site mutant.

(B) Protease activity was determined as the ratio between cleaved (grey arrowhead) and intact (black arrowhead) LTBP1 and was normalized to HTRA1 levels. Data are presented as box-and-whisker plots (median \pm standard error of mean; the activity of

wt HTRA1 was set to 1). Data points represent independent measurements. At least three independent biological replicates were analyzed. 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 alpha smooth muscle actin (SMA) were detected as controls. Shown are representative images. Scale bar: 50µm.

Figure 5. Rare variants in the HTRA1 protease domain associate with multiple neurological and non-neurological phenotypes.

Phenome-wide association study (PheWAS) Manhattan plot for carriers of rare variants in the HTRA1 protease domain. Shown are all standardized Phencodes along with stroke-related phenotypes. 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.

REFERENCES

- Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* 2018; 166: 400-24.
- Armstrong NJ, Mather KA, Sargurupremraj M, Knol MJ, Malik R, Satizabal CL, *et al.* Common Genetic Variation Indicates Separate Causes for Periventricular and Deep White Matter Hyperintensities. *Stroke* 2020; 51(7): 2111-21.
- Barbu MC, Spiliopoulou A, Colombo M, McKeigue P, Clarke TK, Howard DM, *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.
- Beaufort N, Scharrer E, Kremmer E, Lux V, Ehrmann M, Huber R, *et al.* Cerebral small vessel disease-related protease HtrA1 processes latent TGF-beta binding protein 1 and facilitates TGF-beta signaling. *Proc Natl Acad Sci U S A* 2014; 111(46): 16496-501.
- Bipolar Disorder Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* 2018; 173(7): 1705-15 e16.
- Brickman AM, Schupf N, Manly JJ, Luchsinger JA, Andrews H, Tang MX, *et al.* Brain morphology in older African Americans, Caribbean Hispanics, and whites from northern Manhattan. *Arch Neurol* 2008; 65(8): 1053-61.
- Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; 562(7726): 203-9.
- Cho BPH, Nannoni S, Harshfield EL, Tozer D, Graf S, Bell S, *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.
- 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.
- 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.
- Dichgans M, Leys D. Vascular Cognitive Impairment. *Circ Res* 2017; 120(3): 573-91.
- Dichgans M, Mayer M, Uttner I, Bruning R, Muller-Hocker J, Rungger G, *et al.* The phenotypic spectrum of CADASIL: clinical findings in 102 cases. *Ann Neurol* 1998; 44(5): 731-9.
- Dichgans M, Pulit SL, Rosand J. Stroke genetics: discovery, biology, and clinical applications. *Lancet Neurol* 2019; 18(6): 587-99.
- Dufouil C, de Kersaint-Gilly A, Besancon V, Levy C, Auffray E, Brunnereau L, *et al.* Longitudinal study of blood pressure and white matter hyperintensities: the EVA MRI Cohort. *Neurology* 2001; 56(7): 921-6.
- Duperron MG, Tzourio C, Sargurupremraj M, Mazoyer B, Soumare A, Schilling S, *et al.* Burden of Dilated Perivascular Spaces, an Emerging Marker of Cerebral Small Vessel Disease, Is Highly Heritable. *Stroke* 2018; 49(2): 282-7.
- Eigenbrot C, Ultsch M, Lipari MT, Moran P, Lin SJ, Ganesan R, *et al.* Structural and functional analysis of HtrA1 and its subdomains. *Structure* 2012; 20(6): 1040-50.

- Emdin CA, Khera AV, Chaffin M, Klarin D, Natarajan P, Aragam K, *et al.* Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat Commun* 2018; 9(1): 1613.
- Fasano A, Formichi P, Taglia I, Bianchi S, Di Donato I, Battisti C, *et al.* HTRA1 expression profile and activity on TGF-beta signaling in HTRA1 mutation carriers. *J Cell Physiol* 2020; 235(10): 7120-7.
- Fernandez C, Halbert C, De Paula AM, Lacroze V, Froissart R, Figarella-Branger D, *et al.* Non-lethal neonatal neuromuscular variant of glycogenosis type IV with novel GBE1 mutations. *Muscle Nerve* 2010; 41(2): 269-71.
- Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* 2019; 570(7759): 71-6.
- 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(6): 928-39.
- Gaist D, Garde E, Blaabjerg M, Nielsen HH, Kroigard T, Ostergaard K, *et al.* Migraine with aura and risk of silent brain infarcts and white matter hyperintensities: an MRI study. *Brain* 2016; 139(Pt 7): 2015-23.
- 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): 628-34.
- 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-e308.
- Gormley P, Anttila V, Winsvold BS, Palta P, Esko T, Pers TH, *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet* 2016; 48(8): 856-66.
- Gould DB, Phalan FC, van Mil SE, Sundberg JP, Vahedi K, Massin P, *et al.* Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N Engl J Med* 2006; 354(14): 1489-96.
- Grau S, Baldi A, Bussani R, Tian X, Stefanescu R, Przybylski M, *et al.* Implications of the serine protease HtrA1 in amyloid precursor protein processing. *Proc Natl Acad Sci U S A* 2005; 102(17): 6021-6.
- Griffanti L, Zamboni G, Khan A, Li L, Bonifacio G, Sundaresan V, *et al.* BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. *Neuroimage* 2016; 141: 191-205.
- Hamedani AG, Rose KM, Peterlin BL, Mosley TH, Coker LH, Jack CR, *et al.* Migraine and white matter hyperintensities: the ARIC MRI study. *Neurology* 2013; 81(15): 1308-13.
- Hara K, Shiga A, Fukutake T, Nozaki H, Miyashita A, Yokoseki A, *et al.* Association of HTRA1 mutations and familial ischemic cerebral small-vessel disease. *N Engl J Med* 2009; 360(17): 1729-39.
- Hautakangas H, Winsvold BS, Ruotsalainen SE, Bjornsdottir G, Harder AVE, Kogelman LJA, *et al.* Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. *medRxiv* 2021: 2021.01.20.21249647.
- Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 2016; 99(4): 877-85.
- 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.
- Jian X, Satizabal CL, Smith AV, Wittfeld K, Bis JC, Smith JA, *et al.* Exome Chip Analysis Identifies Low-Frequency and Rare Variants in MRPL38 for White Matter Hyperintensities on Brain Magnetic Resonance Imaging. *Stroke* 2018; 49(8): 1812-9.

Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, *et al.* Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature* 1996; 383(6602): 707-10.

Jurgens SJ, Choi SH, Morrill VN, Chaffin M, Pirruccello JP, Halford JL, *et al.* Rare Genetic Variation Underlying Human Diseases and Traits: Results from 200,000 Individuals in the UK Biobank. *bioRxiv* 2020: 2020.11.29.402495.

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; 581(7809): 434-43.

Kruit MC, van Buchem MA, Hofman PA, Bakkers JT, Terwindt GM, Ferrari MD, *et al.* Migraine as a risk factor for subclinical brain lesions. *JAMA* 2004; 291(4): 427-34.

Kurth T, Mohamed S, Maillard P, Zhu YC, Chabriat H, Mazoyer B, *et al.* Headache, migraine, and structural brain lesions and function: population based Epidemiology of Vascular Ageing-MRI study. *BMJ* 2011; 342: c7357.

Ladis Study Group. 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): 577-88.

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.

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): 1658-70.

Lin MK, Yang J, Hsu CW, Gore A, Bassuk AG, Brown LM, *et al.* HTRA1, an age-related macular degeneration protease, processes extracellular matrix proteins EFEMP1 and TSP1. *Aging Cell* 2018; 17(4): e12710.

Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, *et al.* The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun* 2020; 11(1): 2624.

Lopez Chiriboga AS. Teaching NeurolImages: Prominent spinal cord atrophy and white matter changes in adult polyglucosan body disease. *Neurology* 2017; 88(20): e194-e5.

Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018; 50(4): 524-37.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; 26(22): 2867-73.

Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, *et al.* Computationally efficient whole genome regression for quantitative and binary traits. *bioRxiv* 2020: 2020.06.19.162354.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016; 17(1): 122.

Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 2016; 19(11): 1523-36.

Mishra A, Chauhan G, Violleau MH, Vojinovic D, Jian X, Bis JC, *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): 1009-23.

Mochel F, Schiffmann R, Steenweg ME, Akman HO, Wallace M, Sedel F, *et al.* Adult polyglucosan body disease: Natural History and Key Magnetic Resonance Imaging Findings. *Ann Neurol* 2012; 72(3): 433-41.

Monteith T, Gardener H, Rundek T, Dong C, Yoshita M, Elkind MS, *et al.* Migraine, white matter hyperintensities, and subclinical brain infarction in a diverse community: the northern Manhattan study. *Stroke* 2014; 45(6): 1830-2.

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.

Nichol D, Stuhlmann H. EGFL7: a unique angiogenic signaling factor in vascular development and disease. *Blood* 2012; 119(6): 1345-52.

Nozaki H, Kato T, Nihonmatsu M, Saito Y, Mizuta I, Noda T, *et al.* Distinct molecular mechanisms of HTRA1 mutants in manifesting heterozygotes with CARASIL. *Neurology* 2016; 86(21): 1964-74. Palm-Meinders IH, Koppen H, Terwindt GM, Launer LJ, Konishi J, Moonen JM, *et al.* Structural brain changes in migraine. *JAMA* 2012; 308(18): 1889-97.

Parker LH, Schmidt M, Jin SW, Gray AM, Beis D, Pham T, *et al.* The endothelial-cell-derived secreted factor Eglf7 regulates vascular tube formation. *Nature* 2004; 428(6984): 754-8.

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.

Poepsel S, Sprengel A, Sacca B, Kaschani F, Kaiser M, Gatsogiannis C, *et al.* Determinants of amyloid fibril degradation by the PDZ protease HTRA1. *Nat Chem Biol* 2015; 11(11): 862-9.

Rannikmae K, Henshall DE, Thrippleton S, Gijng Kong Q, Chong M, Grami N, *et al.* Beyond the Brain: Systematic Review of Extracerebral Phenotypes Associated With Monogenic Cerebral Small Vessel Disease. *Stroke* 2020; 51(10): 3007-17.

Ravenscroft G, Thompson EM, Todd EJ, Yau KS, Kresoje N, Sivadorai P, *et al.* Whole exome sequencing in foetal akinesia expands the genotype-phenotype spectrum of GBE1 glycogen storage disease mutations. *Neuromuscul Disord* 2013; 23(2): 165-9.

Rutten JW, Hack RJ, Duering M, Gravesteyn G, Dauwerse JG, Overzier M, *et al.* Broad phenotype of cysteine-altering NOTCH3 variants in UK Biobank: CADASIL to nonpenetrance. *Neurology* 2020; 95(13): e1835-e43.

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

Staples J, Qiao D, Cho MH, Silverman EK, University of Washington Center for Mendelian Genomics, Nickerson DA, *et al.* PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am J Hum Genet* 2014; 95(5): 553-64.

Szustakowski JD, Balasubramanian S, Sasson A, Khalid S, Bronson PG, Kvikstad E, *et al.* Advancing Human Genetics Research and Drug Discovery through Exome Sequencing of the UK Biobank. *medRxiv* 2020: 2020.11.02.20222232.

Traylor M, Persyn E, Tomppa L, Klasson S, Abedi V, Bakker MK, *et al.* Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. *The Lancet Neurology* 2021.

Uemura M, Nozaki H, Kato T, Koyama A, Sakai N, Ando S, *et al.* HTRA1-Related Cerebral Small Vessel Disease: A Review of the Literature. *Front Neurol* 2020; 11: 545.

UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019; 47(D1): D506-D15.

Verdura E, Herve D, Bergametti F, Jacquet C, Morvan T, Prieto-Morin C, *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): 741-53.

Verdura E, Herve D, Scharrer E, Amador Mdel M, Guyant-Marechal L, Philippi A, *et al.* Heterozygous HTRA1 mutations are associated with autosomal dominant cerebral small vessel disease. *Brain* 2015; 138(Pt 8): 2347-58.

Wang Q, Dhindsa RS, Carss K, Harper A, Nag A, Tachmazidou I, *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* 2020: 2020.12.13.422582.

Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging. *Lancet Neurol* 2013; 12(5): 483-97.

Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. *Lancet Neurol* 2019; 18(7): 684-96.

Weiss T, Taschner-Mandl S, Janker L, Bileck A, Rifatbegovic F, Kromp F, *et al.* Schwann cell plasticity regulates neuroblastic tumor cell differentiation via epidermal growth factor-like protein 8. *Nat Commun* 2021; 12(1): 1624.

Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, *et al.* Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform* 2019; 7(4): e14325.

Xie F, Zhang LS. A Chinese CARASIL Patient Caused by Novel Compound Heterozygous Mutations in HTRA1. *J Stroke Cerebrovasc Dis* 2018; 27(10): 2840-2.

Zahodne LB, Manly JJ, Narkhede A, Griffith EY, DeCarli C, Schupf NS, *et al.* Structural MRI Predictors of Late-Life Cognition Differ Across African Americans, Hispanics, and Whites. *Curr Alzheimer Res* 2015; 12(7): 632-9.

Zellner A, Scharrer E, Arzberger T, Oka C, Domenga-Denier V, Joutel A, *et al.* CADASIL brain vessels show a HTRA1 loss-of-function profile. *Acta Neuropathol* 2018; 136(1): 111-25.

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): 1423-6.

Figure 1

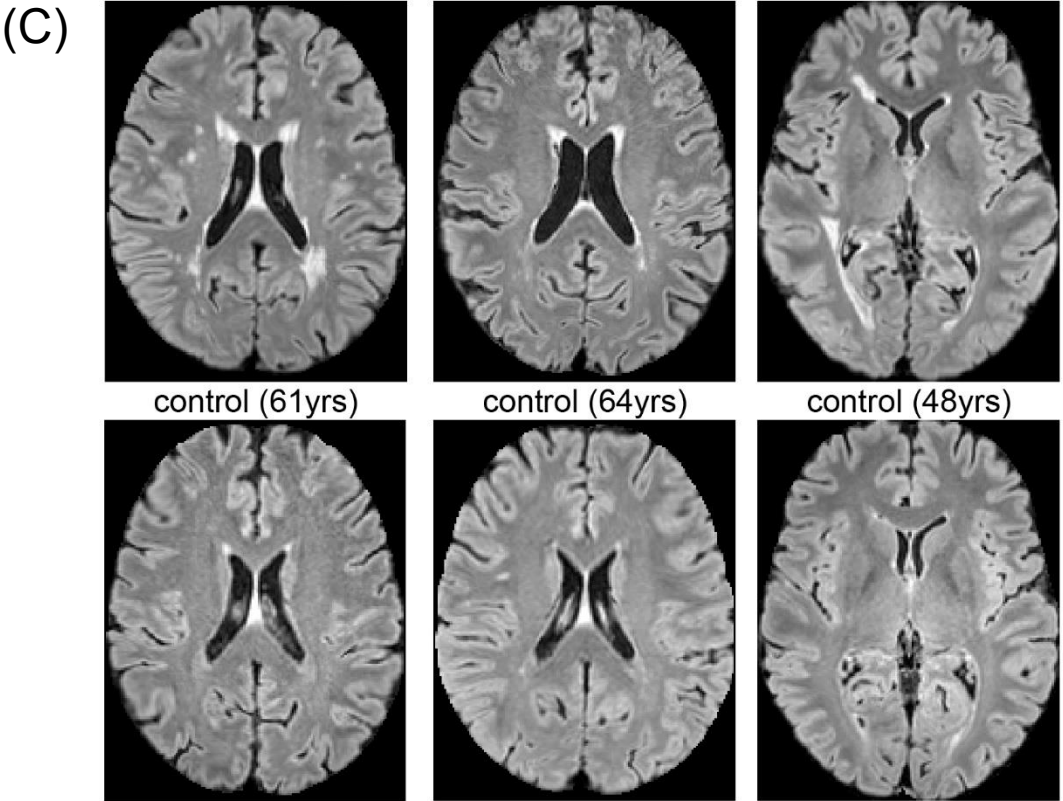
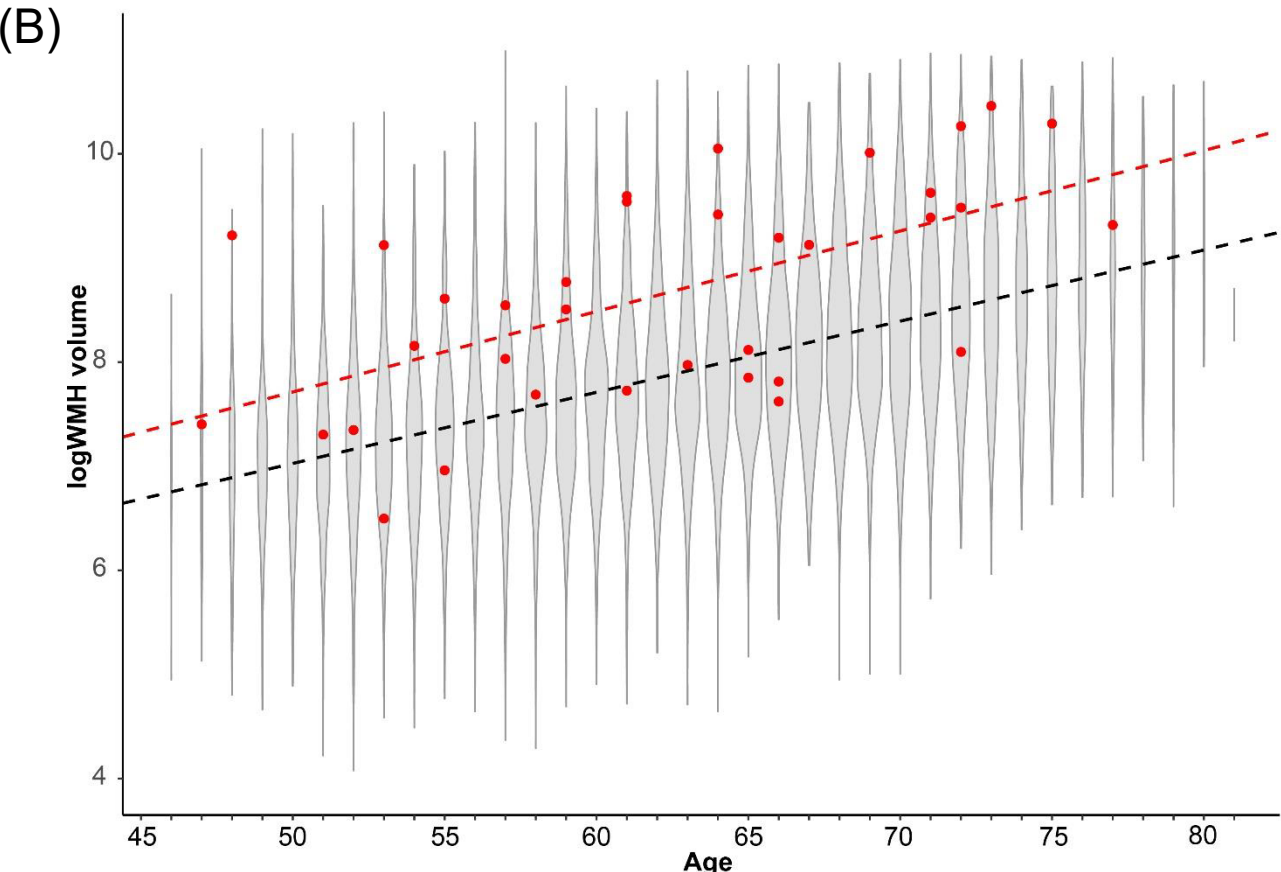
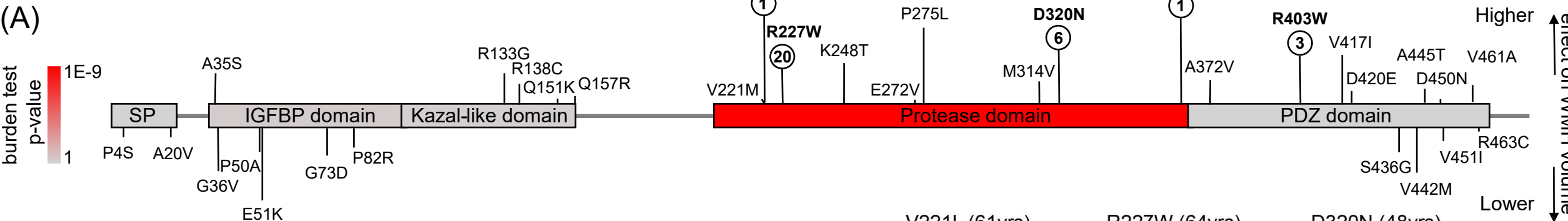


Figure 2

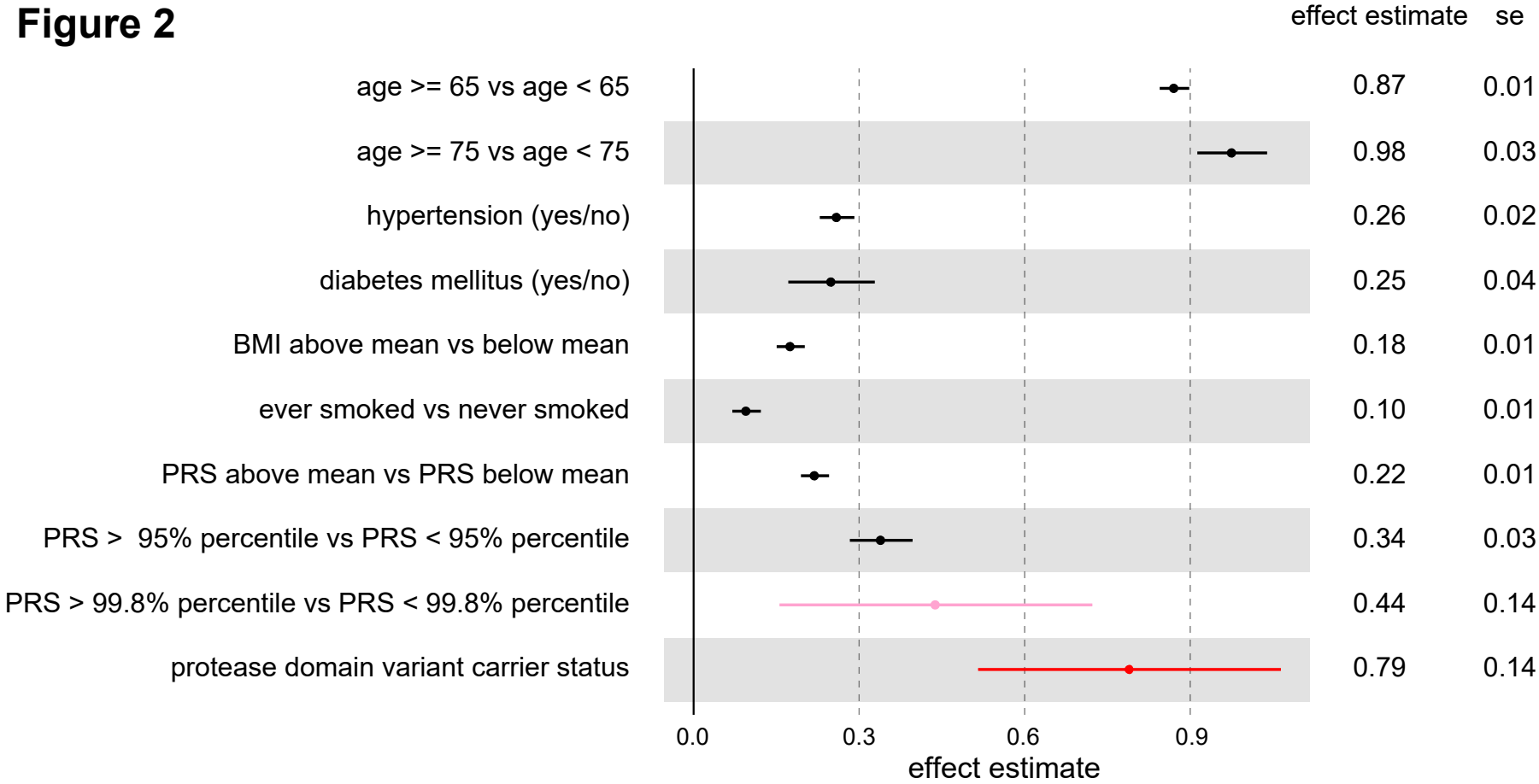


Figure 3

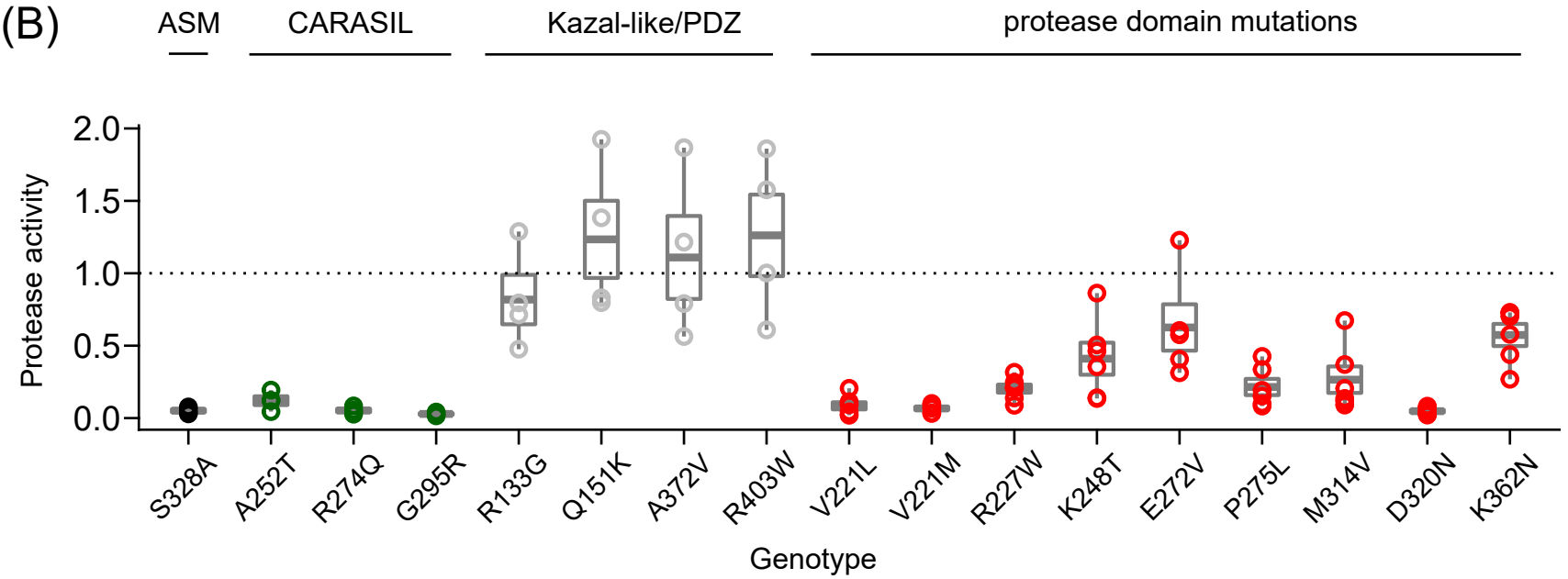
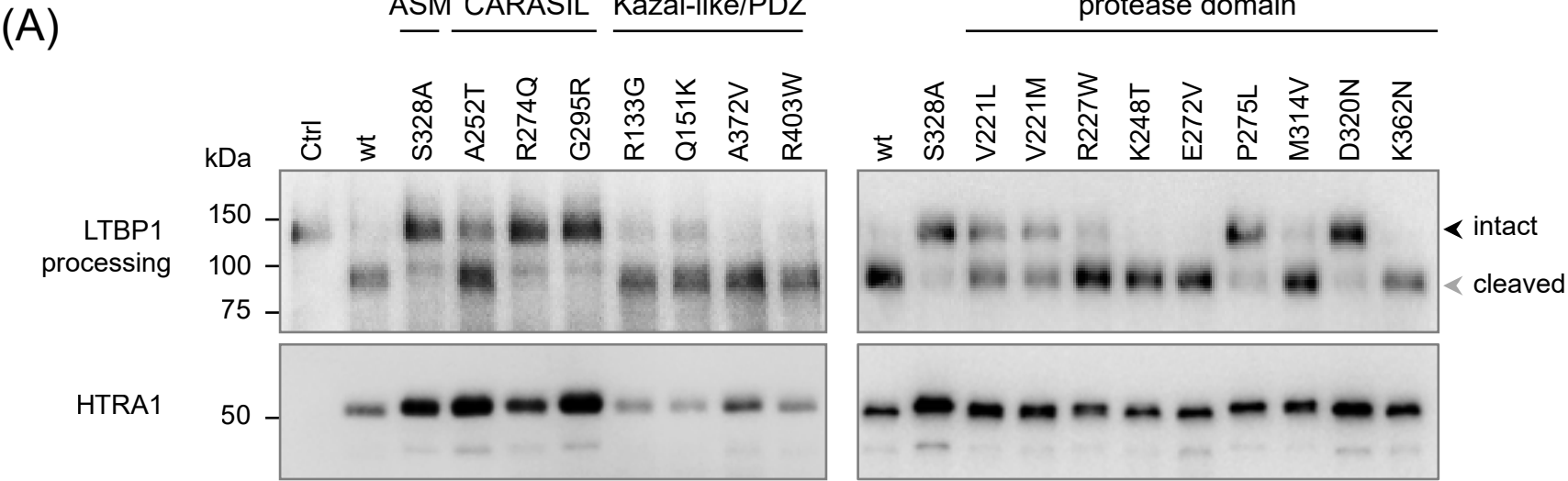
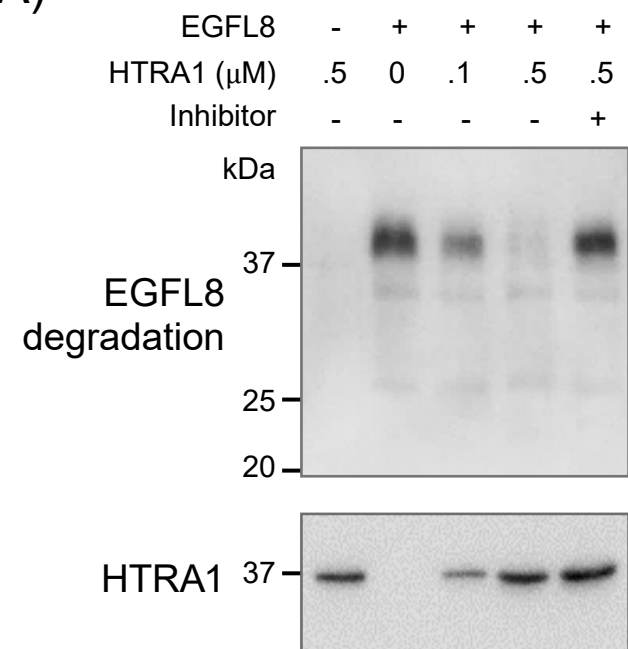


Figure 4

(A)



(B)

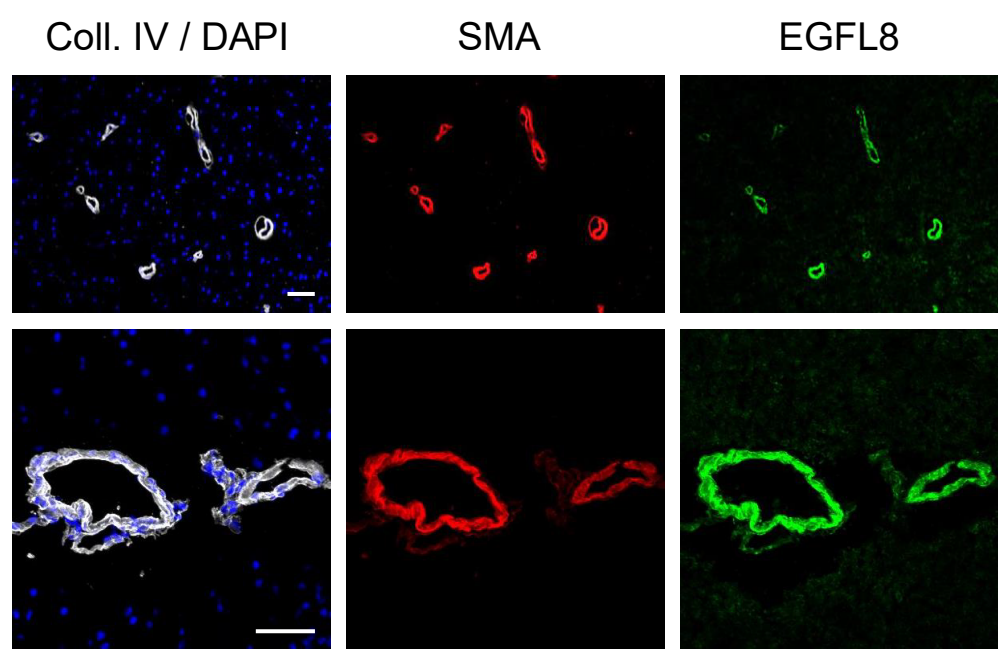


Figure 5

