White matter hyperintensities (WMH) are the most common brain-imaging feature of cerebral small vessel disease (SVD), hypertension being the main known risk factor. Here, we identify 27 genome-wide loci for WMH-volume in a cohort of 50,970 older individuals, accounting for modification/confounding by hypertension. Aggregated WMH risk variants were associated with altered white matter integrity ($p = 2.5 \times 10^{-7}$) in brain images from 1,738 young healthy adults, providing insight into the lifetime impact of SVD genetic risk. Mendelian randomization suggested causal association of increasing WMH-volume with stroke, Alzheimer-type dementia, and of increasing blood pressure (BP) with larger WMH-volume, notably also in persons without clinical hypertension. Transcriptome-wide colocalization analyses showed association of WMH-volume with expression of 39 genes, of which four encode known drug targets. Finally, we provide insight into BP-independent biological pathways underlying SVD and suggest potential for genetic stratification of high-risk individuals and for genetically-informed prioritization of drug targets for prevention trials.
A s a leading cause of stroke, cognitive decline, and dementia, cerebral small vessel disease (SVD) represents a major source of morbidity and mortality in aging populations1–3. Exploring the mechanisms of SVD and their contribution to dementia risk has recently been identified as a priority research area4,5, based on its more frequent recognition with magnetic resonance imaging (MRI), its high prevalence in older community persons6,7 and the demonstration that intensive management of vascular risk factors, especially hypertension, may slow down its progression7,8. The biological underpinnings of SVD are poorly understood and no mechanism-based treatments currently are available. White matter hyperintensities of presumed vascular origin (WMH), the most common MRI-marker of SVD, can be measured quantitatively using automated software. They are highly heritable8, and confer an increased risk of stroke and dementia9, thus making them well-suited to identify potential genetic determinants of SVD and its contribution to stroke and dementia risk. WMH are most often covert, i.e., not associated with a history of clinical stroke. They are highly prevalent in the general population, and much more frequently observed than clinical stroke caused by SVD (which can be both ischemic [small vessel stroke] and hemorrhagic [deep intracerebral hemorrhage]) (Supplementary Fig. 1).

Studying the genomics of SVD also provides a powerful approach to discovery of underlying molecular mechanisms and targets in order to accelerate the development of future therapies, or identify drug repositioning opportunities10–12. Although genomic studies of WMH have been most fruitful for deciphering SVD risk variants compared with other MRI-features of SVD (lacunes, cerebral microbleeds, dilated perivascular spaces)13 or small vessel stroke14, or deep intracerebral hemorrhage15, few risk loci have been identified to date16–18. This is likely due to limited sample size of populations studied and possibly also the failure to take into account the role of hypertension (HTN), the strongest known risk factor for WMH, in confounding or modifying genetic associations. There is also mounting evidence suggesting that early-life factors play a crucial role in the occurrence of late-life vascular and neurological conditions, including SVD19, likely due to both genetic and environmental factors that may intrinsically influence the vascular substrate of SVD or modulate the brain’s resilience to SVD20–22. Identifying these early predictors could have major implications for our understanding of disease mechanisms across the lifespan and for devising effective prevention strategies.

Here, we conduct a large multiancestry meta-analysis of WMH-volume genome-wide association studies (GWAS), accounting for HTN as a potential confounder and effect modifier. We explore association of WMH risk loci with early changes in white matter microstructure on MRI using diffusion tensor imaging (DTI) in young adults in their early twenties. Last, we explore biological pathways underlying the observed genetic associations with SVD and their clinical significance through shared genetic variation and Mendelian randomization (MR) experiments with vascular risk factors and neurological traits, linking them with multiple epigenomic, transcriptomic, and drug-target databases.

Results

Genetic discovery from association analyses. Figure 1 summarizes the overall workflow of our study that included data from 50,970 participants (N = 48,454 Europeans and 2516 African-Americans) from population-based studies taking part in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)23 consortium and from the UK Biobank (Supplementary Data 1). The mean age of participants was 66.0 ± 7.5 years, 53% were women and 52% hypertensive (Methods, Supplementary Methods 1, and Supplementary Data 1 for cohort-specific characteristics). There was no evidence for systematic inflation of SNP-WMH association statistics at the individual cohort or meta-analysis level (Supplementary Data 3 and Supplementary Fig. 2) for the three types of analyses performed (Methods).

In the European-only SNP-main-effects analysis, 22 independent loci harbored common variants associated with WMH volume at genome-wide significance (P < 5 × 10^-8, Table 1, Fig. 2), lead SNPs for each independent locus were confirmed by both LD clumping and GCTA-COJO44. Additionally, the NID2 locus reached genome-wide significance by the joint effect of multiple SNPs (P = 4.87 × 10^-8, Supplementary Data 4), with P = 5.45 × 10^-8 for lead SNP rs72680374, using GCTA-COJO (Methods, Supplementary Fig. 3). The African–American-only analysis identified a genome-wide significant locus at ECHDC3 (Supplementary Data 4). For loci showing heterogeneity in allelic effects across ancestry groups (PHet < 0.01), using MR-MEGA25 the ECHDC3 locus and another locus near ECHDC2 reached genome-wide significance (Table 1, Supplementary Data 5). In the HTN-adjusted model two additional loci were associated with WMH volume at P < 5 × 10^-8 (PKN2 and XKRT6), while three loci were no longer genome-wide significant (Table 1). The 2-df genome-wide gene-HTN interaction joint meta-analysis (JMA) did not identify any additional loci (Table 1, Supplementary Data 6). Five loci reached genome-wide significance in the small African–American-only JMA, but these were not maintained in the fixed-effects multiancestry JMA (Supplementary Data 7).

In total, 27 loci reached genome-wide significance in association with WMH volume in at least one of the aforementioned analyses, of which 18 have not previously been reported (Table 1, Fig. 2). Associations with WMH volume at these loci were similar in participants with and without HTN and when stratifying on quartiles of genetically predicted SBP and DBP levels (Methods, Supplementary Data 8-10). In aggregate, however, a weighted genetic risk score of independent WMH risk loci (WMH wGRS, Methods) showed significant 1-df interaction with HTN in association with WMH volume (P_GRSxHTN = 0.015, P_GRSxHTN = 0.009, Supplementary Fig. 4). One previously described risk locus for WMH burden did not reach genome-wide significance in the current analysis (near PMF1, P = 3.9 × 10^-4). Of note one genome-wide significant locus (COL4A2) and one suggestive locus (HTRA1, P < 5 × 10^-6, Supplementary Data 11), involve genes implicated in monogenic forms of SVD26,27.

Additional, gene-based tests using MAGMA28 yielded 49 genome-wide significant associations (P < 2.8 × 10^-5), of which 13 were outside GWAS loci, including the APOE gene (Methods, Supplementary Data 12). Using the Heritability Estimator from Summary Statistics (HESS)29 we found that 29 ± 2% of WMH-volume variance is explained by common and low frequency variants across the genome, the amount of heritability attributable to loci containing GWAS index SNPs being 2.4 ± 0.1%.

Implications of WMH genes across the lifespan. To examine the lifetime impact of WMH risk variants on brain structure, we explored the association of the WMH wGRS with MRI-markers of white matter microstructure in 1738 young healthy adults participating in the i-Share cohort (mean age 22.1 ± 2.3 years, 72% women). Integrity of the white matter microstructure was measured on diffusion tensor imaging (DTI) using the following metrics: fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), axial diffusivity (AD), and the recently described peak width of skeletonized mean diffusivity (PSMD)30. These MRI-markers are associated with the maturation and aging
process of white matter microstructure\textsuperscript{31,32}, and alterations in DTI metrics were shown to precede the occurrence of WMH lesions in older patients with SVD\textsuperscript{30,33}. The WMH wGRS showed significant associations with higher MD, RD, and PSMD and lower FA values in i-Share; four WMH risk loci individually showed significant associations with at least one of the DTI parameters (\textit{SH3PD2A, NMT1, KLHL24}, and \textit{VCAN}, Table 2). Increasing values of PSMD (but not other DTI markers) shows a trend towards association with slower information processing speed on the Stroop test in i-Share participants (\textit{N} = 1,401, effect estimate \( \pm SE: 0.085 \pm 0.040, \ P = 0.031 \)), which did not survive \( P \geq 0.01 \), Methods) was observed between WMH volume and increased risk of PPA\textsuperscript{3} (\textit{P} \leq 3.6 \times 10^{-3}, Methods, Fig. 3, Supplementary Data 14). Using the Bayesian pairwise GWAS (GWAS-PW) approach\textsuperscript{36}, significant regional overlap (posterior probability of association for model 3, \textit{PPA3} \geq 0.90, Methods) was observed between WMH volume and general cognitive function and between WMH volume and stroke, especially ischemic and small vessel stroke (Supplementary Data 15). This included regions previously implicated in complex and monogenic forms of stroke (\textit{FOXF2/FOXQ1}, \textit{HTRA1}\textsuperscript{27,37}, and cardiovascular disease (NO\textit{S}))\textsuperscript{39}.

Using two-sample MR, which implements the inverse-variance weighting (IVW) method, we observed evidence for significant causal associations after Bonferroni correction for multiple testing (\( P \leq 3.6 \times 10^{-3} \)) between WMH volume and increased risk of Alzheimer-type dementia, with no statistical evidence of horizontal pleiotropy using Cochran’s Q statistic (Q-PHET\textsuperscript{20,01}, Fig. 4, Supplementary Data 16). We also observed evidence for significant causal association of WMH volume with risk of any stroke, ischemic stroke, small vessel stroke, and deep intracerebral hemorrhage. There was some evidence for horizontal pleiotropy
Table 1. Loci reaching genome-wide significance with WMH burden in one or more genetic association model.

| Region | Position | Nearest Gene | Analysis | Local r² | Function | SNP | EAF | OA | EA | P | SE | P | SE | P |
|--------|----------|-------------|----------|----------|----------|-----|-----|----|----|---|----|---|----|---|
| Novel  | 43144218  | NM1F        | EUR/TRANS| 0.23%    | Intronic | c/t | 0.63 | 0.052 | 0.006 | 3.43E-19 | 0.054 | 0.006 | 4.85E-20 | 9.17E-20 |
|        | 16644702  | C6orf52     | EUR/TRANS| 0.12%    | Intronic | a/t | 0.58 | 0.050 | 0.006 | 3.96E-17 | 0.049 | 0.006 | 4.91E-17 | 1.44E-16 |
|        | 13720902  | L1H4C        | EUR/TRANS| 0.12%    | Intronic | a/c | 0.58 | 0.050 | 0.007 | 4.46E-17 | 0.046 | 0.007 | 3.96E-17 | 1.44E-16 |
|        | 13411207  | CCDC28C      | EUR/TRANS| 0.03%    | Intronic | c/a | 0.71 | 0.046 | 0.007 | 2.59E-08 | 0.049 | 0.007 | 4.91E-11 | 1.44E-11 |
|        | 16567132  | AL4G05050      | EUR/TRANS| 0.02%    | Intronic | c/a | 0.71 | 0.046 | 0.007 | 2.59E-08 | 0.049 | 0.007 | 4.91E-11 | 1.44E-11 |
|        | 11458987  | GLI1        | EUR/TRANS| 0.02%    | Intergenic | c/a | 0.55 | 0.034 | 0.006 | 3.76E-10 | 0.034 | 0.006 | 3.76E-10 | 1.44E-10 |
|        | 90624874  | SIGMAR1      | EUR/TRANS| 0.02%    | Intergenic | c/a | 0.55 | 0.034 | 0.006 | 3.76E-10 | 0.034 | 0.006 | 3.76E-10 | 1.44E-10 |
|        | 12088048  | SH3PXD2A-AS1 | EUR/TRANS| 0.02%    | Intergenic | c/a | 0.55 | 0.034 | 0.006 | 3.76E-10 | 0.034 | 0.006 | 3.76E-10 | 1.44E-10 |

For each locus, the value in the P-value column is the lowest P-value in the fixed-effects transancestral meta-analysis or the fixed-effects Europeans-only meta-analysis, respectively, is shown. GWA results have been filtered for a single nucleotide polymorphism (SNP) with a minor allele frequency (MAF) ≥ 0.01. Loci reaching genome-wide significance (GWS, P-value < 5.00E-08) are highlighted in bold. Additional loci reaching GWS significance in the SC-TACO meta-analysis for the main-effects model.
(Q-PHet<0.01) for associations with stroke. However, after removing influential outlier variants associations remained significant and the MR-Egger intercept (indicating average pleiotropic effects) did not significantly differ from zero with $Q_a$ values close to 1, indicating goodness of fit of the IVW method (Methods, Supplementary Data 16).

**Shared genetic risk with vascular traits.** To assess whether genetic associations with WMH reflect known vascular mechanisms we systematically explored the shared genetic variation between WMH burden and related vascular traits, comprising established risk factors for vascular disease (systolic blood pressure [SBP], diastolic blood pressure [DBP], pulse pressure [PP], type 2 diabetes [T2D], low-density lipoprotein [LDL] cholesterol, high-density lipoprotein [HDL] cholesterol, triglycerides [TG], body mass index [BMI], glycated hemoglobin levels [HbA1c], and lifetime smoking index—a composite measure capturing smoking heaviness and duration as well as smoking initiation [SMKindex]), as well as putative risk factors for other disorders including venous thromboembolism (VTE) and migraine, using summary statistics from the most recent GWAS. Some of the latter were made available through collaborations with the relevant research consortia when the data were not publicly available (Supplementary Table 2).

First, we looked up associations of the 27 WMH risk loci with related vascular traits, including the lead WMH risk variants and nearby variants (±250 kb) in moderate to high LD ($r^2 > 0.5$). After correcting for the number of independent loci and traits tested ($P < 1.3 \times 10^{-4}$, Methods), 20 of the 27 WMH risk loci (74%) showed significant association with at least one other trait and/or vascular risk factors. For 13 of these, associations were found at a genome-wide significant level (Fig. 2, Supplementary Data 13). Blood pressure (BP) traits showed by far the largest number of
significant associations with WMH risk variants, 16 loci (59.3%) being associated with SBP, DBP, and/or PP. Further significant associations with WMH risk variants were observed for BMI (8 loci), T2D (5 loci), SMKindex (3 loci), and lipid traits (3 loci), one locus (at XKR6) being notably shared with all these risk factors. Seven loci (C16orf95, ECHDC3, MIN1, NID2, SALL1, VCAN, KCNK2), none of which were reported previously as WMH risk loci, appear not to be associated with any of the vascular traits explored, suggesting other underlying biological pathways.

Second, we explored the genome-wide and regional genetic overlap between WMH volume and related vascular traits. Mean \(X^2\) ranged between 1.06 and 3.99 suggesting strong polygenicity for all investigated traits. The impact of possible sample overlap was estimated to be negligible using LDSR\(^3\) (Supplementary Data 14). We observed significant \((P < 3.6 \times 10^{-3})\) genetic correlation of larger WMH volume with higher SBP, DBP, SMKindex, BMI and increased risk of VTE. Using GWAS-PW\(^3\) and HESS\(^4\) (Methods), we identified 16 genomic regions harboring shared genetic risk variants with at least one other vascular trait, predominantly BP traits, but also BMI, lipid levels and SMKindex (PPA3 ≥ 0.90, Fig. 3, Supplementary Data 15).

Third, we explored the causal relations between the aforementioned vascular traits and WMH volume using two-sample MR\(^4\) (RadialMR\(^5\), Methods), implementing the IVW method. We observed significant \((P < 3.6 \times 10^{-3})\) association of genetically predicted SBP, DBP, PP, SMKindex and T2D with larger WMH volume and of genetically predicted migraine with smaller WMH volume (Fig. 4, Supplementary Data 17). After removal of potentially pleiotropic outlier variants; for SBP, DBP, PP and SMKindex the MR-Egger intercept was nonsignificant, indicating no residual pleiotropy and suggesting causal association with WMH volume (Methods). For migraine and T2D in contrast there was evidence of residual pleiotropic effects (significant MR-Egger intercept, Supplementary Data 17) after removal of potentially pleiotropic outlier variants, and the association became only nominally significant for migraine. Importantly, associations of genetically predicted SBP and DBP with WMH volume remained significant after adjustment for HTN, and in participants with and without HTN (Supplementary Data 17), highlighting that higher levels of BP are likely causally associated with larger WMH volume even below BP thresholds typically used for the definition of hypertension (SBP ≥ 140 mmHg or DBP ≥ 90 mmHg or antihypertensive drug intake)\(^4\).

### Biological interpretation of association signals

We used EPIGWAS\(^4\) to test for cell-type enrichment of WMH association signals using chromatin marks previously shown to be cell-type specific and associated with active gene-regulation (Methods). WMH risk loci were significantly enriched in enhancer (H3K4me1) and promoter sites (H3K4me3) in cell-types derived from the brain, neurosphere (developing brain), vascular tissue, digestive, epithelial and muscle tissues, as well as human embryonic stem cells after removing WMH risk loci associated with BP (Supplementary Data 19). Analysis of brain-specific single-cell expression data in mice using MAGMA.celldtyping\(^4\) (Methods) revealed significant enrichment of highly cell-type-specific genes in endothelial mural cells and nominally significant enrichment for vascular leptomeningeal cells, oligodendrocytes, oligodendrocyte precursors, and ependymal astrocytes; results were substantially unchanged after removing WMH risk loci associated with BP (Supplementary Data 20).

To functionally characterize and prioritize individual WMH genomic risk loci we performed transcriptome-wide association studies (TWAS) using TWAS-Fusion\(^6\), WMH-SNP association statistics from the main effects (EUR-only) and weights from 23

### Table 2 Association trend of WMH risk loci with white matter microstructures in young adults.

| Region | Nearest gene | WMH association trend | FA | MD | RD | AD |
|--------|--------------|-----------------------|----|----|----|----|
|        |              |                       |    |    |    |    |
| rs10976722 | SK3P2O2A | 0.022 | 0.015 | 0.012 | 0.015 | 0.012 | 0.015 |
| rs6530417 | VCAN | 0.019 | 0.015 | 0.012 | 0.015 | 0.012 | 0.015 |
| rs8307079 | VCAN | 0.022 | 0.015 | 0.012 | 0.015 | 0.012 | 0.015 |
| rs17205972 | VCAN | 0.022 | 0.015 | 0.012 | 0.015 | 0.012 | 0.015 |
gene-expression reference panels from blood, arterial, and brain tissues (Supplementary Methods 2). We also included non-publicly available gene-expression weights from the dorsolateral prefrontal cortex (DLPFC) of 494 older community-dwelling participants (Methods) 47,48. TWAS-Fusion identified 201 transcriptome-wide significant associations with WMH, conditionally significant on the predicted expression of a TWAS-associated gene, including 21 with splicing quantitative trait loci (sQTLs) regulating highly tissue-specific gene isoforms in DLPFC (Fig. 5, Supplementary Data 21). To rule out that observed associations reflect the random overlap between expression (eQTLs) and noncausal WMH risk variants, a colocalization analysis (COLOC)49 was performed at each significant locus, to estimate the posterior probability of a shared causal variant (PP4 ≥ 75%) between the gene expression and trait association (Methods). Colocalization was observed for 96 TWAS significant eQTLs (48%, Fig. 5): of these, 54 mapped to 8 WMH genome-wide risk loci and 16 expression-associated genes (eGenes), while 42 mapped to 12 distinct loci that were not genome-wide significant in the WMH GWAS and 23 eGenes. These additional putative WMH risk loci require confirmation in follow-up studies. Leveraging histone regulatory mark information from blood, arterial, and brain tissues (Methods, Supplementary Data 21), we observed that the majority (89%) of TWAS signals overlapped with enhancer and/or promoter elements, including eQTLs exhibiting weaker colocalization probability (PP4 < 75%). Larger WMH volume was associated with either upregulated or downregulated gene expression, the directionality being mostly consistent across broad tissue categories (Fig. 5). We found evidence for colocalization of WMH risk variants with eQTLs in brain tissues (28 eGenes), vascular tissues (15 eGenes), and blood (6 eGenes). Some eGenes (KLHL24, NMT1, DCAKD, KANSL1, AMZ2P1) showed evidence for colocalization in multiple tissues, and for some WMH risk loci (chr2q33.2, chr17q25.1, chr10q26.31). Related traits

Fig. 3 Shared genetic architecture of WMH at genome-wide and regional level. Color coded for the direction of effect (Green: Positive genetic correlation; Red: Negative genetic correlation). The LD-score regression (LDSR) axis shows evidence for genome-wide correlations (after Bonferroni correction for multiple testing *P* < 3.6 × 10⁻³, Methods), with the size of the nodes corresponding to the level of significance of the association. The GWAS-pairwise (PW) axis shows evidence for regional level overlap of association signals between WMH burden and related vascular and neurological traits (PPA3 ≥ 0.90, Methods). For any given region, the nearest gene (in brackets) to the top SNP associated with WMH is shown. Bivariate heritability estimator from summary statistics (p-HESS) was used to infer directionality of shared association signals (Methods) and asterisks denote an unexpected directionality of association. SBP systolic blood pressure, DBP diastolic blood pressure, PP pulse pressure, AS all stroke, IS ischemic stroke, SVS small vessel stroke, CE cardioembolic stroke, BMI body mass index, HDL high-density lipoprotein, LDL low-density lipoprotein, VTE venous thromboembolism, GCF general cognitive function, SMKindex lifetime smoking index, WMH white matter hyperintensity.
chrl7q21.31) colocalized variants associated with multiple eGenes with distinct tissue specificities (Fig. 5, Supplementary Data 21). WMH risk variants at the chr2q33 locus for instance showed evidence for colocalization with eQTLs for NBEAL1 in nerve and arterial tissues, for ICAIL and KRTBP15 in brain tissues, and for CARFI in right atrial appendage.

Among the 39 WMH eGenes with high colocalization probability (Fig. 5, Supplementary Data 21), 4 (MAPT, CRHR1, CALCRL, NOTCH4) are registered as targets of approved drugs in the DrugBank database and the Therapeutic Target Database (Supplementary Data 22).

Discussion
This largest genetic study to date on complex SVD13,14,16–18 leveraging genetic and brain-imaging information from 50,970 older community persons, triples the number of genetic loci associated with cerebral SVD and shows that this genetic risk results in detectable brain changes among asymptomatic young adults in their twenties. We further demonstrate the importance of higher BP as a risk factor for WMH even below clinical thresholds for HTN. MR analysis provides strong evidence for causal links of genetically determined WMH volume with risk of ischemic stroke, intracerebral haemorrhage, and Alzheimer-type dementia in later life. Importantly, we also provide insight into molecular pathways underlying SVD, highlighting relevant tissue and cell types, and suggest potential for genetic stratification of high-risk individuals and for genetically informed prioritization of drug targets for prevention trials.

Our approach focusing on the most common brain-imaging feature of SVD appears to be more powerful than GWAS of the small vessel stroke subtype to identify risk loci for SVD. Indeed, no new small vessel stroke risk locus was identified in MEGASTROKE, the largest stroke GWAS meta-analysis to date14. We show a strong association between genetically determined WMH burden and risk of stroke in the general population, notably both risk of ischemic stroke and of intracerebral hemorrhage. While corroborating epidemiological observations3,50, this has never been demonstrated using genetic instrumental variables, providing evidence for causality. This prompts greater caution with the common empirical prescription of antiplatelet therapy in persons with extensive WMH in the absence of clinical stroke3, given the potential detrimental effects on intracerebral hemorrhage risk, and suggests the need for randomized clinical trials to determine the risk/benefit ratio of antiplatelet therapy in this setting.

The significant association we describe between genetically determined WMH burden and Alzheimer-type dementia also has potential important implications for prevention. It strengthens recent epidemiological evidence that WMH is associated not only with an increased risk of all and vascular dementia, but also of neurodegenerative Alzheimer’s disease, T2D type II diabetes, SMKindex lifetime smoking index, WMH white matter hyperintensity, Model 1 Main effects adjusted for age, sex, principal components for population stratification, intracranial volume; Model 2 Model 1 + hypertension status.

Fig. 4 Mendelian randomization results of vascular risk factors with WMH burden (box A) and WMH burden with neurological traits (box B). Point estimates and confidence intervals (blue) from the inverse-variance weighted (IVW) method are shown along with the point estimates and 95% confidence interval (black) from sensitivity analyses after filtering out potentially pleiotropic outlier variants. The intercept and p-value from the MR-Egger method is displayed on the far right (an intercept term significantly differing from zero at the conservative threshold of \( P < 0.05 \) suggests the presence of directional pleiotropy). SBP systolic blood pressure, DBP diastolic blood pressure, PP pulse pressure, HTN hypertensive, NT normotensive, Str. stratum, AS all stroke, IS ischemic stroke, SVS small vessel stroke, ICH intracerebral hemorrhage, AD Alzheimer’s disease, T2D type II diabetes, SMKindex lifetime smoking index, WMH white matter hyperintensity, Model 1 Main effects adjusted for age, sex, principal components for population stratification, intracranial volume; Model 2 Model 1 + hypertension status.
absence of other efficient therapies. Our results suggest that WMH should be considered a major target for preventative interventions, to mitigate not only the risk of stroke and vascular cognitive impairment but also of Alzheimer-type dementia, and support the rationale of innovative trials using WMH progression as a surrogate or intermediate endpoint for cognitive decline and dementia.

Over half of identified WMH risk loci are associated with higher BP levels. Moreover, using MR we provide evidence for a causal association between higher BP and larger WMH volume, notably even in participants without clinically defined HTN at the time of the MRI. Indeed, associations of genetically predicted SBP and DBP with WMH volume remained significant in participants without HTN, highlighting that higher levels of BP are likely causally associated with larger WMH volume even below BP thresholds typically used for the definition of HTN. Considering the recent conclusions from the SPRINT-MIND trial suggesting that more drastic lowering of BP in persons with HTN is warranted. We additionally show strong causal association between increased exposure to cigarette smoking over the lifetime (lifetime smoking index) and increased WMH burden, as has recently been described in relation with stroke risk, providing some additional evidence for the relevance of smoking cessation to prevent vascular brain injury and specifically SVD.

Importantly, a quarter of the identified WMH risk loci reflect molecular mechanisms that are not mediated by BP or other known vascular risk factors, two of these (NID2, VCAN), along with the COL4A2 and EFEMP1 locus, implicate genes encoding matrisome proteins, involved in cell membrane structure and representing core components of the extracellular matrix (ECM). Converging evidence from experimental models for monogenic SVD suggest that perturbations of the matrisome play a central role in disease pathophysiology. Our findings suggest that these could also be relevant for sporadic forms of SVD. NID2 encodes nidogen, an ECM glycoprotein and a major component of basement membranes and is recognized as having a role in post-stroke angiogenesis. Mutations in COL4A1/2, encoding collagen another basement membrane component already are known causing monogenic SVD, VCAN, which we also found to be associated with white matter microstructure in young adults, encodes versican, a proteoglycan involved in cell adhesion and ECM assembly. In CARASIL, a monogenic SVD caused by HTRA1 mutations, accumulation of versican in the thickened arterial wall was observed. Versican also can form complexes that inhibit oligodendrocyte maturation and remyelination. EFEMP1 encodes fibulin 3, an ECM glycoprotein localised in the basement membrane, and a proteolytic target of serine protease HTRA1. Other previously unreported WMH risk loci that we have identified include KCN2 that encodes Twik-related K+ channel (TREK1), a 2-pore-domain background ATP-sensitive potassium channel expressed throughout the central nervous system, more prominently in fetal than in adult brain. ECHDC3, near a distinct locus previously implicated in Alzheimer disease, which previously has been causally related to familial meningiomas, and XKR6, which has been associated with risk of systemic lupus erythematosus.

Our results provide important insight into the lifetime impact of genetic risk for SVD. Indeed, WMH risk variants observed in older adults were already associated with changes in DTI markers of white matter integrity in young adults in their early twenties. Of these, PSMD, a DTI metric recently described to be more
strongly correlated with cognitive performance in older persons (patients with sporadic or monogenic SVD and older community persons) than any other MRI-marker of SVD, was already showing nominal association with lower cognitive performance in young adults. This finding requires confirmation in future independent samples. The association of the WMH wGRS with subtle changes in white matter microstructure in young adults, if confirmed in independent samples, has potential important implications for the timing and paradigm of prediction and prevention of SVD progression and complications. It could reflect that biological pathways contributing to WMH at an older age already have a significant impact on brain microstructure in young adults, possibly reflecting a very early stage of SVD (typically characterized by reduced FA and increased MD and PSD)\(^\text{33}\). DTI changes and WMH have been suggested to be dependent physiological processes occurring within consecutive temporal windows in older patients with SVD\(^\text{30,33,66}\). Alternatively, observed associations might also reflect pleiotropy between SVD genes and genes influencing brain maturation, as the mean age of i-Share participants corresponds to the peak of white matter maturation\(^\text{36}\). On average FA tends to increase during childhood, adolescence, and early adulthood and then decline in middle-age, while the reverse is observed for MD\(^\text{32,68}\). Hence the association of the WMH wGRS with lower FA and higher MD could also reflect an impaired or delayed maturation or a premature aging process. The significant association of the WMH wGRS with RD but not \(\text{A}_2\text{D}\) could potentially suggest that this is predominantly reflecting an impact on myelination of fiber tracts\(^\text{69}\) in line with involvement of oligodendroglial dysfunction in early SVD pathology\(^\text{70}\). Future follow-up studies in a longitudinal setting are warranted to better understand the impact of genetically predicted WMH burden on the progression of white matter microstructural changes observed already in young adults and on their link with SVD and its complications.

Functional characterization revealed enrichment of WMH risk variants in regulatory marks in brain and neurosphere and in single-cell gene-expression levels in endothelial mural cells (as for clinical stroke)\(^\text{71}\). Gene prioritization using TWAS revealed that several WMH risk loci colocalized with eQTL for multiple genes with distinct tissue specificities. This pattern could potentially partly explain why association of such loci with WMH volume remained unchanged after controlling for the presence of HTN, although they were associated at genome-wide significant level with both BP and WMH. Of the 39 eGenes identified by TWAS four encode known drug targets. \(\text{MAPT}\) is a drug target under investigation for neurodegenerative disorders: the eQTL colocalizing with the WMH risk variant is an sQTL for the MAPT isoform in DLPFC and TWAS suggest that larger WMH volume is associated with upregulated \(\text{MAPT}\) expression. \(\text{CALCRL}\) encodes a component of the Calcitonin Gene Related Peptide receptor. TWAS suggest that lower abundance of the \(\text{CALCRL}\) transcript in arterial and nerve tissue and higher abundance in blood are associated with larger WMH volume. Monoclonal antibodies against \(\text{CALCRL}\) have recently been developed for the treatment of migraine\(^\text{72}\).

We acknowledge limitations. We were underpowered for detecting additional risk variants for WMH after accounting for presence of HTN in the 2-df JMA gene-HTN interaction model. Recognizing that blood pressure is also highly variable and that a one-time blood pressure measurement may not reflect the long-term exposure of participants to high blood pressure levels, we conducted secondary analyses stratifying on quartiles of genetically predicted SBP and DBP levels, yielding similar results. In aggregate, a weighted genetic risk score of independent genome-wide significant WMH risk loci showed a significant 1-df interaction with HTN status in association with WMH volume, suggesting that effect modification of genetic associations by HTN may exist, but that to discover them at the level of individual loci likely will require much larger samples. While we were able to use gene-expression data from many tissues for TWAS, such data are lacking for certain tissues that may be relevant for WMH (e.g., small brain vessels, microglia). Finally, our study population is predominantly of European ancestry (95%) limiting our ability to extrapolate our conclusions to other ancestry groups.

In conclusion we have identified 27 genetic risk loci for WMH volume, of which two thirds are not previously reported, and provided additional insight into their association with structural brain changes in very young adults, their clinical significance and the importance of high BP as a risk factor below clinical thresholds. Our results also point to molecular pathways underlying SVD that are not mediated by vascular risk factors and suggest potential for genetic stratification of high-risk individuals and genetically informed prioritization of drug targets for prevention trials.

**Methods**

**Study population.** The study population comprised 23 population-based studies from the CHARGE consortium comprising a total of 24,182 individuals of European (\(N = 21,666\)) and African–American (\(N = 2516\)) ancestry, along with 26,788 community participants of European origin from the UK Biobank. In total, 50,970 participants were available for testing main genetic effects and 48,524 participants for the gene-hypertension interaction analysis (information on HTN status was missing in 2446 participants). Individuals with a history of stroke (or \(\text{MRI}\)-defined brain infarcts involving the cortical gray matter), or other pathologies that may influence the measurement of WMH (e.g., brain tumor, head trauma, etc.), at the time of \(\text{MRI}\) were excluded from analyses. Study participants in all participating cohorts gave written informed consent for phenotype quantification and use of genetic material (Supplementary Methods 1, Supplementary Table 1).

**Phenotypes.** MRI scans were obtained from scanners with field strengths ranging mostly from 1.5 to 3.0 Tesla and interpreted using a standardized protocol blinded to clinical or demographic features. In addition to T1 and T2 weighted scans along the axial plane, some cohorts included fluid-attenuated inversion recovery (FLAIR) and/or proton density (PD) sequences for better differentiation of WMH from cerebrospinal fluid. The vast majority of participating cohorts (>92% of all participants) used fully automated software to quantify WMH volume, with two cohorts using validated, visually guided semi-quantitative scales in older study subsets (Supplementary Table 1). WMH volume measures were inverse normal transformed to correct for skewness and account for differences in WMH quantification methods. Blood pressure measurements that are closest to the MRI scan were used to define HTN status. Participants with a SBP \(\geq 140\) mmHg, a DBP \(\geq 90\) mmHg, or taking antihypertensive medication were classified as having HTN.

**Genotyping and imputation.** Genome-wide genotyping platforms are described in Supplementary Data 2. Prior to imputation, sample-specific quality control (QC) on heterozygosity, missingness, gender mismatch, cryptic relatedness, and analysis of principal components (PC) for population stratification, as well as SNP-level QC on genotyping call rate and Hardy–Weinberg equilibrium were applied (Supplementary Data 2). Samples and SNPs passing the cohort-specific QC measures were then imputed to the 1000 genomes cosmopolitan panel phase 1 version 3 (1000 G p13) for CHARGE cohorts, while for the UK Biobank the dataset version 3 was imputed to the combined UK10k and HaploType Reference Consortium (HRC) reference panels.

**Genome-wide association analyses.** Each participating study conducted ancestry-specific analyses using linear regression and assuming additive genetic effects under three models: (1) marginal genetic association test of WMH volume (SNP-main effect); (2) SNP-main effect adjusted for HTN status; and (3) joint association test of both SNP-main and SNP-by-HTN interaction effects in relation with WMH volume:

\[
Y = \beta_0 + \beta_1 \text{SNP} + \beta_2 \text{Cov} + \epsilon
\]

(1)

\[
Y = \beta_0 + \beta_1 \text{SNP} + \beta_2 \text{Cov} + \beta_3 \text{HTN} + \epsilon
\]

(2)

\[
Y = \beta_0 + \beta_1 \text{SNP} + \beta_2 \text{Cov} + \beta_3 \text{HTN} \times \text{SNP} + \beta_4 \text{HTN} + \epsilon
\]

(3)

where SNP corresponds to the dosage of a given genetic (G) variant, Cov is the dichotomous variable for HTN status, \(\text{Cov}\) is the vector of covariates, \(\text{G}\) is the vector of maternal SNPs, and \(\epsilon\) values are the corresponding regression
coefficients and error covariance ($\epsilon$) of $\beta$. The joint model (Model 3) provides effect estimates of $G$ and $G\epsilon$, their robust standard errors ($SE$) and robust covariance matrices, and a $P$ value from a $2$-degree-of-freedom (df) Wald test. Robust estimates of $SE$s and covariance matrices were used.

All analyses were adjusted for age, sex, PCs of population stratification and intracranial volume (ICV). Adjustment for ICV was not performed for studies using visual grading of WMH burden, as visual grades are inherently normalized for brain size (Supplementary Data 1, Supplementary Methods 1).

**Genome-wide association meta-analyses.** A custom harmonization script along with the R package EasyQC23 was used to perform the QC of cohort-specific GWAS results. SNPs with minor allele frequency (MAF) lower than 1%, poor imputation quality ($R^2 < 0.80$), or a product of MAF, $R^2$ and sample size less than $10$, or $15$ for the $2$-df interaction analysis were excluded.

Inverse-variance weighted meta-analysis was conducted using METAL75, first within each cohort followed by a meta-analysis of the ancestry-specific results. A patch implemented in the METAL75 software was used to perform a $2$-df joint meta-analysis (JMA) with inverse-variance weighting. For cohort-specific GWAS results with genomic inflation factors ($\lambda$) exceeding $1$, genomic control (GC) correction was applied to correct for any residual population stratification. The effect allele for each risk variant was defined as the allele associated with larger WMH volume. For associations with individual SNPs the significance threshold was set at $P \leq 2 \times 10^{-3}$ ($0.00525$). The aggregate effect of $25$ WMH risk variants with DTI metrics was estimated using the GTX package in R70. The association of FA, MD, RD, AD, and PSMD with reaction time on the Stroop test, reflecting information processing speed, was examined using linear regression in $i$-Share participants who underwent both MRI and cognitive testing ($N = 1401$). Analyses were adjusted for age, sex, ICV, study-curriculum and ethnic origin. The association $p$ value was adjusted for the number of independent comparisons made ($n = 3$), estimated based on the correlation matrix between the DTI traits from $i$-Share and by applying the Matrix Spectral Decomposition (matSpDlite) method (http://neurogenetics.qimrberghofer.edu.au/matSpDlite/).

**Shared genetic architecture of WMH with related traits.** We systematically explored the genetic overlap of WMH SNP-main-effects (in the European-only analysis) with ($i$) neurological traits (any stroke, ischemic stroke, small vessel stroke, large vessel stroke, cardiomyocyte stroke; any, deep, and lobar intracerebral hemorrhage; general cognitive ability and Alzheimer's disease) and ($ii$) vascular risk factors and traits (SBP, DBP, PP, HDL-cholesterol, LDL-cholesterol, TG, BMI, T2D, Hba1c, SMKindex, VTE, and migraine). We acquired summary statistics of European-only analyses for these traits, using the latest largest GWAS, seeking collaboration with the relevant consortia when the data were not publicly available.

We first explored the association of lead WMH risk variants ($n = 27$) with related vascular and neurological traits. For each of the related traits, association statistics of SNPs falling in a window of $\pm 250$kb around each of the lead WMH SNP were retrieved and SNPs satisfying the multiplicity testing threshold defined by $P$ value for the effective number of independent SNPs ($n_{indep}$) as implemented in Genetic Type 1 error calculator (GEO)90, were retained. Only SNPs showing an association with a related vascular or neurological trait at $P < 1.3 \times 10^{-4}$ (accounting for $14$ independent traits and $27$ independent loci) and in moderate to high lead with the WMH SNP ($r^2 > 0.50$) are reported. The correlation matrix estimated between the traits using individual-level DTI variants implemented from the $3$C study77 was used to estimate the number of independent traits by applying the Matrix Spectral Decomposition (matSpDlite) method (http://neurogenetics.qimrberghofer.edu.au/matSpDlite/).

Using LDSR91, genetic correlation estimates between WMH and the aforementioned neurological and vascular traits were obtained. $p$ values $< 3.6 \times 10^{-3}$ correcting for $14$ independent phenotypes was considered significant. As genome-wide correlation estimates may miss significant correlations at the regional level (balancing-effect)41, the Bayesian pairwise GWAS approach (GWAS-PW) was applied92. GWAS-PW identifies trait pairs with high posterior probability of association of a given trait with a window of LD independent markers per locus, as implemented in Genetic Type 1 error calculator (GEO)90, were retained. Only SNPs showing an association with a related vascular or neurological trait at $P < 1.3 \times 10^{-4}$ (accounting for $14$ independent traits and $27$ independent loci) and in moderate to high lead with the WMH SNP ($r^2 > 0.50$) are reported. The correlation matrix estimated between the traits using individual-level DTI variants implemented from the $3$C study77 was used to estimate the number of independent traits by applying the Matrix Spectral Decomposition (matSpDlite) method (http://neurogenetics.qimrberghofer.edu.au/matSpDlite/).

Using LDSR91, genetic correlation estimates between WMH and the aforementioned neurological and vascular traits were obtained. $p$ values $< 3.6 \times 10^{-3}$ correcting for $14$ independent phenotypes was considered significant. As genome-wide correlation estimates may miss significant correlations at the regional level (balancing-effect)41, the Bayesian pairwise GWAS approach (GWAS-PW) was applied92. GWAS-PW identifies trait pairs with high posterior probability of association of a given trait with a window of LD independent markers per locus, as implemented in Genetic Type 1 error calculator (GEO)90, were retained. Only SNPs showing an association with a related vascular or neurological trait at $P < 1.3 \times 10^{-4}$ (accounting for $14$ independent traits and $27$ independent loci) and in moderate to high lead with the WMH SNP ($r^2 > 0.50$) are reported. The correlation matrix estimated between the traits using individual-level DTI variants implemented from the $3$C study77 was used to estimate the number of independent traits by applying the Matrix Spectral Decomposition (matSpDlite) method (http://neurogenetics.qimrberghofer.edu.au/matSpDlite/).

Using LDSR91, genetic correlation estimates between WMH and the aforementioned neurological and vascular traits were obtained. $p$ values $< 3.6 \times 10^{-3}$ correcting for $14$ independent phenotypes was considered significant. As genome-wide correlation estimates may miss significant correlations at the regional level (balancing-effect)41, the Bayesian pairwise GWAS approach (GWAS-PW) was applied92. GWAS-PW identifies trait pairs with high posterior probability of association of a given trait with a window of LD independent markers per locus, as implemented in Genetic Type 1 error calculator (GEO)90, were retained. Only SNPs showing an association with a related vascular or neurological trait at $P < 1.3 \times 10^{-4}$ (accounting for $14$ independent traits and $27$ independent loci) and in moderate to high lead with the WMH SNP ($r^2 > 0.50$) are reported. The correlation matrix estimated between the traits using individual-level DTI variants implemented from the $3$C study77 was used to estimate the number of independent traits by applying the Matrix Spectral Decomposition (matSpDlite) method (http://neurogenetics.qimrberghofer.edu.au/matSpDlite/).
variant (j) over corresponding beta-coefficients from the SNP-exposure associations (β). The ratio estimates from each genetic variant were averaged after weighting on the inverse variance (Wj) of across L uncorrelated SNPs, implemented as an R package RadialMR (available through CRAN repositories). The effect alleles for each risk variant were defined as the allele associated with increasing in the corresponding trait values. A p value < 3.6 × 10^{-3} correcting for 14 independent traits was considered significant. Cochran’s Q statistic was used to test for the presence of heterogeneity (Q Cochran < 0.01) due to horizontal pleiotropy that occurs when instruments exert an effect on the outcome and exposure through independent pathways. Influenntial outlier SNPs that have the largest Z-score contribution to the global Cochran’s Q statistic are identified by regressing the predicted causal estimate against the inverse-variance weights. After excluding the influential outlier SNPs, the IVW test was repeated along with MR-Egger regression. Relative goodness of fit of the MR-Egger over the IVW approach was quantified using the Qne statistics, which is the ratio of the statistical heterogeneity around the MR-Egger fitted slope divided by the statistical heterogeneity around the IVW slope. A Qne close to 1 indicates that MR-Egger is not a better fit to the data and therefore offers no benefit over IVW. Non-significant MR-Egger intercept was used as an indicator to formally rule out horizontal pleiotropy.

Cell and tissue type enrichment analysis. Association statistics from the WMH SNP-main effects (European-only) were used to test cell/tissue-specific enrichment. First, we used the EPIGWAS software and histone marks for promoters (H3K4me3) and enhancers (H3K4me1) from publicly available information on tissue-specific histone regulatory marks (Supplementary Methods 2). EPIGWAS calculates specificity scores for the lead WMH risk variant and its proxies (r² ≥ 0.80, 1000 G EUR based on the distance to the strongest CHIP-seq signal) and tests enrichment significance by comparing the relative proximity and specificity of the test set with 10,000 sets of matched background (using permutation). Bonferroni correction for the number of histone marks tested was applied (P < 2.5 × 10^{-3}). Second, we used MAGMA (gene-property analysis) and differentially expressed gene sets from the single-cell transcriptomic (scRNA) data in mouse brain from the Karolinska Institute. MAGMA generates gene-level association statistics by combining SNP p-values in a specified window (10 kb upstream and 1.5 kb downstream of each gene) accounting for LD (1000 G EUR) and under a linear regression framework performs a one-sided test between the association of genes with WMH volume and cell specificity. Using the MAGMA celltyping R package, scRNA expression values were obtained from five different mouse brain regions (neocortex, hippocampus, hypothalamus, striatum, and midbrain). A gene-expression specificity metric for each cell-type was calculated by dividing the expression level in a given cell type by the sum of the expression levels from all cell types (i.e., genes with high expression levels in two or more cell types will get a lower specificity measure than genes with high expression levels in a single-cell type), followed by binning the metric value to 40 equally sized bins. The MAGMA one-sided test was then used to test for enrichment between the top 10 percentile bins (bins with higher cell-specificity) in each cell type. Bonferroni correction for the number of cell types tested was applied (P < 2.1 × 10^{-3}).

Transcriptome-wide association study and colocalization. We performed transcriptome-wide association studies (TWAS) using the association statistics from the WMH SNP-main effects (European-only) and weights from 22 publicly available gene-expression reference panels (Supplementary Methods 2) from blood (Netherlands Twin Registry, NTR; Young Finns Study, YFS), arterial (Genotype-Tissue Expression, GTEx), brain (GTEx, CommonMind Consortium, CMC) and peripheral nerve tissues (GTEx). For each gene in the reference panel, precomputed SNP-expression weights in the 1-Mb window were obtained (Supplementary Methods 2), including the highly tissue-specific splicing QT (qTL) information on gene isoforms in the dorsolateral prefrontal cortex (DLPFC) derived from the CMC. Additionally, publicly available genome-wide expression weight (E) across the curated drug-target list DLPFC of 494 older individuals from two large community-based studies (the Religious Order Study [ROS] and the Rush Memory Aging Project [MAP]) were obtained. TWAS-Fusion was used to estimate the TWAS Z score (association statistic between predicted expression and WMH), derived from the SNP-expression weights, SNP-WMH effect estimates and the SNP correlation matrix. Transcriptome-wide (TW) significant genes (eGenes) and the corresponding QTs (eQTs) were determined using Bonferroni correction in each reference panel, based on the average number of features (4360 genes) tested across all the reference panels. eGene regions with eQTs not reaching genome-wide significance in association with WMH, and not in LD (r² < 0.01) with the lead SNP for genome-wide significant WMH risk loci, were considered as novel. Finally, a colocalization analysis (COLOC) was carried out at each locus to estimate the posterior probability of shared causal variance (PP4 ≥ 0.75) between the gene expression and trait association, using a prior probability of 1.1 × 10^{-4} for the WMH association. Furthermore, the functional validation of the eGenes was performed by testing for positional overlap of the best eQTs from TWAS with enhancer (H3K4me1) and/or promoter (H3K4me3) elements across a broad category of relevant tissue types (blood, brain/neurological, heart/arterial) using Haploreg V4.1.86. A value of 1 was assigned to eQTs with regulatory epigenome overlap in at least one tissue.

Drug-target enrichment. The Genome for REPositioning drugs (GReP) tool26 was used to quantify the enrichment of eGenes emerging from the TWAS with high probability of colocalization (PP4 ≥ 0.75) in the curated drug-target list classified based on the International Classification of Diseases 10 (ICD10). GReP provides as an output the names of the drug(s) targeting a given gene set along with the drug category. Moreover, by performing a series of Fisher’s exact tests GReP formally tests whether the gene set is enriched in genes targeted by drugs in a specific clinical indication category to treat a certain disease or condition.

Reporting summary. Further information on research design is available in the Nature Research Life Sciences Reporting Summary linked to this article.

Data availability
Summary statistics for the GWAS meta-analysis of the CHARGE cohorts and the UK Biobank on WMH burden main-effects generated and analyzed in the downstream analyses are deposited in a public repository (dbGAP: https://www.ncbi.nlm.nih.gov/gap/) under the accession number: phs002227.v1.p1. All other data supporting the findings of this study are available either within the article, the supplementary information and supplementary data files, or from the authors upon request. Publicly available resources used for this study were the UK Biobank [http://www.ukbiobank.ac.uk/]. Gene-expression weights for TWAS [https://gusevlab.org/projects/fusion/]. Magma.Celltyping [https://github.com/NathanSkene/MAGMA_Celltyping]; Histone regulatory marks [http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/]. Genome-wide summary statistics for other complex disorders were downloaded from public repositories [Supplementary Methods 2 for URL links].

Received: 26 January 2020; Accepted: 10 September 2020;
Published online: 08 December 2020

References
1. Debette, S. & Markus, H. S. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. BMJ 341, c666 (2010).
2. Shi, Y. et al. Cerebral blood flow in small vessel disease: a systematic review and meta-analysis. J. Cereb. Blood Flow Metab. 36, 1653–1667 (2016).
3. Debette, S., Schilling, S., Duperron, M. G., Larsson, S. C. & Markus, H. S. Clinical significance of magnetic resonance imaging markers of vascular brain injury: a systematic review and meta-analysis. JAMA Neurol. 76, 81–94 (2019).
4. Corriveau, R. A. et al. Alzheimer’s disease-related Dementias Summit 2016: National research priorities. Neurology 89, 2381–2391 (2017).
5. The Lancet, N. Vascular disease and neurodegeneration: advancing together. Lancet Neurol. 16, 333 (2017).
6. Alber, J. et al. White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VICID): Knowledge gaps and opportunities. Alzheimers Dement. (N. Y) 5, 107–117 (2019).
7. Dufouil, C. et al. Effects of blood pressure lowering on cerebral white matter hyperintensities in patients with stroke: the PROGRESS (Perindopril Protection Against Recurrent Stroke Study) Magnetic Resonance Imaging Substudy. Circulation 112, 1644–1650 (2005).
8. Nasrallah, I. M. et al. Association of intensive vs standard blood pressure control with cerebral white matter lesions. JAMA 322, 524–534 (2019).
9. Duperron, M. G. et al. Burden of dilated perivascular spaces, an emerging marker of cerebral small vessel disease, is highly heritable. Stroke 49, 282–287 (2018).
10. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. Nat. Genet. 47, 856–860 (2015).
11. Ference, B. A. et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. N. Engl. J. Med. 375, 2144–2153 (2016).
12. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376–381 (2014).
13. Chauhan, G. et al. Genetic and lifestyle risk factors for MCI-defined brain infarcts in a population-based setting. Neurology 92, 486–503 (2019).
14. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat. Genet. 50, 524–537 (2018).
15. Won, D. et al. Meta-analysis of genome-wide association studies identifies lq2 as a susceptibility locus for intracerebral hemorrhage. Am. J. Hum. Genet. 94, 511–521 (2014).
16. Traylor, M. et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. Neurology 86, 146–153 (2016).
74. Willer, J. C., Li, Y. & Abecasis, R. G. A. META: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

75. Manning, A. K. et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet. Epidemiol.* **35**, 11–18 (2011).

76. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

77. The 3C Study Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuropediatrics* **22**, 316–325 (2003).

78. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 190–196 (2014).

79. International Consortium for Blood Pressure Genome-Wide Association, S. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).

80. Li, M.-X., Yeung, Y. J. M., Cherny, S. S. & Sham, C. P. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131**, 747–756 (2012).

81. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

82. Sverdlov, D. I. et al. Selecting instruments for Mendelian randomization in the wake of genomewide association studies. *Int. J. Epidemiol.* **45**, 1600–1616 (2016).

83. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

84. Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).

85. Ward, L. D. & Kells, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).

86. Sakase, S. & Okada, Y. GREP: genome for REPositioning drugs. *Bioinformatics* **35**, 3821–3823 (2019).

Acknowledgements

We thank all the participating cohorts from the CHARGE consortium and the UK biobank participants for contributing to this study. Detailed study-specific acknowledgements are provided in the supplementary note.

### Author contributions

W.T.L., J.J.L., M.L., S.Seshadri, C.T., H.H.A., P.M.M., M.F., and S.D. jointly supervised research. M.S., W.T.L., L.J.L., M.L., S.Seshadri, C.T., H.H.A., P.M.M., M.F., and S.D. wrote and edited the manuscript.

M.S., H.Suzuki, J.I., C.S., T.E.E., G.E., S.Sakaue, N.T., M.H., W.Z., N.J.A., E.H., L.R.Y., S.P.H., R.B.K., E.B.V., R.E.M., S.T., A.M., Y.S., C.L.S., G.B., L.F., A.Tsuchida, L.Z., S.Schilling, S.Sigurdsson, R.F.G., C.E.I., N.T.A., O.L.L., J.A.S., M.C.V., J.V., M.J.W., M.J.K., M.D., R.J.T., C.B., M.G.D., A.V.S., D.S.K., P.J.S., D.A.E., J.I.R., A.S.B., S.M.M., M.B., J.T., D.J.S., M.W.V., K.W., J.N., A.S., E.B., S.Sidney, S.T.T., G.D., A.Thalamuthu, U.V., M.A.V., N.R.B., J.D., M.E.B., D.A., A.Tseumer, P.A., I.B.K., R.B., J.I.D., P.R.S., H.B., J.I., Y.T., K.S., S.M., K.Y., M.N., Y.K., F.M., B.M.P., D.A.B., P.L.D., T.H.M., P.S.S., R.S., H.R.W., E.E., D.T., M.A.I., W.V., C.D., V.K.S., J.Wouter.J., E.P.S., S.I.R.K., Y.O., R.M., J.M.W., P.A., K.A.M., H.J.G., H.Schmidt, and V.G. conducted cohort-wise GWAS analysis. M.S. performed shared genetic architecture and functional characterization analysis. M.S., C.B., and A.S. performed lifetime impact analysis. M.S. and S.Sakaue performed drug-target enrichment. M.S., W.T.L., J.J.L., M.L., S.Seshadri, C.T., H.H.A., P.M.M., M.F., and S.D. wrote and edited the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41467-020-19111-2.

**Correspondence** and requests for materials should be addressed to M.F. or S.D.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer review reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Authors (2020)
Aicha Soumamé, Eric Boerwinkle, Stephen Sidney, Stephen T. Turner, Gail Davies, Anupamal Thalamuthu, Uwe Völker, Mark A. van Buchem, R. Nick Bryan, Josée Dupuis, Mark E. Bastin, David Ames, Alexander Teumer, Philippe Amouyel, John B. Kwock, Robin Bülow, Ian J. Deary, Peter R. Schofield, Henry Brodaty, Jiyang Jiang, Yasuharu Tabara, Kazuaya Setoh, Susumu Miyamoto, Kazumichi Yoshida, Manabu Nagata, Yoichiro Kamatani, Fumihiko Matsuda, Bruce M. Psaty, David A. Bennett, Philip L. De Jager, Thomas H. Mosley, Perminer S. Sachdev, Reinhold Schmidt, Helen R. Warren, Evangelos Evangelou, International Network against Thrombosis (INVENT) Consortium, International Headache Genomics Consortium (IHGC), Mohammad A. Ikrami, Wei Wen, Charles DeCarli, Velandai K. Srikanta, J. Wouter Jukema, Eline P. Slagboom, Sharon L. R. Kardia, Yukinori Okada, Bernard Mazoyer, Joanna M. Wardlaw, Paul A. Nyquist, Karen A. Mather, Hans J. Grabe, Helena Schmidt, Cornelia M. Van Duijn, Vilmundur Gudnason, William T. Longstreth Jr., Lenore J. Launer, Mark Lathrop, Sudha Seshadri, Christophe Tzourio, Hieab H. Adams, Paul M. Matthews, Myriam Fornage, & Stéphanie Debette.

1University of Bordeaux, Inserm, Bordeaux Research Health Center, team VINTAGE, UMR 1219, 33000 Bordeaux, France. 2Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo, Aoba, Sendai 980-8573, Japan. 3Department of Cardiovascular Medicine, Tohoku University Hospital, 1-1, Seiryo, Aoba, Sendai 980-8574, Japan. 4Department of Brain Sciences, Imperial College London, London W12 ONN, UK. 5University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX 77030, USA. 6Glenn Biggs Institute for Alzheimer’s & Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, TX 78229, USA. 7Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA. 8Department of Clinical Genetics, Erasmus MC, 3015 GE Rotterdam, The Netherlands. 9Department of Radiology & Nuclear Medicine, Erasmus MC, 3015 GE Rotterdam, The Netherlands. 10Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA. 11Icelandic Heart Association, IS-201 Kópavogur, Iceland. 12Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan. 13Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Tsukuba, Ibaraki, 305-0002, Japan. 14Department of Allergy and Rhematology, Graduate School of Medicine, the University of Tokyo, Tokyo 113-0033, Japan. 15Department of Epidemiology, Erasmus MC, 3015 GE Rotterdam, The Netherlands. 16Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. 17Institute for Community Medicine, Medical University Greifswald, 17475 Greifswald, Germany. 18Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109-2029, USA. 19Mathematics and Statistics, Murdoch University, Murdoch, WA 6150, Australia. 20Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, 8036 Graz, Austria. 21Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, 8036 Graz, Austria. 22GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. 23Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK. 24Social Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London SE5 8AF, UK. 25Department of Public Health Sciences, University of California at Davis, Davis, CA 95616, USA. 26Section of Molecular Epidemiology, Biomedical Sciences, Leiden university medical Center, 2333 ZA Leiden, The Netherlands. 27Pattern Recognition & Bioinformatics, Delft University of Technology, Delft, NL 2629 HS, USA. 28Leiden Computational Biology Centre, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands. 29Centre for Law and Genetics, Faculty of Law, University of Tasmania, Hobart, TAS 7005, Australia. 30Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7000, Australia. 31Department of Internal Medicine, section of gerontology and geriatrics, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands. 32Department of Cardiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands. 33Gottfried Schatz Research Center, Department of Molecular Biology and Biochemistry, Medical University of Graz, 8010 Graz, Austria. 34Boston University and the NHLBI’s Framingham Heart Study, Boston, MA 02215, USA. 35Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA. 36University of Bordeaux, IMN, UMR 5293, 33000 Bordeaux, France. 37Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. 38University of Alabama at Birmingham School of Medicine, Birmingham, AL 35223, USA. 39Department of Neurological Sciences, Rush University Medical Center, Chicago, IL 60612, USA. 40Department of Neurology and Psychiatry, Boston University School of Medicine, Boston, MA 02118, USA. 41Survey Research Center, Institute of Social Research, University of Michigan, Ann Arbor, MI 48104, USA. 42Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH16 4SB, UK. 43Row Fogo Centre for Ageing and the Brain, University of Edinburgh, Edinburgh EH9 1JZ, UK. 44Department of Radiology, Leiden University medical center, 2333 ZA Leiden, The Netherlands. 45Queenland Brain Institute, The University of Queensland, St Lucia, QLD 4072, Australia. 46Department of Internal Medicine B, University Medicine Greifswald, 17475 Greifswald, Germany. 47DZHK (German Center for Cardiovascular Research), partner site Greifswald, 17475 Greifswald, Germany. 48Centre for Research in Mathematics and Data Science, Western Sydney University, Penrith, NSW 2751, Australia. 49Mayo Clinic, Rochester, MN 55905, USA. 50University of Minnesota School of Public Health, Minneapolis, MN 55455, USA. 51Department of Internal Medicine, Rush University Medical Center, Chicago, IL 60612, USA. 52Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Pediatrics at Harbor-UCLA Medical Center, Torrance, CA 90502, USA. 53Centre for Healthy Brain Ageing, School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia. 54Department of Developmental Disability Neuropsychiatry, School of Psychiatry, University of New South Wales, Sydney, NSW 2052, Australia. 55Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK. 56Department for Neurodegenerative Diseases (DZNE), Rostock/Greifswald, 17489 Greifswald, Germany. 57Faculty of Applied Sciences, Delft University of Technology, Delft, NL 2629, USA. 58University of Texas Health Science Center at Houston School of Public Health, Houston, TX 77030, USA.
International Network against Thrombosis (INVENT) Consortium

Philippe Amouyel, Marie de Andrade, Saonli Basu, Claudine Berr, Jennifer A. Brody, Daniel I. Chapman, Jean-Francois Dartigues, Aaron R. Folsom, Marine Germain, Hugoline de Haan, John Heit, Jeanine Houwing-Duistermaat, Christopher Kabrhel, Peter Kraft, Grégoire Legal, Sara Lindström, Ramin Monajemi, Pieter-Emmanuel Morange, Bruce M. Psaty, Pieter H. Reitsma, Frits R. Rosendaa, Noémie Saut, Eline Slagboom, Paul M. Matthews, Myriam Fornage, Stéphanie Debette.

*Lists of authors and their affiliations appear at the end of the paper.

Email: myriam.fornage@uth.tmc.edu; stephanie.denette@u-bordeaux.fr
International Headache Genomics Consortium (IHGC)

Padhraig Gormley137,138,139,140, Verner Anttila138,139,141, Bendik S. Winsvold142,143,144, Priit Palta145, Tonu Esko138,146,147, Tune H. Pers138,147,148,149, Kai-How Farh138,141,150, Ester Cuenca-Leon137,138,139,151, Mikko Muona145,152,153,154, Nicholas A. Furlotte155, Tobias Kurth156,157, Andres Ingason158, George McMahon159, Lannie Ligthart160,161, Mikko Kallela162, Tobias M. Freilinger163,164, Caroline Ran165, Scott G. Gordon166, Anine H. Stam167, Stacy Steinberg158, Guntram Borck167, Markku Koiranen168,169, Lydia Quaye170, Hieab H. Adams171,172, Terho Lehtimäki173, Antti-Pekka Sarin145,146, Juho Wedenjo174, David A. Hinds155, Julie E. Buring157,175, Markus Schürks176, Paul M. Ridker157,175, Maria Gudlaug Hrafnsdottir177, Hreinn Stefansson158, Susan M. Ring159, Jouke-Jan Hottenga160, Brenda W. J. H. Penninx178, Markus Färkkilä162, Ville Arto162, Mari Kaunisto170, Salli Vepsäläinen162, Rainer Malik163, Andrew C. Heath179, Pamela A. F. Madden179, Nicholas G. Martin166, Grant W. Montgomery166, Mitja Kurki137,138,139, Mart Kala146, Reedik Mägi146, Kalle Pärn146, Eija Hämäläinen145, Hailiang Huang138,139,141, Andrea E. Byrnes138,139,141, Lude Franke180, Jie Huang140, Evie Stergiakouli159, Philip H. Lee137,138,139, Cynthia Sander181, Caleb Webber181, Zameel Cader182,183, Bertram Muller-Mylsko184, Stefan Schreiber185, Thomas Meitinger186, Johan G. Eriksson187,188, Veikko Salomaa188, Kauko Heikkilä189, Elizabeth Loehrer171,190, Andre G. Uitterlinden191, Albert Hofman171, Cornelia M. van Duijn171, Lynn Cherkas170, Linda M. Pedersen142, Audun Stubhaug192,193, Christopher S. Nielsen192,194, Minna Männikkö168,169, Eelin Mihailov146, Lili Milani146, Hartmut Göbel195, Ann-Louise Esserlind196, Anne Francke Christensen196, Thomas Folkman Hansen197, Thomas Werge198,199,200, Jaakko Kaprio145,174,201, Arpo J. Aromaa188, Olli Raitakari202,203, M. Arfan Iqram171,172,203,204, Tim Spector170, Marjo-Riitta Järvelin168,169,205,206,207,208, Andres Metspalu146, Christian Kubisch120, David P. Strachan210, Michel D. Ferrar161, Andrea C. Belin165, Martin Dichgans163,211, Majia Wessman145,152, Arn M. J. van den Maagdenberg161,212, John-Anker Zwart142,143,144, Dorret I. Boomsma160, George Davey Smith159, Kari Stefansson158,213, Nicholas Eriksson155, Mark J. Daly138,139,141, Benjamin M. Neale138,139,141, Jes Olesen196, Daniel I. Chasman157,175, Dale R. Nyholt214 & Aarno Polatii137,138,139,140,141,145,215

137Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. 138Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. 139Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. 140Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, UK. 141Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. 142GORMI, Oslo University Hospital, P.O. 4956 Nydalen, 0424 Oslo, Norway. 143Department of Neurology, Oslo University Hospital, P.O. 4956 Nydalen, 0424 Oslo, Norway. 144Institute of Clinical Medicine, University of Oslo, P.O. 1171 Blindern, 0318 Oslo, Norway. 145Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. 146Estonian Genome Center, University of Tartu, Tartu, Estonia. 147Division of Endocrinology, Boston Children’s Hospital, Boston, MA, USA. 148Statens Serum Institut, Dept of Epidemiology Research, Copenhagen, Denmark. 149Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark. 150Illinima, 5200 Illinima Way, San Diego, CA, USA. 151Vall d’Hebron Research Institute, Pediatric Neurology, Barcelona, Spain. 152Folkhälso Institute of Genetics, FI-00290 Helsinki, Finland. 153Neuroscience Center, University of Helsinki, FI-00014 Helsinki, Finland. 154Research Programs Unit, Molecular Neurology, University of Helsinki, FI-00014 Helsinki, Finland. 15523andMe, Inc., 899 W. Evelyn Avenue, Mountain View, CA, USA. 156deCODE Genetics, 101 Reykjavik, Iceland. 157Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK. 158deCODE Genetics, 101 Reykjavik, Iceland. 159MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. 15623andMe, Inc., 899 W. Evelyn Avenue, Mountain View, CA, USA. 157Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK. 158deCODE Genetics, 101 Reykjavik, Iceland. 159MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. 158deCODE Genetics, 101 Reykjavik, Iceland. 159MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
