

Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes

Stroke has multiple etiologies, but the underlying genes and pathways are largely unknown. We conducted a multiancestry genome-wide-association meta-analysis in 521,612 individuals (67,162 cases and 454,450 controls) and discovered 22 new stroke risk loci, bringing the total to 32. We further found shared genetic variation with related vascular traits, including blood pressure, cardiac traits, and venous thromboembolism, at individual loci ($n=18$), and using genetic risk scores and linkage-disequilibrium-score regression. Several loci exhibited distinct association and pleiotropy patterns for etiological stroke subtypes. Eleven new susceptibility loci indicate mechanisms not previously implicated in stroke pathophysiology, with prioritization of risk variants and genes accomplished through bioinformatics analyses using extensive functional datasets. Stroke risk loci were significantly enriched in drug targets for antithrombotic therapy.

Stroke is the second leading cause of death and disability-adjusted life years worldwide^{1,2}. Characterized by a neurological deficit of sudden onset, stroke is primarily caused by brain infarction (ischemic stroke) and, less often, by intracerebral hemorrhage (ICH). Common etiological subtypes of ischemic stroke include large-artery atherosclerotic stroke (LAS), cardioembolic stroke (CES), and stroke caused by small-vessel disease (small-vessel stroke (SVS)), which is also the leading cause of ICH. Previous genome-wide association studies (GWAS) in predominantly European-ancestry groups have identified ten loci robustly associated with stroke^{3–12}. In most instances, the associations with stroke were attributed to individual subtypes of ischemic stroke, such as LAS^{5,8,9}, CES^{3,4}, and SVS^{10,12}, or of ICH⁶, although some loci were associated with two or more stroke subtypes^{7,9,11,13} or with any stroke¹⁰. We hypothesized that combining a substantially larger sample size with a transancestral analytic approach would identify additional risk loci and improve fine mapping of causal variants. Hence, we combined all available stroke samples with published or unpublished GWAS data, including samples of non-European ancestry that were underrepresented in previous GWAS. We further hypothesized that stroke shares genetic influences with vascular risk factors, intermediate phenotypes for stroke (for example, carotid artery plaque (cPL)), and related phenotypes (for example, coronary artery disease (CAD)) and that a systematic approach to identify genetic influences shared among these traits would provide insights into stroke pathophysiology.

Results

We tested ~8 million SNPs and indels with minor-allele frequency (MAF) ≥ 0.01 in up to 67,162 stroke cases and 454,450 controls for association with stroke. One analysis involved European participants only (40,585 cases; 406,111 controls), and a second involved participants of European, East Asian (17,369; 28,195), African (5,541; 15,154), South Asian (2,437; 6,707), mixed Asian (365; 333), and Latin American (865; 692) ancestry (Fig. 1). Participants were drawn from 29 studies with genome-wide genotypes imputed to 1000 Genomes Project (1000G) phase 1v3 or similar¹⁴ (MEGASTROKE consortium; Supplementary Note and Supplementary Tables 1 and 2). Ancestry-specific meta-analyses and subsequent fixed-effects transancestral meta-analyses and MANTRA transancestral meta-analyses were

conducted¹⁵. Analyses were performed for any stroke (AS), comprising ischemic stroke, ICH, and stroke of unknown or undetermined type ($n=67,162$); any ischemic stroke (AIS) regardless of subtype ($n=60,341$); and ischemic stroke subtypes (LAS, $n=6,688$; CES, $n=9,006$; SVS, $n=11,710$).

New genome-wide-significant stroke loci. We identified 32 genome-wide significant loci, 22 of which were novel (Table 1, Fig. 2, Supplementary Tables 3 and 4, and Supplementary Figs. 1–7). Of the 22 novel loci, 18 were identified by transancestral meta-analyses (fixed-effects $P < 5.0 \times 10^{-8}$ or MANTRA $\log_{10}(\text{Bayes factor (BF)}) > 6$) (Fig. 2 and Supplementary Figs. 1–5), and the remaining four loci were identified by the ancestry-specific meta-analysis in European samples (fixed-effects $P < 5.0 \times 10^{-8}$) (Fig. 2 and Supplementary Figs. 1–5). Apart from two novel loci with a MAF between 0.01 and 0.05 and large effect-size estimates (odds ratios (ORs) of 2.33 and 1.95), the remaining 20 novel loci contained common variants (MAF 0.16–0.48) with observed ORs between 1.05 and 1.20 (Table 1). Comparison of the 32 loci across Europeans and East Asians, the two largest ancestral subgroups, demonstrated significant correlations of risk-allele frequencies and ORs between populations (Supplementary Fig. 8), although six loci exhibited population-specific association (defined as $P < 5.0 \times 10^{-8}$ in Europeans and $P > 0.05$ in East Asians or MAF in East Asians < 0.01) (Supplementary Table 5). Estimates for the phenotypic variance explained by the 32 lead variants ranged between 0.6% and 1.8% (Supplementary Table 6).

Gene-based tests using VEGAS2 (ref. ¹⁶) (Supplementary Fig. 9) confirmed the loci identified by the GWAS analyses above and yielded a novel significant ($P < 2.02 \times 10^{-6}$, Bonferroni corrected for the number of genes) association of the neighboring genes *ICA1L* and *WDR12* with SVS (Supplementary Table 7 and Supplementary Figs. 9 and 10). Prior studies have demonstrated that variants in this region are associated with white-matter hyperintensity (WMH) burden¹⁷, a brain magnetic resonance imaging marker of small-vessel disease (SVD).

Twenty-one additional loci met a less stringent threshold for suggestive evidence of association ($\log_{10}(\text{BF}) > 5.0$ or $P < 1.0 \times 10^{-6}$ in the transancestral fixed-effects analysis) (Supplementary Table 8), including three loci previously implicated in Mendelian stroke (*HTRA1*, *COL4A1*, and *COL4A2*)^{18–21}.

Associations with etiological stroke subtypes. Genome-wide significance was reached for 18 loci (12 novel) for AS, 20 (12 novel) for AIS, 6 (3 novel) for LAS, 4 (2 novel) for CES, and 2 (*ICA1L-WDR12* novel, discovered in gene-based tests) for SVS (Fig. 2, Table 1, and Supplementary Figs. 1–5 and 10). Several loci reaching genome-wide significance for one of the ischemic stroke subtypes were also genome-wide significant for AIS or AS, whereas none reached genome-wide significance for multiple ischemic stroke subtypes (Fig. 2 and Supplementary Table 9). For some novel loci, the association was strictly confined to a single subtype ($P > 0.5$ for other stroke subtypes): *EDNRA* and *LINC01492* showed association with LAS only, thus suggesting mechanisms limited to atherosclerosis, and *NKX2-5* showed association with CES only, thus suggesting that the association may be primarily mediated by cardioembolism. We also found subtype specificity for previously described loci (*TSPAN2* for LAS and *PITX2* for CES). We further investigated shared genetic influences of individual loci on different stroke subtypes by using gwas-pw analyses²², which estimate the posterior probability that a specified genomic region influences two different traits. By applying a posterior-probability cutoff of 90% for shared contribution at a given locus (model 3), we found shared genetic influence between LAS and SVS at *SH2B3*, and between LAS and CES at *ABO* (Supplementary Table 10 and Supplementary Fig. 11).

Conditional analysis to identify independent signals within loci.

When conditioning all SNPs in a ± 0.5 -Mb window on the lead SNPs in the Europeans-only analysis, we found two additional independent genome-wide signals at the *PITX2* locus for CES, in agreement with known multiple independent loci at *PITX2* for atrial fibrillation (AF)²³, thus suggesting that a similar genetic architecture at this locus influences both conditions (Supplementary Fig. 12). We further found suggestive independent signals at *MMP12*, *SH2B3*, and *HDAC9-TWIST1* that did not reach genome-wide significance (Supplementary Table 11).

Association of individual stroke risk variants with related vascular traits.

Several of our loci are in the genomic vicinity of established risk loci for vascular risk factors (for example, blood pressure (BP)), and related vascular phenotypes affecting the heart (for example, CAD), vasculature (for example, carotid intima media thickness (cIMT)), or the brain (WMH). To systematically explore the genetic overlap between stroke and these traits, we surveyed published GWAS for BP, blood lipids, type 2 diabetes (T2D), cIMT, cPL, AF, venous thromboembolism (VTE), CAD, and WMH, assembled through the IGEN-BP²⁴, ENGAGE²⁵, DIAGRAM²⁶, CHARGE^{27,28}, AFGEn²⁹, INVENT³⁰, and CARDIoGRAMplusC4D³¹ consortia (Supplementary Table 12). When constructing sets of index SNPs of the nonstroke phenotypes (Bonferroni-adjusted $P < 1.3 \times 10^{-4} = 0.05/32$ loci/12 related vascular traits) and SNPs in high linkage disequilibrium (LD) ($r^2 > 0.9$ in the 1000G European-ancestry dataset (EUR)) with those index variants, 17 of the 32 stroke lead variants showed overlap with these sets (Fig. 3 and Supplementary Table 13). Fourteen loci reached genome-wide significance ($P < 5.0 \times 10^{-8}$) for association with one or more of the following phenotypes: BP (five loci), CAD (five loci), AF (two loci), VTE (two loci), low-density lipoprotein (LDL) cholesterol (two loci), cPL (one locus), and WMH (one locus). Among the 21 additional subthreshold loci for stroke (Supplementary Table 8), six loci have previously been associated with related vascular traits, including AF (*PRRX* and *CAV1-CAV2*)³², VTE (*F11*)³⁰, CAD (*SWAP70* and *LPA*)³¹, blood lipids (*LPA*)³¹, and WMH (*ICA1L-WDR12*)²⁸.

Association of genetic risk scores of related vascular traits. Second, we generated weighted genetic risk scores (wGRS) for VTE, BP-related traits, blood lipids, T2D, and CAD by using the lead SNPs

from published GWAS and tested these wGRS for association with each stroke phenotype, implementing the inverse-variance weighting approach (Methods and Supplementary Table 14). We found significant associations ($P < 5.6 \times 10^{-3}$, correcting for nine independent phenotypes; Methods) with wGRS for all traits examined, except for triglyceride and LDL-cholesterol levels, and observed clear differences between stroke subtypes (Fig. 4). The strongest association was between the wGRS for CAD and LAS, in agreement with shared pathophysiology through atherosclerosis. We further found associations of all stroke subtypes with wGRS for BP traits. The wGRS for VTE was significantly associated with both LAS and CES (all $P < 1.0 \times 10^{-4}$) but not SVS. The wGRS for high-density lipoprotein (HDL) cholesterol showed a significant inverse association with SVS.

In the present setting, the wGRS analysis was used primarily to explore the genetic overlap with related vascular traits rather than as a tool for establishing causal inference. In sensitivity analyses, we conducted an MR-Egger regression to explore whether any of the significant associations between vascular wGRS and stroke might be partly driven by directional pleiotropy. There was no indication of directional pleiotropy except for the association between the SBP wGRS and AS (MR-Egger intercept estimate $P = 0.015$), which was no longer significant after removal of 6 of 37 SNPs appearing as outliers from the leave-one-out analysis (Methods), thus leading to causal estimates in broad agreement across regression techniques (Supplementary Table 15).

Shared genetic contribution to stroke and related vascular traits genome wide.

Third, we applied LD-score regression to quantify the extent of shared genetic contributions between traits at a genome-wide level^{33,34}. Using available GWAS results from individuals of European ancestry, we found significant positive correlations ($r_g > 0$; $P < 5.6 \times 10^{-3}$, correcting for nine independent phenotypes), mostly corroborating the wGRS results (Fig. 4 and Supplementary Table 16). In addition, we found significant genetic overlap between triglyceride levels and AIS, and similar results were obtained in available GWAS datasets from individuals of East Asian ancestry (Supplementary Table 16). The results did not substantially change after removal of genome-wide signals for stroke and related vascular traits and their proxies ($r^2 \geq 0.8$ in 1000G EUR).

Global epigenetic patterns at the 32 stroke risk loci. To test for cell-specific enrichment in chromatin marks that were previously shown to be phenotypically cell-type specific in the Encyclopedia of DNA Elements (ENCODE)/RoadMap (histone H3 modifications H3K4me1, H3K4me3, and H3K9ac)³⁵, we implemented the epigwas tool³⁵ and the narrow peak information from the latest RoadMap dataset (127 tissues)³⁶. Epigwas estimates the enrichment score (ratio of the height of the nearest narrow peak to the distance to the peak) for the lead variant and proxies ($r^2 \geq 0.8$ in the 1000G cosmopolitan panel) and calculates statistical significance by examining the relative proximity and specificity of the test SNP set with 10,000 sets of matched background. The analysis showed significant enrichment of enhancer and promoter sites (marked by H3K4me1 and H3K4me3) in mesenchymal stem cells, embryonic stem cells, epithelial cells, and blood and T cells, and of active promoters (marked by H3K9ac) in embryonic stem cells and digestive tissue (Supplementary Table 17).

Pathway analyses. To identify pathways overrepresented in the stroke association results, we used the DEPICT gene-set enrichment tool³⁷, using all SNPs with $\log_{10}(\text{BF}) > 5$ for the respective stroke subtype. We found three gene sets to be significantly (false discovery rate (FDR) $< 5\%$) associated with AS: enlarged heart, decreased cardiac muscle contractility, and oxaloacetate metabolic process (Supplementary Table 18). Next, we used Ingenuity Pathway Analysis (IPA; URLs), examining genes within the 53 stroke loci with $\log_{10}(\text{BF}) > 5$. The extended gene list ($r^2 > 0.5$ in 1000G Europeans or East Asians, or

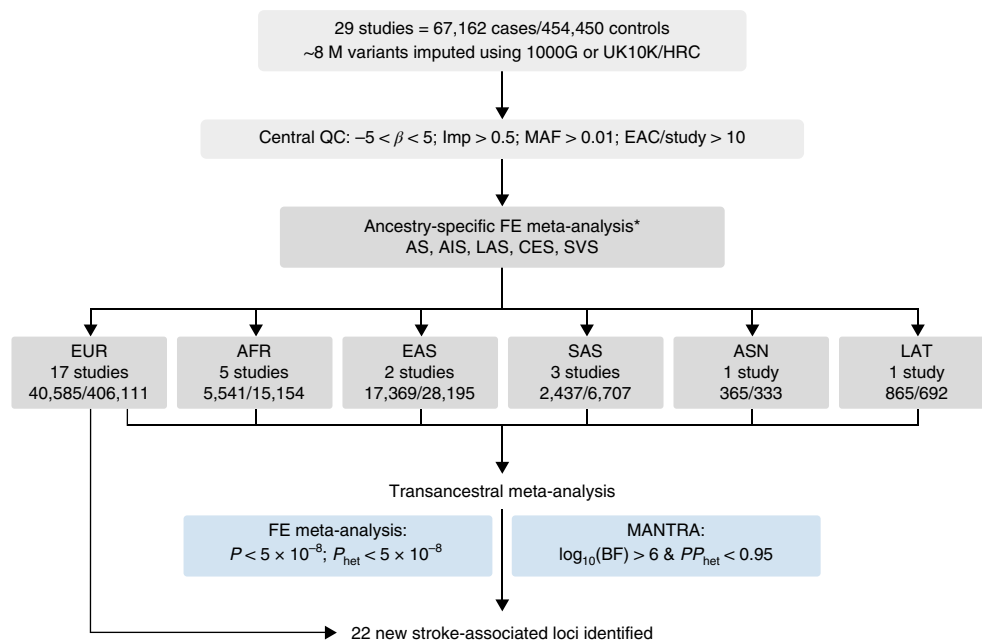


Fig. 1 | MEGASTROKE study design. Variants were retained that passed central quality control (QC) criteria (Methods). The numbers of cases and controls are listed for each ancestry group. HRC, Haplotype Reference Consortium; imp, measure of imputation quality (Methods); FE, fixed effects; EUR, European ancestry; AFR, African ancestry; EAS, East Asian ancestry; SAS, South Asian ancestry; ASN, mixed Asian ancestry; LAT, Latin American ancestry; P_{het} , heterogeneity P value; PP_{het} , posterior probability of heterogeneity. *The ASN and LAT ancestries were composed of a single study and hence did not require ancestry-specific meta-analysis.

located within 50 kb of the lead SNP) consisted of 214 genes. We found the coagulation system to be the most significant canonical pathway, followed by cardiomyocyte differentiation via bone-morphogenetic-protein receptors (FDR of 5%) (Supplementary Table 19). Finally, we tested enrichment of VEGAS2-derived gene-based P values in expert-curated and computationally predicted Biosystem gene sets³⁸, adapting VEGAS2Pathway³⁹, and identified significant association with 18 pathways, including various cardiac pathways, muscle-cell fate commitment, and nitric oxide metabolic process with CES (FDR of 5%) (Supplementary Table 20).

Fine mapping derived from credible SNP-set analyses. To decrease the number of candidate variants per locus to the most noteworthy associations, we constructed 95% credible SNP sets for each of the 32 loci (lead SNP and proxy SNPs $r^2 > 0.1$ in 1000G panels), assuming one causal SNP per locus and uniform priors⁴⁰. Credible SNP sets were generated in all stroke phenotypes and for European, East Asian, and African ancestries separately. We found a marked decrease in credible SNP sets for most loci, a result expectedly most pronounced for the phenotype showing the strongest association signal (Supplementary Table 21). The greatest refinement was observed at *RGS7*, *HDAC9-TWIST1*, and *SH2B3*, where the lead SNP was the only SNP contained in the 95% credible set for the stroke phenotype showing the strongest association.

Stroke loci with nonsynonymous or predicted deleterious variants. To determine SNPs with protein-altering effects, we annotated all SNPs by using ANNOVAR⁴¹. Of the 32 lead SNPs, three were exonic, of which two were nonsynonymous: rs3184504 (p.Arg262Trp) in *SH2B3* and rs1052053 (p.Gln75Arg) in *PMF1*. *SH2B3* p.Arg262Trp is a loss-of function variant that leads to expansion of hematopoietic stem cells and enhanced megakaryopoiesis in humans⁴². Both variants are predicted to be benign or tolerated by PolyPhen⁴³ and SIFT⁴⁴. In addition, we identified a proxy SNP ($r^2 = 0.99$ in 1000G EUR) for another lead SNP that was

nonsynonymous, rs6050 (p.Thr331Ala) in *FGA*, also predicted to be benign or tolerated.

Investigation of eQTLs, meQTLs, and pQTLs in different tissues. To determine whether stroke risk SNPs influenced the cis regulation of nearby genes, we interrogated genome-wide quantitative information (expression quantitative trait loci (eQTLs), methylation quantitative trait loci (meQTLs), and protein-expression quantitative trait loci (pQTLs)) in extensive publicly and nonpublicly available datasets. These datasets encompass numerous tissues and cell types, including cardiac, vascular, and brain tissue; circulating cells; and vascular endothelial cells (Methods). These comprised the following: for eQTLs, GTEx V6 (ref. 45), an expanded version of GRASP2 (refs 46,47), HGVD⁴⁸, BIOS⁴⁹, Blueprint epigenome project (subset)⁵⁰, STARNET⁵¹, and the human aortic endothelial cell study⁵²; for meQTLs, the Blueprint epigenome project (subset)⁵⁰ and the ARIC cohort⁵³; and for pQTLs, the KORA cohort⁵⁴. Only cis eQTLs, meQTLs, and pQTLs were considered.

We found that in 18 of the 32 stroke risk loci, the lead stroke risk variant either overlapped or was in moderate to high LD ($r^2 > 0.8$) with the most significant QTL variant for a nearby gene in at least one tissue or cell type (Supplementary Tables 22 and 23). For seven loci, we observed association of the lead SNP and proxies with expression of a single gene (or methylation or protein level), sometimes the nearest gene (*LRCH1*, *CDK6*, *CDKN2B*, *PRPF8*, and *MMP12*), and sometimes a more distant nearby gene (*ZCCHC14* for the *ZCCHC14* locus, and *TWIST1* for the *HDAC9-TWIST1* locus), within the datasets explored. Associations were found primarily in stroke-relevant tissues and cell types, including vascular tissues, aortic endothelial cells, brain, blood, and immune cells. In most instances (11 loci, 61.1%), the risk SNP affected expression of multiple genes, thus suggesting that at individual loci, pleiotropic mechanisms, which might differ according to tissue/cell type, may in some instances influence stroke susceptibility^{55,56}. For several of these loci, there was a clear predominance of eQTL associations

Table 1 | Results from the MANTRA (transancestral) and METAL fixed-effects (transancestral and Europeans-only) GWAS meta-analyses

| rsID | Chromosome | Gene(s) | Location relative to gene | Risk allele/reference allele | Risk-allele frequency (%) | Phenotype | Analysis | OR | 95% CI | P value | log ₁₀ (BF) |
|--------------------------------------|------------|-----------------------|---------------------------|------------------------------|---------------------------|-----------|----------|------|-----------|--------------------------|------------------------|
| Novel associations | | | | | | | | | | | |
| rs880315 | 1p36 | CASZ1 | Intronic | C/T | 40 | AS | TRANS | 1.05 | 1.04–1.07 | 3.62 × 10 ⁻¹⁰ | 8.09 |
| rs12037987 | 1p13 | WNT2B | Intronic | C/T | 16 | AS | TRANS | 1.07 | 1.05–1.10 | 2.73 × 10 ⁻⁸ | 6.33 |
| rs146390073 | 1q43 | RGS7 | Intronic | T/C | 2 | CES | EUR | 1.95 | 1.54–2.47 | 2.20 × 10 ⁻⁸ | NA ^a |
| rs12476527 | 2p23 | KCNK3 | 5'-UTR | G/T | 48 | AS | TRANS | 1.05 | 1.03–1.07 | 6.44 × 10 ⁻⁸ | 6.47 |
| rs7610618 | 3q25 | <i>TM4SF4-TM4SF1</i> | Intergenic | T/C | 1 | LAS | EUR | 2.33 | 1.74–3.12 | 1.44 × 10 ⁻⁸ | NA ^b |
| rs34311906 | 4q25 | <i>ANK2</i> | Intergenic | C/T | 41 | AIS | EUR | 1.07 | 1.04–1.09 | 1.07 × 10 ⁻⁸ | 5.67 |
| rs17612742 | 4q31 | EDNRA | Intronic | C/T | 21 | LAS | TRANS | 1.19 | 1.13–1.26 | 1.46 × 10 ⁻¹¹ | 9.47 |
| rs6825454 | 4q31 | <i>FGA</i> | Intergenic | C/T | 31 | AIS | TRANS | 1.06 | 1.04–1.08 | 7.43 × 10 ⁻¹⁰ | 7.53 |
| rs11957829 | 5q23 | LOC100505841 | Intronic | A/G | 82 | AIS | TRANS | 1.07 | 1.05–1.10 | 7.51 × 10 ⁻⁹ | 6.67 |
| rs6891174 | 5q35 | <i>NKX2-5</i> | Intergenic | A/G | 35 | CES | TRANS | 1.11 | 1.07–1.16 | 5.82 × 10 ⁻⁹ | 6.96 |
| rs16896398 | 6p21 | <i>SLC22A7-ZNF318</i> | Intergenic | T/A | 34 | AS | TRANS | 1.05 | 1.03–1.07 | 1.30 × 10 ⁻⁸ | 6.60 |
| rs42039 | 7q21 | CDK6 | 3'-UTR | C/T | 77 | AIS | TRANS | 1.07 | 1.04–1.09 | 6.55 × 10 ⁻⁹ | 6.84 |
| rs7859727 | 9p21 | Chr9p21 | ncRNA intronic | T/C | 53 | AS | TRANS | 1.05 | 1.03–1.07 | 4.22 × 10 ⁻¹⁰ | 8.01 |
| rs10820405 | 9q31 | LINC01492 | ncRNA intronic | G/A | 82 | LAS | EUR | 1.20 | 1.12–1.28 | 4.51 × 10 ⁻⁸ | 4.74 |
| rs2295786 | 10q24 | <i>SH3PXD2A</i> | Intergenic | A/T | 60 | AS | TRANS | 1.05 | 1.04–1.07 | 1.80 × 10 ⁻¹⁰ | 8.34 |
| rs7304841 | 12p12 | PDE3A | Intronic | A/C | 59 | AIS | TRANS | 1.05 | 1.03–1.07 | 4.93 × 10 ⁻⁸ | 5.87 |
| rs35436 | 12q24 | <i>TBX3</i> | Intergenic | C/T | 62 | AS | TRANS | 1.05 | 1.03–1.06 | 2.87 × 10 ⁻⁸ | 6.29 |
| rs9526212 | 13q14 | LRCH1 | Intronic | G/A | 76 | AS | TRANS | 1.06 | 1.04–1.08 | 5.03 × 10 ⁻¹⁰ | 7.97 |
| rs4932370 | 15q26 | <i>FURIN-FES</i> | Intergenic | A/G | 33 | AIS | TRANS | 1.05 | 1.03–1.07 | 2.88 × 10 ⁻⁸ | 6.05 |
| rs11867415 | 17p13 | PRPF8 | Intronic | G/A | 18 | AIS | TRANS | 1.09 | 1.06–1.13 | 4.81 × 10 ⁻⁸ | 6.06 |
| rs2229383 | 19p13 | ILF3-SLC44A2 | Exonic; synonymous | T/G | 65 | AIS | TRANS | 1.05 | 1.03–1.07 | 4.72 × 10 ⁻⁸ | 6.02 |
| rs8103309 | 19p13 | <i>SMARCA4-LDLR</i> | Intergenic | T/C | 65 | AS | TRANS | 1.05 | 1.03–1.07 | 3.40 × 10 ⁻⁸ | 5.85 |
| Previously known associations | | | | | | | | | | | |
| rs12124533 | 1p13 | <i>TSPAN2</i> | Intergenic | T/C | 24 | LAS | TRANS | 1.17 | 1.11–1.23 | 1.22 × 10 ⁻⁸ | 6.60 |
| rs1052053 | 1q22 | PMF1-SEMA4A | Exonic; nonsynonymous | G/A | 40 | AS | TRANS | 1.06 | 1.05–1.08 | 2.70 × 10 ⁻¹⁴ | 11.92 |
| rs13143308 | 4q25 | <i>PITX2</i> | Intergenic | T/G | 28 | CES | TRANS | 1.32 | 1.27–1.37 | 1.86 × 10 ⁻⁴⁷ | 45.10 |
| rs4959130 | 6p25 | <i>FOXF2</i> | Intergenic | A/G | 14 | AS | TRANS | 1.08 | 1.05–1.11 | 1.42 × 10 ⁻⁹ | 7.52 |
| rs2107595 | 7p21 | <i>HDAC9-TWIST1</i> | Intergenic | A/G | 24 | LAS | TRANS | 1.21 | 1.15–1.26 | 3.65 × 10 ⁻¹⁵ | 12.99 |
| rs635634 | 9q34 | <i>ABO</i> | Intergenic | T/C | 19 | AIS | EUR | 1.08 | 1.05–1.11 | 9.18 × 10 ⁻⁹ | 4.99 |
| rs2005108 | 11q22 | <i>MMP12</i> | Intergenic | T/C | 12 | AIS | TRANS | 1.08 | 1.05–1.11 | 3.33 × 10 ⁻⁸ | 6.12 |
| rs3184504 | 12q24 | SH2B3 | Exonic; nonsynonymous | T/C | 45 | AIS | TRANS | 1.08 | 1.06–1.10 | 2.17 × 10 ⁻¹⁴ | 12.04 |
| rs12932445 | 16q22 | ZFHX3 | Intronic | C/T | 21 | CES | TRANS | 1.20 | 1.15–1.25 | 6.86 × 10 ⁻¹⁸ | 15.49 |
| rs12445022 | 16q24 | <i>ZCCHC14</i> | Intergenic | A/G | 31 | AS | TRANS | 1.06 | 1.04–1.08 | 1.05 × 10 ⁻¹⁰ | 8.57 |

For each locus, the variant reaching the highest BF in the MANTRA or the lowest *P* value in the fixed-effects transancestral meta-analysis or the fixed-effects Europeans-only meta-analysis, respectively, is shown, and the respective stroke phenotype showing the strongest association is specified. Gene names in bold indicate that the variant is located within the gene; in other cases, the first gene corresponds to the closest gene, whereas additional gene names indicate eQTL signals from multiple studies, or from both eQTLs and meQTLs, or genes previously suspected to be causal (*LDLR*), with a maximum of two genes reported. The lead SNPs in *ILF3-SLC44A2* and *SMARCA4-LDLR* are in low LD ($r^2 = 0.082$). TRANS, MANTRA transancestral meta-analysis; EUR, Europeans-only fixed-effects meta-analysis; OR, odds ratio; CI, confidence interval; NA, not assessed; UTR, untranslated region. ^a rs146390073 did not meet the MAF threshold of 0.01 in samples other than those of European ancestry. ^b rs7610618: The transancestral meta-analysis results showed high heterogeneity ($PP_{het} = 0.96$) and were thus excluded.

with one gene in stroke-relevant tissues, such as *ZNF318* (6p21), *AL049919* (12q24), and *FES* (15q26) in brain tissues (Supplementary Tables 22 and 23).

At some loci, meQTLs and eQTLs provided complementary information on the regulatory pattern. For instance, for the *SH3PXD2A* locus, SNPs in high LD with the lead stroke risk variant were found to be eQTLs for multiple genes (*SH3PXD2A*, *SLK*, *GSTO1*, *GSTO2*, and *LOC729081*), whereas several high-LD

proxies ($r^2 > 0.96$) functioned as the most significant meQTL for CpG probes located in the promoter region of *SH3PXD2A* and not any of the other genes.

For the 149 genes located in the 32 genome-wide-significant loci ($r^2 > 0.5$ in Europeans or East Asians, or located ± 50 kb from the lead SNP; Methods), we assigned an empirical functional score based on the presence and number of eQTLs, meQTLs, pQTLs, and other biological criteria^{57,58} (Methods and Supplementary

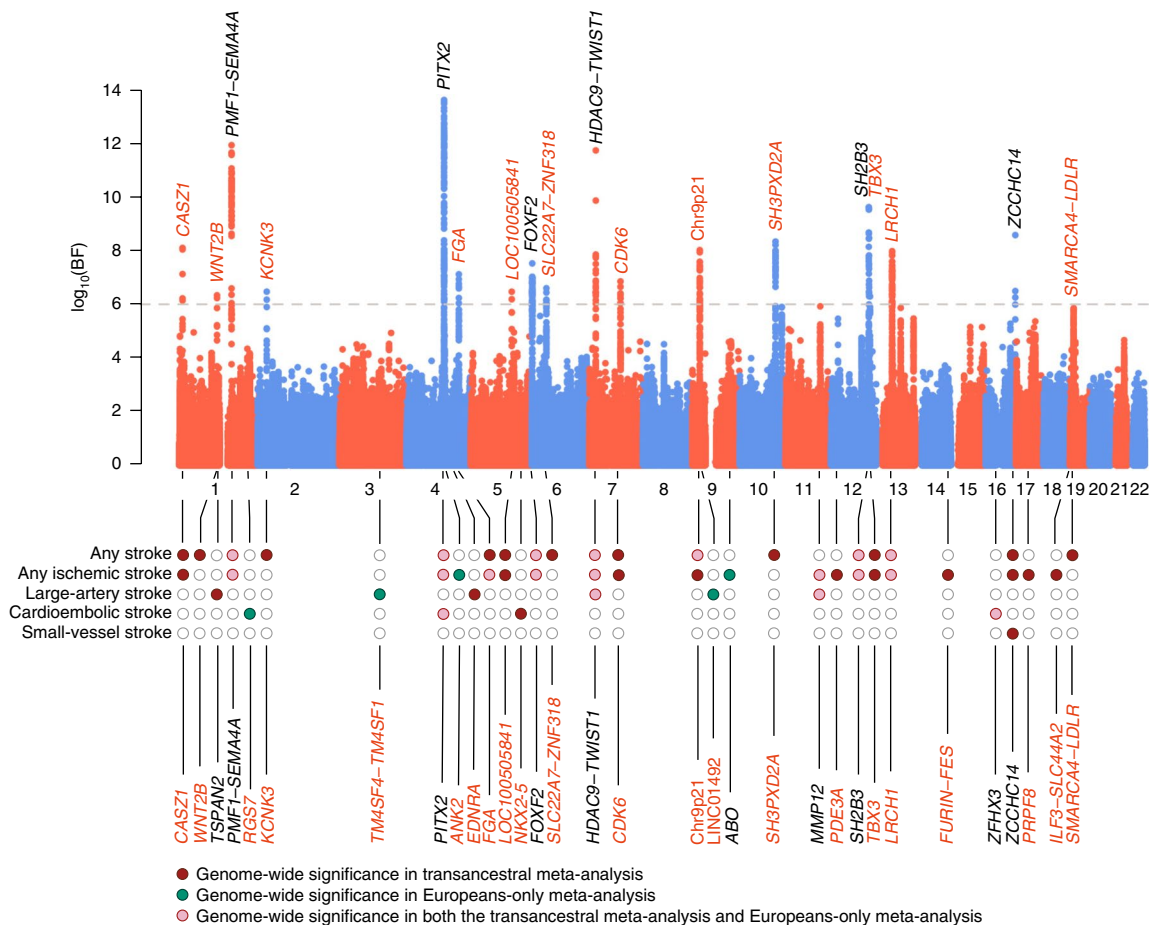


Fig. 2 | Association results of the transancestral GWAS meta-analysis and the prespecified ancestry-specific meta-analysis in European samples.

Shown are novel (red) and known (black) genetic loci associated with any stroke or stroke subtypes. Top, Manhattan plot from the MANTRA transancestral GWAS meta-analysis for any stroke. The dotted line marks the threshold of statistical significance ($\log_{10}(\text{BF}) > 6.0$).

Table 24), reasoning that genes with a higher functional score would be more likely to be causal, although this score requires validation by experimental data.

Joint modeling of epigenetic marks and association statistics.

In an additional approach to identify the most plausible causal variants and genes, we used RiVIERA⁵⁹, which jointly models summary association statistics and corresponding epigenetic regulatory information in a Bayesian framework to estimate the posterior probability of association (PPA). RiVIERA uses the RoadMap epigenome data of 127 tissue types and information on chromatin (H3K4me1, H3K4me3, H3K36me3, H3K27me3, H3K9me3, H3K27ac, and H3K9ac) and DNA-accessibility (DNase I) marks. Three of the stroke risk loci (*PMF1-SEMA4A*, *SH3PXD2A*, and *EDNRA*) displayed a pattern in which the association statistics and epigenetic regulatory information jointly contributed to the modeling of the RiVIERA credible SNP set (the minimum number of SNPs whose PPA, accounting for both association statistics and epigenetic regulatory information, sum to $\geq 95\%$) (Supplementary Fig. 13). The variants identified by RiVIERA as having the highest PPA were in moderate to high LD in the 1000G cosmopolitan panel with the respective lead SNP (rs7534434 for *PMF1-SEMA4A*, $r^2 = 0.79$ with lead SNP; rs11191829 for *SH3PXD2A*, $r^2 = 0.99$ with lead SNP; rs4835084 for *EDNRA*, $r^2 = 0.35$ with lead SNP). Two of these (at *PMF1-SEMA4A* and *SH3PXD2A*) were significantly enriched in RNA polymerase II binding in ENCODE cell types⁶⁰, including H1 human embryonic stem cells (Supplementary Fig. 13).

Enrichment in drug-target genes. Given the previous evidence of the utility of GWAS in drug discovery and drug repositioning^{57,61,62}, we evaluated the overlap between stroke-associated genes and known drug targets. Among the 149 genes located within the 32 stroke risk loci, 16 (11%) were registered as targets of currently approved drugs in the DrugBank database and the Therapeutic Target Database (Supplementary Table 25). Of these, two genes (*FGA* and *PDE3A*) were targets of approved drugs for antithrombotic therapy (ATC B01), i.e., alteplase, tenecteplase, reteplase, and anistreplase for *FGA*, and cilostazol for *PDE3A* (enrichment $\text{OR} = 5.46$, $P = 0.0369$; Fig. 5). This enrichment was strengthened after removal of the locus with the largest number of genes (*SH2B3*, 73 genes) ($\text{OR} = 8.89$, $P = 0.0166$) and after addition of 65 genes in 21 suggestive stroke risk loci ($\text{OR} = 7.83$, $P = 0.00606$).

Discussion

The current transancestral meta-analysis more than triples the number of stroke risk loci and identifies novel loci for AS, AIS, and all major subtypes of ischemic stroke. Our results highlight several major features of stroke genomics: (i) Approximately half of the identified stroke loci showed shared genetic association with other vascular traits, and the largest genetic correlation was found for blood pressure. We also identified shared genetic association with VTE, and distinct patterns of individual stroke subtypes provided further mechanistic insight. (ii) Eleven of the novel stroke risk loci (*ANK2*, *CDK6*, *KCNK3*, *LINC01492*, *LRCH1*, *NKX2-5*, *PDE3A*, *PRPF8*, *RGS7*, *TM4SF4-TM4SF1*, and *WNT2B*) suggest mechanisms not previously implicated in stroke pathophysiology; some

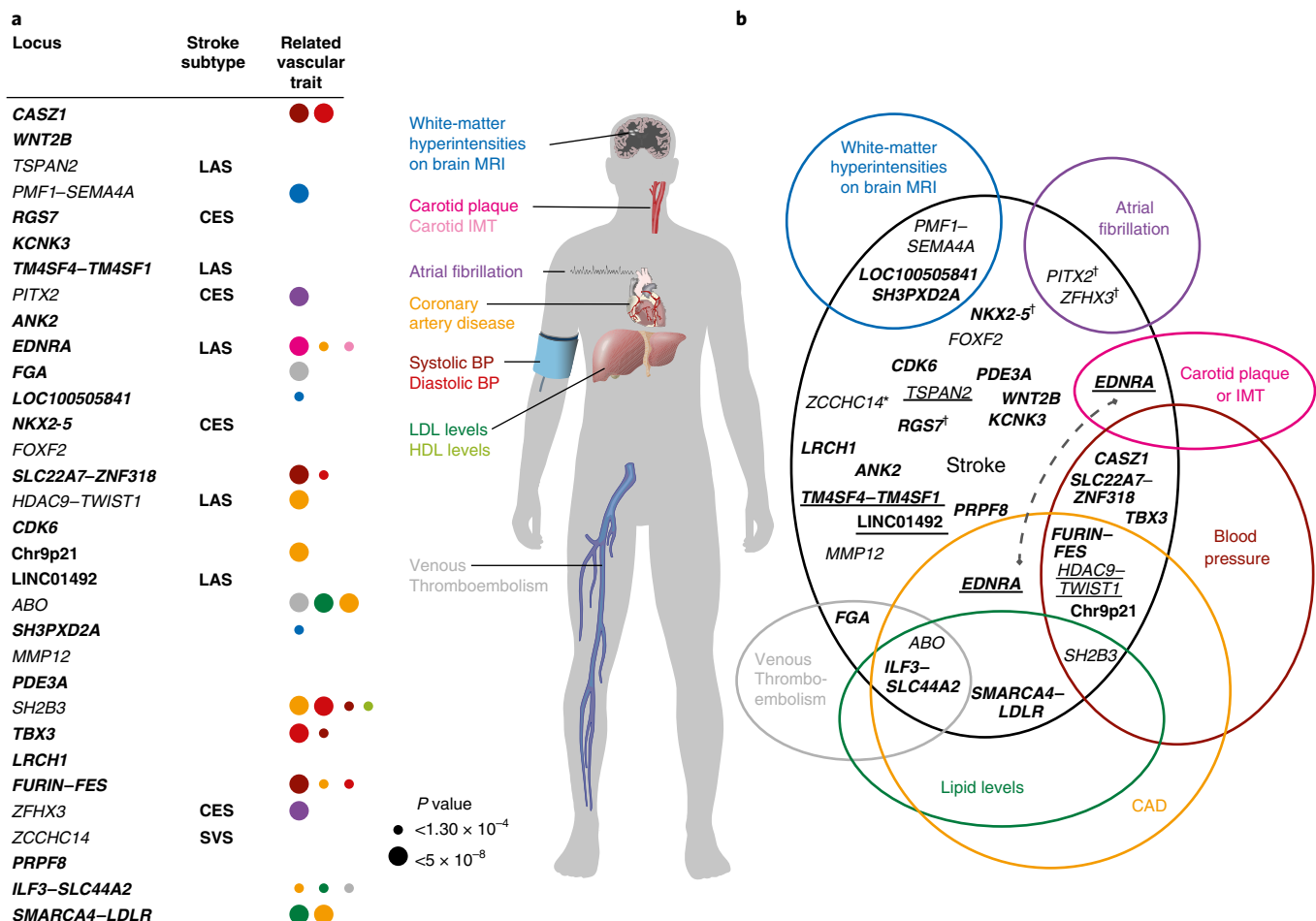


Fig. 3 | Genetic overlap between stroke and related vascular traits at the 32 genome-wide-significant loci for stroke. a, Association results from the look-ups in published GWAS data for related vascular traits. Symbol sizes reflect *P* values for association with the related trait. **b**, Venn diagram. Loci reaching genome-wide significance for association with stroke subtypes are marked with a dagger symbol (for CES), underlined (for LAS), or marked with an asterisk (for SVS). Novel loci are in bold. *SH3PXD2A*, *WNT2B*, *PDE3A*, and *OBFC1* have previously been associated with AF (*SH3PXD2A*)⁶⁵, or diastolic (*WNT2B* and *PDE3A*)^{24,72} or systolic (*OBFC1*)⁷³ BP, but the respective lead SNPs were in low LD ($r^2 < 0.1$ in the 1000G cosmopolitan panel) with variants associated with stroke in the current GWAS. MRI, magnetic resonance imaging; IMT, intima-media thickness; LDL, low-density lipoprotein; HDL, high-density lipoprotein. The lead variant for *TBX3* is not included in the original datasets for BP traits (SBP and DBP). Results are based on a perfect proxy SNP (rs35432, $r^2 = 1$ in the European 1000G phase 3 reference).

of these suggest a strong link with cardiac mechanisms beyond those expected from established sources of cardioembolism. (iii) The 32 stroke risk loci were significantly enriched in drug targets for antithrombotic therapy—one for an approved thrombolytic drug (alteplase) and the other for an antiplatelet agent (cilostazol) approved for stroke prevention in Asia. (iv) Through incorporation of extensive functional datasets and bioinformatics analyses, we provide detailed information on prioritization of stroke risk variants and genes as a resource for further experimental follow-up.

Most of the genome-wide associations were identified with both AS and AIS. Although this result relates in part to a greater statistical power compared with that in subtype analysis, we also found shared genetic influences between stroke subtypes, as exemplified by the gwas-pw analyses (*SH2B3* and *ABO*). A notable finding was the identification of *PMF1-SEMA4A* as a risk locus for AIS. *PMF1-SEMA4A* is an established risk locus for nonlobar ICH⁶ and thus is, to our knowledge, the first reported locus reaching genome-wide significance for ischemic as well as hemorrhagic stroke. *PMF1-SEMA4A* further reached genome-wide association for WMH burden²⁸ (Fig. 3), an established marker for SVD, and showed a strong signal in the SVS subtype, thus suggesting that the association with

stroke is at least in part mediated by SVD. The underlying biological pathways do not seem to involve known vascular risk factors and may thus identify new targets for stroke prevention.

Among the novel loci showing associations restricted to specific stroke subtypes, *EDNRA* is consistent with atherosclerotic mechanisms, given its association with LAS, cPI²⁷, and CAD³¹ (Fig. 3). *LINC01492* and the previously reported *TSPAN2* locus likewise displayed associations restricted to LAS but showed no association with related phenotypes in our look-ups and in prior literature, thus evidencing mechanisms more specific for LAS. *NKX2-5*, showing association restricted to CES, has previously been reported as a genome-wide risk locus for heart rate and PR interval^{63,64} but not consistently for AF^{63,65}, thus implicating cardiac mechanisms other than AF.

Although the number of loci reaching genome-wide significance for association with SVS remained low, our results suggest an important role of common genetic variation in SVS. First, several of the associations with AS or AIS, including those at novel loci (*CASZ1*, *LOC100505841*, *SH3PXD2A*, and *ICA1L-WDR12*), showed predominant association with the SVS subtype (Supplementary Tables 7 and 9). Second, three of the top loci (*PMF1-SEMA4A*, *LOC100505841*, and *SH3PXD2A*) showed genetic overlap with

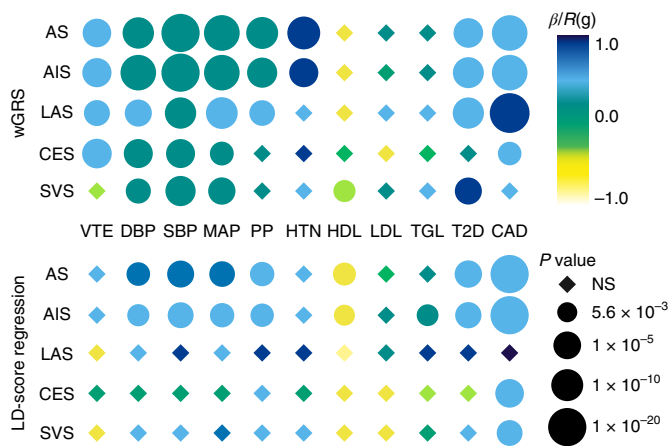


Fig. 4 | Shared genetic contribution between stroke and related vascular traits. Contributions determined by weighted genetic risk scores (wGRS, top) and LD-score regression analysis (bottom). Effect sizes and significance levels are represented by color and symbol size. β , wGRS effect size; $R(g)$, genetic correlation. DBP, diastolic blood pressure; SBP, systolic blood pressure; MAP, mean arterial pressure; PP, pulse pressure; HTN, hypertension; TGL, triglyceride level. Sample sizes for related vascular traits are displayed in Supplementary Table 12. NS, nonsignificant.

loci for WMH. Third, several suggestive loci ($\log_{10}(\text{BF}) \geq 5$) for AS and SVS contained genes implicated in monogenic SVD (*HTRA1*, *COL4A1*, and *COL4A2*) (Supplementary Table 8).

Our extensive exploration of shared genetic variation between stroke and related vascular traits found the most widespread correlations with BP phenotypes, in agreement with epidemiological data showing that high BP is the leading risk factor for stroke. A quarter of the 32 genome-wide-significant stroke loci were BP loci, most of which were novel with respect to stroke risk and showed association with risk of AS or AIS. Aside from the expected genetic overlap between LAS and CAD, we identified significant overlap between a wGRS for VTE and both LAS, and CES, but not SVS (Fig. 4 and Supplementary Table 14) despite a greater statistical power for this subtype, thus potentially suggesting that thrombotic processes play a less important role in SVS.

Three of our novel loci (*NKX2-5*, *ANK2*, and *LRCH1*) have previously been associated with cardiac pacing^{63,64,66}. *NKX2-5* and *ANK2* have been further implicated in familial forms of cardiac disease^{67–70}, but none of the three loci were associated with AF or CAD in the latest published GWAS^{31,65}. Apart from *NKX2-5*, these loci were not specifically associated with CES, thus possibly indicating an involvement of the underlying genes in roles beyond cardiac development and function. rs9526212, the lead variant in *LRCH1*, was an eQTL for *LRCH1* in multiple tissues, including the left ventricle, atherosclerotic aorta, atherosclerotic-lesion-free arteries, and blood (Supplementary Table 22). Pathway analyses further supported a strong link with cardiac mechanisms.

The extensive in silico functional annotation of identified stroke risk loci provides informative elements for future prioritization and follow-up of the most compelling biological candidates. In some instances, the eQTL, meQTL, and pQTL information strongly supports involvement of one gene over others in the region, for example, for *SH3PXD2A*, encoding SH3 and PX-domain-containing protein 2A, an adaptor protein involved in formation of invadopodia and podosomes as well as extracellular-matrix degradation. For some loci, joint analysis of epigenetic regulatory effects and association statistics enabled prioritization of credible SNPs. When exploring the overall epigenetic patterns of identified stroke risk loci, we observed some enrichment in enhancer and promoter sites in developmental tissues, thus suggesting that some associations

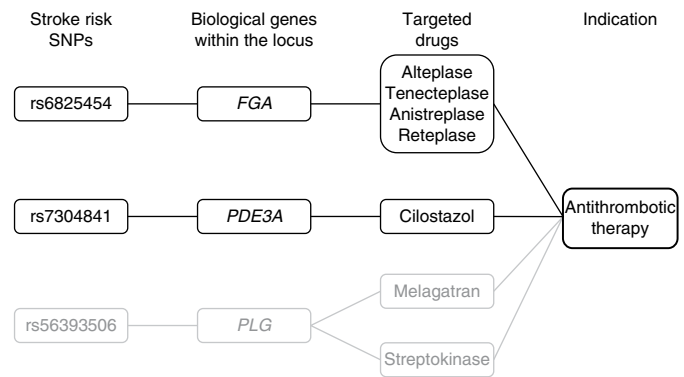


Fig. 5 | Connection between stroke risk genes and approved drugs for antithrombotic therapy. Shown are the connections among lead SNPs at stroke risk loci, biological stroke risk genes, and individual targeted drugs. Lead SNPs reaching suggestive evidence for association (MANTRA transancestral meta-analysis $\log_{10}(\text{BF}) > 5$) are shown in gray.

may be driven by developmental effects, as has recently been proposed for the *FOXF2* locus¹⁰.

RGS7 and *TM4SF4–TM4SF1* showed low MAFs, high heterogeneity, poor imputation quality in non-Europeans, and large effect-size estimates, and they must therefore be interpreted with caution. Moreover, although our extensive functional exploration provides guidance on gene prioritization for further exploration, additional experiments are required to identify the causal genes and variants. Several studies have provided limited information on stroke subtypes. Hence, the sample sizes for ischemic stroke subtypes were still relatively small. In addition, the proportion of the phenotypic variance explained by the 32 lead SNPs was relatively small but comparable to that in other complex diseases⁷¹. Collectively, these aspects highlight the potential for gene discovery in the future.

In conclusion, we identified 22 novel stroke risk loci and demonstrated shared genetic variation with multiple related vascular traits. We further identified new loci offering mechanisms not previously implicated in stroke pathophysiology and provided a framework for prioritization of stroke risk variants and genes for further functional and experimental follow-up. Stroke risk loci were significantly enriched in drug targets for antithrombotic therapy, thus highlighting the potential of stroke genetics for drug discovery. Collectively, these findings represent a major advance in understanding the genetic underpinnings of stroke.

URLs. Ingenuity Pathway Analysis, <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41588-018-0058-3>.

Received: 19 May 2017; Accepted: 8 January 2018;
Published online: 12 March 2018

References

- GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1603–1658 (2016).
- GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1459–1544 (2016).
- Gudbjartsson, D. F. et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353–357 (2007).

4. Gudbjartsson, D. F. et al. A sequence variant in *ZFHX3* on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat. Genet.* **41**, 876–878 (2009).
5. International Stroke Genetics Consortium (ISGC) et al. Genome-wide association study identifies a variant in *HDAC9* associated with large vessel ischemic stroke. *Nat. Genet.* **44**, 328–333 (2012).
6. Woo, D. et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am. J. Hum. Genet.* **94**, 511–521 (2014).
7. Kilarski, L. L. et al. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology* **83**, 678–685 (2014).
8. Traylor, M. et al. A novel *MMP12* locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet.* **10**, e1004469 (2014).
9. NINDS, Stroke Genetics Network (SiGN) & International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol.* **15**, 174–184 (2016).
10. Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Stroke Genetics Network (SiGN) & the International Stroke Genetics Consortium (ISGC). Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **15**, 695–707 (2016).
11. Malik, R. et al. Low-frequency and common genetic variation in ischemic stroke: the METASTROKE collaboration. *Neurology* **86**, 1217–1226 (2016).
12. Traylor, M. et al. Genetic variation at 16q24.2 is associated with small vessel stroke. *Ann. Neurol.* **81**, 383–394 (2017).
13. Williams, F. M. et al. Ischemic stroke is associated with the *ABO* locus: the EuroCLOT study. *Ann. Neurol.* **73**, 16–31 (2013).
14. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
15. Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
16. Mishra, A. & Macgregor, S. VEGAS2: software for more flexible gene-based testing. *Twin Res. Hum. Genet.* **18**, 86–91 (2015).
17. Traylor, M. et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology* **86**, 146–153 (2016).
18. Hara, K. et al. Association of *HTRA1* mutations and familial ischemic cerebral small-vessel disease. *N. Engl. J. Med.* **360**, 1729–1739 (2009).
19. Verdura, E. et al. Heterozygous *HTRA1* mutations are associated with autosomal dominant cerebral small vessel disease. *Brain* **138**, 2347–2358 (2015).
20. Gould, D. B. et al. Role of *COL4A1* in small-vessel disease and hemorrhagic stroke. *N. Engl. J. Med.* **354**, 1489–1496 (2006).
21. Jeanne, M. et al. *COL4A2* mutations impair *COL4A1* and *COL4A2* secretion and cause hemorrhagic stroke. *Am. J. Hum. Genet.* **90**, 91–101 (2012).
22. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
23. Lubitz, S. A. et al. Independent susceptibility markers for atrial fibrillation on chromosome 4q25. *Circulation* **122**, 976–984 (2010).
24. Kato, N. et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat. Genet.* **47**, 1282–1293 (2015).
25. Surakka, I. et al. The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47**, 589–597 (2015).
26. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
27. Bis, J. C. et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat. Genet.* **43**, 940–947 (2011).
28. Verhaaren, B. F. et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ. Cardiovasc. Genet.* **8**, 398–409 (2015).
29. Sinner, M. F. et al. Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation* **130**, 1225–1235 (2014).
30. Germain, M. et al. Meta-analysis of 65,734 individuals identifies *TSPAN15* and *SLC44A2* as two susceptibility loci for venous thromboembolism. *Am. J. Hum. Genet.* **96**, 532–542 (2015).
31. Nikpay, M. et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
32. Ellinor, P. T. et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat. Genet.* **44**, 670–675 (2012).
33. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
34. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
35. Trynka, G. et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
36. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
37. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
38. Geer, L. Y. et al. The NCBI BioSystems database. *Nucleic Acids Res.* **38**, D492–D496 (2010).
39. Mishra, A. & MacGregor, S. A novel approach for pathway analysis of GWAS data highlights role of BMP signaling and muscle cell differentiation in colorectal cancer susceptibility. *Twin Res. Hum. Genet.* **20**, 1–9 (2017).
40. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
41. Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* **10**, 1556–1566 (2015).
42. Wang, W. et al. LNK/SH2B3 loss of function promotes atherosclerosis and thrombosis. *Circ. Res.* **119**, e91–e103 (2016).
43. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30**, 3894–3900 (2002).
44. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
45. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
46. Eicher, J. D. et al. GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.* **43**, D799–D804 (2015).
47. Leslie, R., O'Donnell, C. J. & Johnson, A. D. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* **30**, i185–i194 (2014).
48. Higasa, K. et al. Human genetic variation database, a reference database of genetic variations in the Japanese population. *J. Hum. Genet.* **61**, 547–553 (2016).
49. Bonder, M. J. et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* **49**, 131–138 (2017).
50. Adams, D. et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30**, 224–226 (2012).
51. Franzen, O. et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
52. Erbilgin, A. et al. Identification of CAD candidate genes in GWAS loci and their expression in vascular cells. *J. Lipid Res.* **54**, 1894–1905 (2013).
53. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am. J. Epidemiol.* **129**, 687–702 (1989).
54. Suhre, K. et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
55. Brønne, I. et al. Prediction of causal candidate genes in coronary artery disease loci. *Arterioscler. Thromb. Vasc. Biol.* **35**, 2207–2217 (2015).
56. Flister, M. J. et al. Identifying multiple causative genes at a single GWAS locus. *Genome Res.* **23**, 1996–2002 (2013).
57. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
58. Kemp, J. P. et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat. Genet.* **49**, 1468–1475 (2017).
59. Li, Y. & Kellis, M. Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. *Nucleic Acids Res.* **44**, e144 (2016).
60. Lee, B. K. et al. Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Res.* **22**, 9–24 (2012).
61. Sanseau, P. et al. Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* **30**, 317–320 (2012).
62. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
63. den Hoed, M. et al. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat. Genet.* **45**, 621–631 (2013).
64. Pfeufer, A. et al. Genome-wide association study of PR interval. *Nat. Genet.* **42**, 153–159 (2010).
65. Christophersen, I. E. et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* **49**, 946–952 (2017).
66. Verweij, N. et al. Genetic determinants of P wave duration and PR segment. *Circ. Cardiovasc. Genet.* **7**, 475–481 (2014).
67. Le Scouarnec, S. et al. Dysfunction in ankyrin-B-dependent ion channel and transporter targeting causes human sinus node disease. *Proc. Natl. Acad. Sci. USA* **105**, 15617–15622 (2008).
68. Schott, J. J. et al. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108–111 (1998).

69. Ellesøe, S. G. et al. Familial atrial septal defect and sudden cardiac death: identification of a novel NKX2-5 mutation and a review of the literature. *Congenit. Heart Dis.* **11**, 283–290 (2016).
70. Mohler, P. J. et al. Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* **421**, 634–639 (2003).
71. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).
72. Kato, N. et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat. Genet.* **43**, 531–538 (2011).
73. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).

Acknowledgements

A full list of Acknowledgements appears in the Supplementary Note.

Author contributions

Writing and editing the manuscript: R.M., G.C., M.T., M.S., Y.O., S.D., and M.D. Study design/conception: R.M., M.D., S.D., B.M.P., G.J.F., J.W.J., J.I.R., J.G.W., M.F., H.I.Y., C.J., S. Seshadri, W.T.L., B.B.W., B.D.M., S.J.K., H.S.M., J.D., J.R., K.S., and O.M. Statistical analysis: A.-K.G., G.J.F., M.F., C.D.L., Y.O., E.L., B.R.S., R.M., M.S., M.T., A. Mishra, E.G.H., C.D.A., T.M.B., C. Carrera, I.C., W.-Y.L., S.L.P., K. Rannikmäe, K. Rice, S. Tiedt, J.C.C., A.D.J., P.L.W.d.B., S.W.v.d.L., P. Almgren, S. Gretarsdottir, and F.T. Sample/phenotype contribution: M.D., S.D., C.D.A., C. Cruchaga, I.C., H.I.H., J.W.J., N.S.R., A.S.B., A.C., A.S., A.S.H., A.P.R., A.L.D., A. Rolfs, A. Ruusalepp, A.G.L., A. Manichaikul, B.M.K., C.L.C., C.R., C.K., C. Tanislav, C. Tzourio, C.M.v.D., D.I.C., D.W., D.A.T., D.O.K., D.K.S., D.L., E.S.T., E.E.S., E.I., F.-C.H., G.P., H.A., H.H.H.A., H.S.M., I.E.C., J. Haessler, J. He, J. Hata, J.F.M., J.S.K., J.-M.L., J.D., J.W.C., J.R., J.J.-C., J.A.J., K.S., K.M.R., K.L.K., K.L.W., L.J.L., L.A.L., M.A.N., M.A.I., M.d.H., M.R.I., M.J.O., M. Kanai, M. Kubo, M.W., M.M.S., N.J.W., N.K., O.R.B., P.F.M., P.T.E., P.K.M., P.E., P. Amouyel, P.v.d.H., Q.D., Q.Y., R.P.G., R.L.S., R.F.G., R.S., S.Y., S.K., S.T.E., S.B., S.A.L., S.J.K., S.R.H., S.W.-S., T.B.H., T.R., T.H.M., T.P., T.T., U.S., U.T., V.C., V.G., W.-M.C., V.N.S.T., X.J., B.M.P., J.I.R., J.G.W., O.M., C.J., J.C.H., S. Seshadri, T.A., G.B.B., R.D.B., A.H., N.L.S., R.L., C.M.L., T.N., P. M. Ridker, P. M. Rothwell, V.S., C.O.S., P.S., C.L.M.S., K.D.T., M. Civelek, D. Saleheen, D. Strbian, S. Sakaue, S. Gustafsson, S. Tiedt, S. Trompet, and I.E.-C. Critical revision of article: R.M., M.D., S.D., B.M.P., C.J., J.I.R., O.M., S. Seshadri, G.J.F., J.W.J., W.T.L., C.D.A., D. Strbian, E.G.H., I.E.-C., S. Tiedt,

C.L.M.S., C.O.S., C. Cruchaga, G.B.B., I.C., J.C.B., J. Hata, K. Rice, S.L.P., N.S.R., S.S.R., T.A., T.N., J.M.M.H., T.M.B., and V.S. Supervision: M.D., S.D., C.D.A., J.M.M.H., J.I.R., S. Seshadri, C.M.L., C.L.M.S., J.W.J., V.S., and J.C.B. GWAS analyses: R.M., G.C., M.T., S. Gretarsdottir, G.T., J. Hata, A.K.G., M. Chong, J.L.M.B., C. Carrera, A.H., G.J.F., and Y.K. Functional annotation: M.S., A. Mishra, R.M., G.C., M.T., L.R.-J., and A.K.G. Gene-based analysis: A. Mishra. Pathway analyses: A. Mishra, R.M., M. Chong, and K. Rice. Drug-target analysis: Y.O. Scoring method: M.S., R.M., S.D., and M.D. wGRS analysis: M.S. and R.M. LD-score regression analysis: R.M., M.S., and Y.K. Credible-SNP-set analysis: R.M., G.C., and M.S. Data for GWAS analysis, cross-phenotype analysis or QTL analysis: AFGen Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, iGen-BP Consortium, INVENT Consortium, STARNet, and Biobank Japan Cooperative Hospital Group. Consortia providing stroke data: COMPASS Consortium, EPIC-CVD Consortium, EPIC-InterAct Consortium, ISGC, METASTROKE Consortium, Neurology Working Group of the CHARGE Consortium, NINDS-SiGN, UK Young Lacunar DNA Study, and MEGASTROKE Consortium. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute or the National Institute of Neurological Disorders and Stroke.

Competing interests

S. Gretarsdottir, G.T., U.T., and K.S. are all employees of deCODE Genetics/Amgen, Inc. M.A.N. is an employee of Data Tecnica International. P.T.E. is the PI on a grant from Bayer HealthCare to the Broad Institute, focused on the genetics and therapeutics of atrial fibrillation. S.A.L. receives sponsored research support from Bayer HealthCare, Biotronik, and Boehringer Ingelheim, and has consulted for St. Jude Medical and Quest Diagnostics. E.I. is a scientific advisor for Precision Wellness, Cellink and Olink Proteomics for work unrelated to the present project. B.M.P. serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. The remaining authors have no disclosures.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41588-018-0058-3>.

Reprints and permissions information is available at www.nature.com/reprints.

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

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

Rainer Malik^{1,185}, Ganesh Chauhan^{2,3,185}, Matthew Traylor^{4,185}, Muralidharan Sargurupremraj^{3,5,185}, Yukinori Okada^{6,7,8,185}, Aniket Mishra^{3,5}, Loes Rutten-Jacobs⁴, Anne-Katrin Giese⁹, Sander W. van der Laan¹⁰, Solveig Gretarsdottir¹¹, Christopher D. Anderson^{12,13,14}, Michael Chong¹⁵, Hieab H. H. Adams^{16,17}, Tetsuro Ago¹⁸, Peter Almgren¹⁹, Philippe Amouyel^{20,21}, Hakan Ay^{13,22}, Traci M. Bartz²³, Oscar R. Benavente²⁴, Steve Bevan²⁵, Giorgio B. Boncoraglio²⁶, Robert D. Brown Jr²⁷, Adam S. Butterworth^{28,29}, Caty Carrera^{30,31}, Cara L. Carty^{32,33}, Daniel I. Chasman^{34,35}, Wei-Min Chen³⁶, John W. Cole³⁷, Adolfo Correa³⁸, Ioana Cotlarciuc³⁹, Carlos Cruchaga^{40,41}, John Danesh^{28,42,43,44}, Paul I. W. de Bakker^{45,46}, Anita L. DeStefano^{47,48}, Marcel den Hoed⁴⁹, Qing Duan⁵⁰, Stefan T. Engelter^{51,52}, Guido J. Falcone^{53,54}, Rebecca F. Gottesman⁵⁵, Raji P. Grewal⁵⁶, Vilmundur Gudnason^{57,58}, Stefan Gustafsson⁵⁹, Jeffrey Haessler⁶⁰, Tamara B. Harris⁶¹, Ahamad Hassan⁶², Aki S. Havulinna^{63,64}, Susan R. Heckbert⁶⁵, Elizabeth G. Holliday^{66,67}, George Howard⁶⁸, Fang-Chi Hsu⁶⁹, Hyacinth I. Hyacinth⁷⁰, M. Arfan Ikram¹⁶, Erik Ingelsson^{71,72}, Marguerite R. Irvin⁷³, Xueqiu Jian⁷⁴, Jordi Jiménez-Conde⁷⁵, Julie A. Johnson^{76,77}, J. Wouter Jukema⁷⁸, Masahiro Kanai^{6,7,79}, Keith L. Keene^{80,81}, Brett M. Kissela⁸², Dawn O. Kleindorfer⁸², Charles Kooperberg⁶⁰, Michiaki Kubo⁸³, Leslie A. Lange⁸⁴, Carl D. Langefeld⁸⁵, Claudia Langenberg⁸⁶, Lenore J. Launer⁸⁷, Jin-Moo Lee⁸⁸, Robin Lemmens^{89,90}, Didier Leys⁹¹, Cathryn M. Lewis^{92,93}, Wei-Yu Lin^{28,94}, Arne G. Lindgren^{95,96}, Erik Lorentzen⁹⁷, Patrik K. Magnusson⁹⁸, Jane Maguire⁹⁹, Ani Manichaikul³⁶, Patrick F. McArdle¹⁰⁰, James F. Meschia¹⁰¹, Braxton D. Mitchell^{100,102}, Thomas H. Mosley^{103,104}, Michael A. Nalls^{105,106},

Toshiharu Ninomiya¹⁰⁷, Martin J. O'Donnell^{15,108}, Bruce M. Psaty^{109,110,111,112}, Sara L. Pulit^{45,113}, Kristiina Rannikmäe^{114,115}, Alexander P. Reiner^{65,116}, Kathryn M. Rexrode¹¹⁷, Kenneth Rice¹¹⁸, Stephen S. Rich³⁶, Paul M. Ridker^{34,35}, Natalia S. Rost^{9,13}, Peter M. Rothwell¹¹⁹, Jerome I. Rotter^{120,121}, Tatjana Rundek¹²², Ralph L. Sacco¹²², Saori Sakaue^{7,123}, Michele M. Sale¹²⁴, Veikko Salomaa⁶³, Bishwa R. Sapkota¹²⁵, Reinhold Schmidt¹²⁶, Carsten O. Schmidt¹²⁷, Ulf Schminke¹²⁸, Pankaj Sharma³⁹, Agnieszka Slowik¹²⁹, Cathie L. M. Sudlow^{114,115}, Christian Tanislav¹³⁰, Turgut Tatlisumak^{131,132}, Kent D. Taylor^{120,121}, Vincent N. S. Thijs^{133,134}, Gudmar Thorleifsson¹¹, Unnur Thorsteinsdottir¹¹, Steffen Tiedt¹, Stella Trompet¹³⁵, Christophe Tzourio^{3,136,137}, Cornelia M. van Duijn^{138,139}, Matthew Walters¹⁴⁰, Nicholas J. Wareham⁸⁶, Sylvia Wassertheil-Smoller¹⁴¹, James G. Wilson¹⁴², Kerri L. Wiggins¹⁰⁹, Qiong Yang⁴⁷, Salim Yusuf¹⁵, AFGen Consortium¹⁴³, Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium¹⁴³, International Genomics of Blood Pressure (iGEN-BP) Consortium¹⁴³, INVENT Consortium¹⁴³, STARNET¹⁴³, Joshua C. Bis¹⁰⁹, Tomi Pastinen¹⁴⁴, Arno Ruusalepp^{145,146,147}, Eric E. Schadt¹⁴⁸, Simon Koplev¹⁴⁸, Johan L. M. Björkegren^{148,149,150,151}, Veronica Codoni^{152,153}, Mete Civelek^{124,154}, Nicholas L. Smith^{65,155,156}, David A. Trégouët^{152,153}, Ingrid E. Christophersen^{54,157,158}, Carolina Roselli⁵⁴, Steven A. Lubitz^{54,157}, Patrick T. Ellinor^{54,157}, E. Shyong Tai¹⁵⁹, Jaspal S. Kooner¹⁶⁰, Norihiro Kato¹⁶¹, Jiang He¹⁶², Pim van der Harst¹⁶³, Paul Elliott¹⁶⁴, John C. Chambers^{165,166}, Fumihiko Takeuchi¹⁶¹, Andrew D. Johnson^{48,167}, BioBank Japan Cooperative Hospital Group¹⁴³, COMPASS Consortium¹⁴³, EPIC-CVD Consortium¹⁴³, EPIC-InterAct Consortium¹⁴³, International Stroke Genetics Consortium (ISGC)¹⁴³, METASTROKE Consortium¹⁴³, Neurology Working Group of the CHARGE Consortium¹⁴³, NINDS Stroke Genetics Network (SiGN)¹⁴³, UK Young Lacunar DNA Study¹⁴³, MEGASTROKE Consortium¹⁶⁸, Dharambir K. Sanghera^{125,169,170}, Olle Melander¹⁹, Christina Jern¹⁷¹, Daniel Strbian^{172,173}, Israel Fernandez-Cadenas^{30,31}, W. T. Longstreth Jr^{65,174}, Arndt Rolfs¹⁷⁵, Jun Hata¹⁰⁷, Daniel Woo⁸², Jonathan Rosand^{12,13,14}, Guillaume Pare¹⁵, Jemma C. Hopewell¹⁷⁶, Danish Saleheen¹⁷⁷, Kari Stefansson^{11,178,186}, Bradford B. Worrall^{179,186}, Steven J. Kittner^{37,186}, Sudha Seshadri^{48,180,186}, Myriam Fornage^{74,181,186}, Hugh S. Markus^{4,186}, Joanna M. M. Howson^{28,186}, Yoichiro Kamatani^{6,182,186}, Stephanie Debette^{3,5,186*} and Martin Dichgans^{1,183,184,186*}

¹Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany. ²Centre for Brain Research, Indian Institute of Science, Bangalore, India. ³INSERM U1219 Bordeaux Population Health Research Center, University of Bordeaux, France. ⁴Stroke Research Group, Division of Clinical Neurosciences, University of Cambridge, Cambridge, UK. ⁵Department of Neurology, Institute for Neurodegenerative Disease, Bordeaux University Hospital, Bordeaux, France. ⁶Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. ⁷Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. ⁸Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. ⁹Department of Neurology, Massachusetts General Hospital (MGH), Harvard Medical School, Boston, MA, USA. ¹⁰Laboratory of Experimental Cardiology, Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands. ¹¹deCODE genetics/AMGEN Inc., Reykjavik, Iceland. ¹²Center for Genomic Medicine, MGH, Boston, MA, USA. ¹³J. Philip Kistler Stroke Research Center, Department of Neurology, MGH, Boston, MA, USA. ¹⁴Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. ¹⁵Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada. ¹⁶Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. ¹⁷Department of Radiology and Nuclear Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. ¹⁸Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. ¹⁹Department of Clinical Sciences, Lund University, Malmö, Sweden. ²⁰INSERM, Institut Pasteur de Lille, LabEx DISTALZ-UMR1167, Risk Factors and Molecular Determinants of Aging-Related Diseases, Université Lille, Lille, France. ²¹Centre Hospitalier Université Lille, Epidemiology and Public Health Department, Lille, France. ²²AA Martinos Center for Biomedical Imaging, Department of Radiology, MGH, Harvard Medical School, Boston, MA, USA. ²³Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of Washington, Seattle, WA, USA. ²⁴Division of Neurology, Faculty of Medicine, Brain Research Center, University of British Columbia, Vancouver, British Columbia, Canada. ²⁵School of Life Science, University of Lincoln, Lincoln, UK. ²⁶Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. ²⁷Department of Neurology, Mayo Clinic Rochester, Rochester, MN, USA. ²⁸MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ²⁹National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK. ³⁰Neurovascular Research Laboratory, Vall d'Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona, Vall d'Hebrón Hospital, Barcelona, Spain. ³¹Stroke Pharmacogenomics and Genetics, Fundació Docència

i Recerca MutuaTerrassa, Terrassa, Spain. ³²Children's Research Institute, Children's National Medical Center, Washington, DC, USA. ³³Center for Translational Science, George Washington University, Washington, DC, USA. ³⁴Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. ³⁵Harvard Medical School, Boston, MA, USA. ³⁶Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA. ³⁷Department of Neurology, University of Maryland School of Medicine and Baltimore VAMC, Baltimore, MD, USA. ³⁸Departments of Medicine, Pediatrics and Population Health Science, University of Mississippi Medical Center, Jackson, MS, USA. ³⁹Institute of Cardiovascular Research, Royal Holloway University of London, London, UK, and Ashford and St Peters Hospital, Surrey, UK. ⁴⁰Department of Psychiatry, Hope Center Program on Protein Aggregation and Neurodegeneration (HPAN), Washington University School of Medicine, St. Louis, MO, USA. ⁴¹Department of Developmental Biology, Washington University School of Medicine, St. Louis, MO, USA. ⁴²NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁴³Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. ⁴⁴British Heart Foundation, Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, UK. ⁴⁵Department of Medical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands. ⁴⁶Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands. ⁴⁷Boston University School of Public Health, Boston, MA, USA. ⁴⁸Framingham Heart Study, Framingham, MA, USA. ⁴⁹Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ⁵⁰Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. ⁵¹Department of Neurology and Stroke Center, Basel University Hospital, Basel, Switzerland. ⁵²Neurorehabilitation Unit, University of Basel and University Center for Medicine of Aging and Rehabilitation Basel, Felix Platter Hospital, Basel, Switzerland. ⁵³Department of Neurology, Yale University School of Medicine, New Haven, CT, USA. ⁵⁴Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. ⁵⁵Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁵⁶Neuroscience Institute, SF Medical Center, Trenton, NJ, USA. ⁵⁷Icelandic Heart Association Research Institute, Kopavogur, Iceland. ⁵⁸Faculty of Medicine, University of Iceland, Reykjavik, Iceland. ⁵⁹Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ⁶⁰Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ⁶¹Laboratory of Epidemiology and Population Science, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. ⁶²Department of Neurology, Leeds General Infirmary, Leeds Teaching Hospitals NHS Trust, Leeds, UK. ⁶³National Institute for Health and Welfare, Helsinki, Finland. ⁶⁴FIMM-Institute for Molecular Medicine Finland, Helsinki, Finland. ⁶⁵Department of Epidemiology, University of Washington, Seattle, WA, USA. ⁶⁶Public Health Stream, Hunter Medical Research Institute, New Lambton, New South Wales, Australia. ⁶⁷Faculty of Health and Medicine, University of Newcastle, Newcastle, New South Wales, Australia. ⁶⁸School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. ⁶⁹Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. ⁷⁰Aflac Cancer and Blood Disorder Center, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA. ⁷¹Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA. ⁷²Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ⁷³Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. ⁷⁴Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA. ⁷⁵Neurovascular Research Group (NEUVAS), Neurology Department, Institut Hospital del Mar d'Investigació Mèdica, Universitat Autònoma de Barcelona, Barcelona, Spain. ⁷⁶Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL, USA. ⁷⁷Division of Cardiovascular Medicine, College of Medicine, University of Florida, Gainesville, FL, USA. ⁷⁸Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands. ⁷⁹Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, USA. ⁸⁰Department of Biology, East Carolina University, Greenville, NC, USA. ⁸¹Center for Health Disparities, East Carolina University, Greenville, NC, USA. ⁸²University of Cincinnati College of Medicine, Cincinnati, OH, USA. ⁸³RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. ⁸⁴Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA. ⁸⁵Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. ⁸⁶MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge, UK. ⁸⁷Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. ⁸⁸Department of Neurology, Radiology, and Biomedical Engineering, Washington University School of Medicine, St. Louis, MO, USA. ⁸⁹Department of Neurosciences, Experimental Neurology, KU Leuven-University of Leuven, Leuven, Belgium. ⁹⁰VIB Center for Brain & Disease Research, University Hospitals Leuven, Department of Neurology, Leuven, Belgium. ⁹¹INSERM U 1171, CHU Lille, Université Lille, Lille, France. ⁹²Department of Medical and Molecular Genetics, King's College London, London, UK. ⁹³SGDP Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. ⁹⁴Northern Institute for Cancer Research, Newcastle University, Newcastle, UK. ⁹⁵Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden. ⁹⁶Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden. ⁹⁷Bioinformatics Core Facility, University of Gothenburg, Gothenburg, Sweden. ⁹⁸Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ⁹⁹University of Technology Sydney, Faculty of Health, Ultimo, New South Wales, Australia. ¹⁰⁰Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. ¹⁰¹Department of Neurology, Mayo Clinic, Jacksonville, FL, USA. ¹⁰²Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA. ¹⁰³Division of Geriatrics, School of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. ¹⁰⁴Memory Impairment and Neurodegenerative Dementia Center, University of Mississippi Medical Center, Jackson, MS, USA. ¹⁰⁵Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. ¹⁰⁶Data Tecnica International, Glen Echo, MD, USA. ¹⁰⁷Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. ¹⁰⁸Clinical Research Facility, Department of Medicine, NUI Galway, Galway, Ireland. ¹⁰⁹Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. ¹¹⁰Department of Epidemiology, University of Washington, Seattle, WA, USA. ¹¹¹Department of Health Services, University of Washington, Seattle, WA, USA. ¹¹²Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. ¹¹³Brain Center Rudolf Magnus, Department of Neurology, University Medical Center Utrecht, Utrecht, the Netherlands. ¹¹⁴Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ¹¹⁵Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. ¹¹⁶Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA, USA. ¹¹⁷Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. ¹¹⁸Department of Biostatistics, University of Washington, Seattle, WA, USA. ¹¹⁹Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. ¹²⁰Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA. ¹²¹Division of Genomic Outcomes, Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA. ¹²²Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA. ¹²³Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. ¹²⁴Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. ¹²⁵Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA. ¹²⁶Department of Neurology, Medical University of Graz, Graz, Austria. ¹²⁷Institute for Community Medicine, SHIP-KEF, University Medicine Greifswald, Greifswald, Germany. ¹²⁸Department of Neurology, University Medicine Greifswald, Greifswald, Germany. ¹²⁹Department of Neurology, Jagiellonian University, Krakow, Poland. ¹³⁰Department of Neurology, Justus Liebig University, Giessen, Germany. ¹³¹Department of Clinical Neurosciences/Neurology, Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden. ¹³²Sahlgrenska University Hospital, Gothenburg, Sweden. ¹³³Stroke Division, Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Victoria, Australia. ¹³⁴Austin Health, Department of Neurology, Heidelberg, Victoria, Australia. ¹³⁵Department of Internal Medicine,

Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. ¹³⁶INSERM, U1219 Bordeaux, France. ¹³⁷Department of Public Health, Bordeaux University Hospital, Bordeaux, France. ¹³⁸Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. ¹³⁹Center for Medical Systems Biology, Leiden, the Netherlands. ¹⁴⁰School of Medicine, Dentistry and Nursing at the University of Glasgow, Glasgow, UK. ¹⁴¹Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY, USA. ¹⁴²Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA. ¹⁴³A list of members and affiliations appears in the Supplementary Note. ¹⁴⁴Department of Human Genetics, McGill University, Montreal, Quebec, Canada. ¹⁴⁵Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, University of Tartu, Tartu, Estonia. ¹⁴⁶Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia. ¹⁴⁷Clinical Gene Networks AB, Stockholm, Sweden. ¹⁴⁸Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹⁴⁹Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, Biomeedikum, University of Tartu, Tartu, Estonia. ¹⁵⁰Integrated Cardio Metabolic Centre, Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge, Sweden. ¹⁵¹Clinical Gene Networks AB, Stockholm, Sweden. ¹⁵²UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Sorbonne Universités, Paris, France. ¹⁵³CAN Institute for Cardiometabolism and Nutrition, Paris, France. ¹⁵⁴Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA. ¹⁵⁵Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA. ¹⁵⁶Seattle Epidemiologic Research and Information Center, VA Office of Research and Development, Seattle, WA, USA. ¹⁵⁷Cardiovascular Research Center, MGH, Boston, MA, USA. ¹⁵⁸Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust, Gjøttum, Norway. ¹⁵⁹Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. ¹⁶⁰National Heart and Lung Institute, Imperial College London, London, UK. ¹⁶¹Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan. ¹⁶²Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA. ¹⁶³Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ¹⁶⁴MRC-PHE Centre for Environment and Health, School of Public Health, Department of Epidemiology and Biostatistics and the NIHR Imperial Biomedical Research Centre, Imperial College London, London, UK. ¹⁶⁵Department of Epidemiology and Biostatistics, Imperial College London, London, UK. ¹⁶⁶Department of Cardiology, Ealing Hospital NHS Trust, Southall, UK. ¹⁶⁷National Heart, Lung and Blood Research Institute, Division of Intramural Research, Population Sciences Branch, Framingham, MA, USA. ¹⁶⁸A list of members and affiliations appears at the end of the paper. ¹⁶⁹Department of Pharmaceutical Sciences, College of Pharmacy, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA. ¹⁷⁰Oklahoma Center for Neuroscience, Oklahoma City, OK, USA. ¹⁷¹Department of Pathology and Genetics, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden. ¹⁷²Department of Neurology, Helsinki University Hospital, Helsinki, Finland. ¹⁷³Clinical Neurosciences, Neurology, University of Helsinki, Helsinki, Finland. ¹⁷⁴Department of Neurology, University of Washington, Seattle, WA, USA. ¹⁷⁵Albrecht Kossel Institute, University Clinic of Rostock, Rostock, Germany. ¹⁷⁶Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. ¹⁷⁷Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ¹⁷⁸Faculty of Medicine, University of Iceland, Reykjavik, Iceland. ¹⁷⁹Departments of Neurology and Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA. ¹⁸⁰Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, San Antonio, TX, USA. ¹⁸¹Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA. ¹⁸²Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. ¹⁸³Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. ¹⁸⁴German Center for Neurodegenerative Diseases (DZNE), Munich, Germany. ¹⁸⁵These authors contributed equally: Rainer Malik, Ganesh Chauhan, Matthew Traylor, Muralidharan Sargurupremraj and Yukinori Okada. ¹⁸⁶These authors jointly supervised this work: Kari Stefansson, Bradford B. Worrall, Steven J. Kittner, Sudha Seshadri, Myriam Fornage, Hugh S. Markus, Joanna M. M. Howson, Yoichiro Kamatani, Stephanie Dobbie and Martin Dichgans. *e-mail: stephanie.debette@u-bordeaux.fr; martin.dichgans@med.uni-muenchen.de

MEGASTROKE Consortium:

Rainer Malik¹, Ganesh Chauhan², Matthew Traylor³, Muralidharan Sargurupremraj^{4,5}, Yukinori Okada^{6,7,8}, Aniket Mishra^{4,5}, Loes Rutten-Jacobs³, Anne-Katrin Giese⁹, Sander W. van der Laan¹⁰, Solveig Gretarsdottir¹¹, Christopher D. Anderson^{12,13,14}, Michael Chong¹⁵, Hieab H. H. Adams^{16,17}, Tetsuro Ago¹⁸, Peter Almgren¹⁹, Philippe Amouyel^{20,21}, Hakan Ay^{13,22}, Traci M. Bartz²³, Oscar R. Benavente²⁴, Steve Bevan²⁵, Giorgio B. Boncoraglio²⁶, Robert D. Brown Jr²⁷, Adam S. Butterworth^{28,29}, Caty Carrera^{30,31}, Cara L. Carty^{32,33}, Daniel I. Chasman^{34,35}, Wei-Min Chen³⁶, John W. Cole³⁷, Adolfo Correa³⁸, Ioana Cotlarciuc³⁹, Carlos Cruchaga^{40,41}, John Danesh^{28,42,43,44}, Paul I. W. de Bakker^{45,46}, Anita L. DeStefano^{47,48}, Marcel den Hoed⁴⁹, Qing Duan⁵⁰, Stefan T. Engelter^{51,52}, Guido J. Falcone^{53,54}, Rebecca F. Gottesman⁵⁵, Raji P. Grewal⁵⁶, Vilmundur Gudnason^{57,58}, Stefan Gustafsson⁵⁹, Jeffrey Haessler⁶⁰, Tamara B. Harris⁶¹, Ahamad Hassan⁶², Aki S. Havulinna^{63,64}, Susan R. Heckbert⁶⁵, Elizabeth G. Holliday^{66,67}, George Howard⁶⁸, Fang-Chi Hsu⁶⁹, Hyacinth I. Hyacinth⁷⁰, M. Arfan Ikram¹⁶, Erik Ingelsson^{71,72}, Marguerite R. Irvin⁷³, Xueqiu Jian⁷⁴, Jordi Jiménez-Conde⁷⁵, Julie A. Johnson^{76,77}, J. Wouter Jukema⁷⁸, Masahiro Kanai^{6,7,79}, Keith L. Keene^{80,81}, Brett M. Kissela⁸², Dawn O. Kleindorfer⁸², Charles Kooperberg⁶⁰, Michiaki Kubo⁸³, Leslie A. Lange⁸⁴, Carl D. Langefeld⁸⁵, Claudia Langenberg⁸⁶, Lenore J. Launer⁸⁷, Jin-Moo Lee⁸⁸, Robin Lemmens^{89,90}, Didier Leys⁹¹, Cathryn M. Lewis^{92,93},

Wei-Yu Lin^{28,94}, Arne G. Lindgren^{95,96}, Erik Lorentzen⁹⁷, Patrik K. Magnusson⁹⁸, Jane Maguire⁹⁹, Ani Manichaikul³⁶, Patrick F. McArdle¹⁰⁰, James F. Meschia¹⁰¹, Braxton D. Mitchell^{100,102}, Thomas H. Mosley^{103,104}, Michael A. Nalls^{105,106}, Toshiharu Ninomiya¹⁰⁷, Martin J. O'Donnell^{15,108}, Bruce M. Psaty^{109,110,111,112}, Sara L. Pulit^{45,113}, Kristiina Rannikmäe^{114,115}, Alexander P. Reiner^{65,116}, Kathryn M. Rexrode¹¹⁷, Kenneth Rice¹¹⁸, Stephen S. Rich³⁶, Paul M. Ridker^{34,35}, Natalia S. Rost^{9,13}, Peter M. Rothwell¹¹⁹, Jerome I. Rotter^{120,121}, Tatjana Rundek¹²², Ralph L. Sacco¹²², Saori Sakaue^{7,123}, Michele M. Sale¹²⁴, Veikko Salomaa⁶³, Bishwa R. Sapkota¹²⁵, Reinhold Schmidt¹²⁶, Carsten O. Schmidt¹²⁷, Ulf Schminke¹²⁸, Pankaj Sharma³⁹, Agnieszka Slowik¹²⁹, Cathie L. M. Sudlow^{114,115}, Christian Tanislav¹³⁰, Turgut Tatlisumak^{131,132}, Kent D. Taylor^{120,121}, Vincent N. S. Thijs^{133,134}, Gudmar Thorleifsson¹¹, Unnur Thorsteinsdottir¹¹, Steffen Tiedt¹, Stella Trompet¹³⁵, Christophe Tzourio^{5,136,137}, Cornelia M. van Duijn^{138,139}, Matthew Walters¹⁴⁰, Nicholas J. Wareham⁸⁶, Sylvia Wassertheil-Smoller¹⁴¹, James G. Wilson¹⁴², Kerri L. Wiggins¹⁰⁹, Qiong Yang⁴⁷, Salim Yusuf¹⁵, Najaf Amin¹⁶, Hugo S. Aparicio^{48,187}, Donna K. Arnett¹⁸⁸, John Attia¹⁸⁹, Alexa S. Beiser^{47,48}, Claudine Berr¹⁹⁰, Julie E. Buring^{34,35}, Mariana Bustamante¹⁹¹, Valeria Caso¹⁹², Yu-Ching Cheng¹⁹³, Seung Hoan Choi^{48,194}, Ayesha Chowhan^{48,187}, Natalia Cullell³¹, Jean-François Dartigues^{195,196}, Hossein Delavaran^{95,96}, Pilar Delgado¹⁹⁷, Marcus Dörr^{198,199}, Gunnar Engström¹⁹, Ian Ford²⁰⁰, Wander S. Gurpreet²⁰¹, Anders Hamsten^{202,203}, Laura Heitsch²⁰⁴, Atsushi Hozawa²⁰⁵, Laura Ibanez²⁰⁶, Andreea Ilinca^{95,96}, Martin Ingelsson²⁰⁷, Motoki Iwasaki²⁰⁸, Rebecca D. Jackson²⁰⁹, Katarina Jood²¹⁰, Pekka Jousilahti⁶³, Sara Kaffashian³, Lalit Kalra²¹¹, Masahiro Kamouchi²¹², Takanari Kitazono²¹³, Olafur Kjartansson²¹⁴, Manja Kloss²¹⁵, Peter J. Koudstaal²¹⁶, Jerzy Krupinski²¹⁷, Daniel L. Labovitz²¹⁸, Cathy C. Laurie¹¹⁸, Christopher R. Levi²¹⁹, Linxin Li²²⁰, Lars Lind²²¹, Cecilia M. Lindgren^{222,223}, Vasileios Lioutas^{48,224}, Yong Mei Liu²²⁵, Oscar L. Lopez²²⁶, Hirata Makoto²²⁷, Nicolas Martinez-Majander¹⁷², Koichi Matsuda²²⁷, Naoko Minegishi²⁰⁵, Joan Montaner²²⁸, Andrew P. Morris^{229,230}, Elena Muiño³¹, Martina Müller-Nurasyid^{231,232,233}, Bo Norrving^{95,96}, Soichi Ogishima²⁰⁵, Eugenio A. Parati²³⁴, Leema Reddy Peddareddygar⁵⁶, Nancy L. Pedersen^{98,235}, Joanna Pera¹²⁹, Markus Perola^{63,236}, Alessandro Pezzini²³⁷, Silvana Pileggi²³⁸, Raquel Rabionet²³⁹, Iolanda Riba-Llena³⁰, Marta Ribasés²⁴⁰, Jose R. Romero^{48,187}, Jaume Roquer^{241,242}, Anthony G. Rudd^{243,244}, Antti-Pekka Sarin^{245,246}, Ralhan Sarju²⁰¹, Chloe Sarnowski^{47,48}, Makoto Sasaki²⁴⁷, Claudia L. Satizabal^{48,187}, Mamoru Satoh²⁴⁷, Naveed Sattar²⁴⁸, Norie Sawada²⁰⁸, Gerli Sibolt¹⁷², Ásgeir Sigurdsson²⁴⁹, Albert Smith²⁵⁰, Kenji Sobue²⁴⁷, Carolina Soriano-Tárraga²⁴², Tara Stanne²⁵¹, O. Colin Stine²⁵², David J. Stott²⁵³, Konstantin Strauch^{231,254}, Takako Takai²⁰⁵, Hideo Tanaka^{255,256}, Kozo Tanno²⁴⁷, Alexander Teumer²⁵⁷, Liisa Tomppo¹⁷², Nuria P. Torres-Aguila³¹, Emmanuel Touze^{258,259}, Shoichiro Tsugane²⁰⁸, Andre G. Uitterlinden²⁶⁰, Einar M. Valdimarsson²⁶¹, Sven J. van der Lee¹⁶, Henry Völzke²⁵⁷, Kenji Wakai²⁵⁵, David Weir²⁶², Stephen R. Williams²⁶³, Charles D. A. Wolfe^{243,244}, Quenna Wong¹¹⁸, Huichun Xu¹⁹³, Taiki Yamaji²⁰⁸, Dharambir K. Sanghera^{125,169,170}, Olle Melander¹⁹, Christina Jern¹⁷¹, Daniel Strbian^{172,173}, Israel Fernandez-Cadenas^{30,31}, W. T. Longstreth Jr^{65,174}, Arndt Rolfs¹⁷⁵, Jun Hata¹⁰⁷, Daniel Woo⁸², Jonathan Rosand^{12,13,14}, Guillaume Pare¹⁵, Jemma C. Hopewell¹⁷⁶, Danish Saleheen¹⁷⁷, Kari Stefansson^{11,178}, Bradford B. Worrall¹⁷⁹, Steven J. Kittner³⁷, Sudha Seshadri^{48,180}, Myriam Fornage^{74,181}, Hugh S. Markus³, Joanna M. M. Howson²⁸, Yoichiro Kamatani^{6,182}, Stephanie Debette^{4,5} and Martin Dichgans^{1,183,184}

¹⁸⁷Boston University School of Medicine, Boston, MA, USA. ¹⁸⁸University of Kentucky College of Public Health, Lexington, KY, USA. ¹⁸⁹University of Newcastle and Hunter Medical Research Institute, New Lambton, New South Wales, Australia. ¹⁹⁰INSERM, U1061, Université Montpellier, Montpellier, France. ¹⁹¹Centre for Research in Environmental Epidemiology, Barcelona, Spain. ¹⁹²Department of Neurology, Università degli Studi di Perugia, Umbria, Italy. ¹⁹³Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. ¹⁹⁴Broad Institute, Cambridge, MA, USA. ¹⁹⁵Bordeaux Population Health Research Center, INSERM, UMR 1219, Université Bordeaux, Bordeaux, France. ¹⁹⁶Department of Neurology, Memory Clinic, Bordeaux University Hospital, Bordeaux, France. ¹⁹⁷Neurovascular Research Laboratory, Vall d'Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona, Vall d'Hebrón Hospital, Barcelona, Spain. ¹⁹⁸Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. ¹⁹⁹DZHK, Greifswald, Germany. ²⁰⁰Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK. ²⁰¹Hero DMC Heart Institute, Dayanand Medical College & Hospital, Ludhiana, India. ²⁰²Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. ²⁰³Karolinska Institutet, Stockholm, Sweden. ²⁰⁴Division of Emergency Medicine, and Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA. ²⁰⁵Tohoku Medical Megabank Organization, Sendai, Japan. ²⁰⁶Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. ²⁰⁷Department of Public Health and Caring Sciences/Geriatrics, Uppsala University, Uppsala, Sweden. ²⁰⁸Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan. ²⁰⁹Department of Internal Medicine and the Center for Clinical and Translational Science, Ohio State University, Columbus, OH, USA. ²¹⁰Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Goteborg, Sweden. ²¹¹Department of Basic and Clinical Neurosciences, King's College London, London, UK. ²¹²Department of Health Care Administration and Management, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. ²¹³Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. ²¹⁴Departments of Neurology & Radiology, Landspítali National University Hospital, Reykjavik, Iceland. ²¹⁵Department of Neurology, Heidelberg University Hospital, Heidelberg, Germany. ²¹⁶Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands. ²¹⁷Hospital Universitari Mutua Terrassa, Terrassa (Barcelona), Spain. ²¹⁸Montefiore Medical Center, Albert Einstein College of Medicine, New York, NY, USA. ²¹⁹John Hunter Hospital, Hunter Medical Research Institute and University of Newcastle, Newcastle, New South Wales, Australia. ²²⁰Centre for Prevention of Stroke and Dementia, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. ²²¹Department of Medical Sciences, Uppsala University, Uppsala, Sweden. ²²²Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. ²²³Wellcome Trust Centre for Human Genetics, Oxford, UK. ²²⁴Beth Israel Deaconess Medical Center, Boston, MA, USA. ²²⁵Wake Forest School of Medicine, Wake Forest, NC, USA. ²²⁶Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA. ²²⁷BioBank Japan, Laboratory of Clinical Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan. ²²⁸Neurovascular Research Laboratory, Vall d'Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona, Vall d'Hebrón Hospital, Barcelona, Spain. ²²⁹Department of Biostatistics, University of Liverpool, Liverpool, UK. ²³⁰Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. ²³¹Institute of Genetic Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany. ²³²Department of Medicine I, Ludwig-Maximilians-Universität, Munich, Germany. ²³³DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany. ²³⁴Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. ²³⁵MEB, Karolinska Institutet, Stockholm, Sweden. ²³⁶Estonian Genome Center, University of Tartu, Tartu, Estonia. ²³⁷Department of Clinical and Experimental Sciences, Neurology Clinic, University of Brescia, Brescia, Italy. ²³⁸Translational Genomics Unit, Department of Oncology, IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy. ²³⁹Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain. ²⁴⁰Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addictions, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Spain. ²⁴¹Department of Neurology, IMIM-Hospital del Mar, and Universitat Autònoma de Barcelona, Barcelona, Spain. ²⁴²IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. ²⁴³National Institute for Health Research Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust and King's College London, London, UK. ²⁴⁴Division of Health and Social Care Research, King's College London, London, UK. ²⁴⁵FIMM-Institute for Molecular Medicine Finland, Helsinki, Finland. ²⁴⁶THL-National Institute for Health and Welfare, Helsinki, Finland. ²⁴⁷Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan. ²⁴⁸BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow, UK. ²⁴⁹deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. ²⁵⁰Icelandic Heart Association, Reykjavik, Iceland. ²⁵¹Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Goteborg, Sweden. ²⁵²Department of Epidemiology, University of Maryland School of Medicine, Baltimore, MD, USA. ²⁵³Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow, UK. ²⁵⁴IBE, Faculty of Medicine, LMU Munich, Munich, Germany. ²⁵⁵Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. ²⁵⁶Department of Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan. ²⁵⁷Institute for Community Medicine, SHIP-KEF, University Medicine Greifswald, Greifswald, Germany. ²⁵⁸Department of Neurology, Caen University Hospital, Caen, France. ²⁵⁹University of Caen Normandy, Caen, France. ²⁶⁰Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. ²⁶¹Landspítali University Hospital, Reykjavik, Iceland. ²⁶²Survey Research Center, University of Michigan, Ann Arbor, MI, USA. ²⁶³Department of Neurology, University of Virginia, Charlottesville, VA, USA.

Methods

Study design and phenotyping. A detailed description of the study design, participating studies, and phenotype definitions for stroke and stroke subtypes is provided in the Supplementary Note. Characteristics of study participants are shown in Supplementary Table 2 for each study. All participants provided written informed consent, and local research ethics committees and institutional review boards approved the individual studies.

Genotyping, imputation, and quality control. Genotyping platforms and imputation methods for each participating study are described in Supplementary Table 2. All studies used imputed genotypes based on at least the 1000G phase 1 multiethnic reference panel and conducted logistic regression analyses (or Cox regression for longitudinal population-based cohort studies) for five stroke traits (AS, AIS, LAS, CES, and SVS) with all measured and imputed genetic variants in dosage format by using appropriate software under an additive genetic model with a minimum of sex and age as covariates. Information on additional covariates is given in Supplementary Table 2.

Before ancestry-specific meta-analysis, QC was performed on each study by two independent researchers following a standardized protocol based on the suggestions of Winkler et al.⁷⁴. Marker names and alleles were harmonized across studies. Meta-analyses were restricted to autosomal biallelic markers from the 1000G phase 1 v3. Duplicate markers were removed from each study. *P*-*Q* plots, *Q*-*Q* plots and allele-frequency-plots were constructed for each study. After visual inspection, analysis and QC were repeated if deemed necessary. QC was conducted independently for all participating studies in at least two sites.

Individual study-level filters were set to remove extreme effect values ($\beta > 5$ or $\beta < -5$), rare SNPs (MAF < 0.01) and variants with low imputation accuracy (oevar_imp or info score < 0.5). The effective allele count was defined as twice the product of the MAF, imputation accuracy (r^2 , info score or oevar_imp), and number of cases. Variants with an effective allele count < 10 were excluded⁷⁴. The number of SNPs passing QC for each study is given in Supplementary Table 26.

Genome-wide-association meta-analyses. The overall analytical strategy is shown in Fig. 1. We first conducted fixed-effects inverse-variance-weighted meta-analysis with METAL⁷⁵ in each ancestral group (EUR, EAS, AFR, SAS, LAT, and other ASN), then performed meta-analysis of the ancestry-specific meta-analysis results. We constructed two versions of each meta-analysis: one with single genomic control applied and one without genomic control (for LD-score regression analysis).

The EUR-specific and transancestral fixed-effects meta-analyses were further filtered for heterogeneity ($P_{\text{het}} < 5.0 \times 10^{-8}$) and for the number of cases included for a specific marker ($< 50\%$ of stroke cases were excluded). In addition, we ran a transancestral GWAS meta-analysis, using MANTRA¹⁵, which was based on ancestry-specific meta-analysis results. The final MANTRA results were filtered for a MANTRA posterior-probability heterogeneity $P < 0.95$. SNPs with $\log_{10}(\text{BF}) > 6$ were considered to be genome-wide significant, whereas SNPs with $6 > \log_{10}(\text{BF}) > 5$ were considered to show suggestive association. We used a method based on summary statistics⁷⁶ to estimate the variance in liability explained by each lead variant. Disease prevalence was set to 5.5% for AS, to 4.4% for AIS, and to 0.11% for IS subtype in Europeans⁷⁷. Disease prevalence was set to 2.97% for AIS, to 0.91% for LAS, to 0.24% for CES, and to 1.76% for SVS in East Asians (Hisayama study, J. Hata unpublished data and ref.⁹⁰). We used summary statistics from the Europeans-only fixed-effects meta-analysis and the East Asian-only fixed-effects meta-analysis. Genomic inflation was calculated as lambda in the GenABEL package (available through CRAN repositories). In addition, we calculated the LD-score-regression intercepts for the Europeans-only fixed-effects meta-analysis, using European LD scores.

Shared genetic influences of individual loci on mechanistically defined stroke subtypes

We used gwas-pw²² to detect shared genetic influences of LAS, CES, and SVS, aiming to identify genetic variants that influence respective pairs of these traits. Gwas-pw estimates the PPA for four models. Model 3 is the model in which a given genomic region contains a genetic variant that influences both traits. We used the fixed-effects transancestral meta-analysis results as input, transforming results into signed *Z* scores based on the *P* value and sign of the log(OR). The chunk size (number of SNPs included in each chunk analyzed) was set automatically by using an approximately independent block file (ld-select), as provided by the software. Correlation was set to reflect the overlap in controls. We deemed the results of model 3 with a PPA > 0.9 significant²².

Conditional analysis. We used GCTA-COJO⁷⁸ to perform conditional association analysis in each of the stroke loci in Europeans. We first fit a stepwise joint regression model including all SNPs with joint $P < 5.0 \times 10^{-8}$. In instances in which regions included only one SNP, we fit a model including the top two SNPs from each region. The models made use of (i) summary statistics from the Europeans-only meta-analysis presented herein and (ii) genotype data for 3,291 stroke cases and 11,820 controls of North European ancestry from NINDS-SiGN as an LD reference for each region.

Gene-based analysis. We performed gene-based tests by using the VEGAS approach⁷⁹ implemented in VEGAS2 software¹⁶. We used 24,769 autosomal refseq genes to perform gene-based association studies. To perform gene-based association tests, we used the 1000G phase 3 super populations African (AFR), East Asian (EAS), European (EUR), American (AMR) and South Asian (SAS) as a reference to compute the pairwise LD between variants residing within a gene. We performed gene-based tests, using the ‘-top 10’ parameter in VEGAS2, which tests enrichment of the top 10% of association *P* values within a gene. To maintain specificity while including cis-regulatory variants, we included variants located within 10 kb of a gene’s 3’ and 5’ UTRs. We performed 1×10^6 simulations to compute empirical *P* values for association with each gene. For genes with $P < 1 \times 10^{-5}$, we increased the number of simulations to 1×10^8 to increase the accuracy of the association *P* values. For individual stroke subtypes, we performed ancestry-specific gene-based association followed by meta-analysis of gene association *P* values by using Stouffer’s method, based on sample size.

Association of individual stroke risk variants with related vascular traits

We systematically explored genetic overlap with AF, CAD, cIMT, cPL, diastolic BP, systolic BP, HDL-cholesterol levels, LDL-cholesterol levels, triglyceride levels, T2D, VTE, and WMH. First, we acquired summary statistics from the appropriate consortia (Supplementary Table 12). For each of the nonstroke phenotypes, we constructed a SNP set including the index variant of the nonstroke phenotype with $P < 1.3 \times 10^{-4}$ plus all variants in high LD (r^2 in 1000G EUR > 0.9 with this index variant). If the MEGASTROKE lead SNP was included in this set of SNPs, we deemed the overlap with the nonstroke phenotype to be significant. We show two different tiers: (i) variants that showed genome-wide significance in the related vascular trait ($P < 5.0 \times 10^{-8}$) and (ii) variants that were not genome-wide significant but passed Bonferroni correction ($P = 1.3 \times 10^{-4}$).

Association of genetic risk scores of related vascular traits with stroke and stroke subtypes

Genetic risk scores generated from variants shown to have genome-wide association with various vascular risk factors (VTE, DBP, SBP, MAP, PP, HTN, HDL cholesterol, LDL cholesterol, triglycerides, T2D, and CAD) were used to estimate the overlap between vascular traits and stroke and its subtypes. The effect allele for each risk-factor variant was defined as the allele associated with increased risk-factor levels. The corresponding allele information, β coefficients and standard errors from different stroke subtypes were extracted and used as input. Association was tested with the inverse-variance weighting (IVW) method implemented as an R package gtx V 0.0.8 (available through CRAN repositories).

We further conducted sensitivity analyses, using the MR-Egger method implemented as an R package (TwoSampleMR, available through CRAN repositories)⁸⁰, which, unlike the IVW method, estimates the intercept term as part of the analysis. An intercept term significantly differing from zero suggests the presence of directional pleiotropy. We used a conservative significance threshold of $P < 0.05$ for the intercept. In the presence of directional pleiotropy, leave-one-out analysis was carried out by retesting the association of the vascular GRS with the outcome (stroke), leaving out each SNP in turn to determine whether a single SNP drives the association. We manually identified outlier SNPs that might drive the observed directional pleiotropy and then repeated the analyses (IVW and MR-Egger) after excluding the variants exhibiting directional pleiotropy.

The selection of SNPs for the vascular GRS was based on literature (PubMed) searches and the GWAS catalog (<http://www.ebi.ac.uk/gwas/>), and was used to identify studies that performed GWAS of the various risk factors. The most recent and largest GWAS of each risk factor was selected, and the associated variant details were retrieved. For the GRS analysis, only independent variants ($r^2 < 0.01$, based on the 1000G EUR panel) were used for the analysis (Supplementary Table 27). Risk-variant selection for BP traits (SBP, DBP, MAP, and PP) was further extended to studies with gene-centric chips. We used β coefficients extracted from the summary statistics of the International Consortium of BP GWAS^{81,82} as weights for this GRS analysis. A *P*-value $< 5.6 \times 10^{-3}$ correcting for nine independent phenotypes was considered significant. The number of independent vascular phenotypes, taking into account the correlation between the phenotypes considered, was estimated on the basis of individual-level data from the 3C study by using the online tool matSpDlite (<http://neurogenetics.qimrberghofer.edu.au/matSpDlite/>).

Shared genetic contribution to stroke and related vascular traits at the genome-wide level

We used LD-score regression to estimate the genetic correlation between stroke and related vascular traits^{33,34}. We conducted analyses on the European and East Asian stroke GWAS summary statistics only. Summary statistics from the GWAS meta-analyses for vascular risk factors and intermediate or related vascular phenotypes (BP, blood lipids, T2D, cIMT, cPL, AF, VTE, CAD, and WMH) were acquired from the respective consortia, as detailed in Supplementary Table 12. For LD-score regression in East Asians, we further received prepublication access to summary statistics of GWAS for blood lipids conducted in BioBank Japan⁹¹, as described in the Supplementary Note. For each trait, we filtered the summary statistics to the subset of HapMap 3 SNPs to decrease the potential for bias due to poor imputation quality. Analyses were performed separately by using summary statistics from the European- and East Asian-specific

meta-analysis. We used the European or East Asian LD-score files calculated from the 1000G reference panel and provided by the developers. A P value $<5.6 \times 10^{-3}$ correcting for nine independent phenotypes was considered significant. All analyses were performed with the ldsc package (<https://github.com/bulik/ldsc/>).

Global epigenetic patterns at the 32 stroke risk loci. We used the epigwas tool³⁵ to test for cell-specific enrichment in chromatin marks that have previously been shown to be phenotypically cell-type specific in ENCODE and/or RoadMap epigenome data (H3K4me1, H3K4me3, and H3K9ac)³⁵, leveraging the recent release of ENCODE/RoadMap epigenome data from 127 tissue types³⁶. Histone ChIP-seq data for narrow contiguous regions of enrichment were used to calculate the enrichment score (height of the nearest tall peak/distance to the peak) for the lead variant and proxies ($r^2 > 0.8$ in the 1000G cosmopolitan panel). Significance was estimated by examining the relative proximity and specificity of the test SNP set with 10,000 sets (permutation) of matched background. In addition, Bonferroni correction for the number of chromatin marks tested was applied.

Pathway analyses. To identify pathways overrepresented in the stroke association results, we used data-driven expression-prioritized integration for complex traits (DEPICT³⁷), IPA (<https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>), and VEGAS2Pathway³⁹. DEPICT version 1 release 194 was used to identify biological pathways, tissues, and cell types enriched among suggestive associations ($\log_{10}(\text{BF}) > 5$) for any stroke and stroke subtypes in the MANTRA transancestral GWAS. Results are presented for the MANTRA transancestral analysis. We deemed DEPICT pathways with an FDR < 0.05 statistically significant.

IPA was conducted by using an extended list comprising 214 genes located in the boundaries defined by $r^2 > 0.5$ with the lead SNP in Europeans or East Asians, or located ± 50 kb from the lead SNP, for all suggestive loci reaching $P < 1.0 \times 10^{-5}$ or $\log_{10}(\text{BF}) > 5$ (Supplementary Table 25). This gene list was taken as an input for IPA using only findings from human and experimentally verified results. Otherwise, standard parameters were used for the analysis. We corrected canonical pathway P -value analysis with the Benjamini–Hochberg method and deemed an FDR < 0.05 significant.

We performed gene-wide gene-set enrichment analysis, using the VEGAS2Pathway approach³⁹ to test which Biosystem terms³⁸ were enriched with VEGAS2-derived gene association P values for stroke subtypes. VEGAS2Pathway performs a competitive gene-set enrichment test while accounting for gene density in LD blocks (or correlated association P values of neighboring genes), SNP density, and pathway size by using a resampling strategy.

For individual stroke subtypes, we performed separate ancestry-specific gene-set enrichment analysis. Next, we combined the gene-set-enrichment association P values across ancestries by using Stouffer's method for sample-size-weighted combinations of P values. For each stroke subtype, we tested the association of 9,981 Biosystem gene-set terms.

Fine mapping derived from credible SNP-set analyses. We implemented the method of Maller et al.⁸³, converting our ancestry-specific meta-analysis P values to Bayes factors through Wakefield's approximation⁴⁰ in all stroke phenotypes in the EU- only, EAS-only, and AFR-only analysis. We used all SNPs in LD with the lead SNP ($r^2 > 0.1$, ancestry specific). The Bayes factors were then used to calculate posterior probabilities on the basis of the assumption of a single causal SNP in each region. For all regions, we constructed 95% credible sets of potentially causal SNPs.

Investigation of eQTLs, pQTLs, meQTLs, and regulatory marks in different tissues. The following datasets, covering a large variety of tissue and cell types, were interrogated for eQTLs, pQTLs, and meQTLs:

1. The Genotype-Tissue Expression (GTEx-V6) project data, providing significant eQTL information from 44 postmortem tissues (449 individuals) (<http://biorgx.org/content/early/2016/09/09/074450/>), with significance based on a gene-specific P -value threshold that is permutation-adjusted for multiple SNPs per gene
2. The Genome-wide Repository of Associations between SNPs and Phenotypes, build 2.0 (GRASP2)^{46,47}, as well as a collected expression and epigenetic QTL database of > 100 sources covering a wide range of cell and tissue types (Supplementary Note), using $P < 5 \times 10^{-6}$ as a significance threshold for association with expression of a transcript in the original study
3. The Human Genetic Variation Database (HGVD)⁴⁸, providing eQTL information from peripheral-blood cells in a Japanese population ($n = 1,208$), with significance defined by FDR $< 5\%$
4. The Biobank-based Integrative Omics Studies (BIOS), providing eQTLs from peripheral-blood RNA-seq data in 2,116 unrelated individuals⁴⁹, with significance defined by FDR $< 5\%$
5. A subset of the Blueprint epigenome project⁵⁰ with eQTL, meQTL, and histone-modification data (H3K4me1 and H3K27ac) in CD14⁺ monocytes, CD16⁺ neutrophils, and CD4⁺ naive T cells from 197 individuals; these were mapped through the classical QTL association test, allele-specific-expression test, and combined haplotype test, with significance defined by FDR $< 5\%$

6. The Stockholm–Tartu Atherosclerosis Reverse Networks Engineering Task study (STARNET)⁵¹, providing eQTL data from vascular and metabolic tissues in 600 patients with CAD, with Benjamini–Hochberg–corrected association P values ($P < 0.05$)
7. The aortic endothelial cell study⁵², providing eQTL data from human aortic endothelial cells in 147 individuals, with Bonferroni multiple testing correction for the number of independent SNPs ($P < 1.0 \times 10^{-4}$)
8. The ARIC cohort⁵³, providing meQTL information from peripheral blood in 794 individuals of European ancestry and 784 individuals of African American ancestry, with multiple testing correction for the number of unique CpG probes in the look-up
9. The Cooperative Health Research in the Region of Augsburg (KORA) cohort, with pQTL information from the human blood plasma proteome⁵⁴, measuring 1,124 proteins on the SomaSCAN platform in 1,000 participants; significance for each association was set at $P < 5.0 \times 10^{-8}$

In each of these datasets, we report the most significant cis-QTL, meQTL, or pQTL surpassing a study-specific predefined significance level or FDR, considering only QTLs in LD with the lead stroke SNP at an $r^2 > 0.8$ (in 1000G, as well as queries of multiple builds of SNAP⁸⁴ and SNIpA⁸⁵), thus suggesting high concordance. The results are presented grouped per tissue or cell type (Supplementary Table 23), or per stroke risk locus (Supplementary Table 22). In addition, we also systematically report the association of the top QTL with stroke risk and of the lead stroke risk variant with the corresponding transcript expression, methylation level, or protein level (Supplementary Table 23).

In addition, we used a subset of the Blueprint epigenome project in CD14⁺ monocytes, CD16⁺ neutrophils, and CD4⁺ naive T cells from 197 individuals⁵⁰ and Haploreg V4 (ref. ⁸⁶) to annotate the lead variants and proxies for enrichment in specific histone-modification marks for the chromatin state, on the basis of ChIP-seq data from multiple cell/tissue types from ENCODE⁸⁷ and NIH RoadMap epigenome⁸⁸. The results for each of the lead SNPs and its proxies are displayed in detail in Supplementary Table 22.

Integration of association statistics and in silico functional information in RiVIERA-beta. To identify the most plausible causal variants and genes, we used RiVIERA software⁵⁹, which jointly models the summary association statistics and the corresponding epigenetic regulatory information in a Bayesian framework to estimate the PPA. The empirical prior of a variant to be associated with the respective trait through regulatory features was generated by using the 848 tissue-specific epigenomic data in seven chromatin (H3K4me1, H3K4me3, H3K36me3, H3K27me3, H3K9me3, H3K27ac, and H3K9ac) and DNA-accessibility (DNase I) marks from the ENCODE/RoadMap epigenome data. Binary epigenomic annotation matrices of a variant overlapping the narrow peaks were generated. For inferring the causal region, RiVIERA-beta performs a repeated ($n = 1,000$) random-sampling step per locus, with the step size set to 1.0×10^{-4} . Iteration is performed until convergence (acceptance rate $> 60\%$) is achieved, which is critical for the accurate estimation of PPA. We generated 95% credible sets in each region on the basis of the PPA. Regional plots were generated by using the association statistics and the PPA. Epigenetic enrichment over a fixed window size (50 bp) per tissue group was generated by taking the cumulative sum of empirical prior weighted global epigenetic enrichment. Tissues were divided into 19 groups, as defined in the NIH RoadMap epigenome project.

Scoring method. To prioritize the most likely biological-candidate genes, we integrated functional and biological information into an empirical score for each of the genes residing in the 32 genome-wide-significant loci. These comprised 149 genes within the region defined by an $r^2 > 0.5$ in any of the 1000G European or East Asian populations or physical distances of ± 50 kb from the lead SNP of the respective locus (Supplementary Table 25). A score of 1 was assigned for being the nearest gene to the lead SNP, for containing a missense variant, for containing histone-mark H3K4me3, H3K9ac, and H3K4me1 peaks in cell types that showed significant enrichment in epigwas analysis, and for functioning as an eGene for an eQTL, meQTL, or pQTL (one point for each) in at least one study and one cell/tissue type. In addition, a score of 1 was assigned for each stroke phenotype showing evidence of being a drug-target gene in the DrugBank database (ATC-C and ATC-B01) and the Therapeutic Target Database (Supplementary Table 25), and for overlap with biological pathways in DEPICT, IPA, or VEGAS2 (Supplementary Tables 18–20).

Drug-target gene-enrichment analysis. For each locus containing a variant with $\log_{10}(\text{BF}) > 5$ in the MANTRA analysis, we annotated the genes by considering LD structures ($r^2 > 0.5$ in any of 1000G EUR or ASN populations) or physical distances (± 50 kb) from the lead SNP of the respective locus. Drug-target genes were extracted from the DrugBank database⁸⁸ (considering those registered as pharmacological active targets; <https://www.drugbank.ca/>) and Therapeutic Target Database⁸⁹ (TTD; http://bidb.nus.edu.sg/group/cjtd/TTD_HOME.asp), thus resulting in a list of 1,123 genes (and corresponding proteins) annotated to currently approved drugs indicated for any diseases (Supplementary Table 25). Drugs indicated for antithrombotic therapy ($n = 69$) and cardiovascular diseases

($n = 324$) were curated from Anatomical Therapeutic Chemical (ATC) codes (Supplementary Table 25). Enrichment of overlap between stroke-associated genes with drug targets for antithrombotic therapy and cardiovascular diseases was assessed with Fisher's exact test.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. The datasets generated and/or analyzed during the current study are available from the corresponding authors upon reasonable request.

References

74. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
75. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
76. So, H. C., Gui, A. H., Cherny, S. S. & Sham, P. C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.* **35**, 310–317 (2011).
77. Feigin, V. L., Lawes, C. M., Bennett, D. A. & Anderson, C. S. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol.* **2**, 43–53 (2003).
78. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012). S1–S3.
79. Liu, J. Z. et al. A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* **87**, 139–145 (2010).
80. 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).
81. International Consortium for Blood Pressure Genome-Wide Association Studies et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
82. Wain, L. V. et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* **43**, 1005–1011 (2011).
83. Wellcome Trust Case Control Consortium et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
84. Johnson, A. D. et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).
85. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmüller, G. SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334–1336 (2015).
86. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D1, D877–D881 (2016).
87. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
88. Wishart, D. S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
89. Yang, H. et al. Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res.* **44**, D1069–D1074 (2016).
90. Hachiya, T. et al. Genetic predisposition to ischemic stroke: a polygenic risk score. *Stroke* **48**, 253–258 (2017).
91. Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* <https://doi.org/10.1038/s41588-018-0047-6> (2018).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

No pre-defined power analysis was carried out. We used all currently available stroke samples and controls for this study, therefore reaching the maximum sample size possible.

2. Data exclusions

Describe any data exclusions.

Inclusion and exclusion criteria for each study in MEGASTROKE are given in the Supplementary Material, Section 3, pages 64-97. Due to the extensive nature of our analysis, we cannot provide all details in this form.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Experimental replication was not performed

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Cases were defined as ischemic stroke (IS) or intracerebral hemorrhage (ICH) based on clinical and imaging criteria. IS was further subdivided into the following categories mostly using the Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria: i) large vessel ischemic stroke (LV-IS); ii) cardioembolic ischemic stroke (CE-IS); iii) small vessel ischemic stroke (SV-IS). Subarachnoid hemorrhages were excluded from all analyses. Controls were stroke-free. Where possible, cases and controls were matched for relevant covariates. Association models were additionally corrected for relevant covariates.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding was not relevant for this study, as group allocation was based on clinical and imaging criteria

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

| | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The <u>exact</u> sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly. |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement indicating how many times each experiment was replicated |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Clearly defined error bars |

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

R 3.2.5 (<https://www.r-project.org/>)
 VEGAS2 (<https://vegas2.qimrberghofer.edu.au/>)
 METAL (<http://csg.sph.umich.edu/abecasis/metal/>)
 MANTRA
 LD score regression (<https://github.com/bulik/ldsc>)
 epigwas (<http://archive.broadinstitute.org/mpg/epigwas/>)
 DEPICT (<https://github.com/perslab/depict>)
 Ingenuity Pathway Analysis (<https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>)
 gwas-pw (<https://github.com/joepickrell/gwas-pw>)
 GCTA (<http://cns.genomics.com/software/gcta/>)
 RiVIERA beta (<https://yueli-compbio.github.io/RiVIERA-beta/>)
 GenABEL (<http://www.genabel.org>)
 gtx 0.0.8 (<https://cran.r-project.org/web/packages/gtx/index.html>)
 matSpDlite (<http://neurogenetics.qimrberghofer.edu.au/matSpDlite/>)
 TwoSampleMR (<https://github.com/MRCIEU/TwoSampleMR>)

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* [guidance for providing algorithms and software for publication](#) may be useful for any submission.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used

10. Eukaryotic cell lines

- State the source of each eukaryotic cell line used.
- Describe the method of cell line authentication used.
- Report whether the cell lines were tested for mycoplasma contamination.
- If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No eukaryotic cell lines were used

No eukaryotic cell lines were used

No eukaryotic cell lines were used

No cell lines were used

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Ascertainment criteria for each study in MEGASTROKE are given in the Supplementary Note, Section 3.1. An overview of population characteristics for each study can be seen in Supplementary Table 2. Due to the extensive nature of these data, please refer to this table.