Published in final edited form as:

Nat Genet. 2018 April ; 50(4): 524–537. doi:10.1038/s41588-018-0058-3.

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

A full list of authors and affiliations appears at the end of the article.

Abstract

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 ICH[6],

---

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. stephanie.debette@u-bordeaux.fr; martin.dichgans@med.uni-muenchen.de.

[185] These authors contributed equally: Rainer Malik, Ganesh Chauhan, Matthew Traylor, Muralidharan Sargurupremraj and Yukinori Okada.

[186] These 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 Debette and Martin Dichgans.

[168] A list of members and affiliations appears at the end of the paper

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.

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

although some loci were associated with two or more stroke subtypes[7,9,11,13] or with any stroke[10]. 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 similar[14] (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 conducted[15]. 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 $<5.0 \times 10^{-8}$ or MANTRA $\log_{10}$(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) burden[17], 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 ($\log_{10}(BF) > 5.0$ 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 analyses[22], 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-BP[24], ENGAGE[25], DIAGRAM[26], CHARGE[27,28], AFGen[29], INVENT[30], and CARDIoGRAMplusC4D[31] 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 >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 tool[35] 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 ≥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 tool[37], using all SNPs with $\log_{10}$(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 $\log_{10}$(BF) >5. The extended gene list (r^2 >0.5 in 1000G Europeans or East Asians, or 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 sets[38], adapting VEGAS2Pathway[39], 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 priors[40]. 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 ANNOVAR[41]. 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 humans[42]. Both variants are predicted to be benign or tolerated by PolyPhen[43] and SIFT[44]. 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 proteinexpression 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]), HGVD[48], BIOS[49], Blueprint epigenome project (subset)[50], STARNET[51], and the human aortic endothelial cell study[52]; for meQTLs, the Blueprint epigenome project (subset)[50] and the ARIC cohort[53]; and for pQTLs, the KORA cohort[54]. 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 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 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 RiVIERA[59], 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 ≥ 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 types[60], 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 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) (OR = 8.89, P=0.0166) and after addition of 65 genes in 21 suggestive stroke risk loci (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 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 ICH[6] 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 burden[28] (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, cPL[27], and CAD[31] (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 loci for WMH.

Third, several suggestive loci ($\log_{10}$(BF) ≥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 may be driven by developmental effects, as has recently been proposed for the *FOXF2* locus[10].

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 diseases[71]. 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.

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 multiancestral 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–Z plots, QQ 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 excluded[74]. 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 METAL[75] 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 MANTRA[15], 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 $\log_{10}$ (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 statistics[76] 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 Europeans[77]. 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-pw[22] 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 significant[22].

Conditional analysis

We used GCTA-COJO[78] 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 approach[79] implemented in VEGAS2 software[16]. 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 genomewide 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 Japan [91], 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 tool[35] 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 types[36]. 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 (DEPICT[37]), IPA (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/), and VEGAS2Pathway[39]. DEPICT version 1 release 194 was used to identify biological pathways, tissues, and cell types enriched among suggestive associations ($\log_{10}$(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 $\log_{10}$(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 approach[39] to test which Biosystem terms[38] 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 geneset 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 approximation[40] 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://biorxiv.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 individuals[49], with significance defined by FDR <5%
5. A subset of the Blueprint epigenome project[50] 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 study[52], 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 cohort[53], 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 proteome[54], 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 SNAP[84] and SNiPA[85]), 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 individuals[50] 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 ENCODE[87] and NIH RoadMap epigenome[36]. 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 software[59], 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 $\log_{10}(BF) > 5$ 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 database[88] (considering those registered as pharmacological active targets; https://www.drugbank.ca/) and Therapeutic Target Database[89] (TTD; http://bidd.nus.edu.sg/group/cjttd/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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Rainer Malik[1,185], Ganesh Chauhan[2,3,185], Matthew Traylor[4,185], Muralidharan Sargurupremraj[3,5,185], Yukinori Okada[67,8,185], Aniket Mishra[3,5], Loes Rutten-Jacobs[4], Anne-Katrin Giese[9], Sander W. van der Laan[10], Solveig Gretarsdottir[11], Christopher D. Anderson[12,13,14], Michael Chong[15], Hieab H. H. Adams[16,17], Tetsuro Ago[18], Peter Almgren[19], Philippe Amouyel[20,21], Hakan Ay[13,22], Traci M. Bartz[23], Oscar R. Benavente[24], Steve Bevan[25], Giorgio B. Boncoraglio[26], Robert D. Brown Jr[27], Adam S. Butterworth[28,29], Caty Carrera[30,31], Cara L. Carty[32,33], Daniel I. Chasman[34,35], Wei-Min Chen[36], John W. Cole[37], Adolfo Correa[38], Ioana Cotlarciuc[39], Carlos Cruchaga[40,41], John Danesh[28,42,43,44], Paul I. W. de Bakker[45,46], Anita L. DeStefano[47,48], Marcel den Hoed[49], Qing Duan[50], Stefan T. Engelter[51,52], Guido J. Falcone[53,54], Rebecca F. Gottesman[55], Raji P. Grewal[56], Vilmundur Gudnason[57,58], Stefan Gustafsson[59], Jeffrey Haessler[60], Tamara B. Harris[61], Ahamad Hassan[62], Aki S. Havulinna[63,64], Susan R. Heckbert[65], Elizabeth

G. Holliday[66,67], George Howard[68], Fang-Chi Hsu[69], Hyacinth I. Hyacinth[70], M. Arfan Ikram[16], Erik ingelsson[71,72], Marguerite R. Irvin[73], Xueqiu Jian[74], Jordi Jimenez-Conde[75], Julie A. Johnson[76,77], J. Wouter Jukema[78], Masahiro Kanai[67,79], Keith L. Keene[80,81], Brett M. Kissela[82], Dawn O. Kleindorfer[82], Charles Kooperberg[60], Michiaki Kubo[83], Leslie A. Lange[84], Carl D. Langefeld[85], Claudia Langenberg[86], Lenore J. Launer[87], Jin-Moo Lee[88], Robin Lemmens[89,90], Didier Leys[91], Cathryn M. Lewis[92,93], Wei-Yu Lin[28,94], Arne G. Lindgren[95,96], Erik Lorentzen[97], Patrik K. Magnusson[98], Jane Maguire[99], Ani Manichaikul[36], Patrick F. McArdle[100], James F. Meschia[101], Braxton D. Mitchell[100,102], Thomas H. Mosley[103,104], Michael A. Nalls[105,106], Toshiharu Ninomiya[107], 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. Rexrode[117], Kenneth Rice[118], Stephen S. Rich[36], Paul M. Ridker[34,35], Natalia S. Rost[9,13], Peter M. Rothwell[119], Jerome I. Rotter[120,121], Tatjana Rundek[122], Ralph L. Sacco[122], Saori Sakaue[7,123], Michele M. Sale[124], Veikko Salomaa[63], Bishwa R. Sapkota[125], Reinhold Schmidt[126], Carsten O. Schmidt[127], Ulf Schminke[128], Pankaj Sharma[39], Agnieszka Slowik[129], Cathie L. M. Sudlow[114,115], Christian Tanislav[130], Turgut Tatlisumak[131,132], Kent D. Taylor[120,121], Vincent N. S. Thijs[133,134], Gudmar Thorleifsson[11], Unnur Thorsteinsdottir[11], Steffen Tiedt[1], Stella Trompet[135], Christophe Tzourio[3,136,137], Cornelia M. van Duijn[138,139], Matthew Walters[140], Nicholas J. Wareham[86], Sylvia Wassertheil-Smoller[141], James G. Wilson[142], Kerri L. Wiggins[109], Qiong Yang[47], Salim Yusuf[15], AFGen Consortium[143], Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium[143], International Genomics of Blood Pressure (iGEN-BP) Consortium[143], INVENT Consortium[143], STARNET[143], Joshua C. Bis[109], Tomi Pastinen[144], Arno Ruusalepp[145,146,147], Eric E. Schadt[148], Simon Koplev[148], Johan L. M. Björkegren[148,149,150,151], Veronica Codoni[152,153], Mete Civelek[124,154], Nicholas L. Smith[65,155,156], David A. Tregouet[152,153], Ingrid E. Christophersen[54,157,158], Carolina Roselli[54], Steven A. Lubitz[54,157], Patrick T. Ellinor[54,157], E. Shyong Tai[159], Jaspal S. Kooner[160], Norihiro Kato[161], Jiang He[162], Pim van der Harst[163], Paul Elliott[164], John C. Chambers[165,166], Fumihiko Takeuchi[161], Andrew D. Johnson[48,167], BioBank Japan Cooperative Hospital Group[143], COMPASS Consortium[143], EPiC-CVD Consortium[143], EPiC-interAct Consortium[143], International Stroke Genetics Consortium (ISGC)[143], METASTROKE Consortium[143], Neurology Working Group of the CHARGE Consortium[143], NiNDS Stroke Genetics Network (SiGN)[143], UK Young Lacunar DNA Study[143], MEGASTROKE Consortium[168], Dharambir K. Sanghera[125,169,170], Olle Melander[19], Christina Jern[171], Daniel Strbian[172,173], Israel Fernandez-Cadenas[30,31], W. T. Longstreth Jr[65,174], Arndt Rolfs[175], Jun Hata[107], Daniel Woo[82], Jonathan Rosand[12,13,14], Guillaume Pare[15], Jemma C. Hopewell[176], Danish Saleheen[177], 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,*]

Affiliations

[1]Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany [2]Centre for Brain Research, Indian Institute of Science, Bangalore, India [3]INSERM U1219 Bordeaux Population Health Research Center, University of Bordeaux, France [4]Stroke Research Group, Division of Clinical Neurosciences, University of Cambridge, Cambridge, UK [5]Department of Neurology, Institute for Neurodegenerative Disease, Bordeaux University Hospital, Bordeaux, France [6]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan [7]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan [8]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan [9]Department of Neurology, Massachusetts General Hospital (MGH), Harvard Medical School, Boston, MA, USA [10]Laboratory of Experimental Cardiology, Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands [11]deCODE genetics/AMGEN Inc., Reykjavik, Iceland [12]Center for Genomic Medicine, MGH, Boston, MA, USA [13]J. Philip Kistler Stroke Research Center, Department of Neurology, MGH, Boston, MA, USA [14]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA [15]Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada [16]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands [17]Department of Radiology and Nuclear Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands [18]Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan [19]Department of Clinical Sciences, Lund University, Malmö, Sweden [20]INSERM, Institut Pasteur de Lille, LabEx DISTALZ-UMR1167, Risk Factors and Molecular Determinants of Aging-Related Diseases, Université Lille, Lille, France [21]Centre Hospitalier Universite Lille, Epidemiology and Public Health Department, Lille, France [22]AA Martinos Center for Biomedical Imaging, Department of Radiology, MGH, Harvard Medical School, Boston, MA, USA [23]Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of Washington, Seattle, WA, USA [24]Division of Neurology, Faculty of Medicine, Brain Research Center, University of British Columbia, Vancouver, British Columbia, Canada [25]School of Life Science, University of Lincoln, Lincoln, UK [26]Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milan, Italy [27]Department of Neurology, Mayo Clinic Rochester, Rochester, MN, USA [28]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [29]National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK [30]Neurovascular Research Laboratory, Vall d’Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona, Vall d’Hebrón Hospital, Barcelona, Spain [31]Stroke Pharmacogenomics and Genetics, Fundacio Docència i Recerca MutuaTerrassa, Terrassa, Spain [32]Children’s Research Institute, Children’s National Medical Center, Washington,

DC, USA [33]Center for Translational Science, George Washington University, Washington, DC, USA [34]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA [35]Harvard Medical School, Boston, MA, USA [36]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA [37]Department of Neurology, University of Maryland School of Medicine and Baltimore VAMC, Baltimore, MD, USA [38]Departments of Medicine, Pediatrics and Population Health Science, University of Mississippi Medical Center, Jackson, MS, USA [39]Institute of Cardiovascular Research, Royal Holloway University of London, London, UK, and Ashford and St Peters Hospital, Surrey, UK [40]Department of Psychiatry, Hope Center Program on Protein Aggregation and Neurodegeneration (HPAN), Washington University School of Medicine, St. Louis, MO, USA [41]Department of Developmental Biology, Washington University School of Medicine, St. Louis, MO, USA [42]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [43]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK [44]British Heart Foundation, Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, UK [45]Department of Medical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands [46]Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands [47]Boston University School of Public Health, Boston, MA, USA [48]Framingham Heart Study, Framingham, MA, USA [49]Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden [50]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA [51]Department of Neurology and Stroke Center, Basel University Hospital, Basel, Switzerland [52]Neurorehabilitation Unit, University of Basel and University Center for Medicine of Aging and Rehabilitation Basel, Felix Platter Hospital, Basel, Switzerland [53]Department of Neurology, Yale University School of Medicine, New Haven, CT, USA [54]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA [55]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA [56]Neuroscience Institute, SF Medical Center, Trenton, NJ, USA [57]Icelandic Heart Association Research Institute, Kopavogur, Iceland [58]Faculty of Medicine, University of Iceland, Reykjavik, Iceland [59]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden [60]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA [61]Laboratory of Epidemiology and Population Science, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA [62]Department of Neurology, Leeds General Infirmary, Leeds Teaching Hospitals NHS Trust, Leeds, UK [63]National Institute for Health and Welfare, Helsinki, Finland [64]FIMM-Institute for Molecular Medicine Finland, Helsinki, Finland [65]Department of Epidemiology, University of Washington, Seattle, WA, USA [66]Public Health Stream, Hunter Medical Research Institute, New Lambton, New South Wales, Australia [67]Faculty of Health and Medicine, University of Newcastle, Newcastle, New South Wales, Australia

[68]School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA [69]Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA [70]Aflac Cancer and Blood Disorder Center, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA [71]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA [72]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden [73]Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA [74]Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA [75]Neurovascular Research Group (NEUVAS), Neurology Department, Institut Hospital del Mar d'Investigació Mèdica, Universitat Autònoma de Barcelona, Barcelona, Spain [76]Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL, USA [77]Division of Cardiovascular Medicine, College of Medicine, University of Florida, Gainesville, FL, USA [78]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands [79]Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, USA [80]Department of Biology, East Carolina University, Greenville, NC, USA [81]Center for Health Disparities, East Carolina University, Greenville, NC, USA [82]University of Cincinnati College of Medicine, Cincinnati, OH, USA [83]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan [84]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA [85]Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA [86]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge, UK [87]Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA [88]Department of Neurology, Radiology, and Biomedical Engineering, Washington University School of Medicine, St. Louis, MO, USA [89]Department of Neurosciences, Experimental Neurology, KU Leuven-University of Leuven, Leuven, Belgium [90]VIB Center for Brain & Disease Research, University Hospitals Leuven, Department of Neurology, Leuven, Belgium [91]INSERM U 1171, CHU Lille, Universite Lille, Lille, France [92]Department of Medical and Molecular Genetics, King's College London, London, UK [93]SGDP Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK [94]Northern Institute for Cancer Research, Newcastle University, Newcastle, UK [95]Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden [96]Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden [97]Bioinformatics Core Facility, University of Gothenburg, Gothenburg, Sweden [98]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden [99]University of Technology Sydney, Faculty of Health, Ultimo, New South Wales, Australia [100]Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA [101]Department of Neurology, Mayo Clinic, Jacksonville, FL, USA [102]Geriatrics Research and Education Clinical

Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA [103]Division of Geriatrics, School of Medicine, University of Mississippi Medical Center, Jackson, MS, USA [104]Memory Impairment and Neurodegenerative Dementia Center, University of Mississippi Medical Center, Jackson, MS, USA [105]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA [106]Data Tecnica International, Glen Echo, MD, USA [107]Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan [108]Clinical Research Facility, Department of Medicine, NUI Galway, Galway, Ireland [109]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA [110]Department of Epidemiology, University of Washington, Seattle, WA, USA [111]Department of Health Services, University of Washington, Seattle, WA, USA [112]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA [113]Brain Center Rudolf Magnus, Department of Neurology, University Medical Center Utrecht, Utrecht, the Netherlands [114]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK [115]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK [116]Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA, USA [117]Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA [118]Department of Biostatistics, University of Washington, Seattle, WA, USA [119]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK [120]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA [121]Division of Genomic Outcomes, Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA [122]Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA [123]Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan [124]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA [125]Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA [126]Department of Neurology, Medical University of Graz, Graz, Austria [127]Institute for Community Medicine, SHIP-KEF, University Medicine Greifswald, Greifswald, Germany [128]Department of Neurology, University Medicine Greifswald, Greifswald, Germany [129]Department of Neurology, Jagiellonian University, Krakow, Poland [130]Department of Neurology, Justus Liebig University, Giessen, Germany [131]Department of Clinical Neurosciences/Neurology, Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden [132]Sahlgrenska University Hospital, Gothenburg, Sweden [133]Stroke Division, Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Victoria, Australia [134]Austin Health, Department of Neurology, Heidelberg, Victoria, Australia [135]Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands [136]INSERM, U1219 Bordeaux, France [137]Department of Public Health, Bordeaux University Hospital, Bordeaux, France [138]Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the

Netherlands [139]Center for Medical Systems Biology, Leiden, the Netherlands [140]School of Medicine, Dentistry and Nursing at the University of Glasgow, Glasgow, UK [141]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY, USA [142]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA [143]A list of members and affiliations appears in the Supplementary Note [144]Department of Human Genetics, McGill University, Montreal, Quebec, Canada [145]Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, University of Tartu, Tartu, Estonia [146]Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia [147]Clinical Gene Networks AB, Stockholm, Sweden [148]Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology Icahn School of Medicine at Mount Sinai, New York, NY, USA [149]Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, Biomeedikum, University of Tartu, Tartu, Estonia [150]Integrated Cardio Metabolic Centre, Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge, Sweden [151]Clinical Gene Networks AB, Stockholm, Sweden [152]UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Sorbonne Universités, Paris, France [153]I CAN Institute for Cardiometabolism and Nutrition, Paris, France [154]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA [155]Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA [156]Seattle Epidemiologic Research and Information Center, VA Office of Research and Development, Seattle, WA, USA [157]Cardiovascular Research Center, MGH, Boston, MA, USA [158]Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust, Gjettum, Norway [159]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore [160]National Heart and Lung Institute, Imperial College London, London, UK [161]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan [162]Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA [163]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands [164]MRC-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 [165]Department of Epidemiology and Biostatistics, Imperial College London, London, UK [166]Department of Cardiology, Ealing Hospital NHS Trust, Southall, UK [167]National Heart, Lung and Blood Research Institute, Division of Intramural Research, Population Sciences Branch, Framingham, MA, USA [169]Department of Pharmaceutical Sciences, College of Pharmacy, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA [170]Oklahoma Center for Neuroscience, Oklahoma City, OK, USA [171]Department of Pathology and Genetics, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden [172]Department of Neurology, Helsinki University Hospital, Helsinki, Finland

[173]Clinical Neurosciences, Neurology, University of Helsinki, Helsinki, Finland
[174]Department of Neurology, University of Washington, Seattle, WA, USA
[175]Albrecht Kossel Institute, University Clinic of Rostock, Rostock, Germany
[176]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK [177]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[178]Faculty of Medicine, University of Iceland, Reykjavik, Iceland [179]Departments of Neurology and Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA [180]Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, San Antonio, TX, USA [181]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA [182]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan [183]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany [184]German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

Acknowledgments

A full list of Acknowledgements appears in the Supplementary Note.

MEGASTROKE Consortium

Rainer Malik[1], Ganesh Chauhan[2], Matthew Traylor[3], Muralidharan Sargurupremraj[4, 5], Yukinori Okada[6, 7, 8], Aniket Mishra[4, 5], Loes Rutten-Jacobs[3], Anne-Katrin Giese[9], sander W. van der Laan[10], Solveig Gretarsdottir[11], Christopher D. Anderson[12, 13, 14], Michael Chong[15], Hieab H. H. Adams[16, 17], Tetsuro Ago[18], Peter Almgren[19], Philippe Amouyel[20, 21], Hakan Ay[13, 22], Traci M. Bartz[23], Oscar R. Benavente[24], Steve Bevan[25], Giorgio B. Boncoraglio[26], Robert D. Brown Jr[27], Adam S. Butterworth[28, 29], Caty Carrera[30, 31], Cara L. Carty[32, 33], Daniel i. Chasman[34, 35], Wei-Min Chen[36], John W. Cole[37], Adolfo Correa[38], ioana Cotlarciuc[39], Carlos Cruchaga[40, 41], John Danesh[28, 42, 43, 44], Paul i. W. de Bakker[45, 46], Anita L. DeStefano[47, 48], Marcel den Hoed[49], Qing Duan[50], Stefan T. Engelter[51, 52], Guido J. Falcone[53, 54], Rebecca F. Gottesman[55], Raji P. Grewal[56], Vilmundur Gudnason[57, 58], Stefan Gustafsson[59], Jeffrey Haessler[60], Tamara B. Harris[61], Ahamad Hassan[62], Aki S. Havulinna[63, 64], Susan R. Heckbert[65], Elizabeth G. Holliday[66, 67], George Howard[68], Fang-Chi Hsu[69], Hyacinth i. Hyacinth[70], M. Arfan ikram[16], Erik ingelsson[71, 72], Marguerite R. irvin[73], Xueqiu Jian[74], Jordi Jimenez-Conde[75], Julie A. Johnson[76, 77], J. Wouter Jukema[78], Masahiro Kanai[6, 7, 79], Keith L. Keene[80, 81], Brett M. Kissela[82], Dawn O. Kleindorfer[82], Charles Kooperberg[60], Michiaki Kubo[83], Leslie A. Lange[84], Carl D. Langefeld[85], Claudia Langenberg[86], Lenore J. Launer[87], Jin-Moo Lee[88], Robin Lemmens[89, 90], Didier Leys[91], Cathryn M. Lewis[92, 93], Wei-Yu Lin[28, 94], Arne G. Lindgren[95, 96], Erik Lorentzen[97], Patrik K. Magnusson[98], Jane Maguire[99], Ani Manichaikul[36], Patrick F. McArdle[100], James F. Meschia[101], Braxton D. Mitchell[100, 102], Thomas H. Mosley[103, 104], Michael A. Nalls[105, 106], Toshiharu Ninomiya[107], 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. Rexrode[117], Kenneth Rice[118], Stephen S. Rich[36], Paul M. Ridker[34,35], Natalia S. Rost[9,13], Peter M. Rothwell[119], Jerome i. Rotter[120,121], Tatjana Rundek[122], Ralph L. Sacco[122], Saori Sakaue[7,123], Michele M. Sale[124], Veikko Salomaa[63], Bishwa R. Sapkota[125], Reinhold Schmidt[126], Carsten O. Schmidt[127], Ulf Schminke[128], Pankaj Sharma[39], Agnieszka Slowik[129], Cathie L. M. Sudlow[114,115], Christian Tanislav[130], Turgut Tatlisumak[131,132], Kent D. Taylor[120,121], Vincent N. S. Thijs[133,134], Gudmar Thorleifsson[11], Unnur Thorsteinsdottir[11], Steffen Tiedt[1], Stella Trompet[135], Christophe Tzourio[5,136,137], Cornelia M. van Duijn[138,139], Matthew Walters[140], Nicholas J. Wareham[86], Sylvia Wassertheil-Smoller[141], James G. Wilson[142], Kerri L. Wiggins[109], Qiong Yang[47], Salim Yusuf[15], Najaf Amin[16], Hugo S. Aparicio[48,187], Donna K. Arnett[188], John Attia[189], Alexa S. Beiser[47,48], Claudine Berr[190], Julie E. Buring[34,35], Mariana Bustamante[191], Valeria Caso[192], Yu-Ching Cheng[193], Seung Hoan Choi[48,194], Ayesha Chowhan[48,187], Natalia Cullell[31], Jean-Franęois Dartigues[195,196], Hossein Delavaran[95,96], Pilar Delgado[197], Marcus Dörr[198,199], Gunnar Engström[19], ian Ford[200], Wander S. Gurpreet[201], Anders Hamsten[202,203], Laura Heitsch[204], Atsushi Hozawa[205], Laura ibanez[206], Andreea ilinca[95,96], Martin ingelsson[207], Motoki iwasaki[208], Rebecca D. Jackson[209], Katarina Jood[210], Pekka Jousilahti[63], Sara Kaffashian[3], Lalit Kalra[211], Masahiro Kamouchi[212], Takanari Kitazono[213], Olafur Kjartansson[214], Manja Kloss[215], Peter J. Koudstaal[216], Jerzy Krupinski[217], Daniel L. Labovitz[218], Cathy C. Laurie[118], Christopher R. Levi[219], Linxin Li[220], Lars Lind[221], Cecilia M. Lindgren[222,223], Vasileios Lioutas[48,224], Yong Mei Liu[225], Oscar L. Lopez[226], Hirata Makoto[227], Nicolas Martinez-Majander[172], Koichi Matsuda[227], Naoko Minegishi[205], Joan Montaner[228], Andrew P. Morris[229,230], Elena Muiño[31], Martina Müller-Nurasyid[231,232,233], Bo Norrving[95,96], Soichi Ogishima[205], Eugenio A. Parati[234], Leema Reddy Peddareddygari[56], Nancy L. Pedersen[98,235], Joanna Pera[129], Markus Perola[63,236], Alessandro Pezzini[237], Silvana Pileggi[238], Raquel Rabionet[239], iolanda Riba-Llena[30], Marta Ribases[240], Jose R. Romero[48,187], Jaume Roquer[241,242], Anthony G. Rudd[243,244], Antti-Pekka Sarin[245,246], Ralhan Sarju[201], Chloe Sarnowski[47,48], Makoto Sasaki[247], Claudia L. Satizabal[48,187], Mamoru Satoh[247], Naveed Sattar[248], Norie Sawada[208], Gerli Sibolt[172], Asgeir Sigurdsson[249], Albert Smith[250], Kenji Sobue[247], Carolina Soriano-Tarraga[242], Tara Stanne[251], O. Colin Stine[252], David J. Stott[253], Konstantin Strauch[231,254], Takako Takai[205], Hideo Tanaka[255,256], Kozo Tanno[247], Alexander Teumer[257], Liisa Tomppo[172], Nuria P. Torres-Aguila[31], Emmanuel Touze[258,259], Shoichiro Tsugane[208], Andre G. Uitterlinden[260], Einar M. Valdimarsson[261], Sven J. van der Lee[16], Henry Völzke[257], Kenji Wakai[255], David Weir[262], Stephen R. Williams[263], Charles D. A. Wolfe[243,244], Quenna Wong[118], Huichun Xu[193], Taiki Yamaji[208], Dharambir K. Sanghera[125,169,170], Olle Melander[19], Christina Jern[171], Daniel Strbian[172,173], israel Fernandez-Cadenas[30,31], W. T. Longstreth Jr[65,174], Arndt Rolfs[175], Jun Hata[107], Daniel Woo[82], Jonathan Rosand[12,13,14], Guillaume Pare[15], Jemma C. Hopewell[176], Danish Saleheen[177], Kari Stefansson[11,178], Bradford B. Worrall[179], Steven J. Kittner[37], Sudha Seshadri[48,180], Myriam Fornage[74,181], Hugh S. Markus[3], Joanna M. M. Howson[28], Yoichiro Kamatani[6,182], Stephanie Debette[4,5] and Martin Dichgans[1,183,184]

[187]Boston University School of Medicine, Boston, MA, USA. [188]University of Kentucky College of Public Health, Lexington, KY, USA. [189]University of Newcastle and Hunter Medical Research Institute, New Lambton, New South Wales, Australia. [190]INSERM, U1061, Universite Montpellier, Montpellier, France. [191]Centre for Research in Environmental Epidemiology, Barcelona, Spain. [192]Department of Neurology, Universita degli Studi di Perugia, Umbria, Italy. [193]Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [194]Broad Institute, Cambridge, MA, USA. [195]Bordeaux Population Health Research Center, INSERM, UMR 1219, Université Bordeaux, Bordeaux, France. [196]Department of Neurology, Memory Clinic, Bordeaux University Hospital, Bordeaux, France. [197]Neurovascular Research Laboratory. Vall d'Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona. Vall d'Hebrón Hospital, Barcelona, Spain. [198]Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. [199]DZHK, Greifswald, Germany. [200]Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK. [201]Hero DMC Heart Institute, Dayanand Medical College & Hospital, Ludhiana, India. [202]Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. [203]Karolinska Institutet, Stockholm, Sweden. [204]Division of Emergency Medicine, and Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA. [205]Tohoku Medical Megabank Organization, Sendai, Japan. [206]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [207]Department of Public Health and Caring Sciences/Geriatrics, Uppsala University, Uppsala, Sweden. [208]Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan. [209]Department of Internal Medicine and the Center for Clinical and Translational Science, Ohio State University, Columbus, OH, USA. [210]Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Goteborg, Sweden. [211]Department of Basic and Clinical Neurosciences, King's College London, London, UK. [212]Department of Health Care Administration and Management, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [213]Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [214]Departments of Neurology & Radiology, Landspitali National University Hospital, Reykjavik, Iceland. [215]Department of Neurology, Heidelberg University Hospital, Heidelberg, Germany. [216]Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands. [217]Hospital Universitari Mutua Terrassa, Terrassa (Barcelona), Spain. [218]Montefiore Medical Center, Albert Einstein College of Medicine, New York, NY, USA. [219]John Hunter Hospital, Hunter Medical Research Institute and University of Newcastle, Newcastle, New South Wales, Australia. [220]Centre for Prevention of Stroke and Dementia, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [221]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [222]Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [223]Wellcome Trust Centre for Human Genetics, Oxford, UK. [224]Beth Israel Deaconess Medical Center, Boston, MA, USA. [225]Wake Forest School of Medicine, Wake Forest, NC, USA. [226]Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA. [227]BioBank Japan, Laboratory of Clinical Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo,

Tokyo, Japan. [228]Neurovascular Research Laboratory, Vall d’Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona, Vall d’Hebrón Hospital, Barcelona, Spain. [229]Department of Biostatistics, University of Liverpool, Liverpool, UK. [230]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [231]Institute of Genetic Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany. [232]Department of Medicine I, Ludwig-Maximilians-Universität, Munich, Germany. [233]DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany. [234]Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milan, Italy. [235]MEB, Karolinska Institutet, Stockholm, Sweden. [236]Estonian Genome Center, University of Tartu, Tartu, Estonia. [237]Department of Clinical and Experimental Sciences, Neurology Clinic, University of Brescia, Brescia, Italy. [238]Translational Genomics Unit, Department of Oncology, IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy. [239]Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain. [240]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addictions, Vall d’Hebron Research Institute (VHIR), Universitat Autonoma de Barcelona, Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Spain. [241]Department of Neurology, IMIM–Hospital del Mar, and Universitat Autonoma de Barcelona, Barceloina, Spain. [242]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [243]National Institute for Health Research Comprehensive Biomedical Research Centre, Guy’s & St. Thomas’ NHS Foundation Trust and King’s College London, London, UK. [244]Division of Health and Social Care Research, King’s College London, London, UK. [245]FIMM-Institute for Molecular Medicine Finland, Helsinki, Finland. [246]THL-National Institute for Health and Welfare, Helsinki, Finland. [247]Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan. [248]BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow, UK. [249]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [250]Icelandic Heart Association, Reykjavik, Iceland. [251]Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Goteborg, Sweden. [252]Department of Epidemiology, University of Maryland School of Medicine, Baltimore, MD, USA. [253]Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow, UK. [254]IBE, Faculty of Medicine, LMU Munich, Munich, Germany. [255]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. [256]Department of Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan. [257]Institute for Community Medicine, SHIP-KEF, University Medicine Greifswald, Greifswald, Germany. [258]Department of Neurology, Caen University Hospital, Caen, France. [259]University of Caen Normandy, Caen, France. [260]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. [261]Landspitali University Hospital, Reykjavik, Iceland. [262]Survey Research Center, University of Michigan, Ann Arbor, MI, USA. [263]Department of Neurology, University of Virginia, Charlottesville, VA, USA.

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, WT.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.I.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.F.-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.F.-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 AgingResearch 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.

References

1. GBD 2015 DALYs and HALE Collaborators. Global regional and national disability-adjusted life-years (DALYs) for 315 diseases injuries and healthy life expectancy (HALE) 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016; 388:1603–1658. [PubMed: 27733283]

2. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality cause-specific mortality for 249 causes of death 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016; 388:1459–1544. [PubMed: 27733281]
3. Gudbjartsson DF, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature. 2007; 448:353–357. [PubMed: 17603472]
4. Gudbjartsson DF, et al. A sequence variant in *ZFHX3* on 16q22 associates with atrial fibrillation and ischemic stroke. Nat Genet. 2009; 41:876–878. [PubMed: 19597491]
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. 2012; 44:328–333. [PubMed: 22306652]
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. 2014; 94:511–521. [PubMed: 24656865]
7. Kilarski LL, et al. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. Neurology. 2014; 83:678–685. [PubMed: 25031287]
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. 2014; 10:e1004469. [PubMed: 25078452]
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. 2016; 15:174–184. [PubMed: 26708676]
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 small vessel disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2016; 15:695–707. [PubMed: 27068588]
11. Malik R, et al. Low-frequency and common genetic variation in ischemic stroke: the METASTROKE collaboration. Neurology. 2016; 86:1217–1226. [PubMed: 26935894]
12. Traylor M, et al. Genetic variation at 16q24.2 is associated with small vessel stroke. Ann Neurol. 2017; 81:383–394. [PubMed: 27997041]
13. Williams FM, et al. Ischemic stroke is associated with the *ABO* locus: the EuroCLOT study. Ann Neurol. 2013; 73:16–31. [PubMed: 23381943]
14. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. Nature. 2015; 526:68–74. [PubMed: 26432245]
15. Morris AP. Transethnic meta-analysis of genomewide association studies. Genet Epidemiol. 2011; 35:809–822. [PubMed: 22125221]
16. Mishra A, Macgregor S. VEGAS2: software for more flexible gene-based testing. Twin Res Hum Genet. 2015; 18:86–91. [PubMed: 25518859]
17. Traylor M, et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. Neurology. 2016; 86:146–153. [PubMed: 26674333]
18. Hara K, et al. Association of *HTRA1* mutations and familial ischemic cerebral small-vessel disease. N Engl J Med. 2009; 360:1729–1739. [PubMed: 19387015]
19. Verdura E, et al. Heterozygous *HTRA1* mutations are associated with autosomal dominant cerebral small vessel disease. Brain. 2015; 138:2347–2358. [PubMed: 26063658]
20. Gould DB, et al. Role of *COL4A1* in small-vessel disease and hemorrhagic stroke. N Engl J Med. 2006; 354:1489–1496. [PubMed: 16598045]
21. Jeanne M, et al. *COL4A2* mutations impair COL4A1 and COL4A2 secretion and cause hemorrhagic stroke. Am J Hum Genet. 2012; 90:91–101. [PubMed: 22209247]
22. Pickrell JK, et al. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet. 2016; 48:709–717. [PubMed: 27182965]
23. Lubitz SA, et al. Independent susceptibility markers for atrial fibrillation on chromosome 4q25. Circulation. 2010; 122:976–984. [PubMed: 20733104]

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. 2015; 47:1282–1293. [PubMed: 26390057]
25. Surakka I, et al. The impact of low-frequency and rare variants on lipid levels. Nat Genet. 2015; 47:589–597. [PubMed: 25961943]
26. Morris AP, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012; 44:981–990. [PubMed: 22885922]
27. Bis JC, 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. 2011; 43:940–947. [PubMed: 21909108]
28. Verhaaren BF, et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. Circ Cardiovasc Genet. 2015; 8:398–409. [PubMed: 25663218]
29. Sinner MF, et al. Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. Circulation. 2014; 130:1225–1235. [PubMed: 25124494]
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. 2015; 96:532–542. [PubMed: 25772935]
31. Nikpay M, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015; 47:1121–1130. [PubMed: 26343387]
32. Ellinor PT, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. Nat Genet. 2012; 44:670–675. [PubMed: 22544366]
33. Bulik-Sullivan B, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015; 47:1236–1241. [PubMed: 26414676]
34. Bulik-Sullivan BK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015; 47:291–295. [PubMed: 25642630]
35. Trynka G, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet. 2013; 45:124–130. [PubMed: 23263488]
36. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. [PubMed: 25693563]
37. Pers TH, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015; 6:5890. [PubMed: 25597830]
38. Geer LY, et al. The NCBI BioSystems database. Nucleic Acids Res. 2010; 38:D492–D496. [PubMed: 19854944]
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. 2017; 20:1–9. [PubMed: 28105966]
40. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet. 2007; 81:208–227. [PubMed: 17668372]
41. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015; 10:1556–1566. [PubMed: 26379229]
42. Wang W, et al. LNK/SH2B3 loss of function promotes atherosclerosis and thrombosis. Circ Res. 2016; 119:e91–e103. [PubMed: 27430239]
43. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 2002; 30:3894–3900. [PubMed: 12202775]
44. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009; 4:1073–1081. [PubMed: 19561590]
45. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348:648–660. [PubMed: 25954001]
46. Eicher JD, et al. GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. Nucleic Acids Res. 2015; 43:D799–D804. [PubMed: 25428361]
47. Leslie R, O’Donnell CJ, Johnson AD. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. Bioinformatics. 2014; 30:i185–i194. [PubMed: 24931982]

48. Higasa K, et al. Human genetic variation database, a reference database of genetic variations in the Japanese population. J Hum Genet. 2016; 61:547–553. [PubMed: 26911352]
49. Bonder MJ, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017; 49:131–138. [PubMed: 27918535]
50. Adams D, et al. BLUEPRINT to decode the epigenetic signature written in blood. Nat Biotechnol. 2012; 30:224–226. [PubMed: 22398613]
51. Franzen O, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. Science. 2016; 353:827–830. [PubMed: 27540175]
52. Erbilgin A, et al. Identification of CAD candidate genes in GWAS loci and their expression in vascular cells. J Lipid Res. 2013; 54:1894–1905. [PubMed: 23667179]
53. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol. 1989; 129:687–702. [PubMed: 2646917]
54. Suhre K, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017; 8:14357. [PubMed: 28240269]
55. Brænne I, et al. Prediction of causal candidate genes in coronary artery disease loci. Arterioscler Thromb Vasc Biol. 2015; 35:2207–2217. [PubMed: 26293461]
56. Flister MJ, et al. Identifying multiple causative genes at a single GWAS locus. Genome Res. 2013; 23:1996–2002. [PubMed: 24006081]
57. Okada Y, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014; 506:376–381. [PubMed: 24390342]
58. Kemp JP, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. Nat Genet. 2017; 49:1468–1475. [PubMed: 28869591]
59. Li Y, Kellis M. Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. Nucleic Acids Res. 2016; 44:e144. [PubMed: 27407109]
60. Lee BK, et al. Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. Genome Res. 2012; 22:9–24. [PubMed: 22090374]
61. Sanseau P, et al. Use of genome-wide association studies for drug repositioning. Nat Biotechnol. 2012; 30:317–320. [PubMed: 22491277]
62. Nelson MR, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015; 47:856–860. [PubMed: 26121088]
63. den Hoed M, et al. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. Nat Genet. 2013; 45:621–631. [PubMed: 23583979]
64. Pfeufer A, et al. Genome-wide association study of PR interval. Nat Genet. 2010; 42:153–159. [PubMed: 20062060]
65. Christophersen IE, et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. Nat Genet. 2017; 49:946–952. [PubMed: 28416818]
66. Verweij N, et al. Genetic determinants of P wave duration and PR segment. Circ Cardiovasc Genet. 2014; 7:475–481. [PubMed: 24850809]
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. 2008; 105:15617–15622. [PubMed: 18832177]
68. Schott JJ, et al. Congenital heart disease caused by mutations in the transcription factor NKX2-5. Science. 1998; 281:108–111. [PubMed: 9651244]
69. Ellesoe SG, 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. 2016; 11:283–290. [PubMed: 26679770]
70. Mohler PJ, et al. Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. Nature. 2003; 421:634–639. [PubMed: 12571597]
71. Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. Am J Hum Genet. 2016; 99:139–153. [PubMed: 27346688]

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. 2011; 43:531–538. [PubMed: 21572416]
73. Surendran P, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat Genet. 2016; 48:1151–1161. [PubMed: 27618447]
74. Winkler TW, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc. 2014; 9:1192–1212. [PubMed: 24762786]
75. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]
76. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol. 2011; 35:310–317. [PubMed: 21374718]
77. Feigin VL, Lawes CM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. Lancet Neurol. 2003; 2:43–53. [PubMed: 12849300]
78. Yang J, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012; 44:369–375. S1–S3. [PubMed: 22426310]
79. Liu JZ, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010; 87:139–145. [PubMed: 20598278]
80. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015; 44:512–525. [PubMed: 26050253]
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. 2011; 478:103–109. [PubMed: 21909115]
82. Wain LV, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat Genet. 2011; 43:1005–1011. [PubMed: 21909110]
83. Wellcome Trust Case Control Consortium. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet. 2012; 44:1294–1301. [PubMed: 23104008]
84. Johnson AD, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24:2938–2939. [PubMed: 18974171]
85. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics. 2015; 31:1334–1336. [PubMed: 25431330]
86. Ward LD, 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. 2016; 44:D1, D877–D881. [PubMed: 26740669]
87. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
88. Wishart DS, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006; 34:D668–D672. [PubMed: 16381955]
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. 2016; 44:D1069–D1074. [PubMed: 26578601]
90. Hachiya T, et al. Genetic predisposition to ischemic stroke: a polygenic risk score. Stroke. 2017; 48:253–258. [PubMed: 28034966]
91. Kanai, M., et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet. 2018. https://doi.org/10.1038/s41588-018-0047-6

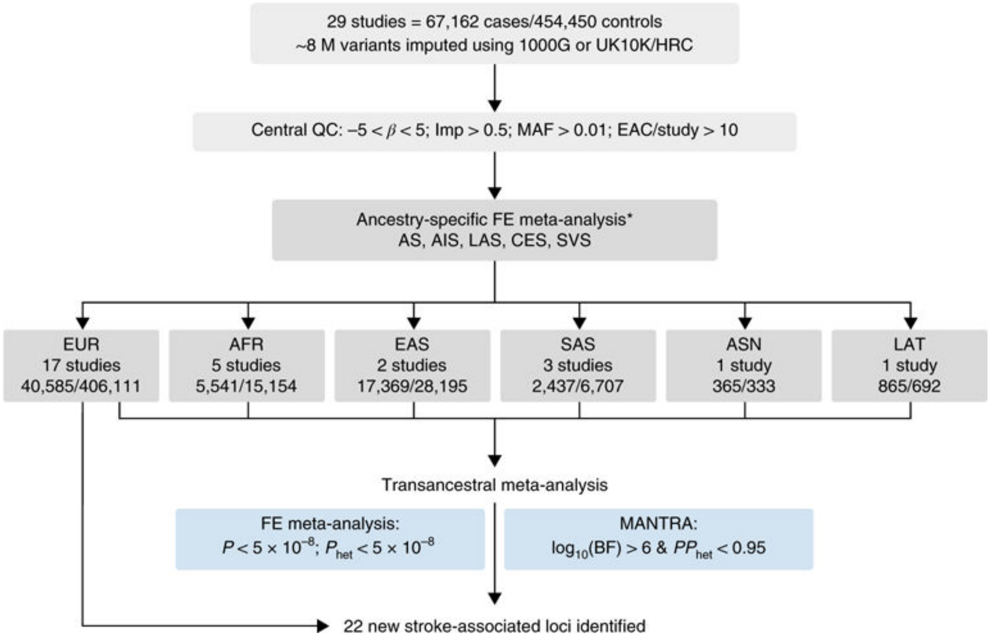

Fig. 1. MEGASTROKE study design

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.

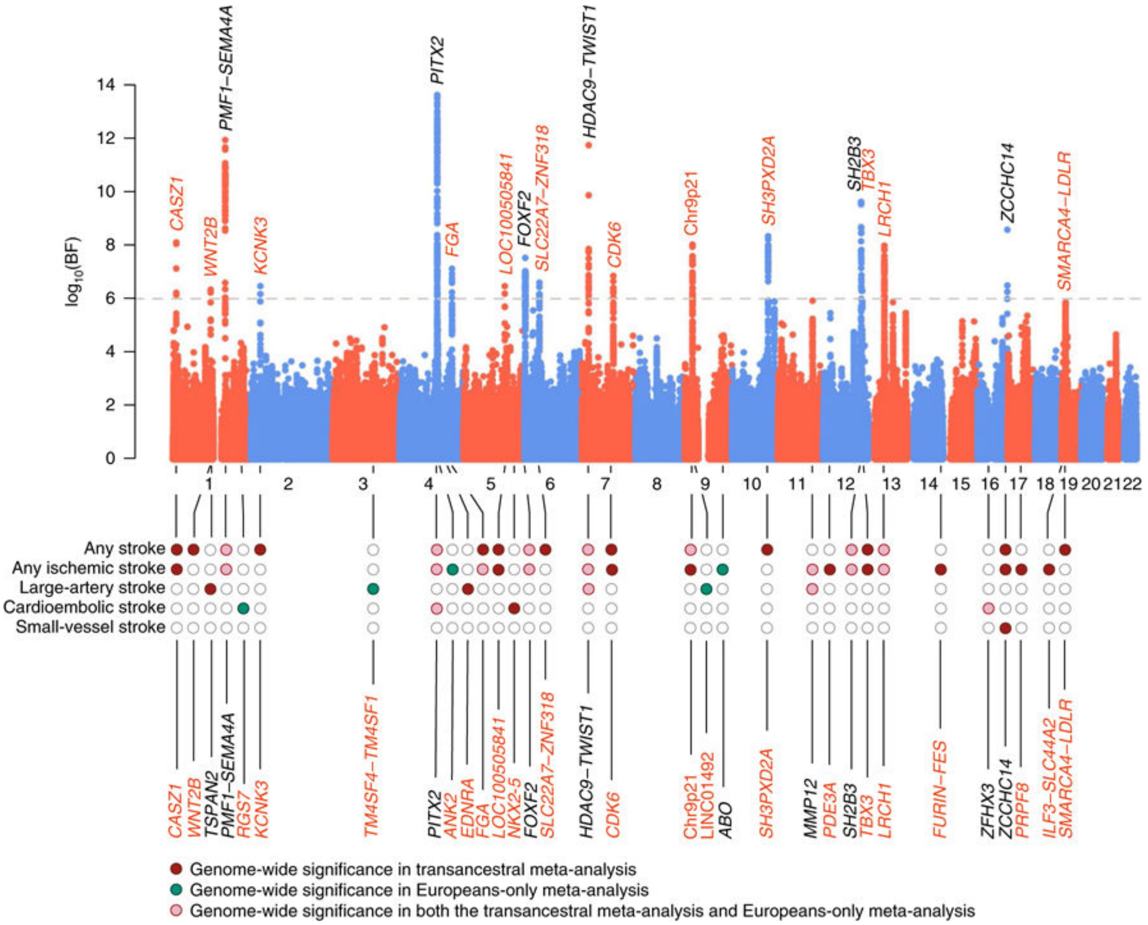

Fig. 2. Association results of the transancestral GWAS meta-analysis and the prespecified ancestry-specific meta-analysis in European samples

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 ($\log_{10}$(BF) >6.0).

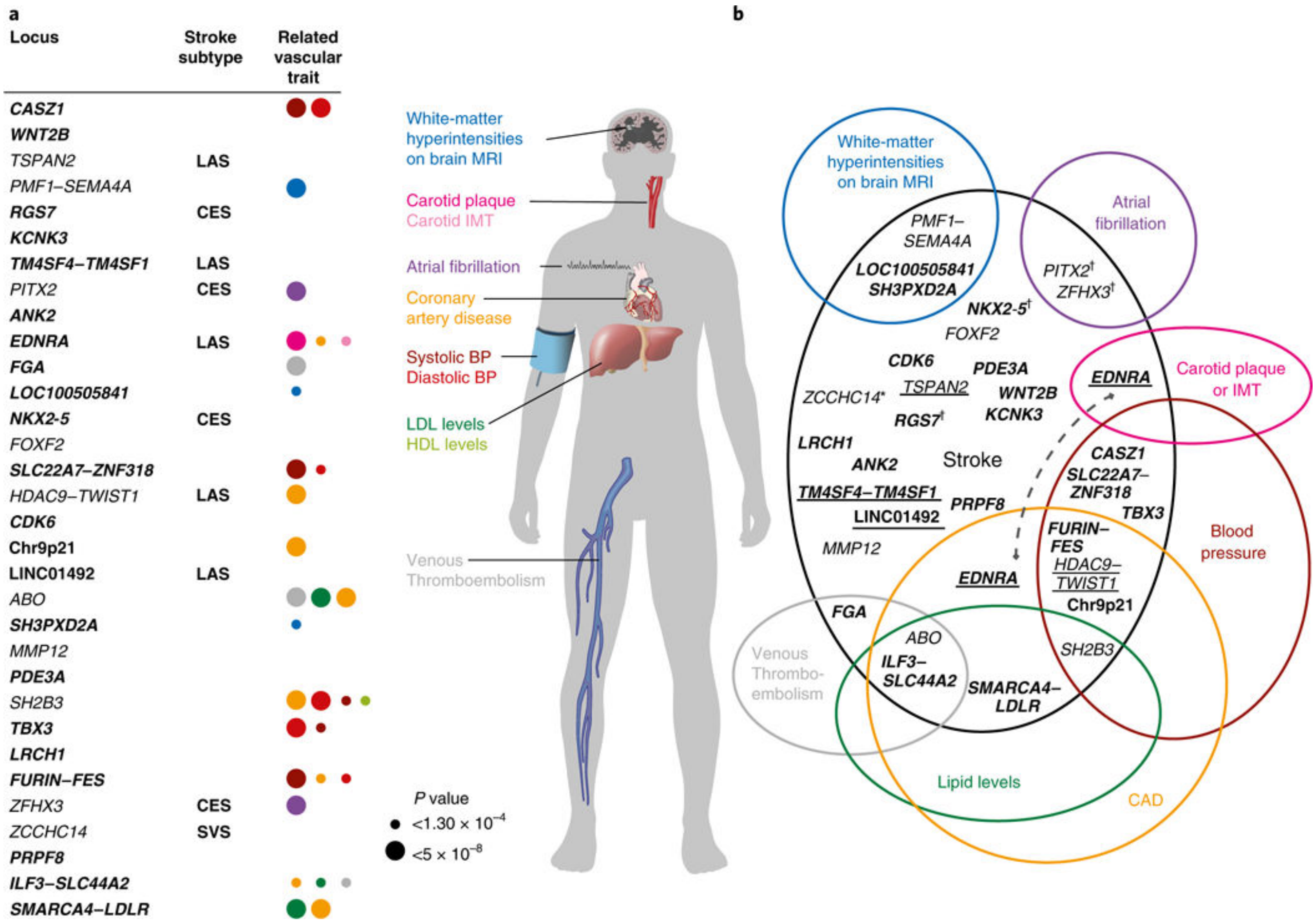

Fig. 3. Genetic overlap between stroke and related vascular traits at the 32 genome-wide-significant loci for stroke

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).

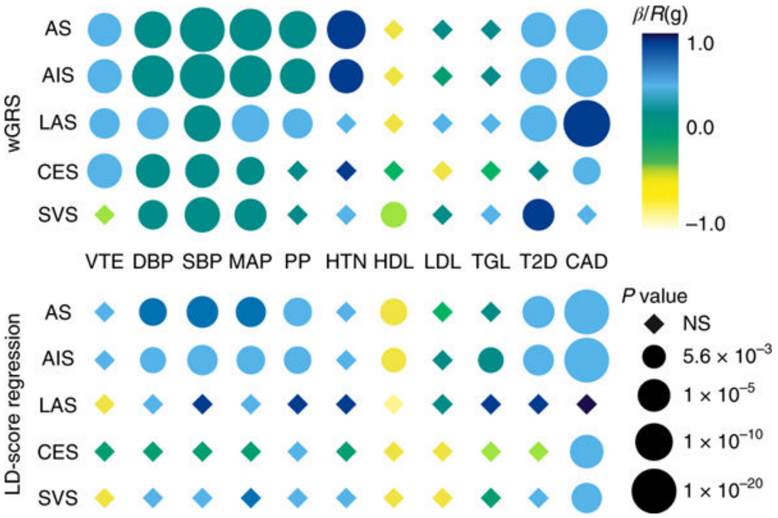

Fig. 4. Shared genetic contribution between stroke and related vascular traits

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.

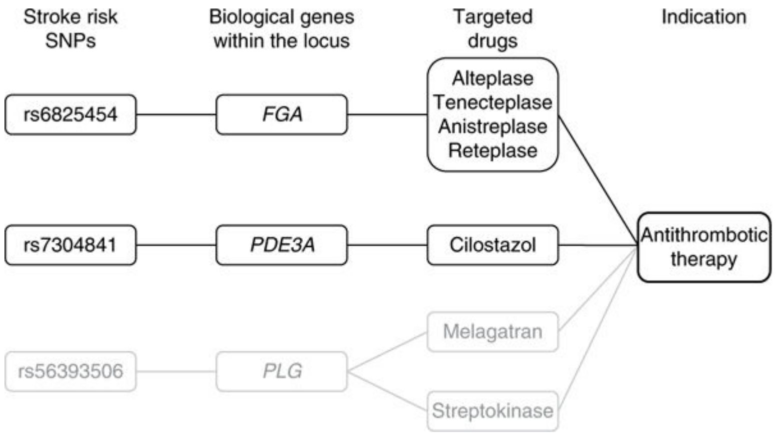

Fig. 5. Connection between stroke risk genes and approved drugs for antithrombotic therapy

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 $\log_{10}(BF) > 5$) are shown in gray.

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	*TM4SF4–TM4Sn*	Intergenic	T/C	1	LAS	EUR	2.33	1.74–3.12	1.44×10^{-8}	NA[b]
rs34311906	4q25	*ANK2*	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	*FGA*	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	*NKX2-5*	Intergenic	A/G	35	CES	TRANS	1.11	1.07–1.16	5.82×10^{-9}	6.96
rs16896398	6p21	*SLC22A7–ZNF318*	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	*SH3PXD2A*	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	*TBX3*	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	*FURIN-FES*	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	*SMARCA4-LDLR*	Intergenic	T/C	65	AS	TRANS	1.05	1.03–1.07	3.40×10^{-8}	5.85
Previously known associations											
rs12124533	1p13	*TSPAN2*	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	*PITX2*	Intergenic	T/G	28	CES	TRANS	1.32	1.27–1.37	1.86×10^{-47}	45.10

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)
rs4959130	6p25	*FOXF2*	Intergenic	A/G	14	AS	TRANS	1.08	1.05–1.11	1.42×10^{-9}	7.52
rs2107595	7p21	*HDAC9–TWIST1*	Intergenic	A/G	24	LAS	TRANS	1.21	1.15–1.26	3.65×10^{-15}	12.99
rs635634	9q34	*ABO*	Intergenic	T/C	19	AIS	EUR	1.08	1.05–1.11	9.18×10^{-9}	4.99
rs2005108	11q22	*MMP12*	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	*ZCCHC14*	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 *SMARCA–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 ($PP_{\text{het}} = 0.96$) and were thus excluded.