Genome-wide association study of intracranial aneurysms reveals 17 risk loci, polygenic architecture, genetic overlap with clinical risk factors, and opportunities for prevention.

Mark K. Bakker[1*], Rick A.A. van der Spek[1], Wouter van Rheenen[1], Sandrine Morel[2,3], Romain Bourcier[4,5], Isabel C. Hostettler[6,7], Varinder S. Alg[8], Kristel R. van Eijk[1], Masaru Koido[9,10], Masato Akiyama[9,11,12], Chikashi Terao[9], Koichi Matsuda[13], Robin G. Walters[14], Kuang Lin[14], Liming Li[15], Iona Y. Millwood[14,16], Zhengming Chen[14], Guy A. Rouleau[17], Sirui Zhou[18], Kristiina Rannikmae[19], Cathie L.M. Sudlow[19,20], Henry Houlden[21], Leonard H. van den Berg[1], Christian Dina[4], Olivier Naggara[22,23], Jean-Christophe Gentric[24], Eimad Shotar[25], Francois Eugene[26], Hubert Desal[4,5], Bendik S. Winsvold[27,28], Sigrid Borte[28,29,30], Marianne Bakke Johnsen[28,29,30], Ben M. Brumpton[28], Marie Softeland Sandvei[31,32], Cristen J. Willer[33], Kristian Hveem[28,34], John-Anker Zwart[27,28,30], W. M. Monique Verschuren[35,36], Christoph M. Friedrich[37], Sven Hirsch[38], Sabine Schilling[38], Jerome Dauvillier[39], Olivier Martin[39], HUNT All-In Stroke[40], China Kadoorie Biobank Collaborative Group[40], BioBank Japan Project Consortium[40], The ICAN Study Group[40], CADISP Group[40], Genetics and Observational Subarachnoid Haemorrhage (GOSH) Study investigators[40], International Stroke Genetics Consortium (ISGC)[40], Gregory T. Jones[41], Matthew J. Bown[42,43], Nerissa U. Ko[44], Helen Kim[45,46,47], Jonathan R.I. Coleman[48,49], Gerome Breen[48,49], Jonathan G. Zaroff[50], Catharina J.M. Klijn[51], Rainer Malik[52], Martin Dichgans[52,53,54], Muralidharan Sargurupremaj[55,56], Turgut Tatlisumak[57], Philippe Amouyel[58], Stephanie Debette[55,56], Gabriel J.E. Rinkel[1], Bradford B. Worrall[59], Joanna Pera[60], Agnieszka Slowik[60], Emilia I. Gaal-Paavola[61,62], Mika Niemela[61], Juha E. Jaaskelainen[63,64], Mikael von Und Zu Fraunberg[63,64], Antti Lindgren[63,64], Joseph P. Broderick[65], David J. Werring[6], Daniel

Woo[65], Richard Redon[4], Philippe Bijlenga[3], Yoichiro Kamatani[13], Jan H. Veldink[1†], Ynte M. Ruigrok[1†*].

1 Department of Neurology and Neurosurgery, University Medical Center Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands.

2 Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, Geneva, Switzerland.

3 Neurosurgery Division, Department of Clinical Neurosciences, Faculty of Medicine, Geneva University Hospitals, Geneva, Switzerland.

4 Université de Nantes, CHU Nantes, INSERM, CNRS, l'institut du thorax, Nantes, France.

5 CHU Nantes, Department of Neuroradiology, Nantes, France.

6 Stroke Research Centre, University College London Queen Square Institute of Neurology, London, UK.

7 Department of Neurosurgery, Klinikum rechts der Isar, Technical University Munich, Munich, Germany.

8 Stroke Research Centre, University College London, Institute of Neurology, London, UK.

9 Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

10 Department of Cancer Biology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

11 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.

12 Department of Ocular Pathology and Imaging Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.

13 Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.

14 Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, U.K.

15 School of Public Health, Peking University Health Science Center, Beijing, China.

16 Medical Research Council Population Health Research Unit, University of Oxford, Oxford, U.K.

17 Montreal Neurological Institute and Hospital, McGill University, Montréal, QC, Canada.

18 Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, QC, Canada.

19 Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK.
20 UK Biobank, Cheadle, Stockport, UK.
21 Neurogenetics Laboratory, The National Hospital of Neurology and Neurosurgery, London, UK.
22 Pediatric Radiology, Necker Hospital for Sick Children, Université Paris Descartes, Paris, France.
23 Department of Neuroradiology, Sainte-Anne Hospital and Université Paris Descartes, INSERM UMR S894, Paris, France.
24 Department of Neuroradiology, University hospital of Brest, Brest, France.
25 Department of Neuroradiology, Pitié-Salpêtrière Hospital, Paris, France.
26 Department of Neuroradiology, University hospital of Rennes, Rennes, France.
27 Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.
28 K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
29 Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.
30 Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway.
31 Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
32 The Cancer Clinic, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.
33 Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, 48109, MI, USA.
34 HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
35 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands.
36 National Institute for Public Health and the Environment, Bilthoven, the Netherlands.
37 University of Applied Science and Arts, Dortmund.
38 Zurich University of Applied Sciences, School of Life Sciences and Facility Management.
39 SIB Swiss Institute of Bioinformatics.

40 A list of members and affiliations appears in the Supplementary Note.

41 Department of Surgery, University of Otago, Dunedin, New Zealand.

42 Department of Cardiovascular Sciences and National Institute for Health Research, University of Leicester, UK.

43 Leicester Biomedical Research Centre University of Leicester Glenfield Hospital Leicester United Kingdom.

44 Department of Neurology, University of California at San Francisco, San Francisco, USA.

45 Department of Anesthesia and Perioperative Care, Center for Cerebrovascular Research, University of California, San Francisco, California, USA.

46 Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA.

47 Institute for Human Genetics, University of California, San Francisco, California, USA.

48 Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, UK.

49 UK National Institute for Health Research (NIHR) Biomedical Research Centre (BRC), South London and Maudsley NHS Foundation Trust, London, SE5 8AF, UK.

50 Division of Research, Kaiser Permanente of Northern California, Oakland, California, USA.

51 Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands.

52 Institute for Stroke and Dementia Research, University Hospital, Ludwig-Maximilians-University, Munich.

53 Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.

54 Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Munich, Germany.

55 INSERM U1219 Bordeaux Population Health Research Center, University of Bordeaux, Bordeaux, France.

56 Department of Neurology, Institute for Neurodegenerative Disease, Bordeaux University Hospital, Bordeaux, France.

57 Department of Clinical Neuroscience at Institute of Neuroscience and Physiology, University of Gothenburg, Sweden.

58 Institut Pasteur de Lille, Inserm U1167 and Lille University Hospital, Lille, France.

59 Departments of Neurology and Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA.

60 Department of Neurology, Faculty of Medicine, Jagiellonian University Medical College, ul. Botaniczna 3, 31-503, Krakow, Poland.

61 Department of Neurosurgery, Helsinki University Hospital, University of Helsinki, Finland.

62 Clinical Neurosciences, University of Helsinki, Topeliuksenkatu 5, 00260, Helsinki, Finland.

63 Neurosurgery NeuroCenter Kuopio University Hospital Kuopio Finland.

64 Institute of Clinical Medicine Faculty of Health Sciences University of Eastern Finland Kuopio Finland.

65 University of Cincinnati College of Medicine, Cincinnati, OH, USA.

†These authors contributed equally

*Corresponding authors: Y.M. Ruigrok. ij.m.ruigrok@umcutrecht.nl and M.K. Bakker. m.k.bakker-25@umcutrecht.nl.

Abstract

An intracranial aneurysm (IA) is present in 3% of the population. Rupture of an IA leads a to subarachnoid hemorrhage, a severe type of stroke. In order to discover new genetic loci and the genetic architecture of IA, we performed a cross-ethnic, genome-wide association study in 10,754 cases and 306,882 controls of European and East Asian ancestry. We discovered 17 risk loci, 11 of which are new. We reveal a polygenic architecture and are able to explain more than half of the disease heritability. We show a high genetic correlation between ruptured and unruptured IAs. A role for endothelial cells in the disease is suggested based on the function of genes mapped to risk loci and on heritability enrichment analysis. Drug target enrichment analysis shows pleiotropic effects between IA and anti-epileptic and sex hormone drugs, which provides insights into IA pathophysiology. Finally, genetic risk for smoking and high blood pressure, the two main clinical risk factors for IA, plays an

important role in IA risk and is the main driver of the genetic correlation between IA and other cerebrovascular traits.

Main

An intracranial aneurysm (IA) is a balloon-shaped dilatation, usually located at a branch of an intracranial artery. It is present in 3% of the population[1]. Rupture of an IA causes an aneurysmal subarachnoid hemorrhage (aSAH), a severe type of stroke. Approximately one third of patients die, and another third remain dependent for daily life activities[2]. IA occurs in relatively young people with a mean age of 50 years and is twice as common in women over 50 years old compared to men of that age. Genetic predisposition plays an important role in the disease with an aSAH heritability of 41%, as estimated in a twin study[3].

Much is still unknown about the genetic architecture of IA[4,5]. Family-based studies identified a number of variants with Mendelian inheritance[6-10], but genome-wide association studies (GWAS) have identified multiple common variants, suggesting a polygenic model of inheritance[5,11-13]. The largest GWAS published to date, involving 2,780 cases and 12,515 controls, identified six risk loci[11,13]. Based on that GWAS, the explained single nucleotide polymorphism (SNP)-based heritability of IA was estimated as being only 4.1-6.1%, depending on population[5].

We aimed to further characterize the genetic architecture of IA by performing a cross-ethnic GWAS meta-analysis on a total of 10,754 cases and 306,882 controls from a wide range of European and East Asian ancestries. We included both cases with unruptured IA and aSAH (i.e. with ruptured IA), enabling us to identify potential risk factors specific for IA rupture. We also looked for genetic similarities between IA and related traits, including other types of stroke, vascular malformations and other aneurysms, and analyzed whether

known risk factors for IA play a causal genetic role. Further, we investigated enrichment of genetic associations in functional genetic regions, tissue subtypes, and drug classes to provide insight into IA pathophysiology.

Results

GWAS of intracranial aneurysms

Our GWAS meta-analysis on IA consisted of two stages. The Stage 1 meta-analysis included all European ancestry individuals and consisted of individual level genotypes from 23 different cohorts, that were merged into nine European ancestry strata, based on genotyping platform and country. These strata were each analyzed in a logistic mixed model[14] and then meta-analyzed, while also including summary statistics from a population-based cohort study: the Nord-Trøndelag Health Study (the HUNT Study). This resulted in 7,495 cases and 71,934 controls and 4,471,083 SNPs passing quality control (QC) thresholds (Online Methods, Supplementary Table 1). Stage 2 was a cross-ethnic meta-analysis including all Stage 1 strata and summary statistics of East Asian individuals from two population-based cohort studies: The Biobank Japan (BBJ) and the China Kadoorie Biobank (CKB). This totaled 10,754 cases and 306,882 controls and 3,527,309 SNPs in Stage 2 (Supplementary Table 1).

The Stage 1 association study resulted in 11 genome-wide significant loci (P-value $\leq 5 \cdot 10^{-8}$, Figure 1, Supplementary Table 2). Transethnic genetic correlation analysis showed a strong correlation between the Stage 1 meta-analysis of European ancestry and an analysis including only East Asian ancestry samples (ρ_g=0.938±0.165, standard error [SE] for genetic impact and 0.908±0.146 for genetic effect, Supplementary Table 3). Stage 2 increased the number of genome-wide significant loci to 17 (Table 1, Figure 1). All but two loci (8q11.23,

rs6997005 and 15q25.1, rs10519203) were also associated with IA in the samples of East Asian ancestry added in Stage 2 (P<0.05/11) and 2 loci were monomorphic in East Asians (Table 1). The Stage 2 loci included 11 novel risk loci and six previously reported risk loci[11]. We used conditional and joint (COJO) analysis to condition the Stage 1 GWAS summary statistics on the lead SNP in each locus. We found that none of the loci consisted of multiple independent SNPs and that each locus tagged a single causal variant (data not shown). Genomic inflation factors ($lambda_{GC}$) were 1.050 for the Stage 1 meta-analysis and 1.065 for Stage 2 (Supplementary Figure 1a-d, Supplementary Table 4). The linkage disequilibrium score regression (LDSR) intercept was 0.957±0.008 (SE) for the Stage 1 meta-analysis and 0.982±0.008 for the East Asian subset. This indicated that in all GWAS analyses, observed inflation was due to polygenic architecture.

Conditioning the Stage 1 GWAS summary statistics on GWAS summary statistics for systolic and diastolic blood pressure (BP, Neale lab summary statistics [URLs]) using multi-trait conditional and joint (mtCOJO) analysis resulted in one additional genome-wide significant locus (rs2616406, P=6.221e-08 in the Stage 1 GWAS, P=4.499e-9 after mtCOJO with BP). mtCOJO with smoking pack-years summary statistics, or including genetic risk scores (GRSs) for smoking (cigarettes per day)[15] or blood pressure related traits[16] did not result in additional loci (data not shown).

Characterization of GWAS loci

An overview of the genic position, alleles, effect size and P-value of the strongest association per locus is shown in Table 1. We used summary statistics-based Mendelian randomization (SMR), co-localization analysis using eCAVIAR, and transcriptome-wide association study (TWAS, URLs) to annotate potential causative genes in these loci (Supplementary Tables 5-9, Supplementary Figure 2). A description of this annotation

process is described in the Supplementary Note. Since SMR, eCAVIAR and TWAS all require LD reference panels, we limited the annotation to the loci identified in the European ancestry Stage 1 GWAS meta-analysis. This resulted in 11 potential causative genes on six unique loci: *SLC22A5/SLC22A4/P4HA2* (chr5), *NT5C2/MARCKSL1P1* (chr10), *FGD6/NR2C1* (chr12), *PSMA4* (chr15) and *BCAR1*/*RP11-252K23.2* (chr16) (Table 1, Supplementary Table 5). Although we did not find evidence for involvement of *SOX17* in the chr8 locus, previous studies did find functional evidence for *SOX17*[17,18]. Therefore, we annotated the chr8 locus as *SOX17*.

In the Stage 2 GWAS, six additional loci were identified: 6q16.1, 10q23.33, 11p15.5, 12p12.2, 12q21.22, and 20p11.23. Due to the combined European and East Asian LD structures, these loci cannot reliably be mapped to genes using the above-mentioned techniques. Of the six additional loci, four have previously been linked to blood pressure, namely 6q16.1 (rs11153071)[19], 10q23.33 (rs11187838)[20], rs11044991 (12p12.2)[21], and rs2681492 (12q21.22)[21,22]. A detailed description of the genes and loci is found in the Supplementary Note.

The potentially causative gene *FGD6*[23] plays a role in angiogenesis and defects may lead to a compromised formation of blood vessels. *FGD6* is a vascular endothelial cell (vEC) signaling gene, involved in stress signaling in vECs[24]. Loss-of-function mutations in *THSD1* and *SOX17,* which both have key roles in vECs, lead to subarachnoid hemorrhage in animal models[7,17,25,26]. *BCAR1* is a ubiquitously expressed gene which protein product is a sensor for mechanical stress[27]. The PSMA4 locus is known for associations with a number of smoking and respiratory system traits[28-31].

Predictors of IA rupture

We assessed whether genetic risk factors differed between ruptured and unruptured IA, using stratified GWAS analysis. The number of cases with unruptured IA was small (N=2070). Therefore, in addition to performing a stratified GWAS on patients with a ruptured aneurysm versus patients with an unruptured IA (aSAH-vs-uIA), we also performed a stratified GWAS on only patients with ruptured IA versus controls (aSAH-only) and a stratified GWAS on only patients with an unruptured IA versus controls (uIA-only) (Supplementary Table 4, Supplementary Figure 1e-j). Overall, 69% of IA cases had a ruptured IA and 28% an unruptured IA while 3.8% had an unknown rupture status. The aSAH-only and uIA-only GWASs identified a number of genome-wide significant loci, all of which reached genome-wide significance in the Stage 1 and 2 GWAS meta-analyses on IA. In the aSAH-vs-uIA GWAS, we found no genome-wide significant loci. Furthermore, genetic correlation analysis showed a high correlation of 0.970±0.133 (SE) between ruptured and unruptured IA (Supplementary Table 3). Together these findings indicate a strong similarity in genetic architecture between ruptured and unruptured IA.

SNP-based heritability

We estimated the SNP-based heritability of IA to be 21.6±2.8% (SE) on the liability scale with LDSR (tool named LDSC[32], URLs) and 29.9±5.4% using SumHer[33] (URLs, Table 2). This corresponds to an explained fraction of the twin-based heritability (h^2=41%[3]) of 53-73% depending on the method used (LDSC or SumHer). We used a life-time risk for unruptured IA of 3%[1] for the conversion to the liability scale. Since this GWAS was an admixture of patients with ruptured and unruptured IA, this prevalence may not be representative of the whole study population. Therefore, we calculated liability scale heritability using a range of life-time risk values (Supplementary Figure 3a). This shows that also when using lower life-

time risk estimates (K), the explained SNP-based heritability is substantial (K=0.02: h^2=19.3±2.5% [LDSC], 26.8±4.8% [SumHer]; K=0.01: 16.3±2.1% [LDSC], 22.6±4.1% [SumHer]).

A substantial SNP-based heritability is also found for ruptured IA (SAH-only, h^2=0.140±0.020) and unruptured IA (uIA-only, h^2=0.223±0.044). The difference between the heritability estimates could suggest differences in genetic architecture, but estimates depend on the prevalence estimate (Supplementary Figure 3b-c), meaning these differences should be interpreted with caution.

Enrichment of genomic regions

To understand the disease mechanisms of IA, we applied several heritability enrichment analyses using LD-score regression (LDSR). Partitioning on functional genomic elements showed a clear enrichment of heritability in regulatory elements, including enhancer and promoter histone marks H3K4me1, H3K27Ac and H3K9Ac, super enhancers, and DNAse I hypersensitivity sites (Figure 2a). Such enrichment of regulatory elements in the genome is also seen in other polygenic traits and indicates that the architecture of IA is also polygenic[34]. Partitioning heritability per chromosome further supported a polygenic architecture as heritability was associated with the number of SNPs on a chromosome (Figure 2b).

Tissue-specific LDSR did not show enrichment for any tissue (Supplementary Tables 10 and 11). We then performed cell-type enrichment analysis using single-cell RNA-sequencing (scRNAseq) reference data derived from mouse brain[35]. No enrichment was found using a scRNAseq dataset of mouse brain blood vessels[36] (Supplementary Table 12). Using a larger dataset defining cell-types in the mouse brain[35], we found enrichment in 'endothelial mural cells', which is a combined set of vascular endothelial and mural cells

(enrichment=2.31±0.41 [SD], $P=1.65\cdot10^{-3}$, Figure 2c), and in midbrain neurons (enrichment=2.23±0.37, $P=6.56\cdot10^{-4}$).

LD-pruned enrichment analysis using GARFIELD showed that genes specific for blood vessels were enriched (Figure 2d, Supplementary Table 13), further supporting the role of promoters and enhancers (Figure 2e).

Causal genetic roles of blood pressure and smoking

To assess which phenotypes causally influence the risk of IA, we performed generalized summary statistics-based Mendelian randomization (GSMR) using summary statistics for all phenotypes available in the UK Biobank (Supplementary Table 14). We used the Stage 1 summary statistics excluding the UK Biobank data as outcome. In this analysis, we chose a stringent value for the multiple testing threshold of 376, which was the number of traits passing the GSMR quality control parameters. Sixteen traits were statistically significant after correction for multiple testing (Figure 3a). All statistically significant traits were related to either smoking or blood pressure (BP), which are the two main clinical risk factors for unruptured IA and aSAH[1,37,38]. To determine whether genetic predisposition for smoking and BP were causal genetic risk factors independent of one another, we conditioned the Stage 1 GWAS summary statistics on GWAS summary statistics for smoking and BP using multi-trait conditional and joint analysis (mtCOJO). We used summary statistics for both systolic BP (SBP) and diastolic BP (DBP) combined to condition on BP and summary statistics for pack-years to condition on smoking (Figure 3a, Supplementary Table 14). All GSMR effects diminished after conditioning on either BP or pack-years, and remained when conditioning on the other risk factor. The mtCOJO method itself did not affect the effect size estimates as conditioning on standing height did not affect the estimates. These findings provide strong

evidence that the genetic predisposition for BP and smoking are independent genetic causes of IA (Figure 3b).

Since the phenotype values of the exposure traits were inverse rank-normalized, the GSMR effect size of SBP (β_{xy} = 1.058±0.187) and pack-years (β_{xy} = 0.973±0.236) cannot easily be interpreted. Therefore, we performed an additional GSMR analysis for BP with an updated version of the UK Biobank GWAS, including raw phenotype values for quantitative traits (Supplementary Table 15). For BP traits, the GSMR analysis resulted in an effect size estimate of 0.095± 0.019 for DBP and 0.047± 0.011 for SBP, meaning an 8-12% increase in IA risk per mmHg increase of DBP and a 3.7-6% increase in IA risk per mmHg increase of SBP, assuming a linear effect of BP on IA liability. In addition, age at high BP diagnosis had a significant GSMR effect ($P= 1.79 \cdot 10^{-4}$, β_{xy}=0.163±0.044), indicating an increase in IA risk of 13-23% for each year of additional high BP exposure. We did not include smoking quantitative traits, because these were not normally distributed (data not shown) and could, therefore, lead to a biased effect estimate.

We then tested whether the effects of smoking and BP were different between ruptured (SAH-only) and unruptured IA (uIA-only, Supplementary Table 16). The GSMR effect sizes followed the same trend for all phenotypes, but 'Hypertension (Self-reported)' had a stronger effect on ruptured IA (SAH-only: b_{xy}=6.74±0.61 [SE], all IA: 2.97±0.42, uIA-only: 2.38±0.70), while amlodipine use had a weaker effect on unruptured IA and became statistically non-significant (uIA-only: b_{xy}=4.77±3.90, P=0.22, all IA: b_{xy}=11.4±2.10, $P=5.25 \cdot 10^{-8}$, SAH-only: b_{xy}=13.1±2.60, $P=5.25 \cdot 10^{-7}$). Although the effect of self-reported hypertension on SAH-only was stronger, conditioning on blood pressure using mtCOJO mitigated the effect (b_{xy}=1.02±0.45, P=0.024, data not shown). Since the power to detect GSMR effects in the uIA-only sample is much lower compared to all IA and SAH-only due to limited sample

size, further investigation is required to make inferences about genetic risk factors for rupture.

Traits influencing female hormones are suggested to play a role in aSAH risk[39]. Only two female hormone-related traits had enough genome-wide significant risk loci to pass GSMR quality control. These were 'age when periods started (menarche)' and 'had menopause'. Neither of these showed a causal relationship with IA in the GSMR analysis (Supplementary Table 14).

Drivers of genetic correlation with vascular traits

To identify traits correlated with IA, we analyzed Stage 1 summary statistics using LDHub[40]. LDHub includes a subset of the summary statistics used for GSMR and a number of summary statistics from publicly available sources. Traits that showed correlations that reached the Bonferroni threshold for multiple testing (p=0.05/464) included several blood pressure (BP)-related traits, including diastolic BP (DBP) (ρ_g=0.223, P=5.40·10^{-9}) and systolic BP (SBP) (ρ_g=0.256, P=1.34·10^{-8}) and smoking traits, such as pack-years (ρ_g=0.330, P=7.87·10^{-8}) (Supplementary Table 17).

We used LDSR to calculate the genetic correlation of IA with other stroke subtypes - ischemic stroke (IS)[41] and intracerebral hemorrhage (ICH) - , with other vascular malformation types - intracranial arteriovenous malformation (AVM)[42] and cervical artery dissection[43] - , and with abdominal aortic aneurysm (AAA)[44]. For IS, a correlation of 0.195±0.079 (P=0.014) was found with IA (Figure 3c, Supplementary Table 3). After conditioning the IA GWAS on either BP or on pack-years, which are clinical risk factors for both IS and IA[1,37,38,45], the correlation was no longer statistically significant and reduced to 0.121±0.081 for BP and 0.147±0.084 for pack-years. The correlation disappeared after conditioning on both risk factors (ρ_g=0.009±0.083, P=0.916). When conditioning on an

unrelated but heritable trait (standing height), the correlation remained (ρ_g=0.238±0.081, P=0.003). No genetic correlation was found for any of the IS subtypes.

We found a statistically significant genetic correlation between IA and ICH (ρ_g= 0.447±0.184, P=0.015), which was mainly driven by deep ICH (ρ_g=0.516±0.198, P=0.009), and not by lobar ICH (P=0.534). After conditioning the IA GWAS on either BP or pack-years, which are also important risk factors for ICH[46], the correlation with deep ICH decreased (ρ_g=0.288±0.189 for BP and 0.234±0.192 for pack-years) and was no longer statistically significant. Conditioning on height had a much smaller effect (ρ_g=0.380±0.196).

A genetic correlation was found between IA and AAA (ρ_g=0.302±0.105, P=0.004). Conditioning on pack-years strongly reduced the correlation between IA and AAA (ρ_g=0.173±0.117, P=0.138), whereas BP did not (ρ_g=0.264±0.117, P=0.024).

There was no genetic correlation between IA and carotid artery dissection (ρ_g=0.151±0.180, P=0.401); whereas for vertebral artery dissection and the combined set of vertebral and carotid artery dissection, a larger, albeit non-statistically significant, estimate was observed (ρ_g=0.281±0.159, P=0.077 and ρ_g=0.174±0.149, P=0.066, respectively) (Supplementary Table 3). For AVM, a negative SNP-based heritability was estimated, which could be due to the small sample size of this GWAS (1,123 cases and 1,935 controls). Therefore, we performed a lookup of all SNPs identified in the Stage 1 and 2 IA GWAS in the summary statistics of the AVM GWAS[42] but were unable to replicate any of these SNP associations (P<0.05/17) (Supplementary Table 18).

Drug target enrichment

To identify pleotropic pathways between IA and other diseases that contain known drug targets, we assessed enrichment in genes targeted by drugs and drug classes[47]. Gene-based P-values were calculated with MAGMA, resulting in 29 genes that passed the Bonferroni

threshold for multiple testing (P<0.05/18106, Supplementary Table 19). The anti-hypertensive drugs ambrisentan and macitentan showed a statistically significant enrichment ($P=1.35\cdot10^{-5}$, Supplementary Table 20) which was driven by a single gene (*EDNRA)*. Drug class enrichment analysis showed that drugs in the classes 'anti-epileptics' were enriched (area under the curve [AUC]=0.675, $P=8\cdot10^{-5}$, Supplementary Table 21). The most statistically significant enriched drugs within this class are blockers of Na^{+} and Ca^{2+} channels, namely phenytoin, zonisamide and topiramate[48] (Supplementary Table 20). These channels are important in blood pressure regulation, as well as in several other biological mechanisms. The other enriched drug class is 'sex hormones + modulators of the genital system' (AUC=0.652, $P=2.02\cdot10^{-4}$). We also used MAGMA to study enrichment in gene pathways, but found no statistically significant results (Supplementary Table 22).

Discussion

We identified 11 novel risk loci for IA and confirmed six previously identified risk loci, making a total of 17 risk loci for IA. A SNP-based heritability of 21.6% was found, explaining over half of the total heritability. We showed strong evidence that the majority of IA heritability is polygenic. Our results further highlight several major features of the genetic architecture of IA. First, we identified endothelial cells as a key cell type in IA risk. Second, we showed that, out of 375 tested traits, smoking and BP predisposition were the main genetic risk factors for IA. Third, we showed that the main drivers of the genetic correlation between IA and other stroke types and between IA and abdominal aortic aneurysms are genetic predisposition for smoking and blood pressure. Last, we found pleiotropic characteristics of anti-epileptic drugs and sex hormones with IA.

Through gene-mapping incorporating gene expression datasets and distinct bioinformatics analyses, we were able to identify 11 potential causative genes within 6 of the Stage 1 risk loci. Many of these genes have known or putative roles in blood vessel function and blood pressure regulation. We found heritability enrichment in genes that are specifically expressed in a combined set of endothelial and mural cells, and not in other vascular cell types. Together, the identified potential causative genes and heritability enrichment analyses suggest an important role of the vascular endothelial cell (vEC) in IA development and rupture.

Through genetic correlation and formal causal inference methods, we established that genetic predisposition for smoking and BP are the most important independent genetic risk factors for IA[1]. First, using causal inference with GSMR, we showed that genetic predisposition for these traits drives a causal increase in IA risk. Then, using multi-trait conditional analysis, we showed that smoking and high BP are causative of IA, independent of one another. By using non-transformed continuous systolic blood pressure (SBP) and diastolic blood pressure (DBP) measures in the UK Biobank, we estimated the increase in IA risk per 1 mmHg increase of SBP to be 3.7-6%, and that of DBP to be 8-12%. These strong effects provide genetic evidence for clinical prevention by lowering blood pressure. Since smoking dose is not normally distributed, we were not able to estimate a quantitative effect of smoking on IA, but this has been done before using non-genetic methods[49-51]. Future studies that model risk prediction using polygenic risk scores should determine whether the polygenic risks of genetic risk factors for IA are clinically relevant risk factors for the disease.

We found that genetic correlations of IA with ischemic stroke (IS) and deep intracerebral hemorrhage (ICH) are mainly driven by genetic predisposition for smoking and BP. For ICH, conditioning on smoking and BP did not completely mitigate the genetic

correlation with IA, suggesting additional shared genetic causes. For vertebral artery dissection, a substantial, but not statistically significant correlation with IA was found, whereas this was absent in carotid artery dissection. We showed that the genetic correlation between IA and AAA was driven by smoking, but not by BP. This implies that IA is more dependent on BP compared to AAA. This observation could be a result of different ratios of unruptured and ruptured aneurysms included in the two GWASs. The AAA GWAS consists of mainly unruptured AAA[44], and while the role of BP on AAA rupture is clear, the effect on developing AAA is a matter of debate[52].

One of the main aims of IA research is to prevent rupture of IA and thus avoid the devastating consequences of aSAH. We performed various analyses in an attempt to identify genetic predictors specific for IA rupture. Instead, we found a very strong genetic correlation between ruptured and unruptured IA. These analyses together indicate that the common variant genetic architecture of ruptured and unruptured aneurysms are strikingly similar.

The heritability of unruptured IA has never been studied in twins, and may, therefore, not be an optimal estimate for IA heritability. One twin study estimated the heritability of aSAH at 41%[3]. Our finding that the genetic architecture of uIA and aSAH are similar suggests that this heritability estimate may also be accurate for unruptured IA. This means that in European ancestry populations, 53 to 73% of the heritability of IA can be explained by variants tagged in this GWAS.

Using transethnic genetic correlation, we found a remarkable similarity of genetic architecture between the European ancestry and East Asian ancestry GWASs of more than 90.8±14.6% (SE). This indicates that the majority of common-variant genetic causes are the same, regardless of ancestry. However, since the LD structures remain distinct, current

methods for summary statistic-based enrichment analysis cannot effectively account for population-specific variation in a cross-ethnic GWAS.

Drug class enrichment showed pleiotropic characteristics of anti-epileptic drugs and sex hormones with the genetic association of IA. It has been suggested that sex hormones might play a role in IA[39], potentially explaining why women have a higher IA risk than men[1]. However, as causal inference analysis with GSMR did not show evidence for the involvement of female hormones, further investigation is required. Enrichment of the anti-epileptic drug class may indicate shared disease mechanisms between IA and epilepsy. The main mechanism of anti-epileptic drugs is through blocking Na^+ and Ca^{2+} ion-channels[48]. Together with other ion channels, these play essential roles in contraction and relaxation of the blood vessels[53]. Mutations in the ion-channel gene *PKD2* (*TRRP2*) are known to cause IA. This gene product, along with other members of the *TRP* gene family, regulates systemic blood pressure through vasoconstriction and vasodilation[54,55]. More research on the effect of anti-epileptics on vascular tension and blood pressure will enhance our understanding of the disease-causing mechanisms. Furthermore, this could help to identify methods of IA prevention using anti-epileptics or related drugs.

In conclusion, we performed a GWAS meta-analysis on IA identifying 11 new risk loci, confirming 6 previously identified risk loci and explaining over half of the heritability of IA. We found strong evidence for a polygenic architecture. Through gene-mapping and heritability enrichment methods, we discovered a possible role for endothelial cells in IA development. We showed that the genetic architecture of unruptured and ruptured aneurysms are very similar. The well-known clinical risk factors, smoking and hypertension, were identified as main genetic drivers of IA. These risk factors also explained most of the similarity to other stroke types, IS and deep ICH, which could open a window for clinical

prevention. We also found pleiotropic effects between IA and anti-epileptic drugs, which require further investigation to understand the shared mechanisms of IA and epilepsy. Our findings represent a major advance in understanding the pathogenesis of IA and a significant step towards the development of effective genetic risk prediction and prevention of IA development and subsequent aSAH in the future.

Online Methods

Recruitment and diagnosis

Detailed cohort descriptions are given in the Supplementary Note. In brief, all IA cases have a saccular IA, in which we included both cases with ruptured-thus with aSAH- and unruptured IAs confirmed using imaging. Patients with conditions known to predispose to IA, including autosomal dominant polycystic kidney disease, Ehlers-Danlos disease and Marfan's syndrome, were excluded. All controls were unselected controls. Controls were matched by genotyping platform and country on cohort-level.

Genotype data quality control

Cohorts for which individual level data were available are specified in Supplementary Table 1. An overview of inclusion and exclusion criteria, data collection and genotyping methods for each cohort are given in the Supplementary Note. Genotypes were lifted to reference genome build GRCh37. An extensive QC was performed on each cohort, described in detail in the Supplementary Note. Cohorts were merged into strata based on genotyping platform and country. An overview of strata compositions is given in Supplementary Table 1. Next, QC was performed on each stratum, outlined in the Supplementary Note. Genotypes were imputed against the Haplotype Reference Consortium (HRC) release 1.1. After imputation, another set of QC steps was taken, which is described in the Supplementary Note. An

overview of the number of SNPs, cases and controls excluded in the QC is shown in Supplementary Table 1.

Individual level association analysis

For each stratum, single-SNP associations were calculated using SAIGE (0.29.3)[14]. SAIGE uses a logistic mixed model to account for population stratification and saddle point approximation to accurately determine P-values even in the presence of case-control imbalance. Details on how these steps were performed are described in the Supplementary Note.

Meta-analysis

We meta-analyzed association statistics from our individual level SAIGE analysis with association statistics prepared by other groups who used the same analysis pipeline. There were two meta-analysis stages: Stage 1, including all individual level data and the European ancestry summary statistics (HUNT Study), and Stage 2 including all individual level data and all summary statistics (HUNT Study, China Kadoorie Biobank, Biobank Japan). Summary statistics that were generated by other groups were cleaned prior to meta-analysis, as described in the Supplementary Note. We used METAL (release 2011-03-25)[56] for the inverse-variance weighted meta-analysis across all studies. Only SNPs present in at least 80% of the strata were included.

Conditional analysis

To investigate whether a genome-wide significant locus consisted of multiple independent signals we used GCTA-COJO. COJO uses GWAS summary statistics and the LD structure of a reference panel to iteratively condition GWAS summary statistics on top SNPs. We used control samples from stratum sNL2 (Doetinchem Cohort Study) as a reference panel for LD

507 estimation. We used a stepwise approach to condition on the top independent SNPs with
508 $P<5\cdot10^{-8}$ and minor allele frequency (MAF) > 0.01. In addition, we conditioned the summary
509 statistics on the identified top independent hits to determine if any additional signal
510 remained.

511 ## Genetic risk score analysis

512 To investigate the effect of genetic risk for blood pressure (BP) and smoking on IA, we used
513 its genetic risk scores (GRS) as covariates in a SAIGE association model. Summary statistics
514 for BP-related traits[16] and cigarettes per day (CPD)[15] were obtained. SNPs to include in the
515 GRS models were determined using different LD thresholds by clumping (R-squared of 0.1,
516 0.2, 0.5, 0.8 or 0.9). Individual level GRS were calculated using plink v1.9. The optimal
517 models were selected based on the highest fraction of variance explained (adj.r.squared
518 from lm() in R/3.6.1). An optimal R-squared of 0.1 and 0.9 were selected for BP and CPD,
519 respectively. A set of 20,000 individuals from the UK Biobank, including all IA cases, was
520 used to train the model. Individual levels GRSs using the optimized set of SNPs was used as a
521 covariate in an association analysis using SAIGE.

522 ## eQTL-based gene mapping

523 We used eCAVIAR[57] to determine colocalization of GWAS hits with eQTLs. Vascular and
524 whole blood eQTLs from GTEx v7 were used. eCAVIAR used SNP Z-scores and LD correlation
525 values to calculate a colocalization posterior probability (CLPP) of a trait GWAS locus and an
526 eQTL. eCAVIAR requires an LD matrix to determine colocalization of eQTLs and GWAS hits.
527 We calculated LD in SNPs 1MB on both sides of the SNPs with lowest Stage 1 GWAS P-value,
528 using European ancestry Health and Retirement Study (HRS dbGaP accession code
529 phs000428.v2.p2) samples as a reference. Multiple causal SNPs were allowed.

TWAS is a method to perform differential expression analysis with eQTL-based predicted transcript levels. We used a summary statistics-based approach integrated in FUSION[58]. We used the 1000 Genomes LD weights provided by FUSION, and vascular and blood eQTL datasets provided on the FUSION reference webpage (URLs). Default settings were used for all other options.

SMR[59] was used to highlight genes the expression of which has a causal influence on IA risk. eQTL reference datasets from vascular tissues and blood provided by the creators of SMR were used. These include: CAGE, GTEx V7 (aorta, coronary artery, tibial artery and whole blood) and Westra (URLs). eQTLs with a p-value below $5 \cdot 10^{-8}$ were selected. The MAF cutoff was set at 0.01. European ancestry samples from the HRS were used as LD reference panel. Both the single SNP and multi-SNP approaches were used.

eCAVIAR, TWAS and SMR results were used to annotate genes to genome-wide significant GWAS loci identified in the Stage 1 GWAS meta-analysis. This approach is explained in more detail in the Supplementary Note.

SNP-based heritability

To calculate SNP-based heritability, we used LDSC (1.0.0)[32] to perform LD-score regression (LDSR), and we used SumHer[33]. LDSC makes the assumption that the contribution of each SNP to the total SNP heritability is normally distributed and not affected by MAF or LD. SumHer is the summary statistics based equivalent of an LD-adjusted kinship (LDAK) method to estimate SNP heritability and, instead, assumes that heritability is higher for low MAF variants and lower in high LD regions. In addition, SumHer models inflation due to residual confounding as a multiplicative parameter, whereas LDSC models this additively (the LDSR intercept). Heritability estimates were converted to the liability scale using effective sample

size. More details and the rationale of these analyses are described in the Supplementary Note.

Functional enrichment analysis using LDSC

To assess enrichment of heritability in functional annotations, tissues, chromosomes and minor allele frequency (MAF) bins, we used stratified LD-score regression with LDSC[60]. When available we used the publicly available partitioned LD scores for pre-defined annotations provided by the LDSC authors (URLs), otherwise we calculated our own LD scores using European ancestry samples from the 1000 Genomes (1000G) project. To further assess cell type-specific enrichment, we used a method introduced by Skene et al[35]. For this analysis, we used single-cell RNA sequencing (scRNAseq) gene expression data derived from mouse brain to define gene sets specific to cell types in brain[35] and brain blood vessels[36]. A detailed description of the rationale and parameters is given in the Supplementary Note.

Functional enrichment analysis using GARFIELD

The GWAS functional enrichment tool GARFIELD v2[61] was used to explore regulatory, functional and tissue-specific enrichment of the GWAS summary statistics. It determines whether GWAS SNPs reaching a certain P-value threshold are enriched in annotations of interest compared to the rest of the genome while accounting for distance to nearest transcription start site, MAF and LD. We used the default annotations provided by the authors to test enrichment in tissues (URLs). We tested enrichment of SNPs passing P-value thresholds for every $\log_{10}$-unit between 0.1 and 10^{-8}. A more detailed description of the method is given in the Supplementary Note.

Genetic correlation

We assessed correlation between IA and other traits using LDHub and LD-score regression (LDSR) with LDSC. To assess genetic correlation between IA and many non-stroke-related traits, we used LD Hub[40]. This platform uses LDSR to assess genetic correlation with a large number of publicly available GWASs. For the correlation of IA and other stroke subtypes, we obtained summary statistics for All Stroke (AS), Cardioembolic Stroke (CE), Any Ischemic Stroke (AnyIS), Large Artery Stroke (LAS), Small Vessel Disease (SVD)[41], Deep, Lobar, and combined Intracerebral Hemorrhage (ICH)[62], carotid- and vertebral artery dissection[43], Arteriovenous Malformation (AVM)[42] and Abdominal Aortic Aneurysms (AAA)[44]. We used LDSC to calculate genetic correlation. LD scores from European ancestry individuals from 1000G were calculated for SNPs in the HapMap 3 SNP set and used to calculate genetic correlation. Since the heritability estimate was negative for AVM, due to the small sample size, we performed a SNP lookup of the Stage 2 IA loci that passed the multiple testing threshold ($P<5\cdot10^{-8}$) from the GWAS of AVM[42].

Conditional genetic correlation

We used mtCOJO to condition Stage 1 IA GWAS summary statistics on summary statistics from the Neale lab UK Biobank GWAS release 1 (URLs) for smoking and blood pressure (BP) following a method described previously[63]. The resulting summary statistics were then used to calculate genetic correlation between IA, conditioned on another trait, and other vascular diseases. LD scores supplied by LDSC (*eur_w_ld_chr/[chr].l2.ldscore.gz*) were used. European ancestry control samples from stratum sNL2 (from the Doetinchem Cohort Study) were used as an LD reference panel. All other settings were left as default.

Trans-ancestry genetic correlation

Popcorn version 0.9.9[64] was used to assess genetic correlation between IA cohorts of European and East Asian ancestry. Popcorn uses separate LD score reference panels per ancestry to account for differences in LD structure between cohorts. We used LD scores provided by the authors of the Popcorn tool (URLs) for European and East Asian descent (EUR_EAS_all_gen_[eff/imp].cscore). We calculated the genetic correlation for both genetic impact and genetic effect.

Mendelian randomization

To infer causal genetic effects of exposure traits on IA (the outcome), we used GSMR[63]. We used a meta-analysis of all European ancestry strata, except the UK biobank (stratum sUK2), as outcome. As exposures we used summary statistics of 2419 traits analyzed using UK Biobank data, prepared by the Neale lab, release 2017 (URLs). For a second GSMR run with raw quantitative phenotypes we used the 2019 GWAS release from the same group. GSMR was run using the GCTA wrapper (v1.92.2). More details on the method and settings are described in the Supplementary Note.

In order to determine which of the top significant GSMR traits were independent genetic causes of IA, the Stage 1 GWAS summary statistics were conditioned on the top traits, i.e. smoking and blood pressure (BP). Conditioning was done using mtCOJO as described in the Conditional genetic correlation section of the Online Methods.

Drug target enrichment

Drug target enrichment analysis was performed according to a previously described method[47]. Gene-wise P-values were calculated with MAGMA v1.06 using a combined approach of average and top P-values per gene region. Gene-set analysis was performed

using MAGMA, with pathways curated from MSigDB[65,66], TargetValidation (URLs), and with drug-target sets described previously[47]. Drug-class enrichment analysis was performed using a Wilcoxon-Mann-Whitney test. Drug gene-set P-values were tested for enrichment in drug-classes. Enrichment was expressed as the area under the curve (AUC). AUCs were compared between drug gene-sets within a drug class and all other drug gene-sets.

References

1. Vlak, M.H., Algra, A., Brandenburg, R. & Rinkel, G.J. Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis. *Lancet Neurol* **10**, 626-36 (2011).
2. Nieuwkamp, D.J. *et al.* Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: a meta-analysis. *Lancet Neurol* **8**, 635-42 (2009).
3. Korja, M. *et al.* Genetic epidemiology of spontaneous subarachnoid hemorrhage: Nordic Twin Study. *Stroke* **41**, 2458-62 (2010).
4. Kurki, M.I. *et al.* High risk population isolate reveals low frequency variants predisposing to intracranial aneurysms. *PLoS Genet* **10**, e1004134 (2014).
5. Yasuno, K. *et al.* Common variant near the endothelin receptor type A (EDNRA) gene is associated with intracranial aneurysm risk. *Proc Natl Acad Sci U S A.* **108**, 19707-12 (2011).
6. Yan, J. *et al.* Genetic study of intracranial aneurysms. *Stroke* **46**, 620-6 (2015).
7. Santiago-Sim, T. *et al.* THSD1 (Thrombospondin Type 1 Domain Containing Protein 1) Mutation in the Pathogenesis of Intracranial Aneurysm and Subarachnoid Hemorrhage. *Stroke* **47**, 3005-13 (2016).
8. Bourcier, R. *et al.* Rare Coding Variants in ANGPTL6 Are Associated with Familial Forms of Intracranial Aneurysm. *Am J Hum Genet.* **102**, 133-141 (2018).
9. Lorenzo-Betancor, O. *et al.* PCNT point mutations and familial intracranial aneurysms. *Neurology* **91**, e2170-e2181 (2018).
10. Zhou, S. *et al.* RNF213 Is Associated with Intracranial Aneurysms in the French-Canadian Population. *Am J Hum Genet* **99**, 1072-1085 (2016).
11. Hussain, I., Duffis, E.J., Gandhi, C.D. & Prestigiacomo, C.J. Genome-wide association studies of intracranial aneurysms: an update. *Stroke* **44**, 2670-5 (2013).
12. Foroud, T. *et al.* Genome-wide association study of intracranial aneurysms confirms role of Anril and SOX17 in disease risk. *Stroke* **43**, 2846-52 (2012).
13. Yasuno, K. *et al.* Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat Genet* **42**, 420-5 (2010).
14. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341 (2018).
15. Tobacco Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).

16. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425 (2018).
17. Lee, S. *et al.* Deficiency of endothelium-specific transcription factor Sox17 induces intracranial aneurysm. *Circulation* **131**, 995-1005 (2015).
18. Laarman, M.D. *et al.* Chromatin Conformation Links Putative Enhancers in Intracranial Aneurysm-Associated Regions to Potential Candidate Genes. *J Am Heart Assoc* **8**, e011201 (2019).
19. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet* **51**, 51-62 (2019).
20. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75 (2019).
21. Takeuchi, F. *et al.* Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nat Commun* **9**, 5052 (2018).
22. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* **49**, 54-64 (2017).
23. Huang, L. *et al.* A missense variant in FGD6 confers increased risk of polypoidal choroidal vasculopathy. *Nat Genet* **48**, 640-7 (2016).
24. Romanoski, C.E. *et al.* Systems genetics analysis of gene-by-environment interactions in human cells. *Am J Hum Genet* **86**, 399-410 (2010).
25. Haasdijk, R.A. *et al.* THSD1 preserves vascular integrity and protects against intraplaque haemorrhaging in ApoE-/- mice. *Cardiovasc Res* **110**, 129-39 (2016).
26. Yang, X. *et al.* Rho Guanine Nucleotide Exchange Factor ARHGEF17 Is a Risk Gene for Intracranial Aneurysms. *Circ Genom Precis Med* **11**, e002099 (2018).
27. Camacho Leal Mdel, P. *et al.* p130Cas/BCAR1 scaffold protein in tissue homeostasis and pathogenesis. *Gene* **562**, 1-7 (2015).
28. Nedeljkovic, I. *et al.* Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics. *Eur J Hum Genet* **26**, 709-722 (2018).
29. David, S.P. *et al.* Genome-wide meta-analyses of smoking behaviors in African Americans. *Transl Psychiatry* **2**, e119 (2012).
30. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244 (2019).
31. Lutz, S.M. *et al.* A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet* **16**, 138 (2015).
32. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
33. Speed, D. & Balding, D.J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat Genet* **51**, 277-284 (2019).
34. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348 (2019).
35. Skene, N.G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat Genet* **50**, 825-833 (2018).
36. He, L. *et al.* Single-cell RNA sequencing of mouse brain and lung vascular and vessel-associated cell types. *Sci Data* **5**, 180160 (2018).

37. Backes, D., Rinkel, G.J., Laban, K.G., Algra, A. & Vergouwen, M.D. Patient- and Aneurysm-Specific Risk Factors for Intracranial Aneurysm Growth: A Systematic Review and Meta-Analysis. *Stroke* **47**, 951-7 (2016).
38. Muller, T.B., Vik, A., Romundstad, P.R. & Sandvei, M.S. Risk Factors for Unruptured Intracranial Aneurysms and Subarachnoid Hemorrhage in a Prospective Population-Based Study. *Stroke* **50**, 2952-2955 (2019).
39. Algra, A.M., Klijn, C.J., Helmerhorst, F.M., Algra, A. & Rinkel, G.J. Female risk factors for subarachnoid hemorrhage: a systematic review. *Neurology* **79**, 1230-6 (2012).
40. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
41. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* **50**, 524-537 (2018).
42. Weinsheimer, S. *et al.* Genome-wide association study of sporadic brain arteriovenous malformations. *J Neurol Neurosurg Psychiatry* **87**, 916-23 (2016).
43. Debette, S. *et al.* Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. *Nat Genet* **47**, 78-83 (2015).
44. Jones, G.T. *et al.* Meta-Analysis of Genome-Wide Association Studies for Abdominal Aortic Aneurysm Identifies Four New Disease-Specific Risk Loci. *Circ Res* **120**, 341-353 (2017).
45. Hankey, G.J. Stroke. *Lancet* **389**, 641-654 (2017).
46. An, S.J., Kim, T.J. & Yoon, B.W. Epidemiology, Risk Factors, and Clinical Features of Intracerebral Hemorrhage: An Update. *J Stroke* **19**, 3-10 (2017).
47. Gaspar, H.A. & Breen, G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci Rep* **7**, 12460 (2017).
48. Rogawski, M.A. & Loscher, W. The neurobiology of antiepileptic drugs. *Nat Rev Neurosci* **5**, 553-64 (2004).
49. Lindbohm, J.V., Kaprio, J., Jousilahti, P., Salomaa, V. & Korja, M. Sex, Smoking, and Risk for Subarachnoid Hemorrhage. *Stroke* **47**, 1975-81 (2016).
50. Vlak, M.H., Rinkel, G.J., Greebe, P. & Algra, A. Risk of rupture of an intracranial aneurysm based on patient characteristics: a case-control study. *Stroke* **44**, 1256-9 (2013).
51. Juvela, S., Poussa, K. & Porras, M. Factors affecting formation and growth of intracranial aneurysms: a long-term follow-up study. *Stroke* **32**, 485-91 (2001).
52. Kobeissi, E., Hibino, M., Pan, H. & Aune, D. Blood pressure, hypertension and the risk of abdominal aortic aneurysms: a systematic review and meta-analysis of cohort studies. *Eur J Epidemiol* **34**, 547-555 (2019).
53. Cheng, J. *et al.* Ion Channels and Vascular Diseases. *Arterioscler Thromb Vasc Biol* **39**, e146-e156 (2019).
54. Bulley, S. *et al.* Arterial smooth muscle cell PKD2 (TRPP1) channels regulate systemic blood pressure. *Elife* **7**(2018).
55. Perrone, R.D., Malek, A.M. & Watnick, T. Vascular complications in autosomal dominant polycystic kidney disease. *Nat Rev Nephrol* **11**, 589-98 (2015).
56. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).

57. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* **99**, 1245-1260 (2016).
58. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
59. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
60. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
61. Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet* **51**, 343-353 (2019).
62. Woo, D. *et al.* Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet* **94**, 511-21 (2014).
63. Zhu, Z.H. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* **9**(2018).
64. Brown, B.C., Asian Genetic Epidemiology Network Type 2 Diabetes, C., Ye, C.J., Price, A.L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am J Hum Genet* **99**, 76-88 (2016).
65. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-73 (2003).
66. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).

Ethical Statement

Approval of data usage was given by all local Data Access Committees. All participants provided written informed consent. The Biobank Research Ethics Committee of the University Medical Center Utrecht reviewed and approved the study protocol (TCBio 17-087).

Acknowledgements

This research has been conducted using the UK Biobank Resource under application number 2532. We acknowledge Dr. R.L. McLaughlin for the advice on population-based heritability analysis. We acknowledge Prof. Dr. M. Gunel and Dr. K. Yasuno for their help with genotyping DNA samples of the Utrecht 1, Finland, and @neurIST cohorts. The authors

thank the staff and participants of all CADISP centers for their important contributions. The authors acknowledge the contribution of participants, project staff, and China National Centre for Disease Control and Prevention (CDC) and its regional offices to the China Kadoorie Biobank. China's National Health Insurance provided electronic linkage to all hospital treatments. The authors acknowledge K.G. Jebsen for genotyping quality control and imputation of the HUNT Study.

For providing clinical information and biological samples collected during the @neurIST project, we thank: Juan Macho, Tamás Dóczi, James Byrne, Paul Summers, Roelof Risselada, Miriam C. J. M. Sturkenboom, Umang Patel, Stuart Coley, Alan Waterworth, Daniel Rüfenacht, Carole Proust, and François Cambien.

Funding

We acknowledge the support from the Netherlands Cardiovascular Research Initiative: An initiative with support of the Dutch Heart Foundation, CVON2015-08 ERASE.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 852173).

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 772376 – EScORIAL).

BioBank Japan project was supported by the Ministry of Education, Culture, Sports, Sciences, and Technology of the Japanese government and the Japan Agency for Medical Research and Development (19km0605001).

809 The CADISP study has been supported by INSERM, Lille 2 University, Institut Pasteur de Lille
810 and Lille University Hospital and received funding from the European Regional Development
811 Fund (FEDER funds) and Région Nord-Pas-de-Calais in the framework of Contrat de Projets
812 Etat-Region 2007–2013 Région Nord-Pas-de-Calais (grant 09120030), Centre National de
813 Génotypage, the Emil Aaltonen Foundation, the Paavo Ilmari Ahvenainen Foundation, the
814 Helsinki University Central Hospital Research Fund, the Helsinki University Medical
815 Foundation, the Päivikki and Sakari Sohlberg Foundation, the Aarne Koskelo Foundation, the
816 Maire Taponen Foundation, the Aarne and Aili Turunen Foundation, the Lilly Foundation,
817 the Alfred Kordelin Foundation, the Finnish Medical Foundation, the Orion Farmos Research
818 Foundation, the Maud Kuistila Foundation, the Finnish Brain Foundation, the Biomedicum
819 Helsinki Foundation, Projet Hospitalier de Recherche Clinique Régional, Fondation de
820 France, Génopôle de Lille, Adrinord, the Basel Stroke Funds, the Käthe-Zingg-
821 Schwichtenberg-Fonds of the Swiss Academy of Medical Sciences and the Swiss Heart
822 Foundation.
823 S.D. has received funding from the French National Funding Agency (ANR), the European
824 Research Council (ERC) under the European Union's Horizon 2020 research and innovation
825 programme (grant agreement No 640643).
826 BioBank Japan project was supported by the Ministry of Education, Culture, Sports,
827 Sciences, and Technology of the Japanese government and the Japan Agency for Medical
828 Research and Development (19km0605001).
829 J.P. was supported by Jagiellonian University Medical College (JUMC) grant K/ZDS/001456.
830 China Kadoorie Biobank was supported as follows: Baseline survey and first re-survey: Hong
831 Kong Kadoorie Charitable Foundation; long-term follow-up: UK Wellcome Trust
832 (202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), National Natural Science Foundation of

China (81390540, 81390541, 81390544), and National Key Research and Development Program of China (2016YFC 0900500, 0900501, 0900504, 1303904). DNA extraction and genotyping: GlaxoSmithKline, UK Medical Research Council (MC_PC_13049, MC-PC-14135). Core funding to the Clinical Trial Service Unit and Epidemiological Studies Unit at Oxford University was provided by The British Heart Foundation, UK MRC, and Cancer Research UK.

S.Z. and G.A.R. received funding from Canadian Institutes of Health Research (CIHR).

This project has received funding from the European Union's Horizon 2020 research and innovation programme (No. 666881), SVDs@target (to M.D.) and No. 667375, CoSTREAM (Common Mechanisms and Pathways in Stroke and Alzheimer's Disease; to M.D.); the DFG (Deutsche Forschungsgemeinschaft) as part of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy—ID 390857198) and the CRC 1123 (B3, to M.D.); the Corona Foundation (to M.D.); the Fondation Leducq (Transatlantic Network of Excellence on the Pathogenesis of Small Vessel Disease of the Brain, to M.D.); the e:Med program (e:AtheroSysMed, to M.D.); and the FP7/2007-2103 European Union project CVgenes@target (grant agreement No. Health-F2-2013-601456, to M.D.).

K.R. is funded by the Health Data Research UK (HDRUK) fellowship MR/S004130/1.

C.L.M.S. was funded by the UK Biobank, Health Data Research UK, and Scottish Funding Council.

I.C.H. received funding from the Alzheimer Research UK and Dunhill Medical Trust Foundation.

J.P.B. and D.W. were supported by NIH Funding.

D.J.W. received funding support from The Stroke Association and the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

The Nord-Trøndelag Health Study (HUNT Study) is a collaboration between the HUNT Research Centre, Faculty of Medicine at the Norwegian University of Science and Technology (NTNU), the Norwegian Institute of Public Health and the Nord- Trøndelag County Council. The genotyping was financed by the National Institute of health (NIH), University of Michigan, The Norwegian Research council, and Central Norway Regional Health Authority and the Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU).

P.B. and C.M.F. were supported by EU commission FP6 – IST - 027703 @neurIST-Integrated biomedical informatics for the management of cerebral aneurysms.

P.B., S.M., S.H., S.S., J.D., and O.M. were supported by the Grant MRD 2014/261 from the Swiss SystemsX.ch initiative and evaluated by the Swiss National Science Foundation (AneuX project).

Author Contributions

J.H.V. and Y.M.R. contributed equally to this study.

Writing and editing the manuscript: M.K.B., Y.M.R., J.H.V.

Supervising the project: Y.M.R., J.H.V.

Designing the study: I.C.H., S.D., B.B.W., J.P., A.S., E.I.G-P., M.N., J.E.J., M.v.U.Z.F., A.L., J.P.B., D.J.W., D.W., R.R., P.B., Y.K., J.H.V., Y.M.R.

Association analyses and scripts: M.K.B., R.A.A.v.d.S., W.v.R.

Functional analyses and scripts: M.K.B., R.A.A.v.d.S., W.v.R.

Phenotype preparation: S.M., R.B., C.M.F., S.H., S.S., J.D., O.M., P.B.

Technical assistance: R.A.A.v.d.S., K.R.v.E.

Phenotype and genotype contributions: M.K.B., Y.M.R., G.J.E.R., J.H.V., L.H.v.d.B., P.B., S.M., E.I.G-P., M.N., J.P., A.S., J.E.E., M.v.U.Z.F., A.L., G.A.R., S.Z., N.U.K., R.M., K.R., C.L.M.S., D.J.W., I.C.H., H.H., V.S.A., J.P.B., D.W., R.R., R.B., C.D., O.N., J.C.G., E.S., F.E., H.D., W.M.M.V. Summary statistic contributions: Y.K., M.K., M.A., C.T., K.M. (BBJ), R.G.W., K.L., L.L., I.Y.M., Z.C. (CKB), B.S.W., S.B., M.B.J., B.M.B., M.S.S., C.J.W., K.H., J.A.Z. (HUNT), M.J.B., G.T.J. (AAA), H.K., J.G.Z., C.J.M.K., N.U.K. (AVM), D.W. (ICH), R.M., M.D. (IS), S.D., T.T., M.S. P.A. (cervical artery dissection).

Drug-target and MAGMA pathway enrichment analyses: J.R.I.C., G.B.

Corresponding author

Ynte M. Ruigrok.: ij.m.ruigrok@umcutrecht.nl

Competing Interests

When this study was conducted, C.L.M.S. was chief scientist for the UK Biobank study.

Data availability

Summary statistics for the Stage 1 and Stage 2 GWAS meta-analyses, the SAH-only, and uIA-only GWAS, and a meta-analysis consisting of only East Asian samples can be accessed upon publication using doi: 10.6084/m9.figshare.11303372. And through the Cerebrovascular Disease Knowledge Portal: http://www.cerebrovascularportal.org.

URLs

COJO and mtCOJO: http://cnsgenomics.com/software/gcta/#Overview

GARFIELD: https://www.ebi.ac.uk/birney-srv/GARFIELD/

GSMR: http://cnsgenomics.com/software/gsmr/

LocusZoom: http://locuszoom.org

LDlink: https://ldlink.nci.nih.gov

LDSC tool: https://github.com/bulik/ldsc

LDSC LD scores: https://data.broadinstitute.org/alkesgroup/LDSCORE/

LDHub: http://ldsc.broadinstitute.org

MSigDB (from GSEA): http://software.broadinstitute.org/gsea/index.jsp

Plink: https://www.cog-genomics.org/plink2/

Popcorn: https://github.com/brielin/Popcorn

SAIGE: https://github.com/weizhouUMICH/SAIGE/wiki/Genetic-association-tests-using-SAIGE

SMR: https://cnsgenomics.com/software/smr/

SMR eQTL datasets: https://cnsgenomics.com/software/smr/#DataResource

TargetValidation: https://www.targetvalidation.org

TWAS/FUSION: http://gusevlab.org/projects/fusion/

UK Biobank GWAS release 1: http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank

UK Biobank GWAS release 2: http://www.nealelab.is/uk-biobank/

Figure legends

Figure 1. GWAS meta-analysis association results.

SAIGE logistic mixed model association P-values of the Stage 1 (upwards direction) and Stage 2 (downwards direction) GWAS meta-analyses. The horizontal axis indicates chromosomal position. The vertical axis indicates -$\log_{10}$(P-value) of the association. The dotted lines indicate the genome-wide significance threshold of $P=5\cdot10^{-8}$. Lead SNPs of each locus are highlighted with a diamond, and SNPs in close proximity (±500Kbp) are colored in pink or purple, depending on chromosome index parity. Labels are gene or locus names annotated using SMR, eCAVIAR and TWAS, or prior information of IA-associated genes. Labels or loci identified only in the Stage 2 GWAS are shown in red.

Figure 2. Heritability and functional enrichment analyses.

a) Partitioned LDSR enrichment analysis of regulatory elements. Labels indicate type of regulatory element or histone mark used to define regulatory region positions. On the horizontal axis, the relative enrichment of an element compared to the rest of the genome is shown. The bold line at enrichment=1 indicates the Null value of no enrichment. Red points indicate a significant enrichment of P<0.05 divided by the number of annotations (52), blue indicates a nominally significant enrichment of P<0.05, and grey indicates P>0.5. **b)** Partitioned LDSR heritability analysis per chromosome. On the horizontal axis the proportion of SNPs per chromosome is shown. On the vertical axis the proportion of SNP-based heritability. The blue line is the regression line of heritability proportion regressed on proportion of SNPs. **c)** Partitioned LDSR enrichment analysis of scRNAseq brain cell types. Coloring and labelling are the same as for a). **d)** GARFIELD analysis of tissues. On the

horizontal axis, the enrichment of annotations is shown; on the vertical axis, the corresponding -$\log_{10}$(P-value). Dashed line indicates the significance threshold of P=0.05 divided by the number of annotations in that category. **e)** GARFIELD analysis of regulatory regions defined by histone modifications. Coloring and labelling are the same as for d). In all sub-plots, error bars denote standard error.

Figure 3. Cross-trait analyses.

a) GSMR analysis of UK Biobank predictors on the Stage 1 IA GWAS, conditioned on traits depicted by column labels with mtCOJO. Numeric values are the GSMR effect sizes. The top 13 traits are pressure-related traits. The bottom three traits are smoking-related. Text is colored black if the GSMR effect was statistically significant beyond the Bonferroni threshold (P<0.05 divided by the number of traits that passed quality control [376]); if not, text is grey. Square fill colors indicate -$\log_{10}$(P-value) of the GSMR effect. All 16 traits that pass the multiple testing threshold for significance in the unconditioned analysis are shown. BP: blood pressure. **b)** Causality d further explaining the analyses of a. GSMR analysis showed that genetic risk for smoking and BP are causative of IA. Using mtCOJO, it was found that the genetic factors associated with BP and smoking cause IA through independent mechanisms. **c)** Genetic correlation analysis with LDSR. Genetic correlation estimates are indicated by color and numeric value. Axis labels on the left denote the trait used for genetic correlation analysis with IA. Labels on the top denote the trait for which the Stage 1 IA GWAS was conditioned using mtCOJO. Text color and symbols indicate P-value bin for genetic correlation. Stars and points indicate significance bin. P<0.001: red and '***', P<0.01: black and '**', P<0.05: black and '*', P<0.1: dark grey and '·', P⩾0.1: light grey. IS: ischemic stroke. ICH: intracerebral hemorrhage. AAA: abdominal aortic aneurysm.

964