

Genetic Overlap Between Diagnostic Subtypes of Ischemic Stroke

Elizabeth G. Holliday, PhD; Matthew Traylor, PhD; Rainer Malik, PhD; Steve Bevan, PhD; Guido Falcone, MD; Jemma C. Hopewell, PhD; Yu-Ching Cheng, PhD; Ioana Cotlarciuc, PhD; Joshua C. Bis, PhD; Eric Boerwinkle, PhD; Giorgio B. Boncoraglio, MD; Robert Clarke, MD; John W. Cole, MD; Myriam Fornage, PhD; Karen L. Furie, MD; M. Arfan Ikram, MD, PhD; Jim Jannes, MBBS; Steven J. Kittner, MD; Lisa F. Lincz, PhD; Jane M. Maguire, PhD; James F. Meschia, MD; Thomas H. Mosley, PhD; Mike A. Nalls, PhD; Christopher Oldmeadow, PhD; Eugenio A. Parati, MD; Bruce M. Psaty, MD, PhD; Peter M. Rothwell, FRCP; Sudha Seshadri, MD; Rodney J. Scott, PhD; Pankaj Sharma, FRCP; Cathie Sudlow, FRCPE; Kerri L. Wiggins, MS; Bradford B. Worrall, MD, MSc; Jonathan Rosand, MD; Braxton D. Mitchell, PhD; Martin Dichgans, MD; Hugh S. Markus, DM; Christopher Levi, MBBS; John Attia, MD, PhD*; Naomi R. Wray, PhD*; on behalf of the Australian Stroke Genetics Collaborative, the Wellcome Trust Case Control Consortium 2, and the International Stroke Genetics Consortium

Background and Purpose—Despite moderate heritability, the phenotypic heterogeneity of ischemic stroke has hampered gene discovery, motivating analyses of diagnostic subtypes with reduced sample sizes. We assessed evidence for a shared genetic basis among the 3 major subtypes: large artery atherosclerosis (LAA), cardioembolism, and small vessel disease (SVD), to inform potential cross-subtype analyses.

Methods—Analyses used genome-wide summary data for 12 389 ischemic stroke cases (including 2167 LAA, 2405 cardioembolism, and 1854 SVD) and 62 004 controls from the Metastroke consortium. For 4561 cases and 7094 controls, individual-level genotype data were also available. Genetic correlations between subtypes were estimated using linear mixed models and polygenic profile scores. Meta-analysis of a combined LAA–SVD phenotype (4021 cases and 51 976 controls) was performed to identify shared risk alleles.

Received October 27, 2014; final revision received December 17, 2014; accepted December 19, 2014.

From the School of Medicine and Public Health (E.G.H., J.A.), School of Biomedical Sciences and Pharmacy (L.F.L., R.J.S.), and School of Nursing and Midwifery (J.M.M.), University of Newcastle, Australia; Clinical Research Design, IT and Statistical Support Unit, Public Health Research Program, Hunter Medical Research Institute, Newcastle, Australia (E.G.H., C.O., J.A.); Department of Clinical Neurosciences, University of Cambridge, UK (M.T., S.B., H.S.M.); Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany (R.M., M.D.); Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA (G.F., J.R.); Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA (G.F., J.R.); Clinical Trial Service Unit and Epidemiological Studies Unit (J.C.H., R.C.) and Stroke Prevention Research Unit, Nuffield Department of Clinical Neurosciences (P.M.R.), University of Oxford, UK; Department of Medicine (Y.-C.C., B.D.M.) and Department of Neurology (J.W.C., S.J.K.), University of Maryland School of Medicine, Baltimore, MD; Imperial College Cerebrovascular Research Unit (ICCRU), Imperial College, London, UK (I.C.); Cardiovascular Health Research Unit, Department of Medicine (J.C.B., B.M.P., K.L.W.) and Department of Epidemiology (B.M.P.), University of Washington, Seattle, WA; Institute of Molecular Medicine and Human Genetics Center, the University of Texas Health Science Center at Houston, TX (E.B., M.F.); Department of Cerebrovascular Disease, IRCCS Istituto Neurologico Carlo Besta, Milan, Italy (G.B.B., E.A.P.); Department of Neurology, Veterans Affairs Medical Center, Baltimore, MD (J.W.C., S.J.K.); Department of Neurology, Rhode Island Hospital, Providence, RI (K.L.F.); Departments of Epidemiology, Neurology, and Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands (M.A.I.); School of Medicine, University of Adelaide, Australia (J.J.); Hunter Haematology Research Group, Calvary Mater Newcastle Hospital, Newcastle, Australia (L.F.L.); Hunter Medical Research Institute, Newcastle, Australia (L.F.L., J.M.M., R.J.S., C.L.); Department of Neurology, Mayo Clinic, Jacksonville, FL (J.F.M.); Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, MS (T.H.M.); Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD (M.A.N.); Group Health Research Institute, Group Health, Seattle, WA (B.M.P.); Department of Neurology, Boston University School of Medicine, Boston, MA (S.S.); Pathology North, NSW Pathology (R.J.S.) and Department of Neurology (C.L.), John Hunter Hospital, Newcastle, Australia; Institute of Cardiovascular Research, Royal Holloway University of London (ICR2UL) & Ashford and St Peters Hospital (P.S.); Division of Clinical Neurosciences, University of Edinburgh, UK (C.S.); Departments of Neurology and Public Health Sciences, University of Virginia, Charlottesville, VA (B.B.W.); Munich Cluster for Systems Neurology (SyNergy), Munich, Germany (M.D.); and Queensland Brain Institute, The University of Queensland, Australia (N.R.W.).

Guest Editor for this article was Jeffrey L. Saver, MD.

*Drs Attia and Wray jointly directed this work.

The online-only Data Supplement is available with this article at <http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.114.007930/-DC1>.

Correspondence to Elizabeth Holliday, PhD, CReDITSS, HMRI, C/- University Dr, Callaghan, NSW 2308, Australia. E-mail liz.holliday@newcastle.edu.au

© 2015 American Heart Association, Inc.

Stroke is available at <http://stroke.ahajournals.org>

DOI: 10.1161/STROKEAHA.114.007930

Results—High genetic correlation was identified between LAA and SVD using linear mixed models ($r_g=0.96$, $SE=0.47$, $P=9\times 10^{-4}$) and profile scores ($r_g=0.72$; 95% confidence interval, 0.52–0.93). Between LAA and cardioembolism and SVD and cardioembolism, correlation was moderate using linear mixed models but not significantly different from zero for profile scoring. Joint meta-analysis of LAA and SVD identified strong association ($P=1\times 10^{-7}$) for single nucleotide polymorphisms near the opioid receptor $\mu 1$ (*OPRM1*) gene.

Conclusions—Our results suggest that LAA and SVD, which have been hitherto treated as genetically distinct, may share a substantial genetic component. Combined analyses of LAA and SVD may increase power to identify small-effect alleles influencing shared pathophysiological processes. (*Stroke*. 2015;46:615-619. DOI: 10.1161/STROKEAHA.114.007930.)

Key Words: atherosclerosis ■ genetic epidemiology ■ lacunar stroke

Ischemic stroke (IS) is a complex disease influenced by numerous clinical, genetic, and lifestyle risk factors. Although conventional factors, such as hypertension, dyslipidaemia, diabetes mellitus, and smoking are well-established, genetic factors contribute $\leq 30\%$ to 40% of risk¹ and are poorly understood. Despite recent advances in high throughput genotyping, gene discovery for IS has progressed slowly because of the technical nature of case ascertainment and etiologic heterogeneity of the IS diagnosis. The latter produces pathophysiological differences and implies genetic differences between patients, complicating efforts to identify susceptibility genes.

To assist diagnosis and clinical management, schemes have been developed to categorise IS into diagnostic subtypes.^{2,3} The major types are large artery atherosclerosis (LAA), cardioembolism, and small vessel (lacunar) disease (SVD). By exploiting these phenotypically more homogeneous categorizations, genome-wide association studies (GWAS) have identified several genetic associations specific to individual subtypes.⁴⁻⁷ In contrast, only 2 genome-wide significant associations have been identified for broadly defined IS,^{8,9} in spite of its ≈ 5 -fold larger sample sizes. Power of association studies is a balance between sample size and the (unknown) effect sizes of risk loci, with estimable effect size depending on genetic homogeneity. To date, GWAS of IS have focussed on individual subtypes which reduce sample size and may reduce power at some loci if there is a shared genetic basis between subtypes.¹⁰

This study aimed to estimate genetic correlations between the 3 major IS subtypes using individual-level GWAS data and meta-analysis summary statistics from the International Stroke Genetics Consortium^{6,7} and Metastroke.⁵ Genetic correlations were estimated using 2 different methods: linear mixed models (LMMs)¹¹ and polygenic profile scoring.¹²

Methods

Data Sources

The Metastroke study included 15 individual studies contributing 12389 total ischemic stroke cases and 62004 controls of European ancestry (Table 1 in the online-only Data Supplement). Details of these 15 studies, including genotyping, phenotyping, and participants' demographic details have been previously described in detail.⁵ Cases and controls did not overlap between studies and were confirmed as unrelated using genotypic data. Stroke subtyping was performed using the TOAST system,² identifying 2167 cases with large artery atherosclerosis (LAA), 2405 with cardioembolism, and 1854 cases with SVD; the remainder had other, undetermined, or cryptogenic pathogenesis. Each study conducted genotype imputation using either HapMap Phase 2 or 1000 Genomes reference panels, fitted

additive logistic regression models for all single nucleotide polymorphisms (SNPs), and provided regression summary statistics for IS and its subtypes (if available). Individual-level GWAS data were also available for 3 of the largest Metastroke cohorts: 2 from the Wellcome Trust Case Control Consortium 2 Study (United Kingdom and Munich)⁶ and the Australian Stroke Genetics Collaborative,⁷ which were genotyped using Illumina arrays with similar content. All studies were approved by appropriate ethics committees and participants provided written informed consent.

Linear Mixed Modeling

Genotype data for the 3 samples with individual-level data were combined to yield a single data set using the software PLINK.¹³ Stringent quality control removed SNPs not directly genotyped in all samples, with $>0.5\%$ missing data, Hardy-Weinberg P value <0.05 , minor allele frequency $<1\%$, or differential missingness ($P<0.05$) between samples. We excluded samples with $>1\%$ data missingness and one from each pair with an absolute value of genome-wide similarity >0.05 .¹⁴ Principal components of ancestry were calculated in the pooled sample after 3 iterations of principal components analysis with outlier removal (>5 standard deviations from the mean on PC1-5).¹⁵

Heritability within and genetic correlations (r_g) between subtypes were estimated using LMMs,¹⁴ adjusting for 20 principal components. Likelihood ratio statistics were used to test whether estimates were significantly different to zero. Heritability estimates were transformed to the liability scale assuming 2% lifetime prevalence for IS, to which the 3 major subtypes each contribute $\approx 20\%$ (total 60%), equating to 0.4% prevalence ($20\%\times 2\%$) for each subtype.^{5,16} The remaining 0.8% prevalence ($40\%\times 2\%$) was assumed to reflect other stroke types.

Polygenic Profile Scoring

The profile scoring approach uses SNP association statistics from a given phenotype to build a linear predictor and tests this for association with the same or a different phenotype in independent data. To facilitate interpretation of cross-subtype analyses, we first assessed association of profile scores within IS and its 3 subtypes. For each, this was performed using leave-one-out validation for the 3 target data sets with individual-level data. In turn, each of the 3 was set aside and a GWAS discovery meta-analysis was performed using all other Metastroke data sets. SNPs with data from ≥ 5 Metastroke studies were retained and pruned for linkage disequilibrium ($r^2>0.2$ within 1Mb) using PLINK's clump algorithm,¹³ which preferentially retains the most associated SNP in a linkage disequilibrium region. From the pruned set, we extracted subsets passing 10 graded significance thresholds ($P_T=1\times 10^{-4}$, 1×10^{-3} , 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1). For subsets passing each threshold, PLINK's score function was used to calculate profile scores for individuals in the left-out data set. These scores represent an average risk allele burden across all SNPs in the score, with weights assigned as the log odds ratio from the discovery meta-analysis.

Associations of profile scores with stroke subtypes were assessed by logistic regression adjusted for 3 ancestry principal components. Variance explained by the score was computed as the difference in

Nagelkerke's pseudo- R^2 between the model including the profile score and principal components and that including only principal components. Results for the 3 target data sets were combined via random-effects meta-analysis to estimate overall significance. Overall variance explained was estimated as the sample-size weighted mean of target data set-specific pseudo- R^2 estimates.

To assess polygenic sharing between subtypes, we used the same approach, with different subtypes alternately specified as discovery and target traits. There was no sample overlap between discovery and target analyses. Using profile score results, genetic correlations (r_g) were estimated using a quantitative genetics framework.¹² At $\alpha=0.05$, we had 98% power to detect polygenic scores explaining $\geq 0.2\%$ of variance in case/control status for any target subtype, 81% to 83% power to detect scores explaining $\geq 0.1\%$ of variance (varying by subtype), and 52% to 54% power for scores explaining $\geq 0.05\%$ of variance.¹²

Joint Meta-Analysis of LAA and SVD

Joint, fixed-effects meta-analysis of allelic effects for LAA and SVD was performed using Meta¹⁷ for 2167 LAA cases, 1854 SVD cases (4021 total cases), and 51976 controls from 12 studies (Table I in the online-only Data Supplement). To control type 1 error caused by overlapping controls for LAA and SVD within cohorts, a covariance correction was applied.¹⁸ Power to detect associated SNPs was calculated¹⁹ assuming an additive model, perfect linkage disequilibrium between risk and marker alleles, and a significance level of $\alpha=5\times 10^{-8}$. For a genetic risk ratio of 1.2, we had 37%, 89%, and 98% to 99% power to identify risk alleles with frequency 0.1, 0.2, and 0.3 to 0.5, respectively. For a true risk ratio of 1.1, power was low, ranging from 0.2% to 10% across allele frequencies.

Results

Linear Mixed Models

After stringent quality control, individual-level genotype data were available for 4561 IS cases and 7094 controls (Table) at 345336 directly genotyped SNPs. Using LMMs, the estimated proportion of variance in case-control status explained by the SNPs (h^2_{SNP}) was significant for all stroke traits (Table II in the online-only Data Supplement). Higher and more significant values were estimated for IS ($h^2_{\text{SNP}}=0.18$; $P=1\times 10^{-14}$), LAA ($h^2_{\text{SNP}}=0.19$; $P=2\times 10^{-5}$), and cardioembolism ($h^2_{\text{SNP}}=0.24$; $P=2\times 10^{-6}$), whereas the estimate for SVD was lower ($h^2_{\text{SNP}}=0.10$; $P=0.04$).

To estimate the genetic correlation (r_g) between subtype cases, controls were randomly allocated to one of the 2 subtype groups in each analysis. This allocation and estimation process was repeated 10 times and the mean and standard deviation of parameter estimates and mean standard errors (SE) derived (Table III in the online-only Data Supplement). The r_g value was highest and significantly different from zero between LAA and SVD at 0.96 (SD=0.059; $P=9\times 10^{-4}$), although the large standard error (0.47) indicates low precision. Reduced,

but nominally significant correlation was observed between LAA and cardioembolism ($r_g=0.39$, SE=0.21, $P=0.024$) and between cardioembolism and SVD ($r_g=0.64$, SE=0.40, $P=0.017$).

Polygenic Profile Scoring

Although LMM analyses were restricted to samples with individual-level data, profile scoring could use Metastroke samples with summary statistics for discovery meta-analyses. In analyses within traits, profile scores for IS, LAA, and cardioembolism showed strong association with the same trait in independent target cohorts (Tables IV–VI and Figure I in the online-only Data Supplement). For IS, strong association was observed across most of the discovery P value distribution with maximum association observed for $P_T<1$ ($P_{\text{score}}=1.1\times 10^{-8}$), typical of relatively small discovery samples.²⁰ There was little effect size heterogeneity across target cohorts. For LAA, maximum association ($P_{\text{score}}=1.7\times 10^{-8}$) was observed for $P_T<0.05$ with no heterogeneity. For cardioembolism, maximum association ($P_{\text{score}}=2\times 10^{-4}$) was observed for predictors including SNPs reaching $P_T<0.001$ and $P_T<0.01$, with no heterogeneity. For all 3 subtypes, profile scores explained a small proportion of observed case-control variance, being highest for LAA (pseudo- $R^2=0.45\%$) and lowest for SVD (pseudo- $R^2=0.05\%$). We also note that SVD-based scores did not associate with SVD in target cohorts and many showed effect heterogeneity between studies (Table VII in the online-only Data Supplement).

Analyses between stroke subtypes detected significant polygenic sharing between LAA and SVD (Tables VIII and IX and Figure II in the online-only Data Supplement). The majority of SVD-based scores were associated with LAA, with no heterogeneity. The highest association was observed for a score including ≈ 36000 SNPs reaching discovery $P_T<0.1$ ($P_{\text{score}}=2\times 10^{-4}$), which explained an estimated 0.19% of observed LAA case/control variance. In the reverse analysis, 3 LAA-based profile scores were associated with SVD at $P<0.05$, for example, the score including ≈ 20000 SNPs reaching $P_T<0.05$ ($P_{\text{score}}=0.032$, $R^2=0.08\%$). In analyses of the other 2 subtype pairs (LAA and cardioembolism, cardioembolism and SVD), no coassociation of profile scores was observed (Tables X–XIII and Figures III and IV in the online-only Data Supplement).

Using profile score results within LAA, the estimated proportion of LAA variance in liability explained by the score most strongly associated with LAA and SVD ($P_T=0.05$) was 12.8% (Table V in the online-only Data Supplement). SVD-based scores did not associate with SVD in target samples, but the score most

Table. Ischemic Stroke Samples Used for Testing Polygenic Scores

Study	IS	LAA	Cardioembolism	SVD	Controls
ASGC	1071	375	226	287	1212
WTCCC2-Munich	1140	338	322	104	775
WTCCC2-UK	2350	494	450	471	5107
Total	4561	1207	998	862	7094

ASGC indicates Australian Stroke Genetics Collaborative; IS, ischemic stroke; LAA, large artery atherosclerosis; SVD, small vessel disease; and WTCCC2, Wellcome Trust Case Control Consortium 2 Study.

significantly associated with LAA ($P_T=0.1$) explained 0.8% of SVD liability variance (Table VII in the online-only Data Supplement). Using these estimates and the observed cross-trait association results, the estimated genetic correlation¹² between LAA and SVD was $r_g=0.72$, which was significantly different from zero (95% confidence interval [CI], 0.52–0.92). The SNP-based correlation was not significantly different from zero for LAA and cardioembolism ($r_g=0.13$, 95% CI, 0–0.56), or cardioembolism and SVD ($r_g=0.64$, 95% CI: 0–0.92).

Quantitative Bias Analysis: Subtype Misclassification

Bias analysis was performed to assess the extent to which the genetic correlation (r_g) between LAA and SVD could result from subtype misclassification²¹ (see Methods and Table XIV in the online-only Data Supplement). Allowing for rates of subtype misclassification consistent with reported values of inter-rater reliability,²² r_g was still significantly different from zero. Assuming all misclassified LAA cases were truly SVD and vice versa, the estimate was $r_g=0.63$ (95% CI, 0.34–0.74). Assuming all misclassified cases were neither LAA nor SVD, the estimate was $r_g=0.75$ (95% CI, 0.43–0.98). This suggests robustness of the observed genetic correlation to likely levels of subtype misclassification.

GWAS Meta-Analysis of LAA and SVD

Given evidence for shared common variants between LAA and SVD, joint meta-analysis of LAA and SVD was performed (Figures V and VI in the online-only Data Supplement). Although no SNPs reached genome-wide significance ($P<5\times 10^{-8}$), suggestive association ($P=1\times 10^{-7}$ at rs17084671; $P=2\times 10^{-7}$ at rs6938958 and rs7763080) was observed for a cluster of SNPs at chromosome 6q25.2 (Table XV in the online-only Data Supplement), ≈ 100 kb upstream of the opioid receptor $\mu 1$ (*OPRM1*) gene.

Discussion

GWAS of IS have revealed the importance of diagnostic subtype classifications. However, exclusive reliance on discrete subtypes reduces sample size and assumes an absence of risk alleles influencing multiple subtypes. Using some of the largest extant GWAS collections and 2 different analytic approaches, this study suggests the presence of extensive genetic overlap between large artery atherosclerotic and small vessel ischemic stroke. We estimated the genetic correlation between these subtypes exceeds 0.7, but larger samples will increase the accuracy of this estimate.

We were careful to eliminate potential sources of bias in our analyses. In individual-level data, European ancestry was strictly defined and principal components of ancestry included as covariates. We also checked that positive score effects were present in multiple target studies and not driven by a single study. Misclassification of TOAST subtypes was considered an important potential source of error because MRI, which increases diagnostic accuracy particularly for SVD, was only used for subtyping $\approx 50\%$ of all cases.⁵ However, sensitivity analyses suggested that typical rates of misclassification would have minimal effects on the estimated correlation.

Although SVD-based profile scores did not show significant association with SVD, they associated with LAA in the cross-trait analysis. In the complementary analysis, which used the more powerful LAA discovery sample, LAA-based profile scores showed significant association with both LAA and SVD. This does not imply lack of consistency or polygenic architecture for SVD. Indeed, our and a previous LMM analysis¹ detected significant SNP-based heritability for SVD. Profile score analyses are influenced both by sample size and phenotypic homogeneity of discovery and target traits. SVD had the lowest case numbers and is also phenotypically heterogeneous. In profile analyses conducted exclusively within SVD, these factors will reduce both the accuracy of polygenic predictors and statistical power in target samples.

The proportion of observed LAA and SVD variance explained by SVD-based and LAA-based profile scores (pseudo- R^2) was 0.2% and 0.08%, respectively. Although these values are low, this does not mean that the true genetic overlap is small. By combining sampling errors in effect estimates across all SNPs in the score, profile scoring produces estimates of explained variance typically lower than true values,^{20,23} but which will increase as sample size increases. Pseudo- R^2 measures for binary traits can be difficult to interpret because they can depend on ascertainment, that is, the proportion of cases in the sample.²⁴ Profile scoring results were used to estimate liability-scale variance explained and genetic correlations using theory that accounts for sample size and ascertainment. For example, within LAA, the pseudo- R^2 for the maximum profile score was 0.48%, but the estimated LAA variance explained by the score, adjusted for sample size and ascertainment, was 12.8%. Thus, although estimates of observed cross-trait variance explained are small, they can signify a higher genetic correlation. When genetic correlation has been estimated from the same data set, the results from the profile score and LMM agree well.²⁵ Here, the use of the profile score method allowed the use of a larger sample via data sets for which only association summary statistics were available.

The SNPs most strongly associated with the joint LAA-SVD trait were near the *OPRM1* gene, alleles within which have previously shown suggestive association with coronary heart disease ($P=5\times 10^{-6}$),²⁶ which has an atherosclerotic pathogenesis. Estimated genetic correlation between LAA and SVD is also consistent with an atherosclerotic pathogenesis in the majority of LAA and a subset of SVD cases. The primary pathophysiological mechanism for LAA is presumed to be atherosclerosis of the large cerebral arteries.² For SVD, pathological and imaging studies suggest the presence of significant disease heterogeneity, with 2 major underlying vascular pathologies being hypothesized.^{27,28} The first involves localized atherosclerosis of the larger perforating arteries, typically resulting in a larger, isolated lacunar infarct. The second involves diffuse, nonatherosclerotic arteriopathy of the smaller perforating arteries, associated with multiple, smaller infarcts and often coexistent radiological leukoaraiosis.²⁸ Earlier risk factor analyses suggested that conventional atherosclerotic factors were more common in the isolated lacunar infarct subtype. This subgroup could thus account for the genetic overlap between the broader SVD category and LAA.

Conclusions

Our analyses strongly suggest that LAA and SVD—which have been hitherto considered genetically distinct—may have a shared genetic pathogenesis. Further investigation of the genetic relationship between ischemic stroke subtypes is merited. Although recent GWAS have identified several subtype-specific genetic associations, the pace of discovery has been constrained by small numbers for individual subtypes. If there exists small-effect variants influencing multiple subtypes, joint subtype analyses will offer higher power to identify these and may also identify biological mechanisms shared by these traditionally distinct clinical diagnoses.

Sources of Funding

E.G. Holliday was supported by the Australian Heart Foundation and National Stroke Foundation (100071). N.R. Wray was supported by the Australian National Health and Medical Research Council (NHMRC, 613602). H.S. Markus is supported by a National Institute for Health Research (NIHR) Senior Investigator award. H.S. Markus and S. Bevan are supported by the Cambridge University Hospital National Institute for Health Research Biomedical Research Centre (NIHR BRC). J. Rosand and B.B. Worrall received Research Funding from the National Institute of Health. Acknowledgments of funding for the individual Metastroke studies are detailed in the original publication by Traylor et al.⁵

Disclosures

Dr Rosand has a consulting relationship with Boehringer Ingelheim. Dr Worrall is an Associate Editor of the American Academy of Neurology (AAN) journal *Neurology*. All the other authors report no conflict.

References

1. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NL, Jackson C, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genome-wide associations. *Stroke*. 2012;43:3161–3167. doi: 10.1161/STROKEAHA.112.665760.
2. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35–41.
3. Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. An evidence-based causative classification system for acute ischemic stroke. *Ann Neurol*. 2005;58:688–697. doi: 10.1002/ana.20617.
4. Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol*. 2008;64:402–409. doi: 10.1002/ana.21480.
5. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al; Australian Stroke Genetics Collaborative, Wellcome Trust Case Control Consortium 2 (WTCCC2); International Stroke Genetics Consortium. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2012;11:951–962. doi: 10.1016/S1474-4422(12)70234-X.
6. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, et al. Genome-wide association study identifies a variant in *hdac9* associated with large vessel ischemic stroke. *Nat Genet*. 2012;44:328–333.
7. Holliday EG, Maguire JM, Evans TJ, Koblar SA, Jannes J, Sturm JW, et al; Australian Stroke Genetics Collaborative; International Stroke Genetics Consortium; Wellcome Trust Case Control Consortium 2. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet*. 2012;44:1147–1151. doi: 10.1038/ng.2397.

8. Kilarski LL, Achterberg S, Devan WJ, Traylor M, Malik R, Lindgren A, et al; GARNET Collaborative Research Group, Wellcome Trust Case Control Consortium 2, Australian Stroke Genetic Collaborative, the METASTROKE Consortium, and the International Stroke Genetics Consortium. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology*. 2014;83:678–685. doi: 10.1212/WNL.0000000000000707.
9. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, et al. Genomewide association studies of stroke. *N Engl J Med*. 2009;360:1718–1728. doi: 10.1056/NEJMoa0900094.
10. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90:7–24. doi: 10.1016/j.ajhg.2011.11.029.
11. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet*. 2011;88:294–305. doi: 10.1016/j.ajhg.2011.02.002.
12. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9:e1003348. doi: 10.1371/journal.pgen.1003348.
13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795.
14. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82. doi: 10.1016/j.ajhg.2010.11.011.
15. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–909. doi: 10.1038/ng1847.
16. Zhang Y, Chapman AM, Plested M, Jackson D, Purroy F. The incidence, prevalence, and mortality of stroke in France, Germany, Italy, Spain, the UK, and the US: a literature review. *Stroke Res Treat*. 2012;2012:436125. doi: 10.1155/2012/436125.
17. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*. 2010;26:2190–2191. doi: 10.1093/bioinformatics/btq340.
18. Lin DY, Sullivan PF. Meta-analysis of genome-wide association studies with overlapping subjects. *Am J Hum Genet*. 2009;85:862–872. doi: 10.1016/j.ajhg.2009.11.001.
19. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003;19:149–150.
20. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.
21. Wray NR, Lee SH, Kendler KS. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur J Hum Genet*. 2012;20:668–674. doi: 10.1038/ejhg.2011.257.
22. Meschia JF, Barrett KM, Chukwudelunzu F, Brown WM, Case LD, Kissela BM, et al; Siblings with Ischemic Stroke Study (SWISS) Investigators. Interobserver agreement in the trial of org 10172 in acute stroke treatment classification of stroke based on retrospective medical record review. *J Stroke Cerebrovasc Dis*. 2006;15:266–272. doi: 10.1016/j.jstrokecerebrovasdis.2006.07.001.
23. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet*. 2013;14:507–515. doi: 10.1038/nrg3457.
24. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol*. 2012;36:214–224. doi: 10.1002/gepi.21614.
25. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry*. 2014;55:1068–1087. doi: 10.1111/jcpp.12295.
26. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, et al. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARE Project. *PLoS Genet*. 2011;7:e1001300. doi: 10.1371/journal.pgen.1001300.
27. Boiten J, Lodder J, Kessels F. Two clinically distinct lacunar infarct entities? A hypothesis. *Stroke*. 1993;24:652–656.
28. Fisher CM. The arterial lesions underlying lacunes. *Acta Neuropathol*. 1968;12:1–15.

Genetic Overlap Between Diagnostic Subtypes of Ischemic Stroke

Elizabeth G. Holliday, Matthew Traylor, Rainer Malik, Steve Bevan, Guido Falcone, Jemma C. Hopewell, Yu-Ching Cheng, Ioana Cotlarciuc, Joshua C. Bis, Eric Boerwinkle, Giorgio B. Boncoraglio, Robert Clarke, John W. Cole, Myriam Fornage, Karen L. Furie, M. Arfan Ikram, Jim Jannes, Steven J. Kittner, Lisa F. Lincz, Jane M. Maguire, James F. Meschia, Thomas H. Mosley, Mike A. Nalls, Christopher Oldmeadow, Eugenio A. Parati, Bruce M. Psaty, Peter M. Rothwell, Sudha Seshadri, Rodney J. Scott, Pankaj Sharma, Cathie Sudlow, Kerri L. Wiggins, Bradford B. Worrall, Jonathan Rosand, Braxton D. Mitchell, Martin Dichgans, Hugh S. Markus, Christopher Levi, John Attia and Naomi R. Wray

on behalf of the Australian Stroke Genetics Collaborative, the Wellcome Trust Case Control Consortium 2, and the International Stroke Genetics Consortium

Stroke. 2015;46:615-619; originally published online January 22, 2015;

doi: 10.1161/STROKEAHA.114.007930

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Copyright © 2015 American Heart Association, Inc. All rights reserved.

Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://stroke.ahajournals.org/content/46/3/615>

Data Supplement (unedited) at:

<http://stroke.ahajournals.org/content/suppl/2015/01/22/STROKEAHA.114.007930.DC1>

<http://stroke.ahajournals.org/content/suppl/2016/04/10/STROKEAHA.114.007930.DC2>

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

Reprints: Information about reprints can be found online at:

<http://www.lww.com/reprints>

Subscriptions: Information about subscribing to *Stroke* is online at:

<http://stroke.ahajournals.org/subscriptions/>

Online Data Supplement

Genetic overlap between diagnostic subtypes of ischemic stroke

Supplementary methods and results

Quantitative bias analysis – subtype misclassification

For given misclassification rates, the greatest bias will arise under two extreme scenarios. First, if all misclassified LAA cases were truly SVD and vice versa (S1), a proportion of r_g will reflect *within*-subtype correlation, biasing r_g upwards. Alternatively, if all misclassified cases were neither LAA nor SVD (S2), the observed r_g will be typically, but not always, biased downwards¹. We re-estimated r_g under S1 and S2 allowing for misclassification rates consistent with reported inter-rater reliability (Cohen's kappa) values for TOAST subtypes. Averaged across subtypes, kappa statistics generally range from 0.5 to 0.8^{2,3}. One study reported subtype-specific values of 0.8 for LAA and 0.53 for SVD³. Since kappa is a correlation measure, these values were specified as assumed correlations of marker effects between discovery and target samples in the *within*-trait analyses of LAA and SVD during estimation of the within-trait variance ($vg1/h^2$) explained by profile scores. This bias-adjusted estimate of $vg1$ was then used to estimate the genetic correlation (r_g) between LAA and SVD using profile scores from the cross-trait analyses (Table XIV). With kappa=0.8 for LAA and 0.5 for SVD, under S1 r_g was reduced, but still significantly different from zero ($r_g=0.63$, 95% CI: [0.34, 0.74]). Under S2 the estimate was slightly higher ($r_g=0.75$, 95% CI: [0.43, 0.98]). The true misclassification scenario will likely lie between the extremes described by S1 and S2, suggesting robustness of the observed genetic correlation to likely levels of subtype misclassification.

Table I. Metastroke study details

Cohort Name	IS – broadly defined	LAA	CE	SVD	Controls
ARIC	385	31	93	63	8,803
ASGC	1,162	421	240	310	1,244
BRAINS	361	120	29	97	444
CHS	454	–	147	73	2,817
deCODE	2,391	255	399	240	26,970
FHS	171	–	48	–	4,164
GEOS	448	37	90	54	498
HPS	578	–	–	–	468
HVH	566	61	88	173	1,290
ISGS/SWISS	1,070	229	247	201	2,329
MGH/GASROS	516	95	169	38	1,202
Milano	372	74	65	25	407
Rotterdam	367	–	–	–	5,396
WTCCC2-Munich	1,174	346	330	106	797
WTCCC2-UK	2,374	498	460	474	5,175
Total	12,389	2,167	2,405	1,854	62,004

Notes: IS – ischaemic stroke; LAA – large artery atherosclerosis; CE – cardioembolic stroke; SVD – small vessel disease.

Table II. Linear mixed modelling: univariate analyses within subtypes

Trait	Cases	Controls	K ^a	h^2 (SE) liability scale ^b	P ^c
Broad IS	4,561	7,094	0.02	0.18 (0.025)	1×10^{-14}
LAA	1,207	7,094	0.004	0.19 (0.046)	2×10^{-5}
CE	998	7,094	0.004	0.24 (0.053)	2×10^{-6}
SVD	862	7,094	0.004	0.10 (0.060)	0.04

^a Lifetime population prevalence. ^b Estimates based on 345,336 genome-wide SNPs.

^c from a likelihood ratio test of the null hypothesis $H_0: h^2=0$.

Table III. Linear mixed modelling: bivariate analyses (between subtypes)

	Trait 1/Trait 2			
	LAA/CE	LAA/SVD	CE/SVD	Control/Control ^d
Cases	1,207/998	1,207/862	998/862	
Controls	3,547/3,547	3,547/3,547	3,547/3,547	
h^2 trait 1 (SE) ^a	0.18 (0.052)	0.17 (0.052)	0.23 (0.059)	0.00 (0.014)
h^2 trait 2 (SE) ^a	0.23 (0.059)	0.10 (0.066)	0.10 (0.066)	–
r_g (SE) ^a	0.39 (0.21)	0.96 (0.47)	0.64 (0.40)	–
SD (r_g)	0.066	0.059	0.18	–
p^b	0.024	9×10^{-4}	0.017	0.43
p^c	0.012	0.44	0.11	

^a Mean estimate from ten permuted assignments of controls to one of the two case traits. Each of the permuted control assignments was used to estimate of h^2 (trait 1), h^2 (trait 2) and r_g for all three trait combinations. ^b P -value for $H_0: r_g=0$, $H_1: r_g>0$. ^c P -value for $H_0: r_g=1$, $H_1: r_g<1$. ^d The permuted control subsets also allowed us to benchmark the expected h^2_{SNP} by assessing genetic sharing within controls. Across ten control subsets the mean was zero, suggesting lack of inflation by factors unrelated to case-control status.

Table IV. Polygenic scoring results for broadly defined ischaemic stroke (IS)

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score) ^a	Het I^2 ^b	P (het) ^c	R^2 (%) ^d	h^2 ^e
0.0001	73	93	85	<u>0.003</u>	0%	0.53	0.06	0.004
0.001	602	619	627	<u>3.2×10^{-6}</u>	0%	0.44	0.16	0.013
0.01	4,710	4,633	4,726	<u>7.2×10^{-5}</u>	42%	0.18	0.22	0.025
0.05	18,432	18,557	18,933	<u>1.8×10^{-7}</u>	0%	0.88	0.19	0.043
0.1	32,876	32,843	33,548	<u>7.1×10^{-7}</u>	0%	0.47	0.17	0.044
0.2	57,318	57,288	57,969	<u>2.2×10^{-7}</u>	0%	0.60	0.19	0.051
0.3	78,183	77,965	78,660	<u>8.8×10^{-8}</u>	0%	0.88	0.20	0.056
0.4	95,954	95,459	96,220	<u>1.6×10^{-8}</u>	0%	0.88	0.22	0.062
0.5	111,005	110,406	110,944	<u>1.7×10^{-8}</u>	0%	0.83	0.22	0.064
1	154,056	152,907	152,394	<u>1.1×10^{-8}</u>	0%	0.79	0.22	0.074

^a From random effects meta-analysis across the three target cohorts (see Table 1, main paper) using all samples listed in Table I above (except the target cohort) in the discovery meta-analysis. ^b Percent of variation of score effects across studies due to heterogeneity, based on Cochran's Q statistic. ^c Significance of Q statistic. ^d Nagelkerke's R^2 : Sample-size weighted mean of estimates from the three individual studies. ^e IS target sample variance explained by the polygenic score. h^2 was calculated using quantitative genetics theory ⁴ with the observed sample sizes, P_T , N_{SNPs} , $P(\text{score})$, assuming genetic correlation =1 between IS in discovery and target samples, and IS lifetime population prevalence = 0.02 ^{5, 6}.

Table V. Polygenic scoring results for large artery atherosclerosis (LAA)

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score) ^a	Het I^2 _b	P (het) ^c	R^2 (%) ^d	h^2 ^e
0.0001	109	90	90	0.36	0%	0.43	0.024	0.005
0.001	689	650	672	0.24	0%	0.60	0.047	0.015
0.01	5,216	5,099	5,336	<u>9.2×10^{-5}</u>	0%	0.52	0.28	0.072
0.05*	20,158	20,258	20,794	<u>1.7×10^{-8}</u>	0%	0.64	0.48	0.128
0.1	35,765	35,815	36,481	<u>1.9×10^{-7}</u>	0%	0.94	0.40	0.128
0.2	62,393	62,099	63,150	<u>2.5×10^{-7}</u>	0%	0.99	0.41	0.138
0.3	84,922	84,770	85,603	<u>1.5×10^{-7}</u>	0%	0.97	0.41	0.150
0.4	104,363	104,249	104,803	<u>5.0×10^{-8}</u>	0%	0.85	0.43	0.163
0.5	121,026	120,984	121,283	<u>3.8×10^{-8}</u>	0%	0.90	0.45	0.172
1	170,896	170,552	169,957	<u>3.4×10^{-8}</u>	0%	0.85	0.45	0.197

^a From random effects meta-analysis across the three target cohorts (see Table 1, main paper) using all samples listed in Table I above (except the target cohort) in the discovery meta-analysis. ^b Percent of variation of score effects across studies due to heterogeneity, based on Cochran's Q statistic. ^c Significance of Q statistic. ^d Nagelkerke's R^2 : Sample-size weighted mean of estimates from the three individual studies. ^e LAA target sample variance explained by the polygenic score. h^2 was calculated using quantitative genetics theory⁴ with the observed sample sizes, P_T , N_{SNPs} , P (score), assuming genetic correlation =1 between LAA in discovery and target samples, and LAA lifetime population prevalence = 0.004^{5,6}.

* Score showing maximum association with both LAA (above) and SVD (Table VII).

Table VI. Polygenic scoring results for cardioembolic stroke (CE)

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score) ^a	Het I^2 ^b	P (het) ^c	R^2 (%) ^d	h^2 ^e
0.0001	107	101	96	0.23	60%	0.08	0.19	0.006
0.001	664	695	727	<u>1.7×10^{-4}</u>	0%	0.91	0.23	0.034
0.01	4,928	4,961	5,031	<u>2.1×10^{-4}</u>	0%	0.89	0.26	0.066
0.05	19,454	19,437	20,000	<u>0.009</u>	28%	0.25	0.30	0.059
0.1	34,326	34,543	35,109	<u>0.017</u>	29%	0.24	0.27	0.058
0.2	59,723	59,554	60,215	<u>0.049</u>	44%	0.17	0.28	0.052
0.3	80,684	80,550	81,223	0.064	43%	0.18	0.26	0.052
0.4	98,386	98,378	98,563	0.16	57%	0.10	0.27	0.041
0.5	113,435	113,474	113,287	0.13	54%	0.11	0.26	0.046
1	156,404	155,514	155,123	0.18	55%	0.11	0.24	0.046

^a From random effects meta-analysis across the three target cohorts (see Table 1, main paper) using all samples listed in Table I above (except the target cohort) in the discovery meta-analysis. ^b Percent of variation of score effects across studies due to heterogeneity, based on Cochran's Q statistic. ^c Significance of Q statistic. ^d Nagelkerke's R^2 : Sample-size weighted mean of estimates from the three individual studies. ^e CE target sample variance explained by the polygenic score. h^2 was calculated using quantitative genetics theory ⁴ with the observed sample sizes, P_T , N_{SNPs} , $P(\text{score})$, assuming genetic correlation =1 between CE in discovery and target samples, and CE lifetime population prevalence = 0.004 ^{5, 6}.

Table VII. Polygenic scoring results for small vessel disease (SVD)

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score) ^a	Het I^2 ^b	P (het) ^c	R^2 (%) ^d	h^2 ^e
0.0001	91	86	114	0.51	0%	0.61	0.018	0.005
0.001	677	620	728	0.95	0%	0.85	0.0019	0.001
0.01	5,114	4,861	5,449	0.92	0%	0.47	0.022	0.003
0.05	20,332	19,636	20,974	0.91	0%	0.67	0.010	0.003
0.1*	35,767	35,057	36,383	0.78	20%	0.29	0.048	0.008
0.2	61,832	60,873	62,892	0.97	38%	0.20	0.058	0.001
0.3	83,691	82,827	84,725	0.76	52%	0.13	0.069	0.010
0.4	102,456	101,697	103,063	0.90	36%	0.21	0.052	0.004
0.5	118,290	117,688	118,447	0.94	34%	0.22	0.046	0.003
1	163,946	163,305	162,329	0.87	42%	0.18	0.053	0.007

^a From random effects meta-analysis across the three target cohorts (see Table 1, main paper) using all samples listed in Table I above (except the target cohort) in the discovery meta-analysis. ^b Percent of variation of score effects across studies due to heterogeneity, based on Cochran's Q statistic. ^c Significance of Q statistic. ^d Nagelkerke's R^2 : Sample-size weighted mean of estimates from the three individual studies. ^e SVD target sample variance explained by the polygenic score. h^2 was calculated using quantitative genetics theory ⁴ with the observed sample sizes, P_T , N_{SNPs} , $P(\text{score})$, assuming genetic correlation =1 between SVD in discovery and target samples, and SVD lifetime population prevalence = 0.004 ^{5, 6}.

* Score showing maximum association with LAA (Table IX).

Table VIII. Pleiotropy analyses: association of LAA-based polygenic scores with SVD

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score)	Het f^2	P (het)	Nagelkerke's R^2 (%)
0.0001	109	90	90	0.35	0%	0.54	0.031
0.001	689	650	672	0.28	0%	0.50	0.029
0.01	5,216	5,099	5,336	0.15	34%	0.22	0.090
0.05	20,158	20,258	20,794	<u>0.032</u>	0%	0.59	0.082
0.1	35,765	35,815	36,481	0.060	0%	0.39	0.075
0.2	62,393	62,099	63,150	0.069	36%	0.21	0.104
0.3	84,922	84,770	85,603	0.064	13%	0.32	0.080
0.4	104,363	104,249	104,803	<u>0.039</u>	0%	0.37	0.080
0.5	121,026	120,984	121,283	<u>0.038</u>	0%	0.42	0.085
1	170,896	170,552	169,957	0.054	0%	0.46	0.069

Table IX. Pleiotropy analyses: association of SVD-based polygenic scores with LAA

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score)	Het f^2	P (het)	Nagelkerke's R^2 (%)
0.0001	91	86	114	0.75	41%	0.18	0.038
0.001	677	620	728	0.51	0%	0.67	0.010
0.01	5,114	4,861	5,449	0.23	0%	0.94	0.022
0.05	20,332	19,636	20,974	<u>2.4×10^{-4}</u>	0%	0.93	0.197
0.1	35,767	35,057	36,383	<u>2.1×10^{-4}</u>	0%	0.81	0.192
0.2	61,832	60,873	62,892	<u>0.001</u>	0%	0.96	0.161
0.3	83,691	82,827	84,725	<u>0.001</u>	0%	0.86	0.150
0.4	102,456	101,697	103,063	<u>0.001</u>	0%	0.88	0.154
0.5	118,290	117,688	118,447	<u>0.001</u>	0%	0.88	0.153
1	163,946	163,305	162,329	<u>0.001</u>	0%	0.80	0.141

Table X. Pleiotropy analyses: association of LAA-based polygenic scores with CE

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score)	Het f^2	P (het)	Nagelkerke's R^2 (%)
0.0001	109	90	90	0.58	73%	0.03	0.083
0.001	689	650	672	0.32	0%	0.64	0.018
0.01	5,216	5,099	5,336	0.48	0%	0.55	0.023
0.05	20,158	20,258	20,794	0.18	0%	0.95	0.029
0.1	35,765	35,815	36,481	0.27	13%	0.32	0.048
0.2	62,393	62,099	63,150	0.21	0%	0.53	0.037
0.3	84,922	84,770	85,603	0.19	0%	0.71	0.035
0.4	104,363	104,249	104,803	0.15	0%	0.80	0.032
0.5	121,026	120,984	121,283	0.13	0%	0.88	0.036
1	170,896	170,552	169,957	0.15	0%	0.97	0.038

Table XI. Pleiotropy analyses: association of CE-based polygenic scores with LAA

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score)	Het f^2	P (het)	Nagelkerke's R^2 (%)
0.0001	107	101	96	0.63	0%	0.48	0.015
0.001	664	695	727	0.99	0%	0.81	0.003
0.01	4,928	4,961	5,031	0.88	0%	0.79	0.001
0.05	19,454	19,437	20,000	0.73	42%	0.18	0.040
0.1	34,326	34,543	35,109	0.83	68%	0.04	0.065
0.2	59,723	59,554	60,215	0.85	41%	0.18	0.031
0.3	80,684	80,550	81,223	0.99	26%	0.26	0.023
0.4	98,386	98,378	98,563	0.71	44%	0.17	0.038
0.5	113,435	113,474	113,287	0.57	42%	0.18	0.034
1	156,404	155,514	155,123	0.70	51%	0.13	0.042

Table XII. Pleiotropy analyses: association of CE-based polygenic scores with SVD

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score)	Het f^2	P (het)	Nagelkerke's R^2 (%)
0.0001	107	101	96	0.068	0%	0.43	0.105
0.001	664	695	727	0.31	8%	0.34	0.060
0.01	4,928	4,961	5,031	0.32	0%	0.85	0.018
0.05	19,454	19,437	20,000	0.85	0%	0.48	0.021
0.1	34,326	34,543	35,109	0.62	50%	0.14	0.082
0.2	59,723	59,554	60,215	0.63	53%	0.12	0.105
0.3	80,684	80,550	81,223	0.33	40%	0.19	0.124
0.4	98,386	98,378	98,563	0.50	58%	0.09	0.170
0.5	113,435	113,474	113,287	0.45	57%	0.10	0.170
1	156,404	155,514	155,123	0.36	54%	0.11	0.176

Table XIII. Pleiotropy analyses: association of SVD-based polygenic scores with CE

P_T	N_{SNPs} ASGC	N_{SNPs} WTCCC2-Mun	N_{SNPs} WTCCC2-UK	P (score)	Het f^2	P (het)	Nagelkerke's R^2 (%)
0.0001	91	86	114	0.15	0%	0.60	0.034
0.001	677	620	728	0.74	52%	0.13	0.057
0.01	5,114	4,861	5,449	0.89	22%	0.28	0.033
0.05	20,332	19,636	20,974	0.42	0%	0.46	0.026
0.1	35,767	35,057	36,383	0.53	5%	0.35	0.031
0.2	61,832	60,873	62,892	0.57	19%	0.29	0.029
0.3	83,691	82,827	84,725	0.56	0%	0.41	0.021
0.4	102,456	101,697	103,063	0.56	0%	0.38	0.023
0.5	118,290	117,688	118,447	0.58	0%	0.43	0.020
1	163,946	163,305	162,329	0.60	0%	0.50	0.015

Table XIV. Bias analysis: genetic correlations between LAA and SVD allowing for varying levels of subtype misclassification

Kappa LAA	Kappa SVD	Scenario 1 ^a	Scenario 2 ^b
		r_g (95% CI)	r_g (95% CI)
1	1	0.72 (0.51,0.92) ^c	0.72 (0.51,0.92)
0.9	0.9	0.68 (0.51,0.86)	0.71 (0.51,0.90)
0.9	0.8	0.67 (0.51,0.84)	0.72 (0.51,0.92)
0.9	0.7	0.66 (0.50,0.81)	0.73 (0.52,0.95)
0.9	0.6	0.65 (0.43,0.79)	0.75 (0.48,0.99)
0.9	0.5	0.64 (0.36,0.77)	0.78 (0.41,0.99)
0.8	0.8	0.65 (0.51,0.80)	0.69 (0.51,0.87)
0.8	0.7	0.64 (0.47,0.78)	0.71 (0.51,0.90)
0.8	0.6	0.63 (0.40,0.76)	0.72 (0.51,0.93)
0.8	0.5	0.62 (0.33,0.74)	0.75 (0.43,0.98)
0.7	0.7	0.62 (0.44,0.74)	0.68 (0.51,0.85)
0.7	0.6	0.62 (0.37,0.73)	0.70 (0.51,0.88)
0.7	0.5	0.60 (0.31,0.71)	0.71 (0.45,0.92)
0.6	0.6	0.60 (0.35,0.69)	0.67 (0.51,0.82)
0.6	0.5	0.59 (0.29,0.68)	0.68 (0.49,0.86)
0.5	0.5	0.57 (0.26,0.65)	0.65 (0.51,0.80)

Notes: Each r_g is the average of estimates from two directional polygenic profiling analyses: i) LAA used as derivation trait and profile scores assessed for association with SVD, and; ii) SVD used as derivation trait and profile scores assessed for association with LAA.

^a Under Scenario 1, all misclassified LAA cases were truly SVD, and vice versa. For the derivation trait, the relevant kappa was specified as the within-trait correlation of genetic effect sizes. This resulted in proportional inflation of the total fraction of trait1 variance explained by genetic effects (vg1) in the profile score. That is, when misclassification exists, the true proportion of trait1 variance explained by the score would be higher. The genetic correlation between genetic effects across traits (r_g) was estimated using the adjusted vg1 and an adjusted estimate of trait2 variance explained by the score (vg2) which was similarly inflated upwards by that trait's kappa value. That is, under this scenario of misclassification, the estimate of trait 2 variance explained by the score would be higher since a component of vg2 now reflects spurious correlation of genetic effects within trait 1. This proportional inflation of both vg1 and vg2 reduces the estimated genetic correlation consistent with the observed association statistics.

^b Under Scenario 2, all misclassified cases were neither LAA nor SVD. For the derivation trait, vg1 was proportionally inflated as before. For the validation trait, vg2 was reduced to the proportion represented by the relevant kappa value. That is, under this misclassification scenario, the estimate of trait 2 variance explained by the score is lower, since the misclassified proportion of cases are of a different subtype to the derivation trait.

^c Reported value, assuming no subtype misclassification

Table XV. Association results for SNPs reaching $P < 1 \times 10^{-5}$ in joint meta-analysis of LAA and SVD

SNP	CHR	BP	A1	Freq	OR	95% CI	P	Het r^2 (%)	Het P	Nearest Gene
rs11155944	6	154,245,875	t	0.24	0.84	(0.78, 0.89)	1.6E-07	21	0.25	<i>OPRM1</i>
rs932671	6	154,258,051	a	0.25	0.84	(0.79, 0.89)	2.3E-07	21	0.25	<i>OPRM1</i>
rs9371764	6	154,259,797	t	0.25	0.86	(0.81, 0.92)	5.1E-06	57	0.01	<i>OPRM1</i>
rs17084671	6	154,261,451	a	0.24	0.84	(0.78, 0.89)	1.3E-07	15	0.31	<i>OPRM1</i>
rs6938958	6	154,262,120	a	0.76	1.19	(1.11, 1.26)	1.8E-07	20	0.26	<i>OPRM1</i>
rs7763080	6	154,263,321	a	0.25	0.84	(0.78, 0.89)	1.7E-07	17	0.28	<i>OPRM1</i>
rs12429886	13	46,092,686	t	0.09	1.27	(1.15, 1.39)	1.4E-06	48	0.04	<i>LRCH1</i>
rs7983635	13	46,093,268	a	0.92	0.79	(0.71, 0.87)	7.2E-06	52	0.03	<i>LRCH1</i>
rs12427953	13	46,101,896	t	0.93	0.79	(0.71, 0.87)	6.0E-06	50	0.03	<i>LRCH1</i>
rs12429970	13	46,138,327	c	0.93	0.78	(0.69, 0.86)	2.3E-06	52	0.03	<i>LRCH1</i>
rs1483968	14	22,021,733	t	0.075	1.35	(1.18, 1.53)	9.4E-06	0	0.44	<i>TCRA</i>
rs8072419	17	46,982,134	a	0.20	1.16	(1.08, 1.23)	6.5E-06	0	0.80	<i>CA10</i>
rs134197	22	26,993,507	a	0.12	1.22	(1.12, 1.33)	4.8E-06	33	0.14	<i>TTC28</i>
rs6519761	22	27,431,600	a	0.86	1.20	(1.10, 1.29)	9.4E-06	0	0.92	<i>CHEK2</i>
rs1884816	22	27,436,733	t	0.84	1.20	(1.11, 1.29)	2.7E-06	0	0.95	<i>CHEK2</i>

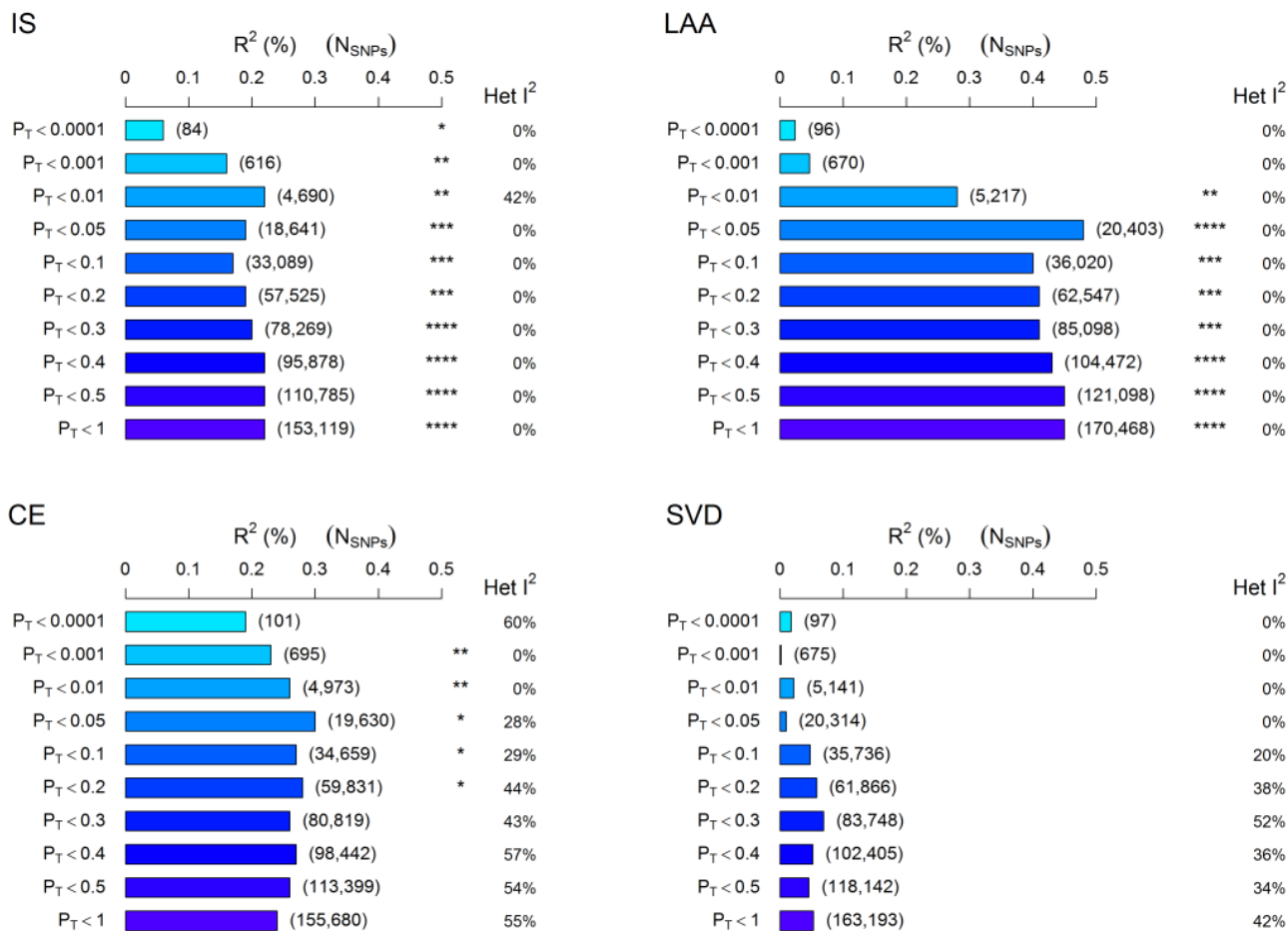


Figure I. Results of polygenic profile scoring within ischemic stroke (IS) and its three major subtypes: large artery atherosclerosis (LAA), cardioembolism (CE) and small vessel disease (SVD). P_T : discovery threshold for including SNPs in the score. R^2 (%): percentage of observed case-control variance explained by the score in target samples. N_{SNPs} : number of approximately independent SNPs included in the score. I^2 : heterogeneity of score effects between target cohorts. * $P \leq 0.05$, ** $P \leq 0.001$, *** $P \leq 1 \times 10^{-5}$, **** $P \leq 1 \times 10^{-7}$.

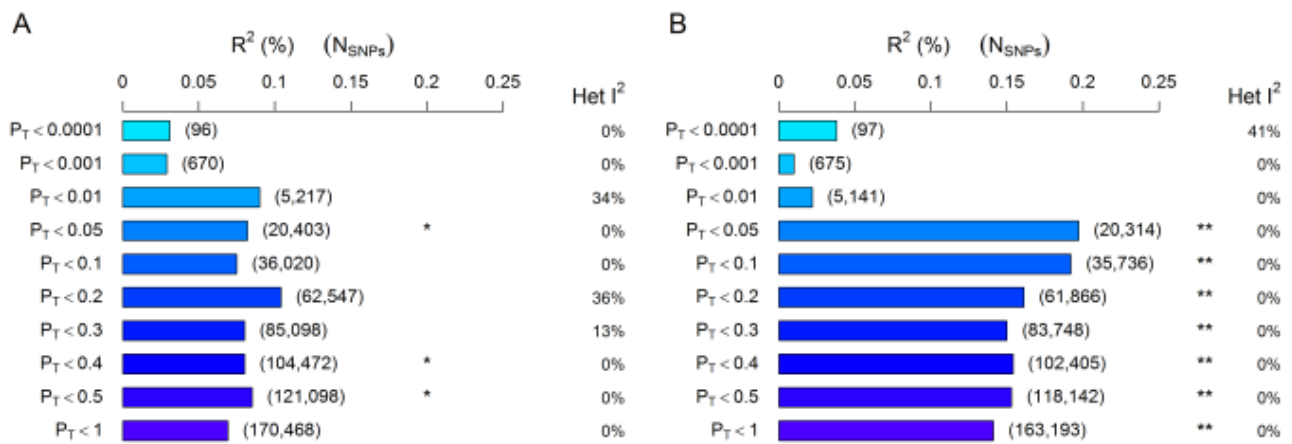


Figure II. Polygenic profile scoring between large artery atherosclerotic stroke (LAA) and small vessel disease (SVD). **A.** Scores derived from meta-analyses of LAA and tested for association with SVD. **B.** Profile scores derived from meta-analyses of SVD and tested for association with LAA. PT: threshold for including SNPs in the score. R² (%): estimated percentage of target sample case-control variance explained by the score using Nagelkerke’s pseudo-R² measure. N_{SNPs}: number of approximately independent SNPs included in the score. I²: heterogeneity of score effects between target cohorts. * $P \leq 0.05$, ** $P \leq 0.001$

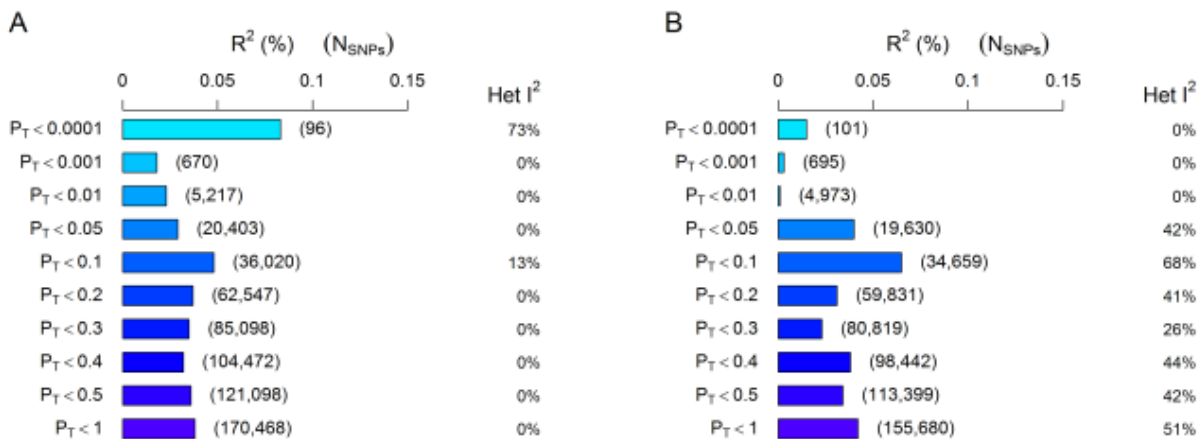


Figure III. Polygenic profile sharing results for LAA and CE. **A:** Scores derived from meta-analyses of LAA and tested for association with CE. **B:** Scores derived from meta-analyses of CE and tested for association with LAA. No score associated with the target trait at $P < 0.05$.

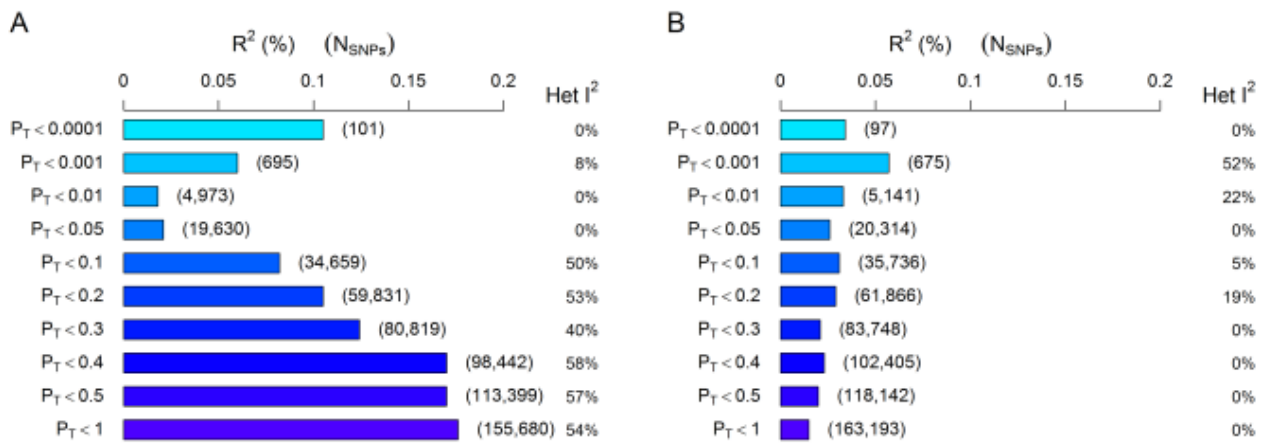


Figure IV. Polygenic profile sharing results for CE and SVD. **A:** meta-analysis performed for CE and profile scores tested for association with SVD. **B:** meta-analysis performed for SVD and derived profile scores tested for association with CE. No score associated with the target trait at $P < 0.05$.

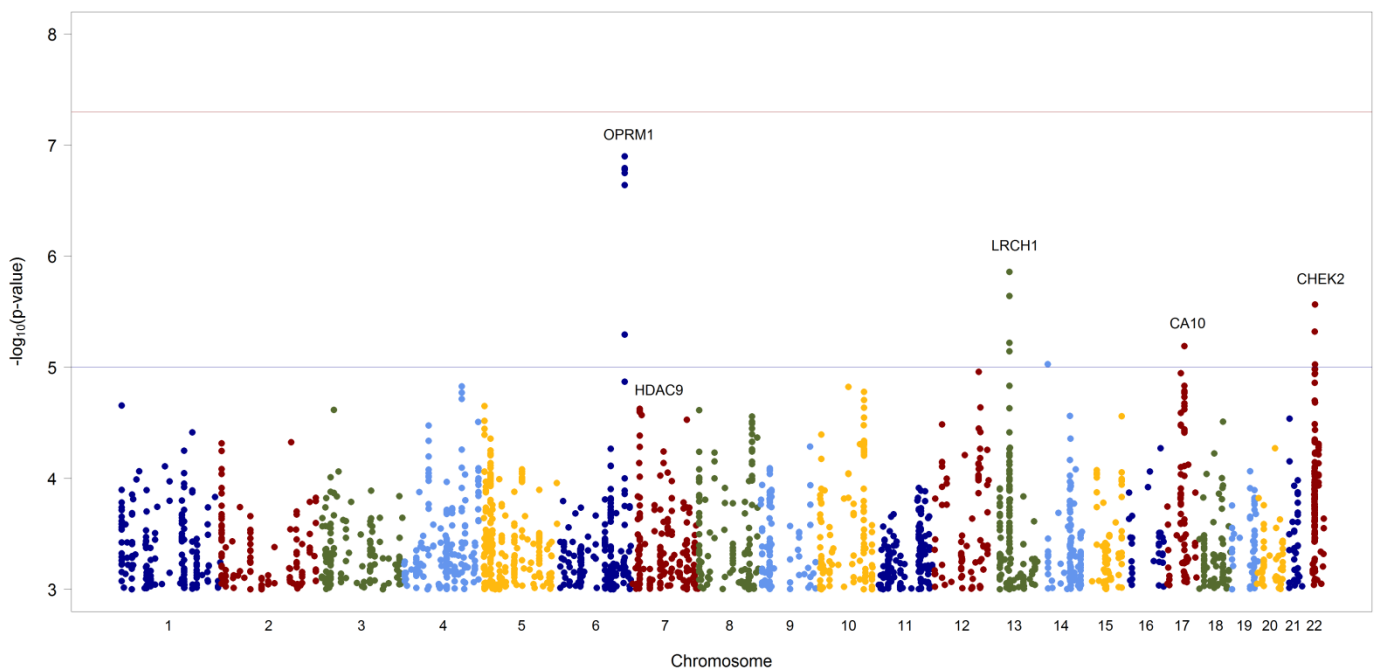


Figure V. Association results for the joint LAA-SVD meta-analysis including 4,021 LAA/SVD cases and 51,976 controls. The plot shows results for SNPs reaching $P \leq 1 \times 10^{-3}$. The upper red line corresponds to $P = 5 \times 10^{-8}$ and the lower grey line corresponds to $P = 1 \times 10^{-5}$.

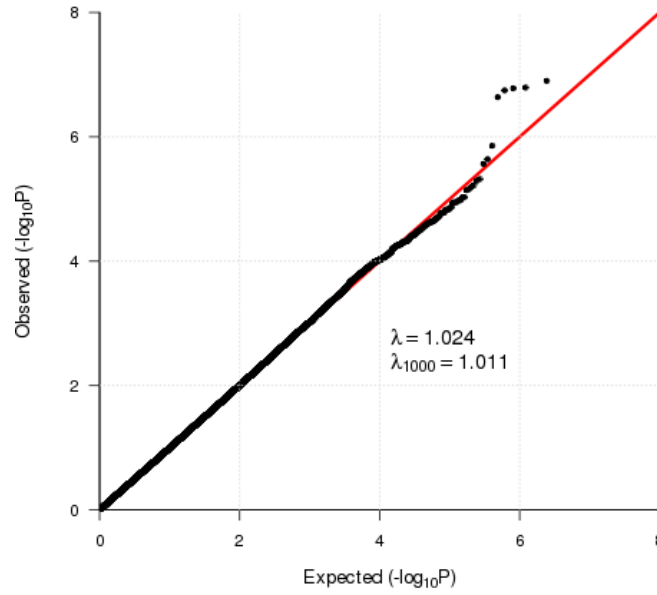


Figure VI. QQ-plot of association results from the joint meta-analysis of LAA and SVD. λ_{1000} is the genomic control factor scaled to 1000 cases and 1000 controls.

References

1. Rothman KJ, Greenland S, Lash TL. In: Rothman KJ, ed. *Modern epidemiology*. Philadelphia, PA USA: Lippincott Williams & Wilkins; 2008:138-139.
2. Goldstein LB, Jones MR, Matchar DB, Edwards LJ, Hoff J, Chilukuri V, et al. Improving the reliability of stroke subgroup classification using the trial of org 10172 in acute stroke treatment (toast) criteria. *Stroke*. 2001;32:1091-1098
3. Meschia JF, Barrett KM, Chukwudelunzu F, Brown WM, Case LD, Kissela BM, et al. Interobserver agreement in the trial of org 10172 in acute stroke treatment classification of stroke based on retrospective medical record review. *J Stroke Cerebrovasc Dis*. 2006;15:266-272
4. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9:e1003348
5. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al. Genetic risk factors for ischaemic stroke and its subtypes (the metastroke collaboration): A meta-analysis of genome-wide association studies. *Lancet Neurol*. 2012;11:951-962
6. Zhang Y, Chapman AM, Plested M, Jackson D, Purroy F. The incidence, prevalence, and mortality of stroke in france, germany, italy, spain, the uk, and the us: A literature review. *Stroke Res Treat*. 2012;2012:436125

Генетическое сходство между пациентами с различными вариантами ишемического инсульта

Источник: E.G. Holliday, M. Traylor, R. Malik, S. Bevan, G. Falcone, J.C. Hopewell, Y.-C. Cheng, I. Cotlarciuc, J.C. Bis, E. Boerwinkle, G.B. Boncoraglio, R. Clarke, J.W. Cole, M. Fornage, K.L. Furie, M.A. Ikram, J. Jannes, S.J. Kittner, L.F. Lincz, J.M. Maguire, J.F. Meschia, T.H. Mosley, M.A. Nalls, C. Oldmeadow, E.A. Parati, B.M. Psaty, P.M. Rothwell, S. Seshadri, R.J. Scott, P. Sharma, C. Sudlow, K.L. Wiggins, B.B. Worrall, J. Rosand, B.D. Mitchell, M. Dichgans, H.S. Markus, C. Levi, J. Attia, N.R. Wray, on behalf of the Australian Stroke Genetics Collaborative, the Wellcome Trust Case Control Consortium 2, and the International Stroke Genetics Consortium. Genetic overlap between diagnostic subtypes of ischemic stroke. *Stroke* 2015;46:3:615–619

School of Medicine and Public Health, School of Biomedical Sciences and Pharmacy, and School of Nursing and Midwifery, University of Newcastle, Australia; Clinical Research Design, IT and Statistical Support Unit, Public Health Research Program, Hunter Medical Research Institute, Newcastle, Australia; Department of Clinical Neurosciences, University of Cambridge, UK; Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany; Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; Clinical Trial Service Unit and Epidemiological Studies Unit and Stroke Prevention Research Unit, Nuffield Department of Clinical Neuroscience, University of Oxford, UK; Department of Medicine and Department of Neurology, University of Maryland School of Medicine, Baltimore, MD; Imperial College Cerebrovascular Research Unit, Imperial College, London, UK; Cardiovascular Health Research Unit, Department of Medicine and Department of Epidemiology, University of Washington, Seattle, WA; Institute of Molecular Medicine and Human Genetics Center, the University of Texas Health Science Center at Houston, TX; Department of Cerebrovascular Disease, IRCCS Istituto Neurologico Carlo Besta, Milan, Italy; Department of Neurology, Veterans Affairs Medical Center, Baltimore, MD; Department of Neurology, Rhode Island Hospital, Providence, RI; Departments of Epidemiology, Neurology, and Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands; School of Medicine, University of Adelaide, Australia; Hunter Haematology Research Group, Calvary Mater Newcastle Hospital, Newcastle, Australia; Hunter Medical Research Institute, Newcastle, Australia; Department of Neurology, Mayo Clinic, Jacksonville, FL; Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, MS; Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD; Group Health Research Institute, Group Health, Seattle, WA; Department of Neurology, Boston University School of Medicine, Boston, MA; Pathology North, NSW Pathology and Department of Neurology, John Hunter Hospital, Newcastle, Australia; Institute of Cardiovascular Research, Royal Holloway University of London (ICR2UL) & Ashford and St Peters Hospital; Division of Clinical Neurosciences, University of Edinburgh, UK; Departments of Neurology and Public Health Sciences, University of Virginia, Charlottesville, VA; Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; and Queensland Brain Institute, The University of Queensland, Australia.

Приглашенный редактор: Jeffrey L. Saver, MD.

Дополнительные данные доступны on-line по адресу:

<http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.114.007930/-/DC1>

Предпосылки и цель исследования. Несмотря на умеренную наследуемость, фенотипическая гетерогенность ишемического инсульта препятствует открытию каузативных генов, тем самым стимулируя проведение анализа диагностических вариантов при небольших размерах выборки. Мы оценили доказательства общей генетической основы трех основных вариантов инсульта: атеросклероза крупных артерий (АКА), кардиоэмболии и болезни мелких сосудов (БМС) для проведения потенциальных перекрестных анализов вариантов инсульта. **Методы.** Для проведения анализа использовали полногеномные обобщенные данные 12 389 пациентов с ишемическим инсультом (в т.ч. 2167 случаев АКА, 2405 случаев кардиоэмболии и 1854 случаев БМС) и 62 004 лиц контрольной группы из консорциума Metastroke. Также были доступны данные индивидуальных генотипов 4561 пациента с инсультом и 7094 лиц контрольной группы. Генетические корреляции между отдельными вариантами инсульта изучали с использованием линейных смешанных моделей и оценки полигенных профилей. Для выявления общих аллелей риска провели мета-анализ сочетанного фенотипа АКА-БМС (4021 пациент и 51 976 лиц контрольной группы). **Результаты.** Высокую генетическую корреляцию выявили между АКА и БМС с использованием линейных смешанных моделей ($r_g=0,96$, $CO=0,47$, $p=9 \times 10^{-4}$) и оценки профилей ($r_g=0,72$; 95% доверительный интервал от 0,52 до 0,93). Между АКА и кардиоэмболией и БМС и кардиоэмболией корреляция была умеренной при использовании линейных смешанных моделей, но незначительно отличалась от нуля при использовании оценки профилей. При проведении обобщенного мета-анализа АКА и БМС выявили сильную ассоциацию ($p=1 \times 10^{-7}$) для однонуклеотидного полиморфизма около гена опиоидных рецепторов $\mu 1$ (OPRM1). **Выводы.** Наши результаты показывают, что АКА и БМС, которые до сих пор считали генетически различными заболеваниями, могут иметь существенный общий генетический компонент. Комбинированные анализы АКА и БМС позволяют увеличить мощность исследований для определения аллелей с небольшим эффектом, влияющих на общие патофизиологические процессы.

Ключевые слова: атеросклероз (atherosclerosis), генетическая эпидемиология (genetic epidemiology), лакунарный инсульт (lacunar stroke)

Ишемический инсульт (ИИ) является сложным заболеванием, на развитие которого оказывают влияние многочисленные клинические, генетические факторы и факторы риска образа жизни. Хорошо известны традици-

онные факторы риска, такие как артериальная гипертензия, дислипидемия, сахарный диабет и курение, однако генетические факторы, на долю которых приходится до 30–40% риска [1], изучены недостаточно. Несмотря на последние достижения в области генотипирования с высокой пропускной способностью, прогресс в области открытия гена ИИ развивается медленно из-за технического характера регистрации случаев ИИ и его этиологи-

© American Heart Association, Inc., 2015

Адрес для корреспонденции: Elizabeth Holliday, PhD, CReDITSS, HMRI, C/- University Dr, Callaghan, NSW 2308, Australia.

E-mail: liz.holliday@newcastle.edu.au

ческой гетерогенности. Последнее обуславливает патофизиологические различия и подразумевает наличие генетических различий между пациентами, что затрудняет выявление генов предрасположенности к ИИ.

Для облегчения диагностики и клинического ведения были разработаны схемы классификации ИИ на диагностические варианты [2, 3]. К основным вариантам относятся атеросклероз крупных артерий (АКА), кардиоэмболия и болезнь мелких сосудов (лакунарный ИИ, БМС). При использовании этой достаточно фенотипически однородной категоризации в исследованиях полногеномных ассоциаций (ИПГА) выявили несколько генетических ассоциаций, характерных для отдельных вариантов ИИ [4–7]. В отличие от этого при более обобщенном определении ИИ были выявлены только 2 значимые полногеномные ассоциации [8, 9], несмотря на использование выборок с 5-кратным увеличением размера. Мощность исследования ассоциации заключается в равновесии между размером выборки и размером эффекта локуса риска, при этом генетическая однородность оказывает влияние на показатели размера эффекта. На сегодняшний день ИПГА ИИ были сосредоточены на отдельных вариантах инсульта, что привело к уменьшению размера выборок и снижению мощности некоторых локусов при наличии общей генетической основы между вариантами ИИ [10].

Цель настоящего исследования заключалась в изучении генетических корреляций между тремя основными вариантами ИИ с использованием данных ИПГА отдельных пациентов и обобщенных статистических показателей мета-анализа международного консорциума по генетике инсульта [6, 7] и консорциума *MetaStroke* [5]. Генетические корреляции изучали с использованием двух различных методов: линейные смешанные модели (ЛСМ) [11] и оценка полигенных профилей [12].

■ МЕТОДЫ

Источники данных

В исследование *MetaStroke* включили 15 отдельных исследований с участием 12 389 пациентов с ИИ и 62 004 лиц контрольной группы европеоидной расы (таблица I в дополнительных данных *on-line*). Детали проведения этих 15 исследований, в т.ч. методы генотипирования, фенотипирования и демографические данные участников были подробно описаны ранее [5]. Пациенты и лица контрольной группы не участвовали в проведении каких-либо исследований, что было подтверждено данными генотипирования. Случаи инсульта классифицировали по вариантам с использованием системы TOAST [2], при этом идентифицировали 2167 случаев ИИ на фоне АКА, 2405 случаев кардиоэмболии и 1854 случаев БМС; у остальных пациентов был другой, неопределенный или криптогенный генез ИИ. В каждом исследовании определяли генотип пациента с использованием стандартизированных панелей *HarMap Phase 2* или 1000 *Genomes*, соответствующие дополнительным моделям логистической регрессии для всех однонуклеотидных полиморфизмов (ОНП), и представили регрессионные статистические данные для ИИ и его вариантов

(по возможности). Также были доступны данные ИПГА отдельных пациентов в трех крупнейших выборках консорциума *MetaStroke*: 2 из *Wellcome Trust Case Control Consortium 2 Study* (Великобритания и Мюнхен) [6] и *Australian Stroke Genetics Collaborative* [7], которые генотипировали с использованием панелей *Illumina* с аналогичным содержанием. Все исследования были одобрены соответствующими комитетами по этике, и участники дали письменное информированное согласие.

Линейные смешанные модели

Данные о генотипах из трех выборок с данными отдельных пациентов объединили для получения единого набора данных с помощью программного обеспечения *PLINK* [13]. Соблюдая строгий контроль качества удалили ОНП, которые не генотипировали во всех выборках, с $>0,5\%$ недостающих данных, значением p Харди-Вайнберга $<0,05$, частотой минорных аллелей $<1\%$ или дифференциальным пропуском ($p < 0,05$) между выборками. Мы исключили выборки с $>1\%$ недостающих данных и одно из каждой пары с абсолютным значением полногеномного сходства $>0,05$ [14]. Основные компоненты происхождения рассчитали в обобщенной выборке после трех повторений анализа главных компонентов с удалением резко отклоняющихся значений (>5 стандартных отклонений от среднего на PC1–5) [15].

Наследственность и генетические корреляции (r) между подтипами оценивали с использованием *LSCM* [14] с внесением поправок на 20 главных компонентов. Статистику отношения правдоподобия применяли для проверки значимого отклонения показателей от нуля. Показатели наследуемости были преобразованы в шкалу ответственности, предполагающей распространенность заболевания в течение жизни 2% для ИИ, при этом на долю каждого из трех основных подтипов приходится по $\approx 20\%$ (всего 60%), соразмерных $0,4\%$ распространенности ($20\% \times 2\%$) каждого подтипа [5, 16]. Оставшаяся распространенность $0,8\%$ ($40\% \times 2\%$) предположительно отражает другие подтипы инсульта.

Оценка полигенного профиля

При оценке полигенного профиля использовали статистику ассоциации ОНП из представленного фенотипа для построения линейного предиктора и проверки этой ассоциации с таким же или другим фенотипом в независимой выборке данных. Для облегчения интерпретации результатов перекрестного анализа вариантов инсульта мы сначала оценили ассоциацию оценки профилей с ИИ и его тремя вариантами. В каждом случае для этого использовали проверку с исключением по одному из трех наборов данных генотипов отдельных пациентов. В свою очередь каждый из трех вариантов не принимали в расчет и проводили мета-анализ данных ИПГА с использованием всех других наборов данных консорциума *MetaStroke*. ОНП с данными из ≥ 5 исследований *MetaStroke* сохраняли и упрощали для неравновесного сцепления ($r^2 > 0,2$ в *1Mb*), используя групповой алгоритм программного обеспечения *PLINK* [13], который сохраняет наиболее ассоциированный ОНП в области неравновесного сцепления. Из упрощенного набора данных мы извлекли подмно-

жества, прошедшие 10 дифференцированных пороговых значений ($p_T=1 \times 10^{-4}$, 1×10^{-3} , 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 1). Для подмножеств, прошедших каждое пороговое значение, функцию оценки программного обеспечения PLINK использовали для расчета оценки профилей для лиц из набора данных с пропущенными значениями. Эти показатели представляют собой среднюю нагрузку аллелей риска всех ОНП в оценке, с долями, выраженными в виде логарифмически преобразованных отношений шансов из мета-анализа обнаружения.

Ассоциации оценок профилей с вариантами инсульта оценивали с использованием логистической регрессии с внесением поправок на 3 главных компоненты происхождения. Дисперсию, объясняемую оценками, рассчитали как разность псевдо- R^2 Найджелкерка между моделью, в которую включили оценку профилей и основные компоненты, и моделью с включением только основных компонент. Результаты, полученные при анализе трех целевых наборов данных, объединили с помощью мета-анализа случайных эффектов для оценки общей значимости. Общую, объясняемую дисперсию оценивали как среднее взвешенное значение выборки псевдо- R^2 показателей целевого набора данных.

Для оценки полигенной общности между пациентами с различными вариантами ИИ мы использовали тот же подход, при этом различные варианты поочередно считали открытием или специфическими чертами. При проведении анализов обнаружения и целевых анализов выборки не совпадали. Используя результаты оценки профилей оценили генетические корреляции (r_g) с помощью концепции генетики количественных признаков [12]. При $\alpha=0,05$ мы получили мощность 98% для обнаружения полигенных оценок, объясняющих $\geq 0,2\%$ дисперсии при сравнении типа случай/контроль для любого из изучаемых вариантов инсульта, мощность 81–83% для обнаружения оценок, объясняющих $\geq 0,1\%$ дисперсии (в зависимости от варианта), и мощность 52–54% для обнаружения оценок, объясняющих $\geq 0,05\%$ дисперсии [12].

Обобщенный мета-анализ АКА и БМС

Обобщенный мета-анализ фиксированных аллельных эффектов для АКА и БМС проводили с использованием Metal [17] для 2167 случаев АКА, 1854 случаев БМС (всего 4021 случай) и данных 51 976 лиц контрольной группы из 12 исследований (таблица I в дополнительных данных *on-line*). Для контроля ошибки 1-го типа, связанной с совпадением данных лиц контрольной группы для АКА и БМС в пределах выборок, применяли коррекцию ковариации [18]. Рассчитали мощность для обнаружения ассоциированных ОНП [19], предпо-

лагая дополнительную модель, идеальное неравновесное сцепление между аллелями риска и маркерными аллелями при уровне значимости $\alpha=5 \times 10^{-8}$. Для генетического отношения рисков в 1,2 мы получили мощность 37%, 89, 98 и 99% для выявления аллелей риска с частотой 0,1, 0,2 и 0,3 до 0,5 соответственно. Для истинного отношения рисков в 1,1 мощность была низкой и варьировалась от 0,2 до 10% в зависимости от частоты встречаемости аллелей.

РЕЗУЛЬТАТЫ

Линейные смешанные модели

После строгого контроля качества были доступны данные генотипа отдельных пациентов для 4561 случая ИИ и 7094 лиц контрольной группы (таблица) на 345 336 непосредственно генотипированных ОНП. При использовании ЛСМ рассчитанная доля дисперсии в модели случай/контроль объясняемая ОНП ($h^2_{ОНП}$) была значимой для всех вариантов инсульта (таблица II в дополнительных данных *on-line*). Более высокие и более значимые значения были получены для ИИ ($h^2_{ОНП}=0,18$; $p=1 \times 10^{-14}$), АКА ($h^2_{ОНП}=0,19$; $p=2 \times 10^{-5}$) и кардиоэмболии ($h^2_{ОНП}=0,24$; $p=2 \times 10^{-8}$) в то время, как показатель для БМС был ниже ($h^2_{ОНП}=0,10$; $p=0,04$).

Для оценки генетической корреляции (r_g) между вариантами инсульта данные лиц контрольной группы случайным образом распределили в одну из двух групп пациентов с различными вариантами инсульта в каждом анализе. Процесс распределения и анализ повторяли 10 раз и получили среднее значение и стандартное отклонение показателей оценки и средние стандартные ошибки (СО) (таблица III в дополнительных данных *on-line*). Значение r_g было самым высоким и значительно отличалось от нуля между АКА и БМС и составило 0,96 (СО=0,059, $p=9 \times 10^{-4}$), хотя большая стандартная ошибка (0,47) указывает на низкую точность. Низкая, но номинально значимая корреляция наблюдалась между АКА и кардиоэмболией ($r_g=0,39$, СО=0,21, $p=0,024$) и между кардиоэмболией, и БМС ($r_g=0,64$, СО=0,40, $p=0,017$).

Оценка полигенного профиля

Хотя ЛСМ анализы были ограничены выборками с данными генотипов отдельных пациентов, для оценки профилей можно использовать выборки консорциума MetaStroke с обобщенной статистикой мета-анализов обнаружения. При проведении анализов в рамках вариантов ИИ, оценки профилей для ИИ, АКА и кардиоэмболии продемонстрировали наличие сильной ассоциации с тем же признаком в независимых целевых

Таблица. Выборки пациентов с ишемическим инсультом, использованные для проверки оценки полигенных профилей

Исследование	ИИ	АКА	Кардиоэмболия	БМС	Контрольная группа
ASGC	1071	375	226	287	1212
WTCCC2-Munich	1140	338	322	104	775
WTCCC2-UK	2350	494	450	471	5107
Всего	4561	1207	998	862	7094

Примечание. ASGC – Australian Stroke Genetics Collaborative; ИИ – ишемический инсульт, АКА – атеросклероз крупных артерий, БМС – болезнь мелких сосудов, WTCCC2 – Wellcome Trust Case Control Consortium 2 Study.

выборках (таблицы IV–VI и рис. I в дополнительных данных *on-line*). Для ИИ сильную связь наблюдали почти во всем распределении значения p . При этом максимальная ассоциация наблюдалась при $p_T < 1$ ($p_{\text{score}} = 1,1 \times 10^{-8}$), типичной для относительно небольшой выборки обнаружения [20]. В целевых выборках существовала небольшая гетерогенность. Для АКА максимальная ассоциация ($p_{\text{score}} = 1,7 \times 10^{-8}$) наблюдалась при $p_T < 0,05$ с отсутствием гетерогенности. Для кардиоэмболии максимальная ассоциация ($p_{\text{score}} = 2 \times 10^{-4}$) наблюдалась в отношении предикторов, включая ОНП, достигших $p_T < 0,001$ и $p_T < 0,01$ без гетерогенности. Для всех трех вариантов ИИ оценка профилей объясняла небольшую часть наблюдаемой дисперсии в модели случай/контроль и была самой высокой для АКА (псевдо- $R^2 = 0,45\%$) и самой низкой для БМС (псевдо- $R^2 = 0,05\%$). Следует также отметить, что оценки с учетом БМС не были ассоциированы с БМС в целевых выборках и на многие из них оказывала влияния гетерогенность между исследованиями (таблица VII в дополнительных данных *on-line*).

При проведении анализов по вариантам инсульта обнаружили значимую общность полигенных профилей между АКА и БМС (таблицы VIII и IX и рис. II в дополнительных данных *on-line*). Большинство оценок профилей при БМС были ассоциированы с АКА без гетерогенности. Наибольшая ассоциация наблюдалась при оценке с включением ≈ 36 000 ОНП, достигших p выявления $< 0,1$ ($p_{\text{score}} = 2 \times 10^{-4}$), которые объясняли 0,19% наблюдаемой дисперсии в модели случай/контроль при АКА. При проведении обратного анализа 3 оценки профилей при АКА были ассоциированы с БМС при $p < 0,05$, например оценка с включением ≈ 20 000 ОНП, достигших $p_T < 0,05$ ($p_{\text{score}} = 0,032$, $R^2 = 0,08\%$). По результатам анализов других двух пар вариантов ИИ (АКА и кардиоэмболия, кардиоэмболия и БМС) не выявили совместной ассоциации оценок профилей (таблицы X–XIII и рис. III и IV в дополнительных данных *on-line*).

При использовании результатов оценки профилей при АКА, рассчитанная доля дисперсии АКА, объясняемая оценками, наиболее тесно связанными с АКА и БМС ($p_T = 0,05$), составила 12,8% (таблица V в дополнительных данных *on-line*). Оценки профилей при БМС не были ассоциированы с БМС в целевых выборках, но оценка, в наибольшей степени ассоциированная с АКА ($p_T = 0,1$), объясняла 0,8% дисперсии ответственности при БМС (таблица VII в дополнительных данных *on-line*). При использовании этих оценок и результатов наблюдаемой ассоциации между вариантами ИИ, рассчитанная генетическая корреляция [12] между АКА и БМС составила $r_g = 0,72$ и достоверно отличалась от нуля (95% доверительный интервал [ДИ] от 0,52 до 0,92). Корреляция с учетом ОНП несущественно отличалась от нуля для АКА и кардиоэмболии ($r_g = 0,13$, 95% ДИ от 0 до 0,56), или кардиоэмболии и БМС ($r_g = 0,64$, 95% ДИ от 0 до 0,92).

Количественный анализ систематической ошибки: ошибочная классификация пациентов с различными вариантами ИИ

Анализ систематической ошибки провели с целью оценки степени, в которой генетическая корреляция

(r_g) между АКА и БМС может быть следствием ошибочной классификации пациентов по вариантам инсульта [21] (см. раздел «Методы» и таблицу XIV в дополнительных данных *on-line*). Учитывая, что ошибочная классификация пациентов по вариантам ИИ соответствует приведенным показателям межэкс-пертной надежности [22], r_g по-прежнему значительно отличалась от нуля. При предположении, что все неправильно классифицированные случаи АКА были, действительно, БМС, и, наоборот, показатель составил $r_g = 0,63$ (95% ДИ от 0,34 до 0,74). При предположении, что все неправильно классифицированные случаи были ни АКА, ни БМС, показатель составил $r_g = 0,75$ (95% ДИ от 0,43 до 0,98). Это свидетельствует о надежности наблюдаемой генетической корреляции с вероятными уровнями ошибочной классификации пациентов в группы по вариантам ИИ.

Мета-анализ ИПГА АКА и БМС

Учитывая доказательства наличия общих распространенных генетических вариантов между пациентами с АКА и БМС, провели обобщенный мета-анализ АКА и БМС (рис. V и VI в дополнительных данных *on-line*). Несмотря на то что ОНП не достигли полногеномной значимости ($p < 5 \times 10^{-8}$), значительная ассоциация ($p = 1 \times 10^{-7}$ в rs17084671; $p = 2 \times 10^{-7}$ в rs6938958 и rs7763080) наблюдалась для кластера ОНП в хромосоме 6q25.2 (таблица XV в дополнительных данных *on-line*), на ≈ 100 Кб выше гена опиоидных рецепторов $\mu 1$ (OPRM1).

■ ОБСУЖДЕНИЕ

ИПГА ИИ выявили важность диагностической классификации ИИ на варианты. Тем не менее использование данных пациентов с отдельными вариантами ИИ приводит к уменьшению размера выборки и предполагает отсутствие аллелей риска, влияющих на риск развития сразу нескольких вариантов ИИ. С использованием некоторых из крупнейших существующих коллекций ИПГА и двух различных аналитических подходов в настоящем исследовании продемонстрировали наличие выраженной общности генов у пациентов с ИИ на фоне АКА и пациентов с поражением мелких сосудов. Согласно нашим оценкам, генетическая корреляция между этими подтипами превышает 0,7, но при использовании более крупных выборок точность этой оценки повысится.

Мы соблюдали осторожность в устранении потенциальных источников систематической ошибки при проведении анализов. Для генетических данных отдельных пациентов главным критерием было их европейское происхождение, а основные варианты происхождения включили в качестве ковариат. Мы также проверили, что положительные эффекты оценки присутствовали в нескольких целевых исследованиях и не были получены в одном единственном исследовании. Ошибочную классификацию случаев инсульта по вариантам, согласно критериям TOAST, считали основным потенциальным источником ошибок, потому что МРТ, которое позволяет повысить точность диагностики, особенно

при БМС, использовали для определения варианта ИИ только в $\approx 50\%$ случаев [5]. Однако по результатам анализа чувствительности предположили, что типичная частота ошибочной классификации будет оказывать минимальное воздействие на предполагаемые корреляции.

Несмотря на то что оценки профилей при БМС не продемонстрировали значимой ассоциации с БМС, они были связаны с АКА по результатам перекрестного анализа вариантов ИИ. При проведении дополнительного анализа, в котором использовали более мощную выборку для обнаружения АКА, оценки профилей при АКА показали наличие значимой ассоциации с АКА и БМС. Это не означает отсутствие согласованности или полигенной архитектуры БМС. Действительно, по результатам нашего и ранее проведенного анализа с использованием ЛСМ [1] обнаружили значимую ОНП-основанную наследуемость для БМС. На результаты оценки профилей оказывали влияние размер выборки и фенотипическая однородность целевых вариантов. Число случаев БМС было самым небольшим, кроме того, они отличались фенотипической гетерогенностью. При проведении анализов профилей исключительно в рамках БМС эти факторы снижают точность полигенных предикторов и статистическую мощность в целевых выборках.

Доля наблюдаемой дисперсии АКА и БМС, объясняемая оценками профилей БМС и АКА (псевдо- R^2), составила 0,2 и 0,08% соответственно. Хотя эти показатели являются низкими, это не означает, что истинная генетическая общность незначительна. Путем объединения ошибок выборок в оценке эффекта всех ОНП в общую оценку, оценки профилей позволяют получить показатели объясняемой дисперсии, которые, как правило, ниже, чем истинные значения [20, 23], с тенденцией к росту по мере увеличения размера выборки. Псевдо- R^2 показатели для бинарных вариантов будет трудно интерпретировать, поскольку они могут зависеть от факта регистрации случаев, т.е. от доли случаев в выборке [24]. Результаты оценки профилей использовали для оценки объясняемой дисперсии в масштабе ответственности и генетических корреляций с использованием теории, учитывающей размер выборки и установление случаев. Например, при АКА, псевдо- R^2 для максимальной оценки профиля составила 0,48%, но рассчитанная дисперсия АКА, объясняемая оценкой с внесением поправок на размер выборки и установлением случаев, составила 12,8%. Таким образом, показатели объясняемой наблюдаемой перекрестной корреляции вариантов инсульта малы, но они могут означать и более высокую генетическую корреляцию. При оценке генетической корреляции в том же наборе данных отметили четкую согласованность между результатами оценки профилей и ЛСМ [25]. В настоящем исследовании применение метода оцен-

ки профилей позволило использовать более широкую выборку с помощью наборов данных, в которых были доступны только обобщенные статистические данные об ассоциации.

ОНП, наиболее сильно ассоциированные с обобщенным вариантом АКА-БМС, были расположены рядом с геном *OPRM1*, аллелями, для которых ранее была предположена ассоциация с ишемической болезнью сердца ($p=5 \times 10^{-6}$) [26], имеющей атеросклеротический генез. Рассчитанная генетическая корреляция между АКА и БМС также согласуется с атеросклеротическим генезом большинства случаев АКА и подмножества случаев БМС. Предполагается, что основным патофизиологическим механизмом развития инсульта при АКА является атеросклероз крупной церебральной артерии [2]. При БМС патологоанатомические и визуализационные исследования свидетельствуют о наличии выраженной гетерогенности заболевания, гипотетически с двумя основополагающими типами сосудистой патологии [27, 28]. К первому типу предположительно относится локализованный атеросклероз крупных перфорантных артерий, как правило, приводящий к развитию крупного изолированного лакунарного инфаркта. Ко второму относятся диффузные, неатеросклеротические артерио-патии мелких перфорантных артерий, ассоциированные с развитием множества мелких инфарктов и часто сопутствующего лейкоареоза [28]. Ранее проведенные анализы факторов риска свидетельствуют о том, что традиционные факторы риска развития атеросклероза более распространены при таком варианте инсульта, как изолированный лакунарный инфаркт. Таким образом, эта подгруппа пациентов может иметь генетическую общность с более широкой выборкой пациентов с ИИ вариантов БМС и АКА.

■ ВЫВОДЫ

Проведенные нами анализы четко свидетельствуют о том, что ИИ вариантов АКА и БМС, которые до сих пор считали генетически различными, могут иметь общий генетический патогенез. Необходимо дальнейшее исследование генетической связи между пациентами с различными вариантами ИИ. Несмотря на то что в последних ИПГА выявили несколько генетических ассоциаций, характерных для отдельных вариантов ИИ, темпы обнаружения ассоциаций были ограничены небольшим числом пациентов с отдельными вариантами ИИ. Если существуют варианты малых эффектов, влияющие на несколько вариантов ИИ, обобщенные анализы вариантов ИИ будут обеспечивать более высокую мощность для их выявления, а также способствовать выявлению биологических механизмов, общих для этих традиционно различных клинических синдромов.

ЛИТЕРАТУРА

1. Bevan S., Traylor M., Adib-Samii P., Malik R., Paul N.L., Jackson C., et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke*. 2012;43:3161–3167. doi: 10.1161/STROKEAHA.112.665760.
2. Adams H.P. Jr, Bendixen B.H., Kappelle L.J., Biller J., Love B.B.,

- Gordon D.L., et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35–41.
3. Ay H., Furie K.L., Singhal A., Smith W.S., Sorensen A.G., Koroshetz W.J. An evidence-based causative classification system for acute ischemic stroke. *Ann Neurol*. 2005;58:688–697. doi: 10.1002/ana.20617.
 4. Gretarsdottir S., Thorleifsson G., Manolescu A., Styrkarsdottir U., Helgadóttir A., Gschwendtner A., et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol*. 2008;64:402–409. doi: 10.1002/ana.21480.
 5. Traylor M., Farrall M., Holliday E.G., Sudlow C., Hopewell J.C., Cheng Y.C., et al; Australian Stroke Genetics Collaborative, Wellcome Trust Case Control Consortium 2 (WTCCC2); International Stroke Genetics Consortium. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2012;11:951–962. doi: 10.1016/S1474-4422(12)70234-X.
 6. Bellenguez C., Bevan S., Gschwendtner A., Spencer C.C., Burgess A.I., Pirinen M., et al. Genome-wide association study identifies a variant in *hdac9* associated with large vessel ischemic stroke. *Nat Genet*. 2012;44:328–333.
 7. Holliday E.G., Maguire J.M., Evans T.J., Koblar S.A., Jannes J., Sturm J.W., et al; Australian Stroke Genetics Collaborative; International Stroke Genetics Consortium; Wellcome Trust Case Control Consortium 2. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet*. 2012;44:1147–1151. doi: 10.1038/ng.2397.
 8. Kilarski L.L., Achterberg S., Devan W.J., Traylor M., Malik R., Lindgren A., et al; GARNET Collaborative Research Group, Wellcome Trust Case Control Consortium 2, Australian Stroke Genetic Collaborative, the METASTROKE Consortium, and the International Stroke Genetics Consortium. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology*. 2014;83:678–685. doi: 10.1212/WNL.0000000000000707.
 9. Ikram M.A., Seshadri S., Bis J.C., Fornage M., DeStefano A.L., Aulchenko Y.S., et al. Genomewide association studies of stroke. *N Engl J Med*. 2009;360:1718–1728. doi: 10.1056/NEJMoa0900094.
 10. Visscher P.M., Brown M.A., McCarthy M.I., Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90:7–24. doi: 10.1016/j.ajhg.2011.11.029.
 11. Lee S.H., Wray N.R., Goddard M.E., Visscher P.M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet*. 2011;88:294–305. doi: 10.1016/j.ajhg.2011.02.002.
 12. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9:e1003348. doi: 10.1371/journal.pgen.1003348.
 13. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795.
 14. Yang J., Lee S.H., Goddard M.E., Visscher P.M. GCTA: a tool or genomewide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82. doi:10.1016/j.ajhg.2010.11.011.
 15. Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A., Reich D. Principal components analysis corrects for stratification in genomewide association studies. *Nat Genet*. 2006;38:904–909. doi: 10.1038/ng1847.
 16. Zhang Y., Chapman A.M., Plested M., Jackson D., Purroy F. The incidence, prevalence, and mortality of stroke in France, Germany, Italy, Spain, the UK, and the US: a literature review. *Stroke Res Treat*. 2012;2012:436125. doi: 10.1155/2012/436125.
 17. Willer C.J., Li Y., Abecasis G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191. doi: 10.1093/bioinformatics/btq340.
 18. Lin D.Y., Sullivan P.F. Meta-analysis of genome-wide association studies with overlapping subjects. *Am J Hum Genet*. 2009;85:862–872. doi:10.1016/j.ajhg.2009.11.001.
 19. Purcell S., Cherny S.S., Sham P.C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003;19:149–150.
 20. Purcell S.M., Wray N.R., Stone J.L., Visscher P.M., O'Donovan M.C., Sullivan P.F., et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.
 21. Wray N.R., Lee S.H., Kendler K.S. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur J Hum Genet*. 2012;20:668–674. doi: 10.1038/ejhg.2011.257.
 22. Meschia J.F., Barrett K.M., Chukwudelunzu F., Brown W.M., Case L.D., Kissela B.M., et al; Siblings with Ischemic Stroke Study (SWISS) Investigators. Interobserver agreement in the trial of org 10172 in acute stroke treatment classification of stroke based on retrospective medical record review. *J Stroke Cerebrovasc Dis*. 2006;15:266–272. doi: 10.1016/j.jstrokecerebrovasdis.2006.07.001.
 23. Wray N.R., Yang J., Hayes B.J., Price A.L., Goddard M.E., Visscher P.M. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet*. 2013;14:507–515. doi: 10.1038/nrg3457.
 24. Lee S.H., Goddard M.E., Wray N.R., Visscher P.M. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol*. 2012;36:214–224. doi: 10.1002/gepi.21614.
 25. Wray N.R., Lee S.H., Mehta D., Vinkhuyzen A.A., Dudbridge F., Middeldorp C.M. Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry*. 2014;55:1068–1087. doi: 10.1111/jcpp.12295.
 26. Lettre G., Palmer C.D., Young T., Ejebe K.G., Allayee H., Benjamin E.J., et al. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARE Project. *PLoS Genet*. 2011;7:e1001300. doi: 10.1371/journal.pgen.1001300.
 27. Boiten J., Lodder J., Kessels F. Two clinically distinct lacunar infarct entities? A hypothesis. *Stroke*. 1993;24:652–656.
 28. Fisher C.M. The arterial lesions underlying lacunes. *Acta Neuropathol*. 1968;12:1–15.