Published in final edited form as:
Stroke. 2015 March ; 46(3): 615–619. doi:10.1161/STROKEAHA.114.007930.

Genetic overlap between diagnostic subtypes of ischemic stroke

Elizabeth G Holliday, PhD[1,2], **Matthew Traylor, PhD**[3], **Rainer Malik, PhD**[4], **Steve Bevan, PhD**[3], **Guido Falcone, MD**[5,6], **Jemma C Hopewell, PhD**[7], **Yu-Ching Cheng, PhD**[8], **Ioana Cotlarciuc, PhD**[9], **Joshua C Bis, PhD**[10], **Eric Boerwinkle, PhD**[11], **Giorgio B Boncoraglio, MD**[12], **Robert Clarke, MD**[7], **John W Cole, MD**[13,14], **Myriam Fornage, PhD**[11], **Karen L Furie, MD**[15], **M Arfan Ikram, MD, PhD**[16], **Jim Jannes, MBBS**[17], **Steven J Kittner, MD**[13,14], **Lisa F Lincz, PhD**[18,19,20], **Jane M Maguire, PhD**[19,21], **James F Meschia, MD**[22], **Thomas H Mosley, PhD**[23], **Mike A Nalls, PhD**[24], **Christopher Oldmeadow, PhD**[2], **Eugenio A Parati, MD**[12], **Bruce M Psaty, MD, PhD**[10,25,26], **Peter M Rothwell, FRCP**[27], **Sudha Seshadri, MD**[28], **Rodney J Scott, PhD**[19,20,29], **Pankaj Sharma, FRCP**[30], **Cathie Sudlow, FRCPE**[31], **Kerri L Wiggins, MS**[10], **Bradford B Worrall, MD MSc**[32], **Jonathan Rosand, MD**[5,6], **Braxton D Mitchell, PhD**[8], **Martin Dichgans, MD**[4,33], **Hugh S Markus, DM**[3], **Christopher Levi, MBBS**[19,34], **John Attia, MD, PhD**[*,1,2], **Naomi R Wray, PhD**[*,35], and **on behalf of the Australian Stroke Genetics Collaborative, the Wellcome Trust Case Control Consortium 2 and the International Stroke Genetics Consortium**

[1] School of Medicine and Public Health, University of Newcastle, Australia [2] Clinical Research Design, IT and Statistical Support Unit, Hunter Medical Research Institute, Newcastle, Australia [3] Department of Clinical Neurosciences, University of Cambridge, UK [4] Institute for Stroke and Dementia Research, Klinikum der Universität Mün-chen, Ludwig-Maximilians-Universität, Munich, Germany [5] Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA USA [6] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA USA [7] Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, UK [8] Department of Medicine, University of Maryland School of Medicine, Baltimore, MD USA [9] Imperial College Cerebrovascular Research Unit (ICCRU), Imperial College, London, UK [10] Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA USA [11] Institute of Molecular Medicine and Human Genetics Center, the University of Texas Health Science Center at Houston, TX USA [12] Department of Cerebrovascular Disease, IRCCS Istituto Neurologico Carlo Besta, Milan, Italy [13] Department of Neurology, University of Maryland School of Medicine, Baltimore, MD USA [14] Department of Neurology, Veterans Affairs Medical Center, Baltimore, MD USA [15] Department of Neurology, Rhode Island Hospital, Providence, RI USA [16] Departments of Epidemiology, Neurology, and Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands [17] School of Medicine, University of Adelaide, Australia [18] Hunter Haematology Research Group, Calvary Mater Newcastle Hospital, Newcastle, Australia [19] Hunter Medical Research Institute, Newcastle, Australia [20] School of

Correspondence: Elizabeth Holliday CReDITSS, HMRI C/- University Drive, Callaghan, NSW, 2308 Australia Ph: +61-2-40420508 Fax: +61-2-40420039 liz.holliday@newcastle.edu.au.
*These authors jointly directed this work

Disclosures

J.R. has a consulting relationship with Boehringer Ingelheim. B.B.W. is an Associate Editor of the American Academy of Neurology (AAN) journal 'Neurology'.

Biomedical Sciences and Pharmacy, University of Newcastle, Australia [21] School of Nursing and Midwifery, University of Newcastle, Australia [22] Department of Neurology, Mayo Clinic, Jacksonville, FL USA [23] Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, MS USA [24] Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD USA [25] Department of Epidemiology, University of Washington, Seattle, WA USA [26] Group Health Research Institute, Group Health, Seattle, WA USA [27] Stroke Prevention Research Unit, Nuffield Department of Clinical Neuroscience, University of Oxford, UK [28] Department of Neurology, Boston University School of Medicine, Boston, MA USA [29] Pathology North, NSW Pathology, John Hunter Hospital, Newcastle, Australia [30] Institute of Cardiovascular Research, Royal Holloway University of London (ICR2UL) & Ashford and St Peters Hospital [31] Division of Clinical Neurosciences, University of Edinburgh, UK [32] Departments of Neurology and Public Health Sciences, University of Virginia, Charlottesville, VA USA [33] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany [34] Department of Neurology, John Hunter Hospital, Newcastle, Australia [35] Queensland Brain Institute, The University of Queensland, Australia

Abstract

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 shared genetic etiology among the three major subtypes: large artery atherosclerosis (LAA), cardioembolism (CE) 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 2,167 LAA, 2,405 CE and 1,854 SVD) and 62,004 controls from the Metastroke consortium. For 4,561 cases and 7,094 controls, individual-level genotype data were also available. Genetic correlations between subtypes were estimated using linear mixed models (LMM) and polygenic profile scores. Meta-analysis of a combined LAA-SVD phenotype (4,021 cases, 51,976 controls) was performed to identify shared risk alleles.

Results—High genetic correlation was identified between LAA and SVD using LMM (r_g=0.96, SE=0.47, P=9×10^{-4}) and profile scores (r_g=0.72; 95% CI: 0.52 – 0.93). Between LAA and CE, and SVD and CE, correlation was moderate using LMM but not significantly different from zero for profile scoring. Joint meta-analysis of LAA and SVD identified strong association (P=1×10^{-7}) for SNPs near the opioid receptor μ1 (*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.

Keywords

ischemic stroke; genetic epidemiology; atherosclerosis; lacunar stroke

Introduction

Ischemic stroke (IS) is a complex disease influenced by numerous clinical, genetic and lifestyle risk factors. While conventional factors such as hypertension, dyslipidaemia, diabetes and smoking are well-established, genetic factors contribute up to 30-40% of risk [1] and are poorly understood. Despite recent advances in high throughput genotyping, gene discovery for IS has progressed slowly due to the technical nature of case ascertainment and etiological 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 (CE) and small vessel (lacunar) disease (SVD). By exploiting these phenotypically more homogeneous categorisations, genome-wide association studies (GWAS) have identified a number of genetic associations specific to individual subtypes [4-7]. In contrast, only two 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 reduces 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 correlation between the three major IS subtypes using individual-level GWAS data and meta-analysis summary statistics from the International Stroke Genetics Consortium (ISGC) [6, 7] and Metastroke [5]. Genetic correlations were estimated using two different methods: linear mixed models [11] and polygenic profile scoring [12].

Methods

Data Sources

The Metastroke study included 15 individual studies contributing 12,389 total ischemic stroke cases and 62,004 controls of European ancestry (Data Supplement, Table I). 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 2,167 cases with large artery atherosclerosis (LAA), 2,405 with cardioembolism (CE) and 1,854 cases with small vessel disease (SVD); the remainder had other, undetermined, or cryptogenic etiology. Each study conducted genotype imputation using either HapMap Phase 2 or 1000 Genomes reference panels, fitted additive logistic regression models for all SNPs and provided regression summary statistics for IS and its subtypes (if available). Individual-level GWAS data was also available for three of the largest Metastroke cohorts: two from the Wellcome Trust Case Control Consortium 2 Study (WTCCC2-UK and WTCCC2-Munich) [6] and the Australian Stroke Genetics Collaborative (ASGC) [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 modelling

Genotype data for the three samples with individual-level data were combined to yield a single dataset using 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 following three 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 linear mixed models [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 three major subtypes each contribute approximately 20% (total 60%), equating to 0.4% prevalence (20% ×2%) for each subtype [5, 16]. The remaining 0.8% prevalence (40% × 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 three subtypes. For each, this was performed using “leave-one-out” validation for the three “target” datasets with individual-level data. In turn, each of the three was set aside and a GWAS “discovery” meta-analysis was performed using all other Metastroke datasets. SNPs with data from at least five Metastroke studies were retained and pruned for linkage disequilibrium (LD: $r^2>0.2$ within 1Mb) using PLINK's --clump algorithm [13], which preferentially retains the most associated SNP in an LD region. From the pruned set we extracted subsets passing ten graded significance thresholds ($P_T = 1\times10^{-4}$, 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” dataset. 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 three 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 three target datasets were combined via random-effects meta-analysis to estimate overall significance. Overall variance explained was estimated as the sample-size weighted mean of target dataset-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 α=0.05, we had 98% power to detect polygenic scores explaining ≥0.2% of variance in case/control status for any target subtype, 81-83% power to detect scores explaining ≥0.1% of variance (varying by subtype) and 52-54% power for scores explaining ≥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 Metal [17], for 2,167 LAA cases, 1,854 SVD cases (4,021 total cases) and 51,976 controls from 12 studies (Data Supplement, Table I). To control type 1 error due to overlapping controls for LAA and SVD within cohorts, a covariance correction was applied [18]. Power to detect associated SNPs was calculated [19] assuming an additive model, perfect LD between risk and marker alleles and a significance level of $\alpha=5\times10^{-8}$. For a genetic risk ratio of 1.2, we had 37%, 89% and 98-99% power to identify risk alleles with frequency 0.1, 0.2 and 0.3-0.5 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 was available for 4,561 IS cases and 7,094 controls (Table 1) at 345,336 directly genotyped SNPs. Using linear mixed models (LMM), the estimated proportion of variance in case-control status explained by the SNPs (h^2_{SNP}) was significant for all stroke traits (Data Supplement, Table II). Higher and more significant values were estimated for IS (h^2_{SNP} =0.18; $P=1\times10^{-14}$), LAA (h^2_{SNP}=0.19; $P=2\times10^{-5}$) and CE (h^2_{SNP}=0.24; $P=2\times10^{-6}$), while the estimate for SVD was lower (h^2_{SNP}=0.10; P=0.04).

To estimate the genetic correlation (r_g) between subtype cases, controls were randomly allocated to one of the two subtype groups in each analysis. This allocation and estimation process was repeated ten times and the mean and standard deviation of parameter estimates, and mean standard errors (SE) derived (Data Supplement, Table III). The r_g value was highest and significantly different from zero between LAA and SVD at 0.96 (SD=0.059; $P=9\times10^{-4}$), although the large standard error (0.47) indicates low precision. Reduced, but nominally significant correlation was observed between LAA and CE (r_g=0.39, SE=0.21, P=0.024) and between CE and SVD (r_g=0.64, SE=0.40, P=0.017).

Polygenic profile scoring

While LMM analyses were restricted to samples with individual-level data, profile scoring could utilise Metastroke samples with summary statistics for "discovery" meta-analyses. In analyses within traits, profile scores for IS, LAA and CE showed strong association with the same trait in independent target cohorts (Data Supplement, Tables IV-VI, Figure I). For IS, strong association was observed across most of the discovery P-value distribution with

maximum association observed for $P_T<1$ ($P_{score}=1.1\times10^{-8}$), typical of relatively small discovery samples [20]. There was little effect size heterogeneity across target cohorts. For LAA, maximum association ($P_{score}=1.7\times10^{-8}$) was observed for $P_T<0.05$ with no heterogeneity. For CE, maximum association ($P_{score}=2\times10^{-4}$) was observed for predictors including SNPs reaching $P_T<0.001$ and $P_T<0.01$, with no heterogeneity. For all three 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 (Data Supplement, Table VII).

Analyses between stroke subtypes detected significant polygenic sharing between LAA and SVD (Data Supplement Tables VIII, IX, Figure II). The majority of SVD-based scores associated with LAA, with no heterogeneity. The highest association was observed for a score including ~36,000 SNPs reaching discovery $P_T<0.1$ ($P_{score}=2\times10^{-4}$) which explained an estimated 0.19% of observed LAA case/control variance. In the reverse analysis, three LAA-based profile scores associated with SVD at P<0.05, e.g., the score including ~20,000 SNPs reaching $P_T<0.05$ (P_{score}=0.032, R^2 =0.08%). In analyses of the other two subtype pairs (LAA and CE, CE and SVD), no co-association of profile scores was observed (Data Supplement Tables X-XIII, Figures III-IV).

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% (Data Supplement, Table V). SVD-based scores did not associate with SVD in target samples, but the score most significantly associated with LAA (P_T=0.1) explained 0.8% of SVD liability variance (Data Supplement, Table VII). Using these estimates and the observed cross-trait association results, the estimated genetic correlation [12] between LAA and SVD was r_g=0.72, which was significantly different from zero (95% CI: 0.52, 0.92). The SNP-based correlation was not significantly different from zero for LAA and CE (r_g=0.13, 95% CI: 0, 0.56), or CE 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 misclassification [21] (see Data Supplement, Methods and Table XIV). 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 (Data Supplement, Figures V, VI). While no SNPs reached genome-wide significance ($P<5\times10^{-8}$), suggestive association ($P=1\times10^{-7}$, at rs17084671; $P=2\times10^{-7}$ at rs6938958 and rs7763080) was observed for a cluster of SNPs at chromosome

6q25.2 (Data Supplement, Table XV), ~100kb upstream of the opioid receptor μ1 (*OPRM1*) gene.

Discussion

Genome-wide association studies of ischemic stroke (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 two 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 since MRI, which increases diagnostic accuracy particularly for SVD, was only used for subtyping approximately 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 analysis [1] 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 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, i.e., 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, while 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 dataset 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 datasets 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 (CHD: $P=5\times10^{-6}$), [26] which has an atherosclerotic etiology. Estimated genetic correlation between LAA and SVD is also consistent with an atherosclerotic etiology 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 two major underlying vascular pathologies being hypothesised [27, 28]. The first involves localised 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 co-existent radiological leukoaraiosis [28]. Earlier risk factor analyses suggested 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.

Summary

Our analyses strongly suggest that LAA and SVD – which have been hitherto considered genetically distinct – may have a shared genetic etiology. 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 exist 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding

EGH was supported by the Australian Heart Foundation and National Stroke Foundation (100071). NRW was supported by the Australian National Health and Medical Research Council (NHMRC: 613602). H.M. is supported by a National Institute for Health Research (NIHR) Senior Investigator award. H.M. and S.B. are supported by the Cambridge University Hospital National Institute for Health Research Biomedical Research Centre (NIHR BRC). J.R. and B.B.W. received Research Funding from the NIH. Acknowledgements of funding for the individual Metastroke studies are detailed in the original publication by Traylor et al [5].

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 genomewide associations. Stroke. 2012; 43:3161–3167. [PubMed: 23042660]
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. [PubMed: 7678184]
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. [PubMed: 16240340]
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. [PubMed: 18991354]
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. [PubMed: 23041239]
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. [PubMed: 22306652]
7. Holliday EG, Maguire JM, Evans TJ, Koblar SA, Jannes J, Sturm JW, et al. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. Nat Genet. 2012; 44:1147–1151. [PubMed: 22941190]
8. Kilarski LL, Achterberg S, Devan WJ, Traylor M, Malik R, Lindgren A, et al. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. Neurology. 2014; 83:678–685. [PubMed: 25031287]
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. [PubMed: 19369658]
10. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of gwas discovery. Am J Hum Genet. 2012; 90:7–24. [PubMed: 22243964]
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. [PubMed: 21376301]
12. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013; 9:e1003348. [PubMed: 23555274]
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. [PubMed: 17701901]
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. [PubMed: 21167468]
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. [PubMed: 16862161]
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. [PubMed: 22550614]
17. Willer CJ, Li Y, Abecasis GR. Metal: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]
18. Lin DY, Sullivan PF. Meta-analysis of genome-wide association studies with overlapping subjects. Am J Hum Genet. 2009; 85:862–872. [PubMed: 20004761]
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. [PubMed: 12499305]

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. [PubMed: 19571811]
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. [PubMed: 22258521]
22. 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. [PubMed: 17904086]
23. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from snps. Nature reviews. Genetics. 2013; 14:507–515. [PubMed: 23774735]
24. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. Genet Epidemiol. 2012; 36:214–224. [PubMed: 22714935]
25. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. Journal of child psychology and psychiatry, and allied disciplines. 2014; 55:1068–1087.
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. [PubMed: 21347282]
27. Boiten J, Lodder J, Kessels F. Two clinically distinct lacunar infarct entities? A hypothesis. Stroke. 1993; 24:652–656. [PubMed: 8488519]
28. Fisher CM. The arterial lesions underlying lacunes. Acta neuropathologica. 1968; 12:1–15. [PubMed: 5708546]

Table 1

Ischaemic stroke samples used for testing polygenic scores

Study	IS	LAA	CE	SVD	Controls
ASGC	1,071	375	226	287	1,212
WTCCC2-Munich	1,140	338	322	104	775
WTCCC2-UK	2,350	494	450	471	5,107
Total	**4,561**	**1,207**	**998**	**862**	**7,094**