Shared Genetic Contributions to Atrial Fibrillation and Ischemic Stroke Risk

Sara L. Pulit, PhD[1,2,3]; Lu-Chen Weng, PhD[3,4]; Patrick F McArdle, PhD[5]; Ludovic Trinquart, PhD[6,7]; Seung Hoan Choi, PhD[3]; Braxton D. Mitchell, PhD, MPH[5,8]; Jonathan Rosand, MD, MSc[3,9,10]; Paul I W de Bakker, PhD[1,11]; Emelia J Benjamin, MD, ScM[6,12,13]; Patrick T Ellinor, MD, PhD[3,4,14], Steven J Kittner, MD, MPH[15,16]; Steven A Lubitz, MD, MPH*[3,4,14]; Christopher D Anderson, MD, MMSc*[3,9,10], on behalf of the Atrial Fibrillation Genetics Consortium and the International Stroke Genetics Consortium.

1. Department of Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
2. Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of Oxford, Oxford, UK
3. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA
4. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
5. Department of Medicine, Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland, United States of America
6. National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, USA
7. Department of Biostatistics, Boston University School of Public Health, Boston, MA
8. Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA
9. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
10. Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
11. Computational Genomics, Vertex Pharmaceuticals, Boston, MA, USA (current address)
12. Cardiovascular Medicine and Preventive Medicine Sections, Evans Department of Medicine, Boston University School of Medicine, Boston, MA, USA
13. Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA
14. Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, MA, USA
15. Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA
16. Department of Neurology, Veterans Affairs Medical Center, Baltimore, MD, USA

* These authors contributed equally to supervision of this study

Corresponding authors

Dr. Steven A. Lubitz
Cardiovascular Research Center and Cardiac Arrhythmia Service
Massachusetts General Hospital
55 Fruit Street
GRB 109
Boston, MA, 02114, USA
slubitz@mgh.harvard.edu

Dr. Christopher D. Anderson
Center for Genomic Medicine
Massachusetts General Hospital
185 Cambridge Street, CPZN-6818
Boston, MA, 02114, USA
cdanderson@mgh.harvard.edu

Abstract

Background

Atrial fibrillation is a heritable disorder and leading cause of stroke. The extent to which cardioembolic strokes are driven by inherited atrial fibrillation risk could inform whether atrial fibrillation genetic risk is a clinically-relevant risk factor for stroke or could be used to profile stroke patients to determine stroke mechanisms.

Methods

From the 13,390 ischemic stroke cases genotyped by the Stroke Genetics Network (SiGN), we identified 3,190 atrial fibrillation cases and 3,000 cardioembolic stroke cases. We performed genome-wide association testing for each condition using 28,026 controls. We evaluated genetic overlap of atrial fibrillation and cardioembolic stroke by calculating correlations between the single nucleotide polymorphism (SNP) effects in both genome-wide scans. We also estimated the proportion of variance explained by additive genetic variation (SNP-based heritability) for atrial fibrillation and cardioembolic stroke. Lastly, we constructed an atrial fibrillation polygenic risk score from a previously-derived set of SNPs to estimate how strongly such a score would associate to cardioembolic stroke and other stroke subtypes.

Results

SNP-based heritability was nearly identical for atrial fibrillation and cardioembolic stroke (20.0% and 19.5%, respectively). The genetic basis of atrial fibrillation and cardioembolic stroke was highly correlated (Pearson's $r = 0.77$ across SNPs with $p < 4.4 \times 10^{-4}$ in a previous atrial fibrillation meta-analysis). An atrial fibrillation polygenic risk score was strongly associated with cardioembolic stroke ($p = 1.01 \times 10^{-65}$, odds ratio (OR) per 1 standard deviation (sd) = 1.39 [95% confidence interval: 1.34 - 1.44]) after adjustment for clinical atrial fibrillation risk factors, and explained ~20% of the variance in cardioembolic stroke. The atrial fibrillation polygenic risk score did not associate with the other primary stroke subtypes (p >

0.1 for each), but was modestly associated with stroke of undetermined cause (p = 0.004, OR per 1 sd = 1.07 [95% CI: 1.02 - 1.13]).

Conclusions

Genetic risk for atrial fibrillation increases cardioembolic stroke risk. Given the high correlation between atrial fibrillation and cardioembolic stroke genetic risk, clinical studies are warranted to assess whether genetic risk information can be used to identify strokes caused by atrial fibrillation.

Introduction

Atrial fibrillation affects nearly 34 million individuals worldwide[1] and is associated with a five-fold increased risk of ischemic stroke,[2] a leading cause of death and long-term disability.[3,4] Strokes caused by atrial fibrillation are classified as cardioembolic in nature, as atrial fibrillation promotes the formation of blood clots in the heart which can embolize distally to the brain. Cardioembolic stroke is often, though not always, caused by atrial fibrillation. However, whether cardioembolic stroke risk is substantially influenced by genetic susceptibility to atrial fibrillation remains an open question.

Recent genome-wide association studies (GWAS) of atrial fibrillation[5] and ischemic stroke indicate that both traits are common complex disorders with polygenic architectures.[6,7] The top loci for cardioembolic stroke, on chromosome 4q25 upstream of *PITX2* and on 16q22 near *ZFHX3*, are both leading risk loci for atrial fibrillation.[8–10] Although the leading genetic loci for atrial fibrillation and cardioembolic stroke overlap, the genetic susceptibility to each is likely to involve the aggregate contributions of hundreds or thousands of genetic loci, consistent with other polygenic traits.[11] Understanding the degree to which the genetic architectures of these diseases overlap may elucidate the extent to which genetic risk factors for atrial fibrillation contribute to cardioembolic stroke risk, and thus whether genetic risk for atrial fibrillation may represent a useful candidate biomarker for the cardioembolic stroke subtype.

We therefore analyzed 13,390 ischemic stroke cases and 28,026 referents from the NINDS-Stroke Genetics Network (SiGN)[12] with genome-wide genotyping data to address two specific questions. First, we assessed the degree of overlap between the genetic architectures of atrial fibrillation in the general population and atrial fibrillation in patients with stroke, leveraging additional GWAS data from the Atrial Fibrillation Genetics Consortium (AFGen). Second, we compared the genetic architectures of atrial fibrillation and cardioembolic stroke to ascertain the extent to

which heritable risk of cardioembolic stroke is explained by genetic risk factors for atrial fibrillation.

Methods

Data access

Code, supporting data, and downloadable supplemental tables from our analyses are available here: https://github.com/UMCUGenetics/Afib-Stroke-Overlap. Please see the **Supplementary Information** for additional information regarding access to genotypes, dbGaP data identifiers, additional methods, and links to downloadable summary-level data from our analyses.

The Stroke Genetics Network (SiGN) and genome-wide association study of ischemic stroke subtypes

The Stroke Genetics Network (SiGN) was established with the aim of performing the largest genome-wide association study (GWAS) of ischemic stroke to date. The complete design of the study, including sample collection, phenotyping, and data processing, have been described previously.[12] Briefly, ischemic stroke cases were collected from cohorts located in the United States, Europe and Australia. Stroke cases included in SiGN were classified into stroke subtypes using the Causative Classification System (CCS), which subtypes cases through an automated, web-based system that accounts for clinical data, test results, and imaging information.[13,14] Within CCS, there are two sub-categories: CCS causative, which does not allow for competing subtypes in a single sample; and CCS phenotypic, which does.

SiGN samples represent three continental populations (European-ancestry; African-ancestry; and non-European ancestry and non-African ancestry samples, primarily of admixed ancestry from Latin American populations, labelled 'Hispanic'). In total, the study contains 13 case-referent analysis groups: 10 of European ancestry, two of African ancestry, and one Hispanic.[12] Due to data sharing restrictions, only the European- and African-ancestry samples are analyzed here.

After quality control (QC), the SiGN dataset comprised 16,851 ischemic stroke cases and 32,473 stroke-free controls. Further details on the SiGN study and a complete breakdown of case and referent information, including counts by stroke subtype, are included in the **Supplementary Information** and **Supplementary Table 1**.

All samples in SiGN were phenotyped using the Causative Classification of Stroke (CCS) system. Additionally, ~74% of samples were subtyped using the TOAST classification system.[15,16] The complete details of phenotyping in SiGN can be found in the **Supplementary Information** and in the original publication**.**[12]

Identifying atrial fibrillation cases and controls in SiGN

To perform association testing and downstream analyses, we sought to identify a set of atrial fibrillation cases and suitably-matched referents available in the SiGN dataset. We defined atrial fibrillation on the basis of five variables available in the CCS phenotyping system: (i) atrial fibrillation, (ii) paroxysmal atrial fibrillation, (iii) atrial flutter, (iv) sick sinus syndrome, and (v) atrial thrombus. This definition yielded 3,190 atrial fibrillation cases for analysis. For quality control purposes, we also ran an initial genome-wide association study in atrial fibrillation only (N = 1,751 cases) to ensure overlap with the broader atrial fibrillation definition (**Supplementary Information**).

Given that the vast majority of stroke-free referents in SiGN were collected from publicly-available sources, available atrial fibrillation information for these samples is limited. From the 28,026 controls available for analysis, we established a set of 3,861 individuals for whom atrial fibrillation information was available and who did not have atrial fibrillation. For the remaining subjects, we assumed that individuals did not have atrial fibrillation.

Genome-wide association testing of ischemic stroke subtypes and atrial fibrillation in SiGN

We merged all genotype dosages available from the European-ancestry and African-ancestry analysis groups in SiGN. SNPs with imputation quality (info score) < 0.3 and minor allele frequency (MAF) < 1% were removed from the data. We additionally cleaned SNPs based on frequencies in genotyped versus imputed data (**Supplementary Information**).

We then implemented a linear mixed model to perform association testing using BOLT-LMM.[17] Linear mixed models can account for structure in the data, such as that due to (familial or cryptic) relatedness and population structure, while improving power for discovery.[18–20] Due to extensive structure in the SiGN data,[12] induced by both study design and population ancestry, we adjusted the BOLT-LMM model for the top ten principal components (PCs) and sex, in addition to the genetic relationship matrix used as a random effect in the linear mixed model.[17] We calculated PCs in EIGENSTRAT[21] using a similar set of SNPs to that used in the genetic relationship matrix (GRM, **Supplementary Information**) but using a missingness threshold of 0.1%. To construct the GRM, we first identified the set of SNPs with imputation quality > 0.8 and MAF > 1%. More than 5.5M SNPs passed these QC criteria, so we randomly selected 20% of the data (~1.1M SNPs) for computational efficiency in calculating the GRM. We also identified SNPs outside the MHC and *LCT* regions, outside the inversions on chromosomes 8 and 17, and LD pruned (r^2 = 0.2). These filtering steps resulted in ~250,000 SNPs available for the GRM. We used Plink 1.9[22,23] to convert imputed dosages to best-guess genotypes and then compute the GRM.

We used this approach to perform GWAS in atrial fibrillation and each of the stroke subtypes available in the SiGN dataset.

Heritability calculations in ischemic stroke subtypes

To assess the aggregate genetic contributions of common variants to disease susceptibility, we calculated additive SNP-based heritability (h_g^2) estimates for

ischemic stroke, within ischemic stroke subtypes, and for atrial fibrillation using restricted maximum likelihood as implemented in BOLT-REML.[17] We used the GRM from our GWAS analyses (described in the section above) to estimate heritability. We adjusted all heritability analyses for 10 PCs and sex. To test the robustness of our heritability estimates, we calculated three additional GRMs to re-estimate heritability, and additionally estimated heritability using a second software (GCTA[24]). Additional details are provided in the **Supplementary Information**.

Genetic correlation between atrial fibrillation and ischemic stroke subtypes

We used summary-level data from the latest Atrial Fibrillation Genetics (AFGen) Consortium meta-analysis of atrial fibrillation[5] to calculate a z-score for each SNP in that GWAS. Additionally, we calculated a z-score for each SNP from our GWAS of each stroke subtype in SiGN as well as in the GWAS of atrial fibrillation we performed in the SiGN data. Finally, as a null comparator, we downloaded SNP z-scores from a GWAS of educational attainment[25] available through LDHub (http://ldsc.broadinstitute.org/, accessed 11-1-2017). We applied additional quality control based on SNP frequency differences (**Supplementary Information**) and calculated correlations (Pearson's r) between z-scores from two traits to evaluate correlation. We also evaluated correlation between the phenotypic data and provide this in the **Supplementary Information**.

Constructing an atrial fibrillation polygenic risk score

To construct an atrial fibrillation polygenic risk score (PRS), we used SNPs from a previously-derived atrial fibrillation PRS.[26] Briefly, the PRS was derived using results from a recent GWAS of atrial fibrillation, comprised of 17,931 cases and 115,142 referents[5] and testing various sets of SNPs based on their p-value from that GWAS (varying from $p < 5 \times 10^{-8}$ to $p < 0.001$) and using varied linkage disequilibrium thresholds (0.1 - 0.9).[26] These sets of SNPs were used to generate various PRSs, which were then independently tested for association to atrial fibrillation in an independent sample from the UK Biobank; the best-performing PRS (defined as the

PRS with the lowest Akaike's Information Criterion) comprised 1,168 SNPs with $p < 1 \times 10^{-4}$ in the atrial fibrillation GWAS and LD pruned at an r^2 threshold of 0.5.[26]

Of these 1,168 SNPs, we identified 934 SNPs in the SiGN dataset with imputation info > 0.8 and MAF > 1%. We used these 934 SNPs to construct the atrial fibrillation PRS in the SiGN dataset by weighting the imputed number of risk-increasing alleles carried by an individual at a given SNP (i.e., 0-2 risk-increasing alleles) and then weighting the dosage by the effect of the allele, as determined by the most recent GWAS.[5] We computed the final PRS for each individual by summing across all of the weighted genotypes.

Testing an atrial fibrillation polygenic risk score in ischemic stroke subtypes

To test the association between the atrial fibrillation PRS and ischemic stroke subtypes, we performed a logistic regression implemented in R. Phenotypes were coded as 0 (unaffected) or 1 (affected), and we included sex and the top 10 PCs as additional covariates. We also optionally adjusted the association tests for the following clinical covariates: age, diabetes mellitus status, cardiovascular disease status, smoking status (current smoker, former smoker, or never smoked), and hypertension status. To ensure that our analyses of the PRS were robust to ancestral heterogeneity, we additionally tested the PRS in the subset of European-ancestry samples only (**Supplementary Information**). We calculated the odds ratio of the PRS for an increase of one standard deviation in the score by first converting the PRS per individual to a z-score, where:

$$\mathrm{PRS}_{\text{z-score}} = \frac{PRS - mean(PRS)}{standard\ deviation\ (PRS)}$$

We then recalculated the association between $\mathrm{PRS}_{\text{z-score}}$ and the phenotype, and converted the resulting regression coefficients (i.e., betas) of the PRS to odds ratios.

To determine the variance explained by the atrial fibrillation PRS in cardioembolic stroke, we constructed a model in BOLT-LMM that consisted of two variance components: (1) a variance component made up of SNPs for the genetic relationship matrix, and (2) a variance component made up of SNPs from the PRS. After computing the estimated variance explained for each component in BOLT-LMM, we converted the estimate to the liability score (**Supplementary Information**).

Results

We began by testing our ability to rediscover known atrial fibrillation genetic associations in the SiGN dataset, assembled to study the genetics of ischemic stroke. Of the 13,390 European- and African-descent stroke cases in SiGN, we identified 3,190 cases with atrial fibrillation or paroxysmal atrial fibrillation, as well as other diagnoses suggestive of underlying atrial fibrillation, including atrial flutter,[27] left atrial thrombus, and sick sinus syndrome[28] (**Table 1** and **Supplementary Table 1**). SiGN also includes 28,026 European- and African-descent controls, the majority of which are publicly-available referent sets with minimal medical history information.

We first performed a genome-wide association study (GWAS) of the 3,190 atrial fibrillation cases and 28,026 referents (**Supplementary Figure 1**). The results were highly concordant with the latest GWAS in atrial fibrillation performed by the Atrial Fibrillation Genetics (AFGen) Consortium. Specifically, the GWAS in SiGN identified SNPs at the *PITX2* and *ZFHX3* loci[8,10] (index SNPs rs6843082 and rs2106261, respectively) that were associated with atrial fibrillation beyond a genome-wide significance threshold ($p < 5 \times 10^{-8}$) and with similar effect sizes compared to those reported in the AFGen GWAS[5] (**Supplementary Table 2**). We additionally looked up 26 variants most recently reported by AFGen as being associated to atrial fibrillation[5]; of the 25 available in the SiGN data, 24 showed a consistent direction of effect (binomial $p = 7.7 \times 10^{-7}$), and 14 showed nominal association signal ($p < 0.05$; **Supplementary Table 2**). The only SNP without a consistent direction of effect in both analyses had originally been identified through an eQTL analysis.[5]

Next, we evaluated heritability within cardioembolic stroke and atrial fibrillation, with the specific goal of evaluating the proportion of genetic variation in cardioembolic stroke that can be explained by genetic susceptibility to atrial fibrillation. As a basis for comparison, we also calculated heritability in the two additional primary stroke subtypes[15] (large artery atherosclerosis and small artery

occlusion), as well as in 'undetermined' stroke, a highly heterogeneous subset of cases[29,30] that cannot be classified by standard subtyping systems.[13,15] This latter category could potentially contain misclassified cardioembolic strokes caused by occult atrial fibrillation.[31] We analyzed common variants (minor allele frequency > 1%) in SiGN to evaluate heritability of atrial fibrillation, followed by the major stroke subtypes (cardioembolic, large artery atherosclerosis, and small artery occlusion), and finally in strokes of undetermined cause (**Methods**). We performed restricted maximum likelihood analysis implemented in BOLT-LMM[17] (i.e., BOLT-REML) to calculate genetic heritability using a high-quality set of linkage disequilibrium (LD)-pruned imputed variants (imputation info score > 0.8, minor allele frequency >1%, LD r^2 threshold = 0.2) converted to best-guess genotypes (**Methods**). We additionally corrected the REML analyses for sex and the top ten principal components and converted the observed-scale heritability estimates produced by BOLT-REML to the liability scale (**Supplementary Information**). Using the 3,190 atrial fibrillation cases and 28,026 referents available in SiGN, we estimated heritability of atrial fibrillation at 20.0%, consistent with previous estimates in larger sample sizes (**Supplementary Figure 2**).[32,33]

We next evaluated SNP-based heritability in the stroke subtypes. We focus these and all subsequent analyses on phenotypes resulting from the CCS causative (CCSc) subtyping algorithm, one of the three stroke subtyping systems available in SiGN[34]; heritability results from all three subtyping systems are provided in **Supplementary Table 3**. Our analyses estimate cardioembolic stroke heritability at 19.5% (**Supplementary Figure 2**), a heritability highly consistent with that of atrial fibrillation. We found large artery stroke and small artery occlusion to have similarly modest heritability (15.5% and 23.0%, respectively; **Supplementary Figure 2** and **Supplementary Table 3**), consistent with the hypothesis that stroke risk is modulated partially by biological factors, but with a large contribution from non-genetic variables as well.

Despite their presumed etiologic and phenotypic heterogeneity, the undetermined categories of stroke showed heritability measures comparable to those of the major

stroke subtypes (**Supplementary Figures 2-3** and **Supplementary Table 3**). Specifically, heritability in all undetermined strokes (N_{cases} = 4,574) was 18.2%. The mutually exclusive subsets of all undetermined stroke, categorized as incomplete/unclassified strokes (N_{cases} = 2,280) and cryptogenic/cardioembolic minor strokes (N_{cases} = 2,294), were also moderately heritable (28.2% and 23.7%, respectively; **Supplementary Figure 2**). We repeated all of the heritability calculations in GCTA[24] to evaluate the stability of the estimates across methods (**Supplementary Information** and **Supplementary Table 3**); estimates from GCTA were primarily similar to those from BOLT (**Supplementary Figure 4**). These heritability estimates suggest that, despite the fact that recent GWAS of undetermined stroke have not revealed any robustly-associated loci,[12,35] the phenotype(s) captured by this subtype are heritable, and likely represent either improperly-classified cases from the primary subtypes or novel stroke phenotypes awaiting further delineation.

Given the known phenotypic overlap of cardioembolic stroke and atrial fibrillation (89.5% of cardioembolic stroke cases in SiGN also have atrial fibrillation, **Supplementary Table 1**) as well as the similar heritable components of the two traits, we next aimed to evaluate how genetic susceptibility to atrial fibrillation might contribute to cardioembolic stroke. To measure correlation between the genetic architectures of atrial fibrillation and stroke, we first performed a series of GWAS in the SiGN data for each of the stroke subtypes as well as all stroke, using BOLT-LMM[17] (**Methods**). We then calculated the z-score (beta/standard error) of each SNP in each stroke subtype, all stroke, and from our analysis of atrial fibrillation described above. Next, we used summary-level results available from the most recent GWAS of atrial fibrillation[5] from AFGen and calculated the z-score for each SNP in that independent dataset. Measuring Pearson's correlation (r) between AFGen z-scores and z-scores from our own GWAS of atrial fibrillation, we found only a modest correlation between atrial fibrillation as measured in the two GWAS (r = 0.07 across ~7.8M SNPs, **Figure 1**). However, when we iteratively subset the AFGen GWAS results by the (absolute values of) z-scores of the SNPs, we found that correlation with our own atrial fibrillation GWAS in SiGN improved as the z-

score threshold became more stringent. For example, for those ~4.5M SNPs with |z| > 1 in AFGen, correlation with the SiGN atrial fibrillation GWAS was 0.12; for those ~1.9M SNPs with |z| > 3.5 in AFGen, correlation with the SiGN atrial fibrillation GWAS rose to 0.77 (**Figure 1** and **Supplementary Table 4**). Removing ±2Mb around the *PITX2* and *ZFHX3* loci only modestly impacted the correlation between AFGen and atrial fibrillation in SiGN (r = 0.63 for SNPs with |z| > 3.5; **Supplementary Figure 5** and **Supplementary Table 4**). As a null comparator, we performed correlations between the AFGen results with z-scores derived from the latest GWAS of educational attainment[25] and found that correlation remained at approximately zero regardless of the z-score threshold used (**Figure 1** and **Supplementary Table 4**).

To investigate the overlap of atrial fibrillation with stroke and its subtypes, we repeated these same correlation calculations, comparing AFGen results to GWAS results from each of the primary stroke subtypes (cardioembolic stroke, large artery atherosclerosis, and small artery occlusion) and the undetermined subtypes (**Figure 1**). The correlation between cardioembolic stroke and AFGen was highly similar to that of the results with atrial fibrillation in SiGN (r = 0.76 for AFGen SNPs with |z| > 3.5, likely due to high concordance between the atrial fibrillation and cardioembolic stroke phenotypes in SiGN; **Figure 1** and **Supplementary Figure 6**) whereas correlation between AFGen and either large artery atherosclerosis or small artery occlusion was essentially zero (**Figure 1**). Notably, the correlation between atrial fibrillation and the undetermined stroke subtypes increased steadily as we partitioned the AFGen data by z-score (all undetermined vs. AFGen r = 0.04 for AFGen SNPs with |z| > 1 and r = 0.16 for AFGen SNPs with |z| > 3.5; **Figure 1** and **Supplementary Table 4**), indicating that genome-wide, there is residual genetic correlation between atrial fibrillation and the undetermined stroke categories, some of which could represent causal atrial fibrillation stroke mechanisms in that subgroup.

To explicitly test what proportion of cardioembolic stroke heritability could be explained by atrial fibrillation loci, we used SNPs from an atrial fibrillation polygenic

risk score (PRS) independently derived from the AFGen GWAS[5] (**Methods**). Of the 1,168 SNPs used to generate this pre-established PRS, we identified 934 in the SiGN dataset with imputation quality > 0.8 and minor allele frequency >1%. To compute the PRS per individual, we weighted the imputed dosage of the risk allele (a number between 0 and 2) by the effect of the SNP (i.e., the beta coefficient) as reported in AFGen,[5] repeated this weighting across all SNPs, and then summed these weighted effects.

We tested the association of the atrial fibrillation PRS with cardioembolic stroke, using a logistic regression and adjusting for the top ten principal components and sex (**Methods**). We found the PRS to be strongly associated with cardioembolic stroke (odds ratio (OR) per 1 standard deviation (SD) of the PRS = 1.93 [95% confidence interval (CI): 1.34 - 1.44], $p = 1.01 \times 10^{-65}$; **Figure 2** and **Supplementary Table 5**), confirming the high genetic concordance of these phenotypes across SNPs which, individually, confer only a modest average association with atrial fibrillation. Expanding the set of SNPs used to construct the PRS to the original 934 SNPs ±25kb, ±50kb, and ±100kb (and subsequently linkage disequilibrium pruned at $r^2 = 0.2$ to ensure independence between markers) revealed a persistently strong though somewhat attenuated association between the PRS and cardioembolic stroke (PRS ± SNPs within 100Kb, $p = 4.47 \times 10^{-44}$, **Supplementary Table 6**). Additionally, adjusting the association with the original PRS for a set of clinical covariates (**Supplementary Table 7**) associated with atrial fibrillation including age, type 2 diabetes, cardiovascular disease, smoking, and presence of hypertension[36] only modestly reduced the persistent PRS signal in cardioembolic stroke (OR per 1 SD of the PRS = 1.40 [95% CI: 1.34 - 1.47], $p = 1.45 \times 10^{-48}$; **Supplementary Tables 5** and **6**).

None of the other subtypes we tested were significantly associated (Bonferroni-corrected $p = 5 \times 10^{-3}$, adjusting for five subtypes and the TOAST and CCS classification systems) to the PRS (all $p > 0.013$, **Figure 2** and **Supplementary Figure 6**).

As we had been leveraging referent samples for which atrial fibrillation status was mostly unknown, cryptic atrial fibrillation cases in the referent group were potentially negatively confounding associations between the AF PRS and stroke subtype. Of all available referents, 4,038 had information on atrial fibrillation diagnosis and of these samples, 3,861 are reported as not having atrial fibrillation. We therefore refined the set of referents by selecting only the 3,861 referents in whom atrial fibrillation status was non-missing and who did not have a reported atrial fibrillation diagnosis. Using this restricted set of samples, we observed a persistently strong association between the PRS and cardioembolic stroke (OR per 1 SD of the PRS = 1.44 [95% CI: 1.37 - 1.52], $p = 1.66 \times 10^{-42}$; **Supplementary Table 5**). The lack of association between the PRS and large artery atherosclerotic stroke or small artery occlusion strokes also persisted when using these refined controls ($p = 0.60$ and $p = 0.17$, respectively). Interestingly, we found the atrial fibrillation risk score to be modestly associated ($p < 5 \times 10^{-3}$) with the overall undetermined subtype (OR per 1 SD of the PRS = 1.07 [95% CI: 1.02 - 1.13], $p = 4.15 \times 10^{-3}$) (**Figure 2** and **Supplementary Table 5**). Further examination of the two mutually exclusive subgroups of the undetermined group revealed that the PRS associated significantly with the incomplete/unclassified categorization (OR per 1 SD = 1.09 [95% CI: 1.03 - 1.16], $p = 3.17 \times 10^{-3}$) (**Figure 2**) but not with cryptogenic/cardioembolic minor classification (OR per 1 SD = 1.06 [95% CI: 1.00 - 1.13], $p = 5.10 \times 10^{-2}$). Correcting for clinical covariates modestly changed the signal in the incomplete/unclassified phenotype ($p = 9.7 \times 10^{-3}$, **Figure 2**).

Lastly, we created a model in BOLT-LMM, fitting two genetic variance components: one component including SNPs for the genetic relationship matrix, and the second component including the original PRS SNPs from the atrial fibrillation PRS (including ±100kb around these SNPs, to include a sufficient number of markers to estimate variance explained). We found that the SNPs from the atrial fibrillation PRS explained 4.1% of the total 20.0% heritability in atrial fibrillation. In evaluating how much of cardioembolic stroke heritability that atrial fibrillation risk-increasing SNPs explain, we found a similar result: the component representing the atrial fibrillation risk score explained 4.5% (s.e. = 1.00%) of the total 19.5% genetic heritability in

cardioembolic stroke. Thus, atrial fibrillation genetic risk accounts for 23.1%, or approximately one-fifth, of the total heritability of cardioembolic stroke.

Discussion

We leveraged genome-wide common variation data from two large collaborative datasets to compare the genetic signatures shared between atrial fibrillation and stroke. Despite the fact that cardioembolic stroke often affects older adults with multiple clinical comorbidities[37] that could increase risk for atrial fibrillation due to non-genetic factors, we observed that nearly 20% of the variance in cardioembolic stroke risk can be explained by genetic risk factors for atrial fibrillation. Our results from genome-wide correlation analysis and construction of a polygenic risk score further indicate that cardioembolic stroke and atrial fibrillation share a highly similar genetic architecture, extending our understanding of the morbid consequences of heritable forms of the arrhythmia and highlighting an opportunity to leverage atrial fibrillation genetics to potentially improve treatment in ischemic stroke.

The use of polygenic risk scores for complex traits has proved an efficient means of understanding how genetic predisposition to diseases can overlap. Given the onslaught of genotyping data available for common diseases, polygenic risk scores are beginning to be used as a means of stratifying patients by risk (e.g., in breast cancer[38,39]) or predicting outcome (e.g., in neuropsychiatric disease[38,40]). More recently, polygenic risk scores derived from common genetic markers have been used to identify individuals in the general population with a four-fold risk for coronary disease,[41] proposed for inclusion in clinical workups of individuals with early-onset coronary artery disease,[42] and used to identify patients for whom lifestyle changes or statin intervention would be particularly beneficial.[43,44] Future work in atrial fibrillation and cardioembolic stroke may reveal how polygenic scores based on variants associated with atrial fibrillation could be used either to predict individuals at highest risk of cardioembolic stroke (and who therefore could be more closely monitored), or help distinguish between clinical etiopathologic subtypes of stroke.

Though our analysis was aimed at understanding the genetic overlap between cardioembolic stroke and atrial fibrillation, we additionally observed genome-wide

correlation between atrial fibrillation and undetermined stroke and an association between the atrial fibrillation polygenic risk score and stroke of unknown cause. Occult atrial fibrillation is well-known to cause strokes,[31,45] and identifying stroke patients with atrial fibrillation is an important clinical challenge since such patients are at high risk for recurrent stroke, which is preventable with anticoagulation.[46,47] Perhaps contrary to expectation, we specifically found the atrial fibrillation polygenic risk score to be more strongly associated with the subset of etiology-undetermined strokes with an incomplete clinical evaluation, as opposed to those with cryptogenic stroke of a presumed, but not demonstrated, embolic source. These associations could be due to physician biases in diagnostic workups leading to subtype assignment in this observational study, rather than supporting a low prevalence of occult atrial fibrillation in presumed embolic strokes of undetermined source. In aggregate, we submit that our findings indicate that atrial fibrillation genetic risk may augment clinical algorithms to determine stroke etiology, but will require further dedicated study.

While our work presented here has helped to define the overlapping common genetic risk factors for both atrial fibrillation and cardioembolic stroke, some limitations remain. First, given the phenotypic heterogeneity of stroke, genetic and environmental factors are not expected to carry a similar proportion of risk throughout an individual's lifespan. Furthermore, due to its late onset, stroke often co-occurs with a number of other common conditions[29] including hypertension, type 2 diabetes, and cardiovascular disease, the genetic contributions of which have not been explored herein. Further compounding the genetic study of ischemic stroke is the lifetime prevalence of stroke (>15%[48]), which increases the likelihood that referents studied here will have a stroke in the future; and sample size, which has lagged behind other GWAS efforts now interrogating tens or hundreds of thousands of samples,[49–52] a challenge further compounded by subtyping (nearly one-third of all cases fail to fall into one of the primary subtypes and are instead categorized as undetermined[34]). Reduced sample sizes impact power for discovery and make other analytic approaches — such as standard approaches for measuring trait correlation[17,24,53] —

unfeasible. Finally, the most recent GWAS of atrial fibrillation revealed a role for rare variation, untested here due to limitations in reliably imputing rare variants in the SiGN samples[12]; similarly, we did not test other types of genetic variation, such as insertions and deletions or larger structural variants.

We have shown that the genetic liability to atrial fibrillation is the same in individuals presenting with stroke as in primary studies of incident and prevalent atrial fibrillation. Additionally, we have provided updated heritability estimates of ischemic stroke and its subtypes, and demonstrated that genetic loci that fall far below the genome-wide significance threshold in studies of atrial fibrillation carry a substantial association with cardioembolic stroke risk. Although atrial fibrillation genetic risk did not associate with other primary stroke subtypes, we found an association with strokes of the 'undetermined' subtype, supporting the notion that undetected atrial fibrillation underlies a proportion of the stroke risk in these individuals. Given the increased prevalence of stroke in non-European populations, and especially in African Americans for whom risk of stroke is double that of individuals of European descent,[54] future studies will be needed to extend our results into populations of non-European ancestry. Further work will also need to incorporate emerging discoveries of rare genetic variants in atrial fibrillation, and explore the potential for genotype-based tools to assist in the diagnostic workup of individuals with ischemic stroke.

Tables and Figures

Table 1 | Atrial fibrillation and stroke cases in SiGN. Of the 13,390 stroke cases available in the SiGN dataset, a total of 3,190 cases had atrial fibrillation or other suggestive diagnoses. While the majority of these cases were subtyped as having a cardioembolic (CE) stroke, a fraction was distributed among the other stroke subtypes. Samples can appear more than once per row, but totals represent the number of unique atrial fibrillation samples in each stroke subtype. There are no subjects with atrial fibrillation or equivalent subtyped as "cryptogenic/CE minor" because such a diagnosis would remove them from this category.

	Phenotype	Total	Ischemic stroke subtype				
			Cardioembolic	Large artery atherosclerosis	Small artery occlusion	Undetermined: incomplete/ unclassified	Undetermined: cryptogenic/ CE minor
Atrial fibrillation diagnosis	Atrial fibrillation	1,751	1,495	63	32	151	0
	Paroxysmal atrial fibrillation	1,315	1,088	52	23	138	0
	Left atrial thrombus	48	37	3	3	4	0
	Sick sinus syndrome	79	65	5	3	4	0
	Atrial Flutter	106	90	4	2	10	0
	Total	3,190	2,684	123	61	298	0

Figure 1 | Genetic correlation between atrial fibrillation in the Atrial Fibrillation Genetics (AFGen) Consortium meta-analysis and GWAS of atrial fibrillation and ischemic stroke subtypes analysed in SiGN. Pearson's r correlation between SNP z-scores in the AFGen GWAS of atrial fibrillation and in GWAS of selected traits performed in the SiGN data. (a) GWAS of atrial fibrillation in AFGen and in SiGN correlate with increasing strength as SNP z-scores in AFGen increase. Correlation with educational attainment (performed separately, shown here as a null comparator) remains approximately zero across all z-score thresholds. (b) SNP effects in AFGen also correlate strongly with cardioembolic stroke in SiGN, but not with the other primary stroke subtypes. (c) Undetermined subtypes of stroke also show modest correlation to the genetic architecture of atrial fibrillation in AFGen. Panels d-f show genome-wide z-score distributions underlying correlations.

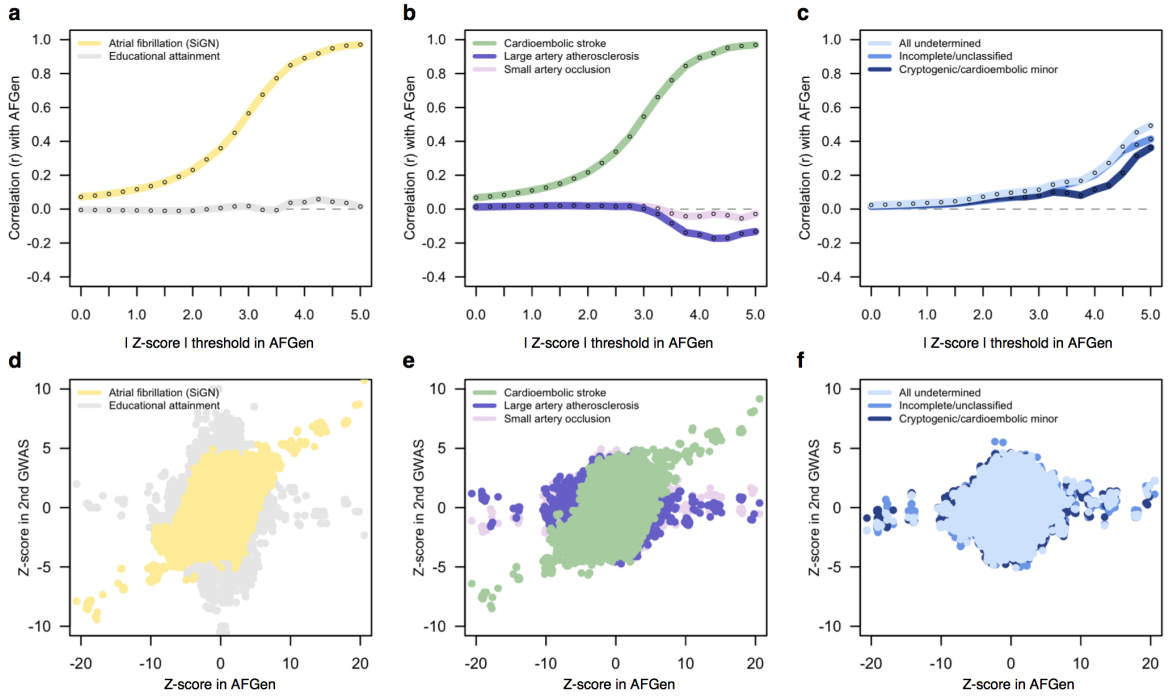
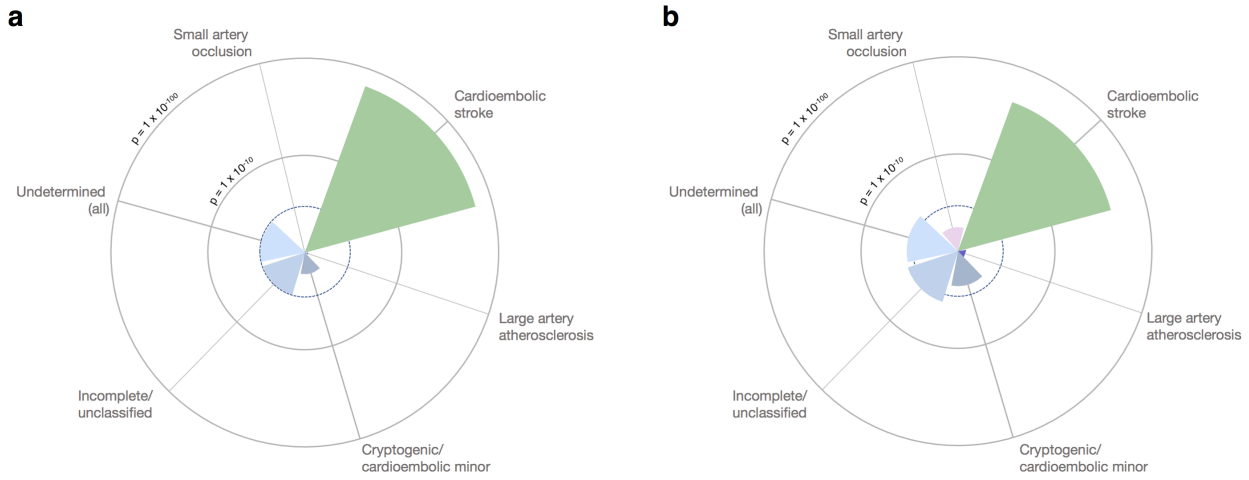


Figure 2 | Association of atrial fibrillation polygenic risk score in ischemic stroke subtypes. We constructed an independent polygenic risk score (PRS) from atrial fibrillation-associated SNPs identified in the AFGen GWAS, and tested associations between this PRS and ischemic stroke subtypes using (a) all available referents (N = 28,026) and (b) referents without atrial fibrillation (N = 3,861). The PRS strongly associated with cardioembolic stroke in both sets of samples. In the atrial fibrillation-free set of controls (panel b) we observed association of the PRS ($p < 5 \times 10^{-3}$, after adjusting for five subtypes and two sets of referents; indicated by the dashed dark blue line) with incomplete/unclassified stroke as well.



Acknowledgements

This work uses data from the National Institute of Neurological Disorders and Stroke - Stroke Genetics Network (NINDS-SiGN) Consortium and the Atrial Fibrillation Genetics (AFGen) Consortium. Members of NINDS-SiGN, the International Stroke Genetics Consortium, and AFGen are provided as an appendix to this paper.

Drs. Pulit, Mitchell, McArdle, and Kittner are supported by NIH grant R01NS100178. The NINDS-SiGN Consortium is supported by the NIH grants R01NS100178 and U01NS069208.

Dr. Anderson is supported in part by K23NS086873, R01NS103924 and a Massachusetts General Hospital Center for Genomic Medicine Catalyst Award.

Dr. Lubitz is supported by NIH grants K23HL114724 and a Doris Duke Charitable Foundation Clinical Scientist Development Award 2014105.

Dr. Weng is supported by an American Heart Association Postdoctoral Fellowship Award 17POST33660226.

Drs. Ellinor and Benjamin are supported by the NIH grants R01HL092577 and R01HL128914. Dr. Benjamin is additionally supported by the NIH grants 1RC1HL101056 and 1R01HL102214. Dr. Ellinor is additionally supported by the NIH grants R01HL104156 and K24HL105780; the National Heart, Lung, and Blood Institute (NHLBI); American Heart Association Established Investigator Awards 13EIA14220013; and the Fondation Leducq 14CVD01.

Disclosures/Competing Financial Interests

Dr. Lubitz receives sponsored research support from Boehringer Ingelheim, Bristol Myers Squibb, Bayer HealthCare, Biotronik, and has consulted for St. Jude Medical / Abbott and Quest Diagnostics. Dr. Anderson has consulted for ApoPharma. Dr. Ellinor is a principal investigator on a Bayer HealthCare grant to the Broad Institute related to the genetics and development of new therapeutics for cardiovascular diseases.

References

1. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, Gillum RF, Kim Y-H, McAnulty JH Jr, Zheng Z-J, Forouzanfar MH, Naghavi M, Mensah GA, Ezzati M, Murray CJL. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation*. 2014;129:837–847.

2. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke*. 1991;22:983–988.

3. World Health Organization. WHO | The top 10 causes of death. 2014. Available from: http://www.who.int/mediacentre/factsheets/fs310/en/

4. Feigin VL, Lawes CM, Bennett D a., Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol*. 2009;8:355–369.

5. Christophersen IE, Rienstra M, Roselli C, Yin X, Geelhoed B, Barnard J, Lin H, Arking DE, Smith AV, Albert CM, Chaffin M, Tucker NR, Li M, Klarin D, Bihlmeyer NA, Low S-K, Weeke PE, Müller-Nurasyid M, Smith JG, Brody JA, Niemeijer MN, Dörr M, Trompet S, Huffman J, Gustafsson S, Schurmann C, Kleber ME, Lyytikäinen L-P, Seppälä I, Malik R, Horimoto ARVR, Perez M, Sinisalo J, Aeschbacher S, Thériault S, Yao J, Radmanesh F, Weiss S, Teumer A, Choi SH, Weng L-C, Clauss S, Deo R, Rader DJ, Shah SH, Sun A, Hopewell JC, Debette S, Chauhan G, Yang Q, Worrall BB, Paré G, Kamatani Y, Hagemeijer YP, Verweij N, Siland JE, Kubo M, Smith JD, Van Wagoner DR, Bis JC, Perz S, Psaty BM, Ridker PM, Magnani JW, Harris TB, Launer LJ, Shoemaker MB, Padmanabhan S, Haessler J, Bartz TM, Waldenberger M, Lichtner P, Arendt M, Krieger JE, Kähönen M, Risch L, Mansur AJ, Peters A, Smith BH, Lind L, Scott SA, Lu Y, Bottinger EB, Hernesniemi J, Lindgren CM, Wong JA, Huang J, Eskola M, Morris AP, Ford I, Reiner AP, Delgado G, Chen LY, Chen Y-DI, Sandhu RK, Li M, Boerwinkle E, Eisele L, et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet*. 2017;57:289.

6. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NLM, Jackson C, Farrall M, Rothwell PM, Sudlow C, Dichgans M, Markus HS. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genome-wide associations. *Stroke*. 2012;43:3161–3167.

7. Lubitz SA, Ozcan C, Magnani JW, Kääb S, Benjamin EJ, Ellinor PT. Genetics of atrial fibrillation: implications for future research directions and personalized medicine. *Circ Arrhythm Electrophysiol*. 2010;3:291–299.

8. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir

R, Sverrisson JT, Kostulas K, Ng MCY, Baum L, So WY, Wong KS, Chan JCN, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, Smith EE, Rosand J, Hillert J, Ma RCW, Ellinor PT, Thorgeirsson G, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature*. 2007;448:353–357.

9. Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, Kostulas K, Kuhlenbäumer G, Bevan S, Jonsdottir T, Bjarnason H, Saemundsdottir J, Palsson S, Arnar DO, Holm H, Thorgeirsson G, Valdimarsson EM, Sveinbjörnsdottir S, Gieger C, Berger K, Wichmann HE, Hillert J, Markus H, Gulcher JR, Ringelstein EB, Kong A, Dichgans M, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol*. 2008;64:402–409.

10. Gudbjartsson DF, Holm H, Gretarsdottir S, Thorleifsson G, Walters GB, Thorgeirsson G, Gulcher J, Mathiesen EB, Njølstad I, Nyrnes A, Wilsgaard T, Hald EM, Hveem K, Stoltenberg C, Kucera G, Stubblefield T, Carter S, Roden D, Ng MCY, Baum L, So WY, Wong KS, Chan JCN, Gieger C, Wichmann H-E, Gschwendtner A, Dichgans M, Kuhlenbäumer G, Berger K, Ringelstein EB, Bevan S, Markus HS, Kostulas K, Hillert J, Sveinbjörnsdóttir S, Valdimarsson EM, Løchen M-L, Ma RCW, Darbar D, Kong A, Arnar DO, Thorsteinsdottir U, Stefansson K. A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet*. 2009;41:876–878.

11. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017;101:5–22.

12. Pulit SL, McArdle PF, Wong Q, Malik R, Gwinn K, Achterberg S, Algra A, Amouyel P, Anderson CD, Arnett DK, Arsava EM, Attia J, Ay H, Bartz TM, Battey T, Benavente OR, Bevan S, Biffi A, Bis JC, Blanton SH, Boncoraglio GB, Brown Jr RD, Burgess AI, Carrera, Smith SNC, Chasman DI, Chauhan G, Chen WM, Cheng YC, Chong M, Cloonan LK, Cole JW, Cotlarciuc I, Cruchaga C, Cuadrado-Godia E, Dave T, Dawson J, Debette S, Delavaran H, Dell CA, Dichgans M, Doheny KF, Dong C, Duggan DJ, Engström G, Evans MK, Pallejà XE, Faul JD,Fernández-Cadenas I, Fornage M, Frossard PM, Furie K, Gamble DM, Gieger C, Giese AK, Giralt-Steinhauer E, González HM, Goris A, Gretarsdottir S, Grewal RP, Grittner R, Gustafsson S, Han B, Hankey GJ, Heitsch L, Higgins P, Hochberg MC, Holliday E, Hopewell EC, Horenstein RB, Howard G, Ikram MA, Ilinca A, Ingelsson E, Irvin MR, Jackson RD, Jern C, Conde JJ, Johnson JA, Jood K, Kahn MS, Kaplan R, Kappelle LJ, Kardia SLR, Keene KL, Kissela BM, O Kleindorfer D, Koblar S, Labovitz D, Launer LJ, Laurie CC, Laurie CA, Lee CH, Lee JM, Lehm M, Lemmens R, Levi C, Leys D, Lindgren A, Longstreth Jr WT, Maguire J, Manichaikul Ani, Markus HS, McClure LA, McDonough CW, Meisinger C, Melander O, Meschia JF, Mola-Caminal M, Montaner J, Mosley TH, Müller-Nurasyid M, Nalls MA, O'Connell JR, O'Donnell M, Ois A, Papanicolaou GJ, Paré G, Peddareddygari LR, Pedersén A, Pera J, Peters A, Poole D, Psaty BM, Rabionet R, Raffeld MR, Rannikmäe K, Rasheed A, Redfors P, Reiner AP,

Rexrode K, Ribasés M, Rich SS, Robberecht W, Rodriguez-Campello A, Rolfs A, Roquer J, Rose LM, Rosenbaum D, Rost NS, Rothwell PM, Rundek T, Ryan KA, Sacco RL, Sale MM, Saleheen D, Salomaa V, Sánchez-Mora C, Schmidt CO, Schmidt H, Schmidt R, Schürks M, Scott R, Segal HC, Seiler S, Seshadri S, Sharma P, Shuldiner AR, Silver B, Slowik A, Smith JA, Söderholm M, Soriano C, Sparks MJ, Stanne T, Stefansson K, Stine OC, Strauch K, Sturm J, Sudlow CLM, Tajuddin SM, Talbert RL, Tatlisumak T, Thijs V, Thorleifsson G, Thorsteindottir U, Tiedt S, Traylor M, Trompet S, Valant V, Waldenberger M, Walters M, Wang L, Wassertheil-Smoller S, Weir DR, Wiggins KL, Williams SR, Wloch-Kopec D, Woo D, Woodfield D, Wu O, Xu H, Zonderman AB, Australian Stroke Genetics Consortium, Cervical Artery Dissection and Ischemic Stroke Patients (CADISP) study, Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, Consortium of Minority Population genome-wide Association Studies of Stroke (COMPASS), METASTROKE consortium, Wellcome Trust Case-Control Consortium, Worrall BB, de Bakker PIW, Kittner SJ, Mitchell BD, Rosand J. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurology*. 2015;15:174–184.

13. Ay H, Benner T, Arsava EM, Furie KL, Singhal AB, Jensen MB, Ayata C, Towfighi A, Smith EE, Chong JY, Koroshetz WJ, Sorensen AG. A computerized algorithm for etiologic classification of ischemic stroke: The causative classification of stroke system. *Stroke*. 2007;38:2979–2984.

14. Arsava EM, Ballabio E, Benner T, Cole JW, Delgado-Martinez MP, Dichgans M, Fazekas F, Furie KL, Illoh K, Jood K, Kittner S, Lindgren AG, Majersik JJ, MacLeod MJ, Meurer WJ, Montaner J, Olugbodi AA, Pasdar A, Redfors P, Schmidt R, Sharma P, Singhal AB, Sorensen AG, Sudlow C, Thijs V, Worrall BB, Rosand J, Ay H. The causative classification of stroke system: An international reliability and optimization study. *Neurology*. 2010;75:1277–1284.

15. Adams HP Jr, Birgitte H. Bendixen P, Jaap Kappelle L, Biller J, Love BB, Gordon DL, Eugene MarshIII E, and the TOAST Investigators. Classification of Subtype of Acute Ischemic Stroke: Definitions for Use in a Multicenter Clinical Trial. *Stroke*. 1993;24:35–41.

16. Goldstein LB, Jones MR, Matchar DB, Edwards LJ, Hoff J, Chilukuri V, Armstrong SB, Horner RD. Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. *Stroke*. 2001;32:1091–1098.

17. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, Patterson N, Price AL. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47:284–290.

18. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods*. 2011;8:833–835.

19. Widmer C, Lippert C, Weissbrod O, Fusi N, Kadie C, Davidson R, Listgarten J, Heckerman D. Further improvements to linear mixed models for genome-wide association studies. *Sci Rep*. 2014;4:6874.

20. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*. 2014;46:100–106.

21. 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.

22. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:1–16.

23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a. R, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575.

24. 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.

25. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen G-B, Emilsson V, Meddens SFW, Oskarsson S, Pickrell JK, Thom K, Timshel P, de Vlaming R, Abdellaoui A, Ahluwalia TS, Bacelis J, Baumbach C, Bjornsdottir G, Brandsma JH, Pina Concas M, Derringer J, Furlotte NA, Galesloot TE, Girotto G, Gupta R, Hall LM, Harris SE, Hofer E, Horikoshi M, Huffman JE, Kaasik K, Kalafati IP, Karlsson R, Kong A, Lahti J, Lee SJ van der, deLeeuw C, Lind PA, Lindgren K-O, Liu T, Mangino M, Marten J, Mihailov E, Miller MB, van der Most PJ, Oldmeadow C, Payton A, Pervjakova N, Peyrot WJ, Qian Y, Raitakari O, Rueedi R, Salvi E, Schmidt B, Schraut KE, Shi J, Smith AV, Poot RA, St Pourcain B, Teumer A, Thorleifsson G, Verweij N, Vuckovic D, Wellmann J, Westra H-J, Yang J, Zhao W, Zhu Z, Alizadeh BZ, Amin N, Bakshi A, Baumeister SE, Biino G, Bønnelykke K, Boyle PA, Campbell H, Cappuccio FP, Davies G, De Neve J-E, Deloukas P, Demuth I, Ding J, Eibich P, Eisele L, Eklund N, Evans DM, Faul JD, Feitosa MF, Forstner AJ, Gandin I, Gunnarsson B, Halldórsson BV, Harris TB, Heath AC, Hocking LJ, Holliday EG, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 2016;533:539.

26. Weng L-C, Preis SR, Hulme OL, Larson MG, Choi SH, Wang B, Trinquart L, McManus DD, Staerk L, Lin H, Lunetta KL, Ellinor PT, Benjamin EJ, Lubitz SA. Genetic Predisposition, Clinical Risk Factor Burden, and Lifetime Risk of Atrial Fibrillation. *Circulation*. Epub ahead of print, 2017.

27. Franz MR, Karasik PL, Li C, Moubarak J, Chavez M. Electrical remodeling of the human atrium: similar effects in patients with chronic atrial fibrillation and atrial

flutter. *J Am Coll Cardiol*. 1997;30:1785–1792.

28. Lamas GA, Lee KL, Sweeney MO, Silverman R, Leon A, Yee R, Marinchak RA, Flaker G, Schron E, Orav EJ, Hellkamp AS, Greer S, McAnulty J, Ellenbogen K, Ehlert F, Freedman RA, Estes NAM 3rd, Greenspon A, Goldman L, Mode Selection Trial in Sinus-Node Dysfunction. Ventricular pacing or dual-chamber pacing for sinus-node dysfunction. *N Engl J Med*. 2002;346:1854–1862.

29. Petty GW, Brown RD, Whisnant JP, Sicks JD, Michael O'Fallon W, Wiebers DO. Ischemic Stroke Subtypes. *Stroke*. 1999;30:2513–2516.

30. Kolominsky-Rabas PL, Weber M, Gefeller O, Neundoerfer B, Heuschmann PU. Epidemiology of Ischemic Stroke Subtypes According to TOAST Criteria. *Stroke*. 2001;32:2735–2740.

31. Sanna T, Diener H-C, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, Rymer MM, Thijs V, Rogers T, Beckers F, Lindborg K, Brachmann J, CRYSTAL AF Investigators. Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med*. 2014;370:2478–2486.

32. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, Singer DE. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA*. 2001;285:2370–2375.

33. Weng L-C, Choi SH, Klarin D, Gustav Smith J, Loh P-R, Chaffin M, Roselli C, Hulme OL, Lunetta KL, Dupuis J, Benjamin EJ, Newton-Cheh C, Kathiresan S, Ellinor PT, Lubitz SA. Heritability of Atrial Fibrillation. *Circ Cardiovasc Genet*. 2017;10:e001838.

34. NINDS Stroke Genetics Network (SiGN), International Stroke Genetics Consortium. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol*. 2015;15:4–7.

35. Malik R, Traylor M, Pulit SL, Bevan S, Hopewell JC, Holliday EG, Zhao W, Abrantes P, Amouyel P, Attia JR, Battey TWK, Berger K, Boncoraglio GB, Chauhan G, Cheng Y-C, Chen W-M, Clarke R, Cotlarciuc I, Debette S, Falcone GJ, Ferro JM, Gamble DM, Ilinca A, Kittner SJ, Kourkoulis CE, Lemmens R, Levi CR, Lichtner P, Lindgren A, Liu J, Meschia JF, Mitchell BD, Oliveira SA, Pera J, Reiner AP, Rothwell PM, Sharma P, Slowik A, Sudlow CLM, Tatlisumak T, Thijs V, Vicente AM, Woo D, Seshadri S, Saleheen D, Rosand J, Markus HS, Worrall BB, Dichgans M, ISGC Analysis Group, METASTROKE collaboration, Wellcome Trust Case Control Consortium 2 (WTCCC2), NINDS Stroke Genetics Network (SiGN). Low-frequency and common genetic variation in ischemic stroke: The METASTROKE collaboration. *Neurology*. 2016;86:1217–1226.

36. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, Sinner MF, Sotoodehnia N, Fontes JD, Janssens ACJW, Kronmal RA, Magnani JW, Witteman JC, Chamberlain AM, Lubitz SA, Schnabel RB, Agarwal SK, McManus DD, Ellinor PT, Larson MG, Burke GL, Launer LJ, Hofman A, Levy D, Gottdiener JS, Kääb S,

Couper D, Harris TB, Soliman EZ, Stricker BHC, Gudnason V, Heckbert SR, Benjamin EJ. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc*. 2013;2:e000102.

37. Henninger N, Goddeau RP Jr, Karmarkar A, Helenius J, McManus DD. Atrial Fibrillation Is Associated With a Worse 90-Day Outcome Than Other Cardioembolic Stroke Subtypes. *Stroke*. 2016;47:1486–1492.

38. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome Med*. 2017;9:96.

39. Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Luben R, Brown J, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Czene K, Darabi H, Eriksson M, Peto J, Dos-Santos-Silva I, Dudbridge F, Johnson N, Schmidt MK, Broeks A, Verhoef S, Rutgers EJ, Swerdlow A, Ashworth A, Orr N, Schoemaker MJ, Figueroa J, Chanock SJ, Brinton L, Lissowska J, Couch FJ, Olson JE, Vachon C, Pankratz VS, Lambrechts D, Wildiers H, Van Ongeval C, van Limbergen E, Kristensen V, Grenaker Alnæs G, Nord S, Borresen-Dale A-L, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Burwinkel B, Marme F, Schneeweiss A, Sohn C, Trentham-Dietz A, Newcomb P, Titus L, Egan KM, Hunter DJ, Lindstrom S, Tamimi RM, Kraft P, Rahman N, Turnbull C, Renwick A, Seal S, Li J, Liu J, Humphreys K, Benitez J, Pilar Zamora M, Arias Perez JI, Menéndez P, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Bogdanova NV, Antonenkova NN, Dörk T, Anton-Culver H, Neuhausen SL, Ziogas A, Bernstein L, Devilee P, Tollenaar RAEM, Seynaeve C, van Asperen CJ, Cox A, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107.

40. Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, O'Reilly P, Curtis C, Kolliakou A, Patel H, Newhouse S, Traylor M, Ajnakina O, Mondelli V, Marques TR, Gardner-Sood P, Aitchison KJ, Powell J, Atakan Z, Greenwood KE, Smith S, Ismail K, Pariante C, Gaughran F, Dazzan P, Markus HS, David AS, Lewis CM, Murray RM, Breen G. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. *Biol Psychiatry*. 2017;81:470–477.

41. Khera AV, Chaffin M, Aragam KG, Emdin CA, Klarin D, Haas ME, Roselli C, Natarajan P, Kathiresan S. Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. bioRxiv. 2017.

42. Thériault S, Lali R, Chong M, Velianou JL, Natarajan MK, Paré G. Polygenic Contribution in Individuals With Early-Onset Coronary Artery Disease. *Circ Genom Precis Med*. 2018;11:e001849.

43. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DI, Baber U, Mehran R, Rader DJ, Fuster V, Boerwinkle E, Melander O, Orho-

Melander M, Ridker PM, Kathiresan S. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med*. 2016;375:2349–2358.

44. Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield M, Devlin JJ, Nordio F, Hyde C, Cannon CP, Sacks F, Poulter N, Sever P, Ridker PM, Braunwald E, Melander O, Kathiresan S, Sabatine MS. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet*. 2015;385:2264–2271.

45. Gladstone DJ, Spring M, Dorian P, Panzov V, Thorpe KE, Hall J, Vaid H, O'Donnell M, Laupacis A, Côté R, Sharma M, Blakely JA, Shuaib A, Hachinski V, Coutts SB, Sahlas DJ, Teal P, Yip S, Spence JD, Buck B, Verreault S, Casaubon LK, Penn A, Selchen D, Jin A, Howse D, Mehdiratta M, Boyle K, Aviv R, Kapral MK, Mamdani M, EMBRACE Investigators and Coordinators. Atrial fibrillation in patients with cryptogenic stroke. *N Engl J Med*. 2014;370:2467–2477.

46. ACTIVE Investigators, Connolly SJ, Pogue J, Hart RG, Hohnloser SH, Pfeffer M, Chrolavicius S, Yusuf S. Effect of clopidogrel added to aspirin in patients with atrial fibrillation. *N Engl J Med*. 2009;360:2066–2078.

47. Diener H-C, Eikelboom J, Connolly SJ, Joyner CD, Hart RG, Lip GYH, O'Donnell M, Hohnloser SH, Hankey GJ, Shestakovska O, Yusuf S, AVERROES Steering Committee and Investigators. Apixaban versus aspirin in patients with atrial fibrillation and previous stroke or transient ischaemic attack: a predefined subgroup analysis from AVERROES, a randomised trial. *Lancet Neurol*. 2012;11:225–231.

48. Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Au R, Kannel WB, Wolf PA. The Lifetime Risk of Stroke: Estimates From the Framingham Study. *Stroke*. 2006;37:345–350.

49. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J 'an, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46:1173–1186.

50. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, Strawbridge RJ, Pers TH, Fischer K, Justice AE, Workalemahu T, Wu JMW, Buchkovich ML, Heard-Costa NL, Roman TS, Drong AW, Song C, Gustafsson S, Day FR, Esko T, Fall T, Kutalik Z, Luan J 'an, Randall JC, Scherag A, Vedantam S, Wood AR, Chen J, Fehrmann R, Karjalainen J, Kahali B, Liu C-T, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bragg-Gresham JL, Buyske S, Demirkan A, Ehret GB, Feitosa MF, Goel A, Jackson AU, Johnson T, Kleber ME, Kristiansson K, Mangino M, Mateo Leach I, Medina-Gomez C, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Stančáková A, Ju Sung Y, Tanaka T, Teumer A, Van Vliet-Ostaptchouk JV, Yengo L, Zhang W, Albrecht E, Ärnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Böhringer S, Bonnet F, Böttcher Y, Bruinenberg M, Carba DB, Caspersen IH, Clarke R, Daw EW, Deelen J, Deelman E, Delgado G, Doney ASF, Eklund N, Erdos MR, Estrada K, Eury E, Friedrich N, Garcia ME, Giedraitis V, Gigante B, Go AS, Golay A, Grallert H, Grammer TB, Gräßler J, Grewal J, Groves CJ, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518:187–196.

51. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J 'an, Mägi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Zhao JH, Zhao W, Chen J, Fehrmann R, Hedman ÅK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Leach IM, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stančáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Ärnlöv J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Blüher M, Böhringer S, Bonnycastle LL, Böttcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Chen Y-DI, Clarke R, Daw EW, de Craen AJM, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197–206.

52. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans P a., Lee P, Bulik-Sullivan B, Collier D a., Huang H, Pers TH, Agartz I, Agerbo E, Albus M, Alexander M, Amin F, Bacanu S a., Begemann M, Belliveau R a. Jr, Bene J, Bergen SE, Bevilacqua E, Bigdeli TB, Black DW, Bruggeman R, Buccola NG, Buckner RL, Byerley W, Cahn W, Cai G, Campion D, Cantor RM, Carr VJ, Carrera N, Catts SV, Chambert KD, Chan RCK, Chen RYL, Chen EYH, Cheng W, Cheung EFC, Ann Chong S, Robert Cloninger C, Cohen D, Cohen N, Cormican P, Craddock N, Crowley JJ, Curtis D, Davidson M, Davis KL, Degenhardt F, Del Favero J, Demontis D, Dikeos D, Dinan T, Djurovic S, Donohoe G, Drapeau E, Duan J, Dudbridge F, Durmishi N, Eichhammer P, Eriksson J, Escott-Price V, Essioux L, Fanous AH, Farrell MS, Frank J, Franke L, Freedman R, Freimer NB, Friedl M, Friedman JI, Fromer M, Genovese G, Georgieva L, Giegling I, Giusti-

Rodríguez P, Godard S, Goldstein JI, Golimbet V, Gopal S, Gratten J, de Haan L, Hammer C, Hamshere ML, Hansen M, Hansen T, Haroutunian V, Hartmann AM, Henskens F a., Herms S, Hirschhorn JN, Hoffmann P, Hofman A, Hollegaard MV, Hougaard DM, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.

53. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Consortium SWG of TPG, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–295.

54. Gorelick PB. Cerebrovascular Disease in African Americans. *Stroke*. 1998;29:2656–2664.