

1 Atrial Fibrillation Genetic Risk Differentiates 2 Cardioembolic Stroke from other Stroke Subtypes

3
4 Sara L. Pulit, PhD^{1,2,3}; Lu-Chen Weng, PhD^{3,4}; Patrick F McArdle, PhD⁵; Ludovic
5 Trinquart, PhD^{6,7}; Seung Hoan Choi, PhD³; Braxton D. Mitchell, PhD^{5,8}; Jonathan
6 Rosand, MD^{3,9,10}; Paul I W de Bakker, PhD^{1,11}; Emelia J Benjamin, MD, ScM^{6,12,13};
7 Patrick T Ellinor, MD, PhD^{3,4,14}, Steven J Kittner, MD^{15,16}; Steven A Lubitz, MD,
8 MPH^{*3,4,14}; Christopher D Anderson, MD^{*3,9,10}, on behalf of the Atrial Fibrillation
9 Genetics Consortium and the International Stroke Genetics Consortium.

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

33
34 * These authors contributed equally to supervision of this study
35

36
37
38 Supplemental data: Figures e1-7, Tables e1-8, and Supplementary Methods

39
40 Title: 95 characters
41 Paper: 3,411 words
42 Abstract: 224 words
43 Figures: 2
44 Tables: 1
45 References: 39
46

1 Publication history: This manuscript was previously published in bioRxiv:

2 <https://doi.org/10.1101/239269>

3 Corresponding authors

4

5 Dr. Steven A. Lubitz

6 Cardiovascular Research Center and Cardiac Arrhythmia Service

7 Massachusetts General Hospital

8 55 Fruit Street

9 GRB 109

10 Boston, MA, 02114, USA

11 slubitz@mgh.harvard.edu

12

13 Dr. Christopher D. Anderson

14 Center for Genomic Medicine

15 Massachusetts General Hospital

16 185 Cambridge Street, CPZN-6818

17 Boston, MA, 02114, USA

18 cdanderson@mgh.harvard.edu

19

20

21

1 Author Contributions

2

3 All named authors have contributed meaningfully to the present study. Specific
4 contributions for each author are described below.

5

6 **S.L. Pulit:** conception of research design, data analysis, drafting of manuscript,
7 critical revision of manuscript

8 **L.C. Weng:** data analysis, critical revision of manuscript

9 **P.F. McArdle:** data acquisition, data analysis, critical revision of manuscript

10 **L. Trinquart:** data acquisition, critical revision of manuscript

11 **S.H. Choi:** data acquisition, critical revision of manuscript

12 **B.D. Mitchell:** data acquisition, study supervision, critical revision of manuscript

13 **J. Rosand:** data acquisition, study supervision, critical revision of manuscript

14 **P.I.W. de Bakker:** study supervision, critical revision of manuscript

15 **E.J. Benjamin:** data acquisition, study supervision, critical revision of manuscript

16 **P.T. Ellinor:** data acquisition, study supervision, critical revision of manuscript

17 **S.J. Kittner:** data acquisition, study supervision, critical revision of manuscript

18 **S.A. Lubitz:** conception of research design, study supervision, drafting of
19 manuscript, critical revision of manuscript

20 **C.D. Anderson:** conception of research design, study supervision, drafting of
21 manuscript, critical revision of manuscript

22

23

24 Disclosures/Competing Financial Interests

25

26 Dr. Pulit reports no disclosures. Dr. Weng reports no disclosures. Dr. McArdle reports
27 no disclosures. Dr. Trinquart reports no disclosures. Dr. Choi reports no disclosures.
28 Dr. Mitchell reports no disclosures. Dr. Rosand reports no disclosures. Dr. de Bakker
29 reports no disclosures. Dr. Benjamin reports no disclosures. Dr. Ellinor is a principal
30 investigator on a Bayer HealthCare grant to the Broad Institute related to the genetics
31 and development of new therapeutics for cardiovascular diseases. Dr. Kittner reports
32 no disclosures. Dr. Lubitz receives sponsored research support from Boehringer
33 Ingelheim, Bristol Myers Squibb, Bayer HealthCare, Biotronik, and has consulted for
34 St. Jude Medical / Abbott and Quest Diagnostics. Dr. Anderson has consulted for
35 ApoPharma.

36

37

38

39

1 Abstract

2

3 *Objective:* We sought to assess whether genetic risk factors for atrial fibrillation can
4 explain cardioembolic stroke risk.

5

6 *Methods:* We evaluated genetic correlations between a prior genetic study of AF and
7 AF in the presence of cardioembolic stroke using genome-wide genotypes from the
8 Stroke Genetics Network (N = 3,190 AF cases, 3,000 cardioembolic stroke cases, and
9 28,026 referents). We tested whether a previously-validated AF polygenic risk score
10 (PRS) associated with cardioembolic and other stroke subtypes after accounting for
11 AF clinical risk factors.

12

13 *Results:* We observed strong correlation between previously reported genetic risk for
14 AF, AF in the presence of stroke, and cardioembolic stroke (Pearson's $r=0.77$ and
15 0.76 , respectively, across SNPs with $p < 4.4 \times 10^{-4}$ in the prior AF meta-analysis). An
16 AF PRS, adjusted for clinical AF risk factors, was associated with cardioembolic stroke
17 (odds ratio (OR) per standard deviation (sd) = 1.40 , $p = 1.45 \times 10^{-48}$), explaining
18 $\sim 20\%$ of the heritable component of cardioembolic stroke risk. The AF PRS was also
19 associated with stroke of undetermined cause (OR per sd = 1.07 , $p = 0.004$), but no
20 other primary stroke subtypes (all $p > 0.1$).

21

22 *Conclusions:* Genetic risk for AF is associated with cardioembolic stroke, independent
23 of clinical risk factors. Studies are warranted to determine whether AF genetic risk
24 can serve as a biomarker for strokes caused by AF.

25

26

1 Introduction

2

3 Atrial fibrillation affects nearly 34 million individuals worldwide¹ and is associated with
4 a five-fold increased risk of ischemic stroke,² a leading cause of death and disability.^{3,4}
5 Atrial fibrillation promotes blood clot formation in the heart which can embolize
6 distally, and is a leading cause of cardioembolism. Secondary prevention of
7 cardioembolic stroke is directed at identifying atrial fibrillation as a potential cause,
8 and initiating anticoagulation to prevent recurrences. Yet atrial fibrillation can remain
9 occult even after extensive workup owing to the paroxysmal nature and fact that it
10 can be asymptomatic. Since both atrial fibrillation and stroke are heritable, and since
11 there is a compelling clinical need to determine whether stroke survivors have atrial
12 fibrillation as an underlying cause, we sought to determine whether genetic risk of
13 cardioembolic stroke can be approximated by measuring genetic susceptibility to
14 atrial fibrillation.

15

16 Recent genome-wide association studies (GWAS) have demonstrated that both atrial
17 fibrillation⁵ and ischemic stroke^{6,7} are complex disorders with polygenic architectures.
18 The top loci for cardioembolic stroke, on chromosome 4q25 upstream of *PITX2* and
19 on 16q22 near *ZFHX3*, are both leading risk loci for atrial fibrillation.⁸⁻¹⁰ Despite
20 overlap in top risk loci, the genetic susceptibility to both atrial fibrillation and
21 cardioembolic stroke is likely to involve the aggregate contributions of hundreds or
22 thousands of loci, consistent with other polygenic conditions.¹¹

23

24 To understand whether genetic risk for atrial fibrillation is an important and
25 potentially useful determinant of overall cardioembolic stroke risk, we analyzed
26 13,390 ischemic stroke cases and 28,026 referents from the NINDS-Stroke Genetics
27 Network (SiGN)¹² with genome-wide genotyping data. First, we assessed whether
28 stroke patients with atrial fibrillation have a genetic predisposition to the arrhythmia,
29 leveraging additional GWAS data from the Atrial Fibrillation Genetics Consortium
30 (AFGen). Second, we compared genetic risk factors for atrial fibrillation and stroke to
31 ascertain the extent to which heritable risk of cardioembolic stroke is explained by
32 genetic risk factors for atrial fibrillation.

1 Methods

2

3 *The Stroke Genetics Network (SiGN)*

4

5 The Stroke Genetics Network (SiGN) was established with the aim of performing the
6 largest genome-wide association study (GWAS) of ischemic stroke to date. **The study**
7 **design has been described previously¹² and is summarized in the **Supplementary****
8 **Methods.** Briefly, subjects in SiGN were classified into stroke subtypes using the
9 Causative Classification System (CCS), which subtypes cases through an automated,
10 web-based system that accounts for clinical data, test results, and imaging
11 information.^{13,14} Within CCS, there are two sub-categories: CCS causative, which
12 does not allow for competing subtypes in a single sample; and CCS phenotypic, which
13 does. Additionally, ~74% of samples were subtyped using the TOAST subtyping
14 system.¹⁵ After quality control (QC), the SiGN dataset comprised 16,851 ischemic
15 stroke cases and 32,473 stroke-free controls (**Supplementary Methods** and
16 **Supplementary Table 1**). Here, we analyze only the European- and African-
17 ancestry samples (13,390 cases and 28,026 controls).

18

19 *Standard Protocol Approvals, Registrations, and Patient Consents*

20

21 All cohorts included in the SiGN dataset received approval from the cohort-specific
22 ethical standards committee. Cohorts received written informed consent from all
23 patients or guardians of patients participating in the study, where applicable. Details
24 on sample collection have been described previously.¹²

25

26 *Identifying atrial fibrillation cases and controls*

27

28 We defined atrial fibrillation in SiGN on the basis of five variables available in the CCS
29 phenotyping system: (i) atrial fibrillation, (ii) paroxysmal atrial fibrillation, (iii) atrial
30 flutter, (iv) sick sinus syndrome, and (v) atrial thrombus. This definition yielded 3,190
31 atrial fibrillation cases for analysis. We also defined a strict case set based on "atrial

1 fibrillation" only (N = 1,751 cases) for sensitivity analyses (**Supplementary**
2 **Methods** and **Supplementary Figure 1**).

3
4 From the 28,026 controls, we established a set of 3,861 control individuals in whom
5 atrial fibrillation was indicated as not present. For the remaining subjects, we
6 assumed that individuals did not have atrial fibrillation since atrial fibrillation status
7 for most control samples in SiGN is unknown.

8
9 *Genome-wide association testing of ischemic stroke subtypes and atrial fibrillation in*
10 *SiGN*

11
12 We merged genotype dosages together and kept SNPs with imputation quality > 0.8
13 and minor allele frequency (MAF) > 1% (**Supplementary Methods**). We performed
14 association testing using a linear mixed model implemented in BOLT-LMM.¹⁶ We
15 adjusted the model for the top ten principal components (PCs) and sex, in addition
16 to the genetic relationship matrix (GRM; **Supplementary Methods**).¹⁶ We
17 performed GWAS in atrial fibrillation and each of the stroke subtypes available in
18 SiGN. Results were unadjusted for age, as adjusting for age in the atrial fibrillation
19 GWAS gave results highly concordant with the age-unadjusted results
20 (**Supplementary Results**).

21
22 *Heritability calculations*

23
24 We calculated additive SNP-based heritability estimates for ischemic stroke, stroke
25 subtypes, and atrial fibrillation using restricted maximum likelihood implemented in
26 BOLT-REML (**Supplementary Methods**).¹⁶

27
28 *Genetic correlation between atrial fibrillation and ischemic stroke subtypes*

29
30 We used summary-level data from a prior Atrial Fibrillation Genetics (AFGen)
31 Consortium meta-analysis of atrial fibrillation⁵ to calculate a z-score for each SNP in
32 that GWAS. Additionally, we calculated a z-score for each SNP from our SiGN GWAS

1 of each stroke subtype and atrial fibrillation. As a null comparator, we downloaded
2 SNP z-scores from a GWAS of educational attainment¹⁷ available through LDHub
3 (<http://ldsc.broadinstitute.org/>, accessed 11-1-2017). We calculated Pearson's r
4 between z-scores from two traits to evaluate correlation (**Supplementary Methods**
5 and **Supplementary Figure 3**).

6

7 *Constructing an atrial fibrillation polygenic risk score*

8

9 To construct an atrial fibrillation polygenic risk score (PRS), we used SNPs from a
10 previously-derived atrial fibrillation PRS (**Supplementary Methods**).¹⁸ Briefly, the
11 PRS was derived from an atrial fibrillation GWAS of 17,931 cases and 115,142
12 controls.⁵ This PRS comprised 1,168 SNPs with $p < 1 \times 10^{-4}$ and LD pruned at an r^2
13 threshold of 0.5.¹⁸ Of these 1,168 SNPs, we identified 934 SNPs in the SiGN dataset
14 with imputation info > 0.8 and MAF $> 1\%$. We used these 934 SNPs to construct the
15 atrial fibrillation PRS in the SiGN dataset. Additional details on the PRS construction
16 can be found in the **Supplementary Methods**.

17

18 *Testing an atrial fibrillation polygenic risk score in ischemic stroke subtypes*

19

20 We tested for association between the atrial fibrillation PRS and stroke subtypes using
21 logistic regression (**Supplementary Methods**). We included sex and the top 10 PCs
22 as additional covariates. We optionally adjusted the association tests for age,
23 diabetes mellitus, cardiovascular disease, smoking status (current smoker, former
24 smoker, or never smoked), and hypertension.

25

26 We calculated the variance explained by the atrial fibrillation PRS in cardioembolic
27 stroke by constructing a model in BOLT-REML that consisted of: (1) a variance
28 component made up of SNPs for the GRM, and (2) a variance component made up of
29 SNPs from the PRS (**Supplementary Methods**).

30

31 *Data availability*

32

1 Code, supporting data, and downloadable supplemental tables are available here:
2 <https://github.com/UMCUGenetics/Afib-Stroke-Overlap>. The **Supplementary**
3 **Information** contains additional information regarding data access, methods, and
4 links to summary-level data.

5

6

1 Results

2

3 We began by testing our ability to rediscover known atrial fibrillation genetic
4 associations in the SiGN dataset, assembled to study the genetics of ischemic stroke.
5 We ran a genome-wide association study (GWAS) in SiGN using 3,190 cases with
6 atrial fibrillation or paroxysmal atrial fibrillation, as well as other diagnoses suggestive
7 of underlying atrial fibrillation^{19,20} (**Methods, Table 1** and **Supplementary Table 1**)
8 and 28,026 controls (**Supplementary Figure 1**). We found the top-associated SNPs
9 to be highly concordant with a prior GWAS of atrial fibrillation performed by the Atrial
10 Fibrillation Genetics (AFGen) Consortium (**Supplementary Table 2**). **Adjusting the**
11 **GWAS for age did not substantially change our findings ($r = 0.83$ between SNP effects**
12 **from the age-unadjusted and age-adjusted GWAS).**

13

14 Extending our analysis beyond these top associations, we next assessed whether
15 stroke patients with atrial fibrillation have a similar overall genetic predisposition to
16 the arrhythmia as seen in the independent AFGen GWAS. Additionally, we assessed
17 the overlap between genetic predisposition to atrial fibrillation and each stroke
18 subtype, allowing for the known phenotypic concordance between cardioembolic
19 stroke and atrial fibrillation (89.5% of cardioembolic stroke cases in SiGN also have
20 atrial fibrillation, **Supplementary Table 1**). We performed a series of GWAS in the
21 SiGN data for atrial fibrillation and each of the stroke subtypes using BOLT-LMM¹⁶
22 (**Methods**), and calculated the z-score (beta/standard error) of each SNP in each
23 phenotype. We then used summary-level results available from the prior
24 (independent) GWAS of atrial fibrillation⁵ (from AFGen) and calculated the z-score for
25 each SNP in that dataset.

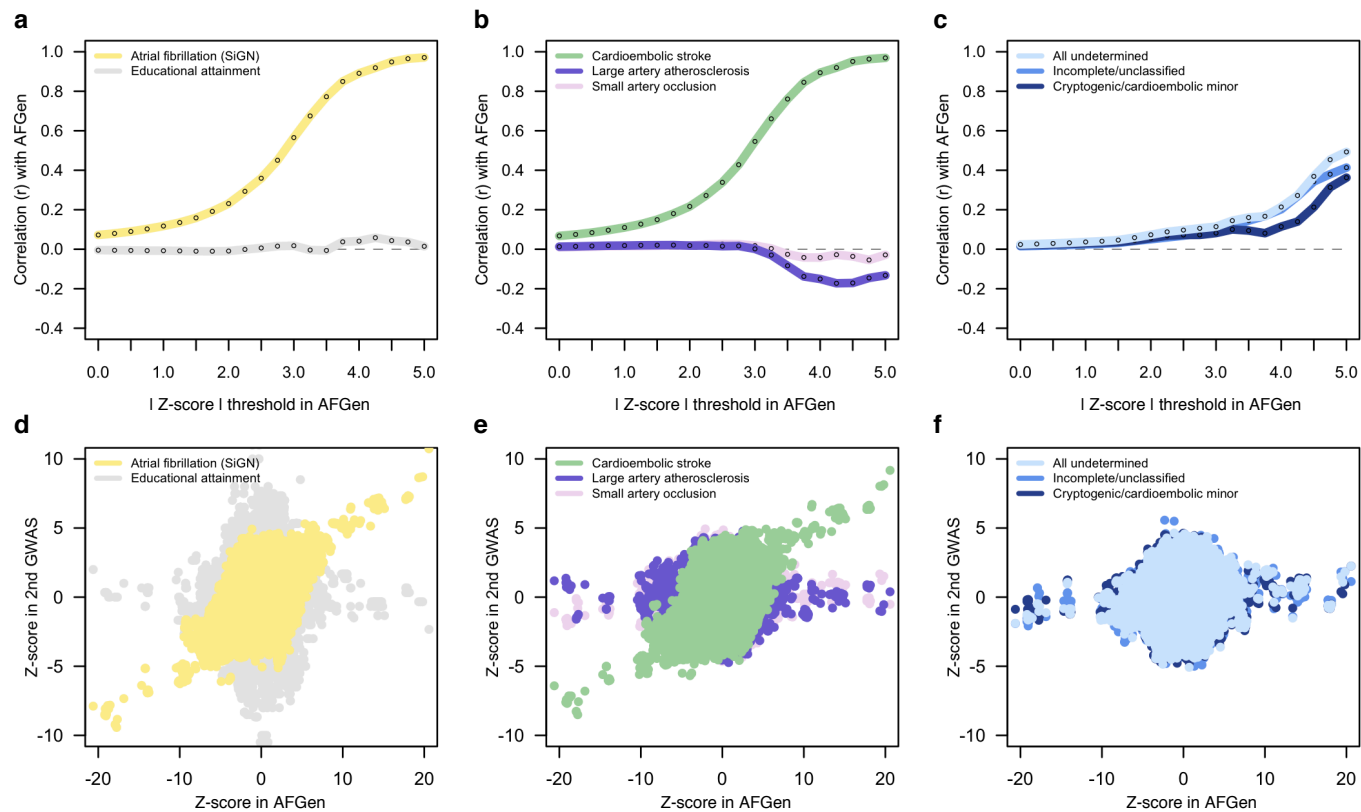
26

27 Measuring Pearson's correlation (r) between AFGen z-scores and z-scores from the
28 atrial fibrillation GWAS in SiGN, we found only a modest correlation ($r = 0.07$ across
29 ~ 7.8 M SNPs, **Figure 1**). However, when we iteratively subsetted the AFGen GWAS
30 results by the (absolute values of) z-scores of the SNPs, we found that correlation
31 with the atrial fibrillation GWAS in SiGN increased as the z-score threshold became
32 more stringent. For example, for those ~ 4.5 M SNPs with $|z| > 1$ in AFGen, correlation

1 with atrial fibrillation SNPs in SiGN was 0.12; for those ~ 1.9 M SNPs with $|z| > 3.5$ in
2 AFGen, correlation with the SiGN atrial fibrillation GWAS rose to 0.77 (**Figure 1** and
3 **Supplementary Table 3**). These correlations, calculated to include even modestly-
4 associated SNPs, indicate that atrial fibrillation in AFGen and atrial fibrillation in stroke
5 (SiGN) share a large proportion of genetic risk factors. Removing ± 2 Mb around the
6 *PITX2* and *ZFHX3* loci only modestly impacted the correlation between AFGen and
7 atrial fibrillation in SiGN ($r = 0.63$ for SNPs with $|z| > 3.5$; **Supplementary Figure**
8 **2** and **Supplementary Table 3**). Correlations between AFGen and cardioembolic
9 stroke in SiGN were unsurprisingly highly similar to that of the results with atrial
10 fibrillation in SiGN ($r = 0.77$ for AFGen SNPs with $|z| > 3.5$), likely due to the high
11 concordance between the atrial fibrillation and cardioembolic stroke phenotypes
12 (**Figure 1** and **Supplementary Figure 3**).

13

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



1 Continuing this analysis across the other stroke subtypes (large artery
2 atherosclerosis, small artery occlusion, and undetermined stroke; **Figure 1**), we
3 found near-zero correlation between AFGen and either large artery atherosclerosis or
4 small artery occlusion (**Figure 1**) indicating no genetic overlap between the
5 phenotypes. However, the correlation between atrial fibrillation and the
6 undetermined stroke subtypes (a highly heterogeneous subset of cases^{21,22} that
7 cannot be classified with standard subtyping systems^{13,15}) increased steadily as we
8 partitioned the AFGen data by z-score (all undetermined vs. AFGen $r = 0.04$ for
9 AFGen SNPs with $|z| > 1$ and $r = 0.16$ for AFGen SNPs with $|z| > 3.5$; **Figure 1** and
10 **Supplementary Table 3**), indicating that genome-wide, there is residual genetic
11 correlation between atrial fibrillation and the undetermined stroke categories, some
12 of which could represent causal atrial fibrillation stroke mechanisms in that subgroup.
13 As an additional null comparator, we performed correlations between the AFGen
14 results with z-scores derived from the latest GWAS of educational attainment¹⁷ and
15 found that correlation remained at approximately zero regardless of the z-score
16 threshold used (**Figure 1** and **Supplementary Table 3**).

17
18 To further understand the overlap between genetic risk factors for atrial fibrillation
19 and cardioembolic stroke and to evaluate the degree to which cardioembolic stroke
20 is comprised of risk factors beyond those for atrial fibrillation, we performed a
21 restricted maximum likelihood analysis implemented in BOLT-REML¹⁶ to estimate
22 SNP-based heritability of atrial fibrillation and cardioembolic stroke. Using
23 phenotypes derived from the CCS subtyping algorithm²³ (**Methods**), we estimated
24 heritability of atrial fibrillation and cardioembolic stroke at 20.0% and 19.5%,
25 respectively. These estimates are consistent with previous estimates in larger
26 samples (**Supplementary Figure 4**),^{24,25} and the similar heritabilities suggest that
27 cardioembolic stroke does not have a substantial heritable component beyond the
28 primary atrial fibrillation risk factor. For comparison, we calculated heritability in the
29 other stroke subtypes¹⁵ and found estimates to be similarly modest (range: 15.5% -
30 23.0%; **Supplementary Figures 4-6** and **Supplementary Table 4**).

31

1 Up to this point, our results indicated that atrial fibrillation in ischemic stroke is
2 genetically similar to that discovered in prior genetic studies of atrial fibrillation alone,
3 and that the bulk of the genetic risk for cardioembolic stroke appears attributable to
4 atrial fibrillation genetic risk factors. Next, we sought to explicitly test what proportion
5 of cardioembolic stroke risk could be explained by atrial fibrillation loci, independent
6 of known clinical risk factors for atrial fibrillation. First, we identified SNPs from an
7 atrial fibrillation polygenic risk score (PRS) independently derived from the AFGen
8 GWAS⁵ (**Methods**). Of the 1,168 SNPs used to generate this pre-established PRS,
9 we identified 934 in the SiGN dataset with imputation quality > 0.8 and minor allele
10 frequency >1%. We computed the PRS per individual (**Methods**), weighting the
11 imputed dosage of each risk allele by the effect of the SNP (i.e., the beta coefficient)
12 as reported in AFGen⁵.

13

14 We tested the association of the atrial fibrillation PRS with cardioembolic stroke, using
15 a logistic regression and adjusting for the top ten principal components and sex
16 (**Methods**). As expected from our earlier results, we found the PRS to be strongly
17 associated with cardioembolic stroke (odds ratio (OR) per 1 standard deviation (sd)
18 of the PRS = 1.93 [95% confidence interval (CI): 1.34 - 1.44], $p = 1.01 \times 10^{-65}$;
19 **Figure 2** and **Supplementary Table 5**), confirming the high genetic concordance
20 of these phenotypes across SNPs which, individually, confer only a modest average
21 association with atrial fibrillation. Next, we adjusted the association model for clinical
22 covariates associated with atrial fibrillation including age, diabetes mellitus,
23 cardiovascular disease, smoking, and hypertension.²⁶ Using a (smaller) set of cases
24 and controls with complete clinical risk factor information, we found that inclusion of
25 these clinical risk factors in the model only modestly reduced the PRS signal in
26 cardioembolic stroke (OR per 1 sd = 1.40 [95% CI: 1.34 - 1.47], $p = 1.45 \times 10^{-48}$;
27 **Supplementary Tables 5-7**). These results indicate a strong relationship between
28 atrial fibrillation genetic risk factors and cardioembolic stroke risk, independent of the
29 clinical factors that associate with atrial fibrillation. Expanding the set of SNPs used
30 to construct the PRS to the original 934 SNPs $\pm 25\text{kb}$, $\pm 50\text{kb}$, and $\pm 100\text{kb}$ (**Methods**)
31 revealed a persistently strong, though somewhat attenuated, association between
32 the PRS and cardioembolic stroke (PRS including SNPs within 100kb, $p = 4.47 \times 10^{-$

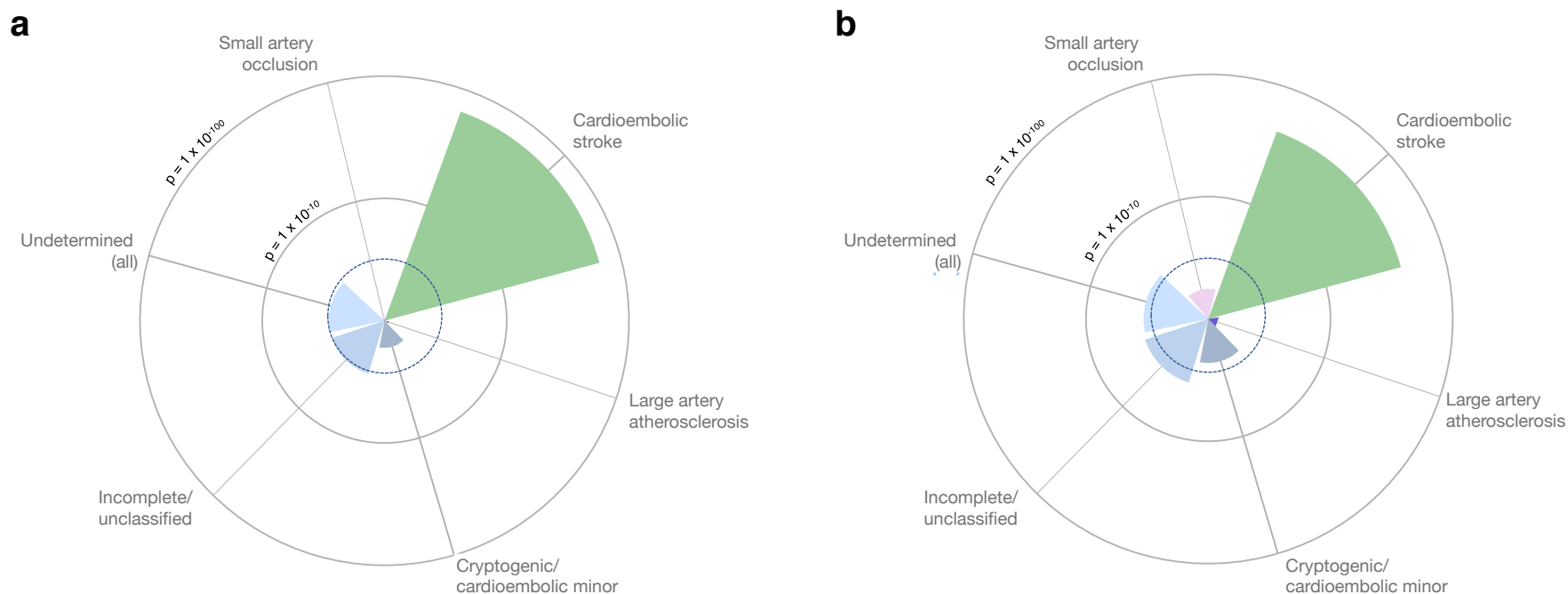
1 ⁴⁴, **Supplementary Table 6**). None of the other stroke subtypes were significantly
2 associated with the atrial fibrillation PRS (all $p > 0.013$, **Figure 2** and
3 **Supplementary Figure 6**).

4
5 Because atrial fibrillation status was missing for most controls in the SiGN dataset,
6 we performed sensitivity analyses using only the 3,861 controls confirmed as having
7 no atrial fibrillation. While reducing the set of controls to this refined group did not
8 substantially change results for the primary stroke subtypes, we found the atrial
9 fibrillation PRS was modestly associated ($p < 5 \times 10^{-3}$, after adjusting for five
10 subtypes and two control groups) with the overall undetermined subtype (OR per 1
11 sd = 1.07 [95% CI: 1.02 - 1.13], $p = 4.15 \times 10^{-3}$) (**Figure 2** and **Supplementary**
12 **Table 5**). Further examination of the two mutually exclusive subgroups of the
13 undetermined group revealed that the PRS associated significantly with the
14 incomplete/unclassified categorization (OR per 1 sd = 1.09 [95% CI: 1.03 - 1.16], p
15 = 3.17×10^{-3}) (**Figure 2**) but not with cryptogenic/cardioembolic minor (OR per 1 sd
16 = 1.06 [95% CI: 1.00 - 1.13], $p = 5.10 \times 10^{-2}$). Correcting for clinical covariates only
17 modestly changed the signal in the incomplete/unclassified phenotype ($p = 9.7 \times 10^{-$
18 ³, **Figure 2**), supporting the robustness of the observed association, independent of
19 clinical risk factors.

20
21 Lastly, we created a model in BOLT-LMM, fitting two genetic variance components:
22 one component including SNPs for the genetic relationship matrix, and the second
23 component including the original PRS SNPs from the atrial fibrillation PRS (including
24 ± 100 kb around these SNPs, to include a sufficient number of markers to estimate
25 variance explained). We found that the SNPs from the atrial fibrillation PRS explained
26 4.1% of the total (20.0%) heritability in atrial fibrillation. In evaluating variance
27 explained in cardioembolic stroke, we found a nearly identical result: the component
28 representing the atrial fibrillation risk score explained 4.5% (s.e. = 1.00%) of the
29 total 19.5% genetic heritability in cardioembolic stroke. Thus, atrial fibrillation
30 genetic risk accounts for 23.1%, or approximately one-fifth, of the total heritability
31 of cardioembolic stroke.

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

8



9

10

11

1 Discussion

2
3 Our results suggest that individuals with cardioembolic strokes have an enrichment
4 for atrial fibrillation genetic risk, despite the fact that cardioembolic stroke often
5 affects older adults with multiple clinical comorbidities²⁷ that could increase risk for
6 atrial fibrillation due to non-genetic factors. The fact that cardioembolic stroke and
7 atrial fibrillation share a highly-similar genetic architecture extends our
8 understanding of the morbid consequences of heritable forms of the arrhythmia.
9 Furthermore, the observation that atrial fibrillation genetic risk was only associated
10 with cardioembolic stroke, and (consistently) lacked association in large artery
11 atherosclerosis or small artery occlusion,²⁸ raises the possibility that atrial fibrillation
12 genetic risk may be informative in the management of ischemic stroke survivors in
13 whom the mechanism may be unclear.

14
15 The use of polygenic risk scores for complex traits has proved an efficient means of
16 understanding how genetic predisposition to diseases can overlap. Given the
17 onslaught of genotyping data available for common diseases, PRS's can now be used
18 to stratify patients by risk (e.g., in breast cancer^{29,30}) or predict outcome (e.g., in
19 neuropsychiatric disease²⁹). More recently, PRS's have been used to identify
20 individuals in the general population with a four-fold risk for coronary disease,³¹
21 proposed for inclusion in clinical workups of individuals with early-onset coronary
22 artery disease,³² and used to identify patients for whom lifestyle changes or statin
23 intervention would be beneficial.^{33,34} While previous work has also shown an
24 association between an atrial fibrillation PRS and cardioembolic stroke,²⁸ we have
25 extended this work to formally quantify the extent to which an atrial fibrillation PRS
26 captures genetic risk for cardioembolic stroke. These findings lay the groundwork for
27 future work that can potentially leverage this overlap to develop atrial fibrillation
28 PRS's that could be used to predict individuals at highest risk of cardioembolic stroke
29 (to improve diagnostic resource allocation) or help distinguish between clinical
30 subtypes of stroke.

31

1 Though our analysis was aimed at understanding the genetic overlap between
2 cardioembolic stroke and atrial fibrillation, we additionally observed genetic
3 correlation between atrial fibrillation and undetermined stroke, **a finding not observed**
4 **in a previous investigation of atrial fibrillation PRS in ischemic stroke subtypes, albeit**
5 **in a smaller sample.**²⁸ Perhaps contrary to expectation, we specifically found the atrial
6 fibrillation polygenic risk score to be more strongly associated with the subset of
7 etiology-undetermined strokes with an incomplete clinical evaluation, as opposed to
8 those with cryptogenic stroke of a presumed, but not demonstrated, embolic source.
9 These associations could be due to physician biases in diagnostic workups, rather
10 than supporting a low prevalence of occult atrial fibrillation in presumed embolic
11 strokes of undetermined source. Identifying stroke patients with atrial fibrillation is
12 an important clinical challenge, as occult atrial fibrillation is well-known to cause
13 strokes,^{35,36} and since such patients are at high risk for recurrent stroke, which is
14 preventable with anticoagulation.^{37,38} Together, our findings indicate that atrial
15 fibrillation genetic risk may augment clinical algorithms to determine stroke etiology,
16 but will require further study.

17

18 **The work presented here benefits from a number of improvements, including**
19 **increased sample size; analysis of samples from a multicenter consortium, potentially**
20 **enhancing the generalizability of the findings; and use of the CCS subtyping system,**
21 **which provides more nuanced phenotyping, particularly in the cryptogenic subtype.**
22 Nevertheless, some limitations remain. Stroke is a heterogeneous condition that
23 occurs later in life and has high lifetime prevalence (>15%³⁹), features that can
24 reduce statistical power. Further, sample sizes have lagged behind other GWAS
25 efforts, a challenge further compounded by subtyping (nearly one-third of all cases
26 are categorized as undetermined²³). Reduced sample sizes impact power for
27 discovery and make other analytic approaches – such as standard approaches for
28 measuring trait correlation¹⁶ – unfeasible. Also, our sample is primarily comprised of
29 European-ancestry samples, and work in non-Europeans, **particularly in African-**
30 **ancestry samples where risk of stroke is double that of European samples, is crucial.**
31 Finally, the current analysis does not analyze rare variation, which also likely
32 contributes to disease susceptibility.⁵

1
2
3
4
5
6
7
8
9
10
11
12
13

We have shown that the cumulative genetic risk for atrial fibrillation in individuals with a stroke is similar to that reported in a larger population-based cohort.²⁵ Genome-wide variation related to atrial fibrillation is substantially associated with cardioembolic stroke risk. Moreover, atrial fibrillation genetic risk was specific for cardioembolic stroke, and was not associated with the other primary stroke subtypes. The observation that atrial fibrillation genetic risk associated with strokes of undetermined cause supports the notion that undetected atrial fibrillation underlies a proportion of stroke risk in these individuals. Further work will need to incorporate emerging discoveries of rare genetic variants in atrial fibrillation, and explore the potential for genetic risk tools, including PRS's performed via clinical-grade genotyping, to assist in the diagnostic workup of individuals with ischemic stroke.

1 Tables

2

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

12

	Phenotype	Total	Ischemic stroke subtype				
			Primary subtypes			Undetermined subtypes	
			Cardioembolic	Large artery atherosclerosis	Small artery occlusion	Incomplete/unclassified	Cryptogenic/cardioembolic minor
	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 atrial fibrillation cases	3,190	2,684	123	61	298	0
	No atrial fibrillation	--	316	2,262	2,201	1,982	2,294

13

14

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. Pulit is supported by Veni Fellowship 016.186.071 (ZonMW) from the Dutch Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO).

Dr. Anderson is supported in part by K23NS086873, R01NS103924, an American Heart Association Strategically Focused Research Network in Atrial Fibrillation Award, and a Massachusetts General Hospital Center for Genomic Medicine Catalyst Award.

Dr. Lubitz is supported by NIH grants K23HL114724, an American Heart Association Strategically Focused Research Network in Atrial Fibrillation Award, 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, and an American Heart Association Strategically Focused Research Network in Atrial Fibrillation Award. 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 Award 13EIA14220013; and the Fondation Leducq 14CVD01.

1 References

- 2
- 3 1. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial
4 fibrillation: a Global Burden of Disease 2010 Study. *Circulation*. 2014;129:837–
5 847.
- 6 2. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor
7 for stroke: the Framingham Study. *Stroke*. 1991;22:983–988.
- 8 3. World Health Organization. WHO | The top 10 causes of death [online]. World
9 Health Organization; 2014. Accessed at:
10 <http://www.who.int/mediacentre/factsheets/fs310/en/>.
- 11 4. Feigin VL, Lawes CM, Bennett D a., Barker-Collo SL, Parag V. Worldwide stroke
12 incidence and early case fatality reported in 56 population-based studies: a
13 systematic review. *Lancet Neurol*. Elsevier Ltd; 2009;8:355–369.
- 14 5. Christophersen IE, Rienstra M, Roselli C, et al. Large-scale analyses of common
15 and rare variants identify 12 new loci associated with atrial fibrillation. *Nat*
16 *Genet*. 2017;57:289.
- 17 6. Bevan S, Traylor M, Adib-Samii P, et al. Genetic heritability of ischemic stroke
18 and the contribution of previously reported candidate gene and genome-wide
19 associations. *Stroke*. 2012;43:3161–3167.
- 20 7. Lubitz SA, Ozcan C, Magnani JW, Käåb S, Benjamin EJ, Ellinor PT. Genetics of
21 atrial fibrillation: implications for future research directions and personalized
22 medicine. *Circ Arrhythm Electrophysiol*. 2010;3:291–299.
- 23 8. Gudbjartsson DF, Arnar DO, Helgadottir A, et al. Variants conferring risk of
24 atrial fibrillation on chromosome 4q25. *Nature*. 2007;448:353–357.
- 25 9. Gretarsdottir S, Thorleifsson G, Manolescu A, et al. Risk variants for atrial
26 fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol*.
27 2008;64:402–409.
- 28 10. Gudbjartsson DF, Holm H, Gretarsdottir S, et al. A sequence variant in ZFX3
29 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet*.
30 2009;41:876–878.
- 31 11. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology,
32 Function, and Translation. *Am J Hum Genet*. 2017;101:5–22.
- 33 12. Pulit SL, McArdle PF, Wong Q, Malik R, Gwinn K, Achterberg S, Algra A,
34 Amouyel P, Anderson CD, Arnett DK, Arsava EM, Attia J, Ay H, Bartz TM, Battey
35 T, Benavente OR, Bevan S, Biffi A, Bis JC, Blanton SH, Boncoraglio GB, Brown
36 Jr RD, Burgess AI, Carrera, Smith SNC, Chasman DI, Chauhan G, Chen WM,
37 Cheng YC, Chong M, Cloonan LK, Cole JW, Cotlarciuc I, Cruchaga C, Cuadrado-

- 1 Godia E, Dave T, Dawson J, Debette S, Delavaran H, Dell CA, Dichgans M,
2 Doheny KF, Dong C, Duggan DJ, Engström G, Evans MK, Pallejà XE, Faul
3 JD, Fernández-Cadenas I, Fornage M, Frossard PM, Furie K, Gamble DM, Gieger
4 C, Giese AK, Giralt-Steinhauer E, González HM, Goris A, Gretarsdottir S, Grewal
5 RP, Grittner R, Gustafsson S, Han B, Hankey GJ, Heitsch L, Higgins P, Hochberg
6 MC, Holliday E, Hopewell EC, Horenstein RB, Howard G, Ikram MA, Ilinca A,
7 Ingelsson E, Irvin MR, Jackson RD, Jern C, Conde JJ, Johnson JA, Jood K, Kahn
8 MS, Kaplan R, Kappelle LJ, Kardia SLR, Keene KL, Kissela BM, O Kleindorfer D,
9 Koblar S, Labovitz D, Launer LJ, Laurie CC, Laurie CA, Lee CH, Lee JM, Lehm M,
10 Lemmens R, Levi C, Leys D, Lindgren A, Longstreth Jr WT, Maguire J,
11 Manichaikul Ani, Markus HS, McClure LA, McDonough CW, Meisinger C,
12 Melander O, Meschia JF, Mola-Caminal M, Montaner J, Mosley TH, Müller-
13 Nurasyid M, Nalls MA, O'Connell JR, O'Donnell M, Ois A, Papanicolaou GJ, Paré
14 G, Peddareddygar LR, Pedersén A, Pera J, Peters A, Poole D, Psaty BM,
15 Rabionet R, Raffeld MR, Rannikmäe K, Rasheed A, Redfors P, Reiner AP,
16 Rexrode K, Ribasés M, Rich SS, Robberecht W, Rodriguez-Campello A, Rolfs A,
17 Roquer J, Rose LM, Rosenbaum D, Rost NS, Rothwell PM, Rundek T, Ryan KA,
18 Sacco RL, Sale MM, Saleheen D, Salomaa V, Sánchez-Mora C, Schmidt CO,
19 Schmidt H, Schmidt R, Schürks M, Scott R, Segal HC, Seiler S, Seshadri S,
20 Sharma P, Shuldiner AR, Silver B, Slowik A, Smith JA, Söderholm M, Soriano C,
21 Sparks MJ, Stanne T, Stefansson K, Stine OC, Strauch K, Sturm J, Sudlow CLM,
22 Tajuddin SM, Talbert RL, Tatlisumak T, Thijs V, Thorleifsson G, Thorsteindottir
23 U, Tiedt S, Traylor M, Trompet S, Valant V, Waldenberger M, Walters M, Wang
24 L, Wassertheil-Smoller S, Weir DR, Wiggins KL, Williams SR, Wloch-Kopec D,
25 Woo D, Woodfield D, Wu O, Xu H, Zonderman AB, Australian Stroke Genetics
26 Consortium, Cervical Artery Dissection and Ischemic Stroke Patients (CADISP)
27 study, Cohorts of Heart and Aging Research in Genomic Epidemiology
28 (CHARGE) consortium, Consortium of Minority Population genome-wide
29 Association Studies of Stroke (COMPASS), METASTROKE consortium, Wellcome
30 Trust Case-Control Consortium, Worrall BB, de Bakker PIW, Kittner SJ, Mitchell
31 BD, Rosand J. Loci associated with ischaemic stroke and its subtypes (SiGN): a
32 genome-wide association study. *Lancet Neurology*. 2015;15:174–184.
- 33 13. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic
34 classification of ischemic stroke: The causative classification of stroke system.
35 *Stroke*. 2007;38:2979–2984.
- 36 14. Arsava EM, Ballabio E, Benner T, et al. The causative classification of stroke
37 system: An international reliability and optimization study. *Neurology*.
38 2010;75:1277–1284.
- 39 15. Adams HP Jr, Birgitte H. Bendixen P, Jaap Kappelle L, et al. Classification of
40 Subtype of Acute Ischemic Stroke: Definitions for Use in a Multicenter Clinical
41 Trial. *Stroke*. 1993;24:35–41.
- 42 16. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model
43 analysis increases association power in large cohorts. *Nat Genet*. 2015;47:284–
44 290.

- 1 17. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study
2 identifies 74 loci associated with educational attainment. *Nature*. Nature
3 Publishing Group, a division of Macmillan Publishers Limited. All Rights
4 Reserved.; 2016;533:539.
- 5 18. Weng L-C, Preis SR, Hulme OL, et al. Genetic Predisposition, Clinical Risk Factor
6 Burden, and Lifetime Risk of Atrial Fibrillation. *Circulation* [online serial]. Epub
7 2017 Nov 12. Accessed at:
8 <http://dx.doi.org/10.1161/CIRCULATIONAHA.117.031431>.
- 9 19. Franz MR, Karasik PL, Li C, Moubarak J, Chavez M. Electrical remodeling of the
10 human atrium: similar effects in patients with chronic atrial fibrillation and atrial
11 flutter. *J Am Coll Cardiol*. 1997;30:1785–1792.
- 12 20. Lamas GA, Lee KL, Sweeney MO, et al. Ventricular pacing or dual-chamber
13 pacing for sinus-node dysfunction. *N Engl J Med*. 2002;346:1854–1862.
- 14 21. Petty GW, Brown RD, Whisnant JP, Sicks JD, Michael O’Fallon W, Wiebers DO.
15 Ischemic Stroke Subtypes. *Stroke*. American Heart Association, Inc.;
16 1999;30:2513–2516.
- 17 22. Kolominsky-Rabas PL, Weber M, Gefeller O, Neundoerfer B, Heuschmann PU.
18 Epidemiology of Ischemic Stroke Subtypes According to TOAST Criteria. *Stroke*.
19 American Heart Association, Inc.; 2001;32:2735–2740.
- 20 23. NINDS Stroke Genetics Network (SiGN), International Stroke Genetics
21 Consortium. Loci associated with ischaemic stroke and its subtypes (SiGN): a
22 genome-wide association study. *Lancet Neurol*. Elsevier Ltd; 2015;15:4–7.
- 23 24. Go AS, Hylek EM, Phillips KA, et al. Prevalence of diagnosed atrial fibrillation in
24 adults: national implications for rhythm management and stroke prevention:
25 the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA*.
26 2001;285:2370–2375.
- 27 25. Weng L-C, Choi SH, Klarin D, et al. Heritability of Atrial Fibrillation. *Circ*
28 *Cardiovasc Genet*. American Heart Association, Inc.; 2017;10:e001838.
- 29 26. Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of
30 atrial fibrillation in a racially and geographically diverse population: the
31 CHARGE-AF consortium. *J Am Heart Assoc*. 2013;2:e000102.
- 32 27. Henninger N, Goddeau RP Jr, Karmarkar A, Helenius J, McManus DD. Atrial
33 Fibrillation Is Associated With a Worse 90-Day Outcome Than Other
34 Cardioembolic Stroke Subtypes. *Stroke*. 2016;47:1486–1492.
- 35 28. Lubitz SA, Parsons OE, Anderson CD, et al. Atrial Fibrillation Genetic Risk and
36 Ischemic Stroke Mechanisms. *Stroke*. 2017;48:1451–1456.
- 37 29. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine.
38 *Genome Med*. 2017;9:96.

- 1 30. Mavaddat N, Pharoah PDP, Michailidou K, et al. Prediction of breast cancer risk
2 based on profiling with common genetic variants. *J Natl Cancer Inst* [online
3 serial]. 2015;107. Accessed at: <http://dx.doi.org/10.1093/jnci/djv036>.
- 4 31. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic score to
5 identify a monogenic risk-equivalent for coronary disease [online]. *bioRxiv*
6 2017. Accessed at: <http://dx.doi.org/10.1101/218388>.
- 7 32. Thériault S, Lali R, Chong M, Velianou JL, Natarajan MK, Paré G. Polygenic
8 Contribution in Individuals With Early-Onset Coronary Artery DiseaseClinical
9 Perspective. *Circ Genom Precis Med*. Lippincott Williams & Wilkins;
10 2018;11:e001849.
- 11 33. Khera AV, Emdin CA, Drake I, et al. Genetic Risk, Adherence to a Healthy
12 Lifestyle, and Coronary Disease. *N Engl J Med*. 2016;375:2349–2358.
- 13 34. Mega JL, Stitzel NO, Smith JG, et al. Genetic risk, coronary heart disease
14 events, and the clinical benefit of statin therapy: an analysis of primary and
15 secondary prevention trials. *Lancet*. 2015;385:2264–2271.
- 16 35. Gladstone DJ, Spring M, Dorian P, et al. Atrial fibrillation in patients with
17 cryptogenic stroke. *N Engl J Med*. 2014;370:2467–2477.
- 18 36. Sanna T, Diener H-C, Passman RS, et al. Cryptogenic stroke and underlying
19 atrial fibrillation. *N Engl J Med*. 2014;370:2478–2486.
- 20 37. ACTIVE Investigators, Connolly SJ, Pogue J, et al. Effect of clopidogrel added to
21 aspirin in patients with atrial fibrillation. *N Engl J Med*. 2009;360:2066–2078.
- 22 38. Diener H-C, Eikelboom J, Connolly SJ, et al. Apixaban versus aspirin in patients
23 with atrial fibrillation and previous stroke or transient ischaemic attack: a
24 predefined subgroup analysis from AVERROES, a randomised trial. *Lancet*
25 *Neurol*. 2012;11:225–231.
- 26 39. Seshadri S, Beiser A, Kelly-Hayes M, et al. The Lifetime Risk of Stroke:
27 Estimates From the Framingham Study. *Stroke*. 2006;37:345–350.

Supplementary information for

Shared Genetic Contributions to Atrial Fibrillation and Ischemic Stroke Risk

Sara L. Pulit, Lu-Chen Weng, Patrick F McArdle, Ludovic Trinquart, Seung Hoan Choi, Braxton D. Mitchell, Jonathan Rosand, Paul I W de Bakker, Emelia J Benjamin, Patrick T Ellinor, Steven J Kittner, Steven A Lubitz*, Christopher D Anderson*, on behalf of the Atrial Fibrillation Genetics Consortium and the International Stroke Genetics Consortium.

Code and data release

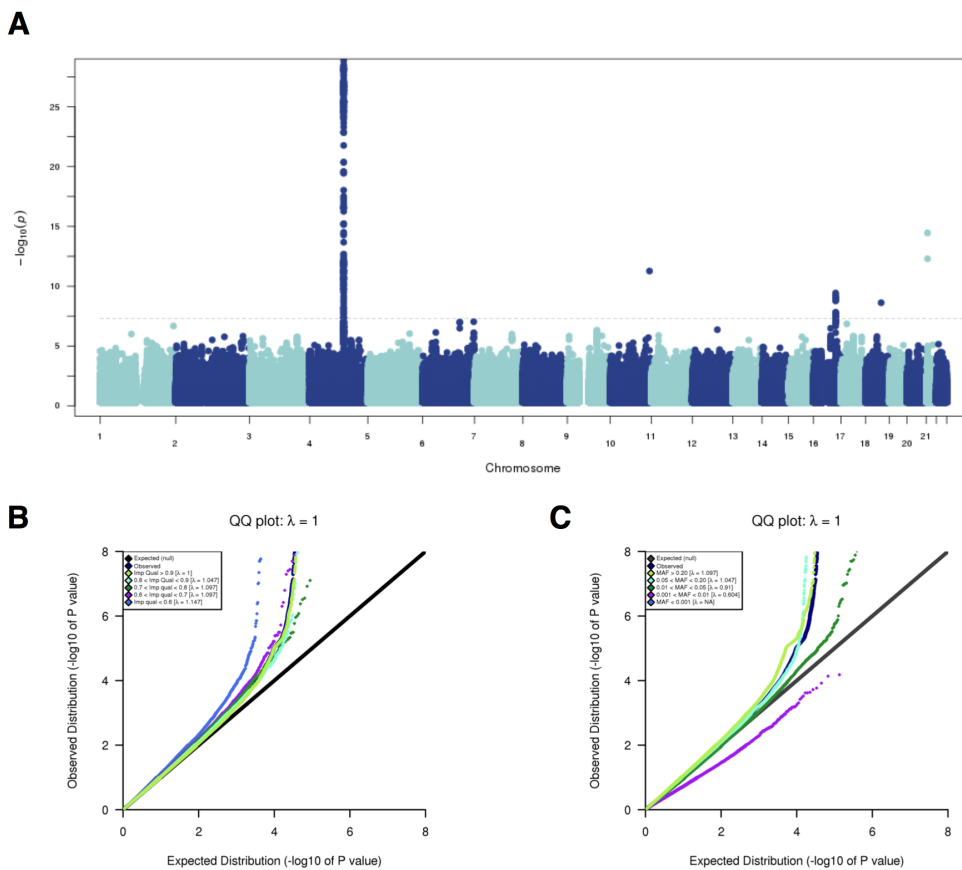
For access to information related to this project, including code, sample identifiers, SNP identifiers, links to summary-level data, and SNP weights used in the construction of the polygenic risk score, please see this GitHub repository: <https://github.com/UMCUGenetics/Afib-Stroke-Overlap>.

Table of Contents

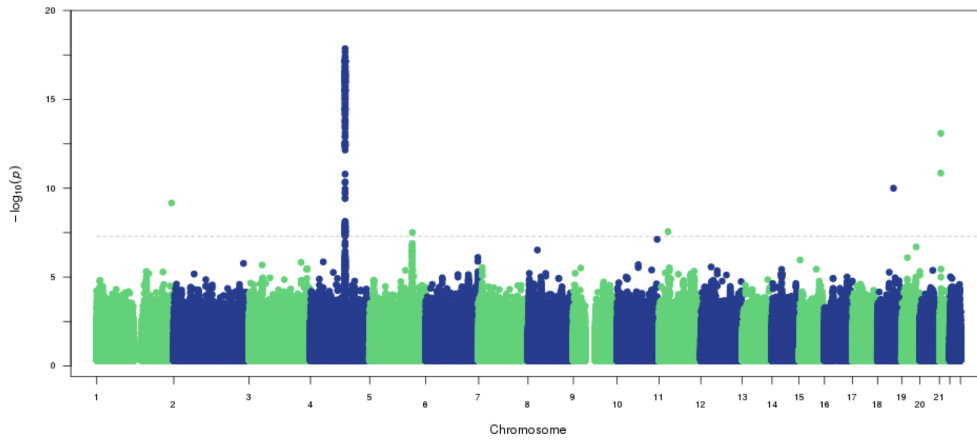
Supplementary Figures	Pages 3-11
Supplementary Tables	Pages 12-25
Supplementary Methods	Pages 26-33
Supplementary Results	Page 34
Appendix I: Atrial Fibrillation Genetics (AFGen) Consortium members	Pages 35-38
Appendix II: Stroke Genetics Network (SiGN) Consortium members	Pages 39-42
References	Pages 43-45

Supplementary Figures

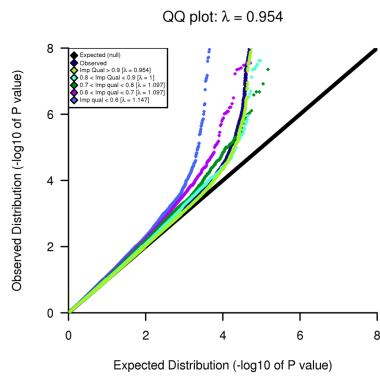
Supplementary Figure 1 | Genome-wide association study (GWAS) of atrial fibrillation in SiGN. (A) We performed a GWAS of 3,190 cases with atrial fibrillation, or paroxysmal atrial fibrillation, as well as other diagnoses suggestive of underlying atrial fibrillation, including left atrial thrombus, sick sinus syndrome, and atrial flutter. We additionally included 28,026 referents. We used a linear mixed model and adjusted the model for principal components and sex. The majority of atrial fibrillation risk loci identified through previous GWAS efforts were identified here at nominal significance or better (see **Supplementary Table 2**). The Manhattan plot only shows QC-passing SNPs with minor allele frequency > 1% and imputation quality score > 0.8. (B) Quantile-quantile (QQ) plot indicating SNPs stratified by minor allele frequency and the corresponding genomic inflation factor (λ) for each stratum. (C) QQ plot showing SNPs stratified by imputation quality and the corresponding λ for each stratum. Figures D-F are identical to those of A-C, but for the analysis performed in atrial fibrillation cases only ($N = 1,751$). We performed this is an internal sensitivity analysis only, to ensure that more broadly defining the atrial fibrillation phenotype was not introducing additional phenotypic noise.



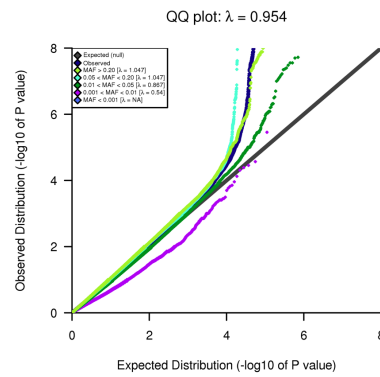
D



E

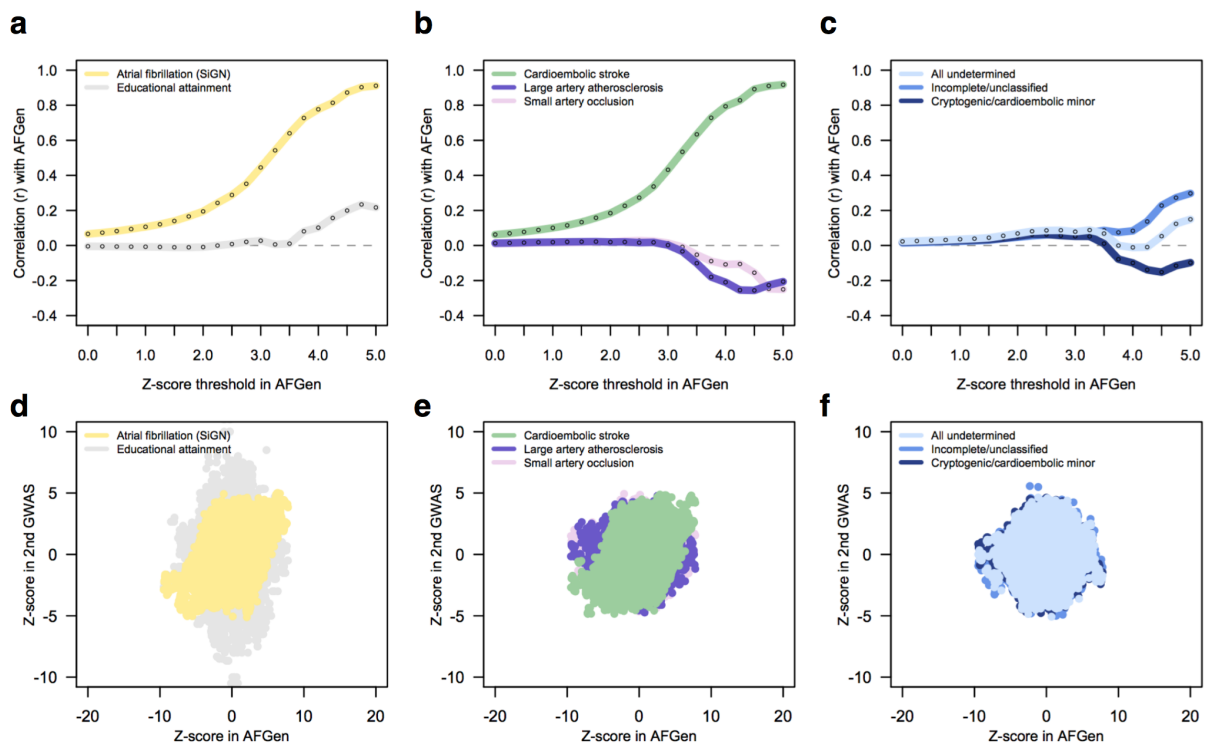


F



Supplementary Figure 2 | Genetic correlations between atrial fibrillation and ischemic stroke subtypes. To estimate genetic correlation between atrial fibrillation and ischemic stroke subtypes, we calculated Pearson's r between SNP z-scores in the AFGen GWAS of atrial fibrillation and in GWAS of ischemic stroke subtypes and atrial fibrillation performed here in the SiGN data. Here, we present data identical to that shown in Figure 2 of the main manuscript, but removing $\pm 2\text{Mb}$ around the two most significant loci discovered in atrial fibrillation and cardioembolic stroke: the region around *PITX2* (chromosome 4) and the region around *ZFHX3* (chromosome 16). (a) Genome wide, atrial fibrillation in AFGen and in SiGN correlate with increasing strength as the z-score in AFGen increases. Educational attainment is included here as a null comparator. (b) Genetic signal in cardioembolic stroke also correlates strongly with atrial fibrillation genetic signal in AFGen, but we do not observe correlation between atrial fibrillation and the other primary stroke subtypes. (c) Removing the *PITX2* and *ZFHX3* regions leaves only somewhat modest correlation between the incomplete/unclassified undetermined subtype and atrial fibrillation. Panels (d-f) show underlying data.

Correlations restricted to those SNPs used in the polygenic risk score for atrial fibrillation were: AFGen vs atrial fibrillation in SiGN, $r = 0.78$; AFGen vs. cardioembolic stroke in SiGN, $r = 0.75$.

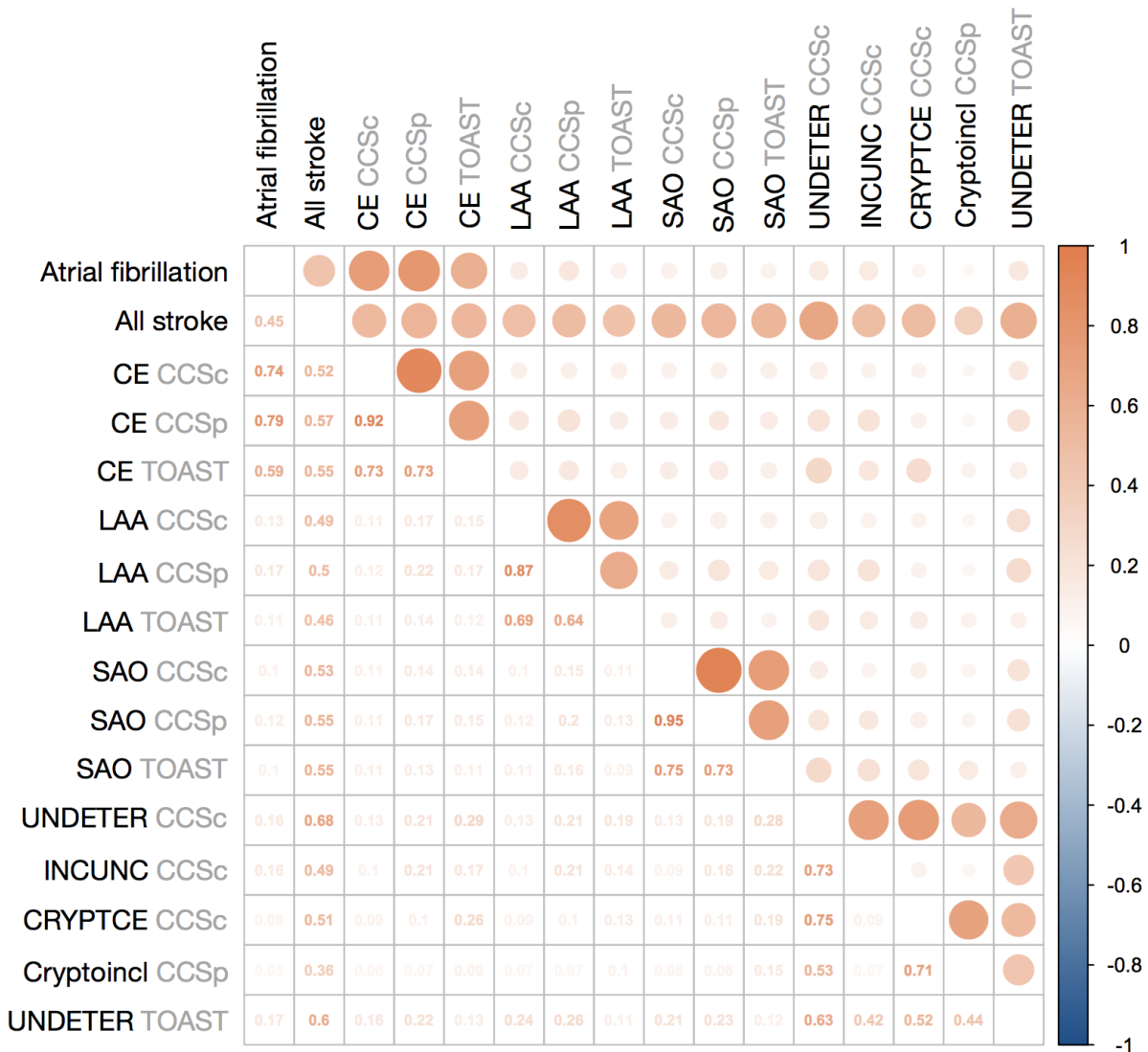


Supplementary Figure 3 | Genetic correlation and phenotypic correlation of atrial fibrillation and stroke subtypes in SiGN.

(a) Using genome-wide SNP effects extracted from GWAS of atrial fibrillation, all stroke, and stroke subtypes, we calculated the Pearson's correlation (r) between each pair of available phenotypes (blue indicates strong negative correlation; orange indicates strong positive correlation). Here, we show all correlations. Correlations are indicated by circle size in the upper half of the square, and the exact correlation values are shown in the lower half of the square.

CE, cardioembolic stroke; LAA, large artery atherosclerosis; SAO, small artery occlusion; UNDETER, undetermined; INCUNC, incomplete/unclassified; CRYPTCE, cryptogenic and CE minor; Cryptoincl, cryptogenic; CCSc, CCS Causative subtyping system; CCSp, CCS Phenotypic subtyping system; TOAST, TOAST subtyping system.

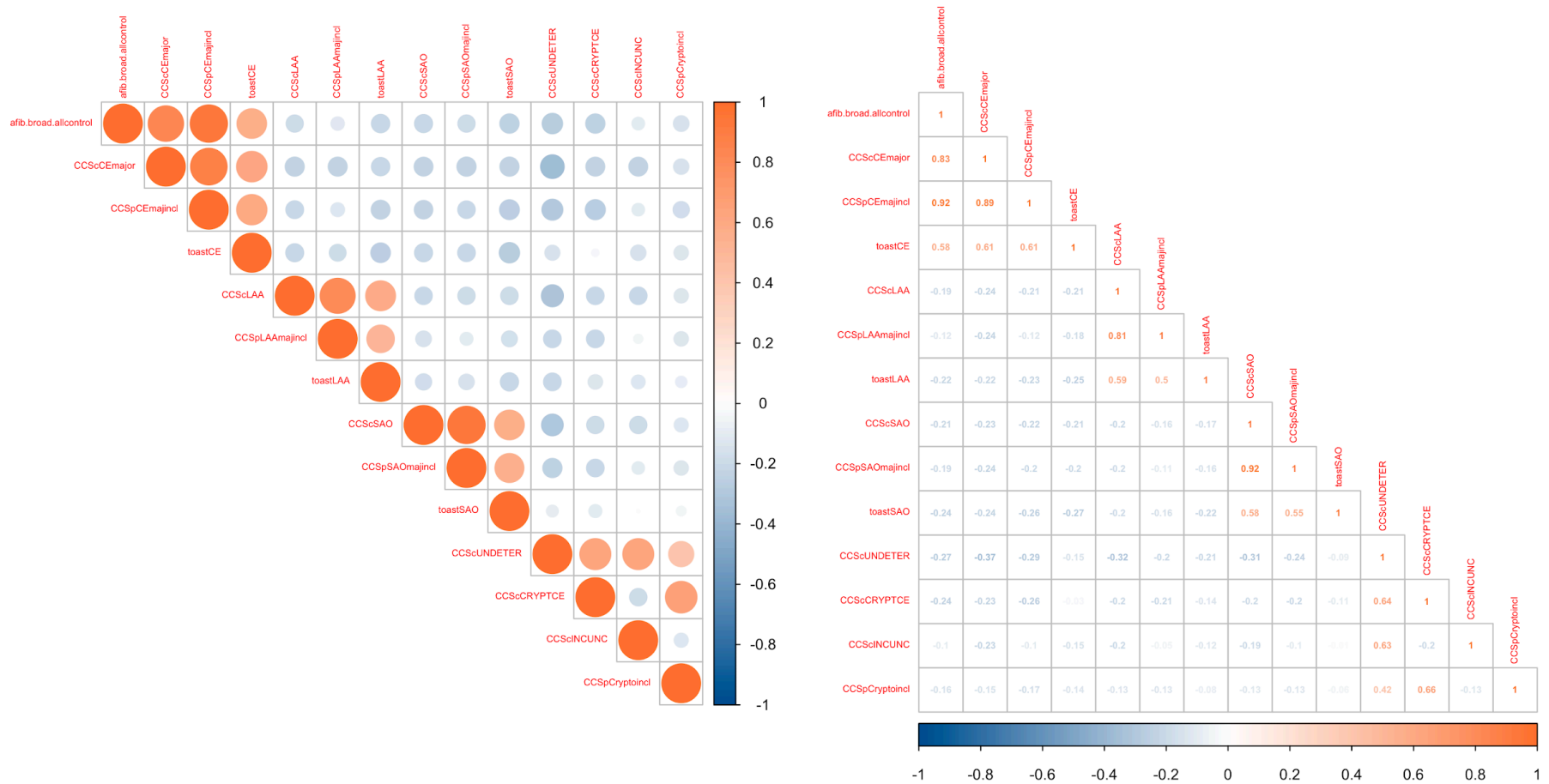
a.



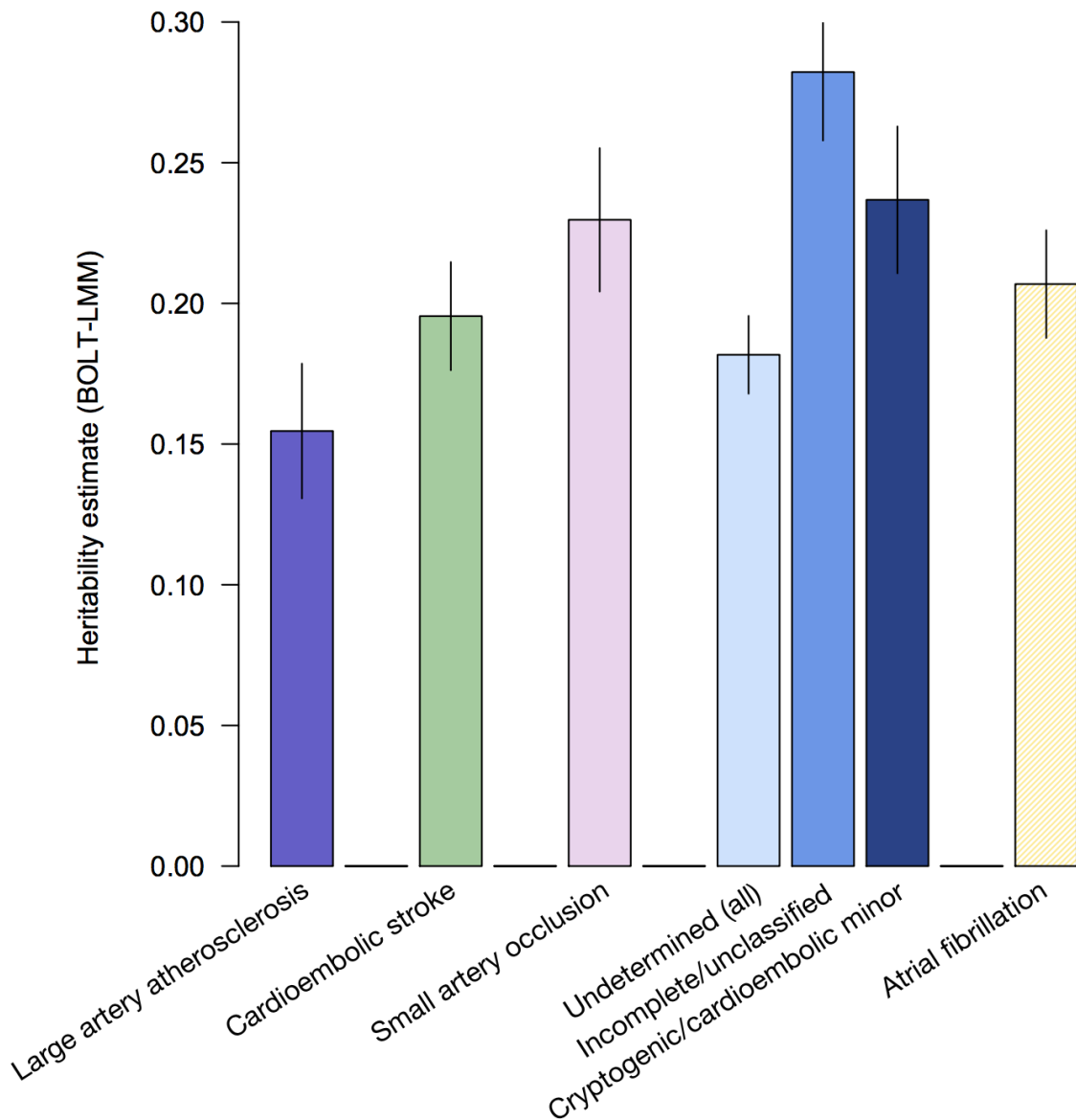
b. Same correlation calculations as in (a), but this time using the phenotypic data only (and looking in cases only, as all controls have the same phenotype). Note that the atrial fibrillation phenotypes and cardioembolic stroke phenotypes are highly correlated in the SiGN data ($r = 0.83$ between atrial fibrillation and cardioembolic stroke as determined by the CCS Causative subtype system).

Supplementary Information: Shared Genetic Contributions to Atrial Fibrillation and Ischemic Stroke Risk

CE, cardioembolic stroke; LAA, large artery atherosclerosis; SAO, small artery occlusion; UNDETER, undetermined; INCUNC, incomplete/unclassified; CRYPTCE, cryptogenic and CE minor; Cryptoincl, cryptogenic; CCSc, CCS Causative subtyping system; CCSp, CCS Phenotypic subtyping system; TOAST, TOAST subtyping system.

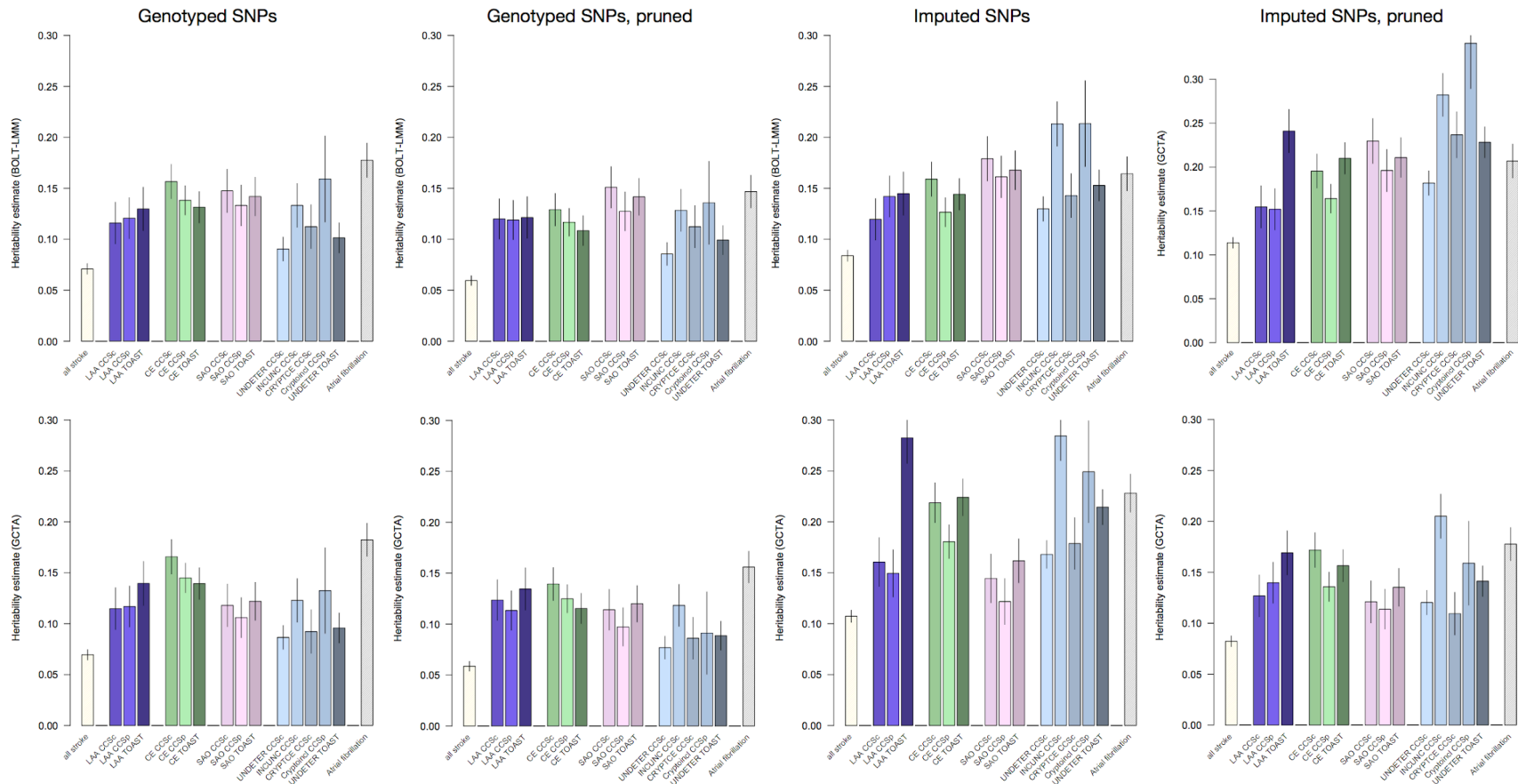


Supplementary Figure 4 | Estimated heritability of ischemic stroke subtypes and atrial fibrillation. Using all available stroke cases in SiGN, we estimated SNP-based heritability of the ischemic stroke subtypes (as sub-typed by the CCS Causative subtyping system) and atrial fibrillation (using the subset of 3,190 cases with atrial fibrillation) using BOLT-LMM and a genetic relationship matrix of high-quality SNPs converted to best-guess genotypes (imputation quality > 0.8, minor allele frequency > 0.01, and pruned at a linkage disequilibrium threshold of 0.2). We assumed a trait prevalence of 1% for all phenotypes. We found heritability estimates in cardioembolic stroke (green) and atrial fibrillation (yellow) to be approximately similar.



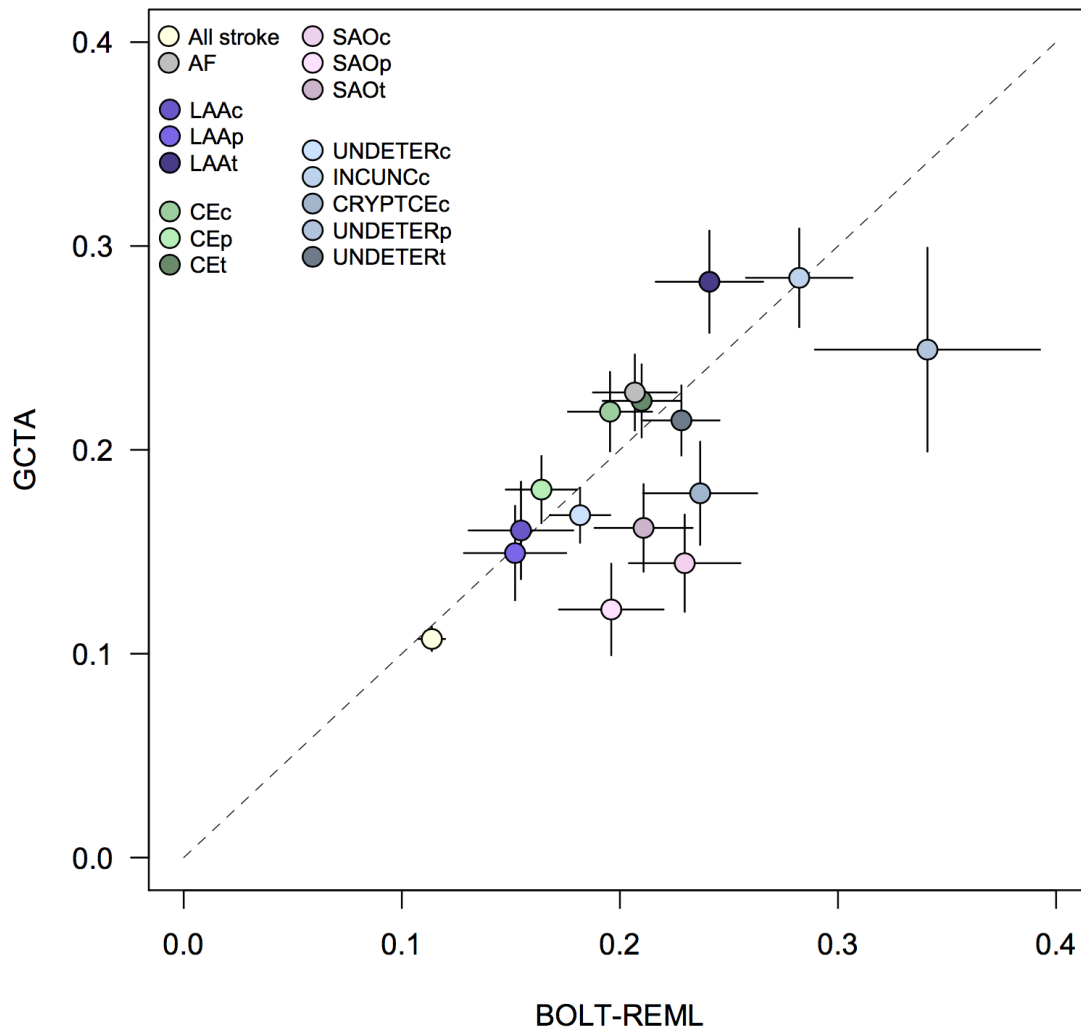
Supplementary Figure 5 | Heritability of ischemic stroke, its subtypes, and atrial fibrillation. We computed the SNP-based heritability of all stroke, all stroke subtypes, and atrial fibrillation using BOLT-LMM (top row) and GCTA (bottom row). All SNPs used for analysis had a minor allele frequency > 1% and imputation quality > 0.8 (for imputed SNPs). Imputed SNPs were converted to best-guess genotypes. We assumed a trait prevalence of 1% for all phenotypes and tested the robustness of h_g^2 estimates to SNPs included in the GRM by using four different GRMs: (a) genotyped SNPs only; (b) genotyped, pruned, and filtered (see **Supplemental Methods**); (c) imputed; and (d) imputed, pruned, and filtered. We converted the imputed SNPs to hard-call genotypes before performing heritability analyses. Estimates are shown below, including error bars. The underlying data for these figures are provided in **Supplementary Table 3**.

LAA, large artery atherosclerosis; CE, cardioembolic stroke; SAO, small artery occlusion; UNDETER, undetermined; INCUNC, incomplete/unclassified; CRYPTCE, cryptogenic and CE minor; Cryptoincl, cryptogenic; CCSc, CCS Causative subtyping system; CCSp, CCS Phenotypic subtyping system; TOAST, TOAST subtyping system.



Supplementary Figure 6 | Comparison of heritability estimates from BOLT-LMM and GCTA. We computed the heritability of all stroke, all stroke subtypes, and atrial fibrillation using BOLT-LMM and GCTA, as shown in **Supplementary Figure 2**. Below, you will find a comparison of the two methods, with BOLT-REML on the x-axis and GCTA estimates on the y-axis. Error bars are shown for the respective estimates.

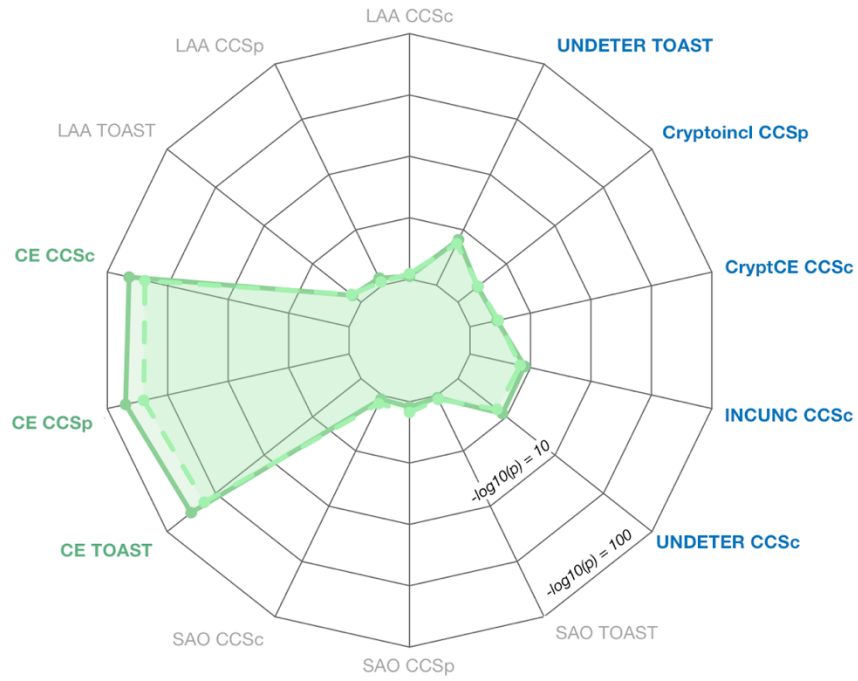
AF, atrial fibrillation; CE, cardioembolic stroke; LAA, large artery atherosclerosis; SAO, small artery occlusion; UNDETER, undetermined; INCUNC, incomplete/unclassified; CRYPTCE, cryptogenic/CE minor; c, CCS Causative; p, CCS Phenotypic; t, TOAST.



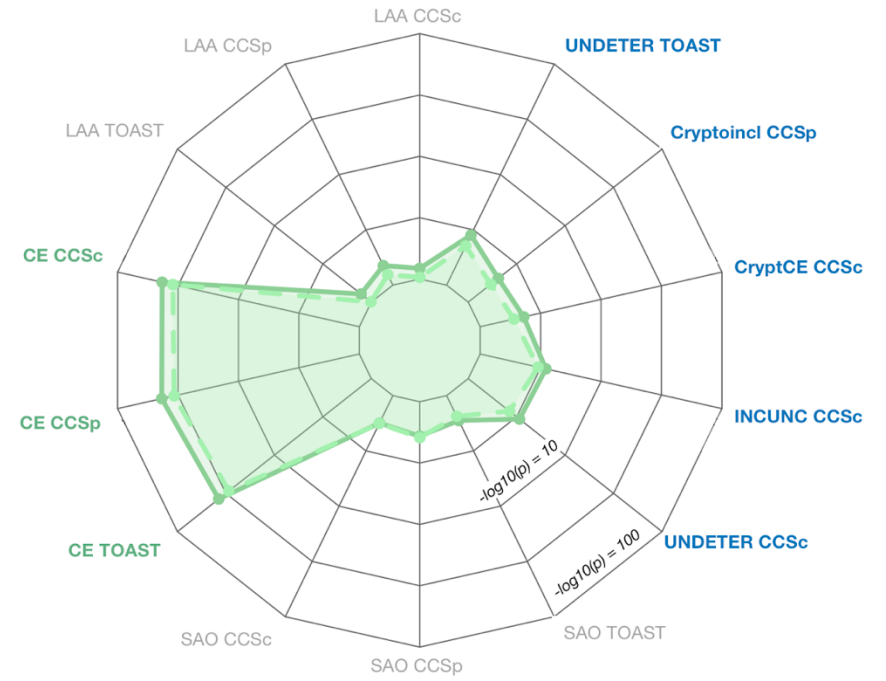
Supplementary Figure 7 | Association of atrial fibrillation polygenic risk score in ischemic stroke subtypes. We constructed a polygenic risk score (PRS) from atrial fibrillation-associated SNPs, and tested for association between the score and ischemic stroke subtypes using (a) all available controls (N = 28,026) and (b) controls without atrial fibrillation (N = 3,861). All subtypes from all available subtyping systems are shown here. The PRS strongly associated to cardioembolic stroke (subtypes highlighted in green font) in both sets of controls. In the atrial fibrillation-free set of controls (b) we observed nominal association of the PRS to incomplete/unclassified stroke. Undetermined subtypes are indicated in blue font.

CE, cardioembolic stroke; LAA, large artery atherosclerosis; SAO, small artery occlusion; UNDETER, undetermined; INCUNC, incomplete/unclassified; CRYPTCE, cryptogenic and CE minor; Cryptoincl, cryptogenic; CCSc, CCS Causative subtyping system; CCSp, CCS Phenotypic subtyping system; TOAST, TOAST subtyping system.

a All controls



b AF-free controls



Supplementary Tables

Supplementary Table 1 | Atrial fibrillation cases and controls available from the Stroke Genetics Network (SiGN) Consortium.

As classified by the CCS Causative system (note that this table is a repeat of **Table 1** from the main manuscript):

Phenotype	Total	Cardioembolic	Large artery atherosclerosis	Small artery occlusion	Undetermined	
					Incomplete/unclassified	Cryptogenic/CE minor
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

As classified by the CCS Phenotypic system (note that this system allows a case to be classified into more than one subtype):

Phenotype	Total	Cardioembolic	Large artery atherosclerosis	Small artery occlusion	Undetermined
Atrial fibrillation	1,751	1,751	161	58	0
Paroxysmal atrial fibrillation	1,315	1,315	126	61	0
Left atrial thrombus	48	48	7	4	0
Sick sinus syndrome	79	79	8	4	0
Atrial Flutter	106	106	11	3	0
Total	3,190	3,190	302	126	0

As classified by the TOAST system:

Phenotype	Total	Cardioembolic	Large artery atherosclerosis	Small artery occlusion	Undetermined
Atrial fibrillation	1,751	1,254	26	23	170
Paroxysmal atrial fibrillation	1,315	880	25	19	178
Left atrial thrombus	48	35	1	1	9
Sick sinus syndrome	79	48	0	1	13
Atrial Flutter	106	75	2	3	12
Total	3,190	2,207	54	47	371

Overlap of atrial fibrillation and cardioembolic stroke in the three subtyping systems in SiGN (CCSc, CCS Causative; CCSp, CCS Phenotypic; TOAST):

Phenotype	CCSc Cardioembolic	CCSp Cardioembolic	TOAST Cardioembolic
Atrial fibrillation	1,495	1,751	1,254
Paroxysmal atrial fibrillation	1,088	1,315	880
Left atrial thrombus	37	48	35
Sick sinus syndrome	65	79	48
Atrial Flutter	90	106	75
No atrial fibrillation phenotypes	316	418	903
Total	3,000	3,608	3,333

Supplementary Table 2 | Look-up of previously-associated atrial fibrillation SNPs in SiGN. After performing a GWAS of atrial fibrillation in the SiGN data, we looked up the 26 known genetic risk loci for atrial fibrillation, as identified in the latest GWAS.¹ Twenty-four of the 25 signals present in the SiGN data were directionally consistent with the previous GWAS. The only signal not directionally consistent was discovered through eQTL analysis. One signal, a rare variant burden signal, was absent from our data (all SNPs here have allele frequency > 1%).

Supplementary Table 2 is provided as a separate, downloadable Excel spreadsheet as well as a tab-delimited text available at the project GitHub repository (download: <https://github.com/saralpulit/Afib-Stroke-Overlap/blob/master/SupplementaryTable2.afib.hits.SiGN-lookup.txt>). The first 14 columns are taken from *Christophersen, et al.*¹ Those columns are:

SNP	single-nucleotide polymorphism; rs identifier
CHR	chromosome
BP	basepair (hg19)
Genes	Closest gene(s)
Location	Where the SNP resides relative to the listed gene
Risk	Risk allele
Ref	Reference allele
RAF	Risk allele frequency
OR	Odds ratio
CI95_1	95% confidence interval for the odds ratio (lower bound)
CI95_2	95% confidence interval for the odds ratio (upper bound)
Pval	Association p-value
Mean_imp	Imputation quality
Analysis	The analysis the variant or gene was discovered in (ExWAS, expression QTL analysis; Meta, meta-analysis; RVAS, rare variant association study)

The remaining columns provided are data points extracted from the atrial fibrillation GWAS in SiGN. They are:

SiGN_RAF	Risk allele frequency in SiGN
SiGN_INFO	Imputation quality (info score) in SiGN
SiGN_BOLT_BETA	Beta of the SNP taken from BOLT-LMM; note that this is a beta that results from a linear mixed model
SiGN_LIAB_BETA	The beta, converted to the liability scale
SiGN_OR	Odds ratio in SiGN
SiGN_SE	Standard error of SiGN_BOLT_BETA
SiGN_P_BOLT	P-value from BOLT-LMM (for the infinitesimal model only)

Supplementary Table 3 | Genetic correlations between atrial fibrillation and ischemic stroke subtypes. To estimate genetic correlation between atrial fibrillation and ischemic stroke subtypes, we calculated Pearson's r between SNP z-scores in the Atrial Fibrillation Genetics (AFGen) GWAS of atrial fibrillation and in GWAS of ischemic stroke subtypes and atrial fibrillation performed here in the SiGN data. The correlation calculations are provided in this table, which is split into two parts and is available to download in text format here:

Part A: correlations calculated across all genome-wide SNPs

<https://github.com/saralpulit/Afib-Stroke-Overlap/blob/master/SuppTable4.partA.SiGN.AFGen.trait.correlations.txt>

Part B: correlations calculated across all genome-wide SNPs except those ± 2 Mb from the PITX2 and ZFH3 index SNPs provided in Supplementary Table 2

<https://github.com/saralpulit/Afib-Stroke-Overlap/blob/master/SuppTable4.partB.SiGN.AFGen.trait.correlations.drop-pitx2-zfhx3.txt>

The headers of the two files are exactly the same:

Column	Definition
Z.threshold	Z-score threshold used to subset AFGen SNPs
EduYrs.Z	Correlation with z-scores from educational attainment GWAS
afib.broad.Z	Correlation with z-scores from atrial fibrillation (broadly defined phenotype) GWAS
allstroke.Z	Correlation with z-scores from all stroke GWAS
CCScCEmajor.Z	Correlation with z-scores from CCSc CE GWAS
CCScCRYPTCE.Z	Correlation with z-scores from CCSc CRYPTCE GWAS
CCScINCUNC.Z	Correlation with z-scores from CCSc INCUNC GWAS
CCScLAA.Z	Correlation with z-scores from CCSc LAA GWAS
CCScSAO.Z	Correlation with z-scores from CCSc SAO GWAS
CCScUNDETER.Z	Correlation with z-scores from CCSc UNDETER GWAS
CCSpCEmajincl.Z	Correlation with z-scores from CCSp CE GWAS
CCSpCryptoincl.Z	Correlation with z-scores from CCSp Cryptogenic GWAS
CCSpLAAmajincl.Z	Correlation with z-scores from CCSp LAA GWAS
CCSpSAOmajincl.Z	Correlation with z-scores from CCSp SAO GWAS
toastCE.Z	Correlation with z-scores from TOAST CE GWAS
toastLAA.Z	Correlation with z-scores from TOAST LAA GWAS
toastSAO.Z	Correlation with z-scores from TOAST SAO GWAS
toastUNDETER.Z	Correlation with z-scores from TOAST UNDETER GWAS

CCSc, CCS Causative subtyping system; CCSp, CCS Phenotypic subtyping system; TOAST, TOAST subtyping system; CE, cardioembolic stroke; LAA, large artery atherosclerosis; SAO, small artery occlusion; UNDETER, undetermined; INCUNC, incomplete/unclassified; CRYPTCE, cryptogenic and CE minor.

Supplementary Table 4 | Heritability calculations in atrial fibrillation and ischemic stroke subtypes. (a) We calculated the SNP-based heritability (h_g^2) of atrial fibrillation, all ischemic stroke, and the stroke subtypes using GCTA². All SNPs used had minor allele frequency > 1% and imputation quality > 0.8 (for imputed SNPs). Imputed SNPs were converted to best-guess genotypes. We assumed a trait prevalence of 1% for all phenotypes and tested the robustness of h_g^2 estimates to SNPs included in the GRM by using four different GRMs: (i) genotyped only; (ii) genotyped, pruned, and filtered (see **Supplemental Methods**); (iii) imputed; and (iv) imputed, pruned, and filtered. (b) We performed the exact same analysis but using BOLT-LMM to estimate h_g^2 . BOLT-LMM estimates were converted to the liability scale (see **Supplemental Methods**).

Geno, genotyped; SE, standard error; CCSc, CCS Causative; CCSp, CCS Phenotypic

a. h_g^2 estimates in GCTA

Subtype	Subtyping system	Cases	Geno h_g^2 (SE)	Geno, filtered h_g^2 (SE)	Imputed h^2 (SE)	Imputed, filtered h_g^2 (SE)
Large artery atherosclerosis	CCSc	2,385	0.115 (0.020)	0.124 (0.020)	0.127 (0.020)	0.160 (0.024)
	CCSp	2,449	0.117 (0.020)	0.113 (0.019)	0.140 (0.020)	0.149 (0.023)
	TOAST	2,318	0.139 (0.021)	0.135 (0.021)	0.169 (0.022)	0.282 (0.025)
Cardio-embolic	CCSc	3,000	0.166 (0.017)	0.139 (0.016)	0.172 (0.017)	0.219 (0.019)
	CCSp	3,608	0.145 (0.014)	0.125 (0.014)	0.136 (0.014)	0.181 (0.016)
	TOAST	3,333	0.139 (0.015)	0.115 (0.015)	0.156 (0.016)	0.224 (0.018)
Small artery occlusion	CCSc	2,262	0.118 (0.021)	0.114 (0.020)	0.121 (0.021)	0.144 (0.024)
	CCSp	2,419	0.106 (0.020)	0.097 (0.019)	0.114 (0.019)	0.122 (0.022)
	TOAST	2,631	0.122 (0.019)	0.120 (0.018)	0.135 (0.019)	0.162 (0.021)
Undetermined	CCSc	4,574	0.087 (0.012)	0.077 (0.011)	0.120 (0.012)	0.168 (0.014)
	CCSc (INCUNC)	2,280	0.123 (0.021)	0.118 (0.021)	0.205 (0.022)	0.284 (0.024)
	CCSc (CRYPTCE)	2,294	0.092 (0.021)	0.086 (0.020)	0.109 (0.021)	0.179 (0.025)
	CCSp	1,096	0.132 (0.042)	0.091 (0.040)	0.159 (0.041)	0.249 (0.050)
	TOAST	3,479	0.096 (0.015)	0.089 (0.014)	0.141 (0.015)	0.214 (0.017)
--	All stroke	13,390	0.069 (0.005)	0.059 (0.005)	0.082 (0.005)	0.107 (0.006)
--	Atrial fibrillation	3,190	0.182 (0.016)	0.156 (0.015)	0.178 (0.016)	0.228 (0.019)

b. h_g^2 estimates in BOLT-LMM

Subtype	Subtyping system	Cases	Geno h_g^2 (SE)	Geno, filtered h_g^2 (SE)	Imputed h^2 (SE)	Imputed, filtered h_g^2 (SE)
Large artery atherosclerosis	CCSc	2,385	0.116 (0.020)	0.120 (0.020)	0.120 (0.020)	0.155 (0.024)
	CCSp	2,449	0.121 (0.020)	0.119 (0.019)	0.142 (0.020)	0.152 (0.023)
	TOAST	2,318	0.130 (0.021)	0.121 (0.020)	0.145 (0.021)	0.241 (0.025)
Cardio-embolic	CCSc	3,000	0.157 (0.017)	0.129 (0.016)	0.159 (0.017)	0.195 (0.019)
	CCSp	3,608	0.138 (0.014)	0.117 (0.014)	0.127 (0.014)	0.164 (0.016)
	TOAST	3,333	0.131 (0.015)	0.108 (0.015)	0.144 (0.015)	0.210 (0.018)
Small artery occlusion	CCSc	2,262	0.147 (0.021)	0.151 (0.020)	0.179 (0.022)	0.230 (0.026)
	CCSp	2,419	0.133 (0.020)	0.127 (0.019)	0.161 (0.020)	0.196 (0.024)
	TOAST	2,631	0.142 (0.019)	0.142 (0.018)	0.168 (0.019)	0.211 (0.022)
Undetermined	CCSc	4,574	0.090 (0.012)	0.086 (0.011)	0.130 (0.012)	0.182 (0.014)
	CCSc (INCUNC)	2,280	0.133 (0.021)	0.118 (0.021)	0.128 (0.021)	0.282 (0.024)
	CCSc (CRYPTCE)	2,294	0.112 (0.021)	0.112 (0.021)	0.143 (0.021)	0.237 (0.026)
	CCSp	1,096	0.159 (0.042)	0.136 (0.041)	0.213 (0.042)	0.341 (0.052)
	TOAST	3,479	0.101 (0.015)	0.099 (0.014)	0.153 (0.015)	0.228 (0.017)
--	All stroke	13,390	0.169 (0.012)	0.059 (0.005)	0.084 (0.005)	0.114 (0.006)
--	Atrial fibrillation	3,190	0.169 (0.016)	0.140 (0.015)	0.156 (0.016)	0.200 (0.018)

Supplementary Table 5 | Association between the atrial fibrillation polygenic risk score and ischemic stroke subtypes. We tested the association between a polygenic risk score (PRS) constructed from atrial fibrillation-associated SNPs and all stroke subtypes. The results of those association tests are shown here. We used two groups of controls: all available controls (N = 28,026 in the model without clinical covariates; N = 14,357 in the model with clinical covariates) and all controls that were free of atrial fibrillation (AF, N = 3,860 in the model without clinical covariates; N = 3,786 in the model with clinical covariates). All analyses were adjusted for sex and principal components (PCs). Regression analyses were optionally adjusted for clinical covariates (age, cardiovascular disease, type 2 diabetes status, smoking status, and hypertension).

Significant results ($p = 0.0062$, Bonferroni-corrected for four subtype groups and two independent subtyping classifications -- CCS and TOAST -- are bolded).

SE, standard error; CCSc, CCS Causative; CCSp, CCS Phenotypic; covar, covariates.

Large artery atherosclerosis (LAA):

All controls included in model without clinical covariates, N = 28,026; with clinical covariates, N = 14,357

Non-AF controls included in model without clinical covariates, N = 3,860; with clinical covariates, N = 3,786

Case definition	Control definition	Cases		Logistic regression, adjusted for PCs and sex			Logistic regression, adjusted for PCs, sex, and clinical covariates		
		w/out clinical covars	with clinical covars	Beta	SE	P-value	Beta	SE	P-value
CCSc LAA	Non-AF controls	2,385	2,093	0.008	0.015	0.600	0.002	0.018	0.929
CCSc LAA	All controls	2,385	2,093	-0.002	0.012	0.885	-0.004	0.013	0.786
CCSp LAA	Non-AF controls	2,449	2,149	0.016	0.016	0.315	0.010	0.018	0.570
CCSp LAA	All controls	2,449	2,149	0.004	0.011	0.694	0.002	0.013	0.850
TOAST LAA	Non-AF controls	2,318	1,884	0.010	0.016	0.528	0.000	0.018	0.980
TOAST LAA	All controls	2,318	1,884	-0.006	0.012	0.594	-0.008	0.014	0.550
<i>Results after standardizing PRS to a z-score</i>									
CCSc LAA	Non-AF controls	2,385	2,093	0.016	0.030	0.600	0.003	0.035	0.929
CCSc LAA	All controls	2,385	2,093	-0.003	0.022	0.885	-0.007	0.026	0.786
CCSp LAA	Non-AF controls	2,449	2,149	0.031	0.030	0.315	0.020	0.035	0.570
CCSp LAA	All controls	2,449	2,149	0.009	0.022	0.694	0.005	0.026	0.850
TOAST LAA	Non-AF controls	2,318	1,884	0.019	0.031	0.528	-0.001	0.036	0.980
TOAST LAA	All controls	2,318	1,884	-0.012	0.023	0.594	-0.016	0.027	0.550

Cardioembolic stroke (CE):

All controls included in model without clinical covariates, N = 28,026; with clinical covariates, N = 14,357

Non-AF controls included in model without clinical covariates, N = 3,860; with clinical covariates, N = 3,786

Case definition	Control definition (N)	Cases		Logistic regression, adjusted for PCs and sex			Logistic regression, adjusted for PCs, sex, and clinical covariates		
				Beta	SE	P-value	Beta	SE	P-value
CCSc CE	Non-AF (3,869)	3,000	2,725	0.187	0.014	1.59E-42	0.218	0.018	1.40E-34
CCSc CE	All (28,026)	3,000	2,725	0.169	0.010	1.01E-65	0.173	0.012	1.45E-48
CCSp CE	Non-AF (3,869)	3,608	3,281	0.178	0.013	6.98E-43	0.203	0.017	8.34E-34
CCSp CE	All (28,026)	3,608	3,281	0.161	0.009	2.43E-70	0.163	0.011	1.05E-49
TOAST CE	Non-AF (3,869)	3,333	3,074	0.171	0.013	3.17E-37	0.172	0.015	3.22E-29
TOAST CE	All (28,026)	3,333	3,074	0.149	0.009	3.00E-56	0.146	0.011	4.43E-41
<i>Results after standardizing PRS to a z-score</i>									
CCSc CE	Non-AF (3,869)	3,000	2,725	0.365	0.027	1.59E-42	0.425	0.035	1.40E-34
CCSc CE	All (28,026)	3,000	2,725	0.329	0.019	1.01E-65	0.337	0.023	1.45E-48
CCSp CE	Non-AF (3,869)	3,608	3,281	0.348	0.025	6.98E-43	0.397	0.033	8.34E-34
CCSp CE	All (28,026)	3,608	3,281	0.315	0.018	2.43E-70	0.318	0.021	1.05E-49
TOAST CE	Non-AF (3,869)	3,333	3,074	0.334	0.026	3.17E-37	0.335	0.030	3.22E-29
TOAST CE	All (28,026)	3,333	3,074	0.291	0.018	3.00E-56	0.284	0.021	4.43E-41

Small artery occlusion (SAO):

All controls included in model without clinical covariates, N = 28,026; with clinical covariates, N = 14,357

Non-AF controls included in model without clinical covariates, N = 3,860; with clinical covariates, N = 3,786

Case definition	Control definition (N)	Cases		Logistic regression, adjusted for PCs and sex			Logistic regression, adjusted for PCs, sex, and clinical covariates		
				Beta	SE	P-value	Beta	SE	P-value
CCSc SAO	Non-AF (3,869)	2,262	2,124	0.023	0.017	0.170	0.026	0.019	0.163
CCSc SAO	All (28,026)	2,262	2,124	0.002	0.012	0.842	0.006	0.013	0.660
CCSp SAO	Non-AF (3,869)	2,419	2,267	0.025	0.016	0.124	0.029	0.018	0.109
CCSp SAO	All (28,026)	2,419	2,267	0.003	0.012	0.787	0.007	0.013	0.602
TOAST SAO	Non-AF (3,869)	2,631	2,415	0.021	0.016	0.209	0.019	0.018	0.289
TOAST SAO	All (28,026)	2,631	2,415	0.001	0.011	0.902	0.003	0.013	0.826
<i>Results after standardizing PRS to a z-score</i>									
CCSc SAO	Non-AF (3,869)	2,262	2,124	0.046	0.033	0.170	0.051	0.036	0.163
CCSc SAO	All (28,026)	2,262	2,124	0.005	0.023	0.842	0.012	0.026	0.660
CCSp SAO	Non-AF (3,869)	2,419	2,267	0.049	0.032	0.124	0.057	0.035	0.109
CCSp SAO	All (28,026)	2,419	2,267	0.006	0.023	0.787	0.013	0.025	0.602
TOAST SAO	Non-AF (3,869)	2,631	2,415	0.040	0.032	0.209	0.037	0.035	0.289
TOAST SAO	All (28,026)	2,631	2,415	0.003	0.022	0.902	0.005	0.025	0.826

Undetermined strokes:

All controls included in model without clinical covariates, N = 28,026; with clinical covariates, N = 14,357

Non-AF controls included in model without clinical covariates, N = 3,860; with clinical covariates, N = 3,786

Case definition	Control definition (N)	Cases		Logistic regression, adjusted for PCs and sex			Logistic regression, adjusted for PCs, sex, and clinical covariates		
				Beta	SE	P-value	Beta	SE	P-value
CCSc UNDETER	Non-AF (3,869)	4,574	4,169	0.036	0.013	0.004	0.031	0.014	0.022
CCSc UNDETER	All (28,026)	4,574	4,169	0.021	0.009	0.013	0.021	0.010	0.030
CCSc INCUNC	Non-AF (3,869)	2,280	2,093	0.046	0.016	0.003	0.045	0.017	0.010
CCSc INCUNC	All (28,026)	2,280	2,093	0.028	0.012	0.015	0.029	0.013	0.025
CCSc CRYPTCE	Non-AF (3,869)	2,294	2,076	0.030	0.016	0.051	0.026	0.017	0.124
CCSc CRYPTCE	All (28,026)	2,294	2,076	0.015	0.012	0.212	0.017	0.013	0.192
CCSp Crypto	Non-AF (3,869)	1,096	972	0.035	0.020	0.090	0.029	0.022	0.195
CCSp Crypto	All (28,026)	1,096	972	0.019	0.016	0.258	0.021	0.018	0.245
TOAST UNDETER	Non-AF (3,869)	3,479	3,216	0.033	0.013	0.015	0.028	0.014	0.055
TOAST UNDETER	All (28,026)	3,479	3,216	0.021	0.010	0.027	0.022	0.011	0.042
<i>Results after standardizing PRS to a z-score</i>									
CCSc UNDETER	Non-AF (3,869)	4,574	4,169	0.071	0.025	0.004	0.061	0.027	0.022
CCSc UNDETER	All (28,026)	4,574	4,169	0.041	0.017	0.013	0.041	0.019	0.030
CCSc INCUNC	Non-AF (3,869)	2,280	2,093	0.090	0.030	0.003	0.088	0.034	0.010
CCSc INCUNC	All (28,026)	2,280	2,093	0.055	0.023	0.015	0.056	0.025	0.025
CCSc CRYPTCE	Non-AF (3,869)	2,294	2,076	0.059	0.030	0.051	0.051	0.033	0.124
CCSc CRYPTCE	All (28,026)	2,294	2,076	0.028	0.023	0.212	0.033	0.025	0.192
CCSp Crypto	Non-AF (3,869)	1,096	972	0.068	0.040	0.090	0.057	0.044	0.195
CCSp Crypto	All (28,026)	1,096	972	0.036	0.032	0.258	0.041	0.035	0.245
TOAST UNDETER	Non-AF (3,869)	3,479	3,216	0.064	0.026	0.015	0.054	0.028	0.055
TOAST UNDETER	All (28,026)	3,479	3,216	0.042	0.019	0.027	0.042	0.021	0.042

UNDETER, undetermined; INCUNC, incomplete and unclassified; CRYPTCE, cryptogenic and CE minor; Crypto, cryptogenic

Supplementary Table 6 | Sensitivity analysis for the atrial fibrillation polygenic risk score. As a sensitivity analysis for the polygenic risk score (PRS), we constructed 3 additional PRSs, including SNPs +/- 25kb, +/- 50kb, and +/- 100kb from the SNPs included in the original score. All scores remain highly significant when tested for association with cardioembolic stroke (using a logistic regression model). P-values after additionally adjusting for clinical covariates are also shown. Clinical covariates: age, cardiovascular disease, type 2 diabetes status, smoking status, and hypertension.

PCs, principal components; MAF, minor allele frequency; INFO, imputation (info) score.

PRS SNPs	Filters	Total SNPs	PRS p-value	
			<i>Adjusted for PCs, sex</i>	<i>Adjusted for PCs, sex, clinical covariates</i>
Original SNPs	MAF > 1% Info > 0.8	975	1.01×10^{-65}	1.44×10^{-48}
Original SNPs +/- 25kb	MAF > 1% Info > 0.8	146,631	9.13×10^{-50}	1.32×10^{-37}
Original SNPs +/- 50kb	MAF > 1% Info > 0.8	258,870	5.76×10^{-48}	1.40×10^{-36}
Original SNPs +/- 100kb	MAF > 1% Info > 0.8	462,146	4.47×10^{-44}	1.77×10^{-32}

Supplementary Table 7 | Clinical covariates available in the SiGN data. We adjusted our analyses of a polygenic risk score for a series of clinical covariates that are associated with atrial fibrillation. Summary-statistics on these covariates are shown below for those samples classified as (a) cardioembolic stroke or (b) undetermined stroke. The number of samples with missing data are provided in parentheses where relevant.

Cardioembolic

Phenotype	CCS Causative	CCS Phenotypic	TOAST
Female	1,588	1,859	1,618
Male	1,247	1,541	1,520
Age: mean (sd)	74.7 (12.4)	74.5 (12.3)	71.0 (15.1)
Hypertensive (missing)	2,195 (18)	2,665 (21)	2,272 (16)
Diabetes mellitus (missing)	763 (26)	950 (29)	799 (8)
CAD (missing)	989 (64)	1206 (83)	911 (119)
Smoking			
Current	379	468	513
Former	694	865	776
Never	1,737	2,055	1,905
Total	3,000	3,608	3,333

Undetermined

Phenotype	CCS Causative	CCS Causative	CCS Causative	CCS Phenotypic	TOAST
Female	1,880	1,024	856	420	1,445
Male	2,151	1,014	1,137	543	1,635
Age: mean (sd)	63.9 (15.4)	67.7 (13.9)	69.0 (15.9)	58.9 (15.7)	63.7 (16.1)
Hypertensive (missing)	2,833 (23)	1,512 (14)	1,321 (9)	612 (3)	2,110 (29)
Diabetes mellitus (missing)	958 (26)	513 (14)	445 (12)	202 (4)	708 (25)
CAD (missing)	739 (169)	421 (86)	318 (83)	115 (46)	573 (100)
Smoking					
Current	1,090	582	508	239	813
Former	1,050	516	534	235	772
Never	2,202	1,081	1,121	548	1,711
Total	4,574	2,280	2,294	1,096	3,479

Supplementary Table 8: Variance explained by the atrial fibrillation polygenic risk score in cardioembolic stroke.

To determine the variance explained by the atrial fibrillation polygenic risk score (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. Below is variance explained for each of the cardioembolic stroke phenotypes as determined by the three subtyping systems available in SiGN: CCS Causative, CCS Phenotypic, and TOAST. Standard errors of each estimate appear in parentheses. Explained variance is shown for a PRS including the *PITX2* (chromosome 4) and *ZFHX3* (chromosome 16) loci, as well as excluding ± 2 Mb around these loci (see <https://github.com/UMCUGenetics/Afib-Stroke-Overlap> for lists of SNPs that fall in these regions). Because a large number of SNPs is needed to construct a variance component to calculate variance explained, we performed the calculation using the atrial fibrillation PRS including SNPs ± 100 kb from the original PRS SNPs, and then pruning SNPs a linkage disequilibrium of 0.2.

CE, cardioembolic; PRS, polygenic risk score; AF, atrial fibrillation

Subtyping System	h_g^2 CE stroke	h_g^2 atrial fibrillation PRS ± 100 kb	Proportion of CE h_g^2 explained by AF PRS
<i>PRS including the PITX2 and ZFHX3 loci</i>			
CCSc	0.195 (0.019)	0.045 (0.010)	23.1%
CCSp	0.164 (0.016)	0.040 (0.008)	24.4%
TOAST	0.210 (0.018)	0.051 (0.01)	24.3%
<i>PRS excluding the PITX2 and ZFHX3 loci</i>			
CCSc	0.195 (0.019)	0.037 (0.010)	19.0%
CCSp	0.164 (0.016)	0.032 (0.008)	19.5%
TOAST	0.210 (0.018)	0.044 (0.009)	21.0%

Supplementary Methods

GitHub repository and data availability

1. GitHub repository and additional supporting data

Relevant code for the analyses performed in this paper can be found here: <https://github.com/saralpulit/Afib-Stroke-Overlap>.

This repository primarily consists of:

- Call to BOLT-LMM to run GWAS

- Call to GCTA and BOLT-LMM to calculate heritability

- Call to PLINK^{3,4} to calculate the polygenic risk score (PRS)

- An R script for converting observed heritability in BOLT-LMM to the liability scale (see below)

- A script in R to check association between the PRS and various phenotypes.

- A call to PLINK^{3,4} to calculate a GRM to run GCTA

- Sample identifiers for those individuals analyzed in this paper

- SNP identifiers and weights for those markers included in the construction of the polygenic risk score

A complete README accompanies the GitHub repository.

2. Sample and SNP identifiers used in these analyses

A file containing:

- the dbGaP sample identifiers

- the cohort the sample is drawn from

- the continental group the sample is in (as determined in the first SiGN GWAS effort⁵)

- a list of quality control-passing SNPs used in the initial GWAS

is available on this paper's GitHub repository.

3. Downloadable summary-level genome-wide association study data

The summary-level data from the original SiGN GWAS has been made publicly available through the Cerebrovascular Disease Knowledge Portal, which can be accessed here: <http://www.cerebrovascularportal.org/>

These summary-level results are available for cardioembolic stroke (CE), large artery atherosclerosis (LAA), small artery occlusion (SAO), and undetermined (UNDETER) stroke, for three different subtyping systems (TOAST, CCS Causative, CCS Phenotypic).

The summary-level results for the atrial fibrillation genome-wide association studies (performed in broadly-defined or strictly-defined cases versus all controls) are available here:

Broadly-defined atrial fibrillation cases vs. all referents:

<https://doi.org/10.5281/zenodo.1035871>

Strictly-defined atrial fibrillation cases vs. all referents:

<https://doi.org/10.5281/zenodo.1035873>

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

The full list of cohorts that are included in the SiGN genome-wide association study can be found in the Supplementary Material of "Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study,"⁵ which can be accessed here: <https://paperpile.com/shared/nvNXQf>.

SiGN is comprised of several case cohorts with pre-existing genotyping data. Newly-collected cases, as well as a small number of matched referents, were genotyped on the Illumina 5M array⁶. The majority of referents included were drawn from publicly-available genotyping data.

1. Referent (control) datasets

Referent datasets downloaded from the Database of Genotypes and Phenotypes (dbGaP) are:

	dbGAP accession #
Genetics Resource with the Health and Retirement Study	phs000428.v2.p2
Whole Genome Association Study of Visceral Adiposity in the HABC study	phs000169.v1.p1

2. Case datasets

A large number of cases and a small number of controls (from Belgium and Poland) were genotyped at the initiation of the SiGN GWAS. These data have been uploaded to dbGaP and are available here:

The National Institute of Neurological Disorders and Stroke (NINDS) Stroke Genetics Network (SiGN) (phs000615.v1.p1)

3. Phenotyping in SiGN

There are three primary subtype definitions of ischemic stroke: cardioembolic stroke, large artery atherosclerotic stroke, and small artery occlusion. The SiGN consortium used the CCS system to attempt to assign each case to one of these three categories. Additionally, ~74% of cases were also classified using the Trial of Org 10 172 in Acute Stroke Treatment (TOAST)^{7,8} system, which classifies stroke cases based on clinical decision-making and clinically-ascertained information. The CCS and TOAST subtyping systems yield moderately-to-strongly correlated phenotyping results (**Supplementary Figure 5**)⁹. Use of these traits in a GWAS setting also yields concordant association results, as previously shown⁶. These subtypes are similarly defined in CCS and TOAST, though determined differently across the two subtyping systems.

In addition to the three primary subtypes, both the CCS and TOAST classification systems generate two additional subtypes: "undetermined" and "other." The "other" classification was small in sample size ($N_{\text{cases}} = 595, 719$ and 374 in CCS Causative, CCS Phenotypic and TOAST, respectively), and was therefore not included in the original SiGN GWAS and was not tested here⁶. The "undetermined" classification, though named the same in CCS and TOAST, is defined differently across the two subtyping systems^{8,10}. In TOAST, patients with conflicting subtype classifications are

placed in the undetermined category^{6,8}. In contrast, the CCS undetermined classification includes patients with cryptogenic embolism, other cryptogenic cases, patients with an incomplete evaluation, or samples with competing subtypes¹⁰.

4. Brief summary of data quality control in SiGN

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 Hispanic⁶.

For quality control (QC) and downstream association testing, cases and referents were matched by genotyping array and PCA-determined ancestry. European-ancestry samples were imputed with IMPUTE2¹¹ using a reference panel built from whole-genome sequence data collected by the 1000 Genomes Project (Phase 1)¹² and the Genome of the Netherlands¹³ project; African-ancestry and Hispanic samples were imputed with the 1000 Genomes Project data only.¹² Due to data-sharing restrictions regarding the referents used for the Hispanic set of samples, only the European- and African-ancestry samples were analyzed here, totaling 13,390 cases and 28,026 referents distributed across 12 case-control analysis groups.

Before performing genome-wide association testing, for those SNPs that were genotyped in a subset of the SiGN study strata but imputed in others, we compared the frequency of the SNP across the various strata. We removed any SNP with a frequency difference > 15% within ancestral group or >50% across ancestral groups comparing imputed and genotyped data, likely induced by sequencing errors in the imputation reference panel(s).

Constructing a genetic relationship matrix for genome-wide association testing in BOLT-LMM

To construct the genetic relationship matrix (GRM) implemented in BOLT-LMM, we used SNPs that were (i) common (MAF > 5%), (ii) with missingness < 5%, (iii) linkage disequilibrium (LD) pruned at an r^2 threshold of 0.2, (iv) on the autosomal chromosomes only, (v) and not in stratified areas of the genome (i.e., not in the major histocompatibility complex (MHC), the inversions on chromosomes 8 and 17, or in the lactase (*LCT*) locus on chromosome 2). After association testing, we additionally removed SNPs with imputation quality (info score) < 0.8, due to excess inflation of the test statistic in those SNPs (**Supplementary Figure 1**).

Running a genome-wide association study using BOLT-LMM

We implemented a linear mixed model to perform association testing using BOLT-LMM.¹⁴ 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.¹⁵⁻¹⁷ Due to extensive structure in the SiGN data,⁶ 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.¹⁴ We calculated PCs in EIGENSTRAT¹⁸ using a similar set of SNPs to that used in the genetic relationship matrix 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^{3,4} to convert imputed dosages to best-guess genotypes and then compute the GRM.

SNP-based heritability calculations in GCTA and BOLT-LMM

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 (GCTA²).

To check the robustness of the heritability calculations to the SNPs included in the GRM, we calculated heritability using the GRM described above, as well as three additional GRMs: (i) using the ~1.1M SNPs with imputation quality > 0.8 and MAF > 1% (and without LD pruning); (ii) using the SNPs that were genotyped across all study strata (~155,000 SNPs); and (iii) the set of genotyped SNPs with the MHC, *LCT* locus, inversions on chromosomes 8 and 17 removed, and LD pruned at $r^2 = 0.2$.

Additionally, we computed heritability in GCTA² using the same GRMs and assuming a trait prevalence of 1%. We compared the results to the BOLT-based h_g^2 estimates (**Supplementary Table 3** and **Supplementary Figures 2-3**). As genome-wide heritability estimates need a large number of SNPs to be accurate, we report in the paper all estimates using a GRM containing imputed, pruned SNPs. Estimates resulting from all GRMs are presented here, in the **Supplementary Information**.

To test the effect of changing the GRM (referred to by the --bfile and 'modelSNPs' option in BOLT-LMM), we selected SNPs for the GRM in four ways:

- (1) Genotyped SNPs only (minor allele frequency > 1%) (115,553 SNPs total)

- (2) Genotyped SNPs, pruned at a linkage disequilibrium threshold (r^2 threshold) of 0.2, and removing the MHC, *LCT* locus, and two chromosomal inversions. (60,432 SNPs total)
- (3) Imputed SNPs (minor allele frequency > 1% and imputation info > 0.8) converted to best-guess genotypes. (1,128,985 SNPs total)
- (4) Imputed SNPs (minor allele frequency > 1% and imputation info > 0.8); pruned at a linkage disequilibrium threshold (r^2 threshold) of 0.2; removing the MHC, *LCT* locus, and two chromosomal inversions; and converted to best-guess genotypes. (250,209 SNPs total)

The GRM in (4) is the GRM used for all heritability results presented in the main manuscript.

As calculating GRMs in GCTA can be extremely computationally intensive, we calculated the GRMs using PLINK 1.9 and then used those GRMs to estimate heritability. A script that shows how to do this is included in the GitHub repository noted above.

The genomic locations (hg19) for excluded markers are as follows:

The lactase (<i>LCT</i>) locus	Chromosome 2 positions 129,883,530 - 140,283,530
The major histocompatibility complex (MHC)	Chromosome 6 positions 24,092,021 - 38,892,022
Inversion 1	Chromosome 8 positions 6,612,592 - 13,455,629
Inversion 2	Chromosome 17 positions 40,546,474 - 44,644,684
All non-autosomal SNPs	--

BOLT-LMM produces heritability estimates on the observed scale. To convert to the liability scale (i.e., the scale on which GCTA produces heritability estimates) we performed a conversion in R. Running the conversion requires knowing the trait prevalence, total cases analyzed, total controls analyzed, and the heritability on the observed scale. This code snippet is available in the accompanying GitHub repository for this paper.

Quality control in genome-wide data for correlation calculations

We used summary-level data from the latest Atrial Fibrillation Genetics (AFGen) Consortium meta-analysis of atrial fibrillation¹ to calculate a z-score for each SNP in that GWAS. Additionally, we calculated a z-score for each SNP in a 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 attainment¹⁹ available through LDHub (<http://ldsc.broadinstitute.org/>, accessed 11-1-2017). We aligned z-score signs based on the risk allele reported in each study. SNPs with an allele frequency difference >5% between AFGen and SiGN (all stroke analysis) were removed from the AFGen data (25,784 SNPs); similarly, SNPs with an allele frequency difference >5% between the educational attainment GWAS and SiGN (all stroke) were also removed (27,866 SNPs). Finally, we calculated Pearson's r between z-scores from two traits to evaluate correlation.

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.²⁰ Briefly, the PRS was derived using results from a recent GWAS of atrial fibrillation, comprised of 17,931 cases and 115,142 referents¹ 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).²⁰ 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 0.5.²⁰

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.¹ We computed the final PRS for each individual by summing across all of the weighted genotypes and performed association testing in R.

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:

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

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

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 (the data were essentially identical to our finding in the complete sample and are therefore not provided).

Supplementary Results

Including age as a covariate in the GWAS of atrial fibrillation

To check for the effects of age on our initial GWAS findings, we ran a GWAS of atrial fibrillation including age as a covariate. Controls without age information were dropped from this analysis. Given the structure of the SiGN dataset -- which includes groups of cases and controls that have been carefully matched on genotyping array and ancestry -- we also dropped the cases for which their matched controls were missing age information.

Our age-adjusted analysis included 2,487 atrial fibrillation cases and 22,072 controls. We performed the GWAS in BOLT-LMM, adjusting for 10 PCs, sex and age. We then checked the correlation between the SNP effects (betas) from the GWAS unadjusted for age and the SNP effects from the GWAS adjusted for age. Correlation was strong ($r = 0.83$).

Appendix I

Members of the Atrial Fibrillation Genetics (AFGen) Consortium

Please note that the AFGen Consortium participants evolve over time. Further information on the AFGen Consortium can be found at www.afgen.org.

Ingrid E. Christophersen, MD, PhD¹⁻³ Joylene E. Siland, BSc,⁴ Marju Orho-Melander, PhD⁹²
 Michiel Rienstra, MD, PhD⁴ Michiaki Kubo, MD, PhD⁵³ Anders Hamsten, MD, PhD⁹³
 Carolina Roselli, MSc^{1,5,6} Jonathan D. Smith, PhD⁹ Jan Heeringa, MD, PhD²²
 Xiaoyan Yin, PhD^{7,8} David R. Van Wagener, PhD⁹ Joshua C. Denny, MD¹⁶
 Bastiaan Geelhoed, PhD⁴ Joshua C. Bis, PhD²¹ Jennifer Kriebel, PhD^{54,63,69}
 John Barnard, PhD⁹ Siegfried Perz, MSc⁵⁴ Dawood Darbar, MD⁹⁴
 Honghuang Lin, PhD^{7,8} Bruce M. Psaty, MD, PhD^{21,55-57} Christopher Newton-Cheh, MD, MPH^{1,2}
 Dan E. Arking, PhD¹⁰ Paul M. Ridker, MD, MPH¹³ Christian Shaffer, BS,¹⁶
 Albert V. Smith, PhD^{11,12} Jared W. Magnani, MD, MSc^{7,58} Peter W. Macfarlane, PhD, DSc⁹⁵
 Christine M. Albert, MD, MPH¹³ Tamara B. Harris, MD, MS⁵⁹ Stefanie Heilmann, PhD^{96,97}
 Mark Chaffin, MSc¹ Lenore J. Launer, PhD⁵⁹ Peter Almgren, MSc⁹²
 Nathan R. Tucker, PhD^{1,2} M. Benjamin Shoemaker, MD, MSCI¹⁶ Paul L. Huang, MD, PhD²
 Molong Li, MD² Sandosh Padmanabhan, MD⁶⁰ Nona Sotoodehnia, MD, MPH⁹⁸
 Derek Klarin, MD¹ Jeffrey Haessler, MS⁶¹ Elsayed Z. Soliman, MD, MSc, MS⁹⁹
 Nathan A. Bihlmeyer, BS,¹⁴ Traci M. Bartz, MS⁶² Andre G. Uitterlinden, PhD¹⁰⁰
 Siew-Kee Low, PhD¹⁵ Melanie Waldenberger, PhD^{19,54,63} Albert Hofman, MD, PhD²²
 Peter E. Weeke, MD, PhD^{16,17} Peter Lichtner, PhD⁶⁴ Oscar H. Franco, MD, PhD²²
 Martina Müller-Nurasyid, PhD^{5,18,19} Marina Arendt, MSc⁶⁵ Uwe Völker, PhD^{24,42}
 J. Gustav Smith, MD, PhD^{1,20} Jose E. Krieger, MD, PhD³³ Karl-Heinz Jöckel, PhD⁶⁵
 Jennifer A. Brody, BA²¹ Mika Kähönen, MD, PhD⁶⁶ Moritz F. Sinner, MD, MPH^{18,19}
 Maartje N. Niemeijer MD²² Lorenz Risch, MD, MPH⁶⁷ Henry J. Lin, MD⁴⁰
 Marcus Dörr, MD^{23,24} Alfredo J. Mansur, MD, PhD⁶⁸ Xiuqing Guo, PhD⁴⁰
 Stella Trompet, PhD²⁵ Annette Peters, PhD^{19,54,69} Martin Dichgans, MD^{32,101,102}
 Jennifer Huffman, PhD²⁶ Blair H. Smith, MD⁷⁰ Erik Ingelsson, MD, PhD^{27,103}
 Stefan Gustafsson, PhD²⁷ Lars Lind, MD, PhD⁷¹ Charles Kooperberg, PhD⁶¹
 Claudia Schurmann, PhD^{28,29} Stuart A. Scott, PhD⁷² Olle Melander, MD, PhD¹⁰⁴
 Marcus E. Kleber, PhD³⁰ Yingchang Lu, MD, PhD^{28,29} Ruth J. F. Loos, PhD^{28,29,105}
 Leo-Pekka Lyytikäinen, MD³¹ Erwin B. Bottinger, MD^{28,73} Jari Laurikka, MD, PhD¹⁰⁶
 Ilkka Seppälä, MD³¹ Jussi Hernesniemi, MD, PhD^{31,74} David Conen, MD, MPH³⁶⁻³⁸
 Rainer Malik, PhD³² Cecilia M. Lindgren, PhD⁷⁵ Jonathan Rosand, MD, MSc^{1,41}
 Andrea R. V. R. Horimoto, PhD³³ Jorge A. Wong, MD⁷⁶ Pim van der Harst, MD, PhD⁴
 Marco Perez, MD³⁴ Jie Huang, MD, MPH⁷⁷ Marja-Liisa Lokki, PhD⁹¹
 Juha Sinisalo, MD, PhD³⁵ Markku Eskola, MD, PhD⁷⁴ Sekar Kathiresan, MD¹
 Stefanie Aeschbacher, MSc^{36,37} Andrew P. Morris, PhD^{75,78} Alexandre Pereira, MD, PhD¹⁰⁷
 Sébastien Thériault, MD, MSc^{38,39} Ian Ford, PhD⁷⁹ J. Wouter Jukema, MD, PhD^{25,108,109}
 Jie Yao, MS⁴⁰ Alex P. Reiner, MD, MSc^{61,80} Caroline Hayward, PhD²⁶
 Farid Radmanesh, MD, MPH^{1,41} Graciela Delgado, MSc³⁰ Jerome I. Rotter, MD¹¹⁰
 Stefan Weiss, PhD^{24,42} Lin Y. Chen, MD, MS⁸¹ Winfried März, MD¹¹¹
 Alexander Teumer, PhD^{24,43} Yii-Der Ida Chen, PhD⁴⁰ Terho Lehtimäki, MD, PhD³¹
 Seung Hoan Choi, PhD¹ Roopinder K. Sandhu, MD, MPH⁸² Bruno H. Stricker, MD, PhD¹¹²
 Lu-Chen Weng, PhD^{1,2} Man Li, PhD^{83,84} Mina K. Chung, MD⁹
 Sebastian Clauss, MD^{2,18} Eric Boerwinkle, PhD⁸⁵ Stephan B. Felix, MD^{23,24}
 Rajat Deo, MD, MTR⁴⁴ Lewin Eisele, MD⁶⁵ Vilmundur Gudnason, MD, PhD^{11,12}
 Daniel J. Rader, MD⁴⁴ Lars Lannfelt, MD, PhD⁸⁶ Alvaro Alonso, MD, PhD¹¹³
 Svati Shah, MD, MHS,⁴⁵ Natalia Rost, MD, MPH, FAAN,^{1,87} Dan M. Roden, MD¹⁶
 Albert Sun, MD⁴⁵ Christopher D. Anderson, MD, MMSc^{1,41} Stefan Kääh, MD, PhD^{18,19}
 Jemma C. Hopewell, PhD⁴⁶ Kent D. Taylor, PhD⁴⁰ Daniel I. Chasman, PhD^{1,114}
 Stephanie Debette, MD, PhD⁴⁷⁻⁵⁰ Archie Campbell, MA,⁸⁸ Susan R. Heckbert, MD, PhD^{55,56}
 Ganesh Chauhan, PhD^{47,48} Patrik K. Magnusson, PhD⁸⁹ Emelia J. Benjamin, MD, ScM^{7,58,115}
 Qiong Yang, PhD⁵¹ David Porteous, PhD⁸⁸ Toshihiro Tanaka, MD, PhD^{116,117}
 Bradford B. Worrall, MD, MSc⁵² Lynne J. Hocking, PhD⁹⁰ Kathryn L. Lunetta, PhD^{7,8}
 Guillaume Paré, MD, MSc^{38,39} Efthymia Vlachopoulou, PhD⁹¹ Steven A. Lubitz, MD, MPH^{1,2,118}
 Yoichiro Kamatani, MD, PhD¹⁵ Nancy L. Pedersen, MA, PhD⁸⁹ Patrick T. Ellinor, MD, PhD^{1,2,118}
 Yanick P. Hagemeijer, MSc⁴ Kjell Nikus, MD, PhD⁷⁴

AFGen Consortium Member Affiliations

1. Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA.
2. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA.
3. Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust, Norway.
4. Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
5. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
6. Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany.
7. NHLBI and Boston University's Framingham Heart Study, Framingham, MA, USA.
8. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.
9. Departments of Cardiovascular Medicine, Cellular and Molecular Medicine, Molecular Cardiology, and Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA.
10. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
11. Icelandic Heart Association, Kopavogur, Iceland.
12. Faculty of Medicine, University of Iceland, Reykavik, Iceland.
13. Divisions of Preventive and Cardiovascular Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA.
14. Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
15. Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.
16. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA.
17. The Heart Centre, Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark.
18. Department of Medicine I, University Hospital Munich, Ludwig-Maximilians-University, Munich, Germany.
19. DZHK (German Centre for Cardiovascular Research), partner site: Munich Heart Alliance, Munich, Germany.
20. Molecular Epidemiology and Cardiology, Clinical Sciences, Lund University, Lund, Sweden.
21. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA.
22. Department of Epidemiology, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands.
23. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany.
24. DZHK (German Centre for Cardiovascular Research), partner site: Greifswald, Germany.
25. Department of Cardiology, Leiden University Medical Center, The Netherlands.
26. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, UK.
27. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.
28. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
29. The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
30. Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Germany.
31. Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere, Finland.
32. Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-University, München, Germany.
33. Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of Sao Paulo, Sao Paulo, Brazil.
34. Stanford University, Stanford, CA, USA.
35. Heart and Lung Center HUS, Helsinki University Central Hospital, Helsinki, Finland.
36. University Hospital Basel, Switzerland.
37. Cardiovascular Research Institute Basel, Switzerland.
38. Population Health Research Institute, Hamilton, Canada.
39. Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Canada.
40. Institute for Translational Genomics and Population Sciences, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA.
41. Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA, USA.
42. Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany.
43. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany.
44. Division of Cardiovascular Medicine, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.
45. Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC, USA.
46. CTSU - Nuffield Department of Population Health, University of Oxford, Oxford, UK.
47. Inserm Center U1219 (Bordeaux Population Health Centre), Bordeaux, France.
48. University of Bordeaux, Bordeaux, France.

49. Department of Neurology, Bordeaux University Hospital, Bordeaux, France.
50. Department of Neurology, Boston University School of Medicine, Boston, MA, USA.
51. Biostatistics Department, School of Public Health, Boston University, Boston, MA, USA.
52. University of Virginia Health System, Departments of Neurology and Public Health Science, Charlottesville, VA, USA.
53. RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.
54. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
55. Department of Epidemiology and Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA.
56. Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA.
57. Department of Health Services, University of Washington, Seattle, WA, USA.
58. Department of Medicine, Boston University School of Medicine, Boston, MA, USA.
59. Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, MD, USA.
60. Institute of Cardiovascular and Medical Sciences, BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, UK.
61. Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA, USA.
62. Cardiovascular Health Research Unit, Departments of Medicine and Biostatistics, University of Washington, Seattle, WA, USA.
63. Research unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
64. Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
65. Institute for Medical Informatics, Biometry, and Epidemiology, University Hospital, University Duisburg-Essen, Germany.
66. Department of Clinical Physiology, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland.
67. University Institute of Clinical Chemistry, University of Bern, Switzerland and labormedizinisches zentrum Dr. Risch, Schaan, Liechtenstein.
68. Heart Institute, University of Sao Paulo, Sao Paulo, Brazil.
69. German Center for Diabetes Research, Neuherberg, Germany.
70. Division of Population Health Sciences, University of Dundee, Scotland, UK.
71. Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden.
72. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
73. Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
74. Department of Cardiology, Heart Hospital, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland.
75. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
76. Division of Cardiology, Hamilton Health Sciences, McMaster University, Hamilton, Ontario, Canada.
77. Boston VA Research Institute, Inc., Boston, MA, USA.
78. Department of Biostatistics, University of Liverpool, Liverpool, UK.
79. Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK.
80. Department of Epidemiology, University of Washington, Seattle, WA, USA.
81. Cardiovascular Division, Department of Medicine, University of Minnesota Medical School, Minneapolis, MN, USA.
82. Division of Cardiology, University of Alberta, Edmonton, Canada.
83. Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA.
84. Division of Nephrology & Hypertension, Internal Medicine, School of Medicine, University of Utah, UT, USA.
85. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.
86. Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden.
87. Acute Stroke Services, Massachusetts General Hospital, Boston, MA, USA.
88. Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, UK.
89. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
90. Musculoskeletal Research Programme, Division of Applied Medicine, University of Aberdeen, Aberdeen, UK.
91. Transplantation Laboratory, Medicum, University of Helsinki, Helsinki, Finland.
92. Department of Clinical Sciences, Lund University, Malmö, Sweden.
93. Cardiovascular Genetics and Genomics Group, Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.
94. University of Illinois, Chicago, IL, USA.
95. Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, UK.
96. Institute of Human Genetics, University of Bonn, Germany.
97. Department of Genomics, Life & Brain Research Center, University of Bonn, Germany.

98. Cardiovascular Health Research Unit, University of Washington Medical Center, Seattle, WA, USA.
99. Epidemiological Cardiology Research Center (EPICARE), Wake Forest School of Medicine, Winston Salem, NC, USA.
100. Department of Epidemiology and Internal Medicine, Erasmus University Medical Center Rotterdam, the Netherlands.
101. Munich Cluster for Systems Neurology (SyNergy), München, Germany.
102. German Center for Neurodegenerative Diseases (DZNE), Munich, Germany.
103. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA.
104. Department of Internal Medicine, Clinical Sciences, Lund University, Malmö, Sweden.
105. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
106. Department of Cardio-Thoracic Surgery, Heart Hospital, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland.
107. Laboratory of Genetics and Molecular Biology, Heart Institute, University of Sao Paulo, Sao Paulo, Brazil and Department of Genetics, Harvard Medical School, Boston, MA, USA.
108. Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands.
109. Interuniversity Cardiology Institute of the Netherlands, Utrecht, The Netherlands.
110. Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA.
111. Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria and Synlab Academy, Synlab Services GmbH, Mannheim, Germany.
112. Department of Epidemiology and Internal Medicine, Erasmus University Medical Center Rotterdam, the Netherlands and Inspectorate of Health Care, Utrecht, the Netherlands.
113. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA.
114. Divisions of Preventive Medicine and Genetics, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA.
115. Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA.
116. Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.
117. Department of Human Genetics and Disease Diversity, Tokyo Medical and Dental University Graduate School of Medical and Dental Sciences, Tokyo, Japan.
118. Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, MA, USA.

Appendix II

Members of the International Stroke Genetics Consortium (ISGC)

Please note that ISGC participants evolve over time. Further information on the ISGC can be found at <http://www.strokegenetics.org/>.

Sylvia Smoller, PhD ¹	Joanna Pera, MD ²⁹	Andreas Gschwendtner, MD, PhD ³⁵
John Sorkin, MD ²	Agnieszka Slowik, MD ²⁹	Zhen Deng, MD ⁴⁹
Xingwu Wang, MD ³	Karol Jozef Juchniewicz, MD ²⁹	Taura Barr, MD ⁵⁰
Magdy Selim, MD, PhD ⁴	Christopher Levi, MD ³⁰	Katrina Gwinn, MD ⁵⁰
Aleksandra Pikula, MD, PhD ⁵	Paul Nyquist, MD, PhD ³¹	Roderick Corriveau, MD ⁵⁰
Philip Wolf, MD, PhD ⁵	Iscia Cendes, MD ³²	Andrew Singleton, MD, PhD ⁵⁰
Stephanie Debette, MD ⁵	Norberto Cabral, MD ³²	Salina Waddy, MD ⁵⁰
Sudha Seshadri, MD ⁵	Paulo Franca, MD ³²	Lenore Launer, MD ⁵⁰
Paul de Bakker, PhD ⁶	Anderson Goncalves, MD ³²	Christopher Chen, MD ⁵¹
Sara L. Pulit, PhD ⁶	Lina Keller, MD ³³	Kim En Le, MD ⁵¹
Daniel Chasman, MD ⁷	Milita Crisby, MD ³³	Wei Ling Lee, MD ⁵¹
Kathryn Rexrode, MD ⁷	Konstantinos Kostulas, MD ³³	Eng King Tan, MD ⁵¹
Ida Chen, MD ⁸	Robin Lemmens, MD ³⁴	Akintomi Olugbodi, MD ⁵²
Jerome Rotter, MD ⁸	Kourosh Ahmadi, MD ³⁴	Peter Rothwell, MD, PhD ⁵³
May Luke, MD ⁹	Christian Opherk, MD ³⁵	Sabrina Schilling, MD ⁵⁴
Michelle Sale, MD ¹⁰	Marco Duering, MD ³⁵	Vincent Mok, MD ⁵⁵
Tsong-Hai Lee, MD ¹¹	Martin Dichgans, MD ³⁵	Elena Lebedeva, MD ⁵⁶
Ku-Chou Chang, MD ¹¹	Rainer Malik, PhD ³⁵	Christina Jern, MD ⁵⁷
Mitchell Elkind, MD, MS ¹²	Mariya Gonik, MD ³⁵	Katarina Jood, MD ⁵⁷
Larry Goldstein, MD, PhD ¹³	Julie Staals, MD ³⁶	Sandra Olsson, MD ⁵⁷
Michael Luke James, MD ¹³	Olle Melander, MD, PhD ³⁷	Helen Kim, MD ⁵⁸
Monique Breteler, MD ¹⁴	Philippe Burri, MD ³⁷	Chaeyoung Lee, MD ⁵⁹
Chris O'Donnell, MD ¹⁵	Ariane Sadr-Nabavi, MD ³⁸	Laura Kilarski, MD ⁶⁰
Didier Leys, MD ¹⁶	Javier Romero, MD, PhD ³⁹	Hugh Markus, MD ⁶⁰
Cara Carty, MD ¹⁷	Alessandro Biffi, MD ³⁹	Jennifer Peycke, MD ⁶⁰
Chelsea Kidwell, MD ¹⁸	Chris Anderson, MD ³⁹	Steve Bevan, PhD ⁶⁰
Jes Olesen, MD ¹⁹	Guido Falcone, MD ³⁹	Wayne Sheu, MD ⁶¹
Pankaj Sharma, MD, PhD ²⁰	Bart Brouwers, MD ³⁹	Hung Yi Chiou, MD ⁶²
Stephen Rich, MD, PhD ²¹	Jonathan Rosand, MD, MSc ³⁹	Joseph Chern, MD ⁶²
Turgot Tatlisumak, MD ²²	Natalia Rost, MD, MSc ³⁹	Elias Giraldo, MD ⁶³
Olli Hapola, MD ²²	Rose Du, MD ³⁹	Muhammad Taqi, MD ⁶³
Philippe Bijlenga, MD ²³	Christina Kourkoulis, BA ³⁹	Vivek Jain, MD ⁶⁴
Carolina Soriano, MD ²⁴	Thomas Battey, BA ³⁹	Olivia Lam, MD ⁶⁵
Eva Giralt, MD ²⁴	Steven Lubitz, MD, PhD ³⁹	George Howard, MD ⁶⁶
Jaume Roquer, MD ²⁴	Bertram Mueller-Myhsok, MD ⁴⁰	Daniel Woo, MD ⁶⁷
Jordi Jimenez-Conde, MD ²⁴	James Meschia, MD ⁴⁰	Steven Kittner, MD ⁶⁸
Ioana Cotlarcius, MD ²⁵	Thomas Brott, MD, PhD ⁴¹	Braxton Mitchell, PhD, MPH ⁶⁸
John Hardy, MD ²⁶	Guillaume Pare, MD ⁴²	John Cole, MD ⁶⁸
Michal Korostynski, MD ²⁷	Alexander Pichler, MD ⁴³	Jeff O'Connell, MD ⁶⁸
Giorgio Boncoraglio, MD ²⁸	Christian Enzinger, MD ⁴³	Dianna Milewicz, MD ⁶⁹
Elena Ballabio, MD ²⁸	Helena Schmidt, MD ⁴³	Kachikwu Illoh, MD ⁷⁰
Eugenio Parati, MD ²⁸	Reinhold Schmidt, MD ⁴³	Bradford Worrall, MD ²¹
Adamski Mateusz, MD ²⁹	Stephan Seiler, MD ⁴³	Colin Stine, MD ⁷⁰
Andrzej Urbanik, MD ²⁹	Susan Blanton, MD ⁴⁴	Bartosz Karaszewski, MD ⁷¹
Tomasz Dzedzic, MD ²⁹	Yoshiji Yamada, MD ⁴⁵	David Werring, MD ⁷¹
Jeremiasz Jagiella, MD ²⁹	Anna Bersano, MD ⁴⁶	Reecha Sofat, MD ⁷¹
Jerzy Gasowski, MD ²⁹	Tatjana Rundek, MD ⁴⁷	June Smalley, MD ⁷¹
Marcin Wnuk, MD ²⁹	Ralph Sacco, MD ⁴⁷	Arne Lindgren, MD ⁷²
Rafael Olszanecki, MD ²⁹	Yu-Feng Yvonne Chan, MD ⁴⁸	Bjorn Hansen, BA ⁷²

Bo Norrving, MD ⁷²	Cathie Sudlow, MD ⁸³	Anne-Katrin Giese, MD ⁹³
Gustav Smith, MD ⁷²	Kristiina Rannikmae, MD ⁸³	Myriam Fornage, MD ⁹⁴
Juan Jose Martin, MD ⁷³	Scott Silliman, MD ⁸⁴	Jennifer Majersik, MD ⁹⁵
Vincent Thijs, MD ⁷⁴	Caitrin McDonough, MD ⁸⁴	Mary Cushman, MD ⁹⁶
Karin Klijn, MD ⁷⁵	Matthew Walters, MD ⁸⁵	Keith Keene, MD ⁹⁷
Femke van't Hof, MD, PhD ⁷⁵	Annie Pedersen, MD ⁸⁶	Siiri Bennett, MD ⁹⁸
Ale Algra, MD ⁷⁵	Kazuma Nakagawa, MD ⁸⁷	David Tirschwell, MD, MSc ⁹⁸
Mary Macleod, MD ⁷⁶	Christy Chang, MD ⁸⁸	Bruce Psaty, MD ⁹⁸
Rodney Perry, MD ⁷⁷	Mark Dobbins, MD ⁸⁸	Alex Reiner, MD ⁹⁸
Donna Arnett, MD ⁷⁷	Patrick McArdle, PhD ⁸⁸	Will Longstreth, MD ⁹⁹
Alessandro Pezzini, MD ⁷⁸	Yu-Ching Chang, MD ⁸⁸	David Spence, MD ¹⁰⁰
Alessandro Padovani, MD ⁷⁸	Robert Brown, MD ⁸⁹	Joan Montaner, MD ¹⁰¹
Steve Cramer, MD, PhD ⁷⁹	Devin Brown, MD ⁸⁹	Israel Fernandez-Cadenas, MD ¹⁰²
Mark Fisher, MD ⁷⁹	Elizabeth Holliday, MD ⁹⁰	Carl Langefeld, MD ¹⁰²
Danish Saleheen, MD ⁸⁰	Raj Kalaria, MD ⁹¹	Cheryl Bushnell, MD ¹⁰²
Joseph Broderick, MD ⁸¹	Jane Maguire, MD ⁹¹	Laura Heitsch, MD ¹⁰³
Brett Kissela, MD ⁸¹	John Attia, MD ⁹¹	Jin-Moo Lee, MD, PhD ¹⁰³
Alex Doney, MD ⁸²	Martin Farrall, MD ⁹²	Kevin Sheth, MD ¹⁰⁴

ISGC Consortium Member Affiliations

1. Albert Einstein College of Medicine, Bronx, NY, USA
2. Baltimore VA Medical Center, Baltimore, MD, USA
3. Beijing Hypertension League Institute, Beijing, China
4. Beth Israel Deaconess Medical Center, Boston, MA, USA
5. Boston University Medical Center, Boston, MA, USA
6. University Medical Center Utrecht, Utrecht, The Netherlands
7. Brigham and Women's Hospital, Boston, MA, USA
8. Cedars Sinai Medical Center, Los Angeles, CA, USA
9. Celera, Alameda, CA, USA
10. University of Virginia, Charlottesville, VA, USA
11. Chang Gung Memorial Hospital, Linkou Medical Center, Guishan District, Taoyuan City, Taiwan
12. Columbia University, New York, NY, USA
13. Duke University, Durham, NC, USA
14. Erasmus University, Rotterdam, Zuid Holland, The Netherlands
15. Framingham Heart Study, Framingham, MA, USA
16. Université du Droit et de la Santé Lille, Lille, France
17. Fred Hutchinson Cancer Research Center, Seattle, WA, USA
18. Georgetown University, Georgetown, MD, USA
19. Glostrup Hospital, Glostrup, Denmark
20. Hammersmith Hospitals & Imperial College London, London, UK
21. University of Virginia Health System, Charlottesville, VA, USA
22. Helsinki University Central Hospital, Helsinki, Finland
23. Hôpitaux Universitaires de Genève, Geneva, Switzerland
24. IMIM-Hospital del Mar, Barcelona, Spain
25. Imperial College London, London, UK
26. Institute of Neurology, University College London, London, UK
27. Institute of Pharmacology, Krakow, Poland
28. IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
29. Jagiellonian University, Krakow, Poland
30. John Hunter Hospital, University of Newcastle, Newcastle, New-South-Wales, Australia
31. Johns Hopkins School of Medicine, Baltimore, MD, USA
32. Joinville Biobank, Joinville, Brazil
33. Karolinska Institutet, Karolinska, Sweden
34. Leuven University, Leuven, Belgium
35. Ludwig-Maximilians-Universität, München, Germany

36. Maastricht University Medical Centre, Maastricht, the Netherlands
37. Malmö University Hospital, Malmö, Sweden
38. Mashhad University of Medical Sciences, Mashhad, Iran
39. Massachusetts General Hospital, Boston, MA, USA
40. Max Planck Institute of Psychiatry, Munich, Germany
41. Mayo Clinic, Rochester, MN, USA
42. McMaster University, Hamilton, Canada
43. Medical University Graz, Graz, Austria
44. Miami Institute of Human Genomics, University of Miami Miller School of Medicine, Miami, FL
45. Mie University, Tsu, Japan
46. Milan University, Milan, Italy
47. University of Miami, Miami, FL, USA
48. Mount Sinai Medical Center, Miami Beach, FL, USA
49. Nanfang Hospital, Southern Medical University, Guangdong, China
50. National Institutes of Health, Bethesda, MD, USA
51. National Neuroscience Institute, Singapore General Hospital, Singapore
52. Obafemi Awolowo University, Ile-Ife, Nigeria
53. Radcliffe Infirmary, Oxford University, Oxford, UK
54. University of Bordeaux, Bordeaux, France
55. Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong
56. Ulm University, Ulm, Germany
57. Sahlgrenska University Hospital, Gothenburg, Sweden
58. Center for Cerebrovascular Research, San Francisco General Hospital, San Francisco, CA, USA
59. Soongsil University, Seoul, South Korea
60. St. George's University of London, London, UK
61. Taichung Veterans General Hospital, Taichung City, Taiwan
62. Taipei Medical University, Taipei City, Taiwan
63. The University of Tennessee Health Science Center at Memphis, Memphis, TN, USA
64. University of California Irvine Medical Center, Irvine, CA, USA
65. University of California San Francisco, San Francisco, CA, USA
66. University of Alabama School of Public Health
67. University of Cincinnati, Cincinnati, OH, USA
68. University of Maryland School of Medicine, Baltimore, MD, USA
69. University of Texas Medical School at Houston, Houston, TX, USA
70. University of Texas-Houston, Houston, TX, MA
71. University College London, London, UK
72. University Hospital Lund, Lund, Sweden
73. University Hospital Sanatorio Allende, Cordoba, Argentina
74. University Hospital Leuven, Leuven, Belgium
75. University Medical Center Utrecht, Utrecht, The Netherlands
76. University of Aberdeen, Aberdeen, Scotland
77. University of Alabama at Birmingham School of Public Health, Birmingham, AL, USA
78. University of Brescia, Brescia, Italy
79. University of California Irvine, Irvine, CA, USA
80. University of Pennsylvania, Philadelphia, PA, USA
81. University of Cincinnati, Cincinnati, OH, USA
82. University of Dundee, Dundee, Scotland
83. University of Edinburgh, Western General Hospital, Edinburgh, Scotland
84. University of Florida, Gainesville, FL, USA
85. University of Glasgow, Glasgow, Scotland, UK
86. University of Gothenburg, Gothenburg, Sweden
87. University of Hawaii, Honolulu, HI, USA
88. University of Maryland, Baltimore, MD, USA
89. University of Newcastle, New-South-Wales, Australia
90. Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, UK
91. University of Rostock, Rostock, Germany
92. University of Texas-Houston, Health Sciences Center, Houston, TX, USA
93. University of Utah, Salt Lake City, UT, USA

94. University of Vermont and Fletcher Allen Health Care, Burlington, VT, USA
95. University of Virginia, Charlottesville, VA, USA
96. University of Washington, Seattle, WA, USA
97. University of Washington, Harborview Medical Center, Seattle, WA, USA
98. University of Western Ontario, Robarts Research Institute, Ontario, Canada
99. Vall d'Hebron Hospital, Barcelona, Spain
100. Wake Forest University, Winston-Salem, NC, USA
101. Washington University of St. Louis, St. Louis, MO, USA
102. Yale New Haven Hospital, Yale School of Medicine, Yale, CT, USA

References

1. Christophersen IE, Rienstra M, Roselli C, et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet.* 2017;57:289.
2. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet.* The American Society of Human Genetics; 2011;88:76–82.
3. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–575.
4. 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.
5. 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.* Elsevier Ltd; 2015;15:4–7.
6. 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, Peddareddygar 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.
7. Goldstein LB, Jones MR, Matchar DB, et al. Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. *Stroke*. 2001;32:1091–1098.
 8. Adams HP Jr, Birgitte H. Bendixen P, Jaap Kappelle L, et al. Classification of Subtype of Acute Ischemic Stroke: Definitions for Use in a Multicenter Clinical Trial. *Stroke*. 1993;24:35–41.
 9. McArdle PF, Kittner SJ, Ay H, et al. Agreement between TOAST and CCS ischemic stroke classification: the NINDS SiGN study. *Neurology*. 2014;83:1653–1660.
 10. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: The causative classification of stroke system. *Stroke*. 2007;38:2979–2984.
 11. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. Nature Publishing Group; 2012;44:955–959.
 12. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
 13. Francioli LC, Menelaou A, Pulit SL, et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet*. 2014;46:818–825.
 14. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47:284–290.
 15. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods*. Nature Publishing Group; 2011;8:833–835.
 16. Widmer C, Lippert C, Weissbrod O, et al. Further improvements to linear mixed models for genome-wide association studies. *Sci Rep*. 2014;4:6874.

17. 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.
18. 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.
19. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature.* Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2016;533:539.
20. Weng L-C, Preis SR, Hulme OL, et al. Genetic Predisposition, Clinical Risk Factor Burden, and Lifetime Risk of Atrial Fibrillation. *Circulation* [online serial]. Epub 2017 Nov 12. Accessed at: <http://dx.doi.org/10.1161/CIRCULATIONAHA.117.031431>.