nature genetics

LETTERS

https://doi.org/10.1038/s41588-018-0133-9

Multi-ethnic genome-wide association study for atrial fibrillation

Atrial fibrillation (AF) affects more than 33 million individuals worldwide[1] and has a complex heritability[2]. We conducted the largest meta-analysis of genome-wide association studies (GWAS) for AF to date, consisting of more than half a million individuals, including 65,446 with AF. In total, we identified 97 loci significantly associated with AF, including 67 that were novel in a combined-ancestry analysis, and 3 that were novel in a European-specific analysis. We sought to identify AF-associated genes at the GWAS loci by performing RNA-sequencing and expression quantitative trait locus analyses in 101 left atrial samples, the most relevant tissue for AF. We also performed transcriptome-wide analyses that identified 57 AF-associated genes, 42 of which overlap with GWAS loci. The identified loci implicate genes enriched within cardiac developmental, electrophysiological, contractile and structural pathways. These results extend our understanding of the biological pathways underlying AF and may facilitate the development of therapeutics for AF.

Atrial fibrillation (AF) is the most common heart rhythm disorder, and is a leading cause of heart failure and stroke[3]. Prior genome-wide association studies (GWAS) have identified at least 30 loci associated with AF[4–9]. We conducted a large-scale analysis with more than half a million participants, including 65,446 with AF, from more than 50 studies. Our AF sample was composed of 84.2% European, 12.5% Japanese, 2% African American and 1.3% Brazilian and Hispanic populations (Supplementary Table 1). We used the Haplotype Reference Consortium (HRC) reference panel to impute variants from SNP array data for 75% of the samples (Fig. 1). In the remainder, we included HRC overlapping variants from 1000 Genomes imputed data, or from a combined reference panel. We analyzed 8,328,530 common variants (minor allele frequency (MAF) > 5%), 2,884,670 low-frequency variants (1%<MAF≤5%) and 936,779 rare variants (MAF ≤1%).

The combined-ancestry meta-analysis revealed 94 AF-associated loci, 67 of which were novel at genome-wide significance (*P* value (*P*) $< 1\times10^{-8}$). This conservative threshold accounts for testing independent variants with MAF ≥0.1% using a Bonferroni correction, while use of a more commonly utilized threshold of 5×10^{-8} resulted in the identification of an additional 10 loci (Supplementary Table 2). The majority of sentinel variants ($n=92$) were common (MAF > 5%), with relative risks ranging from 1.04 to 1.55. Two low-frequency sentinel variants were identified within the genes *C1orf185* and *UBE4B* (Fig. 2, Table 1, Supplementary Table 3 and Supplementary Fig. 1).

We then conducted a gene set enrichment analysis with the results from the combined-ancestry meta-analysis using MAGENTA. We identified 55 enriched gene sets or pathways that largely fall into cardiac developmental, electrophysiological, and cardiomyocyte contractile or structural functional groups (Supplementary Table 4). In total, 48 of the 67 novel loci contain one or more genes within 500 kilobases (kb) of the sentinel variant that were part of an enriched gene set or pathway (Supplementary Fig. 2).

Next, we performed ancestry-specific meta-analyses. Among individuals of European ancestry, we identified three additional loci associated with AF, each of which had a subthreshold association ($P < 1\times10^{-6}$) in the combined-ancestry meta-analysis. These loci were located close to or within the genes *CDK6*, *EPHA3* and *GOSR2* (Supplementary Table 5 and Supplementary Figs. 3 and 4). The region most significantly associated with AF in European, Japanese and African American populations (Supplementary Figs. 5 and 6) was on chromosome 4q25, upstream of the gene *PITX2* (Supplementary Fig. 7). We did not observe significant heterogeneity of effect estimates across ancestries for most associations, suggesting that top genetic susceptibility signals for AF have a relatively constant effect across ancestries (Table 1, Supplementary Table 3 and Supplementary Fig. 8). The proportion of heritability explained by the loci from the European ancestry analysis was 42%, compared to the previously reported 25% (ref. [10] and Supplementary Table 6).

In conditional and joint analyses of the European ancestry results, we found 11 loci with multiple, independent AF-associated signals. At a locus centered on a cluster of sodium-channel genes, we identified three regions that independently associate with AF within *SCN10A*, *SCN5A* and a third signal between both genes. At the previously described *TBX5* locus[8], we detected a novel independent signal close to *TBX3*. Pairwise linkage disequilibrium (LD) estimates between the independent variants at both loci were extremely low ($r^2 < 0.03$; Supplementary Table 7).

For 13 AF loci, the sentinel variant or a proxy ($r^2 > 0.6$) was a missense variant. A missense variant (rs11057401) in *CCDC92* was predicted to be damaging by four of five in silico prediction algorithms (Supplementary Table 8); and was previously associated with coronary artery disease[11]. Since most AF-associated variants reside in non-coding regions, we sought to determine whether the sentinel variants or their proxies ($r^2 > 0.6$) fell within regulatory regions in heart tissues based on chromatin states from the Roadmap Epigenomics Consortium. At 64 out of 67 novel loci, variants were located within regulatory elements (Supplementary Table 9); AF-associated loci were also significantly enriched within regulatory elements (Supplementary Fig. 9).

We then sought to link risk variants to candidate genes by assessing their effect on gene expression levels. First, since AF often arises from the pulmonary veins and left atrium (LA), we performed RNA-sequencing, genotyping and expression quantitative trait locus (eQTL) analyses in 101 human LA samples without structural heart disease from the Myocardial Applied Genomics Network repository. Second, we identified eQTLs from right atrial (RA) and left ventricular (LV) cardiac tissue from the Genotype-Tissue Expression (GTEx) project. Finally, we performed a transcriptome-wide analysis using the MetaXcan[12] method, which infers the association between genetically predicted gene expression and disease risk.

We observed eQTLs to 1 or more genes at 17 novel loci. Of the ten eQTLs detected in LA tissue, eight were also detected in RA or LV, with consistent directionality. For example, we observed that rs4484922 was an eQTL for *CASQ2* only in LA tissue. Although we detected more AF loci with eQTLs in the RA or LV data, for many of these ($n=8$) the results pointed to multiple genes per locus (Supplementary Tables 10–12). LA eQTL studies may facilitate the prioritization of candidate genes, but are currently limited by sample size.

A full list of author and affiliations appears at the end of the paper.

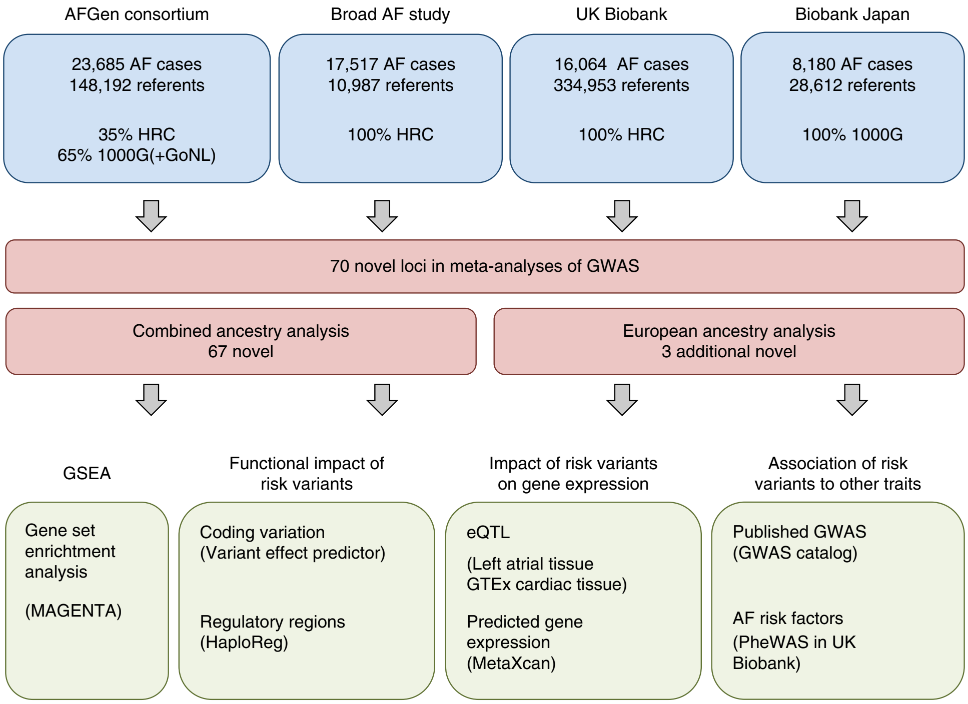

Fig. 1 | Study and analysis flowchart. Top, overview of the participating studies, number of AF cases and referents, and the percentage of samples imputed with each reference panel. Middle, summary of the primary analyses and the newly discovered loci for AF. Bottom, overview of the secondary analyses to evaluate AF risk variants and loci.

For the transcriptome-wide analyses, we used GTEx human atrial and ventricular expression data as a reference. We identified 57 genes significantly associated with AF. Of these, 42 genes were located at AF loci, whereas the remaining 15 were >500 kb from an AF sentinel variant (Supplementary Table 13 and Fig. 3). The probable candidate genes at each locus are summarized in Supplementary Table 12. For example, at the locus with the lead variant rs4484922, we observed results from all downstream analyses pointing towards the nearest gene *CASQ2*, at rs12908437 towards the gene *IGFR1*, and at rs113819537 towards the gene *SSPN*. However, for many loci, the evaluation of candidate genes remains challenging.

We then sought to assess the pleiotropic effects of the identified AF risk variants. First, we queried the NHGRI-EBI GWAS Catalog to detect associations to other phenotypes (Supplementary Table 14). Second, using the UK Biobank[13], we performed a phenome-wide association study (PheWAS) for 12 AF risk factors (Supplementary Table 15). As illustrated in Fig. 4, distinct clusters of variants were associated with AF as well as height, body mass index and hypertension. For example, we observed a pleiotropic effect at rs880315 (*CASZ1*) for blood pressure[14] and hypertension[14], which was also observed in the UK Biobank (association with hypertension, $P = 2.56 \times 10^{-34}$).

In sum, we identified a total of 97 distinct AF loci from 65,446 AF cases and more than 522,000 referents. A recent study reported 111 loci from 60,620 AF cases and more than 970,000 referents[15], including more than 18,000 AF cases from our previous report[8]. We therefore performed a preliminary meta-analysis for the top loci in non-overlapping participants from these two large efforts, with a resulting total of more than 93,000 AF cases and more than 1 million referents. In aggregate, we identified at least 134 distinct AF-associated loci (Supplementary Table 16).

Four major themes emerge from the identified AF loci. First, two AF loci contain genes that are primary targets for current antiarrhythmic medications used to treat AF. The *SCN5A* gene encodes a sodium channel in the heart, the target of sodium-channel blockers such as flecainide and propafenone. Similarly, *KCNH2* encodes the alpha subunit of the potassium channel complex, the target of potassium-channel-inhibiting medications such as amiodarone, sotalol and dofetilide. *SCN5A* and *KCNH2* have previously been implicated in AF through GWAS[8], candidate gene analysis[16] and family-based studies[17,18].

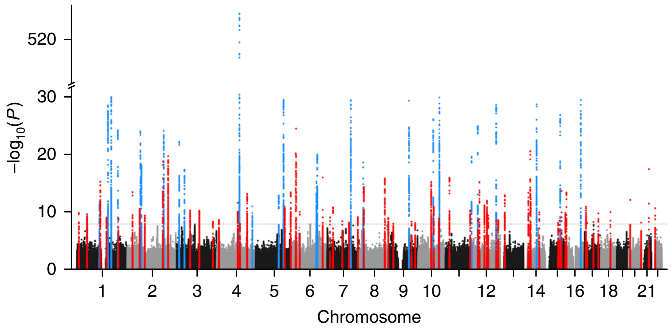

Fig. 2 | Manhattan plot of combined-ancestry meta-analysis. The plot shows 67 novel (red) and 27 known (blue) genetic loci associated with AF at a significance level of $P < 1 \times 10^{-8}$ (dotted line), for the combined-ancestry meta-analysis ($n = 588,190$). The significance level accounts for multiple testing of independent variants with MAF ≥0.1% using a Bonferroni correction. P values (two-sided) were derived from a meta-analysis using a fixed-effects model with an inverse-variance weighted approach. The y axis has a break between $-\log_{10}(P)$ of 30 and 510 to emphasize the novel loci.

Second, transcriptional regulation appears to be a key feature of AF etiology. *TBX3* and the adjacent gene *TBX5* encode transcription factors that have been shown to regulate the development of the cardiac conduction system[19]. Similarly, *NKX2-5* encodes a transcription factor that is an early cue for cardiac development and has been associated

Table 1 | Novel loci in combined-ancestry meta-analysis

Rsid	Chr	hg19	Risk/ref allele	RAF (%)	RR	95% CI	P_{META}	Nearest gene(s)[a]	Func	imp Qual	I^2_{HET}	P_{HET}
rs187585530	1	10167425	A/G	0.5	1.55	1.36–1.77	1.18×10^{-10}	*UBE4B*	Missense	0.81	0.0	1.000
rs880315	1	10796866	C/T	37.4	1.04	1.03–1.06	5.04×10^{-9}	*CASZ1*	Intronic	0.97	40.7	0.150
rs146518726	1	51535039	A/G	2.6	1.18	1.12–1.24	2.05×10^{-10}	*C1orf185*	Intronic	0.96	0.0	1.000
rs4484922	1	116310818	G/C	68.3	1.07	1.05–1.08	4.57×10^{-16}	*CASQ2*	Intronic	0.98	0.0	0.689
rs79187193	1	147255831	G/A	94.8	1.12	1.08–1.16	8.07×10^{-10}	*GJA5*	Upstream	0.97	39.8	0.190
rs4951261	1	205717823	C/A	38.2	1.05	1.03–1.06	1.17×10^{-9}	*NUCKS1*	Intronic	0.99	0.0	0.788
rs6546620	2	26159940	C/T	75.3	1.07	1.05–1.09	2.96×10^{-14}	*KIF3C*	Intronic	0.95	33.0	0.201
rs6742276	2	61768745	A/G	61.2	1.05	1.03–1.06	2.42×10^{-11}	*XPO1*	Upstream	0.99	0.0	0.731
rs72926475	2	86594487	G/A	87.0	1.07	1.05–1.10	3.49×10^{-10}	*REEP1,KDM3A*	Intergenic	0.97	38.7	0.180
rs56181519	2	175555714	C/T	74.0	1.08	1.06–1.10	1.52×10^{-19}	*WIPF1,CHRNA1*	Intergenic	0.94	0.0	0.519
rs295114	2	201195602	C/T	59.7	1.07	1.05–1.09	1.76×10^{-20}	*SPATS2L*	Intronic	1.00	21.9	0.275
rs2306272	3	66434643	C/T	31.8	1.05	1.04–1.07	4.54×10^{-11}	*LRIG1*	Missense	0.99	30.6	0.218
rs17490701	3	111587879	G/A	85.7	1.07	1.05–1.10	5.43×10^{-11}	*PHLDB2*	Intronic	0.97	46.8	0.111
rs4855075	3	179170494	T/C	14.3	1.06	1.04–1.08	4.00×10^{-9}	*GNB4*	Upstream	0.95	10.1	0.348
rs3822259	4	10118745	T/G	67.9	1.05	1.03–1.06	1.93×10^{-9}	*WDR1*	Upstream	0.96	0.0	0.922
rs3960788	4	103915618	C/T	42.4	1.05	1.04–1.07	2.09×10^{-12}	*SLC9B1*	Intronic	0.98	35.7	0.183
rs55754224	4	114428714	T/C	25.0	1.05	1.03–1.07	9.25×10^{-9}	*CAMK2D*	Intronic	0.99	0.0	0.511
rs10213171	4	148937537	G/C	8.2	1.11	1.08–1.14	6.09×10^{-14}	*ARHGAP10*	Intronic	0.96	0.0	0.584
rs174048	5	142650404	C/T	15.7	1.07	1.05–1.09	1.05×10^{-11}	*ARHGAP26,NR3C1*	Intergenic	0.99	0.0	0.852
rs6882776	5	172664163	G/A	67.2	1.06	1.05–1.08	3.18×10^{-14}	*NKX2-5*	Upstream	0.95	0.0	0.858
rs73366713	6	16415751	G/A	86.2	1.11	1.09–1.14	5.80×10^{-21}	*ATXN1*	Intronic	0.94	0.0	0.879
rs34969716	6	18210109	A/G	31.1	1.09	1.07–1.11	2.91×10^{-25}	*KDM1B*	Intronic	0.80	19.5	0.290
rs3176326	6	36647289	G/A	80.4	1.06	1.04–1.08	7.95×10^{-11}	*CDKN1A*	Intronic	0.95	0.0	0.450
rs117984853	6	149399100	T/G	8.9	1.12	1.09–1.15	8.38×10^{-17}	*UST*	Downstream	0.83	56.5	0.100
rs55734480	7	14372009	A/G	26.6	1.05	1.03–1.07	7.34×10^{-10}	*DGKB*	Intronic	0.94	0.0	0.441
rs6462078	7	28413187	A/C	74.7	1.06	1.04–1.08	1.35×10^{-11}	*CREB5*	Intronic	0.98	22.2	0.278
rs74910854	7	74110705	G/A	6.9	1.10	1.07–1.13	3.36×10^{-9}	*GTF2I*	Intronic	0.74	24.4	0.265
rs62483627	7	106856002	A/G	23.5	1.05	1.03–1.07	5.17×10^{-9}	*COG5*	Intronic	0.98	15.1	0.318
rs7789146	7	150661409	G/A	80.3	1.06	1.04–1.08	6.51×10^{-10}	*KCNH2*	Intronic	0.96	66.0	0.019
rs7846485	8	21803735	C/A	86.8	1.09	1.07–1.12	3.71×10^{-15}	*XPO7*	Intronic	0.99	0.0	0.676
rs62521286	8	124551975	G/A	6.7	1.13	1.10–1.16	1.24×10^{-16}	*FBXO32*	Intronic	0.96	0.0	0.678
rs35006907	8	125859817	A/C	32.9	1.05	1.03–1.06	2.76×10^{-9}	*MTSS1,LINC00964*	Regulatory reg.	0.97	0.0	0.542
rs6993266	8	141762659	A/G	53.8	1.05	1.03–1.06	9.73×10^{-10}	*PTK2*	Intronic	0.99	5.7	0.374
rs4977397	9	20235004	A/G	57.0	1.04	1.03–1.06	8.60×10^{-9}	*SLC24A2,MLLT3*	Intergenic	0.95	38.3	0.166
rs4743034	9	109632353	A/G	23.4	1.05	1.03–1.07	3.98×10^{-9}	*ZNF462*	Intronic	1.00	0.0	0.963
rs10760361	9	127178266	G/T	64.7	1.04	1.03–1.06	7.03×10^{-9}	*PSMB7*	Upstream	0.97	0.0	0.680
rs7919685	10	65315800	G/T	53.3	1.06	1.04–1.07	5.00×10^{-16}	*REEP3*	Intronic	1.00	49.2	0.097
rs11001667	10	77935345	G/A	22.2	1.06	1.05–1.08	1.06×10^{-11}	*C10orf11*	Intronic	0.98	26.8	0.243
rs1044258	10	103605714	T/C	66.2	1.05	1.03–1.06	1.07×10^{-9}	*C10orf76*	3′ UTR	0.98	14.0	0.325
rs1822273	11	20010513	G/A	27.1	1.07	1.05–1.09	8.99×10^{-17}	*NAV2*	Intronic	0.98	0.0	0.764
rs949078	11	121629007	C/T	27.1	1.05	1.04–1.07	4.77×10^{-11}	*SORL1,MIR100HG*	Intergenic	0.97	0.0	0.600
rs113819537	12	26348429	C/G	74.3	1.05	1.03–1.07	2.23×10^{-9}	*SSPN*	Upstream	0.98	0.0	0.597
rs12809354	12	32978437	C/T	14.7	1.08	1.06–1.11	5.48×10^{-16}	*PKP2*	Intronic	0.97	31.5	0.211
rs7978685	12	57103154	T/C	27.9	1.06	1.04–1.07	5.99×10^{-12}	*NACA*	Downstream	0.98	2.4	0.393
rs35349325	12	70097464	T/C	54.1	1.05	1.04–1.07	9.04×10^{-13}	*BEST3*	Upstream	0.96	0.0	0.863
rs11180703	12	76223817	G/A	56.0	1.05	1.03–1.06	3.58×10^{-10}	*KRR1,PHLDA1*	Intergenic	0.97	0.0	0.482
rs12810346	12	115091017	T/C	14.9	1.07	1.05–1.09	2.34×10^{-9}	*TBX5-AS1,TBX3*	Intergenic	0.84	0.0	0.428
rs12298484	12	124418674	C/T	67.4	1.05	1.03–1.06	2.05×10^{-9}	*DNAH10*	Intronic	1.00	0.0	0.973
rs9580438	13	23373406	C/T	32.5	1.06	1.04–1.07	1.01×10^{-13}	*LINC00540,BASP1P1*	Intergenic	0.98	0.0	0.485
rs28631169	14	23888183	T/C	19.9	1.07	1.05–1.09	3.80×10^{-14}	*MYH7*	Intronic	0.97	14.5	0.319

Continued

Table 1 | Novel loci in combined-ancestry meta-analysis (continued)

Rsid	Chr	hg19	Risk/ref allele	RAF (%)	RR	95% CI	P_{META}	Nearest gene(s)[a]	Func	imp Qual	I^2_{HET}	P_{HET}
rs2145587	14	32981484	A/G	28.1	1.08	1.06–1.10	2.32×10^{-21}	*AKAP6*	Intronic	0.94	0.0	0.888
rs73241997	14	35173775	T/C	16.4	1.07	1.05–1.10	1.10×10^{-13}	*SNX6,CFL2*	Intergenic	0.98	62.2	0.032
rs10873299	14	77426711	A/G	38.4	1.05	1.03–1.07	9.62×10^{-11}	*LRRC74,IRF2BPL*	Intergenic	0.96	4.4	0.381
rs62011291	15	63800013	G/A	22.9	1.05	1.04–1.07	6.14×10^{-9}	*USP3*	Intronic	0.96	0.0	0.727
rs12591736	15	70454139	G/A	82.0	1.06	1.04–1.08	2.47×10^{-9}	*TLE3,UACA*	Intergenic	0.92	0.0	0.966
rs12908004	15	80676925	G/A	15.9	1.08	1.06–1.10	1.95×10^{-14}	*LINC00927,ARNT2*	Intronic	0.96	57.4	0.052
rs12908437	15	99287375	T/C	39.2	1.05	1.03–1.06	1.25×10^{-10}	*IGF1R*	Intronic	0.98	0.0	0.818
rs2286466	16	2014283	G/A	80.9	1.07	1.05–1.09	3.53×10^{-14}	*RPS2*	Synonymous	0.92	0.0	0.882
rs8073937	17	7435040	G/A	36.6	1.05	1.04–1.07	1.02×10^{-11}	*POLR2A,TNFSF12*	Intergenic	0.96	12.3	0.335
rs72811294	17	12618680	G/C	88.7	1.07	1.05–1.09	6.87×10^{-9}	*MYOCD*	Intronic	0.95	32.3	0.206
rs242557	17	44019712	G/A	61.3	1.04	1.03–1.06	4.35×10^{-9}	*MAPT*	Intronic	0.94	62.1	0.032
rs7219869	17	68337185	G/C	43.9	1.05	1.03–1.06	1.49×10^{-10}	*KCNJ2,CASC17*	Intergenic	0.99	16.1	0.312
rs9953366	18	46474192	C/T	65.5	1.05	1.04–1.07	9.03×10^{-11}	*SMAD7*	Intronic	0.93	0.0	0.565
rs2145274	20	6572014	A/C	91.3	1.11	1.08–1.14	6.97×10^{-13}	*CASC20,BMP2*	Regulatory reg.	0.96	19.0	0.295
rs7269123	20	61157939	C/T	58.5	1.05	1.03–1.06	5.59×10^{-9}	*C20orf166*	Intronic	0.85	68.7	0.012
rs2834618	21	36119111	T/G	89.8	1.12	1.09–1.14	2.93×10^{-18}	*LOC100506385*	Intronic	0.93	21.6	0.277
rs465276	22	18600583	G/A	61.5	1.05	1.04–1.07	1.84×10^{-11}	*TUBA8*	Intronic	0.90	0.0	0.654

Sentinel variants at novel genetic loci associated with AF at a significance level of $P<1\times10^{-8}$, for the combined-ancestry meta-analysis ($n=588{,}190$). The significance level accounts for multiple testing of independent variants with MAF ≥ 0.1% using a Bonferroni correction. P_{META} (two-sided) was derived from a meta-analysis using a fixed-effects model with an inverse-variance weighted approach. P_{HET} was derived from a Cochran's Q-test (two-sided) for heterogeneity. Chr, chromosome; CI, confidence interval; Func, functional consequence (most severe consequence by variant effect predictor); HET, heterogeneity; I^2, I-square; impQual, average imputation quality; META, meta-analysis; P, P value; RAF, risk allele frequency; reg, region; ref, reference; RR, relative risk. [a]Reported is either the gene that overlaps with the sentinel variant or the nearest gene(s) up- and downstream of the sentinel variant (separated by a comma).

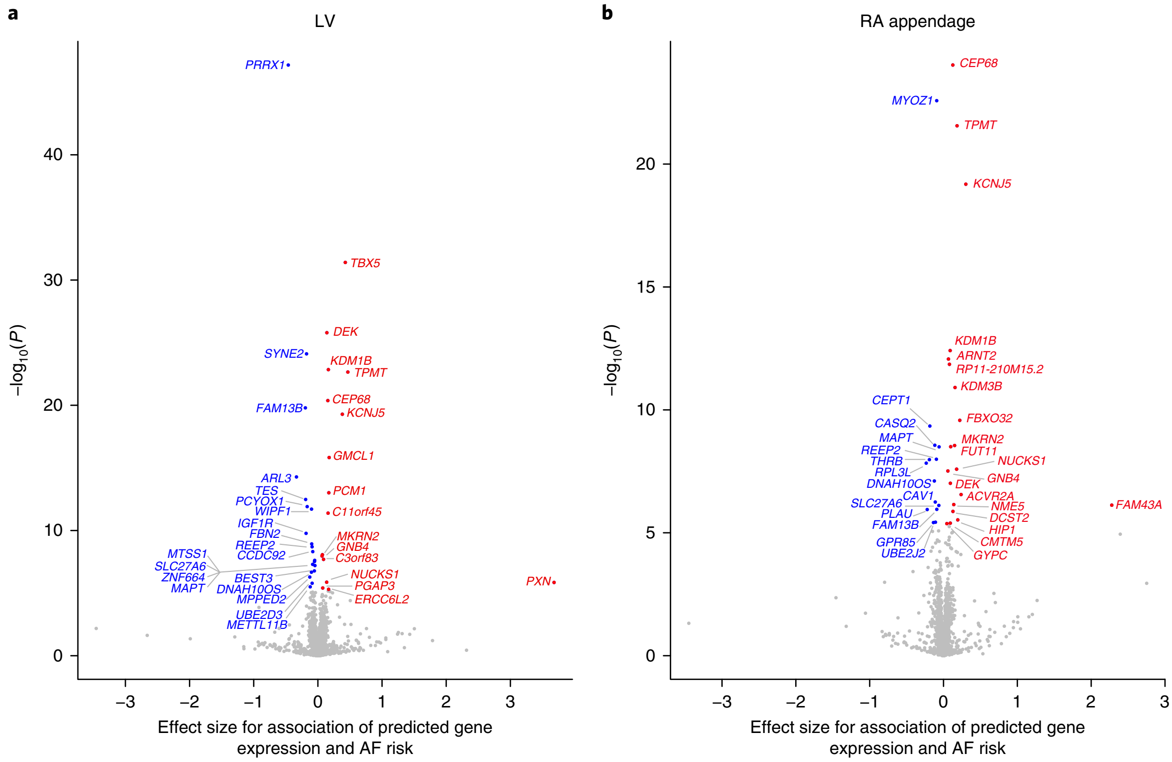

Fig. 3 | Volcano plot of transcriptome-wide analysis from human heart tissues. a,**b**, Plots showing the results from the transcriptome-wide analysis based on LV (**a**, $n=190$) and RA appendage (**b**, $n=159$) tissue from GTEx, calculated with the MetaXcan method based on the combined-ancestry summary level results ($n=588{,}190$). Each plotted point represents the association results for an individual gene. The x axis shows the effect size for associations of predicted gene expression and AF risk for each tested gene. The y axis shows the $-\log_{10}(P)$ for the associations per gene. Genes with a positive effect (red) showed an association of increased predicted gene expression with AF risk. Genes with a negative effect (blue) showed an association of decreased predicted gene expression with AF risk. The highlighted genes are significant after Bonferroni correction for all tested genes and tissues with a P value $<5.36\times10^{-6}$. The result for one gene for the RA appendage (**b**) is not shown (*SNX4*, effect = 6.94, $P=0.2$).

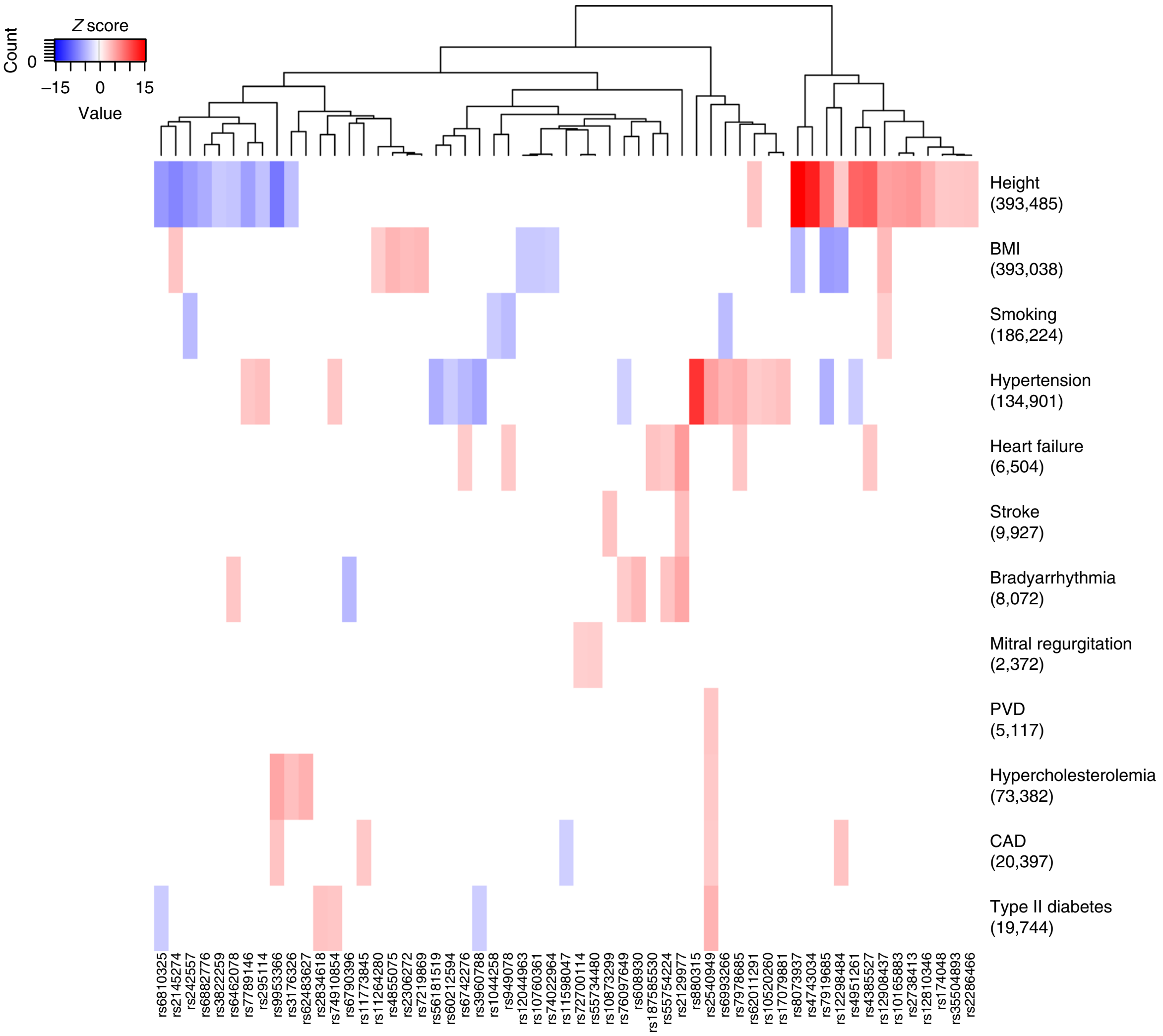

Fig. 4 | Cross-trait associations of AF risk variants with AF risk factors in the UK Biobank. The heatmap shows associations of novel and known sentinel variants at AF risk loci from the combined-ancestry meta-analysis. Shown are variants and phenotypes with significant associations after correcting for 12 phenotypes via Bonferroni correction with $P < 4.17 \times 10^{-3}$. P values (two-sided) were derived from linear and logistic regression models. Listed next to each trait is the number of cases for binary traits or the total sample size for quantitative traits. Hierarchical clustering was performed on a variant level using the complete linkage method based on Euclidian distance. Coloring represents Z scores for each respective trait or disease, oriented toward the AF risk allele. Red indicates an increase in the trait or disease risk while blue indicates a decrease in the trait or disease risk. BMI, body mass index; CAD, coronary artery disease; PVD, peripheral vascular disease.

with congenital heart disease[20] and heart rate[21] (Supplementary Table 14). Further, reduced function of the transcription factor encoded by *PITX2* has been associated with AF, shortening of the left atrial action potential and modulation of sodium-channel-blocker therapy in the adult LA[22–24]. A transcriptional co-regulatory network governed by transcription factors encoded by *TBX5* and *PITX2* has been shown to be critical for atrial development[25].

Third, the transcriptome-wide analyses revealed a number of compelling findings. Decreased expression of *PRRX1* associated with AF, a result consistent with findings where reduction of *PRRX1* in zebrafish and stem cell-derived cardiomyocytes was associated with action potential shortening[26]. Further, increased expression of *TBX5* and *KCNJ5* was associated with AF, a finding consistent with gain-of-function mutations in *TBX5* reported in a family with Holt–Oram syndrome and a high penetrance of AF[27]. Similarly, *KCNJ5* encodes a potassium channel that underlies a component of the I_{KAch} current, a channel that is upregulated in AF. Thus, previous studies support both the role of *PRRX1*, *TBX5* and *KCNJ5* in AF and the observed directionality.

Fourth, many of the novel loci implicate genes that underlie Mendelian forms of arrhythmia syndromes. Mutations in *CASQ2* lead to catecholaminergic polymorphic ventricular tachycardia[28,29]. Pathogenic variants in *PKP2* impair cardiomyocyte communication and structural integrity, and are a common cause of arrhythmogenic right ventricular cardiomyopathy[30,31]. Mutations in *GJA5*, *KCNH2*, *SCN5A*, *KCNJ2*, *MYH7* and *NKX2-5* have been mapped in a variety of inherited arrhythmia, cardiomyopathy or conduction system diseases[32]. Our observations highlight the pleiotropy of variation in genes specifying cardiac conduction, morphology and function, and underscore the complex, polygenic nature of AF.

In conclusion, we conducted the largest AF meta-analysis to date and report a more than threefold increase in the number of loci associated with this common arrhythmia. Our results lay the groundwork for functional evaluations of genes implicated by AF risk loci. Our findings also broaden our understanding of biological pathways involved in AF and may facilitate the development of therapeutics for AF.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0133-9.

Received: 27 October 2017; Accepted: 18 April 2018;

References

1. Chugh, S. S. et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 study. *Circulation* **129**, 837–847 (2014).
2. Lubitz, S. A. et al. Association between familial atrial fibrillation and risk of new-onset atrial fibrillation. *JAMA* **304**, 2263–2269 (2010).
3. January, C. T. et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: executive summary. *J. Am. Coll. Cardiol.* **64**, 2071–2104 (2014).
4. Benjamin, E. J. et al. Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat. Genet.* **41**, 879–881 (2009).
5. Ellinor, P. T. et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat. Genet.* **44**, 670–675 (2012).
6. Sinner, M. F. et al. Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation* **130**, 1225–1235 (2014).
7. Ellinor, P. T. et al. Common variants in KCNN3 are associated with lone atrial fibrillation. *Nat. Genet.* **42**, 240–244 (2010).
8. Christophersen, I. E. et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* **49**, 946–952 (2017).
9. Low, S.-K. et al. Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat. Genet.* **49**, 953–958 (2017).
10. Weng, L.-C. et al. Heritability of atrial fibrillation. *Circ. Cardiovasc. Genet.* **10**, e001838 (2017).
11. Klarin, D. et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat. Genet.* **49**, 1392–1397 (2017).
12. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
13. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
14. Lu, X. et al. Genome-wide association study in Chinese identifies novel loci for blood pressure and hypertension. *Hum. Mol. Genet.* **24**, 865–74 (2015).
15. Nielsen, J. B. et al. Genome-wide association study of 1 million people identifies 111 loci for atrial fibrillation. Preprint at https://www.biorxiv.org/content/early/2018/01/04/242149 (2018).
16. Sinner, M. F. et al. The non-synonymous coding IKr-channel variant KCNH2-K897T is associated with atrial fibrillation: results from a systematic candidate gene-based analysis of KCNH2 (HERG). *Eur. Heart J.* **29**, 907–914 (2008).
17. Olson, T. M. et al. Sodium channel mutations and susceptibility to heart failure and atrial fibrillation. *JAMA* **293**, 447–454 (2005).
18. McNair, W. P. et al. SCN5A mutation associated with dilated cardiomyopathy, conduction disorder, and arrhythmia. *Circulation* **110**, 2163–2167 (2004).
19. van Weerd, J. H. et al. A large permissive regulatory domain exclusively controls Tbx3 expression in the cardiac conduction system. *Circ. Res.* **115**, 432–441 (2014).
20. Schott, J. J. et al. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108–101 (1998).
21. den Hoed, M. et al. Identification of heart rate–associated loci and their effects on cardiac conduction and rhythm disorders. *Nat. Genet.* **45**, 621–631 (2013).
22. Kirchhof, P. et al. PITX2c is expressed in the adult left atrium, and reducing Pitx2c expression promotes atrial fibrillation inducibility and complex changes in gene expression. *Circ. Cardiovasc. Genet.* **4**, 123–133 (2011).
23. Wang, J. et al. Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proc. Natl. Acad. Sci. USA* **107**, 9753–9758 (2010).
24. Syeda, F. et al. PITX2 modulates atrial membrane potential and the antiarrhythmic effects of sodium-channel blockers. *J. Am. Coll. Cardiol.* **68**, 1881–1894 (2016).
25. Nadadur, R. D. et al. Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Sci. Transl. Med.* **8**, 354ra115 (2016).
26. Tucker, N. R. et al. Diminished PRRX1 expression is associated with increased risk of atrial fibrillation and shortening of the cardiac action potential. *Circ. Cardiovasc. Genet.* **10**, e001902 (2017).
27. Postma, A. V. et al. A gain-of-function TBX5 mutation is associated with atypical Holt–Oram syndrome and paroxysmal atrial fibrillation. *Circ. Res.* **102**, 1433–1442 (2008).
28. Lahat, H. et al. A missense mutation in a highly conserved region of CASQ2 is associated with autosomal recessive catecholamine-induced polymorphic ventricular tachycardia in Bedouin families from Israel. *Am. J. Hum. Genet.* **69**, 1378–1384 (2001).
29. Lahat, H. et al. Autosomal recessive catecholamine- or exercise-induced polymorphic ventricular tachycardia: clinical features and assignment of the disease gene to chromosome 1p13-21. *Circulation* **103**, 2822–2827 (2001).
30. Corrado, D., Link, M. S. & Calkins, H. Arrhythmogenic right ventricular cardiomyopathy. *N. Engl. J. Med.* **376**, 61–72 (2017).
31. Gerull, B. et al. Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat. Genet.* **36**, 1162–1164 (2004).
32. Ackerman, M. J. et al. HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies. *Heart Rhythm* **8**, 1308–1339 (2011).

Acknowledgements

A full list of acknowledgments appears in the Supplementary Notes.

Author contributions

C.R., M.D.C., E.J.B., K.L.L., S.A.L., P.T.E. and H.L. drafted and finalized the manuscript. H.J.C., E.A.D., B.L.K., B. Weijs, S. Kääb, M.M.-N., B.N., K.S., M.F.S., J.L., A.A., L.Y.C., K.L., S.A., D.C., G.P., L. Risch, S. Thériault, T.T., C. Schurman, S.A.S., J.C.D., D.M.R., Q.S.W., C.R., M.D.C., K.G.A., B.R.D., N.G., S. Kathiresan, L.M., P.L.H., J.B., M.K.C., J.D.S., H. Sun, D.R.V.W., T.M.B., J.C.B., J.A.B., G.A., M.S.O., L. Refsgaard, J.H.S., D.F., R.J., S. Shah, P.K., R.B.S., T.E., M.T.-L., E.J.B., B. Wang, K.L.L., M. Kähönen, T.L., I.E.C., I.C.V.G., B.G., M. Rienstra, J.E.S., P.V.D.H., N.V., H.L.B., S.C.D., R.G., B.L., S. Saba, A.A.S., R.W., H.C., R.N.L., N.L.S., K.L.W., S.R.H., B.M.P., N.S., J. Carlquist, M.J.C., S. Knight, M.E.K., W.M., P.A., O.M., M.O.-M., X.G., H.J.L., J.I.R., K.D.T., S.H.C., N.R.T., S.A.L., P.T.E., C.N.-C., M.A.R., C.D.A., P.N., J.J.G.S., H. Schunkert, T.P.C., K.B.M., I.F., J.J.W.J., P.W.M., R.N., S. Trompet, O.H.F., A. Hofman, M. Kavousi, M.N.N., B.H. Stricker, A.G.U., R.P.G., J.J.-C., S.L.P., S.M., A. Hamsten, J.P.K., G.M.M., C.R.P., A.P.M., S.G., E. Ingelsson, H.L., D.D., J.A.M., M.M.B.S., Z.T.Y., C. Shaffer, P.E.W., C.M.A., D.I.C., R.K.S., J.W., M. Dichgans and R.M. contributed to and revised the manuscript. H.J.C., E.A.D., B.L.K., B. Weijs, S. Kääb, M.M.-N., B.N., K.S., M.F.S., V.G., T.B.H., L.J.L., A.V.S., M.E., J. Hernesniemi, J.L., I.S., A.A., D.E.A., N.A.B., E.B., L.Y.C., M.L., E.Z.S., S.A., D.C., G.P., L. Risch, S. Thériault, K.I., Y.K., M. Kubo, S.-K.L., T.T., E.B.B., R.J.F.L., Y.L., C. Schurman, S.A.S., J.C.D., D.M.R., Q.S.W., C.R., M.D.C., L.-C.W., K.G.A., N.G., S. Kathiresan, L.M., P.L.H., J.B., M.K.C., J.D.S., H. Sun, D.R.V.W., T.M.B., J.C.B., J.A.B., M.-L.L., J. Sinisalo, E.V., G.A., M.S.O., L. Refsgaard, J.H.S., D.F., R.J., A. Sun, P.K., H.O., R.B.S., T.Z., T.E., M.T.-L., E.J.B., B. Wang, K.L.L., M. Kähönen, T.L., L.-P.L., K.N., I.E.C., A. Tveit, B.G., J.E.S., N.V., H.L.B., S.C.D., R.G., B.L., S. Saba, A.A.S., R.W., A.C., C.H., L.J.H., J. Huffman, S.P., D.P., B.H. Smith, H.C., E. Ipek, S.N., R.N.L., N.L.S., K.L.W., S.R.H., B.M.P., N.S., J. Carlquist, M.J.C., S. Knight, E.-K.C., H.E.L., H.-N.P., J. Shim, P.-S.Y., G.D., J. Huang, M.E.K., P.A., O.M., M.O.-M., Y.-D.C., X.G., K.D.T., J.Y., S.A.L., P.T.E., C.N.-C., M.A.R., J.R., N.R., C.D.A., P.N., J.J.G.S., A.K., T.K., H. Schunkert, L.Z., T.P.C., S.M.D., K.B.M., M.P.M., D.J.R., I.F., J.J.W.J., S. Trompet, O.H.F., A. Hofman, M. Kavousi, M.N.N., B.H. Stricker, A.G.U., M. Dörr, S.B.F., A. Teumer, U.V., S.W., J.W.C., R.P.G., J.J.-C., P.K.-W., J.P., S.L.P., M. Ribasés, A. Slowik, D.W., B.B.W., A.R.V.R.H., J.E.K., A.J.M., A.P., S.M., A.N., A. Hamsten, P.K.M., N.L.P., J.P.K., G.M.M., C.R.P., J. Cook, L.L., C.M.L., A.M., A.P.M., S.G., E. Ingelsson, N.E., K.T., H.L., D.D.M., D.D., J.A.M., M.M.B.S., Z.T.Y., C. Shaffer, P.E.W., C.M.A., D.I.C., P.M.R., M. Dichgans and R.M. contributed to study-specific GWAS by providing phenotype data or performing data analyses. C.R., M.D.C. and S.L.P. performed meta-analyses. N.R.T., P.T.E., T.P.C., K.B.M., M.P.M. and H.L. contributed samples sequencing or performed left atrial eQTL analyses. C.R., M.D.C., L.-C.W., K.L.L., S.H.C., N.R.T. and H.L. performed downstream analyses. K.I., T.T., K.L.L., S.R.H., S.A.L. and P.T.E. conceived designed and supervised the overall project.

Competing interests

P.T.E is the PI on a grant from Bayer to the Broad Institute focused on the genetics and therapeutics of AF. B.M.P. serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. P.K. receives research support from the European Union, the British Heart Foundation, the Leducq Foundation, the Medical Research Council (UK) and the German Centre for Cardiovascular Research, and from several drug and device companies active in AF, and has received honoraria from several such companies. P.K. is also listed as an inventor on two patents held by University of Birmingham (Atrial Fibrillation Therapy WO 2015140571, Markers for Atrial Fibrillation WO 2016012783). K.L. is an employee of Bayer. The genotyping of participants in the Broad AF study and the expression analysis of LA tissue samples were supported by a grant from Bayer to the Broad Institute. S.N. is a consultant to Biosense Webster, Siemens and Cardiosolv. S.N. also receives research grants from NIH/NHLBI, Siemens, Biosense Webster and Imricor. S. Kathiresan has received grant support from Bayer and Amarin; holds equity in San Therapeutics and Catabasis; and has received personal fees for participation in scientific advisory boards for Catabasis, Regeneron Genetics Center, Merck, Celera, Genomics PLC, Corvidia Therapeutics and Novo Ventures. S. Kathiresan also received personal fees for consulting services from Novartis, AstraZeneca, Alnylam, Eli Lilly Company, Leerink Partners, Merck, Noble Insights, Bayer, Ionis Pharmaceuticals, Novo Ventures, Haug Partners LLC and Genetic Modifiers Newco, Inc. S.A.L. receives sponsored research support from Bristol Myers Squibb, Bayer, Biotronik and Boehringer Ingelheim, and has consulted for St. Jude Medical/Abbott and Quest Diagnostics. The remaining authors have no disclosures.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0133-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.T.E.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Carolina Roselli[1,160], Mark D. Chaffin[1,160], Lu-Chen Weng[1,2,160], Stefanie Aeschbacher[3,4], Gustav Ahlberg[5,6,7], Christine M. Albert[8], Peter Almgren[9], Alvaro Alonso[10], Christopher D. Anderson[1,11], Krishna G. Aragam[1,11], Dan E. Arking[12], John Barnard[13], Traci M. Bartz[14], Emelia J. Benjamin[15,16,17], Nathan A. Bihlmeyer[18], Joshua C. Bis[19], Heather L. Bloom[20], Eric Boerwinkle[21], Erwin B. Bottinger[22,23], Jennifer A. Brody[19], Hugh Calkins[24], Archie Campbell[25], Thomas P. Cappola[26], John Carlquist[27,28], Daniel I. Chasman[1,29], Lin Y. Chen[30], Yii-Der Ida Chen[31], Eue-Keun Choi[32], Seung Hoan Choi[1], Ingrid E. Christophersen[1,2,33], Mina K. Chung[13], John W. Cole[34,35], David Conen[3,4,36], James Cook[37], Harry J. Crijns[38], Michael J. Cutler[27], Scott M. Damrauer[39,40], Brian R. Daniels[1], Dawood Darbar[41], Graciela Delgado[42], Joshua C. Denny[43], Martin Dichgans[44,45,46], Marcus Dörr[47,48], Elton A. Dudink[38], Samuel C. Dudley[49], Nada Esa[50], Tonu Esko[1,51], Markku Eskola[52], Diane Fatkin[53,54,55], Stephan B. Felix[47,48], Ian Ford[56], Oscar H. Franco[57], Bastiaan Geelhoed[58], Raji P. Grewal[59,60], Vilmundur Gudnason[61,62], Xiuqing Guo[31], Namrata Gupta[1], Stefan Gustafsson[63], Rebecca Gutmann[64], Anders Hamsten[65], Tamara B. Harris[66], Caroline Hayward[67], Susan R. Heckbert[68,69], Jussi Hernesniemi[52,70], Lynne J. Hocking[71], Albert Hofman[57], Andrea R. V. R. Horimoto[72], Jie Huang[73], Paul L. Huang[2], Jennifer Huffman[67], Erik Ingelsson[63,74], Esra Gucuk Ipek[24], Kaoru Ito[75], Jordi Jimenez-Conde[76,77], Renee Johnson[53], J. Wouter Jukema[78,79,80], Stefan Kääb[81,82], Mika Kähönen[83], Yoichiro Kamatani[84], John P. Kane[85], Adnan Kastrati[82,86], Sekar Kathiresan[1,11], Petra Katschnig-Winter[87], Maryam Kavousi[57], Thorsten Kessler[86], Bas L. Kietselaer[38], Paulus Kirchhof[88,89,90], Marcus E. Kleber[42], Stacey Knight[27,91], Jose E. Krieger[72], Michiaki Kubo[92], Lenore J. Launer[66], Jari Laurikka[93], Terho Lehtimäki[70], Kirsten Leineweber[94], Rozenn N. Lemaitre[19], Man Li[95,96], Hong Euy Lim[97], Henry J. Lin[31], Honghuang Lin[15,16], Lars Lind[98], Cecilia M. Lindgren[99], Marja-Liisa Lokki[100], Barry London[64], Ruth J. F. Loos[22,101,102], Siew-Kee Low[84], Yingchang Lu[22,101], Leo-Pekka Lyytikäinen[70], Peter W. Macfarlane[103], Patrik K. Magnusson[104], Anubha Mahajan[99], Rainer Malik[44], Alfredo J. Mansur[105], Gregory M. Marcus[106], Lauren Margolin[1], Kenneth B. Margulies[26], Winfried März[107,108], David D. McManus[50], Olle Melander[109], Sanghamitra Mohanty[110,111], Jay A. Montgomery[43], Michael P. Morley[26], Andrew P. Morris[37], Martina Müller-Nurasyid[81,82,112], Andrea Natale[110,111], Saman Nazarian[113], Benjamin Neumann[81], Christopher Newton-Cheh[1,11], Maartje N. Niemeijer[57], Kjell Nikus[52], Peter Nilsson[114], Raymond Noordam[115], Heidi Oellers[116], Morten S. Olesen[5,6,7], Marju Orho-Melander[9], Sandosh Padmanabhan[117], Hui-Nam Pak[118], Guillaume Paré[36,119], Nancy L. Pedersen[104], Joanna Pera[120], Alexandre Pereira[121,122], David Porteous[25], Bruce M. Psaty[69,123], Sara L. Pulit[1,124,125], Clive R. Pullinger[85], Daniel J. Rader[126], Lena Refsgaard[5,6,7], Marta Ribasés[127,128,129], Paul M. Ridker[8], Michiel Rienstra[58], Lorenz Risch[130,131], Dan M. Roden[43], Jonathan Rosand[1,11], Michael A. Rosenberg[11,132], Natalia Rost[1,133], Jerome I. Rotter[134], Samir Saba[135], Roopinder K. Sandhu[136], Renate B. Schnabel[137,138], Katharina Schramm[81,112], Heribert Schunkert[82,86], Claudia Schurman[22,101], Stuart A. Scott[139], Ilkka Seppälä[70], Christian Shaffer[43], Svati Shah[140], Alaa A. Shalaby[135,141], Jaemin Shim[142], M. Benjamin Shoemaker[43], Joylene E. Siland[58], Juha Sinisalo[143], Moritz F. Sinner[81,82], Agnieszka Slowik[120], Albert V. Smith[61,62], Blair H. Smith[144],

J. Gustav Smith[1,145], Jonathan D. Smith[13], Nicholas L. Smith[68,69], Elsayed Z. Soliman[146], Nona Sotoodehnia[147], Bruno H. Stricker[148,149], Albert Sun[140], Han Sun[13], Jesper H. Svendsen[5,7], Toshihiro Tanaka[150], Kahraman Tanriverdi[50], Kent D. Taylor[31], Maris Teder-Laving[51], Alexander Teumer[48,151], Sébastien Thériault[36,119], Stella Trompet[78,115], Nathan R. Tucker[1,2], Arnljot Tveit[33,152], Andre G. Uitterlinden[148], Pim Van Der Harst[58], Isabelle C. Van Gelder[58], David R. Van Wagoner[13], Niek Verweij[58], Efthymia Vlachopoulou[100], Uwe Völker[48,153], Biqi Wang[154], Peter E. Weeke[5,43], Bob Weijs[38], Raul Weiss[155], Stefan Weiss[48,153], Quinn S. Wells[43], Kerri L. Wiggins[19], Jorge A. Wong[156], Daniel Woo[157], Bradford B. Worrall[158], Pil-Sung Yang[118], Jie Yao[31], Zachary T. Yoneda[43], Tanja Zeller[137,138], Lingyao Zeng[86], Steven A. Lubitz[1,2,159,161], Kathryn L. Lunetta[15,154,161] and Patrick T. Ellinor[1,2,159,161]*

---

[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]University Hospital Basel, Basel, Switzerland. [4]Cardiovascular Research Institute Basel, Basel, Switzerland. [5]Laboratory for Molecular Cardiology, The Heart Centre, Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. [6]The Danish National Research Foundation Centre for Cardiac Arrhythmia, Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark. [7]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [8]Divisions of Preventive and Cardiovascular Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA. [9]Department of Clinical Sciences, Lund University, Malmo, Sweden. [10]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [11]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [12]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [13]Departments of Cardiovascular Medicine, Cellular and Molecular Medicine, Molecular Cardiology, and Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA. [14]Cardiovascular Health Research Unit, Departments of Medicine and Biostatistics, University of Washington, Seattle, WA, USA. [15]NHLBI and Boston University's Framingham Heart Study, Framingham, MA, USA. [16]Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [17]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA. [18]Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [19]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [20]Division of Cardiology, Emory University and Atlanta VA Medical Center, Atlanta, GA, USA. [21]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [22]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [23]Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [24]Johns Hopkins University, Baltimore, MD, USA. [25]Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [26]Penn Cardiovascular Institute and Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [27]Intermountain Heart Institute, Intermountain Medical Center, Murray, UT, USA. [28]Division of Cardiovascular Medicine, University of Utah, Salt Lake City, UT, USA. [29]Divisions of Preventive Medicine and Genetics, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA. [30]Cardiovascular Division, Department of Medicine, University of Minnesota Medical School, Minneapolis, MN, USA. [31]Institute for Translational Genomics and Population Sciences, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA. [32]Seoul National University Hospital, Seoul, Korea. [33]Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust, Drammen, Norway. [34]Baltimore Veterans Affairs Medical Center, Department of Neurology, Baltimore, MD, USA. [35]University of Maryland School of Medicine, Department of Neurology, Baltimore, MD, USA. [36]Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada. [37]Department of Biostatistics, University of Liverpool, Liverpool, UK. [38]Maastricht University Medical Center+ and Cardiovascular Research Institute Maastricht, Department of Cardiology, Maastricht, The Netherlands. [39]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [40]Department of Surgery, Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, USA. [41]University of Illinois Chicago, Chicago, IL, USA. [42]Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany. [43]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [44]Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany. [45]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [46]German Center for Neurodegenerative Diseases (DZNE), Munich, Germany. [47]Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. [48]DZHK (German Centre for Cardiovascular Research), partner site: Greifswald, Greifswald, Germany. [49]Cardiovascular Division and Lillehei Heart Institute, University of Minnesota, Minneapolis, MN, USA. [50]University of Massachusetts Medical School Worcester, Worcester, MA, USA. [51]Estonian Genome Center, University of Tartu, Tartu, Estonia. [52]Heart Center, Department of Cardiology, Tampere University Hospital, Finland and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [53]Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales, Australia. [54]St Vincent's Hospital, Darlinghurst, New South Wales, Australia. [55]Faculty of Medicine, University of New South Wales, Kensington, New South Wales, Australia. [56]Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK. [57]Department of Epidemiology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands. [58]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [59]Dept. of Neuroscience, Saint Francis Medical Center, Trenton, NJ, USA. [60]School of Health and Medical Sciences, Seton Hall University, South Orange, NJ, USA. [61]Icelandic Heart Association, Kopavogur, Iceland. [62]Faculty of Medicine, University of Iceland, Reykavik, Iceland. [63]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [64]Division of Cardiovascular Medicine and Abboud Cardiovascular Research Center, University of Iowa, Iowa City, IA, USA. [65]Cardiovascular Genetics and Genomics Group, Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. [66]Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, MD, USA. [67]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [68]Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, WA, USA. [69]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. [70]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [71]Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK. [72]Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of São Paulo, São Paulo, Brazil. [73]Boston VA Research Institute, Inc., Boston, MA, USA. [74]Division of Cardiovascular Medicine,

Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [75]Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [76]Department of Neurology, Neurovascular Research Group IMIM–Hospital del Mar (Institut Hospital del Mar d'Investigacions Médiques), Barcelona, Spain. [77]Universitat Autònoma de Barcelona, Medicine Department, Barcelona, Spain. [78]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands. [79]Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands. [80]Interuniversity Cardiology Institute of the Netherlands, Utrecht, The Netherlands. [81]Department of Medicine I, University Hospital Munich, Ludwig-Maximilians-University, Munich, Germany. [82]DZHK (German Centre for Cardiovascular Research), partner site: Munich Heart Alliance, Munich, Germany. [83]Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [84]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [85]Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA, USA. [86]Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München, Munich, Germany. [87]Department of Neurology, Medical University of Graz, Graz, Austria. [88]Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, UK. [89]Sandwell and West Birmingham Hospitals NHS Trust and University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. [90]AFNET, Muenster, Germany. [91]Department of Medicine, University of Utah, Salt Lake City, UT, USA. [92]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [93]Department of Cardio-Thoracic Surgery, Heart Center, Tampere University Hospital, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [94]Dept. Disease Genomics, Bayer, Wuppertal, Germany. [95]Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA. [96]Division of Nephrology & Hypertension, Internal Medicine, School of Medicine, University of Utah, Salt Lake City, UT, USA. [97]Korea University Guro Hospital, Seoul, Korea. [98]Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden. [99]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [100]Transplantation Laboratory, Medicum, University of Helsinki, Helsinki, Finland. [101]The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [102]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [103]Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [104]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [105]Heart Institute, University of São Paulo, São Paulo, Brazil. [106]Division of Cardiology, University of California, San Francisco, San Francisco, California, USA. [107]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. [108]Synlab Academy, Synlab Services GmbH, Mannheim, Germany. [109]Department of Internal Medicine, Clinical Sciences, Lund University, Malmo, Sweden. [110]Texas Cardiac Arrhythmia Institute, St David's Medical Center, Austin, TX, USA. [111]Dell Medical School, Austin, TX, USA. [112]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [113]University of Pennsylvania, Philadelphia, PA, USA. [114]Department of Clinical Sciences, Lund University and Skåne University Hospital, Malmo, Sweden. [115]Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands. [116]Atrial Fibrillation NETwork, Muenster, Germany. [117]Institute of Cardiovascular and Medical Sciences, BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, UK. [118]Yonsei University Health System, Seoul, Korea. [119]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada. [120]Department of Neurology, Faculty of Medicine, Jagiellonian University Medical College, Krakow, Poland. [121]Laboratory of Genetics and Molecular Biology, Heart Institute, University of São Paulo, São Paulo, Brazil. [122]Department of Genetics, Harvard Medical School, Boston, MA, USA. [123]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA. [124]Department of Genetics, Center for Molecular Medicine, University Medical Centre Utrecht, Utrecht University, Utrecht, The Netherlands. [125]Li Ka Shing Center for Health Information and Discovery, Big Data Institute, Oxford University, Oxford, UK. [126]Division of Cardiovascular Medicine, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [127]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [128]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [129]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. [130]University Institute of Clinical Chemistry, University of Bern, Bern, Switzerland. [131]Labormedizinisches Zentrum Dr. Risch, Schaan, Liechtenstein. [132]University of Colorado School of Medicine, Aurora, CO, USA. [133]J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, Boston, MA, USA. [134]Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA. [135]Division of Cardiology, University of Pittsburgh, Pittsburgh, PA, USA. [136]Division of Cardiology, University of Alberta, Edmonton, Alberta, Canada. [137]Department of General and Interventional Cardiology, University Heart Centre Hamburg, Hamburg, Germany. [138]DZHK (German Centre for Cardiovascular Research), partner site: Hamburg/Kiel/Lübeck, Hamburg, Germany. [139]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [140]Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, North Carolina, USA. [141]Cardiology Division, Pittsburgh VA Healthcare System, Pittsburgh, Pennsylvania, USA. [142]Korea University Anam Hospital, Seoul, Korea. [143]Heart and Lung Center HUS, Helsinki University Central Hospital, Helsinki, Finland. [144]Division of Population Health Sciences, University of Dundee, Dundee, UK. [145]Department of Cardiology, Clinical Sciences, Lund University and Skåne University Hospital, Lund, Sweden. [146]Epidemiological Cardiology Research Center (EPICARE), Wake Forest School of Medicine, Winston Salem, NC, USA. [147]Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle, WA, USA. [148]Department of Epidemiology and Internal Medicine, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands. [149]Inspectorate of Health Care, Utrecht, The Netherlands. [150]Department of Human Genetics and Disease Diversity, Tokyo Medical and Dental University Graduate School of Medical and Dental Sciences, Tokyo, Japan. [151]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [152]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [153]Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany. [154]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [155]Division of Cardiovascular Medicine, The Ohio State University, Columbus, OH, USA. [156]Division of Cardiology, Hamilton Health Sciences, McMaster University, Hamilton, Ontario, Canada. [157]University of Cincinnati College of Medicine, Cincinnati, OH, USA. [158]Departments of Neurology and Public Health Science, University of Virginia Health System, Charlottesville, VA, USA. [159]Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, MA, USA. [160]These authors contributed equally: Carolina Roselli, Mark D. Chaffin, Lu-Chen Weng. [161]These authors jointly supervised this work: Steven A. Lubitz, Kathryn L. Lunetta, Patrick T. Ellinor. *e-mail: ellinor@mgh.harvard.edu

Methods

Samples. Participants from more than 50 studies were included in this analysis. Participants were collected from both case-control studies for AF and population-based studies. The majority of studies were part of the Atrial Fibrillation Genetics (AFGen) consortium and the Broad AF study (Broad AF). Additional summary-level results from the UK Biobank (UKBB) and Biobank Japan (BBJ) were included (Fig. 1). Cases include participants with paroxysmal or permanent AF, or atrial flutter, and referents were free of these diagnoses. Adjudication of AF for each study is described in the Supplementary Notes. Ascertainment of AF in the UKBB includes samples with one or more of the following codes: non-cancer illness code, self-reported (1471, 1483); operation code (1524); diagnoses – main/secondary ICD10 (I48, I48.0-4, I48.9); underlying (primary/secondary) cause of death: ICD10 (I48, I48.0-4, I48.9); diagnoses – main/secondary ICD9 (4273); operative procedures – main/secondary OPCS (K57.1, K62.1-4)[8,10,33]. Baseline characteristics for each study are reported in Supplementary Table 17. We analyzed 55,114 cases and 482,295 referents of European ancestry, 1,307 cases and 7,660 referents of African American ancestry, 8,180 cases and 28,612 referents of Japanese ancestry, 568 cases and 1,096 referents from Brazil, and 277 cases and 3,081 referents of Hispanic ethnicity. Samples from the UKBB, the Broad AF study and some studies from the AFGen consortium (SiGN, EGCUT, PHB and the Vanderbilt Atrial Fibrillation Registry) were previously not included in primary AF GWAS discovery analyses. There is minimal sample overlap from the studies MGH AF, BBJ and AFLMU between this and previous analyses. Ethics approval for participation was obtained individually by each study. All relevant ethical regulations were followed for this work. Written informed consent was obtained from all study participants.

The Institutional Review Board (IRB) at Massachusetts General Hospital reviewed and approved the overall study.

Genotyping and genotype calling. Samples within the Broad AF study were genotyped at the Broad Institute using the Infinium PsychArray-24 v1.2 Bead Chip. They were genotyped in 19 batches, grouped by origin of the samples and with a balanced case-control mix on each array. Common variants (≥1% MAF) were called with GenomeStudio v1.6.2.2 and Birdseed v1.33[34], while rare variants (<1% MAF) were called with zCall[35]. Batch-specific quality control (QC) was performed on each call set including >95% sample call rate, Hardy–Weinberg equilibrium $P > 1 \times 10^{-6}$ and variant call rate >97%. For common variants, a consensus merge was performed between the call sets from GenomeStudio and Birdseed. For each genotype, only concordant calls between the two algorithms were kept. The common variants from the consensus call were then combined with the rare variant calls from the zCall algorithm. Samples from all batches were joined before performing pre-imputation QC steps. Detailed procedures for genotyping and genotype calling for the SiGN study[36], the UKBB[37] and BBJ[9] are described elsewhere. Details on genotyping and calling for all participating studies are listed in Supplementary Table 18.

Imputation. Pre-imputation QC filtering of samples and variants was conducted according to recommended guidelines as described in Supplementary Table 19. QC steps were performed by each study and are described in Supplementary Table 18. Most studies with European ancestry samples performed imputation with the HRC reference v1.1[38] panel on the Michigan Imputation Server v1.0.1[39]. Studies without available HRC imputation were included on the basis of imputation to the 1000 Genomes Phase 1 integrated v3 panel (March 2012)[40]. Participants of the SiGN study were imputed to a combined reference panel consisting of 1000 Genomes phase 1 plus Genome of the Netherlands[41]. Studies from Brazil were imputed with the HRC reference v1.1 panel. Studies of Japanese ancestry or Hispanic ethnicity were imputed to the 1000G Phase 1 integrated v3 panel (March 2012). Studies of African American ancestry were imputed to the HRC reference v1.1 panel or the 1000G Phase 1 integrated v3 panel (March 2012). Studies were advised to use the HRC preparation and checking tool (http://www.well.ox.ac.uk/~wrayner/tools/) before imputation. Prephasing and imputation methods for each study are described in Supplementary Table 18.

Primary statistical analyses. Genome-wide association testing on autosomal chromosomes was performed using an additive genetic effect model based on genotype probabilities. Each ancestry group was analyzed separately for each study. For the Broad AF study, the primary statistical analysis was performed jointly on unrelated individuals, excluding one of each pair for related samples with PI_HAT >0.2 as calculated in PLINK v1.90[42]. Samples with sex mismatches and sample call rate <97% were excluded. Ancestry groups were defined with ADMIXTURE[43] based on genotyped, independent and high-quality variants, using the supervised method with 1000 Genomes phase 1 v3 samples as a reference. A cutoff of 80% European ancestry was used to define the European subset and a cutoff of 60% African ancestry was used to define the African American subset. A Brazilian cohort within the Broad AF study was analyzed separately. Principal components were calculated within each ancestry group with the smartpca program from EIGENSOFT v6.1.1[44]. For the UKBB, a European subset was selected within samples with self-reported white race (British, Irish or other) and similar genetic ancestry. Genetic similarity was defined with the aberrant[45] package in R based on principal components, following the same method as described for the UKBB[37]. We excluded samples with sex mismatches, outliers in heterozygosity and missing rates, samples that carry sex chromosome configurations other than XX or XY and samples that were excluded from the kinship inference procedure as flagged in the UKBB QC file. We further removed one sample for each pair of third-degree or closer relatives (kinship coefficient >0.0442), preferentially keeping samples with AF case status. Primary analyses for all other studies were performed at the study sites and the summary level data of the results were provided. Prevalent cases were analyzed in a logistic regression model and most incident cases were analyzed in a Cox proportional hazards model. Studies with both prevalent and incident cases analyzed these either separately using a logistic regression model or a Cox proportional hazards model respectively, or jointly in a logistic regression model. The following tools were used for primary GWAS: ProbABEL[46], SNPTEST[47], FAST[48], mach2dat (http://www.unc.edu/~yunmli/software.html), R[49], EPACTS (http://genome.sph.umich.edu/wiki/EPACTS), Hail (https://github.com/hail-is/hail) and PLINK[42] (Supplementary Table 18). Summary-level results were filtered, keeping variants with imputation quality >0.3 and MAF × imputation quality × N events ≥ 10. We performed post-analysis QC of the summary-level results for each study. We checked reported allele frequencies against allele frequencies from imputation reference panels by ancestry. We inspected Manhattan plots for spurious associations and quantile–quantile plots to identify genomic inflation. We also calculated the genomic inflation factor (λ_{GC}) for each study (Supplementary Table 18). Furthermore, we plotted the reported P value versus the P value derived from the Z score (effect/SE) to check for consistency of results. We checked the distribution of effect estimates and standard errors and confirmed that known AF risk variants[5] showed consistent directionality with reported effect estimates.

Meta-analyses. Summary-level results were meta-analyzed jointly with METAL (released on March 25, 2011) using a fixed-effects model with an inverse-variance weighted approach, correcting for genomic control[50]. Separate meta-analyses were conducted for each ancestry. The results for the Japanese-specific[9] and Hispanic-specific[8] analyses have previously been reported and therefore their ancestry-specific results are not shown. Variants were included if they were present in at least two studies and showed an average MAF ≥ 0.1%. To correct for multiple testing, a genome-wide significance threshold of $P < 1 \times 10^{-8}$ was applied for each analysis. This threshold is based on a naive Bonferroni correction for independent variants with MAF ≥ 0.1%, using an LD threshold of $r^2 < 0.8$ to estimate the number of independent variants based on European-ancestry LD[51]. As these meta-analyses are based on effect estimates and standard errors from both logistic regression and Cox proportional hazards regression, we report variant effects as relative risk, calculated as the exponential of effect estimates. For sentinel variants reaching genome-wide significance in the combined ancestry meta-analysis, we assessed whether effect estimates were homogeneous across ancestries by calculating an I^2 statistic[52] across ancestry-specific meta-analyses. We account for multiple testing across 94 variants using a Bonferroni correction, resulting in a significance threshold of $P < 5.32 \times 10^{-4}$ for the heterogeneity test.

Broad AF LD reference and proxies. A LD reference file was created including 26,796 European ancestry individuals from the Broad AF study. The LD reference was based on HRC-imputed genotypes. Monomorphic variants and variants with imputation quality <0.1 were removed before conversion to hard calls. A genotype probability threshold filter of >0.8 was applied during hard call conversion. For multi-allelic sites, the more common alleles were kept. Variants were included in the final reference file if the variant call rate was >70%.

We identified proxies of sentinel variants as variants in LD of $r^2 > 0.6$ based on the Broad AF LD reference file, using PLINK v1.90[42].

Meta-analysis of provisional loci. We meta-analyzed 111 variants from externally reported[15] provisional loci within predominantly non-overlapping samples from the Broad AF study, BBJ, EGCUT, PHB, SiGN and the Vanderbilt AF Registry with METAL (released on March 25, 2011)[50]. The predominantly non-overlapping samples included a total of 32,957 AF cases and 83,546 referents, with minimal overlap from the studies MGH AF, BBJ and AFLMU. We subsequently meta-analyzed these results with the reported provisional results with METAL using a fixed-effects model with an inverse-variance weighted approach. We analyzed a total of 93,577 AF cases and 1,053,762 referents. We compared our discovery results with the provisional loci using the same significance cutoff of $P < 5 \times 10^{-8}$ for both results. Overlapping loci were identified if the reported sentinel variants were located within 500 kb of each other. For overlapping loci with differing sentinel variants, we calculated the LD between the sentinel variants, based on the Broad AF LD reference panel of European ancestry.

Variant consequence on protein-coding sequence. The most severe consequence for variants was identified with the Ensembl Variant Effect Predictor version 89.7 using RefSeq as a gene reference and the option 'pick' to identify one consequence per variant with the default pick order[53]. We queried sentinel variants and their proxies to identify tagged variants with HIGH and MODERATE impact including the following consequences: 'transcript_ablation', 'splice_acceptor_variant', 'splice_donor_variant', 'stop_gained', 'frameshift_variant', 'stop_lost', 'start_lost', 'transcript_amplification', 'inframe_insertion', 'inframe_deletion', 'missense_variant' and ‘protein_altering_variant’. We evaluated each identified consequence on the protein-coding sequence with in silico prediction tools to assess potentially damaging effects. The evaluation included MutationTaster[54] (disease-causing automatic or disease-causing), SIFT[55] (damaging), LRT[56] (deleterious) and Polyphen2[57] prediction based on HumDiv and HumVar (probably damaging or possibly damaging).

Chromatin states. *Chromatin state annotation.* We identified chromatin states for sentinel variants and their proxies from the Roadmap Epigenomics Consortium 25-state model (2015)[58] using HaploReg v4[59]. We looked for chromatin states occurring in any included tissues as well as chromatin states occurring in heart tissue. Heart tissues include: E065, aorta; E083, fetal heart; E095, left ventricle; E104, right atrium; E105, right ventricle.

Regulatory region enrichment. One thousand sets of control loci were generated by matching SNPs to sentinel variants from the AF combined-ancestry analysis, with the SNPSnap[60] tool. We used the European 1000 Genomes Phase 3 population to match via MAF, gene density, distance to nearest gene and LD buddies using $r^2 > 0.6$ as the LD cutoff and otherwise default settings. We excluded input SNPs and HLA SNPs from the matched SNPs. Loci were defined as SNPs and their proxies with $r^2 > 0.6$ based on LD from the European 1000 Genomes Phase 3 population. We identified SNPs in regulatory regions across all tissues and in cardiac tissues (E065, E095, E104 and E105) based on the Roadmap Epigenomics Consortium 25-state model (2015)[58] using HaploReg v4[59]. Regulatory regions included the following states: 2_PromU, 3_PromD1, 4_PromD2, 9_TxReg, 10_TxEnh5, 11_TxEnh3, 12_TxEnhW, 13_EnhA1, 14_EnhA2, 15_EnhAF, 16_EnhW1, 17_EnhW2, 18_EnhAc, 19_DNase, 22_PromP and 23_PromBiv. We calculated the percentage of overlap of each annotation per locus, defined as the number of SNPs per locus that fall in regulatory regions divided by the total number of SNPs per locus. Statistical significance was calculated with a permutation test from the perm package in R[61].

eQTL. Variants identified from GWAS were assessed for overlap with eQTLs from two sources.

LA tissue from the Myocardial Applied Genomics Network (MAGNet) repository. We performed RNA-sequencing on 101 LA tissue samples from the MAGNet repository (http://www.med.upenn.edu/magnet/) on the Illumina HiSeq 4000 platform at the Broad Institute Genomic Services. LA tissue was obtained at the time of cardiac transplantation from normal donors with no evidence of structural heart disease. All left atrial samples were from individuals of European ancestry. A summary of the clinical characteristics for these samples is shown in Supplementary Table 20. Reads were aligned to the reference genome by STAR v2.4.1a[62] and assigned to genes based on the GENCODE gene annotation[63]. Gene expression was measured in fragments per kilobase of transcript per million mapped reads and subsequently quantile-normalized and adjusted for age, sex and the first ten principal components. Genotyping was performed on the Illumina OmniExpressExome-8v1 array and imputed to the HRC reference panel. Principal components were calculated with the smartpca program from EIGENSOFT v6.1.1[44] and European ancestry was confirmed by assessing principal components in the samples combined with 1000 Genomes European samples[40]. Associations between gene expression and genotypes were tested in a linear regression model with QTLtools v1.0[64], to detect *cis*-eQTLs, defined as eQTLs within 1 Mb of the transcription start site of a gene. To account for multiple testing, an empirical false discovery rate (FDR) was used to identify significant eQTLs with a FDR < 5%.

GTEx project. We queried the GTEx[65] version 6p database for *cis*-eQTLs with significant associations to gene expression levels in the two available heart tissues: LV and RA appendage[66].

Association between predicted gene expression and risk of atrial fibrillation. To investigate transcriptome-wide associations between predicted gene expression and AF disease risk, we employed the method MetaXcan v0.3.5[12]. MetaXcan extends the previous method PrediXcan[67] to predict the association between gene expression and a phenotype of interest, using summary association statistics. Gene expression prediction models were generated from eQTL data sets using Elastic-Net to identify the most predictive set of SNPs. Only models that significantly predict gene expression in the reference eQTL data set (FDR < 0.05) were considered. Pre-computed MetaXcan models for the two available heart tissues (LV and RA appendage) in the GTEx project version 6p[66] were used to predict the association between gene expression and risk of AF. Summary-level statistics from the combined ancestry meta-analysis were used as input. A total of 4,859 genes were tested for LV and 4,467 genes were tested for RA appendage. Bonferroni correction was applied to account for the number of genes tested across both tissues, resulting in a significance threshold of $P < 5.36 \times 10^{-6}$, calculated as $0.05/(4{,}859 + 4{,}467)$.

Conditional and joint analyses. Conditional and joint analyses[68] of GWAS summary statistics were performed with Genome-wide Complex Trait Analysis (GCTA v1.25.2)[69] using a stepwise selection procedure to identify independently associated variants on each chromosome. We used the Broad AF LD reference file for LD calculations.

Gene set enrichment analysis. A meta-analysis gene-set enrichment of variant associations (MAGENTA) v2.4[70] was performed with a combined gene set input database (GO_PANTHER_INGENUITY_KEGG_REACTOME_BIOCARTA) based on publicly available data. The analysis was conducted using the summary-level results from the combined ancestry meta-analysis. A total of 4,045 gene sets were included and multiple testing was corrected via FDR. Gene sets were manually assigned to one or more of the following functional groups: developmental, electrophysiological, contractile/structural, and other. Genes within 500 kb of a sentinel variant were identified on the basis of the longest spanning transcribed region in the RefSeq gene reference. For each gene set, genes close to significant loci were listed. The selected genes were assigned to one or more functional groups based on their affiliation to gene sets. Functional groups from gene sets with a single label were preferentially assigned.

Association with other phenotypes. To determine whether the sentinel AF risk variants had associations with other phenotypes, two sources of data were used.

GWAS catalog. We queried the NHGRI-EBI Catalog of published GWAS[71,72] (accessed August 31, 2017) to detect associations of AF risk variants with other phenotypes.

UKBB PheWAS. A PheWAS was conducted in the UKBB in European-ancestry individuals. Ancestry definition and sample QC exclusions were performed in the same manner as for the primary statistical analysis, as described above. We further removed one sample for each pair of second-degree or closer relatives (kinship coefficient >0.0884), preferentially keeping the sample with case status or non-missing phenotype. We included the following phenotypes: height, body mass index, smoking, hypertension, heart failure, stroke, mitral regurgitation, bradyarrhythmia, peripheral vascular disease, hypercholesterolemia, coronary artery disease and type II diabetes. Phenotype definitions are shown in Supplementary Table 21. The number of samples analyzed, and case and referent counts for each phenotype, are listed in Supplementary Table 22. Binary phenotypes were analyzed with a logistic regression model and quantitative phenotypes with a linear regression model using imputed genotype dosages in PLINK 2.00[42]. As covariates, we included sex, age at first visit, genotyping array and the first ten principal components.

Proportion of heritability explained. We calculated SNP-heritability (h^2_g) of AF-associated loci with the REML algorithm in BOLT-LMM v2.2[73] in 120,286 unrelated samples of European ancestry from a subset of the UKBB data set comprising a prior interim release as previously described in separate work from our group[10]. We defined loci on the basis of a 1 Mb (±500 kb) window around 84 sentinel variants from the European-ancestry meta-analysis. We transformed the h^2_g estimates into liability scale (AF prevalence = 2.45% in the UKBB). We then calculated the proportion of h^2_g explained at AF loci by dividing the h^2_g estimate of AF-associated loci by the total h^2_g for AF, which was based on 811,488 LD-pruned and hard-called common variants (MAF ≥ 1%)[10].

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The data sets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request. The results of this study are available on the Cardiovascular Disease Knowledge Portal (http://www.broadcvdi.org/). The left atrial RNA-sequencing data can be accessed via dbGaP under the accession number phs001539.

References

33. Weng, L.-C. et al. Genetic predisposition, clinical risk factor burden, and lifetime risk of atrial fibrillation. *Circulation* **137**, 1027–1038 (2017).
34. Korn, J. M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).
35. Goldstein, J. I. et al. zCall: a rare variant caller for array-based genotyping. *Bioinformatics* **28**, 2543–2545 (2012).
36. Pulit, S. L. et al. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol.* **15**, 174–184 (2016).
37. Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. Preprint at https://www.biorxiv.org/content/early/2017/07/20/166298 (2017).
38. The Haplotype Reference Consortium et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
39. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

40. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
41. Francioli, L. C. et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).
42. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
43. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
44. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
45. Bellenguez, C., Strange, A., Freeman, C., Donnelly, P. & Spencer, C. C. A. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* **28**, 134–135 (2012).
46. Aulchenko, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
47. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
48. Chanda, P., Huang, H., Arking, D. E. & Bader, J. S. Fast association tests for genes with FAST. *PLoS One* **8**, e68585 (2013).
49. R Core Team. R: A Language and Environment for Statistical Computing. http://www.r-project.org/ (2015).
50. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
51. Fadista, J., Manning, A. K., Florez, J. C. & Groop, L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.* **24**, 1202–1205 (2016).
52. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
53. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
54. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
55. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
56. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
57. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
58. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* **33**, 364–376 (2015).
59. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
60. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31**, 418–420 (2015).
61. Fay, M. P. & Shaw, P. A. Exact and asymptotic weighted logrank tests for interval censored data: the interval R package. *J. Stat. Softw.* **36**, 1–34 (2010).
62. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
63. Harrow, J. et al. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
64. Delaneau, O. et al. A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* **8**, 15452 (2017).
65. GTEx Consortium. The Genotype–Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science **348**, 648–660 (2015).
66. Aguet, F. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
67. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet*. **47**, 1091–1098 (2015).
68. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
69. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
70. Segrè, A. V. et al. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
71. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
72. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
73. Loh, P.-R. et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).

natureresearch

Corresponding Author: Patrick T. Ellinor

Date: Mar 8, 2018

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.	We analyzed the largest sample size available from the AFGen consortium, the Broad AF study, the UK Biobank and the Biobank Japan, including 65,446 individuals with atrial fibrillation (AF) and 522,744 referents. No method was applied to predetermine sample size.

2. Data exclusions

Describe any data exclusions.	Samples and variants were excluded during the pre-imputation quality control procedure. Sample exclusion criteria included: sample call rate, heterozygosity outliers, ancestry outliers, related individuals and sex mismatches. Variant exclusion criteria included: variant call rate, deviations from Hardy-Weinberg, high discordance rates, excess of Mendelian inconsistencies and rare variants. Variant exclusions for the summary level results, prior to meta-analysis, included an imputation quality filter > 0.3 and a score of MAF * imputation quality * N events ≥ 10. Variants available in just 1 study were excluded from the meta-analysis.

3. Replication

Describe whether the experimental findings were reliably reproduced.	Experimental replication was not attempted.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.	AF cases: participants with paroxysmal or permanent atrial fibrillation, or atrial flutter. Referents: participants free of these diagnoses. Participants were grouped for analysis by study and by ancestry.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.	Blinding was not relevant to our study. The participants from the included studies were sampled by multiple different research centers. The meta-analysis was conducted centrally on summary level results from each study.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

METAL (released on 2011-03-25): https://genome.sph.umich.edu/wiki/METAL_Documentation
Michigan Imputation Server (v1.0.1): https://imputationserver.sph.umich.edu
HaploReg (v4): http://archive.broadinstitute.org/mammals/haploreg/haploreg.php
MetaXcan (v0.3.5): https://github.com/hakyimlab/MetaXcan/wiki
GCTA (v1.25.2): http://cnsgenomics.com/software/gcta/
MAGENTA (v2.4): https://software.broadinstitute.org/mpg/magenta/
LocusZoom (v1.3): https://genome.sph.umich.edu/wiki/LocusZoom_Standalone
VEP (v89.7): http://www.ensembl.org/info/docs/tools/vep/script/
R (v3.2.1, v3.2.3): https://www.r-project.org/
PLINK (v2.00, v1.90): https://www.cog-genomics.org/plink/
EIGENSOFT (v6.1.1): https://www.hsph.harvard.edu/alkes-price/software/
QTLtools (v1.0): https://qtltools.github.io/qtltools/
STAR (v2.4.1a): http://code.google.com/p/rna-star/
ProbABEL (v0.5.0): http://www.genabel.org/packages/ProbABEL
SNPTEST (v2.4.1, v2.5, v2.5.2, v2.5.4): https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html
mach2dat (v1.2.4): https://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output
EPACTS (v3.2.6): https://genome.sph.umich.edu/wiki/EPACTS
Hail (v0.1): https://github.com/hail-is/hail
MaCH (v1.0.16, v1.0.151): http://csg.sph.umich.edu/abecasis/mach/
Minimac (v3): https://genome.sph.umich.edu/wiki/Minimac
ShapeIT (v2.r790, v1.r532, v2.r837): https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
IMPUTE2 (v2.3.0, v2.1.0, v2.3.2, v2.2.2): http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
Eagle (v2.3): https://data.broadinstitute.org/alkesgroup/Eagle/
FAST: https://bitbucket.org/baderlab/fast/wiki/Home
BOLT-LMM (v2.2): https://data.broadinstitute.org/alkesgroup/BOLT-LMM/
SNPSnap: https://data.broadinstitute.org/mpg/snpsnap/

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No eukaryotic cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Sample for GWAS analyses: The sample was composed of 91% European, 6% Japanese, 1.5% African American, 0.3% Brazilian and 0.6% Hispanic ancestry. In total 46% of the participants were male with an average age of 58 years at DNA draw and an average BMI of 27. Detailed population characteristics for each study are provided in Supplementary Table S17. Sample for left atrial eQTL analyses: The participants were of European ancestry. In total 44% of the participants were male, with an average age of 59, including 59% participants with hypertension, 20% with diabetes and 14% with history of atrial fibrillation.