Supplemental Material

Table of Contents

Supplementary Methods

Study sample

Individual-level data used in the present study were from the UK Biobank, a prospective cohort study comprising 502,629 individuals aged 40 to 69 at enrollment from the United Kingdom during 2006-2010.[9,10] Briefly, phenotypic data comprising survey, longitudinal electronic health record, national registry, imaging, and other bioassay data were collected. In total 488,377 participants underwent genome-wide genotyping as summarized below. Of these, 39,986 individuals who underwent 12-lead ECG collection and had complete data to estimate the CHARGE-AF score, were without a history of AF at the time of ECG acquisition, and in whom data passed quality control (QC) procedures, were included in the genome-wide association studies (GWAS) of predicted AF risk using either ECG-AI or CHARGE-AF. Of the remaining 448,391 individuals without ECGs, 424,411 that passed genome-wide genotyping QC procedures were included in polygenic risk score analyses. The study overview and analytic workflow can be found in **Figure 1**.

Predicted risk of AF

Our study outcome was the predicted 5-year risk of AF. The 5-year risk of AF was derived from both an ECG-AI model, and separately from the CHARGE-AF clinical risk factor model. The development and validation of the ECG-AI model has previously been reported.[7] Briefly, we trained a convolutional neural network (CNN) to predict 5-year risk of AF using 12-lead ECGs in a longitudinal patient cohort derived from the MGB network.[34] ECG-AI uses an encoding and loss function that takes into account both time to event (i.e., AF) and missingness introduced by censoring (death or loss of follow-up) to estimate a 5-year survival probability of AF. The

predicted AF risk was then calculated as one minus the survival directly outputted from the ECG-AI model. We assessed discrimination of ECG-AI predicted risk for 5-year AF by calculating Harrell's c-index[35] and found a moderate discrimination of 0.70 (95% confidence interval: 0.67-0.72) in the 39,986 GWAS discovery set.

We calculated the CHARGE-AF predicted 5-year risk of AF using the published equation $1-0.9718412736^{\exp(X-12.5815600)}$, where X is the CHARGE-AF score for each individual.[8] Clinical risk factors in the score include baseline age, race, height, weight, systolic and diastolic blood pressure, smoking status, antihypertensive medication use, diabetes, heart failure, and myocardial infarction. Clinical variables were based on self-report (race, smoking status, antihypertensive medication use), measurements at the ECG visit in the UK Biobank (age at ECG, height, weight, systolic and diastolic blood pressure), or inpatient international classification of diseases (ICD) 9 or 10 codes (diabetes, heart failure, myocardial infarction) as has been described previously.[7] CHARGE-AF has been validated widely and shown to have good discrimination with C-statistics ranging from 0.66 to 0.81.[8,36,37] The ECG-AI and CHARGE-AF score linear predictors have previously been reported to have modest correlation (Pearson r 0.66) in an independent dataset, and the CHARGE-AF score exhibits similar predictive performance as ECG-AI.[7] For survival analysis in the present study, incident AF at 5 years was ascertained using follow-up information and a previously published definition comprising inpatient ICD-9 and -10 codes, and procedural codes.[38] Follow-up information in the UK Biobank depends on the availability of linked electronic health records and hence varies by enrollment sites. Censoring occurred at death, withdrawal of consent, or last availability of hospital linked data.[7]

Genetic data

Genotyping was carried out using two different arrays (UK BiLEVE Axiom Array and UK Biobank Axiom Array) with 95% shared marker content. After variants and sample QC processes, these markers were phased and imputed to the Haplotype Reference Consortium (HRC) and the merged UK10K and 1000 Genomes phase 3 reference panels, resulting in ~96 million variants in the final dataset.[10] We used the QC metrics provided by the UK Biobank to perform sample QC. Participants who had mismatch between self-reported sex and genetically inferred sex, high genotype missingness rate, high heterozygosity rate, or sex chromosome aneuploidy were removed from the analysis. UK Biobank calculated and provided the top 40 principal components for the entire cohort using 407,219 unrelated, high-quality samples and 147,604 high-quality, pruned markers. The PCA analysis was performed using the fastPCA[39] algorithm. Subsequently, all samples were projected onto the PCA space to obtain values for the entire cohort.[39] We used genotyped markers to fit null models for the GWAS. Variants were filtered out if MAF was < 0.1%, minor allele count (MAC) was < 100, missing call rate was > 0.1, or Hardy-Weinberg equilibrium exact test $P < 1\times10^{-5}$. In addition, samples with a missingness call rate > 0.1 were excluded from the analysis. We then performed the association tests using imputed genetic data, and variants included were those with a minor allele frequency (MAF) > 0.1% and an information score (imputation quality) > 0.3.

Common variant analyses

We conducted a GWAS of ECG-AI predicted 5-year AF risk, and a separate GWAS of CHARGE-AF predicted 5-year AF risk as a comparator. Rank-based inverse normal transformation (R-INT) was applied to each phenotype before association testing. We used the software Regenie,[40] which implements an efficient whole-genome regression approach, for this

analysis assuming an additive genetic model for each single nucleotide variant (SNV). Covariates adjusted in statistical models included age at the time of ECG, sex, genotyping array, and the first 20 principal components of ancestry. We also conducted conditional analyses[41] using the GCTA software to select secondary independent significant SNVs at each locus. In addition, as a comparator, we performed a GWAS of observed 5-year incident AF in the same dataset using Cox Mixed-Effects models[42] implemented in the *coxmeg* R package with the same covariates. Due to computational burden, we only tested variants with MAF > 10% and information score (imputation quality) > 0.9. Genome-wide significant variants were considered those with $P < 5\times10^{-8}$. The mapped gene at each risk locus was either the nearest gene or the trait-associated gene within 500 kb of the lead SNV reported in the GWAS catalog.

Heritability and genetic correlation

We calculated the additive heritability (h^2) for predicted risk of AF using a variance-components analysis with individual-level genetic and phenotypic data in our GWAS set, adjusting for age at the time of ECG, sex, genotyping array, and the first 20 principal components of ancestry using BOLT-REML.[43] We also used the same method to estimate the genetic correlation between the GWAS of predicted AF risk and 5-year observed incident AF. BOLT-REML implements a Monte Carlo average information restricted maximum-likelihood algorithm to estimate the genetic variance-covariance matrix. We then utilized GWAS summary statistics with LD score regression (LDSC)[14] to corroborate the heritability and genetic correlation estimates. GWAS results of predicted AF risk from the above common variant analyses were compared to a prior large-scale GWAS of AF.[11] We used pre-computed LD-scores provided by the LDSC software package. Here we report the raw genetic correlation estimates without any conversion since these parameters accurately reflect the genetic correlation of underlying disease liabilities.[44]

Heritability estimates were also not converted from observed scale to liability scale since predicted risks were normally distributed continuous traits after rank-based inverse normal transformation.

Polygenic risk scores

We derived polygenic risk scores (PRS) for ECG-AI and CHARGE-AF predicted risk of AF using the GWAS results and then applied these to the remaining participants without ECG data (**Figure 1**). In addition to the sample QC procedures described above, we further excluded participants with > 3rd degree relatedness in the GWAS set. The SNV effect sizes used in score calculations were obtained by using a fully Bayesian approach, PRS-CS,[45] which applies shrinkage factors to the original GWAS effect sizes to avoid overfitting. The polygenic score was calculated as the PRS-CS adjusted SNV effect size-weighted sum of the individual's genetic dosage at effect alleles. Score calculations were carried out using PLINK 2.0 [https://www.cog-genomics.org/plink/2.0/].

Associations between polygenic risk scores and incident AF

We tested the association between each polygenic risk score and 5-year incident AF using a Cox proportional hazards model adjusting for age at enrollment, sex, genotyping array, and the first 20 principal components of ancestry. We also fit another model containing both PRSs with the same covariates to estimate the effect of each PRS on 5-year incident AF risk after conditioning on the other PRS. We then further extended the model to include an interaction term between the two PRSs to assess for effect modification. We plotted the cumulative incidence of AF stratified by distribution of PRS risk (highest 10%, middle 80%, and lowest 10%). We performed the above analyses in the entire testing set as well as a subset of White British participants to assess

the properties of PRSs across ancestries. All analyses were conducted using *survival* and *survminer* packages in R version 4.0.

Mendelian Randomization

We employed a two-sample Mendelian randomization (2SMR) approach[46] implemented in the *TwoSampleMR*[47] R package to estimate potential causal effects of exposures on outcomes using GWAS summary statistics. Genetic instruments were chosen from GWAS results of exposure traits if they reached $P < 5\times10^{-8}$. A clumping step was then performed to select independent variants within 10,000 kb distance with a maximum LD r^2 of 0.001. These SNPs were then extracted from our ECG-AI risk summary statistics if they were present in both GWAS. The resulting genetic instruments were harmonized before analysis to ensure the effect alleles were consistent between the two studies. Summary statistics for P wave duration[16] (unit: ms) and height[48] (unit: SD) were chosen from studies that did not include UK Biobank samples. For each exposure, we performed MR analysis using five methods, including inverse variance weighted approach, weighted median approach, MR-Egger,[49] MR-PRESSO,[50] and MR-PRESSO corrected for horizontal pleiotropic outliers.

Expression association analysis

We investigated novel ECG-AI GWAS loci by examining their associations with AF in a prior AF GWAS,[11] and conducted an expression association analysis using the reference AF GWAS summary statistics with the MetaXcan software [https://github.com/hakyimlab/MetaXcan] to assess for associations between predicted gene expression and AF. Specifically, we tested associations between AF and the predicted expression of *VGLL2* and *EXT1* across 49 tissues from the GTEx (Genotype-Tissue Expression project)[52] database using the S-MultiXcan

approach.[51] Briefly the approach calculates associations between AF and predicted gene expression in individual tissues first, then infers the joint estimates of effect sizes of predicted expression on phenotype using the marginal estimates across tissues.

Median waveform analysis

To discern what ECG features were responsible for associations with observed genes, in exploratory analyses we created 4 gene-based polygenic risk scores (PRS) for discovered loci at *TTN*, *SCN5A/10A*, *VGLL2* and *EXT1*, respectively. Specifically, we took the variant weights used for creating the genome-wide $PRS_{ECG\text{-}AI}$ but restricted variants to those residing in the gene regions of interest (defined as 500kb around the gene) to create the gene-based scores. We then selected the top and bottom 1% groups from each of the PRS distributions and plotted the median ECG waveforms within each group.

Supplementary Table I. Risk loci in ECG-AI AF risk GWAS

rsID	Chr-Position (hg19)	Effect / Reference Allele	Effect Allele Frequency	Beta	Se	P	Imp Qual	Function	Nearest gene	eQTL for gene(s) in heart tissues	prior GWAS associations	GWAS associations reported for high LD SNVs
rs2627040	2-179601975	G/A	0.05	0.08	0.01	4.55E-08	0.999	intronic	*TTN*		No association reported in GWAS catalog	QT interval
rs6801957	3-38767315	C/T	0.60	-0.04	0.01	2.68E-12	1.000	intronic	*SCN10A*	*SCN10A*	Atrial fibrillation/atrial flutter, Electrocardiogram morphology (amplitude at temporal datapoints), Electrocardiographic traits (multivariate), PR interval, P wave duration, QT interval, QRS duration, Pulse pressure	Atrial fibrillation, Electrocardiographic traits, Electrocardiographic conduction measures, QRS duration, Heart rate response to recovery post exercise, PR interval, P wave duration, PR segment, QT interval, Pulse pressure, Atrioventricular conduction, Brugada syndrome, Heart rate response to exercise, Resting heart rate
rs9689288	6-117508467	G/C	0.47	0.04	0.01	5.05E-12	0.986	intergenic	*VGLL2*		No association reported in GWAS catalog	
rs35186392	8-118857805	T/A	0.20	0.04	0.01	7.22E-09	0.991	intronic	*EXT1*		No association reported in GWAS catalog	
Secondary association signals												
rs41312411	3-38621237	G/C	0.15	0.06	0.01	2.79E-10	0.981	intronic	*SCN5A*		Electrocardiography, P wave duration, Resting heart rate	PR interval, QT interval, QRS duration, Brugada syndrome

Genetic variants at risk loci independently associated with ECG-AI predicted 5-year AF risk.
rsID: variant ID; **Chr:** chromosome; **hg19:** genomic position based on build hg19; **Beta:** effect size of effect allele; **Se:** standard error; **P:** P-value in GWAS; **Imp Qual:** imputation quality.
eQTL for gene(s) in heart tissues: eQTL reported by the GTEx V8 data in Heart - Atrial Appendage tissue or Heart - Left Ventricle tissue.
GWAS associations reported for high LD SNVs: Traits associated with SNVs in high linkage disequilibrium (r2>=0.7) with the lead SNV as reported in GWAS catalog.

Supplementary Table II. Summary statistics from reference AF GWAS* for SNPs in high LD with top SNPs at *VGLL2* and *EXT1* loci

a. *VGLL2*

Chromosome position	rs ID	Effect allele	Other allele	Effect size	SE	P value	r^2 with rs9689288
6-117460805	rs11153650	T	C	-0.02	0.01	0.01	0.991
6-117522685	rs1405210	T	C	0.02	0.01	0.02	0.985
6-117485742	rs1508305	A	G	-0.02	0.01	0.02	0.995
6-117477632	rs2049528	A	G	0.02	0.01	0.01	0.992
6-117463853	rs3851205	T	C	0.02	0.01	0.01	0.991
6-117510203	rs6901720	T	G	0.02	0.01	0.02	0.996
6-117497691	rs6902752	T	C	0.02	0.01	0.01	0.995
6-117460355	rs6930573	C	G	-0.02	0.01	0.01	0.991
6-117459604	rs9372474	A	G	0.02	0.01	0.01	0.991
6-117469900	rs9374634	A	G	-0.02	0.01	0.01	0.991
6-117465537	rs9387449	A	G	-0.02	0.01	0.01	0.991
6-117483351	rs9387451	C	G	0.02	0.01	0.02	0.995
6-117493311	rs9387452	A	G	0.02	0.01	0.02	0.995
6-117482982	rs9400981	A	G	0.02	0.01	0.02	0.995
6-117487131	rs9400982	A	G	-0.02	0.01	0.02	0.995

b. *EXT1*

Chromosome position	rs ID	Effect allele	Other allele	Effect size	SE	P value	r2 with rs9689288
8-118857704	rs12676684	C	G	0.04	0.01	5.40E-06	0.994
8-118879130	rs12677136	T	C	-0.04	0.01	1.18E-04	0.816
8-118869138	rs12680472	T	C	0.04	0.01	4.96E-05	0.820
8-118844039	rs12682457	T	C	-0.04	0.01	7.61E-05	0.930
8-118863412	rs17430357	A	T	-0.04	0.01	8.16E-06	0.891
8-118863445	rs17430364	A	T	-0.04	0.01	7.40E-06	0.892
8-118838295	rs17439784	T	C	-0.04	0.01	6.14E-05	0.942
8-118842934	rs17439826	A	T	0.04	0.01	5.71E-05	0.931
8-118879552	rs28535273	A	G	0.03	0.01	1.87E-04	0.811
8-118877575	rs4129114	A	T	0.04	0.01	1.02E-04	0.819
8-118877488	rs4129115	A	C	0.04	0.01	1.16E-04	0.818
8-118877198	rs4129116	A	T	0.03	0.01	2.39E-04	0.850
8-118879406	rs62521109	A	G	-0.03	0.01	1.82E-04	0.810
8-118850101	rs66732486	A	G	-0.04	0.01	3.40E-05	0.945
8-118830036	rs7000499	T	C	-0.03	0.01	0.002	0.895

* Reference AF GWAS[11]

Supplementary Table III. Association between AF and predicted gene expression levels of *VGLL2* and *EXT1* across 49 GTEx* tissues

Gene	P value for association between predicted expression and AF across tissues	Number of tissues available for this gene	Most significant tissue	P value in the most significant tissue
VGLL2	0.17	6	Adipose (Visceral Omentum)	0.08
EXT1	0.04	25	Whole Blood	0.05

* GTEx: The Genotype-Tissue Expression (GTEx) project[52]

Supplementary Table IV. Risk loci in CHARGE-AF risk GWAS

rsID	Chr-Position (hg19)	Effect / Reference Allele	Effect Allele Frequency	Beta	Se	P	Imp Qual	Function	Nearest gene	eQTL for gene(s) in heart tissues	prior GWAS associations	GWAS associations reported for high LD SNVs
rs683600	1-172185145	T/A	0.32	0.01	0.002	2.49E-08	0.941	intronic	*DNM3*		No association reported in GWAS catalog	
rs59985551	2-56106928	T/C	0.23	-0.02	0.003	4.97E-08	0.999	intronic	*EFEMP1*		Fat-free body mass, Hip circumference adjusted for BMI, Waist circumference adjusted for body mass index, Height	Inguinal hernia, Height, Waist circumference adjusted for body mass index, Joint mobility (Beighton score), Hip circumference adjusted for BMI, Optic cup area, Waist circumference adjusted for BMI (adjusted for smoking behaviour), Waist circumference adjusted for BMI (joint analysis main effects and smoking interaction), Waist circumference adjusted for BMI in non-smokers, Waist circumference adjusted for BMI in smokers, Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction), Waist circumference adjusted for BMI in active individuals, Lung function (FVC), Body fat distribution (arm fat ratio), Body fat distribution (leg fat ratio), Body fat distribution (trunk fat ratio), FEV1, Lung function (FEV1/FVC), Carpal tunnel syndrome, Fat-free mass
3:141112859_CTT_C	3-141112859	C/CTT	0.43	0.02	0.002	5.80E-11	0.998	intronic	*ZBTB38*		No association reported in GWAS catalog	Myopia (age of diagnosis), Spherical equivalent or myopia (age of diagnosis), Height, Waist circumference, Estimated glomerular filtration rate, Blond vs. brown/black hair color, Body fat distribution (arm fat ratio), Body fat distribution (leg fat ratio), Body fat distribution (trunk fat ratio), Prostate cancer, Pulmonary function in asthmatics, Infant length, Hip circumference adjusted for BMI, Crohn's disease, Fat-free mass, Hip circumference, Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction), Waist circumference adjusted for BMI in active individuals, Waist circumference adjusted for BMI in inactive individuals, Waist circumference adjusted for body mass index, Eczema, Red blood cell count, Highest math class taken (MTAG), Waist circumference adjusted for BMI (adjusted for smoking behaviour), Waist circumference adjusted for BMI (joint analysis main effects and smoking interaction), Waist circumference adjusted for BMI in non-smokers, Waist circumference adjusted for BMI in smokers, Hair color, Prostate-specific antigen levels (conditioned on lead SNPs), Heel bone mineral density, Multiple sclerosis, Lymphocyte counts, Lymphocyte percentage of white cells, Educational attainment (years of education), Educational attainment (MTAG), Male-pattern baldness, Balding type 1, Chronic obstructive pulmonary disease, Lung function (FVC)
rs140218463	3-100283593	T/G	0.001	0.23	0.039	4.51E-09	0.980	intronic	*TMEM45A*		No association reported in GWAS catalog	
rs10946808	6-26233387	G/A	0.28	-0.02	0.003	3.27E-13	1.000	Noncoding transcript exon variant	*H2AC9P*		Height, Physical activity measurement	Body fat percentage, Height, Cigarettes smoked per day (MTAG), Smoking behaviour (cigarettes smoked per day), Birth weight, Serum urate levels in chronic kidney disease, Hip circumference, Hip circumference adjusted for BMI, Waist circumference, Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction), Waist circumference adjusted for BMI in active individuals, Waist circumference adjusted for body mass index, Parental longevity (combined parental attained age, Martingale residuals), Strenuous sports or other exercises, Body fat

												distribution (leg fat ratio), Body fat distribution (trunk fat ratio), Worry too long after an embarrassing experience
rs202228093	6-34205465	GGAGCCC/G	0.90	-0.03	0.004	1.49E-10	0.975	intronic	*HMGA1*		Hip circumference adjusted for BMI, Waist circumference adjusted for body mass index, Height	Height, Waist-to-hip ratio adjusted for body mass index, Waist circumference adjusted for body mass index, Waist circumference adjusted for BMI (adjusted for smoking behaviour), Waist circumference adjusted for BMI (joint analysis main effects and smoking interaction), Waist circumference adjusted for BMI in non-smokers, Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction), Waist circumference adjusted for BMI in active individuals, Anthropometric traits, Anthropometric traits (multi-trait analysis)
rs7740107	6-130374461	A/T	0.74	-0.02	0.003	3.92E-09	0.998	intronic	*L3MBTL3*	*L3MBTL3*	Height, Red blood cell count, Hematocrit, Appendicular lean mass, Estimated glomerular filtration rate, Body mass index, Hemoglobin, Liver enzyme levels (alkaline phosphatase), Immature fraction of reticulocytes, Hip circumference adjusted for BMI , Lymphocyte counts, Chronic kidney disease, White blood cell count	Fat-free mass, Height, Body mass index, Hip circumference, Lymphocyte counts, Breast cancer, Breast cancer (estrogen-receptor negative), Hair color, Waist-hip ratio, Hip circumference adjusted for BMI, Spherical equivalent or myopia (age of diagnosis), Heel bone mineral density, C-reactive protein levels, Hematocrit, White blood cell count, Red blood cell count, Multiple sclerosis
rs55901622	6-127157438	G/A	0.27	-0.02	0.003	5.04E-09	0.998	intronic	*HLA-DMB*		Hip circumference adjusted for BMI	Bone mineral density (paediatric, upper limb), Intelligence, Femoral neck bone mineral density, Bone mineral density, Bone mineral density (paediatric, total body less head), Total body bone mineral density
rs552707	7-28205303	C/T	0.71	-0.02	0.003	8.39E-14	0.999	intronic	*JAZF1*		BMI-adjusted waist circumference, Physical activity measurement, Height	Height, Hip circumference adjusted for BMI, Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction), Waist circumference adjusted for BMI in active individuals, Waist circumference adjusted for body mass index, Body fat distribution (arm fat ratio), Body fat distribution (leg fat ratio), Body fat distribution (trunk fat ratio), Fat-free mass
rs481237	9-98316094	A/G	0.09	-0.02	0.004	3.47E-08	0.948	intergenic	*PTCH1*		BMI-adjusted hip circumference, BMI-adjusted waist circumference	
rs10878349	12-66327632	G/A	0.51	-0.01	0.002	4.20E-09	0.994	intronic	*HMGA2*		Systolic blood pressure, Cortical surface area	Brain structure, Pulse pressure, Systolic blood pressure, Height, Birth length, Hip circumference, Hip circumference adjusted for BMI, Birth weight, Head circumference (infant), White blood cell count, Chronic obstructive pulmonary disease or resting heart rate (pleiotropy), Infant length, Waist circumference adjusted for body mass index, Waist circumference adjusted for BMI (adjusted for smoking behaviour), Waist circumference adjusted for BMI (joint analysis main effects and smoking interaction), Waist circumference adjusted for BMI in non-smokers, Fat-free mass
rs28559926	15-89400043	C/G	0.04	-0.04	0.006	1.13E-09	0.926	missense variant, NP_001126.3:p.Glu1409Asp	*ACAN*		BMI-adjusted waist circumference	Height, Hip circumference adjusted for BMI, Fat-free mass, Body fat distribution (arm fat ratio), Body fat distribution (leg fat ratio), Body fat distribution (trunk fat ratio)
15:84575867_AAG_A	15-84575867	A/AAG	0.53	0.01	0.002	1.22E-08	0.996	intronic	*ADAMTSL3*		No association reported in GWAS catalog	Height, Hip circumference adjusted for BMI, Waist circumference adjusted for body mass index, Waist circumference adjusted for BMI (adjusted for smoking behaviour), Waist circumference adjusted for BMI (joint analysis main effects and smoking interaction), Waist circumference adjusted for BMI in non-smokers, Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction), Waist circumference adjusted for BMI in active individuals, Hip circumference, Body fat distribution (arm fat ratio), Body fat distribution (leg fat ratio),

												Body fat distribution (trunk fat ratio), Appendicular lean mass, Lean body mass, Waist circumference
rs79461387	17-29168077	T/G	0.26	-0.02	0.003	1.76E-08	0.986	intronic	*ATAD5*		Neutrophil percentage of white cells	
rs4392169	18-20724931	T/A	0.79	0.02	0.003	4.48E-12	0.999	intronic	*CABLES1*		No association reported in GWAS catalog	Heel bone mineral density, Lung function (FVC), FEV1, Height, Scalp hair shape, Fat-free mass, Body fat distribution (leg fat ratio), Body fat distribution (trunk fat ratio), Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction), Waist circumference adjusted for BMI in active individuals, Lung function (FEV1), Telomere length, Waist circumference adjusted for body mass index, Waist circumference adjusted for BMI (adjusted for smoking behaviour), Waist circumference adjusted for BMI (joint analysis main effects and smoking interaction), Waist circumference adjusted for BMI in non-smokers, Waist circumference adjusted for BMI in smokers, Sitting height ratio, Hip circumference adjusted for BMI, Initial pursuit acceleration, Initial pursuit acceleration in psychotic disorders, Waist-hip ratio
rs73047219	19-41875919	G/C	0.08	0.03	0.004	3.99E-09	0.994	intronic	*TMEM91*		No association reported in GWAS catalog	
rs117700741	19-8637832	T/C	0.04	-0.04	0.006	2.65E-08	0.913	intronic	*MYO1F*		No association reported in GWAS catalog	
rs143384	20-34025756	G/A	0.41	0.02	0.002	1.39E-20	1.000	5 Prime UTR variant	*GDF5*	*UQCC1, RPL36P4, NORAD, EDEM2*	BMI-adjusted hip circumference, Height, Body fat distribution, Fat-free body mass, BMI-adjusted waist-hip ratio, Knee osteoarthritis, Hip circumference, Appendicular lean mass, Weight, grip strength measurement, Lung function, Joint mobility, forced expiratory volume, Peak expiratory flow, Subcortical volume, Walking pace, Femoral neck size, Trochanter size, Intertrochanteric region size, Spine bone size, Hip bone size, Knee pain, Brain region volumes, Developmental dysplasia of the hip, Anthropometric traits, Infant length	Height, Waist-to-hip ratio adjusted for BMI (joint analysis for main effect and physical activity interaction), Waist-to-hip ratio adjusted for BMI in active individuals, QRS interval (sulfonylurea treatment interaction), Spine bone size, Developmental dysplasia of the hip, Osteoarthritis, Waist-to-hip ratio adjusted for body mass index, Waist-to-hip ratio adjusted for BMI (additive genetic model), Waist-to-hip ratio adjusted for BMI, Waist-to-hip ratio adjusted for BMI x sex x age interaction (4df test), Pulse pressure, Infant length, Joint mobility (Beighton score), Hip circumference, Hip circumference adjusted for BMI, Knee osteoarthritis, Fat-free mass, Lung function (FVC), Waist-hip ratio, Body fat distribution (leg fat ratio), Body fat distribution (trunk fat ratio), FEV1, Peak expiratory flow, Anthropometric traits, Anthropometric traits (multi-trait analysis)
rs73901538	20-20383074	T/C	0.002	0.22	0.041	4.26E-08	0.623	intronic	*RALGAPA2*		No association reported in GWAS catalog	

Genetic variants at risk loci independently associated with ECG-AI predicted 5-year AF risk.

rsID: variant ID; **Chr:** chromosome; **hg19:** genomic position based on build hg19; **Beta:** effect size of effect allele; **Se:** standard error; **P:** P-value in GWAS; **Imp Qual:** imputation quality.

eQTL for gene(s) in heart tissues: eQTL reported by the GTEx V8 data in Heart - Atrial Appendage tissue or Heart - Left Ventricle tissue.

GWAS associations reported for high LD SNVs: Traits associated with SNVs in high linkage disequilibrium (r2>=0.7) with the lead SNV as reported in GWAS catalog.

Supplementary Table V. Traits associated with lead SNVs identified in the CHARGE GWAS

CHARGE-AF variables	SNVs	Associated traits	PMID
Age			
Race			
Height	**rs10946808**	Height	18391951, 18391950, 19343178
	rs7740107	Height, Body mass index	25282103
	rs552707	Height	25282103
	rs143384	Height	20881960, 25282103
	rs4146922 (rs59985551)	Height	25429064
	rs3791679 (rs59985551)	Height	18391951, 20189936, 23563607, 25282103
	rs3791675 (rs59985551)	Height	18391952, 19396169, 19893584, 20397748, 20881960, 25429064
	rs6440003 (3:141112859_CTT_C)	Height	18391952, 21998595
	rs6763931 (3:141112859_CTT_C)	Height	18391951, 19343178, 25429064
	rs724016 (3:141112859_CTT_C)	Height	18391950, 18193045, 20881960, 25282103
	rs1991431 (3:141112859_CTT_C)	Height	23563607, 28270201
	rs7766641 (rs10946808)	Height	28270201
	rs806794 (rs10946808)	Height	20881960, 23563607, 25429064, 25282103
	rs12214804 (rs202228093)	Height	25282103
	rs1776897 (rs202228093)	Height	18391951, 19343178, 20397748, 25429064
	rs2780226 (rs202228093)	Height	20881960
	rs1150781 (rs202228093)	Height	21998595
	rs6569648 (rs7740107)	Height	20881960
	rs6899976 (rs7740107)	Height	18391951
	rs849141 (rs552707)	Height	19343178
	rs1029534 (rs552707)	Height	23563607
	rs1708299 (rs552707)	Height	20881960
	rs1351394 (rs10878349)	Height	20881960, 23563607
	rs1042725 (rs10878349)	Height	18391950, 18391952, 17767157, 25429064
	rs8756 (rs10878349)	Height	18391951, 19343178, 20397748, 25282103
	rs7968682 (rs10878349)	Height	22021425
	rs7162542 (15:84575867_AAG_A)	Height	25282103
	rs2401171 (15:84575867_AAG_A)	Height	25429064
	rs10906982 (15:84575867_AAG_A)	Height	18391952, 20397748

	rs7183263 (15:84575867_AAG_A)	Height	20397748
	rs11259933 (15:84575867_AAG_A)	Height	23563607
	rs11259936 (15:84575867_AAG_A)	Height	20881960
	rs4842838 (15:84575867_AAG_A)	Height	19343178, 20189936
	rs16942341 (rs28559926)	Height	20881960
	rs8096254 (rs4392169)	Height	28270201
	rs4800148 (rs4392169)	Height	18391951
	rs4800452 (rs4392169)	Height	20881960
	rs4369779 (rs4392169)	Height	20189936, 25429064, 25282103
	rs6060355 (rs143384)	Height	28270201
	rs6060369 (rs143384)	Height	18391950, 18193045, 25429064
	rs6060373 (rs143384)	Height	18391952
	rs6088813 (rs143384)	Height	19343178
	rs6060402 (rs143384)	Height	28270201
	rs224329 (rs143384)	Height	25429064
	rs224333 (rs143384)	Height	23563607
Weight	**rs59985551**	Fat-free body mass	30593698
	rs143384	Fat-free body mass	30593698
	rs724016 (3:141112859_CTT_C)	Fat-free mass	30593698
	rs7740188 (rs7740107)	Fat-free mass	30593698
	rs6569648 (rs7740107)	Body mass index	25673413, 30108127
	rs9375702 (rs7740107)	Body mass index	30595370
	rs508347 (rs552707)	Fat-free mass	30593698
	rs9669278 (rs10878349)	Fat-free mass	30593698
	rs16942341 (rs28559926)	Fat-free mass	30593698
	rs4800148 (rs4392169)	Fat-free mass	30593698
Systolic Blood Pressure	rs10784502 (rs10878349)	Systolic blood pressure, Pulse pressure	27841878
	rs7970350 (rs10878349)	Systolic blood pressure	30595370
	rs113744258 (rs10878349)	Systolic blood pressure, Pulse pressure	30578418
Diastolic Blood Pressure	rs10784502 (rs10878349)	Pulse pressure	27841878
	rs113744258 (rs10878349)	Pulse pressure	30578418
	rs7312464 (rs10878349)	Pulse pressure	28135244
	rs6142381 (rs143384)	Pulse pressure	30578418

Smoking	rs7766641 (rs10946808)	Smoking behaviour (cigarettes smoked per day)	30643251
Antihypertensive medication use			
Diabetes			
Heart failure			
Myocardial infarction			

Supplementary Table VI. Summary statistics of the independent significant lead single nucleotide variants (SNVs) from each GWAS in the other GWAS.

SNVs in ECG-AI risk GWAS	Genome Position (hg19)	Beta (se) in ECG-AI risk GWAS	P in ECG-AI risk GWAS	Beta (se) in CHARGE risk GWAS	P in CHARGE risk GWAS
rs2627040	2-179601975	0.077 (0.014)	4.55×10^{-8}	0.004 (0.005)	0.47
rs41312411	3-38621237	0.055 (0.009)	2.79×10^{-10}	-0.005 (0.003)	0.17
rs6801957	3-38767315	-0.044 (0.006)	2.68×10^{-12}	-0.007 (0.002)	0.003
rs9689288	6-117508467	0.043 (0.006)	5.05×10^{-12}	0.009 (0.002)	2.97×10^{-4}
rs35186392	8-118857805	0.045 (0.008)	7.22×10^{-9}	-0.0004 (0.003)	0.90
SNVs in CHARGE risk GWAS	**Genome Position (hg19)**	**Beta (se) in CHARGE risk GWAS**	**P in CHARGE risk GWAS**	**Beta (se) in ECG-AI risk GWAS**	**P in ECG-AI risk GWAS**
rs683600	1-172185145	0.015 (0.003)	2.49×10^{-8}	-0.006 (0.007)	0.40
rs59985551	2-56106928	-0.016 (0.003)	4.97×10^{-8}	-0.005 (0.007)	0.50
3:141112859_CTT_C	3-141112859	0.016 (0.002)	5.80×10^{-11}	-0.016 (0.006)	0.01
rs140218463	3-100283593	0.228 (0.039)	4.51×10^{-9}	0.145 (0.100)	0.15
rs10946808	6-26233387	-0.020 (0.003)	3.27×10^{-13}	-0.006 (0.007)	0.37
rs202228093	6-34205465	-0.025 (0.004)	1.49×10^{-10}	-0.020 (0.010)	0.05
rs7740107	6-130374461	-0.016 (0.003)	3.92×10^{-9}	-0.011 (0.007)	0.12
rs55901622	6-127157438	-0.016 (0.003)	5.04×10^{-9}	-0.014 (0.007)	0.04
rs552707	7-28205303	-0.020 (0.003)	8.39×10^{-14}	-0.004 (0.007)	0.58
rs481237	9-98316094	-0.024 (0.004)	3.47×10^{-8}	-0.025 (0.011)	0.03
rs10878349	12-66327632	-0.014 (0.002)	4.20×10^{-9}	-0.011 (0.006)	0.09
rs28559926	15-89400043	-0.038 (0.006)	1.13×10^{-9}	-0.024 (0.016)	0.13
15:84575867_AAG_A	15-84575867	0.014 (0.002)	1.22×10^{-8}	0.008 (0.006)	0.20
rs79461387	17-29168077	-0.015 (0.003)	1.76×10^{-8}	-0.010 (0.007)	0.15
rs4392169	18-20724931	0.020 (0.003)	4.48×10^{-12}	0.001 (0.007)	0.91
rs73047219	19-41875919	0.026 (0.004)	3.99×10^{-9}	0.018 (0.011)	0.11
rs117700741	19-8637832	-0.036 (0.006)	2.65×10^{-8}	-0.035 (0.017)	0.04
rs143384	20-34025756	0.023 (0.002)	1.39×10^{-20}	-0.0001(0.006)	0.98
rs73901538	20-20383074	0.223 (0.041)	4.26×10^{-8}	0.259 (0.104)	0.01

Supplementary Table VII. Heritability and genetic correlation estimates for ECG-AI, CHARGE-AF, and previously reported AF GWAS.

Traits	**h^2 (GWAS sumstats)**	**h^2 (individual-level data)**
ECG-AI risk	7.0% (1.3%)	13.0% (1.4%)
CHARGE risk	24.3% (2.1%)	36.5% (1.4%)
	r_g (GWAS sumstats)	**r_g (individual-level data)**
AF and ECG-AI risk	41.9% (7.3%)	35.3% (13.7%)
AF and CHARGE risk	28.2% (3.6%)	18.9% (8.6%)
ECG-AI risk and CHARGE risk	50.1% (7.4%)	39.3% (4.5%)

h^2: heritability; sumstats: summary statistics; r_g: genetic correlation. The previously reported AF GWAS summary statistics are derived from Roselli et al (2018).[11]

Supplementary Table VIII. C-index of Cox Proportional Hazard models testing polygenic risk scores of ECG-AI and CHARGE-AF scores

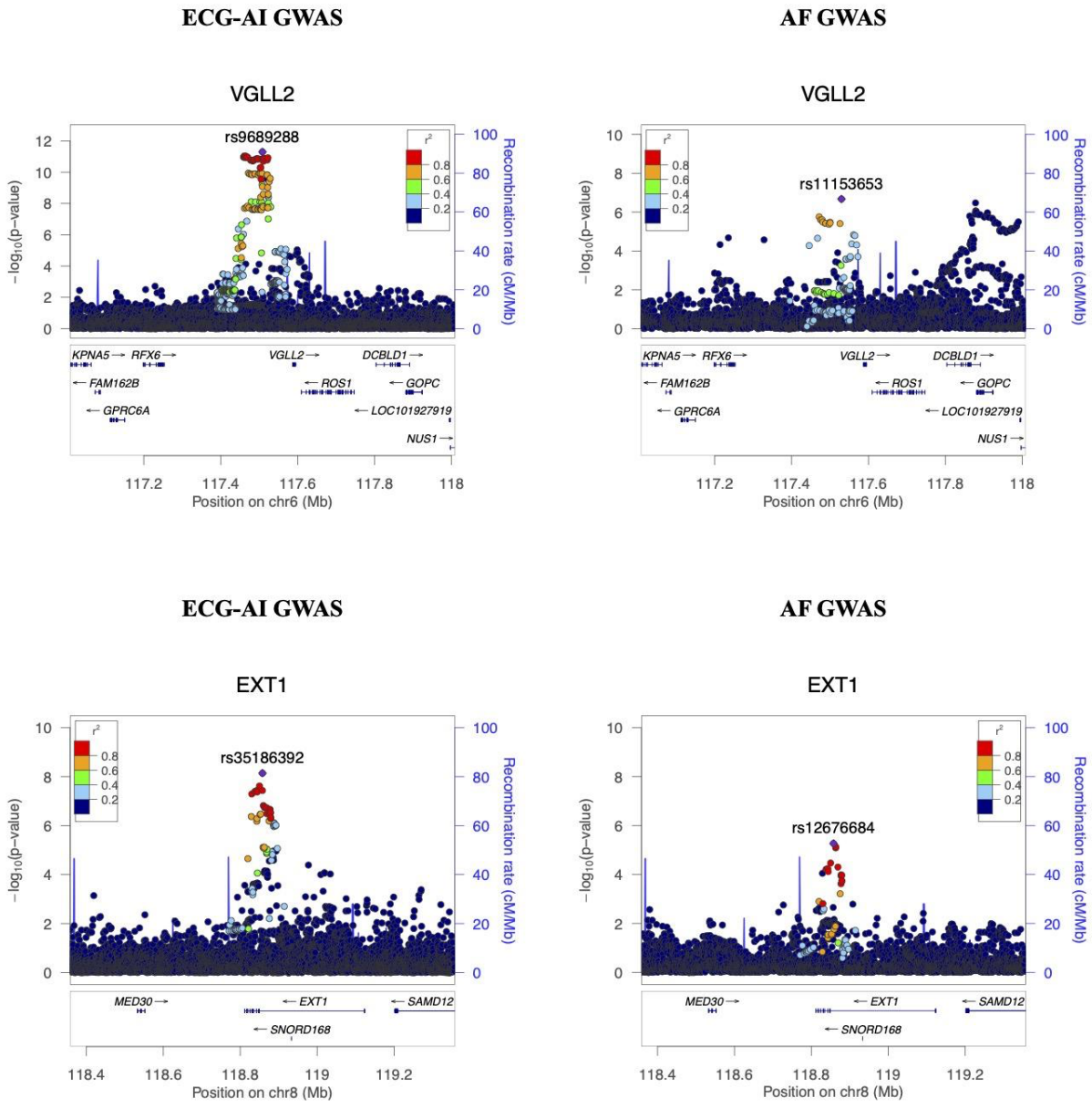
Model	C-index
Baseline (age + sex)	0.7373
Baseline + ECG-AI PRS	0.7380
Baseline + CHARGE-AF PRS	0.7390
Baseline + ECG-AI PRS + CHARGE-AF PRS	0.7394

Supplementary Table IX. Polygenic risk and incident AF in a subset of White British participants (N=370,121).

	Individual models		Interaction model	
	HR (95% CI)	**P-value**	**HR [95% CI]**	**P-value**
$PRS_{ECG\text{-}AI}$	1.08 (1.05-1.10)	2.41×10^{-9}	1.07 (1.04-1.10)	1.29×10^{-7}
$PRS_{CHARGE\text{-}AF}$	1.12 (1.09-1.15)	3.57×10^{-19}	1.11 (1.09-1.14)	1.75×10^{-17}
Interaction term			1.00 (0.98-1.02)	0.98

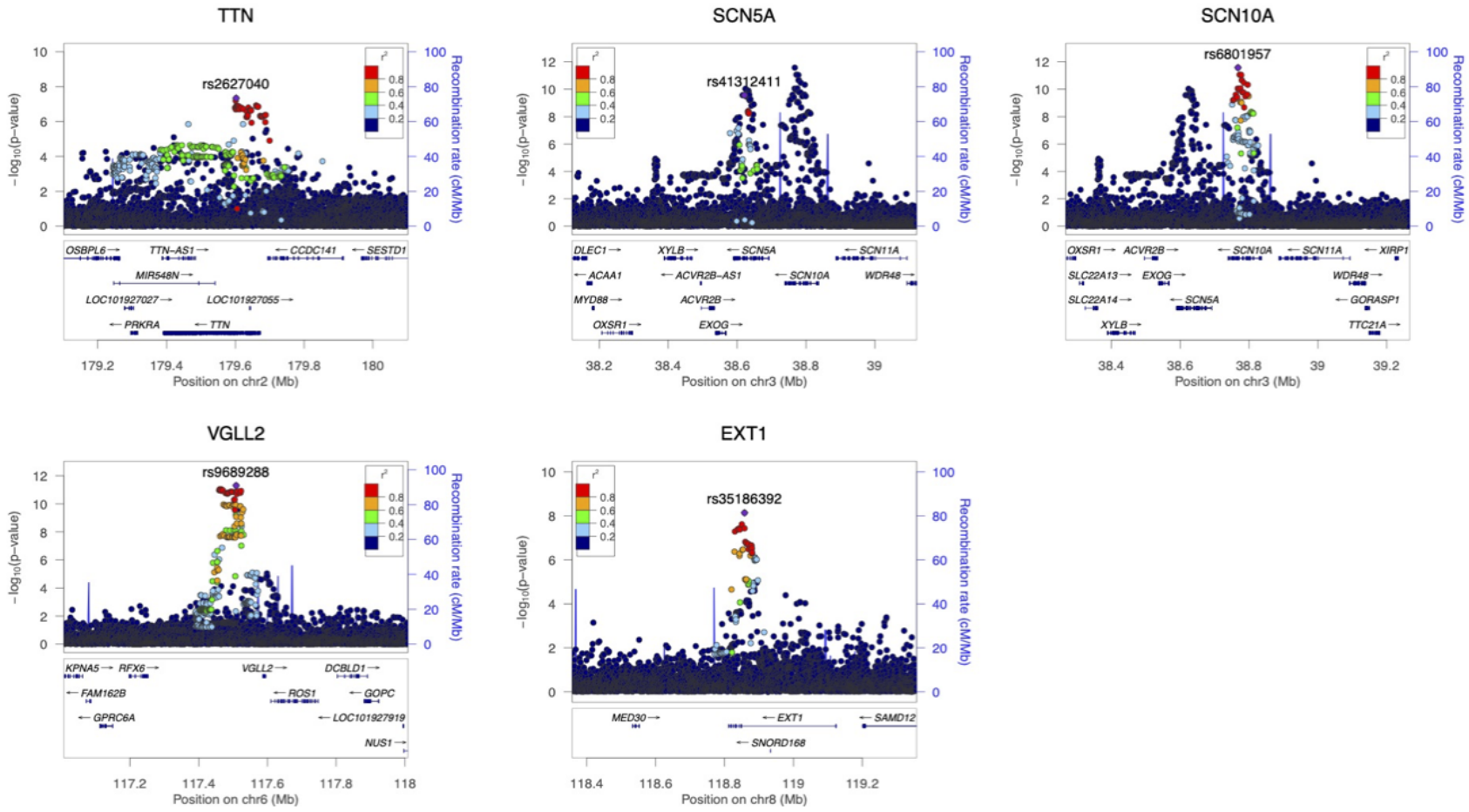
In the individual models, each PRS was tested separately as a predictor of AF, adjusting for age at enrollment, sex, genotyping array, and the first 20 principal components of ancestry. In the interaction model, main effect terms for each PRS were included along with a multiplicative interaction term between each PRS, adjusting for the above covariates. HR: hazard ratio; CI: confidence interval.

Supplementary Figure I. LocusZoom plots comparing ECG-AI GWAS and reference AF GWAS* at *VGLL2* and *EXT1* loci



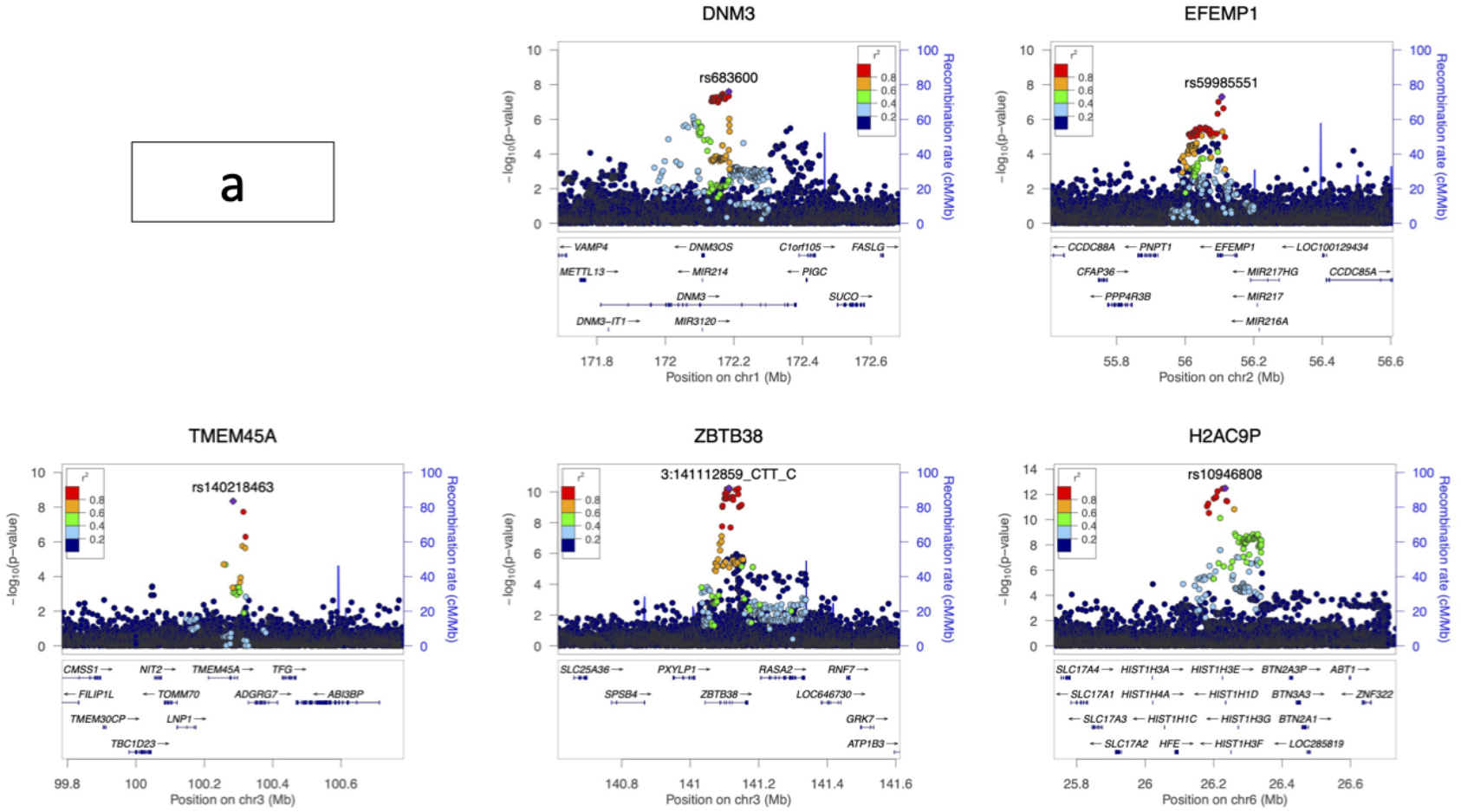
* Reference AF GWAS[1]

Supplementary Figure II. LocusZoom plots of significant risk loci in the ECG-AI predicted 5-year AF risk GWAS.

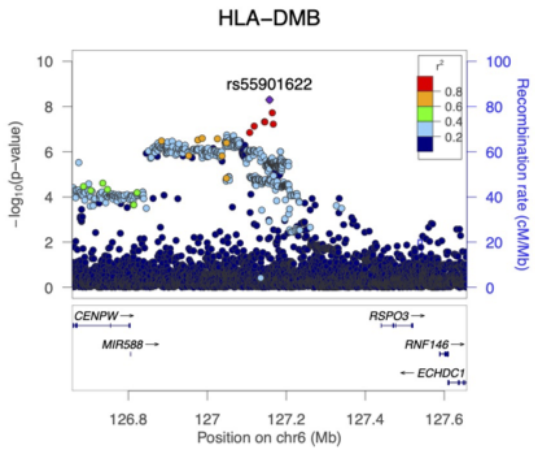


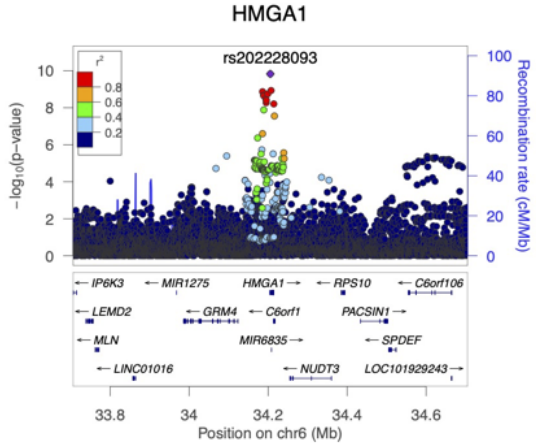
Supplementary Figure III. LocusZoom plots of significant risk loci in the CHARGE-AF predicted 5-year AF risk GWAS.

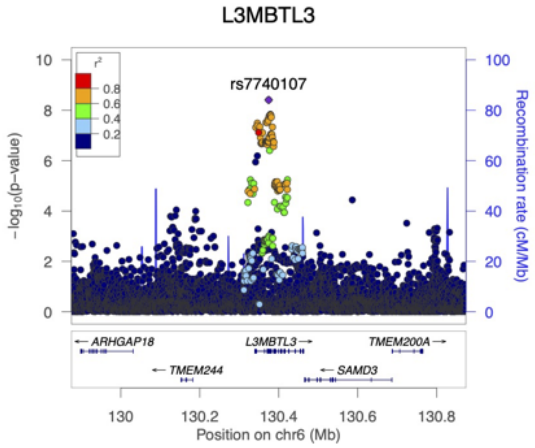
a

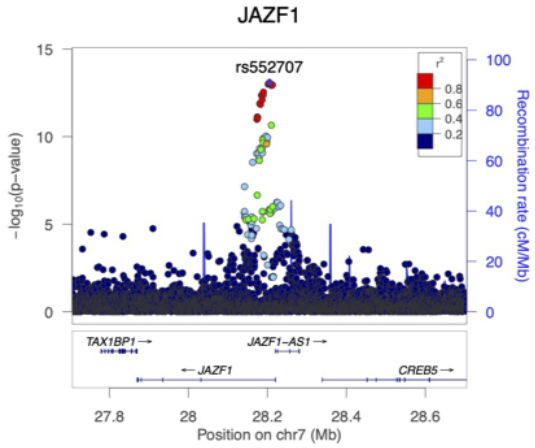


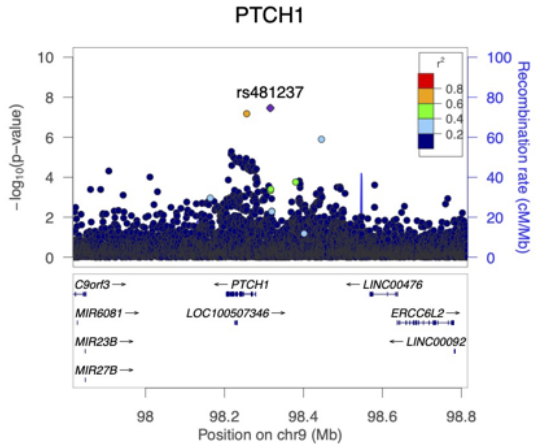
b


HLA–DMB
rs55901622
−log10(p−value)
Recombination rate (cM/Mb)
CENPW
RSPO3
MIR588
RNF146
ECHDC1
126.8
127
127.2
127.4
127.6
Position on chr6 (Mb)

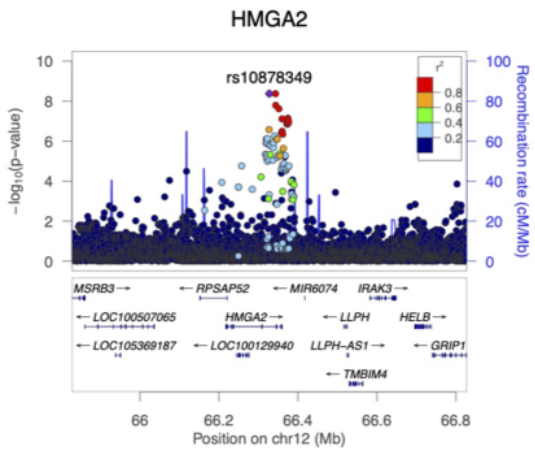

HMGA1
rs202228093
−log10(p−value)
Recombination rate (cM/Mb)
IP6K3
MIR1275
HMGA1
RPS10
C6orf106
LEMD2
GRM4
C6orf1
PACSIN1
MLN
MIR6835
SPDEF
LINC01016
NUDT3
LOC101929243
33.8
34
34.2
34.4
34.6
Position on chr6 (Mb)

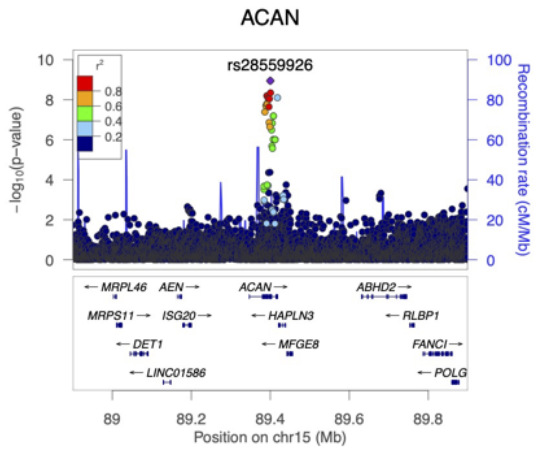

L3MBTL3
rs7740107
−log10(p−value)
Recombination rate (cM/Mb)
ARHGAP18
L3MBTL3
TMEM200A
TMEM244
SAMD3
130
130.2
130.4
130.6
130.8
Position on chr6 (Mb)

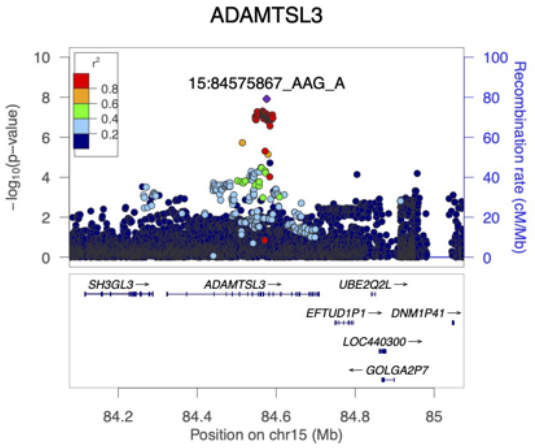

JAZF1
rs552707
−log10(p−value)
Recombination rate (cM/Mb)
TAX1BP1
JAZF1−AS1
JAZF1
CREB5
27.8
28
28.2
28.4
28.6
Position on chr7 (Mb)

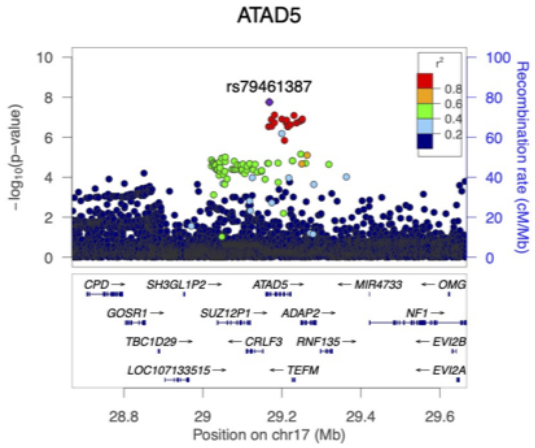

PTCH1
rs481237
−log10(p−value)
Recombination rate (cM/Mb)
C9orf3
PTCH1
LINC00476
MIR6081
LOC100507346
ERCC6L2
MIR23B
LINC00092
MIR27B
98
98.2
98.4
98.6
98.8
Position on chr9 (Mb)

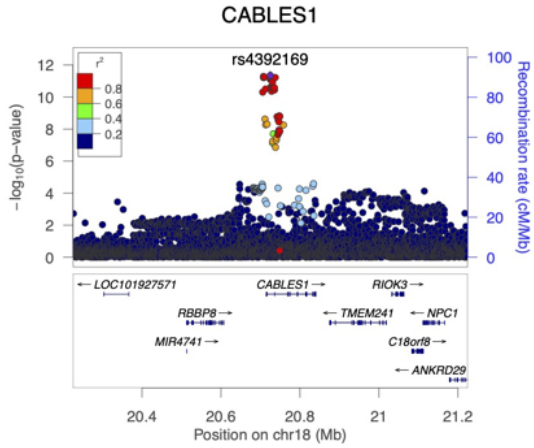
C


HMGA2
rs10878349
MSRB3 RPSAP52 MIR6074 IRAK3
LOC100507065 HMGA2 LLPH HELB
LOC105369187 LOC100129940 LLPH-AS1 GRIP1
TMBIM4
Position on chr12 (Mb)
Recombination rate (cM/Mb)

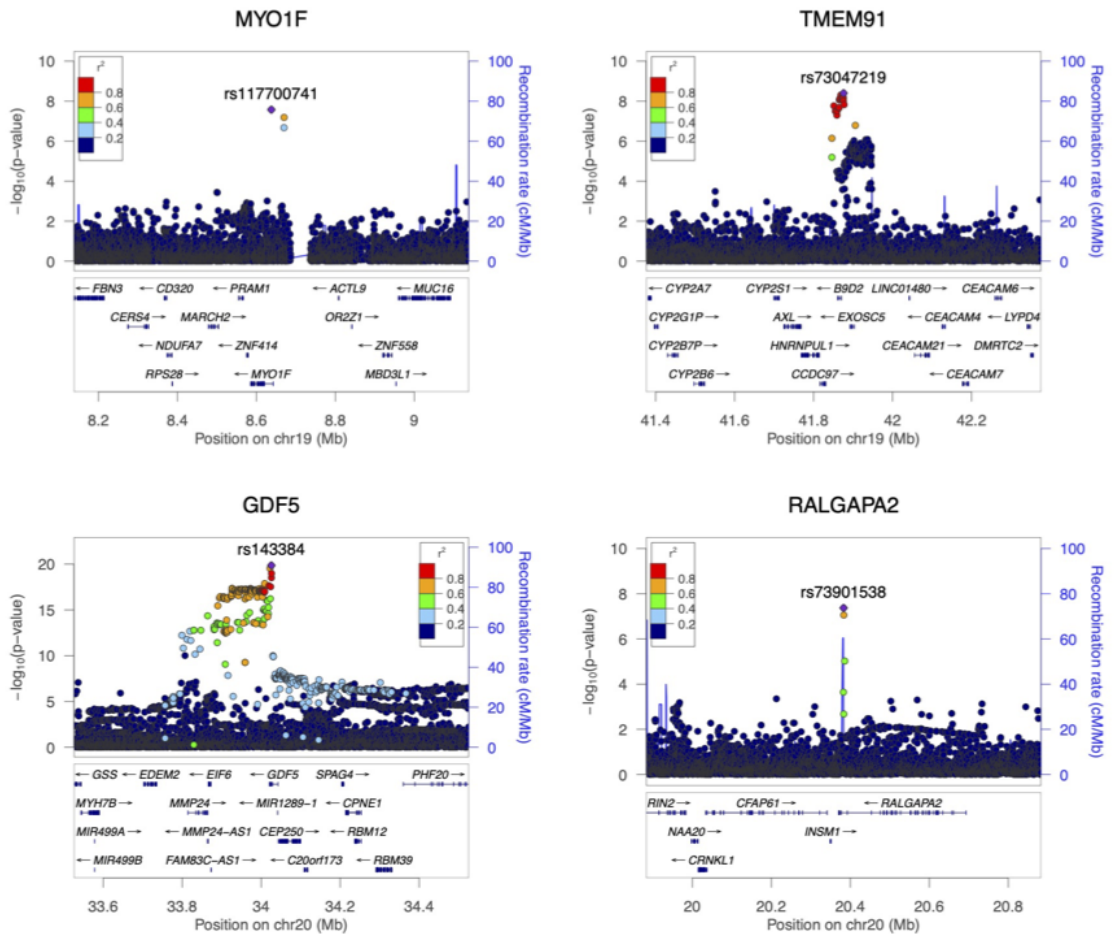

ACAN
rs28559926
MRPL46 AEN ACAN ABHD2
MRPS11 ISG20 HAPLN3 RLBP1
DET1 MFGE8 FANCI
LINC01586 POLG
Position on chr15 (Mb)
Recombination rate (cM/Mb)


ADAMTSL3
15:84575867_AAG_A
SH3GL3 ADAMTSL3 UBE2Q2L
EFTUD1P1 DNM1P41
LOC440300
GOLGA2P7
Position on chr15 (Mb)
Recombination rate (cM/Mb)

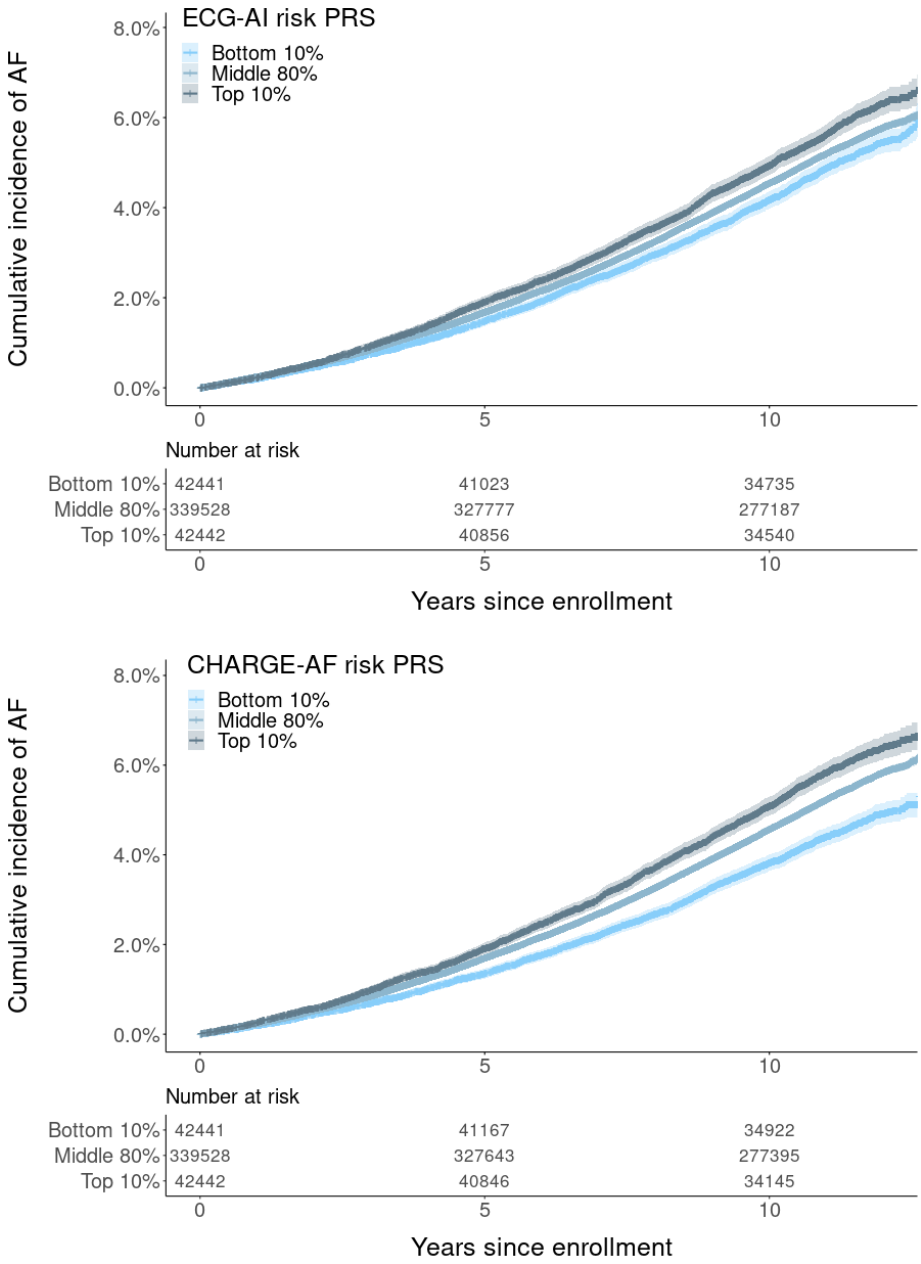

ATAD5
rs79461387
CPD SH3GL1P2 ATAD5 MIR4733 OMG
GOSR1 SUZ12P1 ADAP2 NF1
TBC1D29 CRLF3 RNF135 EVI2B
LOC107133515 TEFM EVI2A
Position on chr17 (Mb)
Recombination rate (cM/Mb)


CABLES1
rs4392169
LOC101927571 CABLES1 RIOK3
RBBP8 TMEM241 NPC1
MIR4741 C18orf8
ANKRD29
Position on chr18 (Mb)
Recombination rate (cM/Mb)

d

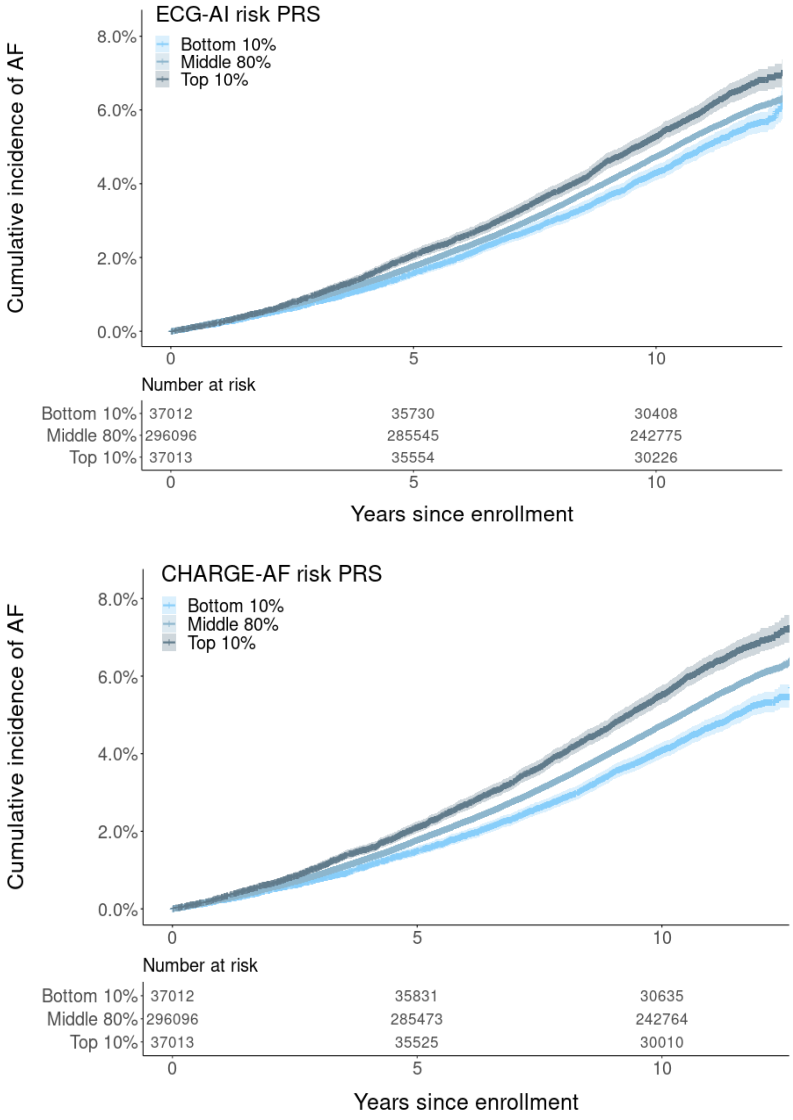

MYO1F
rs117700741
FBN3 CD320 PRAM1 ACTL9 MUC16
CERS4 MARCH2 OR2Z1
NDUFA7 ZNF414 ZNF558
RPS28 MYO1F MBD3L1
Position on chr19 (Mb)
TMEM91
rs73047219
CYP2A7 CYP2S1 B9D2 LINC01480 CEACAM6
CYP2G1P AXL EXOSC5 CEACAM4 LYPD4
CYP2B7P HNRNPUL1 CEACAM21 DMRTC2
CYP2B6 CCDC97 CEACAM7
Position on chr19 (Mb)
GDF5
rs143384
GSS EDEM2 EIF6 GDF5 SPAG4 PHF20
MYH7B MMP24 MIR1289-1 CPNE1
MIR499A MMP24-AS1 CEP250 RBM12
MIR499B FAM83C-AS1 C20orf173 RBM39
Position on chr20 (Mb)
RALGAPA2
rs73901538
RIN2 CFAP61 RALGAPA2
NAA20 INSM1
CRNKL1
Position on chr20 (Mb)
Recombination rate (cM/Mb)

Supplementary Figure IV. ECG-AI and CHARGE-AF polygenic risk and cumulative incidence of AF.



The cumulative incidence of AF is plotted for participants in three risk groups (lowest 10%, middle 80%, and highest 10% of the PRS distribution). Shaded regions around Kaplan-Meier survival estimates represent the 95% confidence intervals.

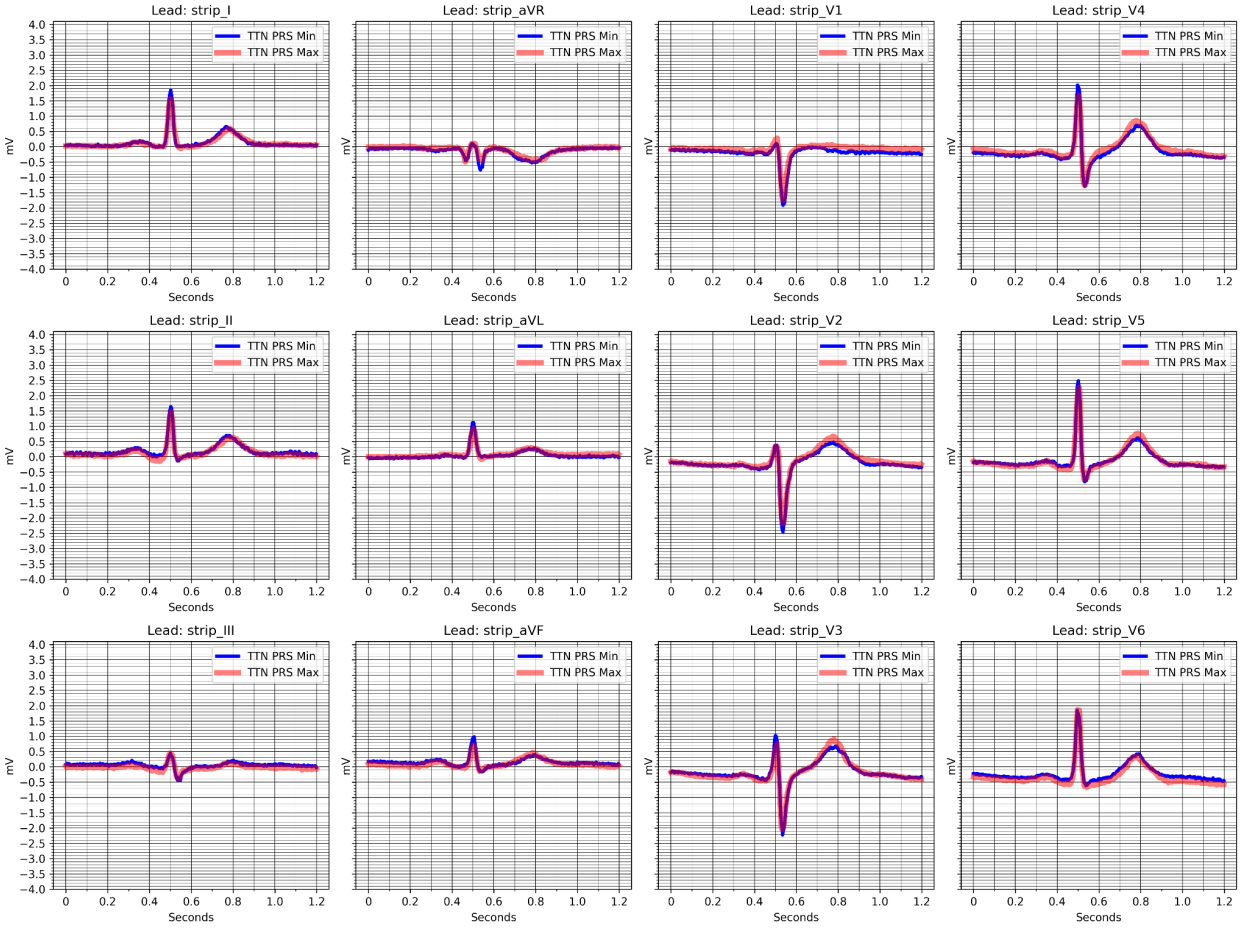
Supplementary Figure V. ECG-AI and CHARGE-AF polygenic risk and cumulative incidence of AF in a subset of White British participants (N=370,121).



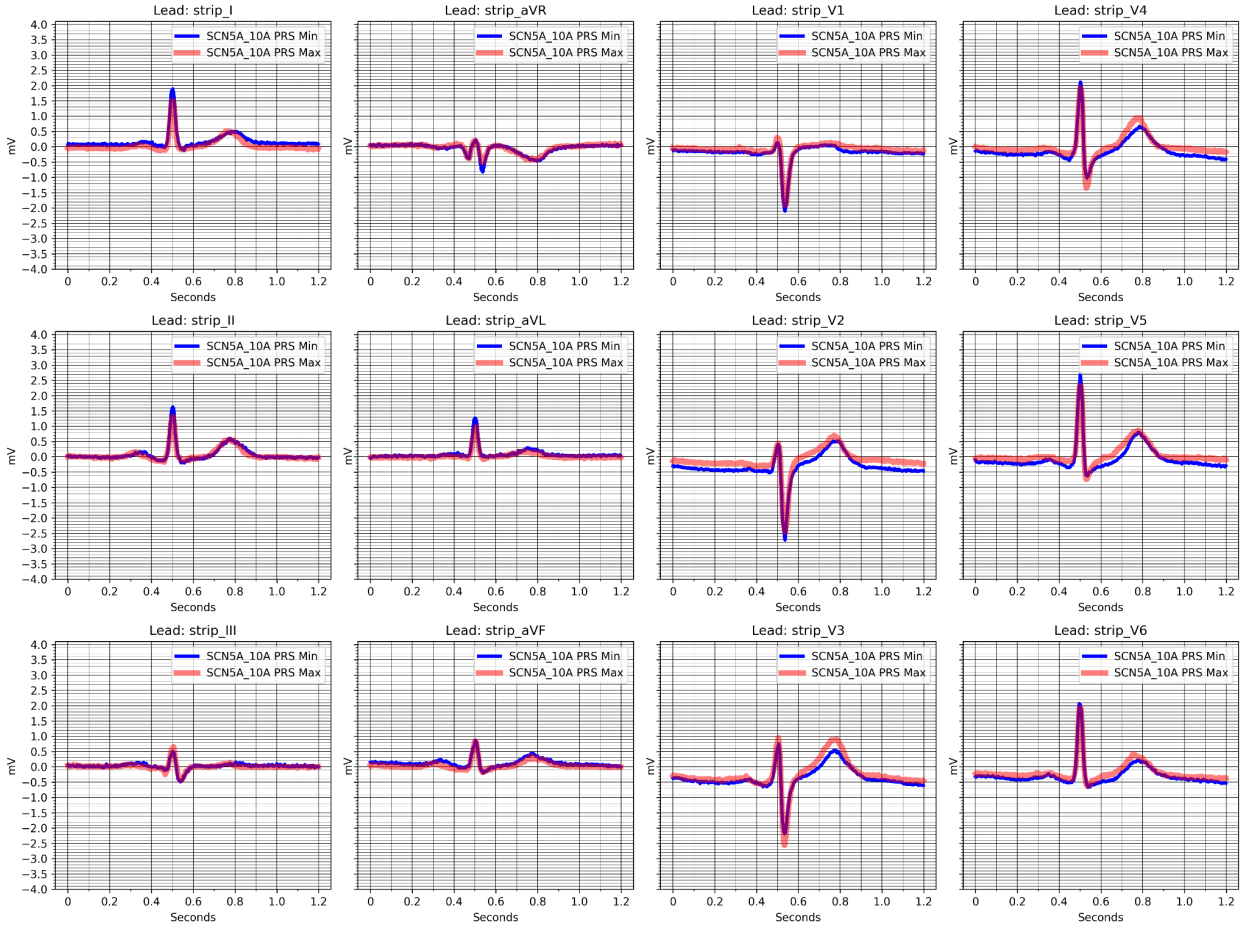
The cumulative incidence of AF is plotted for participants in three risk groups (lowest 10%, middle 80%, and highest 10% of the PRS distribution). Shaded regions around Kaplan-Meier survival estimates represent the 95% confidence intervals.

Supplementary Figure VI. Median ECG waveforms for individuals at highest and lowest ECG-AI genetic risk at significant loci implicated in ECG-AI GWAS

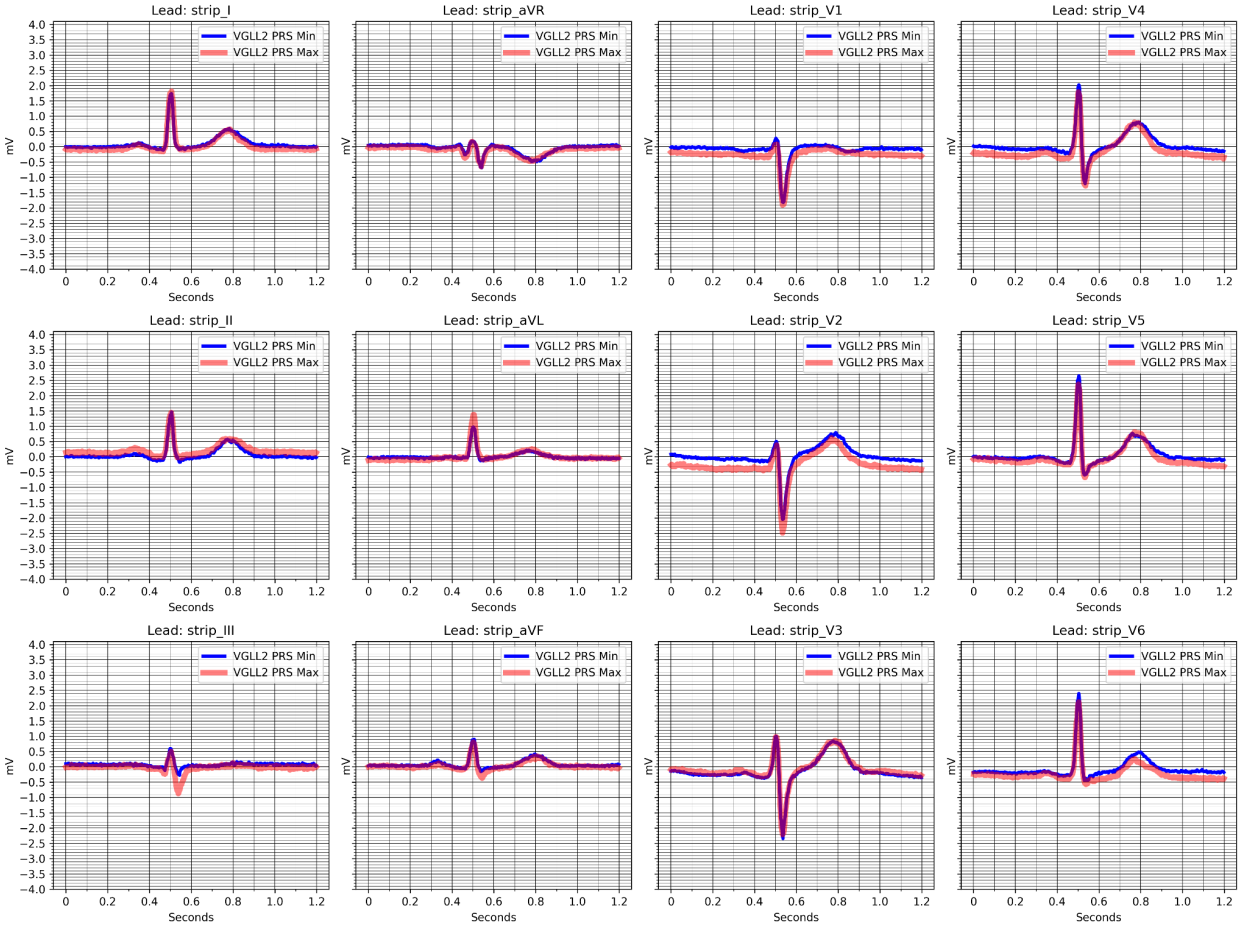
a. ***TTN*** **region**



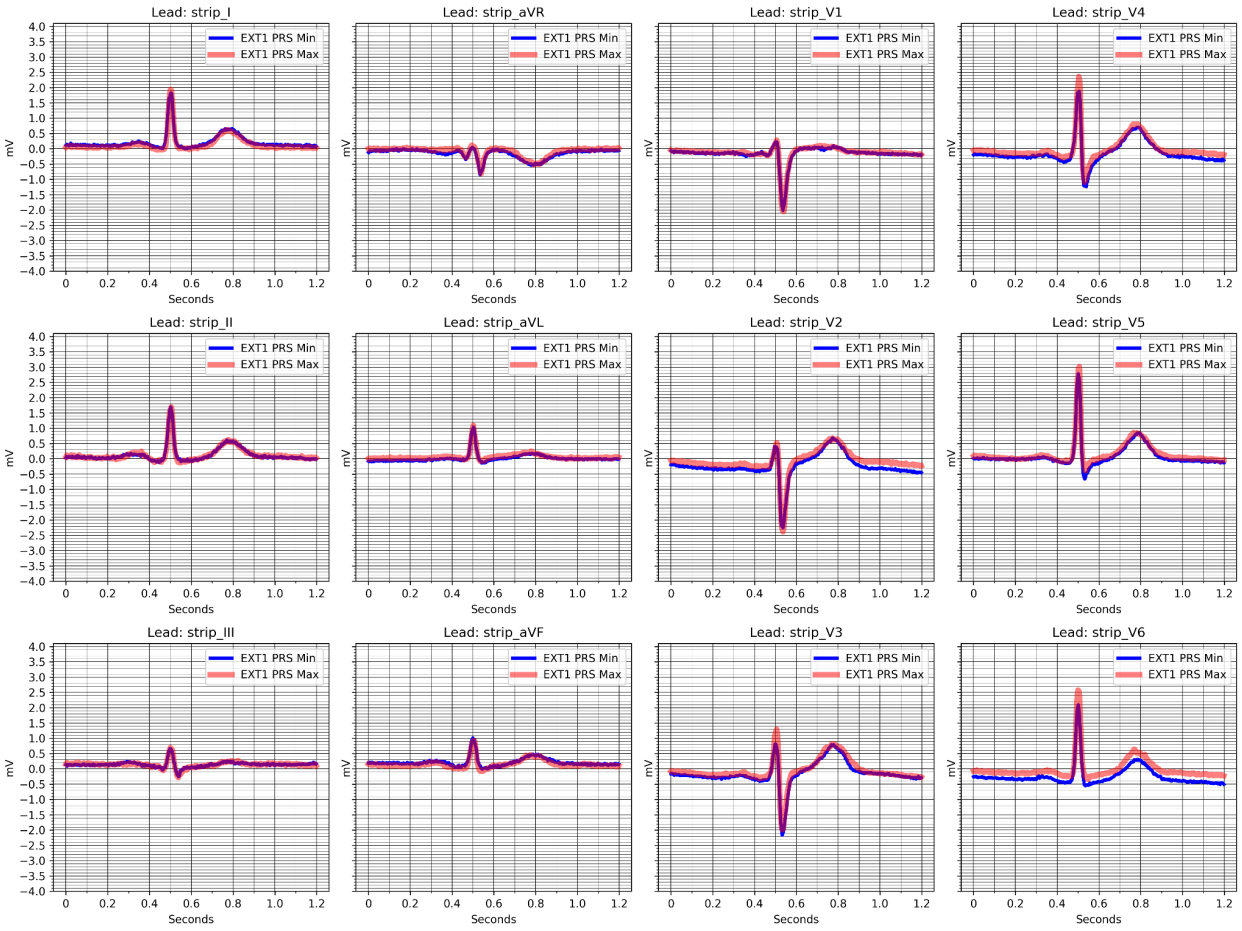
b. *SCN5A/10A* region



c. *VGLL2* region



d. *EXT1* region



Red ECGs represent individuals in the highest 1% risk group and blue ECGs represent individuals in the lowest 1% risk group. The risk groups were defined using a PRS reflecting the SNPs from the ECG-AI GWAS in the corresponding gene region ± 500 kb.