Genetic Susceptibility to Atrial Fibrillation Identified via Deep Learning of 12-lead Electrocardiograms

Xin Wang, MPH^{1,2}; Shaan Khurshid, MD, MPH^{1,2,3}; Seung Hoan Choi, PhD²; Sam Friedman, PhD⁴; Lu-Chen Weng, PhD^{1,2}; Christopher Reeder, PhD⁴; James P. Pirruccello, MD^{1,2,3}; Pulkit Singh, BA⁴; Emily S. Lau, MD^{1,2,3}; Rachael Venn, MD^{1,3}; Nate Diamant, BS⁴; Paolo Di Achille, PhD⁴; Anthony Philippakis, MD, PhD^{4,9}; Christopher D. Anderson, MD^{6,7,8}; Jennifer E. Ho, MD¹⁰; Patrick T. Ellinor, MD, PhD^{1,2,5}; Puneet Batra, PhD⁴; Steven A. Lubitz, MD, MPH^{1,2,5}

¹Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA

²Cardiovascular Disease Initiative, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

³Division of Cardiology, Massachusetts General Hospital, Boston, Massachusetts, USA

⁴Data Sciences Platform, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

⁵Demoulas Center for Cardiac Arrhythmias, Massachusetts General Hospital, Boston, Massachusetts, USA

⁶Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, USA

⁷Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

⁸Henry and Allison McCance Center for Brain Health, Massachusetts General Hospital, Boston, Massachusetts, USA

⁹Eric and Wendy Schmidt Center, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

¹⁰CardioVascular Institute and Division of Cardiology, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

Corresponding author: Steven A. Lubitz, MD, MPH; Demoulas Center for Cardiac

Arrhythmias and Cardiovascular Research Center, Massachusetts General Hospital, 55 Fruit

Street, GRB 109, Boston, MA 02114; P:617-643-7339; F:617-726-3852;

slubitz@mgh.harvard.edu

Abstract

Artificial intelligence (AI) models applied to 12-lead electrocardiogram (ECG) waveforms can predict atrial fibrillation (AF), a heritable and morbid arrhythmia. We hypothesized that there may be a genetic basis for ECG-AI based risk estimates. We applied an ECG-AI model for predicting incident AF to ECGs from 39,986 UK Biobank participants without AF. We then performed a genome-wide association study (GWAS) of the predicted AF risk. We identified three signals ($P < 5 \times 10^{-8}$) at established AF susceptibility loci marked by the sarcomeric gene *TTN*, and sodium channel genes *SCN5A* and *SCN10A*. We also identified two novel loci near the genes *VGLL2* and *EXT1*. In contrast, a GWAS of risk estimates from a clinical variable model indicated a different genetic profile. Predicted AF risk from an ECG-AI model is influenced by genetic variation implicating sarcomeric, ion channel, and height pathways. ECG-AI models may identify individuals at risk for disease via specific biological pathways.

Introduction

Atrial fibrillation (AF) is a heritable arrhythmia associated with substantial morbidity, including stroke, heart failure, dementia, and mortality.^{1–3} Identifying individuals at high risk of developing AF may enable early detection via cardiac rhythm monitoring and treatment, or behavioral modification to prevent AF altogether. Artificial intelligence (AI) algorithms applied to 12-lead electrocardiogram (ECG) waveforms can predict AF.^{4–6} Algorithms that predict AF risk from ECGs have practical appeal given the ubiquity and inexpensive nature of ECGs, and lack of requirement for manual data input for risk estimation. Whether risk estimates derived from AI algorithms reflect specific underlying genetic pathways that increase susceptibility to AF is unclear.

Understanding the biological basis for risk estimates from machine learning models could aid model interpretability, rationalize model outputs, promote clinician confidence, and potentially enable identification of individuals with specific mechanistic pathways that lead to AF. We recently developed and validated an AI algorithm for predicting the 5-year risk of new-onset AF using 12-lead ECGs (“ECG-AI”).⁷ In the present study, we conducted genetic association testing with AF risk estimates generated from the ECG-AI model to assess the genetic underpinnings reflected by the output. As a comparator, we assessed the genetic basis of a widely validated clinical risk factor model for predicting AF, the CHARGE-AF score (Cohorts for Aging and Research in Genomic Epidemiology–AF).⁸

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**. All UK Biobank participants provided electronic signed consent at recruitment, and the study protocol was approved by the UK Biobank Research Ethics Committee (reference number 11/NW/0382). Use of data (under UK Biobank application 7089) for the current study was approved by the Mass General Brigham (MGB) Institutional Review Board.

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.¹¹ 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-index¹² 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.^{8,13,14} 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.¹⁵ 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. We used genotyped markers to fit null models for the GWAS. Variants were filtered out if MAF was $< 0.1\%$, minor allele counts (MAC) was < 100 , missing call rate was > 0.1 , or Hardy-Weinberg equilibrium exact test $P < 1 \times 10^{-5}$. 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,¹⁶ 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 analyses¹⁷ 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 models¹⁸ implemented in the *coxme* 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 $P < 5 \times 10^{-8}$. 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.¹⁹ 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)²⁰ 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.²¹ 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.²² 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,²³ 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.

Mendelian Randomization

We employed a two-sample Mendelian randomization (2SMR) approach²⁴ implemented in the *TwoSampleMR*²⁵ 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 $P < 5 \times 10^{-8}$. 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 duration²⁶ (unit: ms) and height²⁷ (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,²⁸ MR-PRESSO,²⁹ and MR-PRESSO corrected for horizontal pleiotropic outliers.

Results

Sample characteristics

The mean age of the 39,986 participants included in the GWAS was 64.0 +/- 7.7 years at the time of ECG acquisition and 52% were female. The median follow-up time (starting from ECG visit) for this subset was 2.8 years (quartile 1: 1.9, quartile 3: 4.3). 510 individuals developed incident AF within 5 years of follow-up, corresponding to a cumulative incidence of AF of 2.12%. The mean age of the 424,411 participants included in the polygenic risk score application analysis who did not have ECGs was 57.1 +/- 8.1 years at the time of study enrollment and 55% were female. The median follow-up time (starting from study enrollment) for this subset was 11.1 years (quartile 1: 10.4, quartile 3: 11.8). 7,077 individuals developed incident AF within 5 years of follow up, corresponding to a cumulative AF incidence of 1.70%. Participant characteristics are presented in **Table 1**.

Genome-wide association analyses

The GWAS of ECG-AI predicted AF risk did not demonstrate any inflation ($\lambda_{gc}=1.04$). Four genome-wide significant ($P < 5 \times 10^{-8}$) loci were identified (**Figure 2** and **Supplementary Table 1**). Two of the top SNVs were in close proximity to genes previously reported to be associated with AF, including *TTN* and *SCN10A*.^{21,30} The nearest genes at the other two loci were *VGLL2* and *EXT1*. A conditional analysis detected an additional independent association signal at the

SCN5A locus (**Supplementary Table 1**), which has also been reported to be associated with AF in previous studies.^{21,31}

In the GWAS of CHARGE-AF predicted risk, minimal genomic inflation was observed ($\lambda_{gc}=1.10$) which was likely due to polygenicity rather than population stratification, as implicated by the LD score regression intercept (1.0107). Nineteen loci were identified in the GWAS of CHARGE-AF predicted risk (**Figure 2** and **Supplementary Table 2**), none of which have previously been reported in association with AF. Traits associated with lead SNVs at these loci mainly consist of body size measurements and phenotypes related to the clinical factors included in the CHARGE-AF score calculation (**Supplementary Table 2** and **Supplementary Table 3**). No secondary association signals were detected in a conditional analysis.

The LocusZoom plots for risk loci identified in the two GWAS were presented in **Supplementary Figure 1** and **Supplementary Figure 2**, respectively. We also compared the summary statistics of the independent significant lead SNVs in the ECG-AI risk GWAS to that in the CHARGE-AF risk GWAS (**Supplementary Table 4**). No variants exceed the genome-wide significance threshold in the GWAS of 5-year incident AF with the same covariates (**Figure 2**).

Heritabilities and genetic correlations

Using individual level genomic and phenotypic data, the estimated heritability (h^2) was 13.0% (s.e. 1.4%) for ECG-AI risk and 36.5% (s.e. 1.4%) for CHARGE-AF risk. Genetic correlations with AF were estimated to be 35.3% (s.e. 13.7%) for ECG-AI risk and 18.9% (s.e. 8.6%) for CHARGE-AF risk. We further estimated the genetic correlation between the predicted AF risks from ECG-AI and CHARGE-AF, and found a significant correlation of 39.3% (s.e. 4.5%). As a

comparator, we also calculated the heritabilities and genetic correlations using GWAS summary statistics with LD score regression,²⁰ and the estimates were similar in magnitude to those calculated from individual-level data. Detailed results are provided in **Figure 3** and **Supplementary Table 5**.

Polygenic risk scores and incident AF

We calculated two polygenic risk scores (PRS) for the 424,411 eligible participants using predicted AF risk GWAS results. Each 1-SD increase in the PRS of R-INT ECG-AI risk (PRS_{ECG-AI}) and the PRS of R-INT CHARGE-AF risk (PRS_{CHARGE-AF}) were significantly associated with 5-year incident AF (PRS_{ECG-AI} hazard ratio [HR] 1.07, 95% CI 1.04 - 1.09, $P = 3.0 \times 10^{-8}$; and PRS_{CHARGE-AF} HR 1.12, 95% CI 1.09 - 1.14, $P = 3.4 \times 10^{-19}$). When included in the same model, both remained significantly associated with 5-year incident AF (PRS_{ECG-AI} HR 1.06, 95% CI 1.04 - 1.09, $P = 1.1 \times 10^{-6}$; and PRS_{CHARGE-AF} HR 1.11, 95% CI 1.08 - 1.14, $P = 1.1 \times 10^{-17}$). We did not observe a significant interaction between the two PRSs ($P = 0.97$). We also plotted the cumulative risk of AF stratified by high (10%), middle (80%), and low (10%) groupings of the PRS distributions and observed separation between groups (**Supplementary Figure 3**). Due to a difference in the ancestral composition of the GWAS discovery set (White British: 96.6%) and PRS testing set (White British: 87.2%), we repeated the above analysis in the subset of the PRS testing set comprising White British participants only. We observed similar results (**Supplementary Table 6** and **Supplementary Figure 4**).

Contributing components of ECG-AI predicted AF risk

Informed by the genetic signals from our common variant analysis, we tested the causal effects of P wave duration and body height on ECG-AI predicted risk using a two sample Mendelian Randomization (MR) approach. We extracted GWAS summary statistics from studies that did not include UK Biobank samples for P wave duration and body height (see **Methods**). The P wave was selected because (1) it has been linked to *SCN5A/SCN10A* in previous genetic studies,²⁶ and (2) it had the greatest impact on ECG-AI predicted risk indicated by saliency maps and a median waveform analysis in our previous study.⁷ Body height was selected because (1) it has been linked to *VGLL2* and *EXT1* loci,^{32,33} and (2) is an established risk factor for AF.^{34,35}

A significant and plausible causal effect of P wave duration on ECG-AI risk was supported by four out of the five methods we used, with MR effect sizes ranging from 0.017 to 0.023. The causal effect of height on ECG-AI risk was supported by all five methods, with MR effect sizes ranging from 0.085 to 0.123. Results are presented in **Figure 4**.

Discussion

We assessed the genetic basis of AF risk estimates generated from a validated deep-learning model using 12-lead ECGs. Despite the fact that individuals did not have AF at the time of ECG acquisition, we identified variants at three established AF susceptibility loci – *TTN*, *SCN5A*, and *SCN10A* – and at two novel loci that implicate body size measurements – *VGLL2* and *EXT1*. In contrast, our GWAS of CHARGE-AF derived AF risk did not identify any signals previously reported in a GWAS for AF, but identified loci linked to component clinical risk factors for AF that are included in the risk model. Our MR analyses provide supporting evidence that P wave

duration and body height are causal risk factors for AF that mediate the ECG-AI risk estimates.

Broadly, our findings imply that estimates of disease risk from deep learning models that use raw physiologic data, and risk scores more generally, are influenced by genetic susceptibility. In turn, such deep learning models may have the potential to identify individuals at risk for disease via specific genetic pathways.

Deep learning models of 12-lead ECGs for the identification of individuals with a high likelihood of AF have been reported^{5–7} but the interpretability and representations that underlie the risk estimates generated by the models have not been explained. We previously observed that ECG-AI estimates for AF risk are largely influenced by the P wave, a period corresponding to atrial depolarization and repolarization.⁷ Moreover, we have previously reported that both ECG-AI and clinical risk for AF are complementary.⁷ Here, we extend these observations by identifying genetic signals that have been associated with P wave duration,³⁶ and documenting the distinct genetic profiles underlying risk estimates generated by ECG-AI and a clinical risk factor model.

Our findings have two major implications. First, risk estimates from the ECG-AI model are influenced by genetic mechanisms that are more specific to AF than are those from a clinical risk factor model. Specifically, loss-of-function variants in *TTN* have been associated with a substantially increased risk for AF^{37,38} and common variants at this locus have been associated with AF.²¹ *TTN* encodes titin, an integral protein involved in sarcomere development, structural integrity, and contractility.³⁹ The ECG-AI GWAS also identified variation at the *SCN5A* and *SCN10A* loci. Both *SCN5A* and *SCN10A* encode the alpha subunits of voltage gated sodium channels and *SCN5A* is essential for myocyte depolarization. Genetic variants at these loci have been described in association with AF and ECG traits in prior GWAS^{21,26,40,41} and in rare familial

forms of AF.^{31,42–44} *VGLL2* encodes for the vestigial like family member 2a protein, is critical for skeletal muscle development and contains a interacting domain for TEAD1, a member of the Hippo signaling pathway.⁴⁵ *EXT1* encodes a protein involved in the production of heparan sulfate³⁶, and has been implicated in the development of the outflow tract.⁴⁷ We note that the *EXT1* and *VGLL2* loci are relative gene-dense loci, and we have focused here on the nearest genes as a means of prioritization.

In contrast, the genetic signals for CHARGE-AF risk were not specific to AF but reflected the diverse genetic mechanisms underlying the component risk factors in the model, including body height, body weight, blood pressure, and smoking status. Notably, the ECG-AI model loci linked to body size – *VGLL2* and *EXT1* – differ from the CHARGE-AF model loci linked to body size, implying that the manifestation of body size on the ECG may reflect different dimensions of body size from those measured conventionally using height and weight. The relative specificity of the ECG-AI model for genetic mechanisms underlying AF is further supported by our observation that the genetic correlation with AF from a prior GWAS was greater with ECG-AI predicted risk than with CHARGE-AF predicted risk.

Overall, our finding that predicted risk of AF from an ECG-based deep learning model is influenced by inherited susceptibility raises the possibility that an individual's genetic predisposition to AF is inferrable using their raw ECG data alone, even prior to disease onset. Indeed, the GWAS of incident AF did not identify any significant genetic susceptibility loci, underscoring the power of performing a GWAS of models trained to predict risk of a disease, rather than of the disease itself. Future analysis is warranted to assess whether ECG-AI, and other risk models more broadly, can be used to specifically predict which individuals are predisposed to diseases via particular biological mechanisms. Such insights could theoretically

have important implications for the personalization of prevention and therapeutics, identification of individuals with genetic disorders, and the use of deep learning or other risk models as digital biomarkers in addition to general risk prediction models.

Second, the genetic analysis of risk estimates generated by AI models may facilitate the interpretation of output, and underlying representations, of the models which may otherwise be difficult to interpret. Our GWAS, and MR analyses, highlight the fact that our ECG-AI model is influenced by heritable factors related to the P-wave and body size. The fact that ECG-AI can infer anthropometric traits⁴⁸ suggests that deep learning model estimates from raw physiologic data can be influenced by numerous variables which manifest on the data modality under study. Height has been causally linked to AF risk previously.^{34,35} We propose the transfer of clinically-derived models to datasets with genomic information and subsequent genetic association testing, as a means to explore the factors which influence risk estimates from deep learning models.

We further note that whereas our approach focused on a single model task – the predicted risk of AF – examining the genetic architecture of the model latent space itself may reveal the genetic basis for the ECG representations learned by the model. As the number of samples with both ECG and genomic data increase, we anticipate greater statistical power to identify genetic signals underlying predicted disease risk, including rare large-effect variants. Moreover, we submit that the increasing availability of large-scale biobank data with raw data acquisition amenable to deep learning will enable examination of the genetic basis of other disease predictions. We anticipate that improved understanding of the relations between biological pathways and deep learning models may facilitate model application in clinical practice by enhancing clinician confidence in model outputs, and facilitating the inference of biological pathways that lead to disease risk in specific individuals.

Our study should be interpreted in the context of the design. First, our phenotyping algorithm used hospitalization and death records to ascertain disease status, which may lead to disease misclassification. However, before applying the ECG-AI model in UKBB, we omitted individuals that were not identified as prevalent AF by the phenotyping algorithm but were indicated as AF by their diagnostic statement accompanying ECG data, to reduce the impact of misclassification. Second, the MGH dataset we used to train the ECG-AI model and the UKBB dataset we used to perform genetic analyses were predominantly White British, which may limit the generalizability of our findings to populations of other ancestries. It is unclear if the results of the GWAS of ECG-AI reflect the underlying composition of the sample in which it was trained – future analysis of other ECG-AI models for AF prediction are warranted. Third, widespread replication of different model based GWAS will be necessary as additional and larger biorepository datasets emerge. Fourth, risk discrimination of the ECG-AI model in UKBB is moderate. A model with greater discrimination in the discovery sample may increase yield of genetic associations. Lastly, given roughly three years of follow-up after ECG in the UK Biobank, the power of our incident AF analysis may be limited.

In conclusion, we have shown that ECG-AI predicted AF risk reflects inherited predisposition to AF with a genetic background that is more specific for AF risk loci when compared to that for a clinical model. A polygenic risk score constructed using common variants associated with ECG-AI risk was significantly associated with incident AF in a prospective cohort. The interpretability of the ECG-AI model was improved by genetic analyses indicating that P wave duration and body height are likely to be two contributing factors forming the basis of AF risk predictions.

Funding Sources

Dr. Lubitz is supported by NIH grant 1R01HL139731 and American Heart Association 18SFRN34250007. Dr. Anderson is supported by NIH grants R01NS103924 and U01NS069763 and American Heart Association grants 18SFRN34250007 and 21SFRN812095. Dr. Weng is supported by National Institutes of Health (NIH) grant 1R01HL139731. Dr. Choi is supported by the NHLBI BioData Catalyst Fellows program. Dr. Ho is supported by the NIH (R01HL134893, R01HL140224, K24HL153669). Dr. Ellinor is supported by the NIH (1R01HL092577, K24HL105780), AHA (18SFRN34110082) and by MAESTRIA (965286). Dr. Lau is supported by the American Heart Association (853922).

Disclosures

Dr. Lubitz receives sponsored research support from Bristol Myers Squibb / Pfizer, Bayer AG, Boehringer Ingelheim, Fitbit, and IBM, and has consulted for Bristol Myers Squibb / Pfizer, Blackstone Life Sciences, and Invitae. Dr. Anderson receives sponsored research support from Bayer AG and Massachusetts General Hospital and has consulted for ApoPharma. Dr. Weng receives sponsored research support from IBM to the Broad Institute. Dr. Ho has received sponsored research support from Bayer AG and research supplies from EcoNugenics, Inc. Dr. Ellinor has received sponsored research support from Bayer AG and IBM Health, and he has consulted for Bayer AG, Novartis and MyoKardia. Dr. Batra, Dr. Reeder and Dr. Friedman have received sponsored research support from Bayer AG and IBM Health, and Dr. Batra has consulted for Novartis and Prometheus Biosciences.

References

1. Weng LC, Choi SH, Klarin D, et al. Heritability of Atrial Fibrillation. *Circ Cardiovasc Genet*. 2017;10(6):e001838. doi:10.1161/CIRCGENETICS.117.001838
2. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide Epidemiology of Atrial Fibrillation. *Circulation*. 2014;129(8):837-847. doi:10.1161/CIRCULATIONAHA.113.005119
3. Virani SS, Alonso A, Aparicio HJ, et al. Heart Disease and Stroke Statistics—2021 Update. *Circulation*. 2021;143(8):e254-e743. doi:10.1161/CIR.0000000000000950
4. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*. 2019;394(10201):861-867. doi:10.1016/S0140-6736(19)31721-0
5. Artificial Intelligence—Electrocardiography to Predict Incident Atrial Fibrillation. doi:10.1161/CIRCEP.120.009355
6. Raghunath S, Pfeifer JM, Ulloa-Cerna AE, et al. Deep Neural Networks Can Predict New-Onset Atrial Fibrillation From the 12-Lead ECG and Help Identify Those at Risk of Atrial Fibrillation—Related Stroke. *Circulation*. 2021;143(13):1287-1298. doi:10.1161/CIRCULATIONAHA.120.047829
7. Khurshid S, Friedman S, Reeder C, et al. Electrocardiogram-based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation. *Circulation*. 0(0). doi:10.1161/CIRCULATIONAHA.121.057480
8. Alonso A, Krijthe BP, Aspelund T, et al. Simple Risk Model Predicts Incidence of Atrial Fibrillation in a Racially and Geographically Diverse Population: the CHARGE-AF Consortium. *J Am Heart Assoc Cardiovasc Cerebrovasc Dis*. 2013;2(2):e000102. doi:10.1161/JAHA.112.000102
9. Sudlow C, Gallacher J, Allen N, 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*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
10. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
11. Khurshid S, Reeder C, Harrington LX, et al. Cohort Design and Natural Language Processing to Reduce Bias in Electronic Health Records Research: The Community Care Cohort Project. *medRxiv*. Published online January 1, 2021:2021.05.26.21257872. doi:10.1101/2021.05.26.21257872
12. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
13. Alonso A, Roetker NS, Soliman EZ, Chen LY, Greenland P, Heckbert SR. Prediction of Atrial Fibrillation in a Racially Diverse Cohort: The Multi-Ethnic Study of Atherosclerosis (MESA). *J Am Heart Assoc*. 2016;5(2):e003077. doi:10.1161/JAHA.115.003077
14. Pfister R, Brägelmann J, Michels G, Wareham NJ, Luben R, Khaw KT. Performance of the CHARGE-AF risk model for incident atrial fibrillation in the EPIC Norfolk cohort. *Eur J Prev Cardiol*. 2015;22(7):932-939. doi:10.1177/2047487314544045
15. Khurshid S, Choi SH, Weng LC, et al. Frequency of Cardiac Rhythm Abnormalities in a Half Million Adults. *Circ Arrhythm Electrophysiol*. 2018;11(7):e006273. doi:10.1161/CIRCEP.118.006273
16. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021;53(7):1097-1103. doi:10.1038/s41588-021-00870-7
17. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369-375. doi:10.1038/ng.2213

18. He L, Kulminski AM. Fast Algorithms for Conducting Large-Scale GWAS of Age-at-Onset Traits Using Cox Mixed-Effects Models. *Genetics*. 2020;215(1):41-58. doi:10.1534/genetics.119.302940
19. Loh PR, Bhatia G, Gusev A, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet*. 2015;47(12):1385-1392. doi:10.1038/ng.3431
20. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211
21. Roselli C, Chaffin MD, Weng LC, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet*. 2018;50(9):1225-1233. doi:10.1038/s41588-018-0133-9
22. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. 2012;28(19):2540-2542. doi:10.1093/bioinformatics/bts474
23. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1):1776. doi:10.1038/s41467-019-09718-5
24. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89-98. doi:10.1093/hmg/ddu328
25. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Loos R, ed. *eLife*. 2018;7:e34408. doi:10.7554/eLife.34408
26. Christophersen IE, Magnani JW, Yin X, et al. Fifteen Genetic Loci Associated With the Electrocardiographic P Wave. *Circ Cardiovasc Genet*. 2017;10(4):e001667. doi:10.1161/CIRCGENETICS.116.001667
27. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46(11):1173-1186. doi:10.1038/ng.3097
28. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525. doi:10.1093/ije/dyv080
29. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693-698. doi:10.1038/s41588-018-0099-7
30. Christophersen IE, Rienstra M, Roselli C, et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet*. 2017;49(6):946-952. doi:10.1038/ng.3843
31. D D, Pj K, Bs D, et al. Cardiac sodium channel (SCN5A) variants associated with atrial fibrillation. *Circulation*. 2008;117(15). doi:10.1161/CIRCULATIONAHA.107.757955
32. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010;467(7317):832-838. doi:10.1038/nature09410
33. Kichaev G, Bhatia G, Loh PR, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet*. 2019;104(1):65-75. doi:10.1016/j.ajhg.2018.11.008
34. Rosenberg MA, Patton KK, Sotoodehnia N, et al. The impact of height on the risk of atrial fibrillation: the Cardiovascular Health Study. *Eur Heart J*. 2012;33(21):2709-2717. doi:10.1093/eurheartj/ehs301
35. Levin MG, Judy R, Gill D, et al. Genetics of height and risk of atrial fibrillation: A Mendelian randomization study. *PLOS Med*. 2020;17(10):e1003288. doi:10.1371/journal.pmed.1003288
36. Weng LC, Hall AW, Choi SH, et al. Genetic Determinants of Electrocardiographic P-Wave Duration and Relation to Atrial Fibrillation. *Circ Genomic Precis Med*. 2020;13(5):387-395. doi:10.1161/CIRCGEN.119.002874
37. Choi SH, Weng LC, Roselli C, et al. Association Between Titin Loss-of-Function Variants and Early-Onset Atrial Fibrillation. *JAMA*. 2018;320(22):2354-2364. doi:10.1001/jama.2018.18179
38. Choi SH, Jurgens SJ, Weng LC, et al. Monogenic and Polygenic Contributions to Atrial Fibrillation

- Risk. *Circ Res*. 2020;126(2):200-209. doi:10.1161/CIRCRESAHA.119.315686
39. Hinson JT, Chopra A, Nafissi N, et al. HEART DISEASE. Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. *Science*. 2015;349(6251):982-986. doi:10.1126/science.aaa5458
40. Nielsen JB, Thorolfsson RB, Fritsche LG, et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet*. 2018;50(9):1234-1239. doi:10.1038/s41588-018-0171-3
41. Tereshchenko LG, Sotoodehnia N, Sitlani CM, et al. Genome-Wide Associations of Global Electrical Heterogeneity ECG Phenotype: The ARIC (Atherosclerosis Risk in Communities) Study and CHS (Cardiovascular Health Study). *J Am Heart Assoc*. 2018;7(8):e008160. doi:10.1161/JAHA.117.008160
42. Olson TM, Michels VV, Ballew JD, et al. Sodium channel mutations and susceptibility to heart failure and atrial fibrillation. *JAMA*. 2005;293(4):447-454. doi:10.1001/jama.293.4.447
43. Ellinor PT, Nam EG, Shea MA, Milan DJ, Ruskin JN, MacRae CA. Cardiac sodium channel mutation in atrial fibrillation. *Heart Rhythm*. 2008;5(1):99-105. doi:10.1016/j.hrthm.2007.09.015
44. Makiyama T, Akao M, Shizuta S, et al. A novel SCN5A gain-of-function mutation M1875T associated with familial atrial fibrillation. *J Am Coll Cardiol*. 2008;52(16):1326-1334. doi:10.1016/j.jacc.2008.07.013
45. Maeda T, Chapman DL, Stewart AFR. Mammalian vestigial-like 2, a cofactor of TEF-1 and MEF2 transcription factors that promotes skeletal muscle differentiation. *J Biol Chem*. 2002;277(50):48889-48898. doi:10.1074/jbc.M206858200
46. Lin X, Wei G, Shi Z, et al. Disruption of gastrulation and heparan sulfate biosynthesis in EXT1-deficient mice. *Dev Biol*. 2000;224(2):299-311. doi:10.1006/dbio.2000.9798
47. Zhang R, Cao P, Yang Z, et al. Heparan Sulfate Biosynthesis Enzyme, Ext1, Contributes to Outflow Tract Development of Mouse Heart via Modulation of FGF Signaling. *PloS One*. 2015;10(8):e0136518. doi:10.1371/journal.pone.0136518
48. Li X, Patel KHK, Sun L, Peters NS, Ng FS. Neural networks applied to 12-lead electrocardiograms predict body mass index, visceral adiposity and concurrent cardiometabolic ill-health. *Cardiovasc Digit Health J*. 2021;2(6):S1-S10. doi:10.1016/j.cvdhj.2021.10.003

Table 1. Characteristics of the UK Biobank cohort participants.

Variable	GWAS set (N=39,986)	PRS testing set (N=424,411)
Enrollment age	55.6 (7.6)	57.1 (8.1)
ECG age	64.0 (7.7)	Not applicable
Female	20,809 (52.0%)	232,654 (54.8%)
Race (White British)	38,617 (96.6%)	370,121 (87.2%)
Height (cm)	169.1 (9.2)	168.3 (9.3)
Weight (kg)	76.0 (15.2)	78.0 (16.0)
Systolic blood pressure	138.3 (18.6)	138.1 (18.7)
Diastolic blood pressure	79.1 (10.1)	82.4 (10.2)
Smoking (current)	1,446 (3.6%)	46,433 (11.0%)
Antihypertensive medication use (Yes)	4,275 (10.7%)	88,887 (18.6%)
Diabetes (Yes)	1,546 (3.9%)	11,089 (2.6%)
Heart failure (Yes)	186 (0.5%)	1,701 (0.4%)
Myocardial infarction (Yes)	916 (2.3%)	9,657 (2.3%)

Clinical variables were ascertained at baselines. For the genome-wide association studies

(GWAS) discovery set, baseline refers to the time of electrocardiogram (ECG) visit. For the

polygenic risk scores (PRS) testing set, baseline refers to the time of study enrollment.

Descriptive statistics of enrollment age, female, race, and history of diabetes, heart failure and

myocardial infarction for the PRS testing set were calculated using the complete sample (N=424,411). Height, weight, systolic blood pressure and diastolic blood pressure were summarized in a subset with the data available (N=423,044). The sample size for participants who self-reported their smoking status and medication use were 423,919 and 477,644, respectively.

Figure 1. Study overview.

We applied our validated ECG-AI model to samples with ECG data in the UK Biobank. After excluding participants who withdrew consent, had missing CHARGE-AF components, were diagnosed with atrial fibrillation (AF) before ECG examination, did not have follow-up information or failed sample QC procedures, 39,987 remained in the discovery set, in which we performed GWAS and post-GWAS analyses. Among the 446,963 remaining genotyping samples, 424,411 did not withdraw consent, were not < 3rd-degree relatives with individuals in the GWAS set, were not diagnosed with AF before enrollment, had follow-up information, and did not fail sample QC procedures. We calculated polygenic risk scores (PRS) using the GWAS results and associated the PRS with incident AF in this subset.

Figure 2. Manhattan plots of genome-wide association studies of ECG-AI and CHARGE-AF predicted risk of AF, and observed 5-year incident AF in the UK Biobank.

Chromosomal variant positions are plotted on the x-axis. The $-\log_{10}(P \text{ values})$ are plotted on the y-axis. The genome-wide significance threshold (5×10^{-8}) is indicated by the horizontal dotted line. Variants are colored red near loci that have been reported in a prior atrial fibrillation (AF) GWAS,²¹ and are colored dark blue near loci that have not been reported previously in association with AF. Panels display associations with (a) ECG-AI predicted 5-year risk of AF, (b) CHARGE-AF predicted 5-year risk of AF, and (c) observed incident AF at 5-years in the UK Biobank. 5-year AF risk estimates were rank-based inverse normal transformed prior to analysis (see text).

Figure 3. Heritability and genetic correlation estimates for model predicted 5-year atrial fibrillation (AF) risk.

Heritability (h^2) derived from the ECG-AI and CHARGE-AF GWAS are displayed on the left panel. Genetic correlation (r_g) comparing both the ECG-AI and CHARGE-AF GWAS with a prior independent large-scale GWAS of AF,²¹ is displayed on the right panel. ‘Individual-level’ refers to estimates generated from individual-level genetic and phenotypic data. ‘Summary statistics’ refers to estimates generated from GWAS results. Summary statistics for ECG-AI risk and CHARGE-AF risk were extracted from the GWAS in the present study.

Figure 4. Mendelian Randomization analysis between P wave duration, body height, and atrial fibrillation (AF).

The figure displays Mendelian randomization results assessing ECG P wave duration (top panel) and body height (bottom panel) for relations with genetically predicted ECG-AI risk. Mendelian randomization effect sizes are graphed on the x-axis. Dots represent point estimates and bars represent 95% confidence intervals. Dashed gray lines represent zero effect sizes. The five Mendelian randomization methods used in this analysis are shown in the y-axis.