**Electrical Phenotypes of Genetic Risk for Atrial Fibrillation**

**Proposal for Validation Cohort**

AF is a multifactorial disease that is both highly prevalent and burdensome, with a lifetime incidence of 1 out of 3-4 individuals. Although We have a limited understanding of the clinical factors that lead to atrial fibrillation (**AF**) pathogenesis, there is growing evidence of the genetic underpinnings of AF (Weng et al. 2020; Choi et al. 2020). It is clear that AF maintenance and pro-generation are a property of the atrial substrate, which can likely be further phenotyped based on how the atria are remodeled. The relationship of the atrial substrate to underlying genetic factors has somewhat been seen in studies evaluating electrocardiogram (ECG) associations with genetic variants (Verweij et al. 2020; Weng et al. 2020; Holmqvist et al. 2015; Husser et al. 2009; Choi et al. 2021). In this proposal, I plan to correlate the genetic manifestations of AF with the electrical properties of through electrocardiographic (**ECG**) evaluation of atrial and ventricular depolarization and repolarization using a deep neural network (**DNN**).

Relevant work previously is primarily by Verweij et al. (2020) and van de Leur et al. (2021), who have evaluated ECG relationships with genetic variants identified in whole exome sequencing data (**WES**) in the UK Biobank. Verweij et al. (2020) noted that specific temporal features of the ECG could be associated with genetic abnormalities ([Figure 1](#fig-ecg-variants)). van de Leur et al. (2021) expanded on this to understand the electrical phenotype of a specific disease, in this case, phosphalamban cardiomyopathy, and demonstrated that a neural network could help identify morphological features of the ECG that were associated with the disease, best seen in [Figure 2](#fig-ecg-dnn). Additional work has been done that evaluates the polygenic risk score (PGS) for AF, which appears to be additive to that of monogenic risk for AF (Khera et al. 2018; Marston et al. 2023; Choi et al. 2020).

Our proposal varies from the previous work as the goal is instead to evaluate the potential mechanisms of AF, instead of novel diagnosis of specific genetic variants. By comparing ECG of individuals with known AF into specific endotypes based on genetic variants associated with their AF, we can understand how variants may lead to AF through electrophyioslogical manifestations. As such, I have characterized genes putatively associated with cardiovascular disease into 4 categories:

1. Structural genes (e.g. TTN, PITX2, LMNA, NUP155, GJA1/5)
2. Ion channel genes (e.g. KCNE1-5, KCNQ1, SCN5A), subdivided by current type
3. Metabolic genes
4. Inflammatory genes

These categories, from a causal perspective, share mechanistic similarities in how they may lead to AF. To identify if electrophysiological manifestations exist, in a sample with documented paroxysmal AF, individuals that carry a loss-of-function (**LOF**) in one of the above gene categories, or they may be *wild-type* (no known deleterious AF-associated genes). We will also look at the ECG-contribution to the PGS of AF, which has been shown to have independent associations with future AF risk (Wang et al. 2023). Instead of predicting AF, we will attempt to categorize the contributions to the PGS of AF based on 12-lead ECG.

We have completed a feasibility study using the *TTN* gene as a proof-of-concept, seen in [Figure 3](#fig-shah-ttn), using a convolutional neural network with bidirectional memory. The model was trained on a cohort of n=200 individuals with known paroxysmal AF, with n=100 being reserved as a testing data set. Of those, using the strictest variant effect prediction definition for LOF, n=4 individuals were considered as cases (roughly 2% of the cohort). Each individual contributed 1 to 10 ECGs, with each ECG being segmented into individual sinus beats, and each sinus beat used as a training (or test) data array. We had a sensitivity and specificity of 98% and 95%, with a balanced accuracy of 96% and detection prevalence of 95%.

We aim to expand this project to better understand the electrocardiographyic phenotype of AF, to better understand mechanisms behind AF pathogenesis, and identify individuals at different degrees of risk for AF. My aims are as follow:

1. **Monogenic Aim**: Validate a DNN model to identify deleterious gene variants associated with AF in individuals with known AF, with cases defined as those with: a) sarcomeric or structural proteins, b) ion channel-related proteins, c) metabolic proteins, and d) inflammatory proteins that have a known association with cardiovascular disease.
2. **Polygenic Aim**: Train a DNN model to assess the polygenic risk (PGS) of AF in individuals with known AF and dense SNP genotyping data, based on recent, validated AF-GWAS PGS (Khera et al. 2018; Marston et al. 2023). The model would be trained to predict the percent PGS of AF, however the input is a 12-lead ECG.

We utilized participants from the UIC Multi-Ethnic AF Registry as the primary study population, who have a combination of WES and dense SNP genotyping data. This allowed me to generate a small DNN model for both monogenic and polygenic patterns. However, the sample size was small, reducing generalizability. We propose to utilize a large, external cohort, such the UK Biobank (n=68,091 with complete data), as well as additional populations with WES data available. Our work will be the next logical step in elucidating genetic mechanisms of AF, and provides a novel approach to understand biological endotypes of the disease.

|  |
| --- |
| Figure 1: Verweij et al. (2020) shows that different time-based features of a sinus beat are associated with specific genetic variants in known CVD-related genes. Their approach utilizes cross-sectional information about the median sinus beat in those with and without deleterious genetic variants. |

|  |
| --- |
| Figure 2: van de Leur et al. (2021) showed that a neural network can be trained to help understand and identify the potential electrical manifestations of a gene mutation in the phospholamban protein. The convolutional neural network allows for explainability of the model through class activation mapping. |

|  |
| --- |
| Figure 3: Our recent work is demonstrated here, where we identify ECGs that are associated with *TTN* LOF in individuals with AF. The *red* line indicates the median signal seen in cases, while the *blue* represents those that serve as controls (those with no known genetic variants in sarcomeric proteins). The yellow gradient activation notes the regions of the ECG that were most informative in differentiating *TTN* cases from controls. Notably, in this model, the repolarization phase is most strongly associated with *TTN* cases in AF. |

# References

1. Choi, Seung Hoan, Sean J. Jurgens, Christopher M. Haggerty, Amelia W. Hall, Jennifer L. Halford, Valerie N. Morrill, Lu-Chen Weng, et al. 2021. “Rare Coding Variants Associated With Electrocardiographic Intervals Identify Monogenic Arrhythmia Susceptibility Genes.” *Circulation. Genomic and Precision Medicine* 14 (4): e003300. <https://doi.org/10.1161/CIRCGEN.120.003300>.
2. Choi, Seung Hoan, Sean J. Jurgens, Lu-Chen Weng, James P. Pirruccello, Carolina Roselli, Mark Chaffin, Christina J.-Y. Lee, et al. 2020. “Monogenic and Polygenic Contributions to Atrial Fibrillation Risk.” *Circulation Research* 126 (2): 200–209. <https://doi.org/10.1161/CIRCRESAHA.119.315686>.
3. Holmqvist, Fredrik, Sunghee Kim, Benjamin A. Steinberg, James A. Reiffel, Kenneth W. Mahaffey, Bernard J. Gersh, Gregg C. Fonarow, et al. 2015. “Heart Rate Is Associated with Progression of Atrial Fibrillation, Independent of Rhythm.” *Heart (British Cardiac Society)* 101 (11): 894–99. <https://doi.org/10.1136/heartjnl-2014-307043>.
4. Husser, Daniela, Martin Stridh, Leif Sörnmo, Dan M. Roden, Dawood Darbar, and Andreas Bollmann. 2009. “A Genotype-Dependent Intermediate ECG Phenotype in Patients With Persistent Lone Atrial Fibrillation.” *Circulation: Arrhythmia and Electrophysiology* 2 (1): 24–28. <https://doi.org/10.1161/CIRCEP.108.799098>.
5. Khera, Amit V., Mark Chaffin, Krishna G. Aragam, Mary E. Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, et al. 2018. “Genome-Wide Polygenic Scores for Common Diseases Identify Individuals with Risk Equivalent to Monogenic Mutations.” *Nature Genetics* 50 (9, 9): 1219–24. <https://doi.org/10.1038/s41588-018-0183-z>.
6. Leur, Rutger R. van de, Karim Taha, Max N. Bos, Jeroen F. van der Heijden, Deepak Gupta, Maarten J. Cramer, Rutger J. Hassink, et al. 2021. “Discovering and Visualizing Disease-Specific Electrocardiogram Features Using Deep Learning.” *Circulation: Arrhythmia and Electrophysiology* 14 (2): e009056. <https://doi.org/10.1161/CIRCEP.120.009056>.
7. Marston, Nicholas A, Amanda C Garfinkel, Frederick K Kamanu, Giorgio M Melloni, Carolina Roselli, Petr Jarolim, David D Berg, et al. 2023. “A Polygenic Risk Score Predicts Atrial Fibrillation in Cardiovascular Disease.” *European Heart Journal* 44 (3): 221–31. <https://doi.org/10.1093/eurheartj/ehac460>.
8. Verweij, Niek, Jan-Walter Benjamins, Michael P. Morley, Yordi J. Van De Vegte, Alexander Teumer, Teresa Trenkwalder, Wibke Reinhard, Thomas P. Cappola, and Pim Van Der Harst. 2020. “The Genetic Makeup of the Electrocardiogram.” *Cell Systems* 11 (3): 229–238.e5. <https://doi.org/10.1016/j.cels.2020.08.005>.
9. Wang, Maggie, Chang Liu, Anish S. Shah, Yi-An Ko, Yan Sun, Rachel J. Lampert, Mariana Garcia, et al. 2023. “Heart Rate Variability During Mental Stress Predicts Heart Failure Outcomes In Patients With Stable Coronary Artery Disease.” In *Journal of the American College of Cardiology*, 81:469. ACC.23. New Orleans, LA. <https://doi.org/10.1016/S0735-1097(23)00913-0>.
10. Weng, Lu-Chen, Amelia Weber Hall, Seung Hoan Choi, Sean J. Jurgens, Jeffrey Haessler, Nathan A. Bihlmeyer, Niels Grarup, et al. 2020. “Genetic Determinants of Electrocardiographic P-Wave Duration and Relation to Atrial Fibrillation.” *Circulation. Genomic and Precision Medicine* 13 (5): 387–95. <https://doi.org/10.1161/CIRCGEN.119.002874>.