Examining Rare and Low-Frequency Genetic Variants Previously Associated with Lone or Familial Forms of Atrial Fibrillation in an Electronic Medical Record System: A Cautionary Note

Running title: Weeke et al.; Rare and Low-Frequency Genetic Variants and AF

Peter Weeke, MD¹,²; Joshua C. Denny, MD¹,³; Lisa Basterache, MS³; Christian Shaffer, BS¹;
Erica Bowton, PhD⁴; Christie Ingram, BS¹; Dawood Darbar, MD¹; Dan M. Roden, MD¹

¹Department of Internal Medicine, ²Department of Cardiology, ³Department of Biomedical Informatics, ⁴Institute for Clinical & Translational Research, Vanderbilt University Medical Center, Nashville, TN

Correspondence:
Dan M. Roden, MD
Professor of Medicine and Pharmacology
Assistant Vice-Chancellor for Personalized Medicine
Vanderbilt University School of Medicine
1285 Medical Research Building IV
Nashville, TN 37232
Tel: 615-322-0067
Fax: 615-343-4522
E-mail: dan.roden@vanderbilt.edu

Abstract:

**Background** - Studies in individuals or small kindreds have implicated rare variants in 25 different genes in lone and familial atrial fibrillation (AF) using linkage and segregation analysis, functional characterization, and rarity in public databases. Here we used a cohort of 20,204 patients of European or African ancestry with electronic medical records (EMRs) and exome chip data to compare the frequency of AF among carriers and non-carriers of these rare variants.

**Methods and Results** - The exome chip included 19/115 rare variants, in 9 genes, previously associated with lone or familial AF. Using validated algorithms querying a combination of clinical notes, structured billing codes, ECG reports, and procedure codes, we identified 1,056 AF cases (>18 years) and 19,148 non-AF controls (>50 years) with available genotype data on the Illumina HumanExome BeadChip v.1.0 in the Vanderbilt electronic medical record-linked DNA repository, BioVU. Known correlations between AF and common variants at 4q25 were replicated. None of the 19 variants previously associated with AF were overrepresented among AF cases (P >0.1 for all), and the frequency of variant carriers among non-AF controls was >0.1% for 14/19. Repeat analyses using non-AF controls aged >60 (n=14,904) >70 (n=9,670), and >80 (n=4,729) years old did not influence these findings.

**Conclusions** - Rare variants previously implicated in lone or familial forms AF present on the exome chip are detected at low frequencies in a general population but are not associated with AF. These findings emphasize the need for caution when ascribing variants as pathogenic or causative.

**Key words:** Genome Wide Association Study, single nucleotide polymorphism, genetics, association studies, atrial fibrillation, gene mutation, family study, penetrance, exome, genome, GWAS
Atrial fibrillation (AF) is the most prevalent arrhythmia encountered in clinical practice and is significantly associated with increased risks of stroke, heart failure, and death.\(^1\) While there are many clinical correlates of AF (e.g. hypertension and diabetes), genetic variability is being increasingly recognized as an important determinant for developing AF.\(^2\)\(^-\)\(^11\)\(^12\)\(^-\)\(^16\) To date, rare genetic variants in the exonic regions of 25 genes encoding cardiac ion channels, gap junction proteins, and signaling molecules have been linked with lone or familial forms of AF.\(^2\)\(^-\)\(^11\) In addition, genome-wide association studies (GWAS) have associated multiple common variants with AF in a general population.\(^13\)\(^-\)\(^16\) Overall, the effects of common variants on AF susceptibility are small, whereas the effects of rare variants on AF susceptibility have not been determined, although they are presumed to be large.\(^11\)\(^14\) Establishing the prevalence of such variants in large population studies with available phenotypic information can be useful i) to estimate the effect sizes associated with genetic variants; ii) to avoid falsely predicting individuals being at risk for disease; and iii) to evaluate penetrance. In the present study, we studied a large cohort of patients with available genotyping data from the Illumina HumanExome BeadChip v.1.0 genotyping array ("the exome chip") that interrogates \(~250,000\) rare and common exonic variants and linked to de-identified electronic medical records (EMR).\(^18\) We tested the hypothesis that AF is more prevalent in carriers of rare variants previously designated as pathogenic for AF.

**Methods**

**BioVU and Exome Chip Study cohorts**

BioVU links DNA samples extracted from discarded blood samples to de-identified EMRs, as previously described.\(^18\) In 2012-2013, \(~40,000\) subjects in BioVU were genotyped on the exome chip.
Subjects for exome chip genotyping were drawn from the following BioVU cohorts:

1. An elderly cohort: subjects > 75 years of age with ≥3 years of follow-up.
3. A longitudinal cohort:
   a. Healthy longitudinal set = subjects with 1 or 2 notes/year for at least 5 years.
   b. Sick longitudinal set = subjects with 3, 4, 5, or 6 notes/year for at least 5 years.
4. The Vanderbilt Electronic Systems for Pharmacogenomic Assessment (VESPA) cohort: subjects with drug response phenotypes.19
5. A cancer cohort: subjects with cancer according to the Vanderbilt Tumor Registry.
6. A Pediatric cohort: subjects <18 years old with a minimum threshold of clinical record data (≥1 pediatric visit).

For the present study, we included all subjects of European American (EA) and African American (AA) ancestry from groups 1-5 above.

AF phenotype

To identify AF cases, we used a validated algorithm20 that uses a combination of queries of structured billing codes (i.e. International Classification of Diseases, version 9-CM [ICD-9], procedure codes from Current Procedural Terminology [CPT]), and unstructured “natural language” from clinical notes and electrocardiograms. In brief, AF cases were defined as individuals who were >18 years old, had an ICD-9 diagnosis for AF or flutter (ICD-9: 427.3, 427.31, and 427.32) or a cardiologist diagnosis of AF as identified by a natural language processing tool from the unstructured free text of the “ECG impression”. In all instances, patients with a history of a heart transplant were excluded (CPT: 33935, 3394, and 580; ICD-9: V42.1,
For the present study, non-AF controls were defined as individuals who were >50 years old and who were not identified as a case by the case algorithm. For sensitivity analyses, we identified three additional control cohorts of individuals who were not identified as cases and who were >60 years, >70 years, and >80 years old. In addition, we also defined a “definite non-AF” control cohort as individuals who had ≥1 ECG that did not show AF, had no ICD-9 codes for AF or atrial flutter (ICD-9: 427.3, 427.31, and 427.32), had no free text mentions AF in their clinical notes, and no free-text references (including synonyms) to direct-current cardioversion, atrial tachycardia or multifocal atrial tachycardia, atrioventricular nodal ablation, as previously described.\(^{20}\) AF case and control numbers are derived from algorithm deployment into BioVU on January 15, 2014. We stratified all analyses by AA and EA ancestry, as described further below.

**Genotyping and Quality Control**

We excluded samples with genotype-phenotype sex discrepancies, samples with genotyping efficiency <98%, related samples as determined by identity-by-descent, and unintended duplicates. We removed SNPs that were monomorphic, had low call-rates (<98%), or were common (minor allele frequency (MAF) >5%) and were out of Hardy-Weinberg equilibrium (HWE) (HWE P<0.0001).

**Ancestry**

In order to reduce the influence of population stratification on our results, we identified and analyzed separately individuals of EA and AA ancestry by calculating principal components using EIGENSOFT\(^{21}\); the first and second principal components were retained for further analysis.
SNP identification

We performed a literature review to identify studies implicating rare genetic variants with lone or familial forms of AF using the following search terms in PubMed: ((atrial fibrillation (mesh)) OR (atrial fibrillation)) AND ((genetics (mesh)) OR (genetic*)) AND ((mutation (mesh)) OR (mutation*) OR (polymorphism, single nucleotide (mesh)) OR (polymorphism, single nucleotide) OR (monogenic*) OR (gwas)). In total, we identified 565 articles meeting these search criteria as of January 12, 2014. These articles implicated 115 rare genetic variants that had evidence of i) co-segregation with AF; ii) supporting electrophysiological studies or; iii) were associated with AF based on large studies of lone AF probands and/or SNP rarity. Supplemental Table 1 lists additional rare variants not included in a recent review by Olesen et al. We cross-referenced the 115 rare SNPs with the exome chip. Overlapping SNPs were retained for further analysis.

As a positive control for our experiment, we also examined the relationship between AF/non-AF phenotype status and previously-described common AF variants at the 4q25 locus among EAs (rs6843082, rs2200733, and rs17042171).13, 16, 23

Statistical Analyses

Dichotomous and continuous variables were compared with Fisher’s exact test and the Kruskal-Wallis test, respectively. Continuous values are presented as medians and interquartile ranges (IQR). We determined SNP associations with AF using exact logistic regression. We performed both unadjusted and adjusted analyses (age, sex, first and second principal components [PC]), which yielded similar estimates unless otherwise noted.

We also performed a series of interaction analyses using logistic regression analysis with 2 main model effect terms (“presence or absence of rare AF variants” and “presence or absence
of common AF variants”) and a multiplicative interaction term (“presence or absence of rare AF variants” X “presence or absence of common AF variants”) to test the hypothesis that the co-occurrence of both common AF variants and rare AF variants modulates the risk of developing AF (i.e. SNP X SNP interaction). We tested the co-occurrence of each of common vs. each rare variant and we also collapsed rare and common variants into an allelic test based on the presence and absence of the variant. We only evaluated SNP X SNP interactions for SNPs that were identified from studies based on EA populations.

To reduce the risk of misclassification due to age-dependent development of AF, we repeated all association analyses using additional non-AF control cohorts, with no AF as described above at age cut-offs of >60, >70, and >80 years. Conservation for single base variants and prediction of functional effects were assessed using PhastCons, Genome Evolution Rate Profiler (GERP), Grantham scores, and PolyPhen2.

All analyses were performed using PLINK/SEQ (http://atgu.mgh.harvard.edu/plinkseq) and SAS, version 9.2 (SAS institute Inc., Cary, NC, USA). A two-sided p-value <0.05 was considered to be statistically significant.

Ethics

Approval for the present study was obtained from the Vanderbilt Institutional Review Board (IRB).

Results

We identified 20,204 individuals (18,424 EAs and 1,780 AAs) with available genotype data and also passed all QC parameters. Table 1 lists the clinical characteristics for AF cases and the non-AF controls ≥50 years old, stratified by EA and AA ancestry. Overall, the AF cases were older (77.8 [IQR: 69.4-83.8] years) than the non-AF controls ≥50 years old (69.5 [IQR: 60.6-79.4]
years, P<0.001). The median age was 74.2 (IQR: 66.6-81.5) years for non-AF controls aged ≥60 years (n=13,715), 79.4 (IQR: 74.8-84.6) years for the non-AF controls aged ≥70 years (n=8,856), and 84.6 (IQR: 81.9-88.5) years for the non-AF controls aged ≥80 years (n=4,384). For the definite non-AF controls (n=4,514), the median age was 78.0 (IQR: 66.6-83.9).

Previously-described associations between common genetic variants at the 4q25 locus and AF were replicated in the present study in both adjusted (rs6843082, odds ratio [OR]=1.28, 95% confidence interval [CI]: 1.15-1.42, p=5.3x10^{-6}; rs2200733, OR=1.33, CI: 1.16-1.50, p=3.0x10^{-5}; rs17042171, OR=1.33, CI: 1.17-1.51, p=1.9x10^{-5}) and unadjusted additive models (data not shown).

Cross referencing the 115 rare SNPs previously implicated with lone or familial forms of AF with the exome chip yielded a total of 19 overlapping SNPs, in 9 genes: NPPA (n=1), KCNE4 (n=1), GREM2 (n=1), SCN5A (n=11), GATA4 (n=1), KCNJ8 (n=1), KCNJD3 (n=1), KCNA5 (n=1), and KCNJ2 (n=1). In the parent studies suggesting a role for these SNPs in lone or familial forms of AF, 14 SNPs were identified among individuals of EA ancestry, 4 SNPs in individuals of AA ancestry, and 1 SNP in Asian subjects.

None of the 19 SNPs were significantly overrepresented among EA/AA AF cases compared to the respective non-AF controls (P>0.1 for all; Table 2). Of the 19 SNPs, 14/19 (10 in EAs and 4 in AAs) were present in ≥0.1% of the EA/AA non-AF controls (number of risk allele carriers range among 17,416 EA non-AF controls, range: 19-179 [0.11-1.03%]; number of risk allele carriers range among 1,732 AA non-AF controls, range: 24-112 [1.39-6.47%]).

Four of the 10 EA SNPs that were common among the non-AF controls (≥0.1%) were very rare or absent in the 1,008 EA AF cases (range:0-1), and an additional 4 EA SNPs were rare among both EA AF cases (range: 0-1 [0-0.1%]) and EA non-AF controls (range: 2-14 [0.01-
0.08%]). All four SNPs originally reported to be of AA ancestry were common among both AA AF cases and AA non-AF controls (Table 2).

Supplemental Table 2 presents the in silico functional predictions of the 19 SNPs and the derived EA and AA MAFs from the exome sequencing project of 6500 individuals. Of the 19 variants, 9 were predicted to disrupt protein function by PolyPhen2 (the PolyPhen2 score ranges from 0 to 1, with the threshold for probably damaging at 0.85).

Repeat analyses using the definite EA and AA non-AF controls (n=4514 and n= 541, respectively), the EA and AA non-AF controls aged >60 years (n=13,715 and n=1,189, respectively), the EA and AA non-AF controls aged >70 years (n=8,856 and n=714, respectively), and the EA and AA non-AF controls aged >80 years (n=4,388 and n=341, respectively), did not influence the results.

SNP x SNP interaction

Table 3 presents the number of rare variant carriers as a function of common AF variant carrier status and AF status. Here there was no significant interaction between co-occurrence of common AF variants (rs6843082, rs2200733, and rs17042171) and rare genetic variants and AF susceptibility (P=0.47). We also found no association between overall rare AF variant carrier status (ignoring common variants) and AF susceptibility (P=0.81).

Discussion

In the present study we found that the rare or low-frequency variants interrogated by the exome chip and previously implicated in lone or familial forms of AF were not associated with AF in a general population; a substantial proportion of the non-AF controls carried rare or low frequency variants previously labeled as being AF causative or pathogenic. Moreover, co-occurrence of common variants and rare variants on the exome chip did not appear to modulate AF
susceptibility risk. These data stress the need for caution when assigning pathogenicity of rare variants.

We evaluated 19 rare variants previously implicated with lone or familial forms of AF by linking genotypic information with phenotypic data from the Vanderbilt EMR. Notably, we did not find any of the 19 rare variants previously implicated with lone or familial forms AF to be enriched among patients with AF in either EA or AA populations. In fact, 14/19 variants were relatively common among non-AF controls (>0.1%), which suggests that these variants may not be causative (i.e. the monogenic cause of AF). It is possible that these variants are associated with very low penetrance or contribute in some as yet undefined fashion to AF susceptibility. For example, we can postulate that we did not identify a true pathogenic effect of the variants because of a missing “second hit”; the “two-hit” hypothesis wherein the combination of a rare genetic variant and another risk factor (e.g. a common AF variant or a clinical risk factor [e.g. hypertension]) modifies the risk of AF.26 We recently reported an interaction between rare and common AF risk variants modulating the penetrance of familial AF,27 but this signal was not present in the general population studied here. While we determined that 14/19 variants are not strongly correlated with AF, we are unable to determine the correlation with AF for 5/19 as they were very rare or absent among both cases and controls.

For most genetic studies, replication is vital as it adds confidence and lessens the risk of a false positive chance finding. However, for many studies implicating rare genetic variants, replication is not possible given the rarity of the phenotype in question (unable to accrue enough cases), the rarity of the variant in question (investigators would need to screen thousands of subjects to gain adequate statistical power), or both. In the absence of readily available replication cohorts for rare variants, variant function or causality can be inferred by providing
convincing functional data. For rhythm disorders such as AF, function or causality is commonly inferred through strong genotype-phenotype co-segregation patterns\textsuperscript{2, 3, 7, 9, 28}, evidence of electrophysiological effects\textsuperscript{3, 5-10, 29}, and/or rarity of the variant in question in various control populations under the assumption that truly pathogenic variants are under a strong negative selection pressure.\textsuperscript{2, 4, 8, 9, 28} While such methods represent contemporary approaches for inferring function or causality in studies involving arrhythmia susceptibility, they provide no information regarding variant effects on disease susceptibility.

The idea that rare variants previously associated with disease may represent false positive or at best weak associations, as has been suggested by others using public databases or isolated cases of normal individuals harboring previously-reported pathogenic variants. For example, Ploski et al. recently reported a patient with a normal phenotype and the Q247X deletion in muscle ring finger protein 1 (\textit{TRIM63}), a variant which has previously been reported by family and functional studies as a cause of hypertrophic cardiomyopathy (HCM).\textsuperscript{30, 31} In addition, several studies have questioned the pathogenicity of genetic variants previously associated with rare monogenic syndromes such as catecholaminergic polymorphic ventricular tachycardia, Brugada syndrome, HCM, dilated cardiomyopathy, and arrhythmogenic right ventricular cardiomyopathy,\textsuperscript{32-34} based on variant prevalence reported by the exome variant server (http://evs.gs.washington.edu/EVS/).\textsuperscript{25} However, this resource does not include phenotypic information, and did include some subjects with cardiovascular diseases such as DCM or extreme lipid values.\textsuperscript{35, 36} This study suggests broadly-phenotyped EMR-linked genotype data may be a useful resource for evaluating rare variants for pathogenicity.

\textbf{Limitations}

The rarity of the SNPs investigated limits the power of our study. In addition, while stratifying
by race is important to reduce the likelihood of ancestral informative markers driving and
association, it does mean an additional loss of power in our analysis of individuals of both EA
and AA ancestry. The rarity of the SNPs investigated limits the power of our study. Hence, we
acknowledge the fact that we may be underpowered to truly detect an association between rare
genetic variants previously associated with monogenic or familial forms of AF. A power
calculation is depicted in the supplemental appendix (Supplemental Figure 1). While we tried to
eliminate the effects associated with population stratification by performing separate analysis for
EAs and AA, our results may have been influenced by differences in population substructure.
Although our results suggest that caution should be exercised before extrapolating study findings
on genetic correlates of lone or familial forms of AF to a general form of AF, it is important to
note that this conclusion is based on the findings examining 19 SNPs (with one originally
reported in Asian subjects and so not meaningfully evaluated here). Also, low numbers for the
AA AF cases may have influenced our results. The HumanExome BeadChip v.1.0 includes rare
variants (nonsynonymous variants ≥3 times and in ≥2 call sets, splice and stop variants seen ≥2
times and in ≥2 call sets) observed among ~9000 individuals of EA ancestry, ~2000 of AA
ancestry, and ~500 samples of Hispanic and Asian ancestry
(http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Thus, singletons, which may be
associated with AF in families, are not interrogated. In the present study we did not perform
orthogonal validation of genotyped variants on the exome chip; genotype data from this platform
is known to yield reliable results.37,38

While we in the present manuscript were able to replicate known AF associations which
do add confidence in the data, we acknowledge the fact that our study cohort represents patients
either hospitalized or seen in a hospital setting which may have affected our results. Hence,
extrapolation of our results to other study cohorts should be done with caution.

**Conclusion**

A set of rare variants previously implicated with lone or familial forms AF were detected at low frequencies in a general population, where they were not associated with AF. Collectively, our results suggest that evaluating the contribution of rare variants to human phenotypes continues to require assessment of variant frequencies and associated phenotypes in the general population. Moreover, our data also suggest that the variants present on the exome chip may be too common to represent truly causative mutations in the setting of diseases of a presumed monogenic etiology.

**Funding Sources:** The datasets used for the analyses described were obtained from Vanderbilt University Medical Center’s BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. This work was also in part supported by NIH grants U19 HL65962, and R01 HL092217). PW was funded by an unrestricted research grant from the Tryg Foundation (J.nr. 7343-09, TrygFonden, Denmark).

**Conflict of Interest Disclosures:** None

**References:**


**Table 1:** Characteristics EA/AA AF cases and non-AF controls ≥50 years

<table>
<thead>
<tr>
<th></th>
<th>EA AF Cases</th>
<th>EA non-AF Controls</th>
<th>AA AF Cases</th>
<th>AA non-AF Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1,008</td>
<td>17,416</td>
<td>48</td>
<td>1,732</td>
</tr>
<tr>
<td>Age, (IQR)</td>
<td>78.1 (69.7-83.8)</td>
<td>69.9 (60.9-79.6)</td>
<td>71.3 (58.3-81.6)</td>
<td>65.5 (57.6-77.5)</td>
</tr>
<tr>
<td>Female gender, (%)</td>
<td>458 (46.0)</td>
<td>9088 (52.2)</td>
<td>26 (54.2)</td>
<td>1063 (61.4)</td>
</tr>
</tbody>
</table>

AF, atrial fibrillation; EA, European American; AA, African American, IQR, interquartile range.
### Table 2: Prevalence of 19 Variants Previously Associated with Lone or Familial Forms of AF

<table>
<thead>
<tr>
<th>Gene</th>
<th>Amino Acid Change</th>
<th>Genotype Phenotype Segregation</th>
<th>EP function</th>
<th>Discovery &quot;cohort&quot; ancestry</th>
<th>EA MAF</th>
<th>Risk allele carriers among EA AF Cases, (%)</th>
<th>Risk allele carriers among EA non-AF Controls*, (%)</th>
<th>EA P†</th>
<th>AA MAF</th>
<th>Risk allele carriers among AA AF Cases, (%)</th>
<th>Risk allele carriers among AA non-AF Controls*, (%)</th>
<th>AA P†</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NPPA</strong></td>
<td>Ser64Arg</td>
<td>S</td>
<td>LoF</td>
<td>EA</td>
<td>0.003</td>
<td>5/1008 (0.50)</td>
<td>120/17415 (0.69)</td>
<td>0.69</td>
<td>0.0006</td>
<td>0/48</td>
<td>2/1732 (0.11)</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td><strong>GREM2</strong></td>
<td>Gln76Glu</td>
<td>Increase-inhibitory-effect</td>
<td>EA</td>
<td>0.004</td>
<td>10/1008 (1.00)</td>
<td>133/17415 (0.76)</td>
<td>0.46</td>
<td>0</td>
<td>0/48</td>
<td>0/1732</td>
<td></td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td><strong>KCNE4</strong></td>
<td>Glu141Ala</td>
<td>Possible-change</td>
<td>EA</td>
<td>0.0005</td>
<td>0/1008</td>
<td>19/17414 (0.11)</td>
<td>0.62</td>
<td>0</td>
<td>0/48</td>
<td>0/1732</td>
<td></td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Phe2004Leu</td>
<td>GoF</td>
<td>EA</td>
<td>0.0009</td>
<td>1/1008</td>
<td>35/17416 (0.20)</td>
<td>0.26</td>
<td>0.0003</td>
<td>0/48</td>
<td>1/1732 (0.06)</td>
<td>0</td>
<td>2, 24</td>
<td></td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Thr1304Met</td>
<td>GoF</td>
<td>EA</td>
<td>0.0010</td>
<td>4/1008</td>
<td>33/17416 (0.19)</td>
<td>0.14</td>
<td>0</td>
<td>0/48</td>
<td>0/1732</td>
<td></td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Arg1193Gln</td>
<td>GoF</td>
<td>EA</td>
<td>0.0110</td>
<td>6/1008</td>
<td>100/17416 (0.56)</td>
<td>0.09</td>
<td>0</td>
<td>0/48</td>
<td>2/1732 (0.12)</td>
<td>1</td>
<td>8, 11, 24</td>
<td></td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Leu618Phe</td>
<td>-</td>
<td>AA</td>
<td>0</td>
<td>0/1006</td>
<td>1/17381</td>
<td>0.071</td>
<td>1/48</td>
<td>24/1725 (1.39)</td>
<td>0.50</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Ser242Tyr</td>
<td>-</td>
<td>AA</td>
<td>0.0005</td>
<td>0/1008</td>
<td>18/17416 (0.10)</td>
<td>0.62</td>
<td>0.0328</td>
<td>5/48</td>
<td>112/1732 (6.47)</td>
<td>0.25</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Arg481Trp</td>
<td>-</td>
<td>AA</td>
<td>0.0100</td>
<td>8/1008</td>
<td>3/17416 (0.02)</td>
<td>0.00</td>
<td>0</td>
<td>2/48</td>
<td>34/1732 (1.96)</td>
<td>0.25</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Leu461Val</td>
<td>S</td>
<td>EA/AA</td>
<td>0.0001</td>
<td>1/1008</td>
<td>1/17382 (0.02)</td>
<td>0.20</td>
<td>0.0127</td>
<td>0/48</td>
<td>44/1727 (2.55)</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Thr220Ile</td>
<td>LoF</td>
<td>EA</td>
<td>0.0010</td>
<td>1/1008</td>
<td>35/17416 (0.20)</td>
<td>0.72</td>
<td>0.0006</td>
<td>0/48</td>
<td>2/1732 (0.12)</td>
<td>1</td>
<td>8, 11, 24</td>
<td></td>
</tr>
<tr>
<td><strong>SCN5A</strong></td>
<td>Ser216Ile</td>
<td>GoF</td>
<td>EA</td>
<td>0.0015</td>
<td>3/1008</td>
<td>53/17416 (0.30)</td>
<td>1.00</td>
<td>0</td>
<td>0/48</td>
<td>0/1732</td>
<td></td>
<td>-</td>
<td>2, 24</td>
</tr>
<tr>
<td><strong>GATA4</strong></td>
<td>Ala411Val</td>
<td>-</td>
<td>EA</td>
<td>0.0015</td>
<td>1/1008</td>
<td>56/17410 (0.32)</td>
<td>0.37</td>
<td>0.0011</td>
<td>1/48</td>
<td>3/1731 (0.17)</td>
<td>0.10</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><strong>KCNJ8</strong></td>
<td>Ser422Leu</td>
<td>GoF</td>
<td>EA</td>
<td>0.0010</td>
<td>2/1008</td>
<td>36/17416 (0.21)</td>
<td>1.00</td>
<td>0</td>
<td>0/48</td>
<td>0/1732</td>
<td></td>
<td>-</td>
<td>5, 6</td>
</tr>
</tbody>
</table>

*Free from AF at 50 years old
†Unadjusted p-values shown; the adjusted p-values (age, sex, PC1, and PC2) yielded similar estimates
S, Genotype-phenotype segregation; EP, electrophysiological; LoF, loss-of-function; GoF, gain-of-function; EA, European-American; AA, African-American; AF, atrial fibrillation
Table 3: No interaction between common AF risk variants at the 4q25 locus and 14 rare variants previously implicated with lone or familial forms of AF among EAs

<table>
<thead>
<tr>
<th></th>
<th>No AF, (%)</th>
<th>AF, (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absence of common 4q25 variants*</td>
<td>459 (60.5)</td>
<td>25 (3.3)</td>
</tr>
<tr>
<td>Presence of common 4q25 variants*</td>
<td>259 (34.2)</td>
<td>15 (2.0)</td>
</tr>
</tbody>
</table>

* rs6843082, rs2200733, and rs17042171
AF, atrial fibrillation; SNP, single nucleotide polymorphism, EAs, European Americans
Chi-square statistic = 0.85
Examining Rare and Low-Frequency Genetic Variants Previously Associated with Lone or Familial Forms of Atrial Fibrillation in an Electronic Medical Record System: A Cautionary Note

Peter Weeke, Joshua C. Denny, Lisa Basterache, Christian Shaffer, Erica Bowton, Christie Ingram, Dawood Darbar and Dan M. Roden

Circ Cardiovasc Genet. published online November 19, 2014;
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/early/2014/11/19/CIRCGENETICS.114.000718

Data Supplement (unedited) at:
http://circgenetics.ahajournals.org/content/suppl/2014/11/19/CIRCGENETICS.114.000718.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org//subscriptions/
Supplemental Material
## Supplemental Table 1 – Additional Rare and low frequency Variants Associated with Lone or Familial Forms of AF

<table>
<thead>
<tr>
<th>Gene</th>
<th>Amino Acid Change</th>
<th>CHR:POS</th>
<th>REF/ALT</th>
<th>Accession Number</th>
<th>Rs#</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>KCNQ1</td>
<td>p.Gly229Asp</td>
<td>11:2593245</td>
<td>G/A</td>
<td>NM_000218.2</td>
<td>rs199472708</td>
<td>1</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>p.Val241Phe</td>
<td>11:2593280</td>
<td>G/T</td>
<td>NM_000218.2</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>KCNJ2</td>
<td>p.Glu229Val</td>
<td>17:68172076</td>
<td>A/T</td>
<td>NM_000891.2</td>
<td>NA</td>
<td>3</td>
</tr>
<tr>
<td>JPH2</td>
<td>p.Glu169Lys</td>
<td>20:42788922</td>
<td>C/T</td>
<td>NM_020433.4</td>
<td>NA</td>
<td>4</td>
</tr>
</tbody>
</table>
### Supplemental Table 2—*in-silico* Functional Predictions of the 19 Variants Previously Associated with Lone or Familial Forms of AF

<table>
<thead>
<tr>
<th>Gene</th>
<th>Amino Acid Change</th>
<th>CHR:POS</th>
<th>REF/ALT</th>
<th>Accession Number</th>
<th>Rs#</th>
<th>PP2</th>
<th>Grantham Score</th>
<th>Phast Cons</th>
<th>GERP</th>
<th>ESP EA</th>
<th>ESP AA</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variants present in ≥0.1% of EA/AA non-AF controls:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPPA</td>
<td>Ser64Arg</td>
<td>1:11907430</td>
<td>T/G</td>
<td>NM_006172.3</td>
<td>61757261</td>
<td>0.024</td>
<td>110</td>
<td>0</td>
<td>3.54</td>
<td>G=24/T=8576</td>
<td>G=2/T=4402</td>
<td>5</td>
</tr>
<tr>
<td>GREM2</td>
<td>Gln76Glu</td>
<td>1:240656550</td>
<td>G/C</td>
<td>NM_022469.3</td>
<td>142343894</td>
<td>0.437</td>
<td>29</td>
<td>0</td>
<td>5.03</td>
<td>C=27/G=8573</td>
<td>C=4/G=4402</td>
<td>6</td>
</tr>
<tr>
<td>KCNE4</td>
<td>Glu141Ala</td>
<td>2:223917970</td>
<td>A/C</td>
<td>NM_080671.2</td>
<td>149110444</td>
<td>1</td>
<td>107</td>
<td>1</td>
<td>6.17</td>
<td>C=5/A=8595</td>
<td>C=0/A=4406</td>
<td>7</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Phe2004Leu</td>
<td>3:38591853</td>
<td>A/G</td>
<td>NM_198056.2</td>
<td>41311117</td>
<td>0</td>
<td>22</td>
<td>0.992</td>
<td>-1.63</td>
<td>G=26/A=8342</td>
<td>G=2/A=4026</td>
<td>8, 9</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Thr1304Met</td>
<td>3:38603958</td>
<td>G/A</td>
<td>NM_198056.2</td>
<td>199473603</td>
<td>1</td>
<td>81</td>
<td>0.89</td>
<td>0.94</td>
<td>A=4/G=8464</td>
<td>A=1/G=4175</td>
<td>10</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Arg1193Gln</td>
<td>3:38616876</td>
<td>G/T</td>
<td>NM_198056.2</td>
<td>41261344</td>
<td>0.467</td>
<td>43</td>
<td>0.988</td>
<td>1.58</td>
<td>T=11/C=8589</td>
<td>T=0/C=440</td>
<td>8, 9</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Leu618Phe</td>
<td>3:38645241</td>
<td>G/A</td>
<td>NM_198056.2</td>
<td>45488304</td>
<td>0.999</td>
<td>22</td>
<td>0.806</td>
<td>4.18</td>
<td>A=0/G=8416</td>
<td>A=27/G=4095</td>
<td>8</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Ser524Tyr</td>
<td>3:38645522</td>
<td>G/T</td>
<td>NM_198056.2</td>
<td>41313691</td>
<td>0.998</td>
<td>144</td>
<td>1</td>
<td>4.06</td>
<td>T=7/G=8375</td>
<td>T=135/G=3945</td>
<td>8</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Arg481Trp</td>
<td>3:38646297</td>
<td>G/A</td>
<td>NM_198056.2</td>
<td>144511230</td>
<td>0.999</td>
<td>101</td>
<td>0.902</td>
<td>3.29</td>
<td>A=0/G=8270</td>
<td>A=39/G=3897</td>
<td>8</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Leu461Val</td>
<td>3:38646357</td>
<td>A/C</td>
<td>NM_198056.2</td>
<td>41313697</td>
<td>0.285</td>
<td>32</td>
<td>0.998</td>
<td>-0.65</td>
<td>C=2/A=8296</td>
<td>C=34/A=3914</td>
<td>8</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Thr220Ile</td>
<td>3:38655278</td>
<td>G/A</td>
<td>NM_198056.2</td>
<td>45620037</td>
<td>1</td>
<td>89</td>
<td>1</td>
<td>4.3</td>
<td>A=4/G=8404</td>
<td>A=0/G=4032</td>
<td>8, 9, 11</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Ser266Leu</td>
<td>3:38655290</td>
<td>G/A</td>
<td>NM_198056.2</td>
<td>41276525</td>
<td>1</td>
<td>145</td>
<td>1</td>
<td>4.3</td>
<td>A=11/G=8421</td>
<td>A=1/G=4089</td>
<td>9, 10</td>
</tr>
<tr>
<td>GATA4</td>
<td>Ala1115887</td>
<td>8:11615887</td>
<td>C/T</td>
<td>NM_002052.3</td>
<td>55633527</td>
<td>0.012</td>
<td>64</td>
<td>0.001</td>
<td>4.83</td>
<td>T=12/C=8588</td>
<td>T=10/C=4396</td>
<td>12</td>
</tr>
<tr>
<td>KCNJ8</td>
<td>Ser422Leu</td>
<td>12:21918667</td>
<td>G/A</td>
<td>NM_004982.3</td>
<td>72554071</td>
<td>0.013</td>
<td>145</td>
<td>0.038</td>
<td>4.45</td>
<td>A=19/G=8581</td>
<td>A=1/G=4405</td>
<td>13, 14</td>
</tr>
</tbody>
</table>

| **Variants present in <0.1% of EA/AA non-AF controls:** |
| KCND3  | Lys214Arg         | 1:112524708 | T/C     | NM_172198.2      | 142744204 | 0      | 26             | 1          | 5.51    | C=4/T=8596 | C=0/T=4406 | 7    |
| SCN5A  | Arg1897Trp        | 3:38592174 | G/A     | NM_198056.2      | 45465995  | 1      | 101             | 0.988      | 3.97    | A=3/G=8413 | A=0/G=4252 | 10   |
| SCN5A  | Glu428Lys         | 3:38647498 | C/T     | NM_198056.2      | 199473111 | 0.993  | 56             | 0.899      | 4.67    | T=2/C=8372 | T=0/C=4150 | 8    |
| KCNA5  | Gly568Val         | 12:5155016 | G/T     | NM_002234.3      | 71581017  | 0      | 109             | 0.001      | -1.72   | T=3/G=8597 | T=0/G=4406 | 7    |
| KCNJ2  | Val93Ile          | 17:68171457 | G/A      | NM_000891.2      | 147750704 | 0.004  | 29              | 0.026      | 3.48    | A=2/G=8598 | A=1/G=4405 | 15   |

PP2, polyphen2; GERP, genomic evolution rate profiler; ESP, exome sequencing project; EA, European-American; AA, African-American; AF, atrial fibrillation

PolyPhen-2: scores are evaluated as 0.000 (most probably benign) to 0.999 (most probably damaging).
Supplemental Figure 1 - Power calculation

Figure Legend: We employed a 1:18 case control ratio based on our study cohort (1,056 AF cases and 19,148 non-AF controls) assuming an additive gene-only model with a MAF=0.01 using a nominal two-sided p-value of <0.05 as our alpha. The effect sizes are reported as beta coefficient. Power calculations were performed using Quanto, (http://hydra.usc.edu/gxe/).