Targeted Deep Sequencing Reveals No Definitive Evidence for Somatic Mosaicism in Atrial Fibrillation

Jason D. Roberts, MD; James Longoria, MD; Annie Poon, PhD; Michael H. Gollah, MD; Thomas A. Dewland, MD; Pui-Yan Kwok, MD, PhD; Jeffrey E. Olgin, MD; Rahul C. Deo, MD, PhD; Gregory M. Marcus, MD, MAS

Background—Studies of ≤15 atrial fibrillation (AF) patients have identified atrial-specific mutations within connexin genes, suggesting that somatic mutations may account for sporadic cases of the arrhythmia. We sought to identify atrial somatic mutations among patients with and without AF using targeted deep next-generation sequencing of 560 genes, including genetic culprits implicated in AF, the Mendelian cardiomyopathies and channelopathies, and all ion channels within the genome.

Methods and Results—Targeted gene capture and next-generation sequencing were performed on DNA from lymphocytes and left atrial appendages of 34 patients (25 with AF). Twenty AF patients had undergone cardiac surgery exclusively for pulmonary vein isolation and 17 had no structural heart disease. Sequence alignment and variant calling were performed for each atrial–lymphocyte pair using the Burrows–Wheeler Aligner, the Genome Analysis Toolkit, and MuTect packages. Next-generation sequencing yielded a median 265-fold coverage depth (interquartile range, 64–369). Comparison of the 3 million base pairs from each atrial–lymphocyte pair revealed a single potential somatic missense mutation in 3 AF patients and 2 in a single control (12 versus 11%; P=1). All potential discordant variants had low allelic fractions (range, 2.3%–7.3%) and none were detected with conventional sequencing.

Conclusions—Using high-depth next-generation sequencing and state-of-the art somatic mutation calling approaches, no pathogenic atrial somatic mutations could be confirmed among 25 AF patients in a comprehensive cardiac arrhythmia genetic panel. These findings indicate that atrial-specific mutations are rare and that somatic mosaicism is unlikely to exert a prominent role in AF pathogenesis. (Circ Cardiovasc Genet. 2015;8:50-57. DOI: 10.1161/CIRCGENETICS.114.000650.)

Key Words: arrhythmias, cardiac • atrial fibrillation • cardiac electrophysiology • computational biology • genetics • mosaicism

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, and affected patients experience an increased risk of heart failure, stroke, and death.1,2 Despite its clinical importance, current treatment strategies for the arrhythmia, including both antiarrhythmic drugs and catheter ablation, have relatively modest long-term efficacy.3,4 The lack of definitive therapies for AF likely stems from a limited understanding of its underlying pathophysiology, emphasizing a need for novel insights.5 Recent work has increasingly highlighted a genetic contribution to the arrhythmia, especially when AF occurs in the absence of structural heart disease.6–14

Editorial see p 2

Clinical Perspective on p 57

Although a positive family history of the arrhythmia is a major risk factor for AF in the absence of overt cardiovascular disease,15 a substantial proportion of cases are sporadic.16 Given that these cases develop in the absence of identifiable risk factors, it is still probable that genetics play a role. The apparent lack of family history may be secondary to complex polygenic interactions that may lead to correspondingly complex patterns of inheritance. An alternative mechanism accounting for these sporadic cases may be de novo mutations occurring within germine or somatic cells that give rise to the atria.

A somatic mutation that develops within a myocardial progenitor cell will be absent from peripheral lymphocytes, precluding its detection on routine genetic testing. The resultant cardiac mosaicism, referring to the mutation being confined to a proportion of cells in the heart, has the potential to result in regional electric heterogeneity within the atria that could serve as an ideal substrate for the initiation and maintenance of AF.17 Guided by this concept, investigators...
identified somatic mutations within connexin genes, the molecular constituents of gap junctions, in early onset, sporadic AF patients (n=15 and n=10) who had no evidence of structural heart disease or AF risk factors.8,9 Given the high yield of screening, the investigators hypothesized that cardiac mosaicism may be a common cause of sporadic AF within structurally normal hearts.

The advent of next-generation sequencing has revolutionized cancer diagnostics, a condition whose underlying pathophysiology is largely driven by somatic mutations.20 Through the ability to rapidly screen large numbers of genes in a cost-effective manner, next-generation sequencing has led to the identification of novel genetic culprits and has improved insight into the burden of somatic mutations within tumors.21 Equally as important and in parallel with these technological advances, increasingly sophisticated probabilistic variant calling approaches have been developed to maximize sensitivity and specificity of detected variants. We sought to extend the use of these advances to the heart to evaluate the burden of atrial somatic mutations and investigate their potential impact on AF.

### Methods

#### Study Population

Consecutive consenting adult patients aged ≥18 years undergoing cardiac surgery with left atrial appendage excision at Sutter Hospital, Sacramento Medical Center, were recruited between October 1, 2010, and November 1, 2012. Patients were excluded if they had congenital heart disease, any history of rheumatic valve disease, or mitral stenosis, if a right thoracotomy approach was used, if they were unable to provide informed and witnessed signed consent, or if they were pregnant or incarcerated. Participant demographics and medical details were obtained using a study questionnaire and were verified with a subsequent chart review. All study participants provided informed written consent under protocols that were approved by the University of California, San Francisco and Sutter Hospital, Sacramento, CA.

#### Custom-Targeted Genetic Panel

The genetic panel was designed to include all ion channels within the genome and genes previously implicated in Mendelian forms of cardiac disease as of November, 2013. The list of ion channels was constructed through a search of the Uniprot Knowledgebase using the terms ion channel and human. The 502 candidates were further manually curated to verify that the listed gene encoded an ion channel. This strategy led to the identification of 398 separate genes that were incorporated into the genetic panel (Table I in the Data Supplement).

An additional 162 genes were selected based on their documented or potential involvement in primary cardiac disease (Table I in the Data Supplement). We constructed this aspect of the genetic panel through a review of the genetic culprit associated with the following conditions: long QT syndrome, short QT syndrome, Brugada syndrome, catecholaminergic polymorphic ventricular tachycardia, early repolarization syndrome, idiopathic ventricular fibrillation, arrhythmogenic right ventricular cardiomyopathy, hypertrophic cardiomyopathy, dilated cardiomyopathy, restrictive cardiomyopathy, left ventricular noncompaction, and mitochondrial cardiomyopathy. In addition, we included all genes implicated by proxy in the pathogenesis of AF from genome-wide association studies and other genes whose protein products have been implicated in the pathophysiology of the arrhythmia.

We extracted all known exons of the 560 genes using the Ensembl General Transfer Format (gtf) file annotating all transcripts in the human genome (release 68). To obtain exhaustive coverage of protein coding regions of genes of interest, a customized set of hybridization probes was designed and constructed using the Nimblegen SeqCap EZ Library kit (Roche NimbleGen, Madison, WI). In total, 3218.095 of the 3330918 bases of interest were covered by 21 probes.

#### DNA and Library Preparation

Intraoperatively, left atrial appendage samples were immediately flash-frozen in liquid nitrogen in a sterile fashion. Genomic DNA was isolated from atrial tissue using the AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA). Matching lymphocyte DNA was purified from the buffy coat using the GentraPuregene Blood Kit (Qiagen) obtained from phlebotomy performed before surgery.

To generate sequencing libraries, 1 μg of DNA from each sample was randomly sheared to ~200 base pairs using a Covaris S2 Ultrasonicator (Covaris, Woburn, MA). Subsequent library preparation was performed using the KAPA Library Preparation Kit (Kapa Biosystems, Wilmington, MA). Briefly, genomic DNA fragments were end-repaired and underwent A-tailing before adaptor ligation with 24 unique NEXTflex DNA Barcodes (Bioo Scientific, Austin, TX). Library enrichment was then performed through polymerase chain reaction amplification, followed by analysis of the size and quantity of ligated fragments using the Agilent 2100 Bioanalyzer (Santa Clara, CA). Recommended clean-up was performed at each step using Agencourt AMPure XP Beads (Beckman Coulter, Indianapolis, IN).

#### Targeted Gene Capture and Sequencing

The barcoded DNA library for each sample was then pooled with equal quantities of 23 other unique barcoded libraries to a total of 1 μg. The corresponding 24 NEXTflex DNA Barcode Blockers and COT human DNA were added to each pooled sample and then heat dried using a DNA vacuum concentrator. Hybridization of the genomic libraries with the custom-designed genetic panel was then performed consistent with manufacturer specifications.

After hybridization, the targeted fragments were pulled down and recovered using streptavidin-coupled dynabeads (Life Technologies, Grand Island, NY). Library enrichment and product analysis were repeated as detailed above. A total of 101 base pair paired-end sequencing was performed on an Illumina HiSeq 2500 sequencer with 24 samples to each lane of a flow cell. Samples were demultiplexed before analysis.

#### Analysis

The Burrows–Wheeler Aligner was used to align paired-end short reads to the human genome, whereas the Genome Analysis Toolkit was used for local realignment and recalibration.22,23 Sequencing quality metrics, including number of mapped reads, number of duplicate reads, and number of mappable bases were performed with samtools, Picard tools, and Genome Analysis Toolkit (Figures I and II in the Data Supplement).24 A total of 90 genes contained isolated base pair regions with inadequate coverage precluding reliable variant calling, collectively accounting for no >0.7% of the region of interest (Figure III in the Data Supplement).

As a quality assurance measure, we compared single-nucleotide polymorphism (SNP) minor allele frequencies observed from our data with those from the 1000 Genomes CEU population.25 We identified 2369 SNPs within our coverage area with minor allele frequency estimates of ≥5%. As inclusion of non-European individuals would generate inconsistent frequency estimates, we performed principal components analysis and removed 3 individuals based on the first principal component. Minor allele frequency estimates were generated using the remaining samples and plotted against 1000 Genomes CEU frequencies.

To improve variant calling, an estimate of cross-contamination of each sample was first obtained using the ContEst program (Figure IV in the Data Supplement).26 Per sample estimates were then input into MuTect, allowing more accurate models for variant detection. MuTect is a state-of-the-art somatic mutation caller designed to maximize sensitivity and minimize the impact of technical sources.
of false-positives. The software program uses a probabilistic model to call differences based on number of reads, mapping quality, strand bias, and estimates of cross-contamination. Importantly, tumor and normal sources are analyzed in parallel, preventing the false-positives and false-negatives that typically arise from performing a post hoc comparison of variant files after variant calling has already been performed on the samples independently.

We applied the default MuTect parameters to identify potentially discordant variants between atrial and lymphocyte DNA. We observed that the overwhelming majority of variants that emerged from this analysis corresponded to G>T transversions (Figure V in the Data Supplement). Low-level G>T transversions have become recognized as an important artifactual change that occurs with high coverage next-generation sequencing and are felt to arise secondary to oxidative damage that occurs during acoustic shearing of genomic DNA during sample preparation. This oxidative damage occurs regardless of DNA source (ie, tissue versus lymphocyte). Although not an issue for germline mutation calling because of their trace quantities not being consistent with a heterozygous state, their low levels may be confused with somatic mutations. As a result, the recommended bioinformatic approach to G>T transversions within the cancer literature has been to filter them from the analysis given the overwhelming probability that they represent sequence artifacts. Consistent with this methodologic approach, we also elected to filter G>T transversions from our analysis in an effort to minimize false-positive findings.

Among the final discordant atrial variants, SnpEff (v. 3.3) was used to assess their impact on protein coding sequence. Atrial–lymphocyte discordant variants expected to change protein sequence were further scrutinized with manual annotation, including examining their frequency in control populations, visualizing mapped reads, and analyzing whether the regions of interest map ambiguously in the genome. Discordant variants were also analyzed for their presence in multiple participants within the cohort and for evidence of occurrence in the reverse direction (absent in atrial cells and present within lymphocytes) among other members of the cohort, findings suggestive of systematic sequencing errors. A summary of our overall analytic approach to discordant variant calling is outlined in Figure 1. Somatic fractions, defined as the percentage of total reads within an atrial sample, were determined for each remaining discordant atrial variant.

Attempted verification of potential discordant variants was pursued with Sanger sequencing of atrial DNA from relevant study participants. Amplification of targeted genomic regions was performed using polymerase chain reaction (primer sequences provided in the Data Supplement) followed by DNA sequencing using the ABI PRISM dye terminator method (Applied Biosystems, Foster City, CA).

**Somatic Mutation Rate**

Because of the sensitivity of the sequencing used, it is not possible to definitively distinguish potential false-positive discordant variants that pass through our filtering protocol from actual somatic mutations with low somatic fractions. To obtain a conservative estimate, we based our calculations of somatic mutation rates on the assumption that all possible somatic mutations that passed our a priori filtering processes were real. These rates were calculated by dividing the total number of somatic mutations by the total number of nucleotides examined in both AF cases and controls. The mean somatic mutation rate was reported as the number of somatic mutations per 100 million nucleotides.

**Statistical Analysis**

Normally distributed continuous variables are presented as mean±SD and the Student t test. Comparison of categorical values was performed using the χ² and Fisher exact tests. SNP minor allele frequencies within our cohort and the 1000 Genomes Project CEU subpopulation were compared using the Pearson pairwise correlation coefficient. To evaluate the possibility that filtering of G>T transversions may have resulted in a reduced sensitivity for detecting bona fide somatic mutations, we conducted a Monte Carlo simulation analysis to estimate the anticipated number of G>T transversions that would result in functional coding changes. Given that there are 12 nucleotide changes that could be observed for somatic mutations, with each assumed to be equally likely, we were able to generate a robust bootstrap estimate of the number of G>T changes expected under the assumption of no oxidative artifact. We used the total number of prefiltered observed changes (synonymous and nonsynonymous) as an empirical distribution and performed 10000 random draws. For each draw and for each potential variant, we estimated a probability of nonsense mutation (using the observed missense rate and allowing some uncertainty) and totaled the result. A posterior mode and 95% credible interval were computed from the posterior distribution.

Two-tailed P values <0.05 were considered statistically significant. Statistical analyses were performed using Stata version 12 (College Station, TX, USA) and R.
Results

Patient Characteristics

A total of 34 patients undergoing cardiac surgery with left atrial appendage excision at Sutter Hospital, Sacramento, provided both atrial tissue and peripheral blood for sequencing analysis. Twenty-five had a history of AF and 20 were undergoing cardiac surgery exclusively for a minimally invasive AF ablation. Among the participants with AF, the mean age at diagnosis was 63.0±11.9 years and 16 (64%) were men. Seventeen participants had AF in the absence of structural heart disease, 13 of whom had no family history of the arrhythmia. There were 11 individuals with AF in the absence of all known AF risk factors (including hypertension), 9 in the absence of a known family history. The remaining baseline characteristics of the participants are summarized in Table 1.

Next-Generation Sequencing

Targeted gene capture and high-throughput sequencing permitted alignment of 16.5 million reads per sample (interquartile range [IQR], 12.2–19.5 million) at a median 265-fold coverage depth (IQR, 164–369; Figure 2). With respect to GJA1 and GJA5, the median fold-coverage depths were 267 (IQR, 171–422) and 234 (IQR, 149–369), respectively. The median number of mapped bases per sample was 3.25 million (IQR, 3.246–3.259 million) or 99.3% of the bases of interest. Comparison of the minor allele frequencies of 2369 SNPs from our cohort and the 1000 Genomes Project CEU subpopulation revealed a strong correlation (r=0.96; Figure VI in the Data Supplement).

Analysis and Filtering Steps

Bioinformatic analysis using the somatic mutation caller MuTect initially identified 8710 discordant base calls when treating lymphocytes as the reference (germline) DNA source and atrial tissue as the somatic DNA source. Notably, 8604 (98.8%) of these discordant base calls represented G>T transversions, an aforementioned common source of artificial DNA mutations arising secondary to oxidative damage during sample preparation. No G>T transversions resulting in nonsynonymous missense mutations were observed within GJA1 or GJA5. Selective filtering of false-positive G>T transversions was precluded by the absence of previously reported contextual and strand bias. Monte Carlo simulation analysis revealed a 63% probability that none of the previously filtered G>T transversions reflected bona fide functional somatic mutations (95% confidence interval, 0–2; Figure VII in the Data Supplement). Classification of G>T transversions as false-positives reduced the list of discordant base calls to 106. From this list, an additional 5 were flagged by MuTect as having poor coverage and thus reduced reliability for variant calling.

Potential Atrial-Specific Variants

The above filters resulted in a total of 101 potential somatic atrial variants. A reverse analysis, treating atrial samples as reference and lymphocytes as the somatic tissue, revealed a comparable number (93) of variants. Within the overall list, 12 represented nonsynonymous SNPs predicted to impact the protein coding regions of a total of 11 genes. The remaining discordant base calls represented synonymous SNPs or were located within intronic regions.

Of the 12 potential nonsynonymous SNPs observed within atria and not in lymphocytes, an additional 7 were found to be consistent with sequencing artifact on the basis of their (1) being observed in multiple participants (a systematic error associated with the sequencing protocol was felt to be the likely explanation, particularly given that certain of these variants were present within highly repetitive regions of DNA prone to alignment errors); (2) being observed in the reverse direction (present within lymphocytes and absent from atrial cells) among other participants; and (3) having a >50% carrier frequency within the general population. The 5 remaining discordant genetic variants did not have population data frequency available indicating that they were either rare or novel (Table 2). The discordant variants were carried by 3 of the 25 participants with AF and 2 were present within a single control participant with no prior history of the arrhythmia (12% versus 11%; P=1).

Somatic Fractions and Sanger Sequencing

The somatic fractions, defined as the percentage of total reads within the relevant atrial sample, for the 5 remaining potential nonsynonymous cardiac somatic mutations ranged from 2.3% to 7.3%. Sanger sequencing of atrial samples from each patient carrying a potential discordant variant yielded electropherograms with no evidence of a somatic mutation (Figure 3).

Somatic Mutation Rate

Among the 25 AF cases, there were 3 potential somatic mutations and an average of 3.25 million mapped base pairs per sample corresponding to an average somatic mutation rate of 4 per 100 million nucleotides (range, 0–31; SD, 10). A total of 2 potential somatic mutations were observed among the 9 control participants, and the average mapped base pairs per sample was also 3.25 million. This yields an average somatic mutation rate among controls of 7 per 100 million (range, 0–62; SD, 21).

Table 1. Clinical Characteristics of Study Participants

<table>
<thead>
<tr>
<th></th>
<th>AF (n=25)</th>
<th>No AF (n=9)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>64.2±12.0</td>
<td>78.4±8.2</td>
<td>0.003</td>
</tr>
<tr>
<td>Male</td>
<td>16 (64.0)</td>
<td>6 (66.7)</td>
<td>0.886</td>
</tr>
<tr>
<td>White race</td>
<td>24 (96.0)</td>
<td>8 (88.9)</td>
<td>0.380</td>
</tr>
<tr>
<td>Hypertension</td>
<td>13 (52.0)</td>
<td>6 (66.7)</td>
<td>0.447</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>5 (20.0)</td>
<td>1 (11.1)</td>
<td>0.549</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>5 (20.0)</td>
<td>4 (44.4)</td>
<td>0.154</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>3 (12.0)</td>
<td>1 (11.1)</td>
<td>0.943</td>
</tr>
<tr>
<td>Indication for surgery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>20 (80.0)</td>
<td>0 (0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Coronary artery bypass grafting</td>
<td>2 (8.0)</td>
<td>4 (44.4)</td>
<td>0.014</td>
</tr>
<tr>
<td>Aortic valve replacement</td>
<td>1 (4.0)</td>
<td>4 (44.4)</td>
<td>0.003</td>
</tr>
<tr>
<td>Mitral valve surgery</td>
<td>2 (8.0)</td>
<td>2 (22.2)</td>
<td>0.256</td>
</tr>
</tbody>
</table>

Data are n (%) or mean±SD. AF indicates atrial fibrillation.
Discussion

Our next-generation sequencing study targeting 560 genes found no evidence to support a role for somatic mosaicism in the pathogenesis of AF among 25 affected patients (17 with no structural heart disease and 11 with no AF risk factors including hypertension) and 9 control participants. Our study screening for atrial somatic mutations is the largest to date and is the first to assess a large number of genes. We found no missense somatic mutations within the GJA5 and GJA1 genes and no difference in the frequency of potential somatic mutations between AF cases and controls within our cohort (12% versus 11%; \( P = 1 \)). Our findings also suggest that atrial somatic mutations are rare, further reinforcing the notion that atrial mosaicism exerts a minimal role in AF pathogenesis.

Our findings contrast with previous work suggesting that \( \approx 20\% \) of nonfamilial AF occurring in the absence of structural heart disease may be secondary to somatic mutations within GJA5 and GJA1, encoding connexin 40 and 43, respectively.\(^{18,19} \) Given a \( \approx 20\% \) yield from screening just 2 genes, it was reasonable to speculate that somatic mosaicism may reflect a common underlying pathophysiology in AF with important clinical implications. In addition to failing to detect any evidence of somatic mutations within either GJA5 or GJA1, the overall somatic mutation rate among AF cases in our study was also low (4 per 100 million nucleotides). We found only 3 discordant atrial–lymphocyte genetic variants projected to result in functional changes and each had a low atrial somatic fraction. Notably, none could be confirmed by traditional Sanger sequencing. Indeed, even if real, it is unclear whether such low somatic fractions (from 2.3% to 7.3%) would have any meaningful clinical relevance. These findings argue that atrial mosaicism is unlikely to exert a prominent role in AF pathophysiology.

There are 2 possible explanations for the contrasting results between the current and previous studies, namely patient selection and sequencing artifact secondary to formalin fixation. In both previous connexin studies, the investigators restricted their cohort to individuals with early onset (age

---

**Table 2. Possible Discordant Nonsynonymous Atrial/Lymphocyte Genetic Variants**

<table>
<thead>
<tr>
<th>Genomic Position</th>
<th>Gene</th>
<th>Nucleotide Change</th>
<th>Amino Acid Change</th>
<th>Atrial Somatic Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1:27440338</td>
<td>SLC9A1</td>
<td>C&gt;A</td>
<td>L264F</td>
<td>0.073</td>
</tr>
<tr>
<td>chr2:96781756</td>
<td>ADRA2B</td>
<td>C&gt;A</td>
<td>A45S</td>
<td>0.044</td>
</tr>
<tr>
<td>chr2:166894395</td>
<td>SCN1A</td>
<td>C&gt;A</td>
<td>R918L</td>
<td>0.029</td>
</tr>
<tr>
<td>chr15:78921890</td>
<td>CHRNBA4</td>
<td>C&gt;A</td>
<td>V253F</td>
<td>0.030</td>
</tr>
<tr>
<td>chr1:152826150</td>
<td>ATP2B3</td>
<td>C&gt;A</td>
<td>D938E</td>
<td>0.023</td>
</tr>
</tbody>
</table>

*ADRA2B and SCN1A were both identified in a single control participant. Three different study participants with atrial fibrillation carried the remaining three potential discordant variants.*
<55 years), sporadic AF. The mean age at AF diagnosis in the GJA5 study was 45.1±5.9 years and patients were free of comorbidities, whereas our AF cohort was older (63.0±11.9 years), ≥50% had hypertension, and 7 of the 25 study participants with AF had an affected family member. It is conceivable that the selection criteria for the previous cohorts may have resulted in patients who had a higher burden of atrial-specific mutations within cardiac genes.

Patient selection alone, however, is unlikely to account for our discordant findings in relation to the prior AF somatic mutation reports. The previously reported high rate of atrial somatic mutations was likely affected by polymerase chain reaction artifacts after DNA extraction from formalin-fixed paraffin-embedded tissue. A growing number of articles within the oncology literature have warned about the potential for erroneously identifying somatic mutations in tumor samples previously fixed with formalin. Within the field of cardiology, initial reports suggesting that NKX2-5 somatic mutations were a common cause of congenital heart disease were subsequently shown to likely reflect false-positives secondary to formalin fixation. Of note, our atrial DNA samples were obtained from tissue that was flash-frozen immediately after excision, whereas the previous connexin mutations were identified after DNA extraction from formalin-fixed and paraffin-embedded left atrial appendage tissue. Despite the apparent overestimation of the role of atrial-specific connexin mutations in AF, it should be noted that the reported atrial somatic GJA1 mutation has also been reported as a somatic mutation from a flash-frozen gastrointestinal tumor source. It is also important to emphasize that these somatic mutations were described alongside the first connexin germline mutation associated with AF, findings subsequently substantiated by the role of rare connexin 40 mutations in familial AF.

Although it is conceivable that our bioinformatic methods for detecting somatic mutations with next-generation sequencing may have underestimated their true prevalence, it should be noted that the approach used has been validated extensively and shown to have extremely high sensitivity and specificity. At our median sequencing depth (265-fold), the estimated sensitivity for detecting somatic mutations with an atrial somatic fraction of 10% is 99.999%. When the atrial somatic fraction drops to 5%, 2%, and 1%, our sensitivity for detection correspondingly falls to 99.2%, 76.8%, and 26.9%, respectively. These sensitivity estimates far exceed those for Sanger sequencing, the approach used for identifying the previous somatic mutations in AF patients. The atrial somatic fractions for the previously documented GJA5 and GJA1 somatic mutations implicated in AF were estimated to range from 20% to 34% based on the results of allelic subcloning. Given our anticipated sensitivity, it is unlikely that we failed to detect somatic mutations with atrial somatic fractions in that range.

The challenges of recognizing artifactual mutations in previous small-scale studies highlight one of the main strengths of our study: the analysis of an unprecedented number of bases within multiple genes across multiple affected and control individuals. Focus on a small number of genes or samples would have failed to detect systematic biases, such as the G>T oxidative changes, or sample cross-contamination, and these would have been erroneously interpreted as somatic mutations. In addition, the availability of control samples allowed filtering of recurrent artifactual mutations that likely arise from alignment errors. In fact, the somatic mutation calling pipeline of most large tumor sequencing centers includes a critical filtering step whereby all variants previously observed in a large panel of hundreds of control samples are removed, as these are likely to represent artifact arising from ≥1 steps in the variant calling process.

The absence of evidence to support a role for atrial somatic mutations in AF within our cohort should not be viewed as evidence to completely rule out somatic mosaicism as a pathological mediator of AF. Although our results suggest that such a mechanism is likely rare, at least 1 other study identified a potential disease-causing somatic mutation in an arrhythmic disease. However, our results suggest that the vast majority of cases of sporadic, AF occurring in the absence of overt cardiovascular disease develop secondary to either another genetic mechanism or some as yet unknown exposure. Because this form of AF accounts for ≤30% of all AF cases and the majority seem to be sporadic, these patients comprise a substantial number in the population. Furthermore, because understanding the pathogenesis of the disease in these individuals should uncover mechanisms unique to AF itself (rather than simply an AF risk factor, such as congestive heart failure), it is critical to assure that research efforts are on the right track. Therefore, although our results are negative, we think that this comprehensive investigation is sufficiently

![Figure 3. Possible low-level somatic variants identified with next-generation sequencing failed detection with Sanger sequencing.](http://circgenetics.ahajournals.org/Content/full/17/55/Figure3.png)
robust to steer the field toward examining novel polygenic or gene–environment interactions, as well as potential behaviors or environmental influences that may be important.

Limitations

Although our study examining for atrial somatic mutations involves both the largest number of patients and genes tested to date, our cohort size of 25 AF patients and 9 controls is still modest. In addition, our bioinformatic methods for identifying somatic mutations with next-generation sequencing, although highly sensitive and state-of-the-art, could potentially have failed to identify bona fide atrial somatic mutations (particularly those with an atrial somatic fraction <2% when our sensitivity is estimated to drop <95% given our median 265-fold coverage depth). Although the frequency of potential somatic mutations was similar in both cases and controls, we cannot exclude the possibility that the discordant variants among the AF cases were pathogenic, whereas those in controls were benign. Finally, although our genetic panel covered >3 million base pairs and an exhaustive number of genes related to cardiac pathophysiology (including the genes previously implicated in somatic mutations), it remains possible that genetic mosaicism involving undiscovered variants related to AF could yet be important. We chose to restrict our analysis to 560 genes to assure high-depth coverage, thereby maximizing our sensitivity and specificity for accurately identifying somatic mutations.

Conclusions

Using high-depth next-generation sequencing and state-of-the-art somatic mutation identification approaches, we found no evidence to support a role for pathogenic atrial somatic mutations in AF using a comprehensive cardiac genetic arrhythmia panel. These findings indicate that atrial-specific mutations are rare and suggest that somatic mosaicism likely exerts a minimal role in the pathogenesis of AF.

Acknowledgments

We thank Barry S. Taylor, PhD, for his invaluable advice on somatic mutation calling.

Sources of Funding

This work was made possible by grants to Dr Marcus (American Heart Association National Innovative Research Grant and the Joseph Drown Foundation) and Dr Deo (National Institutes of Health K08 HL093861 and DP2 HL123228). Dr Roberts is supported by a Research Fellowship Grant from the Heart and Stroke Foundation of Canada.

Disclosures

None.

References

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia. Genetic differences have been demonstrated to modulate the risk of developing AF. Nonetheless, the majority of AF cases are sporadic, without evidence of familial disease. Recent work has suggested that genetics may still mediate arrhythmia development among sporadic cases secondary to somatic mutations. A somatic mutation, in contrast to a germline mutation, develops after fertilization and becomes confined to a subgroup of cells within the body. Restriction of mutations to a subgroup of organs or cells within an organism is referred to as mosaicism. Atrial mosaicism has been hypothesized as a potential cause of regional structural and electric heterogeneity that can serve as an ideal substrate for the initiation and maintenance of AF. Based on promising results from the sequencing of 2 candidate genes, investigators have hypothesized that somatic mutations may represent a common cause of sporadic AF among younger, otherwise healthy individuals. Our study sought to characterize the burden of somatic mutations within the atria. Using a sensitive form of next-generation sequencing, we sequenced 560 genes (=3 million base pairs) within atrial tissue and blood from 35 individuals (26 with AF and 9 controls). We found somatic mutations to be exceedingly rare, with indistinguishable mutation rates for cases and controls. Our findings suggest that atrial somatic mutations are unlikely to represent a common cause of sporadic AF.
Targeted Deep Sequencing Reveals No Definitive Evidence for Somatic Mosaicism in Atrial Fibrillation

Jason D. Roberts, James Longoria, Annie Poon, Michael H. Gollob, Thomas A. Dewland, Pui-Yan Kwok, Jeffrey E. Olgin, Rahul C. Deo and Gregory M. Marcus

_Circ Cardiovasc Genet._ 2015;8:50-57; originally published online November 18, 2014; doi: 10.1161/CIRCGENETICS.114.000650

_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/8/1/50

Data Supplement (unedited) at:
http://circgenetics.ahajournals.org/content/suppl/2014/11/18/CIRCGENETICS.114.000650.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/
Supplementary Material
**Primers for Sanger Sequencing**

**ADRA2B**
- Forward: 5’ - GGTGCGCTTGAGTTGTACT - 3’
- Reverse: 5’ – GACCACCAGGACCCCTACTC - 3’

**SCN1A**
- Forward: 5’ - TGTGCCATGCTGGTATTT - 3’
- Reverse: 5’ - GCTGCGAGTTTTCAAGTTGG - 3’

**SLC9A1**
- Forward: 5’ - TGTCTCCCCTGGGAAAATGAG - 3’
- Reverse: 5’- TCCTGGACACCTGCTTT - 3’

**ATP2B3**
- Forward: 5’ - AGGGTGTCAGACCTCTGTG - 3’
- Reverse: 5’ - AGATGCGGTTGCTGAAGATG - 3’

**CHRNB4**
- Forward: 5’ - CTGGTGACGAGTTGGGAGGT - 3’
- Reverse: 5’ - AGTTCCGCTCCTGGACCTAT - 3’
Supplemental Figure 1: Number of Mapped Bases Per Sample
Supplemental Figure 2: Number of Callable Bases Per Sample
Supplemental Figure 3: Genes Containing Isolated Base Pair Regions with Inadequate Coverage for Variant Calling
Supplemental Figure 4: Estimates of Cross-Contamination Among Samples using ContEst
Supplemental Figure 5: Discordant Variants Calls with an Anticipated Overwhelming Predominance of G>T Transversions Secondary to Oxidative DNA Damage
**Supplemental Figure 6:** Correlation of Single Nucleotide Polymorphism Minor Allele Frequencies from our Study Cohort and the 1000 Genomes Project CEU Population.

SNP = single nucleotide polymorphism, MAF = minor allele frequency
Supplemental Figure 7: Monte Carlo Simulation Probability Estimate for Missed Somatic Mutations Among Filtered G > T Transversions.
**Supplemental Table 1**: Cardiac Arrhythmia Genetic Panel

**Genes Implicated with AF (rare variants)**
- KCNQ1
- KCNJ2
- KCNE2
- KCNE5
- KCNA5
- KCNH2
- GJA5
- GJA1
- SCN5A
- SCN1B
- SCN1Bb
- SCN2B
- SCN3B
- NPPA
- ANK2
- GATA4
- NUP155

**Genes implicated with AF (common variants) - restricted to studies involving >1000 subjects**
- KCNH2
- GJA1
- IL6

**Loci implicated with AF - restricted to studies involving >1000 subjects**
- 4q25 - PITX2, ENPEP
- 16q22 - ZFHX3
- 1q21 - KCNN3, PMVK
- 1q24 - PRRX1
- 7q31 - CAV1, CAV2
- 9q22 - C9orf3, FBP1, FBP2
- 10q22 - SYNP02L, MYOZ1
- 14q23 - SYNE2, ESR2
- 15q24 - HCN4
- 12p12 - SOX5
- 6p21 - CDKN1A, PXT1
- 3p22 - SCN10A, SCN11A
- 5q34 - NKX2.5
- 12q24 - TBX5
- 1q21 - IL6R
Other Genes Implicated with Primary Arrhythmic Disorders That Are Not Ion Channels

Long QT Syndrome
- CAV3
- AKAP9
- SNTA1

Brugada Syndrome
- GPD1L
- MOG1
- SLMAP

CPVT
- CALM1

Idiopathic Ventricular Fibrillation
- DPP6

Autonomic

Adrenergic Receptors
- ADRA1D
- ADRA1B
- ADRA1A
- ADRA2A
- ADRA2B
- ADRA2C
- ADRB1
- ADRB2
- ADRB3

Inflammation/Fibrosis
- TGFBR1
- TGFBR2
- TGFBR3
- PDGFA
- PDGFB
- PDGFC
- PDGFD
- PDGFRα
- PDGFRβ
- CRP
- TNF
NFkB1
NFkB2
IFNG
CTGF

RAAS Genes
ACE
ACE2
ACE3
REN
AGT
AGTR1
CYP11B2

Natriuretic Hormones/Receptors
NPPA
NPPB
NPPC
NPR1
NPR2
NPR3

MMPs/TIMPs
MMP1
MMP2
MMP3
MMP9
MMP13
TIMP1
TIMP2
TIMP3
TIMP4

Cardiomyopathies (DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, RCM = Restrictive Cardiomyopathy, LVNC = Left Ventricular Non-Compaction, Mito = Mitochondrial Myopathy, ARVC = Arrhythmogenic Right Ventricular Cardiomyopathy)

Cardiomyopathies
BAG3
DCM
RBM20
DCM
RBM24
DCM
MYH7
DCM, HCM, RCM, LVNC
TNNT2           DCM,HCM,RCM
TPM1            DCM,HCM
MYBPC3          DCM,HCM
MYL3            HCM
MYL2            HCM
PSEN1           DCM
PSEN2           DCM
MYOZ2           HCM
TNNI3           DCM,HCM,RCM
ACTC1           DCM,HCM,LVNC
TTN             DCM,HCM
TNNC1           DCM,HCM
MYH6            HCM
CSRP3           DCM,HCM
TCAP            DCM,HCM
VCL             DCM,HCM
JPH2            HCM
OBSCN           HCM
DES             DCM,RCM
LMNA            DCM
SGCD            DCM
PLN             DCM
ACTN2           DCM
LDB3            DCM,LVNC
CRYAB           DCM,HCM
FHL2            DCM
ILK             DCM
LAMA4           DCM
MYPN            DCM
DMD             DCM
EMD             DCM
TAZ             DCM,LVNC,Mito
FKTN            DCM
DSP             DCM,ARVC
JUP             DCM,ARVC
PKP2            ARVC
TMEM43          ARVC
TGFB3           ARVC
DSG2            ARVC
DSC2            ARVC
DTNA            LVNC
PRKAG2          HCM
GLA             HCM (Fabry)
Sodium Channel Genes

NALCN
SCN10A
SCN11A
SCN1A
SCN1B
SCN2A
SCN2B
SCN3A
SCN3B
SCN4A
SCN4B
SCN5A
SCN7A
SCN8A
SCN9A
SCNN1A
SCNN1B
SCNN1D
SCNN1G

Potassium Channel Genes
ABCC8
ABCC9
KCNA1
KCNA10
KCNA2
KCNA3
KCNA4
KCNA5
KCNA6
KCNA7
KCNA8
KCNA9
KCNAB1
KCNAB2
KCNAB3
KCNB1
KCNB2
KCNB3
KCNB4
KCNB5
KCNB6
KCNB7
KCNB8
KCNB9
KCNC1
KCNC2
KCNC3
KCNC4
KCNC5
KCNC6
KCNE1
KCNE2
KCNE3
KCNE4
KCNE5
KCNE6
KCNE7
KCNE8
KCNE9
KCNE10
KCNE11
KCNE12
KCNE13
KCNE14
KCNE15
KCNF1
KCNG1
KCNG2
KCNG3
KCNG4
KCNH1
KCNH2
KCNH3
KCNH4
KCNH5
KCNN1
KCNN2
KCNN3
KCNN4
KCQ1
KCQ2
KCQ3
KCQ4
KCQ5
KC51
KC52
KC53
KCNT1
KCNT2
KCNU1
KCNV1
KCNV2

Calcium Channel Genes
ATP2A1
ATP2A2
ATP2A3
ATP2B1
ATP2B2
ATP2B3
ATP2B4
ATP2C1
CACFD1
CACNA1A
CACNA1B
CACNA1C
CACNA1D
CACNA1E
CACNA1F
CACNA1G
CACNA1H
CACNA1I
CACNA1S
CACNA2D1
CACNA2D2
CACNA2D3
CACNA2D4
CACNB1
CACNB2
CACNB2b
CACNB3
CACNB4
CACNG1
CACNG2
CACNG3
CACNG4
CACNG5
CACNG6
CACNG7
CACNG8
CALHM1
CATSPER1
CATSPER2
CATSPER3
CATSPER4
CATSPERB
CATSPERD
CATSPERG
ITPR1
ITPR2
ITPR3
MCU
ORAI1
ORAI2
ORAI3
PKD1
PKD1L2
PKD1L3
PKD2
PKD2L1
PKD2L2
TPCN1
TPCN2

Acid Sensing Ion Channel Genes
ASIC1
ASIC2
ASIC3
ASIC4
ASIC5
Ion Exchanger Genes
SLC11A2
SLC17A3
SLC24A2
SLC26A7
SLC26A8
SLC26A9
SLC30A1
SLC35F1
SLC4A11
SLC4A7
SLC8A1
SLC8A3
SLC9A1
SLC9C2

Ammonium Ion Transporting Genes
RHAG
RHCG

Transient Receptor Potential Cation Channel Genes
MCOLN1
MCOLN2
MCOLN3
TRPA1
TRPC1
TRPC3
TRPC4
TRPC5
TRPC6
TRPC7
TRPM1
TRPM2
TRPM3
TRPM4
TRPM5
TRPM6
TRPM7
TRPM8
TRPV1
TRPV2
TRPV3
TRPV4
TRPV5
TRPV6

Bestrophin Ion Channel Genes
BEST1
BEST2
BEST3
BEST4

Cyclic Nucleotide-gated Cation Channel Genes
CNGA1
CNGA2
CNGA3
CNGA4
CNGB1
CNGB3
HCN1
HCN2
HCN3
HCN4

GABA-gated Ion Channel Genes
GABRA1
GABRA2
GABRA3
GABRA4
GABRA5
GABRA6
GABRB1
GABRB2
GABRB3
GABRD
GABRE
GABRG1
GABRG2
GABRG3
GABRP
GABRQ
GABRR1
GABRR2
GABRR3
KCTD12
Chloride Channel Genes
ANO1
ANO10
ANO2
ANO3
ANO4
ANO5
ANO6
ANO7
ANO8
ANO9
BSND
CACA3P
CFTR
CLCA1
CLCA2
CLCA4
CLCC1
CLCN1
CLCN2
CLCN3
CLCN4
CLCN5
CLCN6
CLCN7
CLCNKA
CLCNKB
CLIC1
CLIC2
CLIC3
CLIC4
CLIC5
CLIC6
CLNS1A
TTYH1
TTYH2
TTYH3

Glutamate-gated Ion Channel Genes
GRIA1
GRIA2
GRIA3
GRIA4
GRID1
GRID2
GRIK1
GRIK2
GRIK3
GRIK4
GRIK5
GRIN1
GRIN2A
GRIN2B
GRIN2C
GRIN2D
GRIN3A
GRIN3B
GRINA

Gap Junction Ion Channel Genes
GJA1
GJA3
GJA4
GJA5
GJC1
GJD3
PANX1
PANX2
PANX3

Nicotinic-gated Cholinergic Ion Channel Genes
CHRFAM7A
CHRNA1
CHRNA10
CHRNA2
CHRNA3
CHRNA4
CHRNA5
CHRNA6
CHRNA7
CHRNA9
CHRN21
CHRN22
CHRN23
CHRN24
CHRN25
CHRNE
CHRNG

**FXYD Ion Transport Regulator Genes**

*FXYD1*
*FXYD2*
*FXYD3*
*FXYD4*
*FXYD5*
*FXYD6*
*FXYD7*

**Hydrogen Ion Channel Genes**

*ATP1B1*
*ATP5A1*
*ATP5B*
*ATP5C1*
*ATP5D*
*ATP5E*
*ATP5EP2*
*ATP5F1*
*ATP5G1*
*ATP5G2*
*ATP5G3*
*ATP5H*
*ATP5I*
*ATP5J*
*ATP5J2*
*ATP5L*
*ATP5L2*
*ATP5O*
*ATP5S*
*ATP6*
*ATP6V0C*
*ATP6VOA2*
*ATP6VOA4*
*ATP8*
*HVCN1*
*TCIRG1*

**Glycine-gated Ion Channel Genes**

*GLRA1*
*GLRA2*
GLRA3
GLRA4
GLRB

Golgi pH Regulator Ion Channel Genes
GPR89A
GPR89B

Mitochondrial Voltage Dependent Anion Channel Genes
VDAC1
VDAC2
VDAC3

Ryanodine Receptor & Related Genes
CASQ1
CASQ2
JPH1
JPH2
JPH3
JPH4
RYR1
RYR2
RYR3
TRDN
ASPH

Piezo-type Mechanosensitive Ion Channel Genes
PIEZO1
PIEZO2

Serotonin-gated Ion Channel Receptor Genes
HTR3A
HTR3B
HTR3C
HTR3D
HTR3E

Transmembrane Channel-like Genes
TMC4
TMC5
TMC7
Trimeric Intracellular Cation Channel Genes

*TMEM37*
*TMEM38A*
*TMEM38B*

Zinc Activated Ligand Gated Ion Channel Genes

*ZACN*

Purinergic Receptor Genes

*P2RX1*
*P2RX2*
*P2RX3*
*P2RX4*
*P2RX5*
*P2RX6*
*P2RX7*

Other Calcium Genes

*CALM2*
*CALM3*
*CAMK2D*

Phosphodiesterase Genes

*PDE4A*
*PDE4B*
*PDE4C*
*PDE4D*

Other

*NOS1AP (QT-interval)*