DOI: 10.1161/CIRCGENETICS.114.000650

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

Running title: *Roberts et al.; Somatic Mosaicism and Atrial Fibrillation*

Jason D. Roberts, MD[1]; James Longoria, MD[2]; Annie Poon, PhD[3]; Michael H. Gollob, MD[4]; Thomas A. Dewland, MD[1]; Pui-Yan Kwok, MD, PhD[3]; Jeffrey E. Olgin, MD[1]; Rahul C. Deo, MD, PhD[3,5]; Gregory M. Marcus, MD, MAS[1]

[1]Section of Cardiac Electrophysiology, Division of Cardiology, Department of Medicine, [3]Cardiovascular Research Institute, [5]Department of Medicine, California Institute for Quantitative Biosciences and Institute for Human Genetics, University of California San Francisco, San Francisco; [2]Division of Cardiovascular Surgery, Sutter Health, Sacramento, CA; [4]Arrhythmia Research Laboratory, University of Ottawa Heart Institute, Ottawa, Ontario, Canada

Correspondence:

Gregory M. Marcus, MD, MAS
University of California San Francisco
505 Parnassus Ave, M1180B
San Francisco, CA 94143-0124
Tel: (415) 476-5706
Fax: (415) 476-3505
E-mail: marcusg@medicine.ucsf.edu

Rahul C. Deo, MD, PhD
University of California San Francisco
555 Mission Bay Boulevard South
San Francisco, CA 94158
Tel: (415) 476-9593
Fax: (866) 861-0066
E-mail: rahul.deo@ucsf.edu

Journal Subject Codes: [5] Arrhythmias, clinical electrophysiology, drugs, [109] Clinical genetics

DOI: 10.1161/CIRCGENETICS.114.000650

Abstract:

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 (IQR 164-369). 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 vs. 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.

Key words: atrial fibrillation, genetics, bioinformatics, arrhythmia, cardiac electrophysiology, mosaicism, somatic mutation

DOI: 10.1161/CIRCGENETICS.114.000650

Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, and affected patients suffer from an increased risk of heart failure, stroke, and death.[1,2] Despite its clinical importance, current treatment strategies for the arrhythmia, including both anti-arrhythmic 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]

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 germline 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 electrical 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.[18,19] 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 in order to evaluate the burden of atrial somatic mutations and investigate their potential impact on AF.

Methods

Study Population

Consecutive consenting adult patients ≥ 18 years old 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 employed, 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 (UCSF) and Sutter Hospital, Sacramento, CA.

DOI: 10.1161/CIRCGENETICS.114.000650

Custom Targeted Genetic Panel

The genetic panel was designed in order 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 (Data Supplement, Table 1).

An additional 162 genes were selected based on their documented or potential involvement in primary cardiac disease (Data Supplement, Table 1). We constructed this aspect of the genetic panel through a review of the genetic culprits 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 Non-Compaction, 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). In order 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, 3,218,095 of the 3,330,918 bases of interest were covered by one or more probes.

DOI: 10.1161/CIRCGENETICS.114.000650

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 prior to surgery.

In order to generate sequencing libraries, 1 microgram of DNA from each sample was randomly sheared to approximately 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 prior to adaptor ligation with 24 unique NEXTflex DNA Barcodes (Bioo Scientific, Austin, TX). Library enrichment was then performed through polymerase chain reaction (PCR) 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 & 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 microgram. 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.

Following hybridization, the targeted fragments were pulled down and recovered using

Streptavidin-coupled Dynabeads (Life Technologies, Grand Island, NY). Library enrichment and product analysis was repeated as detailed above. 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 prior to analysis.

Analysis

The Burrows-Wheeler Aligner (*bwa*) was used to align paired-end short reads to the human genome, while the Genome Analysis Toolkit (GATK) 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 GATK (Data Supplement, Figures 1 and 2).[23,24] A total of 90 genes contained isolated base pair regions with inadequate coverage precluding reliable variant calling, collectively accounting for no more than 0.7% of the region of interest (Data Supplement, Figure 3).

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 at least 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 (Data Supplement, Figure 4).[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.[27] 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 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 (Data Supplement, Figure 5). 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.[28] This oxidative damage occurs regardless of DNA source (ie. tissue versus lymphocyte). Although not an issue for germ-line mutation calling due to 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.[28] 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.[29] 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

DOI: 10.1161/CIRCGENETICS.114.000650

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 analytical 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, USA).

Somatic Mutation Rate

Because of the sensitivity of the sequencing employed, 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. In order 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 means $\pm$ standard deviation and the

Student's t-test. Comparison of categorical values was performed using the Chi-squared and Fisher's Exact tests. SNP minor allele frequencies within our cohort and the 1000 Genomes Project CEU subpopulation were compared using the Pearson pairwise correlation coefficient. In order 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 pre-filtered observed changes (synonymous and non-synonymous) as an empirical distribution and performed 10,000 random draws. For each draw and for each potential variant, we estimated a probability of missense 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 $\pm$ 11.9 years and 16 (64%) were male. 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 (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 (rho = 0.96; Data Supplement, Figure 6).

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 artifactual DNA mutations arising secondary to oxidative damage during sample preparation.[28] No G>T transversions resulting in non-synonymous 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% CI: 0-2) (Data Supplement, Figure 7). 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 non-synonymous 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 non-synonymous 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; 3) having a greater than 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% vs. 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 non-synonymous 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, standard deviation: 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, standard deviation:21).

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% vs. 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 approximately 20% of non-familial 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 nearly 20% yield from screening just two 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 very 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 very low atrial somatic fraction. Notably, none could be confirmed by traditional Sanger sequencing. Indeed, even if "real", it is unclear if such low somatic fractions (from 2.3-7.3%) would have any meaningful clinical relevance. These findings argue that atrial mosaicism is unlikely to exert a prominent role in AF pathophysiology.

There are two 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<55), sporadic AF.[18,19] The mean age at AF diagnosis in the *GJA5* study was 45.1 $\pm$ 5.9 years and patients were free of co-morbidities, while our AF cohort was older (63.0 $\pm$ 11.9 years), approximately 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 that 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 impacted by PCR artifacts following DNA extraction from formalin-fixed paraffin-embedded tissue.[30] A growing number of papers within the oncology literature have warned about the potential for erroneously identifying somatic mutations in tumor

samples previously fixed with formalin.[30–33] 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.[34–36] Of note, our atrial DNA samples were obtained from tissue that was flash-frozen immediately after excision, whereas the previous connexin mutations were identified following DNA extraction from formalin-fixed and paraffin-embedded left atrial appendage tissue.[18,19] 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.[37] 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.[38,39]

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 utilized has been extensively validated and shown to have extremely high sensitivity and specificity.[20,40,41] 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.[27] These sensitivity estimates far exceed those for Sanger sequencing, the approach utilized 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.[18,19] Given our anticipated sensitivity, it is very 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 highlights 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. Additionally, 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 one or more steps in the variant calling process.[42]

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 pathophysiological mediator of AF. Although our results suggest that such a mechanism is likely rare, at least one other study identified a potential disease-causing somatic mutation in an arrhythmic disease.[43] 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 up to 30% of all AF cases and the majority appear to be sporadic, these patients comprise a substantial number in the population.[16,44] Furthermore, because understanding the etiology 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 believe this comprehensive investigation is sufficiently robust to steer the field towards 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 $\leq$ 2% when our sensitivity is estimated to drop below 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 while those in controls were benign. Finally, although our genetic panel covered more than 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 in order 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: The authors thank Barry S. Taylor, PhD for his invaluable advice regarding somatic mutation calling.

Funding Sources: This work was made possible by grants to G.M.M. (American Heart Association National Innovative Research Grant and the Joseph Drown Foundation) and R.C.D (NIH K08 HL093861 and DP2 HL123228). J.D.R. is supported by a Research Fellowship Grant from the Heart & Stroke Foundation of Canada.

Conflict of Interest Disclosures: None.

References:

1. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke*. 1991;22:983–988.

2. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. *Circulation*. 1998;98:946–952.

3. Singh BN, Singh SN, Reda DJ, Tang XC, Lopez B, Harris CL, et al. Amiodarone versus sotalol for atrial fibrillation. *N Engl J Med*. 2005;352:1861–1872.

4. Weerasooriya R, Khairy P, Litalien J, Macle L, Hocini M, Sacher F, et al. Catheter ablation for atrial fibrillation: are results maintained at 5 years of follow-up? *J Am Coll Cardiol*. 2011;57:160–166.

5. Nattel S. New ideas about atrial fibrillation 50 years on. *Nature*. 2002;415:219–226.

6. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature*. 2007;448:353–357.

7. Gudbjartsson DF, Holm H, Gretarsdottir S, Thorleifsson G, Walters GB, Thorgeirsson G, et al. A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet*. 2009;41:876–878.

8. Benjamin EJ, Rice KM, Arking DE, Pfeufer A, van Noord C, Smith AV, et al. Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet*.

2009;41:879–881.

9. Ellinor PT, Lunetta KL, Glazer NL, Pfeufer A, Alonso A, Chung MK, et al. Common variants in KCNN3 are associated with lone atrial fibrillation. *Nat Genet*. 2010;42:240–244.

10. Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet*. 2012;44:670–675.

11. Marcus GM, Alonso A, Peralta CA, Lettre G, Vittinghoff E, Lubitz SA, et al. European ancestry as a risk factor for atrial fibrillation in African Americans. *Circulation*. 2010;122:2009–2015.

12. Chen Y-H, Xu S-J, Bendahhou S, Wang X-L, Wang Y, Xu W-Y, et al. KCNQ1 gain-of-function mutation in familial atrial fibrillation. *Science*. 2003;299:251–254.

13. Ellinor PT, Shin JT, Moore RK, Yoerger DM, MacRae CA. Locus for atrial fibrillation maps to chromosome 6q14-16. *Circulation*. 2003;107:2880–2883.

14. Lubitz SA, Yin X, Fontes JD, Magnani JW, Rienstra M, Pai M, et al. Association between familial atrial fibrillation and risk of new-onset atrial fibrillation. *JAMA*. 2010;304:2263–2269.

15. Wyse DG, Van Gelder IC, Ellinor PT, Go AS, Kalman JM, Narayan SM, et al. Lone atrial fibrillation: does it exist? *J Am Coll Cardiol*. 2014;63:1715–1723.

16. Marcus GM, Smith LM, Vittinghoff E, Tseng ZH, Badhwar N, Lee BK, et al. A first-degree family history in lone atrial fibrillation patients. *Heart Rhythm*. 2008;5:826–830.

17. Roberts JD, Gollob MH. Impact of genetic discoveries on the classification of lone atrial fibrillation. *J Am Coll Cardiol*. 2010;55:705–712.

18. Gollob MH, Jones DL, Krahn AD, Danis L, Gong X-Q, Shao Q, et al. Somatic mutations in the connexin 40 gene (GJA5) in atrial fibrillation. *N Engl J Med*. 2006;354:2677–2688.

19. Thibodeau IL, Xu J, Li Q, Liu G, Lam K, Veinot JP, et al. Paradigm of genetic mosaicism and lone atrial fibrillation: physiological characterization of a connexin 43-deletion mutant identified from atrial tissue. *Circulation*. 2010;122:236–244.

20. Watson IR, Takahashi K, Futreal PA, Chin L. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet*. 2013;14:703–718.

21. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010;11:685–696.

22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.

23. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303.

24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.

25. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.

26. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics*. 2011;27:2601–2602.

27. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31:213–219.

28. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res*. 2013;41:e67.

29. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92.

30. Quach N, Goodman MF, Shibata D. In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. *BMC Clin Pathol*. 2004;4:1.

31. Wong C, DiCioccio RA, Allen HJ, Werness BA, Piver MS. Mutations in BRCA1 from fixed, paraffin-embedded tissue can be artifacts of preservation. *Cancer Genet Cytogenet*. 1998;107:21–27.

32. Williams C, Pontén F, Moberg C, Söderkvist P, Uhlén M, Pontén J, et al. A high frequency of sequence alterations is due to formalin fixation of archival specimens. *Am J Pathol*. 1999;155:1467–1471.

33. Akbari M, Hansen MD, Halgunset J, Skorpen F, Krokan HE. Low copy number DNA template can render polymerase chain reaction error prone in a sequence-dependent manner. *J Mol Diagn JMD*. 2005;7:36–39.

34. Reamon-Buettner SM, Hecker H, Spanel-Borowski K, Craatz S, Kuenzel E, Borlak J. Novel NKX2-5 mutations in diseased heart tissues of patients with cardiac malformations. *Am J Pathol*. 2004;164:2117–2125.

35. Draus JM Jr, Hauck MA, Goetsch M, Austin EH 3rd, Tomita-Mitchell A, Mitchell ME. Investigation of somatic NKX2-5 mutations in congenital heart disease. *J Med Genet*. 2009;46:115–122.

36. Salazar M, Consoli F, Villegas V, Caicedo V, Maddaloni V, Daniele P, et al. Search of somatic GATA4 and NKX2.5 gene mutations in sporadic septal heart defects. *Eur J Med Genet*. 2011;54:306–309.

37. Dubina MV, Iatckii NA, Popov DE, Vasil'ev SV, Krutovskikh VA. Connexin 43, but not connexin 32, is mutated at advanced stages of human sporadic colon cancer. *Oncogene*. 2002;21:4992–4996.

38. Yang Y-Q, Liu X, Zhang X-L, Wang X-H, Tan H-W, Shi H-F, et al. Novel connexin40 missense mutations in patients with familial atrial fibrillation. *Europace*. 2010;12:1421–1427.

39. Sun Y, Yang Y-Q, Gong X-Q, Wang X-H, Li R-G, Tan H-W, et al. Novel germline GJA5/connexin40 mutations associated with lone atrial fibrillation impair gap junctional intercellular communication. *Hum Mutat*. 2013;34:603–609.

40. Gonzalez-Perez A, Mustonen V, Reva B, Ritchie GRS, Creixell P, Karchin R, et al. Computational approaches to identify functional genetic variants in cancer genomes. *Nat Methods*. 2013;10:723–729.

41. Wang Q, Jia P, Li F, Chen H, Ji H, Hucks D, et al. Detecting somatic point mutations in cancer genome sequencing data: a comparison of mutation callers. *Genome Med*. 2013;5:91.

42. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505:495–501.

43. Lerman BB, Dong B, Stein KM, Markowitz SM, Linden J, Catanzaro DF. Right ventricular outflow tract tachycardia due to a somatic cell mutation in G protein subunitalphai2. *J Clin Invest*. 1998;101:2862–2868.

44. Chugh SS, Blackshear JL, Shen WK, Hammill SC, Gersh BJ. Epidemiology and natural history of atrial fibrillation: clinical implications. *J Am Coll Cardiol*. 2001;37:371–378.

DOI: 10.1161/CIRCGENETICS.114.000650

Table 1: Clinical Characteristics of Study Participants

	AF n= 25	No AF n= 9	p value
Age (years)	64.2 ± 12.0	78.4 ± 8.2	0.003
Male	16 (64.0)	6 (66.7)	0.886
White Race	24 (96.0)	8 (88.9)	0.380
Hypertension	13 (52.0)	6 (66.7)	0.447
Diabetes Mellitus	5 (20.0)	1 (11.1)	0.549
Coronary Artery Disease	5 (20.0)	4 (44.4)	0.154
Congestive Heart Failure	3 (12.0)	1 (11.1)	0.943
Indication For Surgery			
AF Ablation	20 (80.0)	0 (0)	<0.001
Coronary Artery Bypass Grafting	2 (8.0)	4 (44.4)	0.014
Aortic Valve Replacement	1 (4.0)	4 (44.4)	0.003
Mitral Valve Surgery	2 (8.0)	2 (22.2)	0.256

Data are n (%) or mean ± standard deviation

DOI: 10.1161/CIRCGENETICS.114.000650

Table 2: Possible Discordant Non-Synonymous Atrial/Lymphocyte Genetic Variants

Genomic Position	Gene	Nucleotide Change	Amino Acid Change	Atrial Somatic Fraction
chr1:27440338	*SLC9A1*	C > A	L264F	0.073
chr2:96781756	*ADRA2B**	C > A	A45S	0.044
chr2:166894395	*SCN1A**	C > A	R918L	0.029
chr15:78921890	*CHRNB4*	C > A	V253F	0.030
chrX:152826150	*ATP2B3*	C > A	D938E	0.023

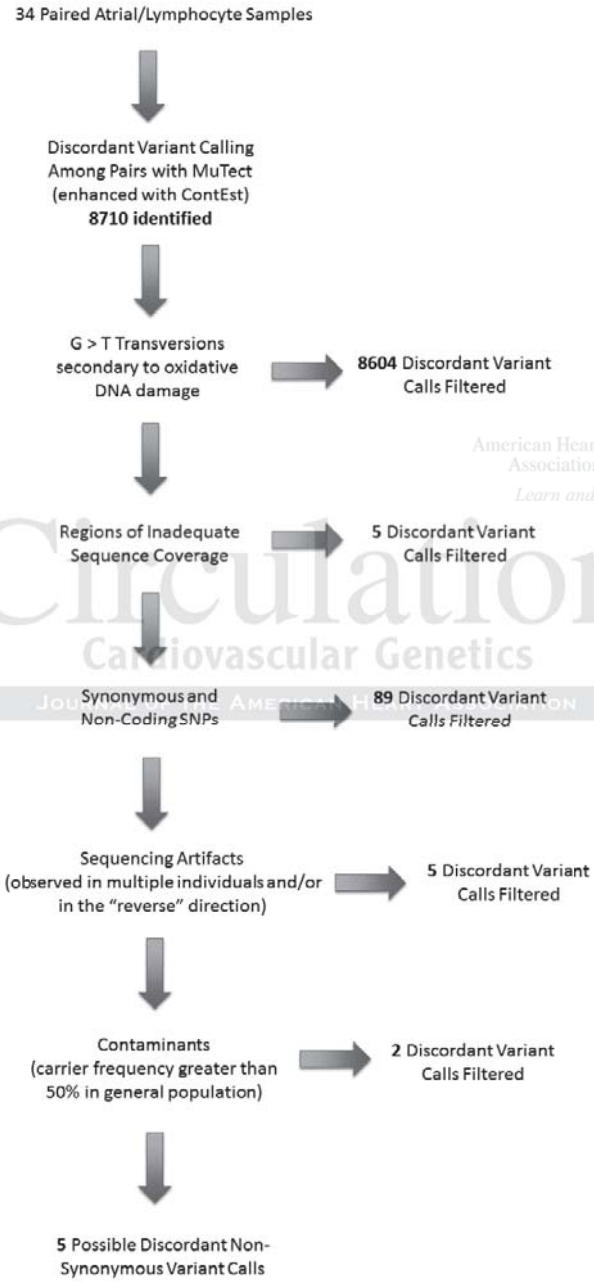
* *ADRA2B* and *SCN1A* were both identified in a single control participant. 3 different study participants with atrial fibrillation carried the remaining 3 potential discordant variants.

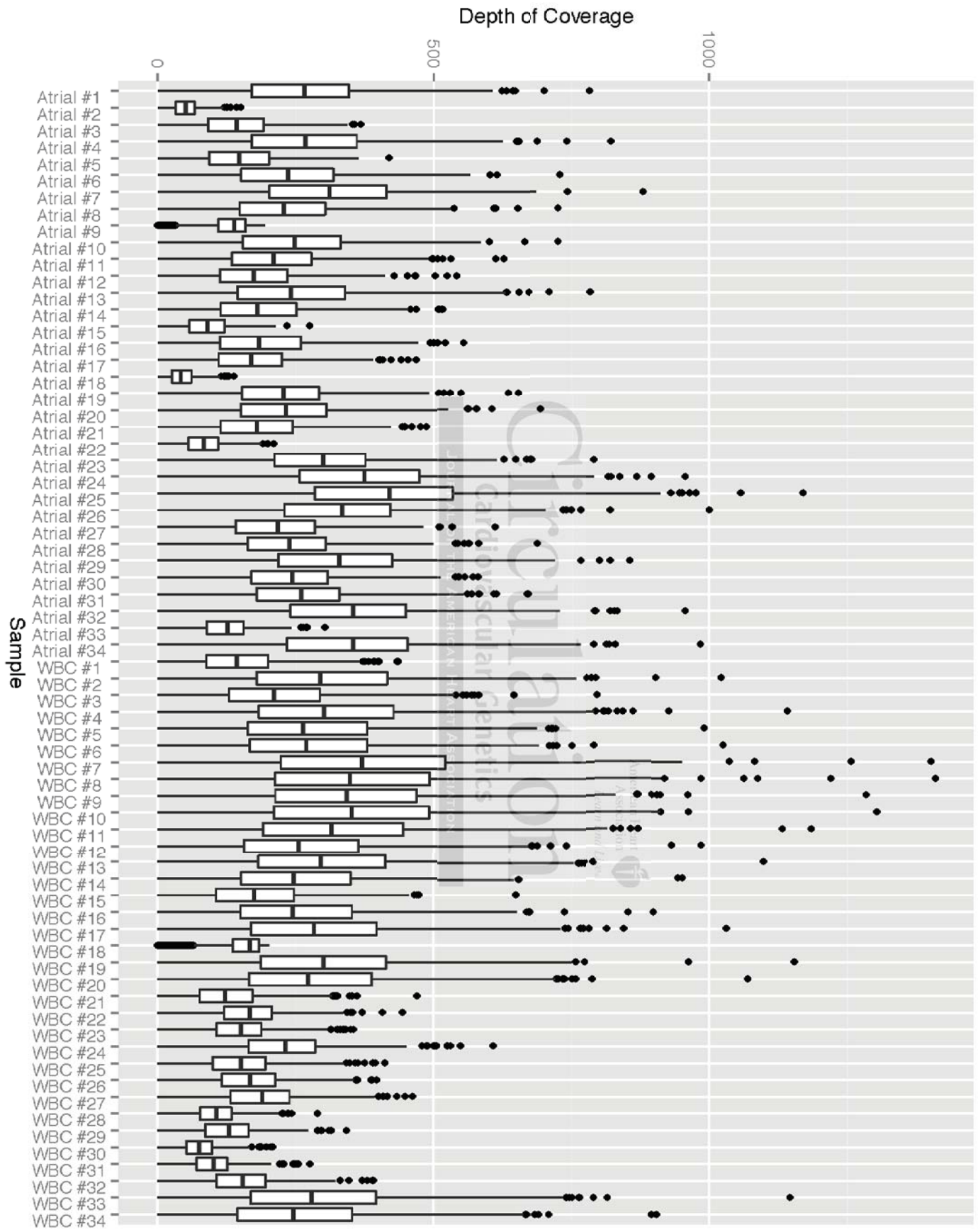
Figure Legends:

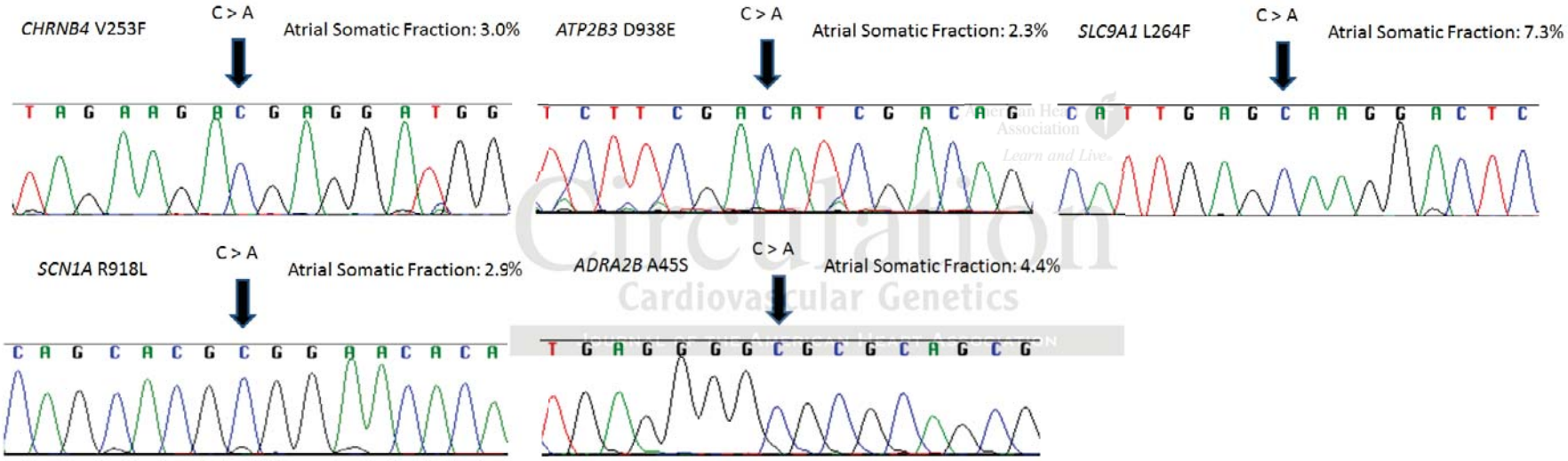
Figure 1: Stepwise Filtering Approach for the Identification of Potential Pathogenic Atrial Somatic Mutations.

Figure 2: Sequencing Coverage Depth Among Atrial and Lymphocyte Samples. Boxes represent 25th to 75th quartiles and lines within boxes represent median values. Outliers are displayed by distinct dots.

Figure 3: Possible Low Level Somatic Variants Identified with Next-Generation Sequencing Failed Detection with Sanger Sequencing.






CHRNB4 V253F
C > A
Atrial Somatic Fraction: 3.0%
T A G A A G A C G A G G A T G G
ATP2B3 D938E
C > A
Atrial Somatic Fraction: 2.3%
T C T T C G A C A T C G A C A G
SLC9A1 L264F
C > A
Atrial Somatic Fraction: 7.3%
C A T T G A G C A A G G A C T C
SCN1A R918L
C > A
Atrial Somatic Fraction: 2.9%
C A G C A C G C G G A A C A C A
ADRA2B A45S
C > A
Atrial Somatic Fraction: 4.4%
T G A G G G C G C G C A G C G