Atrial fibrillation (AF) is the most common arrhythmia and is associated with increased morbidity. As the population ages and the prevalence of AF continues to rise, the socioeconomic consequences of AF will become increasingly burdensome. Although there are well-defined clinical risk factors for AF, a significant heritable component is also recognized. To identify the molecular basis for the heritability of AF, investigators have used a combination of classical Mendelian genetics, candidate gene screening, and genome-wide association studies. However, these avenues have, as yet, failed to define the majority of the heritability of AF. The goal of this review is to describe the results from both candidate gene and genome-wide studies, as well as to outline potential future avenues for creating a more complete understanding of AF genetics. Ultimately, a more comprehensive view of the genetic underpinnings for AF will lead to the identification of novel molecular pathways and improved risk prediction of this complex arrhythmia. (Circ Res. 2014;114:1469-1482.)

Key Words: arrhythmias, cardiac ■ atrial fibrillation ■ genetics
Atrial Fibrillation

Nonstandard Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF</td>
<td>atrial fibrillation</td>
</tr>
<tr>
<td>ANP</td>
<td>atrial natriuretic peptide</td>
</tr>
<tr>
<td>eQTL</td>
<td>expression quantitative trait loci</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association studies</td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
</tr>
<tr>
<td>SNP</td>
<td>single-nucleotide polymorphism</td>
</tr>
</tbody>
</table>

series of related families with early onset AF. Although the specific causative gene at this locus remains unknown, this study helped to establish a genetic basis for some patients with AF (Figure 1).

In a seminal article published in Science in 2003, Chen et al. identified the first gene for familial AF. Using a large Chinese kindred with autosomal dominant AF, they found a gain-of-function mutation in KCNQ1 or the gene encoding the \( \alpha \) subunit of the slowly repolarizing potassium channel current, \( I_{Ks} \). The identification of a well-known ion channel mutation for AF quickly led many groups to turn to candidate gene screening of a wide range of cardiac genes. Indeed, several additional gain-of-function variants have been identified in KCNQ1. A challenge with the interpretation of these candidate gene studies is that most lack convincing genetic support in the form of variant transmission in extended families. With this limitation in mind, we have provided an overview of the genes related to AF in the following section, and we have included a detailed compendium of known AF variants in Table 1.

Ion Channel Variation in AF

In the broadest terms, the majority of functionally validated, AF-associated potassium channel variants have a gain of channel function, with an expected shortening of the atrial action potential duration and atrial refractory period. In addition to \( \text{KCNQ1} \), mutations have been identified in potassium channel genes, including \( \text{KCN5} \), \( \text{KCN3} \), and \( \text{KCN2} \) and accessory subunits \( \text{KCN1} \), \( \text{KCN2} \), \( \text{KCN3} \), and \( \text{KCN5} \). Alternatively, it has also been demonstrated that prolongation of atrial action potentials caused by loss-of-function potassium channel mutations can lead to early afterdepolarizations and AF. After an initial description by the Olson laboratory, additional mutations in \( \text{KCN5} \) of the \( I_{Ks} \) current have been reported in subsequent years.

Variation in sodium channel subunits has also been identified as an important factor in the development of familial AF. Voltage-gated sodium channels (NaV) are responsible for initiating the upstroke during phase 0 of cardiac action potential and for the coordinated propagation of the action potential throughout the atria. Cardiac sodium channels are composed of a pore-forming \( \alpha \) subunit, and \( \beta \) subunits, which can alter channel trafficking and inactivation kinetics. To date, AF-causing variants have been observed in both the major cardiac sodium channel, encoded by \( \text{SCN5A} \), and 4 of its associated \( \beta \) subunits. Similar to reports of potassium channel variation, both loss-of-function and gain-of-function variations seem to be capable of creating a proarrhythogenic substrate.

Other Genes Discovered in Individuals and Families With AF

Several variants have also been identified in genes that do not directly alter the atrial action potential, but instead would be expected to instigate the onset of AF through alternative mechanisms. For example, Gollob et al. discovered a series of somatic mutations in \( \text{GJA5} \), which encodes the gap junctional protein, Connexin 40. Interestingly, although this mutation was observed in atrial biopsies, it was not found in DNA isolated from blood. The extent to which somatic mutation or mosaicism contribute to the AF is unclear, and further study is often limited by the difficulty in obtaining primary samples. Furthermore, lending support to \( \text{GJA5} \) as an AF candidate gene, recently, several reports have identified additional \( \text{GJA5} \) loss-of-function variants that associate with disease. Because gap junctions are responsible for propagation of action potentials between cardiomyocytes, disruption of these complexes can result in reduced conduction velocity throughout the atrium, conditions that would be predicted to promote re-entry.

Another study identified a frameshift mutation that resulted in early truncation of \( \text{NPPA} \) in an extensive family with lone AF. \( \text{NPPA} \) encodes the precursor for atrial natriuretic peptide (ANP), an important factor in the regulation of sodium homeostasis and, by association, blood pressure. This mutation was shown to increase the resistance of ANP to degradation, in essence causing an increase in ANP-mediated signaling. In this study, when the mutant, mature ANP was perfused in a rat, whole-heart Langendorff model, there was significant shortening of the atrial action potential duration. Although the action potential duration shortening may be the major phenotype observed after the acute treatment, prolonged systemic exposure to the mutant ANP could also cause AF-inducing structural remodeling, as seen in canine models, and could also be supported by the recent identification of an autosomal recessive mutation in \( \text{NPPA} \) in a family with severe atrial dilated cardiomyopathy.

Finally, genes broadly characterized under the umbrella of developmentally related cardiac transcription factors have also been identified as being associated with AF. Specifically, genetic variation in \( \text{NKX2.5} \), \( \text{PITX2} \), \( \text{GATA4} \), \( \text{GATA5} \), and \( \text{GATA6} \) has been described although the mechanisms whereby these lead to disease have remained unclear.

GWAS of AF

Until the mid-2000s, linkage and candidate gene sequencing methods were the predominant approaches used to identify AF genes. In 2005, a novel technique, termed a GWAS, was used to identify genetic loci associated with age-related macular degeneration. A GWAS relies on the unbiased comparison of common single-nucleotide polymorphisms (SNPs) throughout the genome. SNPs that occur with different frequency in individuals with a disease versus controls can localize disease-related genetic loci. Although a potentially powerful tool for identifying genetic variation associated with common diseases, careful correction for multiple testing is necessary. Since that initial publication, >1700 GWAS have been published listing associations at ≈12000 SNPs. Among cardiovascular diseases, this technique has
successfully identified the risk loci for premature myocardial infarction, hypertension, lipid levels, and electrocardiographic intervals, among others.

Initial GWAS of AF
The first GWAS performed for AF was published in 2007 and identified a region on chromosome 4q25 (sentinel SNP rs2200733), which was associated with AF in those of European and Asian descent. Subsequently, these findings were broadly replicated in individuals of European, Asian, and African descent. Further analysis also identified the same genomic region as being associated with an increased risk of cardioembolic stroke and a prolonged PR interval. In a recent meta-analysis of AF GWAS data, carriers of a single copy of the 4q25 variant had a ≈65% increased risk of AF (odds ratio [OR], 1.64 for rs2634073; P=1.8×10−74). A follow-up fine-mapping study of the 4q25 locus identified ≥3 independent association signals within this region. When these 3 signals are considered together, there is a subset of ≈1% of the population who has all 6 risk alleles and a =6-fold risk of AF (OR, 6.02; P=1.2×10−36).

The 4q25 risk region lies in a relatively gene-sparse intergenic region ≈150-kb upstream from the PITX2 gene. Although at present there are no data linking the SNPs in this region to the expression levels of PITX2, our current understanding of PITX2 function suggests a plausible link with AF. PITX2 encodes the paired-like homeodomain 2 protein, a transcription factor that is crucial during embryogenesis and, notably for AF, cardiogenesis. PITX2 expression is near the closing stages of the left/right asymmetry program in vertebrates, with 100-fold higher expression in the left versus the right atrium. Critical roles for PITX2 have also been identified for the formation of the atrial septum, outflow tract, SA node, and the pulmonary vein myocardial sleeves. The last of these is of particular note given the prevalence of ectopic electric foci arising from the pulmonary vein in patients with AF and the common approach of electrically isolating the pulmonary veins to treat recurrent AF.
Table 1. Compendium of Atrial Fibrillation Genetic Variants Identified in Families and Individuals

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Name</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC9</td>
<td>ATP-binding cassette, subfamily C, member 9</td>
<td>$I_{Na}$ current</td>
<td>15</td>
</tr>
<tr>
<td>GATA4</td>
<td>Transcription factor GATA-4</td>
<td>Cardiac development</td>
<td>16–19</td>
</tr>
<tr>
<td>GATA5</td>
<td>Transcription factor GATA-5</td>
<td>Cardiac development</td>
<td>20–22</td>
</tr>
<tr>
<td>GATA6</td>
<td>Transcription factor GATA-6</td>
<td>Cardiac development</td>
<td>23–25</td>
</tr>
<tr>
<td>GJA5</td>
<td>Connexin 40</td>
<td>$I_{Ca}$ current</td>
<td>26–31</td>
</tr>
<tr>
<td>GREM2</td>
<td>Gremlin-2</td>
<td>BMP antagonist</td>
<td>32</td>
</tr>
<tr>
<td>HCN4</td>
<td>Hyperpolarization activated cyclic nucleotide-gated potassium channel 4</td>
<td>$I_{Ca}$ current</td>
<td>33</td>
</tr>
<tr>
<td>JPH2</td>
<td>Junctophilin-2</td>
<td>Ca$^{2+}$ homeostasis</td>
<td>34</td>
</tr>
<tr>
<td>KCN5</td>
<td>Potassium voltage-gated channel, subfamily H (eag related), member 2</td>
<td>$I_{Na}$ current</td>
<td>35–38</td>
</tr>
<tr>
<td>KCN10</td>
<td>Potassium inwardly rectifying channel, subfamily J, member 2</td>
<td>$I_{Na}$ current</td>
<td>39</td>
</tr>
<tr>
<td>KCN15</td>
<td>Potassium inwardly rectifying channel, subfamily J, member 5</td>
<td>$I_{Na}$ current</td>
<td>46, 47</td>
</tr>
<tr>
<td>KCNE2</td>
<td>Potassium voltage-gated channel, subfamily J, member 2</td>
<td>$I_{Na}$ current</td>
<td>48</td>
</tr>
<tr>
<td>KCNE3</td>
<td>Potassium voltage-gated channel, subfamily J, member 3</td>
<td>$I_{Na}$ current</td>
<td>49</td>
</tr>
<tr>
<td>KCNE5</td>
<td>Potassium voltage-gated channel, subfamily J, member 1</td>
<td>$I_{Na}$ current</td>
<td>8–14</td>
</tr>
<tr>
<td>LMNA</td>
<td>Lamin A/B</td>
<td>Nuclear envelope structure</td>
<td>50, 51</td>
</tr>
<tr>
<td>NKX2.5</td>
<td>Homeobox protein Nkx2.5</td>
<td>Cardiac development</td>
<td>52</td>
</tr>
<tr>
<td>NPPA</td>
<td>Natriuretic peptide precursor A</td>
<td>Systemic sodium homeostasis</td>
<td>53, 54</td>
</tr>
<tr>
<td>NUP155</td>
<td>Nucleoporin 155</td>
<td>Nuclear pore formation</td>
<td>55</td>
</tr>
<tr>
<td>PITX2c</td>
<td>Paired-like homeodomain 2c</td>
<td>Great vein development, left–right asymmetry</td>
<td>56</td>
</tr>
<tr>
<td>RYR2</td>
<td>Ryanodine receptor 2</td>
<td>Ca$^{2+}$ release from sarcoplasmic reticulum</td>
<td>57</td>
</tr>
<tr>
<td>SCN1B</td>
<td>Sodium channel, voltage-gated, type I, β subunit</td>
<td>$I_{Na}$ current modulation</td>
<td>58, 59</td>
</tr>
<tr>
<td>SCN2B</td>
<td>Sodium channel, voltage-gated, type II, β subunit</td>
<td>$I_{Na}$ current modulation</td>
<td>59</td>
</tr>
<tr>
<td>SCN3B</td>
<td>Sodium channel, voltage-gated, type III, β subunit</td>
<td>$I_{Na}$ current modulation</td>
<td>60, 61</td>
</tr>
<tr>
<td>SCN4B</td>
<td>Sodium channel, voltage-gated, type IV, β subunit</td>
<td>$I_{Na}$ current modulation</td>
<td>62</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Sodium channel, voltage-gated, type V, α subunit</td>
<td>$I_{Na}$ current modulation</td>
<td>63–68</td>
</tr>
</tbody>
</table>

Evaluation of Pitx2 knockout mice has also been informative for potential mechanisms, whereby misregulation of Pitx2 could contribute to AF. Specifically, whereas homozygous knockout is embryonic lethal, haploinsufficiency of the predominant cardiac isoform, Pitx2c, results in a shortened atrial action potential and an increased susceptibility to AF after burst pacing. The same study also identified continued expression of Pitx2c in the left atrial myocardium, but whether altered adult expression in the myocardium contributes to the causation of AF in the absence of developmental differences is unclear. Atrial-specific conditional knockout of Pitx2c also results in perturbation of the action potential and resting membrane potential. Furthermore, deletion of Pitx2c expression results in diminished expression of cardiac sodium and potassium channels.

After this initial study, the need for greater statistical power was recognized and led to the formation of the CHARGE-AF (Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium) or AFGen Consortium. In 2009, 2 groups independently identified a second locus for AF at 16q22 in Europeans and Han Chinese. These results were later replicated in individuals of black descent. The AF risk SNP at this locus is intronic to the gene ZFHX3, alternatively known as ATBF1, that encodes a zinc finger homeobox transcription factor. ZFHX3 expression has been identified as a factor in the terminal differentiation of both neuronal and striated muscle tissues and also reported as a putative tumor suppressor gene. Given these roles in other tissues, and its apparent expression within cardiac tissues, a developmental role in the atria is possible. However, the lack of availability of model systems with altered ZFHX3 expression has limited the understanding of its potential role in AF. The development of these resources will undoubtedly aid in the discovery of the potential mechanisms whereby this gene, and this susceptibility locus, may be related to AF.

In a separate GWAS from the CHARGE-AF Consortium, patients with early onset AF were used in hopes of minimizing any sample heterogeneity that may have been seen in previous analyses. In a meta-analysis of 5 GWAS studies with early onset AF, a region intronic to the KCNN3 gene was identified. Similar to the majority of targets identified in candidate gene studies in familial AF, KCNN3 encodes a
potassium channel responsible for membrane repolarization. The encoded protein, the SK3 channel, is a calcium-activated, small conductance potassium channel that has largely been studied for its role in neuronal electrophysiology. In neurons, SK3 acts in late repolarization to reduce excitability of neurons after repeated stimulation, a phenomenon termed afterhyperpolarization.111 The role of the KCNN3 in the heart is much less clear, but some evidence exists for a role of SK family members in AF pathogenesis. Among these, studies about the deletion of the SK2 channel in mice found a prolongation in cardiac action potentials and increased susceptibility to AF.112 Furthermore, blockade of the SK family-mediated I_{\text{KCa}} current also confers an increased risk of atrial arrhythmias in rodents113 and canines.114 Finally, recent reports using a mouse model of altered SK3 expression demonstrated alterations in atrial myocyte repolarization115 and an increased incidence of inducible atrial arrhythmias.116 Together, these data suggest a mechanism whereby altered expression of SK3 may have important implications on the electric stability of the atrium.

Meta-Analysis Identification of Novel AF Loci

In 2010, the AFGen Consortium published a meta-analysis of GWAS data from 16 different studies in which 6 novel AF loci were identified in individuals of European and Japanese descent (Table 2).91 The following section will detail the identified loci and possible mechanisms of how they might contribute to AF.

Genetic variants at the 1q24 locus, 46-kb upstream of the PRRX1 gene, were associated with a modest, 14% increased risk of AF (P=8.4×10\(^{-14}\)). PRRX1 encodes a member of the paired-related homeobox gene family; transcription factors that broadly contribute to the differentiation and developmental patterning. In humans, mutations in PRRX1 lead to agnathia-oticcephaly,117,118 a generally fatal condition characterized by severely altered craniofacial development. In rodent models, homozygous deletion of PRRX1 results in early postnatal death and abnormal development of craniofacial, limb, and vertebral structures.119 In addition, PRRX1 is highly expressed in the developing great vessels and is essential to the proper formation of the pulmonary vein.120,121 As discussed for PITX2, ectopic depolarizations within the pulmonary venous regions are often responsible for the initiation of AF. It remains unclear whether PRRX1 regulatory variation is related to congenital alterations in the pulmonary vein structure or function during development or is instead associated with altered activity later in life.

Another association signal was localized intronic to HCN4 on 15q24. HCN4 is highly expressed in both sinoatrial and atrioventricular nodes and is responsible for the funny current (I_f) that controls cardiac pacemaking. Interestingly, mutations in HCN4 have been found in individuals and families with sick sinus syndrome,122-124 tachy-brady syndrome, and AF.33 Whether the risk locus for AF alters overall HCN4 expression levels to a sufficient extent to confer a risk for AF or whether this region results in critical expression differences in a tissue-specific manner remains to be determined.

A novel locus was also located on 7q31 intronic to the CAV1 gene that encodes Caveolin-1, a protein essential for the formation and maintenance of caveolae. The caveolae are regions of the membrane with unique phospholipid composition, which act as mediators of clathrin-independent endocytosis and as scaffolds for cellular, particularly integrin-mediated, signaling. In addition to these roles, caveolae also harbor many ion channels,125 including those responsible for all phases of the cardiac action potential. Studies of cardiovascular function in CAV1-null mice reported aberrant calcium signaling and an alteration in myogenic tone.126 Dilated cardiomyopathy, right ventricular hypertrophy, and pulmonary hypertension have also been observed.127 Further evaluation of the risk locus may aid in the determination of the tissue-localized effect of CAV1 that leads to an increased risk of AF.

SNPs significantly associated with AF were identified on chromosome 14q23 intronic to SYNE2. The SYNE2 gene encodes Nesprin2, a KASH protein family member that localizes to the nuclear outer membrane. Through its binding with the cytoskeleton, Nesprins are thought to provide a stable nuclear localization in the cell128,129 and also are crucial for microtubule-mediated migration of the nucleus during differentiation.130 Missense mutations in SYNE2 have been reported to cause familial Emery-Dreifuss muscular dystrophy,131 a disease that is also characterized by a spectrum of arrhythmic disorders, including AF.

On chromosome 9q22, an association signal with AF was identified within the gene C9ORF3 (rs10821415; OR, 1.11; P=4.2×10^{-11}). However, this region is relatively gene rich, with 3 additional genes and 3 identified non-coding RNAs.

Table 2. Genome-Wide Association Studies–Derived Risk Loci for Atrial Fibrillation

<table>
<thead>
<tr>
<th>Locus</th>
<th>Sentinel SNP</th>
<th>RR</th>
<th>P Value</th>
<th>Nearest Gene Symbol</th>
<th>Relative Location</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>4q25</td>
<td>rs8187105</td>
<td>1.64</td>
<td>1.8×10^{-4}</td>
<td>PITX2</td>
<td>150-kb upstream</td>
<td>83–87, 91, 92, 103, 104, 110</td>
</tr>
<tr>
<td>16q22</td>
<td>rs2106261</td>
<td>1.24</td>
<td>3.2×10^{-6}</td>
<td>ZFHX3</td>
<td>Intronic</td>
<td>87, 91, 104</td>
</tr>
<tr>
<td>1q21</td>
<td>rs6666258</td>
<td>1.18</td>
<td>2.0×10^{-4}</td>
<td>KCN3</td>
<td>Intronic</td>
<td>87, 91, 110</td>
</tr>
<tr>
<td>1q24</td>
<td>rs3903239</td>
<td>1.14</td>
<td>8.4×10^{-4}</td>
<td>PRRX1</td>
<td>46-kb upstream</td>
<td>91</td>
</tr>
<tr>
<td>7q31</td>
<td>rs3807989</td>
<td>0.90</td>
<td>3.6×10^{-7}</td>
<td>CAV1</td>
<td>Intronic</td>
<td>91</td>
</tr>
<tr>
<td>14q23</td>
<td>rs1152591</td>
<td>1.13</td>
<td>5.8×10^{-3}</td>
<td>SYNE2</td>
<td>Intronic</td>
<td>91</td>
</tr>
<tr>
<td>9q22</td>
<td>rs10821415</td>
<td>1.11</td>
<td>4.2×10^{-3}</td>
<td>C9orf3</td>
<td>Intronic</td>
<td>91</td>
</tr>
<tr>
<td>15q24</td>
<td>rs7164883</td>
<td>1.19</td>
<td>2.8×10^{-7}</td>
<td>HCN4</td>
<td>Intronic</td>
<td>91</td>
</tr>
<tr>
<td>10q22</td>
<td>rs10824026</td>
<td>0.87</td>
<td>4.0×10^{-4}</td>
<td>MYOZ1</td>
<td>20-kb upstream</td>
<td>91</td>
</tr>
</tbody>
</table>

RR indicates relative risk; and SNP, single nucleotide polymorphism.
within 300 kb of the sentinel SNP. Because none of the genes in the region have an obvious relationship with cardiovascular function or development, further investigation of this locus will be necessary.

Genetic variants associated with AF were also localized to an intergenic region between 2 genes known to play crucial roles in striated muscle physiology, SYNPO2L and MYOZ1 (10q22; rs10824026; OR, 0.87; \( P = 4.0 \times 10^{-9} \)). This SYNPO2L/MYOZ1 locus illustrates the use of expression quantitative trait loci (eQTL) data to identify a disease-associated gene. Many intronic and intergenic SNPs identified by GWAS are thought to mediate their effects by regulating the transcription of a gene in the region. Sometimes there can be many genes at a locus so it can be difficult to know which is related to disease. Therefore, in eQTL mapping, one examines the relationship between an SNP genotype and transcript levels of all genes at the locus, ideally from a relevant tissue. If a disease-related SNP is associated with transcriptional differences in a gene, this cis-eQTL association provides compelling support for the role of this gene in disease.

Initially, an SNP in linkage disequilibrium with the sentinel SNP at SYNPO2L/MYOZ1 locus was found to correlate with alterations in the expression of both genes. However, these data were derived from lymphoblastoid cell lines, a tissue type unlikely to reflect the transcriptional alterations associated with AF truly. Recently, an eQTL analysis from left atrial tissue found that the AF SNP was associated with transcriptional differences in MYOZ1 expression alone. Therefore, it is likely that MYOZ1 is the AF-related gene at this locus. The encoded protein, myozenin 1, is a cardiac-enriched, z-disk localized protein that aids in the binding of \( \alpha \)-actinin and \( \gamma \)-filamin to confer stable sarcomeric organization. Although no known disease-causing variants of MYOZ1 have been identified, mutations in MYOZ2 result in familial hypertrophic cardiomyopathy, and replication of these mutations or ablation of expression in a murine model resulted in hypertrophic program activation and disruption of z-disk structure in ventricular myocytes.

### Integrating GWAS Data to Stratify AF Risk

In summary, GWAS have identified 9 genetic loci associated with AF (Table 2). Although the ORs for any given region are modest, the potential risk in a given individual may be much higher when the AF SNPs are considered together. Ultimately, using these combined data would be important in a clinical setting, where risk could be stratified based on a combination of genetic and clinical risk factors. Along these lines, Everett et al derived a clinical risk score for AF in women without previous cardiovascular disease. The addition of a genetic risk score, consisting of the top 9 GWAS variants, improved AF risk prediction, but it did alter the reclassification into 10-year risk categories.

Interestingly, a recent large-scale conditional analysis in 17 cohorts from the AFGen Consortium found that there are differences in MYOZ1 expression alone. Therefore, it is likely that MYOZ1 is the AF-related gene at this locus. The encoded protein, myozenin 1, is a cardiac-enriched, z-disk localized protein that aids in the binding of \( \alpha \)-actinin and \( \gamma \)-filamin to confer stable sarcomeric organization. Although no known disease-causing variants of MYOZ1 have been identified, mutations in MYOZ2 result in familial hypertrophic cardiomyopathy, and replication of these mutations or ablation of expression in a murine model resulted in hypertrophic program activation and disruption of z-disk structure in ventricular myocytes.

### Integrating GWAS Data to Stratify AF Risk

In summary, GWAS have identified 9 genetic loci associated with AF (Table 2). Although the ORs for any given region are modest, the potential risk in a given individual may be much higher when the AF SNPs are considered together. Ultimately, using these combined data would be important in a clinical setting, where risk could be stratified based on a combination of genetic and clinical risk factors. Along these lines, Everett et al derived a clinical risk score for AF in women without previous cardiovascular disease. The addition of a genetic risk score, consisting of the top 9 GWAS variants, improved AF risk prediction, but it did alter the reclassification into 10-year risk categories.

Interestingly, a recent large-scale conditional analysis in 17 cohorts from the AFGen Consortium found that there are differences in MYOZ1 expression alone. Therefore, it is likely that MYOZ1 is the AF-related gene at this locus. The encoded protein, myozenin 1, is a cardiac-enriched, z-disk localized protein that aids in the binding of \( \alpha \)-actinin and \( \gamma \)-filamin to confer stable sarcomeric organization. Although no known disease-causing variants of MYOZ1 have been identified, mutations in MYOZ2 result in familial hypertrophic cardiomyopathy, and replication of these mutations or ablation of expression in a murine model resulted in hypertrophic program activation and disruption of z-disk structure in ventricular myocytes.
≥4 different risk alleles at the 4q25/PITX2 locus for AF. Consideration of these PITX2 SNPs plus the other 8 GWAS SNPs resulted in a ≈5-fold gradient in the risk of AF among individuals of European descent (Figure 2, European). The application of these same SNPs to a large Japanese population provided similar results (Figure 2, Japanese). As discussed below, with such a marked variation in AF risk in the population, it will be possible to identify individuals with both a marked increased and decreased risk for AF. Such genetic stratification of AF risk may ultimately enable an improved assessment of different treatment approaches or outcomes based on one’s risk.

**Future Directions for the Genetics of AF**

Great strides have been made in determining the genetic risk for the development of AF; however, many challenges remain. In the following section, we outline a series of selected potential future directions for genetics studies of AF (Figure 3).

**Identification of Additional AF Genetic Loci**

A qualitative viewing of the Manhattan plot from the latest publication by the AFGen Consortium reveals several association signals that rise well above the milieu of background noise but do not exceed a genome-wide significance threshold (Figure 3). A logical extension of this work would then be to determine if these subthreshold loci are additional potential AF genetic risk loci. Genotyping AF-associated SNPs from these subthreshold loci in a larger number of patient samples would likely lead a strengthening of an association signal for some loci. Although genotyping these SNPs in additional cases is straightforward, the subthreshold loci that are found are likely to contribute to an ever decreasing fraction of AF risk. Thus, although newly identified loci are unlikely to have a large effect on clinical risk prediction, they could still be helpful to identifying more members of the molecular pathways that underlie AF.

Future work should also focus on the identification of AF risk variants in different races and ethnicities. To date, the majority of discovery has been performed in populations of European descent, with limited work being done in individuals of Asian and black descent. Because AF prevalence varies greatly among races, it remains unclear whether the results from the studies of Europeans translate to other races or whether a different combination of risk loci is instead present. Studies in...
other races and ethnicities will be particularly important for the future application of genetic data to clinical care.

The Challenge of Causal Variant Identification at GWAS Loci
There are currently 9 identified genetic loci that are significantly associated with AF. However, despite the publication of the 4q25/PITX2 locus 26 years ago, the causative variants at all of the AF loci remain unknown. One challenge is the sheer size of these genomic regions because the PITX2 locus alone comprises a region of ~150 kb. Another challenge is that the top SNP identified by GWAS is rarely the causative variant, rather it is usually serving as a surrogate for a nearby causal variant. A final challenge is that the genetic mechanism for the association with AF is also unknown. We typically assume that AF risk is mediated by an SNP, but it is also possible that the association with AF could be because of a noncoding insertion or deletion, a genetic rearrangement, a variation in copy number, or an epigenetic modification.

To address these challenges, a combination of techniques will be required (Figure 3). One approach could be to refine the genetic signal by fine mapping or increasing the density of SNPs within a target locus. This could be done directly genotyping more SNPs at a locus in a large population of cases and controls. Such an approach was used to identify multiple susceptibility signals at the 4q25. However, with the coverage of current genotyping platforms used for GWAS that consist of 1- to 5-million SNPs and the increasing resolution provided by the 1000 Genomes project, additional genotyping may have a limited incremental benefit.

Because the turnaround time from submission to results-in-hand is now measured in weeks to months and the cost continues to drop, a viable complementary approach would be to sequence an entire disease locus. Importantly, sequencing would provide nucleotide-level resolution of the genetic architecture within AF risk loci. Thus, it would be expected that sequencing of AF risk loci in a large number of cases and controls will aid the identification of the causative haplotypes and variants associated with AF. Because the genetic variants identified by GWAS are markers of an association rather than a causative variant, one would anticipate that a causative SNP identified by sequencing would have a greater effect size and significance than the original GWAS signal. Sequencing a locus could also identify insertions/deletions or copy number variants that associate with disease and may be poorly described in current public databases. Finally, it is possible that sequencing could reveal multiple causative variants within a given locus, something that may not be identifiable by fine mapping. Although such large-scale sequencing is currently feasible, the overall benefits remain unclear particularly given the significant cost of such projects.

In addition to refinement of the loci with fine mapping and sequencing, it will be essential to integrate the vast amount of emerging regulatory data. Although coding variation is a possibility at some loci, a large majority of GWAS loci reside in intergenic or noncoding areas of the genome. This observation led to the assumption that these associations may be because of alterations of regulatory elements, such as enhancers or promoters, which, in turn, alter the activity of distant genes. Indeed some GWAS loci have been shown to alter transcription factor–binding sites that, in turn, lead to differential expression of an adjacent gene. For this purpose, in silico analyses of data provided by the ENCODE (Encyclopedia of DNA Elements) project can be incredibly useful for determining the causal mechanism of variation at a genetic locus. Genomic regions with high mammalian conservation, increased DNase hypersensitivity, increased H3K27-acetylation, and identification of transcription factor–binding sites through chromatin immunoprecipitation sequencing can prove to be useful in identifying altered functional elements within a risk locus.

Although both sequencing and in silico analyses can provide a higher resolution map of a genetic locus, there may still be many candidate regulatory regions across the locus. Studies that can identify the functional role of a regulatory region will be a critical next step. For example, one could postulate that, at the 4q25/PITX2 locus, sequencing would allow the identification of the critical haplotypes that are associated with an AF. An in silico analysis would then identify several highly conserved regions with enhancer activity. One could then examine these potential enhancers for activity in a model system, such as mice, zebrafish, or in an atrial or cardiomyocyte cell line. The causative genetic variant would then likely be one that is both significantly associated with disease and results in an alteration in enhancer activity.

Although methods currently exist for each of the steps outlined above, sequencing, in silico analyses, and functional follow-up are expensive, slow, and challenging. The limited number of causative variants that have been identified at GWAS loci is not a problem specific to AF. Indeed, thousands of GWAS loci have been described, but causative variants have only been identified at a handful. Ultimately, a large-scale effort to identify causative variants at GWAS loci will be necessary to overcome the obstacles faced by any single laboratory systematically.

Atrial and Pulmonary Vein–Specific eQTL Maps
As detailed above for the MYOZI locus for AF, eQTL maps, which examine the changes in tissue-specific expression of nearby genes when a given SNP genotype is present, can provide a useful link between GWAS loci and potential gene targets. Such analyses of gene expression have been useful in studies of atrial identity and other cardiovascular traits. Although these eQTL associations at genetic loci can be helpful if they are present, the tissue specificity of an eQTL signal is critical. Current publically available data sets, such as the eQTL browser or the Genotype-Tissue Expression repository, have a limited tissue composition that reflects the challenge in obtaining relevant human tissue samples, but they are quickly expanding. For AF, it would be ideal to have eQTL data from much more specific tissue sources that are more plausibly involved in the pathogenesis of the arrhythmia. One would expect that the generation of publically available left atrial, pulmonary venous, or AV nodal eQTL data sets would greatly aid in the discovery of the mechanism of causal variation in AF.

Exome Chip Will Enable Large-Scale Assessment of Rare Coding Variation in AF
The evaluation of GWAS loci discussed above was focused on noncoding regions, but it is important to realize that many loci
are in linkage disequilibrium, and thus effectively overlap, with the coding region of 1 or more genes. In these cases, it is possible that the GWAS SNP is a marker or proxy for a coding SNP that actually underlies the association signal. SNPs within a gene could have many potential effects—including nonsynonymous variation that directly alters protein function; synonymous variation that affects splicing, affects transcript stability, or influences codon efficiency; or untranslated region variation that affects translational efficiency or interactions with noncoding regulatory RNAs.

Once a locus is identified that overlaps with a gene, one could genotype every SNP within the gene in a large number of cases and controls to see whether it has a stronger association with AF than that identified by GWAS. Although straightforward in concept, in practice, GWAS loci are large, they may contain many genes each of which can have many rare and common variants and the cost of genotyping remains relatively expensive. One solution to address this issue has been the development of an exome genotyping array or exome chip.

In a GWAS genotyping array, SNPs are captured throughout the entire genome, whereas in an exome array, the focus is largely on coding SNPs. Current exome arrays include more than a quarter of a million SNPs that essentially capture almost all of the common and rare coding variants for every gene in the genome. Within the past year, hundreds of thousands of individuals have been genotyped with these arrays. Much like a GWAS analysis, by comparing a large number of cases and controls, one can quickly identify any coding changes associated with AF. The exome chip analysis can be considered with the GWAS results to identify the coding variants simultaneously within all of the known AF loci.

Exome genotyping arrays will be incredibly powerful at systematically identifying any known coding variation for AF, and we can expect to see the initial results of these studies within the next year. However, because these arrays are only genotyping known SNPs, they would not be useful for studying sporadic or novel genetic variation in an individual or family. Detection of such variation would require direct sequencing of individual genes, exomes, or genomes.

**Candidate Gene Screening Will Be Replaced by Exome and Genome Sequencing**

As described earlier, many mutations described for AF have been identified using a candidate gene approach. In brief, the coding region of a gene is sequenced in AF cases, a unique variant is identified, and that variant is then shown to alter the function of a protein. Although such studies are straightforward, they are limited by (1) the time and cost restraints of sequencing that restrict the analysis to a small number of genes, (2) the inability to detect polygenic causative variation, (3) the focus on coding variation, and perhaps most importantly, (4) the limited likelihood that a particular candidate gene or variant within a gene is pathologically related to AF.

In the upcoming years, the continually decreasing cost and improving quality of next-generation sequencing will enable the widespread adoption of sequencing the exome or protein-coding region of the genome. We can expect that exome sequencing of cohorts of individuals with early onset AF will provide a more comprehensive initial approach for relating rare genetic variation to AF; however, several challenges remain. For every individual sequenced, one can expect to find hundreds of unique nonsynonymous SNPs or insertions/deletions that have never been described in publicly available resources, such as the Exome Variant Server. Thus, determining which variants are truly related to AF and which are genetic noise can be difficult. The identification of multiple hits in a given gene or genetic pathway across individuals can provide compelling evidence for the role of the gene or pathway in AF, yet large, well-powered studies will be required to make definitive conclusions. Improvements in the yield of such efforts may come from sequencing extremes of a phenotype, such as cases of early onset AF.

Because costs continue to decrease further, genome sequencing will also become more realistic in cohort studies, yet with an even greater number of variants identified, assigning causality to noncoding variants will prove even more difficult. Given the continued challenges with large-scale sequencing approaches, an important step forward would be the creation of a centralized repository of exome and genome sequencing-derived variants identified in patients with AF. Comparison of variation in larger data sets on the scale of thousands rather than tens or hundreds of patients will aid in determining variants and genes that are truly causative for the arrhythmia.

**Families Can Provide a Unique Window Into the Mechanisms of AF**

Although much of our discussion has focused on using genetics to identify risk markers for AF in populations, familial forms of AF remain an important investigational tool. Although families with autosomal dominant AF are rare, even a single family can shed light on the underlying molecular mechanisms for AF. To date, convincing evidence from families has identified the role of KCNQ1 and ANP in AF. Challenges with using families to identify AF genes include the rarity of the families, the limited number of individuals with AF, and the difficulty in ensuring that all family members have a common genetic basis for the disease. The last point is particularly pertinent given that the background prevalence of AF can be as high as 10%. Traditional linkage analysis has become increasingly easy to perform using SNP chips to genotype family members at a high density. Furthermore, exome and genome sequencing can be quickly performed in affected family members. Although with exome sequencing one will still identify hundreds of variants in each individual, the familial transmission of disease enables a focus on those variants shared among all affected family members. A combination of using linkage analysis to identify a genetic locus and exome or genome sequencing can further narrow the search for an underlying mutation. Ultimately, once identified, functional evaluation of a mutation on protein function will be necessary to provide convincing evidence of the role of gene in the pathogenesis of AF.

Given that only a handful of causative mutations have been identified in families with AF, the current Heart Rhythm Society/European Heart Rhythm Association consensus guideline states that there is no clinical use for screening known AF-associated genes in patients with AF.142 This includes the use
of any currently available commercial testing panels for AF genes and risk loci. Because these gene panels are systematically tested in larger cohorts of individuals with familial AF, future evidence may emerge on the use of commercial testing.

**Other Forms of Genetic Variation**

In addition to analyses of common and rare genetic variants described above, there are multiple other potential genetic analyses that could be considered to identify more of the heritability of AF. Variations in copy number have not been systematically examined in patients with AF. High-resolution detection of deletions, insertions, and duplications either in coding or noncoding regions has become increasingly straightforward using array-based or next-generation sequencing methods. The major barriers to analyses of copy number variation at present are largely centered on cost and sample size necessary to ensure adequate statistical power.

The detection of epigenetic DNA methylation patterns in a tissue is also a straightforward technique; however, because DNA methylation is a highly tissue-specific process, multiple challenges exist with respect to AF. Ideally one would want to analyze left atrial or pulmonary venous tissue from both patients with and without lone AF, yet it is not practical to obtain these samples. Rather, most samples are obtained at the time of cardiac surgery for coronary disease, valvular heart disease, or transplant, and as such the analyses are limited by the inherent comorbidities present with each type of patient population.

Given that AF increases in prevalence with age, it is possible that somatic or acquired mutations underlie some portion of the heritability of AF. In an intriguing article, Gollob et al\(^8\) found somatic mutations in GJA5 among patients with lone AF. Presently, one could identify total somatic variation by whole-genome or whole-exome sequencing rather than on a tissue level. With each successive generation, there is a background rate of spontaneous genetic variation that occurs. By performing exome or genome sequencing in an affected child and unaffected parents, it is possible to identify the handful of novel coding variants present in the child but not in the parents, who may be associated with a disease. Recently, such an approach has identified a novel pathway for autism spectrum disorders.\(^{143}\)

**Integration of Genetic Data to Predict AF and Outcomes**

One ultimate goal of research into the genetic basis of AF is the potential return of this data to clinical arena. It is hoped that with the current trajectory of novel findings and the integration of the additional studies outlined above, that SNP data could be clinically useful in the near future. However, it is important to note that, in addition to not recommending the testing of known AF genes, the current Heart Rhythm Society/European Heart Rhythm Association guidelines recommend against the testing of individual GWAS-associated SNPs in patients with AF. This decision was likely based on the small number of AF SNPs that had been identified at that time and the limited data on the clinical use of these variants.

Since the publication of these guidelines, there have been many studies examining the relationship between AF SNPs and treatment outcomes. Specifically, the risk of AF recurrence after cardioversion,\(^{144}\) pulmonary vein isolation,\(^{145,146}\) or the initiation of antiarrhythmic medication\(^{147}\) has been studied; however, the observed sample and effect sizes have not allowed the applicability of these results to the broader population. More compelling results have been seen in patients with stroke. Interestingly, the top 2 genetic variants identified in a large GWAS for cardioembolic stroke are also the top 2 regions (\textit{PITX2} and \textit{ZFHX3}) associated with AF.\(^{86,88,89,104}\)

During the last 5 years, it has become clear that clinical risk factors,\(^{148}\) biomarkers,\(^{149}\) and now genetic variants can all help to identify individuals at risk for AF. Rather than using any single one of these approaches alone, we should seek to combine each of these risk factors to enhance the detection of AF. One could imagine that in high-risk populations, such as cryptogenic stroke, we will be able to stratify patients into varying degrees of AF risk and, in turn, to consider alternative strategies to AF monitoring or anticoagulation.

**Conclusions**

Recent studies have identified several rare and common genetic variants associated with AF. However, the present data only account for a limited percentage of the heritability of AF. Integration of next-generation sequencing technologies, improved gene expression data repositories, the identification of additional AF risk loci, and a more complete understanding of causative mechanisms behind AF risk loci will be required. Ultimately, with a more complete picture of the genetic risk for AF, we can seek to develop genetically driven clinical interventions and treatment strategies.

**Acknowledgments**

We thank Julie Herndon for the creation of original artwork used in accompanying figures.

**Sources of Funding**

Dr Tucker was supported by a T32 training grant from the National Heart, Lung, and Blood Institute (T32HL007208). Dr Ellinor was supported by grants from the National Institutes of Health (R01HL092577, R01HL104156, and 1K24HL105780) and the American Heart Association (13EIA14220013).

**Disclosures**

None.

**References**

A novel GATA5 loss-of-function mutation underlies lone atrial fibrillation.


Cardiac sodium channel (SCN5A) variants are associated with atrial fibrillation. Circulation. 2010;41:1119–1126.


Emerging Directions in the Genetics of Atrial Fibrillation
Nathan R. Tucker and Patrick T. Ellinor

_Circ Res._ 2014;114:1469-1482
doi: 10.1161/CIRCRESAHA.114.302225

_Circulation Research_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/114/9/1469

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation Research_ 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 Research_ is online at:
http://circres.ahajournals.org/subscriptions/