Published in final edited form as:
Circ Cardiovasc Genet. 2013 August ; 6(4): 362–371. doi:10.1161/CIRCGENETICS.113.000133.

A Weighted Gene Co-Expression Network Analysis of Human Left Atrial Tissue Identifies Gene Modules Associated with Atrial Fibrillation

Nicholas Tan, BS[1], **Mina K. Chung, MD**[2], **Jonathan D. Smith, PhD**[3], **Jeffrey Hsu, BS**[3], **David Serre, PhD**[4], **David W. Newton, MD**[2], **Laurie Castel, BS**[5], **Edward Soltesz, MD, MPH**[6], **Gosta Pettersson, MD, PhD**[6], **A. Marc Gillinov, MD**[6], **David R. Van Wagoner, PhD**[5], and **John Barnard, PhD**[7]

[1]Cleveland Clinic Lerner College of Medicine, Cleveland Clinic Lerner Research Institute, Cleveland, OH

[2]Dept of Cardiovascular Medicine, Cleveland Clinic Lerner Research Institute, Cleveland, OH

[6]Dept of Thoracic & Cardiovascular Surgery, Cleveland Clinic Foundation; Cleveland Clinic Lerner Research Institute, Cleveland, OH

[3]Dept of Cellular & Molecular Medicine, Cleveland Clinic Lerner Research Institute, Cleveland, OH

[4]Genomic Medicine Institute, Cleveland Clinic Lerner Research Institute, Cleveland, OH

[5]Dept of Molecular Cardiology, Cleveland Clinic Lerner Research Institute, Cleveland, OH

[7]Dept of Quantitative Health Sciences, Cleveland Clinic Lerner Research Institute, Cleveland, OH

Abstract

Background—The genetic mechanisms of atrial fibrillation (AF) remain incompletely understood. Previous differential expression studies in AF were limited by small sample size and provided limited understanding of global gene networks, prompting the need for larger-scale, network-based analyses.

Methods and Results—Left atrial tissues from Cleveland Clinic cardiac surgery patients were assayed using Illumina Human HT-12 mRNA microarrays. The dataset included three groups based on cardiovascular co-morbidities: mitral valve (MV) disease without coronary artery disease (CAD) (n=64); CAD without MV disease (n=57); and lone AF (LAF) (n=35). Weighted gene co-expression network analysis was conducted in the MV group to detect modules of correlated genes. Module preservation was assessed in the other two groups. Module eigengenes were regressed on AF severity or atrial rhythm at surgery. Modules whose eigengenes correlated with either AF phenotype were analyzed for gene content. 14 modules were detected in the MV group; all were preserved in the other two groups. One module (124 genes) was associated with AF severity and atrial rhythm across all groups. Its top hub gene, *RCAN1*, is implicated in calcineurin-

Correspondence to: Nicholas Tan, BS Cleveland Clinic Lerner College of Medicine 9500 Euclid Avenue NA21 Cleveland OH 44195 Tel: 216-445-7170 Fax 216-636-3206 tann@ccf.org.

Conflict of Interest Disclosures: A. Marc Gillinov receives consulting/speaking fees from AtriCure Inc. and Edwards Lifesciences LLC, and royalty payments from Clear Catheter Systems Inc.

dependent signaling and cardiac hypertrophy. Another module (679 genes) was associated with atrial rhythm in the MV and CAD groups. It was enriched with cell signaling genes and contained cardiovascular developmental genes including *TBX5*.

Conclusions—Our network-based approach found two modules strongly associated with AF. Further analysis of these modules may yield insight into AF pathogenesis by providing novel targets for functional studies.

Keywords

atrial fibrillation; genetics, bioinformatics; genetics, microarrays; arrhythmias; gene network; gene co-expression

Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, with a prevalence of about 1-2% in the general population[1, 2]. Although AF may be an isolated condition ("lone AF"), it often occurs concomitantly with other cardiovascular diseases such as coronary artery disease (CAD) and/or valvular heart disease[1]. In addition, stroke risk is increased 5-fold among patients with AF, and ischemic strokes attributed to AF are more likely to be fatal[1]. Current anti-arrhythmic drug therapies are limited in terms of efficacy and safety[1, 3, 4]. Thus, there is a need to develop better risk prediction tools as well as mechanistically targeted therapies for AF. Such developments can only come about through a clearer understanding of its pathogenesis.

Family history is an established risk factor for AF. A Danish Twin Registry study estimated AF heritability at 62%, indicating a significant genetic component [5]. Substantial progress has been made to elucidate this genetic basis. For example, genome-wide association studies (GWAS) have identified several susceptibility loci and candidate genes linked with AF. Initial studies conducted in European populations found three AF-associated genomic loci[6-9]. Of these, the most significant single nucleotide polymorphisms (SNPs) mapped to an intergenic region of chromosome 4q25. The closest gene in this region, *PITX2*, is crucial in left-right asymmetrical development of the heart and thus appears promising as a major player in initiating AF[10, 11]. A large-scale GWAS meta-analysis discovered 6 additional susceptibility loci, implicating genes involved in cardiopulmonary development, ion transport, and cellular structural integrity[12].

Differential expression studies have also provided insight into the pathogenesis of AF. A study by Barth et. al.[13] found that about two-thirds of the genes expressed in the right atrial appendage were downregulated during permanent AF, and that many of these genes were involved in calcium-dependent signaling pathways. In addition, ventricular-predominant genes were upregulated in right atrial appendages of AF subjects[13]. Another study showed that inflammatory and transcription-related gene expression was increased in right atrial appendages of AF subjects versus controls[14]. These results highlight the adaptive responses to AF-induced stress and ischemia taking place within the atria.

Despite these advances, much remains to be discovered regarding the genetic mechanisms of AF. The AF-associated SNPs found thus far only explain a fraction of its heritability[15]; furthermore, the means by which the putative candidate genes cause AF have not been fully established[9, 15, 16]. Additionally, prior differential expression studies in human tissue were limited to the right atrial appendage, had small sample sizes, and provided little understanding of global gene interactions[13, 14]. Weighted gene co-expression network analysis (WGCNA) is a technique to construct gene modules within a network based on correlations in gene expression (i.e. co-expression)[17, 18]. WGCNA has been used to study

genetically complex diseases such as metabolic syndrome[19], schizophrenia[20], and heart failure[21]. Here we obtained mRNA expression profiles from human left atrial appendage tissue and implemented WGCNA to identify gene modules associated with AF phenotypes.

Methods

Subject recruitment

From 2001 to 2008, patients undergoing cardiac surgery at the Cleveland Clinic were prospectively screened and recruited. Informed consent for research use of discarded atrial tissues was obtained from each patient by a study coordinator during the pre-surgical visit. Demographic and clinical data were obtained from the Cardiovascular Surgery Information Registry and by chart review. Use of human atrial tissues was approved by the Institutional Review Board of the Cleveland Clinic.

RNA microarray isolation and profiling

Left atria appendage specimens were dissected during cardiac surgery and stored frozen at −80°C. Total RNA was extracted using the Trizol technique. RNA samples were processed by the Cleveland Clinic Genomics Core. For each sample, 250 ng RNA was reverse transcribed into cRNA and biotin-UTP labeled using the TotalPrep RNA Amplification Kit (Ambion, Austin, TX). cRNA was quantified using a Nanodrop spectrophotometer and cRNA size distribution was assessed on a 1% agarose gel. cRNA was hybridized to Illumina Human HT12 Expression BeadChip arrays (v.3). Arrays were scanned using a BeadArray reader.

Expression data pre-processing

Raw expression data was extracted using the “beadarray” package in R and bead-level data was averaged after log base-2 transformation. Background correction was conducted by fitting a normal-gamma deconvolution model using the “NormalGamma” R package[22]. Quantile normalization and batch effect adjustment with the ComBat method were performed using R[23]. Probes that were not detected (at a $p < 0.05$ threshold) in all samples as well as probes with relatively lower variances (inter-quartile range $<= \log_2(1.2)$) were excluded.

The WGCNA approach requires that genes be represented as singular nodes in such a network. However, a small proportion of the genes in our data have multiple probe mappings. To facilitate the representation of singular genes within the network, one probe must be selected to represent its associated gene. Hence, for genes that mapped to multiple probes, the probe with the highest mean expression level was selected for analysis (which often selects the splice isoform with the highest expression and signal-to-noise ratio), resulting in a total of 6168 genes.

Defining training and test sets

Currently, no large external mRNA microarray data from human left atrial tissues are publicly available. To facilitate internal validation of results, we divided our dataset into three groups based on cardiovascular co-morbidities: mitral valve (MV) disease without CAD (MV group) (n=64); CAD without MV disease (CAD group) (n=57); and lone AF (LAF group) (n=35). Lone AF was defined as the presence of AF without concomitant structural heart disease, according to the guidelines set by the European Society of Cardiology[1]. The MV group, which was the largest and had the most power for detecting significant modules, served as the training set for module derivation, while the other two groups were designated test sets for module reproducibility. To minimize the effect of population stratification, the dataset was limited to Caucasian subjects. Differences in

clinical characteristics among the groups were assessed using Kruskal-Wallis rank-sum tests for continuous variables and Pearson's χ^2 test for categorical variables.

Weight gene co-expression network analysis (WGCNA)

WGCNA is a systems-biology method to identify and characterize gene modules whose members share strong co-expression. We applied previously validated methodology in this analysis[17]. Briefly, pair-wise gene (Pearson's) correlations were calculated using the MV group dataset. A weighted adjacency matrix was then constructed:

$$a_{ij}=|c_{ij}|^{\beta} \quad (1)$$

where c_{ij} = Pearson correlation between gene i and gene j and a_{ij} = adjacency between gene i and gene j. β is a soft-thresholding parameter that provides emphasis on stronger correlations over weaker and less meaningful ones while preserving the continuous nature of gene-gene relationships. β = 3 was selected in this analysis based on the criterion outlined by Zhang et. al. (see Supplemental Material)[17].

Next, the topological overlap-based dissimilarity matrix was computed from the weighted adjacency matrix:

$$t_{ij}=1-\frac{\sum_u (a_{iu}a_{uj})+a_{ij}}{min\{k_i,k_j\}+1-a_{ij}} \quad (2)$$

where $k_i = \sum_{y\neq i} a_{iy}$ represents the total connectivity of gene i with all other genes in the network. The topological overlap, developed by Ravasz et. al.[24], reflects the relative interconnectedness (i.e. shared neighbors) between two genes[17]. Hence, construction of the network dendrogram based on this dissimilarity measure allows for the identification of gene modules whose members share strong interconnectivity patterns. The WGCNA cutreeDynamic R function was used to identify a suitable cut height for module identification via an adaptive cut height selection approach[18]. Gene modules, defined as branches of the network dendrogram, were assigned colors for visualization.

Network preservation analysis

Module preservation between the MV and CAD groups as well as the MV and LAF groups was assessed using network preservation statistics as described in Langfelder et. al.[25]. Module density-based statistics (to assess if genes in each module remain highly connected in the test set) and connectivity-based statistics (to assess if connectivity patterns between genes in the test set remain similar compared to the training set) were considered in this analysis[25]. In each comparison, a Z-statistic representing a weighted summary of module density and connectivity measures was computed for every module ($Z_{summary}$). The $Z_{summary}$ score was used to evaluate module preservation, with values of 8 and above indicating strong preservation, as proposed by Langfelder et. al.[25]. The WGCNA R function networkPreservation was used to implement this analysis[25].

Clinical significance of preserved modules

Principal component analysis (PCA) of the expression data for each gene module was conducted. The first principal component of each module, designated the eigengene, was identified for the three cardiovascular disease groups; this served as a summary expression measure that explained the largest proportion of the variance of the module[26]. Multivariate linear regression was performed with the module eigengenes as the outcome variables and AF severity (no AF; paroxysmal AF; persistent AF; permanent AF) as the predictor of interest (adjusting for age and sex). A similar regression analysis was performed with atrial

rhythm at surgery (no AF history; AF history in sinus rhythm; AF history in AF rhythm) as the predictor of interest. The false discovery rate (FDR) method was used to adjust for multiple comparisons. Modules whose eigengenes associated with AF severity and/or atrial rhythm were identified for further analysis.

In addition, hierarchical clustering of module eigengenes and selected clinical traits (age, sex, hypertension, cholesterol, left atrial size, AF state, and atrial rhythm) was utilized to identify additional module-trait associations. Clusters of eigengenes/traits were detected based on a dissimilarity measure D, as given by:

$$D=1-cor\left(V_i,V_j\right),i\neq j \quad (3)$$

where V = the eigengene or clinical trait.

Enrichment analysis

Gene modules significantly associated with AF severity and/or atrial rhythm were submitted to Ingenuity Pathway Analysis (IPA®) to determine enrichment for functional/disease categories. IPA® is an application of gene set over-representation analysis; for each disease/functional category annotation, a p-value is calculated (using Fisher's exact test) by comparing the number of genes from the module of interest which participate in the said category against the total number of participating genes in the background set[27]. All 6168 genes in the current dataset served as the background set for the enrichment analysis.

Hub gene analysis

"Hub genes" are defined as genes that have high intramodular connectivity[17, 20]:

$$k_i=\sum_{j\neq i}^{n}a_{ij} \qquad i,j\in Module \quad q \quad (4)$$

where k_i = intramodular connectivity of gene i, and a_{ij} = adjacency between genes i and j.

Alternatively, they may also be defined as genes with high module membership[21, 25]:

$$MM_i^{(q)}=cor\left(x_i,E^{(q)}\right) \quad (5)$$

where MM_i = module membership of gene i (in module q), x_i = expression profile of gene i, and $E^{(q)}$ = module eigengene of module q. Both definitions were used to identify the hub genes of modules associated with AF phenotype.

To confirm that the hub genes identified were themselves associated with AF phenotype, the expression data of the top 10 hub genes (by intramodular connectivity) were regressed on atrial rhythm (adjusting for age and sex). In addition, eigengenes of AF-associated modules were regressed on their respective (top 10) hub gene expression profiles and the model R^2 indices were computed.

Membership of AF-associated candidate genes from prior studies

Prior GWAS studies identified multiple AF-associated SNPs[8, 9, 12, 15, 28]. We selected candidate genes closest to or containing these SNPs and identified their module locations as well as their closest within-module partners (absolute Pearson's correlations).

Sensitivity analysis of soft thresholding parameter

To verify that the key results obtained from the above analysis were robust with respect to the chosen soft-thresholding parameter (β=3), we repeated the module identification process using β=5. The eigengenes of the detected modules were computed and regressed on atrial rhythm (adjusting for age and sex). Modules significantly associated with atrial rhythm in at least two dataset groups were compared with the AF phenotype-associated modules from the original analysis.

Results

Subject characteristics

Table 1 describes the clinical characteristics of the cardiac surgery patients in the study. Subjects in the LAF group were generally younger and less likely to be a current smoker ($p = 2.0\times10^{-4}$ and 0.032 respectively). Subjects in the MV group had lower BMIs ($p=2.7\times10^{-6}$) and a larger proportion had paroxysmal AF compared to the other two groups (p=0.033).

Gene co-expression network construction and module identification

14 modules were detected using the MV group dataset (Figure 1), with module sizes ranging from 83 genes to 1512 genes. 38 genes did not share similar co-expression with the other genes in the network and were therefore not included in any of the identified modules.

Network preservation analysis revealed strong preservation of all modules between the training and test sets

All 14 modules showed strong preservation across the CAD and LAF groups in both comparisons, with $Z_{summary}$ scores of above 10 in most modules (Figure 2). No major deviations in the $Z_{summary}$ score distributions for the two comparisons were noted, indicating that modules were preserved to a similar extent across the two groups.

Regression analysis of module eigengene profiles identified two modules associated with AF severity and/or atrial rhythm

Table S4 summarizes the proportion of variance explained by the first 3 principal components for each module. On average, the first principal component (i.e. the eigengene) explained approximately 18% of the total variance of its associated module. For each group, the module eigengenes were extracted and regressed on AF severity (with age and sex as covariates). The salmon module (124 genes) eigengene was strongly associated with AF severity in the MV and CAD groups ($p = 1.7\times10^{-6}$ and 5.2×10^{-4} respectively); this association was less significant in the LAF group ($p = 9.0\times10^{-2}$). Eigengene levels increased with worsening AF severity across all 3 groups, with the greatest step-wise change taking place between the paroxysmal AF and persistent AF categories (Figure 3A). When the module eigengenes were regressed on atrial rhythm, the salmon module eigengene showed significant association in all groups (p-values: MV – 1.1×10^{-14}; CAD – 1.36×10^{-6}; LAF – 2.1×10^{-4}). Eigengene levels were higher in the “AF history in AF rhythm” category (Figure 3B).

The regression analysis also revealed statistically significant associations between the tan module (679 genes) eigengene and atrial rhythm in the MV and CAD groups ($p = 5.8\times10^{-4}$ and 3.4×10^{-2} respectively). Eigengene levels were lower in the “AF history in AF rhythm” category compared to the “AF history in sinus rhythm” category (Figure 4); this trend was also observed in the LAF group, albeit with weaker statistical evidence (p=0.15).

Hierarchical clustering of eigengene profiles with clinical traits

Hierarchical clustering was performed to identify relationships between gene modules and selected clinical traits. The salmon module clustered with AF severity and atrial rhythm; in addition, left atrial size was found in the same cluster, suggesting a possible relationship between salmon module gene expression and atrial remodeling (Figure 5A). Although the tan module was in a separate cluster from the salmon module, it was negatively correlated with both atrial rhythm and AF severity (Figure 5B).

IPA® enrichment analysis of salmon and tan modules

The salmon module was enriched in genes involved in "Cardiovascular Function and Development" (smallest $p = 4.4\times10^{-4}$) and "Organ Morphology" (smallest $p = 4.4\times10^{-4}$). In addition, the top disease categories identified included "Endocrine System Disorders" (smallest $p = 4.4\times10^{-4}$) and "Cardiovascular Disease" (smallest $p = 2.59\times10^{-3}$).

The tan module was enriched in genes involved in "Cell-To-Cell Signaling and Interaction" (smallest $p = 8.9\times10^{-4}$) and "Cell Death and Survival" (smallest $p = 1.5\times10^{-3}$). Enriched disease categories included "Cancer" (smallest $p = 2.2\times10^{-4}$) and "Cardiovascular Disease" (smallest $p = 4.5\times10^{-4}$).

See the Supplemental Material file for the full functional enrichment results.

Hub gene analysis of salmon and tan modules

We identified hub genes in the two modules based on intramodular connectivity and module membership. For the salmon module, the gene *RCAN1* exhibited the highest intramodular connectivity and module membership. The top 10 hub genes (by intramodular connectivity) were significantly associated with atrial rhythm, with FDR-adjusted p-values ranging from 1.5×10^{-5} to 4.2×10^{-12}. These hub genes accounted for 95% of the variation in the salmon module eigengene.

In the tan module, the top hub gene was CPEB3. The top 10 hub genes (by intramodular connectivity) correlated with atrial rhythm as well, though the statistical associations in the lower-ranked hub genes were relatively weaker (FDR-adjusted p-values ranging from 1.1×10^{-1} to 3.4×10^{-4}). These hub genes explained 94% of the total variation in the tan module eigengene.

The names and connectivity measures of the hub genes found in both modules are presented in Table 2. A visualization of the salmon module is shown using the Cytoscape tool (Figure 6). A full list of the genes in the salmon and tan modules is provided in the Supplemental Material.

Membership of AF-associated candidate genes from prior studies

The tan module contained *MYOZ1*, which was identified as a candidate gene from the recent AF meta-analysis. *PITX2* was located in the green module (n=349) and *ZFHX3* was in the turquoise module (n=1512). The locations of other candidate genes (and their closest partners) are reported in the Supplemental Material.

Sensitivity analysis of key results

We repeated the WGCNA module identification approach using a different soft thresholding parameter (β=5). 1 module (n=121) was found to be strongly associated with atrial rhythm at surgery across all 3 dataset groups, while another module (n=244) was associated with atrial rhythm at surgery in the MV and CAD groups. The first module overlapped significantly

with the salmon module in terms of gene membership, while most of the second modules' genes were contained within the tan module. The top hub genes found in the salmon and tan modules remained present and highly connected in the two new modules identified with the different soft-thresholding parameter.

Discussion

To our knowledge, our study is the first implementation of an unbiased, network-based analysis in a large sample of human left atrial appendage gene expression profiles. We found two modules associated with AF severity and/or atrial rhythm in two to three of our cardiovascular co-morbidity groups. Functional analyses revealed significant enrichment of cardiovascular-related categories for both modules. In addition, several of the hub genes identified are implicated in cardiovascular disease and may play a role in AF initiation and progression.

In our study, WGCNA was utilized to construct modules based on gene co-expression, thereby reducing the network's dimensionality to a smaller set of elements[17, 21]. Relating module-wise changes to phenotypic traits allowed statistically significant associations to be detected at a lower FDR compared to traditional differential expression studies. Furthermore, shared functions/pathways among genes in the modules could be inferred via enrichment analyses.

We divided our dataset into 3 groups to verify the reproducibility of the modules identified by WGCNA. 14 modules were identified in the MV group in our gene network. All were strongly preserved in the CAD and LAF groups, suggesting that gene co-expression patterns are robust and reproducible despite differences in cardiovascular co-morbidities.

The use of module eigengene profiles as representative summary measures has been validated in a number of studies[20, 26]. Additionally, we found that the eigengenes accounted for a significant proportion (average 18%) of gene expression variability in their respective modules. Regression analysis of the module eigengenes found two modules associated with AF severity and/or atrial rhythm in at least two dataset groups. The association between the salmon module eigengene and AF severity was statistically weaker in the LAF group (adjusted $p = 9.0\times10\text{-}2$). This was probably due to its significantly smaller sample size compared to the MV and CAD groups. Despite this weaker association, the relationship between the salmon module eigengene and AF severity remained consistent among the three groups (Figure 3A). Similarly, the lack of statistical significance for the association between the tan module eigengene and atrial rhythm at surgery in the LAF group was likely driven by the smaller sample size and (by definition) lack of samples in the "No AF" category.

A major part of our analysis focused on the identification of module hub genes. Hubs are connected with a large number of nodes; disruption of hubs therefore leads to widespread changes within the network. This concept has powerful applications in the study of biology, genetics, and disease[29, 30]. While mutations of "peripheral" genes can certainly lead to disease, gene network changes are more likely to be motivated by changes in hub genes, making them more biologically interesting targets for further study[17, 29, 31]. Indeed, the hub genes of the salmon and tan modules accounted for the vast majority of the variation in their respective module eigengenes, signaling their importance in driving gene module behavior.

The hub genes identified in the salmon and tan modules were significantly associated with AF phenotype overall. It was noted that this association was statistically weaker for the lower-ranked hub genes in the tan module. This highlights an important aspect and strength of WGCNA – to be able to capture module-wide changes with respect to disease despite potentially weaker associations among individual genes.

The implementation of WGCNA necessitated the selection of a soft-thresholding parameter β. Unlike hard-thresholding (where gene correlations below a certain value are shrunk to zero), the soft-thresholding approach gives greater weight to stronger correlations while maintaining the continuous nature of gene-gene relationships. We selected a β value of 3 based on the criteria outlined by Zhang et. al.[17]. His team and other investigators have demonstrated that module identification is robust with respect to the β parameter[17, 19-21]. In our data, we were also able to reproduce the key findings reported with a different, larger β value, thereby verifying the stability of our results with regards to β.

The salmon module (124 genes) was associated with both AF phenotypes; furthermore, IPA® analysis of its gene contents suggested enrichment in cardiovascular development as well as disease. Its eigengene increased with worsening AF severity, with the largest step-wise change occurring between the paroxysmal AF and persistent AF categories (Figure 3). Hence, the gene expression changes within the salmon module may reflect the later stages of AF pathophysiology.

The top hub gene of the salmon module was *RCAN1* (regulator of calcineurin 1) Calcineurin is a cytoplasmic Ca^{2+}/calmodulin-dependent protein phosphatase that stimulates cardiac hypertrophy via its interactions with NFAT and L-type Ca^{2+} channels[32, 33]. *RCAN1* is known to inhibit calcineurin and its associated pathways[32, 34]. However, some data suggests that *RCAN1* may instead function as a calcineurin activator when highly expressed and consequently potentiate hypertrophic signaling[35]. Thus, perturbations in *RCAN1* levels (due to genetic variants or mutations) may cause an aberrant switching in function, which in turn triggers atrial remodeling and arrhythmogenesis.

Other hub genes found in the salmon module are also involved in cardiovascular development/function and may be potential targets for further study. *DNAJA4* (DnaJ homolog, subfamily A, member 4) regulates the trafficking and maturation of *KCNH2* potassium channels, which have a prominent role in cardiac repolarization and which are implicated in the long QT syndromes[36]. *FHL2* (Four-and-a-half LIM domain protein 2) interacts with numerous cellular components including actin cytoskeleton, transcription machinery, and ion channels[37]. *FHL2* was shown to enhance the hypertrophic effects of isoproterenol, indicating that *FHL2* may modulate the effect of environmental stress on cardiomyocyte growth[38]. *FHL2* also interacts with several potassium channels in the heart, such as *KCNQ1*, *KCNE1*, and *KCNA5*[37, 39]. Additionally, *BVES* (blood vessel epicardial substance) and other members of its family were shown to be highly expressed in cardiac pacemaker cells. *BVES* knockout mice exhibited sinus nodal dysfunction, suggesting that *BVES* regulates the development of the cardiac pacemaking and conduction system[40] and may therefore be involved in the early phase of AF development.

The tan module (679 genes) eigengene was negatively correlated with atrial rhythm in the MV and CAD groups (Figure 4); this may indicate a general decrease in gene expression of its members in fibrillating atrial tissue. IPA® analysis revealed enrichment in genes involved in cell signaling as well as apoptosis. The top-ranked hub gene, *CPEB3* (cytoplasmic polyadenylation element binding protein 3), regulates mRNA translation and has been associated with synaptic plasticity and memory formation[41]. The role of *CPEB3* in the heart is currently unknown, so further exploration via animal model studies may be warranted. *NPPB* (Natriuretic peptide-precursor B), another highly interconnected hub gene, produces a precursor peptide of BNP (brain natriuretic peptide), which regulates blood pressure through natriuresis and vasodilation[42]. *NPPB* gene variants have been linked with diabetes mellitus, though associations with cardiac phenotypes are less clear[42]. *TBX5* and *GATA4*, which play important roles in the embryonic heart development[43], were members of the tan module. Although not hub genes, they may also contribute towards developmental susceptibility of

AF. In addition, *TBX5* was previously reported to be near a SNP associated with PR interval and AF in separate large-scale GWAS studies[12, 28]. *MYOZ1*, another candidate gene identified in the recent AF GWAS meta-analysis, was found to be a member as well; it associates with proteins found in the Z-disc of skeletal and cardiac muscle and may suppress calcineurin-dependent hypertrophic signaling[12].

Some but not all of the candidate genes found in prior GWAS studies were located in the AF-associated modules. One possible explanation for this could be the difference in sample sizes. The meta-analysis involved thousands of individuals, while the current study had less than 100 in each dataset group, which limited the power to detect significant differences between levels of AF phenotype even with the module-wise approach. Additionally, transcription factors like *PITX2* are most highly expressed during the fetal phase of development. Perturbations in these genes (due to genetic variants or mutations) may therefore initiate the development of AF at this stage and play no significant role in adults (when we obtained their tissue samples).

We noted several limitations in this study. Firstly, no human left atrial mRNA dataset of adequate size currently exists publicly. Hence, we were unable to validate our results with an external, independent dataset. However, the network preservation assessment conducted within our dataset showed strong preservation in all modules, indicating that our findings are robust and reproducible.

Although the module eigengenes captured a significant proportion of module variance, a large fraction of variability did remain unaccounted for, which may limit their use as representative summary measures.

We extracted RNA from human left atrial appendage tissue, which consists primarily of cardiomyocytes and fibroblasts. Atrial fibrosis is known to occur with AF-associated remodeling[44]. As such, the cardiomyocyte:fibroblast ratio is likely to change with different levels of AF severity, which in turn influences the amount of RNA extracted from each cell type. Hence, true differences in gene expression (and co-expression) within cardiomyocytes may be confounded by changes in cellular composition due to atrial remodeling. Also, there may be significant regional heterogeneity in the left atrium with respect to structure, cellular composition, and gene expression[45], which may limit the generalizability of our results to other parts of the left atrium.

All subjects in the study were Caucasians to minimize the effects of population stratification. However, it is recognized that the genetic basis of AF may differ among ethnic groups[9]. Thus, our results may not be generalizable to other ethnicities.

Finally, it is possible for genes to be involved in multiple processes/functions that require different sets of genes. However, WGCNA does not allow for overlapping modules to be formed. Thus, this limits the method's ability to characterize such gene interactions.

Conclusions

In summary, we constructed a weighted gene co-expression network based on RNA expression data from the largest collection of human left atrial appendage tissue specimens to date. We identified two gene modules significantly associated with AF severity or atrial rhythm at surgery. Hub genes within these modules may be involved in the initiation or progression of AF and may therefore be candidates for functional studies. Future steps include comparing co-expression networks between different ethnicities and the use of other network-based tools (e.g. differential co-expression network analysis) to identify novel AF-associated genes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to members of the AF Genetics lab group for their valuable support. We also like to thank the research coordinators and nurses for their efforts in subject recruitment.

Funding Sources: This research was supported by NIH/ NHLBI grants HL090620 and HL111314 to Drs. Chung, Barnard, Smith, and Van Wagoner. Nicholas Tan receives a stipend from the American Heart Association (1-Year Predoctoral Fellowship Award).

References

1. European Heart Rhythm Association, European Association for Cardio-Thoracic Surgery. Camm AJ, Kirchhof P, Lip GY, Schotten U, et al. Guidelines for the management of atrial fibrillation: The task force for the management of atrial fibrillation of the european society of cardiology (ESC). Eur Heart J. 2010; 31:2369–2429. [PubMed: 20802247]
2. Lemmens R, Hermans S, Nuyens D, Thijs V. Genetics of atrial fibrillation and possible implications for ischemic stroke. Stroke Res Treat. 2011; 2011:208694. [PubMed: 21822468]
3. Wann LS, Curtis AB, January CT, Ellenbogen KA, Lowe JE, Estes NA 3rd, et al. 2011 ACCF/AHA/HRS focused update on the management of patients with atrial fibrillation (updating the 2006 guideline): A report of the american college of cardiology foundation/american heart association task force on practice guidelines. J Am Coll Cardiol. 2011; 57:223–242. [PubMed: 21177058]
4. Dobrev D, Carlsson L, Nattel S. Novel molecular targets for atrial fibrillation therapy. Nat Rev Drug Discov. 2012; 11:275–291. [PubMed: 22460122]
5. Christophersen IE, Ravn LS, Budtz-Joergensen E, Skytthe A, Haunsoe S, Svendsen JH, et al. Familial aggregation of atrial fibrillation: A study in danish twins. Circ Arrhythm Electrophysiol. 2009; 2:378–383. [PubMed: 19808493]
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. [PubMed: 17603472]
7. 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. [PubMed: 20173747]
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. [PubMed: 19597492]
9. Sinner MF, Ellinor PT, Meitinger T, Benjamin EJ, Kaab S. Genome-wide association studies of atrial fibrillation: Past, present, and future. Cardiovasc Res. 2011; 89:701–709. [PubMed: 21245058]
10. Clauss S, Kaab S. Is Pitx2 growing up? Circ Cardiovasc Genet. 2011; 4:105–107. [PubMed: 21505199]
11. Kirchhof P, Kahr PC, Kaese S, Piccini I, Vokshi I, Scheld HH, et al. PITX2c is expressed in the adult left atrium, and reducing Pitx2c expression promotes atrial fibrillation inducibility and complex changes in gene expression. Circ Cardiovasc Genet. 2011; 4:123–133. [PubMed: 21282332]
12. 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. [PubMed: 22544366]
13. Barth AS, Merk S, Arnoldi E, Zwermann L, Kloos P, Gebauer M, et al. Reprogramming of the human atrial transcriptome in permanent atrial fibrillation: Expression of a ventricular-like genomic signature. Circ Res. 2005; 96:1022–1029. [PubMed: 15817885]

14. Ohki R, Yamamoto K, Ueno S, Mano H, Misawa Y, Fuse K, et al. Gene expression profiling of human atrial myocardium with atrial fibrillation by DNA microarray analysis. Int J Cardiol. 2005; 102:233–238. [PubMed: 15982490]
15. Mahida S, Ellinor PT. New advances in the genetic basis of atrial fibrillation. J Cardiovasc Electrophysiol. 2012; 23:1400–1406. [PubMed: 23066792]
16. Dobrev D, Nattel S. New insights into the molecular basis of atrial fibrillation: Mechanistic and therapeutic implications. Cardiovasc Res. 2011; 89:689–691. [PubMed: 21296897]
17. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005:4. Article17.
18. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. [PubMed: 19114008]
19. Min JL, Nicholson G, Halgrimsdottir I, Almstrup K, Petri A, Barrett A, et al. Coexpression network analysis in abdominal and gluteal adipose tissue reveals regulatory genetic loci for metabolic syndrome and related phenotypes. PLoS Genet. 2012; 8:e1002505. [PubMed: 22383892]
20. de Jong S, Boks MP, Fuller TF, Strengman E, Janson E, de Kovel CG, et al. A gene co- expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. PLoS One. 2012; 7:e39498. [PubMed: 22761806]
21. Dewey FE, Perez MV, Wheeler MT, Watt C, Spin J, Langfelder P, et al. Gene coexpression network topology of cardiac development, hypertrophy, and failure. Circ Cardiovasc Genet. 2011; 4:26–35. [PubMed: 21127201]
22. Chen M, Xie Y, Story M. An exponential-gamma convolution model for background correction of illumina BeadArray data. Commun Stat Theory Methods. 2011; 40:3055–3069. [PubMed: 21769162]
23. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012; 28:882–883. [PubMed: 22257669]
24. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. Science. 2002; 297:1551–1555. [PubMed: 12202830]
25. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011; 7:e1001057. [PubMed: 21283776]
26. Langfelder P, Horvath S. Eigengene networks for studying the relationships between coexpression modules. BMC Syst Biol. 2007; 1:54. [PubMed: 18031580]
27. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: Current approaches and outstanding challenges. PLoS Comput Biol. 2012; 8:e1002375. [PubMed: 22383865]
28. Holm H, Gudbjartsson DF, Arnar DO, Thorleifsson G, Thorgeirsson G, Stefansdottir H, et al. Several common variants modulate heart rate, PR interval and QRS duration. Nat Genet. 2010; 42:117–122. [PubMed: 20062063]
29. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: A network-based approach to human disease. Nat Rev Genet. 2011; 12:56–68. [PubMed: 21164525]
30. Albert R. Scale-free networks in cell biology. J Cell Sci. 2005; 118:4947–4957. [PubMed: 16254242]
31. Lusis AJ, Weiss JN. Cardiovascular networks: Systems-based approaches to cardiovascular disease. Circulation. 2010; 121:157–170. [PubMed: 20048233]
32. Tandan S, Wang Y, Wang TT, Jiang N, Hall DD, Hell JW, et al. Physical and functional interaction between calcineurin and the cardiac L-type Ca2+ channel. Circ Res. 2009; 105:51–60. [PubMed: 19478199]
33. MacDonnell SM, Weisser-Thomas J, Kubo H, Hanscome M, Liu Q, Jaleel N, et al. CaMKII negatively regulates calcineurin-NFAT signaling in cardiac myocytes. Circ Res. 2009; 105:316–325. [PubMed: 19608982]
34. Liu Q, Busby JC, Molkentin JD. Interaction between TAK1-TAB1-TAB2 and RCAN1-calcineurin defines a signalling nodal control point. Nat Cell Biol. 2009; 11:154–161. [PubMed: 19136967]
35. Shin SY, Yang HW, Kim JR, Heo WD, Cho KH. A hidden incoherent switch regulates RCAN1 in the calcineurin-NFAT signaling network. J Cell Sci. 2011; 124:82–90. [PubMed: 21172821]

36. Ahrendt E, Braun JE. Channel triage: Emerging insights into the processing and quality control of hERG potassium channels by DnaJA proteins 1, 2 and 4. Channels (Austin). 2010; 4:335–336. [PubMed: 20699637]
37. Johannessen M, Moller S, Hansen T, Moens U, Van Ghelue M. The multifunctional roles of the four-and-a-half-LIM only protein FHL2. Cell Mol Life Sci. 2006; 63:268–284. [PubMed: 16389449]
38. Kong Y, Shelton JM, Rothermel B, Li X, Richardson JA, Bassel-Duby R, et al. Cardiac-specific LIM protein FHL2 modifies the hypertrophic response to beta-adrenergic stimulation. Circulation. 2001; 103:2731–2738. [PubMed: 11390345]
39. Chu PH, Chen J. The novel roles of four and a half LIM proteins 1 and 2 in the cardiovascular system. Chang Gung Med J. 2011; 34:127–134. [PubMed: 21539754]
40. Froese A, Breher SS, Waldeyer C, Schindler RF, Nikolaev VO, Rinne S, et al. Popeye domain containing proteins are essential for stress-mediated modulation of cardiac pacemaking in mice. J Clin Invest. 2012; 122:1119–1130. [PubMed: 22354168]
41. Vogler C, Spalek K, Aerni A, Demougin P, Muller A, Huynh KD, et al. CPEB3 is associated with human episodic memory. Front Behav Neurosci. 2009; 3:4. [PubMed: 19503753]
42. Lanfear DE. Genetic variation in the natriuretic peptide system and heart failure. Heart Fail Rev. 2010; 15:219–228. [PubMed: 18850266]
43. Munshi NV. Gene regulatory networks in cardiac conduction system development. Circ Res. 2012; 110:1525–1537. [PubMed: 22628576]
44. Burstein B, Nattel S. Atrial fibrosis: Mechanisms and clinical relevance in atrial fibrillation. J Am Coll Cardiol. 2008; 51:802–809. [PubMed: 18294563]
45. Yeh YH, Kuo CT, Lee YS, Lin YM, Nattel S, Tsai FC, et al. Region-specific gene expression profiles in the left atria of patients with valvular atrial fibrillation. Heart Rhythm. 2013; 10:383–391. [PubMed: 23183193]

Atrial fibrillation is the most common sustained cardiac arrhythmias in the United States. The genetic and molecular mechanisms governing its initiation and progression are complex, and our understanding of these mechanisms remains incomplete despite recent advances via GWAS, animal model experiments, and differential expression studies. In this study, we utilized weighted gene co-expression network analysis (WGCNA) to identify gene modules significantly associated with atrial fibrillation in a large sample of human left atrial appendage tissues. We further identified highly interconnected genes (i.e. hub genes) within these gene modules that may be novel candidates for functional studies. The discovery of the AF-associated gene modules and their corresponding hub genes provide novel insight into the gene network changes that occur with AF, and closer study of these findings can lead to more effective targeted therapies for disease management.

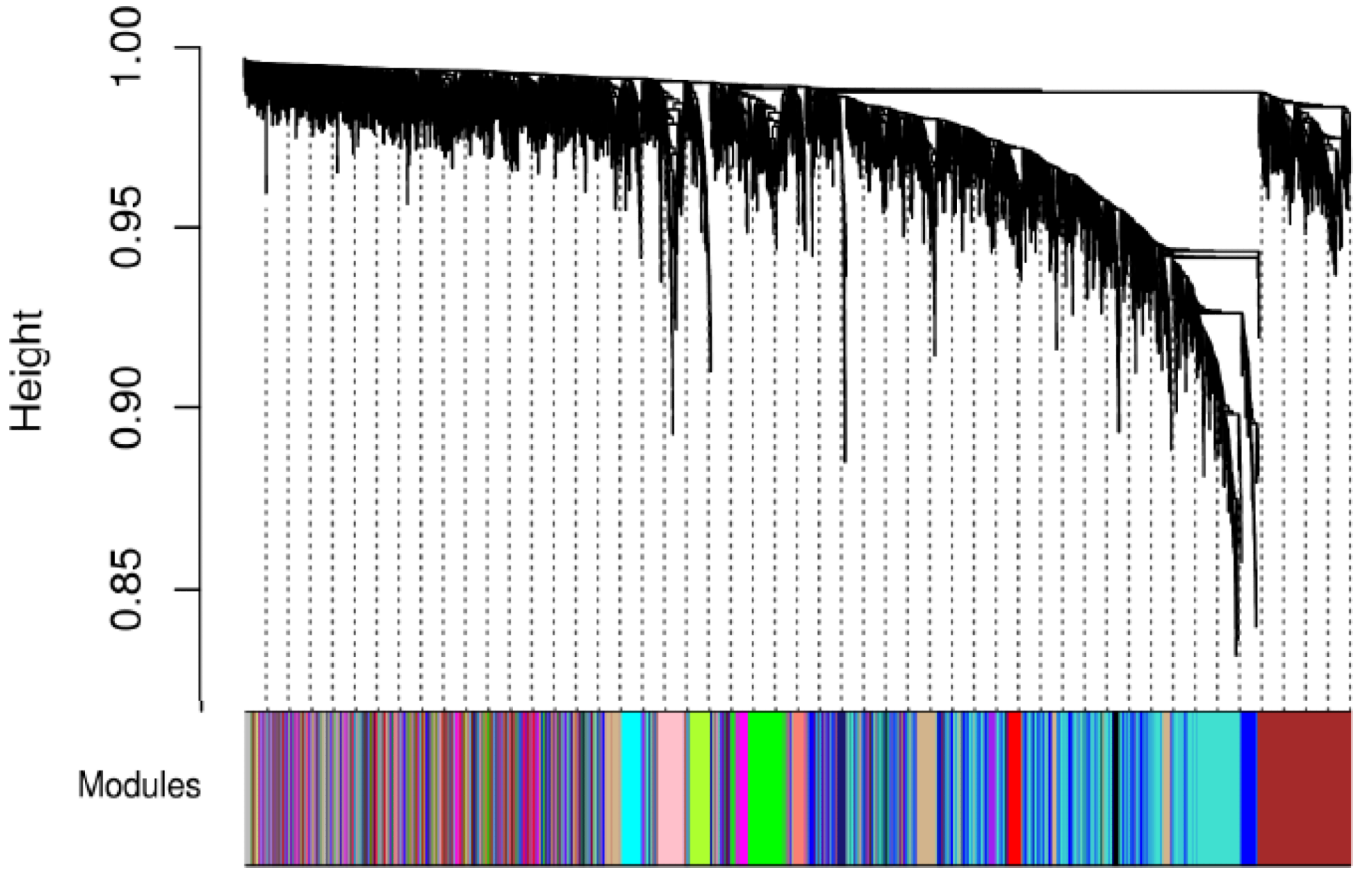

Figure 1.
Network dendrogram (top) and colors of identified modules (bottom). The dendrogram was constructed using the topological overlap matrix (TOM) as the similarity measure. Modules corresponded to branches of the dendrogram and were assigned colors for visualization.

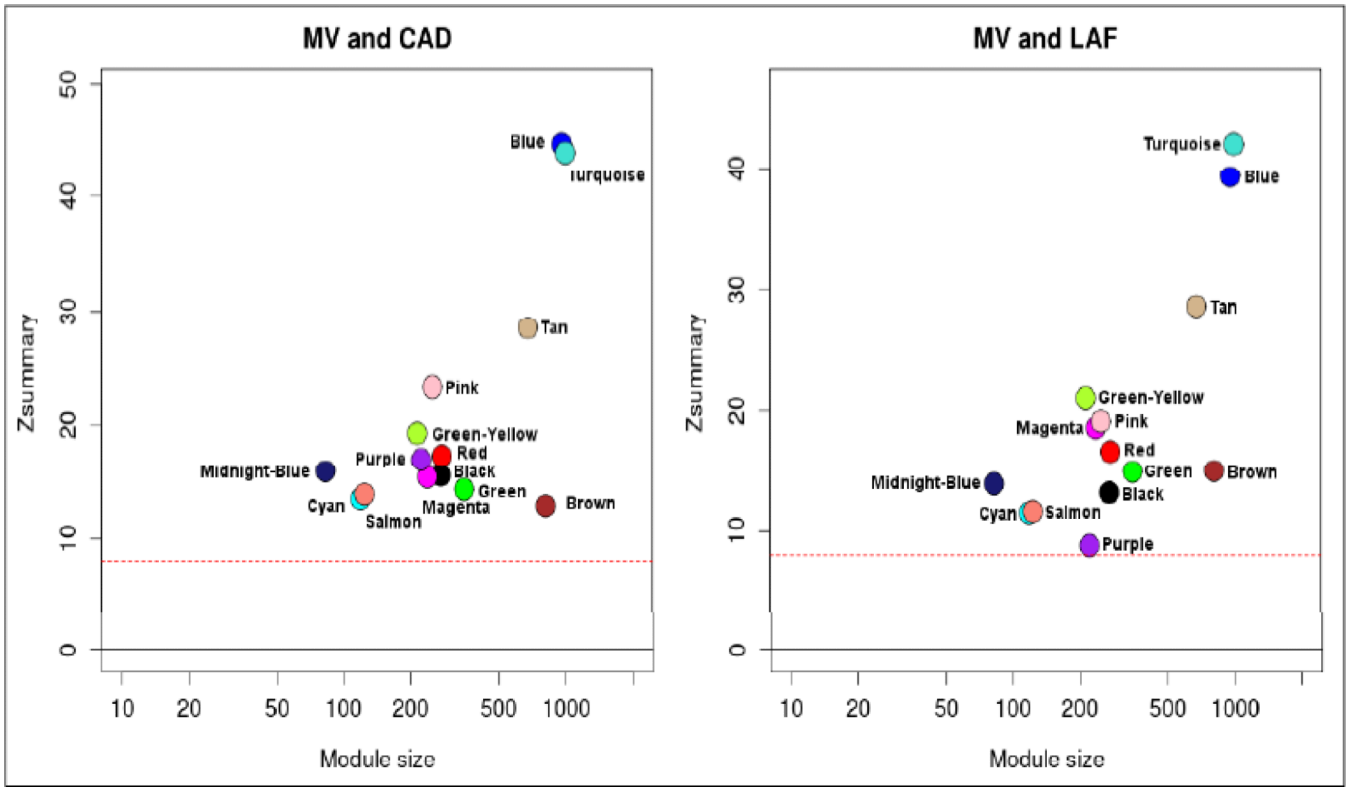

Figure 2.

Preservation of modules between MV and CAD groups (left), and MV and LAF groups (right). A $Z_{summary}$ statistic was computed for each module as an overall measure of its preservation with regards to density and connectivity. All modules showed strong preservation in both comparisons with $Z_{summary}$ scores above 8 (red dotted line).

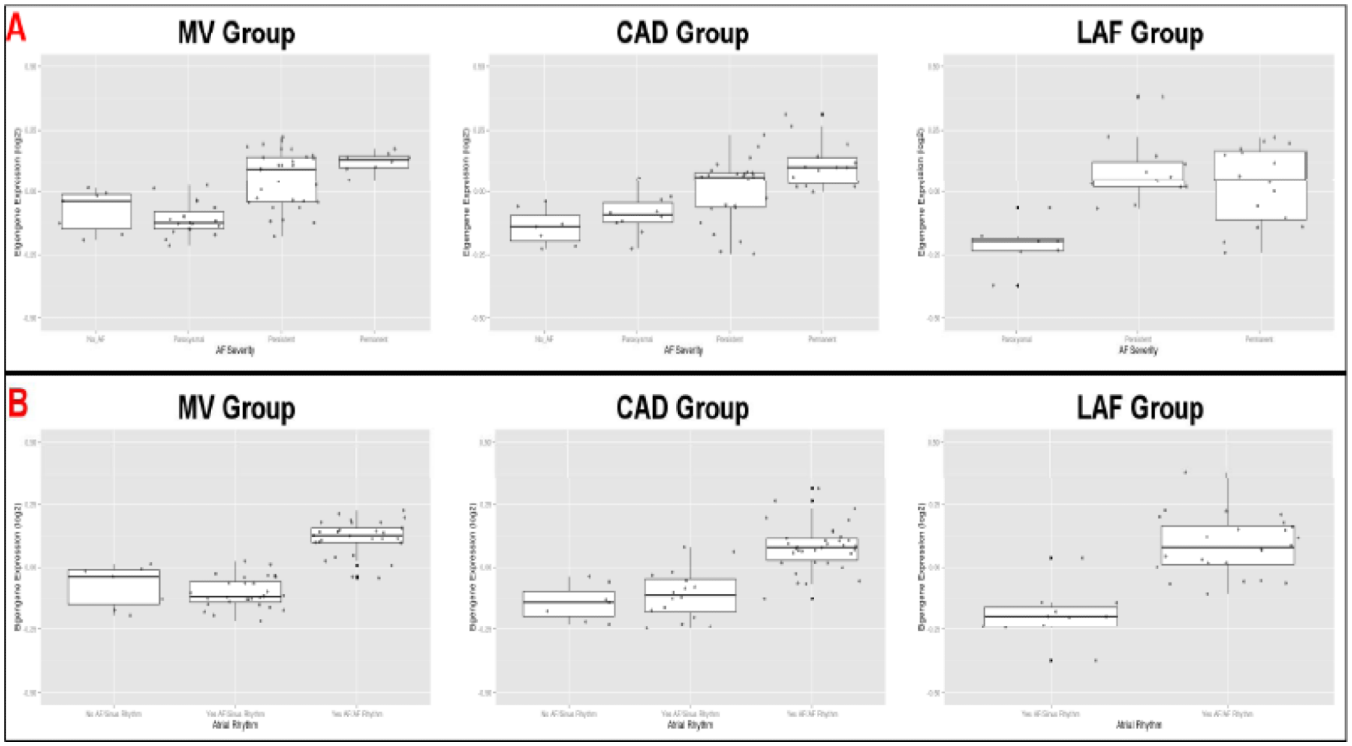

Figure 3.

Boxplots of salmon module eigengene expression levels with respect to AF severity (A; top) and atrial rhythm (B; bottom). A: Eigengene expression correlated positively with AF severity, with the largest step-wise increase between the paroxysmal AF and permanent AF categories. B: Eigengene expression was highest in the “AF History in AF Rhythm” category in all three groups.

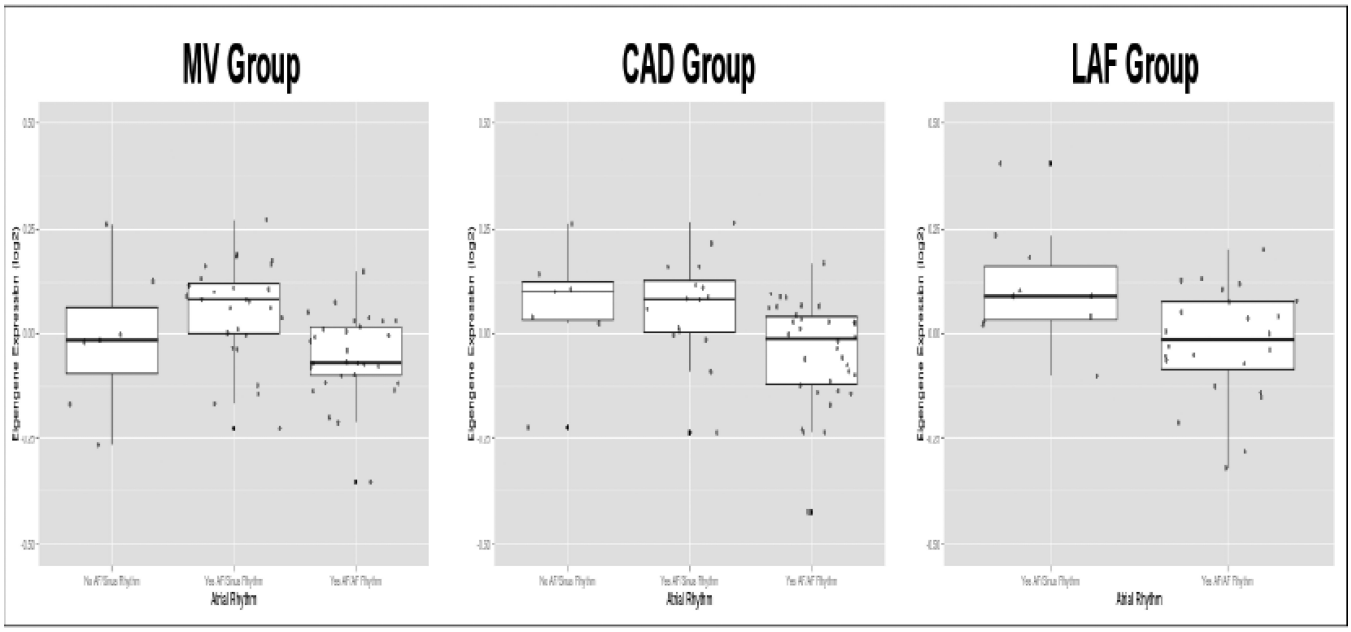

Figure 4.
Boxplots of tan module eigengene expression levels with respect to atrial rhythm. Eigengene expression levels were lower in the “AF History in AF Rhythm” category compared to the “AF History in Sinus Rhythm” category.

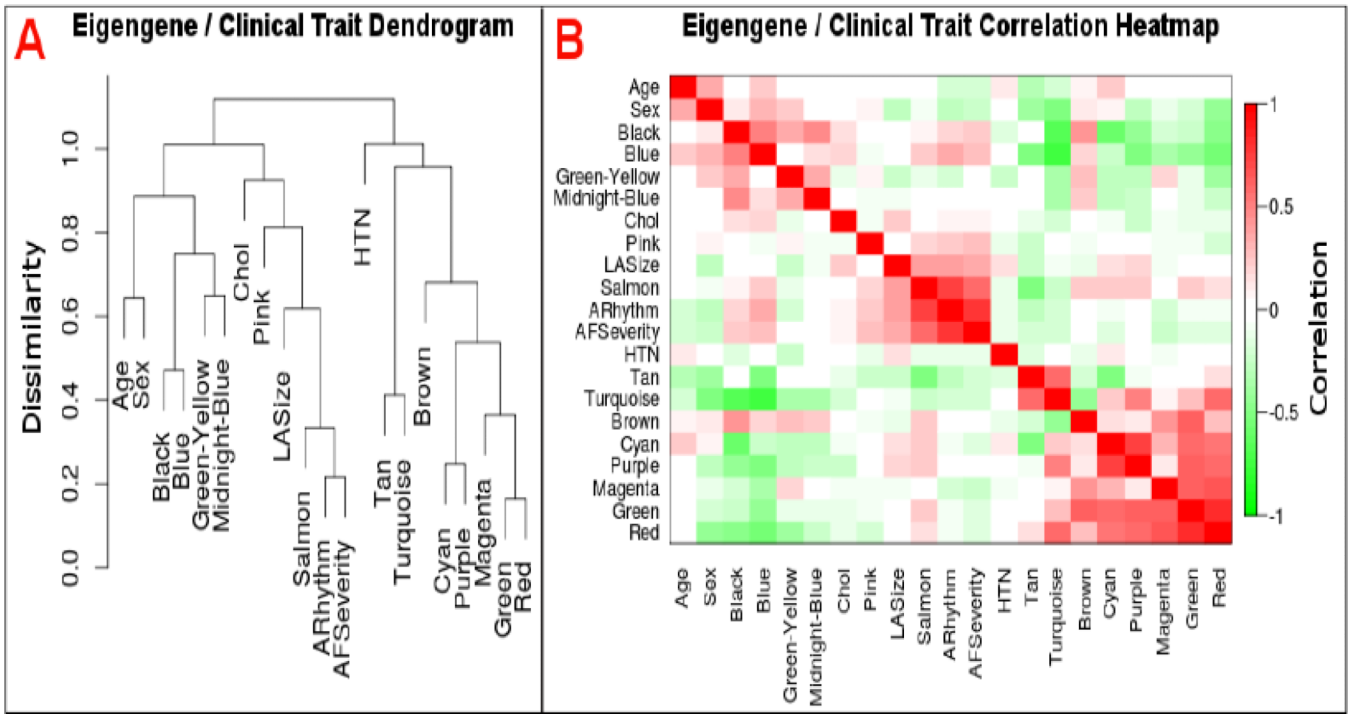

Figure 5.

Dendrogram (A; left) and correlation heatmap (B; right) of module eigengenes and clinical traits. A: The salmon module eigengene but not the tan module eigengene clustered with AF severity, atrial rhythm, and left atrial size. B: AF severity and atrial rhythm at surgery correlated positively with the salmon module eigengene and negatively with the tan module eigengene. Chol – cholesterol; LASize – left atrial size; ARhythm – atrial rhythm at surgery; HTN – hypertension.

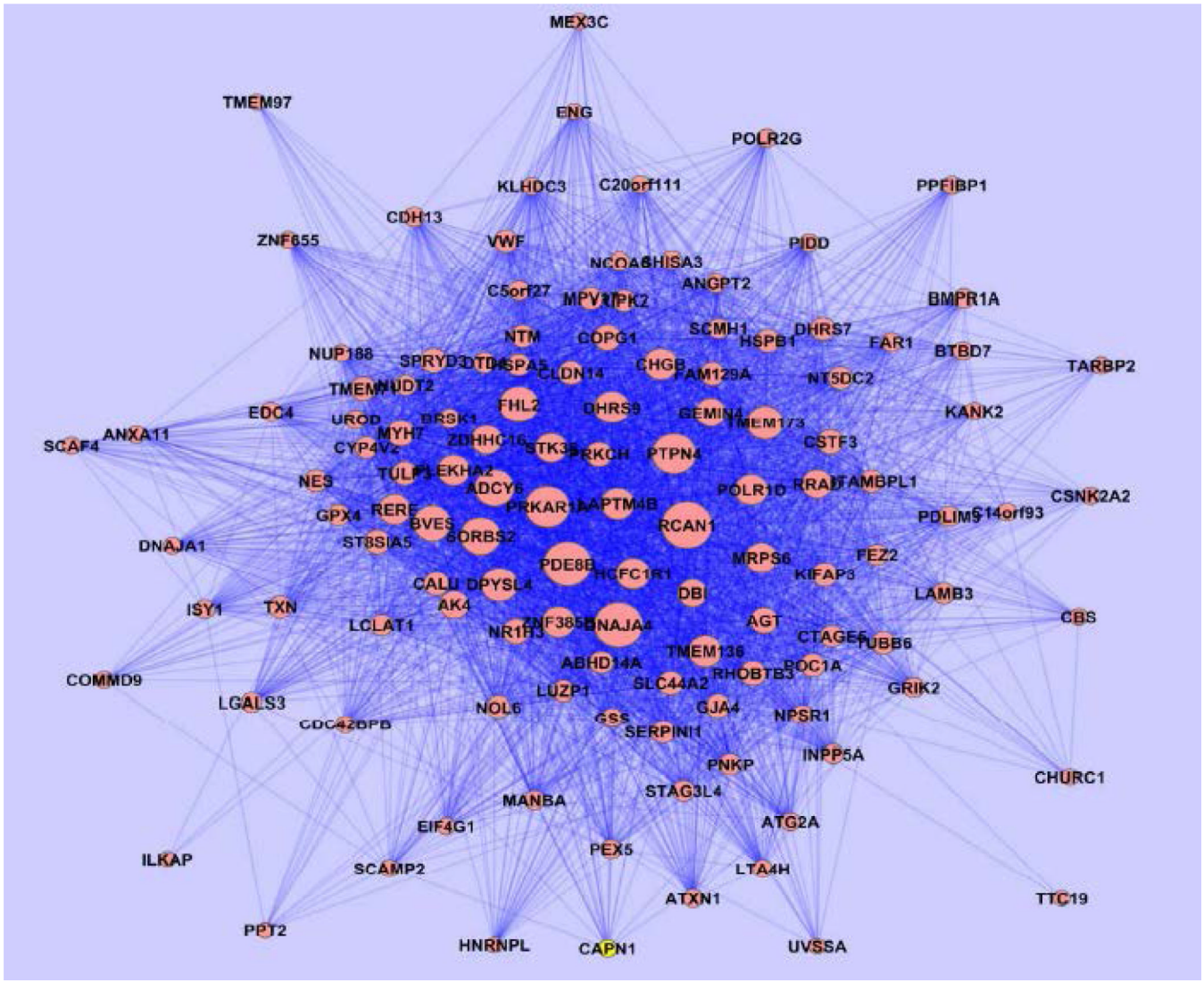

Figure 6.

Cytoscape visualization of genes in salmon module. Nodes representing genes with high intramodular connectivities, such as RCAN1 and DNAJA4, appear larger in the network. Strong connections are visualized with darker lines while weak connections appear more translucent.

Table 1

Clinical characteristics of study subjects.

Characteristic		MV Group (n=64)	CAD Group (n=57)	LAF Group (n=35)	P-Value*
Age, median years (1st, 3rd quartiles)		60 (51.75, 67.25)	64 (58.00, 70.00)	56 (45.50, 60.50)	2.0×10^{-4}
Sex, female (%)		19 (29.7%)	6 (10.5%)	7 (20.0%)	0.033
BMI, median (1st, 3rd quartiles)		25.97 (24.27, 28.66)	29.01 (27.06, 32.11)	29.71 (26.72, 35.10)	2.7×10^{-6}
Current Smoker (%)		29 (45.3%)	35 (61.4%)	12 (21.1%)	0.032
Hypertension (%)		21 (32.8%)	39 (68.4%)	16 (45.7%)	4.4×10^{-4}
AF Severity	No AF	7 (10.9%)	7 (12.3%)	0 (0.0%)	0.033
	Paroxysmal	19 (29.7%)	10 (17.5%)	7 (20.0%)	
	Persistent	30 (46.9%)	26 (45.6%)	15 (42.9%)	
	Permanent	8 (12.5%)	14 (24.6%)	13 (37.1%)	
Atrial rhythm at surgery	No AF History in Sinus Rhythm	7 (10.9%)	7 (12.3%)	0 (0%)	0.065
	AF History in Sinus Rhythm	28 (43.8%)	16 (28.1%)	11 (31.4%)	
	AF History in AF Rhythm	29 (45.3%)	34 (59.6%)	24 (68.6%)	

*P-values were computed using Kruskal-Wallis rank-sum tests for continuous variables and Pearson's χ^2-test for categorical variables.

Table 2

Top 10 hub genes in the salmon (left) and tan (right) modules as defined by intramodular connectivity (IMC) and module membership (MM).

Salmon Module						Tan Module				
Gene	**IMC**		**Gene**	**MM**		**Gene**	**IMC**		**Gene**	**MM**
RCAN1	8.2		RCAN1	0.81		CPEB3	43.3		CPEB3	0.85
DNAJA4	7.7		DNAJA4	0.81		CPLX3	42.4		CPLX3	0.84
PDE8B	7.7		PDE8B	0.80		NEDD4L	40.8		NEDD4L	0.83
PRKAR1A	6.9		PRKAR1A	0.77		SGSM1	40.7		SGSM1	0.82
PTPN4	6.7		PTPN4	0.75		UCKL1	39.0		UCKL1	0.81
SORBS2	6.0		FHL2	0.69		SOSTDC1	37.2		SOSTDC1	0.79
ADCY6	5.7		ADCY6	0.69		PRDX1	35.5		RCOR2	0.78
FHL2	5.7		SORBS2	0.68		RCOR2	35.4		EEF2K	0.77
BVES	5.4		DHRS9	0.67		NPPB	35.3		PRDX1	0.76
TMEM173	5.3		LAPTM4B	0.65		LRRN3	34.6		MMP11	0.76