

1 **Identifying functional genes and pathways towards a unifying model for atrial fibrillation**

2 Sojin Youn Wass, MD,^{1*} Erik J. Offerman, MD,^{6*} Han Sun, MS,³ Jeffrey Hsu, PhD,¹ Julie H. Rennison,
3 PhD,¹ Catherine C. Cantlay BS¹, Meghan L. McHale, BA,¹ A. Marc Gillinov, MD,^{2,4} Christine Moravec,
4 PhD,^{1,2} Jonathan D. Smith, PhD,^{1,2,5*} David R. Van Wagoner, PhD,^{1,2*} John Barnard, PhD,^{3*} Mina K.
5 Chung, MD^{1,2,5*}

6 From: The Departments of ¹Cardiovascular and Metabolic Sciences and ³Quantitative Health Sciences,
7 Lerner Research Institute, and the Departments of ⁴Cardiothoracic Surgery and ⁵Cardiovascular Medicine,
8 Heart & Vascular Institute,² Cleveland Clinic Lerner College of Medicine of Case Western Reserve
9 University, Cleveland Clinic, Cleveland, OH, ⁶ Department of Internal Medicine, New York University
10 School of Medicine, New York, NY. *Equal contributions

11
12 Short title: AF gene coexpression and pathway analyses

13 Subject codes: Atrial fibrillation, Mechanisms, Translational Studies

14 Word count: 7625

15
16 Send correspondence to:

17 Mina K. Chung, MD

18 Department of Cardiovascular Medicine

19 Cleveland Clinic

20 9500 Euclid Ave., J2-2

21 Cleveland, OH 44195

22 chungm@ccf.org

23 Phone: 216-444-2290

24

1 **ABSTRACT**

2

3 **Rationale:** Genome wide association studies (GWAS) have associated >100 genetic loci with
4 atrial fibrillation (AF), yet the biological pathways of AF remain elusive.

5 **Objective:** To determine candidate causal genes associated with AF risk loci and their
6 coexpression partners, modules, biologic and mechanistic pathways.

7 **Methods and Results:** Cis-expression quantitative trait loci (eQTLs) were identified for
8 candidate genes near AF risk single nucleotide polymorphisms (SNPs) in human left atrial
9 tissues. Genes were categorized into 3 sets according to likelihood of being a causative AF gene:
10 1) All Candidate Genes (with significant eQTLs or previously prioritized); 2) Any eQTL Genes
11 (with ≥ 1 significant eQTL); and 3) Top GWAS SNP eQTL Genes (top SNP within the top 10 eQTL
12 SNPs). Coexpression partners were identified for each candidate gene. Weighted gene coexpression
13 network analysis (WGCNA) identified modules and modules with overrepresentation of candidate AF
14 genes. Ingenuity Pathway Analysis (IPA) was applied to the coexpression partners of each
15 candidate gene, and IPA and gene set enrichment analysis (GSEA) to each WGCNA module.
16 166 AF-risk SNPs were located in 135 distinct loci. The All Candidate Genes group contained
17 233, the Any eQTL Genes group 131 (83 novel), and the Top GWAS SNP eQTL Genes group
18 37 genes. IPA identified mitochondrial dysfunction, oxidative stress, epithelial adherens junction
19 signaling, and sirtuin signaling as the most frequent pathways. WGCNA characterized 64
20 modules; candidate AF genes were overrepresented in 8. Modules were represented by cell
21 injury, death, stress, developmental, metabolic/mitochondrial, transcription/translation, and
22 immune activation/inflammation regulatory pathways.

23 **Conclusions:** AF candidate gene coexpression analyses suggest significant roles for cellular
24 stress and remodeling in AF. We propose a dual risk model for AF: Genetic susceptibility to AF

1 may not manifest until later in life, when cellular stressors overwhelm adaptive responses. These
2 analyses provide a resource for further functional studies on potential causal AF genes.

3

4 **Keywords:** atrial fibrillation, cellular stress, remodeling

1 **INTRODUCTION**

2

3 Atrial fibrillation (AF) affects over 33.5 million worldwide, is strongly associated with aging and
4 obesity, and is projected to increase further with the aging population and increasing prevalence
5 of obesity¹. AF is often initiated by triggers from the pulmonary vein ostia, which are targets for
6 AF ablation, and maintained by vulnerable atrial substrates that promote AF propagation and
7 persistence². Atrial substrates become vulnerable from structural, electrical, and autonomic
8 remodeling promoted by conditions that can cause stretch, inflammation, oxidative stress, or
9 calcium overload².

10

11 Genome-wide association studies (GWAS) have identified multiple loci associated with AF risk;
12 the top locus is on chromosome 4q25 near *PITX2*, which has been implicated in pulmonary vein
13 formation³⁻⁶. Using a collection of human left atrial appendage (LAA) tissue samples, we
14 performed RNA sequencing (RNASeq) to identify cis-expression quantitative trait loci (cis-
15 eQTLs), where AF GWAS SNP genotypes correlated with the expression of nearby genes; we
16 identified candidate genes and top eQTL SNPs for 12 of 23 AF GWAS loci known at that
17 time³. Since then, larger GWAS meta-analyses have identified over 100 AF risk loci, most of
18 which have not yet undergone functional analyses^{5,6}.

19

20 Although some of the identified candidate genes have transcriptional products with known
21 function, the biological function of many remain poorly understood in the heart. Whole-genome
22 coexpression and pathway analysis of the candidate gene(s) associated with a target locus can
23 provide insights into their mechanisms and regulation that can guide further functional studies⁷.

1 To this end, we aimed to: 1) determine candidate causal genes associated with the expanded list
2 of AF risk loci by identification of cis-eQTLs in human LAA tissues; 2) identify coexpression
3 partners of AF candidate genes; and 3) identify biological pathways by which these genes might
4 function, focusing on candidate genes whose functional connections to AF are poorly
5 understood.

6

7 **METHODS**

8

9 **Left atrial tissue collection and processing.**

10 Human LAA tissues were obtained from 265 patients, including 251 who underwent elective
11 cardiac surgery and 14 from non-failing unused transplant donors. Prior to 2008 verbal consent
12 was obtained and documented in a process approved by the Cleveland Clinic Institutional
13 Review Board. After this time written informed consent was obtained. Consent to use donor
14 hearts was obtained from the family. The patient population, tissue processing, genomic DNA
15 isolation, single nucleotide polymorphism (SNP) genotyping, RNA isolation and sequencing
16 have been previously reported³. Briefly, DNA was genotyped using Illumina Hap550v3 and
17 Hap610-quad SNP beadchips and SNP data imputed to the 1000 Genomes Project phase 2³.
18 RNA unstranded 100-bp paired-end sequencing was performed on the Illumina HiSeq 2000
19 platform. Gene read counts are available in the Gene Expression Omnibus database (GSE69890).

20

21 ***Cis*-eQTL analyses**

22 RNASeq and eQTL statistical analyses were previously described, with cis-eQTLs defined as
23 being ± 250 kb from the SNP³. Expression values were obtained from aligned files using HTseq⁸.

1 Reads were quantile normalized, and gene counts for eQTL analysis were variance-stabilized
2 transformed using the R package Deseq2. Samples were analyzed in R (version 3.4.4) using the
3 surrogate variable analysis (SVA) and Limma R packages. eQTL analyses were performed for
4 235 samples of European descent and can be found at <http://afeqtls.lerner.ccf.org>³.

6 **Determination of candidate AF risk genes**

7 Using our mRNA-Seq eQTL database, we determined candidate genes near the top SNPs at each
8 locus published from two recent GWAS meta-analyses^{5,6}. Genes were categorized into 3 sets
9 according to likelihood of being a causative AF gene, from weakest to strongest: 1) All potential
10 candidate AF genes (All Candidate Genes), consisting of genes with significant cis-eQTLs (q
11 value <0.05) from the two GWAS meta-analyses, or if no significant eQTL was identified, then
12 the prioritized genes by Nielsen, et al⁶; 2) Genes with at least one AF SNP associated with its
13 gene expression with q value <0.05 from our eQTL study (Any eQTL Genes); and 3) eQTL
14 genes with the top GWAS SNP within the top 10 eQTL SNPs (Top GWAS SNP eQTL Genes).

15
16 All Candidate Genes group is the most liberal and comprehensive list, including all cis-eQTL
17 genes and non-cis-eQTL genes prioritized by Nielsen et al. (including prioritized genes that were
18 not expressed in our LA mRNA-Seq database). This list allows for inspection of potential AF
19 genes, such as *PITX2* or other developmentally active genes that are no longer highly expressed
20 in adult left atrial appendage tissue, but which are still likely to play a significant role in AF
21 development. Any eQTL Genes group includes those showing evidence of a gene expression
22 relationship with an AF risk SNP. Top AF GWAS SNP eQTL Genes group is the most stringent

1 list in level of specificity, but is also the lowest in sensitivity, including genes with the highest
2 level of evidence for future functional studies.

3

4 **Gene coexpression analyses**

5 Coexpression analysis was performed for all genes in the All Candidate Genes group across the
6 Ensembl genome database version 71. Each gene was included in individual linear models as a
7 predictor with all other Ensembl genes as outcome, additionally controlling for AF rhythm, sex,
8 age, and surrogate variables. The number of surrogate variables was determined using the R
9 surrogate variable analysis (SVA) package; surrogate variables were recalculated for each of the
10 genes. The SVA used unsupervised models to correct for the heterogeneity in gene expressions,
11 such as batch effects, unknown factors, library depth, or comorbidities like AF status, history of
12 structural heart disease, etc. We controlled the false discovery rate (FDR) of the multiple testing
13 with the R q value package. Genes were considered significantly coexpressed if q value was
14 <0.05 (Supplemental Excel 1).

15

16 Expression correlations were also performed, limiting analyses to the genes within each gene
17 group, i.e. the correlation of each candidate AF risk gene with all other AF candidate genes was
18 calculated. Genes with a q value <0.05 were considered significantly coexpressed. For each gene
19 group, agglomerative hierarchical clustering with the complete linkage method was used, where
20 genes start as singletons and are successively joined to form one cluster that contains all the
21 genes.

22

23 **Weighted Gene Coexpression Network Analysis (WGCNA)**

1 WGCNA is a method to cluster genes into distinct modules by degree of connectivity determined
2 by commonly coexpressed (as measured by correlation) genes. To construct the required
3 correlation matrix for WGCNA, first the conditional mean (log₂ scale) of LAA gene expression
4 was modeled using the limma-voom⁹ approach with an extensive set of known AF risk factors
5 and other covariates. Residuals were calculated from this model to remove mean effects and used
6 as inputs to the RSVP¹⁰ approach for estimating correlation matrices in the presence of latent
7 factors. The RSVP adjusted correlation matrix was then normalized using the spqn approach¹¹ to
8 remove the mean-variance bias common in RNA-seq data due to differential measurement error.
9 A weighted gene correlation network was built using the RSVP-spqn adjusted and normalized
10 correlation matrix using the CEMiTool R package¹². Hub genes were defined as genes in the top
11 1% for intra-module connectivity (Supplemental Excel 2).

13 **Candidate pathway identification**

14 Canonical pathways, upstream regulators, and toxicity lists were identified using the Ingenuity
15 Pathway Analysis (IPA) software package (QIAGEN, Inc.). IPA uses gene expression data to
16 predict biological pathways that are activated by comparing known genes within a set of
17 pathways to expression patterns in the imported dataset. Ensemble ID number, q-value, and log
18 fold change values for coexpressed genes were imported into IPA. Coexpressed genes with q
19 values <0.05 were used for analysis. Significance of association between the data and canonical
20 pathways was determined by the ratio of the number of overlapping genes from our dataset to the
21 total number of IPA dataset genes in a particular pathway, to determine the most significantly
22 affected pathways. Fisher's exact test was used to determine the p-value of association between
23 genes and the canonical pathway, where p <0.05 was considered significant. Toxicity lists are

1 generated based on sets of molecules that are known to be involved in a particular type of
2 toxicity. IPA analyses were performed for the coexpression partners of each gene in the All
3 Candidate Genes group (Supplemental Excel 3-5), and for each module from the WGCNA
4 (Supplemental Excel 6). Over representation analysis (ORA) was performed on each WGCNA
5 module using the CEMiTool package¹² and the Molecular Signatures Database¹³ v7.1
6 (Supplemental Excel 7).

7

8 **RESULTS**

9

10 Patient characteristics of the 235 subjects have been previously described³.

11

12 ***AF GWAS Loci SNPs.*** AF GWAS risk SNPs from two recently published meta-analyses were
13 aggregated to create a comprehensive list of 166 SNPs: 142 risk variants were identified by
14 Nielsen, et al⁶, with an additional 24 independent risk variants identified by Roselli, et al⁵. The
15 166 SNPs were located in 135 distinct loci (Supplemental Excel 8). The locus with the most
16 independent risk SNPs⁶ was at 4q25, near *PITX2*, which had 10; loci labelled *SCN10A*, *SCN5A*,
17 *KCNN3*, *NEURL1*, *RPL3L*, and *LINC01142* each had 3; loci *LRRC10*, *NKX2-5*, *ZFHX3*,
18 *LINC00477*, *HSF2*, *NACA*, *AKAP6*, RP11_380D231 (AC098798.1), *FBXO32*, *MEX3C* had 2.
19 All remaining loci were associated with 1 risk SNP.

20

21 ***AF Candidate Gene Lists*** (Table 1). Of the 166 risk variant SNPs from the published meta-
22 analyses, 80 SNPs had significant cis-eQTL association ($q < 0.05$) with 131 genes from our LAA
23 mRNA-Seq eQTL European- ancestry database comprising the Any eQTL Genes group. Of the

1 131 genes in this group, 83 genes had not been previously annotated by Nielsen, et al⁶, as
2 potential AF genes (Table 2). The functions of many of these genes are not yet known.

3
4 The All Candidate Genes group comprised the 131 genes from the Any eQTL Genes group and
5 102 genes prioritized by Nielsen, et al⁶ (Table 1). Of the 233 genes, 111 are known protein-
6 coding genes (Supplemental Figure 1), encoding mostly enzymes and transcription factors with
7 predominant localization in the cytoplasm (32%) or nucleus (22%) and the remaining in the
8 plasma membrane (17%), extracellular space (7%) or other (22%). By function (Supplemental
9 Figure 1), 12% encode transcription regulators, 12% enzymes, 5% transporters, 4% ion channels,
10 4% kinases, 3% peptidases, and 2% transmembrane receptors.

11
12 ***Genes Coexpressed with Each Candidate Gene.*** Of 233 candidate genes in the All Candidate
13 Genes group, 5 AF candidate genes were not significantly expressed in our human adult LAA
14 tissues (MIR6500, *PANDAR*, MIR54810, MIR548AZ, LOC101926943) and 3
15 (ENSG00000269728, ENSG00000255864, ENSG00000270028) were retired by Ensembl.
16 Coexpression partners of the remaining 228 AF candidate genes ranged from 0 to 1928 genes.
17 C4orf32 had the highest number of coexpressed genes (1928 genes), followed by *FXR2*, which
18 was significantly coexpressed with 1889 genes.

19
20 ***Novel candidate genes identified by coexpression analyses.*** Adenosine monophosphate (AMP)
21 activated protein kinase (AMPK) alpha 2 catalytic subunit (*PRKAA2*) and calcium/calmodulin-
22 dependent protein kinase II beta (*CAMK2B*) were the most common genes co-expressed with the
23 AF candidate genes that are not themselves GWAS-identified. Other commonly coexpressed

1 genes included prostaglandin F2 receptor inhibitor (*PTGFRN*), talin 1 (*TLN1*), sarcolemma
2 associated protein (*SLMAP*), PDZ and LIM domain 5 (*PDLIM5*), and methionine sulfoxide
3 reductase B3 (*MSRB3*) (Supplemental Figure 2).

4
5 ***Top Canonical Pathways of Coexpressed Genes.*** IPA was used to identify top candidate
6 pathways of each group of genome-wide coexpressed partners for each gene in the All Candidate
7 Genes group (Supplemental Excel Folder10). After ranking IPA canonical pathways for each
8 candidate gene by p-value, the top 5 significant canonical pathways were listed for each gene
9 (Supplemental Excel 9). Mitochondrial dysfunction, oxidative phosphorylation, epithelial
10 adherens junction signaling, sirtuin signaling pathway and actin cytoskeleton pathway were the
11 most frequent pathways listed among the top 5 canonical pathways (Supplemental Excel 3). For
12 the Any eQTL Genes group (Supplemental Excel 4), epithelial adherens junction signaling,
13 mitochondrial dysfunction and oxidative phosphorylation, germ cell-Sertoli cell junction
14 signaling and axonal guidance signaling were the most frequent pathways listed. For the Top
15 GWAS SNP eQTL Genes group (Supplemental Excel 5), EIF2 signaling, axonal guidance
16 signaling, sertoli cell-sertoli cell junction signaling, regulation of eIF4 and p70S6K signaling,
17 oxidative phosphorylation, and mitochondrial dysfunction were the most frequent pathways
18 listed among the top 5 canonical pathways. Overall, the most frequent pathways were those
19 involved in energy and metabolism signaling, including mitochondrial dysfunction, oxidative
20 phosphorylation, and sirtuin signaling. Stress and injury signaling pathways were also
21 represented and included EIF2 signaling pathways, regulation of eIF4 and p70S6K signaling,
22 protein ubiquitination and cardiac hypertrophy signaling.

23

1 Top canonical pathways, top upstream regulators and top networks from pathway analyses
2 within each of the gene groups are shown in Supplemental Table 1.

3
4 **WGCNA Modules.** WGCNA characterized 64 distinct modules from 17,357 genes expressed in
5 the LAA (Supplemental Excel 2). 48 of these modules, including a module containing
6 uncorrelated genes ('not_correlated'), contained at least one AF candidate gene. Candidate AF
7 genes were overrepresented in 8 modules ($p < 0.05$; modules 21, 2, not correlated, 25, 11, 42, 39,
8 24). Hierarchical clustering (Figure 1) of the modules' top eigengene produced 15 clusters. A list
9 of genes for each module, alongside with hub genes that were determined by module
10 connectivity for each module are available in Supplemental Excel file 2.

11
12 Concordant with IPA analyses, the modules with AF risk gene overrepresentation suggest that
13 AF genes are over represented in modules related to intracellular ion homeostasis,
14 metabolic/energy utilization, gene expression regulation, and signaling pathways of stress and
15 unfolded protein responses.

16
17 The clustering of modules 25, 39, and 42 (Figure 1), suggests that genes in these three modules
18 are closely related. Gene set enrichment analyses of these 3 modules (Supplemental Excels 6 &
19 7) suggest that genes in these modules are involved in remodeling of the heart through a
20 combination of myogenesis and the unfolded protein response.

21
22 **Biological pathways of top AF candidate genes.** Coexpression analyses can provide insight into
23 biological function of top AF candidate genes. These include transcription factors such as *PITX2*,

1 *ZFHX3*, top implicated genes by AF GWAS for which function remains elusive, as well as
2 candidate causal genes in the Top GWAS SNP eQTL Genes group. IPA pathways for *PITX2*
3 (Figure 2) and *ZFHX3* (Figure 3) are shown.

5 **DISCUSSION**

6 Despite having identified over 135 genetic loci associated with AF from prior meta-analyses,
7 establishing functional connections and causal genes contributing to AF pathophysiology
8 remains challenging. Here we utilized human atrial eQTL analyses to identify top candidate
9 causal AF genes. To advance our understanding of the biological mechanisms of these candidate
10 genes, we identified gene coexpression partners for each candidate gene and WGCNA modules
11 that inform functional characterization using pathway and gene set analyses. This application
12 provides insight into the function of genes nearest the top AF GWAS loci, for example *PITX2*
13 and *ZFHX3*, for which connections to AF mechanisms remain challenging.

15 *PITX2*

16 The top locus implicated by AF risk GWAS is on chromosome 4q25. The nearest gene is
17 *PITX2*^{5,6}, which encodes a transcription factor that regulates left-right asymmetry during organ
18 morphogenesis, suppresses a sinus node program in the LA, and is critical to formation of the
19 pulmonary veins¹⁴, making it an attractive candidate causal gene. *PITX2* is active during cardiac
20 development and decreases with maturation, likely contributing to our failure to identify
21 significant eQTLs in our adult LAA tissues. Nevertheless, *PITX2* remains expressed in our adult
22 LA tissues, though in lower levels in patients with AF,³ suggesting an additional function(s) in
23 the adult LA. Top canonical pathways of genes coexpressed with *PITX2* (Figure 2) include

1 amino acid metabolism (e.g. alanine degradation, valine, 4-aminobutyrate and glutamine
2 degradation, driven by *ABAT* and *ALDH6A1*), and major signaling pathways, such as PI3K/AKT,
3 G protein coupled receptor signaling, IGF-1, PDGF signaling, renin-angiotensin and AMPK
4 signaling pathways, all commonly mediated by the phosphoinositide-3-kinase regulatory subunit
5 1 (*PIK3R1*), and also AMPK non-catalytic subunit γ 2 (*PRKAG2*). *PIK3R1* encodes for a subunit
6 of PIK3, which is critical in insulin resistance¹⁵, and in the regulation of myogenic
7 differentiation, myoblast proliferation, and myotube hypertrophy^{16,17}. The PI3K-Akt signaling
8 pathway is activated by insulin like growth factor-1 (IGF-1)¹⁷ which is critical in regulation of
9 energy metabolism in cardiomyocytes¹⁸, and decreases autophagy and apoptosis¹⁹⁻²². Inhibition
10 of the PI3K-Akt signaling in the heart is thought to contribute to AF in patients on ibrutinib²³, an
11 immunotherapeutic agent that acts as a Bruton's tyrosine kinase inhibitor²⁴. *PRKAG2* is one of
12 three AMPK regulatory subunits, and is abundant in the heart²⁵. *PRKAG2* gain of function
13 mutations increase AMPK activity and promote accumulation of glycogen in cardiac tissues via
14 transcriptional changes in glucose handling regulators²⁶. *PRKAG2* mutations are considered a
15 metabolic storage disease and not a disease of the sarcomeres - this is further evidenced by the
16 absence of fibrosis or myocyte disarray in patients with a mutation presenting with either
17 hypertrophic cardiomyopathy and/or conduction abnormalities²⁷. Increased AMPK activity and
18 the resultant glycogen accumulation are also correlated with changes in metabolic pathways and
19 AKT phosphorylation²⁶. The AKT signaling pathway also regulates myocyte hypertrophy²⁸.
20 These findings suggest that *PITX2* may play a role in metabolic energy utilization and/or
21 regulation in the heart, and dysregulation may cause metabolic dysregulation and imbalance in
22 autophagy, with subsequent disease.

23

1 IPA analysis also highlighted fibrotic signaling pathways, driven by transmembrane protein 67
2 (*TMEM67*), cannabinoid receptor 1 (*CNR1*), collagen type IV α 2 chain (*COL4A2*), insulin like
3 growth factor binding protein 5 (*IGFBP5*), adrenergic receptor β 2 (*ADRB2*), plasminogen
4 activator, urokinase (*PLAU*), angiotensin I converting enzyme (*ACE*), neurofibromin 1 (*NF*), and
5 Xin actin binding repeat containing 1 (*XIRP1*). AMPK activation inhibits Transforming Growth
6 Factor (TGF) beta signaling and target transcripts of pathways associated with cardiac fibrosis²⁶.
7 Although *PITX2* had no prior known role in fibrosis signaling, this network pattern suggests that
8 *PITX2* may suppress general fibrosis and tissue remodeling pathways, perhaps by the AMPK
9 pathway; *PITX2* loss of function may promote activation of fibrosis pathways, which may
10 predispose to AF.

11

12 *PITX2* is most strongly coexpressed with *PITX2* adjacent non-coding RNA (*PANCR*), which we
13 reported is an upstream regulator of *PITX2*²⁹. Both are located in module 58, which contains
14 genes encoding ion channels (*SCN4B*, *ANO1*, *KCNJ8*, *CACNA1H*, *P2RX3*, *KCNT1*, *SCN3A*),
15 cytoskeletal proteins (*MYH9*, *ADCY3*, *MYO1B*, *MYO6*), and G protein coupled receptors. These
16 associations suggest further functional impact of *PITX2* on these pathways in adult atria.

17

18 *ZFHX3*

19 *ZFHX3* is associated with the regulation of numerous pathways, including growth and
20 differentiation of neuronal tissue³⁰, regulation of inflammation³¹, and tumor suppression through
21 regulation of *MYC* transcription³². *ZFHX3* has been implicated in calcium homeostasis, and
22 knock down of *ZFHX3* promotes arrhythmogenesis in atrial myocytes,³³ mediated by increased
23 expression of sarcoplasmic/endoplasmic reticulum Ca²⁺ATPase 2a (*SERCA2a*), ryanodine

1 receptors, ultra-rapid delayed rectifier potassium currents, transient outward currents and
2 acetylcholine-sensitive potassium currents.

3
4 WGCNA analysis associates *ZFHX3* with module 2, which was overrepresented by AF risk
5 genes. Hub genes of module 2 include rho guanine nucleotide exchange factor 12 (*ARHGEF12*),
6 phosphoinositide kinase, FYVE-type zinc finger containing (*PIKFYVE*), and kinesin family
7 member 1B (*KIF1B*). Gene enrichment analyses suggest module 2 likely contains genes critical
8 in mitochondrial function modulation and apoptosis.

9
10 The top canonical pathways of *ZFHX3* and coexpressed genes include cell proliferation
11 pathways such as acute myeloid leukemia, and prostate cancer signaling pathways, driven by
12 *KRAS*, mitogen-activated protein kinase kinase 6 (*MAP2K6*), phosphatidylinositol-4,5-
13 bisphosphate 3-kinase catalytic subunit α (*PIK3CA*), cAMP responsive element binding protein
14 1 (*CREB1*), and phosphatase and tensin homolog (*PTEN*). *RAS*, *PI3K* and *CREB* gene expression
15 is positively associated with *ZFHX3* (Figure 3). As these genes have been linked to modulation
16 of cardiomyocyte apoptosis,^{19,22,34} *ZFHX3* may also regulate apoptosis and hypertrophy of
17 cardiomyocytes. This is consistent with the top toxicity lists suggested by IPA which include
18 cardiac necrosis/cell death, cardiac fibrosis and hypertrophy, and followed by RAR activation,
19 and NRF2-mediated oxidative stress.

20

21

22 ***Identification of Novel AF Candidate Genes by Coexpression Analyses***

1 Identification of genes coexpressed with GWAS-identified AF risk genes provides novel insights
2 into AF. While not identified by GWAS as AF risk genes, these genes may also play significant
3 roles in AF pathogenesis or yield additional targets for functional studies. Based on frequency of
4 being coexpressed with the putative AF genes, we identified two genes of potential interest in AF
5 pathophysiology: PRKAA2, the $\alpha 2$ catalytic subunit of AMPK, and CAMK2B, a subunit of
6 calcium/calmodulin-dependent protein kinase II (*CAMKII*).

7
8 PRKAA2 is a catalytic subunit of AMPK in the heart³⁵. AMPK plays an important role in
9 cellular energy and metabolism sensing and regulation. Activation of AMPK induces
10 mitochondrial biogenesis, as well as increases insulin sensitivity and fatty acid oxidation³⁶.
11 AMPK activity is induced by drugs, such as metformin and thiazolidinediones³⁷. Metformin
12 affects phosphorylation of PRKAA2³⁷. Activity of PRKAA2 increases with muscle contraction
13 and exercise³⁵. *PRKAA2* deletion causes defects in autophagy, mitochondrial fission and
14 fragmentation, and insulin resistance³⁸ in mice. In the heart, *PRKAA2* deletion increases release
15 of atrial natriuretic peptide, myocardial fibrosis and hypertrophy, with subsequent decrease in
16 ejection fraction³⁹ and arrhythmogenesis⁴⁰. In a study of 645,710 diabetics, metformin was
17 associated with 20% less AF⁴¹. *PRKAA2* is located in module 1. GSEA and IPA analyses suggest
18 genes in module 1 may be involved in cell death signaling.

19
20 CAMK2B is a subunit of calcium/calmodulin-dependent protein kinase II (*CAMKII*) that is
21 expressed primarily in the brain, followed by the heart. In the heart *CAMKII* is activated by β -
22 adrenergic signaling, Gq coupled receptor signaling, presence of reactive oxygen species,
23 hyperglycemia, ischemic injury, and other stimuli that cause increased intracellular calcium⁴².

1 Persistent activation of *CAMKII* promotes arrhythmogenesis, myocardial apoptosis and
2 necroptosis⁴³. In contrast to apoptosis that requires ATP derived from mitochondria, necroptosis
3 is mediated by opening of the mitochondrial permeability transition pore (mPTP), resulting in
4 dissipation of mitochondrial membrane potential and subsequent shutdown of the
5 mitochondria⁴². Necroptosis triggers an inflammatory response that results in tissue injury and
6 likely contributes to many cardiovascular disease processes⁴². *CAMK2B* is located in module
7 22 which is associated with genes involved in myogenesis and development.

8

9 ***WGCNA modules with overrepresentation of AF risk genes***

10 AF genes are over represented in modules related to intracellular ion homeostasis,
11 metabolic/energy utilization, gene expression regulation, and signaling pathways of stress and
12 unfolded protein response (Table 3).

13

14 One of the most statistically significant modules with AF over representation is module 2. Heme
15 metabolism is the hallmark pathway from GSEA for module 2. Iron deficiency is seen in patients
16 with heart failure, and portends higher mortality⁴⁴. In contrast, iron overload can result in AF⁴⁵.
17 Iron homeostasis affects mitochondrial function and ATP- linked respiration by modulating
18 activity of electron train complexes I, II and III⁴⁶. Respirometry studies from human LAA tissues
19 of patients with AF and diabetes undergoing coronary artery bypass graft surgery also implicate
20 impairment of Complexes I and II⁴⁷. In rapidly paced cardiomyocytes, mitochondrial dysfunction
21 has been linked to AF progression, mediated by activation of autophagy and endoplasmic
22 reticulum stress induction⁴⁸. One of the top canonical pathways from IPA analysis of module 1
23 includes Protein Kinase A (PKA) signaling. Oxidative stress activates PKA signaling in cardiac

1 myocytes which triggers apoptosis and reduced mitochondrial respiration⁴⁹. PKA activation
2 indirectly inhibits mitochondrial fission and increases mitochondrial fusion, with resultant
3 superoxide production and altered mitochondrial potential with increase in mitochondrial pore
4 permeability⁵⁰. Analysis results of module 2 from both IPA and GSEA suggest a potential
5 significant role of the mitochondria in AF.

6
7 These examples demonstrate how WGCNA modules and pathway analyses of coexpressed genes
8 can help elucidate potential function of AF risk genes, including those for which mechanistic
9 connections to the disease have remained poorly defined. Our overall findings are consistent
10 with the two meta-analyses reported by Roselli et al⁵, and Nielsen et al.,⁶ but are also distinct
11 from previous findings. Roselli et al⁵ noted themes of cardiac development, electrophysiological,
12 and cardiomyocyte contractile or structural functional groups. Nielsen et al.⁶ noted cardiac
13 structural remodeling as a recurrent theme and postulated that genes active during fetal heart
14 development are reactivated in response to stress in the adult heart. We also see themes related to
15 cardiac structural change, but also discovered involvement of unfolded protein response,
16 mitochondrial and metabolic pathways, especially those of AMPK.

17

18 ***Dual Risk Model for Development of AF***

19 Differences between the All Candidate Genes group and the Any eQTL Genes group highlight
20 the differing expression patterns that may be present in cardiac development compared to later in
21 life, when AF typically presents. The All Candidate Genes group contains GWAS implicated
22 genes active in development. The Any eQTL Genes group indicate genes that are active in adult
23 LAA, whose expression may serve as a barometer of function in the adult atria. These findings

1 support our proposed dual risk model for development of AF (Figure 4)⁵¹. In this model, genes
2 active during cardiac structural development contribute to genetic susceptibility to AF, for
3 example potentially by alterations in pulmonary vein development (e.g. *PITX2*) that may
4 contribute to pulmonary vein triggers. However, with exposure to stresses later in life,
5 transcriptional responses to cellular stress, such as metabolic stress, mitochondrial dysfunction,
6 oxidative stress, sirtuin signaling and abnormal proteostasis may lead to altered, overwhelmed or
7 reduced transcriptional responses, as suggested by our prior study of differential gene expression
8 in AF⁵² and the current analyses. Indeed, AF is associated with heterogeneous cardiovascular
9 stressors, including aging, obesity, hypertension, coronary or valvular disease, heart failure,
10 metabolic syndrome, sleep apnea, excessive alcohol, and extreme exertion. Thus, in genetically
11 susceptible individuals, the imbalance in transcriptional responses to stress may predispose to AF
12 occurring later in life. In addition, the imbalance or overcompensation during response to injury,
13 initially adaptive, may eventually lead to fibrosis and adverse remodeling, with eventual
14 presentation of the disease. This concept may explain why AF tends to occur later in life as
15 evidenced by the top canonical pathways in this study, including pathways involved in
16 mitochondrial dysfunction, oxidative phosphorylation and sirtuin signaling. Metabolic stressors
17 and pathways likely play a strong role in AF and suggest that AF occurrence may reflect crossing
18 the threshold of cellular resilience and adaptation.

19

20 **Limitations and Future Studies**

21 IPA was initially designed for use in cancer data sets and relies upon curated pathways, thus our
22 results may be skewed by the composition of the pathways. Some of these pathways may not
23 accurately reflect the pathology of cardiovascular diseases. Nevertheless, many of the top

1 canonical pathways seem relevant to AF. Pathway analyses may oversimplify mechanisms.
2 Functional studies, including overexpression and knockdown models, are required to establish
3 causation. However, these analyses are intended to help guide such studies toward relevant
4 candidate genes and pathways. Future studies are planned or underway for several of the targets
5 identified. Our samples were from the LAA, rather than the PVs. However, PV tissue is difficult
6 to obtain, and the LAA is directly adjacent to the left PVs and reflects atrial function in the
7 human adult. The left atria obtained for this study were from patients that required open heart
8 surgery, which suggests significant disease comorbidity. Here, we attempted to adjust for
9 heterogeneity of comorbidities with use of surrogate variable adjustments.

10

11 **Conclusions**

12 In conjunction with recent GWAS meta-analyses demonstrating 135 loci associated with risk for
13 AF, our genetic-genomic analyses in human adult LA tissue have identified 233 candidate genes,
14 of which 131 had significant eQTLs. Coexpression analyses of these gene sets identified stress
15 pathways that are implicated in AF risk. These include mitochondrial dysfunction, oxidative
16 stress, sirtuin signaling, EIF2 signaling, and protein ubiquitination pathways. WGCNA modules,
17 and specifically those with AF candidate genes over representation, suggest remodeling in
18 response to injury may be occurring through changes in metabolomics, myogenesis and
19 apoptosis. We propose a dual risk model for AF, where genes active during cardiac development
20 contribute to genetic susceptibility to AF, for example by alterations in PV development that
21 leads to PV triggers. In genetically susceptible individuals, development of stresses later in life,
22 such as those associated with aging or obesity, may lead to inadequate or overwhelmed
23 transcriptional responses to stress that can predispose to AF. Here we also identify multiple

1 genes not previously linked to AF. We have created a resource to help scientists elucidate
2 possible functions of genes through identification of coexpressed genes with each candidate AF
3 gene, and associated modules in the LAA via WGCNA. The components allow for better
4 elucidation of likely AF mechanisms and potential targets for upstream therapies.

5
6 **Sources of Funding:** National Institutes of Health grants R01 HL 090620 and R01 HL 111314;
7 American Heart Association Atrial Fibrillation Strategically Focused Research Network grant
8 18SFRN34110067, 18SFRN34170013, 18SFRN34140065, 18SFRN34170442; the NIH National
9 Center for Research Resources for Case Western Reserve University and Cleveland Clinic
10 Clinical and Translational Science Award UL1-RR024989, the Cleveland Clinic Department of
11 Cardiovascular Medicine philanthropy research funds, and the Tomsich Atrial Fibrillation
12 Research Fund. JH was supported by the National Institutes of Health training grant T32 GM
13 088088. JDS was supported by the Geoffrey Gund Endowed Chair for Cardiovascular Research.

14 **Disclosures:** None.

15
16
17 **References**

- 18
19 1. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial
20 fibrillation: a Global Burden of Disease 2010 Study. *Circulation* 2014;129(8):837-47. (In
21 eng). DOI: 10.1161/CIRCULATIONAHA.113.005119
22 CIRCULATIONAHA.113.005119 [pii].

- 1 2. Iwasaki YK, Nishida K, Kato T, Nattel S. Atrial fibrillation pathophysiology: implications
2 for management. *Circulation* 2011;124(20):2264-74. (In eng). DOI:
3 10.1161/CIRCULATIONAHA.111.019893.
- 4 3. Hsu J, Gore-Panter S, Tchou G, et al. Genetic Control of Left Atrial Gene Expression
5 Yields Insights into the Genetic Susceptibility for Atrial Fibrillation. *Circ Genom Precis*
6 *Med* 2018;11(3):e002107. (In eng). DOI: 10.1161/CIRCGEN.118.002107.
- 7 4. Liu X, Wang F, Knight AC, Zhao J, Xiao J. Common variants for atrial fibrillation: results
8 from genome-wide association studies. *Hum Genet* 2012;131(1):33-9. (In eng). DOI:
9 10.1007/s00439-011-1052-3.
- 10 5. Roselli C, Chaffin MD, Weng LC, et al. Multi-ethnic genome-wide association study for
11 atrial fibrillation. *Nat Genet* 2018;50(9):1225-1233. (In eng). DOI: 10.1038/s41588-018-
12 0133-9.
- 13 6. Nielsen JB, Thorolfsdottir RB, Fritsche LG, et al. Biobank-driven genomic discovery yields
14 new insight into atrial fibrillation biology. *Nat Genet* 2018;50(9):1234-1239. (In eng).
15 DOI: 10.1038/s41588-018-0171-3.
- 16 7. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of
17 conserved genetic modules. *Science* 2003;302(5643):249-55. (In eng). DOI:
18 10.1126/science.1087447.
- 19 8. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput
20 sequencing data. *Bioinformatics* 2015;31(2):166-9. (In eng). DOI:
21 10.1093/bioinformatics/btu638.

- 1 9. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis
2 tools for RNA-seq read counts. *Genome Biol* 2014;15(2):R29. (In eng). DOI: 10.1186/gb-
3 2014-15-2-r29.
- 4 10. Shah RD, Frot B, Thanei G-A, Meinshausen N. Right singular vector projection graphs:
5 fast high dimensional covariance matrix estimation under latent confounding. *Journal of*
6 *the Royal Statistical Society: Series B (Statistical Methodology)* 2020;82(2):361-389. DOI:
7 10.1111/rssb.12359.
- 8 11. Wang Y, Hicks SC, Hansen KD. Co-expression analysis is biased by a mean-correlation
9 relationship. *bioRxiv* 2020:2020.02.13.944777. DOI: 10.1101/2020.02.13.944777.
- 10 12. Russo PST, Ferreira GR, Cardozo LE, et al. CEMiTool: a Bioconductor package for
11 performing comprehensive modular co-expression analyses. *BMC Bioinformatics*
12 2018;19(1):56. (In eng). DOI: 10.1186/s12859-018-2053-1.
- 13 13. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP.
14 Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27(12):1739-40. (In
15 eng). DOI: 10.1093/bioinformatics/btr260.
- 16 14. Logan M, Pagán-Westphal SM, Smith DM, Paganessi L, Tabin CJ. The transcription factor
17 Pitx2 mediates situs-specific morphogenesis in response to left-right asymmetric signals.
18 *Cell* 1998;94(3):307-17. (In eng). DOI: 10.1016/s0092-8674(00)81474-9.
- 19 15. Kwok A, Zvetkova I, Virtue S, et al. Truncation of Pik3r1 causes severe insulin resistance
20 uncoupled from obesity and dyslipidaemia by increased energy expenditure. *Mol Metab*
21 2020:101020. (In eng). DOI: 10.1016/j.molmet.2020.101020.

- 1 16. Madhala-Levy D, Williams VC, Hughes SM, Reshef R, Halevy O. Cooperation between
2 Shh and IGF-I in promoting myogenic proliferation and differentiation via the MAPK/ERK
3 and PI3K/Akt pathways requires Smo activity. *J Cell Physiol* 2012;227(4):1455-64. (In
4 eng). DOI: 10.1002/jcp.22861.
- 5 17. Rommel C, Bodine SC, Clarke BA, et al. Mediation of IGF-1-induced skeletal myotube
6 hypertrophy by PI(3)K/Akt/mTOR and PI(3)K/Akt/GSK3 pathways. *Nat Cell Biol*
7 2001;3(11):1009-13. (In eng). DOI: 10.1038/ncb1101-1009.
- 8 18. Gong Y, Yang J, Liu Q, et al. IGF1 Knockdown Hinders Myocardial Development through
9 Energy Metabolism Dysfunction Caused by ROS-Dependent FOXO Activation in the
10 Chicken Heart. *Oxid Med Cell Longev* 2019;2019:7838754. (In eng). DOI:
11 10.1155/2019/7838754.
- 12 19. Sun XH, Wang X, Zhang Y, Hui J. Exosomes of bone-marrow stromal cells inhibit
13 cardiomyocyte apoptosis under ischemic and hypoxic conditions via miR-486-5p
14 targeting the PTEN/PI3K/AKT signaling pathway. *Thromb Res* 2019;177:23-32. (In eng).
15 DOI: 10.1016/j.thromres.2019.02.002.
- 16 20. Pires KM, Torres NS, Buffolo M, et al. Suppression of Cardiac Autophagy by
17 Hyperinsulinemia in Insulin Receptor-Deficient Hearts Is Mediated by Insulin-Like
18 Growth Factor Receptor Signaling. *Antioxid Redox Signal* 2019;31(6):444-457. (In eng).
19 DOI: 10.1089/ars.2018.7640.
- 20 21. Ren N, Wang M. microRNA-212-induced protection of the heart against myocardial
21 infarction occurs via the interplay between AQP9 and PI3K/Akt signaling pathway. *Exp*
22 *Cell Res* 2018;370(2):531-541. (In eng). DOI: 10.1016/j.yexcr.2018.07.018.

- 1 22. Zhu HH, Wang XT, Sun YH, et al. MicroRNA-486-5p targeting PTEN Protects Against
2 Coronary Microembolization-Induced Cardiomyocyte Apoptosis in Rats by activating the
3 PI3K/AKT pathway. *Eur J Pharmacol* 2019;855:244-251. (In eng). DOI:
4 10.1016/j.ejphar.2019.03.045.
- 5 23. McMullen JR, Boey EJ, Ooi JY, Seymour JF, Keating MJ, Tam CS. Ibrutinib increases the
6 risk of atrial fibrillation, potentially through inhibition of cardiac PI3K-Akt signaling.
7 *Blood* 2014;124(25):3829-30. (In eng). DOI: 10.1182/blood-2014-10-604272.
- 8 24. Leong DP, Caron F, Hillis C, et al. The risk of atrial fibrillation with ibrutinib use: a
9 systematic review and meta-analysis. *Blood* 2016;128(1):138-40. (In eng). DOI:
10 10.1182/blood-2016-05-712828.
- 11 25. Lang T, Yu L, Tu Q, et al. Molecular cloning, genomic organization, and mapping of
12 PRKAG2, a heart abundant gamma2 subunit of 5'-AMP-activated protein kinase, to
13 human chromosome 7q36. *Genomics* 2000;70(2):258-63. (In eng). DOI:
14 10.1006/geno.2000.6376.
- 15 26. Hinson JT, Chopra A, Lowe A, et al. Integrative Analysis of PRKAG2 Cardiomyopathy iPS
16 and Microtissue Models Identifies AMPK as a Regulator of Metabolism, Survival, and
17 Fibrosis. *Cell Rep* 2016;17(12):3292-3304. (In eng). DOI: 10.1016/j.celrep.2016.11.066.
- 18 27. Arad M, Benson DW, Perez-Atayde AR, et al. Constitutively active AMP kinase mutations
19 cause glycogen storage disease mimicking hypertrophic cardiomyopathy. *J Clin Invest*
20 2002;109(3):357-62. (In eng). DOI: 10.1172/JCI14571.
- 21 28. Matsui T, Nagoshi T, Rosenzweig A. Akt and PI 3-kinase signaling in cardiomyocyte
22 hypertrophy and survival. *Cell Cycle* 2003;2(3):220-3. (In eng).

- 1 29. Gore-Panter SR, Hsu J, Hanna P, et al. Atrial Fibrillation associated chromosome 4q25
2 variants are not associated with PITX2c expression in human adult left atrial
3 appendages. *PLoS One* 2014;9(1):e86245. (In eng). DOI: 10.1371/journal.pone.0086245.
- 4 30. Zhang Z, Wei S, Du H, et al. Zfhx3 is required for the differentiation of late born D1-type
5 medium spiny neurons. *Exp Neurol* 2019;322:113055. (In eng). DOI:
6 10.1016/j.expneurol.2019.113055.
- 7 31. Tan KP, Kosuge K, Yang M, Ito S. NRF2 as a determinant of cellular resistance in retinoic
8 acid cytotoxicity. *Free Radic Biol Med* 2008;45(12):1663-73. (In eng). DOI:
9 10.1016/j.freeradbiomed.2008.09.010.
- 10 32. Hu Q, Zhang B, Chen R, et al. ZFH3 is indispensable for ER β to inhibit cell proliferation
11 via MYC downregulation in prostate cancer cells. *Oncogenesis* 2019;8(4):28. (In eng).
12 DOI: 10.1038/s41389-019-0138-y.
- 13 33. Kao YH, Hsu JC, Chen YC, et al. ZFH3 knockdown increases arrhythmogenesis and
14 dysregulates calcium homeostasis in HL-1 atrial myocytes. *Int J Cardiol* 2016;210:85-92.
15 (In eng). DOI: 10.1016/j.ijcard.2016.02.091.
- 16 34. Chen WK, Kuo WW, Hsieh DJ, et al. CREB Negatively Regulates IGF2R Gene Expression
17 and Downstream Pathways to Inhibit Hypoxia-Induced H9c2 Cardiomyoblast Cell Death.
18 *Int J Mol Sci* 2015;16(11):27921-30. (In eng). DOI: 10.3390/ijms161126067.
- 19 35. Musi N, Yu H, Goodyear LJ. AMP-activated protein kinase regulation and action in
20 skeletal muscle during exercise. *Biochem Soc Trans* 2003;31(Pt 1):191-5. (In eng). DOI:
21 10.1042/bst0310191.

- 1 36. Minokoshi Y, Kim YB, Peroni OD, et al. Leptin stimulates fatty-acid oxidation by
2 activating AMP-activated protein kinase. *Nature* 2002;415(6869):339-43. (In eng). DOI:
3 10.1038/415339a.
- 4 37. Sun MW, Lee JY, de Bakker PI, et al. Haplotype structures and large-scale association
5 testing of the 5' AMP-activated protein kinase genes PRKAA2, PRKAB1, and PRKAB2
6 [corrected] with type 2 diabetes. *Diabetes* 2006;55(3):849-55. (In eng). DOI:
7 10.2337/diabetes.55.03.06.db05-1418.
- 8 38. Viollet B, Andreelli F, Jørgensen SB, et al. Physiological role of AMP-activated protein
9 kinase (AMPK): insights from knockout mouse models. *Biochem Soc Trans* 2003;31(Pt
10 1):216-9. (In eng). DOI: 10.1042/bst0310216.
- 11 39. Zhang P, Hu X, Xu X, et al. AMP activated protein kinase- α 2 deficiency exacerbates
12 pressure-overload-induced left ventricular hypertrophy and dysfunction in mice.
13 *Hypertension* 2008;52(5):918-24. (In eng). DOI:
14 10.1161/HYPERTENSIONAHA.108.114702.
- 15 40. Cao H, Wang X, Ying S, Huang C. AMPK α 2 deficiency enhanced susceptibility to
16 ventricular arrhythmias in mice by the role of β -adrenoceptor signaling. *Exp Biol Med*
17 (Maywood) 2018;243(8):708-714. (In eng). DOI: 10.1177/1535370218767389.
- 18 41. Chang SH, Wu LS, Chiou MJ, et al. Association of metformin with lower atrial fibrillation
19 risk among patients with type 2 diabetes mellitus: a population-based dynamic cohort
20 and in vitro studies. *Cardiovasc Diabetol* 2014;13:123. (In eng). DOI: 10.1186/s12933-
21 014-0123-x.

- 1 42. Konstantinidis K, Whelan RS, Kitsis RN. Mechanisms of cell death in heart disease.
2 Arterioscler Thromb Vasc Biol 2012;32(7):1552-62. (In eng). DOI:
3 10.1161/ATVBAHA.111.224915.
- 4 43. Feng N, Anderson ME. CaMKII is a nodal signal for multiple programmed cell death
5 pathways in heart. J Mol Cell Cardiol 2017;103:102-109. (In eng). DOI:
6 10.1016/j.yjmcc.2016.12.007.
- 7 44. Anand IS, Gupta P. Anemia and Iron Deficiency in Heart Failure: Current Concepts and
8 Emerging Therapies. Circulation 2018;138(1):80-98. (In eng). DOI:
9 10.1161/CIRCULATIONAHA.118.030099.
- 10 45. Nomani H, Bayat G, Sahebkar A, et al. Atrial fibrillation in β -thalassemia patients with a
11 focus on the role of iron-overload and oxidative stress: A review. J Cell Physiol
12 2019;234(8):12249-12266. (In eng). DOI: 10.1002/jcp.27968.
- 13 46. Hoes MF, Grote Beverborg N, Kijlstra JD, et al. Iron deficiency impairs contractility of
14 human cardiomyocytes through decreased mitochondrial function. Eur J Heart Fail
15 2018;20(5):910-919. (In eng). DOI: 10.1002/ejhf.1154.
- 16 47. Kanaan GN, Patten DA, Redpath CJ, Harper ME. Atrial Fibrillation Is Associated With
17 Impaired Atrial Mitochondrial Energetics and Supercomplex Formation in Adults With
18 Type 2 Diabetes. Can J Diabetes 2019;43(1):67-75.e1. (In eng). DOI:
19 10.1016/j.jcjd.2018.05.007.
- 20 48. Wiersma M, Meijering RAM, Qi XY, et al. Endoplasmic Reticulum Stress Is Associated
21 With Autophagy and Cardiomyocyte Remodeling in Experimental and Human Atrial
22 Fibrillation. J Am Heart Assoc 2017;6(10) (In eng). DOI: 10.1161/JAHA.117.006458.

- 1 49. Haushalter KJ, Schilling JM, Song Y, et al. Cardiac ischemia-reperfusion injury induces
2 ROS-dependent loss of PKA regulatory subunit R1 α . *Am J Physiol Heart Circ Physiol*
3 2019;317(6):H1231-H1242. (In eng). DOI: 10.1152/ajpheart.00237.2019.
- 4 50. Kamga Pride C, Mo L, Quesnelle K, et al. Nitrite activates protein kinase A in normoxia to
5 mediate mitochondrial fusion and tolerance to ischaemia/reperfusion. *Cardiovasc Res*
6 2014;101(1):57-68. (In eng). DOI: 10.1093/cvr/cvt224.
- 7 51. Chung MK, Youn Wass SJ, Barnard J. Insights From Atrial Fibrillation Genomics: From
8 Bedside to Bench and Back Again. *Cardiol Rev* 2019;27(6):302-307. (In eng). DOI:
9 10.1097/CRD.0000000000000267.
- 10 52. Deshmukh A, Barnard J, Sun H, et al. Left atrial transcriptional changes associated with
11 atrial fibrillation susceptibility and persistence. *Circ Arrhythm Electrophysiol*
12 2015;8(1):32-41. (In eng). DOI: 10.1161/CIRCEP.114.001632.

13
14
15
16
17

1 **Figure Legends**

2 **Figure 1. WGCNA correlation plot.** WGCNA identified 64 distinct modules, which included a
3 ‘non correlated’ module. 8 of the 48 modules had over-representation of AF genes in their
4 respective modules. Hierarchical clustering of the modules show clustering of AF
5 overrepresented modules 25, 39 and 42, and clustering of modules 11, 21, and ‘not correlated’.
6 The size and shade of circle reflect strength of correlation, and color shows direction of
7 correlation.

8

9 **Figure 2. PITX2 and coexpressed genes signaling pathways.** Top canonical pathways include
10 protein handling (e.g. β - alanine degradation, valine, 4-aminobutyrate and glutamine degradation)
11 and and major signaling pathways (e.g. PI3K/AKT, G protein coupled receptor signaling, IGF-1,
12 PDGF signaling, renin-angiotensin and AMPK signaling pathways). PIK3 is critical in insulin
13 resistance and in the regulation of myogenic differentiation, myoblast proliferation, and myotube
14 hypertrophy. Activation of the PI3K-Akt signaling pathway, mediated through IGF-1, decreases
15 autophagy and apoptosis. AMPK activity and the resultant glycogen accumulation is correlated
16 with changes in metabolic pathways and Akt phosphorylation.

17

18 **Figure 3. ZHFX3 and coexpressed genes signaling pathways.** *RAS*, *PI3K* and *CREB* have
19 been linked to modulation of cardiomyocyte apoptosis.

20

21 **Figure 4. Dual Risk Model for AF.** Genes active during cardiac development may increase
22 genetic susceptibility to AF. Stressors, such as aging and obesity, may lead to increased cellular

1 stress, including metabolic, mitochondrial, oxidative, and unfolded protein stress. The
2 combination of genetic developmental susceptibility and accumulated stressors may lead to
3 overwhelmed or inadequate transcriptional responses which predisposes to AF occurrence.

4

1 **Table 1. Candidate AF genes.** Shaded headings indicate the 3 targeted groups of study genes.
 2 Genes are listed in order of statistical significance of eQTL SNPs.

All Candidate Genes (n=233) <i>Predicted by AF GWAS loci with significant cis-eQTL, or prioritized by Nielsen⁶, et al</i>		
Any eQTL Genes (n= 131)		
Top GWAS SNP eQTL Genes (n=37) <i>Ranked top 10</i>	Other left atrial eQTL Genes (n=94) <i>with cis-eQTL q <0.05</i>	Remaining Nielsen prioritized genes <i>without significant cis-eQTL(n=102)</i>
<i>GNB4(1), MYOZ1(1), ARNT2(1), AC016705.2 (1), CTXND1(1), SYNPO2L(1), HEATR2(1), TUBA8(1), RP11-1082L8.1 (1), CASQ2(1), AC100858 (2), ENSG00000269728 – retired (2), WIPF1(2), XYLT1(2), ASAH1(2), SCN10A(2), LINC01426(2), ERBB4(2), LINC01629 (3), AC007620.2(3), SYNE2 (3), CAV1(3), DCST1-AS1(3), RN7SL356P(4), KCND3(4), THRB(4), SSPN(4), CEP68(5), MTSS1(6), PKP2(6), AC113382.1(7), SCN5A (7), GJA5(7), KCNJ5 (8), C21orf2(10), AC113382.2 (10), TBX5(10)</i>	<i>SLC2A9, FAM13B, ADORA1, ENSG00000255864 (retired), GMCL1, CMTM5, AP001062.7, IL25, EXT1, ABHD17C, MYOG, C5orf47, IL6R- ASI, IL6R, METTL1B, KDM1B, LINC00964, USP36, LRCH1, HAND2- ASI, PTK2, TPMT, USP3-ASI, AC083843.3, SLC25A26, IGF1R, DNAH10OS, ZNF664, OR10P1, PCMI, ENSG00000270028 (retired), RAB7L1, ERBB2, GYG1, GSDMB, STAG3L2, PRRX1, CCDC92, PROZ, SPATS2L, COG5, DEK, PCCB, MEX3C, PHLDB2, SLC27A6, MRPL35P2, C11orf45, WNT3, MSL1, FBXO32, CAV2, NSG2, MTHFD1, AC098798.1, SCMHI, HSPG2, NSF, GET4, P2RX2, CUL4A, FKBP7, DMTN, F10, PFDN1, DPF3, SNRNP27, RNF144B, SAT2, PM20D1, C2orf74, USP34, FBXL22, KCNN2, MYRFL, YWHAE, ARL17A, SMIM8, CTPS1, THAP9-ASI, PNMT, AHSA2, PGAP3, GSDMA, AC027117.1, KIAA1841, ZBTB25, MED24, LRRC37A2, RAB3IP, HLA-B, MICA, PSORS1C1, HLA-C</i>	<i>SLC9B1, ARHGAP10, C10orf11, NAV2, RBM20, HIP1R, C9orf3, AKAP6, AGBL4, NEURL1, UST, RPL3L, SLIT3, PHLDA1, HSF2, MYO18B, FGF5, MIR6500, GCOM1/MYZAP, FRMD4B, LHX3, ZFH3, GYPC, CASZ1, CDKN1A, PANDAR, PI16, GATA4, PLN, SLC35F1, SORL1, DGKB, CALU, OPN1SW, CDK6, CREB5, NR3C1, LOC102467213, EPHA3, TEX41, CAMK2D, MYPN, SIRT1, MYOCD, REEP1, CFL2, ATXN1, HCN4, KIF3C, CAND2, KCNH2, LINC00540, LINC00621, SGCG, SMAD7, HAND2, KCNN3, PITX2, NACA, LINC00477, NKX2-5, LINC01142, LRRC10, AGAP5, NUDT13, IRF2BPL, LINC01426, ESR2, MIR548AZ, MYOT, NPY6R, WNT8A, PPFIA4, ANXA4, MYH6, MYH7, CYTH1, LRIG1, HERC1, NUCKS1, SLC41A1, ORMDL3, ZPBP2, GTF2I, GTF2IRD2, LOC101926943, PPP2R3A, PLCXD2, NRBF2, REEP3, FBRSL1, MIR548N, TTN, TTN-ASI, XPO7, FXR2, SOX15, TNFSF12, TNFSF12-TNFSF13, CRK, MYO1C, CGA</i>

3 AF = atrial fibrillation; *cis*-eQTL = cis expression quantitative trait loci; GWAS = genome wide
 4 association study. (in parentheses): rank of GWAS SNP among the top 10 eQTL SNPs.

5
 6

1 **Table 2. Genes identified by significant left atrial cis-eQTLs but not prioritized by Nielsen,**
2 **et al⁶.**

HEATR2, GET4, LINC01314, RP11-210M15.2, RP11-1082L8.1, MTSSI, LINC00964, RP11-145M9.4, AC007620.3, RP11-7F17.1, RP11-7F17.7, RN7SL356P, PCM1, AP000330.8, l. RP11-806O11.1, RP11-307C12.11, AC113382.1, AC083843.1, AC098798.1, ZBTB25, RP11-325L7.2, AP001062.7, RP11-444D3.1, NSG2, IL6R-AS1, C11orf45, C21orf2, SLC2A9, ADORA1, MYOG, SNRNP27, CMTM5, IL25, EXT1, ABHD17C, C5orf47, IL6R, METTL11B, TPMT, RNF144B, LRCH1, USP3-AS1, FBXL22, DNAH10OS, ZNF664, CCDC92, OR10P1, RAB7L1, PM20D1, ERBB2, MSL1, PNMT, PGAP3, GSDMA, MED24, GYG1, STAG3L2, PRRX1, PROZ, F10, COG5, PCCB, MRPL35P2, NSF, ARL17A, LRRC37A2, CTPSI, P2RX2, DMTN, PFDN1, SAT2, C2orf74, AHS2, KIAA1841, MYRFL, RAB3IP, SMIM8, THAP9-AS1, HLA-B, MICA, PSORS1C1, HLA-C, ENSG00000270028 (retired)

3
4
5

1 **Table 3.**

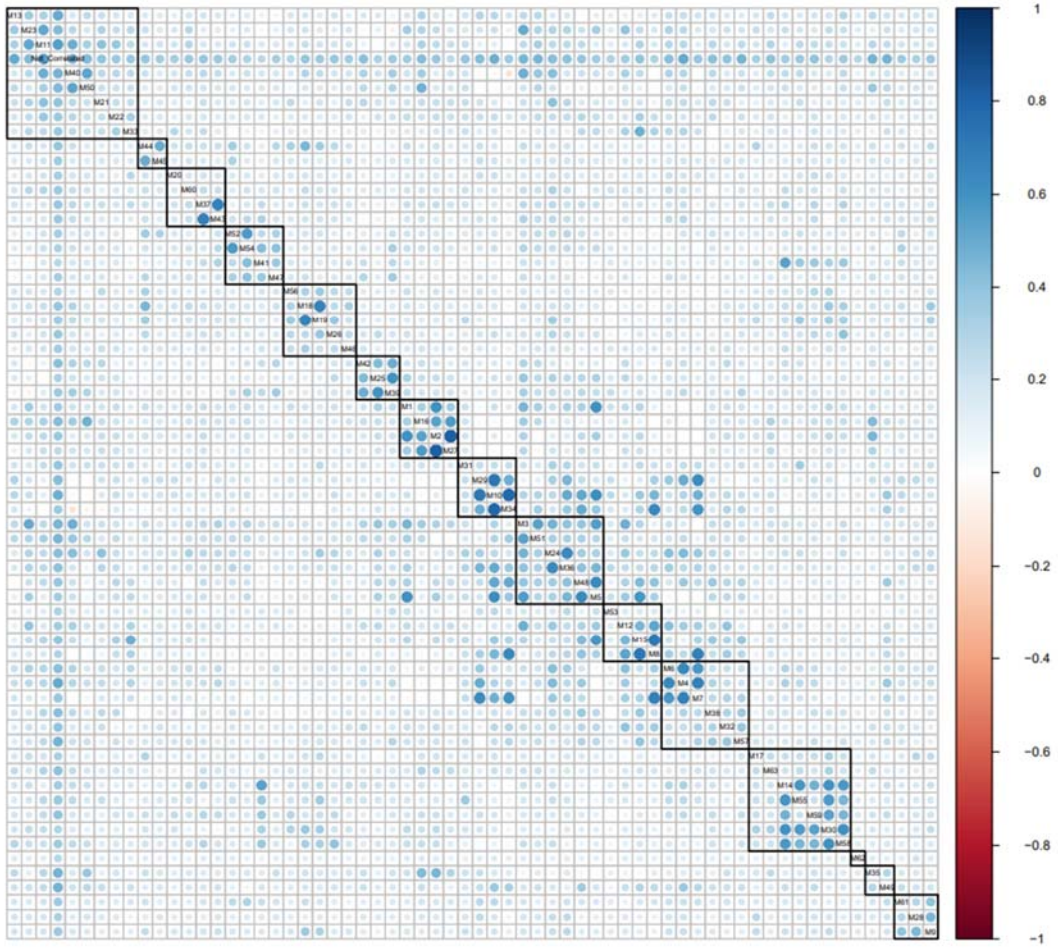
Modules with AF gene overrepresentation ($p < 0.05$)*	Total number of genes in module	# AF genes in module	Hub genes of module	Pathways identified by IPA and GSEA
M21	236	9 (<i>PRRX1HSPG2</i> <i>SLIT3</i> <i>EXT1</i> <i>TBX5</i> <i>HIP1R</i> <i>METTL11B</i> <i>PHLDA1</i> <i>NAV2</i>)	<i>ABTB2</i> ; <i>SLC8A3</i>	Intracellular ion homeostasis Apelin Cardiomyocyte Signaling Pathway
M2	1082	23 (<i>DMTN</i> <i>FAM13B</i> <i>KIAA1841</i> <i>CEP68</i> <i>THAP9-AS1</i> <i>FRMD4B</i> <i>MYH6</i> <i>KCND3</i> <i>XPO7</i> <i>FBXO32</i> <i>SGCG</i> <i>ATXN1</i> <i>PPP2R3A</i> <i>CDK6</i> <i>SORL1</i> <i>MYOCD</i> <i>KCNN3</i> <i>CUL4A</i> <i>IGF1R</i> <i>ZFHX3</i> <i>NR3C1</i> <i>TTN</i>)	<i>ARHGEF12</i> ; <i>PIKFYVE</i> ; <i>KIF1B</i> ; <i>TEAD1</i> ; <i>ICE1</i> ; <i>PBRM1</i> ; <i>UBR2</i> ; <i>NCOA2</i> ; <i>ARFGEF2</i> ; <i>DENND5B</i>	Heme metabolism Protein Kinase A Signaling
Not Correlated	1468	18 (<i>LINC01629</i> <i>AC007686.1</i> <i>TNFSF12-TNFSF13</i> <i>COG5</i> <i>GTF2IRD2</i> <i>C2orf74</i> <i>ARL17A</i> <i>ARNT2</i> <i>MICA</i> <i>NACA</i> <i>RAB31P</i> <i>AC016705.2</i> <i>CCDC92</i> <i>LINC01426</i> <i>LINC01426</i> <i>DGKB</i> <i>DNAH100S</i> <i>KIF3C</i>)		
M25	230	7 (<i>C9orf3</i> <i>MYO1C</i> <i>MYO18B</i> <i>RBM20</i> <i>CAV1</i> <i>CFL2</i> <i>MYPN</i>)	<i>DNAJB5</i> ; <i>SLMAP</i>	Myogenesis Remodeling of Epithelial Adherens Junctions Gap Junction Signaling
M11	361	9 (<i>SYNE2</i> <i>IL6R</i> <i>LRRC10</i> <i>MTSS1</i> <i>PTK2</i> <i>ERBB2</i> <i>ESR2</i> <i>ADORA1</i>)	<i>RCOR2</i> ; <i>ADAMTS7</i> ; <i>CXXC4</i> ; <i>TESK2</i>	Ephrin A Signaling

		<i>HCN4</i>		
M42	148	5 (<i>DNAAF5</i> <i>USP3-AS1</i> <i>FBXL22</i> <i>NEURL1</i> <i>AC083843.2</i>)	<i>PPP1R3B</i>	ILK Signaling Integrin Signaling Actin Nucleation by ARP-WASP Complex
M39	156	5 (<i>CMTM5</i> <i>HLA-C</i> <i>SMIM8</i> <i>CASZ1</i> <i>CREB5</i>)	<i>AC083902.2; ABRA</i>	TNFA signaling via NFKB Endoplasmic Reticulum Stress Pathway Protein Kinase A Signaling
M24	232	6 (<i>CTXND1</i> <i>ORMDL3</i> <i>OPN1SW</i> <i>PM20D1</i> <i>ANXA4</i> <i>CALU</i>)	<i>LRR59; SEC13</i>	Protein Secretion Unfolded protein response Gap Junction Signaling Remodeling of Epithelial Adherens Junctions

*listed in descending order of statistical significance

1
2
3

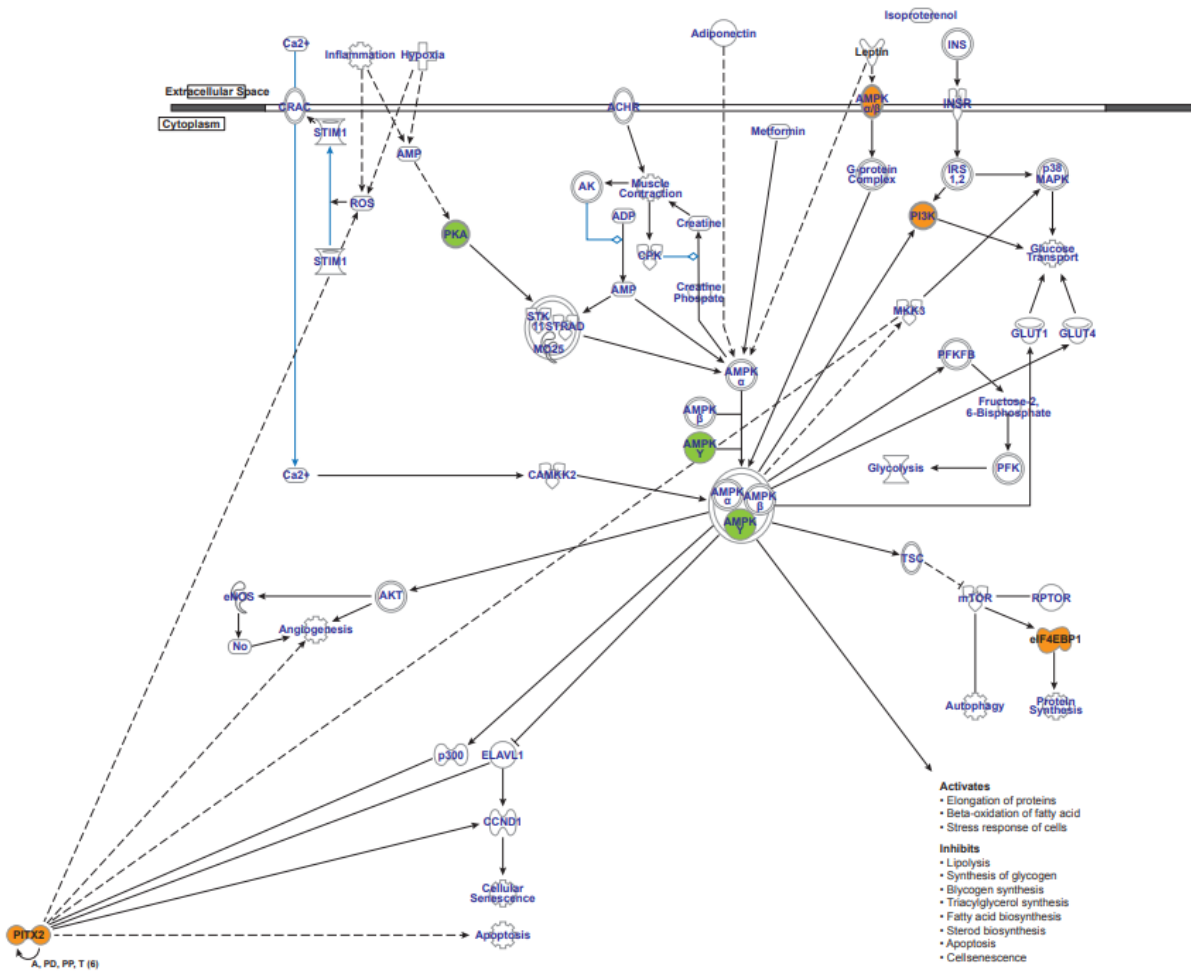
1



2
3
4
5

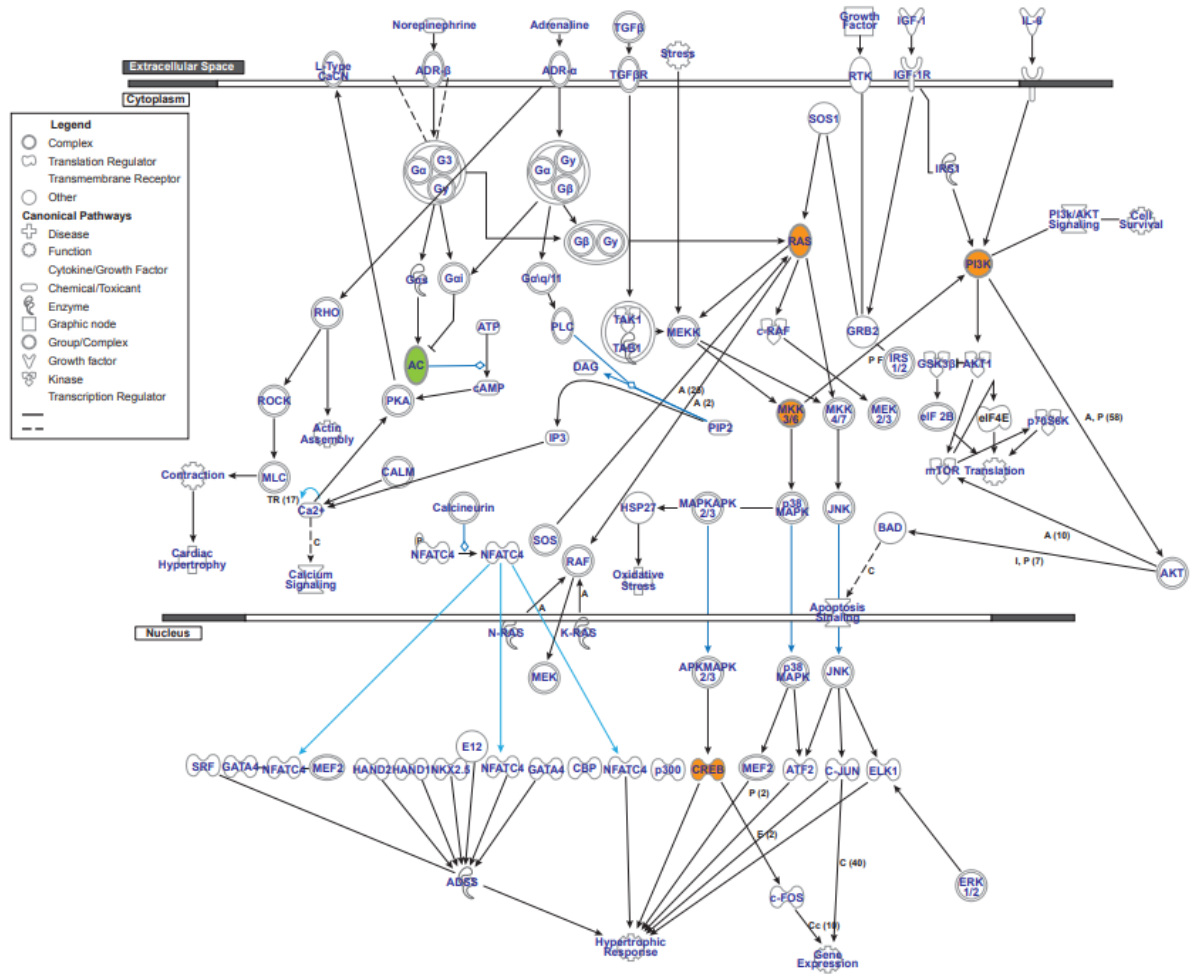
Figure 1. WGCNA correlation plot

1



2
3
4

Figure 2. *PITX2* and coexpressed genes signaling pathways

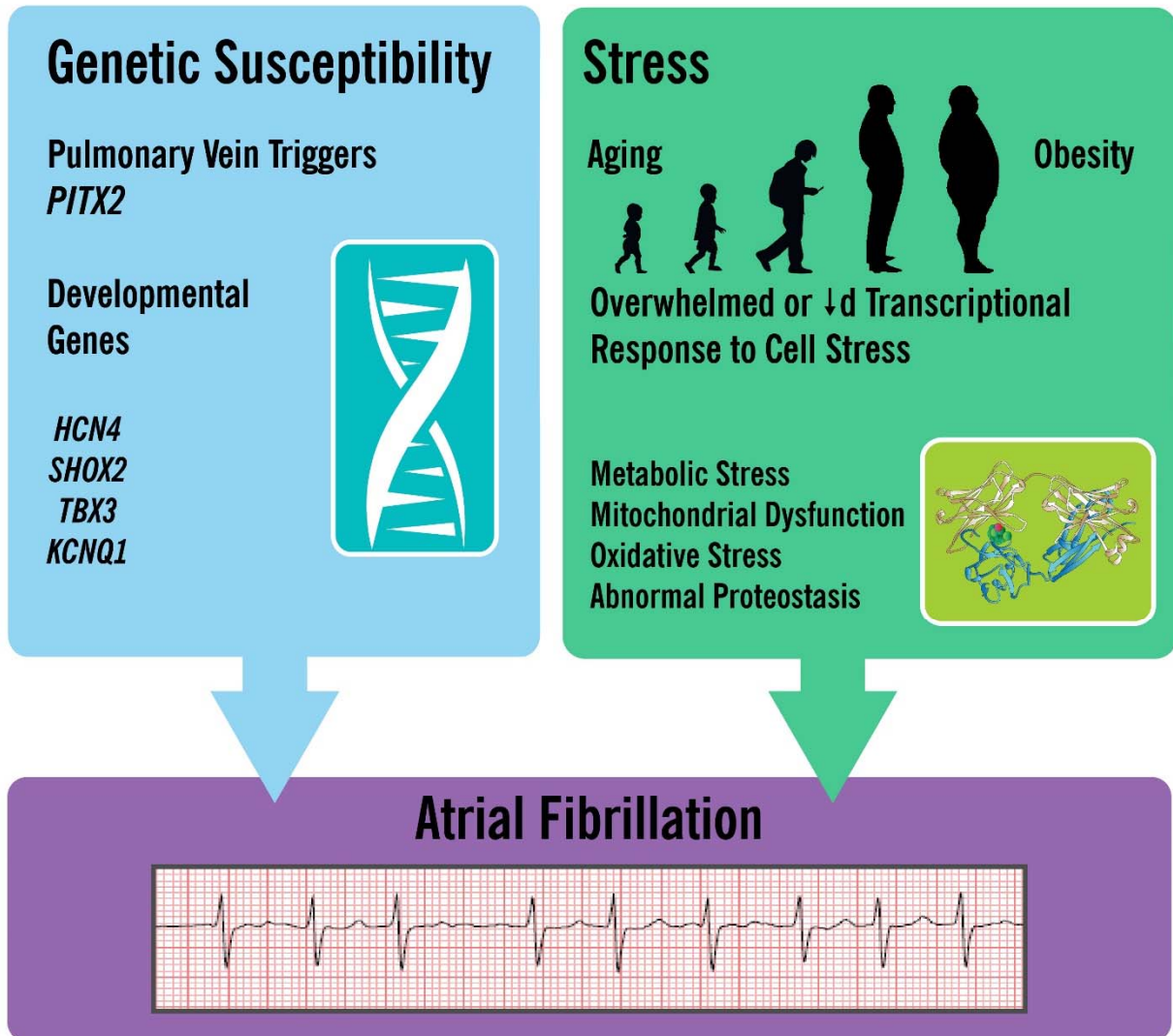


1

2 **Figure 3. *ZHFX3* and coexpressed genes signaling pathways**

Dual Risk Model

In Genetically Susceptible Patients, Inadequate Transcriptional Responses to Stress Predispose to AF in Later Life



1
2
3
4

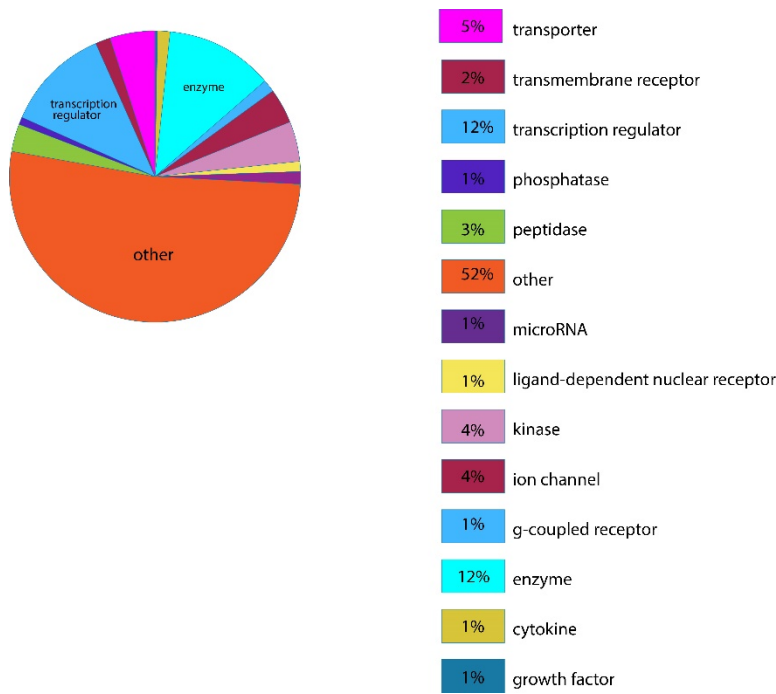
Figure 4. Dual Risk Model for AF.

1 **SUPPLEMENTAL MATERIALS**

- Supplemental Figure 1.** All Candidate genes by function and location
Supplemental Figure 2. Most commonly coexpressed non-GWAS genes
- Supplemental Figure 3.** Top significant IPA® canonical pathways for the coexpression networks of each candidate AF gene group
- Supplemental Table 1.** Top pathways and upstream regulators within each AF candidate gene group
- Supplemental Excel 1.** Coexpression partners of 225 candidate AF genes.
Supplemental Excel 2. WGCNA modules- LA genes and hub genes
Supplemental Excel 3. All significant top canonical pathways for the coexpression networks of the All Candidate Genes (n=233 genes)
Supplemental Excel 4. All significant top canonical pathways for the coexpression networks of the Any eQTL Genes (n=131 genes).
Supplemental Excel 5. All significant top canonical pathways for the coexpression networks of the Top GWAS SNP eQTL Genes (n=37 genes).
Supplemental Excel 6. IPA analysis results of WGCNA modules
Supplemental Excel 7. GSEA of WGCNA Modules
Supplemental Excel 8. eQTL GWAS SNPs
Supplemental Excel 9. Top 5 significant canonical pathways for each candidate AF gene and coexpression partners.
- Supplemental Pdfs** IPA analysis results of candidate AF gene and coexpression partners

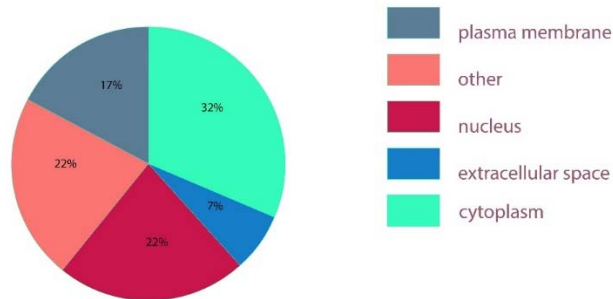
2

1 A.



2

3 B.



4

5 **Supplemental Figure 1. All Candidate genes by function and location. A. Categories of**

6 **Gene Function.** Of the 233 candidate atrial fibrillation genes 12% encode transcription

7 regulators, 12% enzymes, 5% transporters, 4% ion channels, 4% kinases and 3% peptidases. **B.**

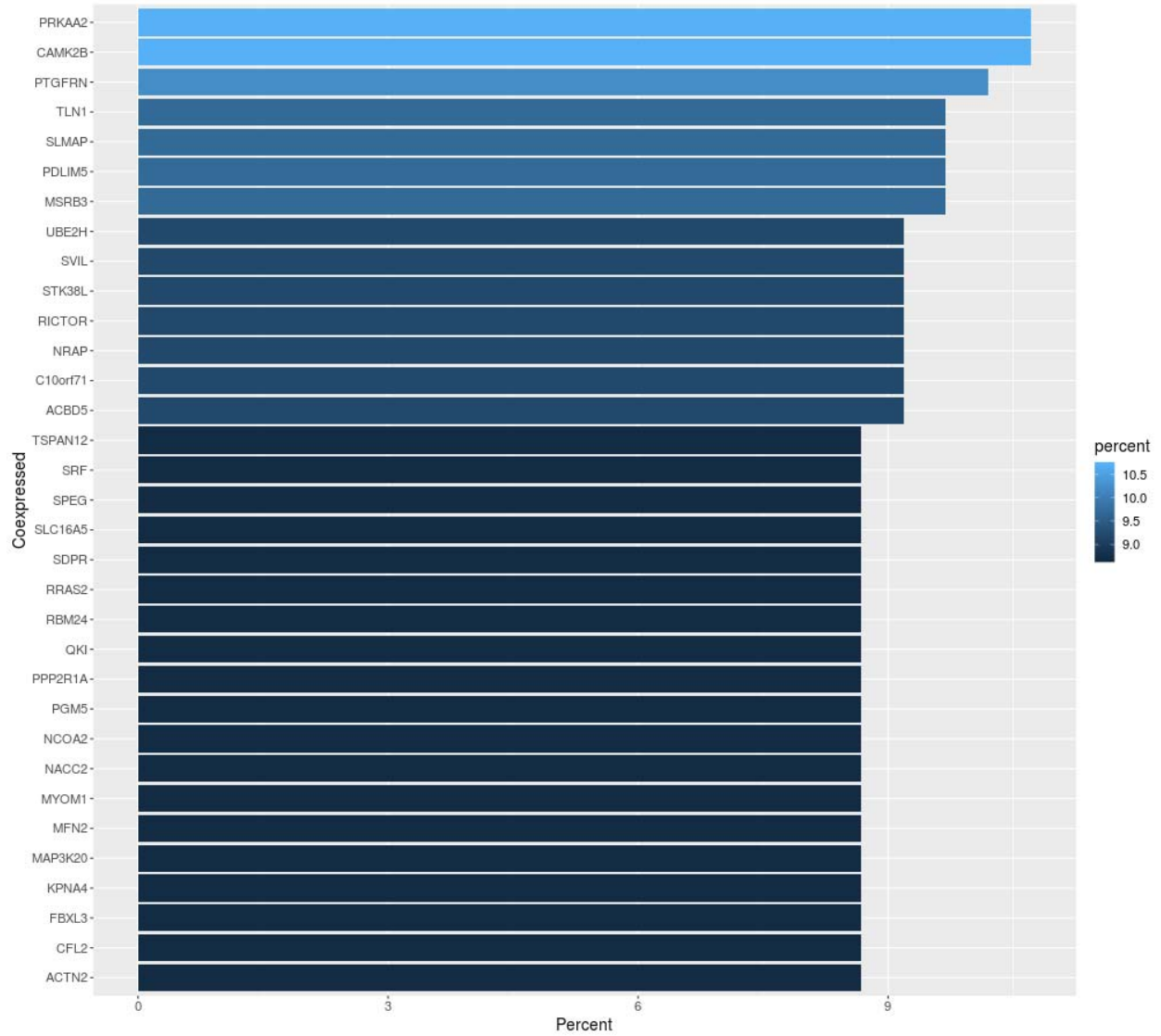
8 **Categories of Gene Location.** Of the 233 candidate atrial fibrillation genes 32% are in the

9 plasma membrane, 22% in the nucleus, 17% in the cytoplasm and 7% in the extracellular space.

10

11

12



1

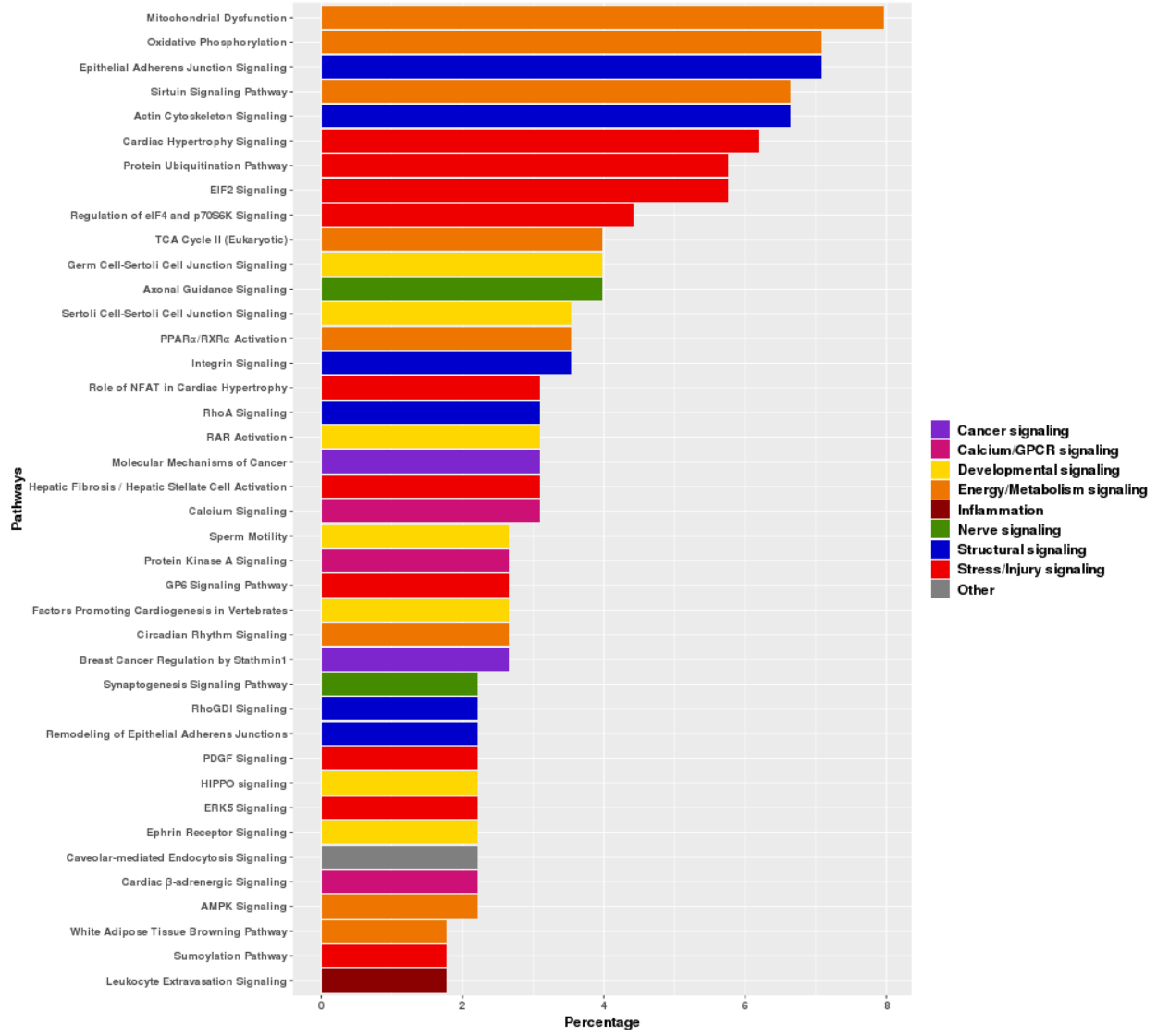
2 **Supplemental Figure 2. Most commonly coexpressed non-GWAS genes.** *PRKAA2* and

3 *CAMK2B* are non AF GWAS genes that are most frequently coexpressed with AF GWAS

4 candidate genes, followed by *PTGFRN*, *TLN1*, *SLMAP*, *PDLIM5* and *MSRB3*.

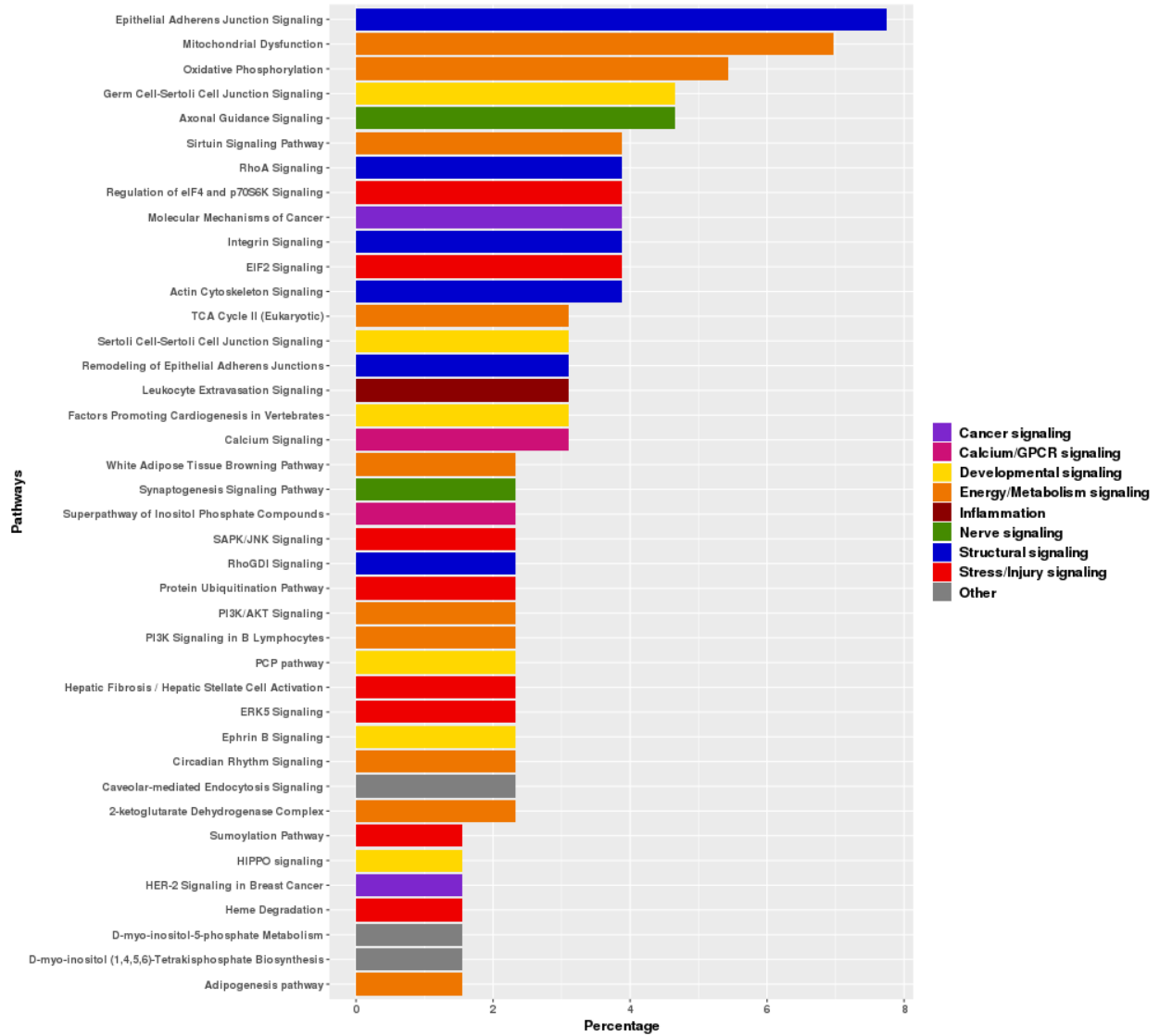
5

1
2 A.



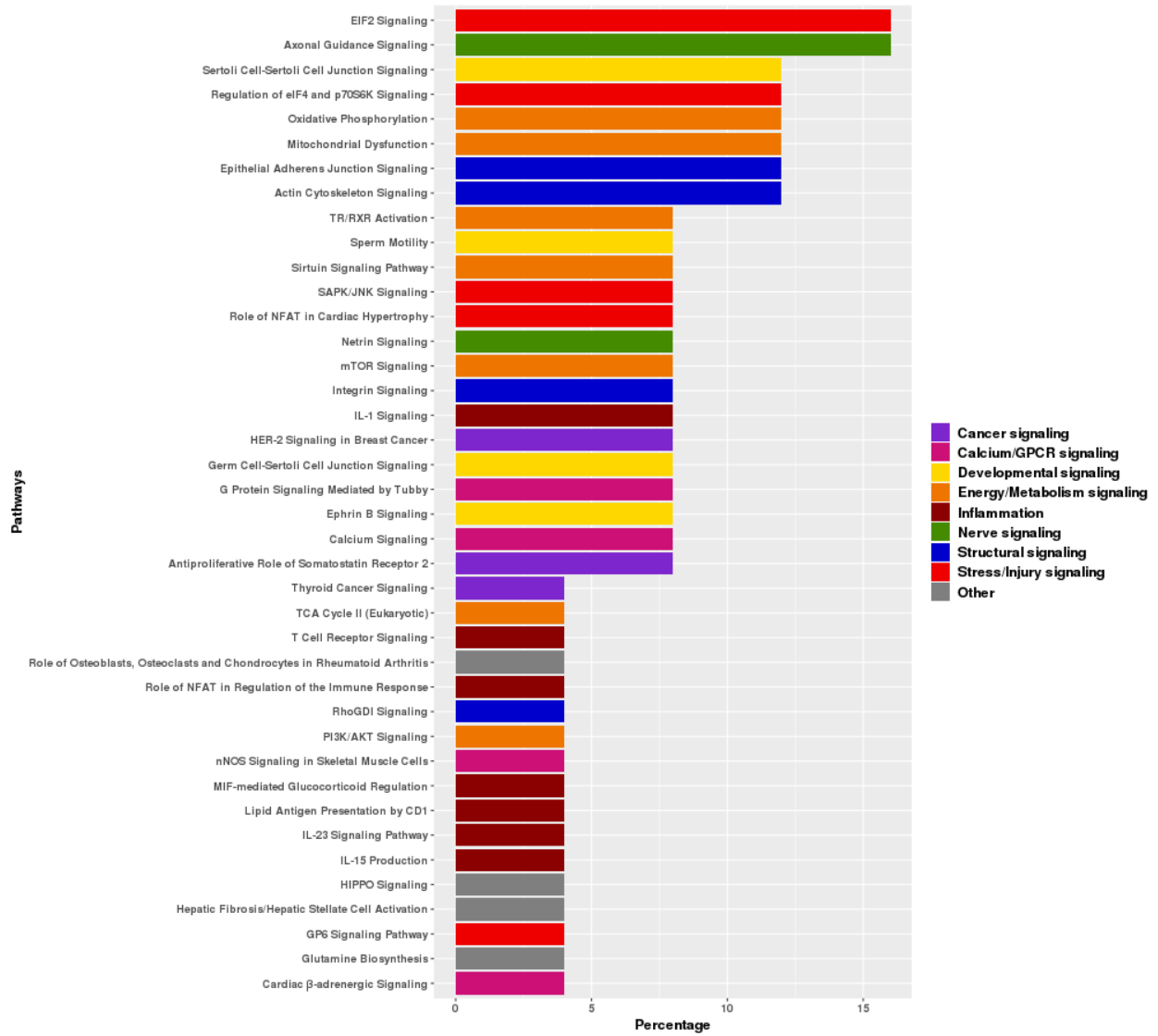
3
4

1 B.



2
3

1 C.



2

1 **D.**

	All Candidate Genes	All eQTL Genes	Top GWAS SNP eQTLGenes
Cancer Signaling	2	2	3
Calcium/GPCR signaling	2	3	4
Developmental signaling	6	7	4
Energy/Metabolism signaling	10	8	7
Inflammation	1	1	7
Nerve signaling	2	2	2
Structural signaling	6	6	4
Stress/Injury signaling	8	10	5
Other	3	1	4

2

3 **Supplemental Figure 3. Top significant IPA® canonical pathways for the coexpression**

4 **networks of each candidate AF gene list.** IPA canonical pathways for each candidate gene

5 were ranked from lowest to highest p-value and the top 5 significant canonical pathways listed

6 for each gene (See Supplemental Excel file 8, Supplemental Figure 3-5). Displayed here is the

7 percentage of genes for which canonical pathways were listed in the top 5. Data are displayed by

8 gene lists. **A. All Candidate Genes** (N=225; data not available for 8 genes, see text).

9 Mitochondrial dysfunction, oxidative phosphorylation, epithelial adherens junction signaling,

10 sirtuin signaling pathway and actin cytoskeleton pathway were the most frequent pathways listed

11 among the top 5 canonical pathways. **B. Any eQTL Genes** (n=131). Epithelial adherens junction

12 signaling, mitochondrial dysfunction and oxidative phosphorylation, germ cell-Sertoli cell

13 junction signaling and axonal guidance signaling were the most frequent pathways listed among

14 the top 5 canonical pathways. **C. Top GWAS SNP eQTL Genes** (n=37). EIF2 signaling, axonal

15 guidance signaling, regulation of eIF4 and p70S6K signaling, oxidative phosphorylation, and

16 mitochondrial dysfunction were the most frequent pathways listed among the top 5 canonical

17 pathways. **D. Frequency of significance of each pathway for the 3 gene lists.**

18 Energy/Metabolism signaling followed by Stress/Injury signaling, followed by Developmental

19 signaling and Structural signaling were the most frequent themes in the All Candidate Genes list.

- 1 Stress/Injury signaling, followed by Energy/Metabolism signaling, Developmental signaling, and
- 2 Structural signaling were the most frequent themes in the All eQTL Genes list.
- 3 Energy/Metabolism signaling and Inflammation most commonly appeared in the Top GWAS
- 4 SNP eQTL Genes list.
- 5

1 **Supplemental Table 1. Top pathways and upstream regulators within each AF candidate**
 2 **gene list**

	Top GWAS SNP eQTL Genes (n=37)	All eQTL Genes (n=131)	All Candidate Genes (n=233)
Top Canonical Pathways	<ul style="list-style-type: none"> • G Beta Gamma Signaling • Ceramide Degradation • Gi Signaling • Sphingosine and Sphingosine -1-phosphate Metabolism • Axonal Guidance Signaling 	<ul style="list-style-type: none"> • IL-15 Production • G Beta Gamma Signaling • Caveolar-mediated Endocytosis Signaling • Phagosome Maturation • Agrin Interactions at Neuromuscular Junction 	<ul style="list-style-type: none"> • Cardiac Hypertrophy Signaling (Enhanced) • Axonal Guidance Signaling • Cardiomyocyte Differentiation via BMP Receptors • Ephrin Receptor Signaling • Factors Promoting Cardiogenesis in Vertebrates
Top upstream regulators	<ul style="list-style-type: none"> • miR-4455 • <i>MEF2C</i> • <i>PTCH1</i> • <i>HEY2</i> 	<ul style="list-style-type: none"> • <i>ESRRA</i> • miR-450a-1-3p • <i>LRIG3</i> • Phenylacetate 	<ul style="list-style-type: none"> • <i>MYOCD</i> • <i>ESRRA</i> • <i>NKX2-5</i> • <i>HEY2</i> • <i>MEF2C</i>
Top Networks	<ul style="list-style-type: none"> • Cancer, Organismal Injury and Abnormalities, Respiratory Disease • Organ Morphology, Organismal Injury and Abnormalities, Skeletal and Muscular Disorders 	<ul style="list-style-type: none"> • Developmental Disorder, Hereditary Disorder, Organismal Injury and Abnormalities • Cardiac Arrhythmia, Cardiovascular Disease, Organismal Injury and Abnormalities • Auditory Disease, Cardiovascular Disease, Dental Disease • Organismal Injury and Abnormalities, Skeletal and Muscular Disorders, Infectious Diseases • Cancer, Organismal Injury and Abnormalities, Respiratory Disease 	<ul style="list-style-type: none"> • Cardiovascular Disease, Organismal Injury and Abnormalities, Skeletal and Muscular Disorders • Cell-To-Cell Signaling and Interaction, Cellular Assembly and Organization, Cancer • Cardiovascular Disease, Hereditary Disorder, Organismal Injury and Abnormalities • Inflammatory Disease, Infectious Diseases, Lipid Metabolism • Cell Cycle, Developmental Disorder, Hereditary Disorder

3

4