Whole Blood Gene Expression and Atrial Fibrillation: The Framingham Heart Study

Citation

Published Version
doi:10.1371/journal.pone.0096794

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:12406873

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Whole Blood Gene Expression and Atrial Fibrillation: The Framingham Heart Study

Honghuang Lin1,2,*, Xiaoyan Yin2, Kathryn L. Lunetta2,3, Josée Dupuis2,3, David D. McManus2,4, Steven A. Lubitz5, Jared W. Magnani6, Roby Joehanes2,7,8, Peter J. Munson2,7,8, Martin G. Larson2,9, Daniel Levy2,8, Patrick T. Ellinor5,10,11,12,*, Emelia J. Benjamin2,6,13,14

1 Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, 2 National Heart Lung and Blood Institute’s and Boston University’s Framingham Heart Study, Framingham, Massachusetts, United States of America, 3 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, 4 Cardiology Division, Department of Medicine, and Epidemiology Division, Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America, 5 Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts, United States of America, 6 Section of Cardiovascular Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, 7 Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, Maryland, United States of America, 8 Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland, United States of America, 9 Department of Mathematics and Statistics, Boston University, Boston, Massachusetts, United States of America, 10 Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 11 Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 12 Harvard Medical School, Boston, Massachusetts, United States of America, 13 Section of Preventive Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, 14 Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, United States of America

Abstract

Background: Atrial fibrillation (AF) involves substantial electrophysiological, structural and contractile remodeling. We hypothesize that characterizing gene expression might uncover important pathways related to AF.

Methods and Results: We performed genome-wide whole blood transcriptomic profiling (Affymetrix Human Exon 1.0 ST Array) of 2446 participants (mean age 66±9 years, 55% women) from the Offspring cohort of Framingham Heart Study. The study included 177 participants with prevalent AF, 143 with incident AF during up to 7 years follow up, and 2126 participants with no AF. We identified seven genes statistically significantly up-regulated with prevalent AF. The most significant gene, PBX1 (P = 2.8×10⁻⁷), plays an important role in cardiovascular development. We integrated differential gene expression with gene-gene interaction information to identify several signaling pathways possibly involved in AF-related transcriptional regulation. We did not detect any statistically significant transcriptomic associations with incident AF.

Conclusion: We examined associations of gene expression with AF in a large community-based cohort. Our study revealed several genes and signaling pathways that are potentially involved in AF-related transcriptional regulation.

Introduction

Atrial fibrillation (AF), the most common cardiac arrhythmia, is characterized by electrophysiological, structural and contractile remodeling typically in the left atrium [1–4]. Several disease pathways might predispose to AF [5–7], and heritability plays a role in its development [8–11]. Recently, genome-wide association studies (GWAS) identified several genetic loci associated with AF [12–14]. However, the mechanisms underlying the genetic predisposition to AF susceptibility remain unknown.

Gene expression is a quantitative and heritable phenotype mediated by the interplay of environmental and genetic factors [15,16]. Both human and animal models indicate that AF has profound effects on global gene expression [17–19]. Many of the differentially expressed genes were involved in ion channels, which play an essential role in the pathophysiology of AF initiation, perpetuation and adaptation [19–22]. Expression of extracellular signal-regulated kinases and angiotensin-converting enzyme were increased in AF [23]. In addition, differential gene expression was observed in patients with heart failure [24] and valvular heart disease [25].

Identification of gene expression signatures may provide insights into AF pathogenesis and suggest targets for therapeutic intervention. Prior transcriptomic studies were generally limited to small groups of selected samples, which might not represent AF in the community. We aimed to investigate associations of AF with gene...
expression in whole blood using a large community-based cohort. We also investigated interactions between genes differentially expressed in AF patients.

**Methods**

**Study Samples**

The Framingham Heart Study is a community-based observational cohort designed to investigate cardiovascular disease and its risk factors. Since the study was initiated in 1948, three generations of participants have been enrolled. A total of 5124 participants were enrolled in 1971 in the Offspring Cohort. It consists of children of the Original Cohort and their spouses [26]. Every 4–8 years, Offspring participants complete a standardized history and 12-lead ECG, blood tests and noninvasive tests. For the present study, we focused on 2446 Offspring participants who had RNA collected at the eighth examination (2005–2008). All participants gave written informed consent. Our study was approved by the Institutional Review Boards of National Human Genome Research Institute and Boston University Medical Center.

**AF Ascertainment**

At each examination, participants were asked about AF status. In addition, their AF status and cardiovascular history were solicited during surveillance interviews biennially [27,28]. We included multiple sources to determine the presence of AF, such as a 12-lead electrocardiogram from Framingham Heart Study examination, and all cardiovascular disease-related hospitalizations and clinic visits. All electrocardiograms available from study visits or in- and outpatient records were reviewed by study cardiologists to adjudicate AF and atrial flutter events.

**Gene Expression Profiling**

The gene expression profiling has been described in detail by Joehanes et al [29]. In brief, fasting peripheral whole blood was collected during clinical visits. Total RNA was isolated from frozen PAXgene blood tubes (PreAnalytiX, Hombrechtikon, Switzerland) and amplified using the WT-Ovation Pico RNA Amplification System (NuGEN, San Carlos, CA) according to the manufacturers’ standard operating procedures. The obtained cDNA was hybridized to the Human Exon 1.0 ST Array (Affymetrix, Inc., Santa Clara, CA). The raw data were quantile-normalized and log2 transformed, followed by summarization using Robust Multiarray Average [30]. The data were further adjusted for batch effects and technical covariates, including the first principle component of the expression data, batch effect, and the all-probeset-mean residual. We also adjusted gene expression for the complete blood counts, which were imputed from more than 2000 samples from the Third Generation cohort as previously described [29]. We also measured the RNA Integrity Number in almost all samples using an automated method (Assuragen Inc, Austin, TX), and tested the inclusion of the RNA Integrity Number and RNA-related measures as technical covariates, but did not find them statistically significant. We therefore did not include them in our models for this study. Our analysis focused on the “core” probe sets, which were defined as the most reliable probe sets derived from RefSeq and GenBank records. The gene and exon annotations were obtained from Affymetrix NetAffx Analysis Center (version 31). We excluded exons with signals lower than the background, and transcript clusters that were not mapped to RefSeq transcripts. We studied 207,329 probesets, corresponding to 209,699 exons and 17,873 distinct transcripts for downstream analysis.

**Statistical Analyses**

Our primary analysis used prevalent AF cases – AF diagnosed before the blood sample for RNA was drawn. The association between gene expression and AF was assessed by linear mixed effect models to account for familial correlation between Framingham Heart Study Offspring participants. We regressed gene expression on AF status, and adjusted for age and sex. Significant hits were adjusted for additional AF risk factors [31], including smoking, height, weight, systolic blood pressure, diastolic blood pressure, prevalent diabetes mellitus, prevalent myocardial infarction, prevalent heart failure and antihypertensive treatment. In secondary analyses, we studied associations of gene expression with incident AF using used Cox proportional hazards models with robust sandwich estimators and clustering by pedigree, censoring at the last follow-up time or death. We also examined the correlation of gene expression with various medications, including warfarin, beta blocker, digoxin, and calcium channel blocker. The gene expression was treated as the outcome, whereas the usage of medication was treated as dichotomous independent variables. To adjust for multiple testing, false discovery rate (FDR) [32] was used, and statistical significance was claimed if FDR was

### Table 1. Clinical characteristics of the sample.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No AF (n = 2,126)</th>
<th>Prevalent AF (n = 177)</th>
<th>Incident AF (n = 143)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women, n (%)</td>
<td>1224 (58%)</td>
<td>63 (36%)</td>
<td>56 (39%)</td>
</tr>
<tr>
<td>Age, year ± SD</td>
<td>66±9</td>
<td>73±8</td>
<td>71±8</td>
</tr>
<tr>
<td>Smoker, n (%)</td>
<td>186 (9%)</td>
<td>7 (4%)</td>
<td>11 (8%)</td>
</tr>
<tr>
<td>Height, inches ± SD</td>
<td>66±4</td>
<td>67±4</td>
<td>66±4</td>
</tr>
<tr>
<td>Weight, pound ± SD</td>
<td>174±39</td>
<td>186±42</td>
<td>181±41</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>128±17</td>
<td>128±20</td>
<td>137±19</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>74±10</td>
<td>69±11</td>
<td>72±10</td>
</tr>
<tr>
<td>Prevalent diabetes mellitus, n (%)</td>
<td>317 (15%)</td>
<td>56 (32%)</td>
<td>40 (28%)</td>
</tr>
<tr>
<td>Prevalent myocardial infarction, n (%)</td>
<td>22 (1%)</td>
<td>41 (23%)</td>
<td>7 (5%)</td>
</tr>
<tr>
<td>Prevalent heart failure, n (%)</td>
<td>67 (3%)</td>
<td>45 (25%)</td>
<td>9 (6%)</td>
</tr>
<tr>
<td>Antihypertensive treatment, n (%)</td>
<td>971 (46%)</td>
<td>129 (73%)</td>
<td>94 (66%)</td>
</tr>
</tbody>
</table>

DOI:10.1371/journal.pone.0096794.t001
less than 0.05. All analyses were performed using R software package (www.r-project.org/).

Construction of Interaction Subnetwork Associated with AF

We applied a dense module searching strategy [33] to identify modules enriched with AF-related genes. The protein-protein interaction data were obtained from the PINA database [34], which was compiled from multiple protein-protein interaction databases. All genes were included into analyses and assigned as the seed gene once. Briefly, each gene was assigned a score to represent its association with AF. The module searching started from a single seed gene. Neighboring genes were added sequentially to the module if the adding increased the overall module score [35], which was defined as $Z_m = \sum_{i=1}^{k} g_i$, where $k$ was the number of genes in the module, and $g_i$ was the score of gene $i$. The searching stopped if no genes could be added. Ingenuity Pathway Analysis tool (Ingenuity Systems, Redwood, CA) was then used to examine the enrichment of all module genes in each canonical pathway, and the significance was claimed if its FDR [32] was less than 0.05.

Results

Differential Gene Expression

Descriptive characteristics for 2446 eligible participants (mean age $66 \pm 9$ years, 55% women) are provided in Table 1. We observed 177 prevalent cases who developed AF before examination 8 when their blood was collected for RNA extraction (prevalent AF), and 143 additional AF cases who developed AF after their RNA was collected (incident AF). The reference group consisted of 2126 participants who had not developed AF by the last follow-up in 2012. Figure 1 is the volcano plot showing the association of all the studied genes with prevalent AF. As listed in Table 2, with FDR $<0.05$, seven genes were significantly associated with prevalent AF; all were over-expressed in prevalent AF. The most significant gene, PBX1 ($P = 2.8 \times 10^{-7}$, FDR = 0.005), encodes a transcriptional factor belonging to the PBX homeobox family. The most significant decreased expression gene was REL ($P = 2.6 \times 10^{-5}$, FDR = 0.14), which did not reach the significance cutoff. We then tested the association of these genes with prevalent AF adjusting for additional AF risk factors [31], including smoking, height, weight, systolic blood pressure, diastolic blood pressure, prevalent diabetes mellitus, prevalent myocardial infarction, prevalent heart failure and antihypertensive treatment. As shown in Table S1 in File S1, the associations of these genes were attenuated, but they still reached the nominal significance cutoff ($P < 0.05$).

In our secondary analysis, we performed the association of gene expression with incident AF adjusted for age and sex. No genes reached the significance cutoff (FDR $<0.05$) (Table S2 in File S1). The most significant gene was GPCPD1 ($P = 1.5 \times 10^{-5}$, FDR = 0.27). The volcano plot of gene expression with incident AF is shown in Figure S1 in File S1.

Since Affymetrix Exon ST1.0 Array is able to profile transcriptomic signatures at the exon level, we also studied the association of each exon expression with prevalent AF. Table S3 in File S1 shows the top prevalent AF-related exons, and the Table S4 in File S1 shows the association of each exon within the seven AF-related genes. Interestingly, the seventh exon of PBX1 was also the most significant exon associated with prevalent AF ($P = 3.5 \times 10^{-7}$), suggesting that the exon has the most significant association with prevalent AF.
significant differential expression between prevalent AF cases and controls.

We also performed additional analyses to study the correlation of gene expression with various medications, including warfarin, beta blocker, digoxin, and calcium channel blocker. We found that warfarin, beta blocker, digoxin, and calcium channel blocker have 3, 287, 0, and 166 significant genes respectively (FDR < 0.05); the most significant genes for each medication are shown in the Table S5 in File S1.

In order to further characterize transcriptomic signatures associated with prevalent AF, we stratified samples with prevalent AF based on their status during the blood draw: sinus rhythm

![Figure 1. Volcano plot of expression association with prevalent AF.](image) Each dot represents one gene. The x-axis represents the effect of each gene, whereas the y-axis represents the log10 (P value). Positive effects represent that the genes were over-expressed in prevalent AF, while negative effects represent that the genes had decreased expression with prevalent AF. Seven genes that reached the significant cutoff (FDR < 0.05) were marked.

doi:10.1371/journal.pone.0096794.g001

**Table 3.** Most significant canonical pathways enriched with genes in the AF subnetwork (FDR < 0.005).

<table>
<thead>
<tr>
<th>Canonical pathway</th>
<th>FDR</th>
<th>Ratio*</th>
<th>Genes in the AF subnetwork</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antiproliferative Role of TOB in T Cell Signaling</td>
<td>3.5 × 10^{-6}</td>
<td>6/26 (0.23)</td>
<td>SMAD2, CCNA2, PABPC4, CUL1, CDKN1B, SKP2</td>
</tr>
<tr>
<td>Hypoxia Signaling in the Cardiovascular System</td>
<td>1.1 × 10^{-6}</td>
<td>6/67 (0.09)</td>
<td>HSP90B1, JUN, COP5, HSP90AA1, HIF1A, UBE2D3</td>
</tr>
<tr>
<td>Cyclins and Cell Cycle Regulation</td>
<td>3.6 × 10^{-6}</td>
<td>6/90 (0.07)</td>
<td>CCNA2, PA2G4, CUL1, BTRC, CDKN1B, SKP2</td>
</tr>
<tr>
<td>Cell Cycle: G1/S Checkpoint Regulation</td>
<td>1.7 × 10^{-6}</td>
<td>5/67 (0.07)</td>
<td>PA2G4, CUL1, BTRC, CDKN1B, SKP2</td>
</tr>
<tr>
<td>Protein Ubiquitination Pathway</td>
<td>5.4 × 10^{-6}</td>
<td>8/264 (0.03)</td>
<td>USP7, HSP90B1, HSPA9, CUL1, HSP90AA1, BTRC, UBE2D3, SKP2</td>
</tr>
<tr>
<td>Aryl Hydrocarbon Receptor Signaling</td>
<td>8.7 × 10^{-6}</td>
<td>6/148 (0.04)</td>
<td>CCNA2, HSP90B1, JUN, NEDD8, HSP90AA1, CDKN1B</td>
</tr>
<tr>
<td>Estrogen-mediated S-phase Entry</td>
<td>2.5 × 10^{-4}</td>
<td>3/27 (0.11)</td>
<td>CCNA2, CDKN1B, SKP2</td>
</tr>
<tr>
<td>Prostate Cancer Signaling</td>
<td>9.1 × 10^{-4}</td>
<td>4/94 (0.04)</td>
<td>HSP90B1, PA2G4, HSP90AA1, CDKN1B</td>
</tr>
<tr>
<td>Cell Cycle: G2/M DNA Damage Checkpoint Regulation</td>
<td>1.5 × 10^{-6}</td>
<td>3/49 (0.06)</td>
<td>CUL1, BTRC, SKP2</td>
</tr>
<tr>
<td>HIF1α Signaling</td>
<td>2.0 × 10^{-2}</td>
<td>4/105 (0.04)</td>
<td>JUN, COP5, HSP90AA1, HIF1A</td>
</tr>
<tr>
<td>Glucocorticoid Receptor Signaling</td>
<td>2.3 × 10^{-6}</td>
<td>6/280 (0.02)</td>
<td>SMAD2, HSP90B1, JUN, HSPA9, PBX1, HSP90AA1</td>
</tr>
<tr>
<td>EIF2 Signaling</td>
<td>2.6 × 10^{-3}</td>
<td>5/192 (0.03)</td>
<td>RPL15, EIF2AK1, RPL4, RPL22, RPS18</td>
</tr>
<tr>
<td>eNOS Signaling</td>
<td>4.8 × 10^{-3}</td>
<td>4/136 (0.03)</td>
<td>CCNA2, HSP90B1, HSPA9, HSP90AA1</td>
</tr>
</tbody>
</table>

*Ratio is the number of genes within AF subnetwork comparing to the total number of genes in the pathway.

doi:10.1371/journal.pone.0096794.t003
(n = 115) and AF (n = 62). As shown in the Table S6 in File S1, all seven prevalent AF-related genes were associated with both sinus rhythm and AF (nominal \( P < 0.05 \)), although the association was attenuated due to the reduced sample size and limited power.

**Gene Interaction Network Associated with AF**

We applied a dense module searching strategy [33] to identify modules containing genes with strong association with AF. Each module represents a group of interacting genes; the nodes represent genes, and the edges represent interactions between genes. A total of 7993 modules were identified. Given the highly overlapping nature of modules and the enrichment of significant modules, we merged the top 1% modules with highest scores as previously suggested [33] and considered it as an AF subnetwork (Figure 2). The subnetwork consisted of 106 nodes and 360 edges. The most significant gene, *PBX1*, was included in the subnetwork.

In order to further characterize the function of AF subnetwork, we applied Ingenuity Pathway Analysis to study the gene enrichment in canonical pathways. Twenty-two canonical pathways were significantly enriched with genes in the AF subnetwork (FDR < 0.05). The most significant pathways were antiproliferative role of transducer of ERBB2 (TOB) in T Cell Signaling (FDR = 3.5 \times 10^{-9}) and hypoxia signaling in the cardiovascular system (FDR = 1.1 \times 10^{-5}) (Table 3).

**Discussion**

Gene expression is considered an important bridge to connect genetic variations and diseases [15,36]. We performed genome-wide whole blood transcriptomic profiling of 177 participants with prevalent AF and more than 2000 AF-free participants from the Framingham Heart Study Offspring cohort. Seven genes were upregulated in samples with prevalent AF after adjustment for multiple testing, but none of them were reported in prior studies [17–23,37], suggesting the complexity of gene regulation network underlying AF. We did not identify any transcriptomic signatures significantly associated with incident AF.

The most significant gene was *PBX1*, which encodes a homeodomain transcription factor. The gene plays an important role in cardiovascular development, and orchestrates separate transcriptional pathways to control great-artery patterning and...
cardiac outflow tract septation [38]. The inactivation of PBX1 results in persistent truncus arteriosus in mice [39]. Another significant gene, SLC7A1, encodes a high affinity cationic amino acid transporter that participates in the total carrier-mediated uptake activity [40]. A polymorphism in the 3’-UTR of SLC7A1 was found to be associated with hypertension and endothelial dysfunction [41], probably due to the binding of miR-122 to the region [42].

Beyond the individual gene association, we also studied the combined association of differential gene expression with AF. Increasing evidence has suggested that the development of AF might result from the disruption of multiple biological pathways [5–7,43]. We constructed an AF subnetwork by taking advantage of a wealth of protein-protein interaction information that was collected in several databases over years. The subnetwork was highly enriched with genes involved in multiple signaling pathways, including hypoxia signaling pathway in the cardiovascular system and the TOB antiproliferative signaling pathway. The hypoxia signaling pathway is key to a series of molecular events to response to the local oxygen tension in the cardiovascular system [44,45]. The pathway is also involved in the myocardial remodeling [46] and the activation of angiogenic growth factors [47]. The TOB antiproliferative signaling pathway regulates T cell proliferation and cytokine transcription, which could avoid excessive T cell response and thus reduce the risk of autoimmune disorders or tissue injury [48]. Multiple genes in the pathways are involved in attenuated cardiomyocyte hypertrophy and congenital heart diseases [49,50]. Our analyses provide an approach to translate individual gene associations into a broader picture of transcriptomic profiles, which might be associated with the prevalence or onset of AF.

There are several limitations in our study. We found no significant transcriptomic signatures associated with incident AF, although the sample size was similar to that of prevalent AF. The mechanisms contributing to the discrepancy are unclear. We speculate that the discrepancy may be due to different signaling pathways underlying the onset and perpetuation of AF; genes associated with incident AF might be responsible for the initiation of AF, whereas genes associated with prevalent AF might be responsible for the maintenance of AF. Given that RNA was only collected at the eighth examination of the Offspring cohort but not any earlier examinations, it is infeasible to examine the longitudinal changes of transcriptomic signatures for the same participants. Gene expression varies in different tissue types. The transcriptional profiles in this study were measured on whole blood, but the ideal tissue for study AF is the left atrium. However, invasive left atrial specimen collection is infeasible in a community-based study, and would result in highly selected samples [51] and few controls. In addition, the cross-hybridization in microarray based platforms might result in non-specific signals for some genes, thus a further validation by RT-PCR or RNA sequencing would be helpful to rule out potential false associations [52,53]. As a proof-of-concept, we performed RT-PCR on 95 coronary heart disease-related genes in 429 samples selected for a pilot study [29]. As shown in the Table S7 in File S1, the expression of 49 out of 95 genes was correlated well between microarray and RT-PCR (Pearson’s correlation coefficient >0.5), and 91 genes have the same direction of effect, suggesting robustness of our assay, although it is not a direct validation of top hits for AF. Given that thousands of tests were conducted and the lack of external replications, it is possible that some findings might be false positives even if we used very stringent Bonferroni correction.

Because of the observational study design, the associations we observed do not prove causation. Moreover, the participants were middle-age to older adults, largely of European descent, raising uncertainty regarding generalizability to other ethnicities/races. Our current study did not distinguish between atrial fibrillation and atrial flutter, nor patterns of AF, which might have unique gene expression profiles. Future analyses with stratification of rhythm types might uncover unique gene expression profiles specific to each rhythm type.

In conclusion, we examined the association of whole blood gene expression with AF in a large community-based cohort. Our study revealed that several genes and signaling pathways might be activated in prevalent AF and may be involved in AF maintenance. Future increase of sample size, longer follow up, and external validation would help to identify additional transcription signatures associated with AF and might provide valuable insights into AF pathogenesis [54].

Supporting Information
File S1 Supplemental Materials. (PDF)

Author Contributions
Conceived and designed the experiments: HL DL PTE EJB. Performed the experiments: HL XY. Analyzed the data: HL XY. Contributed reagents/materials/analysis tools: KLL JD RJ PJM MGL. Wrote the paper: HL EJB. Critical review and suggestion: XY KLL JD DDM SAL RJM RJ PJM MGL DL PTE.

References