Gene expression and genetic variation in human atria

Honghuang Lin, PhD,[1,2] Elena V. Dolmatova, MD,[8] Michael P. Morley, PhD,[4] Kathryn L. Lunetta, PhD,[1,5] David D. McManus, MD, ScM,[1,3] Jared W. Magnani, MD, MSc,[1,6] Kenneth B. Margulies, MD,[4] Hakon Hakonarson, MD, PhD,[4] Federica del Monte, MD, PhD,[10] Emelia J. Benjamin, MD, ScM,[1,6,7,11] Thomas P. Cappola, MD, ScM,[4] Patrick T. Ellinor, MD, PhD[8,9,12,13]

From the [1]National Heart Lung and Blood Institute's and Boston University's Framingham Heart Study, Framingham, Massachusetts, [2]Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, [3]Cardiology Division, Department of Medicine, and Epidemiology Division, Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts, [4]Penn Cardiovascular Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, [5]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, [6]Section of Cardiovascular Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, [7]Section of Preventive Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, [8]Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts, [9]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, [10]Cardiovascular Institute, Beth Israel Deaconess Medical Center, Boston, Massachusetts, [11]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, [12]Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, Massachusetts, and [13]Harvard Medical School, Boston, Massachusetts.

BACKGROUND The human left and right atria have different susceptibilities to develop atrial fibrillation (AF). However, the molecular events related to structural and functional changes that enhance AF susceptibility are still poorly understood.

OBJECTIVE The purpose of this study was to characterize gene expression and genetic variation in human atria.

METHODS We studied the gene expression profiles and genetic variations in 53 left atrial and 52 right atrial tissue samples collected from the Myocardial Applied Genomics Network (MAGNet) repository. The tissues were collected from heart failure patients undergoing transplantation and from unused organ donor hearts with normal ventricular function. Gene expression was profiled using the Affymetrix GeneChip Human Genome U133A Array. Genetic variation was profiled using the Affymetrix Genome-Wide Human SNP Array 6.0.

RESULTS We found that 109 genes were differentially expressed between left and right atrial tissues. A total of 187 and 259 significant *cis*-associations between transcript levels and genetic variants were identified in left and right atrial tissues, respectively. We also found that a single nucleotide polymorphism at a known AF locus, rs3740293, was associated with the expression of *MYOZ1* in both left and right atrial tissues.

CONCLUSION We found a distinct transcriptional profile between the right and left atrium and extensive *cis*-associations between atrial transcripts and common genetic variants. Our results implicate *MYOZ1* as the causative gene at the chromosome 10q22 locus for AF.

KEYWORDS Genetics; Expression quantitative trait loci; Gene expression; Atrial tissue

ABBREVIATIONS **AF** = atrial fibrillation; **eQTL** = expression quantitative trait loci; **FDR** = false discovery rate; **GWAS** = genome-wide association studies; **LA** = left atrium; **MAGNet** = Myocardial Applied Genomics Network; **RA** = right atrium; **SNP** = single nucleotide polymorphism

(Heart Rhythm 2014;11:266–271)

The human left atrium (LA) and right atrium (RA) have distinct electrophysiologic and pathophysiologic differences. Atrial fibrillation (AF) is characterized predominantly by left atrial (LA) structural and contractile remodeling.[1] LA enlargement is also associated with many complex diseases, such as stroke,[2] obesity,[3] and cardiovascular diseases.[4,5] In contrast, typical atrial flutter is known to be an arrhythmia arising from the RA.

Drs. Benjamin, Cappola, and Ellinor contributed equally to this work.

This work is supported by National Institutes of Health Grant R01HL092577 to Drs. Ellinor and Benjamin, 1R01 HL102214 to Dr. Benjamin, R01HL104156 and K24HL105780 to Dr. Ellinor, R01HL105993 to Dr. Margulies and Cappola, and KL2RR031981 and 1U01HL105268-01 to Dr. McManus; and American Heart Association Award 13EIA14220013 to Dr. Ellinor and 09FTF2190028 to Dr. Magnani. **Address reprint requests and correspondence:** Dr. Honghuang Lin, Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, 72 East Concord St, B-616, Boston, MA 02118. E-mail address: hhlin@bu.edu.

1547-5271/$-see front matter http://dx.doi.org/10.1016/j.hrthm.2013.10.051

Thousands of genes are expressed in the human heart,[6] and many genes are well known to be differentially regulated between the atria and ventricles[7,8] as well as in disease states including AF,[9] heart failure,[6,10] and hypertrophy.[11,12] Earlier studies on mice and human atria have found several genes, including *BMP10* and *PITX2*, are differentially expressed between LA and RA.[13,14] However, the sample sizes of these studies were usually small (<15), and the relationship between atrial gene expression and genetic variants has not been systematically investigated.

In the past few years, genome-wide association studies (GWAS) have been used successfully to identify thousands of genetic loci associated with a variety of diseases and phenotypic traits.[15] Unfortunately, many novel candidate loci do not have defined functions, and the mechanisms to confer disease susceptibility remain largely unknown. Expression quantitative trait loci (eQTL) analyses can be used to correlate the relation between single nucleotide polymorphisms (SNPs) and gene expression. Such eQTL analysis is considered an intermediate phenotype between genetic variations and diseases.[16,17] Thus, it would be interesting to determine if any genetic variations were associated with gene expression in atrial tissues.

Our objectives were three-fold: (1) to characterize the expression profiles of the RA and LA; (2) to perform an eQTL analysis in atrial tissue; and (3) to determine, using AF as an example, if GWAS disease variants correlate with atrial gene expression.

Methods

Study samples

We studied samples collected from 64 genetically inferred European ancestry participants in the Myocardial Applied Genomics Network (MAGNet) repository. The tissues were collected from discarded hearts of heart failure patients undergoing transplantation and from unused organ donor hearts with normal ventricular function. Twelve had only LA, 11 had only RA tissue, and 41 individuals had both atrial tissues collected. The sample collection was approved by the Institutional Review Board at the University of Pennsylvania.

Genotyping

Genomic DNA was extracted using the Gentra Puregene Tissue Kit (Qiagen, Gaithersburg, MD), which was then hybridized to the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA) in accordance with the manufacturer's standard recommendations. The SNP calling from raw CEL files were performed using Birdsuite software package.[18] We filtered out SNPs with missing rate higher than 20%, minor allele frequency $<5\%$, or Hardy-Weinberg equilibrium $P < 1 \times 10^{-3}$ (Fisher exact test). At the end, 637,607 SNPs were used for downstream analyses. Quality control was performed using PLINK software package.[19] All the participants were of European ancestry, which was inferred using multidimensional scaling of analysis Affymetrix 6.0 genotypes with similar data from a larger cohort of known ancestry ($N = 340$).

Transcriptional profiling

The RNA was extracted using the Trizol Reagent (Life Technologies, Carlsbad, CA),[20] and cDNA was hybridized to the Affymetrix Human Genome U133A Array (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. The raw CEL files were preprocessed using Bioconductor R package.[21] The data were quantile normalized and log2 transformed, followed by summarization using Robust Multi-array Average.[22] Potential batch effects were corrected by ComBat, an empirical Bayes–based approach.[23] The gene annotation was downloaded from Affymetrix NetAffx Analysis Center (version 32). We excluded transcripts that were not aligned uniquely to the reference genome and those that were not expressed in any of the samples using MAS 5.0 algorithm.[24,25] Many genes were represented by multiple probesets, corresponding to different transcripts. We treated each transcript equally and reported the transcript-based analysis, but the result was interpreted at the gene level. A total of 11,818 transcripts, corresponding to 8644 genes, were used for the downstream analyses after adjusting for age and sex.

Differential gene expression

Differential expression between LA and RA tissues was assessed using the unpaired Student t test. To adjust for multiple testing, we used the highly conservative Bonferroni correction method. Significance was claimed if $P < 4.2 \times 10^{-6}$ (0.05/11,818 transcripts). The differential analysis was performed using R software packages (www.r-project.org/). We also used GOrilla web tool[26] to test the enrichment of differentially expressed genes in Gene Ontology categories. We limited the enrichment analysis on the biologic processes, and the enrichment was claimed if the false discovery rate (FDR) was $<5\%$.[27] Using a similar approach, we also compared the differential expression between participants with and without heart failure or AF.

eQTL Analysis

We used a linear regression model to test the association between genetic variations and gene expression. The significance was indicated by the regression P value. We defined *cis*-eQTLs as transcripts that were associated with SNPs within 1 Mb from the gene boundary and *trans*-eQTLs as transcripts that were associated with SNPs at least 1 Mb far away or on different chromosomes.[28] A total of 7,535,239,526 associations (637,607 SNP $\times$ 11,818 transcripts) were tested. Empirical FDR was estimated by the permutation test (see Online Supplemental Materials).

qPCR confirmation of significant association

A total of 70 human LA samples were used for the replication studies. Surgical samples were obtained at Massachusetts General Hospital during cardiac surgery for