



Integrative Analysis Revealing Human Heart-Specific Genes and Consolidating Heart-Related Phenotypes

Jinsoo Ahn^{1†}, Huiguang Wu^{1,2†} and Kichoon Lee^{1*}

¹ Department of Animal Sciences, The Ohio State University, Columbus, OH, United States, ² College of Veterinary Medicine, Yangzhou University, Yangzhou, China

OPEN ACCESS

Edited by:

Clint L. Miller,
University of Virginia, United States

Reviewed by:

Xia Yang,
University of California, Los Angeles,
United States
Alexander Teumer,
University of Greifswald, Germany

*Correspondence:

Kichoon Lee
lee.2626@osu.edu

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Human Genomics,
a section of the journal
Frontiers in Genetics

Received: 24 September 2019

Accepted: 30 June 2020

Published: 30 July 2020

Citation:

Ahn J, Wu H and Lee K (2020)
Integrative Analysis Revealing Human
Heart-Specific Genes
and Consolidating Heart-Related
Phenotypes. *Front. Genet.* 11:777.
doi: 10.3389/fgene.2020.00777

Elucidating expression patterns of heart-specific genes is crucial for understanding developmental, physiological, and pathological processes of the heart. The aim of the present study is to identify functionally and pathologically important heart-specific genes by performing the Ingenuity Pathway Analysis (IPA). Through a median-based analysis of tissue-specific gene expression based on the Genotype-Tissue Expression (GTEx) data, we identified 56 genes with heart-specific or elevated expressions in the heart (heart-specific/enhanced), among which three common heart-specific/enhanced genes and four atrial appendage-specific/enhanced genes were unreported regarding the heart. Differential expression analysis further revealed 225 differentially expressed genes (DEGs) between atrial appendage and left ventricle. Our integrative analyses of those heart-specific/enhanced genes and DEGs with IPA revealed enriched heart-related traits and diseases, consolidating evidence of relationships between these genes and heart function. Our reports on comprehensive identification of heart-specific/enhanced genes and DEGs and their relation to pathways associated with heart-related traits and diseases provided molecular insights into essential regulators of cardiac physiology and pathophysiology and potential new therapeutic targets for heart diseases.

Keywords: heart-specific genes, differentially expressed genes, ingenuity pathway analysis, gene expression, heart disease

INTRODUCTION

During heart development, conserved transcriptional networks govern cardiac cell fate determination, cardiomyocyte differentiation, and cardiac morphogenesis (Olson, 2006). Identification of cardiac-specific genes and elucidation of their expression patterns are therefore crucial for understanding developmental, physiological, and pathological processes of the heart. Importantly, characteristic transcriptional profiles between atria and ventricles of the heart have been linked to heart region-specific differences in morphogenesis, structure, contractility, and electrophysiological properties (Bao et al., 1999; Barth et al., 2005; Olson, 2006). Based on these significances of abundant or differential expression of genes in the heart, comprehensive screenings of heart-specific genes and differentially expressed genes (DEGs) between heart chambers gain significant attention. To this end, the Genotype-Tissue Expression (GTEx) data can be analyzed to provide up-to-date RNA-seq transcriptomic profiling for various human tissues

(GTEx Consortium, 2017). Recently, a robust approach of identifying tissue-specific genes using GTEx data was developed by our group (Ahn et al., 2019). This approach was further used to comprehensively identify heart-specific genes and DEGs in this study.

In order to obtain insight into functional relevance of genes of interests in organ development and various pathological conditions, molecular interactions and pathways under various developmental and experimental conditions can be analyzed by Ingenuity Pathway Analysis (IPA) (QIAGEN Inc.,¹) as described in a previous publication (Kramer et al., 2014). Furthermore, previous studies documented a large number of candidate genes and loci associated with heart diseases and characteristics of the heart to determine genetic factors associated with physiology and pathophysiology of the heart (Buniello et al., 2019). However, despite these achievements, evidence has yet to be explored to link between genes expressed exclusively or abundantly in the heart (e.g., heart-specific genes and heart-enhanced genes described in the current study) and heart traits or diseases. The objectives of the current study were to identify heart-specific and heart-enhanced genes including functionally unreported novel genes, examine DEGs between atrial appendage and left ventricle, and provide an approach that can narrow down candidate genes for further research by performing IPA-based pathway analysis.

Herein, using gene expression values from atrial appendage and left ventricle of the heart, three novel protein-coding genes that are commonly specific or enhanced in both atrial appendage and left ventricle, four novel protein-coding genes that are specific or enhanced in atrial appendage, and 225 protein-coding DEGs between the two tissues were identified. By integrating data from pathway analysis, evidence of interrelationships between these genes and major heart-related traits or diseases including cardiogenesis, cardiac muscle contraction, fibrosis, heart failure, organismal death, morphogenesis of cardiac muscle, hypertrophy of left ventricle, and calcium signaling were consolidated. Overall, in this study, heart-specific genes and DEGs between atrial appendage and left ventricle and their relations with pathways associated with heart-related traits and diseases were comprehensively identified, providing molecular insights into essential regulators of cardiac physiology and pathophysiology and potential new therapeutic targets for heart diseases.

MATERIALS AND METHODS

Data Collection and Filtering

Comprehensive analyses of human transcriptome across tissues were initiated by downloading on 02/22/2019 public human RNA-seq data (GTEx Analysis v7; dbGaP Accession: phs000424.v7.p2, release date: June 30, 2017) deposited in the GTEx portal²) (Figure 1). Gene expression values in the data were derived from non-diseased normal tissues of

postmortem human donors (aged 21–70) and normalized using transcripts per million (TPM). Before processing the data, tissues from less than 30 donors and non-tissue samples such as cultured transformed fibroblasts and EBV-transformed lymphocytes were excluded from the current study. Then, tissue samples with high quality RNA (RNA integrity number, RIN, of 6.0 or higher) were selected. Finally, the filtered data included 47 human tissues, and the sample size (i.e., donor numbers) for each tissue ranged from 35 to 560 (Supplementary Table S1). In addition, 56,202 human protein-coding and non-coding genes were filtered by general thresholds of at least 20% of samples having TPM > 0.1³) and median TPM > 0.5 in the heart (Ahn et al., 2019), leaving 16,480 and 14,911 protein-coding and non-coding genes for data analysis for the atrial appendage (AA) and left ventricle (LV), respectively, (Figure 2A).

Cumulative Pathway Enrichment

Enrichment analysis on randomly selected genes was conducted as follows. From the above 16,480 (AA) and 14,911 (LV) genes, the top 500 genes were selected by the rank of relative median value. From either 16,480 (or 14,911) genes or the top 500 genes, 5 to 500 genes were randomly selected by adding one gene per new selection, using the sample_n function of the dplyr package (v0.8.5⁴). The same random selection procedure was also applied for AA-specific/enhanced genes (5 to 81 genes) and LV-specific/enhanced genes (5 to 79 genes). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was selected as a reference database, and we ran an Enrichr algorithm⁵ (Chen et al., 2013; Kuleshov et al., 2016) to query to the KEGG 2019 human database (308 terms, gene coverage of 7,802, and 92 genes per term) as presented in a previous publication (Zachariou et al., 2018). In the current study, a Python script (enrichr-cli.py, W. De Coster⁶) was used to run the algorithm and obtain Enrichr combined scores for input gene lists, which are derived from a correction to the Fisher's exact test *p*-value by the *z*-score of deviation from the expected rank (Chen et al., 2013).

Data Processing and Statistical Analysis

Heart-specific genes were defined as genes with fold changes of the median expression levels (FCMs) higher than 5.0 in the heart *versus* all other tissues. Heart-enhanced genes were defined as genes having FCMs higher than 5.0 in the heart *versus* all other tissues, with an exception of 1 or 2 other tissues, and a higher median expression in the heart in the case of exceptions. That is, we selected a trend in which the heart showed a greater expression than any other analyzed tissues even in the case of heart-enhanced genes as well as heart-specific genes.

Using lmFit function in the R/Bioconductor limma package (Ritchie et al., 2015), a linear model was fitted to our data and a contrast matrix was created for all 45 pair-wise comparisons

¹<https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>

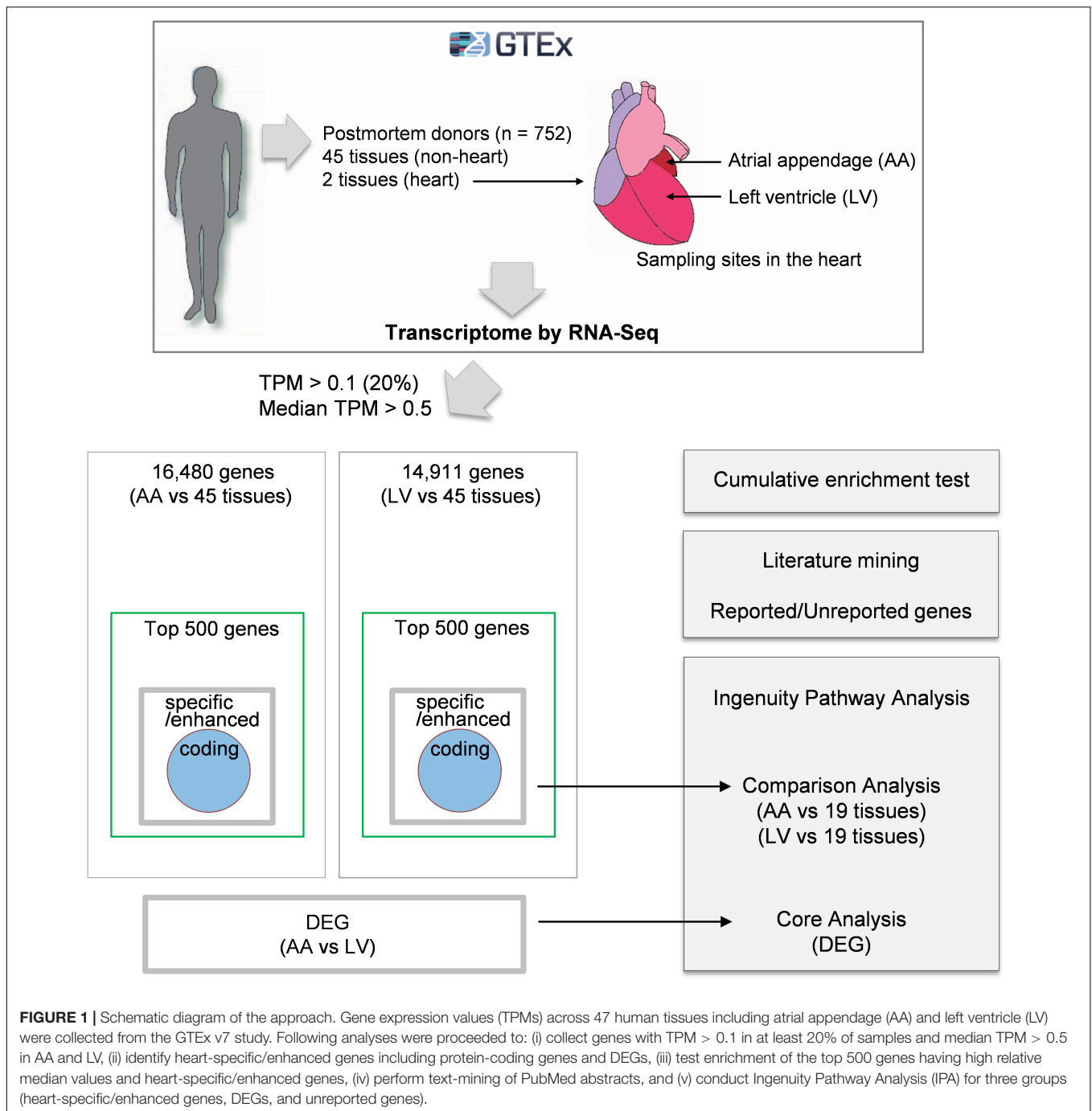
²<https://www.gtexportal.org/home>

³<https://www.gtexportal.org/home/documentationPage>

⁴<https://CRAN.R-project.org/package=dplyr>

⁵<http://amp.pharm.mssm.edu/Enrichr>

⁶https://github.com/wdecoster/enrichr_cli



between each tissue. In limma, the empirical Bayes (eBayes) method (Smyth, 2004) followed by the decideTests function with “global” option was used to test *t*-statistic for each contrast and F-statistics for all 45 contrasts simultaneously, along with multiple testing correction by the Benjamini-Hochberg’s procedure (Benjamini and Hochberg, 1995). As a result, genes with at least a 5-fold higher median expression in the heart and whose expression is significantly different between tissues, formulated as: [FDR of *i* versus *j* pairwise comparison] < 0.01, where *i* = A1 (heart-atrial appendage) or

A2 (heart-left ventricle) and *j* = B, C, D, ..., or AT (non-heart), were selected.

For grouping protein-coding and non-coding genes (81 genes in AA and 79 genes in LV) related to long intergenic non-coding RNA (linc RNA), pseudogene, antisense RNA, and sense intronic transcripts, the Ensembl Gene IDs retrieved from the GTEx portal (hg19/GRCh37) were reannotated using the human reference genome assembly (GRCh38.p12⁷). Heat

⁷<https://www.ensembl.org/info/data/ftp/index.html>

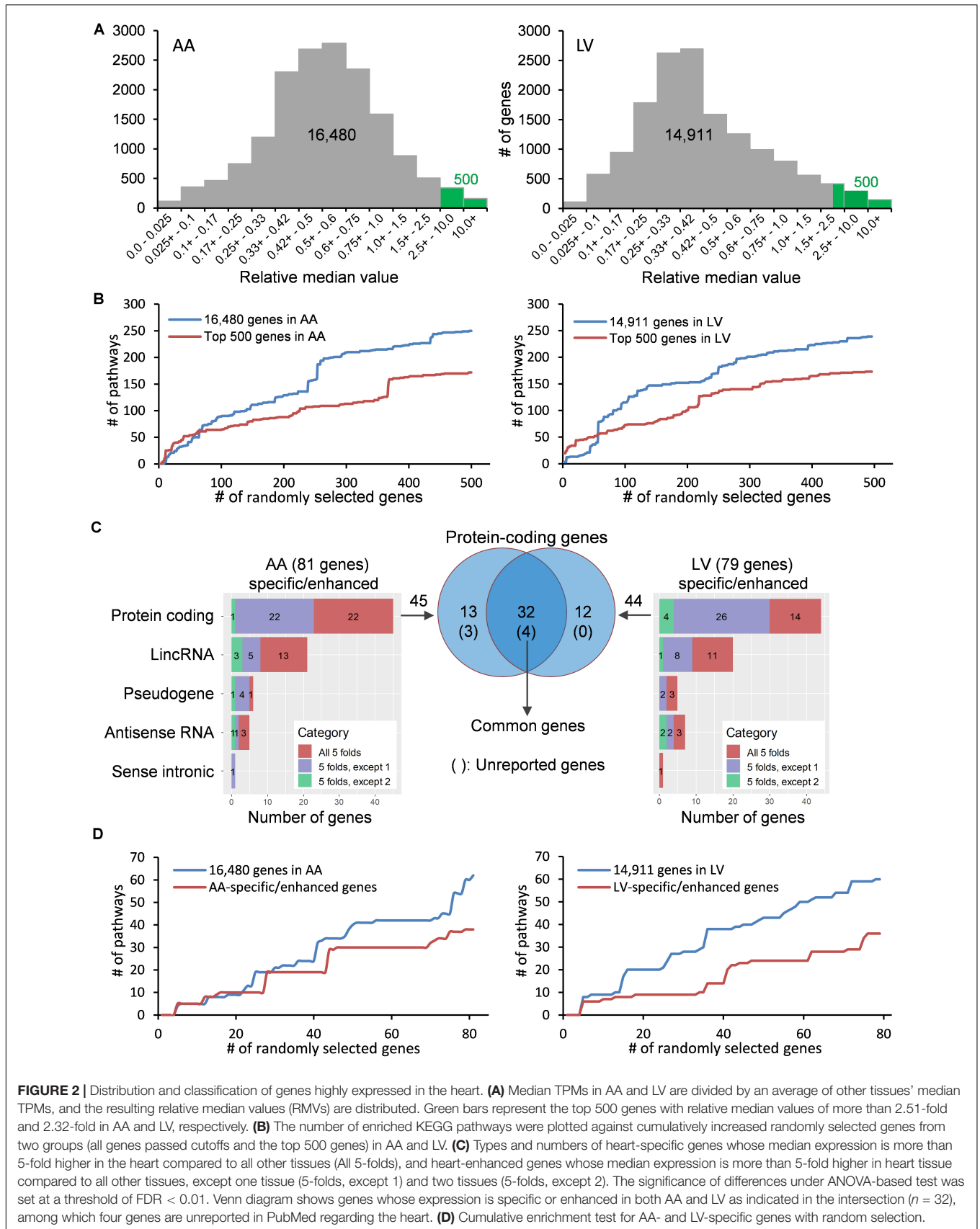


FIGURE 2 | Distribution and classification of genes highly expressed in the heart. **(A)** Median TPMs in AA and LV are divided by an average of other tissues' median TPMs, and the resulting relative median values (RMVs) are distributed. Green bars represent the top 500 genes with relative median values of more than 2.51-fold and 2.32-fold in AA and LV, respectively. **(B)** The number of enriched KEGG pathways were plotted against cumulatively increased randomly selected genes from two groups (all genes passed cutoffs and the top 500 genes) in AA and LV. **(C)** Types and numbers of heart-specific genes whose median expression is more than 5-fold higher in the heart compared to all other tissues (All 5-folds), and heart-enhanced genes whose median expression is more than 5-fold higher in heart tissue compared to all other tissues, except one tissue (5-folds, except 1) and two tissues (5-folds, except 2). The significance of differences under ANOVA-based test was set at a threshold of FDR < 0.01. Venn diagram shows genes whose expression is specific or enhanced in both AA and LV as indicated in the intersection ($n = 32$), among which four genes are unreported in PubMed regarding the heart. **(D)** Cumulative enrichment test for AA- and LV-specific genes with random selection.

maps were generated to visualize relative heart-specific and heart-enhanced expressions using the R package “heatmap3”⁸ (Zhao et al., 2014).

Literature Mining

To access to PubMed⁹ and download abstracts of publications on the heart and heart-related medical conditions, the PubMed literature database was queried on May 26, 2020 using the R package “easyPubMed”¹⁰. Briefly, the PubMed database was queried without restriction on publication date and using search terms independently for each of the 57 heart-specific/enhanced protein-coding genes as follows: ‘heart[TIAB] OR atrial[TIAB] OR ventricle[TIAB] OR cardio[TIAB] OR cardiac[TIAB] OR coronary[TIAB] OR cardium[TIAB] AND (genesymbol [TIAB])’, where TIAB stands for title/abstract and genesymbol was replaced with each different gene symbol. Abstracts were downloaded in XML format. Extraction of biomedical entities such as genes, diseases, species, and PubMed identifier (PMID) from the downloaded abstracts was performed using the R package “pubmed.mineR”¹¹ (Rani et al., 2015). When no abstract was retrieved for a certain gene by easyPubMed, it was referred to as a previously “unreported gene” in PubMed (hereafter unreported gene). When an abstract was retrieved and biomedical entities were extracted, it was referred to as a “reported gene” in PubMed (hereafter, a reported gene). For some cases in which abstracts were retrieved but related diseases were not extracted, a manual review on abstracts and full texts under each extracted PMID was performed as a part of data mining strategy (Pospisil et al., 2006) (e.g., research on non-pathological conditions such as heart rate was regarded as a reported case).

Ingenuity Pathway Analysis

The Ingenuity Pathway Analysis (IPA) was used to perform pathway enrichment analysis as follows. IPA Comparison Analysis was conducted on two groups of 20 tissues (i.e., either AA and 19 other tissues or LV and 19 other tissues) to examine similarities and differences in pathway enrichment. The selection of 20 tissues was due to the maximum analyzing capacity of IPA Comparison Analysis of 20 observations (or experimental groups), and tissues with greater size and larger sample numbers were chosen. Among the top 500 genes, all heart-specific/enhanced genes were analyzed, and their log ratios (i.e., \log_2 RMVs) were used to rank them. For those heart-specific/enhanced genes, adjusted p -values of differential expression between tissue i versus $rest$, where i is one tissue from 20 tissues and $rest$ is 19 other tissues, were obtained from the limma package using a contrast matrix of one versus 19 followed by the eBayes method and t -statistic. Using \log_2 RMVs and adjusted p -values, IPA Core Analysis was performed on each of the 20 tissues and then IPA Comparison Analysis was conducted using 20 Core Analysis results. Enrichment was considered

significant when $p < 0.01$. A cutoff set for activation or inhibition of the pathways predicted by IPA was $|\text{activation } z\text{-score}| \geq 2$.

Differential Expression Between Atrial Appendage and Left Ventricle

Read count matrix from RNA-seq of atrial appendage and left ventricle was downloaded from the GTEx portal on 02/22/2019. Read counts ≥ 6 in at least 20% of samples¹² was satisfied with 23,288 genes in AA and 22,137 genes in LV, and the union (23,610 genes) of the two gene sets were retained. For differential expression analysis, the R/Bioconductor “DESeq2” package (Love et al., 2014) was used to normalize samples for sequencing depth and calculate DEGs between AA and LV. To obtain significant DEGs, combined criteria of $FDR < 0.01$ and absolute value of $\log_2(\text{fold change}) > 3$ were used, where a fold change is defined as the expression in samples of AA divided by that of LV. IPA Core Analysis was conducted to analyze pathway enrichment of those significant DEGs, using $\log_2(\text{fold change})$ and adjusted p -values as input values. The $p < 0.01$ was considered a significant enrichment. The $|\text{activation } z\text{-score}| \geq 2$ was used to predict activation or inhibition of the pathways.

RESULTS

Data Distribution and Cumulative Pathway Enrichment

With the purpose of identifying heart-specific genes and their associations with diseases and traits, our integrative workflow was designed (Figure 1). To initiate the workflow, the human RNA-seq-based transcriptome data (GTEx Analysis v7) were downloaded from the GTEx portal¹³. In each tissue, the medians of gene expression values (TPMs) were obtained in virtue of their robustness to outliers. Then, the median TPM value in the heart was divided by an average of the rest of the median TPMs from other tissues, producing ‘relative median values (RMVs)’, and distribution of these values were plotted against the number of genes (Figure 2A). The plots were relatively symmetric and displayed about one hundred and fifty values with more than 10-fold, indicating the presence of exclusive or abundant expression in the two heart tissues: atrial appendage and left ventricle.

To examine whether the degree of enrichment changes by selecting genes with high RMVs, the number of enriched pathways were queried by increasing the input gene set cumulatively (Figure 2B). At first, five genes were randomly selected from two groups: one is the gene set with all valid genes (i.e., pre-selected 16,480 genes for AA or 14,911 genes for LV) and another is with the top-ranked 500 genes (more than 2.51- and 2.32-fold in AA and LV, respectively). For the cumulative increase, one gene was added into the gene set each time after random selection from either group. The KEGG 2019 human database was selected as a reference pathway database for this test query owing to its comprehensiveness and a wide range of usage (Kanehisa et al., 2017). Upon running through the Enrichr

⁸<https://CRAN.R-project.org/package=heatmap3>

⁹<https://www.ncbi.nlm.nih.gov/pubmed/>

¹⁰<https://CRAN.R-project.org/package=easyPubMed>

¹¹<https://CRAN.R-project.org/package=pubmed.mineR>

¹²<https://www.gtexportal.org/home/documentationPage>

¹³<https://www.gtexportal.org/home>

algorithm, cumulative numbers of total enriched pathways with Enrichr combined scores, which outperform other measurements on enrichment (Kuleshov et al., 2016; Zachariou et al., 2018), were plotted for each gene list with respect to numbers of randomly selected genes (Figure 2B).

As shown in Figure 2B, in either case, with less than about 50 to 70 genes, the number of total enriched pathways was higher in gene sets selected from the top 500 genes compared to the all genes. However, for larger gene sets, the number of enriched pathways for the top 500 gene set became substantially smaller than the all genes, and this pattern was continued in both cases of AA and LV. It indicated that the extent of enrichment was affected by RMVs and choosing genes with high RMVs increased the detection sensitivity. Taken together, highly expressed genes in the heart could be detected by sorting RMVs, and selection of the top 500 genes improved enrichment results.

Identification of Heart-Specific and Heart-Enhanced Genes

In order to identify more specific genes in the heart, we further narrowed down those top 500 genes to “heart-specific genes” and “heart-enhanced genes” using rigorous criteria as described in the section “Materials and Methods.” As a result, we selected a total of 81 protein coding and non-coding genes in AA and 79 in LV (all FCMs > 5 with up to 2 exceptions, FDR < 0.01; Figure 2C and Supplementary Table S1). Among them, the number of protein coding genes were 45 and 44 in AA and LV, respectively, and there were 32 overlapping genes. In addition, with the same random selection strategy used in the case of the top 500 genes (Figure 2B), those AA- and LV-specific/enhanced genes were subjected to the cumulative pathway enrichment test and showed a more refined enrichment than all genes similar to the top 500 genes (Figure 2D). It also indicated that the detection sensitivity increased with AA- and LV-specific/enhanced genes because the refined enrichment occurred with smaller gene sets, compared to the top 500 genes, and these specific/enhanced genes were further studied below.

Among those overlapping genes, three genes (*RD3L*, *FBXO40*, and *SMCO1*) have not been documented in PubMed (i.e., unreported genes) regarding their related diseases and functions in the heart in humans based on our literature search. Genes whose function has been investigated in non-pathological conditions such as *CCDC141* on heart rate (van de Vegte et al., 2019), as well as disease conditions, were classified as reported genes. Among non-overlapping genes, 13 protein-coding genes were AA-specific/enhanced and among them the following four genes were unreported genes: *SBK2*, *PRR32*, and *SBK3* showed exclusive expression (AA-specific), and expression of *CHRNE* was more than 5-fold compared to other tissues, except pituitary (AA-enhanced). In addition, 11 non-overlapping LV-specific/enhanced protein coding genes were found and all of them were reported. Heat maps displayed *z*-scores of RMVs of the heart-specific/enhanced protein-coding genes in various tissues (i.e., each row) with extreme specificity or abundance to AA and LV (Figure 3). Those heat maps were extended for

16,480 genes in AA and 14,911 genes in LV and divided by each chromosome, displaying specificity or abundance of the heart-specific/enhanced protein-coding and non-coding genes in each chromosome compared to the rest of the genes (Supplementary Figures S1, S2).

Identification of Enriched Diseases and Biological Functions

Given the extreme specificity and abundance of the heart-specific/enhanced genes, pathological or functional implications of those genes were explored. Using IPA software, enriched “Diseases and Biological Functions” were identified in the following three groups of gene sets: 58 common-specific/enhanced protein-coding and non-coding genes (α), 81 AA-specific/enhanced protein-coding and non-coding genes ($\alpha + \beta$), and 79 LV-specific/enhanced protein-coding and non-coding genes ($\alpha + \gamma$) (Figure 4A). Based on the aforementioned cumulative enrichment test, more than 50 specific/enhanced genes were selected to produce enrichment results with IPA.

With common-specific/enhanced genes (α), out of a total of approximately 900 terms listed in “Diseases and Biological Functions,” 18 terms had *z*-scores in at least one tissue among 20 tissues. Among them, 5 functional terms (Function of Muscle, Cardiac Contractility, Contractility of Cardiac muscle, Cardiogenesis, and Contraction of Heart) were predicted to be activated (*z*-score ≥ 2) and 4 disease terms (Congestive Heart Failure, Fibrosis, Failure of Heart, and Organismal Death) were predicted to be inhibited (*z*-score ≤ -2), both when common-specific/enhanced genes were compared between AA and other tissues (Figure 4B) and between LV and other tissues (Supplementary Figures S3A,B). Most other tissues showed opposite prediction of activation or inhibition, and prediction was not significant for skeletal muscle (for all of the above 9 terms), testis (8 terms), spleen (5 terms) and coronary artery (2 terms) showing these diseases and biological functions were not oppositely inhibited or activated in those tissues (Figure 4B and Supplementary Figure S3B). Nonetheless, none of the other tissues showed the same pattern of prediction as the heart, indicating the exclusive role of common-specific/enhanced genes in promoting those biological functions and inhibiting those diseases.

With AA-specific/enhanced genes ($\alpha + \beta$), 6 terms were predicted to be activated (*z*-score ≥ 2) and four terms were predicted to be inhibited (*z*-score ≤ -2) in AA (Supplementary Figure S3C), and with LV-specific/enhanced genes ($\alpha + \gamma$), 7 terms were predicted to be activated (*z*-score ≥ 2) and 5 terms were predicted to be inhibited (*z*-score ≤ -2) in LV (Supplementary Figure S3D). Those terms were all related to activation in cardiac muscle contraction and cardiogenesis (functional terms) and inhibition of heart failure, fibrosis and organismal death (disease terms), as shown in the case of common-specific/enhanced genes. It suggested essential roles of common-specific/enhanced genes in promoting contraction of the heart and cardiac development and preventing dysfunction and failure of the heart.

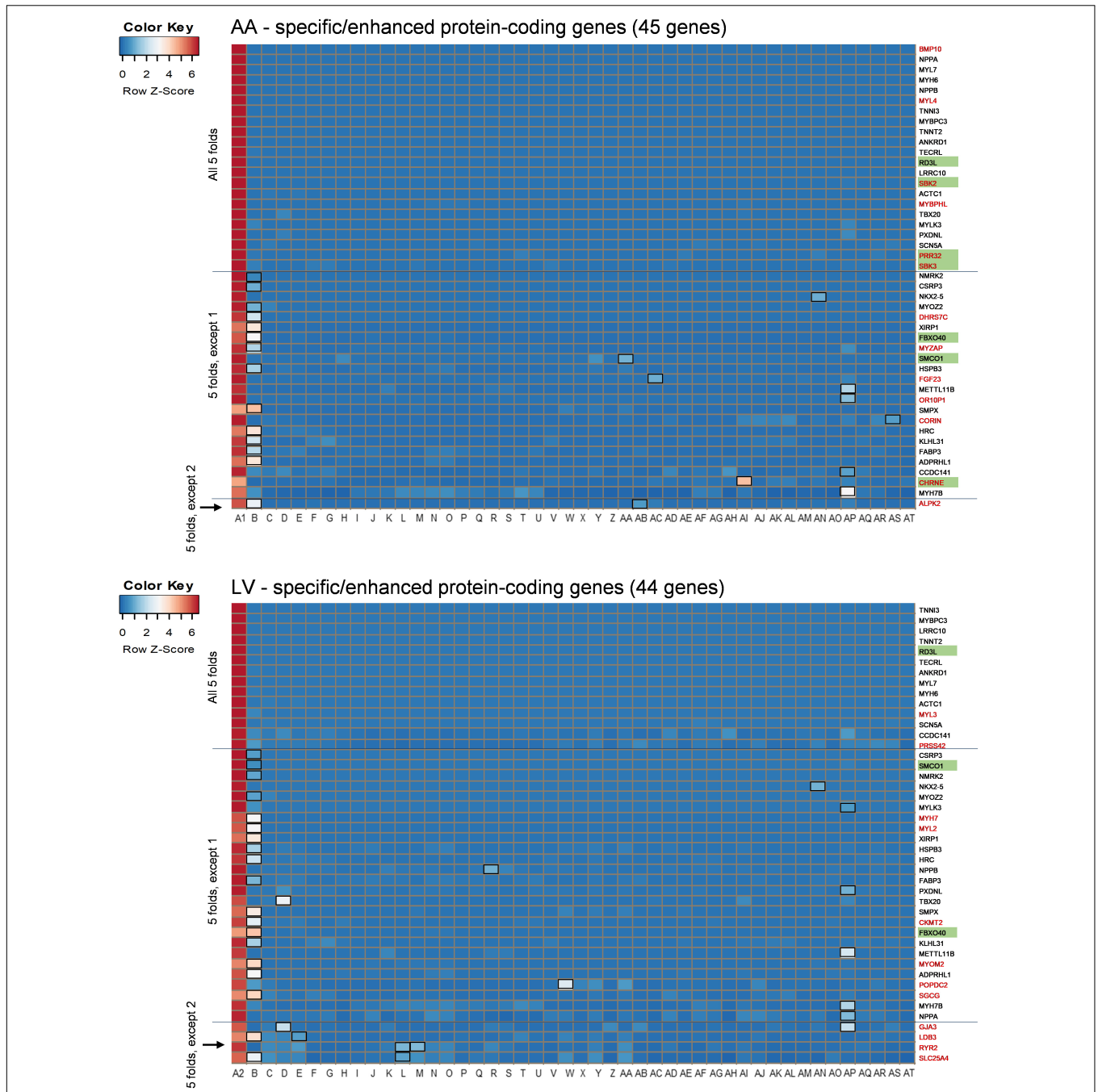
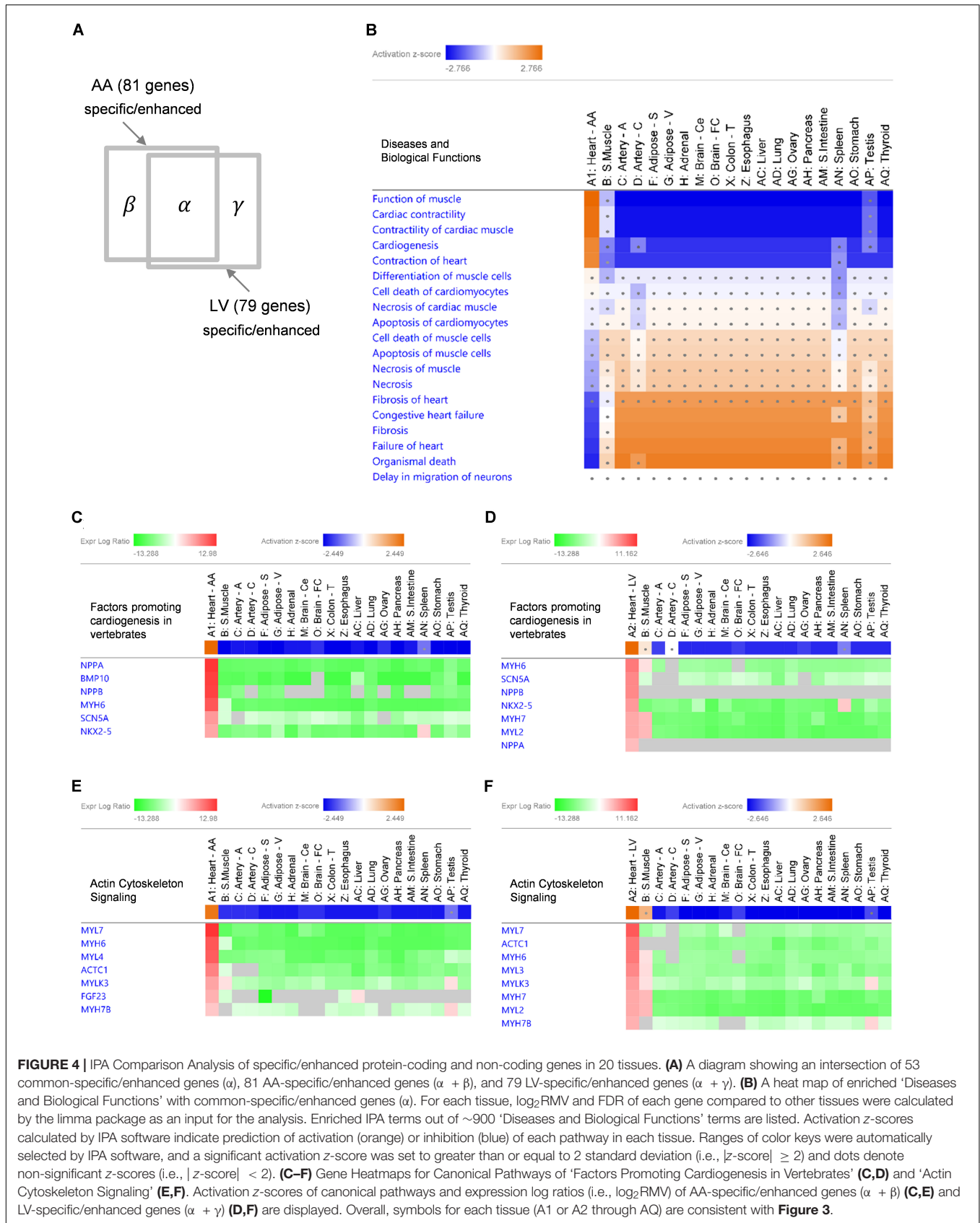


FIGURE 3 | Heat maps of heart-specific and heart-enhanced genes. Visualization is based on relative median expression levels of heart-specific (all median 5-folds) and heart-enhanced (median 5-folds, except 1 & median 5-folds, except 2) genes in 47 human tissues [A1: Heart – Atrial Appendage, A2: Heart – Left Ventricle, and B through AT: 45 non-heart tissues (B: Muscle – Skeletal, C: Artery – Aorta, D: Artery – Coronary, E: Artery – Tibial, F: Adipose – Subcutaneous, G: Adipose – Visceral (Omentum), H: Adrenal Gland, I: Brain – Amygdala, J: Brain – Anterior cingulate cortex (BA24), K: Brain – Caudate (basal ganglia), L: Brain – Cerebellar Hemisphere, M: Brain – Cerebellum, N: Brain – Cortex, O: Brain – Frontal Cortex (BA9), P: Brain – Hippocampus, Q: Brain – Hypothalamus, R: Brain – Nucleus accumbens (basal ganglia), S: Brain – Putamen (basal ganglia), T: Brain – Spinal cord (cervical c-1), U: Brain – Substantia nigra, V: Breast – Mammary Tissue, W: Colon – Sigmoid, X: Colon – Transverse, Y: Esophagus – Gastroesophageal Junction, Z: Esophagus – Mucosa, AA: Esophagus – Muscularis, AB: Kidney – Cortex, AC: Liver, AD: Lung, AE: Minor Salivary Gland, AF: Nerve – Tibial, AG: Ovary, AH: Pancreas, AI: Pituitary, AJ: Prostate, AK: Skin – Not Sun Exposed (Suprapubic), AL: Skin – Sun Exposed (Lower leg), AM: Small Intestine – Terminal Ileum, AN: Spleen, AO: Stomach, AP: Testis, AQ: Thyroid, AR: Uterus, AS: Vagina, AT: Whole Blood)]. Genes specific or enhanced in either atrial appendage or left ventricle are shown in red letters, and common genes overlapping between atrial appendage and left ventricle are denoted by black letters. Functionally unreported heart-specific/enhanced genes are indicated with green filled rectangles. Tissues with less than a 5-fold difference in medians are marked with black unfilled rectangles in the heat maps. Z-scores of RMVs were calculated by heatmap3 for each row (i.e., each gene). By selecting the heart-specific/enhanced genes, heat maps were skewed to positive z-scores compared to **Supplementary Figures S2, S3**. |z-score| ≥ 2 indicates at least 2 standard deviation from the mean.



Characteristics in Enriched Canonical Pathways

Specific canonical pathways were further explored to compare common-specific/enhanced genes with AA- or LV-specific/enhanced genes. From IPA, prediction of activation for “Factors Promoting Cardiogenesis in Vertebrates” were found to be higher in AA (*z*-score 2.45) and LV (*z*-score 2.65), compared to common-specific genes (*z*-score 2.24). The AA- or LV-specific/enhanced genes, that are not common-specific/enhanced and might increase those *z*-scores were *BMP10* in AA and *MYH7* (or β -*MHC*) and *MYL2* (or *MLC2v*) in LV (Figures 4C,D). In this study, the *BMP10* gene showed the highest specificity (Figure 3) but the expression ratio was slightly changed when comparing 20 tissues for IPA ranking *BMP10* the second in Figure 4C. The *MYH7* and *MYL2* genes were LV-enhanced genes (Figure 3).

In addition, in relation to cardiac muscle contraction, prediction of activation for another canonical pathway “Actin Cytoskeleton Signaling” was higher in AA (*z*-score 2.24) and LV (*z*-score 2.65), compared to common-specific genes (*z*-score 2.00). The AA- or LV-specific/enhanced genes, that are not common-specific/enhanced and might increase those *z*-scores were *FGF23* and *MYL4* in AA and *MYL3*, *MYH7* and *MYL2* in LV (Figures 4E,F).

For disease terms, prediction of inhibition was increased the most for “Organismal Death” with AA-specific/enhanced genes (*z*-score -3.05) and LV-specific/enhanced genes (*z*-score -3.07) compared to common-specific/enhanced genes (*z*-score -2.38). Also, prediction of inhibition for “Failure of Heart” (*z*-score -2.219) with common-specific/enhanced genes was increased with AA-specific/enhanced genes (*z*-score -2.43), and “Fibrosis” (*z*-score -2.09) with common-specific/enhanced genes was predicted to be inhibited more with LV-specific/enhanced genes (*z*-score -2.31). It suggests that those above canonical pathways related to cardiogenesis and cardiac muscle contraction and AA- and LV-specific/enhanced genes might be related to suppressing these pathological conditions and promoting organismal survival and function.

Atrial Appendage-Related BMP and FGF Signaling and Differences in Left Ventricle

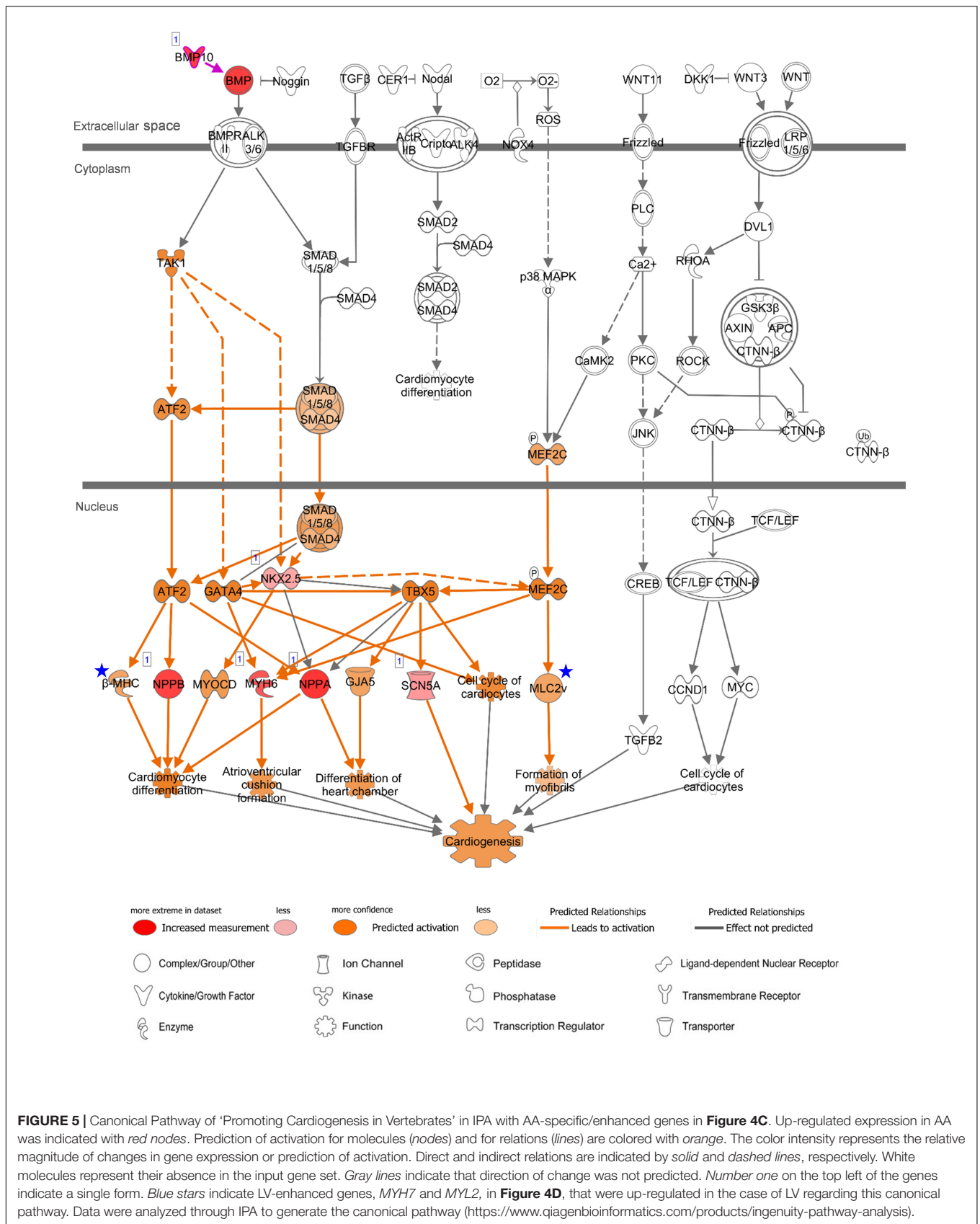
In order to gain functional and mechanistic insights, the above canonical pathways were further investigated. In relation to “Factors Promoting Cardiogenesis in Vertebrates,” bone morphogenetic protein 10 (*BMP10*) is a peptide growth factor and both SMAD-dependent and independent pathways were predicted to be activated via phosphorylation (Figure 5). A subsequent binding of SMAD1/5/8 and co-SMAD (*SMAD4*) and activation of *NKX2.5*, the determinant of cardiac development and the common-enhanced gene in our case, through both SMAD complex and *GATA-4* (Brown et al., 2004) might be a key determination step of cardiogenesis in atrial appendage. The role of *BMP10* in the adult heart remains elusive, but as reported recently *BMP10* may play multifunctional roles not only in promoting cardiac development in prenatal stages but also in inhibiting cardiomyocyte apoptosis and cardiac

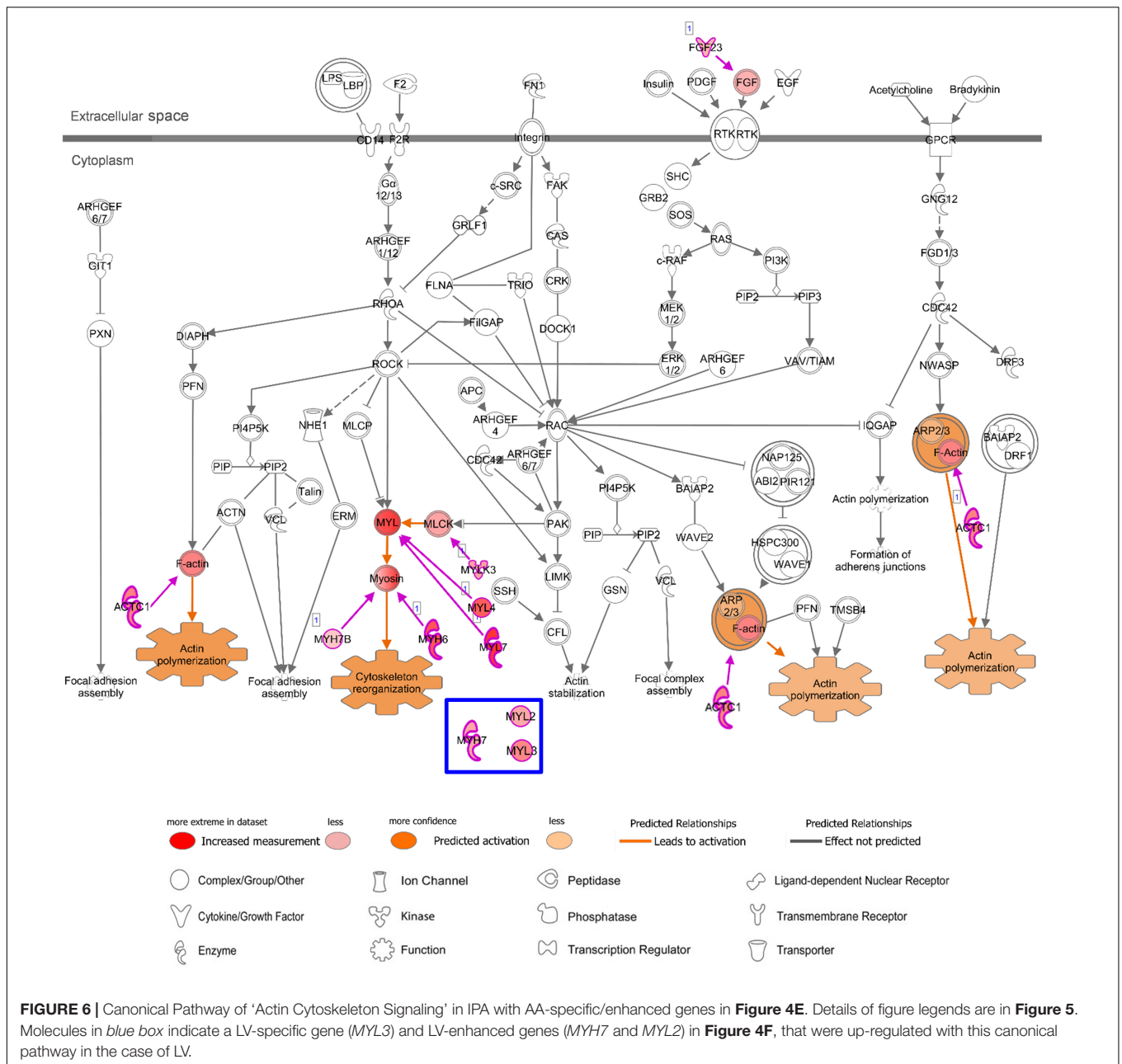
fibrosis in adults (Qu et al., 2019). In the downstream of this canonical pathway, cardiac transcription factors were predicted to activate target genes including AA-specific and LV-enhanced genes: natriuretic peptide B (*NPPB*) and natriuretic peptide A (*NPPA*). These paralogous genes are transcriptionally regulated as a cluster and encode peptide hormones that are secreted by cardiomyocytes (Man et al., 2018), and was predicted to induce cardiomyocyte differentiation (Figure 5). The expression of the other two target genes, sodium voltage-gated channel alpha subunit 5 (*SCN5A*) that encodes an integral membrane protein, a sodium channel, for transmitting electric signals to maintain normal heart contraction (Zaklyazminskaya and Dzemeshkevich, 2016) and myosin heavy chain 6 (*MYH6*) which encodes the sarcomeric filament protein alpha-myosin heavy chain (Posch et al., 2011), were relatively lower than *NPPB* and *NPPA*. In LV, additional genes include myosin heavy chain 7 (*MYH7* or β -*MHC*) that encodes beta-myosin heavy chain and myosin light chain 2 (*MYL2* or *MLC2v*) that encodes regulatory myosin light chain that are up-regulated as LV-enhanced genes.

In addition, regarding “Actin Cytoskeleton Signaling” which was predicted to be activated with heart-specific/enhanced genes, one of the genes that were additionally involved in this enriched pathway in AA was fibroblast growth factor 23 (*FGF23*), which is an AA-enhanced gene (Figure 6). Fibroblast growth factor 23 (*FGF23*) is a circulating hormone that is associated with cardiac hypertrophy and also acutely elevates intracellular Ca^{2+} and increases cardiac contractility (Touchberry et al., 2013). Binding of *FGF23* to receptor tyrosine kinases leads to tyrosine phosphorylation and stimulation of Ras/MARK and PI3K/Akt signaling (Grabner et al., 2015). Another additional gene in AA was myosin light chain 4 (*MYL4*), a previously known atrial-specific myosin light chain, whose mutation causes familial atrial fibrillation (Orr et al., 2016). In LV, *MYH7*, *MYL2*, and myosin light chain 3 (*MYL3*) were additionally involved in this canonical pathway. Although both AA-specific/enhanced genes and LV-specific/enhanced genes were predicted to promote actin polymerization and cytoskeleton reorganization, those different cardiac sarcomere protein genes might give rise to distinct contractile and electrophysiological properties between AA and LV.

Differential Expression Between Atrial Appendage and Left Ventricle

A total of 341 DEGs ($\text{Log}_2(\text{fold change}) \geq 3$, $\text{FDR} < 0.01$) were identified, among which 236 and 105 DEGs showed up-regulated expression in AA and LV, respectively, (Figure 7A and Supplementary Table S2). Protein-coding DEGs were 170 and 55 in atrial appendage and left ventricle, respectively, (Figure 7B). By doing IPA Core Analysis on 341 DEGs, twenty networks regarding diseases and functions were identified. Among them, the top ten networks were scored 19 to 42 [p -score = $-\log_{10}(p\text{-value})$] with 13 to 23 focus molecules (Supplementary Table S3). The first-ranked network was “Cardiovascular System Development and Function, Embryonic Development, Organ Development” with 23 focus molecules (p -score = 42). In this network, a subnetwork with the





most significant enrichment was “Morphogenesis of Cardiac Muscle” ($p = 8.96 \times 10^{-7}$) with 4 molecules: *BMP10*, *HAND1*, *MYL3*, and *XIRP2*. Among those four genes, *BMP10* was upregulated in AA ($\log_2FC = 12.38$), and the other three genes were downregulated in AA (*HAND1* $\log_2FC = -3.03$, *MYL3* $\log_2FC = -4.50$, and *XIRP2* $\log_2FC = -3.05$) and upregulated in LV. Another significant subnetwork was “Hypertrophy of Left Ventricle” ($p = 4.08 \times 10^{-5}$), and related molecules were *IRX4* ($\log_2FC = -7.33$), *MYL3* ($\log_2FC = -4.50$), *NPPA* ($\log_2FC = 6.05$), and *REN* ($\log_2FC = 7.20$).

The network between 23 focus molecules under the first-ranked “Cardiovascular System Development and

Function, Embryonic Development, Organ Development,” including the top two subnetwork molecules (the above seven molecules), was displayed (**Figure 7C**). In addition, the relationship of this network to canonical pathways was examined and the “Calcium Signaling” pathway showed a significant enrichment ($p = 3.31 \times 10^{-5}$) and the highest prediction of activation ($z\text{-score} = 2.24$) with this experimental group (the combination of up-regulated DEGs in AA and down-regulated DEGs in LV) (**Figure 7C**). Eight molecules (*MYL7*, *MYL1*, *MYL3*, *ERK1/2*, *Mlc*, *MYL4*, *MYL2*, and *Rlc*) were involved in this calcium signaling pathway. On the other hand, other canonical pathways that were related to AA- and LV-specific/enhanced genes

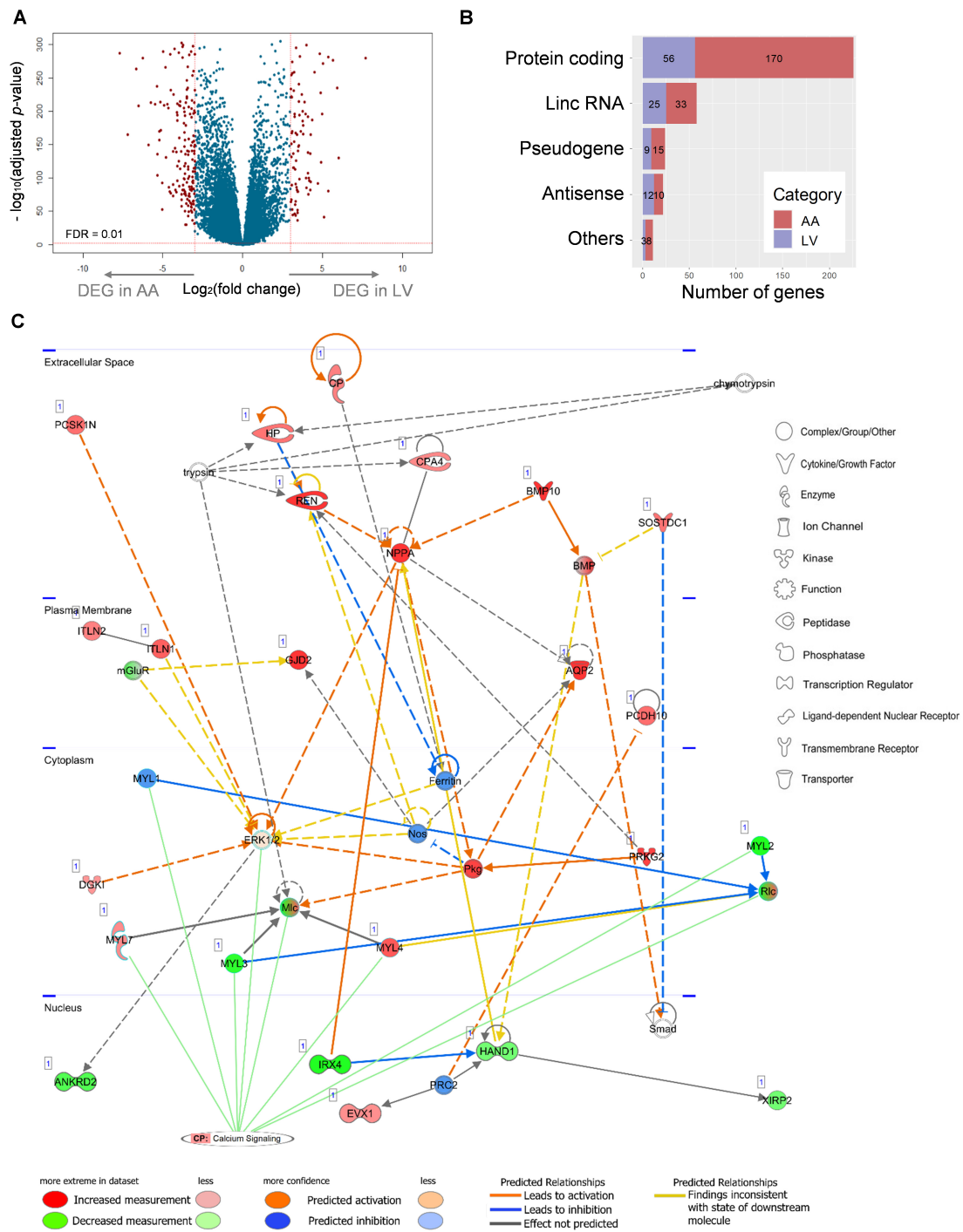


FIGURE 7 | Relationship between a network of DEGs and the canonical calcium signaling pathway. **(A)** Differential expression analysis of 23,610 genes and volcano plot representation of DEGs. The negative log₁₀-transformed adjusted *p*-value was plotted against the log ratios (fold change) between AA and LV. Red dots represent 236 and 105 DEGs in AA and LV, respectively, that show both at least 8-fold changes and high statistical significance (FDR < 0.01). **(B)** DEGs were categorized into five groups including protein-coding, lincRNA, pseudogene, antisense RNA, and others (processed transcript and sense intronic). Numbers of DEGs in each category are marked on the bar graph. **(C)** The top scoring IPANet network related to DEGs: 'Cardiovascular System Development and Function, Embryonic development, Organ development'. Up- (red) and down- (green) regulated genes (nodes) are indicated. Orange and blue lines indicate prediction of activation and inhibition, respectively, while gray lines indicate no prediction for direction of changes. Solid and dashed lines represent direct and indirect relations, respectively. Calcium signaling is shown as a pathway that has the highest prediction of activation. CP, canonical pathway. The networks were generated through IPANet (QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>).

were not significantly enriched or not predicted to be activated or inhibited: “Factors Promoting Cardiogenesis in Vertebrates” ($p = 0.06$, z -score = 0) and “Actin cytoskeleton signaling” ($p = 5.37 \times 10^{-3}$, z -score = -0.447). It suggests that there may be functional differences between heart-specific/enhanced genes and DEGs.

Novel Insights on Unreported Heart-Specific/Enhanced DEGs

In order to find heart-specific/enhanced genes that are differentially expressed between AA and LV and are previously unreported in PubMed, we first compared heart-specific/enhanced genes with DEGs. We found that 9 AA-specific/enhanced protein-coding genes were up-regulated DEGs in AA. Thus, these genes (*BMP10*, *MYBPHL*, *CHRNE*, *NPPA*, *SBK2*, *MYL4*, *OR10P1*, *PRR32*, and *MYL7*) were “AA-specific/enhanced DEGs” meaning that they were specific/enhanced to AA compared to 45 non-heart tissues as well as up-regulated in AA compared to LV (**Supplementary Table S2**). Also, there were 3 LV-specific/enhanced protein-coding genes that were up-regulated DEGs in LV (*MYL2*, *MYL3*, and *MYH7*), and they were named ‘LV-specific/enhanced DEGs’ which were specific/enhanced to LV compared to 45 non-heart tissues as well as up-regulated in LV compared to AA.

Among both AA- and LV-specific/enhanced DEGs, three AA-specific/enhanced DEGs [cholinergic receptor nicotinic epsilon subunit (*CHRNE*), SH3 domain binding kinase family member 2 (*SBK2*), and proline rich 32 (*PRR32*)] were “unreported” genes in PubMed regarding the heart. On the other hand, all the LV-specific/enhanced DEGs are known myosin genes. From the IPA Network Analysis for DEGs, networks for two out of the three AA-specific/enhanced DEGs were found (**Supplementary Table S3**). The *CHRNE* gene was one of the focus molecules under “Developmental Disorder, Hereditary Disorder, Immunological Disease” (p -score = 37 with 21 focus molecules) (**Figure 8**). Also, the *PRR32* gene was one of the focus molecules under “Cell Morphology, Embryonic Development, Hair and Skin Development and Function” (p -score = 19 with 13 focus molecules) (**Figure 9**).

CHRNE is a transmembrane acetylcholine receptor (AChR) and its mutations lead to congenital myasthenic syndrome which is responsible for perturbation on calcium signaling and post-synaptic Ca^{2+} accumulation (Zayas et al., 2007; Shen et al., 2018). Along with calcium voltage-gated channel alpha subunits (*CACNA1D*, *CACNA1G*, and *CACNA2D2* in **Figure 8**), the *CHRNE* expression may induce normal calcium flux. Subsequently, activation of AMPK via phosphorylation by calcium sensitive kinase CAMKK2 (Herzig and Shaw, 2018) may lead to prevention of cardiomyocyte injury in the atrial appendage, as shown in AMPK-mediated prevention of cellular injury (Zhao et al., 2013; Bi et al., 2015).

PRR32 is one of the proline-rich proteins that have been recognized as tanning-binding salivary proteins which counteracts a negative effect of a phenolic compound, tannin, in mammalian herbivores (Shimada, 2006). Although *PRR32* expression was AA-specific in this current study, the role of

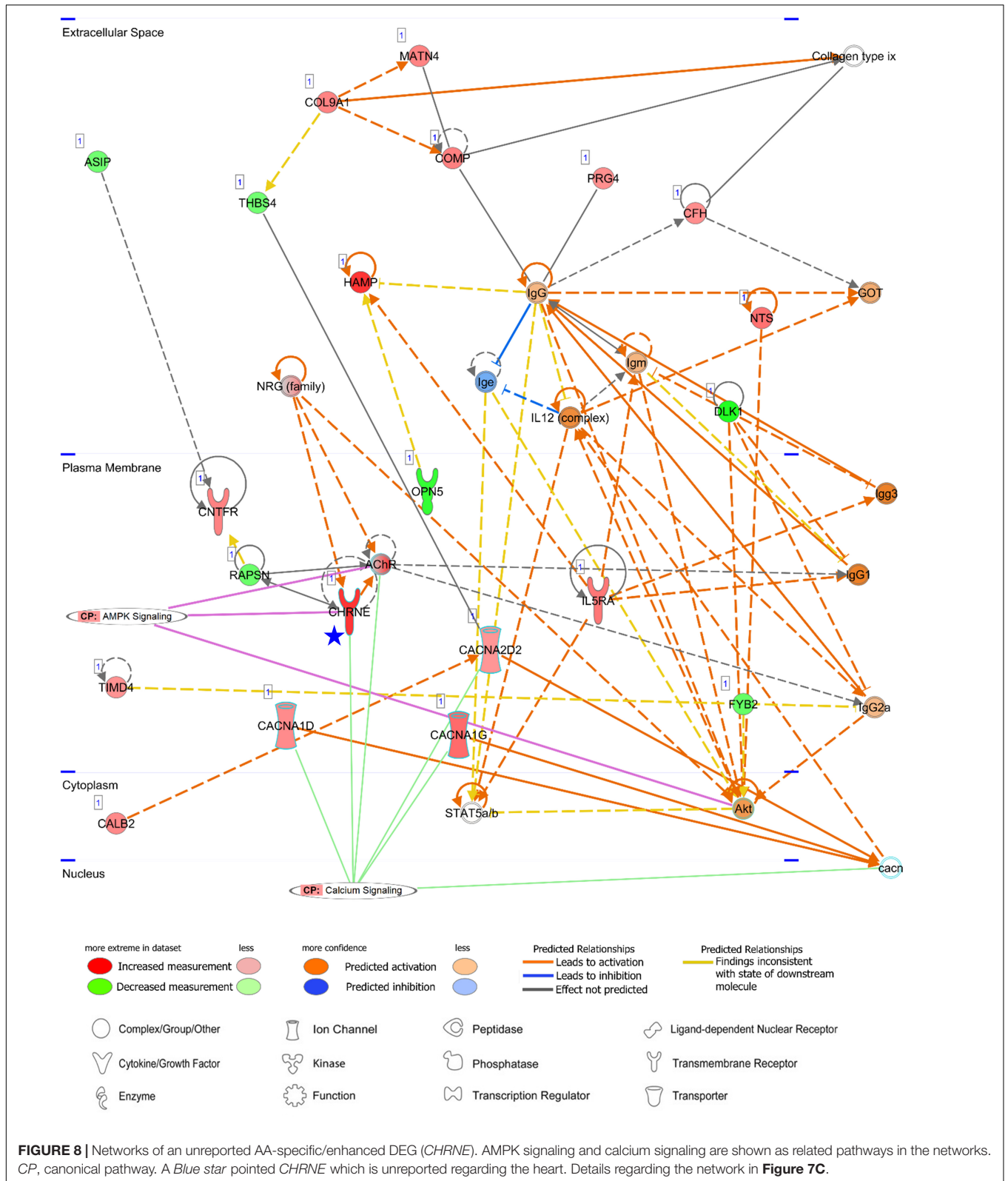
PRR32 is yet to be investigated in the heart, and also the relationship between *PRR32* and *ESR1* has not been established according to the network analysis (**Figure 9**). Estrogen receptor 1 (*ESR1*) is a nuclear receptor of 17beta-estradiol and *ESR1* activation in animal models exhibited anti-inflammatory effects (Puzianowska-Kuznicka, 2012). *ESR1* shared one toxicity-related list, “Cardiac Hypertrophy,” with proinflammatory cytokine tumor necrosis factor (TNF) (**Figure 9**). Because a relationship between *ESR1* and TNF has also not been established (**Figure 9**) but expression of TNF in the heart leads to cardiac pathogenesis, the role of *ESR1* and *PRR32* will need to be further investigated. Additionally, network analyses were conducted for three unreported common-specific/enhanced genes, and results were presented in **Supplementary Figures S4–S6**. Mechanistic insights on these unreported genes were provided in the “Discussion” section.

DISCUSSION

The comprehensive human tissue transcriptome provided by the ongoing GTEx project was used to screen human heart-related genes including heart-specific genes, heart-enhanced genes, and DEGs, based on our previously developed method (Ahn et al., 2019). These transcriptomic data enabled large-scale identification of novel heart-specific/enhanced genes for further investigation on evidence of their relation to heart diseases and heart-related phenotypes. These approaches combined with pathway analysis provided a robust and integrative strategy for the consolidation of heart-related functions and diseases.

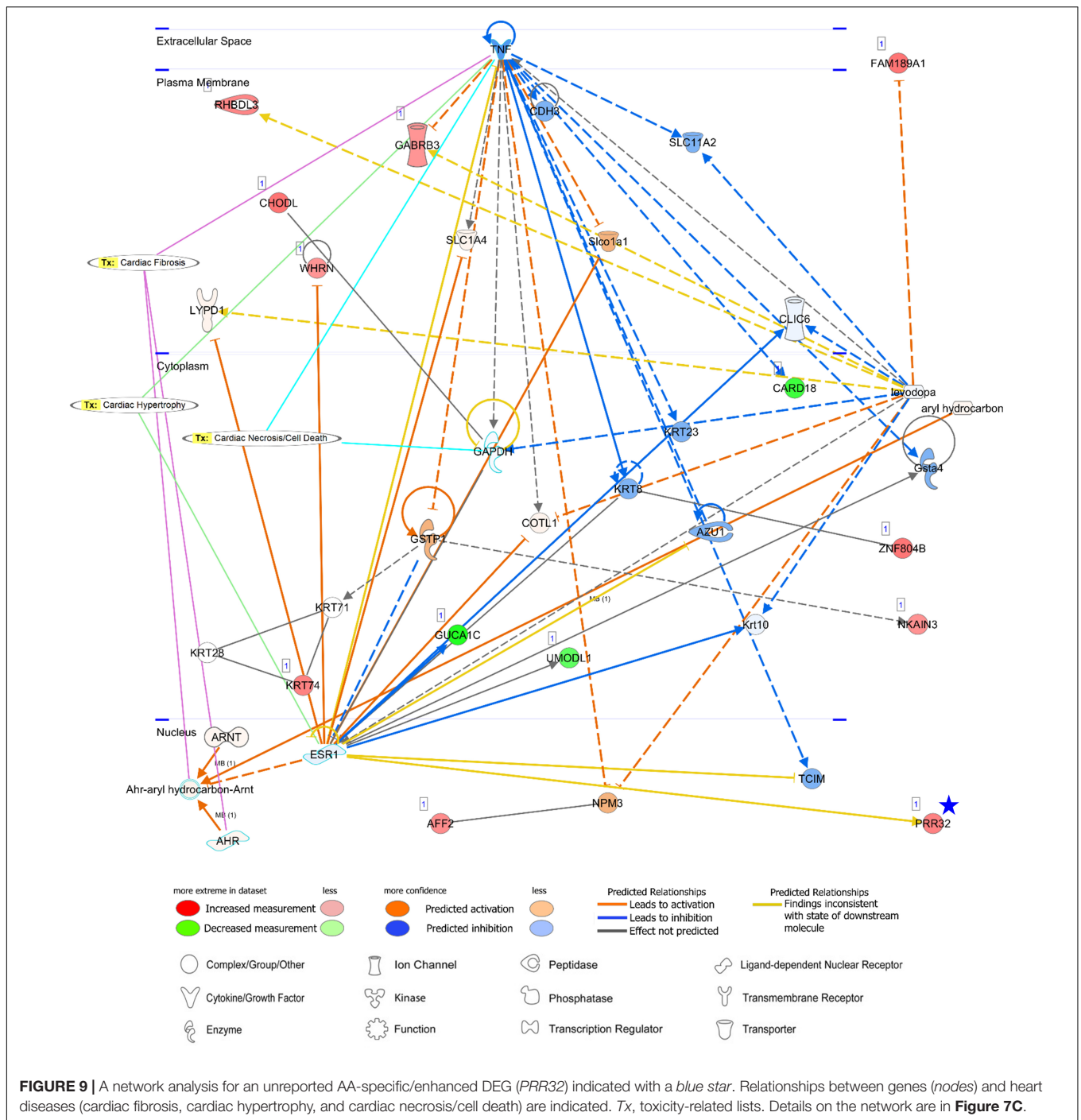
In this study, more than 70% of genes specific/enhanced to atrial appendage (32 out of 45) and left ventricle (32 out of 44) were overlapping genes, suggesting those common heart-specific/enhanced genes are involved in essential biological functions of the heart. According to our literature mining, the majority of those common-specific/enhanced genes have been reported in PubMed in relation to their functions and associated diseases in the heart; however, among the 32 common-specific/enhanced genes, three genes (*FBXO40*, *RD3L*, and *SMCO1*) have not been reported regarding the heart. Among non-overlapping AA-specific/enhanced protein-coding genes, four genes were unreported genes (*SBK2*, *PRR32*, *SBK3*, and *CHRNE*). “Diseases and Biological Functions” enriched with common-specific genes, AA-, and LV-specific/enhanced genes were apparently similar, but their canonical pathways showed differences regarding the significance in prediction of activation or inhibition and participating genes in the pathways. Since our gene sets were highly specific, but relatively small, many of the pathway genes might be excluded from the analysis. However, two distinct upstream genes of canonical pathways were found to be AA-specific/enhanced (*BMP10* and *FGF23*), and differences in downstream molecules were detected which may provide characteristics of the atrial appendage and left ventricle.

In order to increase specificity in either the atrial appendage or left ventricle, we matched specific/enhanced genes with DEGs and found unreported AA-specific and differentially upregulated three genes (*PRR32*, *SBK2*, and *CHRNE*). The epsilon



subunit of acetylcholine receptor, *CHRNE*, is a transmembrane protein and mutations in *CHRNE* have been associated with congenital myasthenic syndrome which is characterized by

compromised neuromuscular transmission and muscle weakness (Shen et al., 2018). This syndrome perturbs the kinetics of synaptic currents, leading to post synaptic Ca^{2+} accumulation



(Zayas et al., 2007). The neurotransmitter, acetylcholine, prevents hypoxia-induced cellular injury by activating AMP-activated protein kinase (AMPK) (Zhao et al., 2013; Bi et al., 2015). AMPK is activated via phosphorylation by calcium sensitive kinase CAMKK2 in response to calcium flux, thus calcium signaling is linked to the regulation of energy metabolism by AMPK (Herzig and Shaw, 2018). Therefore, CHRNE/calcium/AMPK signaling cascades may play a role in preventing cardiomyocyte injury in AA, along with calcium voltage-gated channel alpha subunits

(CACNA1D, CACNA1G, and CACNA2D2) shown in **Figure 8**. Also, activation of AMPK via phosphorylation by calcium sensitive kinase CAMKK2 has been reported to be independent from upstream kinases for AMPK (Herzig and Shaw, 2018).

Regarding a tannin suppressant *PRR32*, *ESR1* activation in animal models of cardiac ischemia showed cardioprotective roles of *ESR1* including reduction of inflammation, oxidative stress, and apoptosis of cardiomyocytes (Puzianowska-Kuznicka, 2012), but currently identified relations between *ESR1* and

PRR32 are inconsistent according to the network analysis. In addition, although *ESR1* might have anti-inflammatory effects, a relationship between *ESR1* and *TNF* is also not established based on the network. *TNF* was linked to three toxicity-related lists generated by IPA: “Cardiac Fibrosis,” “Cardiac Hypertrophy” and “Cardiac Necrosis/Cell Death” and among them “Cardiac Hypertrophy” was shared between *TNF* and *ESR1*. Because expression of proinflammatory cytokine tumor necrosis factor (*TNF*) in the heart leads to cardiac pathogenesis and congestive heart failure, and therefore attenuation or inhibition of *TNF* expression has been investigated as clinical strategies (Feldman et al., 2000; Hartman et al., 2018), the unknown function of *ESR1* and *PRR32* gains attention. In addition, *SBKs* are kinases bind to the Src homology 3 (SH3) domain which is involved in various cellular functions such as cell proliferation, cytoskeletal modifications, and signal transduction (Kurochkina and Guha, 2013). In this study, *SBK2* was more AA-specific than *SBK3*, but pathways related to *SBK2* were not identified. The specific roles of these kinases in the atrial appendage are yet to be investigated.

Among the unreported common-specific/enhanced genes, F-box protein 40 (*FBXO40*) encodes a protein that forms SCF ubiquitin ligase complex, with *SKP1* and Cullin proteins, which is involved in protein ubiquitination (Jin et al., 2004). An isoform of heat shock protein 90 (*HSP90AA1*), that was linked to *FBXO40* (Supplementary Figure S4), plays roles in degrading proteins as a molecular chaperone and stabilizing TGF- β signaling cascade (Garcia et al., 2016). The interaction between *FBXO40* and *HSP90AA1* may occur through the ubiquitin ligase complex (Ehrlich et al., 2009), but their direct roles in the heart need further investigation. Another common-specific gene, retinal degeneration 3 like (*RD3L*), encodes a domain of retinol degeneration (RD) 3 protein which is associated with retinal degeneration (Friedman et al., 2006). The function of *RD3L* remains unknown in the heart, despite its specific cardiac expression shown in this study. In the network analysis, relations were found between *RD3L* and heat shock protein family A (Hsp70) member 9 (*HSPA9*) and between *RD3L* and WD repeat containing, antisense to TP73 (*WRAP73*), but the directions of influences between each other (activation or inhibition) were not predicted (Supplementary Figure S5). Among them, the *HSPA9* gene encodes a mitochondrial chaperone and its mutations were identified in patients with cardiac malformation (Royer-Bertrand et al., 2015), and this gene was linked to eNOS signaling which has a protective function in the cardiovascular system by regulating the diameter of blood vessels (Daiber et al., 2019) (Supplementary Figure S5). However, the relation was not statistically significant. In addition, single-pass membrane protein with coiled-coil domains 1 (*SMCO1*) was related to a transcription factor, cAMP-responsive element-binding protein (CREB), without prediction on the direction of influence (Supplementary Figure S6). The role of *SMCO1* is largely unknown, but its unknown interaction with CREB1 might be important because CREB1 is implicated in the regulation of cardiac pathophysiology (Matus et al., 2007).

Among identified DEGs between AA and LV, *BMP10*, *HAND1*, *MYL3*, and *XIRP2* were significantly enriched with “Morphogenesis of Cardiac Muscle” under the first-ranked

“Cardiovascular System Development and Function, Embryonic Development, Organ Development” network. The expression of *BMP10* was upregulated in AA and the other three genes were upregulated in LV. The *BMP10* gene has been reported as the most highly expressed gene in the right atrial appendage compared to the left atrial appendage (Kahr et al., 2011) and the GTEx sampling site was the right atrial appendage. The AA-specific expression of the *BMP* gene in our results further revealed its specificity in the right atrial appendage compared to all other analyzed tissues including LV. It was also consistent with previous studies that reported *HAND1* is a transcription factor whose expression is restricted to the left ventricle and contributes to chamber formation (Huang et al., 2004; Firulli et al., 2010). The *MYL3* gene is one of the cardiac sarcomeric genes and encodes the ventricular form of myosin light chain (Cohen-Haguenaer et al., 1989). α -actinin binding repeat containing 2 (*XIRP2*) is localized to the intercalate disks (ICDs) in cardiac muscle and is an actin-binding protein that interacts with actin filaments through conserved Xin repeats. *XIRP2* may regulate cardiac ion channel by modulating the actin cytoskeleton, and its deficiency has been shown to be associated with cardiac conduction (Wang et al., 2014; Huang et al., 2018). Another significant subnetwork, “Hypertrophy of Left Ventricle” involved *NPPA* and *REN* (up-regulated in AA) and *IRX4* and *MYL3* (up-regulated in LV). The renin (*REN*) gene encodes protease renin which plays a central role in regulating blood pressure, and activation of the renin-angiotensin system is linked to the development of left ventricular hypertrophy (Cowan and Young, 2009). The role of the iroquois homeobox 4A (*IRX4*) gene has been reported in regulating chamber-specific expression of myosin isoforms and ventricular differentiation in animal models, and potential causal effects of its mutation in a ventricular septal defect in humans have been proposed (Bao et al., 1999; Bruneau et al., 2001; Cheng et al., 2011). As mentioned earlier, *NPPA* forms a cluster with *NPPB* and encodes a peptide hormone, but *NPPA* was more specific to AA compared to *NPPB*. The *MYL3* gene was both up-regulated in DEG in the LV and LV-specific gene, indicating its role as a left ventricle-specific myosin protein. Also, a canonical pathway that was predicted to be activated the most within the first-ranked network was “Calcium Signaling.” Considering the most enriched canonical pathways of the heart-specific/enhanced gene were not predicted to be activated or inhibited with DEGs, functional roles of the gene sets of heart-specific/enhanced genes and DEGs may be distinctive.

Taken together, this study reports on comprehensive identification of heart-specific genes, heart-enhanced genes and DEGs, and their relations with pathways associated with heart-related traits and diseases. Limitations to the present study that needs to be addressed include age of donors and stringent selection criteria. Given the donors aged 21–70 for the GTEx project, our identification of heart-specific genes and DEGs was limited to the adult stage. Thus, potentially important developmental gene expression in earlier stages could not be examined. In addition, we cannot rule out that some potentially important heart-related genes might not be incorporated in the current study due to stringent selection

criteria (median 5-fold; FDR < 0.01); however, heart-specific genes identified in this study will need to be prioritized for further functional evaluations. In conclusion, our findings on novel heart-specific/enhanced genes and their functional implications, together with their pathways, further provide new insights into biological underpinnings of cardiac symptoms and diseases. To enhance our understanding on the genetic architecture underlying cardiac physiology and pathophysiology, continued attempts need to be made on further elucidation of functional aspects of those heart-specific genes while ultimately providing potential targets for cardiac gene therapy.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

KL conceived the study. JA, HW, and KL developed the study designed and analyzed and interpreted the data. JA and HW performed the data processing and statistical analysis and drafted and wrote the manuscript. JA conducted the pathway analysis. All authors read, edited, and approved the final manuscript.

REFERENCES

- Ahn, J., Wu, H., and Lee, K. (2019). Integrative analysis revealing human adipose-specific genes and consolidating obesity loci. *Sci. Rep.* 9:3087.
- Bao, Z. Z., Bruneau, B. G., Seidman, J. G., Seidman, C. E., and Cepko, C. L. (1999). Regulation of chamber-specific gene expression in the developing heart by *Irx4*. *Science* 283, 1161–1164. doi: 10.1126/science.283.5405.1161
- Barth, A. S., Merk, S., Arnoldi, E., Zwermann, L., Kloos, P., Gebauer, M., et al. (2005). Functional profiling of human atrial and ventricular gene expression. *Pflugers. Arch.* 450, 201–208. doi: 10.1007/s00424-005-1404-8
- Benjamini, A., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bi, X., He, X., Xu, M., Zhao, M., Yu, X., Lu, X., et al. (2015). Acetylcholine ameliorates endoplasmic reticulum stress in endothelial cells after hypoxia/reoxygenation via M3 AChR-AMPK signaling. *Cell Cycle* 14, 2461–2472. doi: 10.1080/15384101.2015.1060383
- Brown, C. O. III, Chi, X., Garcia-Gras, E., Shirai, M., Feng, X. H., and Schwartz, R. J. (2004). The cardiac determination factor, *Nkx2-5*, is activated by mutual cofactors *GATA-4* and *Smad1/4* via a novel upstream enhancer. *J. Biol. Chem.* 279, 10659–10669. doi: 10.1074/jbc.m301648200
- Bruneau, B. G., Bao, Z. Z., Fatkin, D., Xavier-Neto, J., Georgakopoulos, D., Maguire, C. T., et al. (2001). Cardiomyopathy in *Irx4*-deficient mice is preceded by abnormal ventricular gene expression. *Mol. Cell. Biol.* 21, 1730–1736. doi: 10.1128/mcb.21.5.1730-1736.2001
- Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012.
- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14:128. doi: 10.1186/1471-2105-14-128
- Cheng, Z., Wang, J., Su, D., Pan, H., Huang, G., Li, X., et al. (2011). Two novel mutations of the *IRX4* gene in patients with congenital heart disease. *Hum. Genet.* 130, 657–662. doi: 10.1007/s00439-011-0996-7

FUNDING

This work was partially supported by the United States Department of Agriculture National Institute of Food and Agriculture Hatch Grant (Project No. OHO01304), the Ohio Agricultural Research and Development Center Research SEEDS funding (Project No. 2017-069), and the National Natural Science Foundation of China (Project No. 31660648).

ACKNOWLEDGMENTS

The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 02/22/2019. We thank Dr. Michael Townsend from The Ohio State University Comprehensive Cancer Center for allowing access to Ingenuity Pathway Analysis. We also thank Ms. Michelle Milligan for her assistance in proofreading this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00777/full#supplementary-material>

- Cohen-Haguener, O., Barton, P. J., Van Cong, N., Cohen, A., Masset, M., Buckingham, M., et al. (1989). Chromosomal assignment of two myosin alkali light-chain genes encoding the ventricular/slow skeletal muscle isoform and the atrial/fetal muscle isoform (*MYL3*, *MYL4*). *Hum. Genet.* 81, 278–282. doi: 10.1007/bf00279004
- Cowan, B. R., and Young, A. A. (2009). Left ventricular hypertrophy and renin-angiotensin system blockade. *Curr. Hypertens. Rep* 11, 167–172.
- Daiber, A., Xia, N., Steven, S., Oelze, M., Hanf, A., Kroller-Schon, S., et al. (2019). New therapeutic implications of endothelial nitric oxide synthase (eNOS) function/dysfunction in cardiovascular disease. *Int. J. Mol. Sci.* 20:187. doi: 10.3390/ijms20010187
- Ehrlich, E. S., Wang, T., Luo, K., Xiao, Z., Niewiadomska, A. M., Martinez, T., et al. (2009). Regulation of Hsp90 client proteins by a Cullin5-RING E3 ubiquitin ligase. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20330–20335. doi: 10.1073/pnas.0810571106
- Feldman, A. M., Combes, A., Wagner, D., Kadakomi, T., Kubota, T., Li, Y. Y., et al. (2000). The role of tumor necrosis factor in the pathophysiology of heart failure. *J. Am. Coll. Cardiol.* 35, 537–544.
- Firulli, A. B., Firulli, B. A., Wang, J., Rogers, R. H., and Conway, S. J. (2010). Gene replacement strategies to test the functional redundancy of basic helix-loop-helix transcription factor. *Pediatr. Cardiol.* 31, 438–448. doi: 10.1007/s00246-010-9669-x
- Friedman, J. S., Chang, B., Kannabiran, C., Chakarova, C., Singh, H. P., Jalali, S., et al. (2006). Premature truncation of a novel protein, RD3, exhibiting subnuclear localization is associated with retinal degeneration. *Am. J. Hum. Genet.* 79, 1059–1070. doi: 10.1086/510021
- Garcia, R., Merino, D., Gomez, J. M., Nistal, J. F., Hurle, M. A., Cortajarena, A. L., et al. (2016). Extracellular heat shock protein 90 binding to TGFbeta receptor I participates in TGFbeta-mediated collagen production in myocardial fibroblasts. *Cell. Signal.* 28, 1563–1579. doi: 10.1016/j.cellsig.2016.07.003
- Grabner, A., Amaral, A. P., Schramm, K., Singh, S., Sloan, A., Yanucil, C., et al. (2015). Activation of cardiac fibroblast growth factor receptor 4 causes left ventricular hypertrophy. *Cell Metab.* 22, 1020–1032.
- GTEx Consortium (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. doi: 10.1038/nature24277

- Hartman, M. H. T., Groot, H. E., Leach, I. M., Karper, J. C., and Van Der Harst, P. (2018). Translational overview of cytokine inhibition in acute myocardial infarction and chronic heart failure. *Trends Cardiovasc. Med.* 28, 369–379. doi: 10.1016/j.tcm.2018.02.003
- Herzig, S., and Shaw, R. J. (2018). AMPK: guardian of metabolism and mitochondrial homeostasis. *Nat. Rev. Mol. Cell Biol.* 19, 121–135. doi: 10.1038/nrm.2017.95
- Huang, F., Wagner, M., and Siddiqui, M. A. (2004). Ablation of the CLP-1 gene leads to down-regulation of the HAND1 gene and abnormality of the left ventricle of the heart and fetal death. *Mech. Dev.* 121, 559–572. doi: 10.1016/j.mod.2004.04.012
- Huang, L., Wu, K. H., Zhang, L., Wang, Q., Tang, S., Wu, Q., et al. (2018). Critical roles of xirp proteins in cardiac conduction and their rare variants identified in sudden unexplained nocturnal death syndrome and brugada syndrome in chinese han population. *J. Am. Heart Assoc.* 7:e006320.
- Jin, J., Cardozo, T., Lovering, R. C., Elledge, S. J., Pagano, M., and Harper, J. W. (2004). Systematic analysis and nomenclature of mammalian F-box proteins. *Genes Dev.* 18, 2573–2580. doi: 10.1101/gad.1255304
- Kahr, P. C., Piccini, I., Fabritz, L., Greber, B., Scholer, H., Scheld, H. H., et al. (2011). Systematic analysis of gene expression differences between left and right atria in different mouse strains and in human atrial tissue. *PLoS One* 6:e26389. doi: 10.1371/journal.pone.0026389
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361.
- Kramer, A., Green, J., Pollard, J. Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97.
- Kurochkina, N., and Guha, U. (2013). SH3 domains: modules of protein-protein interactions. *Biophys. Rev.* 5, 29–39. doi: 10.1007/s12551-012-0081-z
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.
- Man, J., Barnett, P., and Christoffels, V. M. (2018). Structure and function of the Nppa-Nppb cluster locus during heart development and disease. *Cell Mol. Life Sci.* 75, 1435–1444. doi: 10.1007/s00018-017-2737-0
- Matus, M., Lewin, G., Stumpel, F., Buchwalow, I. B., Schneider, M. D., Schutz, G., et al. (2007). Cardiomyocyte-specific inactivation of transcription factor CREB in mice. *FASEB J.* 21, 1884–1892. doi: 10.1096/fj.06-7915com
- Olson, E. N. (2006). Gene regulatory networks in the evolution and development of the heart. *Science* 313, 1922–1927. doi: 10.1126/science.1132292
- Orr, N., Arnaout, R., Gula, L. J., Spears, D. A., Leong-Sit, P., Li, Q., et al. (2016). A mutation in the atrial-specific myosin light chain gene (MYL4) causes familial atrial fibrillation. *Nat. Commun.* 7:11303.
- Posch, M. G., Waldmuller, S., Muller, M., Scheffold, T., Fournier, D., Andrade-Navarro, M. A., et al. (2011). Cardiac alpha-myosin (MYH6) is the predominant sarcomeric disease gene for familial atrial septal defects. *PLoS One* 6:e28872. doi: 10.1371/journal.pone.0028872
- Pospisil, P., Iyer, L. K., Adelstein, S. J., and Kassir, A. I. (2006). A combined approach to data mining of textual and structured data to identify cancer-related targets. *BMC Bioinformatics* 7:354. doi: 10.1186/1471-2105-7-354
- Puzianowska-Kuznicka, M. (2012). ESR1 in myocardial infarction. *Clin. Chim. Acta* 413, 81–87. doi: 10.1016/j.cca.2011.10.028
- Qu, X., Liu, Y., Cao, D., Chen, J., Liu, Z., Ji, H., et al. (2019). BMP10 preserves cardiac function through its dual activation of SMAD-mediated and STAT3-mediated pathways. *J. Biol. Chem.* 294, 19877–19888. doi: 10.1074/jbc.ra119.010943
- Rani, J., Shah, A. B., and Ramachandran, S. (2015). pubmed.mineR: an R package with text-mining algorithms to analyse PubMed abstracts. *J. Biosci.* 40, 671–682. doi: 10.1007/s12038-015-9552-2
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Royer-Bertrand, B., Castillo-Taucher, S., Moreno-Salinas, R., Cho, T. J., Chae, J. H., Choi, M., et al. (2015). Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. *Sci. Rep.* 5:17154.
- Shen, X. M., Brengman, J. M., Shen, S., Durmus, H., Preethish-Kumar, V., Yuceyar, N., et al. (2018). Mutations causing congenital myasthenia reveal principal coupling pathway in the acetylcholine receptor epsilon-subunit. *JCI Insight* 3:e97826.
- Shimada, T. (2006). Salivary proteins as a defense against dietary tannins. *J. Chem. Ecol.* 32, 1149–1163. doi: 10.1007/s10886-006-9077-0
- Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3:3.
- Touchberry, C. D., Green, T. M., Tchikrizov, V., Mannix, J. E., Mao, T. F., Carney, B. W., et al. (2013). FGF23 is a novel regulator of intracellular calcium and cardiac contractility in addition to cardiac hypertrophy. *Am. J. Physiol. Endocrinol. Metab.* 304, E863–E873.
- van de Vegte, Y. J., Teegene, B. S., Verweij, N., Snieder, H., and Van Der Harst, P. (2019). Genetics and the heart rate response to exercise. *Cell. Mol. Life Sci.* 76, 2391–2409.
- Wang, Q., Lin, J. L., Erives, A. J., Lin, C. I., and Lin, J. J. (2014). New insights into the roles of Xin repeat-containing proteins in cardiac development, function, and disease. *Int. Rev. Cell. Mol. Biol.* 310, 89–128. doi: 10.1016/b978-0-12-800180-6.00003-7
- Zachariou, M., Minadakis, G., Oulas, A., Afxenti, S., and Spyrou, G. M. (2018). Integrating multi-source information on a single network to detect disease-related clusters of molecular mechanisms. *J. Proteomics* 188, 15–29. doi: 10.1016/j.jprot.2018.03.009
- Zaklyazminskaya, E., and Dzemeshevich, S. (2016). The role of mutations in the SCN5A gene in cardiomyopathies. *Biochim. Biophys. Acta* 1863, 1799–1805. doi: 10.1016/j.bbamcr.2016.02.014
- Zayas, R., Groshong, J. S., and Gomez, C. M. (2007). Inositol-1,4,5-triphosphate receptors mediate activity-induced synaptic Ca²⁺ signals in muscle fibers and Ca²⁺ overload in slow-channel syndrome. *Cell Calcium* 41, 343–352. doi: 10.1016/j.ceca.2006.07.007
- Zhao, M., Sun, L., Yu, X. J., Miao, Y., Liu, J. J., Wang, H., et al. (2013). Acetylcholine mediates AMPK-dependent autophagic cytoprotection in H9c2 cells during hypoxia/reoxygenation injury. *Cell. Physiol. Biochem.* 32, 601–613. doi: 10.1159/000354464
- Zhao, S., Guo, Y., Sheng, Q., and Shyr, Y. (2014). Advanced heat map and clustering analysis using heatmap3. *Biomed. Res. Int.* 2014:986048.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ahn, Wu and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.