Transcriptomic Predictors of Pulmonary Hypertension Development in Heart Failure Patients

BY

MEGHAN JANE ARWOOD
B.S., University of Florida, 2010
Pharm.D., University of Florida, 2013

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Science in Clinical and Translational Science
in the Graduate College of the
University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:

Jack Zwanziger, Chair and Advisor
Julio Duarte, University of Florida
Roberto Machado, Indiana University

This thesis is dedicated to my soulmate John, my mother Michelle, my father George, and my twin brother Matthew, all of whom provided invaluable emotional support.

ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Julio Duarte, for his unwavering support and mentorship throughout this entire process. His intellect, wit, and passion for pharmacogenomics have made my training with him an incredible experience. I am thankful that he encouraged me to take on such a complex project, as it taught me important bioinformatics and statistical analysis skills. Furthermore, I would like to thank Dr. Jack Zwanziger and Dr. Roberto Machado for serving on my committee and for their valuable guidance in this endeavor.

This study was made possible through extracting medical record data and blood samples from patients within two databases, thus, I would like to acknowledge the following individuals who were instrumental in the design, recruitment, sample collection/processing, data collection, and/or reading echocardiograms for these two repositories: Dr. Larisa Cavallari, Dr. Thomas Stamos, Dr. Christopher Gans, Dr. Vicki Groo, Shitalben Patel, Dr. Ina Liko, Dr. Mayank Kansal, and Dr. Ankit Desai.

Lastly, I would like to thank the individuals who were involved with the RNA sequencing aspect of this project, including Dr. Stefan Green and Dr. Zarema Arbieva for their work in processing and sequencing our samples, Dr. Cécile Pereira, Dr. Ana Conesa, Dr. Ana Caroline Costa Sá, and Dr. Nasim Vahabi for their significant guidance on my scripts for analysis, and Christelle Lteif for her work on validation. Also, I would like to acknowledge the University of Florida Research Computing for providing computational resources and support that have contributed to the research results reported in this thesis.

MJA

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

LIST OF ABBREVIATIONS

CO	Cardiac output
cDNA	Complementary DNA
CpcPH	Combined pre-and post-capillary pulmonary hypertension
CPM	Count per million
DAVID	Database for annotation, visualization and integrated discovery
DPG	Diastolic pulmonary gradient
FDA	Food and Drug Administration
FDR	False discovery rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GAGE	Generally Applicable Gene-set Enrichment
GBPH	Genomics and Biomarkers of Pulmonary Hypertension database
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study
HF	Heart failure
HFGDB	Heart Failure Genomics Database
HFpEF	Heart failure with preserved ejection fraction
HFrEF	Heart failure with reduced ejection fraction
HWE	Hardy-Weinberg equilibrium
IpcPH	Isolated post-capillary pulmonary hypertension
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	Linkage disequilibrium
logFC	Log_2-fold change
MAF	Minor allele frequency
mPAP	Mean pulmonary artery pressure

LIST OF ABBREVIATIONS (continued)

mRNA	Messenger RNA
PAWP	Pulmonary artery wedge pressure
PASP	Pulmonary artery systolic pressure
PAH	Pulmonary arterial hypertension
PBMCs	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PDEI-5s	Phosphodiesterase-5 inhibitors
PH	Pulmonary hypertension
PVR	Pulmonary vascular resistance
RNA-seq	RNA sequencing
rRNA	Ribosomal RNA
SNP	Single nucleotide polymorphism
TGF-β	Transforming growth factor β
TPG	Transpulmonary gradient
TPM	Transcripts per million
TRV	Tricuspid regurgitation velocity

SUMMARY

The aim of this study was to use RNA-sequencing (RNA-seq) data to identify transcriptomic predictors of isolated post-capillary pulmonary hypertension (IpcPH) and combined pre- and post-capillary pulmonary hypertension (CpcPH) pathogenesis in heart failure patients with preserved ejection fraction (HFpEF).

Whole blood samples were selected from two different biobanks, from which peripheral blood mononuclear cells (PBMCs) were extracted. RNA was isolated from PBMCs, converted into complementary DNA libraries, and then sequenced to generate paired-end reads of about 75 bases. Next, these reads were aligned to the human reference genome (hg38). After alignment, the number of reads per gene were quantified on an individual patient basis. Following bias correction and normalization, adjusted pairwise gene expression values were compared among three groups (n=10/group): HFpEF without pulmonary hypertension (HFpEF without PH); IpcPH; and CpcPH. Gene set enrichment analysis was performed to identify groups of differentially expressed genes that were significantly over-represented in certain molecular pathways in relation to all interrogated genes. Network analysis was performed to observe co-expression and genetic interactions among a subset of differentially expressed genes. Finally, expression quantitative trait loci (eQTLs) were analyzed *in silico*. Using the candidate eQTLs from this analysis, a genetic association analysis was performed to determine if the single nucleotide polymorphisms (SNPs) were associated with pulmonary hypertension in a larger dataset of heart failure patients with and without pulmonary hypertension.

After controlling for age, sex, race and smoking status, 152 genes/transcripts were significantly downregulated, and 194 genes/transcripts were significantly upregulated in CpcPH patients compared with HFpEF without PH patients. No significant differentially expressed genes were found in other pairwise comparisons. Based on the enrichment analysis of all 22,418 genes/transcripts, 11 pathways were significantly upregulated in CpcPH compared to

HFpEF without PH, one of which, the cell-cycle, maintained significance after correction for multiple comparisons. Network analysis revealed a high level of co-expression among the candidate genes (false discovery rate < 0.05 and log_2-fold change > 1.5 or < -1.5) and 6 genes had eQTLs in one of the tissues of interest. Eighty six unique candidate eQTLs were identified from the 6 genes; of these, genotype array data was only available for 15 SNPs, none of which showed significance with PH in the larger dataset.

This study confirms that there is a complex network of molecular pathways that modulate the intricate CpcPH phenotype involving inflammation, cell proliferation and mitochondrial dysfunction. In addition, the results suggest some overlap between CpcPH phenotype and cancer etiology, as well as between CpcPH and other pulmonary hypertension subtypes. The candidate genes associated with CpcPH that were identified in this study may provide a new level of molecular targeting that has not been previously published, but which is strongly advocated for by experts in the field. Future work should focus on the translational confirmation of these molecular signatures.

I. INTRODUCTION

A. Heart Failure and Pulmonary Hypertension

By 2030, it is estimated that over 8 million Americans will have heart failure (HF).[1] Heart failure is a complex syndrome characterized by a decreased ability of the heart to fill and/or eject blood proportionate to the body's metabolic needs, which results in a host of signs and symptoms of systemic and pulmonary venous congestion.[2,3] Of HF patients, roughly half have HF with preserved ejection fraction (HFpEF), which compared to HF with reduced ejection fraction (HFrEF), has limited clinical evidence for specific pharmacotherapies.[4,5] Also problematic for these HFpEF patients is that up to 80% of them can develop pulmonary hypertension (PH), which confers additional morbidity and mortality risk.[6-12] The five-year mortality rate for HFpEF-PH patients is alarming – about 50%.[13] The World Health Organization categorizes this condition as Group 2 PH, with an etiology of left ventricular dysfunction and/or valvular heart disease.[14] Other classifications of PH by the World Health Organization are shown in Table I.[15] In order to definitively diagnose Group 2 PH, a right heart catheterization is required, which must reveal the following hemodynamic measurements: mean pulmonary artery pressure (mPAP) ≥ 25 mmHg and pulmonary artery wedge pressure (PAWP) > 15 mmHg.[16]

Within Group 2 PH, the predominant pathophysiology is thought to originate from sustained left atrial pressure, which leads to passive pulmonary venous congestion with high pulmonary pressures.[17] In many HFpEF-PH patients, this most likely accurately reflects the underlying mechanism. These patients are classified as having isolated post-capillary PH (IpcPH), displaying a PAWP > 15 mmHg and diastolic pulmonary gradient (DPG, = diastolic PAP – PAWP) < 7 mmHg or transpulmonary gradient (TPG, = mPAP – PAWP) ≤ 12 mmHg and/or pulmonary vascular resistance (PVR) ≤ 3 wood units.[14] However, up to 20% of HFpEF-PH patients develop an intrinsic pulmonary vascular disease and a disproportionately elevated mPAP that cannot be solely explained by passive venous congestion.[18] This severe progression

TABLE I

WORLD HEALTH ORGANIZATION CLASSIFICATION OF PH

I. Pulmonary arterial hypertension	
1.1	Idiopathic
1.2	Heritable
1.2.1	BMPR2 mutation
1.2.2	Other mutations
1.3	Drugs and toxins induced
1.4	Associated with:
1.4.1	Connective tissue disease
1.4.2	Human immunodeficiency virus infection
1.4.3	Portal hypertension
1.4.4	Congenital heart disease
1.4.5	Schistosomiasis
1'. Pulmonary veno-occlusive disease and/or pulmonary capillary hemangiomatosis	
1". Persistent pulmonary hypertension of the newborn	
2. Pulmonary hypertension due to left heart disease	
2.1	Left ventricular systolic dysfunction
2.2	Left ventricular diastolic dysfunction
2.3	Valvular disease
2.4	Congenital/acquired left heart inflow/outflow tract obstruction and congenital cardiomyopathies
2.5	Congenital /acquired pulmonary veins stenosis
3. Pulmonary hypertension due to lung diseases and/or hypoxia	
3.1	Chronic obstructive pulmonary disease
3.2	Interstitial lung disease
3.3	Other pulmonary diseases with mixed restrictive and obstructive pattern
3.4	Sleep-disordered breathing
3.5	Alveolar hypoventilation disorders
3.6	Chronic exposure to high altitude
3.7	Developmental lung diseases
4. Chronic thromboembolic pulmonary hypertension and other pulmonary artery obstructions	
4.1	Chronic thromboembolic pulmonary hypertension
4.2	Other pulmonary artery obstructions
4.2.1	Angiosarcoma
4.2.2	Other intravascular tumors
4.2.3	Arteritis
4.2.4	Congenital pulmonary arteries stenoses
4.2.5	Parasites (hydatidosis)
5. Pulmonary hypertension with unclear and/or multifactorial mechanisms	
5.1	Hematological disorders: chronic hemolytic anemia, myeloproliferative disorders, splenectomy
5.2	Systemic disorders: sarcoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, neurofibromatosis
5.3	Metabolic disorders: glycogen storage disease, Gaucher disease, thyroid disorders
5.4	5.4 Others: pulmonary tumoral thrombotic microangiopathy, fibrosing mediastinitis, chronic renal failure (with/without dialysis), segmental pulmonary hypertension

is called combined pre- and post-capillary PH (CpcPH, PAWP > 15 mmHg and DPG ≥ 7 mmHg or TPG > 12 mmHg and/or PVR > 3 wood units), and carries a more than two-fold mortality rate compared to IpcPH.[14,19,20]

B. Significance

Due to the fact that over half of HFpEF patients develop PH, which carries an astounding mortality rate, evidence-based treatment for HFpEF-PH is needed, but no FDA-approved treatments currently exist.[16] Consequently, clinicians are limited to treating comorbidities like hypertension and focusing on optimization of volume status and relaxation of the left ventricle.[21] Chronic pressure elevations within the pulmonary capillaries and subsequent remodeling suggest that a critical therapeutic target for HFpEF-PH may be the pulmonary vasculature.[22] Within this vasculature, there is an intricate interplay among a variety of cells (e.g. endothelial, progenitor, pericytes, smooth muscle, myofibroblasts, epithelial, macrophages, lymphocytes), and the dysregulation of this intercellular communication is the origin of PH.[23]

Thus far, most therapies tested in HF-PH patients have targeted endothelial control of vascular tone and permeability,[17] but have failed to show a consistent positive benefit. Treatments effective for pulmonary arterial hypertension (PAH), such as prostacyclins and endothelin receptor antagonists, have shown neutral results and even increased mortality in HF patients.[24,25] More recently, trials have assessed phosphodiesterase-5 inhibitors (PDE5-Is) as a potential therapy in HFpEF-PH patients due to the role of nitric oxide in maintaining endothelial cell quiescence and preventing dysfunction of endothelial cells.[26] In a trial by Guazzi *et. al.* (n=44),[27] patients were randomized to placebo or sildenafil 50mg three times daily for one year. At both six months and one year, sildenafil was shown to significantly improve mPAP and function of the right ventricle. While these results are extremely promising, a trial by Redfield *et. al.* that enrolled a larger number of HFpEF patients (n=216) failed to show any benefit at 24 weeks with use of sildenafil at 60mg three times daily.[28] The lack of strict, invasive hemodynamic characteristics as inclusion criteria for PH patients to enroll into the Redfield *et.*

al.[28] trial may have been the reason for the lack of significant findings despite almost five times the number of patients compared to the positive Guazzi *et. al.* trial.[27] However, the Guazzi *et. al.* study suggests that PDE5-Is may be a treatment option for CpcPH patients. These findings highlight the lack of effective treatments for Group 2 PH patients.

Histological investigation into the impact of chronic HF on the pulmonary vasculature has revealed that persistent elevations in left-sided filling pressures cause pathological changes of the pulmonary arteries, illustrated by intimal fibrosis, medial hypertrophy, and elastic fiber derangements.[29,30] Some postulate that IpcPH is not derived from abnormalities inherent to the pulmonary vascular system, but rather stems from increased left ventricular and atrial pressures, which is why IpcPH is usually considered to be reversible.[21] This paradigm is challenged by findings that some degree of pulmonary capillary remodeling may happen even in IpcPH patients.[18] Combined pre- and post-capillary PH is thought to be the result of sustained elevated pressures in the left atria, which leads to pathological changes in the distal pulmonary arteries and arterioles and causes an unusual increase in TPG.[29] The mPAP is higher than anticipated for the increase in PAWP. At this phase, the pulmonary vascular injury may be permanent. It seems likely that there is interindividual variability in the time course and degree of the development and regression of the disruptive changes seen in HFpEF-PH, which may be linked to constitutional elements apart from genetic predisposition.[21] Despite these known remodeling effects on the pulmonary vasculature, the pathogenesis of IpcPH and CpcPH in HFpEF patients is poorly understood, thus more research is needed to uncover the driving pathophysiological mechanisms.

It is essential to understand these mechanisms behind PH development in HFpEF in order to personalize patients' pharmacotherapy and discover innovative treatment strategies. One way to understand these pathophysiological mechanisms is to identify genes or gene pathways that promote HFpEF-PH development. Studies that have identified a few potential candidate genes for HF-PH development have been limited in scope and have suffered from a

lack of robust genome-wide expression analyses.[31-33] Thus, the objective of this study was to use RNA- sequencing (RNA-seq) data to identify transcriptomic predictors of IpcPH and CpcPH development in HFpEF patients.

C. RNA-sequencing Overview

RNA-seq is a revolutionary tool that enables investigators to study gene expression and identify candidate genes associated with disease or drug response via deep-sequencing technologies and bioinformatics analyses.[34] This approach is attractive as it allows for the combination of gene discovery and quantification in one high-throughput sequencing assay,[35] which is one way that transcriptomics has an advantage over methodologies in genomics that are limited to identifying genetic variants associated with drug response or disease. While the importance of genomics should not be disregarded, solely utilizing one-dimensional genomic information restricts our ability to understand the multidimensional relationship among genes/transcripts, levels of gene/transcript expression, and disease, under different conditions or within specific tissues. RNA-seq enables researchers to catalogue all transcript species, including messenger RNA (mRNA), non-coding RNA, and small RNAs, to ascertain the transcriptional structure of genes (with respect to start sites, 5' and 3' ends, splicing patterns, and other post-transcriptional modifications), and to quantify changing expression levels of all transcripts.[34]

The first RNA-seq study was reported ten years ago,[36] and until then, complementary DNA (cDNA) microarrays were a popular technology for high throughput gene expression profiling. However, microarrays have a number of limitations compared to RNA-seq due to their hybridization-based approach. One of which is their need for prior knowledge about the existing sequence, which restricts the identification of novel genes or transcripts or alternative splicing.[34,37] Furthermore, microarrays have a much smaller dynamic range of expression levels than RNA-seq, which means they lack sensitivity for genes that are expressed at low or high levels.[34] Table II describes these and other key differences between RNA-seq and microarray.

TABLE II

COMPARISON OF RNA-SEQ WITH MICROARRAY

	RNA-seq	Microarray
Methodology	High-throughput sequencing	Hybridization
Need for prior sequence knowledge	No	Yes
Resolution	Single base	>100 base pair
Dynamic range to quantify gene expression	$>10^5$	10^2-10^3
Background noise	Low	High
Ability to detect structural variations or isoforms	Yes	Limited
Data analysis	Complex; no standard pipeline	Straightforward

While either can be used to study the association of genes and complex diseases, RNA-seq provides a far more comprehensive view of the transcriptome, while having the ability to uncover the exact location of transcription boundaries to a single base resolution.[34]

In this past decade, RNA-seq studies have revealed thousands of novel isoforms and given additional insight into the complexity of the protein-coding transcriptome.[38] Further, RNA-seq studies have illustrated the importance of non-coding RNAs, particularly long non-coding RNAs, in physiological processes (e.g. neuron development, endocrine regulation, brown adipose differentiation)[39-41] and dysregulation of these having implications in disease (e.g. cancer, diabetes, and myocardial infarction).[42-45] Mounting evidence suggests that dysregulation of circular and short RNAs also play a role in disease.[46-48] These studies elucidate the potential role of RNA-seq to discover biomarkers that drive disease progression or that can act as therapeutic targets.[49]

II. METHODS

A. Study Population, Data Collection, and Design

Heart failure patients with and without PH were included in this study, and were recruited from the University of Illinois Hospital and Health Science System (UI Health) pulmonary hypertension and heart failure clinics. This study selected patient data and samples from two biobanks, Genomics and Biomarkers of Pulmonary Hypertension (GBPH) and the Heart Failure Genomics Database (HFGDB). Recruitment protocols for these biobanks were approved by the institutional review board at the University of Illinois at Chicago and all enrolled patients provided written, informed consent.

Inclusion and exclusion criteria for HFGDB and GBPH are summarized in Table III. For this study, the following data were available: DNA, peripheral blood mononuclear cells

TABLE III

INCLUSION AND EXCLUSION CRITERIA FOR THE HF AND PH BIOBANKS

	Heart Failure Genomics Database	Genomics And Biomarkers Of Pulmonary Hypertension
Inclusion	Men or women of any race who are 18 years of age or older	Men or women of any race who are 18 years of age or older
	Diagnosis of HF	Suspected or clinically diagnosed with PH
	Echocardiogram within the past two years	
	Ejection fraction ≤ 45% or diastolic dysfunction on echocardiogram	
Exclusion	Documented non-adherence to HF medications	
	Usage of illicit drugs	

(PBMCs), patient characteristics, past medical history, medication regimen, and echocardiogram and right heart catheterization measurements. Patient characteristics included age, sex, race, and ethnicity. Past medical history included hypertension, obstructive sleep apnea, lung disease, asthma, chronic obstructive pulmonary disorder, chronic renal insufficiency, hemodialysis, and smoking status. Echocardiogram measurements included ejection fraction, tricuspid regurgitation velocity (TRV) and pulmonary artery systolic pressure (PASP). Lastly, right heart catheterization measurements included systolic and diastolic PAP, PAWP, and cardiac output (CO). The following hemodynamic measurements were calculated: mPAP (=2/3 * diastolic PAP + 1/3 * systolic PAP), DPG (= diastolic PAP – PAWP), and TPG (=mPAP – PAWP), and PVR (=TPG/CO, in wood units).

In order to search for transcriptomic predictors of IpcPH and CpcPH development in HFpEF patients, our study used RNA-seq analysis to compare pairwise gene expression values among three groups: HFpEF without PH (n=10), IpcPH (n=10), and CpcPH (n=10). Table IV displays the echocardiogram and right heart catheterization measurements used to separate patients into one of these groups from GBPH (n=125) and HFGDB (n=353). Additionally, HFpEF patients had a history of diastolic dysfunction on echocardiogram and an ejection fraction >45%.

TABLE IV

ECHOCARDIOGRAM AND RIGHT HEART CATHETERIZATION MEASUREMENTS FOR COMPARATOR GROUPS

	mPAP (mmHg)		TRV (m/s)		PASP (mmHg)		PAWP (mmHg)		DPG (mmHg)		TPG (mmHg)		PVR (units)
HFpEF (no PH)	≤ 20	or	≤ 3	or	≤ 35								
IpcPH	≥ 25					and	> 15	and	< 7	or	≤ 12	or	≤ 3
CpcPH	≥ 25					and	> 15	and	≥ 7	or	> 12	or	> 3

Baseline patient characteristics were compared among the three experimental groups. The Wilcoxon Mann-Whiney U-test or Kruskal-Wallis test were used to compare continuous data, whereas Fisher's exact test was used to compare categorical data.

B. RNA-sequencing and Analysis

1. Sample Preparation and Sequencing

Peripheral blood mononuclear cells, which were isolated from whole blood according to the manufacturer's protocol,[50] were used as the RNA source from the 30 patients included in this study. RNA was extracted and its quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany). To create a sequencing library, the TruSeq Stranded Total RNA Sample Preparation kit (Illumina, San Diego, CA) was used, in which the first step was bead-based ribosomal RNA (rRNA) depletion. After purification, the RNA was fragmented into small pieces, which were synthesized into double stranded complementary DNA (cDNA). Indexing adapters were ligated to the ends of the cDNA, and these cDNA libraries were polymerase chain reaction (PCR) amplified and purified. In order to reduce batch and lane effects, the indexed libraries prepared from HFpEF without PH, IpcPH, and CpcPH samples were randomized by experimental group and RNA extraction batch and combined into four different pools for sequencing. Total RNA-seq was performed on the Illumina NextSeq 500 platform at the University of Illinois Core Genomics Facility, which generated paired-end reads of about 75 bases.

2. Data Preparation and Alignment

The raw sequencing data in FASTQ file format were uploaded to BaseSpace (Illumina, San Diego, CA), a cloud computing environment, from which it was downloaded and stored on an external hard drive. All files were then transferred to a directory on HiPerGator, the supercomputing cluster at the University of Florida, as cluster computing was used to run programs for some key analysis steps (e.g. quality control, read alignment to the human reference genome, and quantification). Figure 1 displays the key steps in RNA-seq data

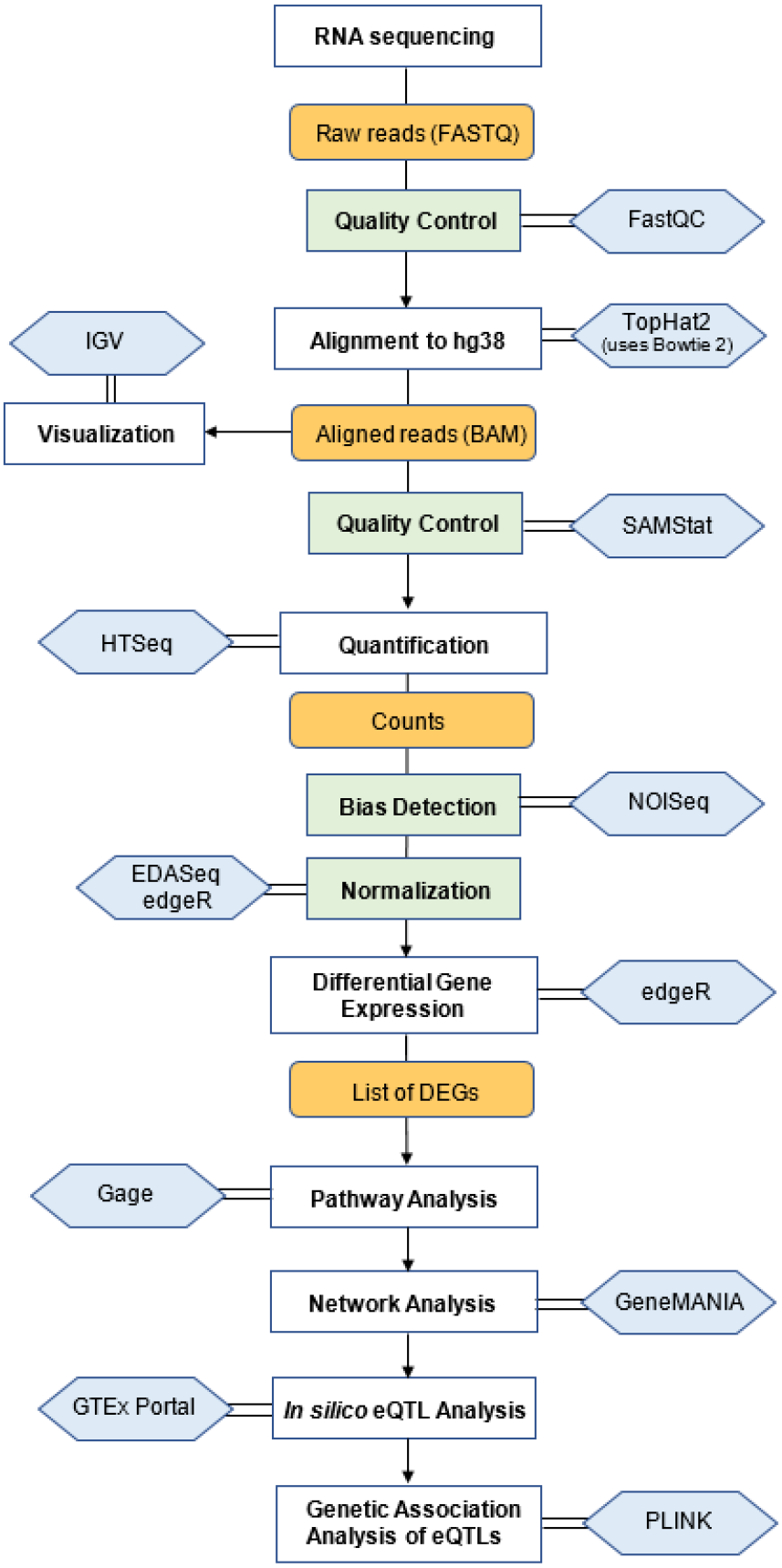

Figure 1. Steps in our RNA-seq data analysis. DEGs- differentially expressed genes; eQTLs- expression quantitative trait loci.

analysis, including file formats and programs used to execute each step. White and green boxes represent steps, orange boxes represent file formats, and blue hexagons represent programs. FastQC[51] was used to assess the quality of the raw sequenced reads (Table V) before proceeding to alignment.

Prior to alignment, Bowtie 2[52] was used to build a reference index (hg38) to be used in the downstream analysis with TopHat. The paired-end reads produced from sequencing were aligned to the human reference genome (hg38) using the splice-aware short read aligner TopHat v2.1.1,[53] with an expected mean inner distance between mate pairs of 175 bases (calculated as mean average library fragment [325 bases] minus 75 bases on each end). We also used a parameter that indicated that the right-most end of fragments (in transcript coordinates) were sequenced first. Alignment data were generated in BAM file format, which were assessed for quality using SAMStat.[54] As an additional quality check, SAMtools[55] was used to create an index of the BAM files, in order to use the Integrative Genomics Viewer[56] to visualize a handful of random genes and the reads mapped to them.

TABLE V

FASTQC: SUMMARY OF EVALUATED ITEMS

Modules
1. Basic Statistics
2. Per base sequence quality
3. Per tile sequence quality
4. Per sequence quality scores
5. Per base sequence content
6. Per sequence GC content
7. Per base N content
8. Sequence Length Distribution
9. Sequence Duplication Levels
10. Overrepresented sequences
11. Adapter Content
12. Kmer Content

3. Quantification and Bias Detection

To prepare the paired-end alignment data for quantification, SAMtools[55] was used to sort the BAM files by read name. In order to estimate expression, HTSeq[57] was used to count the number of reads per gene using the sorted BAM files. In order to run the *htseq-count* function, we set parameters that indicated our input files were BAM, that the right-most end of fragments were sequenced first, that gene ID was the aggregate level at which to return results, that paired end reads were sorted by name, and that mode was set as the intersection of all non-empty sets. The read count files for each patient sample were merged together into one file using R v3.5.1,[58] with gene names as row headers, sample IDs as column headers, and discrete counts by gene for each sample.

In order to detect biases in the gene expression estimates, exploratory analyses were performed on raw count data using NOISeq[59] to assess for: gene length bias, GC content bias, and count distribution bias. Plots from these analyses were generated and used to inform downstream normalization methods. Gene length bias describes the fact that longer genes are more likely to be sequenced and appear to have higher expression, whereas GC content bias describes how genes with low/high GC content (percentage of guanosine or cytosine nucleotides of a given gene) typically have lower expression. Count distribution bias describes a situation where two samples may have the same sequencing depth and expression of gene B, but gene A is highly expressed in sample 1 and not in sample 2; as gene A accumulates a lot of the reads in sample 1, there are less reads remaining for the rest of the genes. This means that even though gene B should have the same number of reads mapped to it in both samples, it will probably have less counts in sample 1.

4. Normalization and Differential Gene Expression Analyses

Prior to differential gene expression analysis, a principle component analysis plot was generated in R to look for any relationships among samples. Next, GC content bias was corrected using loess robust local regression within the EDASeq package[60,61] in R. Then, genes

with low expression were filtered out of the count data in R, retaining genes with a count per million (CPM) value greater than 0.24 in at least 5 samples. In other words, taking into account the sequence depth (minimum library size) of 20,690,000 reads, genes were retained if they had five fragments in at least 5 samples ($CPM = \frac{5}{Min\ (Library\ size)} \times 1{,}000{,}000$). Furthermore, the trimmed mean of M-values (TMM) method[62] was used to normalize the count data for count distribution bias, which was performed in edgeR in R with the function *calcNormFactors*. Design and contrast matrices, which are R objects that describe the experimental design and which comparisons are to be made, respectively, were created in R to facilitate differential gene expression analyses.

Expression levels were compared in pairwise analyses among the three groups of biological replicates: 1) HFpEF without PH vs. IpcPH; 2) HFpEF without PH vs. CpcPH; 3) IpcPH vs. CpcPH. Differences in gene expression were compared using R's BioConductor package edgeR,[63] using a negative binomial, generalized linear model with a quasi-likelihood method and adjusting for age, sex, race (Caucasian, Hispanic/Latino, African American), and smoking status (never smoker, past smoker, current smoker). This method allowed us to account for the uncertainty of the variance and overdispersion in the data, for which the *estimateDisp* function was used. Using the Benjamini-Hochberg method to adjust for multiple comparisons, false discovery rates (FDR) were calculated. Genes were considered to be differentially expressed between two groups if the FDR adjusted p-value < 0.05.

Since the output file from the edgeR analysis only provided Ensembl gene IDs, the BioConductor package biomaRt[64] was used to link Ensembl gene IDs with Entrez gene IDs (needed for downstream pathway analysis) and HUGO Gene Nomenclature Committee gene names (helpful for biological interpretation). Volcano and smear plots were produced in R to visualize the differentially expressed genes, including whether the genes were up or downregulated.

C. Pathway Analysis

Gene set enrichment analysis was performed to identify groups of differentially expressed genes that were significantly over-represented in certain molecular pathways in relation to all interrogated genes. Specifically, a functional class scoring method called Generally Applicable Gene-set Enrichment (GAGE)[65] was used. Since the functional class scoring method does not need an arbitrary *a priori* significance level and uses all available molecular measurements for pathway analysis, this method was preferred over approaches utilizing over-representation analysis, like the commonly used database for annotation, visualization and integrated discovery (DAVID).[66] Using the GAGE package in R, pathway analysis was done by drawing upon the Kyoto Encyclopedia of Genes and Genomes (KEGG) database,[67] a widely used knowledge base that links genomic information with molecular pathways, illustrating higher order biological function.

D. Network Analysis

Next, network analysis was performed to observe co-expression and genetic interactions among a subset of differentially expressed genes using GeneMANIA.[68] Differentially expressed genes (FDR < 0.05) were utilized as input for this analysis if they had a LogFC > 1.5 or < -1.5. In order to infer functional relevance in PH, genes used in the network analysis and revealed from the pathway analysis (P < 0.05 for enriched pathway, FDR < 0.05 for gene) were explored in the literature for previous associations with PH, including the subtype of PH and studied cell type. Furthermore, the literature was searched to identify if these candidate genes had associations with pulmonary disease.

E. *In Silico* Expression Quantitative Trait Loci Analysis

In order to investigate whether the differentially expressed genes have any known expression quantitative trait loci (eQTL) potentially driving their expression, the Genotype-Tissue Expression (GTEx) portal (https://gtexportal.org/home/) was utilized to perform *in silico* eQTL analysis. Of the genes that were differentially expressed (i.e. FDR < 0.05), a subset of

genes were removed that did not have Entrez IDs, as to narrow this list down to genes with more detailed annotation. From here, only a subset of differentially expressed genes were selected if they had a $\log_2$-fold change (logFC) > 1.5 or < -1.5, since these genes are least likely to have significant gene expression values due to chance. The remaining genes were queried in the GTEx portal to determine if they were expressed in our tissues of interest (i.e. lung, atria, left ventricle, aorta, coronary artery). Plots were generated in the GTEx portal to view expression of each of these genes by tissue type, examining the relative expression within our tissues of interest compared to all other GTEx tissues. Genes where lung was ranked within the top 5 tissues with highest expression were noted.

Next, the GTEx portal was used to create a list of genes within our tissues of interest that had eQTLs associated with them. From this list of genes, we created a list of known eQTLs (noting rsID) using GTEx, removing any duplicates. Using SNPclip tool (https://ldlink.nci.nih.gov/?tab=snpclip), linkage disequilibrium (LD) pruning was performed to remove eQTLs with $R^2 > 0.8$ (a value indicating that the loci have a high likelihood of being inherited together) and minor allele frequency < 0.01, using data from the following ancestries: Americans of African ancestry in Southwestern US, Mexican ancestry from Los Angeles, and Utah residents from Northern and Western Europe. The pruned dataset of candidate eQTLs was set aside for further validation analysis within the larger dataset of patients in the HFGDB with available genotype array data (see section II, F). For a representative subgroup of eQTLs, boxplots were viewed in GTEx, which showed gene expression values (y-axis) by genotype group (x-axis).

F. Genetic Association Analysis of Expression Quantitative Trait Loci

Genotyping of genomic DNA from the HFGDB was completed using the Axiom® Genome-Wide Pan-African Array Set (Affymetrix, Inc, Santa Clara, CA), which has provided us genotype data on over 2 million single nucleotide polymorphisms (SNPs) per patient. Due to our unique study population consisting mainly of those of African descent, this platform was ideal. In

an effort to determine if the candidate SNPs from the *in silico* eQTL analysis are associated with PH, a genetic association analysis was attempted in a larger dataset of HF patients with and without PH. For this analysis, the aim was to determine if HF patients with these candidate SNPs were more likely to have PH, controlling for age, sex, race, and smoking status. From the HFGDB, 165 HF patients with genotype array data had hemodynamic or echocardiographic information available to categorize them into one of two groups: Heart failure without PH (mPAP ≤ 20 mmHg or TRV ≤ 3 m/s; n = 80) and HF-PH (mPAP ≥ 25 mmHg and PAWP > 15 mmHg; n = 85). Since right heart catheterization data was not available on many of these HF patients, we also had to use TRV as a marker for PH status. This TRV threshold was used because patients with a TRV ≤ 3 m/s are less likely to have PH than those with TRV > 3 m/s.

First, candidate eQTLs were extracted from the GWAS data for the 165 HF patients using PLINK v1.9.[69] PLINK was used for all quality control procedures and SNP analysis. Remaining candidate SNPs were tested for Hardy-Weinberg equilibrium (HWE). Quality control procedures included removing low quality SNPs with more than 10% of samples missing genotype data, excluding low quality samples with more than 10% missing genotype data, and removing monomorphic SNPs (with minor allele frequency, MAF, equal to 0%). Logistic regression was performed, comparing log odds of PH status to candidate SNP genotypes (e.g. CC, CT, TT coded as 0, 1, 2) assuming an additive inheritance model, adjusting for age, sex, race, and smoking status. SNPs were considered to be significantly associated with log odds of PH if they had a Bonferroni-corrected *P* value < 0.003.

III. RESULTS

A. Patient Characteristics

Table VI shows the demographic, hemodynamic, and clinical characteristics of the total 28 patients by PH group. Two samples collected from IpcPH patients failed quality control, hence they were excluded from all analyses. As expected based on study selection criteria and diagnostic differences between IpcPH and CpcPH, the following measurements were significantly different between the two PH phenotype groups: mPAP, PVR, TPG, and DPG. Likewise, there was no difference in PAWP between the PH patients. Demographics were similar among the three groups, as well as prevalence of hypertension, obstructive sleep apnea, lung disease, and hemodialysis use. Smoking status significantly differed among the three groups, as about a third of HFpEF patients were current smokers and the majority of PH patients were previous smokers. As anticipated, CpcPH patients were younger on average and had a higher prevalence of chronic lung disease compared to IpcPH patients.

B. Sample Preparation, Sequencing, Alignment, Quantification, Bias Detection, Normalization

Prior to sequencing, all isolated samples had RNA integrity numbers above 7, meaning the RNA quality was acceptable and confirming a low contamination with degradation products.[70,71] Most samples (except for three samples with values of 7.7, 7.9, and 7.9) had RNA integrity numbers above 8, indicating good quality.

Upon assessment of sequencing quality, the majority of the sequenced reads had an average quality score of 32-36 (Figure 2), meaning the probability of calling an incorrect base was 0.025-0.063%. Two IpcPH samples failed quality control, while the quality of the other raw data was deemed acceptable to move onto alignment.

At the alignment stage, an average of 54.5 million paired-end reads per sample were mapped to the human reference genome (hg38), of which approximately 95.9% were mapped

TABLE VI

PATIENT CHARACTERISTICS BY GROUP

Baseline Characteristics	HFpEF (no PH) (n=10)	IpcPH (n=8)	CpcPH (n=10)	P-value
Age (years), mean ± SD	58 ± 6	63 ± 12	55 ± 8	0.16
Female sex, %	80	62.5	50	0.38
Race/ethnicity, %				
African American	70	37.5	70	
Hispanic Caucasian	0	37.5	10	0.31
Non-Hispanic Caucasian	20	25	20	
Unknown	10	0	0	
Mean PAP (mmHg), mean ± SD		33.6 ± 6.4	49.7 ± 7.5	0.001
PAWP (mmHg), mean ± SD		25.0 ± 8.5	22.5 ± 3.3	0.89
PVR (wood units), mean ± SD	N/A	1.2 ± 0.7[a]	4.1 ± 1.4	0.002
TPG (mmHg), mean ± SD		8.8 ± 4.1	27.2 ± 5.7	0.0004
DPG (mmHg), mean ± SD		0.5 ± 1.4	11.5 ± 3.3	0.0003
Hypertension, %	90	100	100	1.00
OSA, %	40	37.5	50	0.90
History of lung disease,[b] %	60	12.5	60	0.08
Hemodialysis, %	10	12.5	20	1.00
Smoker, %				
Never	50	37.5	10	
Previous	20	62.5	90	0.01
Current	30	0	0	

OSA- obstructive sleep apnea; SD- standard deviation.

[a] n=6.

[b] Includes asthma, chronic obstructive pulmonary disease, restrictive lung disease, latent tuberculosis, parenchymal lung disease, history of pulmonary embolism (in addition to PH for IpcPH and CpcPH patients.

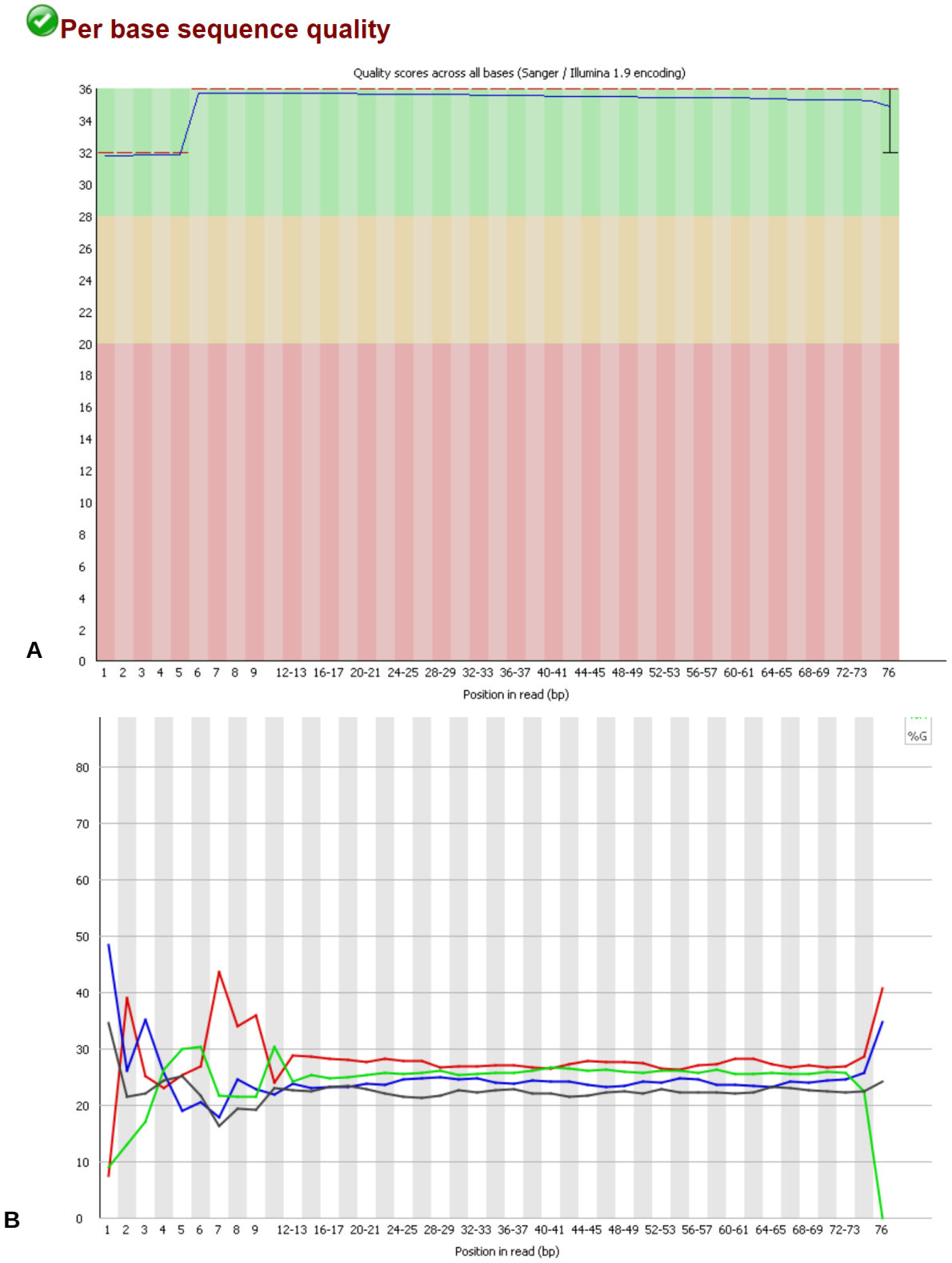

Figure 2. FastQC modules for a representative sample, showing average quality scores per base (A) and sequence content per base (B).

and 85.2% were uniquely mapped (Table VII). The ratio of uniquely mapped reads to multiple mapped reads was 5.8, which is within the acceptable threshold for human data (i.e., < 20). Using HTSeq for quantification and R to merge read count files for each sample, read counts per gene/transcript per sample were generated for a total of 58,051 unique coding or non-coding transcripts. Based on assessment of the mapping statistics and the mean base quality via SAMStat (Figure 3) and visualization with Integrative Genomics Viewer (Figure 4), the data was acceptable for quantification.

Based on exploratory bias analyses in NOISeq, the raw count data showed significant biases in gene length (Figure 5), GC content (Figure 6), and count distribution (Figure 7). As such, GC content and count distribution bias were corrected using EDASeq and the TMM method in R. Sequencing depth bias was addressed using a CPM filter to remove genes with low expression. During normalization, the total gene/transcript count was reduced from 58,051 to 57,785 after GC content bias correction, which was further reduced to 22,418 after using the CPM filter. Hence, GC-corrected and filtered count data underwent TMM normalization for 22,418 genes, which was used for differential gene expression analyses in edgeR. Lastly, a principle component analysis plot revealed lack of defined clustering for the three PH groups (Figure 8), suggesting heterogeneity within and between these groups. Black study IDs indicate HFpEF without PH samples, red study IDs indicate IpcPH samples, and green study IDs indicate CpcPH samples.

C. Differentially Expressed Genes and Associated Pathways

When controlling for age, sex, race, and smoking status, 152 genes/transcripts were significantly downregulated and 194 genes/transcripts were significantly upregulated (FDR < 0.05) in CpcPH patients compared to HFpEF patients without PH (Figure 9). Red and blue dots indicate differentially expressed genes/transcripts with FDR < 0.05, with red illustrating upregulated genes/transcripts and blue illustrating downregulated genes/transcripts. Blue lines are set at logFC values of 1.5 and -1.5. At this significance threshold, there were no differentially

TABLE VII

ALIGNMENT STATISTICS

Characteristics	HFpEF (no PH) (n=10)	IpcPH (n=8)	CpcPH (n=10)	Total (n=28)
Total number of unaligned paired-end reads	581,839,436	451,993,807	556,558,150	1,590,391,393
Number of unaligned reads per sample[a]	58,183,943 (49,950,758-66,813,976)	56,499,225 (49,696,045-65,949,768)	55,655,815 (50,680,843-68,024,333)	56,799,692.6 (49,696,045-68,024,333)
Total number of aligned paired-end reads	558,219,210	433,681,474	533,980,236	1,525,880,920
Number of aligned reads per sample[a]	55,821,921 (46,556,139-64,993,962)	54,210,184 (47,307,761-64,129,854)	53,398,023 (47,946,846-66,160,132)	54,495,747 (46,556,139-66,160,132)
Aligned reads[a] (%)	95.9 (95.0-96.4)	95.9 (95.6-96.4)	95.9 (95.0-96.5)	95.9 (95.0-96.5)
Uniquely mapped reads[a] (%)	85.1 (80.6-88.6)	86.1 (79.9-88.5)	84.5 (75.2-90.8)	85.2 (75.2-90.8)
Multiple mapped reads[a] (%)	14.9 (11.4-19.4)	13.9 (11.5-20.1)	15.5 (9.2-24.8)	14.8 (9.2-24.8)

[a] Denotes mean (range)

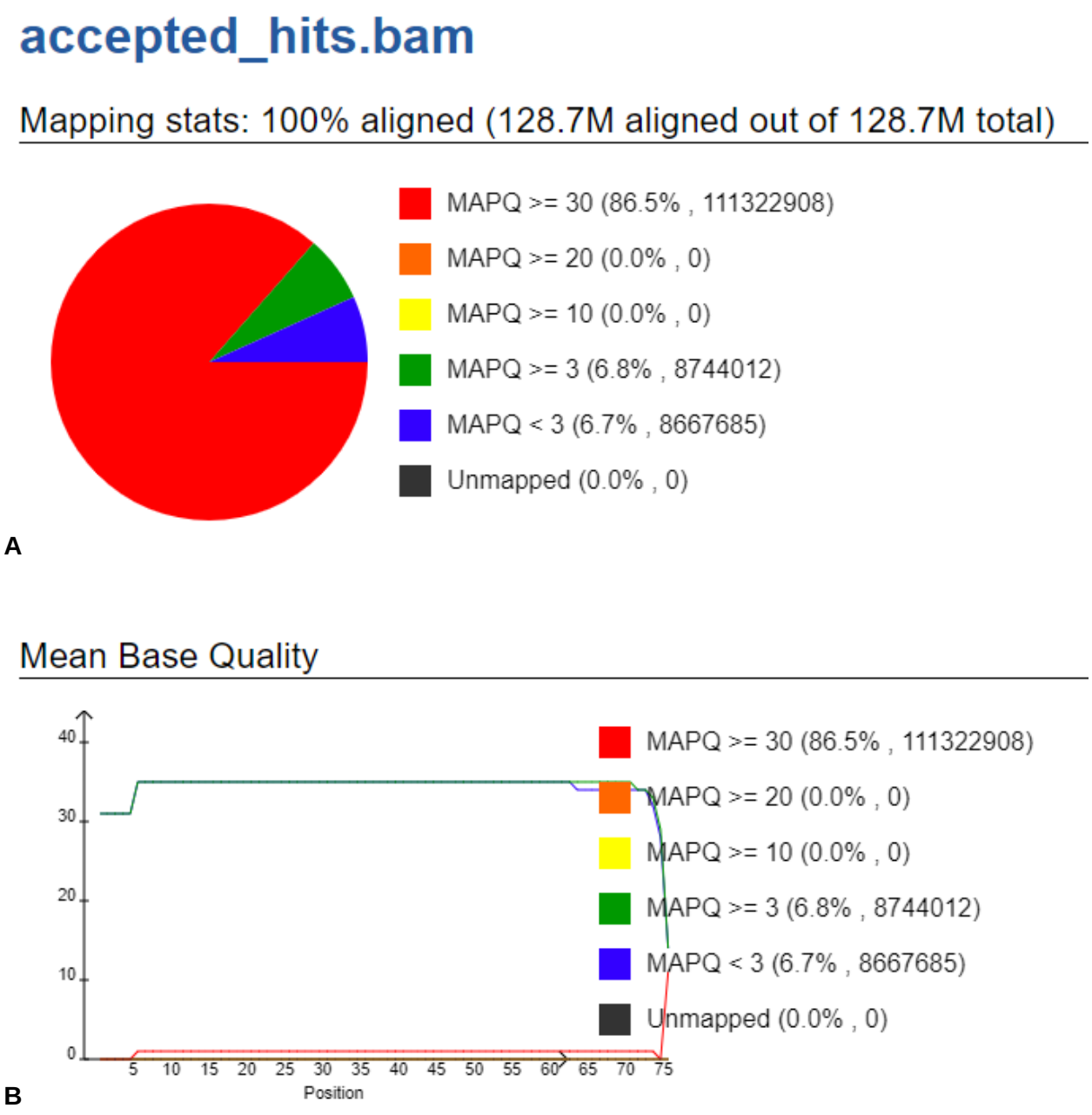

Figure 3. SAMStat modules for a representative sample, showing mapping statistics by quality score (A) and mean base quality (B).

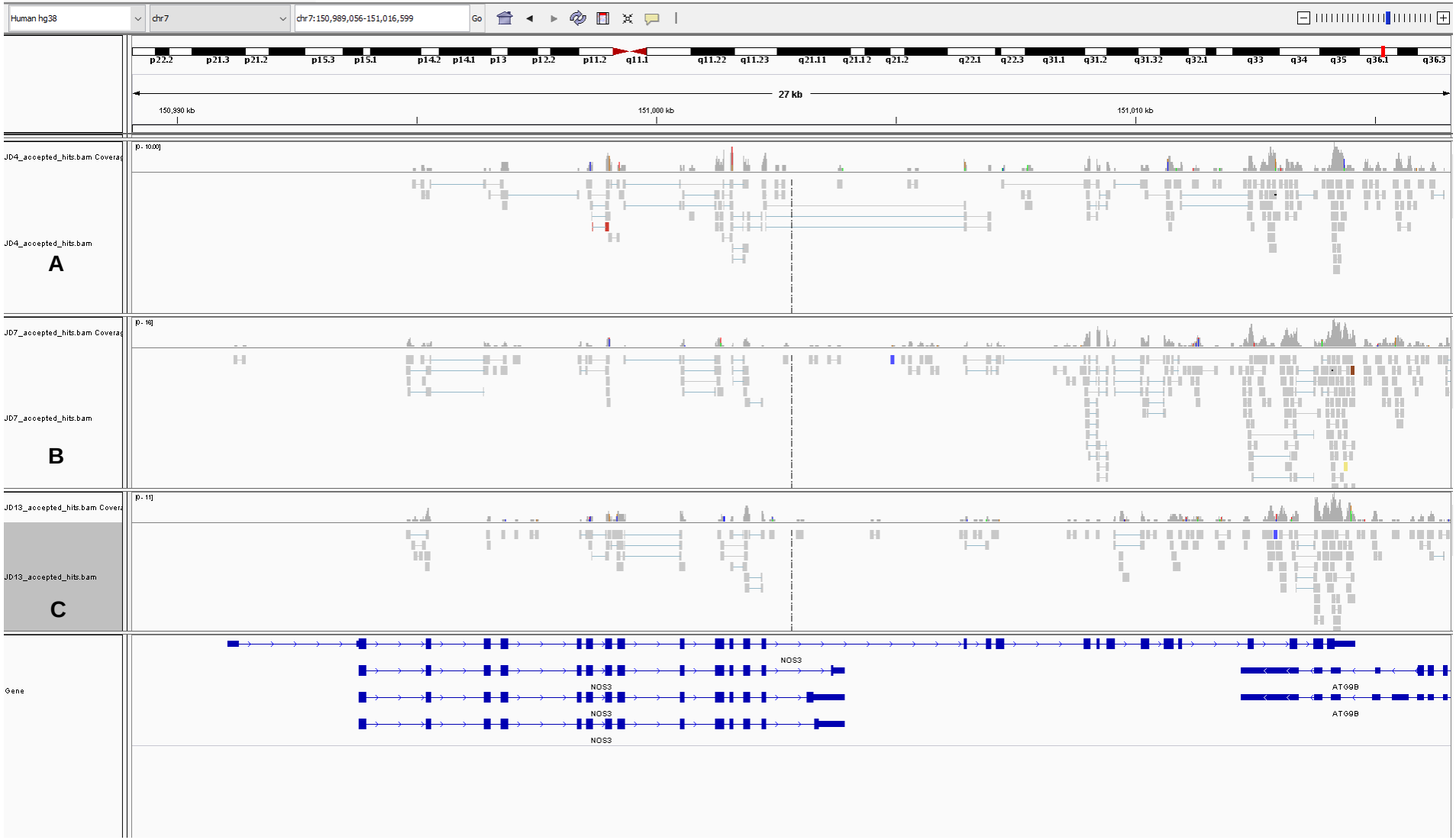

Figure 4. Visualization of mapped reads to *NOS3* within Integrative Genomics Viewer for a CpcPH (A), IpcPH (B), and HFpEF without PH (C) sample.

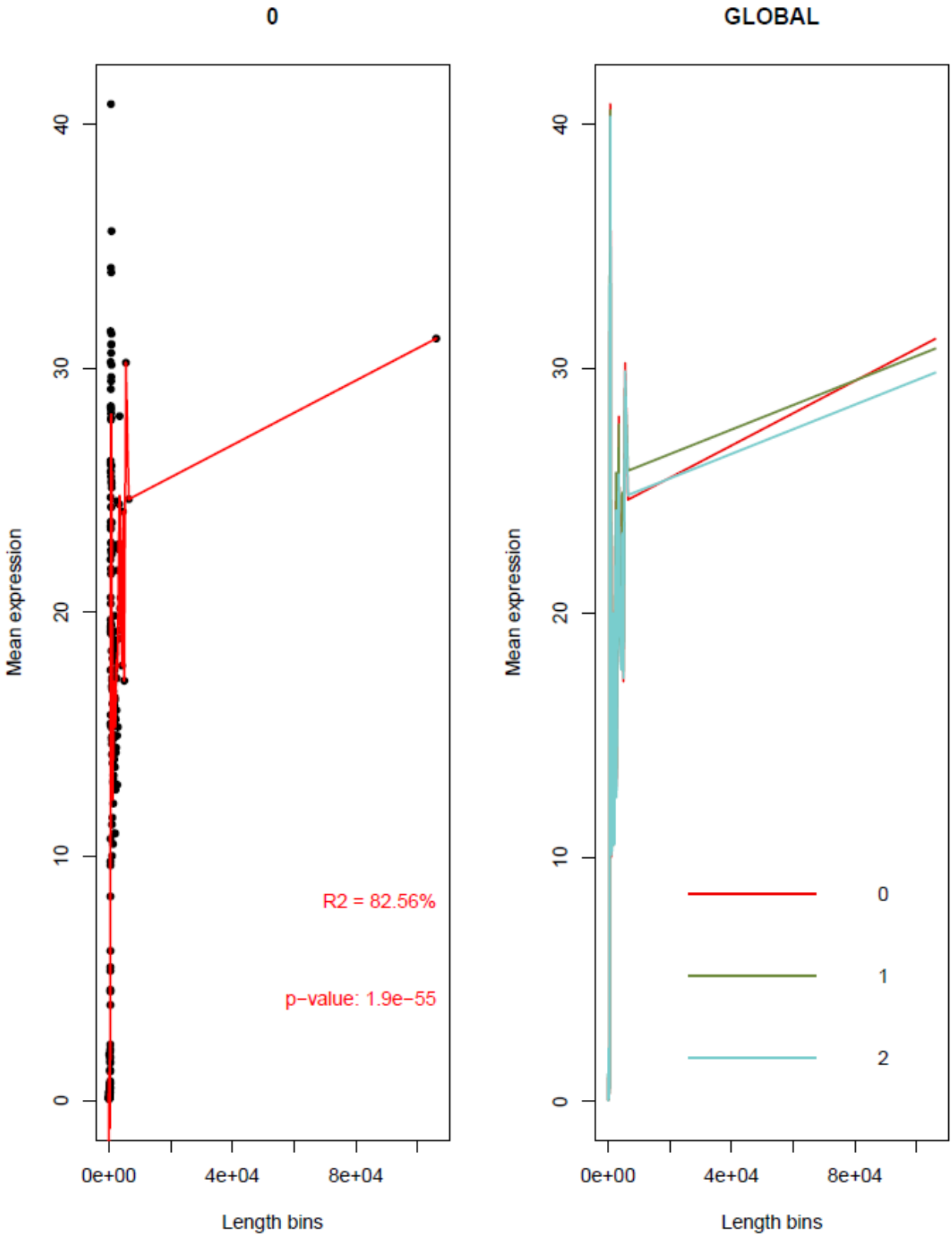

Figure 5. Gene length bias plots from NOISeq.

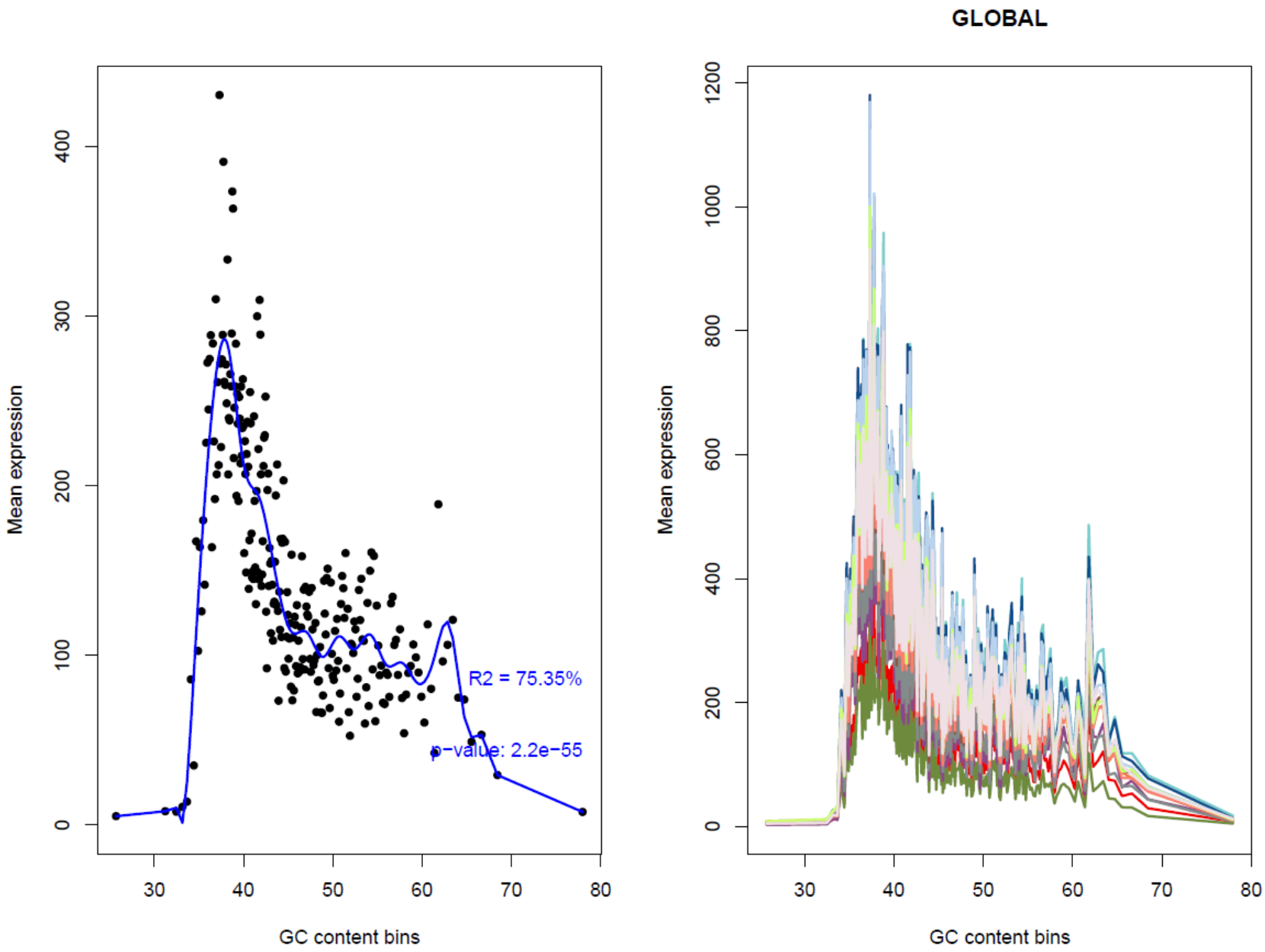

Figure 6. GC content bias plots from NOISeq.

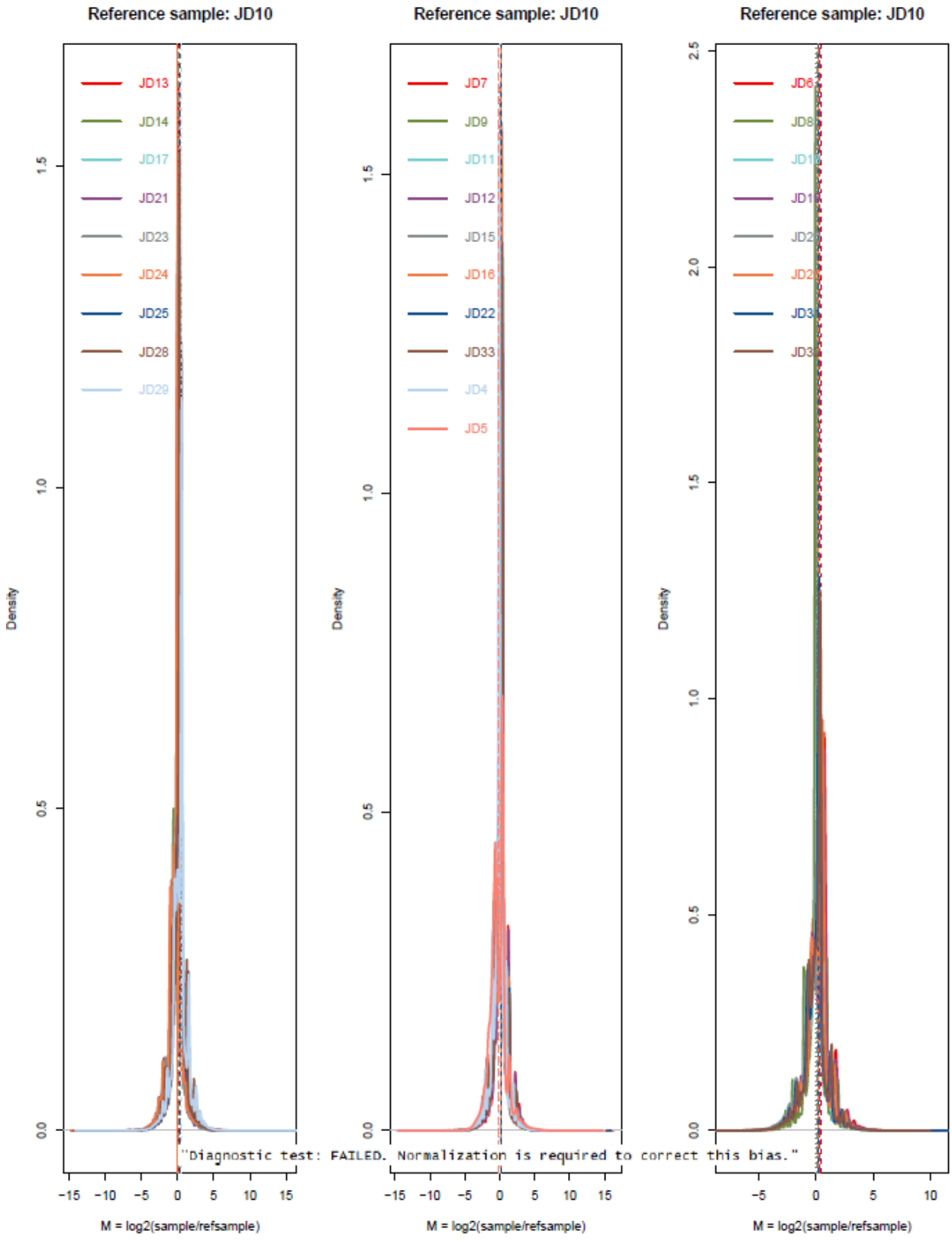

Figure 7. Count distribution bias plots from NOISeq.

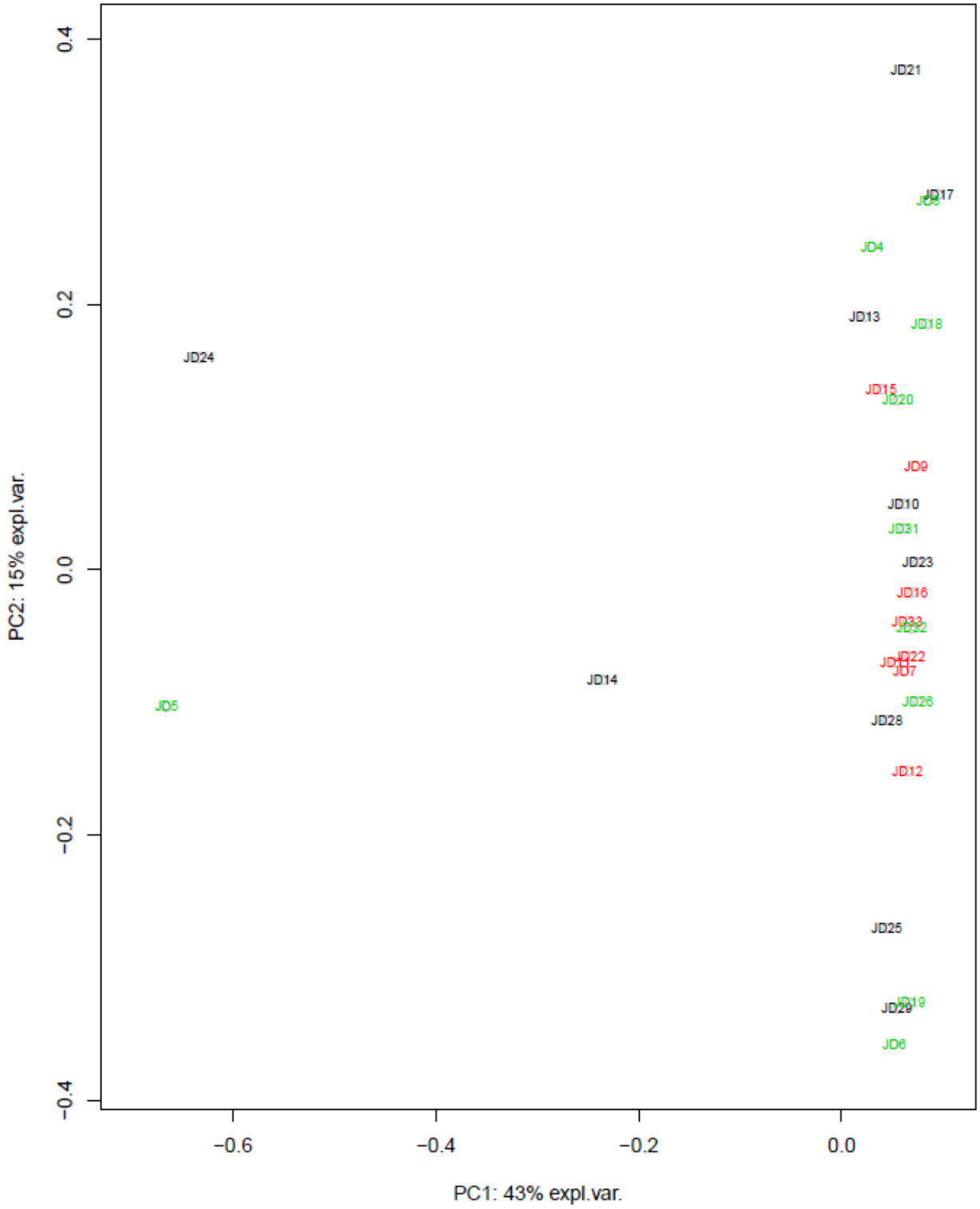

Figure 8. Principle component analysis plot of all samples, illustrating the first two principle components. PC- principle component.

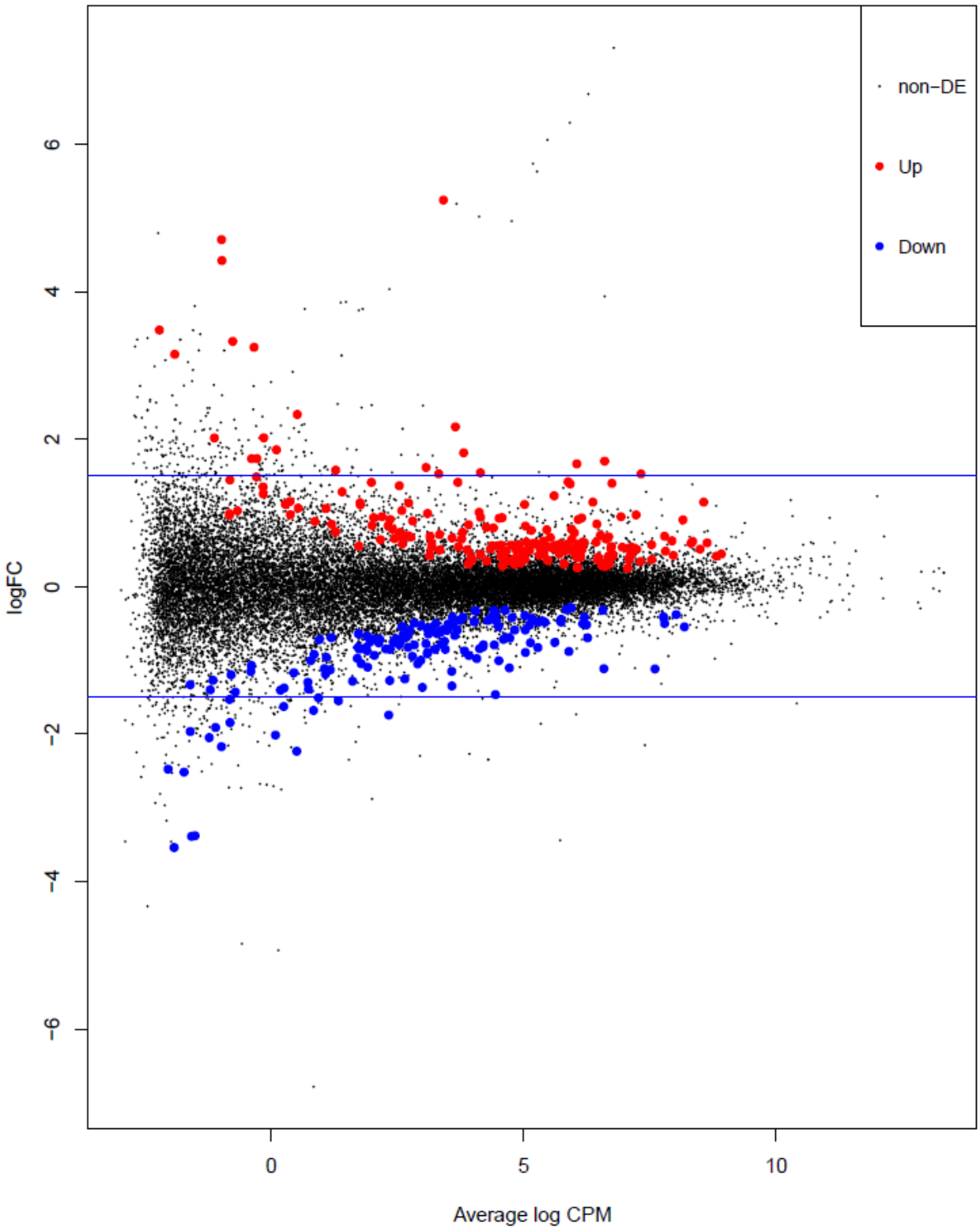

Figure 9. Smear plot comparing log fold change expression values between CpcPH and HFpEF without PH patients by average logCPM. DE- differentially expressed.

expressed genes/transcripts between the other pairwise comparisons (i.e. HFpEF without PH vs. IpcPH and IpcPH vs. CpcPH). Of the 346 differentially expressed genes/transcripts, 26 genes had a logFC > |1.5| (Figure 10 and Table VIII). In Figure 10, blue dots indicate differentially expressed genes/transcripts with FDR < 0.05 and logFC > |1.5|.

Based on the enrichment analysis of all 22,418 genes/transcripts, 11 pathways were significantly upregulated in CpcPH patients compared to HFpEF without PH patients ($P < 0.05$), while one retained significance after Benjamini-Hochberg correction for multiple testing (FDR < 0.05): cell-cycle (FDR = 0.04). Table IX displays these significant pathways, including differentially expressed genes/transcripts within these pathways (FDR < 0.05).

D. Network Analysis

Figure 11 shows a network of genes that may have biological relevance with respect to PH development in HFpEF patients. Of the 346 differentially expressed genes/transcripts between CpcPH and HFpEF without PH (FDR < 0.05), 26 genes/transcripts with LogFC > 1.5 or < -1.5 (Table VIII) were explored in this analysis: Nine genes/transcripts were not found in GeneMANIA (*BNIP3P5*, *RPL32P2*, *SMIM33*, *IGHV7-4-1*, *HSP90AA6P*, *RN7SL749P*, *LOC105371242*, *LOC100506571*, *LOC100422382*) due to lack of substantial information about them in the literature. *OLFM2* and the pseudogene *RIMBP3* did not have connections to other candidate genes, thus are not shown in Figure 11. Hence, 15 of our candidate genes were displayed in this network. The network analysis also revealed 20 additional related genes based on functional association data (Table X). Six of these genes were differentially expressed within our dataset at a FDR < 0.05 (*NFKBIA*, logFC = 1.15, FDR = 0.0010; *DUSP2*, logFC = 1.40, FDR = 0.0013; *HBEGF*, logFC = 1.42, FDR = 0.0015; *PLK3*, logFC = 0.95, FDR = 0.0062; *RGS1*, logFC = 1.22, FDR = 0.0100; *PPP1R15A*, logFC = 0.74, FDR = 0.0114; Table X).

Within the constructed composite network (Figure 11), this analysis uncovered 3 networks, each of which are a weighted sum of individual data sources and are denoted by different colored lines. Each line (link) is weighted by the particular data source. These 3

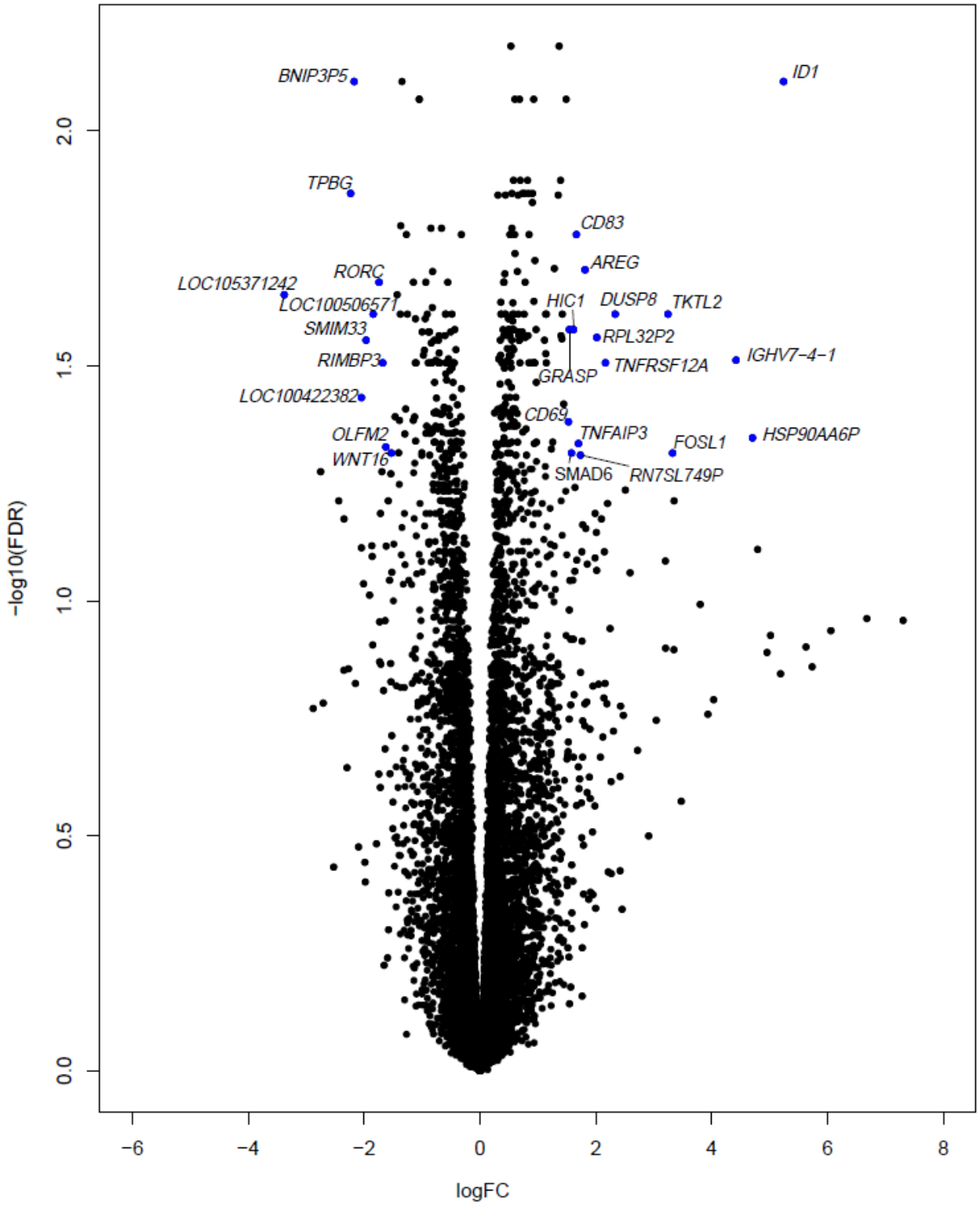

Figure 10. Volcano plot comparing log fold change expression values between CpcPH and HFpEF without PH patients by level of significance.

TABLE VIII

DIFFERENTIALLY EXPRESSED GENES/TRANSCRIPTS BETWEEN HFPEF WITHOUT PH AND CPCPH PATIENTS

Gene Symbol	Gene Name	Gene Family	Chr Location	Length (bases)	LogFC[b]	*P* value	FDR
ID1	inhibitor of DNA binding 1, HLH protein	Basic helix-loop-helix proteins	20q11.21	1,233	5.2	1.75E-06	0.0078
BNIP3P5[a]	BCL2 interacting protein 3 pseudogene 5	N/A – pseudogene	15q15.1	899	-2.2	1.48E-06	0.0078
TPBG	trophoblast glycoprotein	N/A	6q14.1	7,623	-2.2	1.39E-05	0.0136
CD83	CD83 molecule	CD molecules, V-set domain containing	6p23	19,663	1.7	2.92E-05	0.0166
AREG	amphiregulin	Endogenous ligands	4q13.3	9,912	1.8	4.05E-05	0.0197
RORC	RAR related orphan receptor C	Nuclear hormone receptors	1q21	26,168	-1.7	5.00E-05	0.0210
LOC105371242	peptidyl-prolyl isomerase A like 4C	N/A	1q21.1	1,752	-3.4	6.07E-05	0.0223
TKTL2	transketolase like 2	N/A	4q32.2	2,803	3.2	8.17E-05	0.0245
DUSP8	dual specificity phosphatase 8	MAP kinase phosphatases	11p15.5	18,788	2.3	8.60E-05	0.0245
LOC100506571	uncharacterized	N/A	1q42.13	4,098	-1.8	7.55E-05	0.0245
HIC1	HIC ZBTB transcriptional repressor 1	BTB domain containing, Zinc fingers C2H2-type	17p13.3	9,088	1.6	1.16E-04	0.0265
GRASP	general receptor for phosphoinositides 1 associated scaffold protein	PDZ domain containing	12q13.13	8,953	1.5	1.16E-04	0.0265
RPL32P2[a]	ribosomal protein L32 pseudogene 2	N/A – pseudogene	15q14	509	2.0	1.51E-04	0.0275
SMIM33	small integral membrane protein 33	N/A	5q31.2	3,995	-2.0	1.61E-04	0.0279

[a] Non-protein coding.

[b] Log fold change corresponds to gene/transcript expression values from CpcPH patients divided by values from HFpEF without PH patients. Positive values indicate upregulation in CpcPH vs. HFpEF without PH patients, whereas negative values indicate downregulation in CpcPH vs. HFpEF without PH patients (i.e., upregulation in HFpEF without PH patients vs. CpcPH patients).

TABLE VIII (continued)

DIFFERENTIALLY EXPRESSED GENES/TRANSCRIPTS BETWEEN HFPEF WITHOUT PH AND CPCPH PATIENTS

Gene Symbol	Gene Name	Gene Family	Chr Location	Length (bases)	LogFC[b]	*P* value	FDR
IGHV7-4-1	immunoglobulin heavy variable 7-4-1	Immunoglobulin heavy locus at 14q32.33	14q32.33	486	4.4	2.07E-04	0.0307
TNFRSF12A	TNF receptor superfamily member 12A	CD molecules, Tumor necrosis factor receptor superfamily	16p13.3	3,939	2.2	2.40E-04	0.0312
RIMBP3	RIMS binding protein 3	N/A	22q11.21	6,105	-1.7	2.16E-04	0.0312
LOC100422382[a]	nocturnin pseudogene	N/A – pseudogene (lncRNA)	11q23.2	1,062	-2.0	3.36E-04	0.0369
CD69	CD69 molecule	CD molecules, C-type lectin domain containing	12p13.31	8,416	1.5	4.56E-04	0.0416
HSP90AA6P[a]	heat shock protein 90 alpha family class A member 6, pseudogene	N/A – pseudogene	4q33	23,989	4.7	5.66E-04	0.0450
TNFAIP3	TNF alpha induced protein 3	OTU domain containing	6q23.3	16,998	1.7	6.20E-04	0.0463
OLFM2	olfactomedin 2	N/A	19p13.2	82,835	-1.6	6.55E-04	0.0470
FOSL1	FOS like 1, AP-1 transcription factor subunit	Fos transcription factor family, Basic leucine zipper proteins	11q13.1	8,525	3.3	7.20E-04	0.0485
SMAD6	SMAD family member 6	SMAD family	15q22.31	80,749	1.6	7.12E-04	0.0485
WNT16	Wnt family member 16	Wnt family	7q31.31	15,738	-1.5	7.09E-04	0.0485
RN7SL749P[a]	RNA, 7SL, cytoplasmic 749, pseudogene	N/A – pseudogene (lncRNA)	10q26.11	290	1.7	7.45E-04	0.0490

[a] Non-protein coding.

[b] Log fold change corresponds to gene/transcript expression values from CpcPH patients divided by values from HFpEF without PH patients. Positive values indicate upregulation in CpcPH vs. HFpEF without PH patients, whereas negative values indicate downregulation in CpcPH vs. HFpEF without PH patients (i.e., upregulation in HFpEF without PH patients vs. CpcPH patients).

TABLE IX

SIGNIFICANT KEGG PATHWAYS UPREGULATED IN CPCPH PATIENTS VERSUS HFPEF WITHOUT PH PATIENTS USING GAGE METHOD

KEGG Pathway	Gage *P value*	Gage FDR	Differentially Expressed Genes (FDR<0.05)	logFC	*P* value	FDR
Cell cycle hsa04110	0.0003	0.0442[a]	*GADD45B*	0.73	0.0001	0.0268
			DBF4	0.38	0.0004	0.0401
			CDKN1A	1.23	0.0007	0.0475
Oocyte meiosis hsa04114	0.0044	0.3452	–	–	–	–
Complement and coagulation cascades hsa04610	0.0147	0.6011	–	–	–	–
p53 signaling pathway hsa04115	0.0152	0.6011	*GADD45B*	0.73	0.0001	0.0268
			CYCS	0.40	0.0005	0.0423
			PMAIP1	0.94	0.0006	0.0463
			CDKN1A	1.23	0.0007	0.0475
Spliceosome hsa03040	0.0342	0.6986	*SNRPD1*	0.44	1.59E-05	0.0137
			SNRPA1	0.93	0.0002	0.0311
			SRSF4	0.37	0.0004	0.0388
			TRA2B	0.61	0.0005	0.0425
			SRSF10	0.42	0.0006	0.0450
			XAB2	0.52	0.0007	0.0485
Proteasome hsa03050	0.0364	0.6986	*PSMC1*	0.84	1.27E-05	0.0136
			PSMA6	0.74	1.29E-05	0.0136
			PSMC6	0.57	8.51E-05	0.0245
			PSMA2	0.42	0.0003	0.0320
Progesterone-mediated oocyte maturation hsa04914	0.0375	0.6986	–	–	–	–
ECM-receptor interaction hsa04512	0.0441	0.6986	*ITGA10*	-0.87	0.0007	0.0490
Oxidative phosphorylation hsa00190	0.0448	0.6986	–	–	–	–
Vascular smooth muscle contraction hsa04270	0.0453	0.6986	*NPR2*	-0.70	0.0002	0.0297
Gap junction hsa04540	0.0490	0.6986	*NRAS*	0.42	0.0005	0.0429

[a] Significant based on FDR<0.05.

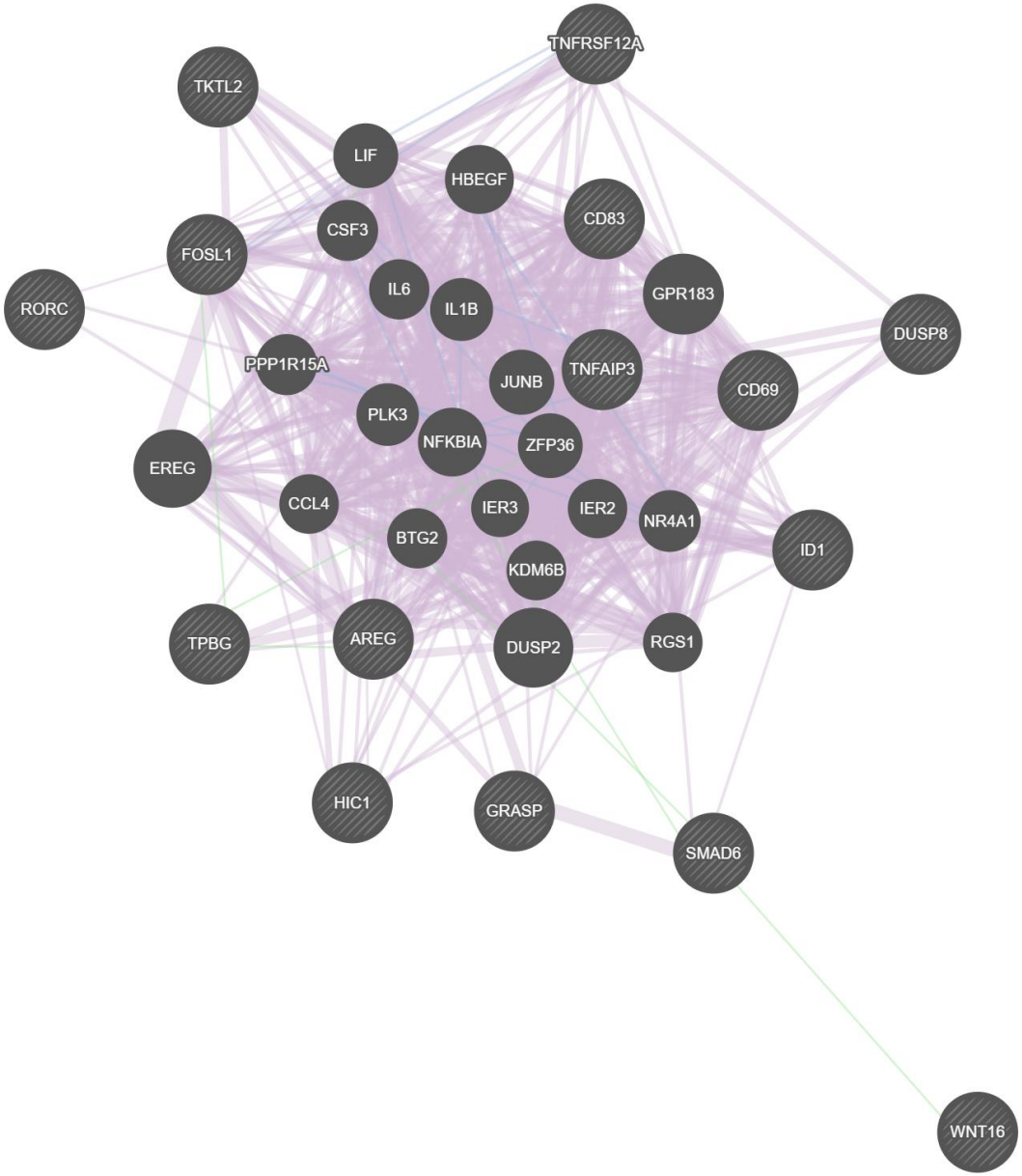

Figure 11. GeneMANIA network analysis. Circles containing diagonal lines represent genes that were input into this analysis and circles without diagonal lines depict genes that were output from this analysis. Networks: co-expression (pink lines); co-localization (blue lines); genetic interactions (green lines).

TABLE X

GENE EXPRESSION DATA FOR THE 20 ADDITIONAL GENES GENERATED FROM THE NETWORK ANALYSIS

Genes	LogFC[a]	*P* value[a]	FDR[b]
NFKBIA	1.15	7.68E-05	0.0010
DUSP2	1.40	0.0001	0.0013
HBEGF	1.42	0.0002	0.0015
PLK3	0.95	0.0013	0.0062
RGS1	1.22	0.0026	0.0100
PPP1R15A	0.74	0.0036	0.0114
EREG	0.64	0.0344	0.0934
JUNB	0.41	0.0570	0.1354
GPR183	-0.35	0.0816	0.1723
NR4A1	0.87	0.1091	0.2073
IER3	0.39	0.1305	0.2254
IL1B	-0.66	0.1322	0.2093
IL6	0.80	0.1908	0.2789
ZFP36	0.26	0.2748	0.3729
IER2	-0.15	0.4595	0.5820
CCL4	-0.19	0.5287	0.6278
BTG2	-0.16	0.5960	0.6661
KDM6B	0.05	0.9126	0.9633
LIF	-0.004	0.9948	0.9948
CSF3	Not found in our data set		

[a] Indicate values from our differential gene expression analysis.
[b] Calculated based on correction for multiple comparisons (19 above genes).

network weights add up to 100%, and indicate the relevance of the data sources for predicting membership in the queried gene list. The co-expression network, denoted by pink lines in Figure 11, was scored 98.10%; genes that are linked by pink lines have similar expression levels across conditions in gene expression studies collected from the Gene Expression Omnibus or other published literature. This means that when two genes are co-expressed, their transcript levels increase and decrease jointly across conditions. The co-localization network, denoted by blue lines in Figure 11, was scored 1.73%; genes that are linked by blue lines are both expressed in the same tissue or have gene products in the same cellular location. Lastly, the genetic interactions network, depicted by green lines in Figure 11, was scored 0.17%; genes that are linked by green lines are functionally associated per the Biological General Repository for Interaction Datasets or primary literature, meaning the impacts of perturbing one gene were found to be changed by perturbations to a second gene.

Based on review of primary literature for the genes used in the network analysis (n= 17) and additional genes revealed from the network analysis (n= 20), 30% (11 of 37) of these candidate genes have previously been associated with PH (Table XI). Of these candidate genes, 78% (29 of 37) have been associated with other pulmonary disease, including lung cancer (21 of 29) and other inflammatory and fibrotic conditions (Table XI). Five candidate genes did not appear to have any prior associations with PH or other pulmonary disease (*TPBG*, *TKTL2*, *DUSP8, SRSF4*, and *PSMC1*), however, they are all expressed in the lung. *TPBG*, *TKTL2*, *DUSP8* have proliferative, apoptotic, and/or oncological functions, whereas *SRSF4* and *PSMC1*A have associations with HF phenotypes (Table XI). Sixteen percent of the candidate genes (6 of 37), including the top signal *ID1*, were within or interact with the transforming growth factor β (TGF- β) signaling pathway (Table XI); dysregulation in this pathway has been linked to several subtypes of PH.

E. *In Silico* Expression Quantitative Trait Loci Findings

In order to perform *in silico* eQTL analyses using the GTEx portal, a stepwise process

TABLE XI
ASSOCIATIONS IN THE LITERATURE FOR DIFFERENTIALLY EXPRESSED GENES (LOGFC >|1.5| OR PATHWAY *P*<0.05)

Descriptor	Genes	LogFC[a]	FDR[a]	Association with PH	Subtype of PH	Studied Cell Type	Association with Lung Cancer	Association with other Lung Conditions	Associations outside the Lung	Within or interacts with TGF-β Pathway
Log_2-fold change >|1.5|	*ID1*	5.2	0.0078	Yes [72-78]	HPAH, hypoxia-induced PH (mice)	PAECs, PASMCs	Yes [79,80]			Yes (within)
	TPBG	-2.2	0.0136						Mediates inhibition of Wnt/ β-catenin signaling [81]	
	CD83	1.7	0.0166	Yes [82]	Sch-APAH	Lung tissue		COPD [83,84]		
	AREG	1.8	0.0197				Yes [85]	IPF [86]		
	RORC	-1.7	0.021					Pulmonary manifestations of Behcet disease,[87] lung response to acute ozone exposure [88]		Yes (downstream)
	TKTL2	3.2	0.0245						Cellular proliferation in uterine cancer [89]	
	DUSP8	2.3	0.0245						VEC apoptosis [90]	
	HIC1	1.6	0.0265	Yes [91]	CTEPH	PASMCs	Yes [92]			
	GRASP	1.5	0.0265					Asthma [93]		
	TNFRSF12A	2.2	0.0312				Yes [94-96]			
	RIMBP3	-1.7	0.0312					IPF with pirfenidone [97]		
	CD69	1.5	0.0416					IPF, pulmonary inflammation, emphysema [98-100]		
	TNFAIP3	1.7	0.0463	Yes [101,102]	SSc-APAH, CHD-APAH	lung tissue		Neutrophilic airway inflammation,[103] endotoxin-induced lung injury [104]		
	OLFM2	-1.6	0.047	Yes [105]	PH in sarcoidosis	PBMCs				Yes (induced by TGF-β)[106]
	FOSL1	3.3	0.0485				Yes [85,107-109]			
	SMAD6	1.6	0.0485	Yes [110]	MCT-induced PAH (rats)	PASMCs	Yes [111,112]	IPF [113]		Yes (within)
	WNT16	-1.5	0.0485				Yes [114]	Bleomycin-induced lung injury and fibrosis [115]		Yes (TGF-β-inhibits WNT16 expression)[116]

CHD-APAH- congenital heart disease-associated PAH; COPD- chronic obstructive pulmonary disease; CTEPH- chronic thromboembolic PH; HPAH- heritable PAH; IPAH- idiopathic PAH; IPF- idiopathic pulmonary fibrosis; LVDD- left ventricular diastolic dysfunction; MCT- monocrotaline; PAECs- pulmonary artery endothelial cells; PASMCs- pulmonary artery smooth muscle cells; PBMCs- peripheral blood mononuclear cells; RVF- right ventricular failure; Sch-APAH- Schistosomiasis-associated PAH; SSc-APAH- Systemic sclerosis-associated PAH; VEC- vascular endothelial cell; VSM-vascular smooth muscle.

[a] Indicate values from our differential gene expression analysis.

TABLE XI (continued)
ASSOCIATIONS IN THE LITERATURE FOR DIFFERENTIALLY EXPRESSED GENES (LOGFC >|1.5| OR PATHWAY *P*<0.05)

Descriptor	Genes	LogFC[a]	FDR[a]	Association with PH	Subtype of PH	Studied Cell Type	Association with Lung Cancer	Association with other Lung Conditions	Associations outside the Lung	Within or interacts with TGF-β Pathway
Cell cycle and p53	*GADD45B*	0.73	0.0268	Yes [117]	IPAH	PAECs	Yes [118-120]			Yes (effector of TGF-β-induced apoptosis)[121]
	DBF4	0.38	0.0401				Yes [122-124]			
	CDKN1A	1.23	0.0475	Yes [125,126]	Hypoxia-induced PH (mice)	PAECs, PASMCs	Yes [127,128]			
	CYCS	0.4	0.0423	Yes [129]	Fenfluramine-associated PAH	PASMCs	Yes [130]			
	PMAIP1	0.94	0.0463				Yes [131]			
VSM contraction	*NPR2*	-0.7	0.0297					OSA [132]		
Spliceosome	*SNRPD1*	0.44	0.0137				Yes [133,134]			
	SNRPA1	0.93	0.0311				Yes [134]			
	SRSF4	0.37	0.0388						Loss of SRSF4 resulted in cardiac hypertrophy, LVDD and increased risk of sudden death[135]	
	TRA2B	0.61	0.0425				Yes [136]			
	SRSF10	0.42	0.0450					Smoking exposure [137]		
	XAB2	0.52	0.0485				Yes [138]			
Proteasome	*PSMC1*	0.84	0.0136						RVF in carvedilol-treated rats with PH [139]	
	PSMA6	0.74	0.0136				Yes [140,141]			
	PSMC6	0.57	0.0245	Yes [142]	PAH (*in vivo*)	PASMC	Yes [143]			
	PSMA2	0.42	0.0320	Yes [142]	PAH (*in vivo*)	PASMC	Yes [141]			
ECM-R Interaction	*ITGA10*	-0.87	0.0490				Yes [144]	Severe COPD [145]		
Gap junction	*NRAS*	0.42	0.0429				Yes [146,147]	Pulmonary Langerhans cell histiocytosis [148]		

CHD-APAH- congenital heart disease-associated PAH; COPD- chronic obstructive pulmonary disease; CTEPH- chronic thromboembolic PH; HPAH- heritable PAH; IPAH- idiopathic PAH; IPF- idiopathic pulmonary fibrosis; LVDD- left ventricular diastolic dysfunction; MCT- monocrotaline; PAECs- pulmonary artery endothelial cells; PASMCs- pulmonary artery smooth muscle cells; PBMCs- peripheral blood mononuclear cells; RVF- right ventricular failure; Sch-APAH- Schistosomiasis-associated PAH; SSc-APAH- Systemic sclerosis-associated PAH; VEC- vascular endothelial cell; VSM-vascular smooth muscle.

[a] Indicate values from our differential gene expression analysis.

was taken to determine which differentially expressed genes to investigate for known eQTLs driving their expression, the process and results of which are shown in Figure 12. In summary, of the 346 differentially expressed genes/transcripts between CpcPH and HFpEF without PH patients, 298 genes/transcripts had Entrez IDs. Of these, 26 genes/transcripts had a logFC>1.5 or <-1.5, and 19 of them were expressed in one of the five tissues of interest to our phenotype (i.e. lung, atria, left ventricle, aorta, coronary artery) based on GTEx portal interrogation. Seven transcripts/genes were not found in GTEx portal. Seventeen of 19 genes had eQTLs in GTEx portal, although only 6 of these genes had eQTLs in one of our tissues of interest. In these 6 genes (i.e. *TPBG, AREG, RN7SL749P, CD83, OLFM2, and WNT16*), there were a total of 1415 eQTLs, which was further reduced to 668 unique eQTLs after removing duplicates across tissues for each gene. After LD pruning, there were 86 candidate eQTLs remaining for further genetic association analysis in HF patients with and without PH to determine whether patients with these variants were more likely to have PH, while controlling for age, sex, race, and smoking status.

Table XII displays the results from these *in silico* eQTL analyses in GTEx portal, which shows the following for 19 genes: median expression values in the lung (in transcripts per million, TPM), instances where lung was ranked within the top 5 tissues with highest expression, and tissues where eQTLs were found and how many candidate eQTLs are known within those tissues. There were 9 genes for which lung tissue was ranked among the top five tissues with highest expression (Table XII), two of which (*ID1* and *SMAD6*) are displayed in Figure 13. In this figure, the upper and lower boundaries of the box plots mark the 75th and 25th percentiles. Circles represent outliers, meaning they are above or below 1.5 multiplied by the interquartile range. The neon green box plots represent lung tissue expression. Some notable eQTL box plots are shown in Figure 14.

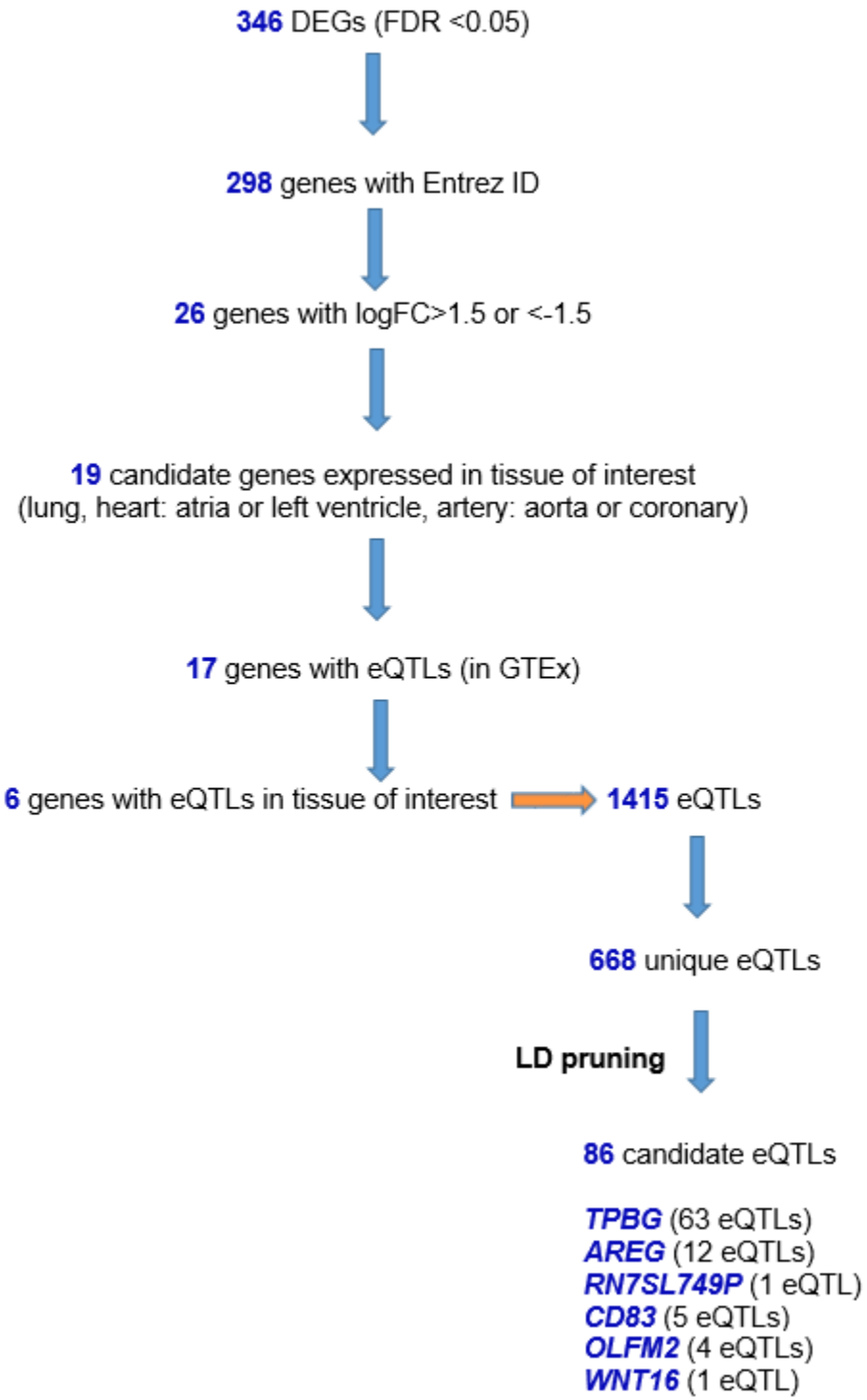

Figure 12. Flowchart showing the stepwise process of determining candidate eQTLs from the list of differentially expressed genes. DEG- differentially expressed gene.

TABLE XII

***IN SILICO* EQTL ANALYSIS RESULTS FROM GTEX PORTAL**

Gene Symbol	Median expression in lung (TPM)	Lung was ranked within top 5 tissues with highest expression	Tissue of interest with eQTLs	# unique eQTLs in tissue of interest
ID1	291.83	Yes, #5	–	–
GRASP	127.14	Yes, #1	–	–
TNFRSF12A	94.88	Yes, #3	–	–
TNFAIP3	74.1	Yes, #3	–	–
SMAD6	52.14	Yes, #1	–	–
CD83	34.62	Yes, #5	atria, LV	31
CD69	27.31	Yes, #3	–	–
DUSP8	18.53	–	–	–
FOSL1	15.08	Yes, #5	–	–
HIC1	14.31	–	–	–
OLFM2	12.14	–	lung, coronary artery	4
TPBG	6.74	–	lung, aorta, coronary artery, atria, LV	469
RORC	5.2	–	–	–
AREG	1.94	Yes, #1	lung, aorta	156
RN7SL749P	0.8	–	lung, aorta	2
WNT16	0.35	–	coronary artery	6
HSP90AA6P	0.09	–	–	–
TKTL2	0.05	–	–	–
RIMBP3	0.05	–	–	–

LV- left ventricle.

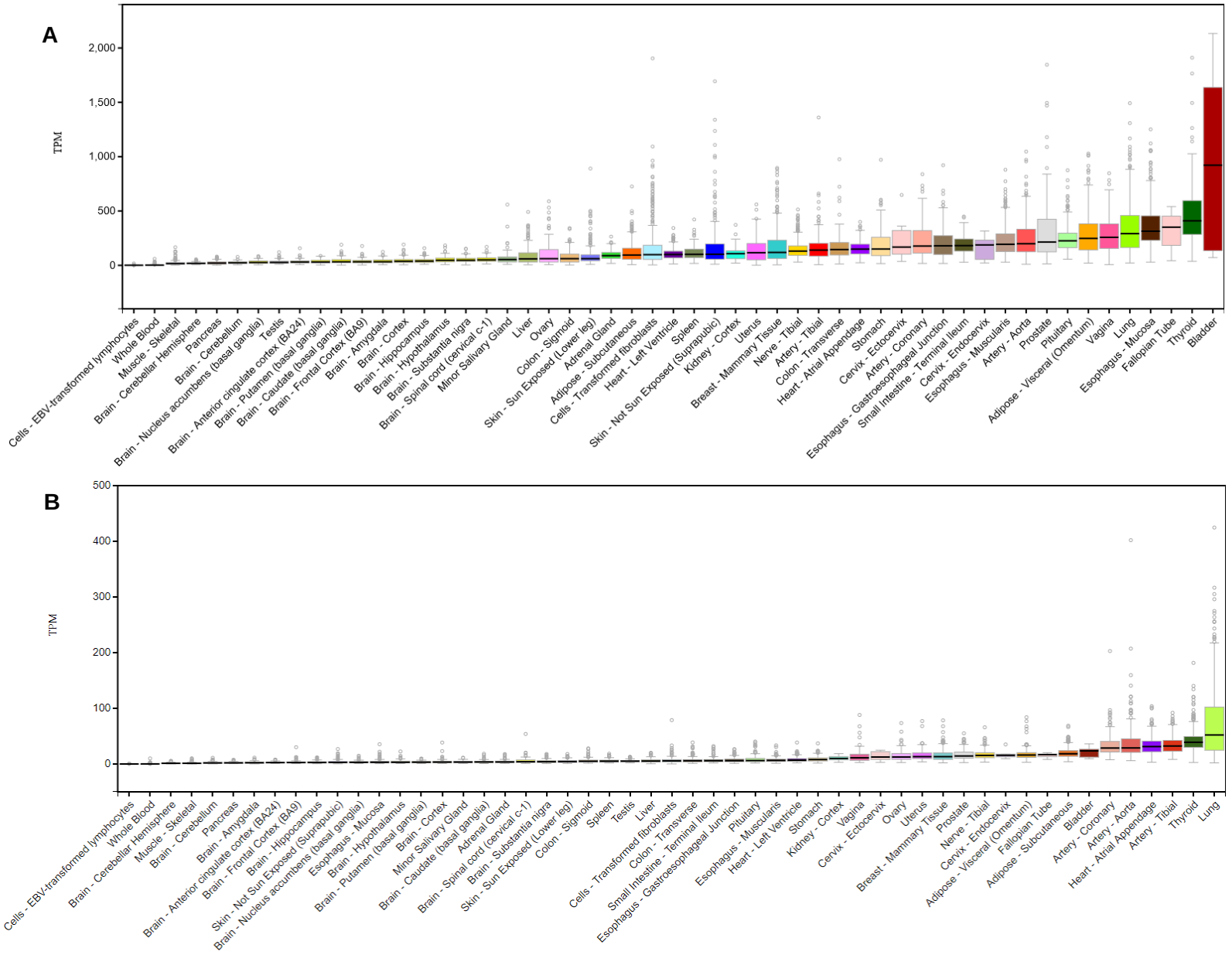

Figure 13. Representative box plots of *ID1* (A) and *SMAD6* (B) expression within all tissues in GTEx portal.

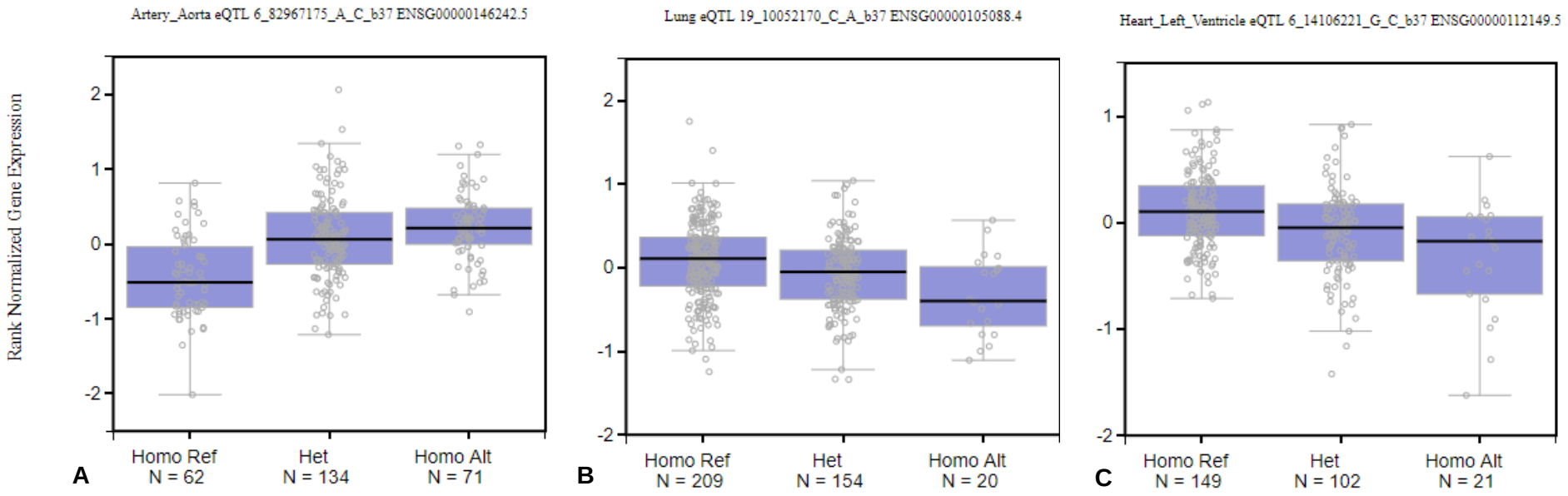

Figure 14. Representative expression quantitative loci box plots from GTEx portal for rs950611 within *TPBG* in the aorta (A), rs36110067 within *OLFM2* in the lung (B), and rs9464657 within *CD83* in the left ventricle.

F. Genetic Association of Expression Quantitative Trait Loci

Of the 86 candidate SNPs discovered from *in silico* eQTL analysis in GTEx portal, genotype array data was available for only 15 SNPs in the 165 HF patients from the HFGDB. All SNPs met the HWE threshold *P* value of above 0.003 (0.05/15). No SNPs had more than 10% of samples missing genotype data. Fifteen samples were excluded due to missing genotype data for more than 10% of SNPs and no SNPs were monomorphic. Logistic regression results for 76 HF-PH and 74 HF without PH patients are displayed in Table XIII, which showed that none of the 15 SNPs located within genes *CD83*, *TPBG*, and *OLFM2* were significantly associated with PH in HF patients in our larger dataset. Expression box plots from GTEx portal for the three SNPs with the strongest associations (i.e. rs950611, rs36110067, and rs9464657) are shown in Figure 14.

TABLE XIII

LOGISTIC REGRESSION RESULTS FROM GENETIC ASSOCIATION ANALYSIS WITHIN HFGDB

Gene	Chr	SNP	A1	A2	MAF	OR	SE	95% CI (lower)	95% CI (upper)	*P* value
CD83	6	rs72836546	T	C	0.09	0.65	0.43	0.28	1.51	0.32
CD83	6	rs9464657	C	G	0.17	0.62	0.34	0.32	1.20	0.15
TPBG	6	rs62429917	G	A	0.30	0.86	0.26	0.51	1.44	0.57
TPBG	6	rs1890118	T	C	0.19	1.30	0.29	0.73	2.31	0.37
TPBG	6	rs950611	A	C	0.33	1.48	0.25	0.91	2.40	0.11
TPBG	6	rs507500	C	T	0.49	0.90	0.23	0.57	1.40	0.63
TPBG	6	rs72904360	G	A	0.30	1.10	0.27	0.65	1.87	0.73
TPBG	6	rs2208422	T	G	0.29	0.71	0.27	0.42	1.20	0.20
TPBG	6	rs6926639	G	A	0.08	0.84	0.45	0.34	2.05	0.70
TPBG	6	rs1321796	G	A	0.26	1.14	0.32	0.61	2.15	0.68
TPBG	6	rs3011887	G	A	0.30	1.11	0.29	0.63	1.95	0.71
TPBG	6	rs1321793	C	T	0.40	1.04	0.26	0.62	1.73	0.90
TPBG	6	rs2458536	T	C	0.48	0.96	0.23	0.61	1.51	0.86
OLFM2	19	rs36110067	A	C	0.31	1.45	0.25	0.89	2.38	0.14
OLFM2	19	rs10425175	T	C	0.29	1.16	0.26	0.70	1.92	0.56

A1- allele 1 (minor allele); A2- allele 2 (major allele); Chr- chromosome; CI: confidence interval; MAF: minor allele frequency; OR- odds ratio (regression coefficient for allele 1); SE: standard error of the odds ratio.

IV. DISCUSSION

With RNA-seq data analysis, there is no agreed upon pipeline and there are certain unavoidable biases, though many can be corrected. Although the per base sequence content chart in FastQC revealed that there was some selection bias in the sequence composition of the first 12 bases of our reads (Figure 2), this is a common bias seen with Illumina sequencing due to random hexamer priming during library generation. Some experts believe this cannot be corrected by trimming and in most instances does not appear to adversely affect downstream analysis.[149] We did not perform quality-based trimming on our reads based on evidence that suggests that untrimmed reads are most likely to result in the highest accuracy for expression estimates from RNA-seq.[150] Further, we did not trim our reads for adapters based on literature that shows that adapter trimming does not significantly improve the significance of RNA-seq biological signals.[151] Since the alignment program TopHat has been designed to handle unmapped portions of the reads, like adapters,[53] the data processed by it may not be remarkably improved by adapter trimming.[151] In addition to this selection bias, we did not correct for gene length bias because count data (rather than Fragments Per Kilobase of transcript per Million mapped reads [FPKM]) is required to use edgeR for differential gene expression. Some experts have shown that many biological replicate data from unrelated samples do not suffer from this bias except for genes with reduced read counts (small gene variance).[152] Since our samples were biological replicates from unrelated patients and we used CPM to filter out genes with low expression, we expect that gene length bias had a relatively small impact on our findings.

The results of this transcriptome-wide study suggest that HFpEF patients with CpcPH may have systemic pathophysiological differences that are measurable in PBMCs, some of which overlap with other subtypes of PH. This study confirms that there is a complex network of molecular pathways that modulate the intricate CpcPH phenotype, involving inflammation, cell proliferation, and mitochondrial dysfunction. To the best of our knowledge, this is the first study

to use RNA-seq data to investigate Group 2 PH. Previous genetic/genomic studies related to this specific phenotype have either utilized a candidate gene approach with genotyping several SNPs,[32] or used GWAS data in mice[33] or humans.[31] Group 2 PH patients are an understudied patient population in genomic/transcriptomic studies, especially compared to their Group 1 PH counterpart (PAH), on which the majority of studies focus their investigations.[117,153-159]

In general, it can be difficult to study Group 2 PH in humans for several reasons: 1) right heart catheterization is required to properly diagnose PH and often to enroll patients in prospective PH trials, but it is an invasive and costly procedure that requires an individual to take a half to full day off of work; 2) there is heterogeneity within the phenotype. With respect to the latter, patients can have more than one type of PH as classified by the World Health Organization (e.g. Group 2 and 3), numerous comorbidities, and/or varying phenotypes of HFpEF.[160] Despite these challenges, we aimed to identify transcriptomic predictors of IpcPH and CpcPH development in HFpEF patients using RNA-seq data analysis, in an attempt to find biological pathways that could be targets for novel treatment strategies, as there are currently no approved treatments for these patients.

Herein, we have shown that 346 genes were differentially expressed (FDR<0.05) between CpcPH and HFpEF without PH, when adjusting for age, sex, race, and smoking status. We did not find any differentially expressed genes between HFpEF without PH and IpcPH nor between IpcPH and CpcPH patients. These lack of significant findings in the IpcPH analyses could suggest that these analyses were underpowered, as a larger sample size may be needed to detect differentially expressed genes due to a likely smaller effect size for IpcPH than seen between CpcPH and HFpEF without PH. The effect size is potentially smaller because IpcPH may be in the middle of the phenotypic spectrum between HFpEF without PH and CpcPH pathogenesis/severity. However, recent findings raise suspicions about the hypothesis that CpcPH is merely a result of sustained exposure to high left-sided pressures. In a study by Assad *et. al.*, CpcPH patients were younger but had more pronounced hemodynamic

imbalances (in right atrial and pulmonary arterial pressures, PVR, and pulmonary arterial compliance) than IpcPH patients, although similar chronicity, severity of left ventricular dysfunction, and co-morbidities (e.g. vascular disease, left heart disease, obesity, and diabetes), which the authors suggested may indicate an alternative reason for the inflated pulmonary vascular disease in CpcPH patients.[31] One such reason could be a biological predisposition, which our findings support with respect to dysregulation of certain gene signaling pathways. Further, the severe hemodynamics in CpcPH patients were similar to those observed in PAH patients, which other studies have corroborated.[161] Thus, the effect size for CpcPH versus HFpEF without PH is likely much larger than that of IpcPH compared to CpcPH or HFpEF, which presumably played a role in our discovery of several significant associations between CpcPH and HFpEF with a relatively small sample size, but none for IpcPH comparisons.

In exploratory genetic findings from the Assad *et. al.* study, 141 genes were associated with shared risk for CpcPH and PAH compared to IpcPH patients, which was a result of testing their hypothesis that CpcPH develops, in part, because of a shared genetic risk with PAH.[31] It seems likely that these investigators found genes associated with risk for CpcPH versus IpcPH, unlike in our study, as they had a larger sample size (139 IpcPH, 36 CpcPH, and 79 PAH), they used a less conservative significance threshold, and they combined CpcPH and PAH patients into one cohort (potentially enabling them to tease out the effects of the pre-capillary component of pulmonary vascular disease). Despite these differences, these findings are interesting as they may implicate genetic similarities between PAH and CpcPH, which our findings support as well, as described later in this discussion. Along the same lines, another Assad *et. al.* study demonstrated that CpcPH patients have a pulmonary vascular physiology that more closely resembles that of PAH patients versus IpcPH patients.[162]

Another potential reason that we were unable to find differentially expressed genes between IpcPH and HFpEF without PH or CpcPH is our lack of defined clustering of the three

groups on our principle component analysis plot. Based on this plot, it appears clear that there is more variability in our CpcPH and HFpEF without PH samples compared to our IpcPH samples. The IpcPH samples displayed the tightest clustering, perhaps indicating that IpcPH patients tend to have less heterogeneity than either CpcPH or HFpEF alone. By the nature of IpcPH presentation, it makes sense that our IpcPH patients had smaller variability in certain hemodynamics (i.e. mPAP, PVR, TPG, and DPG), had a lower prevalence of chronic lung disease, and were older on average.[31,163]

Although not significant, our IpcPH patients had a trend towards significantly less chronic lung disease (12.5%) compared to our HFpEF and CpcPH patients (both 60%). It could be possible that this influenced our lack of significant findings for IpcPH patients, due to the potential increase in the noise of the signals from PH due to signals from other lung conditions. Alternatively, this decreased prevalence of chronic lung disease could have made the signals even stronger, due to reduced inflammatory processes, and therefore lack of significant findings could merely reiterate that the effect size between IpcPH and CpcPH or HFpEF without PH is likely small, warranting larger samples sizes to see an effect. Or perhaps IpcPH has genetic similarities to HFpEF and CpcPH that make it difficult to discern the signals. There are a number of possibilities with respect to pathophysiology of IpcPH and potential links to CpcPH progression, which are incompletely understood.

We took multiple approaches to functionally annotate our gene expression results, including pathway analysis, network analysis, *in silico* eQTL analyses, and genetic association analysis of eQTLs. Additionally, in order to provide insight into whether our findings may be unique to Group 2 PH or common among other PH subtypes, we performed literature searches on differentially expressed candidate genes (FDR< 0.05) with logFC > |1.5| or within significant pathways ($P < 0.05$) (Table XI) to determine if there have been any previous associations between them and PH. With the large number of differentially expressed genes between CpcPH and HFpEF, we set logFC, FDR, and other thresholds in order to decide which genes to

functionally annotate. This may have led us to overlook certain interesting findings, however, herein we will discuss overall trends and focus in on select genes within pathways, the constructed network, or the literature. We plan to further investigate the other significant genes outside these thresholds in future analyses.

Our pathway analysis (Table IX) revealed that the most significantly enriched pathway between our CpcPH and HFpEF without PH controlled the cell cycle. Pulmonary hypertension appears to be a multifaceted pan-vasculopathy that mimics certain characteristics seen in cancer, like inflammation, increased proliferation, resistance to apoptosis, and fibrosis due to drastic remodeling of the extracellular matrix.[164] As such, dysregulation in the ability of cells in the pulmonary vasculature to progress through the phases of the cell cycle could lead to cell proliferation and apoptosis resistance. Genes from our data set within the cell cycle control pathway include *GADD45B, CDKN1A, and DBF4.*

Expression of *GADD45B* (Growth Arrest and DNA Damage Inducible Beta) is induced by stressful growth arrest conditions, and it has been shown to play a role in TGF-β induced apoptosis by acting upstream of p38 activation.[121] *CDKN1A* (Cyclin Dependent Kinase Inhibitor 1A) regulates cell cycle progression at G1 phase, and its expression is tightly controlled by tumor suppressor protein p53; through its inhibition of cyclin-dependent kinase activity, it appears to block cell cycle progression in response to stress stimuli by preventing phosphorylation of important cyclin-dependent kinase substrates. Of interest, hypoxia has been shown to inhibit *CDKN1A*, which would allow cell cycle progression.[125] One proposed mechanism for this hypoxia-induced repression of *CDKN1A* is via the activity of microRNA family miR-130, which decreased *CDKN1A* expression *in vitro* and significantly increased smooth muscle cell proliferation.[125] Alternatively, exogenous nitric oxide upregulates *CDKN1A* in rat pulmonary microvascular smooth muscle cells, a mechanism by which it may wield its anti-proliferative effects.[165] Peculiarly, both *GADD45B* and *CDKN1A* had opposite directions of effect in our data set than previously reported in other PH subtypes, since they were upregulated in

our CpcPH patients and downregulated in Group 1 and 3 PH studies, respectively.[117,125,126] DBF4 (DBF4 Zinc Finger) is a regulatory subunit for CDC7 and plays a key role in DNA replication (S phase) and cell proliferation. In breast cancer, it has been demonstrated that increased *DBF4* expression was correlated with loss of p53, which would have pro-proliferative implications.[123]

Another enriched pathway that overlaps with the cell cycle is the p53 signaling pathway, which, in addition to *GADD45B* and *CDKN1A,* contains *CYCS,* and *PMAIP1* (Table IX). p53 is one of the most known tumor suppressor proteins, as genetic variation and reduced activity of this gene has been associated with cancer.[166] p53 is a critical transcriptional activator of p53-regulated genes, and its activation is induced by various stress stimuli, including oxidative stress, DNA damage, and activated oncogenes.[167] This signaling pathway results in cell cycle arrest, cellular senescence, or apoptosis.[167] Compared with wild type mice, p53 knock out mice developed worse PH under hypoxia.[168] Whereas inactivation of p53 with pifithrin-α induced pulmonary vascular remodeling in rats with monocrotaline-induced PH.[169] One significant gene that we revealed in the p53 signaling pathway was *CYCS* (Cytochrome C, Somatic), which is a key part of the electron transport chain in mitochondria, while *PMAIP1* (Phorbol-12-Myristate-13-Acetate-Induced Protein 1) stimulates changes in the mitochondrial membrane.

Being that mitochondria play an essential role in energy production, respiration, and modulation of calcium signaling, investigators have turned their attention to its role in the pathogenesis of numerous diseases.[170] The “metabolic theory” of disease proposes that changes in an organism’s bioenergetics, shifting from energy production primarily through aerobic respiration to fermentation and glycolysis, can lead to dysfunctional downstream processes, subsequently causing disease.[23] And although it has been explained in most detail in cancer,[171-174] recently it has been expanded to the pathogenesis of PH.[175-177] This metabolic shift (coined the “Warburg effect”) impacts endothelial cells, especially since they are first to detect a low oxygenated environment and they have the ability to signal to cells around them.[23]

The signaling cascade leads to quick vasoconstriction of the pulmonary bed in an effort to maintain ventilation-perfusion matching.[178] Apart from deceased efficiency in energy production, this process leads to rapid changes in the production of reactive oxygen species with worsened handling of oxidative stress, changes in oxygen-sensing potassium channels, resultant changes in cytosolic calcium, and pulmonary vasculature constriction.[23] The mitochondria depend on reactive oxygen species for proper signaling and internal regulation, but, the distinguishing characteristic of mitochondrial or metabolic disease is an imbalance of oxidative stress.[179]

Related to this discussion on the biology of mitochondria, another pathway revealed in our pathway analysis was oxidative phosphorylation, which is a process of energy production in mitochondria through which energy (ATP) is formed via the transfer of electrons. In a review by Marshall *et. al.* on the role of mitochondrial dysfunction in PH,[23] authors explain that HIF-1α (Hypoxia inducible factor-1α) is one of the most studied metabolic pathways in PH. Expression of *HIF-1α* is usually upregulated by hypoxia,[180] and its upregulation activates over 100 genes, controlling energy metabolism, vasomotor tone, apoptosis, angiogenesis, and erythropoiesis.[181] In endothelial cells of PH patients under normoxia, high levels of nitric oxide led to increased *HIF-1α* expression, while low levels of supplemental nitric oxide reduced *HIF-1α*.[182] Increased levels of nitric oxide reduce the need for endothelial cells to produce their own nitric oxide, which alludes to the fact that loss of nitric oxide production, through reduced endothelial nitric oxide synthesis, may lead to activation of HIF-1α under normoxia in PH patients.[182] In our data set, *HIF-1α* was upregulated in CpcPH compared to HFpEF (logFC 0.50, FDR= 0.05), a similar direction of effect to other PH studies, suggesting that mitochondrial dysfunction mediated by hypoxia or nitric oxide levels may be one aspect of CpcPH development in HFpEF patients. Furthermore, *HIF-1α* has been linked in the literature to several other significant genes in our data set (*RORC* [Table VIII] and genes from the proteasome pathway, *PSMC1, PSMA6, PSMC6, PSMA2* [Table IX]). HIF-1α activates transcription of *RORC,* thereby playing a key role in T helper 17 cell differentiation,[183] and HIF-1α is degraded via the proteasome pathway.[184]

These results highlight the pervasive role of HIF-1α in mediating energy metabolism and inflammation.

As discussed earlier, PH has many pathogenic mechanisms similar to cancer.[185] Interestingly, over half of our candidate differentially expressed genes have had previous associations with lung cancer (Table XI). Shared pathological mechanisms respective to cell growth between PH and cancer include sustaining proliferative signaling, evading growth suppressors, resisting cell death, limitless replicative potential, and genome instability and mutation.[185] With respect to the first mechanism, upregulated *EGFR* (Epidermal growth factor receptor) has been implicated in cancer, as it mediates cell proliferation and protection from apoptosis.[186] *EGFR* has been linked to pathogenesis of Group 1 PH in rats[187], as well as Group 2 PH in mice.[33] Although *EGFR* was not present in our data set, six of our 346 differentially expressed genes/transcripts (*MRPL38, PHLDB2, AREG, TGFA, HBEGF*, and *NRAS*) are found within the EGFR signaling pathway per PANTHER pathway analysis.[188] This pathway was tied with the p53 signaling pathway for most significant genes within a pathway. *EGFR* may have been filtered out of our data set due to our CPM filter/normalization or there may have been poor coverage of this gene during sequencing. *EGFR* may play a role in CpcPH development, but more research is needed in this area.

In addition to sustaining proliferative signaling, evading growth suppressors is another pathological mechanism of cancer. p53 has already been discussed as an example of this mechanism, however, TGF-β has tumor suppressor activities as well.[185] TGF-β inhibits growth of numerous cell types by inhibiting cyclin-dependent kinase complexes that inactivate retinoblastoma protein.[189] Inactivating mutations in TGF-β receptors and *SMAD* genes have been implicated in cancer.[190] Group 1 PH has also been associated with dysregulation in the TGF-β signaling pathway.[191] It is possible that the TGF-β signaling pathway has a role in Group 2 PH as well, because our top differentially expressed gene (based on FDR and logFC), *ID1*, is part of this pathway, and some of our other candidate genes are within or interact with this

pathway (*SMAD6, OLFM2, RORC, WNT16, GADD45B,* Table XI). ID1 (inhibitor of DNA binding 1, HLH protein) regulates proliferation of pulmonary artery smooth muscle cells by cell cycle inhibition.[72,75,76] Additionally, *SMAD6* (SMAD family member 6) is an inhibitory SMAD that is induced by TGF-β ligands,[112] which has been shown to repress bone morphogenetic protein-induced *ID1* transcription.[192] Both *ID1* and *SMAD6* had the same direction of effect (i.e., both genes were upregulated) in our data set as in other studies.[72-78,110] *OLFM2* (olfactomedin 2) is involved in TGF-β induced smooth muscle cell differentiation.[106] Lastly, *WNT16* also plays a role in vascular smooth muscle cell activity, as it attenuates TGF-β induced chondrogenic transformation in these cells.[116]

In addition to our results that suggest some overlap between CpcPH development and cancer pathology, a striking finding is that nearly a third of these candidate genes have had previous PH associations (Table XI): 8 genes with Group 1 PH, 2 genes with Group 3 PH, 1 gene with Group 4 PH, and 1 gene with Group 5 PH (PH in sarcoidosis). These findings may support some pathobiological overlap between Group 2 PH and other PH subtypes. This also has been asserted by Assad *et. al.* for Group 1 and Group 2 PH.[31] Although there are some common genes and pathways among the different PH subtypes, a number of genes have not yet been identified in Group 2 PH pathogenesis. This study has provided numerous candidate genes associated with CpcPH development that warrant further investigation.

There appears to be an overlap of pro-proliferative, anti-apoptotic stimuli and signaling pathways among the different PH subtypes, and developing a strategy to target a single receptor or pro-mitogenic factor may have restricted and/or temporary effects.[185] Pullamsetti *et. al.* discuss that the awareness of major signaling pathways cross-talking and inter-controlling each other by common signaling molecules at areas below the receptors has led to an interesting potential treatment strategy to reverse pulmonary vascular remodeling.[185] There may be an opportunity to target downstream effectors (called “signaling hubs”) that incorporate signals from multiple receptors, induce cell cycle entry and progression, and cell proliferation.[185]

In another review article by Pullamsetti *et. al.*, the authors build upon the "signaling hubs" concept, highlighting that numerous, complex signaling pathways modulate the activity of particular sequence-specific DNA-binding transcription factors and coregulators that are crucial for transcriptional regulation of gene expression that contributes to vascular cell phenotypes in PH.[193] Multiple pathologies and stimuli (e.g. oxidative stress, hypoxia, infection) can activate these signaling cascades that lead to PH. However, eventually different signaling pathways unite to target the activity of specific transcription factors.[194] This results in activation of a "vascular gene program" in the nucleus that is displayed ultimately as the vascular phenotype in PH patients.[193] The authors explain that comprehension of the differential gene expression in diseased cells will provide the foundation for innovative treatment plans that involve gene expression manipulation. And that targeting transcription factors is one novel way that can enable alteration of gene activation/suppression.[193]

In line with the importance of transcription factors in facilitating PH-related vascular phenotypes, more than a quarter of our 346 differentially expressed genes/transcripts have the category "binding" for molecular function (Figure 15), of which about half encode sequence-specific DNA- or RNA-binding transcription factors (Figure 15).[188] Some of these DNA-binding TFs include *HIF-1α*, *ID1*, *SMAD6, RORC*, *HIC1*, and *FOSL1*. These genes warrant further investigation to explore their potential role in CpcPH pathogenesis and possible role as therapeutic targets.

Our network analysis (Figure 11) supports the interconnectedness and co-expression of our top signals, and revealed additional related genes of lower effect size (logFC ≤ |1.5|) that may also be of importance in CpcPH pathogenesis (Table X). Of the 20 additional genes revealed from this analysis, 6 were significant in our dataset at FDR < 0.05 (*HBEGF*, *NFKBIA, DUSP2, PLK3, RGS1, PPP1R15A*). HBEGF (Heparin Binding EGF-like Growth Factor) is a growth factor that is a stronger mitogen for smooth muscle cells than EGF and it can bind the

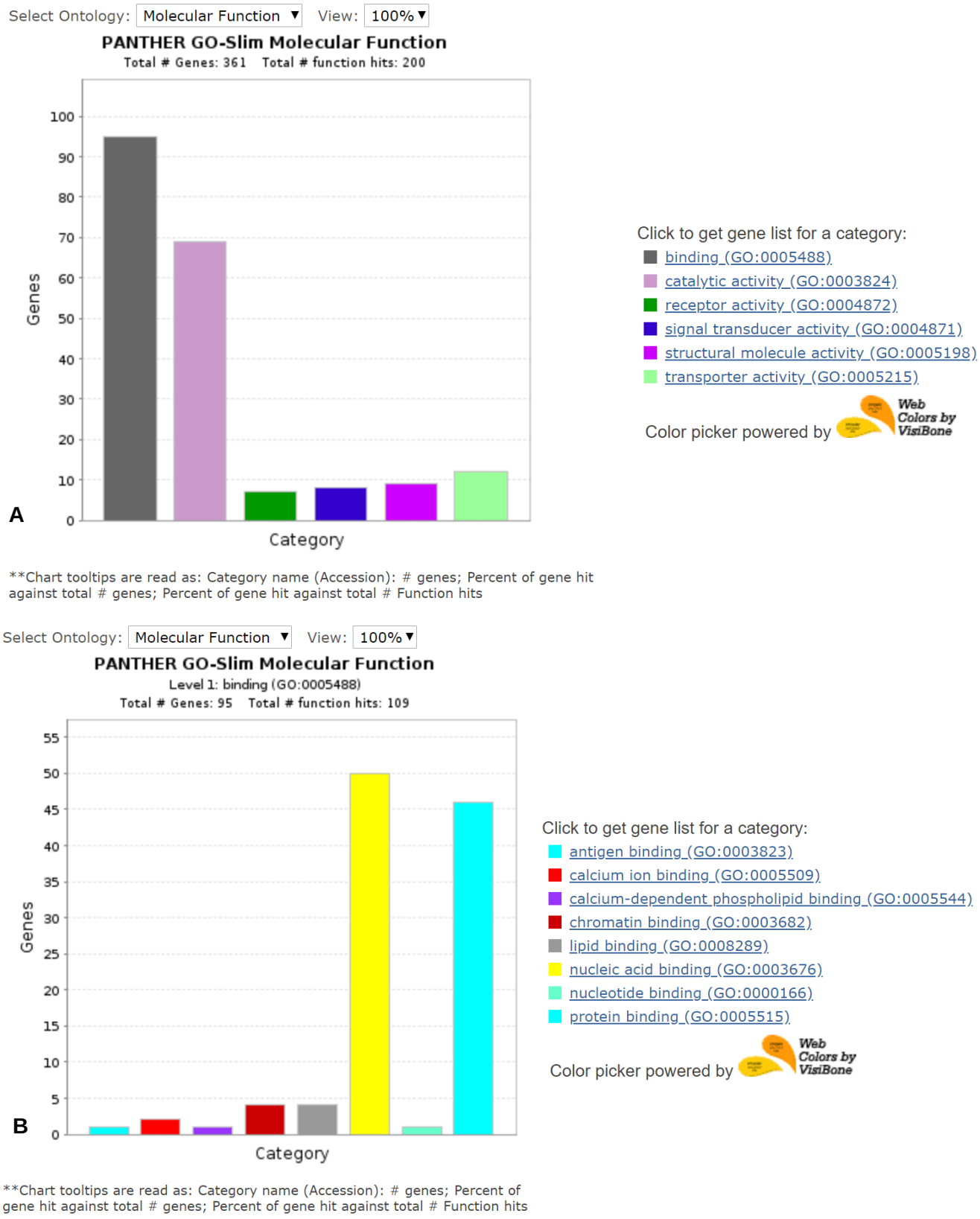

Figure 15. PANTHER gene ontology charts that display molecular function for our differentially expressed genes (A) and the subset involved in “binding” (B).

EGF receptor with greater affinity.[195] Although *HBEGF* has been associated with hypoxia-induced PH in rats[196] and was similarly upregulated in our study, a more recent study that included human lung tissue was unable to validate this.[197] While it is unclear if upregulation of *HBEGF* is unique to CpcPH pathogenesis compared to other PH subtypes, it is a candidate gene worthy of further investigation. Since its affinity for the EGF receptor is higher than EGF itself, perhaps HBEGF has a competitive advantage to sustaining proliferative signaling. As previously mentioned, *EGFR* was not found in our data set, so we are unable to confirm whether this association seen with *HBEGF* and CpcPH holds true for *EGFR* as well.

NFKBIA (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha) is a transcription factor that plays a crucial role in inflammation, cellular proliferation, differentiation, immunity, and survival.[198] It has been previously associated with hypoxia-induced PH *in vitro,*[199] with the same direction of effect in our data set (i.e. upregulated). *DUSP2* (dual-specificity phosphatase 2) negatively regulates members of the mitogen-activated protein kinase superfamily and has been associated with PAH and PH secondary to idiopathic pulmonary fibrosis.[154] Based on the known functions of *HBEGF*, *NFKBIA,* and *DUSP2*, as well as *PLK3, RGS1, PPP1R15A*,[200-203] it seems likely that these associations are not due to chance, thus, they may play a role in CpcPH development in HFpEF patients.

Regarding the *in silico* eQTL analysis (Table XII), after excluding non-coding transcripts or pseudogenes that were not available in GTEx portal, it was discovered that all remaining candidate genes were expressed in the lung. Of these genes, only 6 contained known eQTLs. After LD pruning, 86 unique SNPs were explored for potential association with PH in HF patients. In this genetic association analysis, we did not identify any SNPs associated with Group 2 PH. However, the significant genes in our differential gene expression analysis were associated with CpcPH rather than Group 2 PH in general. Further, HF patients in this analysis included those with HFrEF as well, unlike our RNA-seq analysis that included only HFpEF patients. Our non-significant findings may be explained by the fact that we were underpowered

to detect PH association with our small sample size and/or that our phenotype in this analysis was not as clean as the phenotype in our RNA-seq analysis.

Lastly, we compared a list of differentially expressed genes from Group 2 PH genetic studies (as no published transcriptomic studies were found) and select Group 1 PH transcriptomic studies to our data set to search for concordant genes, assessing direction and effect size (Table XIV). Of the three Group 2 PH studies,[31-33] only one gene was found in common, *ALG11* (Alpha-1,2-Mannosyltransferase). This gene is involved in metabolism and is found on the endoplasmic reticulum.[31] For the comparison with Group 1 PH transcriptomic studies, there were some curious findings. Two of our candidate genes, *SMAD6* and *GADD45B*, were significant in PAH patients in these studies.[117,156] While about half of our concordant genes (*SMAD6,*[156] *IL16,*[154] *IL7R,*[154] and PDE4D[154]) had the same direction of effect (i.e., upregulated) in our CpcPH patients as in patients from studies in other PH subtypes, there were several genes that had opposite directions of effect in our data set compared to these other studies. *PTX3, GADD45B,* and *IL7R* were upregulated in our CpcPH patients, whereas they were downregulated in various Group 1 PH subtypes.[117,204,205] PDE4 is the primary selective cyclic adenosine monophosphate (cAMP) enzyme in immune and inflammatory cells, which is involved in the pathobiologic mechanisms of certain inflammatory lung diseases, like chronic obstructive pulmonary disease and asthma.[206] These findings further support our assertion that there is pathobiological overlap between Group 2 PH and other PH subtypes, particularly Group 1 PH, but they also suggest that these signaling molecules may have different magnitudes or directions of effect depending on the underlying cause of PH. One might theorize that this may be part of the reason why some therapies used to treat one subtype of PH (e.g. endothelin receptor antagonists in PAH) may cause harm to another (e.g. Group 2 PH).

Previous PAH and heritable pulmonary veno-occlusive disease genes (*BMPR2*, *ALK1*, *ENG*, *BMPR1B*, *CAV1*, *EIF2AK4*) were not significantly associated with either PH subtype in our analysis. This may suggest that there are other more important genes with a role in CpcPH

TABLE XIV

CONCORDANT DIFFERENTIALLY EXPRESSED GENES BETWEEN OUR STUDY AND SEVERAL PUBLISHED GENOMIC GROUP 2 PH STUDIES AND TRANSCRIPTOMIC PAH STUDIES

Lead Author [reference]	Type of PH	Study Description	Concordant DEGs between our studies	Fold change comparison between their study and ours
Duarte JD[32]	Group 2 PH	Genotyped 118 HF-PH patient samples for *NOS2* and *NOS3* polymorphisms.	-	-
Kelly NJ[33]	Group 2 PH	Performed a GWAS to find candidate genes associated with development of high-fat diet-induced PH in mice.	-	-
Assad TR[31]	Group 2 PH, PAH	Conducted a GWAS to identify differentially expressed SNPs between 36 CpcPH and 79 PAH patients versus 139 IpcPH control patients. GTEx portal was used to perform eQTL mapping to identify genes associated with shared risk for CpcPH and PAH compared to IpcPH.	*ALG11*	Fold change and FDR not reported in Assad *et. al.*. 1.24 in CpcPH vs. HFpEF without PH (FDR=0.039)
Hoffman J[159,204,205]	PAH	Review of 25 human microarray studies that analyzed expression profiles of PH patients in various tissues.	*IL7R*[a,b]	-1.95 in SSc-APAH vs. SSc without PAH (*P*=3.94E-6) 0.33 in CpcPH vs. HFpEF without PH (FDR=0.061)
			PTX3[a,b]	-2.21 in IPAH vs. no IPAH (FDR=0.020) 2.23 in CpcPH vs. HFpEF without PH (FDR=0.070)
Sasagawa S[156]	PAH	Performed comparative transcriptome analysis of 5 mammalian PAH data sets within a publically available database.	*SMAD6*	0.51 in Human PAH vs. controls (FDR=0.00) 2.99 in CpcPH vs. HFpEF without PH (FDR=0.049)
Rhodes CJ[117]	IPAH	Used RNA-seq to analyze endothelial transcriptomes from the lungs of 7 control patients and 6 patients with IPAH.	*GADD45B*[b]	-0.80 in IPAH vs. no IPAH (*P*=0.05) 1.66 in CpcPH vs. HFpEF without PH (FDR=0.027)
Rajkumar R[154]	PAH	Created the largest data set to date of RNA expression profiles from lung tissue samples from 18 patients with PAH, 8 patients with PH secondary to IPF, and 13 control patients, using genome-wide microarray analysis.	*IL16*	0.27 in PAH vs no PAH (FDR<0.01) 0.69 in CpcPH vs. HFpEF without PH (FDR=0.031)
			IL7R[a]	2.83 in PAH vs no PAH (FDR<0.01) 0.33 in CpcPH vs. HFpEF without PH (FDR=0.061)
			PDE4D[a]	3.48 in PAH vs no PAH (FDR<0.01) 1.87 in CpcPH vs. HFpEF without PH (FDR=0.061)

HPAH- heritable PAH; IPAH- idiopathic PAH; IPF- idiopathic pulmonary fibrosis; SSc-APAH- systemic sclerosis-associated PAH.

[a] FDR was trending in our differential gene expression analysis.

[b] Genes with a fold change in the opposite direction than found in our differential gene expression analysis.

pathogenesis than in PAH development.

Our study has several strengths worthy of mention. Importantly, we utilized RNA-seq analysis to study Group 2 PH, which to our knowledge, has never been done before. Studies thus far investigating this PH subtype have been primarily genetic or genomic in nature,[31-33] with one study performing an *in silico* transcriptomic analysis due to lack of gene expression data to accompany their GWAS data. This current landscape in Group 2 PH research is striking compared to the plethora of transcriptomic studies in PAH, for which there are about 30 or more published at this time.[159,207,208] As previously described in the *Introduction*, RNA-seq allows us to study gene expression at a single base resolution with a high dynamic range and low background noise, as compared to microarray, which has a resolution of 100 or more bases with about half the dynamic range and high background noise (Table II). Microarray analysis represents the majority of existing transcriptomic studies in PAH.[159] Furthermore, RNA-seq enables us to study alternative splicing and detect novel transcripts or genes. Although we did not explore these data in this particular study, this available data presents a wealth of opportunity to further investigate in the future, far beyond what can be studied with genotyping or whole genome sequencing data. Another strength of our study is that we had well-phenotyped patients, since we had extensive clinical history, medication data, and echocardiogram information on all patients, with right heart catheterization data for our PH patients. This detailed phenotype information, in addition to the high quality, deeply sequenced RNA data, likely contributed to our ability to detect differentially expressed genes between CpcPH and HFpEF without PH patients, despite the relatively small sample size. Having the access to echocardiographic and hemodynamic data is crucial to distinguish whether a patient has HFpEF and whether they have developed PH and if so, to what severity.

In addition to our small sample size, our study had some additional limitations. In our genetic association analysis that was informed by the *in silico* eQTL analysis, only 15 of our 86 candidate SNPs were found in our genotype array data. Since only a small percentage of SNPs

were present in our data set, a potential future direction would be to use imputed genotype data, which would allow us to investigate many more of these SNPs for association with PH. Additionally, utilizing RNA isolated from PBMCs for RNA-seq data analysis may have restricted our ability to detect gene expression that may be tissue- or cell-specific to the pulmonary vasculature. Although pulmonary artery endothelial or smooth muscle cells would logically appear to be a better tissue source and cell type for gene expression studies in PH patients, due to the challenging nature of obtaining pulmonary arteries from patients, we used RNA isolated from PBMCs instead. Investigators have had previous successes in finding gene expression signatures from PBMCs,[209,210] thus we felt confident in this approach. Presumably we care about gene expression in the lung vasculature, but PBMCs afford us the opportunity to investigate gene expression from a pathobiologic perspective in the periphery where there are various biomarkers and signaling pathways. Furthermore, PBMC expression data can be used more easily as a clinical biomarker.

Another limitation is that our samples represent a single snapshot in time, as they were collected from patients at different time points after their PH development or diagnosis. If a patient had a right heart catheterization 1 year prior to sample collection that showed they had IpcPH, it is possible that they may no longer have the same pulmonary abnormalities they had when they received the diagnosis. In this scenario, their current gene expression signature would not match their phenotype from the year before. In line with this concept, we also could not definitively say that all HFpEF “without PH” patients did not have some early stages of PH. Right heart catheterization is not available on all HFpEF patients, and often it is a procedure that is performed only once a patient is having issues (e.g. shortness of breath, fatigue, edema). However, to reduce the impact posed by these last two limitations, we did have echocardiogram and other clinical data on all patients, including TRV on most patients. Setting our TRV threshold at less than or equal to 3 m/s helped to ensure that our HFpEF “control” patients were less likely to have PH. Since we found numerous strong signals between CpcPH and HFpEF

patients, lack of right heart catheterization data on HFpEF patients does not appear to have been a huge concern. One limitation inherent to this patient population is that HFpEF patients have been shown to have several clinical phenotypes, which may have impacted the heterogeneity of our sample and hindered our ability to detect true gene signals from the PH subtype rather than HFpEF phenotype that may be more likely to develop PH.[160,211]

With this in mind, we reviewed some of the literature on common biomarkers in HFpEF to look for any overlapping signals with our data set. This search largely turned up empty, but we did find one significant and another trending gene in common with HFpEF: *IL6R* (logFC= -0.43, FDR= 0.047),[212] *PTX3* (logFC= 1.16, FDR= 0.07).[213] Interleukin-6 is part of the cytokine family and it interacts with gp 130 and its receptor, IL-6R, in order to become activated. Once this complex is activated, IL6 can induce cardiomyocyte hypertrophy, although the soluble IL-6R is required for it to interact within the myocardium.[212] Interestingly, in a study that explored interactions between biomarkers in HFrEF and HFpEF, there was a significant interaction between interleukin-6 and PTX-3 that was only observed in HFpEF.[213] While HFpEF revealed interactions primarily associated with inflammation, HFrEF revealed NT-proBNP mediated interactions that are more related to cardiac stretch. These authors noted that biomarker studies like theirs have difficulty distinguishing whether elucidated biomarker differences are the cause or result of HF.[213] This is an astute remark and one that certainly applies to our RNA-seq study as well.

We cannot be certain that our differentially expressed genes in CpcPH patients are the cause or consequence of PH in HFpEF. For this reason, it is essential to validate the differentially expressed genes of interest via *in vitro* or *in vivo* molecular studies or animal models. Validation is currently underway, as real time-quantitative PCR is ongoing, as is work with mouse models, as these models provide a way to correlate differential gene expression in tissues of interest. This is an exciting time for Group 2 PH research as just last year a study by Meng *et. al.* was published on the development of the first Group 2 PH mouse model.[214] AKR/J-

strain mice fed a high fat diet consistently displayed a higher right ventricular systolic pressure, mPAP, left ventricular end diastolic pressure, preserved left ventricular ejection fraction, and left and right ventricular hypertrophy, compared to regular diet-fed mice. Remarkably, elevated PVR and pulmonary vascular proliferative modeling were observed as well.[214] The creation of this genetically modifiable mouse model holds the potential to further explore our RNA-seq findings, mechanisms of HFpEF-PH development, and ultimately possible treatments.

In addition to real time-quantitative PCR validation and preparations for mouse models, there are a number of further analyses that can be performed with our RNA-seq data. These include alternative splicing, novel transcript and isoform detection, and allele specific expression. Lastly, there is an opportunity to try HFpEF phenomapping in our patients,[160] which may help us improve the clustering of our samples and reduce inherent noise in our RNA-seq data due to HFpEF heterogeneity.

As discussed, Group 2 PH is understudied compared to Group 1 PH. Furthermore, clearly there are limitations to the current clinical classification of PH, which has posed challenges for identifying innovative therapeutic strategies.[215] Before focusing on treatment approaches, experts have argued that it may be wise to reclassify PH based on molecular phenotype.[23] Gurtu and Michelakis advocate that a diagnostic, therapeutic, and research-focused approach to PH should mirror the “precision medicine” model of cancer research and treatment.[216] We wholeheartedly agrees with this, which is reflected in our overarching goal of this current study, as we aimed to identify transcriptomic predictors of CpcPH and IpcPH development in HFpEF patients. In our effort, another limitation is that patients can sometimes have more than one subtype of PH, thus it is possible that some of these top associated genes with CpcPH development are confounded by other unknown or unconfirmed concomitant PH subtypes that these patients may have had in addition to Group 2 PH (e.g. Group 3 or 4 PH). Unfortunately, we do not currently have access to pulmonary function tests and/or tests to rule out pulmonary embolism on our PH patients, but a future direction would be to collect this

information and adjust for these other PH subtypes, if possible, in our differential gene expression analyses. This further illustrates the dire need to expand our clinical classification of PH.

In an effort to improve the current knowledge base of the clinical classification of PH using “omics” methodologies, the National Institutes of Health/ National Heart, Lung, and Blood Institute in conjunction with the Pulmonary Hypertension Association has funded a multi-center study called PVDOMICS to implement deep clinical and molecular phenotyping on 1500 patients with PH, as well as healthy and disease comparators.[217,218] Its goal is to enhance the classification of PH and foster new understanding of the molecular pathogenesis of PH. This trial stands to provide a wealth of information on PH etiology to inform therapeutics, however, in the meantime, we have a number of leads to pursue from this RNA-seq study.

CITED LITERATURE

1. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. *Circulation Heart failure.* 2013;6(3):606-619.

2. Jessup M, Brozena S. Heart failure. *The New England journal of medicine.* 2003;348(20):2007-2018.

3. Lam CS, Donal E, Kraigher-Krainer E, Vasan RS. Epidemiology and clinical course of heart failure with preserved ejection fraction. *European journal of heart failure.* 2011;13(1):18-28.

4. Bhuiyan T, Maurer MS. Heart Failure with Preserved Ejection Fraction: Persistent Diagnosis, Therapeutic Enigma. *Current cardiovascular risk reports.* 2011;5(5):440-449.

5. Oktay AA, Shah SJ. Diagnosis and management of heart failure with preserved ejection fraction: 10 key lessons. *Current cardiology reviews.* 2015;11(1):42-52.

6. Bursi F, McNallan SM, Redfield MM, et al. Pulmonary pressures and death in heart failure: a community study. *Journal of the American College of Cardiology.* 2012;59(3):222-231.

7. Ghio S, Gavazzi A, Campana C, et al. Independent and additive prognostic value of right ventricular systolic function and pulmonary artery pressure in patients with chronic heart failure. *Journal of the American College of Cardiology.* 2001;37(1):183-188.

8. Grigioni F, Potena L, Galie N, et al. Prognostic implications of serial assessments of pulmonary hypertension in severe chronic heart failure. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation.* 2006;25(10):1241-1246.

9. Hoeper MM, Barbera JA, Channick RN, et al. Diagnosis, assessment, and treatment of non-pulmonary arterial hypertension pulmonary hypertension. *Journal of the American College of Cardiology.* 2009;54(1 Suppl):S85-96.

10. Lam CS, Roger VL, Rodeheffer RJ, Borlaug BA, Enders FT, Redfield MM. Pulmonary hypertension in heart failure with preserved ejection fraction: a community-based study. *Journal of the American College of Cardiology.* 2009;53(13):1119-1126.

11. Leung CC, Moondra V, Catherwood E, Andrus BW. Prevalence and risk factors of pulmonary hypertension in patients with elevated pulmonary venous pressure and preserved ejection fraction. *The American journal of cardiology.* 2010;106(2):284-286.

12. Redfield MM, Borlaug BA, Lewis GD, et al. PhosphdiesteRasE-5 Inhibition to Improve CLinical Status and EXercise Capacity in Diastolic Heart Failure (RELAX) trial: rationale and design. *Circulation Heart failure.* 2012;5(5):653-659.

13. Salamon JN, Kelesidis I, Msaouel P, et al. Outcomes in World Health Organization group II pulmonary hypertension: mortality and readmission trends with systolic and

preserved ejection fraction-induced pulmonary hypertension. *Journal of cardiac failure.* 2014;20(7):467-475.

14. Vachiery JL, Adir Y, Barbera JA, et al. Pulmonary hypertension due to left heart diseases. *Journal of the American College of Cardiology.* 2013;62(25 Suppl):D100-108.

15. Galie N, Humbert M, Vachiery JL, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Heart J.* 2016;37(1):67-119.

16. McLaughlin VV, Archer SL, Badesch DB, et al. ACCF/AHA 2009 expert consensus document on pulmonary hypertension: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents and the American Heart Association: developed in collaboration with the American College of Chest Physicians, American Thoracic Society, Inc., and the Pulmonary Hypertension Association. *Circulation.* 2009;119(16):2250-2294.

17. Guazzi M, Borlaug BA. Pulmonary hypertension due to left heart disease. *Circulation.* 2012;126(8):975-990.

18. Gerges M, Gerges C, Pistritto AM, et al. Pulmonary Hypertension in Heart Failure. Epidemiology, Right Ventricular Function, and Survival. *American journal of respiratory and critical care medicine.* 2015;192(10):1234-1246.

19. Guazzi M, Naeije R. Pulmonary Hypertension in Heart Failure: Pathophysiology, Pathobiology, and Emerging Clinical Perspectives. *Journal of the American College of Cardiology.* 2017;69(13):1718-1734.

20. Rezaee ME, Nichols EL, Sidhu M, Brown JR. Combined Post- and Precapillary Pulmonary Hypertension in Patients With Heart Failure. *Clinical cardiology.* 2016;39(11):658-664.

21. Guazzi M. Pulmonary hypertension in heart failure preserved ejection fraction: prevalence, pathophysiology, and clinical perspectives. *Circulation Heart failure.* 2014;7(2):367-377.

22. Guazzi M, Arena R. Pulmonary hypertension with left-sided heart disease. *Nature reviews Cardiology.* 2010;7(11):648-659.

23. Marshall JD, Bazan I, Zhang Y, Fares WH, Lee PJ. Mitochondrial dysfunction and pulmonary hypertension: cause, effect, or both. *American journal of physiology Lung cellular and molecular physiology.* 2018;314(5):L782-l796.

24. Califf RM, Adams KF, McKenna WJ, et al. A randomized controlled trial of epoprostenol therapy for severe congestive heart failure: The Flolan International Randomized Survival Trial (FIRST). *American heart journal.* 1997;134(1):44-54.

25. Packer M, McMurray J, Massie BM, et al. Clinical effects of endothelin receptor antagonism with bosentan in patients with severe chronic heart failure: results of a pilot study. *Journal of cardiac failure.* 2005;11(1):12-20.

26. Dixon DD, Trivedi A, Shah SJ. Combined post- and pre-capillary pulmonary hypertension in heart failure with preserved ejection fraction. *Heart failure reviews.* 2016;21(3):285-297.

27. Guazzi M, Vicenzi M, Arena R, Guazzi MD. Pulmonary hypertension in heart failure with preserved ejection fraction: a target of phosphodiesterase-5 inhibition in a 1-year study. *Circulation.* 2011;124(2):164-174.

28. Redfield MM, Chen HH, Borlaug BA, et al. Effect of phosphodiesterase-5 inhibition on exercise capacity and clinical status in heart failure with preserved ejection fraction: a randomized clinical trial. *Jama.* 2013;309(12):1268-1277.

29. Delgado JF, Conde E, Sanchez V, et al. Pulmonary vascular remodeling in pulmonary hypertension due to chronic heart failure. *European journal of heart failure.* 2005;7(6):1011-1016.

30. Moraes DL, Colucci WS, Givertz MM. Secondary pulmonary hypertension in chronic heart failure: the role of the endothelium in pathophysiology and management. *Circulation.* 2000;102(14):1718-1723.

31. Assad TR, Hemnes AR, Larkin EK, et al. Clinical and Biological Insights Into Combined Post- and Pre-Capillary Pulmonary Hypertension. *Journal of the American College of Cardiology.* 2016;68(23):2525-2536.

32. Duarte JD, Kansal M, Desai AA, et al. Endothelial nitric oxide synthase genotype is associated with pulmonary hypertension severity in left heart failure patients. *Pulmonary circulation.* 2018;8(2):2045894018773049.

33. Kelly NJ, Radder JE, Baust JJ, et al. Mouse Genome-Wide Association Study of Preclinical Group II Pulmonary Hypertension Identifies Epidermal Growth Factor Receptor. *American journal of respiratory cell and molecular biology.* 2017;56(4):488-496.

34. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57-63.

35. Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis. *Genome biology.* 2016;17:13.

36. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods.* 2008;5(7):621-628.

37. Shendure J. The beginning of the end for microarrays? *Nature methods.* 2008;5(7):585-587.

38. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology.* 2010;28(5):511-515.

39. Yao B, Christian KM, He C, Jin P, Ming GL, Song H. Epigenetic mechanisms in neurogenesis. *Nature reviews Neuroscience.* 2016;17(9):537-549.

40. Knoll M, Lodish HF, Sun L. Long non-coding RNAs as regulators of the endocrine system. *Nature reviews Endocrinology.* 2015;11(3):151-160.

41. Alvarez-Dominguez JR, Bai Z, Xu D, et al. De Novo Reconstruction of Adipose Tissue Transcriptomes Reveals Long Non-coding RNA Regulators of Brown Adipocyte Development. *Cell Metab.* 2015;21(5):764-776.

42. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature.* 2010;464(7291):1071-1076.

43. Arnes L, Akerman I, Balderes DA, Ferrer J, Sussel L. betalinc1 encodes a long noncoding RNA that regulates islet beta-cell formation and function. *Genes & development.* 2016;30(5):502-507.

44. Moran I, Akerman I, van de Bunt M, et al. Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab.* 2012;16(4):435-448.

45. Ishii N, Ozaki K, Sato H, et al. Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *Journal of human genetics.* 2006;51(12):1087-1099.

46. Chen Y, Li C, Tan C, Liu X. Circular RNAs: a new frontier in the study of human diseases. *Journal of medical genetics.* 2016;53(6):359-365.

47. Li LJ, Huang Q, Pan HF, Ye DQ. Circular RNAs and systemic lupus erythematosus. *Experimental cell research.* 2016;346(2):248-254.

48. Feinberg MW, Moore KJ. MicroRNA Regulation of Atherosclerosis. *Circulation research.* 2016;118(4):703-720.

49. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome biology.* 2017;18(1):83.

50. TECHNOLOGIES SC. SEPMATE™ PROCEDURE. VERSION 2.1.0. 2013; http://www.stemcell.com/~/media/Technical%20Resources/6/D/4/9/D/29251PIS.pdf?la=en. Accessed March 27, 2015.

51. Andrews S. *FastQC A Quality Control tool for High Throughput Sequence Data.* 2014.

52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods.* 2012;9:357.

53. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols.* 2012;7(3):562-578.

54. Lassmann T, Hayashizaki Y, Daub CO. SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics.* 2011;27(1):130-131.

55. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.

56. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics.* 2013;14(2):178-192.

57. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166-169.

58. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

59. Tarazona S, Furio-Tari P, Turra D, et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic acids research.* 2015;43(21):e140.

60. Risso D. *EDASeq: Exploratory data analysis and normalization for RNA-Seq.* 2013.

61. Risso D, Schwartz K, Sherlock G, Dudoit S. GC-content normalization for RNA-Seq data. *BMC bioinformatics.* 2011;12:480.

62. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology.* 2010;11(3):R25.

63. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-140.

64. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols.* 2009;4(8):1184-1191.

65. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics.* 2009;10:161.

66. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS computational biology.* 2012;8(2):e1002375.

67. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research.* 2000;28(1):27-30.

68. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research.* 2010;38(Web Server issue):W214-220.

69. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics.* 2007;81(3):559-575.

70. Fleige S, Pfaffl MW. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular aspects of medicine.* 2006;27(2-3):126-139.

71. Jahn CE, Charkowski AO, Willis DK. Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *Journal of microbiological methods.* 2008;75(2):318-324.

72. Lowery JW, Frump AL, Anderson L, DiCarlo GE, Jones MT, de Caestecker MP. ID family protein expression and regulation in hypoxic pulmonary hypertension. *American journal of physiology Regulatory, integrative and comparative physiology.* 2010;299(6):R1463-1477.

73. Frump AL, Lowery JW, Hamid R, Austin ED, de Caestecker M. Abnormal trafficking of endogenously expressed BMPR2 mutant allelic products in patients with heritable pulmonary arterial hypertension. *PloS one.* 2013;8(11):e80319.

74. Star GP, Giovinazzo M, Langleben D. ALK2 and BMPR2 knockdown and endothelin-1 production by pulmonary microvascular endothelial cells. *Microvascular research.* 2013;85:46-53.

75. Upton PD, Davies RJ, Tajsic T, Morrell NW. Transforming growth factor-beta(1) represses bone morphogenetic protein-mediated Smad signaling in pulmonary artery smooth muscle cells via Smad3. *American journal of respiratory cell and molecular biology.* 2013;49(6):1135-1145.

76. Yang J, Li X, Li Y, et al. Id proteins are critical downstream effectors of BMP signaling in human pulmonary arterial smooth muscle cells. *American journal of physiology Lung cellular and molecular physiology.* 2013;305(4):L312-321.

77. Maruyama H, Dewachter C, Belhaj A, et al. Endothelin-Bone morphogenetic protein type 2 receptor interaction induces pulmonary artery smooth muscle cell hyperplasia in pulmonary arterial hypertension. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation.* 2015;34(3):468-478.

78. Hurst LA, Dunmore BJ, Long L, et al. TNFα drives pulmonary arterial hypertension by suppressing the BMP type-II receptor and altering NOTCH signalling. *Nature communications.* 2017;8:14079.

79. Langenfeld E, Deen M, Zachariah E, Langenfeld J. Small molecule antagonist of the bone morphogenetic protein type I receptors suppresses growth and expression of Id1 and Id3 in lung cancer cells expressing Oct4 or nestin. *Molecular cancer.* 2013;12(1):129.

80. Chen D, Forootan SS, Gosney JR, Forootan FS, Ke Y. Increased expression of Id1 and Id3 promotes tumorigenicity by enhancing angiogenesis and suppressing apoptosis in small cell lung cancer. *Genes & cancer.* 2014;5(5-6):212-225.

81. Zhao Y, Malinauskas T, Harlos K, Jones EY. Structural insights into the inhibition of Wnt signaling by cancer antigen 5T4/Wnt-activated inhibitory factor 1. *Structure (London, England : 1993).* 2014;22(4):612-620.

82. Mauad T, Pozzan G, Lancas T, et al. Immunopathological aspects of schistosomiasis-associated pulmonary arterial hypertension. *The Journal of infection.* 2014;68(1):90-98.

83. Tsoumakidou M, Koutsopoulos AV, Tzanakis N, et al. Decreased small airway and alveolar CD83+ dendritic cells in COPD. *Chest.* 2009;136(3):726-733.

84. Zanini A, Spanevello A, Baraldo S, et al. Decreased maturation of dendritic cells in the central airways of COPD patients is associated with VEGF, TGF-beta and vascularity. *Respiration; international review of thoracic diseases.* 2014;87(3):234-242.

85. Elangovan IM, Vaz M, Tamatam CR, Potteti HR, Reddy NM, Reddy SP. FOSL1 Promotes Kras-induced Lung Cancer through Amphiregulin and Cell Survival Gene Regulation. *American journal of respiratory cell and molecular biology.* 2018;58(5):625-635.

86. Ding L, Liu T, Wu Z, et al. Bone Marrow CD11c+ Cell-Derived Amphiregulin Promotes Pulmonary Fibrosis. *Journal of immunology (Baltimore, Md : 1950).* 2016;197(1):303-312.

87. Hamzaoui K, Berraies A, Kaabachi W, Ammar J, Hamzaoui A. Pulmonary manifestations in Behcet disease: impaired natural killer cells activity. *Multidisciplinary respiratory medicine.* 2013;8(1):29.

88. Gabehart K, Correll KA, Yang J, et al. Transcriptome profiling of the newborn mouse lung response to acute ozone exposure. *Toxicological sciences : an official journal of the Society of Toxicology.* 2014;138(1):175-190.

89. Bentz S, Pesch T, Wolfram L, et al. Lack of transketolase-like (TKTL) 1 aggravates murine experimental colitis. *American journal of physiology Gastrointestinal and liver physiology.* 2011;300(4):G598-607.

90. Miao Y, Cui L, Chen Z, Zhang L. Gene expression profiling of DMU-212-induced apoptosis and anti-angiogenesis in vascular endothelial cells. *Pharmaceutical biology.* 2016;54(4):660-666.

91. Wang Y, Huang X, Leng D, et al. DNA methylation signatures of pulmonary arterial smooth muscle cells in chronic thromboembolic pulmonary hypertension. *Physiological genomics.* 2018;50(5):313-322.

92. Hayashi M, Tokuchi Y, Hashimoto T, et al. Reduced HIC-1 gene expression in non-small cell lung cancer and its clinical significance. *Anticancer research.* 2001;21(1b):535-540.

93. Igartua C, Myers RA, Mathias RA, et al. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. *Nature communications.* 2015;6:5965.

94. Cheng E, Whitsett TG, Tran NL, Winkles JA. The TWEAK Receptor Fn14 Is an Src-Inducible Protein and a Positive Regulator of Src-Driven Cell Invasion. *Molecular cancer research : MCR.* 2015;13(3):575-583.

95. Jandova J, Mason CJ, Pawar SC, Watts GS. Fn14 receptor promotes invasive potential and metastatic capacity of non-small lung adenocarcinoma cells through the up-regulation of integrin alpha6. *Neoplasma.* 2015;62(1):41-52.

96. Whitsett TG, Cheng E, Inge L, et al. Elevated expression of Fn14 in non-small cell lung cancer correlates with activated EGFR and promotes tumor cell migration and invasion. *The American journal of pathology.* 2012;181(1):111-120.

97. Kwapiszewska G, Gungl A, Wilhelm J, et al. Transcriptome profiling reveals the complexity of pirfenidone effects in IPF. *The European respiratory journal.* 2018.

98. Tsuyusaki J, Kuroda F, Kasuya Y, et al. Cigarette smoke-induced pulmonary inflammation is attenuated in CD69-deficient mice. *Journal of receptor and signal transduction research.* 2011;31(6):434-439.

99. Yamauchi K, Kasuya Y, Kuroda F, et al. Attenuation of lung inflammation and fibrosis in CD69-deficient mice after intratracheal bleomycin. *Respiratory research.* 2011;12:131.

100. Fujita T, Yoshioka K, Umezawa H, et al. Role of CD69 in the pathogenesis of elastase-induced pulmonary inflammation and emphysema. *Biochemistry and biophysics reports.* 2016;7:400-407.

101. Park D-W, Song J-M, Han K-H, et al. *Different Gene Expression Patterns in the Lungs of Patients with Secondary Pulmonary Hypertension.* Vol 382008.

102. Dieude P, Guedj M, Wipff J, et al. Association of the TNFAIP3 rs5029939 variant with systemic sclerosis in the European Caucasian population. *Annals of the rheumatic diseases.* 2010;69(11):1958-1964.

103. Vroman H, Das T, Bergen IM, et al. House dust mite-driven neutrophilic airway inflammation in mice with TNFAIP3-deficient myeloid cells is IL-17-independent. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology.* 2018.

104. Yu WW, Lu Z, Zhang H, et al. Anti-inflammatory and protective properties of daphnetin in endotoxin-induced lung injury. *Journal of agricultural and food chemistry.* 2014;62(51):12315-12325.

105. Zhou T, Zhang W, Smith MA, Rhaghavachari N, Nathan SD, Sweiss N, Garcia JGN, Machado RF. Expression Profiling Elucidates A Molecular Gene Signature Which Identifies Pulmonary Hypertension In Sarcoidosis. Am J Respir Crit Care Med 185;2012:A2620.

106. Shi N, Guo X, Chen SY. Olfactomedin 2, a novel regulator for transforming growth factor-beta-induced smooth muscle differentiation of human embryonic stem cell-derived mesenchymal cells. *Molecular biology of the cell.* 2014;25(25):4106-4114.

107. Vallejo A, Perurena N, Guruceaga E, et al. An integrative approach unveils FOSL1 as an oncogene vulnerability in KRAS-driven lung and pancreatic cancer. *Nature communications.* 2017;8:14294.

108. Vallejo A, Valencia K, Vicent S. All for one and FOSL1 for all: FOSL1 at the crossroads of lung and pancreatic cancer driven by mutant KRAS. *Molecular & cellular oncology.* 2017;4(3):e1314239.

109. Keshamouni VG. Excavation of FOSL1 in the Ruins of KRAS-Driven Lung Cancer. *American journal of respiratory cell and molecular biology.* 2018;58(5):551-552.

110. Morty RE, Nejman B, Kwapiszewska G, et al. Dysregulated bone morphogenetic protein signaling in monocrotaline-induced pulmonary arterial hypertension. *Arterioscler Thromb Vasc Biol.* 2007;27(5):1072-1078.

111. Jeon HS, Dracheva T, Yang SH, et al. SMAD6 contributes to patient survival in non-small cell lung cancer and its knockdown reestablishes TGF-beta homeostasis in lung cancer cells. *Cancer research.* 2008;68(23):9686-9692.

112. Afrakhte M, Moren A, Jossan S, et al. Induction of inhibitory Smad6 and Smad7 mRNA by TGF-beta family members. *Biochemical and biophysical research communications.* 1998;249(2):505-511.

113. Shen ZJ, Braun RK, Hu J, et al. Pin1 protein regulates Smad protein signaling and pulmonary fibrosis. *The Journal of biological chemistry.* 2012;287(28):23294-23305.

114. Binet R, Ythier D, Nissou D, et al. WNT16, a new biomarker of senescence involved in lung tumorigenesis. *Cancer research.* 2008;68(9 Supplement):1844.

115. Zhou Q, Chen T, Zhang W, et al. Suppression of von Hippel-Lindau Protein in Fibroblasts Protects against Bleomycin-Induced Pulmonary Fibrosis. *American journal of respiratory cell and molecular biology.* 2016;54(5):728-739.

116. Beazley KE, Nurminsky D, Lima F, Gandhi C, Nurminskaya MV. Wnt16 attenuates TGFbeta-induced chondrogenic transformation in vascular smooth muscle. *Arterioscler Thromb Vasc Biol.* 2015;35(3):573-579.

117. Rhodes CJ, Im H, Cao A, et al. RNA Sequencing Analysis Detection of a Novel Pathway of Endothelial Dysfunction in Pulmonary Arterial Hypertension. *American journal of respiratory and critical care medicine.* 2015;192(3):356-366.

118. Na YK, Lee SM, Hong HS, Kim JB, Park JY, Kim DS. Hypermethylation of growth arrest DNA-damage-inducible gene 45 in non-small cell lung cancer and its relationship with clinicopathologic features. *Molecules and cells.* 2010;30(1):89-92.

119. Hou W, Yin J, Vogel U, Sun Z, Liang D. 19p13.3-GADD45B common variants and 19q13.3-PPP1R13L and 19q13.3-CD3EAP in lung cancer risk among Chinese. *Chemico-biological interactions.* 2017;277:74-78.

120. Jin X, Liu X, Zhang Z, Guan Y, Xv R, Li J. Identification of key pathways and genes in lung carcinogenesis. *Oncology letters.* 2018;16(4):4185-4192.

121. Yoo J, Ghiassi M, Jirmanova L, et al. Transforming Growth Factor-β-induced Apoptosis Is Mediated by Smad-dependent Expression of GADD45b through p38 Activation. *Journal of Biological Chemistry.* 2003;278(44):43001-43007.

122. Barkley LR, Santocanale C. MicroRNA-29a regulates the benzo[a]pyrene dihydrodiol epoxide-induced DNA damage response through Cdc7 kinase in lung cancer cells. *Oncogenesis.* 2013;2:e57.

123. Bonte D, Lindvall C, Liu H, Dykema K, Furge K, Weinreich M. Cdc7-Dbf4 kinase overexpression in multiple cancers and tumor cell lines is correlated with p53 inactivation. *Neoplasia (New York, NY).* 2008;10(9):920-931.

124. Datta A, Ghatak D, Das S, et al. p53 gain-of-function mutations increase Cdc7-dependent replication initiation. *EMBO reports.* 2017;18(11):2030-2050.

125. Brock M, Haider TJ, Vogel J, et al. The hypoxia-induced microRNA-130a controls pulmonary smooth muscle cell proliferation by directly targeting CDKN1A. *The international journal of biochemistry & cell biology.* 2015;61:129-137.

126. Huber LC, Ulrich S, Leuenberger C, et al. Featured Article: microRNA-125a in pulmonary hypertension: Regulator of a proliferative phenotype of endothelial cells. *Experimental biology and medicine (Maywood, NJ).* 2015;240(12):1580-1589.

127. Fukazawa T, Guo M, Ishida N, et al. SOX2 suppresses CDKN1A to sustain growth of lung squamous cell carcinoma. *Scientific reports.* 2016;6:20113.

128. Yu SY, Liao CH, Chien MH, Tsai TY, Lin JK, Weng MS. Induction of p21(Waf1/Cip1) by garcinol via downregulation of p38-MAPK signaling in p53-independent H1299 lung cancer. *Journal of agricultural and food chemistry.* 2014;62(9):2085-2095.

129. Yao W, Mu W, Zeifman A, et al. Fenfluramine-induced gene dysregulation in human pulmonary artery smooth muscle and endothelial cells. *Pulmonary circulation.* 2011;1(3):405-418.

130. Allavena G, Cuomo F, Baumgartner G, et al. Suppressed translation as a mechanism of initiation of CASP8 (caspase 8)-dependent apoptosis in autophagy-deficient NSCLC cells under nutrient limitation. *Autophagy.* 2018;14(2):252-268.

131. Zhao X, Liu X, Su L. Parthenolide induces apoptosis via TNFRSF10B and PMAIP1 pathways in human lung cancer cells. *Journal of experimental & clinical cancer research : CR.* 2014;33:3.

132. Chen YC, Chen TW, Su MC, et al. Whole Genome DNA Methylation Analysis of Obstructive Sleep Apnea: IL1R2, NPR2, AR, SP140 Methylation and Clinical Phenotype. *Sleep.* 2016;39(4):743-755.

133. Quidville V, Alsafadi S, Goubar A, et al. Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. *Cancer research.* 2013;73(7):2247-2258.

134. Zhang F, Chen X, Wei K, et al. Identification of Key Transcription Factors Associated with Lung Squamous Cell Carcinoma. *Medical science monitor : international medical journal of experimental and clinical research.* 2017;23:172-206.

135. Larrasa-Alonso J, Villalba M, Sanchez-Cabo F, et al. 1981Loss of SRSF4 in cardiomyocytes induces hypertrophy, diastolic dysfunction and risk of sudden death. *European heart journal.* 2017;38(suppl_1):ehx502.1981-ehx1502.1981.

136. Li F, Fang Z, Zhang J, et al. Identification of TRA2B-DNAH5 fusion as a novel oncogenic driver in human lung squamous cell carcinoma. *Cell research.* 2016;26(10):1149-1164.

137. Faner R, Gonzalez N, Cruz T, Kalko S, Agusti A. *Systemic Inflammatory Response to Smoking in Chronic Obstructive Pulmonary Disease: Evidence of a Gender Effect.* Vol 92014.

138. Pei N, Cao L, Liu Y, et al. XAB2 tagSNPs contribute to non-small cell lung cancer susceptibility in Chinese population. *BMC cancer.* 2015;15:560.

139. Drake JI, Gomez-Arroyo J, Dumur CI, et al. Chronic carvedilol treatment partially reverses the right ventricular failure transcriptional profile in experimental pulmonary hypertension. *Physiological genomics.* 2013;45(12):449-461.

140. Kakumu T, Sato M, Goto D, et al. Identification of proteasomal catalytic subunit PSMA6 as a therapeutic target for lung cancer. *Cancer science.* 2017;108(4):732-743.

141. Li Y, Huang J, Sun J, et al. The transcription levels and prognostic values of seven proteasome alpha subunits in human cancers. *Oncotarget.* 2017;8(3):4501-4519.

142. Yao C, Yu J, Taylor L, Polgar P, McComb ME, Costello CE. Protein Expression by Human Pulmonary Artery Smooth Muscle Cells Containing a BMPR2 Mutation and the Action of ET-1 as Determined by Proteomic Mass Spectrometry. *International journal of mass spectrometry.* 2015;378:347-359.

143. Xu H, Ma J, Wu J, et al. Gene expression profiling analysis of lung adenocarcinoma. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas.* 2016;49(3).

144. Jin Y, Liu Y, Zhang J, et al. The Expression of miR-375 Is Associated with Carcinogenesis in Three Subtypes of Lung Cancer. *PloS one.* 2015;10(12):e0144187.

145. Sakornsakolpat P, Morrow JD, Castaldi PJ, et al. Integrative genomics identifies new genes associated with severe COPD and emphysema. *Respiratory research.* 2018;19(1):46.

146. Ohashi K, Sequist LV, Arcila ME, et al. Characteristics of lung cancers harboring NRAS mutations. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2013;19(9):2584-2591.

147. Youssef O, Knuuttila A, Piirila P, Bohling T, Sarhadi V, Knuutila S. Hotspot Mutations Detectable by Next-generation Sequencing in Exhaled Breath Condensates from Patients with Lung Cancer. *Anticancer research.* 2018;38(10):5627-5634.

148. Mourah S, How-Kit A, Meignin V, et al. Recurrent NRAS mutations in pulmonary Langerhans cell histiocytosis. *The European respiratory journal.* 2016;47(6):1785-1796.

149. Andrews S. FastQC Help Guide. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3%20Analysis%20Modules/4%20Per%20Base%20Sequence%20Content.html.

150. Williams CR, Baccarella A, Parrish JZ, Kim CC. Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC bioinformatics.* 2016;17:103.

151. Dozmorov MG, Adrianto I, Giles CB, et al. Detrimental effects of duplicate reads and low complexity regions on RNA- and ChIP-seq data. *BMC bioinformatics.* 2015;16 Suppl 13:S10.

152. Yoon S, Nam D. Gene dispersion is the key determinant of the read count bias in differential expression analysis of RNA-seq data. *BMC Genomics.* 2017;18(1):408.

153. Geraci MW, Moore M, Gesell T, et al. Gene expression patterns in the lungs of patients with primary pulmonary hypertension: a gene microarray analysis. *Circulation research.* 2001;88(6):555-562.

154. Rajkumar R, Konishi K, Richards TJ, et al. Genomewide RNA expression profiling in lung identifies distinct signatures in idiopathic pulmonary arterial hypertension and secondary pulmonary hypertension. *American journal of physiology Heart and circulatory physiology.* 2010;298(4):H1235-1248.

155. Yu J, Wilson J, Taylor L, Polgar P. DNA microarray and signal transduction analysis in pulmonary artery smooth muscle cells from heritable and idiopathic pulmonary arterial hypertension subjects. *Journal of cellular biochemistry.* 2015;116(3):386-397.

156. Sasagawa S, Nishimura Y, Sawada H, et al. Comparative Transcriptome Analysis Identifies CCDC80 as a Novel Gene Associated with Pulmonary Arterial Hypertension. *Frontiers in pharmacology.* 2016;7:142.

157. West JD, Austin ED, Gaskill C, et al. Identification of a common Wnt-associated genetic signature across multiple cell types in pulmonary arterial hypertension. *American journal of physiology Cell physiology.* 2014;307(5):C415-430.

158. Damico R, Kolb TM, Valera L, et al. Serum endostatin is a genetically determined predictor of survival in pulmonary arterial hypertension. *American journal of respiratory and critical care medicine.* 2015;191(2):208-218.

159. Hoffmann J, Wilhelm J, Olschewski A, Kwapiszewska G. Microarray analysis in pulmonary hypertension. *The European respiratory journal.* 2016;48(1):229-241.

160. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation.* 2015;131(3):269-279.

161. Caravita S, Faini A, Carolino D'Araujo S, et al. Clinical phenotypes and outcomes of pulmonary hypertension due to left heart disease: Role of the pre-capillary component. *PloS one.* 2018;13(6):e0199164.

162. Assad TR, Brittain EL, Wells QS, et al. Hemodynamic evidence of vascular remodeling in combined post- and precapillary pulmonary hypertension. *Pulmonary circulation.* 2016;6(3):313-321.

163. Naeije R, D'Alto M. The Diagnostic Challenge of Group 2 Pulmonary Hypertension. *Progress in cardiovascular diseases.* 2016;59(1):22-29.

164. D'Alessandro A, El Kasmi KC, Plecita-Hlavata L, et al. Hallmarks of Pulmonary Hypertension: Mesenchymal and Inflammatory Cell Metabolic Reprogramming. *Antioxidants & redox signaling.* 2018;28(3):230-250.

165. Stotz WH, Li D, Johns RA. Exogenous nitric oxide upregulates p21(waf1/cip1) in pulmonary microvascular smooth muscle cells. *Journal of vascular research.* 2004;41(3):211-219.

166. Joerger AC, Fersht AR. The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. *Annual review of biochemistry.* 2016;85:375-404.

167. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research.* 2016;44(D1):D457-462.

168. Mizuno S, Bogaard HJ, Kraskauskas D, et al. p53 Gene deficiency promotes hypoxia-induced pulmonary hypertension and vascular remodeling in mice. *American journal of physiology Lung cellular and molecular physiology.* 2011;300(5):L753-761.

169. Jacquin S, Rincheval V, Mignotte B, et al. Inactivation of p53 Is Sufficient to Induce Development of Pulmonary Hypertension in Rats. *PloS one.* 2015;10(6):e0131940.

170. Duchen MR, Szabadkai G. Roles of mitochondria in human disease. *Essays in biochemistry.* 2010;47:115-137.

171. Dohi T, Beltrami E, Wall NR, Plescia J, Altieri DC. Mitochondrial survivin inhibits apoptosis and promotes tumorigenesis. *The Journal of clinical investigation.* 2004;114(8):1117-1127.

172. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science (New York, NY).* 2009;324(5930):1029-1033.

173. Sutendra G, Dromparis P, Kinnaird A, et al. Mitochondrial activation by inhibition of PDKII suppresses HIF1a signaling and angiogenesis in cancer. *Oncogene.* 2013;32(13):1638-1650.

174. Chourasia AH, Boland ML, Macleod KF. Mitophagy and cancer. *Cancer & metabolism.* 2015;3:4.

175. Michelakis ED, Weir EK. The metabolic basis of vascular oxygen sensing: diversity, compartmentalization, and lessons from cancer. *American journal of physiology Heart and circulatory physiology.* 2008;295(3):H928-h930.

176. Guignabert C, Tu L, Le Hiress M, et al. Pathogenesis of pulmonary arterial hypertension: lessons from cancer. *European respiratory review : an official journal of the European Respiratory Society.* 2013;22(130):543-551.

177. Archer SL. Acquired Mitochondrial Abnormalities, Including Epigenetic Inhibition of Superoxide Dismutase 2, in Pulmonary Hypertension and Cancer: Therapeutic Implications. *Advances in experimental medicine and biology.* 2016;903:29-53.

178. Weir EK, Lopez-Barneo J, Buckler KJ, Archer SL. Acute oxygen-sensing mechanisms. *The New England journal of medicine.* 2005;353(19):2042-2055.

179. Leopold JA, Maron BA. Molecular Mechanisms of Pulmonary Vascular Remodeling in Pulmonary Arterial Hypertension. *International journal of molecular sciences.* 2016;17(5).

180. Archer SL, Gomberg-Maitland M, Maitland ML, Rich S, Garcia JG, Weir EK. Mitochondrial metabolism, redox signaling, and fusion: a mitochondria-ROS-HIF-1alpha-Kv1.5 O2-sensing pathway at the intersection of pulmonary hypertension and cancer. *American journal of physiology Heart and circulatory physiology.* 2008;294(2):H570-578.

181. Tuder RM, Davis LA, Graham BB. Targeting energetic metabolism: a new frontier in the pathogenesis and treatment of pulmonary hypertension. *American journal of respiratory and critical care medicine.* 2012;185(3):260-266.

182. Fijalkowska I, Xu W, Comhair SA, et al. Hypoxia inducible-factor1alpha regulates the metabolic shift of pulmonary hypertensive endothelial cells. *The American journal of pathology.* 2010;176(3):1130-1138.

183. Chou T-F, Chuang Y-T, Hsieh W-C, et al. Tumour suppressor death-associated protein kinase targets cytoplasmic HIF-1α for Th17 suppression. *Nature communications.* 2016;7:11904.

184. Huang LE, Gu J, Schau M, Bunn HF. Regulation of hypoxia-inducible factor 1alpha is mediated by an O2-dependent degradation domain via the ubiquitin-proteasome pathway. *Proceedings of the National Academy of Sciences of the United States of America.* 1998;95(14):7987-7992.

185. Pullamsetti SS, Savai R, Seeger W, Goncharova EA. Translational Advances in the Field of Pulmonary Hypertension. From Cancer Biology to New Pulmonary Arterial Hypertension Therapeutics. Targeting Cell Growth and Proliferation Signaling Hubs. *American journal of respiratory and critical care medicine.* 2017;195(4):425-437.

186. Witsch E, Sela M, Yarden Y. Roles for growth factors in cancer progression. *Physiology (Bethesda, Md).* 2010;25(2):85-101.

187. Merklinger SL, Jones PL, Martinez EC, Rabinovitch M. Epidermal growth factor receptor blockade mediates smooth muscle cell apoptosis and improves survival in rats with pulmonary hypertension. *Circulation.* 2005;112(3):423-431.

188. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic acids research.* 2013;41(Database issue):D377-386.

189. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646-674.

190. Massague J. TGFbeta in Cancer. *Cell.* 2008;134(2):215-230.

191. Hatton N, Frech T, Smith B, Sawitzke A, Scholand MB, Markewitz B. Transforming growth factor signalling: a common pathway in pulmonary arterial hypertension and systemic sclerosis. *International journal of clinical practice Supplement.* 2011(172):35-43.

192. Lin X, Liang Y-Y, Sun B, et al. Smad6 Recruits Transcription Corepressor CtBP To Repress Bone Morphogenetic Protein-Induced Transcription. *Molecular and Cellular Biology.* 2003;23(24):9081-9093.

193. Pullamsetti SS, Perros F, Chelladurai P, Yuan J, Stenmark K. Transcription factors, transcriptional coregulators, and epigenetic modulation in the control of pulmonary vascular cell phenotype: therapeutic implications for pulmonary hypertension (2015 Grover Conference series). *Pulmonary circulation.* 2016;6(4):448-464.

194. Paulin R, Michelakis ED. Addressing Complexity in Pulmonary Hypertension: The FoxO1 Case. *Circulation research.* 2015;116(11):1732-1735.

195. Vinante F, Rigo A. Heparin-binding epidermal growth factor-like growth factor/diphtheria toxin receptor in normal and neoplastic hematopoiesis. *Toxins.* 2013;5(6):1180-1201.

196. Powell PP, Klagsbrun M, Abraham JA, Jones RC. Eosinophils expressing heparin-binding EGF-like growth factor mRNA localize around lung microvessels in pulmonary hypertension. *The American journal of pathology.* 1993;143(3):784-793.

197. Dahal BK, Cornitescu T, Tretyn A, et al. Role of epidermal growth factor inhibition in experimental pulmonary hypertension. *American journal of respiratory and critical care medicine.* 2010;181(2):158-167.

198. Oeckinghaus A, Ghosh S. The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harbor perspectives in biology.* 2009;1(4):a000034.

199. Patel H, Zaghloul N, Lin K, Liu SF, Miller EJ, Ahmed M. Hypoxia-induced activation of specific members of the NF-kB family and its relevance to pulmonary vascular remodeling. *The international journal of biochemistry & cell biology.* 2017;92:141-147.

200. Dengler VL, Galbraith M, Espinosa JM. Transcriptional regulation by hypoxia inducible factors. *Critical reviews in biochemistry and molecular biology.* 2014;49(1):1-15.

201. Sawada H, Saito T, Nickel NP, et al. Reduced BMPR2 expression induces GM-CSF translation and macrophage recruitment in humans and mice to exacerbate pulmonary hypertension. *The Journal of experimental medicine.* 2014;211(2):263-280.

202. Shimodaira K, Okubo Y, Ochiai E, et al. Gene expression analysis of a murine model with pulmonary vascular remodeling compared to end-stage IPAH lungs. *Respiratory research.* 2012;13:103.

203. Wang L, Gao J, Dai W, Lu L. Activation of Polo-like kinase 3 by hypoxic stresses. *The Journal of biological chemistry.* 2008;283(38):25928-25935.

204. Mura M, Anraku M, Yun Z, et al. Gene expression profiling in the lungs of patients with pulmonary hypertension associated with pulmonary fibrosis. *Chest.* 2012;141(3):661-673.

205. Risbano MG, Meadows CA, Coldren CD, et al. Altered immune phenotype in peripheral blood cells of patients with scleroderma-associated pulmonary hypertension. *Clinical and translational science.* 2010;3(5):210-218.

206. Beghe B, Rabe KF, Fabbri LM. Phosphodiesterase-4 inhibitor therapy for lung diseases. *American journal of respiratory and critical care medicine.* 2013;188(3):271-278.

207. Graf S, Haimel M, Bleda M, et al. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nature communications.* 2018;9(1):1416.

208. Potus F, Hindmarch CCT, Dunham-Snary KJ, Stafford J, Archer SL. Transcriptomic Signature of Right Ventricular Failure in Experimental Pulmonary Arterial Hypertension: Deep Sequencing Demonstrates Mitochondrial, Fibrotic, Inflammatory and Angiogenic Abnormalities. *International journal of molecular sciences.* 2018;19(9).

209. Desai AA, Zhou T, Ahmad H, et al. A novel molecular signature for elevated tricuspid regurgitation velocity in sickle cell disease. *American journal of respiratory and critical care medicine.* 2012;186(4):359-368.

210. Singla S, Zhou T, Javaid K, et al. Expression profiling elucidates a molecular gene signature for pulmonary hypertension in sarcoidosis. *Pulmonary circulation.* 2016;6(4):465-471.

211. Shah SJ, Katz DH, Deo RC. Phenotypic spectrum of heart failure with preserved ejection fraction. *Heart failure clinics.* 2014;10(3):407-418.

212. Tanai E, Frantz S. Pathophysiology of Heart Failure. *Comprehensive Physiology.* 2015;6(1):187-214.

213. Tromp J, Khan MA, Klip IT, et al. Biomarker Profiles in Heart Failure Patients With Preserved and Reduced Ejection Fraction. *Journal of the American Heart Association.* 2017;6(4).

214. Meng Q, Lai YC, Kelly NJ, et al. Development of a Mouse Model of Metabolic Syndrome, Pulmonary Hypertension, and Heart Failure with Preserved Ejection Fraction. *American journal of respiratory cell and molecular biology.* 2017;56(4):497-505.

215. Dweik RA, Rounds S, Erzurum SC, et al. An official American Thoracic Society Statement: pulmonary hypertension phenotypes. *American journal of respiratory and critical care medicine.* 2014;189(3):345-355.

216. Gurtu V, Michelakis ED. A Paradigm Shift Is Needed in the Field of Pulmonary Arterial Hypertension for Its Entrance Into the Precision Medicine Era. *Circulation research.* 2016;119(12):1276-1279.

217. Hemnes AR, Beck GJ, Newman JH, et al. PVDOMICS: A Multi-Center Study to Improve Understanding of Pulmonary Vascular Disease Through Phenomics. *Circulation research.* 2017;121(10):1136-1139.

218. Newman JH, Rich S, Abman SH, et al. Enhancing Insights into Pulmonary Vascular Disease through a Precision Medicine Approach. A Joint NHLBI-Cardiovascular Medical Research and Education Fund Workshop Report. *American journal of respiratory and critical care medicine.* 2017;195(12):1661-1670.

VITA

EDUCATION

2014 – Present	Master of Science (degree in progress) – Clinical and Translational Science University of Illinois at Chicago, School of Public Health, Chicago, IL
2009 – 2013	Doctor of Pharmacy University of Florida, College of Pharmacy, Gainesville, FL *Graduated magna cum laude*
2006 – 2009	Bachelor of Science – Food Science and Human Nutrition University of Florida, College of Agricultural and Life Sciences, Gainesville, FL *Graduated cum laude*

POSTDOCTORAL TRAINING

2014 – 2016	*Translational Pharmacogenomics Fellowship* University of Illinois at Chicago College of Pharmacy, Chicago, IL University of Florida College of Pharmacy, Gainesville, FL Program Director: Julio Duarte, PharmD, PhD
2016	*Computational Genomics Training* Community Research Education and Engagement for Data Science (NIH BD2K) Icahn School of Medicine at Mount Sinai, New York, NY Program Directors: Patricia Kovatch, BS, Andrew Sharp, PhD, Luz Claudio, PhD
2013 – 2014	*PGY1 Pharmacy Practice Residency, ASHP Accredited* Holmes Regional Medical Center, Melbourne, FL Program Director: Carlette Norwood-Williams, PharmD Program Coordinator: Jessica Dabady, PharmD

LICENSURE AND CERTIFICATION

Licensure:	Illinois License No. 051.297966 (Issued 2014) Florida License No. PS50521 (Issued 2013)
Board Certification: (Issued 2016)	Board Certified Pharmacotherapy Specialist (BCPS) No. 3153304

PROFESSIONAL APPOINTMENTS

01/2017 – Present	*Clinical Assistant Professor* Center for Pharmacogenomics, Department of Pharmacotherapy and Translational Research, University of Florida, Gainesville, FL

PROFESSIONAL EXPERIENCE

09/2017 – Present	*Clinical Pharmacogenetics Specialist* Pharmacogenetics Consultation Service, Internal Medicine at Tower Hill Clinic, Gainesville, FL
04/2015 – 12/2015	*Clinical Pharmacist* Cardiology Clinic, Outpatient Care Center, University of Illinois Hospital and Health Sciences System, Chicago, IL
07/2014 – 12/2015	*Clinical Pharmacist* Personalized Medicine Program (PMP) University of Illinois Hospital and Health Sciences System, Chicago, IL
07/2013 – 06/2014	*Staff Pharmacist* Holmes Regional Medical Center, Melbourne, FL
05/2010 – 07/2010	*Pharmacist Intern* Walgreens Pharmacy, Lutz, FL
10/2005 – 08/2007	*Pharmacy Technician* Rx Express Pharmacy, Lutz, FL

PGY1 PHARMACY PRACTICE RESIDENCY LEARNING EXPERIENCES

Rotations	Preceptors
Anticoagulation	Ted Heierman, PharmD, CPh
Clinical Pharmacy Practice	Jessica Dabady, PharmD
Advanced Infectious Disease	Heather McCormick, PharmD, BCPS
Medication Safety	Jaclyn Jeffries, PharmD
Pharmacy Informatics	Marta Hamilton, PharmD, John Voorhies, PharmD
Research	Jessica Dabady, PharmD
Critical Care I and II	Kimberly Hunger, PharmD, BCPS
Infectious Disease	Stacey Baggett, PharmD, BCPS
Internal Medicine I and II	Stacey Baggett, PharmD, BCPS, Katherine Imhof, PharmD
Ambulatory Care	Norman Tomaka, M.S. Pharm, CPh, Ted Heierman, PharmD, CPh
Hospital Pharmacy 101	Jason Perry, PharmD

CLINICAL SERVICE DEVELOPMENT

Co-founder, Pharmacogenetics Consultation Service (PGx-CS), 01/2017 – Present
Participated in the design and implementation of this novel service. The PGx-CS is a pilot consult service that is a joint effort of the UF College of Pharmacy and UF Health Internal Medicine Clinic at Tower Hill. In this standalone outpatient clinic, as the lead Clinical Pharmacogenetics Specialist, I provide primary care providers with detailed, customized drug therapy plans for their referred patients based on pharmacogenetic test results.

Aims: (1) implement a novel pharmacogenetics consult service, (2) assess the feasibility of the service, (3) evaluate the clinical usefulness of the service, and (4) evaluate patients' attitudes, beliefs, and satisfaction with the service.

RESEARCH EXPERIENCE

08/2017 – Present
Principal Investigator
Pharmacogenetics Consultation Service: Patient Attitudes and Beliefs Surveys
UF Health Personalized Medicine Program, Gainesville, FL

01/2017 – Present
Co-investigator
UF Health Personalized Medicine Program Implementation Data-CYP2C19, TPMT, IL28B, CYP2D6
UF Health Personalized Medicine Program, Gainesville, FL
Principal Investigator: Kristin Weitzel, PharmD

01/2017 – Present
Co-investigator
Genomic Medicine Implementation: The Personalized Medicine Program: Education Protocol
UF Health Personalized Medicine Program, Gainesville, FL
Principal Investigator: Kristin Weitzel, PharmD

04/2016 – 03/2017,
08/2014 – 12/2015
Co-investigator
Improving Treatment Personalization of Pulmonary Hypertension Associated with Diastolic Heart Failure Study
Department of Pharmacotherapy and Translational Research
University of Florida, College of Pharmacy, Gainesville, FL
Investigators: Julio Duarte, PharmD, PhD, Hassan Alnuaimat, MD, Ali Ataya, MD

Department of Pharmacy Practice
University of Illinois at Chicago, College of Pharmacy, Chicago, IL
Investigators: Julio Duarte, PharmD, PhD, Roberto Machado, MD, Christopher Gans, MD
Responsibilities: Recruited and consented patients, coordinated study visits, and performed data collection.

01/2015 – 12/2015
Co-investigator
DNA Collection in Healthy Subjects
Department of Pharmacy Practice
University of Illinois at Chicago, College of Pharmacy, Chicago, IL
Responsibilities: Wrote study protocol, recruited and consented patients, trained research assistants, and performed data collection.
Investigators: Julio Duarte, PharmD, PhD, Roberto Machado, MD

08/2014 – 12/2015
Co-investigator
Genetic Determinants of Drug Response in Heart Failure Study
Department of Pharmacy Practice
University of Illinois at Chicago, College of Pharmacy, Chicago, IL
Responsibilities: Recruited and consented patients, coordinated study visits, processed blood samples, trained research assistants, performed data collection, submitted protocol amendments to the Institutional Review Board (IRB), performed Continuing Review mandated by the IRB, and performed data analysis.

Investigators: Julio Duarte, PharmD, PhD, Thomas Stamos, MD, Christopher Gans, MD

10/2012 – 11/2012 *Research Assistant*
Clarification of Optimal Anticoagulation through Genetics (COAG) Trial
Department of Pharmacotherapy and Translational Research
University of Florida, College of Pharmacy, Gainesville, FL
Responsibilities: Recruited and consented patients.
Advisors: Rhonda Cooper-DeHoff, PharmD, MS, Julie Johnson, PharmD, BCPS

05/2011 – 12/2011 *Co-investigator*
Pharmacogenomic Evaluation of Antihypertensive Responses (PEAR) Follow-up Trial
Department of Pharmacotherapy and Translational Research
University of Florida, College of Pharmacy, Gainesville, FL
Responsibilities: Genotyped two single nucleotide polymorphisms (SNPs) hypothesized to have an association with new onset Type 2 Diabetes, recruited and consented patients for the PEAR Follow-up trial, and wrote a primary research manuscript to graduate with honors: "TCF7L2 Polymorphisms and New Onset Diabetes within an INVEST Case Control: A Focus on SNPs in Those with African and Hispanic Ancestry."
Advisor: Rhonda Cooper-DeHoff, PharmD, MS

PUBLICATIONS

1. Duarte JD, Kansal M, Desai AA, Riden K, **Arwood MJ**, Yacob AA, Stamos TD, Cavallari LH, Zamanian RT, Shah SJ, Machado RF. Endothelial nitric oxide synthase genotype is associated with pulmonary hypertension severity in left heart failure patients. Pulm Circ. 2018 Apr-Jun;8(2):2045894018773049.

2. Singh S, McDonough CW, Gong Y, Alghamdi WA, **Arwood MJ**, Bargal SA, Dumeny L, Li WY, Mehanna M, Stockard B, Yang G, de Oliveira FA, Fredette NC, Shahin MH, Bailey KR, Beitelshees AL, Boerwinkle E, Chapman AB, Gums JG, Turner ST, Cooper-DeHoff RM, Johnson JA. Genome Wide Association Study Identifies the HMGCS2 Locus to be Associated With Chlorthalidone Induced Glucose Increase in Hypertensive Patients. *J Am Heart Assoc*. 2018 Mar 9;7(6).

3. Mangal N, Hamadeh IS, **Arwood MJ**, Cavallari LH, Samant TS, Klinker KP, Bulitta J, Schmidt S. Optimization of Voriconazole Therapy for the Treatment of Invasive Fungal Infections in Adults. *Clin Pharmacol Ther*. 2018 Jan 9. [Epub ahead of print]

4. **Arwood MJ**, Deng J, Drozda K, Pugach O, Nutescu EA, Schmidt S, Duarte JD, Cavallari LH. Anticoagulation endpoints with clinical implementation of warfarin pharmacogenetic dosing in a real-world setting: A proposal for a new pharmacogenetic dosing approach. *Clin Pharmacol Ther*. 2017 May;101(5):675-683.

5. **Arwood MJ***, Chumnumwat S*, Cavallari LH, Nutescu EA, Duarte JD. Implementing Pharmacogenomics at Your Institution: Establishment and Overcoming Implementation Challenges. *Clin Transl Sci.* 2016 May 23. [Epub ahead of print] *Contributed equally to this work

6. **Arwood MJ**, Cavallari LH, Duarte JD. Pharmacogenomics of hypertension and heart disease. Curr *Hypertens Rep.* 2015 Sep;17(9):586.

7. Karnes JH, Gong Y, **Arwood MJ**, Gums JG, Hall K, Limacher MC, Johnson JA, Cooper-DeHoff RM. Alteration in fasting glucose after prolonged treatment with a thiazide diuretic. *Diabetes Res Clin Pract.* 2014 Jun; 104(3):363-9.

8. Karnes JH, Gong Y, Pacanowski MA, McDonough CW, **Arwood MJ**, Langaee TY, Pepine CJ, Johnson JA, Cooper-DeHoff RM. Impact of TCF7L2 single nucleotide polymorphisms on hydrochlorothiazide- induced diabetes. Pharmacogenet Genomics. 2013 Dec; 23(12):697-705.

ABSTRACTS AND SCIENTIFIC PRESENTATIONS

National

1. Duong BQ, **Arwood MJ**, Hicks JK, Beitelshees AL, Elsey AR, Franchi F, Limdi NA, Mills AE, Owusu-Obeng A, Petry N, Tuteja S, Cavallari LH, Weitzel KW, on behalf of the IGNITE Network. Development of implementation guides to support clinical adoption of genomic medicine: Experiences of the implementing genomics in practice (IGNITE) network. Poster presentation at the American College of Medical Genetics (ACMG) 2018 Annual Meeting. Charlotte, NC. April 13, 2018.

2. **Arwood MJ**, McDonough CW, Cavallari LH, Elsey AR, Frye RF, Gong Y, Langaee T, Johnson JA, Weitzel KW. Preparing students for implementation of genomics: Comparison of teaching and learning strategies to develop clinical decision making skills in pharmacogenomics patient cases. Poster and oral presentation at the Association of Professors of Human and Medical Genetics (APHMG) 23rd Annual Workshop & Special Interest Groups Meetings. Clearwater Beach, FL. May 9, 2017.

3. **Arwood MJ**, McDonough CW, Cavallari LH, Elsey AR, Frye RF, Gong Y, Langaee T, Johnson JA, Weitzel KW. Equipping health professional students to apply pharmacogenomic data to clinical decision making in real-world scenarios: Comparison of an active engagement versus didactic teaching approach. Poster presentation at the Translational Science 2017 Annual Meeting. Washington DC. April 20 2017.

4. **Arwood MJ**, Deng J, Drozda K, Nutescu EA, Schmidt S, Duarte JD, Cavallari LH. Clinical implementation of warfarin pharmacogenetics in a real-world setting: a proposal for a new pharmacogenetic dosing approach for diverse patient populations. *Clin Pharmacol Ther.* 2017;101(S1):S5. (#PT-002). Poster presentation at the 2017 American Society for Clinical Pharmacology and Therapeutics (ASCPT) Annual Meeting. Washington DC. March 15, 2017. ***Winner of Presidential Trainee Award.**

5. Singh S, Alghamdi W, **Arwood MJ**, Bargal SA, de Oliveira F, Dumeny L, Wen-Yi L, Mehanna M, Stockard B, Yang G, Fredette N, Shahin M, Bailey KR, Beitelshees AL, Boerwinkle E, Chapman AB, Gums JG, Turner ST, Gong Y, McDonough CW, Cooper-DeHoff RM, Johnson JA. A Genome Wide Association Study Identifies Pharmacogenomic Variants Associated with Chlorthalidone Induced Glucose Change in African Americans. *Clin Pharmacol Ther.* 2017;101(S1):S9. (#PT-019). Poster presentation at the 2017 American Society for Clinical Pharmacology and Therapeutics (ASCPT) Annual Meeting. Washington DC. March 15, 2017. ***Winner of Presidential Trainee Award.**

6. Duarte JD, **Arwood MJ**, Liko I, Mansour I, Nair V, Kansal M, Stamos T, Cavallari LH, Desai AA. Beta-blocker dose stratifies both mortality risk and circulating procollagen levels in African Americans with heart failure. *Circulation*. 2016;134:A11877. Poster presentation at

the American Heart Association (AHA) Scientific Sessions. New Orleans, LA. November 15, 2016.

7. Deng J, **Arwood MJ**, Duarte JD, Cavallari LH, Schmidt S. Development of a universal pharmacogenetics-guided warfarin dosing nomogram. Poster presentation at the American Conference on Pharmacometrics (ACoP) Annual Meeting. Arlington, VA. October 24, 2016.

8. **Arwood MJ**, Heierman T, Gilbert S, Dabady J. Ventilator-associated pneumonia in trauma patients: gasping for answers about contributing risk factors and the efficacy of the ventilator bundle. Poster Presentation at the American Society of Health-System Pharmacists (ASHP) 2013 Midyear Clinical Meeting. Orlando, FL. December 2013.

Local

1. "Translational Pharmacogenomics: Research and Clinical Practice." Clinical Pharmacogenomics Faculty Candidate Seminar at the Department of Pharmacotherapy and Translational Research, University of Florida, College of Pharmacy. Gainesville, FL. June 8, 2016.

2. "Genome-wide analysis to identify novel treatment targets for pulmonary hypertension in heart failure patients." MS Thesis Project Proposal Defense at the University of Illinois at Chicago, School of Public Health. Chicago, IL. June 2015.

3. "Ventilator-associated pneumonia in trauma patients: gasping for answers about contributing risk factors and the efficacy of the ventilator bundle." PowerPoint presentation at the Florida Society of Health-System Pharmacists (FSHP) Florida Residency Conference. Gainesville, FL. May 2014.

4. "The approach to reversing anticoagulation in intracerebral hemorrhages." Continuing Education Presentation at the 3rd Annual Pharmacy Practice Update & Residency Forum. Palm Beach Atlantic University, Gregory School of Pharmacy. West Palm Beach, FL. January 2014.

5. "Conversions between Anticoagulants and the Approach to Their Reversal." Continuing Education Presentation at the Treasure Coast Society of Health-System Pharmacists' meeting. Melbourne, FL. September 2013.

Invited

1. "Integrating Clinical Pharmacogenetics Into Primary Care: Initial Experiences from the University of Florida." IGNITE Clinical Informatics Webinar Series. May 3, 2018.

2. "Integrating Pharmacogenetics Into Primary Care: Implementation of a Pharmacist-Led, Referral-Based Pharmacogenetics Consult Clinic." Continuing Education Presentation at the University of Florida Primary Care Innovations Conference. Gainesville, FL. April 28, 2018.

3. "Translational Pharmacogenomics." Presentation to the Student Industry Pharmacists Organization (IPhO). Gainesville, FL. March 26, 2018.

4. "Innovative Practice Models for Incorporating Precision Medicine Into Primary Care: Making it Work." Continuing Education Presentation at the University of Florida Precision Medicine Conference. Orlando, FL. March 9, 2018.

5. “Translational Pharmacogenomics: Opportunities, Experience, & Implementation.” Presentation to the Student Personalized Medicine Coalition. Gainesville, FL (broadcast to Orlando, FL). Apr 6, 2017.

6. “Taking a Novel Approach: Update on Innovations in Pharmacogenetics Education.” Continuing Education Presentation at the University of Florida Precision Medicine Conference. Orlando, FL. March 9, 2017. (Co-author to Kristin Weitzel, PharmD)

7. “Advancing Clinical Implementation of Pharmacogenetics: Progress, Needs, and Opportunities.” Continuing Education Presentation at the University of Florida Precision Medicine Conference. Orlando, FL. March 8, 2017. (Co-author to Kristin Weitzel, PharmD)

AWARDS AND HONORS

03/2017	2017 American Society for Clinical Pharmacology and Therapeutics (ASCPT) Presidential Trainee Award
04/2013	2013 Merck Award for Outstanding Research
03/2011	F. Peter Field Scholarship, Kappa Psi Pharmaceutical Society, Inc.
05/2008	Elected into Golden Key International Honour Society
08/2006 – 05/2010	Florida Bright Futures: Florida Academic Scholars Award
05/2006	International Baccalaureate Certificate

PROFESSIONAL AFFILIATIONS

American Society of Human Genetics (ASHG)	2017 – Present
Clinical Pharmacogenetics Implementation Consortium (CPIC)	2017 – Present
Florida Pharmacy Association (FPA)	2017
American Society for Clinical Pharmacology and Therapeutics (ASCPT)	2014 – Present
American Heart Association (AHA)	2014 – Present
Pharmacogenomics Research Network (PGRN)	2014 – Present
Florida Society of Health-System Pharmacists (FSHP)	2013 – 2014, 2017 – Present
American Society of Health-System Pharmacists (ASHP)	2012 –2013, 2017 – Present
American College of Clinical Pharmacy (ACCP)	2011 – 2012, 2015 – Present
Kappa Psi Pharmaceutical Fraternity, Inc (ΚΨ)	2009 – Present

PROFESSIONAL SERVICE

<u>American Society for Clinical Pharmacology and Therapeutics (ASCPT)</u>

04/2018 – Present	Member, 2019 ASCPT Pharmacogenomics Community Programming Committee
08/2015 – 06/2017	Member, Social Media Task Force

<u>American College of Clinical Pharmacy (ACCP)</u>

11/2017 – Present	Member, 2018 ACCP Pharmacokinetics/Pharmacodynamics/ Pharmacogenetics Practice and Research Networks (PRN) Focus Session Planning Committee

<u>Implementing Genomics in Practice (IGNITE)</u>

02/2017 – Present	Member, Common Measures Group

Kappa Psi Pharmaceutical Fraternity, Inc (ΚΨ)

11/2012 – 11/2013	Member, Southeast Province ΚΨ Political Action Committee
08/2011 – 04/2012	Chair, Gamma Sigma ΚΨ Golf Tournament
02/2011 – 02/2012	Vice Regent, Gamma Sigma ΚΨ Pharmaceutical Fraternity
01/2011 – 01/2012	Assistant Webmaster, Province IV ΚΨ Pharmaceutical Fraternity
01/2011 – 01/2012	Member, Province IV ΚΨ Publications Committee
02/2010 – 02/2011	Historian, Gamma Sigma ΚΨ Pharmaceutical Fraternity
01/2010 – 01/2011	Member, Province IV ΚΨ Risk Management Committee

National Human Genome Research Institute (NHGRI)

06/2017	Subject Matter Expert

Manuscript Review: *Cardiovascular Therapeutics* (11/2017), *Currents in Pharmacy Teaching and Learning* (08/2017), *Pharmacotherapy* (12/2015, 05/2018), *Future Medicine* (08/2015)

TEACHING EXPERIENCE

Clinical Assistant Professor, University of Florida, College of Pharmacy, Gainesville, FL
01/2017 – Present

Didactic Teaching

PHA5012: Clinical Applications of Personalized Medicine (Elective)

- Spring 2018, Course co-coordinator and lecturer for P2, P3, and P4 pharmacy students
  - Lecture: *Beyond the CPIC Guidelines: Emerging Gene-Drug Pairs and Clinical Implementation*
  - Lecture: *Clinical Laboratory Testing in Pharmacogenetics*
    - Conducted 6 Active Learning sessions
    - Wrote quiz and test questions, assignments, patient cases
- Spring 2017, Course co-coordinator and lecturer for P2 and P3 pharmacy students
  - Lecture: *Beyond the CPIC Guidelines: Emerging Gene-Drug Pairs and Clinical Implementation*
    - Conducted 4 Active Learning sessions
    - Wrote quiz and test questions, assignments, patient cases

Experiential Teaching

Research Trainee

06/2017 – 08/2017	Hanna Harper, PharmD Candidate 2020

Clinical Trainees

07/2018 – Present	Kelsey Melloy, PharmD, PGY2 Pharmacogenetics Resident
07/2017 – 06/2018	Benjamin Duong, PharmD, PGY2 Pharmacogenetics Resident
01/2016 – 06/2018	D. Max Smith, PharmD, BCPS, PGY2 Pharmacogenetics Resident and Pharmacogenetics Fellow

Postdoctoral Fellow, University of Florida, College of Pharmacy, Gainesville, FL
12/2015 – 12/2016

Didactic Teaching

PHA 5933: Clinical Applications of Personalized Medicine (Elective)

- Fall 2016, Lecturer for P2 and P3 pharmacy students
  - Lecture: *Clinical Laboratory Testing in Pharmacogenetics*

Experiential Teaching
Research Trainee
05/2016 – 08/2016 Justin Immerman, PharmD Candidate 2018

Clinical Instructor, University of Illinois at Chicago, College of Pharmacy, Chicago, IL
07/2014 – 12/2015

Experiential Teaching
Research Trainees
08/2015 – 12/2015 Candice Bundzinski, PharmD Candidate 2019
07/2014 – 12/2015 Ina Liko, PharmD Candidate 2017
05/2015 – 08/2015 Anesia Reticker, PharmD Candidate 2018
05/2015 – 08/2015 Patrick Prunty, PharmD Candidate 2018
01/2015 – 04/2015 Lorena Berrios, PharmD Candidate 2018
01/2015 – 04/2015 Paula Bielnicka, PharmD Candidate 2018
01/2015 – 04/2015 Ali Alobaidi, PharmD Candidate 2018