Integrative Multiomics to Dissect the Lung Transcriptional Landscape of Pulmonary Arterial Hypertension.

Jason Hong[1], Brenda Wong[1], Christopher J. Rhodes[2], Zeyneb Kurt[3], Tae-Hwi Schwantes-An[4], Elizabeth A. Mickler[4], Stefan Gräf[5], Mélanie Eyries[6], Katie A. Lutz[7], Michael W. Pauciulo[7], Richard C. Trembath[8], David Montani[9], Nicholas W. Morrell[5], Martin R. Wilkins[10], William C. Nichols[7], David-Alexandre Trégouët[11], Micheala A. Aldred[12], Ankit A. Desai[4], Rubin M. Tuder[13], Mark W. Geraci[14], Mansoureh Eghbali[15*], Robert S. Stearman[4*], Xia Yang[16*]

[1]Division of Pulmonary and Critical Care Medicine, University of California, Los Angeles, Los Angeles, CA, USA, [2] National Heart and Lung Institute, Imperial College London, London, UK, [3]Northumbria University, Newcastle Upon Tyne, UK, [4]Department of Medicine, Indiana University, Indianapolis, IN, USA, [5]Department of Medicine, University of Cambridge, Cambridge, UK, [6]Hôpital Pitié- Salpêtrière, AP-HP, Département de Génétique, Paris, France, [7]Cincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine, Dept of Pediatrics, Division of Human Genetics, Cincinnati, OH, USA, [8]Department of Medical & Molecular Genetics, Faculty of Life Sciences & Medicine, King's College London, London, UK, [9]AP-HP, Service de Pneumologie, Hôpital Bicêtre, 78 Rue du Général Leclerc, 94270, Le Kremlin Bicêtre, France; Université Paris-Saclay, 78 Rue du Général Leclerc, 94270 Le Kremlin Bicêtre, France; UMR_S 999, Université Paris-Saclay, INSERM, Groupe hospitalier Marie-Lannelongue -Saint Joseph, 133 avenue de la résistance, 92350 Le Plessis-Robinson, France, [10]National Heart and Lung Institute, Imperial College London, London, UK, [11]Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, F-33000

Bordeaux, France, [12]Division of Pulmonary, Critical Care, Sleep and Occupational Medicine, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA, [13]Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, University of Colorado, Aurora, CO, USA, [14]Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA, [15]Department of Anesthesiology & Perioperative Medicine, University of California, Los Angeles, Los Angeles, CA, USA, [16]Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA, USA

*Co-senior authors

Correspondence

Jason Hong, M.D., Ph.D.
Division of Pulmonary and Critical Care Medicine
David Geffen School of Medicine at UCLA
200 UCLA Medical Plaza, Suite 365-B, Box 951693
Los Angeles, CA 90095
E-mail: jasonhong@mednet.ucla.edu

Abstract

Pulmonary arterial hypertension (PAH) remains an incurable and often fatal disease despite currently available therapies. Multiomics systems biology analysis can shed new light on PAH pathobiology and inform translational research efforts. Using RNA sequencing on the largest PAH lung biobank to date (96 disease and 52 control), we aim to identify gene co-expression

network modules associated with PAH and potential therapeutic targets. Co-expression network analysis was performed to identify modules of co-expressed genes which were then assessed for and prioritized by importance in PAH, regulatory role, and therapeutic potential via integration with clinicopathologic data, human genome-wide association studies (GWAS) of PAH, lung Bayesian regulatory networks, single-cell RNA-sequencing data, and pharmacotranscriptomic profiles. We identified a co-expression module of 266 genes, called the pink module, which may be a response to the underlying disease process to counteract disease progression in PAH. This module was associated not only with PAH severity such as increased PVR and intimal thickness, but also with compensated PAH such as lower number of hospitalizations, WHO functional class and NT-proBNP. GWAS integration demonstrated the pink module is enriched for PAH-associated genetic variation in multiple cohorts. Regulatory network analysis revealed that BMPR2 regulates the main target of FDA-approved riociguat, GUCY1A2, in the pink module. Analysis of pathway enrichment and pink hub genes (i.e. ANTXR1 and SFRP4) suggests the pink module inhibits Wnt signaling and epithelial-mesenchymal transition. Cell type deconvolution showed the pink module correlates with higher vascular cell fractions (i.e. myofibroblasts). A pharmacotranscriptomic screen discovered ubiquitin-specific peptidases (USPs) as potential therapeutic targets to mimic the pink module signature. Our multiomics integrative study uncovered a novel gene subnetwork associated with clinicopathologic severity, genetic risk, specific vascular cell types, and new therapeutic targets in PAH. Future studies are warranted to investigate the role and therapeutic potential of the pink module and targeting USPs in PAH.

Introduction

Pulmonary arterial hypertension (PAH) remains an incurable and often fatal disease characterized by irreversible vascular remodeling. Despite the identification of many candidate drugs in the preclinical stage, effective therapies that reverse the underlying disease process are still lacking. A deeper understanding of the molecular and cellular mechanisms in PAH lung tissue is needed to bridge this translational gap.

Data-driven transcriptome-wide studies of PAH lungs have uncovered genes and pathways differentially expressed in PAH(1, 2). However, whether such findings are robust, causal, and cell-specific in disease pathogenesis remain unknown since lung samples are usually from limited numbers of advanced stage PAH patients at the bulk tissue level. Furthermore, typical gene-level differential expression analysis may not reveal upstream causal and regulatory genes and pathways(3, 4). A rigorous systems-level examination of altered transcriptomes in PAH lungs integrating different omics data types is needed to advance our understanding of PAH pathobiology and help inform potential causal genes, regulatory networks and pathways, and therapeutic targets to facilitate translational efforts.

In this study which leverages the transcriptional landscape of a large biorepository of PAH lungs (96 disease vs 52 control), we dissect the gene networks of PAH lungs using an integrative multiomic and systems biology approach to uncover a module of co-expressed genes associated with clinicopathologic severity, genetic risk, and vascular cell specificity in PAH, and further identify novel therapeutic targets by a pharmacotranscriptomic screen for future preclinical studies.

Methods

RNA sequencing and differential expression analysis

Patient enrollment and the standardized tissue-processing protocol for PHBI have been previously described(1, 5). Paired-end 75 base-pair RNA sequencing (RNA-seq) was performed on all available PHBI lung samples using an Illumina sequencer. Samples were sequenced in two batches. Sequencing depth was 20-25 million reads per sample in one batch and 15-20 million reads per sample in the other batch. Reads were mapped to the UCSC human reference genome (version hg19) using STAR(6). Transcripts were assembled and quantified using StringTie(7). Transcript-level abundance estimates were imported and summarized into a counts matrix using tximport(8) which was then input into DESeq2 (9) for differential gene expression analysis using a negative binomial generalized linear model. Potential outliers and batch effects of different covariates (i.e. sequencing batch, sex, age, ethnicity) were assessed by hierarchical clustering and principal component analysis. Two patients with WHO Group IV pulmonary hypertension were excluded from this analysis. Sequencing batch was adjusted for in the DESeq2 model. Sex-stratified differential expression analysis was also performed. Differentially expressed genes with FDR < 0.05 were considered statistically significant.

Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA v. 1.69 R package was used to identify modules of co-expressed genes in PHBI lung RNA-seq samples. We first performed variance stabilizing transformation of the counts matrix using DESeq2. We then adjusted for the effect of the two sequencing batches using an empirical Bayes framework through the ComBat function of the sva v. 3.34 R package. The batch-

corrected expression matrix was then evaluated for potential outlier samples by hierarchal clustering. Two samples were identified as outliers and were removed from downstream analyses: a 55 year-old female control and a 14 year-old male patient with idiopathic PAH. The bottom 25% of genes with the least variation across samples were filtered out. Known PAH genes that were filtered out at this step were added back to the expression matrix (48 out of 582 PAH genes retrieved from disease-gene databases DisGeNET (10) and Comparative Toxicogenomics Database (11) using the Harmonizome portal (12)). A total of 17,564 genes were then included in downstream WGCNA steps. A soft-thresholding power of 3 was selected to power the correlation of genes with the assumption that raising the correlation to a specific power will reduce the noise of the correlations in the adjacency matrix. A soft-thresholding power of 3 was selected to optimize both the scale-free topology index ($R^2 > 0.8$) and mean connectivity ($k = 205$). To minimize effects of noise and spurious associations, the adjacency matrix was transformed into a Topological Overlap Matrix (TOM), and the corresponding dissimilarity matrix was calculated. Hierarchical clustering was then performed on the dissimilarity matrix after which genes were split into 25 modules using the cutreeDynamic function of the dynamicTreeCut R package using the following parameters: DeepSplit = 2, pamRespectsDendro = FALSE, cutHeight = 0.99, minClusterSize = 30. Module eigengenes representing the first principal component of a given module in a given single dataset (i.e. PHBI lung dataset) were calculated using the moduleEigengenes function of WGCNA. Given that dynamicTreeCut may identify modules whose expression profiles are similar in which their genes are highly co-expressed, such modules were assessed for and merged using the following step per the WGCNA analysis pipeline: hierarchical clustering was performed on the dissimilarity of module eigengenes after which similar modules were merged using the

mergeCloseModules function of WGCNA using a cutHeight of 0.15. Merging of similar modules yielded 20 final modules which were the used for downstream analyses. The strongest pairwise gene-gene connection (ANTXR1 and SFRP4) within the pink module was identified by comparing TOM values across all 33,732 pairs of pink genes.

Pathway enrichment analysis

Gene set enrichment analysis (GSEA) using R package fgsea v1.18.0 was performed to identify enriched pathways in the PAH lung signature. The PAH signature represents the differential transcriptome between PAH and control. Genes were ordered by the Wald statistic as determined by DESeq2 of PAH vs. control. Per the GSEA algorithm, genes were not filtered (i.e. by expression, variance, or by a statistical threshold for differential expression) prior to GSEA. The PAH signature was tested for enrichment in Hallmark pathways from Molecular Signature Database (13), as well as in co-expression modules. Sex-stratified analysis was performed using the sex-stratified differential expression analysis results. GSEA was also performed on select co-expression module signatures (i.e. pink, royalblue, greenyellow). Module signatures were defined as the correlation of the module eigengene with the expression of genes across the transcriptome. Similar to the PAH signature analysis, genes were not filtered prior to GSEA. Module signatures were tested for enrichment in Biological Processes from Gene Ontology (GO) (14), Hallmark pathways from Molecular Signature Database (13), and/or known cell-type signatures from Azimuth(15). Enrichment in pathways with FDR < 0.05 were considered statistically significant.

Module-trait correlation analysis

Pearson correlations of module eigengenes with clinical and pathologic characteristics were computed in order to prioritize modules by importance in PAH. Dichotomous categorical variables were coded 0 and 1. WHO functional class was obtained from the New York Heart Association (NYHA) class recorded immediately pre-transplant (i.e. day of transplant) and was coded as integers 1, 2, 3 and 4. If the immediate pre-transplant NYHA class was unavailable for a given patient, then the WHO functional class from their most recent clinic visit was used. Number of hospitalizations due to PAH were counted between the time of diagnostic RHC and lung transplantation. Presence of right heart failure signs such as ascites or leg swelling were recorded at the time of enrollment in the study (i.e. just prior to lung transplantation). Lab values (i.e. NT-proBNP and creatinine) were obtained from the most recent blood draw prior to transplant. The most recent pulmonary function testing and right heart catheterization results prior to transplant were used. Intima and intima plus media thickness were measured by a lung pathologist on explant histological tissue sections by morphometric analysis of volume density of pulmonary arteries. REVEAL lite scores were calculated as per Benza et al(16) using values of NT-proBNP or BNP, six-minute walk distance, WHO functional class, systolic blood pressure, heart rate, and creatinine. Values were obtained from the most recent available assessment prior to transplant. A score of zero was assigned for missing individual assessments as per Benza et al(16). P values < 0.05 were considered statistically significant. To minimize type II error and potential false negative results, results were interpreted using nominal P values given that our module-trait correlation analysis was intended to be exploratory and hypothesis-generating rather than to confirm an a prior hypothesis about module-trait correlations. Results will need to be confirmed in future targeted studies. Multiple testing correction (n=260 comparisons from 13

184 traits and 20 modules) was also performed and correlations with FDR < 0.05 are shown in a
185 supplemental figure.
186
187 *Genome-wide association study (GWAS) enrichment analysis*
188 Enrichment of modules for PAH GWAS single-nucleotide polymorphisms (SNPs) were assessed
189 using two distinct computational methods, MAGMA(17) and GSA-SNP2(18), across four
190 independent PAH GWAS cohorts totaling 11,744 individuals(19, 20): the US PAH Biobank
191 (PAHB), French Pulmonary Hypertension Allele-Associated Risk (PHAAR), British Heart
192 Foundation Pulmonary Arterial Hypertension (BHFPAH), and UK National Institute for Health
193 Research BioResource (NIHRBR) cohorts. SNPs were mapped to genes by chromosomal
194 proximity (within 20 kilobases from the 5' or 3' ends of a gene) and genes were scored for
195 association with PAH based on disease-SNP *P*-value associations from GWAS summary
196 statistics. SNPs were not filtered (i.e. by a specified statistical threshold) prior to input into
197 MAGMA and GSA-SNP2. Gene scores were then used in competitive gene-set analyses to
198 identify module enrichment for PAH common genetic variation. To aggregate genetic variants
199 into a gene score, the mean χ^2 statistic and the log-minimum GWAS *P* value for all SNPs
200 localizing to a gene were used as per MAGMA(17) and GSA-SNP2(18), respectively. To
201 determine significance, MAGMA uses a linear mixed model whereas GSA-SNP2 uses a standard
202 normal distribution. Both methods adjust for gene size and gene density (the number of SNPs
203 assigned to a given gene). The default statistical results were reported for MAGMA (*P* value)
204 and GSA-SNP2 (false discovery rate). MAGMA *P* values were not corrected post-hoc for
205 multiple testing given that this analysis was intended to be exploratory and hypothesis-generating

rather than to confirm an a prior hypothesis about GWAS enrichment. Results will need to be confirmed in future targeted studies.

Bayesian gene regulatory network analysis

In a complementary approach to co-expression analysis to infer co-regulation, we employed Bayesian network (BN) analysis to build a gene regulatory network. Specifically, BNs were constructed using Reconstructing Integrative Molecular Bayesian Network (RIMBANet)(21). For this method, 1000 networks were generated from different random seed genes using continuous and discrete expression data derived from transcriptomes from either GSE23546 (n = 1343) (22), PHBI (n = 146), or GTEx v8 (n = 577) (23). Whole lung-specific cis eQTLs from GTEx v8 (23) and transcription factor-target gene data from HTRI (24), TRRUST (25), and PAZAR (26) databases were used as priors. Then, the final network for each of the 3 datasets was obtained by taking a consensus network from the 1000 randomly generated networks whereby only edges that passed a probability of >30% across the 1000 BNs were kept. Finally, the union of the 3 networks was taken to create a combined gene regulatory network derived from a total of 2,066 human lungs. The network was visualized in Cytoscape(27) where nodes represent genes and edges represent inferred directional gene-gene regulation. Node positions were determined by a prefuse force-directed algorithm. Genes from co-expression modules were projected onto this regulatory network with different color nodes representing module membership. Known PAH genes from disease-gene databases (Comparative Toxicogenomics Database(11) and DisGeNET(10)) were also projected onto the network. Local hub genes of a particular subnetwork (i.e. pink subnetwork) whose neighboring nodes are enriched for genes in

the gene set of interest (i.e. pink gene set) were determined by Key Driver Analysis in the Mergeomics R package(4, 28, 29).

Analysis of public transcriptomic datasets

Public transcriptomic datasets were queried for expression of selected genes (i.e. *GUCY1A2*, *ANTXR1, USP28,* and *USP12*). *GUCY1A2* expression was obtained from two independent RNA-seq datasets: CRISPR/Cas9-induced monoallelic mutations in BMPR2 (n = 6) vs wild-type control (n = 3) human umbilical vein endothelial cells (HUVEC)(30) and endothelial cells derived from induced pluripotent stem cells (iPSCs) of patients with hereditary PAH (HPAH) due to BMPR2 mutations (n = 5) vs control (n = 3)(31). Sequence Read Archive (SRA) data was downloaded from the NIH Gene Expression Omnibus (GEO) database. SRA files were converted to FASTQ files using the SRA Toolkit. Sequences were aligned to the GENCODE human reference genome (v. 32) using HISAT2(7) and transcripts were assembled and quantified using StringTie(7). DESeq2(9) was used to perform differential expression analysis and determine *P* values. *ANTXR1* expression in fibroblasts was obtained from an scRNAseq dataset comparing 3 PAH vs 6 control lungs (32). *ANTXR1* expression counts were averaged across cells within the fibroblast cluster for each sample. Note, myofibroblasts were not subclustered in this dataset. Wilcoxon rank-sum test was used to determine differential expression of *ANTXR1* in PAH vs. control samples. The expression of *USP28* and *USP12* was obtained from a human whole lung microarray of 15 PAH patients vs 11 controls(33) deposited in the NIH GEO database. *P* values were obtained from NCBI's GEO2R.

Quantitative polymerase chain reaction (qPCR)

Pulmonary arterial adventitial cells isolated from idiopathic PAH and control lungs were obtained from PHBI and grown in Human Vascular Smooth Muscle Cell Basal Medium (M231500, ThermoFisher) supplemented with Smooth Muscle Growth Supplement (S00725, ThermoFisher) and Antibiotic-Antimycotic 100X (15240096, ThermoFisher). Both PAH and control cells originated from the lungs of white women age 29 and 33, respectively. Cells were collected in TRIzol (15596018, ThermoFisher) once grown to 100% confluency between passages 6 to 10. Different passages served as biological replicates. RNA was extracted from cells through a series of washes using chloroform (364320025, ThermoFisher), isopropanol (I9516, MillaporeSigma), and 70% ethanol (459844, MillaporeSigma). RNA was resuspended in DEPC-Treated Water (AM9922,ThermoFisher) and then converted to cDNA using a High-Capacity cDNA Reverse Transcription Kit (4368814, ThermoFisher) and a Bio-Rad S1000 Thermal Cycler. A qCPR was run using cDNA, DEPC water, PowerUp™ SYBR™ Green Master Mix (A25743, ThermoFisher) and primers on a Bio-Rad CFX Connect Real-Time PCR Detection System. Primers were used for ANTXR1 (Forward: GAGGAAACGGCTTCCGACAT, Reverse: GAGTGCAGCTTTCATGCCAA) and housekeeping gene RPLP0 (Forward: CAGGTGTTCGACAATGGCAG, Reverse: ACAAGGCCAGGACTCGTTTG).

Deconvolution

To serve as a cell type reference for deconvolution, we integrated seven publicly available human lung single-cell RNAseq datasets(34–40) and identified 37 cell-type clusters using known marker genes from the literature. Within each cell-type cluster, the average expression of gene counts was calculated across cells for each individual sample to create a cell-type signature for

each of the seven datasets. PHBI bulk transcriptomes were deconvoluted with CIBERSORTx(41) with cell-type signatures from each of the seven datasets as a reference. The resulting cell fractions using each of the seven dataset-specific reference signatures served as technical replicates. These technical replicates were then averaged to determine the final estimated cell fractions for each lung sample. Pearson correlation of deconvoluted cell type fractions with PAH vs control status (coded 1 and 0, respectively) was calculated across PHBI lung samples. Wilcoxon rank-sum test was performed on cell fractions between PAH vs. control samples for vascular cell types. Similar to module-trait correlation analysis, modules were correlated with cell fractions by calculating Pearson correlations of module eigengenes with deconvoluted cell fractions across samples.

Pharmacotranscriptomic analysis

Genes differentially expressed between PAH and control in select co-expression modules (i.e. pink, royalblue, greenyellow) were queried against the full Connectivity Map(42) (CMap) database of perturbagen expression signatures induced in human cell lines(42). A less stringent statistical threshold for differentially expressed genes (P value < 0.05) were used as previously described(43) to ensure an adequate number of query signature genes to perform the CMap analysis. A total of 8,559 pharmacologic and genetic perturbagens were screened including both gene overexpression and knockdown by short hairpin RNA (shRNA). Pattern-matching algorithms assessed each reference perturbagen profile for the direction and strength of connectivity with the query signature by a score range of -100 to +100. The summary score was used across 9 cell lines. Perturbagens with strongly positive connectivity scores indicate highly similar signatures that mimic that of the query whereas those perturbagens with strongly negative

scores indicate signatures that strongly reverse that of the query (i.e. genes that are differentially upregulated in the module query are decreased by the perturbagen or vice versa). We also assessed the connectivity scores of a total of 171 CMap classes defined as groups of pharmacologic or genetic CMap perturbagens that share the same mechanism of action or biological function.

The pink signature of differentially expressed genes (PAH vs. control; P value < 0.05) was also queried against the CRISPR knockout (KO) consensus signature database of Library of Integrated Network-based Cellular Signatures (LINCS) L1000 using SigCom LINCS(44). A total of 7,502 genes were screened by CRISPR KO from this database. SigCom outputs a Z score which indicates the degree to which the CRISPR KO signature mimics or reverses the query signature (i.e. pink) by highly positive or negative scores, respectively. The expression of 12,327 genes were profiled and ranked for each CRISPR KO gene where lowly and highly ranked genes indicate downregulation and upregulation, respectively, in the CRISPR KO vs. control.

Results

Characterization of PHBI cohort.

RNA sequencing was performed on a total of 96 explant lungs with pulmonary hypertension (PH) collected at the time of lung transplantation and 52 control lungs from the Pulmonary Hypertension Breakthrough Initiative (PHBI) (Table 1 and Figure 1A). WHO group 1 PAH patients consisted of 94 of 96 PH subjects of which the most common subtypes were idiopathic PAH (IPAH) and associated PAH (APAH) (43% and 40%, respectively). The majority of PH patients were WHO functional class III or IV, had significantly impaired hemodynamics by right

heart catheterization with a mean PVR of 12.7±7.4, and were receiving triple PAH-targeted therapy (73%) including prostacyclin infusion therapy (85%). Unsupervised hierarchical clustering and principal component analysis (PCA) showed separation between PH and control samples suggesting overall differences in the transcriptome between the two groups (Figures 1A-1B). Neither approach showed distinct separation of samples by PAH subgroup suggesting relative lack of subgroup-specific transcriptional heterogeneity. Furthermore, we did not observe significant clustering by age, sex, or race among PH or control samples. Moreover, samples did not cluster by transplant center of tissue origin nor treatment group (Supplemental Figure 1). Given the female predominance of PH subjects compared to control, sex-stratified analyses were performed where appropriate. Two outliers (1 IPAH and 1 control) were removed from downstream analyses (Figures 1A-1B) since network analysis and module detection can be biased by outlier samples(45).

The lung transcriptome is significantly altered in PAH

Consistent with the separation observed between PH and control samples by hierarchical clustering and PCA, differential expression analysis between PAH and control samples yielded 5,253 differentially expressed genes (DEGs; FDR < 0.05) consisting of 22% of the transcriptome of which 2,719 were upregulated and 2,534 were downregulated (Figure 1C). The top upregulated genes were *HBA2*, *HBB*, *LAMP5*, *HBA1*, and *MFAP4* whereas the top downregulated genes were *SIGLEC10*, *PI3*, *SAA2*, *SLC36A1*, and *ALPP*. Epithelial-mesenchymal transition (EMT) was the top enriched pathway among genes upregulated in PAH whereas mTORC1 signaling was the top pathway among downregulated genes (Figure 1D).

Co-expression network analysis reveals modules associated with PAH severity.

We then used weighted gene co-expression network analysis (WGCNA)(46) across all samples to dissect the lung transcriptome into clusters based on gene co-expression, referred to as modules. Modules organize transcriptional changes of individual genes into clusters which represent co-regulation or shared biological functions(47) (Figure 2A-B). We identified 20 gene co-expression modules with a median size of 141 genes (Figure 2C). The expression of genes within a module can be summarized by their eigengene which represents the first principal component of gene expression in the module. Correlation of clinicopathologic characteristics with module eigengenes revealed the pink module of 266 genes to have the most notable pattern of associations (Figure 2D). The pink module was not only strongly associated with PAH diagnosis (Figures 2D-2E, Supplemental Figure 2) but also with physiologic, hemodynamic, and pathologic markers of disease severity based on pulmonary function testing, right heart catheterization, and histologic analysis of vascular remodeling by morphometry. Specifically, the pink module correlated with reduced diffusing capacity for carbon monoxide (DLCO), elevated mean pulmonary artery pressure (mPAP), elevated pulmonary vascular resistance (PVR), and increased intima or intima plus media thickness of pulmonary arteries. However, the pink module also correlated with clinical characteristics and blood tests suggestive of compensated disease, such as lower number of hospitalizations due to PAH, signs of right heart failure, WHO functional class, NT-proBNP, creatinine, and REVEAL lite score. Among other modules positively correlated with PAH diagnosis, the royalblue module of 98 genes shared a similar pattern of clinicopathologic correlation as with the pink module. The greenyellow module of 290 genes had the highest negative correlation with PAH diagnosis (Figure 2F). Pink and royalblue were also the top two modules most strongly enriched for genes upregulated in PAH lungs and

366 greenyellow was most strongly enriched for downregulated genes (Figure 2G, Supplemental
367 Figure 3).
368
369 *The pink module is enriched for PAH genetic variations.*
370 Having identified PAH-relevant modules, we next asked whether these modules might be a cause
371 or consequence of PAH pathogenesis. To infer causality, we integrated PAH genetic association
372 studies with our lung-derived modules. Specifically, we employed two distinct computational
373 approaches, MAGMA(17) and GSA-SNP2(18), to test whether modules were enriched for PAH-
374 associated single nucleotide polymorphisms (SNPs) using the full summary statistics from four
375 independent PAH genome-wide association studies (GWAS) totaling 11,744 individuals(19, 20):
376 the US PAH Biobank (PAHB), French Pulmonary Hypertension Allele-Associated Risk
377 (PHAAR), British Heart Foundation Pulmonary Arterial Hypertension (BHFPAH), and UK
378 National Institute for Health Research BioResource (NIHRBR) cohorts. Despite different
379 statistical methods, MAGMA and GSA-SNP2 captured similar relative associations of genes
380 with PAH genetic variation and neither approach was biased towards gene size or number of
381 SNPs localizing to a gene (Supplemental Figure 4). We found that only the pink module was
382 significantly enriched for PAH-associated SNPs using both approaches and across multiple
383 cohorts (Figures 2H-2I): PAHB and PHAAR cohorts by both methods, and BHFPAH using
384 GSA-SNP2. This finding suggests that pink module genes are not only associated with PAH
385 diagnosis and severity, but also enriched with genetic risk of developing PAH.
386
387 *The pink module is co-regulated with known PAH genes and is enriched for Wnt signaling and*
388 *EMT pathways.*

To delineate the regulatory relationships among genes within co-expression modules, we employed a Bayesian network analysis to build a gene regulatory network of the human lung by incorporating 2,066 lung transcriptomes, lung-specific expression quantitative trait loci (eQTL), and known transcription factor-target gene relationships (Supplemental Figure 5). Projection of co-expression module genes onto this lung regulatory network confirmed that the genes within individual modules are in close neighborhoods in the gene regulatory network analyses (Figure 3A).

We found that a number of established PAH genes co-localize with pink module genes in the Bayesian gene regulatory network, suggesting a regulatory relationship between PAH genes and pink module genes (Figure 3B). For example, BMPR2, the most well-established causal PAH gene, was predicted to regulate GUCY1A2, a pink module gene that is also upregulated in PAH lungs (Supplemental Figure 6). To validate this prediction, we queried public RNA-seq datasets (30, 31) and found that *GUCY1A2* was upregulated in CRISPR-induced BMPR2 mutant endothelial cells (ECs), and in ECs derived from induced pluripotent stem cells (iPSCs) from hereditary PAH patients with BMPR2 mutations (Figures 3C-3D).

We then used gene set enrichment analysis (GSEA) to functionally characterize the pink signature which we defined as the correlation of the pink eigengene with the expression of genes across the transcriptome. We found that regulation of Wnt signaling and epithelial mesenchymal transition (EMT), both important pathways in PAH(48–50), were strongly enriched in the pink module (Figures 3E-G). *ANTXR1* was the top pink hub gene most connected to pink module genes and most correlated with PAH (Figure 3G). While not traditionally associated with Wnt

signaling or EMT, *ANTXR1* has been recently implicated in various cancers through such pathways(51, 52). Furthermore, the strongest pairwise connection among pink genes as determined by WGCNA (out of >30k pairs) was between *ANTXR1* and *SFRP4*, a secreted frizzled-related protein and one of 3 pink genes known to be involved in both Wnt signaling and EMT (Figures 3G-3H). Therefore, the pink module and its top PAH-associated gene *ANTXR1* may play a role in PAH through modulation of Wnt signaling and EMT.

Cell type deconvolution reveals cell-type specificity in PAH lung modules.

Having identified PAH-specific transcriptional changes at the whole lung level, we next asked whether cell-type fractional changes could be inferred from the transcriptomes of PAH and control lungs by deconvolution analysis based on transcriptomic references of 37 lung cell type clusters from seven publicly available human lung single-cell RNAseq datasets (34–40) (Figures 4A-4B; Supplemental Figure 7). Using the cell-type references from this integrated reference atlas, we deconvoluted PHBI bulk transcriptomes using CIBERSORTx(41) and found that PAH samples clustered together based on estimated cell fractions (Supplemental Figure 8) and that specific cell-type fractions correlated positively or negatively with PAH, such as endothelial (i.e. lymphatic and arterial) and myeloid (i.e. interstitial macrophage and classical monocyte) subpopulations, respectively (Figure 4C). In addition to lymphatic and arterial endothelial cells, myofibroblast fractions were particularly abundant in PAH samples relative to control, whereas cell fractions of other vascular mesenchymal subpopulations (i.e. fibroblast, smooth muscle cell and pericyte) were unchanged. Endothelial capillary 1 (EndoCap1) fractions were decreased in PAH lungs consistent with microvascular rarefaction, a known feature of PAH(53).

We then integrated the deconvoluted cell fractions with our bulk lung-derived co-expression modules to decipher cell-type specificity of individual modules. We identified distinct patterns of correlation between cell fractions and module eigengenes where, for example, some modules were highly specific to particular cell types such as yellow to ciliated cells and cyan to $CD8^+$ T cells (Figure 4E). Among the top correlations was the PAH-associated pink module to myofibroblast fractions ($r = 0.78$). In a complementary approach to infer cell specificity, we then performed GSEA using known cell-type signatures of 341 cell types across >9 tissues and found the pink module to be most enriched in the myofibroblast signature (Figure 4F). Given this finding, we then asked whether *ANTXR1*, the top pink gene whose expression is upregulated in PAH lungs, might be upregulated in PAH lung fibroblasts specifically. Supporting our cell-type deconvolution analyses, we found that *ANTXR1* is upregulated in PAH fibroblasts by lung single-cell RNA-sequencing in a published dataset (in which myofibroblasts were not subclustered) (32) and in pulmonary arterial adventitial cells isolated from PAH lungs by qPCR (Figures 4G-4H).

Pharmacotranscriptomics identifies novel therapeutic targets.

Having prioritized the pink module by association with PAH diagnosis, clinicopathologic severity, and genetic risk, we next investigated whether pattern matching the pink module signature with known pharmacologic and genetic perturbation signatures could reveal novel therapeutic targets. We screened the pink signature against 8,559 perturbation signatures from Connectivity Map (CMap) (42). These CMap signatures were grouped into 171 classes that share similar mechanisms of action or biological functions. We found ubiquitin specific peptidase (USP) loss-of-function by short hairpin RNAs (shRNAs) to have the highest connectivity score

of all CMap classes, which suggests that knockdown of USPs induces a transcriptional response highly similar to the pink signature (Figure 5A). Indeed, six different USPs (USP7, USP22, USP12, USP20, USP1, and USP15) were among the top scoring perturbations (Figure 5B). We next queried an independent genetic perturbation screen of 7,502 genes by CRISPR knockout (KO) from the LINCS L1000 database(44) and found that USP28 KO was a top mimicker of the pink signature (Figures 5C-5D). Therefore, targeting members of the USP family by either complete knockout via CRISPR or partial knockdown via shRNA induced transcriptional changes similar to that of pink module genes in PAH lungs.

Interestingly, Janus kinase (JAK) and cyclin-dependent kinase (CDK) inhibitors, both recently studied as potential therapies in PAH(54–56), were also among the top CMap classes whose transcriptional signature matched that of the pink module (Figure 5A). We found a similar connectivity profile for another co-expression module, royalblue, containing a distinct set of 98 genes which also shared with pink a similar pattern of clinicopathologic correlations (Figures 5A and 2D) and was second only to pink as a top enriched module for upregulated PAH DEGs (Figure 2G). This suggests converging and targetable pathways between pink and royalblue genes. In contrast, greenyellow, the most enriched module for downregulated DEGs, did not show a strong CMap matching profile (Figure 5A). Indeed, royalblue was also enriched for Wnt signaling and EMT and a number of its module member genes co-localized to the pink sub-regulatory network (Supplemental Figure 9) such as LTBP2, a recently identified biomarker for PAH(57).

Given that JAK and CDK inhibitors, both of which mimicked the pink signature (Figure 5A), may have a therapeutic role in PAH based on recent preclinical studies(54–56) and given that the pink eigengene correlated with clinical markers of favorable disease prognosis (Figure 2D), we reasoned that the pink signature could be beneficial in PAH. As downregulation of USPs induced transcriptomic signatures that mimicked that of the pink module in our CMap and L1000 analyses, we postulate that targeting USPs might also have a therapeutic role. We found that USP28 and USP12, whose downregulation led to transcriptomic responses mimicking that of the pink module, were upregulated in PAH lungs in our PHBI dataset (Figure 5E) and in an independent microarray (Figures 5F-5G). These USPs may serve as top candidates for PAH therapeutic development.

Discussion

Leveraging the largest PAH lung biobank to date and the first to use RNA-sequencing combined with state-of-the-art multiomic integration and systems biology approaches, we dissected the transcriptional landscape of PAH lungs to uncover a novel gene module enriched in upregulated genes and associated with clinicopathologic severity, genetic risk, specific vascular cell types, and new therapeutic targets in PAH.

We identified by network analysis a set of 266 co-expressed genes called the pink module that was not only associated with objective measures of underlying disease severity such as increased PVR, increased intimal thickness, and reduced DLCO, but also associated with lower risk of mortality by REVEAL lite as well as indicators of clinically compensated PAH such as lower number of PAH hospitalizations, signs of right heart failure, WHO functional class and NT-

proBNP (Figure 2D). We hypothesize that the pink module is active in response to the underlying disease process to counteract disease progression in PAH. Supporting this possibility, JAK and CDK inhibitors, both of which counteract preclinical PAH (54–56), were top perturbagens predicted to mimic the pink signature in our CMap analysis (Figure 5A). Moreover, our regulatory network analysis uncovered a novel connection where deficiency of BMPR2, the most well-established causal PAH gene, leads to an upregulation of pink gene GUCY1A2 (Figures 3B-3D). GUCY1A2 encodes the alpha subunit of soluble guanylate cyclase 1 (GC-1), the primary receptor of nitric oxide and the stimulation of which is the primary mechanism of action of riociguat, an FDA-approved therapy in clinical use to treat PAH patients. Therefore, supporting our hypothesis that the pink module might be a response to PAH to counteract the disease, deficiency of BMPR2, which is causative and harmful in PAH pathogenesis, leads to upregulation of the pink module gene GUCY1A2 which is beneficial in PAH.

EMT and Wnt signaling were top pathways enriched in the pink module, both of which are known to play a critical role in PAH pathobiology and are interrelated (Wnt signaling induces EMT) (Figures 3E-3G)(48, 49, 58, 59). In terms of specific pink genes and their potential role in these pathways, ANTXR1, a transmembrane protein that interacts with extracellular matrix proteins, was the top hub gene most connected to other pink module genes and it was upregulated in not only PAH lungs (Figure 1C) but specifically in PAH lung fibroblasts (Figures 4G-4H). While its role in PAH has not been investigated, one study found that ANTXR1-deficient fibroblasts showed increased expression of EMT markers *Col1a1* and *Fn* raising the possibility that ANXTXR1 might modulate EMT in a beneficial manner in PAH(60). Furthermore, SFRP4, which showed the strongest pairwise correlation with ANTXR1 among all

pink gene pairs (Figure 3H), is a secreted Wnt antagonist that has been shown to inhibit EMT in cancer cells(61, 62). Therefore, the pink module may counteract the PAH disease process by modulation of EMT and Wnt signaling.

The following question then arises- if the pink module and its member genes (i.e. ANTXR1, SFRP4, and GUCY1A2) are beneficial in PAH and upregulated in explant lungs, why did the disease in these patients still progress to the point of needing a lung transplantation? One possibility is that the pink module is activated too little and/or too late in these patients or the activation is not sufficient to counteract the effects of other deleterious pathways. Despite its relatively large size of 266 genes, the pink signature can be leveraged using pharmacotranscriptomic pattern matching algorithms to identify novel therapeutic targets for testing in future investigations with preclinical models. Using such an approach, we identified JAK and CDK inhibitors as well as ubiquitin specific peptidase (USP) loss-of-function as CMap perturbagen classes which induce signatures that match that of pink, but also ubiquitin specific peptidase (USP) loss-of-function as the top CMap class mimicker of the pink signature. As main members of the deubiquitinase family, USPs are involved in diverse processes such as cell cycle progression, apoptosis, EMT, and DNA damage repair and have been strongly implicated in cancer progression(63). Furthermore, USPs regulate PAH-relevant pathways such as NFkB, TGFbeta, and Wnt signaling and have been investigated as therapeutic targets in cancer and other fields(64, 65), yet their role in PAH has not yet been described. We demonstrated that targeting members of the USP family by either complete knockout via CRISPR or partial knockdown via shRNA induced transcriptional changes similar to that of pink DEGs in PAH lungs (Figure 5). Specifically, USP28 and USP12 may serve as particularly attractive targets for

downstream investigation as they were also upregulated in PAH lungs (Figures 5E-5G). While studies on USP12 are limited, USP28 has been shown to activate Wnt signaling(66, 67) and its inhibition blocks EMT progression in cancer cells(68). Thus, inhibition of Wnt signaling and EMT may be the common pathways shared between USP loss-of-function and the pink module.

Given that the PHBI cohort consisted of patients with advanced stage PAH at the time of lung transplantation, our results are likely not representative of the full range of disease and our analysis was limited in discerning cause versus consequence of PAH. However, the deep clinical phenotyping allowed us to make correlations with disease severity, and GWAS integration enabled us to infer causality in PAH pathogenesis. The majority of patients had idiopathic PAH and thus our findings may not be generalizable to other WHO Group 1 PAH subtypes or other WHO groups, and our sample size of other PAH subtypes was insufficiently powered to detect subtype-specific differences. The majority of our patients were also female, reflective of the strong female predominance of PAH. However, our sensitivity analyses did not reveal significant sex-specific differences in the top pathway and module enrichments (Supplemental Figure 10). Finally, while heterogeneity of PAH-targeted therapy in these patients could affect the transcriptome profiles, the majority of patients were on triple therapy including prostacyclin infusion. Thus, we did not explore treatment-specific differences.

In conclusion, our study leverages the largest PAH lung biobank to date to provide an in-depth analysis of the lung transcriptional landscape of PAH using multiomic integration and systems biology approaches. Through this analysis, we uncovered a novel gene network module that is associated with PAH risk and severity, may counteract disease progression through modulation

573 of EMT and Wnt signaling, and may be regulated by USPs. Future experimental studies such as
574 knockdown of USPs in PAH vascular cells are warranted to further investigate the role and
575 therapeutic potential of the pink module and targeting USPs in PAH.
576

577 **Acknowledgements**

578 The authors thank the investigators, personnel, and participants of the PHBI, particularly those
579 involved in the transplant and preparation centers: Allegheny University of the Health Sciences
580 (PI: Raymond L. Benza, M.D.); Baylor College of Medicine (George Noon, M.D.); Cleveland
581 Clinic (PI: Serpil Erzurum, M.D.); Duke University (PI: Pang-Chieh Jerry Eu, M.D.); Stanford
582 University-UCSF (PI: Marlene Rabinovitch, M.D.); University of Alabama (PI: Keith Wille,
583 M.D.; prior PI: Raymond L. Benza, M.D.); University of California, San Diego (PI: Patricia
584 Thistlethwaite, M.D., Ph.D); Vanderbilt University (Barbara Meyrick, Ph.D.).
585 Supported by ALA CA-675591 (J.H.) and NHLBI R01HL162124 (M.E.).
586

587 **References**

588 1. Stearman RS, Bui QM, Speyer G, Handen A, Cornelius AR, Graham BB, *et al.* Systems
589 Analysis of the Human Pulmonary Arterial Hypertension Lung Transcriptome. *Am J Respir*
590 *Cell Mol Biol* 2019;60:637–649.
591 2. Hoffmann J, Wilhelm J, Olschewski A, Kwapiszewska G. Microarray analysis in pulmonary
592 hypertension. *European Respiratory Journal* 2016;48:229–241.

3. Porcu E, Sadler MC, Lepik K, Auwerx C, Wood AR, Weihs A, *et al.* Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat Commun* 2021;12:5647.
4. Shu L, Zhao Y, Kurt Z, Byars SG, Tukiainen T, Kettunen J, *et al.* Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC Genomics* 2016;17:874.
5. Stacher E, Graham BB, Hunt JM, Gandjeva A, Groshong SD, McLaughlin VV, *et al.* Modern Age Pathology of Pulmonary Arterial Hypertension. *American Journal of Respiratory and Critical Care Medicine* 2012;186:261–272.
6. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
7. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 2016;11:1650–1667.
8. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 2015;4:1521.
9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 2014;15:550.
10. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 2017;45:D833–D839.

11. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J, *et al.* The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res* 2019;47:D948–D954.
12. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* 2016;2016:.
13. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–425.
14. The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, Harris NL, *et al.* The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* 2021;49:D325–D334.
15. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573-3587.e29.
16. Benza RL, Kanwar MK, Raina A, Scott JV, Zhao CL, Selej M, *et al.* Development and Validation of an Abridged Version of the REVEAL 2.0 Risk Score Calculator, REVEAL Lite 2, for Use in Patients With Pulmonary Arterial Hypertension. *Chest* 2021;159:337–346.
17. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015;11:e1004219.
18. Yoon S, Nguyen HCT, Yoo YJ, Kim J, Baik B, Kim S, *et al.* Efficient pathway enrichment and network analysis of GWAS summary data using GSA-SNP2. *Nucleic Acids Res* 2018;46:e60.

19. Rhodes CJ, Batai K, Bleda M, Haimel M, Southgate L, Germain M, *et al.* Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *Lancet Respir Med* 2019;7:227–238.
20. Germain M, Eyries M, Montani D, Poirier O, Girerd B, Dorfmüller P, *et al.* Genome-wide association analysis identifies a susceptibility locus for pulmonary arterial hypertension. *Nature Genetics* 2013;45:518–521.
21. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* 2008;40:854–861.
22. Lesseur C, Ferreiro-Iglesias A, McKay JD, Bossé Y, Johansson M, Gaborieau V, *et al.* Genome-wide association meta-analysis identifies pleiotropic risk loci for aerodigestive squamous cell cancers. *PLoS Genet* 2021;17:e1009254.
23. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;369:1318–1330.
24. Bovolenta LA, Acencio ML, Lemke N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* 2012;13:405.
25. Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, *et al.* TRRUST: a reference database of human transcriptional regulatory interactions. *Sci Rep* 2015;5:11432.
26. Portales-Casamar E, Kirov S, Lim J, Lithwick S, Swanson MI, Ticoll A, *et al.* PAZAR: a framework for collection and dissemination of cis-regulatory sequence annotation. *Genome Biology* 2007;8:R207.

656 27. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* Cytoscape: a
657 software environment for integrated models of biomolecular interaction networks. *Genome*
658 *Res* 2003;13:2498–2504.
659 28. Arneson D, Bhattacharya A, Shu L, Mäkinen V-P, Yang X. Mergeomics: a web server for
660 identifying pathological pathways, networks, and key regulators via multidimensional data
661 integration. *BMC Genomics* 2016;17:722.
662 29. Ding J, Blencowe M, Nghiem T, Ha S, Chen Y-W, Li G, *et al.* Mergeomics 2.0: a web server
663 for multi-omics data integration to elucidate disease networks and predict therapeutics.
664 *Nucleic Acids Research* 2021;49:W375–W387.
665 30. Hiepen C, Jatzlau J, Hildebrandt S, Kampfrath B, Goktas M, Murgai A, *et al.* BMPR2 acts as
666 a gatekeeper to protect endothelial cells from increased TGFβ responses and altered cell
667 mechanics. *PLoS Biol* 2019;17:e3000557.
668 31. Gu M, Shao N-Y, Sa S, Li D, Termglinchan V, Ameen M, *et al.* Patient-Specific iPSC-
669 Derived Endothelial Cells Uncover Pathways that Protect against Pulmonary Hypertension in
670 BMPR2 Mutation Carriers. *Cell Stem Cell* 2017;20:490-504.e5.
671 32. Saygin D, Tabib T, Bittar HET, Valenzi E, Sembrat J, Chan SY, *et al.* Transcriptional
672 profiling of lung cell populations in idiopathic pulmonary arterial hypertension. *Pulm Circ*
673 2020;10:.
674 33. Mura M, Cecchini MJ, Joseph M, Granton JT. Osteopontin lung gene expression is a marker
675 of disease severity in pulmonary arterial hypertension. *Respirology* 2019;24:1104–1110.
676 34. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, *et al.*
677 Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology
678 of Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2018;doi:10.1164/rccm.201712-2410OC.

679 35. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, *et al.* Single-cell RNA-
680 seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary
681 fibrosis. *Science Advances* 2020;6:eaba1983.
682 36. Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT,
683 Georgakopoulos N, *et al.* scRNA-seq assessment of the human lung, spleen, and esophagus
684 tissue stability after cold preservation. *Genome Biology* 2019;21:1.
685 37. Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, *et al.* Construction of a human cell
686 landscape at single-cell level. *Nature* 2020;581:303–309.
687 38. Morse C, Tabib T, Sembrat J, Buschur K, Bittar HT, Valenzi E, *et al.* Proliferating
688 SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *European*
689 *Respiratory Journal* 2019;1802441.doi:10.1183/13993003.02441-2018.
690 39. Raredon MSB, Adams TS, Suhail Y, Schupp JC, Poli S, Neumark N, *et al.* Single-cell
691 connectomic analysis of adult mammalian lungs. *Science Advances* 2019;5:eaaw3851.
692 40. Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, *et al.* Single-cell
693 RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in
694 pulmonary fibrosis. *Science Advances* 2020;6:eaba1972.
695 41. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating
696 immune cells with CIBERSORT. *Methods Mol Biol* 2018;1711:243–259.
697 42. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, *et al.* The Connectivity
698 Map: using gene-expression signatures to connect small molecules, genes, and disease.
699 *Science* 2006;313:1929–1935.
700 43. Hong J, Arneson D, Umar S, Ruffenach G, Cunningham CM, Ahn IS, *et al.* Single-Cell
701 Study of Two Rat Models of Pulmonary Arterial Hypertension Reveals Connections to

Human Pathobiology and Drug Repositioning. *Am J Respir Crit Care Med* 2021;203:1006–1022.

44. Evangelista JE, Clarke DJB, Xie Z, Lachmann A, Jeon M, Chen K, *et al.* SigCom LINCS: data and metadata search engine for a million gene expression signatures. *Nucleic Acids Research* 2022;50:W697–W709.
45. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, *et al.* Functional organization of the transcriptome in human brain. *Nat Neurosci* 2008;11:1271–1282.
46. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
47. Weirauch MT. Gene Coexpression Networks for the Analysis of DNA Microarray Data. *Applied Statistics for Network Biology* John Wiley & Sons, Ltd; 2011. p. 215–250.doi:10.1002/9783527638079.ch11.
48. Ranchoux B, Antigny F, Rucker-Martin C, Hautefort A, Péchoux C, Bogaard HJ, *et al.* Endothelial-to-mesenchymal transition in pulmonary hypertension. *Circulation* 2015;131:1006–1018.
49. de Jesus Perez V, Yuan K, Alastalo T-P, Spiekerkoetter E, Rabinovitch M. Targeting the Wnt signaling pathways in pulmonary arterial hypertension. *Drug Discov Today* 2014;19:1270–1276.
50. West JD, Austin ED, Gaskill C, Marriott S, Baskir R, Bilousova G, *et al.* Identification of a common Wnt-associated genetic signature across multiple cell types in pulmonary arterial hypertension. *Am J Physiol Cell Physiol* 2014;307:C415–C430.

723 51. Huang X, Zhang J, Zheng Y. ANTXR1 Is a Prognostic Biomarker and Correlates With
724 Stromal and Immune Cell Infiltration in Gastric Cancer. *Frontiers in Molecular Biosciences*
725 2020;7:.
726 52. Ding C, Liu J, Zhang J, Wan Y, Hu L, Charwudzi A, *et al.* Tumor Endothelial Marker 8
727 Promotes Proliferation and Metastasis via the Wnt/β-Catenin Signaling Pathway in Lung
728 Adenocarcinoma. *Frontiers in Oncology* 2021;11:.
729 53. Chaudhary KR, Taha M, Cadete VJJ, Godoy RS, Stewart DJ. Proliferative Versus
730 Degenerative Paradigms in Pulmonary Arterial Hypertension. *Circulation Research*
731 2017;120:1237–1239.
732 54. Leopold JA. Inhibiting Jak2 Ameliorates Pulmonary Hypertension: Fulfilling the Promise of
733 Precision Medicine. *Am J Respir Cell Mol Biol* 2021;64:12–13.
734 55. Yerabolu D, Weiss A, Kojonazarov B, Boehm M, Schlueter BC, Ruppert C, *et al.* Targeting
735 Jak-Stat Signaling in Experimental Pulmonary Hypertension. *Am J Respir Cell Mol Biol*
736 2021;64:100–114.
737 56. Tabarroki A, Lindner DJ, Visconte V, Zhang L, Rogers HJ, Parker Y, *et al.* Ruxolitinib leads
738 to improvement of pulmonary hypertension in patients with myelofibrosis. *Leukemia*
739 2014;28:1486–1493.
740 57. Boucherat O, Yokokawa T, Krishna V, Kalyana-Sundaram S, Martineau S, Breuils-Bonnet
741 S, *et al.* Identification of LTBP-2 as a plasma biomarker for right ventricular dysfunction in
742 human pulmonary arterial hypertension. *Nat Cardiovasc Res* 2022;1:748–760.
743 58. Stenmark KR, Frid M, Perros F. Endothelial-to-Mesenchymal Transition: An Evolving
744 Paradigm and a Promising Therapeutic Target in PAH. *Circulation* 2016;133:1734–1737.

745 59. Ma J, Sanchez-Duffhues G, Goumans M-J, ten Dijke P. TGF-β-Induced Endothelial to
746 Mesenchymal Transition in Disease and Tissue Engineering. *Frontiers in Cell and*
747 *Developmental Biology* 2020;8:.
748 60. Hu K, Olsen BR, Besschetnova TY. Cell autonomous ANTXR1-mediated regulation of
749 extracellular matrix components in primary fibroblasts. *Matrix Biol* 2017;62:105–114.
750 61. Ford CE, Jary E, Ma SSQ, Nixdorf S, Heinzelmann-Schwarz VA, Ward RL. The Wnt
751 Gatekeeper SFRP4 Modulates EMT, Cell Migration and Downstream Wnt Signalling in
752 Serous Ovarian Cancer Cells. *PLoS One* 2013;8:e54362.
753 62. Warrier S, Balu SK, Kumar AP, Millward M, Dharmarajan A. Wnt antagonist, secreted
754 frizzled-related protein 4 (sFRP4), increases chemotherapeutic response of glioma stem-like
755 cells. *Oncol Res* 2013;21:93–102.
756 63. Young M-J, Hsu K-C, Lin TE, Chang W-C, Hung J-J. The role of ubiquitin-specific
757 peptidases in cancer progression. *Journal of Biomedical Science* 2019;26:42.
758 64. Chen S, Liu Y, Zhou H. Advances in the Development Ubiquitin-Specific Peptidase (USP)
759 Inhibitors. *International Journal of Molecular Sciences* 2021;22:4546.
760 65. Zhang X-W, Feng N, Liu Y-C, Guo Q, Wang J-K, Bai Y-Z, *et al.* Neuroinflammation
761 inhibition by small-molecule targeting USP7 noncatalytic domain for neurodegenerative
762 disease therapy. *Science Advances* 2022;8:eabo0789.
763 66. Chen L, Xu Z, Li Q, Feng Q, Zheng C, Du Y, *et al.* USP28 facilitates pancreatic cancer
764 progression through activation of Wnt/β-catenin pathway via stabilising FOXM1. *Cell Death*
765 *Dis* 2021;12:1–12.

766 67. Sun X, Cai M, Wu L, Zhen X, Chen Y, Peng J, *et al.* Ubiquitin-specific protease 28
767 deubiquitinates TCF7L2 to govern the action of the Wnt signaling pathway in hepatic
768 carcinoma. *Cancer Sci* 2022;113:3463–3475.
769 68. Liu Z, Zhao T, Li Z, Sun K, Fu Y, Cheng T, *et al.* Discovery of [1,2,3]triazolo[4,5-
770 d]pyrimidine derivatives as highly potent, selective, and cellularly active USP28 inhibitors.
771 *Acta Pharmaceutica Sinica B* 2020;10:1476–1491.
772 69. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, *et al.* Protocol Update for Large-
773 scale genome and gene function analysis with PANTHER Classification System (v.14.0).
774 *Nat Protoc* 2019;14:703–721.
775 70. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic
776 data across different conditions, technologies, and species. *Nature Biotechnology*
777 2018;36:411–420.

778

779 **Figure Legends**

780 **Figure 1: The lung transcriptome is significantly altered in PAH.** (A) Unsupervised
781 hierarchical clustering of PHBI transcriptomes: 17,567 genes after filtering the bottom 25% of
782 genes with the least variation across samples. Samples are annotated by age, sex, race, and
783 diagnosis. Two outliers are circled. (B) PCA plot showing PCA of all 23,355 detected genes
784 where samples are colored by diagnosis. The same two outliers as in (A) are circled. (C) Volcano
785 plot showing upregulated genes colored in red and downregulated genes colored in green. Grey
786 dots indicate genes with FDR $\geq$ 0.05. (D) Bar plot showing GSEA results using the Hallmark
787 pathway database. Pathways enriched in genes upregulated in PAH are colored in red and

788 pathways enriched in downregulated genes in green. Only pathways with FDR < 0.05 are shown.
789 IPAH = idiopathic PAH; APAH = associated PAH; HPAH = hereditary PAH; PVOD =
790 pulmonary veno-occlusive disease; WHO4 = WHO group 4 PAH; FDR = false discovery rate;
791 log2FC = log2 fold change; NES = normalized enrichment score; EMT = epithelial-
792 mesenchymal transition.
793
794 **Figure 2: Co-expression network analysis reveals modules associated with PAH severity**
795 **and genetic variants.** (A) Schematic of integrative analytical strategy centered around co-
796 expression modules. (B) Gene clustering dendrogram as determined by WGCNA with color
797 module assignments shown at the bottom. (C) Bar plot showing number of genes in each
798 module. (D) Heatmap showing significant ($P < 0.05$) Pearson correlations of module eigengenes
799 with clinical and pathologic characteristics where red and blue indicate positive and negative
800 correlation, respectively. Larger size dots indicate stronger correlation. No. hospitalizations
801 indicates number of hospitalizations due to PAH between the time of diagnostic RHC and lung
802 transplantation. R heart failure signs indicate signs of right heart failure such as ascites or leg
803 swelling. Intima and intima plus media thickness were determined by morphometric analysis of
804 volume density of pulmonary arteries in histological lung sections. (E-F) Scatter plots where
805 each dot represents a gene in the (E) pink or (F) greenyellow module. Red and blue dots indicate
806 up- or downregulation in PAH lungs, respectively. Genes are plotted by the Pearson correlation
807 of their expression with PAH diagnosis (vs control) on the y-axis and the module eigengene on
808 the x-axis. (G) Dot plot showing the normalized enrichment score (NES) of modules for the PAH
809 lung differential transcriptome as determined by GSEA. Larger size dots indicate stronger FDR
810 value. (H-I) Dot plots showing enrichment of modules for PAH GWAS SNPs using two distinct

811 computational methods, (H) MAGMA, and (I) GSA-SNP2, across four independent PAH
812 GWAS cohorts. Vertical red dotted lines indicate significance threshold. SNPs were mapped to
813 genes by chromosomal proximity (within 20 kilobases from the 5' or 3' ends of a gene) and
814 genes were scored for association with PAH based on disease-SNP p-value associations from
815 GWAS summary statistics. Gene scores were then used in competitive gene-set analyses to
816 identify module enrichment for PAH common genetic variation. To aggregate genetic variants
817 into a gene score, the mean χ^2 statistic and the log-minimum GWAS p-value for all SNPs
818 localizing to a gene were used in MAGMA and GSA-SNP2, respectively. To determine
819 significance, MAGMA uses a linear mixed model whereas GSA-SNP2 uses a standard normal
820 distribution. Both methods adjust for gene size and gene density (the number of SNPs assigned
821 to a given gene). WGCNA = weighted gene co-expression network analysis; GWAS = genome-
822 wide association study; scRNAseq = single-cell RNA sequencing; Dx = diagnosis; PFT =
823 pulmonary function test; RHC = right heart catheterization; Histo = histology; DLCO = diffusing
824 capacity for carbon monoxide; FVC/DLCO = ratio of forced vital capacity to DLCO; mPAP =
825 mean pulmonary artery pressure; PVR = pulmonary vascular resistance; REVEAL = Registry to
826 Evaluate Early and Long-Term PAH Disease Management; cor = correlation; PAHB = US PAH
827 Biobank; PHAAR = French Pulmonary Hypertension Allele-Associated Risk; BHFPAH =
828 British Heart Foundation Pulmonary Arterial Hypertension; UK = UK National Institute for
829 Health Research BioResource (NIHRBR); MAGMA = Multi-marker Analysis of GenoMic
830 Annotation; FDR = false discovery rate.
831
832 **Figure 3: The pink module is co-regulated with known PAH genes and is enriched in Wnt**
833 **signaling and EMT pathways.** (A) Bayesian gene regulatory network constructed from 2,066

lung transcriptomes with incorporation of lung-specific expression quantitative trait loci (eQTL), and known transcription factor-target gene relationships. Nodes represent genes. Edges represent inferred gene-gene regulation. Node positions were determined by a prefuse force-directed algorithm. Genes from co-expression modules are shown projected onto this regulatory network with the different color nodes representing module membership. The largest modules with >3000 genes (turquoise, blue, and brown) are not shown to allow better visualization of other modules. (B) Pink subnetwork where pink genes and known PAH genes (red nodes) from disease-gene databases (Comparative Toxicogenomics Database(11) and DisGeNET(10)) were projected onto the lung Bayesian regulatory network in (A). BMPR2-GUCY1A2 pair is highlighted in the upper right where the arrow represents the predicted directional regulatory relationship. Larger size nodes represent hub genes where node size is proportional to -log10(FDR) as determined by Key Driver Analysis (4, 28, 29). Light grey nodes represent hub genes of the pink subnetwork that are neither pink nor red genes. (C-D) Box plots showing *GUCY1A2* expression two independent RNA-seq datasets: (C) CRISPR/Cas9-induced monoallelic mutations in BMPR2 (n = 6) vs wild-type control (n = 3) human umbilical vein endothelial cells (HUVEC)(30) and (D) endothelial cells (ECs) derived from induced pluripotent stem cells (iPSCs) of patients with hereditary PAH (HPAH) due to BMPR2 mutations (n = 5) vs control (n = 3)(31). *P* values were determined by DESeq2 for (C) and (D). (E-F) Dots plots showing Gene Set Enrichment Analysis (GSEA) of the pink module signature using (E) Gene Ontology (GO) and (F) Hallmark(13) gene sets where y-axis represents normalized enrichment scores (NES) in which scores greater than or less than zero represent gene sets enriched in genes positively or negatively correlated with the pink eigengene, respectively. The x-axis represents gene sets ordered by their enrichment score. Select top gene sets are labeled: Regulation of Wnt signaling (NES score 2.29, NES rank 4 of

6,033) in (E) and Epithelial Mesenchymal Transition (NES score 2.30, NES rank 1 of 50) in (F). Dots larger in size represent higher -log10(*P*) values. (G) Scatter plot showing pink genes where the x- and y-axes represent the absolute correlation of the pink gene with the pink eigengene and PAH diagnosis, respectively. Red and blue dots denote up- and downregulated genes in PAH lungs, respectively. EMT genes from Hallmark(13) and Wnt genes from PANTHER(69) are indicated by green and orange text, respectively, and blue text indicates genes in both gene sets. (H) Scatter plot showing expression of ANTXR1 (x-axis) vs. SFRP4 (y-axis) in PHBI lungs where yellow triangles and grey squares represent control and PAH lungs, respectively. * $P < 0.05$, ** $P < 0.01$. BMPR2mut = BMPR2 mutation; KO = knockout; rE2 = deletion in exon 2; CRISPR = clustered regularly interspaced short palindromic repeats.

Figure 4: Deconvolution reveals cell-type specificity in PAH lungs and modules. (A) Uniform Manifold Approximation and Projection plot showing seven publicly available human lung single-cell RNAseq datasets(34–40) integrated and reclustered in Seurat(70), totaling 559,511 cells from 154 lungs. (B) Heatmap showing scaled expression of cell-type specific marker genes on the x-axis across all cell types in (A) on the y-axis. Larger size dots indicate higher fraction of cells expressing a given gene. (C) Dot plot showing Pearson correlation of deconvoluted cell type fractions with PAH vs control status across PHBI lung samples. Only correlations with $P < 0.05$ are shown. (D) Box plots of deconvoluted vascular cell fractions in PAH vs control lungs. (E) Heatmap showing Pearson correlation of deconvoluted cell fractions (columns) and module eigengenes (rows). Only correlations with $P < 0.05$ are shown. Larger size dots indicate higher absolute correlation. (F) Dot plot showing the normalized enrichment score of the pink module signature (defined as the correlation of the pink eigengene with the

880 expression of genes across the transcriptome) for known cell-type signatures from Azimuth(15)
881 as determined by GSEA. Larger size dots indicate stronger *P* value. (G) Box plot showing
882 averaged *ANTXR1* expression in fibroblasts profiled by scRNAseq from 3 PAH vs 6 control
883 lungs (32). Note, myofibroblasts were not subclustered in this dataset. (H) Box plot showing
884 *ANTXR1* expression by qPCR of pulmonary arterial adventitial cells isolated from PAH vs.
885 control lungs (n = 4 biological replicates/group). Wilcoxon rank-sum test: * $P < 0.05$, ** $P <$
886 0.01, *** $P < 0.001$, **** $P < 0.0001$. AbbBasaloid = aberrant basaloid; AlvProgen = alveolar
887 progenitor; AM = alveolar macrophage; AT1 = alveolar type 1; AT2 = alveolar type 2, cDC =
888 conventional dendritic cell; cMono = classical monocyte; EndoArt = endothelial arterial;
889 EndoBronch = endothelial bronchial; EndoCap1 = endothelial capillary 1; EndoCap2 =
890 endothelial capillary 2; EndoLymph = endothelial lymphatic; EndoVein = endothelial vein; Fb =
891 fibroblast; IM = interstitial macrophage; MacProlif = macrophage proliferating; MyoFb =
892 myofibroblast; ncMono = nonclassical monocyte; NK = natural killer; pDC = plasmacytoid
893 dendritic cell; PNEC = pulmonary neuroendocrine cell; SMC = smooth muscle cell; Tcd8 =
894 $CD8^+$ T cell; Tprolif = proliferating T cell; Treg = regulatory T cell; scRNAseq = single-cell
895 RNA sequencing; qPCR = quantitative polymerase chain reaction.
896
897 **Figure 5: Pharmacotranscriptomics identifies novel therapeutic targets.** (A) Dot plot
898 showing CMap score of 171 CMap classes which are groups of pharmacologic or genetic
899 perturbagens that share the same mechanism of action or biological function. Scores approaching
900 100 or -100 indicate perturbagens predicted to mimic or reverse the query signature,
901 respectively. Larger size dots indicate higher absolute score. Color indicates module
902 membership. Table on the right shows the score of select CMap classes with the rank out of 171

classes shown in parenthesis. (B) Dot plot showing CMap score of 8,559 CMap perturbagens. Table on the right shows CMap scores of USP shRNA targets. (C) Volcano plot of top 100 mimickers and top 100 reversers of the pink PAH DEG signature out of a total of 7,502 genetic targets screened from the CRISPR KO consensus signature database of LINCS L1000 using SigCom(44). Z score on the x-axis indicates the degree to which the target signature mimics or reverses the pink signature. Select top mimickers and reversers that are shown in a heatmap in (D) are labeled by purple and orange, respectively. (D) Heatmap showing the relative expression of select pink lung DEGs (rows) in the signatures of select top CRISPR KO targets (columns) that are mimickers or reversers of the pink signature. Lowly ranked genes are downregulated while highly ranked genes are upregulated in the CRISPR KO vs. control. Rows are annotated on the left by red or blue to indicate upregulated or downregulated in PAH vs. control lungs, respectively. (E) Volcano plot showing upregulated and downregulated USPs in PAH lungs in red and blue, respectively. Grey dots indicate genes with FDR ≥ 0.05. USPs that were both upregulated and top hits in our CMap or CRISPR analyses are boxed. (F-G) Box plots showing mRNA expression of (F) USP28 and (G) USP12 in a human whole lung microarray of 15 PAH patients vs 11 controls(33). *P* values were obtained from NCBI's GEO2R: ***$P < 0.001$, **** $P < 0.0001$. CMap = Connectivity Map; USP = ubiquitin specific peptidase; JAK inh. = Janus kinase inhibitor; CDK inh. = cyclin-dependent kinase inhibitor; shRNA = short hairpin RNA.

Table 1: Patient characteristics.

	PH (n=96)	**Control (n=52)**
Age, years	39.4±16.1	40.6±16.9
Sex, female	73 (76)	18 (35)
PH classification		
Idiopathic PAH	41 (43)	

Associated PAH	38 (40)
CTD	11 (11)
CHD	19 (20)
Drug/toxin	8 (8)
Hereditary PAH	8 (8)
PVOD	7 (7)
WHO group 4	2 (2)
Weeks since diagnosis	415±331
No. hospitalizations	4.5±3.4
R heart failure signs	33 (34)
WHO functional class	
II	8 (8)
III	42 (44)
IV	39 (41)
PFT	
FVC, %	79.4±14.2
DLCO, %	64.2±25.4
RHC	
RA, mm Hg	11.5±6.2
mPAP, mm Hg	60.5±18.1
PCWP, mm Hg	12.7±6.0
PVR, Wood units	12.7±7.4
CO, liters/min	4.2±1.5
CI, liters/min/m^2	2.6±0.8
NT-proBNP, pg/ml	1993±2897
Creatinine	1.0±0.8
PH therapy	
Prostacyclin infusion	82 (85)
Monotherapy	6 (6)
Double therapy	20 (21)
Triple therapy	70 (73)

Available data are presented in numbers (% total), or mean +/- standard deviation

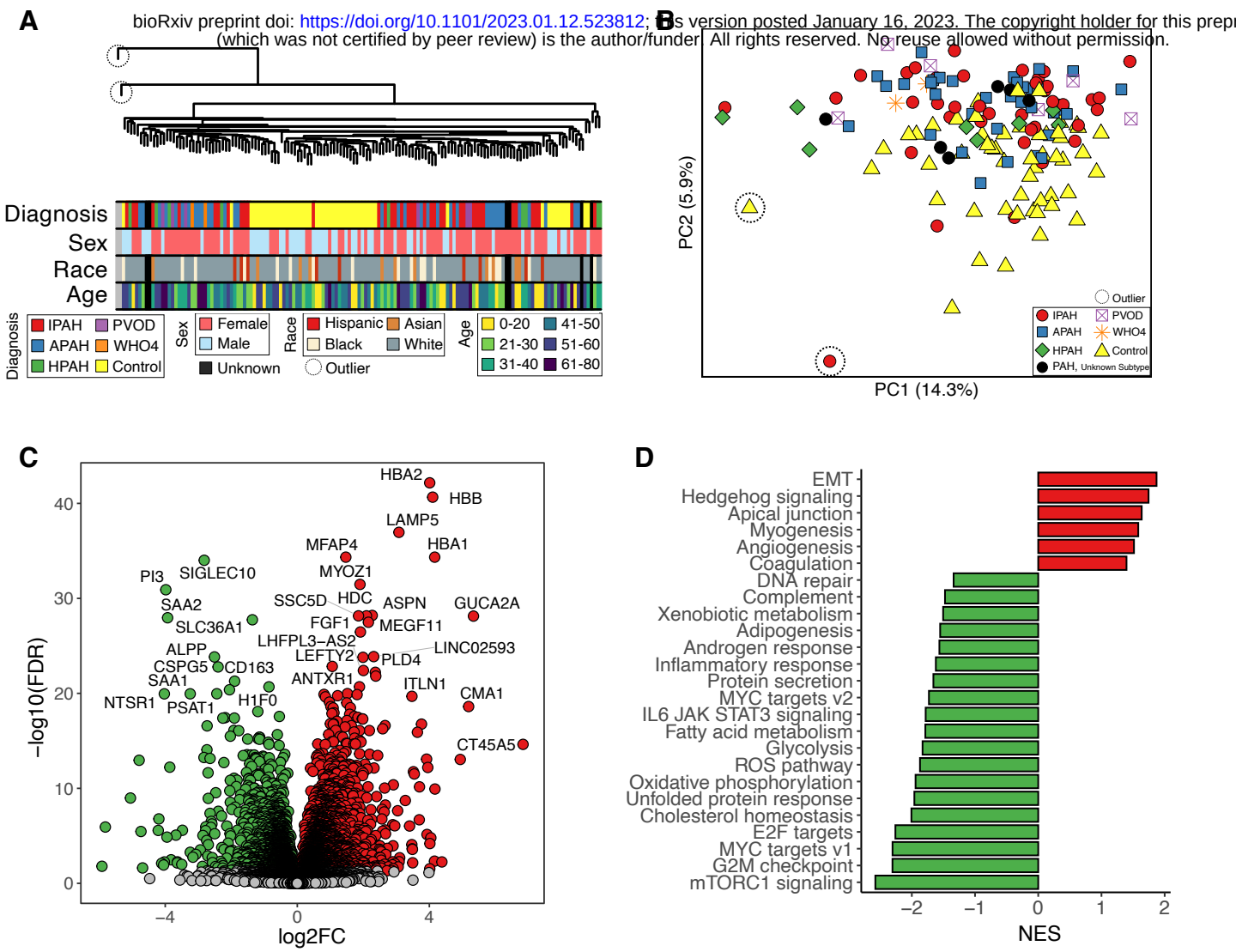


Figure 1

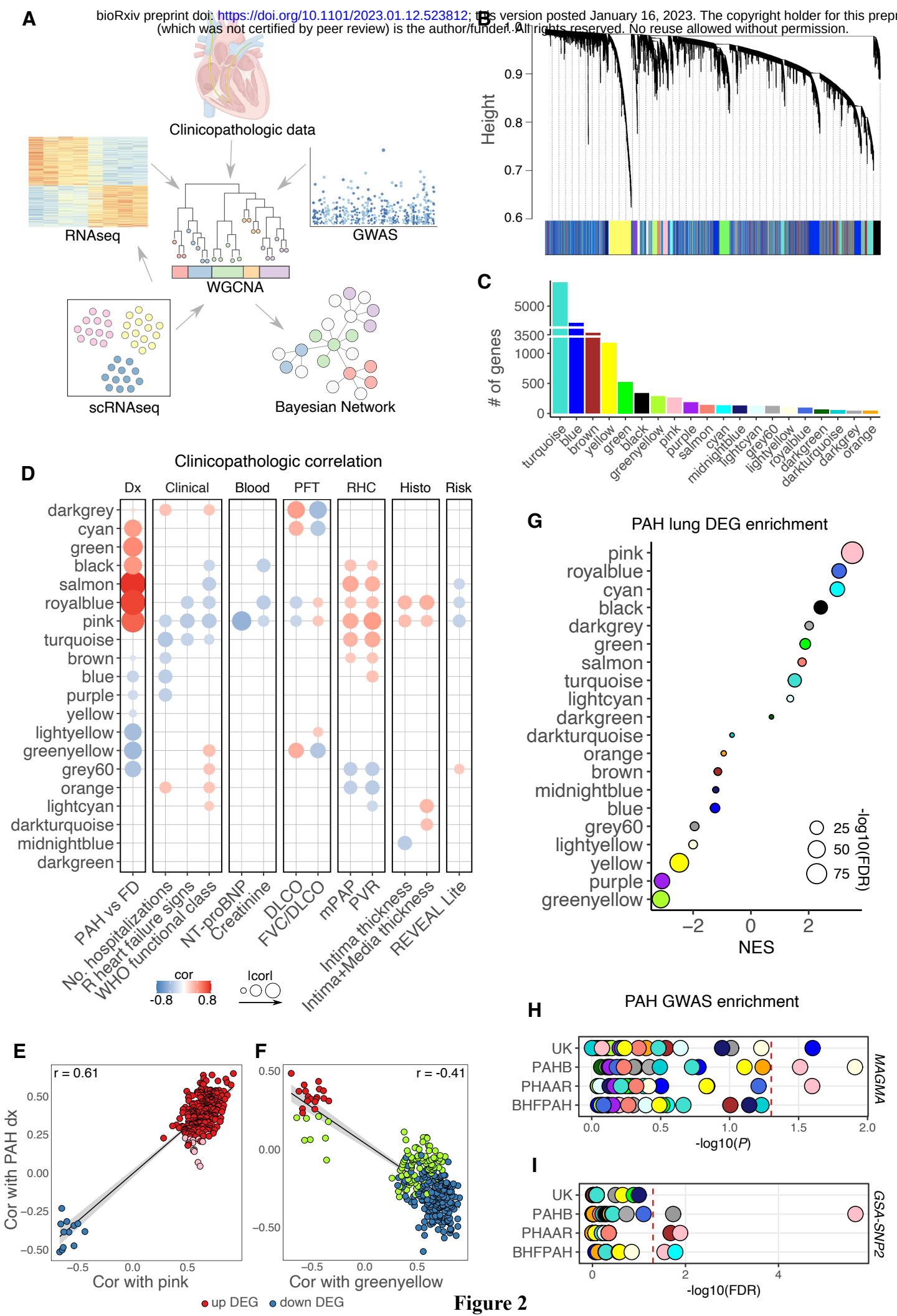


Figure 2

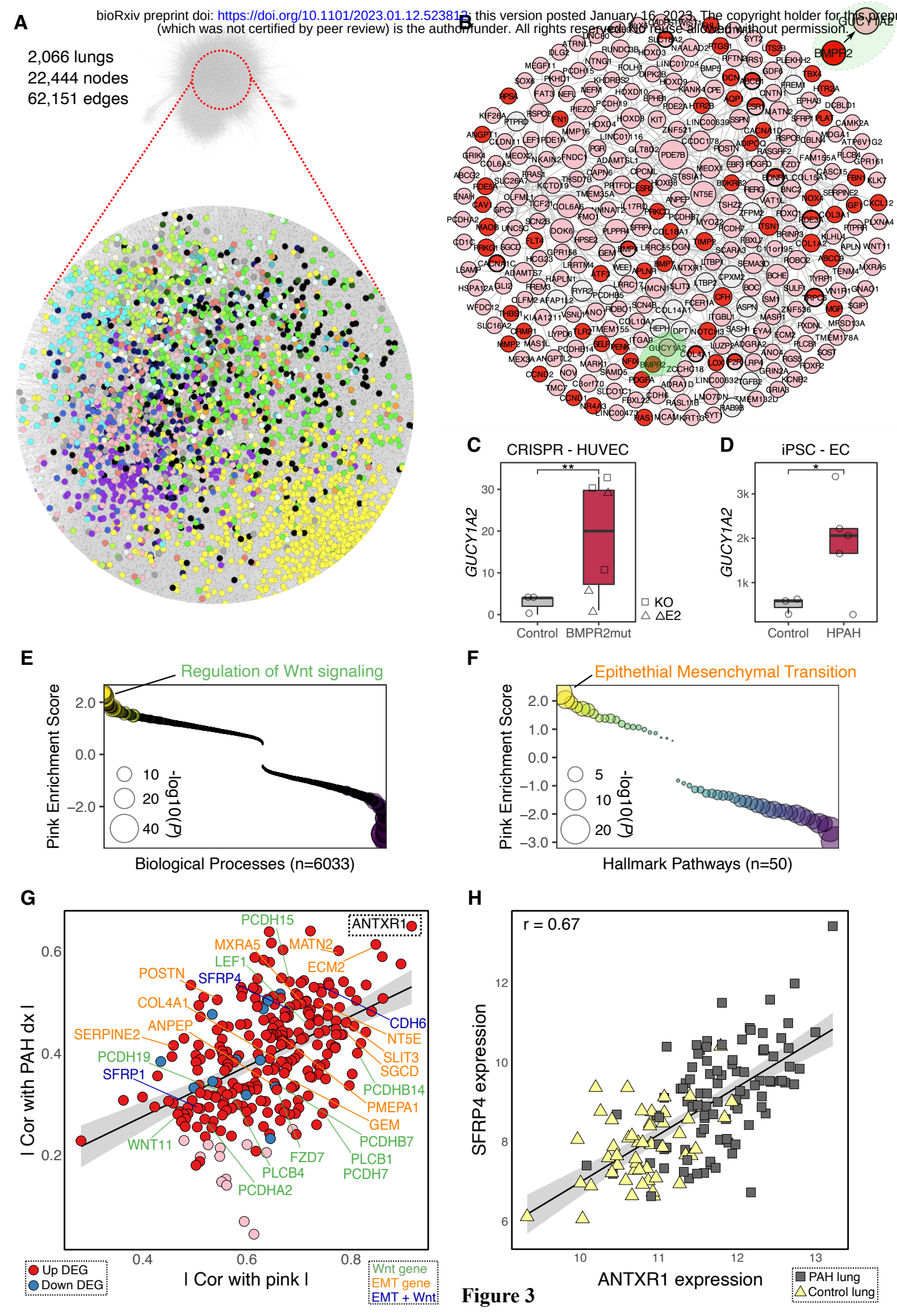


Figure 3

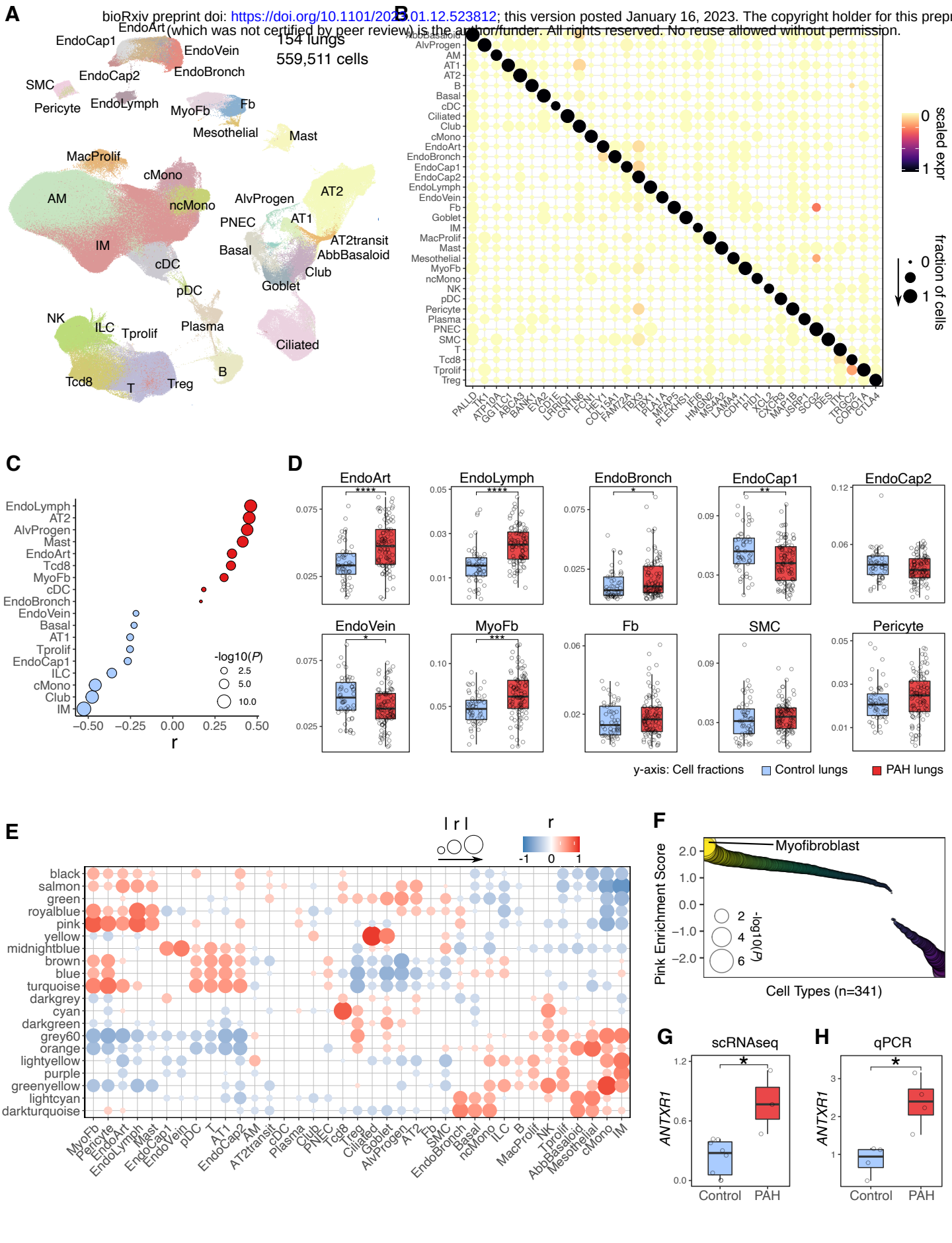
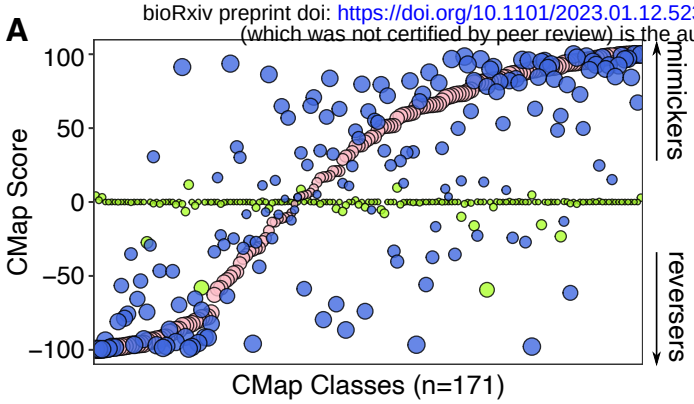
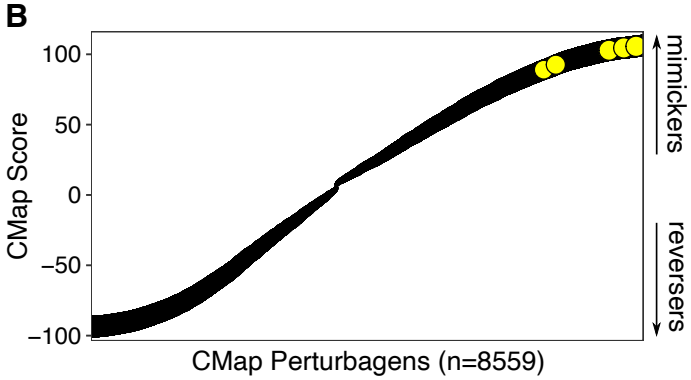


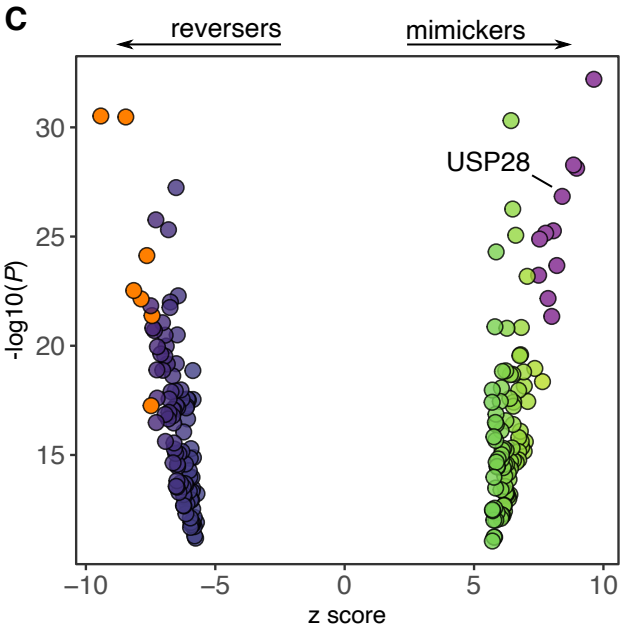
Figure 4

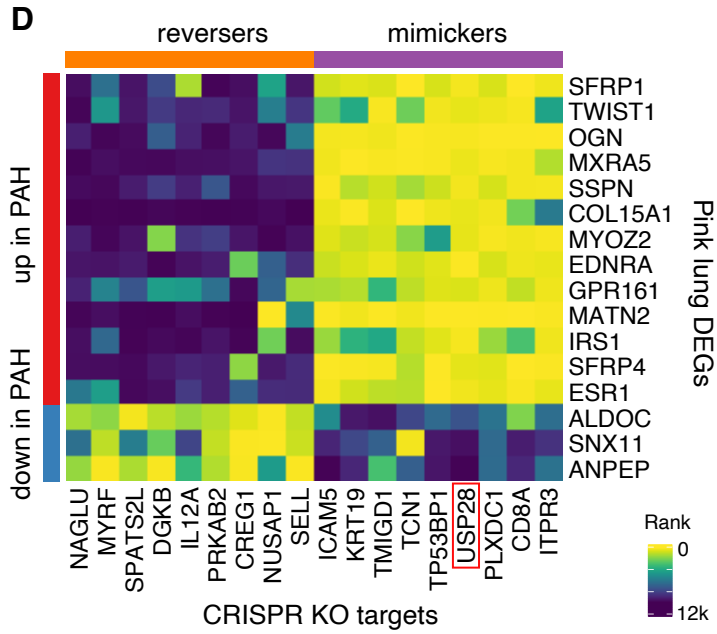


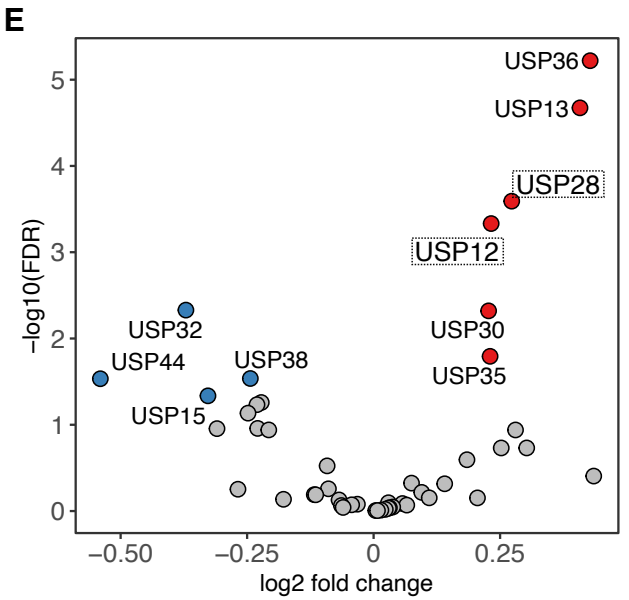
Class	pink	royalblue
USP LOF	99.84 (1)	99.92 (2)
JAK inh.	99.75 (3)	99.93 (1)
CDK inh.	99.62 (5)	96.95 (9)

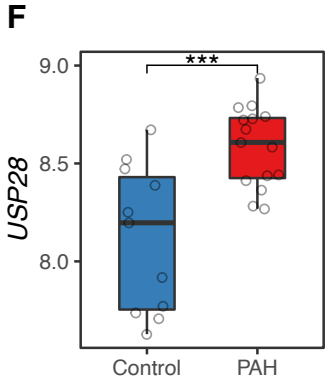


shRNA target	pink	royalblue
USP7	99.45	99.02
USP22	98.32	13.18
USP12	96.67	98.55
USP20	96.55	83.25
USP1	86.38	59.84
USP15	82.83	91.22









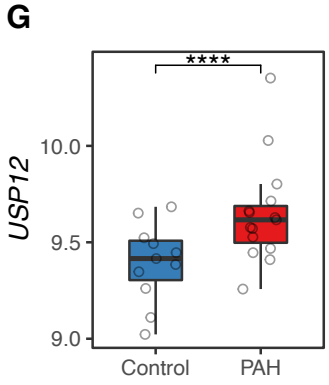


Figure 5

Integrative Multiomics to Dissect the Lung Transcriptional Landscape of Pulmonary Arterial Hypertension.

Jason Hong, Brenda Wong, Christopher J. Rhodes, Zeyneb Kurt, Tae-Hwi Schwantes-An, Elizabeth A. Mickler, Stefan Gräf, Mélanie Eyries, Katie A. Lutz, Michael W. Pauciulo, Richard C. Trembath, David Montani, Nicholas W. Morrell, Martin R. Wilkins, William C. Nichols, David-Alexandre Trégouët, Micheala A. Aldred, Ankit A. Desai, Rubin M. Tuder, Mark W. Geraci, Mansoureh Eghbali, Robert S. Stearman, Xia Yang

Supplemental Figure Legends

Supplemental Figure 1: Lung transcriptomes do not cluster by transplant center nor treatment group. (A-B) PCA plots showing PCA of all 23,355 detected genes where samples are colored by (A) transplant center of tissue origin and (B) treatment group. Control samples are not shown in (B). Treatment groups refer to one or a combination of the three major classes of PAH-targeted drugs: phosphodiesterase type 5 inhibitors, endothelin receptor antagonists, and prostacyclin analogues. AH = Allegheny Hospital; BA = Baylor; CC= Cleveland Clinic, DU=Duke University; ST=Stanford; UA=University of Alabama; UC=University of California, San Diego; UM=University of Michigan; VA= Vanderbilt.

Supplemental Figure 2: Modules correlated with clinical traits after adjustment for multiple testing. Heatmap showing Pearson correlations of module eigengenes with clinical and pathologic characteristics that met FDR threshold of < 0.05 after adjustment for multiple

comparisons (n=260 comparisons from 13 traits and 20 modules). Red and blue dots indicate positive and negative correlation, respectively. Larger size dots indicate stronger correlation. No. hospitalizations indicates number of hospitalizations due to PAH between the time of diagnostic RHC and lung transplantation. R heart failure signs indicate signs of right heart failure such as ascites or leg swelling. Intima and intima plus media thickness were determined by morphometric analysis of volume density of pulmonary arteries in histological lung sections. Dx = diagnosis; PFT = pulmonary function test; RHC = right heart catheterization; Histo = histology; DLCO = diffusing capacity for carbon monoxide; FVC/DLCO = ratio of forced vital capacity to DLCO; mPAP = mean pulmonary artery pressure; PVR = pulmonary vascular resistance; REVEAL = Registry to Evaluate Early and Long-Term PAH Disease Management; cor = correlation.

Supplemental Figure 3: Dysregulated genes of royalblue, greenyellow and pink modules. (A-C) Volcano plots showing dysregulated genes of (A) royalblue, (B) greenyellow, and (C) pink where red, green, and grey indicate upregulation, downregulation, or no change in PAH, respectively. The top 50 (A-B) or 100 genes (C) by the absolute Wald statistic as determined by DESeq2 are labeled. FDR = false discovery rate; log2FC = log2 fold change.

Supplemental Figure 4: Analysis of MAGMA and GSA-SNP2 gene scores. (A) Scatter plot showing gene scores averaged across four PAH GWAS cohorts as determined by MAGMA on the x-axis and GSA-SNP2 on the y-axis. To aggregate genetic variants into a gene score, the mean χ^2 statistic and the log-minimum GWAS p-value for all SNPs localizing to a gene were used in MAGMA and GSA-SNP2, respectively. (B-C) Scatter plots showing gene scores from

(B) GSA-SNP2 and (C) MAGMA plotted against gene length in base pairs. (D-E) Scatter plots showing gene scores from (D) GSA-SNP2 and (E) MAGMA plotted against the average SNP counts across four PAH GWAS cohorts. SNP counts represent the number of SNPs localizing to a given gene within 20 kilobases from the 5' or 3' ends of the gene. (A-E) Colors represent module membership.

Supplemental Figure 5: Bayesian network analysis workflow to construct a gene regulatory network of the human lung. Bayesian networks (BNs) were constructed using Reconstructing Integrative Molecular Bayesian Network (RIMBANet)[19]. For this method, 1000 networks were generated from different random seed genes using continuous and discrete expression data derived from transcriptomes from either GSE23546 (n = 1343) (1), PHBI (n = 146), or GTEx v8 (n = 577) (1). Whole lung-specific cis eQTLs from GTEx v8 (1) and transcription factor-target gene data from HTRI (1), TRRUST (1), and PAZAR (1) databases were used as priors. Then, the final network for each of the 3 datasets was obtained by taking a consensus network from the 1000 randomly generated networks whereby only edges that passed a probability of >30% across the 1000 BNs were kept. Finally, the union of the 3 networks was taken to create a combined gene regulatory network derived from a total of 2,066 human lungs.

Supplemental Figure 6: *GUCY1A2* is upregulated in PAH lungs. Box plot showing *GUCY1A2* expression in WHO Group 1 PAH (n = 93) vs control (n = 51) lungs from PHBI. * FDR < 0.05.

Supplemental Figure 7: Single-cell RNA sequencing and deconvolution analysis scheme. To serve as a cell type reference for deconvolution, we integrated seven publicly available human lung single-cell RNAseq datasets(29–35) and identified 37 cell-type clusters using known marker genes from the literature. Within each cell-type cluster, the average expression of gene counts were calculated across cells for each individual sample to create a cell-type signature for each of the seven datasets. PHBI bulk transcriptomes were deconvoluted with CIBERSORTx(36) with cell-type signatures from each of the seven datasets as a reference. The resulting cell fractions using each of the seven dataset-specific reference signatures served as technical replicates. These technical replicates were then averaged to determine the final estimated cell fractions for each lung sample.

Supplemental Figure 8: PAH lung samples cluster together based on estimated cell fractions. (A) Heatmap showing cell fractions estimated by deconvolution of PHBI lung transcriptomes by CIBERSORTx(1). Dendrograms are shown on the left and top representing hierarchical clustering of cell types (rows) and lung samples (columns), respectively. Lung samples are annotated at the bottom to indicate PAH in red or control in grey. (B) PCA plot showing PCA of estimated cell fractions with samples colored to indicate PAH in red or control in grey.

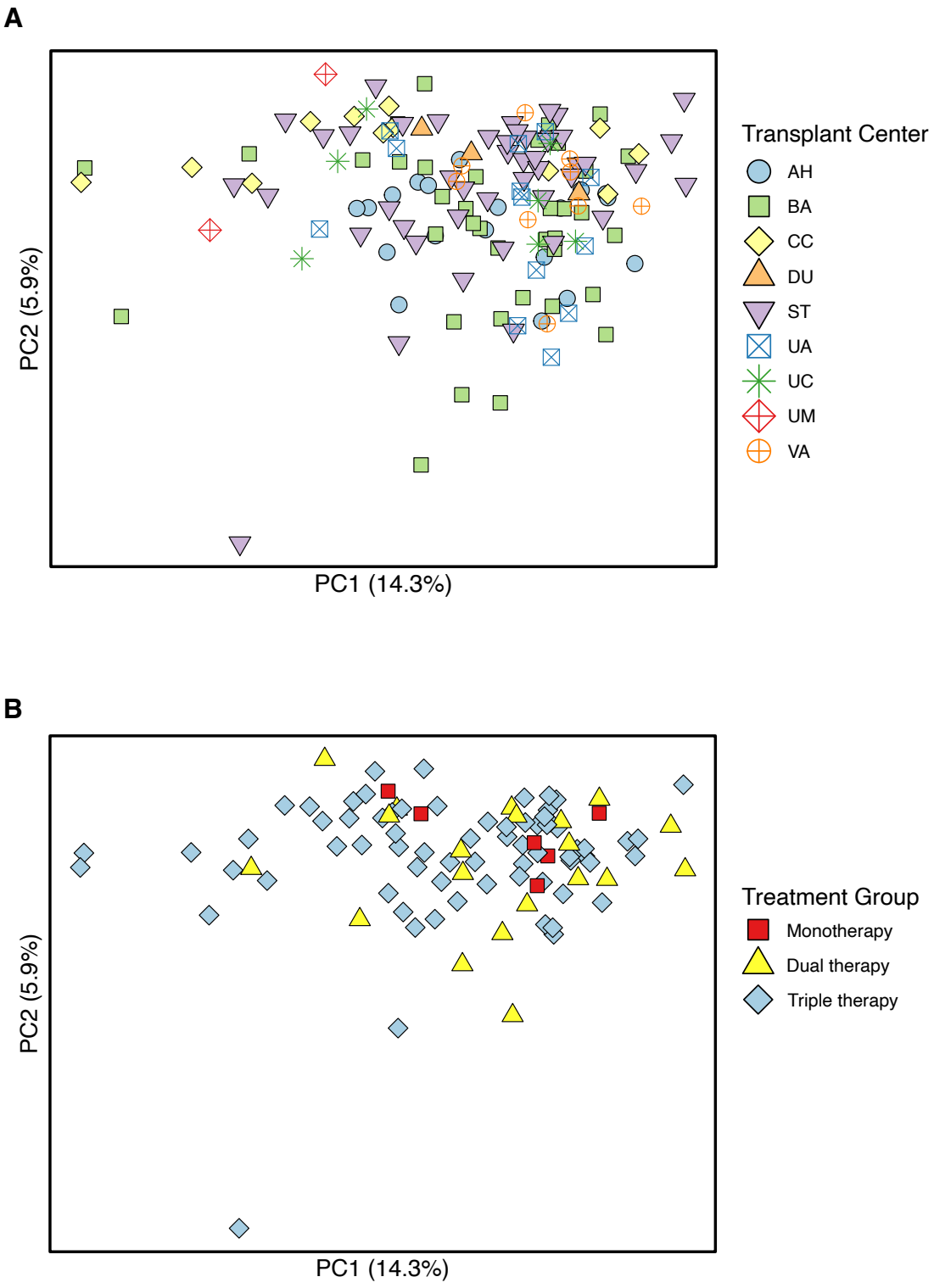
Supplemental Figure 9: Royalblue genes share similar pathways with and are hub nodes of the pink module. (A-B) Dots plots showing Gene Set Enrichment Analysis (GSEA) of the royalblue module signature using (A) Gene Ontology (GO) and (B) Hallmark(1) gene sets where y-axis represents normalized enrichment scores (NES) in which scores greater than or less than

zero represent gene sets enriched in genes positively or negatively correlated with the royalblue eigengene, respectively. The x-axis represents gene sets ordered by their enrichment score. Select top gene sets are labeled: Regulation of Wnt signaling (NES score 2.15, NES rank 14 of 6,033) in (A) and Epithelial Mesenchymal Transition (NES score 2.39, NES rank 1 of 50) in (B). Dots larger in size represent higher -log10(*P*) values. (C) Pink subnetwork where pink genes, royalblue genes, and known PAH genes (red nodes) from disease-gene databases (Comparative Toxicogenomics Database(1) and DisGeNET(1)) were projected onto the lung Bayesian regulatory network in Figure 3A. Larger size nodes represent hub genes where node size is proportional to -log10(FDR) as determined by Key Driver Analysis (1–3). Light grey nodes represent hub genes of the pink subnetwork that are not pink, royalblue, or red genes. Note this is the same subnetwork as in Figure 3B but with royalblue genes displayed here.

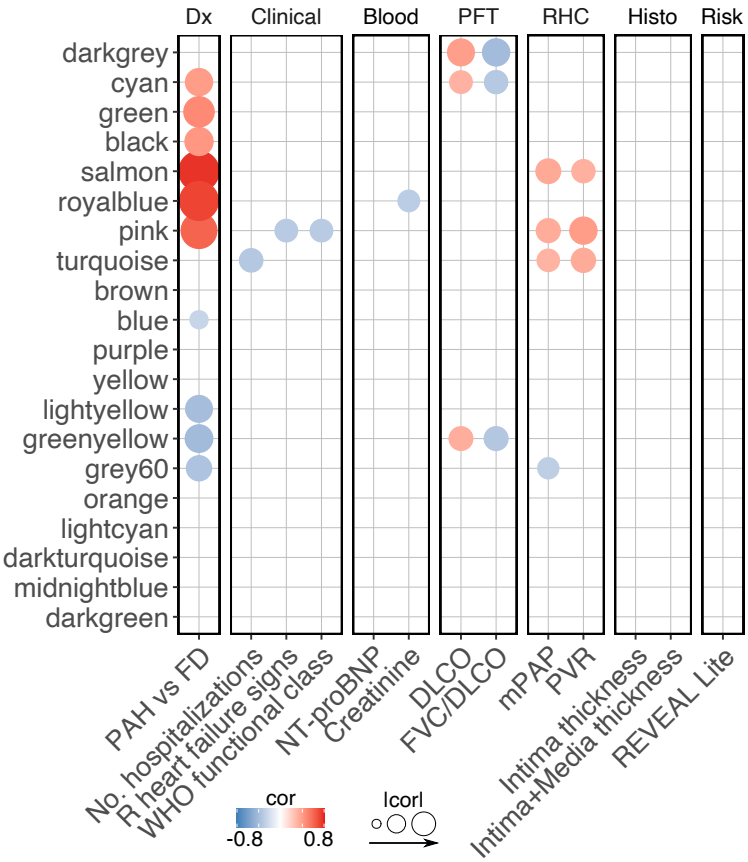
Supplemental Figure 10: Sex-stratified analysis of dysregulated genes, pathways and modules. (A-B) Volcano plots showing upregulated genes in PAH lungs colored in red and downregulated genes colored in green among (A) females (71 PAH vs. 17 control) and (B) males (22 PAH vs. 34 control). Grey dots indicate genes with FDR ≥ 0.05. Select genes are labeled. (C-D) Bar plots showing GSEA results using the Hallmark pathway database and the DEG signature of PAH vs. control among (C) females and (D) males. Pathways enriched in genes upregulated in PAH with normalized enrichment score (NES) > 0 are colored in red and pathways enriched in downregulated genes with NES < 0 in green. Only pathways with FDR < 0.05 are shown. (E-F) Dot plots showing enrichment of modules as determined by GSEA for the PAH lung differential transcriptome among (E) females and (F) males. Larger size dots indicate stronger FDR value.

114 log2FC = log2 fold change; NES = normalized enrichment score; EMT = epithelial-

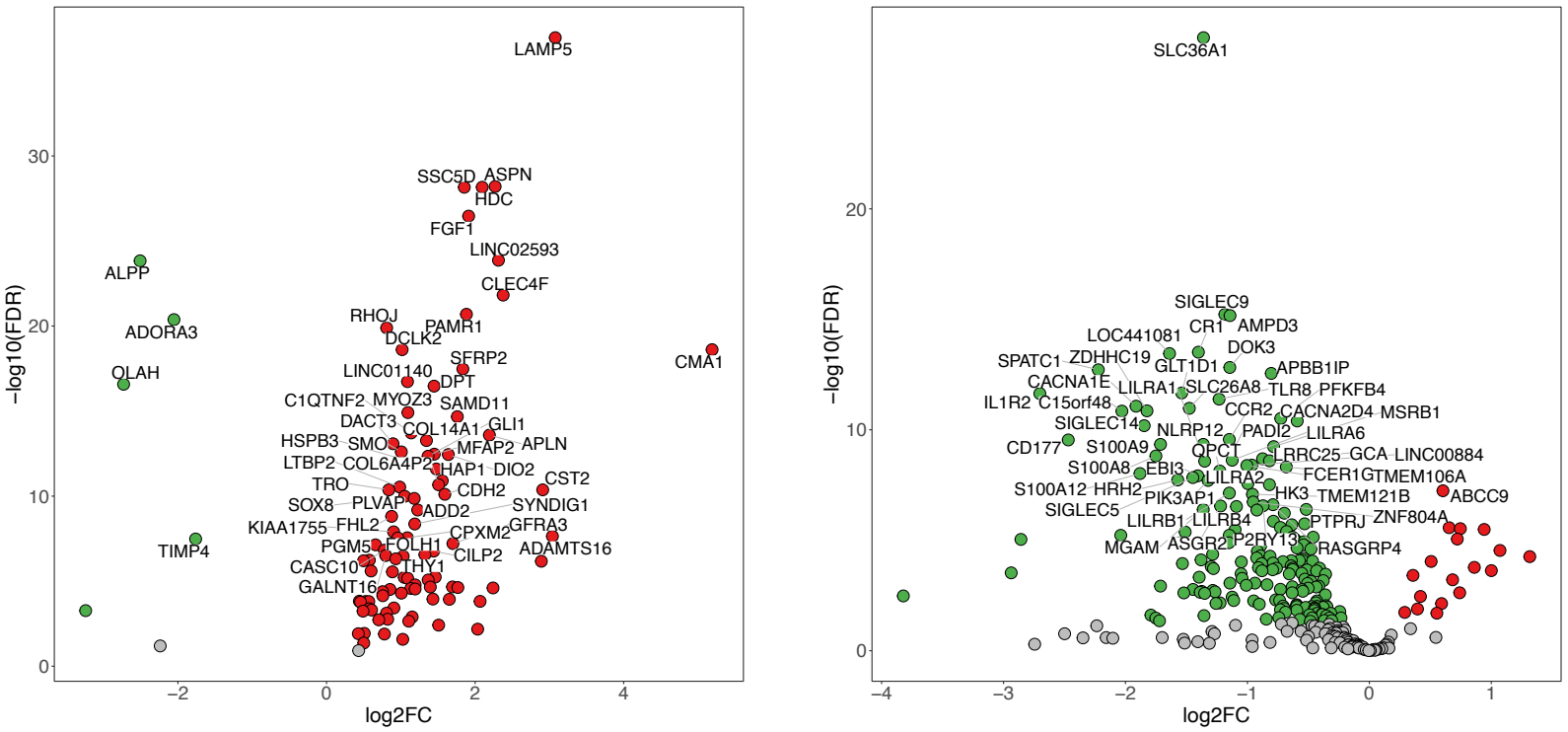
115 mesenchymal transition.

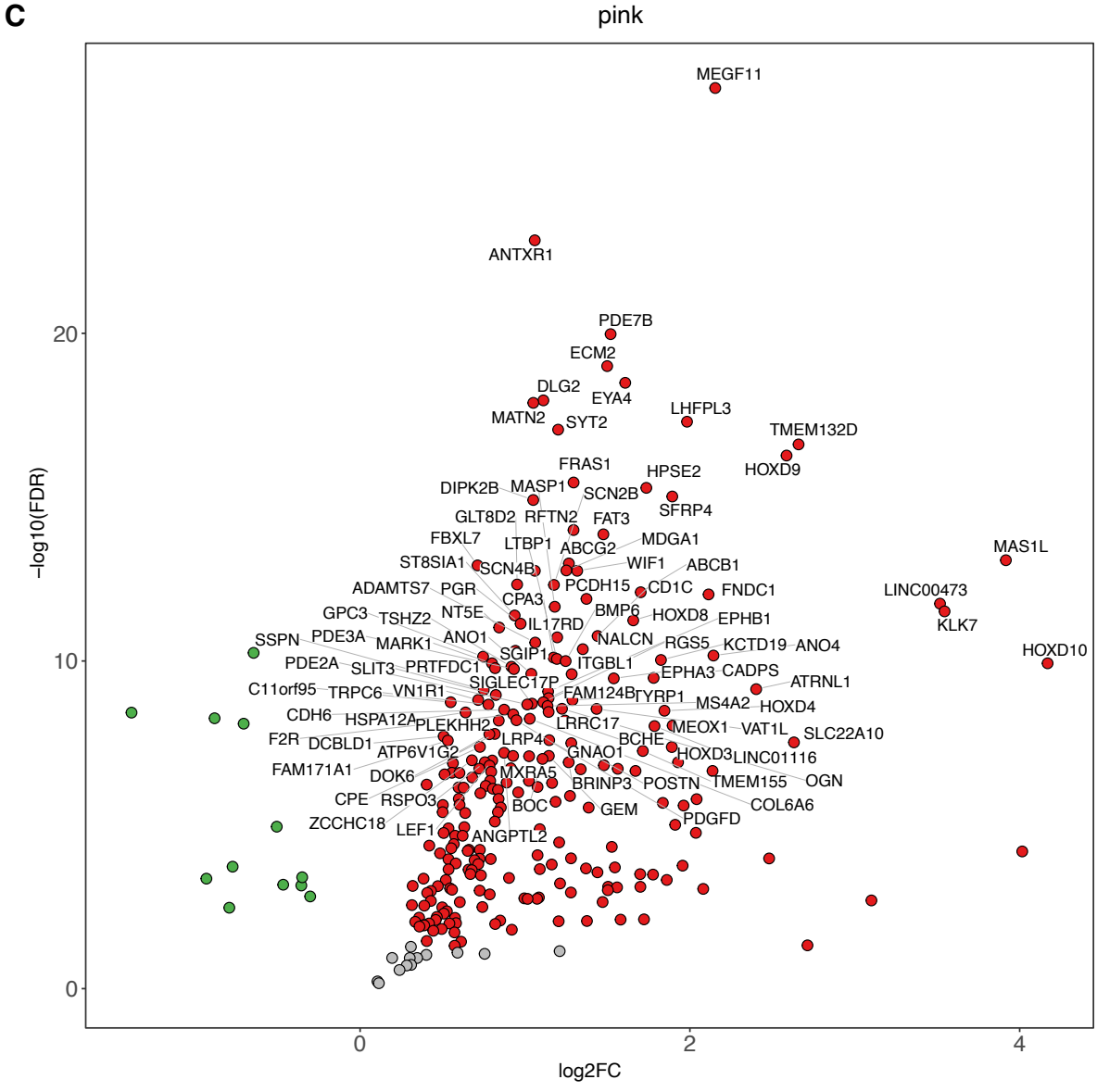


Supplemental Figure 1

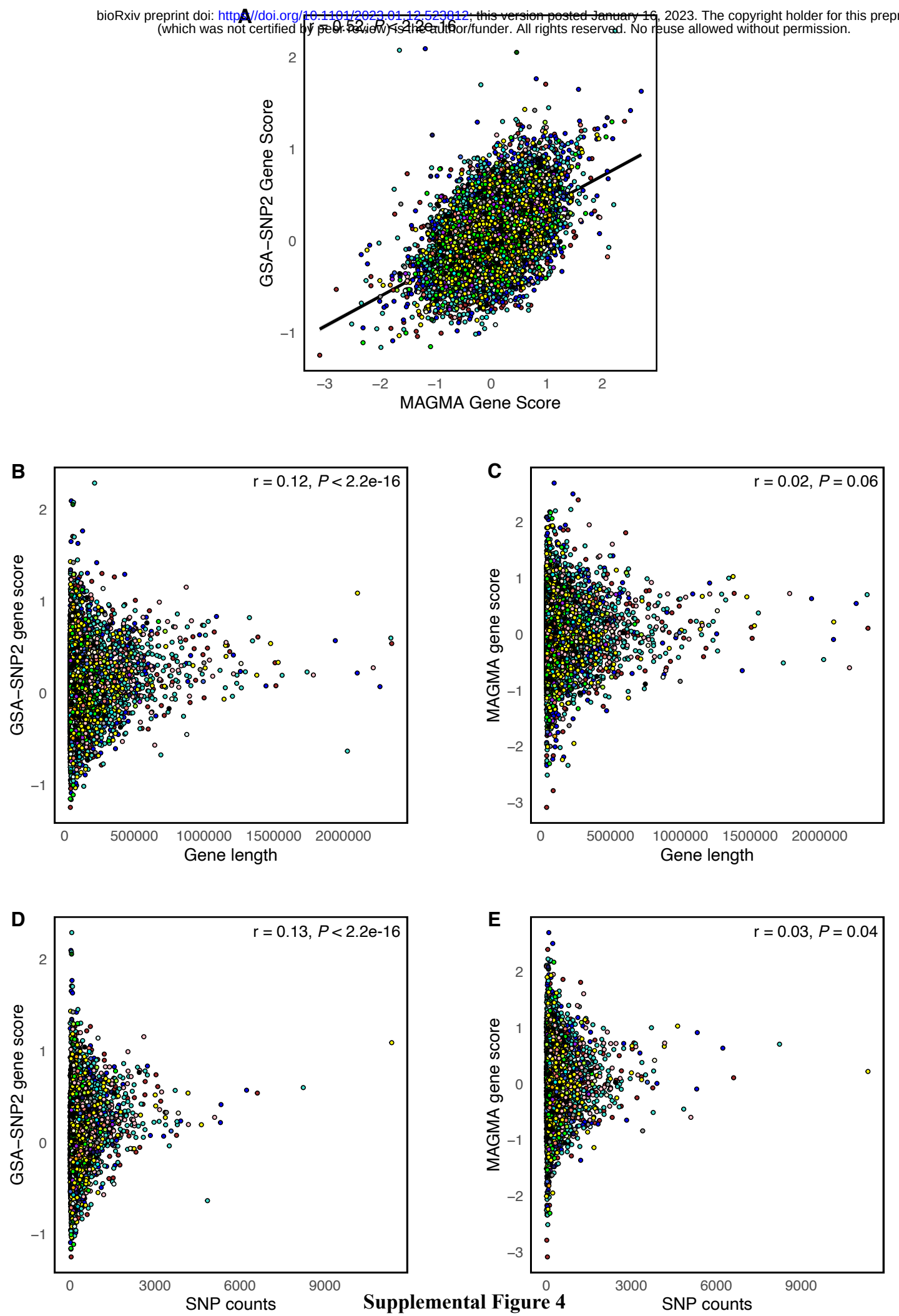


Supplemental Figure 2

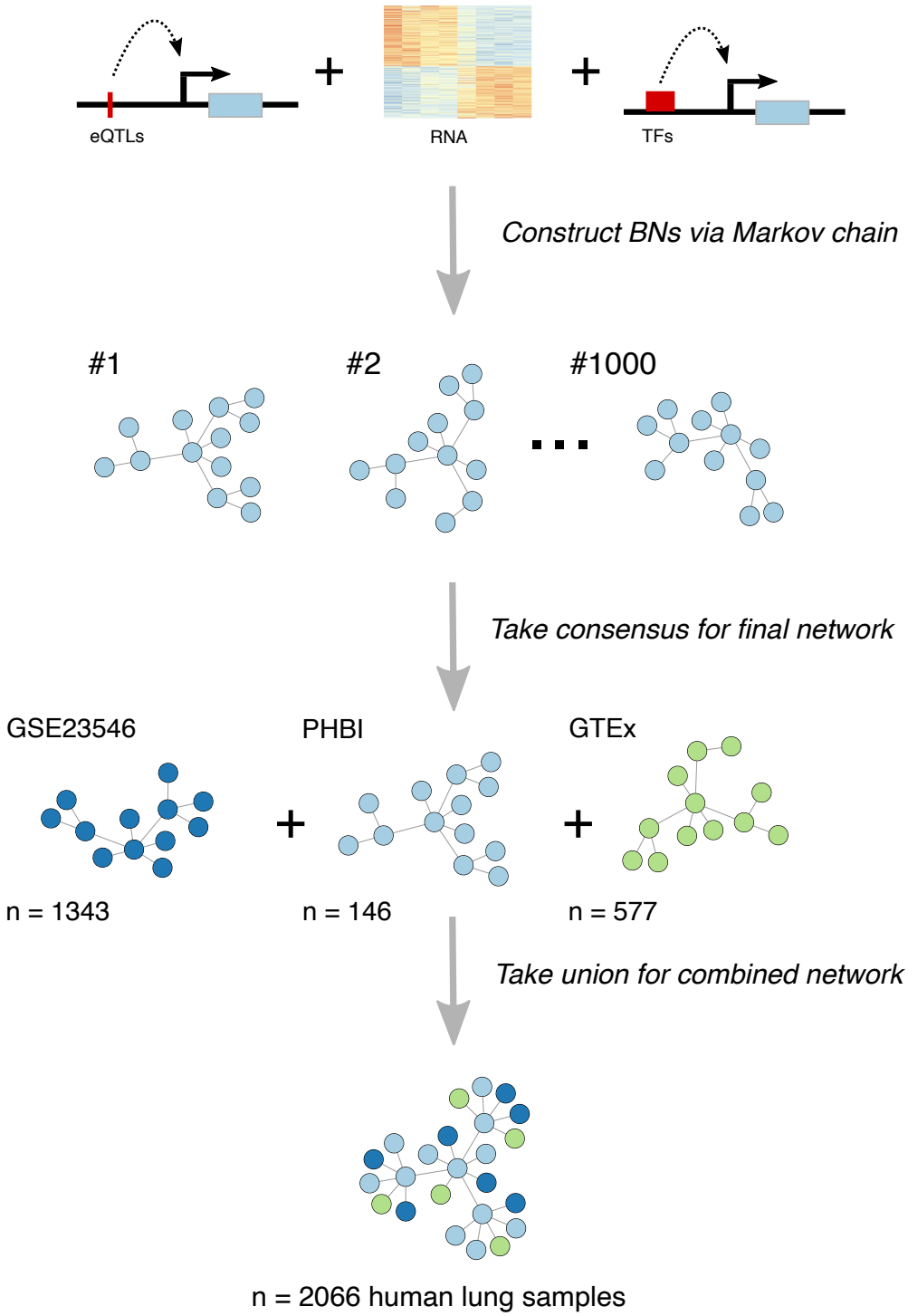




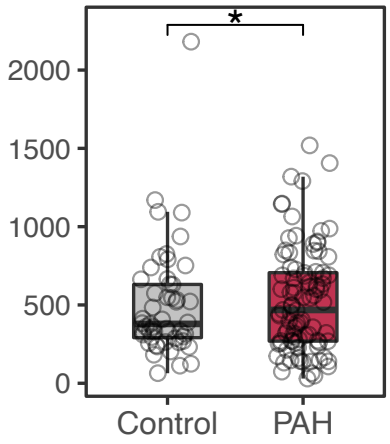
Supplemental Figure 3



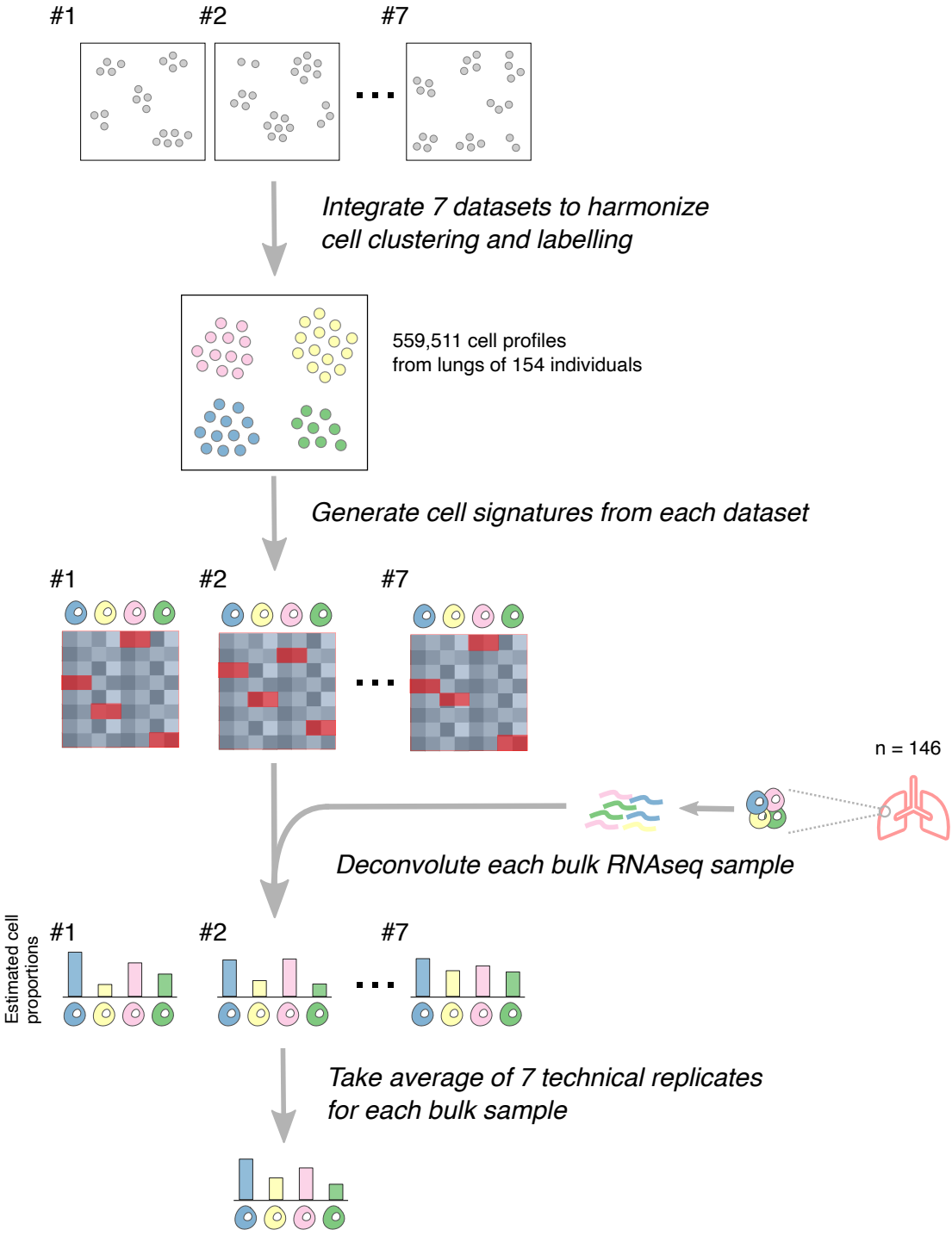
Supplemental Figure 4



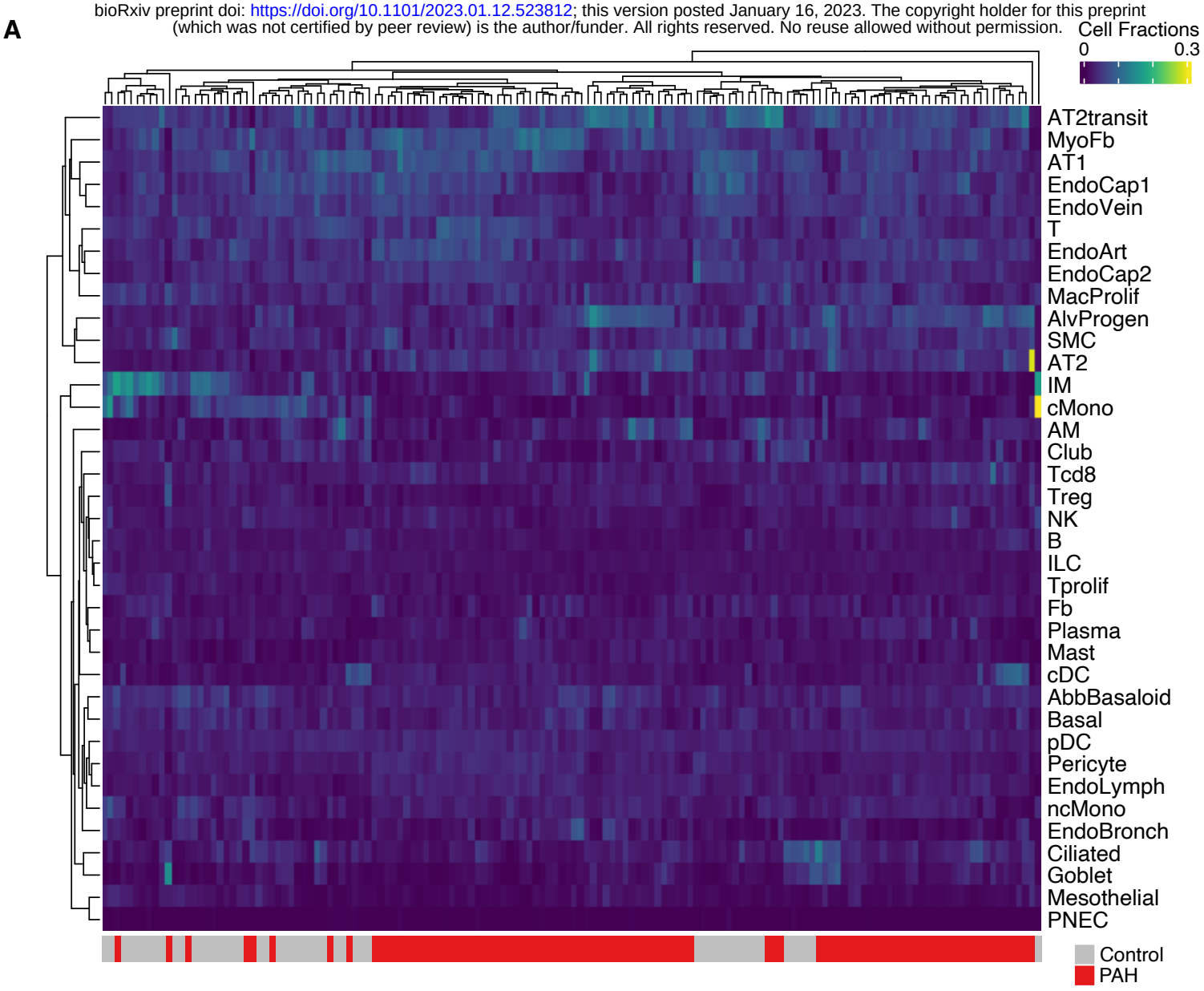
Supplemental Figure 5

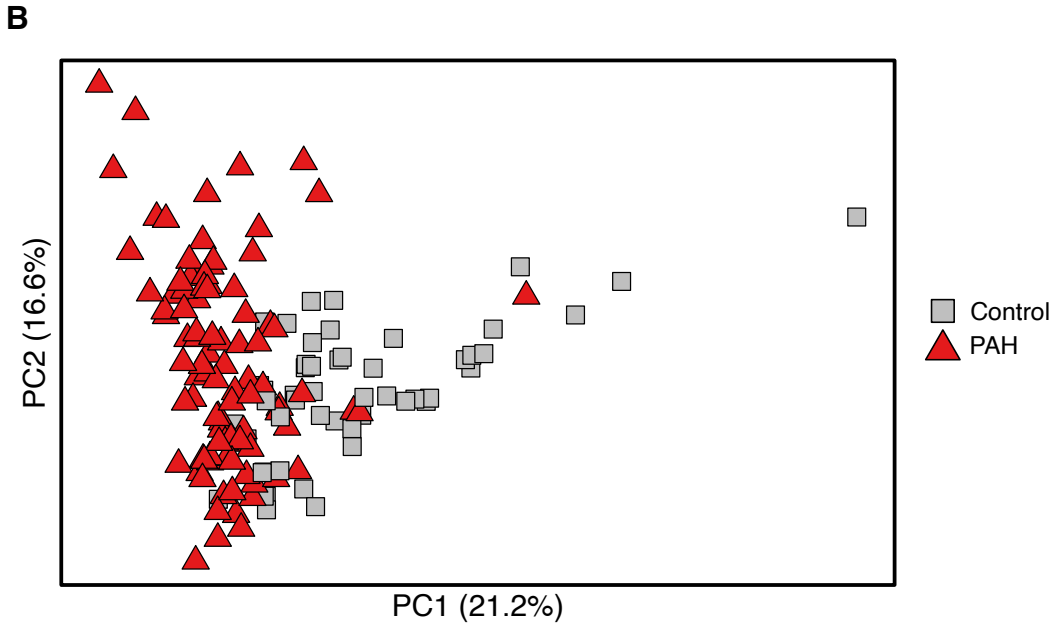


Supplemental Figure 6

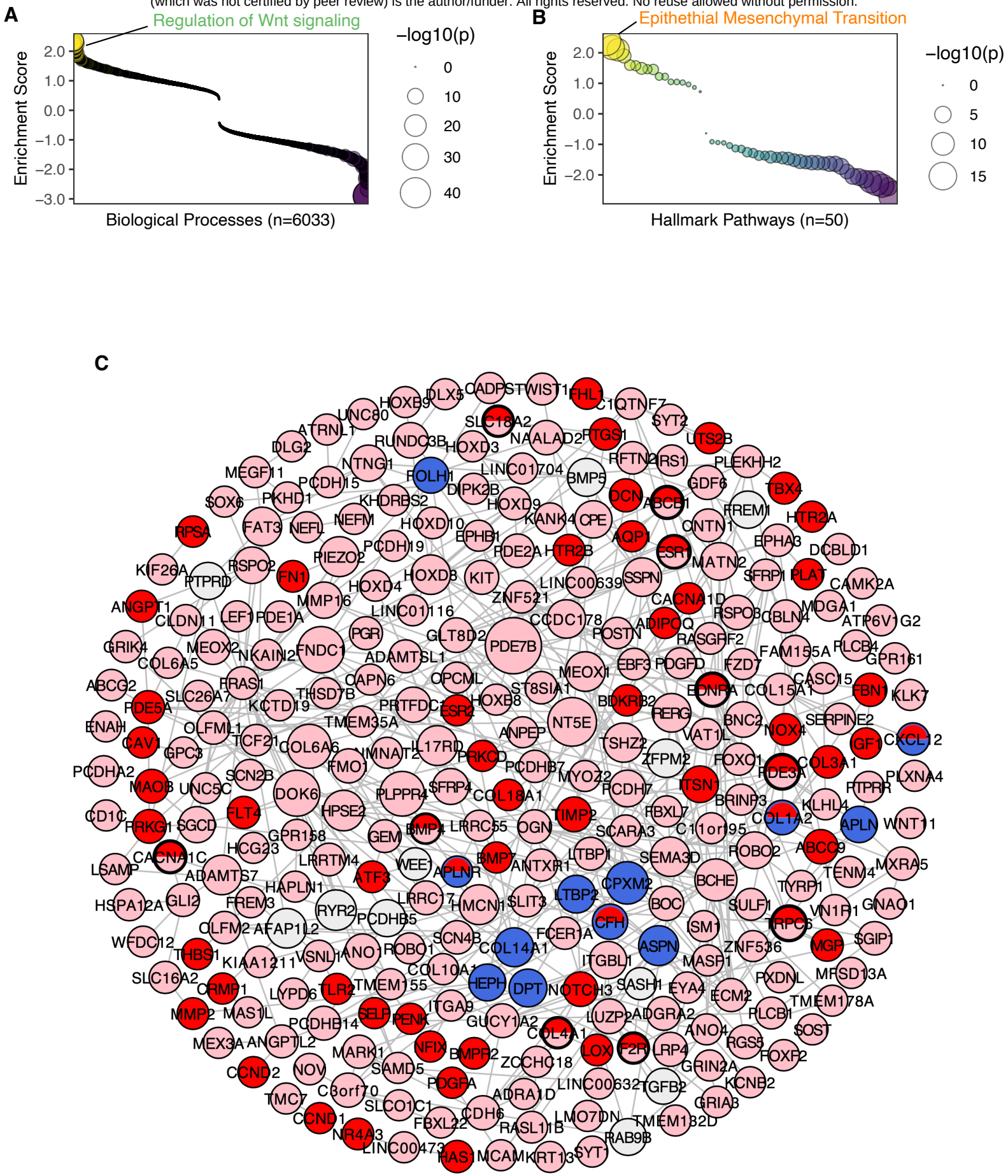


Supplemental Figure 7

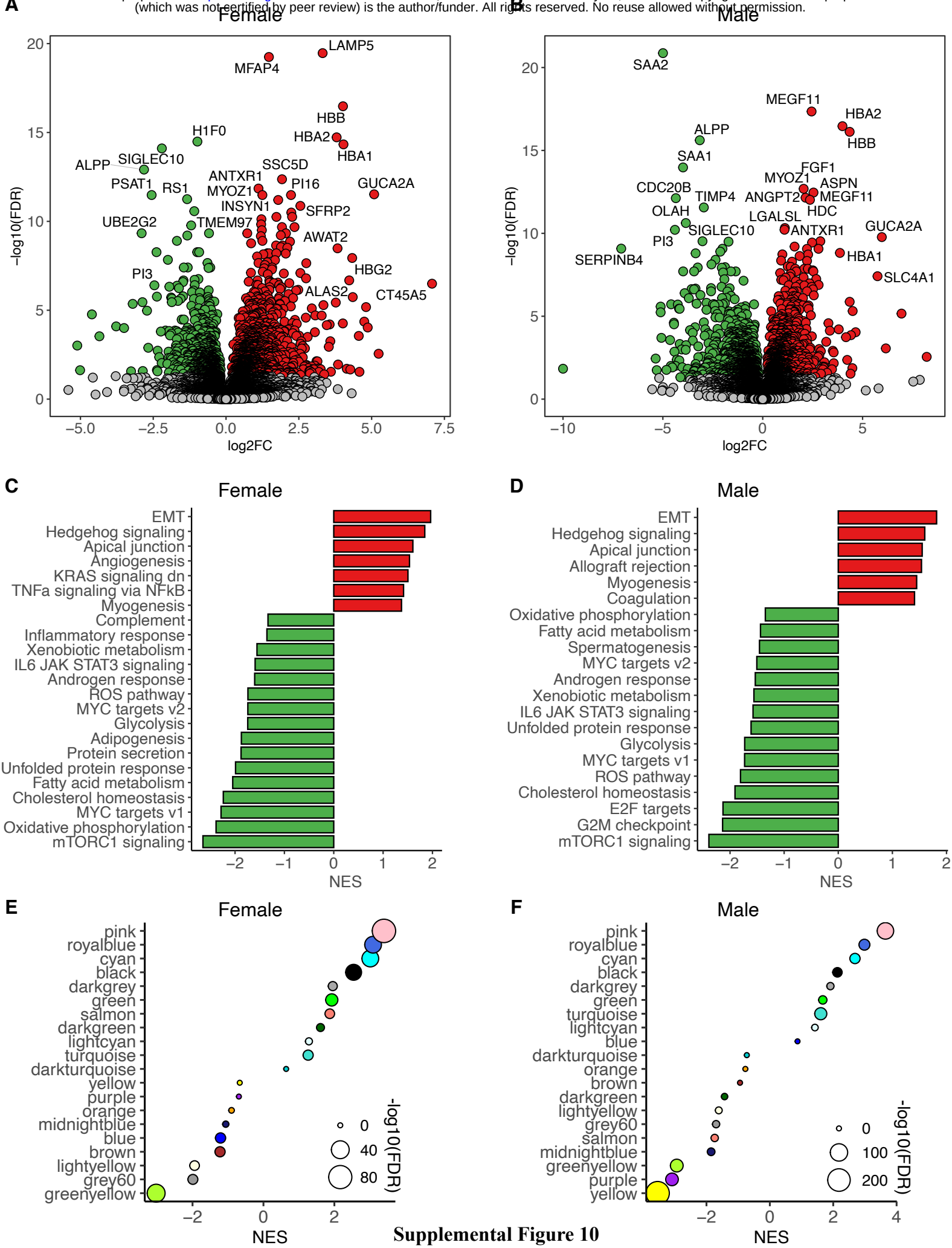




Supplemental Figure 8



Supplemental Figure 9



Supplemental Figure 10