Supplemental Figures

Figure S1: Hierarchical clustering of the gene data for (top) IPF and (bottom) control datasets are shown. Red lines on dendrograms refer to the threshold that was set for identifying outliers. We identified 1 outlier in each IPF group and 3 and 1 outliers in the two control groups.

Figure S2: Module Eigengene clustering for the two IPF groups. Comparison of Module Eigengenes between the two groups shows good preservation.

Figure S3: A t-SNE plot is shown using the 6425 differentially expressed genes we identified for IPF on 3 IPF datasets. (A) GSE53845 consists of 40 IPF samples and 8 control samples. (B) GSE110147 consists of 22 IPF, 10 NSIP and 11 control samples. (C) The complete LTRC dataset comprised of IPF (n=160), COPD (n=220), respiratory bronchiolitis interstitial lung disease (RB-ILD) (n=13), non-specific interstitial pneumonia (NSIP) (n=17), hypersensitivity pneumonitis (HP) (n=30), and control (n=108) subjects. Subjects with IPF were distinctly separated from control samples using these genes in all datasets. There appears to be some overlap between NSIP and HP with IPF but a distinct separation for other disease phenotypes. Ellipses denote a 67% confidence interval for each phenotype.

Figure S4-S7: Correlations between all modules and phenotypic traits are shown for the two control groups (S4 and S5) and two IPF groups (S6 and S7). Numbers are the R-values determined by Spearman's correlation between the module eigenvector and the trait data. Number in parenthesis is the FDR adjusted *P*-value calculated for the correlation.

Figure S8: Module associated with specific cell types by GO pathway analysis was localized to specific cells using a single cell dataset. Markers for specific cell types consist of VWF (endothelial cells), EPCAM (epithelial cells), and FOXJ1 (ciliated epithelial cells). The marker used to identify fibroblasts is collagen 1 which was also the marker used for module ME6. We localized the most highly connected genes for these modules to specific cluster of cells to confirm the validity of the GO pathway analysis. Colour scale ranges from low (purple) to high (yellow) expression.

Figure S9: The transcription factor ZBTB7B was found to be a novel regulator for epithelial cells as it was the regulator of module ME1 with the strongest enrichment score. (A) We examined the human protein atlas and found ZBTB7B mRNA to be highly expressed in all epithelial cell types (red). (B) Human Protein Atlas showed the ZBTB7B protein expression was also present in lung bronchial epithelial cells. Scale bar = 50 µm (C) A single cell dataset we had previously generated showed ZBTB7B (right) as being highly expressed in lung epithelial cells (EPCAM positive stained cells, left).

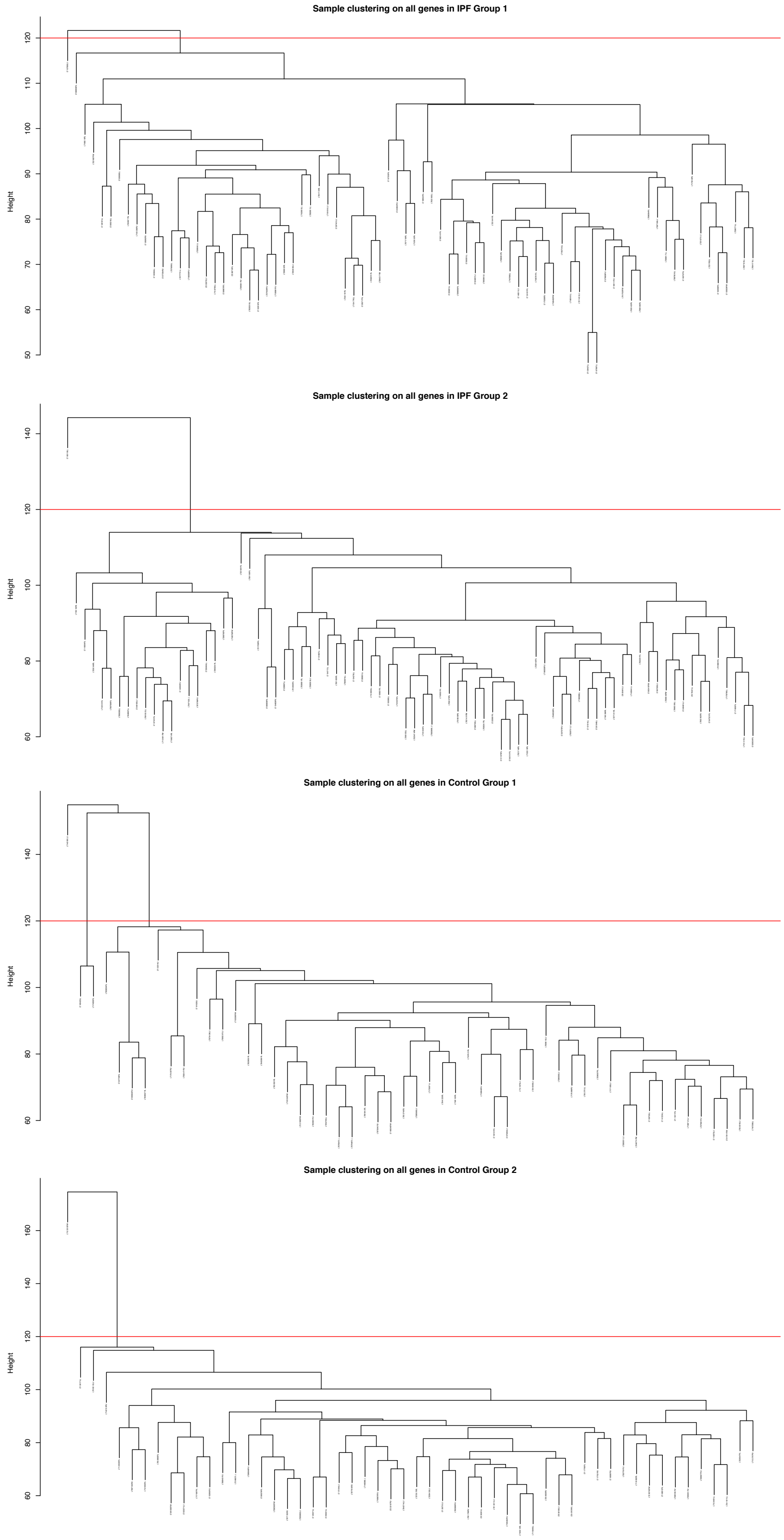
Figure S10: The median value for every gene included in each module was calculated and shown as a boxplot using the peripheral blood sample dataset. While immune response modules showed the highest median gene expression, all modules were detectable in the blood.

Table S1: Transcription factor regulators associated with each module are listed with the number of genes within each module they regulate. Merged number of regulated genes column shows the total number of overlapping genes these transcription factors regulate within the module. Asterisk denotes transcription factors which are also present within the module gene list.

Table S2: A list of miRNAs correlated with each module and the target genes they regulate based on the validated miRNA-target database (miRTarBase). Negative correlation identifies miRNA with a negative correlation to the module Eigengene. MM columns show the module membership of each gene for its respective module in both IPF group 1 and group 2.

Table S3: A listing of all GEO datasets using IPF tissue samples with transcriptomics data, including the dataset used in this study. The number of IPF and Control subjects included in each dataset are shown with the platform used for analysis and the date the data was submitted.

Figure S1



Sample clustering on all genes in IPF Group 1

Sample clustering on all genes in IPF Group 2

Sample clustering on all genes in Control Group 1

Sample clustering on all genes in Control Group 2

Figure S2

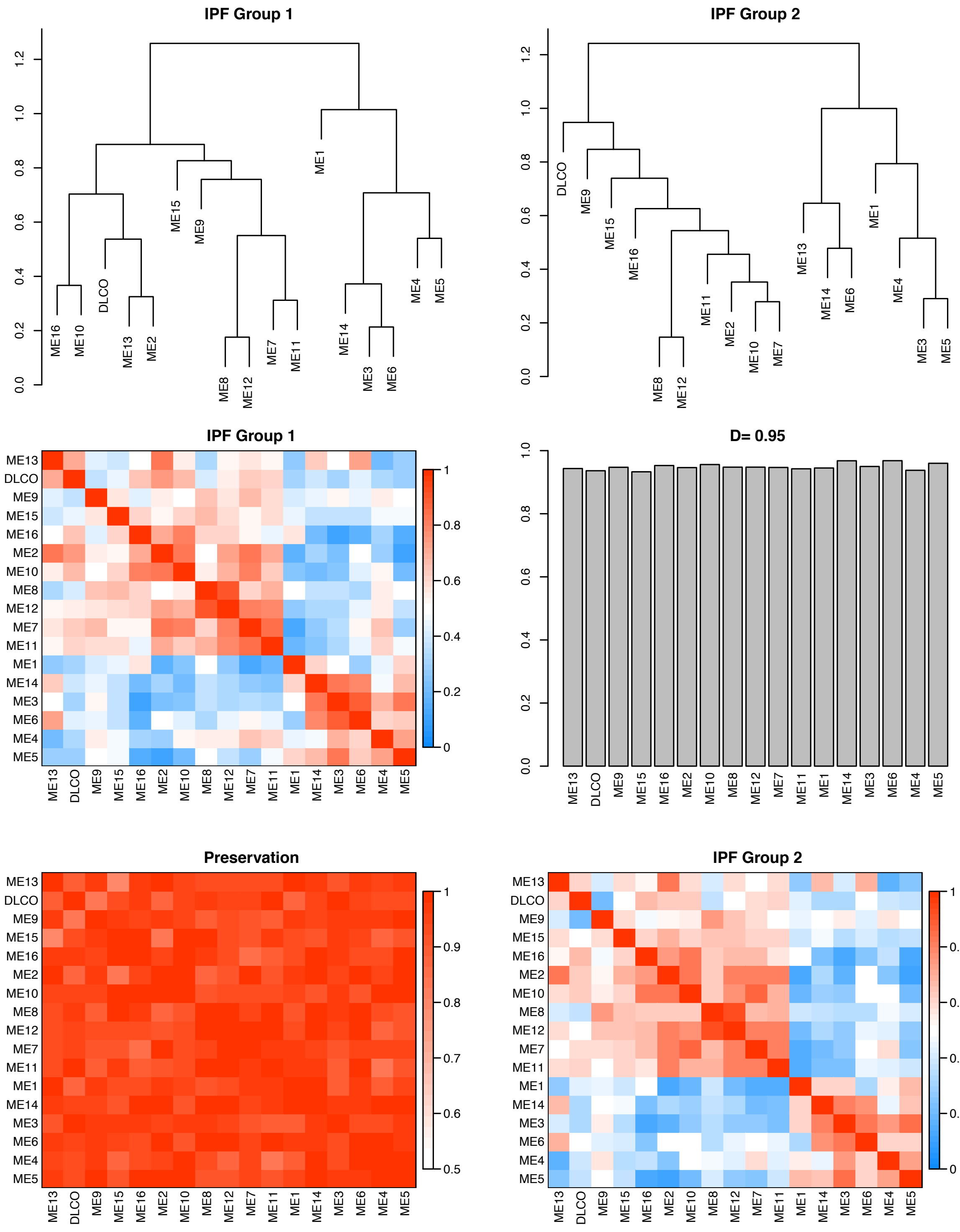

IPF Group 1
0.0
0.2
0.4
0.6
0.8
1.0
1.2
ME16
ME10
DLCO
ME13
ME2
ME15
ME9
ME8
ME12
ME7
ME11
ME1
ME14
ME3
ME6
ME4
ME5
IPF Group 2
DLCO
ME9
ME15
ME16
ME8
ME12
ME11
ME2
ME10
ME7
ME13
ME14
ME6
ME1
ME4
ME3
ME5
IPF Group 1
ME13
DLCO
ME9
ME15
ME16
ME2
ME10
ME8
ME12
ME7
ME11
ME1
ME14
ME3
ME6
ME4
ME5
1
0.8
0.6
0.4
0.2
0
D= 0.95
Preservation
1
0.9
0.8
0.7
0.6
0.5
IPF Group 2

Figure S3

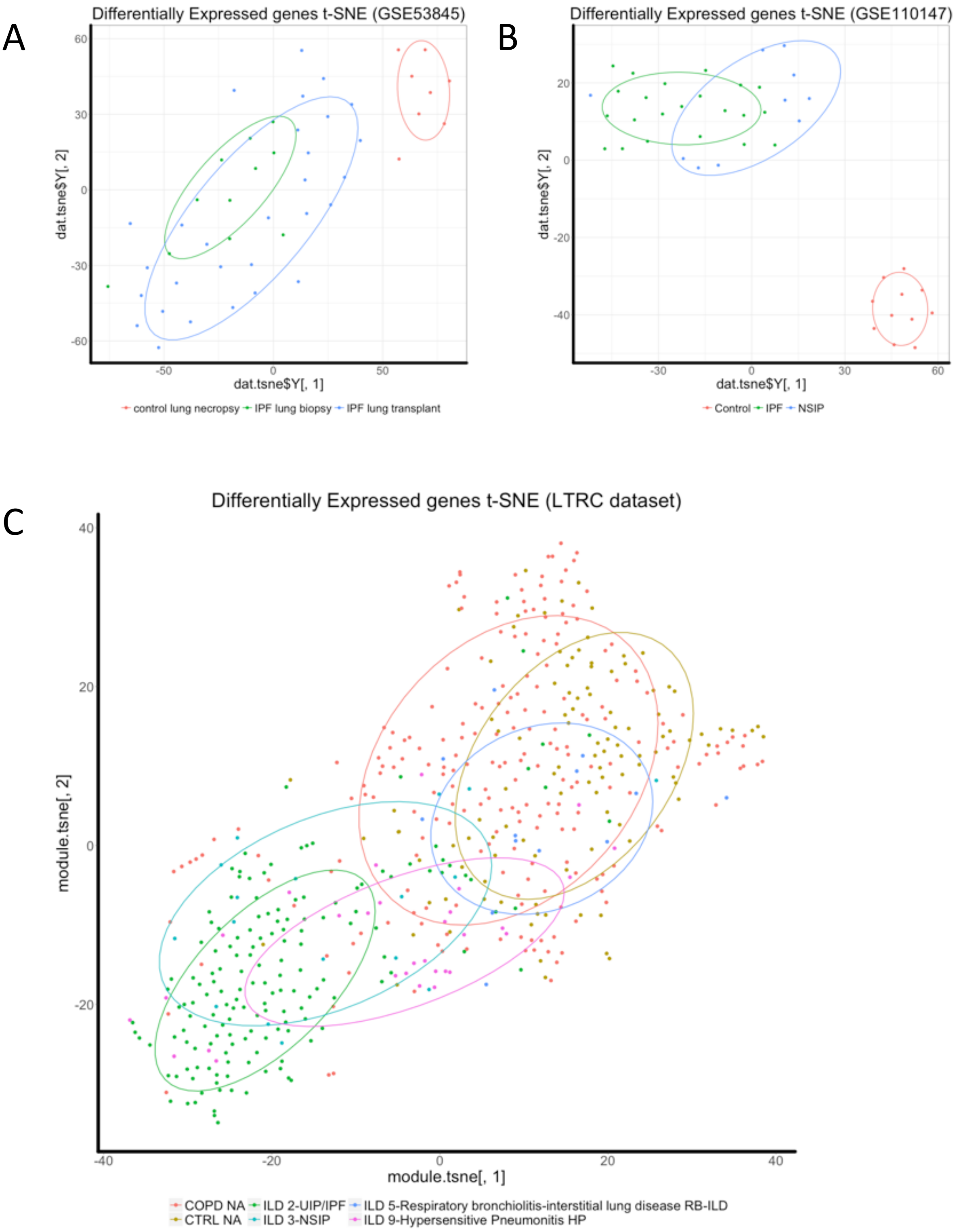

A
Differentially Expressed genes t-SNE (GSE53845)
60
30
0
-30
-60
dat.tsne$Y[, 2]
-50
0
50
dat.tsne$Y[, 1]
control lung necropsy
IPF lung biopsy
IPF lung transplant
B
Differentially Expressed genes t-SNE (GSE110147)
20
0
-20
-40
dat.tsne$Y[, 2]
-30
0
30
60
dat.tsne$Y[, 1]
Control
IPF
NSIP
C
Differentially Expressed genes t-SNE (LTRC dataset)
40
20
0
-20
module.tsne[, 2]
-40
-20
0
20
40
module.tsne[, 1]
COPD NA
CTRL NA
ILD 2-UIP/IPF
ILD 3-NSIP
ILD 5-Respiratory bronchiolitis-interstitial lung disease RB-ILD
ILD 9-Hypersensitive Pneumonitis HP

Figure S4

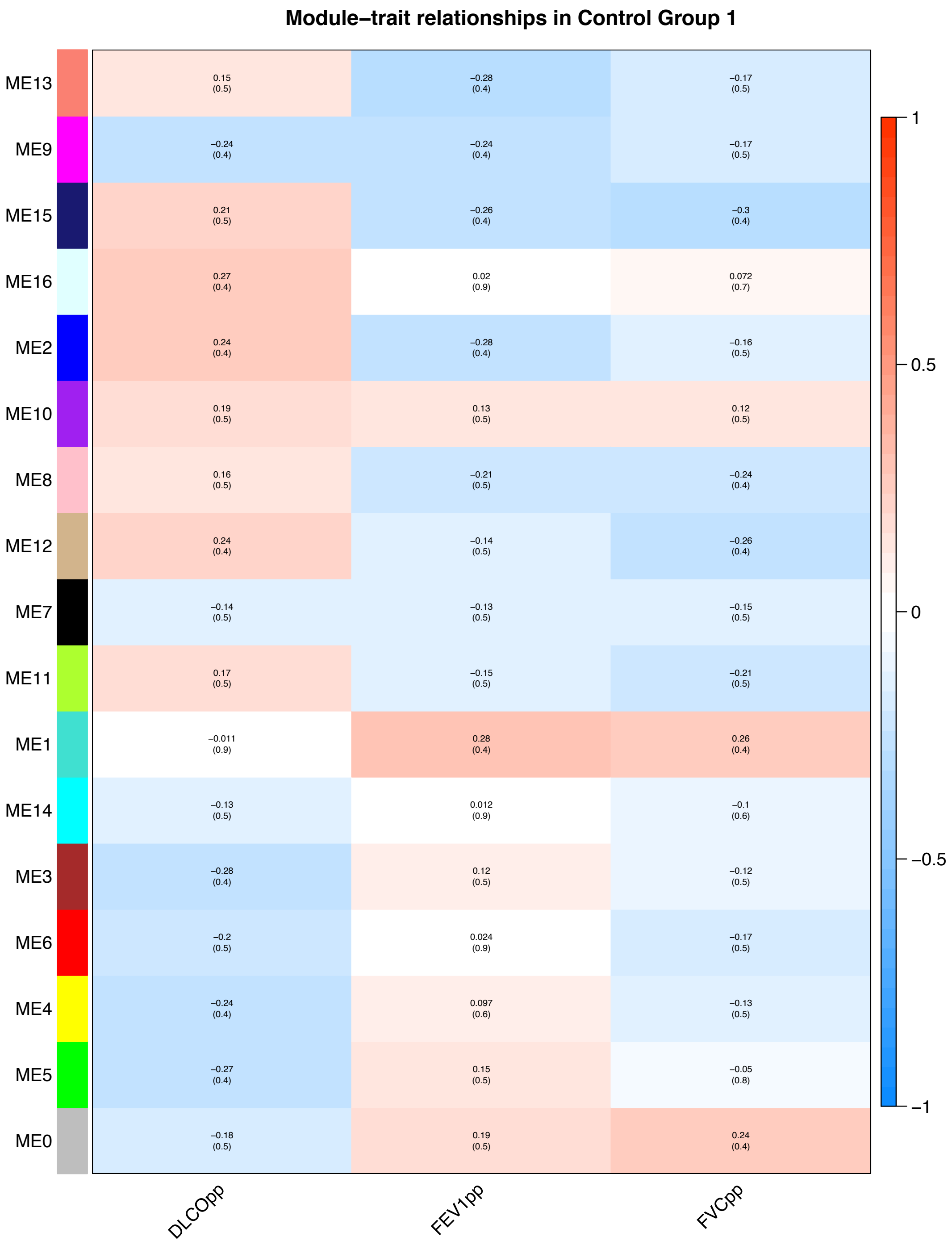
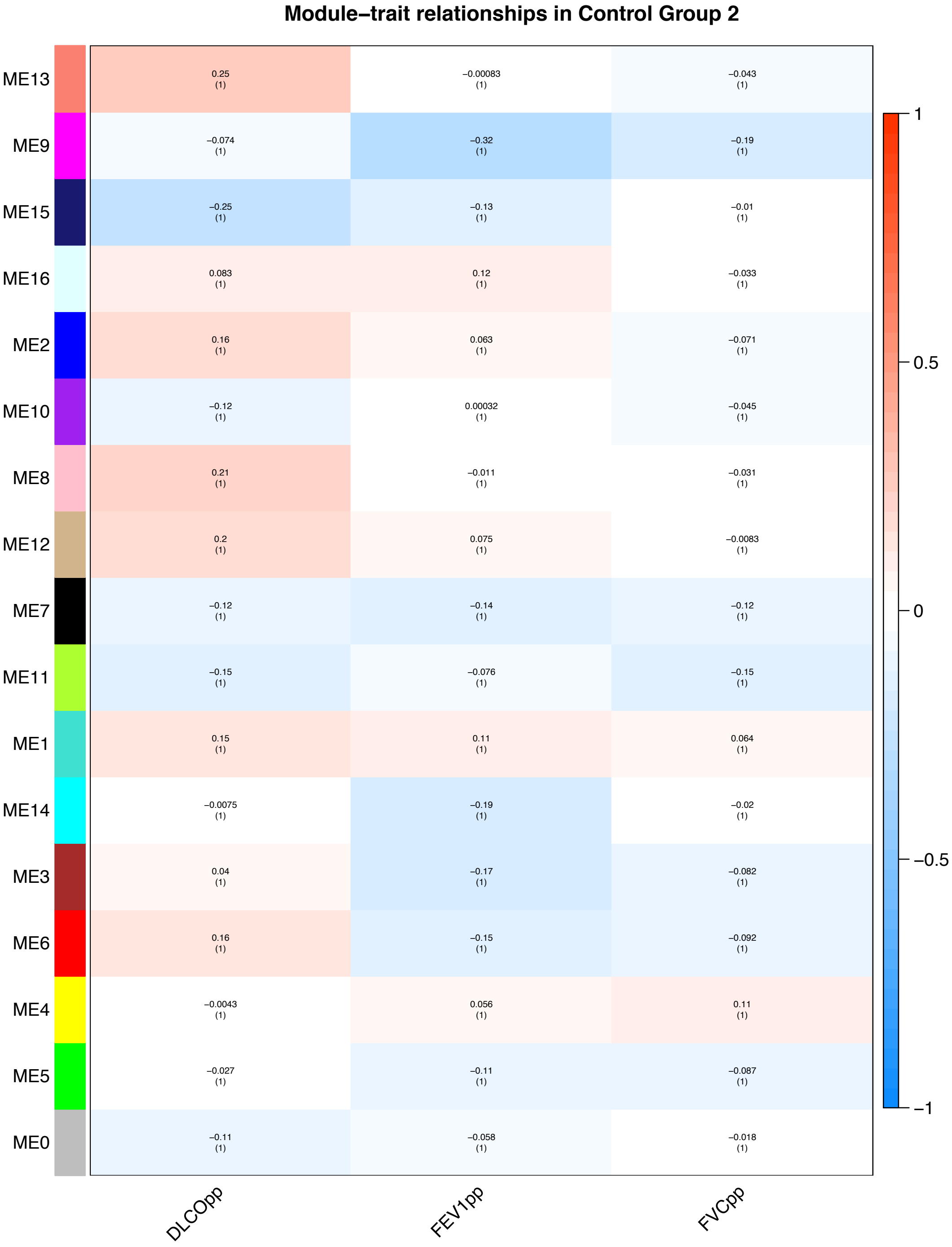

Module–trait relationships in Control Group 1
ME13
0.15 (0.5)
−0.28 (0.4)
−0.17 (0.5)
ME9
−0.24 (0.4)
−0.24 (0.4)
−0.17 (0.5)
ME15
0.21 (0.5)
−0.26 (0.4)
−0.3 (0.4)
ME16
0.27 (0.4)
0.02 (0.9)
0.072 (0.7)
ME2
0.24 (0.4)
−0.28 (0.4)
−0.16 (0.5)
ME10
0.19 (0.5)
0.13 (0.5)
0.12 (0.5)
ME8
0.16 (0.5)
−0.21 (0.5)
−0.24 (0.4)
ME12
0.24 (0.4)
−0.14 (0.5)
−0.26 (0.4)
ME7
−0.14 (0.5)
−0.13 (0.5)
−0.15 (0.5)
ME11
0.17 (0.5)
−0.15 (0.5)
−0.21 (0.5)
ME1
−0.011 (0.9)
0.28 (0.4)
0.26 (0.4)
ME14
−0.13 (0.5)
0.012 (0.9)
−0.1 (0.6)
ME3
−0.28 (0.4)
0.12 (0.5)
−0.12 (0.5)
ME6
−0.2 (0.5)
0.024 (0.9)
−0.17 (0.5)
ME4
−0.24 (0.4)
0.097 (0.6)
−0.13 (0.5)
ME5
−0.27 (0.4)
0.15 (0.5)
−0.05 (0.8)
ME0
−0.18 (0.5)
0.19 (0.5)
0.24 (0.4)
DLCOpp
FEV1pp
FVCpp
1
0.5
0
−0.5
−1

Figure S5



Module–trait relationships in Control Group 2

	DLCOpp	FEV1pp	FVCpp
ME13	0.25 (1)	−0.00083 (1)	−0.043 (1)
ME9	−0.074 (1)	−0.32 (1)	−0.19 (1)
ME15	−0.25 (1)	−0.13 (1)	−0.01 (1)
ME16	0.083 (1)	0.12 (1)	−0.033 (1)
ME2	0.16 (1)	0.063 (1)	−0.071 (1)
ME10	−0.12 (1)	0.00032 (1)	−0.045 (1)
ME8	0.21 (1)	−0.011 (1)	−0.031 (1)
ME12	0.2 (1)	0.075 (1)	−0.0083 (1)
ME7	−0.12 (1)	−0.14 (1)	−0.12 (1)
ME11	−0.15 (1)	−0.076 (1)	−0.15 (1)
ME1	0.15 (1)	0.11 (1)	0.064 (1)
ME14	−0.0075 (1)	−0.19 (1)	−0.02 (1)
ME3	0.04 (1)	−0.17 (1)	−0.082 (1)
ME6	0.16 (1)	−0.15 (1)	−0.092 (1)
ME4	−0.0043 (1)	0.056 (1)	0.11 (1)
ME5	−0.027 (1)	−0.11 (1)	−0.087 (1)
ME0	−0.11 (1)	−0.058 (1)	−0.018 (1)

Figure S6

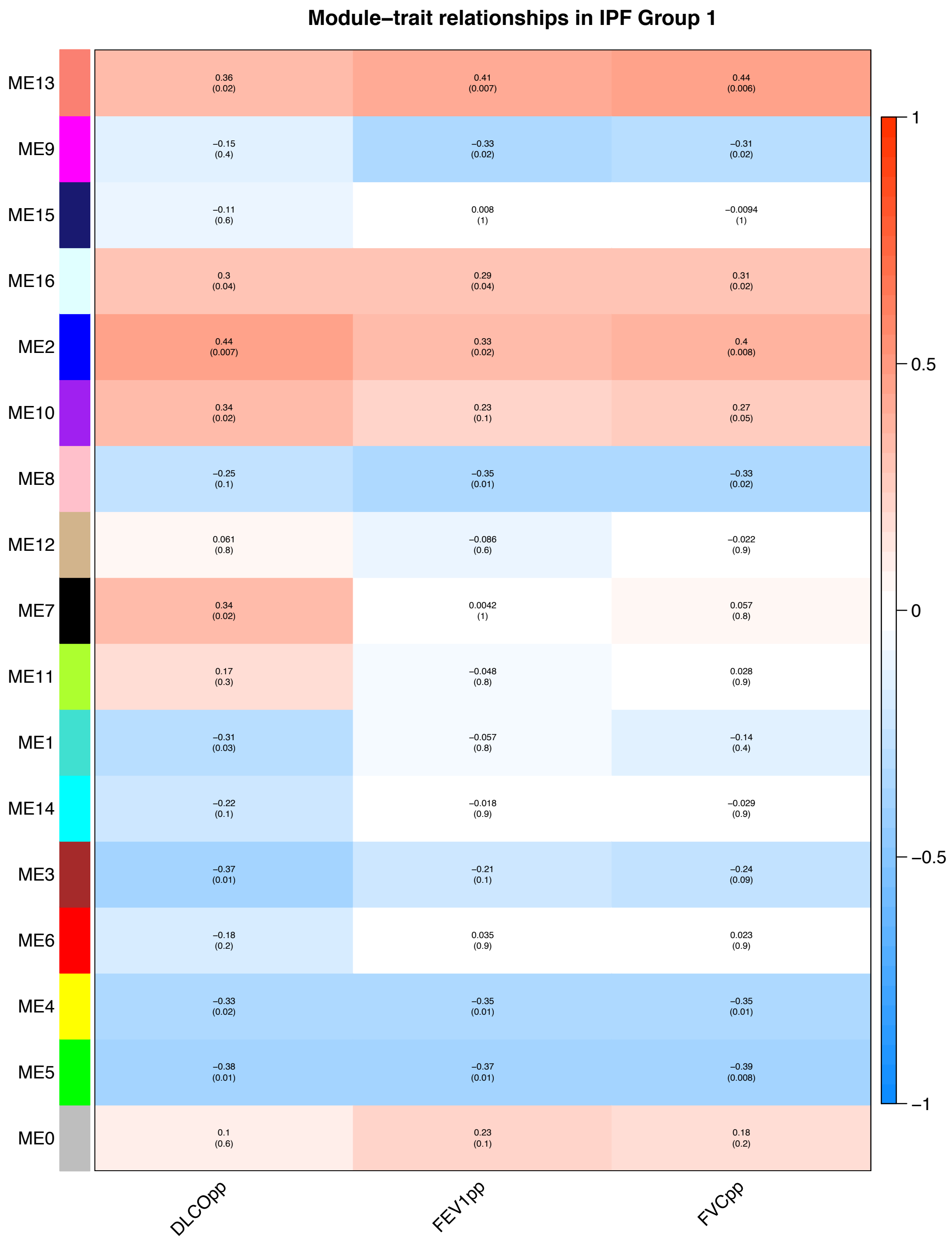

Module–trait relationships in IPF Group 1
ME13
0.36 (0.02)
0.41 (0.007)
0.44 (0.006)
ME9
−0.15 (0.4)
−0.33 (0.02)
−0.31 (0.02)
ME15
−0.11 (0.6)
0.008 (1)
−0.0094 (1)
ME16
0.3 (0.04)
0.29 (0.04)
0.31 (0.02)
ME2
0.44 (0.007)
0.33 (0.02)
0.4 (0.008)
ME10
0.34 (0.02)
0.23 (0.1)
0.27 (0.05)
ME8
−0.25 (0.1)
−0.35 (0.01)
−0.33 (0.02)
ME12
0.061 (0.8)
−0.086 (0.6)
−0.022 (0.9)
ME7
0.34 (0.02)
0.0042 (1)
0.057 (0.8)
ME11
0.17 (0.3)
−0.048 (0.8)
0.028 (0.9)
ME1
−0.31 (0.03)
−0.057 (0.8)
−0.14 (0.4)
ME14
−0.22 (0.1)
−0.018 (0.9)
−0.029 (0.9)
ME3
−0.37 (0.01)
−0.21 (0.1)
−0.24 (0.09)
ME6
−0.18 (0.2)
0.035 (0.9)
0.023 (0.9)
ME4
−0.33 (0.02)
−0.35 (0.01)
−0.35 (0.01)
ME5
−0.38 (0.01)
−0.37 (0.01)
−0.39 (0.008)
ME0
0.1 (0.6)
0.23 (0.1)
0.18 (0.2)
DLCOpp
FEV1pp
FVCpp
1
0.5
0
−0.5
−1

Figure S7

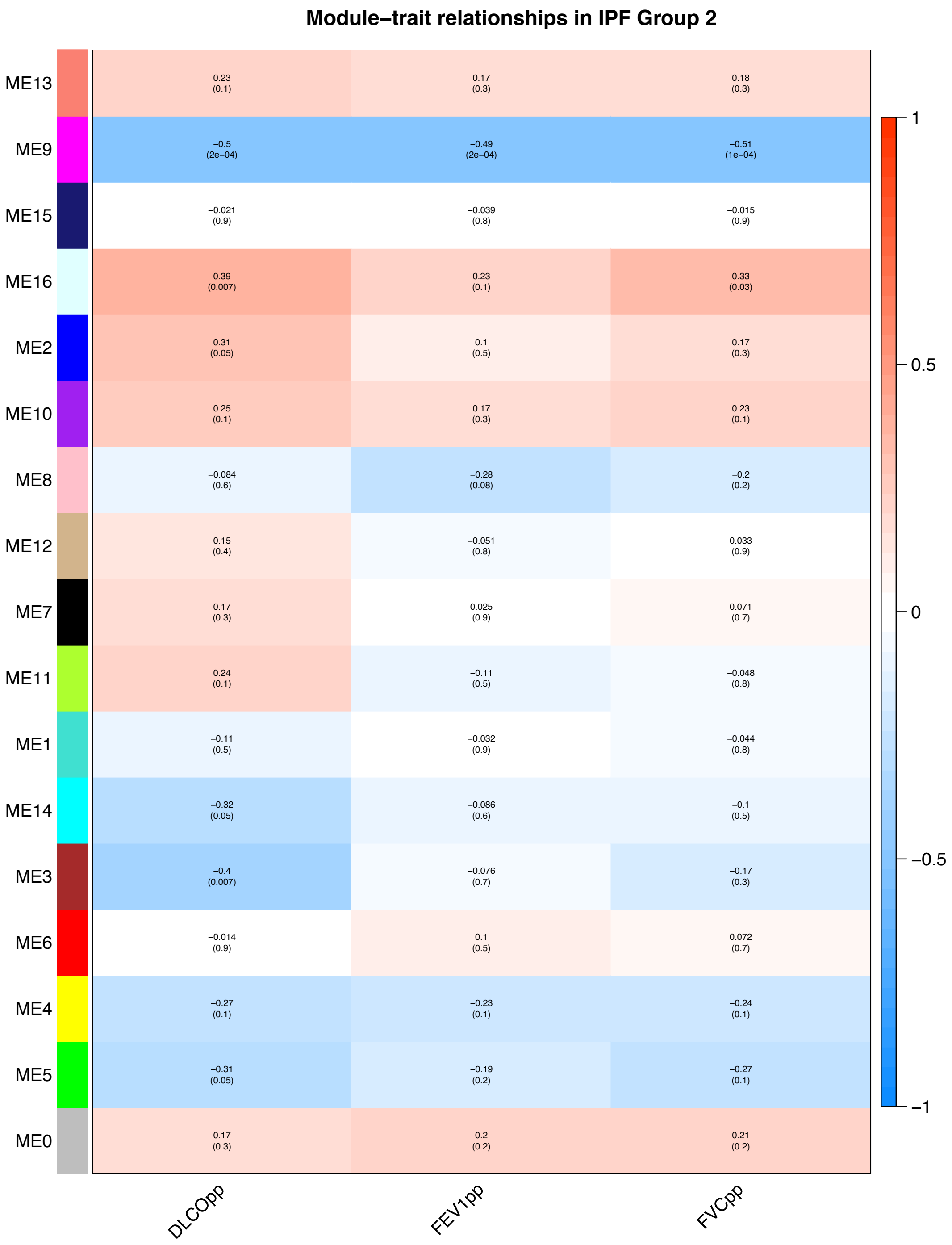

Module–trait relationships in IPF Group 2
ME13
0.23 (0.1)
0.17 (0.3)
0.18 (0.3)
ME9
−0.5 (2e−04)
−0.49 (2e−04)
−0.51 (1e−04)
ME15
−0.021 (0.9)
−0.039 (0.8)
−0.015 (0.9)
ME16
0.39 (0.007)
0.23 (0.1)
0.33 (0.03)
ME2
0.31 (0.05)
0.1 (0.5)
0.17 (0.3)
ME10
0.25 (0.1)
0.17 (0.3)
0.23 (0.1)
ME8
−0.084 (0.6)
−0.28 (0.08)
−0.2 (0.2)
ME12
0.15 (0.4)
−0.051 (0.8)
0.033 (0.9)
ME7
0.17 (0.3)
0.025 (0.9)
0.071 (0.7)
ME11
0.24 (0.1)
−0.11 (0.5)
−0.048 (0.8)
ME1
−0.11 (0.5)
−0.032 (0.9)
−0.044 (0.8)
ME14
−0.32 (0.05)
−0.086 (0.6)
−0.1 (0.5)
ME3
−0.4 (0.007)
−0.076 (0.7)
−0.17 (0.3)
ME6
−0.014 (0.9)
0.1 (0.5)
0.072 (0.7)
ME4
−0.27 (0.1)
−0.23 (0.1)
−0.24 (0.1)
ME5
−0.31 (0.05)
−0.19 (0.2)
−0.27 (0.1)
ME0
0.17 (0.3)
0.2 (0.2)
0.21 (0.2)
DLCOpp
FEV1pp
FVCpp
1
0.5
0
−0.5
−1

Figure S8

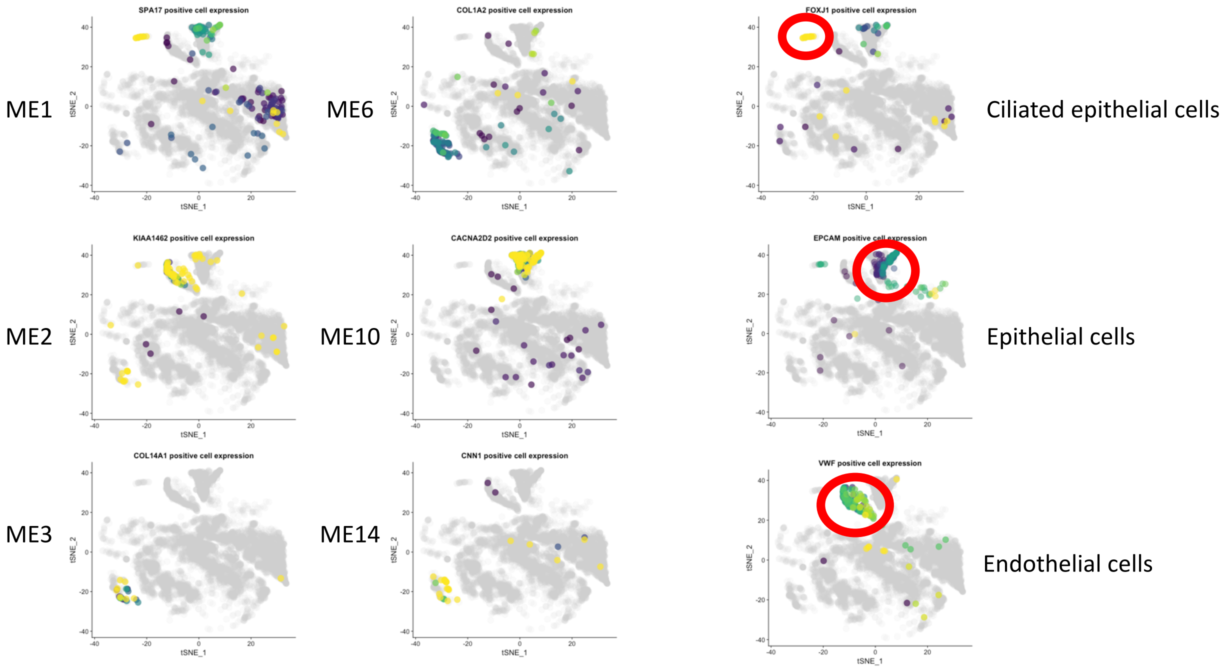

ME1
SPA17 positive cell expression
ME6
COL1A2 positive cell expression
FOXJ1 positive cell expression
Ciliated epithelial cells
ME2
KIAA1462 positive cell expression
ME10
CACNA2D2 positive cell expression
EPCAM positive cell expression
Epithelial cells
ME3
COL14A1 positive cell expression
ME14
CNN1 positive cell expression
VWF positive cell expression
Endothelial cells
tSNE_1
tSNE_2
-40
-20
0
20
40

Figure S9

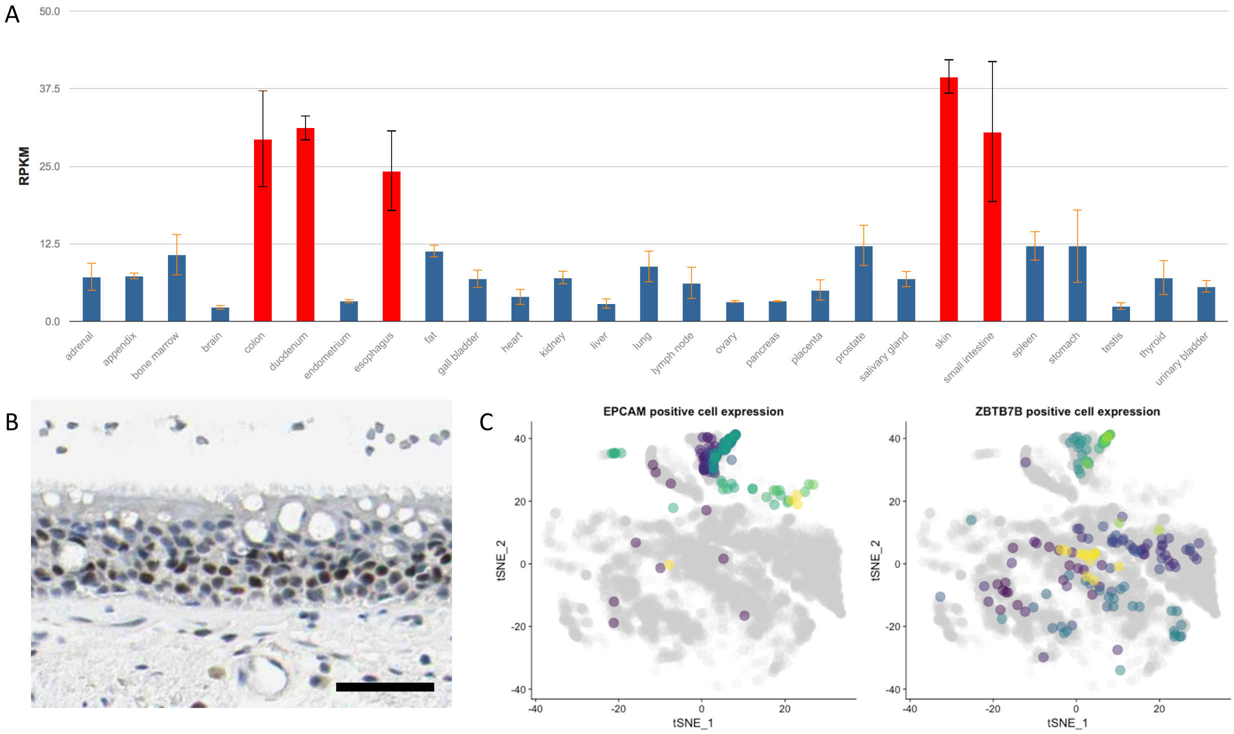

A
50.0
37.5
25.0
12.5
0.0
RPKM
adrenal
appendix
bone marrow
brain
colon
duodenum
endometrium
esophagus
fat
gall bladder
heart
kidney
liver
lung
lymph node
ovary
pancreas
placenta
prostate
salivary gland
skin
small intestine
spleen
stomach
testis
thyroid
urinary bladder
B
C
EPCAM positive cell expression
ZBTB7B positive cell expression
40
20
0
-20
-40
tSNE_2
-40
-20
0
20
tSNE_1

Figure S10

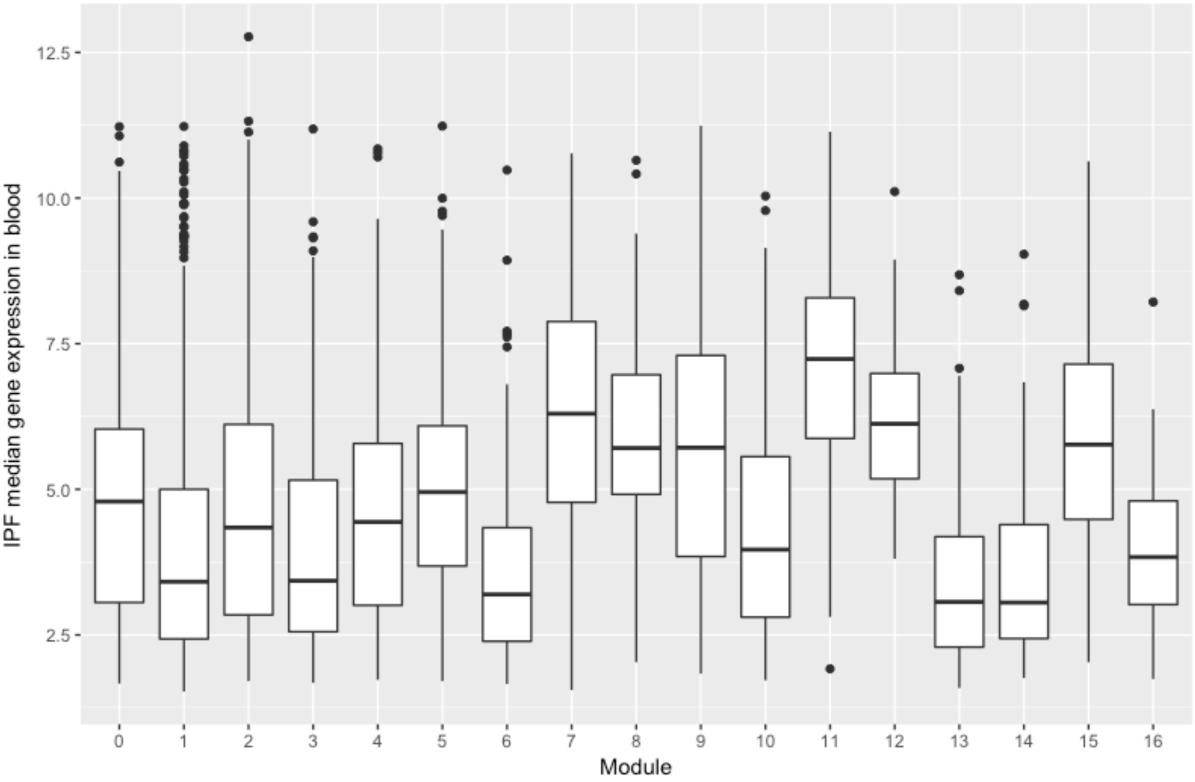

12.5
10.0
7.5
5.0
2.5
IPF median gene expression in blood
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Module

Table S1

Module	Number Genes in Module	Merged Number Regulated Genes	Transcription Factor	Regulated Genes in Module
ME1	1579	541	RFX2	506
			ZBTB7B	376
ME4	380	258	E2F4	190
			SIN3A	161
			TFDP1	123
			MYBL2*	97
			FOXM1*	93
			E2F7	30
ME6	207	89	EP300	74
			TCF12	65
ME7	204	133	SPI1	123
			EBF1	23
			JUN	23
ME8	188	98	CREB1	98
ME9	163	82	NR3C1	57
			CHD1	54
ME10	158	30	SREBF1	18
			ZEB1	13
ME11	116	85	EBF1	65
			STAT1	39
			IRF1*	30
			STAT2*	21
			RFX5	19
ME13	80	26	EB300	23
			TCF12	18
ME14	75	58	SRF	38
			TCF12*	36
ME15	58	15	ATF3	15
ME16	39	24	ZEB1	10
			EP300	22
			TEAD4	8

Table S2

Module	miRNA	Target gene	Negative correlation	MM IPF group 1	MM IPF group 2
ME1	hsa-miR-200a-3p	MACC1		0.314	0.421
ME1	hsa-miR-200a-3p	RPL12		0.526	0.368
ME1	hsa-miR-200a-5p	METTL7A		0.468	0.317
ME1	hsa-miR-200b-3p	AVPR1A		0.362	0.278
ME1	hsa-miR-200b-3p	UNC119B		0.863	0.826
ME1	hsa-miR-200b-5p	C11orf74		0.891	0.897
ME1	hsa-miR-205-5p	TRAF3IP1		0.929	0.927
ME1	hsa-miR-31-5p	FOXC1		0.721	0.739
ME1	hsa-miR-34b-3p	MET		0.727	0.800
ME1	hsa-miR-34b-5p	MET		0.727	0.800
ME1	hsa-miR-34b-5p	RPS19		0.255	0.539
ME1	hsa-miR-34c-3p	GALK2		0.686	0.664
ME1	hsa-miR-34c-5p	RUNX2		0.558	0.590
ME2	hsa-miR-126-3p	KRAS		0.313	0.406
ME2	hsa-miR-126-3p	VEGFA		0.674	0.785
ME2	hsa-miR-126-5p	VEGFA		0.674	0.785
ME2	hsa-miR-132-3p	BDNF	*	0.761	0.817
ME2	hsa-miR-132-3p	PARP10	*	0.290	0.310
ME2	hsa-miR-132-3p	TSPAN12	*	0.839	0.728
ME2	hsa-miR-205-3p	HS2ST1	*	0.254	0.384
ME2	hsa-miR-205-5p	TAF11	*	0.268	0.630
ME2	hsa-miR-30a-3p	FOXP1		0.322	0.467
ME2	hsa-miR-30a-3p	PLA2G16		0.582	0.535
ME2	hsa-miR-30a-3p	SULT1C4		0.628	0.372
ME2	hsa-miR-30a-3p	TMEM2		0.496	0.513
ME2	hsa-miR-30a-5p	MOGAT1		0.696	0.599
ME2	hsa-miR-30a-5p	RAB14		0.647	0.428
ME2	hsa-miR-30a-5p	TBC1D4		0.528	0.597
ME2	hsa-miR-30a-5p	ZNRF1		0.636	0.566
ME2	hsa-miR-30b-3p	C19orf47		0.271	0.396
ME2	hsa-miR-30b-3p	FAM105A		0.852	0.872
ME2	hsa-miR-30b-3p	NCOR2		0.486	0.475
ME2	hsa-miR-30b-3p	PLCD3		0.402	0.286
ME2	hsa-miR-30b-3p	PPFIBP1		0.534	0.581
ME2	hsa-miR-30b-3p	RBMS2		0.399	0.480
ME2	hsa-miR-30d-5p	DCTN4		0.344	0.308
ME2	hsa-miR-30d-5p	IER5		0.616	0.456
ME2	hsa-miR-30d-5p	LDLR		0.485	0.497
ME2	hsa-miR-30d-5p	ZMPSTE24		0.185	0.181
ME2	hsa-miR-31-5p	KLF13	*	0.584	0.657

ME2	hsa-miR-31-5p	MED12	*	0.211	0.230
ME2	hsa-miR-31-5p	PRKCE	*	0.806	0.846
ME2	hsa-miR-31-5p	RDX	*	0.759	0.742
ME2	hsa-miR-31-5p	ZC3H12C	*	0.454	0.593
ME2	hsa-miR-338-3p	RNF8		0.320	0.332
ME3	hsa-miR-127-3p	RGMA		0.590	0.647
ME3	hsa-miR-203a-3p	EGLN1	*	0.595	0.539
ME3	hsa-miR-203a-5p	CALU	*	0.769	0.728
ME3	hsa-miR-203b-3p	SEC24D	*	0.828	0.780
ME3	hsa-miR-203b-3p	TIMM8A	*	0.470	0.658
ME3	hsa-miR-30a-3p	FXN	*	0.467	0.563
ME3	hsa-miR-30a-3p	SLITRK4	*	0.558	0.385
ME3	hsa-miR-30a-5p	CDC123	*	0.279	0.315
ME3	hsa-miR-30a-5p	LOX	*	0.786	0.503
ME3	hsa-miR-30a-5p	PLA2G4A	*	0.397	0.343
ME3	hsa-miR-30a-5p	PPP1CC	*	0.263	0.235
ME3	hsa-miR-30a-5p	SH3GL1	*	0.729	0.748
ME3	hsa-miR-30a-5p	XPOT	*	0.339	0.308
ME3	hsa-miR-30b-5p	ATG12	*	0.284	0.416
ME3	hsa-miR-30b-5p	ELOVL4	*	0.647	0.476
ME3	hsa-miR-30b-5p	TMEM106B	*	0.610	0.497
ME5	hsa-miR-30a-5p	NOL8	*	0.320	0.405
ME5	hsa-miR-30a-5p	NXT1	*	0.398	0.511
ME5	hsa-miR-30a-5p	PAICS	*	0.531	0.520
ME5	hsa-miR-30a-5p	PLIN3	*	0.518	0.481
ME5	hsa-miR-30a-5p	SLC38A1	*	0.462	0.561
ME5	hsa-miR-30a-5p	SPCS3	*	0.901	0.890
ME5	hsa-miR-30a-5p	TNFRSF10B	*	0.515	0.557
ME5	hsa-miR-30a-5p	TXNDC5	*	0.844	0.819
ME5	hsa-miR-30b-3p	COX6B1	*	0.514	0.551
ME5	hsa-miR-30b-3p	CSTF1	*	0.253	0.304
ME5	hsa-miR-30b-3p	DPM2	*	0.492	0.499
ME5	hsa-miR-30b-3p	SNRPD1	*	0.375	0.612
ME5	hsa-miR-30b-3p	TIAL1	*	0.558	0.227
ME5	hsa-miR-30b-3p	TNFRSF13C	*	0.847	0.787
ME5	hsa-miR-30b-5p	CASP3	*	0.779	0.606
ME5	hsa-miR-30b-5p	CHST15	*	0.735	0.717
ME5	hsa-miR-30b-5p	LARP1	*	0.463	0.587
ME5	hsa-miR-30b-5p	SMAD1	*	0.405	0.466
ME10	hsa-miR-338-3p	GPRC5A		0.686	0.707
ME10	hsa-miR-652-3p	QSOX1		0.756	0.702
ME14	hsa-miR-133b	PTBP2		0.473	0.564

Table S3

GEO Dataset	# Subjects	Platform		Submission Date
GSE10667	23 IPF/8 IPF+acute exacerbations; 15 Control	GPL4133	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Feature Number version)	28-Feb-08
GSE21369	11 IPF/5 NSIP/7 uncharacterized ILD; 6 Control	GPL570	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	16-Apr-10
GSE24206	11 IPF; 6 Control	GPL570	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	17-Sep-10
GSE31934	3 IPF; 3 Control	GPL11097	[Glyco_v4_Hs] Custom Affymetrix Glyco v4 GeneChip	06-Sep-11
GSE32537	119 IPF/17 NSIP/31 other ILD; 50 Control (LTRC)	GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	01-Oct-11
GSE35145	4 IPF; 4 Control	GPL10558	Illumina HumanHT-12 V4.0 expression beadchip	17-Jan-12
GSE47460	160 IPF/30 HP/17 NSIP/47 other ILD; 108 Control (LTRC)	GPL6480	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version)	29-May-13
GSE48149	13 IPF; 9 Control	GPL16221	Illumina HumanRef-8 v3.0 expression beadchip (Search Key version)	20-Jun-13
GSE52463	8 IPF; 7 Control	GPL11154	Illumina HiSeq 2000 (Homo sapiens)	18-Nov-13
GSE53845	40 IPF; 8 Control	GPL6480	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version)	06-Jan-14
GSE72073	5 IPF; 3 Control	GPL17586	[HTA-2_0] Affymetrix Human Transcriptome Array 2.0 [transcript (gene) version]	14-Aug-15
GSE83717	7 IPF; 5 Control	GPL11154	Illumina HiSeq 2000 (Homo sapiens)	24-Jun-16
GSE92592	20 IPF; 19 Control (LTRC)	GPL11154	Illumina HiSeq 2000 (Homo sapiens)	19-Dec-16
GSE99621	3 IPF; 3 Control	GPL16791	Illumina HiSeq 2500 (Homo sapiens)	02-Jun-17
GSE110147	22 IPF/10 NSIP/5 mixed IPF-NSIP; 11 Control	GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	05-Feb-18