Supplemental Data
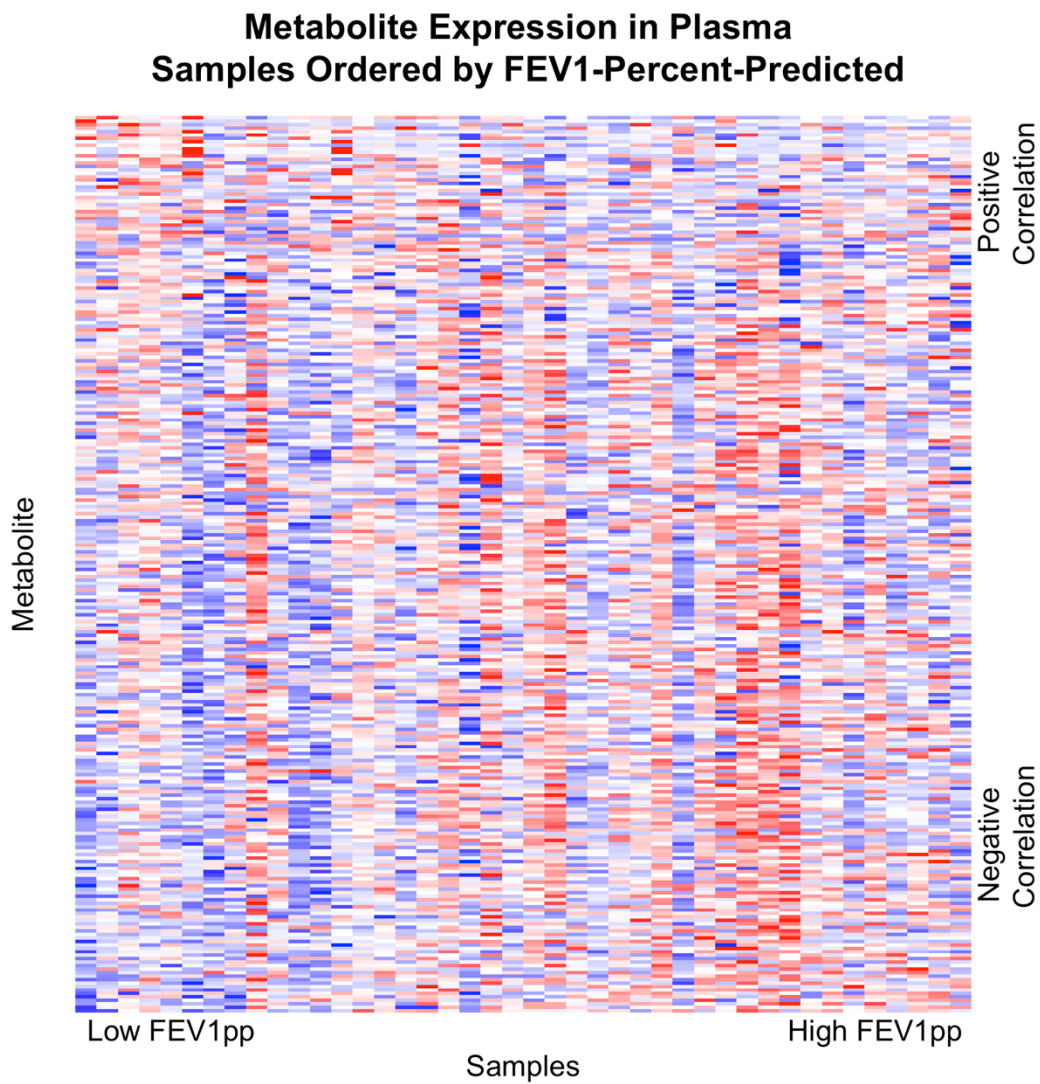
Figure 1S (Supplement): Heatmap of metabolite expression in plasma. Columns reflect samples and rows reflect metabolites. The metabolites are ordered by the direction and magnitude of their correlation with FEV1pp.

Metabolite Expression in Plasma
Samples Ordered by FEV1-Percent-Predicted

Table 1S (Supplement): Top 10 pathways based on the proteins significantly correlated with FEV1pp at the 0.05 level under a permutation testing framework. The column names are: # Overlapping Genes lists the number of significant proteins in our dataset that belonged to each pathway; # All Pathway Genes lists the total number of proteins in our dataset that belong to each pathway and in parentheses is given the total number of proteins in that pathway; P-Value gives the overrepresentation p-value; and Q-Value gives the false-discovery rate adjusted p-value for each pathway.

Pathway	Source	# Overlapping Genes	# All Pathway Genes	P-Value	Q-Value
EPH-Ephrin signaling	Reactome	25	39 (74)	0.000125	0.434
Ephrin signaling	Reactome	11	13 (19)	0.000274	0.434
Salivary secretion - Homo sapiens (human)	KEGG	20	30 (93)	0.000278	0.434
Sema3A PAK dependent Axon repulsion	Reactome	7	7 (16)	0.000549	0.642
Post-translational protein modification	Reactome	167	405 (1425)	0.001040	0.918
Class B/2 (Secretin family receptors)	Reactome	15	22 (91)	0.001180	0.918
Splicing factor NOVA regulated synaptic proteins	Wikipathways	9	11 (42)	0.001690	1.000
EPHB forward signaling	PID	14	21 (39)	0.002400	1.000
VEGFA-VEGFR2 Signaling Pathway	Wikipathways	74	166 (433)	0.003000	1.000

Lysosome - Homo sapiens (human)	KEGG	27	50 (128)	0.003040

Table 2S (Supplement): Top 10 proteins correlated with FEV1/FVC. Proteins are ordered based on FDR-adjusted p-values from permutation testing with the Pearson correlation test. Correlation was calculated by averaging the correlation across all aptamers that map to each protein.

Protein	UniProt ID	Average Correlation	Permutation P-Value	Q-Value
OCRL	Q01968	-0.5813986	0.0001000	0.0019753
CUL3	Q13618	-0.5740182	0.0001000	0.0019753
NFL	P07196	-0.6217485	0.0001000	0.0019753
TM190	Q8WZ59	-0.5814013	0.0001000	0.0019753
RASN	P01111	-0.6181181	0.0001000	0.0019753
PXDC2:ECD	Q6UX71	0.5717866	0.0001000	0.0019753
megalin	P98164	-0.5534422	0.0001000	0.0019753
Serpin B1	P30740	-0.5748941	0.0001000	0.0019753
STX8	Q9UNK0	-0.5168538	0.0001000	0.0019753
NRG3	P56975	0.4639801	0.0001000	0.0019753

Table 3S (Supplement): Top 10 proteins correlated with DLCO-percent-predicted. Proteins are ordered based on FDR-adjusted p-values from permutation testing with the Pearson correlation test. Correlation was calculated by averaging the correlation across all aptamers that map to each protein.

Protein	UniProt ID	Average Correlation	Permutation P-Value	Q-Value
NRG3	P56975	0.6242945	0.0001000	0.3871613
Testican-2	Q92563	0.5712197	0.0002000	0.3871613
CNTN3	Q9P232	0.5254836	0.0005000	0.3871613
DAF	P08174	0.5159539	0.0005000	0.3871613
CNTN6	Q9UQ52	0.5085081	0.0005000	0.3871613
Caspase-14	P31944	0.5201442	0.0005999	0.3871613
Nectin-like protein 2	Q9BY67	0.4933851	0.0010999	0.6083963
Butyrophilin 1A1	Q13410	0.4831189	0.0016998	0.6843036
B3GN2	Q9NY97	0.4686611	0.0019998	0.6843036
FLRT3:ECD	Q9NZU0	0.4629704	0.0022998	0.6843036

Table 3S (Supplement): Demographics of the two K-means clusters determined by BALF metabolite expression. Single asterisk (*) indicates significant differences between clusters at 0.05 level.

	Cluster 1 (N=10)	Cluster 2 (N=42)	Total (N=52)
Sex			
Male	9 (90.0%)	29 (69.0%)	38 (73.1%)
Female	1 (10.0%)	13 (31.0%)	14 (26.9%)
Age*			
Mean (SD)	62.6 (8.78)	55.3 (7.78)	56.7 (8.41)
Median [Min, Max]	62.5 [49.0, 76.0]	55.0 [42.0, 80.0]	56.0 [42.0, 80.0]
Ethnicity			
Black, Non-Hispanic	7 (70.0%)	21 (50.0%)	28 (53.8%)

White, Hispanic/Latino	2 (20.0%)	21 (50.0%)	23 (44.2%)
Asian/Pacific Islander	1 (10.0%)	0 (0%)	1 (1.9%)
Smoker*			
Yes	3 (30.0%)	25 (59.5%)	28 (53.8%)
Former	7 (70.0%)	9 (21.4%)	16 (30.8%)
Never	0 (0%)	8 (19.0%)	8 (15.4%)
Pack Years			
Mean (SD)	31.8 (22.4)	21.1 (23.5)	23.1 (23.4)
Median [Min, Max]	30.7 [5, 80]	15.2 [0, 120]	17.2 [0, 120]
Receiving ART			
Yes	9 (90.0%)	39 (92.9%)	48 (92.3%)
No	1 (10.0%)	3 (7.1%)	4 (7.7%)
Viral Load			
<50 Copies	2 (20.0%)	28 (66.7%)	30 (57.7%)
>50 Copies	1 (10.0%)	4 (9.5%)	5 (9.6%)
Missing	7 (70.0%)	10 (23.8%)	17 (32.7%)
FEV1-Percent-Predicted*			
Mean (SD)	68.7 (27.0)	89.9 (19.6)	85.8 (22.6)
Median [Min, Max]	68.7 [21.0, 108]	89.4 [46.7, 128]	86.2 [21.0, 128]
FEV1*			
Mean (SD)	2.12 (0.844)	2.79 (0.867)	2.66 (0.896)
Median [Min, Max]	2.002 [0.650, 0.362]	2.75 [1.32, 4.77]	2.59 [0.650, 4.77]
FEV1/FVC			
Mean (SD)	0.558 (0.201)	0.704 (0.122)	0.676 (0.15)
Median [Min, Max]	0.546 [0.293, 0.825]	0.705 [0.413, 0.905]	0.684 [0.293, 0.905]
Case-Control Status			
Case	7 (70.0%)	19 (45.2%)	26 (50.0%)
Control	3 (30.0%)	23 (54.8%)	26 (50.0%)
DLCO-Percent-Predicted			
Mean (SD)	63.6 (20.4)	76.0 (25.0)	74.2 (24.6)
Median [Min, Max]	59.0 [43.0, 98.2]	75.0 [14.4, 139]	74.5 [14.4, 139]
Missing	4 (40.0%)	7 (16.7%)	11 (21.2%)

Table 4S (Supplement): Top 10 BALF metabolites differentially expressed between two K-means clusters determined by metabolite expression. A positive test statistic implies average metabolite expression was higher in Cluster 1 compared to Cluster 2.

Metabolite	Test Statistic	P-Value	Q-Value
Asn	-9.5122472	0.0000000	0.0000000
AC(12:0)	8.5161455	0.0000001	0.0000157
AC(4:1-DC)	8.8476460	0.0000008	0.0000450
AC(5:0)	9.5827866	0.0000007	0.0000450
AC(6:0-DC)	8.1804113	0.0000009	0.0000450
AC(5:1)	9.1026218	0.0000014	0.0000580
CE(17:0)	8.1085759	0.0000023	0.0000810
AC(5:0-DC)	7.2955463	0.0000026	0.0000833
AC(4:0-DC)	7.6018715	0.0000047	0.0001309
AC(14:0-OH)	7.4445988	0.0000072	0.0001809

Table 5S (Supplement): Demographics of the six overlapping individuals between smaller clusters determined by the BALF proteome and the BALF metabolome. Single asterisk (*) indicates significant differences between clusters at 0.05 level. Double asterisk (**) indicates significant differences between clusters at the 0.005 level.

	Overlapping (N=6)	Non-Overlapping (N=46)	Total (N=52)
Sex			
Male	6 (100%)	32 (69.6%)	38 (73.1%)
Female	0 (0%)	14 (30.4%)	14 (26.9%)
Age			
Mean (SD)	65.7 (9.71)	55.5 (7.58)	56.7 (8.41)
Median [Min, Max]	67.5 [49.0, 76.0]	55.0 [42.0, 80.0]	56.0 [42.0, 80.0]
Ethnicity*			
Black, Non-Hispanic	6 (100%)	22 (47.8%)	28 (53.8%)
White, Hispanic/Latino	0 (0%)	23 (50.0%)	23 (44.2%)
Asian/Pacific Islander	0 (0%)	1 (2.2%)	1 (1.9%)
Smoking Status			

Yes	2 (33.3%)	26 (56.5%)	28 (53.8%)
Former	4 (66.7%)	12 (26.1%)	16 (30.8%)
Never	0 (0%)	8 (17.4%)	8 (15.4%)
Pack Years			
Mean (SD)	37.7 (26.7)	21.3 (22.6)	23.1 (23.4)
Median [Min, Max]	32.5 [5.00, 80.0]	16.2 [0, 120]	17.2 [0, 120]
Receiving ART			
Yes	5 (83.3%)	43 (93.5%)	48 (92.3%)
No	1 (16.7%)	3 (6.5%)	4 (7.7%)
Viral Load			
<50 Copies		30 (65.2%)	30 (57.7%)
>50 Copies		5 (10.9%)	5 (9.6%)
Missing	6 (100%)	11 (23.9%)	17 (32.7%)
FEV1-Percent-Predicted**			
Mean (SD)	51.8 (18.9)	90.3 (19.0)	85.8 (22.6)
Median [Min, Max]	52.7 [21.0, 79.2]	89.9 [46.7, 128]	86.2 [21.0, 128]
FEV1**			
Mean (SD)	1.66 (0.608)	2.79 (0.846)	2.66 (0.896)
Median [Min, Max]	1.70 [0.650, 2.54]	2.75 [1.32, 4.77]	2.59 [0.650, 4.77]
FEV1/FVC**			
Mean (SD)	0.430 (0.125)	0.708 (0.121)	0.676 (0.150)
Median [Min, Max]	0.437 [0.293, 0.599]	0.722 [0.413, 0.905]	0.684 [0.293, 0.905]
DLCO-Percent Predicted			
Mean (SD)	49.0 (8.49)	75.7 (24.5)	74.2 (24.6)
Median [Min, Max]	49.0 [43.0, 55.0]	74.9 [14.4, 139]	74.5 [14.4, 139]
Missing	4 (66.7%)	7 (15.2%)	11 (21.2%)
Case-Control Status*			
Case	6 (100%)	20 (43.5%)	26 (50.0%)
Control	0 (0%)	26 (56.5%)	26 (50.0%)