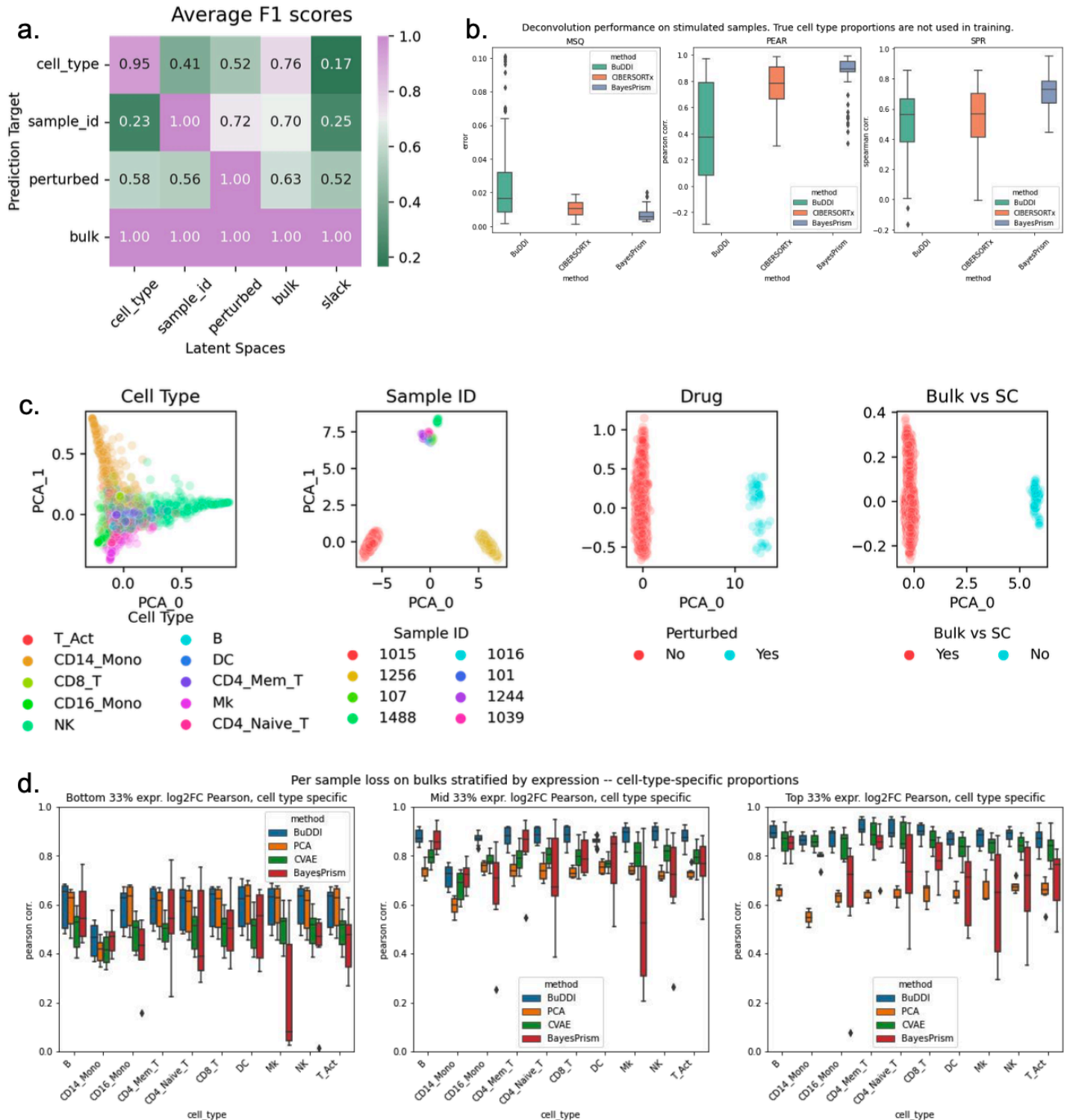# Supplementary Figures and Tables

Supp Figure 1. Latent space analysis of BuDDI on Kang et al. data set with an experimental design where bulk samples are correlated with the sample IDs and perturbation status. **Panel a** depicts that average F1 score of each latent space to predict each source of variation. Midpoint coloration is the average across all observed F1 scores. Panel b compares the performance of BuDDI, CIBERSORTx, and BayesPrism, in estimating the cell type proportions. Panel c depicts each of BuDDI's latent spaces, colored by source of variation. Panel d depicts the Pearson correlation of the simulated perturbation expression, stratified by expression level.
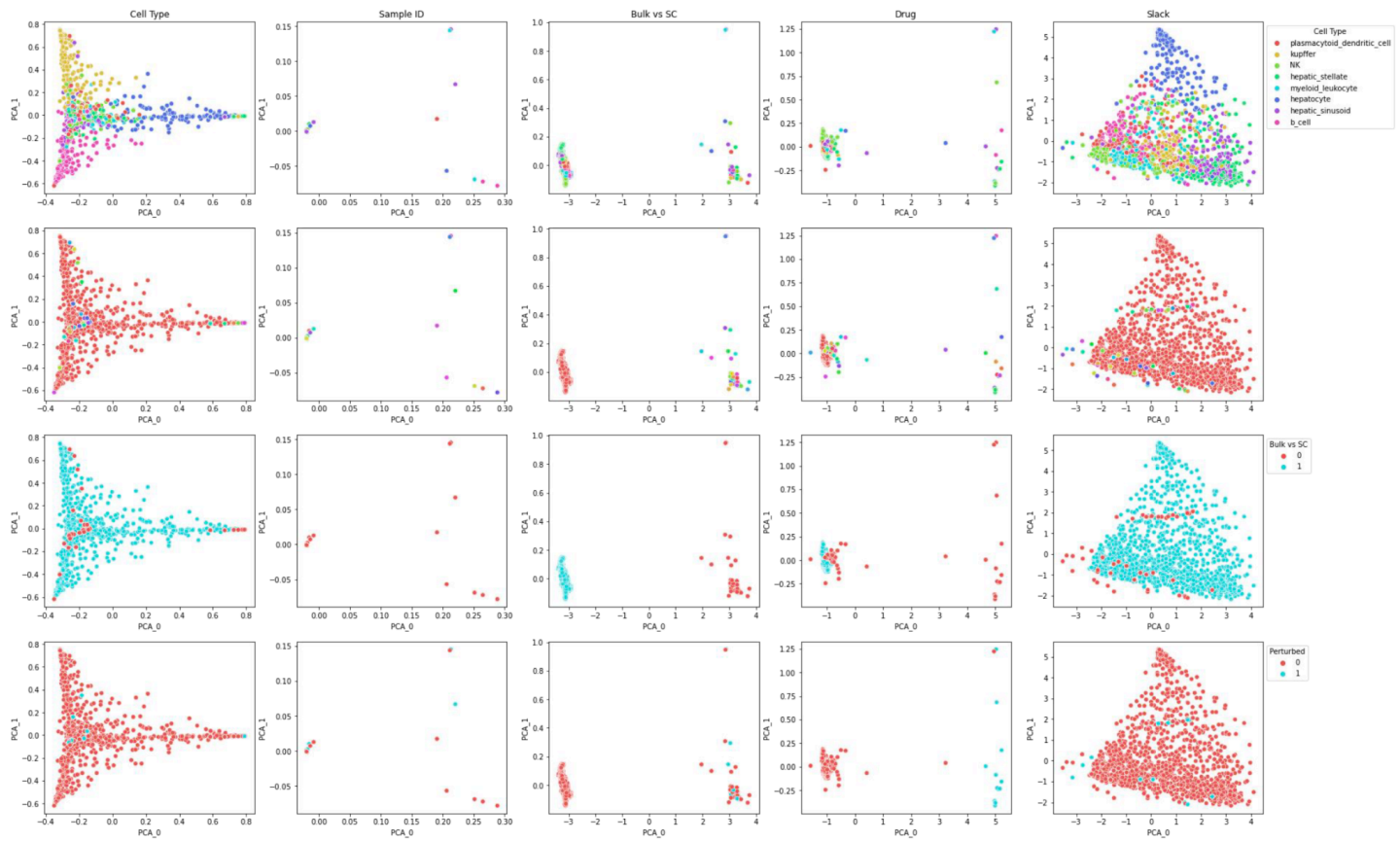
Supp Figure 2. Latent space analysis of BuDDI on Tabula Muris Senis dataset. Each column is a latent space and each row is colored by a source of variation. The second row is colored by sample ID, but due to the number of bulk samples, we omit the sample ID legend.
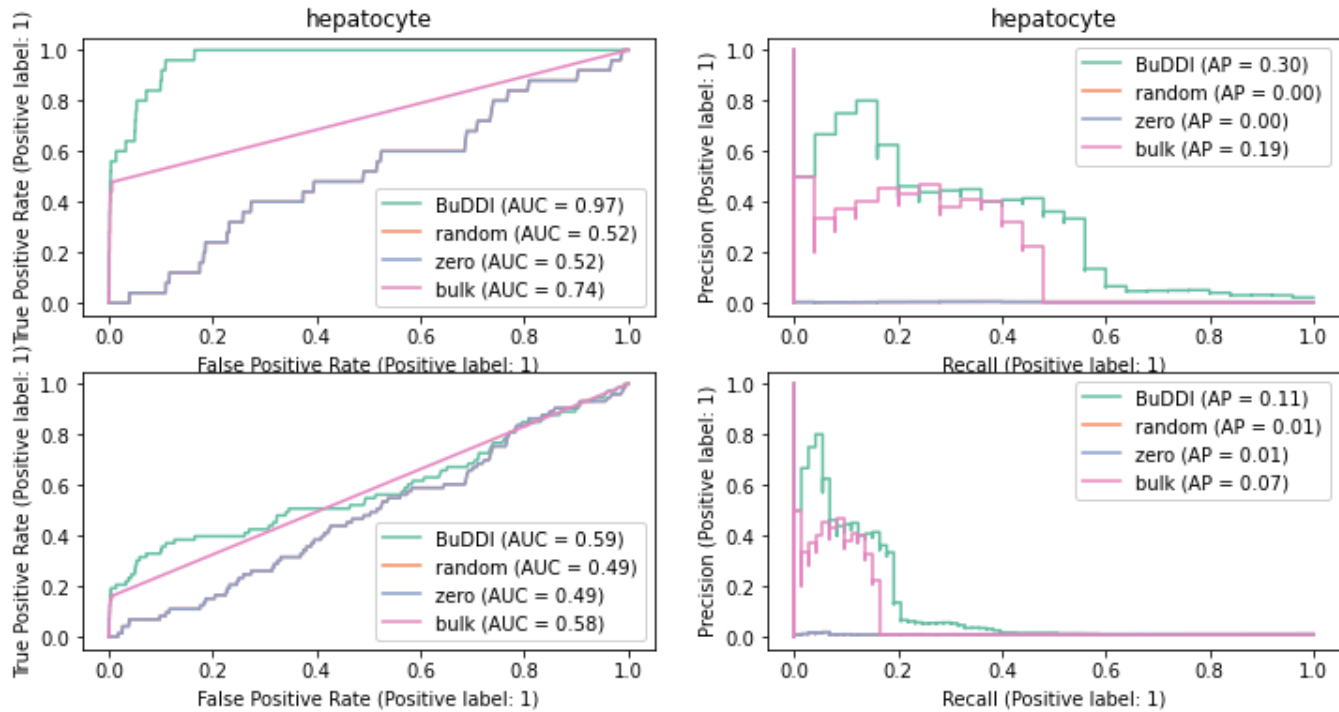
Supp Figure 3. ROC and PR curves for predicting differentially expressed genes between sexes in hepatocytes using BuDDI. Top row uses the differential expressed genes form an independent single-nucleus experiment[46] as the ground truth, bottom row uses the union of the single-nucleus and our calculated single-cell results from Tabula Muris Senis[41,42] as the ground truth.
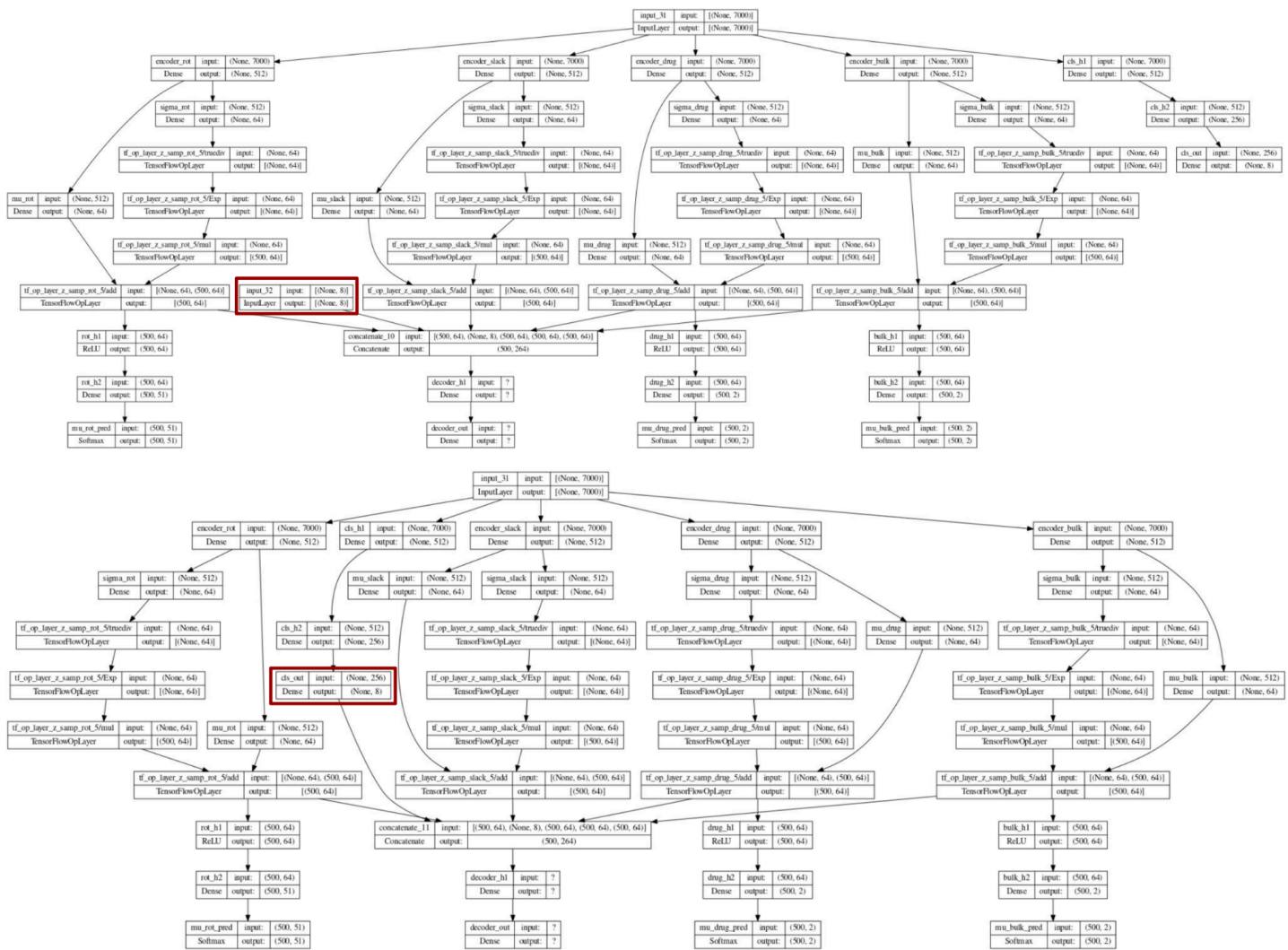
Supp Figure 4. BuDDI model overview for the supervised (top) and unsupervised (bottom) models. The red box highlights the true or estimated cell type proportions used in BuDDI.
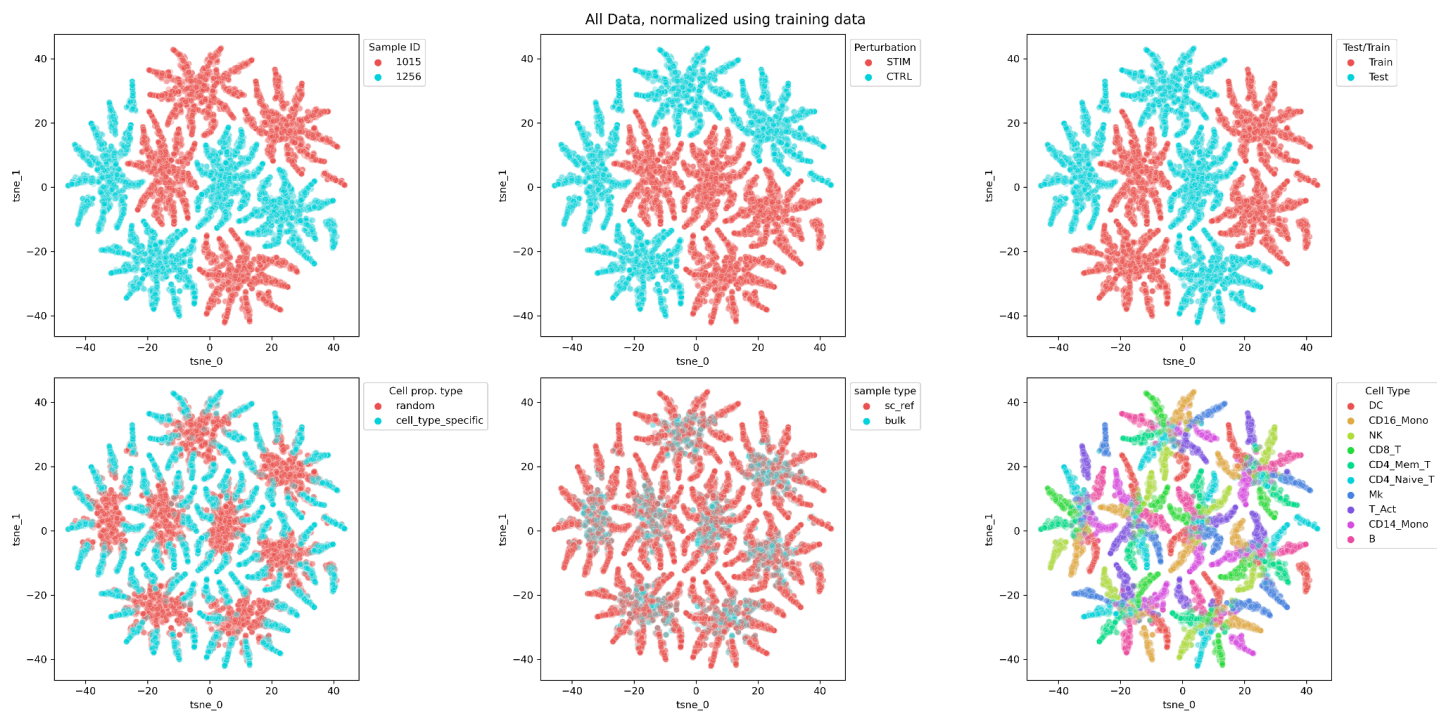
Supp Figure 5. Pseudobulk data generated and colored by source of variation. Our generated data shows independence between, each source of variation, including cell type proportion.
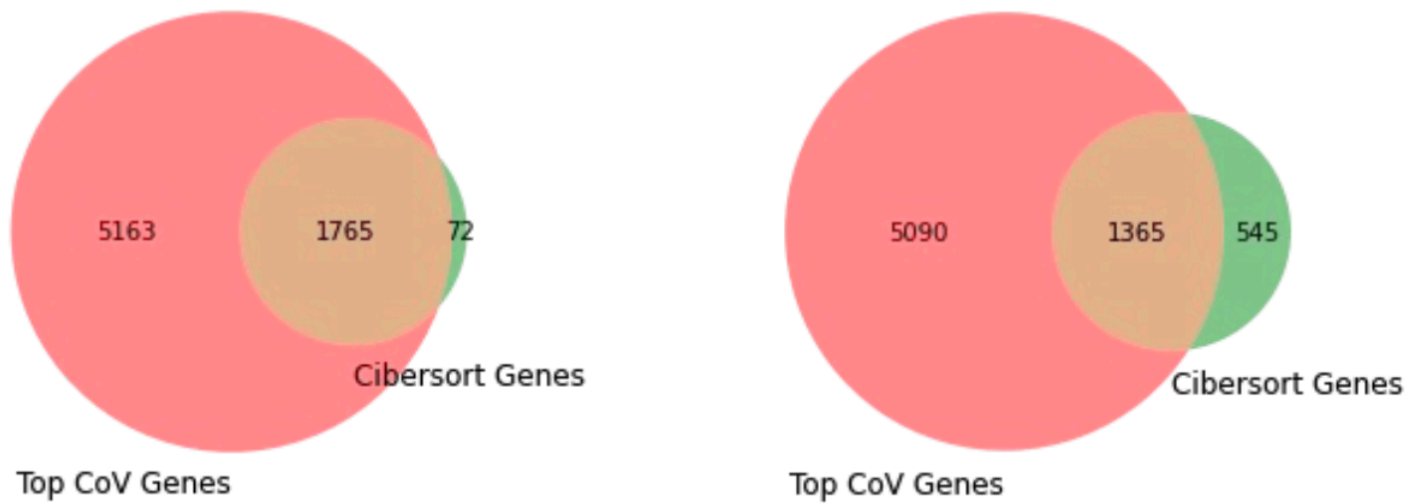
Supp Figure 6. Overlap of top coefficient of variation genes and CIBERSORTx signature genes used in the Kang et al. (left) and sex-dependent liver (right) analyses.
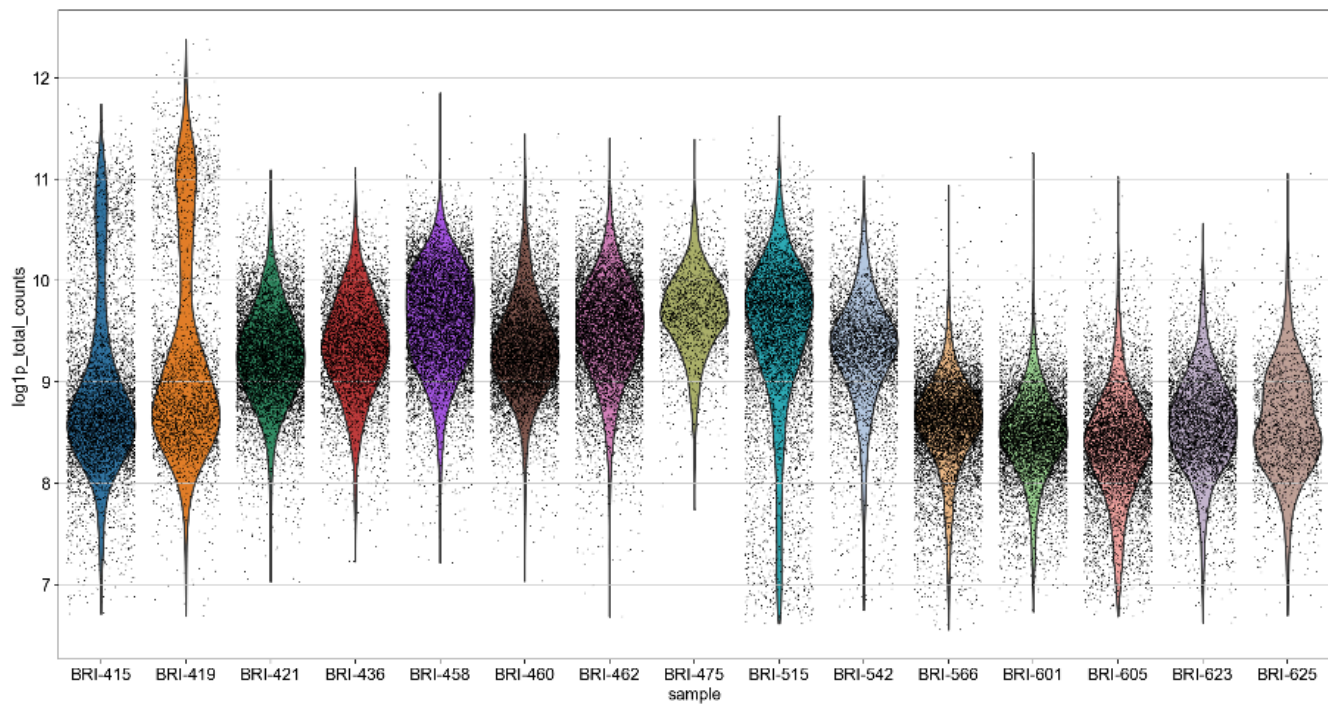
Supp Figure 7. Log total counts for each single-cell synovium sample from Zhang et al.[3]. Only samples with sufficient expression were used in our analysis, this includes samples 421, 436, 458, 460, 462, 475, 515, and 542.

Cell Type in Kang et al.	Number of cells after filtering	Sample ID in Kang et al.	Number of cells after filtering
CD14 Monocyte	4361	1015	3177
CD4 Naive T cells	2504	1256	2396
CD4 Memory T cells	1762	1499	2280
B cells	1363	1244	2031
CD16 Monocytes	1044	1016	1484
CD8 T cells	813	101	1224
T Activated	633	1039	679
NK	619	107	668
CD	604		
Mk	236		

Supp Table 1. Number of cells by cell type and by sample ID in the Kang et al. dataset after filtering.

cell_ontology_class mouse.id	B cell	Kupffer cell	NK cell	duct epithelial cell	endothelial cell of hepatic sinusoid	hepatic stellate cell	hepatocyte	myeloid leukocyte	plasmacytoid dendritic cell
1-M-62	1	5	2	0	25	2	434	1	1
1-M-63	16	609	245	0	495	26	881	34	14
3-F-56	0	12	2	0	9	1	363	0	0
3-F-57	0	1	0	0	5	0	173	0	0
3-M-8/9	0	0	0	2	5	0	453	0	0
18-F-51	24	221	109	0	27	2	267	40	8
21-F-54	39	25	32	0	14	2	250	4	1
24-M-58	0	0	0	0	2	0	74	0	0
24-M-59	3	18	7	0	5	1	4	5	0
30-M-3	33	1608	56	0	3	0	1	177	7
30-M-4	14	3	11	0	5	0	0	12	1
30-M-5	73	44	88	0	80	5	29	34	9

Supp Table 2. Number of cells by sample ID and cell type after filtering and before combining the two cell types "endothelial cell of hepatic sinusoid" and "duct epithelial cell"

characteristics: sex	f	m
characteristics: age		
12	2.0	4.0
15	2.0	4.0
18	2.0	4.0
21	2.0	4.0
24	NaN	3.0
27	NaN	4.0
3	2.0	4.0
6	2.0	4.0
9	2.0	4.0

Supp Table 3. Number of bulk liver samples used in analysis by sample ID and age.

cell_type	B	Endothelial	Fibroblast	Myeloid	NK	T_CD4	T_CD8	T_other
sample								
BRI-415	1830	139	855	800	197	2567	703	279
BRI-419	493	242	558	1385	127	715	270	107
BRI-421	427	283	1795	2935	432	2059	772	328
BRI-436	1121	79	68	741	157	1827	1085	577
BRI-458	1322	319	548	4286	59	1791	743	259
BRI-460	811	903	5378	1363	110	297	182	59
BRI-462	771	250	3216	4923	136	858	346	148
BRI-475	87	90	303	941	183	335	263	56
BRI-515	1721	461	1956	546	384	1675	890	311
BRI-542	594	496	293	1175	161	1246	744	271
BRI-566	455	200	2169	299	407	2895	2574	639
BRI-601	253	180	4032	2477	127	834	229	112
BRI-605	2339	558	2083	924	150	1265	412	240
BRI-623	250	317	82	2926	313	991	832	333
BRI-625	80	451	267	595	119	522	193	96

Supp Table 4. Number of cells by sample ID and cell type from Zhang et. al.