

Supplementary Data

Single-Cell Transcriptome Analysis Identifies Novel Biomarkers Involved in Major Liver Cancer Subtypes

Asish Kumar Swain¹, Prashant Pandey¹, Riddhi Sera¹, and Pankaj Yadav^{1*}

¹Department of Bioscience Bioengineering, Indian Institute of Technology, Jodhpur, 342037,
Rajasthan, India

*Corresponding author: pyadav@iitj.ac.in

Contents

1	Supplementary figures	3
2	Supplementary table	9

1 Supplementary figures

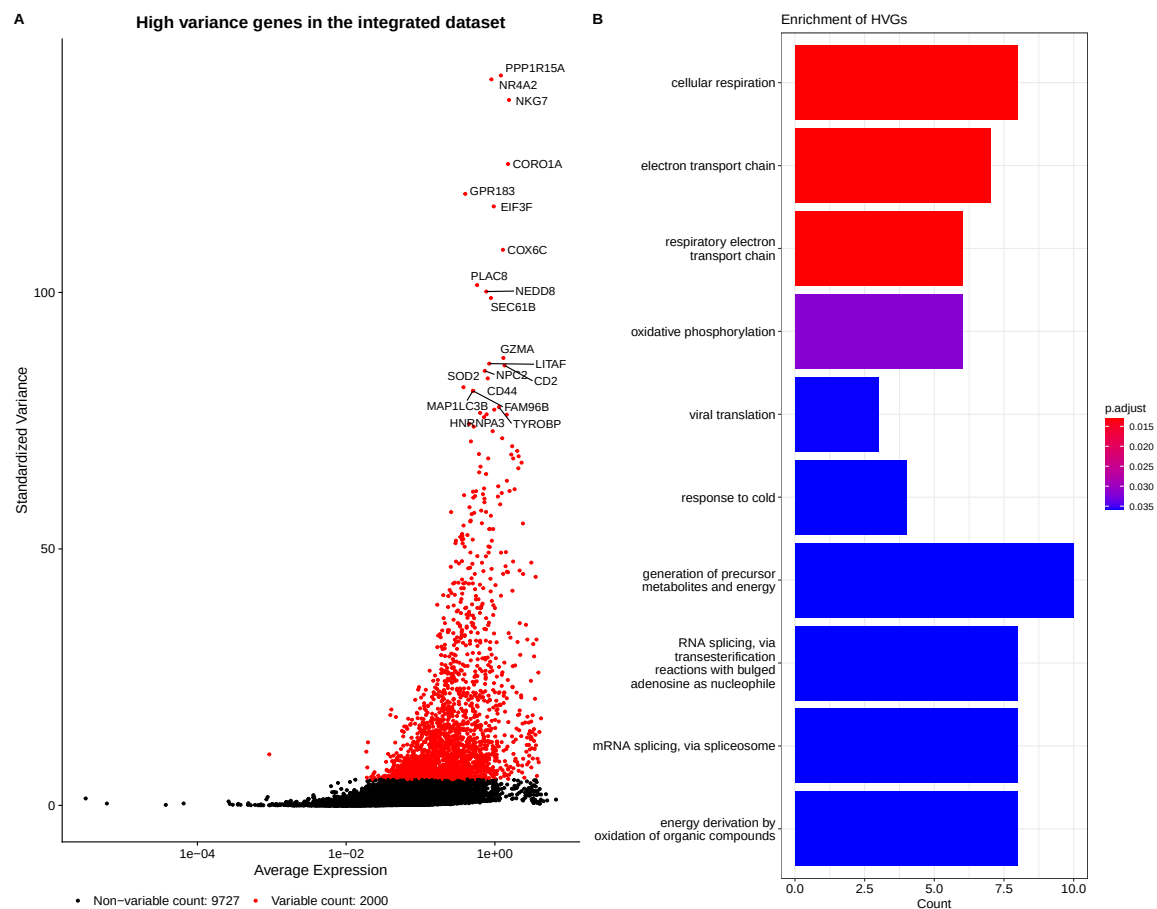


Figure 1: A) Top 2000 high variance genes (in red), where top 20 genes were labelled B) enrichment analysis of top HVGs shows pathways affected by top HVGs in the integrated dataset.

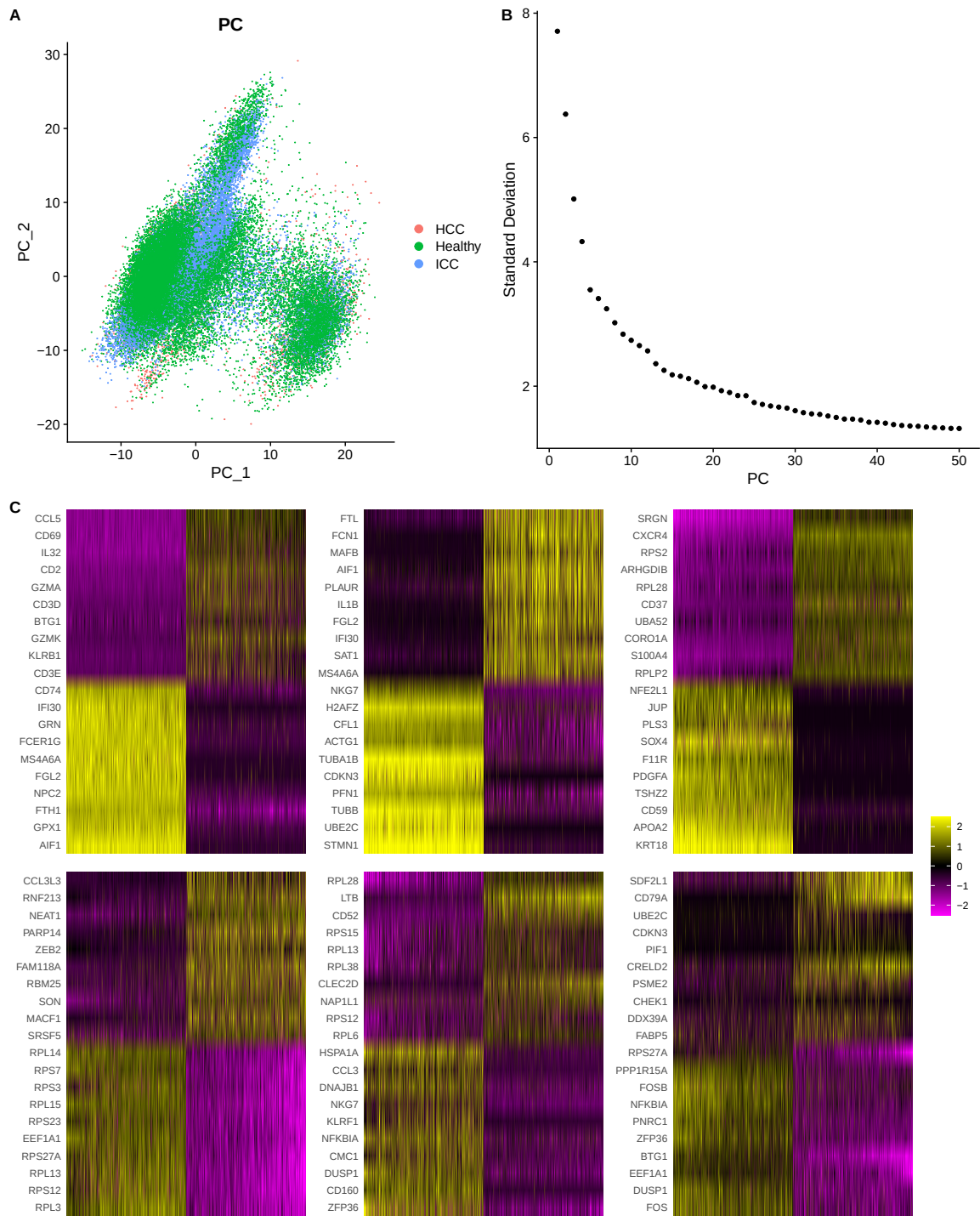


Figure 2: A) After dimension reduction by PCA, scatter plot of expression levels of latent feature PC1 and PC2 B) standard deviation of top 50 principal components of the integrated dataset C) heatmap of genes associated with first 6 PCs.

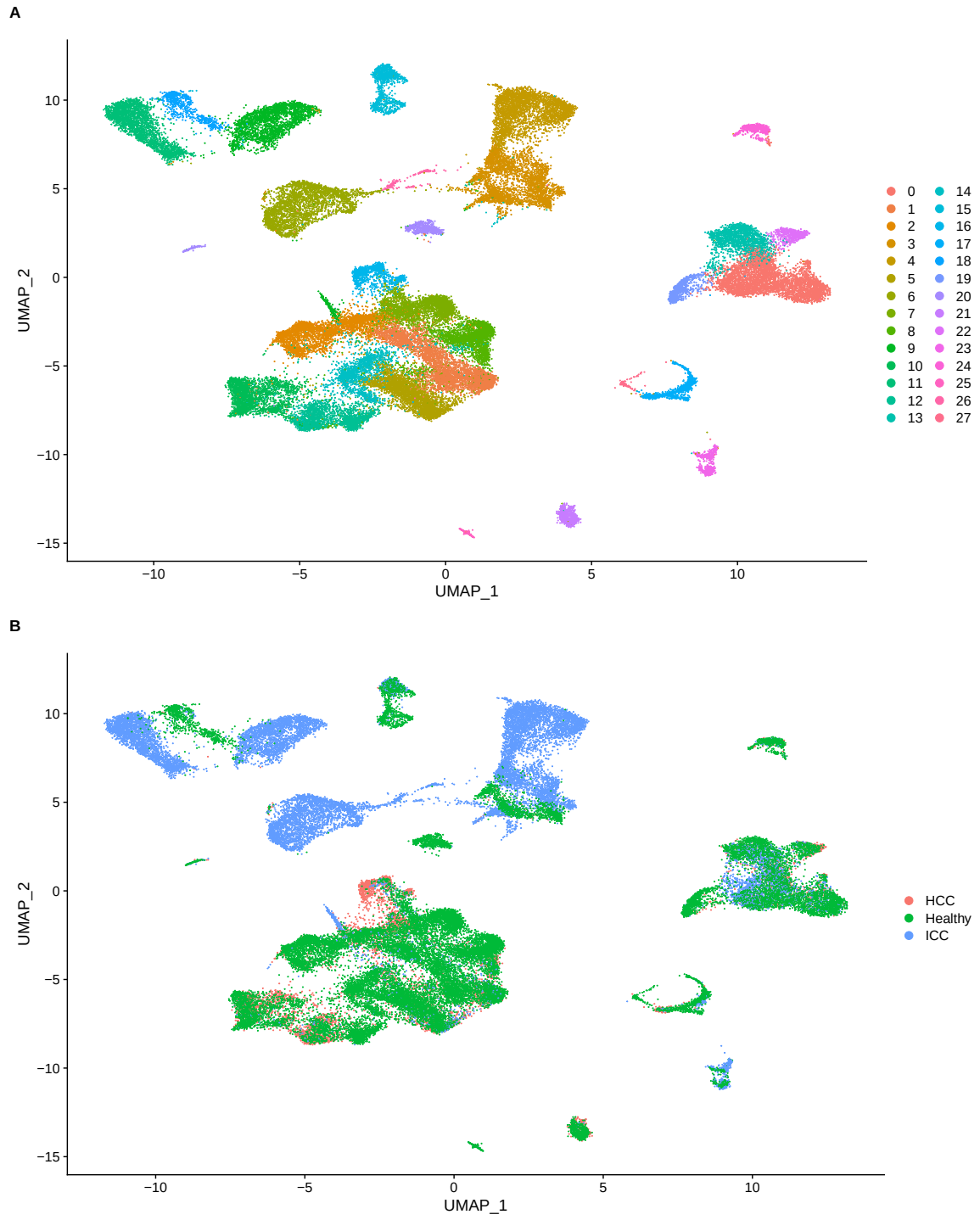


Figure 3: UMAP plot of integrated dataset A) cluster wise B) LC subtype wise.

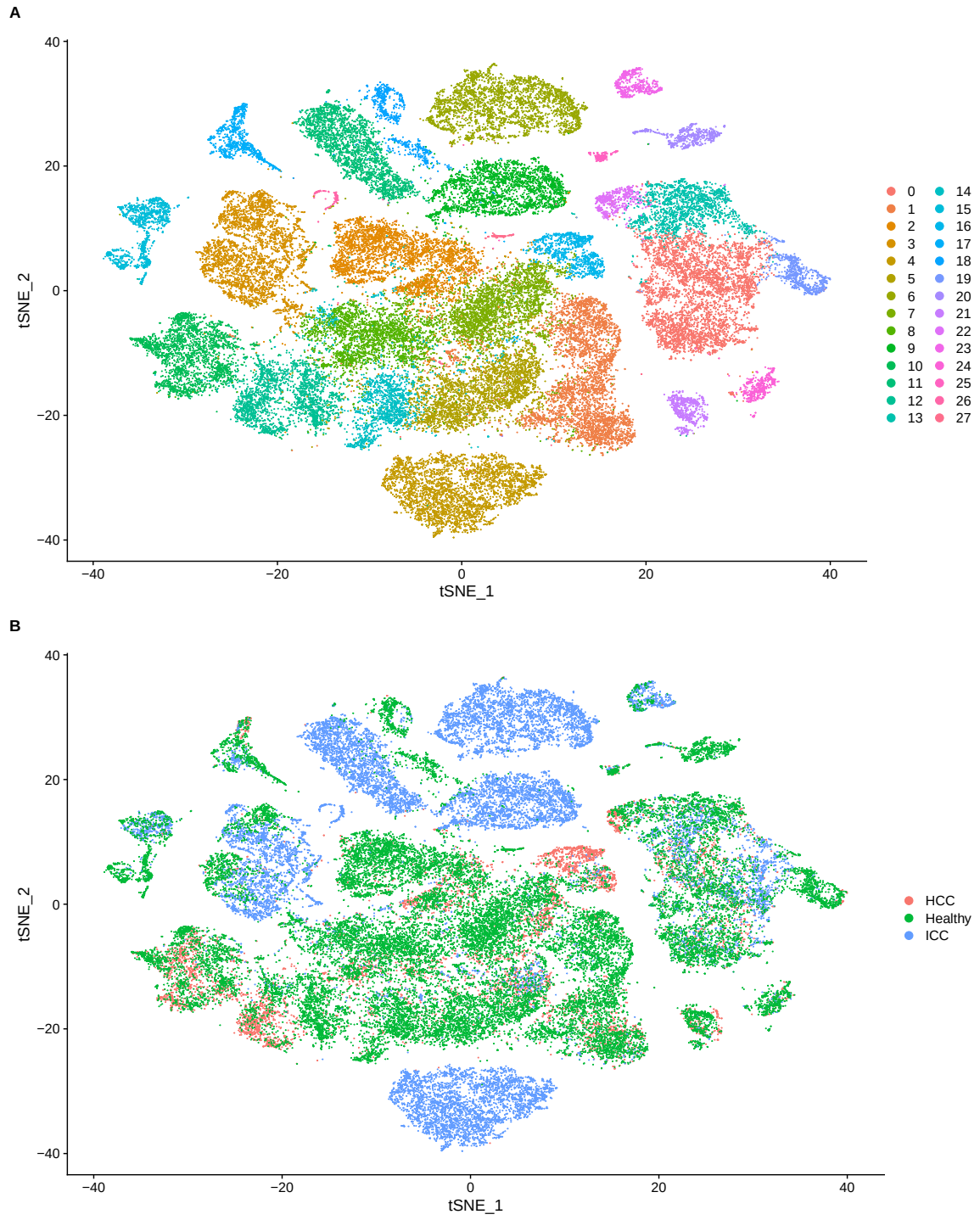


Figure 4: t-SNE plot of integrated object A) cluster wise B) LC subtype wise.

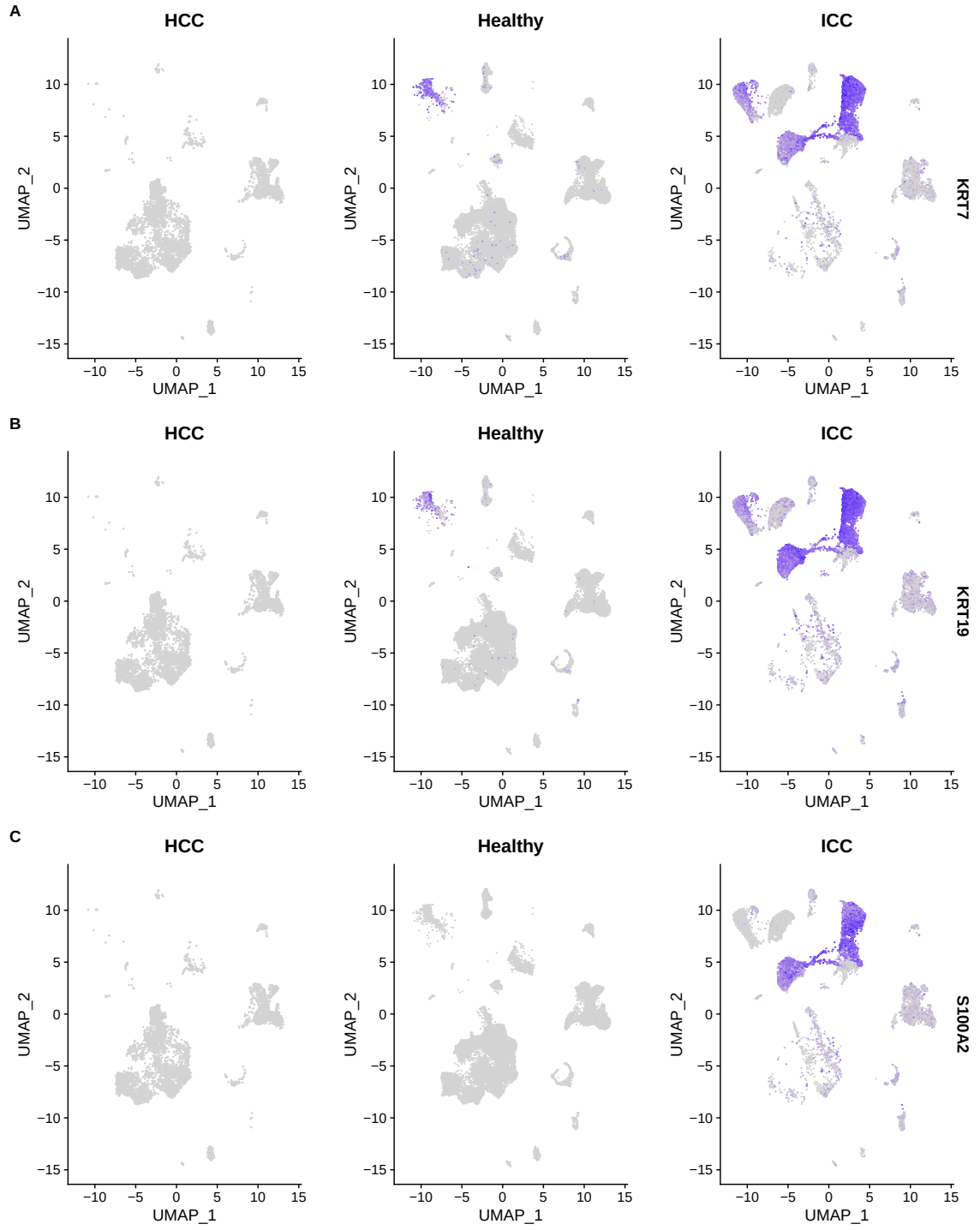
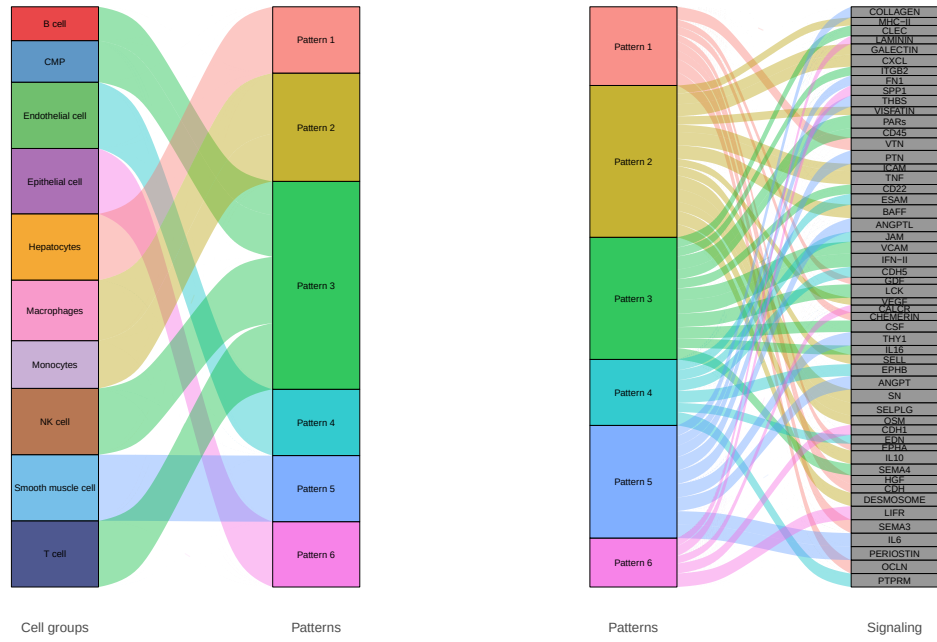


Figure 5: Expression of KRT7, KRT19, and S100A2 in the UMAP plot, where higher expression of these epithelial markers observed in ICC subtype.

A

Outgoing communication patterns of secreting cells



B

Incoming communication patterns of target cells

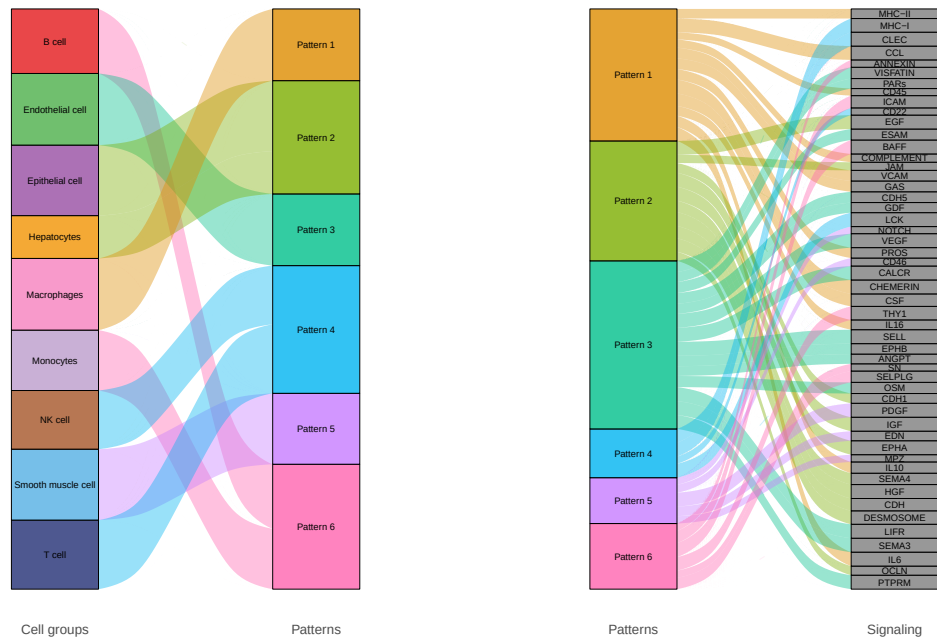


Figure 6: A) Alluvial plot showing outgoing communication pattern of cell types B) plot showing incoming communication patterns of target cells in integrated dataset.

2 Supplementary table

Gene name	p_val	avg_log2FC	pct.1	pct.2	p_val_adj	Sub_type
"HLA-DRB5"	0	-0.512659788979458	0.895	0.143	0	"HCC"
"KLRC1"	0	-0.568181384179069	0.665	0.099	0	"HCC"
"KLRG1"	1.98661169642167e-265	-0.65340418205636	0.895	0.29	2.32969953639369e-261	"HCC"
"TNFRSF18"	1.07624415504335e-237	0.745014569255751	0.31	0.031	1.26211152061934e-233	"HCC"
"CCL3L3"	2.55268167142766e-147	-1.04772718493381	0.877	0.326	2.99352979608322e-143	"HCC"
"HSPA6"	1.24580747628825e-90	1.11145130387718	0.406	0.057	1.46095842744323e-86	"HCC"
"GNLY"	2.41076092173296e-59	-0.601438540839928	0.707	0.219	2.82709933291624e-55	"HCC"
"TNFRSF4"	7.06429565371028e-44	0.75663051472647	0.411	0.028	8.28429951310605e-40	"HCC"
"S100A8"	3.48580185590846e-12	-0.757451591054402	0.451	0.068	4.08779983642386e-08	"HCC"
"CXCL13"	0.000617646084054636	0.529483788992033	0.229	0.008	1	"HCC"
"S100A9"	0.984601578378236	-0.700042608373819	0.506	0.091	1	"HCC"
"GNLY1"	0	-1.17071909987427	0.851	0.764	0	"ICC"
"CXCR4"	0	-1.2559929754671	0.989	0.972	0	"ICC"
"RGS1"	1.64045907878347e-277	-1.26632128650779	0.993	0.929	1.92376636168938e-273	"ICC"
"CD3D"	5.8372391548254e-224	-1.10659001232771	0.942	0.758	6.84533035686375e-220	"ICC"
"CCL3"	7.62900315734336e-98	-0.986650607256853	0.985	0.717	8.94653200261656e-94	"ICC"
"IL7R"	2.02423793561019e-66	-1.17858220774852	0.975	0.779	2.37382382709007e-62	"ICC"
"CST3"	4.75056109615365e-62	-0.989171644551047	0.838	0.529	5.57098299745938e-58	"ICC"
"DUSP2"	4.54494425939185e-46	-0.999021272648927	0.989	0.971	5.32985613298882e-42	"ICC"
"SPP1"	6.44767392136254e-39	0.969202868921665	0.452	0.425	7.56118720758186e-35	"ICC"
"CD69"	1.09191060314982e-24	-1.03879529271115	0.994	0.927	1.28048356431379e-20	"ICC"
"CCL4"	3.94881731947252e-06	-1.35616033990723	0.991	0.903	0.0463077807054543	"ICC"
"GZMK"	0.652323755085457	-1.05242523357603	0.957	0.718	1	"ICC"
"KLRB1"	0.660927863446059	-1.29619571383765	0.959	0.767	1	"ICC"

Table 1: Top differentially expressed genes between healthy and cancer subtypes (HCC and ICC) where avg_log2FC shows the log fold-change of the average expression between the two groups. pct.1 explains the percentage of cells where the feature is detected in the first group, pct.2 explains the percentage of cells where the feature is detected in the second group. p_val_adj explains adjusted p-value, based on Bonferroni correction using all features in the dataset. p-value 2.225e-308 assigned as 0 in R.