Hypothesis

Clinicopathological info of kidney patients

Patient_ID	sex	age	stage	metastases	grade	necrosis	sarcomatoid changes	lymphovascular invasion	Leibovich	Histology	VHL	PBRM1	BAP1	SETD2	TSC2	KDM5C
PD43824	male	41-50	1b	0	2	no	no	no	2	ccRCC	ns-sub					
PD43948	female	71-80	3a	1	4	yes	yes	yes	8	ccRCC					bifs	
PD44714	male	51-60	NA	0	NA				NA	Benign						
PD44966	male	51-60	1a	0	3	no	no	no	1	ccRCC	fs		fs			
PD44967	male	71-80	NA	1	4	no	no	no	8	ccRCC	-	-	-	-	-	-
PD45814	male	61-70	3a	0	4	yes	yes	yes	8	ccRCC	fs	fs				
PD45815	male	51-60	3a	0	2	no	no	no	4	ccRCC	fs					
PD45816	female	71-80	3a	0	4	yes	yes	yes	9	ccRCC			ns-sub			
PD47171	female	51-60	3a	0	4	yes	yes	yes	7	ccRCC	ns-sub	fs	ns-sub			
PD47172	male	51-60	NA	0	NA				NA	oncocytoma						
PD47465	female	61-70	3a	0	3	no	no	yes	5	ccRCC	ns-sub	fs		fs		ns-sub
PD47512	male	51-60	3a	0	4	yes	yes	yes	8	ccRCC	fs	ns-sub		ns-sub		

Malfunction of **VHL**, a crucial driver of ccRCC, leads to accumulation HIF-2α
TSC2 frameshift mutation leads to loss of mTOR pathway suppression

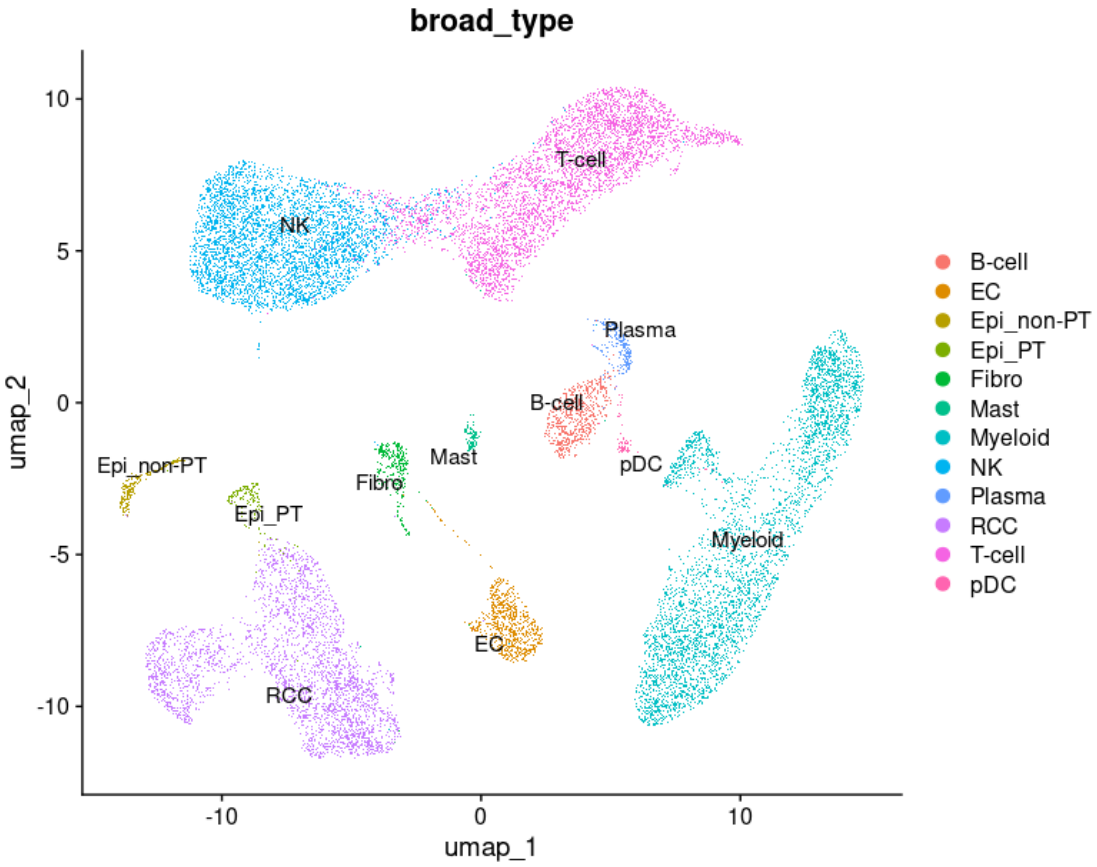
VHL (Von Hippel-Lindau)
TSC2 (Tuberous Sclerosis Complex)

Epi_non-PT : non-proximal tubule epithelial
Epi_PT : proximal tubule epithelial
RCC = renal cell carcinoma

Reference Paper : Chengheng Liao *et al.*, Nature Reviews Urology, 2024
Data retrieved from comprehensive single cell atlas of kidney cancer (Sanger institute)

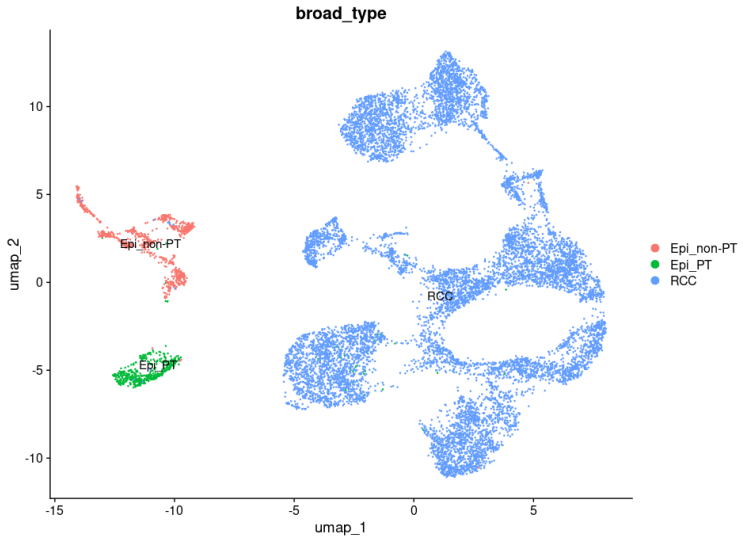
PD43948 also included an adrenal metastasis that was sampled
TSC2 bi-frameshift mutation

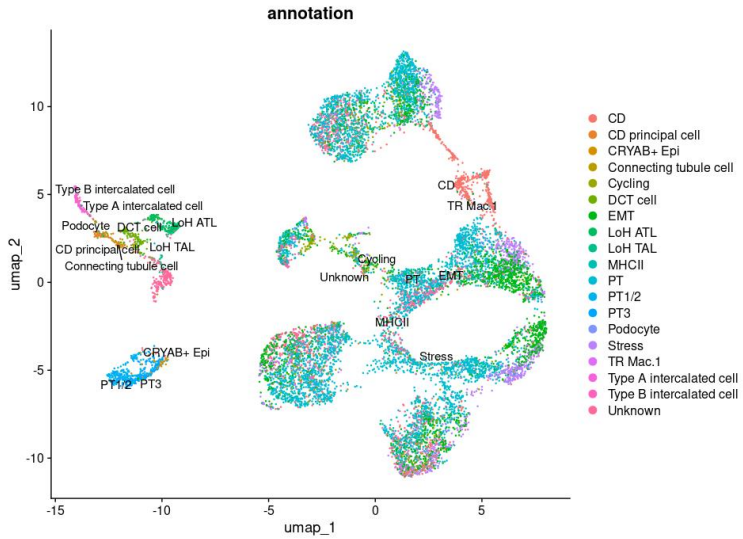
Identification of human ccRCC cell populations

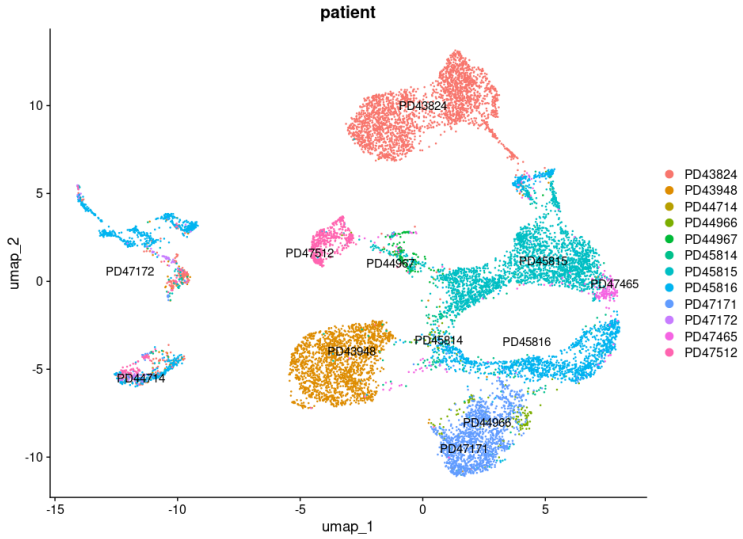


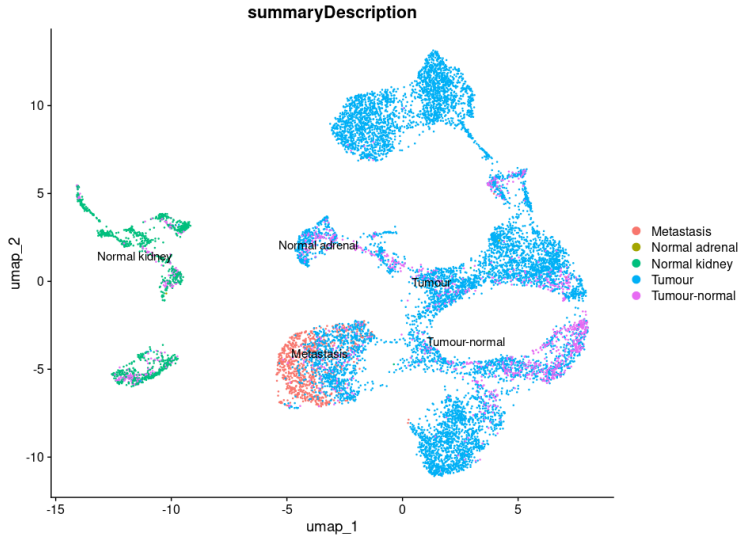
2k HVGs selection followed by HVGs Z-score normalization, Non-harmonized

Epithelial cells of RCC







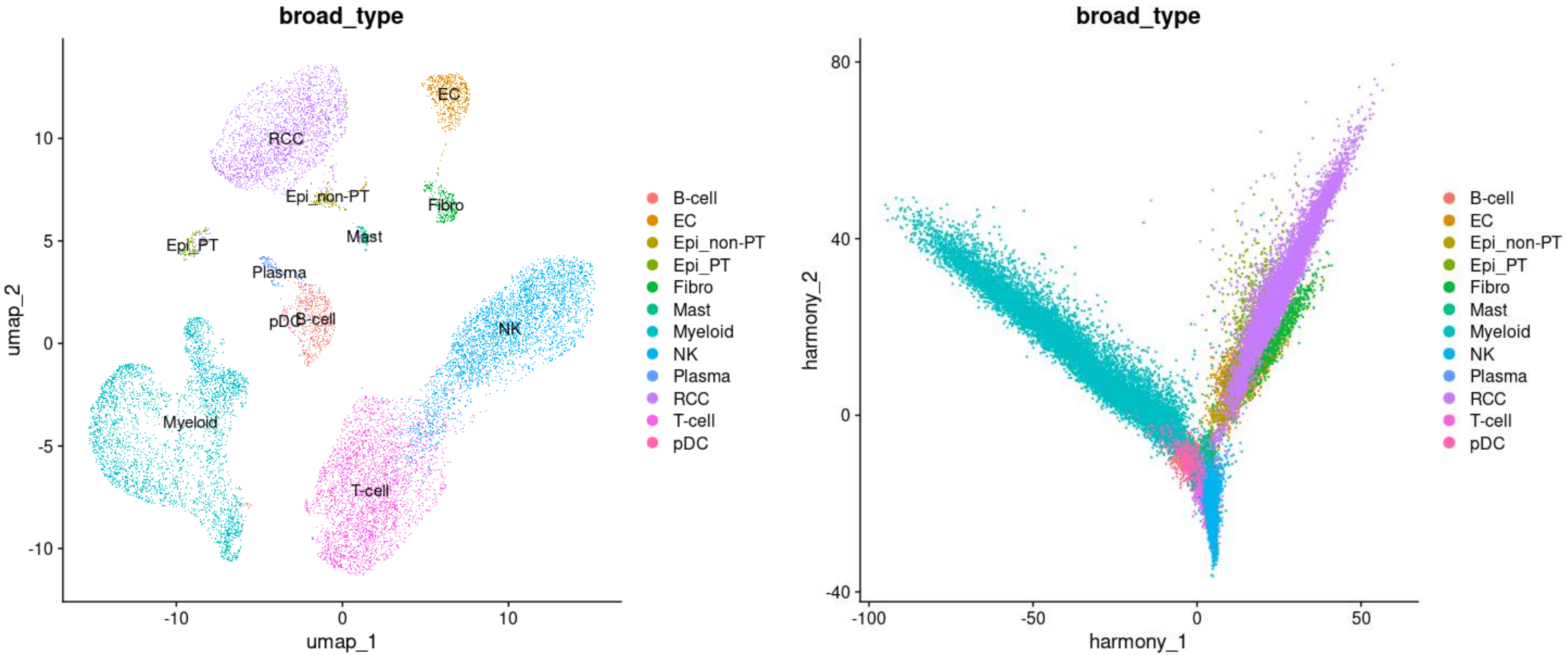


✓ Clusters unique to a donor through heterogenicity can bias the pseudo-time trajectory

→ Harmonization is needed

2k HVGs selection followed by all genes Z- score normalization, Non-harmonized

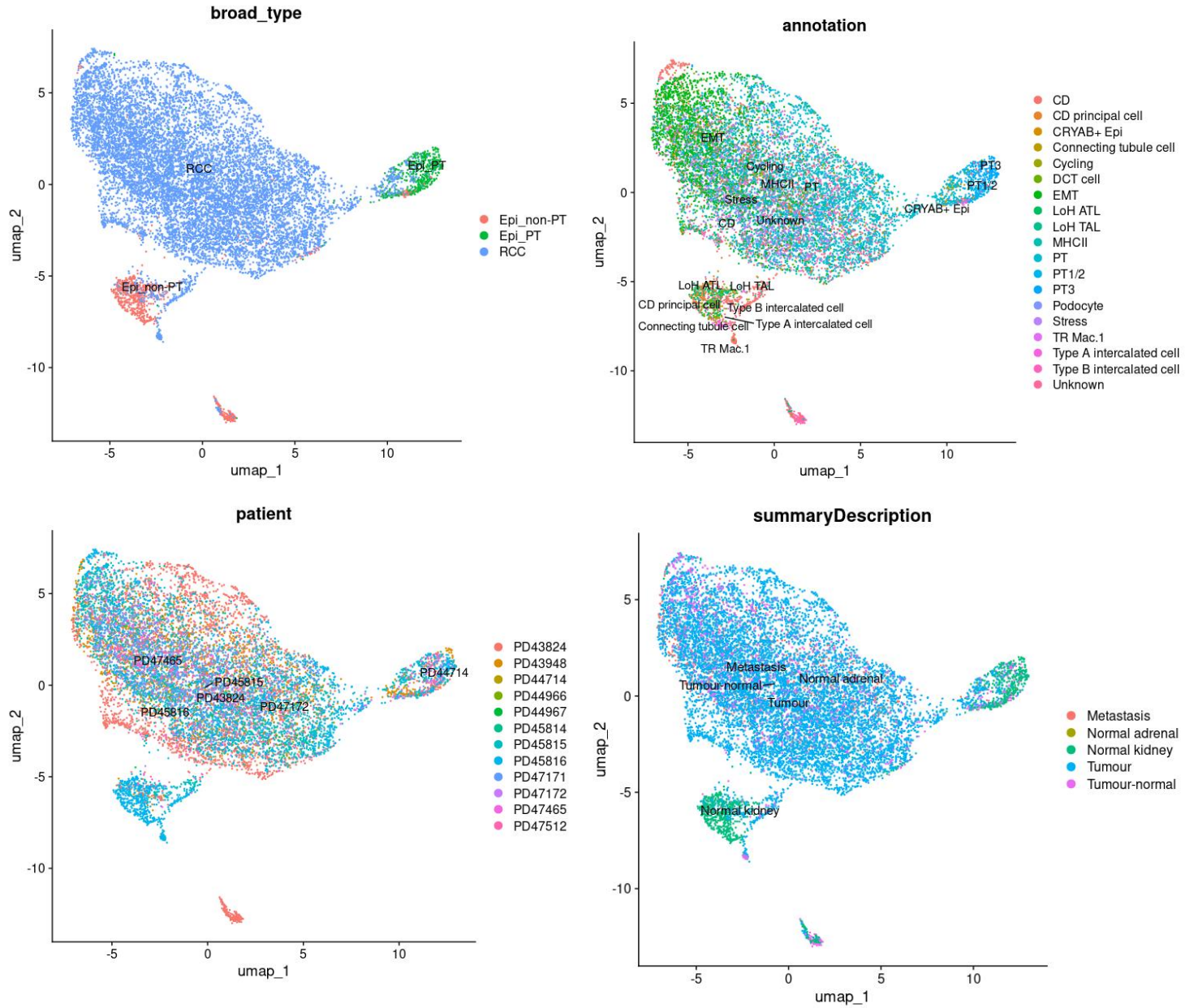
After harmonization



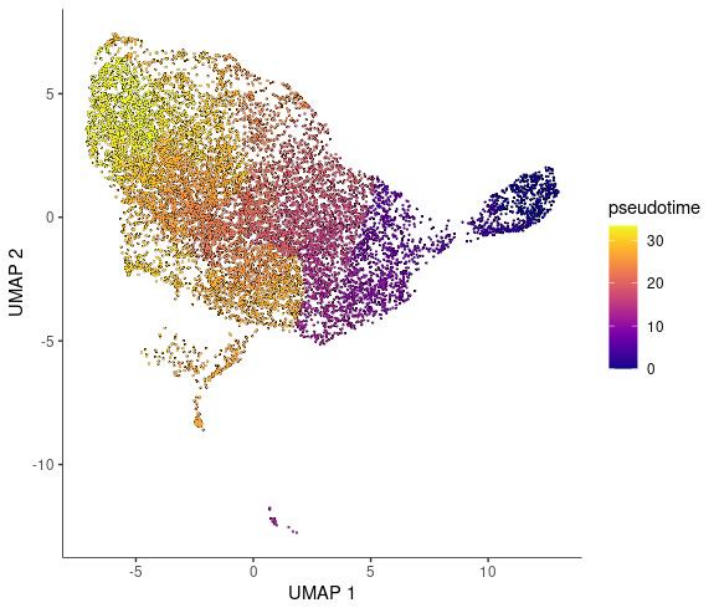
Cancer-Epithelial cells (RCC, Epi_PT, Epi_non-PT, Fibro, EC) & Immune cells (Myeloid, NK, T-cell) are separated clearly

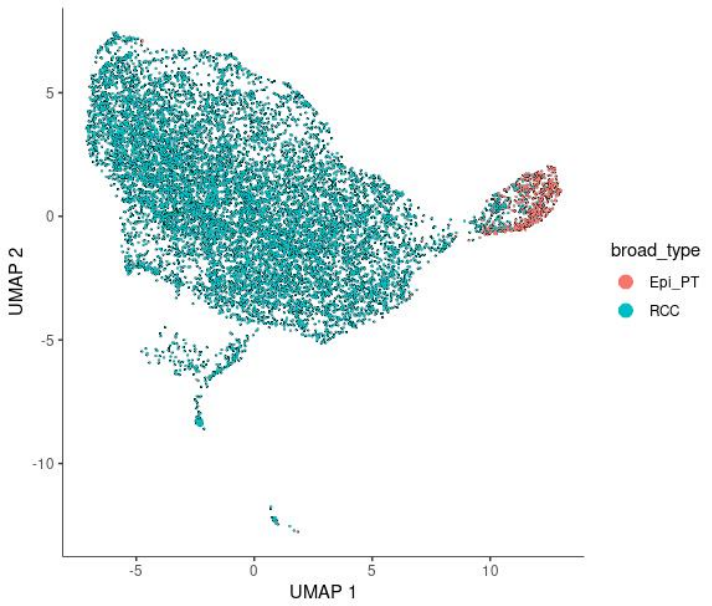
2k HVGs selection followed by HVGs Z-normalization, Harmonized

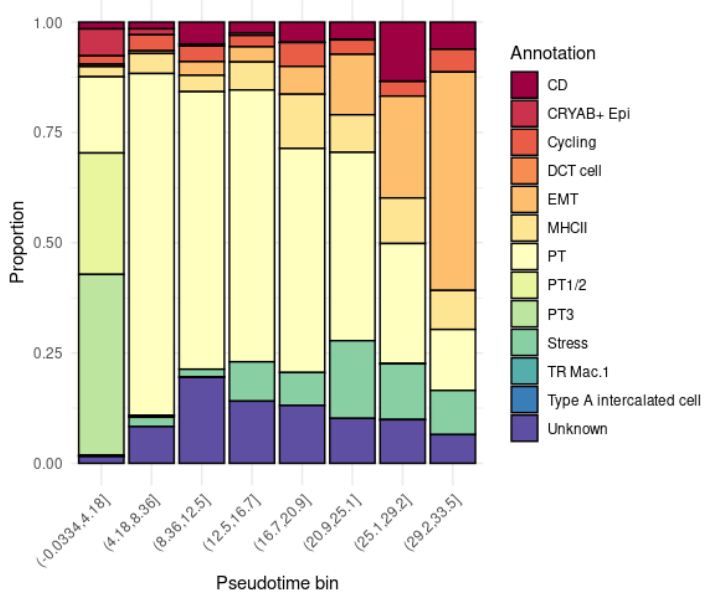
After harmonization

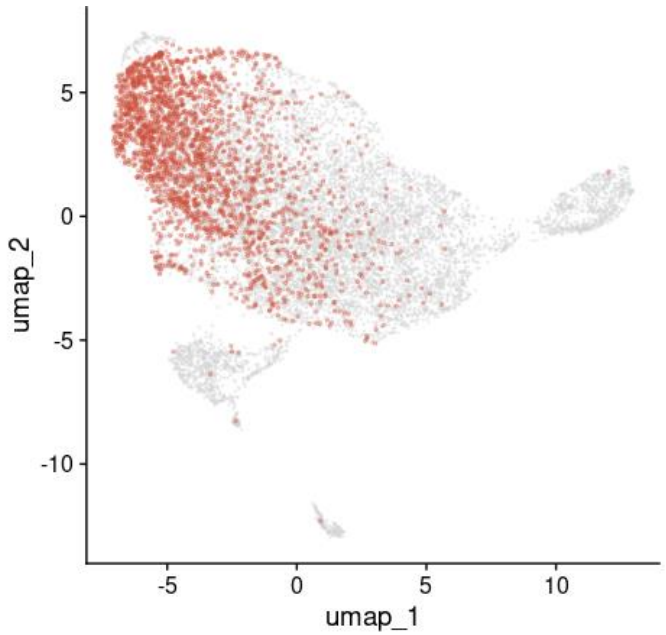


Pseudo-time trajectory: EMT increased & PT decreased







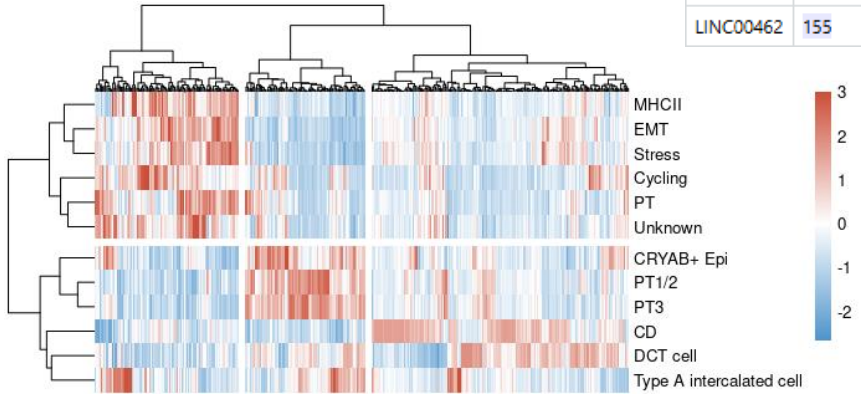


id	module
EFEMP1	306
IL1R2	306
IL1R1	306
GPC1	306
CPE	306
HTRA1	306
SCG5	306
MMP2	306
MT1M	306
MT1A	306
MRC2	306
SLPI	306
CEBPB	306
SPINK1	306
BAALC	306
BTBD16	306
SERPINA3	306
MGAT3	306

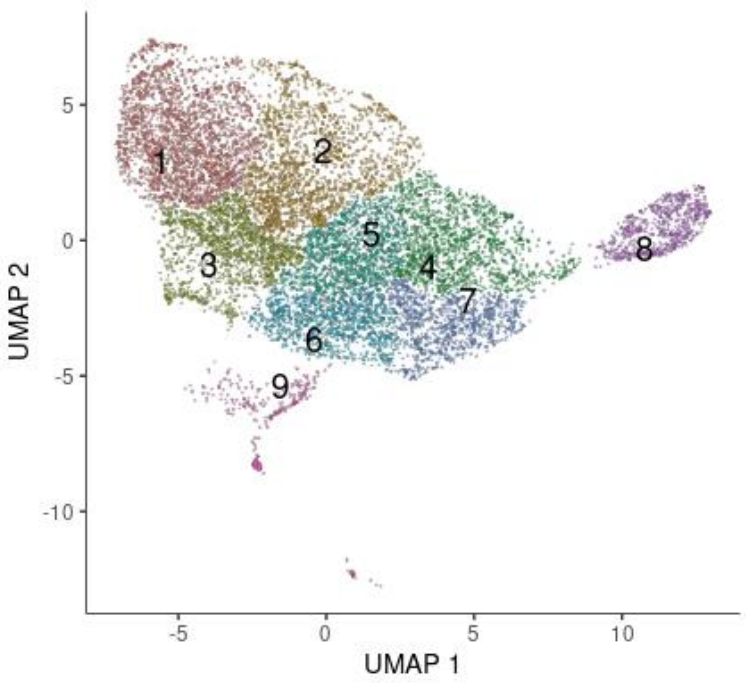
id	module
FHL3	130
F3	130
CD34	130
STEAP3	130
CCL20	130
MAP1B	130
F13A1	130
TREM1	130
SERPINE1	130
SLC38A5	130
LOXL2	130
MMP28	130
COL1A1	130
CAPN12	130
BCAM	130
CHST3	130
COL11A1	130
PODNL1	130

id	module
QPCT	155
TFPI	155
IGFBP2	155
IGFBP5	155
IMPDH2	155
F2R	155
LOX	155
TGFBI	155
SPOCK1	155
C5orf46	155
BET1	155
HILPDA	155
BGN	155
DEFB1	155
SBSPON	155
IARS	155
CRACR2B	155
ADM	155

KDELC2	155
NNMT	155
ACTA2	155
ZFP36L1	155
SERPINA1	155
CCDC64B	155
MT2A	155
MT1E	155
MT1X	155
PMP22	155
TMEM100	155
WFDC2	155
PLTP	155
ZFAS1	155
FSTL3	155
GADD45B	155
DMKN	155
C19orf48	155
LINC01426	155
CADM3	155
SLC34A2	155
LINC00462	155



Clustering



gene_id	gene_short_name	cell_group
MT2A	MT2A	1
MT1E	MT1E	1
MT1X	MT1X	1
RARRES2	RARRES2	2
LDHA	LDHA	2
P4HB	P4HB	2
TMSB10	TMSB10	3
CAV1	CAV1	3
VIM	VIM	3
GSTA2	GSTA2	4
GSTA1	GSTA1	4
CD24	CD24	4
RARRES2	RARRES2	5
VIM	VIM	5
NDUFA4L2	NDUFA4L2	5
LDHA	LDHA	6
CRYAB	CRYAB	6
VIM	VIM	6
TMSB10	TMSB10	7
GNB2L1	GNB2L1	7
GAPDH	GAPDH	7
ALDOB	ALDOB	8
ASS1	ASS1	8
GATM	GATM	8

Normal renal cell (cluster 8)

↓

EMT activation, metabolic reprogramming, hypoxia response (cluster 2~7)

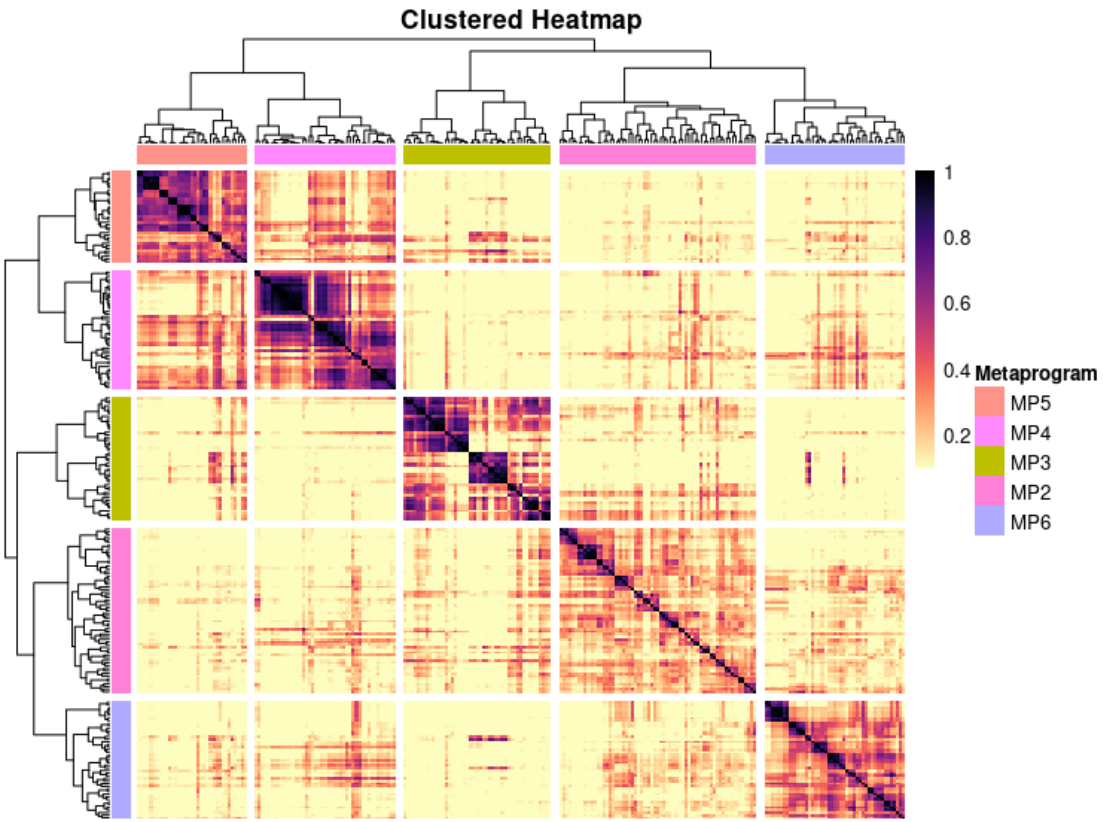
↓

Anti-oxidant defense, Stress-adaptive, adaptive-survival, drug-resistance (cluster 1)

RARRES2: lipid metabolism contributing to tumor growth
LDHA: tumor metastasis facilitated by EMT and Warburg effect
P4HB: cancer progression facilitated by EMT
TMSB10: cancer progression marker regulated by JUN
CAV1: metastasis promotion in RCC

filter(fraction_expressing >= 0.50
top_n(3, pseudo_R2)

NMF for



MP5: IER2, FOS, JUNB, EGR1, FOSB, ZFP36
MP4: NEAT1, MALAT1, DDX17, FOSB, VEGFA, EGR1
MP3: SRGN, HLA-E, CCL5, SH3BGRL3, ZFP36L2, NKG7
MP2: MT1X, MT1G, MT1E, MT1F, MT2A, FOS
MP6: APOE, MT1G, DCXR, ACAA1, PEPD, SUCLG1

↓

Preprocessing for SCENIC analysis

A minimal amount of filtering is needed before running a SCENIC analysis. On a cell level, we examine the number of expressed genes and remove cells that fell into the distribution extremes. In the peripheral blood mononuclear cell (PBMC) study case (Table 1), we discard cells with <200, and more than ~5,000, expressed genes; however, these thresholds must be determined empirically. We further filter out cells that have a large fraction of mitochondrial gene transcripts; these cells are thought to be of lower quality as this is indicative of cell membrane breach[9]. We again use the empirical distribution to select an upper threshold on mitochondrial genes expressed, and this is highly dependent on cell type but typically 5–15%. Finally, on the gene level, genes with low overall expression are removed; with our default settings, we remove genes expressed in fewer than three cells in the data set.

Raw UMI count data without ln(x+1) transformation

↓

Delete ribosomal gene and exclude cell with percent.mt over 10

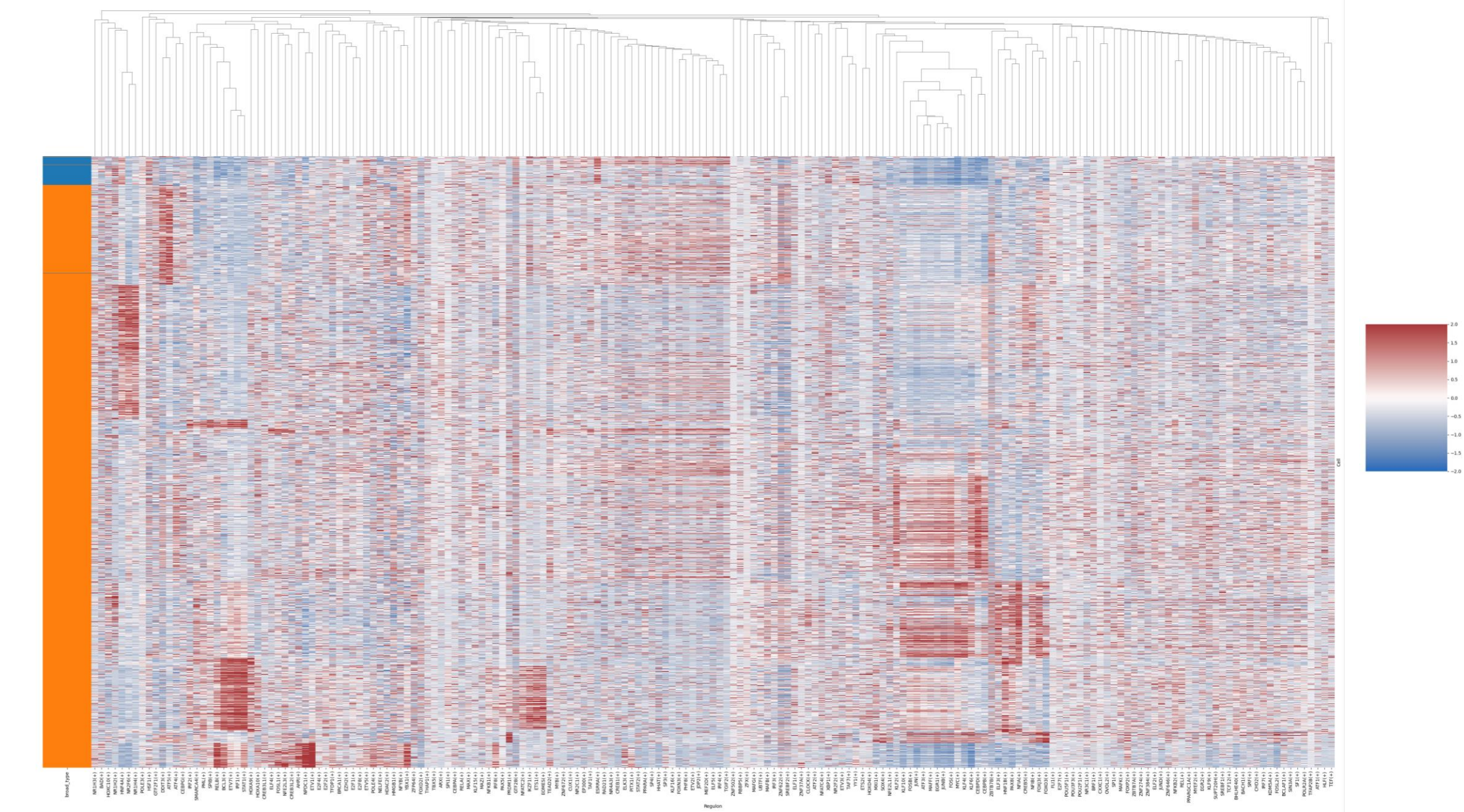
↓

Auxiliary data info :

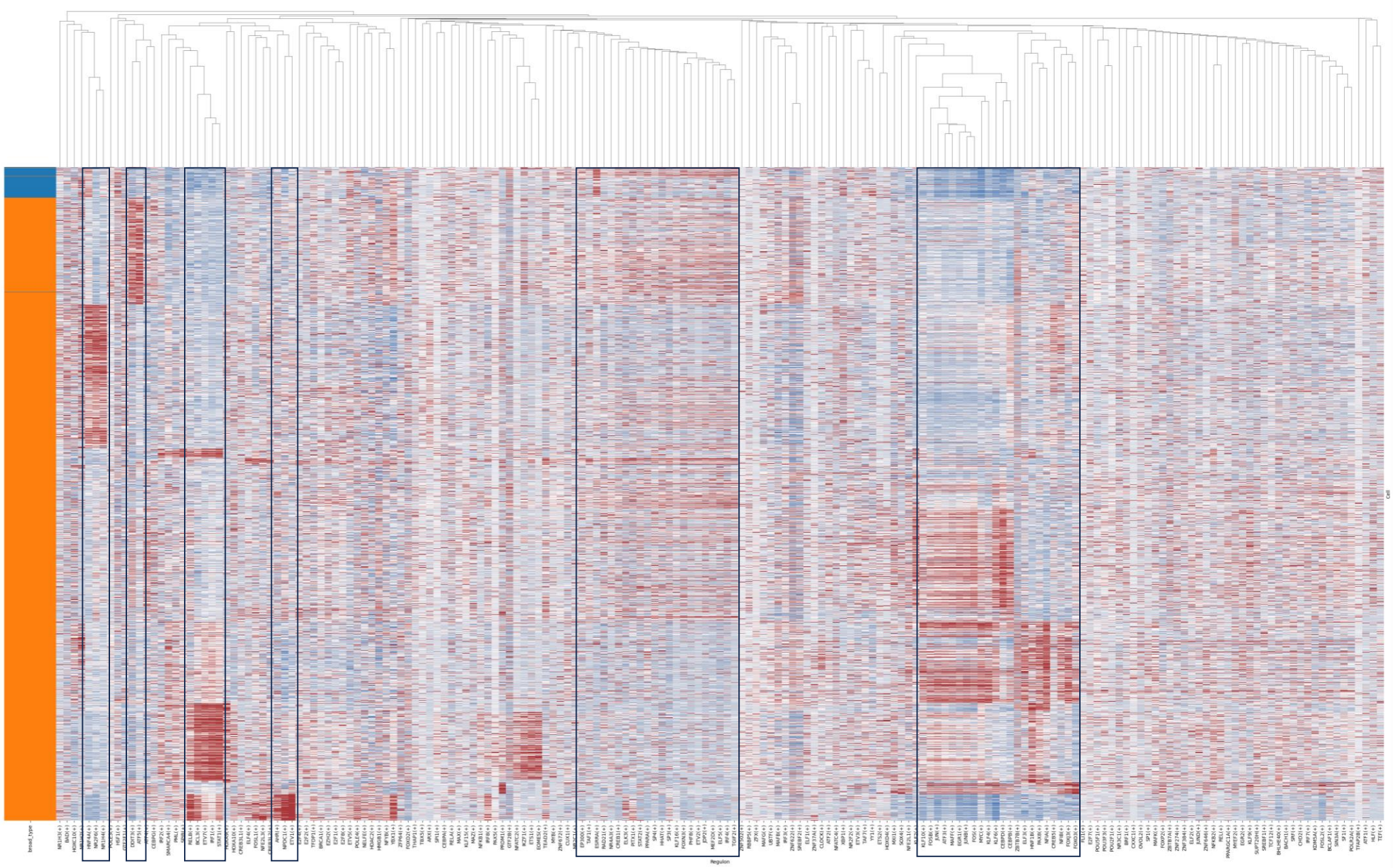
[FeatherRankingDatabase(name="hg38__refseq-r80__10kb_up_and_down_tss.mc9nr.genes_vs_motifs.rankings")]

Bram Van de Sande, *et al*., 2020, Nature Protocols

AUCell : 3 clusters



AUCell : 3 clusters



Differential TFs (Normal → EMT → Stress-adaptive)

ESRRA, RAD21, NR4A3, CREB1, ELK3, PITX1, STAT2, PPARA, SP4, HHAT, SP3, KLF16, FOXN3, PHF8, ETV2, JDP2, MEF2D, ELF5, IRF4, TGIF2

HNF4A, NR2F6, NR1H4, ATF5, ATF4,

KLF10, FOSB, JUN, ATF3, MAFF, EGR1, JUNB, FOS, MYC, KLF4, KLF6, CEBPD, CEBPB, ZBTB7B, ELF3, HNF1B, PAX8, NFIB, FOXJ3, FOXO3

RELB, BCL3, ETV7, IRF1, STAT1, AHR, NPDC1, ETV1

Preprocessing for CellOracle analysis

Raw UMI count data

↓

Delete ribosomal gene and exclude cell with percent.mt over 10

↓

Preprocessing for scVI analysis

Raw UMI count data

↓

Delete ribosomal gene and exclude cell with percent.mt over 10

↓

Preprocessing for scVI analysis