



QCIF
BIOINFORMATICS

Single Cell RNAseq Analysis in R with Seurat

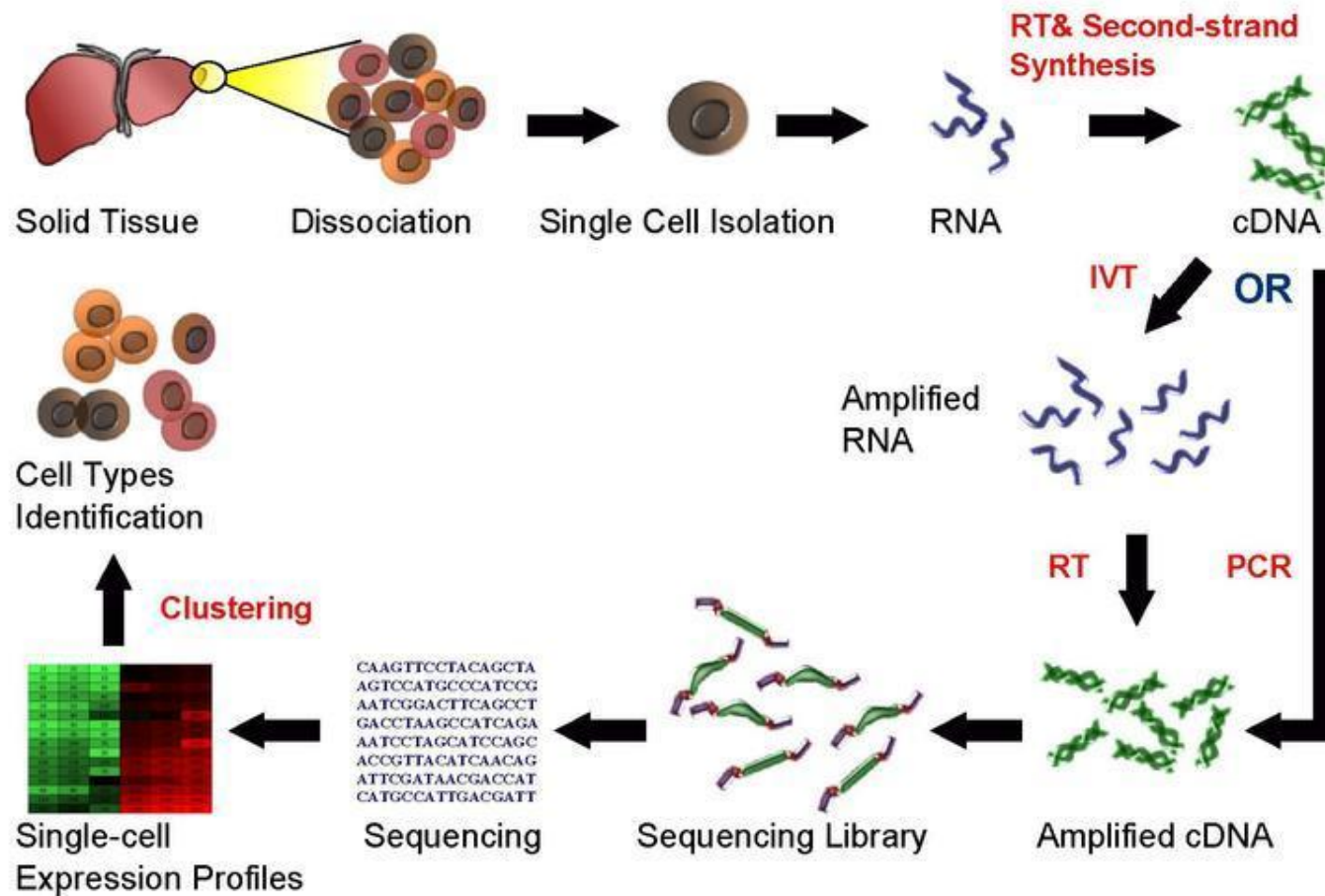
22-23 August 2022



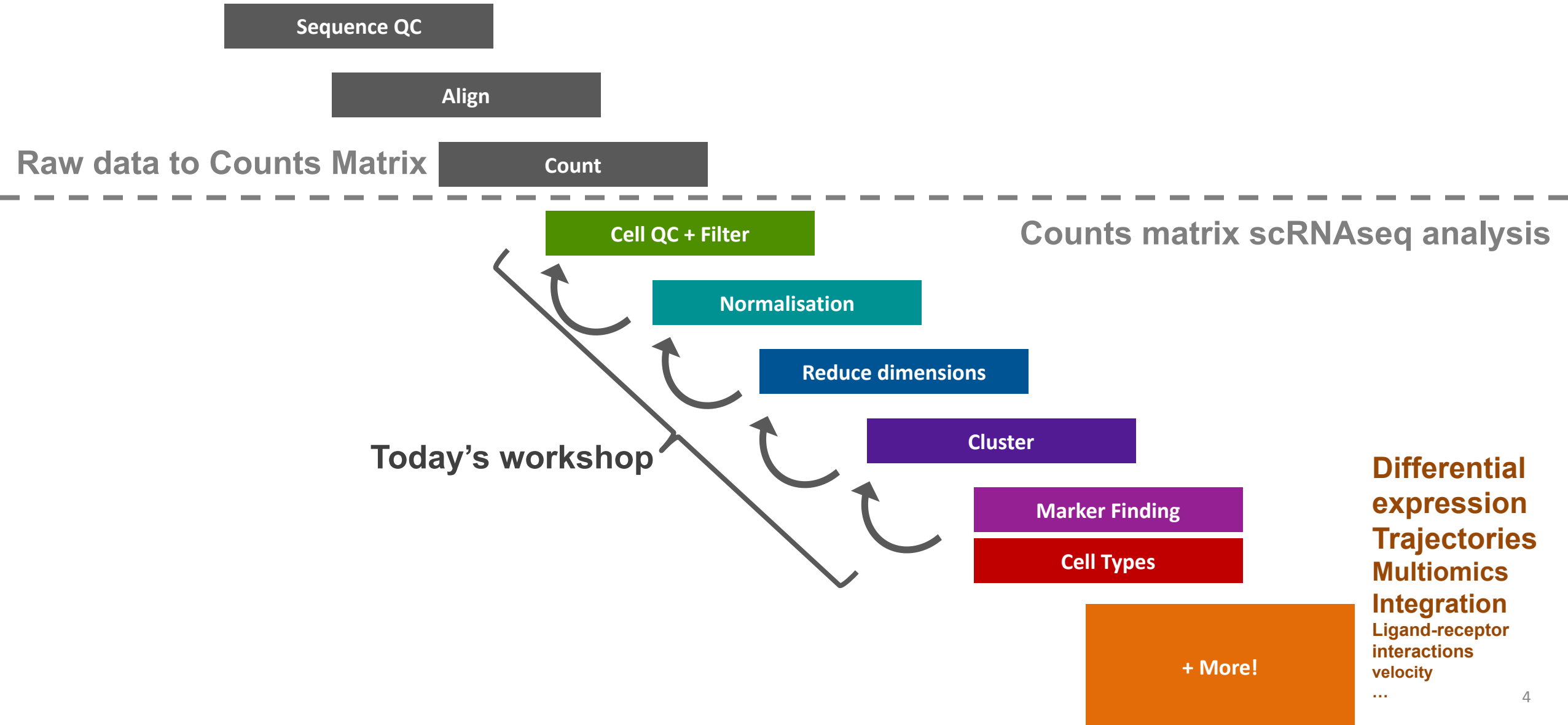
Single cell RNAseq Analysis

Single Cell RNaseq

Single Cell RNA Sequencing Workflow



Single Cell RNAseq Analysis Workflow



Raw data to Counts Matrix

Raw reads
(fastq files)

```
> read1  
CAGCGATTACAC  
> read2  
CACACCGGCAG  
...
```



(aligned to genome)



	Cell01	Cell02	Cell03	Cell04	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell11	Cell12	Cell13	Cell14	Cell15	Cell16	Cell17	Cell18	Cell19
CXCL1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3	2
CXCL8	0	0	2	0	0	0	0	0	0	2	0	0	0	0	0	0	1	3	1
DKK2	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0
EGF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOX11A	0	0	0	0	0	5	1	1	0	0	0	0	2	0	0	0	0	3	2
HBA1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0
IL5	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IL6	4	0	2	1	9	3	0	0	1	1	0	0	2	2	0	3	0	8	0
KIT	1	0	1	0	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1
LRRTM2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFKB1	1	1	0	0	1	0	1	1	0	0	1	1	1	0	1	0	0	1	2
PKD2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
RPS6KA2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SIX2	0	1	1	0	0	0	2	0	1	0	0	0	2	0	0	0	1	0	1
SNCA	0	0	1	0	0	2	2	1	1	1	0	0	0	0	0	0	3	0	0
TLR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TLR3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
ZNF621	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	4	2

Tools:

- starSOLO
- 10X CellRanger
- Alevin

Why start with counts matrix?

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114802>

A standard output of cellRanger (10X pipeline)

Often available for Public data!

Supplementary file	Size	Download	File type/resource
GSE114802_org4_barcode.tsv.gz	6.6 Kb	(ftp)(http)	TSV
GSE114802_org4_counts.csv.gz	5.7 Mb	(ftp)(http)	CSV
GSE114802_org4_genes.tsv.gz	258.3 Kb	(ftp)(http)	TSV
GSE114802_org4_matrix.mtx.gz	16.1 Mb	(ftp)(http)	MTX
GSE114802_org_barcode.tsv.gz	30.1 Kb	(ftp)(http)	TSV
GSE114802_org_counts.csv.gz	20.6 Mb	(ftp)(http)	CSV
GSE114802_org_genes.tsv.gz	258.6 Kb	(ftp)(http)	TSV
GSE114802_org_matrix.mtx.gz	60.6 Mb	(ftp)(http)	MTX

NCBI GEO Gene Expression Omnibus

COVID-19 is an emerging, rapidly evolving situation. Get the latest public health information from CDC: <https://www.coronavirus.gov>. Get the latest research from NIH: <https://www.nih.gov/coronavirus>. Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

NCBI > GEO > Accession Display

Scope: Self Format: HTML Amount: Quick GEO accession: GSE114802

Series GSE114802 Query DataSets for GSE114802

Status Public on Dec 07, 2018
Title Single cell RNA-Seq of four human kidney organoids
Organism [Homo sapiens](#)
Experiment type Expression profiling by high throughput sequencing
Summary These files represent single cell RNA-Seq data generated on a 10x Chromium genomics platform from four biological replicates of iPSC-derived human kidney organoids, in two batches, differentiated according to our published protocol (Takasato et al., Nature Protocols 2016). The aggregated human organoid data contains populations representing endothelial cells, podocytes, stroma, nephron, and off-target populations with similarity to neurons.

Overall design Examination of the cellular composition of human kidney organoids

Contributor(s) [Phipson B](#), [Zappia L](#), [Combes AN](#)
Citation(s) Phipson B, Er PX, Combes AN, Forbes TA et al. Evaluation of variability in human kidney organoids. *Nat Methods* 2019 Jan;16(1):79-87. PMID: 30573816
Combes AN, Zappia L, Er PX, Oshlack A et al. Single-cell analysis reveals congruence between kidney organoids and human fetal kidney. *Genome Med* 2019 Jan 23;11(1):3. PMID: 30674341

Submission date May 23, 2018
Last update date Mar 27, 2019
Contact name Luke Zappia
Organization name Murdoch Childrens Research Institute
Street address 50 Flemington Road
City Parkville
ZIP/Postal code 3052
Country Australia

Platforms (1) [GPL16791](#) Illumina HiSeq 2500 (Homo sapiens)

Samples (4) [GSM3150499](#) C32 Organoid 1
[GSM3150500](#) C32 Organoid 2
[GSM3150501](#) C32 Organoid 3

Relations
BioProject [PRJNA472699](#)
SRA [SRP148773](#)

Download family

Supplementary file	Size	Download	File type/resource
GSE114802_org4_barcode.tsv.gz	6.6 Kb	(ftp)(http)	TSV
GSE114802_org4_counts.csv.gz	5.7 Mb	(ftp)(http)	CSV
GSE114802_org4_genes.tsv.gz	258.3 Kb	(ftp)(http)	TSV
GSE114802_org4_matrix.mtx.gz	16.1 Mb	(ftp)(http)	MTX
GSE114802_org_barcode.tsv.gz	30.1 Kb	(ftp)(http)	TSV
GSE114802_org_counts.csv.gz	20.6 Mb	(ftp)(http)	CSV
GSE114802_org_genes.tsv.gz	258.6 Kb	(ftp)(http)	TSV
GSE114802_org_matrix.mtx.gz	60.6 Mb	(ftp)(http)	MTX

Raw data are available in SRA
Processed data are available on Series record

NLM | NIH | GEO Help | Disclaimer | Accessibility

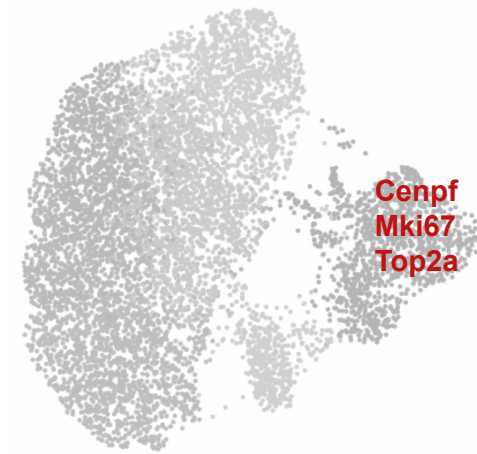
What we will cover

Counts Matrix

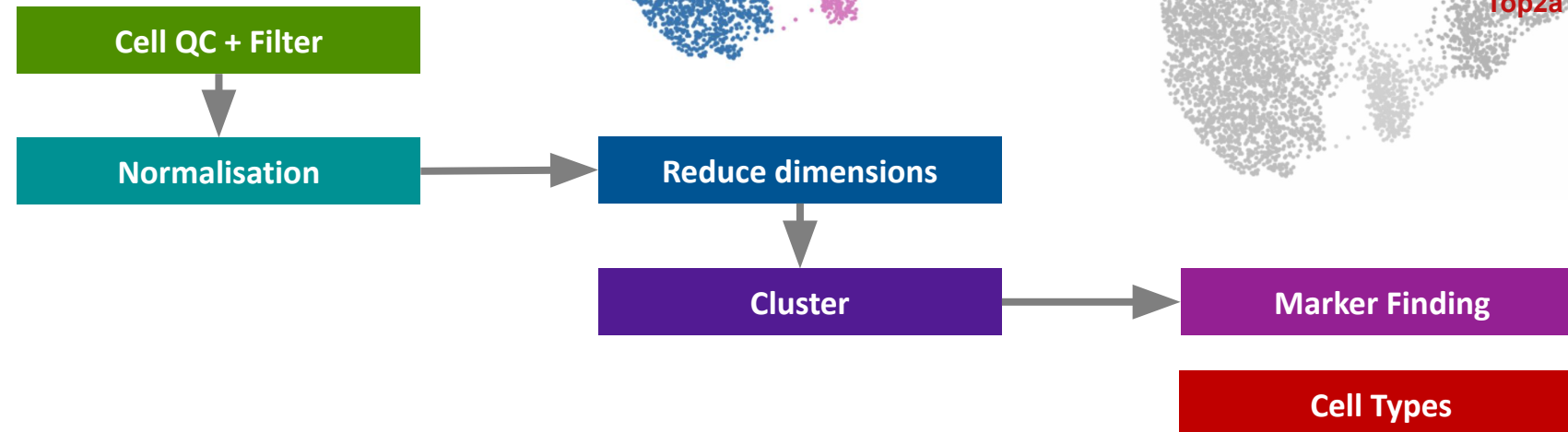
	Cell01	Cell02	Cell03	Cell04	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell11	Cell12	Cell13	Cell14	Cell15	Cell16	Cell17	Cell18	Cell19
CXCL1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3	2
CXCL8	0	0	2	0	0	0	0		0	2	0	0	0	0	0	0	1	3	1
DKK2	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0
EGF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOX11A	0	0	0	0	0	5	1	1	0	0	0	0	2	0	0	0	0	3	2
HBA1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0
IL5	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IL6	4	0	2	1	9	3	0	0	1	1	0	0	2	2	0	3	0	8	0
KIT	1	0	1	0	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1
LRRTM2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFKB1	1	1	0	0	1	0	1	1	0	0	1	1	1	0	1	0	0	1	2
PKD2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
RPS6KA2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SIX2	0	1	1	0	0	0	2	0	1	0	0	0	2	0	0	0	1	0	1
SNCA	0	0	1	0	0	2	2	1	1	1	0	0	0	0	0	0	3	0	0
TLR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TLR3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
ZNF621	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	4	2

Counts matrix to single-cell analysis

	Cell01	Cell02	Cell03	Cell04	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell11	Cell12	Cell13	Cell14	Cell15	Cell16	Cell17	Cell18	Cell19
CXCL1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3	2
CXCL8	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	1	3	1
DKK2	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	1	0
EGF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOX11A	0	0	0	0	0	5	1	1	0	0	0	0	2	0	0	0	0	3	2
HBA1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0
IL5	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IL6	4	0	2	1	9	3	0	0	1	1	0	0	2	2	0	3	0	8	0
KIT	1	0	1	0	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1
LRRTM2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFKB1	1	1	0	0	1	0	1	1	0	0	1	1	1	0	1	0	0	1	2
PKD2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
RPS6KA2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SIX2	0	1	1	0	0	0	2	0	1	0	0	0	2	0	0	0	1	0	1
SNCA	0	0	1	0	0	2	2	1	1	1	0	0	0	0	0	0	3	0	0
TLR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TLR3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
ZNF621	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	4	2



Cenpf
Mki67
Top2a



What we won't cover

Technology landscape

There are a lot more involved analyses that often

More than just RNA

- CITE-seq – protein assays
- VDJ – Immune profiling
- Spatial – 10X visium, CosMX e.t.c
- ATAC – chromatin accessibility
- Variant analyses

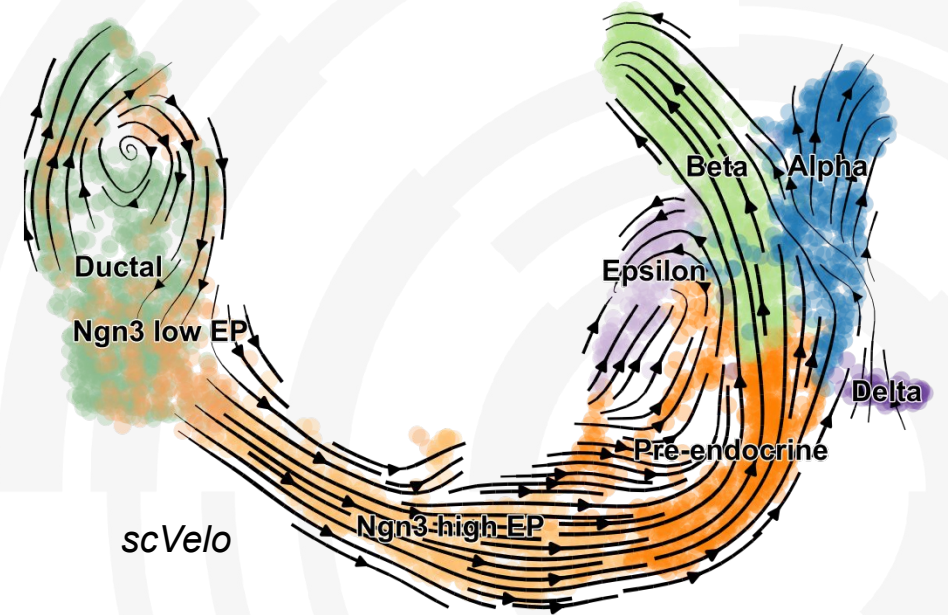
Differences between different technologies

- 10X vs BD rhapsody vs CosMX vs

Downstream analyses

Beyond the basics:

- Developmental trajectories and RNA velocity
- Batch correction
- Interactive visualisations tools
- Multimodal analyses
- Multiple dataset integration
- ...



We are happy to chat about these at the end though 😊

+ More!

Single cell Analysis Ecosystems

Single cell RNAseq Analysis Ecosystems



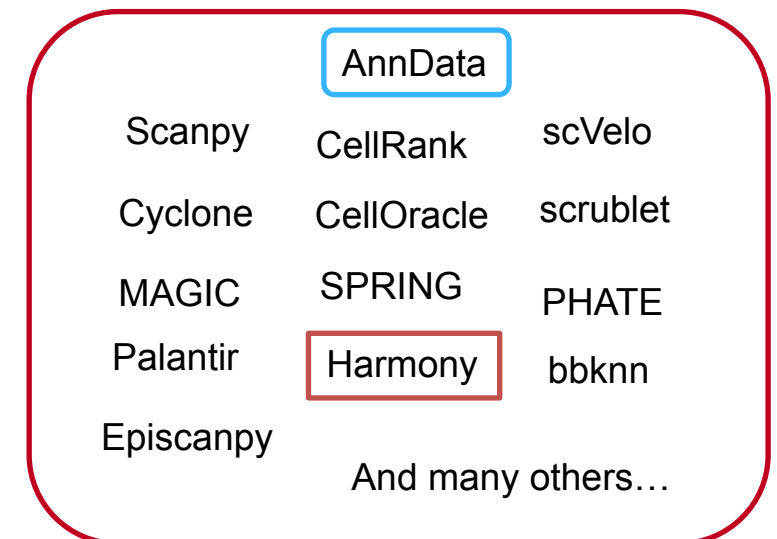
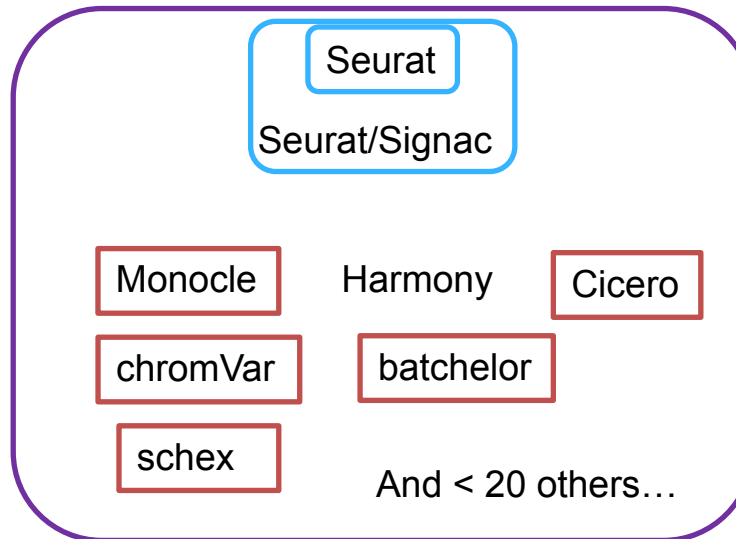
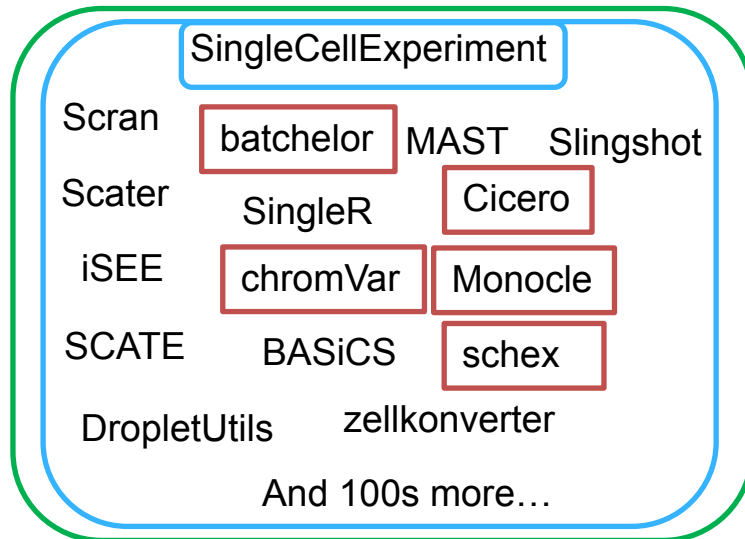
Bioconductor: R repository of many bioinformatics analysis packages. Single cell packages in Bioconductor make use of the *singleCellExperiment* class



•**Seurat:** R packages that has decided to make themselves a one-stop shop for most common single cell analysis tasks, uses the *Seurat* class



Scanpy: Python toolkit for single cell analysis. Uses the *anndata* class. Large ecosystem of tools that integrate with scanpy

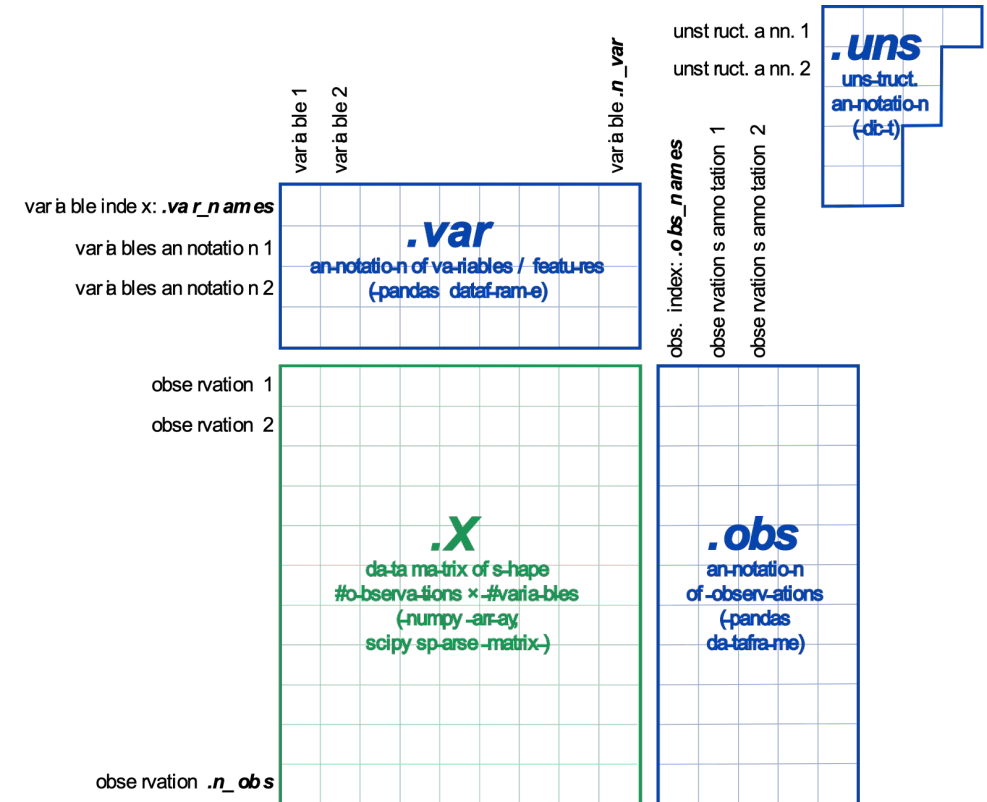
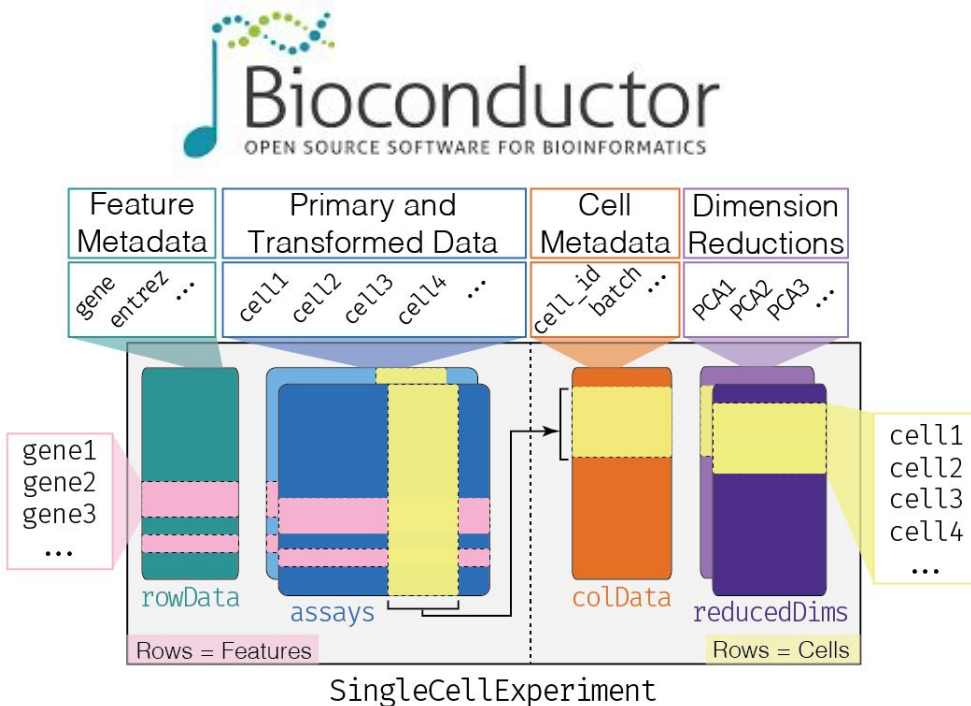


All are good options and well supported!

Single cell RNAseq Analysis Ecosystems

Each system has its **own central data structure** for storing and representing single cell data

Conversion: difficult, but not impossible!



Why Seurat?

- R package
- Widely used
- Fantastic tutorials - We're following one today!
- Excellent workflows for the routine tasks



<https://satijalab.org/seurat/>

	Introductory Vignettes
	PBMC 3K guided tutorial
	Using Seurat with multi-modal data
	Analysis, visualization, and integration of spatial datasets
	Analysis of Image-based Spatial Data in Seurat
	Data Integration
f	Introduction to scRNA-seq integration
	Mapping and annotating query datasets
	Fast integration using reciprocal PCA (RPCA)
jp	Tips for integrating large datasets
ab	Integrating scRNA-seq and scATAC-seq data
en	Multimodal reference mapping
ah	New Statistical Methods
as	Weighted Nearest Neighbor Analysis
ar	Mixscape Vignette
(P	Using sctransform in Seurat
Jlt	SCTransform, v2 regularization
g o	Cross-modality Bridge Integration
ef	Other
eq	Data visualization vignette
g t	Cell-cycle scoring and regression
jsi	Differential expression testing
qu	Demultiplexing with hashtag oligos (HTOs)
vi	Interoperability between single-cell object formats
ha	Parallelization in Seurat with future
im	Dimensional reduction vignette
de	Seurat essential commands list
Jr	Seurat interaction tips
Se	Merging Seurat objects
tit	

Compute requirements

- For today's workshop – an internet connection.
- For a real analysis
 - **RAM heavy** and scales with size e.g. 64Gb +
 - Modern datasets are much bigger than today's demo data (which could run on a laptop!)
 - There are **ways to improve RAM usage** e.g. SeuratDisk
 - Interactivity with your data is important, but sometimes you'll have to move it to a different server to run something overnight.
- If you're making your own counts matrix from Raw data it's a different task.
 - Suited to local HPC clusters
 - More data on disk, less ram needs.

Where to get compute? See what resources are available at your uni, what are your colleagues using?

Workshop Resources

Counts Matrix

	Cell01	Cell02	Cell03	Cell04	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell11	Cell12	Cell13	Cell14	Cell15	Cell16	Cell17	Cell18	Cell19
CXCL1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3	2
CXCL8	0	0	2	0	0	0	0		0	2	0	0	0	0	0	0	1	3	1
DKK2	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0
EGF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOX11A	0	0	0	0	0	5	1	1	0	0	0	0	2	0	0	0	0	3	2
HBA1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0
IL5	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IL6	4	0	2	1	9	3	0	0	1	1	0	0	2	2	0	3	0	8	0
KIT	1	0	1	0	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1
LRRTM2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFKB1	1	1	0	0	1	0	1	1	0	0	1	1	1	0	1	0	0	1	2
PKD2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
RPS6KA2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SIX2	0	1	1	0	0	0	2	0	1	0	0	0	2	0	0	0	1	0	1
SNCA	0	0	1	0	0	2	2	1	1	1	0	0	0	0	0	0	3	0	0
TLR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TLR3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
ZNF621	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	4	2

Loading a counts matrix

matrix.mtx

```
%%MatrixMarket matrix coordinate real
general
%
32738      2700      2286884
32709      1        4
32707      1        1
32706      1       10
32704      1        1
32703      1        5
32702      1        6
32700      1       10
32699      1       25
32698      1        3
32697      1        8
32527      1        1
...
```

genes.tsv

```
ENSG00000243485    MIR1302-10
ENSG00000237613    FAM138A
ENSG00000186092    OR4F5
ENSG00000238009    RP11-34P13.7
ENSG00000239945    RP11-34P13.8
ENSG00000237683    AL627309.1
ENSG00000239906    RP11-34P13.14
ENSG00000241599    RP11-34P13.9
ENSG00000228463    AP006222.2
ENSG00000237094    RP4-669L17.10
ENSG00000235249    OR4F29
ENSG00000236601    RP4-669L17.2
ENSG00000236743    RP5-857K21.15
ENSG00000231709    RP5-857K21.1
...
```

barcodes.tsv

```
AAACATACAACCAC-1
AAACATTGAGCTAC-1
AAACATTGATCAGC-1
AAACCGTGCTTCCG-1
AAACCGTGTATGCG-1
AAACGCACTGGTAC-1
AAACGCTGACCAGT-1
AAACGCTGGTTCTT-1
AAACGCTGTAGCCA-1
AAACGCTGTTTCTG-1
AAACTTGAAAAACG-1
AAACTTGATCCAGA-1
AAAGAGACGAGATA-1
AAAGAGACGCGAGA-1
...
```

30k genes * 10k cells = 300,000,000 measurements
= A lot of RAM!

Filter Cells

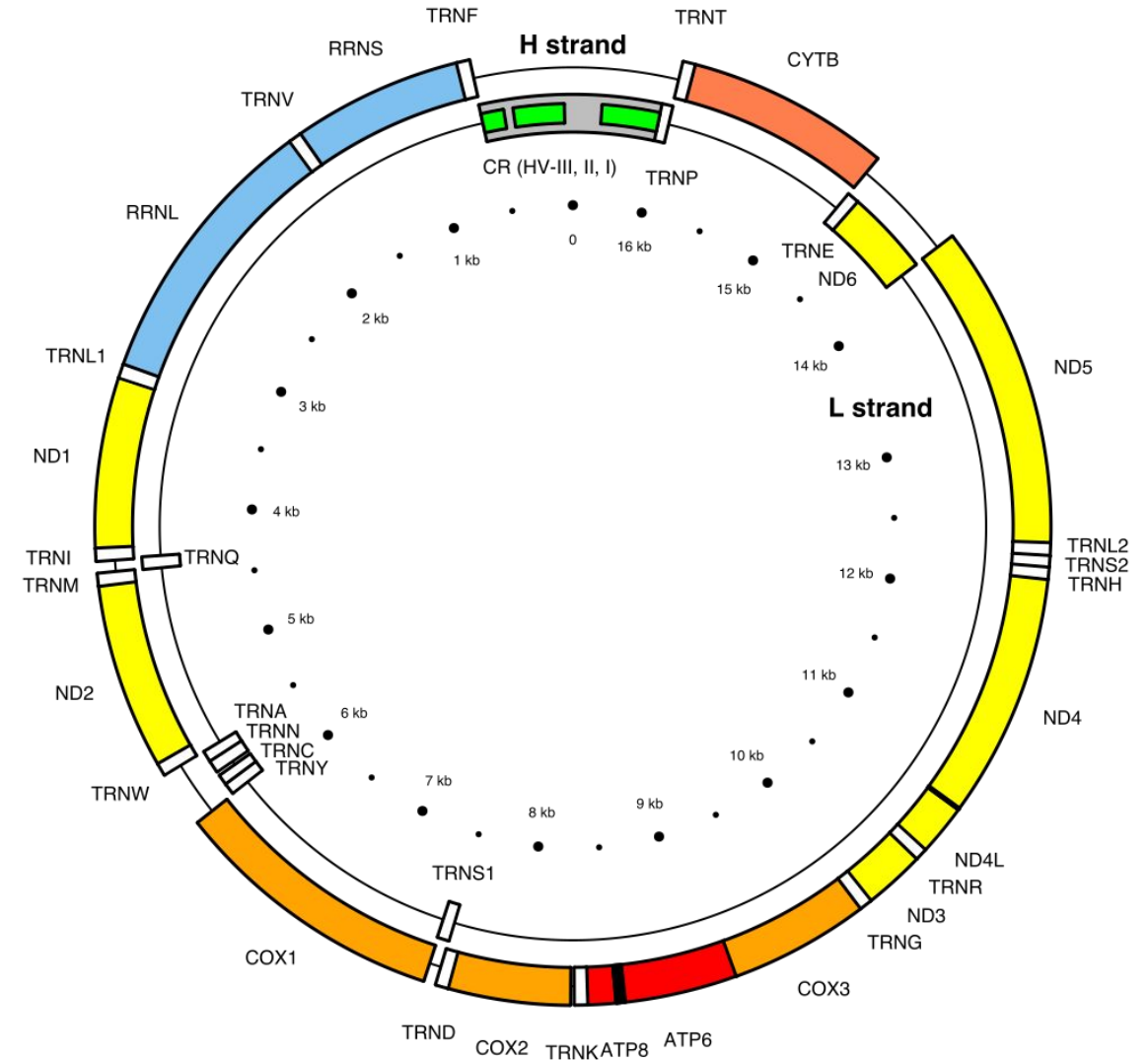
	Cell01	Cell02	Cell03	Cell04	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell11	Cell12	Cell13	Cell14	Cell15	Cell16	Cell17	Cell18	Cell19
CXCL1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3	2
CXCL8	0	0	2	0	0	0	0		0	2	0	0	0	0	0	0	1	3	1
DKK2	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0
EGF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOX11A	0	0	0	0	0	5	1	1	0	0	0	0	2	0	0	0	0	3	2
HBA1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0
IL5	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IL6	4	0	2	1	9	3	0	0	1	1	0	0	2	2	0	3	0	8	0
KIT	1	0	1	0	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1
LRRTM2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFKB1	1	1	0	0	1	0	1	1	0	0	1	1	1	0	1	0	0	1	2
PKD2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
RPS6KA2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SIX2	0	1	1	0	0	0	2	0	1	0	0	0	2	0	0	0	1	0	1
SNCA	0	0	1	0	0	2	2	1	1	1	0	0	0	0	0	0	3	0	0
TLR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TLR3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
ZNF621	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	4	2

Filter Genes

	Cell01	Cell03	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell13	Cell17	Cell18	Cell19
CXCL1	0	1	0	0	0	0	1	0	0	0	3	2
CXCL8	0	2	0	0	0	0	0	2	0	1	3	1
DKK2	0	0	0	0	0	1	1	1	1	0	1	0
EGF	0	0	0	0	0	0	0	0	0	0	0	0
HOX11A	0	0	0	5	1	1	0	0	2	0	3	2
HBA1	0	0	0	0	0	0	0	0	0	0	0	0
IL5	0	1	0	0	0	0	0	0	0	0	0	0
IL6	4	2	9	3	0	0	1	1	2	0	8	0
KIT	1	1	0	1	0	1	0	1	0	0	1	1
LRRTM2	0	0	0	0	0	0	0	0	0	0	0	0
NFKB1	1	0	1	0	1	1	0	0	1	0	1	2
PKD2	0	0	0	0	0	0	0	0	0	0	1	0
RPS6KA2	0	0	0	0	0	0	0	0	0	0	0	0
SIX2	0	1	0	0	2	0	1	0	2	1	0	1
SNCA	0	1	0	2	2	1	1	1	0	3	0	0
TLR2	0	0	0	0	0	0	0	0	0	0	0	0
TLR3	0	0	0	0	0	0	0	0	0	0	1	0
ZNF621	0	0	0	0	0	0	1	1	1	1	4	2

Mitochondrial Content

- High mitochondrial gene content can indicate 'dying' cells
 - Uninteresting, confuses analysis
- Measure the % of counts from mitochondrial genes, and remove cells with > 5%



Filtered Counts matrix

	Cell01	Cell03	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell13	Cell17	Cell18	Cell19
CXCL1	0	1	0	0	0	0	1	0	0	0	3	2
CXCL8	0	2	0	0	0	0	0	2	0	1	3	1
DKK2	0	0	0	0	0	1	1	1	1	0	1	0
HOX11A	0	0	0	5	1	1	0	0	2	0	3	2
IL6	4	2	9	3	0	0	1	1	2	0	8	0
KIT	1	1	0	1	0	1	0	1	0	0	1	1
NFKB1	1	0	1	0	1	1	0	0	1	0	1	2
SIX2	0	1	0	0	2	0	1	0	2	1	0	1
SNCA	0	1	0	2	2	1	1	1	0	3	0	0
ZNF621	0	0	0	0	0	0	1	1	1	1	4	2

Normalisation

	Cell01	Cell03	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell13	Cell17	Cell18	Cell19
CXCL1	1.7	1.3	0.0	0.9	1.7	0.0	1.7	0.0	0.0	8.3	0.0	0.0
CXCL8	1.7	0.0	1.0	0.0	0.0	2.0	1.7	1.4	0.0	10.0	0.0	0.0
DKK2	1.7	0.0	1.0	0.9	0.0	0.0	1.7	1.4	0.0	11.7	0.0	0.0
HOX11A	1.7	2.5	2.0	0.0	1.7	4.0	0.0	1.4	0.0	15.0	0.0	0.0
IL6	0.0	0.0	0.0	0.0	0.0	2.0	5.0	1.4	0.0	10.0	0.0	0.0
KIT	1.7	3.8	8.0	0.9	1.7	0.0	0.0	5.7	0.0	40.0	0.0	0.0
NFKB1	0.0	2.5	0.0	0.9	3.3	2.0	0.0	2.9	0.0	18.3	0.0	0.0
SIX2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SNCA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ZNF621	8.3	17.5	30.0	6.4	13.3	16.0	18.3	14.3	0.0	0.0	0.0	0.0

Highly variable Genes

	Cell01	Cell03	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell13	Cell17	Cell18	Cell19
CXCL1	1.7	1.3	0.0	0.9	1.7	0.0	1.7	0.0	0.0	8.3	0.0	0.0
CXCL8	1.7	0.0	1.0	0.0	0.0	2.0	1.7	1.4	0.0	10.0	0.0	0.0
DKK2	1.7	0.0	1.0	0.9	0.0	0.0	1.7	1.4	0.0	11.7	0.0	0.0
HOX11A	1.7	2.5	2.0	0.0	1.7	4.0	0.0	1.4	0.0	15.0	0.0	0.0
IL6	0.0	0.0	0.0	0.0	0.0	2.0	5.0	1.4	0.0	10.0	0.0	0.0
KIT	1.7	3.8	8.0	0.9	1.7	0.0	0.0	5.7	0.0	40.0	0.0	0.0
NFKB1	0.0	2.5	0.0	0.9	3.3	2.0	0.0	2.9	0.0	18.3	0.0	0.0
SIX2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SNCA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ZNF621	8.3	17.5	30.0	6.4	13.3	16.0	18.3	14.3	0.0	0.0	0.0	0.0

PCA

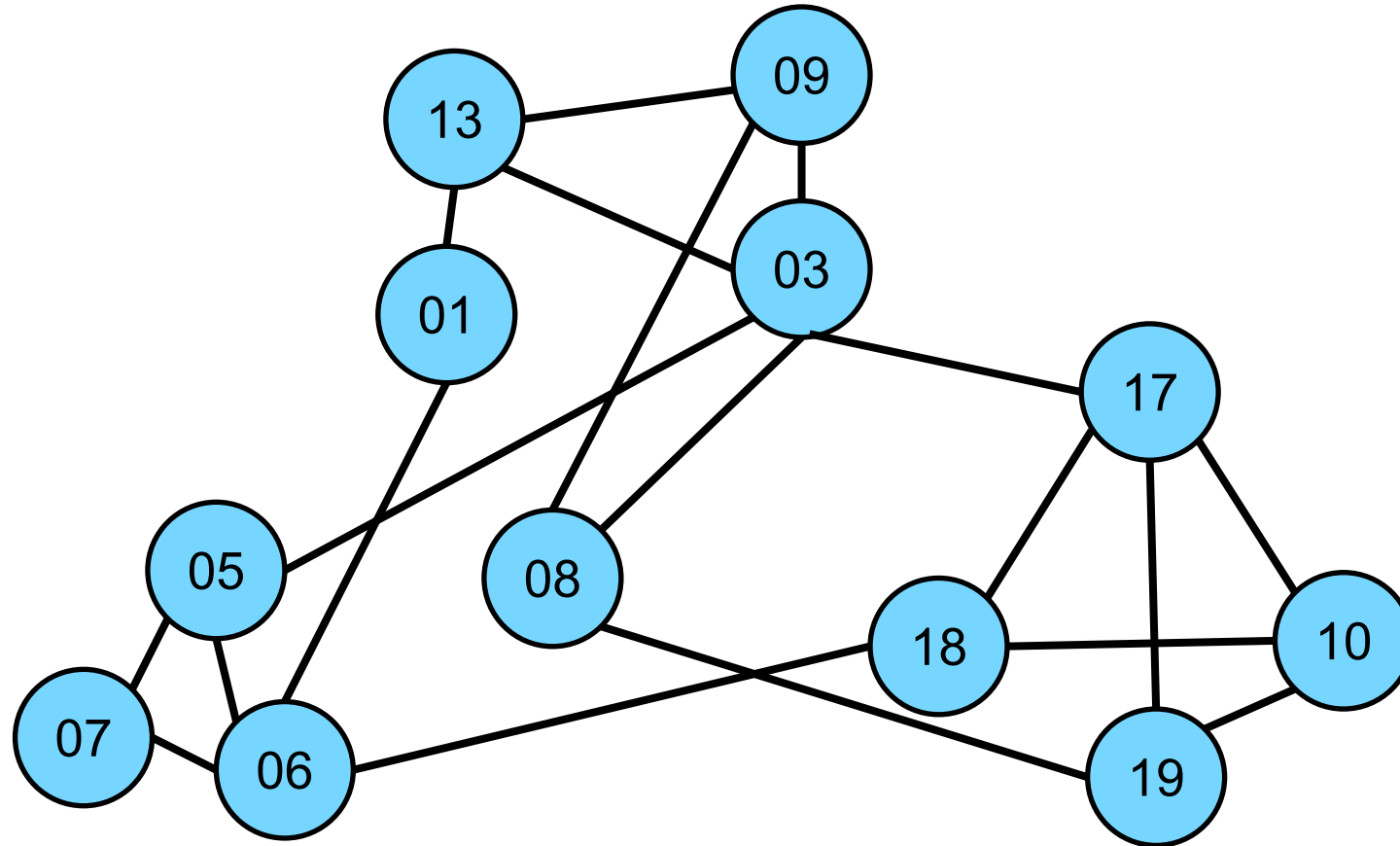
	Cell01	Cell03	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell13	Cell17	Cell18	Cell19
CXCL1	1.7	1.3	0.0	0.9	1.7	0.0	1.7	0.0	0.0	8.3	0.0	0.0
CXCL8	1.7	0.0	1.0	0.0	0.0	2.0	1.7	1.4	0.0	10.0	0.0	0.0
DKK2	1.7	0.0	1.0	0.9	0.0	0.0	1.7	1.4	0.0	11.7	0.0	0.0
HOX11A	1.7	2.5	2.0	0.0	1.7	4.0	0.0	1.4	0.0	15.0	0.0	0.0
IL6	0.0	0.0	0.0	0.0	0.0	2.0	5.0	1.4	0.0	10.0	0.0	0.0
KIT	1.7	3.8	8.0	0.9	1.7	0.0	0.0	5.7	0.0	40.0	0.0	0.0
NFKB1	0.0	2.5	0.0	0.9	3.3	2.0	0.0	2.9	0.0	18.3	0.0	0.0
SIX2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SNCA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ZNF621	8.3	17.5	30.0	6.4	13.3	16.0	18.3	14.3	0.0	0.0	0.0	0.0

	Cell01	Cell03	Cell05	Cell06	Cell07	Cell08	Cell09	Cell10	Cell13	Cell17	Cell18	Cell19
PC1	-1.2	6.5	2.6	0.1	0.3	1.7	4.2	-1.2	-1.2	-1.2	-1.2	-1.2
PC2	0.6	0.1	2.0	0.0	-0.3	0.1	0.4	1.2	-0.5	-0.5	-0.5	-0.5
PC3	-1.1	0.4	-1.1	-1.1	0.4	1.8	1.8	-1.1	-1.1	-1.1	-1.1	-1.1
PC4	-0.8	-0.1	0.1	-1.7	-1.1	1.2	2.4	0.6	-2.0	-2.0	-2.0	-2.0

...

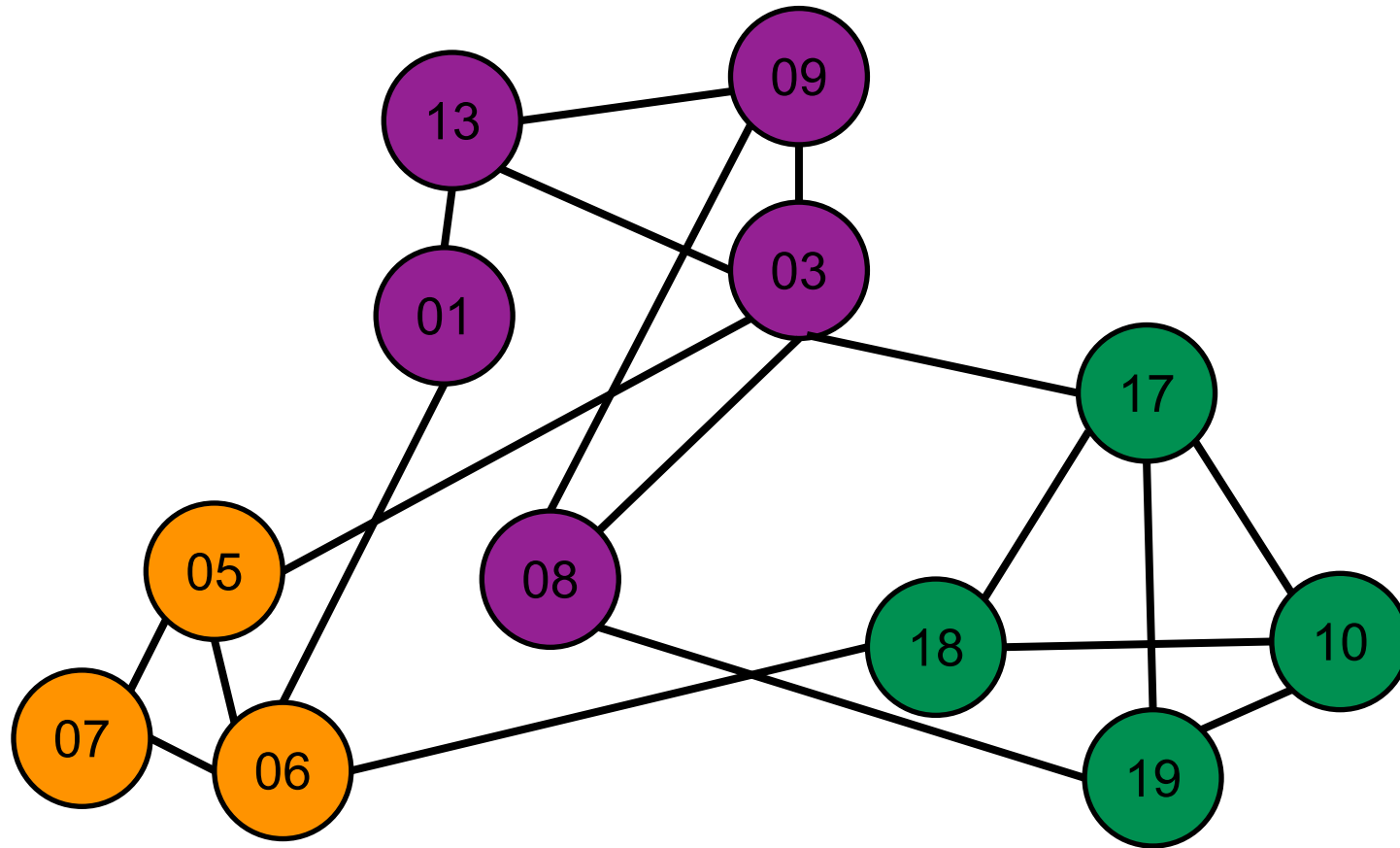
PC50

Graph

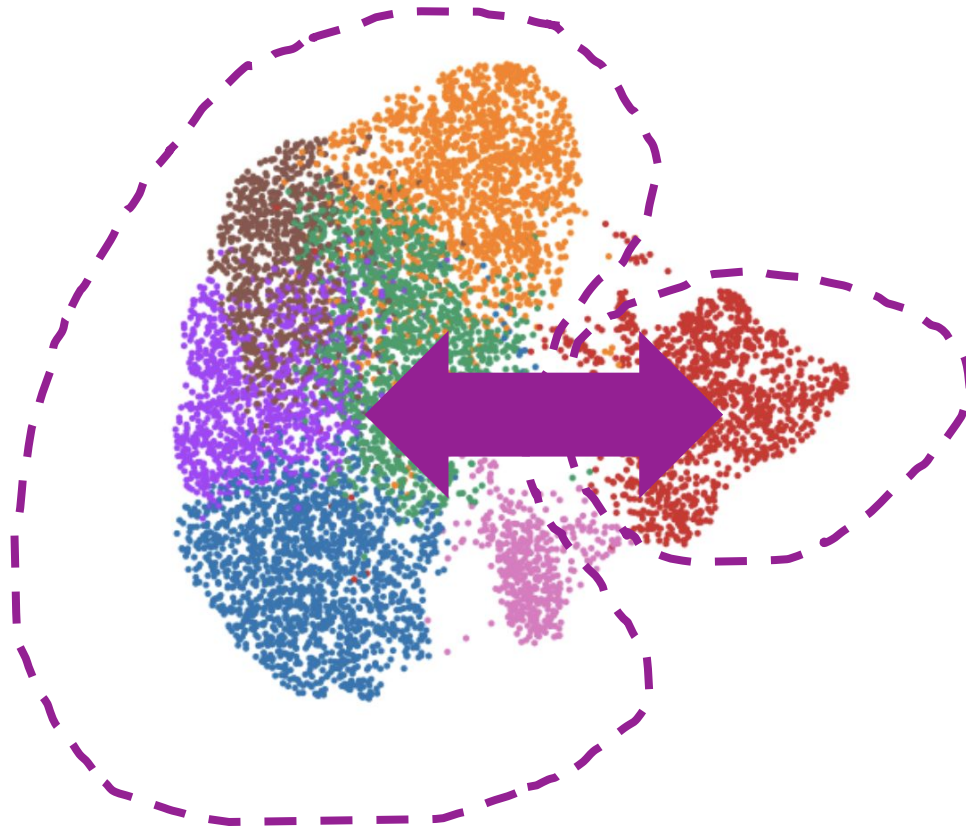


Not the UMAP; *Input* for the UMAP and clustering

Clustering



Differential Expression: Cluster Markers



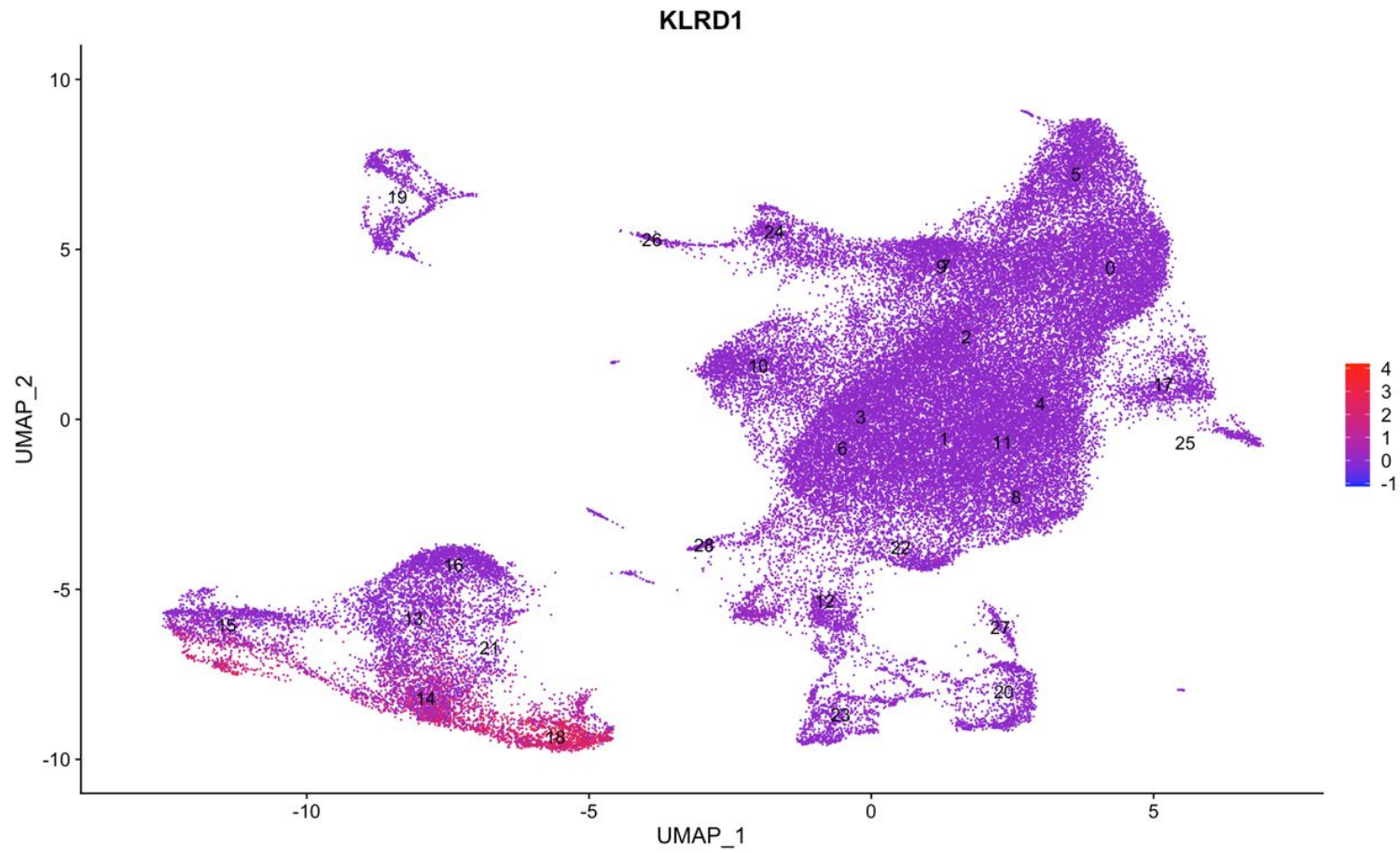
Cell Type Analysis

- Substantial effort to use scRNAseq to identify the type of cells
- Cell can be organized into groups based on similar features;
 - Expression (gene/protein)
 - location
 - Interaction
 - lineage
- Flow cytometry
 - Antibodies
 - Clusters based on expression
- scRNAseq
 - genes
 - Clusters based on expression

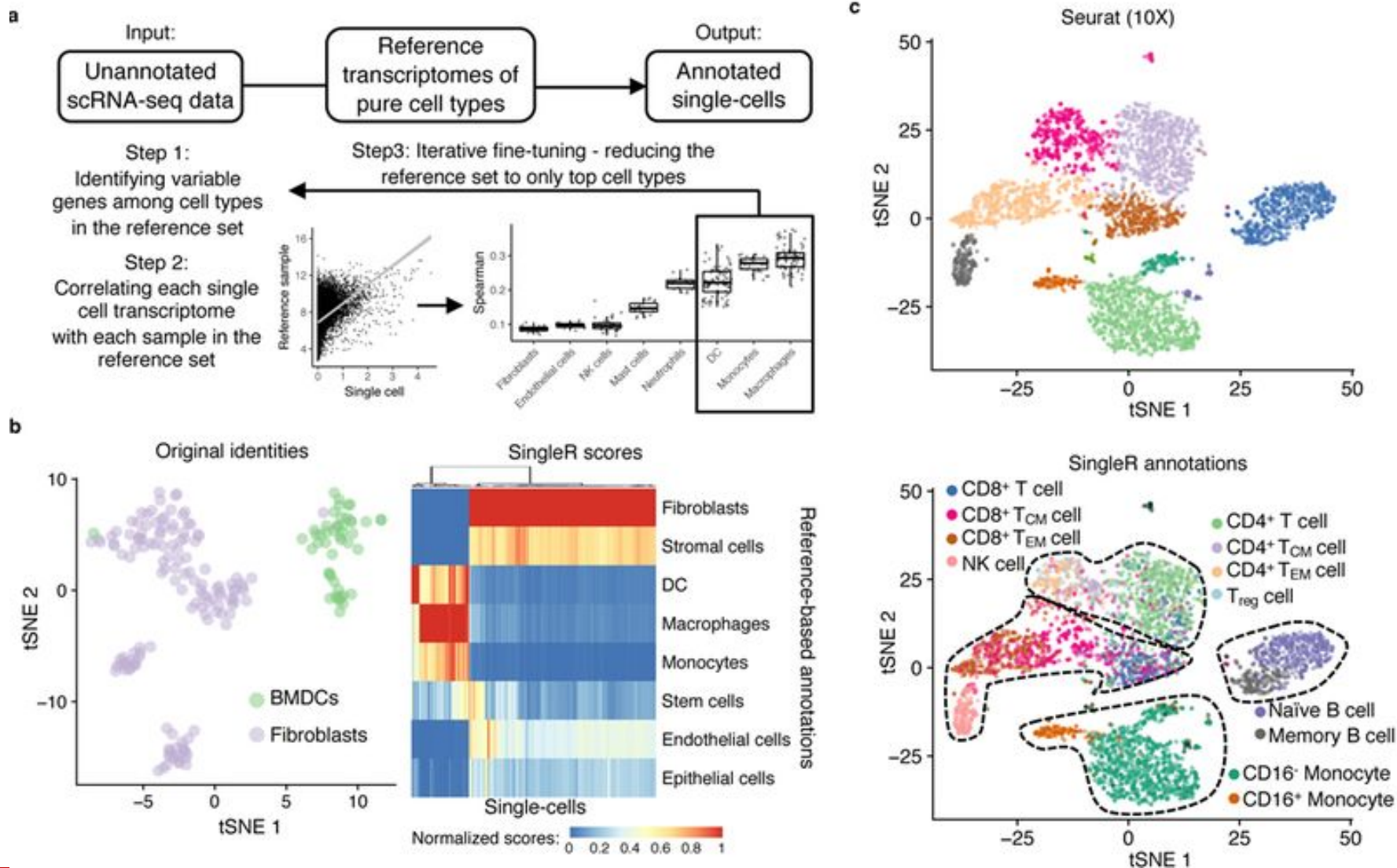
Cell Type Classification

- Unsupervised
 - Clustering methods
 - Known cell marker genes
 - Manual annotation
 - Examples: Graphical models (Seurat clustering), RACEID
- Supervised
 - Reference dataset
 - Model (training/testing)
 - Examples: Correlation based methods, singleR, scPred, CHETAH

Manual annotation

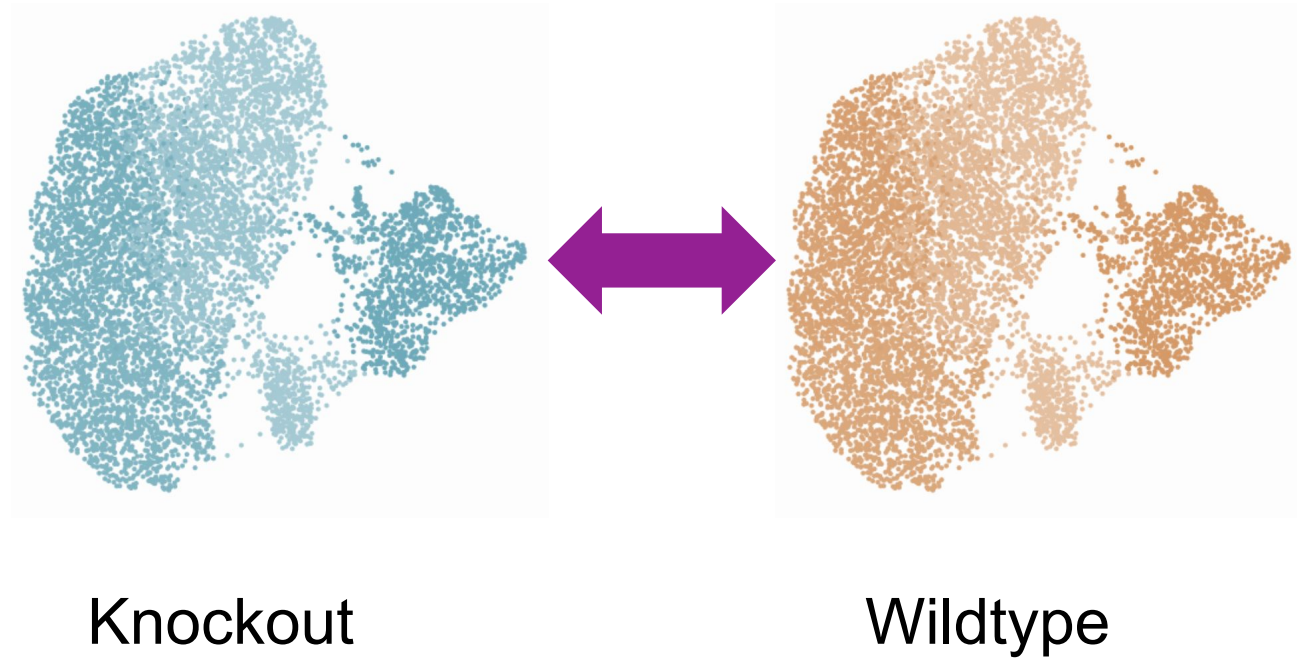


Supervised example - SingleR



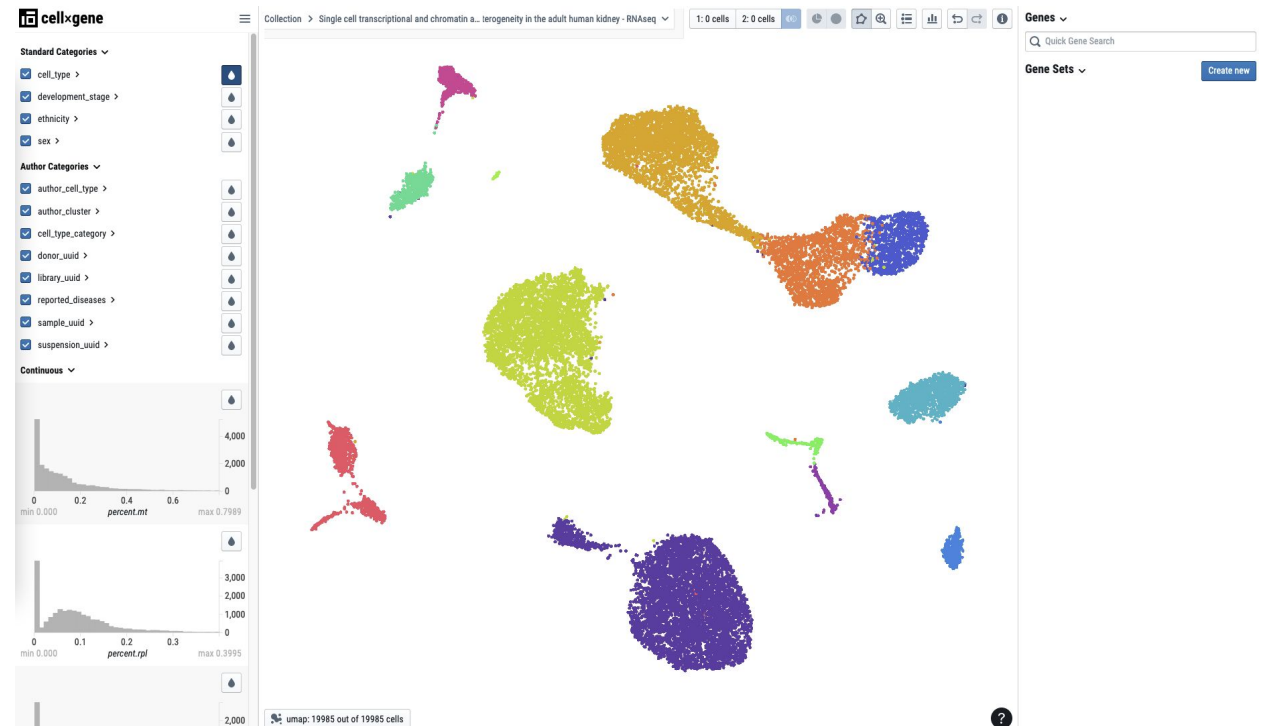
The home for lost slides

Differential Expression: KO vs WT



Single Cell Browsers

- CellXGene
 - In galaxy soon
 - <https://cellxgene.cziscience.com/e/9df60c57-fdf3-4e93-828e-fe9303f20438.cxg/>
- iSEE
 - Requires different file format
- CellLoupe
 - Desktop application
 - Proprietary 10X software for 10X data only



Finishing up

Where to from here?

The data is now ready for biological analysis.

- Keep plotting
- Load annotated data into a cell browser
- Share with a colleague

Trajectories
Differential expression
Multomics
Integration
Ligand-receptor interactions
velocity
...

Rdsfasdf