

High Resolution Technologies

Dr Stevie Pederson (They/Them) stevie.pederson@thekids.org.au

Black Ochre Data Labs
The Kids Research Institute Australia

Acknowledgement Of Country

I'd like to acknowledge the Kaurna people as the traditional owners and custodians of the land we know today as the Adelaide Plains, where I live & work.

I also acknowledge the deep feelings of attachment and relationship of the Kaurna people to their place.

I pay my respects to the cultural authority of Aboriginal and Torres Strait Islander peoples from other areas of Australia, and pay my respects to Elders past, present and emerging, and acknowledge any Aboriginal Australians who may be with us today

High Resolution Technologies

- Microarray and early RNA-Seq analysis used 'bulk' tissues
- Very hard to obtain pure cell-types in most samples
 - T cells sorted by FACS still heterogeneous
 - Cancer biopsises are heterogeneous
 - Immortalised cell-lines (HeLa, HEK293 etc.) differentiate into multiple cell types
- Single-Cell RNA-Seq (scRNA-Seq) \Rightarrow Identify cell-types within a sample

High Resolution Technologies

- Microarray and early RNA-Seq analysis used 'bulk' tissues
- Very hard to obtain pure cell-types in most samples
 - T cells sorted by FACS still heterogeneous
 - Cancer biopsises are heterogeneous
 - Immortalised cell-lines (HeLa, HEK293 etc.) differentiate into multiple cell types
- Single-Cell RNA-Seq (scRNA-Seq) \Rightarrow Identify cell-types within a sample
- Spatial Transcriptomics \Rightarrow Identify cell-types on a slide by location

High Resolution Technologies

- Microarray and early RNA-Seq analysis used 'bulk' tissues
- Very hard to obtain pure cell-types in most samples
 - T cells sorted by FACS still heterogeneous
 - Cancer biopsises are heterogeneous
 - Immortalised cell-lines (HeLa, HEK293 etc.) differentiate into multiple cell types
- Single-Cell RNA-Seq (scRNA-Seq) \Rightarrow Identify cell-types within a sample
- Spatial Transcriptomics \Rightarrow Identify cell-types on a slide by location
 - Doesn't always utilise sequencing

High Resolution Technologies

- Quite a different set of questions to bulk tissue
- Is differential expression large in a subset of cells?

High Resolution Technologies

- Quite a different set of questions to bulk tissue
- Is differential expression large in a subset of cells?
 - Or small in all cells?

High Resolution Technologies

- Quite a different set of questions to bulk tissue
- Is differential expression large in a subset of cells?
 - Or small in all cells?
- How do cells travel through differentiation trajectories?

High Resolution Technologies

- Quite a different set of questions to bulk tissue
- Is differential expression large in a subset of cells?
 - Or small in all cells?
- How do cells travel through differentiation trajectories?
- Which cells are communicating with which other cells?

High Resolution Technologies

- Quite a different set of questions to bulk tissue
- Is differential expression large in a subset of cells?
 - Or small in all cells?
- How do cells travel through differentiation trajectories?
- Which cells are communicating with which other cells?
- Which cells are co-localised within tissue samples?

High Resolution Technologies

- Quite a different set of questions to bulk tissue
- Is differential expression large in a subset of cells?
 - Or small in all cells?
- How do cells travel through differentiation trajectories?
- Which cells are communicating with which other cells?
- Which cells are co-localised within tissue samples?
 - And where are the transcripts within a cell?

Single Cell RNA-Seq

Single Cell RNA-Seq

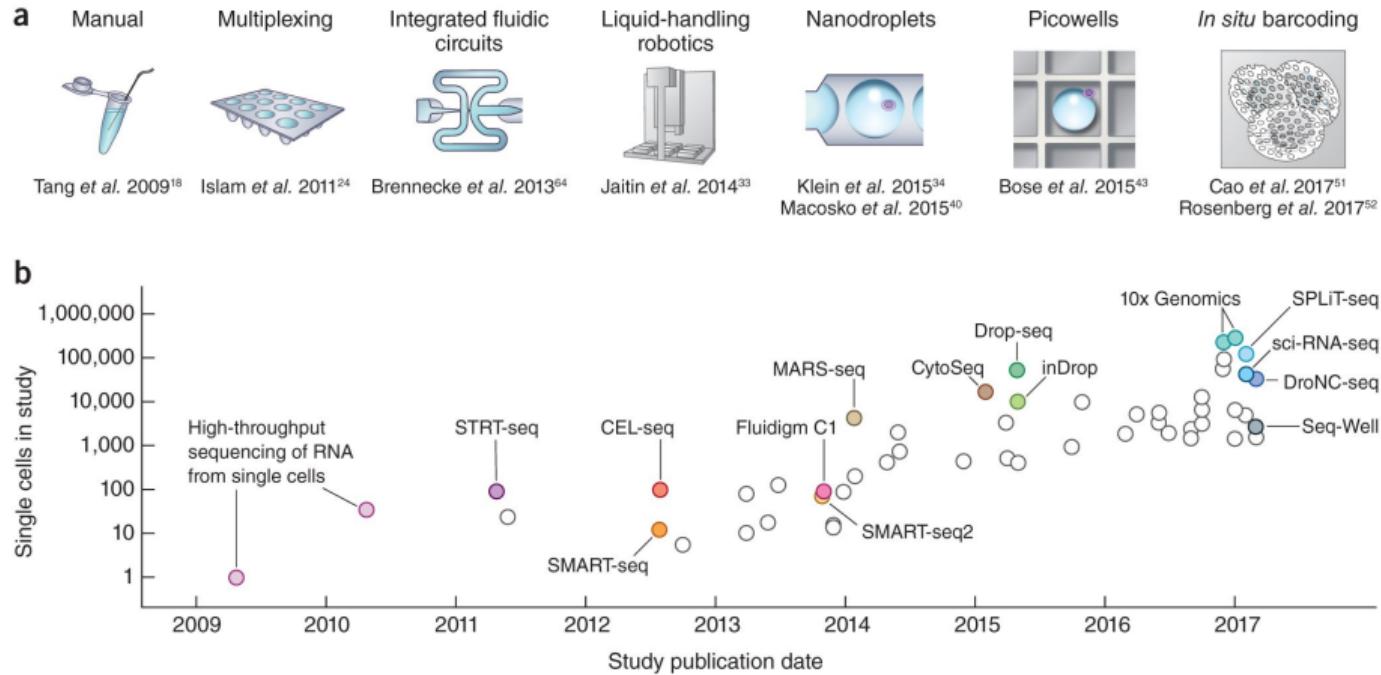
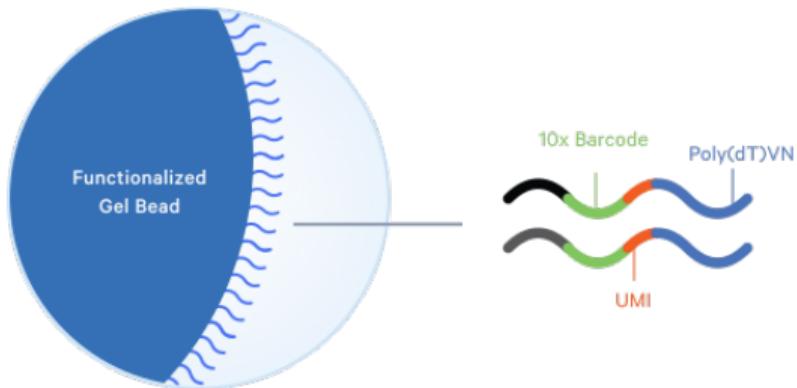


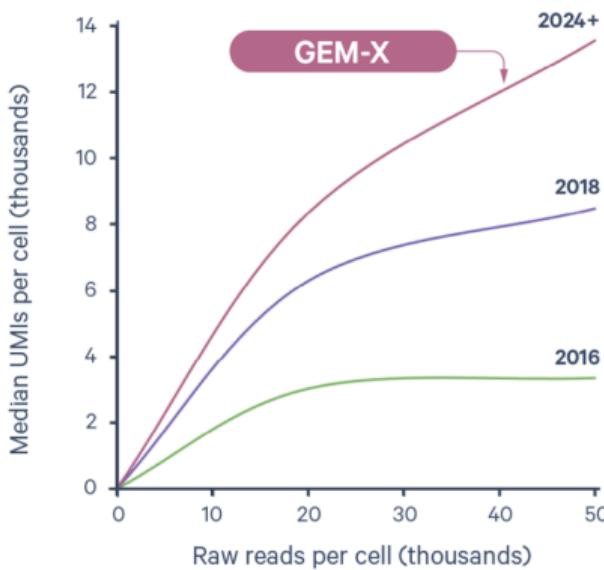
Figure 1: Figure from Svensson, Vento-Tormo, and Teichmann (2018)

Droplet-Based scRNA



- GEM: Gel bead in EMulsion
- Primers can capture 3', 5' or targeted sequences
- Figures from www.10xgenomics.com

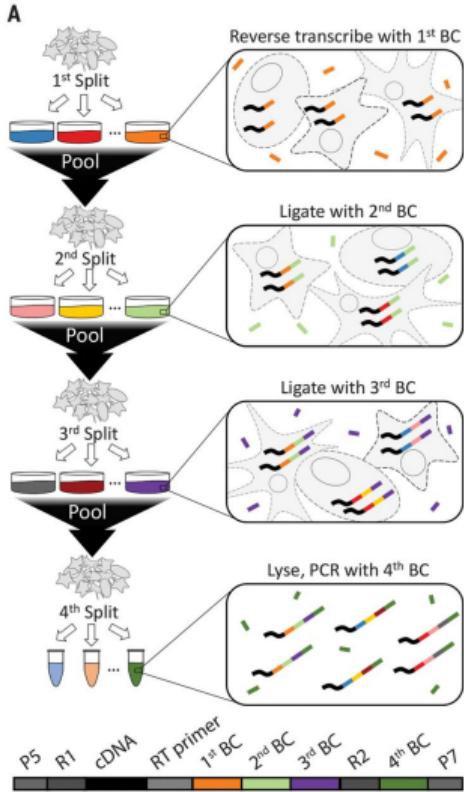
Droplet-Based scRNA



- UMIs represent individual RNA molecules
- Now capturing ~14,000 transcripts/cell

Figure 2: www.10xgenomics.com

SPLIT-Seq



SPLIT-Seq is an alternative, plate based scRNA method

- Cells are split into pools and fixed
- One barcode/pool
- Multiple rounds of pooling and barcoding
- All amplification is *in situ*

Figure 3: Taken from Rosenberg et al. (2018)

Comparison of Select Technologies

Table 1: Data sourced from Haque et al. (2017)

Protocol	C1 (SMART-Seq)	SMART- Seq2	10X Chromium	SPLIT-Seq
<i>Platform</i>	Microfluidics	Plate-based	Droplet	Plate-based
<i>Transcript</i>	Full-length	Full-length	3'-end	3'-end
<i>Cells</i>	$10^2 - 10^3$	$10^2 - 10^3$	$10^3 - 10^4$	$10^3 - 10^5$
<i>Reads/Cell</i>	10^6	10^6	$10^4 - 10^5$	10^4

Data Processing

- Most pre-processing for 10X data is performed using CellRanger
- Handles demultiplexing, alignment (STAR) and quantification (using UMIs)
 - Full-length transcript methods can utilise kallisto/salmon
- We end up with a feature-barcode matrix
 - A barcode represents an individual cell (or a set of reactions)
 - A feature is commonly thought of as a gene in scRNA-Seq
- Similar to counts from bulk RNA-Seq but with many more columns (cells)
 - Also many more zero counts for genes

Filtering Stages

- The aim is to keep the *high quality cells* and discard the dubious cells, such as:
 - ① Low/High read numbers (library sizes)
 - ② Low feature/gene numbers
 - ③ High proportions of mitochondrial RNA = cells broken prior to lysis
 - ④ Doublets (i.e. two cells instead of a single cell)
- Also filter for genes considered as detectable (Average Counts > 1)
 - Some zero counts are from unexpressed genes
 - Others are “dropouts” → RNA capture efficiency between 10-40%

Normalisation

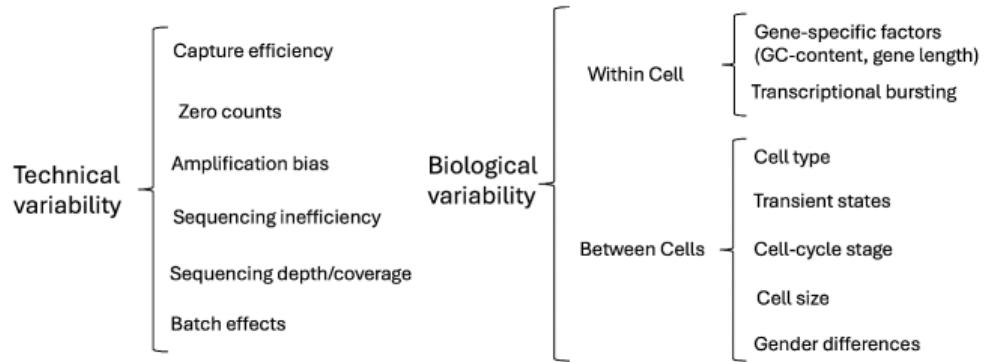


Figure 4: Image from Cuevas-Diaz Duran, Wei, and Wu (2024)

- Bulk-RNA normalisation methods no longer applicable
 - Does each cell (or cell-type) have the same starting amount of RNA?
- Normalisation strategies may be platform dependent
 - Gene length impacts full-length methods but not droplet-based (3' or 5')
- Should missing values be imputed?

Identifying Cell Types

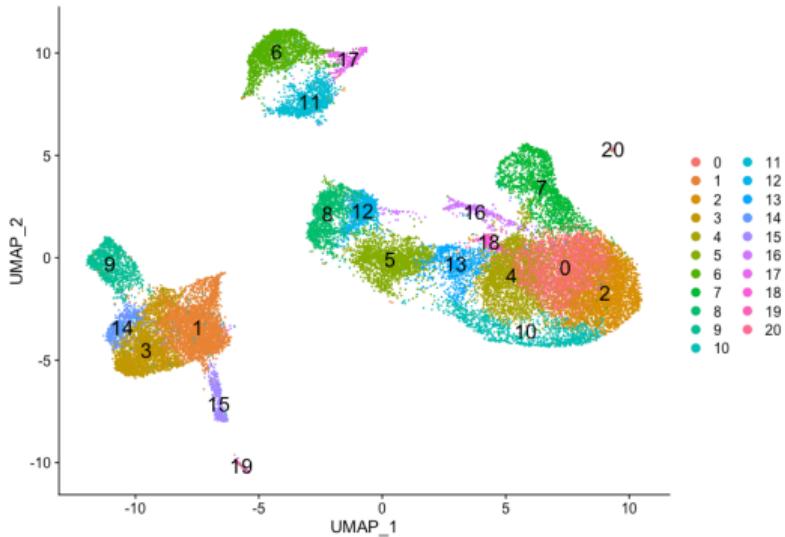


Figure 5: Image courtesy of the Harvard Chan Bioinformatics Core

- A key aim is to identify different cell types in our sample
 - How many cell types are there really?
 - How similar are cell types?
 - What genes drive the differences → Differential Expression
- ① Dimensional Reduction

Identifying Cell Types

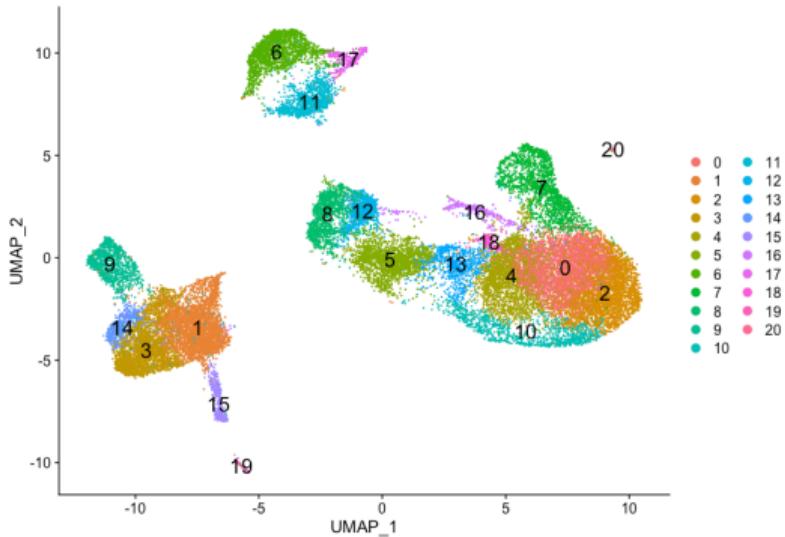


Figure 5: Image courtesy of the Harvard Chan Bioinformatics Core

- A key aim is to identify different cell types in our sample
 - How many cell types are there really?
 - How similar are cell types?
 - What genes drive the differences → Differential Expression
- ① Dimensional Reduction
 - ② Clustering Cells

Principal Component Analysis

- scRNA (+ bulk RNA) are high dimensional datasets
 - e.g. 1,000 cells \times 15,000 genes
- How do we see how similar cells are to each other?
 - Is there a way to summarise distances between cells?
 - Can we visualise in 2-dimensions?
- Principal Component Analysis (PCA) identifies direction of maximum variance

Principal Component Analysis

- scRNA (+ bulk RNA) are high dimensional datasets
 - e.g. 1,000 cells \times 15,000 genes
- How do we see how similar cells are to each other?
 - Is there a way to summarise distances between cells?
 - Can we visualise in 2-dimensions?
- Principal Component Analysis (PCA) identifies direction of maximum variance
 - Known as PC1

Principal Component Analysis

- scRNA (+ bulk RNA) are high dimensional datasets
 - e.g. 1,000 cells \times 15,000 genes
- How do we see how similar cells are to each other?
 - Is there a way to summarise distances between cells?
 - Can we visualise in 2-dimensions?
- Principal Component Analysis (PCA) identifies direction of maximum variance
 - Known as PC1
- Then we find the next most variable (orthogonal) direction \Rightarrow PC2

Principal Component Analysis

- scRNA (+ bulk RNA) are high dimensional datasets
 - e.g. 1,000 cells \times 15,000 genes
- How do we see how similar cells are to each other?
 - Is there a way to summarise distances between cells?
 - Can we visualise in 2-dimensions?
- Principal Component Analysis (PCA) identifies direction of maximum variance
 - Known as PC1
- Then we find the next most variable (orthogonal) direction \Rightarrow PC2
- Data is rotated around PC1, then PC2 etc

Principal Component Analysis

- scRNA (+ bulk RNA) are high dimensional datasets
 - e.g. 1,000 cells \times 15,000 genes
- How do we see how similar cells are to each other?
 - Is there a way to summarise distances between cells?
 - Can we visualise in 2-dimensions?
- Principal Component Analysis (PCA) identifies direction of maximum variance
 - Known as PC1
- Then we find the next most variable (orthogonal) direction \Rightarrow PC2
- Data is rotated around PC1, then PC2 etc
- Comparing PC1 vs PC2, PC2 vs PC3 etc allows 2D (or 3D) visualisation

Principal Component Analysis

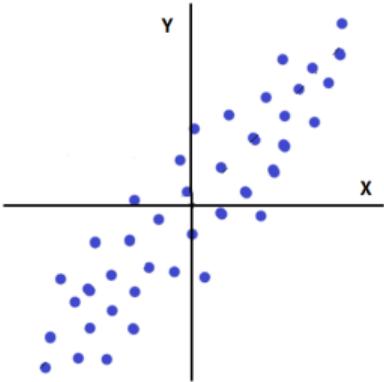


Figure 6: Some 2D data points

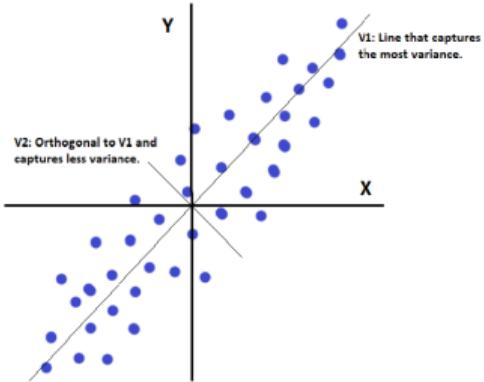


Figure 7: Find the direction of max variance

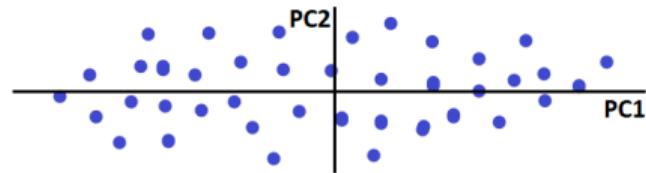


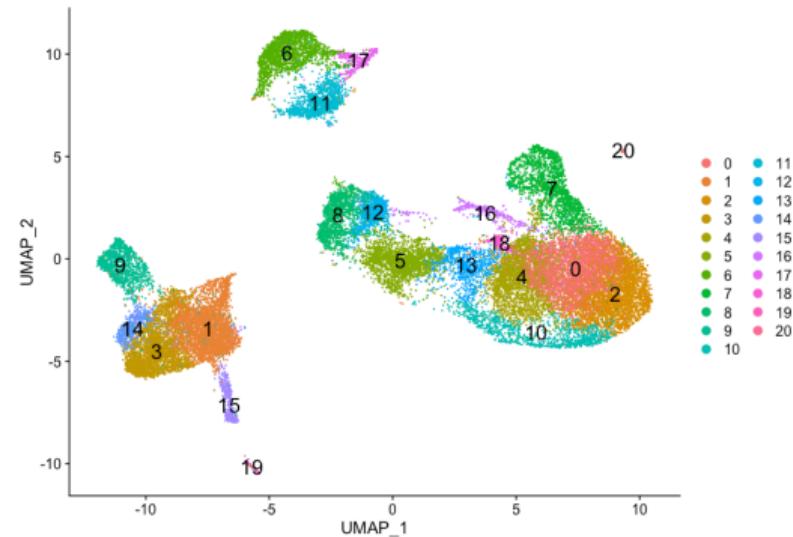
Figure 8: Each point now has a 'rotated' value

- The main directions of variability are hopefully related to the biology
- Rotations along PC1 vs PC2 should show separation by biology of interest

Figures taken from <https://statisticsbyjim.com/basics/principal-component-analysis/>

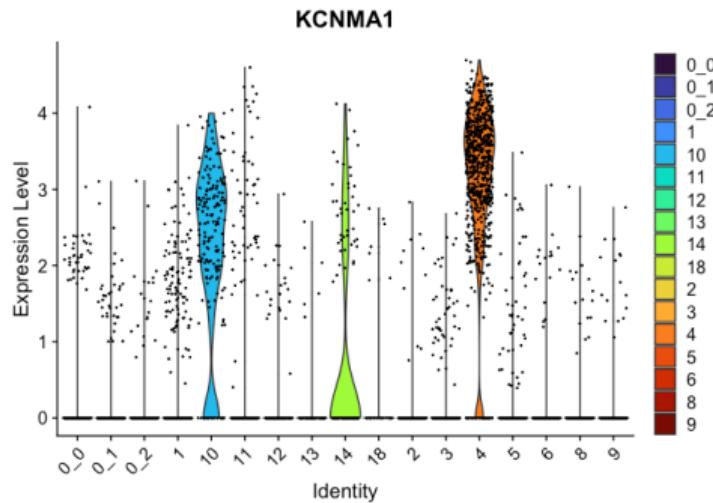
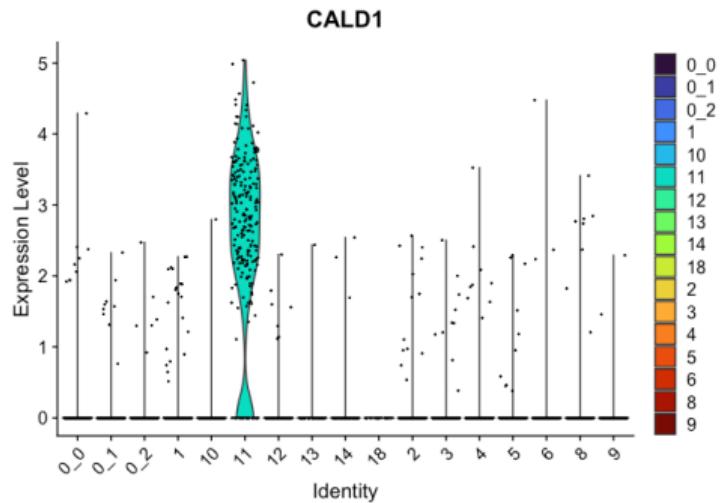
Clustering of Cells

- To form clusters \Rightarrow use top principal components
- Often use k-nearest-neighbour (kNN) graph methods to form clusters
- Visualisation methods like UMAP highlight clusters



- Huge number of possible parameters \Rightarrow subjective approach to data

Marker Detection



- Find genes unique to a given cluster or shared across multiple clusters

Figures taken from <https://ucdavis-bioinformatics-training.github.io/2023-June-Single-Cell-RNA-Seq-Analysis>

Cluster Annotation

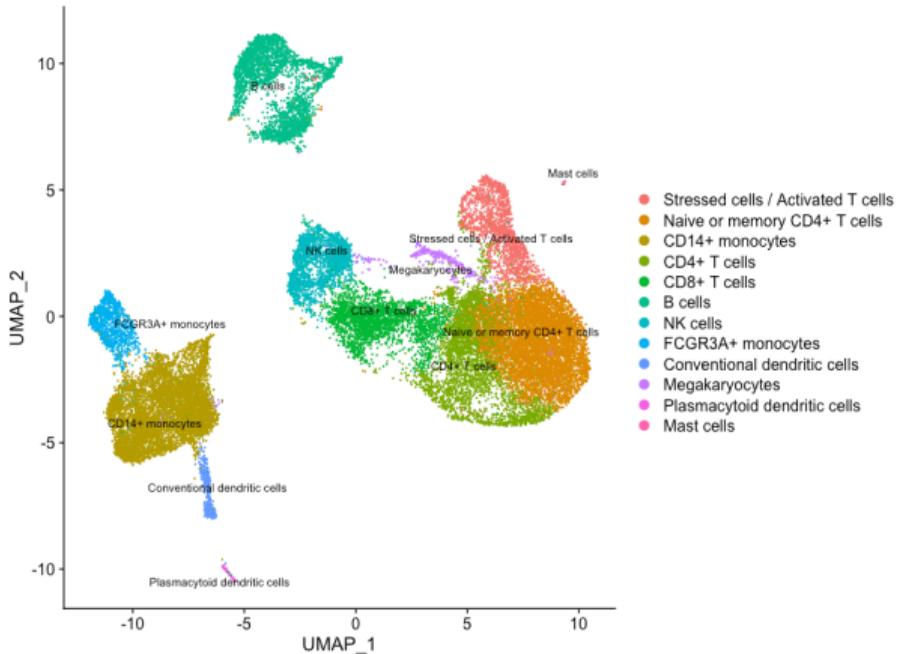


Figure 9: Image courtesy of the Harvard Chan Bioinformatics Core

- Use our expert knowledge of known marker genes to annotate clusters
- CITE-Seq uses cell-surface antibodies to aid this step
- Can drill deeply into cell-type differences & communication
 - Complementary receptor-ligand expression

Spatial Transcriptomics

Spatial Transcriptomics

- Even single cell RNA dissociates cells from position with tissues
- Cells function within highly structured tissue environments
 - Cells are likely to have different communication dynamics based on location
- *Spatial transcriptomics* attempts to resolve this limitation

Early Spatial Transcriptomics

- First appeared using barcoded primers with fixed location (Ståhl et al. 2016)
 - ~100 μm resolution
- Tissue affixed to prepared slide:
 - Stained & Visualised
 - Made Permeable
 - RT-PCR \Rightarrow cDNA
 - Tissue removed enzymatically
 - Barcoded cDNA \Rightarrow sequencing

Approaching Single Cell Resolution

- 10X Visium now down to 100 μm resolution
- *Slide-seq* (Stickels et al. 2021): Barcoded beads $\rightarrow \sim 10\mu\text{m}$ resolution
 - Approaching single cell resolution
 - Deconvolution of beads single cell
- Stereo-Seq (A. Chen et al. 2022): etched grid on slide \implies resolution of $\sim 200\text{nm}$
- Seq-Scope (Cho et al. 2021): Tissue placed directly on modified Illumina flow-cell
 \implies resolution $\sim 500\text{nm}$

Beyond Sequencing

- Barcoded probes via Fluorescence In Situ Hybridisation enable a sequencing-free perspective
- Staining technologies used to identify cell boundaries, nuclei etc
- MERFISH: multiplexed error-robust FISH (K. H. Chen et al. 2015)
 - Sequence specific barcodes able to detect ~10,000 transcripts
 - Directly count specific transcripts
 - Location within cell also identified
- Extremely rapid technological development (also \$\$\$)
- Methods often lag technology somewhat

Examples of 10X Xenium Spatial Transcriptomics

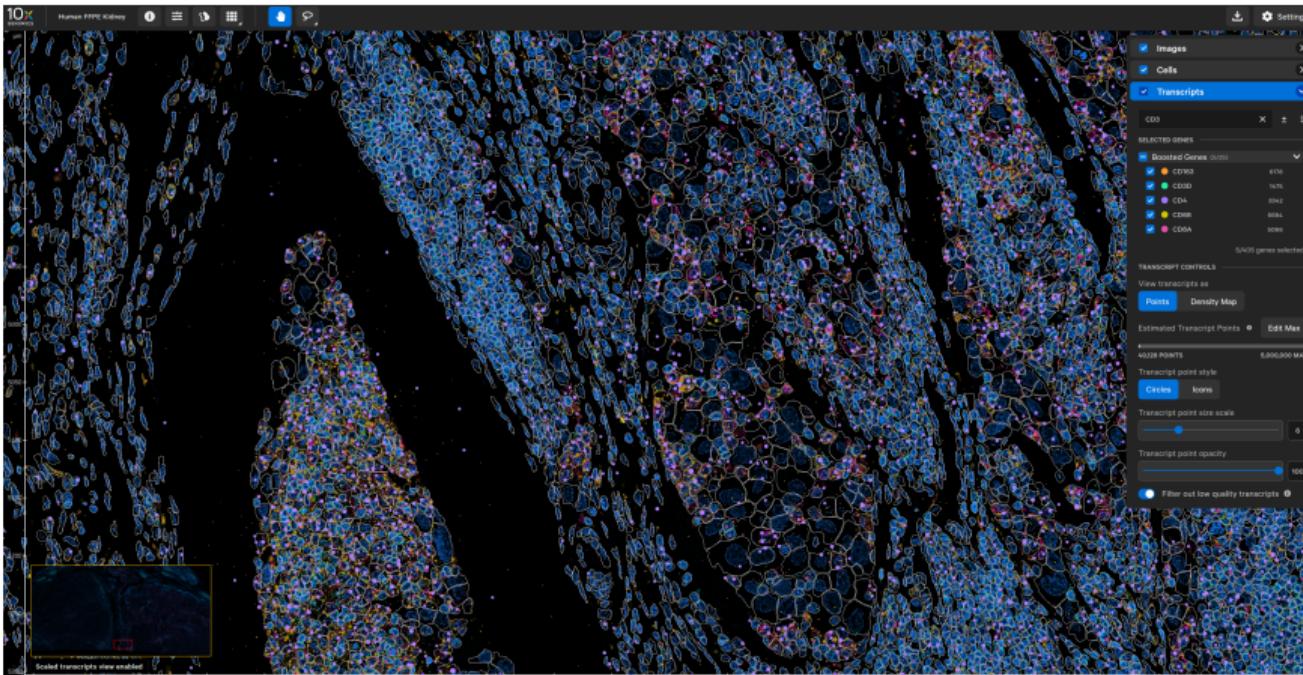
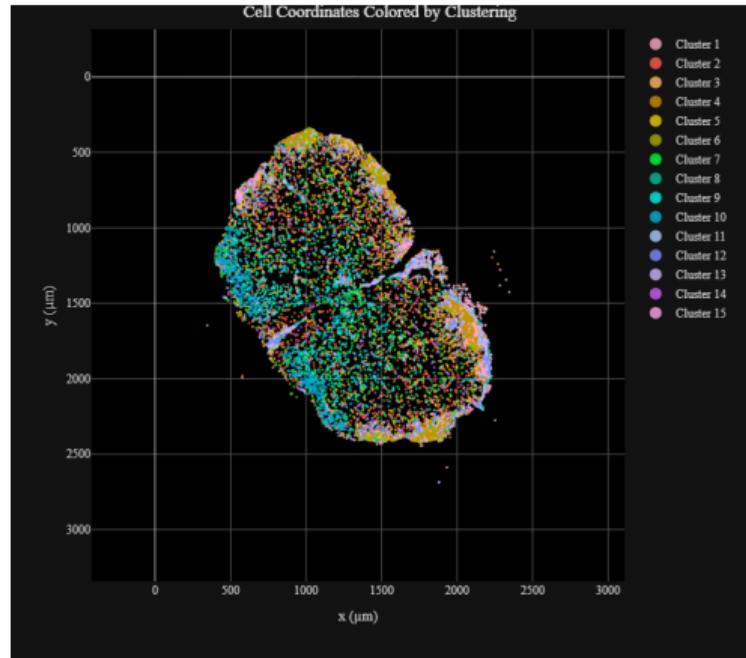
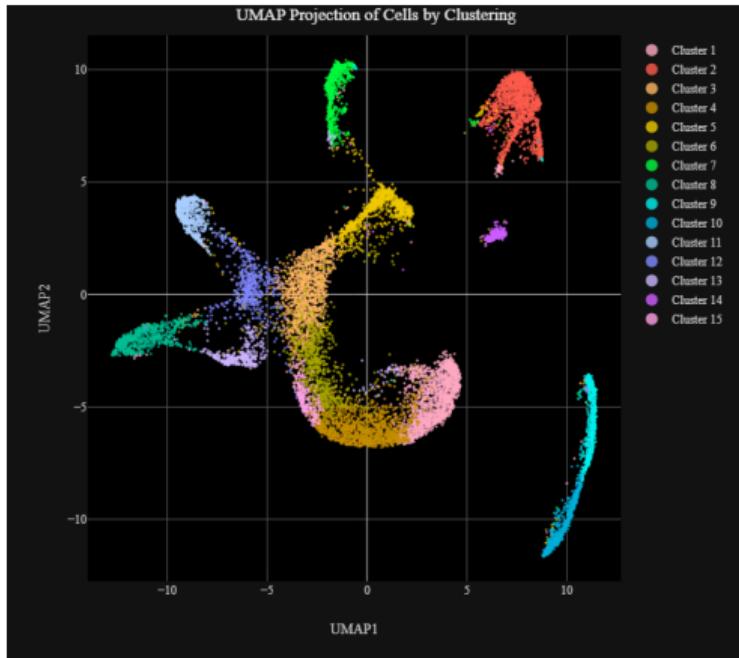


Figure 10: Taken from <https://xenium.10xgenomics.com>. Highlighting CD4 in human kidney

Examples of 10X Xenium Spatial Transcriptomics



Images courtesy of Megan Monaghan, PhD Candidate, Comerford Lab, Adelaide University

Additional Reading

- StatQuest: Principal Component Analysis (PCA)
- StatQuest: K-nearest neighbors, Clearly Explained
- StatQuest: UMAP Dimension Reduction, Main Ideas!!!

References

Chen, Ao, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, et al. 2022. "Spatiotemporal Transcriptomic Atlas of Mouse Organogenesis Using DNA Nanoball-Patterned Arrays." *Cell* 185 (10): 1777–1792.e21.

Chen, Kok Hao, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang, and Xiaowei Zhuang. 2015. "RNA Imaging. Spatially Resolved, Highly Multiplexed RNA Profiling in Single Cells." *Science* 348 (6233): aaa6090.

Cho, Chun-Seok, Jingyue Xi, Yichen Si, Sung-Rye Park, Jer-En Hsu, Myungjin Kim, Goo Jun, Hyun Min Kang, and Jun Hee Lee. 2021. "Microscopic Examination of Spatial Transcriptome Using Seq-Scope." *Cell* 184 (13): 3559–3572.e22.

Cuevas-Diaz Duran, Raquel, Haichao Wei, and Jiaqian Wu. 2024. "Data Normalization for Addressing the Challenges in the Analysis of Single-Cell Transcriptomic Datasets." *BMC Genomics* 25 (1): 444.

Haque, Ashraful, Jessica Engel, Sarah A Teichmann, and Tapio Lönnberg. 2017. "A Practical Guide to Single-Cell RNA-Sequencing for Biomedical Research and Clinical Applications." *Genome Med.* 9 (1): 75.

Rosenberg, Alexander B, Charles M Roco, Richard A Muscat, Anna Kuchina, Paul Sample, Zizhen Yao, Lucas T Graybuck, et al. 2018. "Single-Cell Profiling of the Developing Mouse Brain and Spinal Cord with Split-Pool Barcoding." *Science* 360 (6385): 176–82.

Ståhl, Patrik L, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, et al. 2016. "Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics." *Science* 353 (6294): 78–82.

Stickels, Robert R, Evan Murray, Pawan Kumar, Jilong Li, Jamie L Marshall, Daniela Di Bella, Paola Kröckert, Evan Z Macosko, and Fei Chen. 2021. "Highly Sensitive Spatial Transcriptomics at Near-Cellular Resolution via Siliene V2." *Nat. Biotechnol.* 39 (2): 212–19.



BLACK
OCHRE
DATA
LABS

