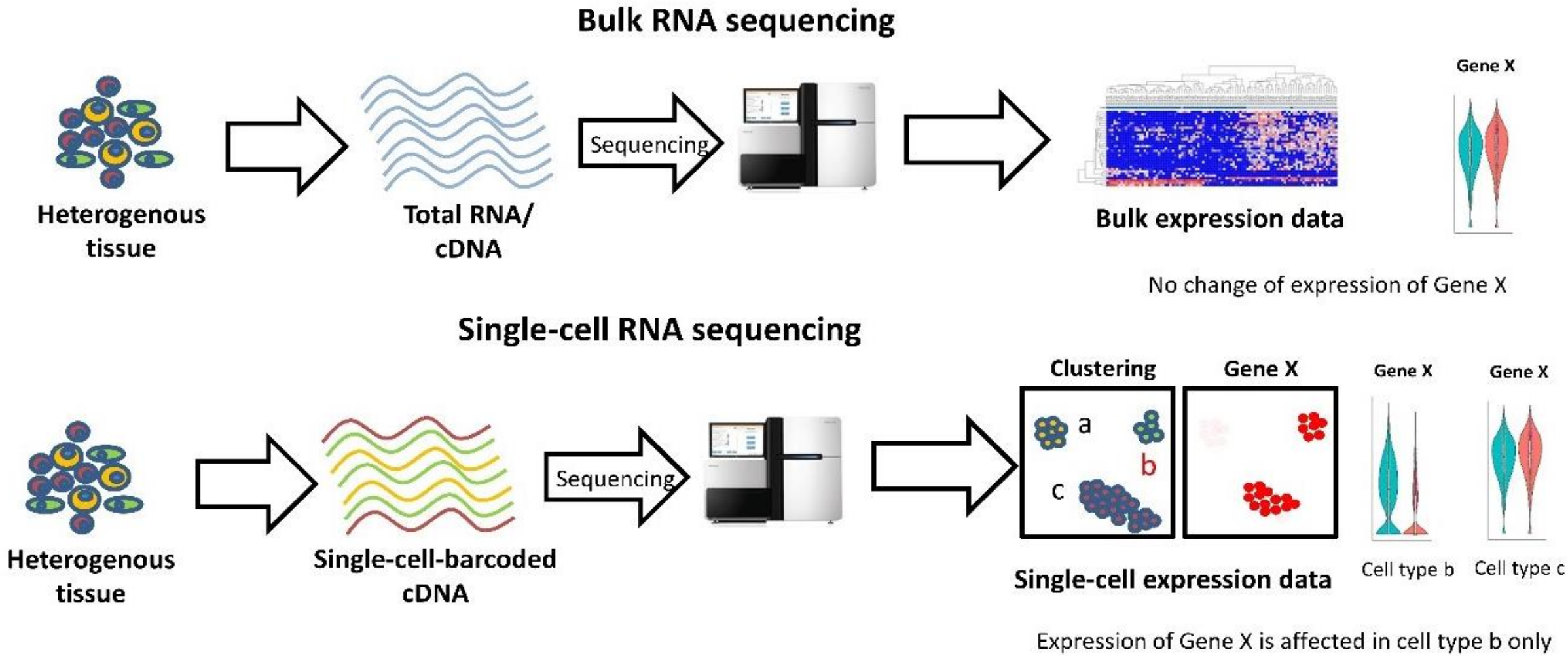
Single cell RNA-sequencing: A primer

CRI Bioinformatics Core Facility
Surajit Bhattacharya
11/21/2019

Single Cell RNA seq: What is it?

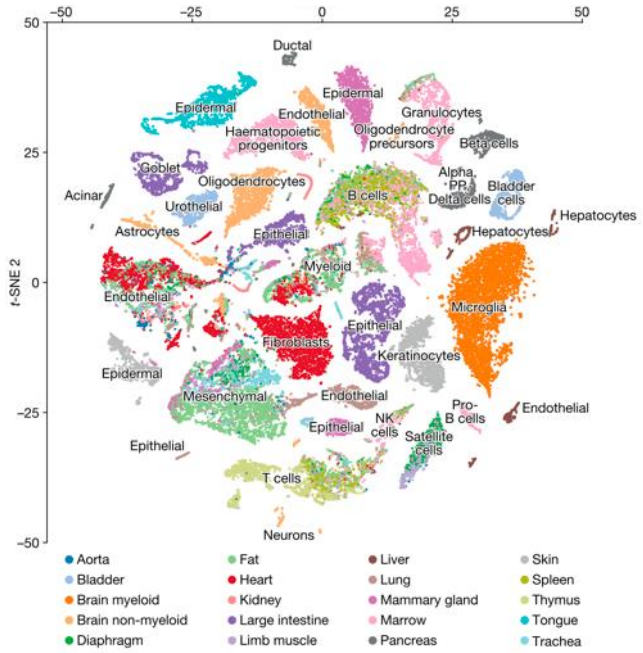


https://www.technologynetworks.com/genomics/articles/recent-advances-in-single-cell-genomics-techniques-324695

Single cell Studies: What we can learn.

Single–cell transcriptomics of 20 mouse organs creates a *Tabula Muris*

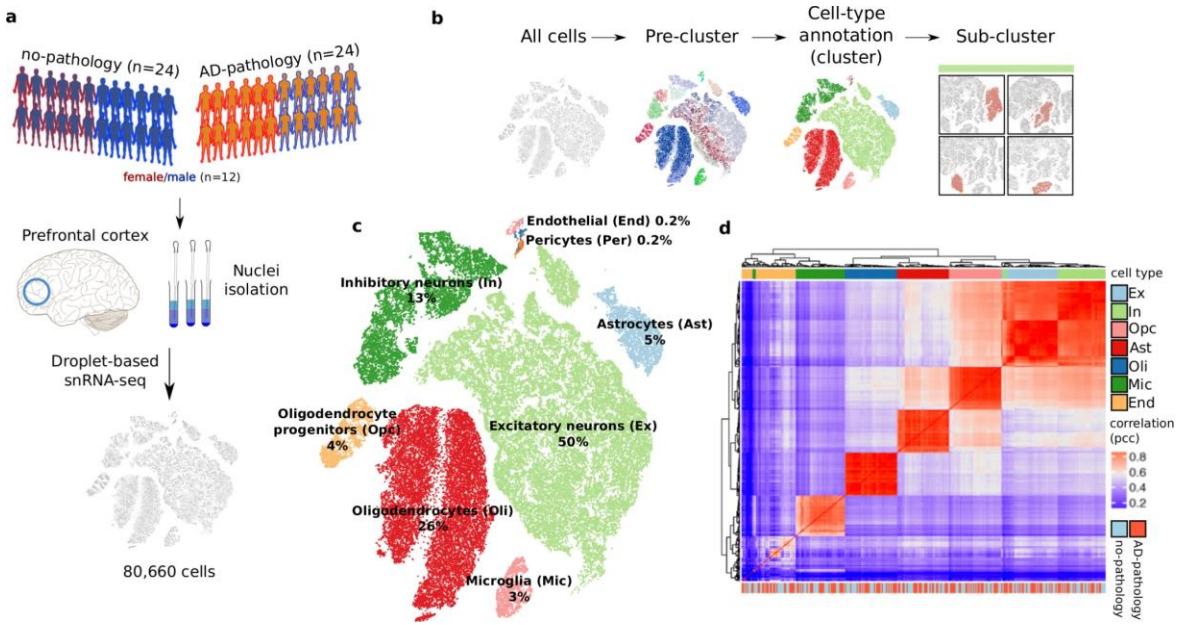
The Tabula Muris Consortium*



- Profiled 100k cells from 20 organs and tissues.
- Approaches used
  - ✓ High throughput 3’-end counting approach.
  - ✓ FACs sorting of full transcriptomics.
- Foundation for a mouse single cell atlas.

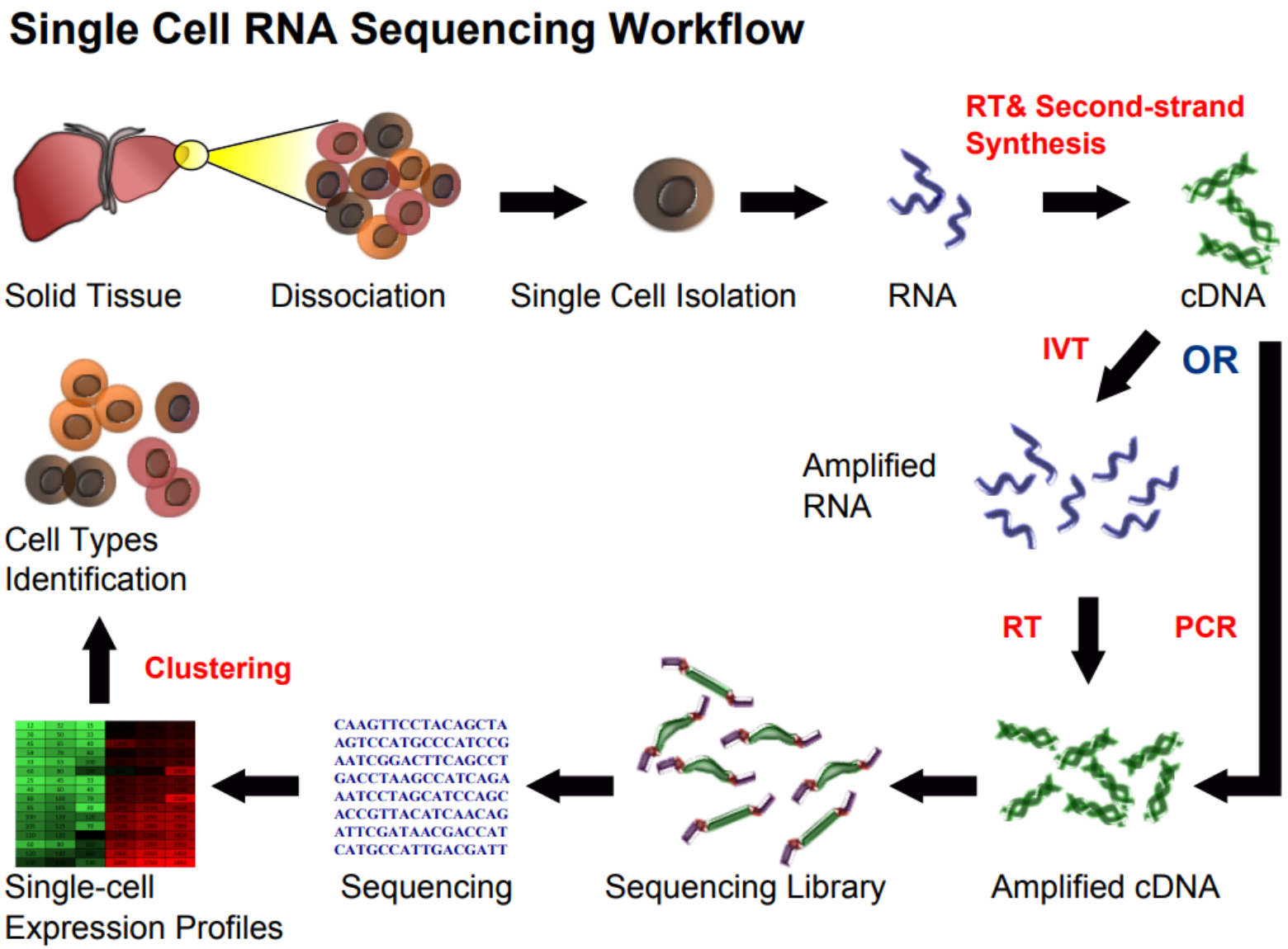
Article | Published: 01 May 2019

Single-cell transcriptomic analysis of Alzheimer’s disease



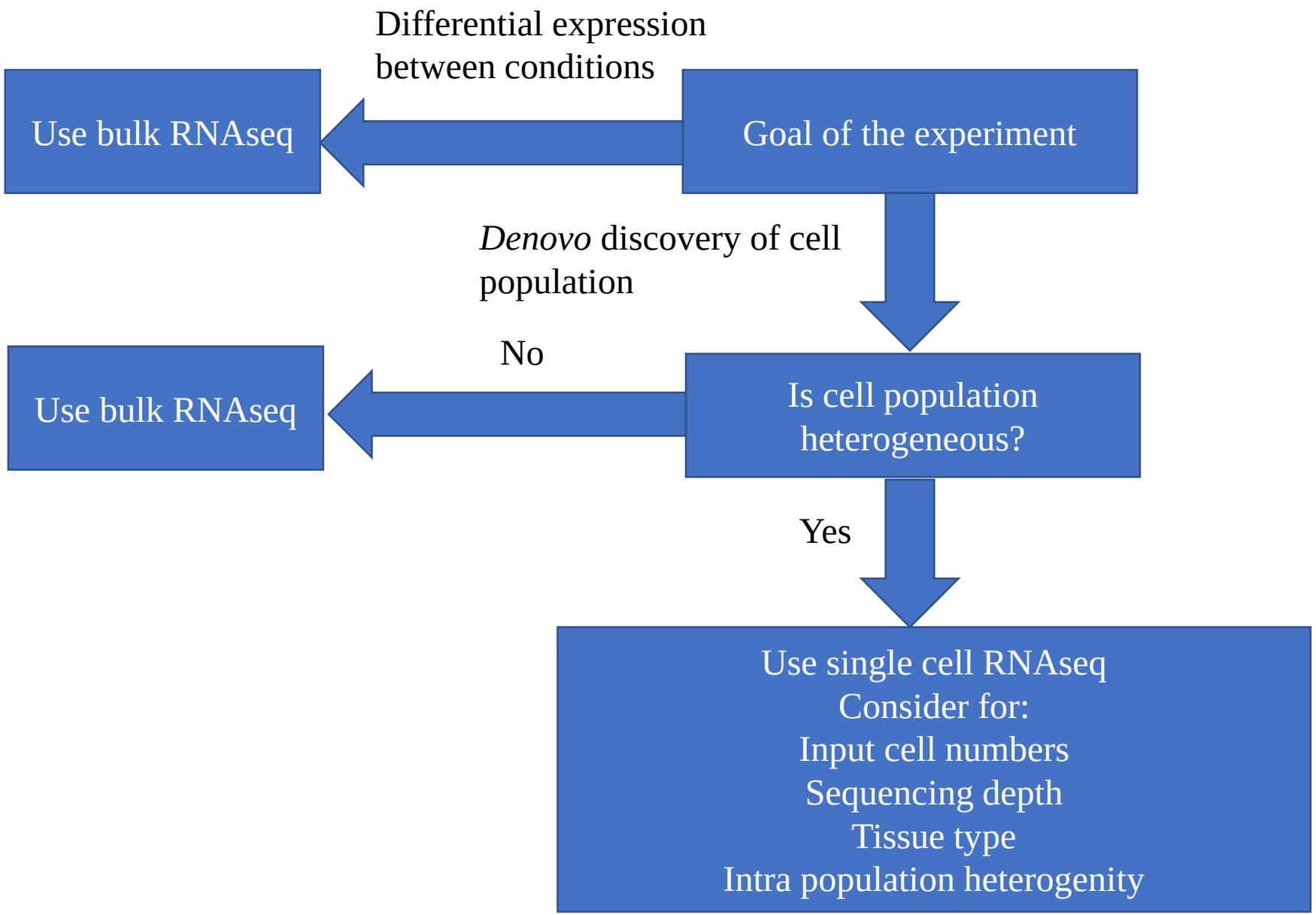
- Profiled 80,600 cells from prefrontal cortex of 48 individuals with varying degrees of Alzheimer’s disease pathology.
- Transcriptionally distinct subpopulations identified.
- Female cells were overrepresented in disease-associated subpopulations and transcriptional responses were substantially different between sexes

Single Cell RNA seq: Workflow

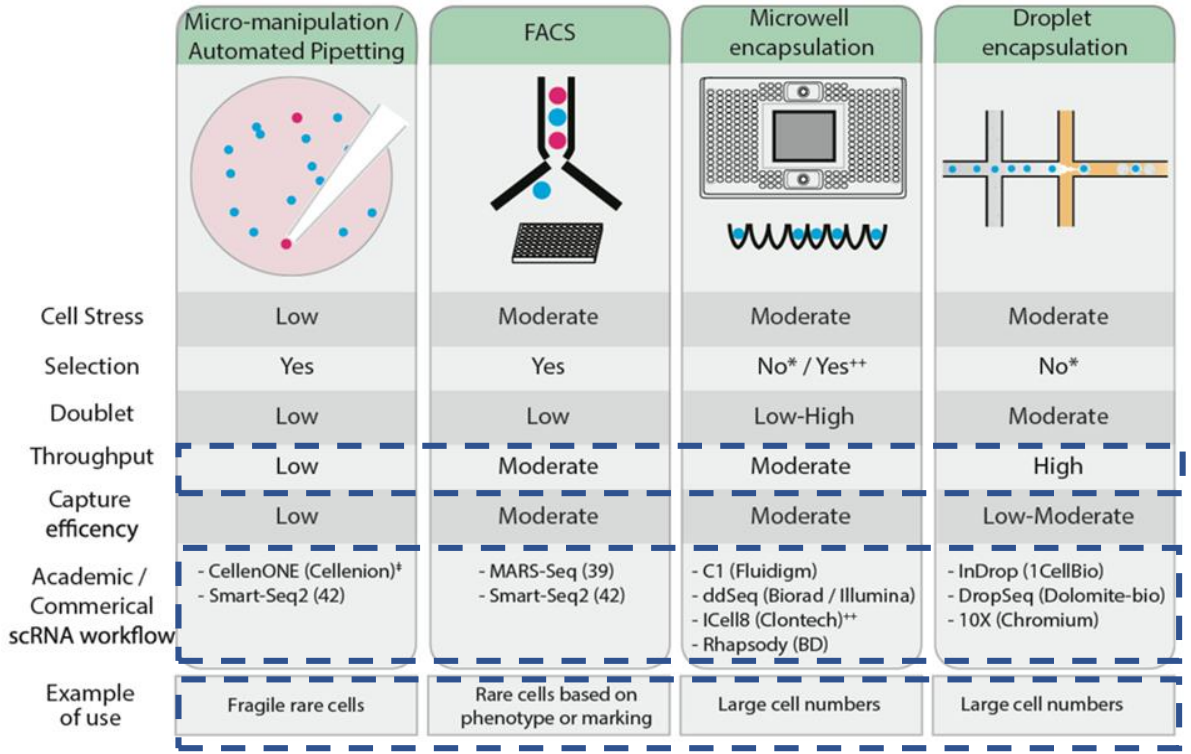


https://upload.wikimedia.org/wikipedia/commons/d/d5/Single_cell_RNA-Seq_workflow.pdf

Single Cell RNA seq: Design of the experiment



Single Cell methods: isolation and library generation

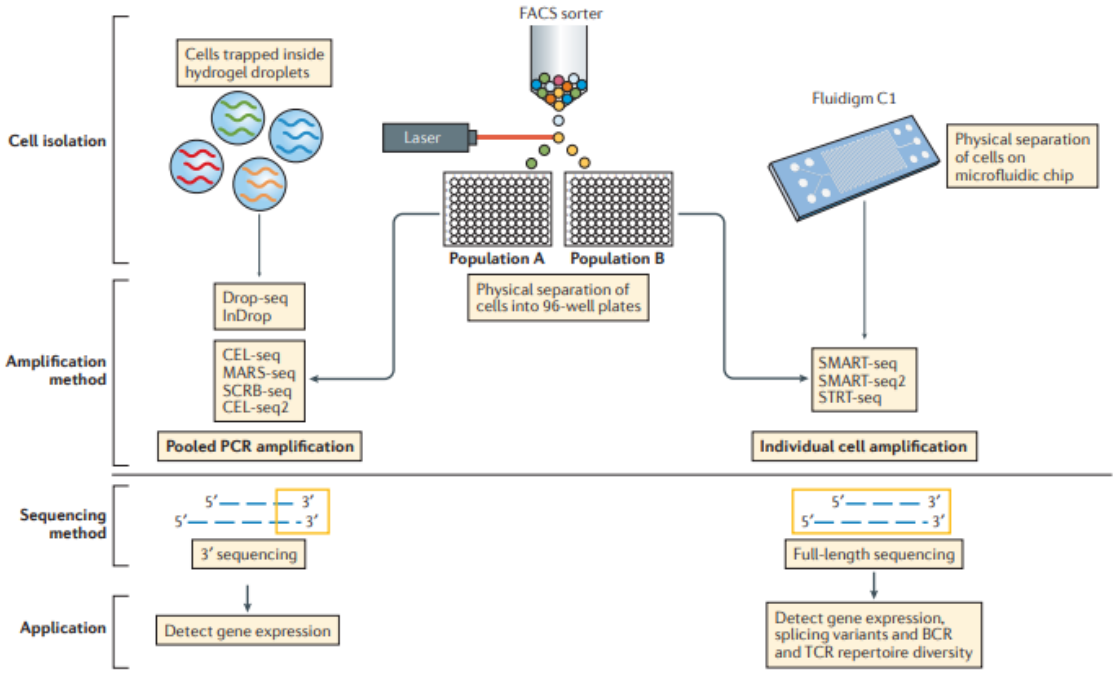


	Micro-manipulation / Automated Pipetting	FACS	Microwell encapsulation	Droplet encapsulation
Cell Stress	Low	Moderate	Moderate	Moderate
Selection	Yes	Yes	No* / Yes++	No*
Doublet	Low	Low	Low-High	Moderate
Throughput	Low	Moderate	Moderate	High
Capture efficency	Low	Moderate	Moderate	Low-Moderate
Academic / Commerical scRNA workflow	- CellenONE (Cellenion)‡ - Smart-Seq2 (42)	- MARS-Seq (39) - Smart-Seq2 (42)	- C1 (Fluidigm) - ddSeq (Biorad / Illumina) - ICell8 (Clontech)++ - Rhapsody (BD)	- InDrop (1CellBio) - DropSeq (Dolomite-bio) - 10X (Chromium)
Example of use	Fragile rare cells	Rare cells based on phenotype or marking	Large cell numbers	Large cell numbers

	FACS		Microwell encapsulation				Droplet encapsulation		
	Smart-Seq2	MARS-Seq	C1	ddSeq	ICell8	Rhapsody	InDrop	DropSeq	10X
Singlet Capture efficency	82%	92%	39%	2.6%	37%++	Not reported	7%	Not reported	50%
Doublet rate	Not reported	2.27%	3-30%	5.8%	1.3-4%	0.6%	4%	0.36-11.3	1.6-3%
Reference	42	39	37 FWP	PB	PB	PB	36	37	26

https://www.frontiersin.org/articles/10.3389/fimmu.2018.01553/full

Single Cell methods: isolation and library generation



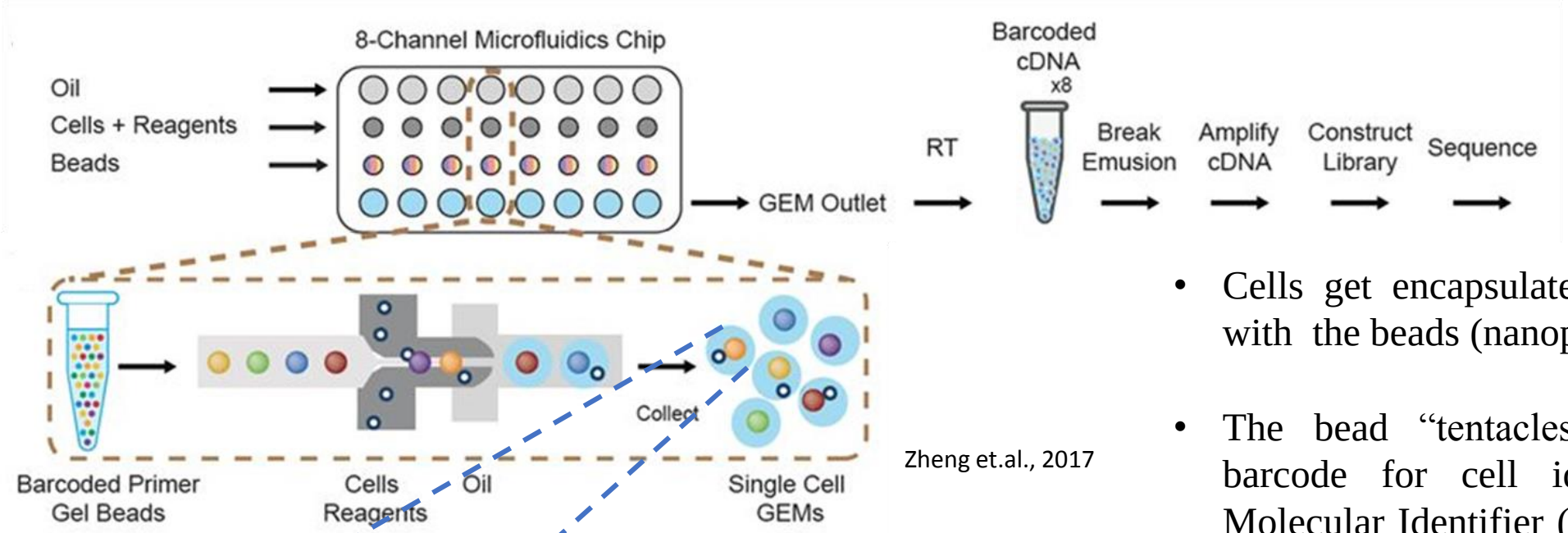
https://www.nature.com/articles/nri.2017.76

	Full length			3′ sequencing and barcoding					
Applications	Gene expression Splice variants and BCR and TCR repertoire diversity			Gene expression					
Costs	High			Low					
	Smart-Seq2	Smarter /ICell8/C1	NuGEN Solo	MARS-Seq	ddSeq	Rhapsody	InDrop	DropSeq	10X*
UMI	-	-	√	√	√	√	√	√	√
mRNA priming (1st strand syn)	poly T	poly T	Random priming & poly T	poly T	poly T	poly T	poly T	poly T	poly T
Template Switching	√	√	-	-	-	-	-	√	√
DNase treatment	-	-	√	√	-	-	-	-	-
cDNA preamplification	PCR	PCR	-	In Vitro Transcription	PCR	PCR	In Vitro Transcription	PCR	PCR
Targeted sequencing	-	-	Depletion	-	-	Enrichment	-	-	-
Library generation	Transposon Tagmentation	Transposon Tagmentation	cDNA fragmentation, adapter ligation & library amp	RNA fragmentation & adapter ligation	Transposon Tagmentation	PCR targeted primer panels	RNA fragmentation & adapter ligation	Transposon Tagmentation	cDNA fragmentation, adapter ligation & library amp
Example of use	Sequencing the TCR of tumour-infiltrating lymphocytes			High-throughput sequencing of large cell numbers from solid organ tumours in large patient cohorts					

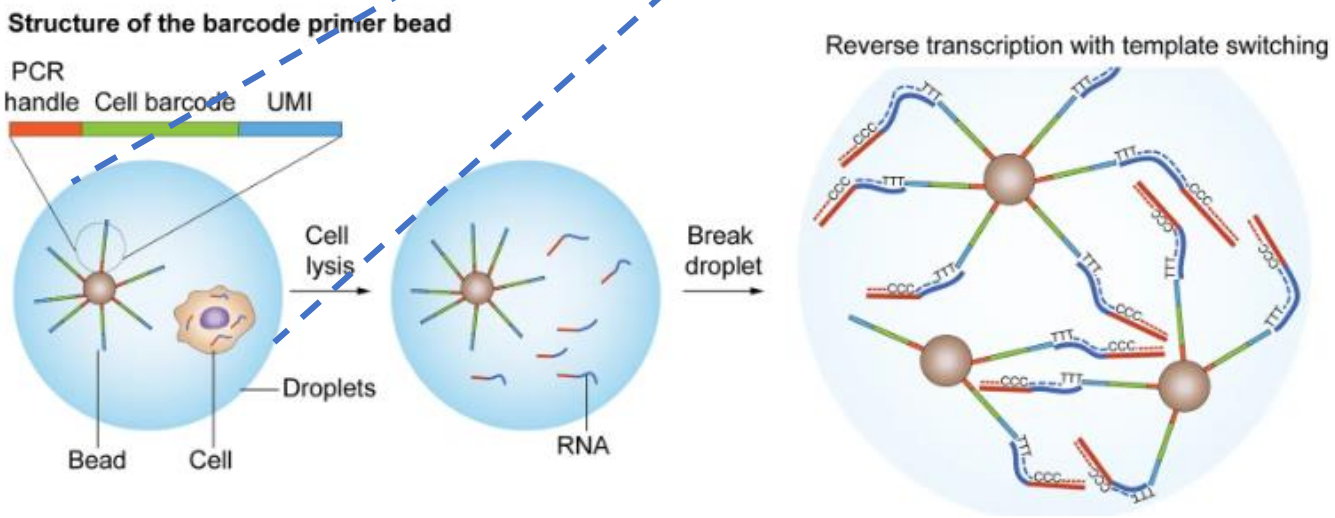
*10X has recently released a 5′ barcoding that allows reconstruction of full length idiotype sequences

https://www.frontiersin.org/articles/10.3389/fimmu.2018.01553/full

Single Cell methods: droplet based method (10X)



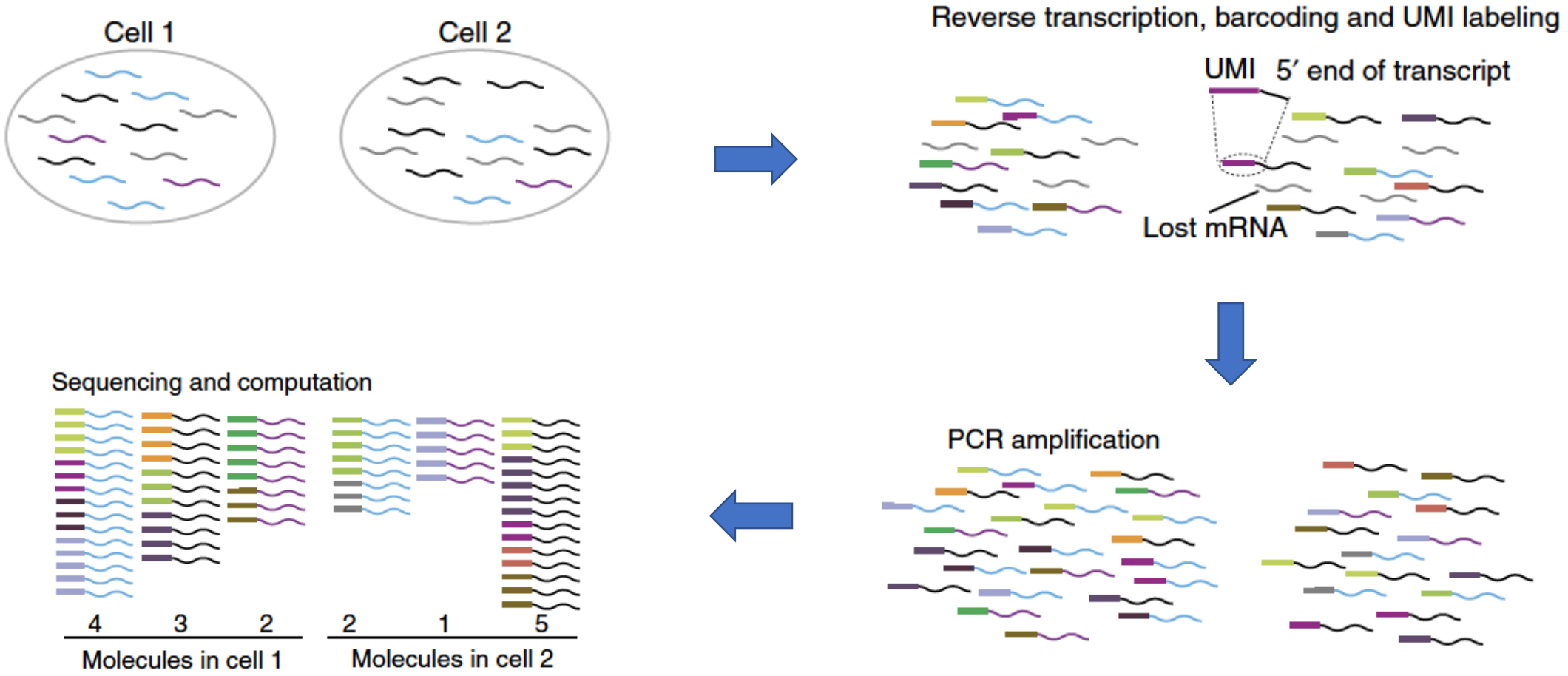
Zheng et.al., 2017



Hwang et.al, 2018

- Cells get encapsulated into oil droplets along with the beads (nanoparticles) and lysis buffer.
- The bead "tentacles" have PCR handle, a barcode for cell identification, an Unique Molecular Identifier (UMI), and a "30 bp oligo-dT to prime polyadenylated RNA transcripts".
- Cells are lysed. Oligo DT primer of the beads, reverse transcribes the poly-A RNA to cDNA.
- cDNA thus formed has a cell barcode, a UMI and template switching oligo at 3'end.
- Emulsion is broken. cDNA extracted and amplified, library prepared and 3'sequenced.

Unique molecular identifiers (UMIs)



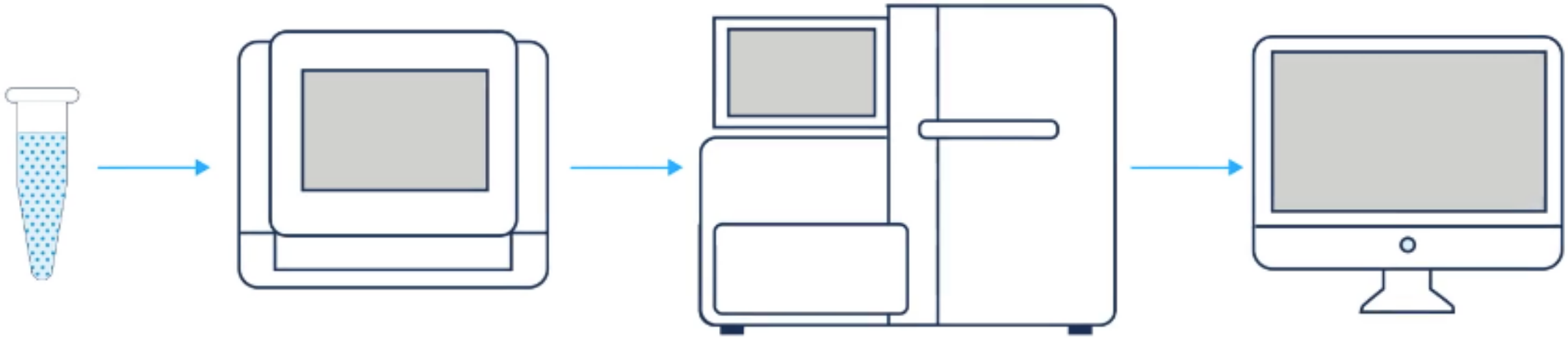
Overview of 10x Single-cell RNA-Seq pipeline

Single cell/nuclei suspension

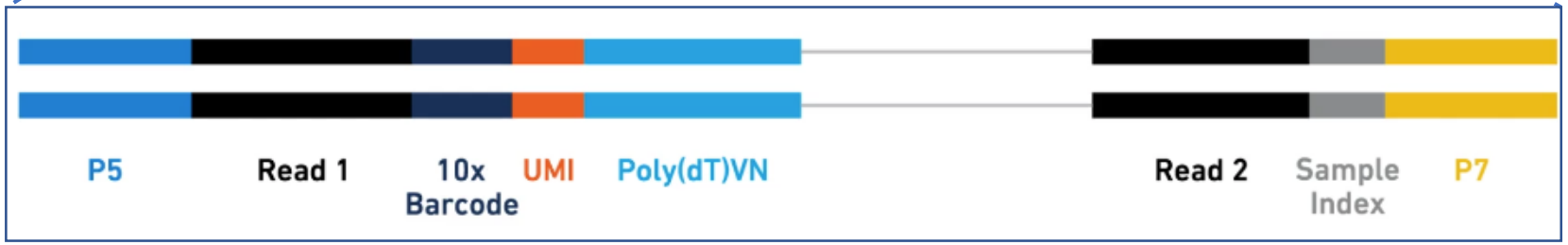
10x Chromium

Illumina Sequencer

Data analysis



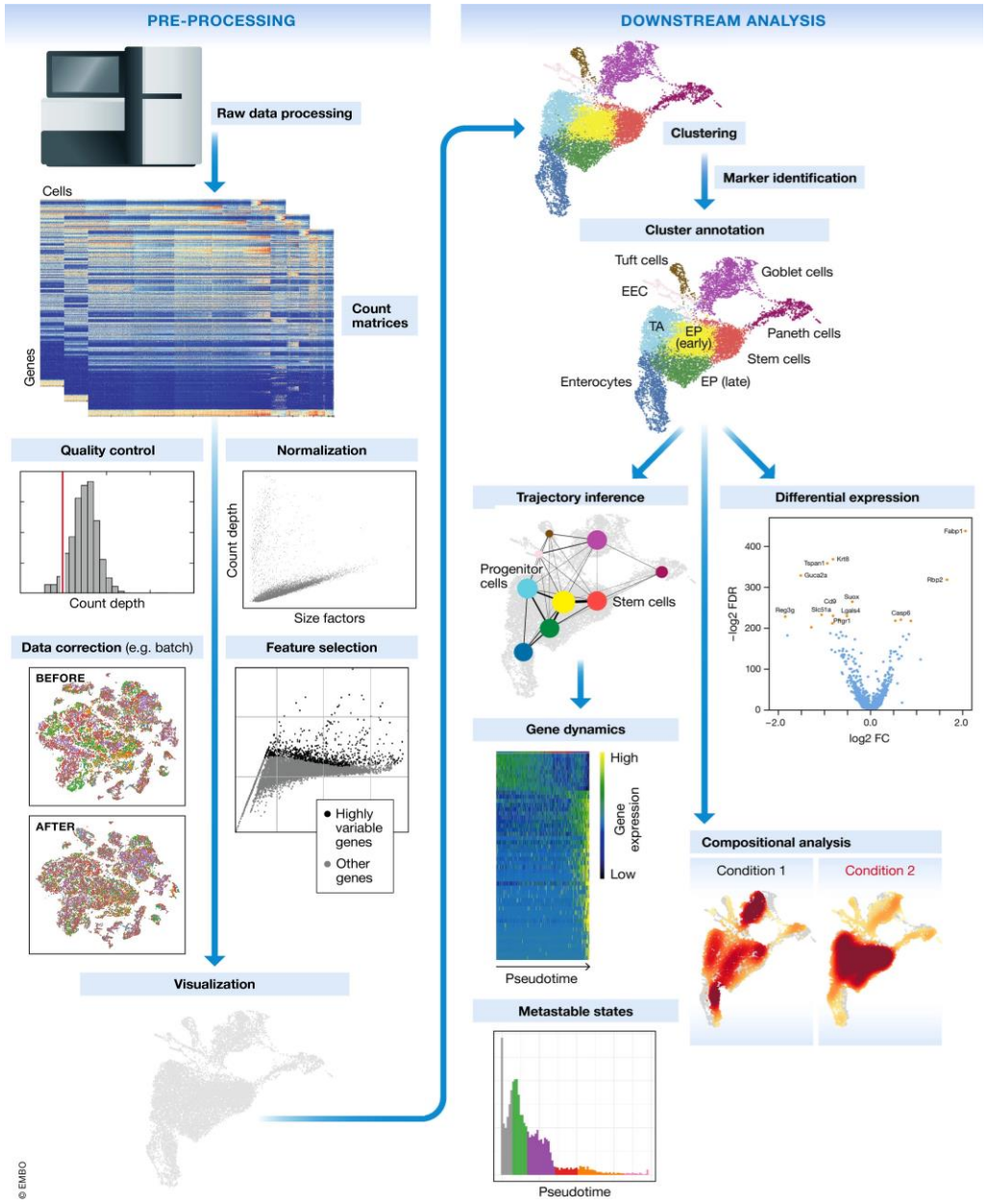
- Read1 consists of Illumina p5 primer, Truseq read1 primer, 10x barcode (16 bases), UMI (12 bases) and Poly dt oligonucleotide.
- Read2 consists of cDNA insert, Truseq read2 primer, Sample index and Illumina p7 primer.



Cost and amount of sample required

- **Single RNA Sequencing**
  - Cell isolation and addition of 10X barcodes: $500 per sample.
  - Sequencing: 8,000 cells you need 288, 000, 000 good reads, with sequencing cost in Novogene approximately $1300
  - Total cost: Approximately $2000 to $3000 per sample depending on number of cells sequenced
  - RIN ≥8, 260/280/230 ratios, 700-1200 cell/ul.
  - Check Cell viability **before** and **after** the nuclei/cell isolation
  - A **picture** of the cell suspension (to make sure that the suspension is free from clumps, aggregates, debris that can cause a clog)
  - Wash twice the cells or nuclei prior to loading on the chip in PBS + BSA 0.04%, sometimes the nuclei especially from brain are very sticky and the tend to clump together, using BSA 2% will help to prevent the clumping.

Bioinformatics pipeline: Single Cell RNA-seq



https://www.embopress.org/doi/10.15252/msb.20188746

Raw reads
(Fastq)

↓

Alignment and Quantification
(Alignment: STAR,
Quantification: CellRanger (10X))

↓

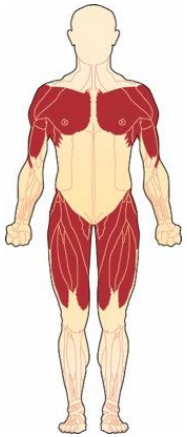
Clustering , Differential Expression,
visualization,
(Seurat and Monocle)

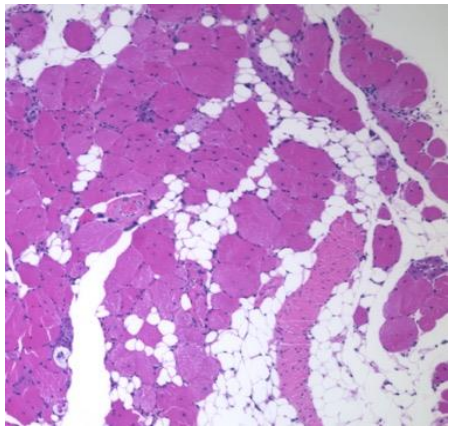
↓

Downstream Functional Analysis
(DAVID)

Case Study: Limb Girdle Muscular Dystrophy

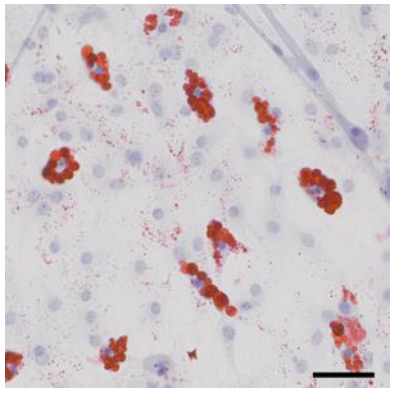
- Mutations in Dysferlin cause Limb Girdle Muscular Dystrophy Type 2B, a rare disease which features progressive muscle weakness in the shoulder and hip joints.
- A major driver of the pathology is the replacement of functional muscle fibers with fat.
- Major Goal of this experiment, **how is the fat forming?**



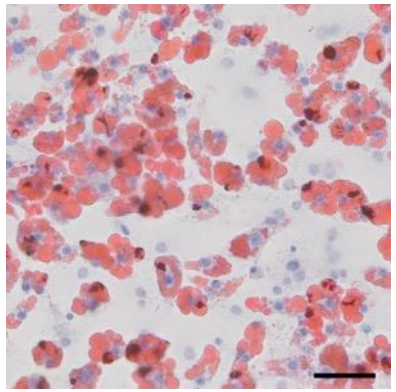


12month old dysferlin knock out (KO) mouse (12Mo B6A/J) quadriceps showing replacement of muscle with fat.

12Mo WT

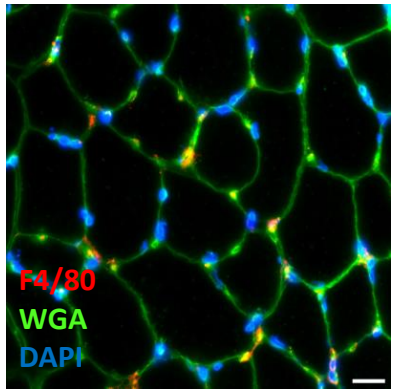


12Mo B6A/J

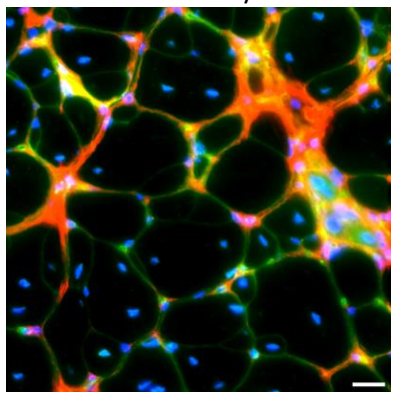


Dysferlin KO muscle shows a progressive accumulation of fibro/adipogenic progenitors (FAPs), compared to Wild Type (WT).

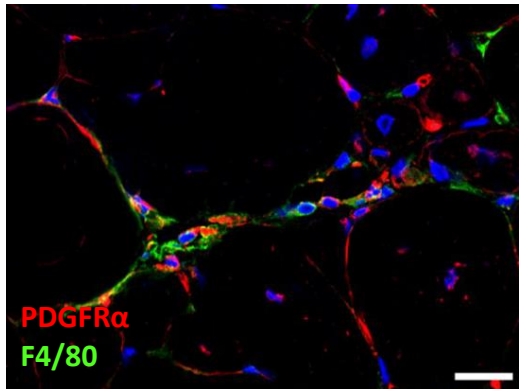
12Mo WT



12Mo B6A/J

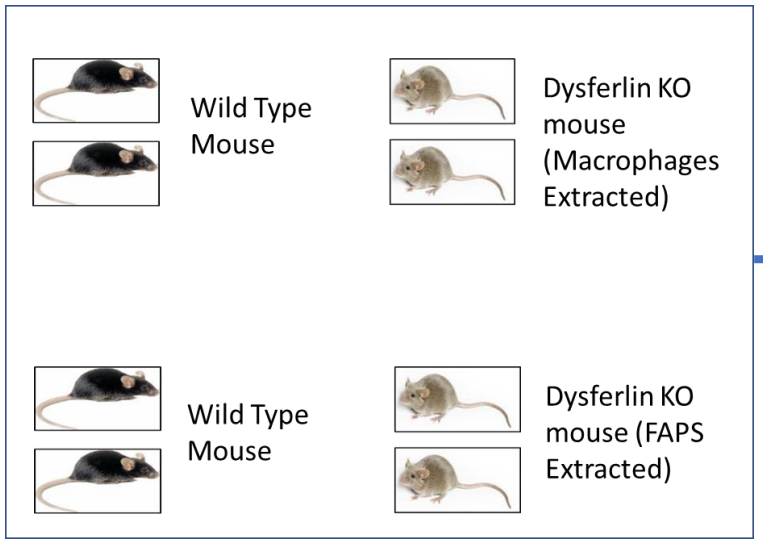


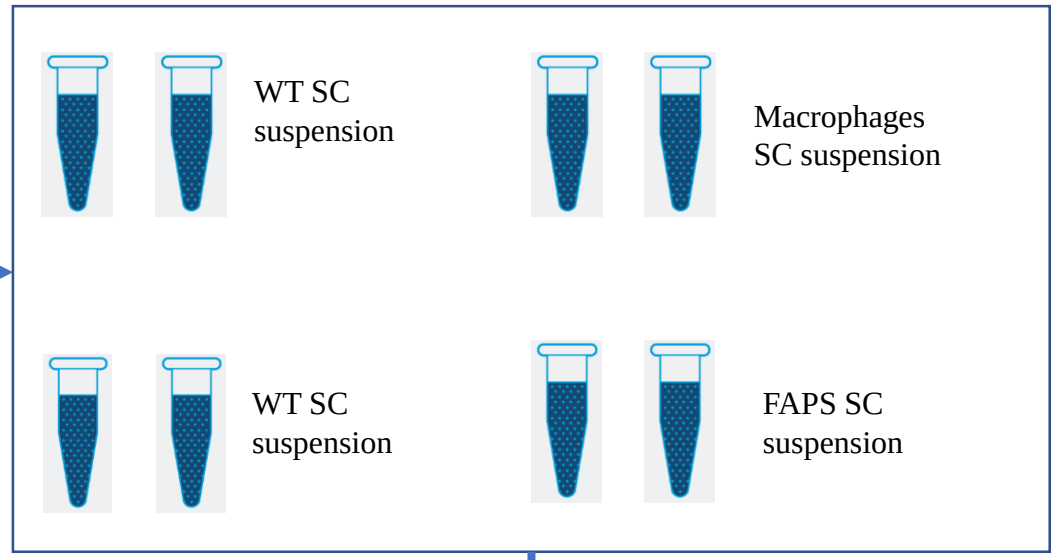
Macrophages are known to influence FAPs. In a Dysferlin KO mouse, macrophages(F4/80) are increased, compared to WT.

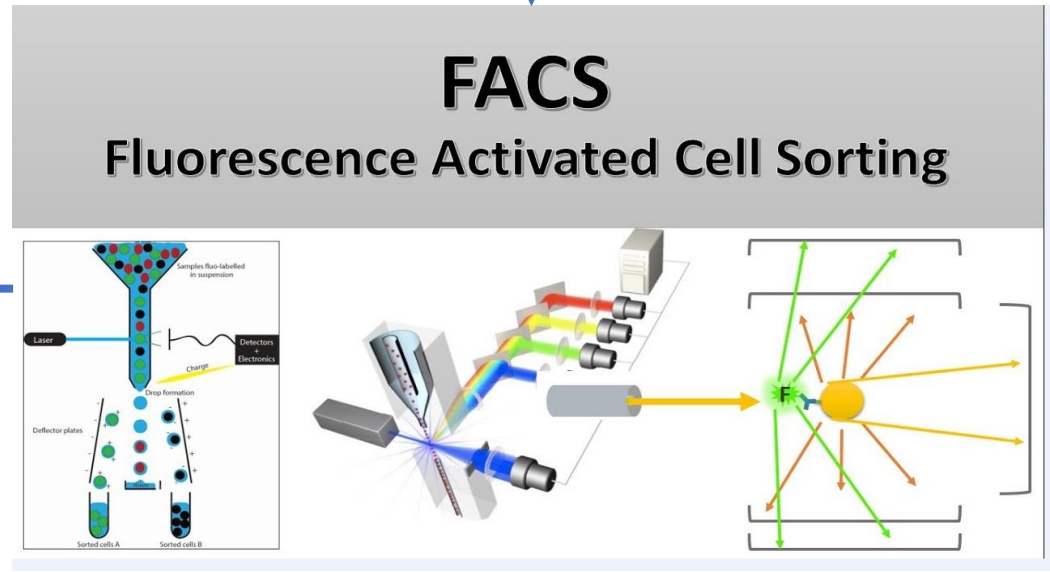


Macrophages and FAPs (PDGFRα) are closely associated in areas of pathology in dysferlin-deficient muscle.

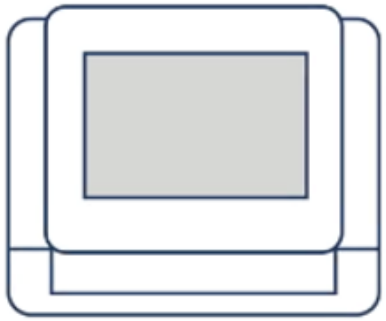
Work Flow of the experiment



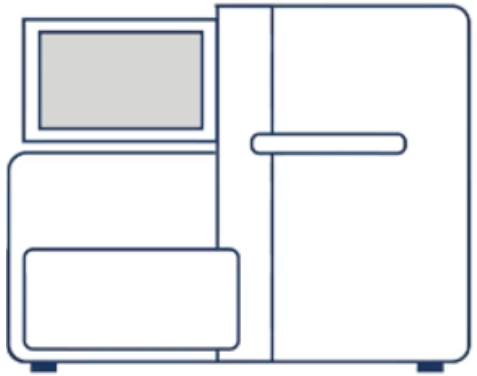




10x Chromium



Illumina Sequencer



Summary of Samples: Quality Metrics: Clean Sample

10x GENOMICS™

Cell Ranger · W1M_exp ·

SUMMARY ANALYSIS

Estimated Number of Cells

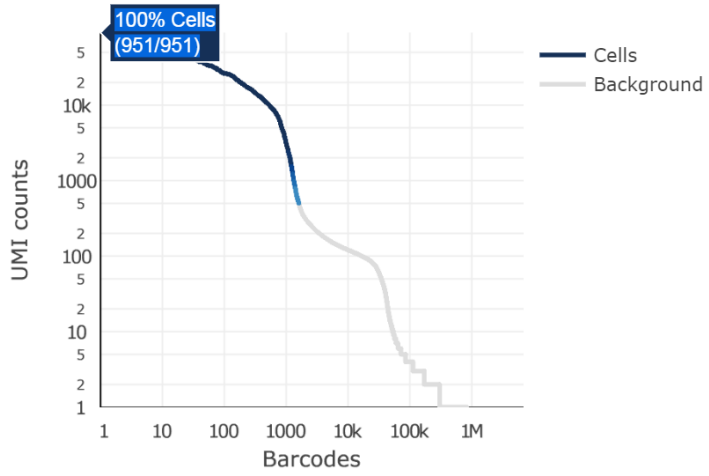
1,424

Mean Reads per Cell	Median Genes per Cell
174,560	2,317

Sequencing

Number of Reads	248,574,748
Valid Barcodes	95.5%
Sequencing Saturation	85.1%

Cells



Sequencing

Number of Reads	248,574,748
Valid Barcodes	95.5%
Sequencing Saturation	85.1%
Q30 Bases in Barcode	97.7%
Q30 Bases in RNA Read	87.9%
Q30 Bases in Sample Index	94.6%
Q30 Bases in UMI	97.4%

Mapping

Reads Mapped to Genome	90.0%
Reads Mapped Confidently to Genome	87.0%
Reads Mapped Confidently to Intergenic Regions	6.7%
Reads Mapped Confidently to Intronic Regions	20.3%
Reads Mapped Confidently to Exonic Regions	60.0%
Reads Mapped Confidently to Transcriptome	55.7%
Reads Mapped Antisense to Gene	1.1%

Estimated Number of Cells	1,424
Fraction Reads in Cells	81.3%
Mean Reads per Cell	174,560
Median Genes per Cell	2,317
Total Genes Detected	16,265
Median UMI Counts per Cell	7,405

Primary Quality matrices are: Estimated Number of cells, Reads Mapped Confidently to Transcriptome, Reads mapped Antisense to gene, Fractions of Reads per cell.

Summary of Samples: Quality Metrics: Contaminated Sample

10x GENOMICS™

Cell Ranger · W2M_exp ·

1 SUMMARY ANALYSIS

The analysis detected some issues. Details »

Estimated Number of Cells

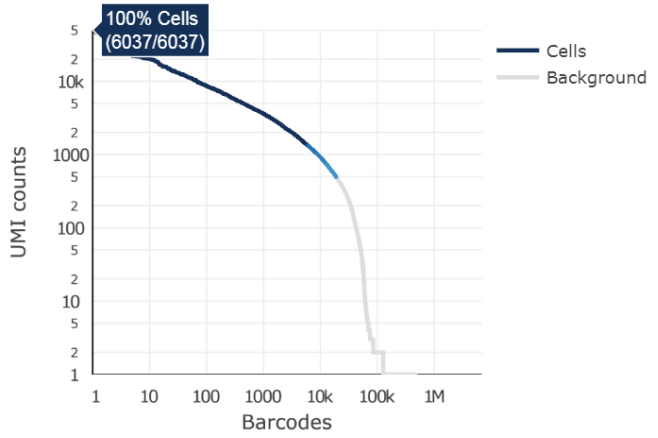
12,077

Mean Reads per Cell	Median Genes per Cell
18,317	741

Sequencing

Number of Reads	221,226,020
Valid Barcodes	96.6%
Sequencing Saturation	67.2%

Cells



Sequencing

Number of Reads	221,226,020
Valid Barcodes	96.6%
Sequencing Saturation	67.2%
Q30 Bases in Barcode	97.9%
Q30 Bases in RNA Read	89.1%
Q30 Bases in Sample Index	95.1%
Q30 Bases in UMI	97.7%

Mapping

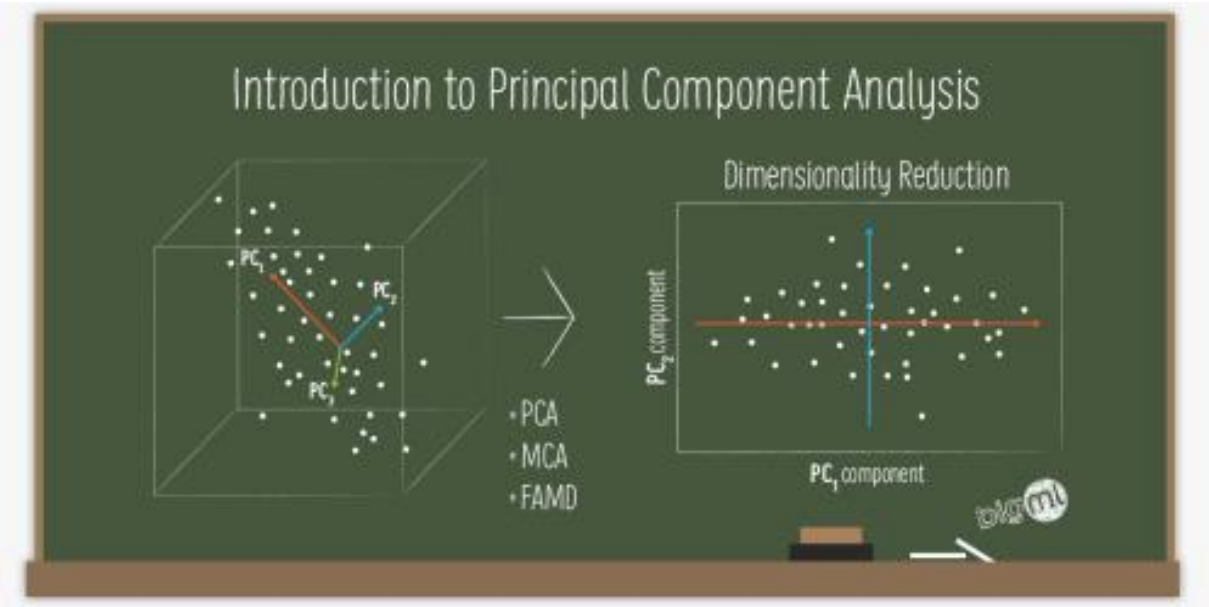
Reads Mapped to Genome	89.6%
Reads Mapped Confidently to Genome	86.0%
Reads Mapped Confidently to Intergenic Regions	7.6%
Reads Mapped Confidently to Intronic Regions	25.5%
Reads Mapped Confidently to Exonic Regions	52.9%
Reads Mapped Confidently to Transcriptome	48.2%
Reads Mapped Antisense to Gene	2.0%

The analysis detected some issues. Details »

Alert	Value	Detail
⚠ **Low Fraction Reads in Cells**	64.0%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

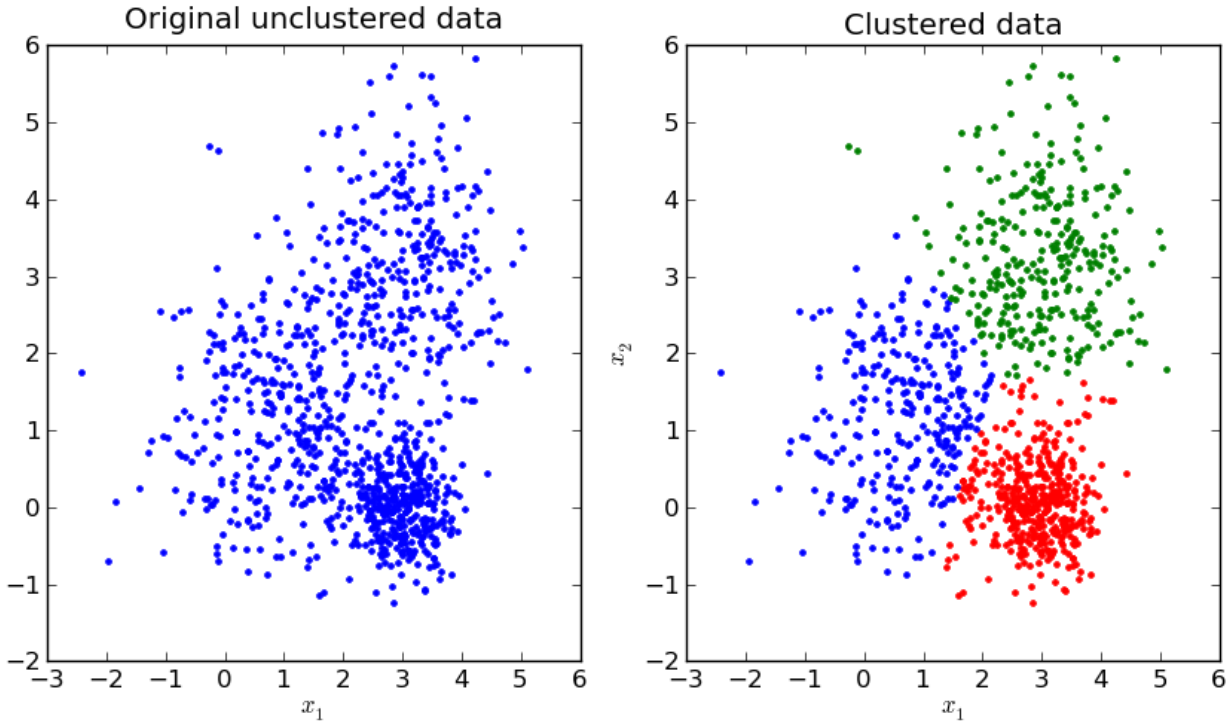
Estimated Number of Cells	12,077
Fraction Reads in Cells	64.0%
Mean Reads per Cell	18,317
Median Genes per Cell	741
Total Genes Detected	17,270
Median UMI Counts per Cell	1,324

Principal component analysis (PCA)



- “In statistics, machine learning, and information theory, dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables.”(Wikipedia)
- In other words we have ~1500 cells per sample and ~750 genes per cells (1500 X 750 dimensions), it is completely impossible to analyze this data with current available computational power we reduce the number of dimensions to more reasonable numbers for ease of computation.
- Principle component Analysis (**PCA**) is one method used for dimensionality reduction.

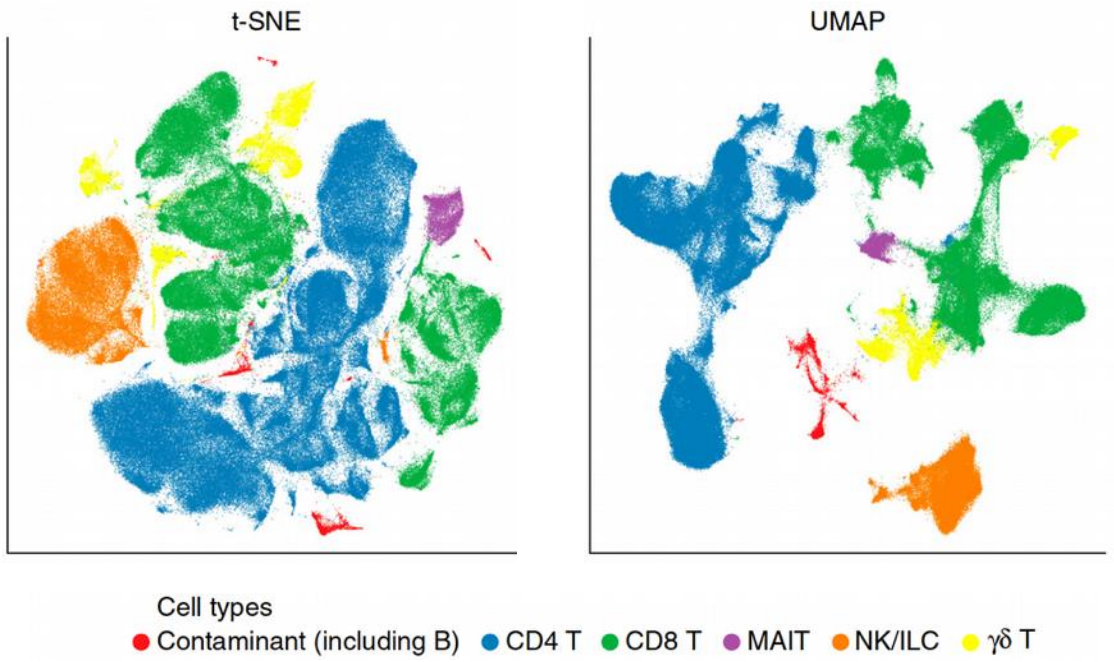
Clustering



https://mubaris.com/posts/kmeans-clustering/

- Clustering is the technique of grouping objects based on certain features.
- They are unsupervised, and are based on the characteristics of the samples and not on any other previous knowledge of the samples.
- The clusters for single cell analysis are based on gene expression patterns, and the clusters **may or may not** represent cell clusters of biological significance.

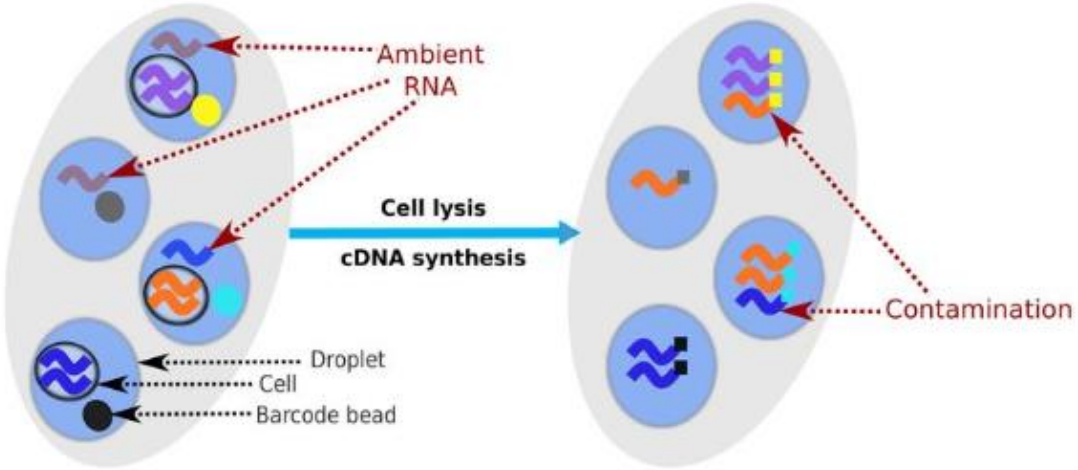
tSNE vs UMAP



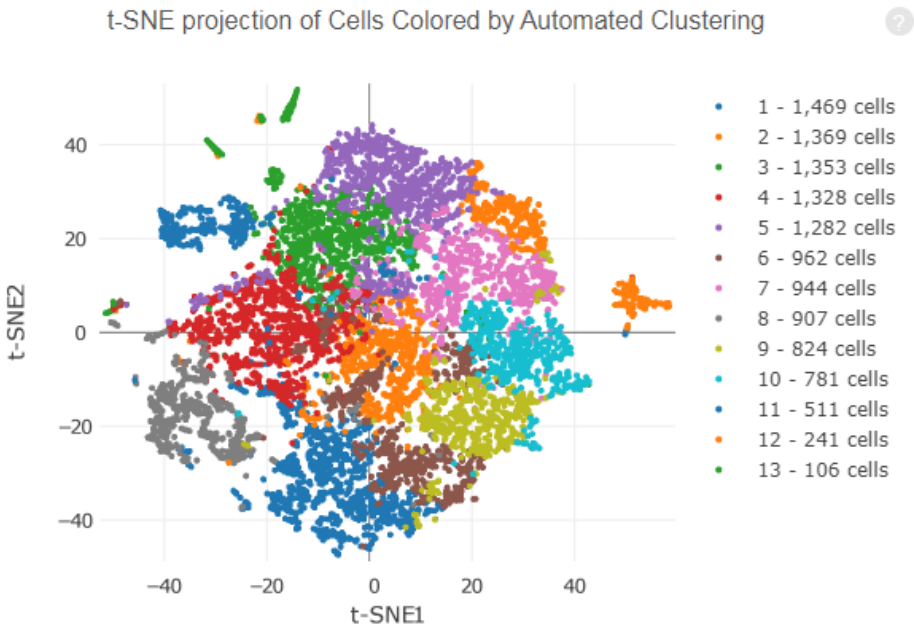
https://www.nature.com/articles/nbt.4314

- In Seurat, we first do PCA reduction, followed by clustering of the data and then another reduction method t-Distributed Stochastic Neighbor Embedding (**t-SNE**), which is the general method of visual representation of the single cell RNAseq data.
- t-SNE, reduces the larger dimension datasets into 2 dimensions. The major goal of t-SNE plot is to put the cells with similar property near each other. It is just a visualization method.
- UMAP is a new popular dimension reduction method, faster too
- Distances between clusters in a tSNE plot does not mean anything, UMAP does!

Contamination: What is it?

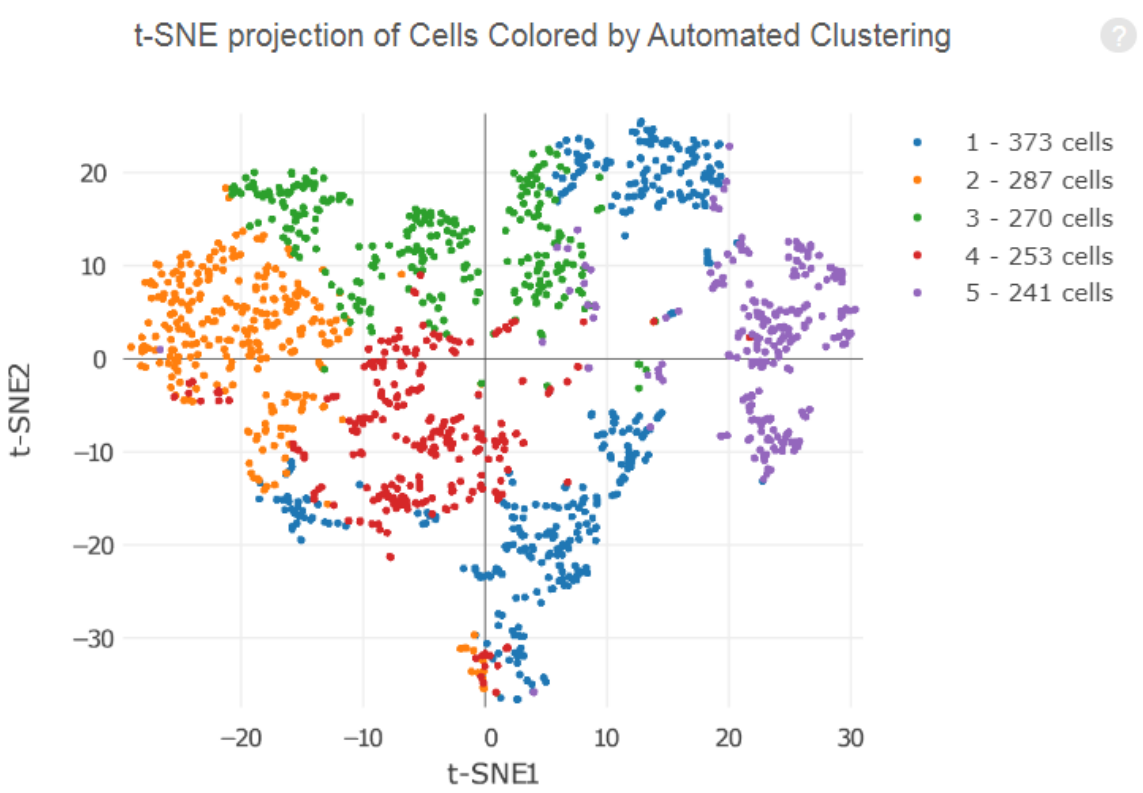


https://www.biorxiv.org/content/10.1101/704015v1.full

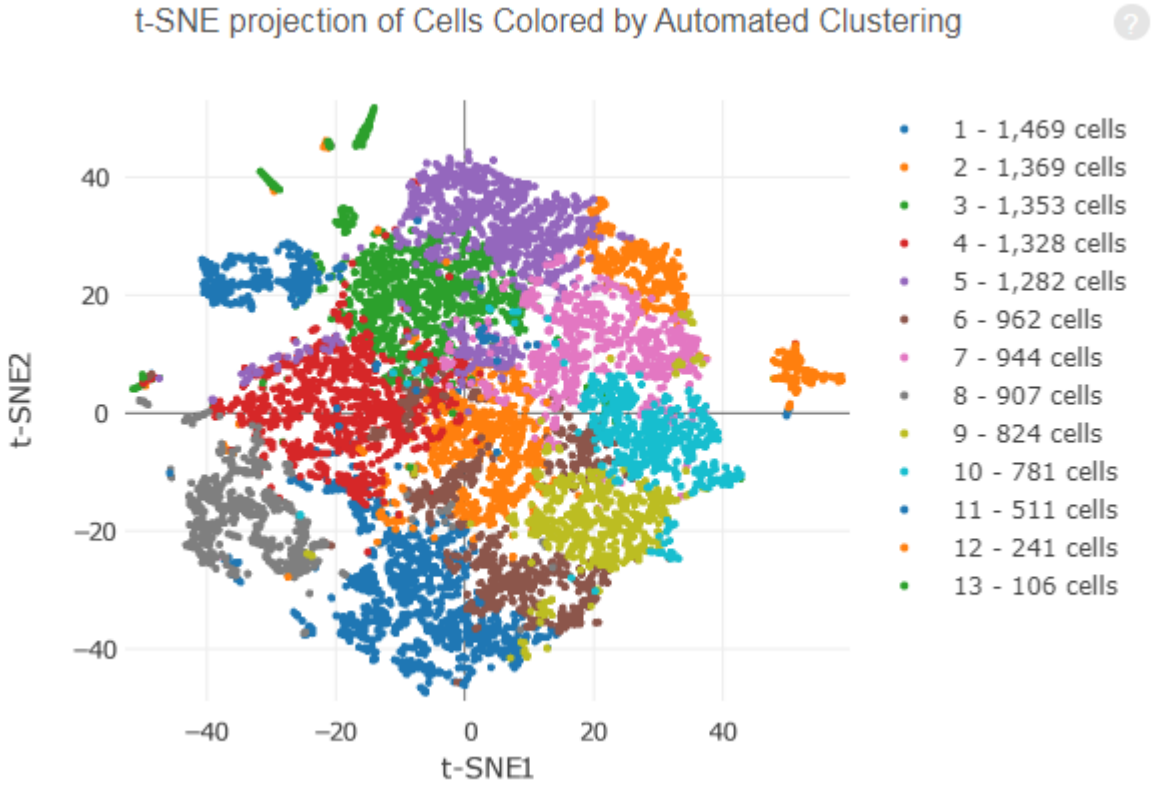


- Single cell RNA-seq experiences contamination due to “ambient RNA”.
- Ambient RNA is a pool of mRNA molecules that have been released in the cell suspension, likely from cells that are stressed or are dead (apoptosis).
- Cross-contamination occurs when the ambient RNA gets incorporated into droplets and is barcoded and amplified along with a cell’s mRNA.
- This leads to a badly clustered t-sne plot.

Clean vs Contaminated sample



Clean Sample



Contaminated Sample

Summary of the Samples: Macrophage

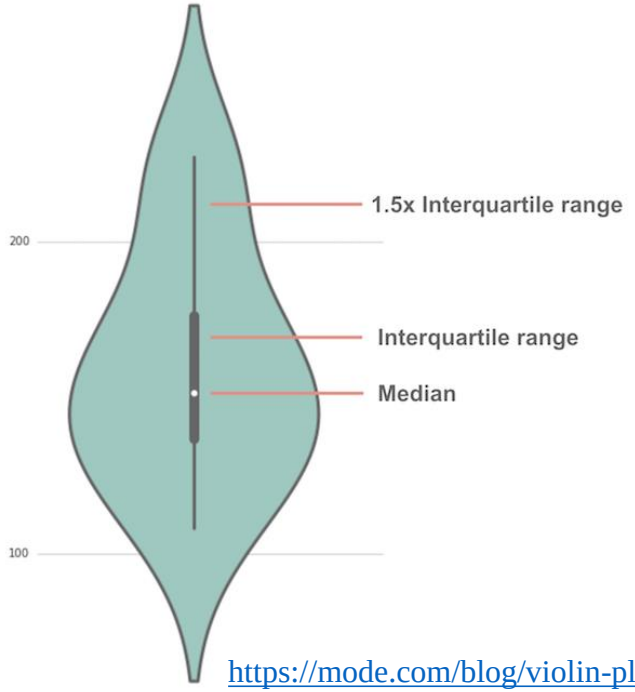
Sample ID	Input Number of cells	Estimated # of cells	Reads Mapped Confidently to Transcriptome	Reads mapped Antisense to gene	Fraction Reads per cell	Quality of the sample
Wild type macrophage (W1M)	8000	1424	55.7%	1.1%	81.3%	Clean
Wild type macrophage (W2M)	8000	12,077	48.2%	2%	64%	Contaminated
Diseased macrophage (B1M)	8000	1453	55.3%	1.1%	81.1%	Clean
Diseased macrophage (B2M)	8000	2054	58.8%	1%	78.8%	Clean

Summary of the Samples: FAPs

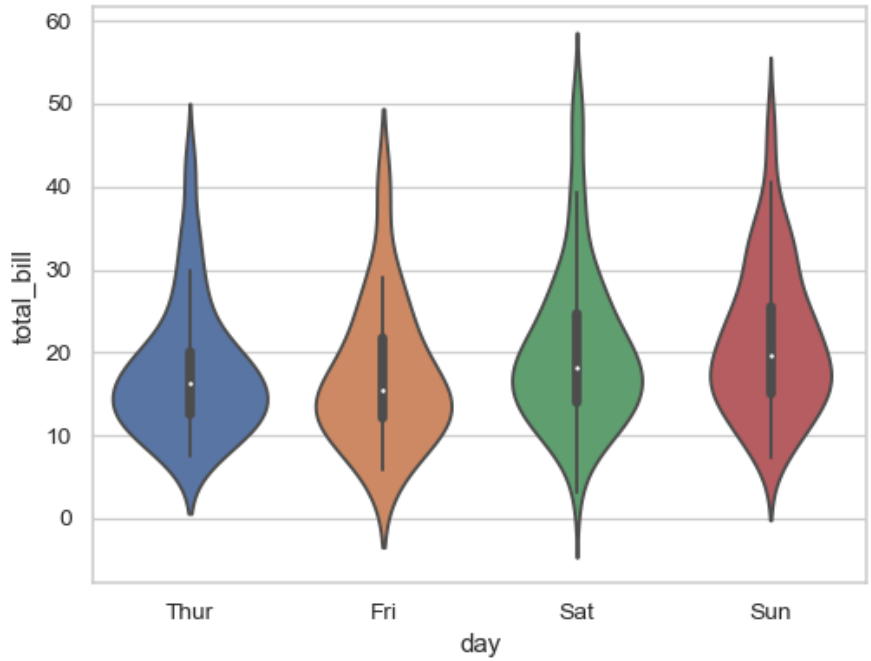
Sample ID	Input Number of cells	Estimated # of cells	Reads Mapped Confidently to Transcriptome	Reads mapped Antisense to gene	Fraction Reads per cell	Quality of the sample
Wild type FAPs (W1F)	8000	14,201	53.2%	3.3%	60.3%	Contaminated
Wild type FAPs (W2F)	8000	2073	59.4%	1.7%	85.3%	Clean
Diseased FAPs (B2F)	8000	1724	50.9%	1.4%	84.9%	Clean

* Sample B1F was not sequenced as it did not meet RNA Quality (RIN) standards.

Violin Plots: What does it represent?



https://mode.com/blog/violin-plot-examples



https://seaborn.pydata.org/generated/seaborn.violinplot.html

- Violin Plot represents the spread of the data, as well as a probability density representation of each element.
- Wider sections of the violin plot represent a higher probability that members of the population will take on the given value; the skinnier sections represent a lower probability.
- For example, maximum people in a restaurant will have a total bill between $10-$20 from Thursday- Sunday, with the probability of people having a bill of 15$ is highest on Thursday.

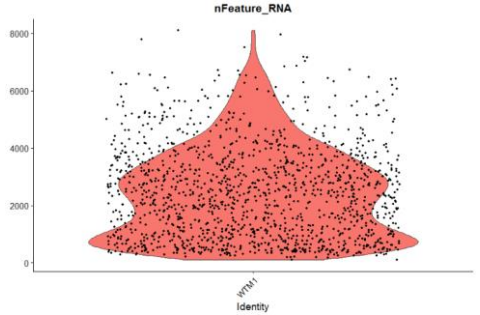
Quality Check: Seurat

Commonly used Quality Check criterions are:

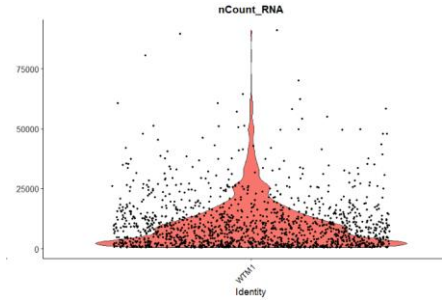
- Low quality cells or empty droplets, will have few genes/UMIs/transcripts, whereas doublets or multiplets (droplets having two or multiple cells) will have more genes. The two features that represent this criteria are

.

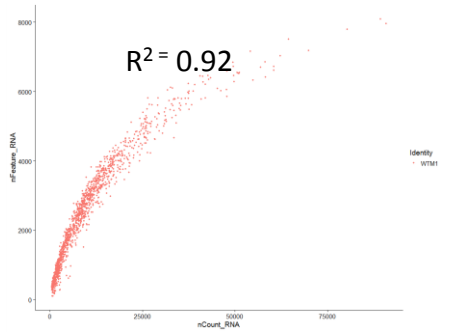
Unique genes (nFeature_RNA)



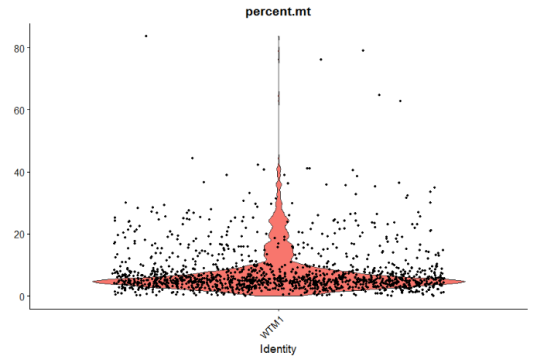
Unique UMIs (nCounts_RNA)



- Correlation between, nFeature RNA and nCount_RNA is generally high for a good sample.

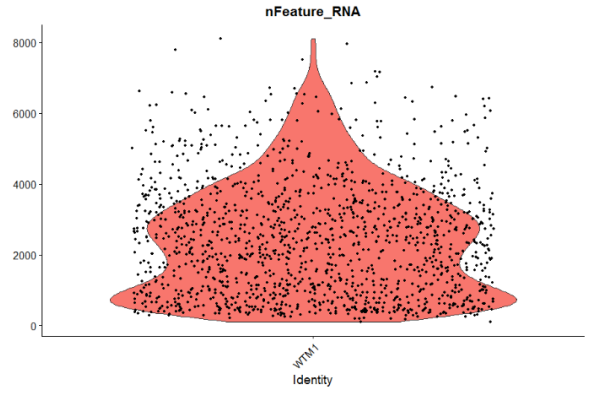


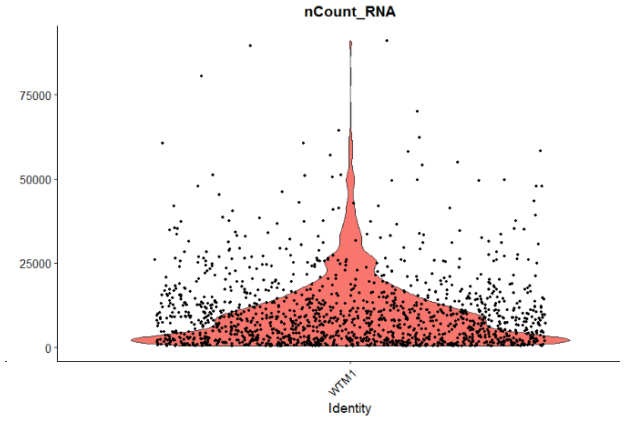
- The percentage of reads that map to the mitochondrial genome.
  Low-quality / dying cells often exhibit extensive mitochondrial contamination

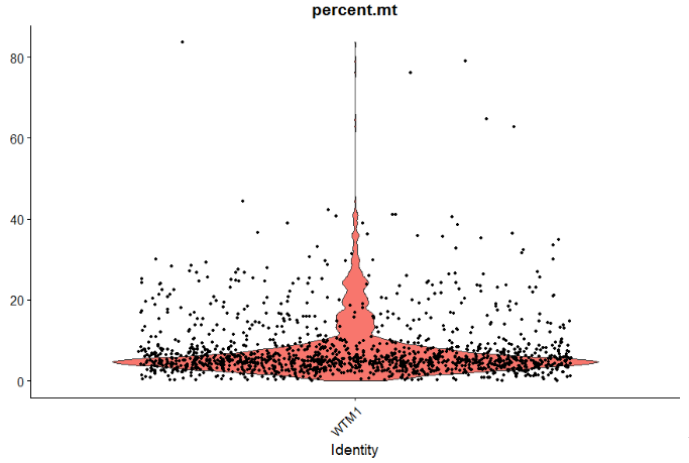


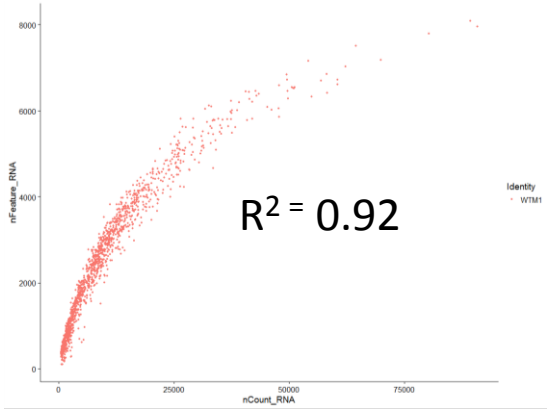
Quality Check: Wild Type Macrophage (W1M) (Before vs after Filtration)

Before Filtration

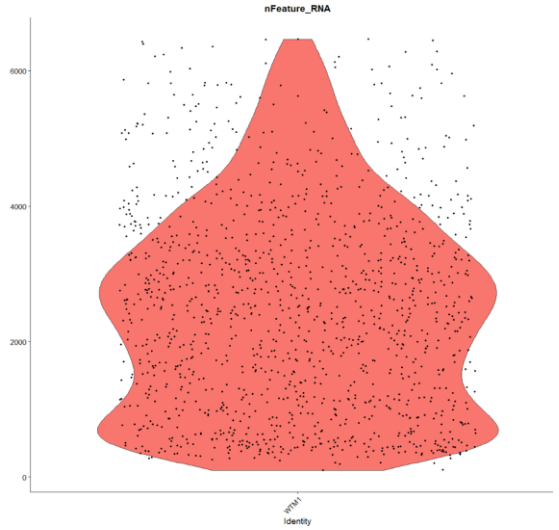


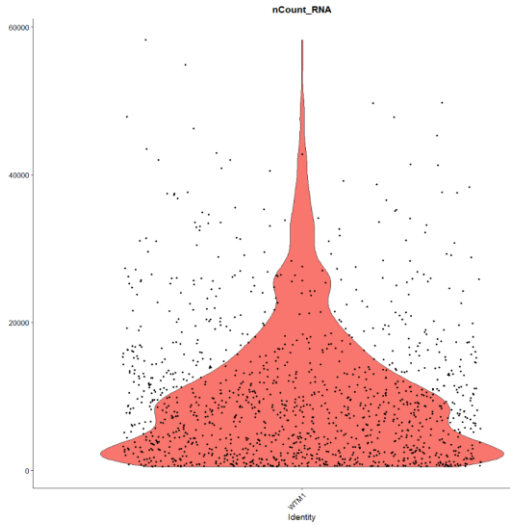


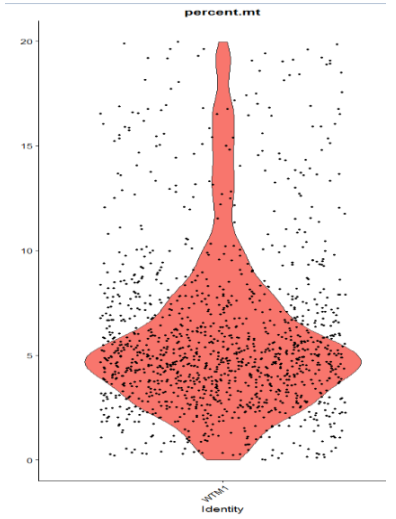


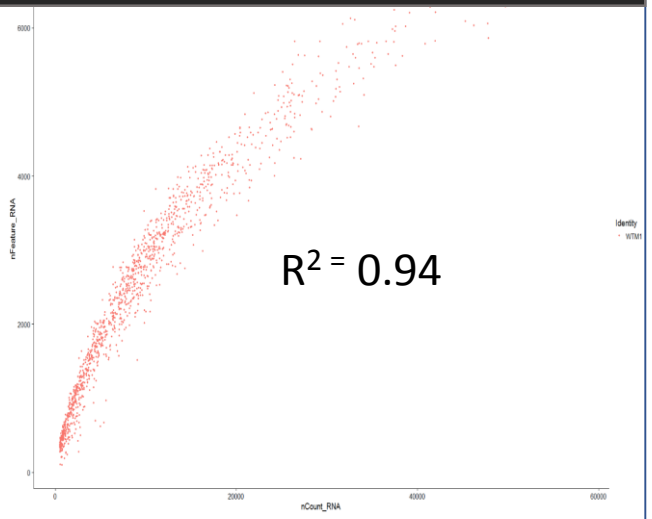


After Filtration

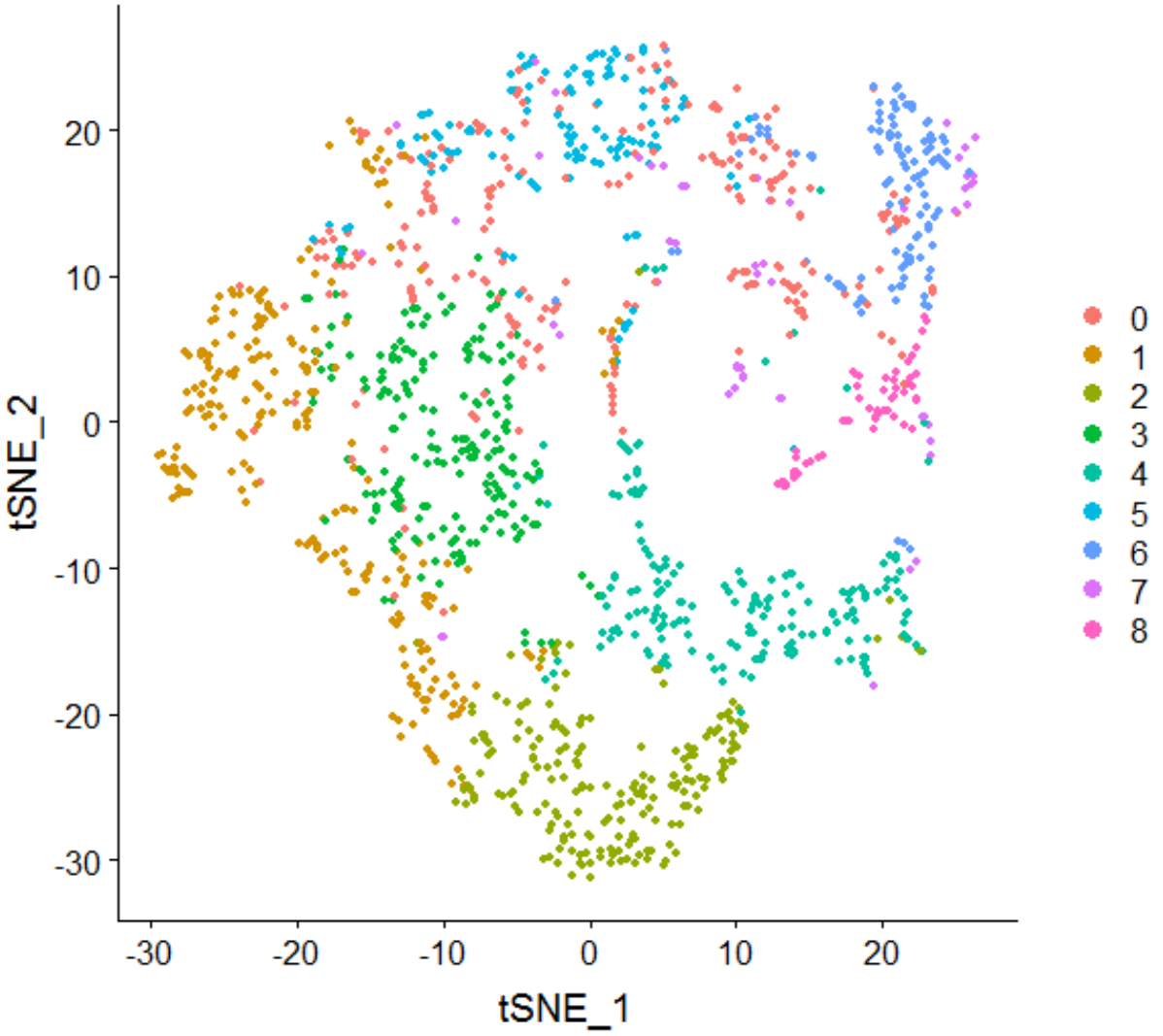


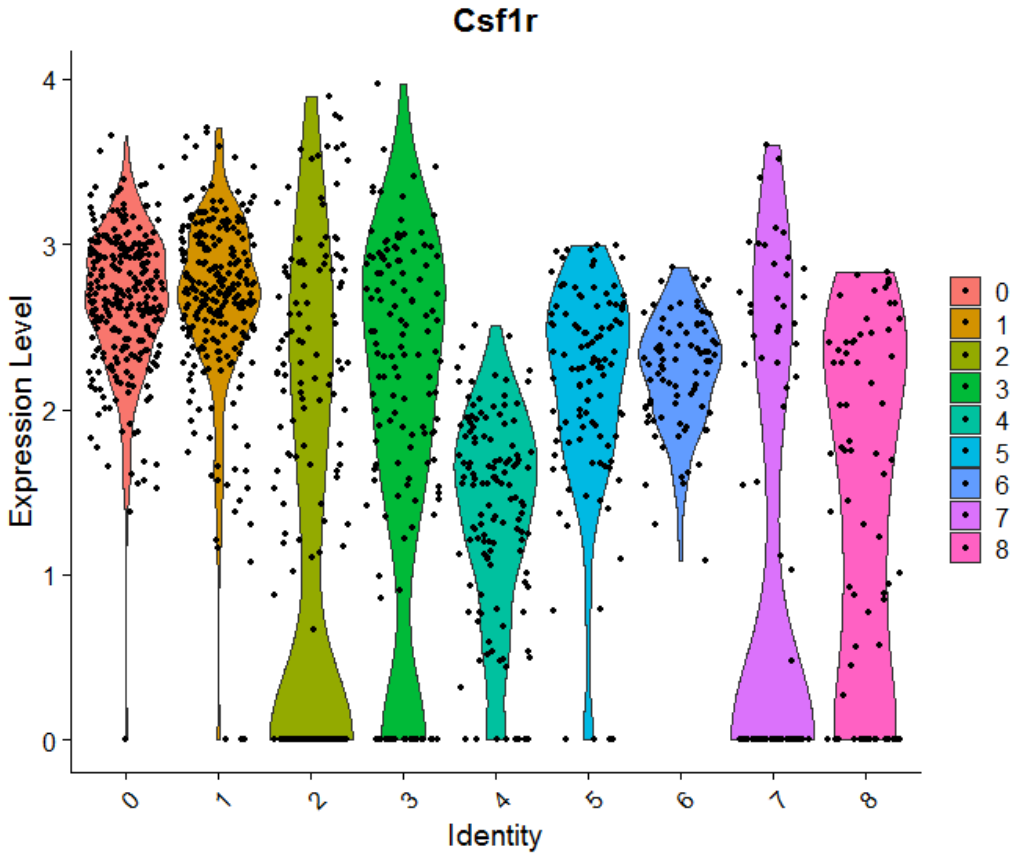






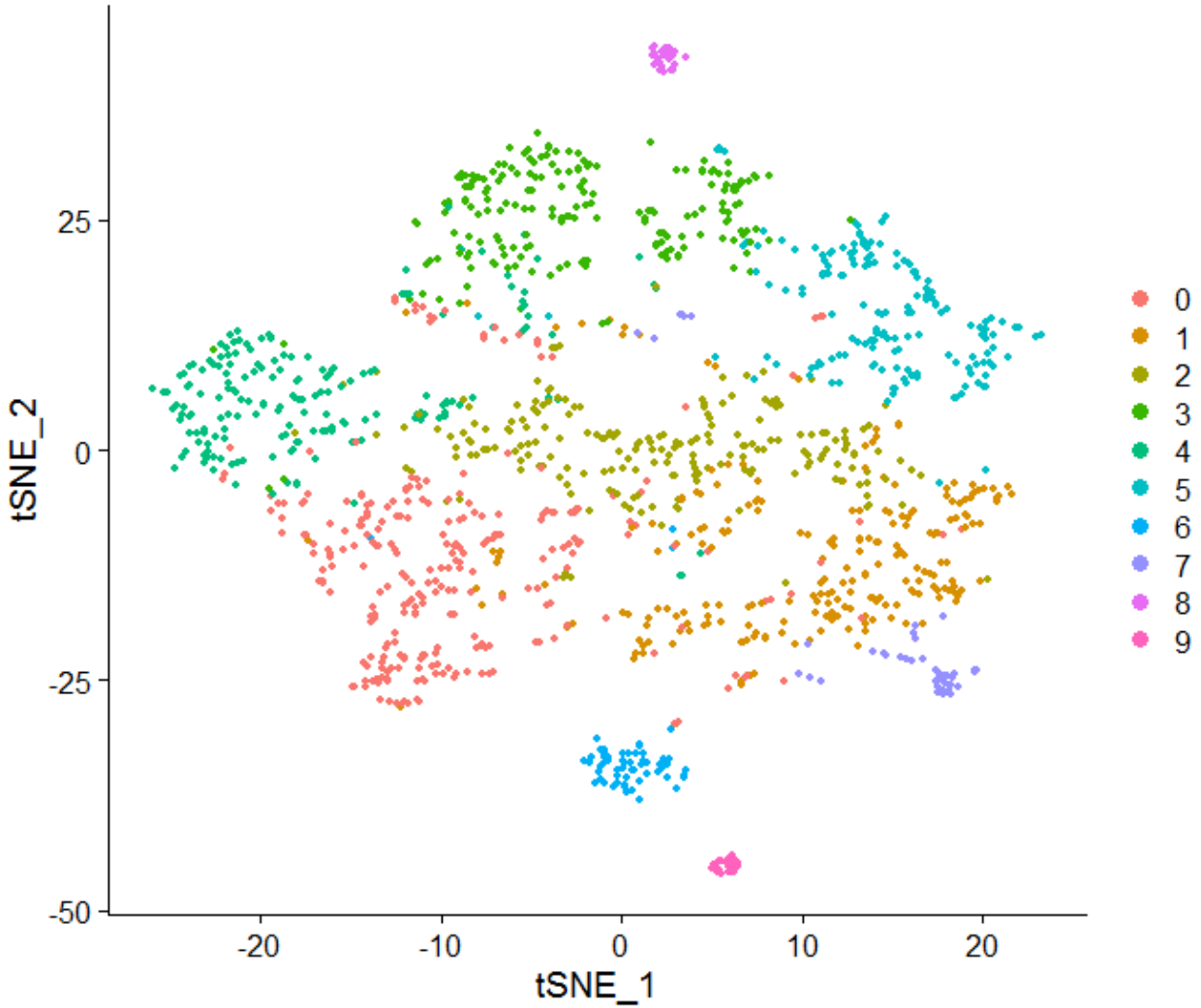
Single cell t-sne plots: W1M

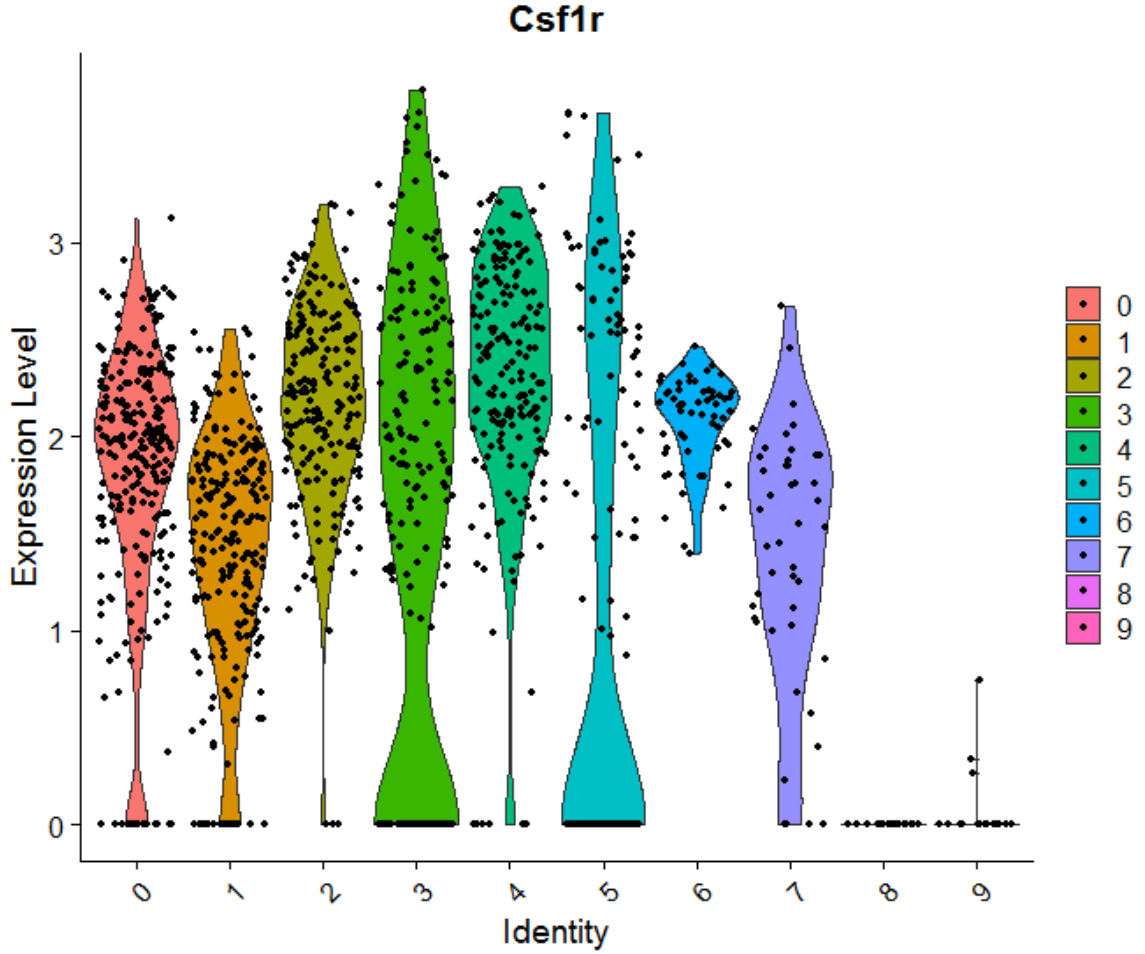




- Marker for macrophages.
- Have a quite significant expression across all clusters.

Single cell t-sne plots: B1M





- Marker for macrophages.
- Has different expression between Diseased and wild type

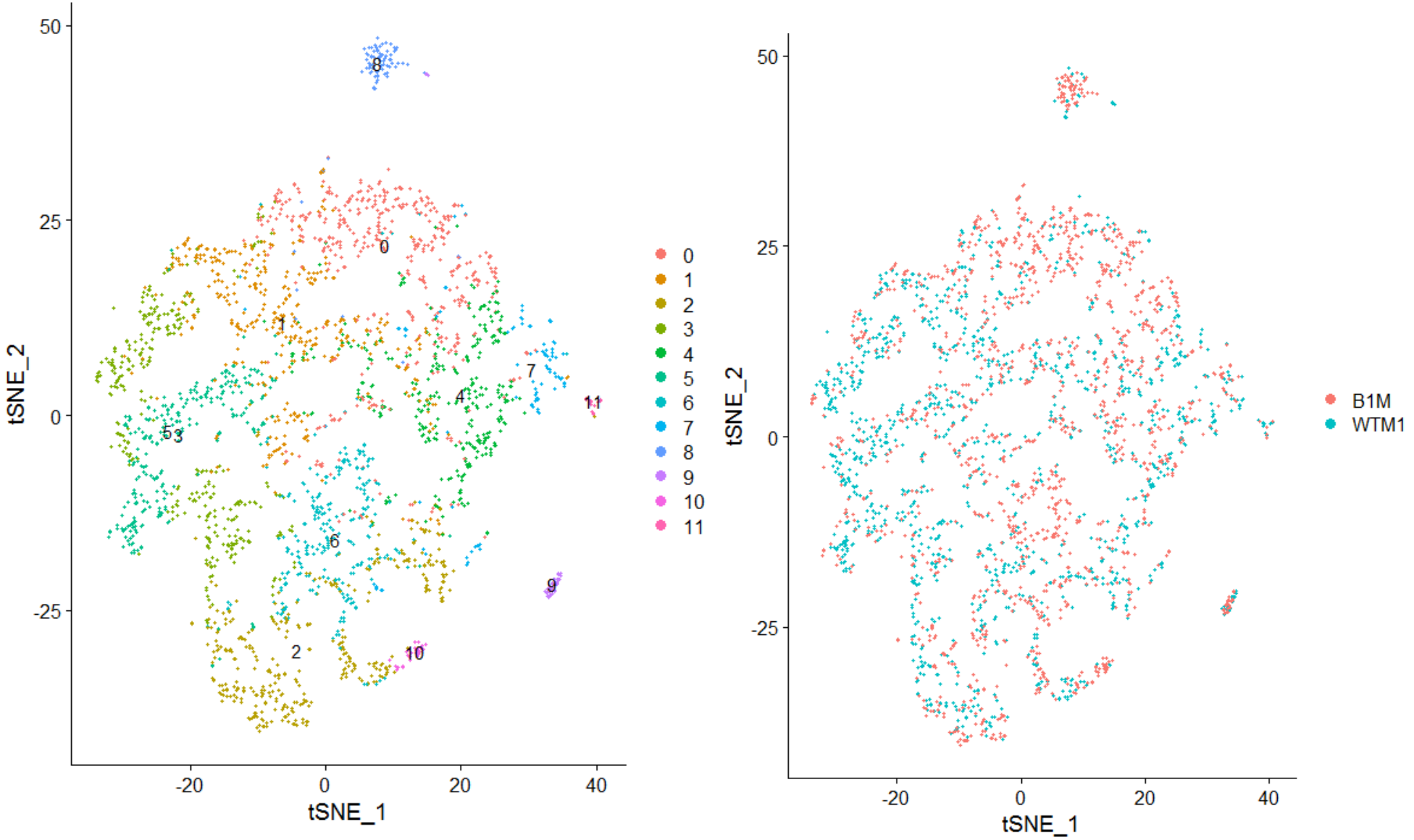
Differential expression (DE) of genes in each cluster: W1M

Top Genes By Cluster (Log2 fold-change, p-value)

		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6		Cluster 7	
Gene ID	Gene name	L2FC	p-value	L2FC	p-value	L2FC	p-value	L2FC	p-value	L2FC	p-value	L2FC	p-value	L2FC	p-value
ENSMUSG00000030787	Lyve1	2.31	2e-32	-2.56	1e-35	-0.51	1e+00	-0.13	1e+00	-4.97	1e+00	-6.35	1e+00	-2.91	1e+00
ENSMUSG00000051504	Siglech	2.31	1e-24	-2.57	1e-26	0.04	1e+00	-0.32	1e+00	-1.41	1e+00	-2.79	1e+00	-1.66	1e+00
ENSMUSG00000009185	Ccl8	2.26	1e-17	-2.18	3e-15	-0.17	1e+00	-1.50	2e-01	-2.72	1e+00	-0.77	1e+00	-0.98	1e+00
ENSMUSG00000022122	Ednrb	2.17	1e-24	-2.63	2e-31	-0.23	1e+00	0.31	8e-01	-2.81	1e+00	-4.18	1e+00	-2.06	1e+00
ENSMUSG00000026938	Fcna	2.14	1e-24	-2.24	2e-24	0.08	1e+00	-0.65	7e-01	-2.22	1e+00	-4.18	1e+00	-3.06	1e+00
ENSMUSG00000051906	Cd209f	2.14	1e-25	-2.01	3e-21	-0.38	1e+00	-1.31	1e-01	-4.69	1e+00	-5.07	1e+00	-3.95	1e+00
ENSMUSG00000055546	Timd4	2.13	1e-22	-2.26	9e-23	0.19	1e+00	-0.57	8e-01	-2.34	1e+00	-2.72	1e+00	-1.59	1e+00
ENSMUSG00000031495	Cd209d	2.10	3e-25	-2.06	1e-22	-0.44	1e+00	-0.79	5e-01	-4.51	1e+00	-4.89	1e+00	-3.77	1e+00
ENSMUSG00000032725	Folr2	2.08	2e-26	-1.98	2e-22	-0.50	1e+00	-1.01	3e-01	-4.05	1e+00	-5.43	1e+00	-4.31	1e+00
ENSMUSG00000079168	Cd209g	2.05	2e-22	-1.99	1e-19	-0.45	1e+00	-0.78	5e-01	-3.10	1e+00	-3.48	1e+00	-2.36	1e+00

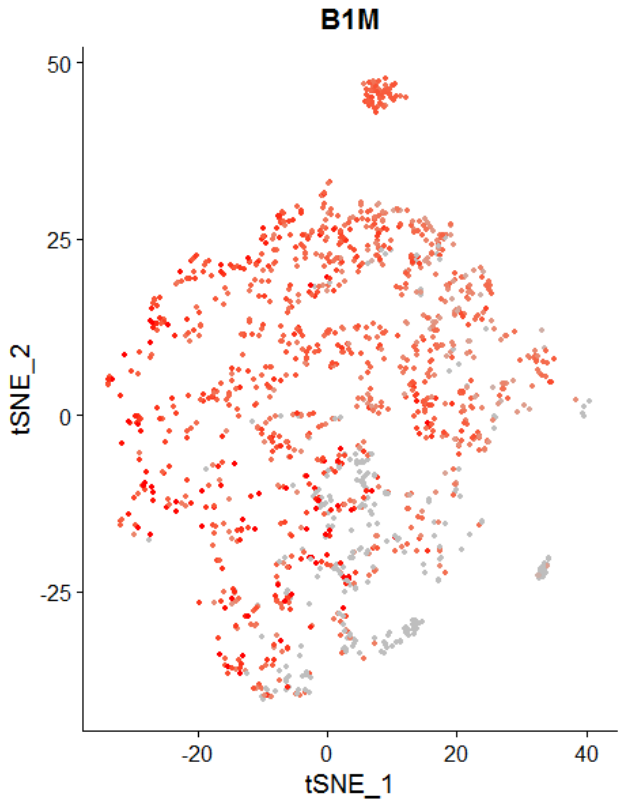
- Differential gene expression between genes of cells in one cluster versus all the other cells in different clusters, gives us a knowledge about the cell.
- For example , gene Lyve1 has log2FC of 2.31 with adjusted p-value of 2e-32, for cluster 1, compared to all cells present in clusters 2-7.
- Functional annotation of these DE genes gives the cell its identity.

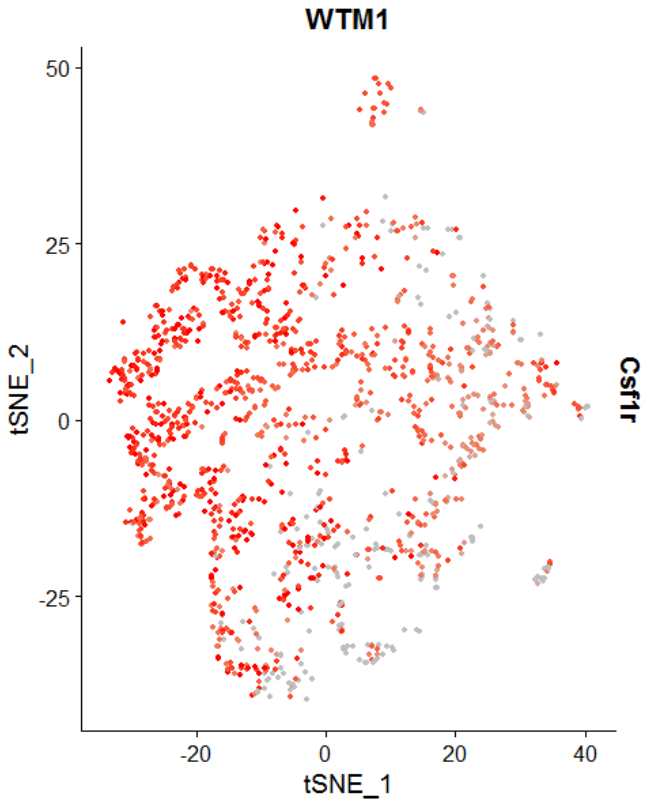
Differential Expression: B1M and W1M

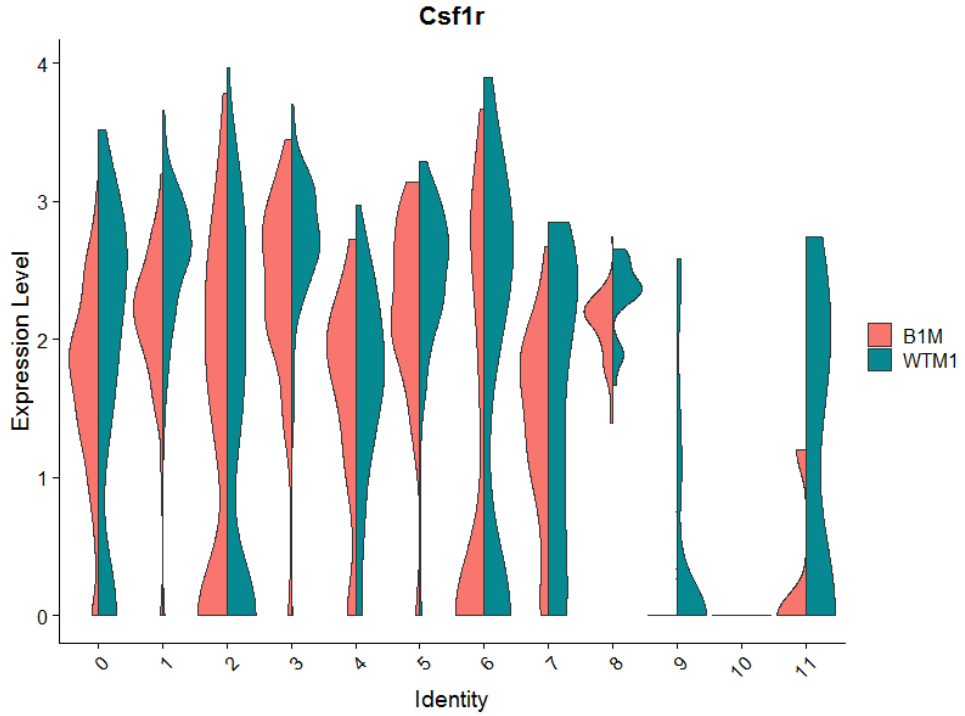


- Differential expression in Seurat is done using DEseq2.
- You can visualize the expression of the genes in each condition for individual clusters
- You can also output the results as a table.

Differential Expression Csf1r: B1M and W1M

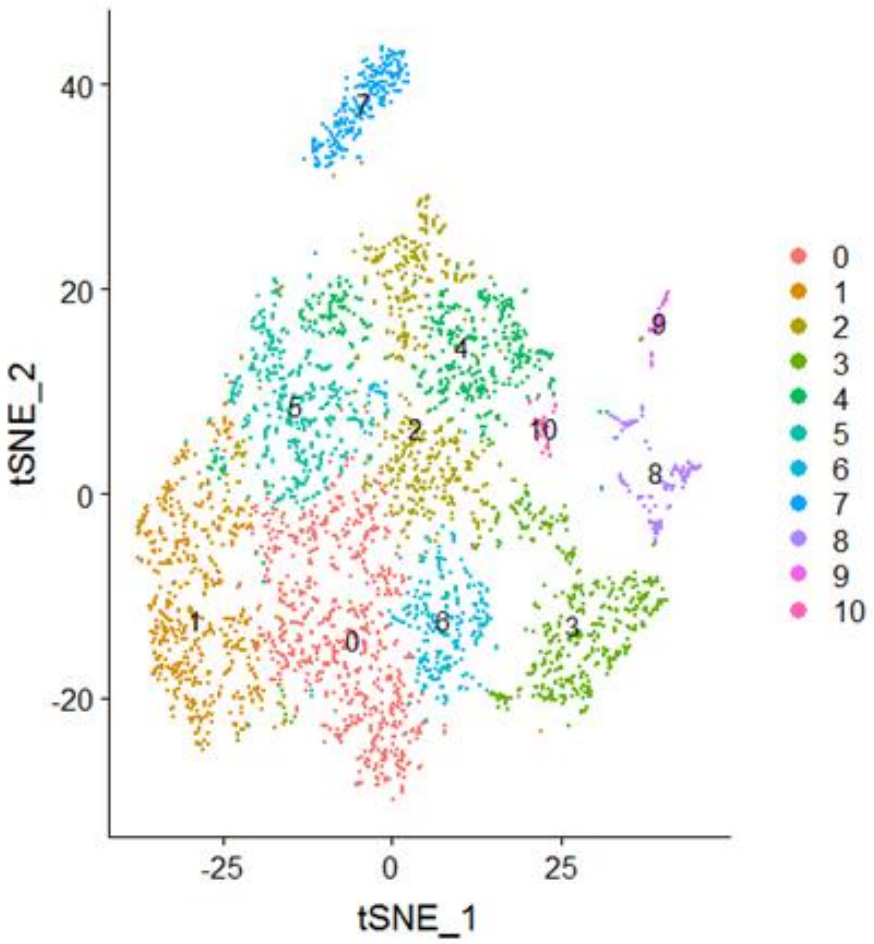


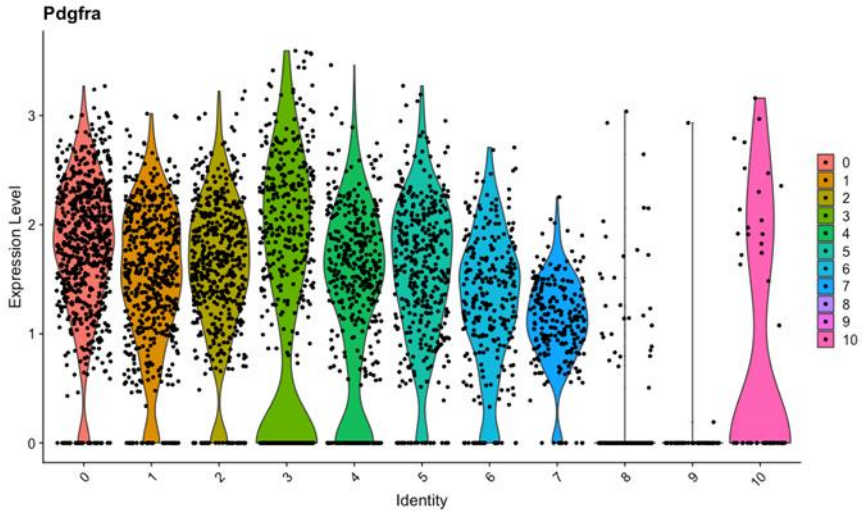


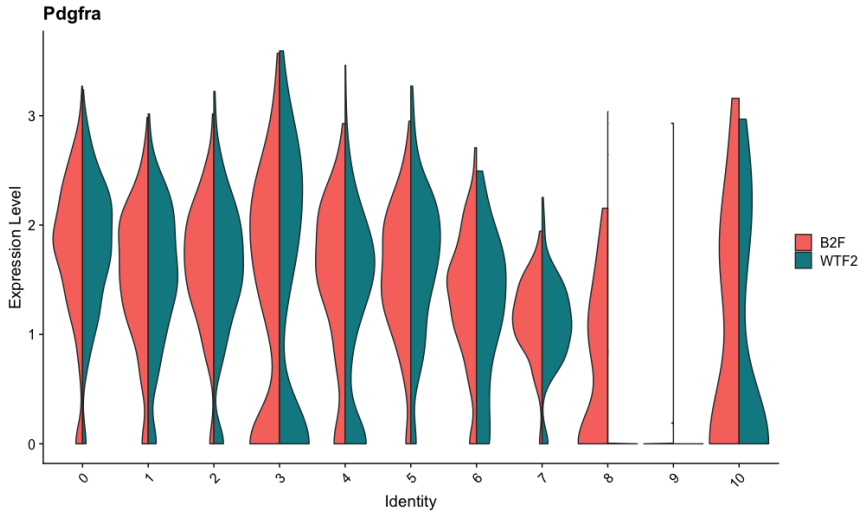


- Csfr1 gene expression remains quite similar across the clusters.
- The difference occurs in cluster 9 and 11, where it is more in Wild type than diseased condition.
- Cluster 10 has no difference for both the conditions.

FAP Marker Expression



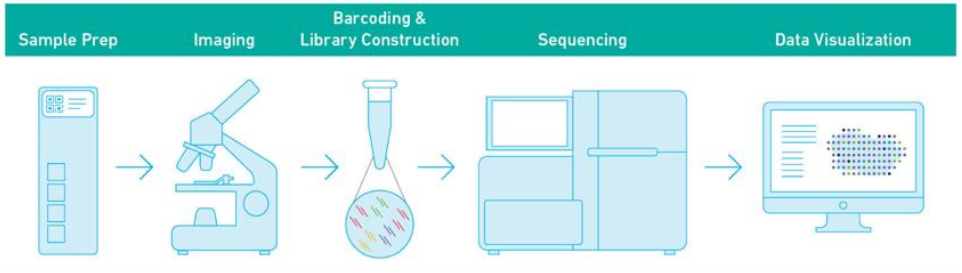


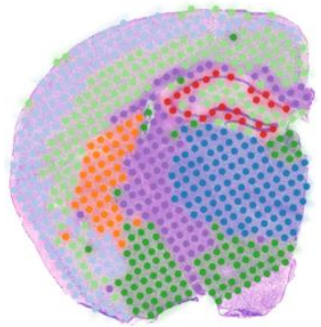


- Pdgfra is a marker for FAPs.
- Not expressed in 2 clusters in WT.
- Can be due to macrophage contaminations
- More markers are being identified for both FAPs and macrophages.

Conclusion

- Single cell RNA-sequencing method is a still evolving field, with the amount of new technologies, and analysis tools being created for the same.
- Before a single cell experiment one has to be cautious of various factors as illustrated in the workflow.
- A future innovation in single cell technique is spatial transcriptomics, where single cell rna-seq result is superimposed on tissue images.





Broad considerations	Model-specific or practical considerations	Resources and References
Characterize specific rare cell population OR de novo discovery?	How many cells? Sequencing depth? Intra-population heterogeneity?	powsimR
Study design		
What tissues? solid vs liquid? fresh vs. archived?	Heterogeneity in cell size? Heterogeneity in cell surface markers? Microanatomical location? Minimization of batch effects?	FACS, microfluidics, others (see Fig. 2)
Single-cell isolation		
well-annotated genome OR de novo genome assembly?	Single or paired end? 3′ or full length transcripts?	(see Fig. 3)
Sequencing		
Supervised OR unsupervised?	Read QC Read quantitation Preprocessing ID subpopulations and gene sets Functional annotation	command-line tools, R/Bioconductor packages
Analysis		
Further experimental validation		

https://www.frontiersin.org/articles/10.3389/fimmu.2018.01553/full

Bioinformatics pipeline: Tools

sc RNA-tools Table Tools Analysis Updates Submit FAQs

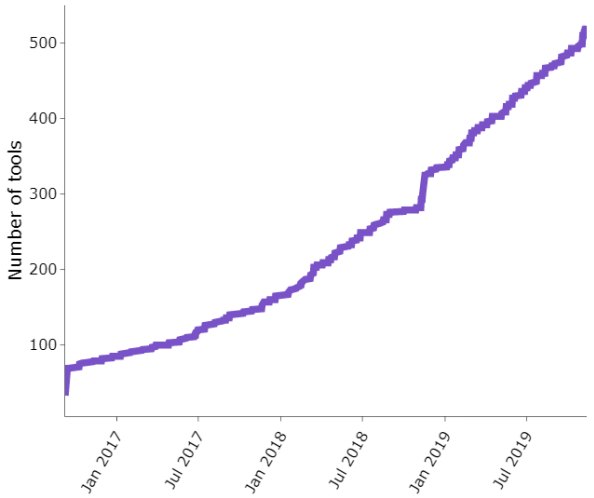
Tools Table

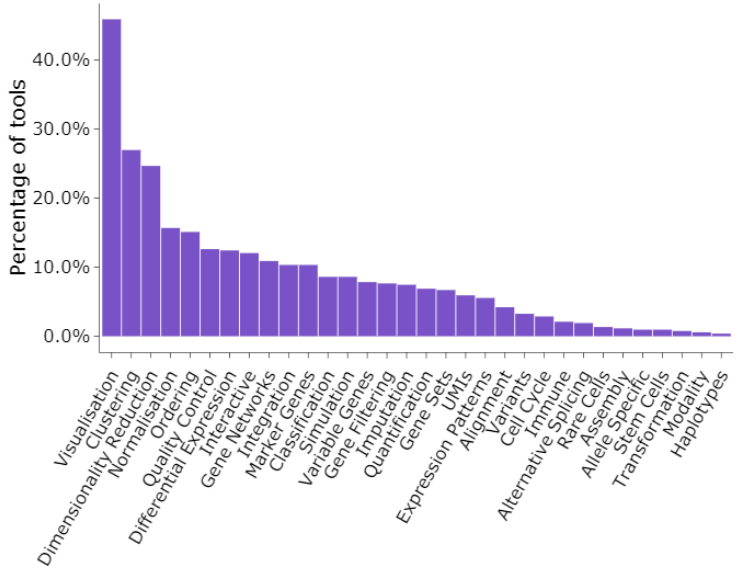
Sort, filter and download the database

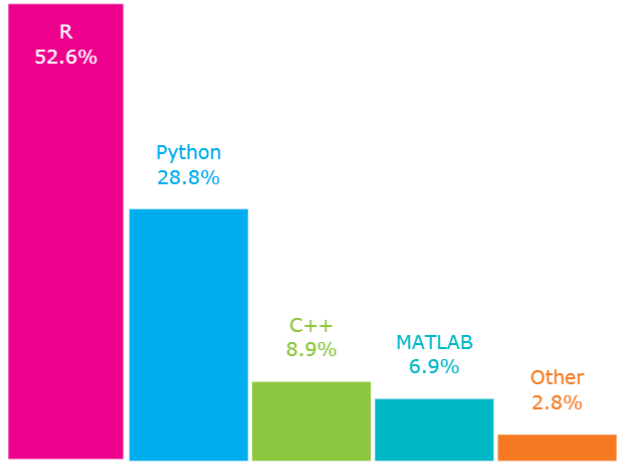
Search

NAME	PLATFORM	DOIS	CITATIONS	LICENSE	CATEGORIES
ACTINN	Python	10.1093/bioinformatics/btz592;10.1101/532093	6	-	Classification
ACTION	C++/R/MATLAB	10.1038/s41467-018-03933-2	10	-	Clustering, Dimensionality Reduction, Gene Networks
alevinQC	R	-	-	MIT	Interactive, Visualisation
ALRA	R	10.1101/397588	11	-	Imputation
AltAnalyze	Python	10.1038/nature19348;10.1101/412080	184	Apache-2.0	Alternative Splicing, Cell Cycle, Classification, Clustering, Differential Expression, Dimensionality Reduction, Gene Filtering, Gene Networks, Gene Sets, Interactive, Marker Genes, Normalisation, Quantification, Visualisation
anchor	Python	10.1016/j.molcel.2017.06.003	39	BSD-3-clause	Modality
ASAP	R/Python	10.1093/bioinformatics/btx337;10.1101/096222	33	GPL-3	Clustering, Differential Expression, Gene Filtering, Interactive, Normalisation, Visualisation
ascend	R	10.1093/gigascience/giz087;10.1101/207704	7	GPL-3	Cell Cycle, Clustering, Differential Expression, Dimensionality Reduction, Gene Filtering, Normalisation, Quality Control, Visualisation

- Single Cell RNA seq analysis tools database.
- A total of 523 tools.
- Maximum tools are currently for visualization (46%).
- Most of the tools are written in R.

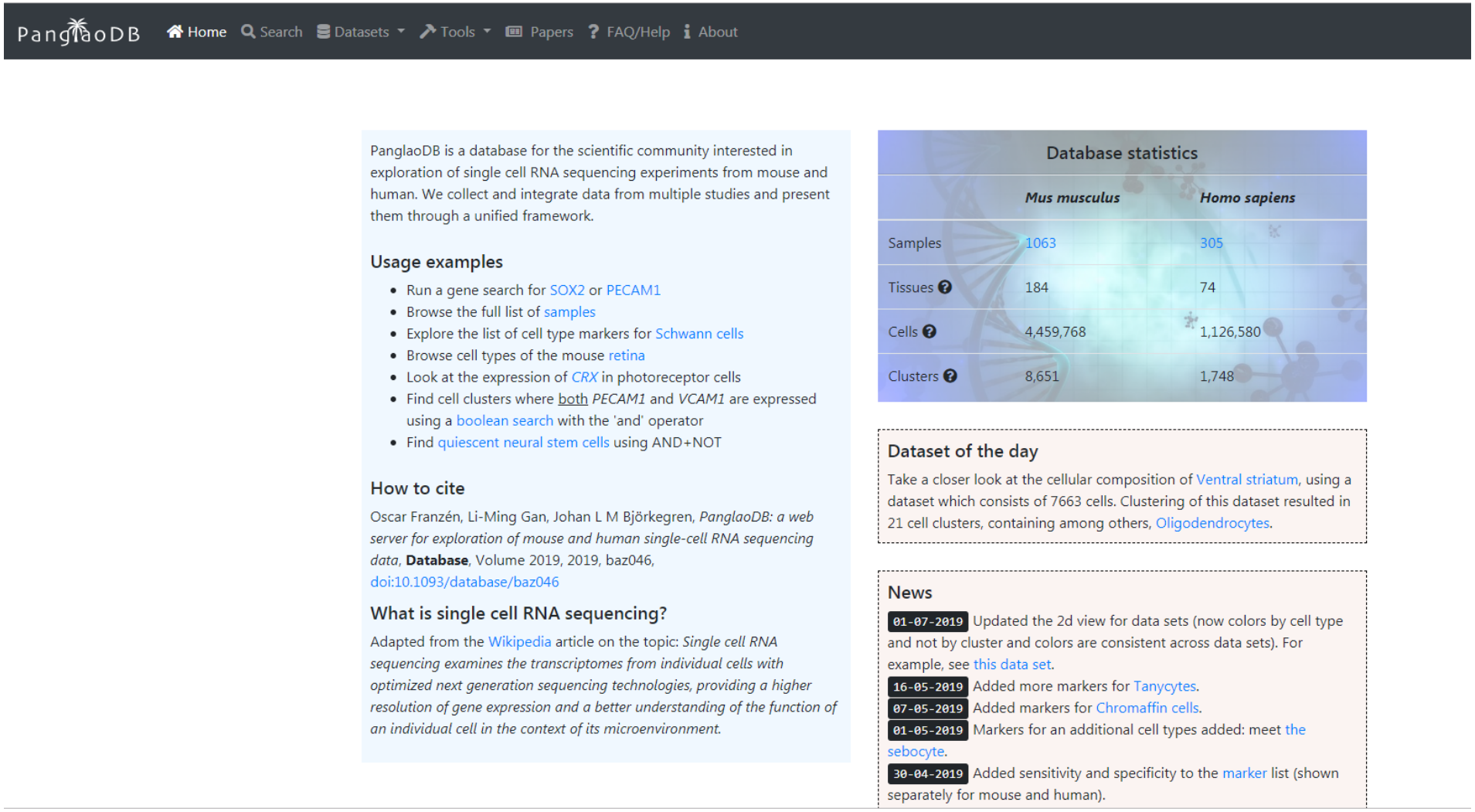






https://www.scrna-tools.org/

Single cell Databases: PanglaoDB



https://panglaodb.se/index.html

Single cell Databases: Allen Brain Institute

http://celltypes.brain-map.org/rnaseq

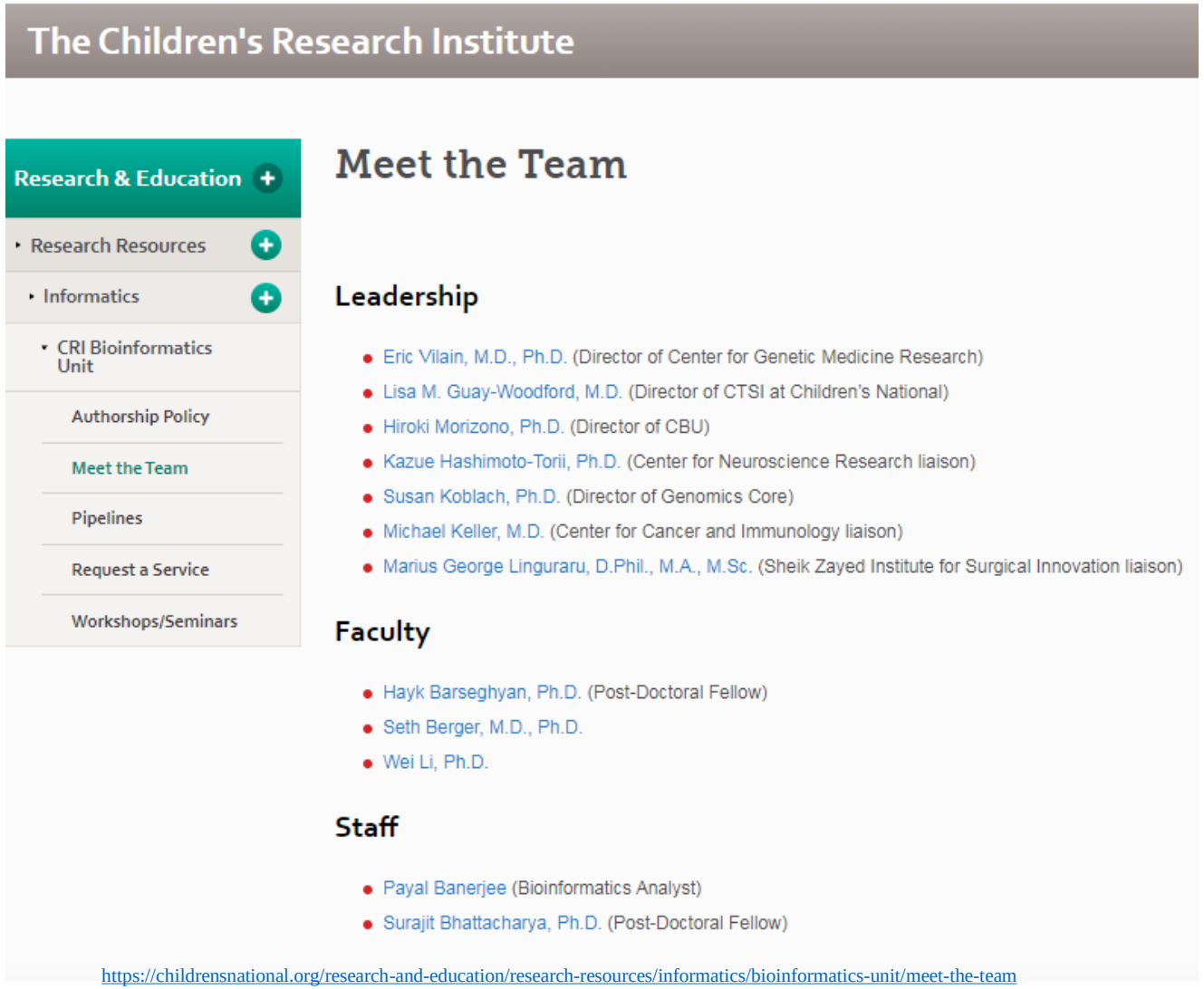
Contacts

- Design of Experiment
  - Susan Knoblach - SKnoblach@childrensnational.org
  - CRI Bioinformatics Unit - https://cri-datacap.org/surveys/?s=3EJP7L8PLK
- Cell Isolation and sample preparation
  - Karuna Panchapakesan - KPanchapakesan@childrensnational.org
- Using 10x system
  - Karuna Panchapakesan - KPanchapakesan@childrensnational.org
- Sequencing
  - Susan Knoblach - SKnoblach@childrensnational.org
- Bioinformatics
  - Payal Banerjee - pbanerjee@childrensnational.org
  - Surajit Bhattacharya - sbhattach2@childrensnational.org

References

- Hwang, B., Lee, J.H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* **50,** 96 (2018) doi:10.1038/s12276-018-0071-8
- Stuart, T., Satija, R. Integrative single-cell analysis. *Nat Rev Genet* **20,** 257–272 (2019) doi:10.1038/s41576-019-0093-7
- Luecken, MD., Theis, FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* **15**, 1744-4292 (2019). doi: https://doi.org/10.15252/msb.20188746
- Vieth, B., Parekh, S., Ziegenhain, C. *et al.* A systematic evaluation of single cell RNA-seq analysis pipelines. *Nat Commun* **10,** 4667 (2019) doi:10.1038/s41467-019-12266-7
- http://www.cellatlassearch.com/
- https://satijalab.org/seurat/ - Seurat
- http://cole-trapnell-lab.github.io/monocle-release/ - Monocle
- https://www.cell.com/molecular-therapy-family/methods/pdf/S2329-0501(18)30066-4.pdf
- https://singlecell.broadinstitute.org/single_cell : Single cell databases
- https://bioturing.com/bbrowser : Single cell databases
- https://www.biorxiv.org/content/10.1101/704015v1.full: decontX

Acknowledgement



https://childrensnational.org/research-and-education/research-resources/informatics/bioinformatics-unit/meet-the-team

- Jyoti Jaiswal
- Marshall Hogarth
- **Brian Uapinyoying**

Questions?