



EST. 1960 AS THE
TERATOLOGY SOCIETY

Technical Considerations for Single-Cell RNAseq Workflows

David Gallegos, PhD

R-Authority Postdoctoral
Researcher

United States Environmental
Protection Agency

Society for Birth Defects Research and Prevention's **2021 Virtual 61st Annual Meeting**

DISCLOSURES

Presenter has no conflicts of interest

This presentation does not necessarily reflect EPA policy. Mention of trade names is not an endorsement or recommendation for use.



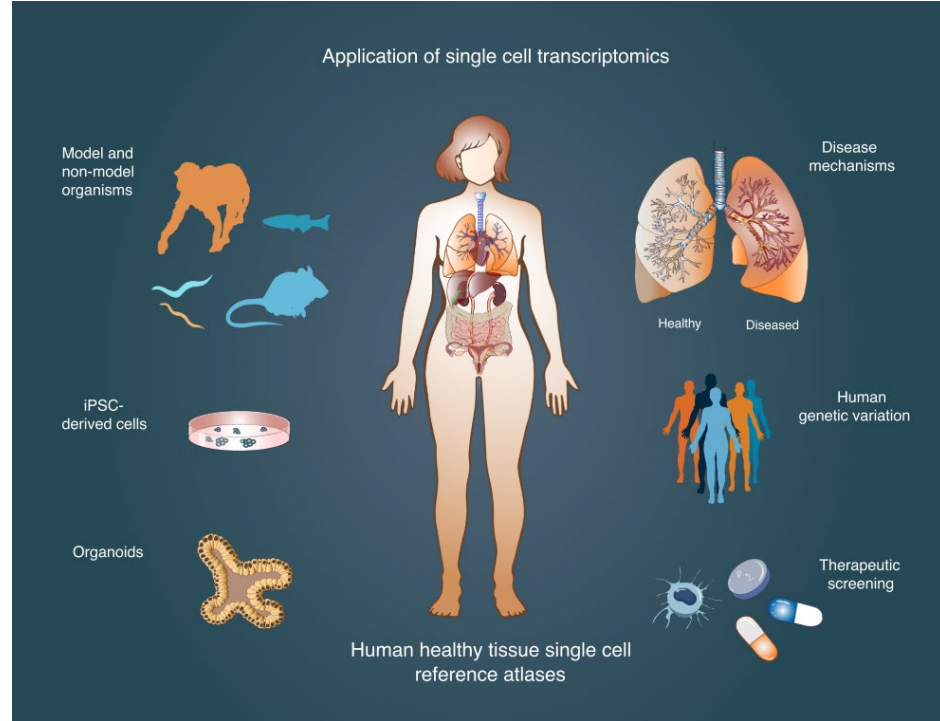
SINGLE-CELL SEQUENCING CONCEPTUAL UTILITY



Input Substrate



Amplification/Analysis Substrate



PRINCIPLES AND STEPS OF SINGLE-CELL RNA-SEQ WORKFLOW

I. Tissue Procurement



Source:

- Primary human
- Model organism
- Cell culture

Key considerations:

- Biological variation
- Sampling/handling variation
- Duration of sourcing

Study design:

- Biological replicates
- Technical replicates
- Cell number calculation
- Workflow optimization

II. Tissue Dissociation



Method:

- Mechanical mincing
- Enzymatic digestion
- Automated blending
- Microfluidics devices

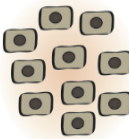
Key considerations:

- Experimental consistency
- Shortest duration
- Highest cell/nucleus quality
- Representation of all cell types

Quality control:

- FACS analysis
- qPCR for marker genes
- Imaging of cell integrity
- RNA quality (RIN)

III. Cell Enrichment (optional)



Method:

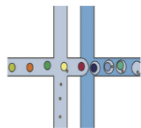
- Differential centrifugation, sedimentation, filtration
- Antibody labeling for positive/negative selection
- Flow cytometry or bead-based enrichment
- Dead cell removal

Key considerations:

- Additional handling
- Longer duration
- Loss of RNA quality
- Transcriptome changes

PRINCIPLES AND STEPS OF SINGLE-CELL RNA-SEQ WORKFLOW

IV. Single Cell RNAseq Platform



Method:

- Droplet-based
- Tube-based after FACS
- Microwell-based
- Microfluidics-enabled

Key considerations:

- Cell throughput and handling time
- Gene coverage and cell type detection
- Whole transcript versus 3'end counting
- Imaging capability for doublet detection

V. Library Sequencing



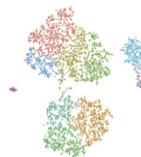
Method:

- Illumina NGS
- Compatible with cDNA library

Sequencing depth considerations:

- 3'end counting: low depth ~50K RPC
- Whole transcript: high depth ~1M RPC
- Alternative splicing: ~20-30M RPC
- Iterative optimization for biological system

VI. Computational Analysis



Key considerations:

- Separation of *batch* and *condition*
- Technical vs. biological variation

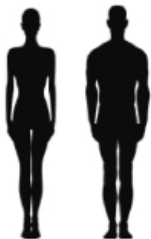
Sample Batch correction approaches:

- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST



TISSUE/SAMPLE PROCUREMENT

I. Tissue Procurement



Source:

- Primary human
- Model organism
- Cell culture

Key considerations:

- Biological variation
- Sampling/handling variation
- Duration of sourcing

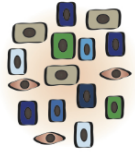
Study design:

- Biological replicates
- Technical replicates
- Cell number calculation
- Workflow optimization



SINGLE CELLS VS. SINGLE NUCLEI

II. Tissue Dissociation



Method:

- Mechanical mincing
- Enzymatic digestion
- Automated blending
- Microfluidics devices

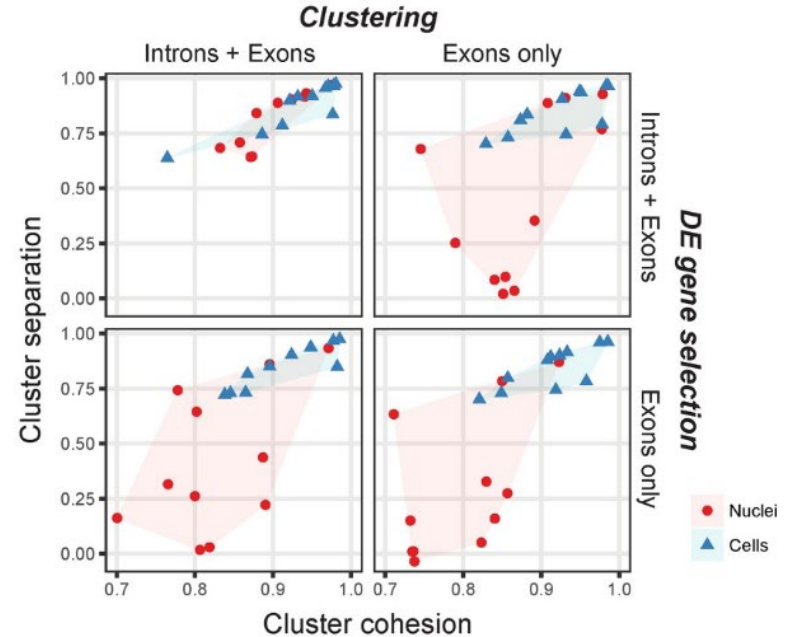
Key considerations:

- Experimental consistency
- Shortest duration
- Highest cell/nucleus quality
- Representation of all cell types

Quality control:

- FACS analysis
- qPCR for marker genes
- Imaging of cell integrity
- RNA quality (RIN)

- Cell State and Harvest Limitations
- Frozen/Post-Mortem Tissue
- Total Detectable Transcripts
- Intronic Sequences
- Nuclear Sequence Proportion



Bakken et al, *PLOS One*, 2018



CELL AND NUCLEUS DISSOCIATION

II. Tissue Dissociation



Method:

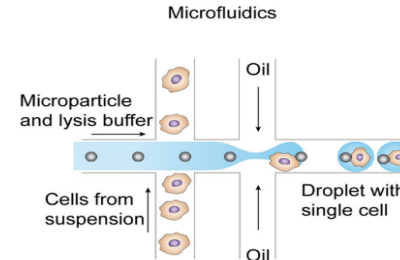
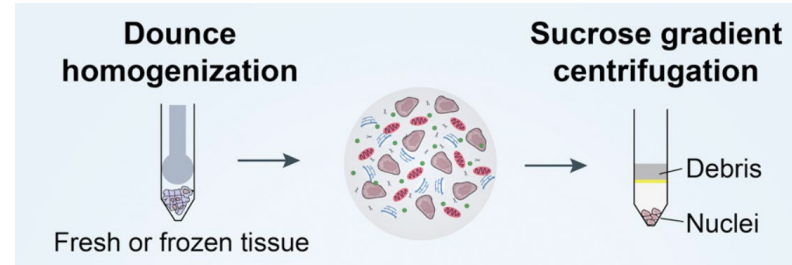
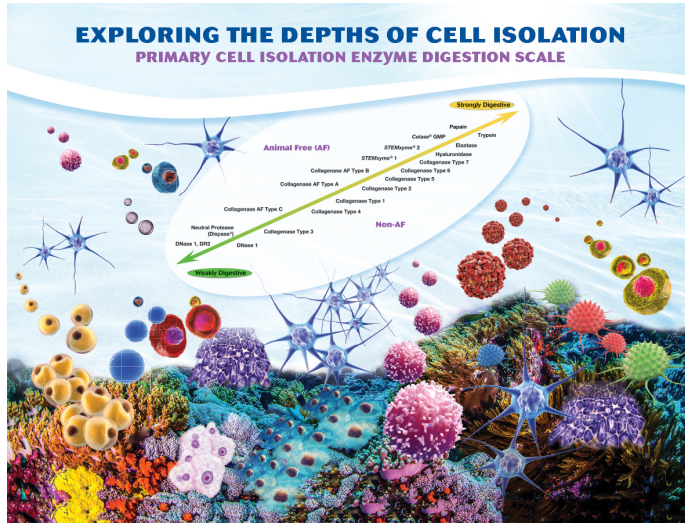
- Mechanical mincing
- Enzymatic digestion
- Automated blending
- Microfluidics devices

Key considerations:

- Experimental consistency
- Shortest duration
- Highest cell/nucleus quality
- Representation of all cell types

Quality control:

- FACS analysis
- qPCR for marker genes
- Imaging of cell integrity
- RNA quality (RIN)



Hu et al, *Molecular Cell*, 2017



CELL AND NUCLEUS DISSOCIATION

II. Tissue Dissociation



Method:

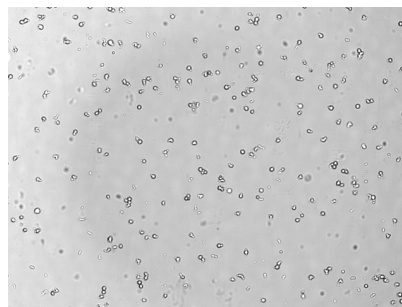
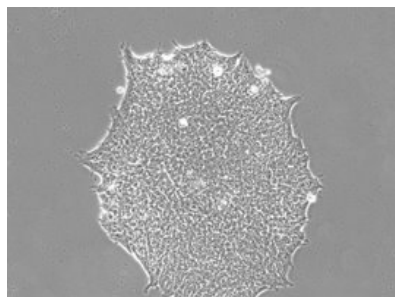
- Mechanical mincing
- Enzymatic digestion
- Automated blending
- Microfluidics devices

Key considerations:

- Experimental consistency
- Shortest duration
- Highest cell/nucleus quality
- Representation of all cell types

Quality control:

- FACS analysis
- qPCR for marker genes
- Imaging of cell integrity
- RNA quality (RIN)



Results:

Count

FL1: 1121
FL2: 43

Viability
96.4%

Concentration

3.24×10^6 cells/mL
 1.22×10^5 cells/mL

Mean Diameter

18.1 micron
11.9 micron

Suspension Buffer Options

Basic: 1XPBS + 0.04% BSA

Enhanced for Transcript Stasis: 1XPBS + 0.04% BSA
+ 100 mM D-APS, 5 mg ml⁻¹ of actinomycin D, 20 mM Triptolide,
10 mg ml⁻¹ of anisomycin

Enhanced for Cell Suspension and Separation: 1X PBS + 0.6% BSA
+ 15% Optiprep Density Gradient Medium

Other Possible Reagents and Changes: 1 nM tetrodotoxin citrate,
Kynurenic Acid, Itraconazole

Key Considerations: Viscosity
Downstream Platforms
Transcriptional/Translational Pausing
Minimizing Processing Artifacts
Cell Clumping



CELL AND NUCLEUS DISSOCIATION

II. Tissue Dissociation



Method:

- Mechanical mincing
- Enzymatic digestion
- Automated blending
- Microfluidics devices

Key considerations:

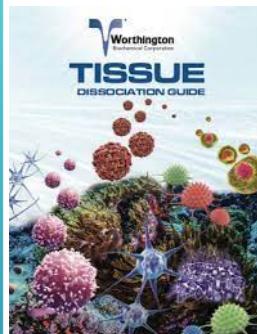
- Experimental consistency
- Shortest duration
- Highest cell/nucleus quality
- Representation of all cell types

Quality control:

- FACS analysis
- qPCR for marker genes
- Imaging of cell integrity
- RNA quality (RIN)

Tissue Tables (references, grouped by tissue type and species)

Adipose/Fat	Adrenal	Bone	Brain
Cartilage	Colon	Endothelial	Epithelial
Eye	Heart	Intestine	Kidney
Liver	Lung	Lymph nodes	Mammary
Miscellaneous	Muscle	Neural	Pancreas
Parotid	Pituitary	Prostate	Reproductive
Scales	Skin	Spleen	Stem
Thymus	Thyroid/Parathyroid	Tonsil	Tumor



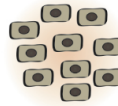
Brain					Brain
Species	Species Detail	Cell(s)	Enzyme(s)	Medium	Reference
Human	Human	Microglia	Collagenase Type 1: 300 u/ml Trypsin: 0.125%	DMEM	Mizee, M., Miedema, S., van der Poel, M., Adelia, S., van Strien, M., Melief, J., Smolders, J., Hendrickx, D., Heutink, K., Hamann, J. and Huitinga, I.: Isolation of Primary Microglia from the Human Post-Mortem Brain: Effects of Ante- and Post- Mortem Variables., <i>Acta Neuropathol</i> 5, 16, 2017 (11604)
	Human, adult	Neuronal	Papain: 20 u/ml	Neurobasal	Spaethling, J., Na, Y., Lee, J., Ulyanova, A., Baltuch, G., Bell, T., Brem, S., Chen, H., Dueck, H., Fisher, S., Garcia, M., Khaladkar, M., Kung, D., Lucas, T., O'Rourke, D. and Stefanik, D.: Primary Cell Culture of Live Neurosurgically Resected Aged Adult Human Brain Cells and Single Cell Transcriptomics., <i>Cell Rep</i> 16, 791-803, 2017 (11673)
	Human, fetal and mature	Astrocytes and neurons	Papain: 7.5-20 u/ml	RPMI	Zhang, Y., Sloan, S., Clarke, L., Caneda, C., Plaza, C., Blumenthal, P., Vogel, H., Steinberg, G., Edwards, M., Li, G., Duncan, J., Cheshier, S., Shuer, L., Chang, E., Grant, G., Gephart, M. and Barres, B.: Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse., <i>Neuron</i> 89, 37-53, 2016 (11490)



CELL ENRICHMENT AND ISOLATION

- Limiting Dilution
- Differential Centrifugation
 - Debris Removal and Cell-Size Separation
- Flow Cytometry/Fluorescent Activated Sorting (FACS)
 - Debris Removal
 - Cell-Type Targeting
 - Doublet/Multiplet Filtering
- Immunoprecipitation (IP)
 - Cell-Type Targeting
- Concern: Time
- Goals:
 - Increase Purity/Viability
 - Decrease Debris
 - Target Specific Cell/Nuclei Populations

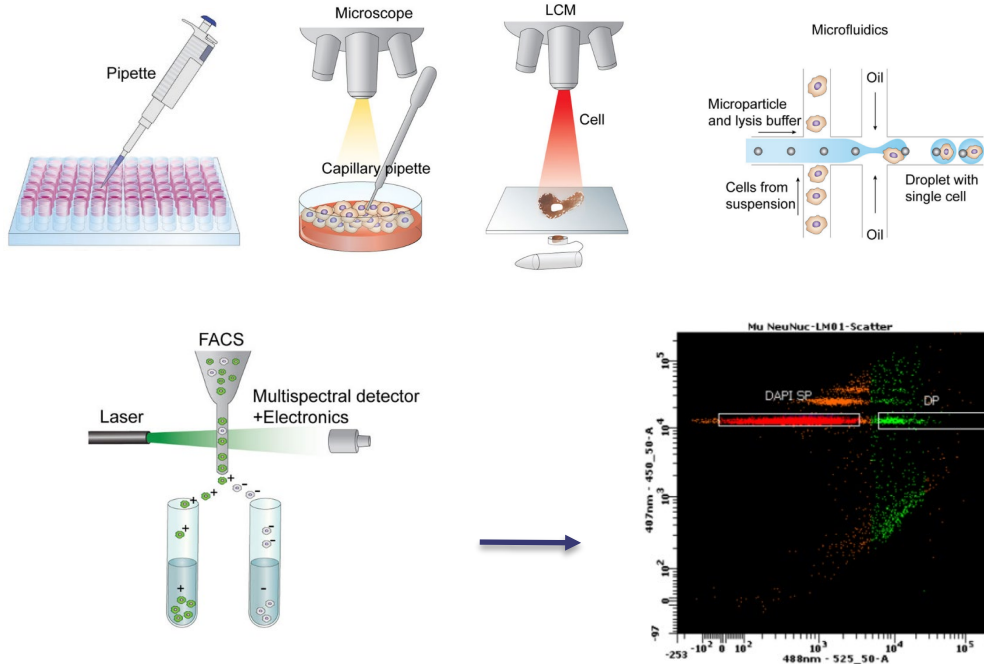
III. Cell Enrichment (optional)



Method:

- Differential centrifugation, sedimentation, filtration
- Antibody labeling for positive/negative selection
- Flow cytometry or bead-based enrichment
- Dead cell removal

- Key considerations:*
- Additional handling
 - Longer duration
 - Loss of RNA quality
 - Transcriptome changes

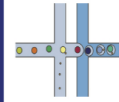


Hwang et al, *Experimental & Molecular Medicine*, 2018



SINGLE-CELL RNA-SEQ PLATFORMS

IV. Single Cell RNAseq Platform

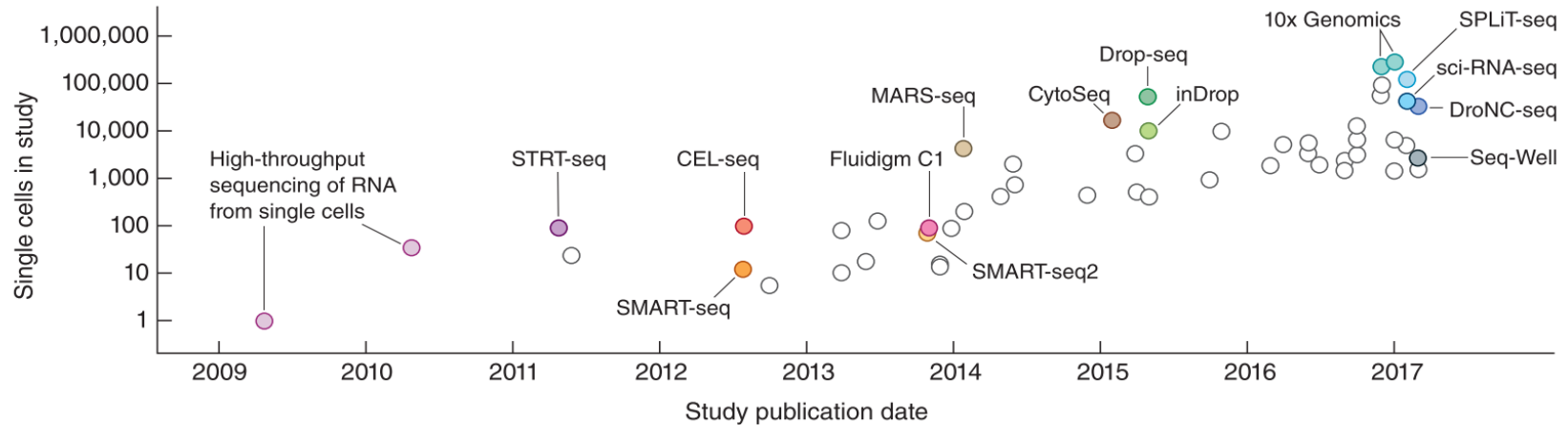
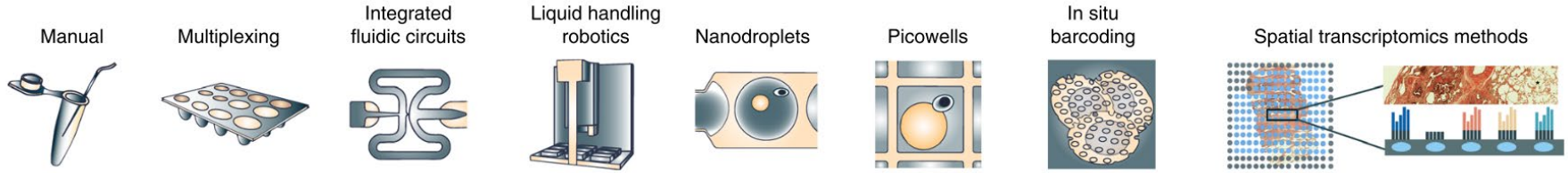


Method:

- Droplet-based
- Tube-based after FACS
- Microwell-based
- Microfluidics-enabled

Key considerations:

- Cell throughput and handling time
- Gene coverage and cell type detection
- Whole transcript versus 3' end counting
- Imaging capability for doublet detection

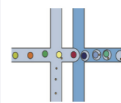


Aldridge and Teichmann, *Nature Communications*, 2020



SINGLE-CELL RNA-SEQ PLATFORMS

IV. Single Cell RNAseq Platform



Method:

- Droplet-based
- Tube-based after FACS
- Microwell-based
- Microfluidics-enabled

Key considerations:

- Cell throughput and handling time
- Gene coverage and cell type detection
- Whole transcript versus 3' end counting
- Imaging capability for doublet detection

3' (and 5') End

Modern Examples: CEL-seq2, MARS-seq, Drop-seq, InDrop, 10X Chromium, SPLiT-seq, Quartz-Seq2, sci-RNA-seq

Features: Generally higher-throughput, higher scalability, lower cost per cell, require lower sequencing depth

Important Controls: Unique Molecular Identifiers (UMIs)

Subpopulation Identification
Differential Expression
Trajectory Inference/Pseudotiming
Network Reconstruction

Full-Length Transcript

Modern Examples: Smart-seq2, SUPeR-seq, MATQ-seq

Features: Tend to detect more expressed genes, better coverage of low-abundance genes, can incorporate nascent transcripts, allows for more downstream analysis

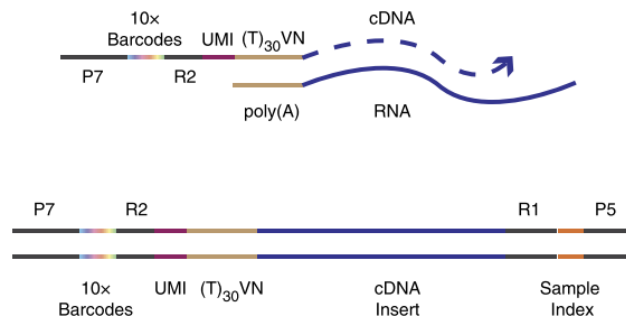
Important Controls: Spike-In RNA

Alternative Splicing
Allelic Expression
RNA Editing



BASIC SC-RNASEQ INDEXING PRINCIPLES

- Applying Barcodes
- UMI use
- Template-Switching



V. Library Sequencing

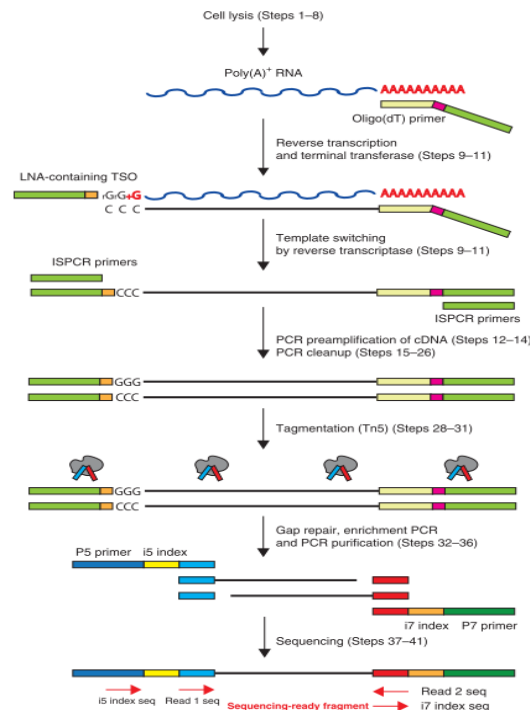


Method:

- Illumina NGS
- Compatible with cDNA library

Sequencing depth considerations:

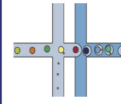
- 3' end counting: low depth ~50K RPC
- Whole transcript: high depth ~1M RPC
- Alternative splicing: ~20-30M RPC
- Iterative optimization for biological system



DROPLET-BASED PLATFORMS

- Reaction Specificity
- Poisson Inclusion
- Ambient RNA correction

IV. Single Cell RNAseq Platform

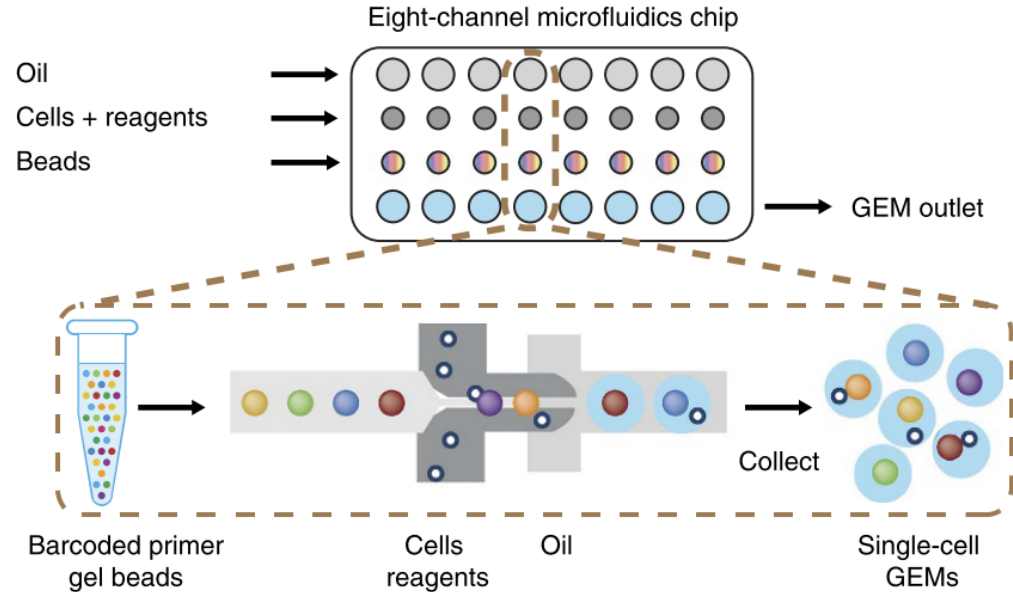


Method:

- Droplet-based
- Tube-based after FACS
- Microwell-based
- Microfluidics-enabled

Key considerations:

- Cell throughput and handling time
- Gene coverage and cell type detection
- Whole transcript versus 3'end counting
- Imaging capability for doublet detection



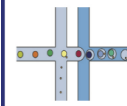
Zheng et al, *Nature Communications*, 2017



COMBINATORIAL INDEXING PLATFORMS

- Highly Scalable
- Simplified Internal Multiplexing
- Maximizes Condition Inclusion

IV. Single Cell RNAseq Platform

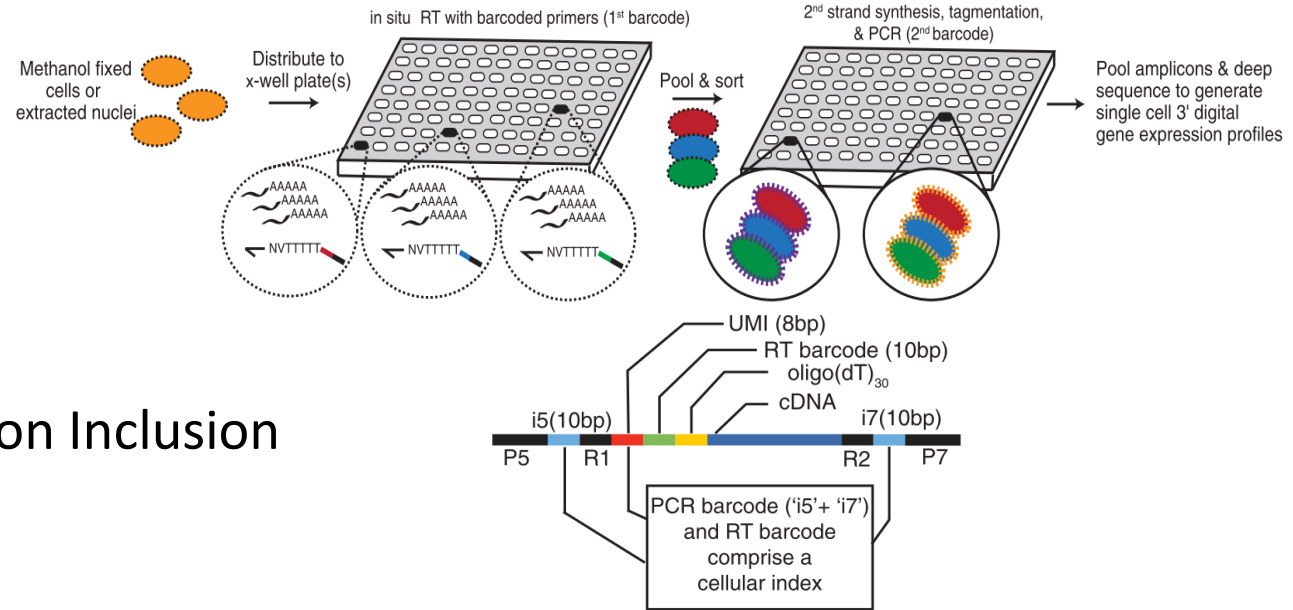


Method:

- Droplet-based
- Tube-based after FACS
- Microwell-based
- Microfluidics-enabled

Key considerations:

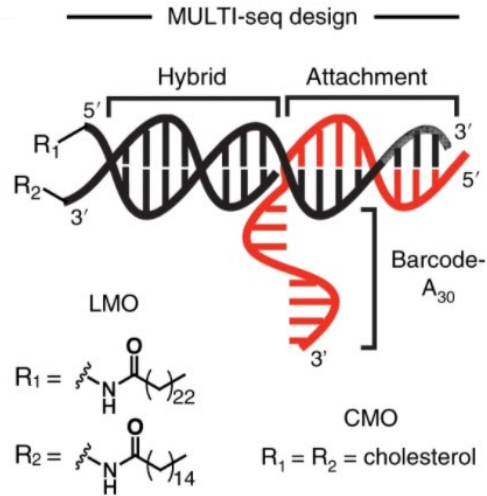
- Cell throughput and handling time
- Gene coverage and cell type detection
- Whole transcript versus 3' end counting
- Imaging capability for doublet detection



Cao et al, *Science*, 2017

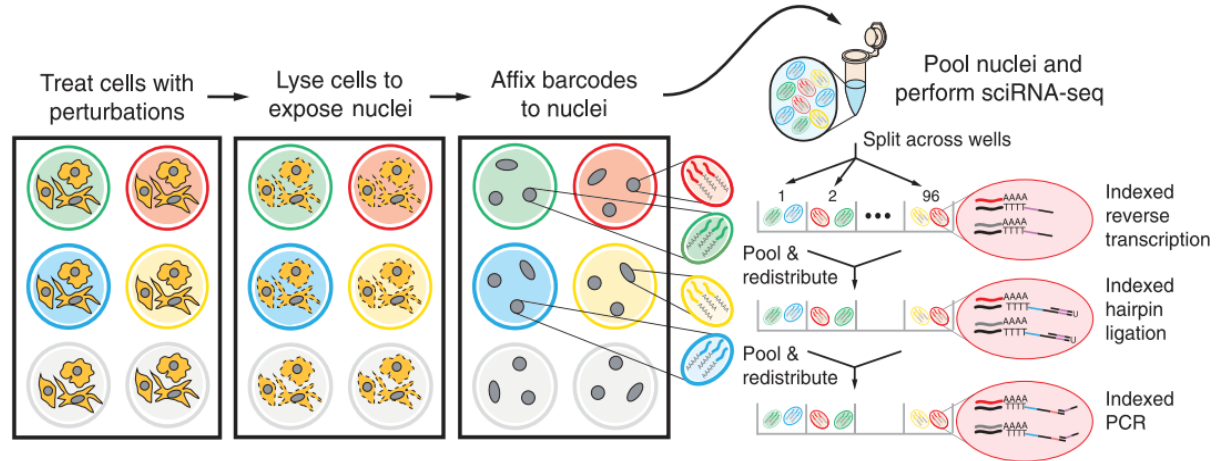


PLATFORM CONSIDERATIONS: STRATEGIC MULTIPLEXING ADVANCES



Multi-Seq

McGinnis et al, *Nature Methods*, 2019

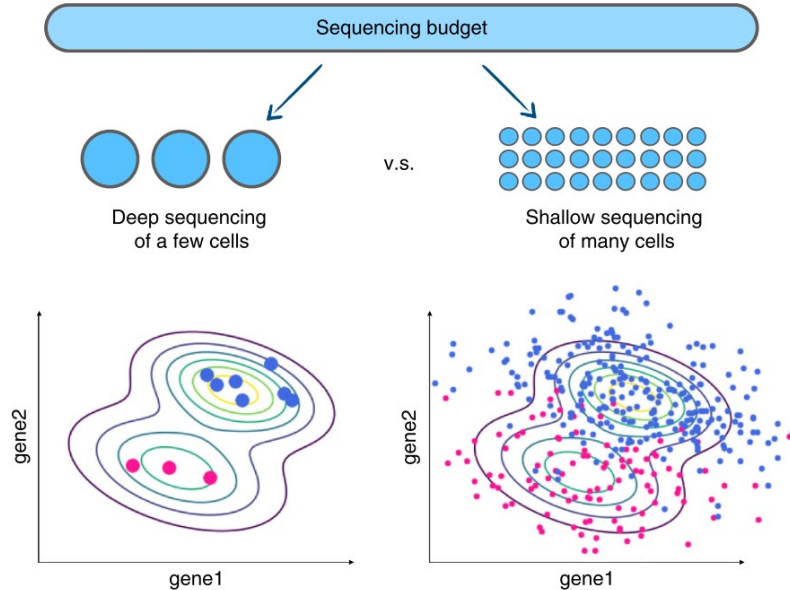


Sci-Plex

Srivatsan et al, *Science*, 2020



SEQUENCING DEPTH CONSIDERATIONS



V. Library Sequencing

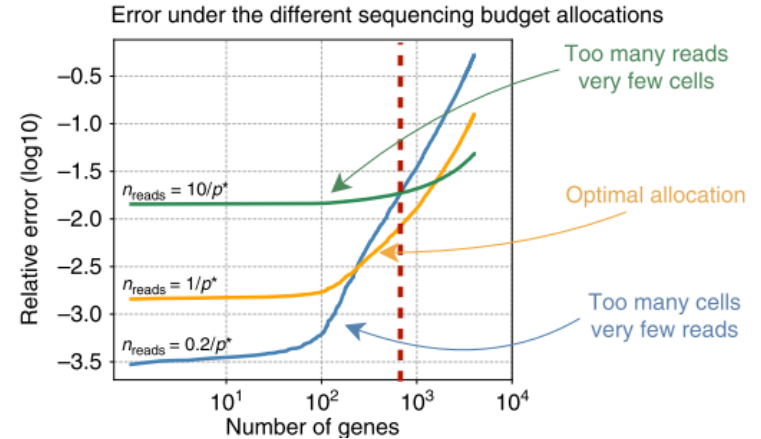


Method:

- Illumina NGS
- Compatible with cDNA library

Sequencing depth considerations:

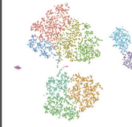
- 3' end counting: low depth ~50K RPC
- Whole transcript: high depth ~1M RPC
- Alternative splicing: ~20-30M RPC
- Iterative optimization for biological system



Zhang et al, *Nature Communications*, 2020



VI. Computational Analysis



Key considerations:

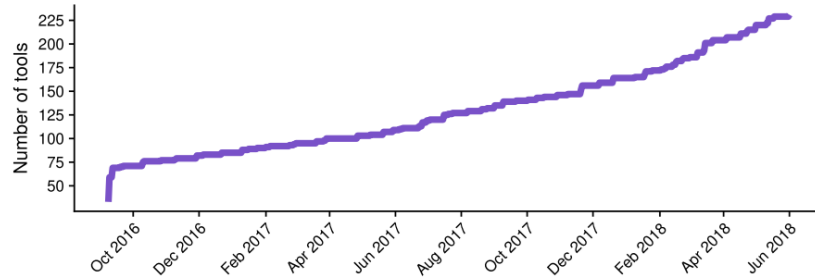
- Separation of *batch* and *condition*
- Technical vs. biological variation

Sample Batch correction approaches:

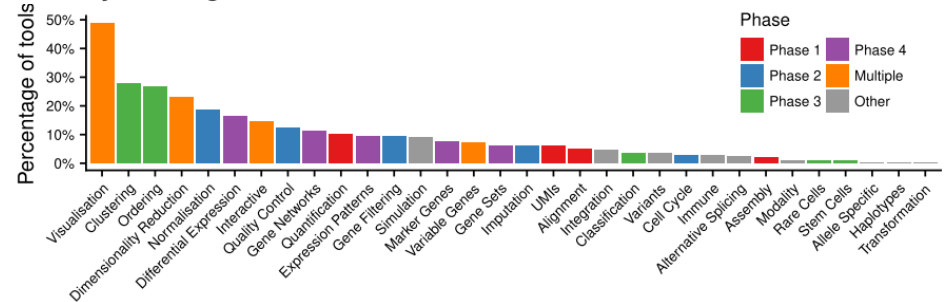
- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST

SC-RNASEQ DATA ANALYSIS: TOOLS

Increase in tools over time

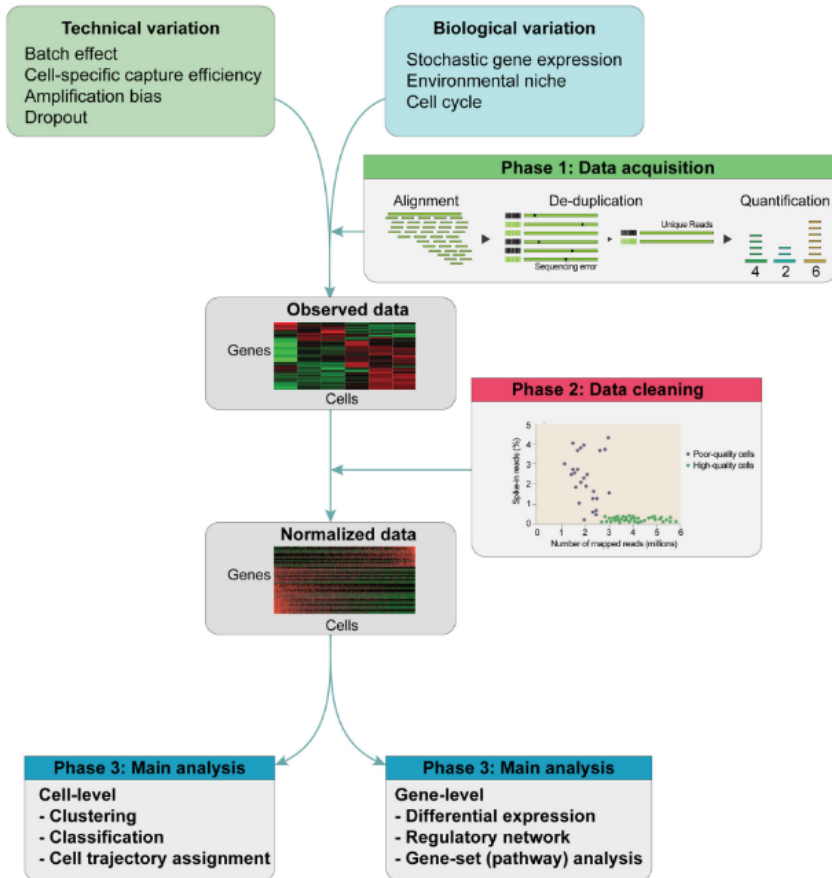


Analysis categories



Zappia et al, *PLOS Computational Biology*, 2018





VI. Computational Analysis

Key considerations:

- Separation of *batch* and *condition*
- Technical vs. biological variation

Sample Batch correction approaches:

- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST

SC-RNASEQ DATA ANALYSIS: PRACTICE

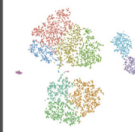


PRE- PROCESSING: QUALITY CONTROL COVARIATES

- Counts per Barcode (Cell)
- Genes per Barcode (Cell)
- Fraction of Mitochondrial Reads
- Batch Effect Corrections
- Multi-Variate Assessment
 - Permissive vs Conservative

Luecken & Theis, *Molecular Systems Biology*, 2019

VI. Computational Analysis

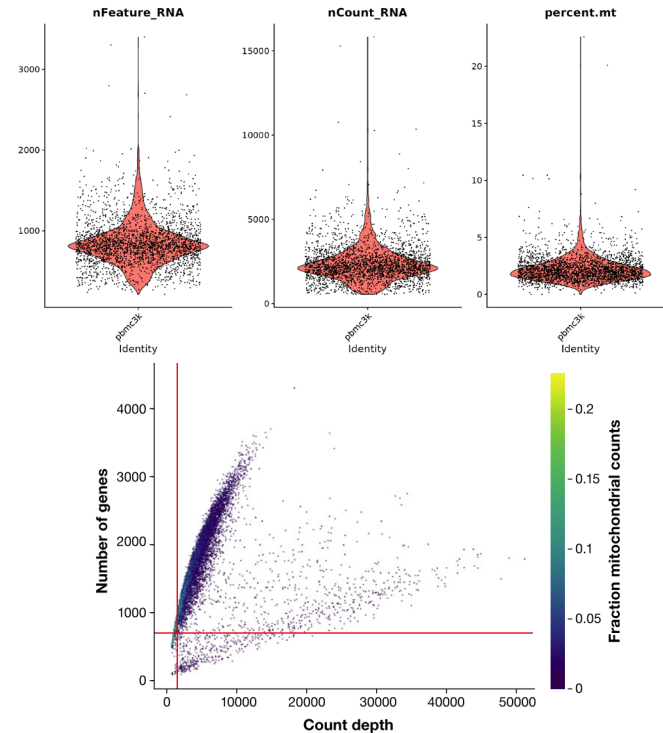


Key considerations:

- Separation of *batch* and *condition*
- Technical vs. biological variation

Sample Batch correction approaches:

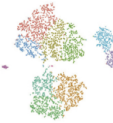
- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST



PRE-PROCESSING: NORMALIZATION AND SCALING

- Critical to obtaining relative gene expression between cells
- Account for variability in count depth and gene dropout
- Within Sample Normalization
 - CPM/C Normalization – scran
 - TPM – full length vs non
 - Non-Linear Normalization
- Scaling Mean and Variance

VI. Computational Analysis

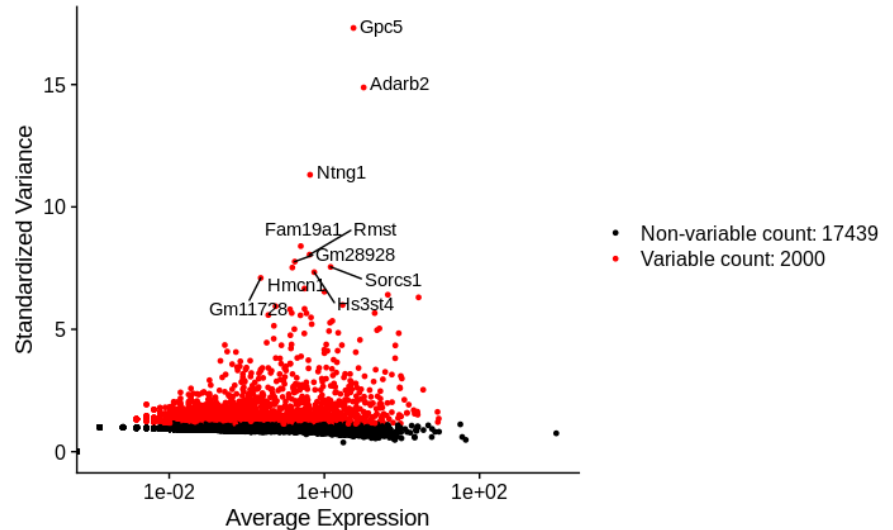


Key considerations:

- Separation of *batch* and *condition*
- Technical vs. biological variation

Sample Batch correction approaches:

- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST



Gallegos et al, *In Review*

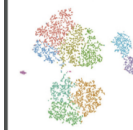


DIMENSIONALITY REDUCTION AND VISUALIZATION

- Feature Selection
- Principal Component Analysis
- Diffusion Maps
- T-Distributed Stochastic Neighbor Embedding
- Uniform Approximation and Projection

Luecken & Theis, *Molecular Systems Biology*, 2019

VI. Computational Analysis

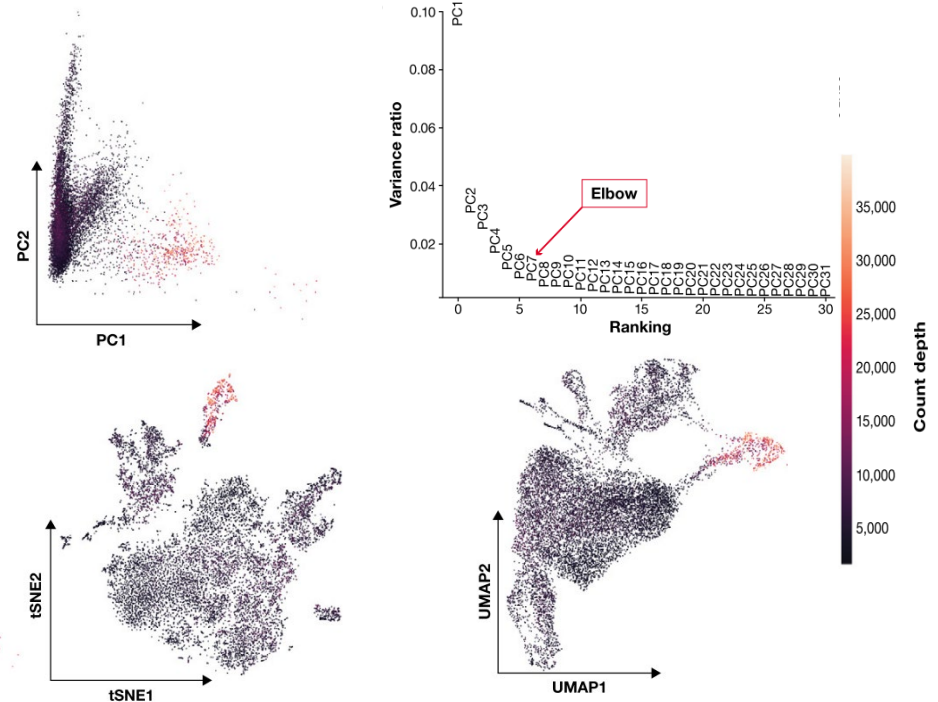


Key considerations:

- Separation of *batch* and *condition*
- Technical vs. biological variation

Sample Batch correction approaches:

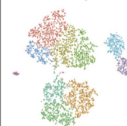
- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST



FEATURE DETECTION AND CLUSTERING

- Unsupervised Feature Detection
 - Highest Variable Genes (Seurat FVG)
 - Spike-In Based (BASiCS)
 - Dropout Based (M3Drop)
- Clustering vs Community Detection
 - K-means
 - K-Nearest Neighbor
- “Marker” Genes and “Cell Types”

VI. Computational Analysis

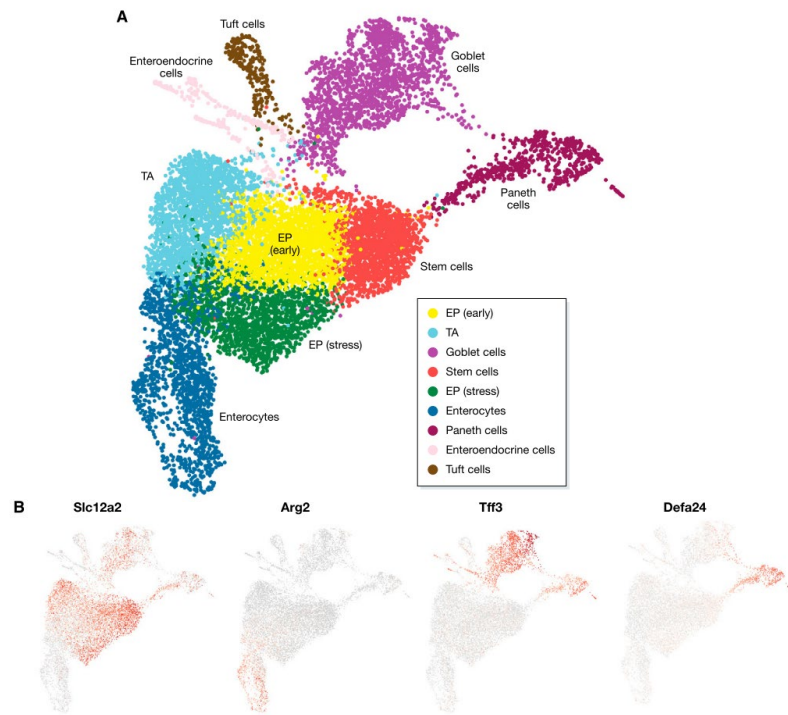


Key considerations:

- Separation of *batch* and *condition*
- Technical vs. biological variation

Sample Batch correction approaches:

- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST



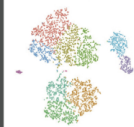
DIFFERENTIAL EXPRESSION AND TRAJECTORY INFERENCE

- Differential Expression
- Trajectory Inference/Pseudotiming

- Bulk Tools vs SC Tools
- MAST
- Differentiation and Progressive Cell Change
- Linear Minimum variability
- Monocle

Luecken & Theis, *Molecular Systems Biology*, 2019

VI. Computational Analysis

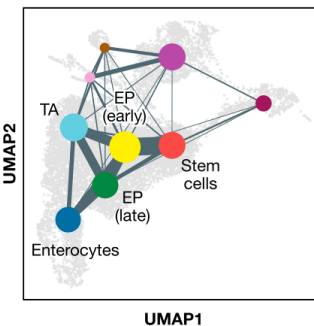
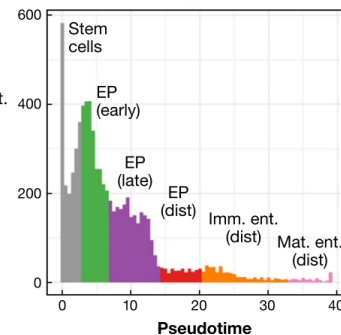
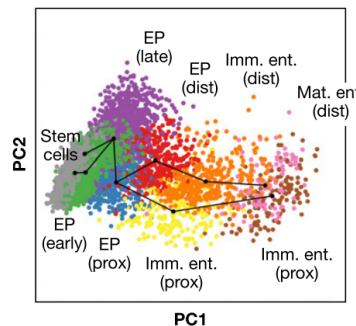
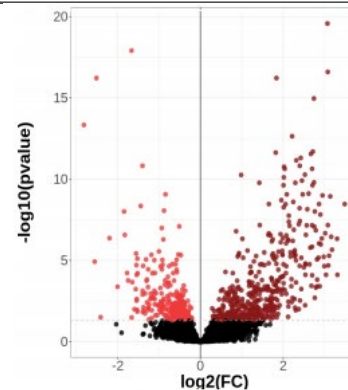


Key considerations:

- Separation of batch and condition
- Technical vs. biological variation

Sample Batch correction approaches:

- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST



COMMON PLATFORMS AND ANALYSIS TOOLS

Methods	Transcript coverage	UMI possibility	Strand specific	References
Tang method	Nearly full-length	No	No	Tang et al., 2009
Quartz-Seq	Full-length	No	No	Sasagawa et al., 2013
SUPeR-seq	Full-length	No	No	Fan X. et al., 2015
Smart-seq	Full-length	No	No	Ramskold et al., 2012
Smart-seq2	Full-length	No	No	Picelli et al., 2013
MATQ-seq	Full-length	Yes	Yes	Sheng et al., 2017
STRT-seq and STRT/C1	5'-only	Yes	Yes	Islam et al., 2011, 2012
CEL-seq	3'-only	Yes	Yes	Hashimshony et al., 2012
CEL-seq2	3'-only	Yes	Yes	Hashimshony et al., 2016
MARS-seq	3'-only	Yes	Yes	Jaitin et al., 2014
CytoSeq	3'-only	Yes	Yes	Fan H.C. et al., 2015
Drop-seq	3'-only	Yes	Yes	Macosko et al., 2015
InDrop	3'-only	Yes	Yes	Klein et al., 2015
Chromium	3'-only	Yes	Yes	Zheng et al., 2017
SPLiT-seq	3'-only	Yes	Yes	Rosenberg et al., 2018
sci-RNA-seq	3'-only	Yes	Yes	Cao et al., 2017
Seq-Well	3'-only	Yes	Yes	Gierahn et al., 2017
DroNC-seq	3'-only	Yes	Yes	Habib et al., 2017
Quartz-Seq2	3'-only	Yes	Yes	Sasagawa et al., 2018

Tools	Category	URL	References
TopHat2	Read mapping	https://ccb.jhu.edu/software/tophat/index.shtml	Kim et al., 2013
STAR	Read mapping	https://github.com/alexdobin/STAR	Dobin and Gingeras, 2015
HISAT2	Read mapping	https://ccb.jhu.edu/software/hisat2/index.shtml	Kim et al., 2015
Cufflinks	Expression quantification	https://github.com/cole-trapnell-lab/cufflinks	Trapnell et al., 2010
RSEM	Expression quantification	https://github.com/deweylab/RSEM	Li and Dewey, 2011
StringTie	Expression quantification	https://github.com/gpertea/stringtie	Pertea et al., 2015

Chen et al, *Frontiers in Genetics*, 2019



COMMON PLATFORMS AND ANALYSIS TOOLS

Methods	URL	References	Methods	Category	URL	References	Tools	Dimensionality reduction	URL	References
SC3	http://bioconductor.org/packages/SC3	Kiselev et al., 2017	ROTS	Single cell	https://bioconductor.org/packages/release/bioc/html/ROTS.html	Seyednasrollah et al., 2016	Monocle	ICA	http://cole-trapnell-lab.github.io/monocle-release/	Trapnell et al., 2014
ZIFA	https://github.com/epierson9/ZIFA	Pierson and Yau, 2015	MAST	Single cell	https://github.com/RGLab/MAST	Finak et al., 2015	Waterfall	PCA	https://www.cell.com/cms/10.1016/j.stem.2015.07.013/attachment/3e966901-034f-418a-a439-996c50292a11/mmc9.zip	Shin et al., 2015
Destiny	https://github.com/theislab/destiny	Angerer et al., 2016	BCseq	Single cell	https://bioconductor.org/packages/devel/bioc/html/bcSeq.html	Chen and Zheng, 2018	Wishbone	Diffusion maps	https://github.com/ManuSetty/wishbone	Setty et al., 2016
SNN-Cliq	http://bioinfo.uncc.edu/SNNCliq/	Xu and Su, 2015	SCDE	Single cell	http://hms-dbmi.github.io/scde/	Kharchenko et al., 2014	GrandPrix	Gaussian Process Latent Variable Model	https://github.com/ManchesterBioInference/GrandPrix	Ahmed et al., 2019
RaceID	https://github.com/dgrun/RaceID	Grun et al., 2015	DEsingle	Single cell	https://bioconductor.org/packages/DEsingle	Miao et al., 2018	SCUBA	t-SNE	https://github.com/gcyuan/SCUBA	Marco et al., 2014
SCUBA	https://github.com/gcyuan/SCUBA	Marco et al., 2014	Census	Single cell	http://cole-trapnell-lab.github.io/monocle-release/	Qiu et al., 2017	DPT	Diffusion maps	https://media.nature.com/original/nature-assets/nmeth/journal/v13/n10/extref/nmeth.3971-S3.zip	Haghverdi et al., 2016
BackSPIN	https://github.com/linnarsson-lab/BackSPIN	Zeisel et al., 2015	D3E	Single cell	https://github.com/hemberg-lab/D3E	Delmans and Hemberg, 2016	TSCAN	PCA	https://github.com/zij90/TSCAN	Ji and Ji, 2016
PAGODA	http://hms-dbmi.github.io/scde/	Fan et al., 2016	BPSC	Single cell	https://github.com/nghiavtr/BPSC	Vu et al., 2016	Monocle2	RGE	http://cole-trapnell-lab.github.io/monocle-release/	Qiu et al., 2017
CIDR	https://github.com/VCCRI/CIDR	Lin et al., 2017	DESeq2	Bulk	https://bioconductor.org/packages/release/bioc/html/DESeq2.html	Love et al., 2014	Slingshot	Any	https://github.com/kstreet13/slshot	Street et al., 2018
pcaReduce	https://github.com/JustinaZ/pcaReduce	Zuraskiene and Yau, 2016	edgeR	Bulk	https://bioconductor.org/packages/release/bioc/html/edgeR.html	Robinson et al., 2010	CellRouter	Any	https://github.com/edroaldo/cellrouter	Lummertz da Rocha et al., 2018
Seurat	https://github.com/satijalab/seurat	Satija et al., 2015	Limma	Bulk	http://bioconductor.org/packages/release/bioc/html/limma.html	Ritchie et al., 2015				
TSCAN	https://github.com/zij90/TSCAN	Ji and Ji, 2016	Ballgown	Bulk	http://www.bioconductor.org/packages/release/bioc/html/ballgown.html	Frazee et al., 2015				



ACKNOWLEDGMENTS

- US EPA – Chorley Lab
 - Dr. Brian Chorley
 - Gleta Carswell
 - Gail Nelson
 - Nyssa Tucker
- US EPA – Hunter Lab
 - Dr. Sid Hunter
 - Susan Jeffay
 - Sarah Park
 - Jessica Conley
- Duke University – West Lab
 - Dr. Anne West
 - Melyssa Minto
 - Luke Bartelt



QUESTIONS

