2018 ABRF Meeting – Satellite Workshop 4
Bridging the Gap: Isolation to Translation (Single Cell RNA-Seq)
Sunday, April 22

Methods, applications & analysis of scRNA-Seq:

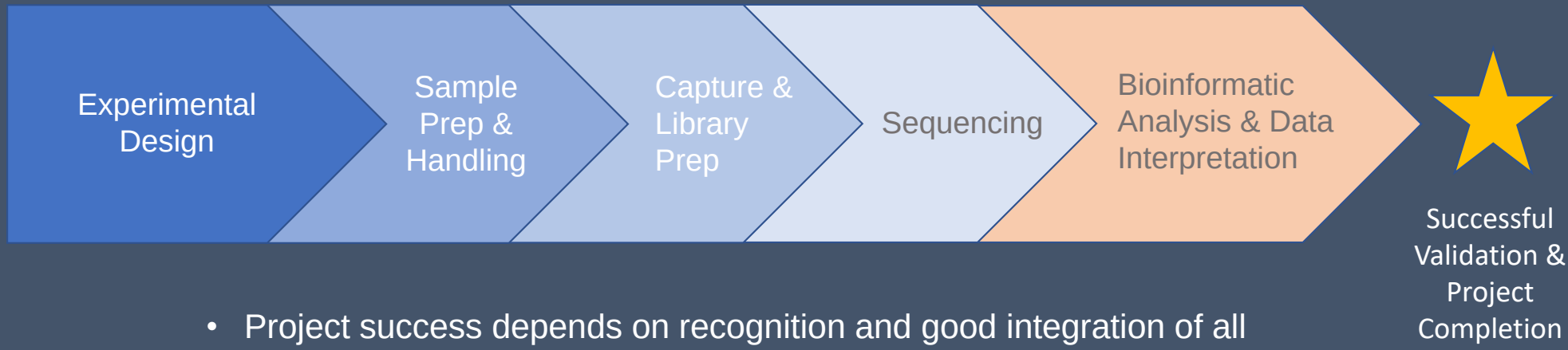
How an integrated understanding of every step makes for a better single cell experiment

Michael Kelly, PhD

Team Lead, NCI Single Cell Analysis Facility

Frederick National Laboratory
for Cancer Research
sponsored by the National Cancer Institute

End-to-End Support for Single Cell RNA-Seq



- Project success depends on recognition and good integration of all aspects of the single cell genomics workflow
- Failures in one aspect can lead to problems downstream
- Lack of integration makes troubleshooting especially difficult
- Relatively “easy” to collect data; potential to get “stuck” at last step

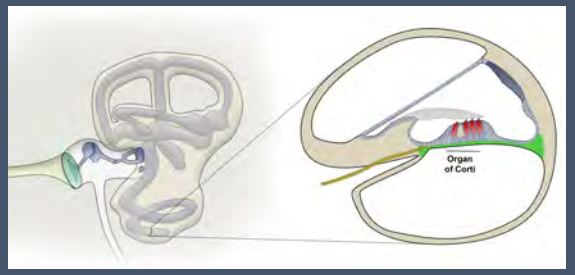
Lessons learned from various scRNA-Seq projects

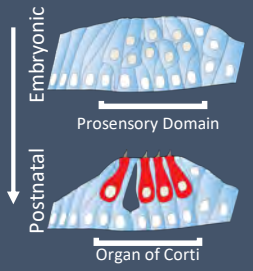
- Selecting the right approach for the project goals
  - Importance of numbers and capture efficiency
  - Full-length transcript versus 3’ (or 5’) end-counting
- Sample prep matters
  - Single nuclei RNA-Seq for specific applications
  - Effects of low viability sample preps on downstream data quality
- Data wrangling
  - What single cell RNA-Seq looks like and why the reference is important
- A proposed model for integrated informatics support

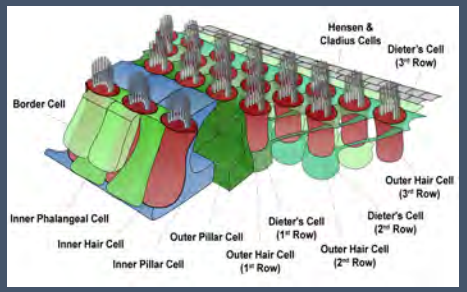
Selecting the right approach for the project goals

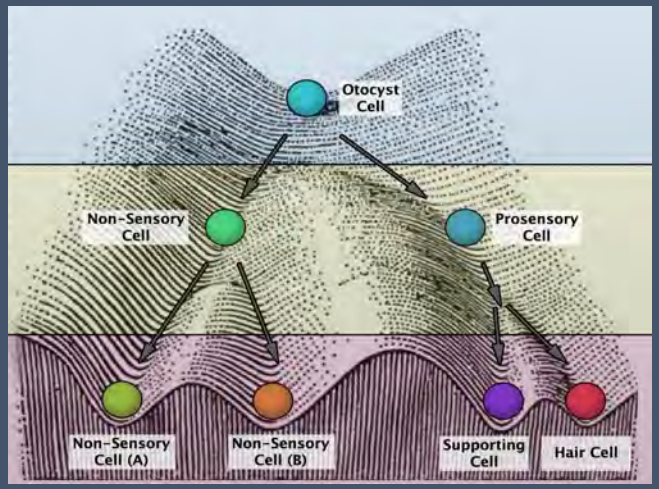
Importance of numbers and capture efficiency

Single Cell RNA-Seq to Identify Cell-Type Specific Transcriptional Programs in Mammalian Cochlea









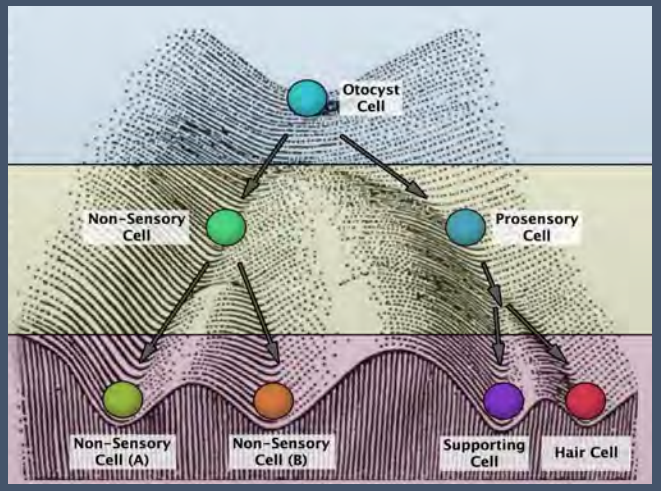
Modified from Waddington 1957

Goals of our single cell RNA-Seq study

- Transcriptional programs
- Understand changes in cellular plasticity
- Novel cell type markers

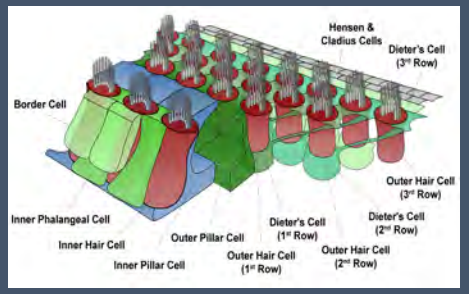
Theoretical principle of trajectory analysis with single cell RNA-Seq data

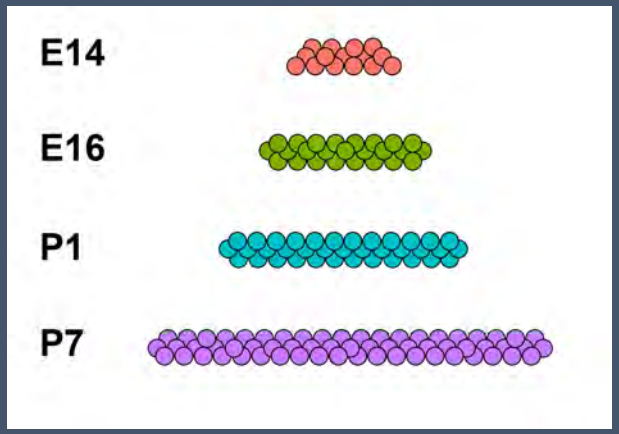
Undifferentiated Precursors



Modified from Waddington 1957

Differentiated Cell Types

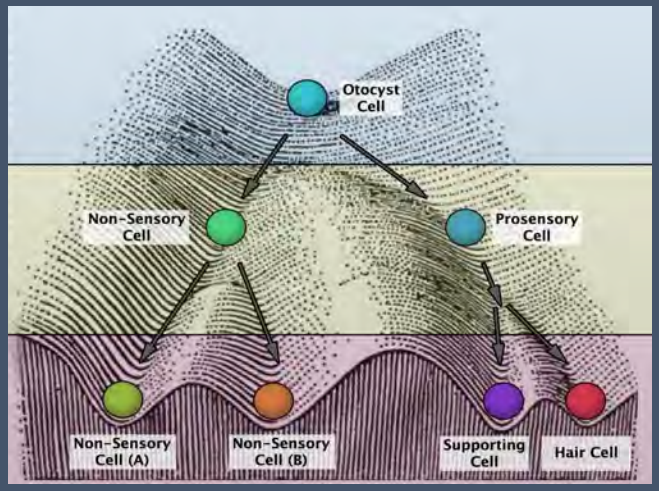




Single cell profiles at multiple profiles provide snapshots of transcriptional expression across many cells

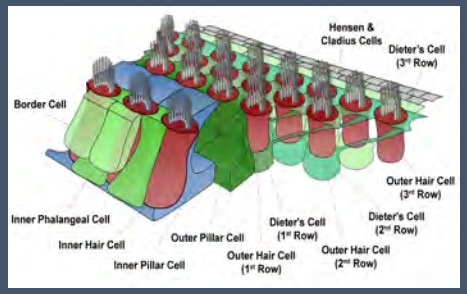
Theoretical principle of trajectory analysis with single cell RNA-Seq data

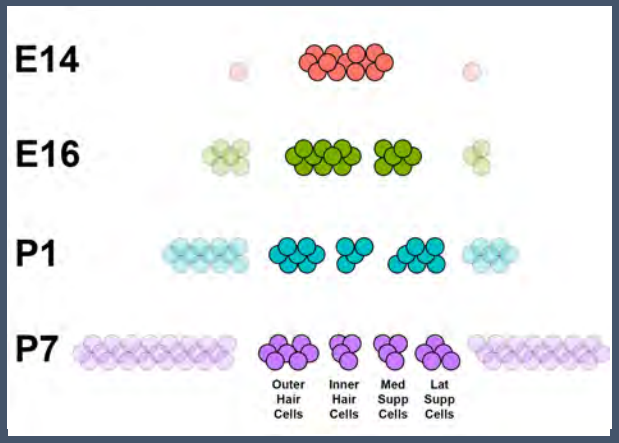
Undifferentiated Precursors



Modified from Waddington 1957

Differentiated Cell Types

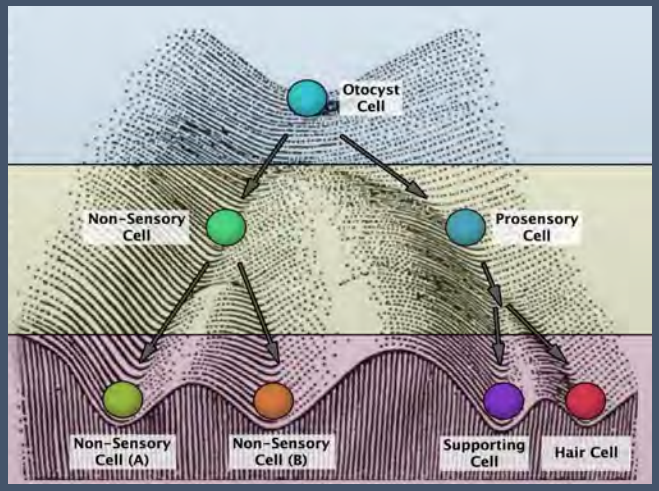




Individual cell types can be identified from each time-point

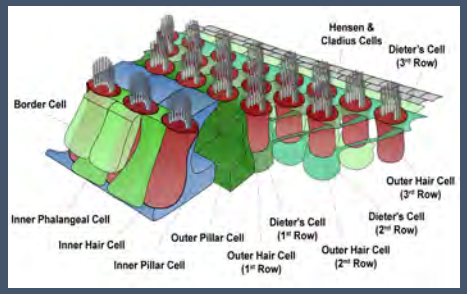
Theoretical principle of trajectory analysis with single cell RNA-Seq data

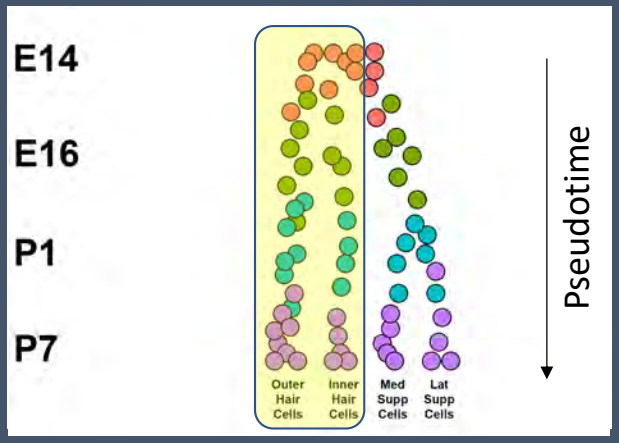
Undifferentiated Precursors



Modified from Waddington 1957

Differentiated Cell Types

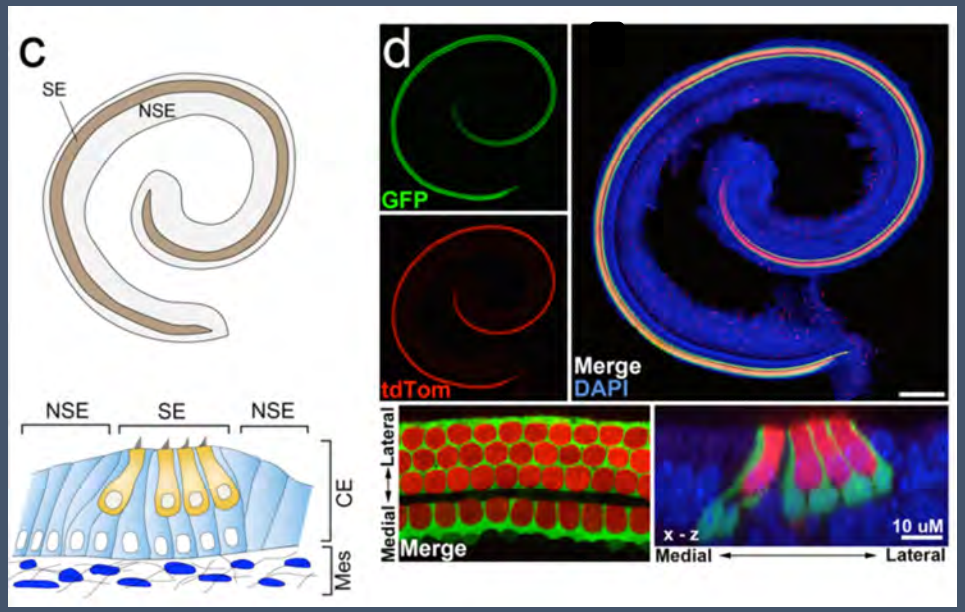


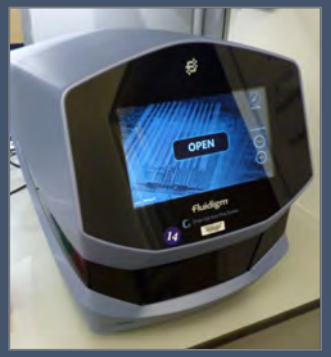


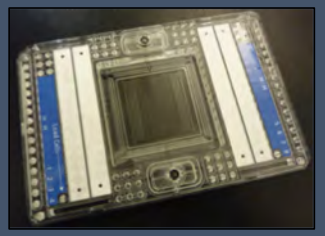
Asynchrony in development allows for generation of a temporal model of expression across ”Pseudotime”

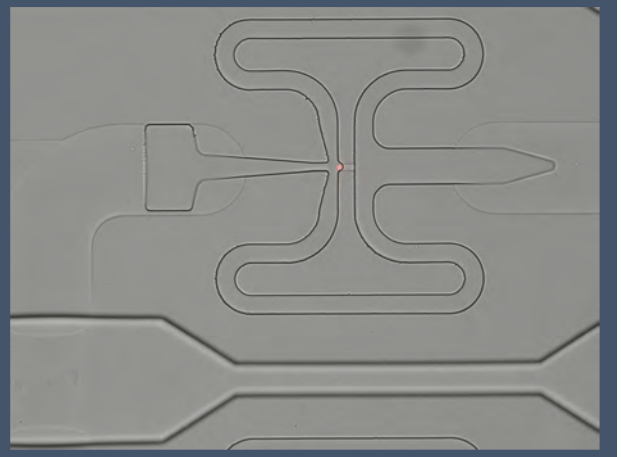
Transcriptional trajectory associated with differentiation of each unique cell type can be analyzed

Single cell profiling with Fluidigm C1 was a good start, but was too low throughput

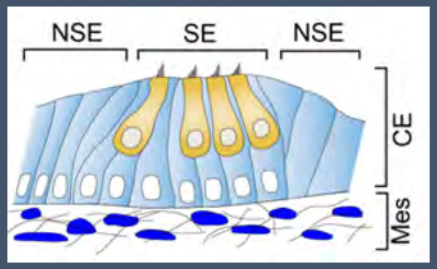


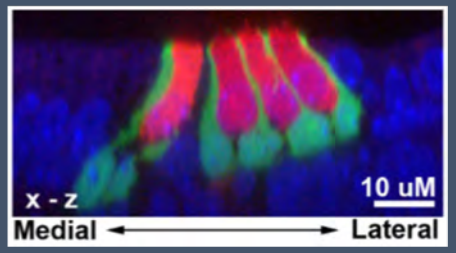


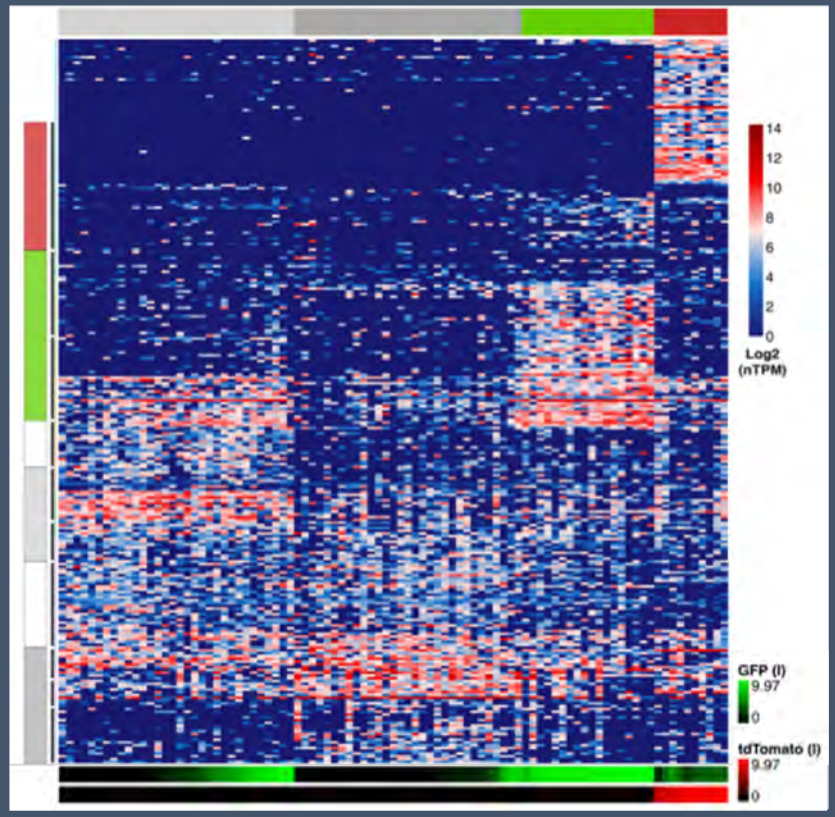




Original single cell dataset could only distinguish major cell type differences

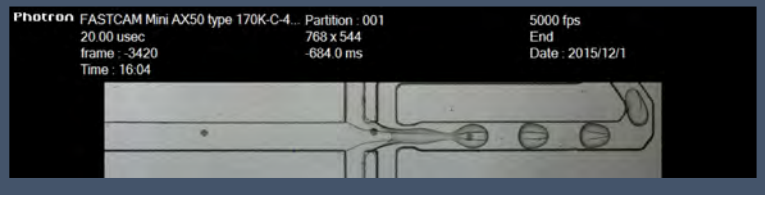


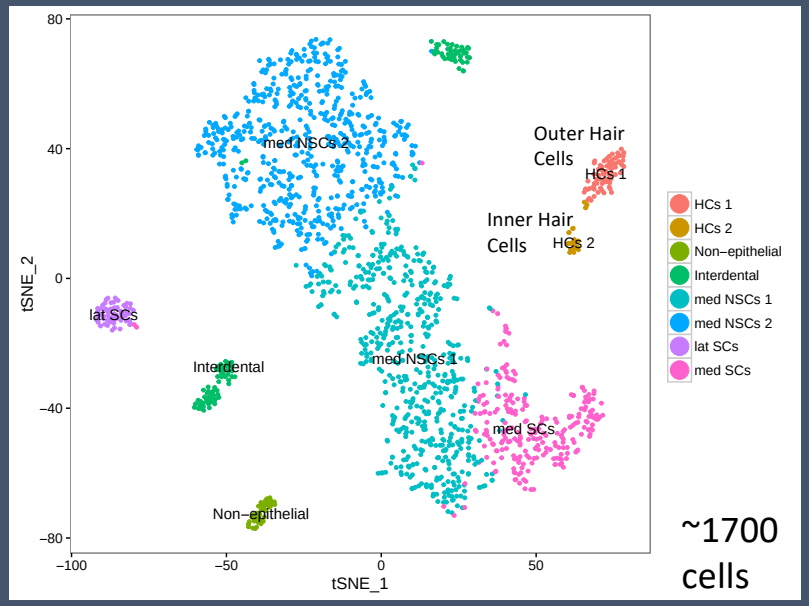




From Burn & Kelly et al 2015

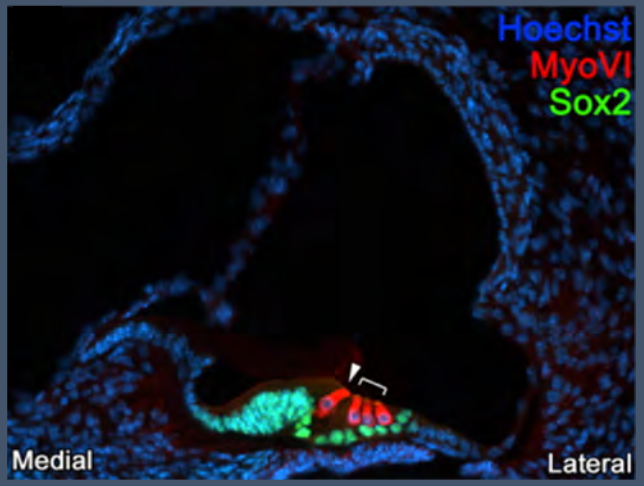
Drop-Seq scRNA-Seq provided better resolution than Fluidigm C1





~1700 cells

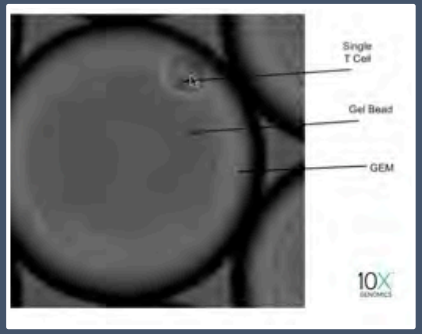
Drop-Seq data generated with Joey Mays (post-bac).

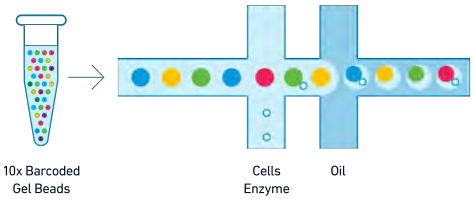


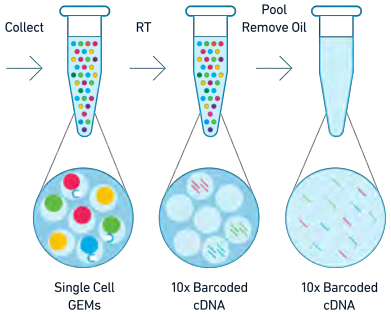
Kelly et al 2012

-Relatively low cost, once established and working consistently (not guaranteed…)
-Considerable troubleshooting required (not plug & play)
-Lower sensitivity (1-2,000 genes detected per cell) than Fluidigm C1 or FAC-Seq
-Limited experimental design (captures back to back)

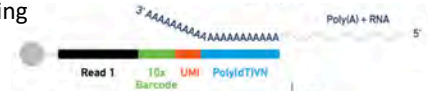
10X Genomics droplet-based single cell RNA-Seq provided high-throughput method to profile large population of cells in unbiased manner





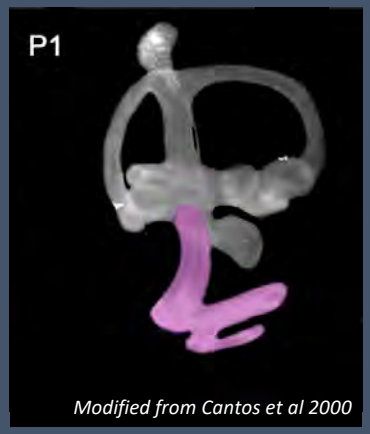


- Cell partitioned with barcoded gel bead
- Cell Barcode & Unique Molecular Identifiers (UMI) on 3'-end of cDNA
- ~2000 median genes detected on average with ~50,000 mean sequencing reads per cells
- Gene-level counts data

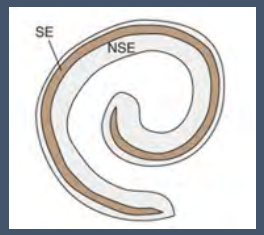


- High efficiency of capture in 10X platform is amenable to smaller input numbers of cells (small tissue or rare population)
- Much better suited to limited samples like mouse cochlea…

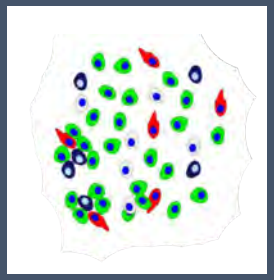
Workflow of droplet-based single cell RNA-Seq profiling of the mammalian cochlea



Extract cochlear epithelium



Dissociate cells

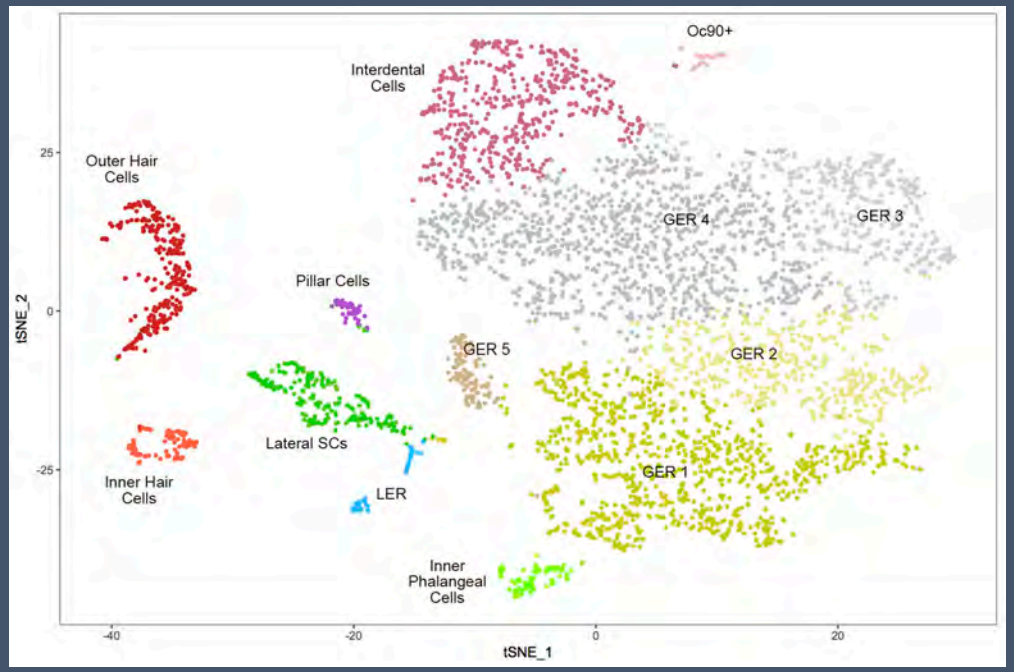


Single cell cDNA library generation & sequencing

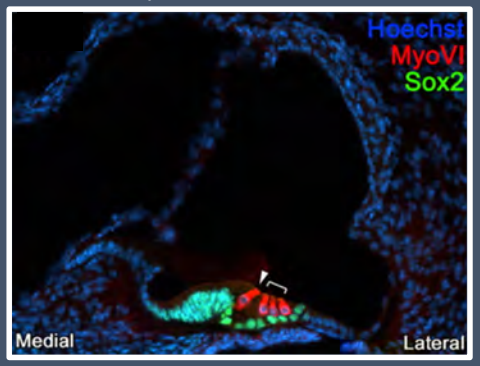
Data Analysis.
And More
Data Analysis.

- Novel Cell Type Markers
- Transcriptional Programs

Single cell RNA-Seq of postnatal day 1 cochlea identifies expected cell subtypes



Cochlear duct cross section
Sensory cells in Red and Green



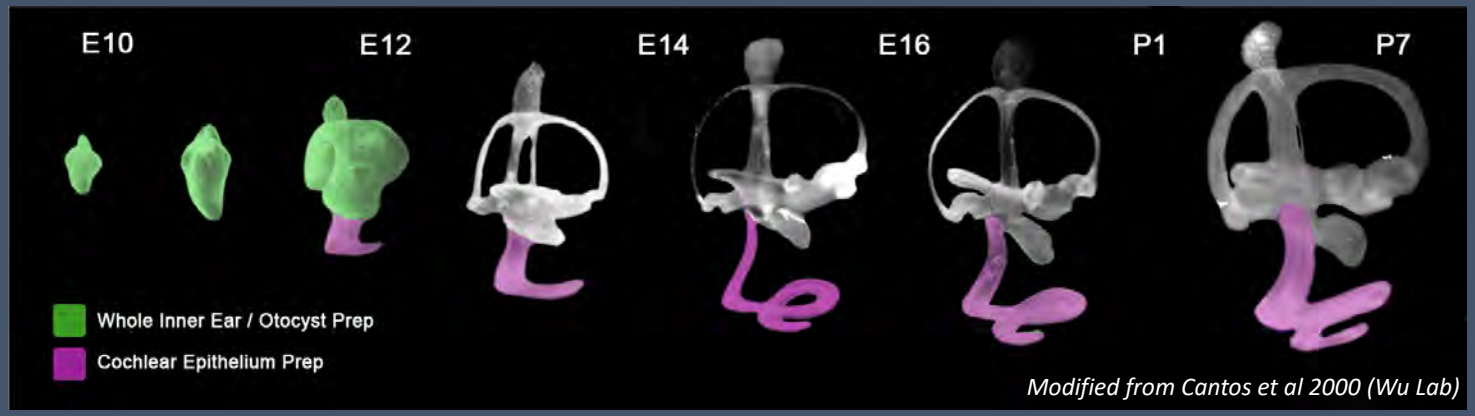
Clusters defined in unbiased manner.

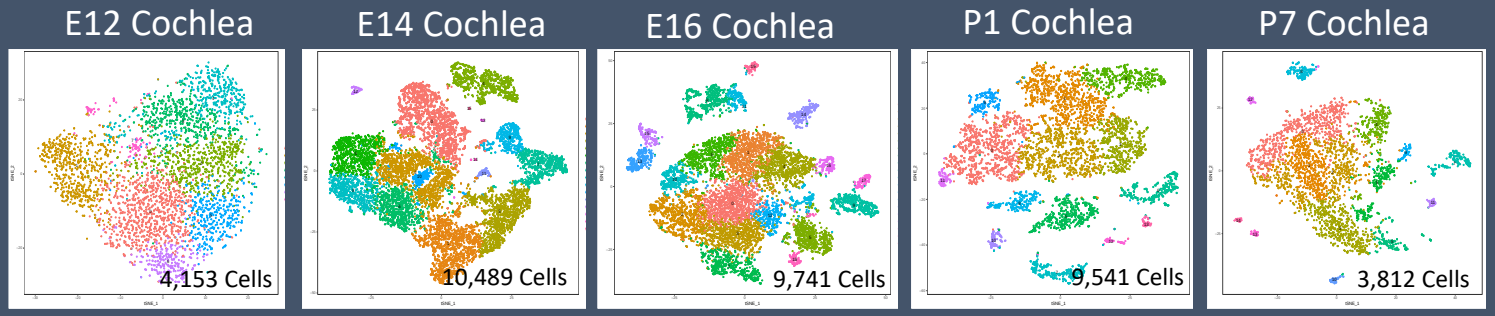
Cluster identities determined from marker gene expression.

- ~5000 P1 cochlear epithelia cells
  - 99 Inner Hair Cells
  - 294 Outer Hair Cells
  - 88 Inner Phalangeal Cells
  - 54 Inner Pillar Cells
  - 229 Lateral Supporting Cells
  - 3753 Medial Non-sensory Cells
  - 74 Lateral Non-Sensory Cells

Differential expression analysis can reveal genes uniquely enriched in each cell subtype (or groups of cells)

Greater number of scRNA-Seq datapoints allows better modeling of dynamic expression changes during differentiation



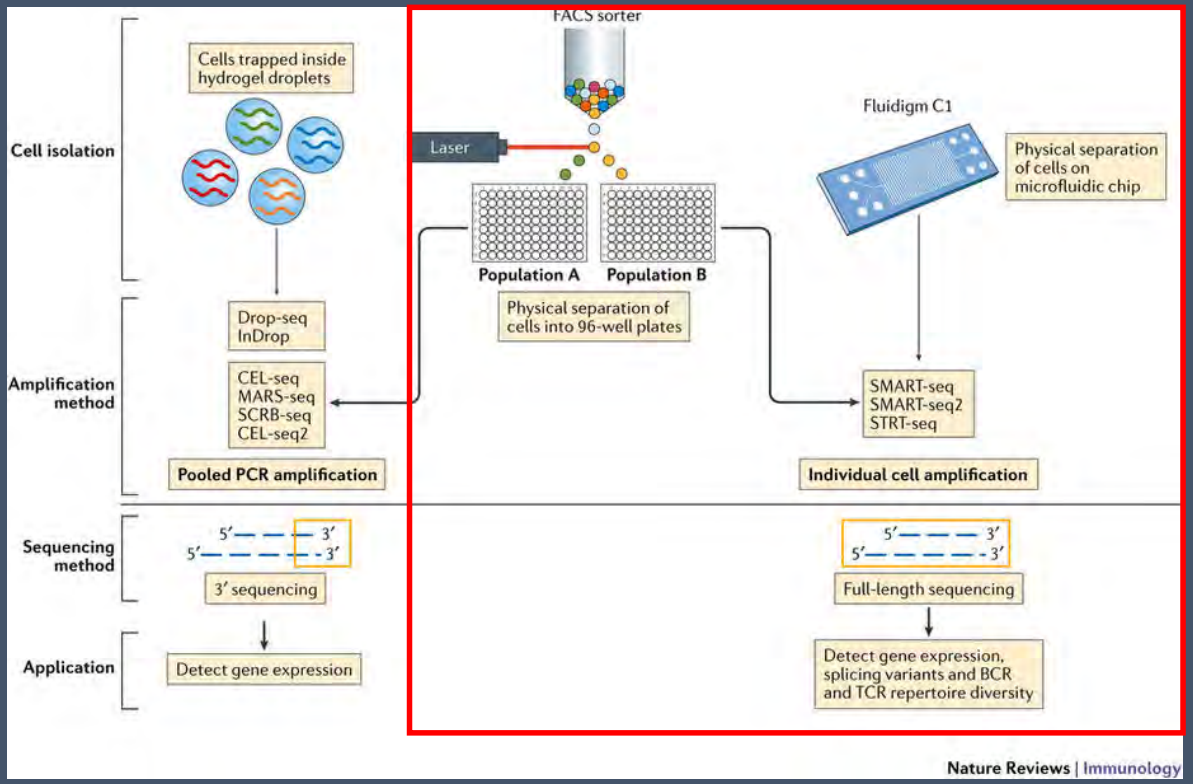


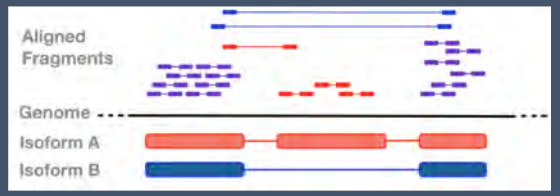
As cells move from undifferentiated precursors to differentiated cell types, they become more transcriptionally distinct

Selecting the right approach for the project goals

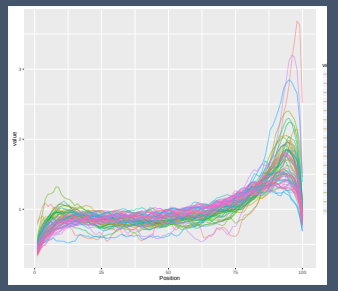
Full-length transcript versus 3’ (or 5’) end-counting

Full-length scRNA-Seq should allow isoform discrimination / quantitation – requires sufficient coverage

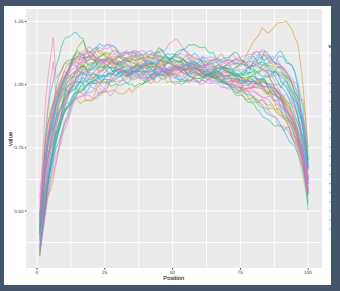




5’ to 3’ Coverage Plots

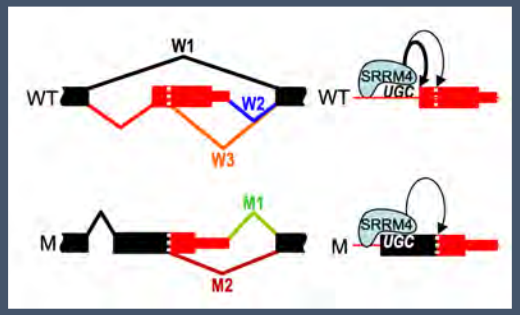


Fluidigm C1 (v1 Chemistry)



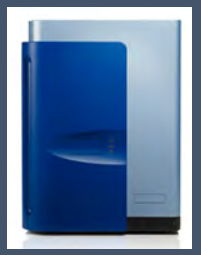
Clontech SMARTer v4 and SMART-Seq2

Single cell gene specific target amplification & qPCR to study crucial differential isoform usage in developing cochlea

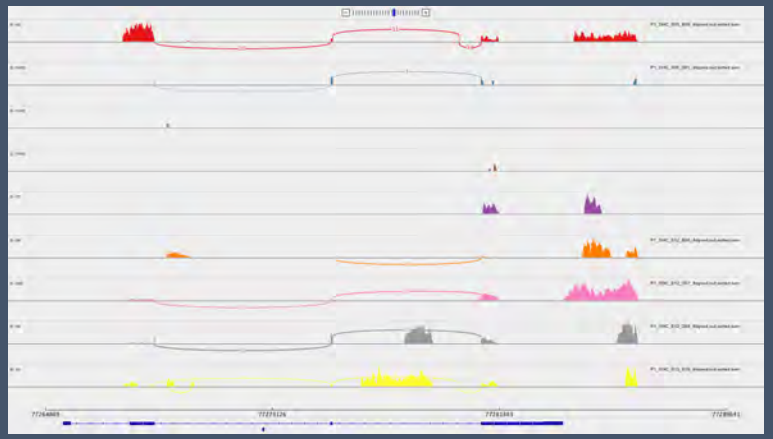


Mutation in mice and humans in intron of gene leads to mis-splicing and deafness. Cell type specific phenotype.

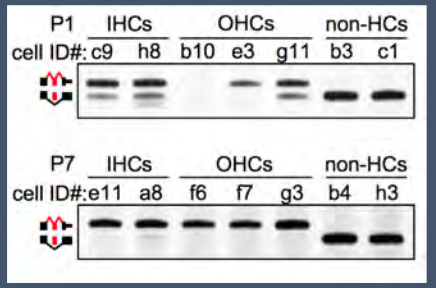
C1



Biomark HD



Whole transcriptome scRNA-Seq data was too sparse and had 3’ bias – not reliable enough. Moved to STA -> qPCR approach

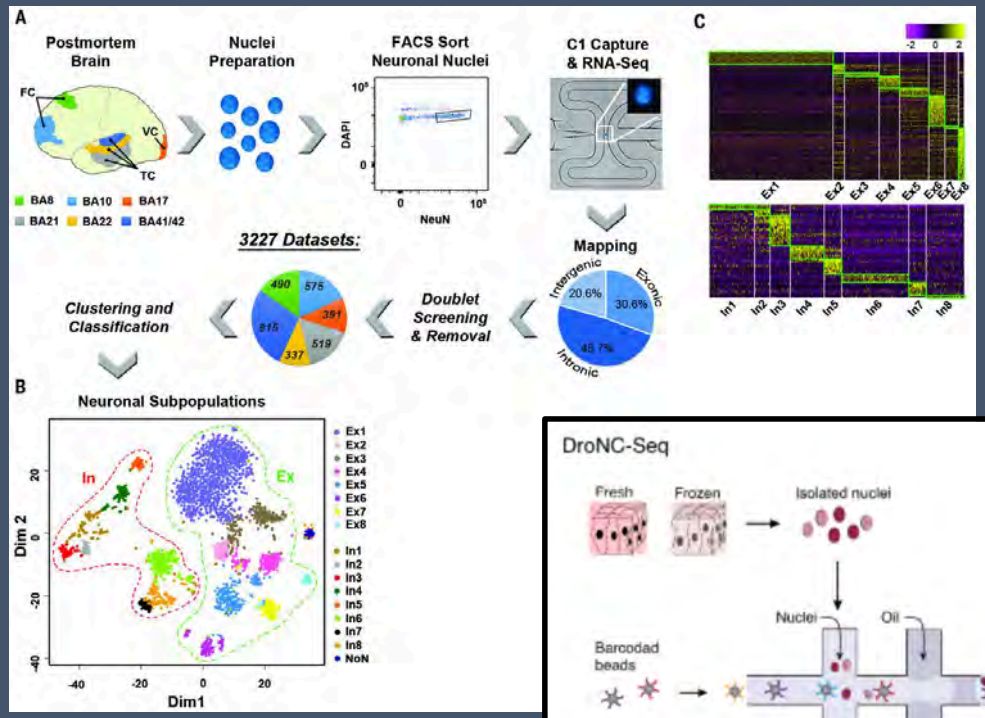


Collaboration with Banfi Lab at Univ of Iowa – Nakano et al paper in final review

Sample prep matters

Single nuclei RNA-Seq for specific applications

Single nucleus RNA-Seq provides a useful alternative to whole single cell RNA-Seq



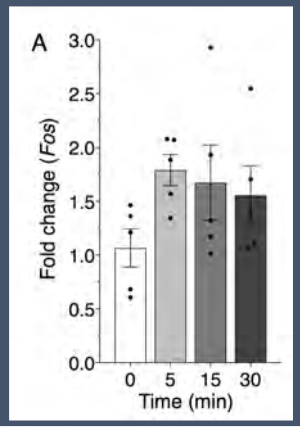
Lake et al 2016 Science

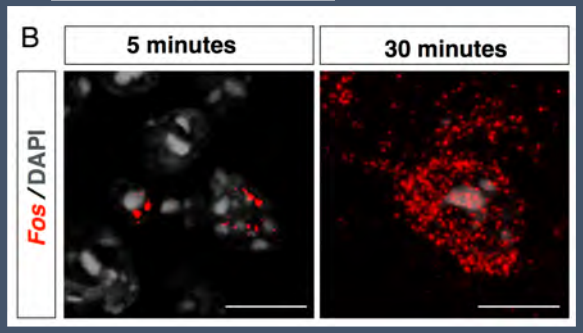
Habib et al 2017 Nature Methods

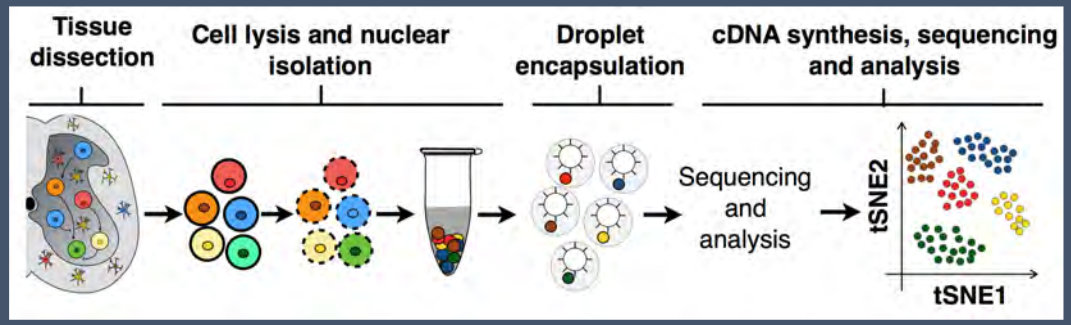
- Allows scRNA-Seq profiling of difficult to dissociate tissues
  - Neuronal tissues
  - Biobanked samples
- Avoid dissociation-associated transcriptional changes
- Nuclear transcriptome representative of whole cell transcriptome
- Lower sensitivity & more intronic reads

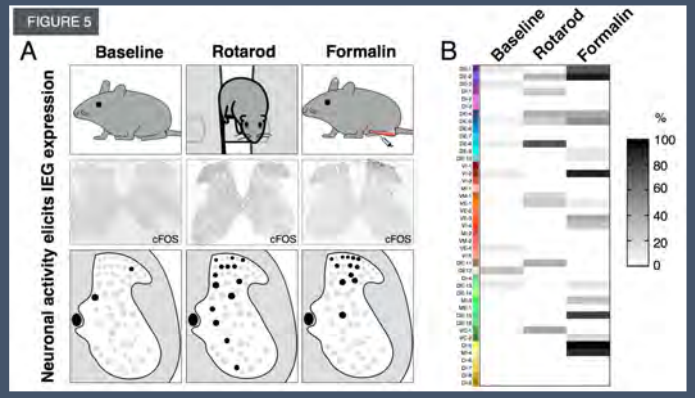
Single nuclei DropSeq to study spinal cord neurons and cell type specific activity in behavior

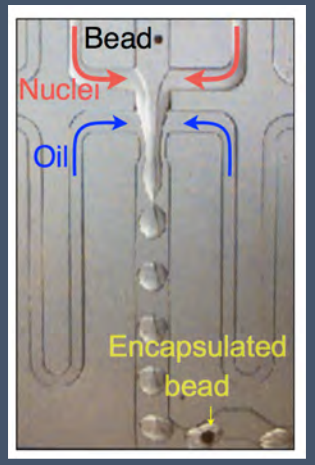
Time post-treatment of target signal upregulation







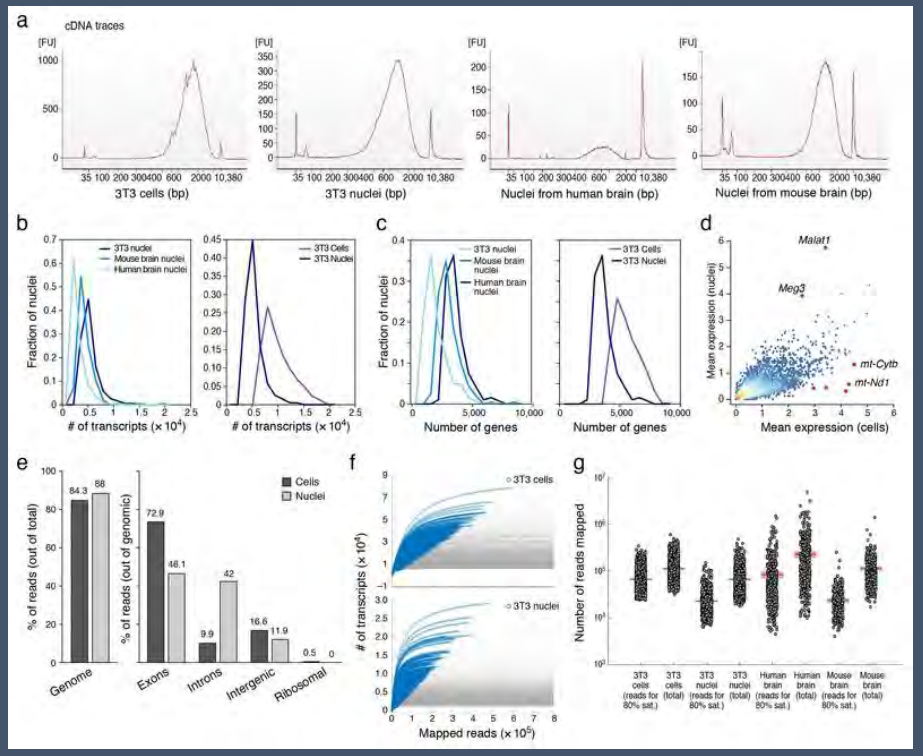




Modified detergent concentration and adjusted flow rates to optimize single nuclei encapsulation efficiency

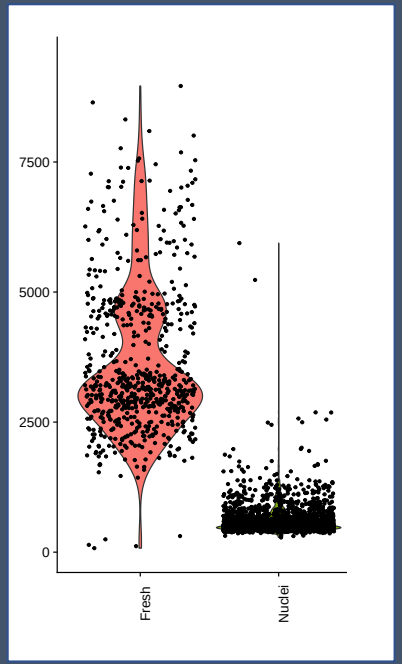
Sathyamurthy & Johnson et al (2018) Cell Reports

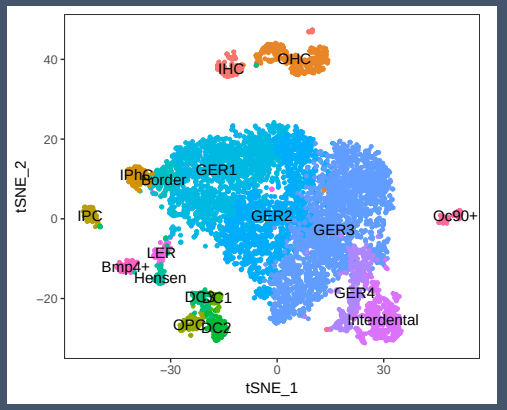
Single nuclei RNA-Seq has limited sensitivity – still good for identifying major cell types

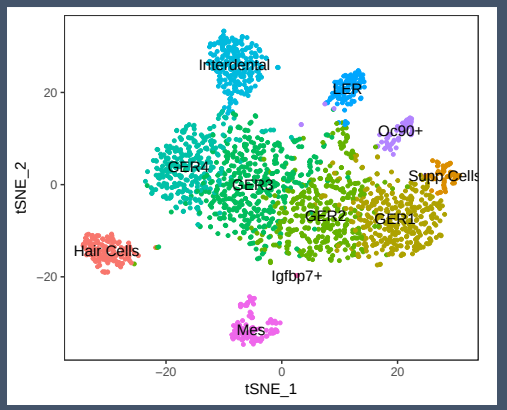


Habib et al 2017 Nature Methods "DroNc-seq"

Evaluating sensitivity and effectiveness with cochlear epithelial cells







Sample prep matters

Effects of low viability sample preps on downstream data quality

scRNA-Seq capture input of a high viability cell preparation is important for

Single Cell Suspension for Optimal Performance

10x Genomics® Single Cell Protocols require a suspension of viable single cells as input. Minimizing the presence of cellular aggregates, dead cells, non-cellular nucleic acids and potential inhibitors of reverse transcription is critical to obtaining high quality data.

1.3. Factors Influencing Cell Recovery

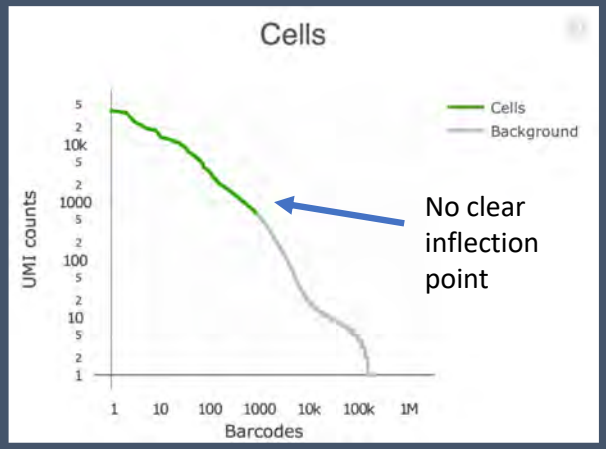
To recover the expected number of cells, it is critical to maximize viability, minimize the cell preparation time, accurately measure the input cell concentration and pipette the correct volume into each reaction.

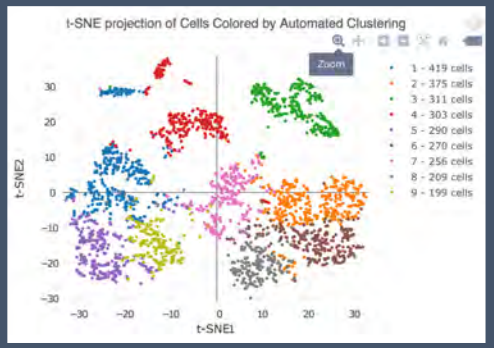
Ideally, input cell suspensions should contain more than 90% viable cells. Non-viable and dying cells generally contain less and more fragmented RNA that may not be efficiently captured by 10x Genomics Single Cell Solutions. The presence of a high fraction of non-viable cells in the input suspension may therefore decrease the apparent efficiency of cell partitioning and recovery.

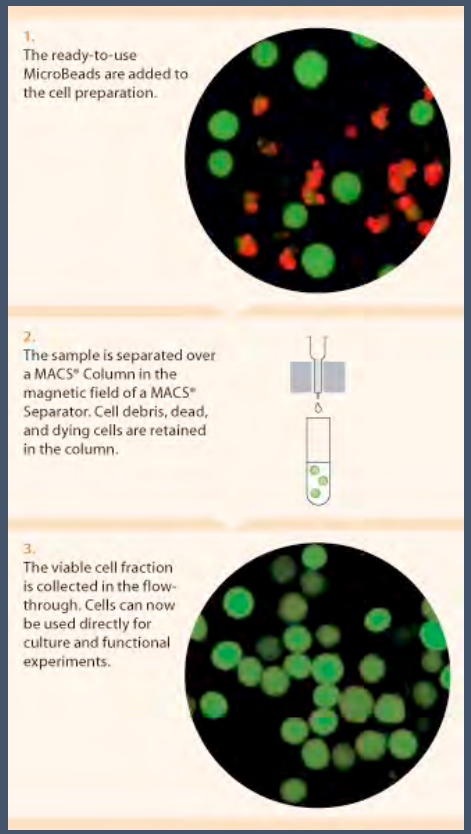
- 10X Genomics recommends loading 90% viable cells or higher
- Often the underlying causes of below target # of cells
- Can contribute a “background signal” – ambient RNA

Effects of low cell viability on downstream and options to enrich for viable cells

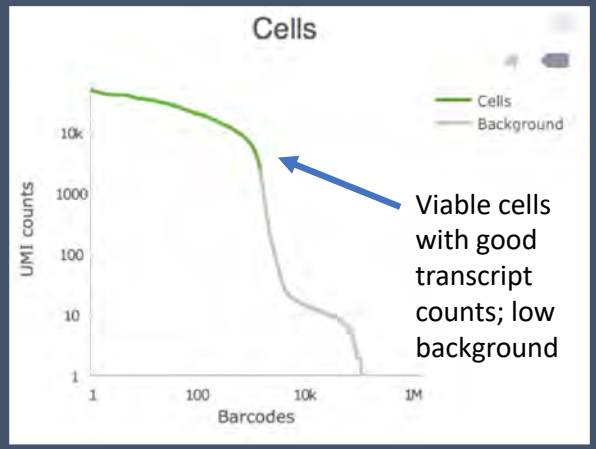
~10% cell viability &
no dead cell removal

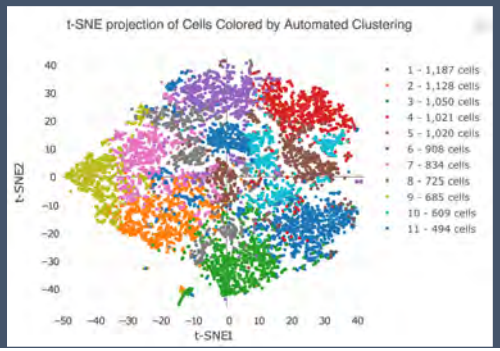




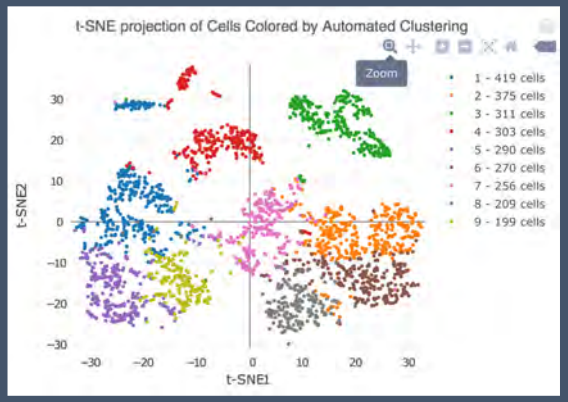


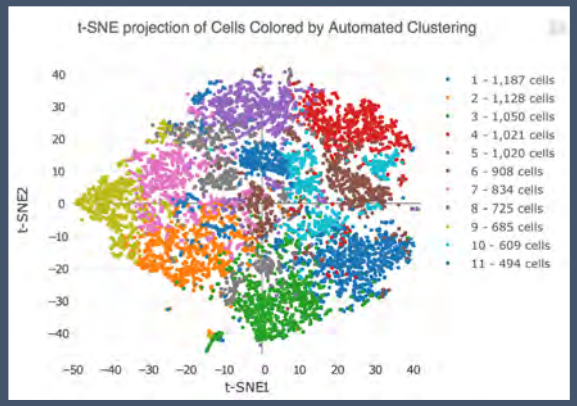
Dead cell removal ->
50% cell viability



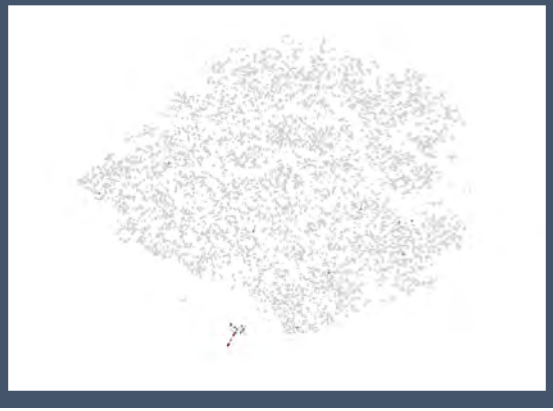


Any selection or manipulation can have the effect of biasing cell populations





CD45 Expression

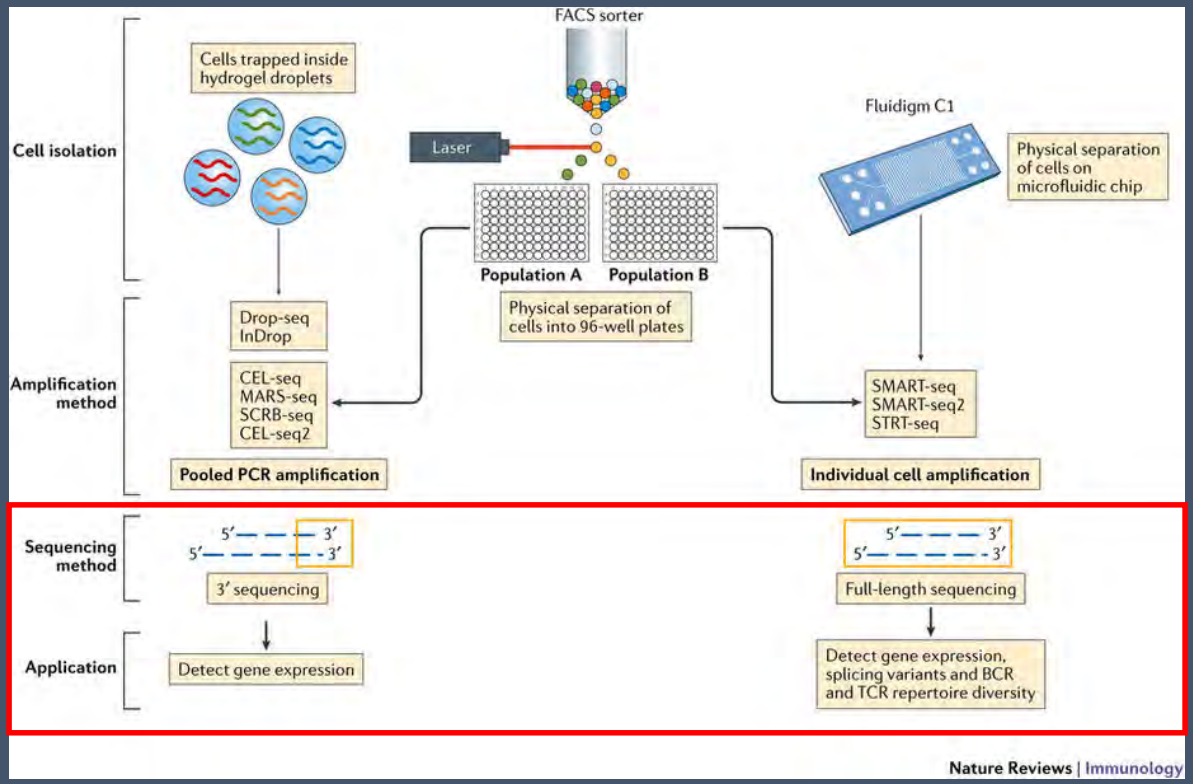


- Selection / enrichment is trade—off between ”cleaner” data and a potential to bias the data
- Alignment of cell rations with histology or flow cytometry can provide insight and confidence

Data wrangling

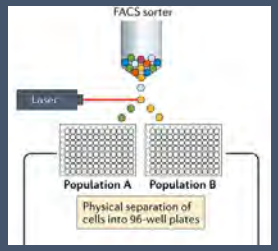
What single cell RNA-Seq looks like and why the reference is important

Type of downstream data depends on scRNA-Seq method - *data defined by cDNA library type*

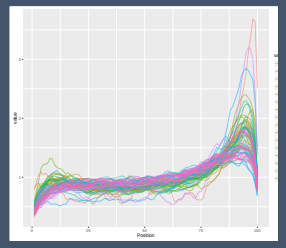


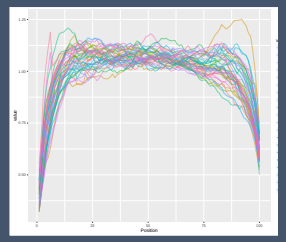
- ”Full-length” scRNA-Seq methods generate reads that can span entire transcript length – might not cover very 5’ or 3’ end
- 3’ or 5’ transcript end enriched scRNA-Seq for gene-level counts (5’ for VDJ methods)

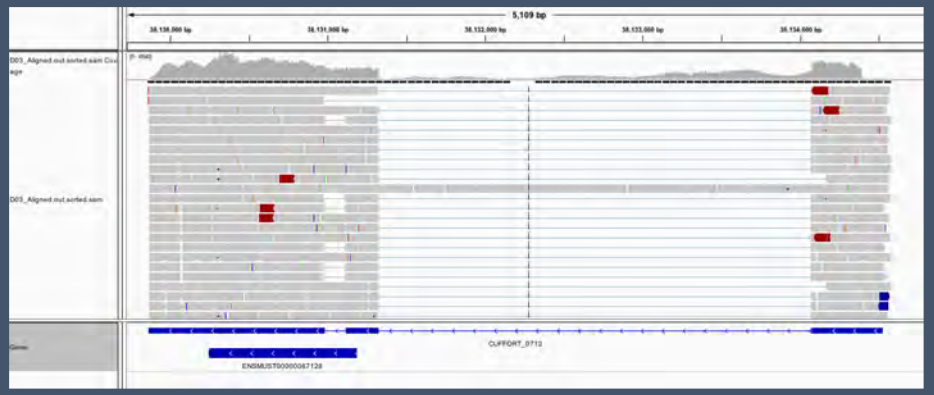
What does ”full-length” single cell RNA-Seq data look like?

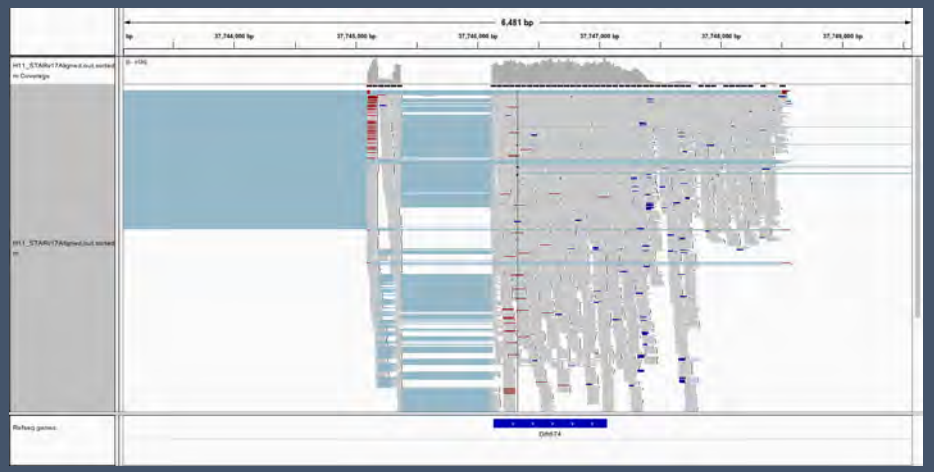


5’ to 3’ Coverage Plots

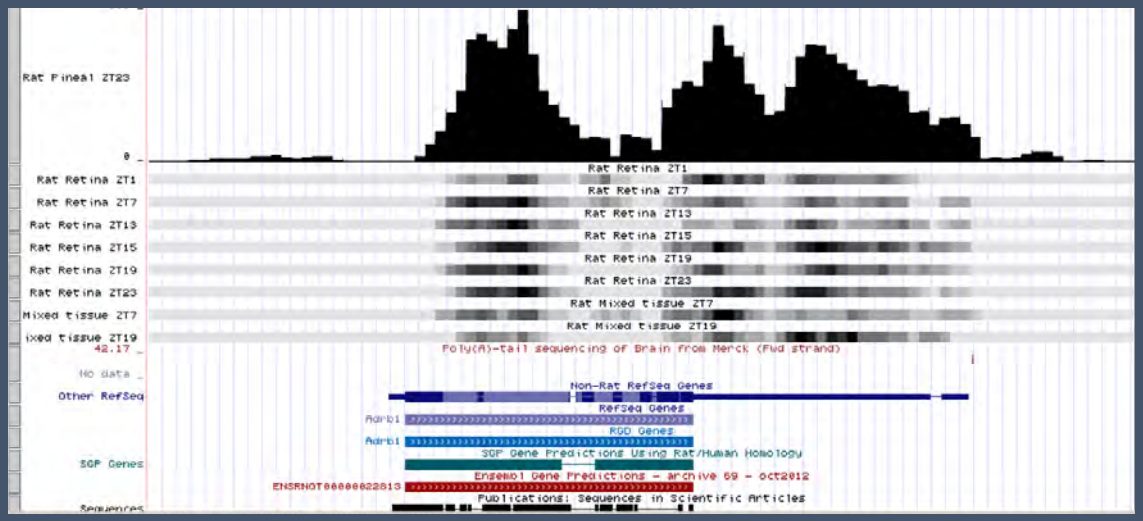




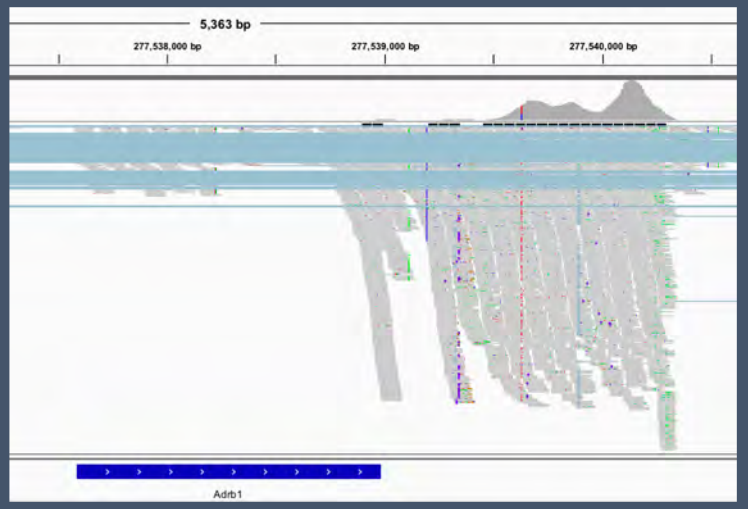




Public reference annotations don’t cover all the possible transcript isoforms and polyA sites



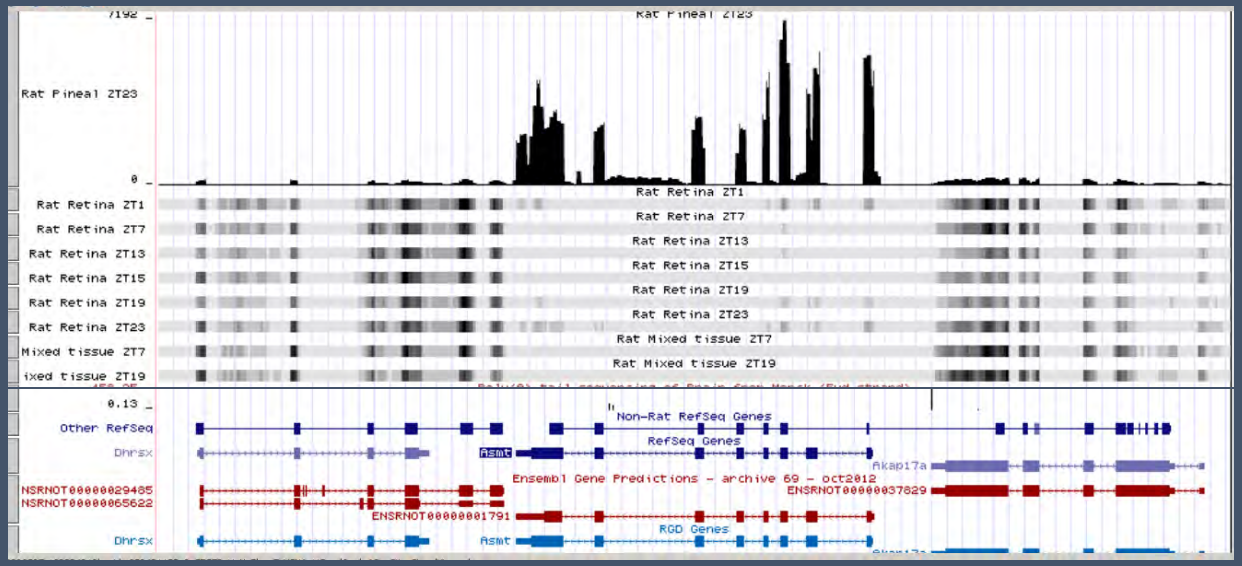
Bulk RNA-Seq Data

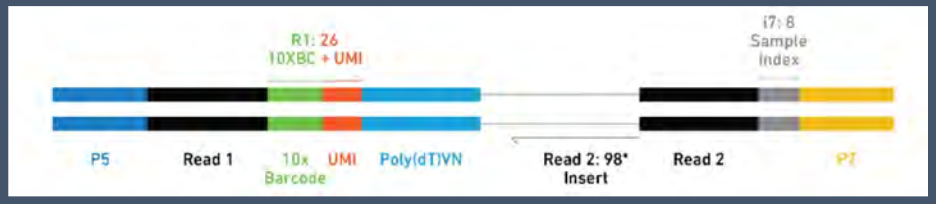


3’ Only scRNA-Seq

- Be aware that you might be missing some of your reads for less mature reference annotations
- 3’ (or 5’) end counting could exacerbate this?
- Cell-type specific rare exon usage of alternative polyA signal?

A complex genome adds some additional challenges – 3’ ends of genes can overlap

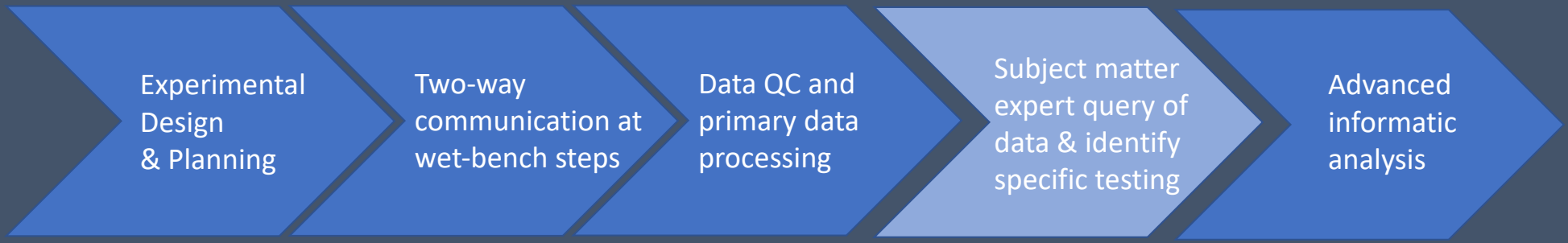




- “Unstranded” RNA-Seq reads aligning to overlapping 3’ end can’t be reliably disambiguated
- Interestingly, 3’-end only (or 5’-only) libraries are inherently “stranded”

A proposed model for integrated informatics support

Guidance, Training & Directed Informatics Support



Core Facility Embedded Bioinformatics

Basic level informatics-trained end-user

Bioinformatic support: Lab-embedded, third-party, or core-embedded

NCI Single Cell Analysis Facility as a collaborative team

(we are hiring a bioinformatic analyst!)

NCI Investigator Labs
(clinical & basic science)
& NIH Intramural Community

Broader Single Cell Community

- Sequencing Production
- Project Tracking
- R&D Innovation
- Analysis Workflows

Sequencing Facility & Genomics Tech Laboratory

Frederick National Lab Campus

Single Cell Analysis Facility

NCI Flow Core

NCI Genomics Core

Collaborative Bioinformatics & NCI Data Science Laboratory

Protein Analysis Core

NIH Bethesda Campus

NCI Single Cell Analysis Facility

NIH Clinical Center on Main Bethesda Campus

Bioinformatic Analyst will be embedded in our Single Cell Analysis Facility and will work closely with SCAF team & NCI Investigator

Acknowledgements

Laboratory of Cochlear Development

Matt Kelley
Joseph Mays
Alejandro Anaya
Mahmoud Khalil
Betsy Driver
Weise Chang
Kathryn Ellis
Hanna Sherrill
Tessa Sanders
Kuni Iwasa

NIDCD Genomics & Computational Biology Core

Robert Morell
Erich Boger

NIDCR Flow Cytometry Core

Mehrnoosh Abshari

Collaborators

Banfi Lab (Univ of Iowa)
Levine Lab (NINDS)
Friedman Lab (NIDCD)
Klein Lab (NICHD)

Decibel Therapeutics

Joe Burns
Adam Palermo
Kathy So

NCI Center for Cancer Research

Office of Science Technology Resources
NCI Genomics Core
Val Bliskovsky, PhD & Liz Conner, PhD

Frederick National Lab - Cancer Technology Research Program

Sequencing Facility & Genomic Technology Lab

Developers of open-source analysis packages

NIH High-performance computing staff

Cochlear scRNA-Seq work was funded by NIDCD Intramural Research Program - DC000021 (MWK) & Other Sources of Funding for Collaborators