Introduction to Single Cell Sequencing

Sarah Boswell

Director of the Single Cell Core, Harvard Medical School
Director of Sequencing Technologies, Laboratory of Systems Pharmacology

http://iccb.med.harvard.edu/single-cell-core
sarah_boswell@hms.harvard.edu
Outline

• Introduction to single cell RNA sequencing

• Overview of most relevant technologies

• Sample preparation

• Analysis
Bulk vs Single Cell RNA-seq (scRNA-seq)

**Bulk RNA-seq**
- Average expression level
- Comparative transcriptomics
- Disease biomarker
- Homogenous systems

**scRNA-seq**
- Population 1
- Population 2
- Population 3
- Population 4
- Define heterogeneity
- Identify rare cell population
- Cell population dynamics
Common applications of scRNA-seq

Studying heterogeneity

Lineage tracing study

Stochastic gene expression study

Heterogenous tissue

Cell differentiation

OFF state

ON state

Component1

Component2

Component1

Component2

Slow transition of promoter

Fast transition of promoter

Rel. freq.

mRNA copies

Bimodal distribution

Unimodal distribution


Junker and van Oudenaarden; Every Cell Is Special: Genome-wide Studies Add a New Dimension to Single-Cell Biology, Cell 2014 (doi: 10.1016/j.cell.2014.02.010)
Transcriptome Coverage (mRNA)

1. mRNA: TruSeq RNA-Seq (Gold Standard)
   • ~20,000 transcripts
     • More when consider splice variants / isoforms
   • Observe 80-95% of transcripts depending on sequencing depth

2. Low input methods ~3000 cells / well
   • 4000-6000 transcripts per sample
     • Limiting to transcripts observed across all samples
   • Observe 20-40% of the transcriptome

3. Single Cell Methods
   • 500 -10,000 transcripts per cell
   • Observe 10-30% of the transcriptome
   • Many transcripts will show up with zero counts in every cell. Even GAPDH.
   • If you only looked at transcripts observed in all cells numbers drop dramatically.
Comparison of Single Cell Methods

- CELL-Seq
- MARS-Seq
- SMART-Seq
- SCRB-Seq
- Chromium (10x)
- Seq-Well
- inDrops
- Drop-Seq
• Lysis and reverse transcription occurs in the beads
• Samples are frozen after RT as RNA:DNA hybrid in gel
• Library prep is based on CEL-Seq method

A. M. Klein et al., Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells, Cell 2015  (doi: 10.1016/j.cell.2015.04.044)
inDrops Method Overview

- Single cell suspension injected at density of ~80,000 cells / ml

- Matching the speed of bead injection with the speed of droplet generation it is possible to set conditions in which nearly every droplet would be loaded.
scRNA-seq Library Structure (inDrops)

Barcode uniquely identifies every cell (or well).

UMI is an N-mer that uniquely identifies every transcript.

<table>
<thead>
<tr>
<th>Read</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read1</td>
<td>61 bp Read 1 transcript</td>
</tr>
<tr>
<td>Read 2</td>
<td>8 bp Index Read 1 (i7) single cell barcode</td>
</tr>
<tr>
<td>Read 3</td>
<td>8 bp Index Read 2 (i5) library index</td>
</tr>
<tr>
<td>Read 4</td>
<td>14 bp Read 2 barcode/UMI</td>
</tr>
</tbody>
</table>
## Comparison of Single Cell Methods

<table>
<thead>
<tr>
<th></th>
<th>inDrops</th>
<th>10x</th>
<th>Drop-seq</th>
<th>Seq-well</th>
<th>SMART-seq</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell capture efficiency</strong></td>
<td>~70-80%</td>
<td>~50-65%</td>
<td>~10%</td>
<td>~80%</td>
<td>~80%</td>
</tr>
<tr>
<td><strong>Time to capture 10k cells</strong></td>
<td>~30min</td>
<td>10min</td>
<td>1-2 hours</td>
<td>5-10min</td>
<td>--</td>
</tr>
<tr>
<td><strong>Encapsulation type</strong></td>
<td>Droplet</td>
<td>Droplet</td>
<td>Droplet</td>
<td>Nanolitre well</td>
<td>Plate-based</td>
</tr>
<tr>
<td><strong>Library prep</strong></td>
<td>CEL-seq</td>
<td>SMART-seq</td>
<td>SMART-seq</td>
<td>SMART-seq</td>
<td>SMART-seq</td>
</tr>
<tr>
<td></td>
<td>Linear amplification by IVT</td>
<td>Linear amplification by IVT</td>
<td>Linear amplification by IVT</td>
<td>Linear amplification by IVT</td>
<td>Linear amplification by IVT</td>
</tr>
<tr>
<td><strong>Commercial</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>--</td>
<td>--</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Cost (~$ per cell)</strong></td>
<td>~0.06</td>
<td>~0.2</td>
<td>~0.06</td>
<td>--</td>
<td>1</td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
<td>• Good cell capture  • Cost-effective  • Real-time monitoring  • Customizable</td>
<td>• Good cell capture  • Fast and easy to run  • Parallel sample collection</td>
<td>• Cost-effective  • Customizable</td>
<td>• Good cell capture  • Cost-effective  • Real-time monitoring  • Customizable</td>
<td>• Good cell capture  • Good mRNA capture  • Full-length transcript  • No UMI</td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
<td>Difficult to run</td>
<td>Expensive</td>
<td>Difficult to run &amp; low cell capture efficiency</td>
<td>Still new!</td>
<td>Expensive</td>
</tr>
</tbody>
</table>

Experimental Decision Making

Is your sample fairly homogeneous?

- Yes: Bulk RNA-Seq
- No: scRNA-seq

Do you want full length transcripts/splice variants?

- Is our sample limited in cell number?
  - Yes: SMART-Seq
  - No: Droplet-based method
Single Cell Core Workflow

- Good sample prep is the key to success.
- A well planned pilot experiment is essential for evaluating sample preparation.
- Do not rush to the final experiment.
Key to Success: Sample Preparation

- High cell viability (>90-95% preferred), no free-floating RNA
- Single cell suspension
  - Enzymatic dissociation
  - FAC sorting
  - Mechanical dissociation
    - Collagenase, hyaluronidase, elastase, papain, trypsin etc. Use more gentle alternatives - accutase, liberase etc.
    - Use for enriching the population of interest. It might stress the cells, include “unsorted” sample control.
    - Gentle way to isolate cells.

- Dissociation protocol is cell type-dependent
- Cryopreservation or Nuc-Seq only works on some sample types
Sample Preparation: increasing cell viability

- Mild dissociation reagent (TrypLE, StemPro, Accutase, Liberase)
- Shorten dissociation time
- Reduce dissociation temperature
- Using ROCK inhibitor/ apoptosis inhibitor (esp. epithelial cells)
- Avoid cell pelleting
- Avoid FACS sorting on more fragile cell types
- Try magnetic activated cell sorting (MACS)
Sample Preparation: cell numbers

• Droplet methods have a 10,000-25,000 cell minimum
  • need ~50-100 cells with a unique transcriptome to identify a population cluster

• Count cells by hemocytometer – do not trust sorter counts
  • counts from the sorter are often ½ of actual cell counts

• Try negative selection to remove unwanted cells

• Sort on a broader marker to increase cell numbers

• For unavoidably low density samples
  • spike the sample with cells with distinct expression profile
Sample Preparation: single cell suspension

- Use appropriate cell strainer to get rid off clumps.
- Use metabolically inert, non-toxic density gradient media to fractionate cells.
- We routinely use 15% Opti-prep to keep cells in single cell suspension while loading sample, but the concentration might vary from cell to cell.
- Make sure final buffer does not contain calcium, EDTA, or heparin (inhibit RT)
Sample Preparation: buffers

- Suggest final sample prepared in:
  - PBS with 0.1-1% BSA
  - Defined media without calcium or EDTA

- 2% FBS in defined media without calcium or EDTA has worked for some users.

- Make sure final buffer does not contain calcium, EDTA, or heparin (inhibit RT)
Sample Preparation: viability checks

- **Check viability of sample over time**
  - If viability decreases over a short period of time this will be reflected in transcriptional data.
  - Will see high mitochondrial read counts.

- **Check single cell suspension supernatant for the presence of free floating RNA** (Ribogreen)
  - Creates background noise in all samples and complicates analysis.

- **Number of trypan positive cells \( \propto \) number of wasted sequencing reads**
  - If 30% of your cells are dead at the time of encapsulation then at most you will be able to use 70% of your sequencing data.
Sample Preparation: cryopreservation

- Several papers have come out using various cryopreservation techniques on samples (PBMC’s or cell lines).
- Success of cryopreservation is dependent on the sample type.
- Have seen this worked well on blood and immune cells.
- Key is the viability of the cells upon rehydration.
- Consider Nuc-Seq as an option from cryopreserved cells.

## Comparison of Single Cell Methods

<table>
<thead>
<tr>
<th></th>
<th>inDrops</th>
<th>10x</th>
<th>Drop-seq</th>
<th>Seq-well</th>
<th>SMART-seq</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell capture efficiency</strong></td>
<td>~70-80%</td>
<td>~50-65%</td>
<td>~10%</td>
<td>~80%</td>
<td>~80%</td>
</tr>
<tr>
<td><strong>Time to capture 10k cells</strong></td>
<td>~30min</td>
<td>10min</td>
<td>1-2 hours</td>
<td>5-10min</td>
<td>--</td>
</tr>
<tr>
<td><strong>Encapsulation type</strong></td>
<td>Droplet</td>
<td>Droplet</td>
<td>Droplet</td>
<td>Nanolitre well</td>
<td>Plate-based</td>
</tr>
<tr>
<td><strong>Library prep</strong></td>
<td>CEL-seq Linear amplification by IVT</td>
<td>SMART-seq Exponential PCR based amplification</td>
<td>SMART-seq Exponential PCR based amplification</td>
<td>SMART-seq Exponential PCR based amplification</td>
<td>SMART-seq Exponential PCR based amplification</td>
</tr>
<tr>
<td><strong>Commercial</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>--</td>
<td>--</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Cost (~$ per cell)</strong></td>
<td>~0.06</td>
<td>~0.2</td>
<td>~0.06</td>
<td>--</td>
<td>1</td>
</tr>
</tbody>
</table>
| **Strengths** | • Good cell capture  
• Cost-effective  
• Real-time monitoring  
• Customizable | • Good cell capture  
• Fast and easy to run  
• Parallel sample collection | • Cost-effective  
• Customizable | • Good cell capture  
• Cost-effective  
• Real-time monitoring  
• Customizable | • Good cell capture  
• Good mRNA capture  
• Full-length transcript  
• No UMI |
| **Weaknesses** | Difficult to run | Expensive | Difficult to run & low cell capture efficiency | Still new! | Expensive |

Doublets / Cell Density

- Rate of doublets depends on the cell density and the flow rate used for encapsulation.

<table>
<thead>
<tr>
<th>Multiplet Rate (%)</th>
<th># of Cells Loaded</th>
<th># of Cells Recovered</th>
</tr>
</thead>
<tbody>
<tr>
<td>~0.4%</td>
<td>~870</td>
<td>~500</td>
</tr>
<tr>
<td>~0.8%</td>
<td>~1700</td>
<td>~1000</td>
</tr>
<tr>
<td>~1.6%</td>
<td>~3500</td>
<td>~2000</td>
</tr>
<tr>
<td>~2.3%</td>
<td>~5300</td>
<td>~3000</td>
</tr>
<tr>
<td>~3.1%</td>
<td>~7000</td>
<td>~4000</td>
</tr>
<tr>
<td>~3.9%</td>
<td>~8700</td>
<td>~5000</td>
</tr>
<tr>
<td>~4.6%</td>
<td>~10500</td>
<td>~6000</td>
</tr>
<tr>
<td>~5.4%</td>
<td>~12200</td>
<td>~7000</td>
</tr>
<tr>
<td>~6.1%</td>
<td>~14000</td>
<td>~8000</td>
</tr>
<tr>
<td>~6.9%</td>
<td>~15700</td>
<td>~9000</td>
</tr>
<tr>
<td>~7.6%</td>
<td>~17400</td>
<td>~10000</td>
</tr>
</tbody>
</table>

- Easy to claim low doublet rate for a particular method, but be aware of these parameters to accurately assess that.

CRISPR Transcript Specific Library

- CRISPR pool vector backbone must contain a transcribed poly-adenylated unique guide index (UGI), which can include a fluorescent marker

- scRNA-seq library to phenotype cellular transcriptome (NextSeq/HiSeq)
- gRNA-targeted library to ensure proper cell identification (MiSeq)

A. Dixit et al., Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens, Cell 2016 (doi10.1016/j.cell.2016.11.038)
inDrops Library Prep

- Second strand synthesis to make full length dsDNA (store -80 C)
- In vitro transcription (IVT) back to RNA off T7 promoter from primer
- RNA Fragmentation
- RT with random hexamer primer containing adaptor
- PCR off adaptors to add index and illumina adaptors
inDrops Transcript Specific Library Prep

- Post-IVT sample

- Perform RT with gene specific primer (external primer)

- PCR amplify with gene specific primer internal to previous primer and containing part of adapter

- Final amplification to add on rest of adapter
inDrops Transcript Specific Library Prep

- Make standard library
  - NextSeq/HiSeq sequencing to identify cell barcodes in sample

- Make transcript specific library with aliquot of initial library
  - MiSeq to identify cells with transcripts of interest

- Match barcodes identified in both sequencing runs
Single Cell Data Analysis

1. Identify cell barcodes and UMI
2. Correct and filter barcodes
3. Align with Kalisto
4. Transcript compatibility counts
5. Quality control and clustering
6. Align with Rapmap
7. Scaled counts per transcript
8. Downstream Analysis

Provided by the Harvard School of Public Health Chan Bioinformatics Core
Data Analysis: Quality Control (QC) metrics

• Reads per cell: How many reads assigned to a given cell barcode

• UMI per cell: “Novelty” score looks for greater diversity genes per UMI

• Genes detected: Genes with a non-zero count measurement per cell

• Mitochondrial counts ratio: Biomarker for cellular stress
Data Analysis QC: reads per cell

- Cut off usually remove any cell with < 10,000 / 20,000 barcodes per cell
- It is normal for single cell RNA-seq data to contain a large number of low complexity barcodes.
- Exact threshold will depend on sample
Data Analysis QC: reads per cell

- Bimodal peak is due to sample type. (infiltrating immune cells in tumor)
- Lower peak gets filtered out in analysis.

Provided by the Harvard School of Public Health Chan Bioinformatics Core
Data Analysis QC: genes detected per cell

- Genes “detected” means genes with a non-zero count measurement per cell.
- Gene detection in the range of 500-5000 is normal for most single-cell experiments.

Provided by the Harvard School of Public Health Chan Bioinformatics Core
Data Analysis QC: mitochondrial counts ratio

- Cells assessed for mitochondrial content
- High mitochondrial content cells are removed at a set threshold from analysis (<0.1)

Provided by the Harvard School of Public Health Chan Bioinformatics Core
Data Analysis QC: novelty score

- Assessment of genes per UMI in aggregate
- Higher novelty means more unique genes per unique UMI (>0.8)

Provided by the Harvard School of Public Health Chan Bioinformatics Core
Data Analysis: filtering & correction

- Remove cells with high mitochondrial RNA
- Remove cells with abnormally low or high genes detected
- Correct total UMI
- Correct genes detected
- Correct/filter low complexity

Filter parameters:
- \( \geq 500 \) UMI counts per cell
- \( \geq 500 \) genes per cell
- \( \leq 5000 \) genes per cell
- \( \leq 0.1 \) relative mitochondrial abundance
- \( \geq 0.8 \) novelty score

Provided by the Harvard School of Public Health Chan Bioinformatics Core
Data Analysis: filtering & correction

- Libraries were of 3,000 cells. Post-filtering retains 50-80% of cells

Provided by the Harvard School of Public Health Chan Bioinformatics Core
Data Analysis: group identification

• tSNE projection of WT and mutant zebrafish kidney marrow
• Clearly see changing populations of cells based on genotype

Q. Tang et al., Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing, JEM 2017 (doi: 10.1084/jem.20170976)
Final thoughts on scRNA-seq

• Practice your sample prep protocol. KEY to SUCCESS
• Start with a pilot sample set to ensure your protocol is working.
• Do not make your scRNA-seq run day the first day you run through the whole protocol.
• Be sure sequencing core understands the specific sequencing parameters needed for your scRNA-seq library.

qPCR

Precise quantitation is key to good clustering / sequencing
Acknowledgements

Single Cell Core
Mandovi Chatterjee
Alex Ratner

Klein Lab
Allon Klein
Rapo Zilionis

ICCB-Longwood
Caroline Shamu
Jennifer Smith
Katrina Rudnicki
Mirela Vaso

Tools and Technology Committee
Contact / References


http://iccb.med.harvard.edu/single-cell-core
sarah_boswell@hms.harvard.edu