Garvan Institute of Medical Research

Single cell experimental design and hypothesis testing

Associate Professor Joseph Powell
Director, Garvan-Weizmann Centre for Cellular Genomics
Deputy Director, UNSW Cellular Genomics Futures Institute

SISG - 2019

Garvan Institute of Medical Research

Key Topics

Garvan Institute of Medical Research

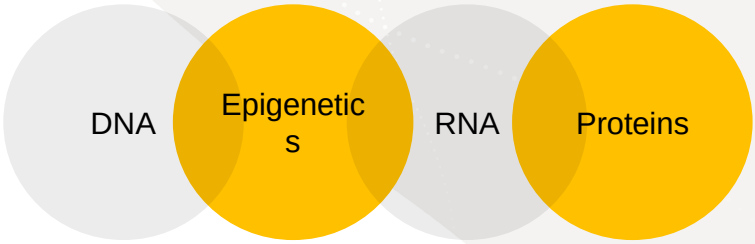
- Introduction to the range of techniques for generating scRNA-seq data
- Key challenges and considerations
- What analysis questions can I address?

3

Cellular genomics

Garvan Institute of Medical Research

Science
Breakthrough technology of 2018

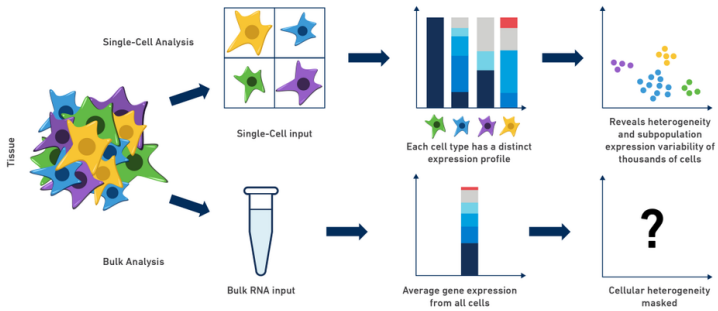


Cellular genomics technology allows you to measure the genomic contents for an individual cell. But at 10,000s of cells at a time.

4

Why single cells?

Garvan Institute of Medical Research



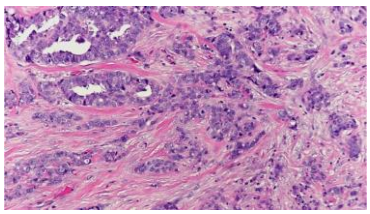
Do we expect cells to have the same transcriptional signatures?

Bulk methods give an average RNA signal

5

A typical challenge that we might face

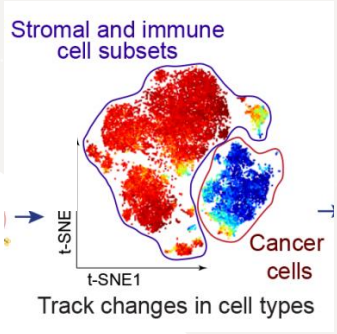
Garvan Institute of Medical Research



A tumor biopsy contains many cell types
- Tumor cells (with different mutations)
- Healthy non-tumor cells
- Immune cells
- Nerves
- Blood vessels
- ….

Traditional genomic assays average the signals across all cells

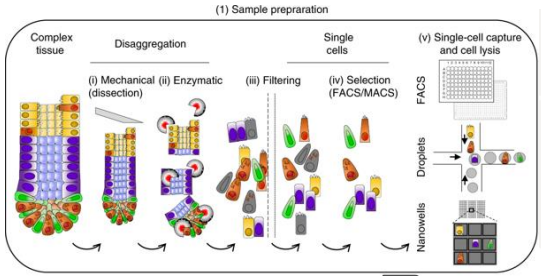
Single cell sequencing enables complete resolution of cell types and their genomic functions



6

Typical workflow

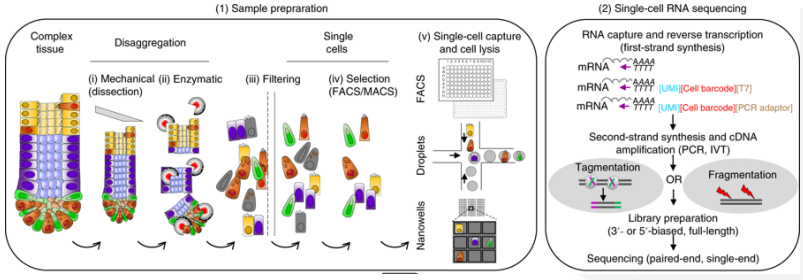
Garvan Institute of Medical Research



7 Lafzi et al. Nature Protocols 2018

Typical workflow

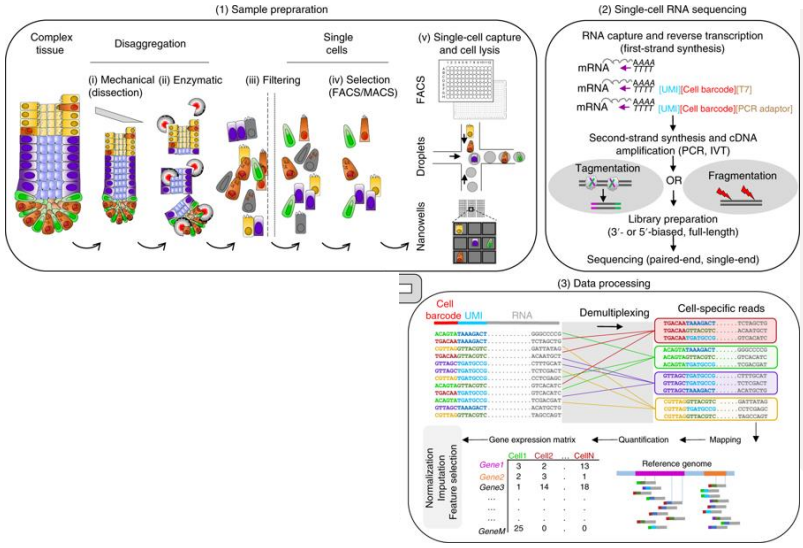
Garvan Institute of Medical Research



8 Lafzi et al. Nature Protocols 2018

Typical workflow

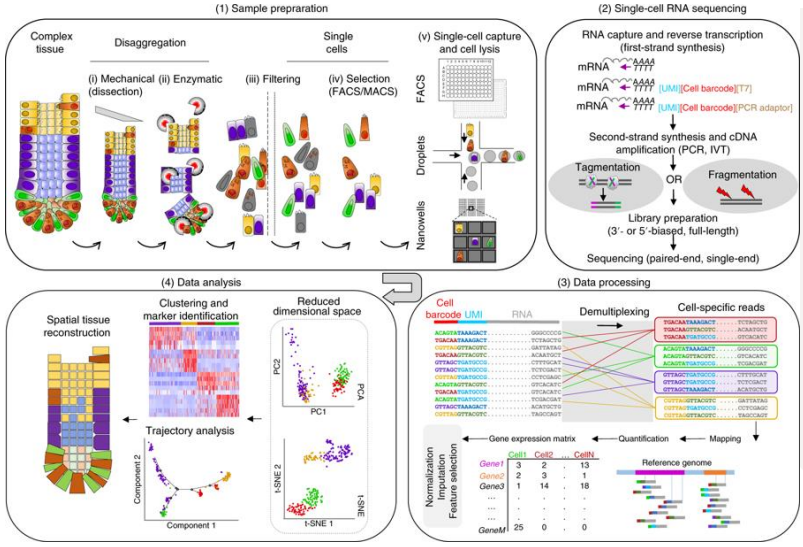
Garvan Institute of Medical Research



9 Lafzi et al. Nature Protocols 2018

Typical workflow

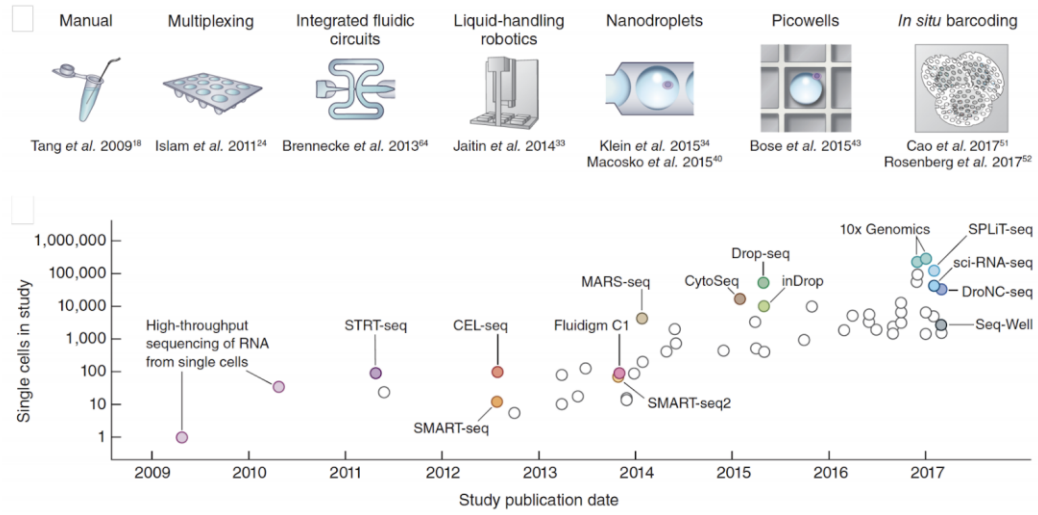
Garvan Institute of Medical Research



10 Lafzi et al. Nature Protocols 2018

Technology has progressed rapidly

Garvan Institute of Medical Research



11

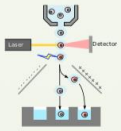
Cell capture defines methods

Garvan Institute of Medical Research

Microfluidic device

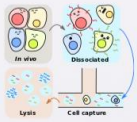
- 96 or 800 well format
- Physically check presence of cells
- High capture efficiency
- Doublet issues
- Expensive
- Full-length cDNA (SMART-seq{2})
- Spike-in control RNA
- **High gene coverage**

Plate-based



- 96 or 384 well format
- Sort specific population(s) of cells
- High capture efficiency
- Experimental design considerations
- Full-length cDNA (SMART-seq(2) or end-tagging; UMIs)
- Spike-in control RNA
- **High gene coverage**

Droplet-based



- 100-1000's of cells
- Doublet issues
- Variable capture efficiency
- Low per-cell cost
- 3' end tag; UMIs
- No spike-in control RNA
- **High cell coverage**

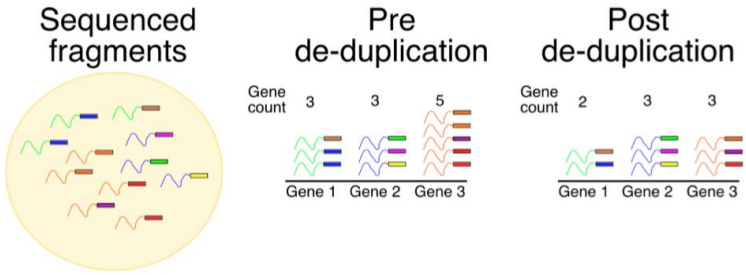
12

What are UMIs?

Garvan Institute of Medical Research

Unique molecular identifiers give (almost) exact molecule counts in sequencing experiments.

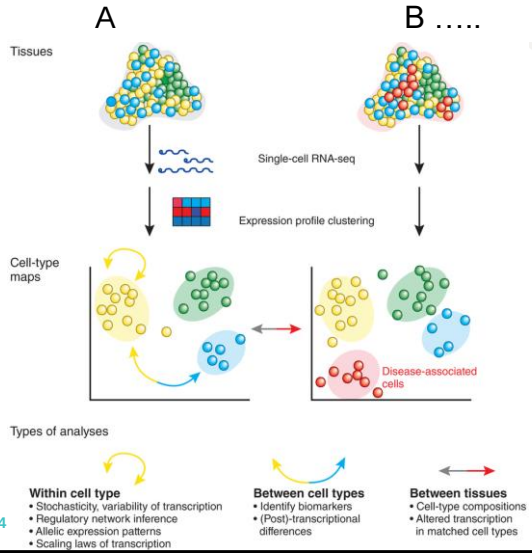
They reduce the amplification noise by allowing (almost) complete de-duplication of sequenced fragments.



13

What questions need to be considered in the experimental design?

Garvan Institute of Medical Research



How common / rare are the cells in am most interested in?

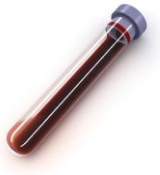
How different are they likely to be from other cell types?

Can the ‘phenotype’ be detected By RNA differences?

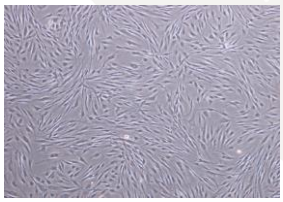
14

Considerations for single cell experiments
Cell Prep

Garvan Institute of Medical Research



Liquid



cell cultures

solid tissue

How much starting material is there?
Human/non-human

The critical point is **getting single cells *and keeping them that way***

Complete dissociation is required

Cell Strainers

Sorting – FACS

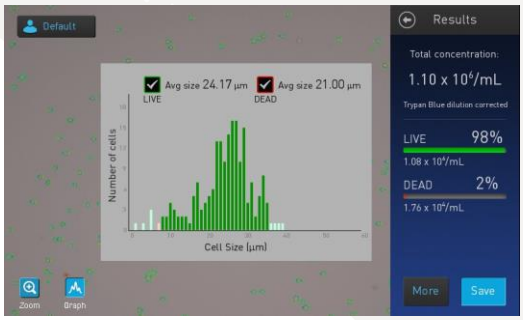
Magnetics beads

15

Considerations for single cell experiments
Cell Prep

Garvan Institute of Medical Research

- How big are they?
- How sensitive are they?
  - Handling (e.g. pipetting force)
  - Enzymatic dissociation (timing and harshness)
  - FACS pressure/nozzle gauge
  - Post-dissociation viability?
- Will they lyse in the reaction buffer?
- Are they sticky/liable to clump?

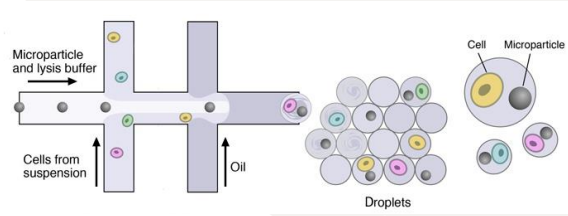


16

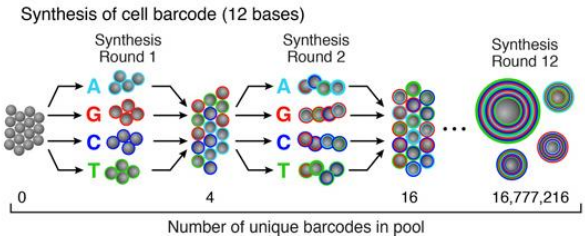
Barcoding

Garvan Institute of Medical Research

Microfluidic



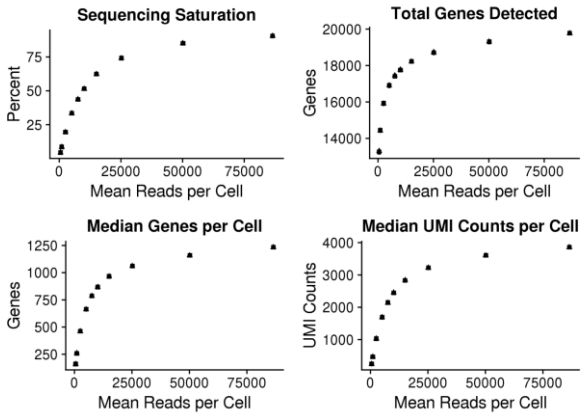
Barcodes



17 Makosco et al Cell 2015

How deep should I sequence?

Garvan Institute of Medical Research



Costs

Per 10,000 reads* per cell

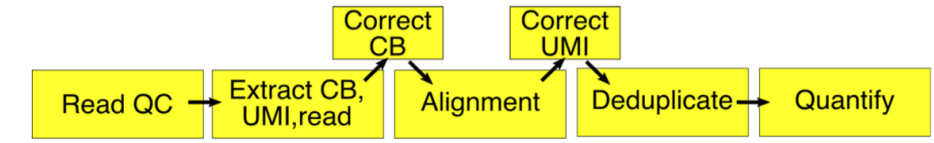
NextSeq	5-10 cents
HighSeq2500	10-15 cents
NovaSeq S2	3-5 cents
NovaSeq S4	2-4 cents

* With 3' scRNA-seq

18

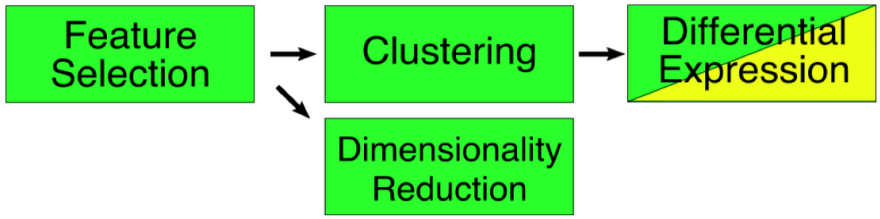
Bioinformatics workflow

Garvan Institute of Medical Research



Generally run in a single pipeline, e.g. Cellranger (10X specific), DropSeq (Macosko *et al.*) or custom (not recommended if just starting).

Single-cell specific tools (generally performed in R; Practical 1)



19

Analysis

Garvan Institute of Medical Research

Starting with a counts matrix:

Cell	A	B	C	D
Gene X	10	20	30	40
Gene Y	15	30	45	60
Gene Z	20	40	60	80
	…	…	…	…

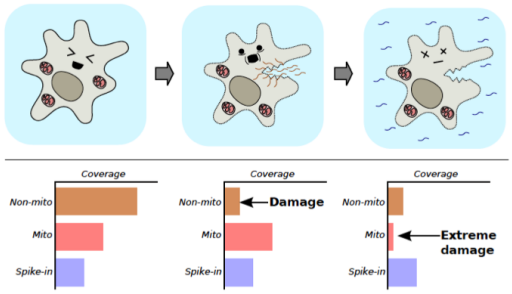
- Quality control
- Normalization
- Batch correction [if required]
- Dimensionality reduction and visualisation (part 2)
- Clustering (part 2)
- Differential expression testing (same as bulk RNA seq… mostly)

20

Dealing with low quality cells

Garvan Institute of Medical Research

- Low sequencing depth
- Low numbers of expressed genes (i.e. any non-zero count)
- High spike-in (if present) or mitochondial content

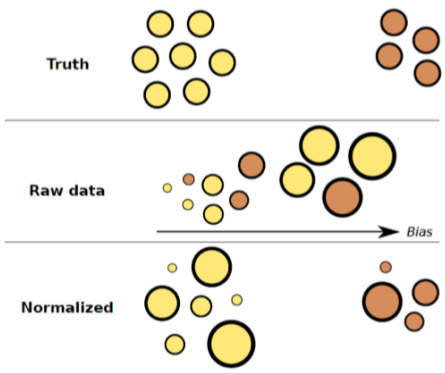


21 From Aaron Lun

Normalization

Garvan Institute of Medical Research

- The aim is bring all cells onto the same *distribution* to remove biases between them
- We want to preserve biological variability, not introduce new technical variation
- Primary source of bias is sequencing depth – scale down counts accordingly
- Need a method that is robust to sparsity and composition bias
  - TMM & DESeq size factors are not!



22 Image from Aaron Lun

Confounding and batch effects

Garvan Institute of Medical Research

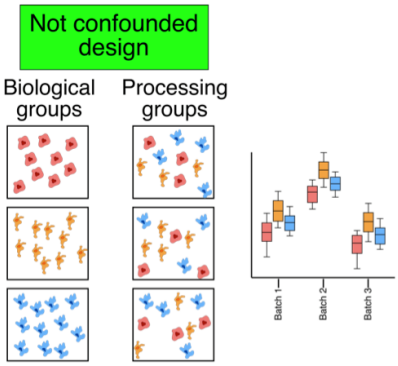
- A segue into proper experimental design
- Some batch effects cannot be avoided
- Some can, make sure you know which is which

A simple batch correction is
To use linear regression

Providing your biological groups and processing groups are not confounded.

But these methods assume most cells are transcriptionally the same.

We will discuss more complex normalization later



23

Resources

Garvan Institute of Medical Research

Single Cell Resources:
Single cell course (Hemberg Lab; Wellcome Sanger Institute):
http://hemberg-lab.github.io/scRNA.seq.course/index.html

Aaron Lun's single cell workflow (very detailed):
https://www.bioconductor.org/packages/release/workflows/html/simpleSingleCell.html

Cellranger pipeline:
https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger

Thank You!

24

Post Doc positions

Garvan Institute of Medical Research

Interested in a single cell postdoc?

My lab is recruiting 3 positions:

- Population genetics at single-cell resolution
- Clinical translation of single-cell seq methods
- Modeling common disease with stem cells at single-cell resolution

Advert on Garvan Jobs

25 https://www.garvan.org.au/research/cancer/computational-genomics