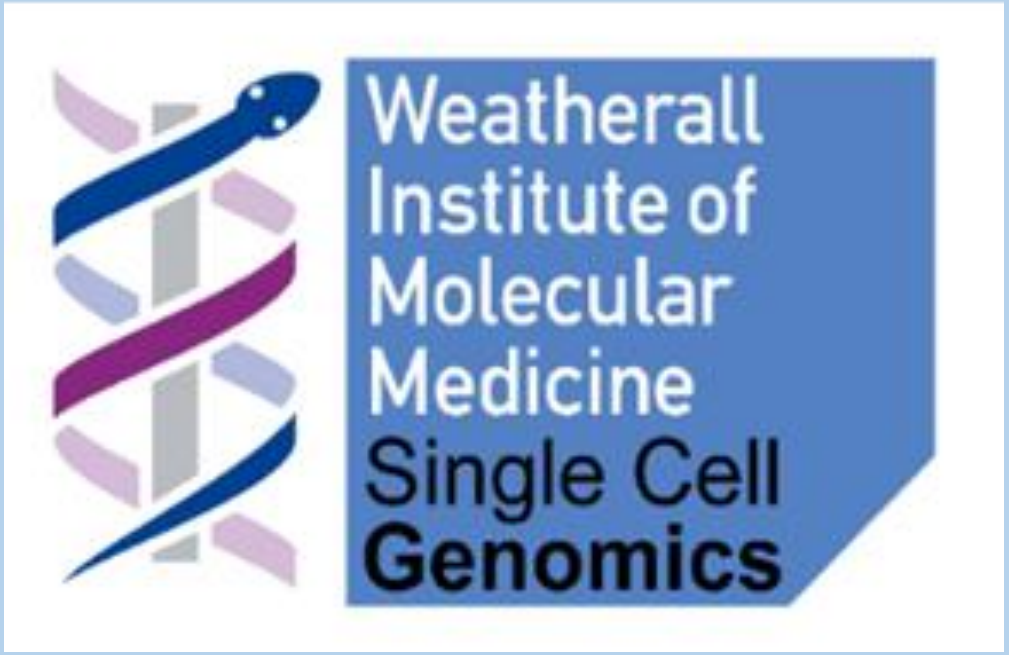
Single Cell Genomic Techniques



Dr. Neil Ashley

Single Cell Genomics Core

Weatherall Institute of Molecular Medicine

What Can We Get from Single Cell Methods ?

- Gene expression, RNA splicing, small RNA, circular RNA.
- Genomic Sequencing e.g. DNA mutations, genotypes, copy variation.
- Epigenomic Information e.g. methylation, chromatin conformation, Hi-C.
- Proteomics – protein expression, protein modification.
- Combinations of the above all from the same cell.

Current Single Cell RNA Seq Methods – there is a lot!

3’ Prime End Transcript Sequencing:

SC3-Seq
InDrop Seq*
Drop-Seq*
Mars-SEQ
SCRB-SEQ
CelSeq2/SORT-SEQ
Quartz-Seq
BAT-Seq
Hi-SCL*
CytoSeq**
Chromium (10x Genomics)*
Wafergen/WellSeq RNAseq**
Precise Assay (**Becton Dickinson**).
SureCell™ WTA 3‘* (**BioRad/Illumina**)
C1 high-throughput IFC (**Fluidigm**)***
*Droplet
**Nanowells
***Nanofluidic traps

Full Transcript Sequencing:

SmartSeq2
STRT-SEQ
SmartSeqV4 (Clonetech)
Phi29 polymerase WTA (Qiagen)

5’ Prime End Transcript Sequencing:

5’ Arguel Assay (2017).

Formaldehyde Fixed Tissue RNA Seq:

FRISCR (Smartseq2)

Circular RNA/Small Non-coding RNA Sequencing:

SUPeR-Seq
scSmall-RNA sequencing

Spatial/InSitu RNA Seq:

CLAP (Biotination of cells)
FISSEQ (in situ sequencing)
TIVA (in cell RNA capture)
PATCH-Seq (SmartSeq2 in patch clamp pipettes)

Basic Steps for Single Cell Genomics

1. Do I need single cell analysis and if I do how many do I need?
2. Isolate single cells
3. Amplify single cells
4. Prepare amplified material for sequencing
5. Sequence single cell libraries.
6. Analyse the data

1. Do I Need Single Cell Genomics?

- Cell populations of interest cannot be experimentally purified from bulk data e.g. stemcells
- My disease of interest is heterogeneous e.g. clonal mutants in cancer or T-clones.
- I’ve done bulk RNAseq and I don’t see any differences between treatment groups (I don’t know which cell type is involved).
- I’m interested in cell lineages.

Many diseases are heterogeneous at the single cell level even when genetically identical.

Mitochondrial DNA (small dots) and nuclear DNA (large blobs) in cultured human patient fibroblasts. (G) Healthy Control Fibroblasts. (H + I). Fibroblasts from patients with mitochondrial DNA depletion syndrome. *Asterisks show cells with mosaic depletion of mitochondrial DNA (Ashley et al, Human Molecular Genetics, 2008).

So how many cells do I need?

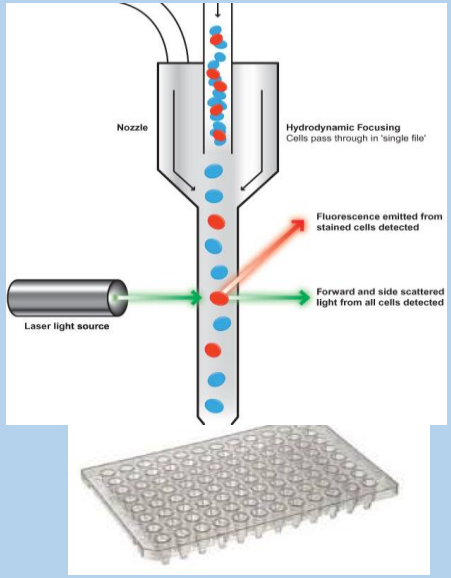
- 500 reads per gene and a minimum coefficient of variation (CV) of 4% under the assumption of the Poisson distribution = 50 single cells need to be pooled in order to achieve this minimum CV value.
- Technical drop outs and noise (biological and technical) usually mean we need lots more.
- What are the cells of interest? Do they have lots of RNA?
- How frequent are my cells of interest?
- How many cells can I get?
- Which technique am I using?
- How many can I afford to process and sequence?

2. Isolating single cells

Method	Description	Advantages	Disadvantages	Number of cells	Examples
Direct Pipetting	Glass pipettes or automated micropipettes.	Good for rare cells or very large cells	Hard to do. Not many cells.	10s-100s	Embryology
Serial Dilution	Dilute cells until you get 1 a drop on average.	Very cheap, pretty easy.	Wastes cells, relies on poison distribution, lots of empty wells	10s	?
Microfluidic traps	Cell capture in microfluidic chips	Cells can be imaged and treated.	Difficult to optimise, expensive, size dependent, doublets	10-100s	Fluidigm C1
Nanowells	Cells enter tiny wells by gravity or are pipetted in.	Can be cheap or expensive. Gentle.	Poisson cell distribution, might need imaging	100s-1000s	Takara Icell, CellSee, SeqWell
Microfluidic Droplets	Cells are encapsulated in oil in emulsion drop.	Easy, versatile	Expensive, limit to cell size, cell doublets can be problem	100s-10000s	10X Genomics Chromium, DROP-SEQ
Optical Tweezers	Light is used to move cells around	Gentle, versatile	Very expensive	10s-1000s	Berkerley Lights
FACS sorting cells or nucleii	Flow cytometric distribution into plates or tubes	Very versatile, low doublets, coupled to protein.	Expensive, processing is intensive without cell barcoding.	1-1000s	SmartSeq2, CelSeq2

Examples of Single Cell Capture Techniques

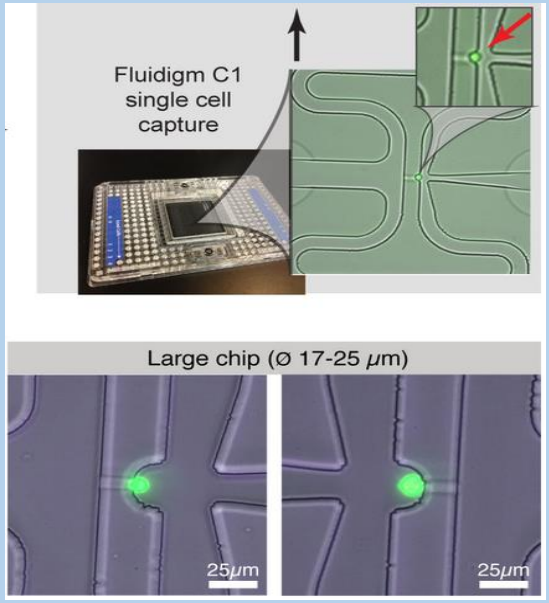
Flow Sorting into Plate or Tube



Pros-
Very common technique
Low cost consumables
Index sorting
Very flexible (Colours/cell numbers etc)
Very reliable single cell sorting
Only needs a 200-1000 cells.

Cons –
Relatively low throughput
Time consuming
Relatively big volumes = more cost

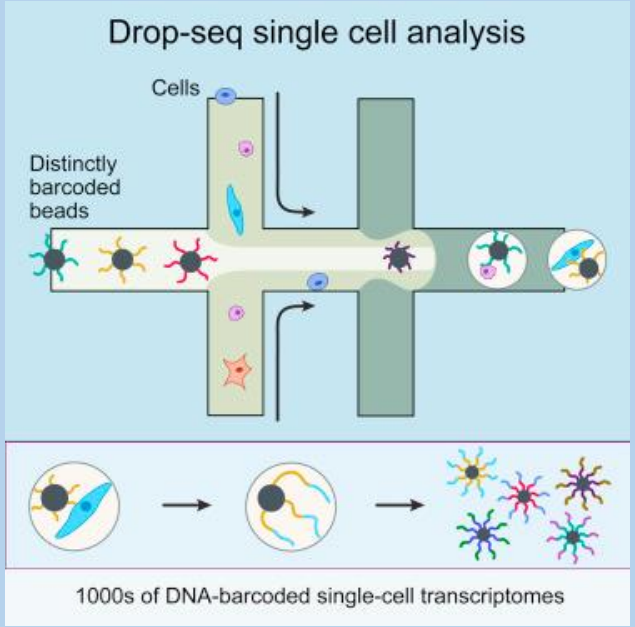
Fluidigm C1/Polaris



Pros-
Very low volumes = low cost reagents
Allows imaging of cell

Cons –
Relatively low throughput (max 800 cells)
Expensive chips
Can get lots of empty wells
Doublets can be an issue
Cell size dependent
Usually requires a imaging step to check viability/capture efficiency.
No index sort.
Needs at least 900 cells.

Droplet based: DropSeq/Chromium/BioRad etc



Pros-
Very high throughput (80,000 cells in one go)
Relatively easy processing as samples are pooled early.
Very quick.

Cons –
Expensive reagents
Expensive to sequence so many cells.
No inherent viability test
No index sort.
Needs at least 1500 cells.
No spike ins.

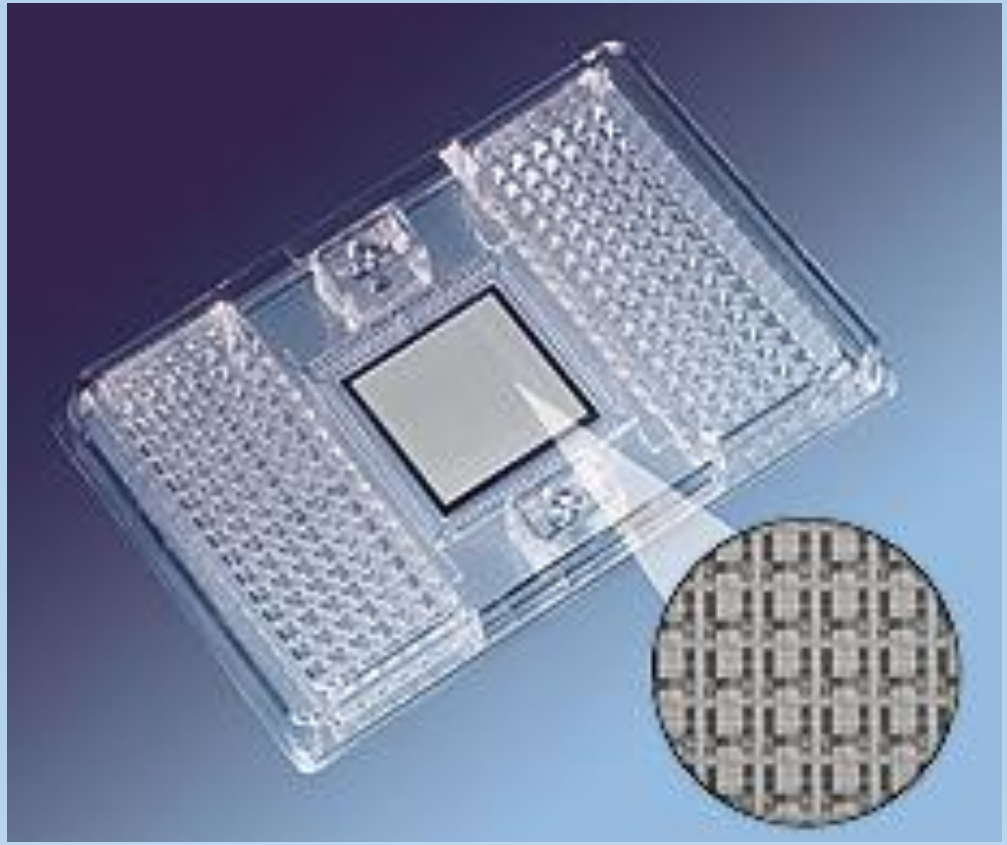
3. Amplification

- There is only 6-10 pg of DNA per cell and 10–30 pg total RNA per cell.
- This needs to be amplified to be able to sequence feature of interest.
- Amplification is most commonly done by PCR, which is generally targeted to poly A transcripts or individual genes.
- Poly A primed cDNA is amplified by PCR for a whole exome approach, or targeted genes can be amplified in parallel using microfluidic platforms.
- Alternatively the whole genome or whole exome can be amplified using isothermal enzymes e.g. Phi29.

Targeted QRT-PCR Analysis of Single Cells using Fluidigm Biomark

Targeted approaches are generally cheaper than whole exome/genome sequencing.

A Biomark 96:96 microfluidic chip for gene expression. Other chip formats are available e.g. 48:48



1. Sort and Lyse cells in PCR plate

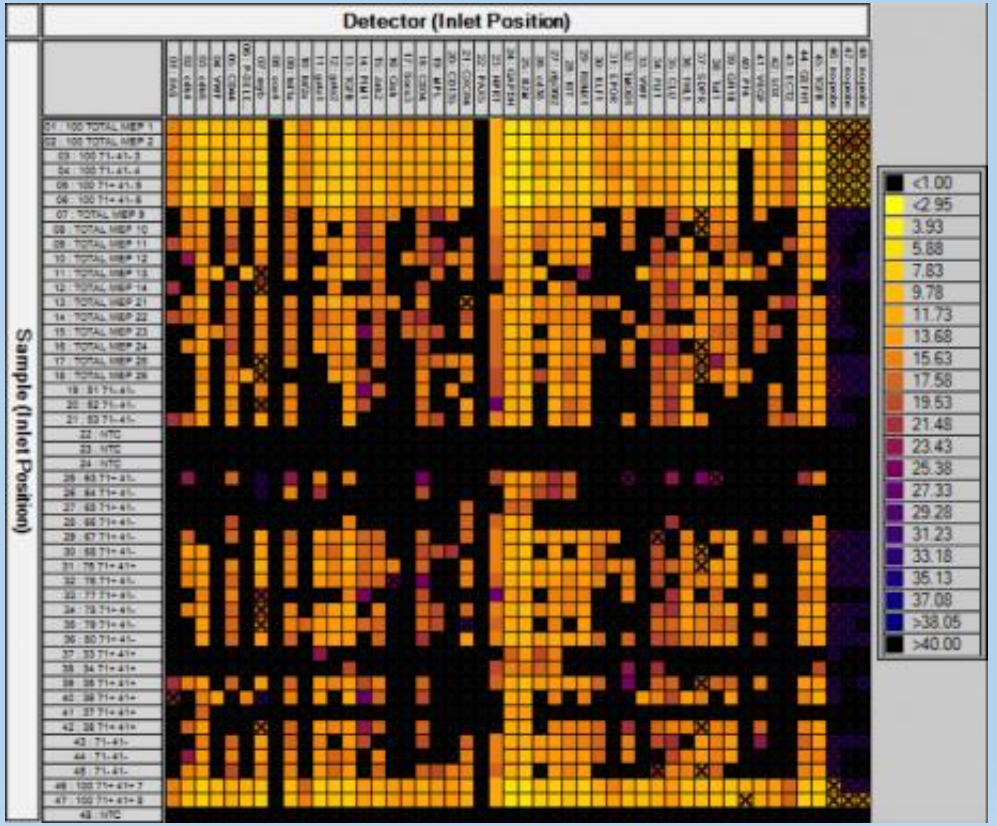
2. RT-PCR with 96 Taqman probe cocktail to generate cDNA.

3. Load cDNA onto chip with 96 probes in individual wells

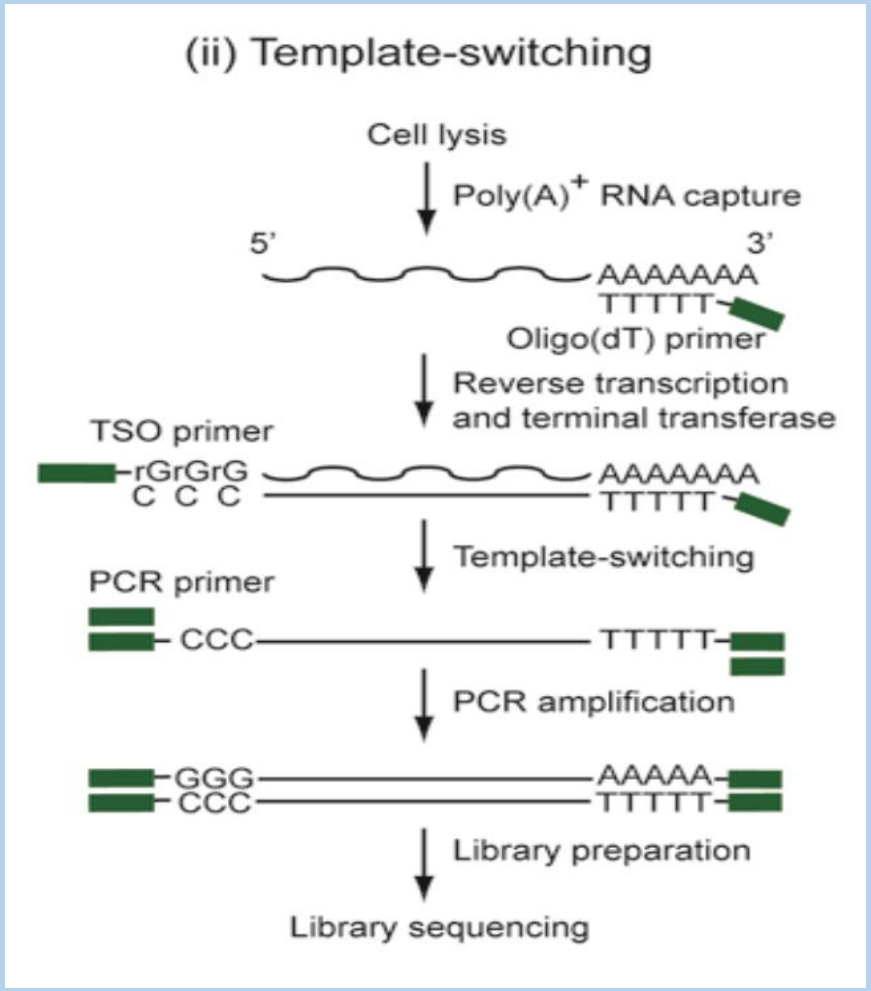
4. cDNA is mixed with probes in tiny chambers.

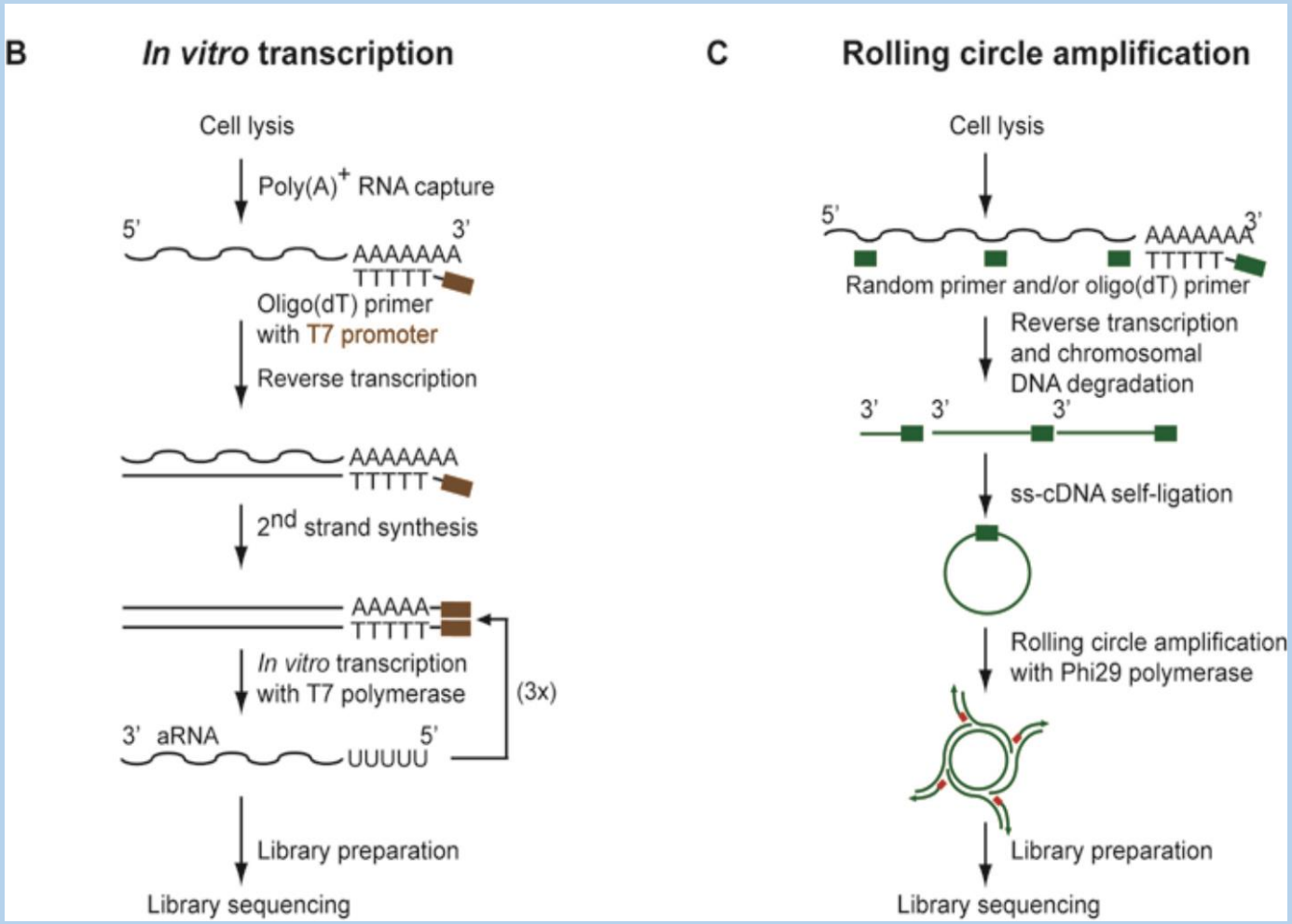
5. QPCR on Biomark

Biomark Data of 96 genes in 96 Single Cells using 96:96 chip.



Common Approaches for RNA/DNA Amplification of Single Cells





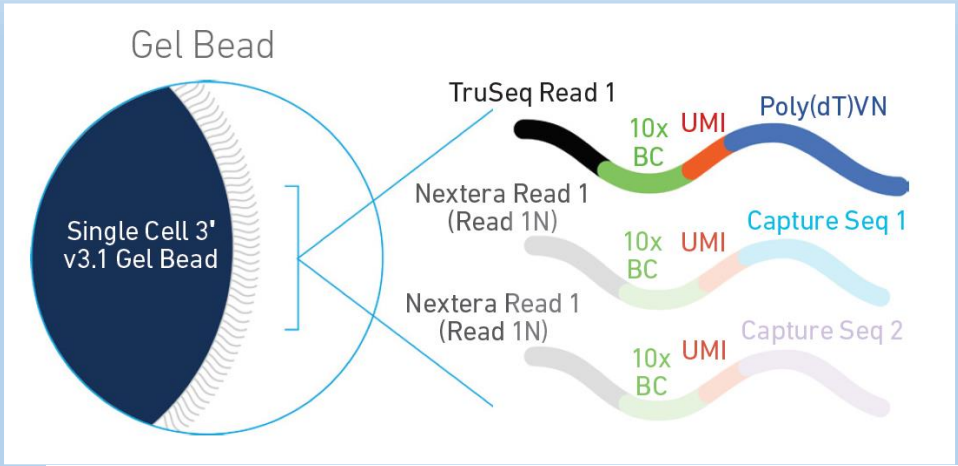
Existing methods to prepare sequencing libraries from a single cell. (A) (i) poly(A) tailing at the 3′ end of the first cDNA strand. (ii) template-switching mechanism involving a template-switching oligonucleotide (TSO). (B) During cDNA synthesis a T7 RNA polymerase primer can be integrated that enables the further amplification of RNA via in vitro transcription. (C) Alternatively, after cDNA synthesis and chromosomal DNA degradation, cDNA can be circularized and amplified by a technique referred to as rolling circle amplification using the Phi29 polymerase.

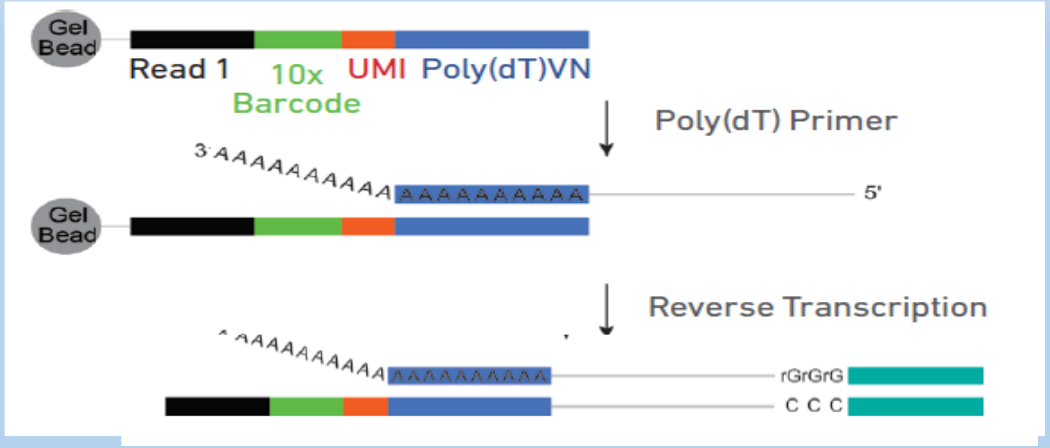
Nucleic Acids Res. 2014;42(14):8845-8860. doi:10.1093/nar/gku555

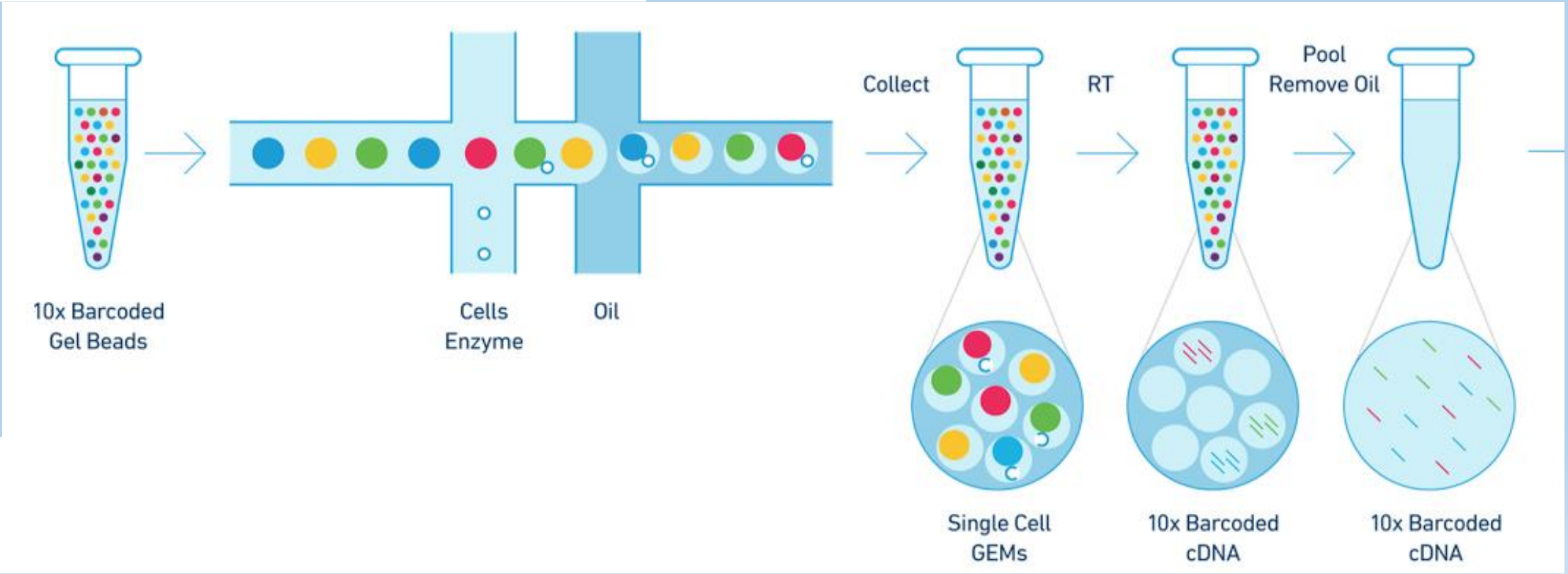
Cell Barcoding and Unique Molecular Indexes (UMI)

- Amplification chemistry is not dependent on isolation method as similar chemistry is used by different isolation methods.
- However samples need to be barcoded at some point to enable pooling.
- Barcodes can be added at different points in the protocol, usually by adding barcoded primers.
- But the earlier the barcoding in the method the easier and cheaper subsequent processing is as samples can be pooled and the more samples can be processed.
- UMIs can be added along with the cell barcode.

Droplet Based DNA/RNA Amplification

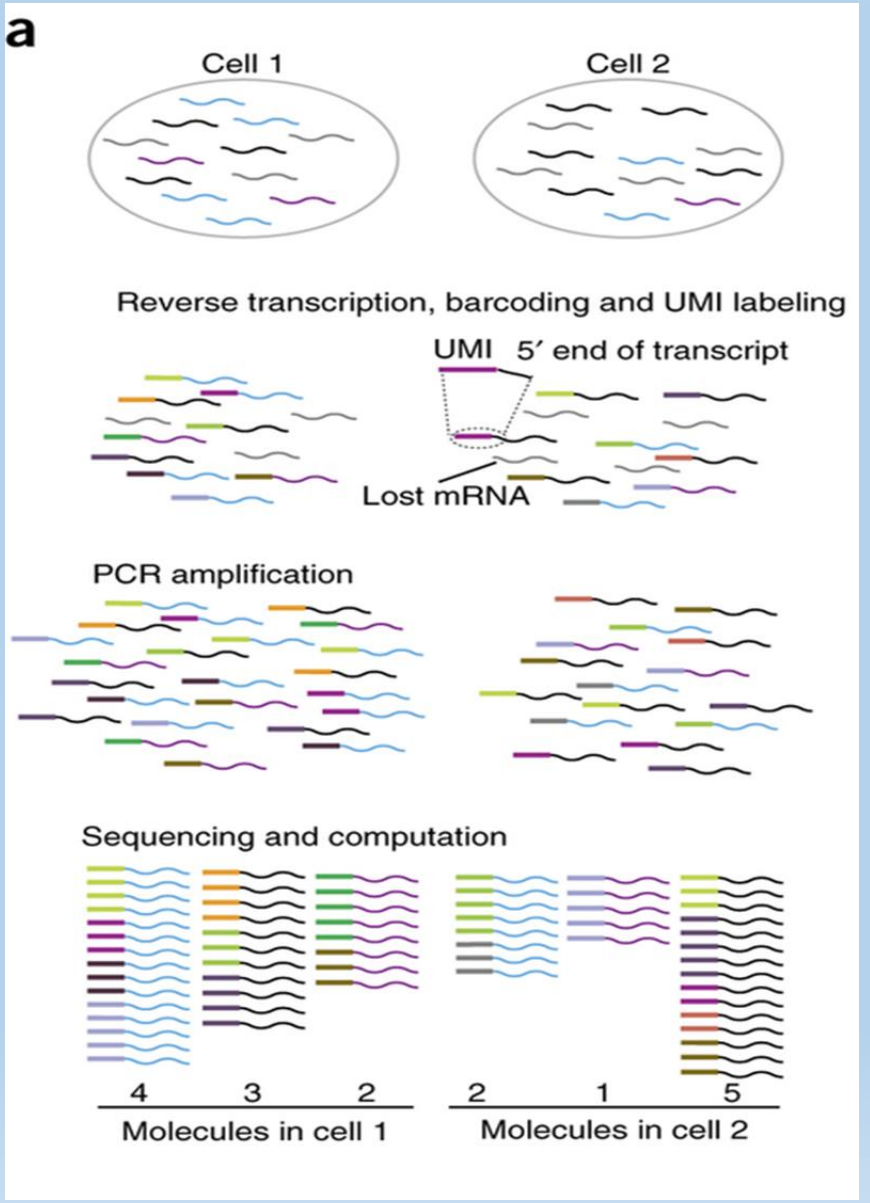






Problems with all that Amplification

- Low abundant transcripts are difficult to detect and data has lots of zeros.
- Ratio of transcripts or genes is often distorted during amplification.
- This can be mitigated by the UMIs. UMI acts as a molecule barcode.
- If each transcript gets a unique UMI then counting the UMI rather than the transcript corrects for amplification distortion.
- For DNA, amplification can lead to allelic imbalance, allelic drop outs, chimeric molecules, base changes.

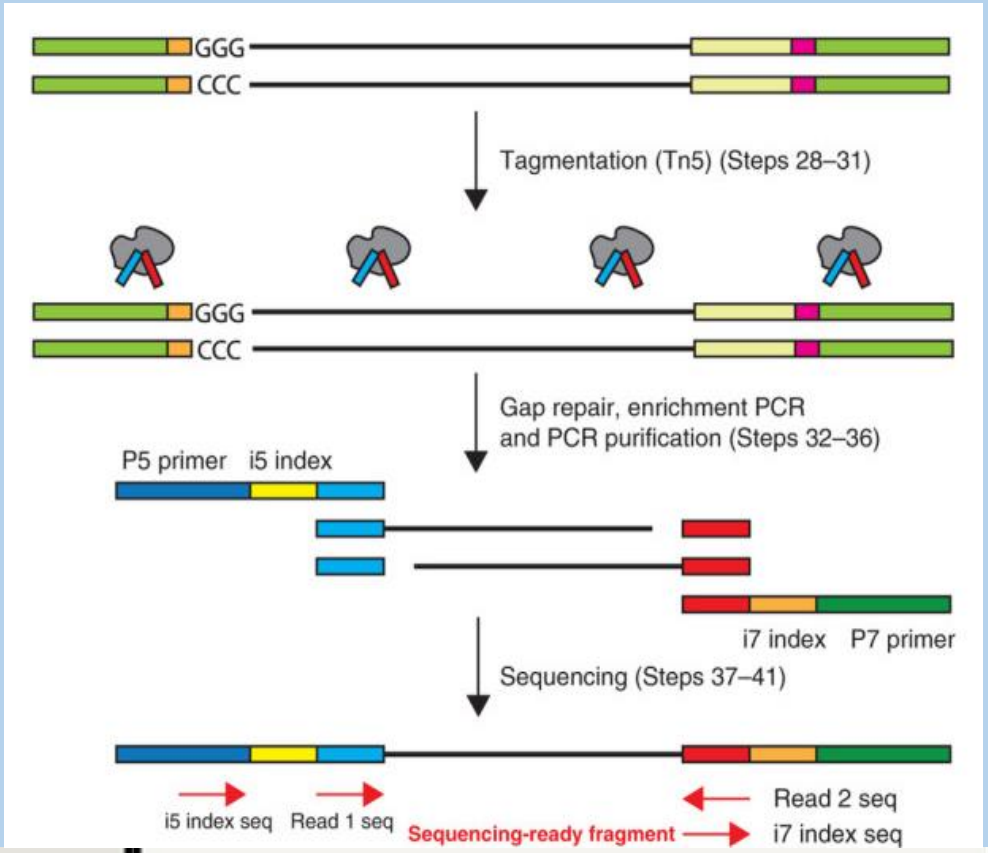


4. Preparing for sequencing

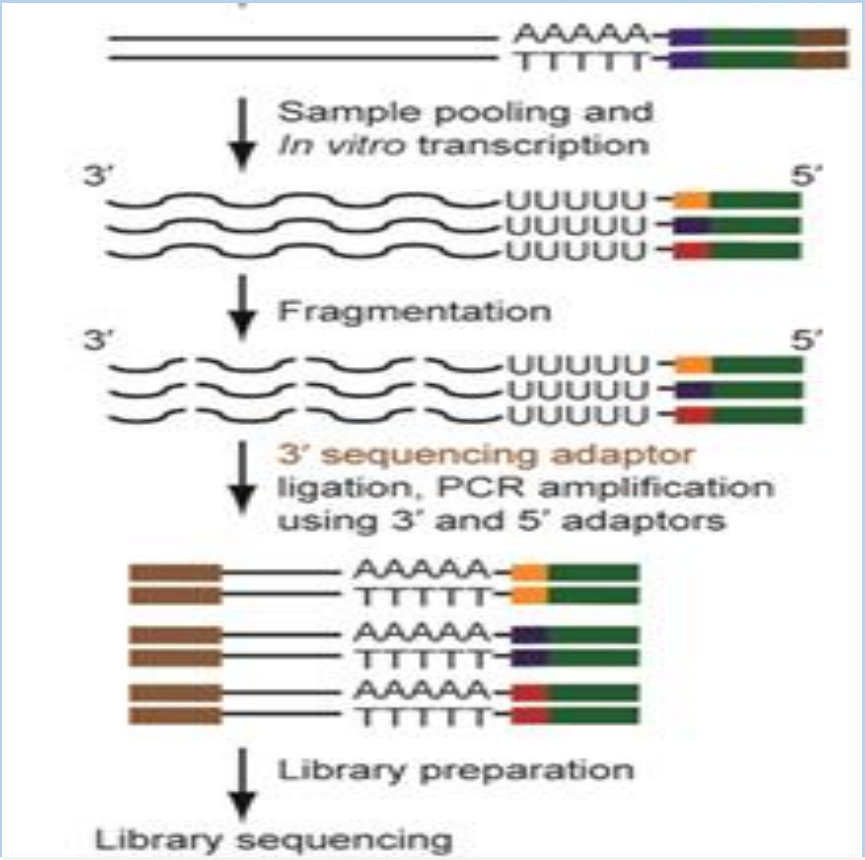
- Transposase based single cell libraries (e.g. ATAC-seq) are often ready to be sequenced.
- But for most techniques, DNA fragmentation and barcoding is required to allow sample pooling and sequencing with short read sequencing.
- Illumina requires fragments of ~300-600 bp, while nanopore fragments can be many kb in length.
- Simplest approach is to use a commercial kit e.g. Nextera.
- However for techniques that don’t include cell barcode e.g Smartseq2, this often expensive due to many parallel reactions.
- This step is often miniaturized to reduce costs.

Incorporated Cell Barcodes Complicate Library Preparation

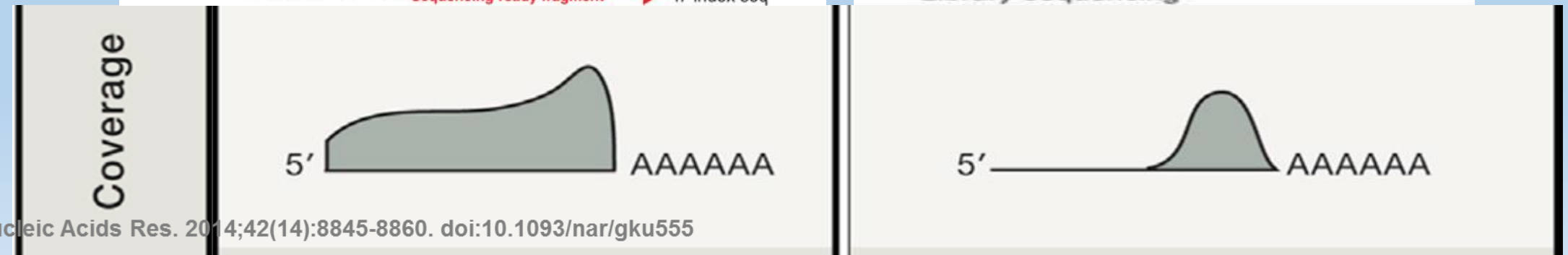
Full length cDNA for Sequencing e.g. Smartseq2



3’ Selection of cDNA for Sequencing e.g. Chromium

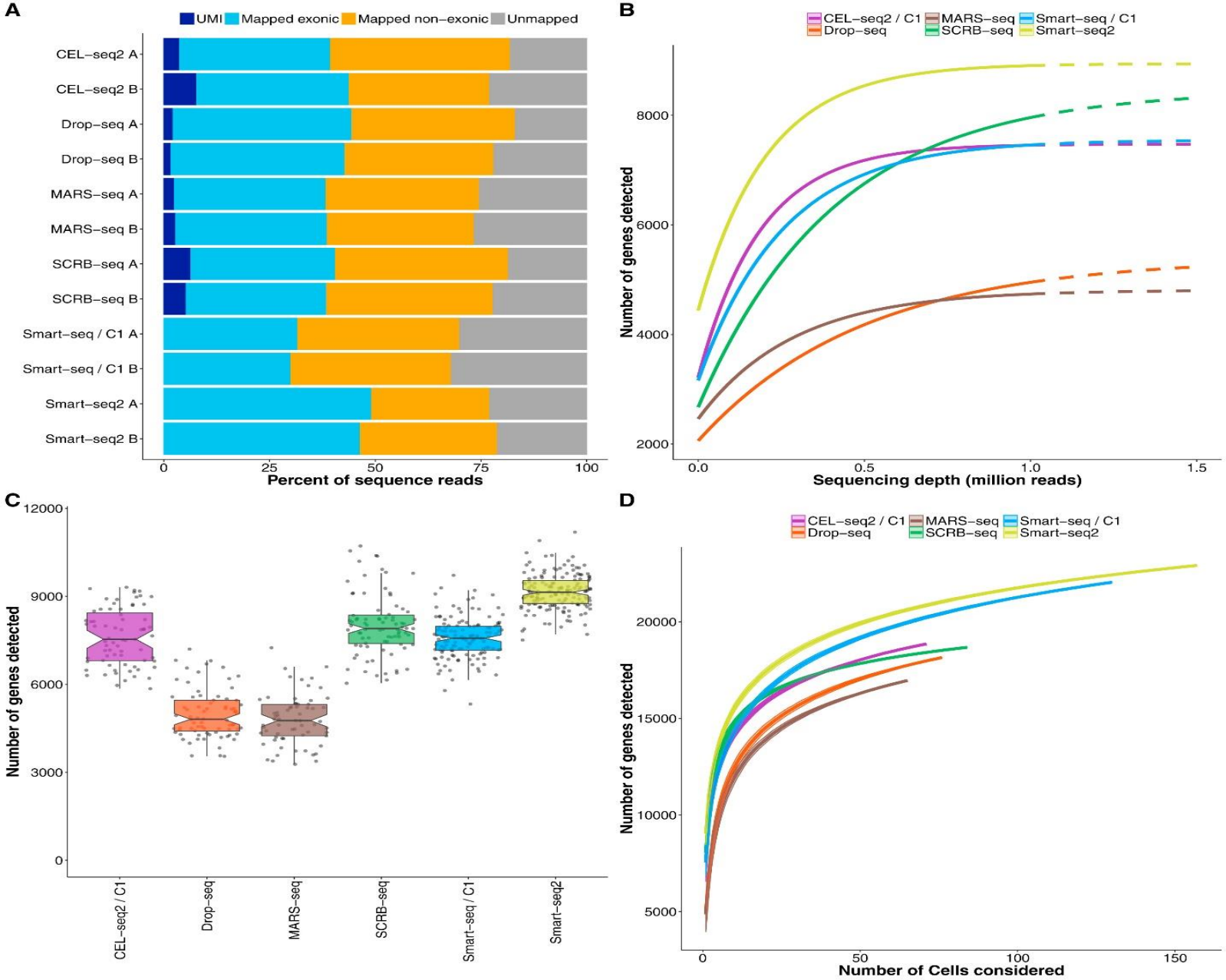


Keeping a cell barcode for sequencing means that the 3’ or 5’ prime part of the transcripts has to be ‘sacrificed’ in order to sequence on a short range sequencers.



Nucleic Acids Res. 2014;42(14):8845-8860. doi:10.1093/nar/gku555

Comparison of Single Cell RNAseq Techniques



Generally plate based full transcript techniques such as SmartSeq2 detect more genes than the droplet techniques such as Chromium.

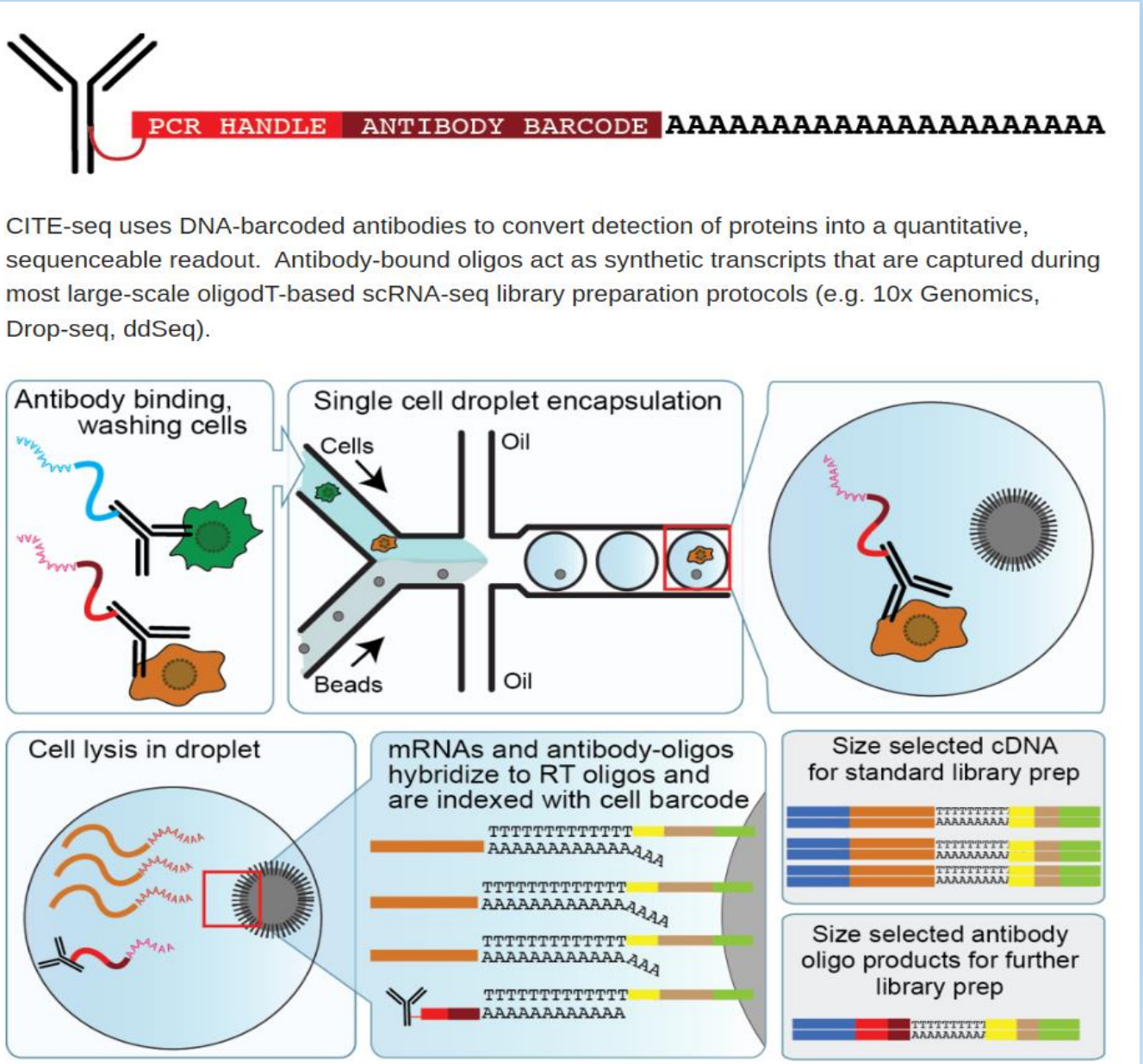
Comparative Analysis of Single-Cell RNA Sequencing Methods. Ziegenhain et al
Molecular Cell, Volume 65, Issue 4, 16 February 2017, Pages 631-643.e4

Single Cell Multi-Omics

Current single cell multi-omics:

- Exome and targeted genome e.g Target-Seq
- Protein/Exome e.g. CITE-Seq
- Protein/DNA e.g. MissionBio Tapestri
- RNA/ATAC-seq
- Methylation/ATAC-seq
- Exome/VDJ e.g Chromium 5’ Kits

Antibody Barcoding (CITE-Seq, Total-Seq etc)



Measurement of cell surface proteins using sequencing in many 1000s of cells.

Barcoding and mixing of different samples together (hashing) by using different barcodes with a common epitope e.g. CD45. Samples are resolved after sequencing by the antibody barcode.

Increasing the cell capacity of droplet RNASeq systems by allowing for improved doublet discrimination (termed ‘superloading’).

Any molecule can be barcoded and therefore measured by sequencing.

Summary

- Partitioning of single cells can be done by microfluidics, FACS or manual pipetting.
- Droplet techniques generally provide the highest throughput, but lowest sensitivity. They are generally most cost effective per cell.
- Plate based techniques offer the highest sensitivity and versatility but have a relatively low throughput. Generally most expensive per cell.
- Incorporation of cell barcodes early in a protocol greatly simplifies processing, and allows many more cells to be processed.
- Main problem with single cell methods is dealing with biological and technical noise e.g. cell cycle and amplification bias.
- Use of barcoded antibodies has many applications in single cell genomics.

The WIMM Single Cell Genomics Facility

The WIMM single cell facility is an ultra-clean environment for the processing and amplification of single cells and small biological samples for sequencing and other molecular analysis.

We offer the following services :

RNAseq (plate based and Fluidigm C1), Whole genome amplication, G&T Seq, ATAC Seq, and single cell targeted gene expression (Fluidigm Biomark, NanoString Ncounter, QPCR), fixed/permebilised cell RNA Seq, laser microdissection RNA Seq.

We also provide training, equipment and space to enable researchers to conduct their own single cell genomics without risk of contamination.

We also have pre- and post PCR robotics for automation of high throughput projects, as well as a TTP Labtech Mosquito robot for miniaturisation of library preps, QPCR and Whole Genome Amplification and other molecular techniques.

http://www.imm.ox.ac.uk/single-cell-facility

5. Single Cell Data Analysis

Analysing scRNA-seq data is harder than for bulk experiments.

Typical single-cell studies capture hundreds or even thousands of cells.

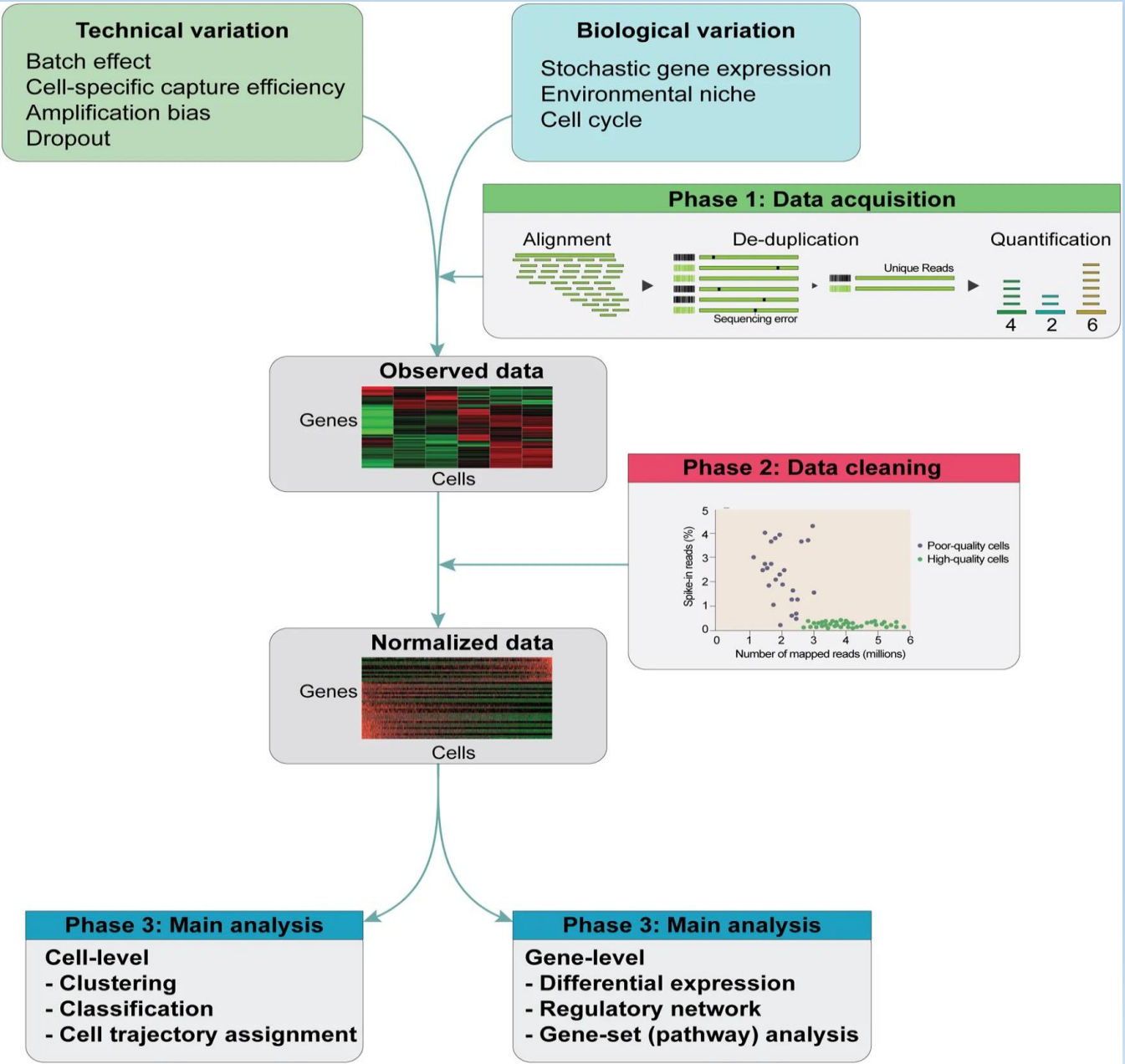
Single-cell expression measurements are often highly variable, and separating technical from biological variability is essential.

Biological noise can be cell cycle, ‘burst transcription’, cell death, senescence, aneuploidy etc.

Technical noise is can be due to loss of material during processing, inefficient reverse transcription/PCR, PCR bias, loss of low abundant genes, batch effects etc.

For techniques such as single cell ATAC seq the data may also be very sparse.

A Typical Single Cell Pipeline



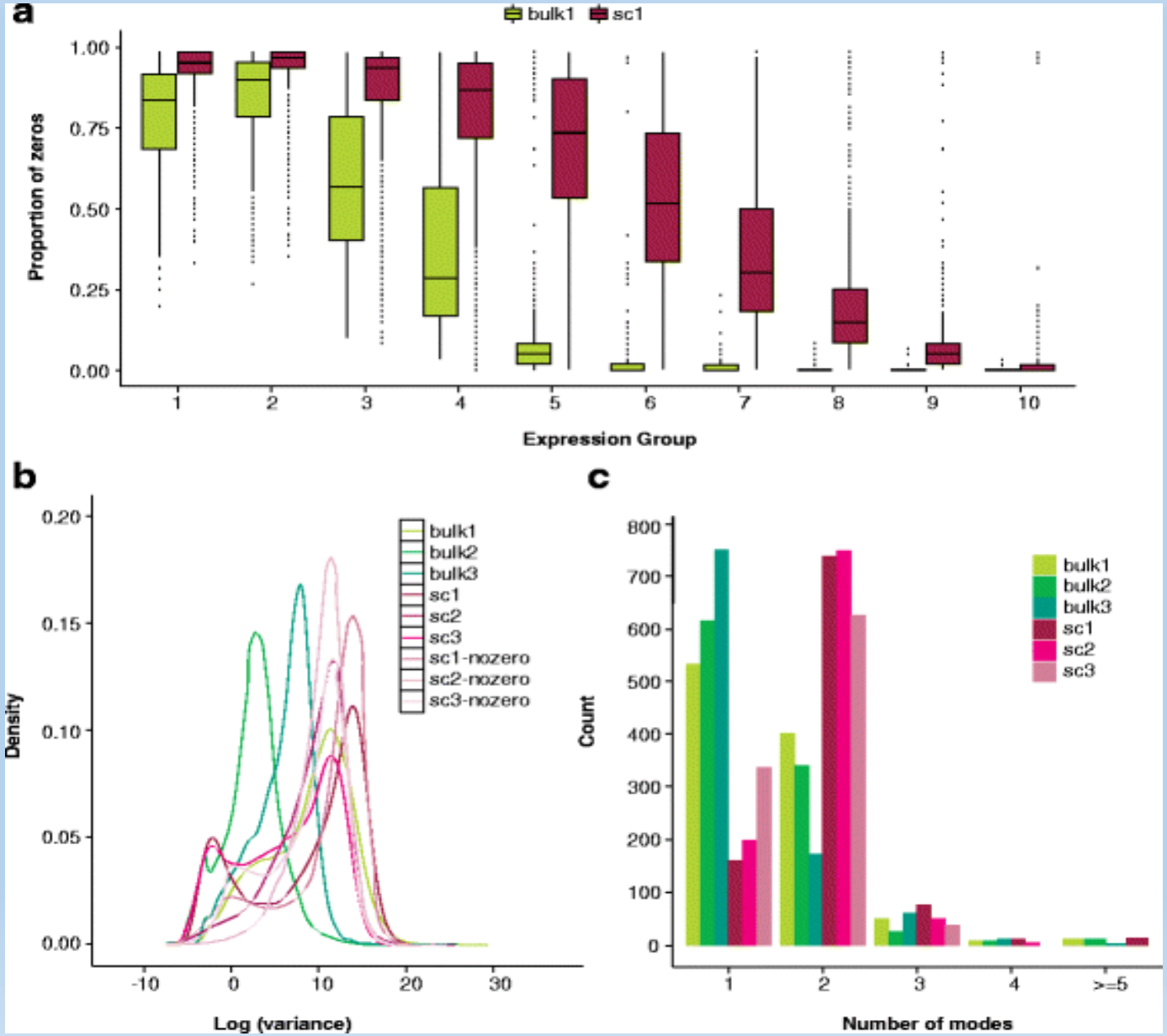
1 Quality control (QC) e.g.FastQC. Low-quality bases (usually at the 3’ end) and adapter sequences can be removed at this pre-processing step.

2 Read alignment. e.g. STAR are the same as those used in the bulk RNA-seq analysis pipeline. Uniquely mapped reads, exonic regions, and expected coverage patterns.

3 Normalization is necessary to remove cell-specific bias, which can affect downstream applications

Single-cell RNA sequencing technologies and bioinformatics pipelines Byungjin Hwang, Ji Hyun Lee & Duhee Bang
Experimental & Molecular Medicine
Volume 50, Article number: 96 (2018)

5. Analysis of Single Cell Data



Prominent features in single-cell RNA-seq data relative to bulk RNA-seq include an abundance of zeros, increased variability, and multi-modal expression distributions.

a Boxplots of the gene-specific proportion of zeros in a bulk (bulk1) and single-cell (sc1) dataset stratified by percentile of median gene expression.

b Densities of gene-specific log variance for all genes in three bulk and three single-cell RNA-seq datasets.

c For each dataset shown in b, 1000 genes were selected at random from the list of genes for which at least 75 % of cells showed non-zero expression.

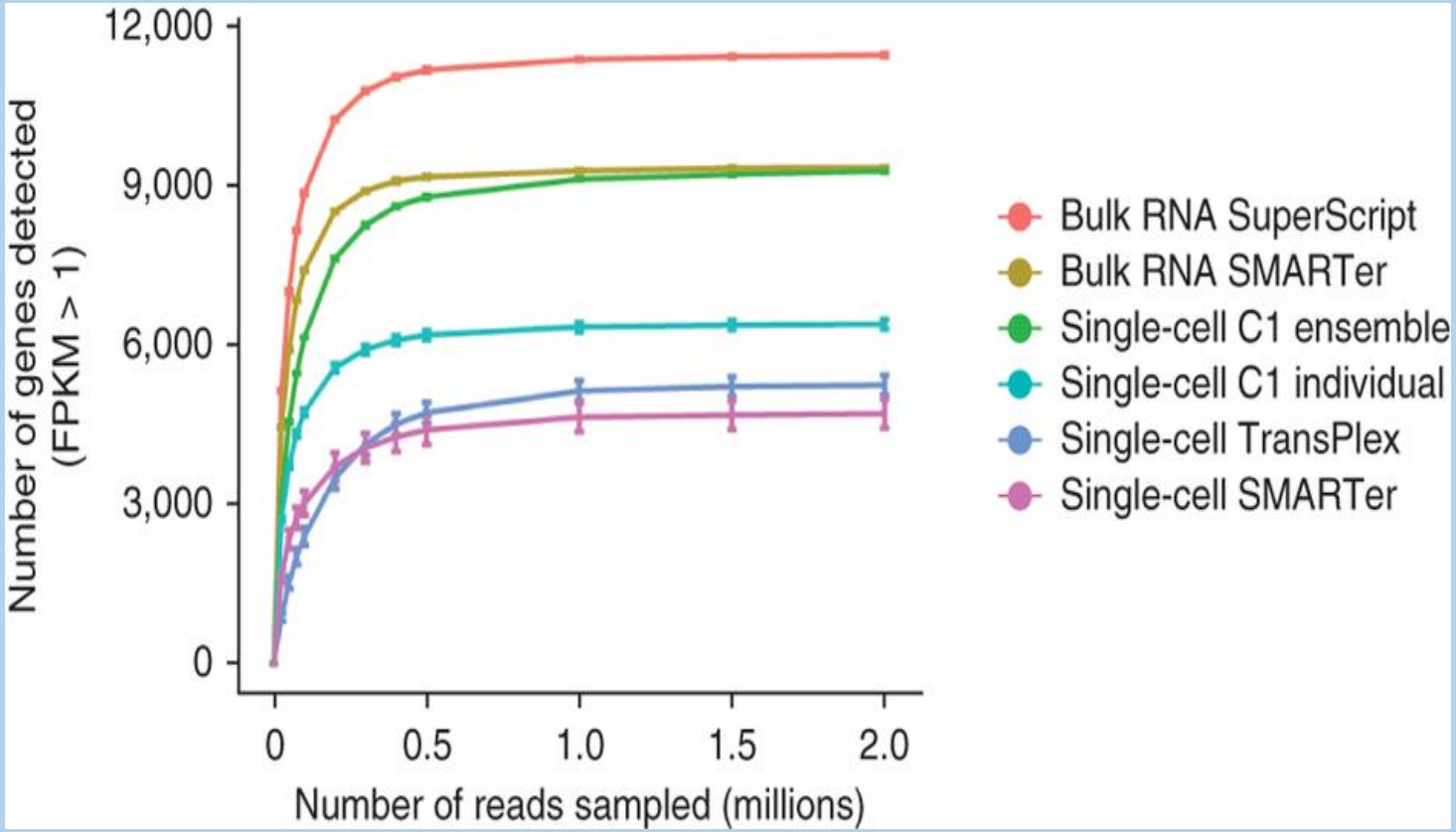
Good Experimental Design is Essential for Single Cell Techniques

- Avoid bias, technical or experimental.
- Consult a statistical biologist before you start!
- Plan efficient sample collection, storage, and sample processing to minimise batch effects at al stages. Process samples in big batches, use automation.
- Include controls, randomizing sample processing and smart management of sequencing runs.
- Use sequins/spike ins and preferably use UMIs if possible. Use appropriate sequence depth.

A survey of best practices for RNA-seq data analysis, Conesa et al, Genome Biology 2016, 17:13

Sequence Depth for single cell libraries

Low complexity of single cell libraries means they only need to be lightly sequenced to gain all the information



Nature Methods 11, 41–46 (2014)