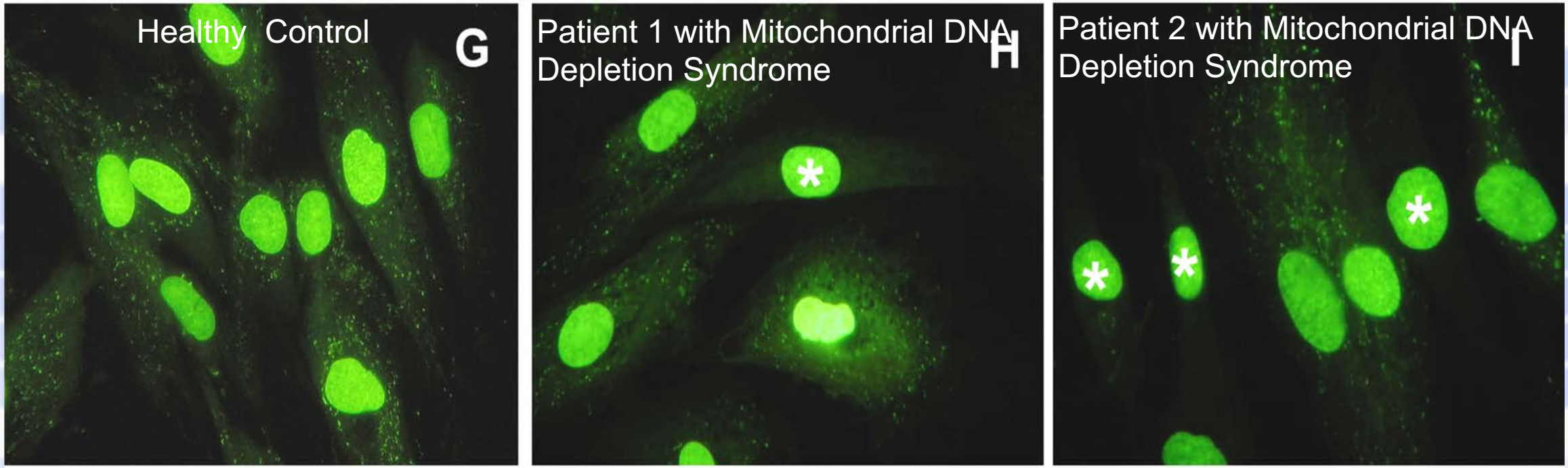
Single Cell Genomics Twitter @NOTWIMM

Dr. Neil Ashley | MRC Weatherall Institute Molecular Medicine | University of Oxford

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



Microscopic methods were first used to study disease at the single cell level. NGS allowed genomics to be applied at the single cell level and recent advances in situ sequencing mean the 2 approaches are merging.

What Can We Get from Single Cell Methods ?

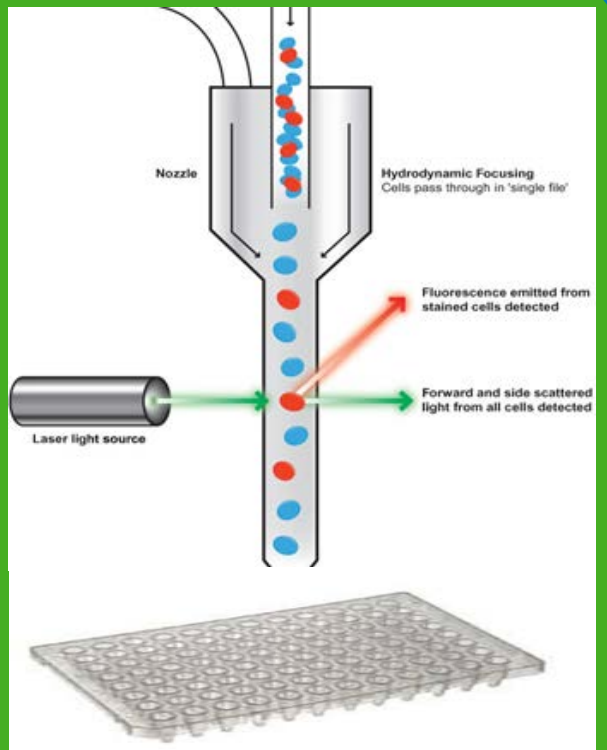
- Gene expression, RNA splicing, small RNA, circular RNA.
- Genome e.g. DNA mutations, genotypes, copy variation, deletions, inversions etc.
- Epigenomic Information e.g. methylation, chromatin conformation, Hi-C.
- Proteomics – protein expression, protein modification.
- Combinations of the above all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

Basic Steps for Single Cell Genomics

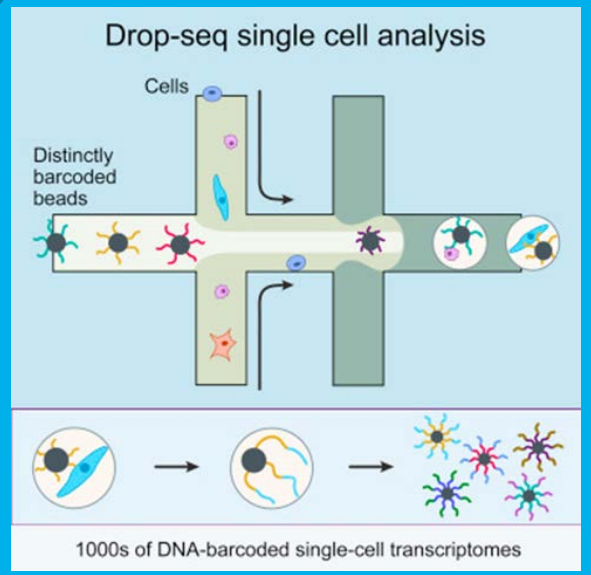
1. Isolate single cells
2. Amplify single cells
3. Prepare amplified material for sequencing
4. Sequence single cell libraries.
5. Analyse the data

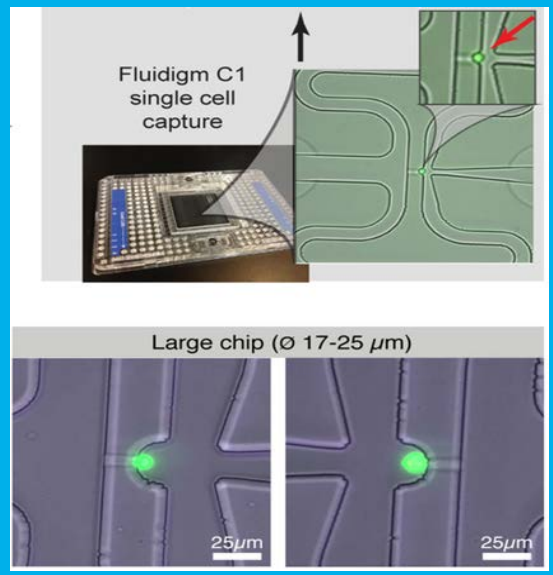
Step 1: Partitioning Single Cells for Amplification

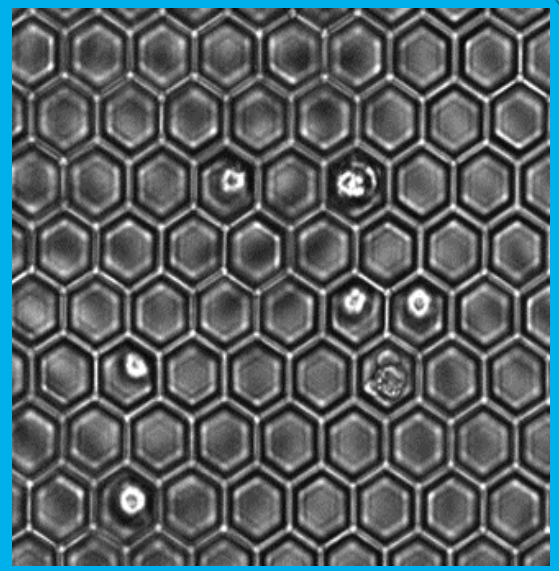
- **The 1st step in any single cell genomics method is to partition the single cell so it can be amplified for the read-out.**
- **This can be achieved in several ways:**



FACS sorting into plates







Microfluidic capture in in oil emulsion drops, fluidic traps, microwells

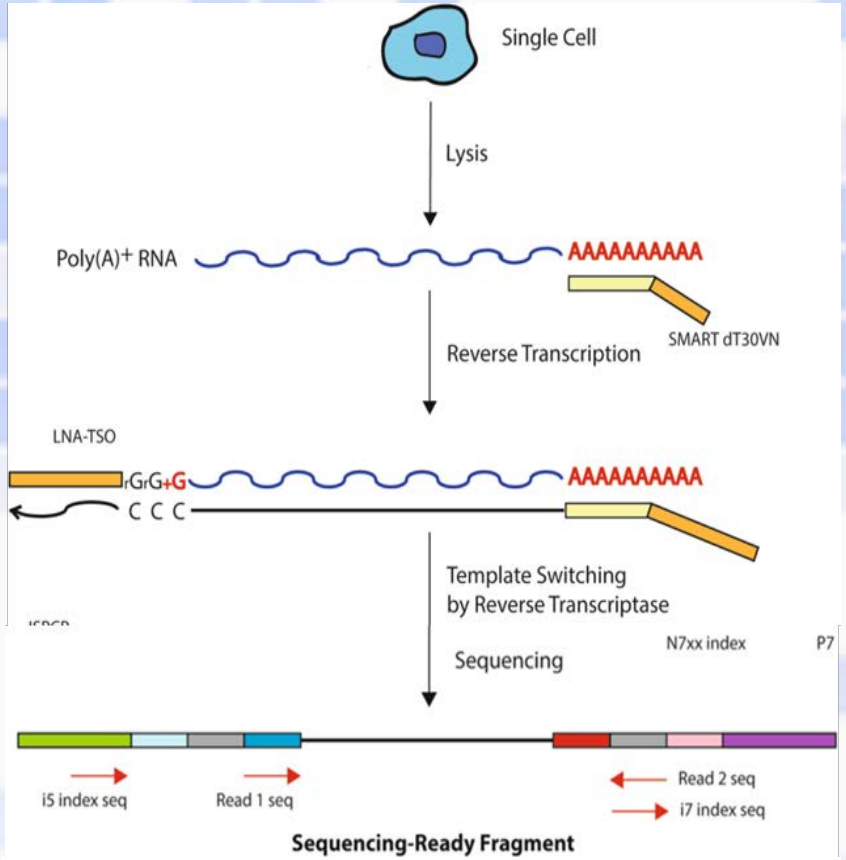
MRC Weatherall Institute of Molecular Medicine | MRC Medical Research Council | UNIVERSITY OF OXFORD

2. Amplification

- There is about 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 (even for many in situ techniques)
- Amplification is most commonly done by PCR, which is generally targeted to poly A transcripts or individual genes.
- Alternatively the whole genome or whole exome can be amplified using isothermal enzymes e.g. Phi29, BST2.

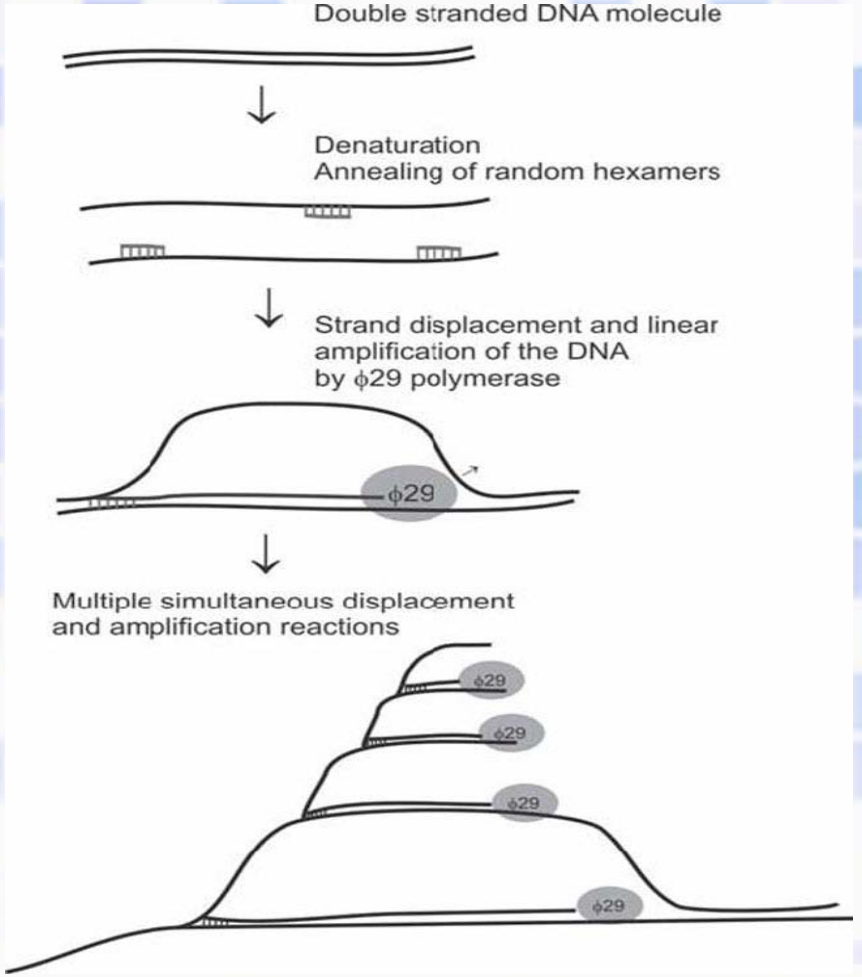
Some Commonly used Single Cell Amplification Methods

RNAseq: Template switching: 2 PCR handles to each cDNA molecule during reverse transcription. The cDNA is then amplified by PCR.



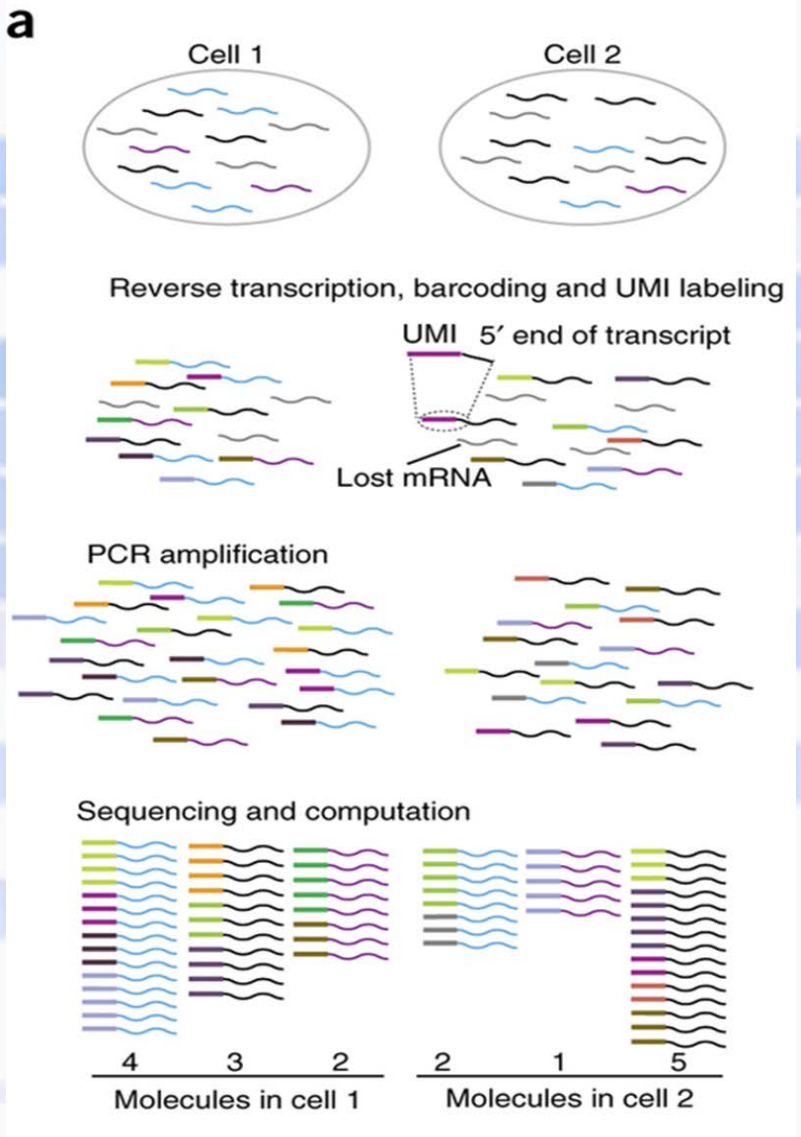
Picelli S, et al Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014 Jan;9(1):171-81.

Whole Genome Amplification : Random hexamer primers anneal to DNA. Phi29 copies the DNA via strand displacement

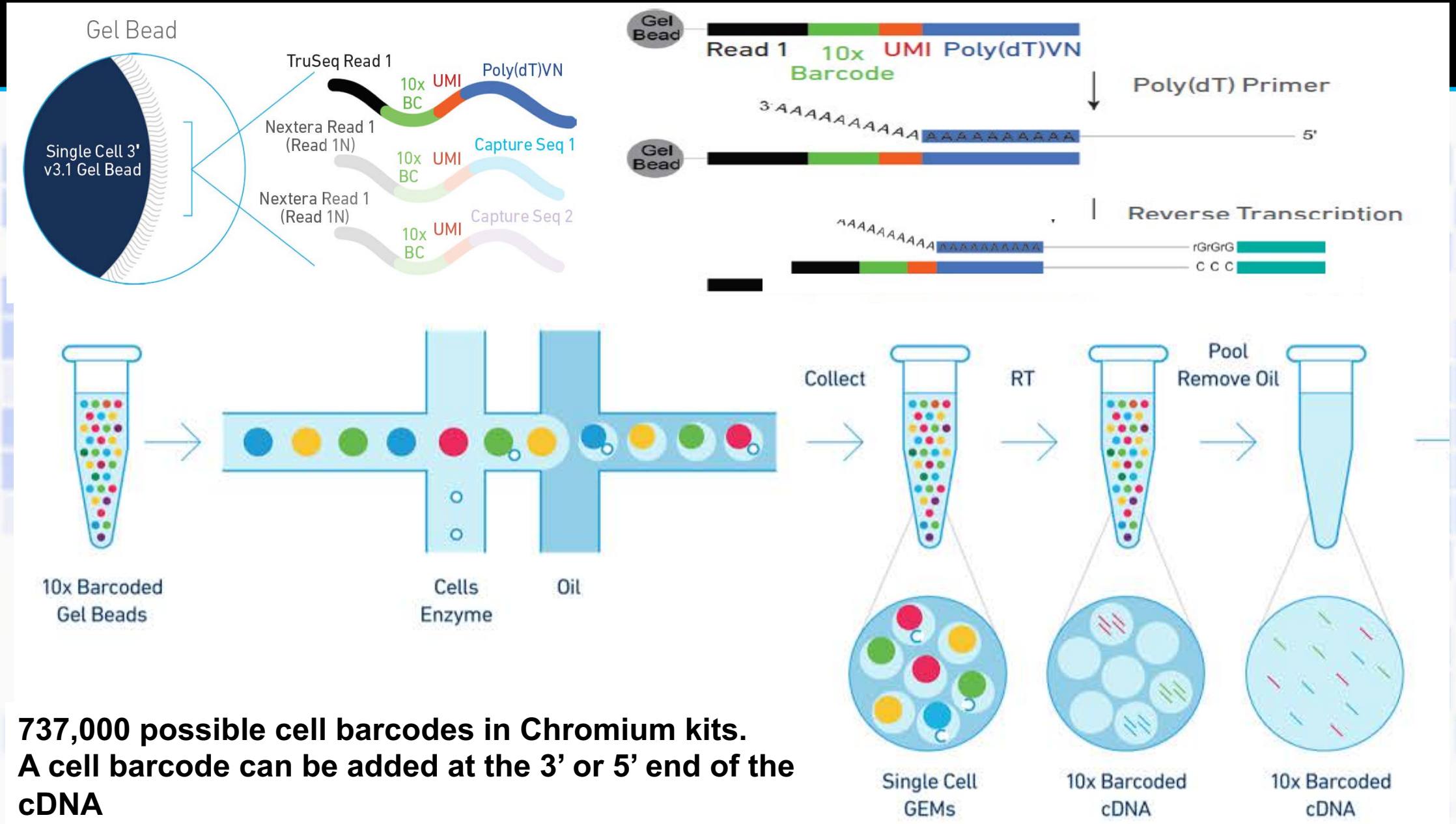


Problems with all that Amplification

- Low abundant transcripts are difficult to detect in single cells.
- 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.



Droplet Based scRNA-Seq – Adding a cell barcode



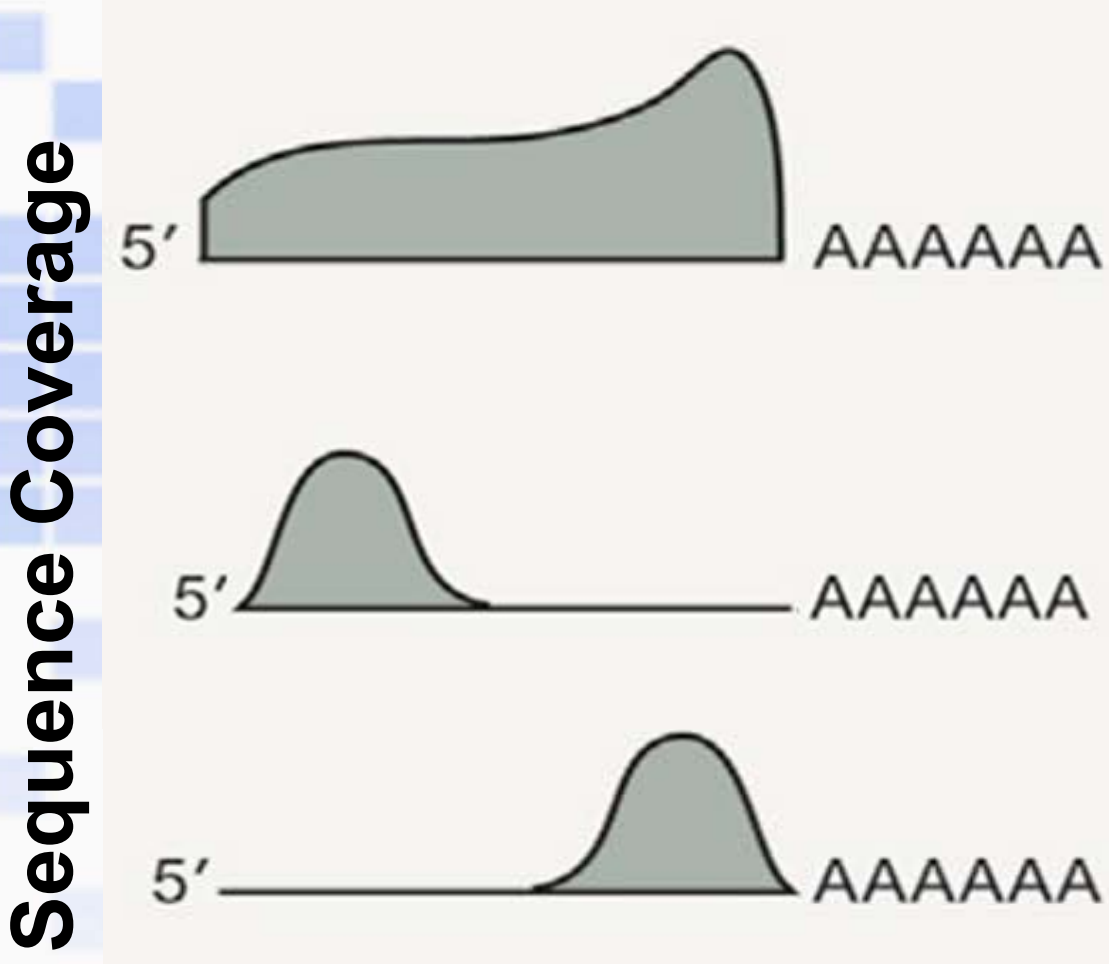
737,000 possible cell barcodes in Chromium kits.
A cell barcode can be added at the 3' or 5' end of the cDNA

Key Advantages of Cell Barcoding

- The ability to barcode cells allows a massive increase in scale as cells can be pooled at an early step in the workflow.
- For example for Chromium/Rhapsody/DropSeq assays upto 60000 cells can be pooled into 1 tube straight after cDNA barcoding.
- Non barcoded single cells have to be processed separately right up to sequencing. This is logistically difficult when doing many cells as it means many plates of cells need to be processed.
- This is expensive for reagent usage and often requires automation.

Single Cell Barcodes & UMIs need to be enriched for short sequencing platforms

Keeping a cell barcode for sequencing means that the 3’ or 5’ prime part of the transcripts have to be enriched so they can be sequenced on a short range sequencers. Otherwise the gene data would lose its cell barcode.

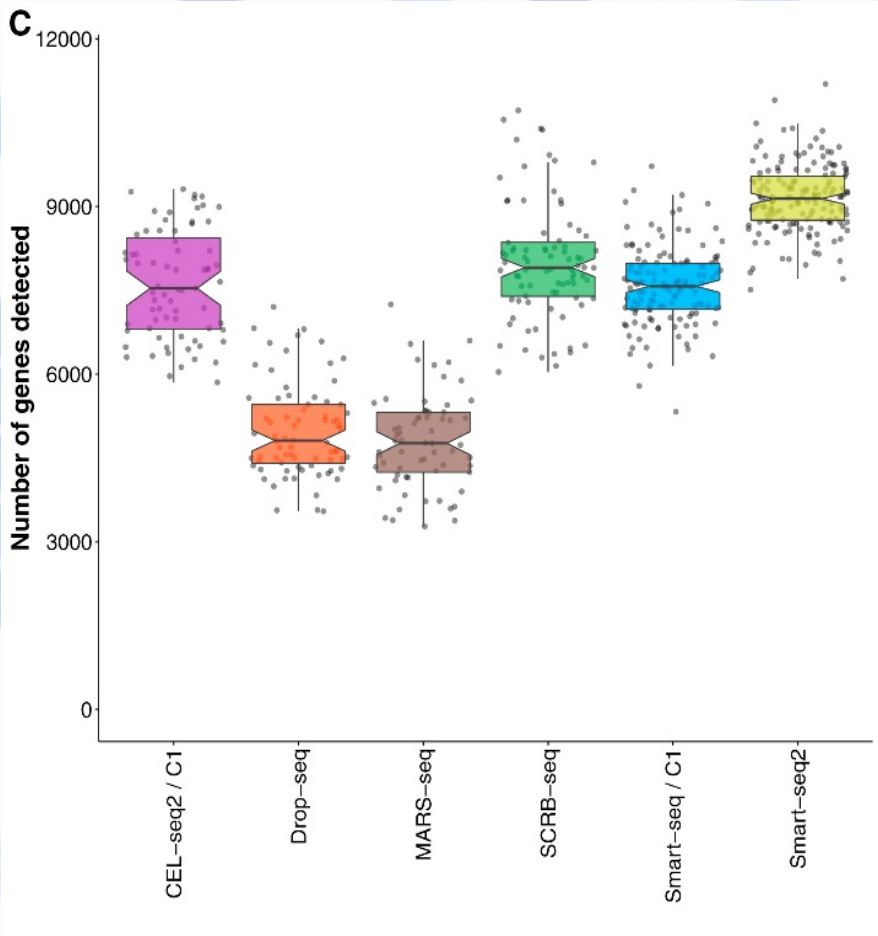


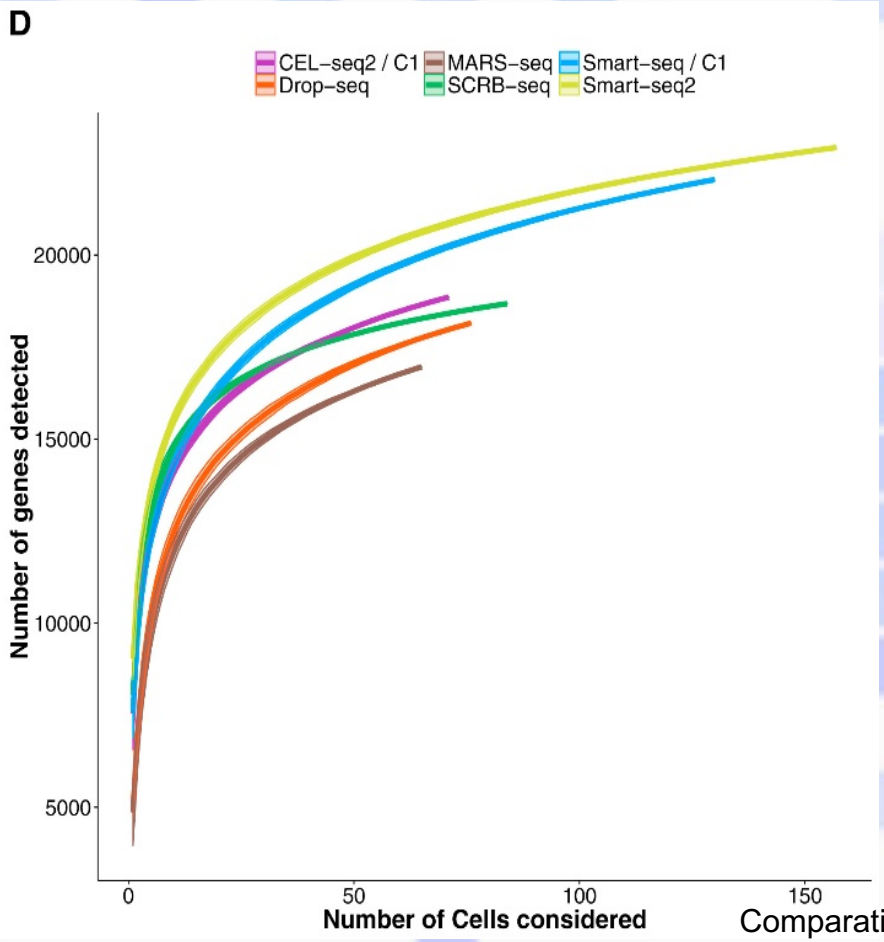
With no UMI or barcode the full length of the cDNA can be sequenced.

For barcodes at the 5’ end (added via TSO) the 5’ end of the cDNA is enriched for Sequencing e.g. Chromium TCR

For barcodes at the 3’ end (added via poly A oligo) the 5’ end of the cDNA is enriched for Sequencing e.g. Chromium TCR

Comparison of Single Cell RNAseq Techniques – plate based techniques are Generally more Sensitive





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

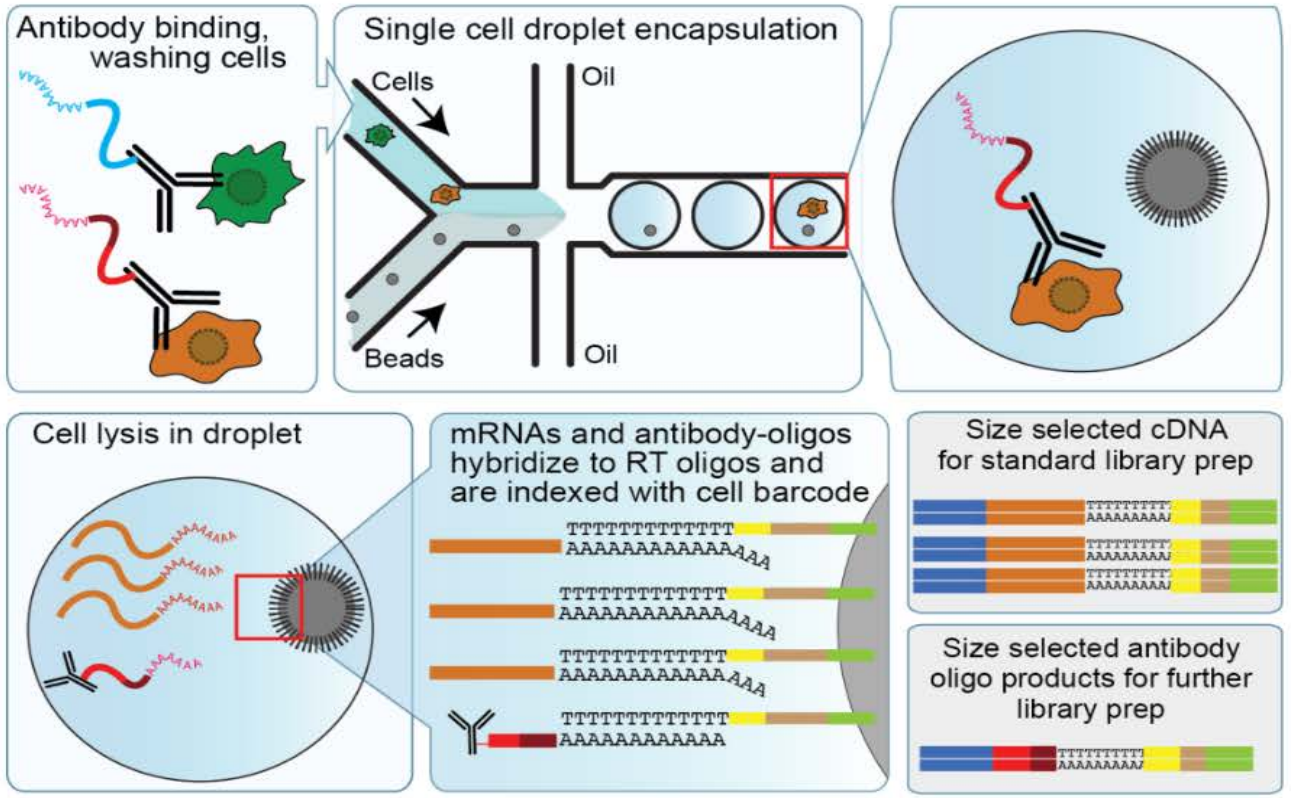
Comparative Analysis of Single-Cell RNA Sequencing Methods. Ziegenhain et al
Molecular Cell, Volume 65, Issue 4, 16 February 2017, Pages 631-643.e4

Single Cell Multi-Omics

- Exome and genome e.g *Target-Seq*
- Protein/genome – *Tapestri*
- RNA/ATAC-seq - *Chromium*
- Methylation/ATAC-seq
- Exome/VDJ e.g *Chromium*
- ATAC/Protein – *ASAP-Seq*
- CRISPR/RNA/Protein/TCR – *ECCITE-SEQ*

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

CITE-seq uses DNA-barcoded antibodies to convert detection of proteins into a quantitative, sequenceable readout. Antibody-bound oligos act as synthetic transcripts that are captured during most large-scale oligodT-based scRNA-seq library preparation protocols (e.g. 10x Genomics, Drop-seq, ddSeq).



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

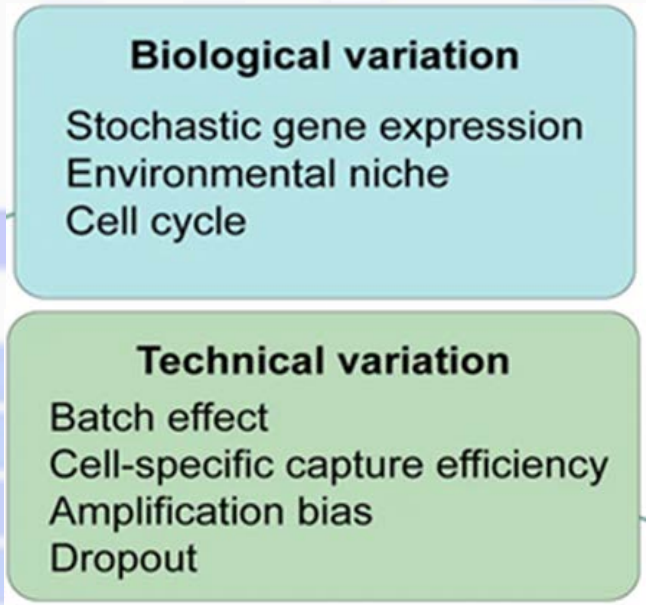
Barcoding and mixing of different samples together (hashing) by using different barcodes with a common epitope e.g. CD45. Samples are resolved after sequencing by the antibody barcode.

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

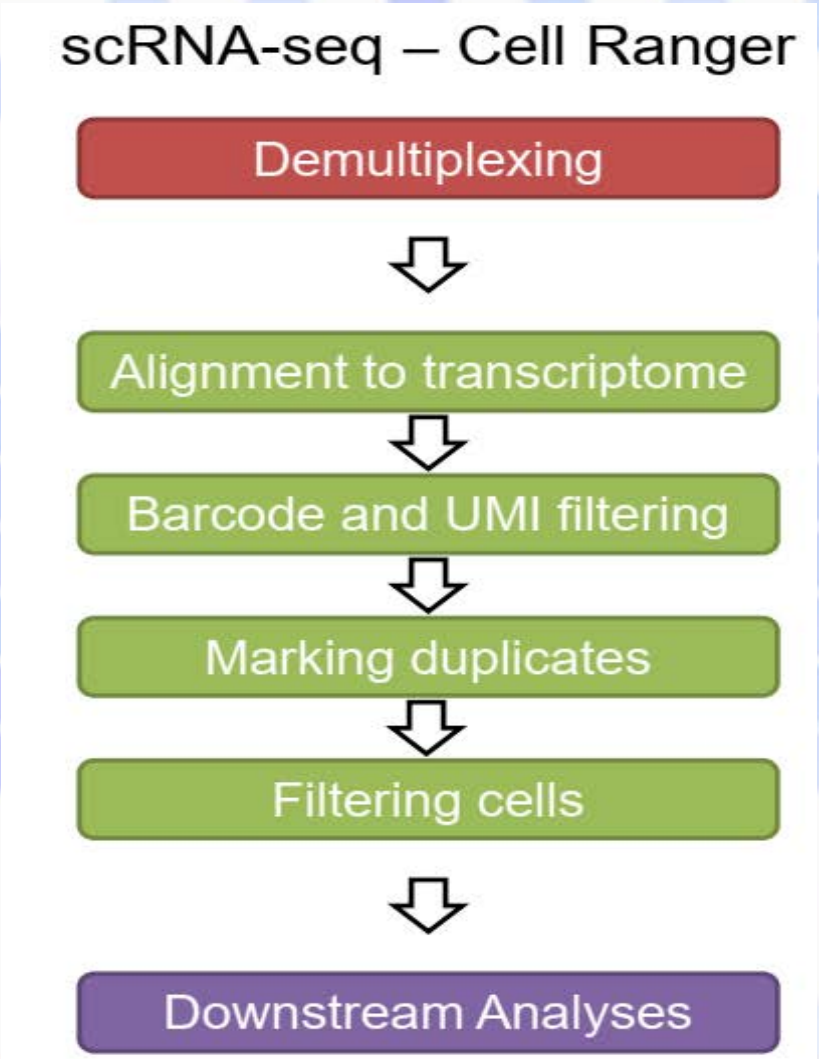
Any molecule can be barcoded and therefore measured by sequencing.

RNASeq Data Analysis

- There are many single cell data analysis packages
- **Suerat** is a popular R package
- Commercial packages include **Partek Flow** are easy to use.
- Most commercial kits have online Bioinformatics support e.g. **CellRanger/Loupe** from 10XGenomics.
- These packages take the raw BCL data and convert it into FASTQ, then on to BAM files.
- They give a number of useful metrics and some simple data presentation.
- More sophisticated analysis may be needed for single cell noise. There are a lot of 0s in single cell data.



Single Cell RNASeq Data Analysis



- Barcode Extraction and filtering
  - Identifies cell level barcodes
- Mapping to reference
  - Uses STAR aligner
- Generate count table
  - UMIs per gene in each cell
- Dimensionality Reduction
  - PCA and tSNE
- Clustering
  - K-means and Graph Based

Typical 10XGenomics Chromium 3’ Data

Estimated Number of Cells

10,140

Mean Reads per Cell	Median Genes per Cell
11,376	2,014

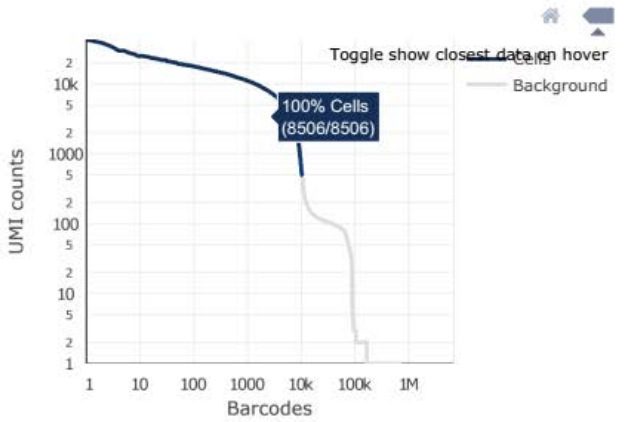
Sequencing

Number of Reads	115,354,960
Valid Barcodes	97.5%
Sequencing Saturation	9.0%
Q30 Bases in Barcode	97.0%
Q30 Bases in RNA Read	89.3%
Q30 Bases in Sample Index	96.7%
Q30 Bases in UMI	96.5%

Mapping

Reads Mapped to Genome	95.9%
Reads Mapped Confidently to Genome	92.6%
Reads Mapped Confidently to Intergenic Regions	5.4%
Reads Mapped Confidently to Intronic Regions	20.8%
Reads Mapped Confidently to Exonic Regions	66.3%
Reads Mapped Confidently to Transcriptome	63.1%
Reads Mapped Antisense to Gene	1.6%

Cells



Estimated Number of Cells	10,140
Fraction Reads in Cells	87.7%
Mean Reads per Cell	11,376
Median Genes per Cell	2,014
Total Genes Detected	20,483
Median UMI Counts per Cell	4,777

Sample

Name	d135control
Description	
Transcriptome	mm10plus
Chemistry	Single Cell 3' v3
Cell Ranger Version	3.0.2

Summary

- Partitioning of single cells can be done by microfluidics or FACS.
- 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.

Oxford University WIMM Single Cell Facility

- Dedicated single cell genomics facility open to all.
- 10X Chromium, BD Rhapsody, 5’ & 3’ Prime RNA-seq, Full Length RNA-Seq Fluidigm.
- Sequencing and library preparation.
- Services or training provided for most single cell techniques.

www.imm.ox.ac.uk/research/facilities/single-cell-facility

neil.ashley@imm.ox.ac.uk Twitter @NOTWIMM

MRC Weatherall Institute of Molecular Medicine

MRC Medical Research Council

UNIVERSITY OF OXFORD