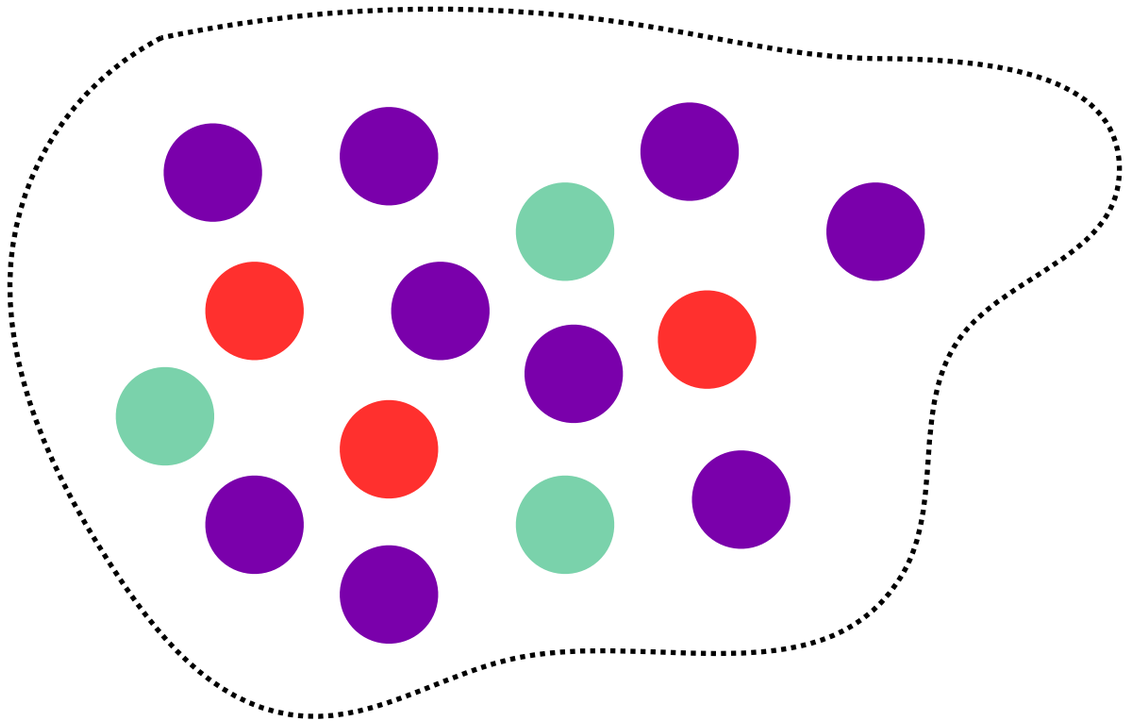
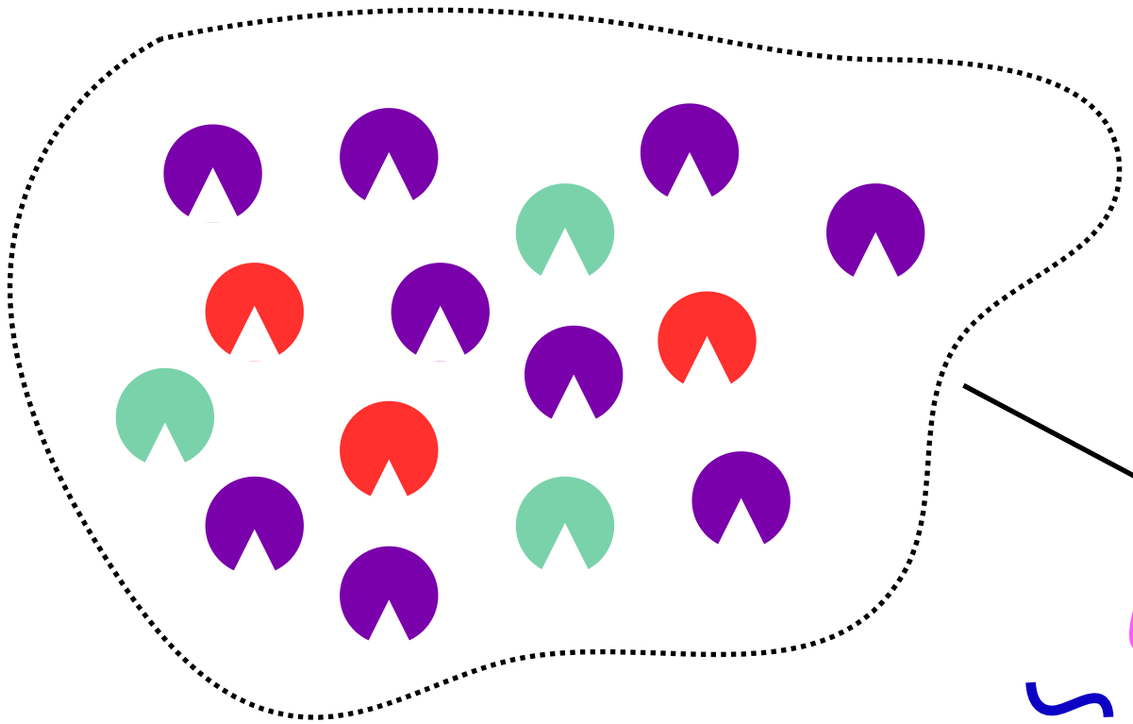
Single-cell RNA-seq (intro)

Charlotte Soneson
CSAMA, Brixen, July 10 2018

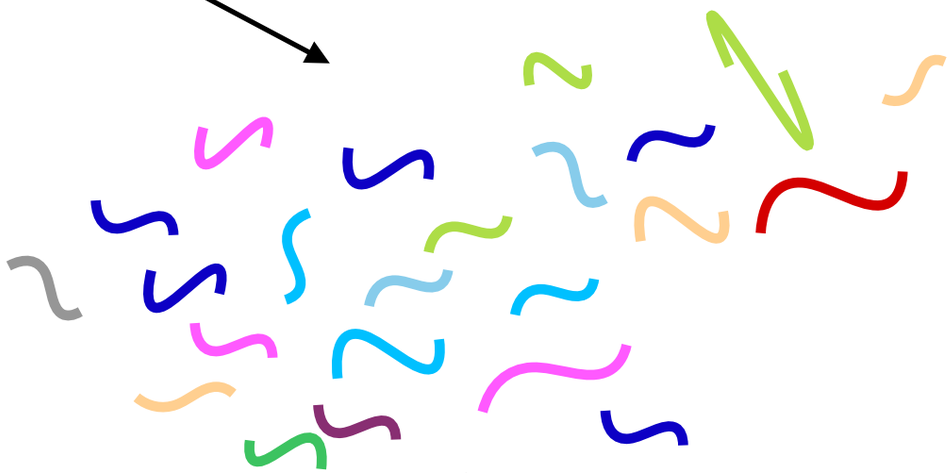
Bulk RNA-seq



Bulk RNA-seq



RNA extraction (polyA selection, ribosomal RNA depletion)



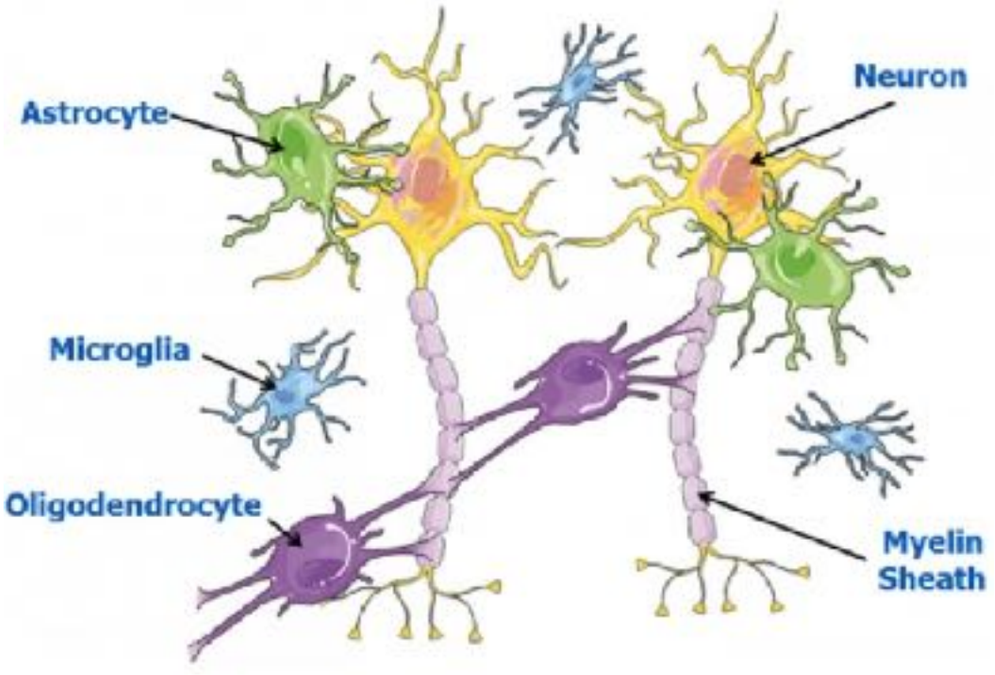
Fragmentation

cDNA synthesis Amplification

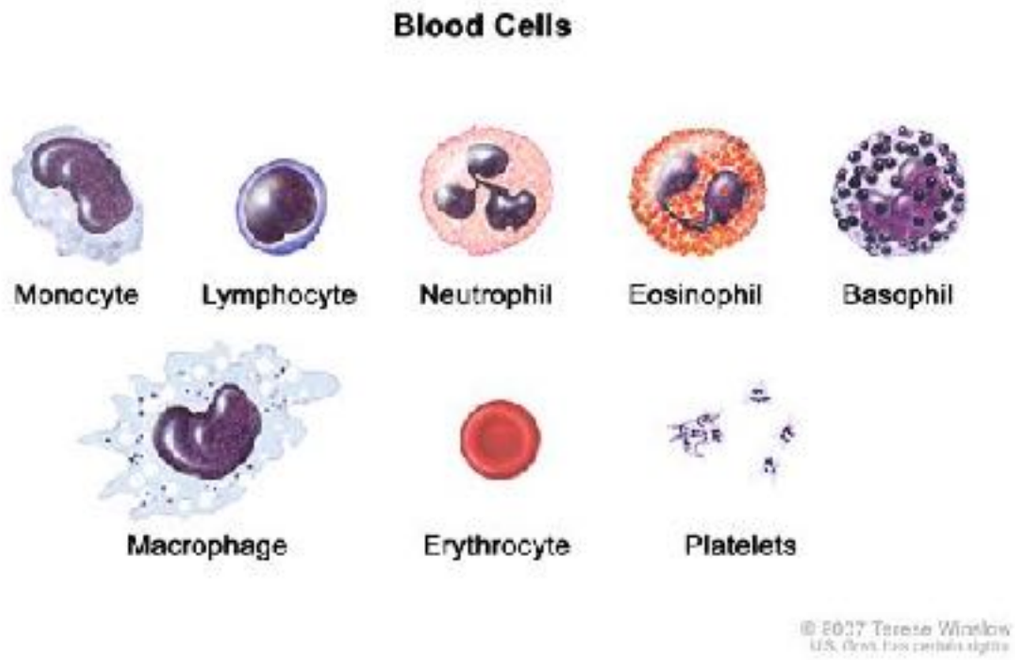
Sequencing Abundance quantification

gene	count
gene1	32
gene2	4
gene3	0
gene4	1928
...	

There are lots of different sorts of cells



- Wikipedia lists >200 cell types present in the human body
- Cell *type* often refers to more “permanent” aspects of a cell’s identity



- Cell *state* instead reflects aspects arising in more transient processes (differentiation, cell cycle, circadian rhythm)
- Samples are often heterogeneous in terms of cell types/states

Wagner et al., Nature Biotechnology 2016

https://www.cancer.gov/publications/dictionaries/cancer-terms/def/red-blood-cell, http://facts.net/brain/

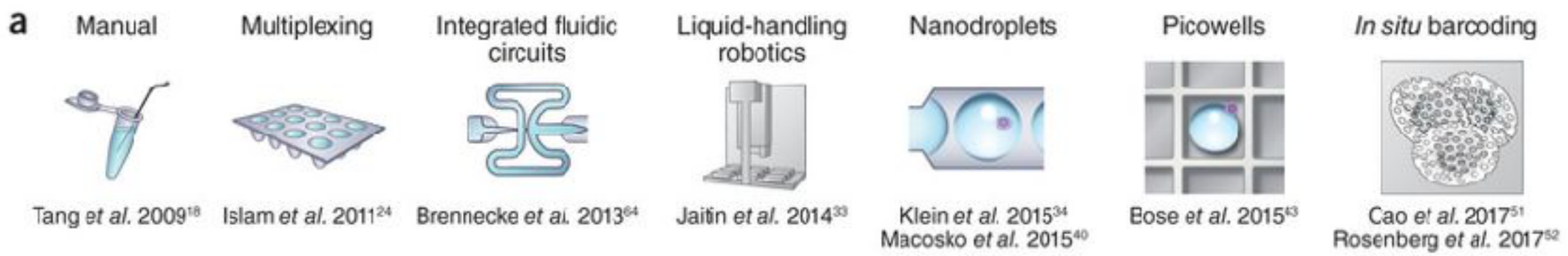
Opportunities with single-cell (RNA-seq) assays

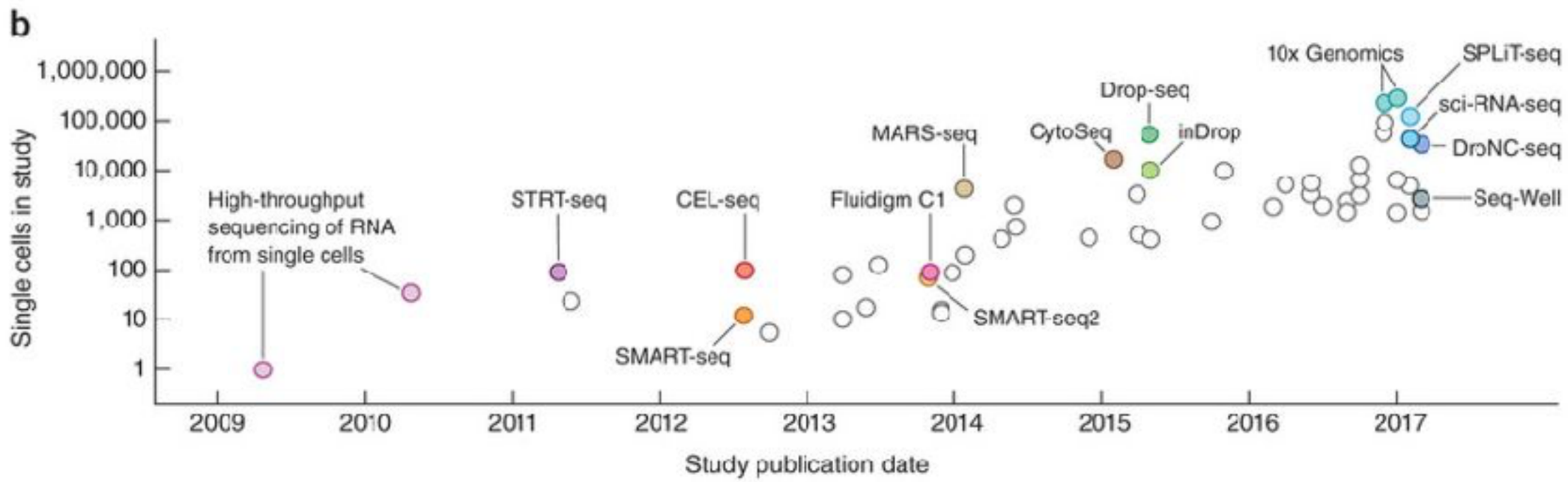
- Quantify abundance of individual cell types, and detect new ones
- Discriminate between differences in gene abundance due to changes in cell type composition, and differences due to changes in the expression levels within a given cell type
- Study the variability (more generally, the distribution) of a gene's expression among cells
- ...

scRNA-seq is not a single assay

- SmartSeq(2)
- CEL-Seq
- 10x Genomics (GemCode, Chromium)
- DropSeq
- STRT-Seq

- MARSseq
- QuartzSeq
- inDrop
- ...





Svensson et al., Nature Protocols 2018

Differences between single-cell protocols

- Throughput (many shallowly sequenced cells or few deeply sequenced ones)

- Full-length coverage vs 3’ (or 5’) end tag sequencing

- Use of UMIs (unique molecular identifiers) to account for PCR amplification artifacts or not

- Possibilities to incorporate external spike-in molecules for normalization and technical variance estimation

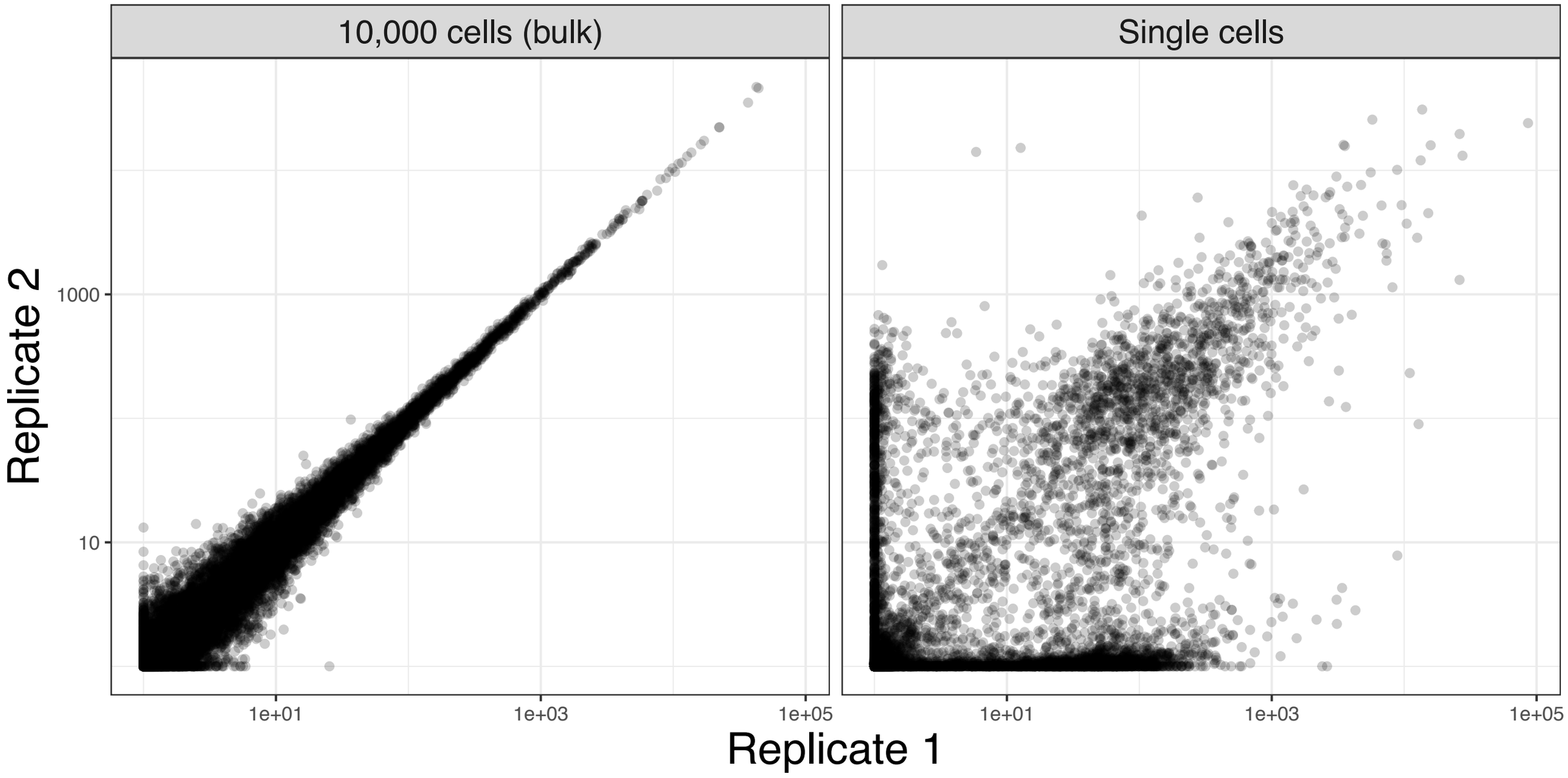
UMIs - accounting for amplification biases

- Low input -> strong amplification needed
- UMIs are short random “barcodes” that are added to transcripts during reverse transcription
- All reads arising from the same transcript molecule will have the same barcode
- Instead of counting reads aligning to a gene, count unique UMIs (“molecules”)

Challenges with scRNA-seq data

- Doublets (can be difficult to identify, depends on number of loaded cells) - single-cell isolation may not be straightforward
- Not all cells are captured for sequencing
- Not all transcripts (only 10-20%) in a cell are captured (reverse transcribed) -> dropouts (zero observed reads for a gene that is actually expressed)
- The same cell can not be sequenced repeatedly to quantify technical variability
- Variability between cells is often high
- Proper experimental design can be challenging

Higher noise level in scRNA-seq than in bulk

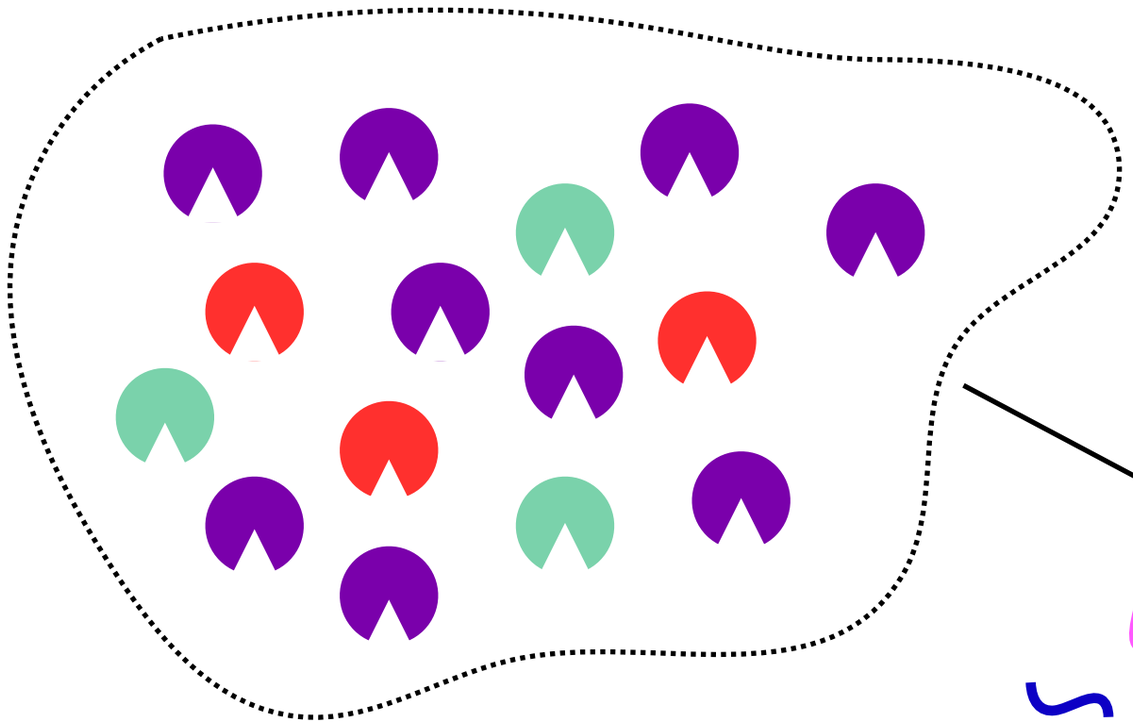


Data from GSE41265

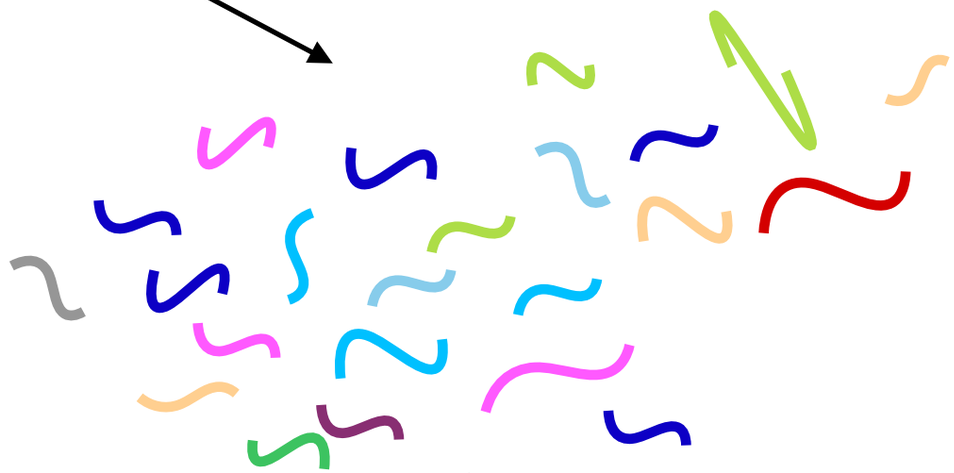
Long-read RNA-seq (intro)

Charlotte Soneson
CSAMA, Brixen, July 10 2018

“Bulk” RNA-seq



RNA extraction (polyA selection, ribosomal RNA depletion)



Fragmentation

cDNA synthesis Amplification

Sequencing Abundance quantification

gene	count
gene1	32
gene2	4
gene3	0
gene4	1928
...	

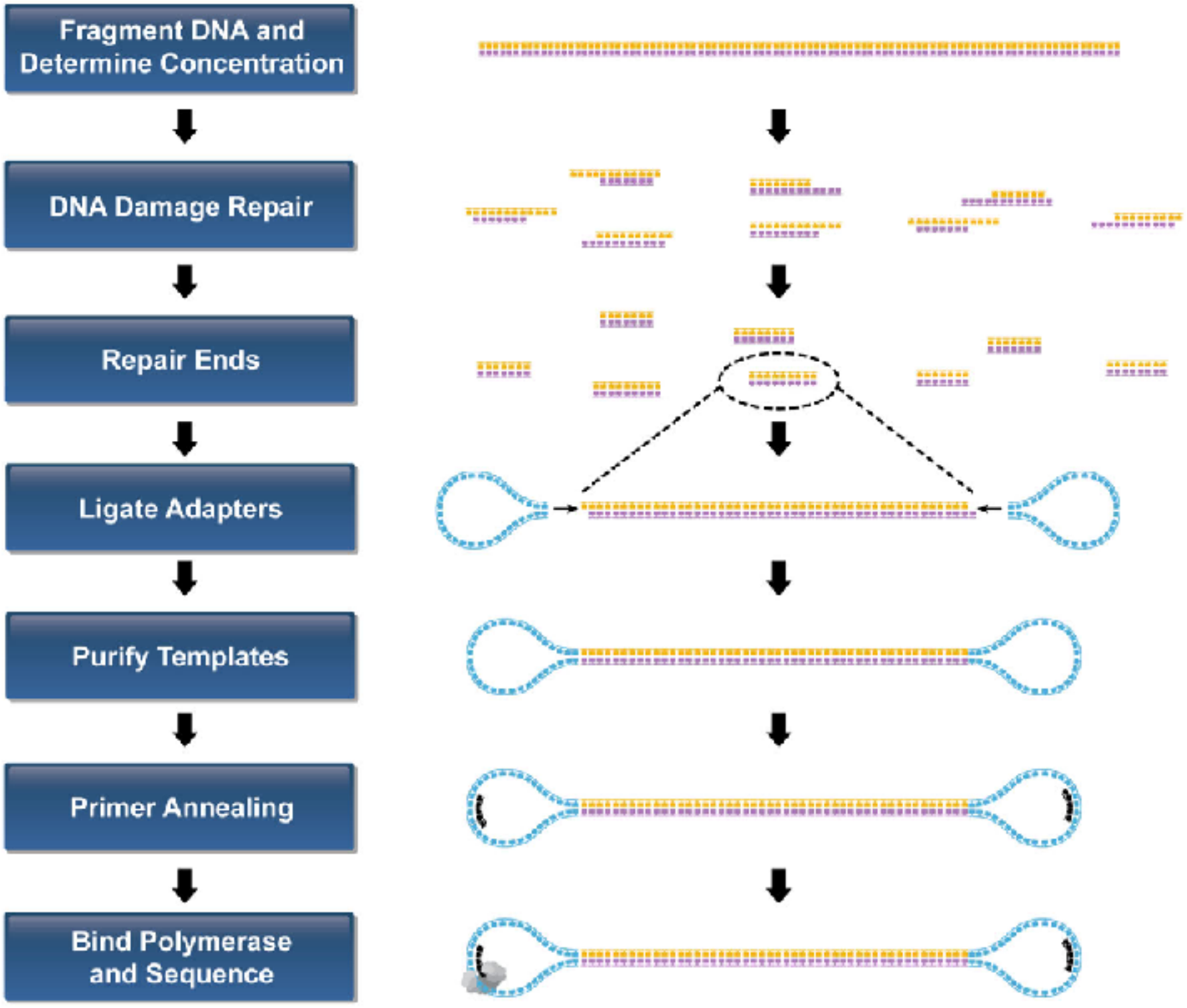
Why long reads?

- With Illumina sequencing, we need to fragment and amplify the (c)DNA before sequencing

- Application areas for long-read technologies:
  - genome assembly
  - metagenomics
  - transcriptomics
  - DNA/RNA modification detection

- Advantages for RNA-seq: identification of splice isoforms, transcriptome annotation

SMRT Sequencing (PacBio)



Kong et al., Standards in Genomic Sciences 2017

SMRT Sequencing (PacBio)

- SMRT cell containing tens of thousands of zero-mode wave guides (ZMWs) - small detection volumes

- ZMWs are illuminated from below

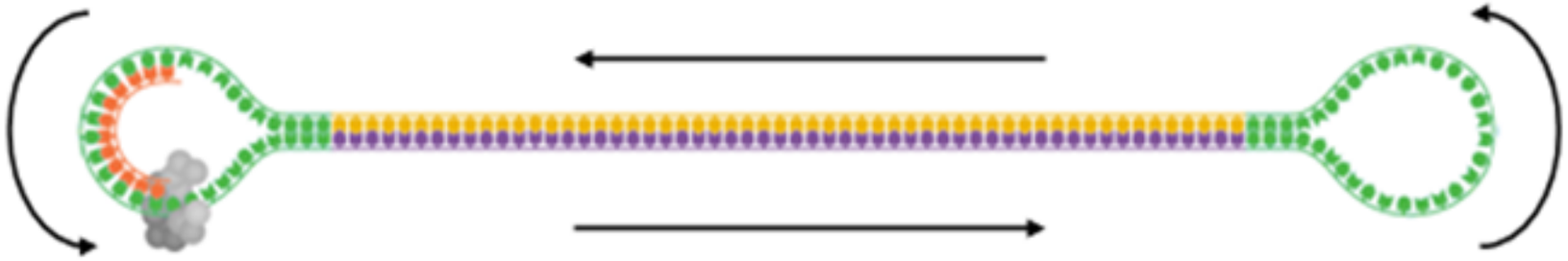
- A DNA template-polymerase complex is immobilized at the bottom of each ZMW

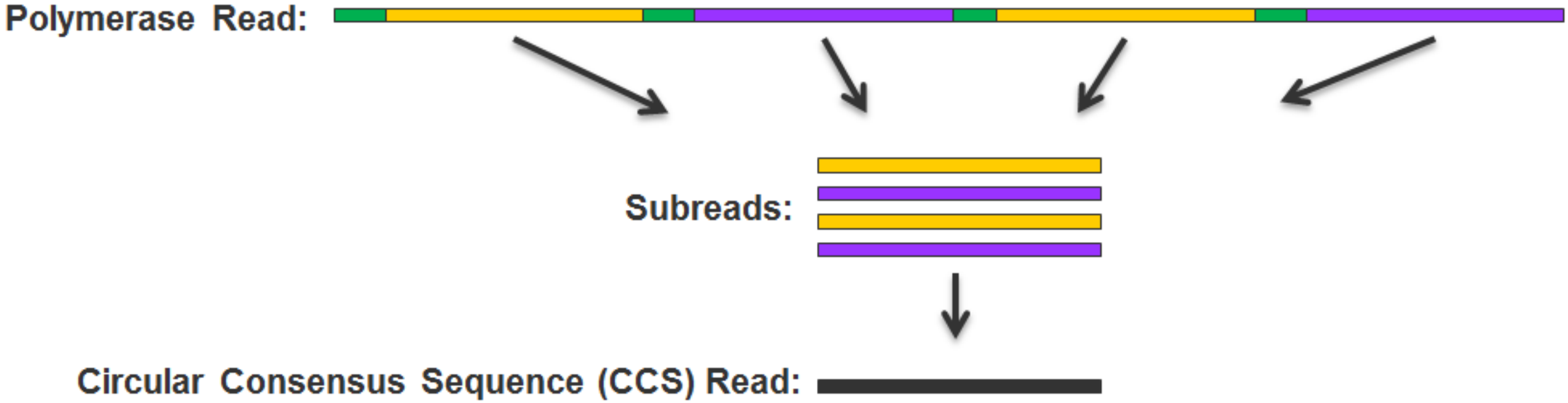
- Phospho-linked nucleotides are introduced in the ZMW

- As a base is held in the detection volume, a light pulse is generated

https://www.pacb.com/smrt-science/smrt-sequencing/, https://www.mscience.com.au/upload/pages/pacbio/pacbio_rs_ii_brochure.pdf

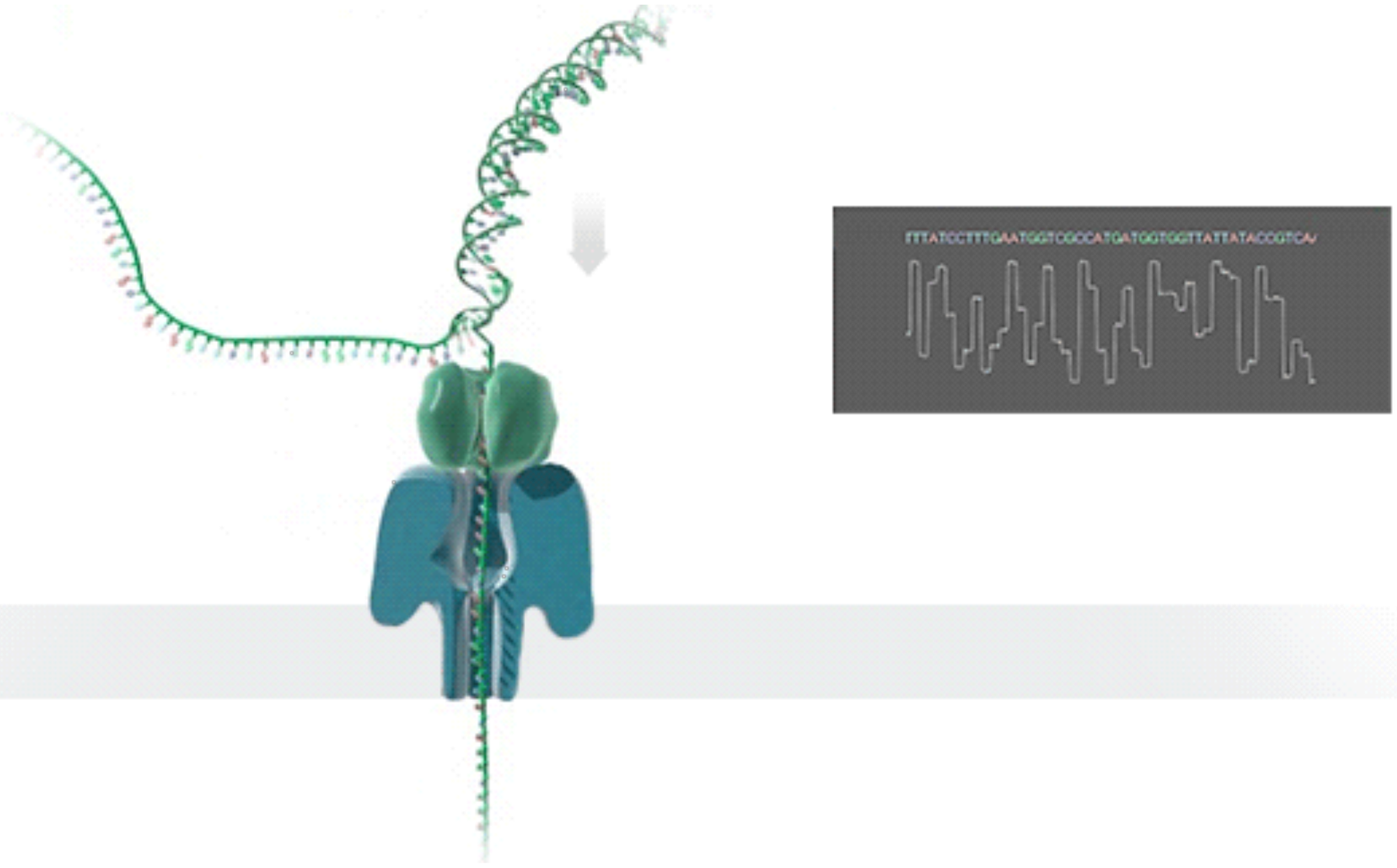
SMRT Sequencing (PacBio)





http://files.pacb.com/software/smrtanalysis/2.2.0/doc/smrtportal/help/!SSL!/Webhelp/Portal_PacBio_Glossary.htm

Oxford Nanopore sequencing



- Pass DNA or RNA through a nanopore
- The current is changed as the bases pass through the pore

https://nanoporetech.com/how-it-works

Challenges with long-read cDNA/RNA seq data

- Individual reads have high error rate (nanopore: 2-20%, PacBio subreads: 13-15%, PacBio CCS: 2%)
- Relatively low throughput
- Unequal read length
- Not all reads represent full-length transcript molecules
- How quantitative?