Generating Single-Cell Data

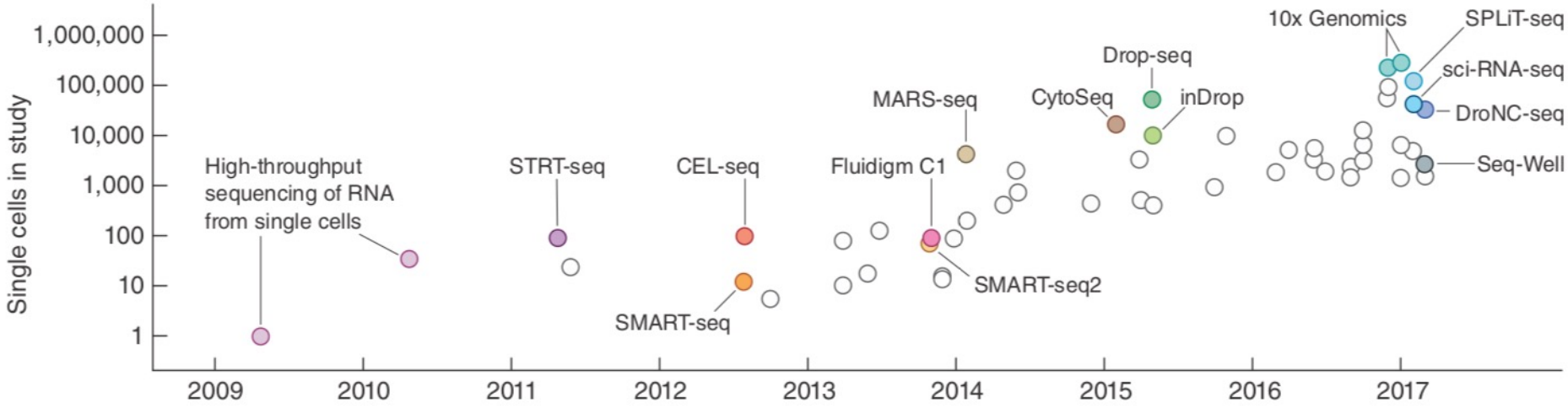
Technical overview of single-cell -omics methods.

1. **scRNA-seq families: Template Switch Approaches**
2. **Droplet-based scRNA-seq: 10x genomics (principle and sample preparation)**
3. **Alternative HT scRNA-seq approaches (combinatorial indexing / nanowells)**
4. **Other sc-omics approaches.**
5. **Integrated multi-omics approaches (scRNA-seq + xxx)**

Sébastien Smallwood – Head Functional Genomics FMI
sebastien.smallwood@fmi.ch

"disclaimer": my own views... what works best for us might not be the best for other labs...

Svensson *et al.* Nat Protocols, 2018



⇒ ***Many methods (60+)***

many "copies": small changes –> new name!

⇒ ***Common features / issues…***

still limited comparisons in 2022…

Resource for molecular biology / sequences:
initiative by Xi Chen (Teichman lab) https://github.com/Teichlab/scg_lib_structs

How to select the right protocol for a given biological question?

nature protocols

REVIEW ARTICLE

https://doi.org/10.1038/s41596-018-0073-y

Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies

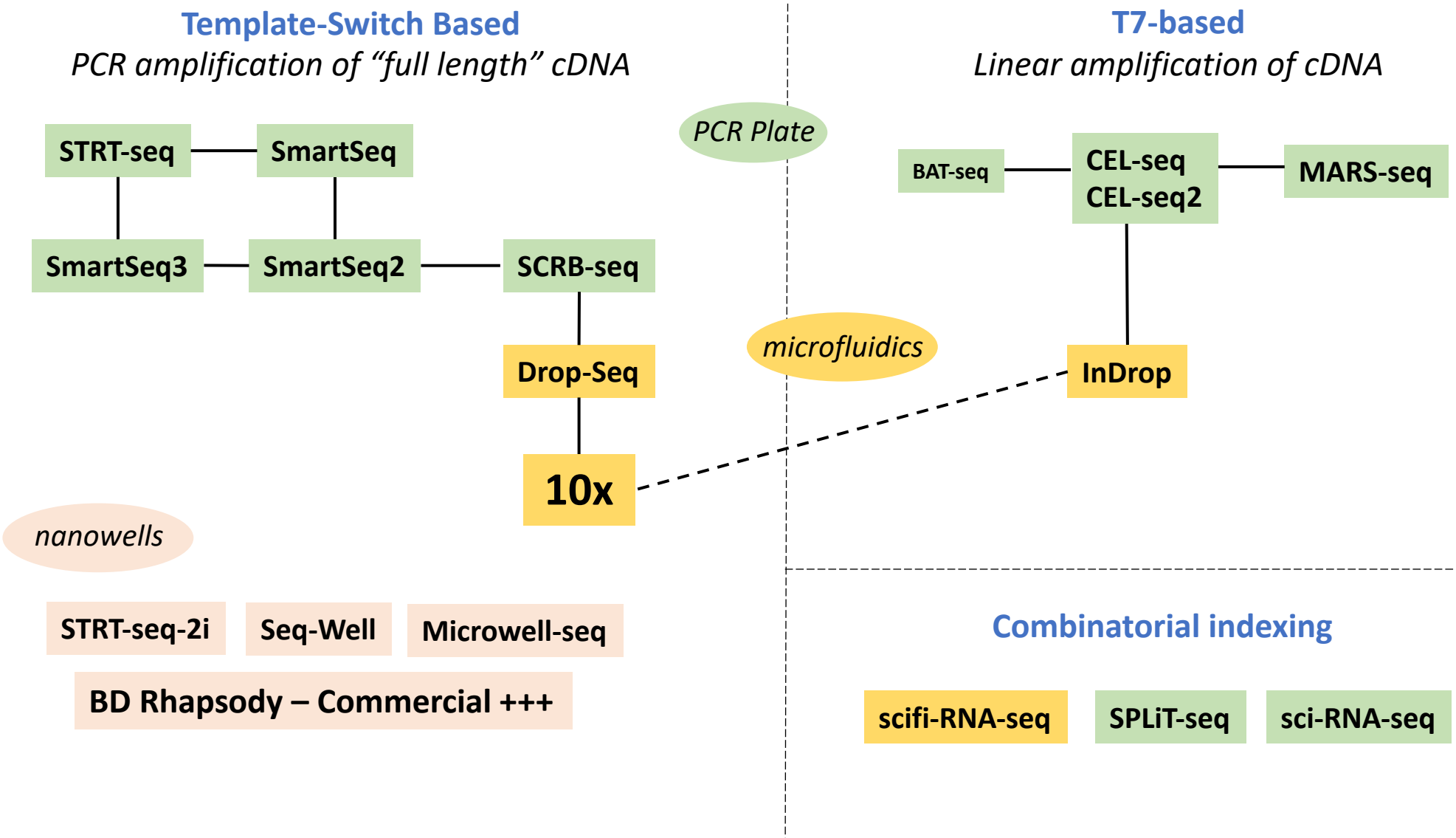
Atefeh Lafzi[1,5], Catia Moutinho[1,5], Simone Picelli[2,4], Holger Heyn[1,3*]

- **How many cells / nuclei do I need to sequence?**
- **Are samples precious or can sample collection be repeated easily?**
- **How can I collect cells / nuclei? How rapidly? How many cells can I collect in one experiment?**
  Can I store my samples and process them later?
- **Available equipment / Cost considerations...**

Plate-based approaches can still be relevant in 2022.
e.g. limited cells + precious sample

Collecting healthy single cells is CRITICAL for all methods.
Can be challenging and optimisations can be required.
in 2022 for the success of scRNA-seq, cell preparation is as important as the actual NGS processing.

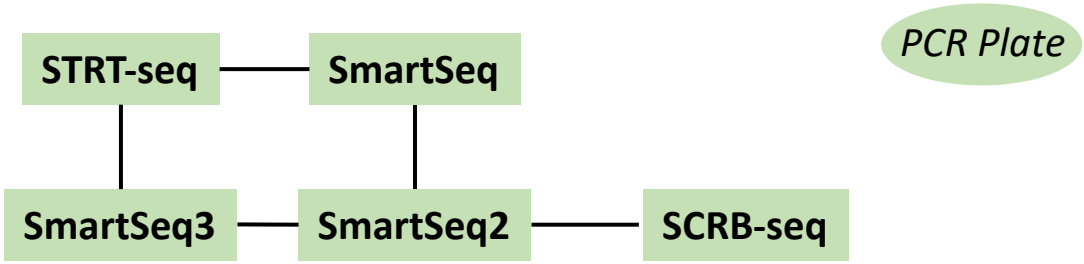
scRNA-seq Methods Families



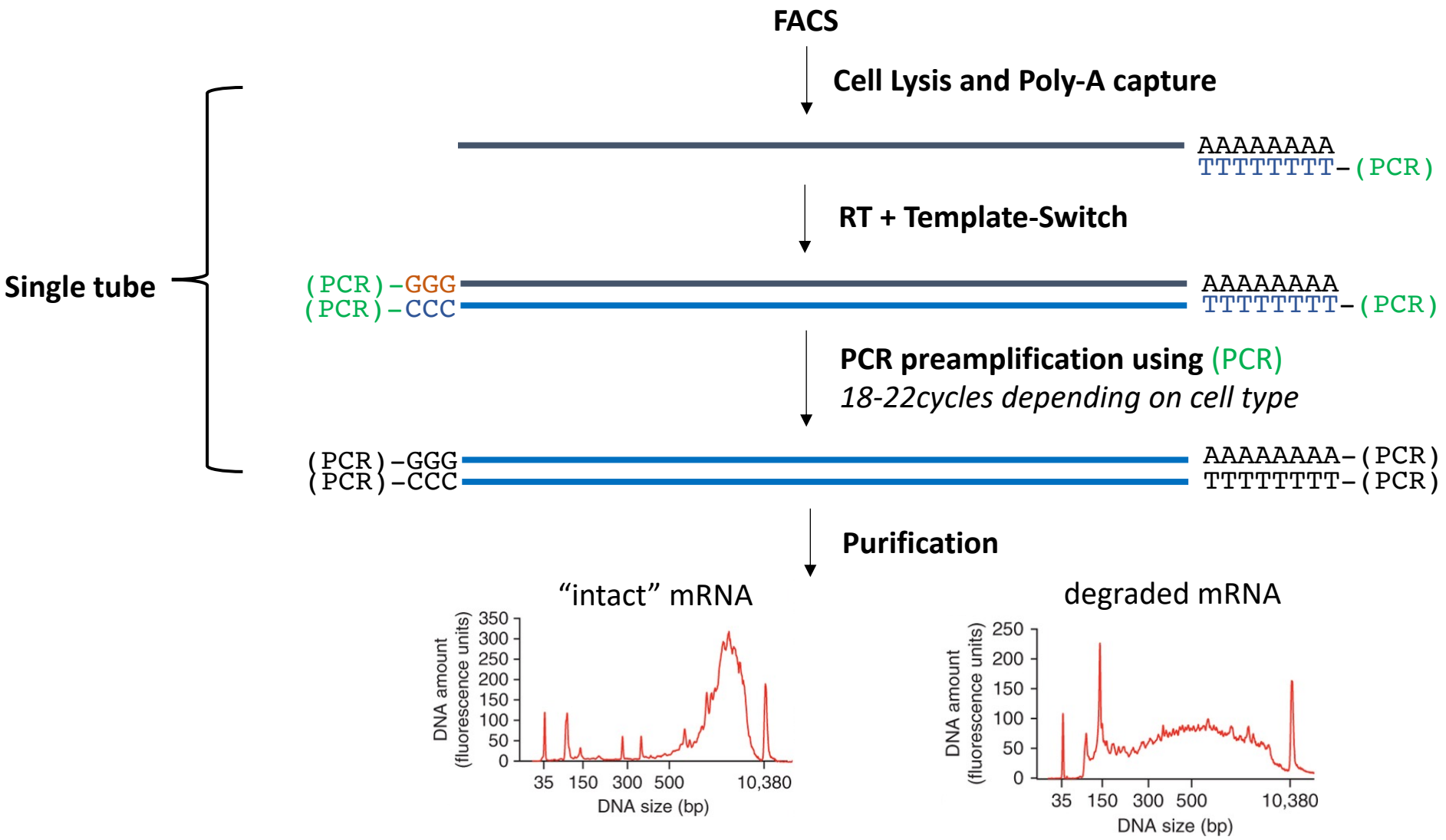
scRNA-seq Methods Families

Template-Switch Based

PCR amplification of "full length" cDNA

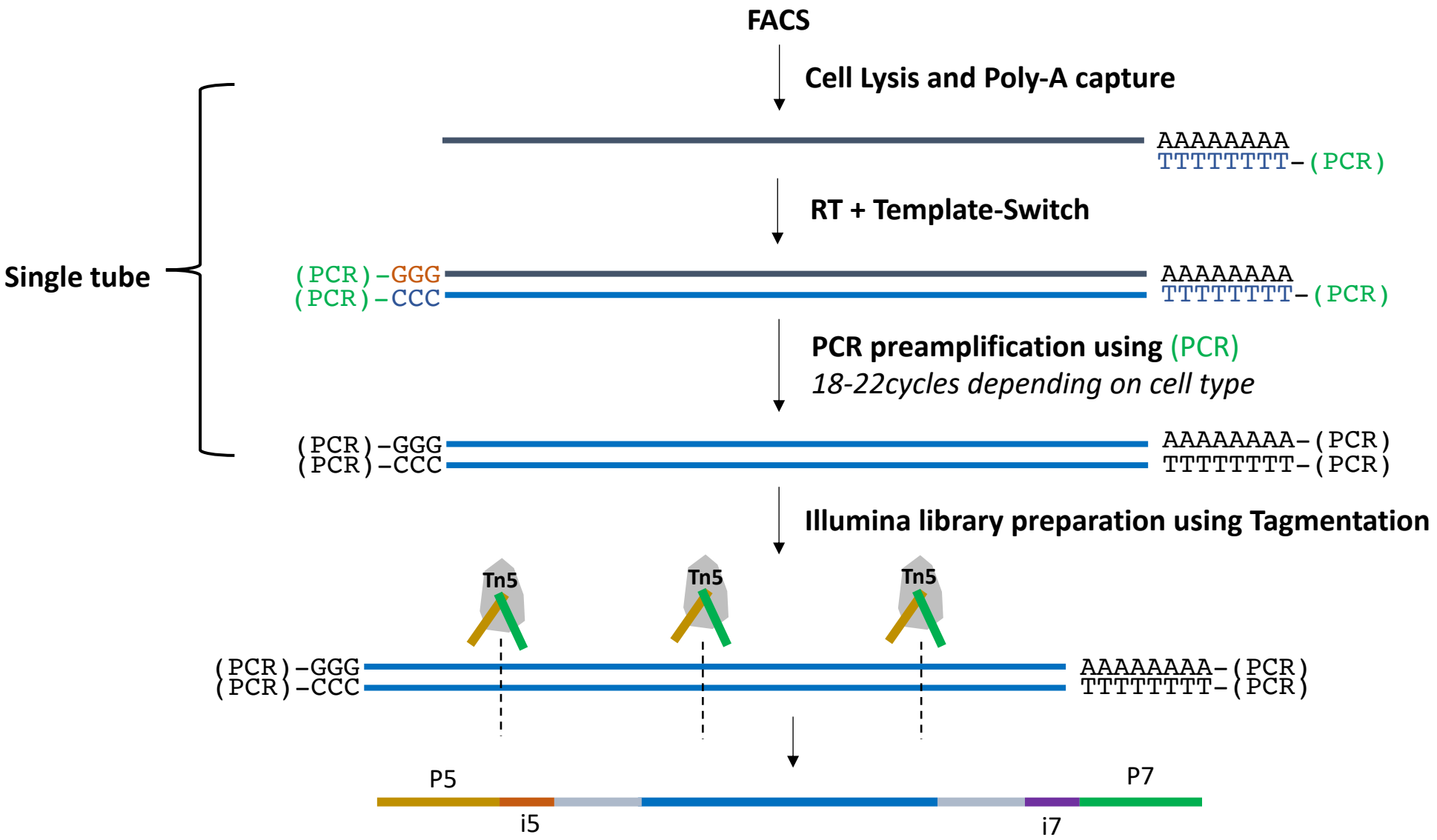


Template-Switch Based approaches with plate format: SmartSeq2



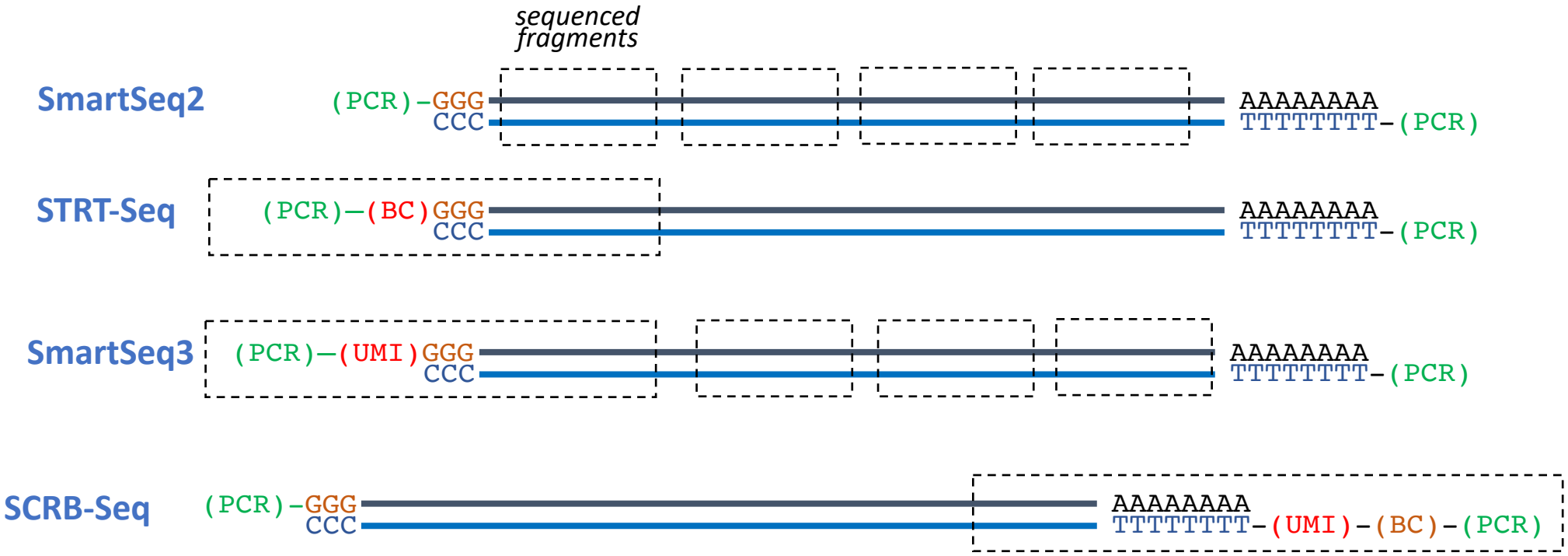
Picelli *et al.* Nat. Methods 2013

Template-Switch Based approaches with plate format: SmartSeq2



Picelli *et al.* Nat. Methods 2013

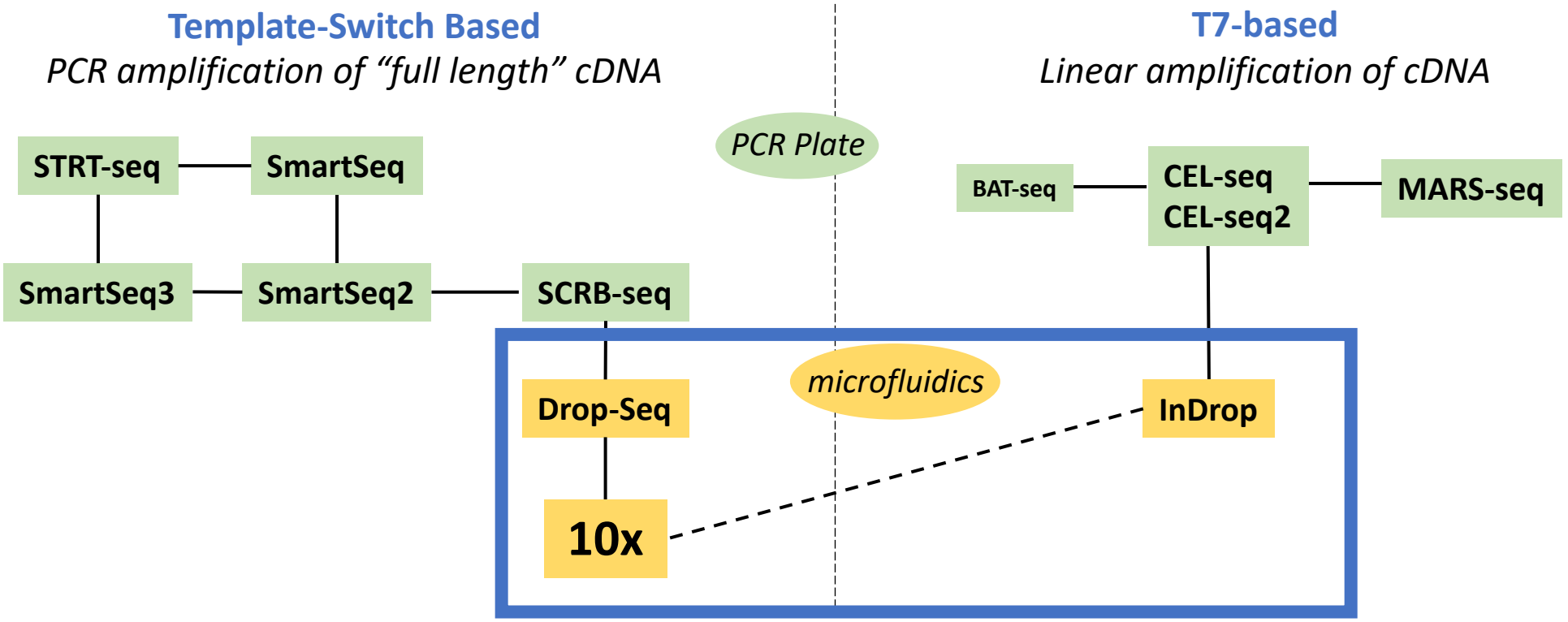
Template-Switch Based approaches: SmartSeq2 / STRT-Seq / SCRB-Seq



Labor intensive: *~1,000-6,000 cells per week depending on methods / resources (automation)*
Costly: *~ processing each cells in a separate tube / well. Plasticware + mol bio reagents $$$*

Islam *et al.* Genome Research 2011 | Soumillon *et al.* bioRxiv 2014 | Bagnoli *et al.* Nat Commun. 2018

scRNA-seq Methods Families

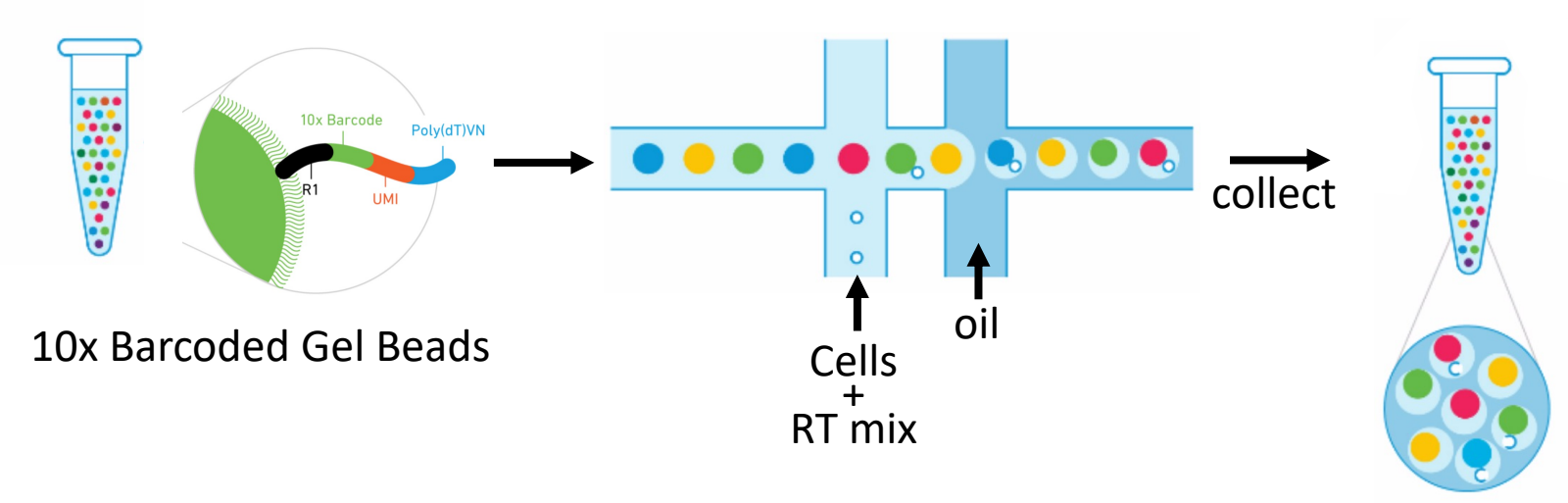


Microfluidics Devices.

-> cells are compartmentalised in nanoliter oil droplets

-> 1000s cells processed in parallel

10x Genomics Chromium



10x Genomics Chromium

1. Droplets collection and RT incubation

"GEMs" → Thermocycler -> start reverse transcription.

↓ Gel Beads dissolve and release barcoded dT oligos

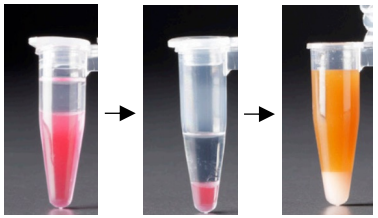
AAAAAAAA
TTTTTTTT[UMI][10xBC](PCR)

↓ RT + Template-Switch

(PCR)–GGG
(PCR)–CCC

AAAAAAAA
TTTTTTTT[UMI][10xBC](PCR)

2. Emulsion break-down and cDNA purification



3. PCR amplification in bulk

amplified full-length cDNA with UMI and Cell barcode information in 3'

(PCR)–GGG
(PCR)–CCC

AAAAAAAA[UMI][10xBC](PCR)
TTTTTTTT[UMI][10xBC](PCR)

10x Genomics Chromium

Pros: Ready-to-use kit -> robust process, rapid (2days without sequencing)
Reproducibility across labs / projects +++ (unlike Drop-seq / InDrop).
High quality data (although slightly lower than some plate based methods).
Good cell capture rate (~50%)
Very High Throughput. Chromium run takes ~15min

Cons: Not so convenient / cost effective for low cells numbers.
Price per cell is low but experiments are very expensive…only sequencing to know the success.
Doublets can be an issue.

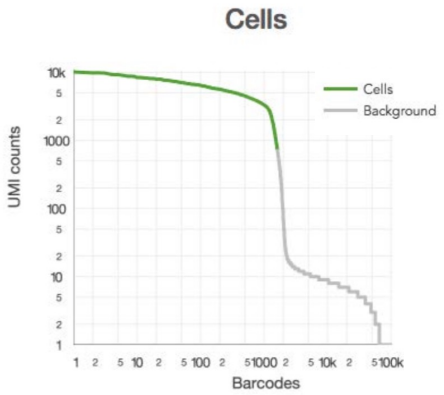
Cell preparation is critical.

Cell counting needs to be accurate and reproducible (new cell types can be challenging).

Cost SCRB-seq (plate): ~3CHF per cell with sequencing.
Cost 10x genomics: ~0.6CHF per cell with sequencing.

10x Genomics Chromium

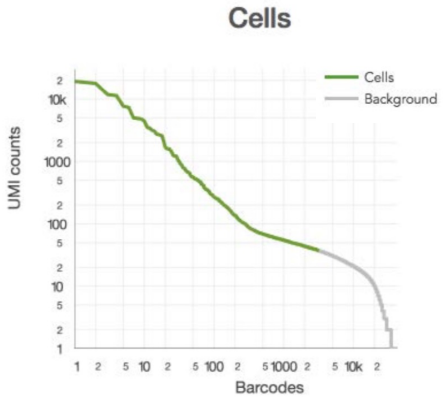
Typical Profile



Defined cliff and knee

Metric	Value
Barcodes	>90,000
Cell Barcodes	>1,000
UMIs	>10,000

Ambient RNA +++ -> cell death



Lack of defined cliff and knee

Metric	Value
Barcodes with >1000 UMIs	Few

10x Genomics: Samples Preparation Considerations

Sample Preservation / Storage.

- Using fresh cells is the best option when possible. Requires good logistics.
- Methanol fixation is a reported option for isolated cells, although no all-rounder protocol
  requires optimisation for each cell type (can be challenging). never worked well in our hands.
- Cryopreservation works well for cultured / primary cell lines.
- For tissue: flash-freezing + nuclei preparation can be a good option (e.g. brain)

Sample Purity.

- Single Cell suspension (no clumps)
- Aggregates and debris needs to be removed (microfluidics channels are narrow <100µm).
- Cells should be healthy:
  - 90% viable cells. dead cells -> ambient RNA issue
  - Some cell types are more fragile than others, and cell viability decreases with time.
    workflows need speed
    new cell types can be challenging / behave differently:
    Optimisation + pilot before large experiment

possibility to remove debris / aggregates / dead cells by a combination of filters and mag. beads not really reproducible + cell loss.

FACS +++ fast / only single cells / live-dead sorting / FACS washes the cells ☺
test that cells are happy post FACS
from FACS to 10x “straight away”(depends on your lab environment)

10x Genomics: Samples Preparation Considerations

Cell Counting

can be tricky and frustrating (only sequencing tells you the number of cells...money + time spent already...)

- accurate quantification of cell concentration is required
  - too many cells loaded -> doublets
  - too little cells loaded -> expensive + not enough cells for biological questions
    large time course with many samples....

- capture rate (difference cells in / out) of 10x is officially 65%
  capture rate is “constant / reproducible”
  in practise the capture rate depends on the cell type to some extent.

- How to count cells?
  manually (haemocytometer): prone to error – know your cells well!
  automated cell counter.
    different counters can generate different results
    some cell types like some cell counters better than others
    very challenging for small cells / nuclei
  FACS.
    bad ”press”
    FACS count does not match the number of cells landing in the tube ultimately.
    Works well in our hands with a ”correction” factor.
    In our hands, FACS count provide the most reproducible results across projects / cell types

scRNA-seq Methods Families

Template-Switch Based

PCR amplification of "full length" cDNA

PCR Plate

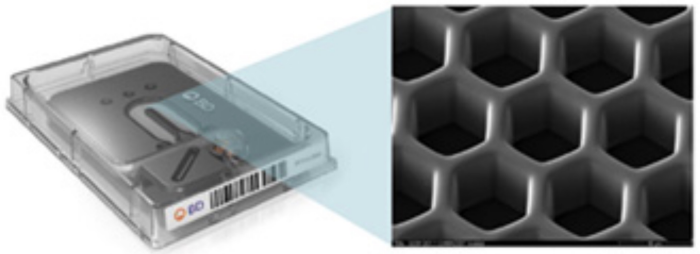
STRT-seq — **SmartSeq**

SmartSeq3 — **SmartSeq2** — **SCRB-seq**

Drop-Seq

10x

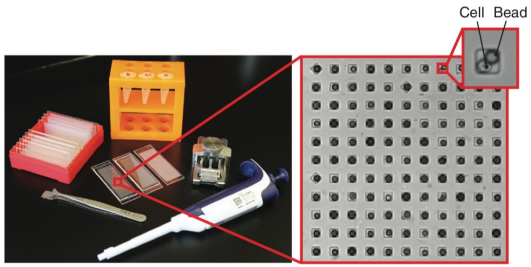
microfluidics



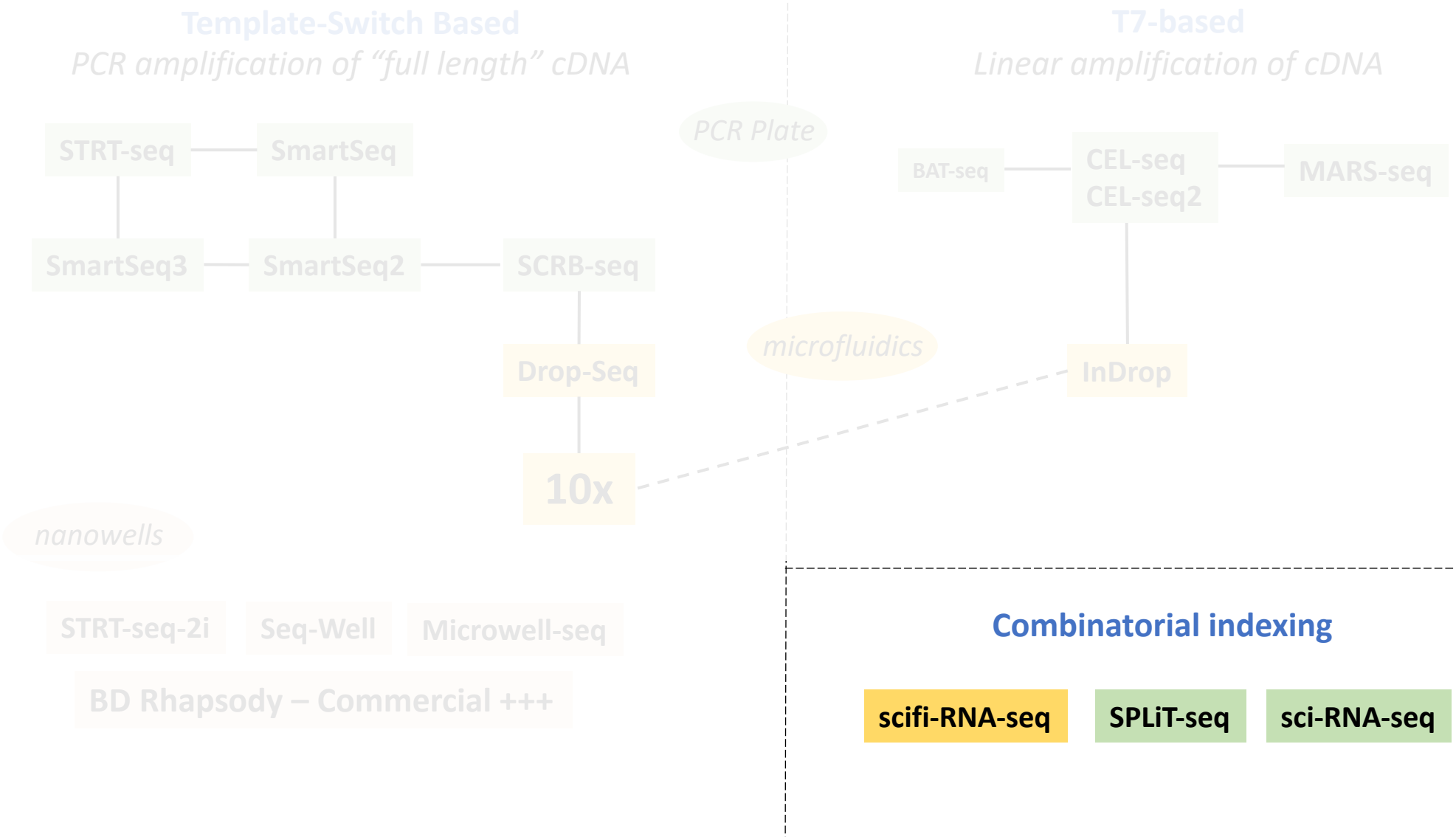
nanowells

STRT-seq-2i **Seq-Well** **Microwell-seq**

BD Rhapsody – Commercial +++

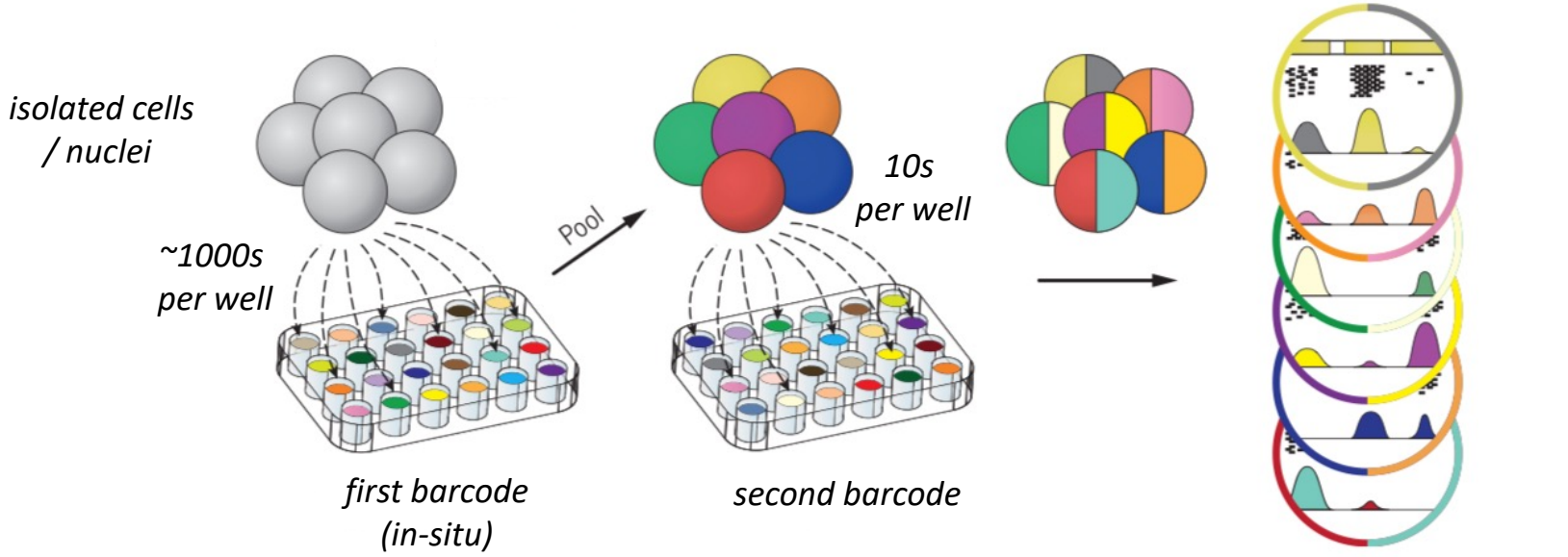


scRNA-seq Methods Families



Combinatorial Indexing Approaches

Shendure & Trapnell labs …..



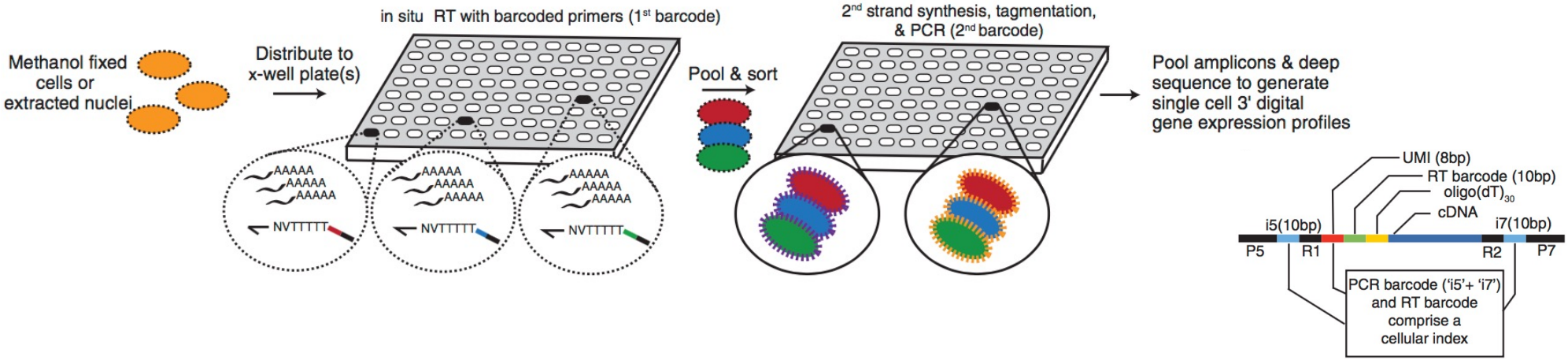
Getting large number of cells without single cell compartmentalisation, only basic labware

General CONS: a lot of cells are needed.

doublets = barcodes collision can be an issue (5-10%)

Combinatorial Indexing Approaches: sci-RNA-seq

Cao *et al*. Science 2017



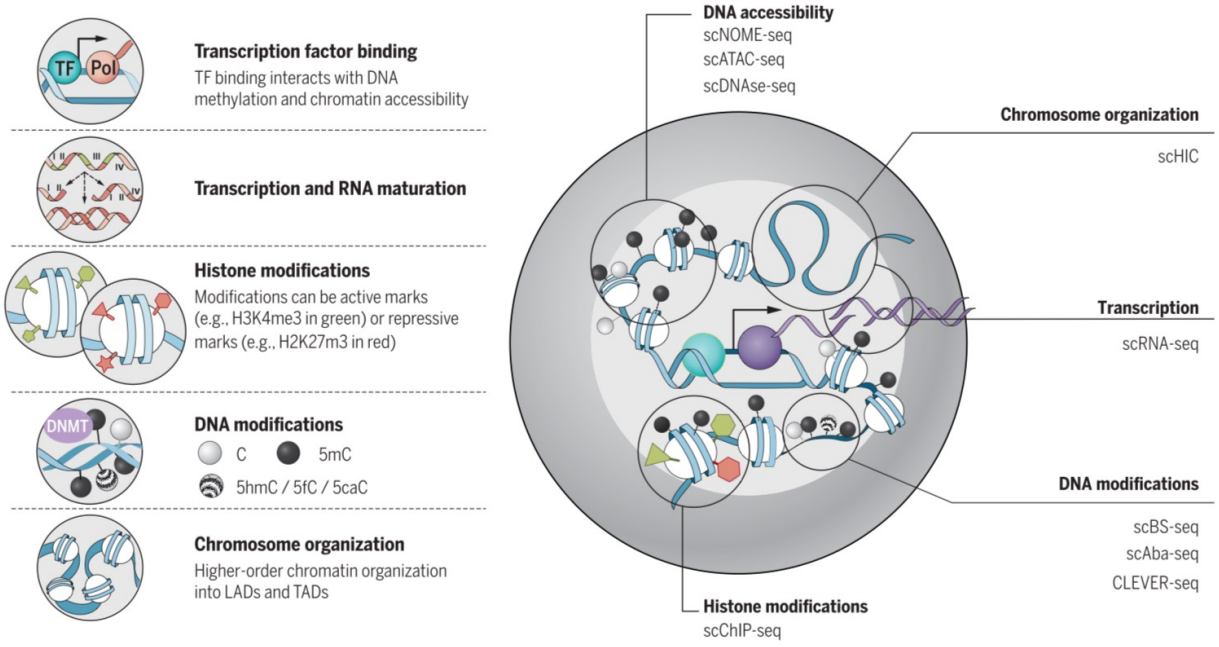
~4500 genes per mouse cell (3T3) (~similar to 10x)
Cao 2017: can process 10,000s per experiments
Cao 2019: sci-RNAv3: 2M cells in one experiment (3 rounds of barcoding)

Pros: - do not require expensive / complicated microfluidics instruments.
- compatible with cells and nuclei. Need fixation so storage +++
Cons: a lot of cells are required.

SPLiT-seq: Rosenberg et al. Science 2018

Not only mRNA.

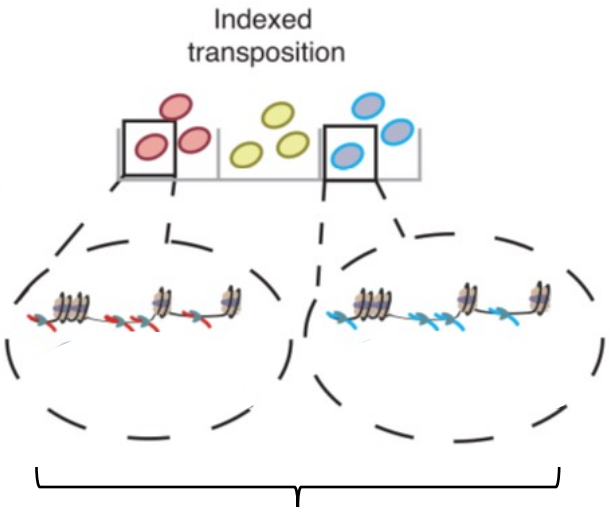
- gDNA
- small RNA-seq, total RNA-seq
- ATAC-seq
- BS-seq (DNA methylation)
- ChIP-seq
- scHiC
- …



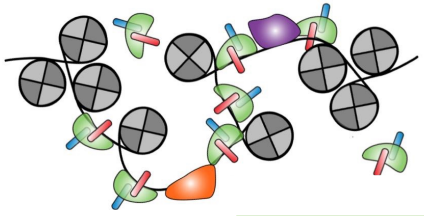
Kelsey et al. Science 2018

Single Cell Chromatin Accessibility (scATAC-seq)

sci-ATAC-seq

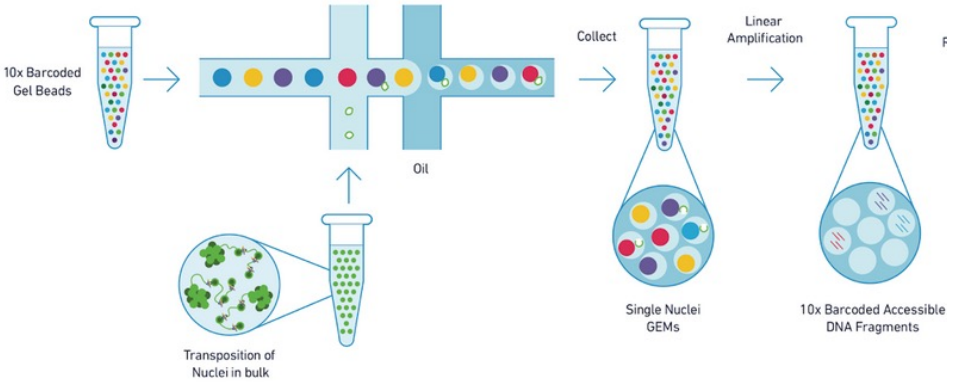


Pool and Split



Tn5

10X scATAC-seq




```
10x beads (P5)[BC]TCGTCGGCAGCGTC
          Tn5 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGXXXXXXXXXXXXX
                            TCTACACATATTCTCTGTC           XXX
```

25-35K unique reads per cell (? v2 coming up)

Cusanovich et al. Science 2015

coverage of 0.3-3% of DHS sites

Cusanovich et al. Cell 2018

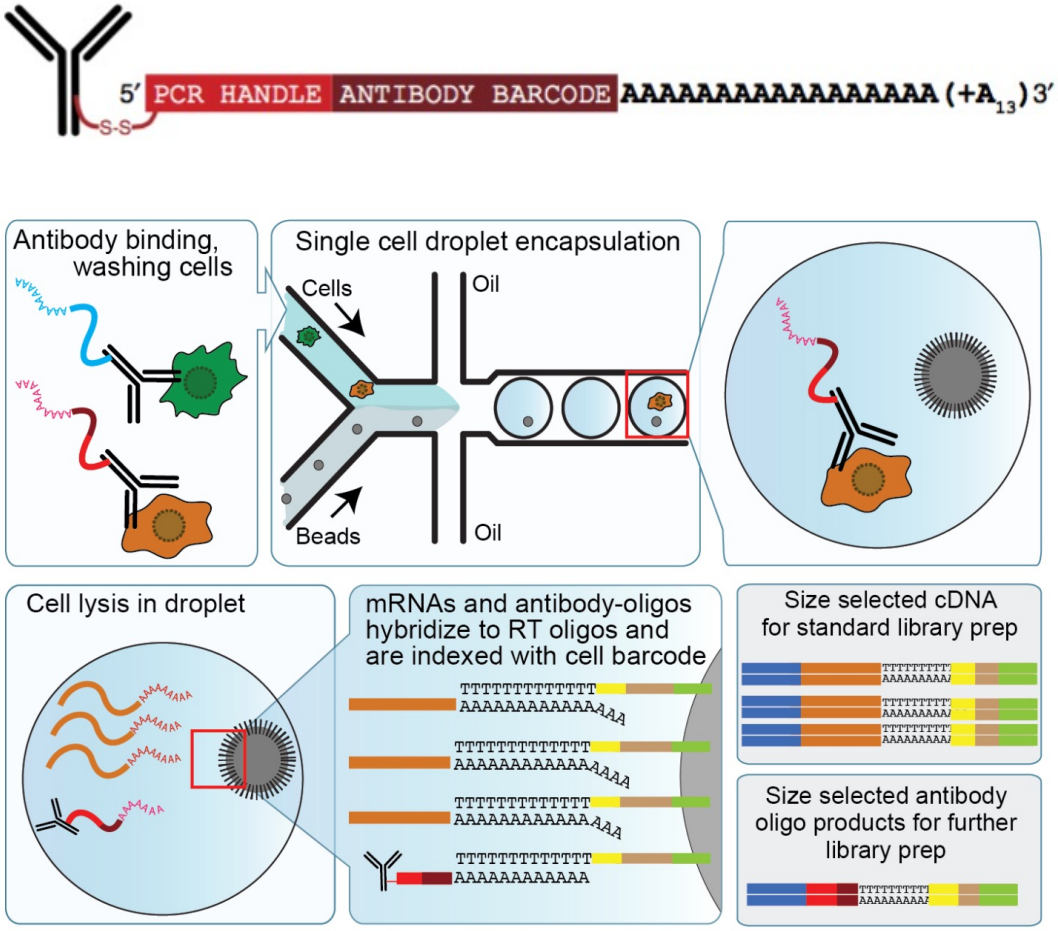
10-25K unique reads per cell (40% in bulk ATAC peaks)

CONS: sparse data + high doublets rate (5-10%)

INTEGRATED MULTI-OMICS:

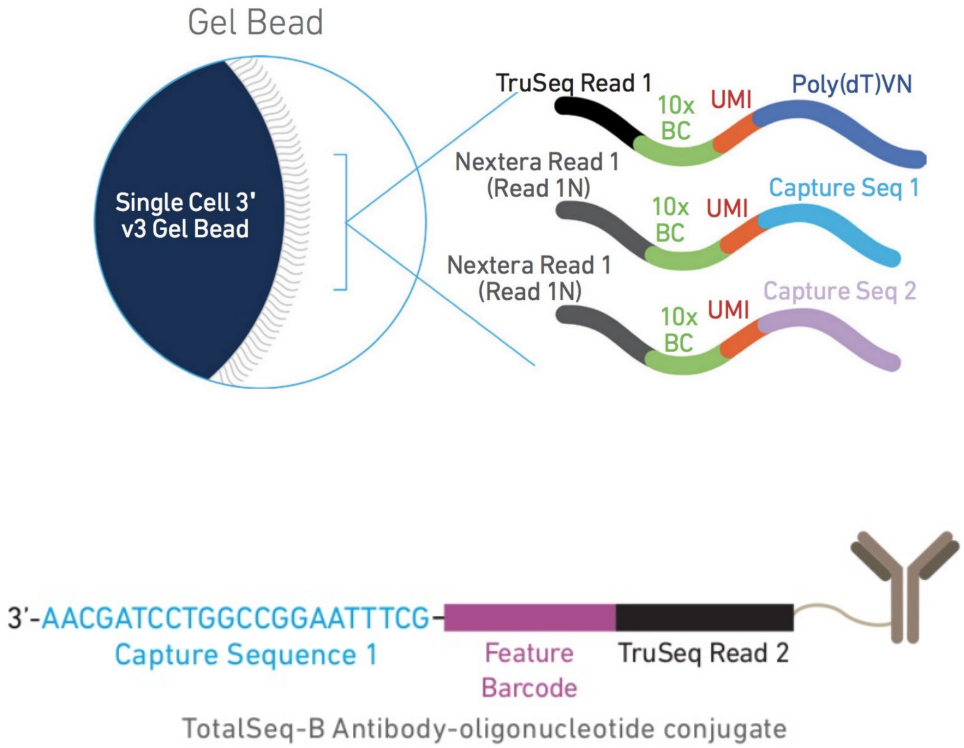
transcriptome + another layer of information from the same cell

CITE-seq / REAP-seq: Cell surface markers + transcriptome



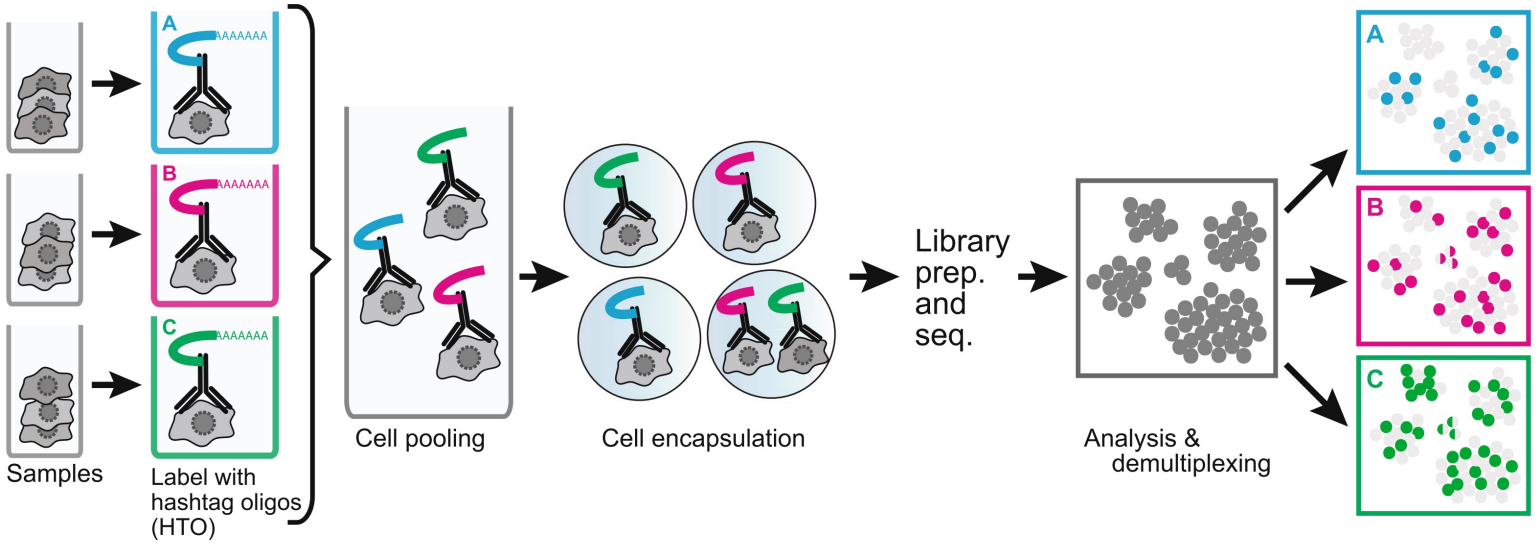
Stoeckius *et al.* Nat. Methods 2017
Peterson *et al.* Nat Biotechnology 2017

CITE-seq / REAP-seq: Cell surface markers + transcriptome



Many commercially available Ab (Hu + Ms)

Cell Hashing / Sample Multiplexing



-> individual sample labelling: replicates, batch effects...
-> “super-load” 10x genomics and discard doublets...

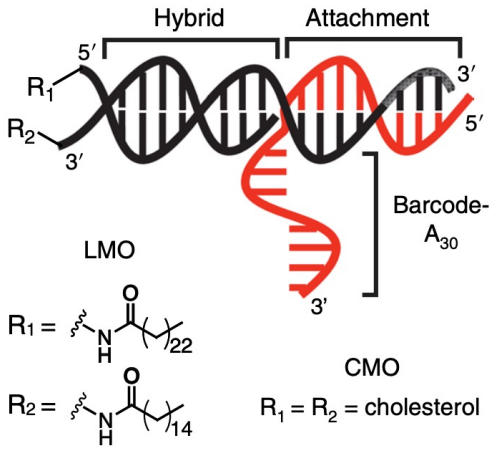
Stoeckius *et al.* Genome Biology 2018

Cell Hashing / Sample Multiplexing

Antibody-free solutions: species-agnostic, lipid-based tagging
-> barcodes embedded into the **cells or nuclei** lipid membrane

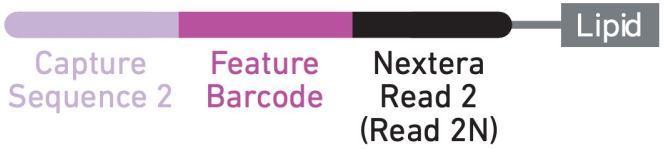
MULTI-Seq

McGinnis *et al.* Nat. Methods 2019



CellPlex 10x genomics Solution

CMOs

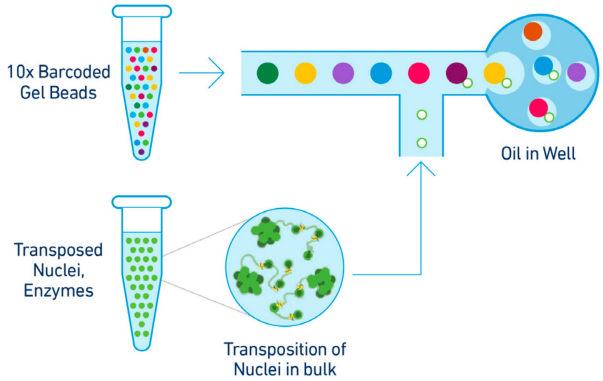


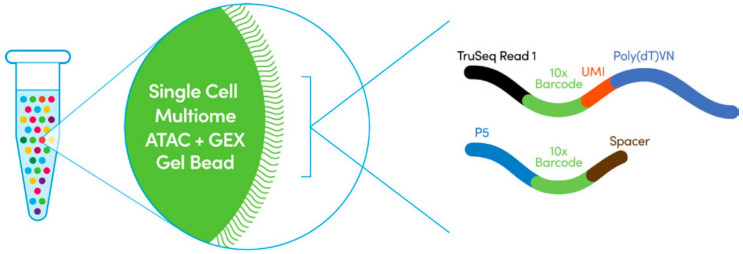
In practise: ~18K cells / nuclei per 10x channel

CONS: require more input cells
depending on the experimental set-up:
100K-800K cells per sample (FMI protocol)

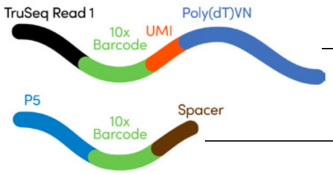
Chromatin Accessibility + Transcriptome

10x genomics Multiome: combined scATAC-seq and snRNA-seq (nuclei)





in droplets



mRNA capture + RT

transposed DNA capture

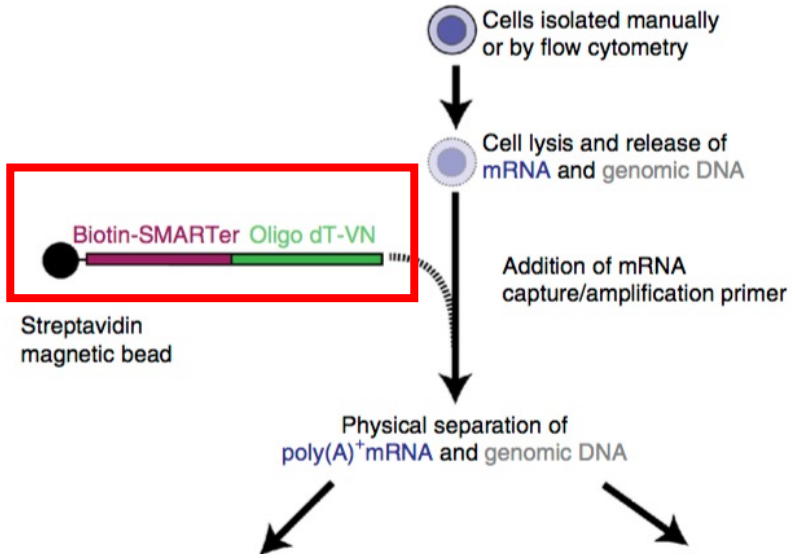
in bulk

Pre-amplification
cDNA +
transposed DNA

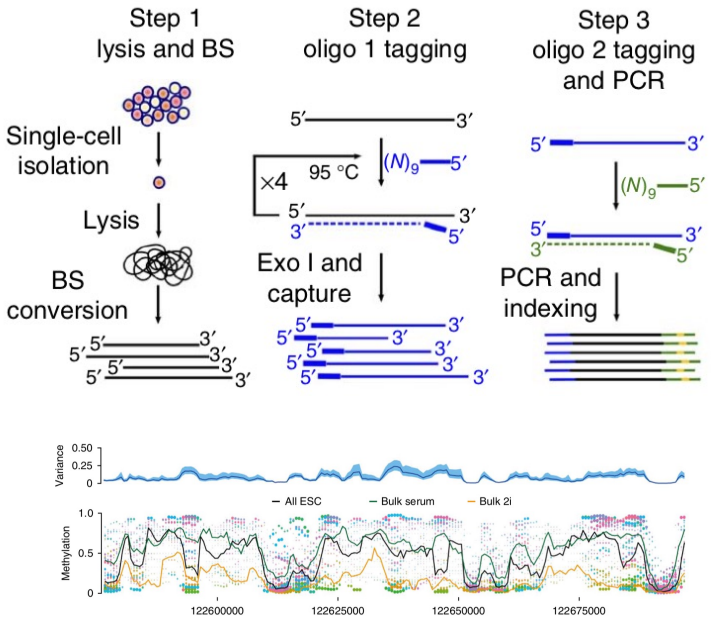
RNA-seq library prep
and sequencing

ATAC-seq library prep
and sequencing

scM&T-seq: Transcriptome and DNA methylome



scBS-seq (PBAT based)



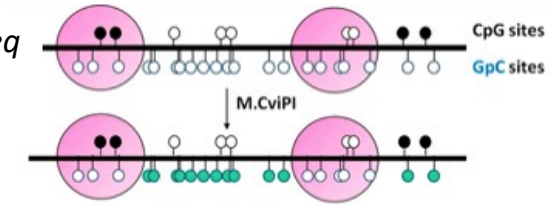
Pros: Quality is high (5000-6000 genes + 20-25% of the DNA methylome)

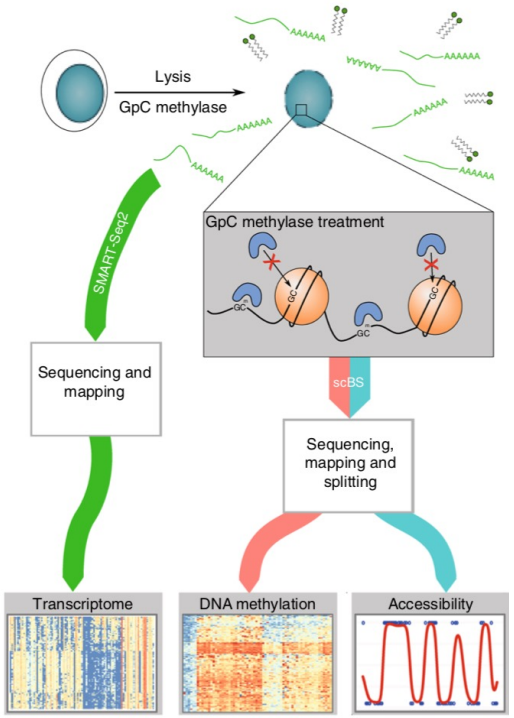
Cons: Relative Low-Throughput + Expensive

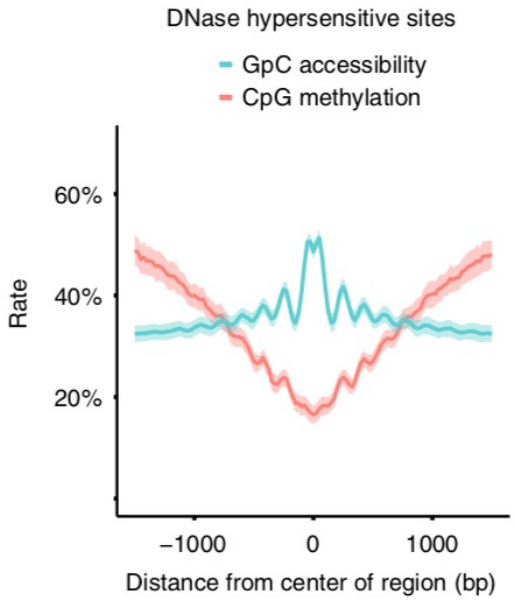
scBS-seq: Smallwood *et al.* Nat. Methods 2014
scG&T: Macaulay *et al.* Nat. Methods 2015
scM&T: Angermueller *et al.* Nat. Methods 2016

scNMT: transcriptome + DNA methylation + chromatin accessibility

NOME-seq





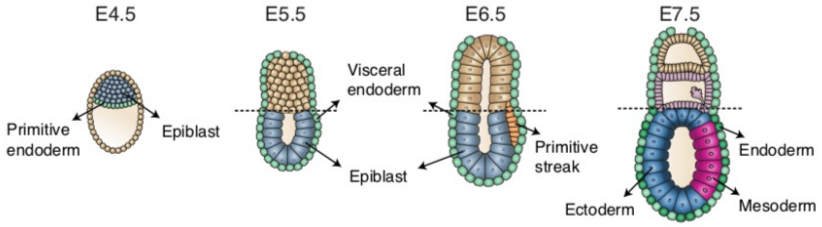


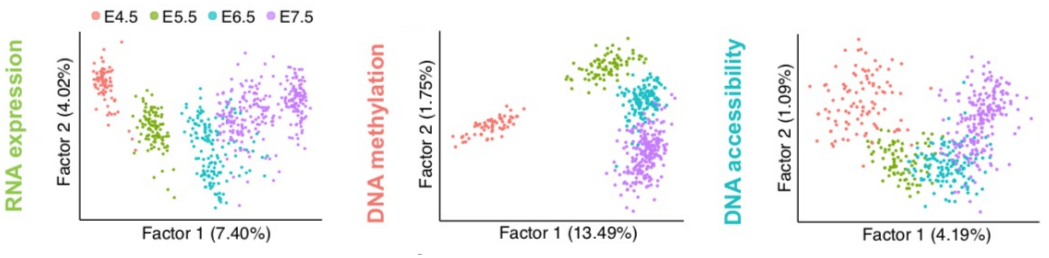
~75% of promoters, ~60% of enhancers are captured in a typical cell

Cons: very expensive: ~50CHF per cell with sequencing.

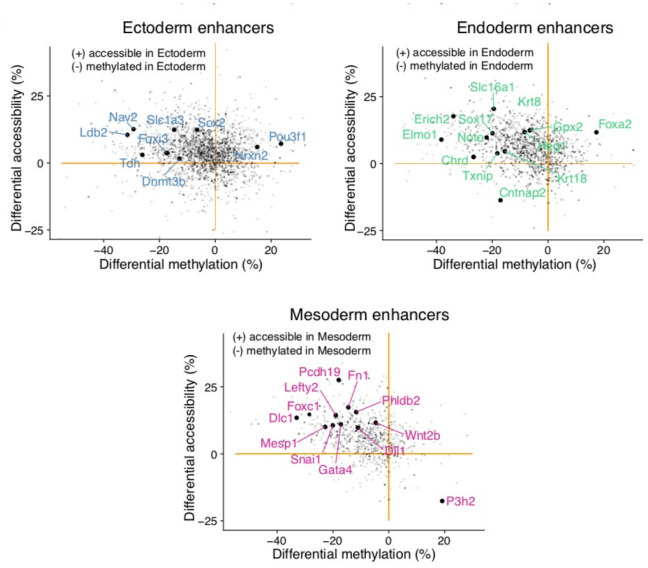
Clark *et al.* Nat Communication 2018

scNMT: transcriptome + DNA methylation + chromatin accessibility

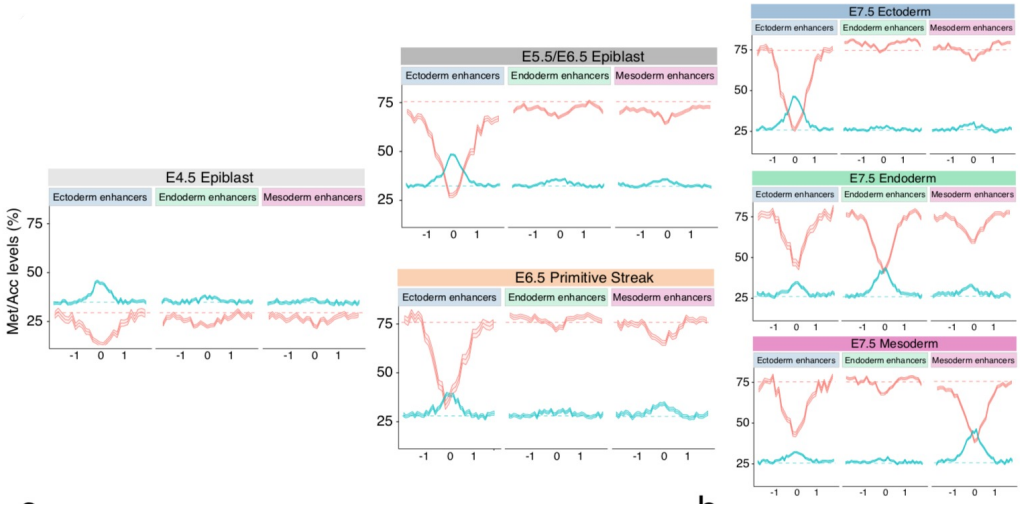




Identify cell lineages (RNA) and specific enhancers



Follow enhancers activity throughout development.



the epigenetic landscape of ectodermal cells is already established in the early epiblast.

Argelaguet *et al.* Nature

Thanks for listening……

Sébastien Smallwood – FMI
sebastien.smallwood@fmi.ch