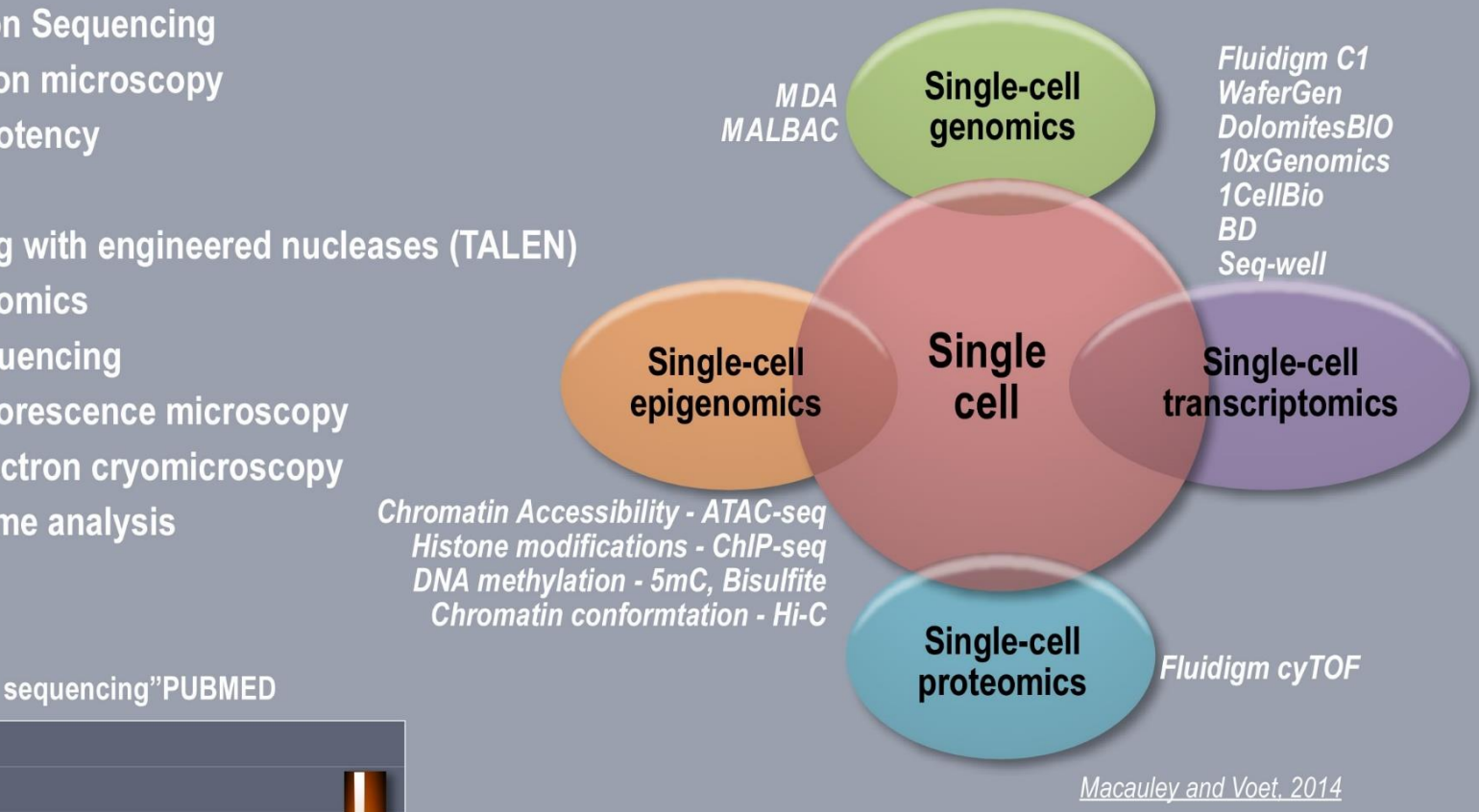
# Singel Cell RNA-seq: Technologies and Experimental Approaches

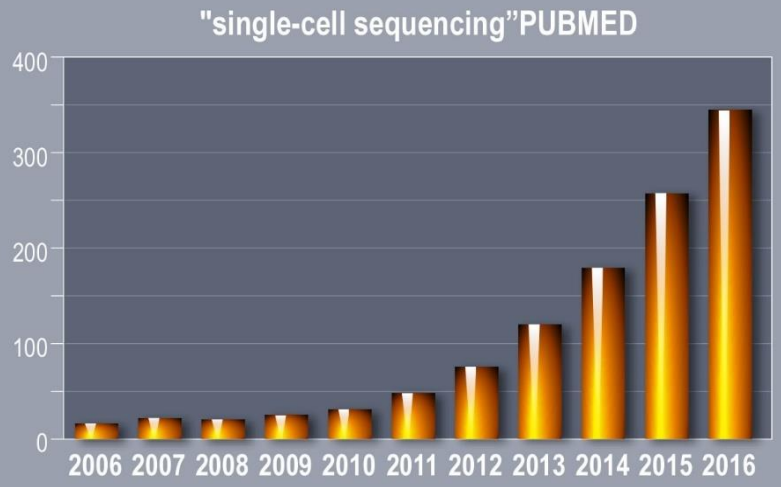
Kévin Lebrigand
UCAGenomix, Nice-Sophia-Antipolis
lebrigand@ipmc.cnrs.fr
@kevinlebrigand
Roscoff, 18 juin 2018

# Methods of the year 2013 (Nature Methods)

2007: Next Generation Sequencing
2008: Super-resolution microscopy
2009: Induced pluripotency
2010: Optogenetics
2011: Genome editing with engineered nucleases (TALEN)
2012: Targeted proteomics
2013: Single cell sequencing
2014: Light-sheet fluorescence microscopy
2015: Cryo-EM or electron cryomicroscopy
2016: Epitranscriptome analysis

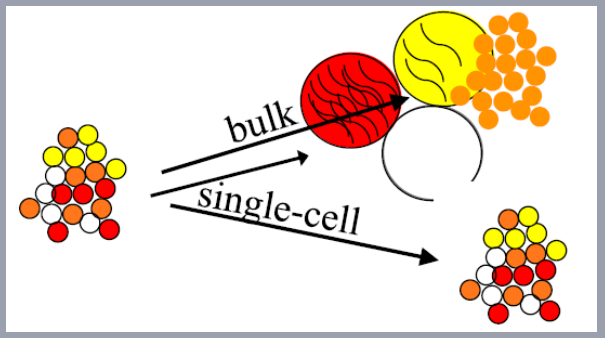



Macauley and Voet, 2014



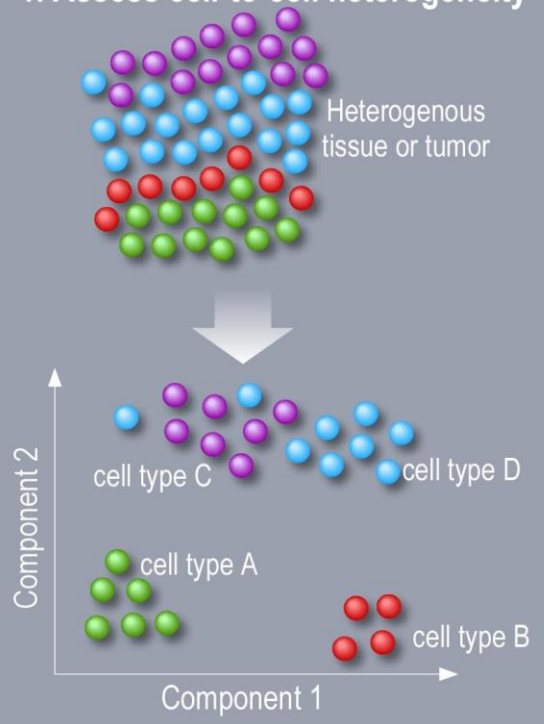
# Why single cell profiling?

Stop measuring gene average as in bulk,
Population sequencing yields average values,
Changes in subpopulation might remain undetected,

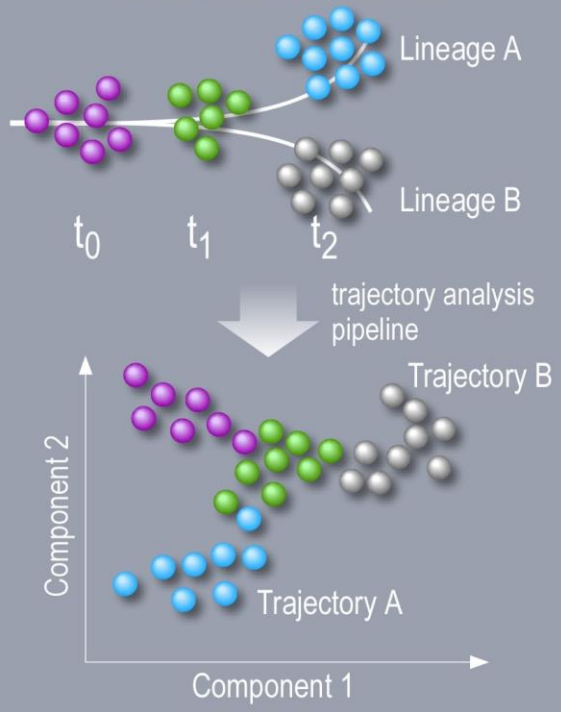



# Why single cell profiling?

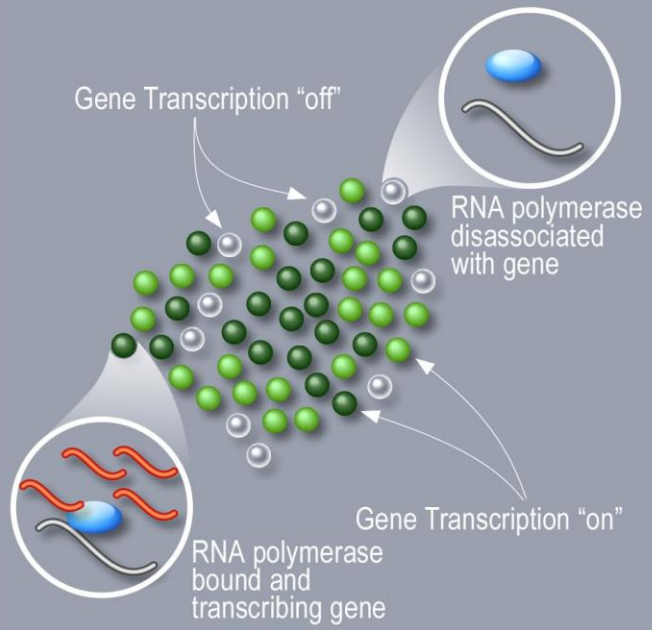
## 1. Assess cell-to-cell heterogeneity



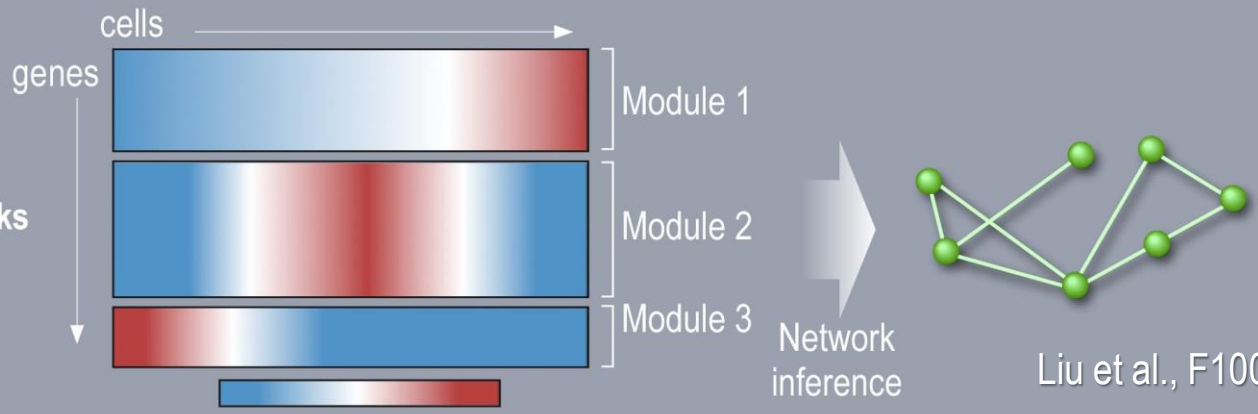

## 2. Map cell trajectories




## 3. Dissect transcriptional mechanics

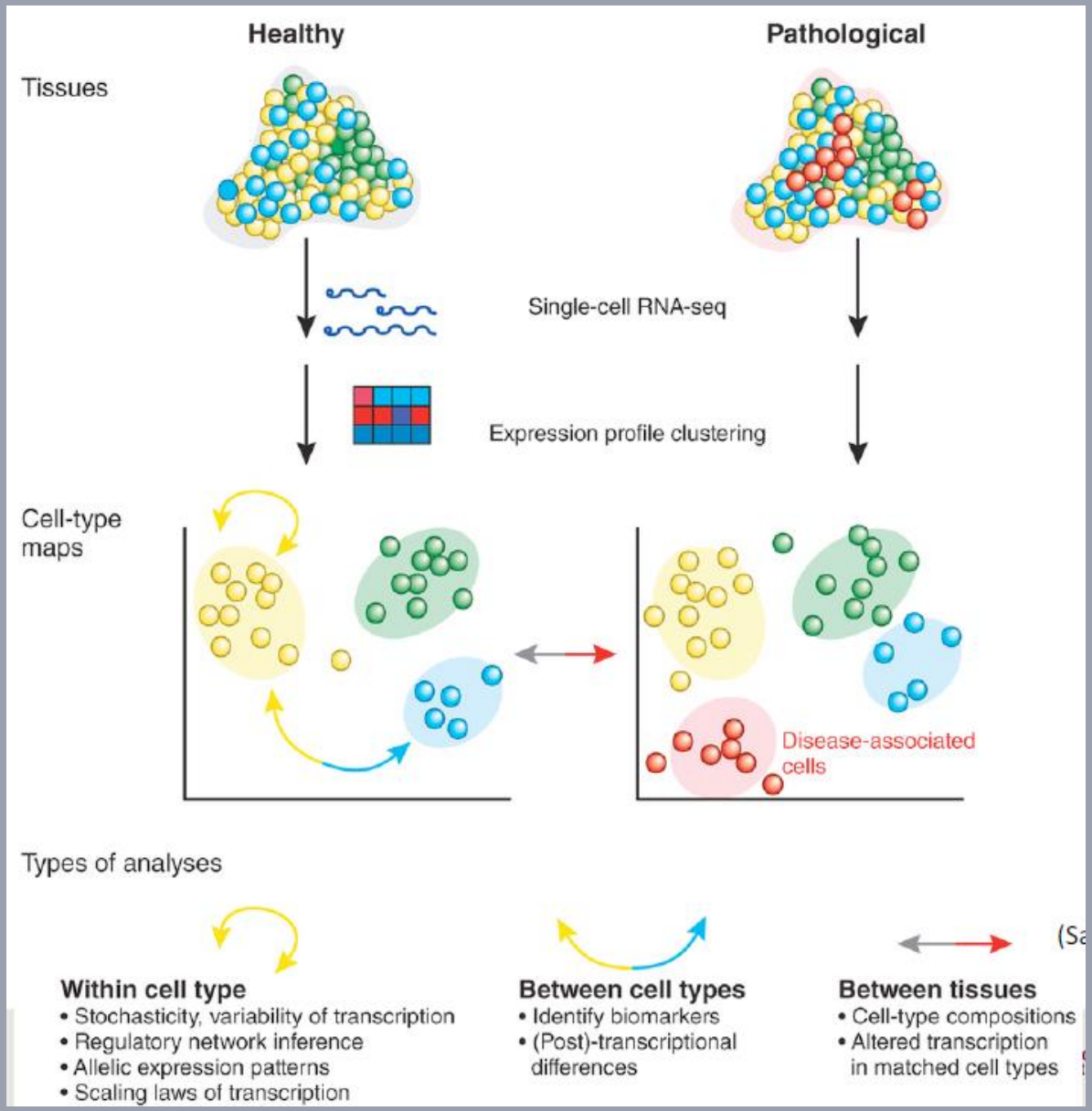



## 4. Infer gene regulatory networks




Liu et al., F1000, 2016

# Why single cell profiling?




- Understanding heterogeneous tissues,
- Identification of **cell types** and **cell states**
- Identification and analysis of **rare cell types**
- **Changes in cellular composition**
- Dissection of temporal changes

Sandberg et al.,
Nature Methods, 2014

# Single cell analysis: the context

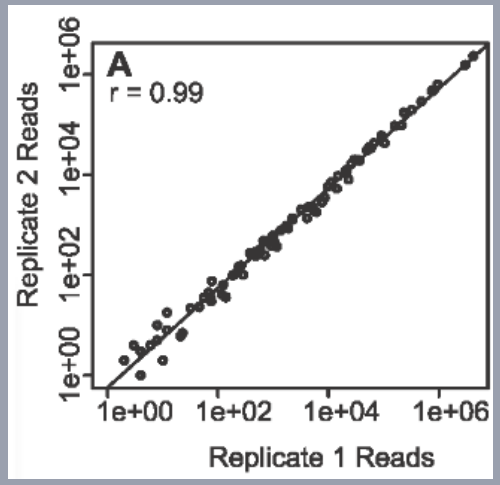
| | |
|---|---|
| mRNA libraries | 100 ng total RNA (10.000 cells) |
| Single cell mRNA libraries | 10 pg total RNA (<<1pg mRNA) |

- Highly efficient library preparation techniques
- Elimination of PCR amplification bias

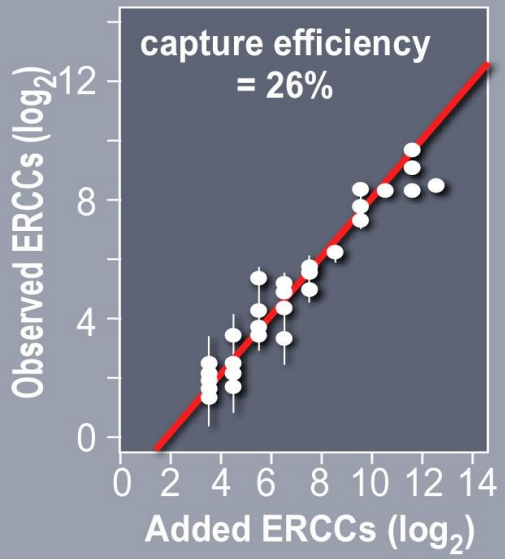
- Spike-in ERCC molecules to evaluate yield and capture efficiency
- Use of Unique Molecular Identifiers to monitor the number of molecules:
  - ***Random multimers to reduce cloning biases during small RNAs profiling*** *(Jayaprakash, NAR, 2011)*
  - ***Detection of rare DNA mutations*** *(Ståhlberg et al, NAR, 2016)*
  - ***Improved accuracy of molecule counting***

# ERCC spike-ins

ERCC (Externals RNA Controls Consortium) spikes is 92 polyadenylated RNA molecules
- different sequences and lengths,
- no homology with known genome sequences,
- 2 mixes (mix1 and mix2) with the 92 sequences but in different amounts
- relative  amounts of spikes in mix1 and mix2 from 1 to 10e6

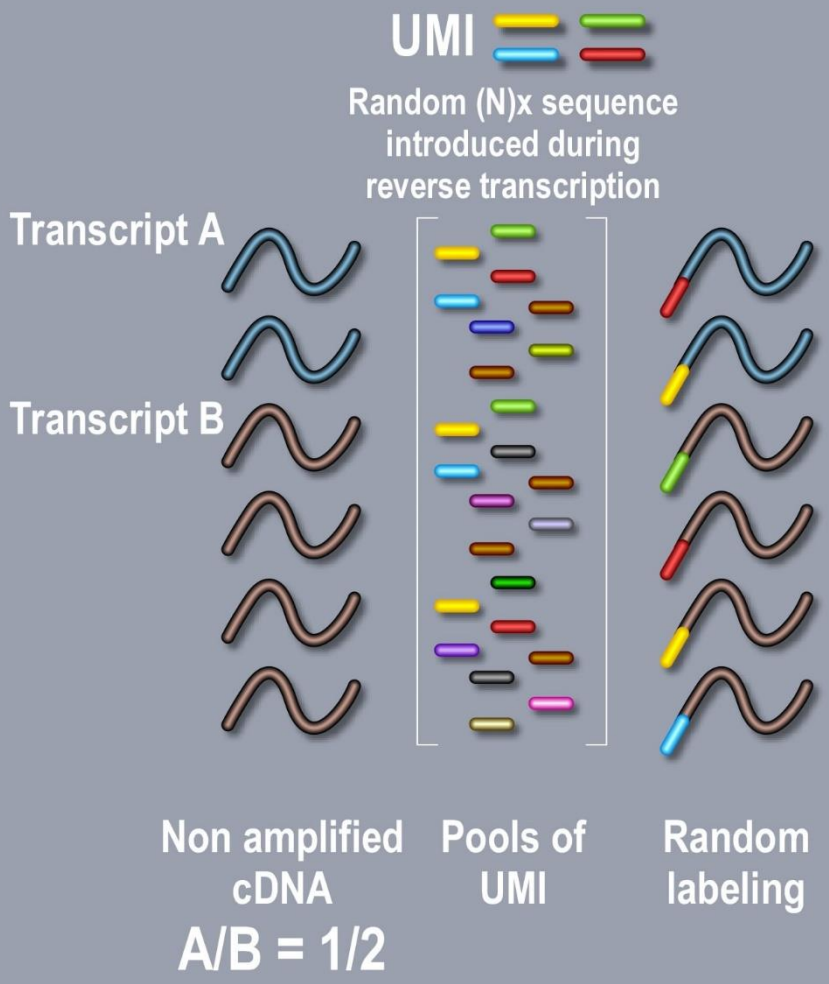



Scatter plot of read counts for each ERCC transcript in 2 RNA-seq libraries
Jiang et al., Gen.Res. (2015)



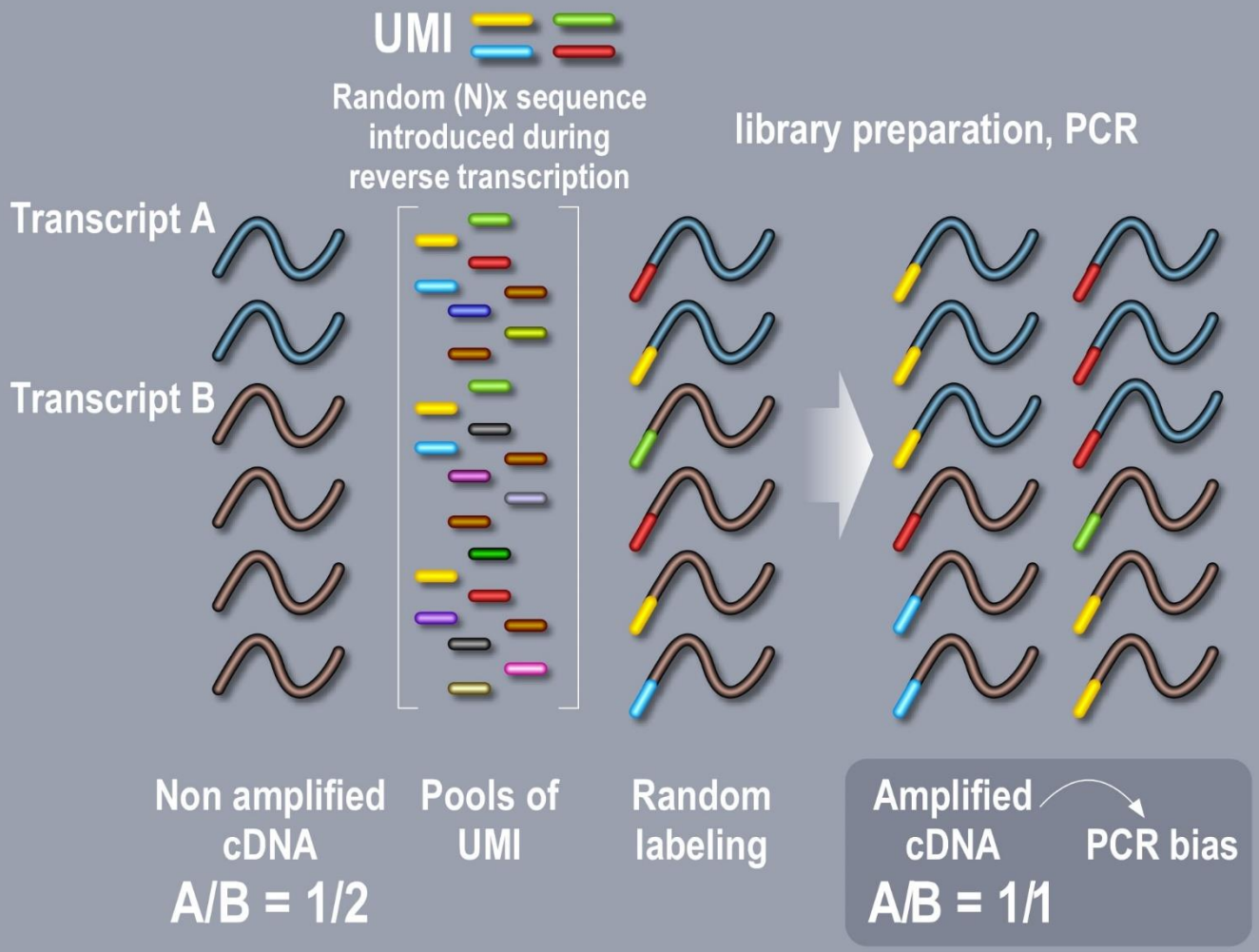

Single Cell capture efficiency
Arguel et al., NAR (2016)

# Unique Molecular Identifier (*Islam et al., Nature Methods, 2014*)



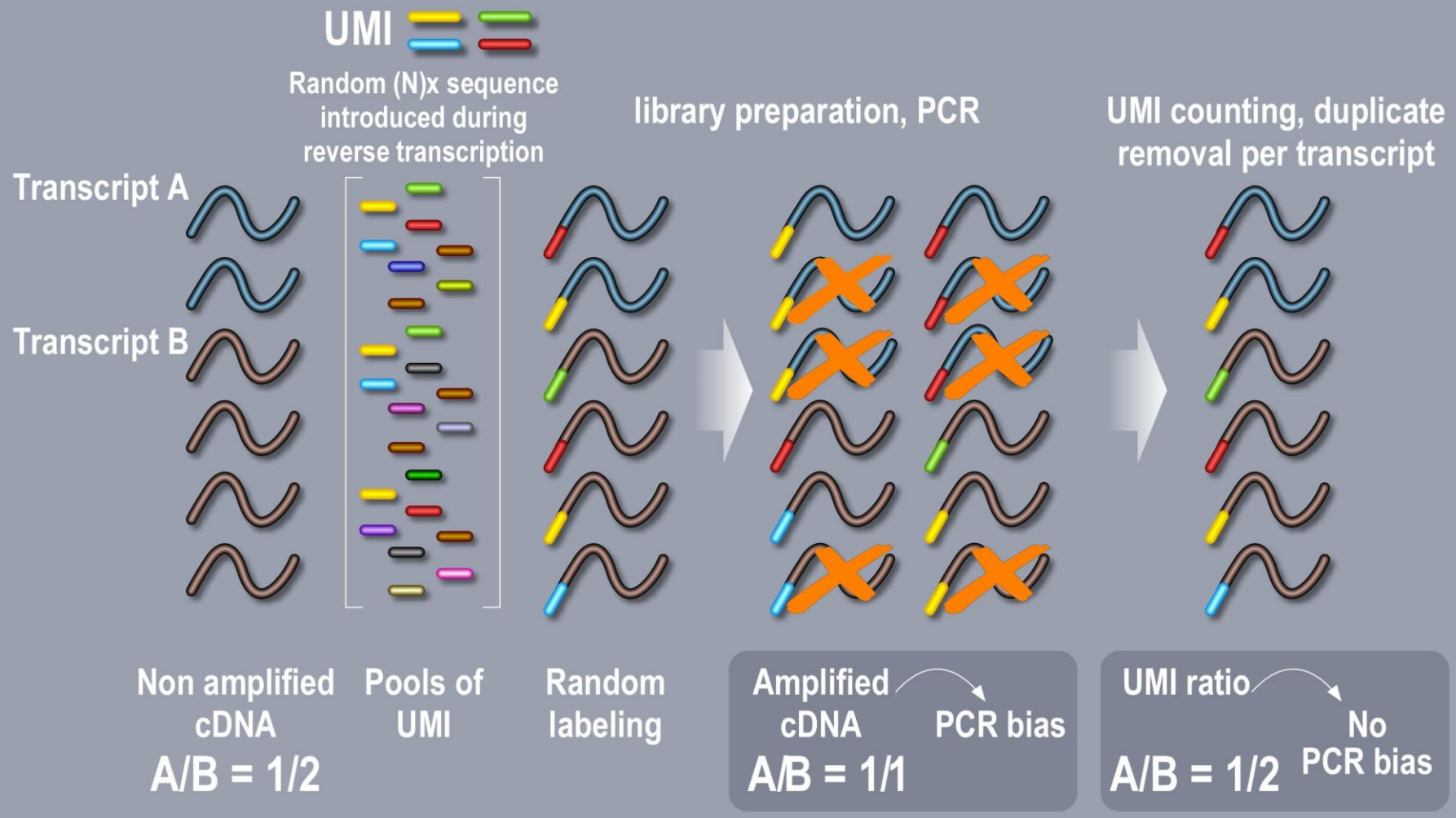

***UMIs :*** *Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Meth 9, 72-74 (2012)*
***UMIs for single cell transcriptome:*** *Islam, S. et al. Quantitative single-cell RNA-seq with UMI . Nat Methods 11, (2014).*

# Unique Molecular Identifier (*Islam et al., Nature Methods, 2014*)



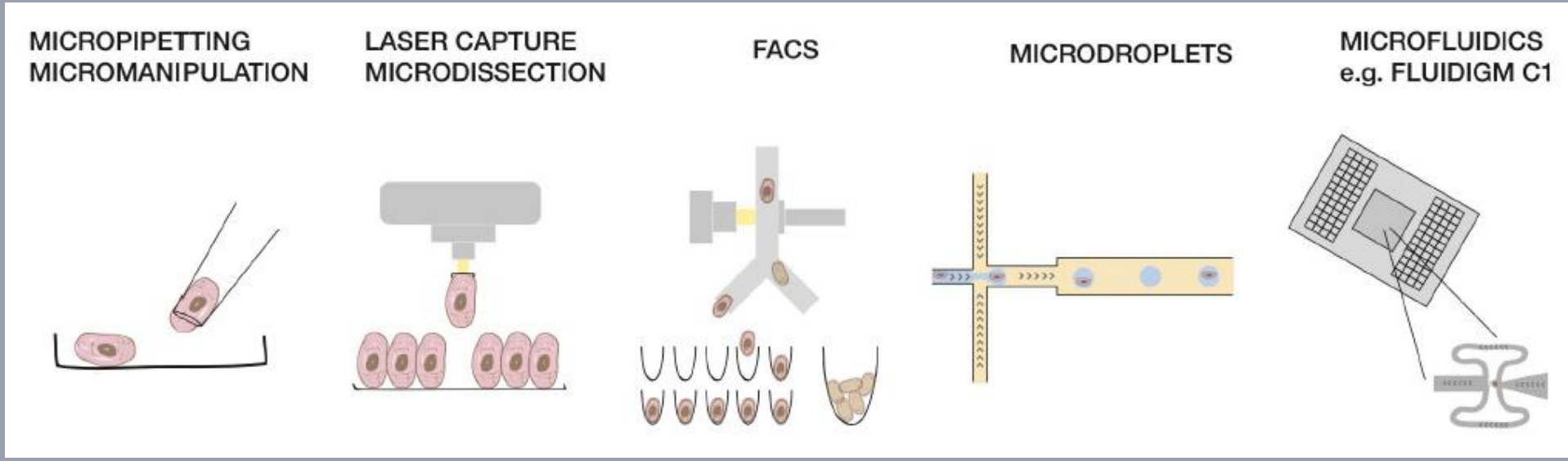

***UMIs :*** *Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Meth 9, 72-74 (2012)*
***UMIs for single cell transcriptome:*** *Islam, S. et al. Quantitative single-cell RNA-seq with UMI . Nat Methods 11, (2014).*

# Unique Molecular Identifier (*Islam et al., Nature Methods, 2014*)

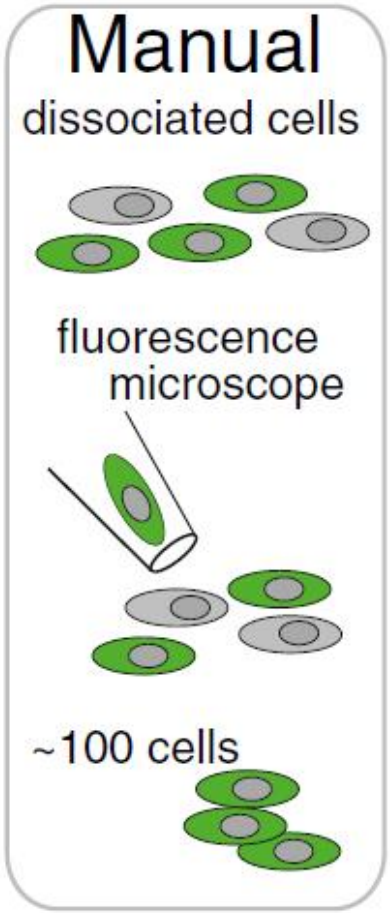


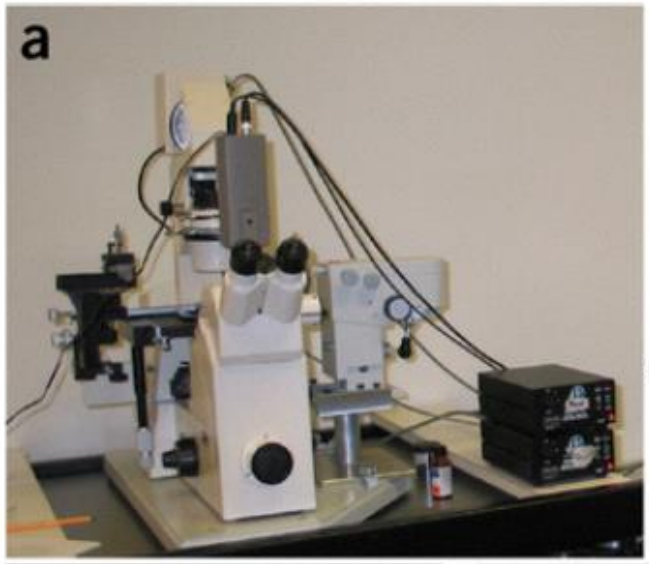

***UMIs :*** *Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Meth 9, 72-74 (2012)*
***UMIs for single cell transcriptome:*** *Islam, S. et al. Quantitative single-cell RNA-seq with UMI . Nat Methods 11, (2014).*

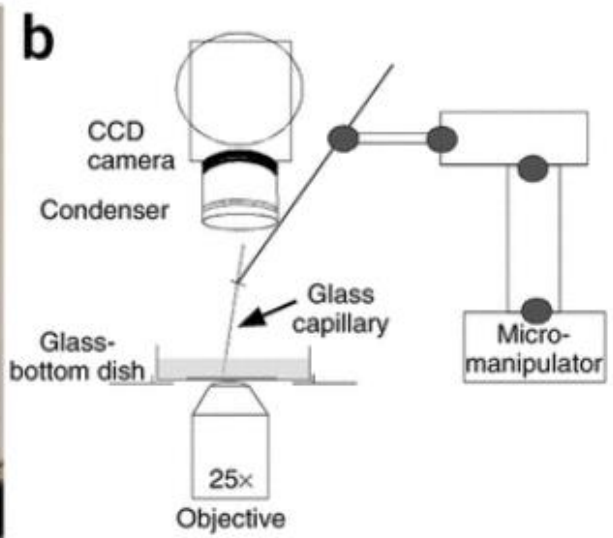
# To measure sequences in individual cells, need methods that capture one cell at a time

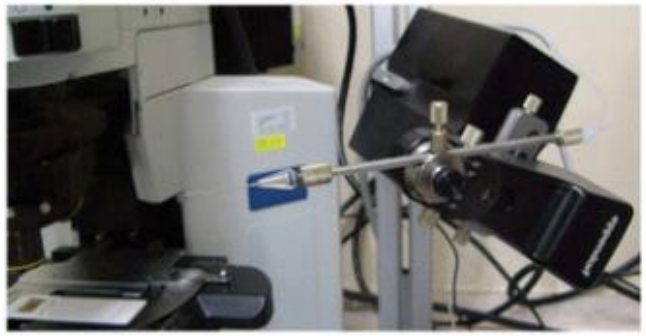


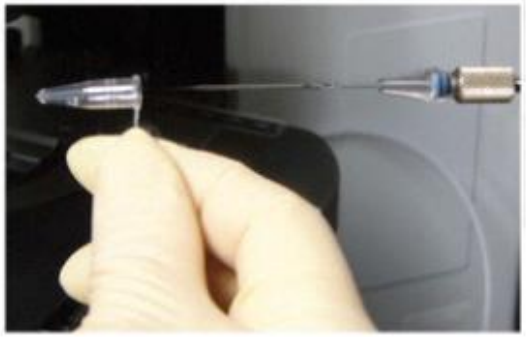
# Manual Cell sorting



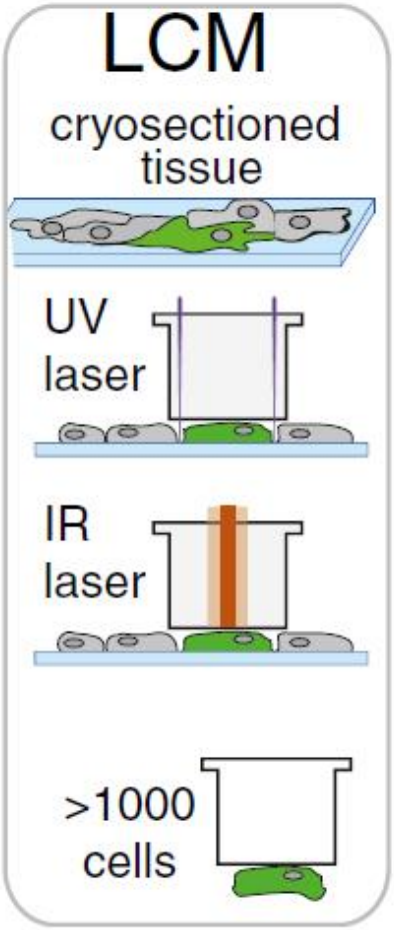



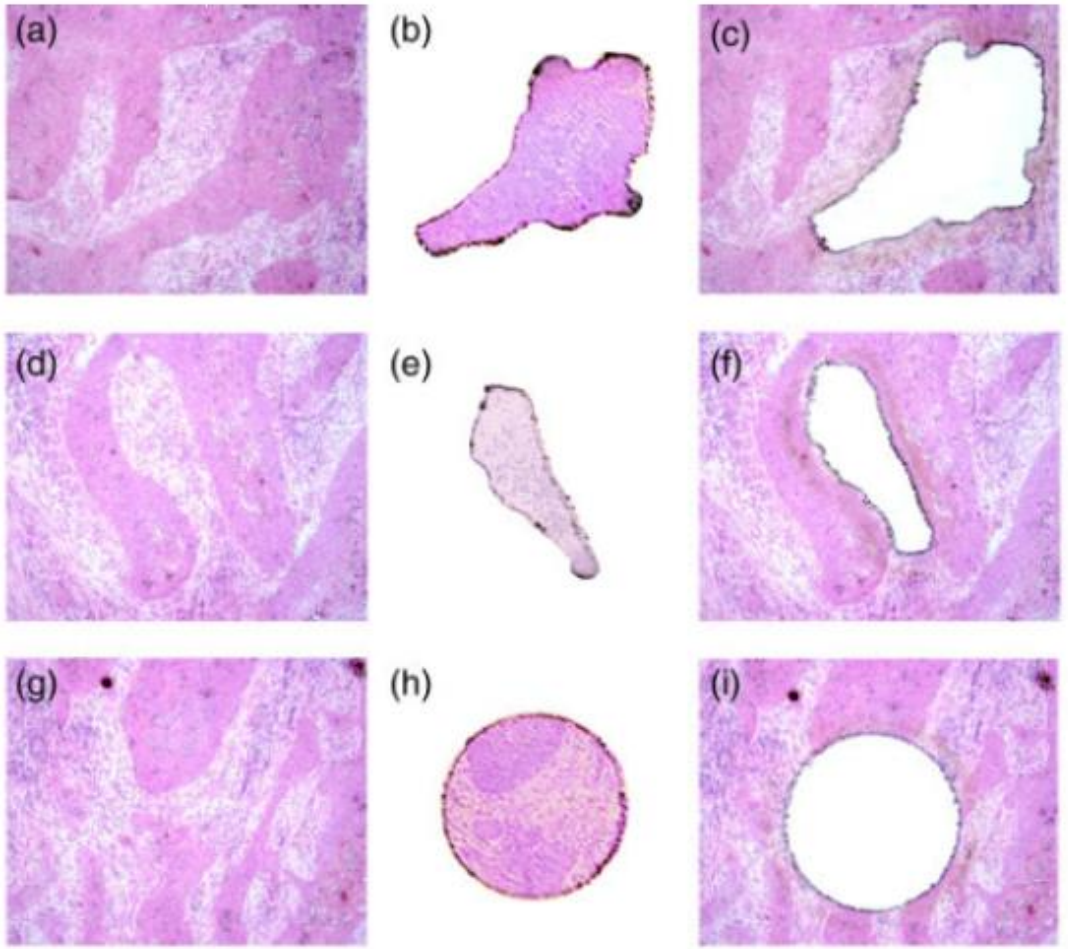



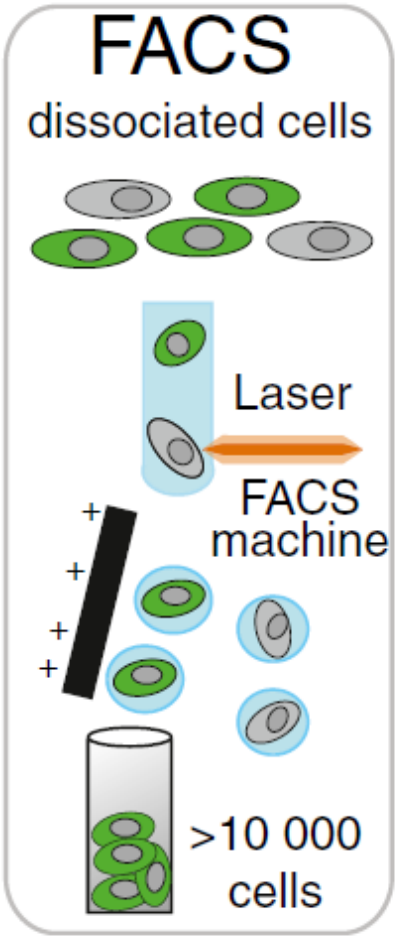

http://www.nature.com/nprot/journal/v6/n5/images_article/nprot.2011.322-F2.jpg

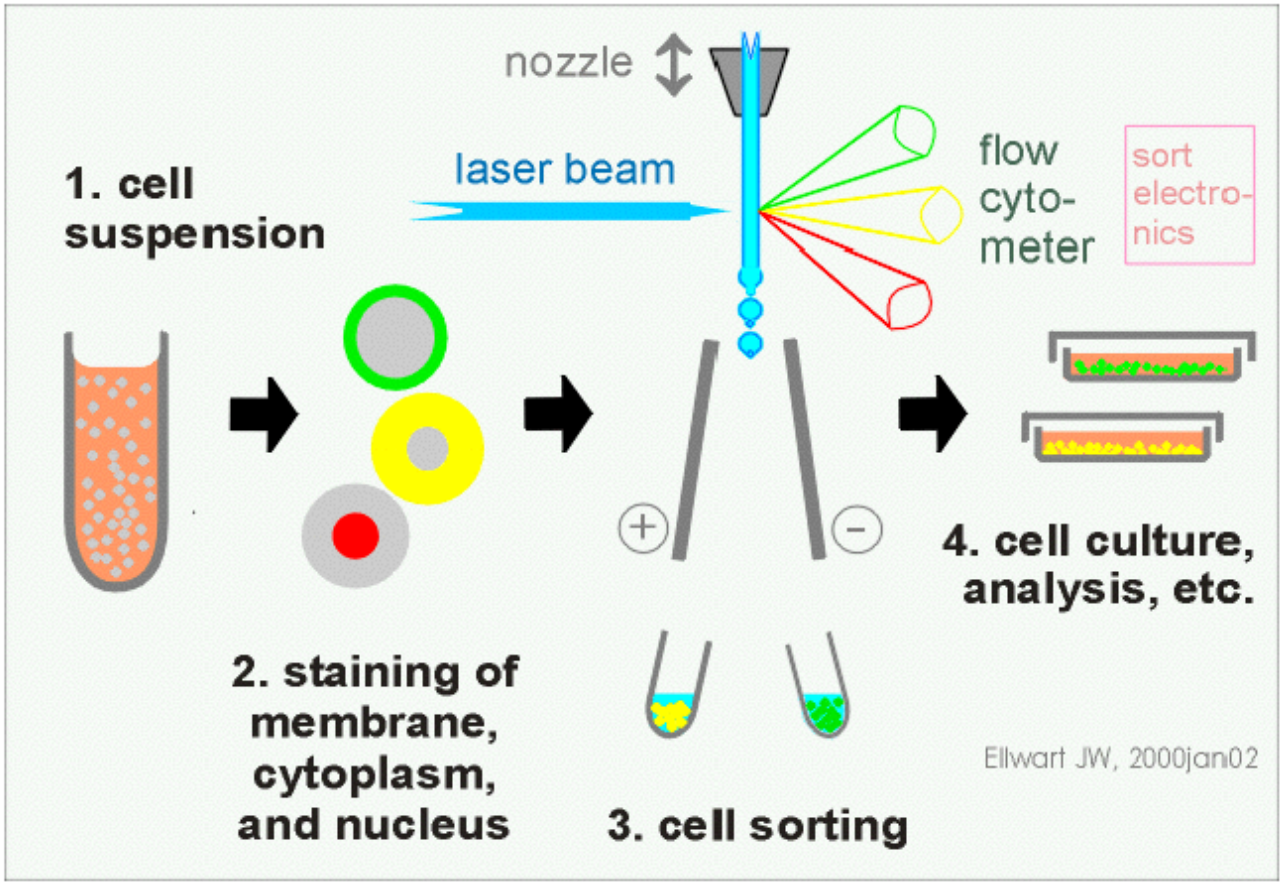
# Laser Capture Microdissection



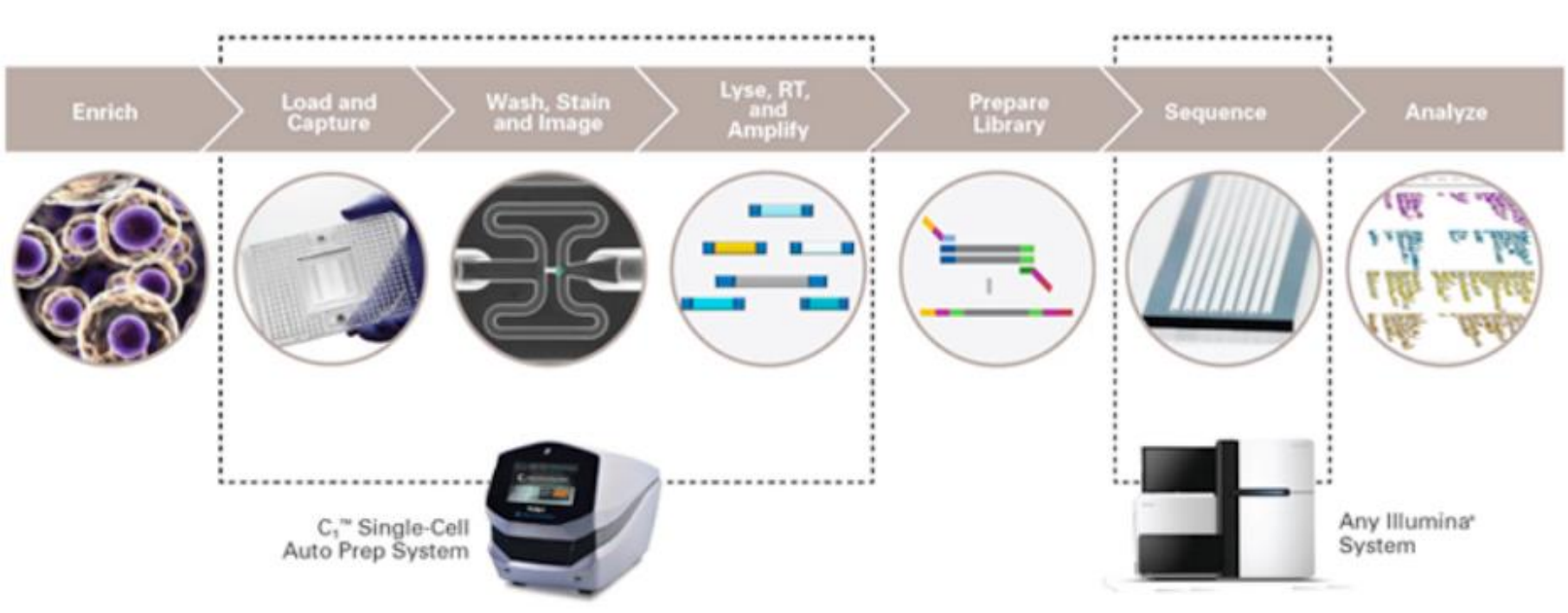

http://www.genomemedicine.com/content/figures/gm247-2-l.jpg




http://www.flowlab-childrens-harvard.com/yahoo_site_admin/assets/images/principle123.285181420_std.gif

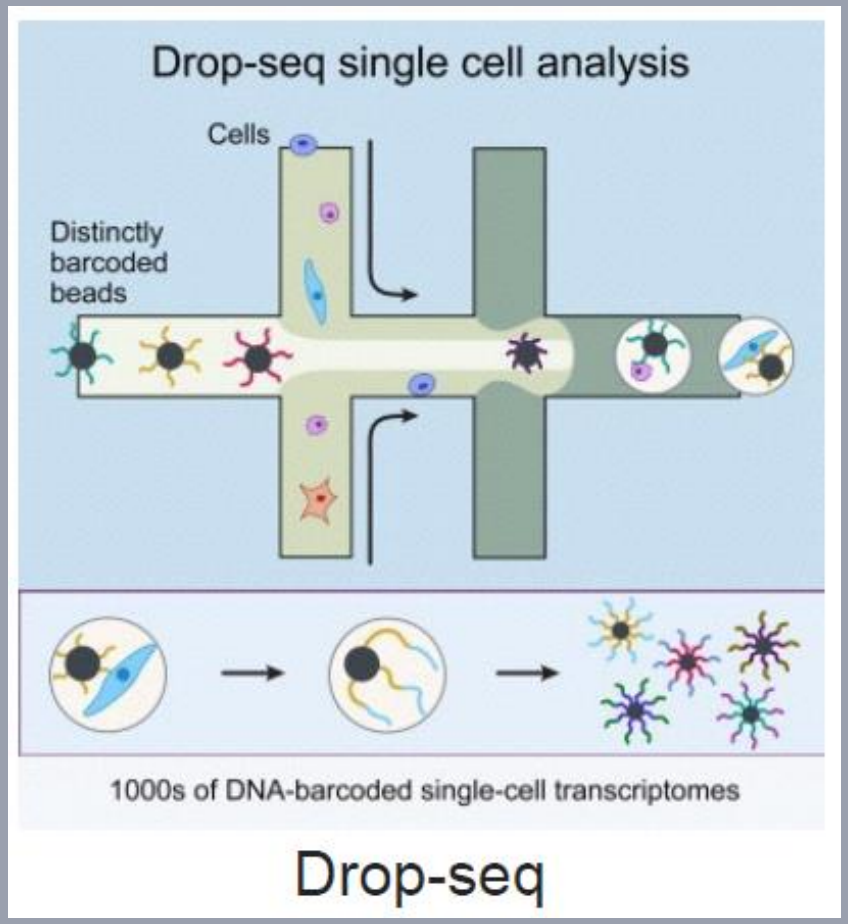
# Microfluidics: Fluidigm C1

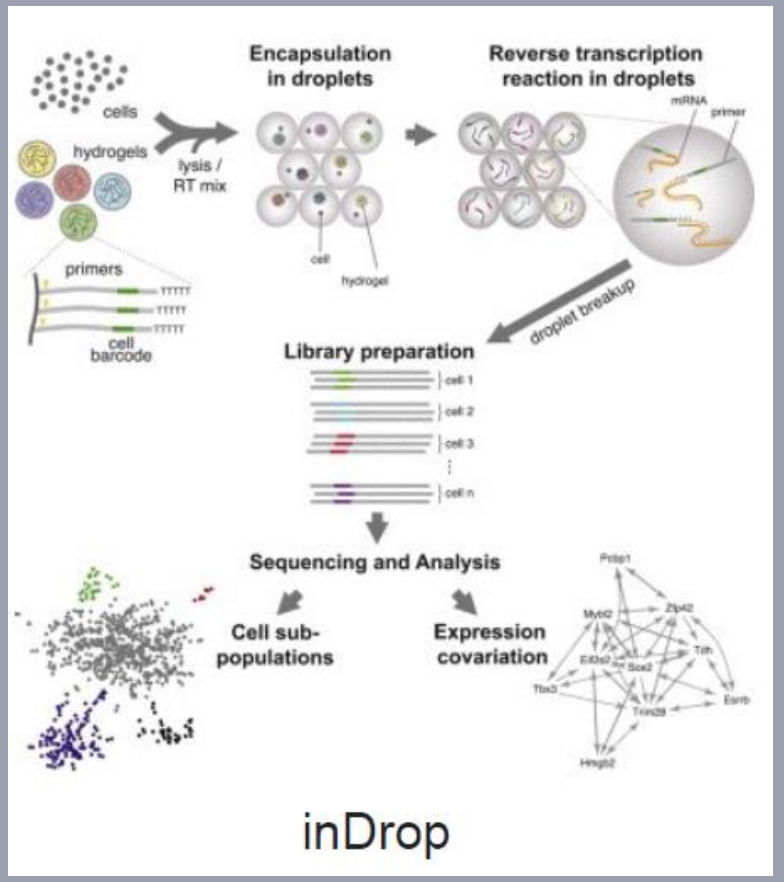



- Limiting factor is size of capture chambers (96 chips: 5-10µm, 10-17µm, >17µm)
- 800 cells chip (10-17 µm diameter cells)
- ScriptHub: protocols for running SMARTer, SmartSeq2, CEL-seq, STRT, …

openApp Chip for custom protocols development

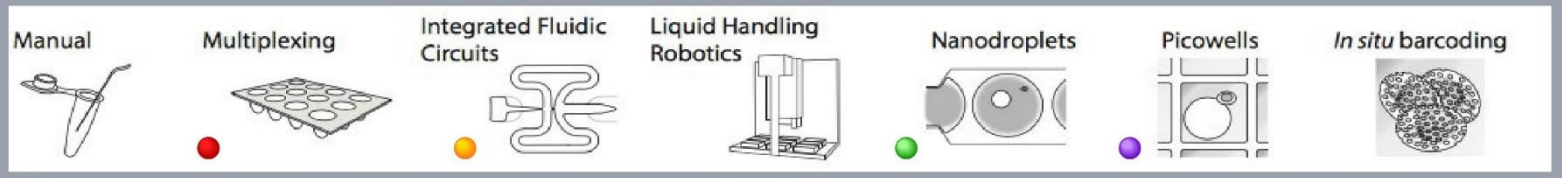
# Microfluidics: Droplets-based approaches, Macosko et al., Klein et al., Cell, 2015

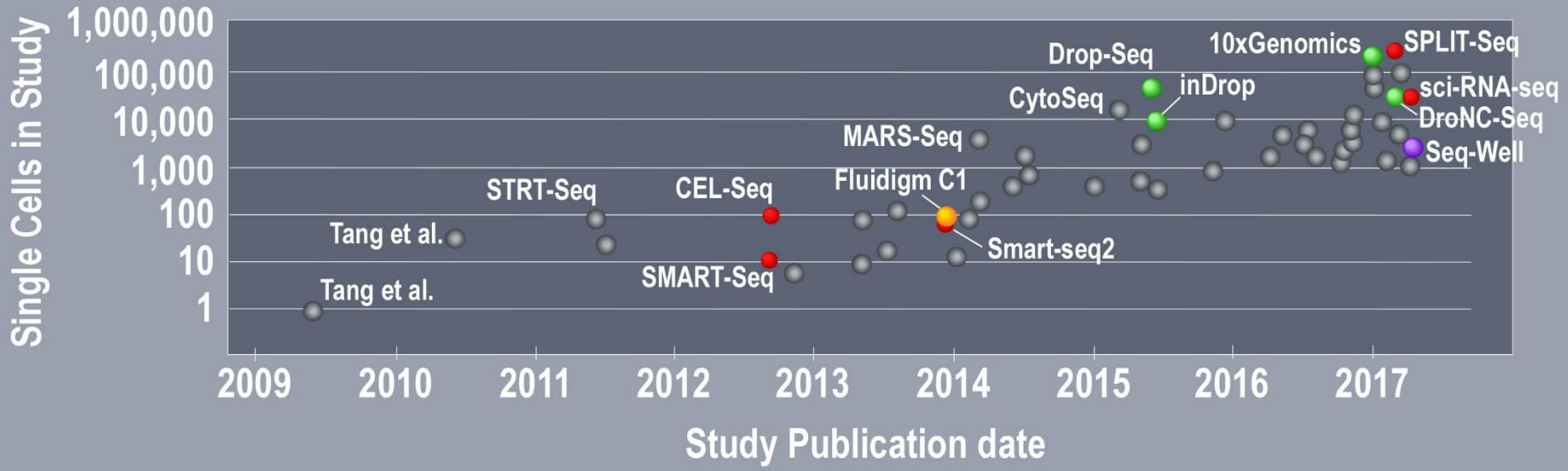



- water-in-oil droplets to capture cells with pre-barcoded (BC+UMIs) beads,
- really really high throughput (40,000 cells),
- No cell size issue

# Scaling Single Cell Transcriptomics (*Svensson et al., Nature Methods, 2017*)

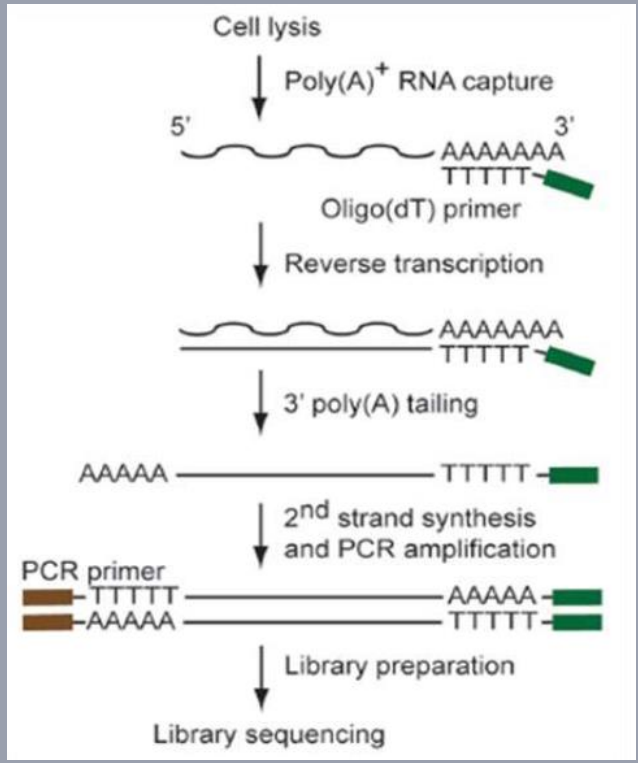


# Tang et al., 2009

## Protocol

- Total RNA is isolated and fragmented,
- Converted to cDNA by using an **oligodT** primer with a specific anchor sequence,
- **Second strand synthesis using a polyT primer** with another anchor sequence,
- **PCR amplified** from primers against the two anchor sequences.



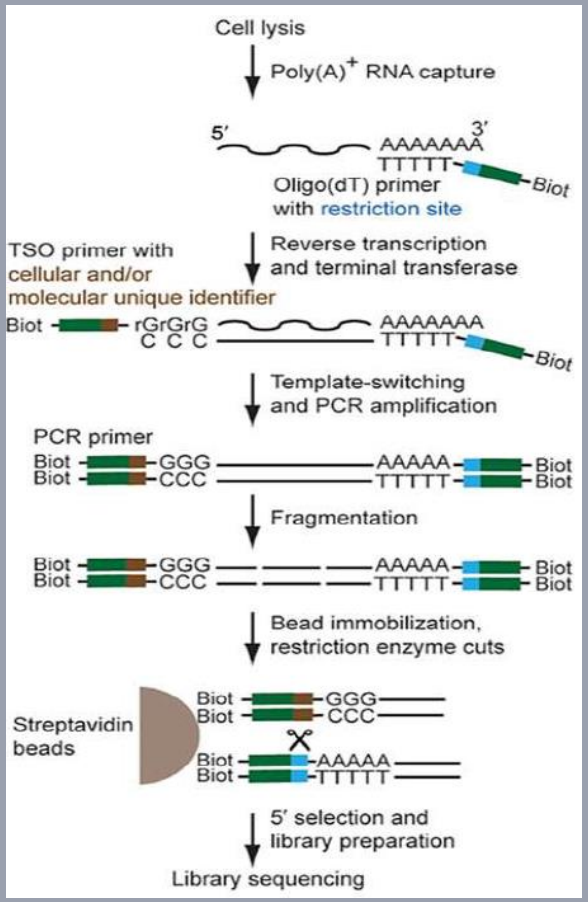

## Drawback

- Premature termination of RT reduces transcript coverage at the 5' end
- Introduction of a polyA tail in addition to its own polyA sequence at the **3' end** of the input RNA causes a **loss of strand information** in the resulting double-stranded cDNA

# STRT: Single cell tagged reverse transcription, Islam, Nat.Prot. (2012) @ Linnarsson's lab

Protocol

- based on **template switching**,
- **5' end cDNA tagged N5 UMI**,
- biotin is introduced at both the 3' and 5' ends via the use of biotinylated primers.
- enzymatic cleavage leads to the selection of only the 5' fragments for library construction.
- sequencing and analysis shows **5' read bias**



# SMART-Seq, Ramskold (2012), SMART-seq2, Picelli (2014) @ Sandberg's lab

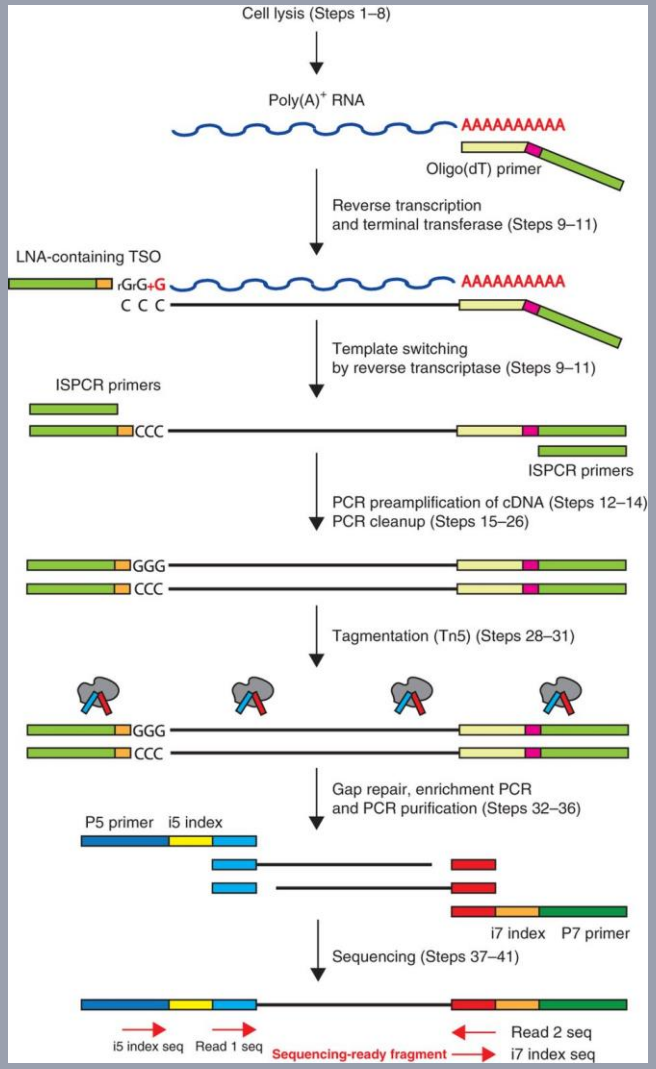
SMART= Switching Mechanism at the end of the 5´-end of the RNA Transcript

## Protocol

- Based on template switching mechanism,
- Anchor a 5' universal seq. along with Locked nucleic acid by reverse transcription,
- cDNA is then **PCR amplified ,**
- Tagmentation is used to construct libraries,
- Generate **full transcript coverage**

## Drawback

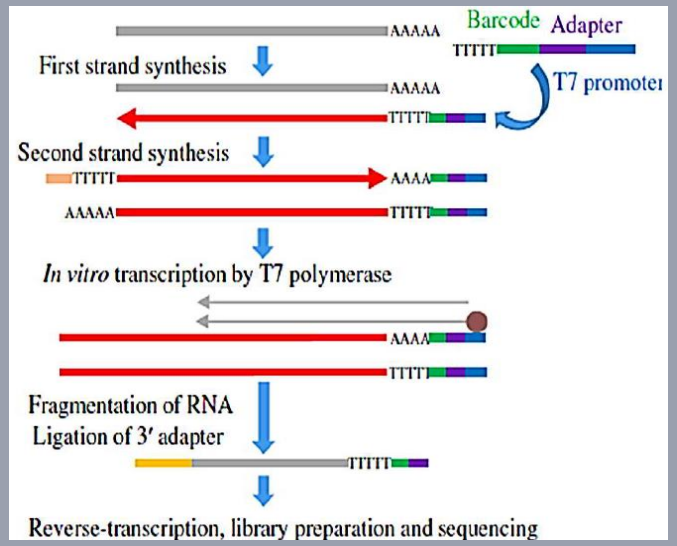
- No UMIs



# Cell expression by linear amplification and sequencing (CEL-seq, CEL-seq2)

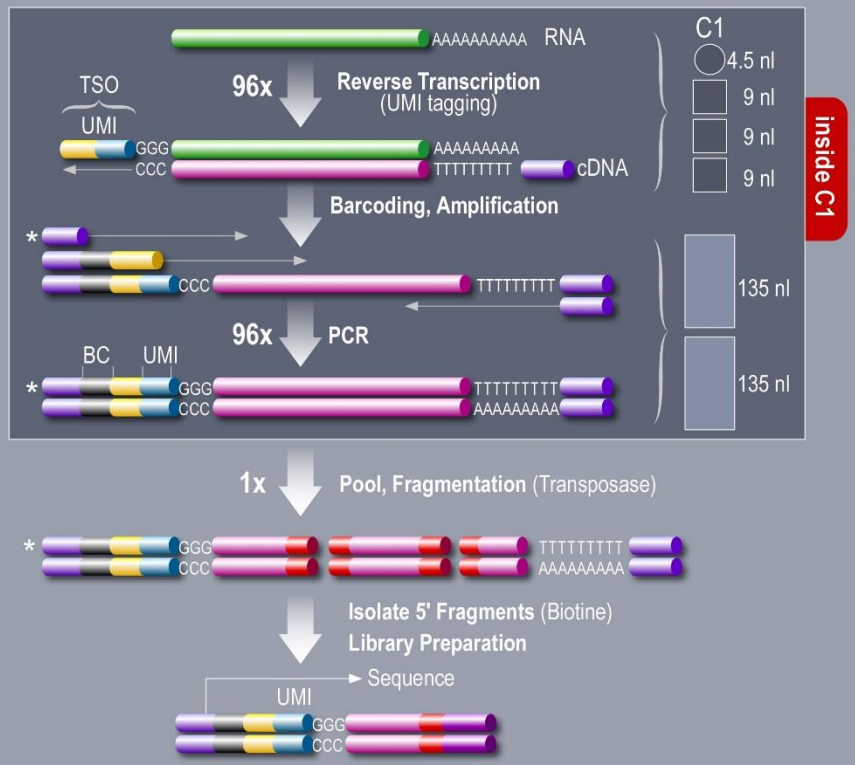
Protocol

- OligodT primer containing the 5' Illumina adaptor, a cell barcode, and a T7promoter (**CEL-seq2 add UMI**),
- RT and second-strand synthesis,
- cDNA from all the cells is pooled and amplified **by in vitro transcription** from the T7 promoter,
- RNA fragmentation, Illumina adaptor ligated at 3' end
- RNA is reverse transcribed, library is prepared then sequencing,
- Sequencing of the **3´terminal fragments**



# A cost effective 5' selective single cell transcriptome profiling approach

- UMIs inside Fluidigm C1, PCR bias removal (molecules couting)
- 60k transcripts in HEK (homogeneity of chemistry)
- Illumina and Ion Torrent sequencing (no paired-end required)
- Capture efficiency 26% (ERCC spike-ins)
- Sequences 5' end of transcripts → TSS identification




*Arguel et al., Nucleic Acid Research, 2016*

**Now available on ScriptHub (Fluidigm)**

Single-cell mRNA Seq with Integrated Barcoding mRNA Sequencing

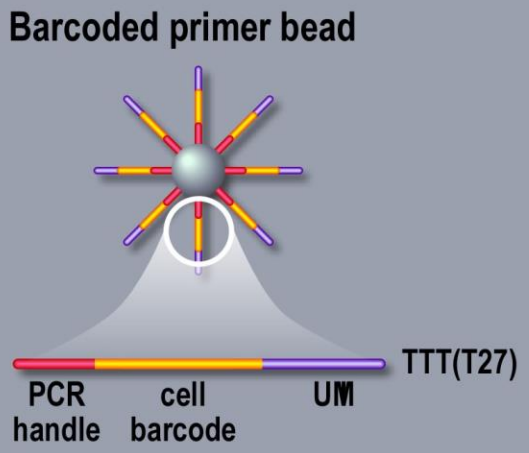
**Arguel et al.**
Pascal Barbry - UCA Genomix - IPMC, CNRS - University of Cote d'Azur

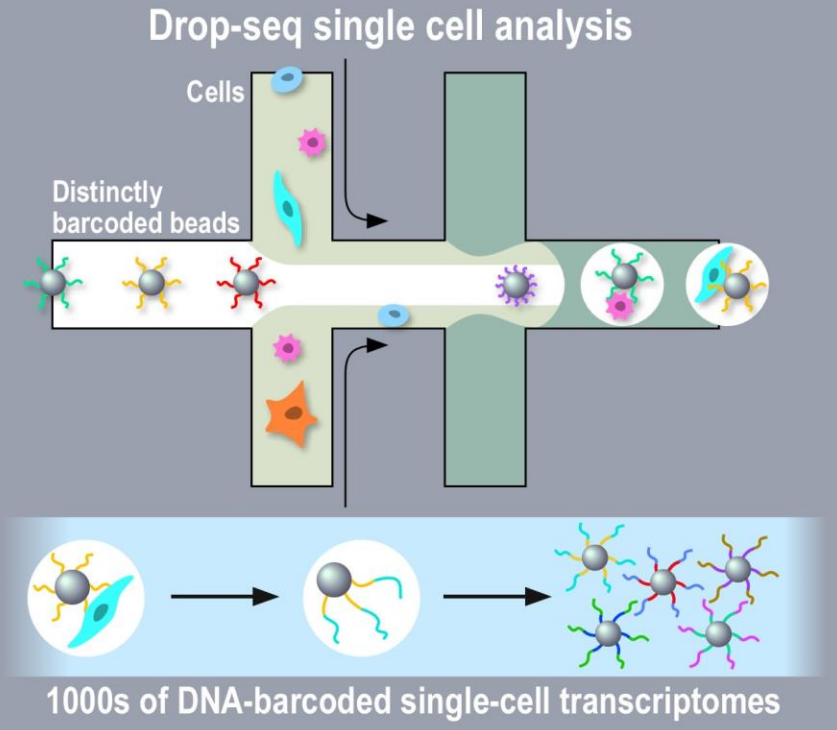
# Drop-seq / inDROP: droplets-based scRNA-seq (*Mascosko et al., Klein et al., CELL, may 2015*)

Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets

Evan Z. Macosko[1, 2, 3], Anindita Basu[4, 5], Rahul Satija[4, 6, 7], James Nemesh[1, 2, 3], Karthik Shekhar[4], Melissa Goldman[1, 2], Itay Tirosh[4], Allison R. Bialas[8], Nolan Kamitaki[1, 2, 3], Emily M. Martersteck[9], John J. Trombetta[4], David A. Weitz[5, 10], Joshua R. Sanes[9], Alex K. Shalek[4, 11, 12], Aviv Regev[4, 13, 14], Steven A. McCarroll[1, 2, 3]

- droplets encapsulation of cells and barcoded beads
- 3' selective single cell RNA-seq
- 12bp cell barcode and 8bp UMI
- capture efficiency 12.5%
- 44,808 mouse retinal cells
- identification of 39 differents cell types

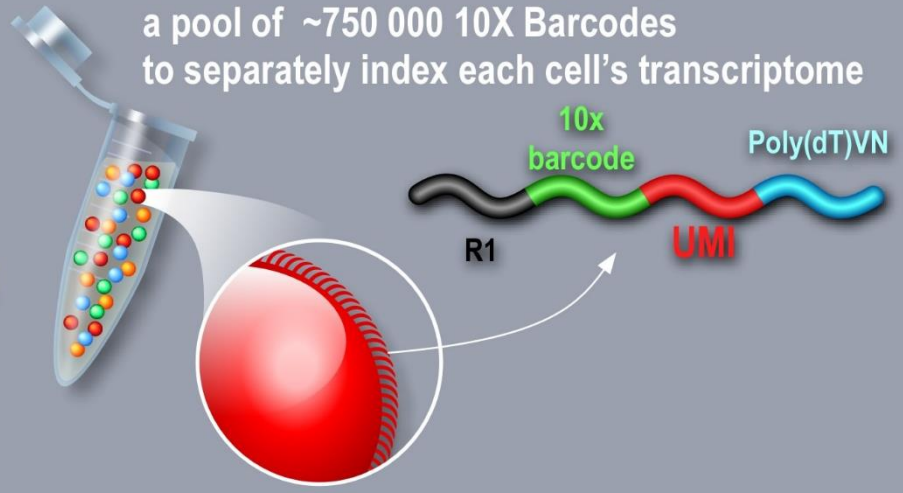


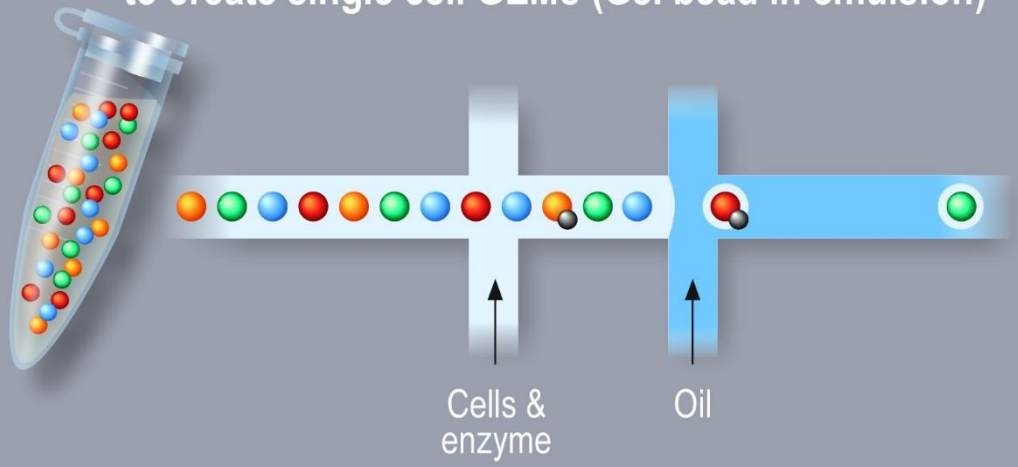
# 10X Genomics Chromium

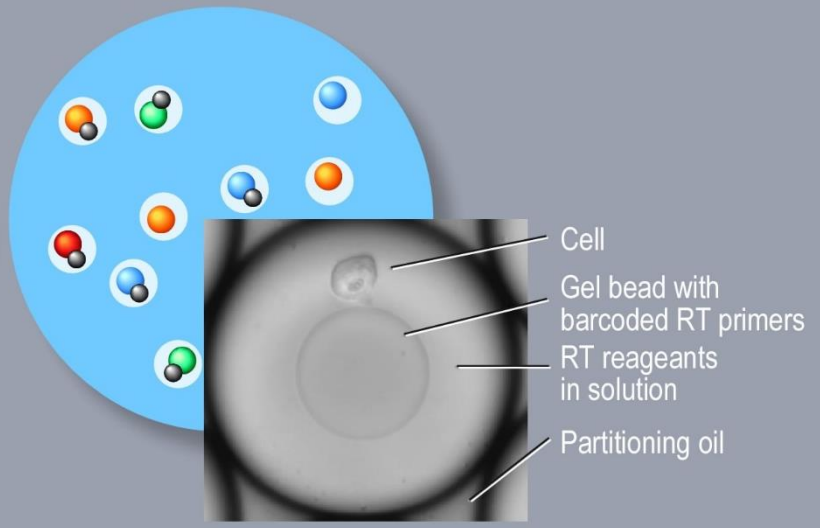

Single-use microfluidics chip

10X GemCode Technology samples
a pool of ~750 000 10X Barcodes
to separately index each cell's transcriptome




10X Barcoded Gel Beads are mixed with cells, enzyme and oil
to create single cell GEMs (Gel bead in emulsion)



# Experimental approaches comparison

| Lab (Ref) Name of approach | Region sequenced | UMI | | Capture efficiency (ERCC) | Amplification | Barcoding | Device | Sequencer |
|---|---|---|---|---|---|---|---|---|
| | | Design | Counting / error correction | | | | | |
| This approach | 5' | N4H4 | Unique (a) / edit distance (b) | 26% | PCR | PCR pre-fragmentation | C1 | Illumina |
| Linnarsson (10) STRT-seq | 5' | N5 | Unique start (c) / percentile (d) | 48% (e) | PCR | Tagmentation *custom barcoded transposons* | C1 | Illumina |
| Linnarsson (7) STRT-seq | | | Unique start (c) / percentile + (df | 22% (b) | | | | |
| Amit (13) Mars-seq | Close to 3' | N4 | edit distance (b) | 2 – 3% | Isothermal | RT | Plates | Illumina |
| McCarol (16) Drop-seq | Close to 3' | N8 | edit distance (b) | 12.8% | PCR | RT | Microdroplets | Illumina |
| Kirschner (17) InDrop | Close to 3' | N6 | (g) | 7.1% | Isothermal | RT | Microdroplets | Illumina |
| Yanai (15) Cel-seq | Close to 3' | No | none | 6% | Isothermal | RT | Plates | Illumina |
| Yanai (15) Cel-seq2 | Close to 3' | N6 | none | 22% in C1 | Isothermal | RT | Plates, C1 | Illumina |
| Smartseq2 | Internal (a) | No | – | n.d. | PCR | – | Tubes | Illumina |

Arguel et al., NAR, 2016