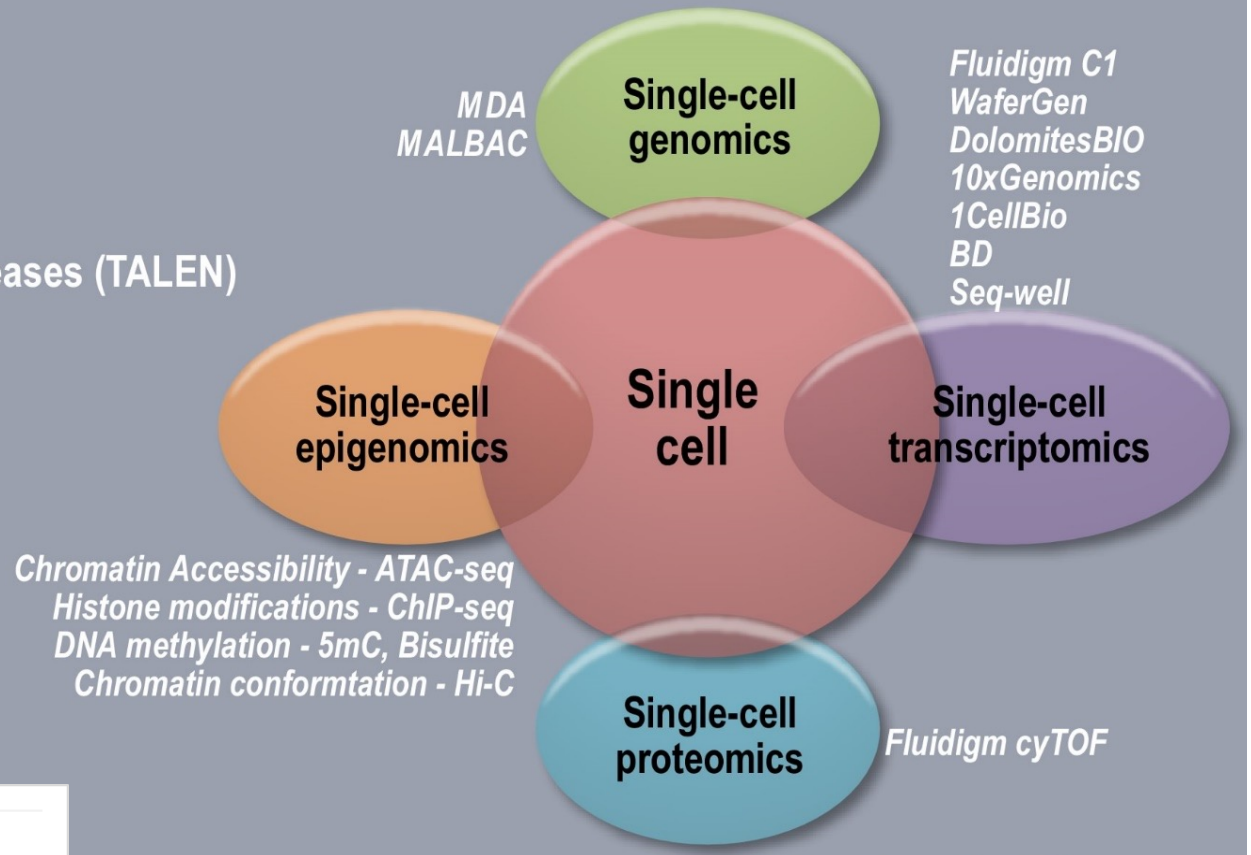
Singel Cell RNA-seq: Technologies and Experimental Approaches

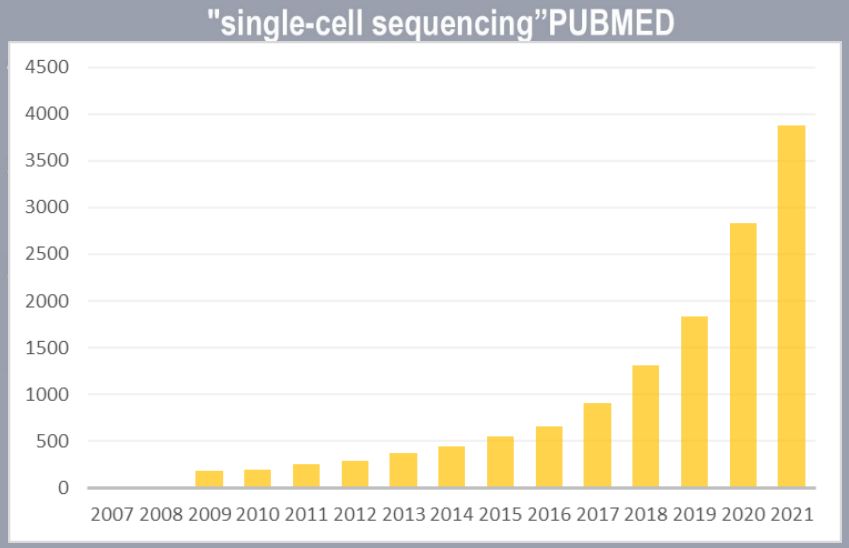
Kévin Lebrigand
UCAGenomix, Nice-Sophia-Antipolis
lebrigand@ipmc.cnrs.fr
@kevinlebrigand

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

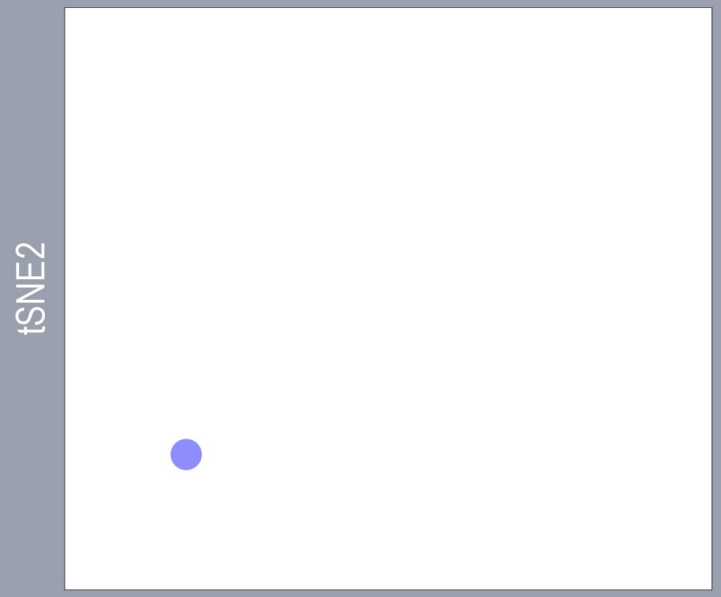


Macauley and Voet, 2014

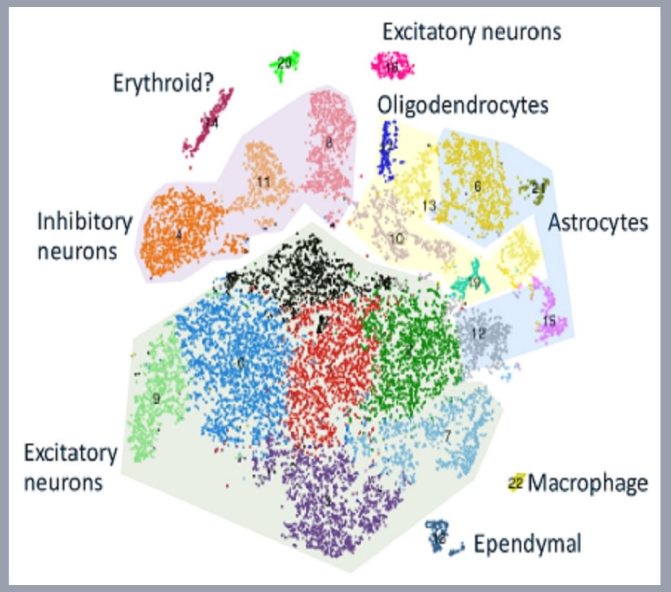


Single cell is a revolution for gene expression studies

Conventional gene expression



Single Cell gene expression

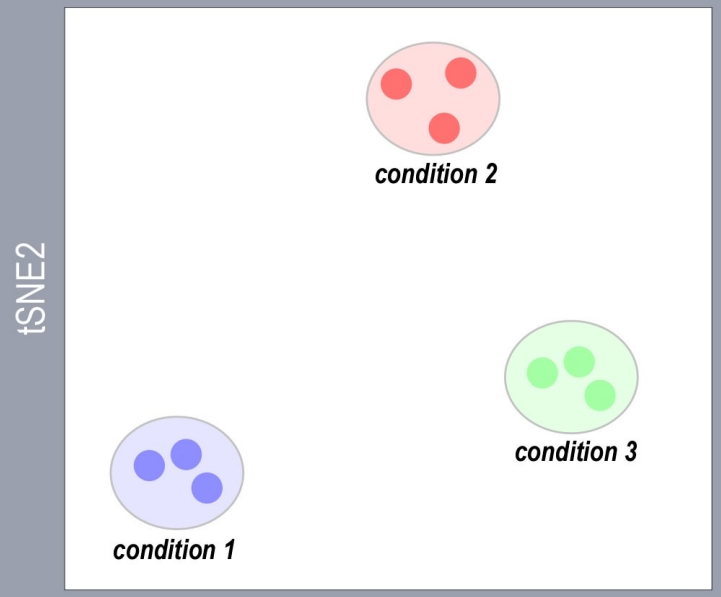


10xGenomics 1.3M single cell mouse brain dataset

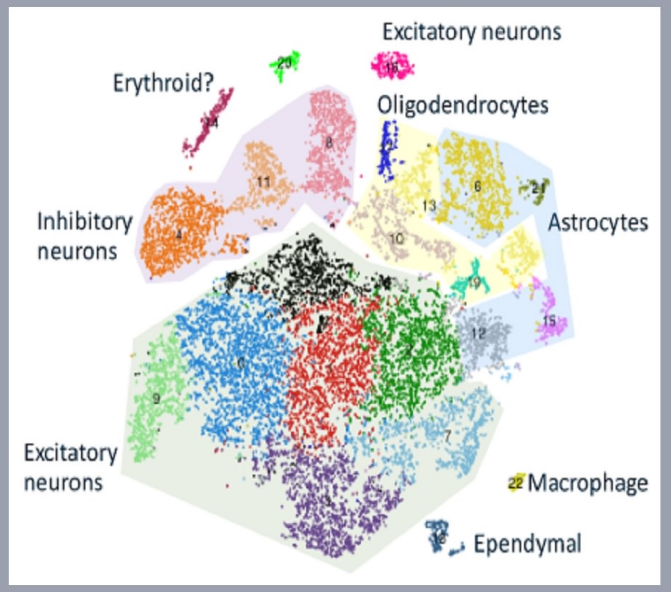
One "average" data point from a mixture of 10.000 single cells

Single cell is a revolution for gene expression studies

Conventional gene expression



Single Cell gene expression

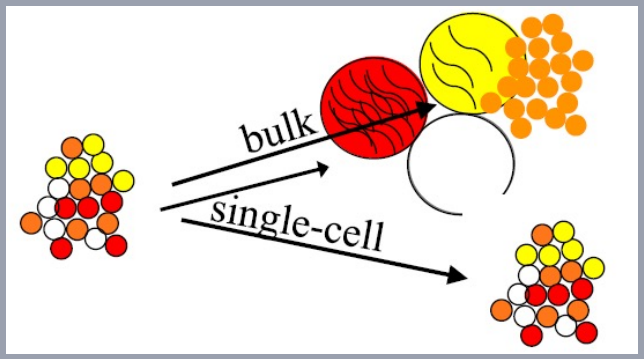


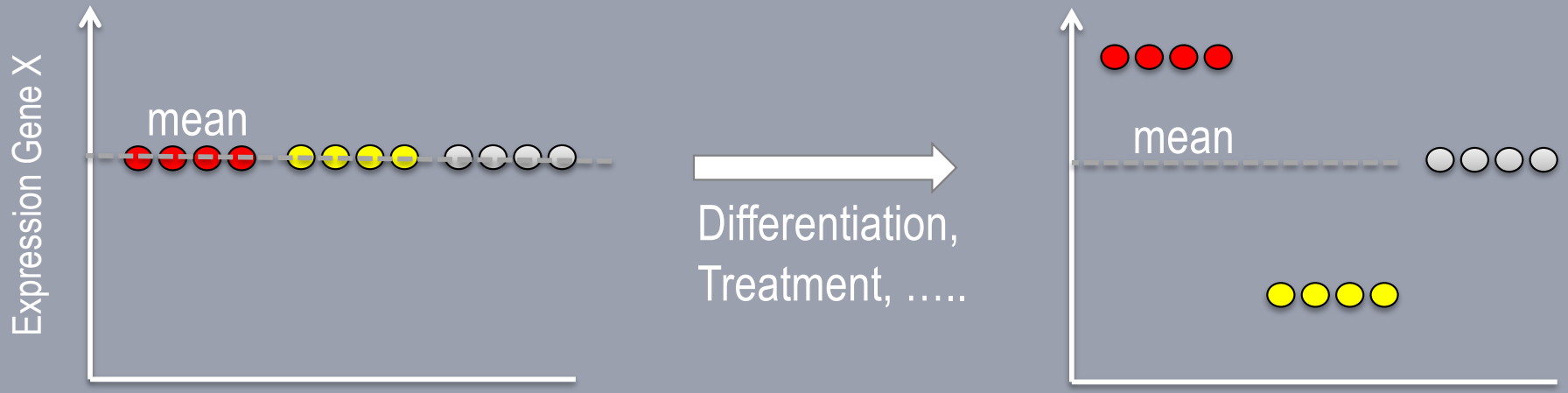
10xGenomics 1.3M single cell mouse brain dataset

One "average" data point from a mixture of 10.000 single cells

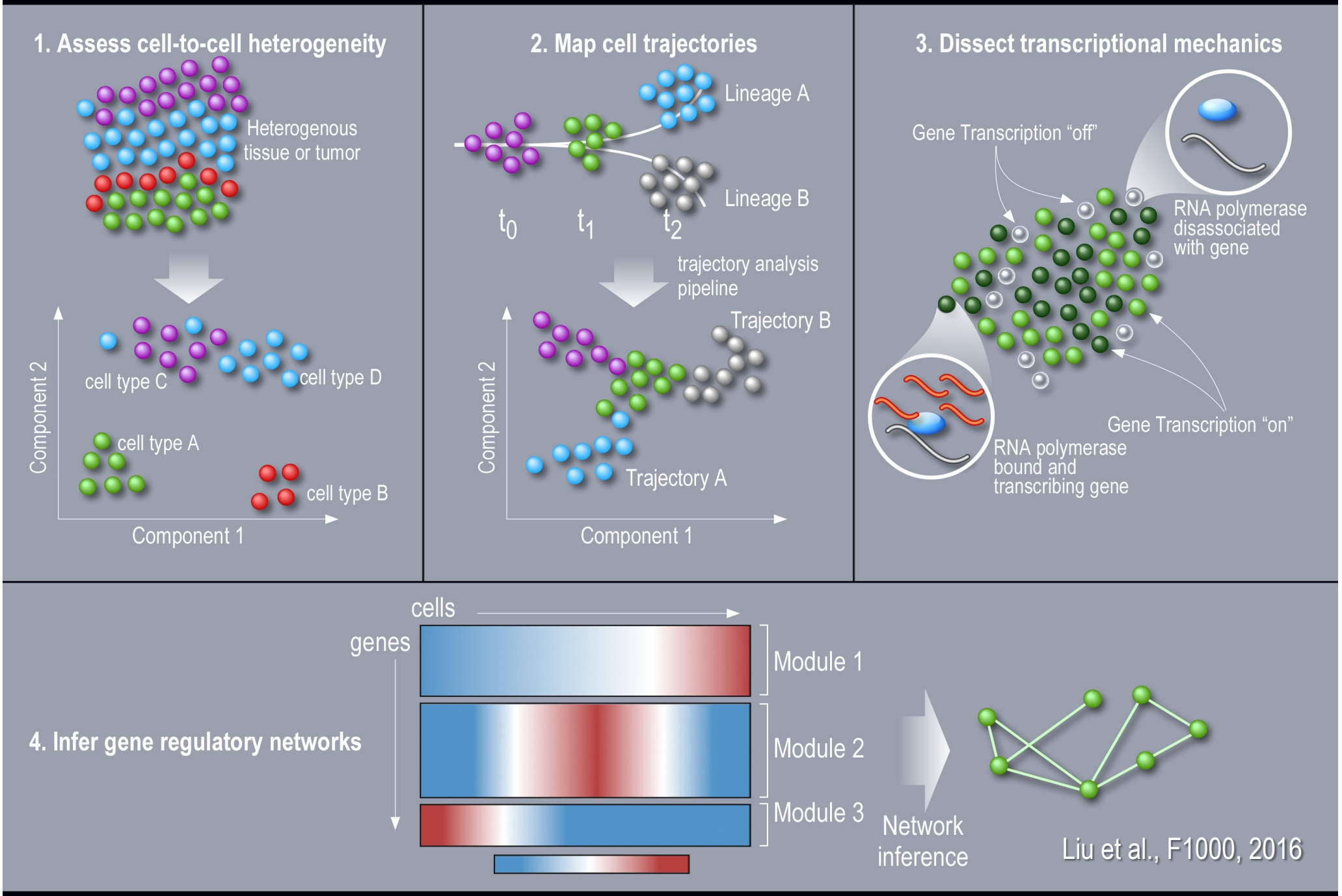
Why single cell profiling?

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





Why single cell profiling?



How much RNA does a typical mammalian cell contain?

https://www.qiagen.com/fr/resources/faq?id=06a192c2-e72d-42e8-9b40-3171e1eb4cb8&lang=en

The RNA content and RNA make up of a cell depend very much on its developmental stage and the type of cell. To estimate the approximate yield of RNA that can be expected from your starting material, we usually calculate that a typical mammalian cell contains **10-30 pg total RNA**.

The majority of RNA molecules are tRNAs and rRNAs. mRNA accounts for only **1–5%** of the total cellular RNA although the actual amount depends on the cell type and physiological state. Approximately **360,000 mRNA** molecules are present in a single mammalian cell, made up of approximately 12,000 different transcripts with a typical length of around 2 kb. Some mRNAs comprise 3% of the mRNA pool whereas others account for less than 0.1%. These rare or low-abundance mRNAs may have a copy number of only 5-15 molecules per cell.

Average total RNA yields

Primary cells (1×10^6 cells)	Total RNA (µg)
Dendritic cells, human	4
Hematopoietic progenitor cells ($CD34^+$), human	1
Fibroblasts, rat	5
PBMC	8

Cell lines (1×10^6 cells)	Total RNA (µg)
Colon carcinoma cells	30
HEK 293 cells	16
HeLa cells	32
HUV-EC-C	38
THP1 cells	16
U937 cells	12

Single cell analysis: the context

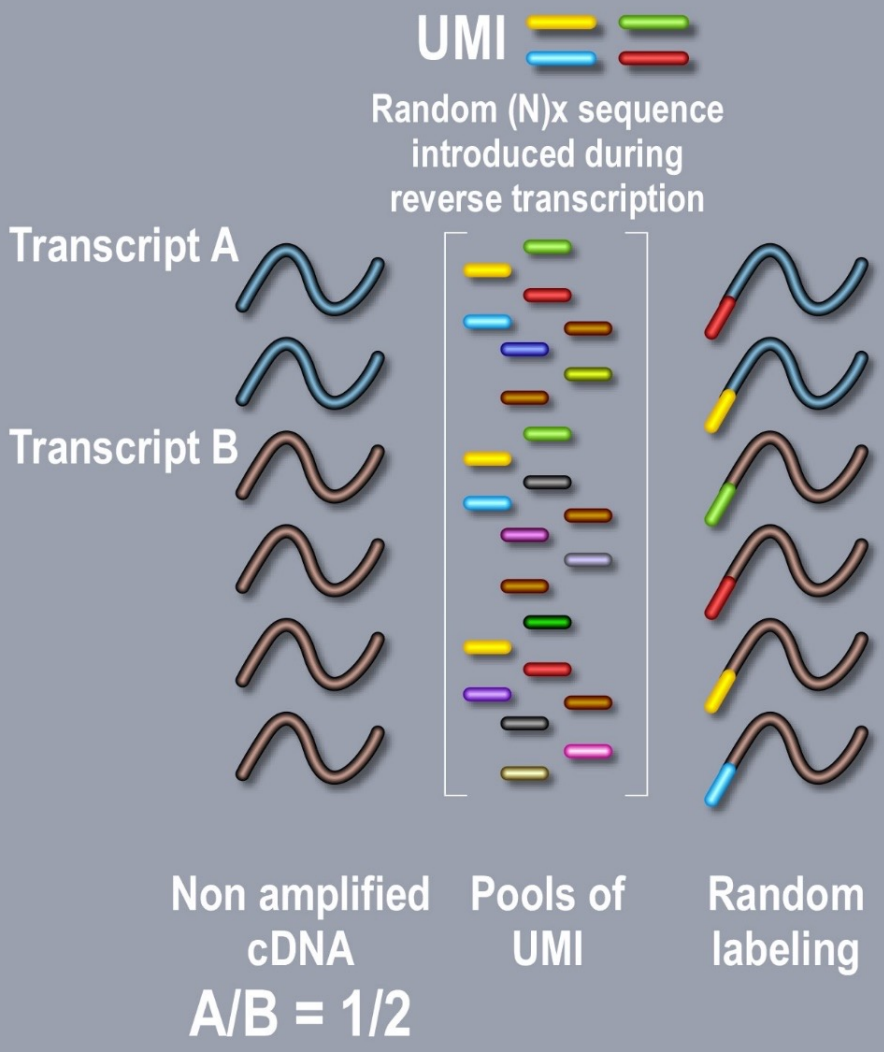
mRNA libraries 100 ng total RNA (10.000 cells)

Single cell mRNA libraries 10 pg total RNA (<<1pg mRNA)

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

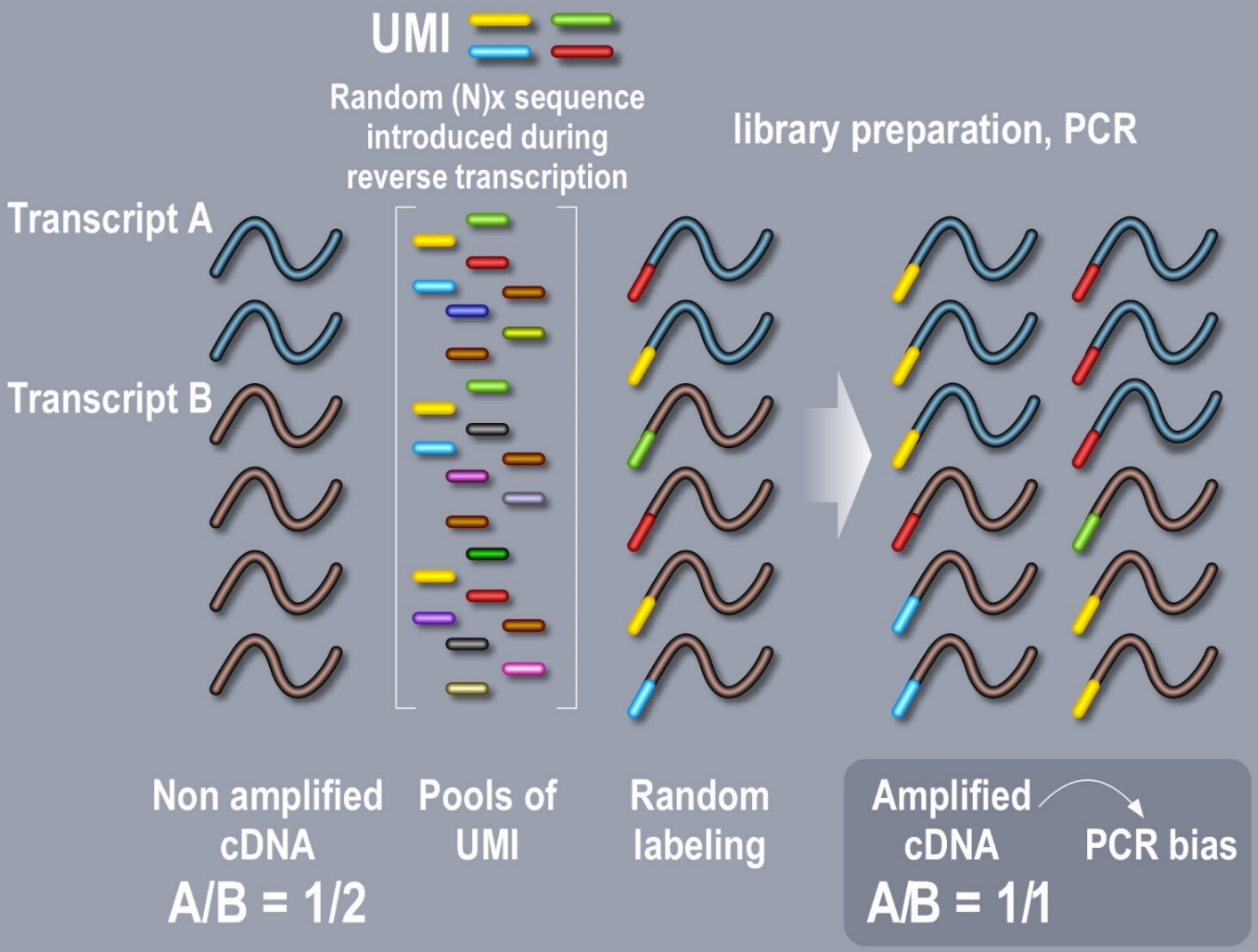
- Spike-in ERCC molecules to evaluate yield and capture efficiency
- Use of Unique Molecular Identifiers to monitor the number of molecules:
  - ***Random multimers to reduce cloning biases during small RNAs profiling*** *(Jayaprakash, NAR, 2011)*
  - ***Detection of rare DNA mutations*** *(Ståhlberg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

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



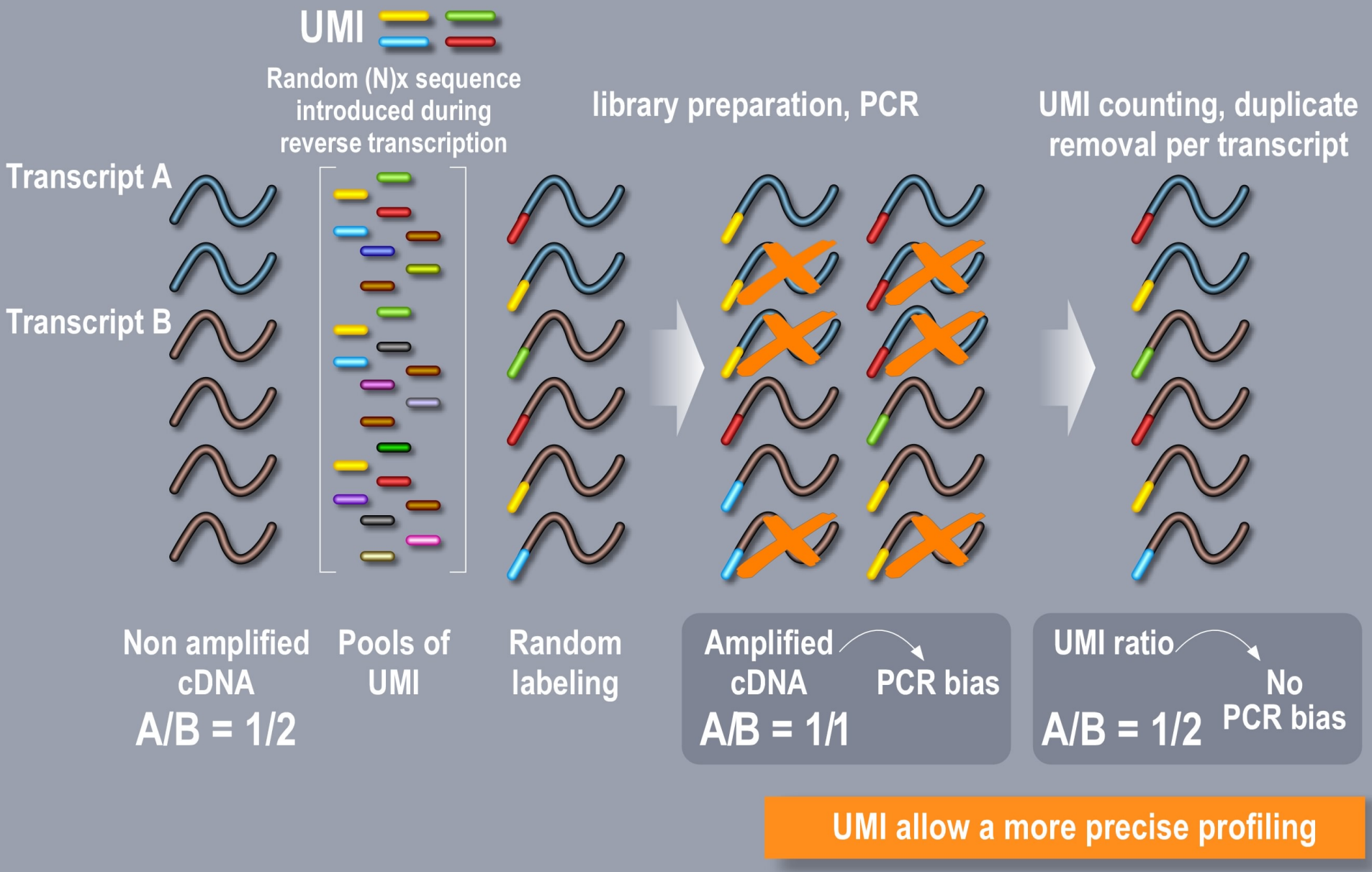
UMIs : *Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Meth 9, 72-74 (2012)*
UMIs for single cell transcriptome: *Islam, S. et al. Quantitative single-cell RNA-seq with UMI . Nat Methods 11, (2014).*

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).*

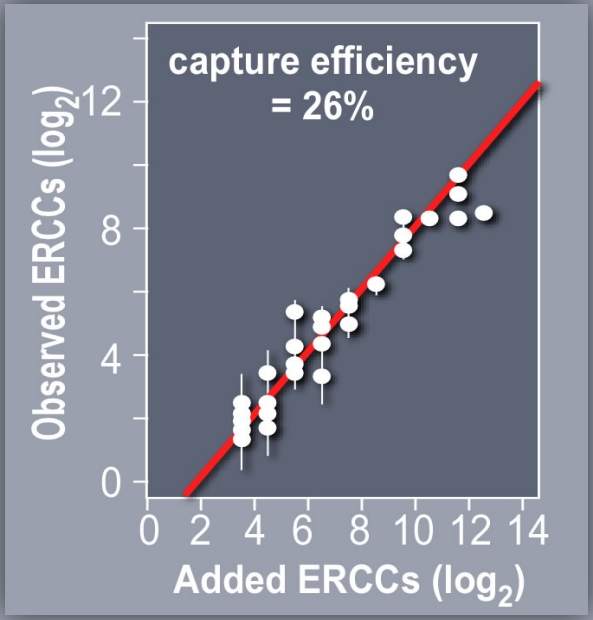
Unique Molecular Identifier (*Islam et al., Nature Methods, 2014*)



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).*

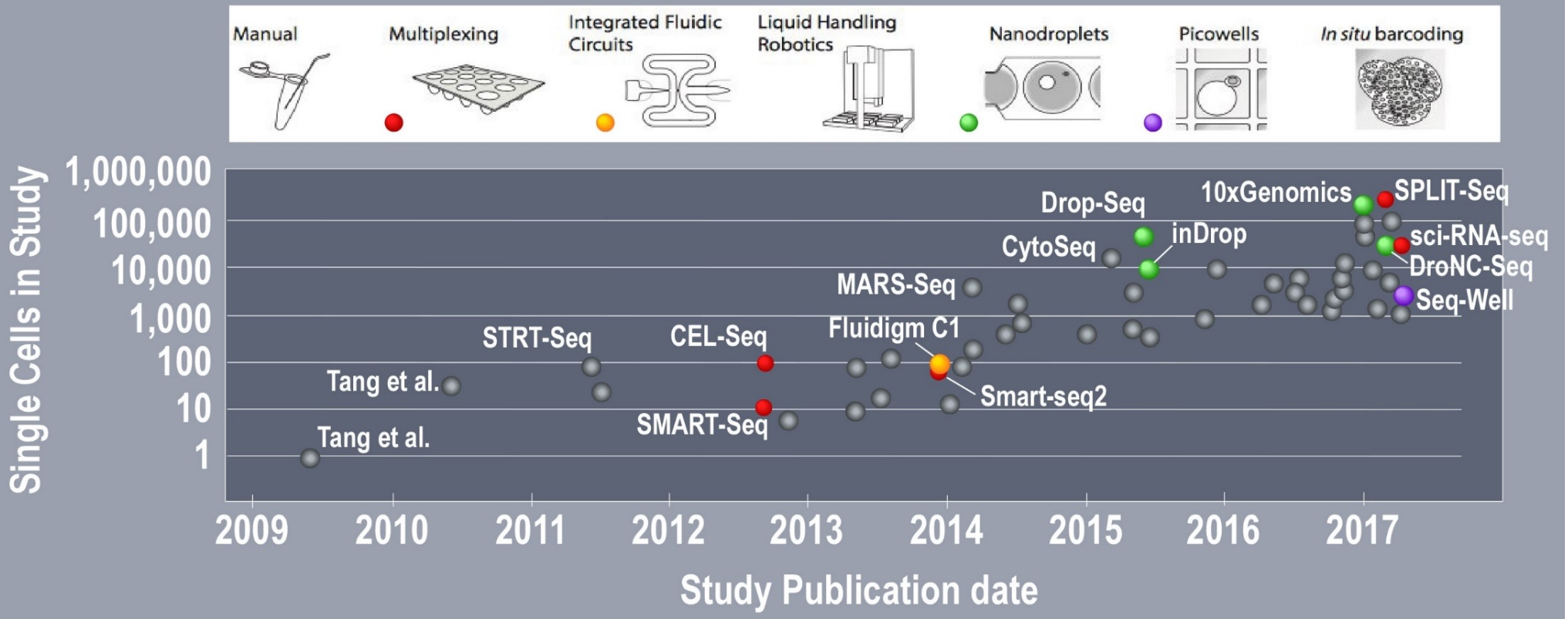
ERCC spike-ins (others set exists: SIRVs, Sequins)

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



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

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



Published: 01 March 2018

Exponential scaling of single-cell RNA-seq in the past decade

Valentine Svensson, Roser Vento-Tormo & Sarah A Teichmann

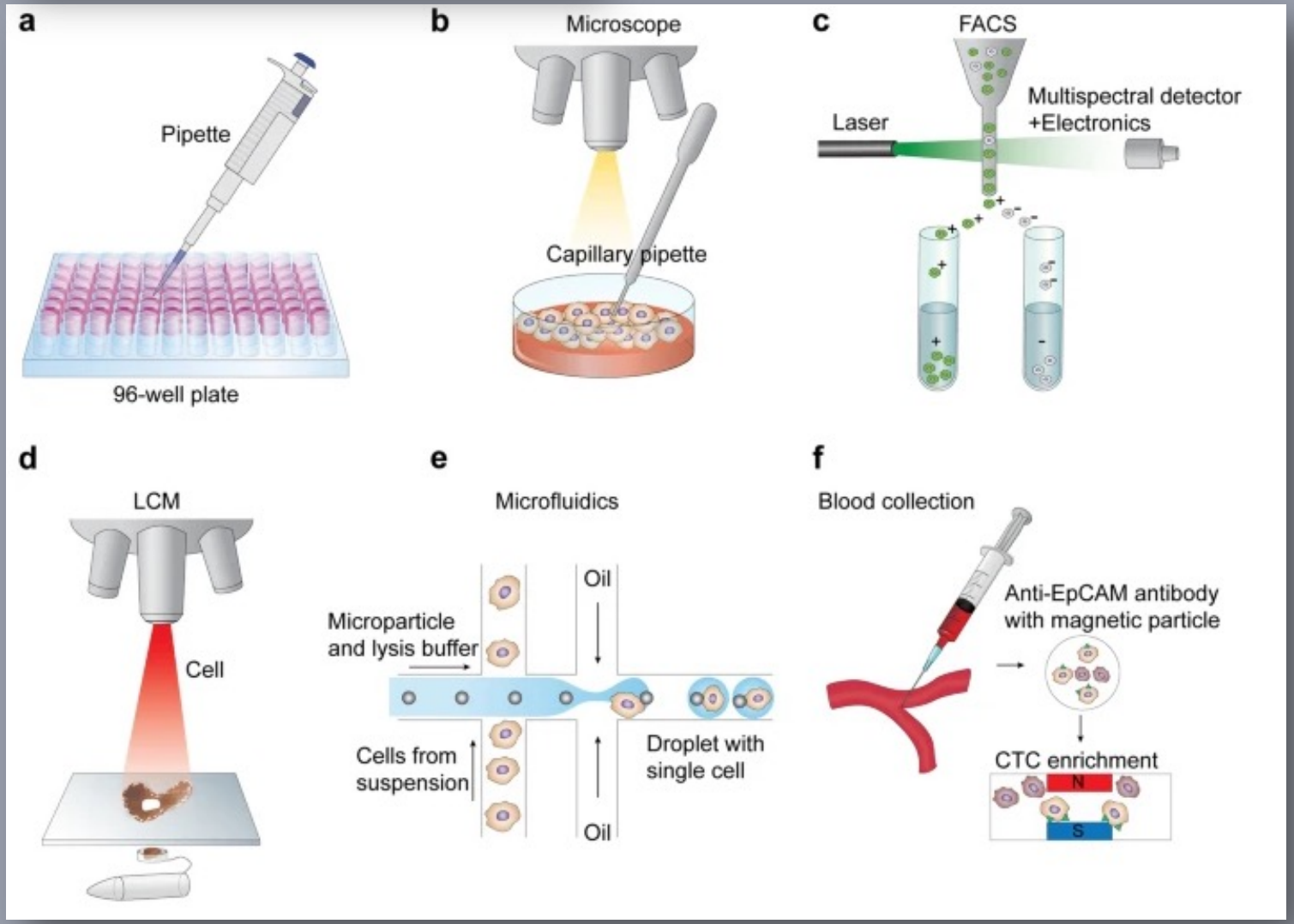
Nature Protocols **13**, 599–604 (2018) | Cite this article

To measure sequences in individual cells, we need method that capture one cell at a time

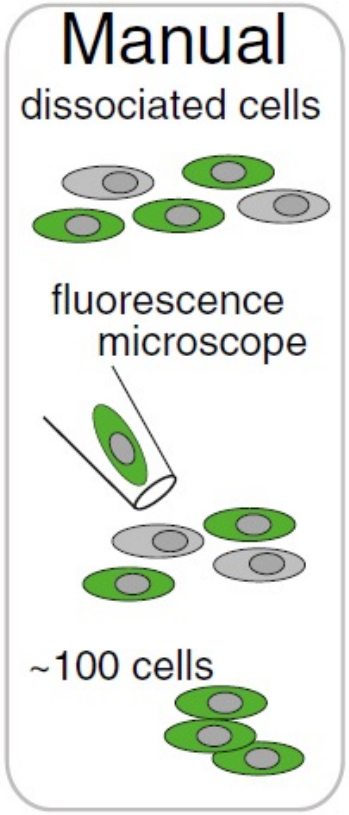
Review Article | Open Access | Published: 07 August 2018

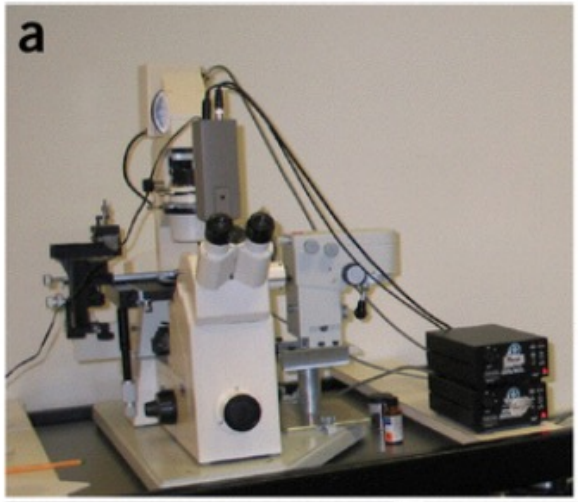
Single-cell RNA sequencing technologies and bioinformatics pipelines

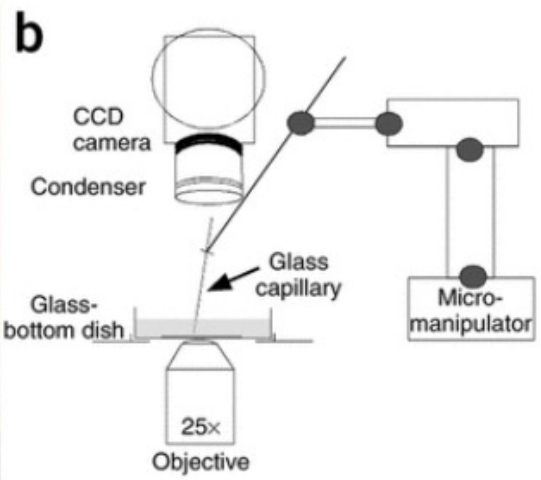
Byungjin Hwang, Ji Hyun Lee ✉ & Duhee Bang ✉

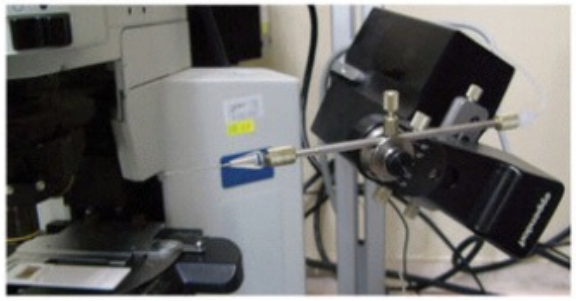


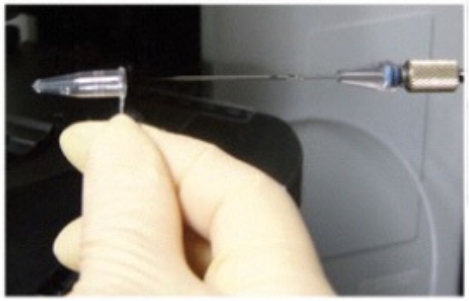
Manual Cell sorting





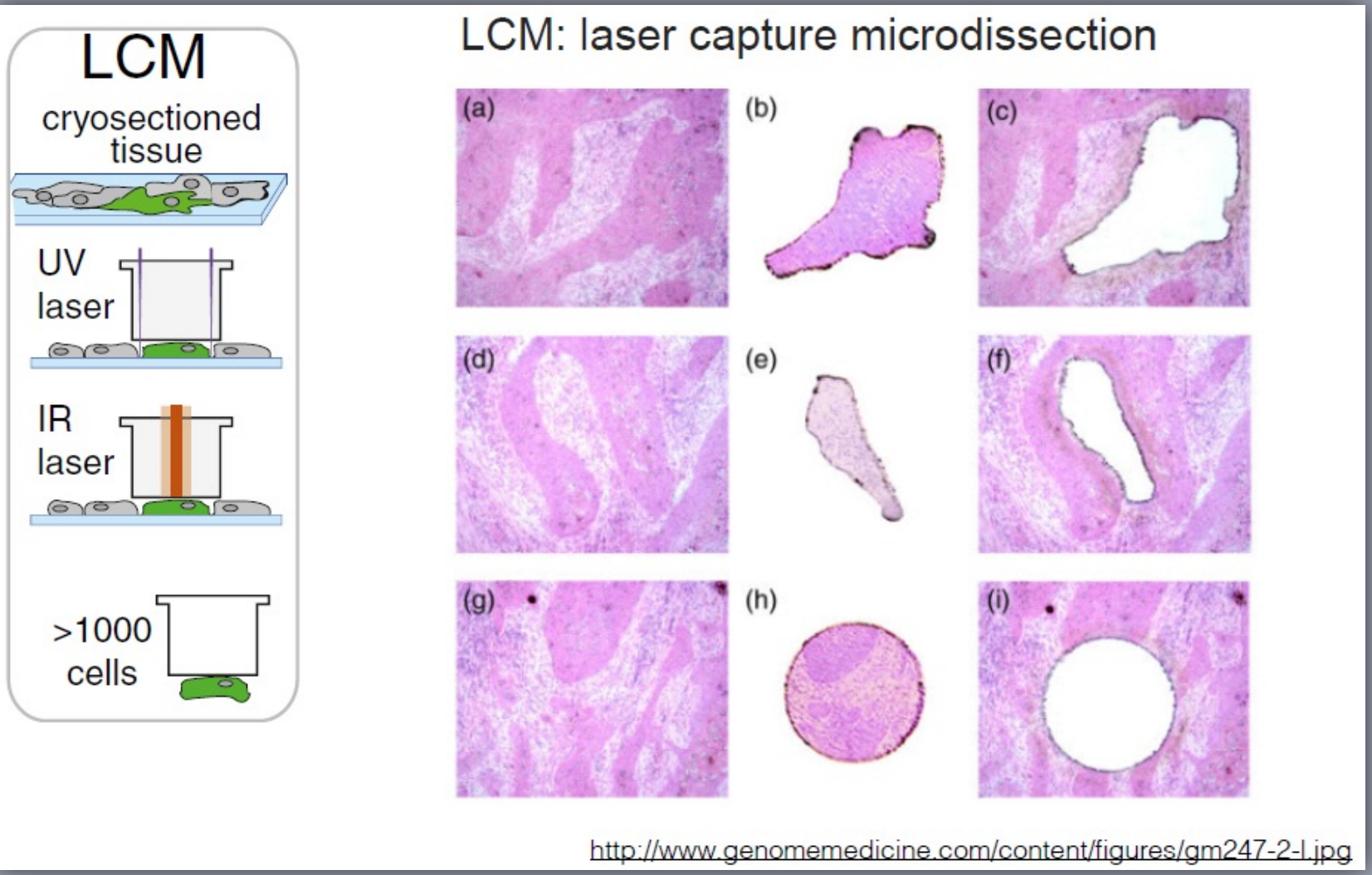




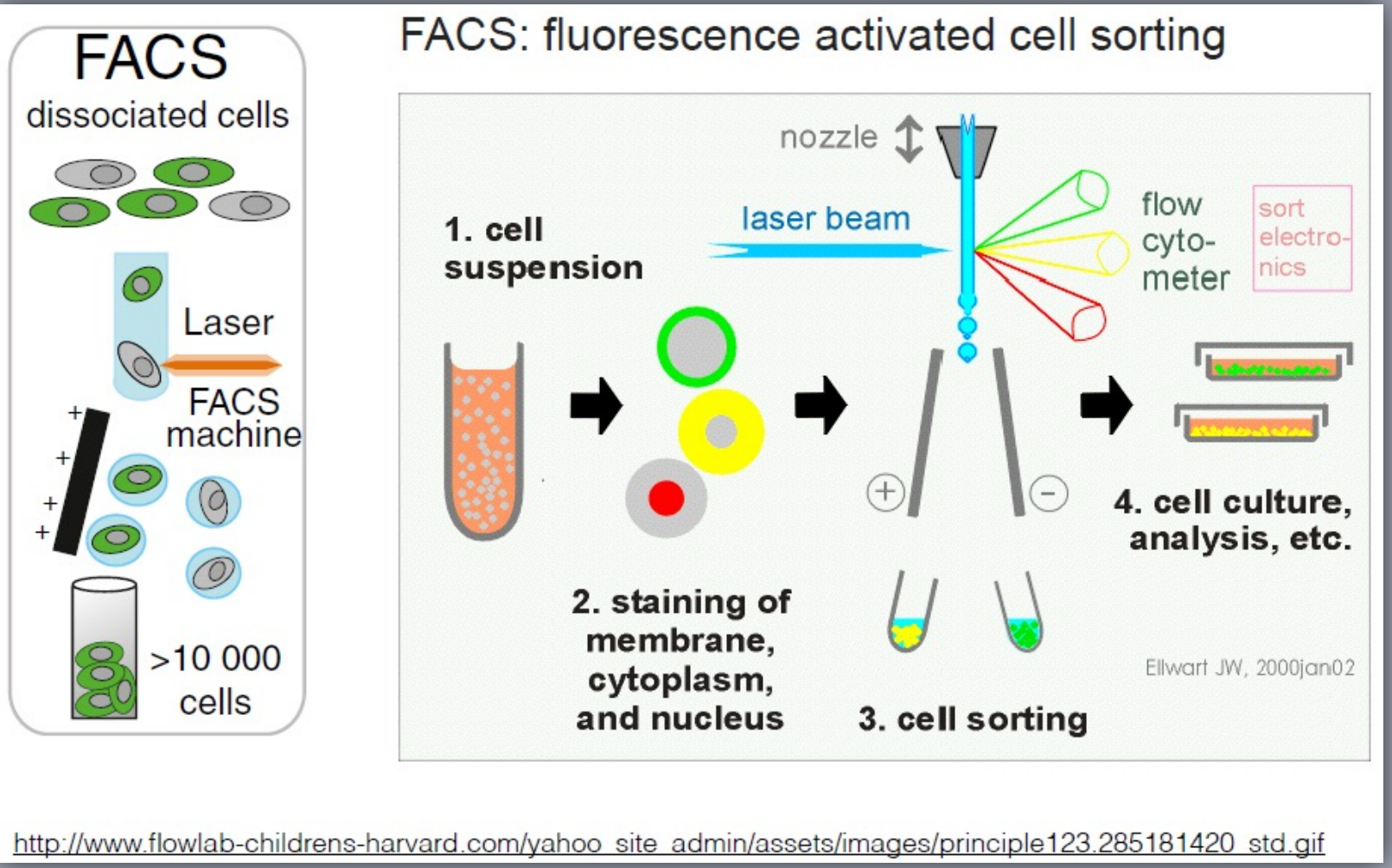


http://www.nature.com/nprot/journal/v6/n5/images_article/nprot.2011.322-F2.jpg

Laser Capture Microdissection



FACS Cell Sorting

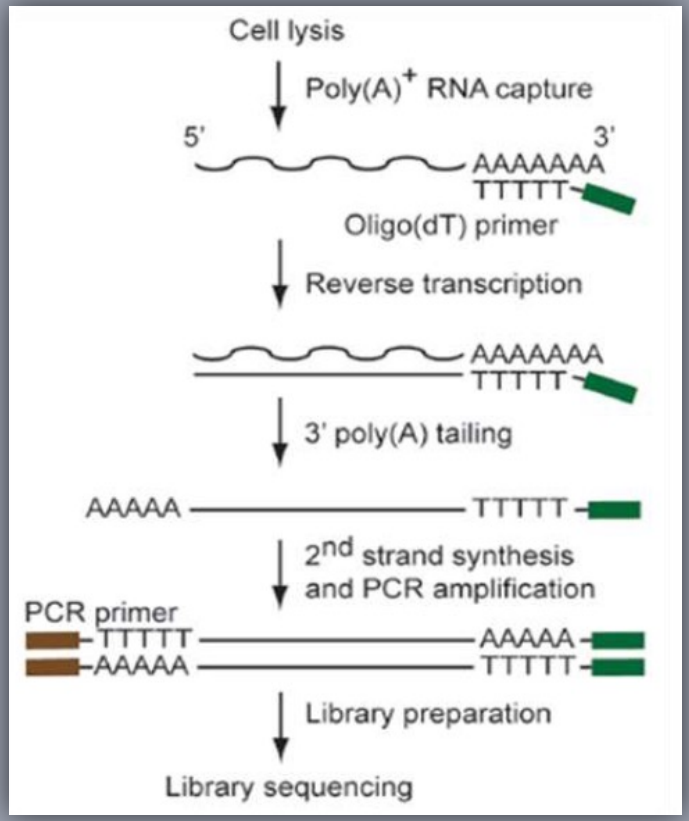


http://www.flowlab-childrens-harvard.com/yahoo_site_admin/assets/images/principle123.285181420_std.gif

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

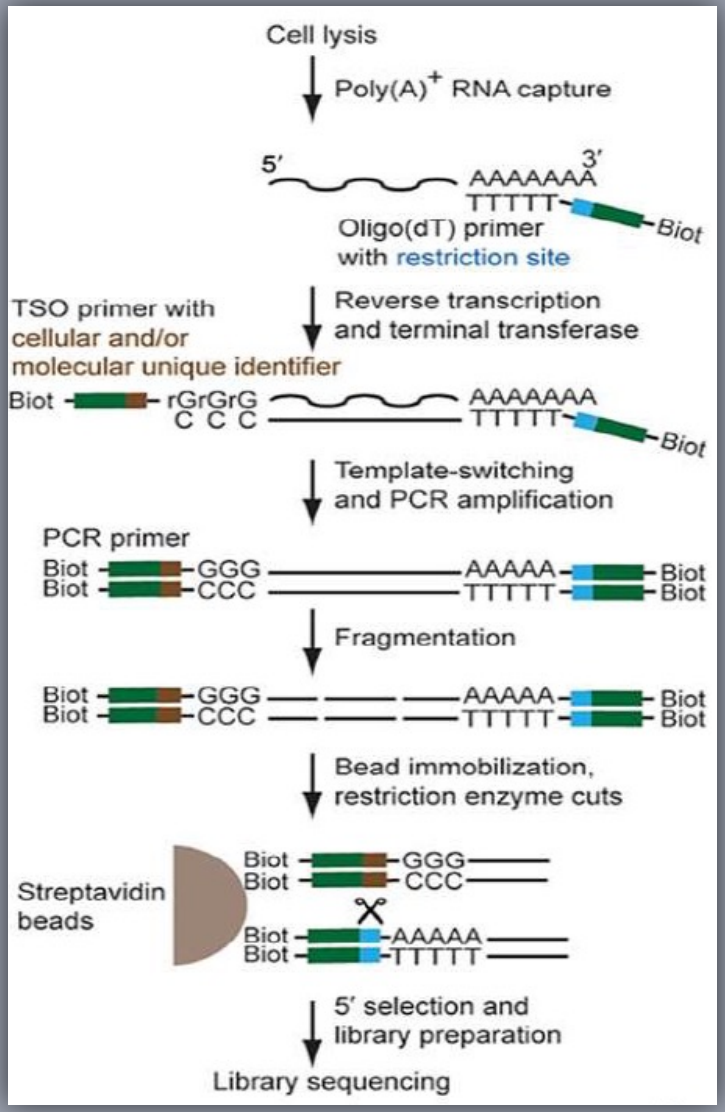
- Premature termination of RT reduces transcript coverage at the 5' end
- Introduction of a polyA tail in addition to its own polyA sequence at the **3' end** of the input RNA causes a **loss of strand information** in the resulting double-stranded cDNA

cDNA expression profile of a single blastomere
→ more than 100 million from the single blastomere
+2 wild-type mature oocytes

STRT: Single cell tagged reverse transcription, Islam (2011)

Protocol

- based on **template switching (TSO)**,
- **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**

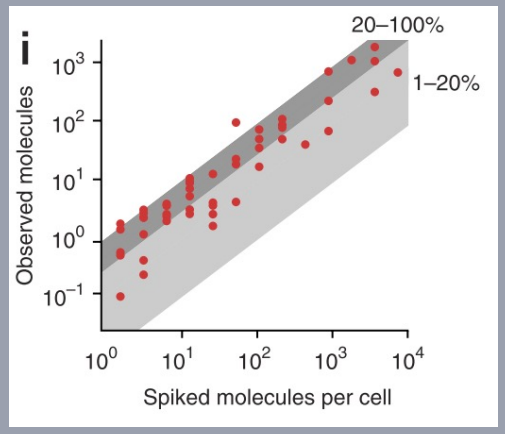


The feasibility of the strategy was demonstrated by analyzing the transcriptomes of 85 single cells of two distinct types, mouse embryonic stem cells / embryonic fibroblasts (MEFs)

STRT: Single cell tagged reverse transcription, Islam (2011), +UMI (2014)

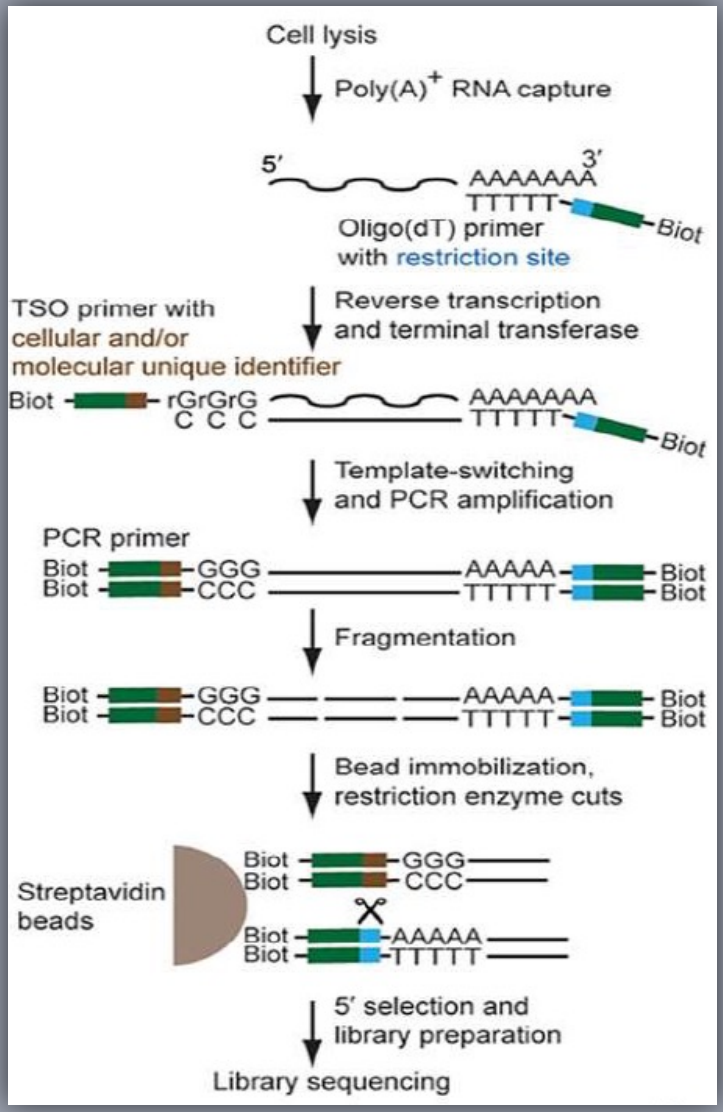
Protocol

- based on **template switching (TSO)**,
- **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**



48 % !!! (41 ES cells)

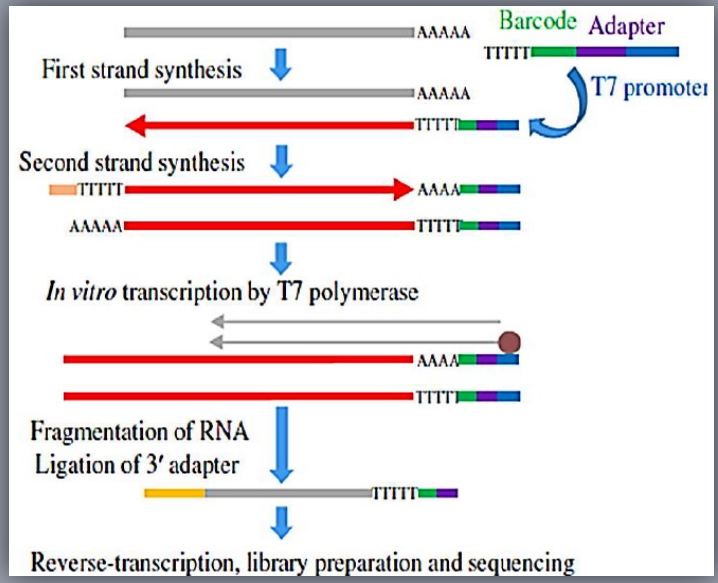
Zeisel et al., (2015), primary somatosensory cortex (S1) and hippocampal CA1 of **3,005 cells**
same protocol →22%
more realistic !



Cell expression by linear amplification and sequencing CEL-seq (2012)

Protocol

- OligodT primer containing the 5' Illumina adaptor, a cell barcode, and a T7promoter,
- 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**

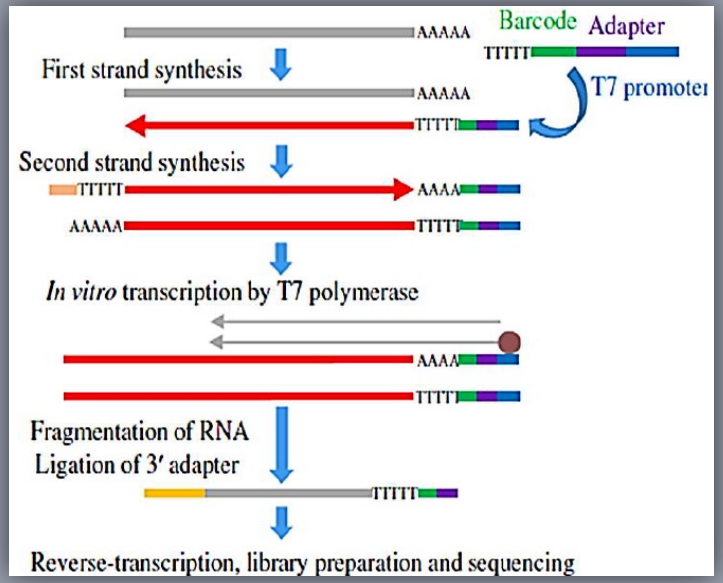


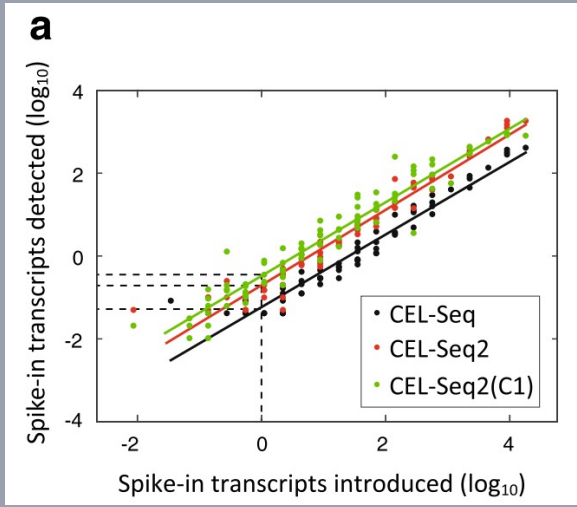
9 mouse embryonic stem (ES) cells and 7 mouse embryonic fibroblasts (MEFs)

Cell expression by linear amplification and sequencing CEL-seq (2012), CEL-seq2 (2016)

Protocol

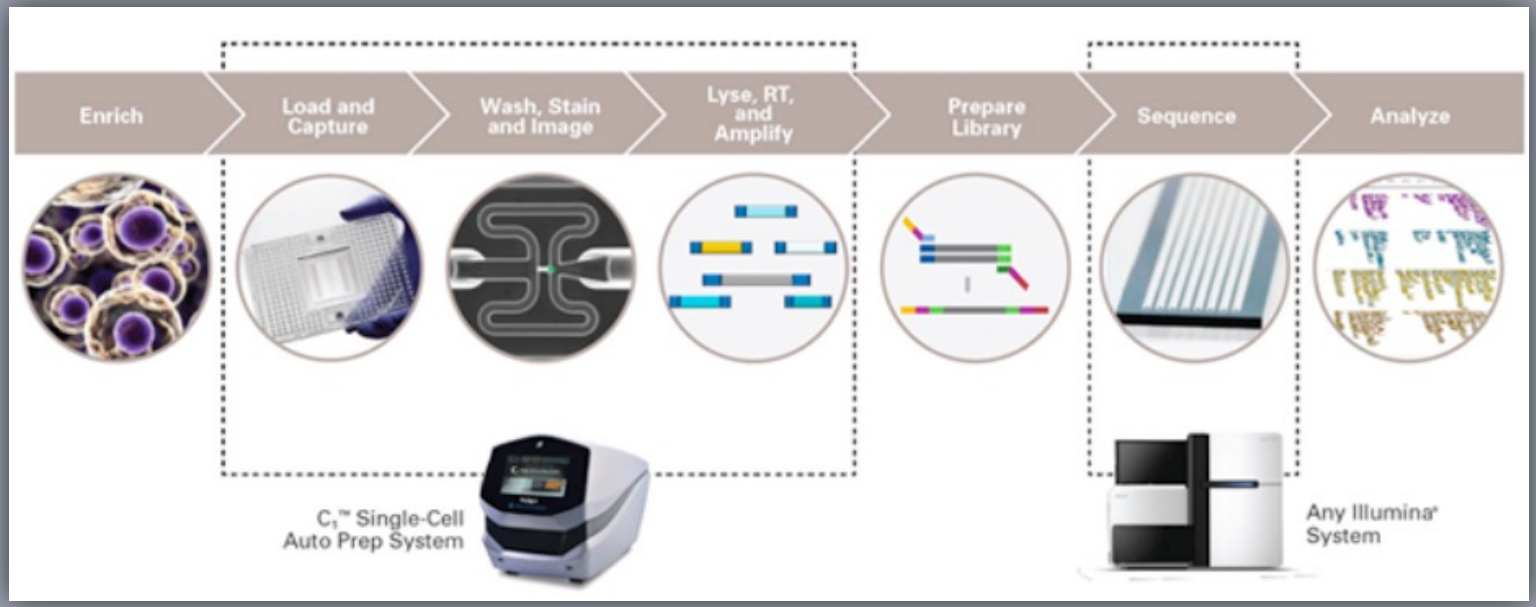
- OligodT primer containing the 5' Illumina adaptor, a cell barcode, and a T7promoter (**CEL-seq2 add UMI**),
- RT and second-strand synthesis,
- cDNA from all the cells is pooled and amplified **by in vitro transcription** from the T7 promoter,
- RNA fragmentation, Illumina adaptor ligated at 3' end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





The intersect with the y-axis provides a measure of the efficiency and brings us to 19.7 %, relative to 5.8 % for CEL-Seq. On the C1 an even higher efficiency is obtained of 22 %.

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

SMART-Seq, Ramskold (2012), SMART-seq2, Picelli (2014)

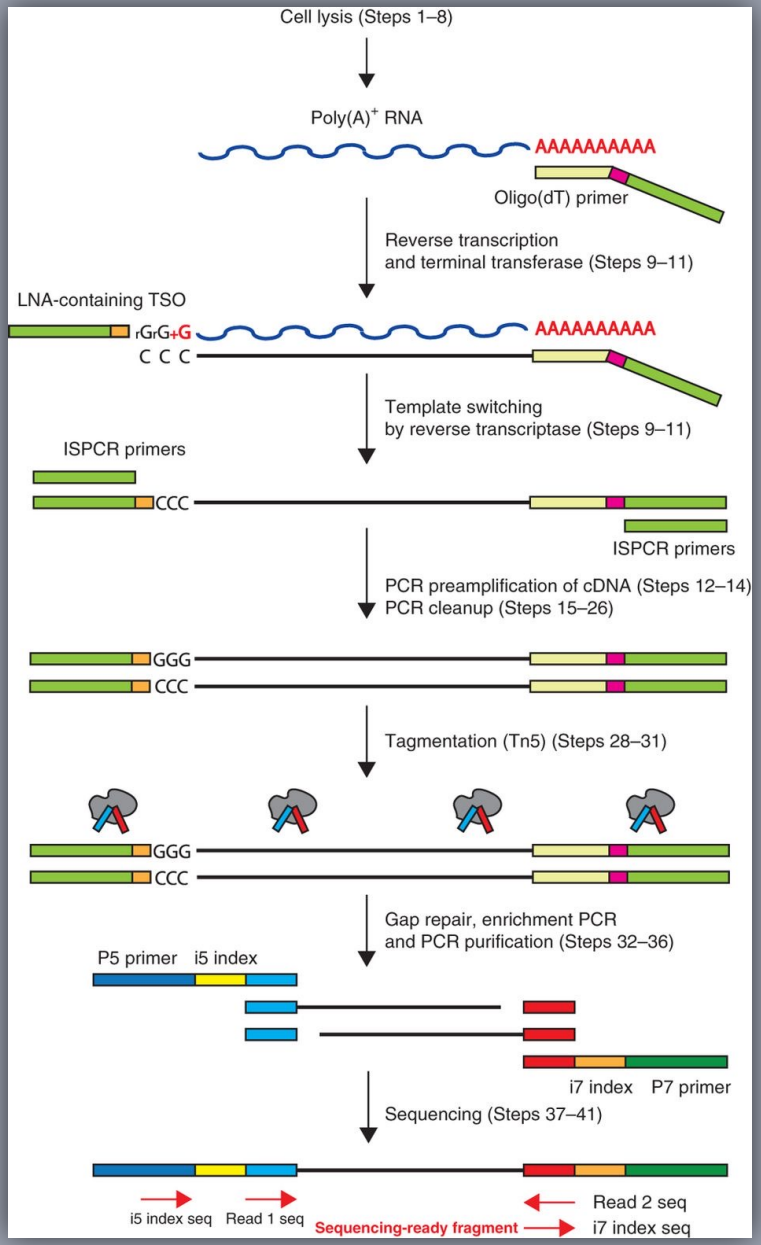
SMART= Switching Mechanism at the end of the 5´-end of the RNA Transcript

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**

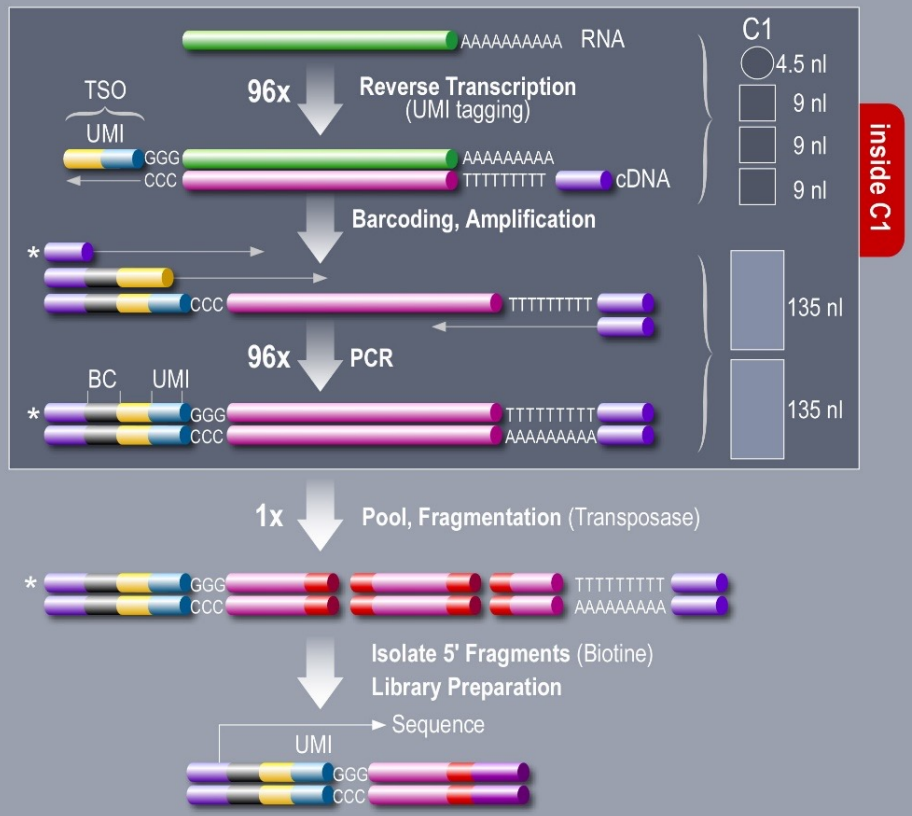
Drawback

- No UMIs



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

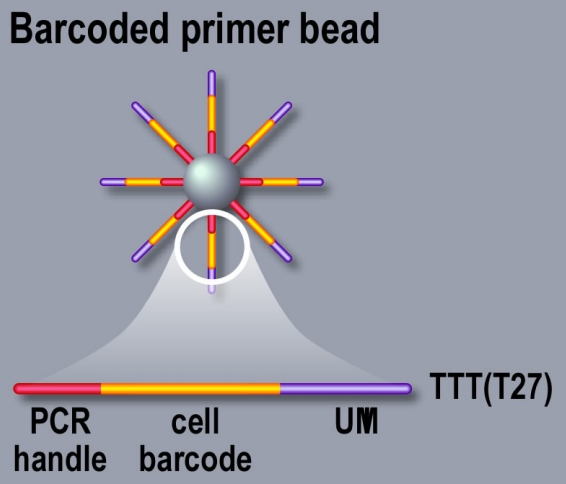
Arguel et al.
Pascal Barbry - UCA Genomix - IPMC, CNRS - University of Cote d'Azur

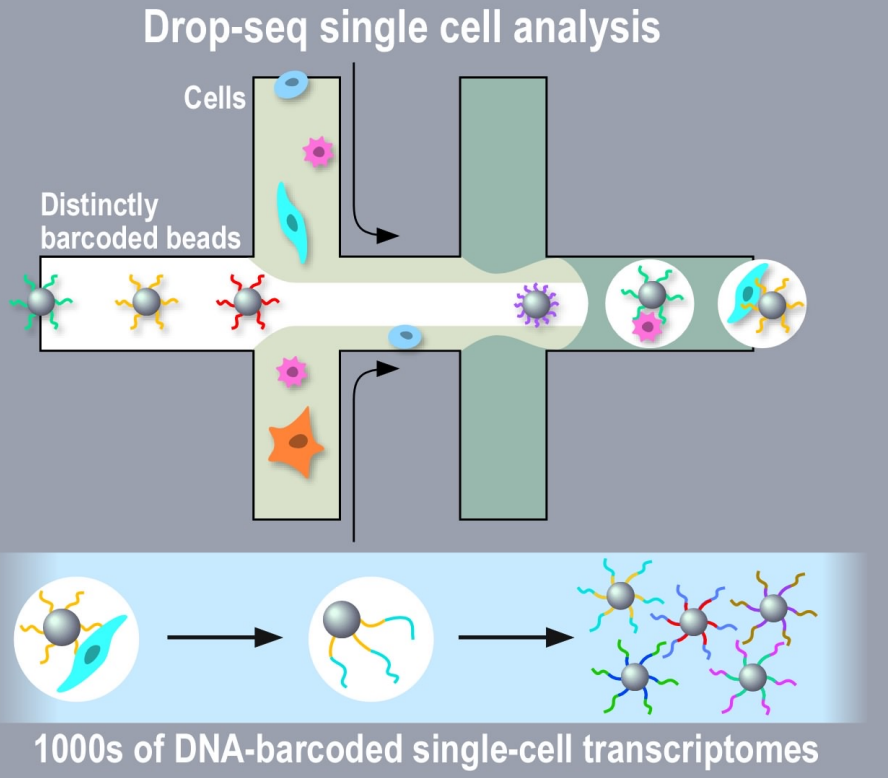
Drop-seq / inDROP: droplets-based scRNA-seq (*Mascosko et al., Klein et al., CELL, may 2015*)

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

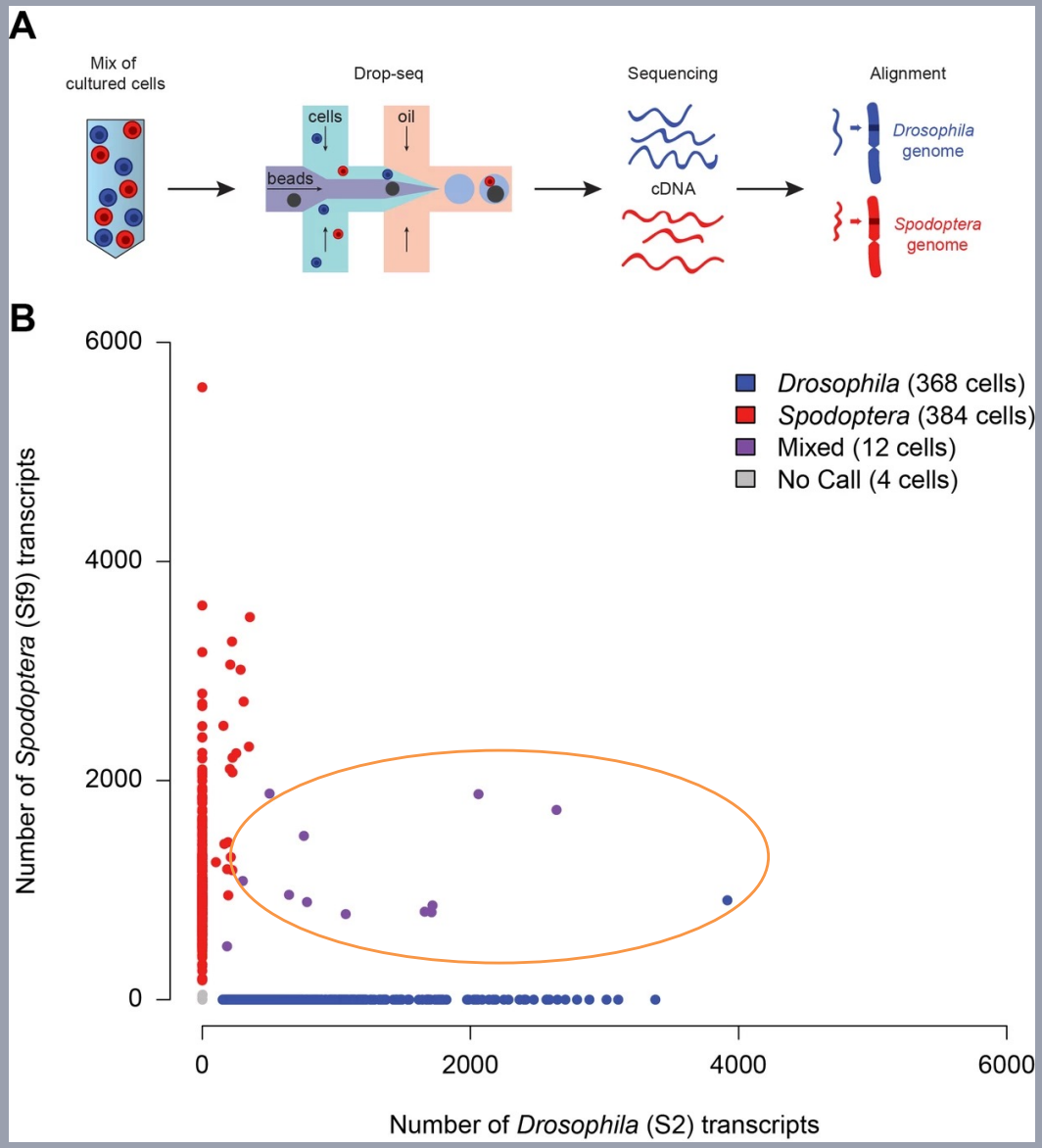
Evan Z. Macosko[1, 2, 3], Anindita Basu[4, 5], Rahul Satija[4, 6, 7], James Nemesh[1, 2, 3], Karthik Shekhar[4], Melissa Goldman[1, 2], Itay Tirosh[4], Allison R. Bialas[8], Nolan Kamitaki[1, 2, 3], Emily M. Martersteck[9], John J. Trombetta[4], David A. Weitz[5, 10], Joshua R. Sanes[9], Alex K. Shalek[4,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





Doublet detection : the barnyard plot



Croset et al. 2018. DOI:10.7554/eLife.34550

% of cells loaded	% of cells Recovered	Multiplet rate (%)
~870	~500	~0.4%
~1.700	~1.000	~0.8%
~3.500	~2.000	~1.6%
~5.300	~3.000	~2.3%
~7.000	~4.000	~3.1%
~8.700	~5.000	~3.9%
~10.500	~6.000	~4.6%
~12.200	~7.000	~5.4%
~14.000	~8.000	~6.1%
~15.700	~9.000	~6.9%
~17.400	~10.000	~7.6%

10x Genomics Chromium