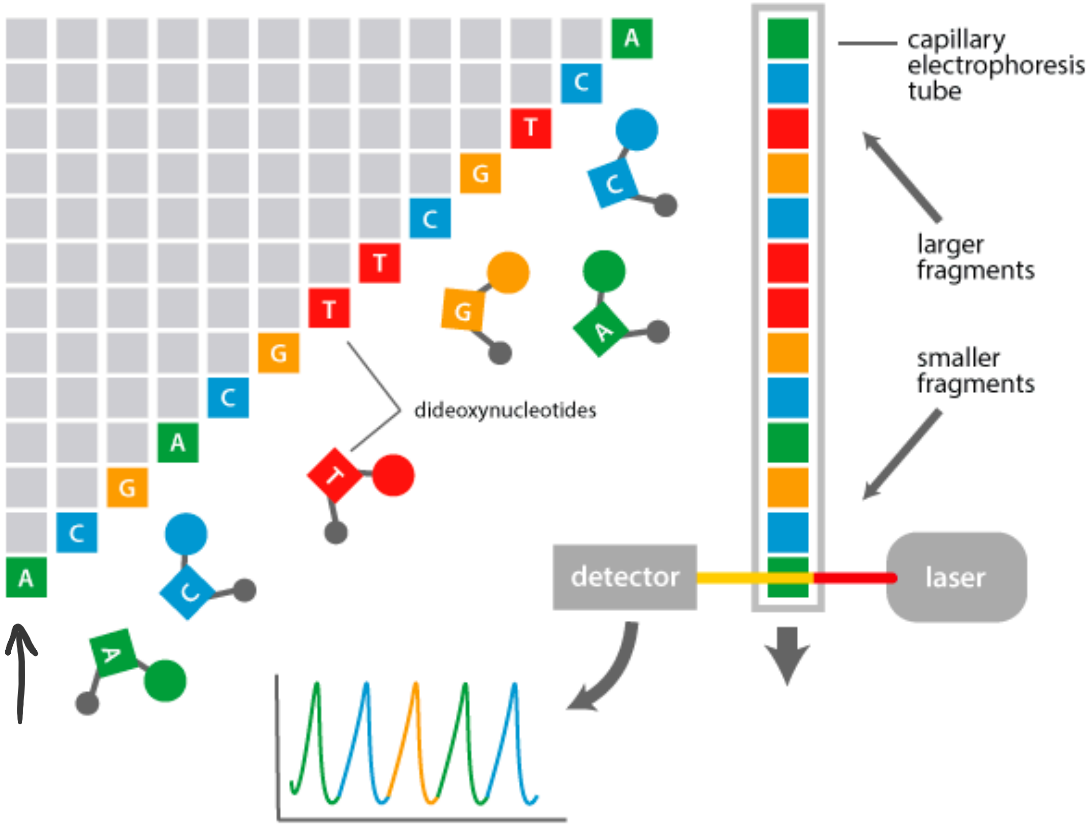
THE OMICS ERA

Genomics and its applications; Proteomics and single-cell technologies

Sanger sequencing

high throughput

dNTP
+
ddNTP



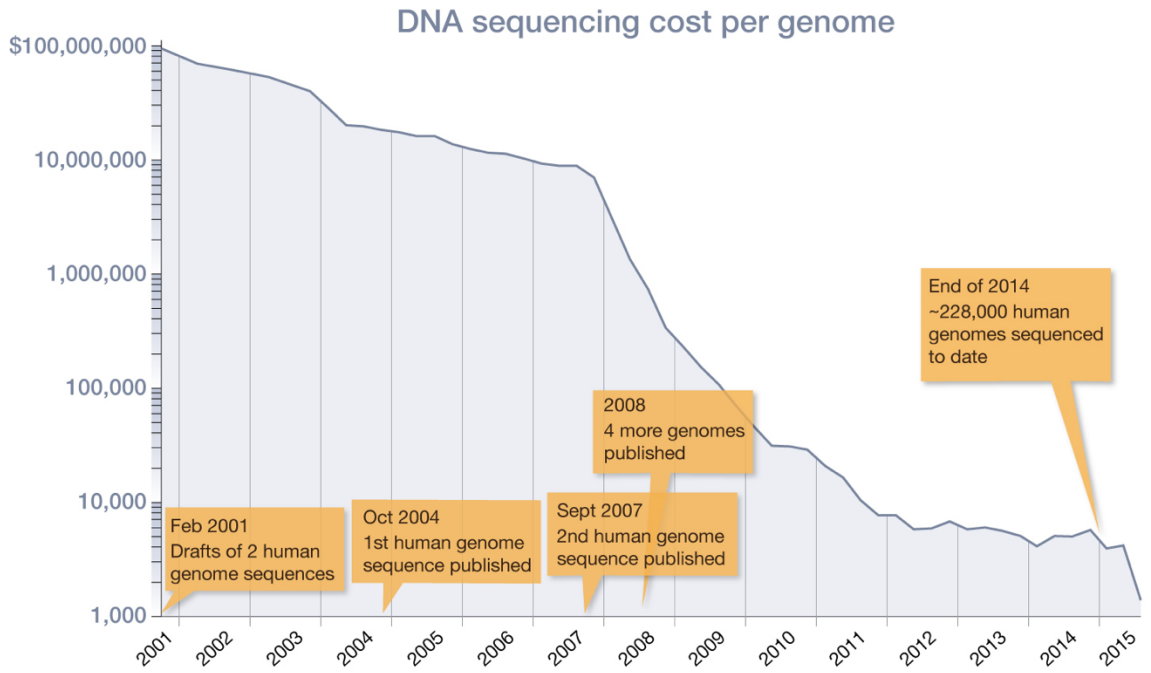
https://www.abmgood.com/marketing/knowledge_base/next_generation_sequencing_introduction.php#sanger

Sanger sequencing

- https://www.youtube.com/watch?v=jFCD8Q6qSTM&list=PL_VcB7OJ1TCAWRXN6vnC5IKbMHjIMtN8P&index=2

Human Genome Project

- Used Sanger Sequencing
- Took ~13 years (started in 1985!!!)
- Spent ~3 billion USD!!!
- Today: ~100 USD



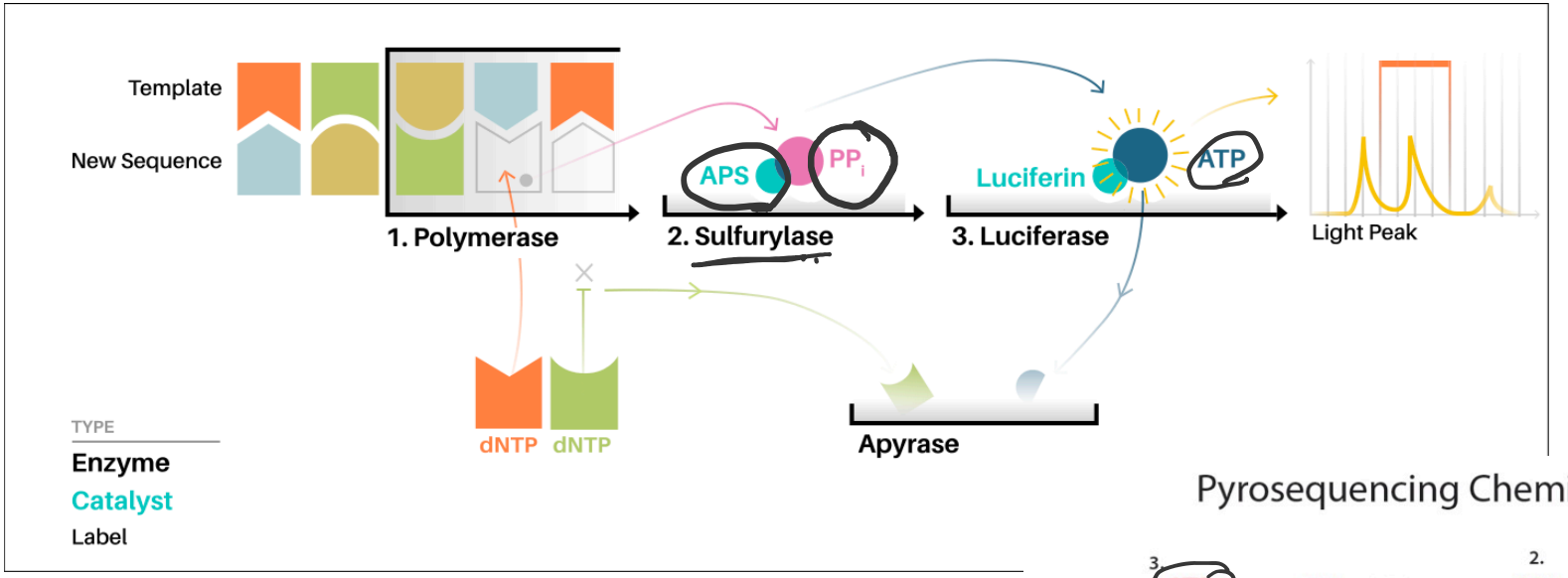
DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). http://www.genome.gov/sequencingcosts/.

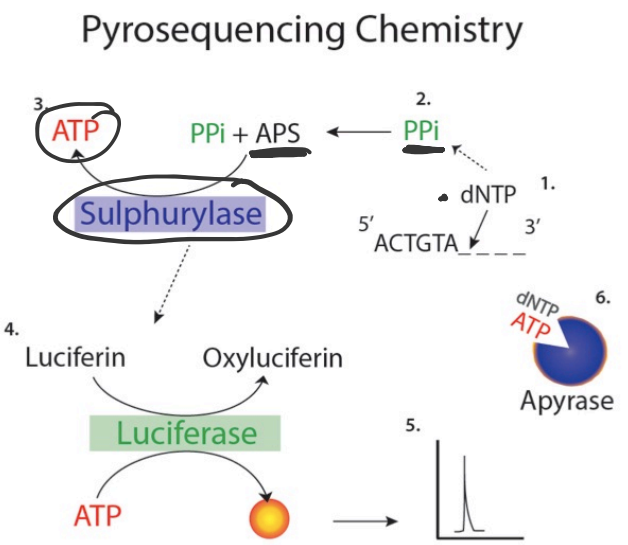
Nature editorial staff (2010). Human genome at ten: The sequence explosion. Nature, 464, 670-671. doi:10.1038/464670a

Next-generation sequencing (NGS)

- Massively parallel
- Huge output of data
- Decreasing costs
- Fast
- https://www.youtube.com/watch?v=jFCD8Q6qSTM&list=PL_VcB7OJ1TCAWRXN6vnC5IKbMHjIMtN8P&index=2

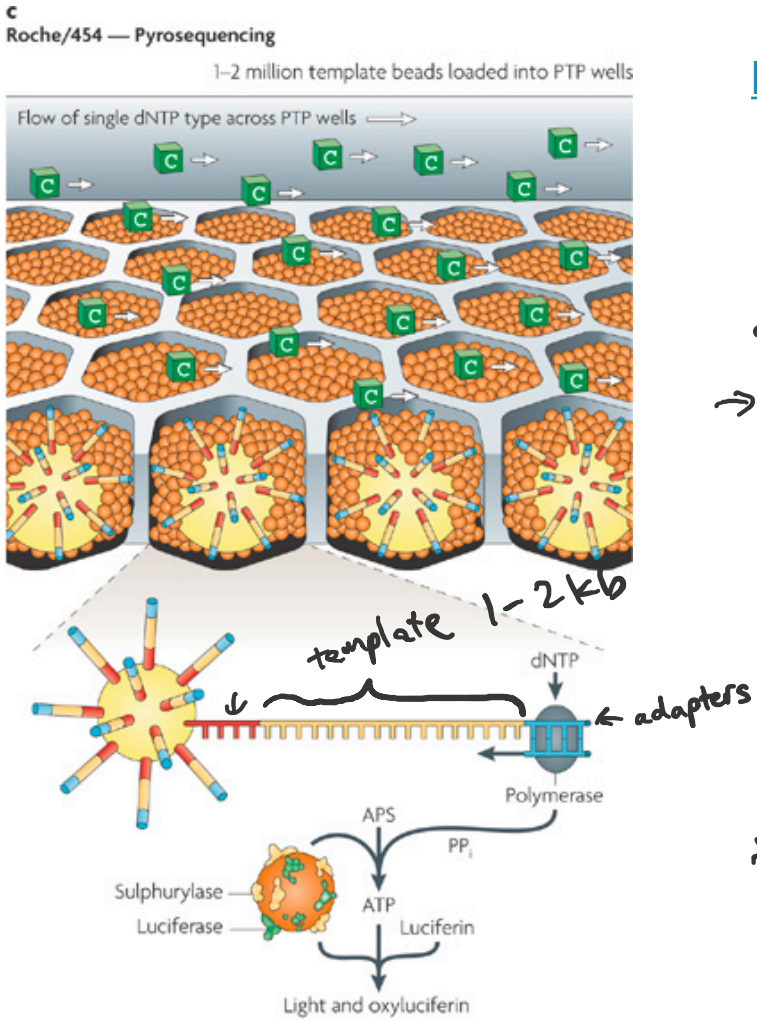
Roche 454: Pyrosequencing



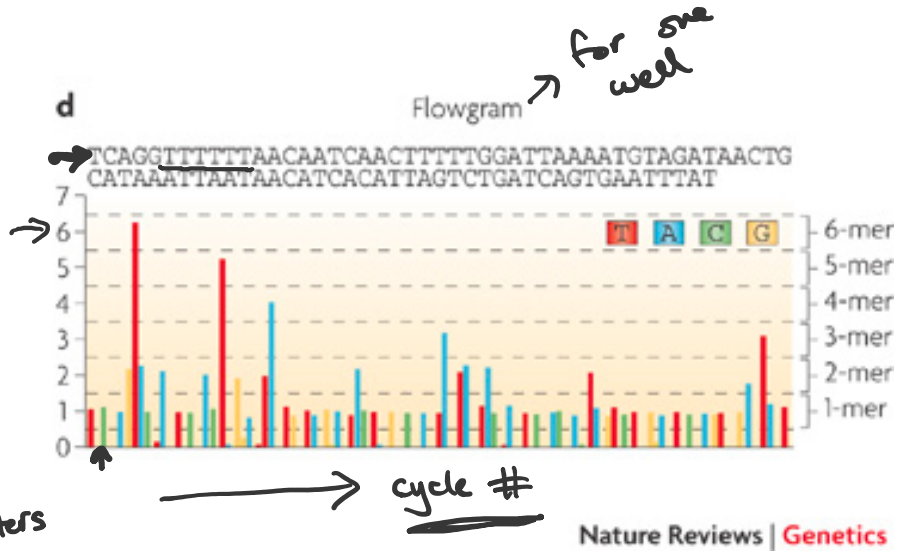


Jacopo Pompilii, DensityDesign Research Labhttps://commons.wikimedia.org/w/index.php?curid=37083509
http://bitesizebio.com/19008/how-bisulfite-pyrosequencing-works/

Roche 454: Pyrosequencing



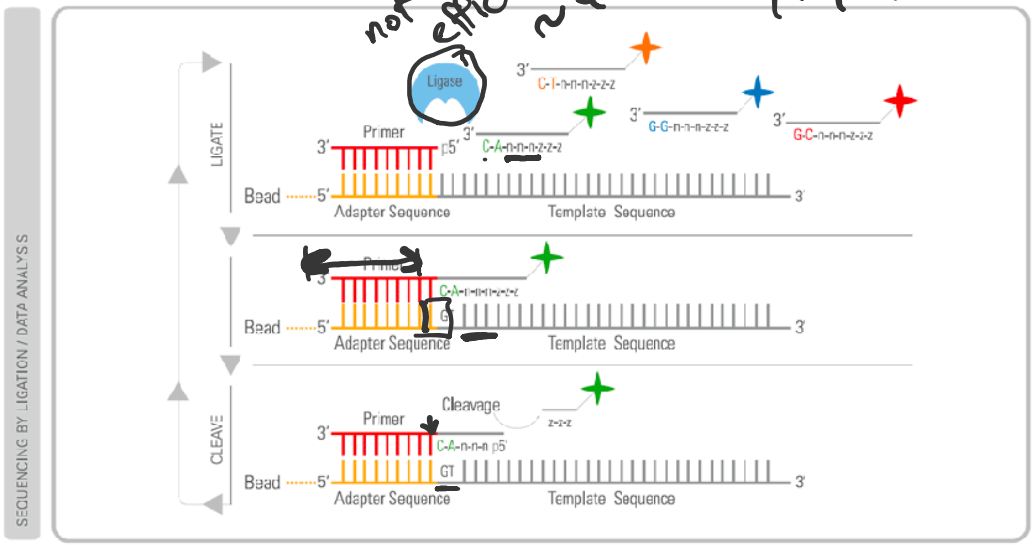
https://youtu.be/jFCD8Q6qSTM?t=3m40s`

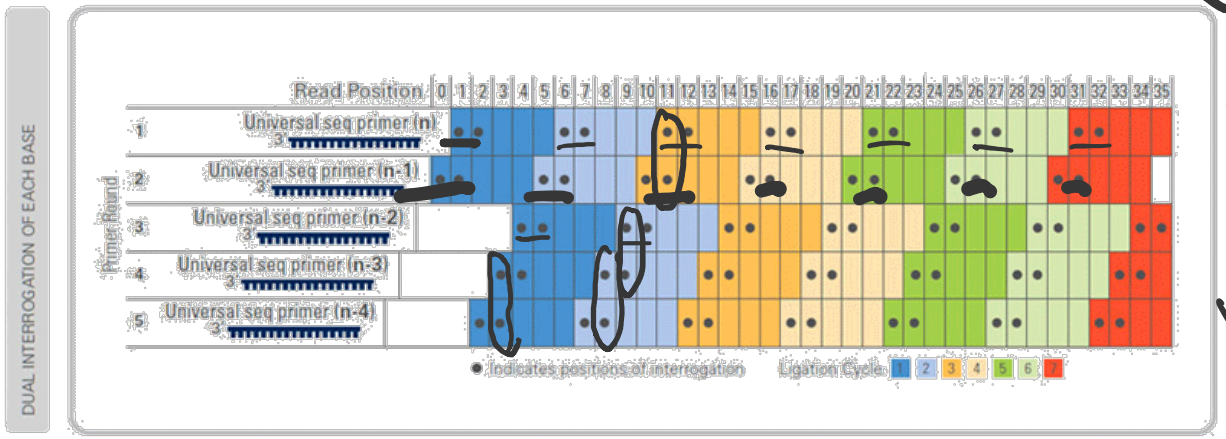


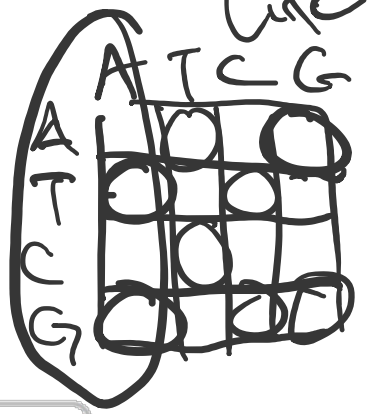
* Homopolymer runs Errors

ML Metzker, Nature Review Genetics, 2009

SOLiD: Sequencing by Ligation





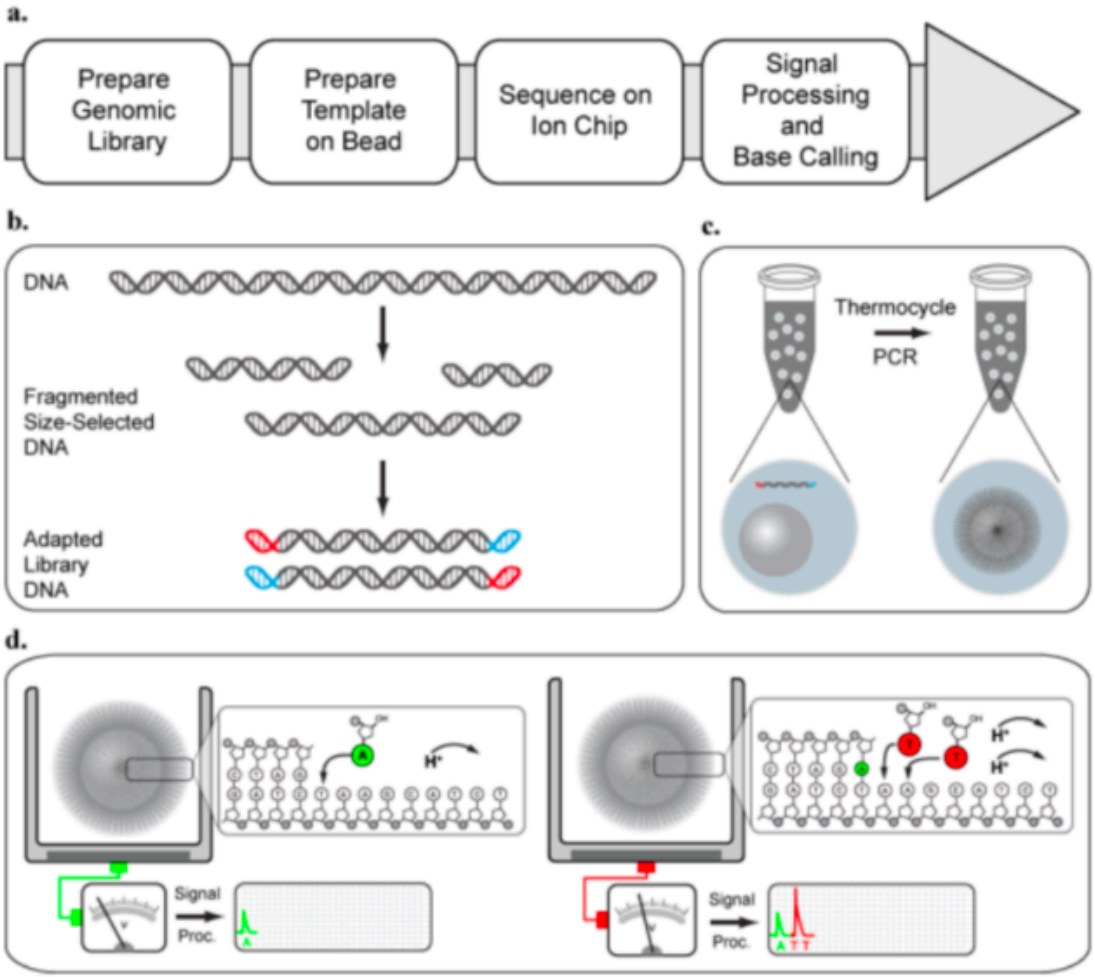


Figures from Applied Biosystems website

https://youtu.be/jFCD8Q6qSTM?t=5m33s

Ion Torrent

https://youtu.be/jFCD8Q6qSTM?t=6m42s

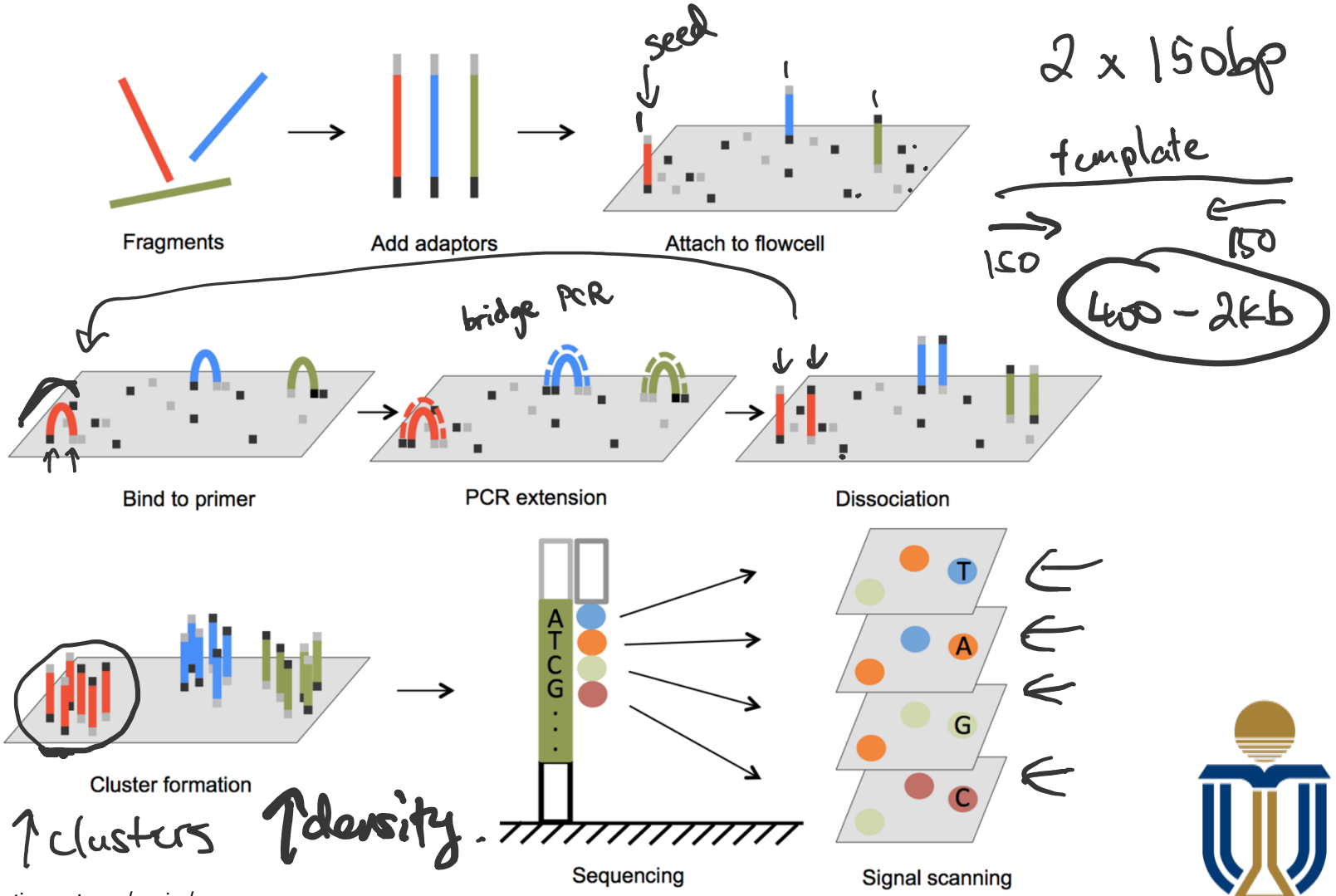


http://www.genomics.cn/en/navigation/show_navigation?nid=2640

Illumina: Sequencing-by-synthesis (SBS)

- Video: https://www.youtube.com/watch?v=HMyCqWhwB8E
- Video: https://www.youtube.com/watch?v=jFCD8Q6qSTM&list=PL_VcB7OJ1TCAWRXN6vnC5IKbMHjlMtN8P&index=2
- Illumina company PDF: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

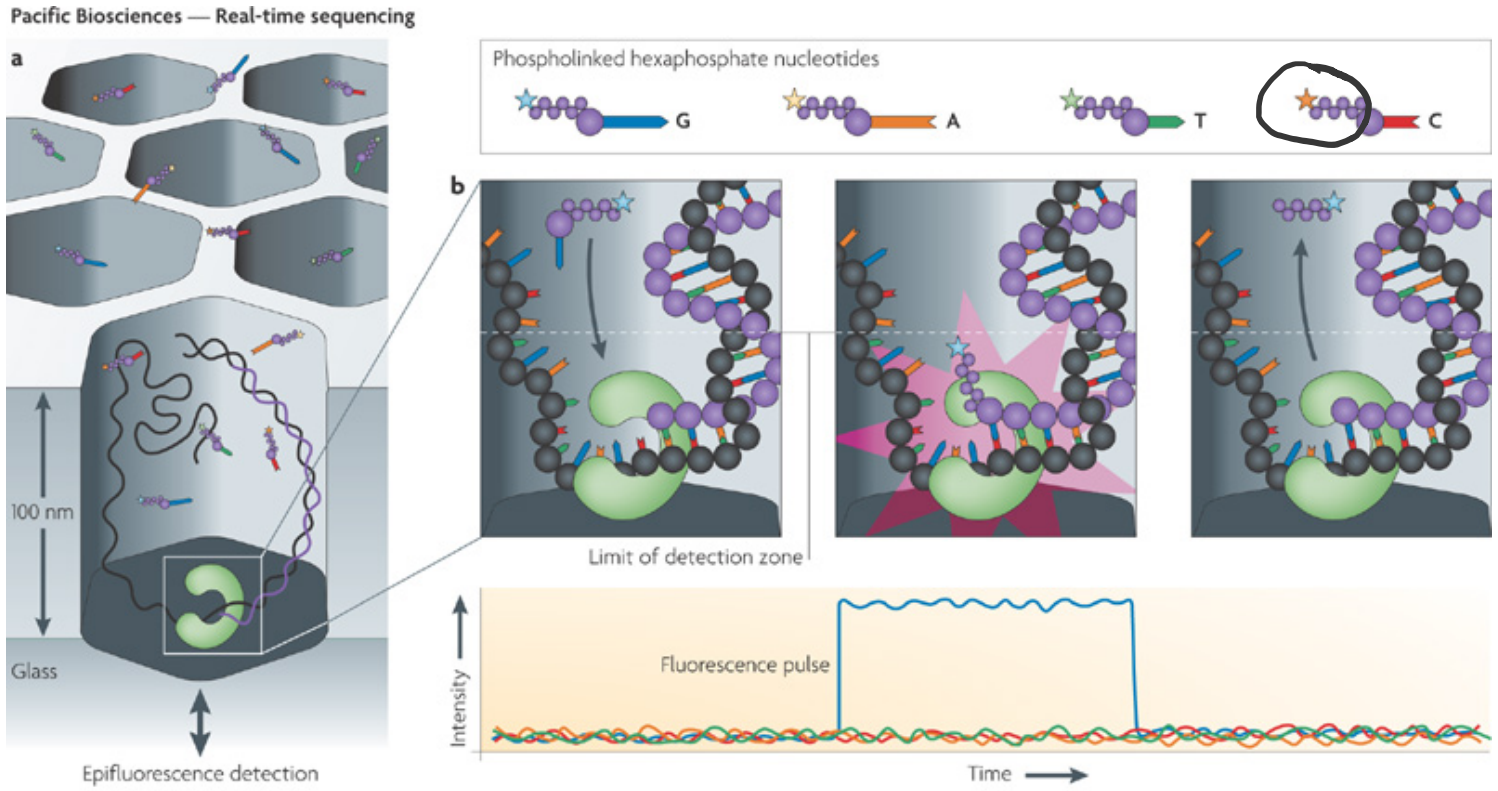
Illumina: Sequencing-by-synthesis (SBS)



http://www.3402bioinformaticsgroup.com/service/

PacBio.

Pacific Biosciences: SMS → single molecule sequencing



ZMW - zero mode waveguide. Read length 2kb.

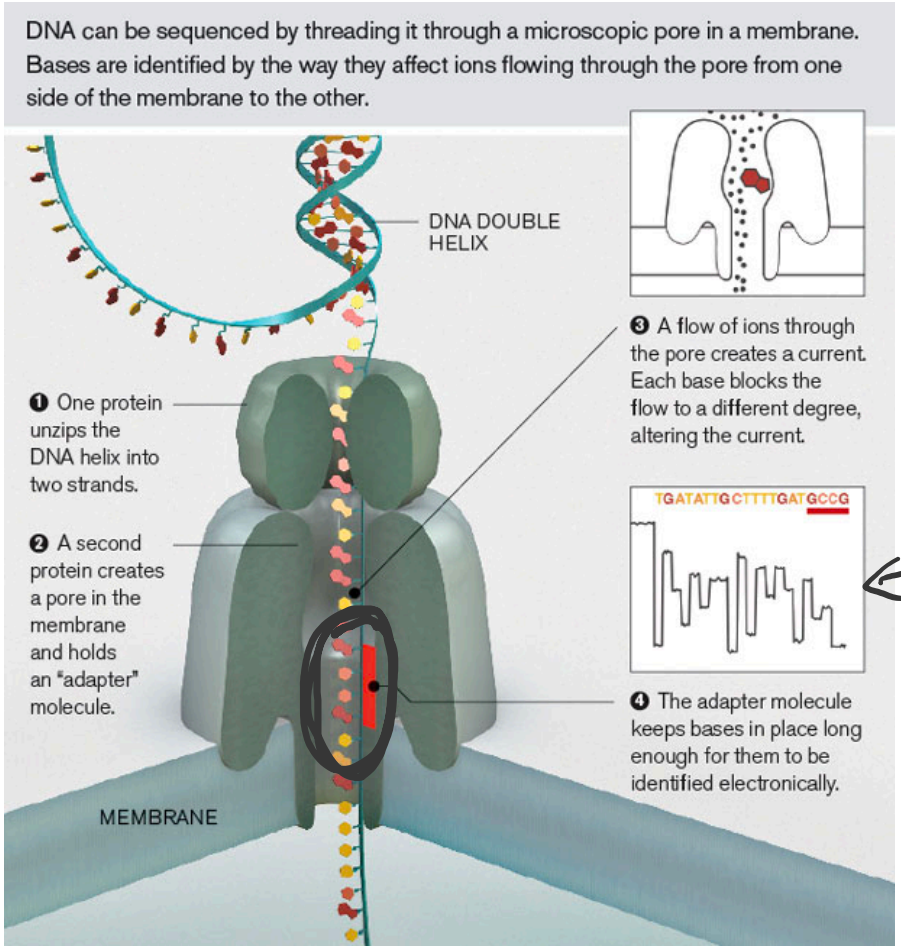
ML Metzker, Nature Review Genetics, 2009

Pacific Biosciences: SMS

- https://www.youtube.com/watch?v=NHCJ8PtYCFc&list=PL_VcB7OJ1TCAWRXN6vnC5IKbMHjIMtN8P&index=4

4th Generation Sequencing!!!

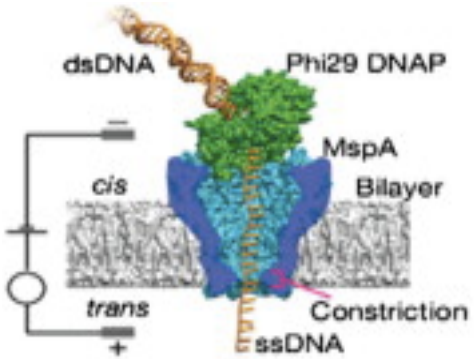
- NANOPORES!!!

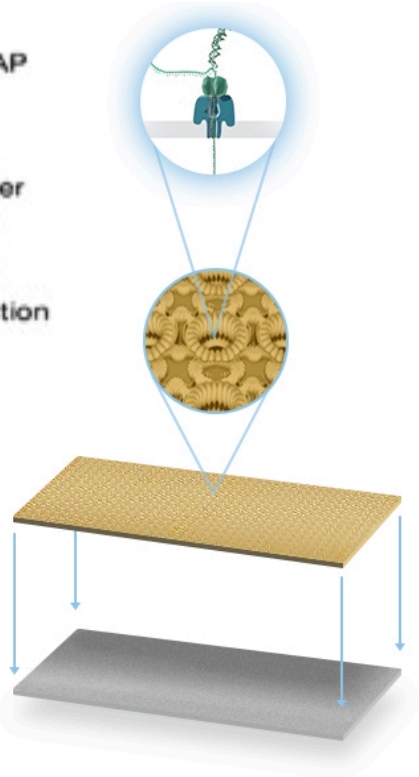


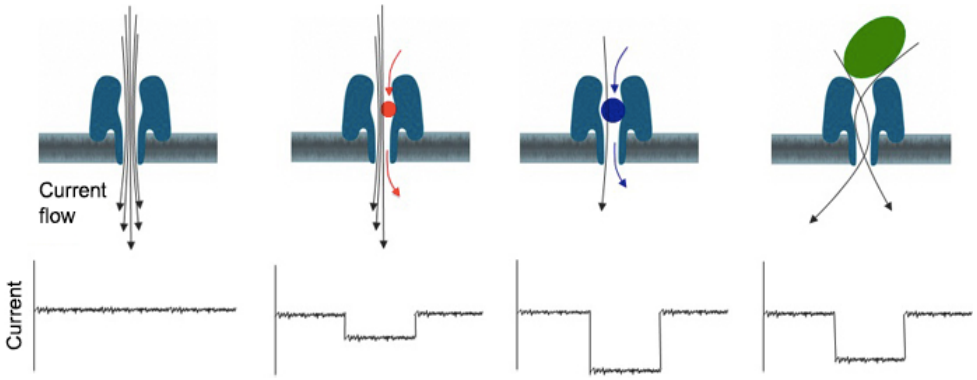
Schaffer, MIT Technology Review, 2012

Nanopore Sequencing

- Rated lifetime of one MinION flow cell: 48 hours of run time







Oxford Nanopore

- https://www.youtube.com/watch?v=CE4dW64x3Ts&index=5&list=PL_VcB7OJ1TCAWRXN6vnC5lKbMHjlMtN8P

5-10%

Summary of NGS technologies

Company (former companies)	Platforms	Library amplification	Carrier of library during sequencing	Sequencing principle	Nucleotide modifications	Signal detection method	Dominant sequencing error	Main advantages	Main disadvantages
Helicos Bioscience (defunct)	Heliscope	None	Flow cell	Sequencing by synthesis	Fluorescently modified nucleotides (cleavable)	High powered optical detection of single fluorescence molecules	Indels	No amplification thereby avoiding biases; higher tolerance of degraded samples	Long time (imaging is slow); high error rate; short read length; huge, $$ machine
Roche (454 until 2006)	454 Titanium 454 FLX+ 454 GS Junior Titanium	emPCR on microbeads	Picotiterplate	Pyrosequencing	None	Optical detection of light (luciferase reaction using PP_i released upon dNTP incorporation)	Indels in homopolymer runs	Longer reads than most other NGS platforms, relatively high fidelity	Shorter read than Sanger; lower output/yield – high price per base
Illumina (Solexa until 2007)	MiniSeq MiSeq NextSeq 500 HiSeq 2500 HiSeq 4000 HiSeq X five/ten	Bridge-PCR on flow cell surface	Flow cell	Reversible terminator sequencing by synthesis	End-blocked fluorescent nucleotides	Optical detection of fluorescence from incorporated nucleotides	Substitution, esp at end of reads	Good support; reasonable read lengths; low cost per read; flexibility in output/scalable; reasonable error rates	Generally higher instrument cost; bigger machines; long sample prep; amplification bias
Thermo Fisher Scientific (Agencourt until 2006, Applied Biosystems until 2008, Life Technologies until 2014;	SOLiD 5500 SOLiD 5500xl SOLiD 5500W	emPCR on microbeads	FlowChip	Sequencing by ligation	2-base encoded fluorescent oligoNTP	Optical detection of fluorescent emission from ligated dye-labeled oligoNTP	Substitution	High accuracy; High throughput of 20-30Gb/day	Relatively short reads; less even data distribution; High capital cost
Ion Torrent until 2010, Life Technologies until 2014	Ion Torrent PGM Ion Torrent Proton Ion Torrent S5/S5xl	emPCR on microbeads	Ion Chip (semiconductor based)	Semiconductor-based sequencing by synthesis	None	Transistor-based detection of H+ shift upon nucleotide incorporation	Indels	Generally moderate cost instrument; easy to use	More hands-on time; higher cost per Mb; small user community
Pacific Biosciences	PacBio RS II PacBio Sequel	None	SMRT cell (zero mode wave guides)	Single-molecule, real-time DNA sequencing by synthesis	Phosphor-linked fluorescent nucleotides	Real-time optical detection of fluorescent dye in polymerase active site during nucleotide incorporation	Indels	Single molecule real-time sequencing; Long read length; can detect base modifications; Short instrument run time; Random error profile; Modest cost per sample	High error rate; Low output; High cost per Mb; High instrument cost
Oxford Nanopore	minION PromethION (coming soon)	None	Flow cell	Single-molecule, real-time direct DNA sequencing	None	Semiconductor-based detection of changes in electron flow through nanopore protein; each base blocks electron flow through the nanopore differently as it passes through	Indels	Very small, low-cost, portable instrument (USB device); very long reads feasible (multiple kb); potentially very fast	High error rate; systematic errors; High cost per read

Latest in sequencing technology – Error Correction Code (ECC) Sequencing

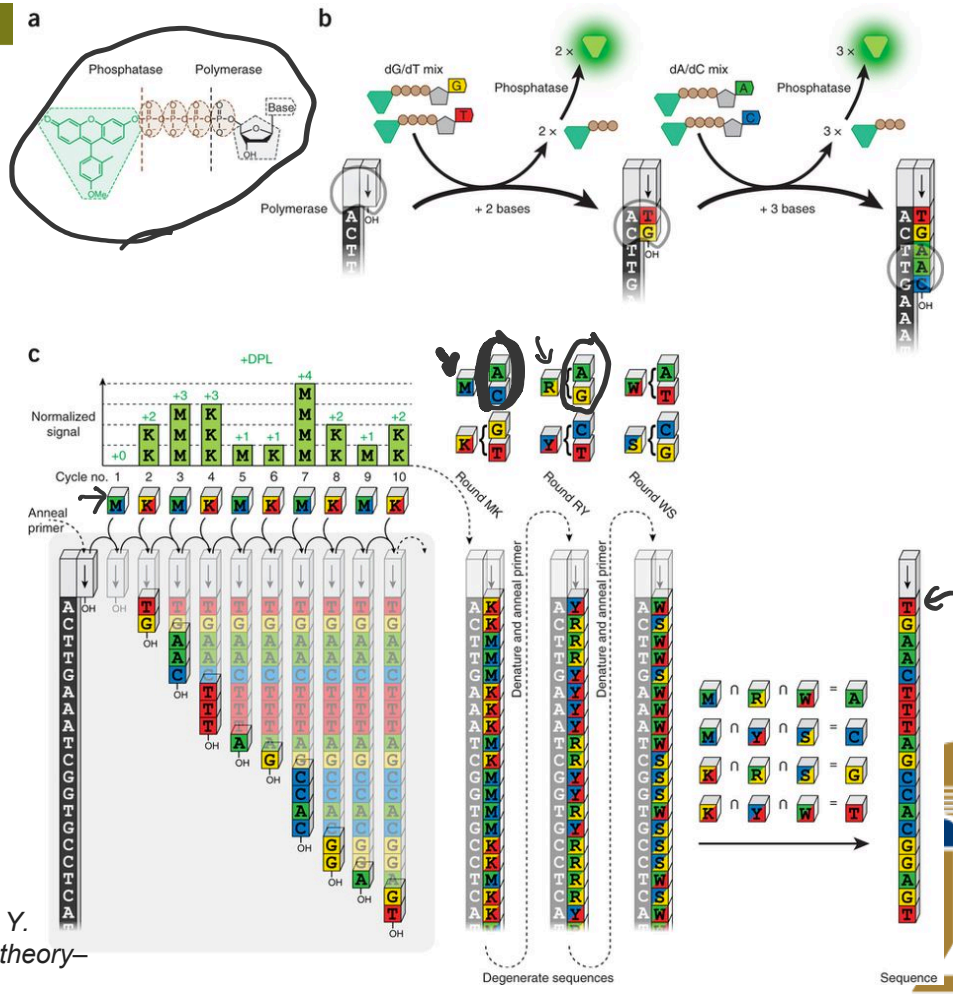
nature biotechnology

Article | Published: 06 November 2017

Highly accurate fluorogenic DNA sequencing with information theory–based error correction

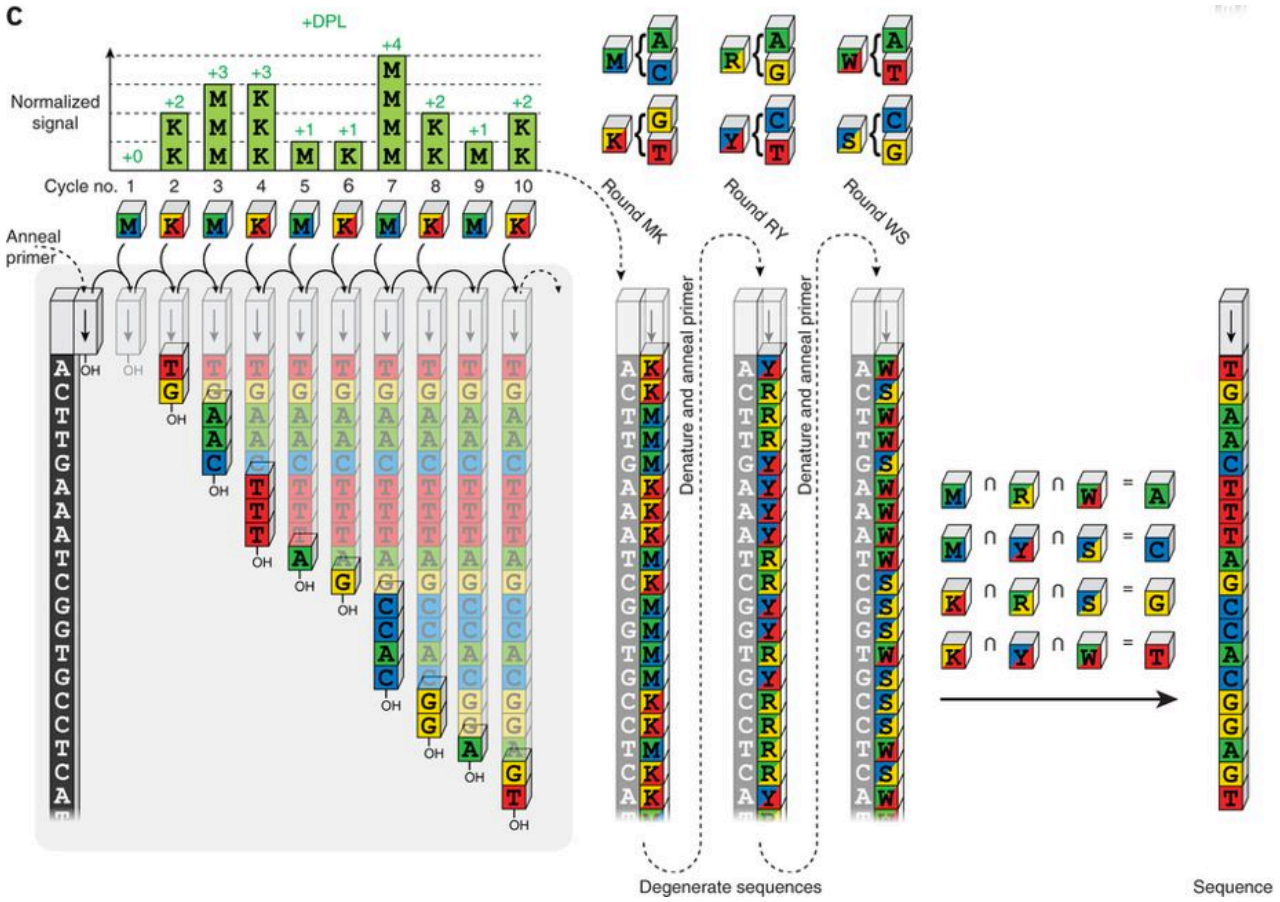
Zitian Chen, Wenxiong Zhou, Shuo Qiao, Li Kang, Haifeng Duan, X Sunney Xie & Yanyi Huang

Nature Biotechnology **35**, 1170–1178 (2017) | Download Citation



Chen, Z., Zhou, W., Qiao, S., Kang, L., Duan, H., Xie, X. S., & Huang, Y. (2017). Highly accurate fluorogenic DNA sequencing with information theory–based error correction. Nature biotechnology, 35(12), 1170.

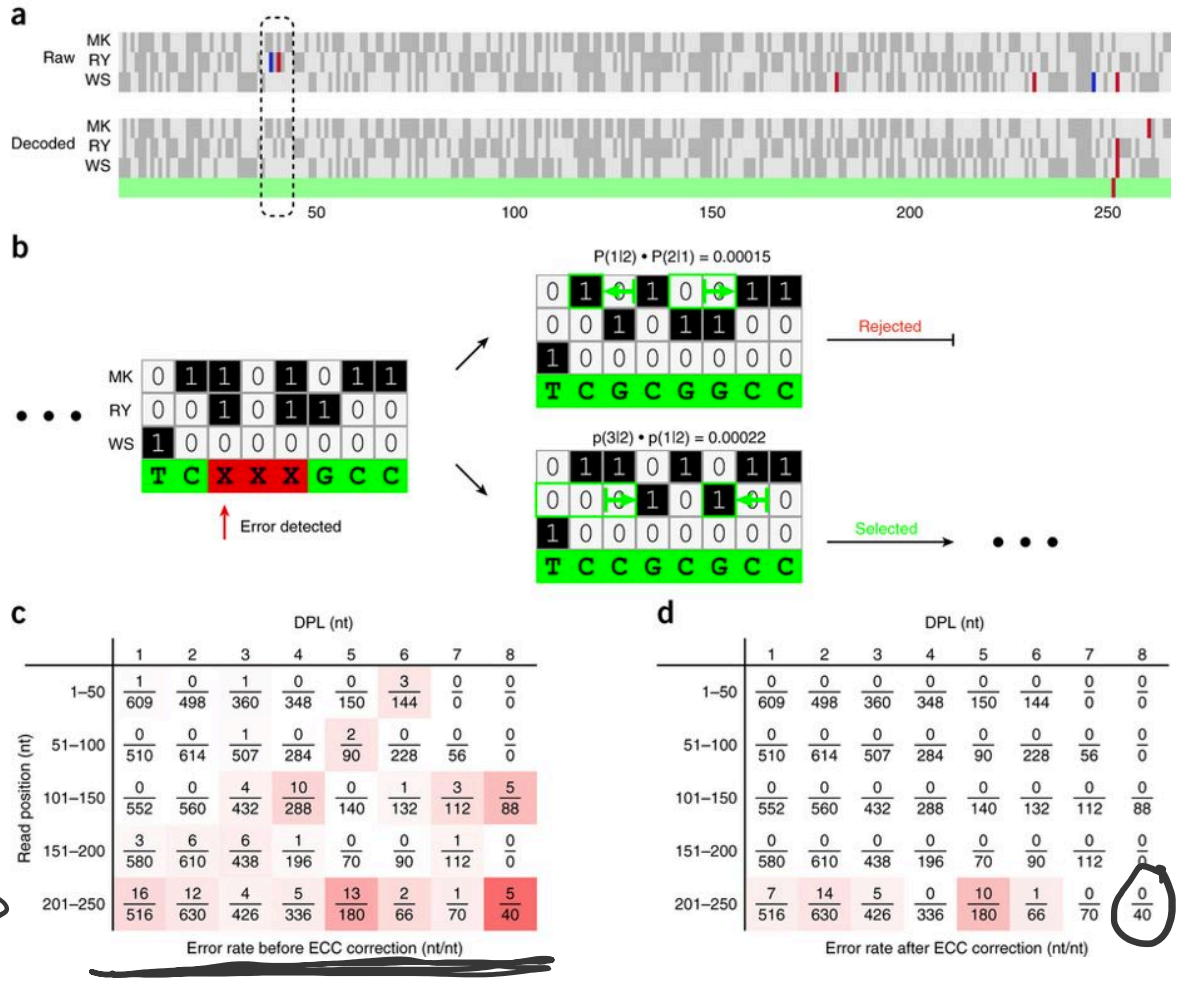
Latest in sequencing technology – Error Correction Code (ECC) Sequencing



Chen, Z., Zhou, W., Qiao, S., Kang, L., Duan, H., Xie, X. S., & Huang, Y. (2017). Highly accurate fluorogenic DNA sequencing with information theory–based error correction. Nature biotechnology, 35(12), 1170.

Latest in sequencing technology – Error Correction Code (ECC) Sequencing

ECC sequencing – error free up to 250 bp!



Chen, Z., Zhou, W., Qiao, S., Kang, L., Duan, H., Xie, X. S., & Huang, Y. (2017). Highly accurate fluorogenic DNA sequencing with information theory–based error correction. Nature biotechnology, 35(12), 1170.

Project	Which sequencing platform(s) would you propose?	Why?
De novo sequencing and assembly of a microbial genome		
Sequencing a plasmid		
Re-sequencing a human genome (e.g. cancer sample) to look for novel mutations		
Rapid diagnosis of a viral infection by sequencing in the field		
Targeted amplicon/exome sequencing		

Applications of NGS

- Your ideas

Applications of NGS

- Cancer research
- Pre-natal diagnostics
- Discovery of new microbial or viral species
- Predicting organ transplant rejection

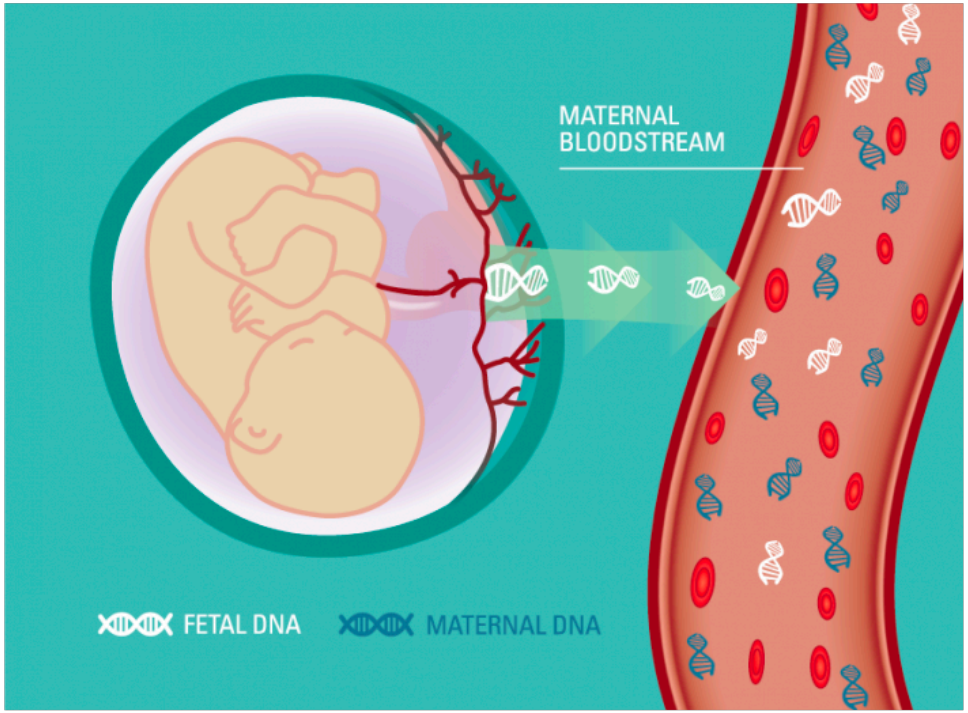
Applications of NGS

- Cancer research and diagnosis
  - Personal cancer genomes
  - RNA-seq comparing normal tissue to cancer tissue

- E.g. Breast Cancer types
  - ER+ or PR+ (drug tamoxifen to block hormone receptors)
  - HER2+ (drug herceptin)
  - Triple positive
  - Triple negative (often BRCA1+; chemo, high chance of relapse)

Applications of NGS

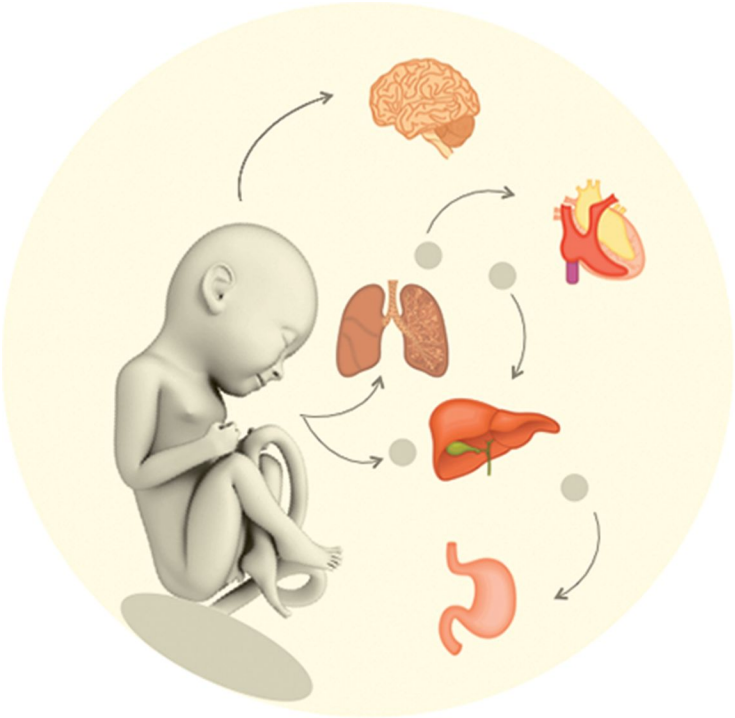
- Pre-natal diagnostics
  - DNA
  - RNA



From Ariosa website

Applications of NGS

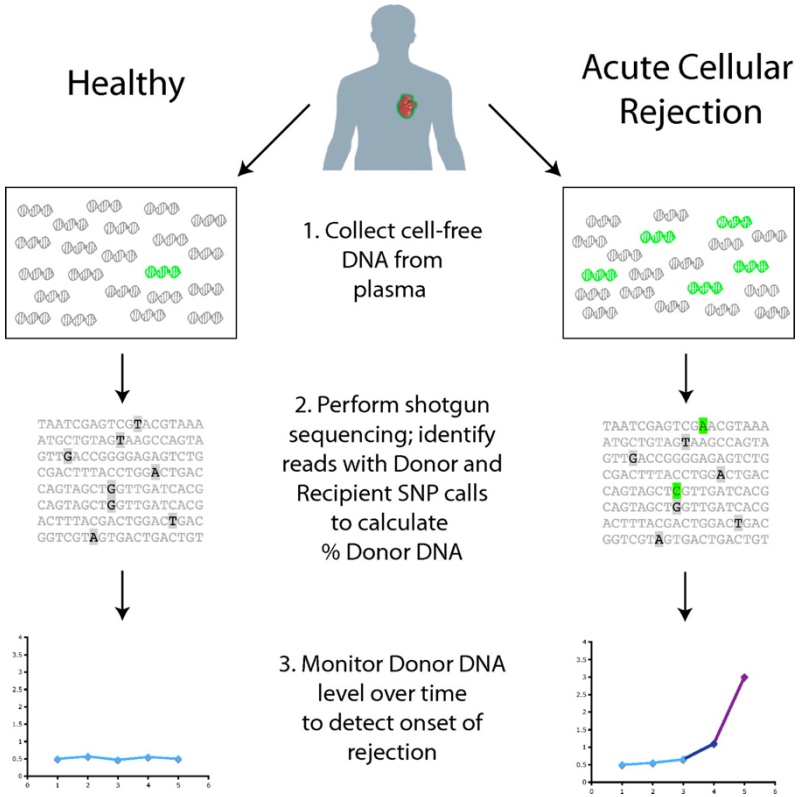
- Pre-natal diagnostics
  - DNA
  - RNA



Lillian M. Zwemer, and Diana W. Bianchi Cold Spring Harb Perspect Med 2015;5:a023101

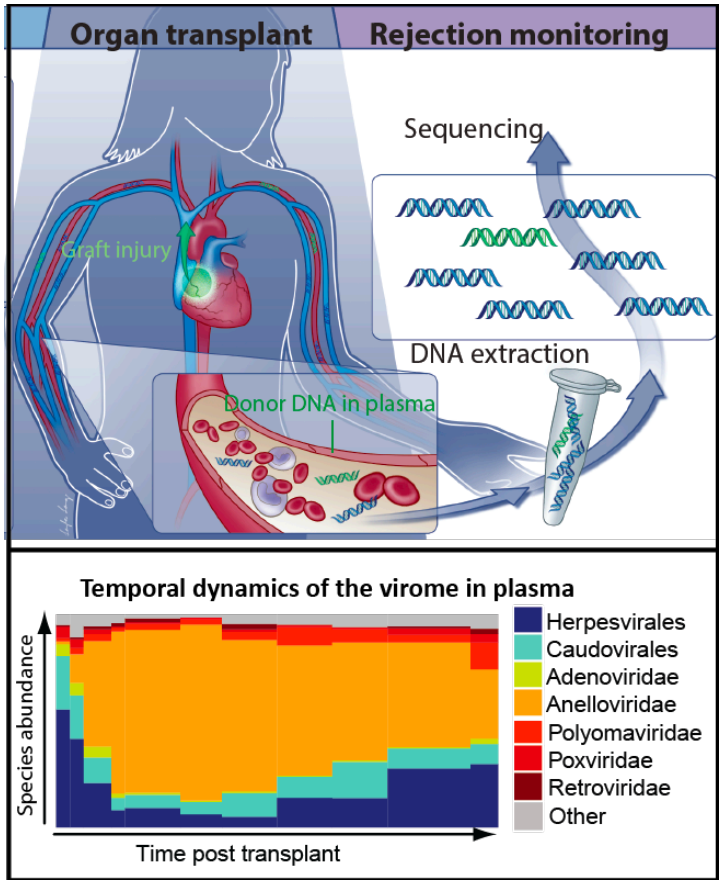
Applications of NGS

- Predicting organ transplant rejection
  - DNA of donor
  - RNA of microbes and viruses



Applications of NGS

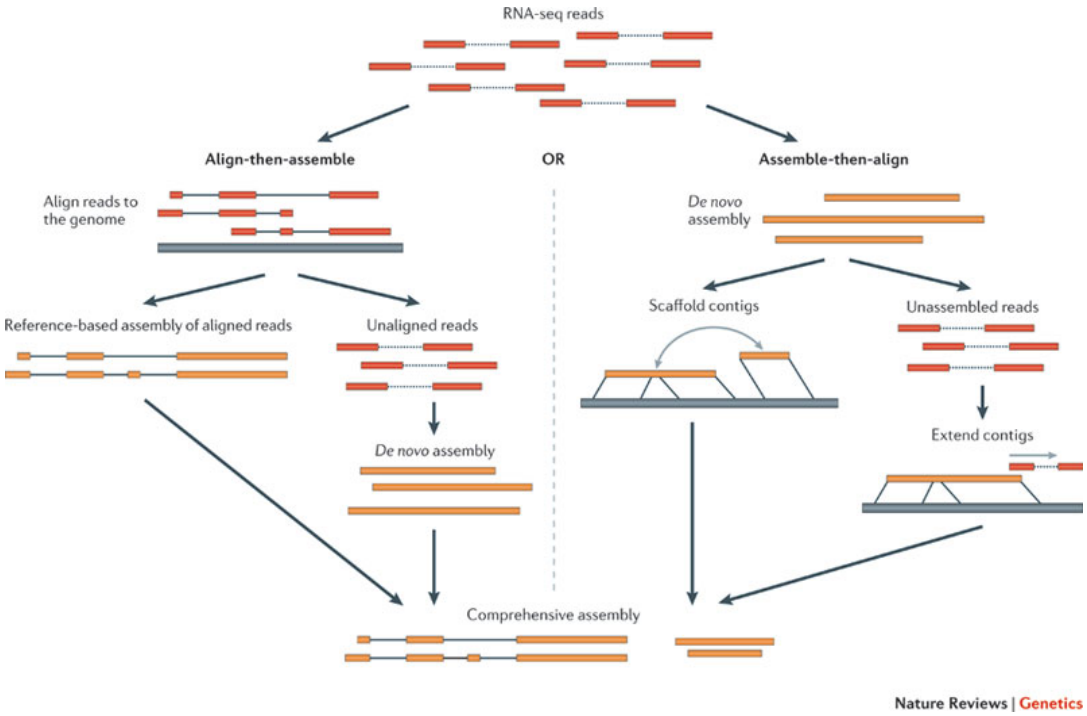
- Predicting organ transplant rejection
  - DNA of donor
  - RNA of microbes and viruses

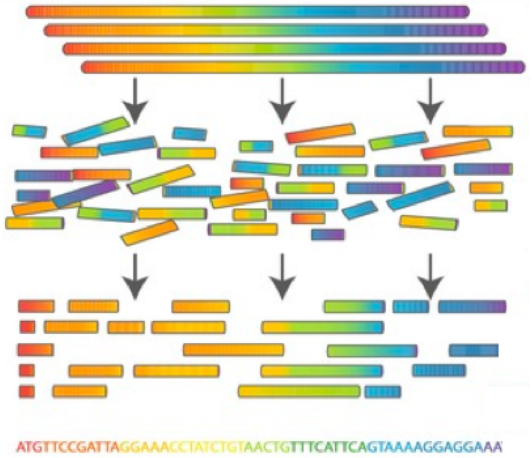


De Vlaminck et al., Science Translational Medicine, 2014

Applications of NGS

- Discovery of new microbial or viral species
  - De novo assembly



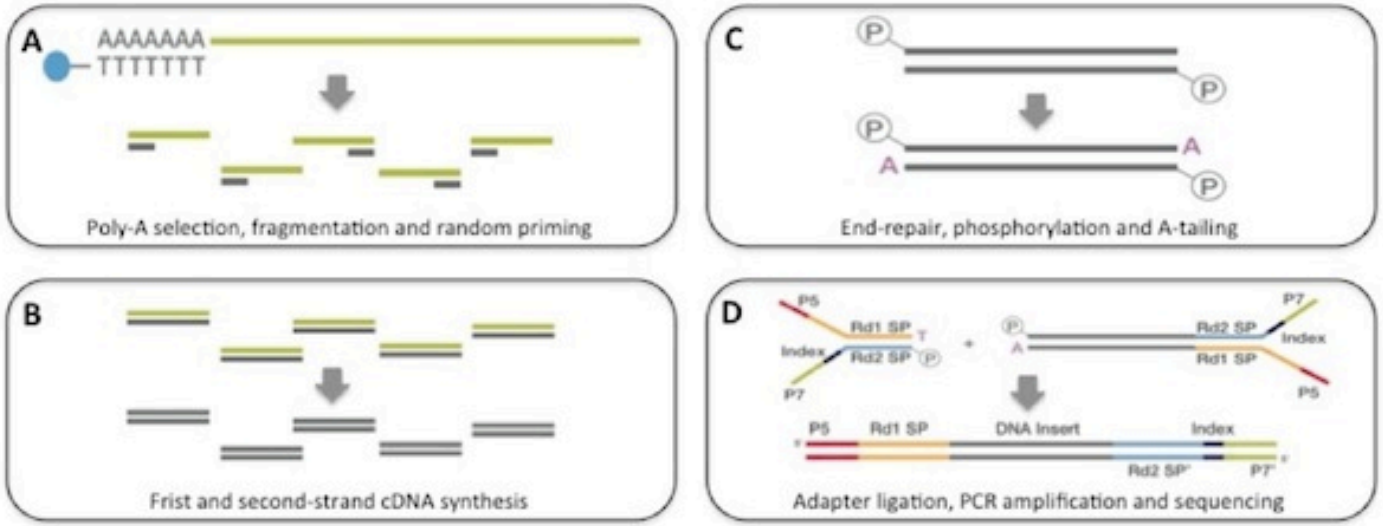


Applications of NGS

- Tools in the research lab: WGS, WES, RNA-seq, ChIP-seq, CHIRP-seq, methyl-seq, Hi-C, PRO-seq, ATAC-seq… etc.

RNA-seq

Illumina Tru-Seq RNA-seq protocol

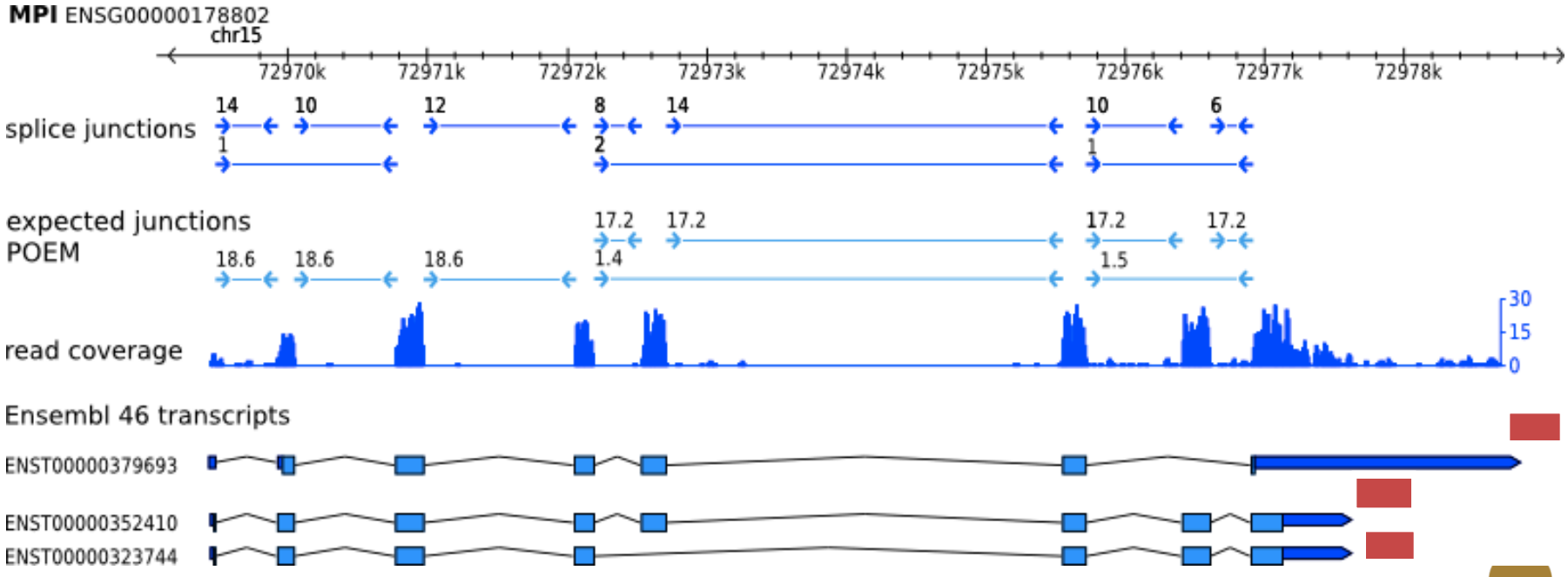


Library prep begins from 100ng-1ug of Total RNA which is poly-A selected (A) with magnetic beads. Double-stranded cDNA (B) is phosphorylated and A-tailed (C) ready for adapter ligation. The library is PCR amplified (D) ready for clustering and sequencing.

Image from BiteSize Bio

RNA-seq

Example of data

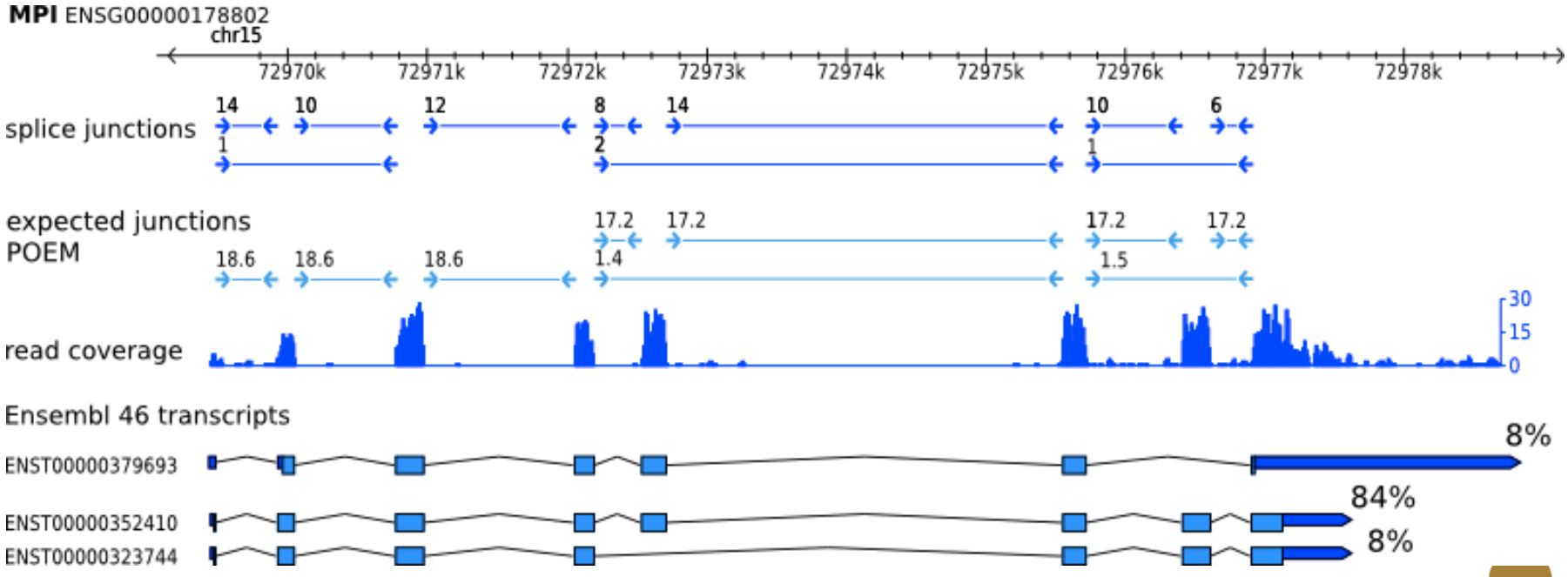


Which of the three transcripts is expressed with highest abundance?

Schulz MH, et al., Bioinformatics, 2012

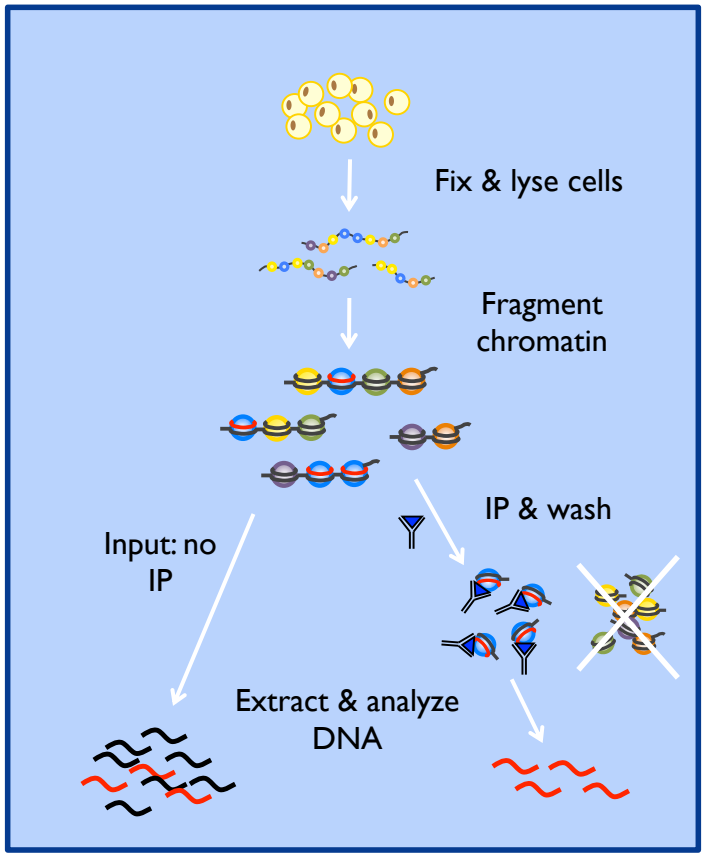
RNA-seq

Example of data

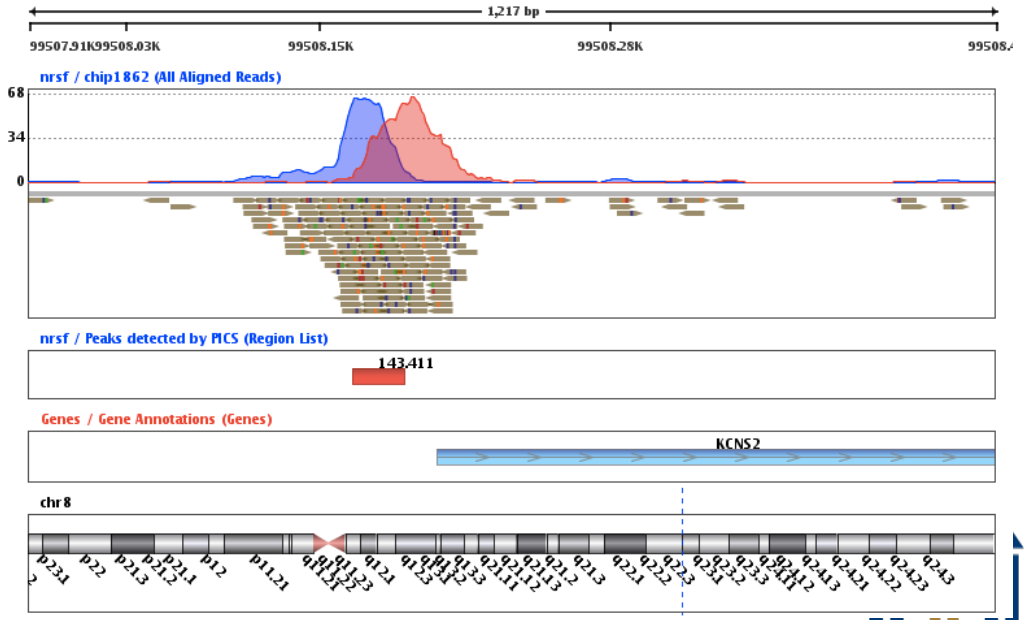


Schulz MH, et al., Bioinformatics, 2012

ChIP-seq



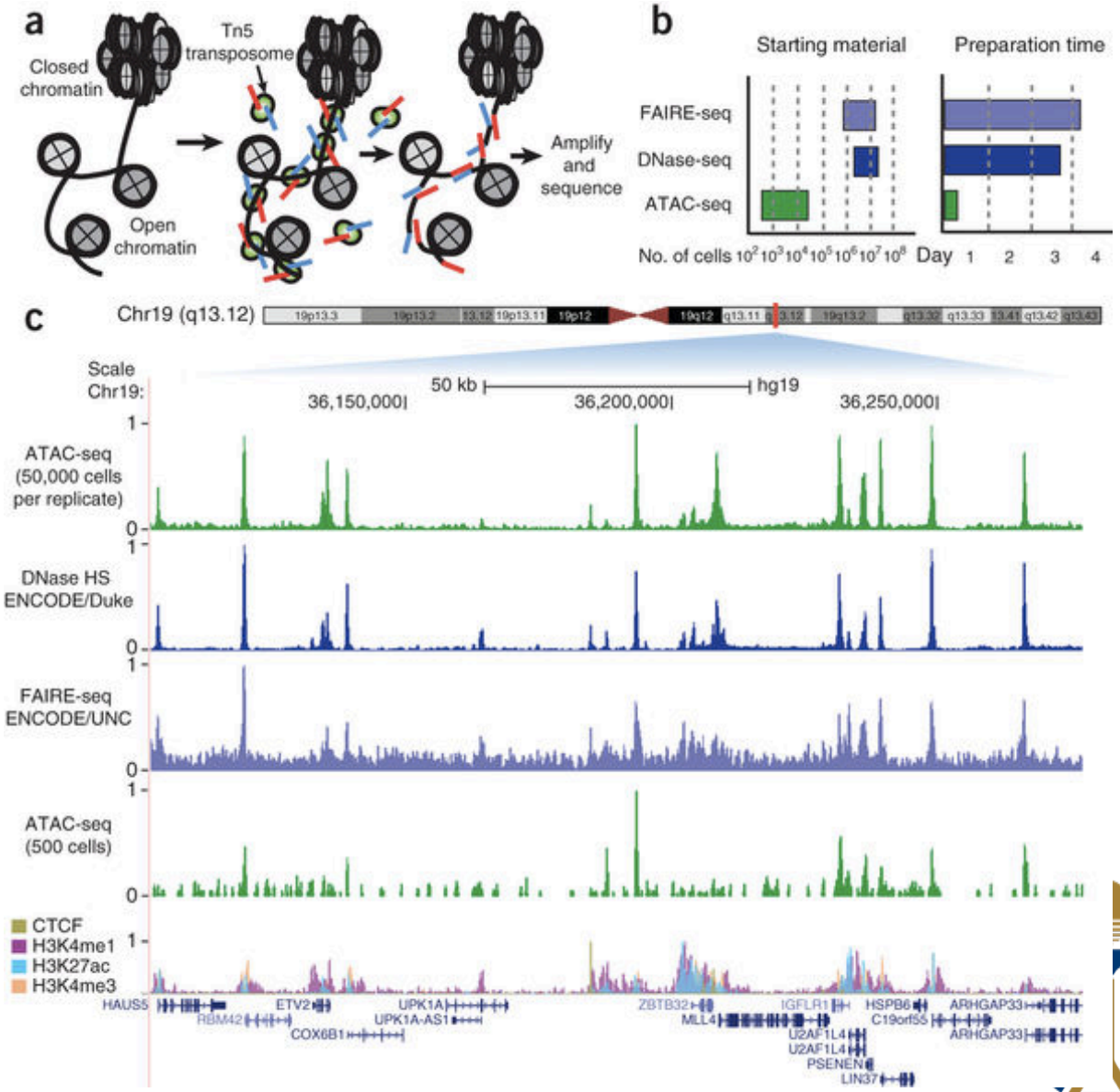
- ChIP: assesses protein-DNA interactions



ATAC-seq

Interrogates chromatin accessibility

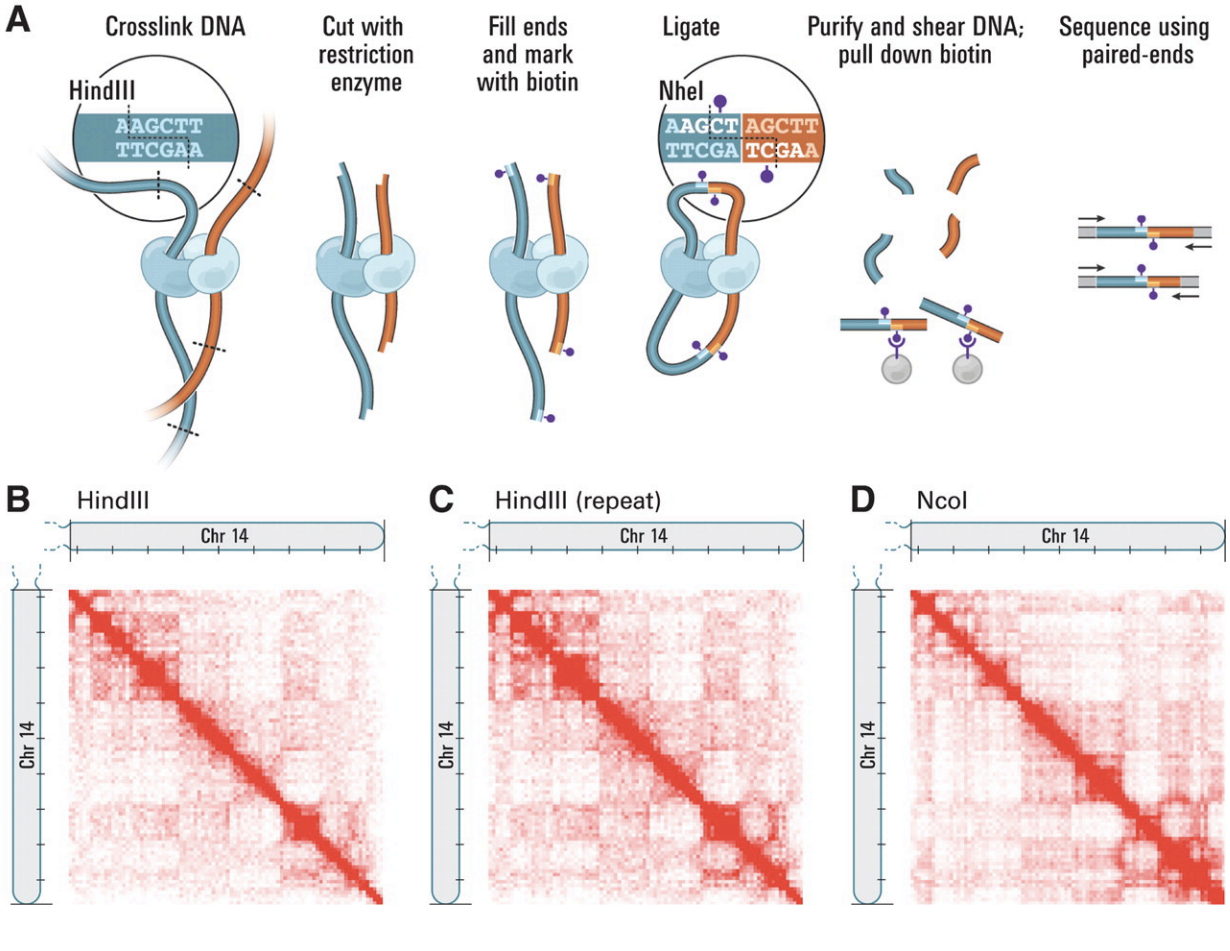
Easy to perform (compared to FAIRE-seq, MNase-seq, DNAse-seq etc.)

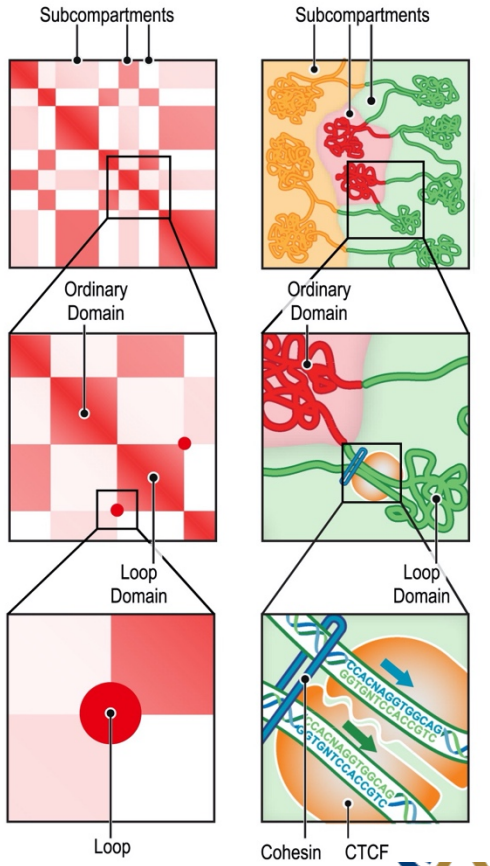


Buenrostro J.D., et al., Nature Methods, 2013

Hi-C

Probes 3D conformation of the genome architecture





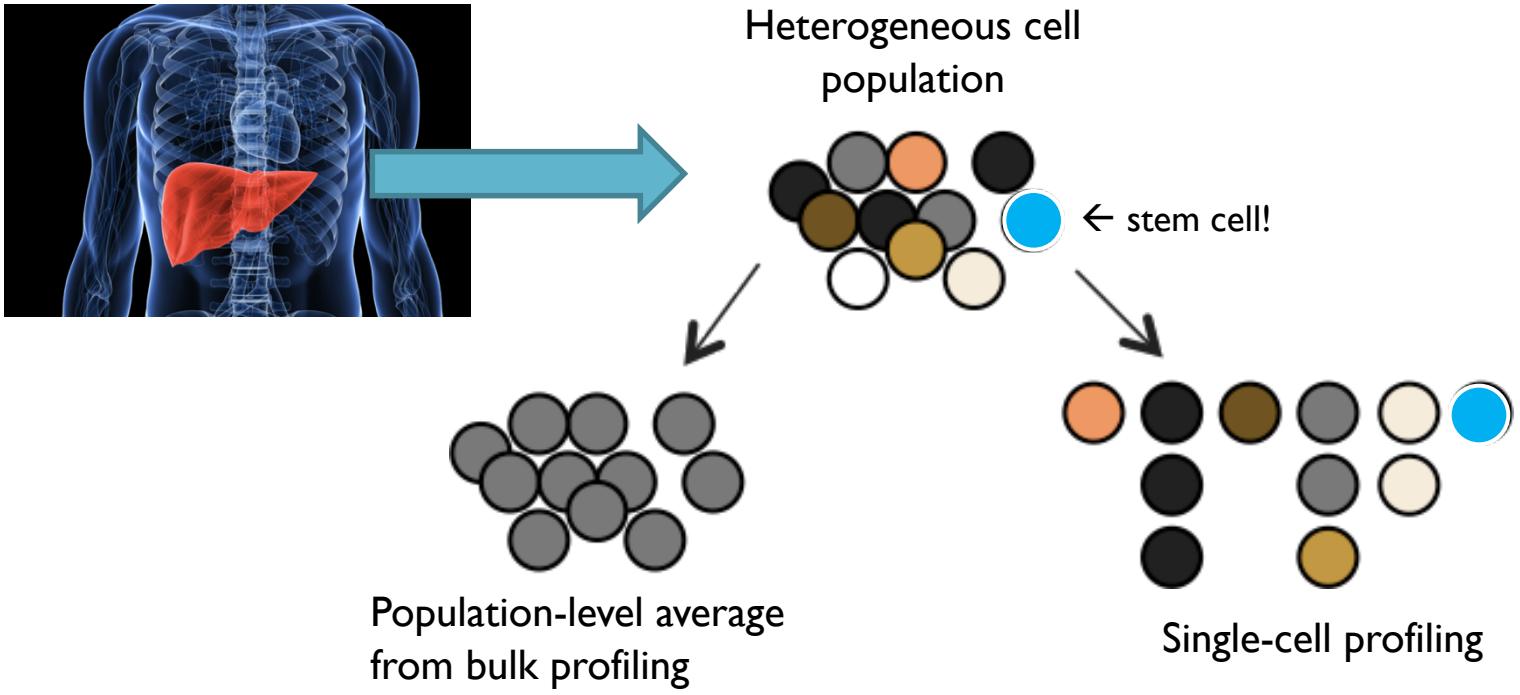
Rao et al., Cell, 2014

SINGLE-CELL OMICS

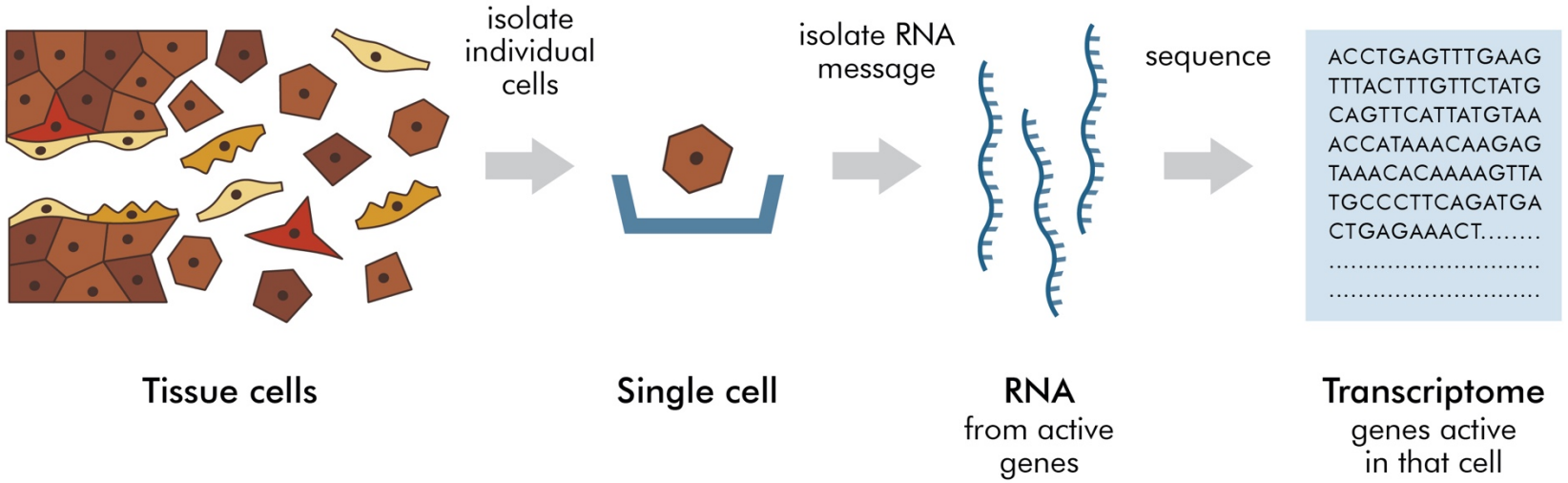
One field (among many) that is greatly enabled by microfluidic technologies

Why single cell?

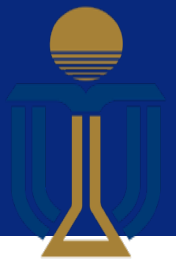
- Tissues consist of ***heterogeneous cell types***
- Method can be used for ***rare/valuable cell types***
  - e.g. circulating tumor cells; primary embryonic tissues



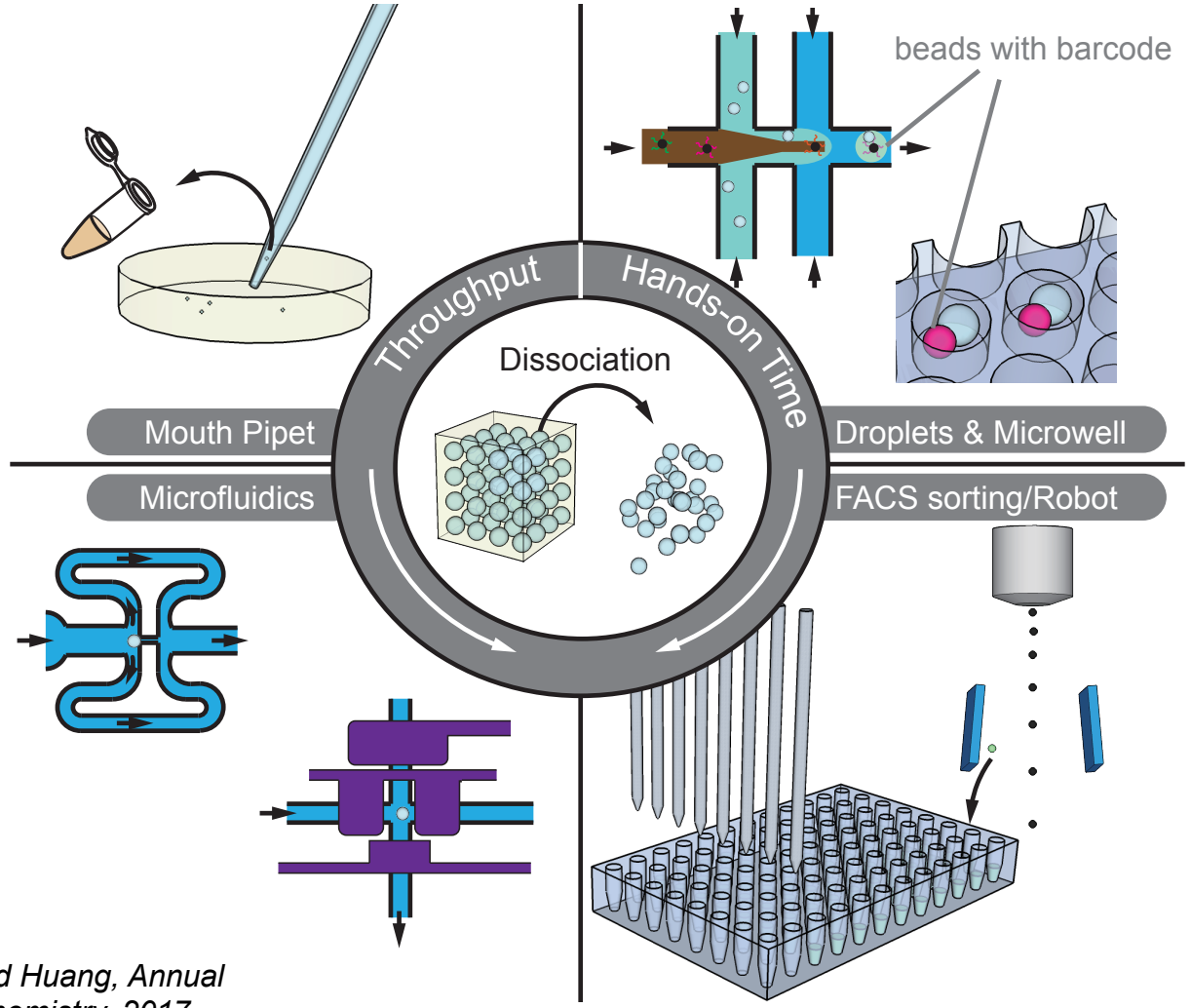
Single-cell genomics workflow



SINGLE CELL RNA-SEQ

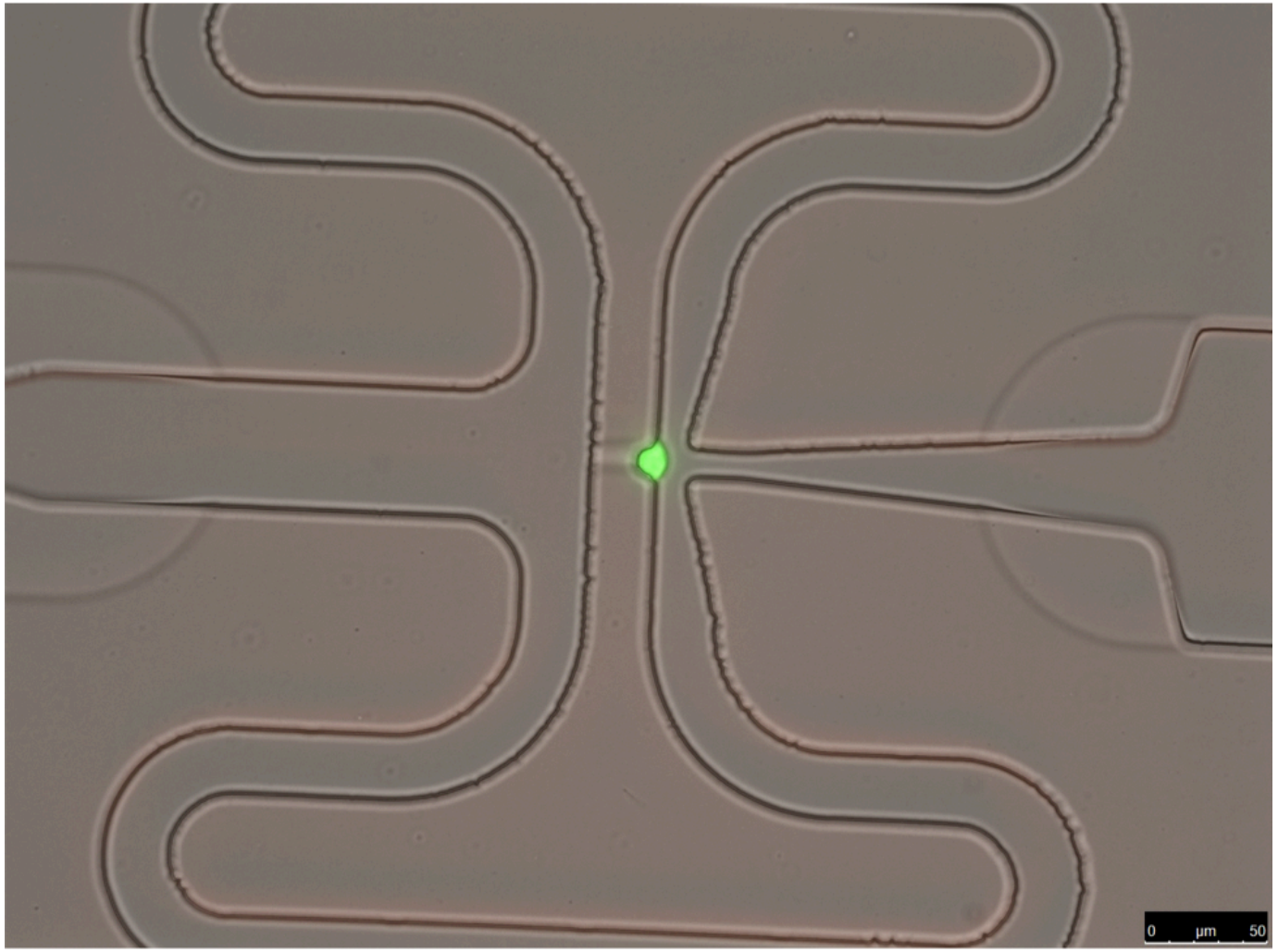


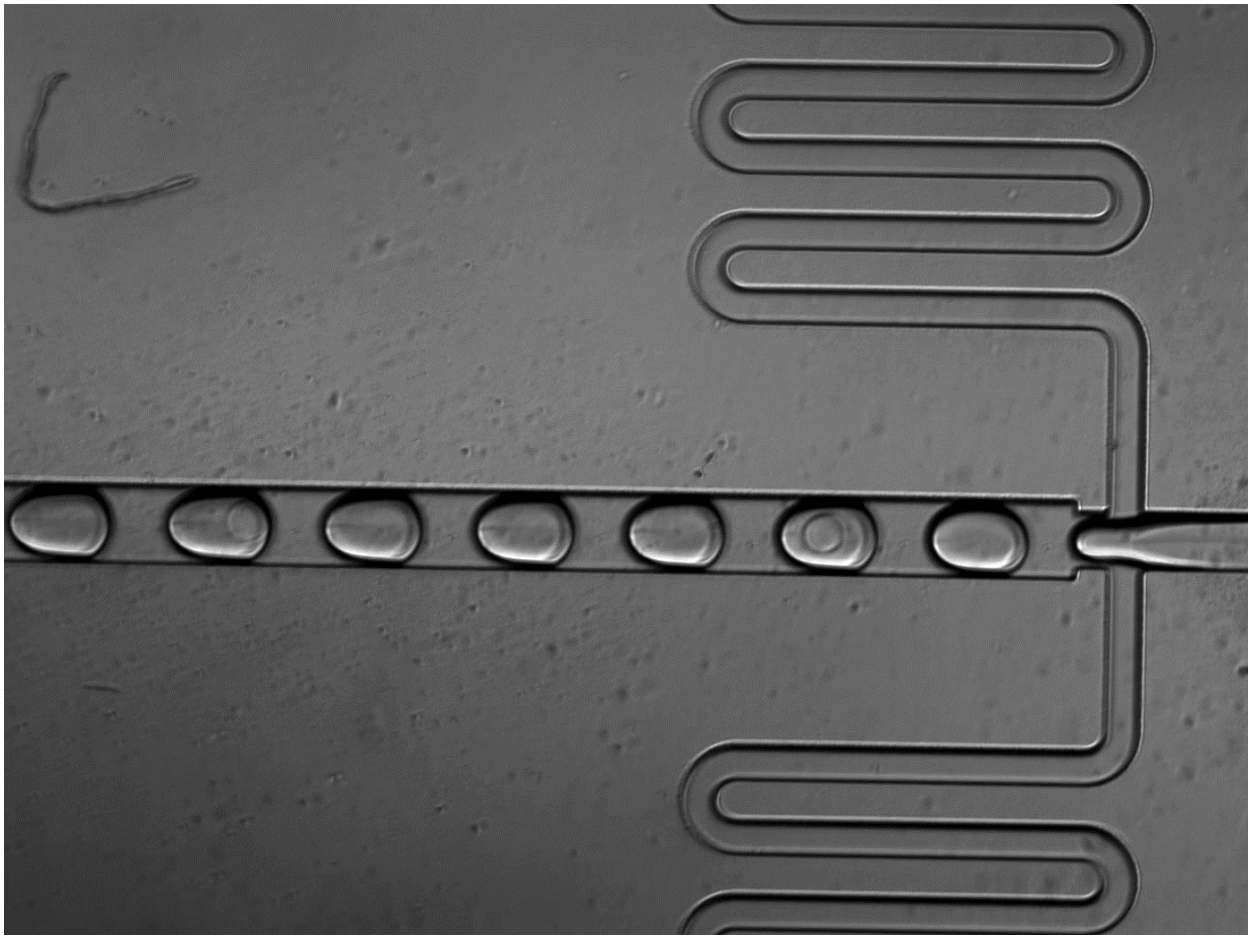
Many technology platforms to choose from



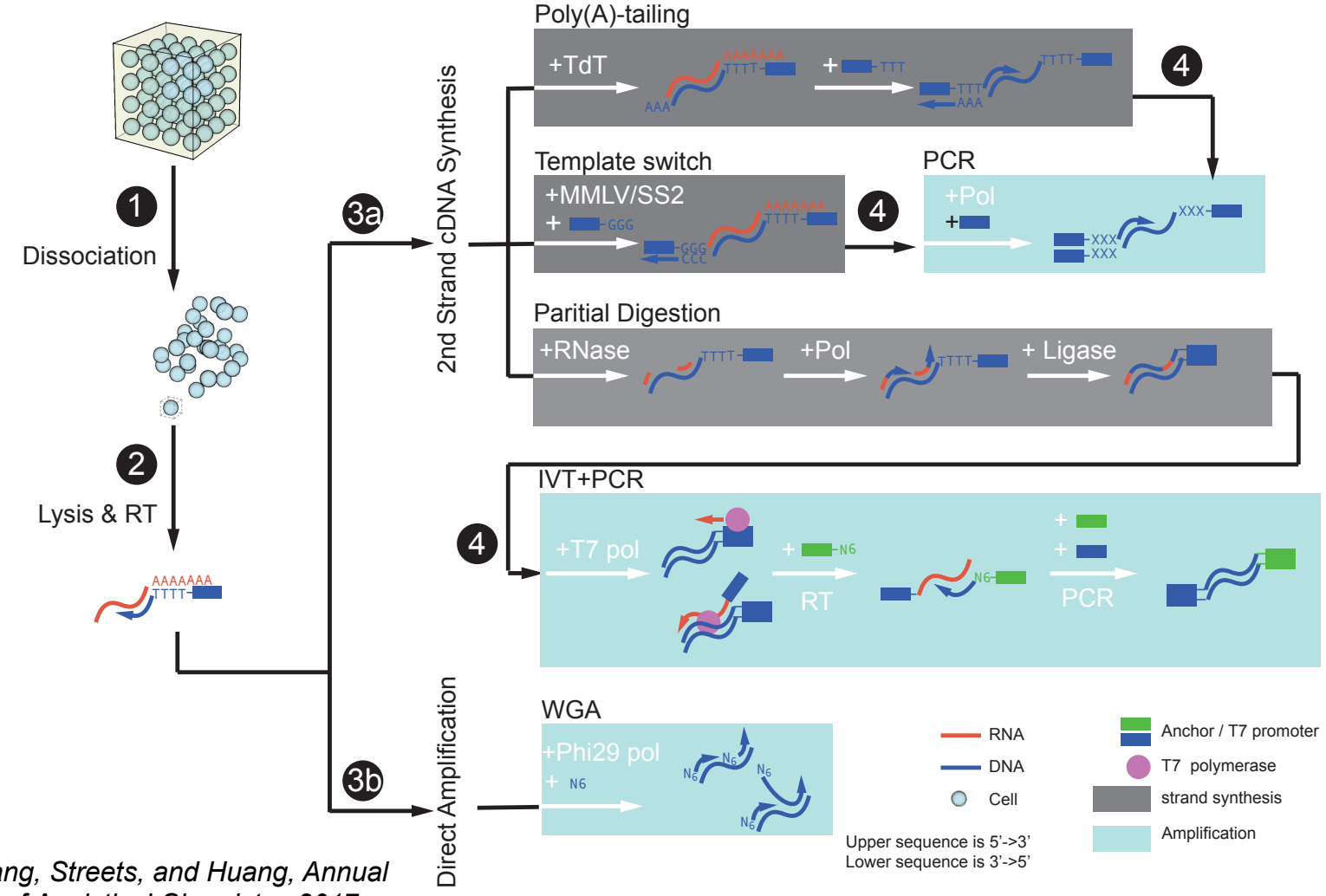
Wu, Wang, Streets, and Huang, Annual Review of Analytical Chemistry, 2017

Captured cells in the C1



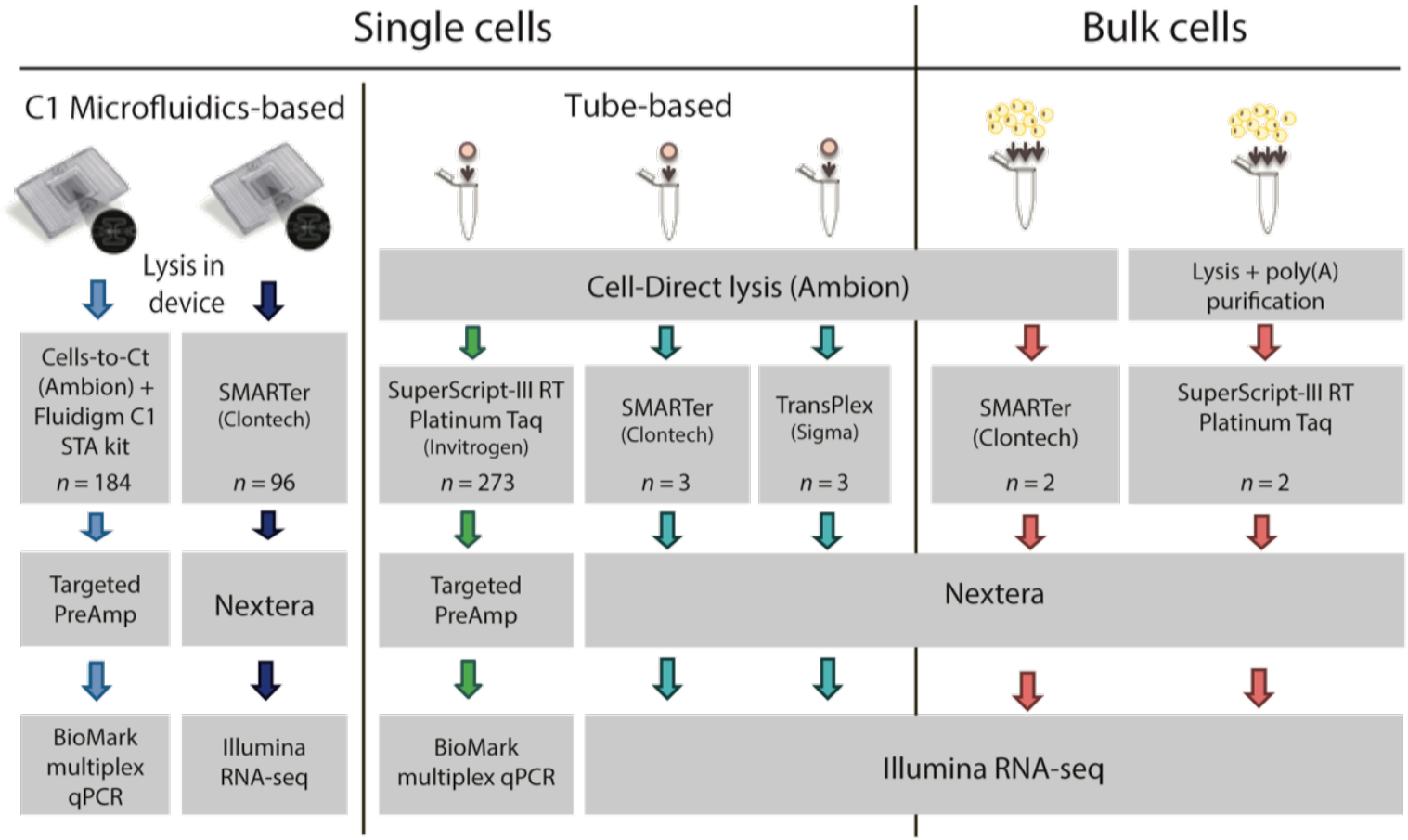


Single cell transcriptomics methodology



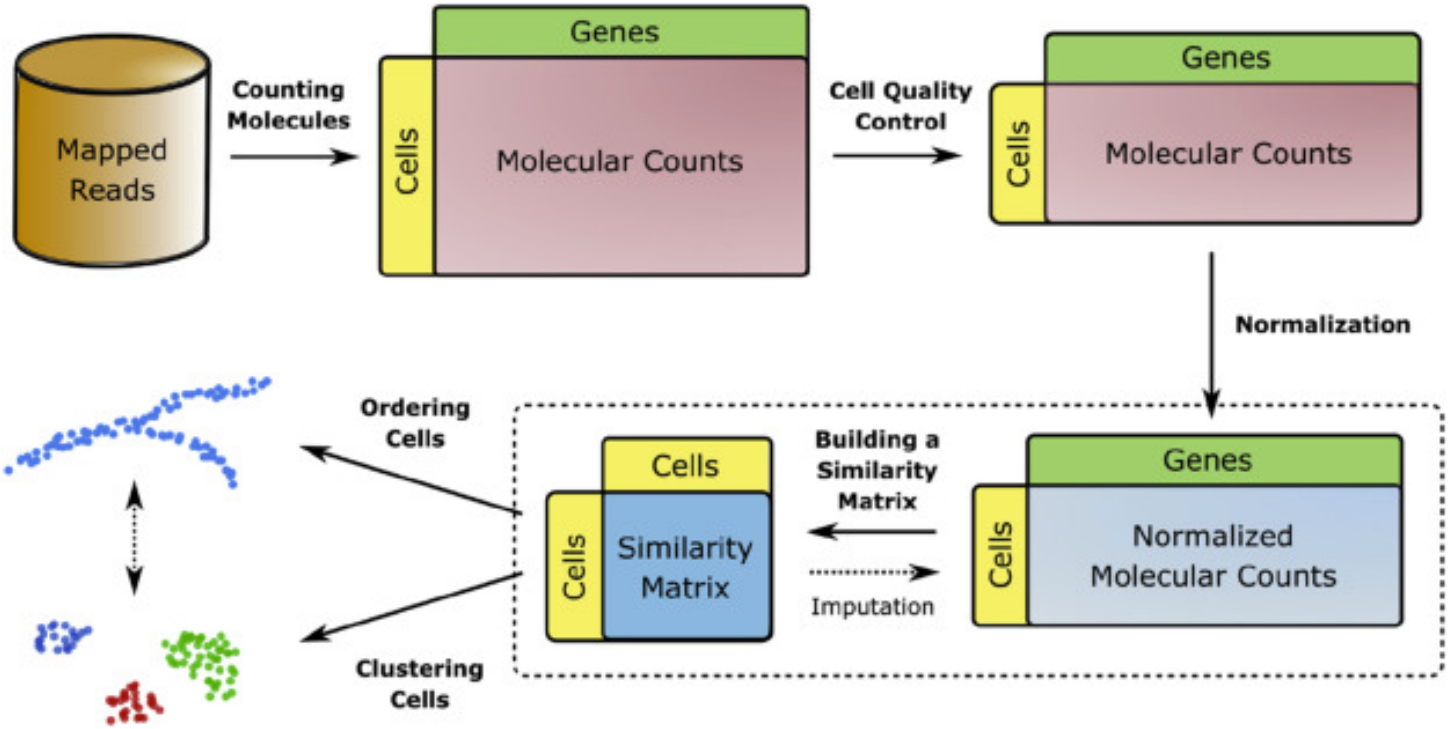
Wu, Wang, Streets, and Huang, Annual Review of Analytical Chemistry, 2017

Benchmarking single-cell RNA-seq vs. other gene expression measurements



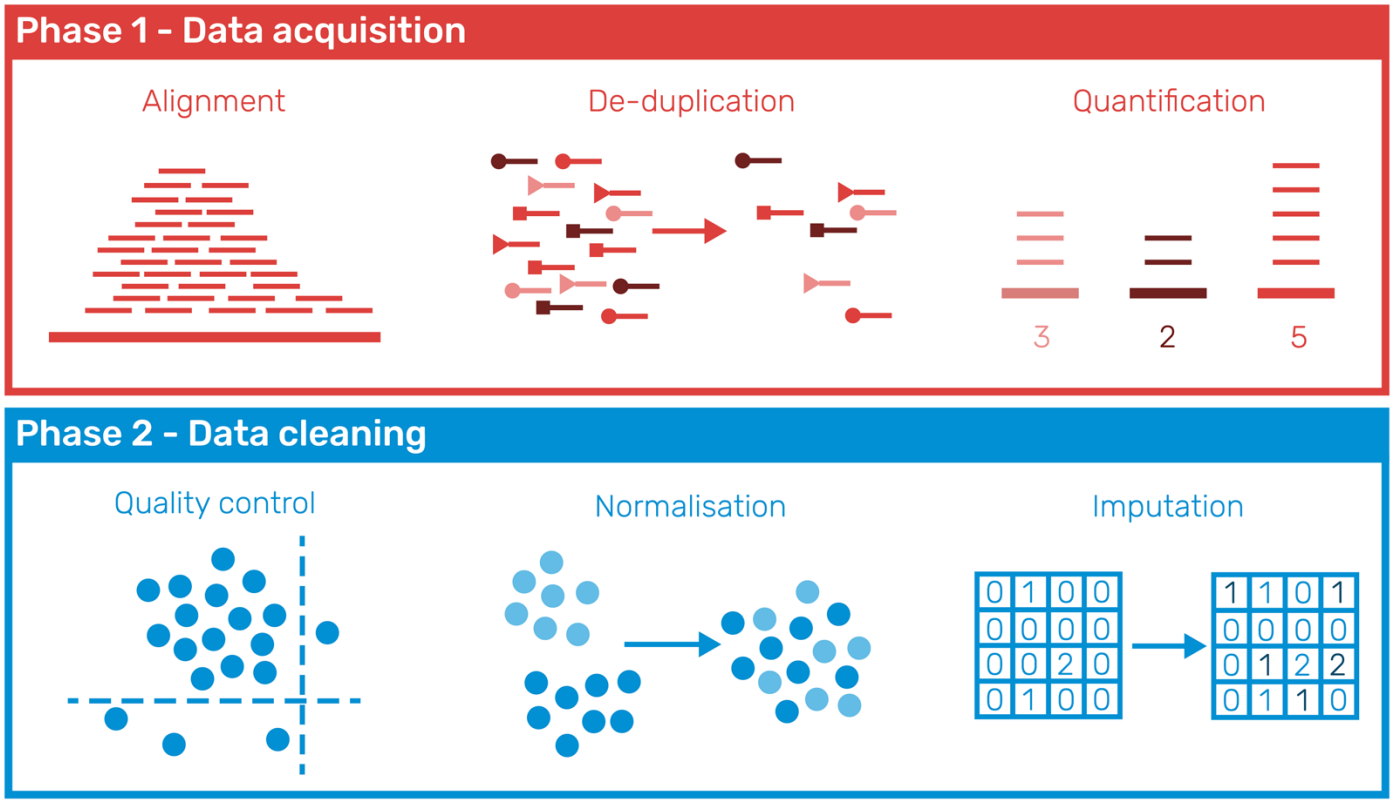
Wu, et al., 2014, Nature Methods

What does the data look like?



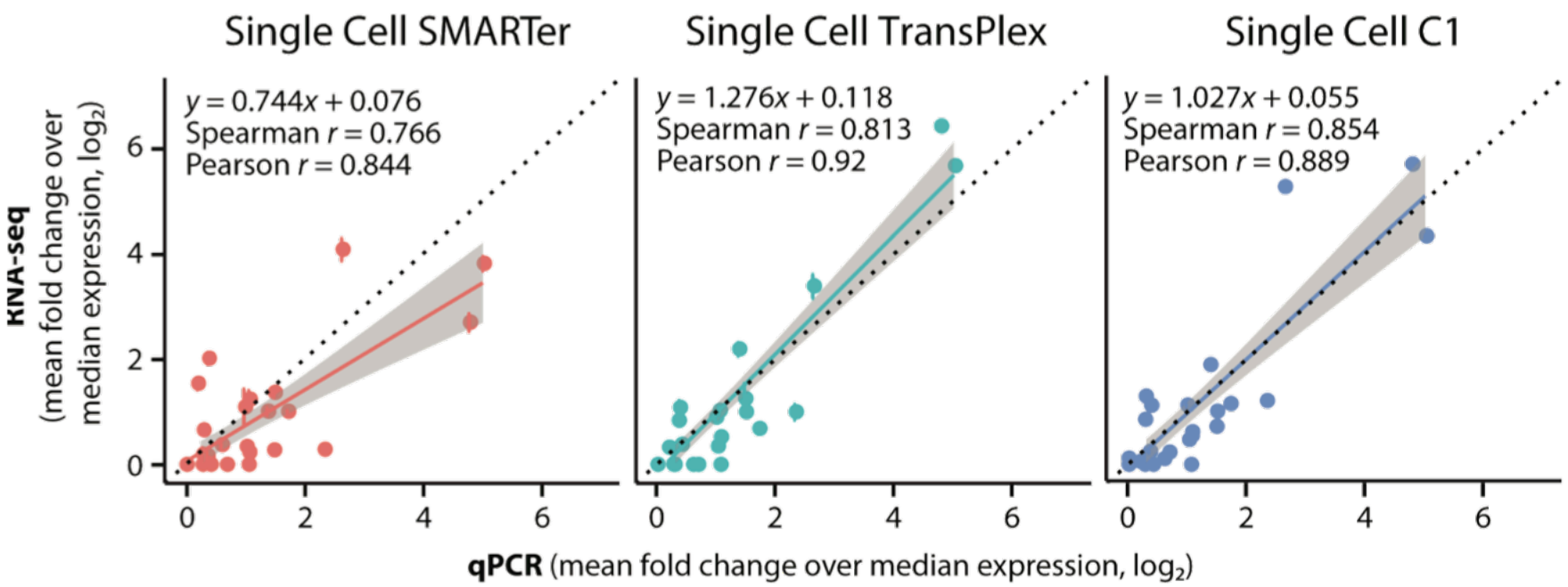
Camara P. G.., 2018, Current Opinions in Systems Biology

Data workflow for single-cell RNA-seq



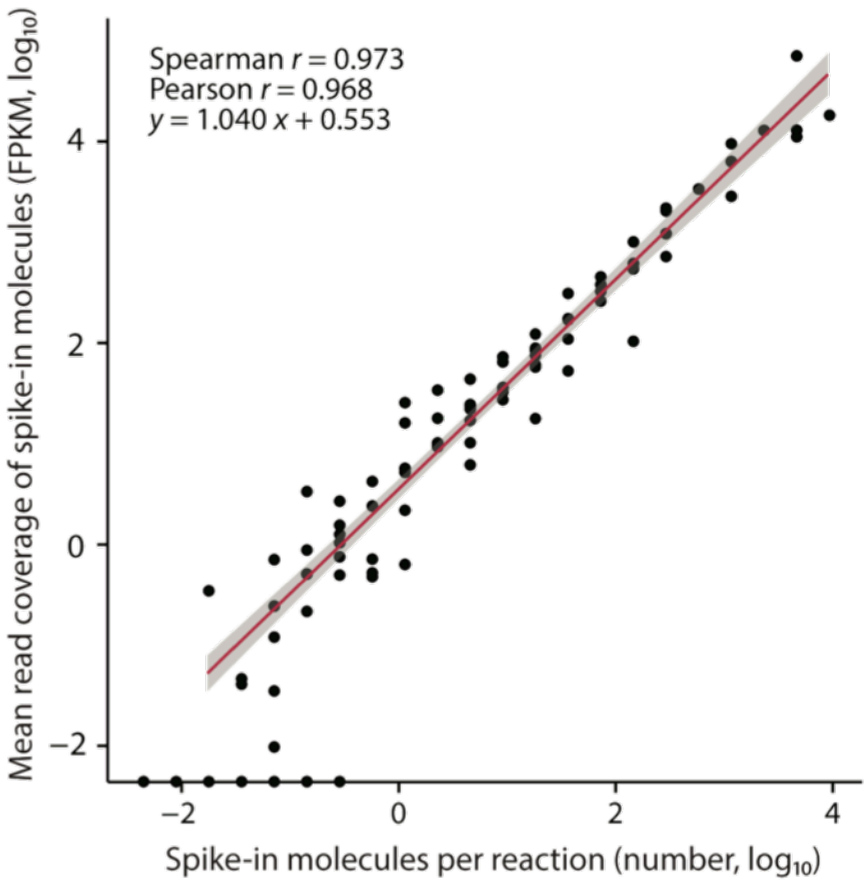
Zappia L., Phipson B., Oshlack A., 2018, PLoS Computational Biology

Accuracy of single cell RNA-seq



Wu, et al., 2014, Nature Methods

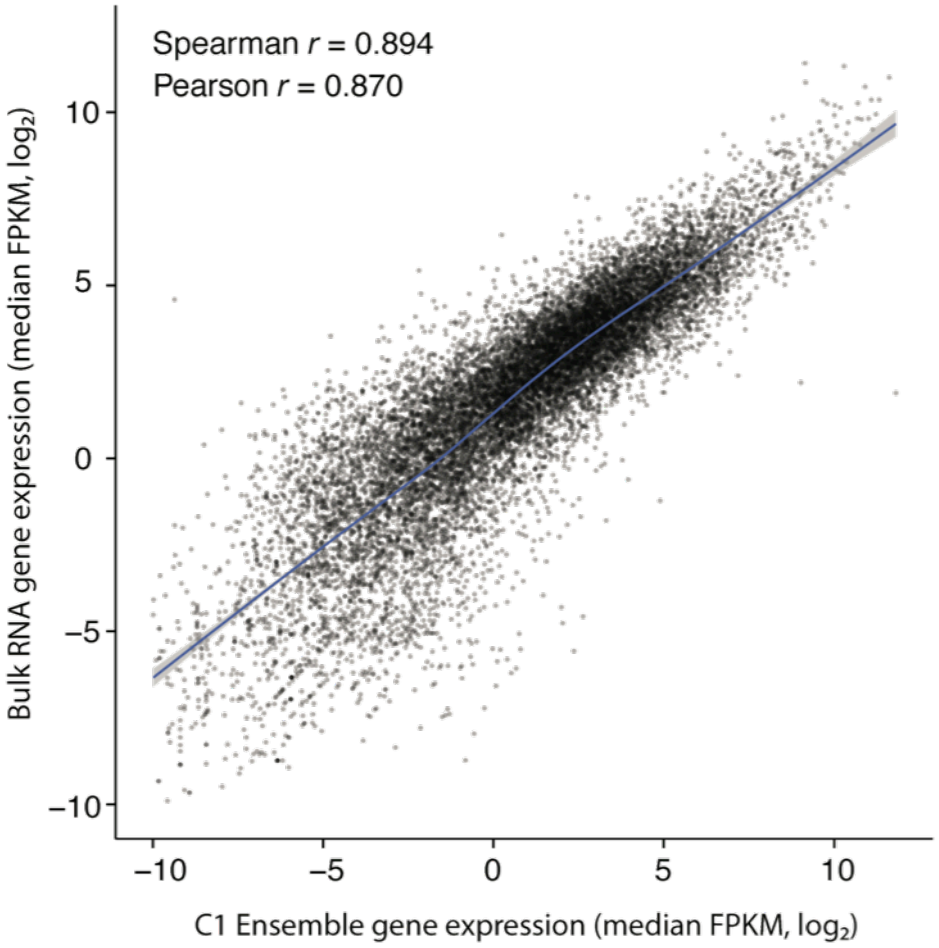
Limit of detection



- Spike-in synthetic sequences with various length, sequence content, concentration. Low homology with mammalian genomes
- **Limit of detection: ~1 molecule per reaction chamber**
- **Detection rate at this conc ~0.4**

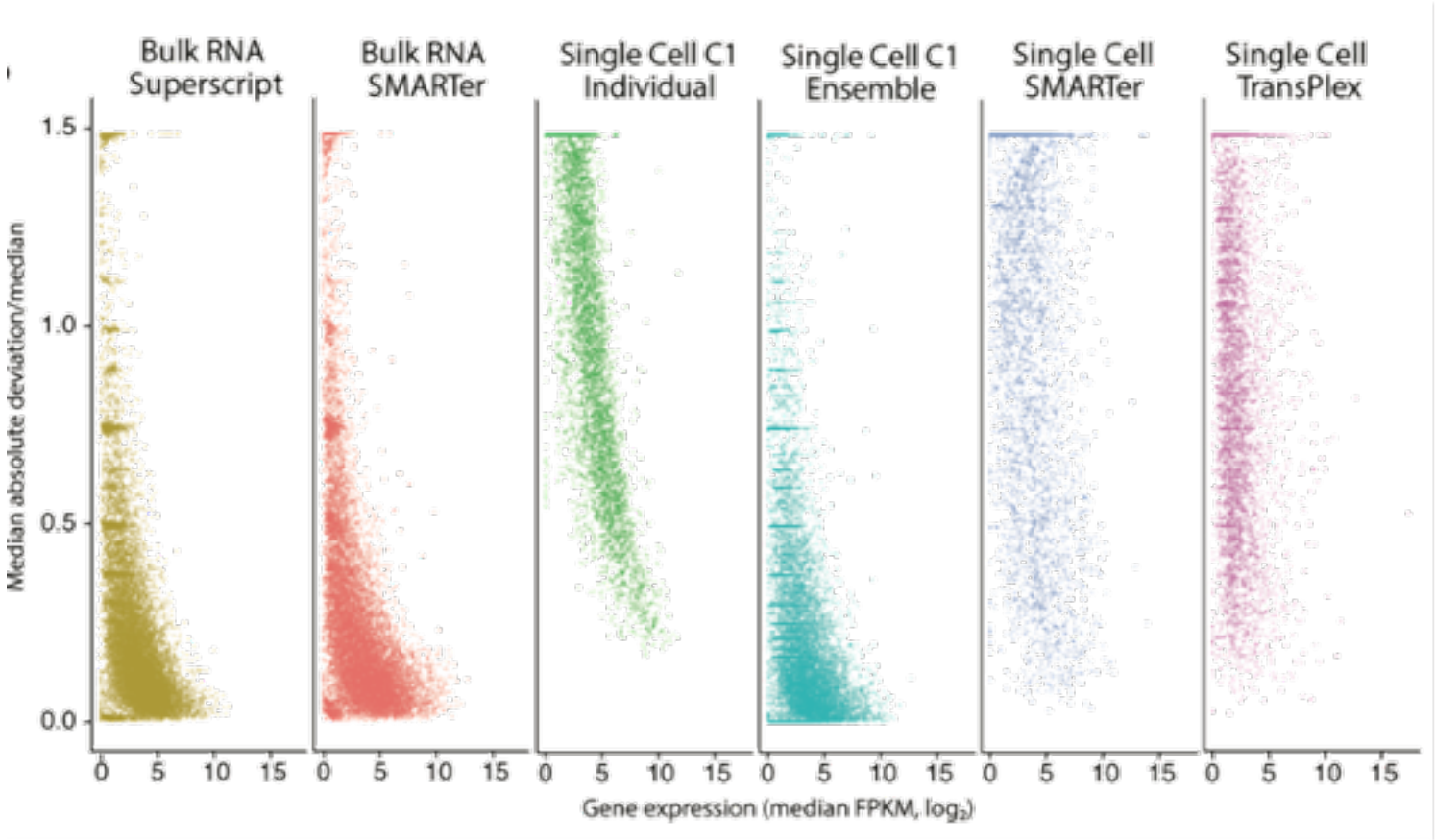
Wu, et al., 2014, Nature Methods

Ensemble of single cells recapitulates bulk population measurement



Wu, et al., 2014, Nature Methods

Ensemble of single cells recapitulates bulk population measurement

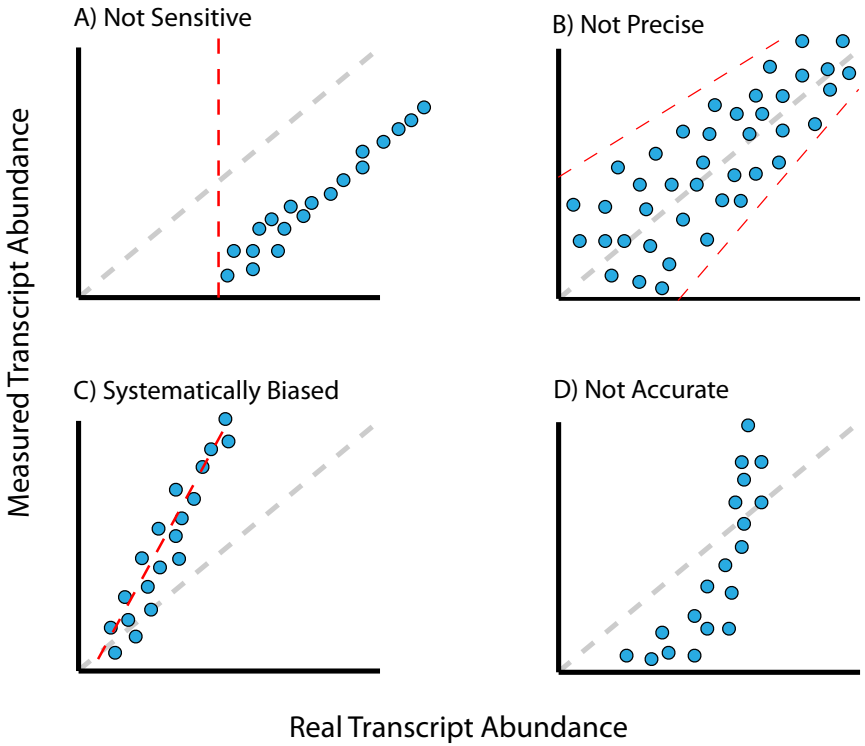


Level of dispersion about the median is similar for synthetic ensemble and bulk samples; single cell samples have relatively higher dispersion for genes with high expression level

Wu, et al., 2014, Nature Methods

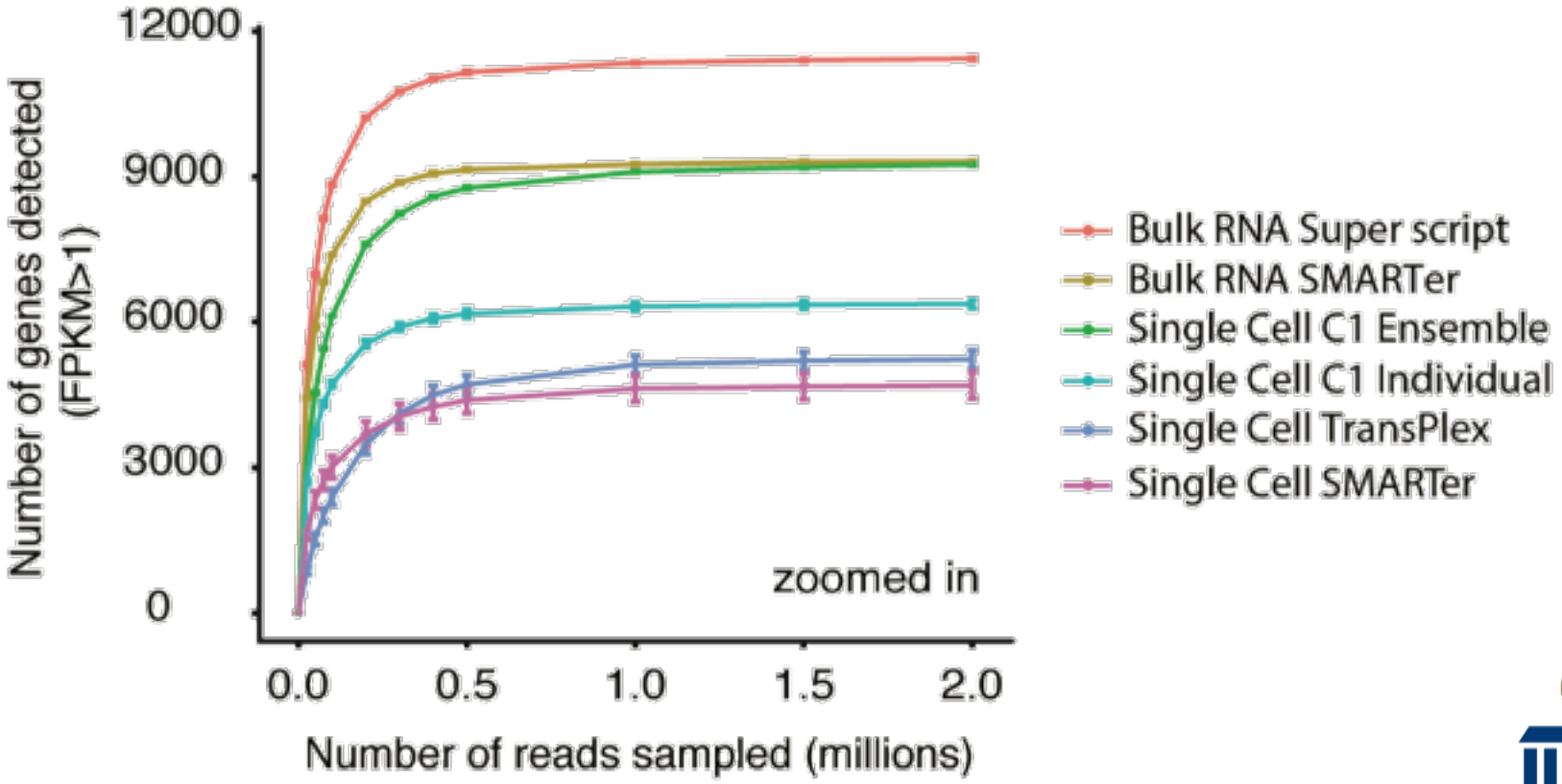
Features of single cell transcriptomic datasets

- ***Accuracy***, or "how quantitative?"
- ***Sensitivity***, or "how deeply do I need to sequence?"
- Technical/Stochastic vs. Biological variation, i.e. ***the noise***



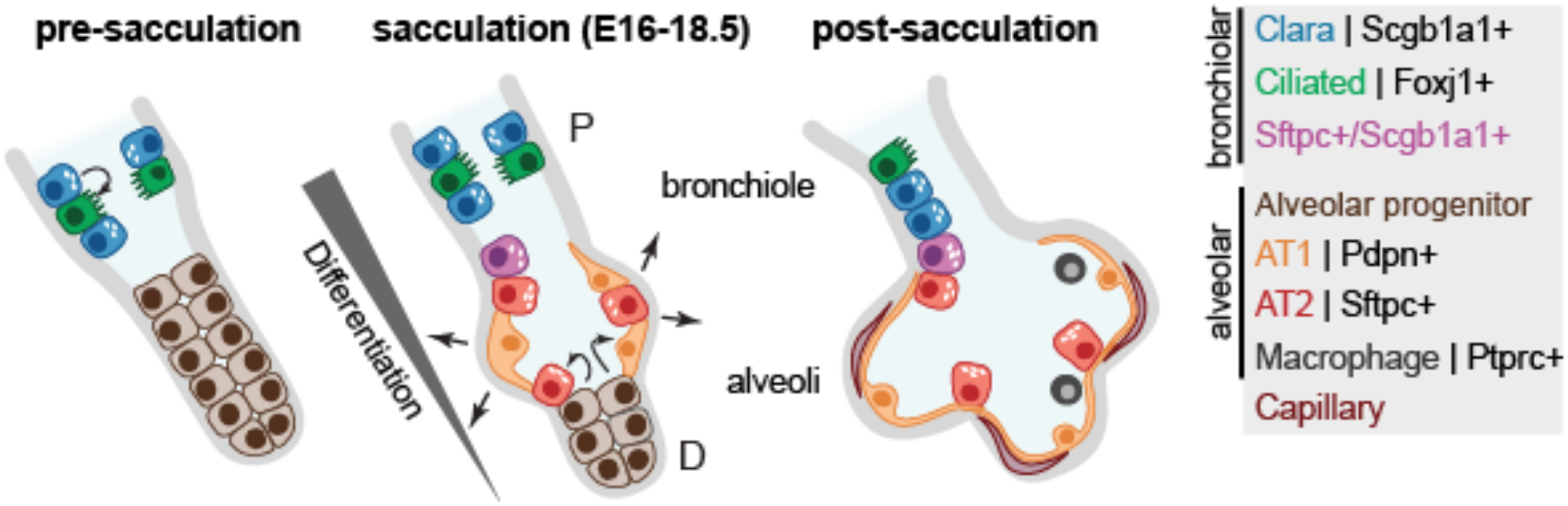
Wu, et al., 2014, Nature Methods

Microfluidics sample preparation improves RNA-seq sensitivity



Wu, et al., 2014, Nature Methods

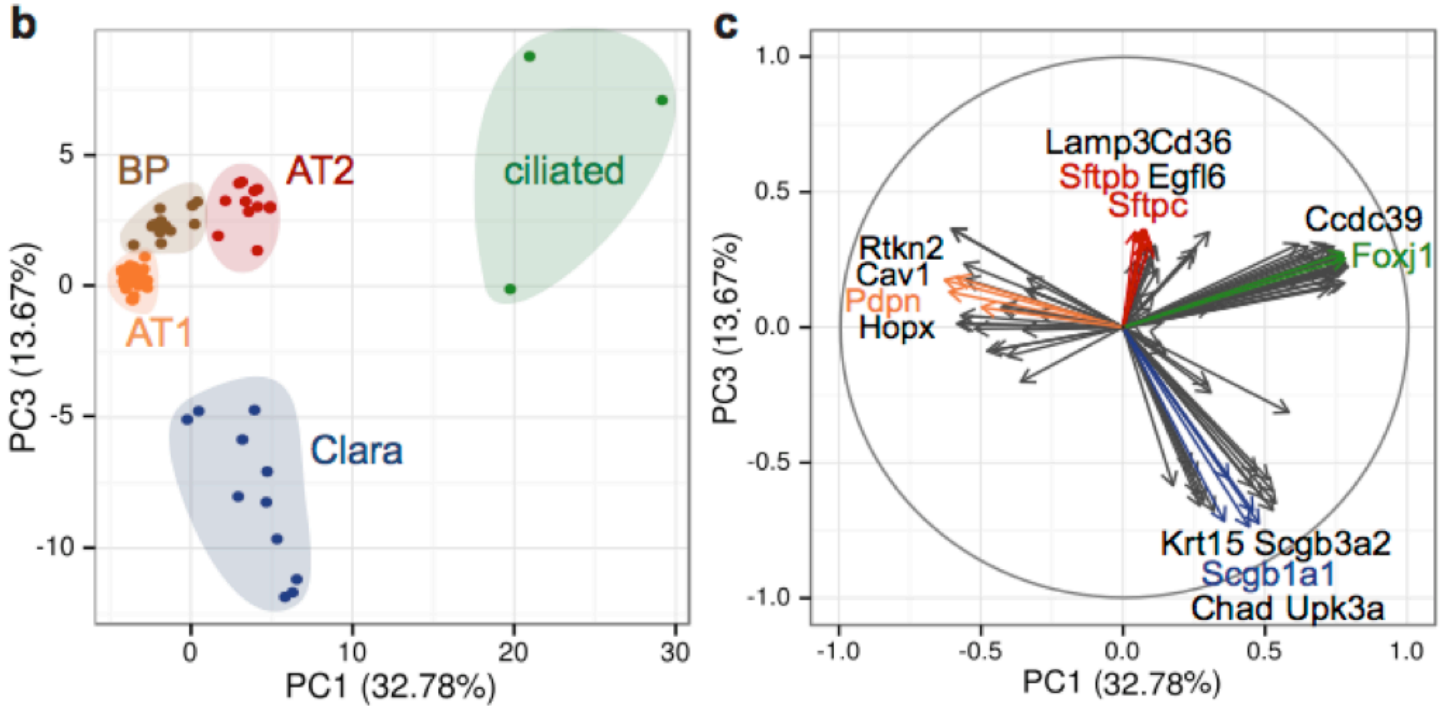
Studying lung development using single-cell gene expression analysis



- Developmental lung biology:
  - Cell differentiation is directional
  - Progenitors persist longest at the tips
  - Widening of airway structures to form alveolar sacs

Treutlein, Brownfield*, Wu, et al., 2014, Nature*

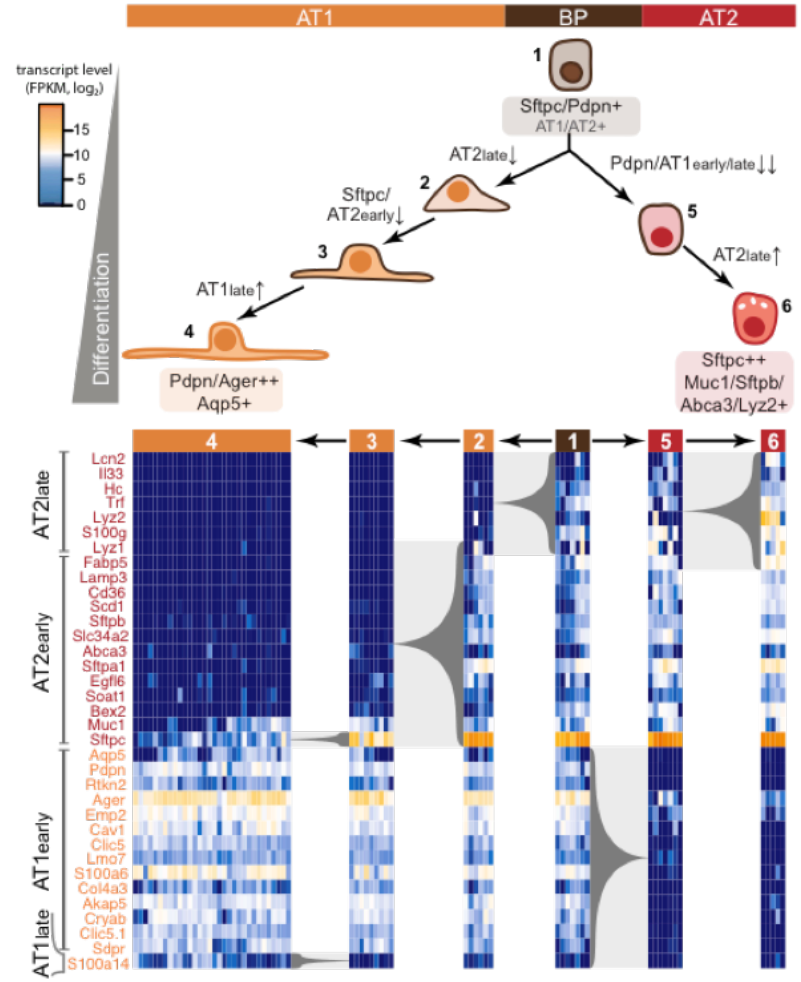
RNA-seq identifies bipotent progenitor cells in alveolar development



- PCA found genes with highest loading at day E18.5 (late sacculation)
- Unsupervised clustering revealed bipotent progenitors

Treutlein, Brownfield*, Wu, et al., 2014, Nature*

Reconstructed differentiation pathway of BPs into AT1 and AT2 lineages



- Using genes identified in BP, AT1, and AT2, individual cells can be classified into sub-populations of intermediate cell types between BP and mature AT1 or AT2
- Reconstruction of lineage differentiation based on gene expression
- Additional support from pathway analysis

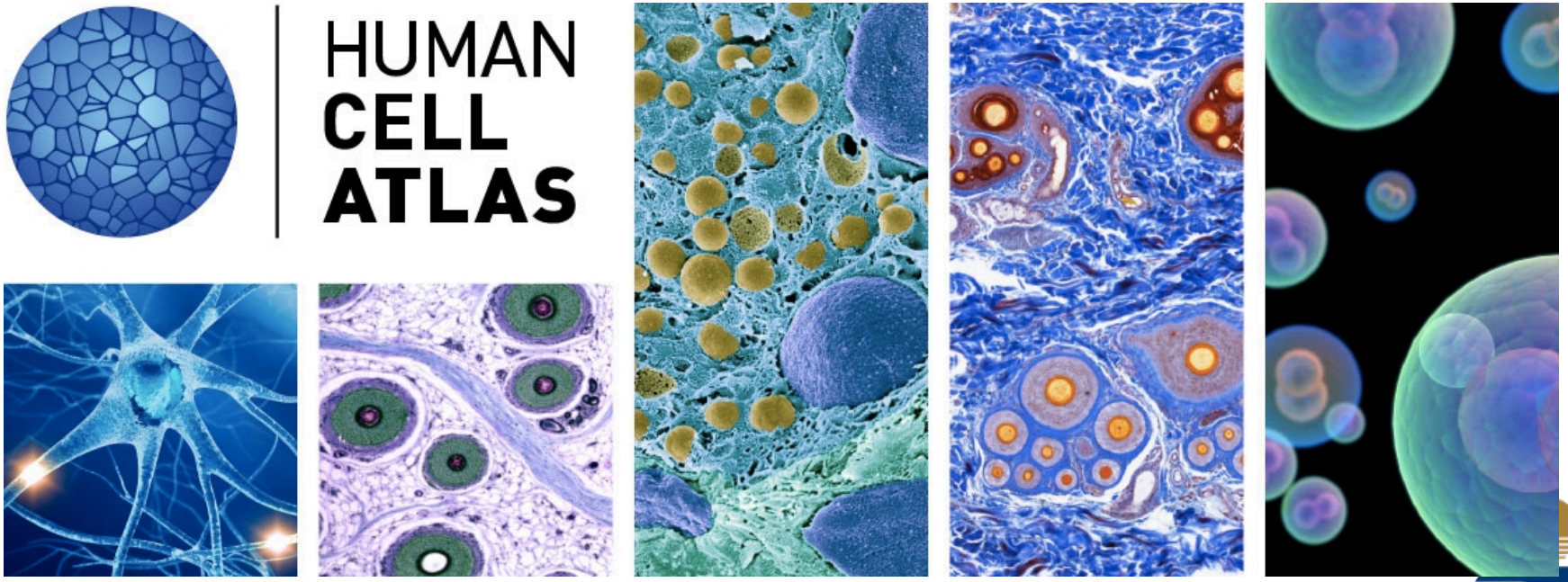
Treutlein, Brownfield*, Wu, et al., 2014, Nature*

What can we do with single-cell RNA-seq?

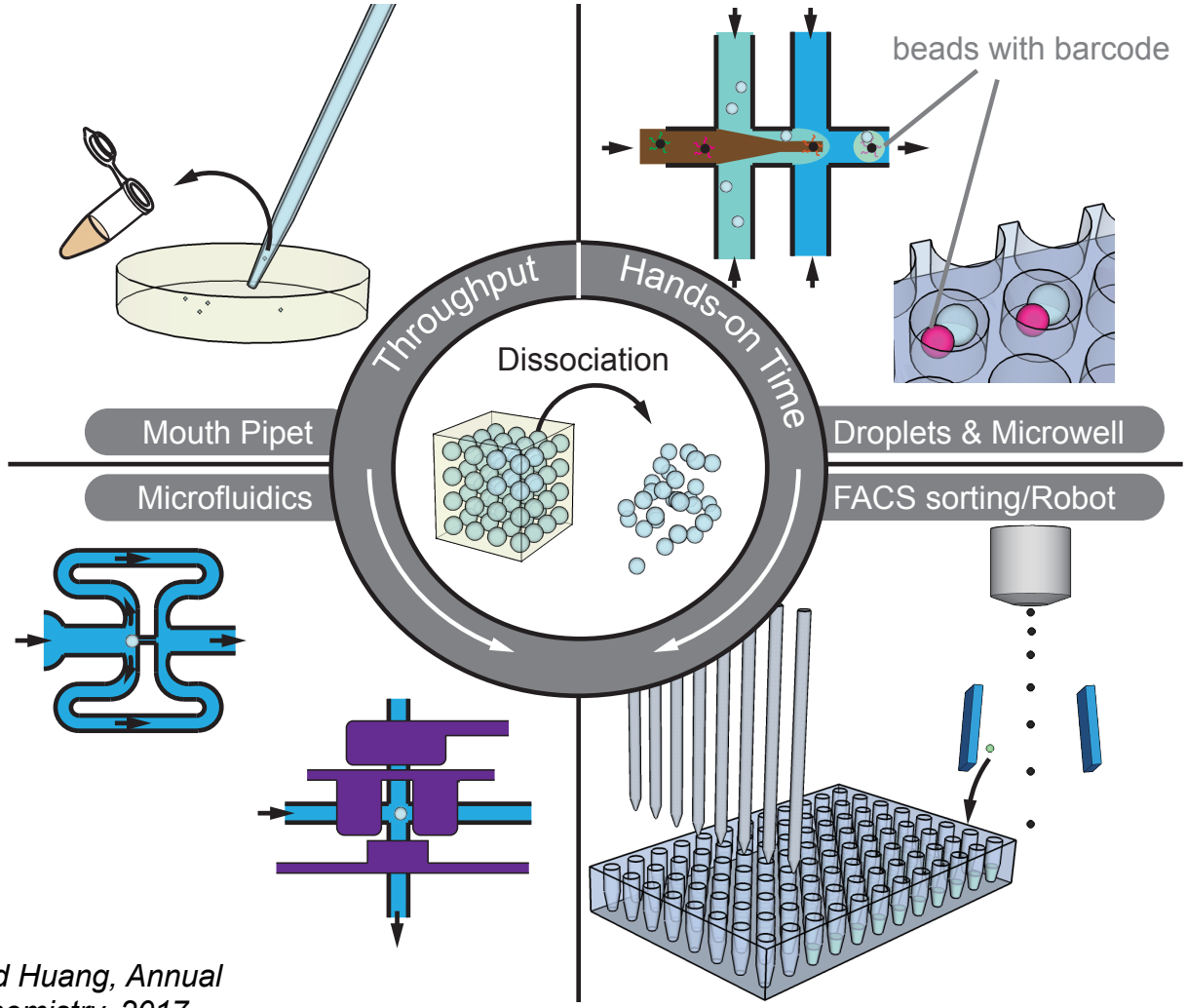
The Human Cell Atlas

A "Google Maps" For the cells in the human body

…Can it really be done? How?



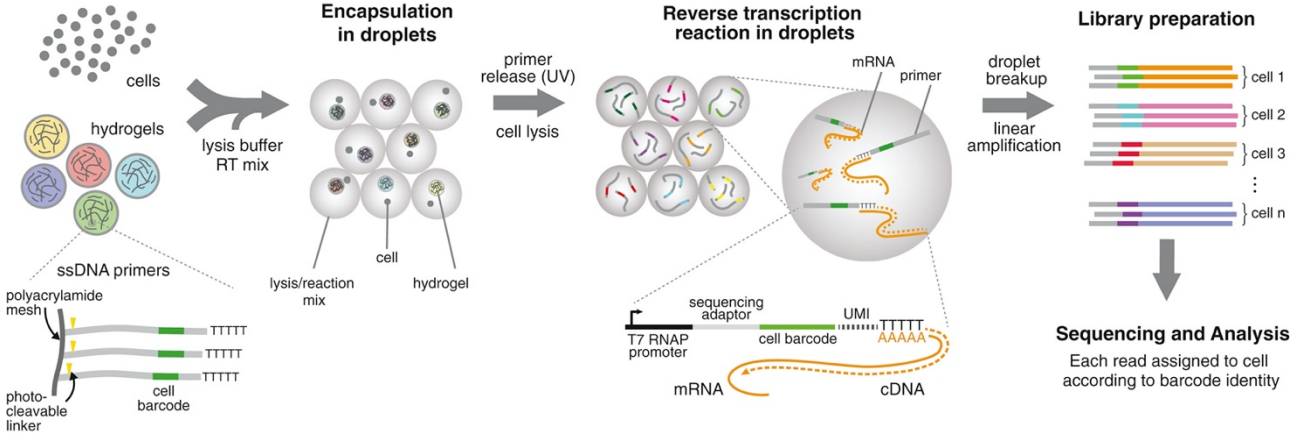
Many technology platforms to choose from

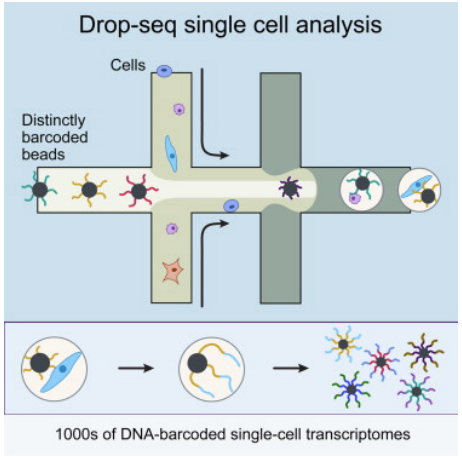


Wu, Wang, Streets, and Huang, Annual Review of Analytical Chemistry, 2017

Microfluidic droplets applied to NGS

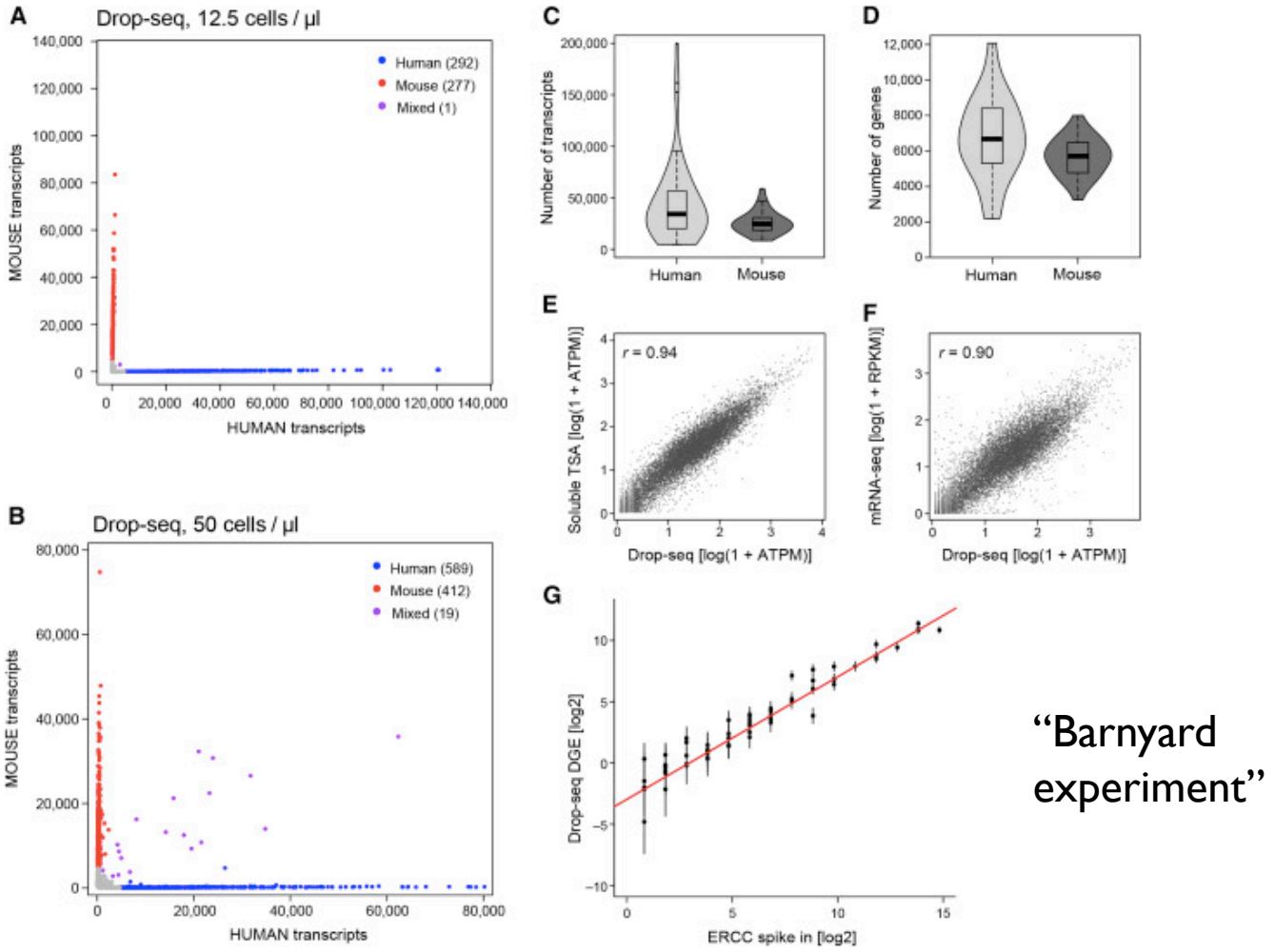
- Using droplets as chambers, we can increase throughput even more, to ~100,000 single cells per run!
- Two Harvard groups published similar technology recently:
  - Drop-seq - https://vimeo.com/128484564
  - inDrop - https://vimeo.com/126829858





Drop-seq: http://www.sciencedirect.com/science/article/pii/S0092867415005498
inDrop: http://www.cell.com/cell/fulltext/S0092-8674(15)00500-0

Microfluidic droplets applied to NGS



"Barnyard experiment"

Single-cell resolution profiling of a whole organism!

Science

RESEARCH ARTICLE

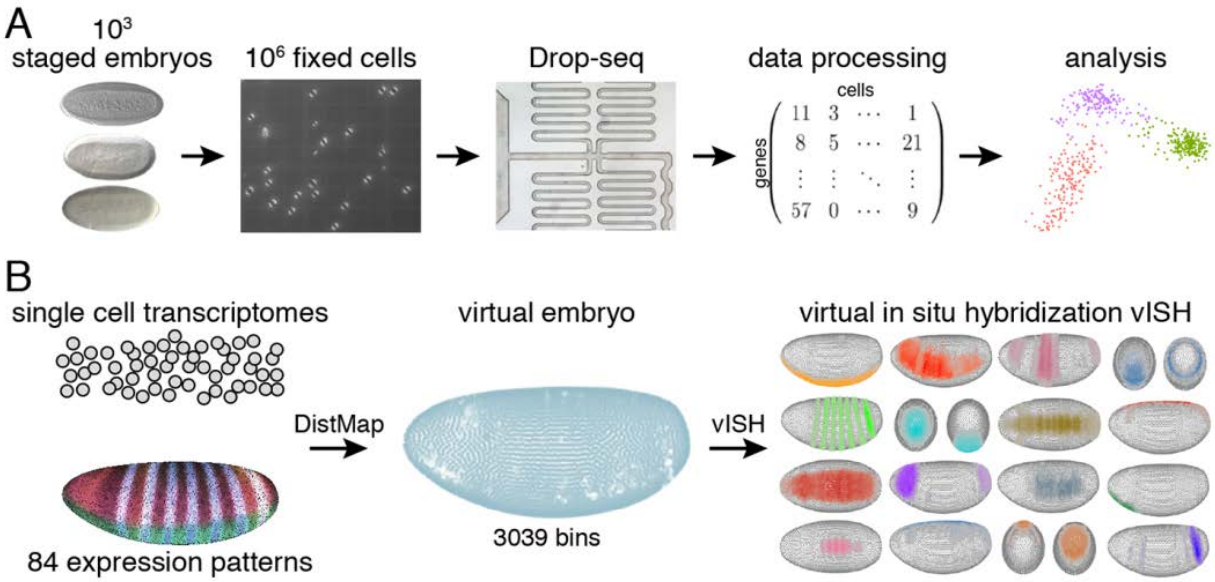
Cite as: N. Karaiskos *et al.*, *Science*
10.1126/science.aan3235 (2017).

The *Drosophila* embryo at single-cell transcriptome resolution

Nikos Karaiskos,[1]* Philipp Wahle,[2]* Jonathan Alles,[1] Anastasiya Boltengagen,[1] Salah Ayoub,[1] Claudia Kipar,[2] Christine Kocks,[1] Nikolaus Rajewsky,[1†] Robert P. Zinzen[2†]

[1]Systems Biology of Gene Regulatory Elements, Berlin Institute for Medical
Association (MDC), 13125 Berlin, Germany. [2]Systems Biology of Neural Tiss
for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, G

*These authors contributed equally to this work.



Single cell resolution profiling of a whole organism!

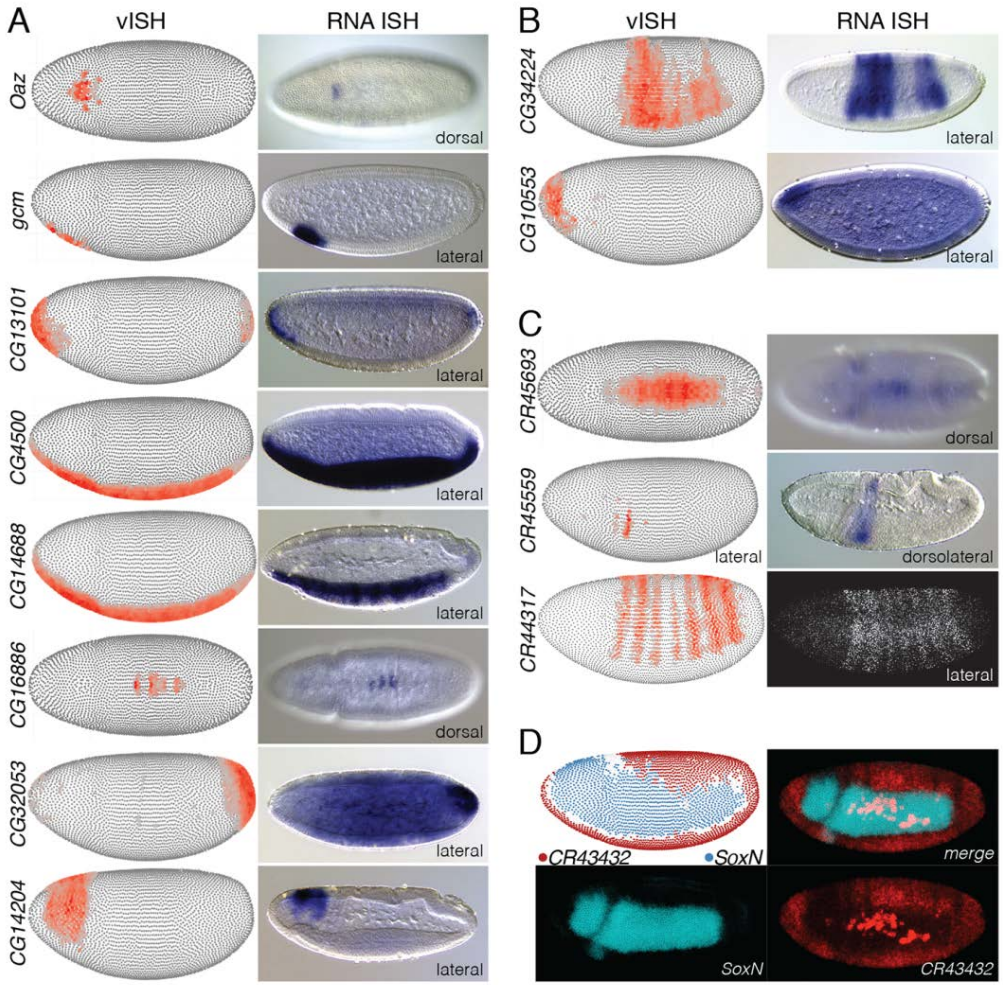


Fig. 5. Prediction accuracy and detection of new regulators. (A) vISH predictions are accurate across a wide variety of expression patterns. Expression of CGs had not been reported previously. (B) Patterned expression of putative transcription factors. (C) Patterned expression of lncRNAs. (D) *CR43432* and pan-neurogenic genes are expressed in complimentary patterns. Dual vISH of *SoxN* and *CR43432* (top left), double in situ hybridization validates the predicted expression. *CR43432* is additionally expressed in yolk nuclei (not shown in vISH).

Single-cell analysis of 20 mouse tissues – mouse cell atlas

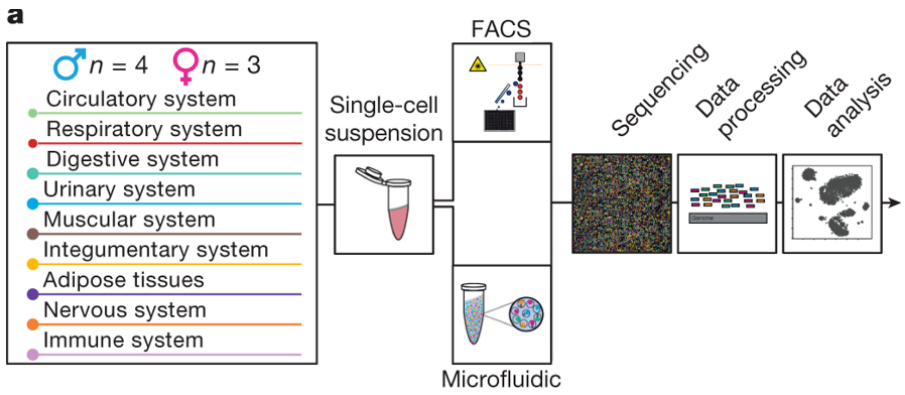
nature
International journal of science

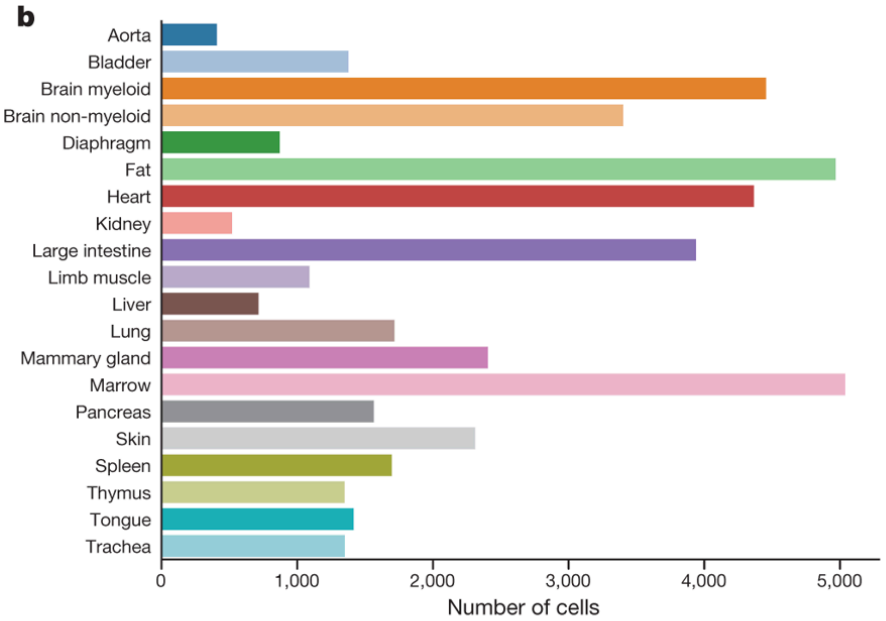
Article | Published: 03 October 2018

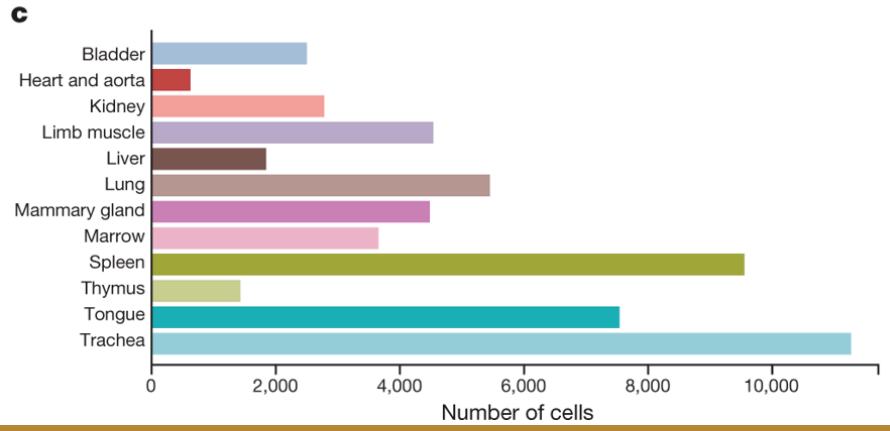
Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*

The Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group & Principal investigators

Nature **562**, 367–372 (2018) | Download Citation

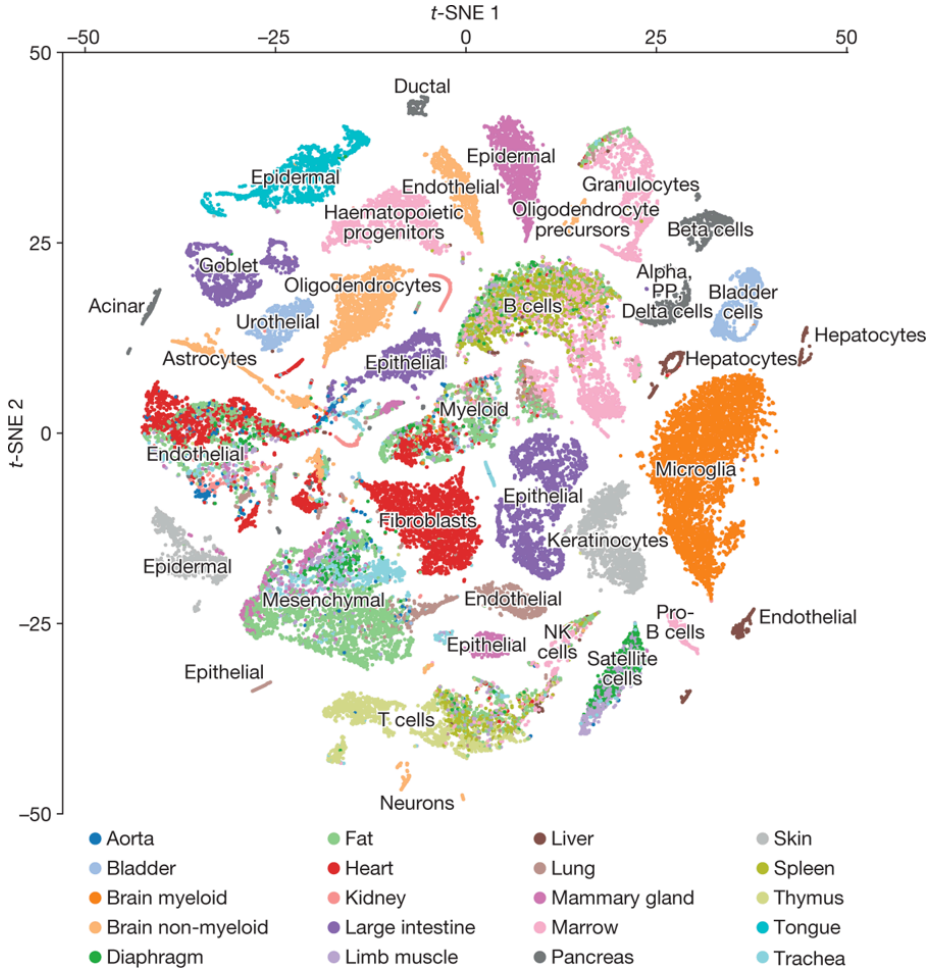






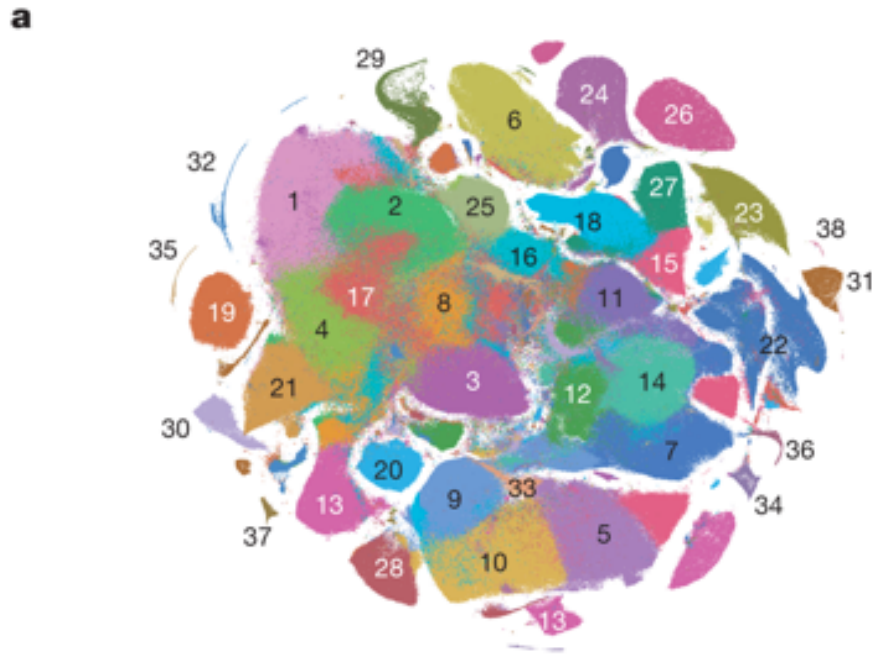
Tabula Muris Consortium. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature, 562(7727), 367.

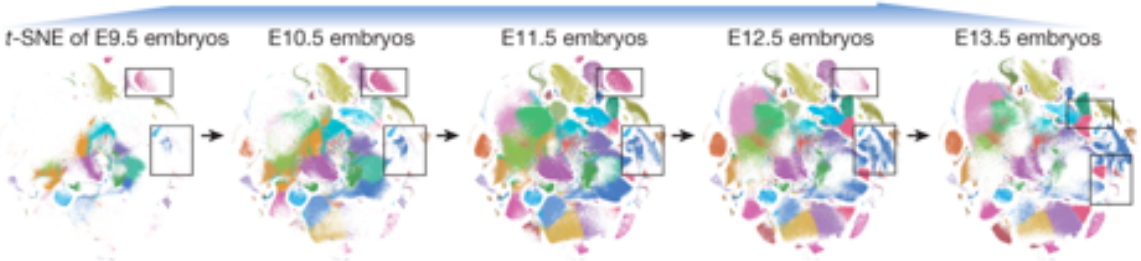
Single-cell analysis of 20 mouse tissues – mouse cell atlas



Tabula Muris Consortium. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature, 562(7727), 367.

Single-cell analysis of 2 million cells from developing mouse embryo





nature
International journal of science

Article | Published: 20 February 2019

The single-cell transcriptional landscape of mammalian organogenesis

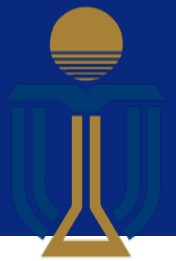
Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J. Steemers, Cole Trapnell & Jay Shendure

Nature **566**, 496–502 (2019) | Download Citation

25
26–Primit
27–Int
30–Notochord
33–
34–Ca

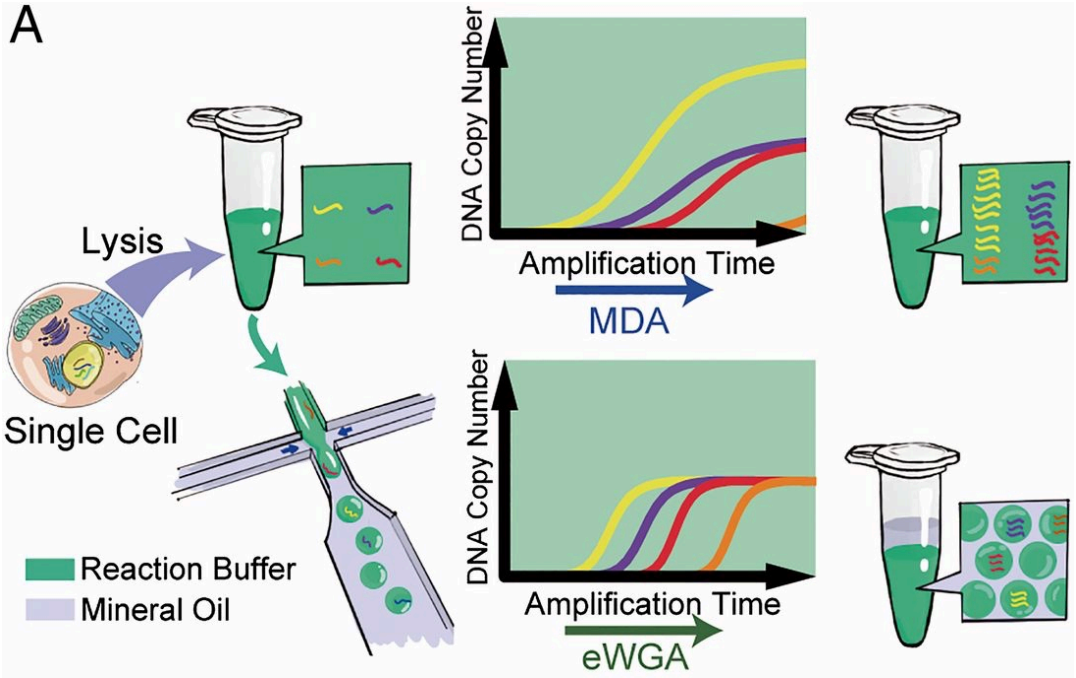
https://tabula-muris.ds.czbiohub.org

SINGLE CELL WHOLE GENOME SEQUENCE (WGS)



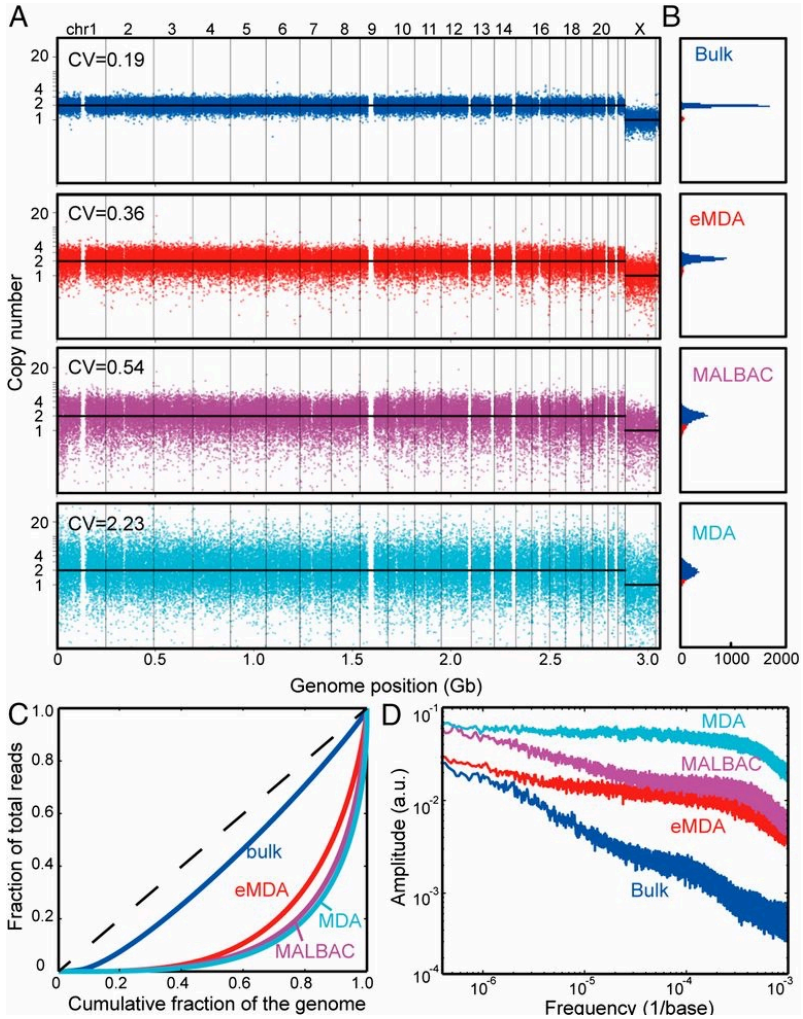
Microfluidic droplets applied to NGS

- Single cell DNA sequencing using Multiple Displacement Amplification (MDA) is known to have problems of amplification bias (e.g. preference for GC rich regions)
- Huang group at Peking University solves this problem using droplet-based MDA (http://www.pnas.org/content/112/38/11923.full)



Fu et al, PNAS, 2015

Microfluidic droplets applied to NGS



Single-cell proteomics

CYTOF – cytometry and time-of-flight

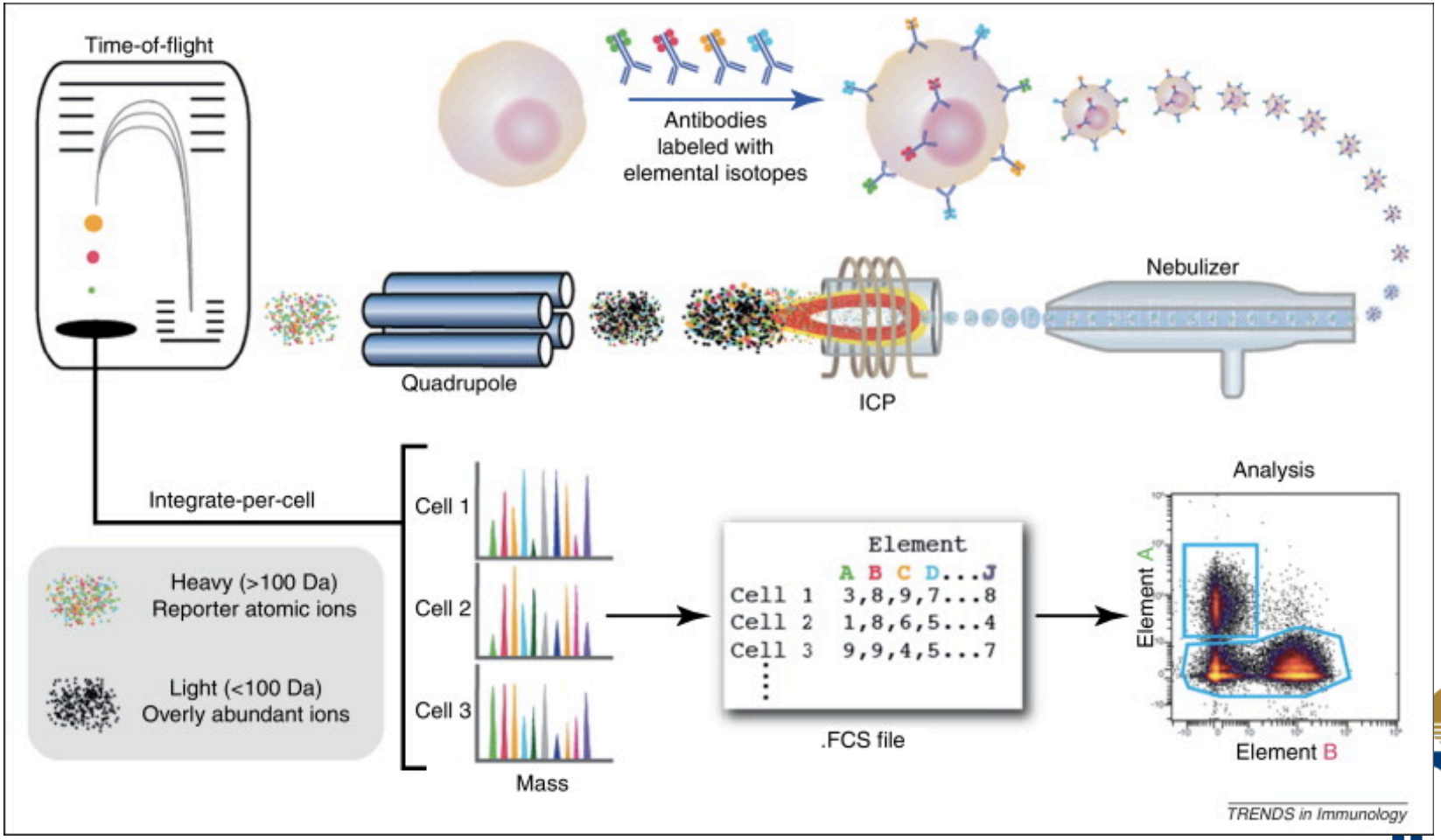
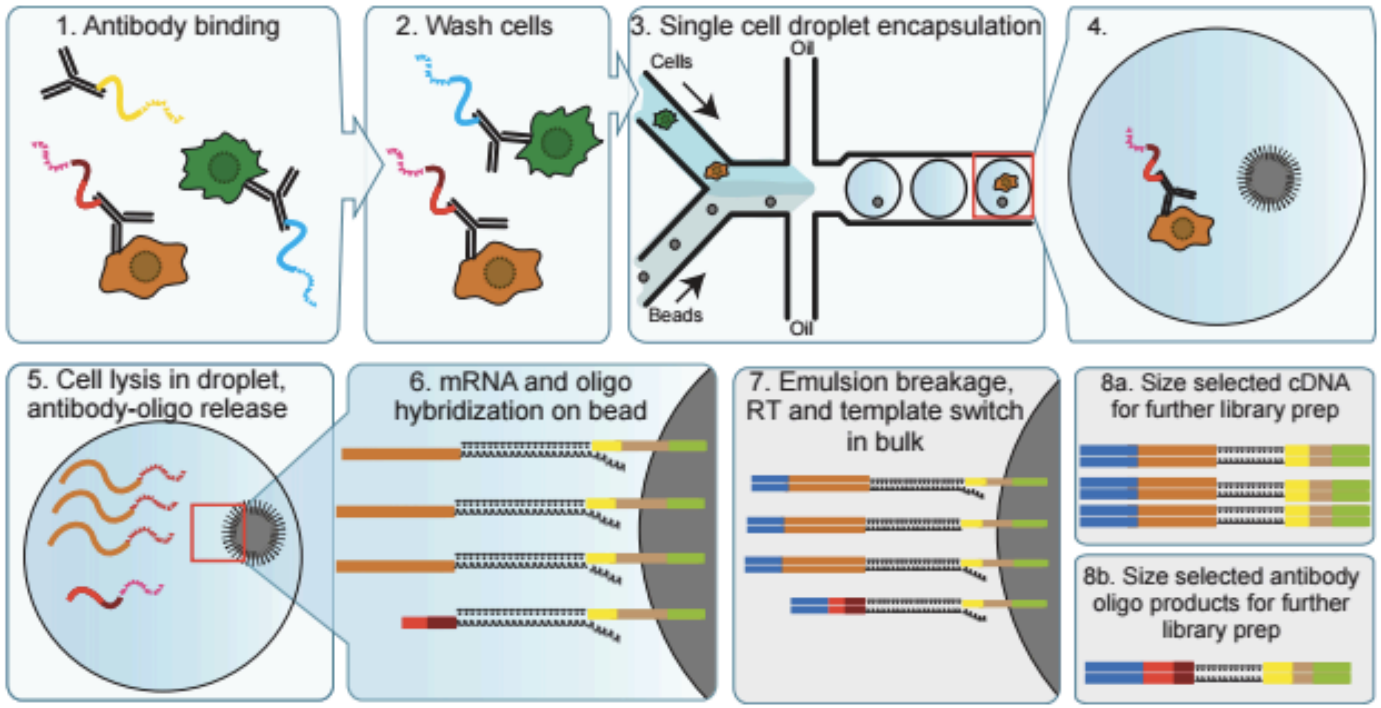


Image: http://cytof.scilifelab.se/homepage/static/images/cytof.jpg

Review: Bendall, Sean C., and Garry P. Nolan. "From single cells to deep phenotypes in cancer." Nature biotechnology 30.7 (2012): 639-647.

Single-cell multi-omics

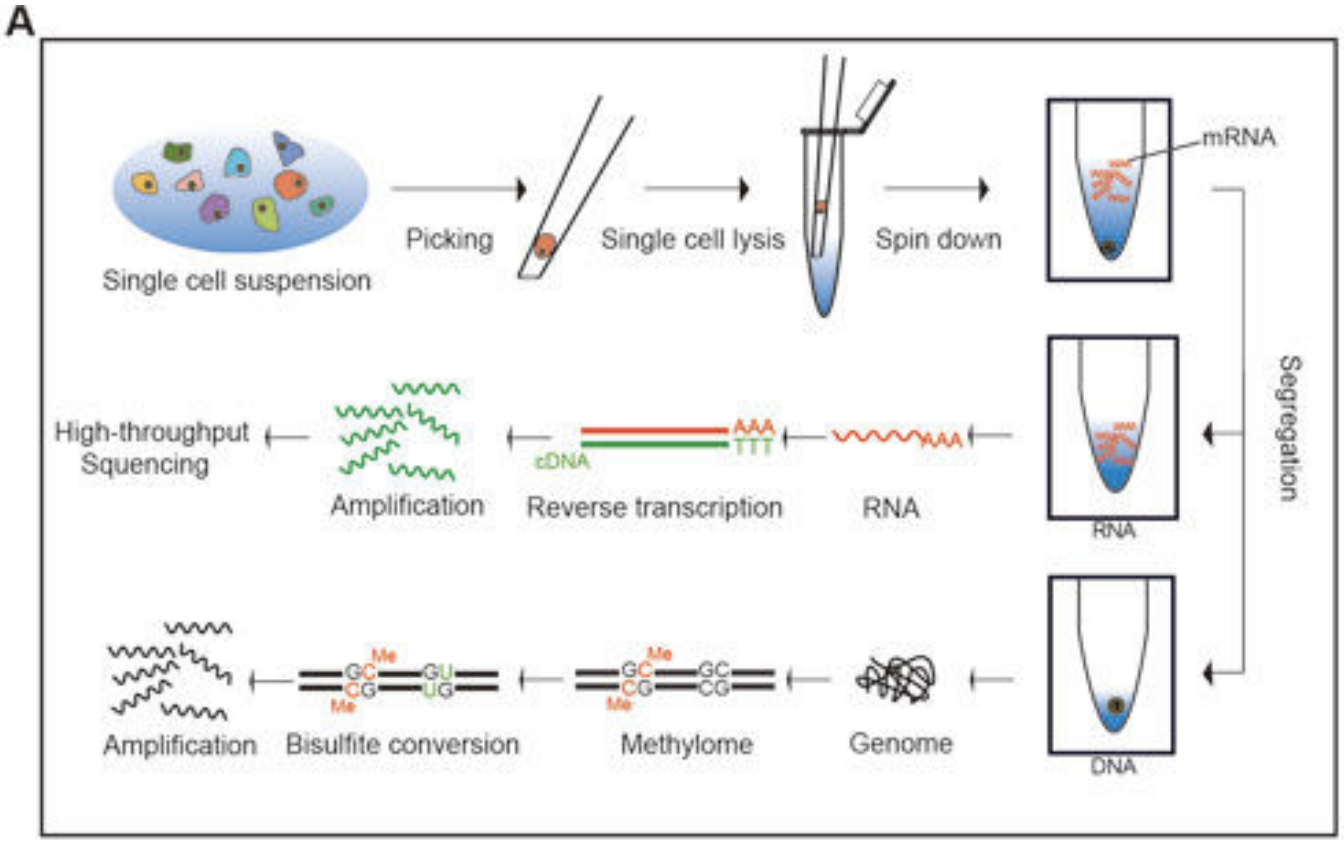
Ab-seq or CITE-seq: cellular indexing of transcriptomes and epitopes by sequencing



Stoeckius, Marlon, et al. "Simultaneous epitope and transcriptome measurement in single cells." Nature 201 (2017): 7.

Single-cell multi-omics

scTrio-seq



Hou, Yu, et al. "Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas." Cell research 26.3 (2016): 304-319.