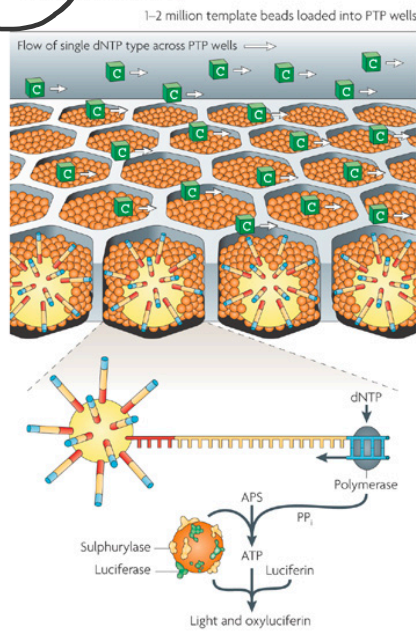
THE OMICS ERA PART 2

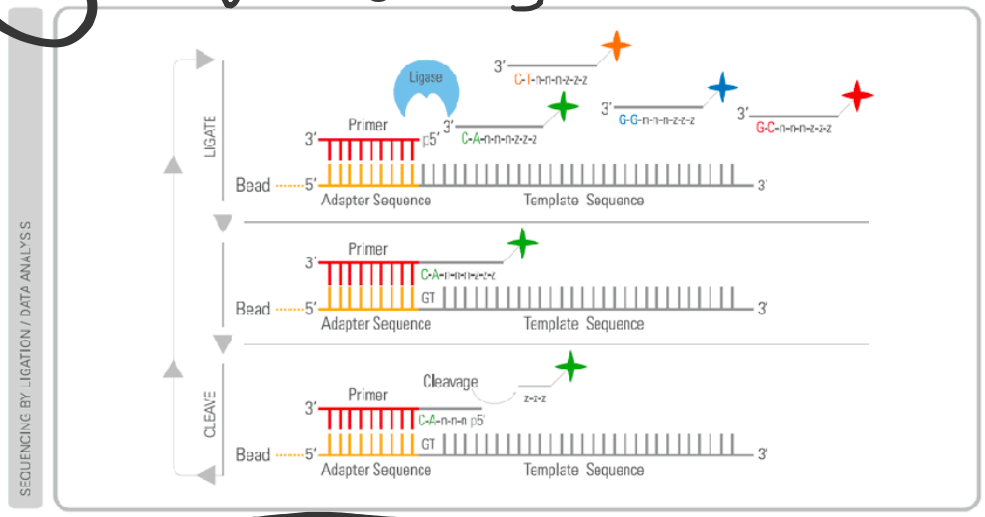
Genomics and its applications; Proteomics and single-cell technologies

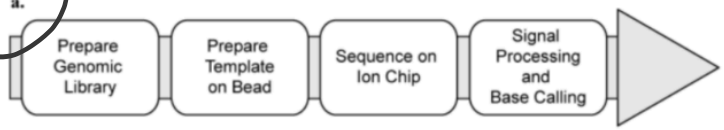
Roche/454 — Pyrosequencing

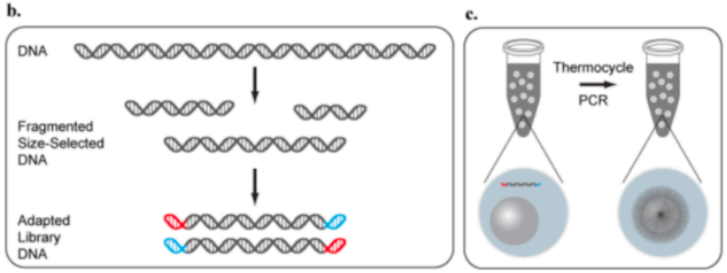


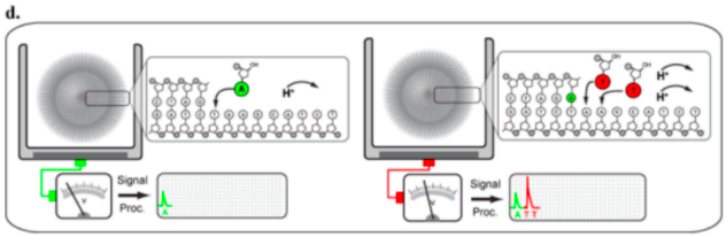
Pyroseq

Seq. by ligation

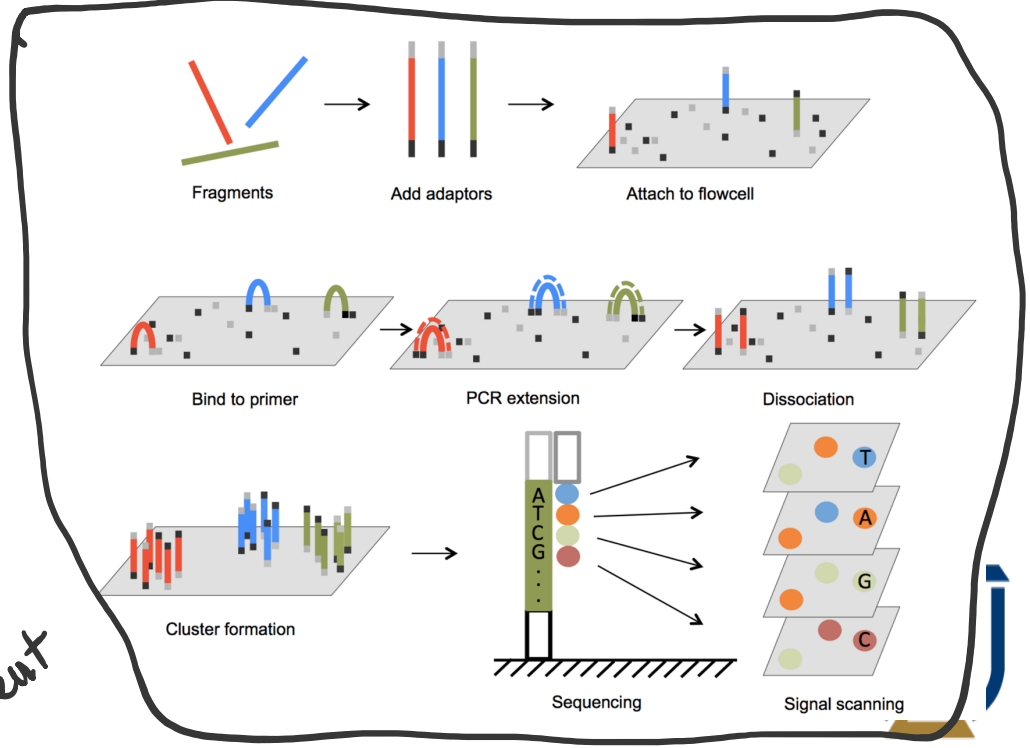




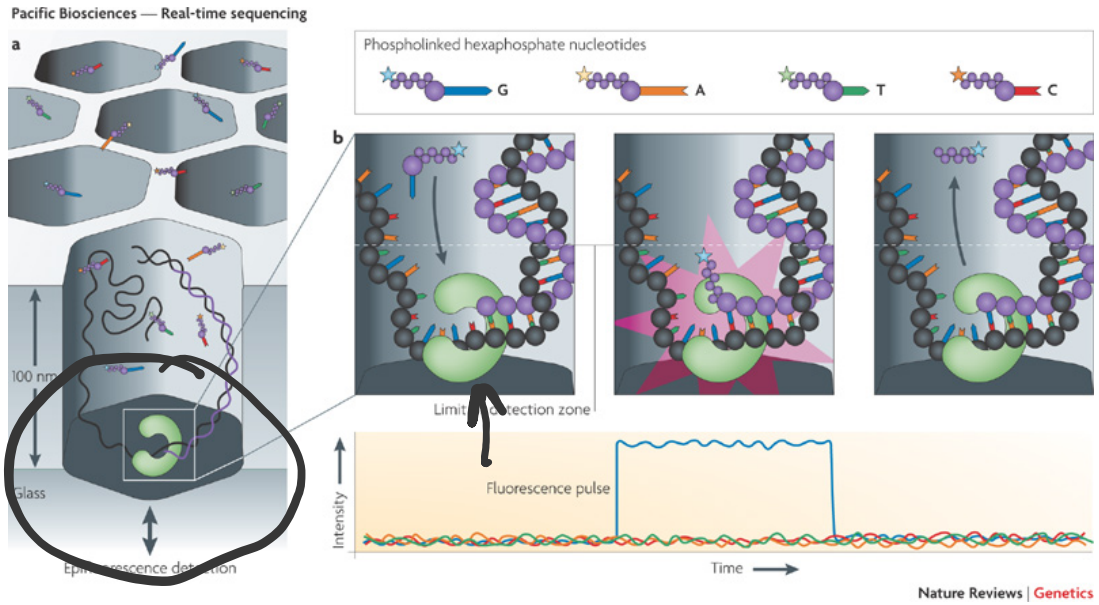




H+ current

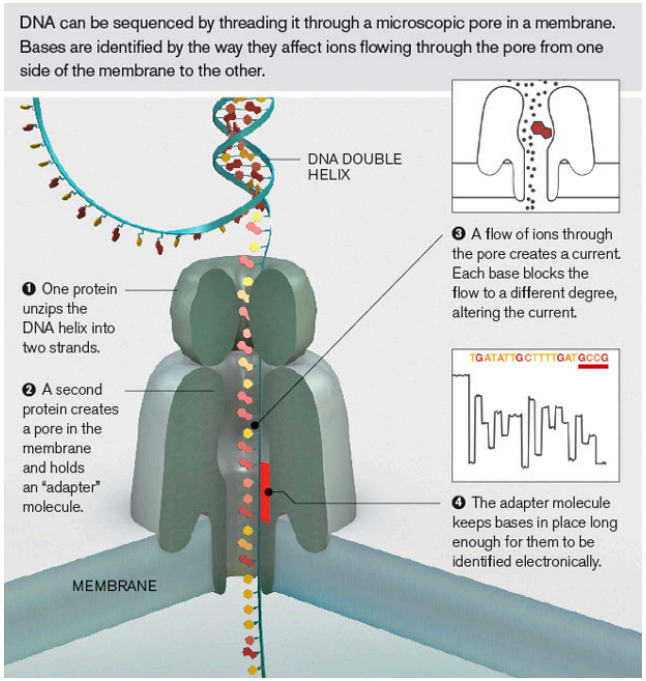


3RD GEN – SMS



ZMW

4TH GEN



Nanopore

Applications of NGS

- Cancer research ← mutations, incorrect transcription/transl.
- 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
    DNA

- 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)

~~2010~~
~ 2010
Seq. patient cancer genomes.

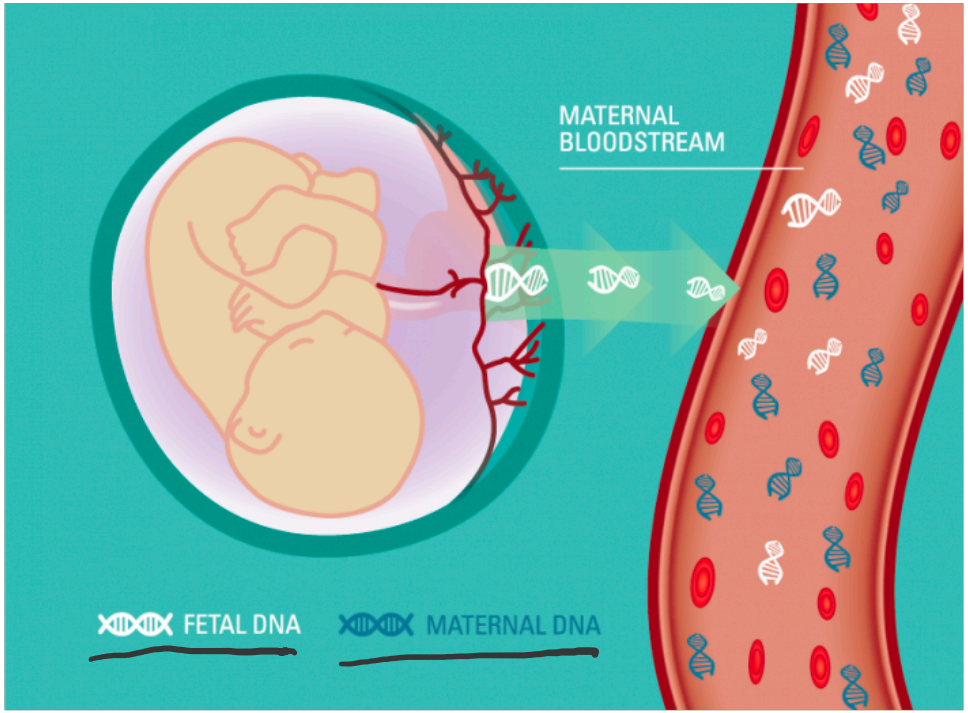
exome - all exons.

Applications of NGS

- Pre-natal diagnostics
  - DNA
  - RNA

cell-free DNA / RNA
very short 150 - 250 bp.

fetal DNA %
~ 1 - 10%
Varies by trimester



From Ariosa website

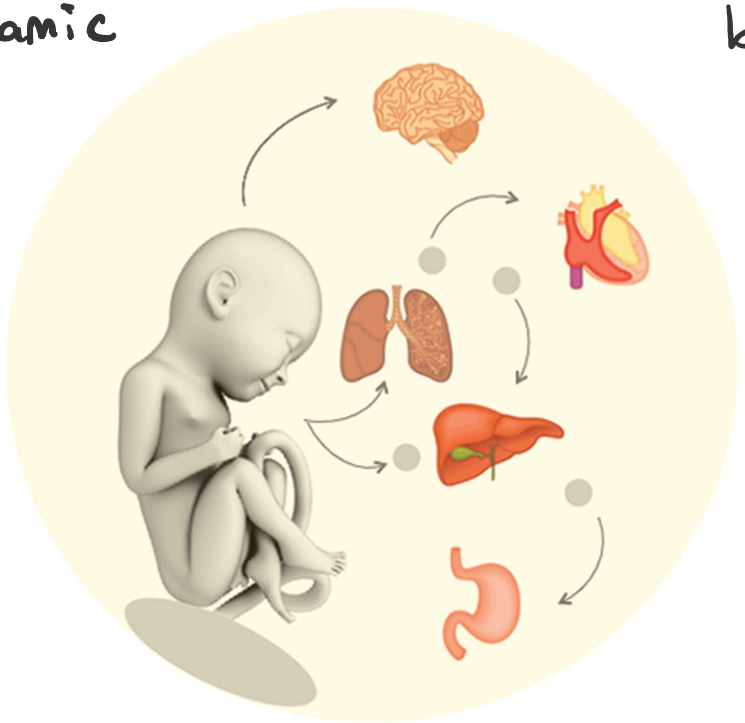
NON-INVASIVE

Applications of NGS

- Pre-natal diagnostics
  - DNA → methylation state
    dynamic
  - RNA

dynamic
always change

fetal liver specific RNA
↓ ↓
brain



organ/tissue specific
contributions.

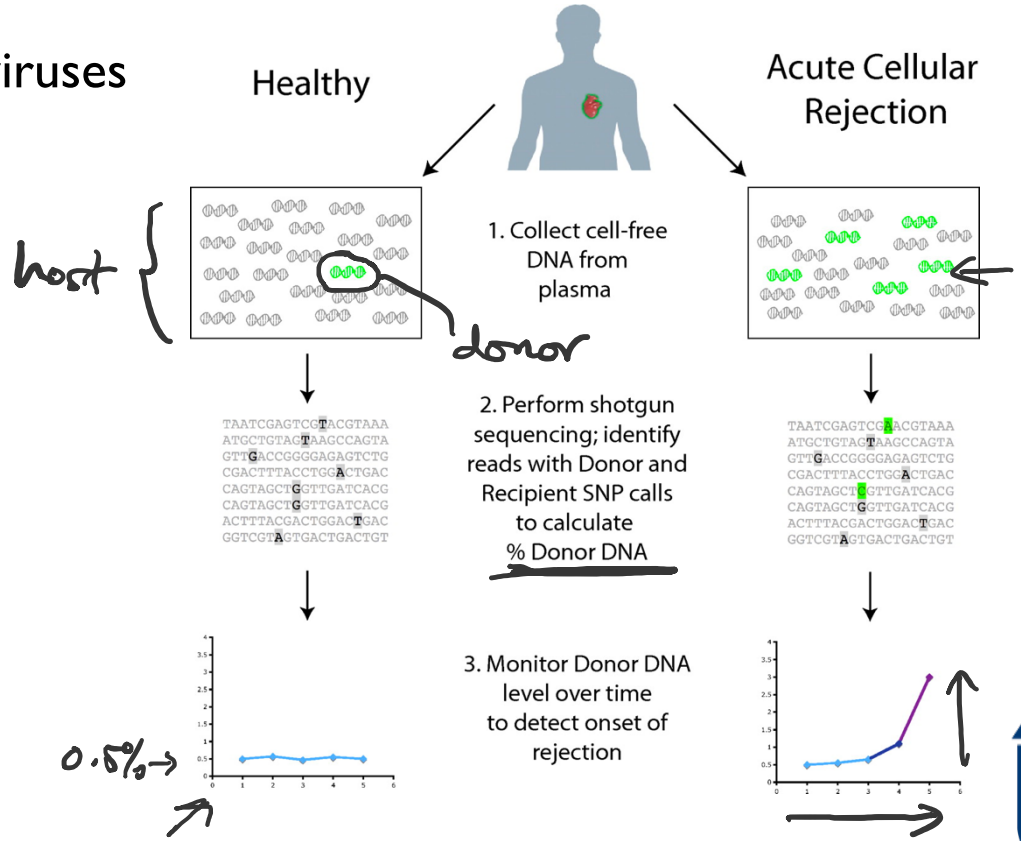
Lillian M. Zwemer, and Diana W. Bianchi Cold Spring Harb Perspect Med 2015;5:a023101

Applications of NGS

cell-free nucleic acids

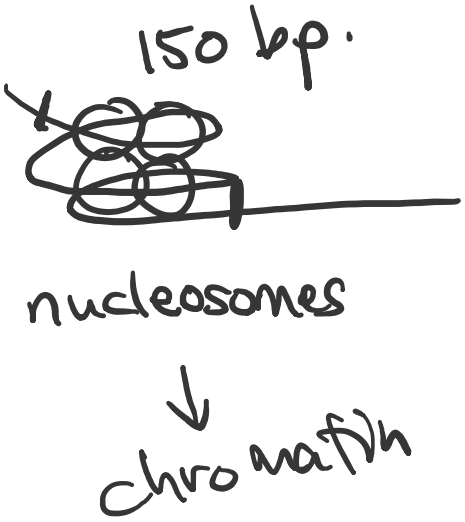
- Predicting organ transplant rejection
  - DNA of donor
  - RNA of microbes and viruses

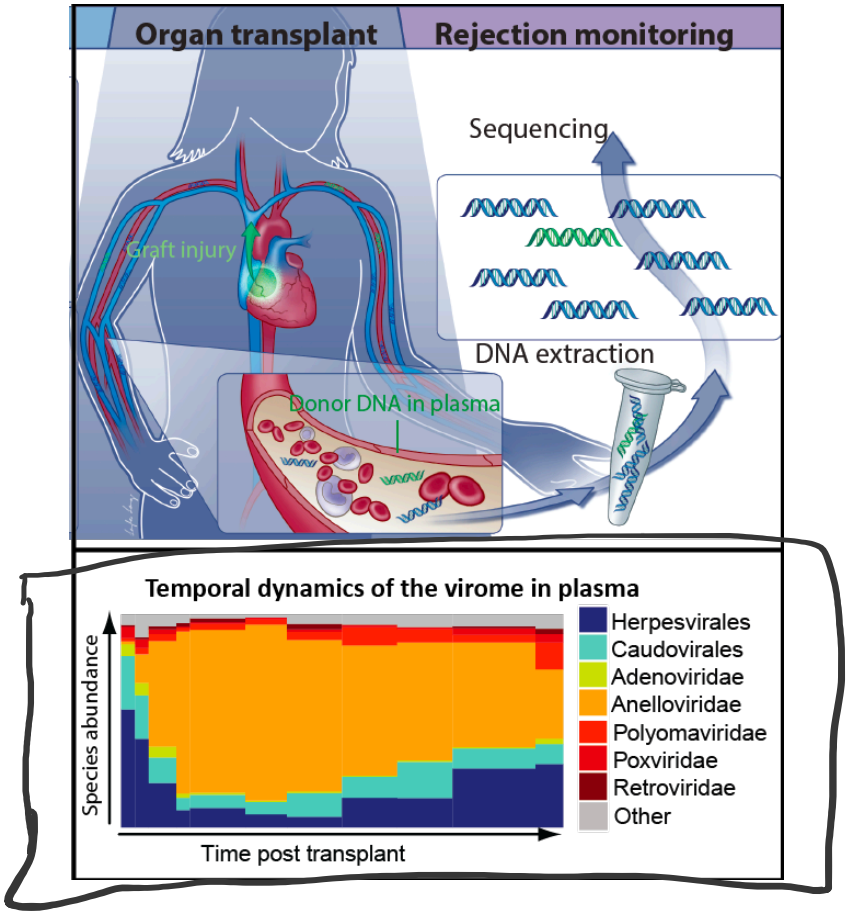
immunosuppressive drug.



Applications of NGS

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





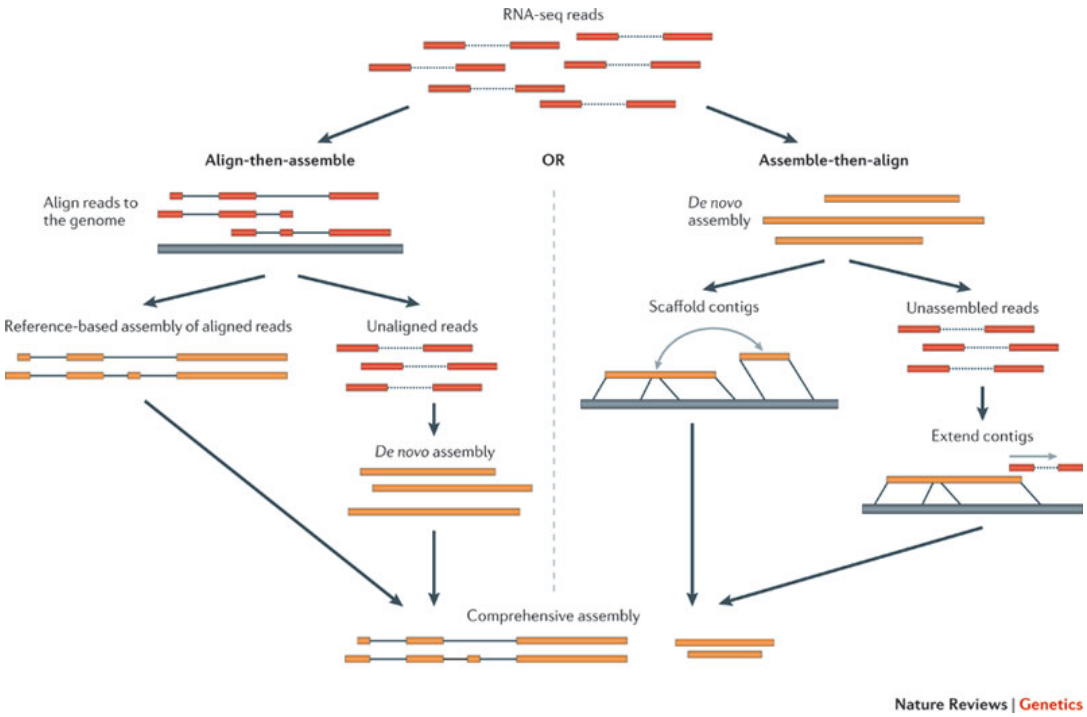
De Vlaminck et al., Science Translational Medicine, 2014

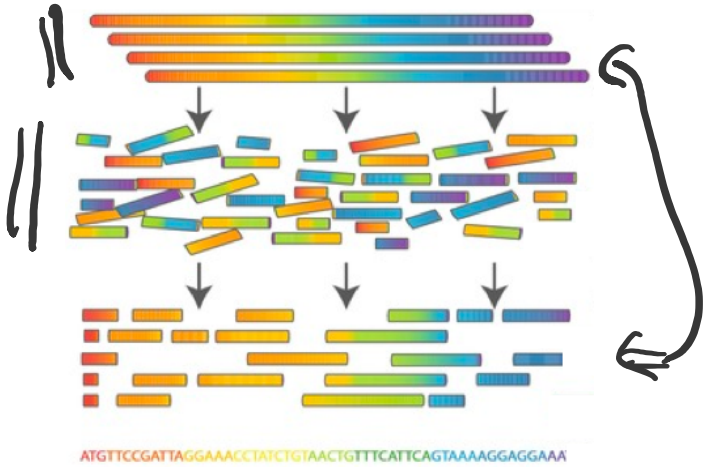
Applications of NGS

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

150bp.
1k - 2k.

brand new

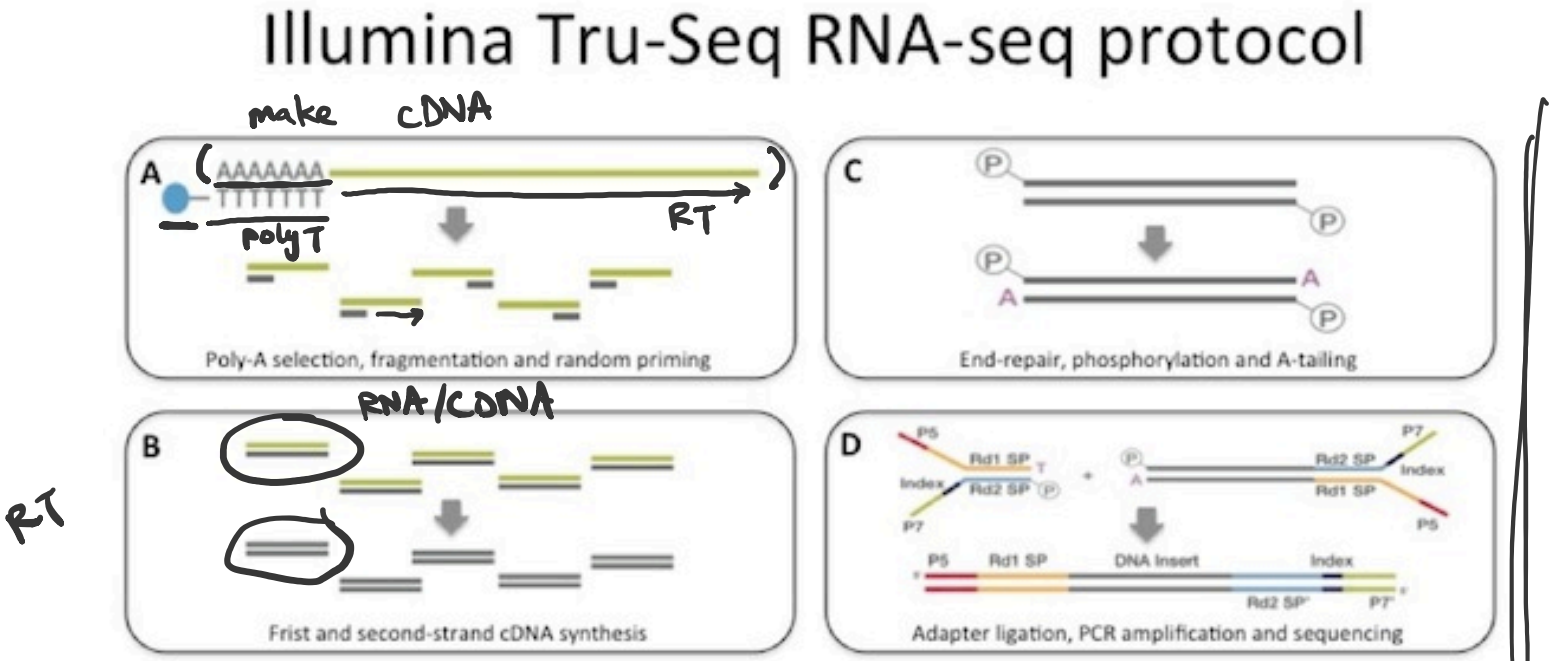




Applications of NGS

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

RNA-seq

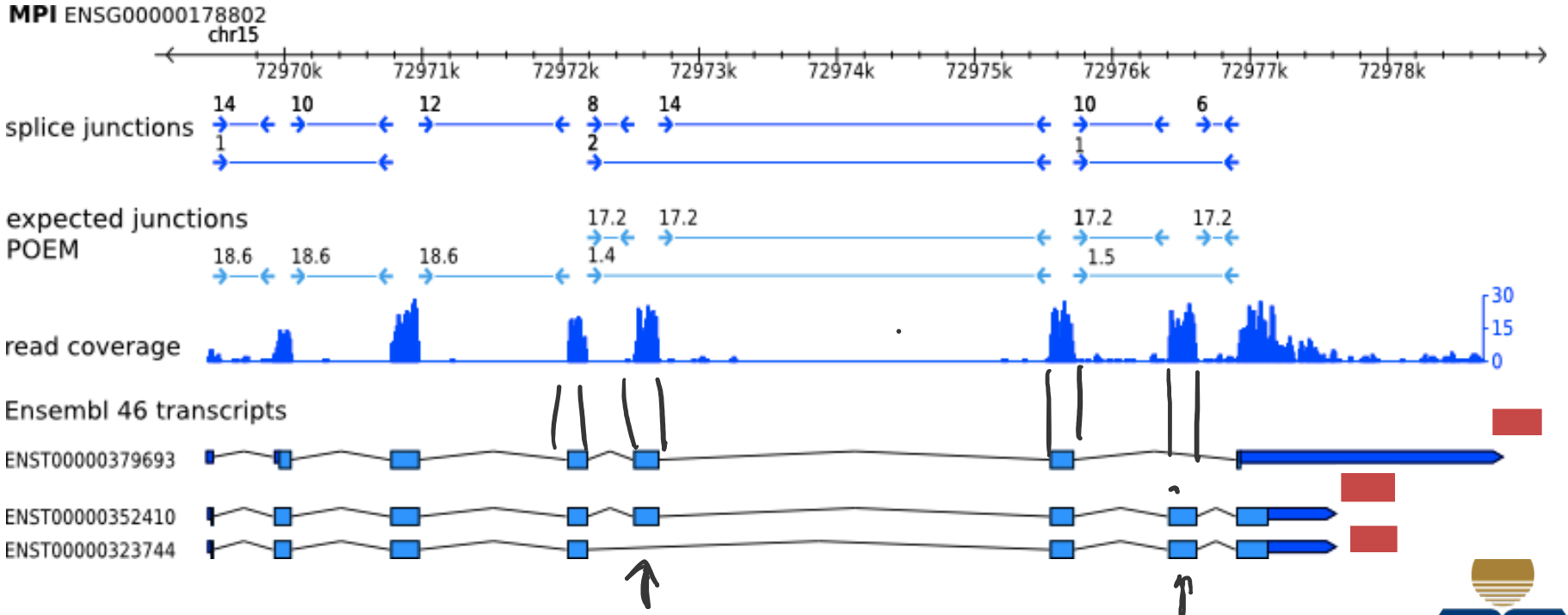


Library prep begins from 100ng-1ug of Total RNA which is poly-A selected (A) with magnetic beads. Double-stranded cDNA (B) is phosphorylated and A-tailed (C) ready for adapter ligation. The library is PCR amplified (D) ready for clustering and sequencing.

Image from BiteSize Bio

RNA-seq

Example of data

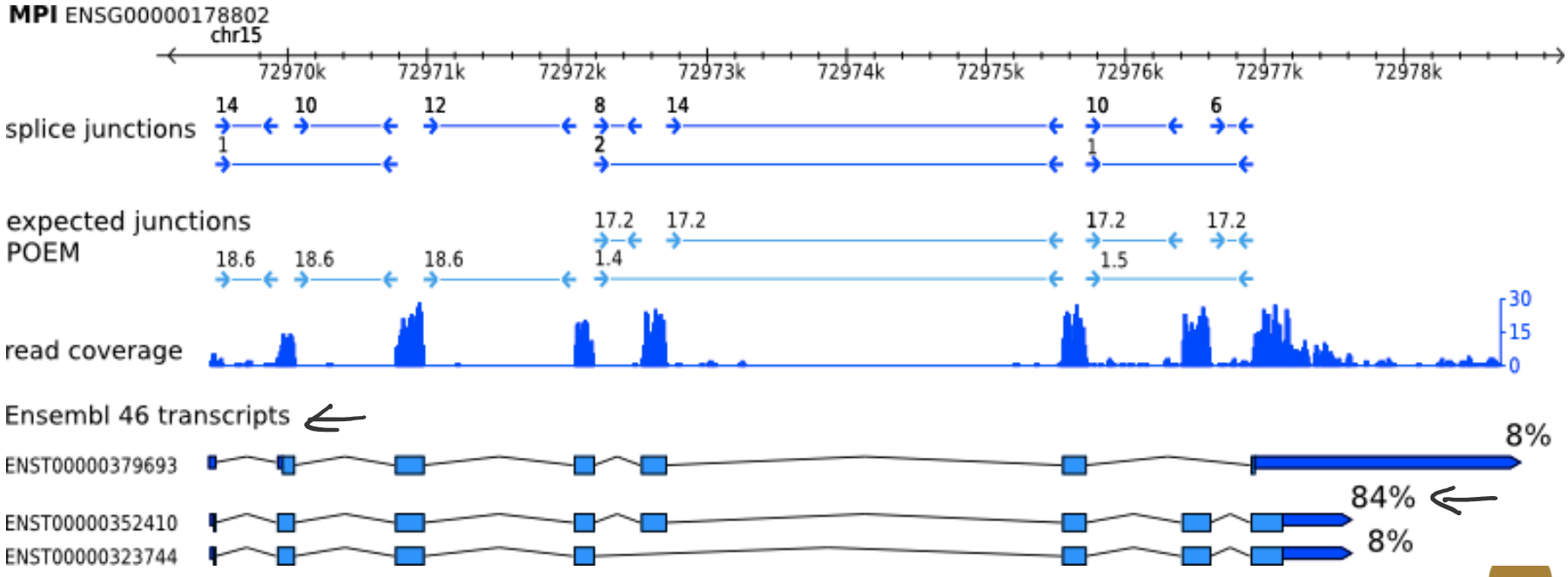


Which of the three transcripts is expressed with highest abundance?

Schulz MH, et al., Bioinformatics, 2012

RNA-seq

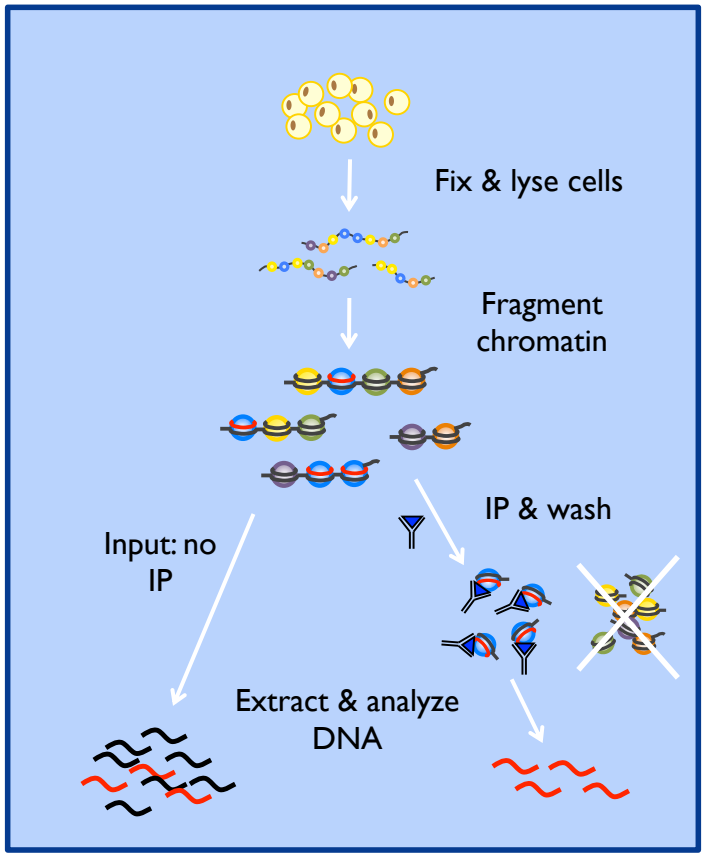
Example of data



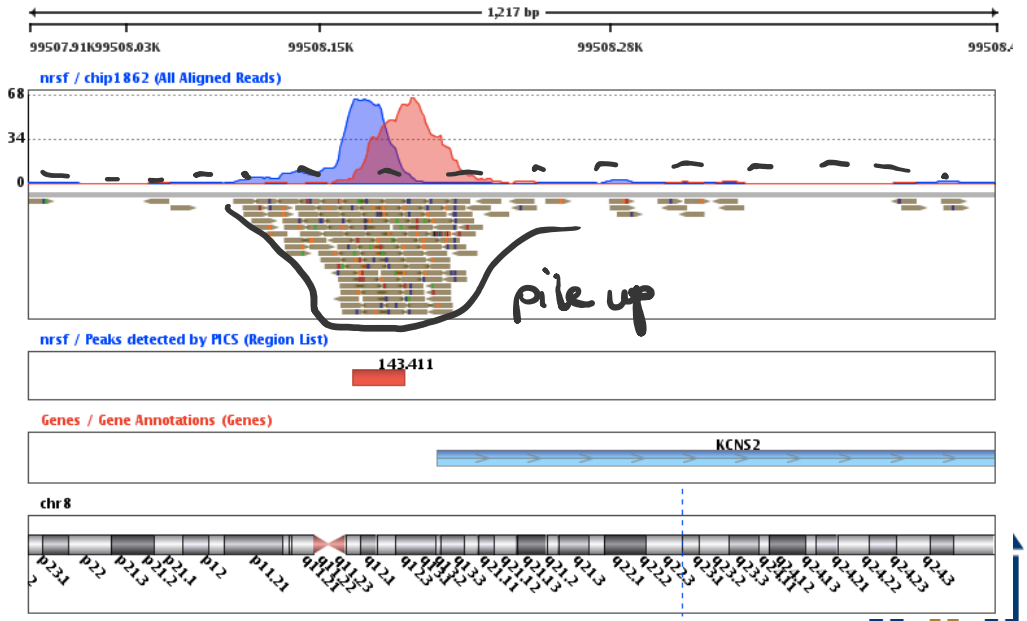
Schulz MH, et al., Bioinformatics, 2012

ChIP-seq

Chromatin ImmunoPrecipitation -seq.



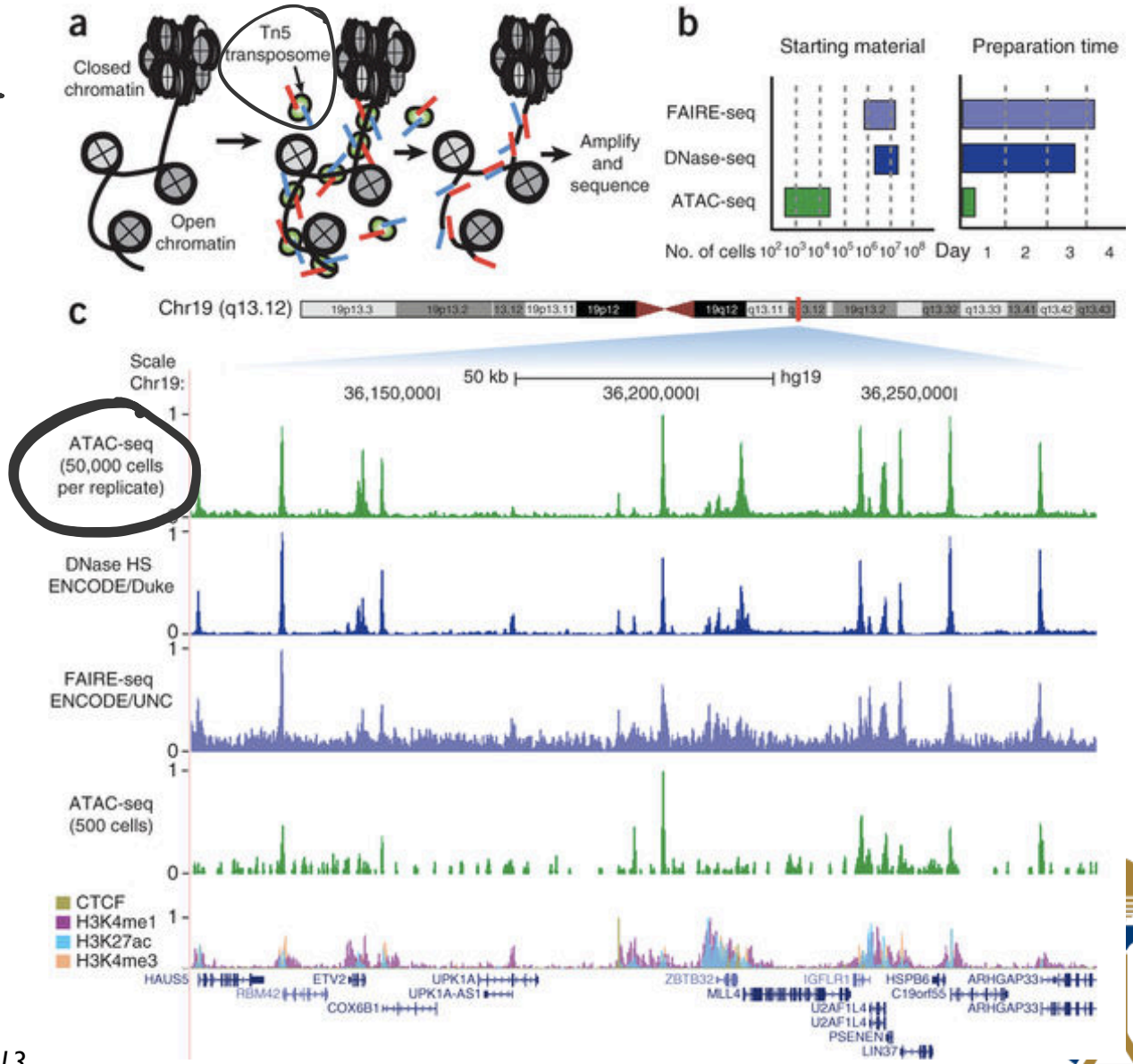
- ChIP: assesses protein-DNA interactions



ATAC-seq

Interrogates chromatin accessibility

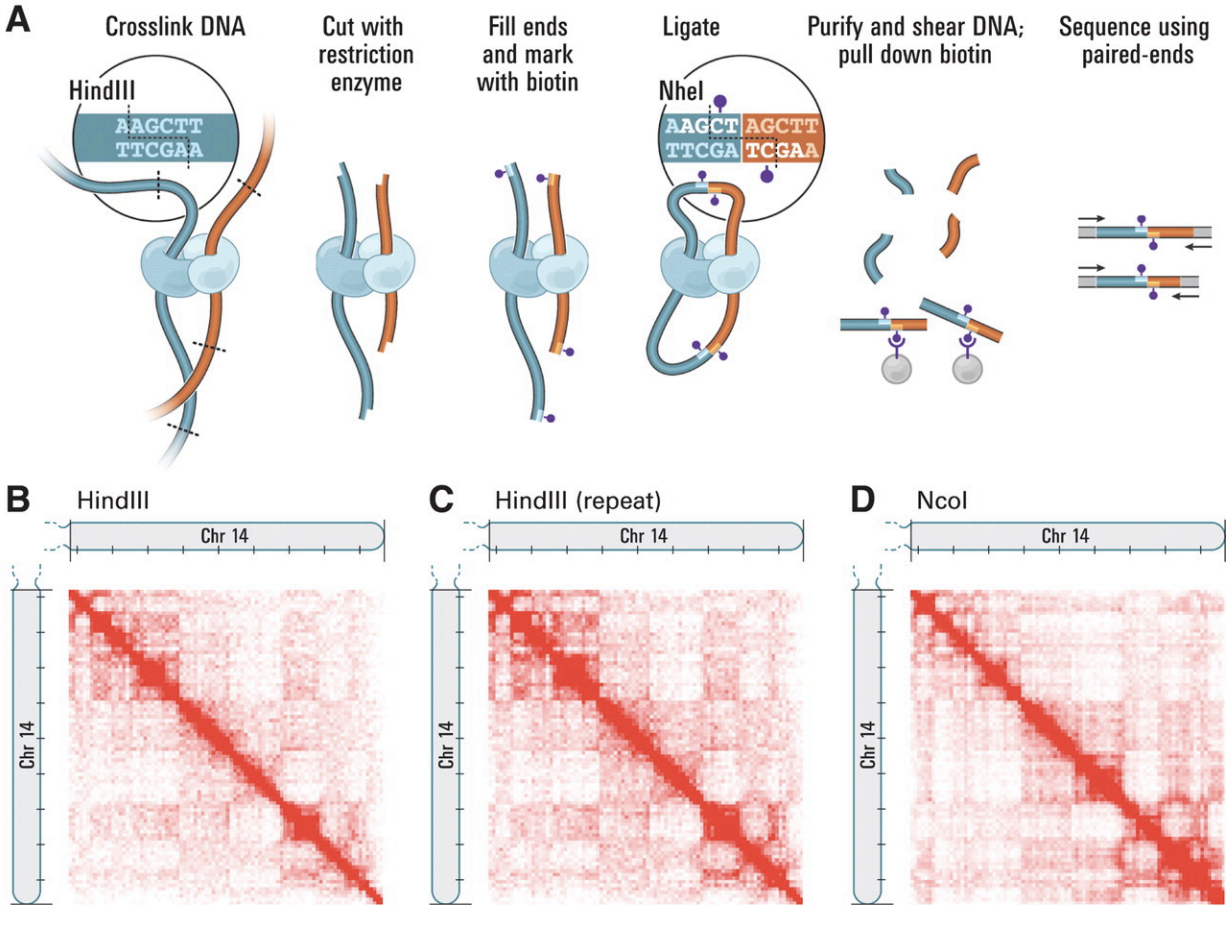
Easy to perform (compared to FAIRE-seq, MNase-seq, DNAse-seq etc.)



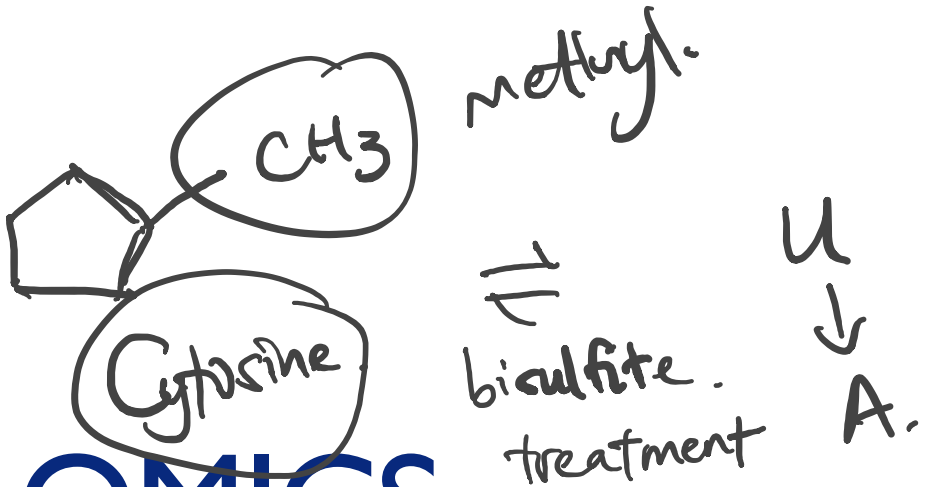
Buenrostro J.D., et al., Nature Methods, 2013

Hi-C

Probes 3D conformation of the genome architecture



Rao et al., Cell, 2014

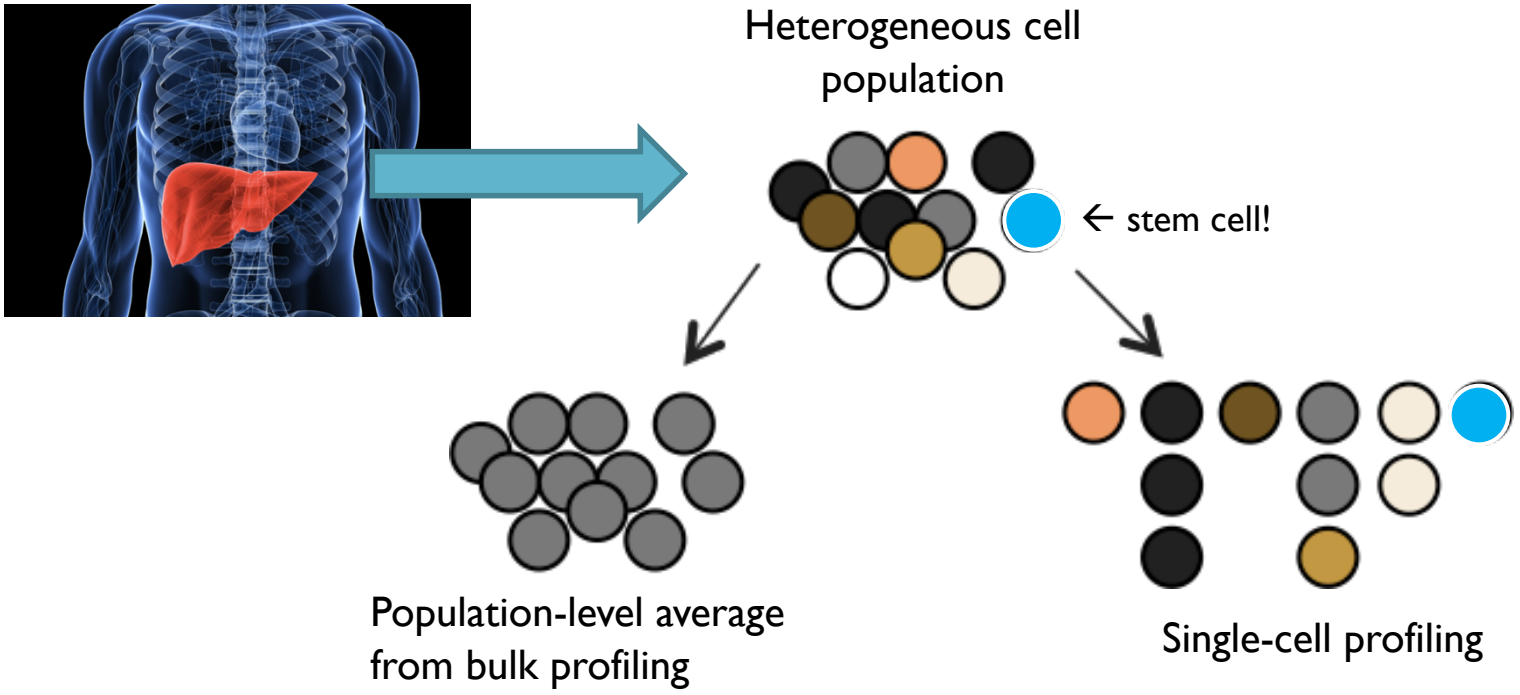


SINGLE-CELL OMICS

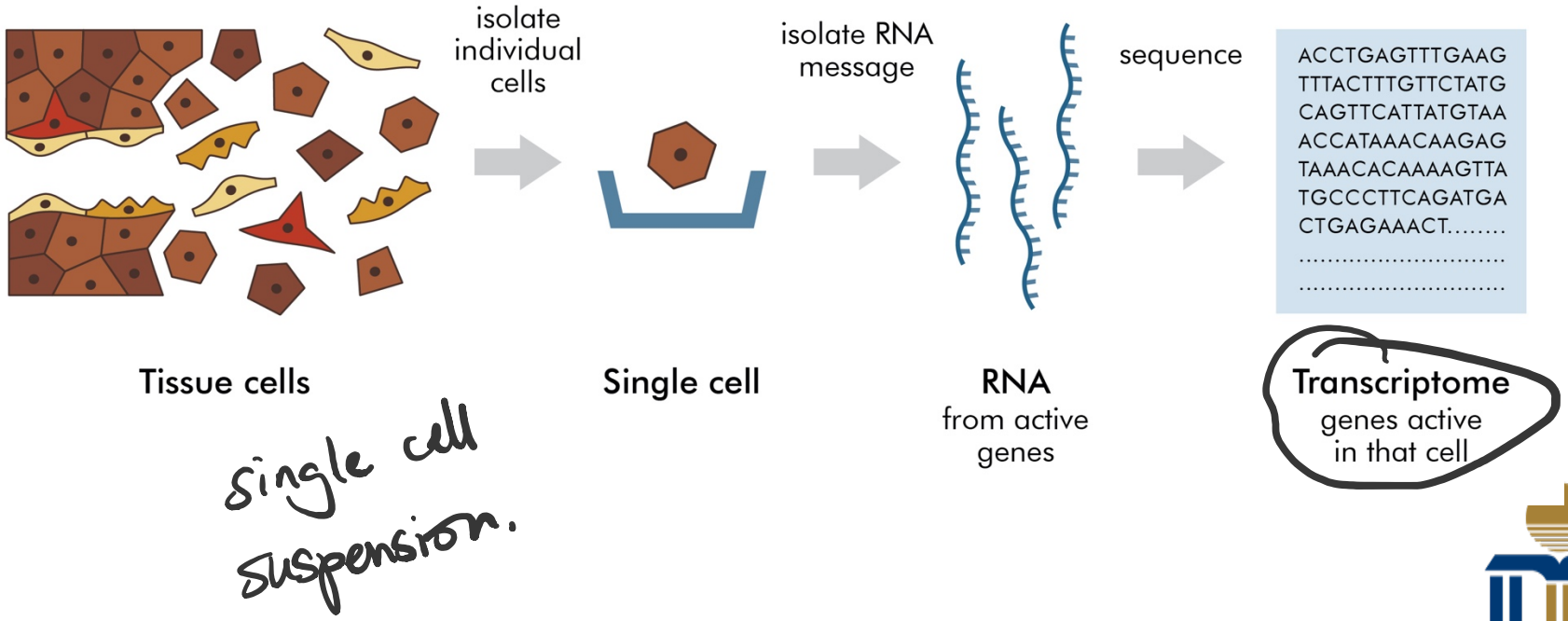
One field (among many) that is greatly enabled by microfluidic technologies

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

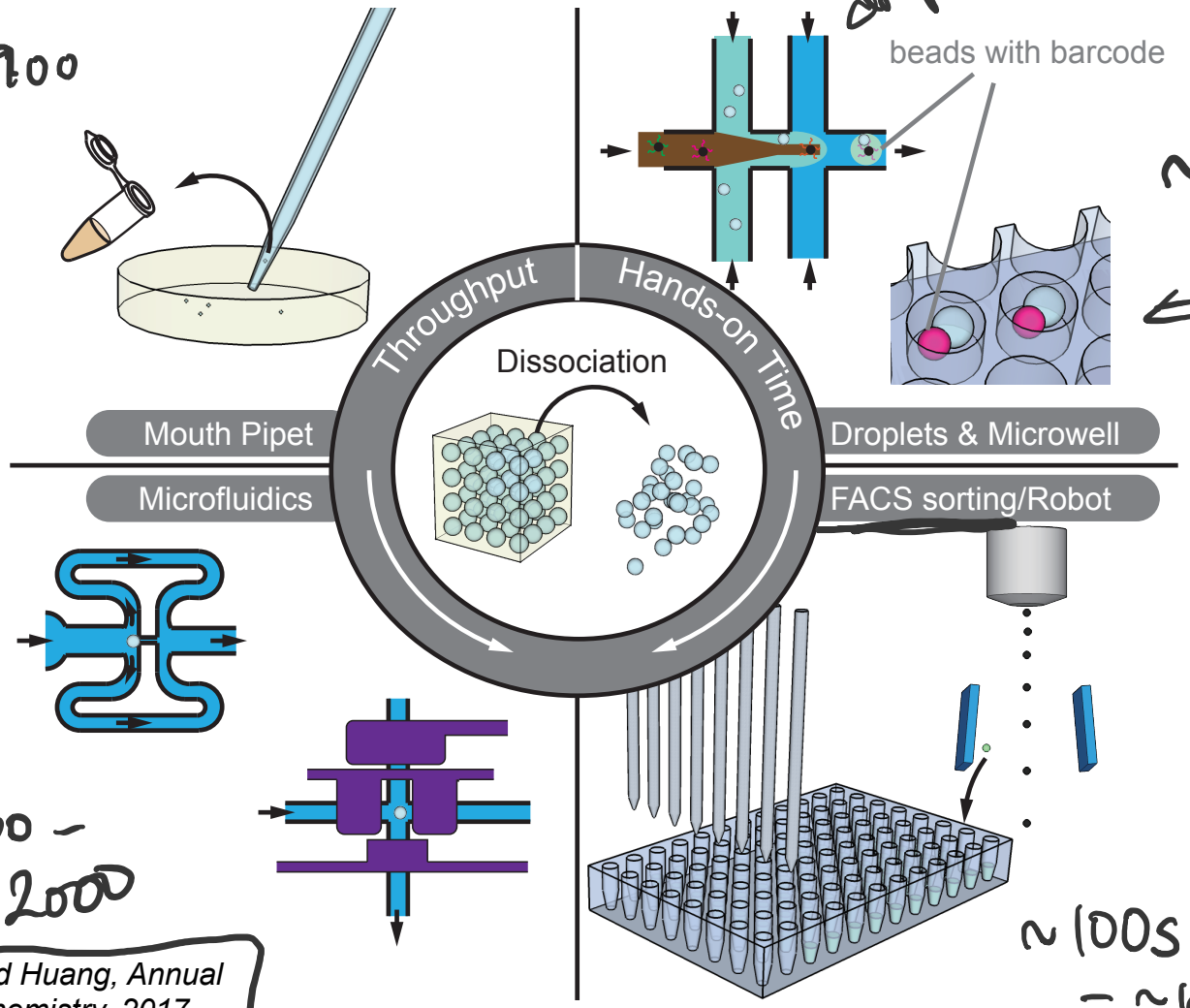


Single-cell genomics workflow



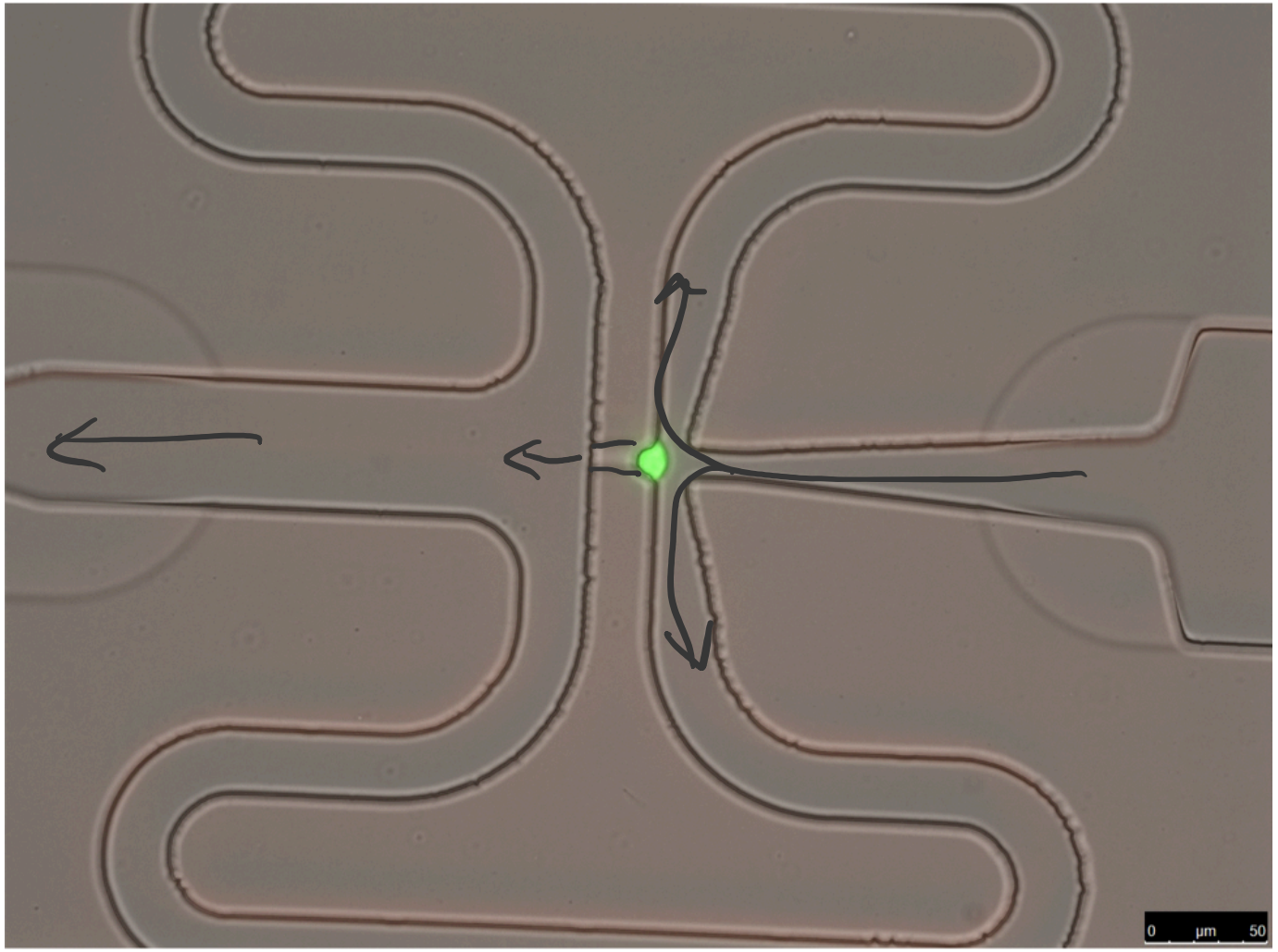
scRNAseq ~ 2009 1st POC. explosion ~2013-2017

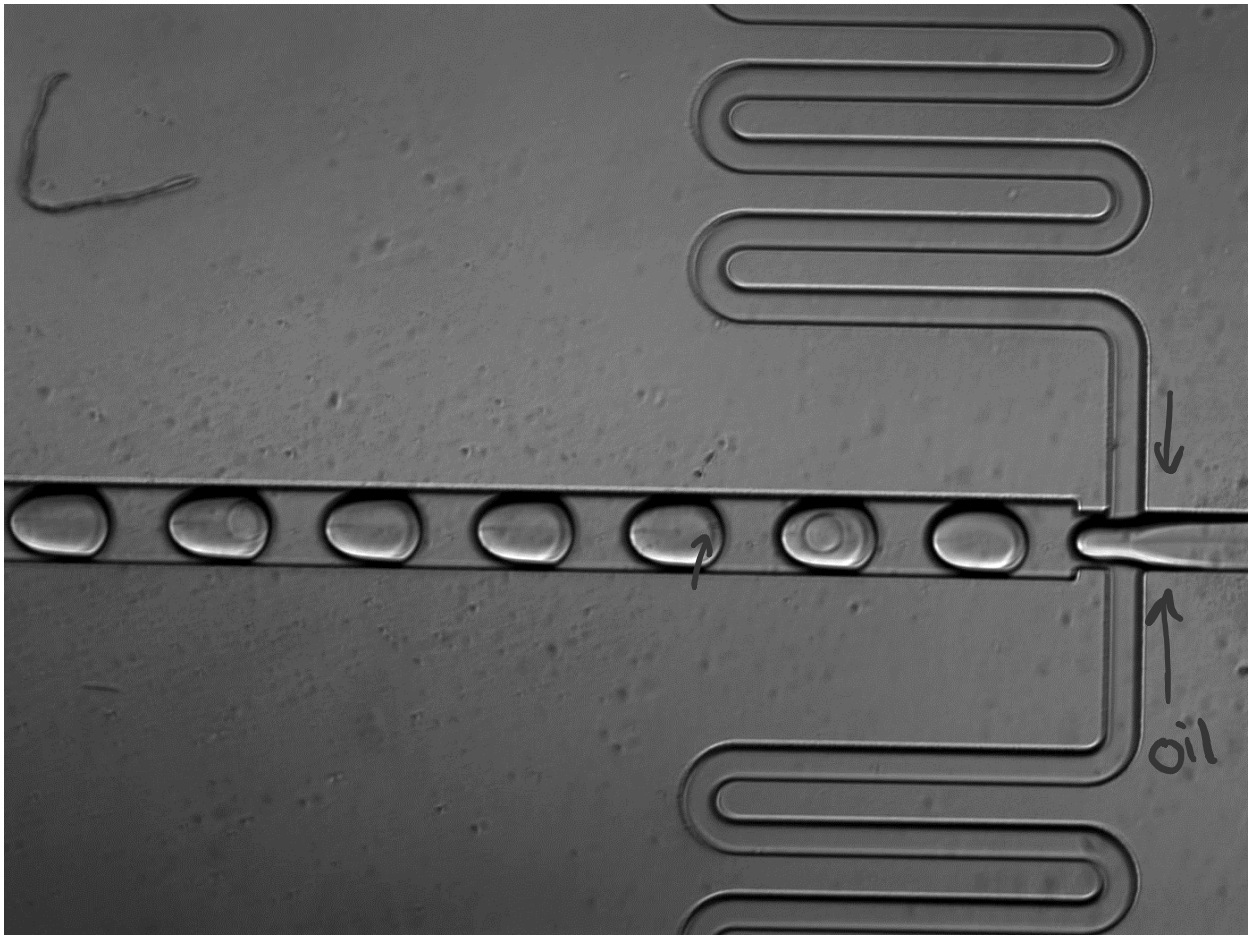
Many technology platforms to choose from



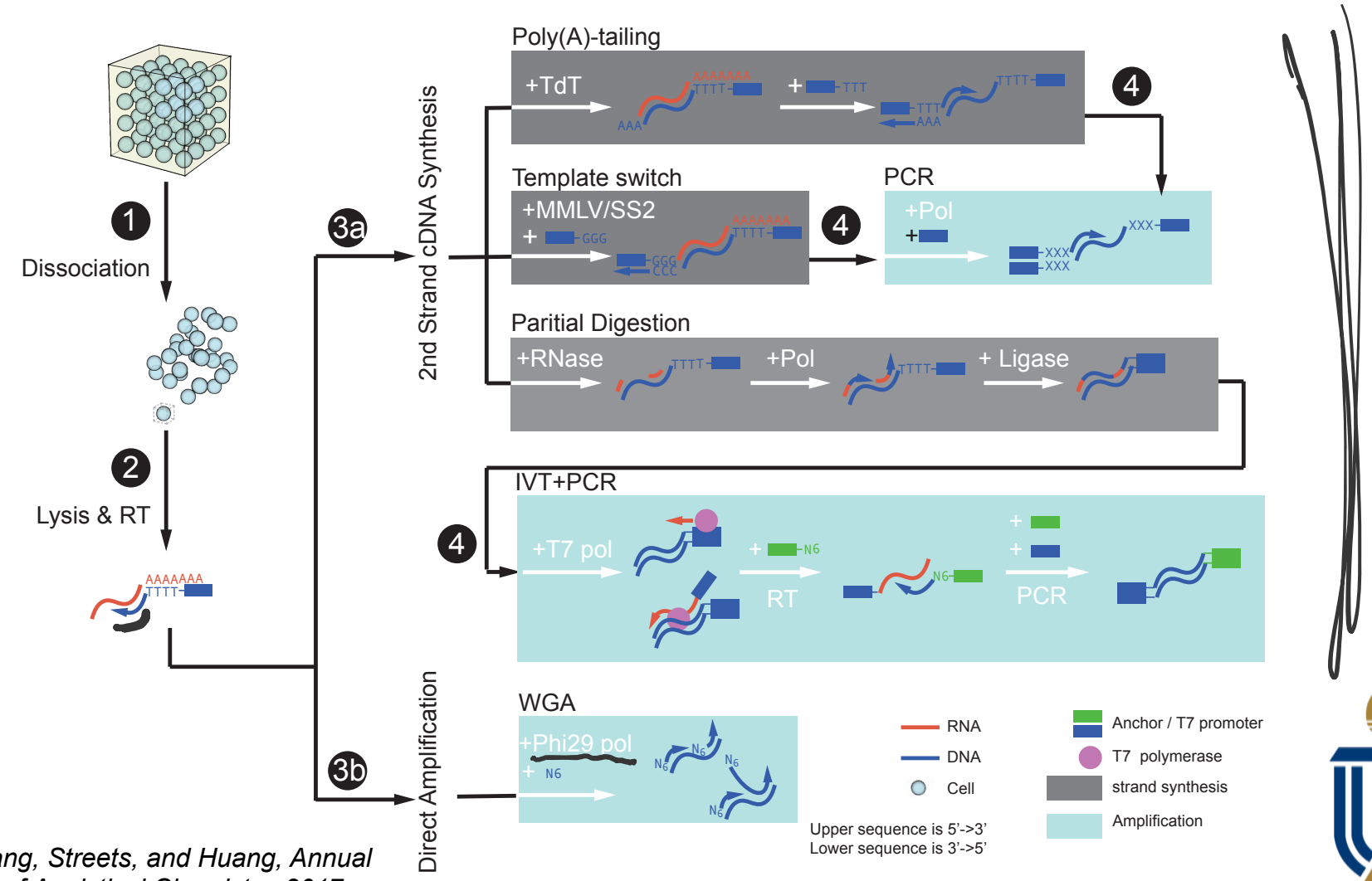
Wu, Wang, Streets, and Huang, Annual Review of Analytical Chemistry, 2017

Captured cells in the C1



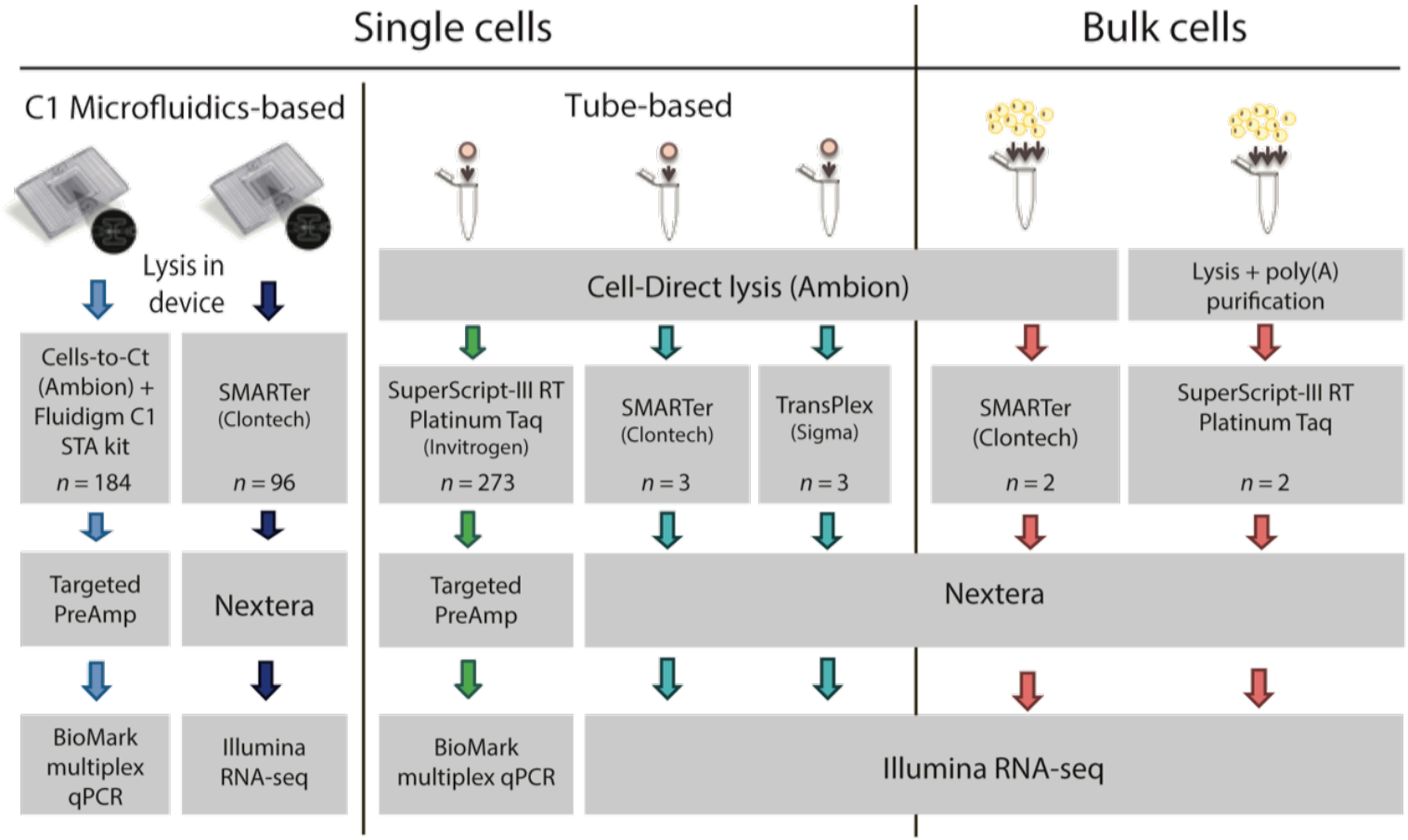

water
oil

Single cell transcriptomics methodology



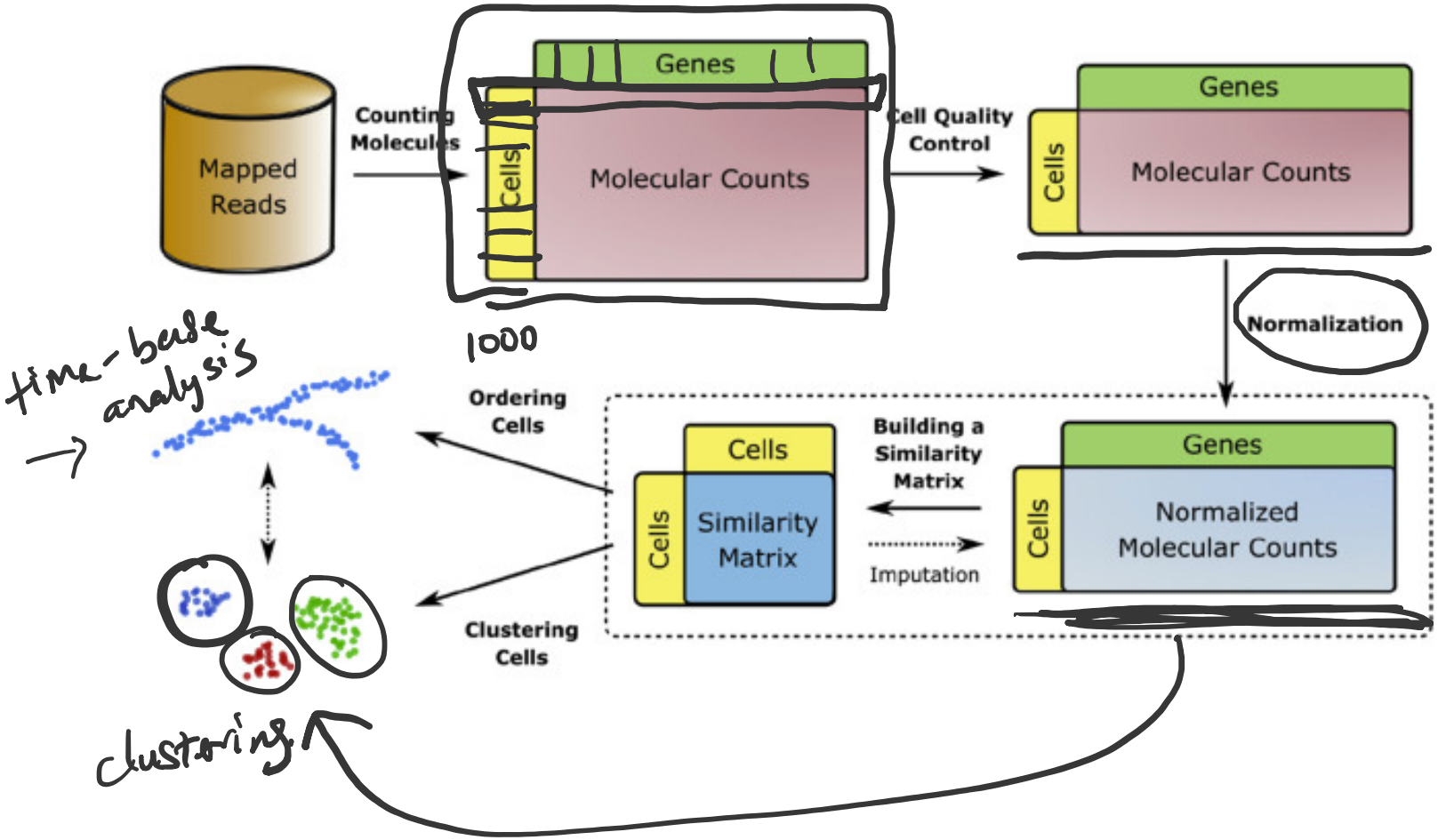
Wu, Wang, Streets, and Huang, Annual Review of Analytical Chemistry, 2017

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



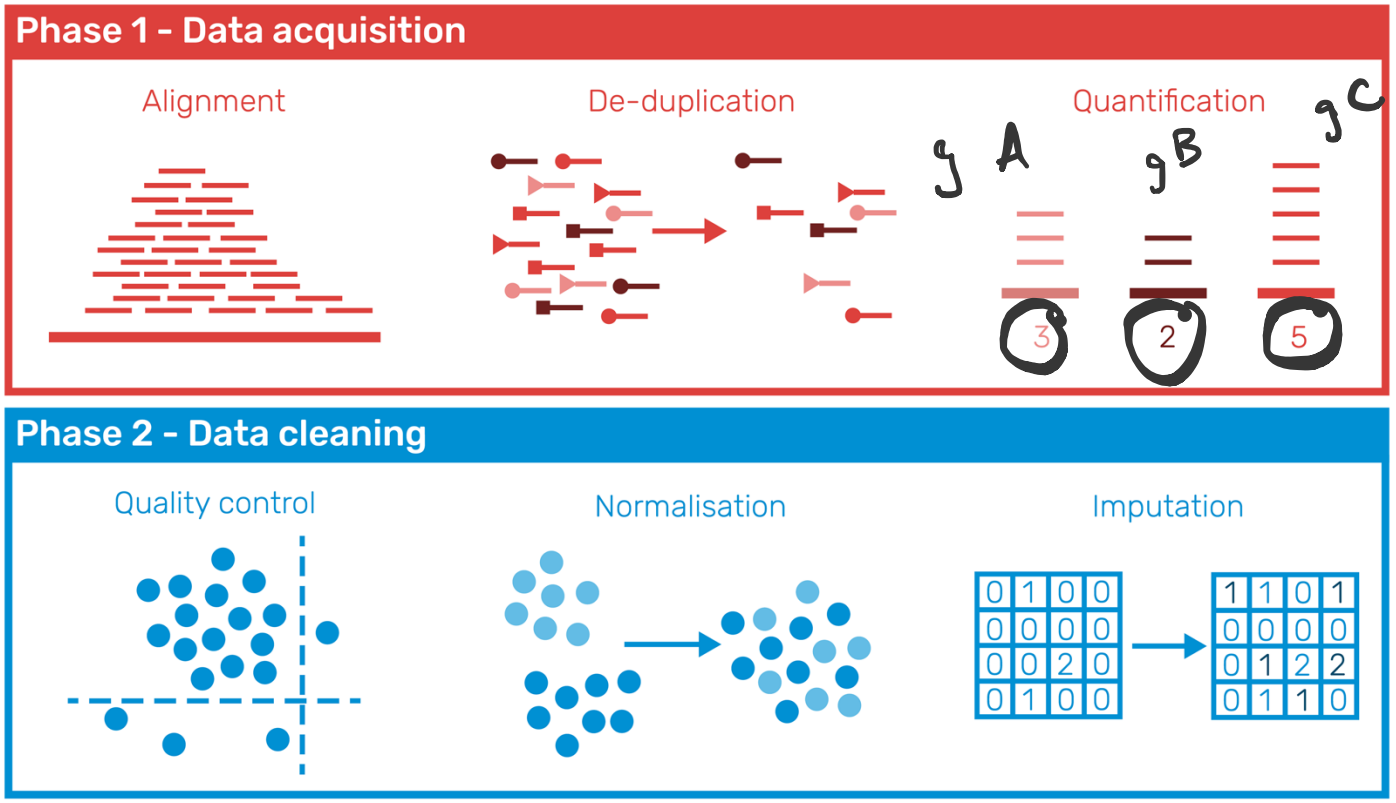
Wu, et al., 2014, Nature Methods

What does the data look like?



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

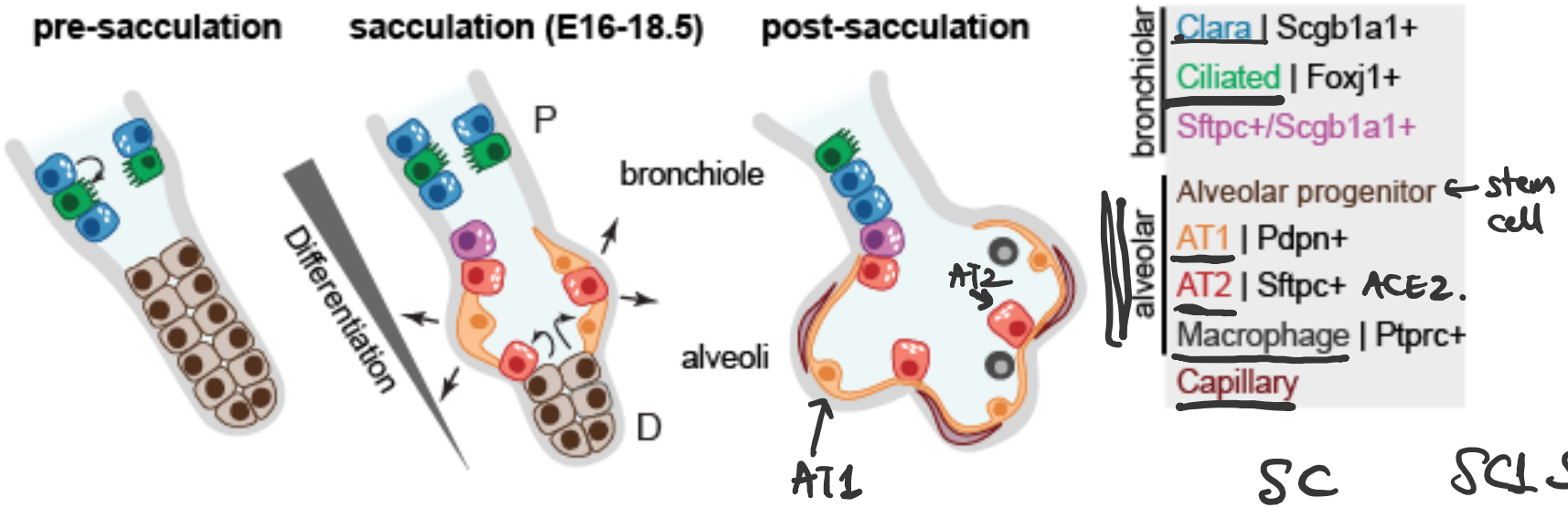
Data workflow for single-cell RNA-seq



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

Studying lung development using single-cell gene expression analysis

~300 cells/paper.
~2 million cells/paper



SC
↙ ↘
AT1 AT2

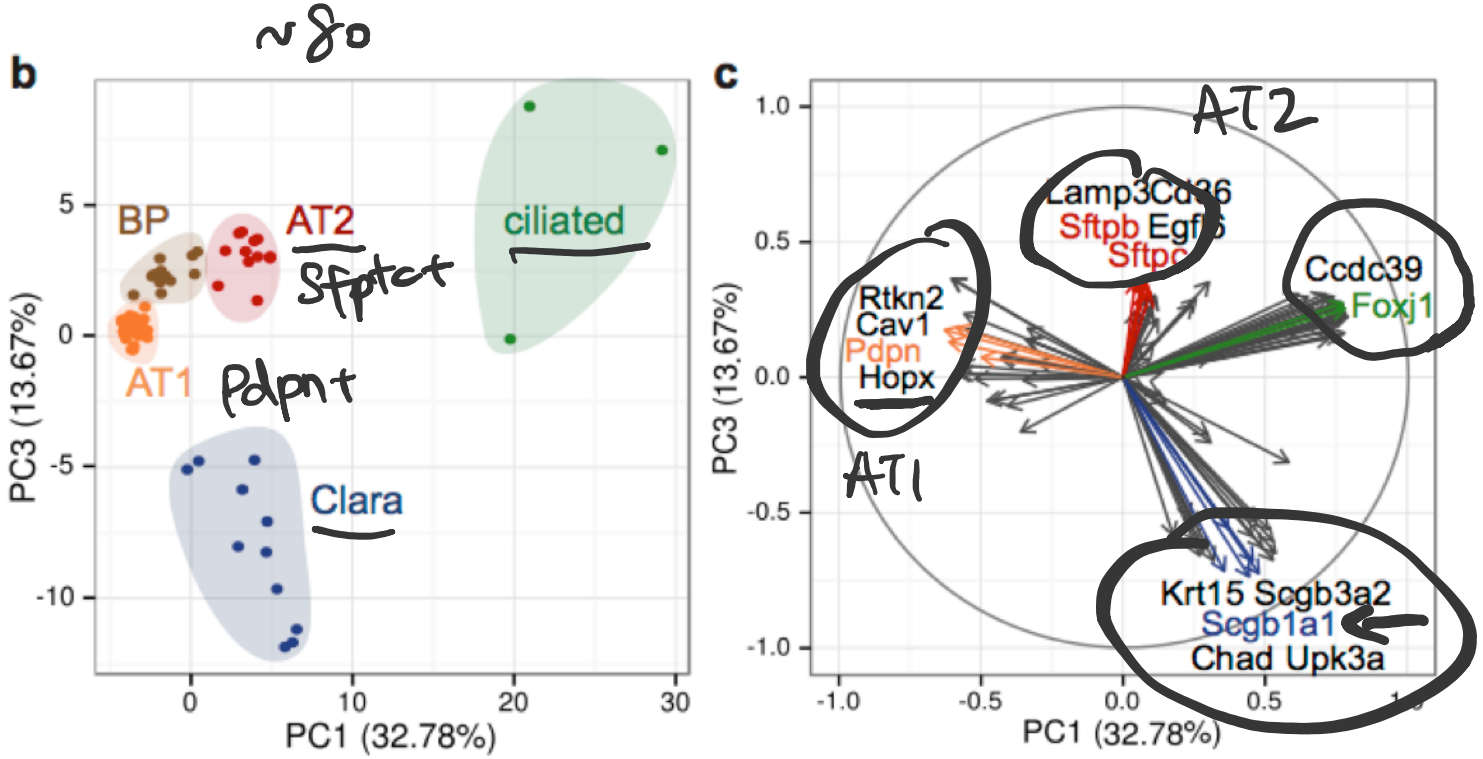
SC
↓
AT1
↓
AT2

SC1 SC2
↙ ↓
AT1 AT2

- 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*

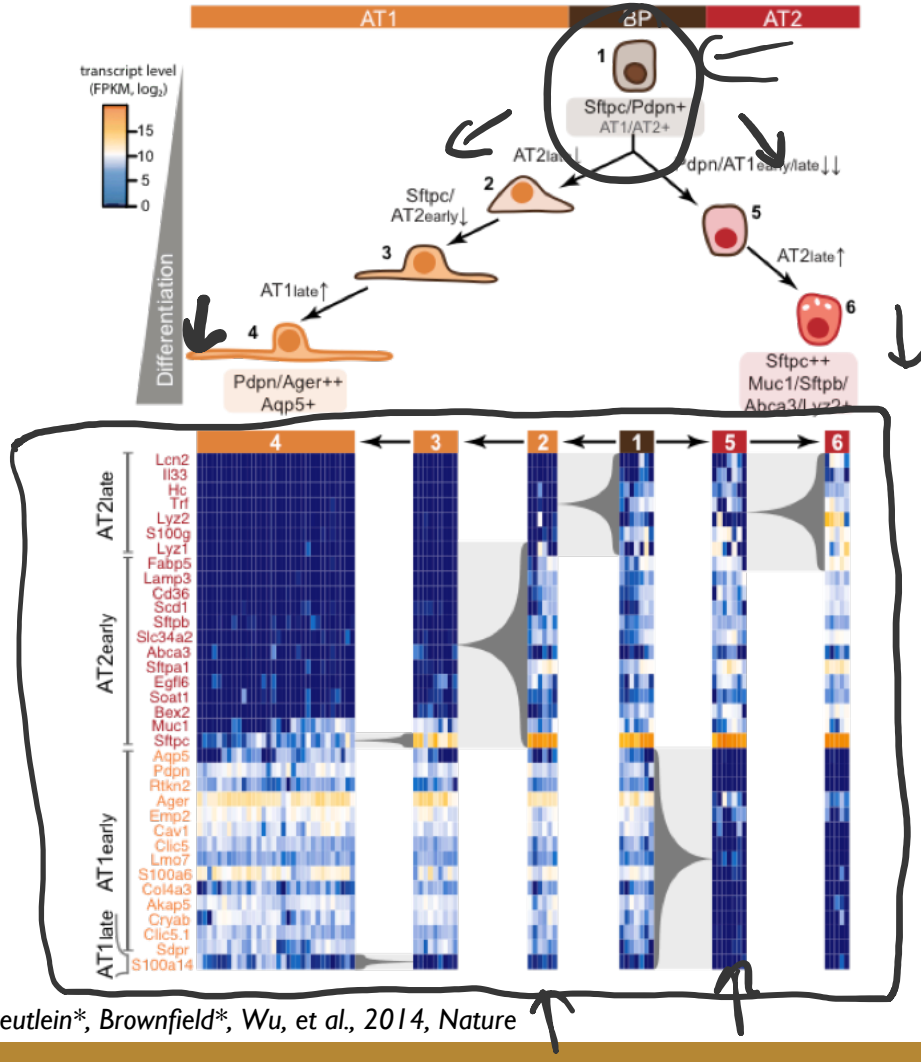
RNA-seq identifies bipotent progenitor cells in alveolar development



- 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



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

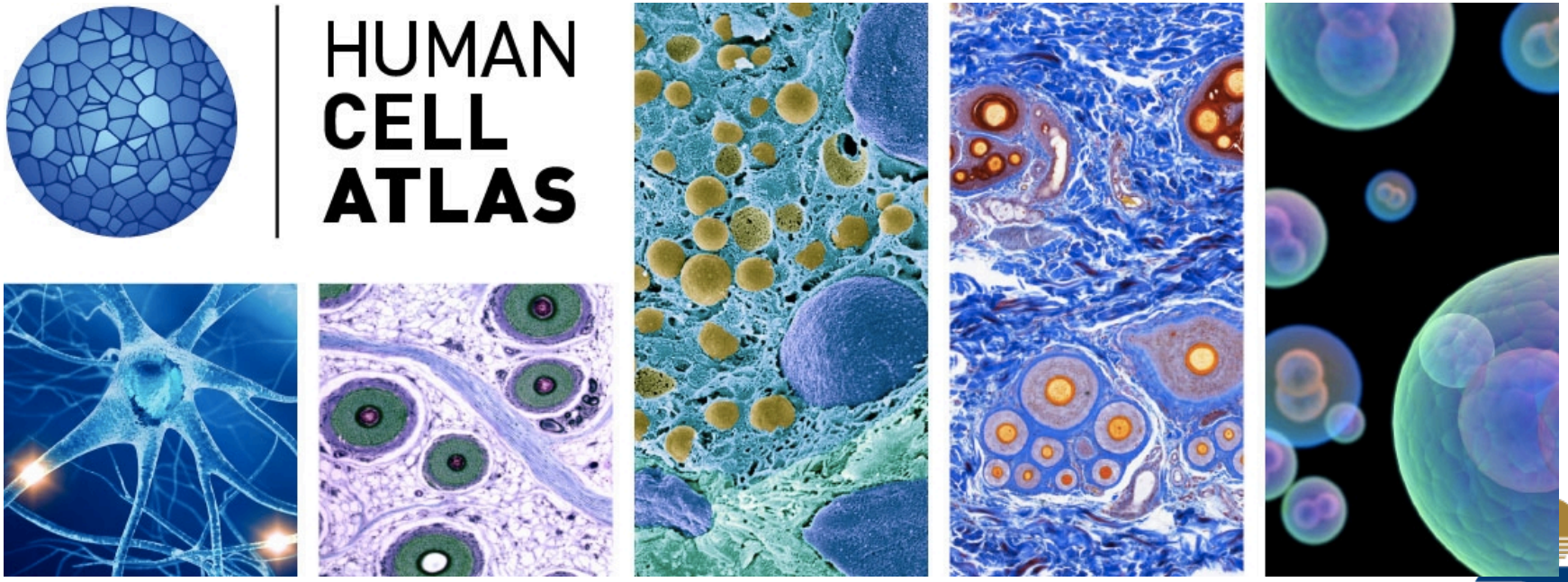
- Using genes identified in BP, AT1, and AT2, individual cells can be classified into sub-populations of intermediate cell types between BP and mature AT1 or AT2
- Reconstruction of lineage differentiation based on gene expression
- Additional support from pathway analysis

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

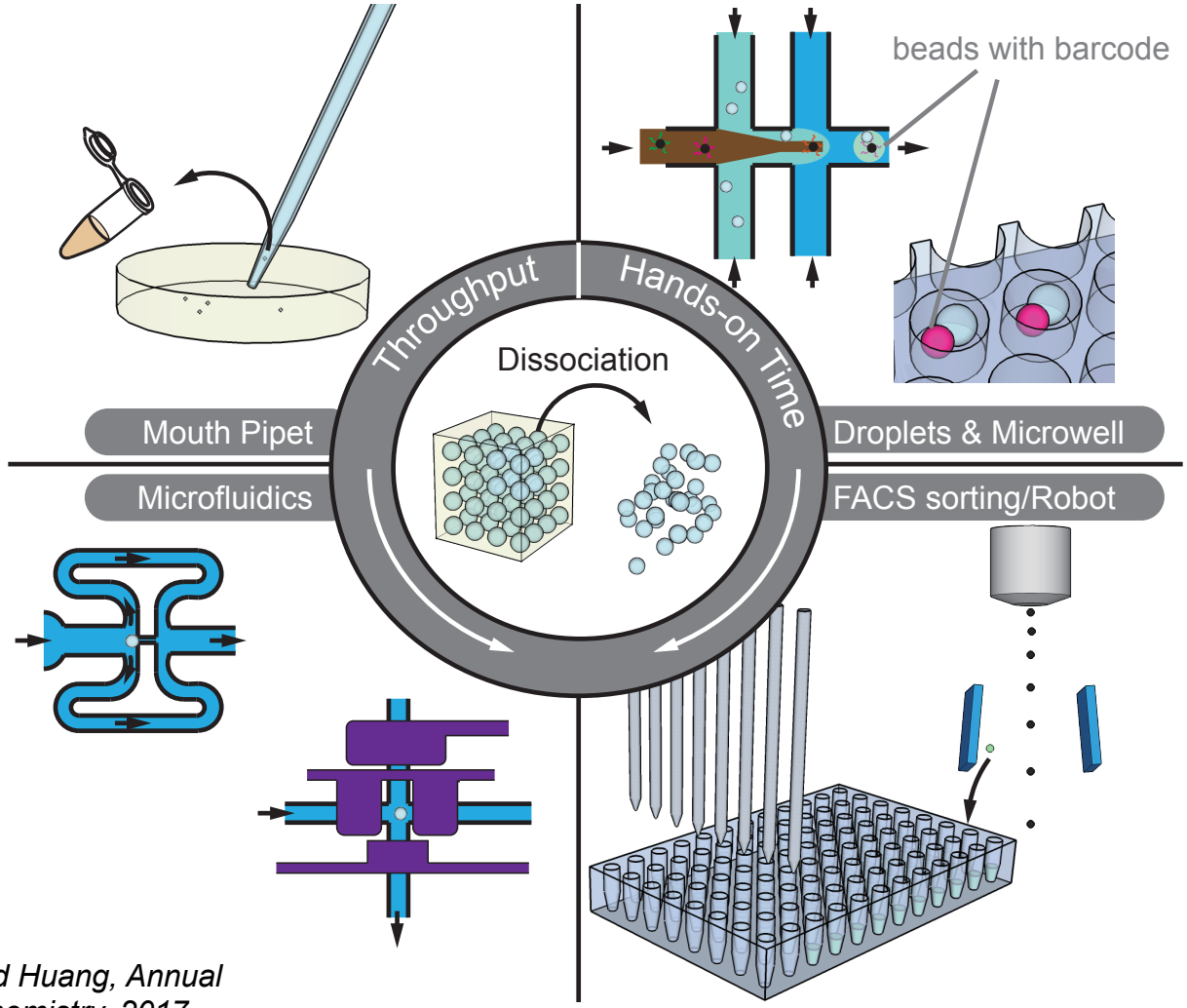
The Human Cell Atlas

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

…Can it really be done? How?



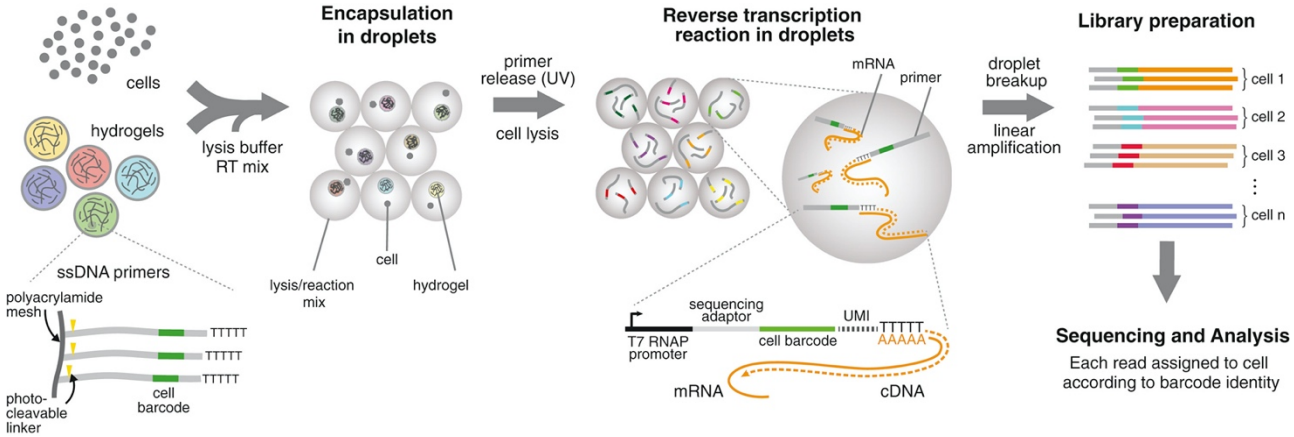
Many technology platforms to choose from

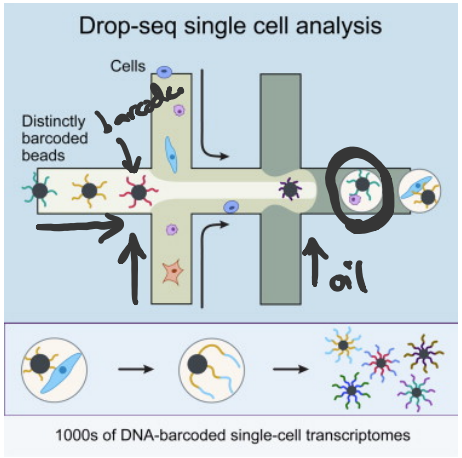


Wu, Wang, Streets, and Huang, Annual Review of Analytical Chemistry, 2017

Microfluidic droplets applied to NGS

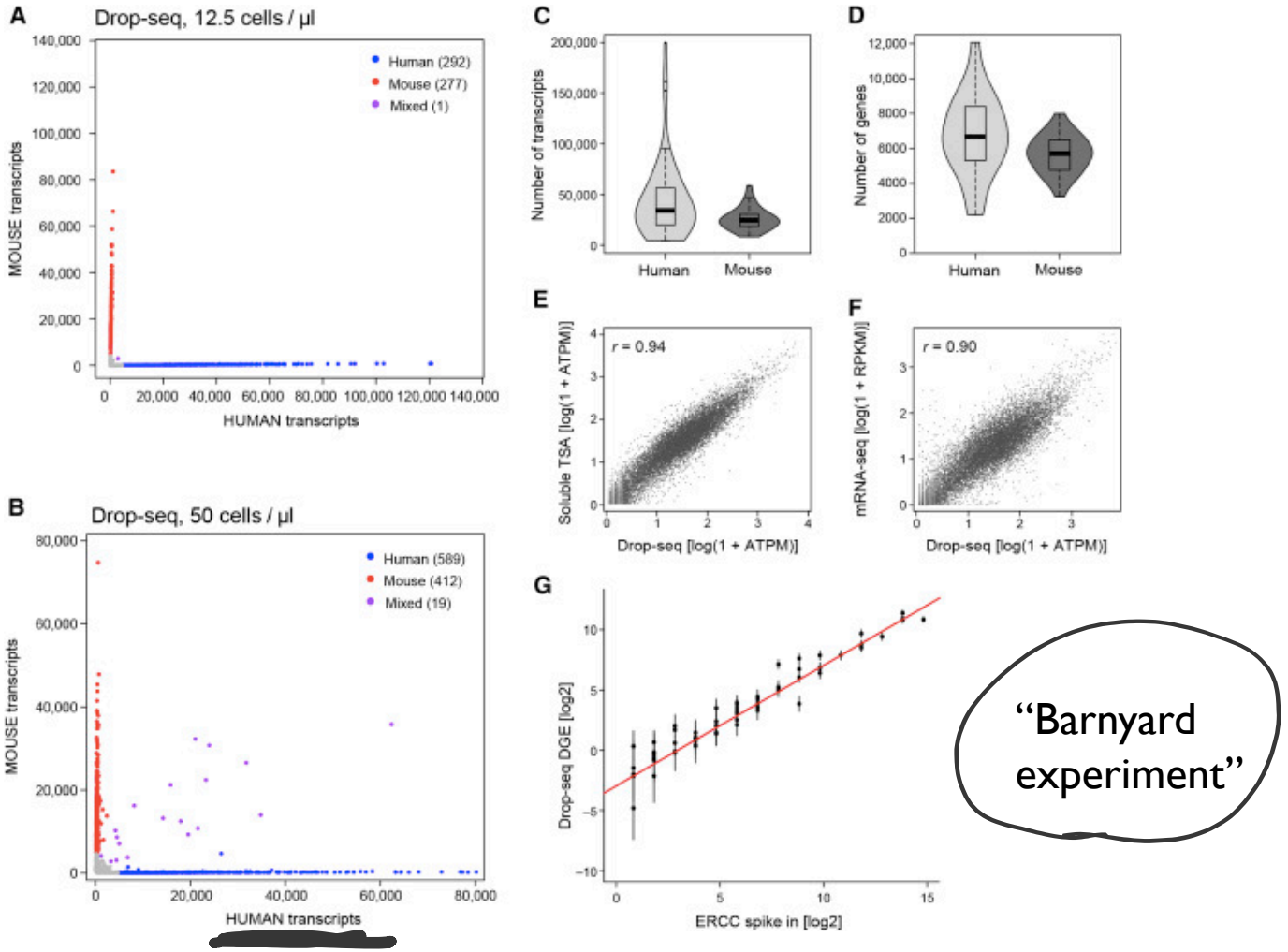
- Using droplets as chambers, we can increase throughput even more, to ~100,000 single cells per run!
- Two Harvard groups published similar technology recently:
  - Drop-seq - https://vimeo.com/128484564
  - inDrop - https://vimeo.com/126829858





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

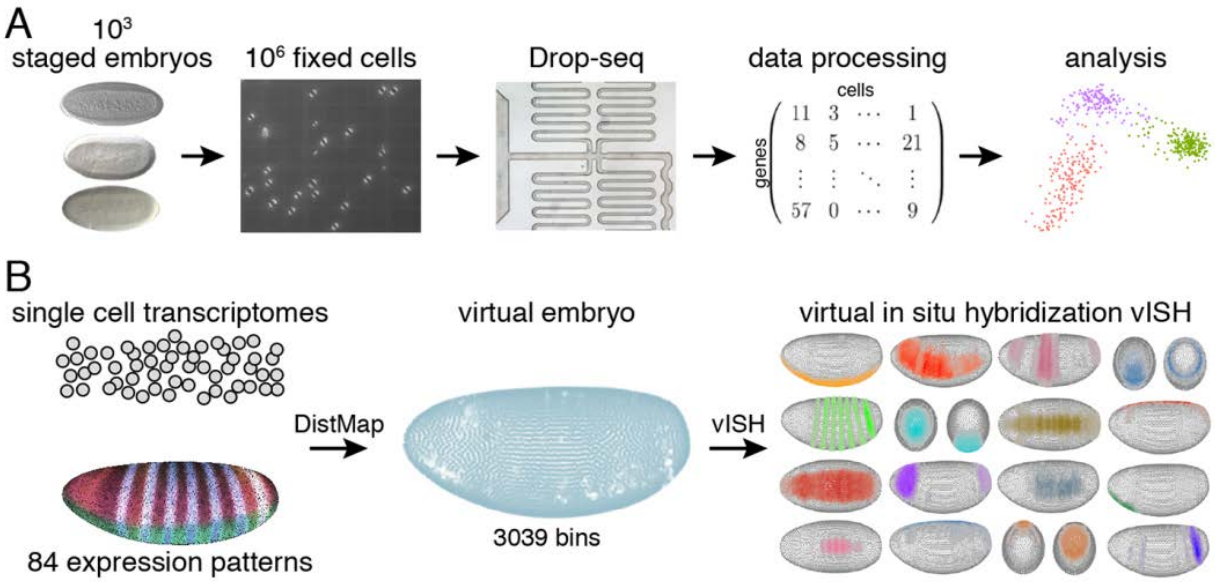
Cite as: N. Karaiskos *et al.*, *Science*
10.1126/science.aan3235 (2017).

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!

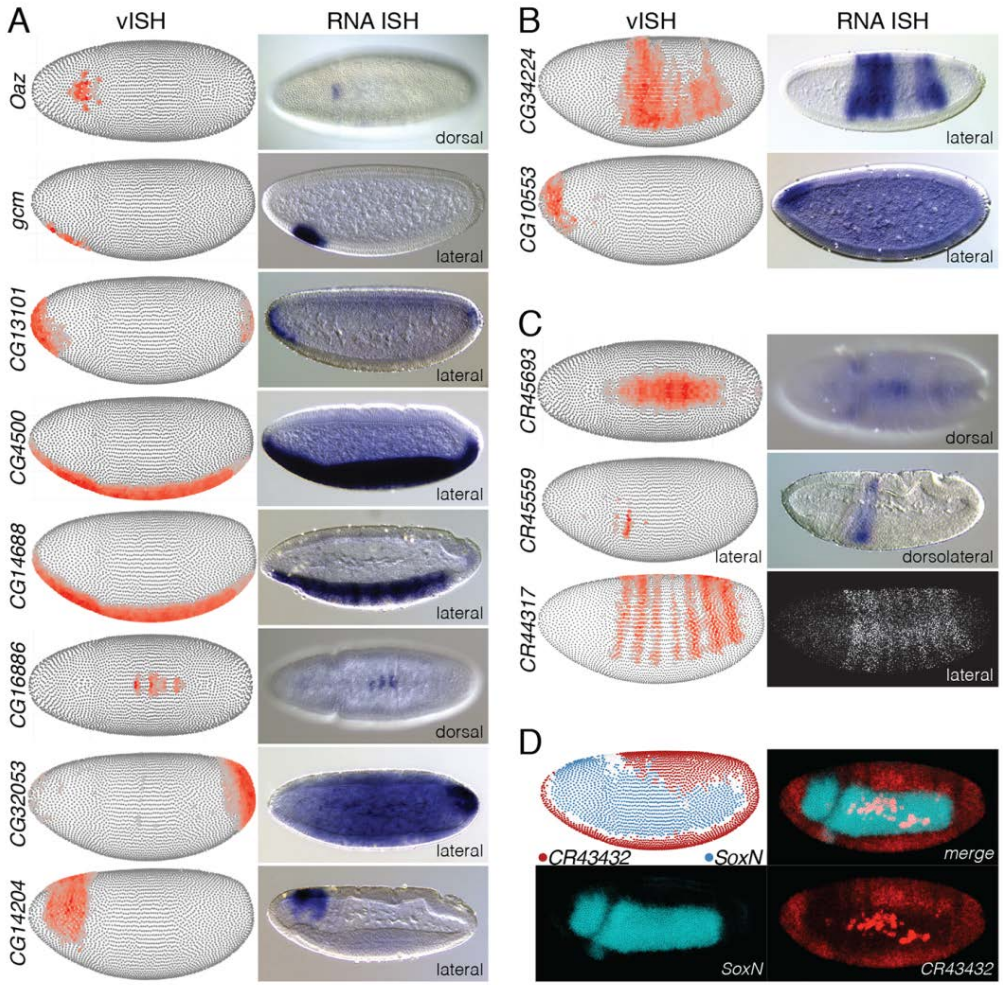


Fig. 5. Prediction accuracy and detection of new regulators. (A) vISH predictions are accurate across a wide variety of expression patterns. Expression of CGs had not been reported previously. (B) Patterned expression of putative transcription factors. (C) Patterned expression of lncRNAs. (D) *CR43432* 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

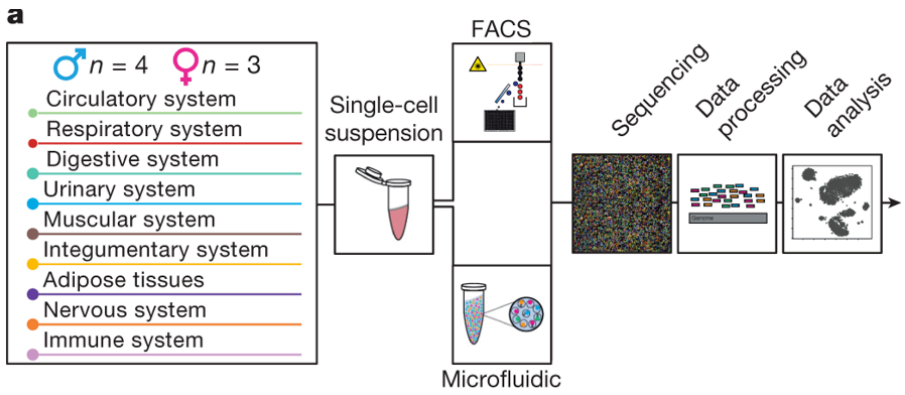
nature
International journal of science

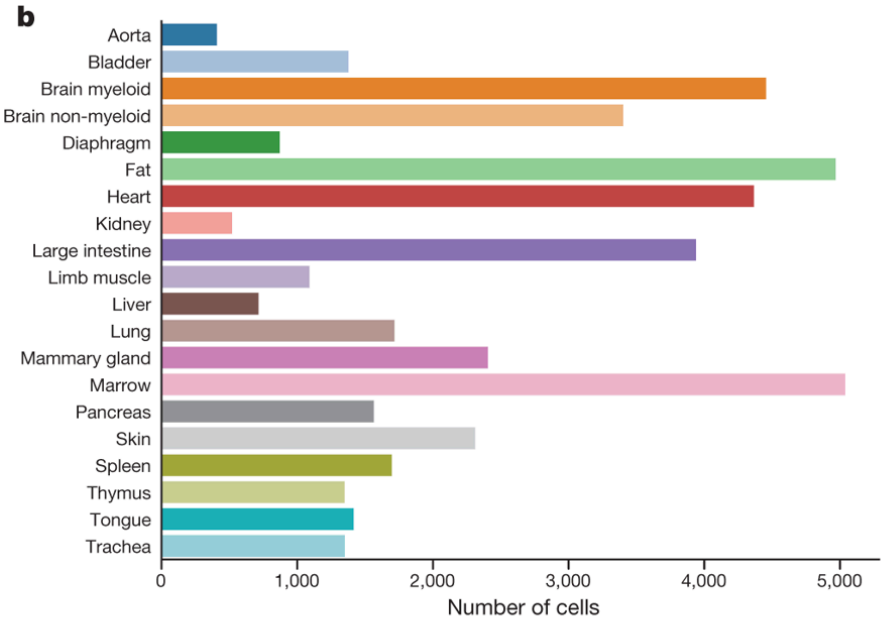
Article | Published: 03 October 2018

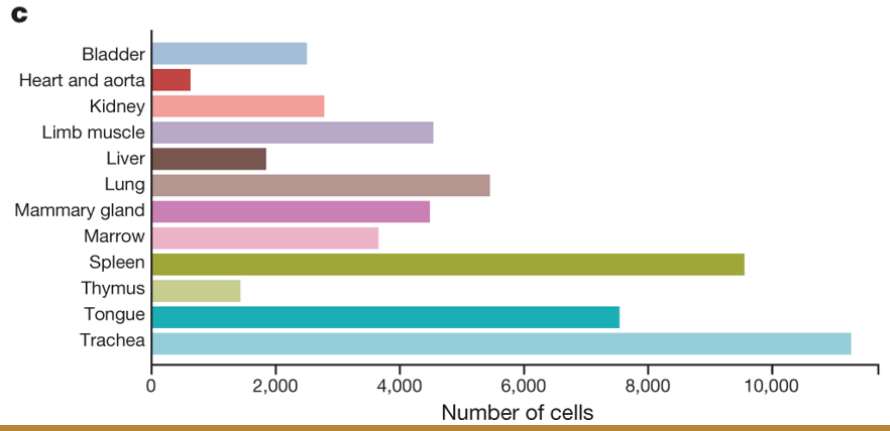
Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*

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

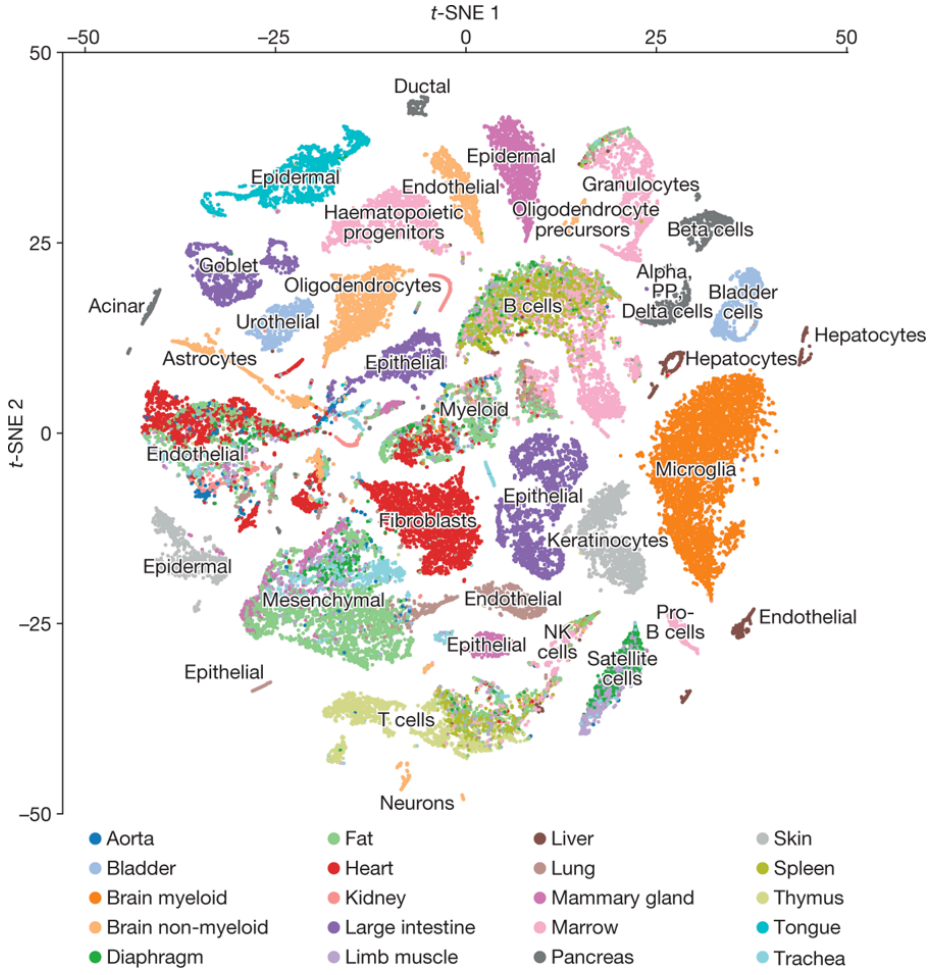






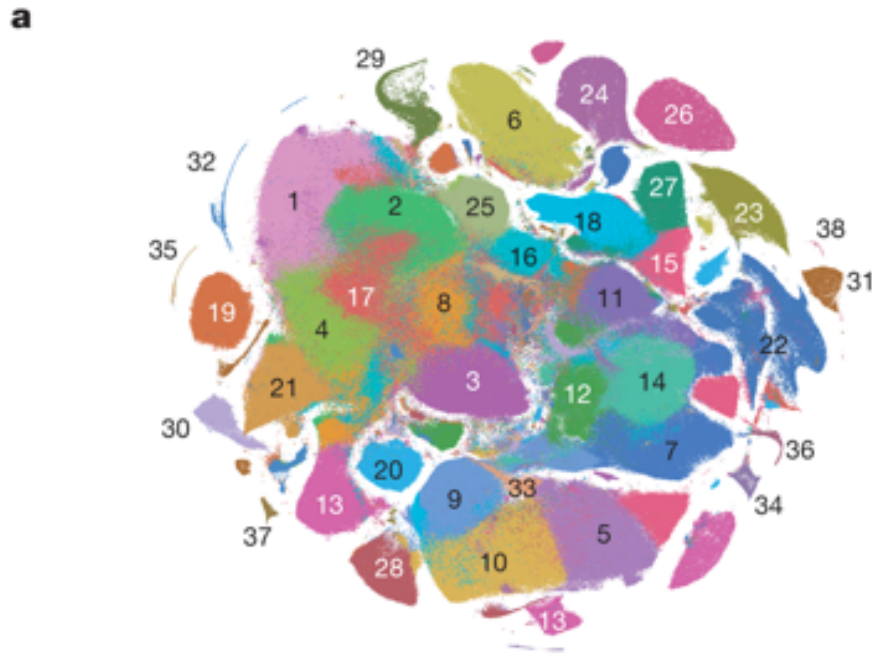
Tabula Muris Consortium. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature, 562(7727), 367.

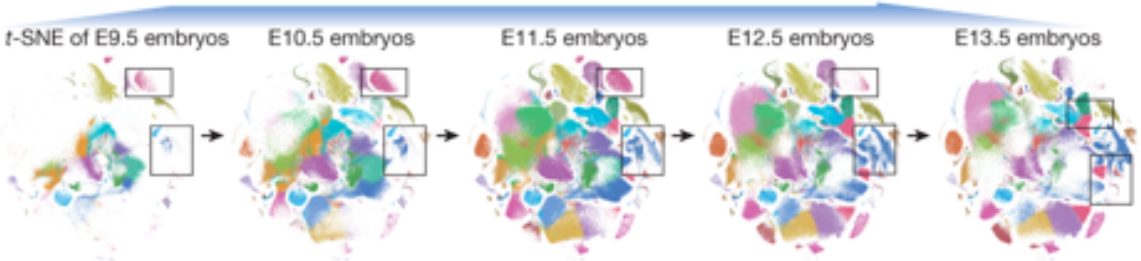
Single-cell analysis of 20 mouse tissues – mouse cell atlas



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

Mouse Lemur Atlas

nature

NEWS FEATURE · 12 JUNE 2019

Small, furry and powerful: are mouse lemurs the next big thing in genetics?

More-human than mice, the world's tiniest primates may just have what it takes to become the next top model organism.

Leslie Roberts

A mouse lemur shows its strength at a field lab in Madagascar before returning to the wild. Credit: Rijasolo/Riva Press

Onja is struggling tonight – her hands keep slipping off a miniature grip bar used to measure her strength. "Come on, you can do better," coos Zeph Pendleton, who is gently supporting the mouse lemur as she tries to get a firm hold. Finally, the animal gets her fingers around the bar and gives it a tug. It records a force of 1 kilogram, impressive for a creature weighing only 41 grams. "Good," says Pendleton, a research assistant who is working here in the rainforest at Centre ValBio, a research station at Ranomafana National Park in Madagascar.

PDF version

RELATED ARTICLES

Hacking conservation: how a tech start-up aims to save biodiversity

Monkey kingdom

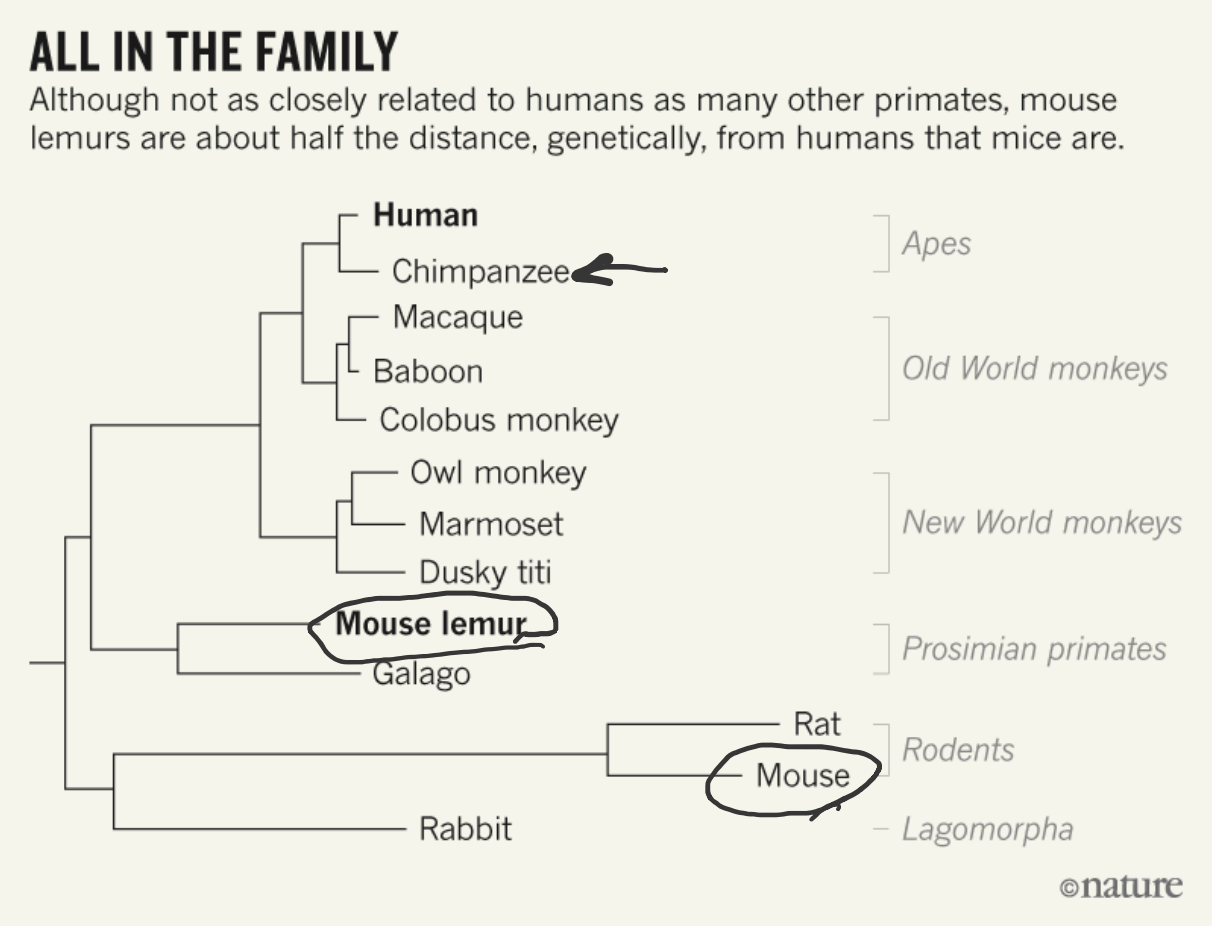
Fate of Madagascar's forests in the hands of incoming president

How to build a human cell atlas

Welcome to the CRISPR zoo

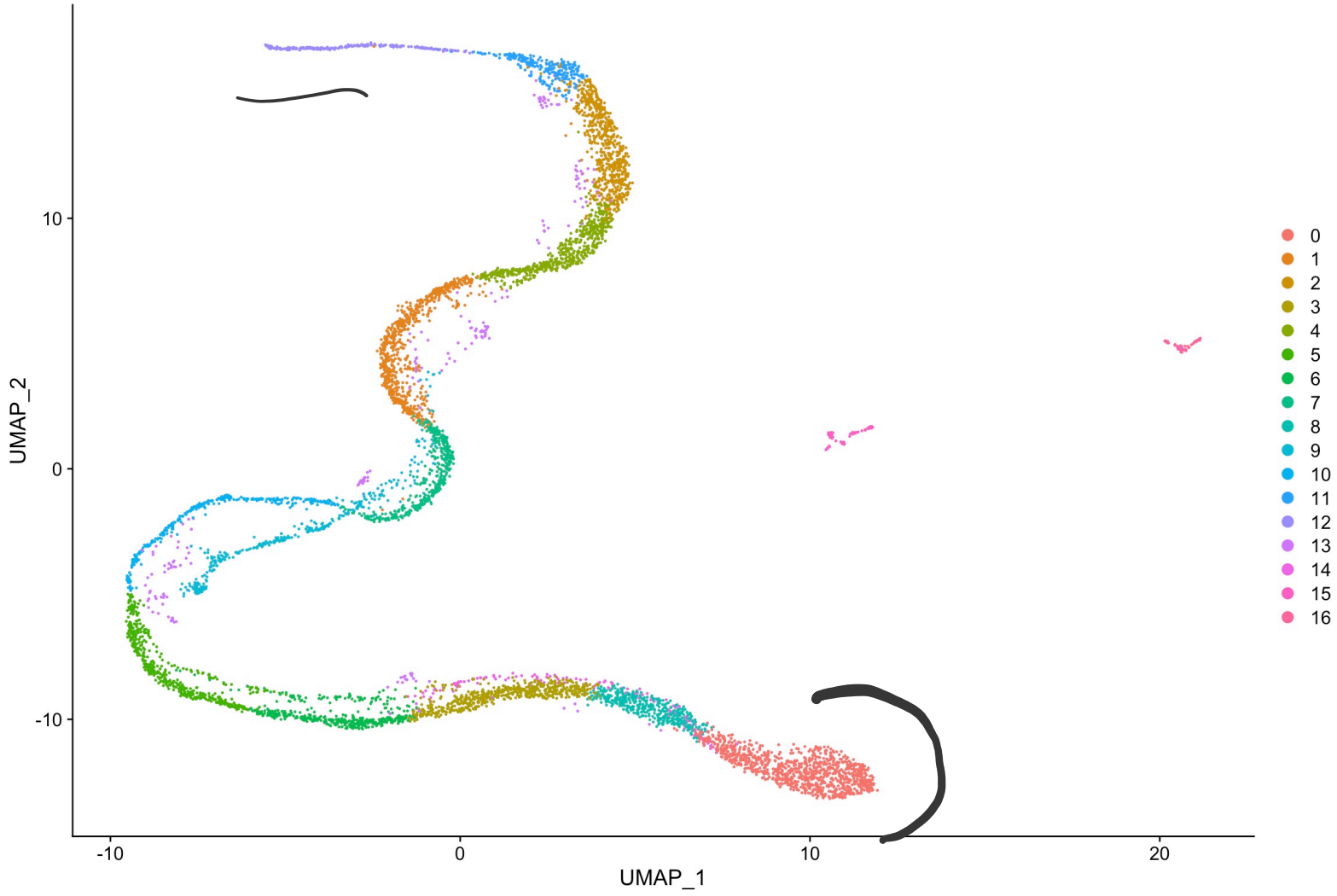
SUBJECTS

Mouse Lemur Atlas

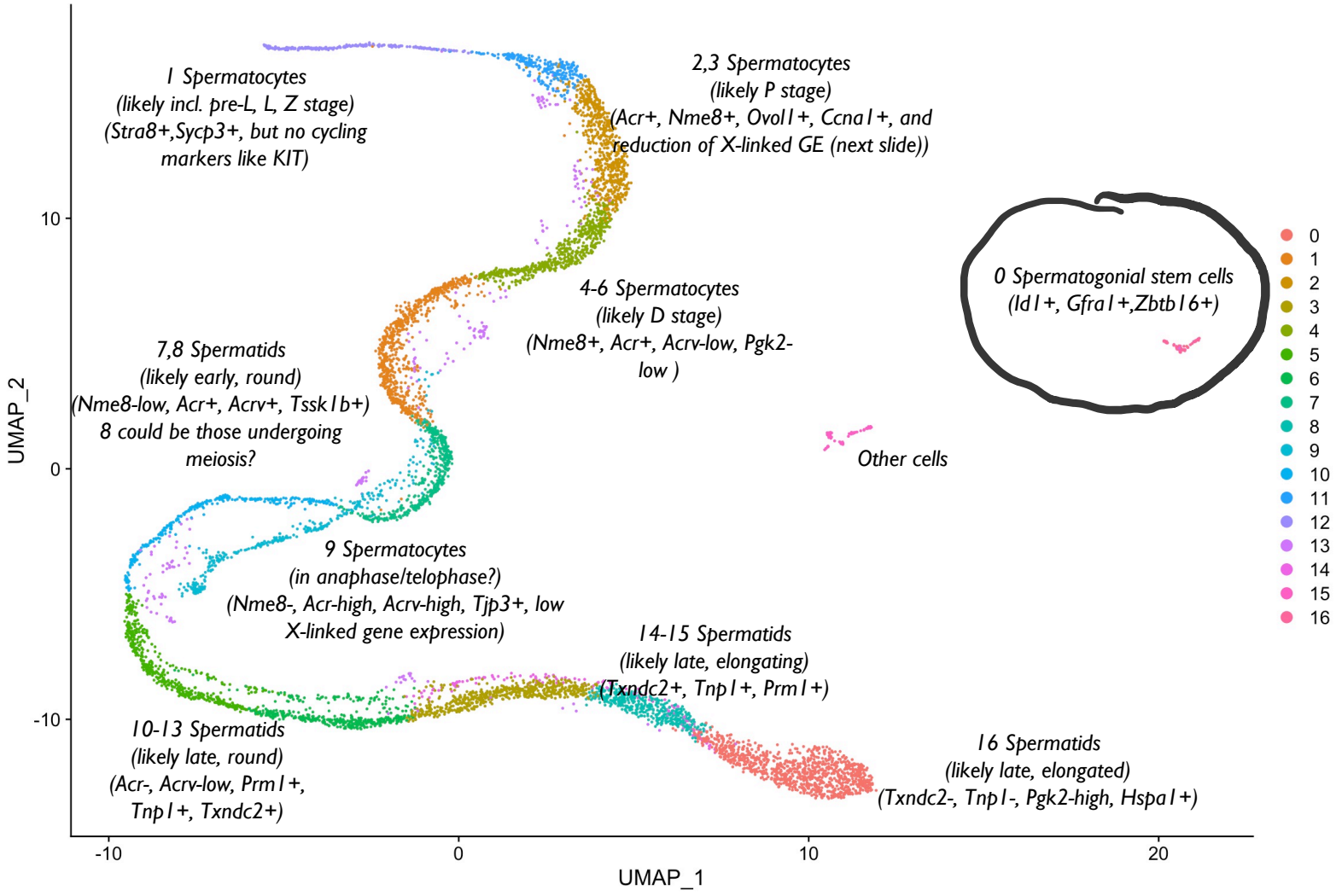


Mouse Lemur Atlas

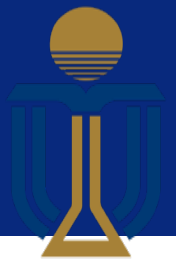
testis



Mouse Lemur Atlas

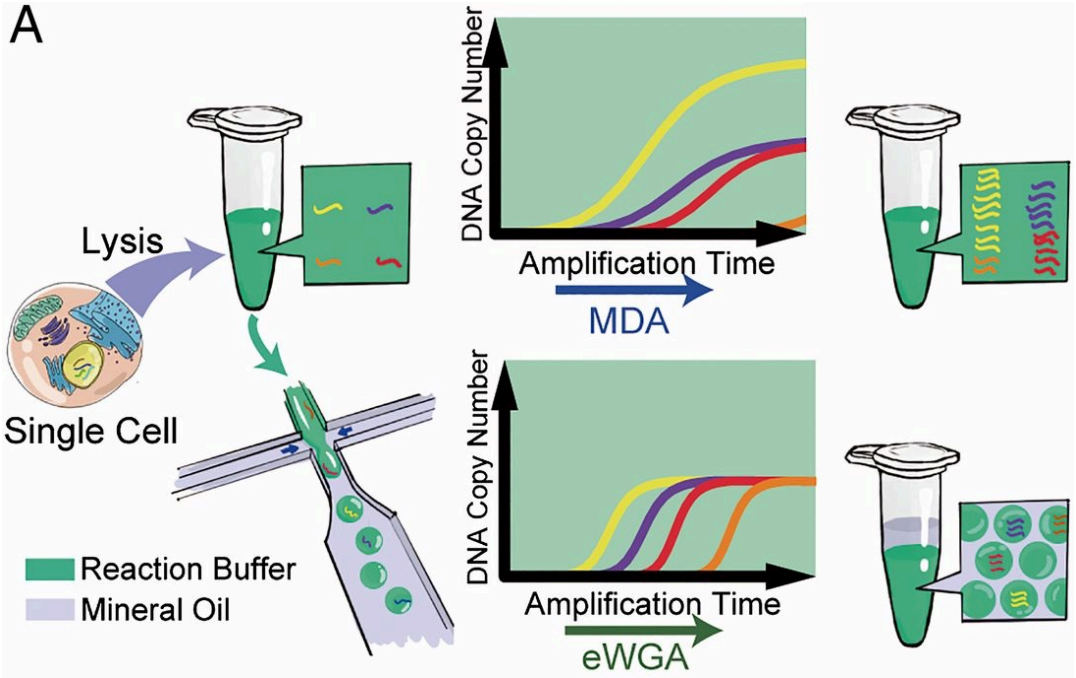


SINGLE CELL WHOLE GENOME SEQUENCE (WGS)



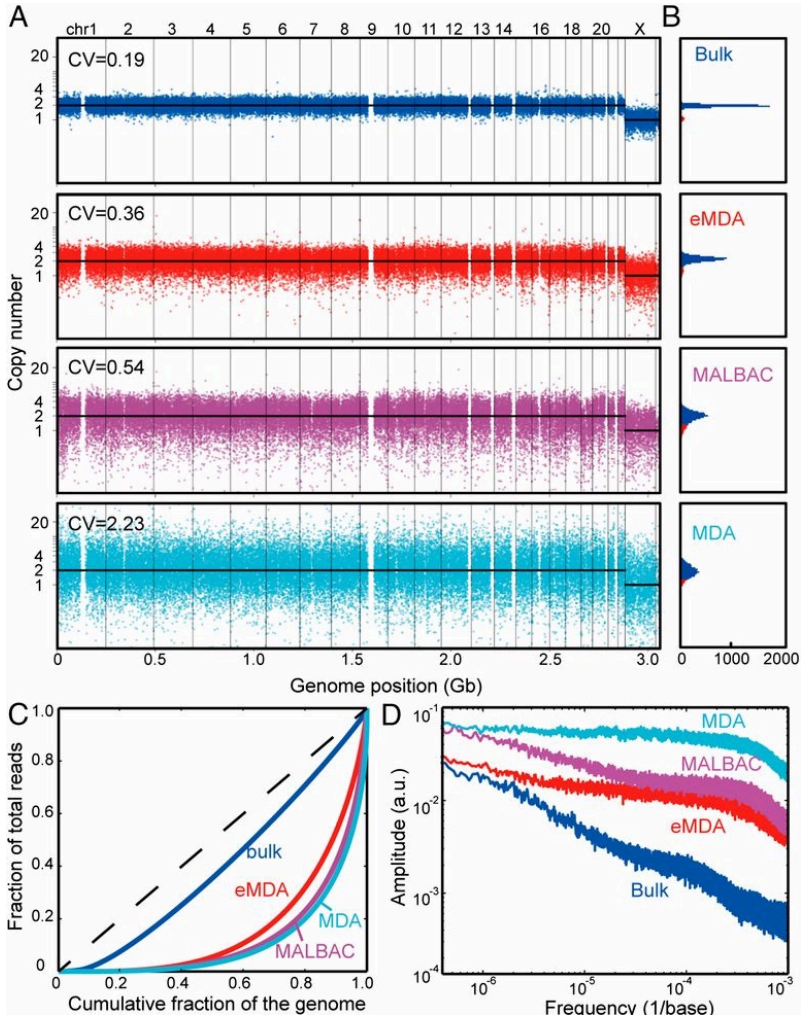
Microfluidic droplets applied to NGS

- 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

→ panel of protein.

CYTOF – cytometry and time-of-flight

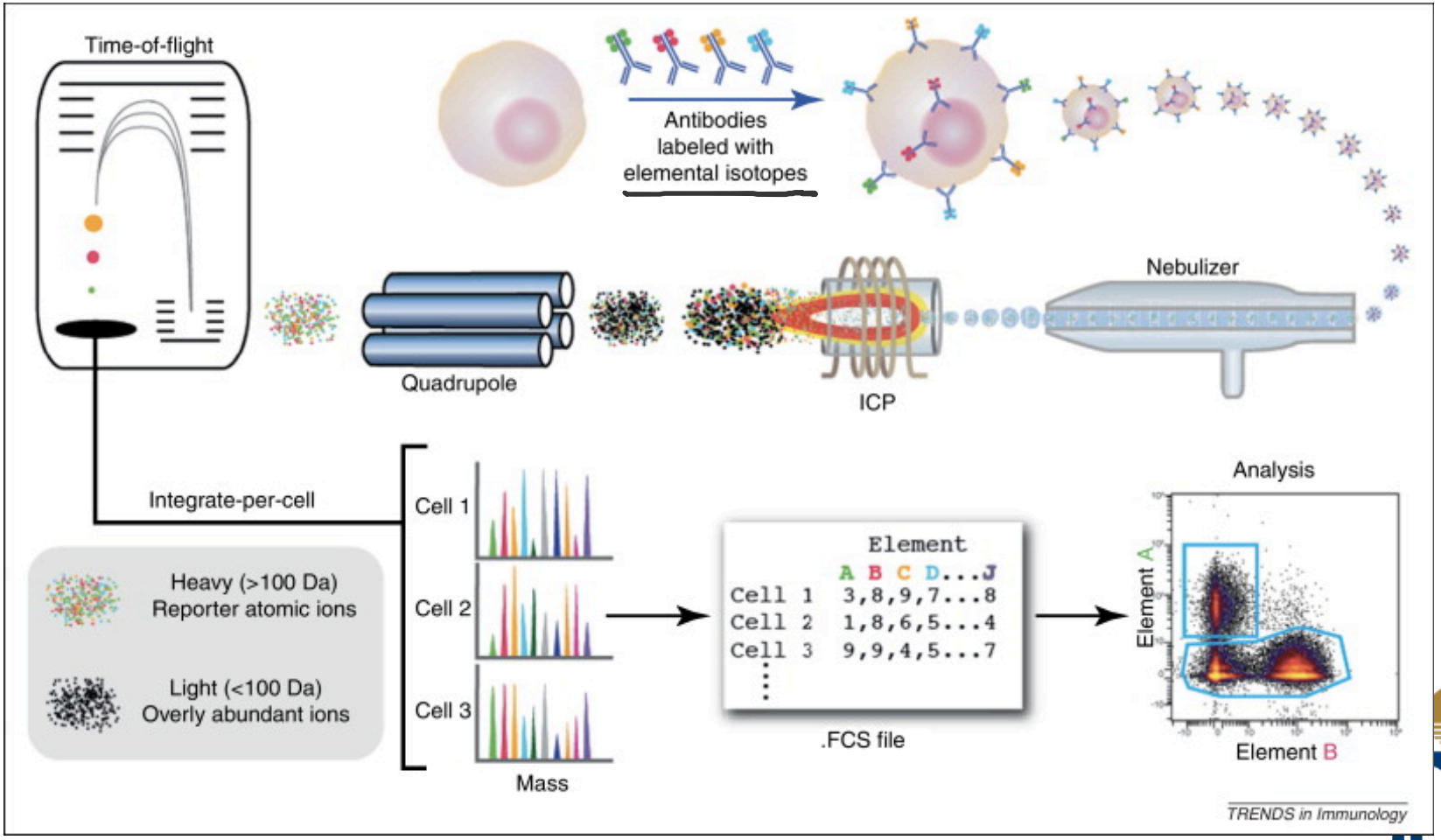
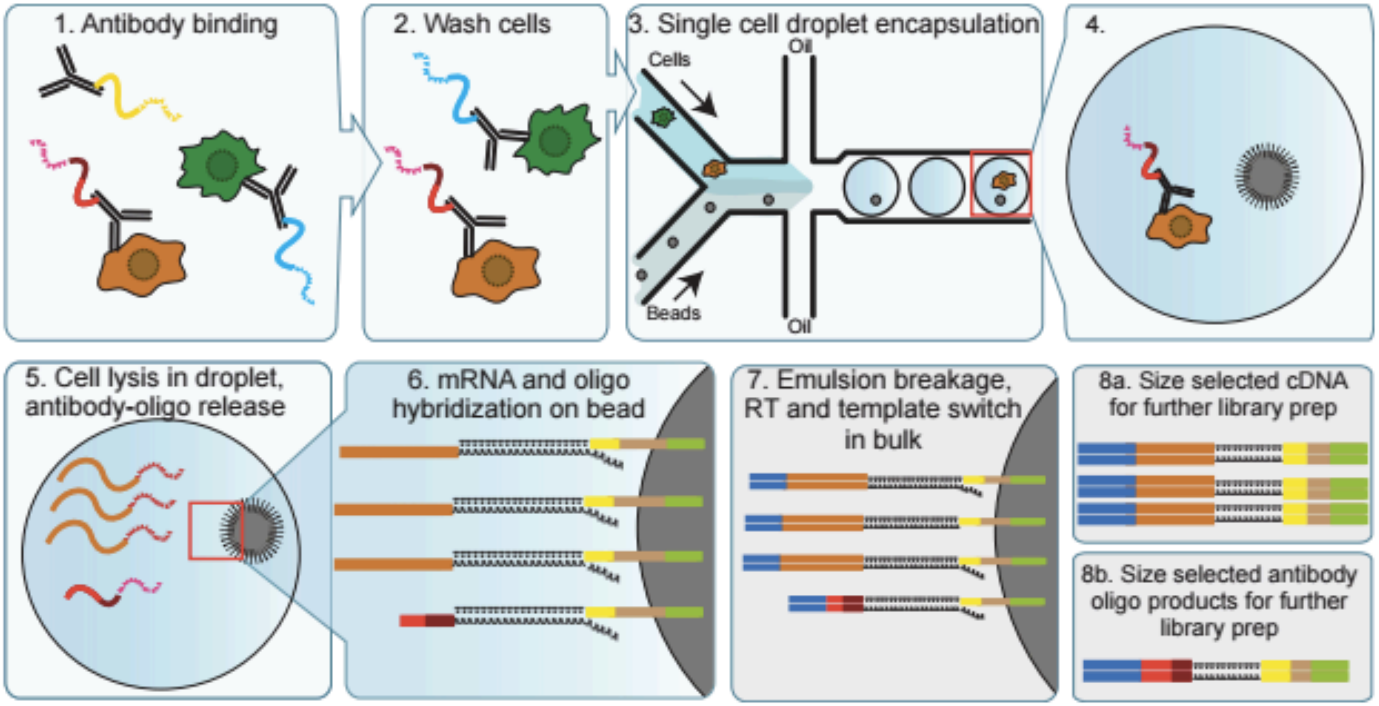


Image: http://cytof.scilifelab.se/homepage/static/images/cytof.jpg

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

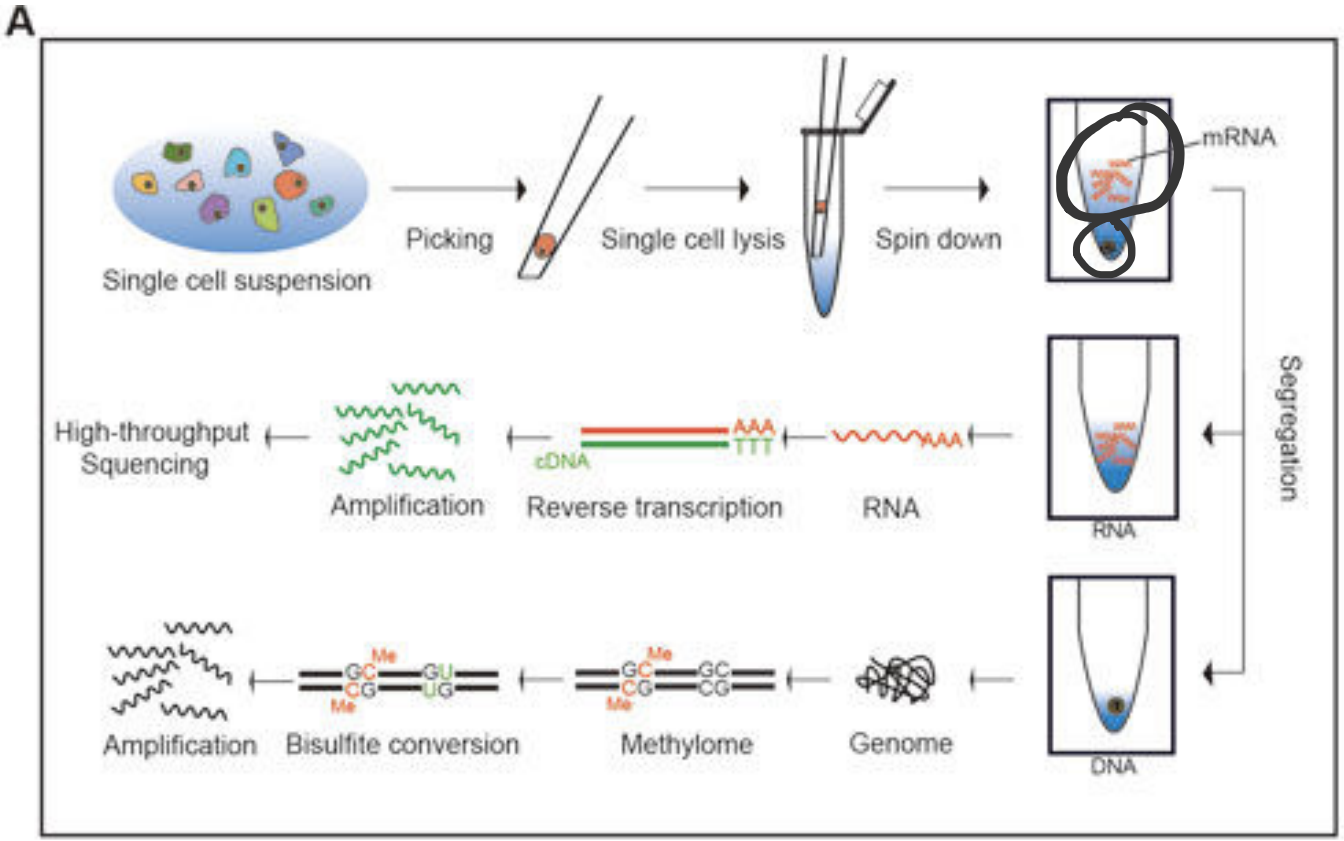
Ab-seq or CITE-seq: cellular indexing of transcriptomes and epitopes by sequencing



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.

PROTEOMICS 101

More details to come in our upcoming guest lecture on May 7!

Mass spectrometry

- A method for determining the chemical components of a sample (could be pure or mixture; gas, liquid, or solid phase)
- "Ionization of the chemical species in the sample, sort the ions into a spectrum based on their mass-to-charge ratio"

Lorentz force

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}$$

Electric force Magnetic force

Newton's second law

$$\vec{F} = m\vec{a}$$

$$\Rightarrow \quad ma = q(\vec{E} + \vec{v}\vec{B})$$

$$\Rightarrow \quad \frac{m}{q} a = \vec{E} + \vec{v}\vec{B}$$

solve for $\frac{m}{q}$ mass to charge.

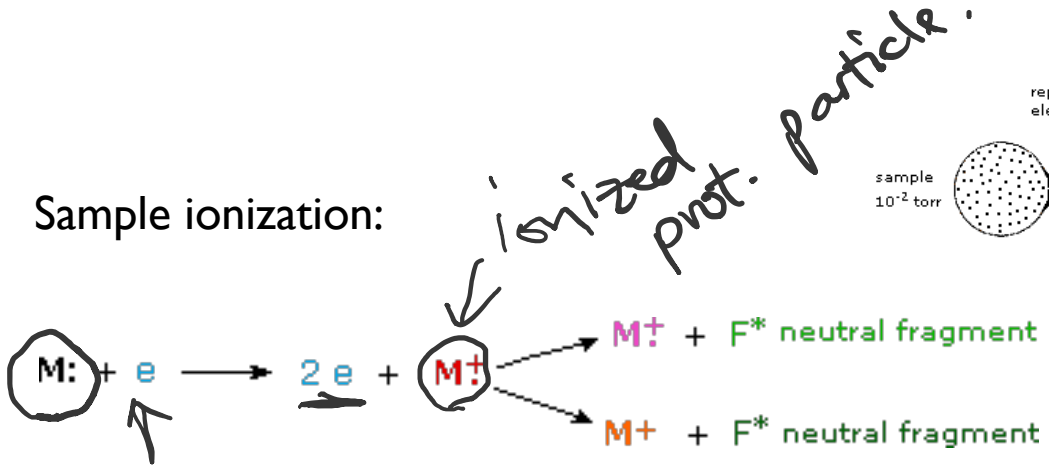
For more derivations, you can check:
http://www.sophphx.caltech.edu/Physics_6/Experiment_09.pdf

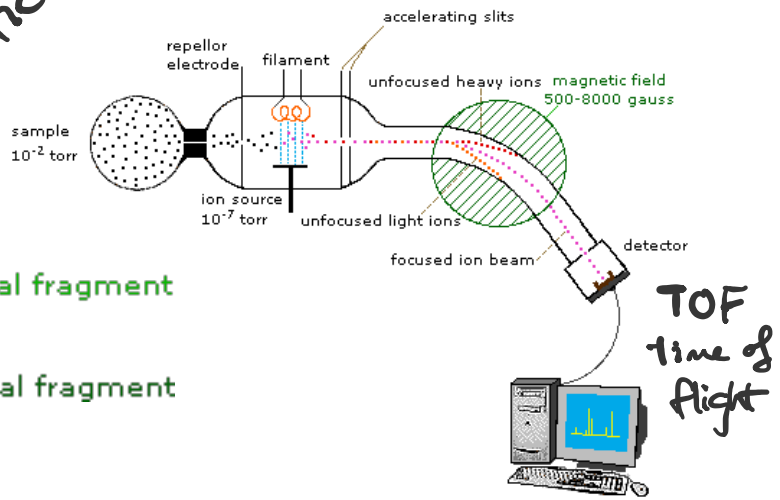
https://en.wikipedia.org/wiki/Mass_spectrometry

Mass spectrometry

- Three components to a MS (modified from https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/spectrpy/massspec/masspec1.htm):
  1. *Sample is ionized, usually to cations by loss of an electron.* ***(The Ion Source)***
  2. *Ions are sorted and separated by mass and charge.* ***(The Mass Analyzer)***
  3. *Separated ions are then measured.* ***(The Detector)***

Sample ionization:

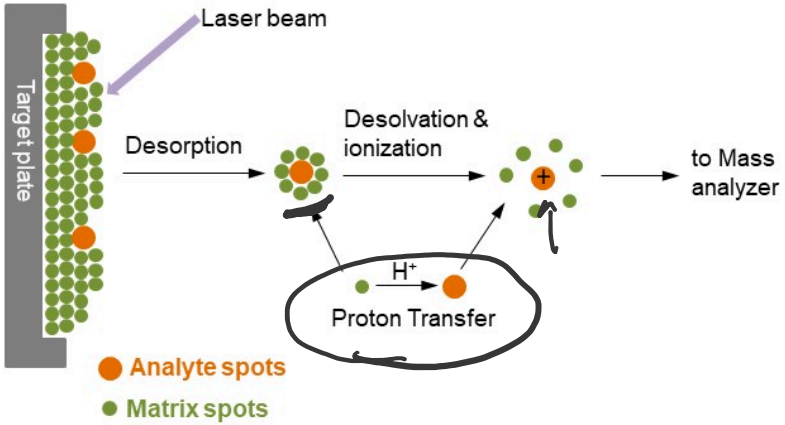




https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/spectrpy/massspec/masspec1.htm

MALDI-TOF

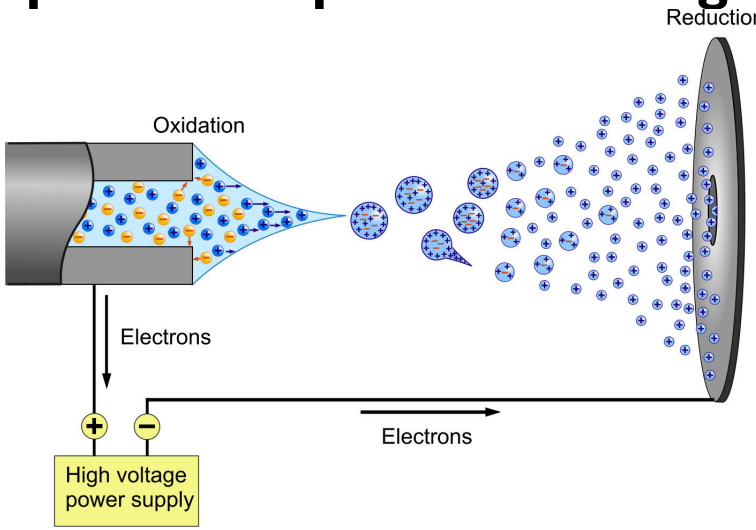
- **Matrix-assisted laser desorption/ionization – time-of-flight**
- **MALDI:** lets us ionize larger molecules (e.g. proteins may fall apart if heated directly, and cannot be ionized directly)
  - The substrate is a solid; usually only singly charged ions are made
- **TOF:** time it takes for ionized particles to reach the detector
  - E.g. if a sample is not pure, then things will be hitting the detector at multiple times, giving a broad peak or multiple peaks



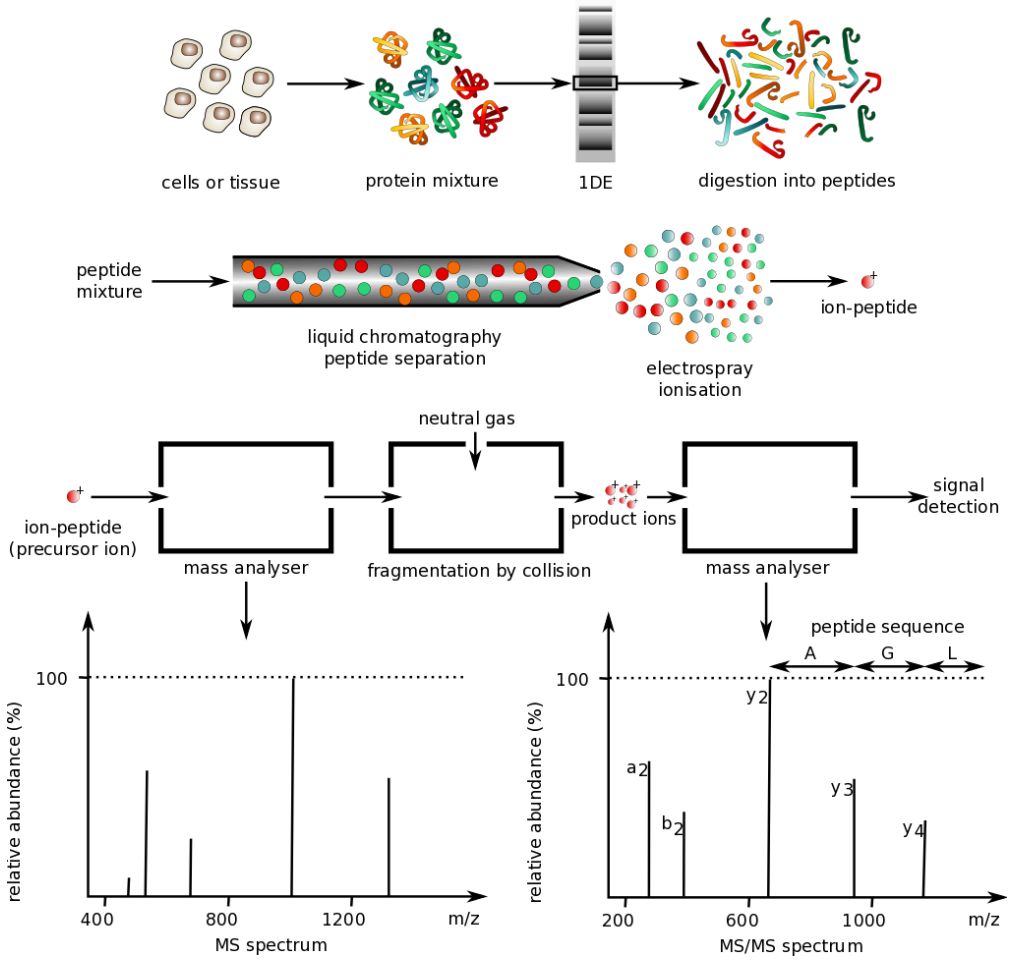
https://www.creative-proteomics.com/technology/maldi-tof-mass-spectrometry.htm

ESI-MS

- **Electrospray ionization:** sample is in a solution that results in ionization by acid/base equilibration; high voltage spraying of the liquid through a tiny capillary nozzle leads to vaporization of the sample followed by detection
- Can produce multiple-charge sample molecules; less chance sample will be (unintentionally) fragmented
- **Can directly couple with liquid chromatography (LC-MS)**



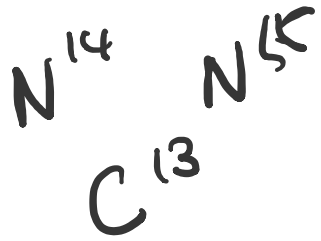
Proteomics approaches



Strategically separate, digest, or fragment the protein into smaller pieces, following by MS on the fragments to generate peptide sequence

2D gel, enzyme digest, fractionation, chromatography

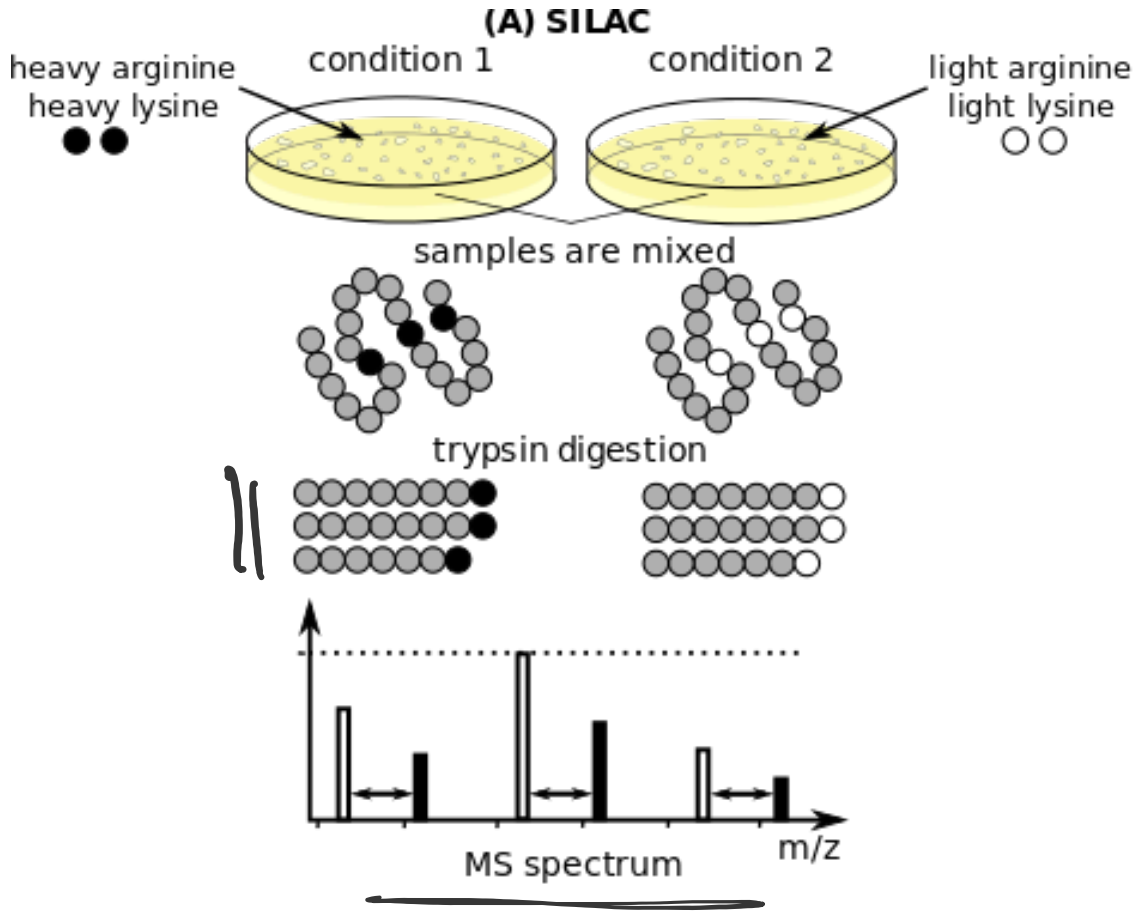
Applications



- **Protein identification** via fingerprinting (esp post-translational modifications); matches MS spectra to existing databases to check for similar peptides
- **Protein structure -** crosslinking then digest/fragment
- **Protein quantification** – makes use of atomic isotopes to label proteins of interest in a system
- **De novo protein sequencing** – more sophisticated computational strategies (Prof Henry Lam in our dept)

https://youtu.be/DAX04wr78Qg

SILAC - Stable isotope labeling by amino acids in cell culture



SYNTHETIC BIOLOGY AND GENOME EDITING

What is synthetic biology and what can we do with it? Also, basics of genome editing

What is synthetic biology?

https://youtu.be/rD5uNAMbDaQ

What is synthetic biology?

- "The element that distinguishes synthetic biology from traditional molecular and cellular biology is the focus on the **design and construction** of core components (parts of enzymes, genetic circuits, metabolic pathways, etc.) that can be modeled, understood, and tuned to **meet specific performance criteria**, and the **assembly** of these smaller parts and devices into larger integrated systems to solve specific problems. Just as engineers now design integrated circuits based on the known physical properties of materials and then fabricate functioning circuits and entire processors (with relatively high reliability), synthetic biologists will soon design and build engineered biological systems."

https://www.ebrc.org/what-is-synbio

Four components of synthetic biology

- Building models of biological systems to test/validate our understanding of biological processes
  - measuring differences between expectation (nature) and observation (model)
- Modifying or manipulating existing living systems to understand biological components/systems; an extension of synthetic chemistry
  - e.g. making synthetic amino acids or modified proteins with new function to replicate/mimic natural phenomenon
- Biology as a technology: creating new biological systems for specific human-oriented needs in informatics, energy production, manufacturing, etc.
  - e.g. engineering bacteria or yeast for production of fuel or medical compounds
  - e.g. genome editing to create chimeras for organ transplant
  - Creating completely new biological organisms

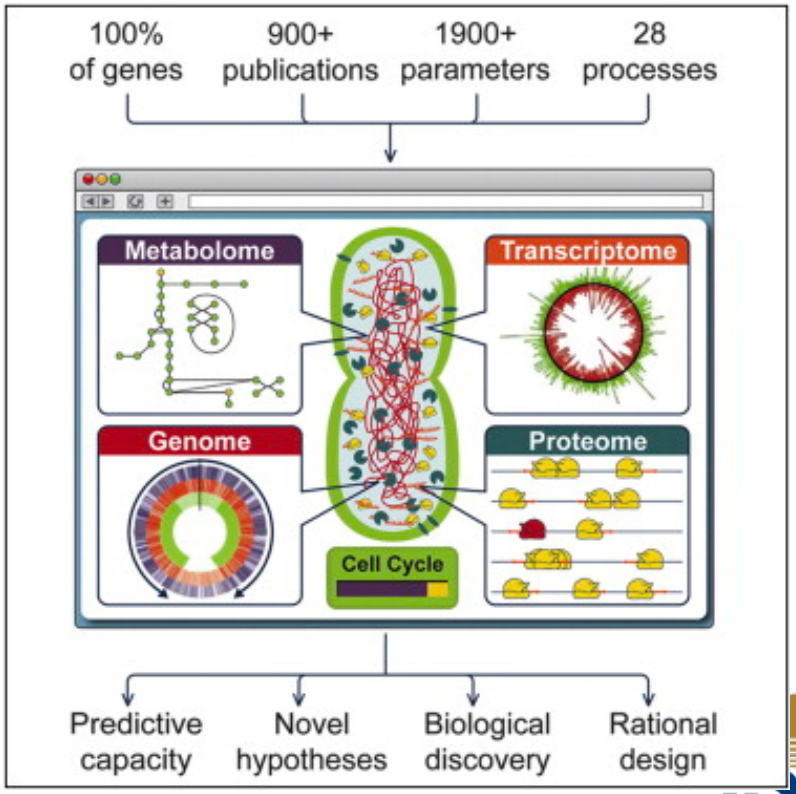
- A "corollary" – creating basic building blocks of biology to make synthetic biology more systematically engineer-able

https://www.ebrc.org/what-is-synbio
https://www.ebrc.org/sites/default/files/Keasling%20sb%20overview.pdf

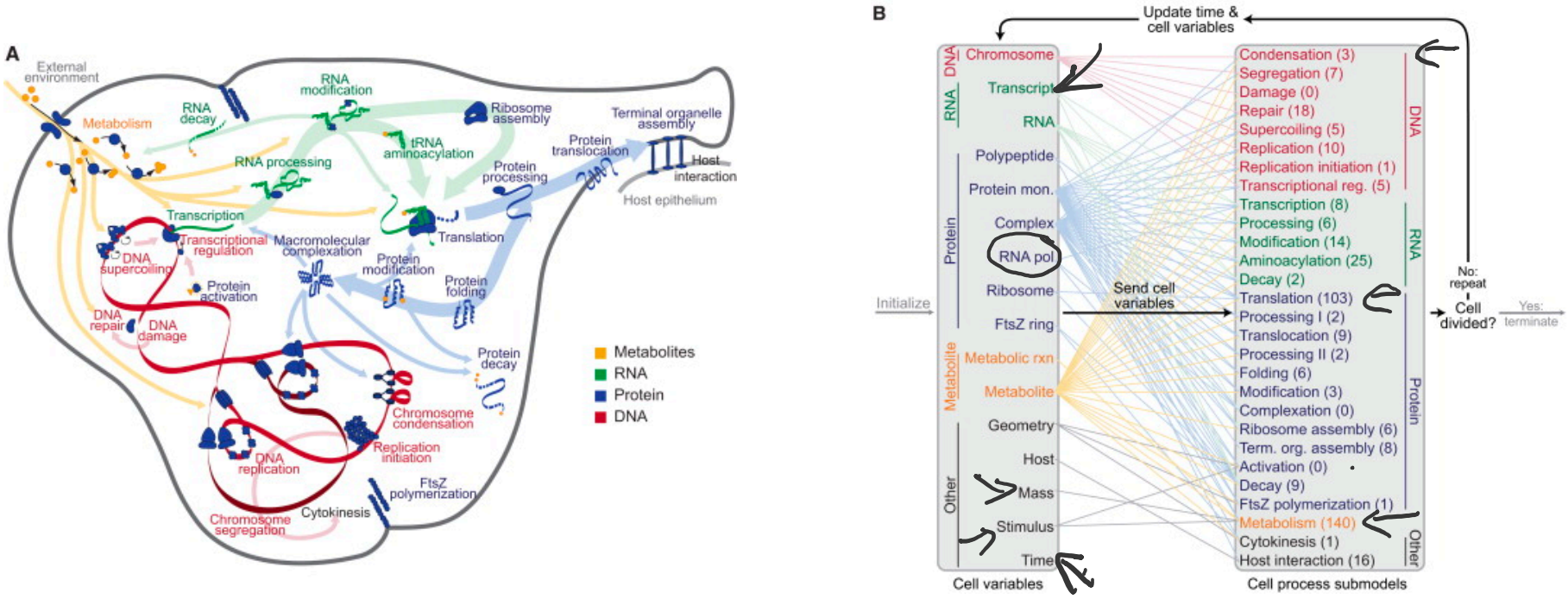
Building models of biological systems to test/validate understanding of biology

mycoplasma.

- Whole cell modelling in-silico – Markus Covert's group, Stanford
- Paper: *JR Karr, JC Sanghvi et al.*, ***A Whole-Cell Computational Model Predicts Phenotype from Genotype***, *Cell, 2012*
- Short talk by the PI: https://www.youtube.com/watch?v=AYC5IE0b8os



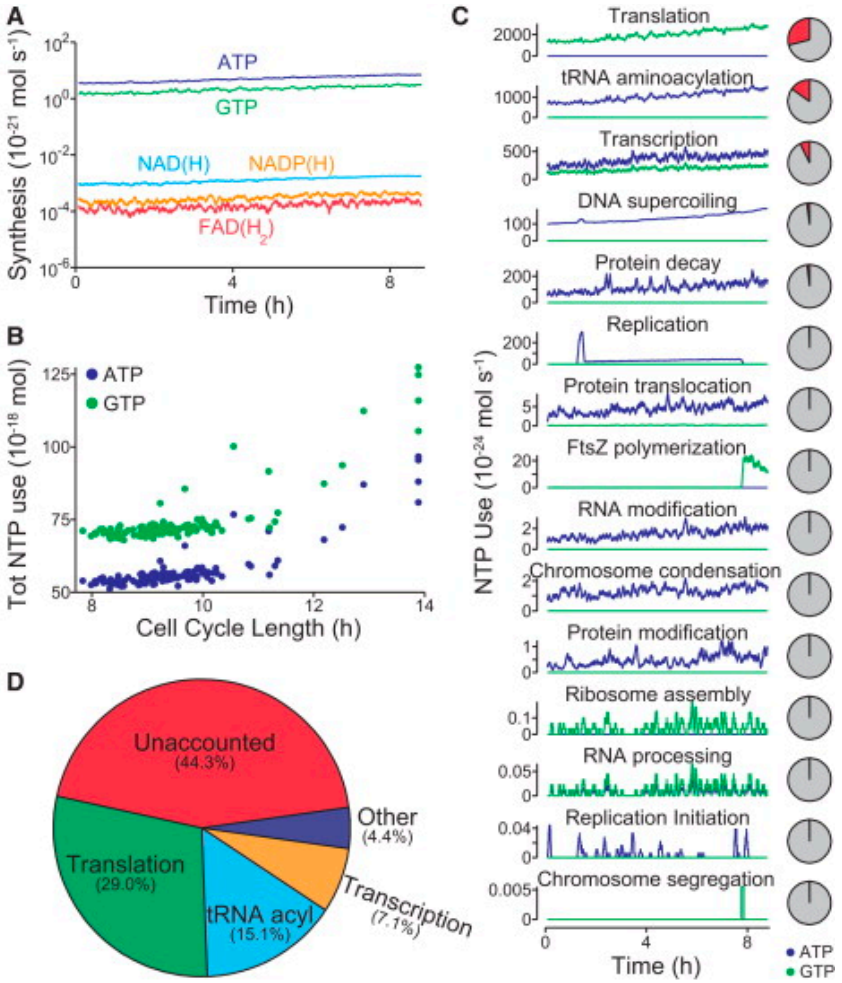
Building models of biological systems to test/validate understanding of biology



JR Karr, JC Sanghvi et al., ***A Whole-Cell Computational Model Predicts Phenotype from Genotype****, Cell, 2012*

Figure 1. (A) 28 submodels grouped by category as metabolic (orange), RNA (green), protein (blue), and DNA (red) in the context of a single M. genitalium cell. Submodels are connected through common metabolites, RNA, protein, and the chromosome. (B) The model integrates cellular function submodels through 16 cell variables. First, simulations are randomly initialized to the beginning of the cell cycle. Next, for each 1 s time step, the submodels retrieve the current values of the cellular variables, calculate their contributions to the temporal evolution of the cell variables, and update the values of the cellular variables. This is repeated thousands of times during the course of each simulation. Simulations are terminated upon cell division when the septum diameter equals zero (right gray arrow).

Building models of biological systems to test/validate understanding of biology



JR Karr, JC Sanghvi et al., ***A Whole-Cell Computational Model Predicts Phenotype from Genotype****, Cell, 2012*

Figure 5. Model Provides a Global Analysis of the Use and Allocation of Energy

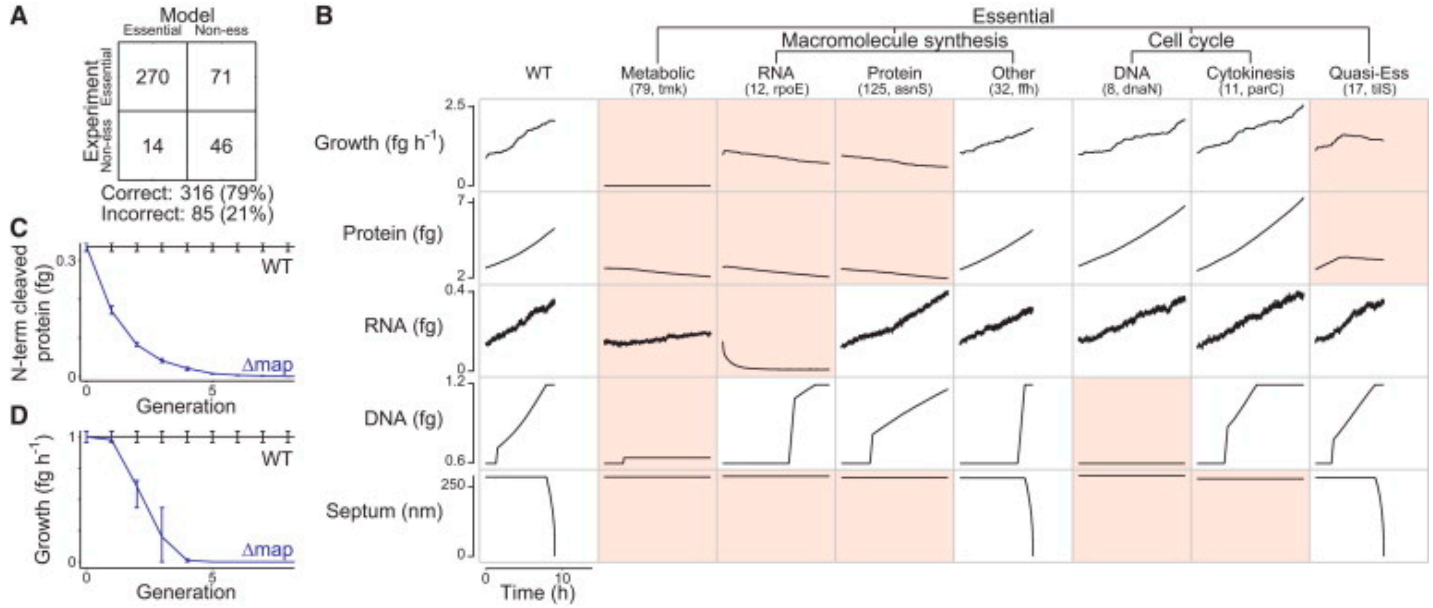
(A) Intracellular concentrations of the energy carriers ATP, GTP, FAD(H_2), NAD(H), and NADP(H) of one in silico cell.

(B) Comparison of the cell-cycle length and total ATP and GTP usage of 128 in silico cells.

(C) ATP (blue) and GTP (green) usage of 15 cellular processes throughout the life cycle of one in silico cell. The pie charts at right denote the percentage of ATP and GTP usage (red) as a fraction of total usage.

(D) Average distribution of ATP and GTP usage among all modeled cellular processes in a population of 128 in silico cells. In total, the modeled processes account for only 44.3% of the amount of energy that has been experimentally observed to be produced during cellular growth.

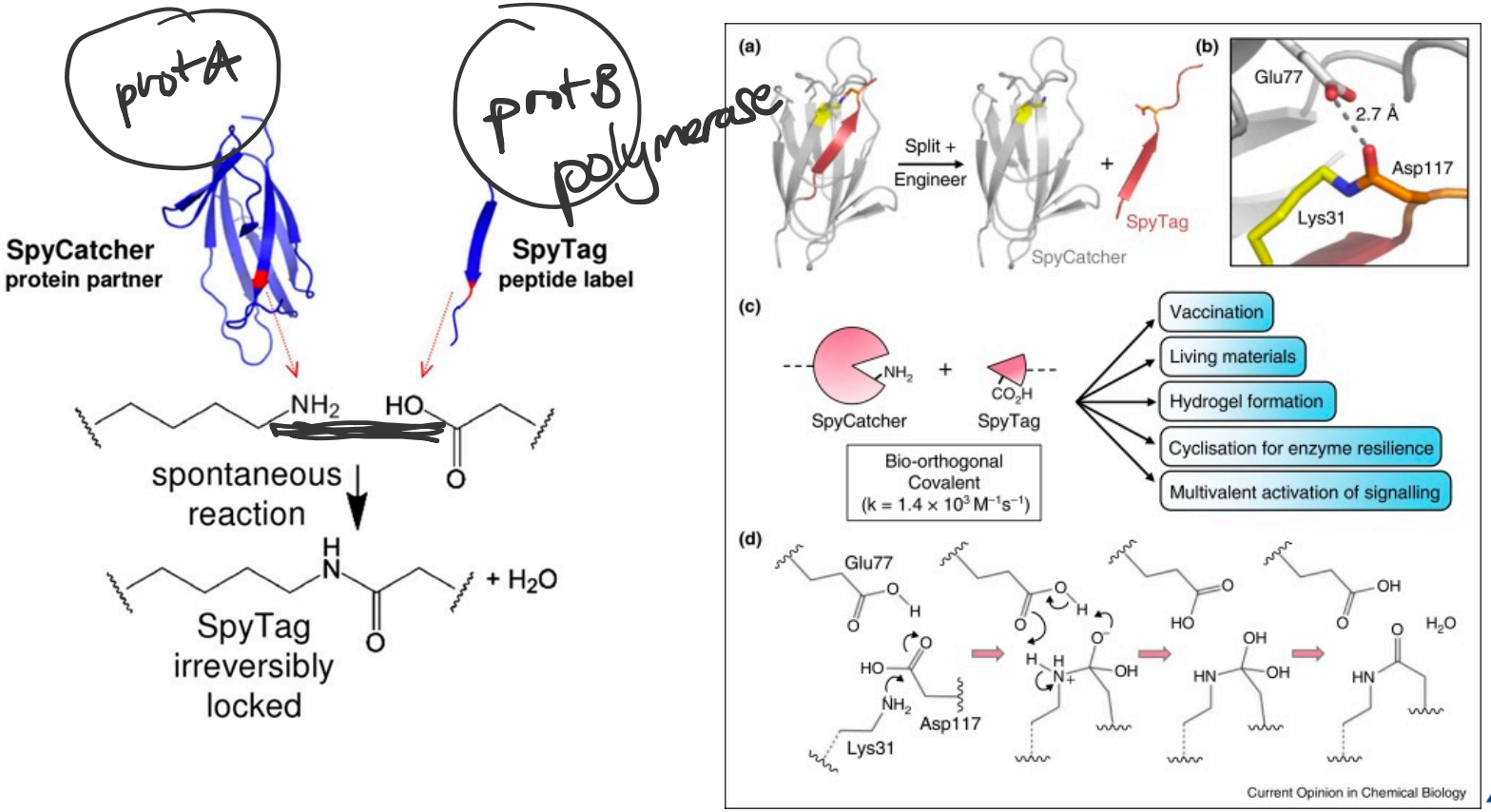
Building models of biological systems to test/validate understanding of biology



JR Karr, JC Sanghvi et al., ***A Whole-Cell Computational Model Predicts Phenotype from Genotype****, Cell, 2012*

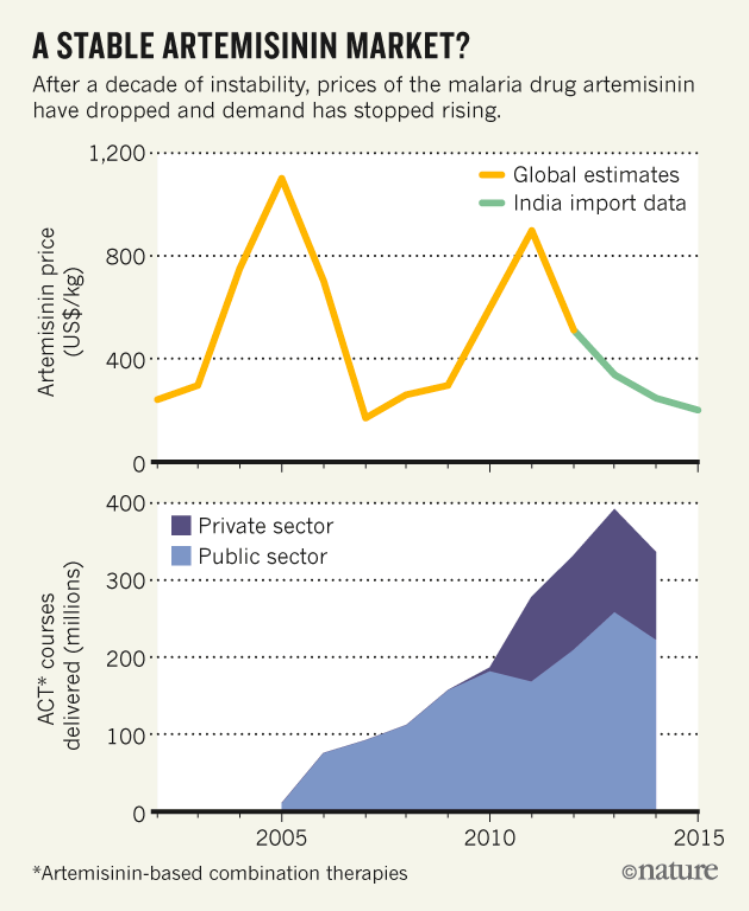
Figure 6. Model Identifies Common Molecular Pathologies Underlying Single-Gene Disruption Phenotypes
(A) Comparison of predicted and observed (Glass et al., 2006) gene essentiality.
(B) Single-gene disruption strains were grouped into phenotypic classes (columns) according to their capacity to grow, synthesize protein, RNA, and DNA, and divide (indicated by septum length). Each column depicts the temporal dynamics of one representative in silico cell of each essential disruption strain class. Disruption strains of nonessential genes are not shown. Dynamics significantly different from wild-type are highlighted in red.
(C and D) Degradation and dilution of N-terminal protein content (C) of methionine aminopeptidase (map, MG172) disrupted cells causes reduced growth (D).

Modifying or manipulating existing living systems; an extension of synthetic chemistry



SC Reddington and M Horwarth, ***Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher****, Curr. Opin. Chem. Biology, 2015*

Biology as a technology: creating new biological systems for human-oriented needs



- Need: stable and inexpensive source of anti-malarial drug Artemisinin (*#NobelPrize*)
- Previously from agricultural sources
- Supply could not meet demand

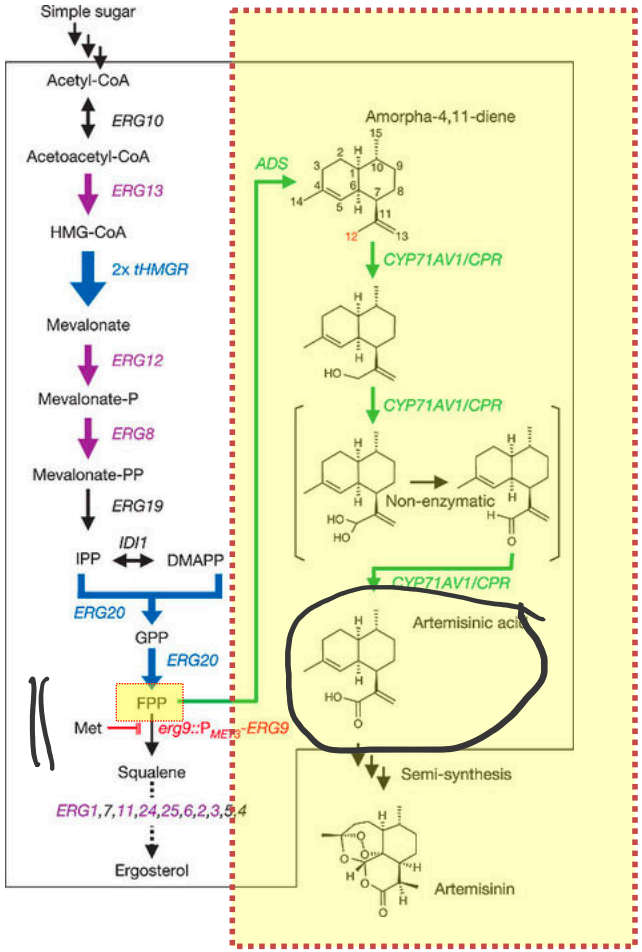
https://www.nature.com/news/synthetic-biology-s-first-malaria-drug-meets-market-resistance-1.19426

Biology as a technology: creating new biological systems for human-oriented needs

Synthetic production of artemisinin in yeast – Jay Keasling's group, UC Berkeley

D Ro et al., ***Production of the antimalarial drug precursor artemisinic acid in engineered yeast****, Nature, 2006*

"We engineered artemisinic-acid-producing yeast in three steps, by (1) engineering the farnesyl pyrophosphate (FPP) biosynthetic pathway to increase FPP production and decrease its use for sterols, (2) introducing the amorphadiene synthase gene (*ADS*) from *A. annua* into the high FPP producer to convert FPP to amorphadiene, and (3) cloning a novel cytochrome P450 that performs a three-step oxidation of amorphadiene to artemisinic acid from *A. annua* and expressing it in the amorphadiene producer."

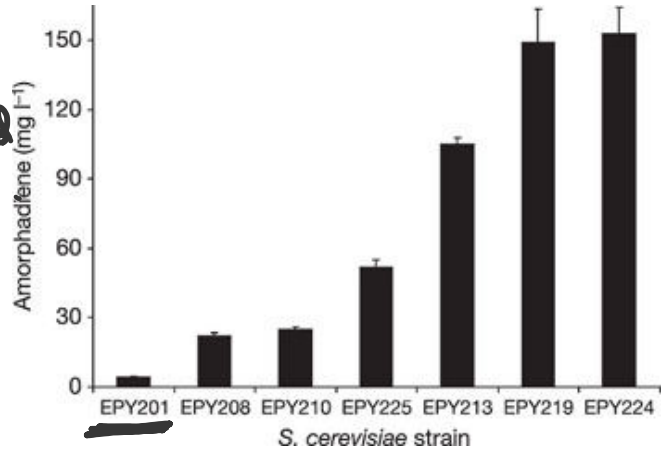


https://www.nature.com/articles/nature04640

Biology as a technology: creating new biological systems for human-oriented needs

D Ro et al., ***Production of the antimalarial drug precursor artemisinic acid in engineered yeast****, Nature, 2006*

"To increase FPP production in *S. cerevisiae*, the expression of several genes responsible for FPP synthesis was upregulated, and one gene responsible for FPP conversion to sterols was downregulated. All of these modifications to the host strain were made by chromosomal integration to ensure the genetic stability of the host strain."



Further optimization was done, reported more recently in Nature; this drug is now in manufacturing production by drug companies:

CJ Paddon et al., ***High-level semi-synthetic production of the potent antimalarial artemisinin****, Nature, 2013*

AMYRIS

"Our results describe for the first time… the expression of the complete pathway for artemisinic acid production, which resulted in a ***greater than tenfold increase*** in artemisinic acid titres. In addition, we demonstrated a significant increase in the efficiency of artemisinic acid conversion to artemisinin compared with earlier work."

https://www.nature.com/articles/nature04640
https://www.nature.com/articles/nature12051

Biology as a technology: creating new biological systems for human-oriented needs

painkillers

Second example: Synthetic production of opioids in yeast – Christina Smolke's group, Stanford

Can This Silicon Valley Startup Bioengineer A Less Addictive Opioid?

Christina Smolke's Stanford team genetically altered yeast to produce opioids. Can her startup, Antheia, make them more efficient to produce and safer to use?

Professor **Christina Smolke** [Photo: Rod Searcey]

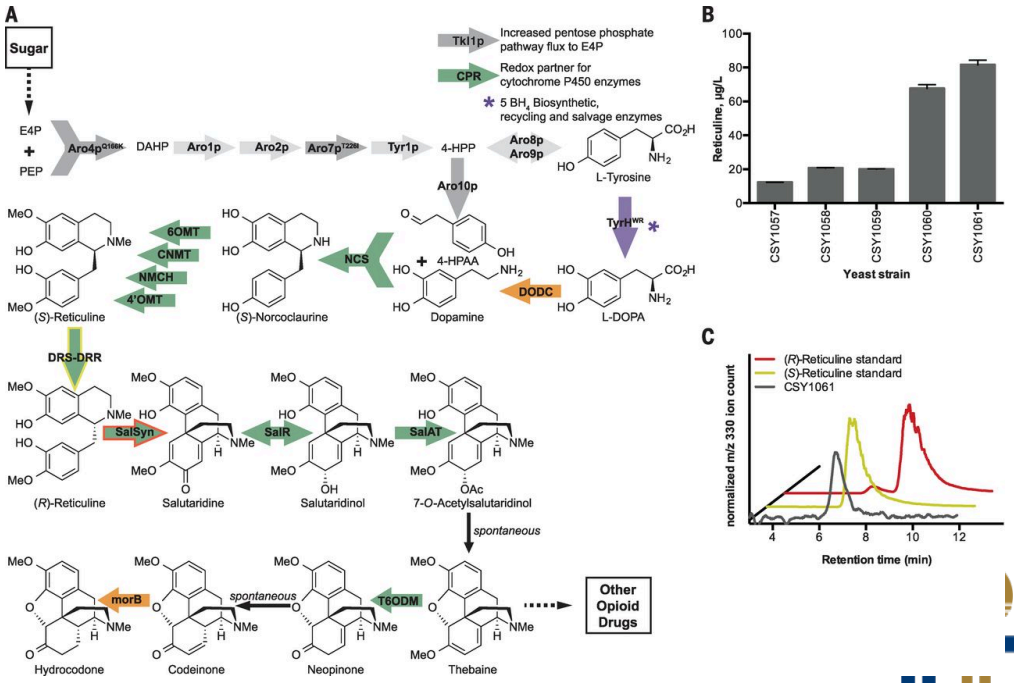
BY MARK SULLIVAN LONG READ

Back in 2015, a 40-year-old synthetic biologist named Christina Smolke, along with a small team of researchers at Stanford, made a huge discovery. They proved that a genetically engineered yeast could produce opioid molecules, the core ingredients of some of the world's most widely prescribed pain medicines.

ANTHEIA

Opinion and controversy:
https://www.nature.com/news/engineered-yeast-paves-way-for-home-brew-heroin-1.17566

S Galanie et al., ***Complete biosynthesis of opioids in yeast***, *Science, 2015*



http://science.sciencemag.org/content/349/6252/1095.full

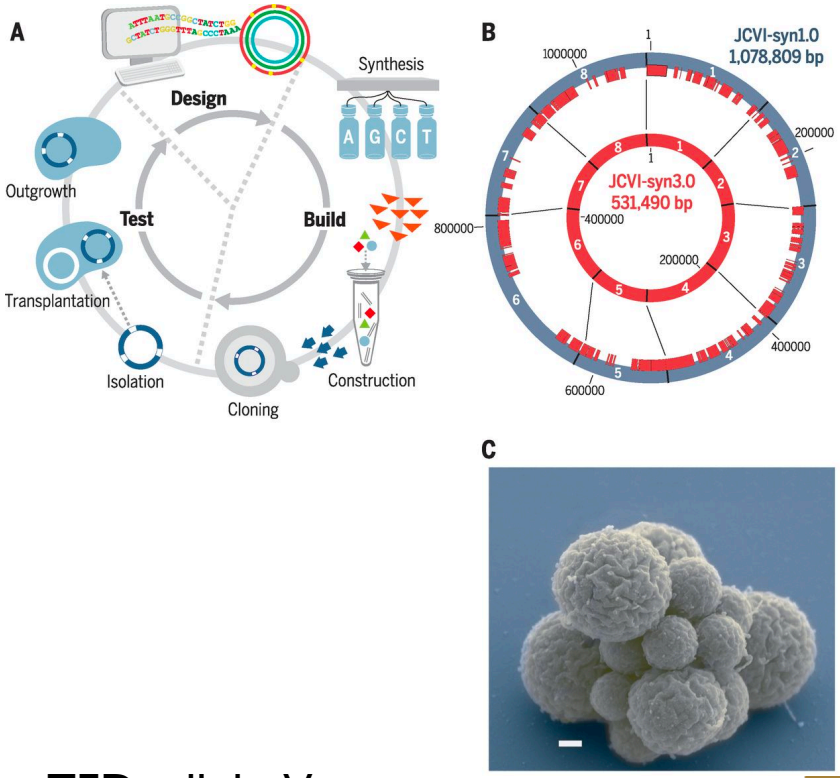
https://www.fastcompany.com/3066129/can-antheias-opioid-producing-yeast-one-day-challenge-big-pharma

Creating new organisms from scratch??

"First synthetic lifeform", a microorganism for producing alternative fuels – Craig Venter's group, J. Craig Venter Institute

CA Hutchison III et al., ***Design and synthesis of a minimal bacterial genome****, Science, 2016*

"Whole-genome design and synthesis were used to minimize the 1079–kilobase pair (kbp) synthetic genome of *M. mycoides* JCVI-syn1.0. An initial design, based on collective knowledge of molecular biology in combination with limited transposon mutagenesis data, failed to produce a viable cell. Improved transposon mutagenesis methods revealed a class of quasi-essential genes that are needed for robust growth, explaining the failure of our initial design. Three more cycles of design, synthesis, and testing, with retention of quasi-essential genes, produced JCVI-syn3.0 (531 kbp, 473 genes). Its genome is smaller than that of any autonomously replicating cell found in nature. JCVI-syn3.0 has a doubling time of ~180 min, produces colonies that are morphologically similar to those of JCVI-syn1.0, and appears to be polymorphic when examined microscopically."



TED talk by Venter:
https://www.youtube.com/watch?v=HdgfzdlgUHw

http://science.sciencemag.org/content/351/6280/aad6253

Creating new organisms from scratch??

Mixed reviews on whether this really counts as a fully synthetic lifeform:

"Bioethicist Arthur Caplan finds the philosophical ramifications of the work fascinating:

> *"Their achievement undermines a fundamental belief about the nature of life that is likely to prove as momentous to our view of ourselves and our place in the Universe as the discoveries of Galileo, Copernicus, Darwin and Einstein." [Nature News]*

But many experts say that since Venter copied a pre-existing genome, he didn't really create a new life form.

> *"To my mind Craig has somewhat overplayed the importance of this,"* said David Baltimore, a leading geneticist at Caltech. Dr. Baltimore described the result as *"a technical tour de force"* but not breakthrough science, but just a matter of scale…. *"He has not created life, only mimicked it,"* Dr. Baltimore said [The New York Times].

In addition, many experts note that the experimenters got a big boost by placing the synthetic genome in a preexisting cell, which was naturally inclined to make sense of the transplanted DNA and to turn genes on and off. Thus, they say, it's not accurate to label the experiment's product a true "synthetic cell.""

http://blogs.discovermagazine.com/80beats/2010/05/20/did-craig-venter-just-create-synthetic-life-the-jury-is-decidedly-out/#.Wua0s9OWTUI

Basic building blocks to make biology more amenable to systematic engineering

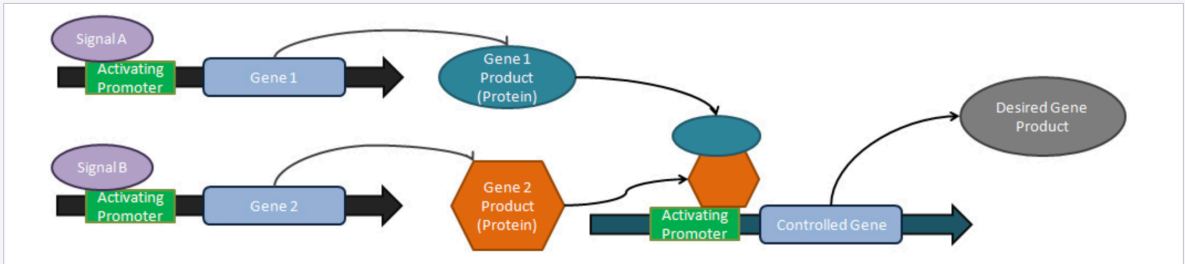
- e.g. gene circuits; biology has often been compared to electronics/circuits

BioBricks FOUNDATION

iGEM

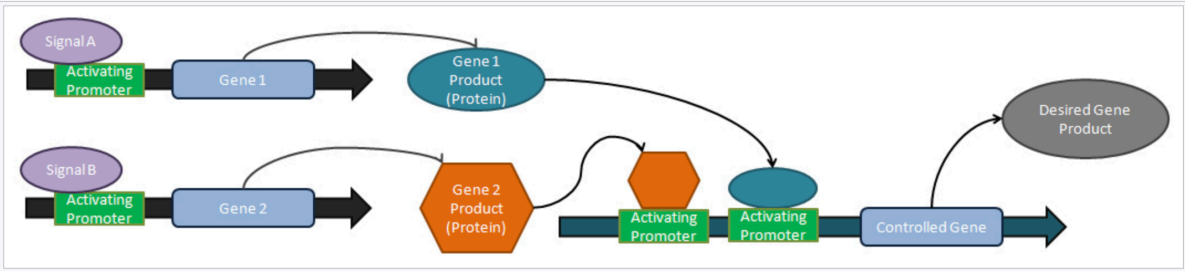
Single biowerks

AND



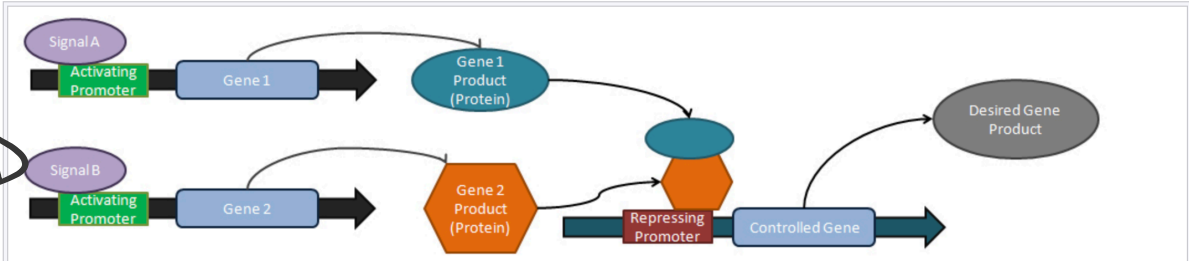
The logical AND gate.[10][11] If Signal A **AND** Signal B are present, then the desired gene product will result. All promoters shown are inducible, activated by the displayed gene product. Each signal activates expression of a separate gene (shown in light blue). The expressed proteins then can either form a complete complex in cytosol, that is capable of activating expression of the output (shown), or can act separately to induce expression, such as separately removing an inhibiting protein and inducing activation of the uninhibited promoter.

OR



The logical OR gate.[10][11] If Signal A **OR** Signal B are present, then the desired gene product will result. All promoters shown are inducible. Either signal is capable of activating the expression of the output gene product, and only the action of a single promoter is required for gene expression. Post-transcriptional regulation mechanisms can prevent the presence of both inputs producing a compounded high output, such as implementing a low binding affinity ribosome binding site.

NAND



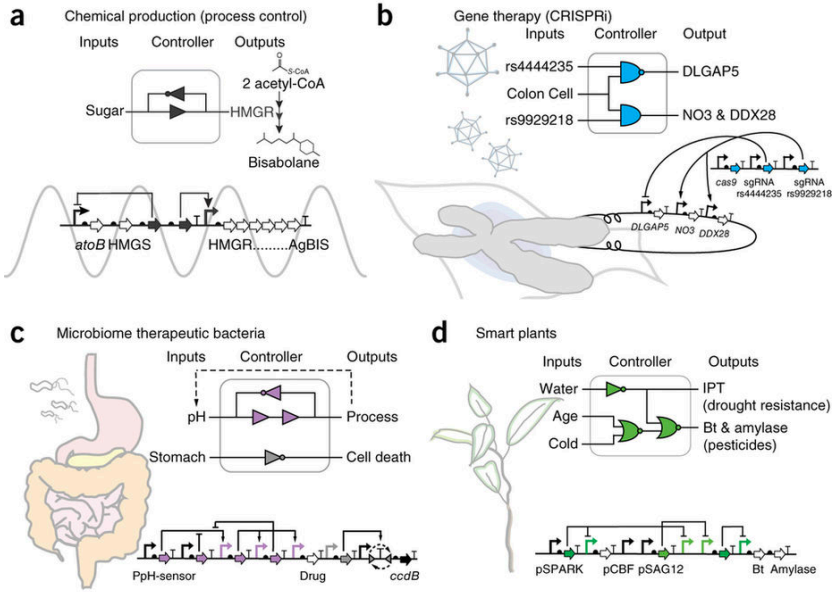
The logical Negated AND gate.[10][11] If Signal A **AND** Signal B are present, then the desired gene product will **NOT** result. All promoters shown are inducible. The activating promoter for the output gene is constitutive, and thus not shown. The constitutive promoter for the output gene keeps it "on" and is only deactivated when (similar to the AND gate) a complex as a result of two input signal gene products blocks the expression of the output gene.

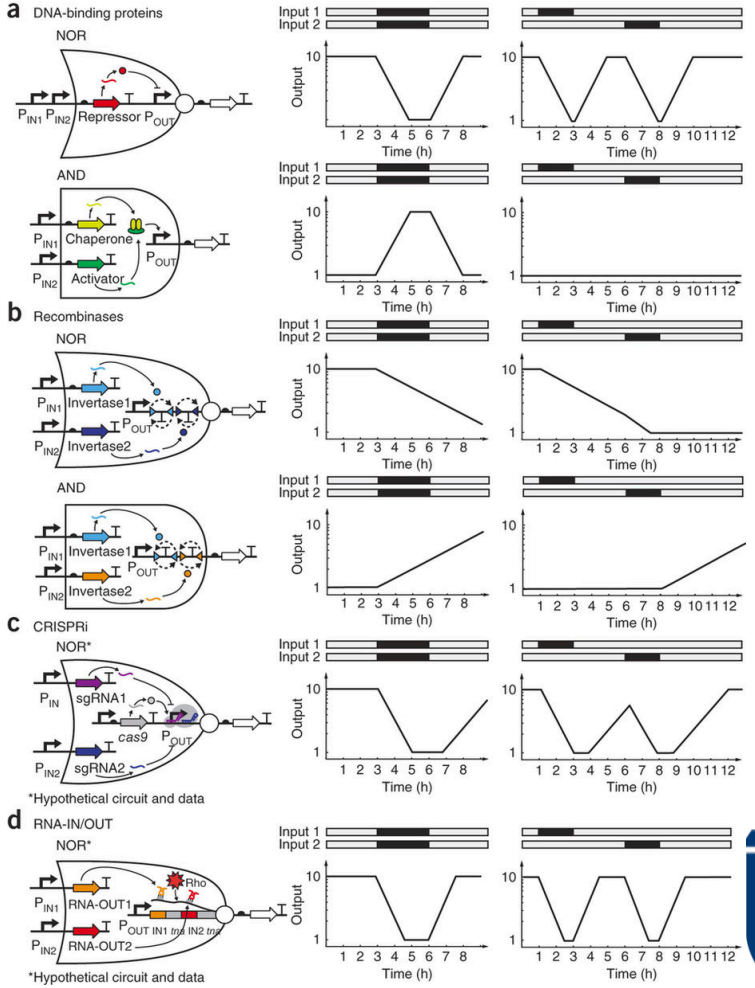
https://en.wikipedia.org/wiki/Synthetic_biological_circuit

Basic building blocks to make biology more amenable to systematic engineering

A nice review article on gene circuits:

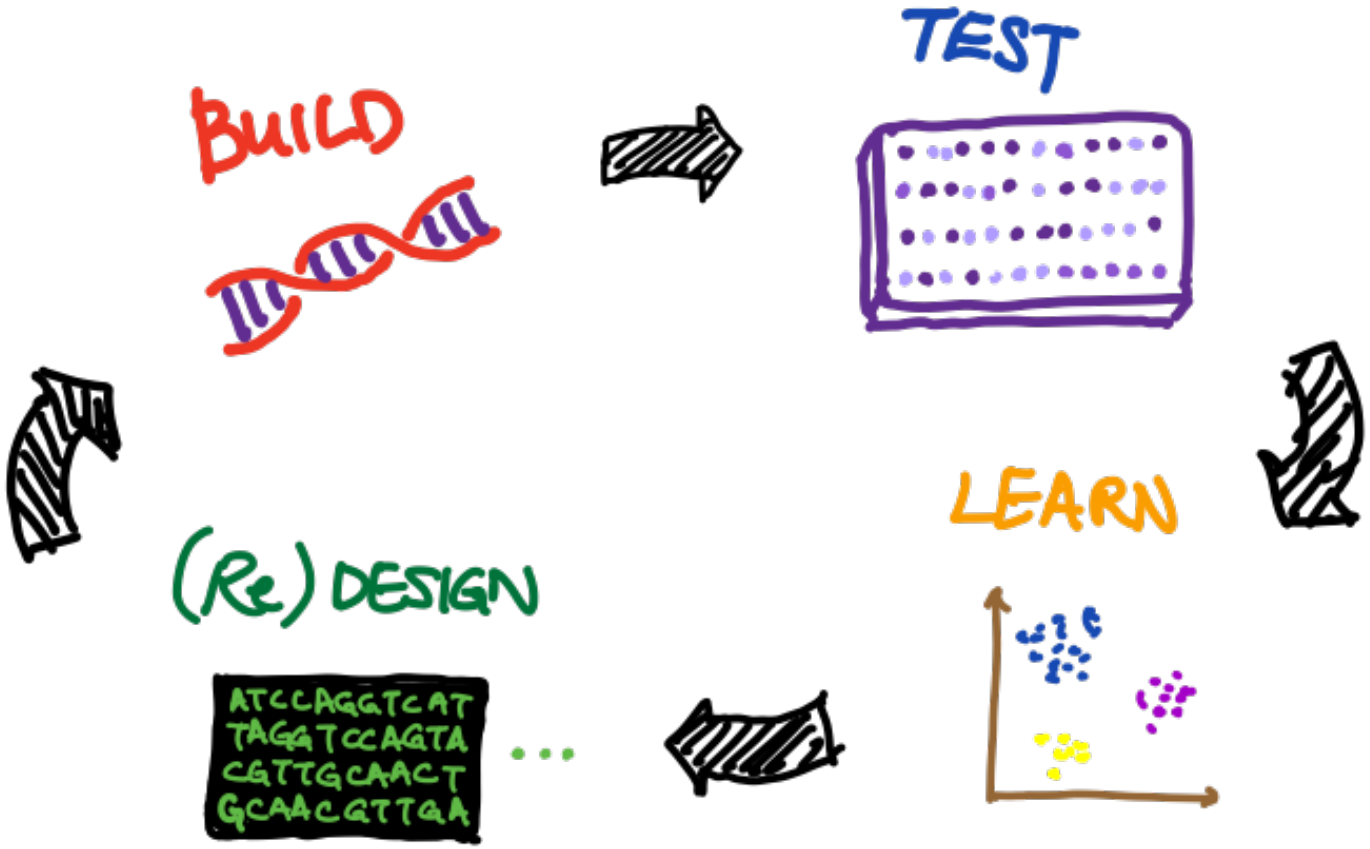
JAN Brophy and CA Voigt, ***Principles of genetic circuit design****, Nature Methods, 2014*





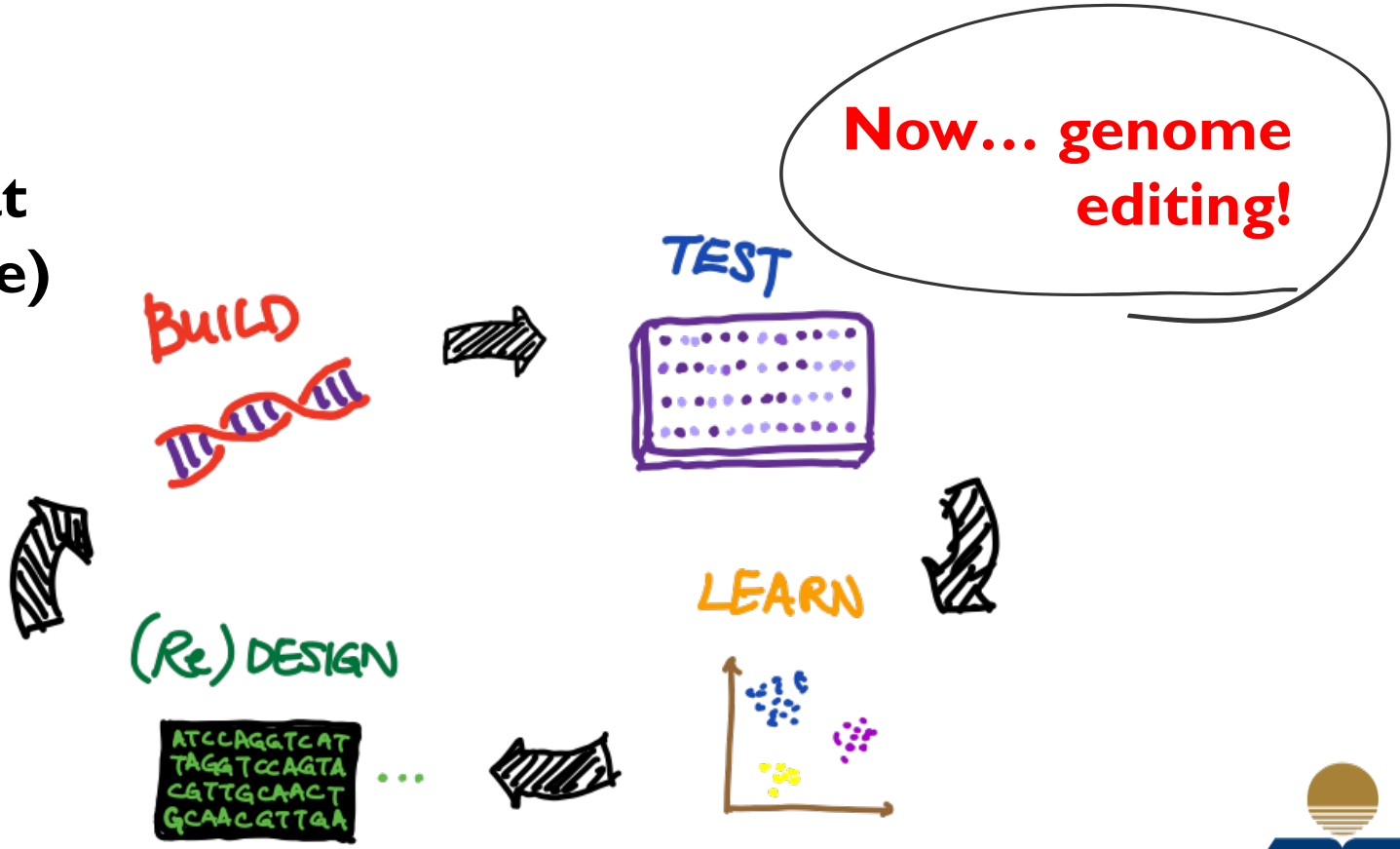
https://www.nature.com/articles/nmeth.2926

The synthetic biology engineering pipeline



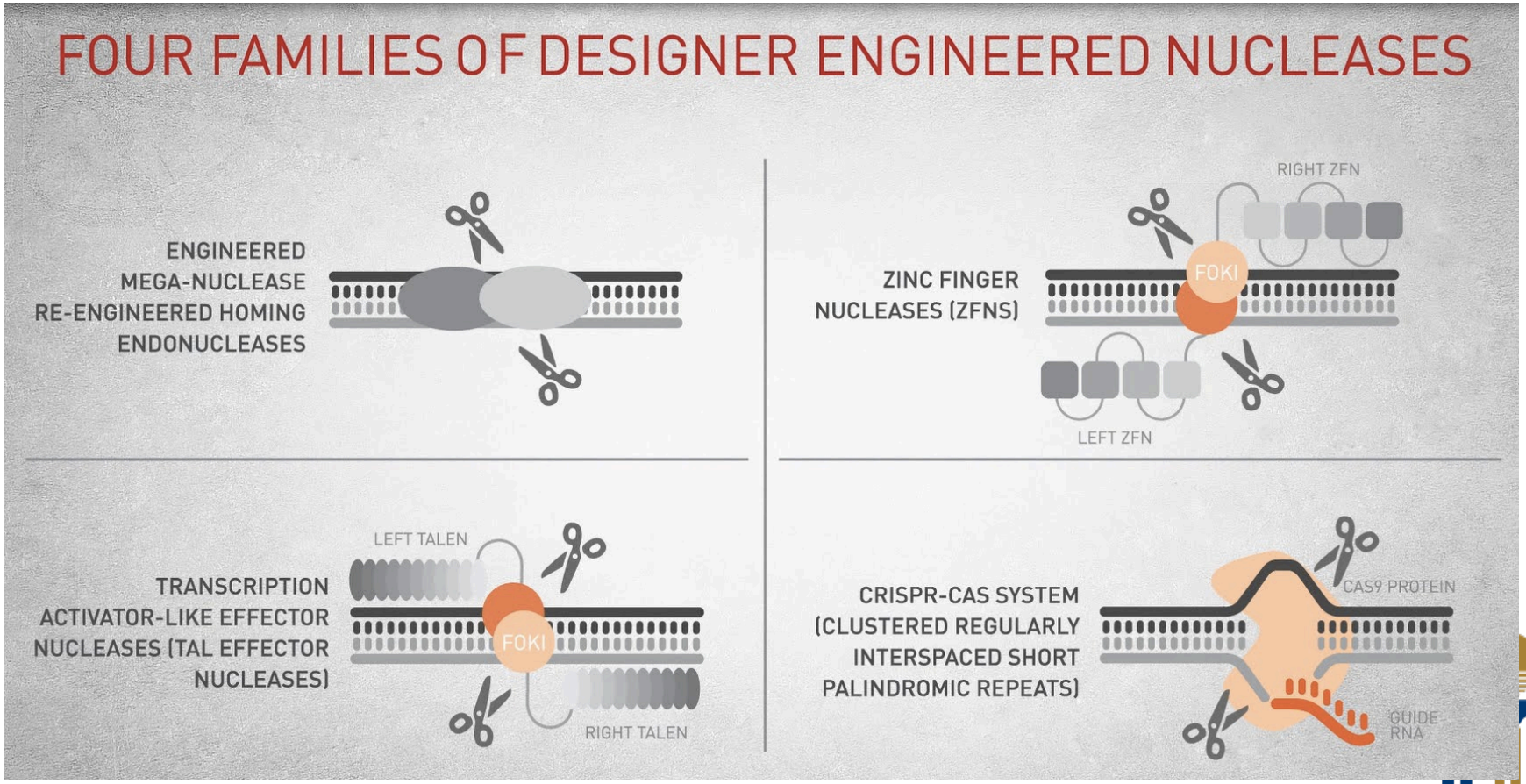
Two major innovations has sped up synthetic biology progress in the last decade

DNA synthesis at scale (write)



NGS (read)

Genome editing has been researched for a while and there are lots of options available…



From ThermoFisherScientific: https://i.ytimg.com/vi/VvwKubUgwDQ/maxresdefault.jpg

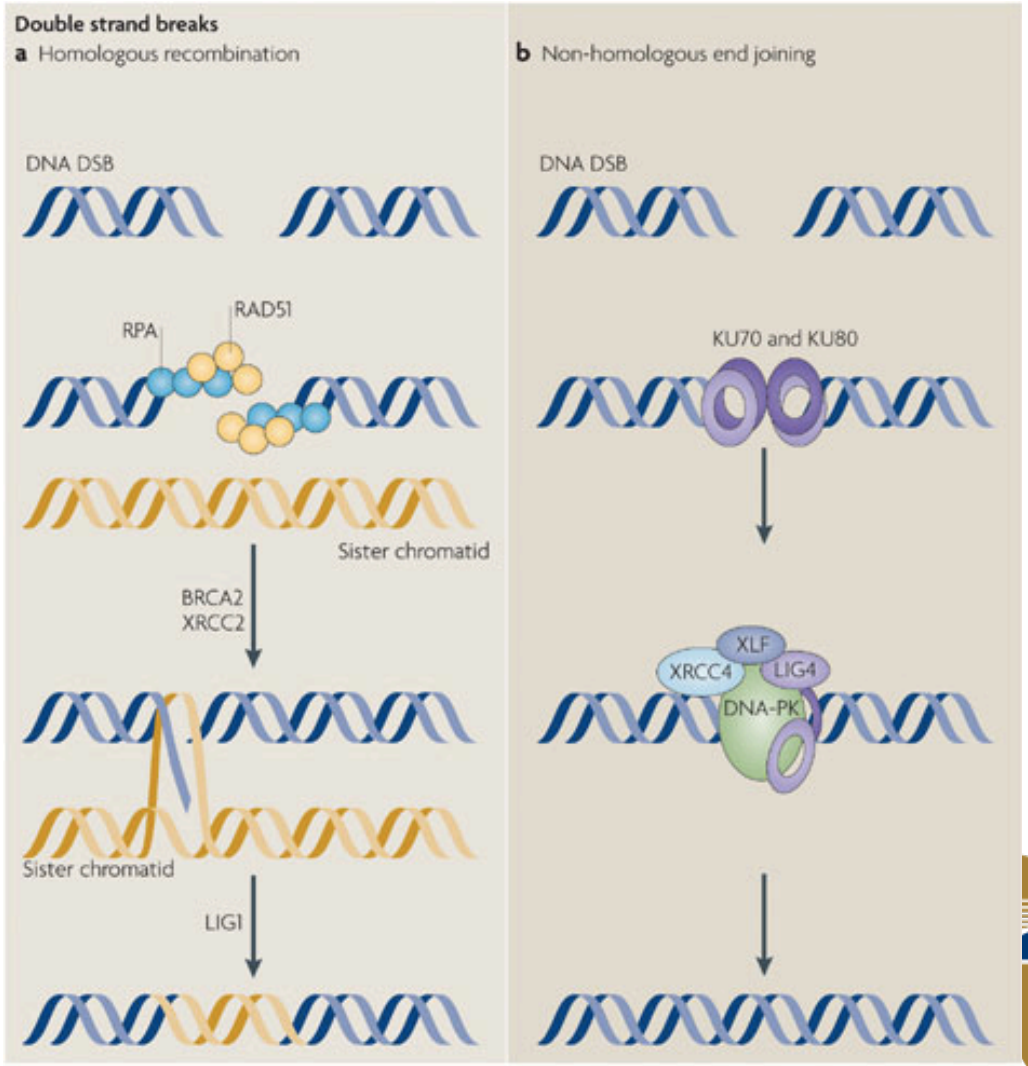
Putting the elephant in the fridge

- Fundamentally, how does one change the genome? It's as easy as putting an elephant in a fridge:
  - 1. Cut the DNA
  - 2. Remove the DNA
  - (Optional: add new sequence)
  - 3. Join the pieces back together

- DNA damage and repair happens in our body all the time, and the body uses TWO main mechanisms to do so.

DNA repair

- Homology directed repair (HDR)
  - More precise; less error
  - Requires homologous DNA to be present
  - https://www.youtube.com/watch?v=86JCMM5kb2A
- Non-homologous end joining (NHEJ)
  - Error prone – could end up with cancer ☹
  - Maybe it's still useful?
  - https://www.youtube.com/watch?v=31stiofJjYw



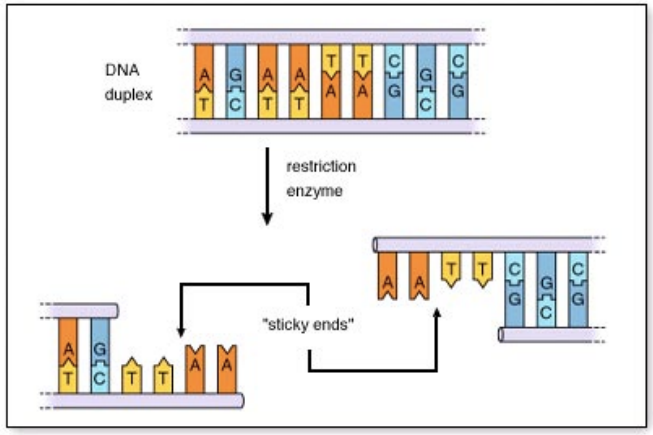
Putting the elephant in the fridge

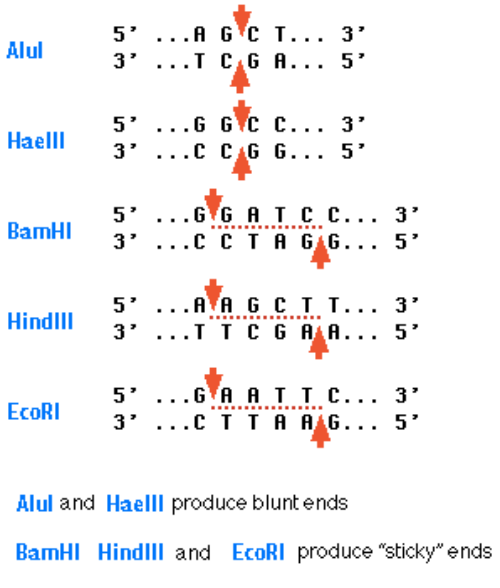
- So let's refine our steps:
  - 1. Cut the DNA: Create a double strand break (DSB)
  - 2. Remove the DNA: end-repair by the cell's natural mechanisms
  - (Optional: add new sequence – use the HDR feature!)
  - 3. Join the pieces back together: either HDR or NHEJ

- From the mechanism, we can see that if we want to add a new sequence, we must use HDR, and not NHEJ!
  - Deliver the desired sequence into the cell, so HDR uses new sequence as template for repair

Genome editing in the lab

- In some cases, the goal is to generate a lot of random mutations to make a model organism:
  - E.g. Genetic screening
  - We randomly introduce DSBs, using mutagenic chemicals added to the cells, or UV light, making random DSB
  - NHEJ will end up making some mutations
- Restriction enzymes offer a bit more sequence specificity than random UV-induced DSBs, but still many locations in the genome for each restriction site sequence

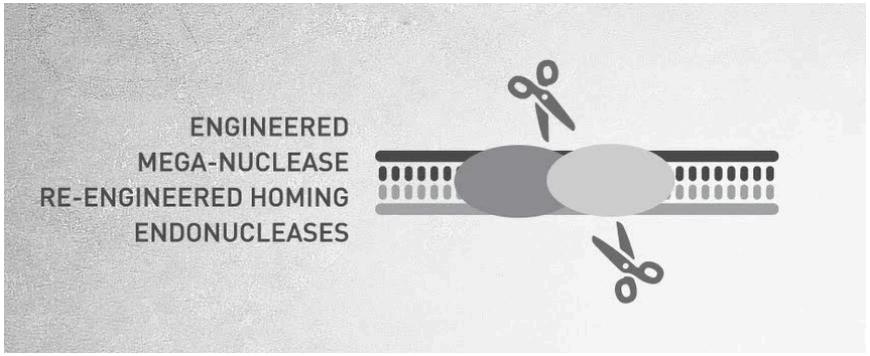




Making the cut

- To be more useful, we need to make a DSB in a specific place!
- Engineered nucleases in different families have different mechanisms of locating and making the DSB

Meganucleases



- Endo-deoxyribonucleases
- 12-40 bp recognition site
  - Question: how many bases do you need for one specific 18bp sequence occur by chance?
- High cleavage efficiency
- HIGHEST SPECIFICITY for genome editing
- Recognition is by specific sequence recognition domain ”hardwired” into the protein
- Protein engineering used to generate new target sites
- What could affect presumed cutting efficiency or specificity of this nuclease?
- Precision Biosciences; Bayer

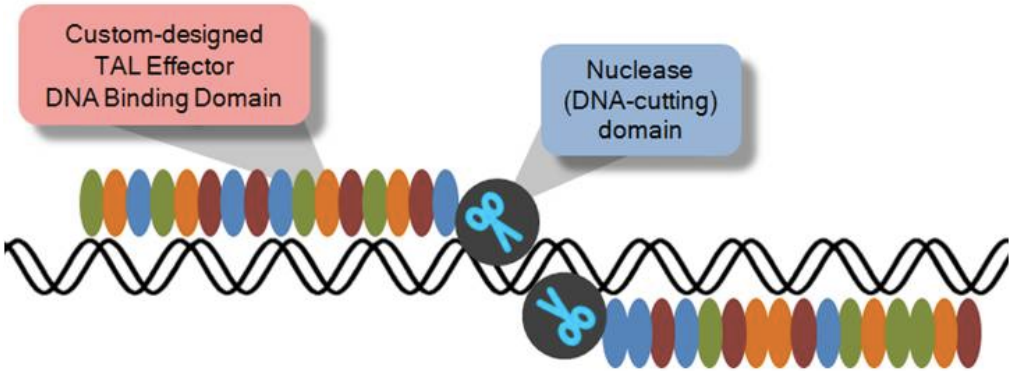
Zinc Finger Nuclease (ZFN)

- DNA recognition domain (ZF) + cutting domain (FokI)
  - GPS + bomb! One to locate, one to execute
- Zinc finger motifs/repeats (https://www.youtube.com/watch?v=yYibFsjX9Zw)
  - Each 'unit' recognizes 3 bp; one ZF protein = 3-6 motifs/repeats
  - That gives 9-18 bp recognition site
- Assembly can be difficult
- Off target effects (less specific)
- Sangamo; Sigma Aldrich

http://www.sigmaaldrich.com/life-science/zinc-finger-nuclease-technology/learning-center/what-is-zfn.html

Transcription activator-like effector nucleases (TALEN)

- DNA recognition domain (TALE) + cutting domain (part of FokI)
- TAL-effector domain and DNA bp has1-to-1 recognition
  - Changing 2 of the amino acids in the TALE protein changes the bp binding
  - Easy to engineer/modify the TALE domain
- Total recognition site ~16bp
- Same FokI dimer cutting



MegaTAL

- A combination of TAL and meganuclease
- TAL recognition domain + meganuclease cutting domain
- Why is this good?

The CRISPR system

- Discovered in bacteria! It is the bacteria's immune system against viruses
- Timeline of discovery: https://www.broadinstitute.org/what-broad/areas-focus/project-spotlight/crispr-timeline
- CRISPR - Clustered regularly-interspaced short palindromic repeats

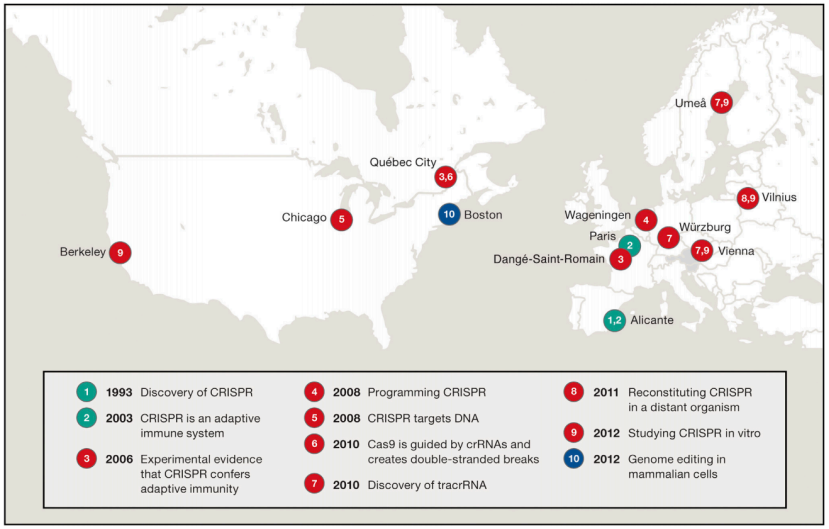
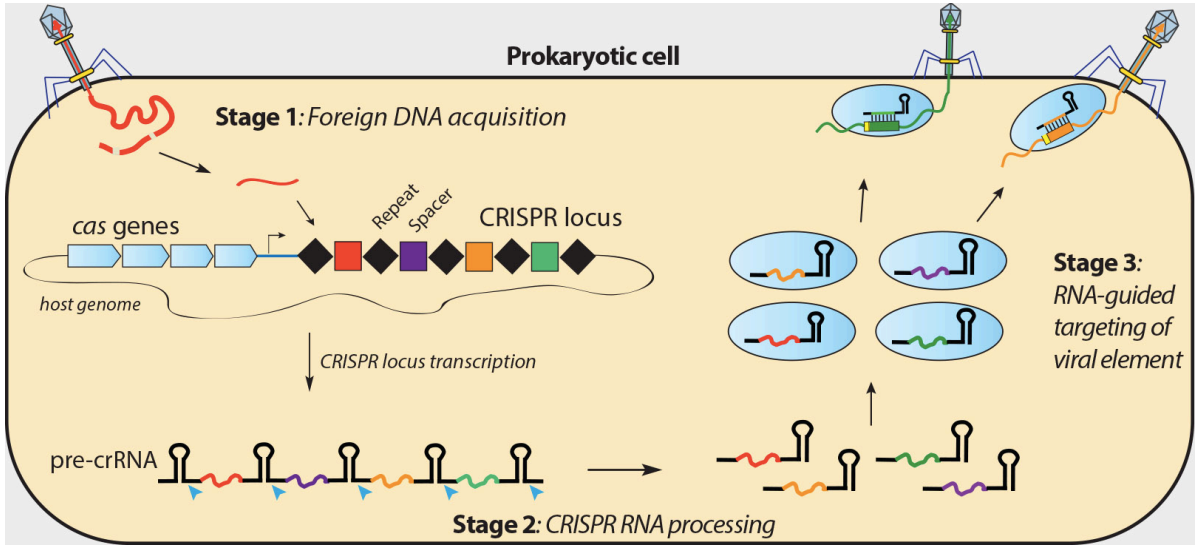


Figure 2. The Twenty-Year Story of CRISPR Unfolded across Twelve Cities in Nine Countries
For each "chapter" in the CRISPR "story," the map shows the sites where the primary work occurred and the first submission dates of the papers. Green circles refer to the early discovery of the CRISPR system and its function; red to the genetic, molecular biological, and biochemical characterization; and blue to the final step of biological engineering to enable genome editing.

https://www.broadinstitute.org/files/news/pdfs/PIIS0092867415017055.pdf

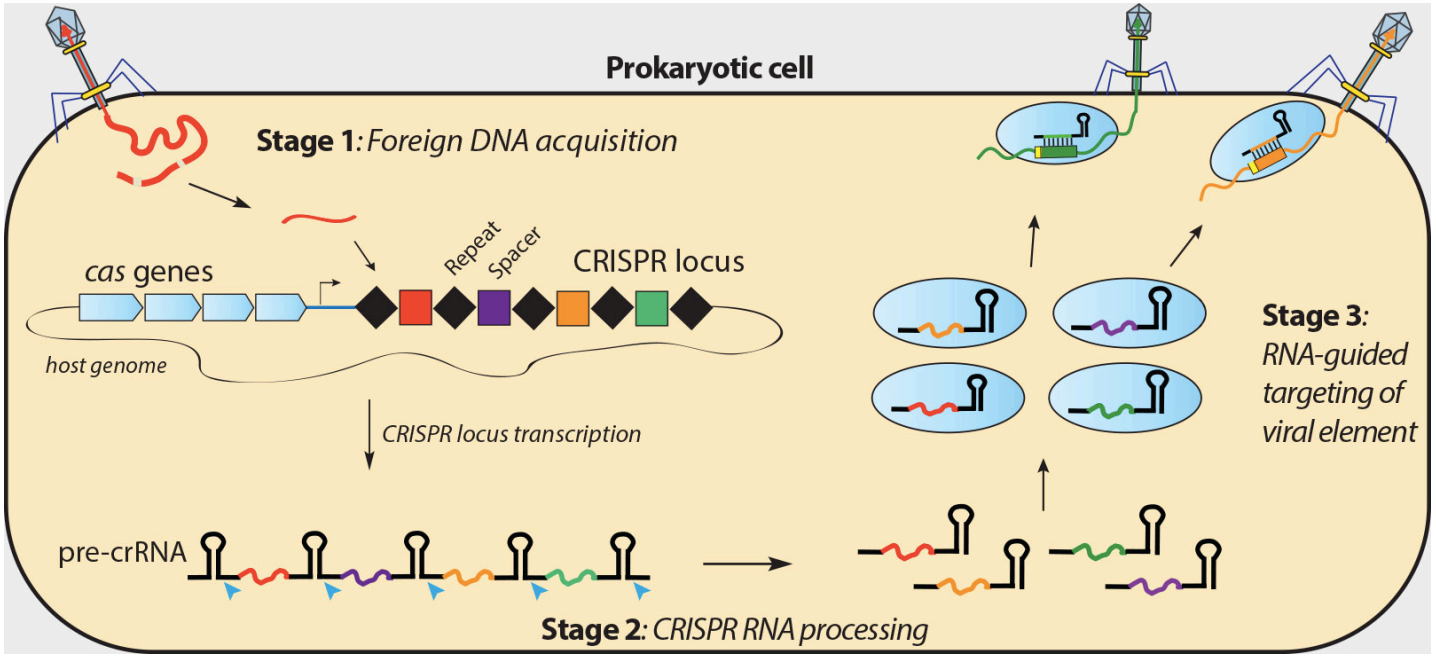
The CRISPR system

- Bacteria chews up the viral genome, and inserts the viral sequences into a "CRISPR locus/array" in its own genome
  - Creates a memory bank of viral sequences
- Only cuts virus around PAM sequences
  - PAM = Protospacer Adjacent Motifs, 3-5 bp long



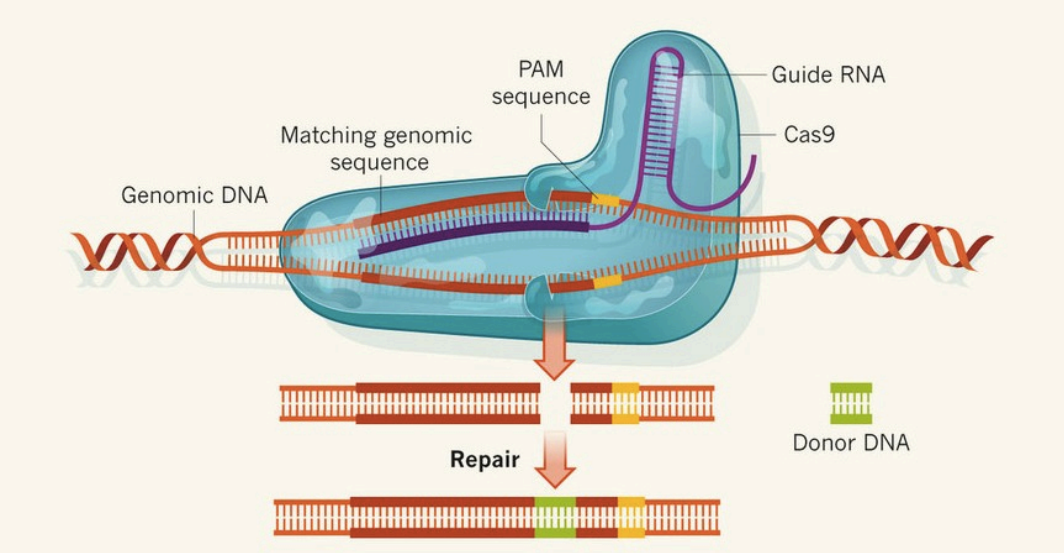
The CRISPR system

- CRISPR array is transcribed, and cut up by cas protein
- Different cas protein picks up the pieces containing the 'memory bank IDs'



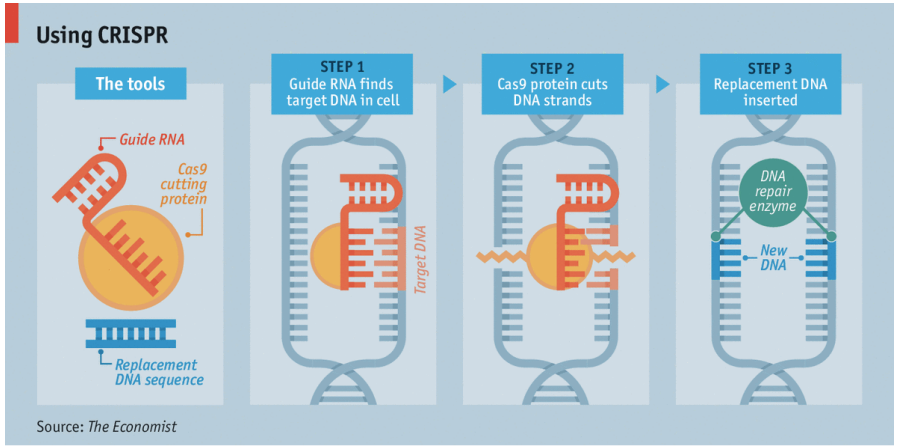
The CRISPR system

- Cas protein with transcribed viral RNA makes a complex
- Complementarity to invading viral sequences guides cas-RNA complex to destroy invader
  - sgRNA- homing device; cas9/Cpf – cleavage
- A good video to review: https://www.youtube.com/watch?v=MnYppmstxIs



The CRISPR system

- Cas9 and Cpf specificity depends on PAM + specific guide RNA sequence (sgRNA)
- Typical recognition site ~20 bp
- Genome editing application has following components
  - Cas/Cpf protein
  - sgRNA
  - Optional – replacement sequence dsDNA
- Easy to customize
- Off target effect is low
- Highly efficient
- Ability to multiplex

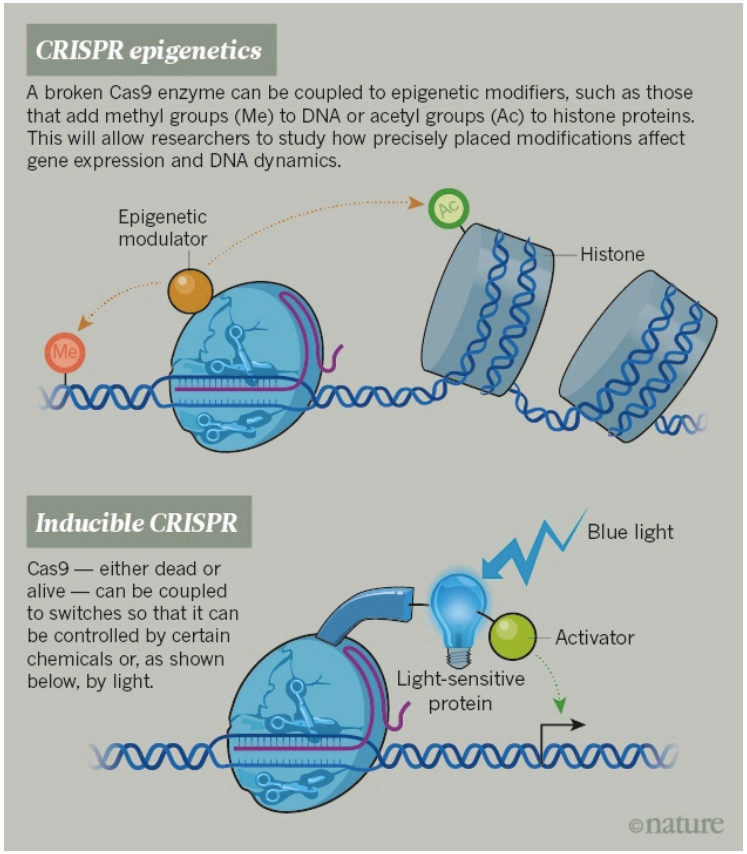


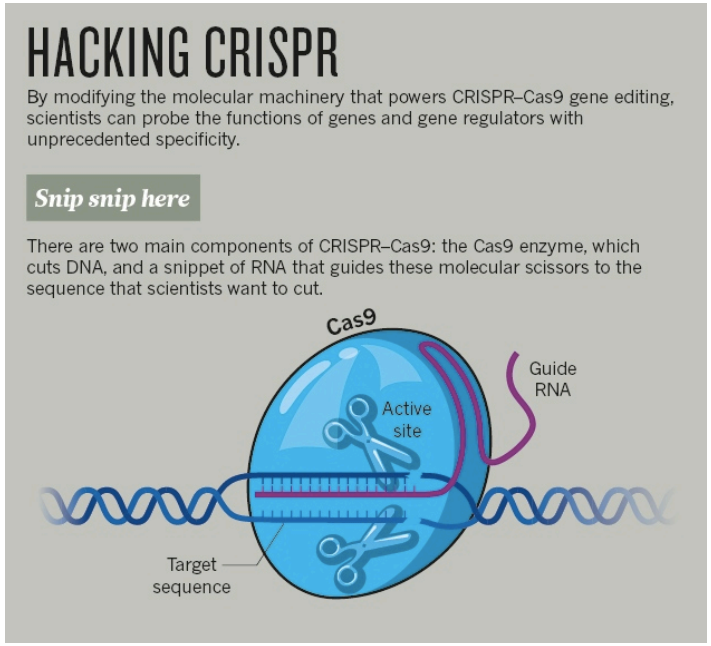
- https://youtu.be/2pp17E4E-O8?t=51

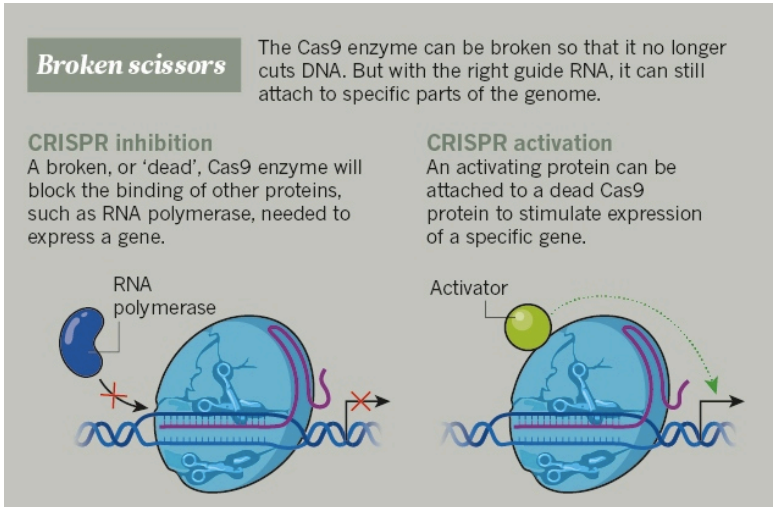
CRISPRi

- People worried about off-target cleavage
- Modified cas protein loses nuclease function, therefore binding only silences the transcription, but doesn't cut
  - Possibly reversible

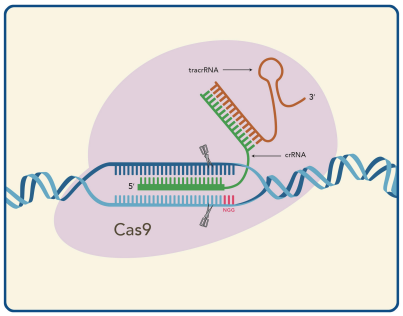
CRISPR variants

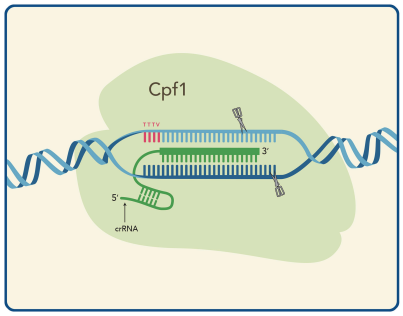


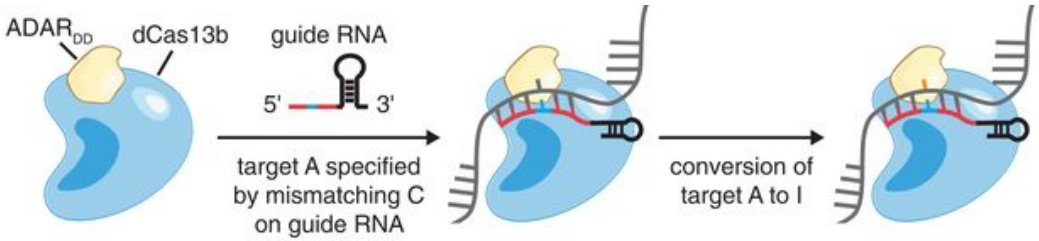




Feature	Cas9	Cpf1	Cas13
Target molecule	DNA	DNA	RNA
Structure	2 RNA required, or 1 fusion transcript (crRNA+tracrRNA=gRNA)	1 RNA required	1 RNA required + fusion protein
Cutting mechanism	Blunt end cuts	Staggered end cuts	Dead nuclease, no cutting
Cutting site	Proximal to recognition site	Distal from recognition site	-
Target sites	G-rich PAM	T-rich PAM	No specified PAM





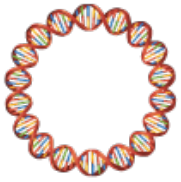


http://science.sciencemag.org/content/358/6366/1019.long
https://www.nature.com/articles/nature24049

Delivery is a key challenge

- How to deliver these nucleases into cells?
  - DNA (plasmid, small fragments) - must go into nucleus
  - RNA - must survive RNAses
  - Protein – larger size poses challenges

- Viral vector delivery?
- Electroporation?
- How to target cell types?

Cas9 Delivery Methods	pDNA	mRNA	Protein
			
High Efficiency	++++	++++	++++
Low Cost	++++	++++	++++
Specificity	++++	++++	++++

Story about CRISPR-cas9

- UC Berkeley

vs

MIT BROAD

Intellia THERAPEUTICS

About to IPO ~$120 MM

CRISPR THERAPEUTICS

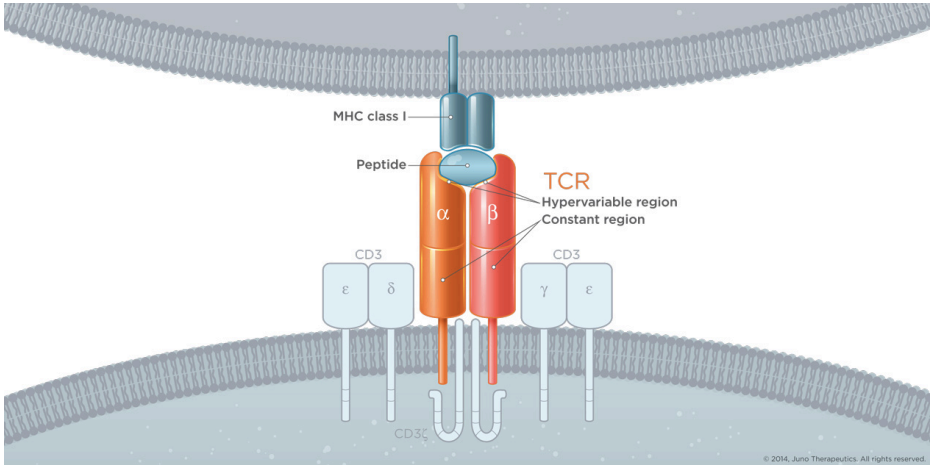
editas MEDICINE

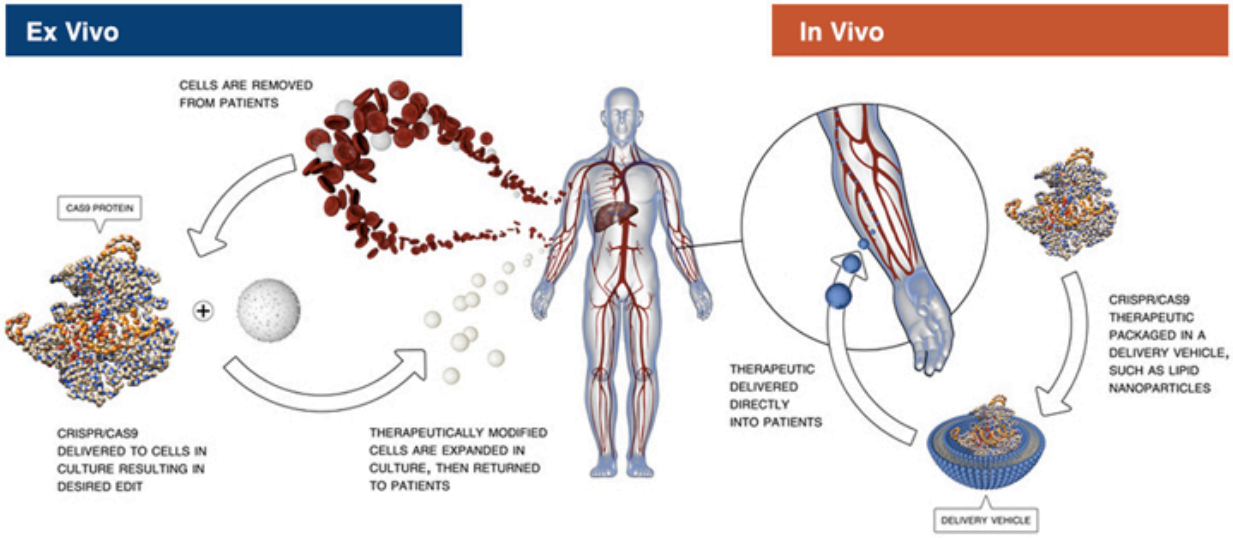
IPO $94 MM

CARIBOU BIOSCIENCES™

Applications of Genome editing

- CAR-T
- https://www.youtube.com/watch?v=TzSurHZjoy0



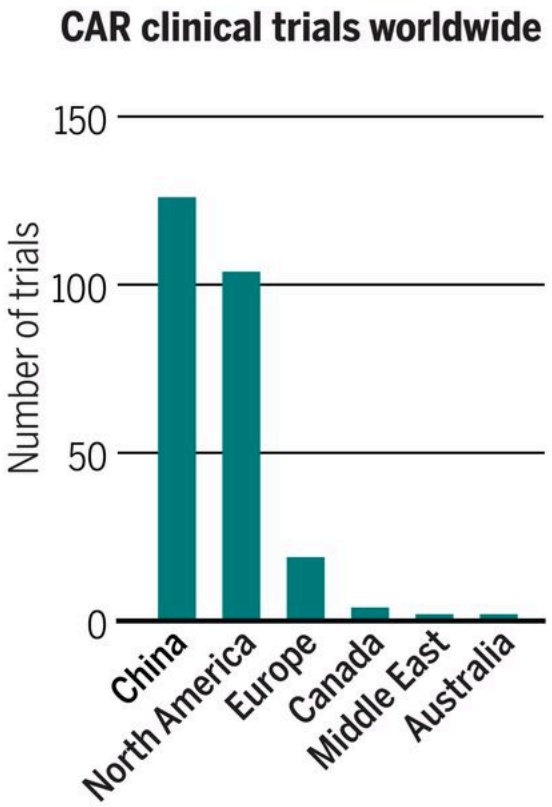


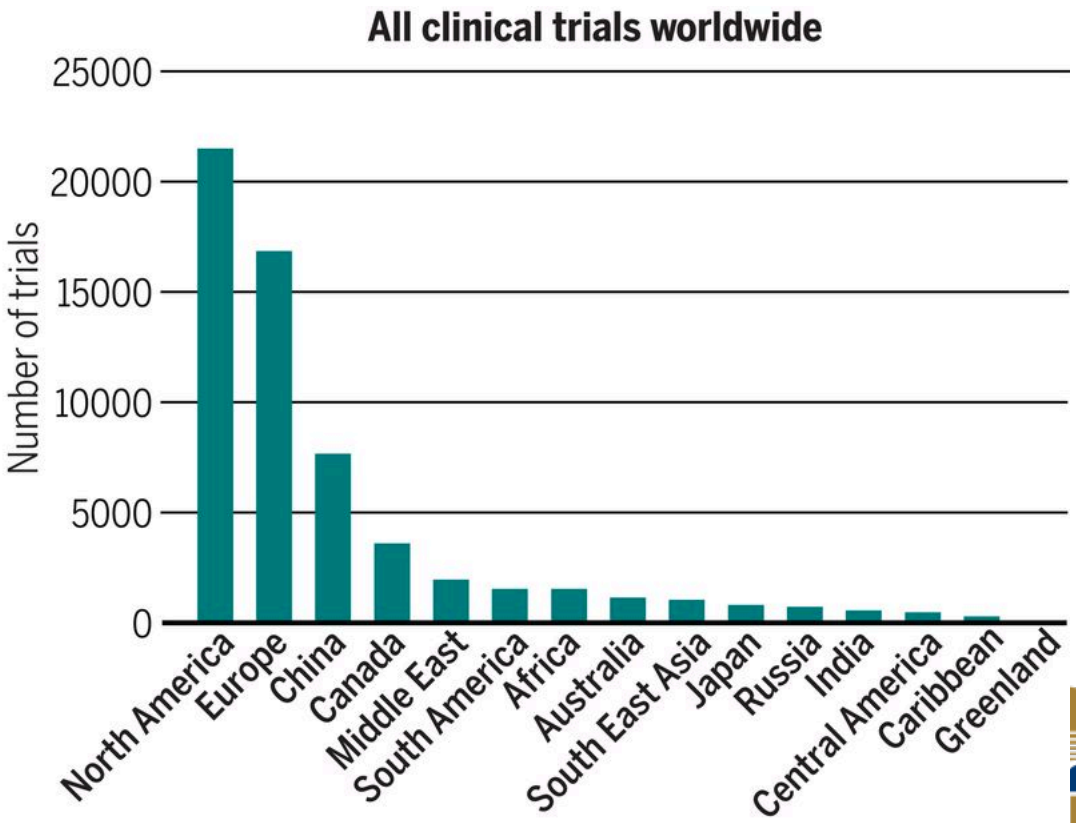
CAR-T trials: now many trials and several approved therapies

- Earliest trials circa March 2017 "Kite Pharma, a US pharmaceutical company, just released the groundbreaking results of their six-month gene therapy trial: terminal cancer patients in complete remission after just a single round."
- "Patients who participated in the trial had one of three types of non-Hodgkin lymphoma. Patients were all given only a few months to live. However, following the first round of gene therapy, which took place nine months after the trial began, half the patients are not only still alive, but a third of them appear to be cured."

https://futurism.com/4-gene-therapy-treatment-can-bring-terminal-cancer-patients-to-complete-remission/
http://ir.kitepharma.com/releasedetail.cfm?ReleaseID=1014817

CAR-T trials: many trials and several approved therapies



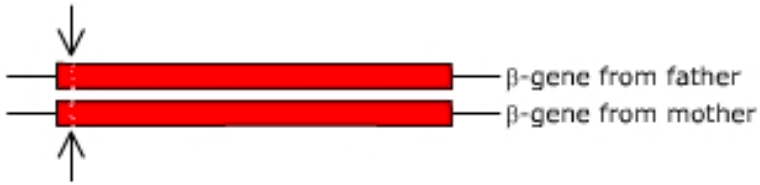


https://science.sciencemag.org/content/359/6382/1361

Treating beta-thalassemia

April 2015 (Sun Yat-sen University, China): CRISPR used to edit a beta- thalassemia gene in non-viable human embryos

With a mutation on one of the two ß-globin genes, a carrier is formed with lower protein production, but enough hemoglobin



Without a mutation enough Hemoglobin

No thalassemia carrier

With one mutation less Hemoglobin

β-thalassemia carrier without illness, but less hemoglobin (slight aneamia)

With two mutations no β-globin

β-thalassemia major patient with severe aneamia

CRISPR prophylactic antiviral strategy

New Results

Comment on this paper

Development of CRISPR as a prophylactic strategy to combat novel coronavirus and influenza

Timothy R. Abbott, Girija Dhamdhere, Yanxia Liu, Xueqiu Lin, Laine Goudy, Leiping Zeng, Augustine Chemparathy, Stephen Chmura, Nicholas S. Heaton, Robert Debs, Tara Pande, Drew Endy, Marie La Russa, David B. Lewis, Lei S. Qi

doi: https://doi.org/10.1101/2020.03.13.991307

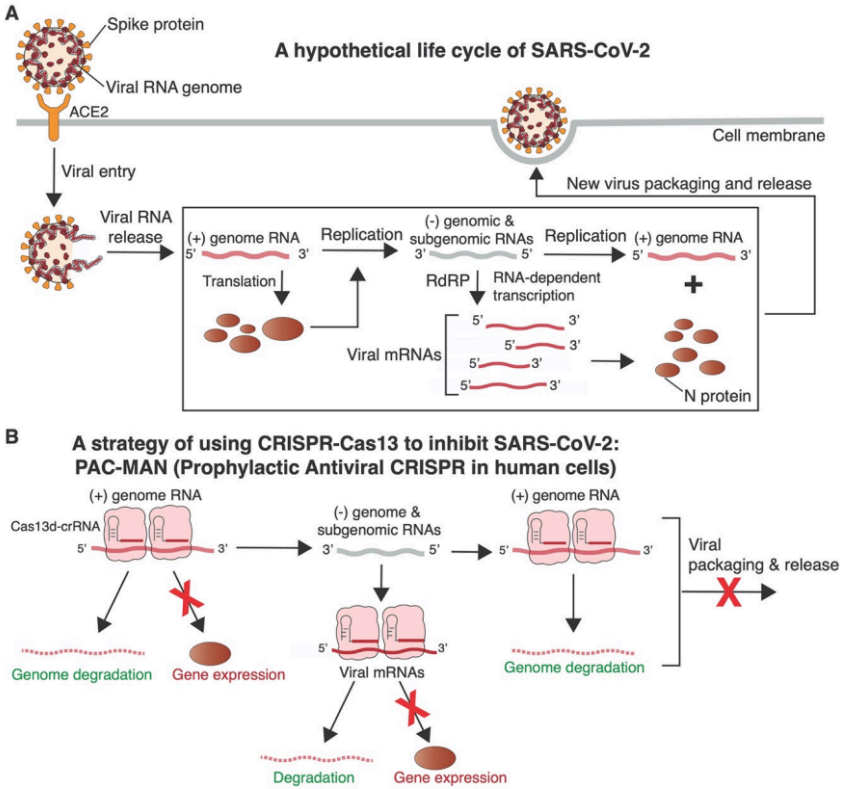


Figure 1 | The hypothetical life cycle of SARS-CoV-2 and the PAC-MAC approach for anti-COVID-19. (A) A hypothetical life cycle of SARS-CoV-2 based on what is known about other coronavirus life cycles. SARS-CoV-2 virions bind to the ACE2 receptor on the surface of cells via interactions with the Spike protein. Upon viral release, the positive strand RNA genome serves as a template to make negative strand genomic and subgenomic templates, which are used to produce more copies of the positive strand viral genome and viral mRNAs. **(B)** Cas13d can inhibit viral function and replication by directly targeting and cleaving all viral positive-sense RNA.

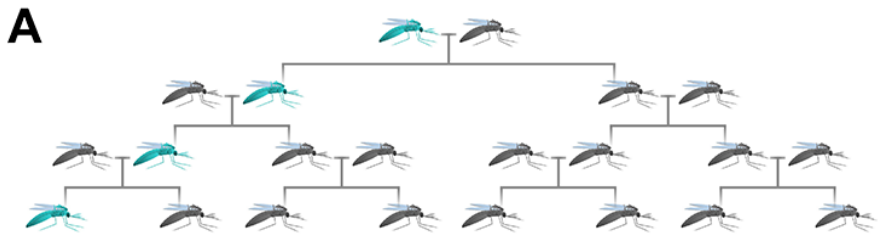
https://www.biorxiv.org/content/10.1101/2020.03.13.991307v1.full.pdf+html

Latest CRISPR trial – Leber congenital amaurosis, first in-vivo CRISPR treatment

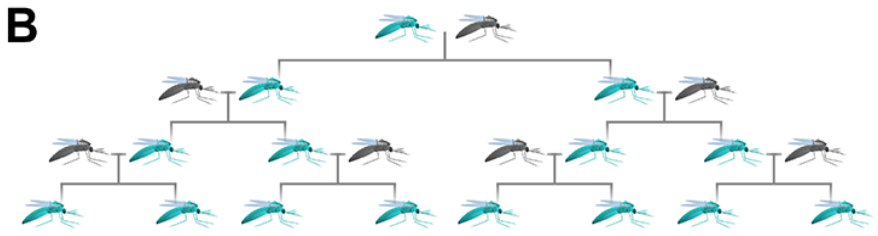
- "AGN-151587 (EDIT-101) is an experimental medicine delivered via sub-retinal injection under development for the treatment of Leber congenital amaurosis 10 (LCA10), an inherited form of blindness caused by mutations in the centrosomal protein 290 (CEP290) gene. The BRILLIANCE clinical trial is a Phase 1/2 study to evaluate AGN-151587 for the treatment of patients diagnosed with LCA10 and is the world's first human study of an in vivo, or inside the body, CRISPR genome editing medicine. The trial will assess the safety, tolerability, and efficacy of AGN-151587 in approximately 18 patients with LCA10."
- "LCA10, is a monogenic disorder caused by mutations in the CEP290 gene and is the cause of disease in approximately 20-30 percent of all LCA patients."
- Uses AAV – adeno-associated virus
- Uses Cas9

Non-human genome editing: gene drives

- Gene drive:
  - Normally, each allele has a 50% chance of being inherited by offspring
  - Genetic systems that violate these rules and increase the probability that a particular gene will be passed to offspring is a gene drive
  - Increased % of inheritance means the gene can be spread to entire population, even if it reduces chance of reproduction



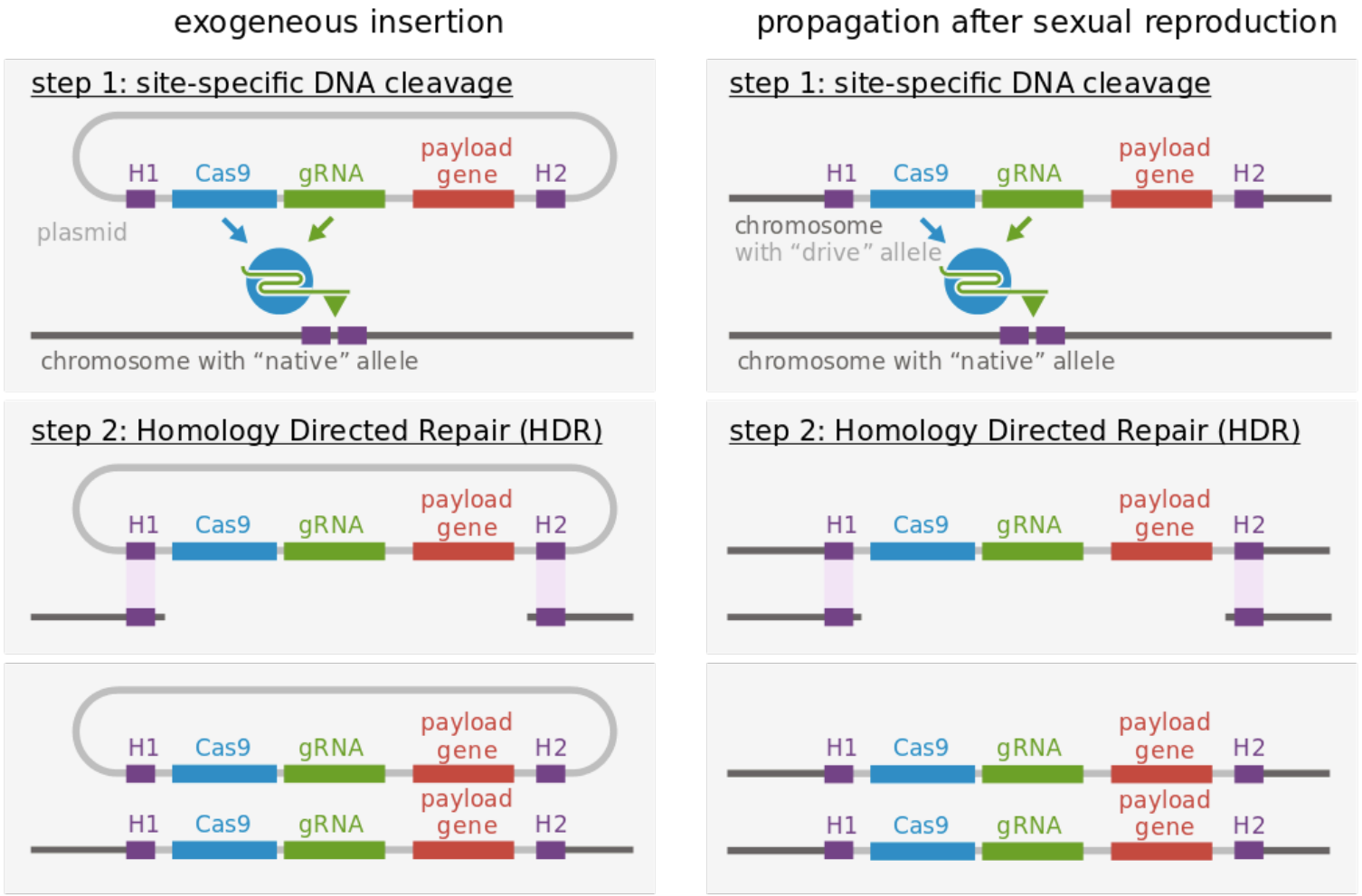
normal



with gene drive

https://wyss.harvard.edu/staticfiles/newsroom/pressreleases/Gene%20drives%20FAQ%20FINAL.pdf

Non-human genome editing: gene drives



https://en.wikipedia.org/wiki/Gene_drive

Additional fun resources to learn about synthetic biology and commentary on it

- Commentary – Five Hard Truths about Synthetic Biology: https://www.nature.com/news/2010/100120/full/463288a.html
- Fun podcast on synthetic biology and genome editing with George Church: https://boingboing.net/2018/04/03/the-astounding-present-and-diz.html

'I want to help humans genetically modify themselves'

Former Nasa biochemist Josiah Zayner became an online sensation by conducting DIY gene therapy on himself. He explains why he did it

▲ Josiah Zayner with his Crispr gene-editing kit. Photograph: Courtesy Josiah Zayner / The ODIN

Josiah Zayner, 36, recently made headlines by becoming the first person to use the revolutionary gene-editing tool Crispr to try to change their own genes. Part way through a talk on genetic engineering, Zayner pulled out a syringe apparently containing DNA and other chemicals designed to trigger a genetic change in his cells associated with dramatically increased muscle mass. He injected the DIY gene therapy into his left arm, live-streaming the procedure on the internet.

https://www.theguardian.com/science/2017/dec/24/josiah-zayner-diy-gene-editing-therapy-crispr-interview

https://futurism.com/bad-idea-biohacker/