RUHR-UNIVERSITÄT BOCHUM

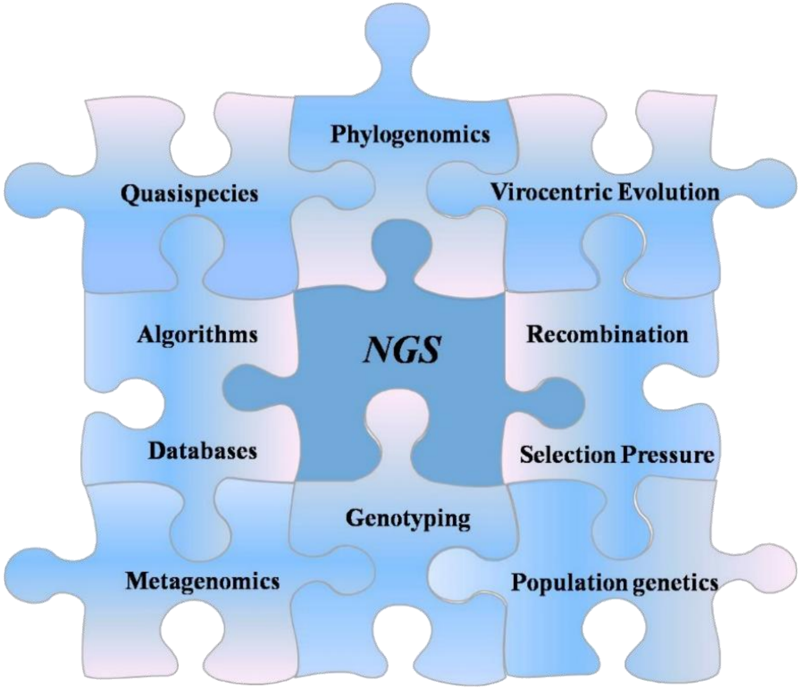
RUB

Next generation sequencing in Virology

Daniel Todt

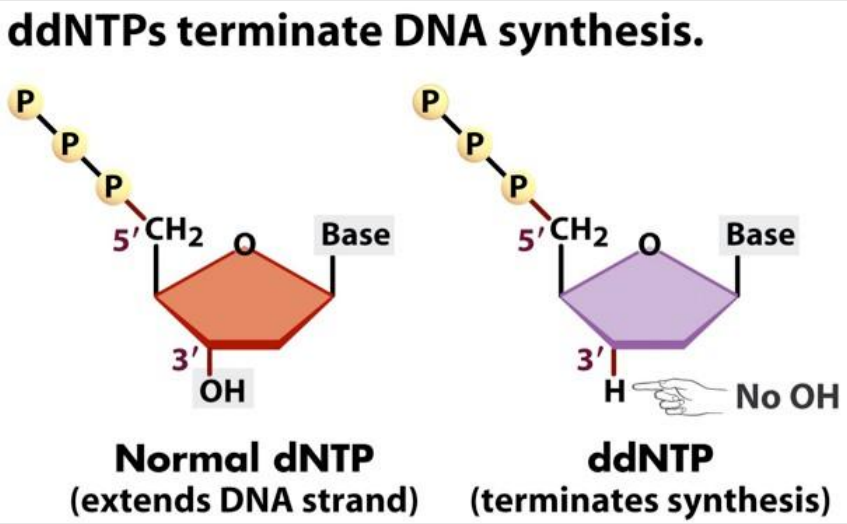
TRACiR

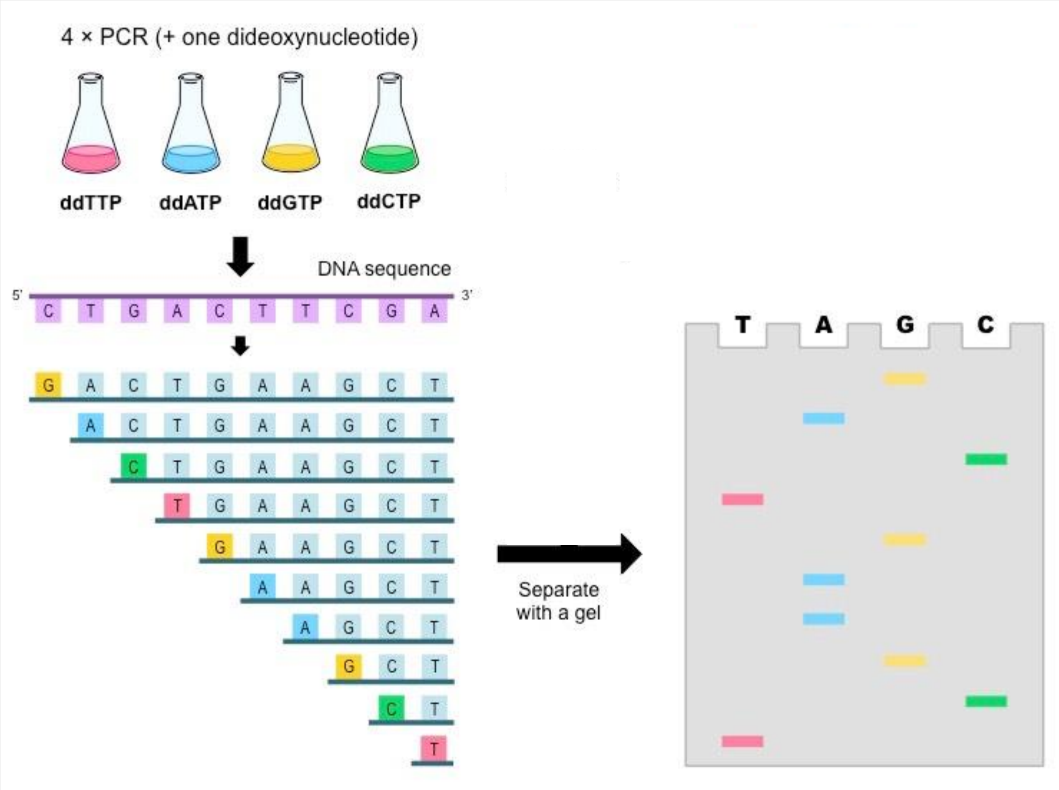
TRANSLATIONAL AND COMPUTATIONAL INFECTION RESEARCH



First generation sequencing

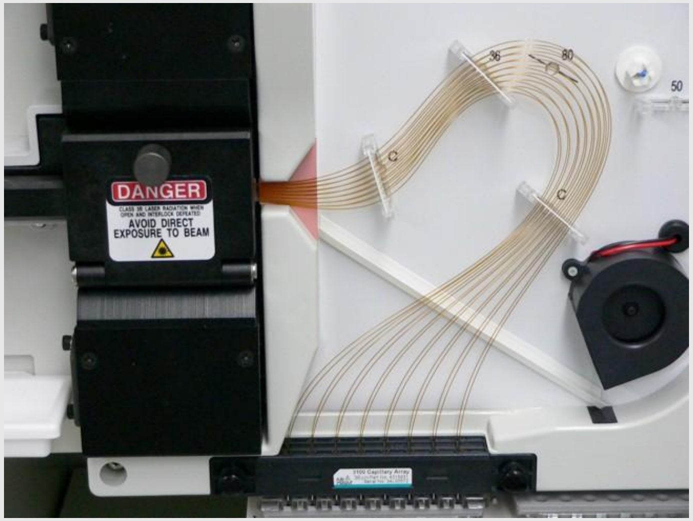
Sanger Sequencing (dideoxy chain termination)

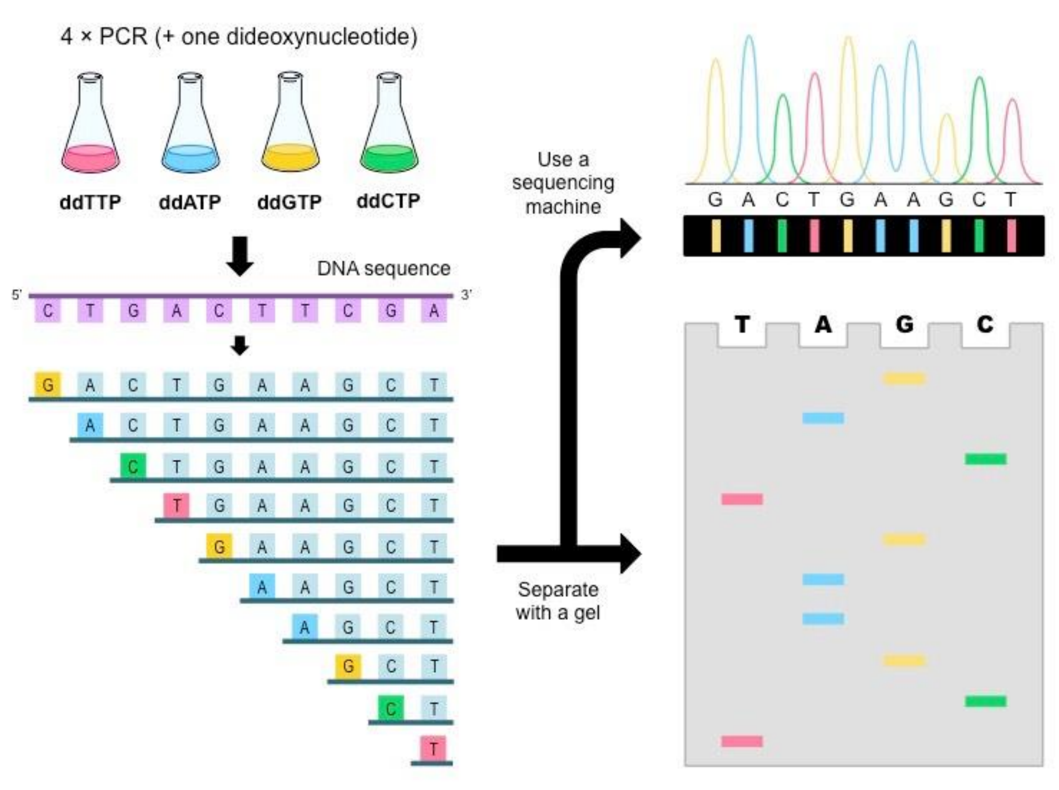




- Developed in 1977 by Fred Sanger
- DNA extended from **radiolabelled** primers using a mix of dNTP and ddNTP nucleotides
- Random **chain termination** upon ddNTP incorporation
- **Separate reaction** for each terminator (ddC-ddT-ddA-ddG)
- DNA fragments resolved on large **polyacrylamide gels** and detected on film by **autoradiography**
- Sequence **read by hand** and typed in
- Labour intensive, slow and expensive

Sanger Sequencing





- Automation developed by Leroy Hood and **Applied BioSystems**
- Improved using **fluorescently-labelled** ddNTP terminators (**one reaction** per sequence)
- Separated by **capillary electrophoresis** in automated sequencing machine
- **Long reads (up to 1100 bp) and low error rate**
- Limited throughput, **expensive** per base but still in wide use.

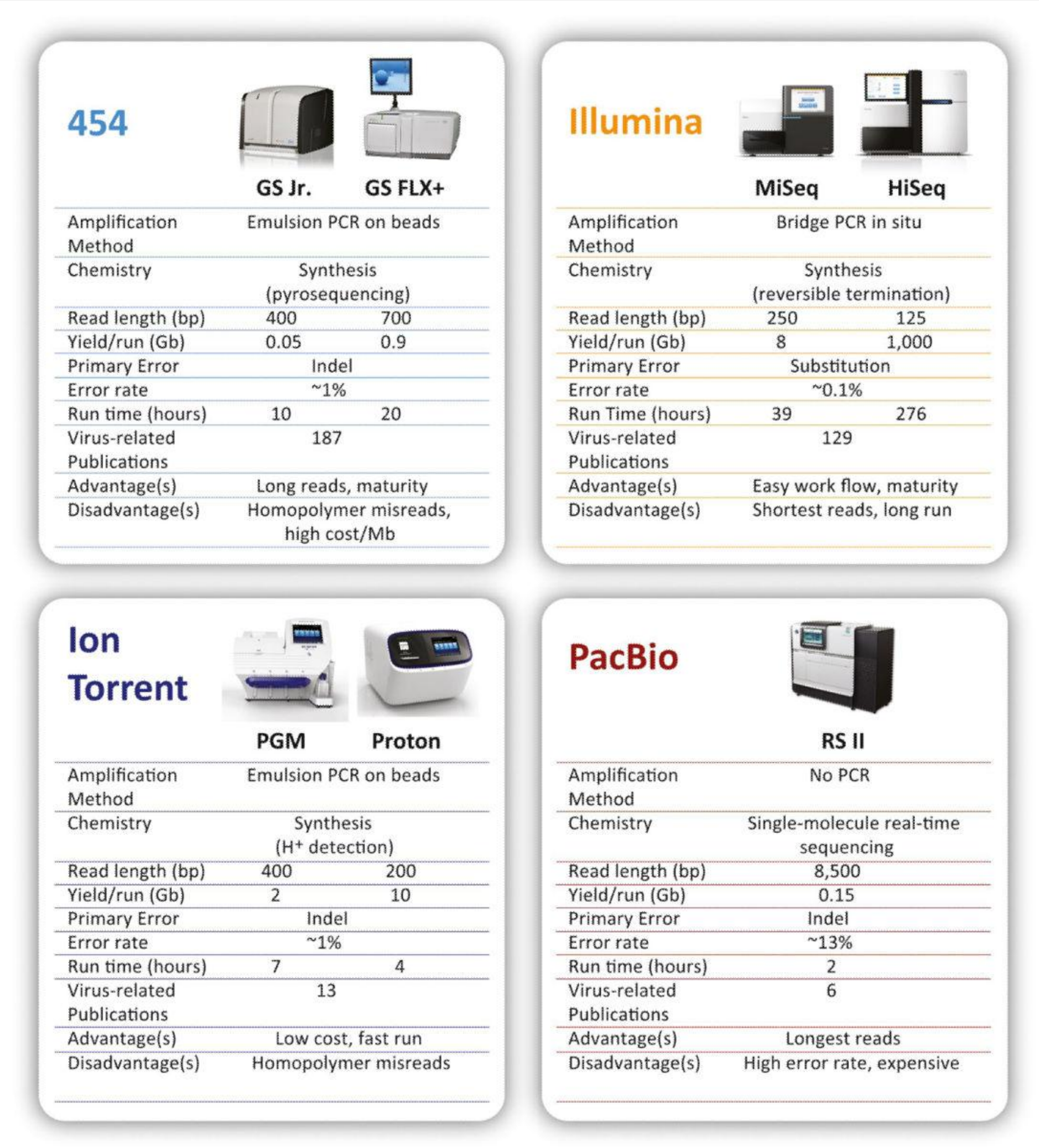
Second generation sequencing

Second Generation Sequencing – High Throughput Sequencing (HTS)

- General characteristics:
  - DNA molecules from a single library are clustered on a planar substrate (**bridge PCR**), or to the surface of micron-scale beads (**emulsion PCR**).
  - Sequencing by **synthesis** or by **ligation**.

- Advantages over first generation sequencing:
  - *In vitro* **clonal amplification** circumvents time consuming steps such as ligation of DNA fragments into a plasmid, transformation of E. coli and colony picking.
  - Array-based sequencing enables higher degree of **parallelism** than conventional capillary-based sequencing.

Principal characteristics of the four most used deep sequencing platforms

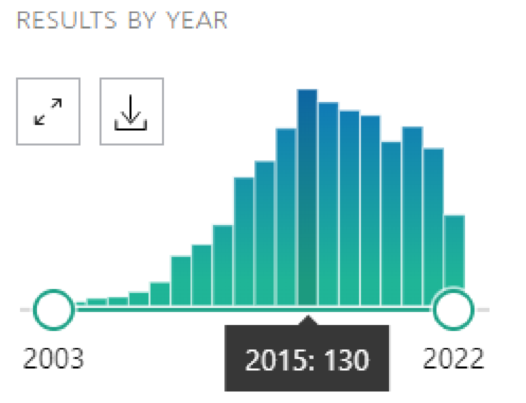


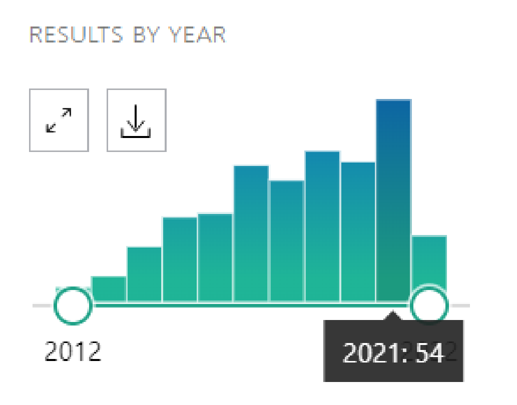
454	GS Jr.	GS FLX+
Amplification Method	Emulsion PCR on beads	
Chemistry	Synthesis (pyrosequencing)	
Read length (bp)	400	700
Yield/run (Gb)	0.05	0.9
Primary Error	Indel	
Error rate	~1%	
Run time (hours)	10	20
Virus-related Publications	187	
Advantage(s)	Long reads, maturity	
Disadvantage(s)	Homopolymer misreads, high cost/Mb	

Illumina	MiSeq	HiSeq
Amplification Method	Bridge PCR in situ	
Chemistry	Synthesis (reversible termination)	
Read length (bp)	250	125
Yield/run (Gb)	8	1,000
Primary Error	Substitution	
Error rate	~0.1%	
Run Time (hours)	39	276
Virus-related Publications	129	
Advantage(s)	Easy work flow, maturity	
Disadvantage(s)	Shortest reads, long run	

Ion Torrent	PGM	Proton
Amplification Method	Emulsion PCR on beads	
Chemistry	Synthesis (H^+ detection)	
Read length (bp)	400	200
Yield/run (Gb)	2	10
Primary Error	Indel	
Error rate	~1%	
Run time (hours)	7	4
Virus-related Publications	13	
Advantage(s)	Low cost, fast run	
Disadvantage(s)	Homopolymer misreads	

PacBio	RS II
Amplification Method	No PCR
Chemistry	Single-molecule real-time sequencing
Read length (bp)	8,500
Yield/run (Gb)	0.15
Primary Error	Indel
Error rate	~13%
Run time (hours)	2
Virus-related Publications	6
Advantage(s)	Longest reads
Disadvantage(s)	High error rate, expensive

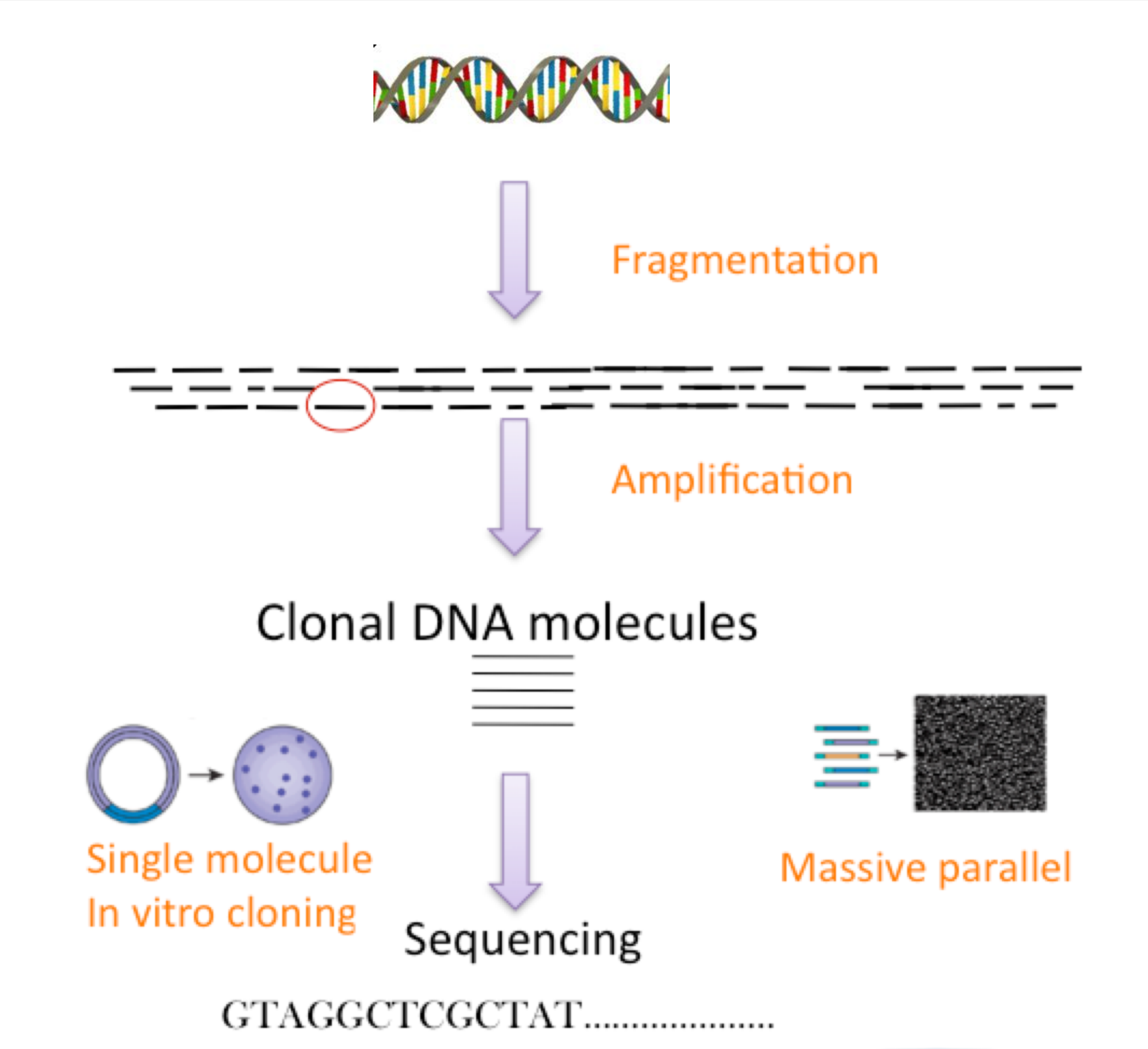




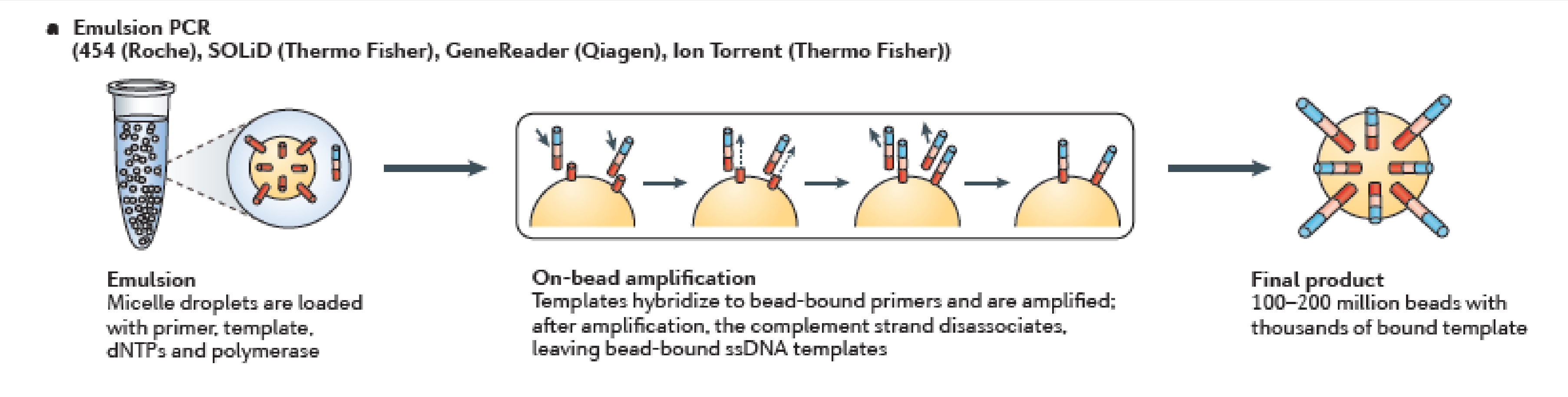
Principal characteristics of the four most used deep sequencing platforms

Method	Read length	Accuracy (single read not consensus)	Reads per run	Time per run	Cost per 1 million bases (in US$)	Advantages	Disadvantages
Single-molecule real-time sequencing (**Pacific Biosciences**)	10,000 bp to 15,000 bp	87% single-read accuracy	500–1000 megabases	30 minutes to 4 hours	$0.13–$0.60	Longest read length. Fast. Detects 4mC, 5mC, 6mA.	Moderate throughput. Equipment can be very expensive.
Ion semiconductor (**Ion Torrent** sequencing)	up to 400 bp	98%	up to 80 million	2 hours	$1	Less expensive equipment. Fast.	Homopolymer errors.
Pyrosequencing (**454**)	700 bp	99.9%	1 million	24 hours	$10	Long read size. Fast.	Runs are expensive. Homopolymer errors.
Sequencing by synthesis (**Illumina**)	MiSeq: 50-600 bp HiSeq: 50-500 bp	99.9% (Phred30)	MiSeq: 1-25 Million; HiSeq: 300 million - 2 billion,	1 to 11 days, depending upon sequencer and specified read length	$0.05 to $0.15	Potential for high sequence yield, depending upon sequencer model and desired application.	Equipment can be very expensive. Requires high concentrations of DNA.
Sequencing by ligation (**SOLiD** sequencing)	50+35 or 50+50 bp	99.9%	1.2 to 1.4 billion	1 to 2 weeks	$0.13	Low cost per base.	Slower than other methods. Has issues sequencing palindromic sequences.
Chain termination (**Sanger** sequencing)	400 to 900 bp	99.9%	N/A	20 minutes to 3 hours	$2400	Long individual reads. Useful for many applications.	More expensive and impractical for larger sequencing projects. This method also requires the time consuming step of plasmid cloning or PCR.

How does high-throughput sequencing work?



Pyrosequencing



a Emulsion PCR
(454 (Roche), SOLiD (Thermo Fisher), GeneReader (Qiagen), Ion Torrent (Thermo Fisher))

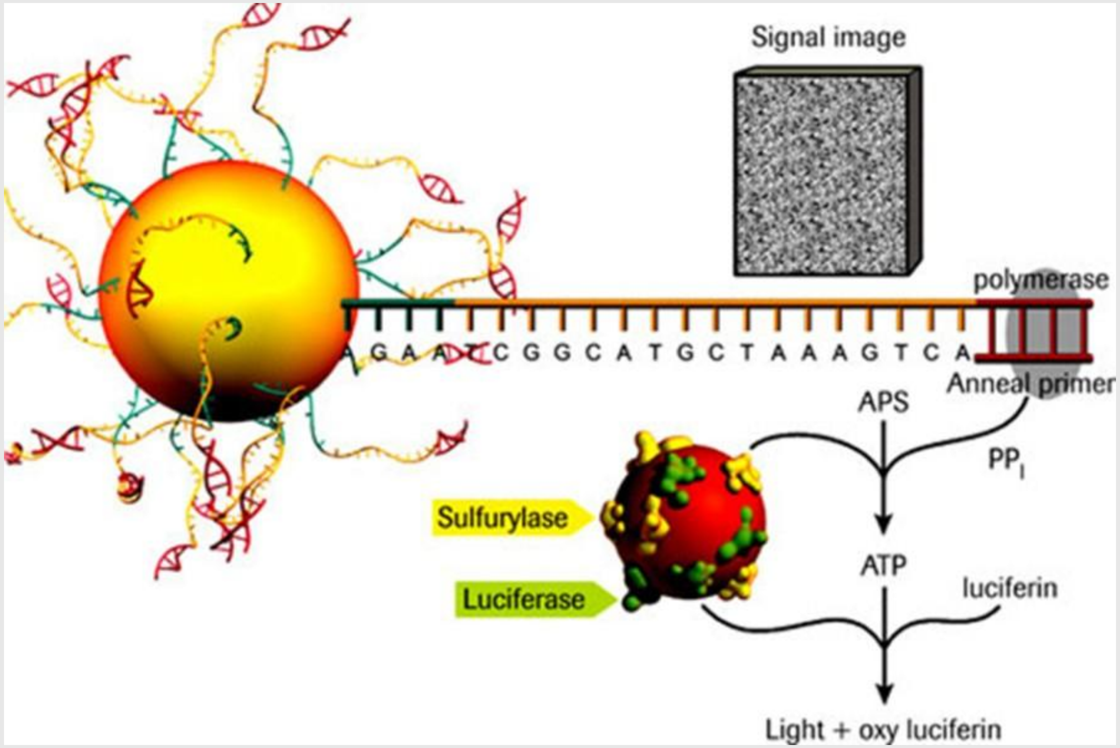
Emulsion
Micelle droplets are loaded with primer, template, dNTPs and polymerase

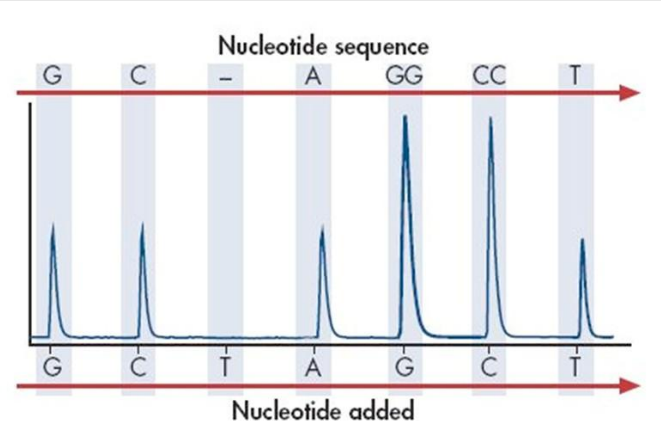
On-bead amplification
Templates hybridize to bead-bound primers and are amplified; after amplification, the complement strand disassociates, leaving bead-bound ssDNA templates

Final product
100–200 million beads with thousands of bound template

- Developed in Uppsala, Sweden. Later acquired by Qiagen, then licensed to Life Sciences (**454**)
  - DNA fragmentation
  - **Adapters** ligated to DNA fragments (**biotin** tag)
  - Bound to **Streptavidin** beads (each fragment, one bead)
  - Amplified by **emulsion PCR**
  - Beads deposited into **separate wells** on PicoTitrePlate with separate pyrosequencing reaction in each well, in a large-scale parallel pyrosequencing system.

Pyrosequencing



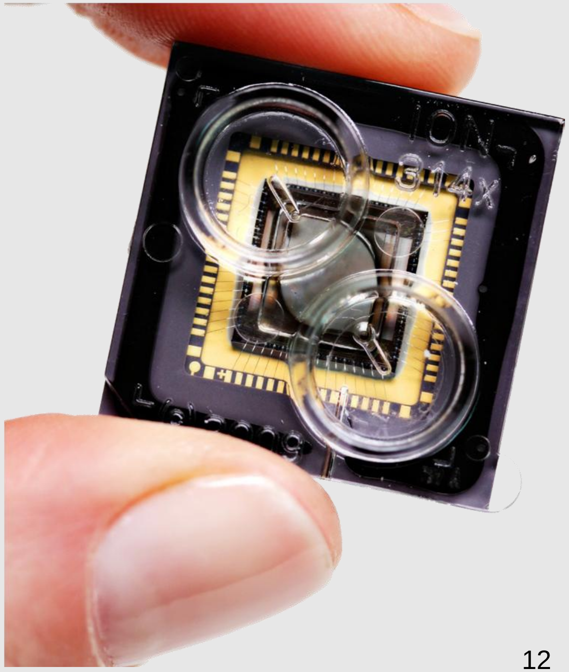


Discontinued in 2014!!!

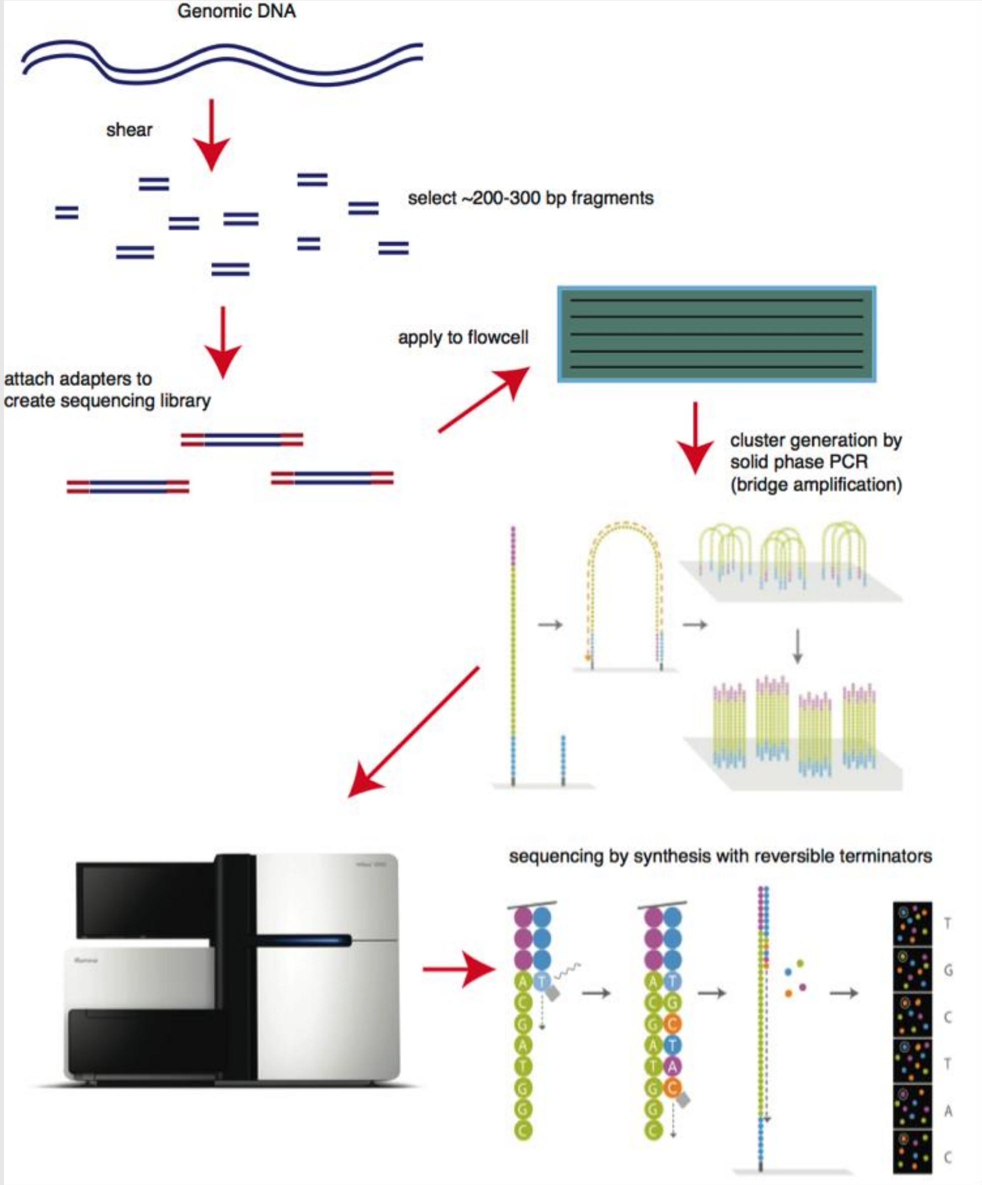
- **Sequencing by synthesis method**
- G-C-T-A nucleotides are added sequentially, dNTP incorporation releases **pyrophosphate** (PPi)
- ATP **sulfurylase** converts dNTP to ATP, acts as a substrate for the **luciferase**
- Generates **light** in amounts that are proportional to the amount of PPi (homopolymer error!)
- Unincorporated nucleotides and ATP are degraded by the **apyrase**
- **Light signal** recorded on camera

Ion Torrent

- Licensed from DNA Electronics Ltd, developed by **Ion Torrent Systems**. Later bought over by **Life Technologies**
  - **Adapters** ligated to DNA fragments
  - Bound to **beads**, amplified by **emulsion PCR**
  - Beads deposited into separate wells on **semiconductor chip** with A-T-C-G nucleotides are added **sequentially**
  - **Sequencing by synthesis**
  - Nucleotide incorporation **releases a proton** and the pH of the well changes. A sensing layer detects the change and translates the chemical signal to a digital signal. (Avoids using optical sensors or fluorescent nucleotides; still with homopolymer errors)



Illumina

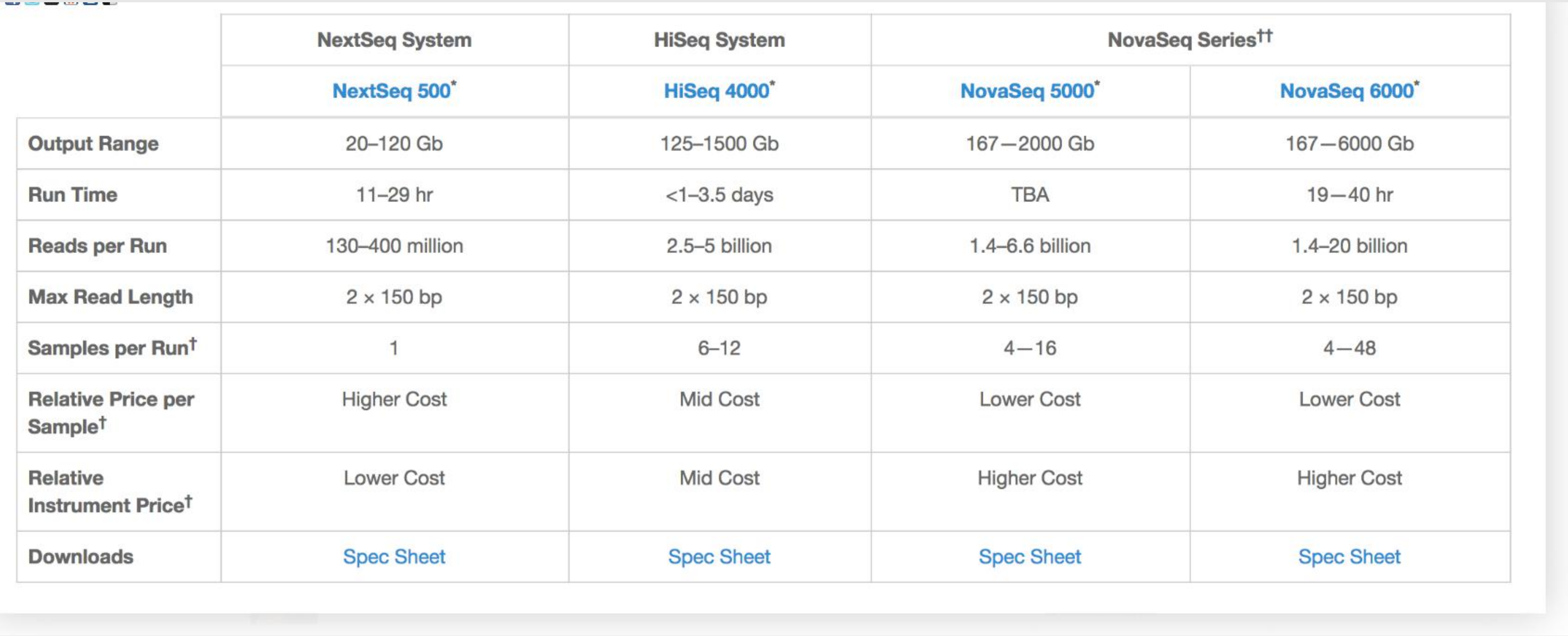

Genomic DNA
shear
select ~200-300 bp fragments
attach adapters to create sequencing library
apply to flowcell
cluster generation by solid phase PCR (bridge amplification)
sequencing by synthesis with reversible terminators
T
G
C
T
A
C

- Developed by Balasubramanian and Klenerman who founded Solexa, later acquired by Illumina
  - **Adapters** ligated to DNA fragments
  - **Flow cell** – glass slide with oligos matching adapters
  - Captured DNA replicated through **bridge amplification** to make identical ‘colonies’
  - **Fluorescent reversible terminators** passed over flow cell
  - **Image** captured, terminator and dye removed (better performance with homopolymers)
  - barcoding and UMI for multiplexing

Illumina

- Low error rate, read lengths have increased to ≥300 bp.
- Currently used for vast majority of sequencing
- Range of machines with different throughput and cost
- Run time is slower than Ion Torrent (days compared to hours)
- Low error rate – 0.1%
- Single or paired end reads

Illumina



	NextSeq System	HiSeq System	NovaSeq Series††	
	NextSeq 500*	HiSeq 4000*	NovaSeq 5000*	NovaSeq 6000*
Output Range	20–120 Gb	125–1500 Gb	167—2000 Gb	167—6000 Gb
Run Time	11–29 hr	<1–3.5 days	TBA	19—40 hr
Reads per Run	130–400 million	2.5–5 billion	1.4–6.6 billion	1.4–20 billion
Max Read Length	2 x 150 bp	2 x 150 bp	2 x 150 bp	2 x 150 bp
Samples per Run†	1	6–12	4—16	4—48
Relative Price per Sample†	Higher Cost	Mid Cost	Lower Cost	Lower Cost
Relative Instrument Price†	Lower Cost	Mid Cost	Higher Cost	Higher Cost
Downloads	Spec Sheet	Spec Sheet	Spec Sheet	Spec Sheet

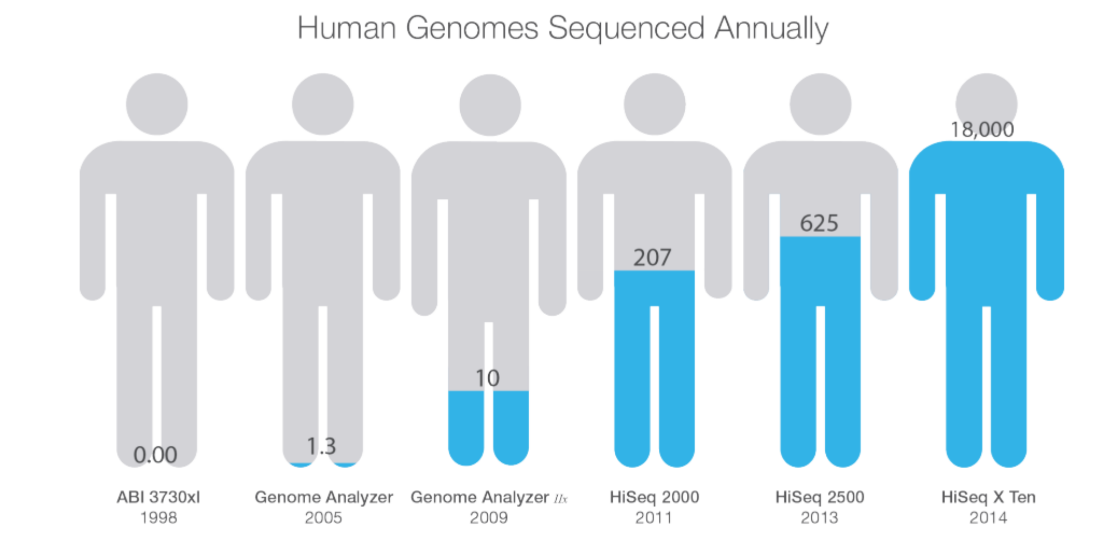
MGI

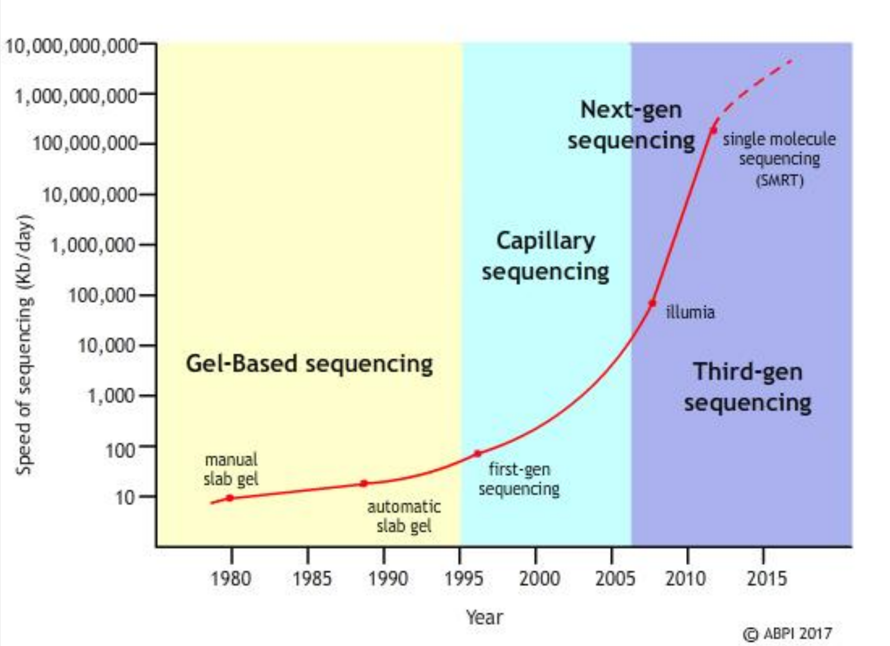
High Speed: 22 hrs ~24 hrs for PE150 sequencing

High Flexibility: 4 Flow Cells, PE150, and PE100 at the same time

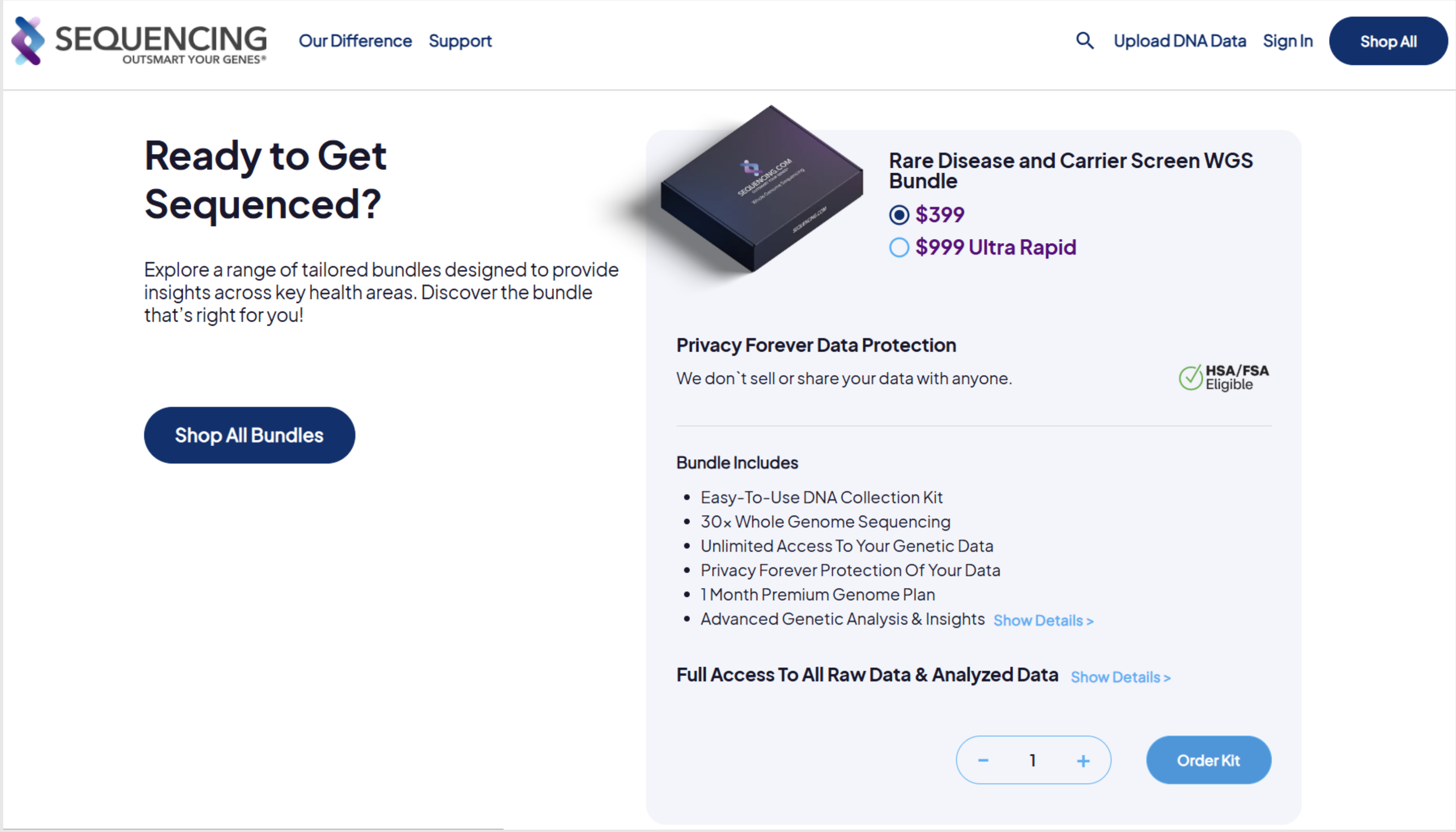
Ultra-high Throughput: 7 Tb per day; high quality data around the clock

Illumina


Human Genomes Sequenced Annually
0.00
ABI 3730xl
1998
1.3
Genome Analyzer
2005
10
Genome Analyzer IIx
2009
207
HiSeq 2000
2011
625
HiSeq 2500
2013
18,000
HiSeq X Ten
2014


Speed of sequencing (Kb/day)
10,000,000,000
1,000,000,000
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
100
10
Gel-Based sequencing
Capillary sequencing
Next-gen sequencing
Third-gen sequencing
manual slab gel
automatic slab gel
first-gen sequencing
illumia
single molecule sequencing (SMRT)
1980
1985
1990
1995
2000
2005
2010
2015
Year
© ABPI 2017

Illumina



Ready to Get Sequenced?

Explore a range of tailored bundles designed to provide insights across key health areas. Discover the bundle that's right for you!

Rare Disease and Carrier Screen WGS Bundle

● $399

○ $999 Ultra Rapid

Privacy Forever Data Protection

We don`t sell or share your data with anyone.

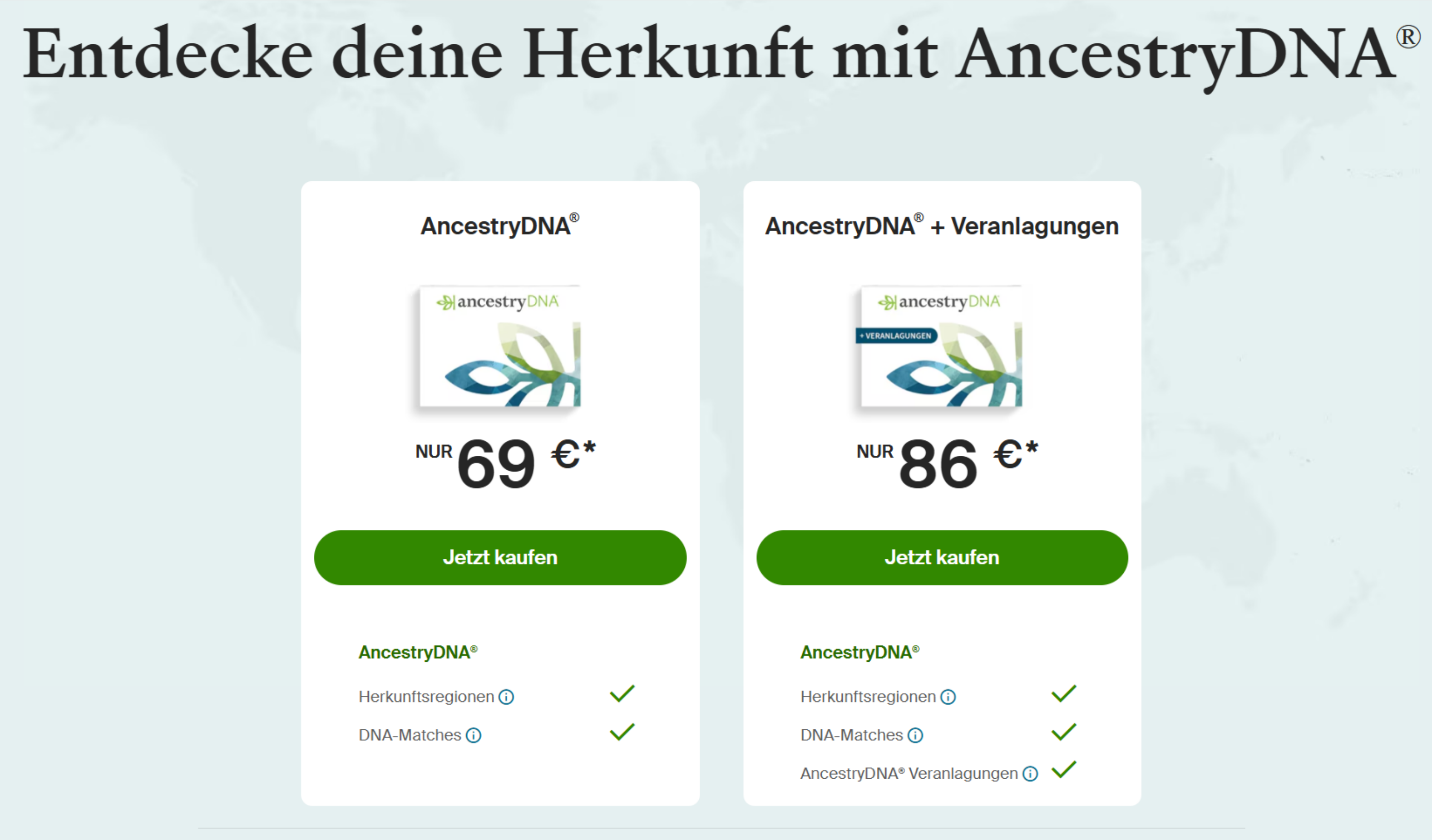
✓ HSA/FSA Eligible

Bundle Includes

- Easy-To-Use DNA Collection Kit
- 30x Whole Genome Sequencing
- Unlimited Access To Your Genetic Data
- Privacy Forever Protection Of Your Data
- 1 Month Premium Genome Plan
- Advanced Genetic Analysis & Insights Show Details >

Full Access To All Raw Data & Analyzed Data Show Details >

Illumina



Entdecke deine Herkunft mit AncestryDNA®

AncestryDNA®

NUR 69 €*

Jetzt kaufen

AncestryDNA®

Herkunftsregionen ⓘ ✓

DNA-Matches ⓘ ✓

AncestryDNA® + Veranlagungen

NUR 86 €*

Jetzt kaufen

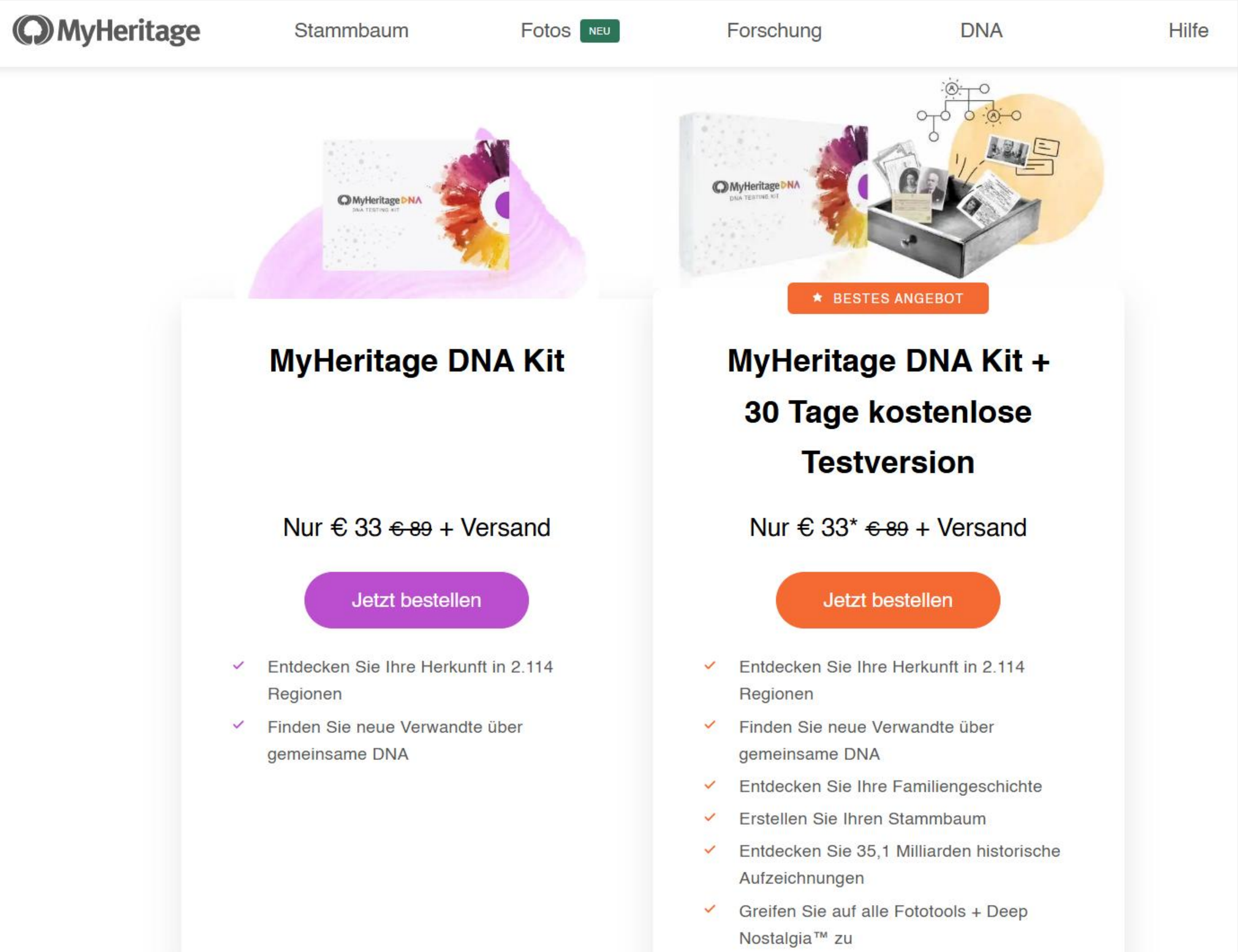
AncestryDNA®

Herkunftsregionen ⓘ ✓

DNA-Matches ⓘ ✓

AncestryDNA® Veranlagungen ⓘ ✓

Illumina



MyHeritage DNA Kit

Nur € 33 ~~€ 89~~ + Versand

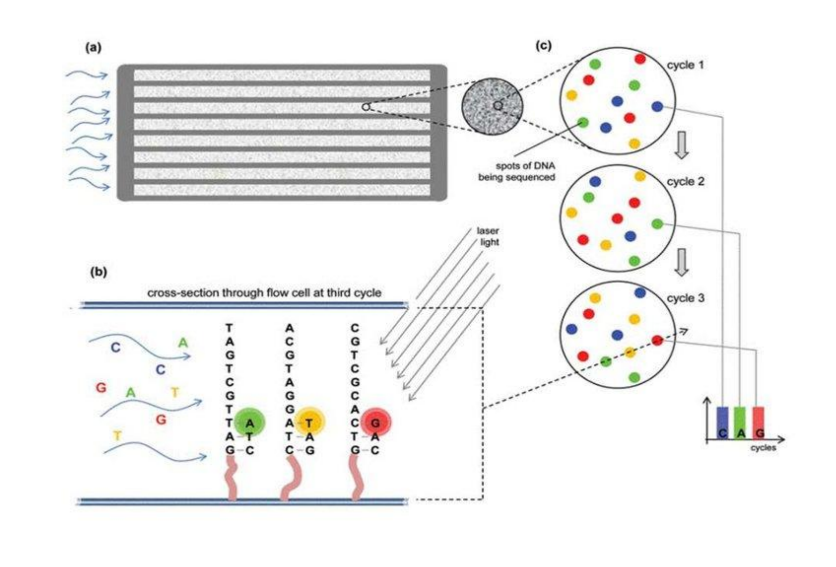
- ✓ Entdecken Sie Ihre Herkunft in 2.114 Regionen
- ✓ Finden Sie neue Verwandte über gemeinsame DNA

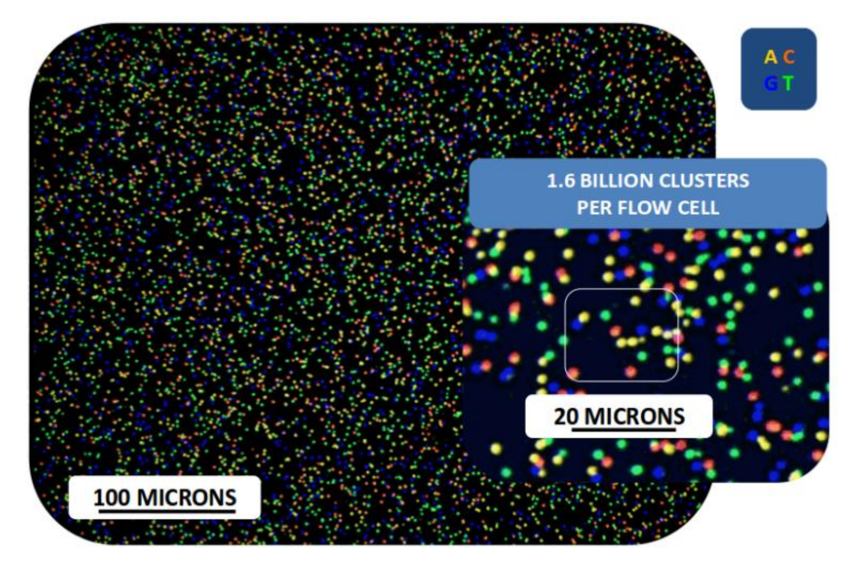
MyHeritage DNA Kit + 30 Tage kostenlose Testversion

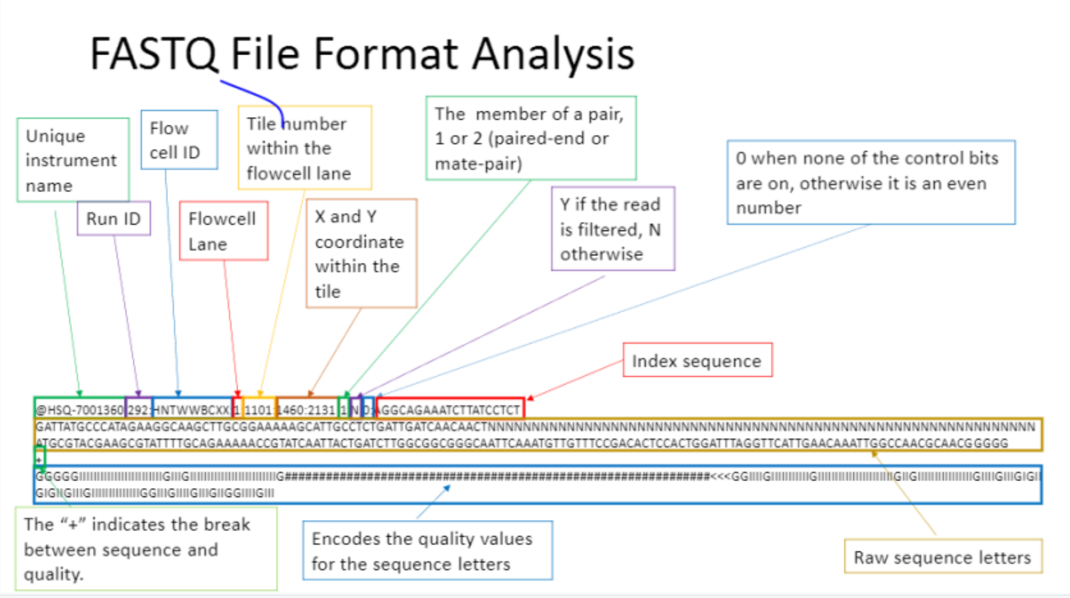
Nur € 33* ~~€ 89~~ + Versand

- ✓ Entdecken Sie Ihre Herkunft in 2.114 Regionen
- ✓ Finden Sie neue Verwandte über gemeinsame DNA
- ✓ Entdecken Sie Ihre Familiengeschichte
- ✓ Erstellen Sie Ihren Stammbaum
- ✓ Entdecken Sie 35,1 Milliarden historische Aufzeichnungen
- ✓ Greifen Sie auf alle Fototools + Deep Nostalgia™ zu

Illumina


(a)
(c)
cycle 1
cycle 2
cycle 3
spots of DNA being sequenced
laser light
(b)
cross-section through flow cell at third cycle
C A G T
cycles





FASTQ File Format Analysis

```
@HSQ-7001360:292:HNTWWBCXX:1:1101:1460:2131 1:N:0:AGGCAGAAATCTTATCCTCT
GATTATGCCCATAGAAGGCAAGCTTGCGGAAAAAGCATTGCCTCTGATTGATCAACAACTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
ATGCGTACGAAGCGTATTTTGCAGAAAAACCGTATCAATTACTGATCTTGGCGGCGGGCAATTCAAATGTTGTTTCCGACACTCCACTGGATTTAGGTTCATTGAACAAATTGGCCAACGCAACGGGGG
+
GGGGGIIIIIIIIIIIIIIIIIIIIIIGIIGIIIIIIIIIIIIIIIIIIIIIIIIG##########################################################<<<GGIIIIGIIIIIIIIIIGIIIIIIIIIIIIIIIIIIIGIIGIIIIIIIIIIIIIIIIGIIIGIIIGIGI
GIGIIGIIIGIIIIIIIIIIIIGGIIIGIIIIGIIIGIIGGIIIIGIII
```

Unique instrument name

Run ID

Flow cell ID

Flowcell Lane

Tile number within the flowcell lane

X and Y coordinate within the tile

The member of a pair, 1 or 2 (paired-end or mate-pair)

Y if the read is filtered, N otherwise

0 when none of the control bits are on, otherwise it is an even number

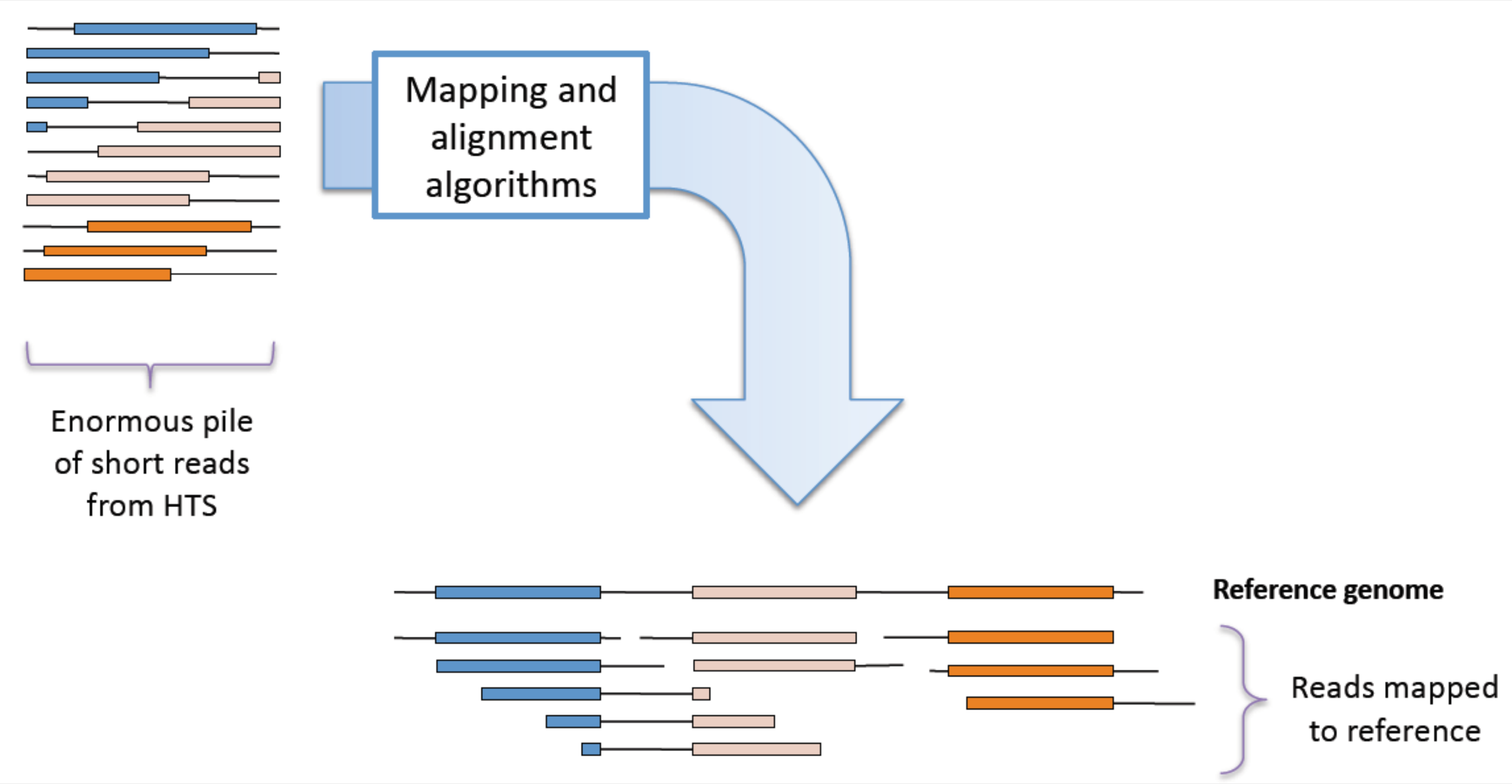
Index sequence

The "+" indicates the break between sequence and quality.

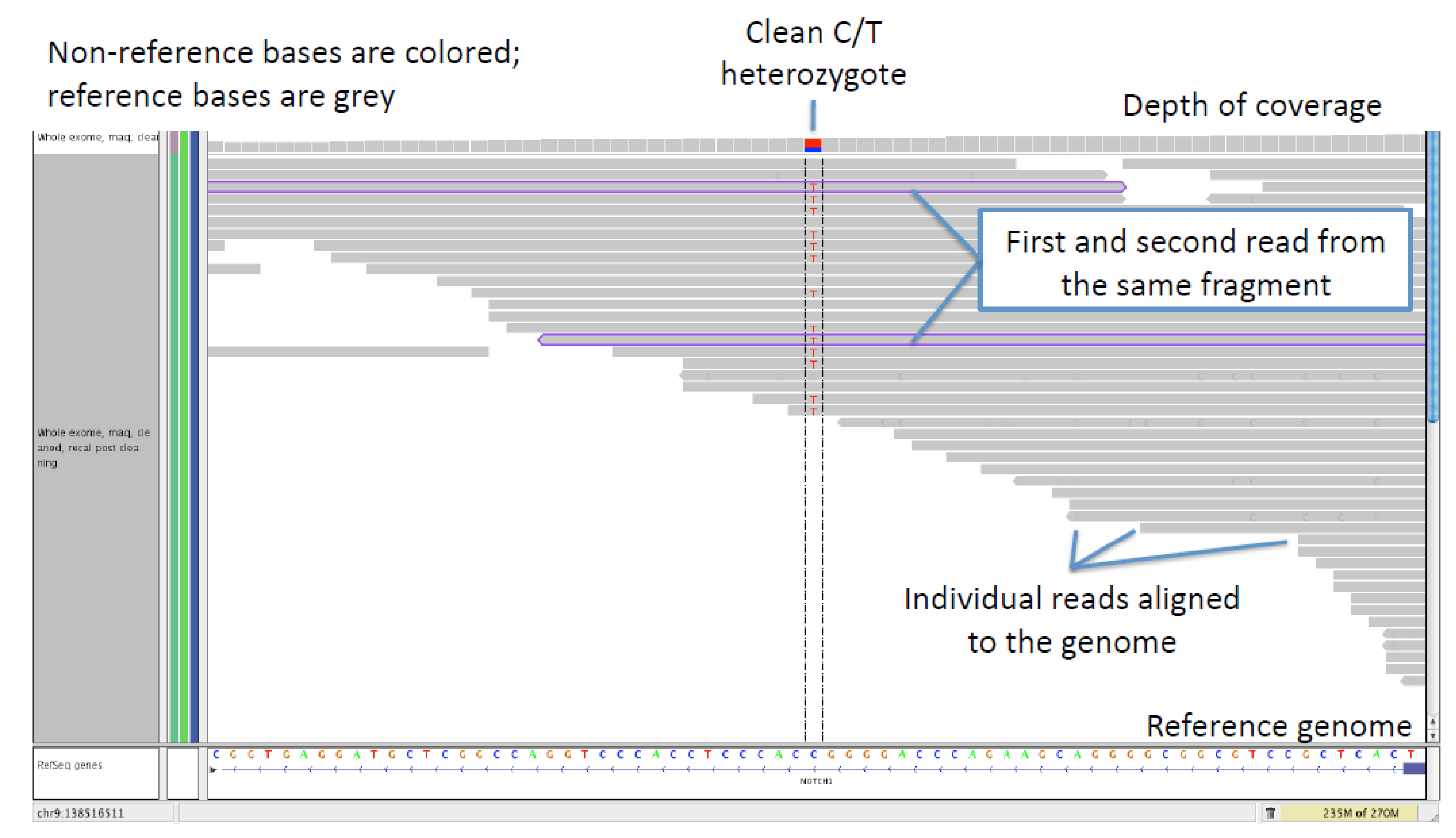
Encodes the quality values for the sequence letters

Raw sequence letters

Instruments generate short reads that must be mapped to the reference


Mapping and alignment algorithms
Enormous pile of short reads from HTS
Reference genome
Reads mapped to reference

Typical screenshot representing aligned HTS reads


Whole exome, maq, clea
Whole exome, maq, cle aned, recal post clea ning
T
RefSeq genes
C G G T G A G G A T G C T C G G C C A G G T C C C A C C T C C C A C C G G G G A C C C A G A A G C A G G G G C G G C G T C C G C T C A C T
NOTCH1
chr9:138516511
235M of 270M

Non-reference bases are colored; reference bases are grey

Clean C/T heterozygote

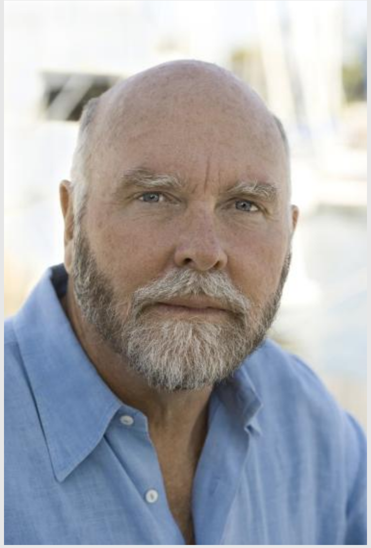
Depth of coverage

First and second read from the same fragment

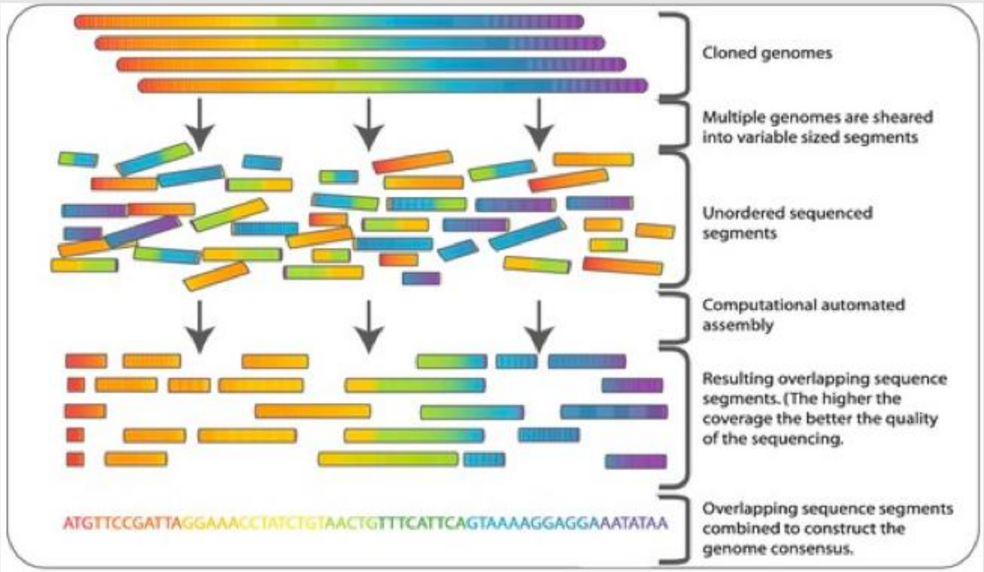
Individual reads aligned to the genome

Reference genome

Human genome project



- 1984: Plan – Sanger sequencing
- 1990: Start at National Institutes of Health (NIH)
- 1998: Craig Venter and Celera Genomics shotgun sequencing
- 2001: Draft(s) published together with Francis Collins of NIH
- 2004: Final published
- Size: ~3 billion base pairs
- Cost: ~$3 billion


Cloned genomes
Multiple genomes are sheared into variable sized segments
Unordered sequenced segments
Computational automated assembly
Resulting overlapping sequence segments. (The higher the coverage the better the quality of the sequencing.
Overlapping sequence segments combined to construct the genome consensus.
ATGTTCCGATTAGGAAACCTATCTGTAACTGTTTCATTCAGTAAAAGGAGGAAATATAA

Human genome XPRIZE

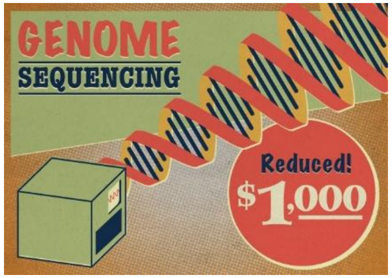
XPRIZE FOUNDATION

- **XPRIZES**: intended to encourage technological development that could benefit mankind
- 1996: Ansari XPRIZE for suborbital spaceflight. Claimed by SpaceShipOne in 2004 ($10 million)
- **2006: Archon Genomics XPRIZE: $10 million will be awarded to the first team to rapidly, accurately and economically sequence 100 whole human genomes to an unprecedented level of accuracy**

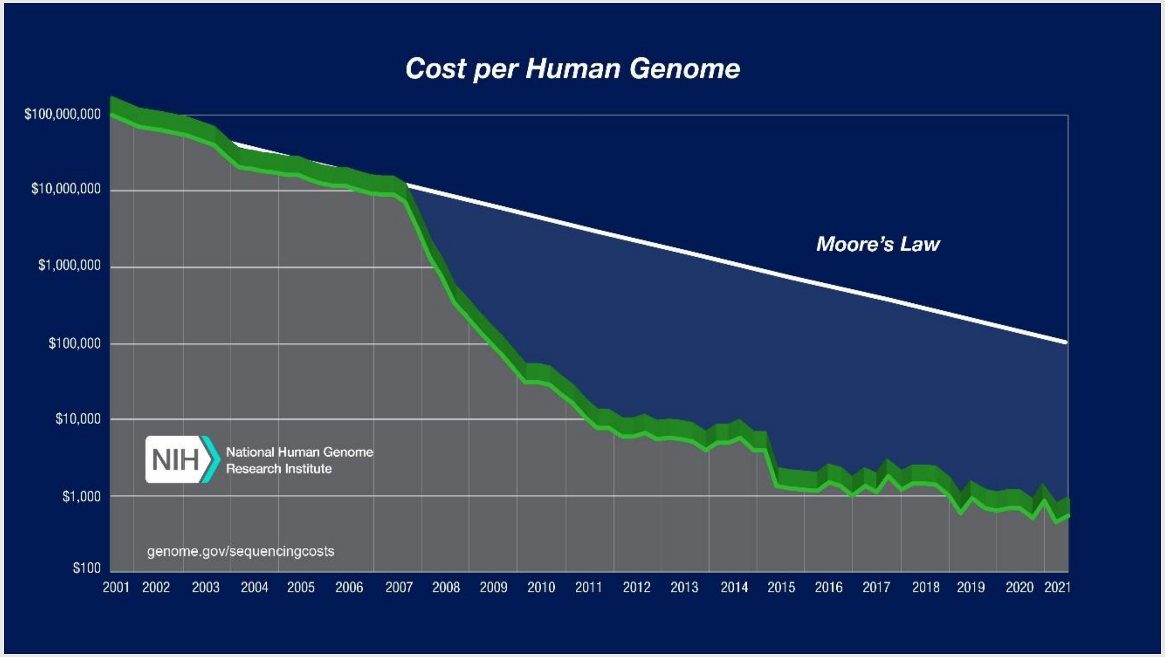
cancelled

- 2007: Google Lunar XPRIZE: $20 million to land a rover on the moon, move more than 500 m, and transmit HD images and video back to earth
- 2011: Tricorder XPRIZE: $10 million for a mobile device that can diagnose patients as accurately as a panel of board-certified physicians

Sequencing costs



	Sanger 3730xl	454 GS FLX	Ion Proton	Illumina HiSeq	Oxford Nanopore
Generation	1	2	2	2	3
Max read length	1,100 bp	700 bp	200 bp	150 bp	15,000 bp
Max output	0.1 Mb	700 Mb	1,000 Mb	1,800,000 Mb	400 Mb
Error rate	0.1% (1:1000)	1% (1:100)	1% (1:100)	0.1% (1:1000)	25% (1:4)
Cost per Mb	£1000	£5	£0.05	£0.005	£0.50



1000 genomes project

- 2008: Launched
- Establish a detailed catalogue of **human genetic variation** correlated with ethnicities
- Sequence 1000 anonymous participants from various ethnic groups within 3 years
- **2012**: 1092 genomes announced
- Each person carries **250-300 loss- of- function variants** in annotated genes
- **50-100 variants** previously implicated in inherited disorders
- Mutation rate of **10-8 per bp per generation** (based on mother-father- child trios)
- 1000 nematode genomes, 1000 plant genomes, Genome 10K project, etc.

Applications

RNA Sequencing

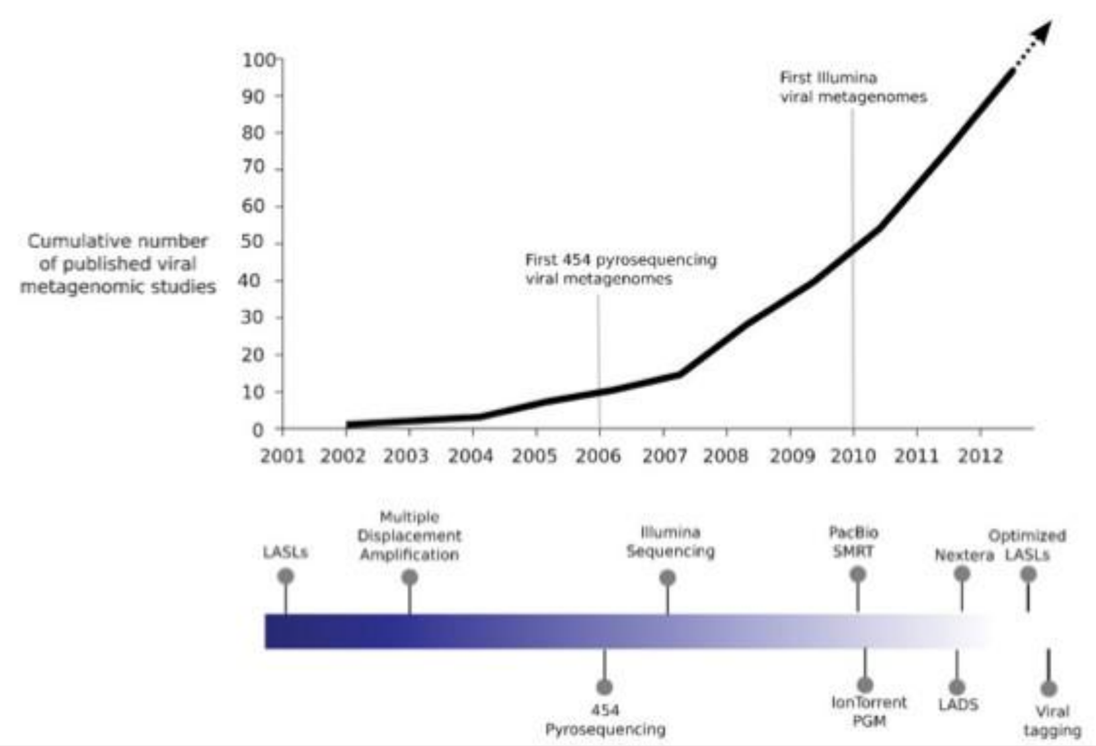
- mRNA Sequencing
- Targeted RNA Sequencing
- Ribosome Profiling
- RNA Exome Capture Sequencing
- Total RNA Sequencing
- Small RNA Sequencing
- Ultra-Low-Input and Single-Cell RNA-Seq

Methylation Sequencing

DNA Sequencing

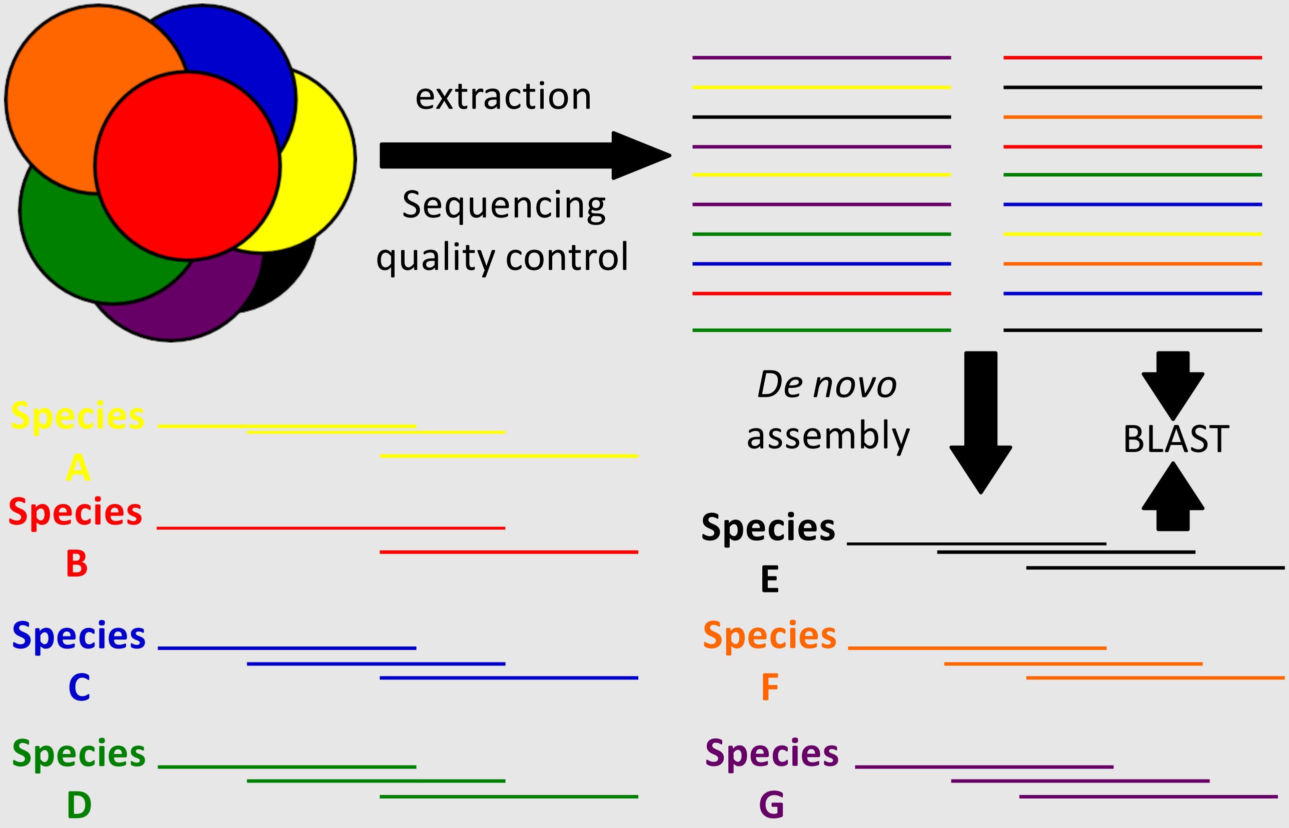
- Whole-Genome Sequencing
- Targeted Sequencing
- ChIP-Seq
- ATAC-Seq

Metagenomics

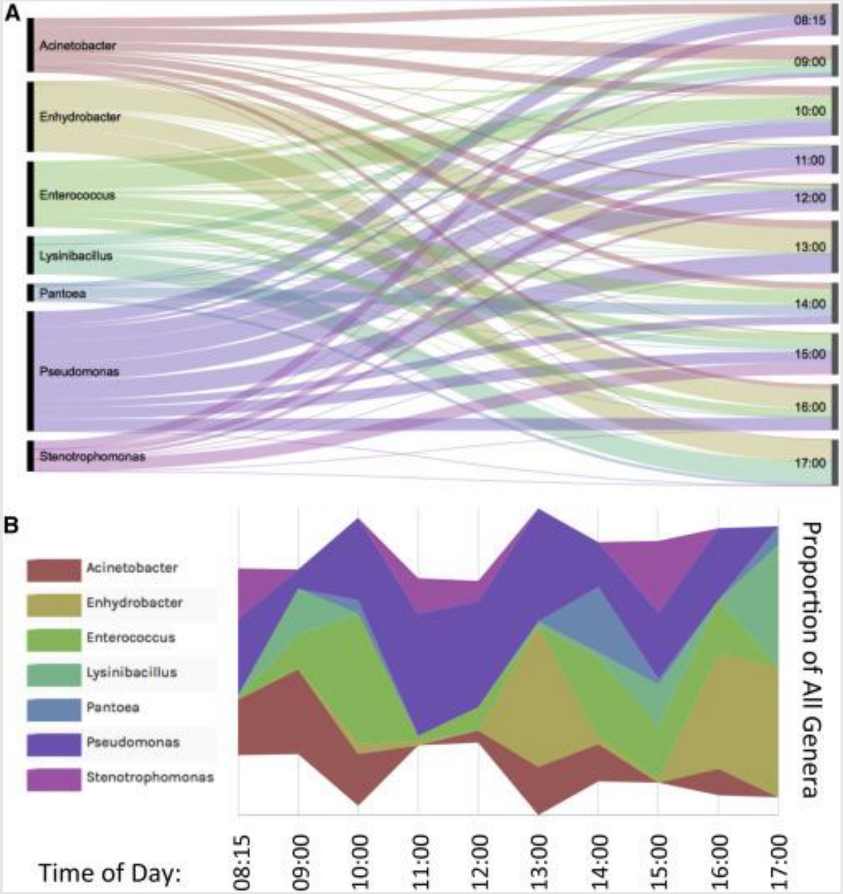


- **Metagenomics** can be defined as the sequenced-based analysis of the **whole collection of genomes** isolated directly from a sample
- The advantage is that **isolation is not needed** – only extraction and sequencing (although there's more to it than that!)
- Bacteria and archaea: 16S rRNA gene, relatively short, often conserved within species, and generally different among species
- **Viruses**: often present with a large excess of host DNA, making their efficient and reliable detection problematic

Metagenomics methods



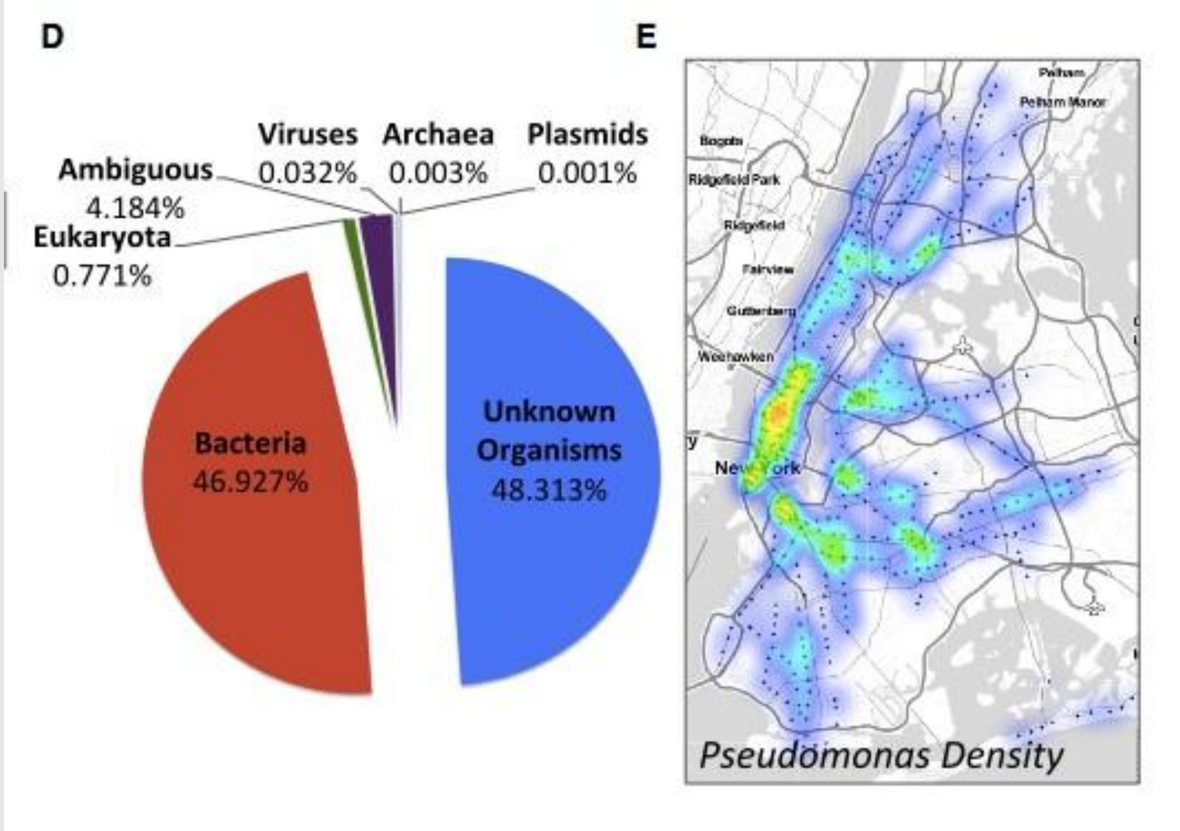
Metagenomics – detection



Analysis of samples collected at Penn Station on one day, compared at each hour

- Saunders et al. (2012): Geospatial resolution of human and bacterial diversity with city-scale metagenomics

Metagenomics – detection



Geospatial analysis of the most prevalent genus, *Pseudomonas*, across the subway system

- Saunders et al. (2012): Geospatial resolution of human and bacterial diversity with city-scale metagenomics

Metagenomics – detection

weird news (http://www.mtv.com/news/collection/weird-news/)

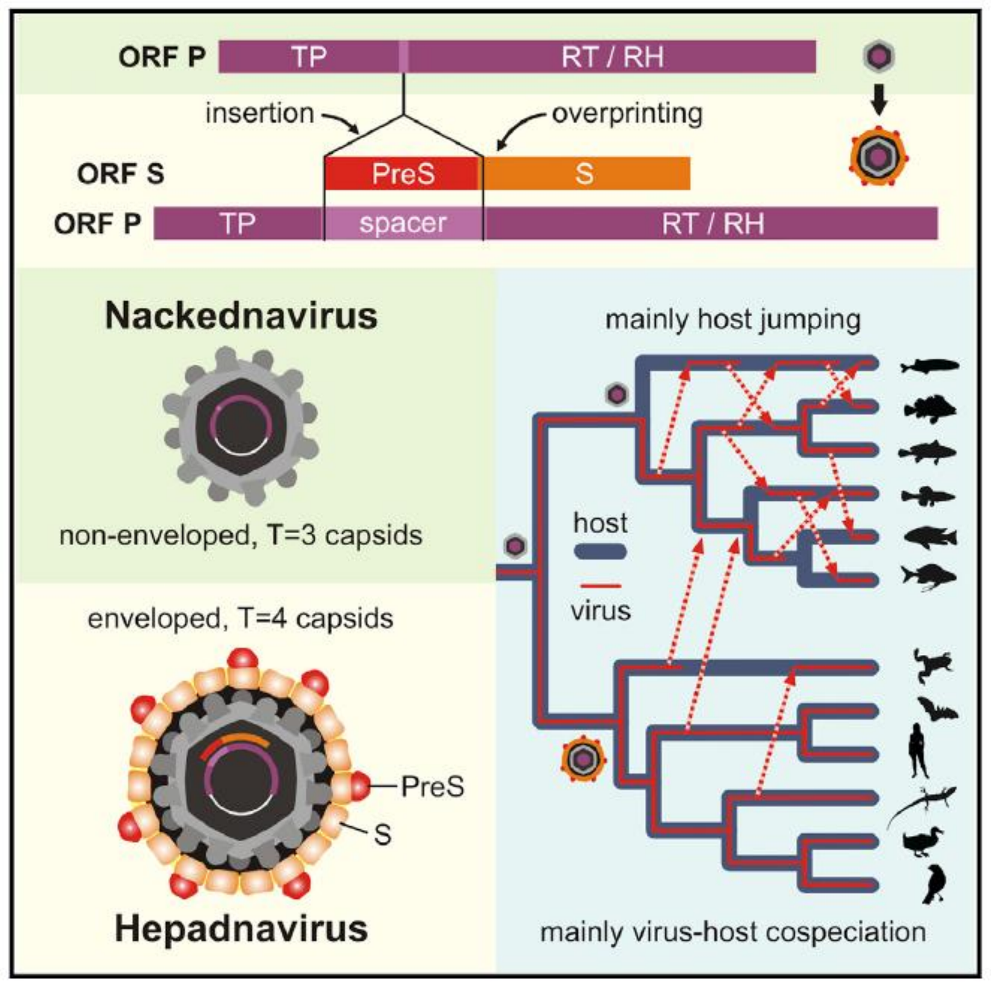
Getty Images

Scientists Basically Just Discovered Alien Life — In The NYC Subway

Nearly half of the germs on the train are unrecognizable even to the experts.

- Saunders et al. (2012): Geospatial resolution of human and bacterial diversity with city-scale metagenomics

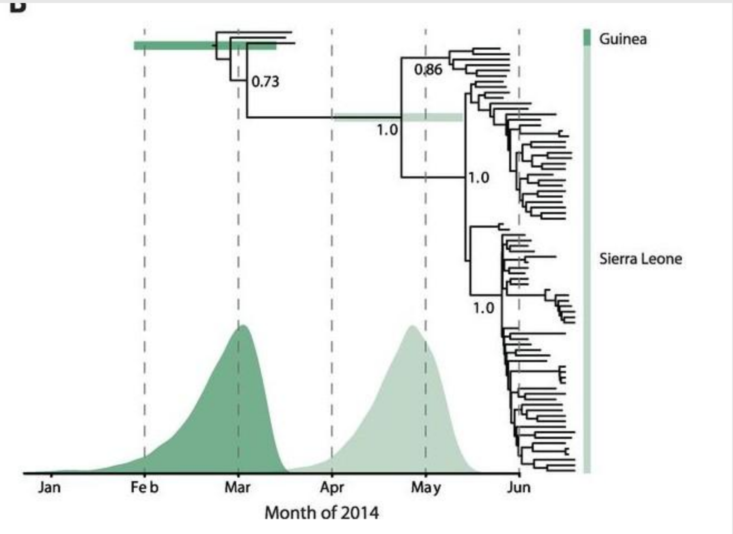
Metagenomics – virus discovery

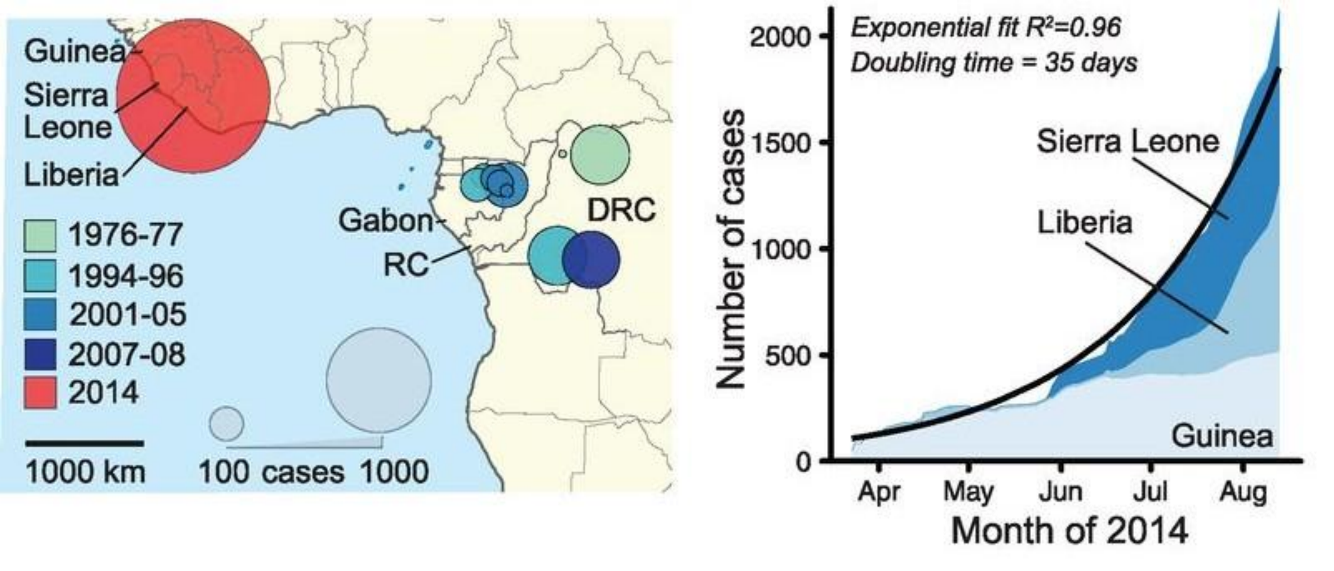


- Lauber & Seitz et al. (2017): Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses

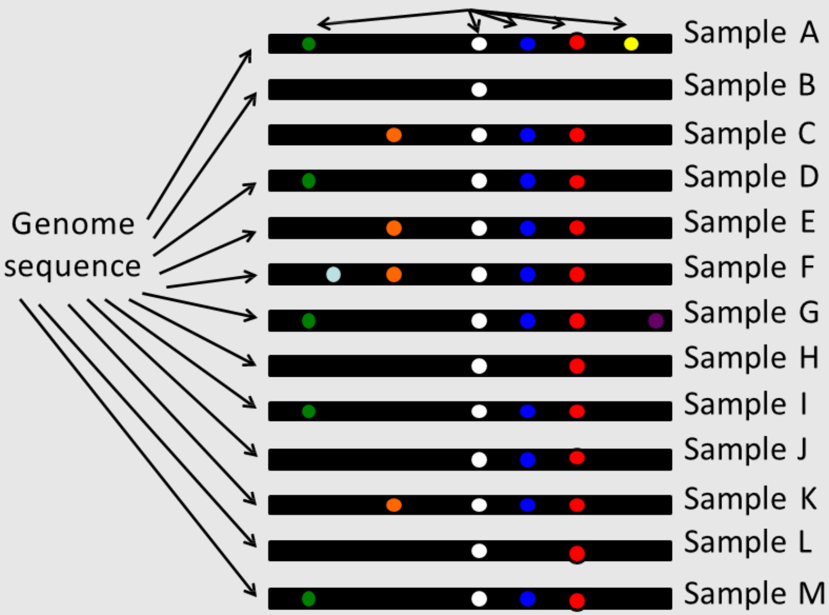
Epidemics – Ebola

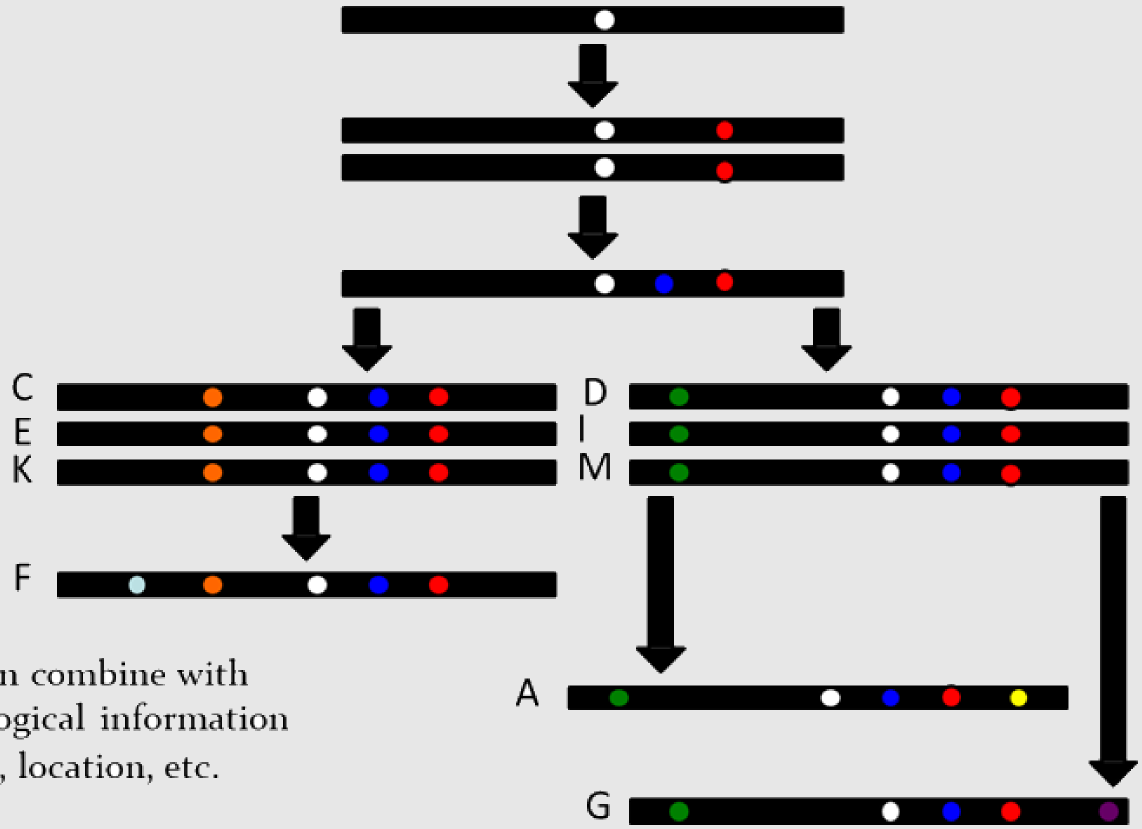
- The 2013-2015 West Africa Ebola epidemic, 26648 cases, 11017 deaths
- HTS used throughout the epidemic to sequence Ebola virus genomes from patient samples
- Used to monitor viral evolution: how fast is it mutating, where is it mutating, which selection pressures are operating





Epidemics – who infected whom?

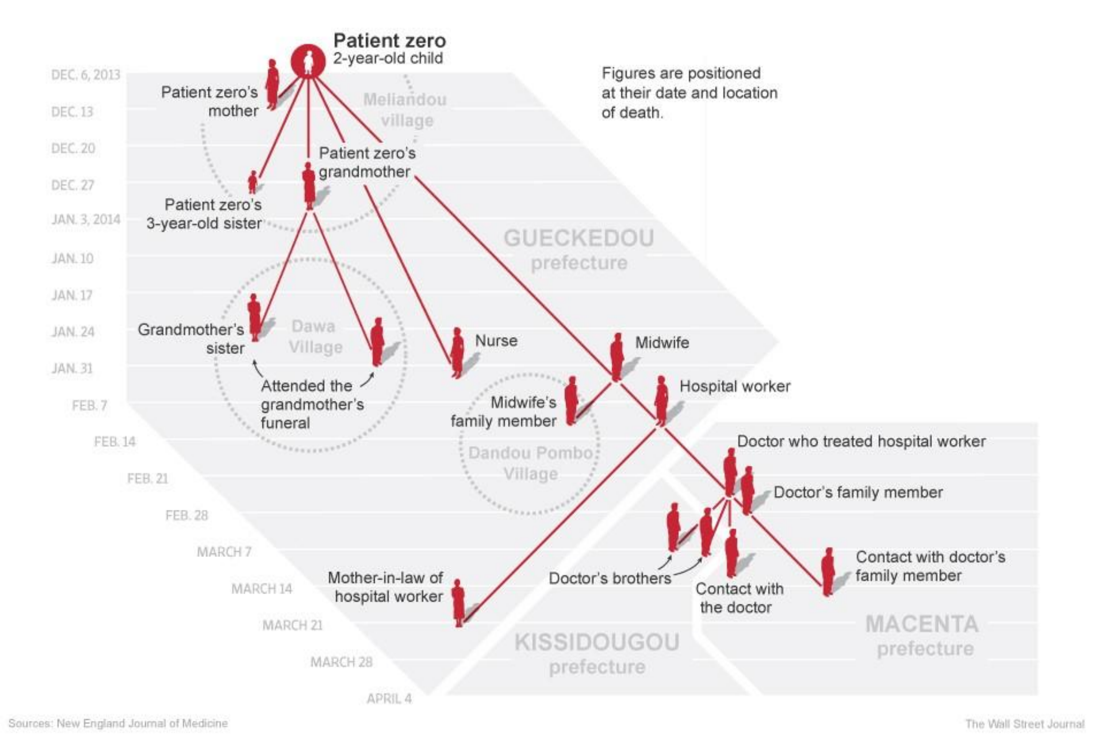




Can then combine with epidemiological information – date, location, etc.

Epidemics – who infected whom?

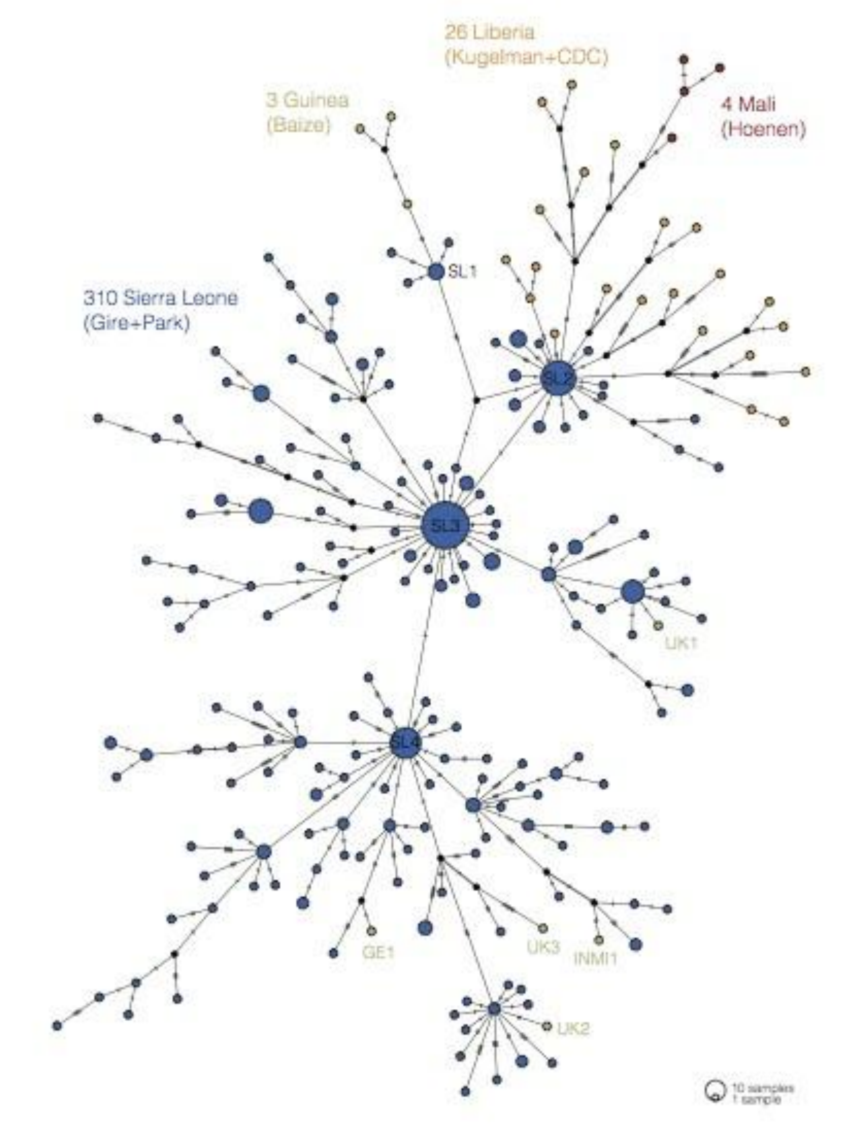
- Identify source of infection
- Identify long transmission events
- Identify super-spreaders – individual or hub level
- Identify new incursions or spillovers



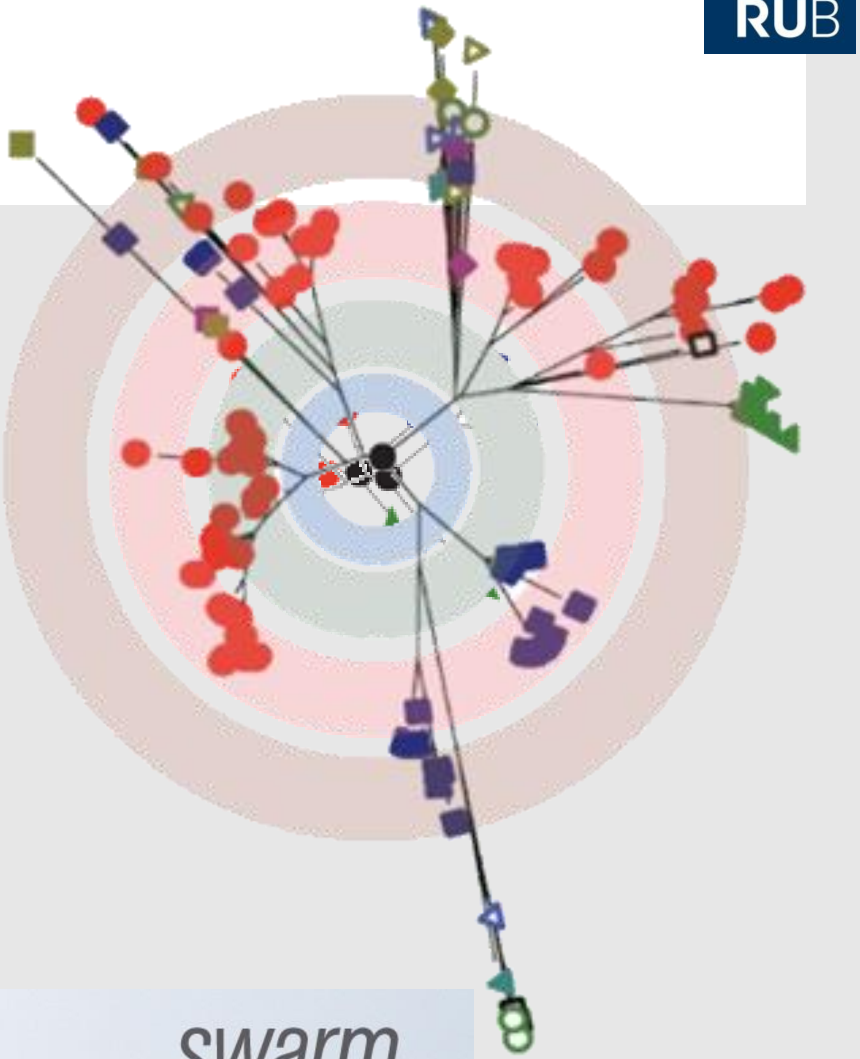
Figures are positioned at their date and location of death.

Sources: New England Journal of Medicine

The Wall Street Journal

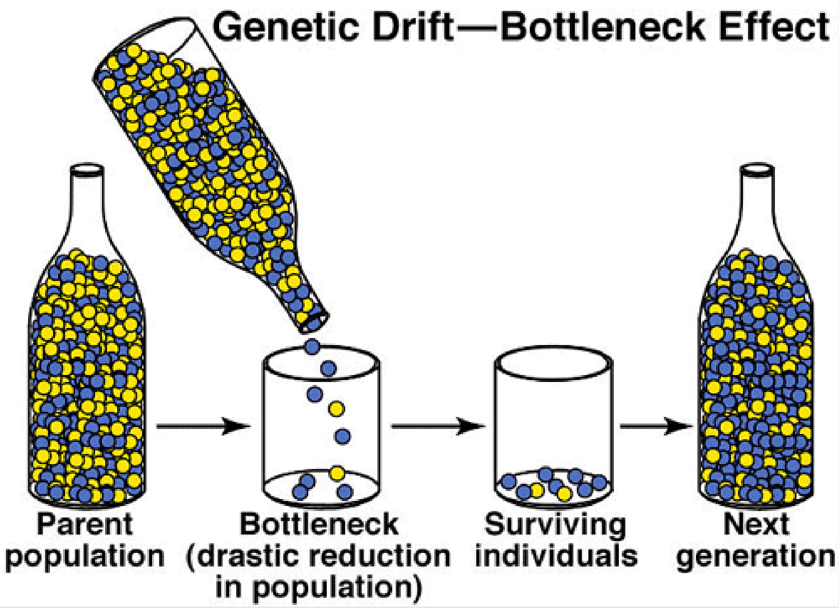


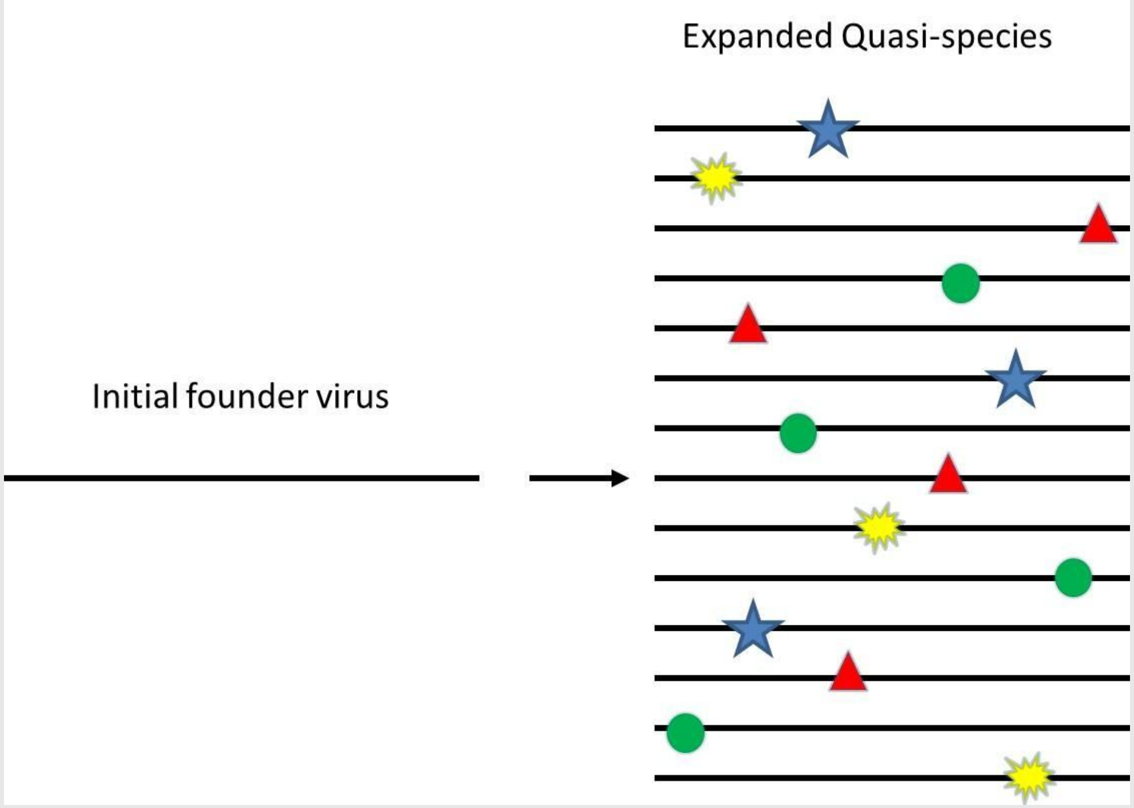
Viral populations



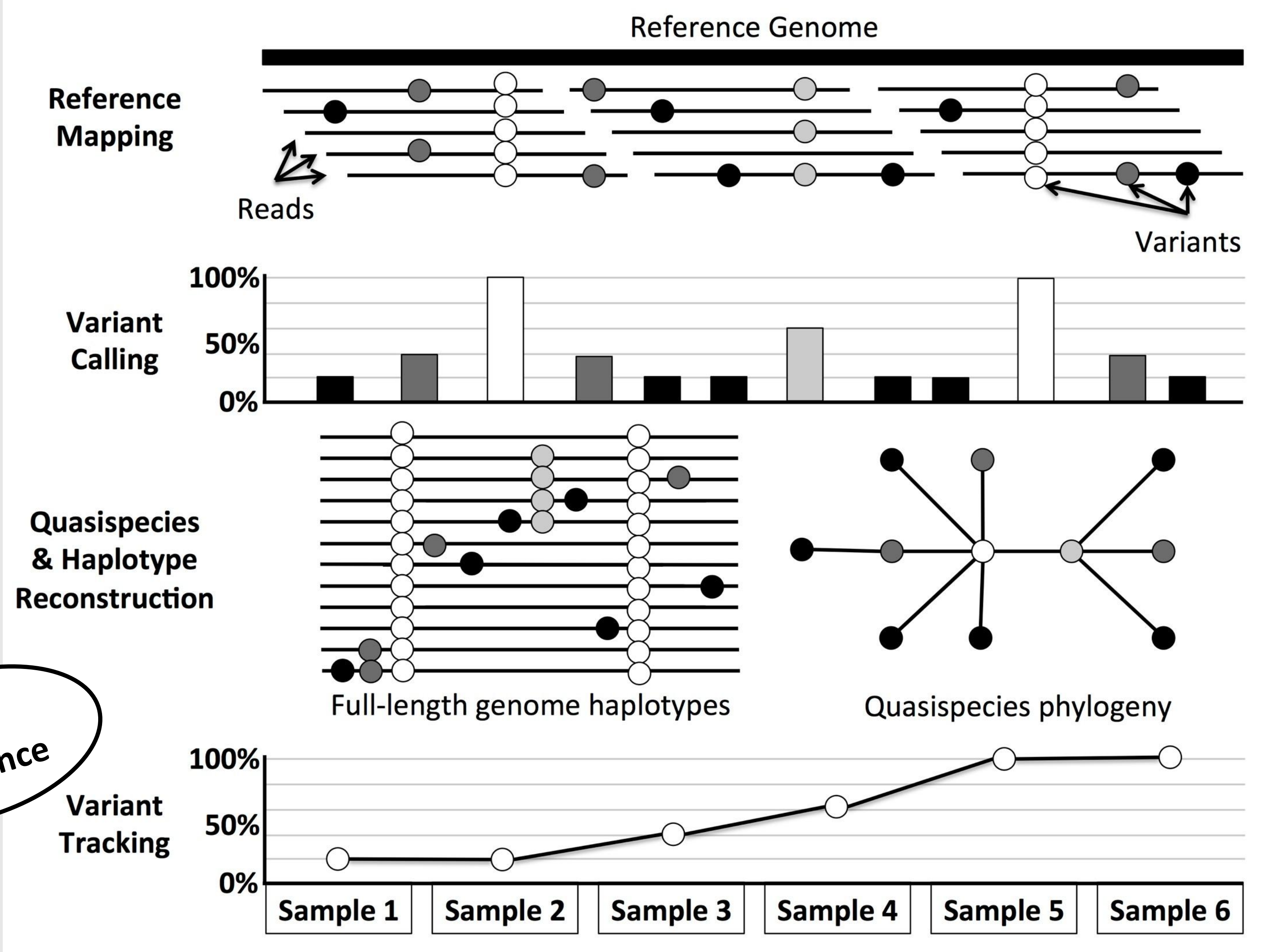
- Viruses **mutate rapidly**
- A single virus can enter a cell, and output tens of thousands of virions within hours
- Every time the genome is copied, **mutations** are introduced
- Enables viruses to **adapt** to change rapidly
- New environments
- New hosts
- Drug and vaccine treatment
- Viruses exist as a large and constantly and rapidly evolving swarm – the **quasispecies**

Bottlenecks and Founder Effect


Genetic Drift—Bottleneck Effect
Parent population
Bottleneck (drastic reduction in population)
Surviving individuals
Next generation



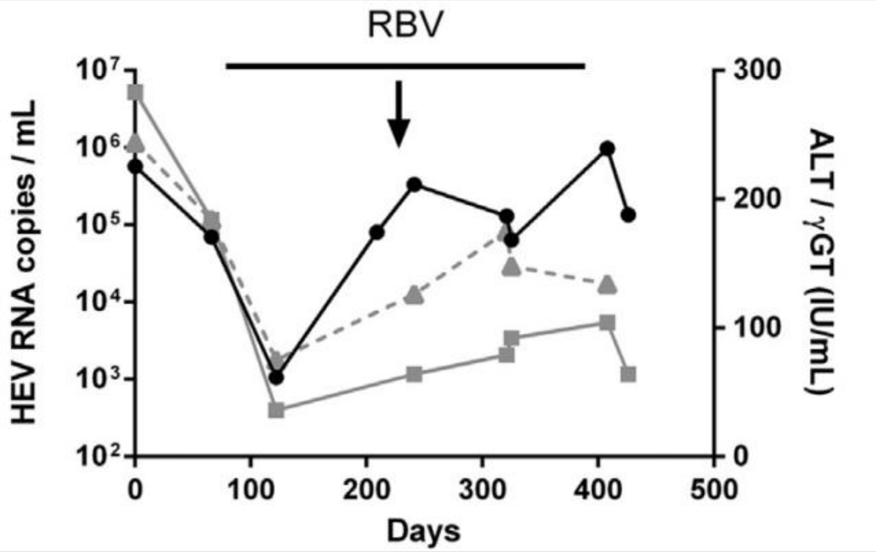
Viral HTS

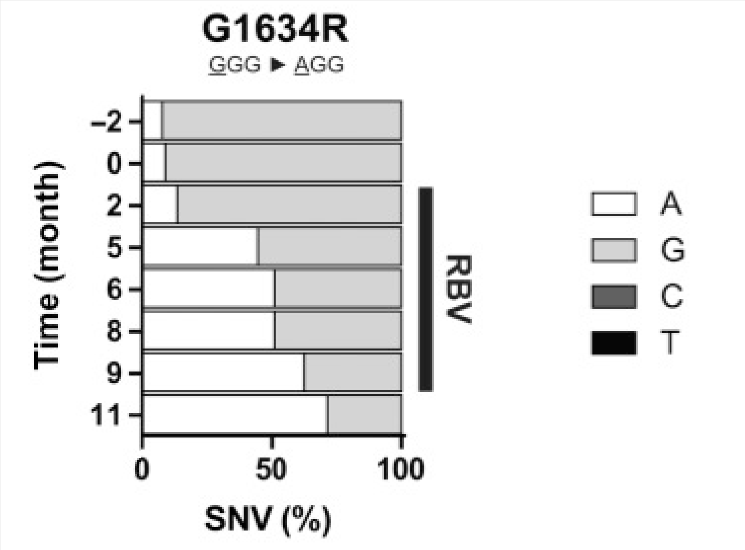


drug resistance

Viral mutation tracking

- Ability to detect mutations at **low levels** in a sample
- Can then examine samples for the presence of important mutations: e.g. **drug resistance**
  - ➢ Hepatitis E virus





- clinical application in HIV diagnostics!

Third generation sequencing

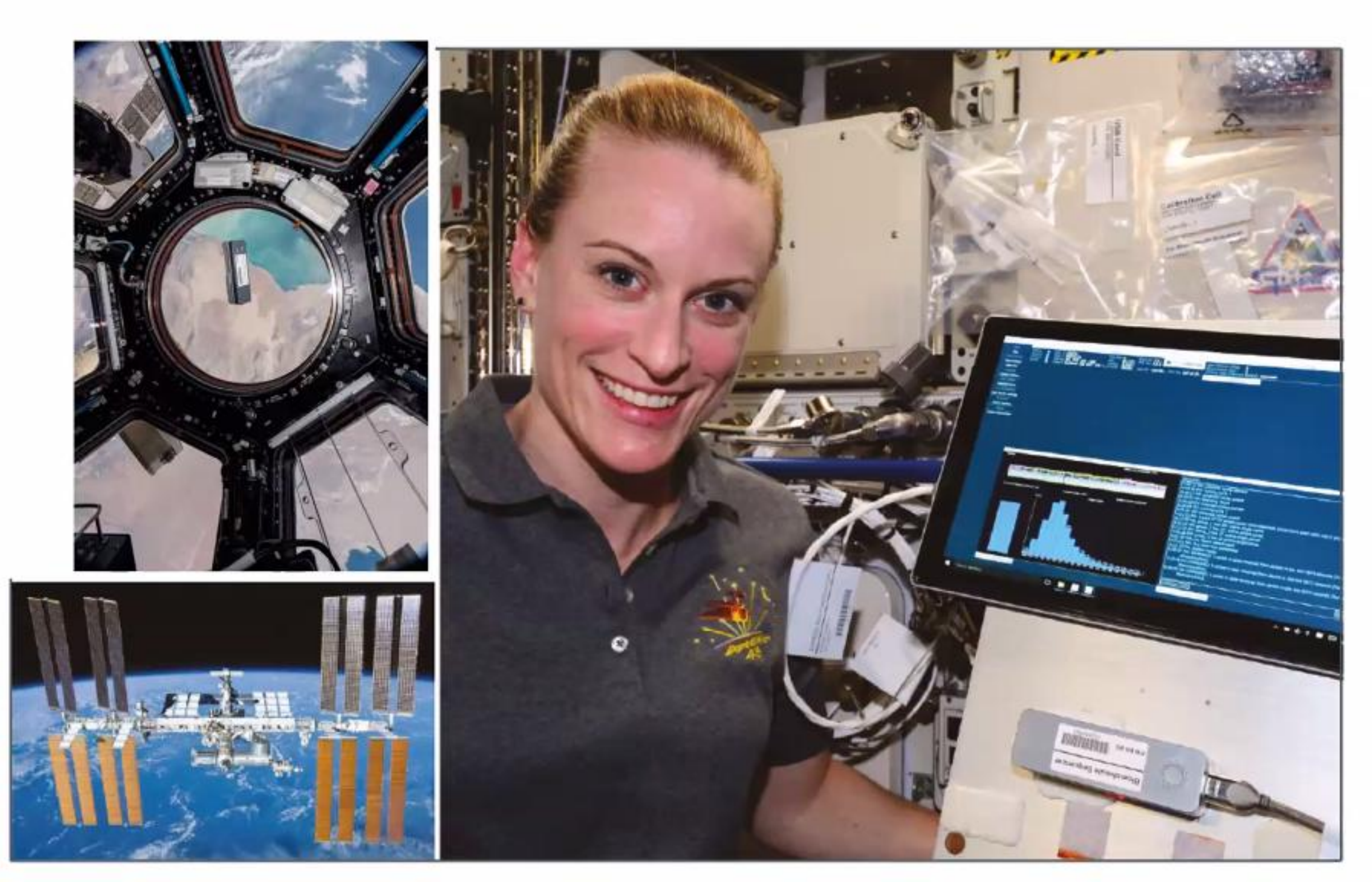
Third generation sequencing

- Advantages over second generation sequencing:
  - Very long reads (Oxford nanopore)
  - Real time output
  - scRNAseq (10x Genomics)

Oxford nanopore sequencing (ONT)

2016 - Kate Rubens becomes the first person to ever sequence in space

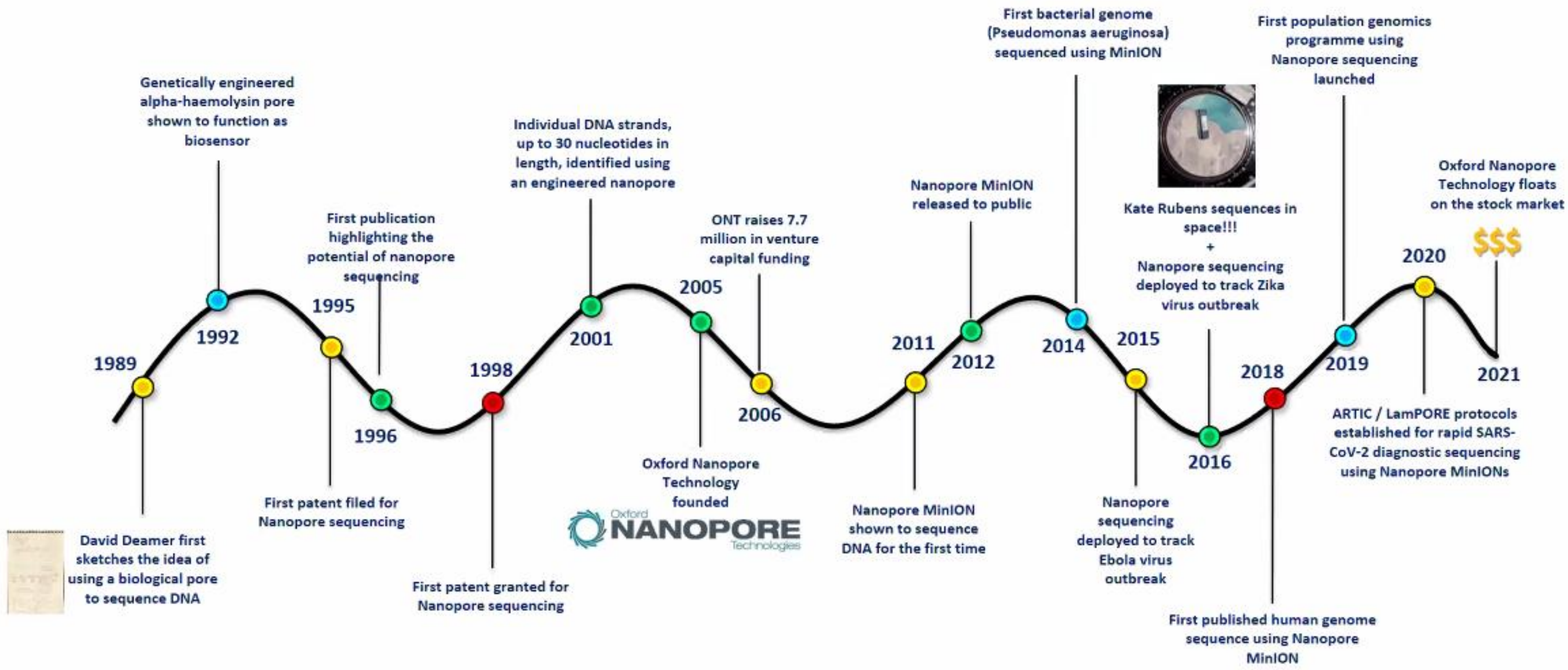
investigated the effects of microgravity on RNA isolation and PCR analysis



+ she is a virologist!!!

Oxford nanopore sequencing (ONT)

A brief history of nanopore sequencing



1989 David Deamer first sketches the idea of using a biological pore to sequence DNA

1992 Genetically engineered alpha-haemolysin pore shown to function as biosensor

1995 First patent filed for Nanopore sequencing

1996 First publication highlighting the potential of nanopore sequencing

1998 First patent granted for Nanopore sequencing

2001 Individual DNA strands, up to 30 nucleotides in length, identified using an engineered nanopore

2005 Oxford Nanopore Technology founded

2006 ONT raises 7.7 million in venture capital funding

2011 Nanopore MinION shown to sequence DNA for the first time

2012 Nanopore MinION released to public

2014 First bacterial genome (Pseudomonas aeruginosa) sequenced using MinION

2015 Nanopore sequencing deployed to track Ebola virus outbreak

2016 Kate Rubens sequences in space!!! + Nanopore sequencing deployed to track Zika virus outbreak

2018 First published human genome sequence using Nanopore MinION

2019 First population genomics programme using Nanopore sequencing launched

2020 ARTIC / LamPORE protocols established for rapid SARS-CoV-2 diagnostic sequencing using Nanopore MinIONs

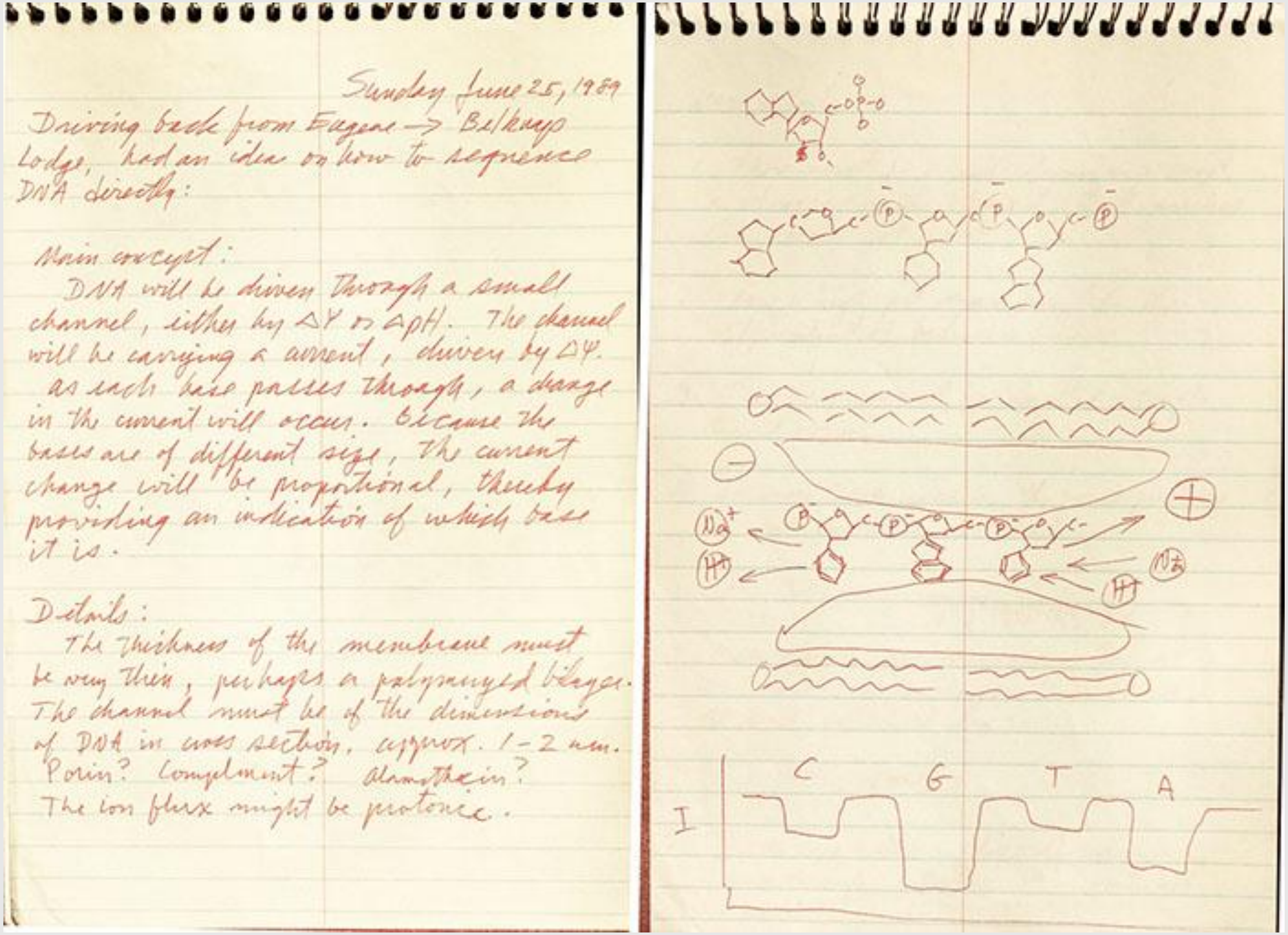
2021 Oxford Nanopore Technology floats on the stock market

Key publications

Kasianowicz et al (1996) Characterization of individual polynucleotide moleculesusing a membrane channel. *PNAS*
Clarke et al (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Biotechnology*
Cherf et al (2012) Automated forward and reverse ratcheting of DNA in a nanopore at five angstrom precision. *Nature Biotechnology*
Deamer et al (2016) Three decades of nanopore sequencing. *Nature Biotechnology*
Garalde et al (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*
Jain et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*

Oxford nanopore sequencing (ONT)

Professor David Deamer’s initial sketch for sequencing DNA using a nanopore



Sunday June 25, 1989

Driving back from Eugene → Belknap Lodge, had an idea on how to sequence DNA directly:

Main concept:

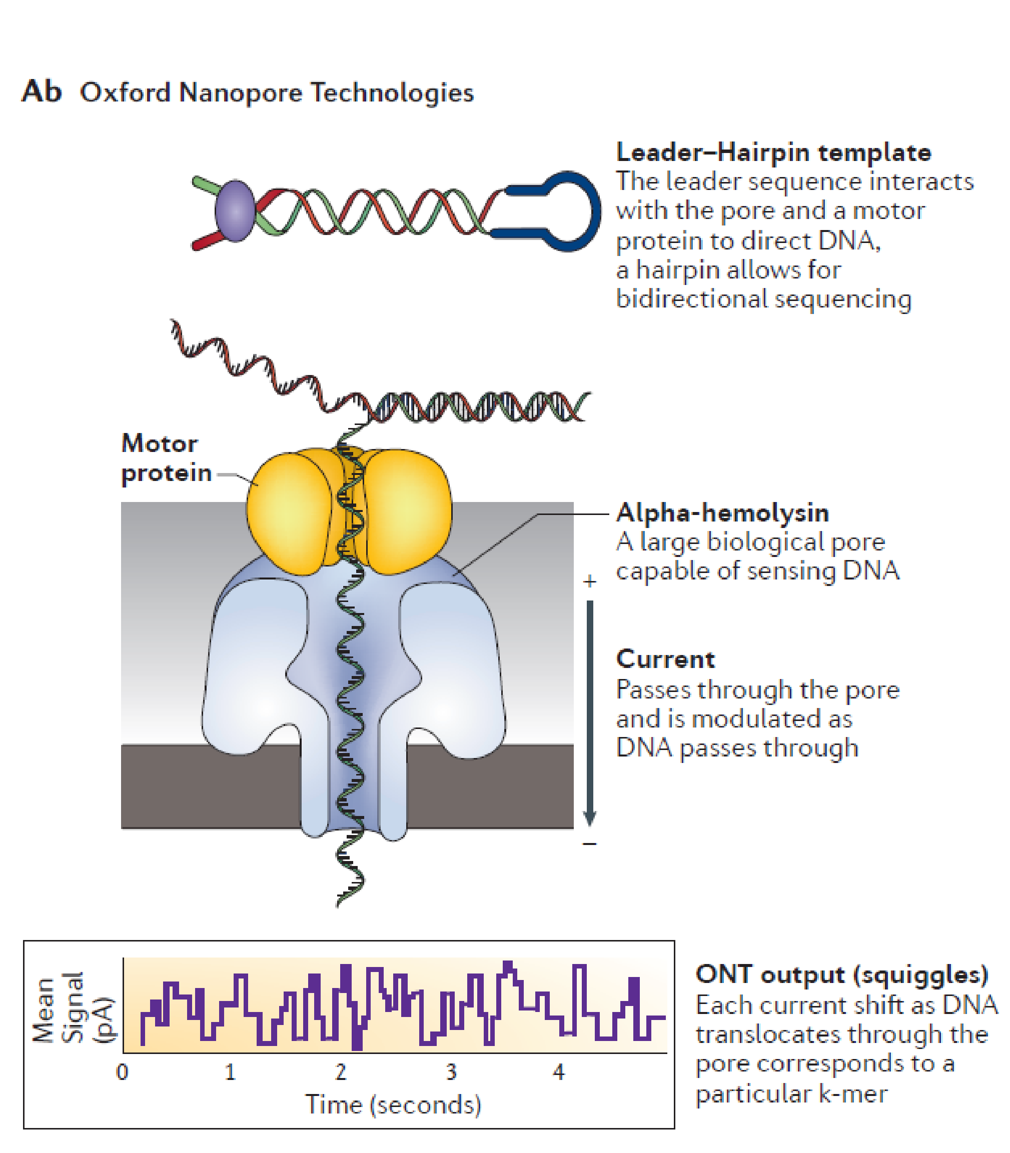
DNA will be driven through a small channel, either by $\Delta\Psi$ or ΔpH. The channel will be carrying a current, driven by $\Delta\Psi$. As each base passes through, a change in the current will occur. Because the bases are of different size, the current change will be proportional, thereby providing an indication of which base it is.

Details:

The thickness of the membrane must be very thin, perhaps a polymerized bilayer. The channel must be of the dimensions of DNA in cross section, approx. 1-2 nm. Porin? Complement? Alamethicin? The ion flux might be protonic.

Oxford Nanopore

- '**Strand sequencing**' is a technique that passes intact DNA polymers through a protein nanopore, sequencing in real time as the DNA translocates the pore
- Simple sample preparation
- Nucleotide base detected as passes through pore (median kmers 5nt)
- Very long reads, up to 15,000 bp
- Small and portable devices useable in field studies (**MinION**), benchtop system for high throughput (**PromethION**) and for use with mobile devices (**SmidgION**)
- High error rate



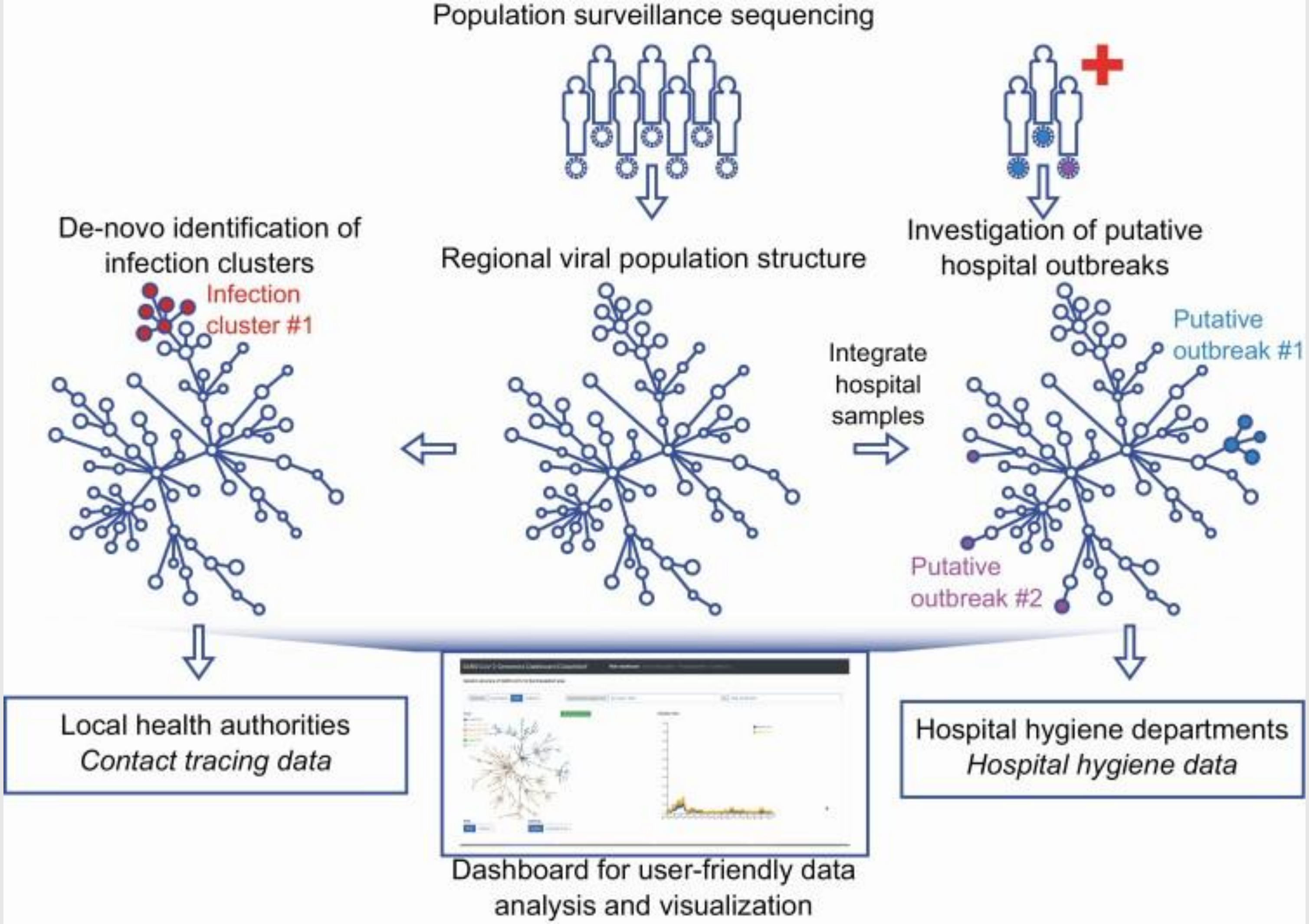
Leader–Hairpin template
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing

Alpha-hemolysin
A large biological pore capable of sensing DNA

Current
Passes through the pore and is modulated as DNA passes through

ONT output (squiggles)
Each current shift as DNA translocates through the pore corresponds to a particular k-mer

Oxford Nanopore



Population surveillance sequencing

De-novo identification of infection clusters

Regional viral population structure

Investigation of putative hospital outbreaks

Infection cluster #1

Integrate hospital samples

Putative outbreak #1

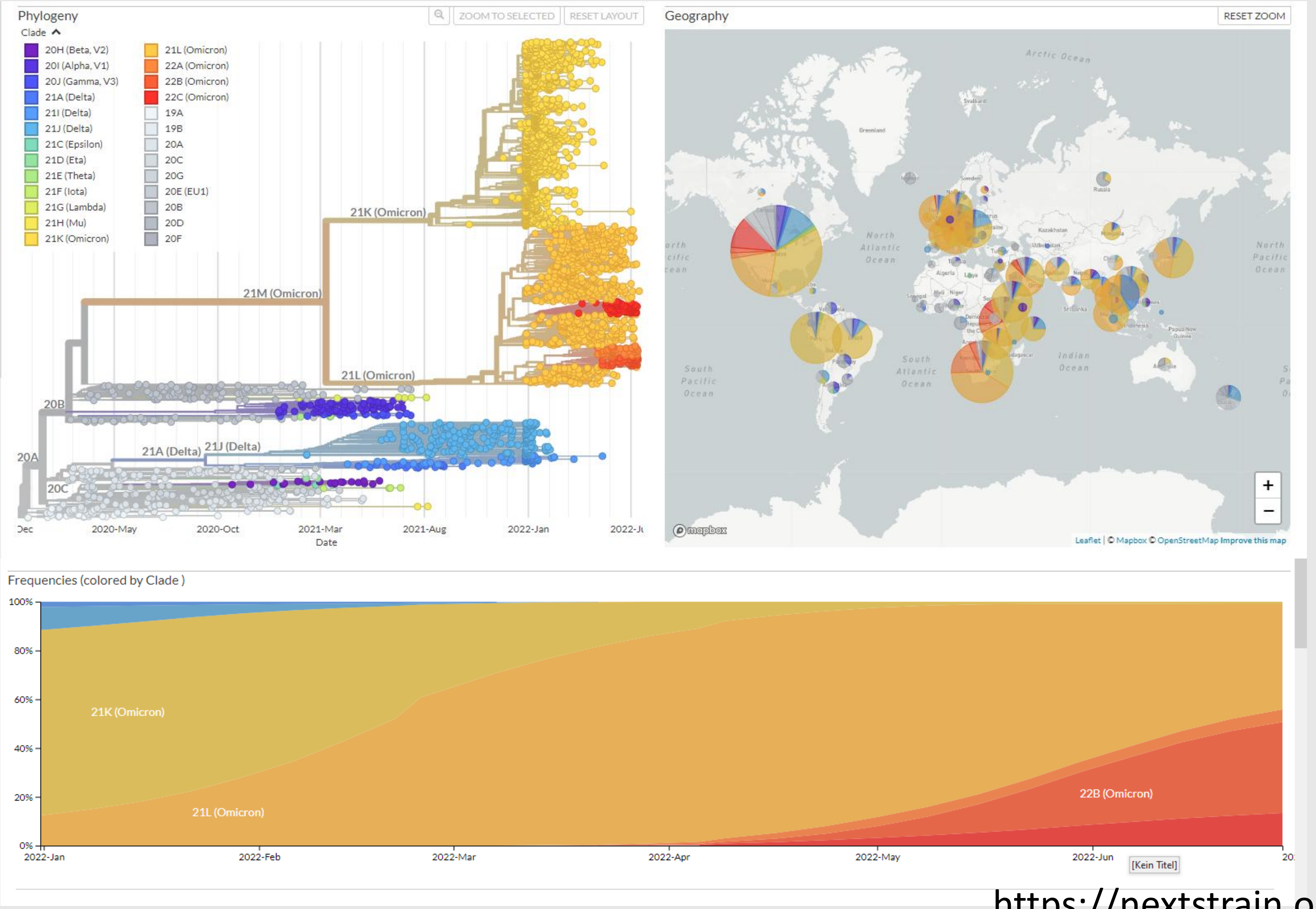
Putative outbreak #2

Local health authorities
Contact tracing data

Hospital hygiene departments
Hospital hygiene data

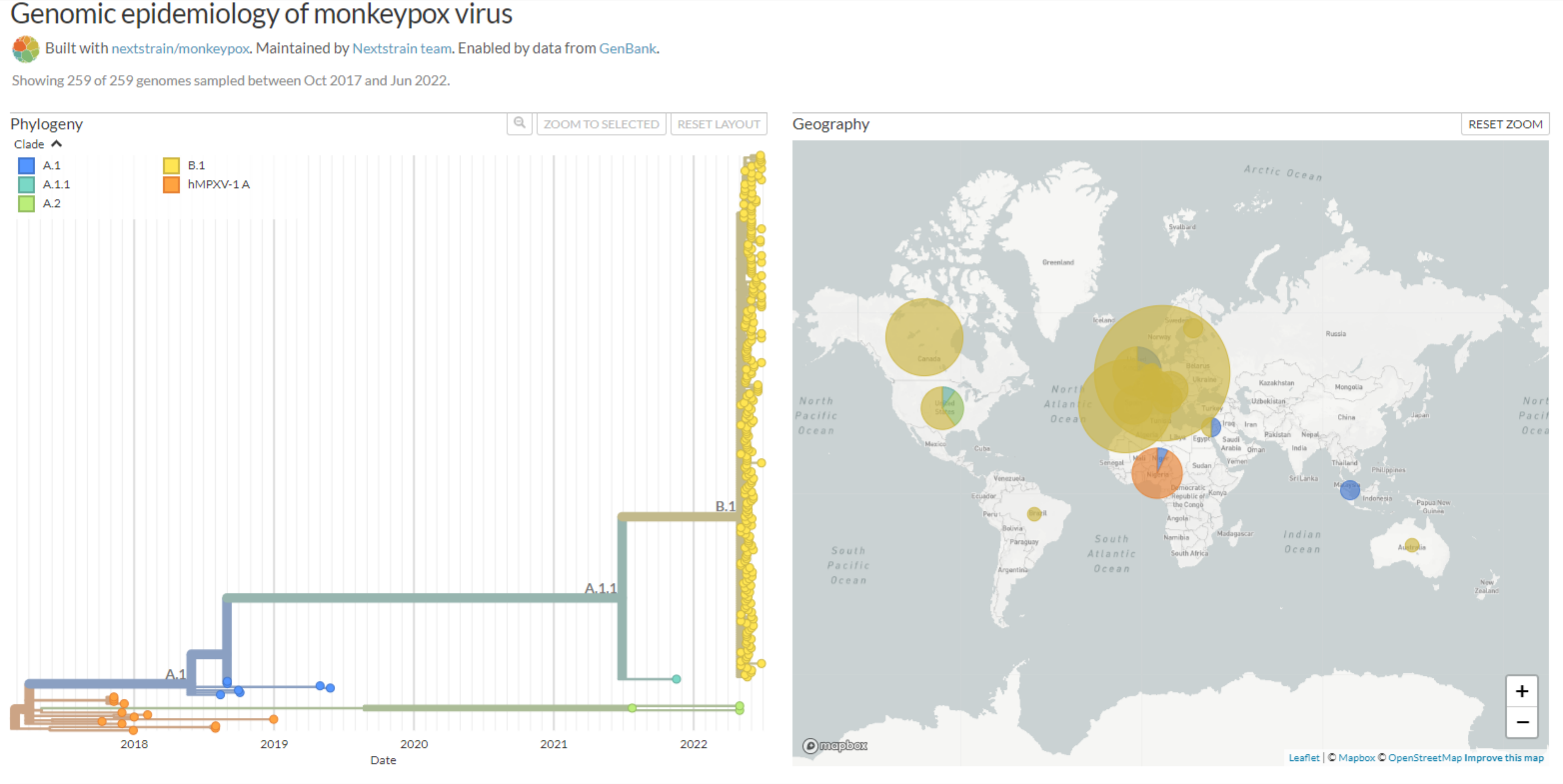
Dashboard for user-friendly data analysis and visualization

Oxford Nanopore



https://nextstrain.org

Oxford Nanopore



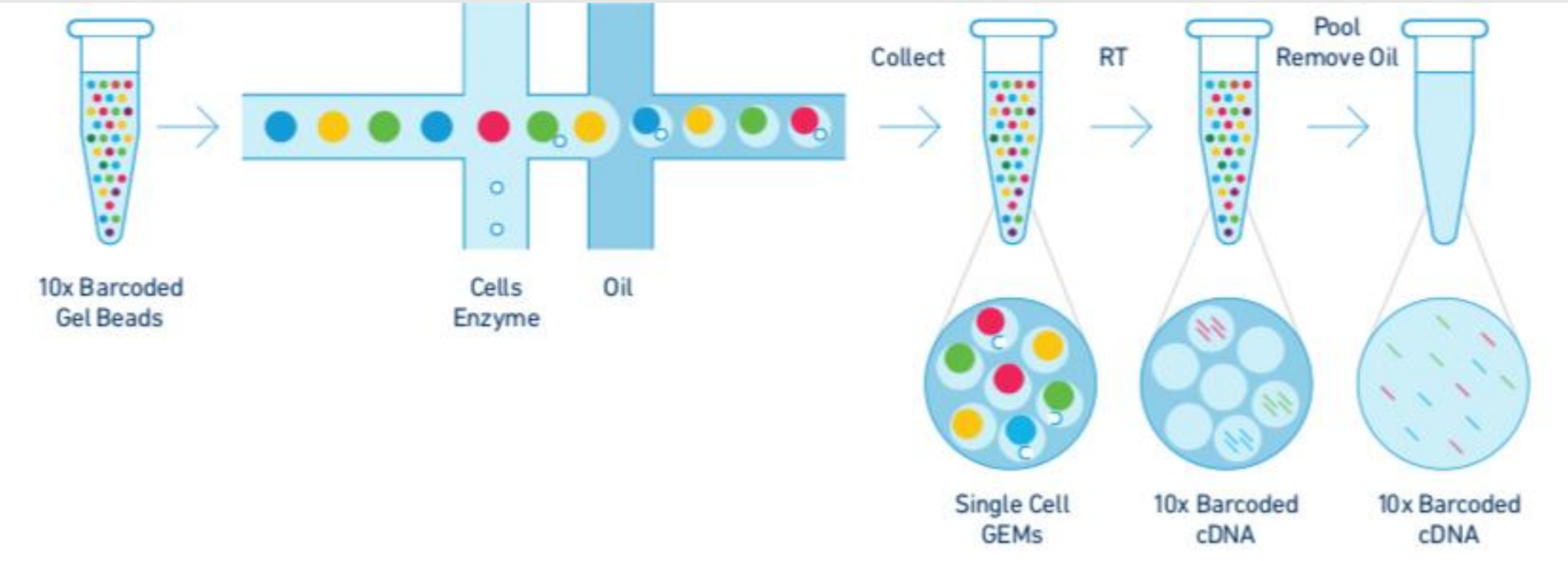
Genomic epidemiology of monkeypox virus

Built with nextstrain/monkeypox. Maintained by Nextstrain team. Enabled by data from GenBank.

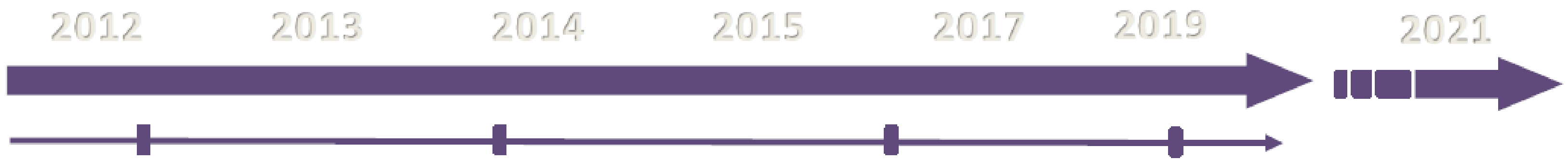
Showing 259 of 259 genomes sampled between Oct 2017 and Jun 2022.

https://nextstrain.org

Single Cell sequencing



Timeline of single cell RNA-seq



2012 2013 2014 2015 2017 2019 2021

100 cells
Smart-seq

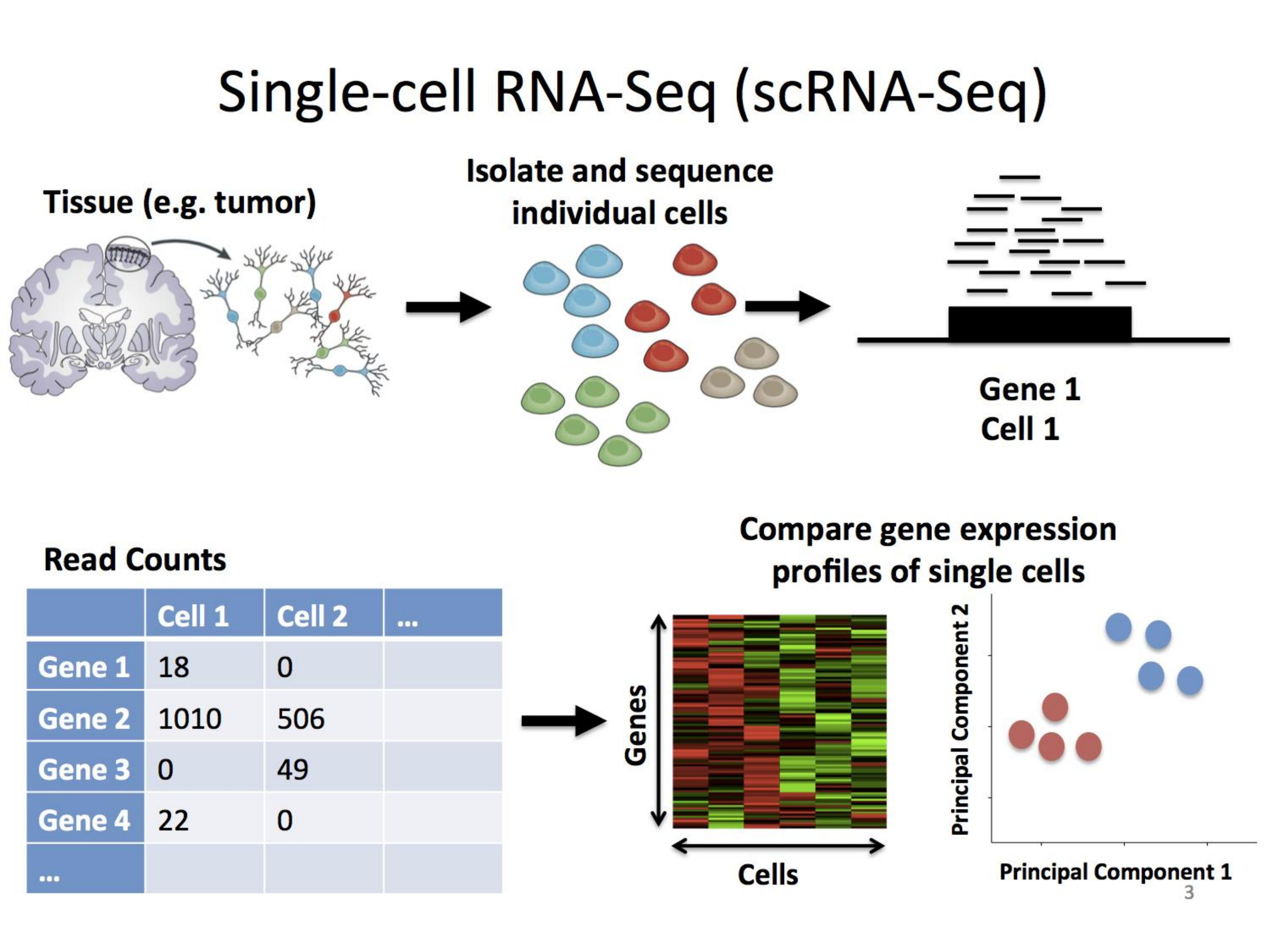
1,000 cells
MARS-seq

10,000 cells
Drop-seq

100,000 cells
Drop-seq

Human Cell Atlas
FET Flagship

Single Cell sequencing



Single-cell RNA-Seq (scRNA-Seq)

Tissue (e.g. tumor)

Isolate and sequence individual cells

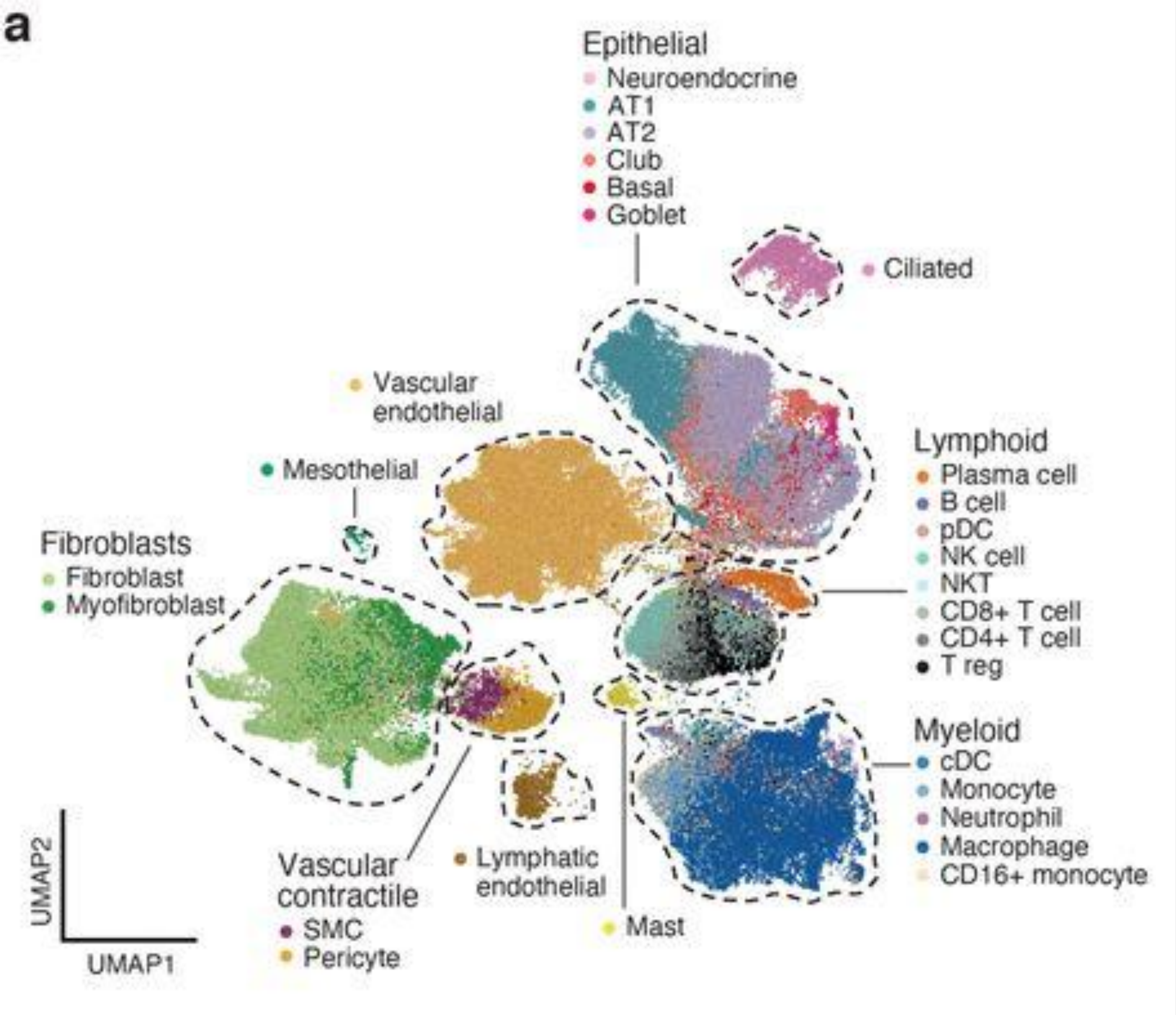
Gene 1
Cell 1

Read Counts

	Cell 1	Cell 2	...
Gene 1	18	0	
Gene 2	1010	506	
Gene 3	0	49	
Gene 4	22	0	
...			

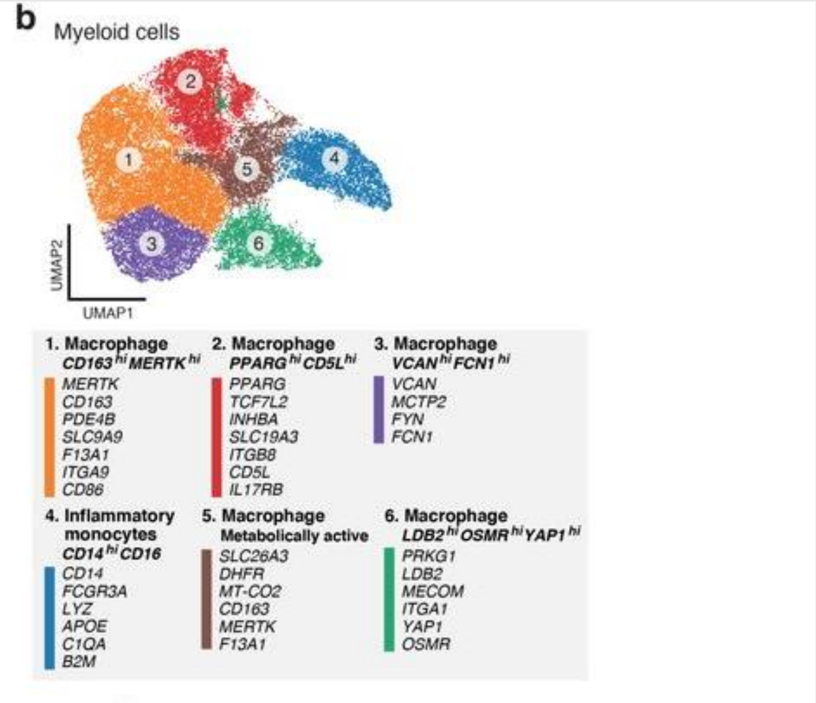
Compare gene expression profiles of single cells

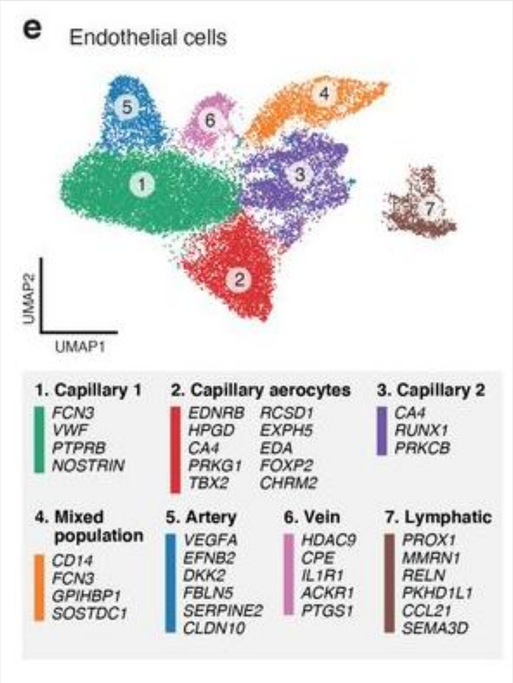
A single cell and single nucleus atlas of COVID-19 lung

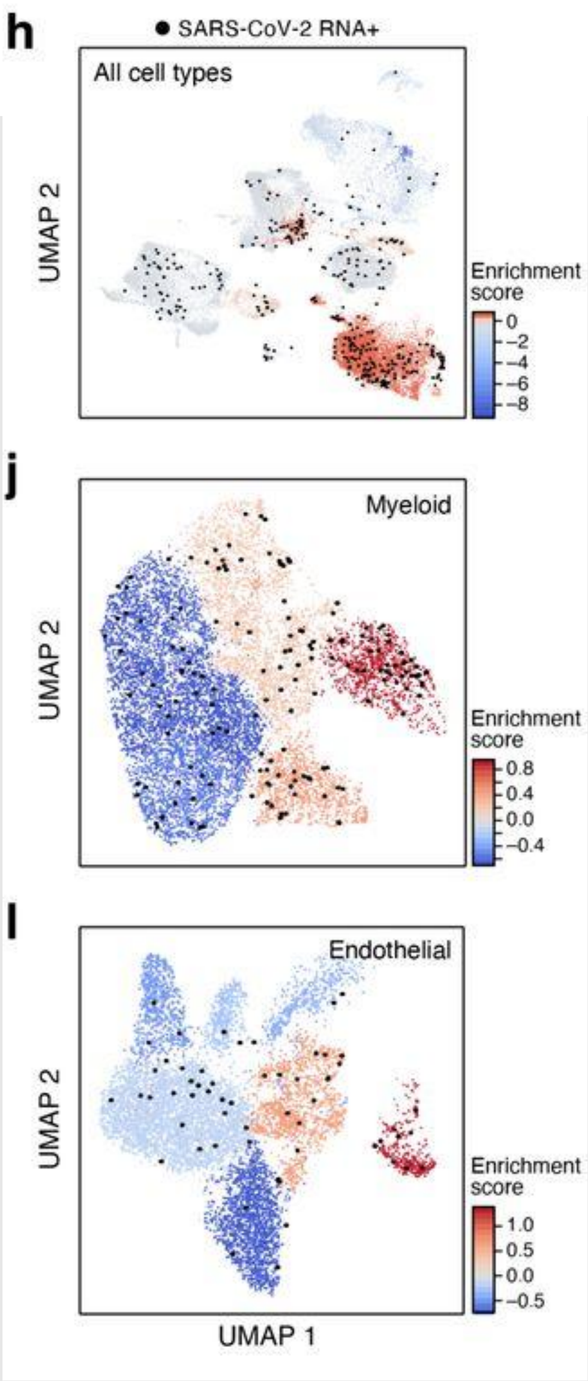


https://doi.org/10.1101/2021.02.25.430130

A single cell and single nucleus atlas of COVID-19 lung

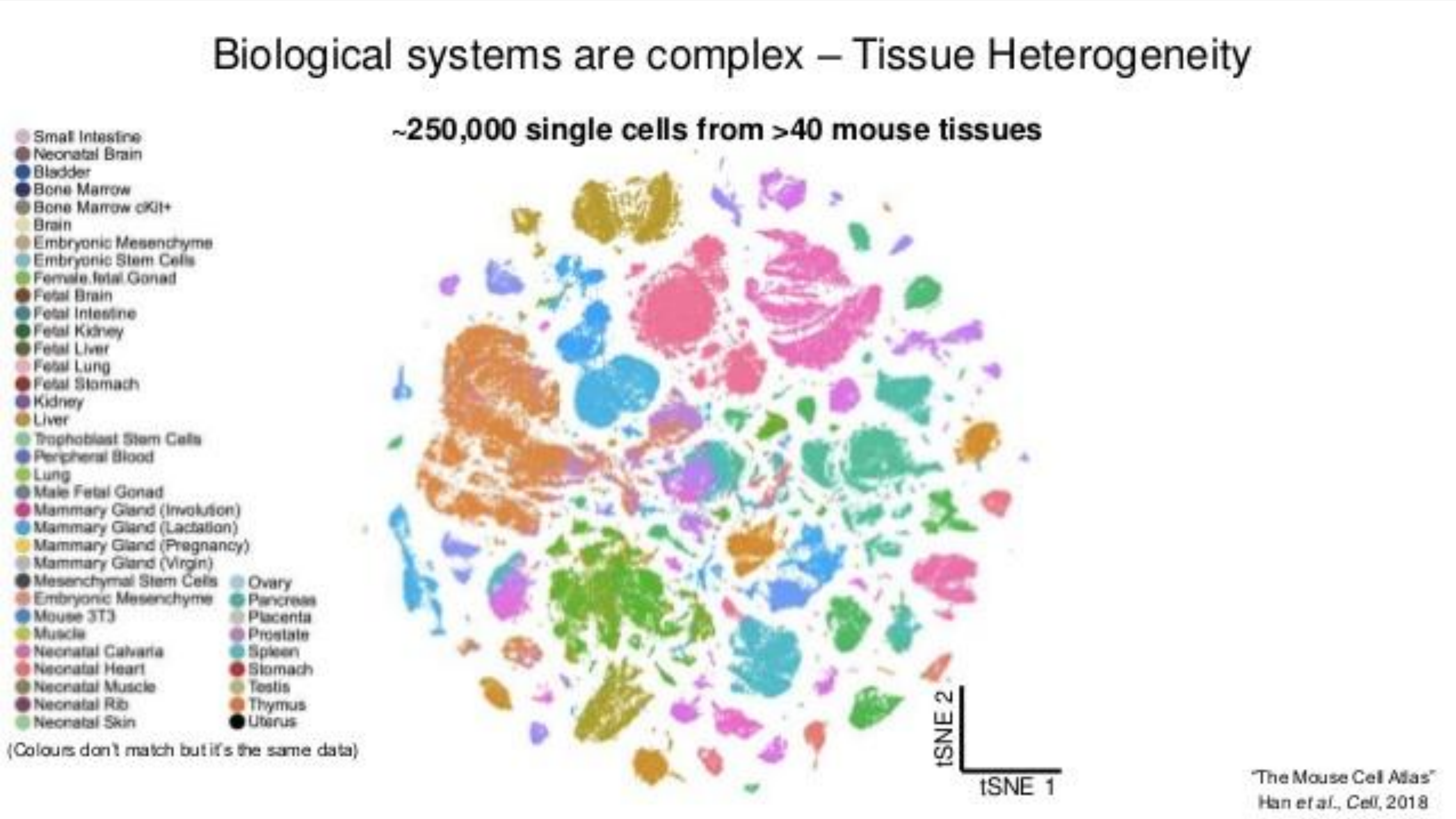






https://doi.org/10.1101/2021.02.25.430130

The Mouse Cell Atlas



Biological systems are complex – Tissue Heterogeneity

~250,000 single cells from >40 mouse tissues

- ● Small Intestine
- ● Neonatal Brain
- ● Bladder
- ● Bone Marrow
- ● Bone Marrow cKit+
- ● Brain
- ● Embryonic Mesenchyme
- ● Embryonic Stem Cells
- ● Female fetal Gonad
- ● Fetal Brain
- ● Fetal Intestine
- ● Fetal Kidney
- ● Fetal Liver
- ● Fetal Lung
- ● Fetal Stomach
- ● Kidney
- ● Liver
- ● Trophoblast Stem Cells
- ● Peripheral Blood
- ● Lung
- ● Male Fetal Gonad
- ● Mammary Gland (Involution)
- ● Mammary Gland (Lactation)
- ● Mammary Gland (Pregnancy)
- ● Mammary Gland (Virgin)
- ● Mesenchymal Stem Cells
- ● Embryonic Mesenchyme
- ● Mouse 3T3
- ● Muscle
- ● Neonatal Calvaria
- ● Neonatal Heart
- ● Neonatal Muscle
- ● Neonatal Rib
- ● Neonatal Skin
- ● Ovary
- ● Pancreas
- ● Placenta
- ● Prostate
- ● Spleen
- ● Stomach
- ● Testis
- ● Thymus
- ● Uterus

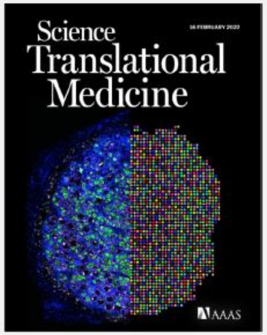
(Colours don't match but it's the same data)

"The Mouse Cell Atlas"
Han *et al.*, *Cell*, 2018

Spatial Transcriptomics

nature
TECHNOLOGY FEATURE | 25 January 2022
Seven technologies to watch in 2022

2022
Spatial Multiomics



Science Translational Medicine
AAAS

2022
1st Visium Cover

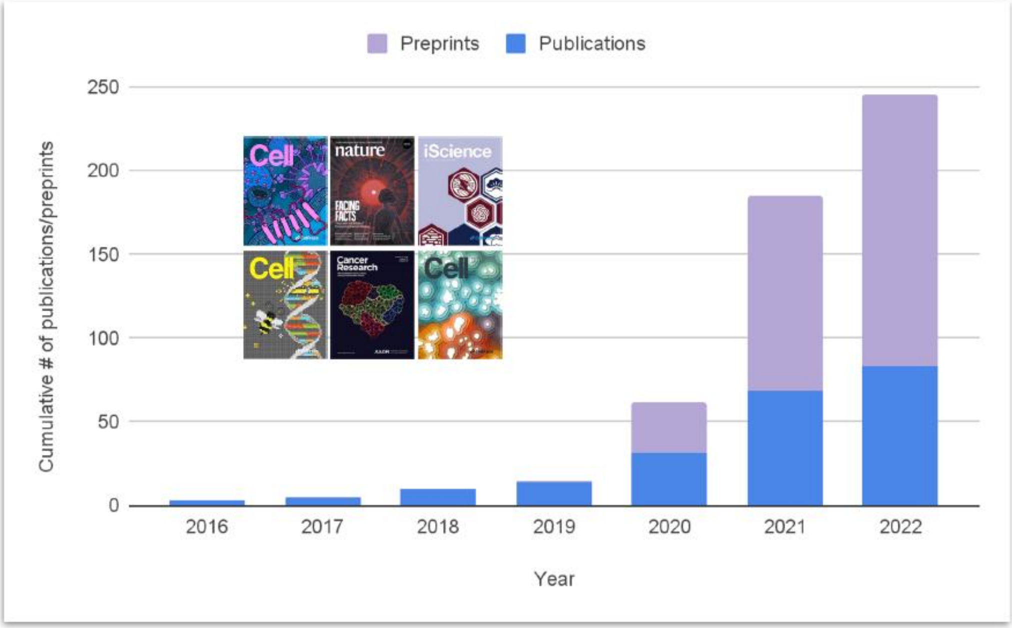
nature methods
Method of the Year 2020: Spatially resolved transcriptomics

2020
Spatially Resolved Transcriptomics

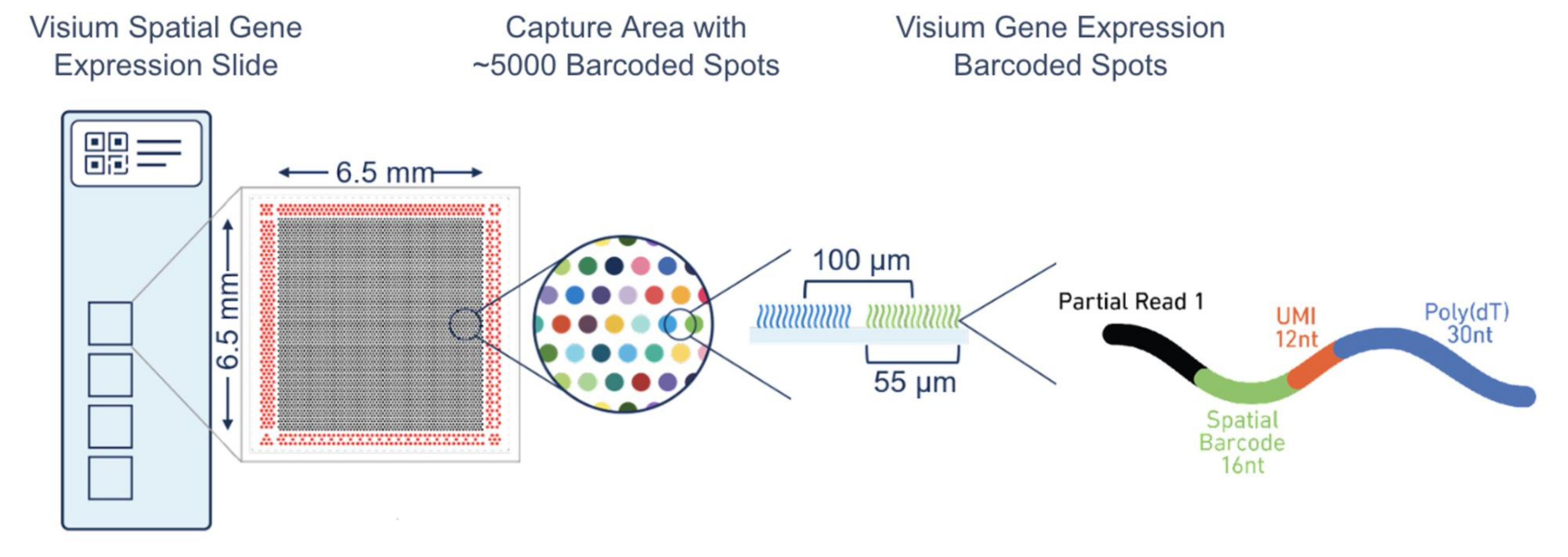
TheScientist
TOP 10
INNOVATIONS

2020
Visium Spatial Gene Expression

200+ Visium Publications and Preprints

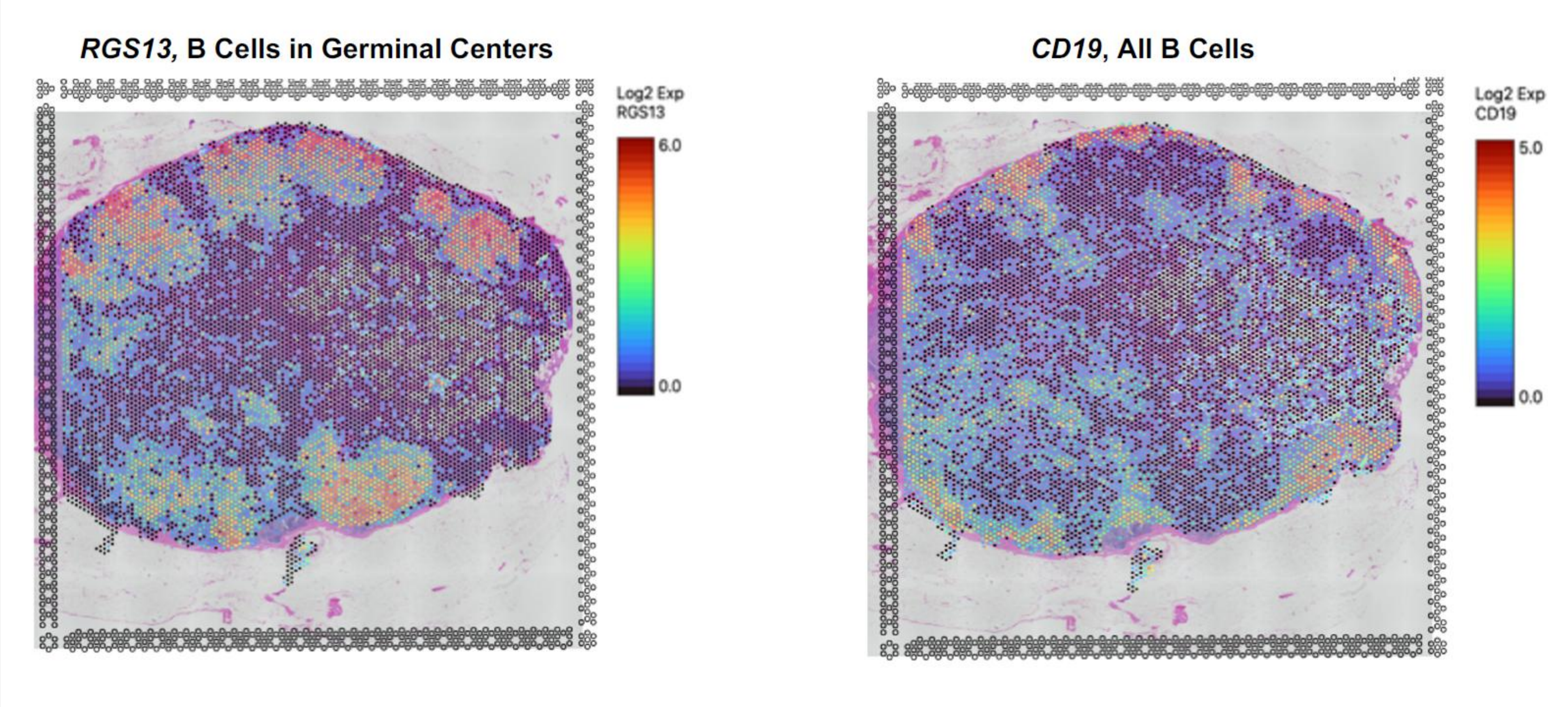


Spatial Transcriptomics


Visium Spatial Gene Expression Slide
Capture Area with ~5000 Barcoded Spots
Visium Gene Expression Barcoded Spots
6.5 mm
6.5 mm
100 µm
55 µm
Partial Read 1
UMI 12nt
Poly(dT) 30nt
Spatial Barcode 16nt

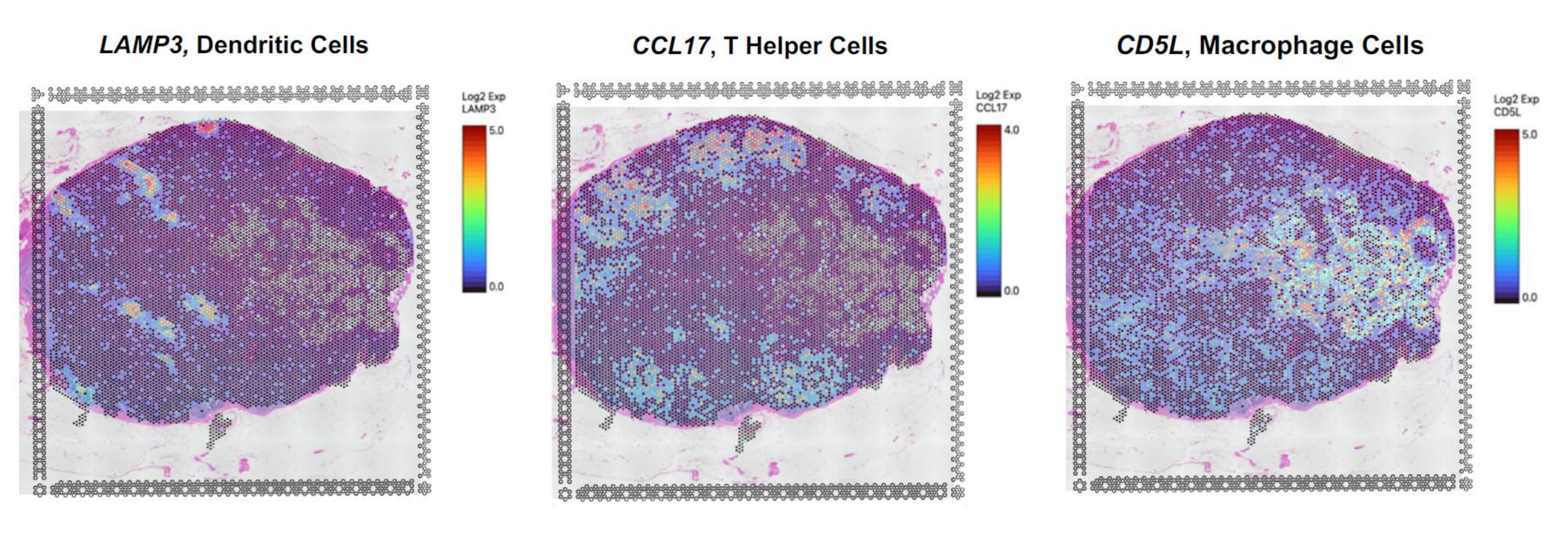
Spatial Transcriptomics

analysis of Human Lymph Nodes



Spatial Transcriptomics

analysis of Human Lymph Nodes

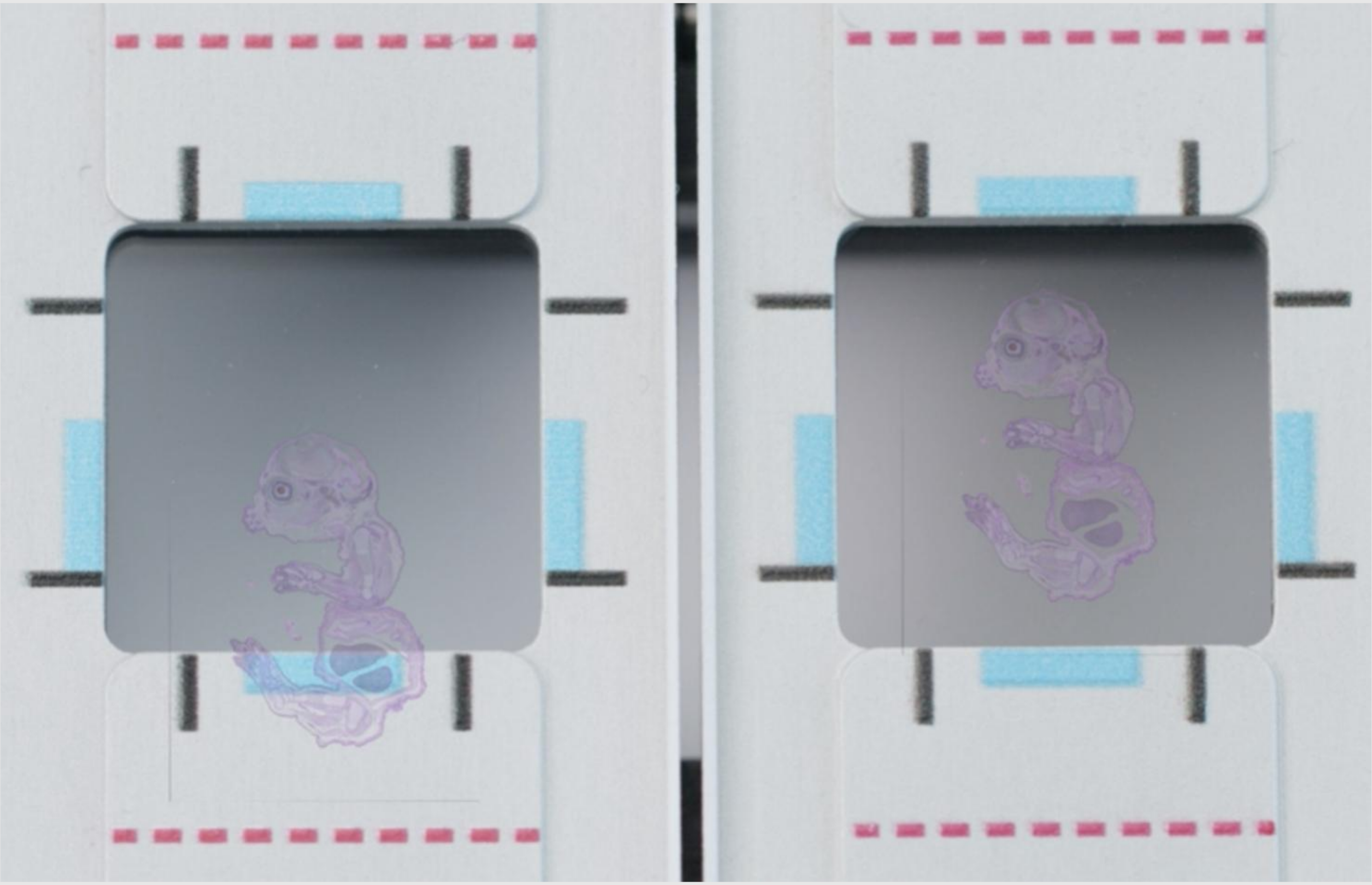


LAMP3, Dendritic Cells

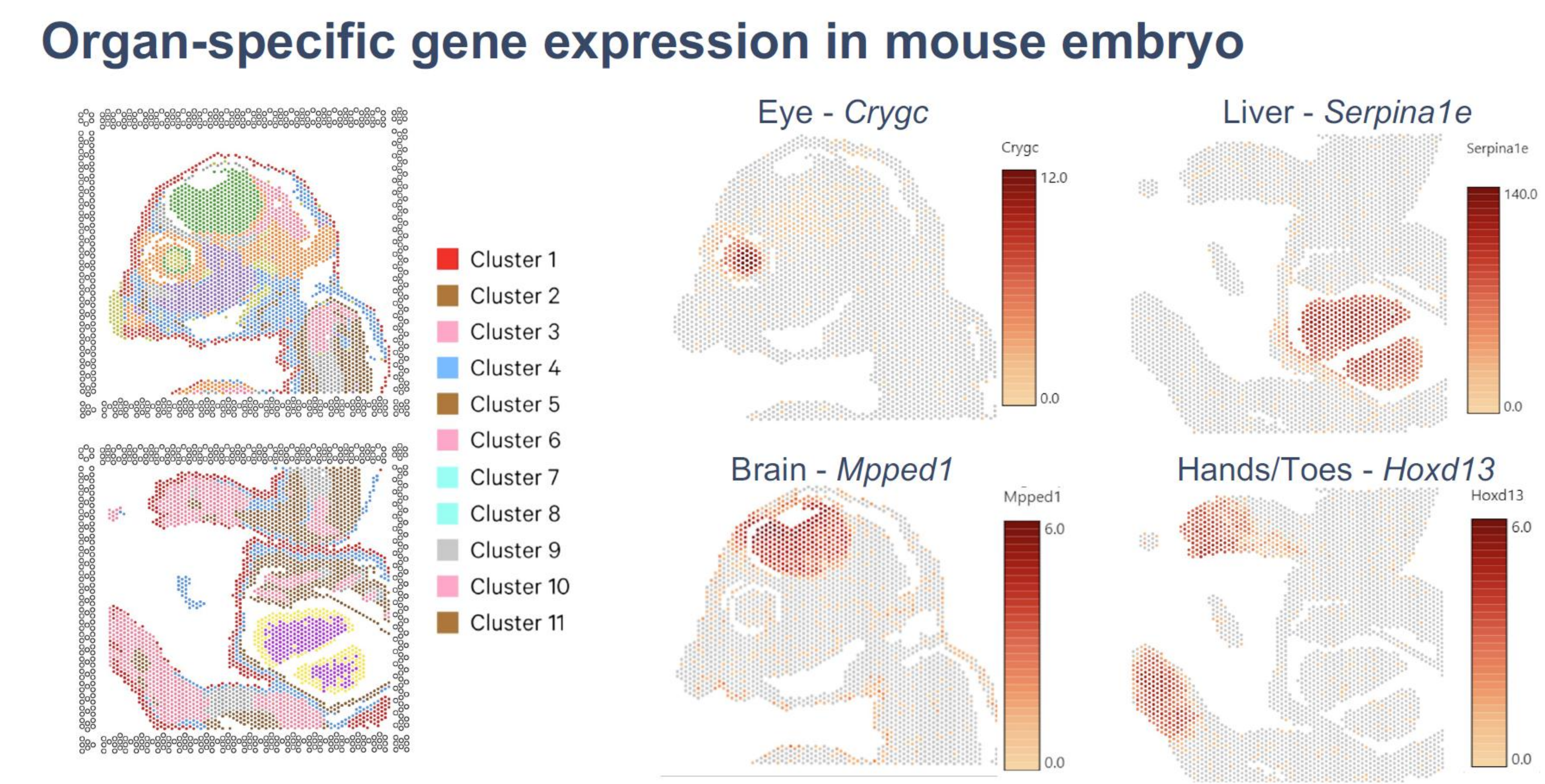
CCL17, T Helper Cells

CD5L, Macrophage Cells

Spatial Transcriptomics



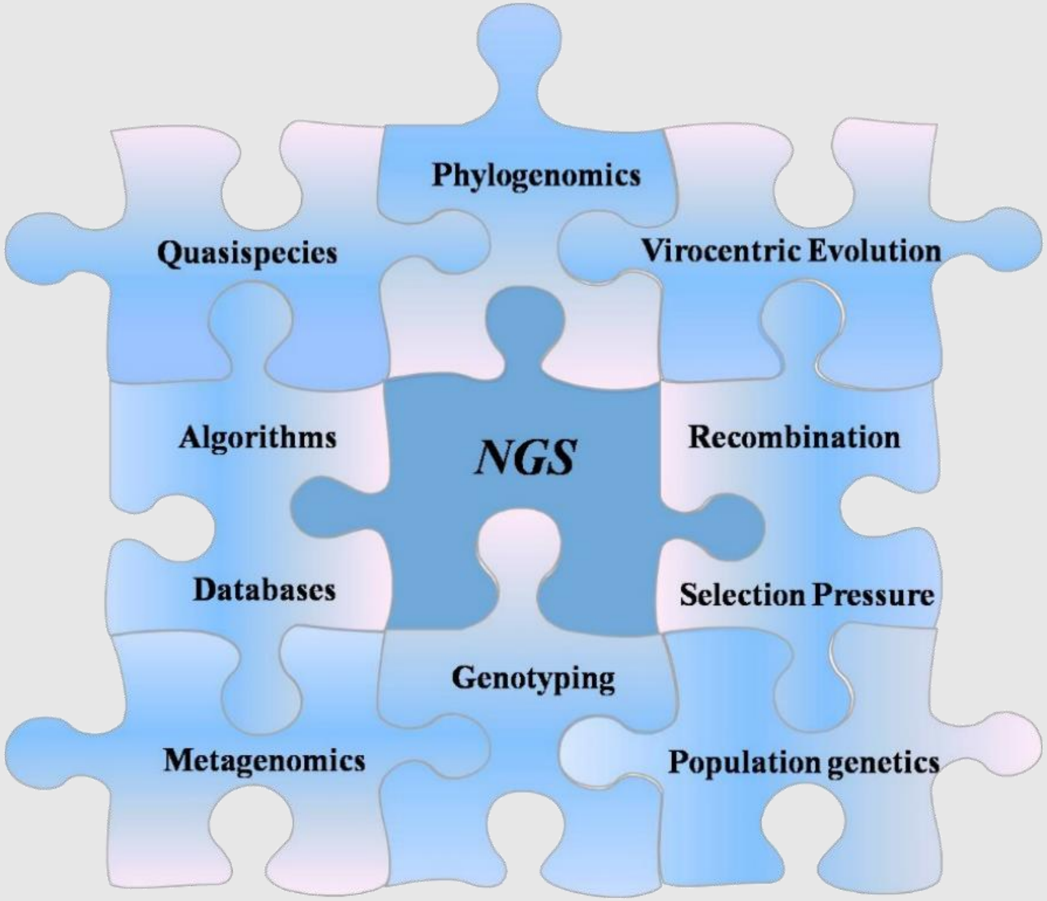
Spatial Transcriptomics



Organ-specific gene expression in mouse embryo

- 🟥 Cluster 1
- 🟫 Cluster 2
- 🟪 Cluster 3
- 🟦 Cluster 4
- 🟫 Cluster 5
- 🟪 Cluster 6
- 🟦 Cluster 7
- 🟦 Cluster 8
- ⬜ Cluster 9
- 🟪 Cluster 10
- 🟫 Cluster 11

Thank you for your attention !!


Phylogenomics
Quasispecies
Virocentric Evolution
Algorithms
NGS
Recombination
Databases
Selection Pressure
Genotyping
Metagenomics
Population genetics