



# 2001: A Base Odyssey

The era of genomics and massive parallel sequencing

---

Matthias Zepper, PhD

February 24, 2025

NGI Stockholm

<https://ngisweden.scilifelab.se>

2001: Draft assemblies of the human genome are published



**Figure 1:** The private company Celera [Venter et al., 2001] and the International Human Genome Sequencing Consortium [Lander et al., 2001] both publish a draft sequence of the euchromatic portion of the human genome.

# The overture to the genomic era



A remake of the opening scene by SumoSebi, CC-BY-SA on Wikimedia Commons

Stanley Kubrick's *2001- A Space Odyssey* premieres 2 April 1968

1968: Nobel prize for the interpretation of the genetic code

## Nobel Prize in Physiology or Medicine 1968



Photo from the Nobel Foundation archive.

Robert W. Holley

Prize share: 1/3



Photo from the Nobel Foundation archive.

Har Gobind Khorana

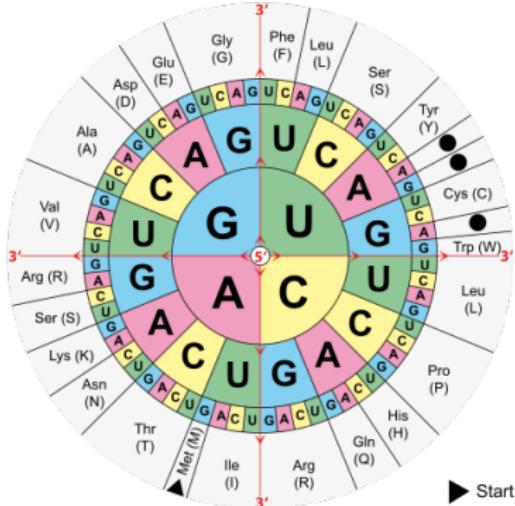
Prize share: 1/3



Photo from the Nobel Foundation archive.

Marshall W. Nirenberg

Prize share: 1/3



► Start  
● Stop

- The genetic code is (almost) universal<sup>[1]</sup>
- It was resolved entirely using synthetic sequences.

[1] <http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=tgencodes>

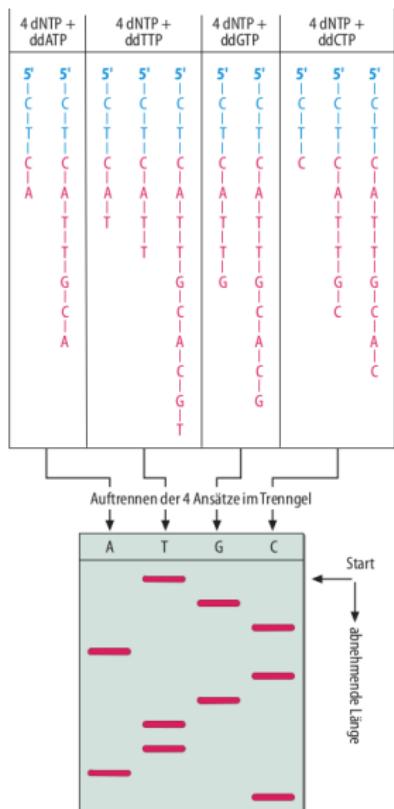
# Encoded information of naturally occurring DNA unknown



- Peptides could be sequenced since the 1950s (Sanger method, Edman degradation).
- Sequencing of DNA was one of the most urgent, unresolved problems in the early 1970s.
- Frederick Sanger (Nobel laureate for sequencing Insulin 1958) started working with DNA.

*F. Sanger*

# 1977: Chain-termination sequencing by Frederick Sanger



- DNA fragments could be separated by size.
- Sanger's method creates sequence-derived length patterns.
- It relies on radioactive labeling and in-vitro amplification of DNA.

## DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage  $\phi$ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge

Figure 2: [Sanger et al., 1977]

# 1980: Nobel prize for DNA sequencing

## Nobel Prize in Chemistry 1980



Photo from the Nobel Foundation archive.

**Paul Berg**

Prize share: 1/2



Photo from the Nobel Foundation archive.

**Walter Gilbert**

Prize share: 1/4



Photo from the Nobel Foundation archive.

**Frederick Sanger**

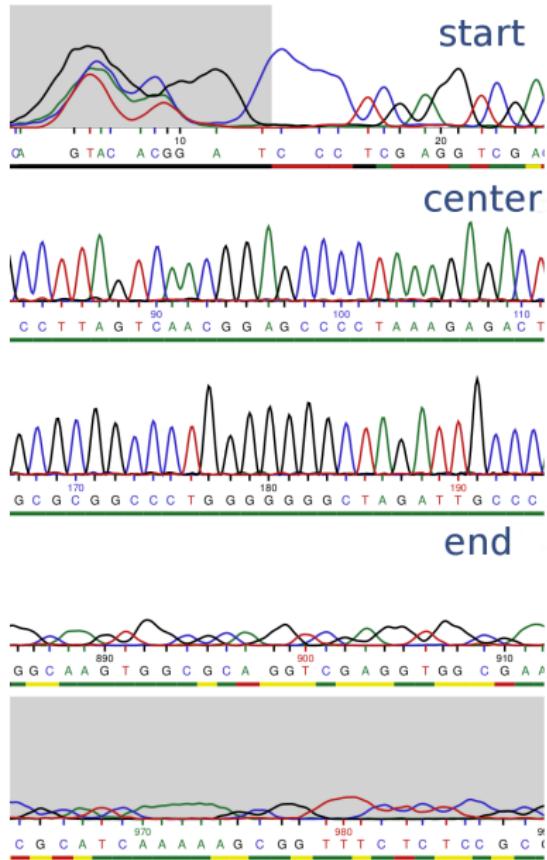
Prize share: 1/4

- Ample DNA input needed  
PCR was introduced in 1989
- Four reactions per sequence
- Read length  $\sim$  200bp



<https://www.nobelprize.org/prizes/chemistry/1980/summary/>

# Advanced Sanger sequencing for the Human Genome Project

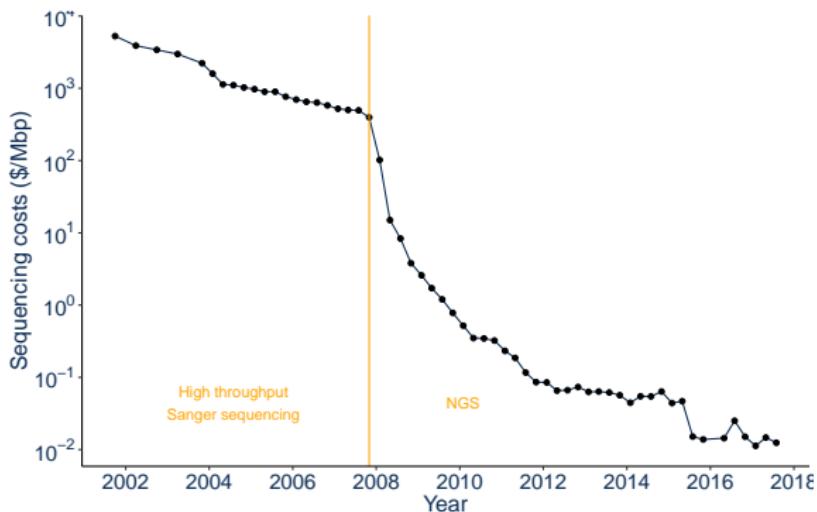


- Fluorescent chain terminators.
- Capillary electrophoresis for size separation of amplicons.
- Parallelized and automated.
- Sequencing technology of the Human Genome Project (1990-2004).

## **Next-generation sequencing**

---

## New high-throughput methods were developed



**Figure 3:** Sequencing costs per one million bases of raw sequence

**1990-2004:** Human Genome Project sequencing: US \$500 million

**2025:** Sequencing of a human genome: ~ US \$100-1000

National Human Genome Research Institute (NHGRI)

<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>

Around 2010: Sanger sequencing was outcompeted by NGS



**ABI 3730xl DNA Sequencer**  
(Sanger Multiplex, 2013)

- ~6912 reads of 400bp
- ~2,76 Mbp / day



**Illumina HiSeq 2500**  
(NGS / MPS, 2013)

- ~600 Million reads of 100bp
- ~60.000 Mbp / day

(depending on settings and sequencing chemistry used)

# National Genomics Infrastructure Sweden

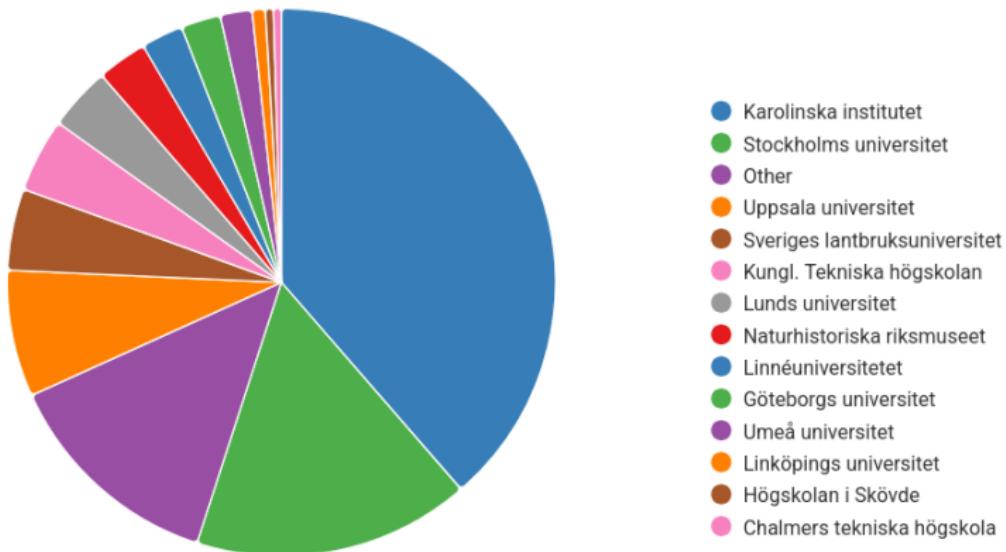
---

DNA sequencing facilities provide sequencing capacity



- DNA sequencing of paramount importance for life science.
- 2013: National Genomics Infrastructure Sweden is founded.
- Our mission is to offer a state-of-the-art infrastructure available to researchers all over Sweden.

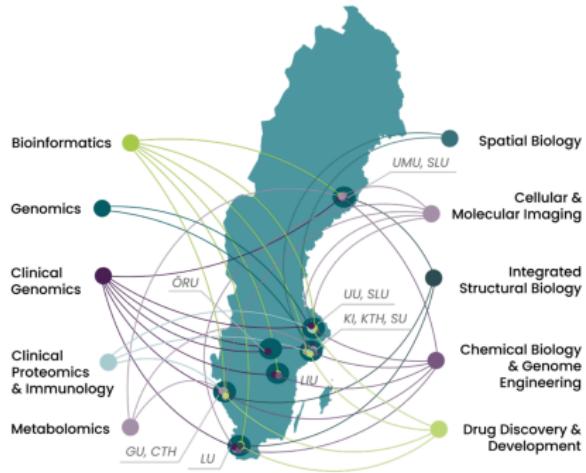
## Project Affiliations in 2024



<https://ngisweden.scilifelab.se/resources/ngi-stockholm-status/>



- NGI is a sequencing facility for *research projects*
- Part of the Genomics Platform at SciLifeLab
- Distributed in 3 nodes:
  - SNP&SEQ Technology Platform, Uppsala
  - Uppsala Genome Center
  - NGI Stockholm + Eukaryotic Single Cell Genomics (ESCG), Solna

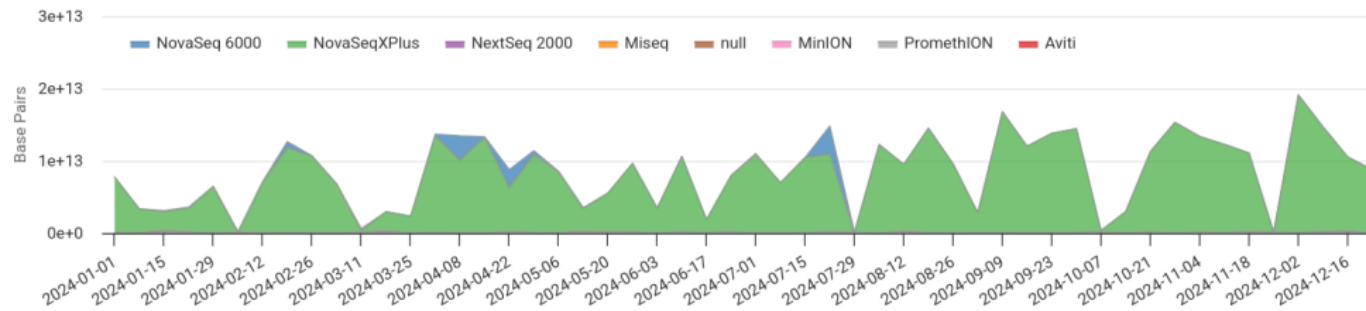


# NGI-S employs various sequencing technologies



## Sequencing Throughput

Average for 52 weeks: 1234 Gbp per day  
(1 Human genome equivalent every 3.77 minutes)



- In 2024, NGI Stockholm sequenced on average 1200 Gbp/day

<https://ngisweden.scilifelab.se/resources/ngi-stockholm-status/>

## Sequencing platforms

---

# Sequencing platforms / technologies since Sanger

## Next generation sequencing

- Roche 454 sequencing (Pyrosequencing)
- Ion semiconductor sequencing
- **Illumina (Solexa) sequencing**
- **PacBio HiFi Sequencing**

## Third generation sequencing

- **Oxford Nanopore sequencing**
- **Element Biosciences Avidite Sequencing**
- Ultima Genomics UG 100 Sequencing
- MGI DNBSEQ Technology
- Singular Genomics G4X

Platforms in **bold** are in use at the National Genomics Infrastructure

# Sequencing platforms / technologies since Sanger

## Sequencing by synthesis

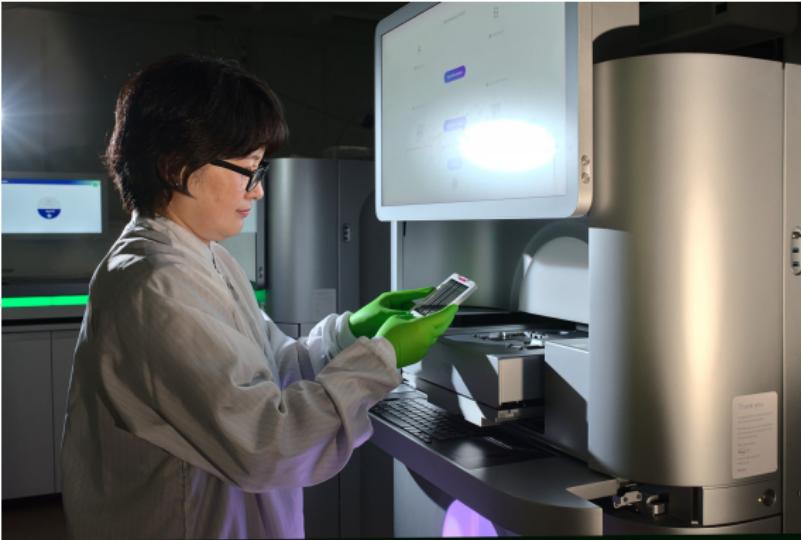
- Roche 454 sequencing (Pyrosequencing)
- Ion semiconductor sequencing
- **Illumina (Solexa) sequencing**
- **PacBio HiFi Sequencing**
- **Element Biosciences Avidite Sequencing**
- Ultima Genomics UG 100 Sequencing
- MGI DNBSEQ Technology
- Singular Genomics G4X

## Direct DNA/RNA sequencing

- **Oxford Nanopore sequencing**

Platforms in **bold** are in use at the National Genomics Infrastructure

# Illumina sequencing is *the* NGS sequencing platform



Illumina's sequencing by synthesis technology is NGI's bread-and-butter platform

# Preparation for sequencing (in the lab)

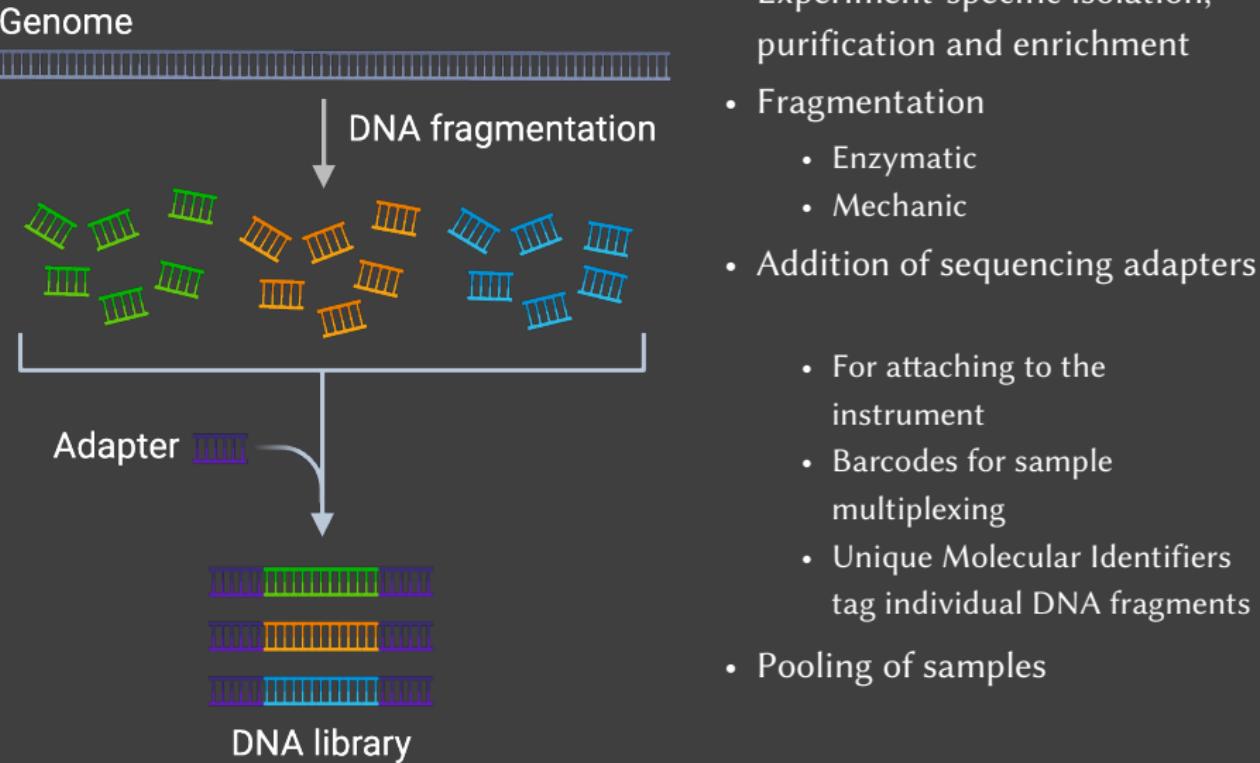


Figure by Anja Mezger

# Preparation for sequencing (on the machine)

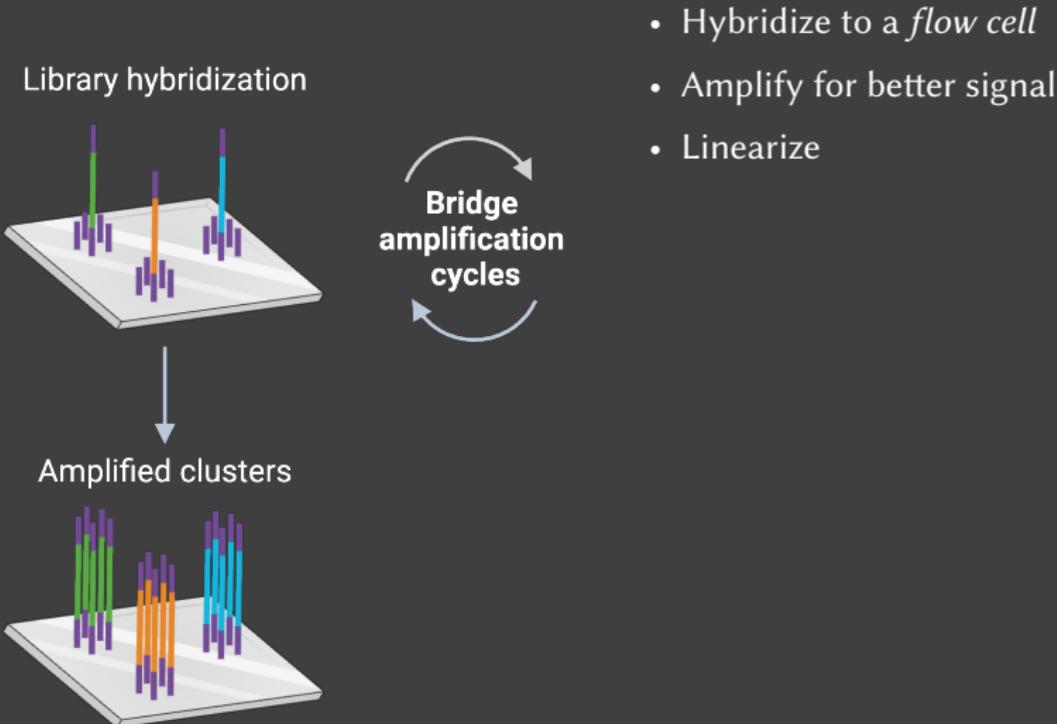
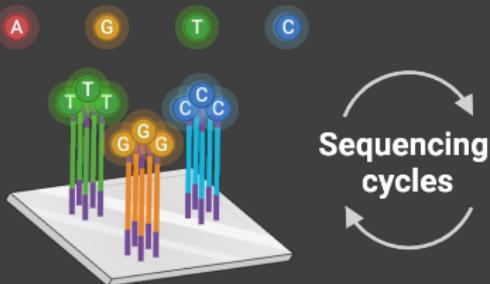


Figure by Anja Mezger

# Illumina: *Sequencing by Synthesis* of DNA clusters

Fluorescently labeled nucleotides



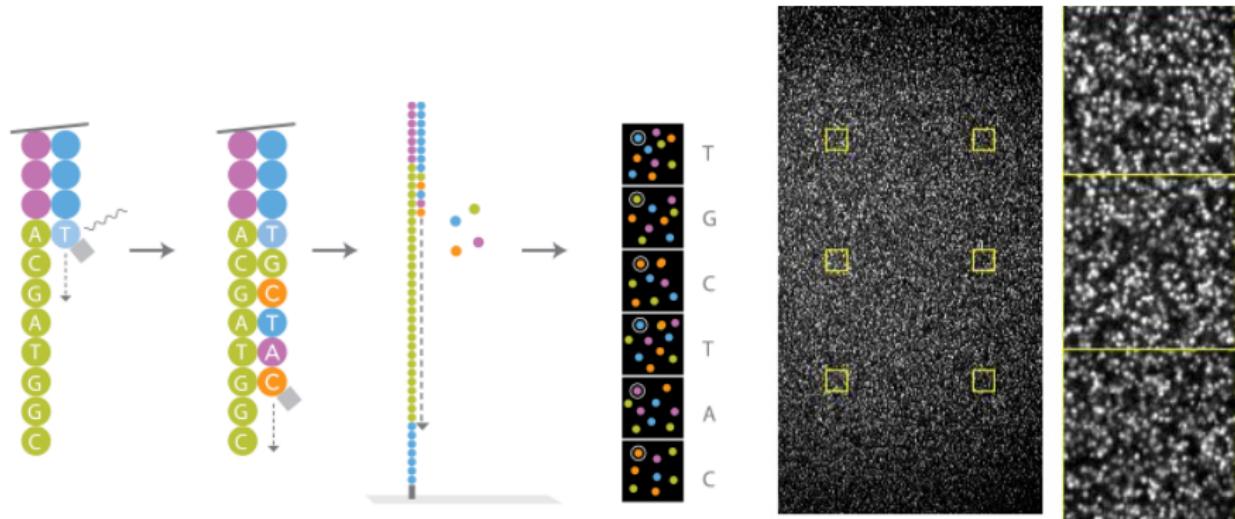
Data collection



- DNA is amplified (again)
- Base integration yields a light signal (details vary among Illumina machines)
- Sequence is derived from a time-series of images

Figure by Anja Mezger

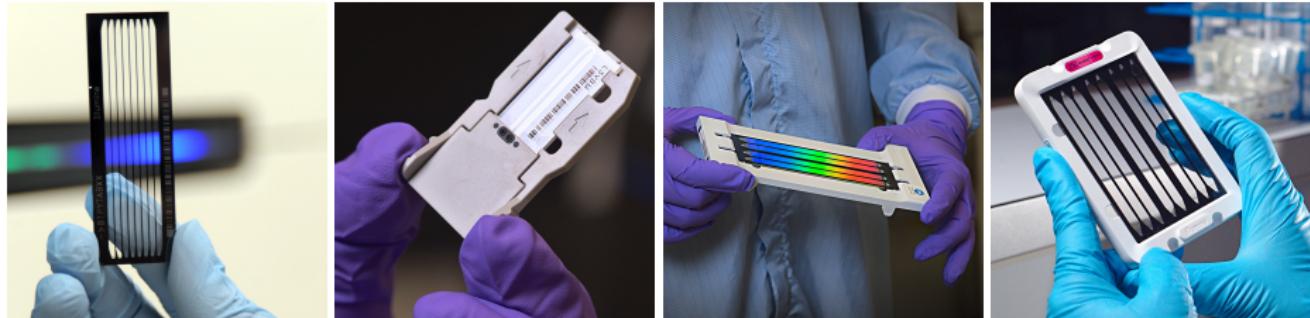
# Illumina: Sequencing by Synthesis of DNA clusters



1. Integration of base is monitored directly
2. Image sequence is recorded
3. For each cluster, the light/dark pattern is converted into a DNA sequence

→ Highly parallelized, direct monitoring as synthesis proceeds

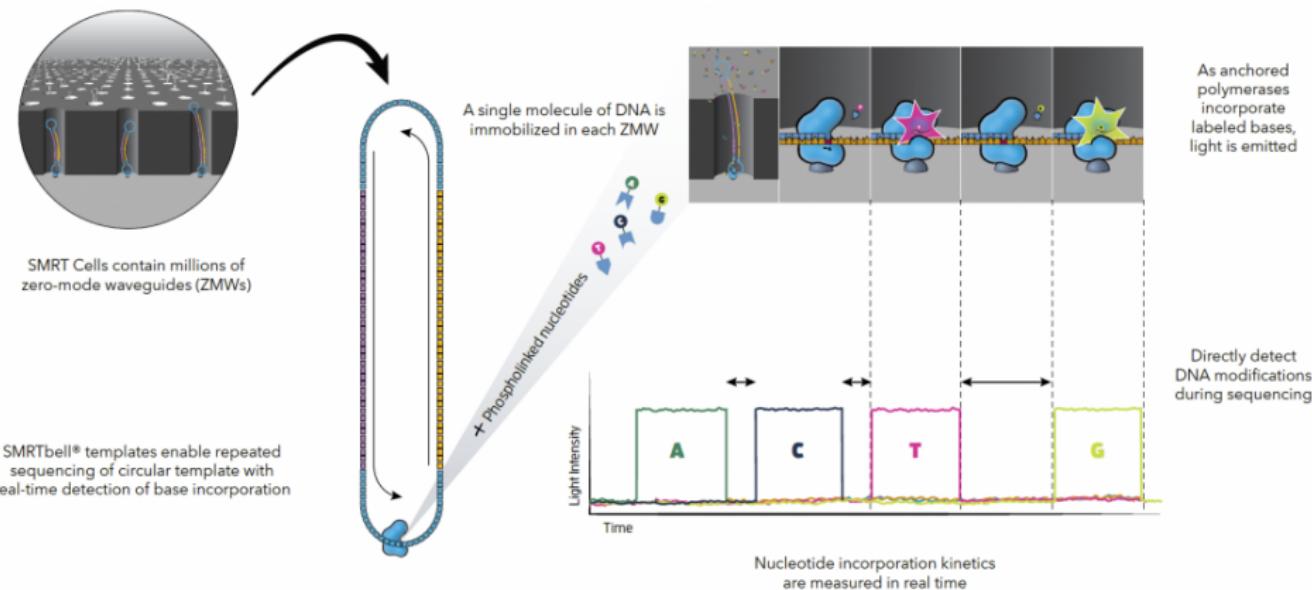
## Flow cells instead of plates: Massive parallel sequencing



**Figure 4:** Various Illumina flow cells

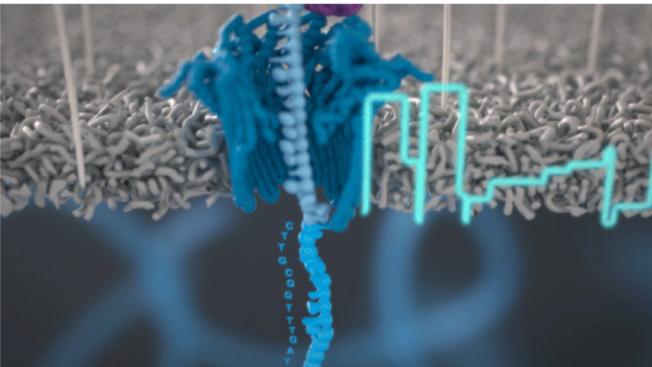
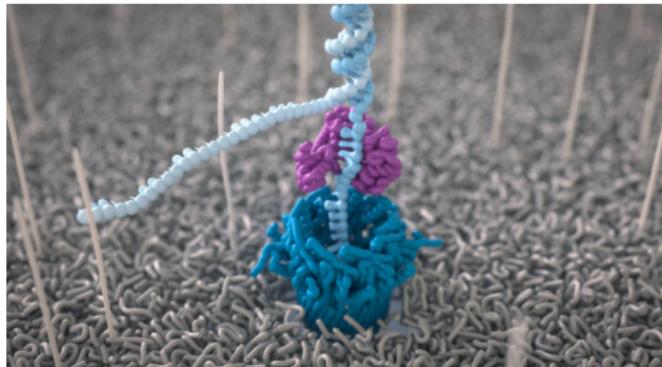
- Illumina's platform produces 2x 150bp reads from a fragment.  
→ **short-read sequencing**
- Instead of 6912 fragments like with Sanger, Illumina machines can sequence Millions to Billions in parallel → **massive-parallel sequencing**

# PacBio: Single-molecule sequencing by synthesis



1. PacBio can generate longer reads than Illumina.
2. Circular libraries, fragment is sequenced repeatedly.

# Oxford Nanopore: Sequencing by electric conductivity



1. DNA is sequenced without amplification
2. A motor protein pulls a DNA strand through a pore (protein channel or solid state)
3. Bases cause specific conductivity changes
4. Direct reading of RNA and detection of methylated bases.

# NGI provides sequencing platforms for every need

## **Standard**

- Illumina sequencing

## **Longer reads, less base-call errors**

- PacBio HiFi Sequencing

## **Much longer reads, many more base-call errors**

- Oxford Nanopore sequencing

## **Short-reads, fewest base-call errors:**

- Element Biosciences Avidite Sequencing

## Sequencing data handling

---

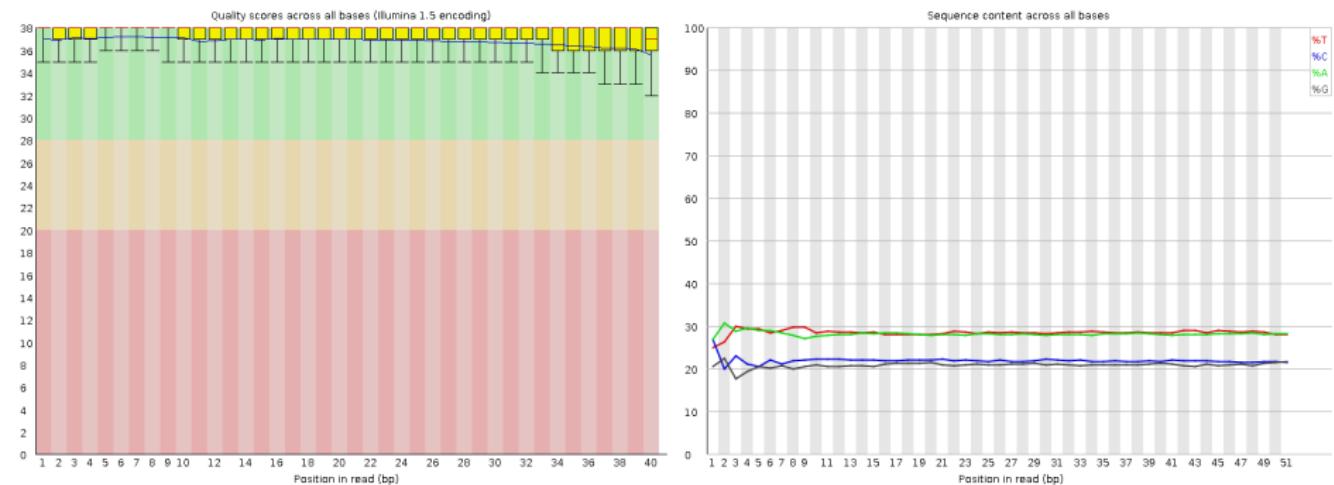
# Sequencing result: Terabytes of data in FastQ-format

## A single read:

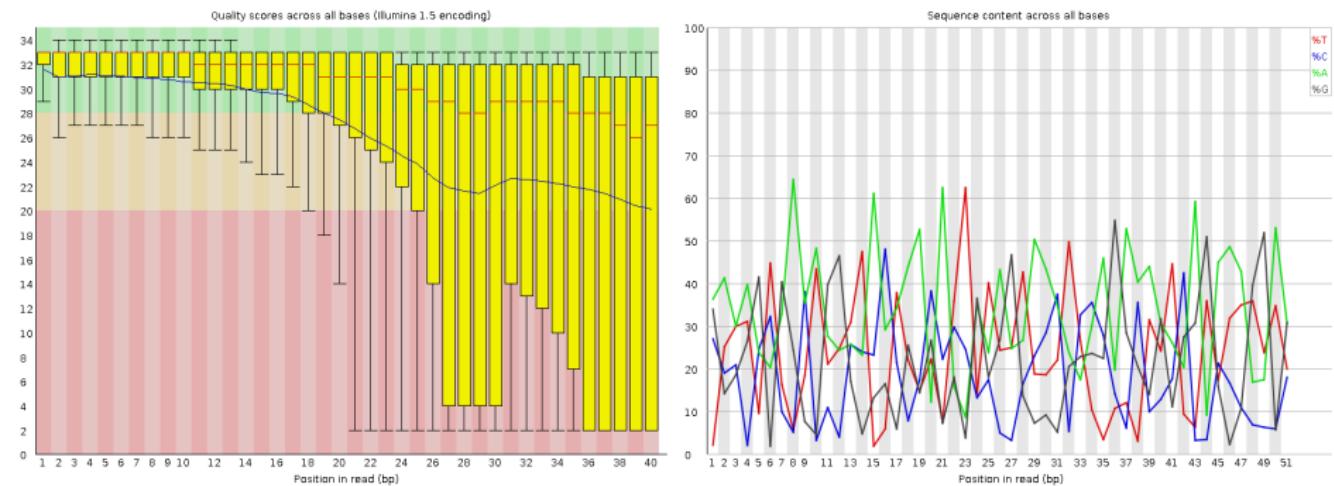
1. Read ID
  2. DNA-sequence
  3. +
  4. Error rate of the base call
- 

```
@M00463:56:000000000-AD76D:1:1101:15189:1873 1:N:0:1
ATAAACACGGTCTTTCCAGGTCAAGCCGGACGGTACCGCCCTGTGGCCATCGAA
+
-86<<F9FB7FFFGGGGGFACGFFDEFGGGEFE>FGGGDGFGFGGCFIG?FF7E@
```

# Quality control: Good data



# Quality control: Poor data



# Common bioinformatic analyses

```
>NC_001422.1 Escherichia phage phiX174
GAGTTTATCGCTTCATGACGAGAATTAACTCTGGATATTCGATGAGTCGAAAAATT/
GATAAAAGCAGGAATTAACTACTGCTTGTTACGAATTAAATCGAAGTGGACTCTGGCGAAAATG/
ATTCGACCTATCCTTGGCAGCTCGAGAAGCTTACTTGCACCTTCGCCATCAACTAACGAT
TCAAAAAACTGACCGGTGGATGAGGAGAAGTGGCTTAATATGCTTGGCACGTCGTCAGGACTG
GATATGAGTCACATTGTTAGAGATTCTTGTGACATTAAAAGAGCTGGATTAA
TGAGTCGATGCTGTTCAACCACTAATAGGTAAGAAATCATGAGTCAGTTACTGAAACATCGT
TCCAGACCCTTGGCTCTATTAGCTCATCGGCTCTGCCGTTTGGATTAACCGAGATG
CGATTTCCTGACGAGTAACAAAGTTGGATTGCTACTGACCGCTCTGCTGCTGCTGGTTG
TGGCTTATGGTACGCTGGACTTGGACCTCCGCTTCCGCTCTGCTGCTGCTGGTTG
TCATGGCTTATTAGTGTACCCGTCAACCTAACAGCGCTGTCTCATGGAAAGCTGTAAT
GGAAAAACATTATAATGGCTCGAGCGTCCGGTAAAGCCGTAATTGTCGCTTACCTGGC
CGCGCAGGAAACACTGACGTTTACTGACGCGAGAAGAAAACGTCGCTAAATACTGTCGGA
TGATGTAATGCTCAAAGTAAAAACGTTCTGGCCTCGCCCTGGTGTCCGAGCCGTTGCGAG
AAAGCCAGCGTAAGGGCGCTCGCTTGGTATGAGGTGTAACAAATTGAGGGCTT
CCCTTACTTGAGGATAAATTATGCTAATATTCAAACACTGGCGCGAGCGTATGCCGATGACCTT
TCTGGCTTCTGCTGGTCAAGATTGGCTGCTTATTACCATTCACACTCGGGTATGCTG
TCCTTGAGAGATTGACGCCGTTGGCGCTCTCGCTTCTCATGGCTGCGCTTGTATGAG
CTGAGACATTAACTTTTATGCTCTCATGCTCACGTTATGGTAACAGTGGATTAAAGTTG
GGATGGTGTAAATGCCACTCTCCCGACTGTTAACACTACTGGTATATTGACCATGCCGTT
GGCACGATTAACCCGTACCAATAAAATCCCTAACGATTTGGTTCAGGGTTATTGAAATCTAT
ACTATTAAAGGCCGCTGGATGCTGACCGTACCGAGGCTAACCTAACTGAGCTTAATCAAGATG
TCGTTATGGTTCCGTTGCTGCCATCTCAAAACATTGGACTGCTCCGCTTCTCTGAGACTG
[...]
```

- **pairwise Alignment:**

Find the exact origin of a short fragment in a long reference.

- **Quasi-mapping:**

Which reference is the most-likely origin?

- **De-novo assembly:**

Create a long reference from short fragments.

# Analyses as if we were back in the sixties

## De novo assembly of contigs (fault-tolerant)

nsver, my friend

owin' in the wind. The ans

e amber my fr

Technical read error / mutation

y friend is blowin'

The answer is blowin' in

my friend, is blow

e wind. The answe

e answer, my fr

in the wind. The answer is blowin'

The answer, my friend, is blowin' in the wind. The answer is blowin' in the wind.

## Alignment (fault-tolerant)

my vriend

Technical read error/ mutation

Theeeeeeee answ\* end

Indel

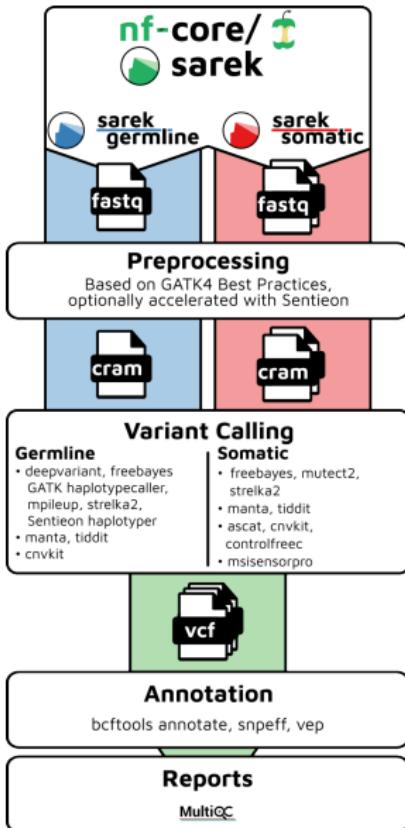
answer



answer

Multimapper

# Analyses as if we are back today



- Data pipelines combine sequential steps.
- Workflow managers execute pipelines and scale analyses to many samples.
- Collaborative communities for workflow development help you to get started:
  - <https://nf-co.re>
  - <https://anvio.org>

## Weblinks

- Own lecture on NGS data analysis  
<https://github.com/MatthiasZepper/Lecture-OmicsDataAnlysis>
- Course Materials on sequencing data science  
<http://data-science-sequencing.github.io>
- DNA Sequencing Coursera class slides  
<https://github.com/BenLangmead/ads1-slides>
- Genome Browser (Easy access to selected genomes)  
<http://genome-euro.ucsc.edu>
- European Nucleotide Archive (Complete genomes and contigs)  
<https://www.ebi.ac.uk/ena>
- Current human reference genome (version 38)  
<http://ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/>

## Sequencing applications

---

# We are surrounded by genetic information



Library of the Human Genome, Wellcome Collection. Photo by Ben Gilbert and Thomas Farnetti

We are surrounded by genetic information: Many applications

**Science**

**RESEARCH ARTICLES**

**Three-dimensional intact-tissue sequencing of single-cell transcriptional states**

Xiao Wang<sup>1\*</sup>, William E. Alles<sup>1,2\*</sup>, Matthew A. Wright<sup>1,3</sup>, Emily L. Sylwestrak<sup>1</sup>, Nikolay Samusik<sup>1</sup>, Sam Vesuna<sup>1</sup>, Karun Chari Ramakrishnan<sup>1</sup>, Jia Liu<sup>1</sup>, Garry P. Nolan<sup>1,2</sup>, Felice-Alessio Rava<sup>1</sup>, Karl Deisseroth<sup>1,3,4,5</sup>

<sup>1</sup>Department of Bioengineering, Stanford University, Stanford, CA 94301, USA; <sup>2</sup>Stanford Bio-X, Stanford, CA 94301, USA; <sup>3</sup>Stanford University School of Medicine, Stanford, CA 94301, USA; <sup>4</sup>Stanford Neurosciences Institute, Stanford, CA 94301, USA; <sup>5</sup>Stanford Department of Psychology, Stanford, CA 94301, USA

**LETTER**

doi:10.1038/nature10554

**Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination**

Eduardo<sup>1</sup>, Zakeya Al Rasbi<sup>1</sup>, Marina Boziki<sup>1</sup>, Caroline Johner<sup>2</sup>, Hartmut Wekerle<sup>1</sup>

**Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing**

Yong-hui Jiang,<sup>1,18</sup> Ryan K.C. Yuen,<sup>2,18</sup> Xin Jin,<sup>3,4,5,18</sup> Mingbang Wang,<sup>3,18</sup> Nong Chen,<sup>3</sup> Xueli Wu,<sup>3</sup> ...  
Jing Wang,<sup>1,6,7,8,18</sup> ...  
Liu Li,<sup>1,9,10,11,12,13,14,15,16,17,18</sup> ...  
ARTICLE

doi:10.1038/nature10555

**A map of human genome variation from population-scale sequencing**

The 1000 Genomes Project Consortium\*

**A Draft Sequence of the Neandertal Genome**

Richard E. Green,<sup>1,2,†‡</sup> Johannes Krause,<sup>1,3,§</sup> Adrian W. Briggs,<sup>1,4,§</sup> Tomislav Maricic,<sup>1,4,§</sup> ...  
Emre S. Maenpaa,<sup>1</sup> Shahrooz Taheri<sup>2</sup>, Edward K. Gibbons<sup>3</sup>, Yanhong Li<sup>2</sup>, Peter D. Pääbo<sup>1</sup>, ...  
Anastasios Tsiplakos<sup>1</sup>, ...  
**Ancient human genome sequence of an extinct Palaeo-Eskimo**

Richness of human gut microbiome correlates with metabolic markers

Emmanuelle Le Chatelier<sup>1,\*</sup>, Trine Nielsen<sup>2,\*</sup>, Junji Qin<sup>3,\*</sup>, Edi Prifti<sup>1\*</sup>, Falk Hildebrand<sup>4,5</sup>, Gwen Falony<sup>4,5</sup>, Mathieu Mavingagu<sup>4,5</sup>, Anne-Sophie Artigues<sup>1</sup>, ...  
Jens Roat Kultima<sup>1</sup>, Daniel Kristoffer Forslund<sup>1</sup>, Shinichi Sunagawa<sup>1</sup>, ...  
Manimozhiyan Arumugam,<sup>1,2,3</sup> Athanasios Tsiplakos<sup>1</sup>, ...  
Country-specific antibiotic use practices in the human gut resistome

Not a finite list, but let's tidy up

### **Applications that characterize genetic (mal)function**

- Gene expression / Transcriptomics (RNA-seq, CAGE-seq)
- Gene regulation / Epigenetics (DNA-Methylation, Histone modifications)
- Gene alterations (Hereditary diseases, cancer biology)

### **Applications that explore what is around us**

- Genome assemblies of other species (biodiversity)
- Metagenomics (environmental DNA)
- Pathogen surveillance (antibiotic resistance, epidemics)

### **Applications to elucidate evolutionary processes**

- Ancient genomes
- Population genomics

**ONE DOES NOT SIMPLY**



**SEQUENCE DNA, EVEN WITH NANOPORE**

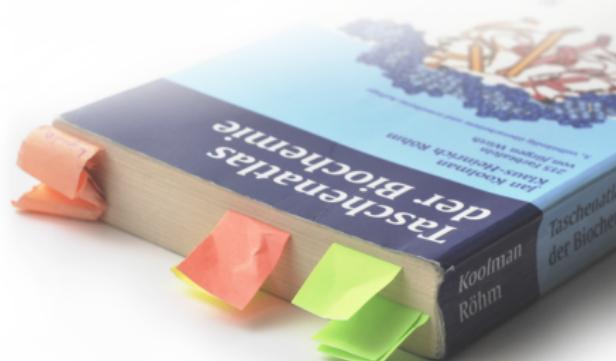
**Sequencing application:** Combination of a  
library preparation method and a suitable  
sequencing technology

## **Applications that characterize genetic (mal)function**

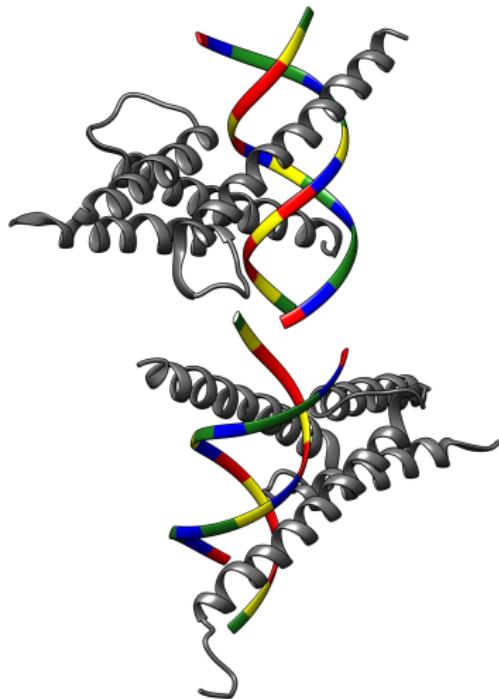
---

# Epigenetics: Regulatory layers of the genome

Cells of an organism contain the identical genome but utilize it in different ways.



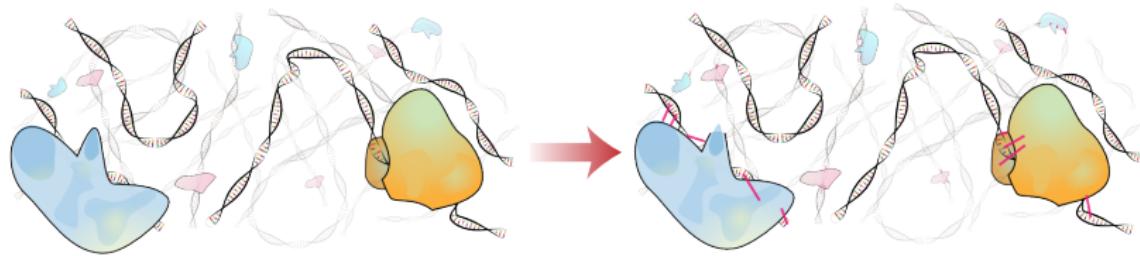
# ChIP-seq: Chromatin Immunoprecipitation Sequencing



- Which genomic sites are bound by a particular protein?
- Can detect transcription factor binding or epigenetic histone modifications.

← Two HLH-motifs (grey) of the helix-loop-helix-transcription factor MyoD are bound to DNA.

# ChIP-seq: Chromatin Immunoprecipitation Sequencing

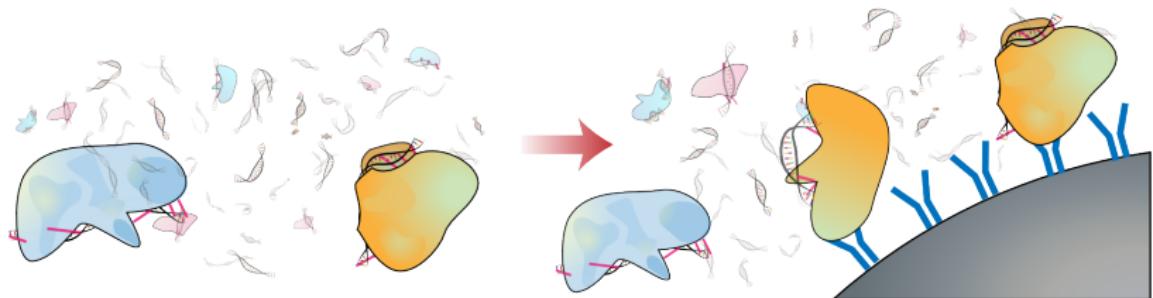


1. Isolation of target cells

2. *Cross-linking:* Stably bind proteins to DNA with covalent bonds

Fig: own derivative work. Original: [Jon Chui, Wikimedia Commons, CC-BY-SA 3.0](#)

# ChIP-seq: Chromatin Immunoprecipitation Sequencing

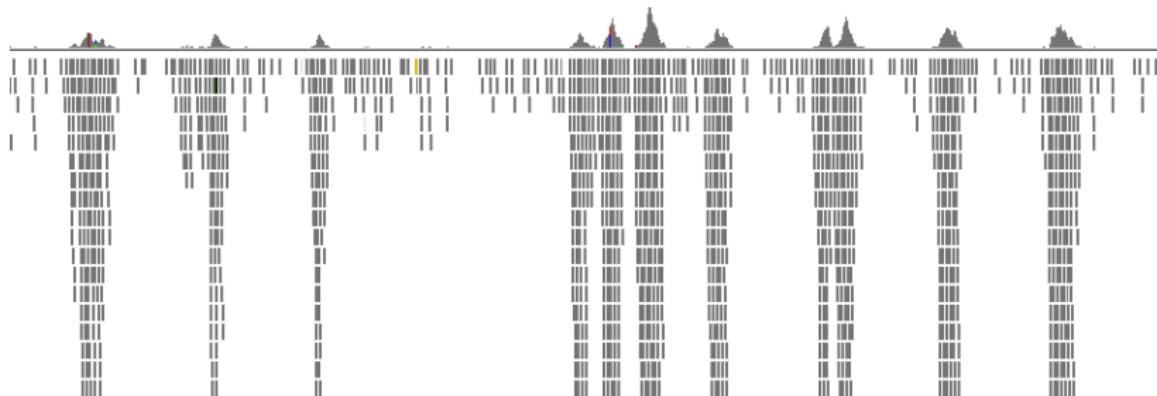


3. *Lysis + Sonication:* Lyse cells and fragment DNA by ultrasonic sound
4. *Precipitation:* Recover target protein and bound DNA from lysate

Fig: own derivative work. Original: [Jon Chui, Wikimedia Commons, CC-BY-SA 3.0](#)

# ChIP-seq: Chromatin Immunoprecipitation Sequencing

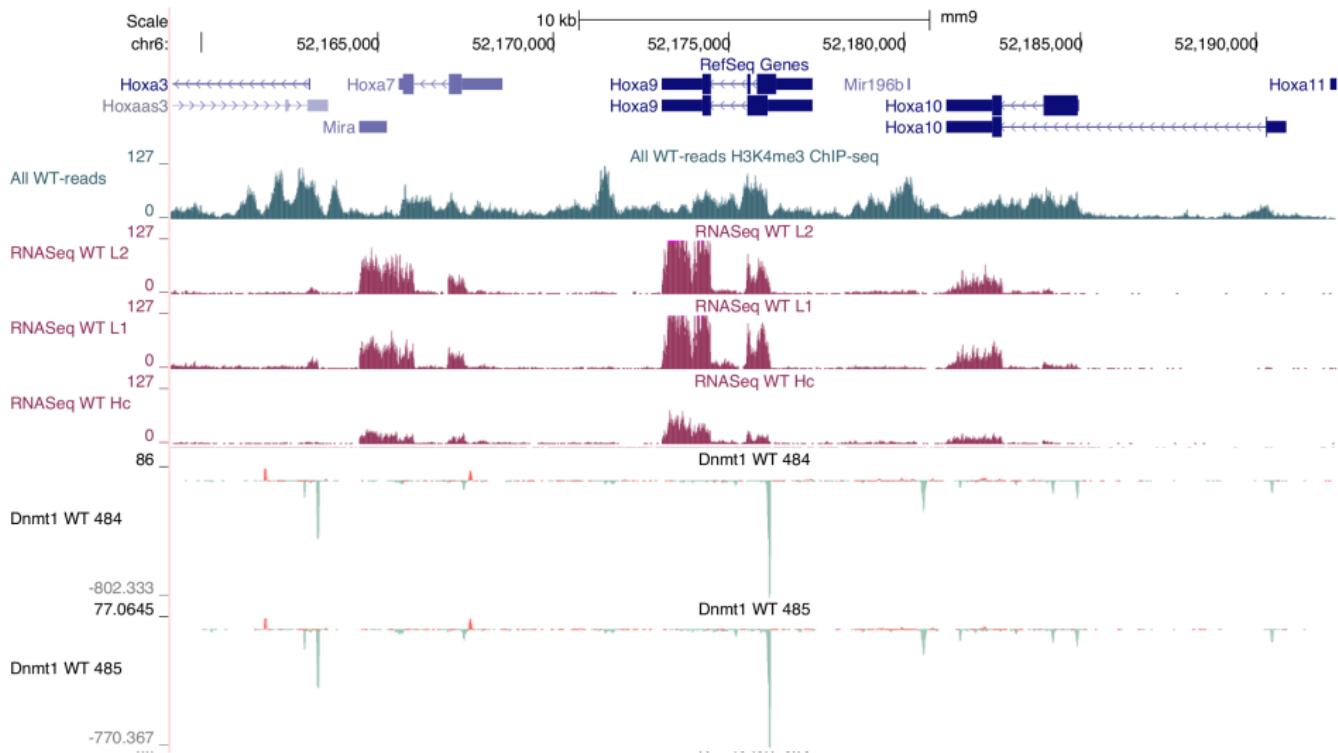
5. Decrosslink and recover formerly bound DNA
6. Sequence DNA
7. Alignment on the reference genome



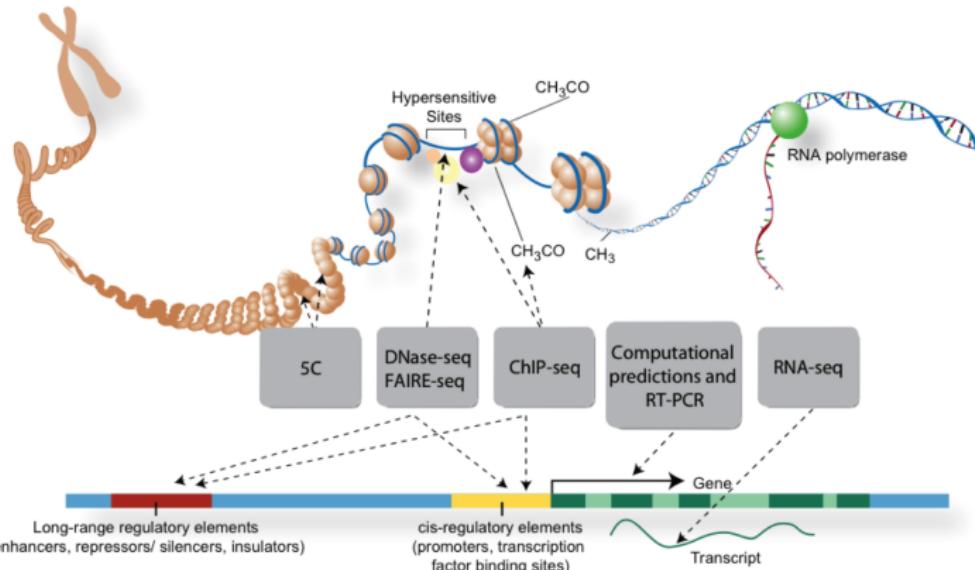
8. Subtract signals from negative control (IgG-control)
9. Peak detection and assignment to genes

Screenshot [Integrative Genomics Viewer \(IGV\)](#)

# Peaks of different applications: Histone ChIP-seq, RNA-seq, CAGE-seq



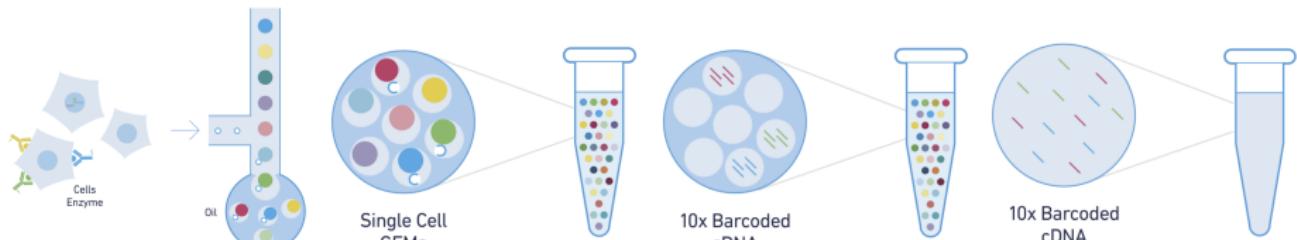
# Big sequencing projects addressed basic research on gene regulation



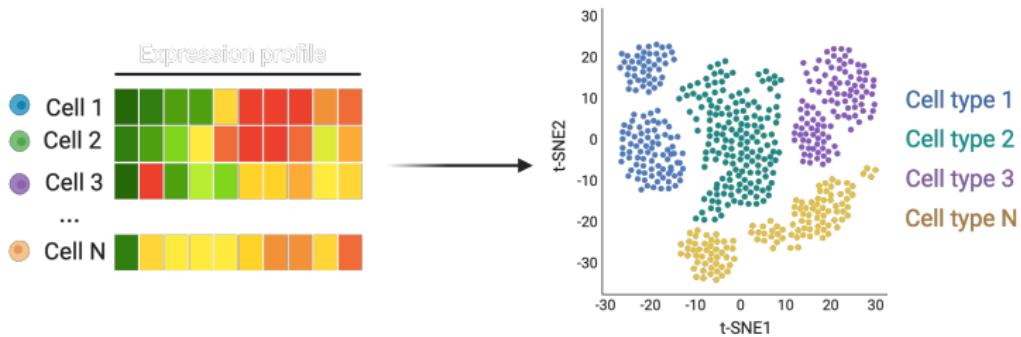
- ENCODE Project: ENCYclopedia Of DNA Elements
- Int. Human Epigenome Consortium

Fig.: Darryl Leja (NHGRI), Ian Dunham (EBI), <http://encodeproject.org/ENCODE/aboutScaleup.html>

# Most methods now on single-cell level



**Figure 5:** Artificial DNA barcode tags allow to determine cell of origin



**Figure 6:** Dimensionality reduction methods (UMAP, t-SNE) separate cell-types

# Spatial transcriptomics: Microscopy + single-cell sequencing

The screenshot shows the Human Cell Atlas website. At the top left is the logo, a blue circular icon with a grid pattern. To its right is the text "HUMAN CELL ATLAS". A horizontal navigation bar follows, containing links: Home, HCA, Areas of Impact, News, Publications, Data Coordination, EC H2020, Join HCA, and Contact.

**MISSION**

To create comprehensive reference maps of all human cells—the fundamental units of **Science**—for the benefit of human health and disease.

**RESEARCH ARTICLES**

**Three-dimensional intact-tissue sequencing of single-cell transcriptional states**

Xiao Wang<sup>1\*</sup>, William E. Allen<sup>1,2\*</sup>, Matthew A. Wright<sup>1,3</sup>, Emily L. Sylwestrak<sup>1</sup>, Nikolay Samusik<sup>4</sup>, Sam Vesuna<sup>1</sup>, Koen De Bruyn<sup>1</sup>, Charu Ramakrishnan<sup>1</sup>, Jia Liu<sup>1</sup>, Garry P. Nolan<sup>4†</sup>, Felice-Alessio Bava<sup>4†</sup>, Karl Deisseroth<sup>1,3,6†</sup>

<sup>1</sup>Department of Bioengineering, Stanford University, Stanford, CA 94301, USA

<https://www.humancellatlas.org>

Wang et al., Jul 2018, Science 361(6400)

SciLifeLab is one of the birthplaces of spatial transcriptomics

**10X GENOMICS**

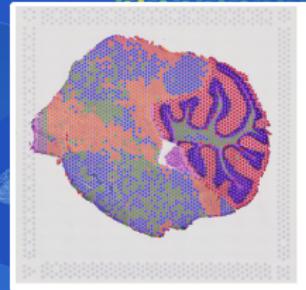
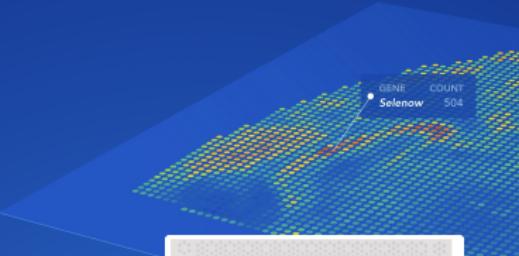
Products    Area of Interest    Resources    Support    Company    Careers   

 Visium Spatial Gene Expression

# Map the whole transcriptome within the tissue context

Visium Spatial Gene Expression is a next-generation molecular profiling solution for classifying tissue based on total mRNA. Map the whole transcriptome with morphological context in FFPE or fresh-frozen tissues to discover novel insights into normal development, disease pathology, and clinical translational research.

[Request Pricing](#)    [See How It Works ▶](#)



<https://www.spatialresearch.org>

# Cancer cell genomes (Recurrent mutations and aberrations)

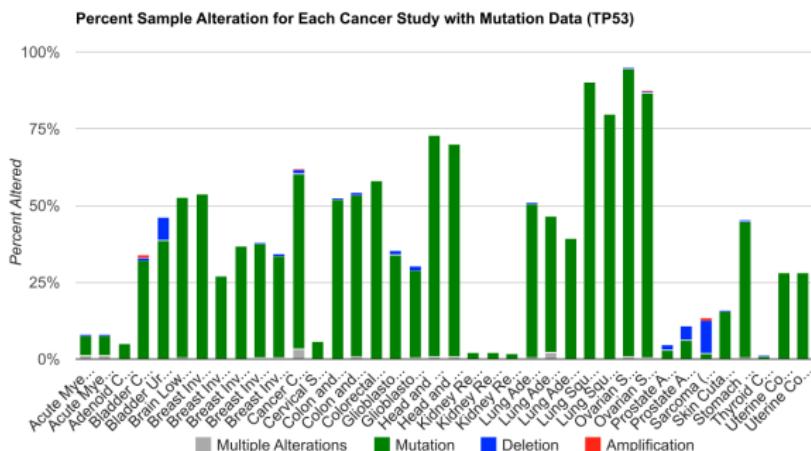
- Cancer Genome Project
- Cancer Genome Atlas
- PanCancerAtlas
- ...

**Cancer Cell**

Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients

André Kahles <sup>15</sup> • Kjøng-Van Lehmann <sup>15</sup> • Nora C. Toussaint • Matthias Hüser • ... Chris Sander • Gunnar Rätsch <sup>16</sup> • The Cancer Genome Atlas Research Network • Show all authors • Show footnotes

Published: August 02, 2018 • DOI: <https://doi.org/10.1016/j.ccr.2018.07.001>



<http://cancergenome.nih.gov>

<http://www.cbiportal.org>

## **Applications that explore what is around us**

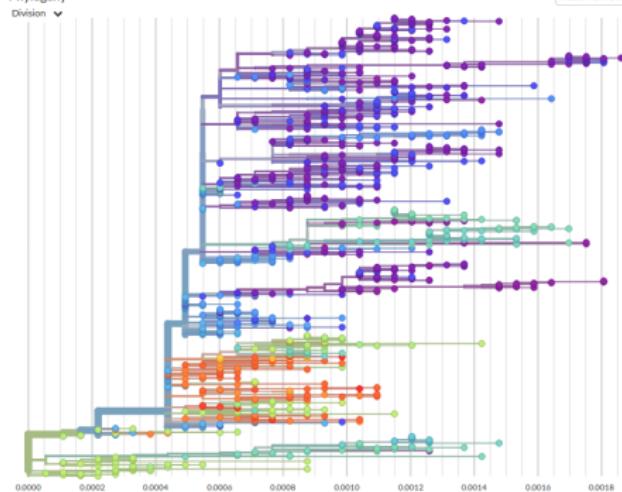
---

# Ebola epidemic 2016: Infectious disease monitoring

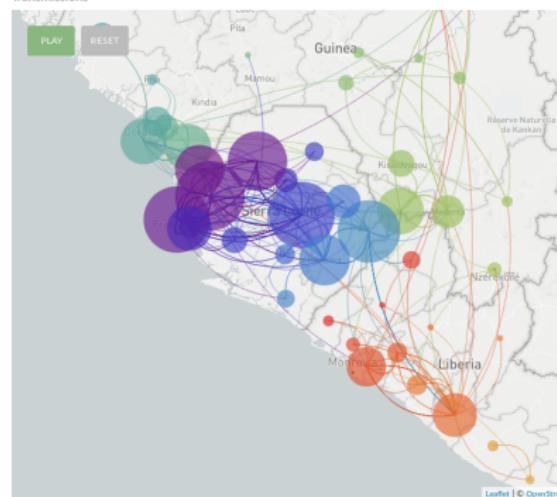
## Genomic epidemiology of the 2013-2016 West African Ebola epidemic

Showing 1238 of 1238 genomes sampled between Mar 2014 and Oct 2015 and comprising 19 authors, 3 countries and 46 divisions. Filtered to Dec 2013 to Oct 2015.

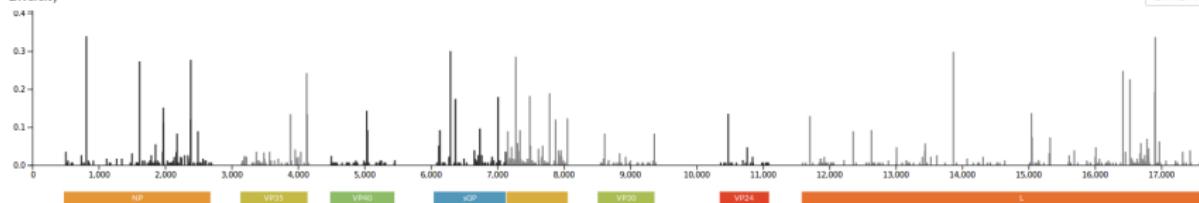
### Phylogeny



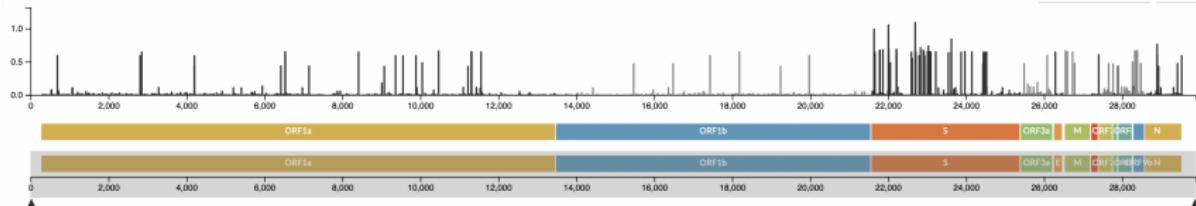
### Transmissions



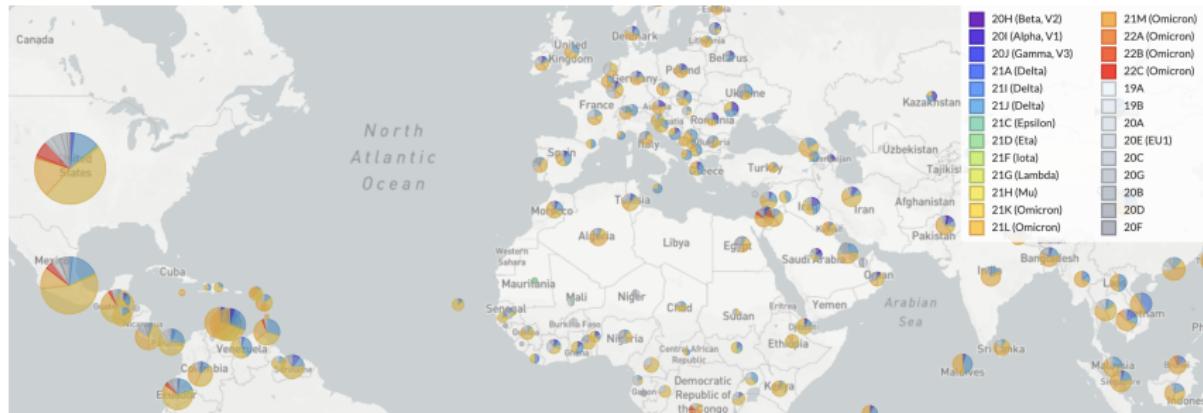
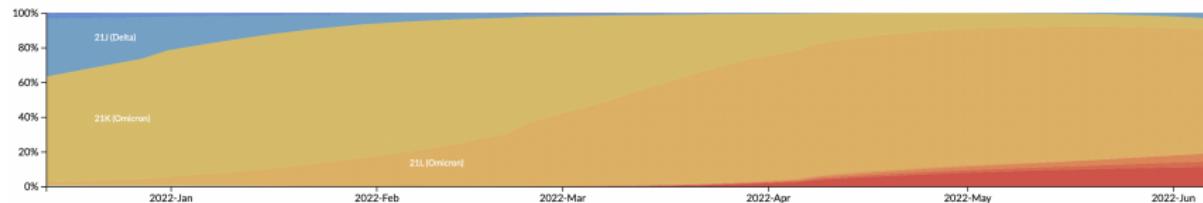
### Diversity



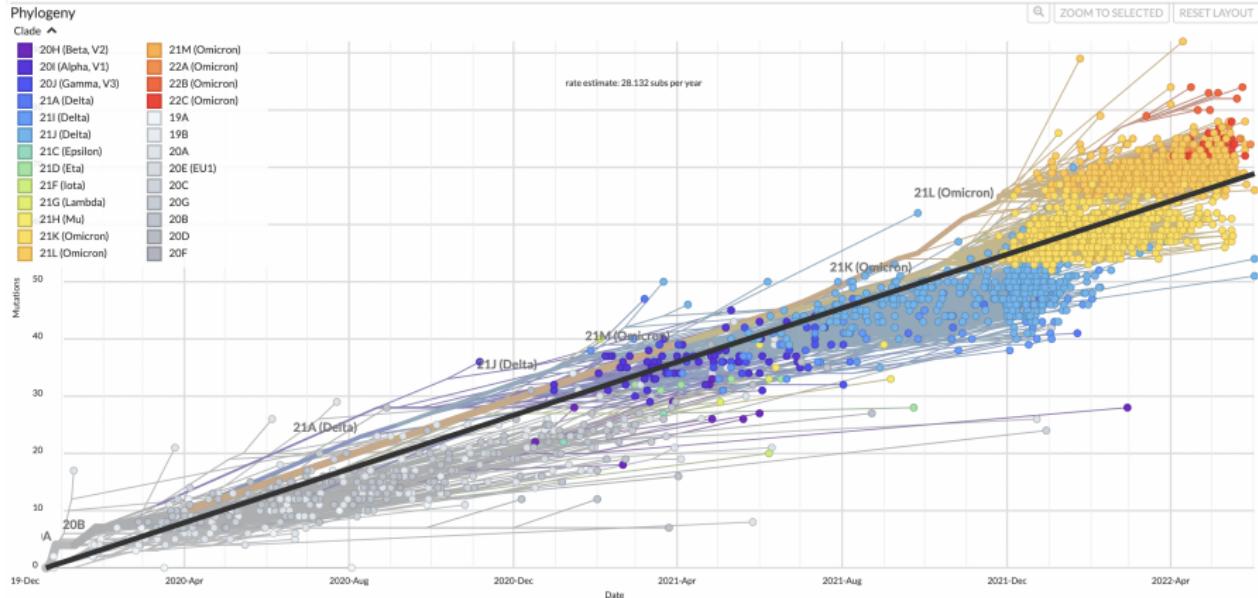
# SARS-CoV-2 pandemic: Global variant monitoring



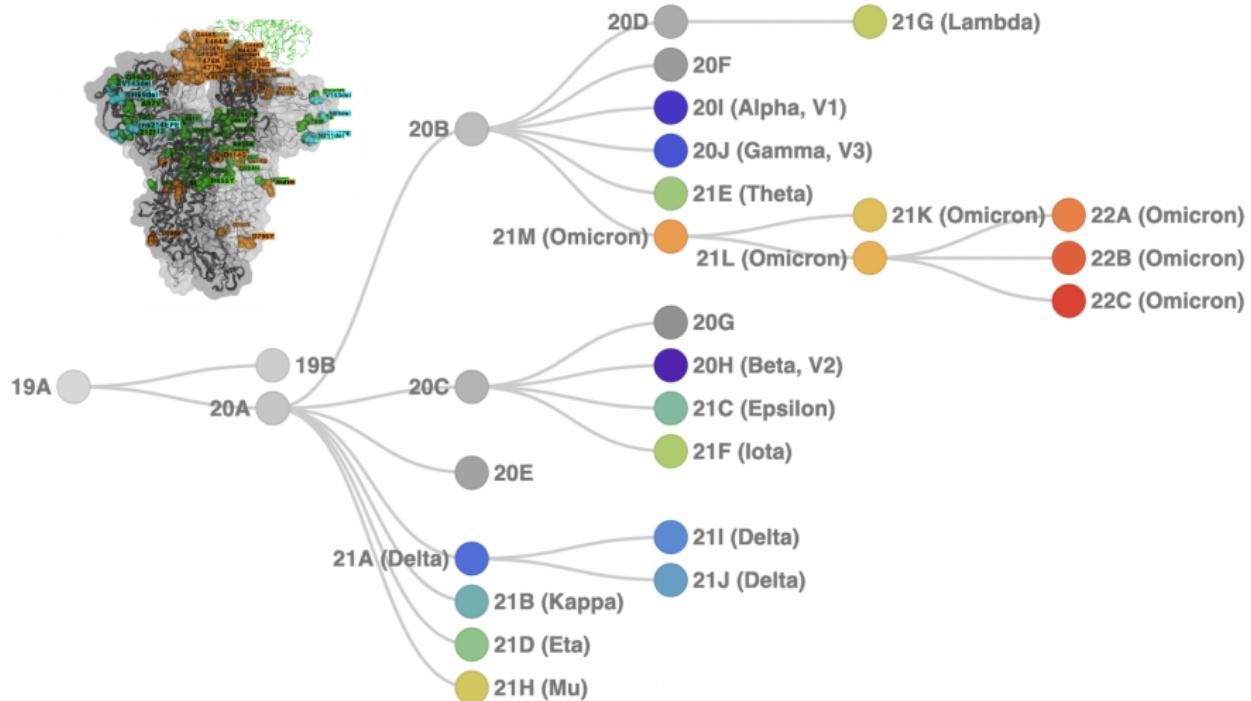
Frequencies (colored by Clade)



Tens of millions of genomes were deposited in GISAID database

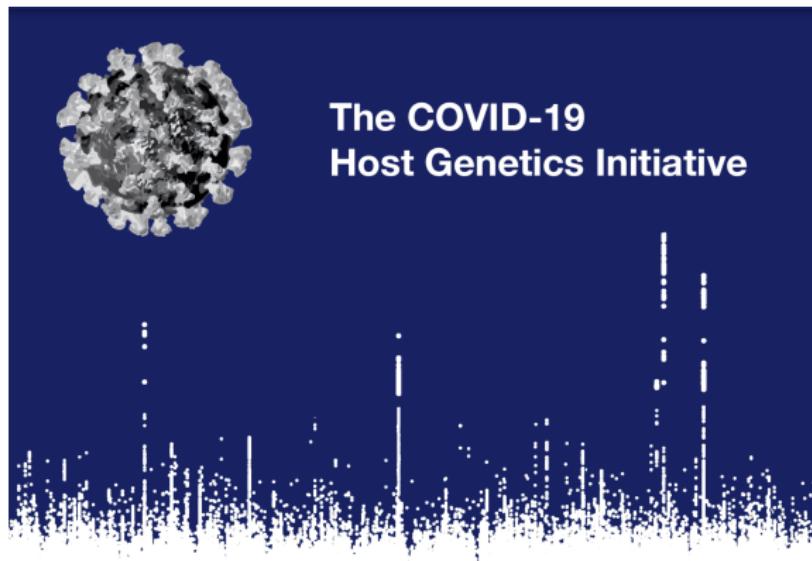


# SARS-CoV-2 pandemic: Risk-variant assessment



<https://nextstrain.org/sars-cov-2/>

# SARS-CoV-2 pandemic: Host genetics

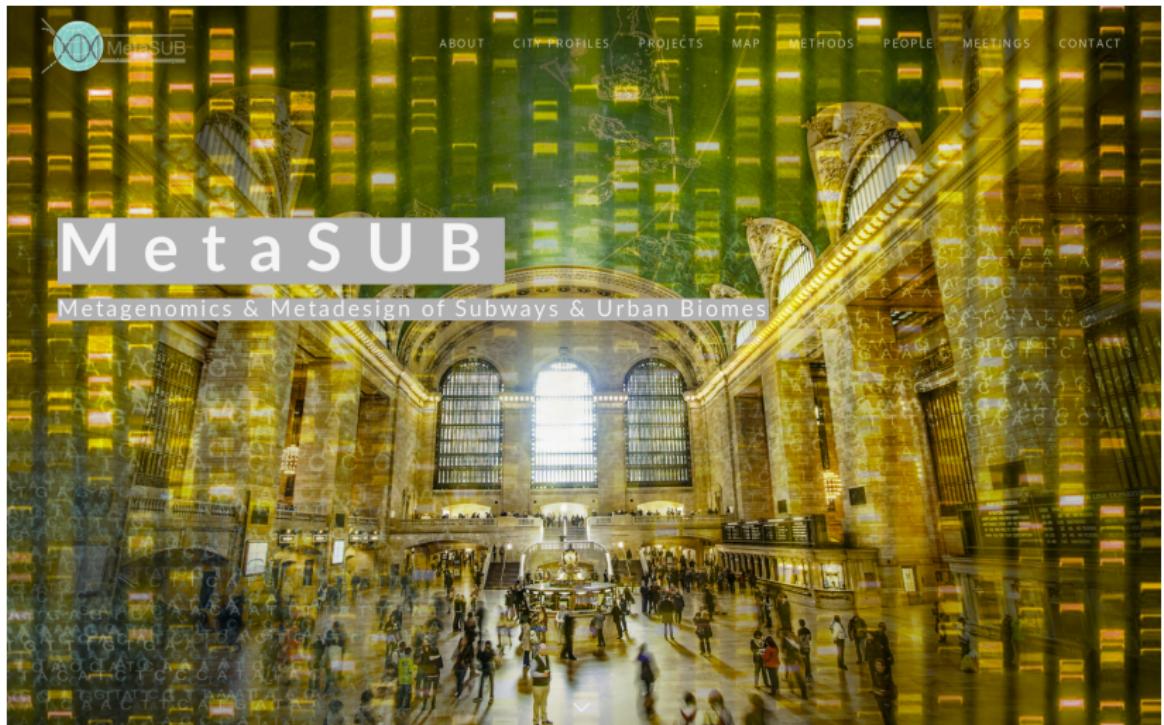


- Severity
- Susceptibility

<https://www.covid19hg.org>

- INF $\alpha$ R2
- OAS1
- DPP9
- FOXP4
- TYK2
- SFTPD
- MUC5B
- ACE2

# eDNA metagenomics: What is around us in the city?



<http://metasub.org>

# eDNA metagenomics: What is around us in the nature?

The screenshot shows a news article from tagesschau.de. At the top, there's a navigation bar with links like ARD Home, Nachrichten, Sport, Börse, Ratgeber, Wissen, Kultur, Kinder, Die ARD, Fernsehen, Radio, ARD Mediathek, and the ARD logo. Below the navigation is the tagesschau.de logo and a search bar. The main headline is "Untersuchung in Schottland DNA-Test bei Nessies Nachbarn". It includes a timestamp "Stand: 28.05.2018 09:27 Uhr" and social media sharing icons for Facebook, Twitter, Google+, and LinkedIn. A large image of a loch is visible in the background.

**Untersuchung in Schottland**  
**DNA-Test bei Nessies Nachbarn**  
Stand: 28.05.2018 09:27 Uhr

Gibt es eine Grundlage für die Untersuchung von Loch Ness? Die Forscher wollen DNA-Spuren aller Lebewesen im See sammeln. Ein internationales Forscherteam will die DNA aus dem Wasser des Loch Ness erfassen. Dafür müssen sie zuerst alle DNA-Spuren gesammelt werden. Geleitet wird das Team von Neil Gemmell. Er hofft sich die Forscher mit dem Ergebnis der Legende vom Ungeheuer zu beschäftigen.

**Neil Gemmell**  
@ProfGemmell

Folgen

Antwort an @jon\_waters\_nz

If I find unicorns, mermaids or Nessie will let you know.

17:19 - 25. Mai 2018 aus Dunedin City, New Zealand

The Guardian article headline is "Scientists to lead DNA hunt for Loch Ness monster". Below it is a sub-headline: "Samples of the Scottish loch will be tested to reveal truth behind legend". To the right of the text is a photograph of a misty, mountainous landscape, likely the area around Loch Ness.

## The Guardian

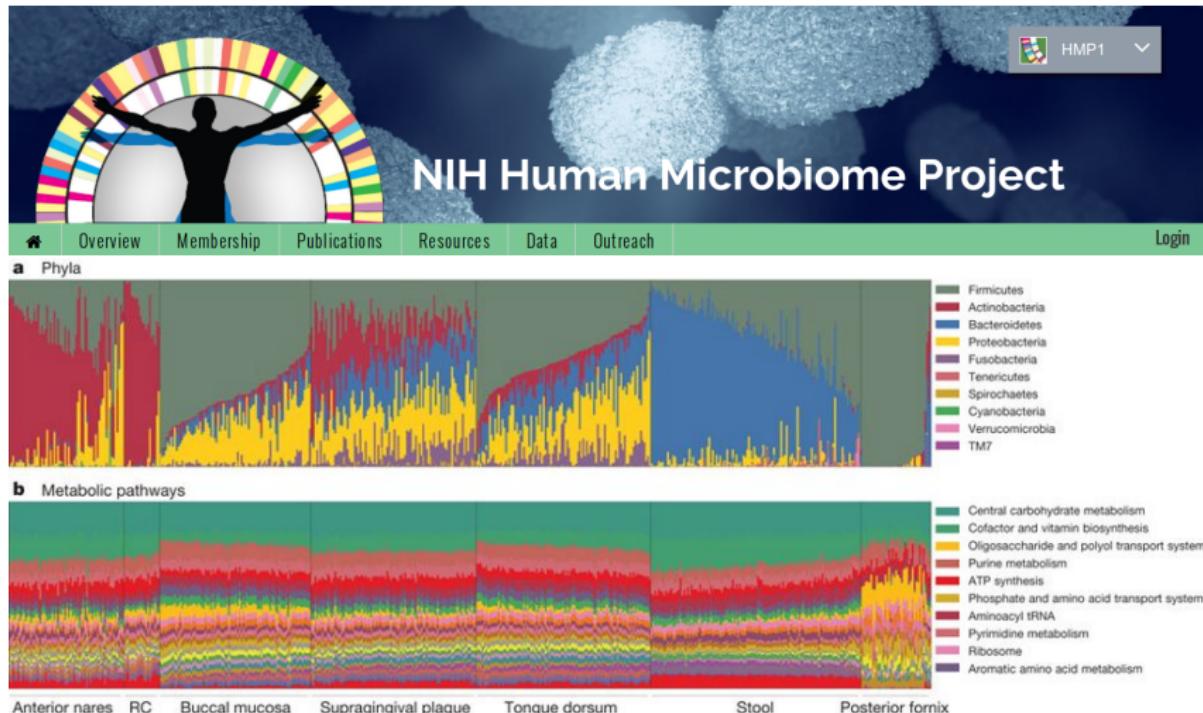
Scientists to lead DNA hunt for Loch Ness monster

Samples of the Scottish loch will be tested to reveal truth behind legend

A tweet from Neil Gemmell (@ProfGemmell) on Twitter. The tweet text is: "If I find unicorns, mermaids or Nessie will let you know." The timestamp is 17:19 - 25. Mai 2018, and the location is Dunedin City, New Zealand.

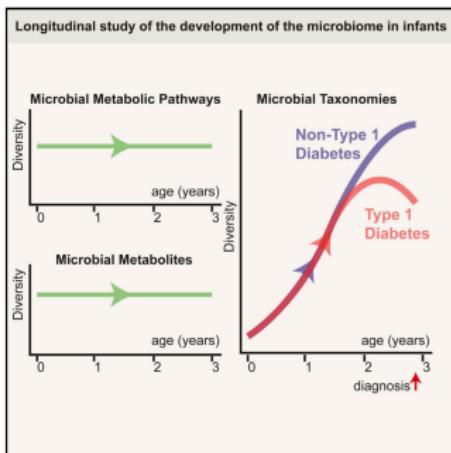
http://gemmell-lab.otago.ac.nz

# What is on and within us: Human Microbiome Project



<https://hmpdacc.org>

## Bacterial dysbiosis facilitates type 1 diabetes



Davis-Richardson et al. 2014. *Frontiers in Microbiology* 5, 678

# Our gut microbiome is associated with...

- Chemosensitivity
- Depression & mental health ("Gut-Brain-Axis")
- Multiple sclerosis, osteoarthritis & other auto-immune diseases

**Cell** Article

## *Fusobacterium nucleatum* Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy

TaChung Yu,<sup>1,3</sup> Fangfang Guo,<sup>1,3</sup> Yanan Yu,<sup>1</sup> Tian Tian Sun,<sup>1</sup> Dan Ma,<sup>1</sup> Jixuan Han,<sup>1</sup> Yun Qian,<sup>1</sup> Ilona Kryczek,<sup>2</sup> Dan Feng Sun,<sup>1,2</sup> Nisha Nagarkoth,<sup>2</sup> Yingxuan Chen,<sup>1,\*</sup> Haoyan Chen,<sup>1,\*</sup> Jie Hong,<sup>1,7</sup> Weiping Zou,<sup>2,4,\*</sup> and Jing-Yuan Fang<sup>1,\*</sup>

<sup>1</sup>State Key Laboratory for Oncogenes and Related Genes, Key Laboratory of Gastroenterology and Hepatology, Shanghai Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China  
<sup>2</sup>Department of Molecular Psychiatry, University of Heidelberg, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom  
<sup>3</sup>Department of Cell Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA  
<sup>4</sup>Department of Internal Medicine, Division of Hematology/Oncology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA  
<sup>5</sup>Department of Cell Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA  
<sup>6</sup>Department of Cell Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA  
<sup>7</sup>Department of Cell Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA  
<sup>8</sup>Department of Cell Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

Molecular Psychiatry (2016), 1–11  
© 2016 Macmillan Publishers Limited. All rights reserved 1359-4184/16

**ORIGINAL ARTICLE**

### Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism

P Zheng,<sup>1,2,3,8</sup> B Zeng,<sup>4,8</sup> C Zhou,<sup>1,2,3,8</sup> M Liu,<sup>1,2,3</sup> Z Fang,<sup>1,2,3</sup> X Xu,<sup>1,2,3</sup> L Zeng,<sup>1,2,3</sup> J Chen,<sup>1,2,3</sup> S Fan,<sup>1,2,3</sup> X Du,<sup>1,2,3</sup> X Zhang,<sup>1,2,3</sup> D Yang,<sup>5</sup> Y Yang,<sup>1,2,3</sup> H Meng,<sup>6</sup> W Li,<sup>1</sup> ND Melgire,<sup>1,2,3</sup> J Licinio,<sup>7,9</sup> H Wei,<sup>4,9</sup> and P Xie,<sup>1,2,3,9</sup>

DOI: 10.1038/s41467-018-05184-7 OPEN

### Gut microbiota associations with common diseases and prescription medications in a population-based cohort

Matthew A. Jackson,<sup>1,2</sup> Serena Verdi,<sup>1</sup> Maria-Emanuela Maxan,<sup>3</sup> Cheol Min Shin,<sup>1,4</sup> Jonas Zierer,<sup>1,5</sup> Ruth C.E. Bowyer,<sup>1</sup> Tinhaiou Martin,<sup>1,6</sup> Francisco J. Lopez,<sup>1,7</sup> Daniel R. Johnson,<sup>1,8</sup> and Michael J. Frazee,<sup>1,9</sup>

Yu et al. 2017, Cell 170, 548–563

Zheng et al., 2016, Mol Psychiatry. 21(6):786-96

Jackson et al. 2018 Nature Communications 9:2655

## **Applications to elucidate evolutionary processes**

---

# Prehistoric and ancient human genomes

## A Draft Sequence of the Neandertal Genome

Richard E. Green,<sup>1,\*†‡</sup> Johannes Krause,<sup>1,§</sup> Adrian W. Briggs,<sup>1,§</sup> Tomislav Maricic,<sup>1,§</sup> Udo Stenzel,<sup>1,§</sup> Martin Kircher,<sup>1,§</sup> Heng Li,<sup>2,¶</sup> Michael Hofreiter,<sup>1,||</sup>

## Ancient human genome sequence of an extinct Palaeo-Eskimo

Morten Rasmussen<sup>1,2,§</sup>, Yingrui Li<sup>2,3,§</sup>, Stinus Lindgreen<sup>1,4,¶</sup>, Jakob Skou Pedersen<sup>1</sup>, Anders Albrechtsen<sup>1</sup>, Ida Moltke<sup>4</sup>, Mait Metspalu<sup>5</sup>, Ene Metspalu<sup>5</sup>, Kivisild<sup>5,6</sup>, Ramneek Gupta<sup>7</sup>, Marcelo Bertalan<sup>7</sup>, Kasper Nielsen<sup>1,2,¶</sup>, M. Thomas P. Gilbert<sup>1,4,¶</sup>

<http://genome-euro.ucsc.edu/Neandertal/>

Green et al. 2010, Science Vol. 328 No. 5979, 710-722

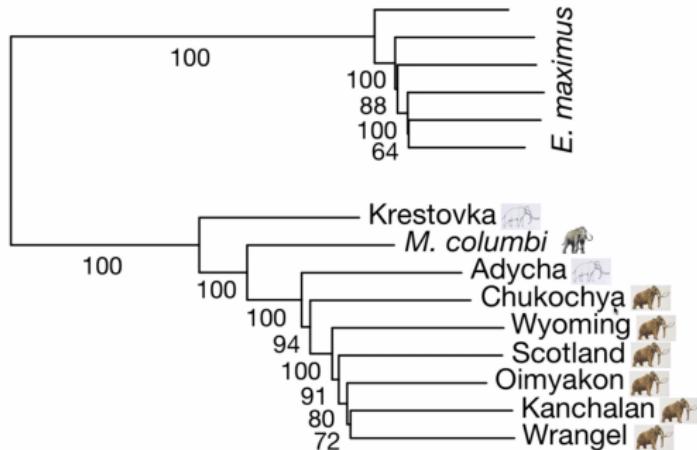
Rasmussen et al. 2010, Nature 463, 757-762



BEST BEFORE  
50.000 y/o

Store in cool, dry place

# Ancient genomes shed light on evolutionary developments



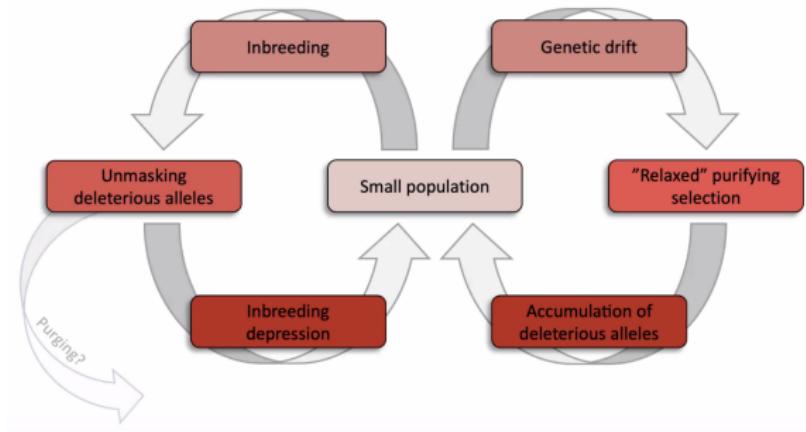
- Hair growth
- Thermal sensation
- Fat deposition
- Thermoregulation
- Circadian rhythm



Van der Falk et al., Nature volume 591, pages 265–269

(2021) 53

# Genomic processes in populations on the brink of extinction



Dussex et al., *Cell Genomics* Volume 1, Issue 1, 13 October 2021  
Liu et al., *Volume 184*, Issue 19, 16 September 2021, Pages 4874-4885.e16

# Genomic adaption in highly invasive species: Marbled crayfish

- Invasive species
  - Emerged 1995 in an aquarium, all females.
  - Parthenogenetic reproduction ( 700 viable eggs, every 3 month)
- Model organism for clonal evolution:
  - 3.5 Gbp (haploid), >21,000 genes
  - Triploid AA'B genotype: 276 chromosomes
  - Phenotypic variation solely by DNA-methylation.



**Figure 7:** Three-month-old marbled crayfish from the same clutch.

Gutekunst et al., Nature Ecology & Evolution volume 2, pages 567-573 (2018)

# Agrigenomics - analysis of cultured plants and farm animals

# Science



## ROAD MAP FOR **WHEAT**

Ordered sequence will  
speed research pp. 635, 661 & 662

### Genome sequence of the soybean

Jeremy Schmutz<sup>1,2</sup>, Steven B. Cai<sup>1</sup>,  
David L. Hyten<sup>1</sup>, Guo-Lin Song<sup>1</sup>,  
BRIAN AUGUST 2008  
AAAS

Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum

Emma S. Mace<sup>1,\*</sup>, Shuaishuai Tai<sup>2,\*</sup>, Edward K. Gilding<sup>3</sup>, Yanhong Li<sup>2</sup>, Peter J. Prentis<sup>4</sup>, Liapl...  
pp. 642 & 656

### The genome of *Theobroma cacao*

Xavier Argout<sup>1,24</sup>, Jerome Salse<sup>2,24</sup>, Jean-Marc Aury<sup>3,5,24</sup>, Mark J. Guillotin<sup>6</sup>,  
Mathilde Allegre<sup>3</sup>, Cristian Chambard<sup>9</sup>, Thierry Legavre<sup>1</sup>, Siela N. Maxine<sup>10</sup>,  
Olivier Fouet<sup>1</sup>, J. Véronique...  
pp. 643 & 657

### Genome sequence and analysis of the tuber crop potato

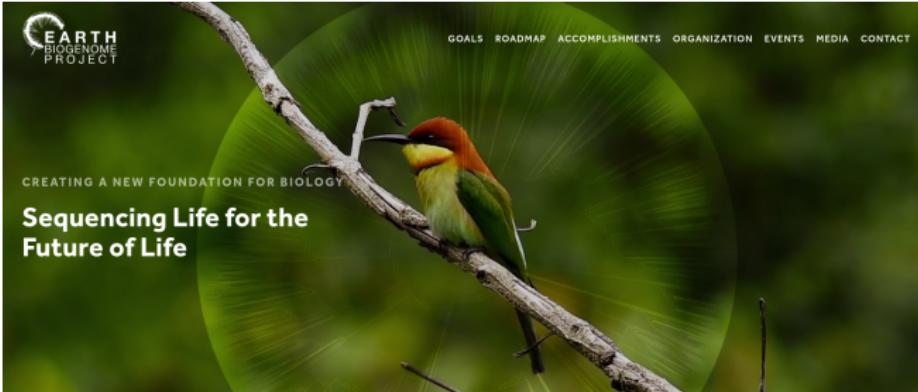
The Potato Genome Sequencing Consortium\*

### The B73 Maize Genome: Complexity, Diver...

Patrick S. Schnable<sup>1,3,6,7</sup>, Doreen Ware<sup>5,6,7</sup>, Robert S. Fulton<sup>1,2,8</sup>,  
Suzanne M. O'Connor<sup>1,2</sup>, Chennalai Liyanage<sup>1</sup>,  
pp. 658 & 663

<https://www.wheatgenome.org>

# Reference genomes of most vertebrate species until 2030



The screenshot shows the homepage of the Earth Biogenome Project. At the top left is the project logo with the text "EARTH BIOTRUST BIOPROJECT". At the top right are links for "GOALS", "ROADMAP", "ACCOMPLISHMENTS", "ORGANIZATION", "EVENTS", "MEDIA", and "CONTACT". Below the header is a large photograph of a chestnut-headed bee-eater perched on a branch. Overlaid on the image is the text "CREATING A NEW FOUNDATION FOR BIOLOGY" and "Sequencing Life for the Future of Life". A green circular graphic is overlaid on the background image. Below the image, the section title "What is the Earth Biogenome Project?" is displayed, followed by a detailed paragraph about the project's goals and the power of modern sequencing technology.

**What is the Earth Biogenome Project?**

Powerful advances in genome sequencing technology, informatics, automation, and artificial intelligence, have propelled humankind to the threshold of a new beginning in understanding, utilizing, and conserving biodiversity. For the first time in history, it is possible to efficiently sequence the genomes of all known species, and to use genomics to help discover the remaining 80 to 90 percent of species that are currently hidden from science.

**Q4/2023** : One annotated reference genome per taxonomic family of eukaryotes (ca. 9500 genomes).

**Q4/2027** : One annotated reference genome of each genus (ca. 180.000 genomes)

**Q4/2030** : Remaining 1.65 million genomes in the last phase.

## Summary

---

# Milestones in DNA sequencing history

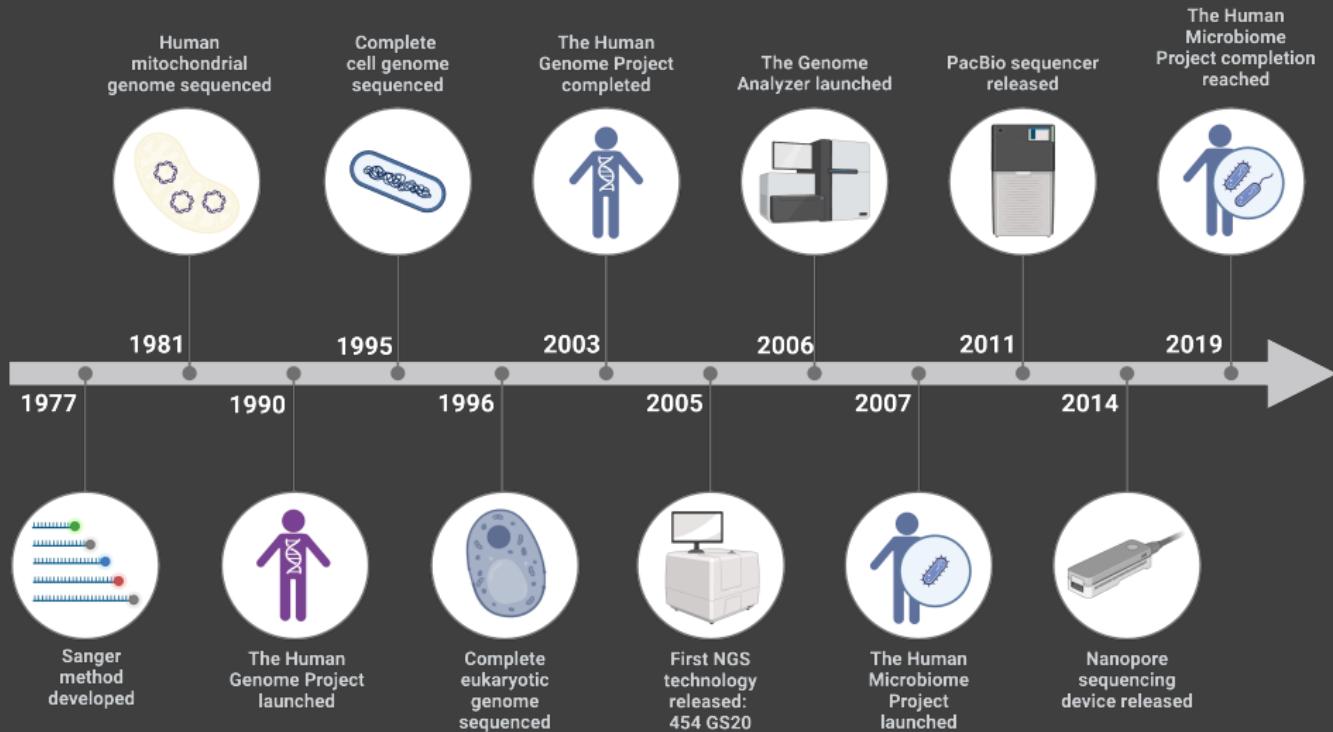


Figure by Anja Mezger

NGI has you covered for your research



<https://ngisweden.scilife-lab.se/resources/getting-started-at-ngi/>

## References

---

## References i

-  Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., ... Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome.. *Nature*, 409, 860–921. <https://doi.org/10.1038/35057062>
-  Sanger, F., Nicklen, S., & Coulson, A. R. (1977). Dna sequencing with chain-terminating inhibitors.. *Proceedings of the National Academy of Sciences of the United States of America*, 74, 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
-  Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., ... Zhu, X. (2001). The sequence of the human genome.. *Science (New York, N.Y.)*, 291, 1304–1351. <https://doi.org/10.1126/science.1058040>