



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

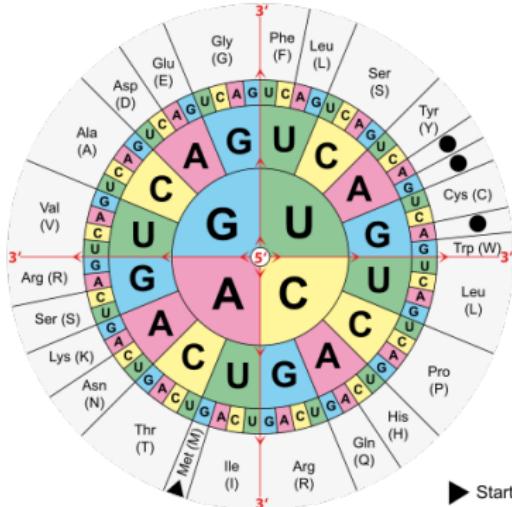
Prize share: 1/3



Photo from the Nobel Foundation archive

Marshall W. Nirenberg

Prize share: 1/3



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

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

<https://www.nobelprize.org/prizes/medicine/1968/summary> | User:Mouagip, Wikimedia Commons, PD

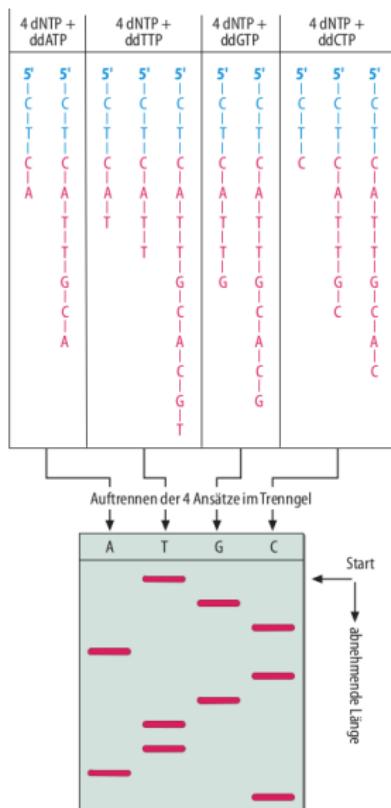
Encoded information of naturally occurring DNA unknown



- 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 ϕ 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

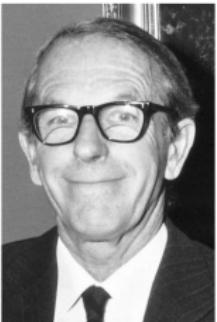


Photo from the Nobel Foundation archive.

Frederick Sanger

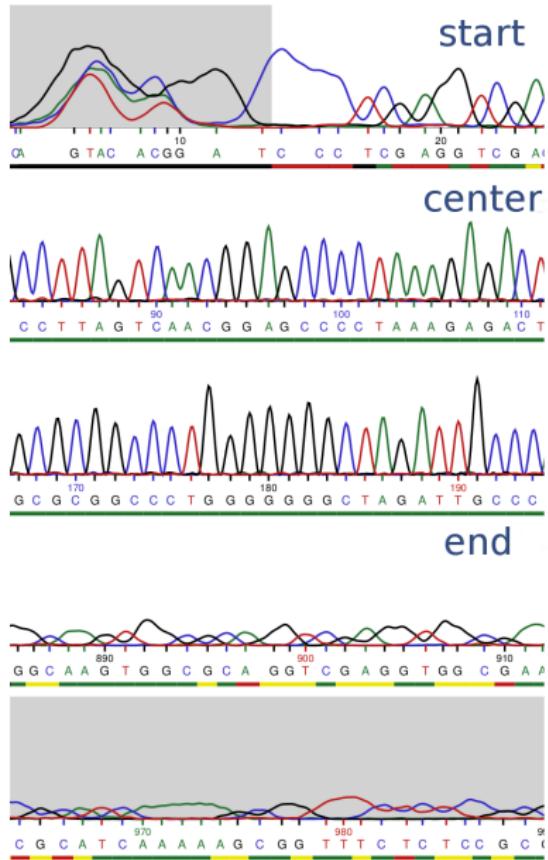
Prize share: 1/4

- 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

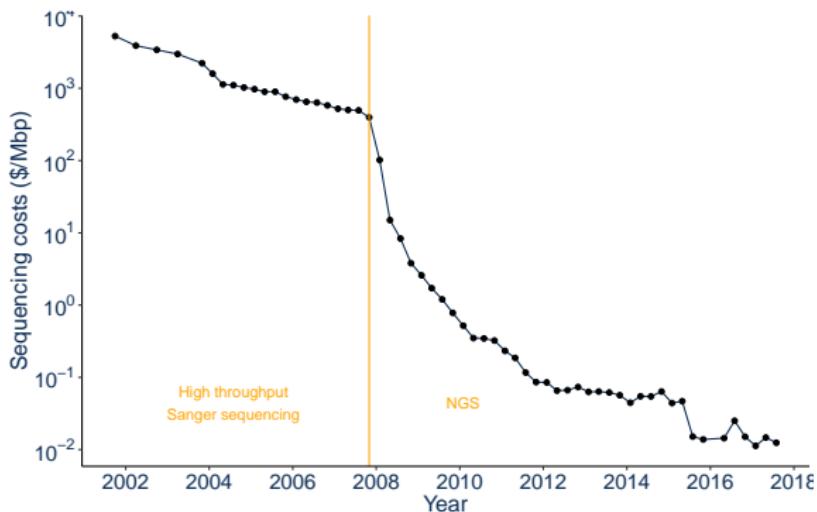


Figure 3: Sequencing costs per one million bases of raw sequence

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

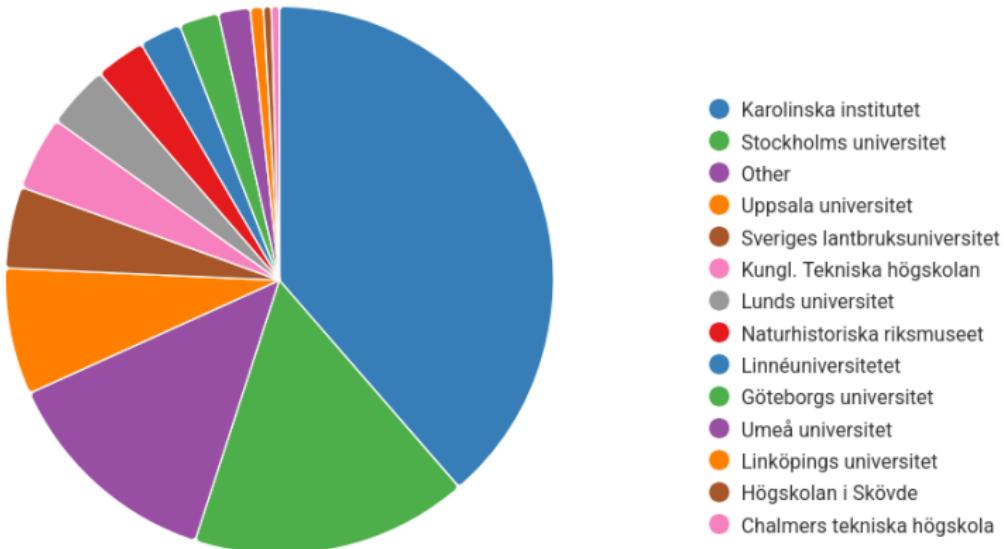
Sweden

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.

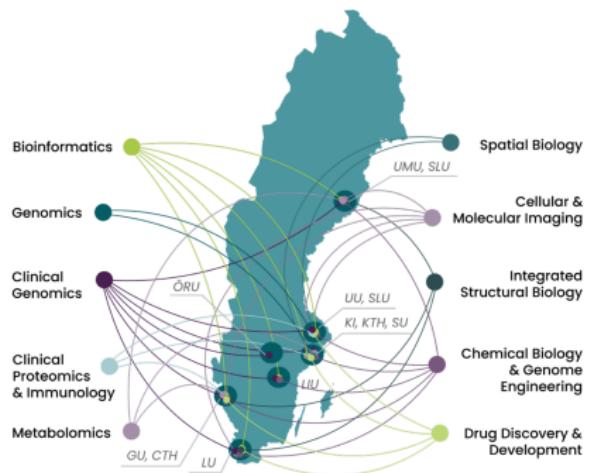
Project Affiliations in 2024



<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

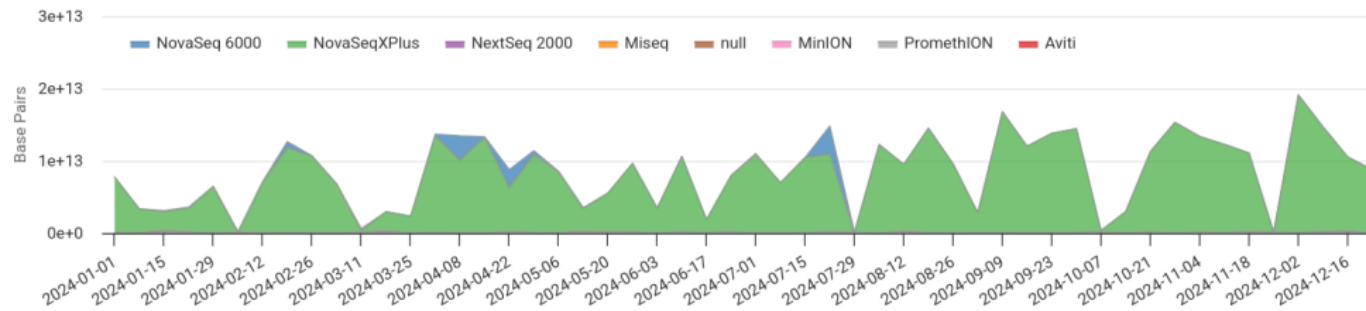


NGI-S employs various sequencing technologies



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

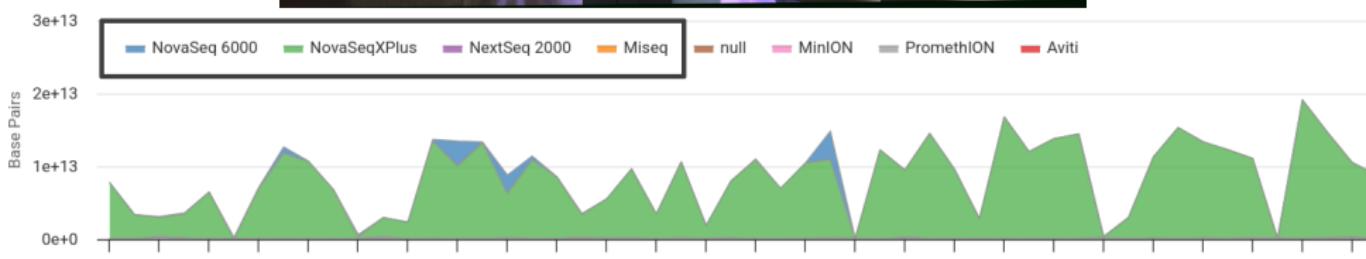
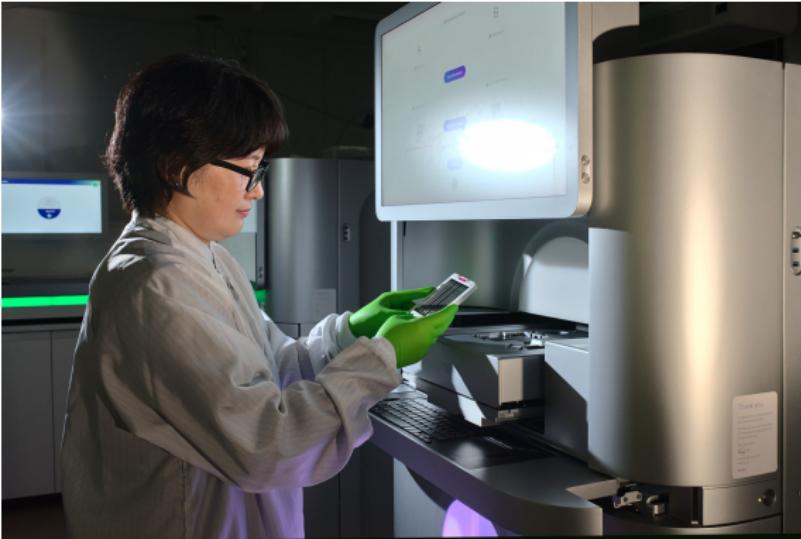
- Roche 454 sequencing (Pyrosequencing)
- Ion semiconductor sequencing
- **Illumina (Solexa) sequencing**
- **PacBio HiFi Sequencing**
- **Element Biosciences Avidite Sequencing**
- Ultima Genomics UG 100 Sequencing
- MGI DNBSEQ Technology
- Singular Genomics G4X

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)

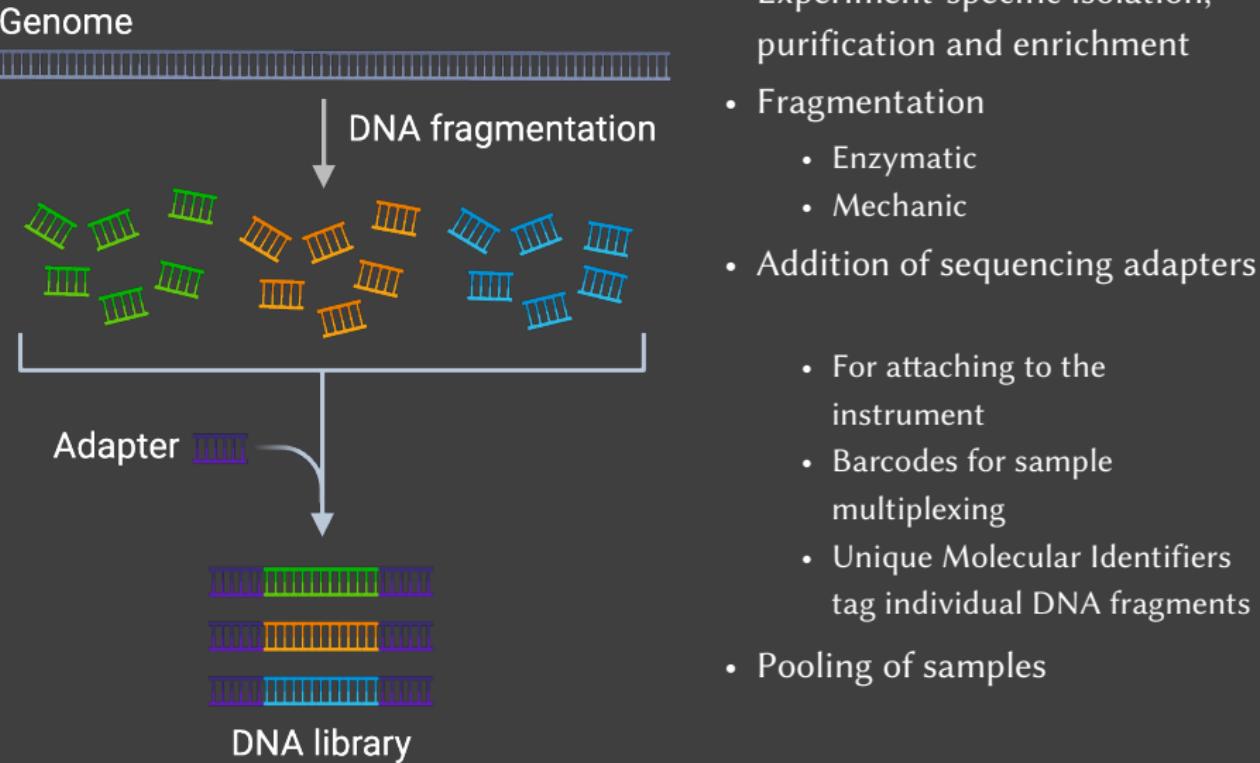
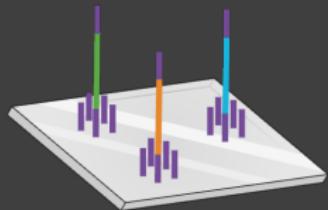


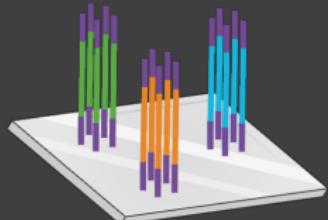
Figure by Anja Mezger

Preparation for sequencing (on the machine)

Library hybridization



Amplified clusters



Bridge
amplification
cycles

- Hybridize to a *flow cell*
- Amplify for better signal
- Linearize

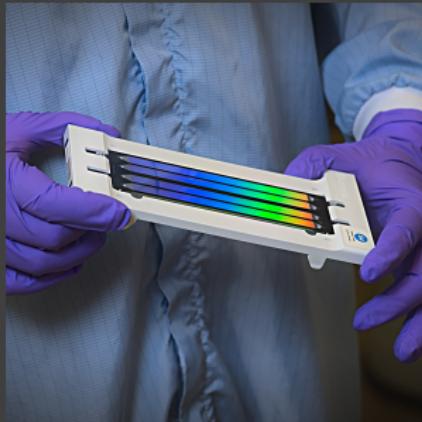
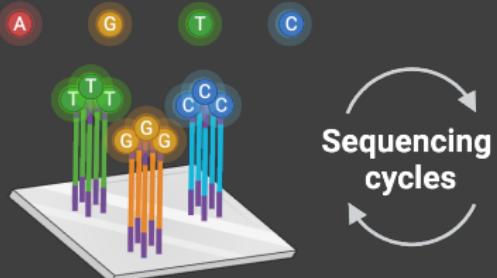


Figure by Anja Mezger

Illumina: *Sequencing by Synthesis* of DNA clusters

Fluorescently labeled nucleotides



Data collection



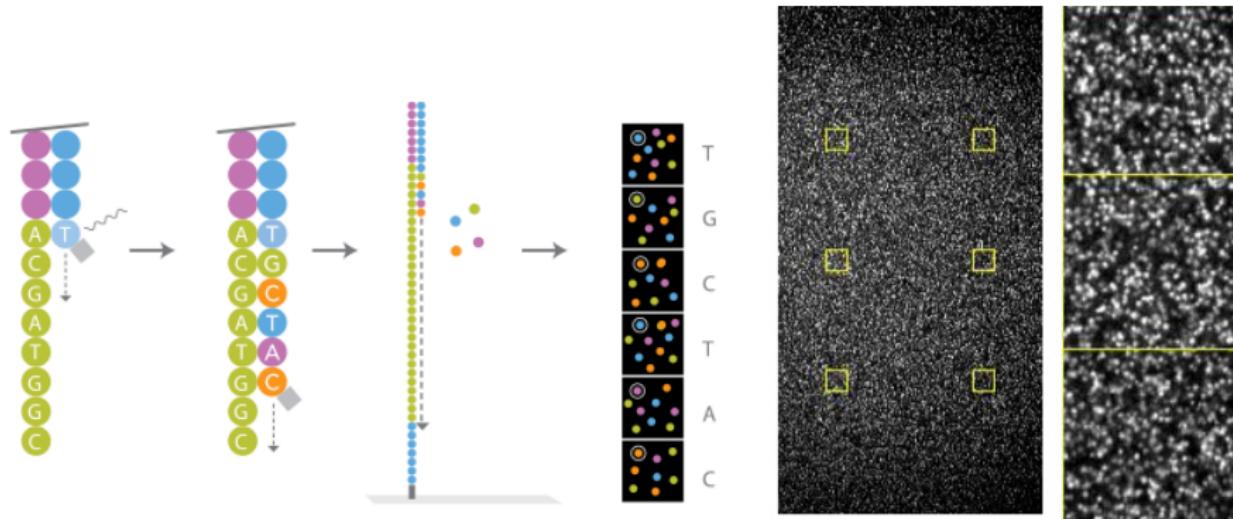
- DNA is amplified (again)
- Base integration yields a light signal (details vary among Illumina machines)
- Sequence is derived from a time-series of images
- NGS was developed from Sanger sequencing, but several key innovations were required.

[Rodriguez and Krishnan, 2023]

doi: 10.1038/s41587-023-01986-3

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

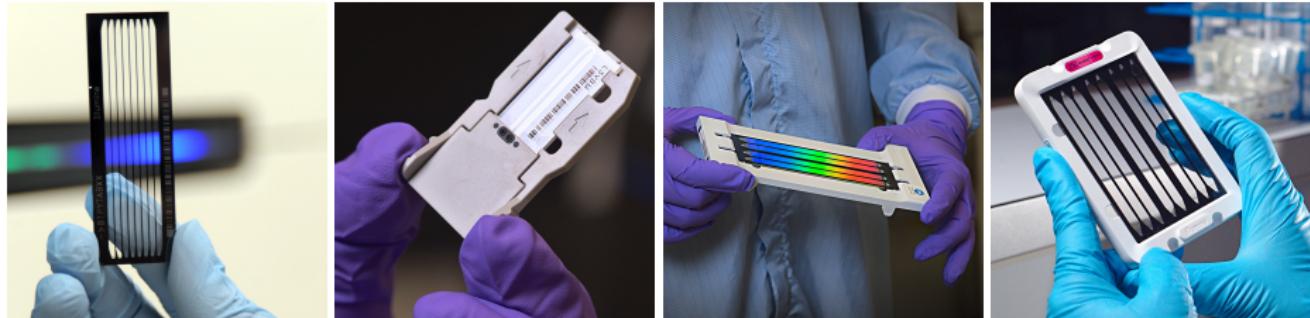
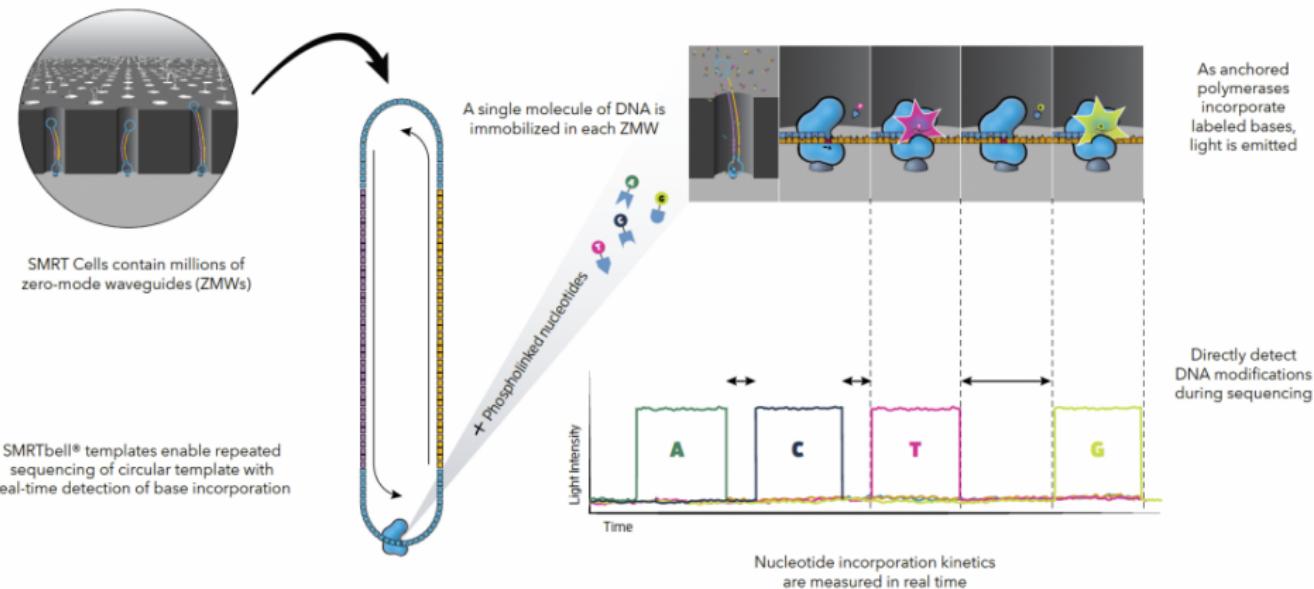


Figure 4: Various Illumina flow cells

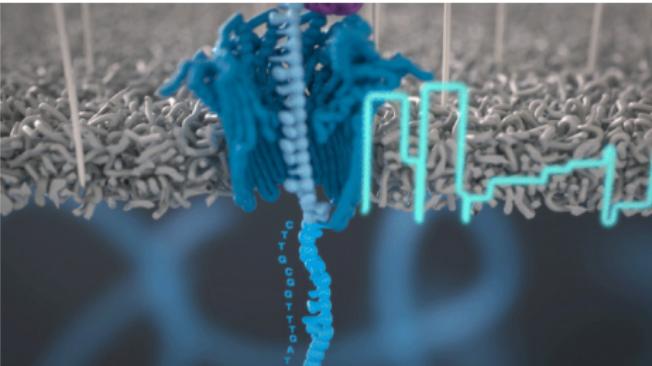
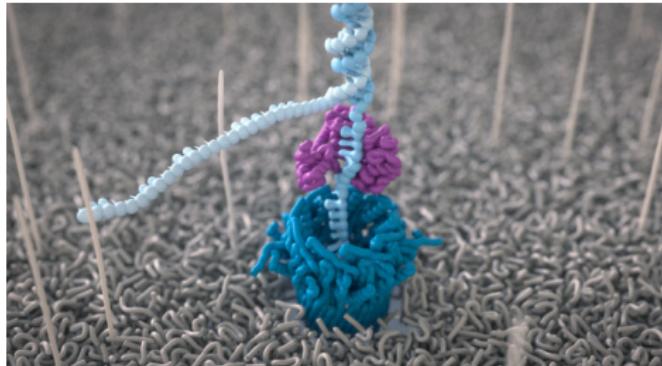
- Illumina's platform produces 2x 150bp reads from a fragment.
→ **short-read sequencing**
- Instead of 6912 fragments like with Sanger, Illumina machines can sequence Millions to Billions in parallel → **massive-parallel sequencing**

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

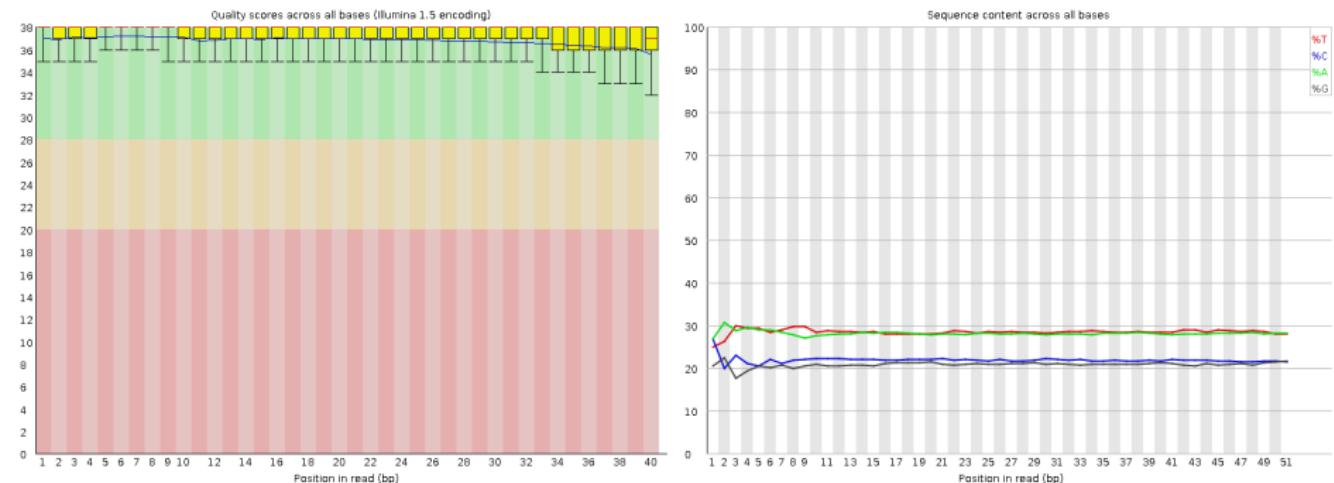
Sequencing result: Terabytes of data in FastQ-format

A single read:

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

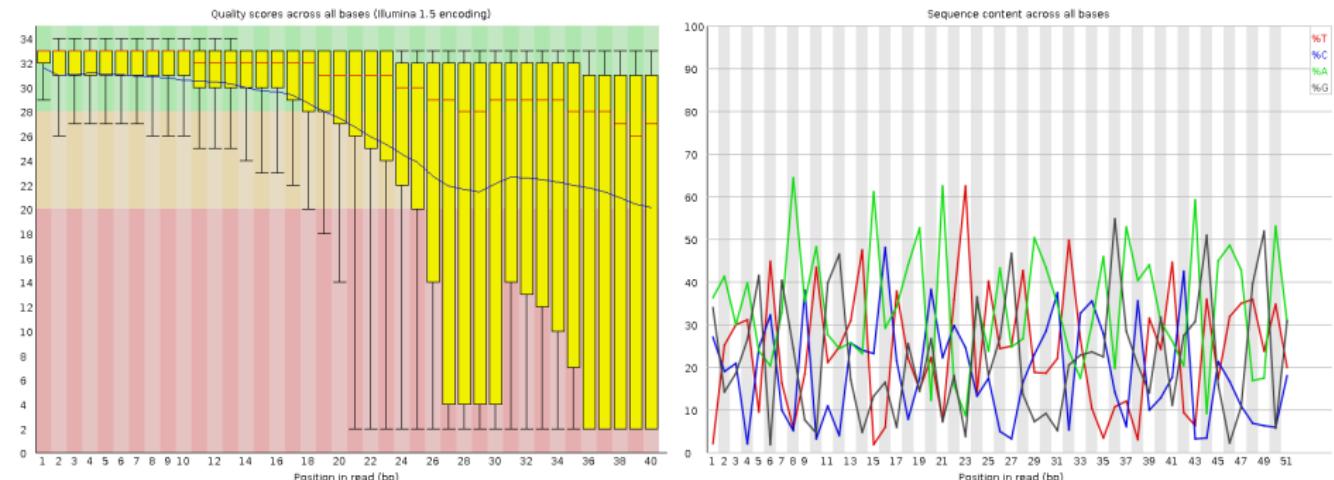
```
@SCILIFELAB:500:NGISTLM:1:1101:31620:1016 1:N:0
ATAAACACGGTCTTTCCAGGTCAAGCCGGACGGTACCGCCCTGTGGCCATCGAA
+
-86<<F9FB7FFFGGGGGGFACGFFDEFGGGEFE>FGGGDGFGGFGGCFG?FF7E@
```

Quality control: Good data



FastQC - A quality control tool for high throughput sequence data

Quality control: Poor data



FastQC - A quality control tool for high throughput sequence data

Common bioinformatic analyses

```
>NC_001422.1 Escherichia phage phiX174
GAGTTTATCGCTTCATGACGAGAATTAACTCTGGATATTCGTGAGTCGAAAAATT/
GATAAAAGCAGGAATTAACTACTGCTTGTTACGAATTAAATCGAAGTGGACTCTGGCGAAAATG/
ATTCGACCTATCCTTGGCAGCTCGAGAAGCTTACTTGCACCTTCGCCATCAACTAACGAT
TCAAAAAACTGACCGTGTGGATGAGGAGAAGTGGCTTAATATGCTTGGCACGTCGTCAGGACTG
GATATGAGTCACATTTGTCATGGTAGAGATTCTTGTGACATTAAAAGAGCTGGATTAA
TGAGTCGATGCTGTTCAACCACTAATAGGTAAGAAATCATGAGTCAGTTACTGAAACATCGT
TCCAGACCCTTGGCTCTATTAGTCATTCAGGCTCTGCCGTTTGGATTAACCGAGATG
CGATTTCCTGACGAGTAACAAAGTTGGATTGACTCTGACCGCTCTGCTGCTGCTGGTTG
TGGCTTATGGTACGCTGGACTTGGACCTCCGCTTCCGCTCTGCTGCTGCTGGTTG
TCATGGCTTATTAGTCATCCCGTCAACCTAACAGCGCTGTGTCATGGAAAGCTGTAAT
GGAAAAACATTATAATGGCTCGAGCTGGCTTAAAGCCGTAATTGTCGCTTACCTGGC
CGCGCAGGAAACACTGACGTTTACTGACGCGAGAAGAAAACGTCGCTCAAATTAACGTCGGA
TGATGTAATGCTCAAAGTAAAAACGTTCTGGCCTCGCCCTGGTGTCCGAGCCGTTGCGAG
AAAGCCAGCGTAAAGGGCTCGCTTGGTATGAGGTGTCACAAATTAAATGCAAGGGCTT
CCCTTACTTGAGGATAAATTATGTCATAATTCAAACACTGGCCGAGCGTATGCCGATGACCTT
TCTGGCTTCTGCTGGTCAAGATTGGTCGCTTATTACCATTCACACTCCGGTTACGTTG
TCCTGGAGATGGACGCCGTTGGCGCTCTCGCTTCTCATTGGCTGCGCTTGTGAA
CTGTAGACATTAACTTTTATGTCCTCATCGTCACGTTATGGTAACAGTGGATTAAAGTTG
GGATGGTGTAAATGCCACTCTCCCGACTGTTAACACTGTTATGGCATGCCGTT
GGCACGATTAACCCGTACCAATAAAATCCCTAACGATTTGGTTCAGGGTTATTTGAATATCTA
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

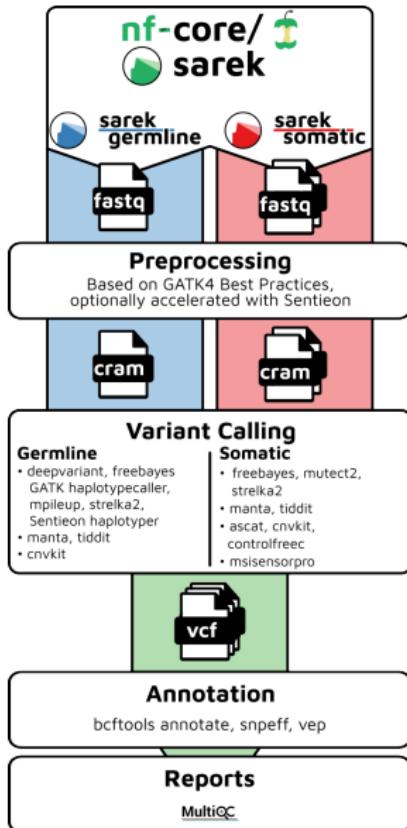
answer

←→

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^{1*}, William E. Alles^{1,2*}, Matthew A. Wright^{1,3}, Emily L. Sybrettak¹, Nikolay Samusik¹, Sam Venugra¹, Li Chen¹, Charu Ramakrishnan¹, Jia Liu¹, Gary P. Nolan^{1*}, Felice-Alessio Rava^{4,5}, Karl Deisseroth^{1,2,3,4}

¹Department of Bioengineering, Stanford University, Stanford, CA 94301



LETTER

doi:10.1038/nature10554

Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination

Antrollos¹, Zakeya Al Rasbi¹, Marina Boziki¹, Caroline Johner¹, Hartmut Wekerle¹

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

Yong-hui Jiang,^{1,18} Ryan K.C. Yuen,^{2,18} Xin Jin,^{3,4,5,18} Mingbang Wang,^{3,18} Nong Chen,³ Xueli Wu,³ ...
Jianguo Zhou,¹ ...
Liu, et al. *Nature* 505, 525–529 (2014); doi:10.1038/nature12879

ARTICLE

A map of human genome variation from population-scale sequencing

The 1000 Genomes Project Consortium*

The 1000 Genomes Project aims to provide a deep characterization of

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

Emma S. Maizel¹, Shashank Tal², Edward K. Gildin³, Yanheng Li², ...
Richard B. Pringault¹, et al. *Nature* 505, 530–534 (2014); doi:10.1038/nature12880

Ancient human genome sequence of an extinct Palaeo-Eskimo

Morten Rasmussen^{1,2*}, Yonina Li^{1,3*}, Stine Lindgreen^{1,4*}, Idaah Skov Pedersen⁴, Anders Albrechtsen⁴,

Emmanuelle Le Chatelier^{1*}, Trine Nielsen^{2*}, Junjie Qin^{3*}, Edi Prifti^{4*}, Falk Hildebrand^{4,5}, Gwen Falony^{4,5}, Mathis Manimozhiyan Arumugam^{3,6}, Jean-Michel

Manimozhiyan Arumugam^{3,6}, Jean-Michel

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



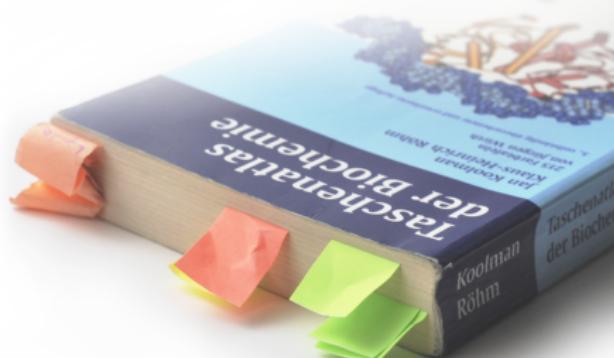
SEQUENCE DNA, EVEN WITH NANOPORE

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

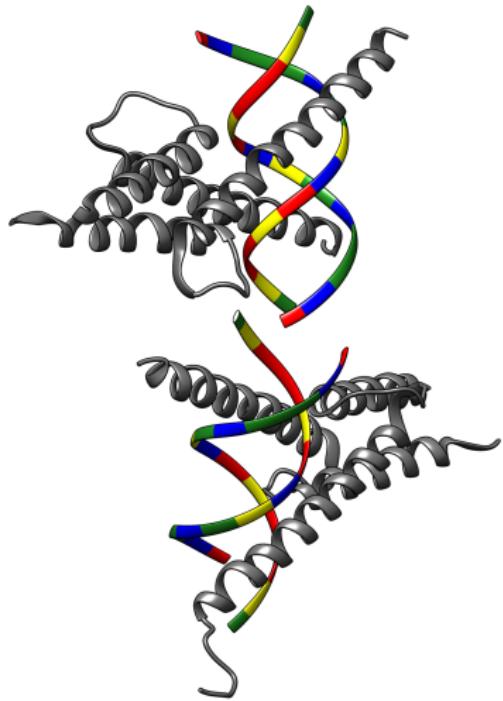
Applications that characterize genetic (mal)function

Epigenetics: Regulatory layers of the genome

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



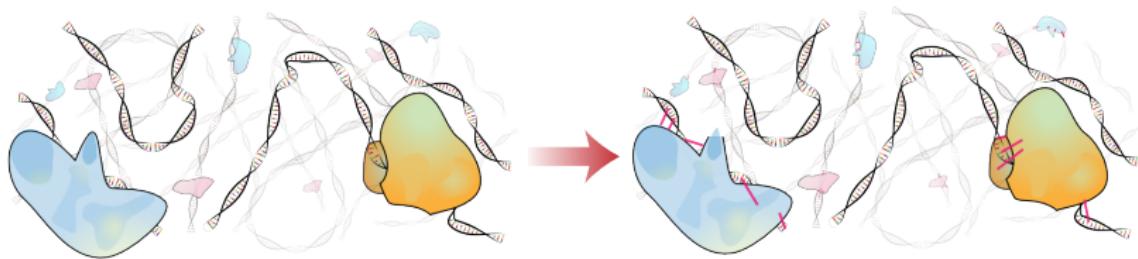
ChIP-seq: Chromatin Immunoprecipitation Sequencing



- 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

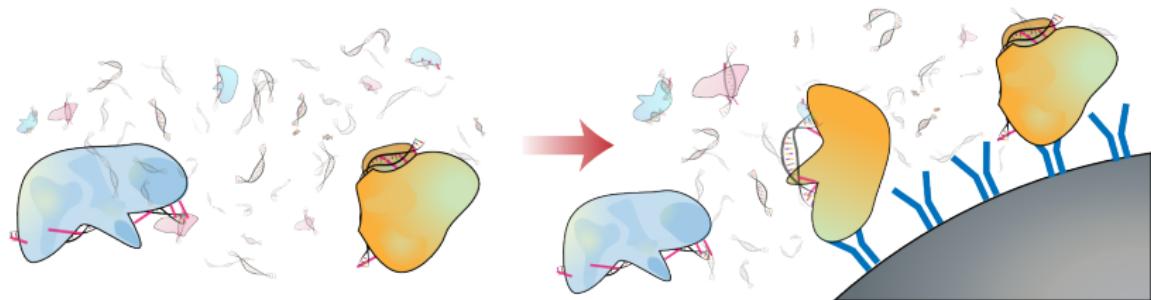


1. Isolation of target cells

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

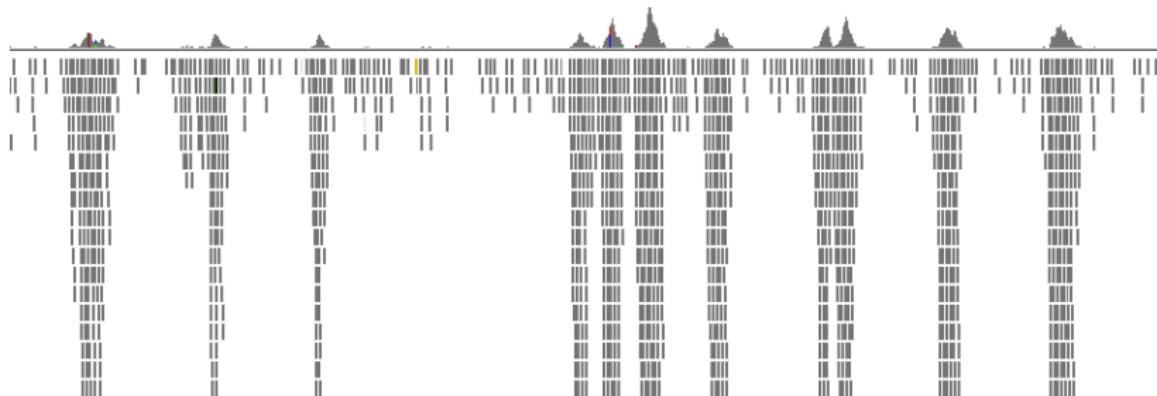


3. *Lysis + Sonication:* Lyse cells and fragment DNA by ultrasonic sound
4. *Precipitation:* Recover target protein and bound DNA from lysate

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

ChIP-seq: Chromatin Immunoprecipitation Sequencing

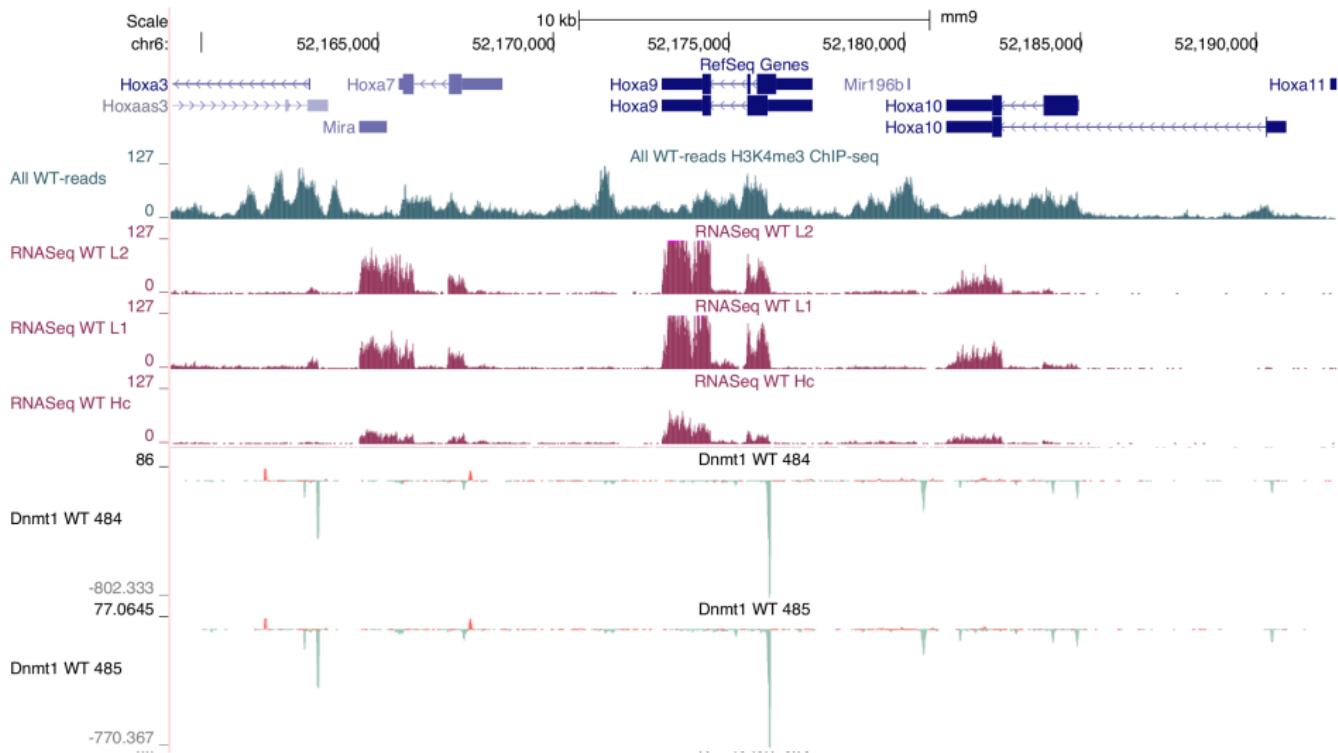
5. Decrosslink and recover formerly bound DNA
6. Sequence DNA
7. Alignment on the reference genome



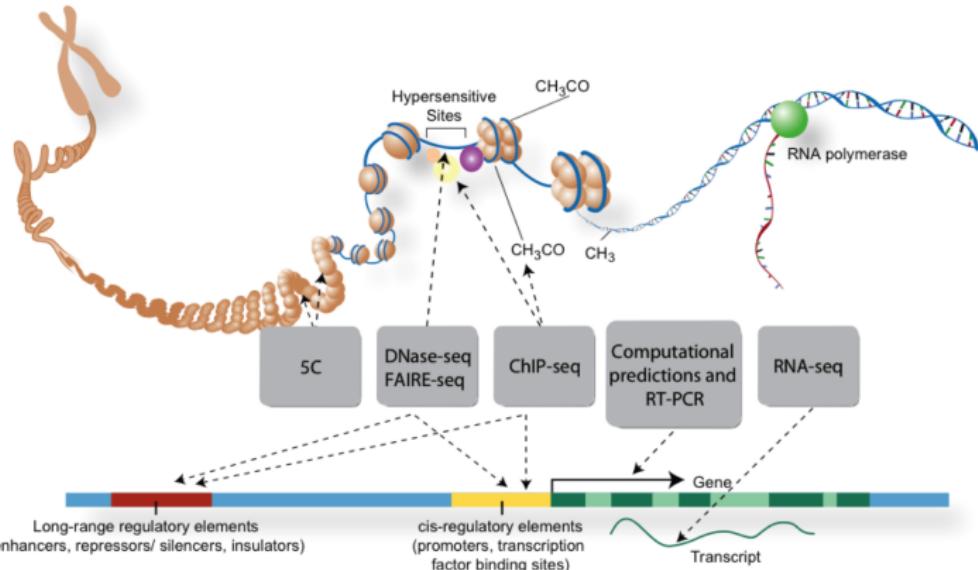
8. Subtract signals from negative control (IgG-control)
9. Peak detection and assignment to genes

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)

Most methods now on single-cell level

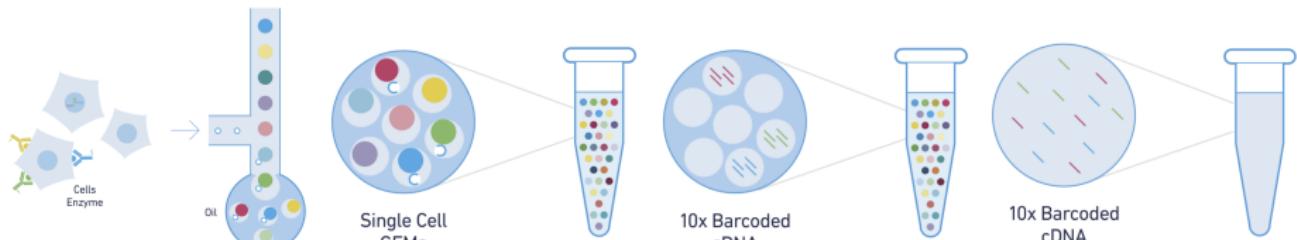


Figure 5: Artificial DNA barcode tags allow to determine cell of origin

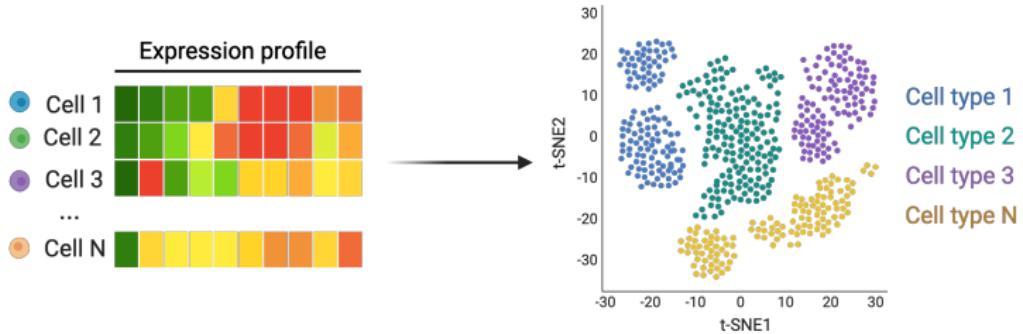


Figure 6: Dimensionality reduction methods (UMAP, t-SNE) separate cell-types

Spatial transcriptomics: Microscopy + single-cell sequencing

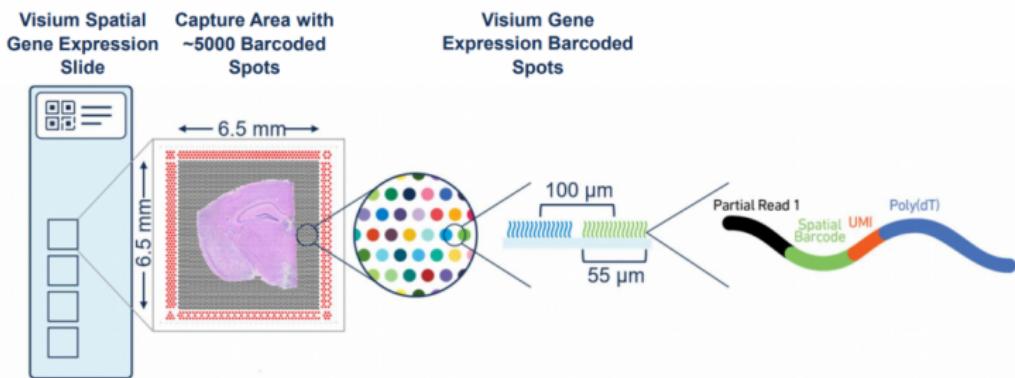


Figure 7: Cell barcodes on microscopy slide

Spatial transcriptomics: Microscopy + single-cell sequencing

The screenshot shows the Human Cell Atlas website. At the top left is the logo 'HUMAN CELL ATLAS' with a blue circular icon. The top navigation bar includes links for 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^{1*}, William E. Allen^{1,2*}, Matthew A. Wright^{1,3}, Emily L. Sylwestrak¹, Nikolay Samusik⁴, Sam Vesuna¹, Karen C. M. van der Loos¹, Charu Ramakrishnan¹, Jia Liu¹, Garry P. Nolan^{4†}, Felice-Alessio Bava^{4†}, Karl Deisseroth^{1,3,6†}

¹Department of Bioengineering, Stanford University, Stanford, CA 94301, USA

<https://www.humancellatlas.org>

[Wang et al., 2018], Science 361(6400)

SciLifeLab is one of the birthplaces of spatial transcriptomics

10X GENOMICS

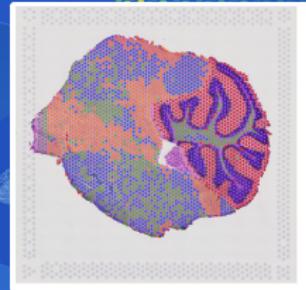
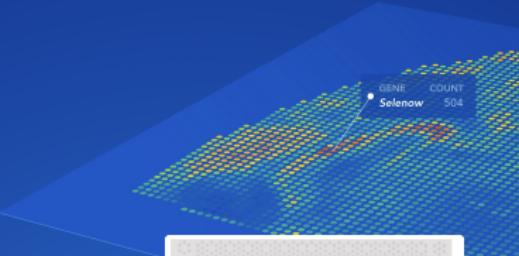
Products Area of Interest Resources Support Company Careers

 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)

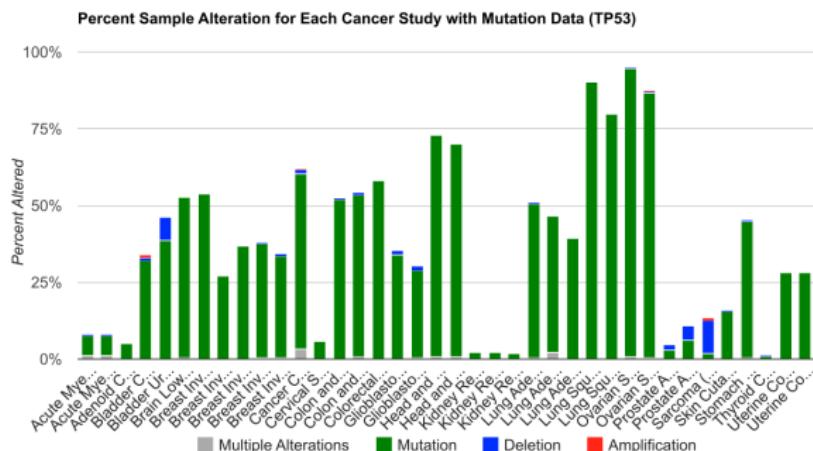
- Cancer Genome Project
- Cancer Genome Atlas
- PanCancerAtlas
- ...

Cancer Cell

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

André Kahles ¹⁵ • Kjøng-Van Lehmann ¹⁵ • Nora C. Toussaint • Matthias Hüser • ... Chris Sander • Gunnar Rätsch ¹⁶ • 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>

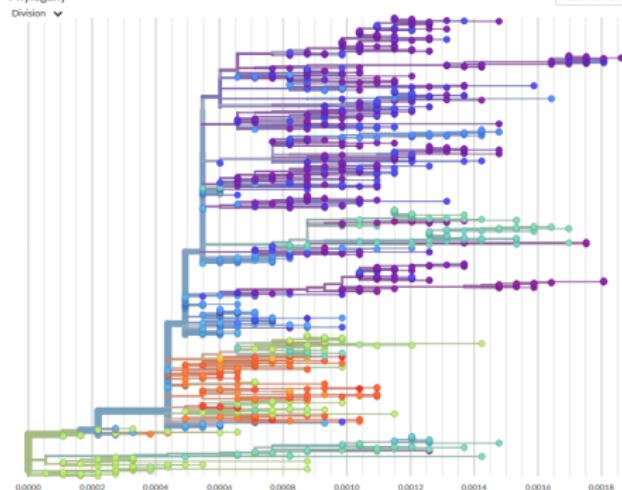
Applications that explore what is around us

Ebola epidemic 2016: Infectious disease monitoring

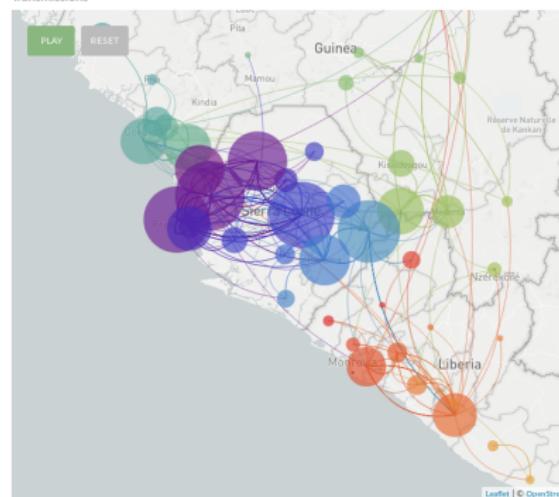
Genomic epidemiology of the 2013-2016 West African Ebola epidemic

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.

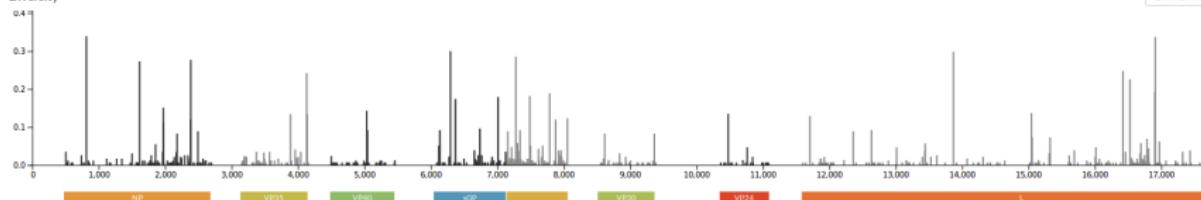
Phylogeny



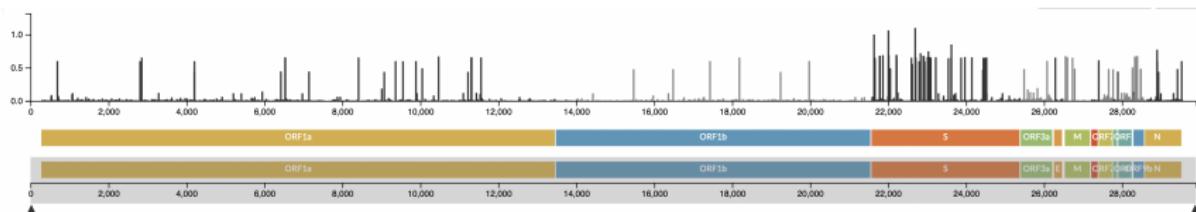
Transmissions



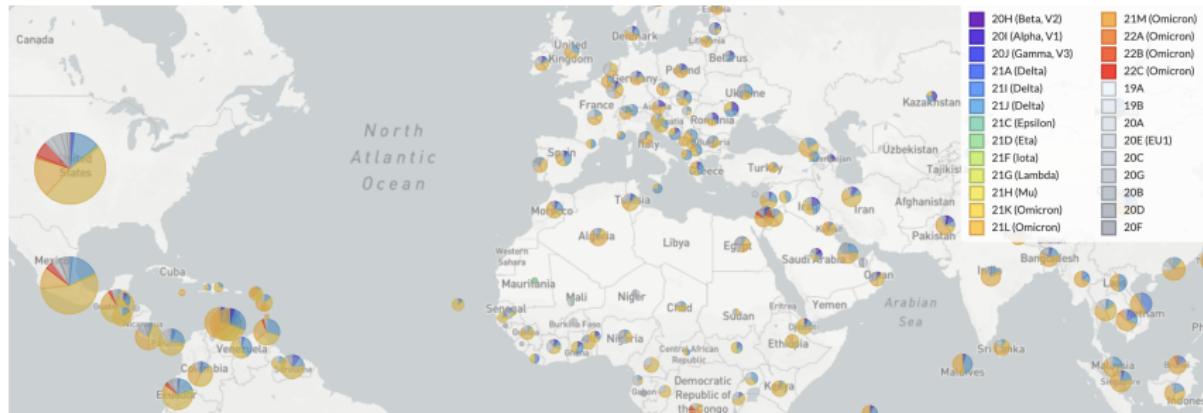
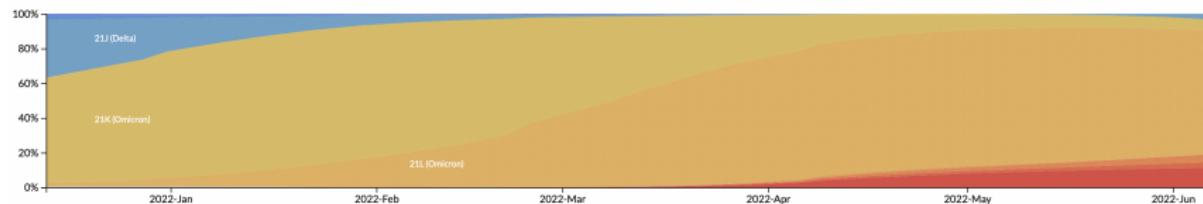
Diversity



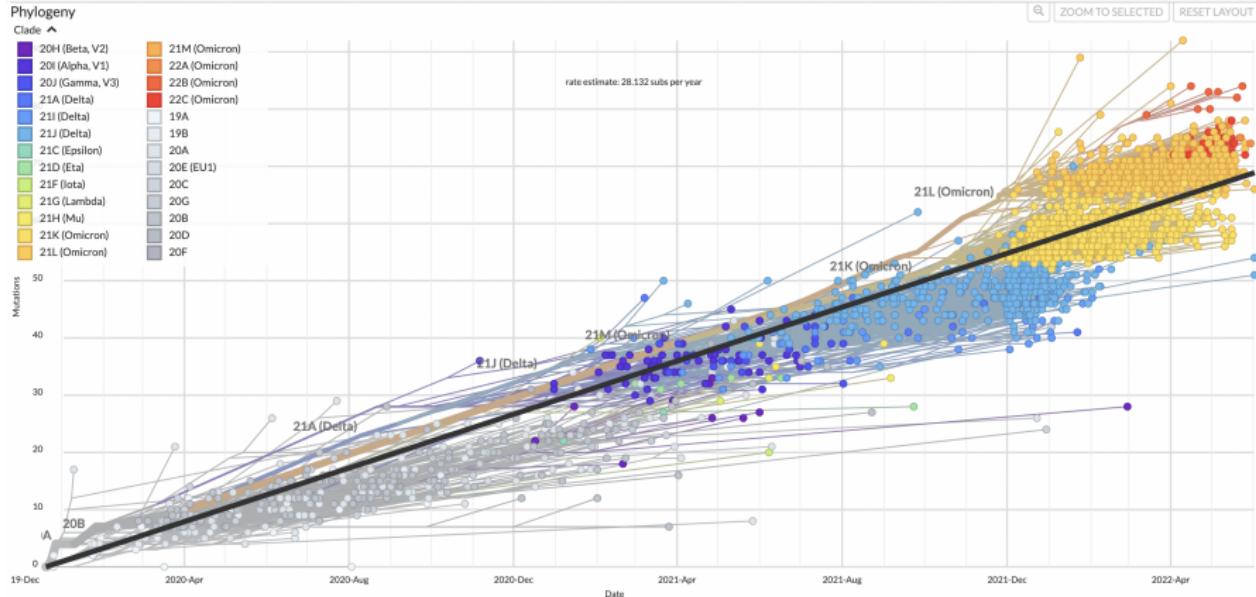
SARS-CoV-2 pandemic: Global variant monitoring



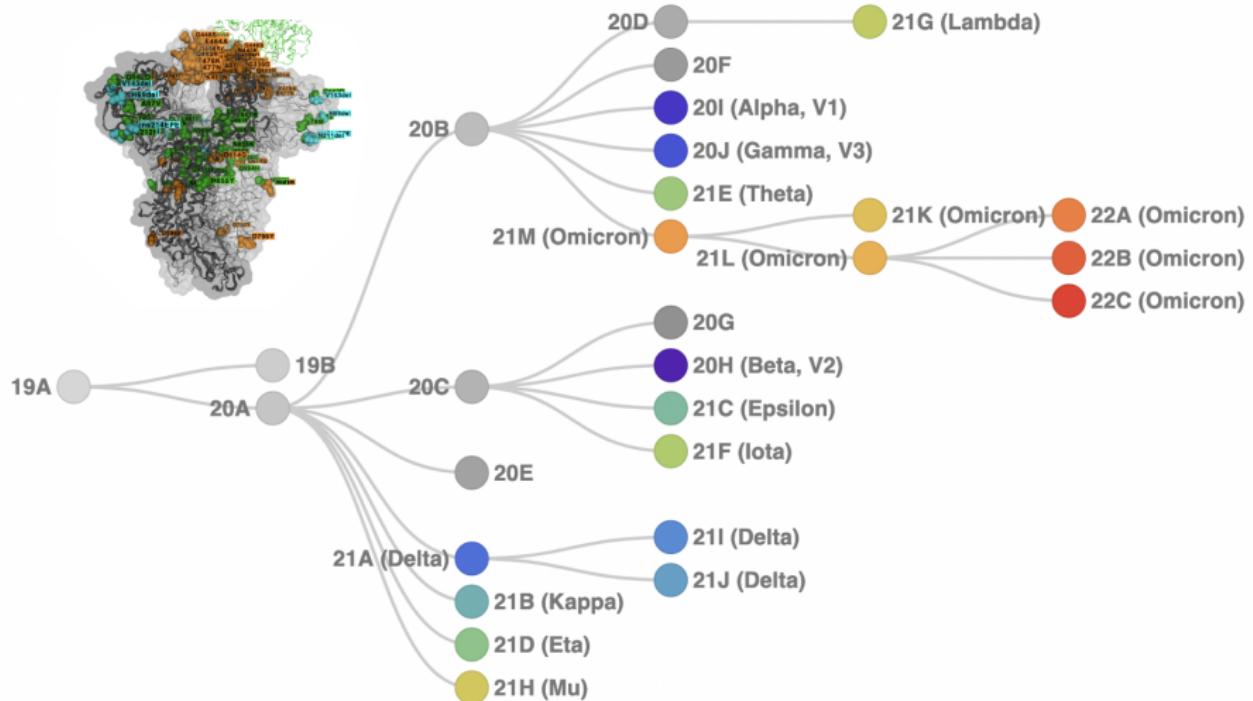
Frequencies (colored by Clade)



Tens of millions of genomes were deposited in GISAID database

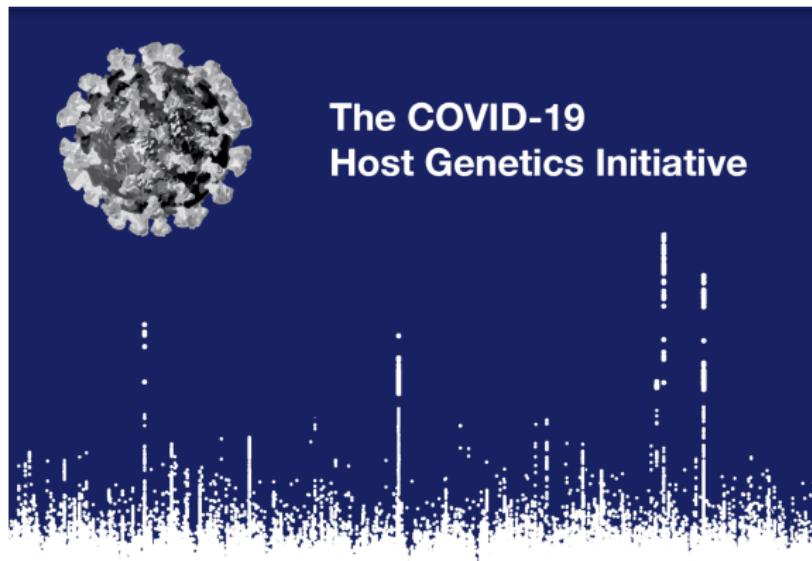


SARS-CoV-2 pandemic: Risk-variant assessment



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

SARS-CoV-2 pandemic: Host genetics

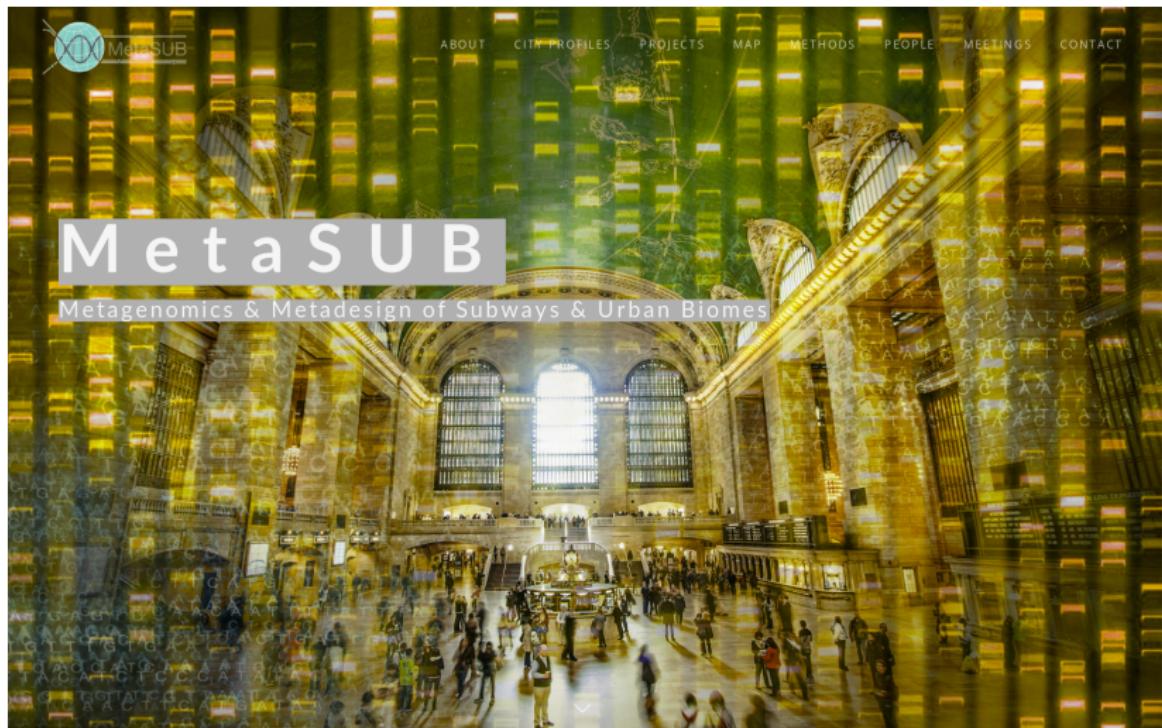


- Severity
- Susceptibility

<https://www.covid19hg.org>

- INF α 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.

DNA-Test bei Nessies Nachbarn

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? Wollen DNA-Spuren aller Lebewesen im See gesammelt werden? Ein internationales Forscher-Team will die Spuren gesammelt werden. Geleitet wird das Team von Neil Gemmell, der Legende vom Unergründlichen.

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

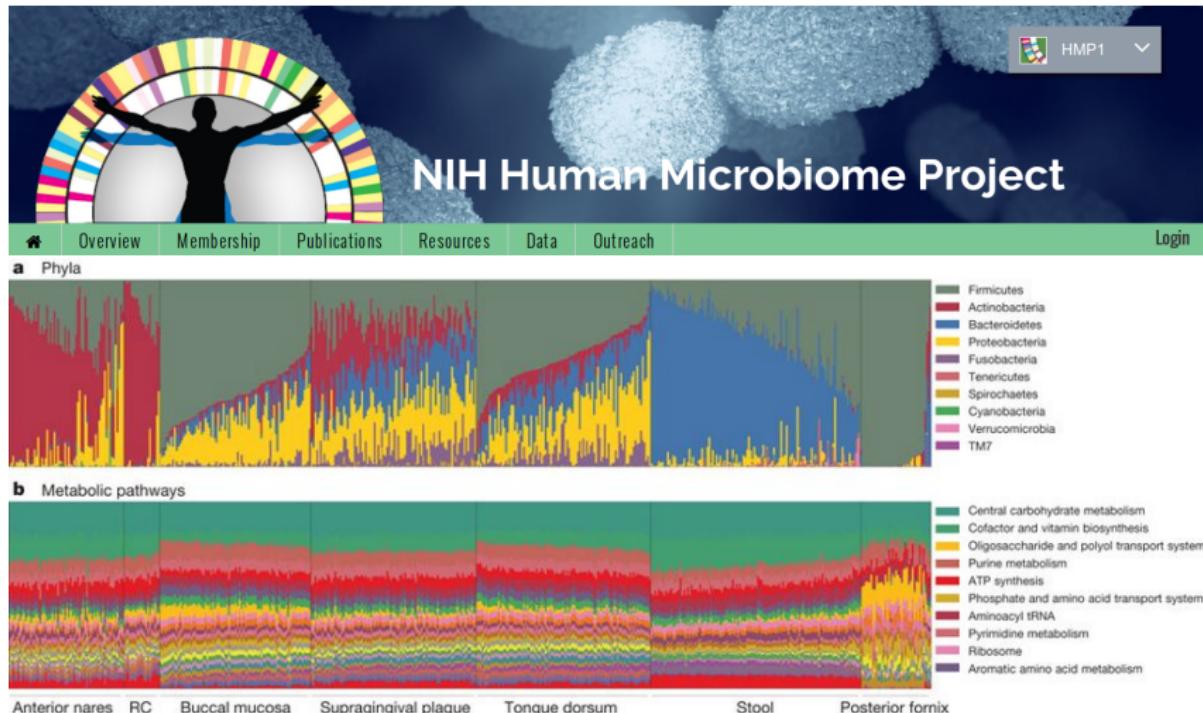
The Guardian

Scientists to lead DNA hunt for Loch Ness monster

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

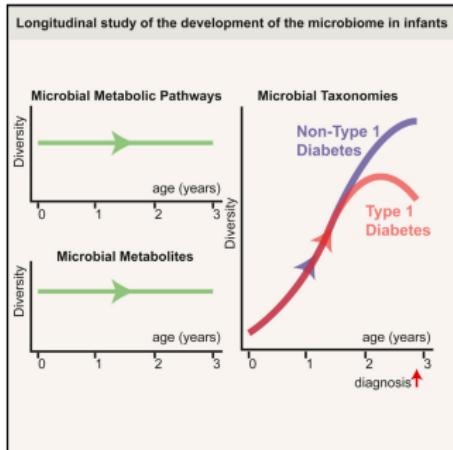
<https://gemmell-lab.otago.ac.nz>

What is on and within us: Human Microbiome Project



<https://hmpdacc.org>

Bacterial dysbiosis facilitates type 1 diabetes



CellPress

Cell Host & Microbe
Resource

The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes

Aleksandar D. Kostic,^{1,2,3} Dirk Gevers,¹ Heli Siljander,^{4,5} Tommi Vatanen,^{1,6} Tuulia Hyötyläinen,^{7,11} Anu-Maria Hämäläinen,⁹ Aleksandr Peet,¹⁰ Vallo Tillmann,¹⁰ Päivi Pöhö,^{11,12} Ismo Mattila,^{7,11} Harri Lähdesmäki,⁶ Eric A. Franzosa,³ Outi Vaarala,⁵ Marcus de Goffau,¹² Hermie Harmsen,¹² Jorma Ilonen,^{13,14} Suvi M. Virtanen,¹⁵ Clary B. Clish,¹ Matej Oresic,^{7,11} Curtis Huttenhower,^{1,3} Mikael Fagerström,¹⁶ MUNE Study Group,¹⁷ and Jari T. Tuomi^{1,2,3} on behalf of the

frontiers in
MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE

published: 10 December 2014
doi: 10.3389/fmicb.2014.00678

Bacteroides dorei dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes

Austin G. Davis-Richardson¹, Alexandria N. Ardissoone¹, Raquel Dias¹, Ville Simell¹, Michael T. Leonard¹, Kaisa M. Kemppainen¹, Jennifer C. Drew¹, Desmond Schatz³, Mark A. Atkinson⁴, Bryan Kolaczkowski¹, Jorma Ilonen^{5,6}, Mikael Knip^{7,8,9}, Jari Tuomi^{1,2,3}, Jorma Nurminen¹⁰, Heikki Hyöty¹⁰, Riitta Veijola¹¹, Tuula Simell¹², Jari Tuomi^{1,2,3}

[Kostic et al., 2015], Cell Host & Microbe 17, 260–273
[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,^{1,3} Fangfang Guo,^{1,3} Yanan Yu,¹ Tian Tian Sun,¹ Dan Ma,¹ Jixuan Han,¹ Yun Qian,¹ Ilona Kryczek,² Dan Feng Sun,^{1,2} Nisha Nagarkoth,² Yingxuan Chen,^{1,*} Haoyan Chen,^{1,*} Jie Hong,^{1,4} Weiping Zou,^{2,4,5} and Jing-Yuan Fang^{1,*}

¹State Key Laboratory for Oncogenes and Related Genes, Key Laboratory of Gastroenterology and Hepatology, Shanghai Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China
²Department of Molecular Psychiatry, University of Heidelberg, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom
³Department of Cell Biology, University of Heidelberg, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom
⁴Department of Cell Biology, University of Heidelberg, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom
⁵Department of Cell Biology, University of Heidelberg, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom

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

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,^{1,2} Serena Verdi,¹ Maria-Emanuela Maxan,³ Cheol Min Shin,^{1,4} Jonas Zierer,^{1,5} Ruth C.E. Bowyer,¹ Tinhaiou Martin,^{1,6} Francisco J. Lopez,^{1,7} Daniel E. Gitter,^{1,8} and David J. Schlessinger,^{1,9}

[Yu et al., 2017], Cell 170, 548–563

[Zheng et al., 2016], Mol Psychiatry. 21(6):786–96

[Berer et al., 2011], Nature 479, 538–542

[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,^{1,*†‡} Johannes Krause,^{1,§} Adrian W. Briggs,^{1,§} Tomislav Maricic,^{1,§} Udo Stenzel,^{1,§} Martin Kircher,^{1,§} Nadin Rohland,^{2,§} Michael Hofreiter,^{2,§} Heng Li,^{2,§} Michael Hofreiter,^{2,§} Zhe Ji,^{3,||}

The complete mitochondrial DNA genome of an unknown hominin from southern Siberia

Johannes Krause¹, Qiaomei Fu¹, Jeffrey M. Good², Bence Viola^{1,3}, Michael V. Shunkov⁴, Anatoli P. Derevianko⁴ & Svante Pääbo¹

Figure 8: Nobel laureate Svante Pääbo. Painted by Sixten Sandra Österberg for the Swedish National Portrait Gallery at Gripsholms slott. →



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

[Green et al., 2010], Science Vol. 328, 710-722

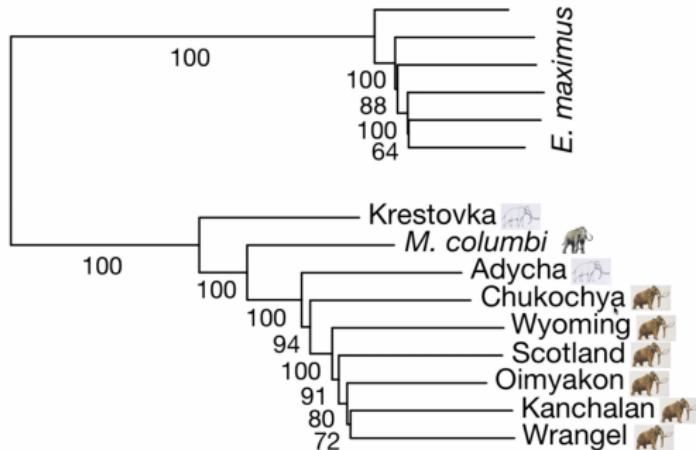
[Krause et al., 2010], Nature Vol. 464, 894-897

BEST BEFORE

50.000 y/o

Store in cool, dry place

Ancient genomes shed light on evolutionary developments

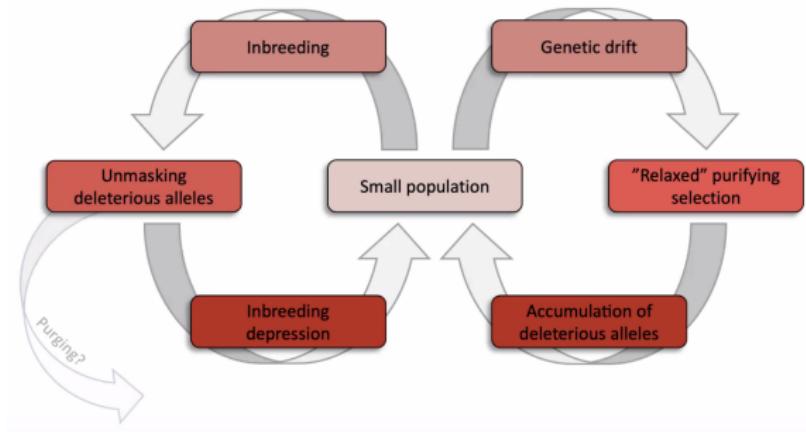


- Hair growth
- Thermal sensation
- Fat deposition
- Thermoregulation
- Circadian rhythm



[van der Valk et al., 2021], Nature 591, pages 265–269

Genomic processes in populations on the brink of extinction



[Dussex et al., 2021], Cell Genomics Volume 1, Issue 1
[Liu et al., 2021], Volume 184, Issue 19, 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 9: Three-month-old marbled crayfish from the same clutch.

[Gutekunst et al., 2018], Nature Ecology & Evolution vol. 2, pages 567-573

Agrigenomics - analysis of cultured plants and farm animals

The tragic ripples of an epic fraud p. 626

Insect pest profits from maize defenses pp. 642 & 648

Science



ROAD MAP FOR WHEAT
Ordered sequence will speed research pp. 635, 661 & 662

Genome sequence of the soybean

Jeremy Schmutz^{1,2}, Steven B. Casper¹, David L. Hyten¹, Michael J. Sorrells³,... 17 AUGUST 2008

AAAS

The genome of *Theobroma cacao*

Xavier Argout^{1,24}, Jerome Salse^{2,24}, Jean-Marc Aury^{3,5,24}, Mark J. Guiltinan⁶, Mathilde Allegre³, Cristian Chambard⁹, Thierry Legavre¹, Sicla N. Maxine¹⁰, Olivier Fouet¹, Jean-Pierre... 17 AUGUST 2008

The genome of woodland strawberry

Vladimir Shulaev^{1*}, Daniel J. Sargent², Ross N. Crowhurst³, Todd C. Mockler⁴, Pankaj Jaiswal⁴, Keithanne Mockaitis⁸, Aaron Liston⁴, Shrinivas... 17 AUGUST 2008

The genome of the cucumber, *Cucumis sativus* L.

Jiang Li^{2,3,19}, Zhonghua Zhang^{1,19}, Li Li^{2,19}, Xingfang Gu^{1,19}, Wei Fan^{2,19}, Zhou Wang¹, Bingyan Xie¹, Peixiang Ni², Yuanyuan Ren²,... 17 AUGUST 2008

A Draft Sequence of the Rice Genome

Darrel Rickett¹, Tien-Hung Lan¹,... 17 AUGUST 2008

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

Emma S. Mace^{1,*}, Shuaishuai Tai^{2,*}, Edward K. Gilding³, Yanhong Li², Peter J. Prentis⁴, Liapin... 17 AUGUST 2008

Genome sequence and analysis of the tuber crop potato

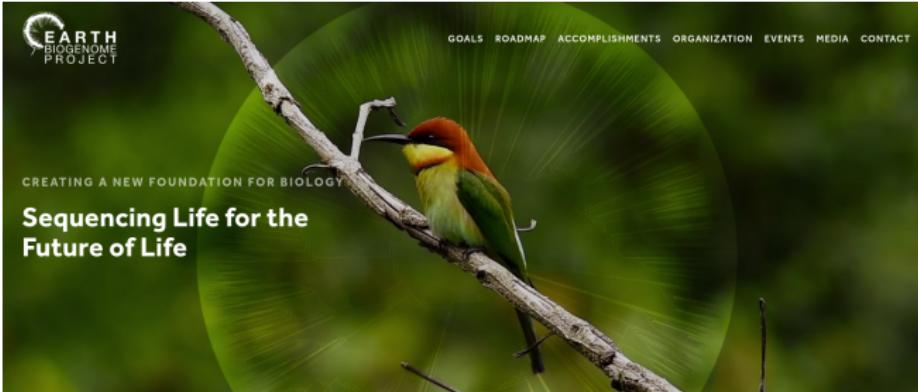
The Potato Genome Sequencing Consortium*

The B73 Maize Genome: Complexity, Diversity,

Patrick S. Schnable^{1,2,3,4,5,6,7}, Doreen Ware^{6,8,9}, Robert S. Fulton^{1,2,3,4,5,6,7,10},... 17 AUGUST 2008

<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

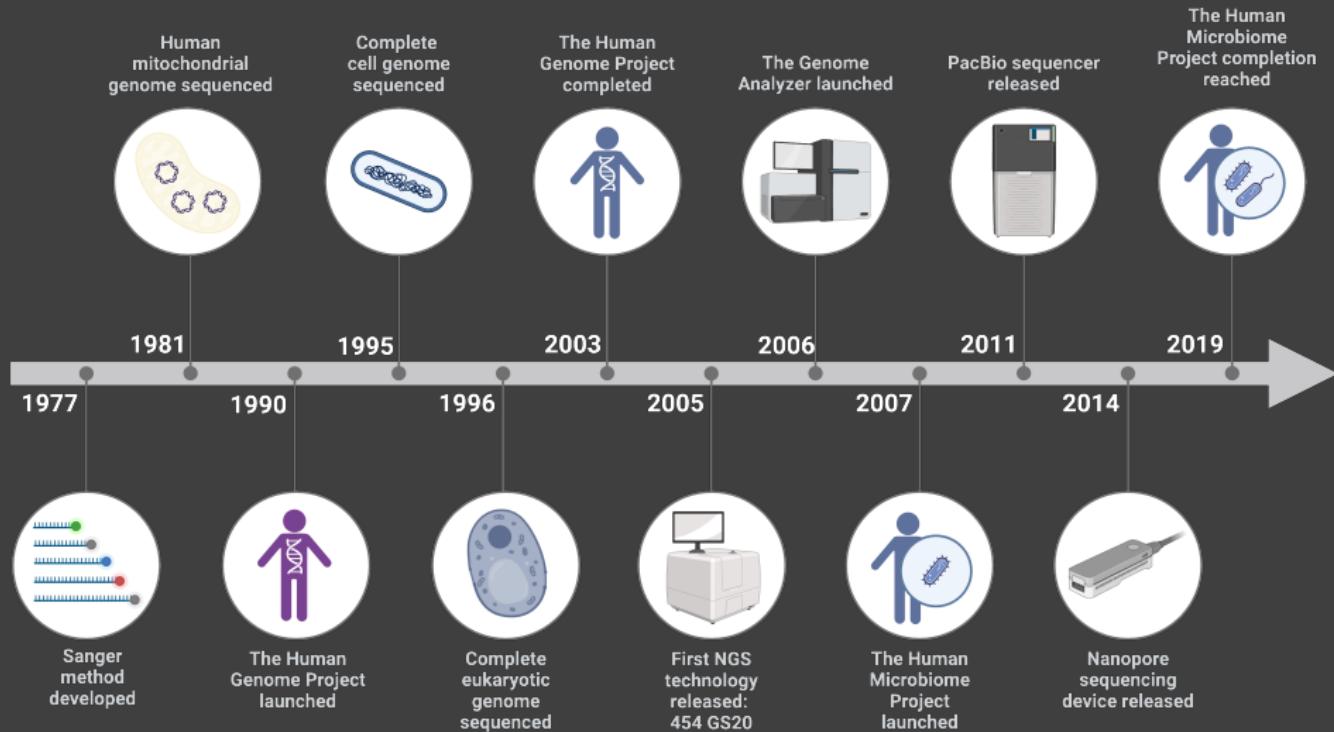


Figure by Anja Mezger

NGI has you covered for your research



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

References

References i

-  Berer, K., Mues, M., Koutrolos, M., Rasbi, Z. A., Boziki, M., Johner, C., Wekerle, H., & Krishnamoorthy, G. (2011). Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination.. *Nature*, 479, 538–541.
<https://doi.org/10.1038/nature10554>
-  Davis-Richardson, A. G., Ardissono, A. N., Dias, R., Simell, V., Leonard, M. T., Kemppainen, K. M., Drew, J. C., Schatz, D., Atkinson, M. A., Kolaczkowski, B., Ilonen, J., Knip, M., Toppari, J., Nurminen, N., Hyöty, H., Veijola, R., Simell, T., Mykkänen, J., Simell, O., & Triplett, E. W. (2014). Bacteroides dorei dominates gut microbiome prior to autoimmunity in finnish children at high risk for type 1 diabetes.. *Frontiers in microbiology*, 5, 678. <https://doi.org/10.3389/fmicb.2014.00678>
-  Dussex, N., van der Valk, T., Morales, H. E., Wheat, C. W., Díez-Del-Molino, D., von Seth, J., Foster, Y., Kutschera, V. E., Guschnski, K., Rhie, A., Phillippy, A. M., Korlach, J., Howe, K., Chow, W., Pelan, S., Mendes Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., ... Dalén, L. (2021). Population genomics of the critically endangered kākāpō.. *Cell genomics*, 1, 100002. <https://doi.org/10.1016/j.xgen.2021.100002>

References ii

-  Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., ... Pääbo, S. (2010). A draft sequence of the neandertal genome.. *Science (New York, N.Y.)*, 328, 710–722. <https://doi.org/10.1126/science.1188021>
-  Gutekunst, J., Andriantsoa, R., Falckenhayn, C., Hanna, K., Stein, W., Rasamy, J., & Lyko, F. (2018). Clonal genome evolution and rapid invasive spread of the marbled crayfish.. *Nature Ecology & Evolution*, 2, 567–573. <https://doi.org/10.1038/s41559-018-0467-9>
-  Jackson, M. A., Verdi, S., Maxan, M.-E., Shin, C. M., Zierer, J., Bowyer, R. C. E., Martin, T., Williams, F. M. K., Menni, C., Bell, J. T., Spector, T. D., & Steves, C. J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort.. *Nature communications*, 9, 2655. <https://doi.org/10.1038/s41467-018-05184-7>

References iii

-  Kostic, A. D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A.-M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., Lähdesmäki, H., Franzosa, E. A., Vaarala, O., de Goffau, M., Harmsen, H., Ilonen, J., Virtanen, S. M., Clish, C. B., Orešić, M., ... Xavier, R. J. (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes.. *Cell Host & Microbe*, 17, 260–273. <https://doi.org/10.1016/j.chom.2015.01.001>
-  Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., & Pääbo, S. (2010). The complete mitochondrial dna genome of an unknown hominin from southern siberia.. *Nature*, 464, 894–897. <https://doi.org/10.1038/nature08976>
-  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>

References iv

-  Liu, S., Westbury, M. V., Dussex, N., Mitchell, K. J., Sinding, M.-H. S., Heintzman, P. D., Duchêne, D. A., Kapp, J. D., von Seth, J., Heiniger, H., Sánchez-Barreiro, F., Margaryan, A., André-Olsen, R., De Cahsan, B., Meng, G., Yang, C., Chen, L., van der Valk, T., Moodley, Y., ... Gilbert, M. T. P. (2021). Ancient and modern genomes unravel the evolutionary history of the rhinoceros family.. *Cell*, 184, 4874–4885.e16. <https://doi.org/10.1016/j.cell.2021.07.032>
-  Rodriguez, R., & Krishnan, Y. (2023). The chemistry of next-generation sequencing.. *Nature biotechnology*, 41, 1709–1715. <https://doi.org/10.1038/s41587-023-01986-3>
-  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>
-  van der Valk, T., Pečnerová, P., Díez-Del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Sağlıcan, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., ... Dalén, L. (2021). Million-year-old dna sheds light on the genomic history of mammoths.. *Nature*, 591, 265–269. <https://doi.org/10.1038/s41586-021-03224-9>

References v

-  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>
-  Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., Nolan, G. P., Bava, F.-A., & Deisseroth, K. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states.. *Science (New York, N.Y.)*, 361. <https://doi.org/10.1126/science.aat5691>
-  Yu, T., Guo, F., Yu, Y., Sun, T., Ma, D., Han, J., Qian, Y., Kryczek, I., Sun, D., Nagarsheth, N., Chen, Y., Chen, H., Hong, J., Zou, W., & Fang, J.-Y. (2017). Fusobacterium nucleatum promotes chemoresistance to colorectal cancer by modulating autophagy.. *Cell*, 170, 548–563.e16. <https://doi.org/10.1016/j.cell.2017.07.008>
-  Zheng, P., Zeng, B., Zhou, C., Liu, M., Fang, Z., Xu, X., Zeng, L., Chen, J., Fan, S., Du, X., Zhang, X., Yang, D., Yang, Y., Meng, H., Li, W., Melgiri, N. D., Licinio, J., Wei, H., & Xie, P. (2016). Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism.. *Molecular psychiatry*, 21, 786–796. <https://doi.org/10.1038/mp.2016.44>