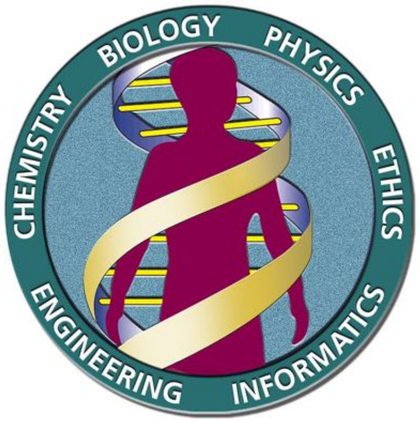
OMICs

The Human Atlas

Steve Knutson December 6, 2024

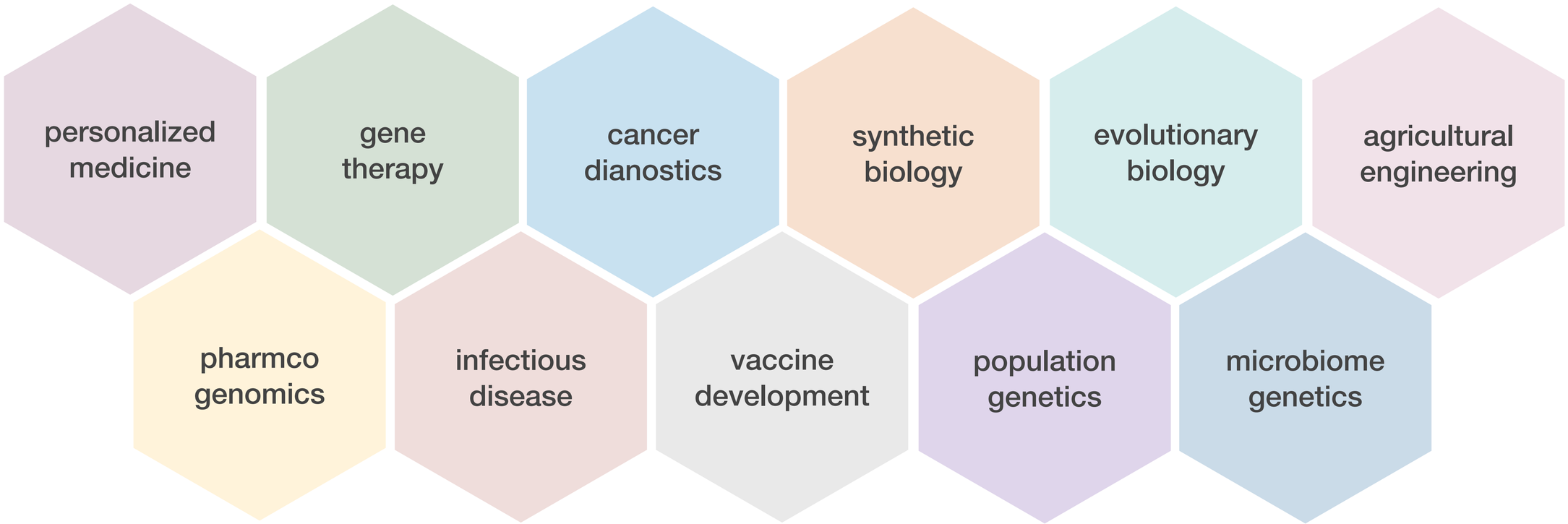
April 14, 2003

The Human Genome Project



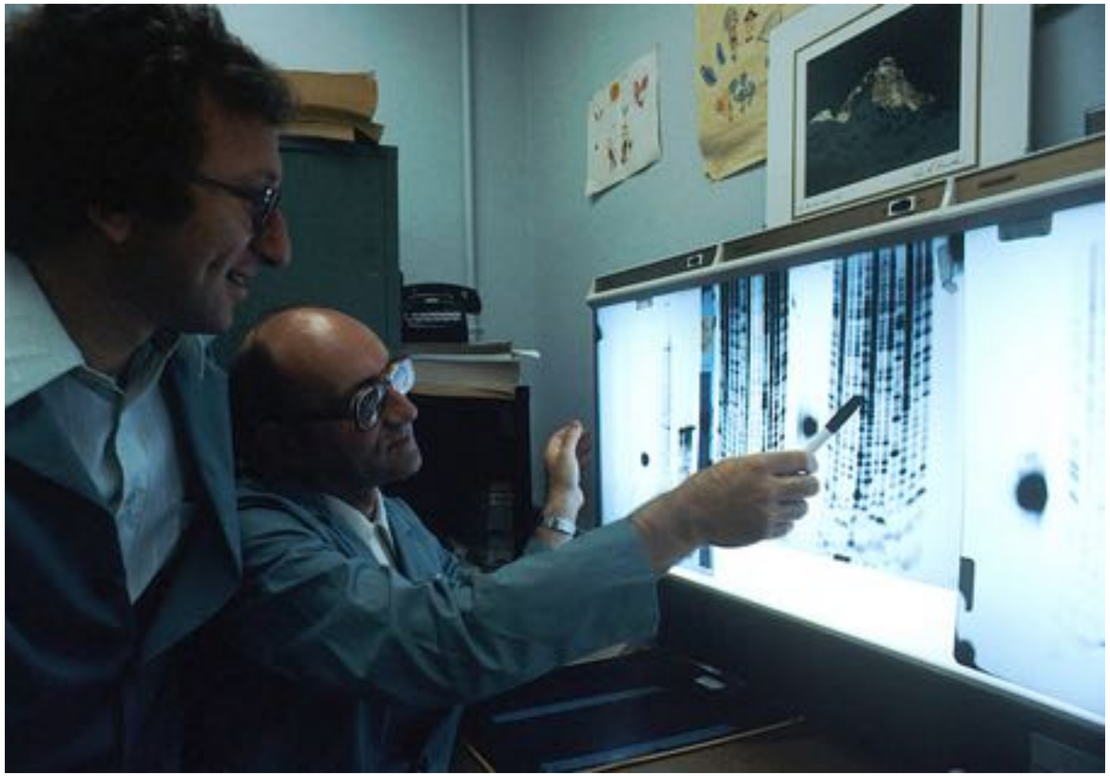
- first map of human genome (92%)
- 15 years, 20 research institutions
- $2.7 billion

“It changed the way we thought biology could be done.”



The technology did not exist

1990s



2005

First generation DNA sequencing

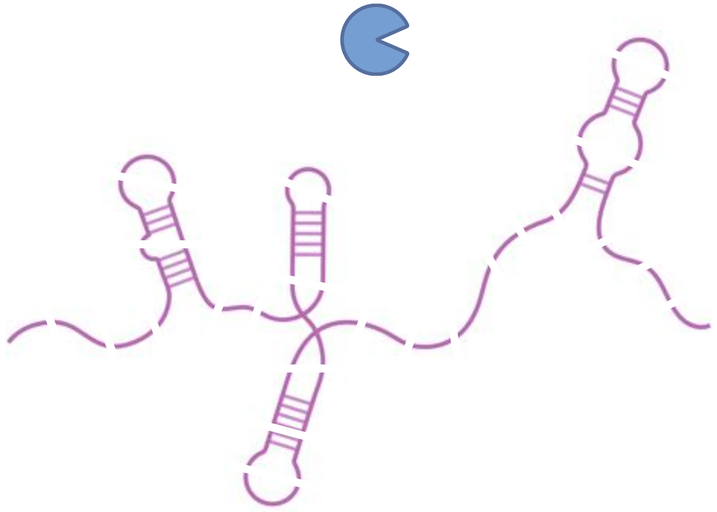
Frederick Sanger

First generation DNA sequencing

The Sequence of 5 s Ribosomal Ribonucleic Acid

G. G. Brownlee, F. Sanger and B. G. Barrell

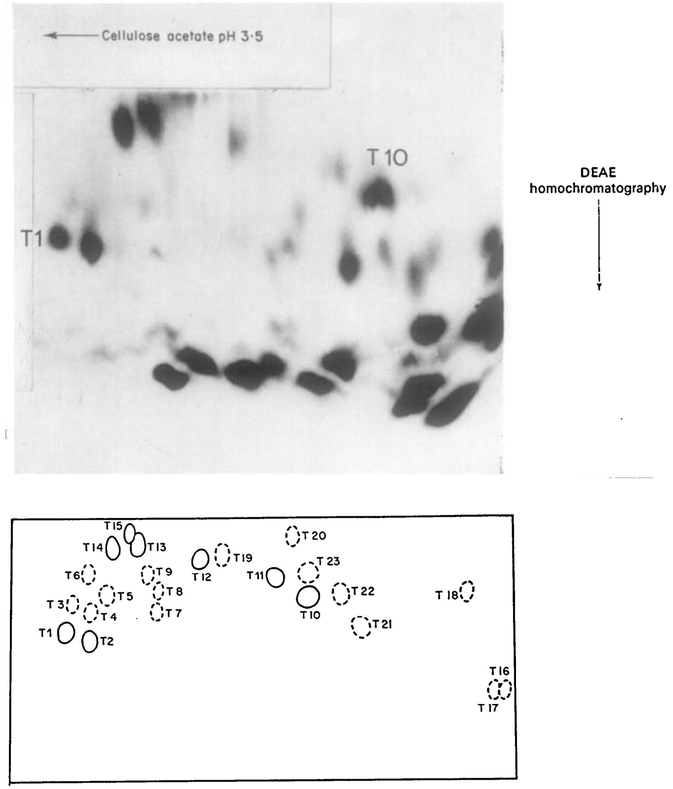
RNase digestion

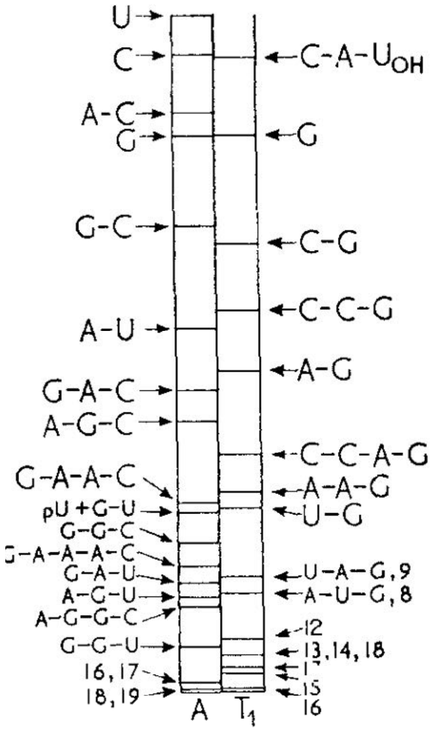


Frederick Sanger

Journal of Molecular Biology 34, no. 3 (1968): 379-412.

First generation DNA sequencing

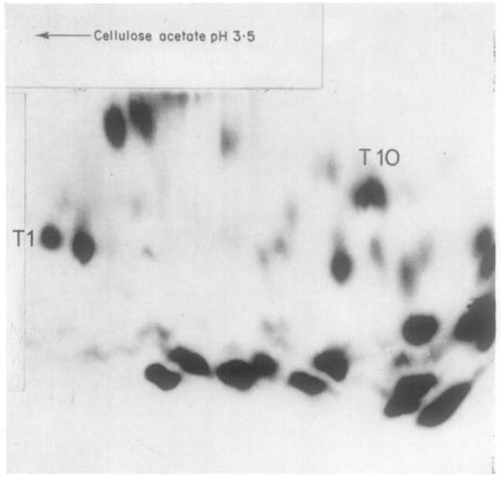




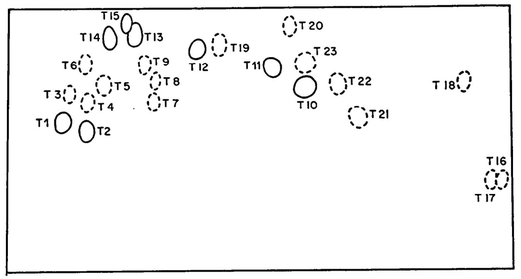
Frederick Sanger

Journal of Molecular Biology 34, no. 3 (1968): 379-412.

First generation DNA sequencing



DEAE
homochromatography



	T_1	A	
T1	UCUCCCCAUG U-G G	G-G-G-G-U G-U A-U G	UGUGGGGUCUCCCCAUG
T2	UCUCCCCAUG U-G G	G-G-G-G-U A-U G	UGGGGUCUCCCCAUG
T3	UCUCCCCAUG U-A-G U-G G	G-G-G-G-U A-G-U G-U A-U G	UAGUGUGGGGUCUCCCCAUG

Frederick Sanger

Journal of Molecular Biology 34, no. 3 (1968): 379-412.

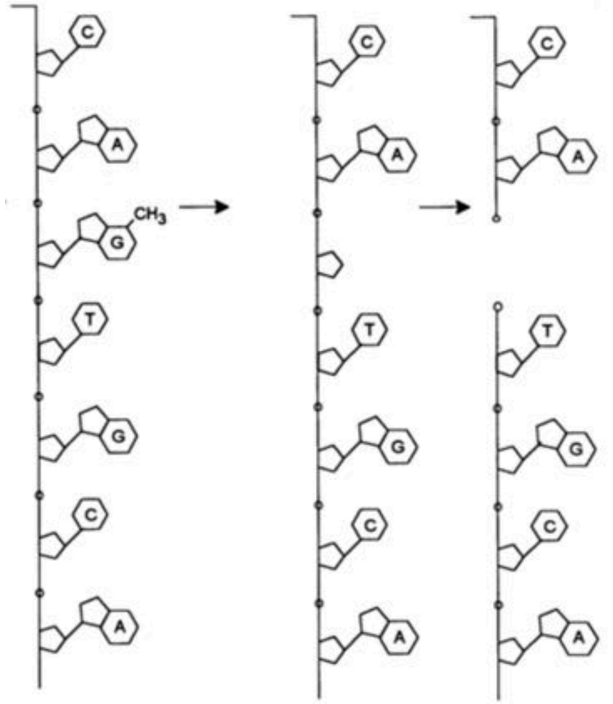
First generation DNA sequencing

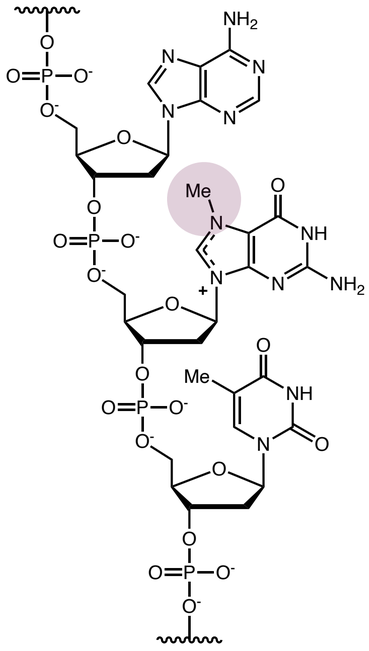
A new method for sequencing DNA

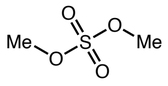
(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT

base modification *depurination* *strand cleavage*







dimethylsulfate

Walter Gilbert

Proc. Natl. Acad. Sci., 1977, *74*(2):560-4.

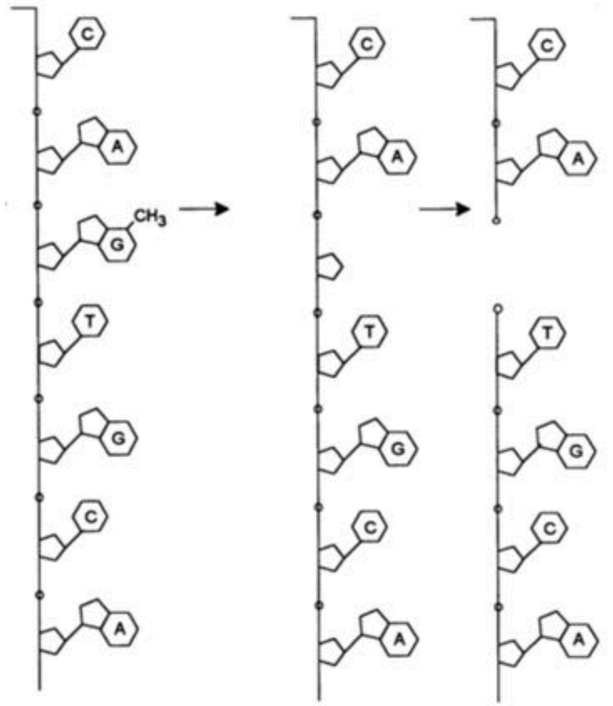
First generation DNA sequencing

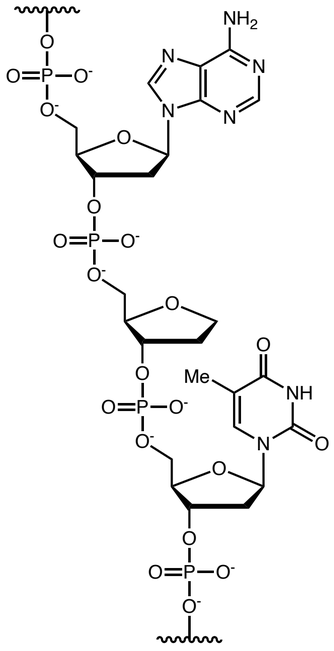
A new method for sequencing DNA

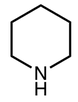
(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT

base modification *depurination* *strand cleavage*







hot piperidine

Walter Gilbert

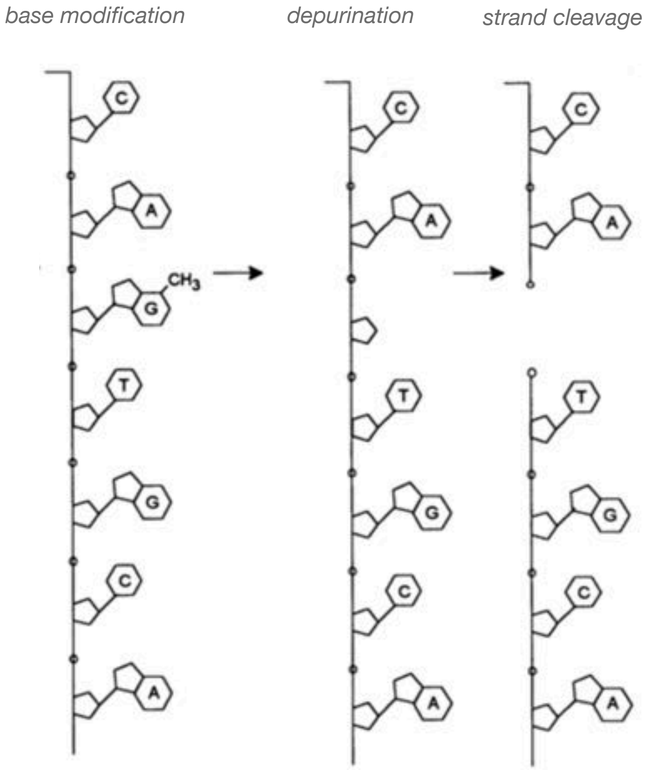
Proc. Natl. Acad. Sci., 1977, *74*(2):560-4.

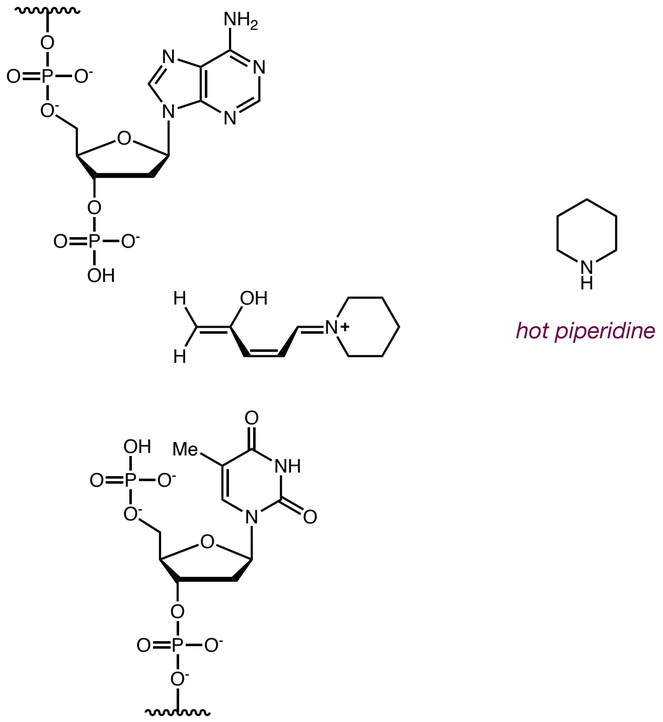
First generation DNA sequencing

A new method for sequencing DNA

(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT



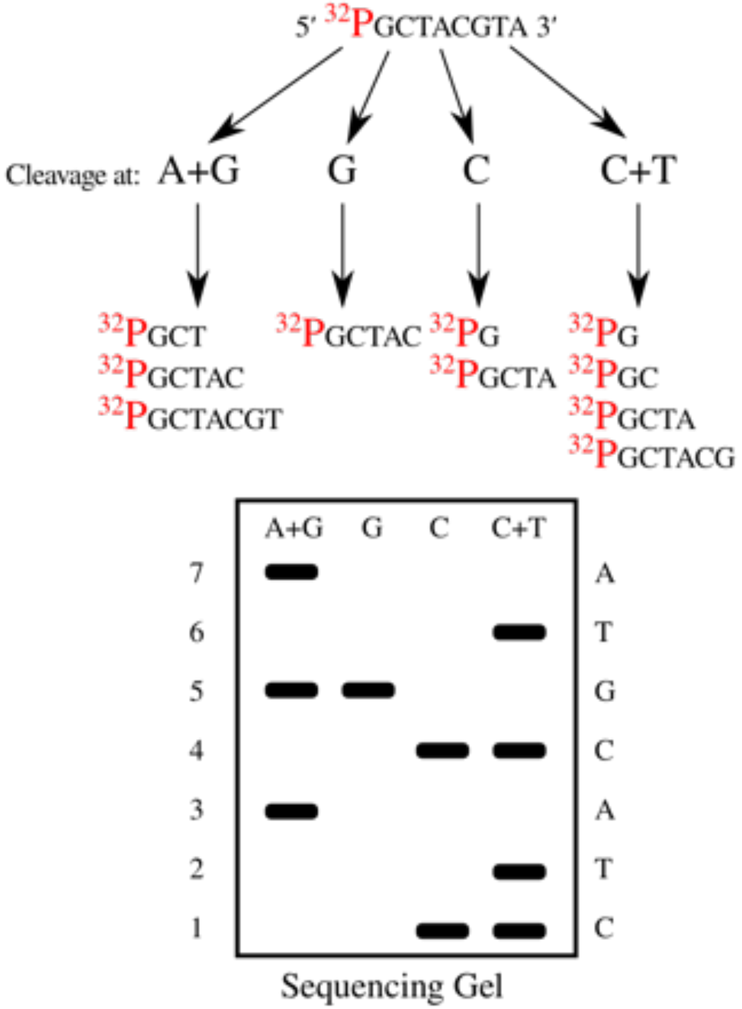


hot piperidine

Walter Gilbert

Proc. Natl. Acad. Sci., 1977, *74*(2):560-4.

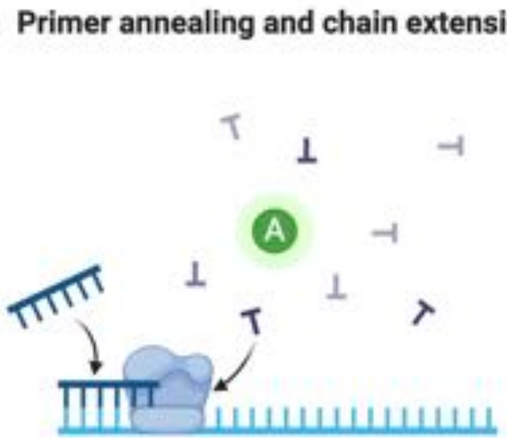
First generation DNA sequencing

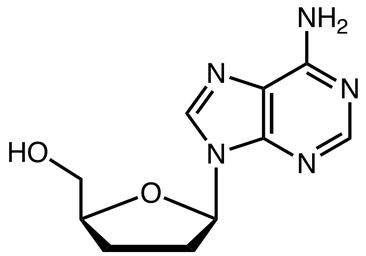


Walter Gilbert

Proc. Natl. Acad. Sci., 1977, *74*(2):560-4.

First generation DNA sequencing

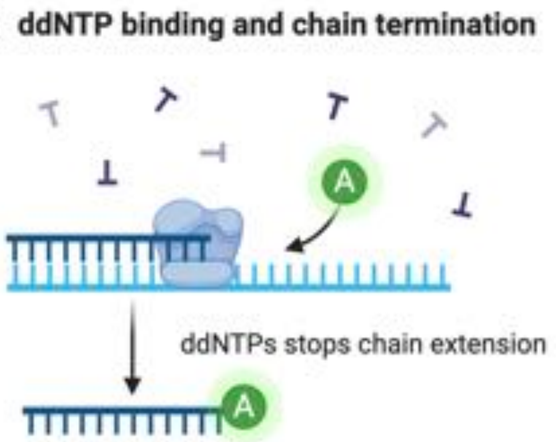


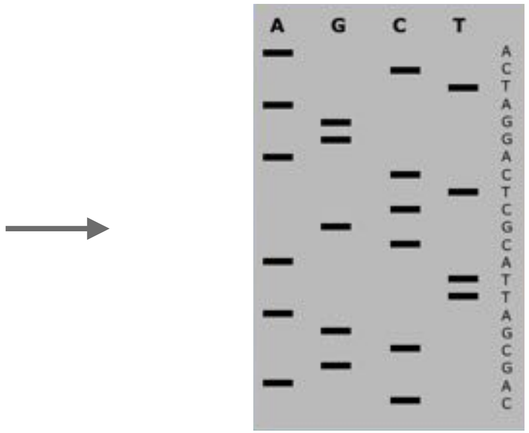


2,3-dideoxyadenosine

- *antibacterial, inhibited DNA synthesis*

Frederick Sanger

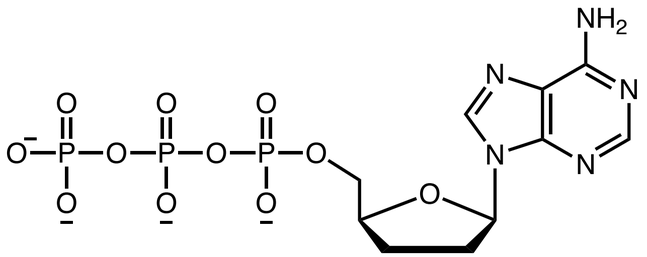




First generation DNA sequencing

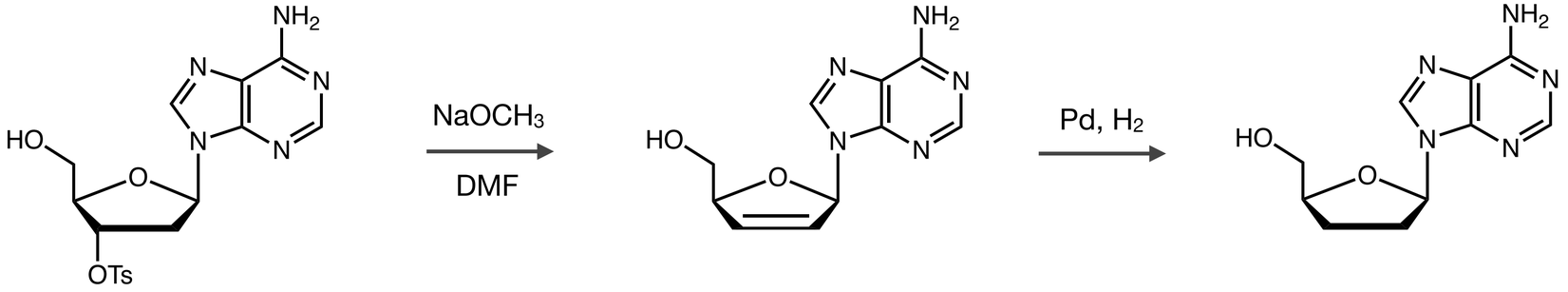
Synthesis of Some Nucleotides Derived from 3′-Deoxythymidine*

Alan F. Russell† and J. G. Moffatt



2,3-dideoxyadenosine triphosphate

Frederick Sanger

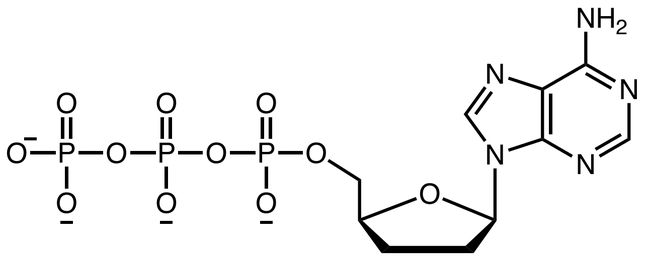


J. Am. Chem. Soc., 1966, *88*(7):1549-53.

First generation DNA sequencing

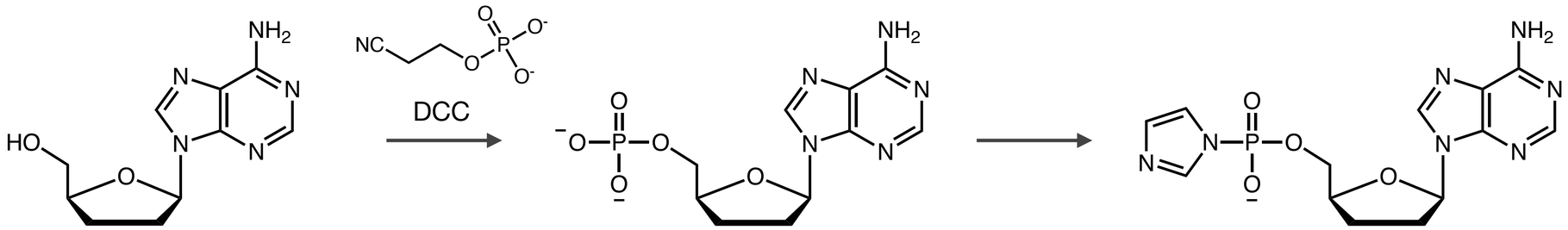
Synthesis of Some Nucleotides Derived from 3′-Deoxythymidine*

Alan F. Russell† and J. G. Moffatt



2,3-dideoxyadenosine triphosphate

Frederick Sanger

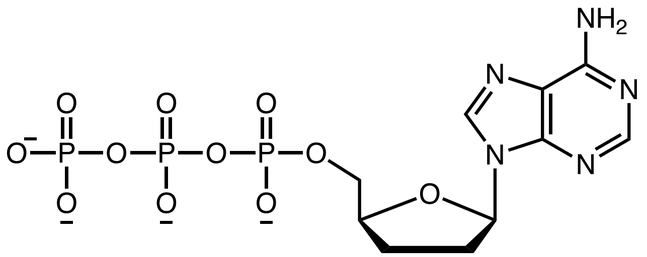


J. Am. Chem. Soc., 1966, *88*(7):1549-53.

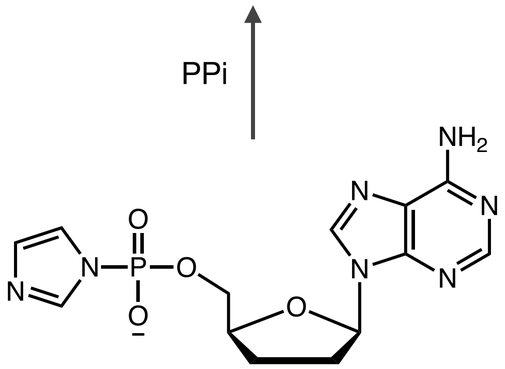
First generation DNA sequencing

Synthesis of Some Nucleotides Derived from 3′-Deoxythymidine*

Alan F. Russell† and J. G. Moffatt



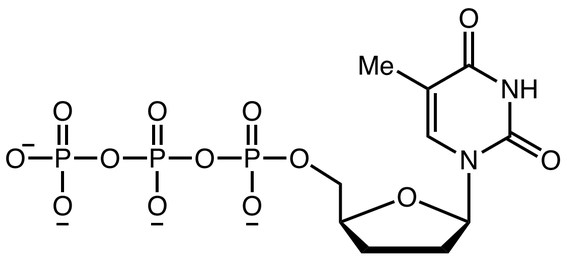
2,3-dideoxyadenosine triphosphate



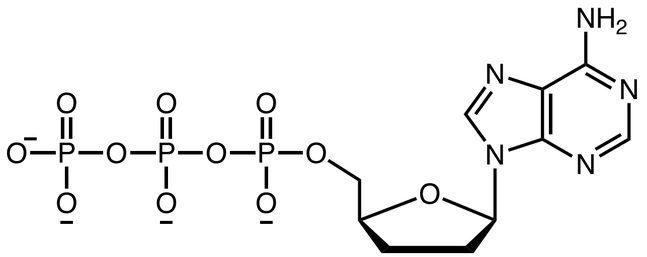
Frederick Sanger

J. Am. Chem. Soc., 1966, *88*(7):1549-53.

First generation DNA sequencing

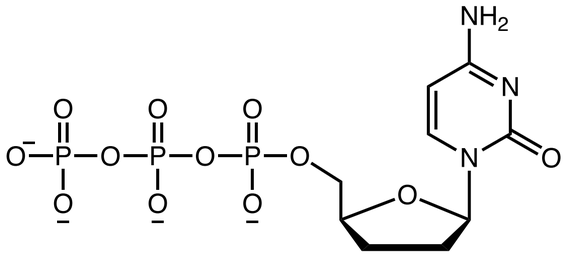


2,3-dideoxythymidine triphosphate

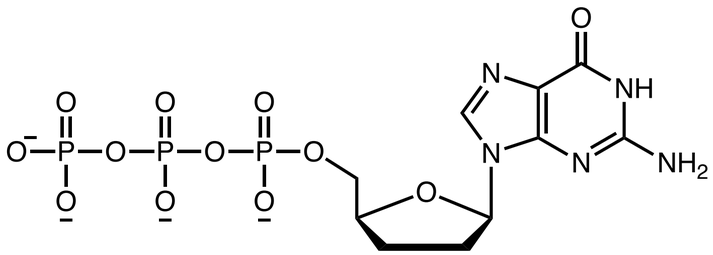


2,3-dideoxyadenosine triphosphate

Frederick Sanger



2,3-dideoxycytidine triphosphate



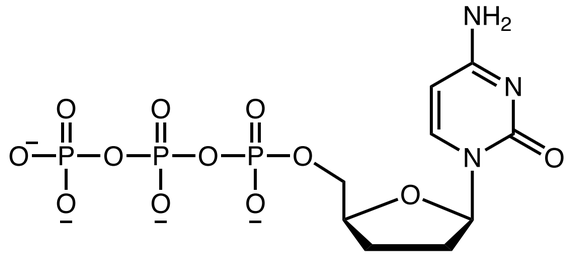
2,3-dideoxyguanosine triphosphate

Alan Coulson

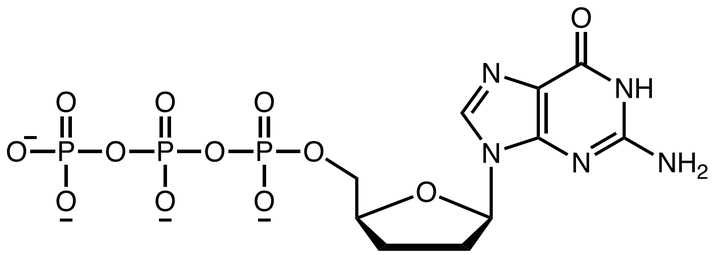
Proc. Natl. Acad. Sci., 1977, *74*(12):5463-7.

First generation DNA sequencing

carbonate at pH 8.4. The preparation of ddGTP and ddCTP has not been described previously; however we applied the same method as that used for ddATP and obtained solutions having the requisite terminating activities. The yields were very low and this can hardly be regarded as adequate chemical characterization. However, there can be little doubt that the activity was due to the dideoxy derivatives.



2,3-dideoxycytidine triphosphate



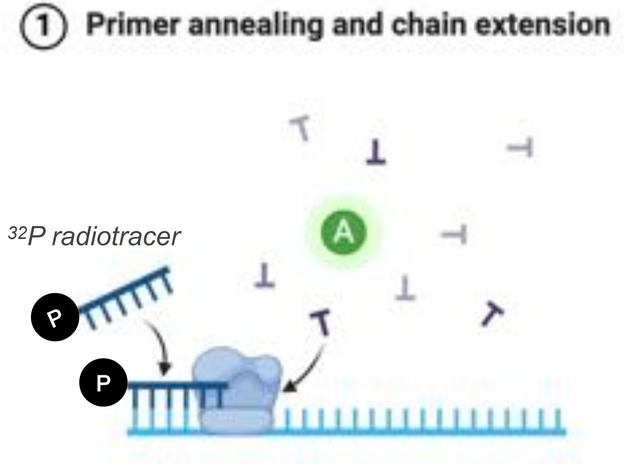
2,3-dideoxyguanosine triphosphate

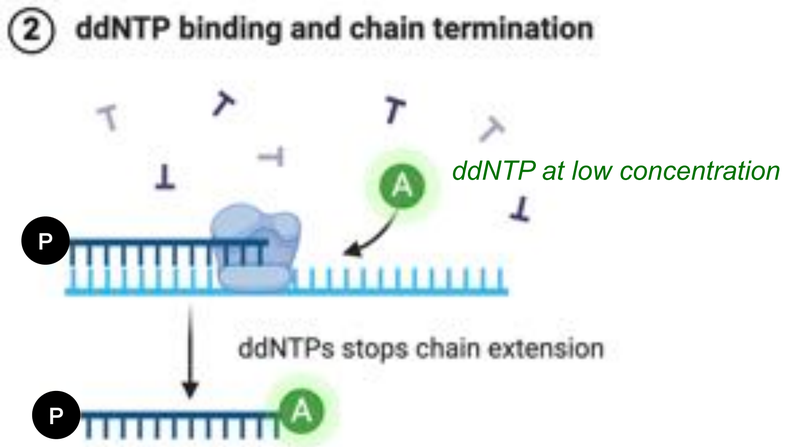
Frederick Sanger

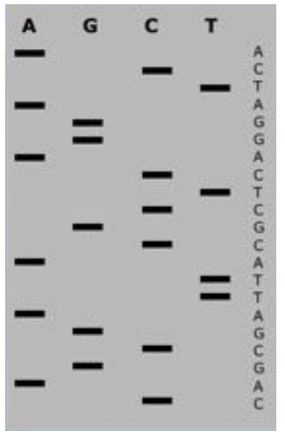
Alan Coulson

Proc. Natl. Acad. Sci., 1977, *74*(12):5463-7.

First generation DNA sequencing







5’ TCGCTCCATGCTTACCTCGATCCG 3’

5’ TCGCTCCATGCTTACCTCG**A**

5’ TCGCTCCATGCTT**A**

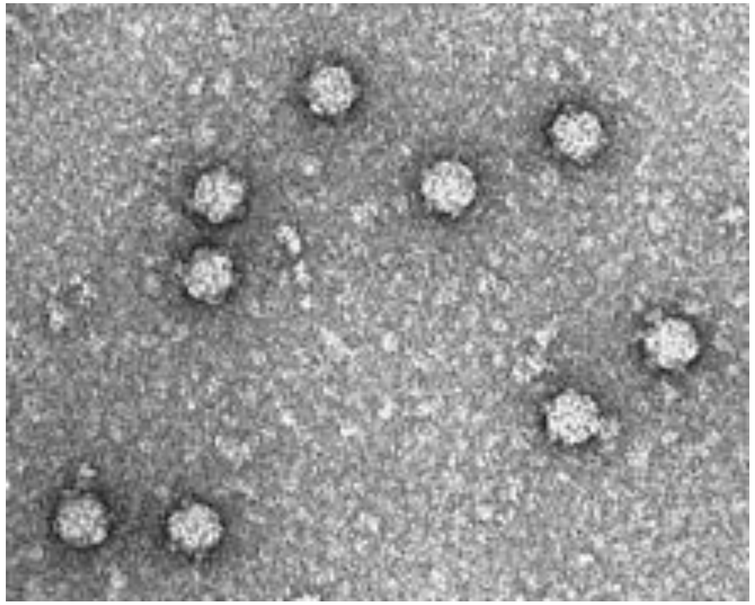
5’ TCGCTCC**A**

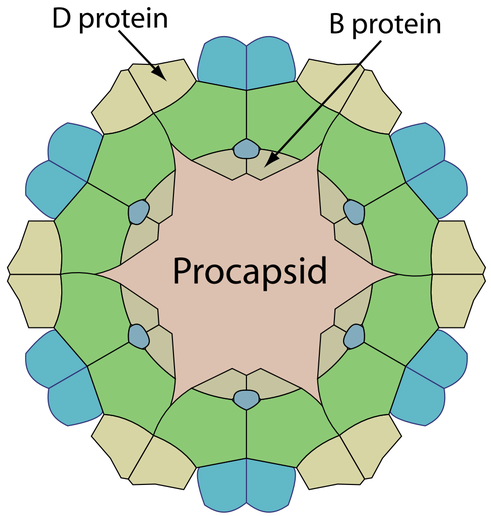
3’ AGCGAGGTACGAATGGAGCTAGGC 5’

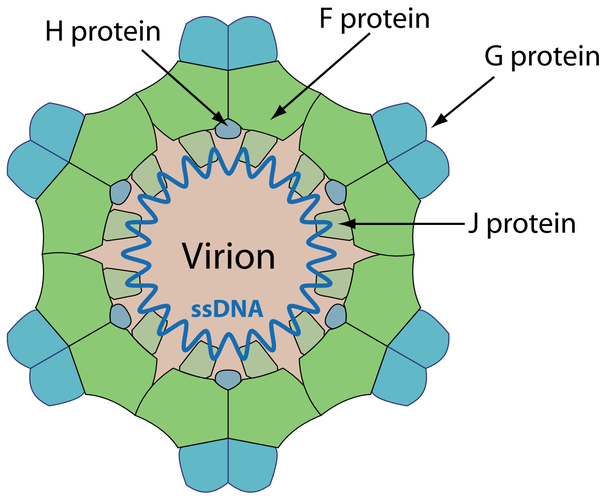
Proc. Natl. Acad. Sci., 1977, *74*(12):5463-7.

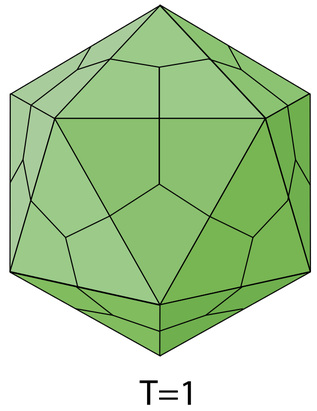
First generation DNA sequencing

Escherichia virus ΦX174

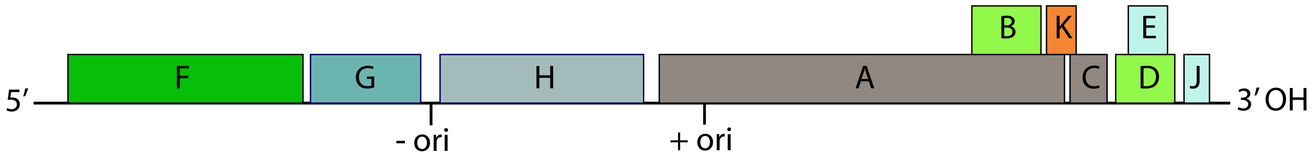






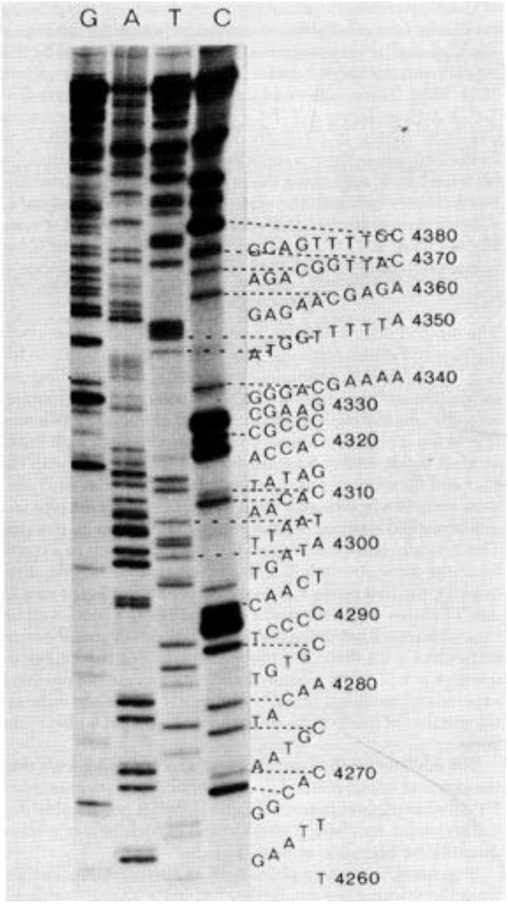


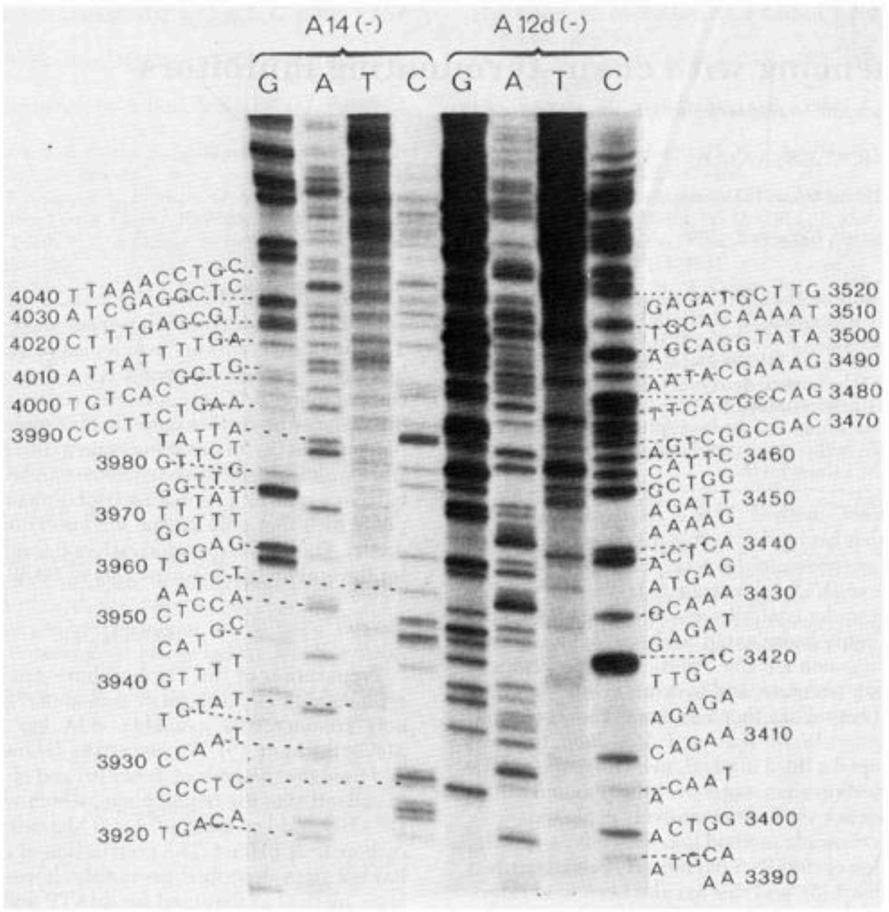
ΦX174 genome ~5500 nt



Proc. Natl. Acad. Sci., 1977, *74*(12):5463-7.

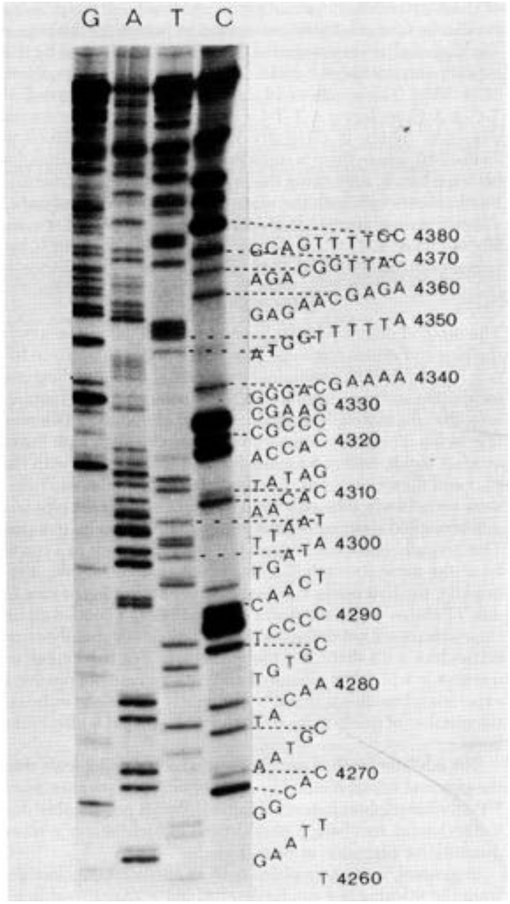
First generation DNA sequencing

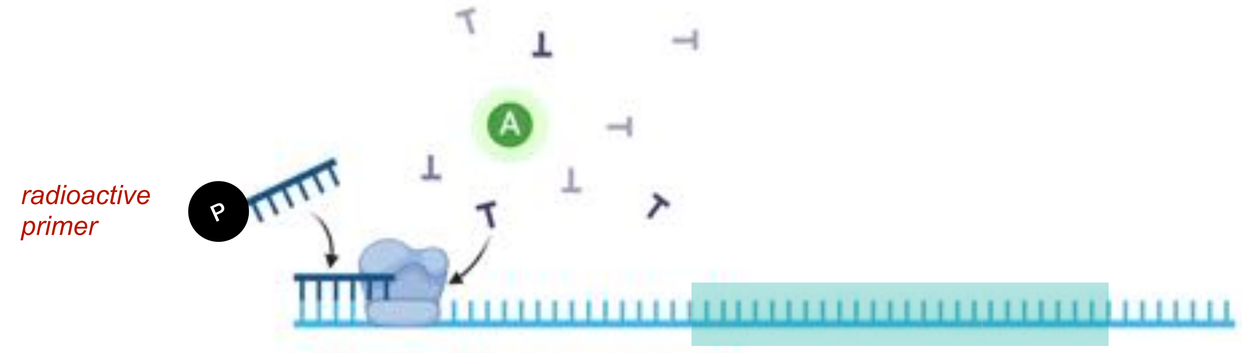




Proc. Natl. Acad. Sci., 1977, *74*(12):5463-7.

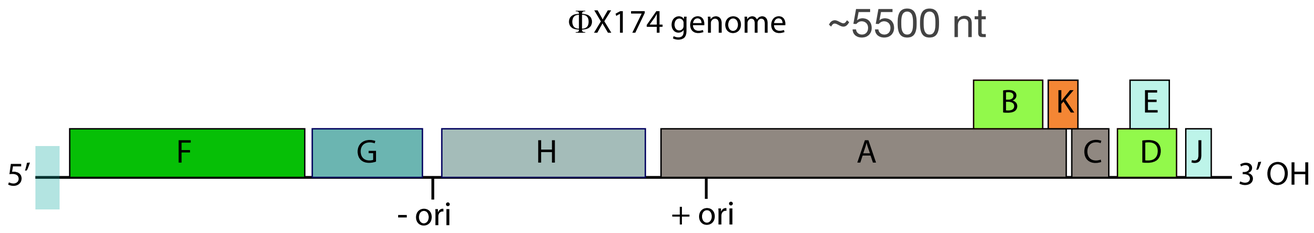
First generation DNA sequencing





+ 80 to about +160 is the "green zone"

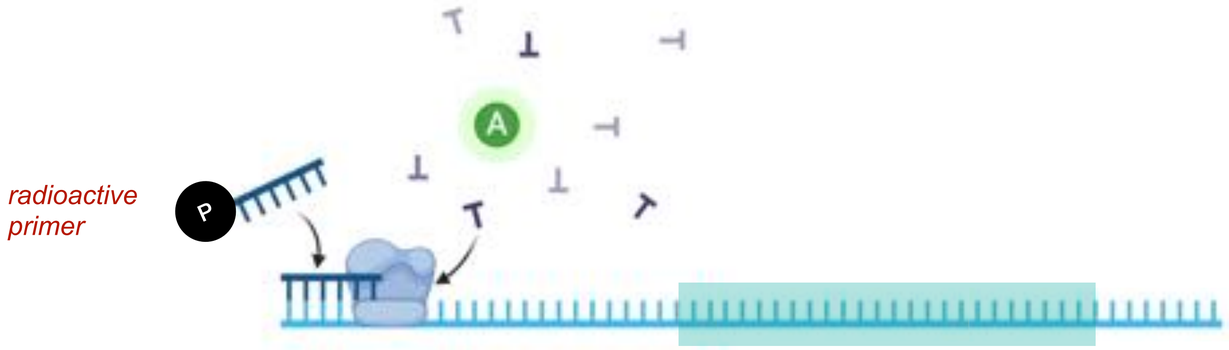
4 x ~60 rxns = 240 times



Proc. Natl. Acad. Sci., 1977, *74*(12):5463-7.

First generation DNA sequencing - improvements

Leroy Hood

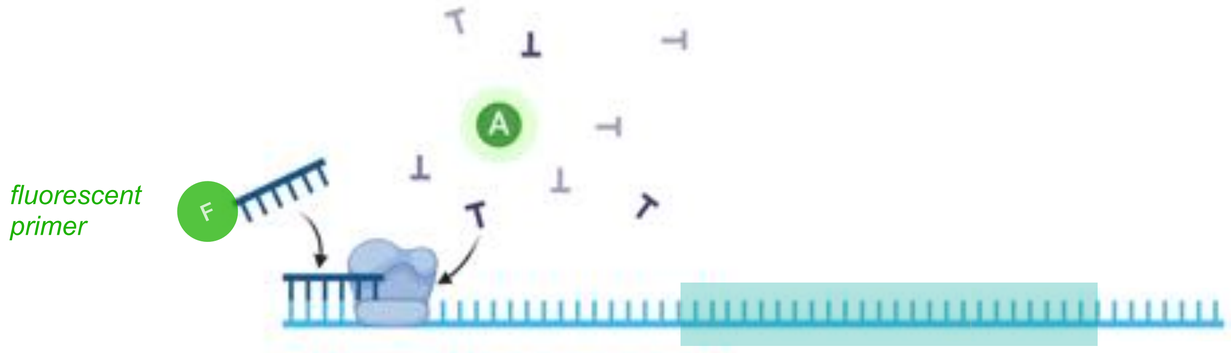


+ 80 to about +160 is the “green zone”

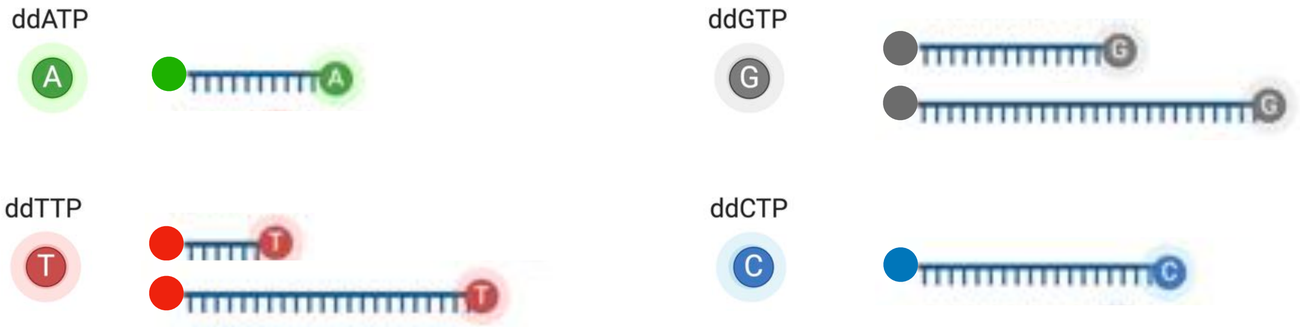
Nature, 1986, 321, 674–679

First generation DNA sequencing - improvements

Leroy Hood



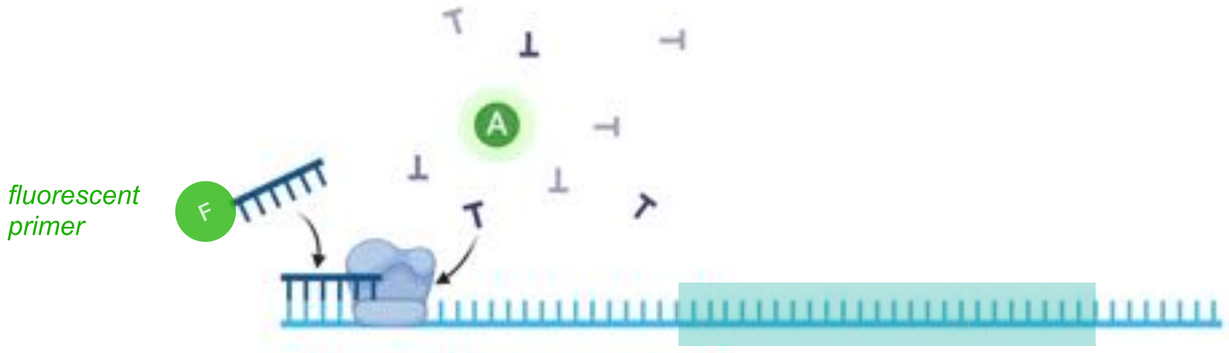
+ 80 to about +160 is the “green zone”



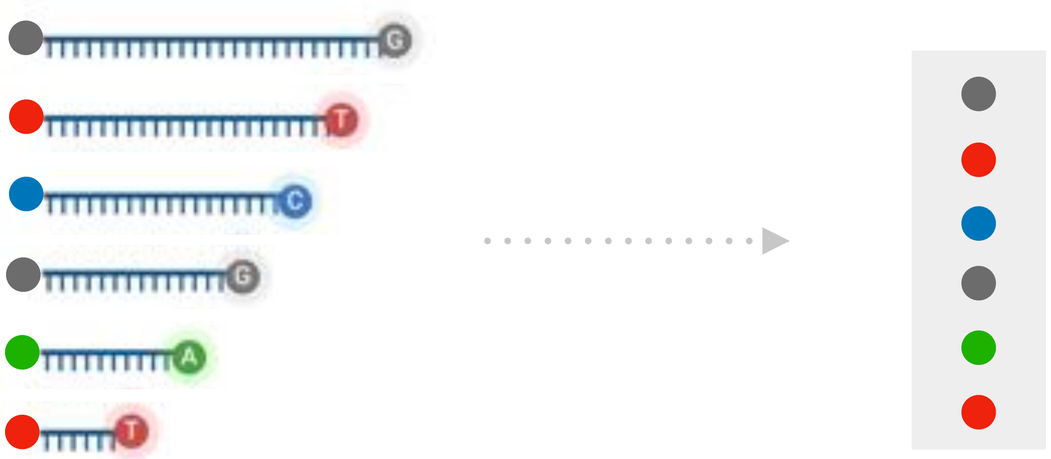
Nature, 1986, 321, 674–679

First generation DNA sequencing - improvements

Leroy Hood



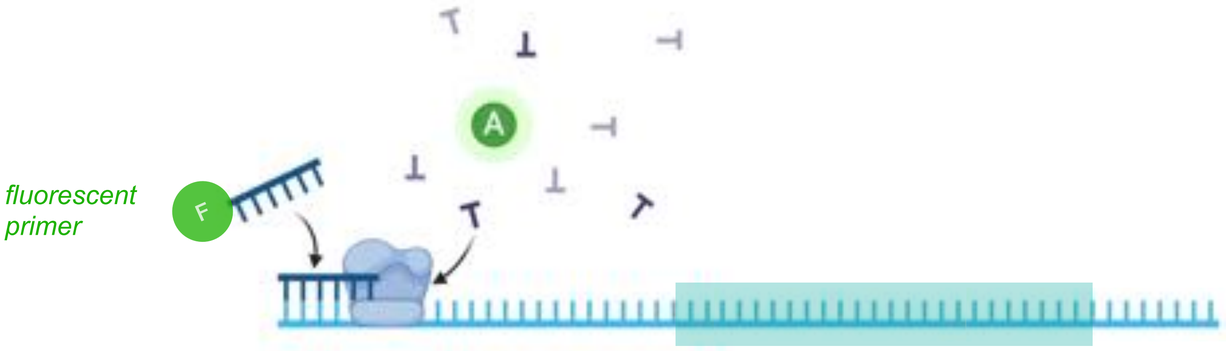
+ 80 to about +160 is the “green zone”



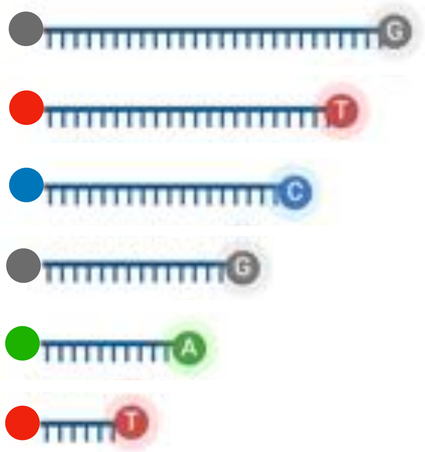
Nature, 1986, 321, 674–679

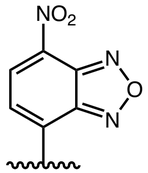
First generation DNA sequencing - improvements

Leroy Hood

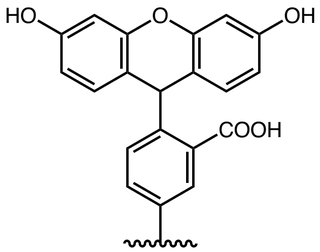


+ 80 to about +160 is the "green zone"

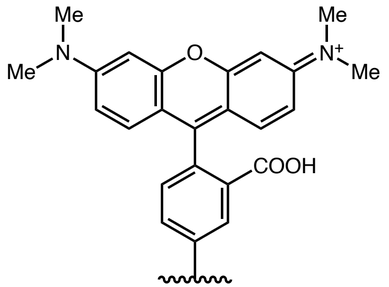




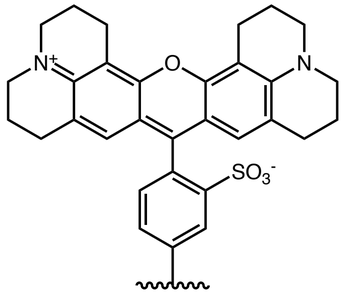
NBD



FITC



TMR



Texas Red

Nature, 1986, 321, 674–679

First generation DNA sequencing - improvements

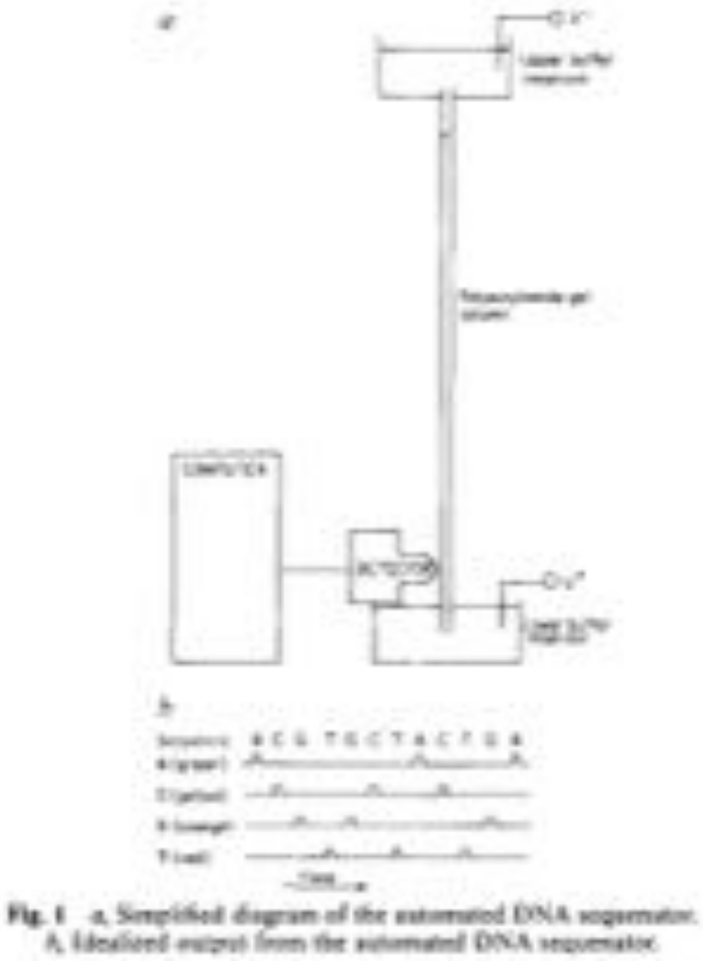
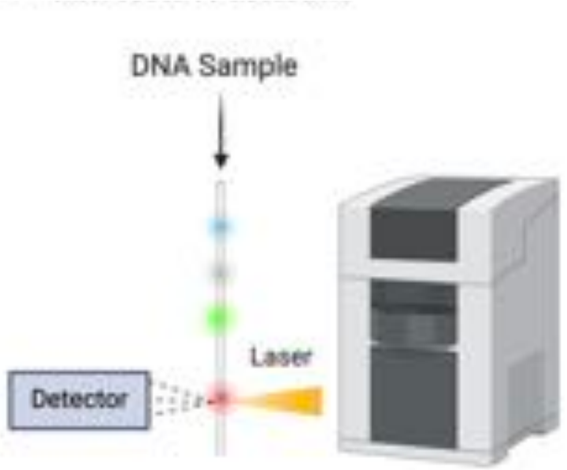
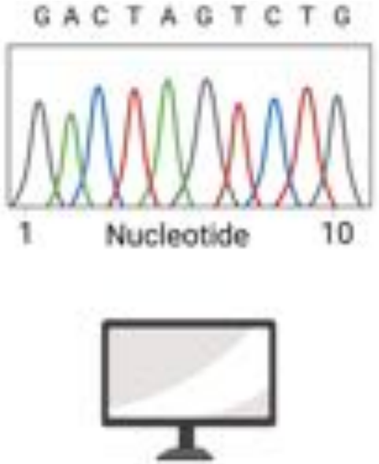


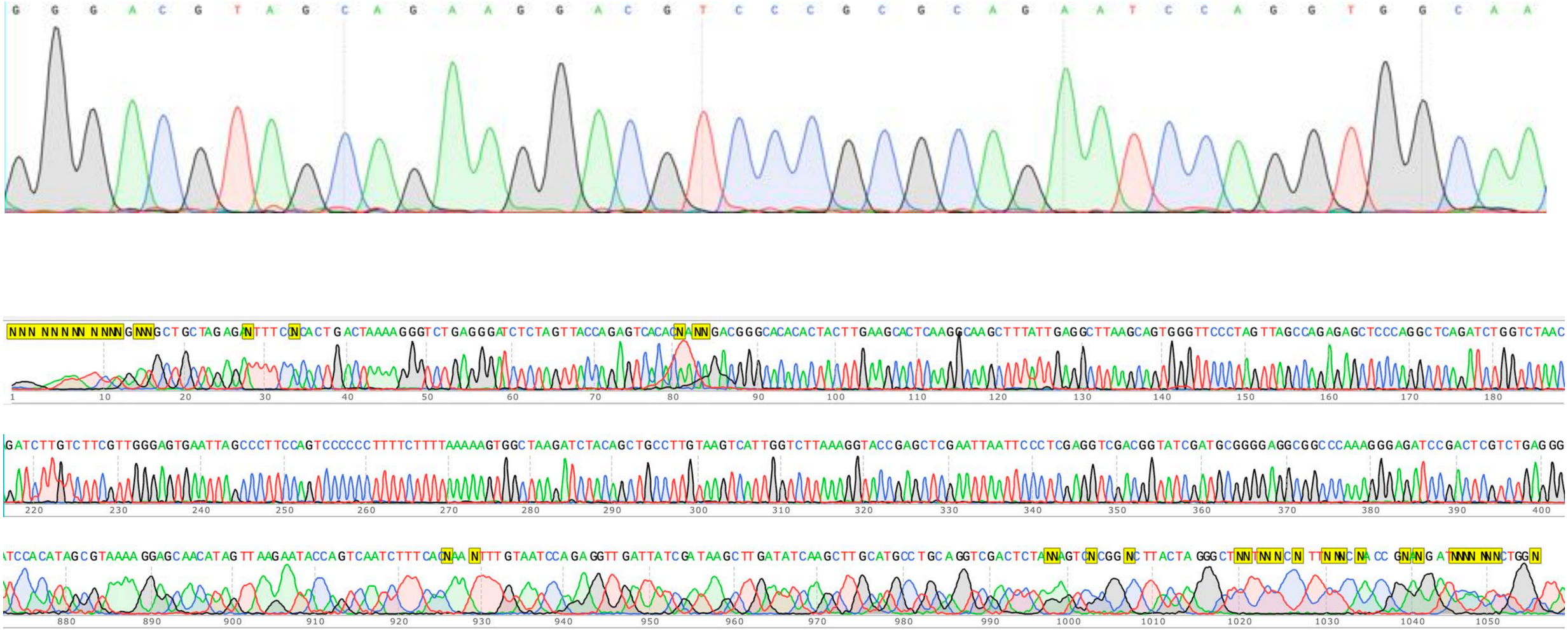
Fig. 1 a, Simplified diagram of the automated DNA sequencator. A, Idealized output from the automated DNA sequencator.



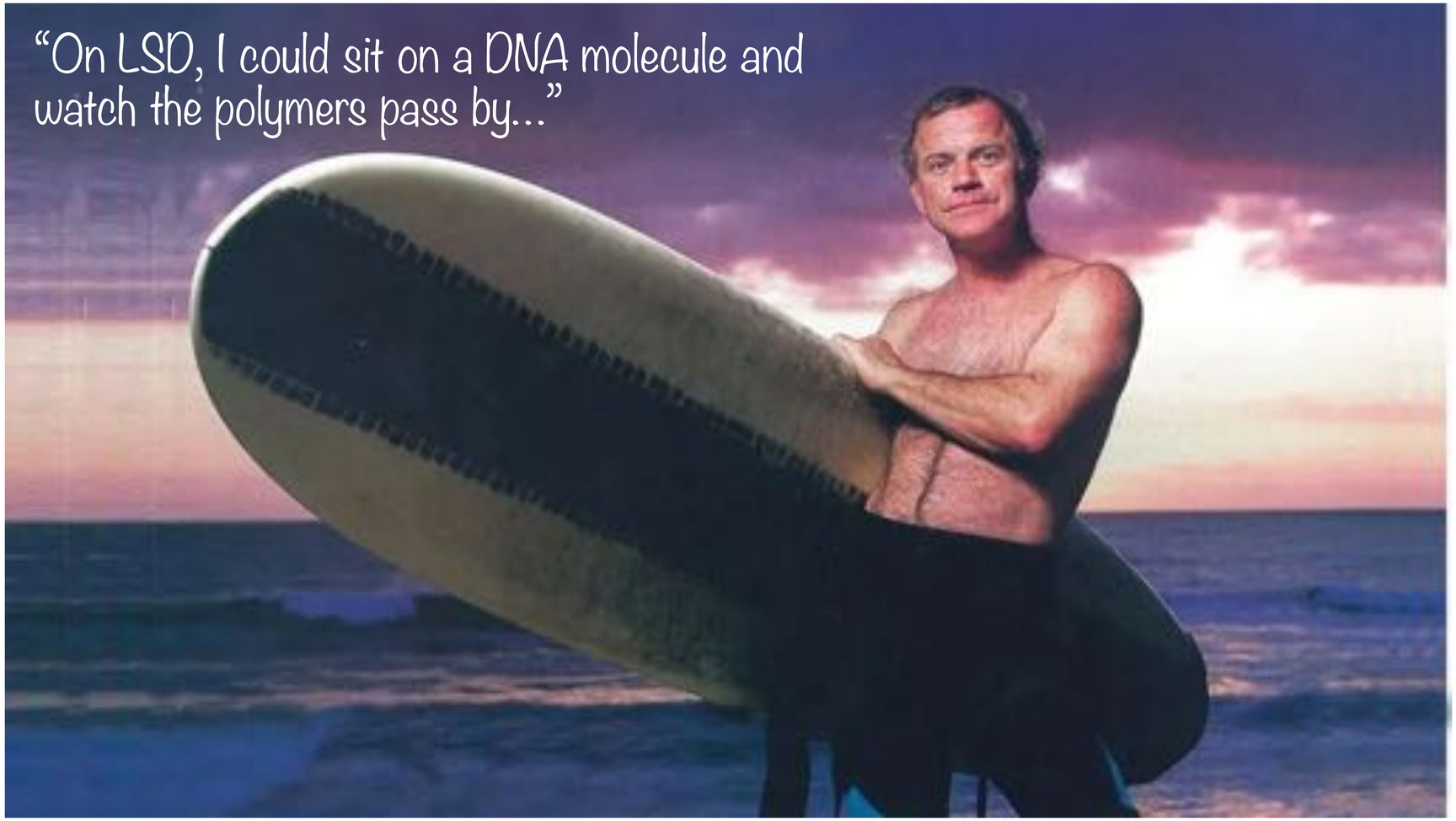


Nature, 1986, 321, 674–679

First generation DNA sequencing - improvements



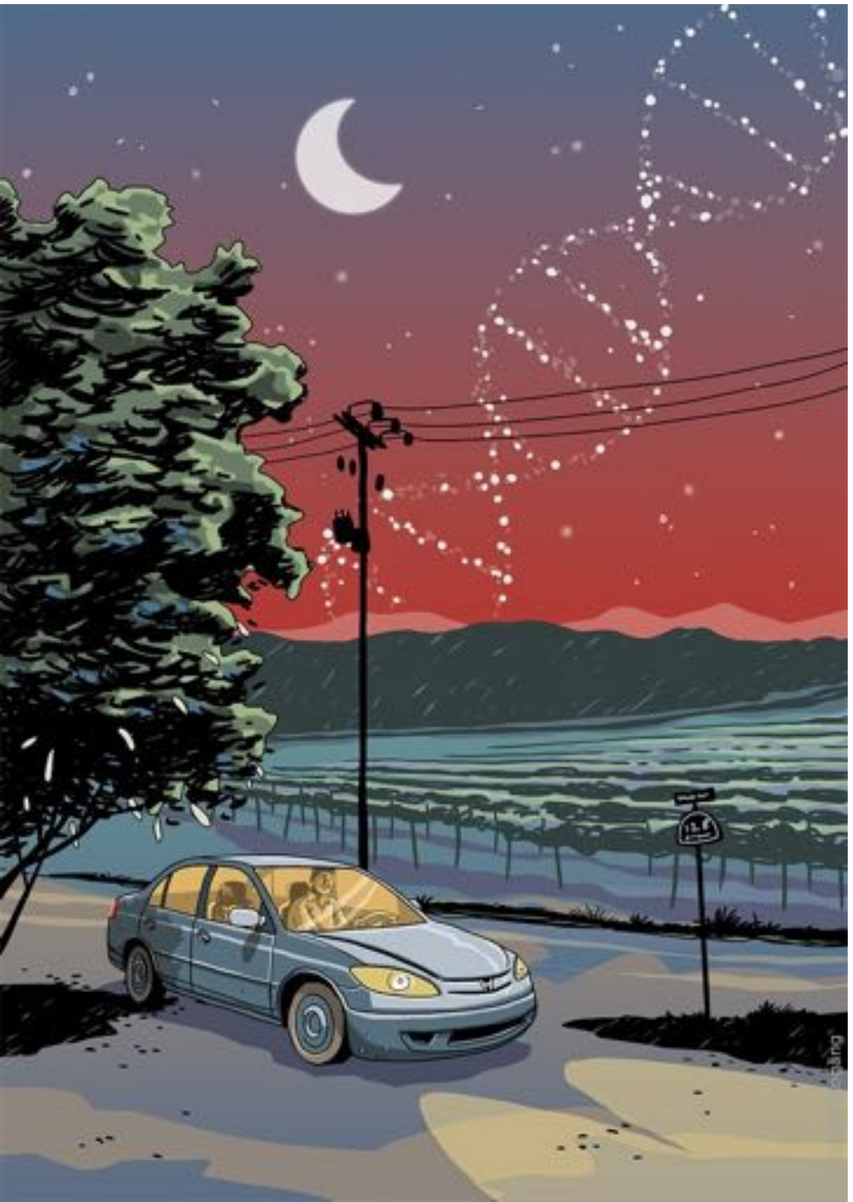
The Deoxyribonucleic Acid Trip

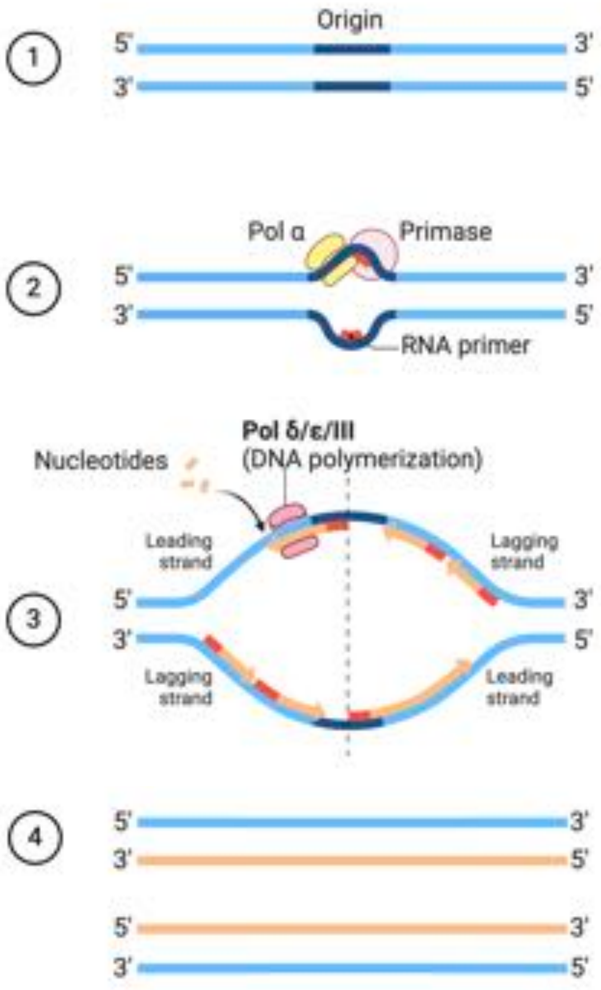


Kary Mullis

"What if I had not taken LSD ever; would I have still invented PCR? I don't know. I doubt it. I seriously doubt it."

The Deoxyribonucleic Acid Trip



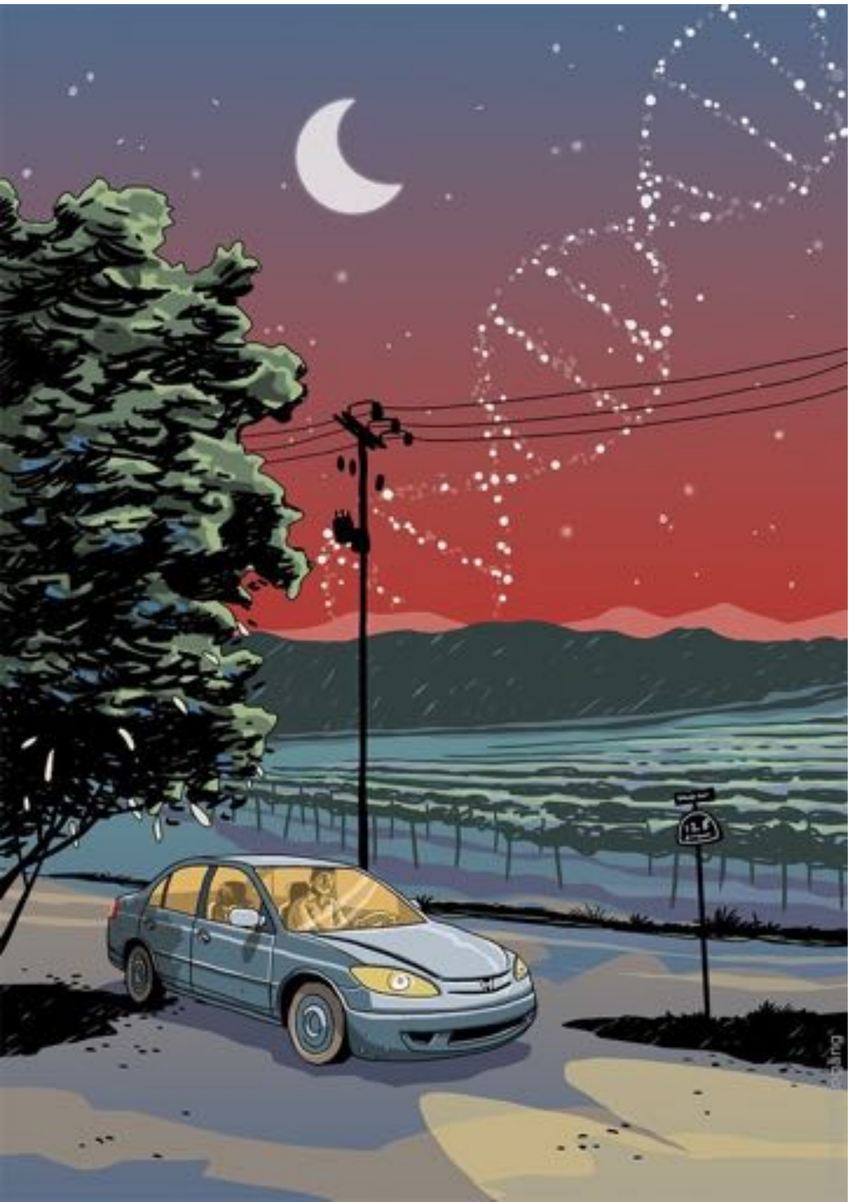


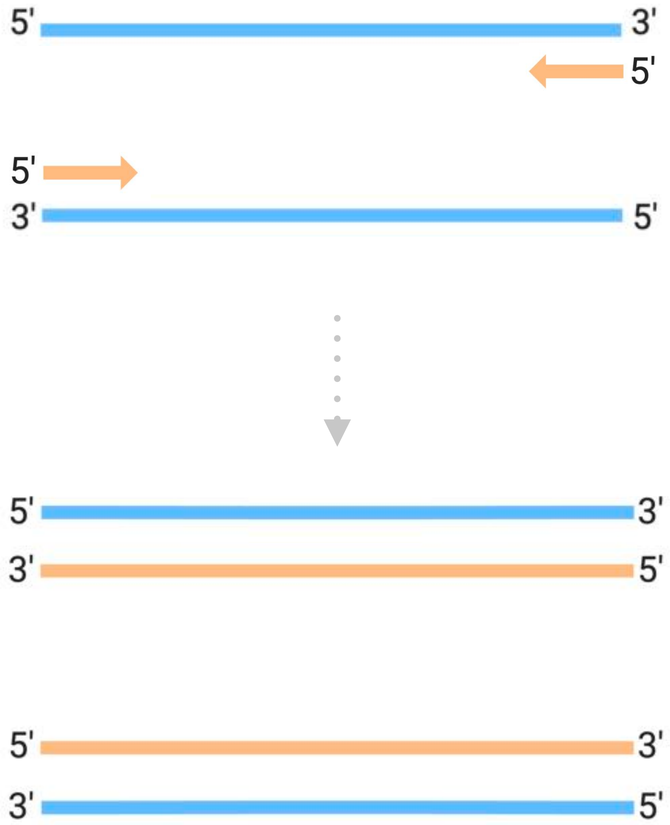
semi-conservative replication

Kary Mullis

"I was "functionally sober", but in a different state of mind entirely."

The Deoxyribonucleic Acid Trip

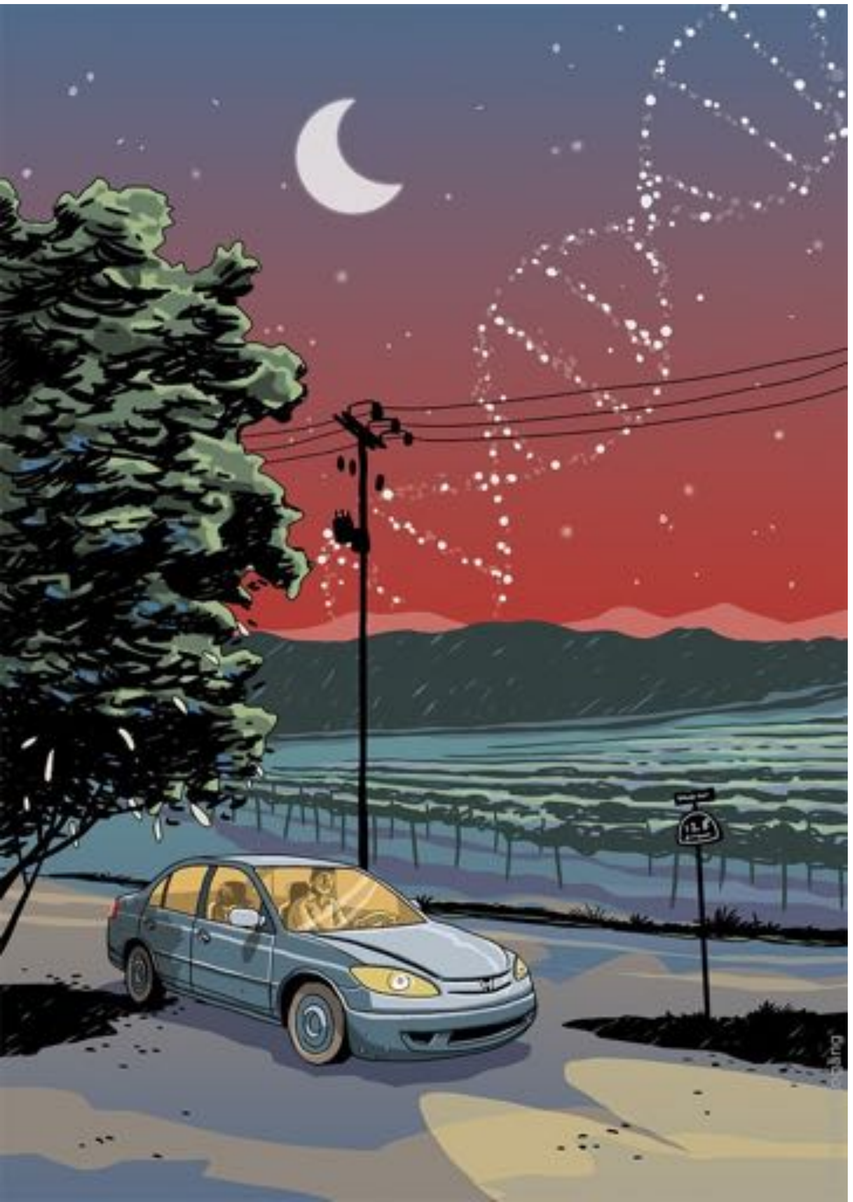


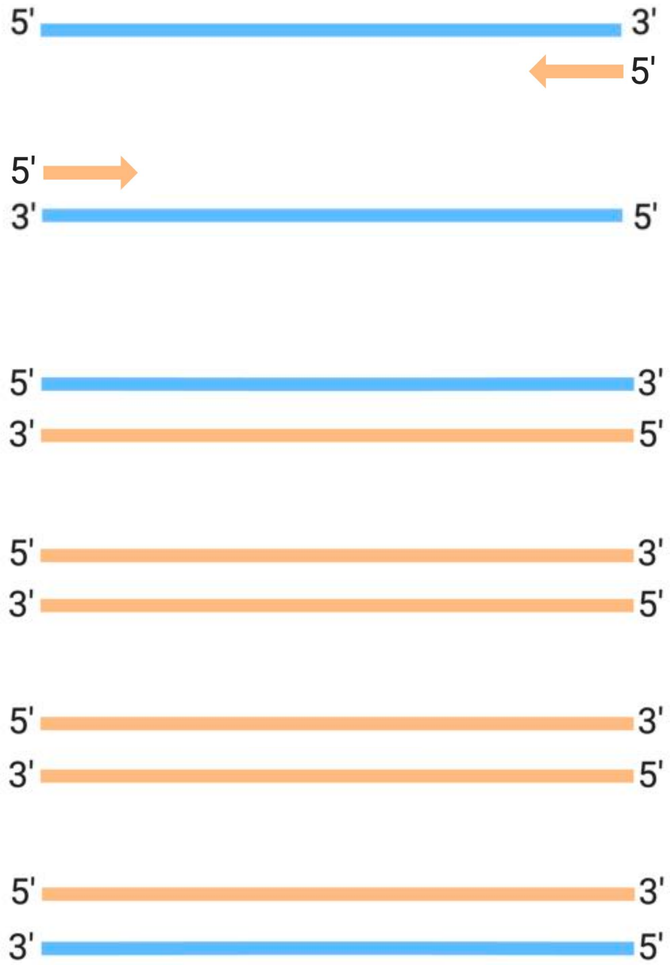


Kary Mullis

“I was “functionally sober”, but in a different state of mind entirely.”

The Deoxyribonucleic Acid Trip



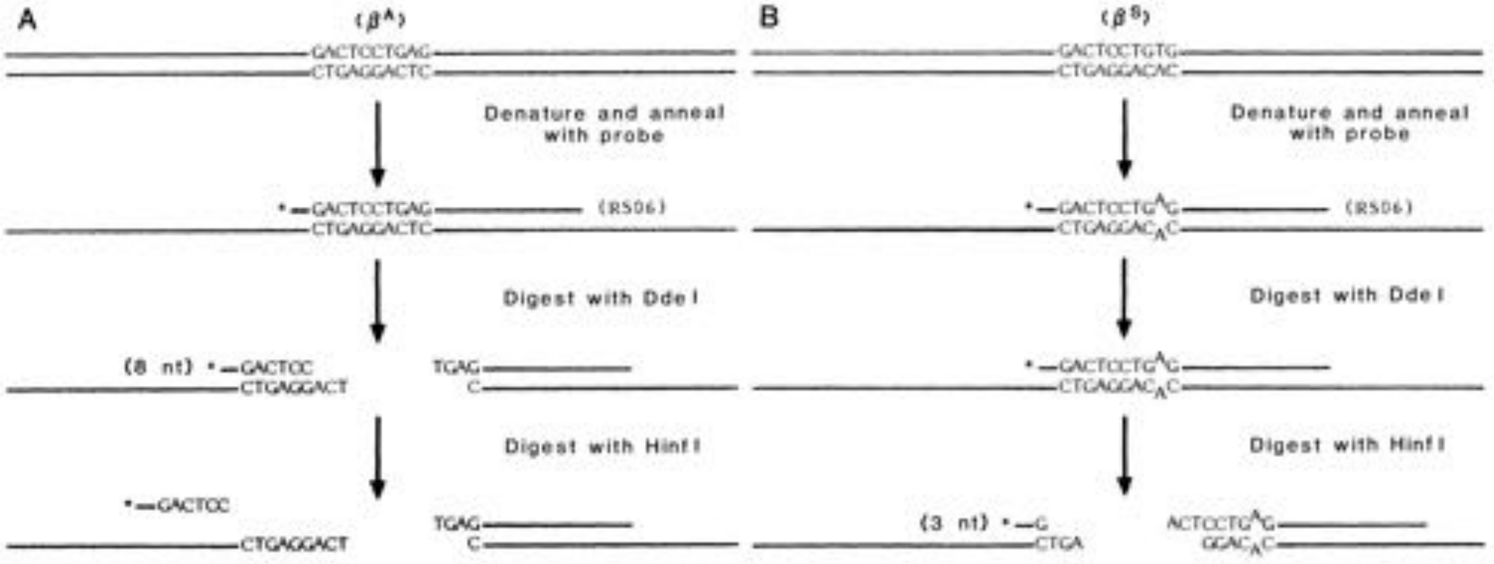


Kary Mullis

“I was “functionally sober”, but in a different state of mind entirely.”

The Deoxyribonucleic Acid Trip

Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia



Science, 1985, *230*(4732),1350-4.

The Deoxyribonucleic Acid Trip

Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase

RANDALL K. SAIKI, DAVID H. GELFAND, SUSANNE STOFFEL, STEPHEN J. SCHARF, RUSSELL HIGUCHI, GLENN T. HORN, KARY B. MULLIS,* HENRY A. ERLICH

A thermostable DNA polymerase was used in an in vitro DNA amplification procedure, the polymerase chain reaction. The enzyme, isolated from *Thermus aquaticus*, greatly simplifies the procedure and, by enabling the amplification reaction to be performed at higher temperatures, significantly improves the specificity, yield, sensitivity, and length of products that can be amplified. Single-copy genomic sequences were amplified by a factor of more than 10 million with very high specificity, and DNA segments up to 2000 base pairs were readily amplified. In addition, the method was used to amplify and detect a target DNA molecule present only once in a sample of 10^5 cells.

Science, 1988, *239*(4839),487-491.

The Human Genome Project

GENOME SEQUENCING WORKSHOP

MARCH 3 & 4, 1986

SANTA FE, NEW MEXICO

HOST

LIFE SCIENCES DIVISION

LOS ALAMOS NATIONAL LABORATORY

NATURE VOL. 326 2 APRIL 1987

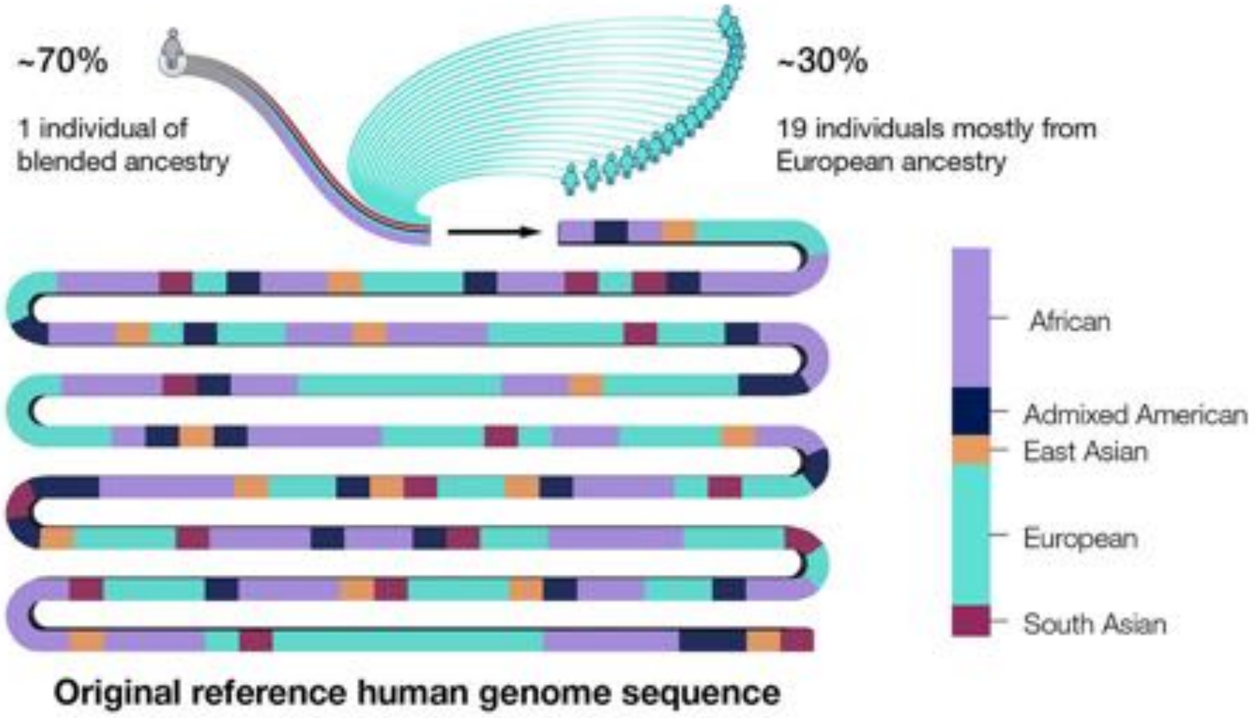
Human genome sequencing plan wins unanimous approval in US

Gaithersburg, Maryland

THE project to map and sequence the human genome is now a big step nearer reality. At a meeting here last week, the Health and Environmental Research Advisory Committee (HERAC) of the Department of Energy (DoE) collectively fully fledged mapping and sequencing project. Few in Congress will be enthusiastic about a new initiative that will cost more than $1,000 million, although this is less upsetting than the estimate of $3,000 million initially quoted for the project (see *Nature* **321**, 371; 1986).

NIH National Human Genome Research Institute

The Human Genome Project



WANTED

20 Volunteers

to participate in the

Human Genome Project

a very large international scientific research effort.

The goal is to decode the human hereditary information (*human blueprint*) that determines all individual traits inherited from parents. The outcome of the project will have tremendous impact on future progress of medical science and lead to improved diagnosis and treatment of hereditary diseases.

Volunteers will receive information about the project from the Clinical Genetics Service at Roswell Park, and sign a consent form before participating.

No personal information will be maintained or transferred.

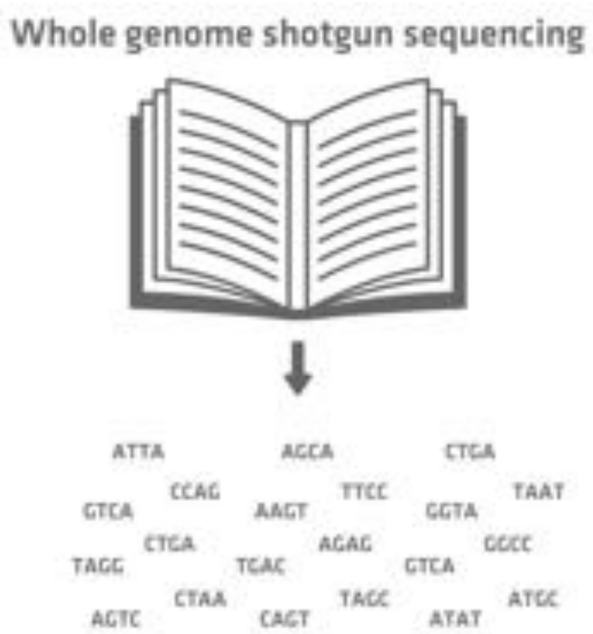
Volunteers will provide a one-time donation of a small blood specimen. A small monetary reimbursement will be provided to the participants for their time and effort.

Individuals must be at least 18 years of age.
Persons who have undergone chemotherapy are not eligible.

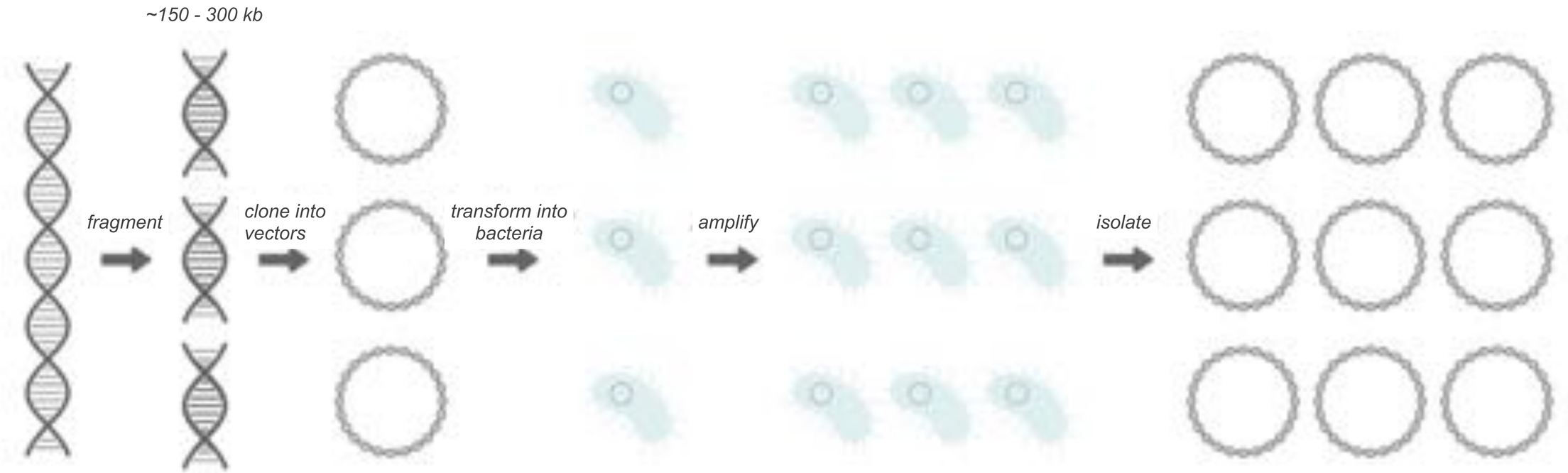
ROSWELL PARK CANCER INSTITUTE

For more information please contact the
Clinical Genetics Service
845-5720 (9:00 am - 3:00 pm)
March 24 - 26, 1997

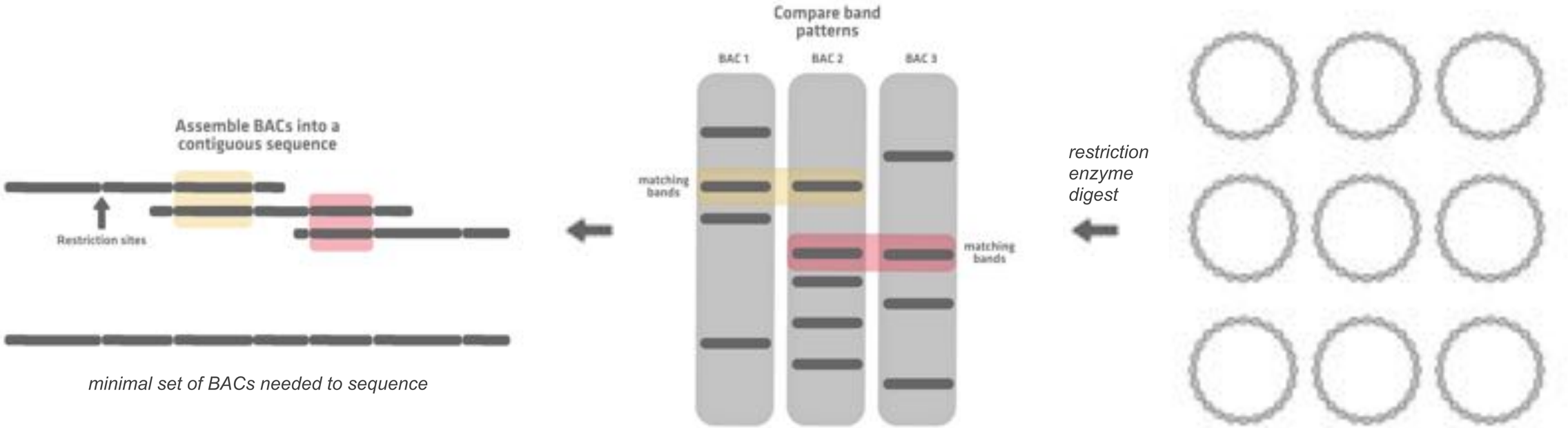
The Shotgun Approach



The Shotgun Approach



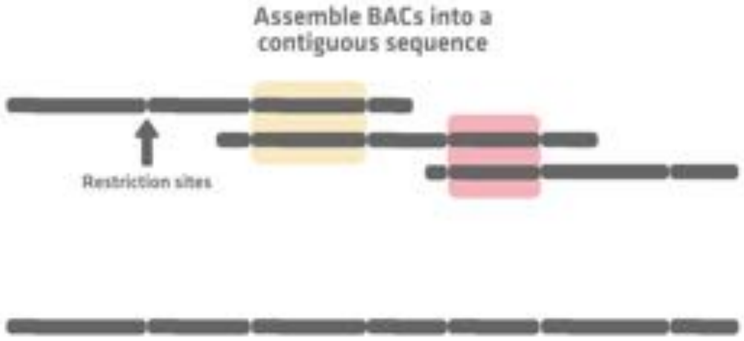
The Shotgun Approach



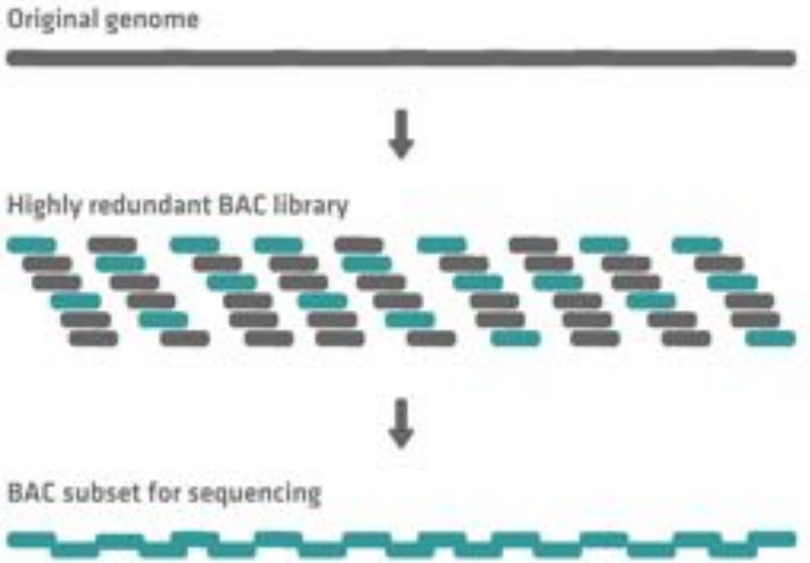
minimal set of BACs needed to sequence

restriction enzyme digest

The Shotgun Approach

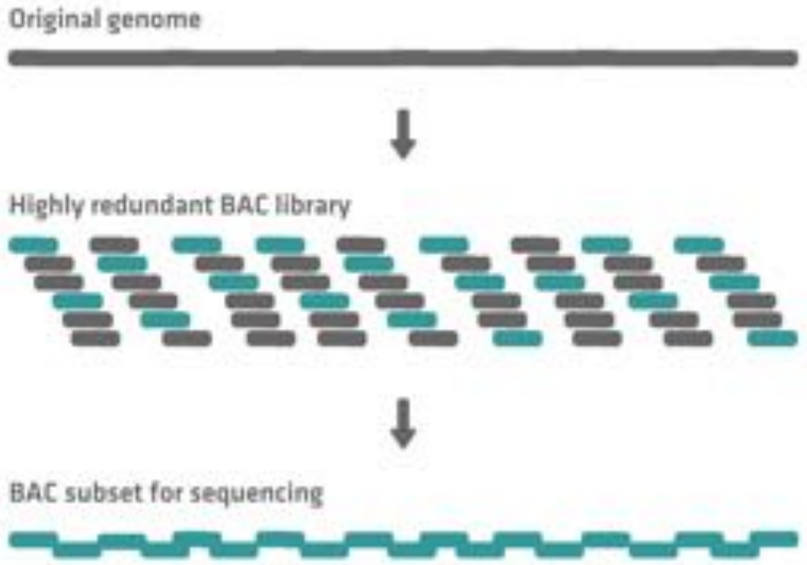


minimal set of BACs needed to sequence



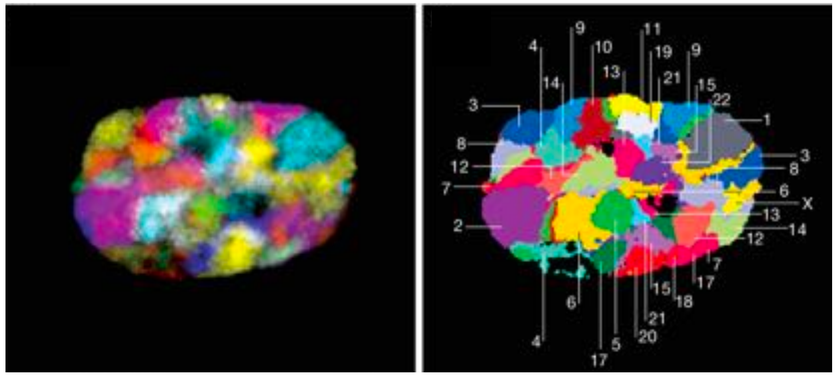
minimal set of BACs needed to sequence

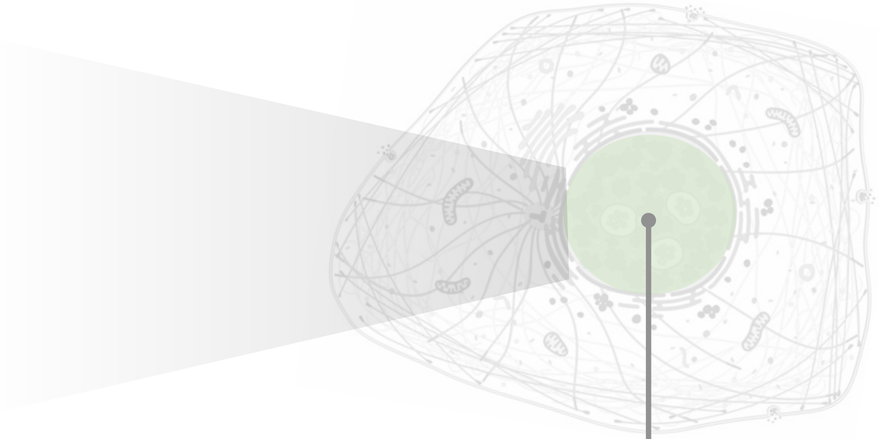
The Shotgun Approach

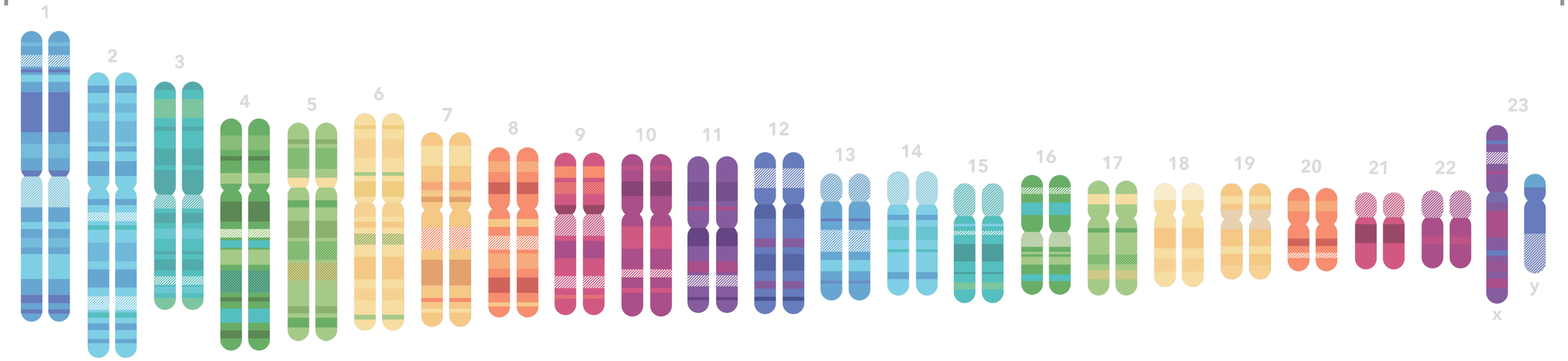


minimal set of BACs needed to sequence

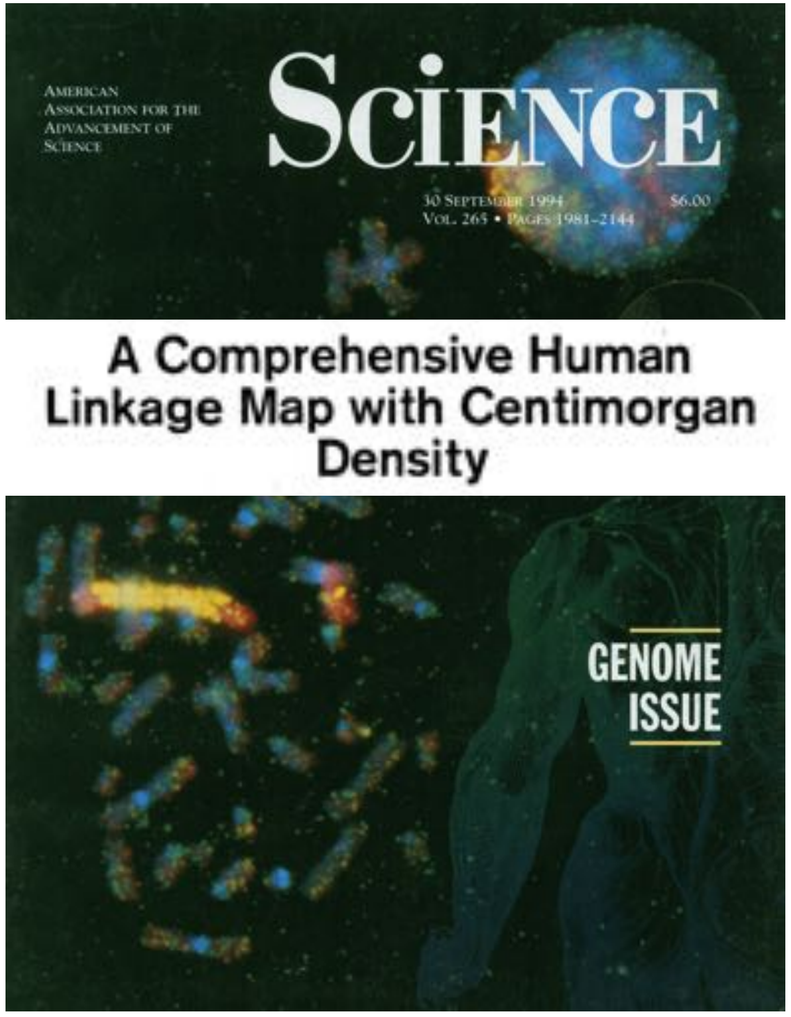
The “You” You Are

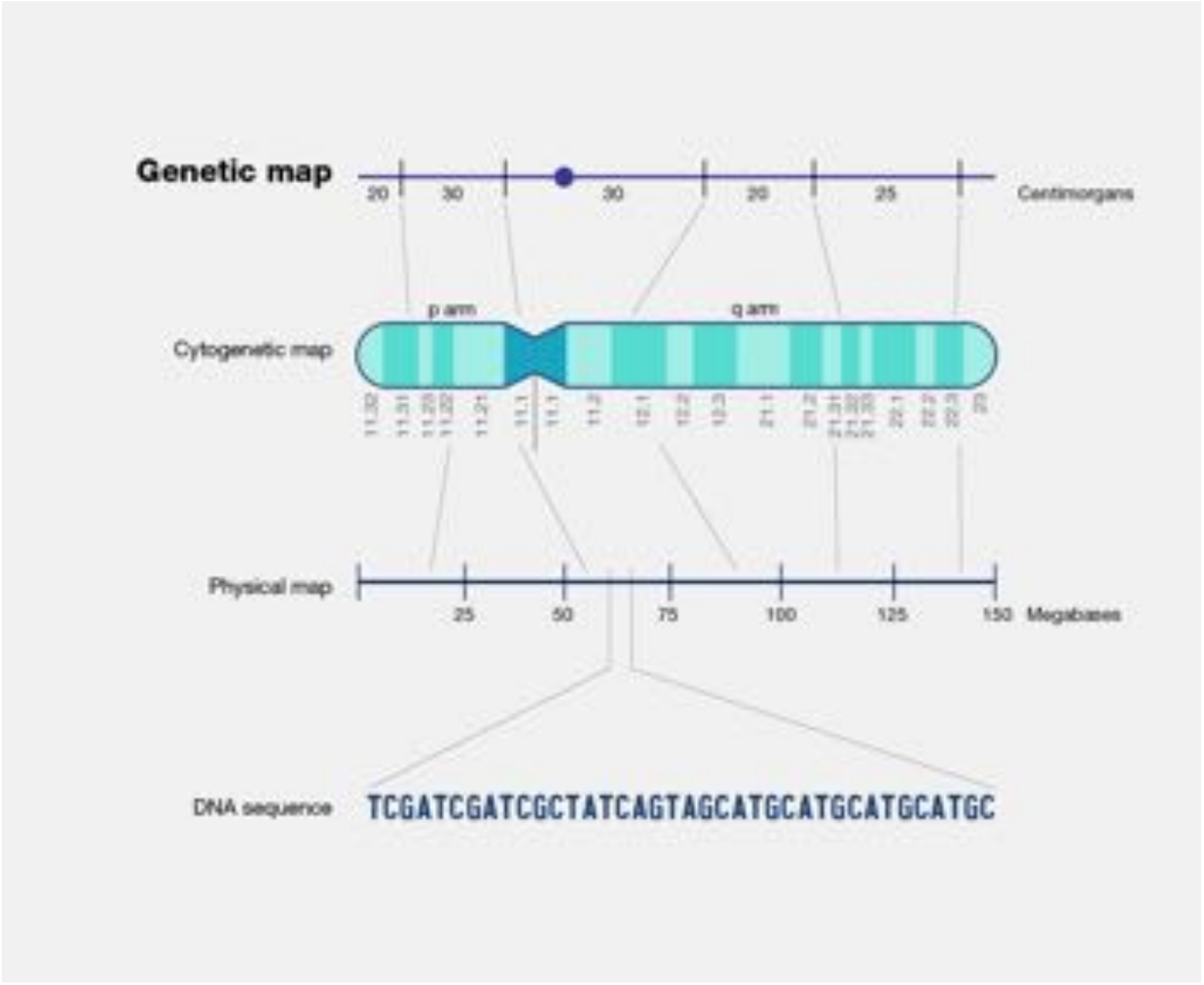






1994 - Basic Mapping of Chromosome Architecture

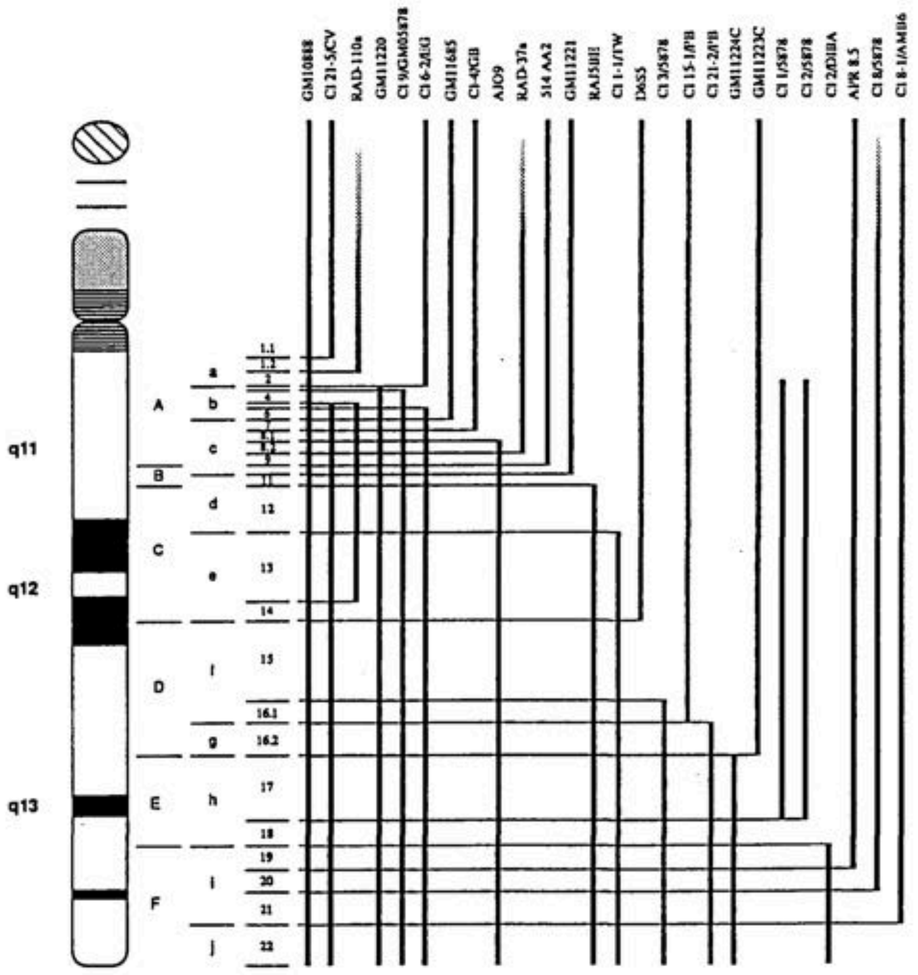




Science, 1994, 265,2049-2054.

1994 - Basic Mapping of Chromosome Architecture

A Comprehensive Human Linkage Map with Centimorgan Density



Science, 1994, 265,2049-2054.

1995 - 2003 Competition, Automation, and a (Rough) Final Draft

‘It’s a G’: the one-billionth nucleotide

NATURE | VOL 402 | 25 NOVEMBER 1999 |

The DNA sequence of human chromosome 22

Nature, 1999, 402, 489–495

The DNA sequence of human chromosome 21

Nature, 2000, 405, 311–319

1995 - 2003 Competition, Automation, and a (Rough) Final Draft

‘It’s a G’: the one-billionth nucleotide

NATURE | VOL 402 | 25 NOVEMBER 1999 |

The DNA sequence of human chromosome 22

Nature, 1999, 402, 489–495

CELERA
A PE Corporation Business

NIH

The DNA sequence of human chromosome 21

Nature, 2000, 405, 311–319

1995 - 2003 Competition, Automation, and a (Rough) Final Draft

CELERA
A PE Corporation Business

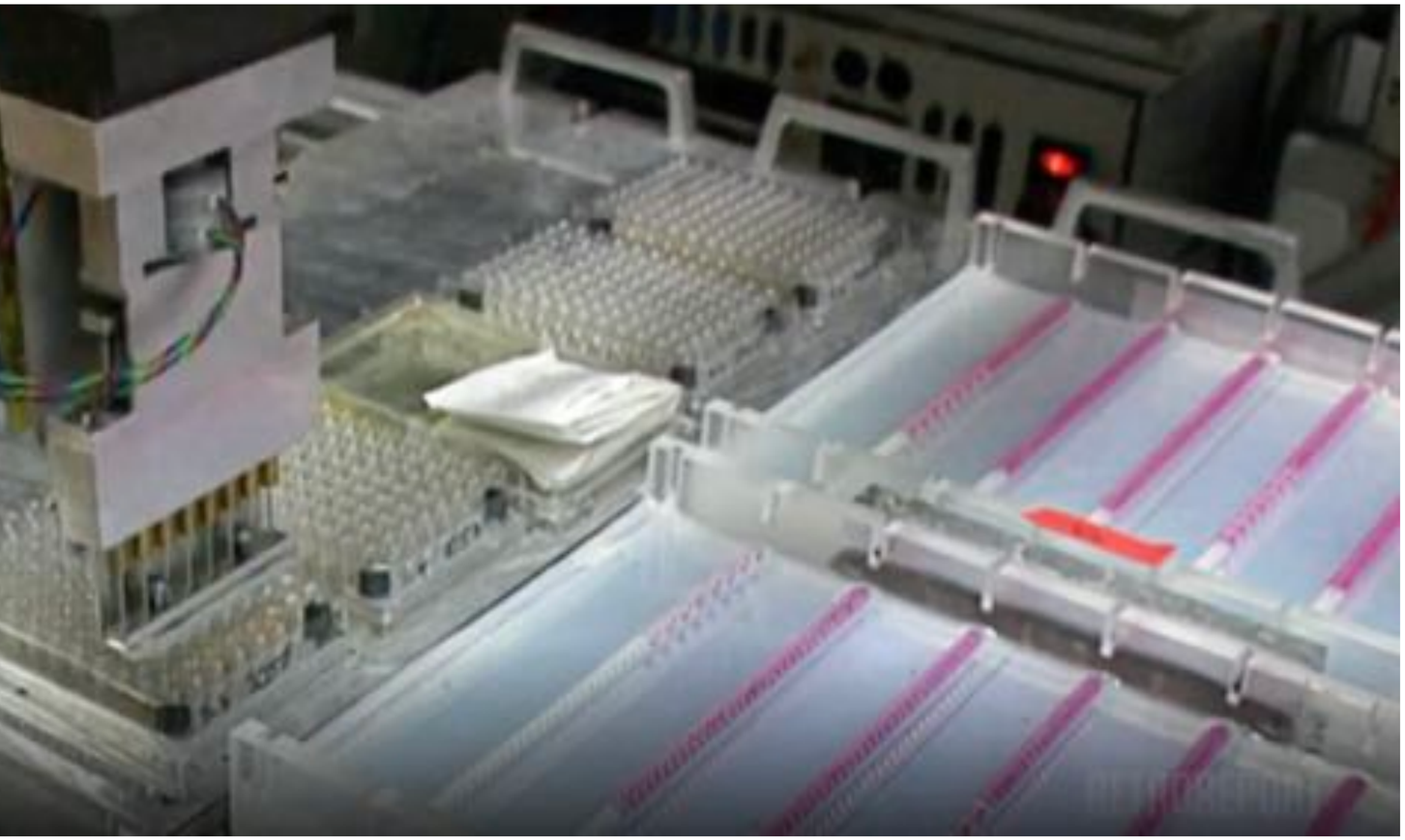
NIH

1995 - 2003 Competition, Automation, and a (Rough) Final Draft

CELERA
A PE Corporation Business

NIH

1995 - 2003 Competition, Automation, and a (Rough) Final Draft



CELERA
A PE Corporation Business

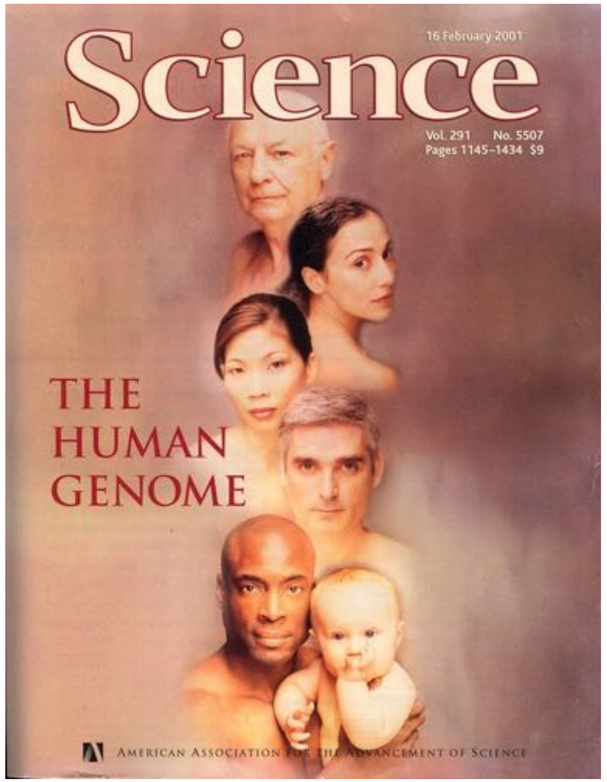
NIH

1995 - 2003 Competition, Automation, and a (Rough) Final Draft

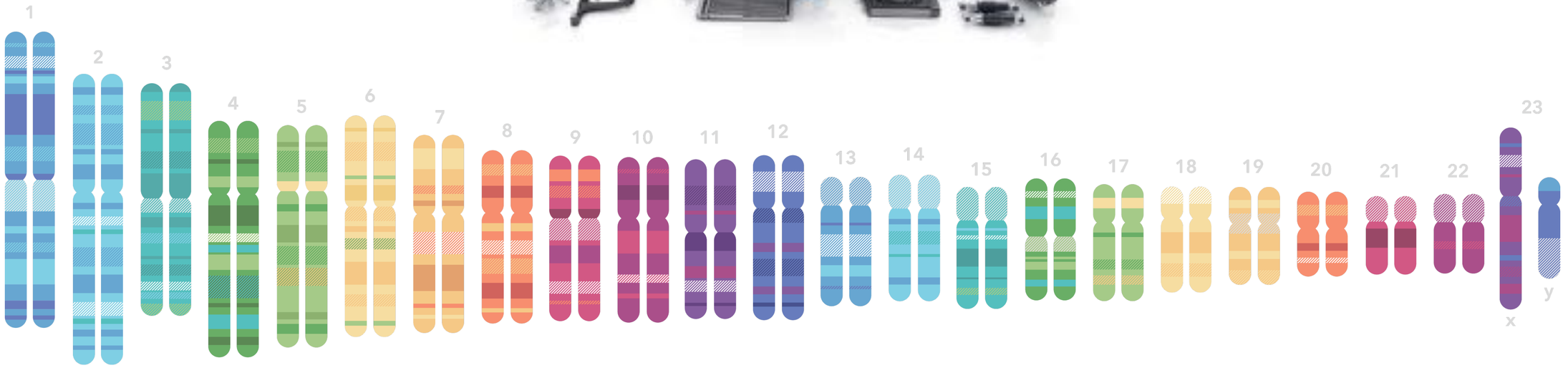
CELERA
A PE Corporation Business

NIH

1995 - 2003 Competition, Automation, and a (Rough) Final Draft

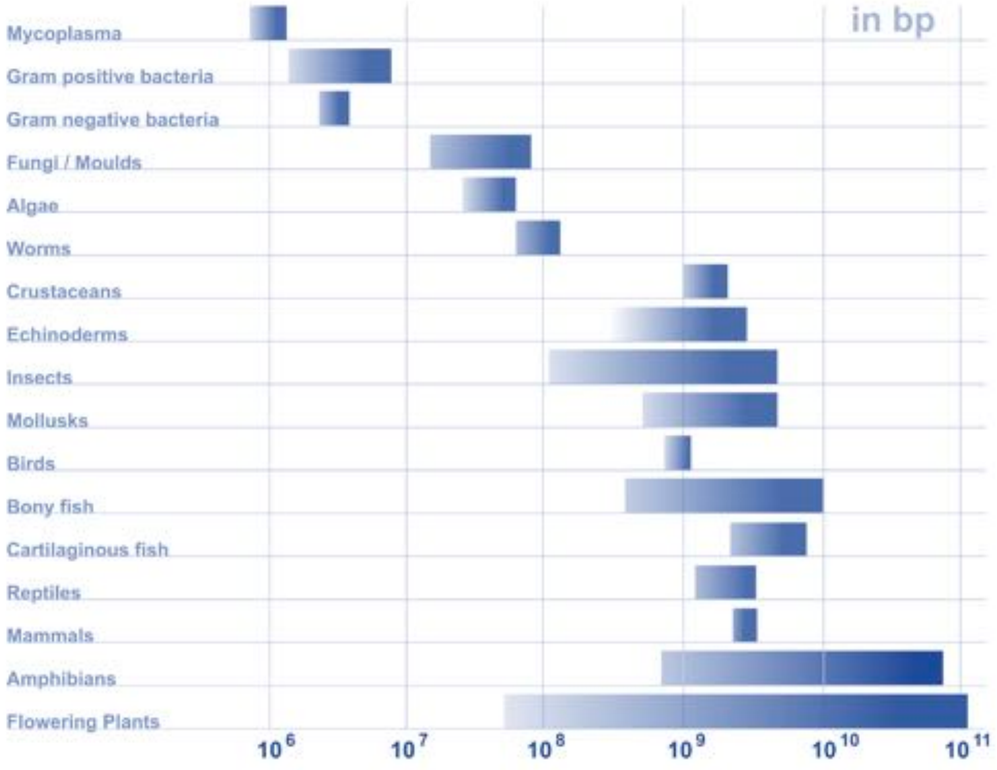


Surprise #1 - We don't have that many "genes"



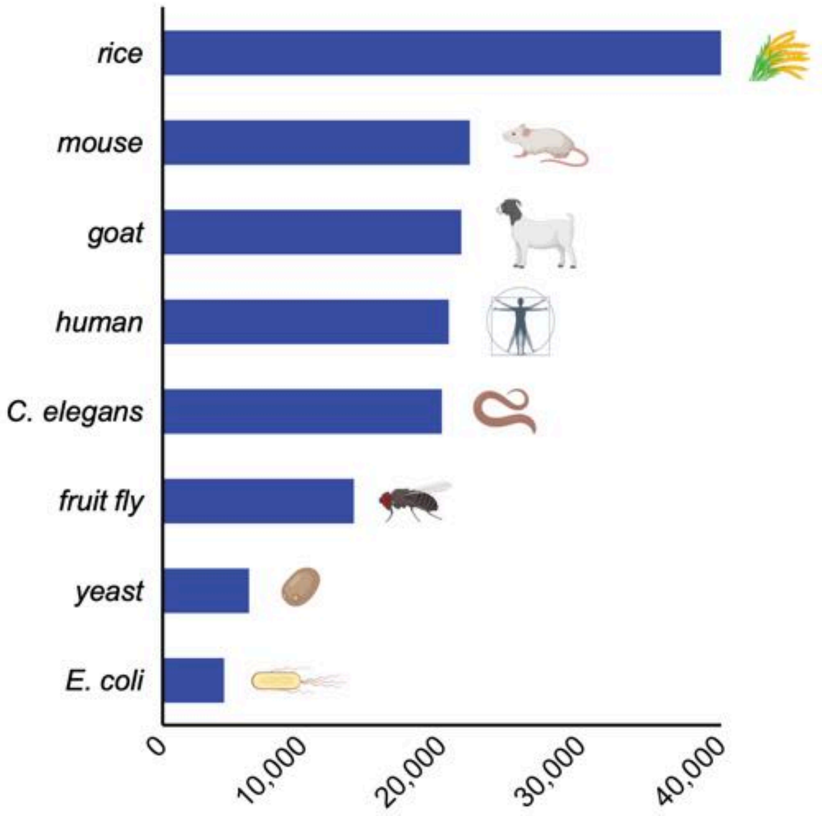
Surprise #1 - We don't have that many "genes"

genome size



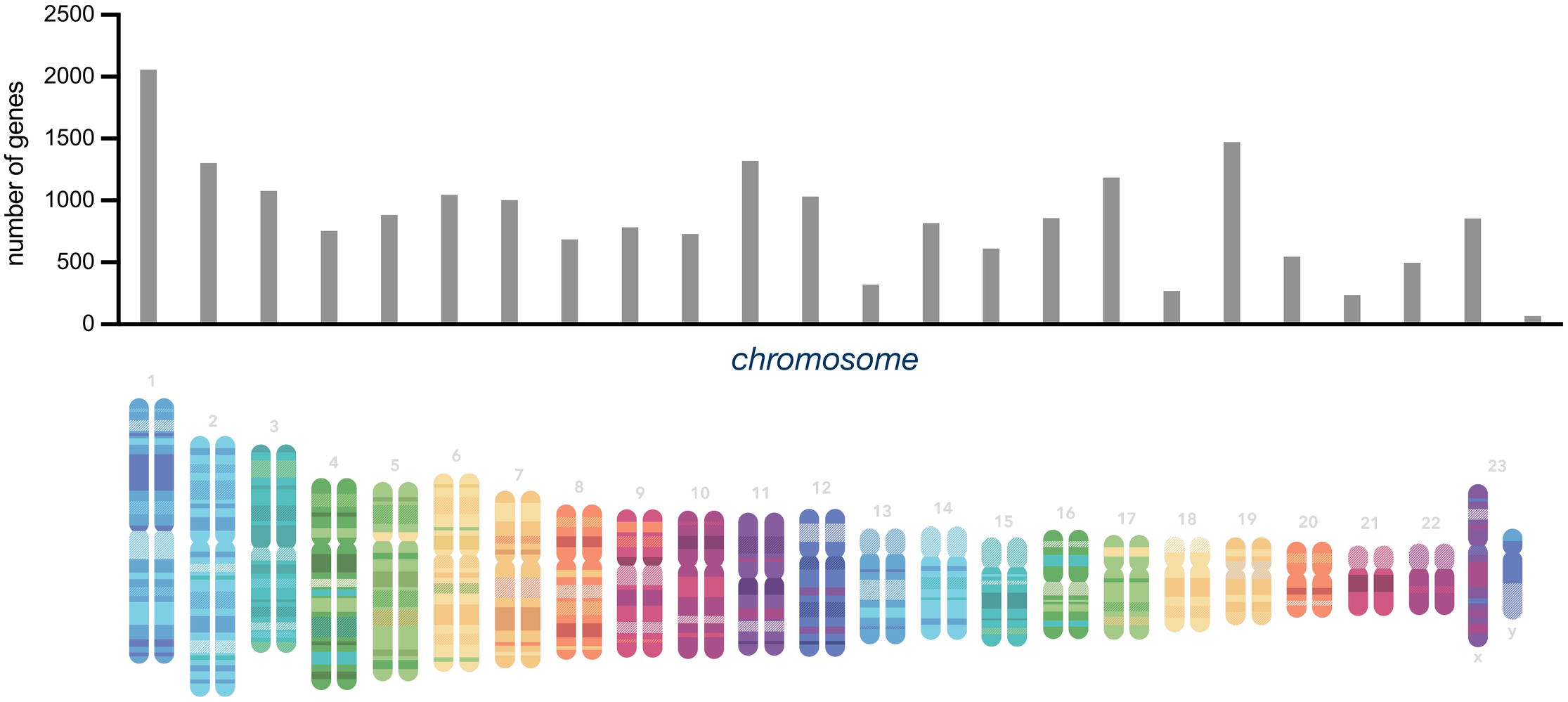
basepairs

genome complexity

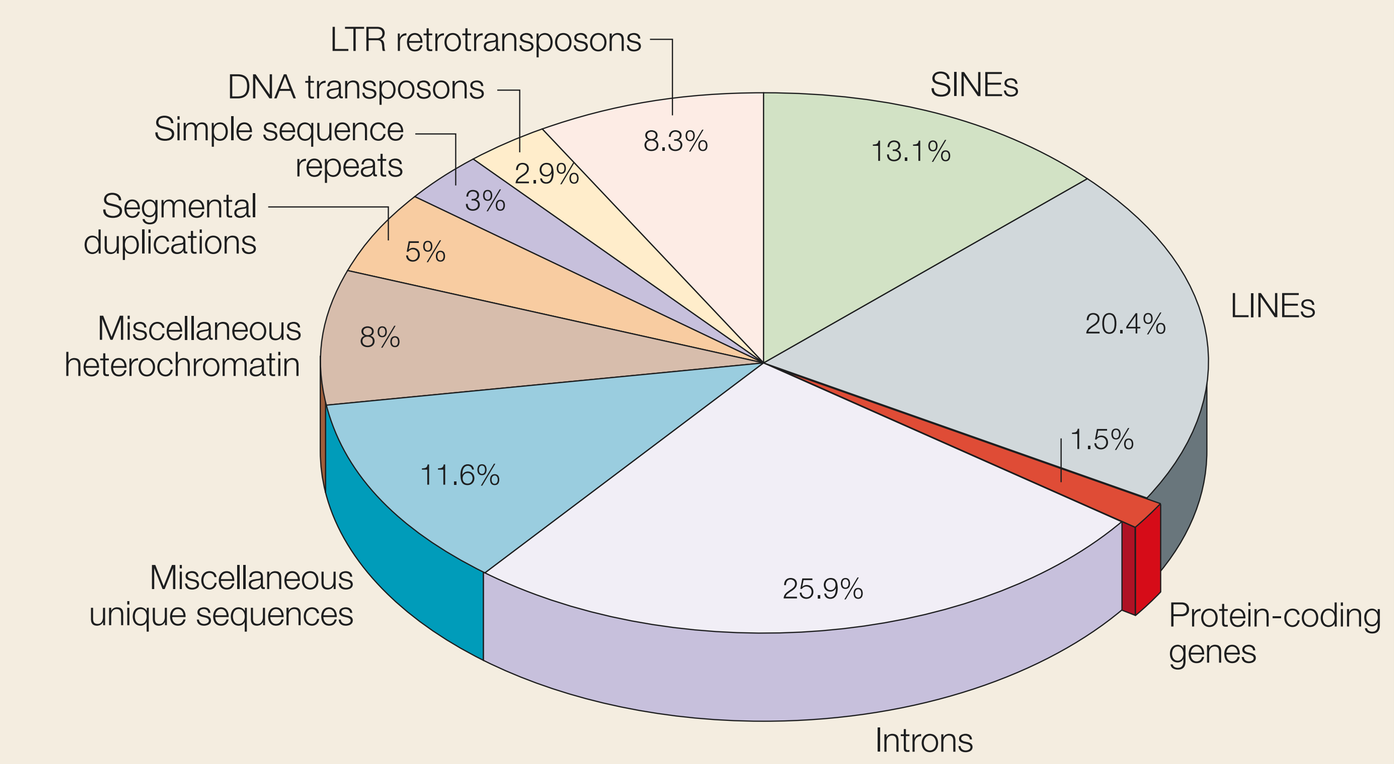


# of genes

Surprise #1 - We don't have that many "genes"

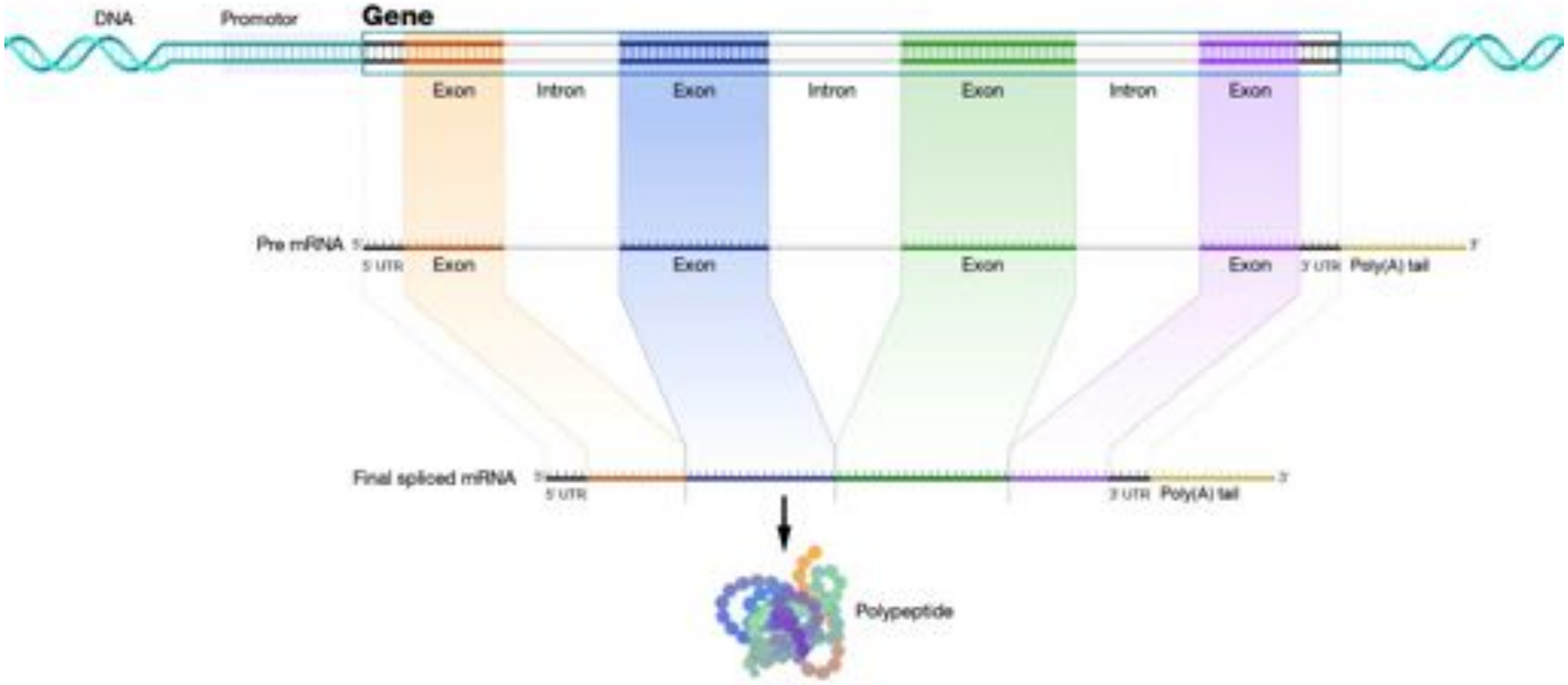


Surprise #2 - It's mostly junk?



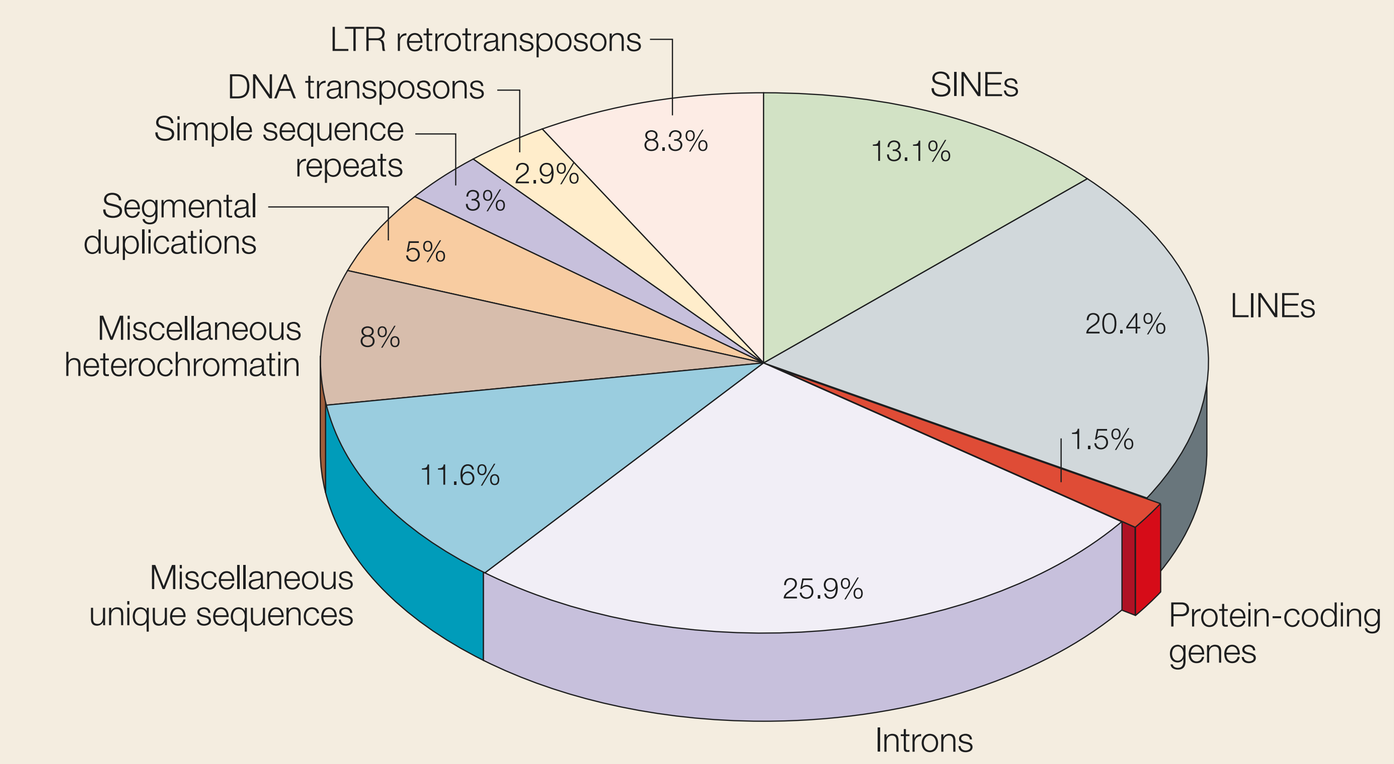
Nat Rev Genetics, 2005, *6*(9):699-708.

Surprise #2 - It's mostly junk?



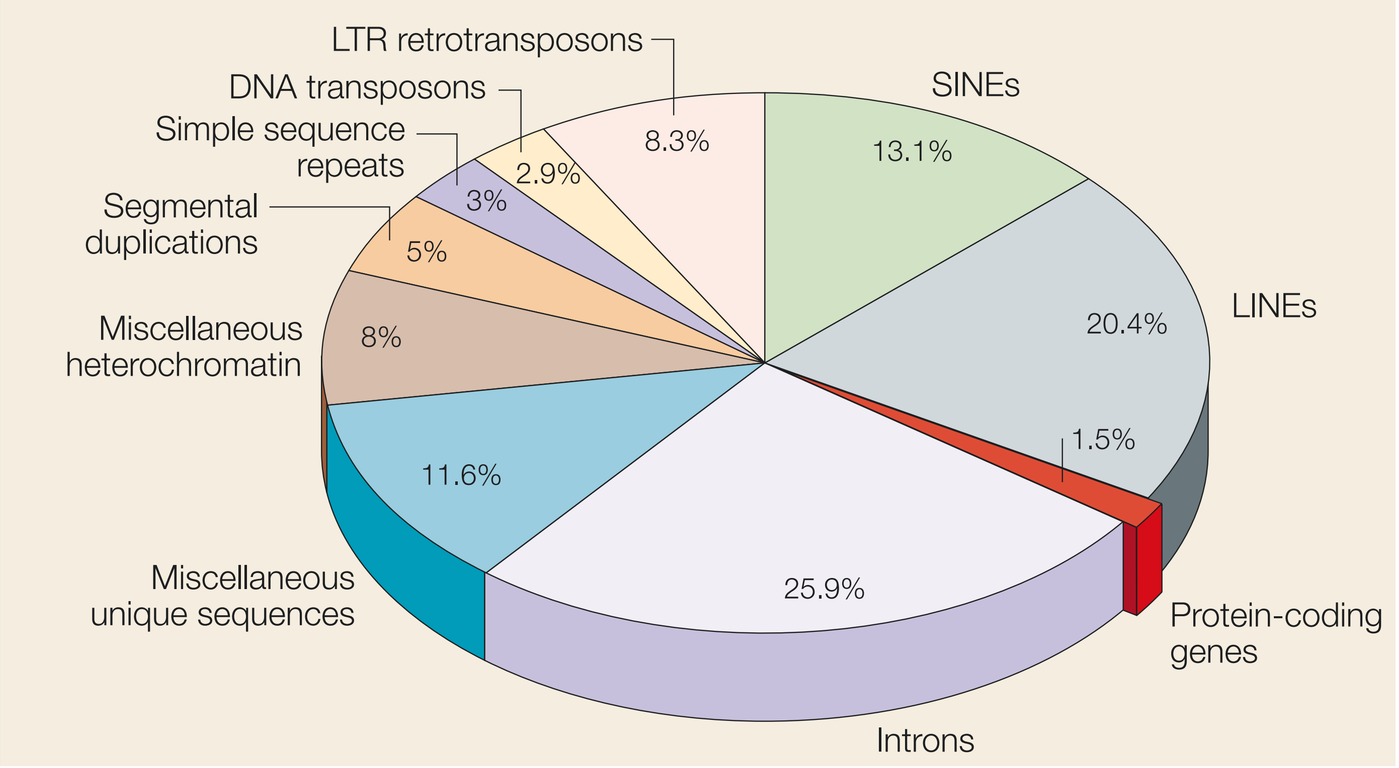
Nat Rev Genetics, 2005, *6*(9):699-708.

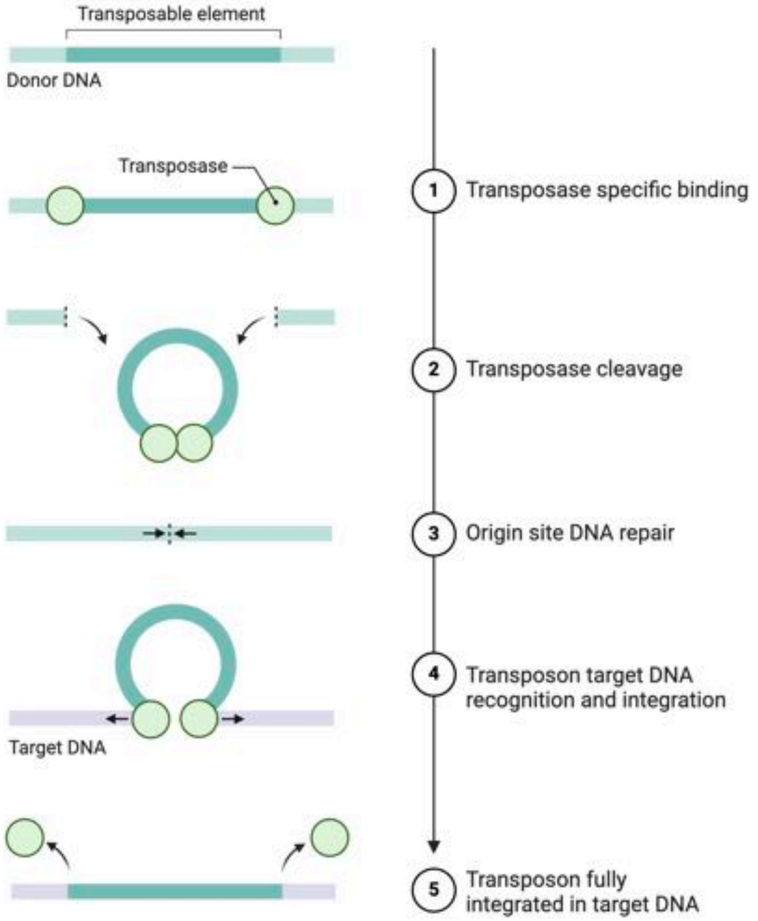
Surprise #2 - It's mostly junk?



Nat Rev Genetics, 2005, *6*(9):699-708.

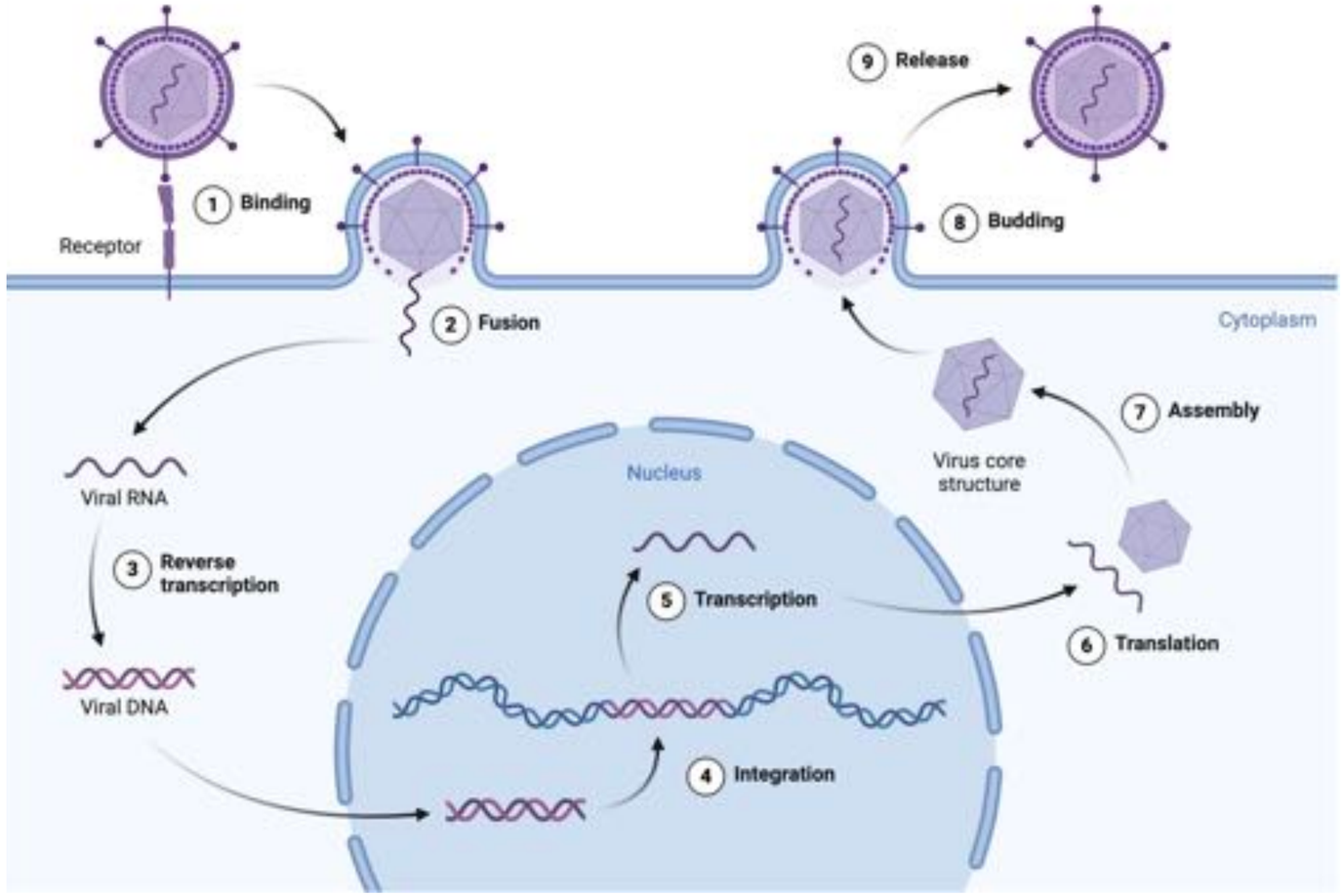
Surprise #3 - It's mostly transposons?

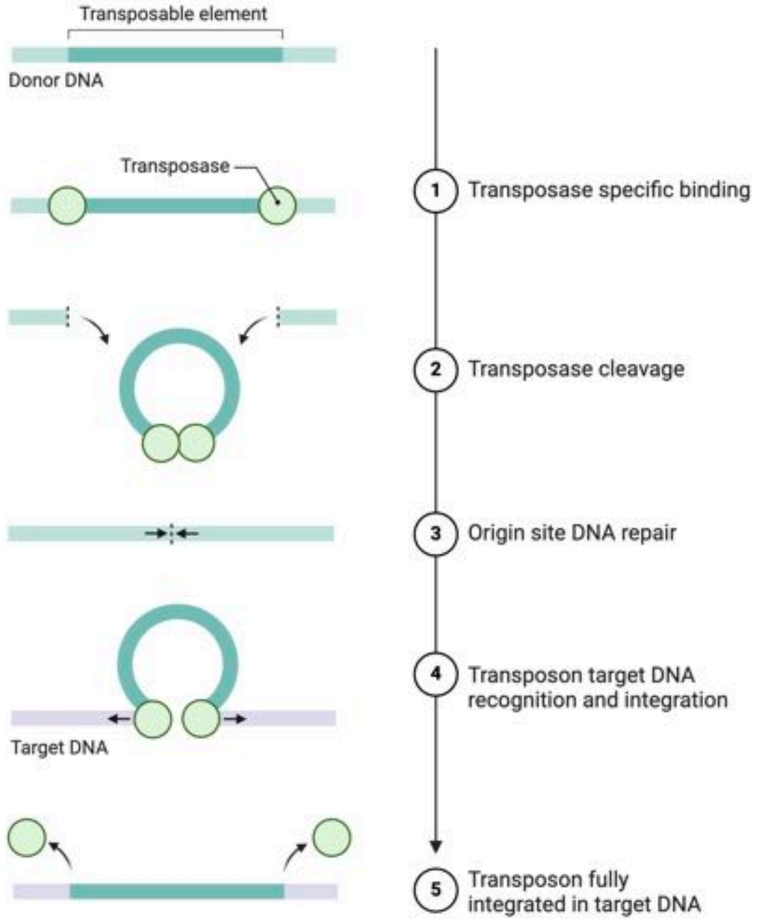




Nat Rev Genetics, 2005, *6*(9):699-708.

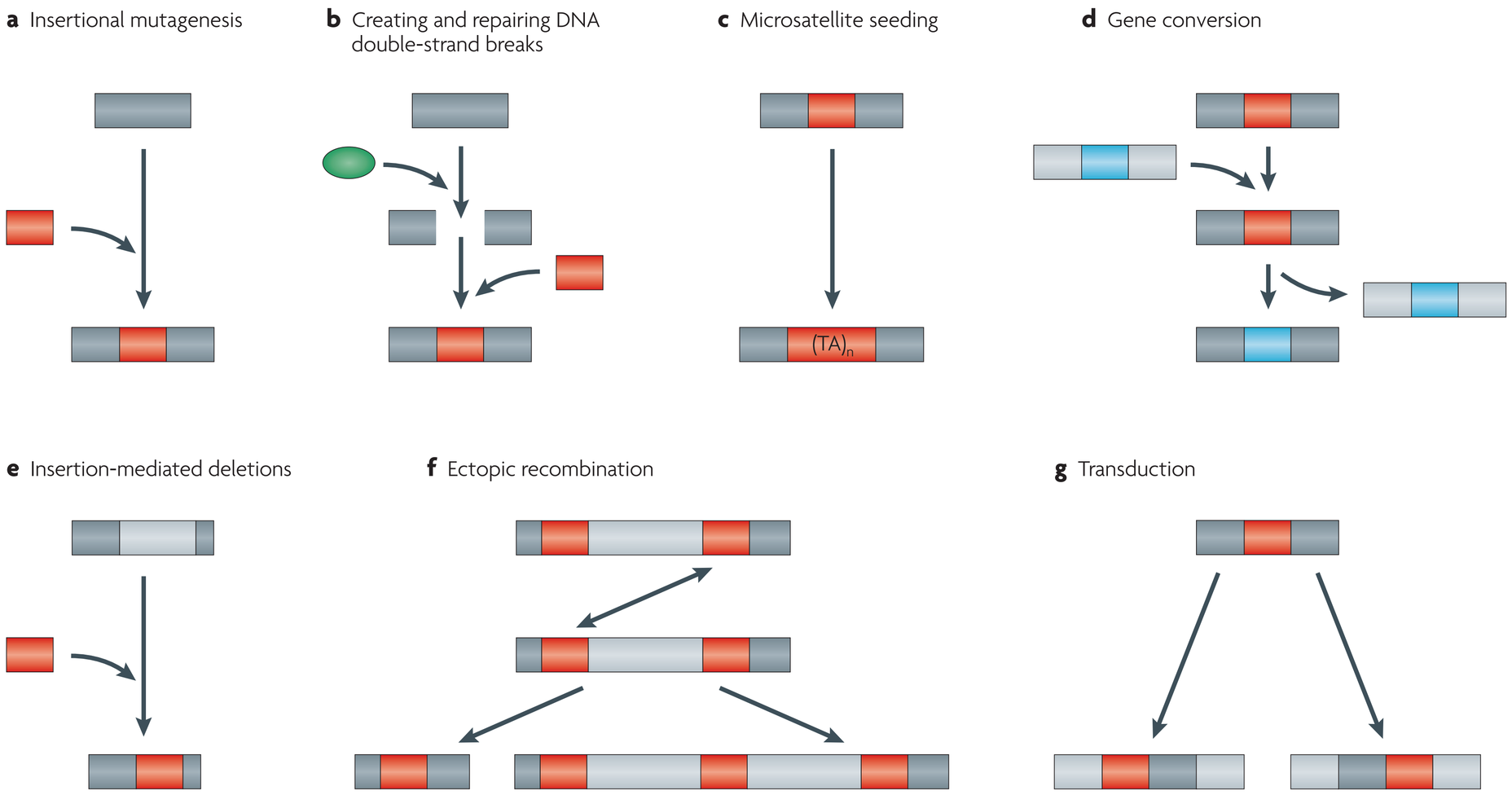
Surprise #3 - It's mostly transposons?





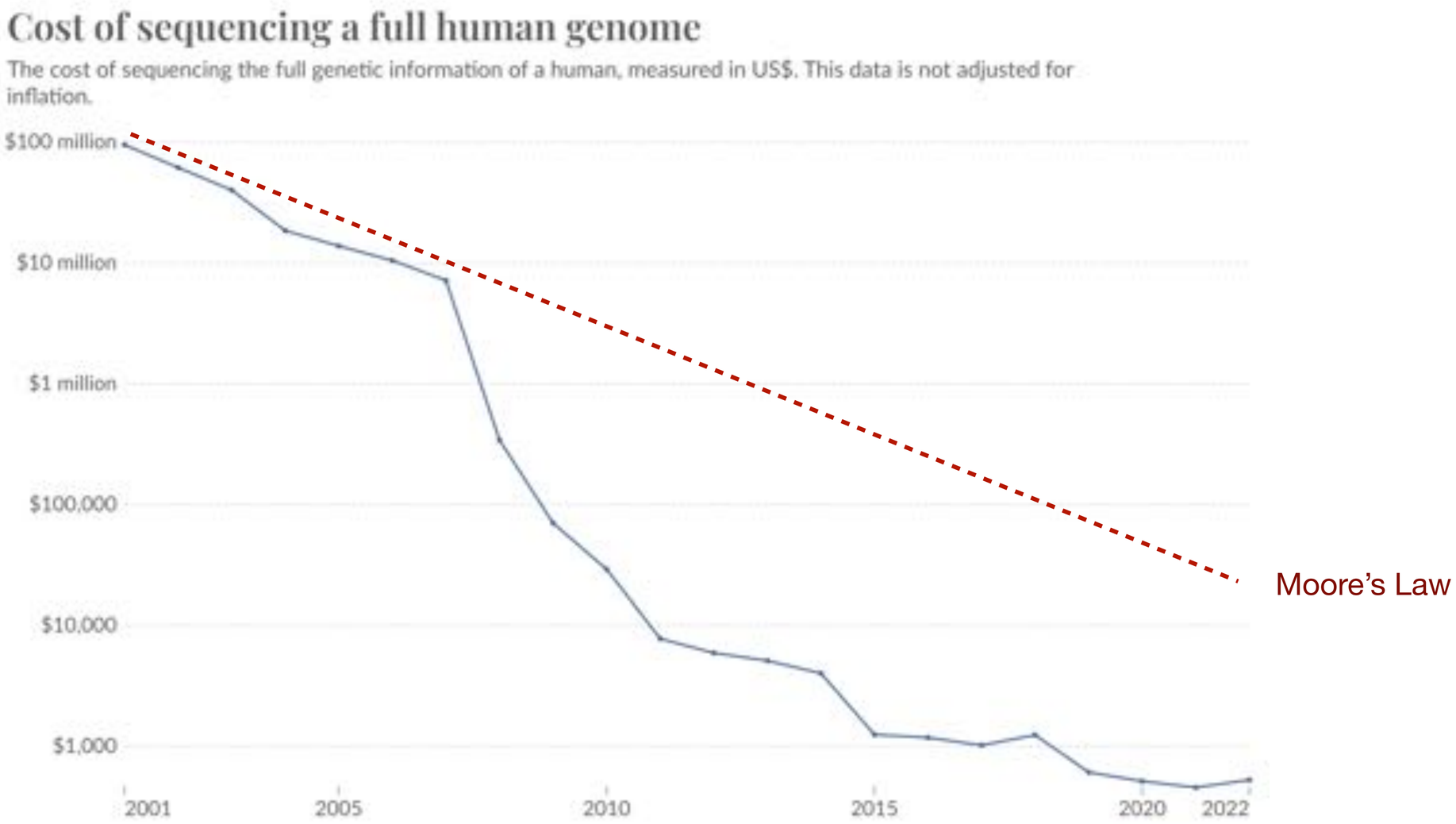
Nat Rev Genetics, 2005, *6*(9):699-708.

Surprise #3 - It's mostly transposons?



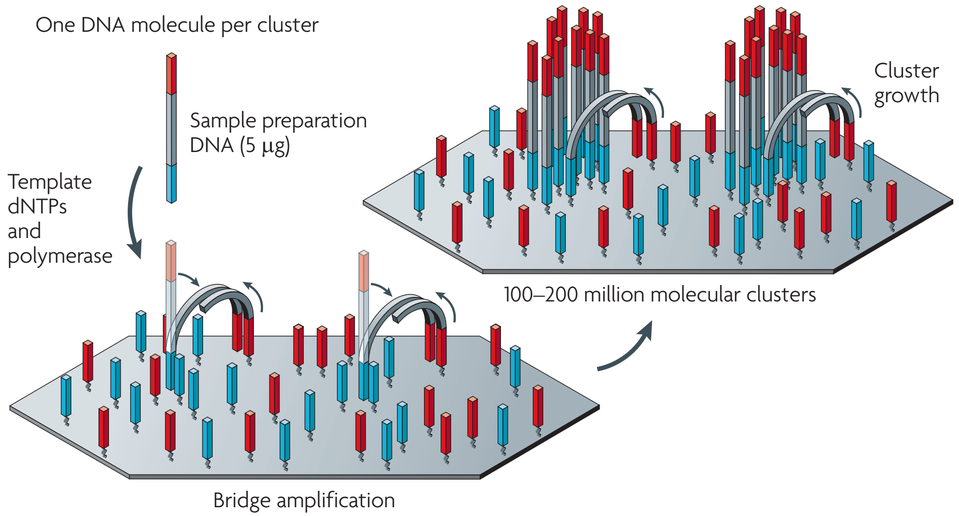
Nat Rev Genetics, 2005, *6*(9):699-708.

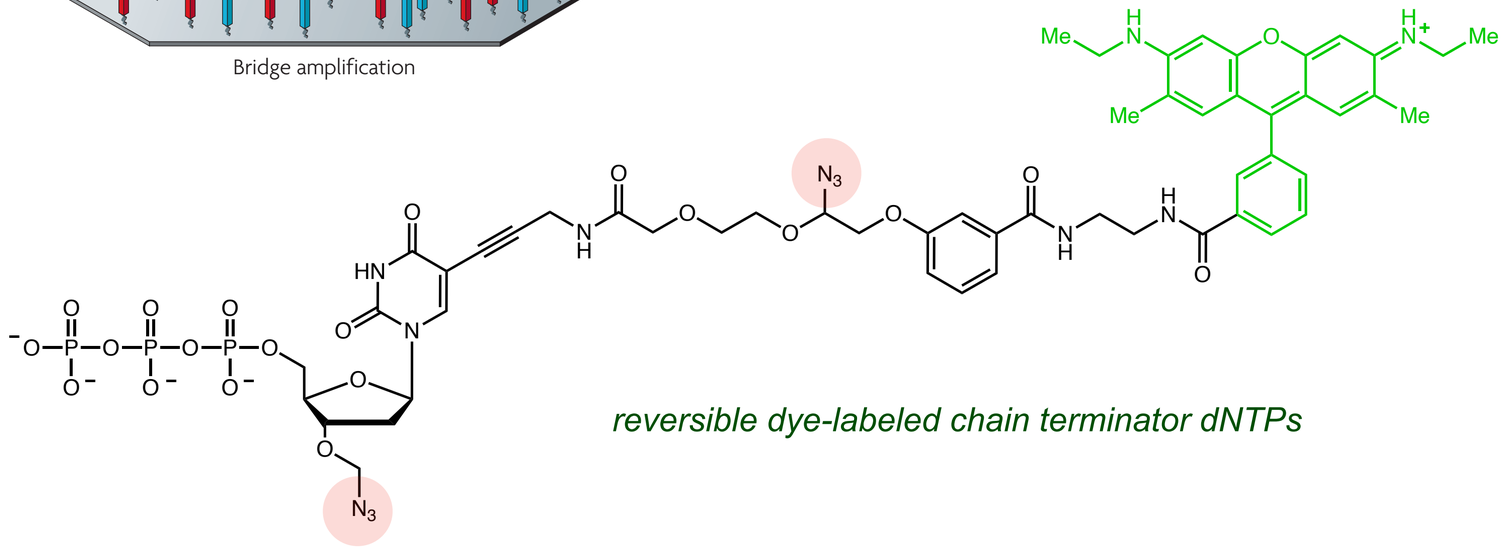
2005 and Onwards - Genome Sequencing Accelerates



2nd (Next) Generation DNA Sequencing

solid-phase bridge amplification





reversible dye-labeled chain terminator dNTPs

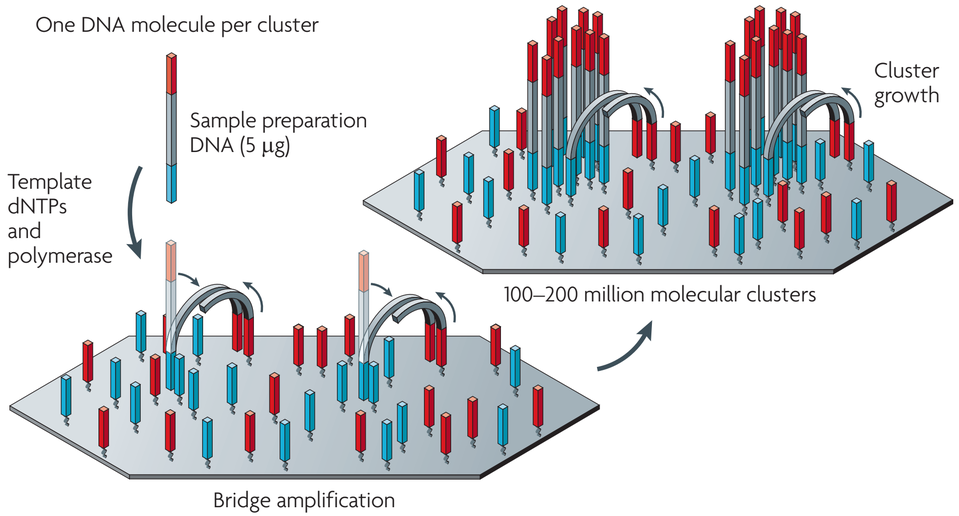
Shankar Balasubramanian

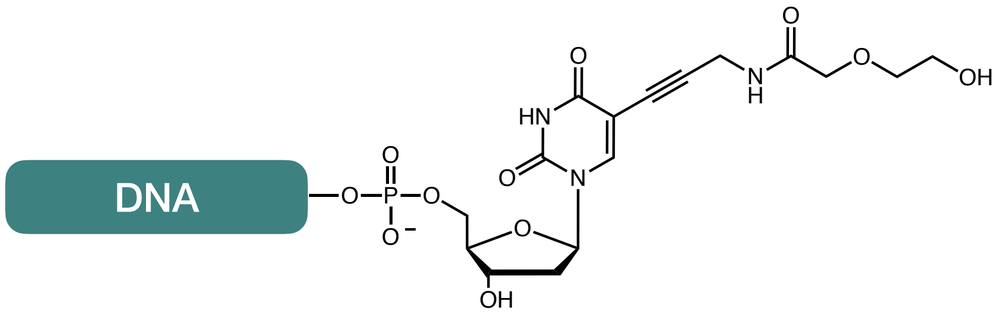
David Klenerman

Nature, 2008, *456,* 53-59.

2nd (Next) Generation DNA Sequencing

solid-phase bridge amplification



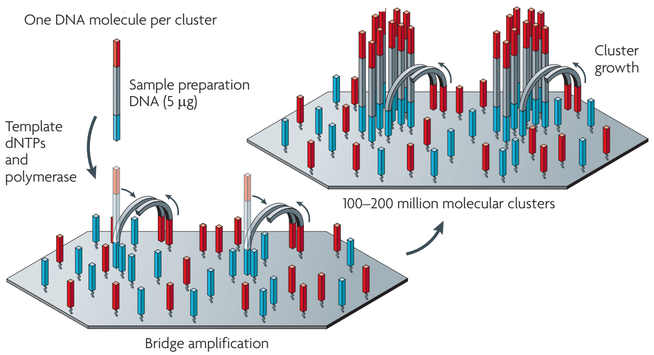


Shankar Balasubramanian

David Klenerman

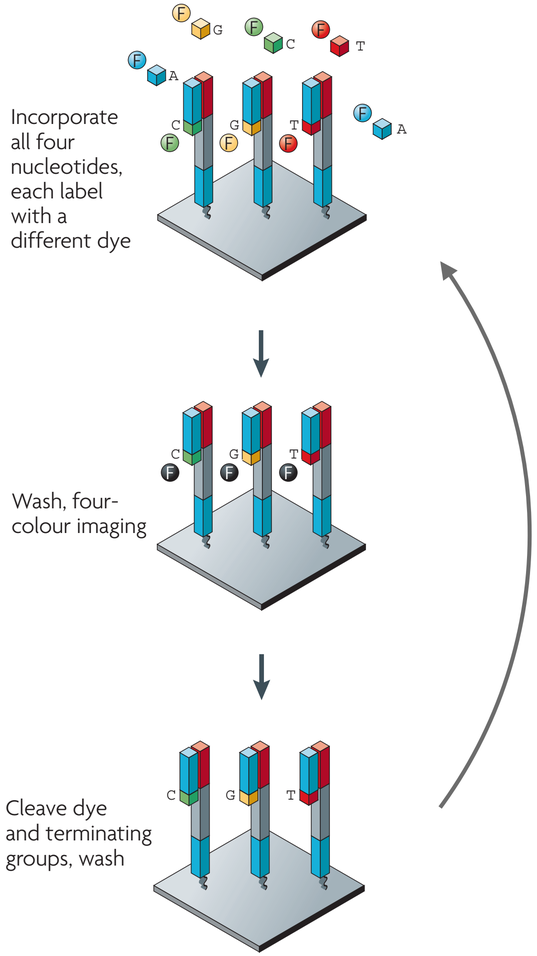
Nature, 2008, *456,* 53-59.

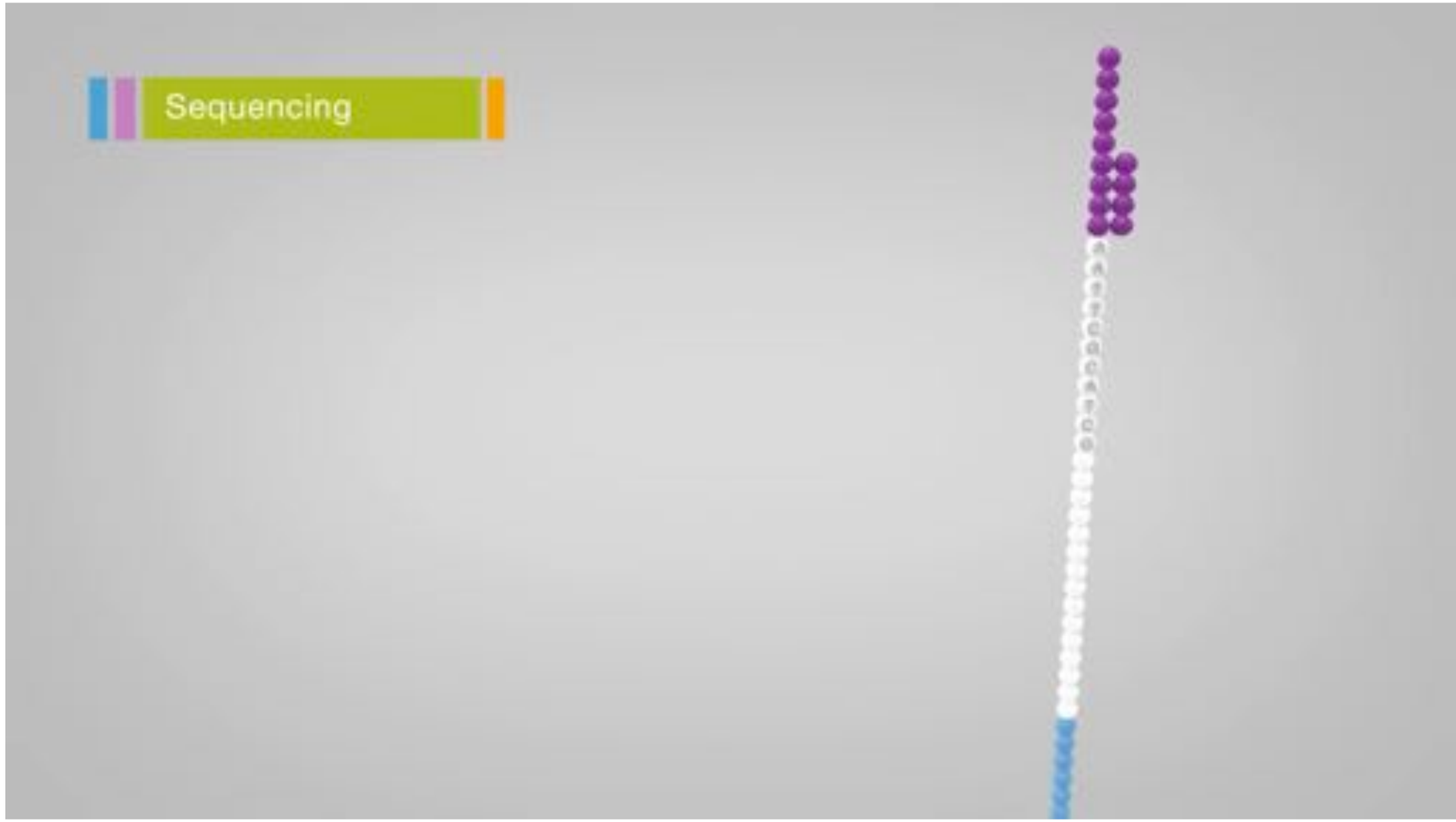
2nd (Next) Generation DNA Sequencing



Nature, 2008, *456,* 53-59.

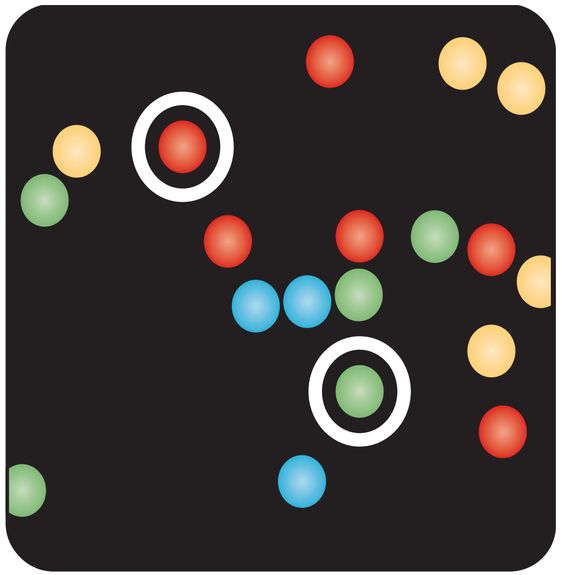
2nd (Next) Generation DNA Sequencing

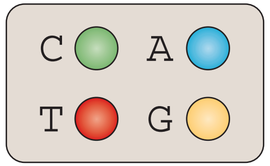




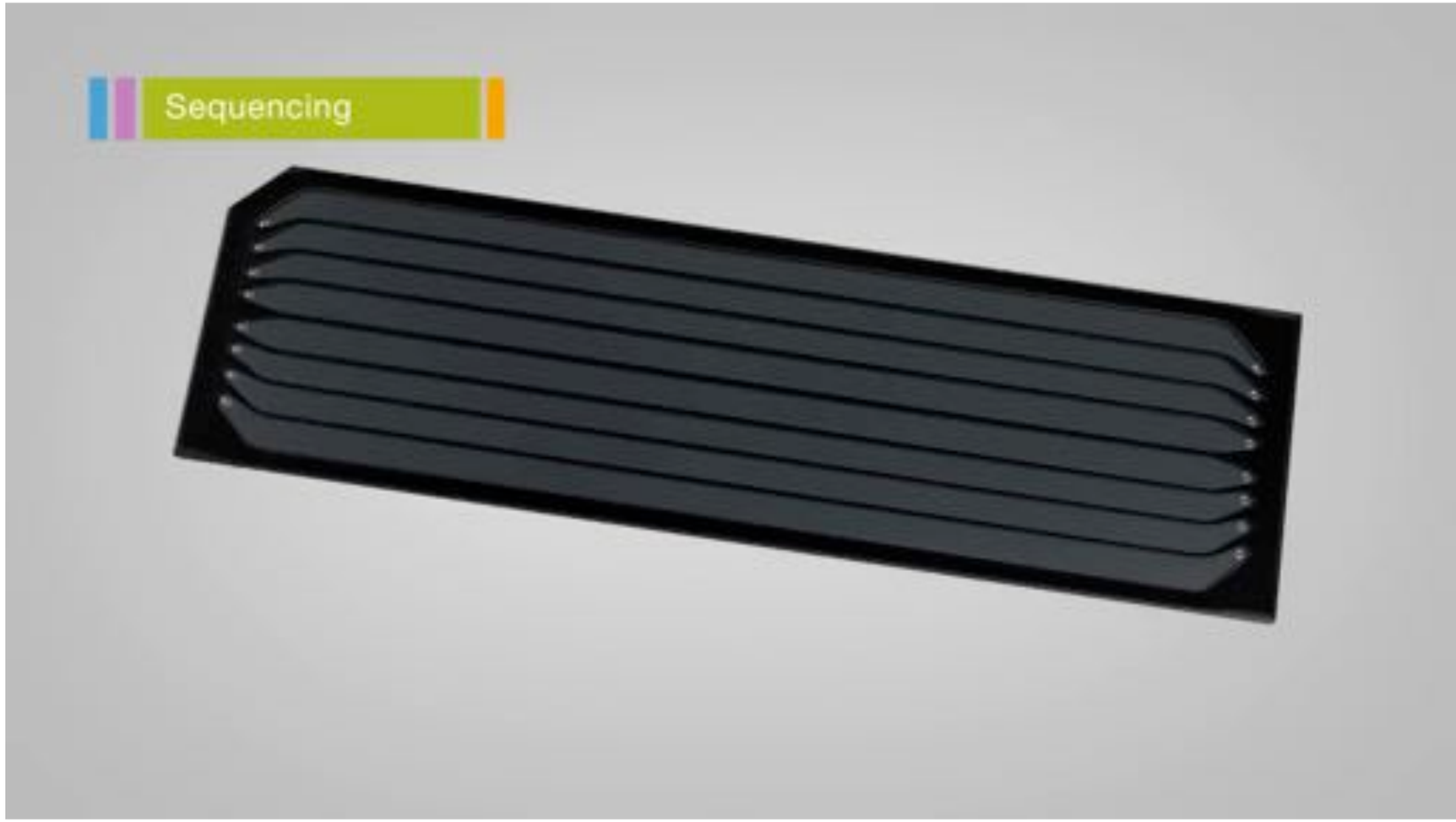
Nature, 2008, *456,* 53-59.

2nd (Next) Generation DNA Sequencing



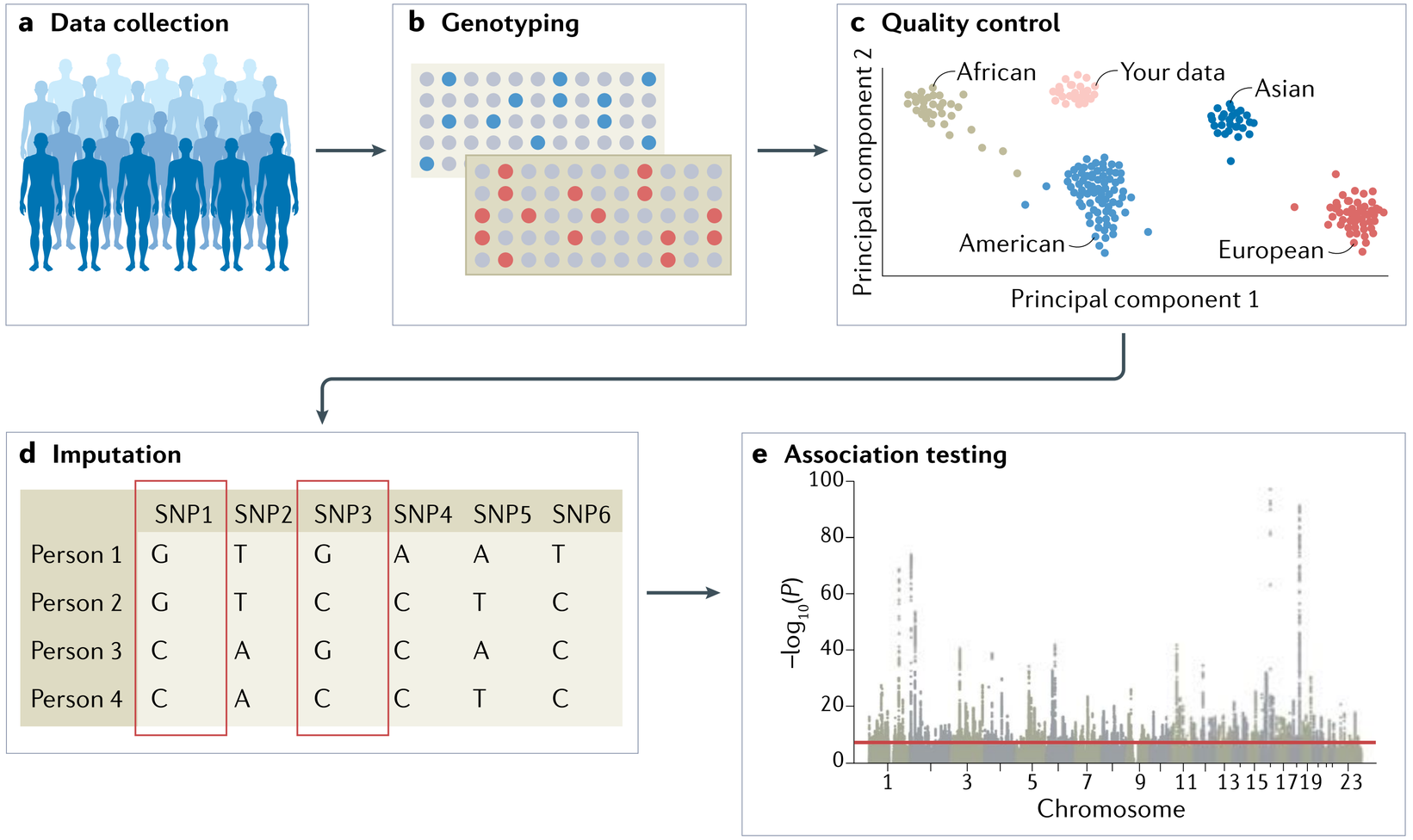


Top: CATCGT
Bottom: CCCCCC



Nature, 2008, *456,* 53-59.

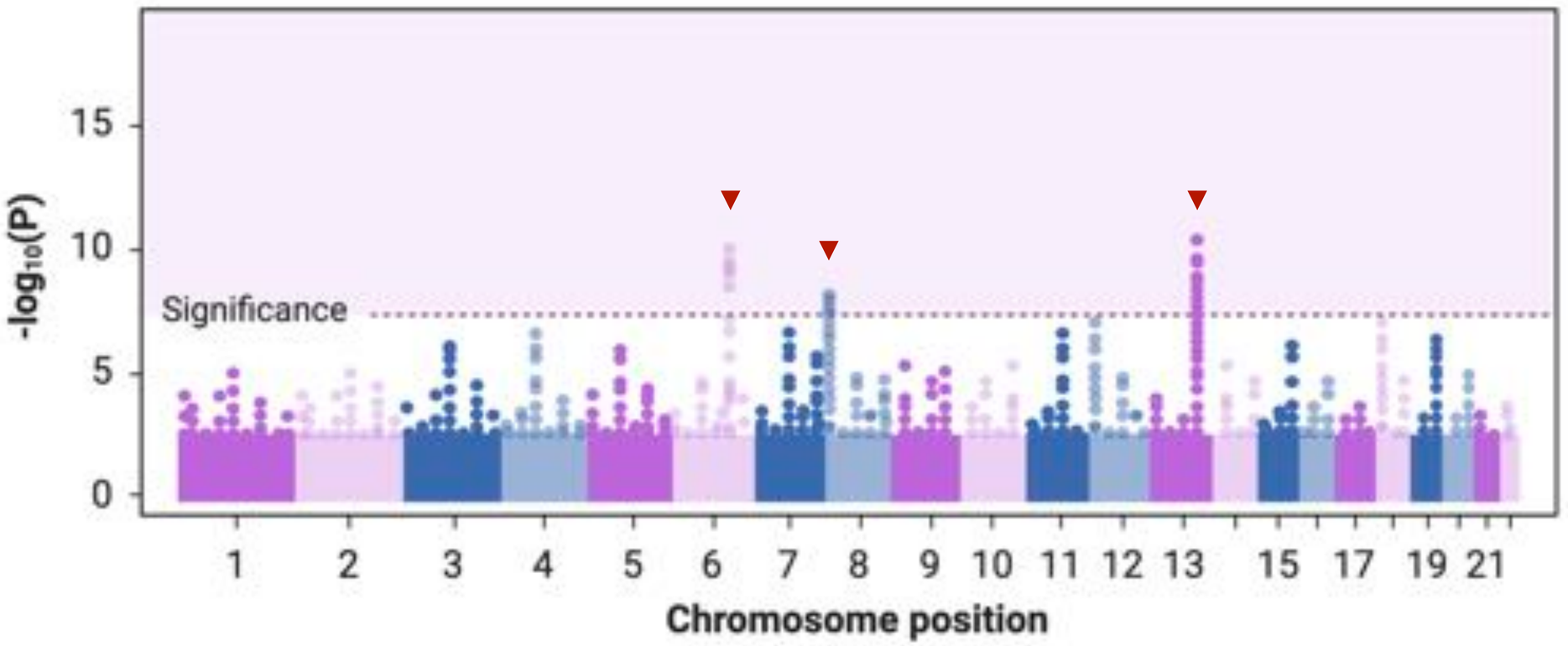
More Genomes Sequenced = Genome-Wide Association Studies (GWAS)



Nat Rev Genetics, 2021, *1,* 59.

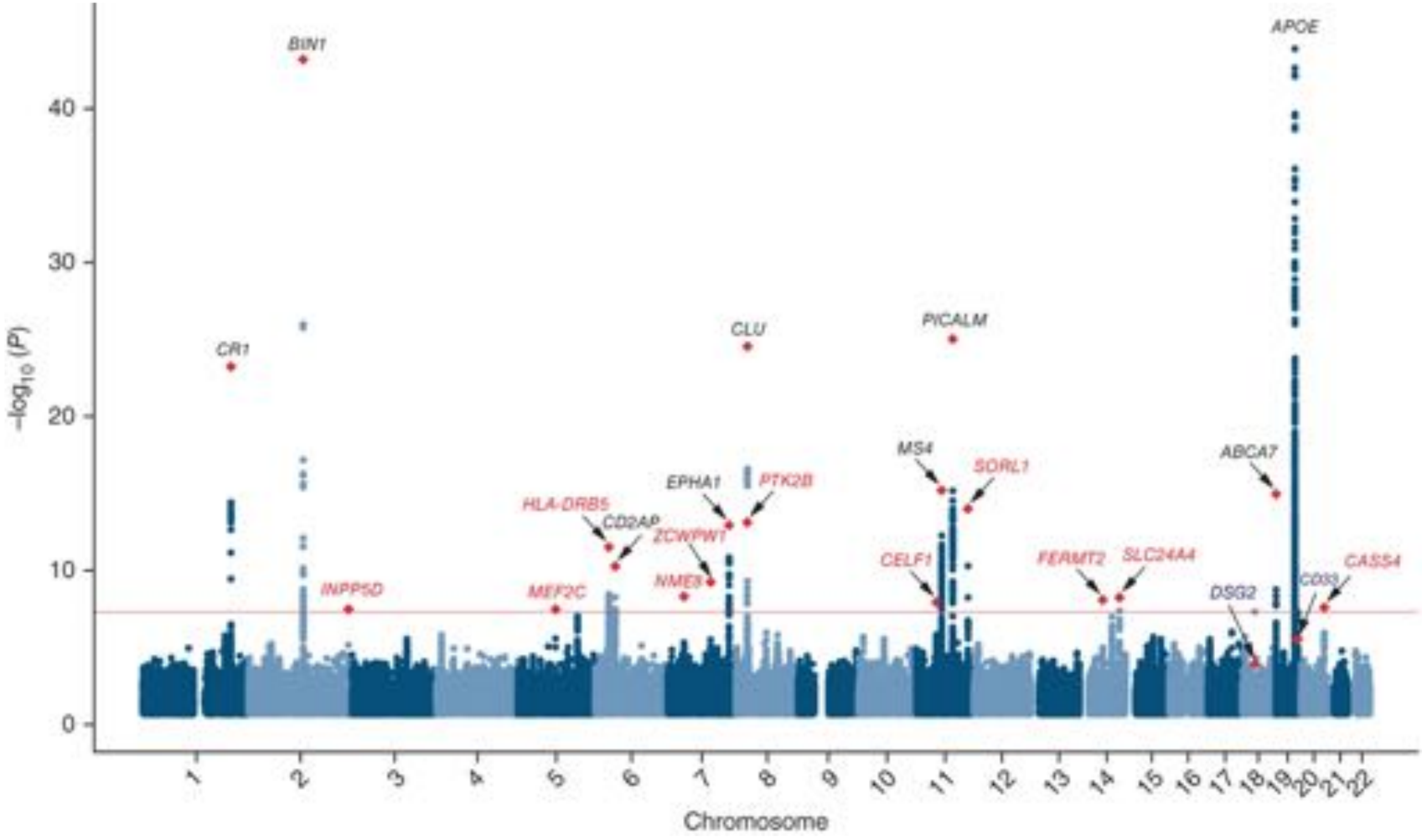
More Genomes Sequenced = Genome-Wide Association Studies (GWAS)

Trait/Disease vs Normal population



Nat Rev Genetics, 2021, *1,* 59.

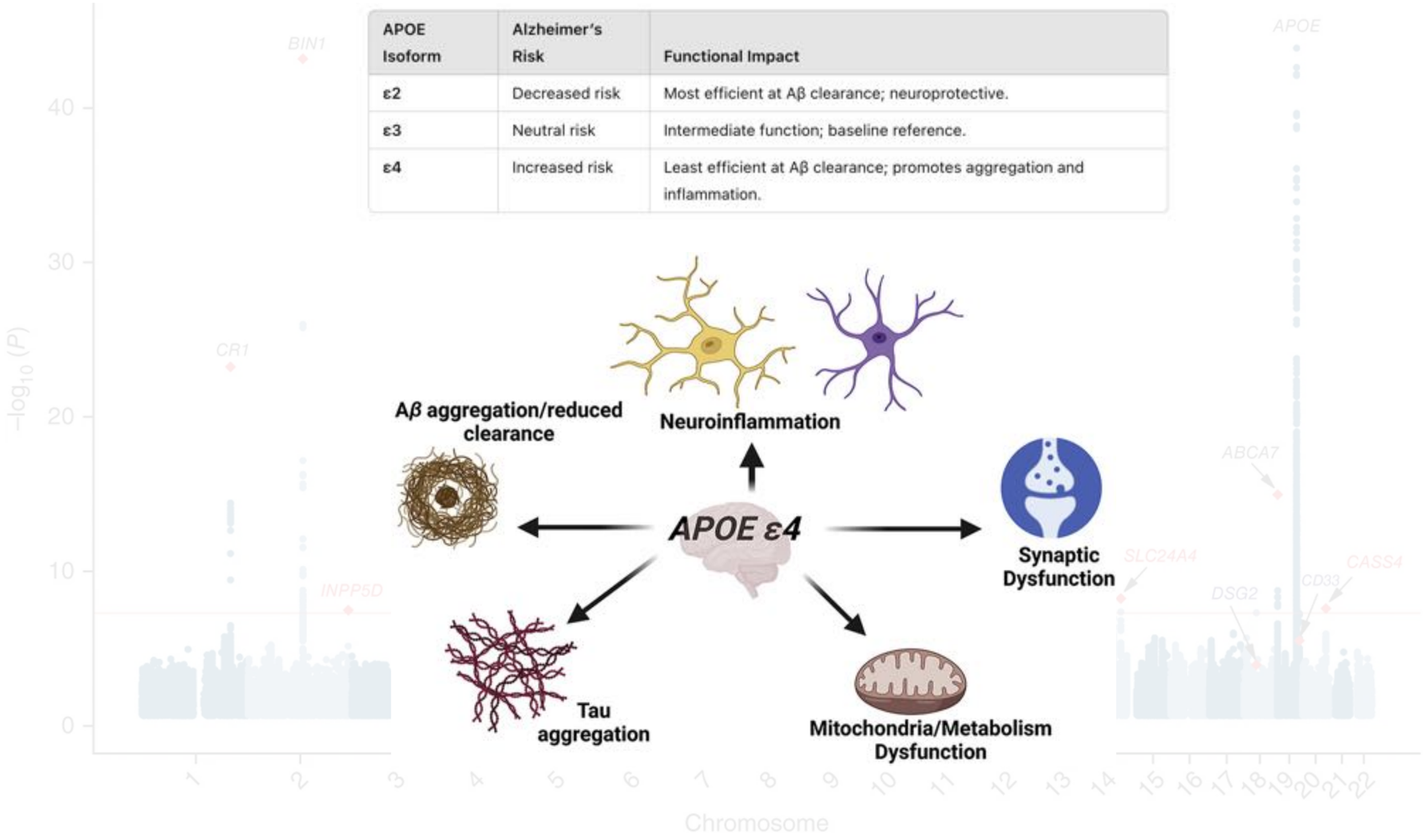
Gene Mutations that Cause (?) Late-Onset Alzheimer's



Nature genetics, 2013, *45*(12), 1452-1458.

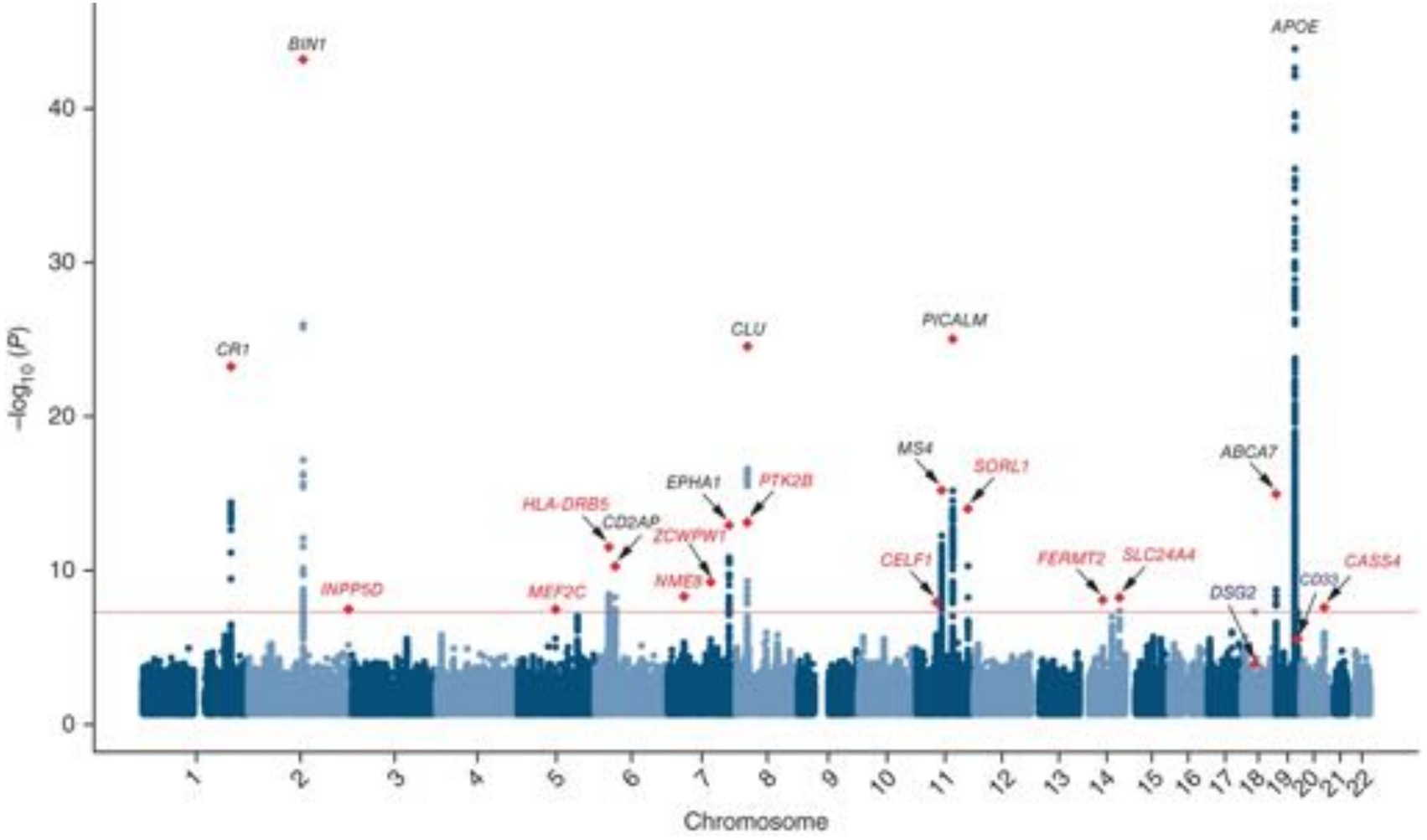
Gene Mutations that Cause (?) Late-Onset Alzheimer's

APOE Isoform	Alzheimer's Risk	Functional Impact
ε2	Decreased risk	Most efficient at Aβ clearance; neuroprotective.
ε3	Neutral risk	Intermediate function; baseline reference.
ε4	Increased risk	Least efficient at Aβ clearance; promotes aggregation and inflammation.



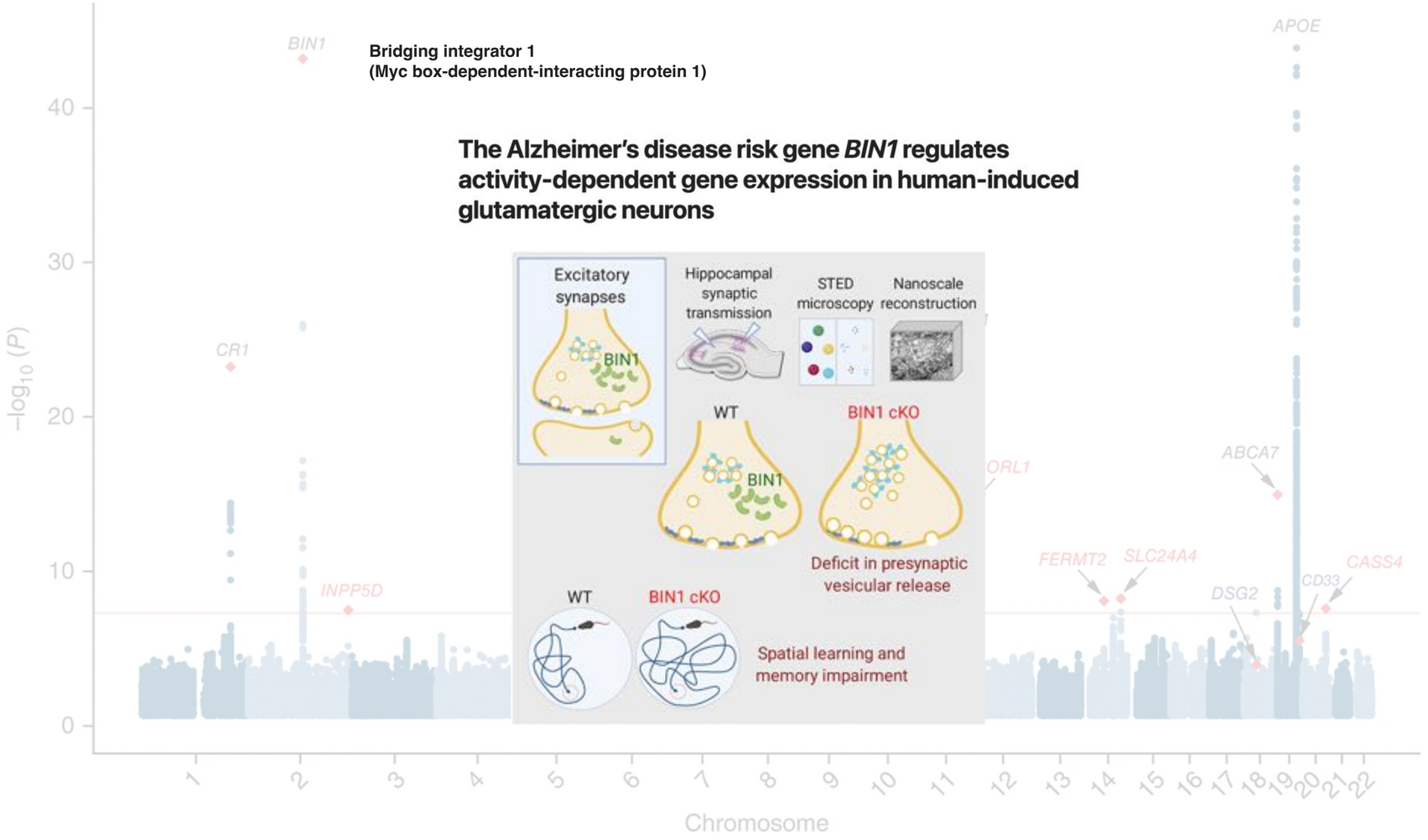
Nature genetics, 2013, *45*(12), 1452-1458.

Gene Mutations that Cause (?) Late-Onset Alzheimer's



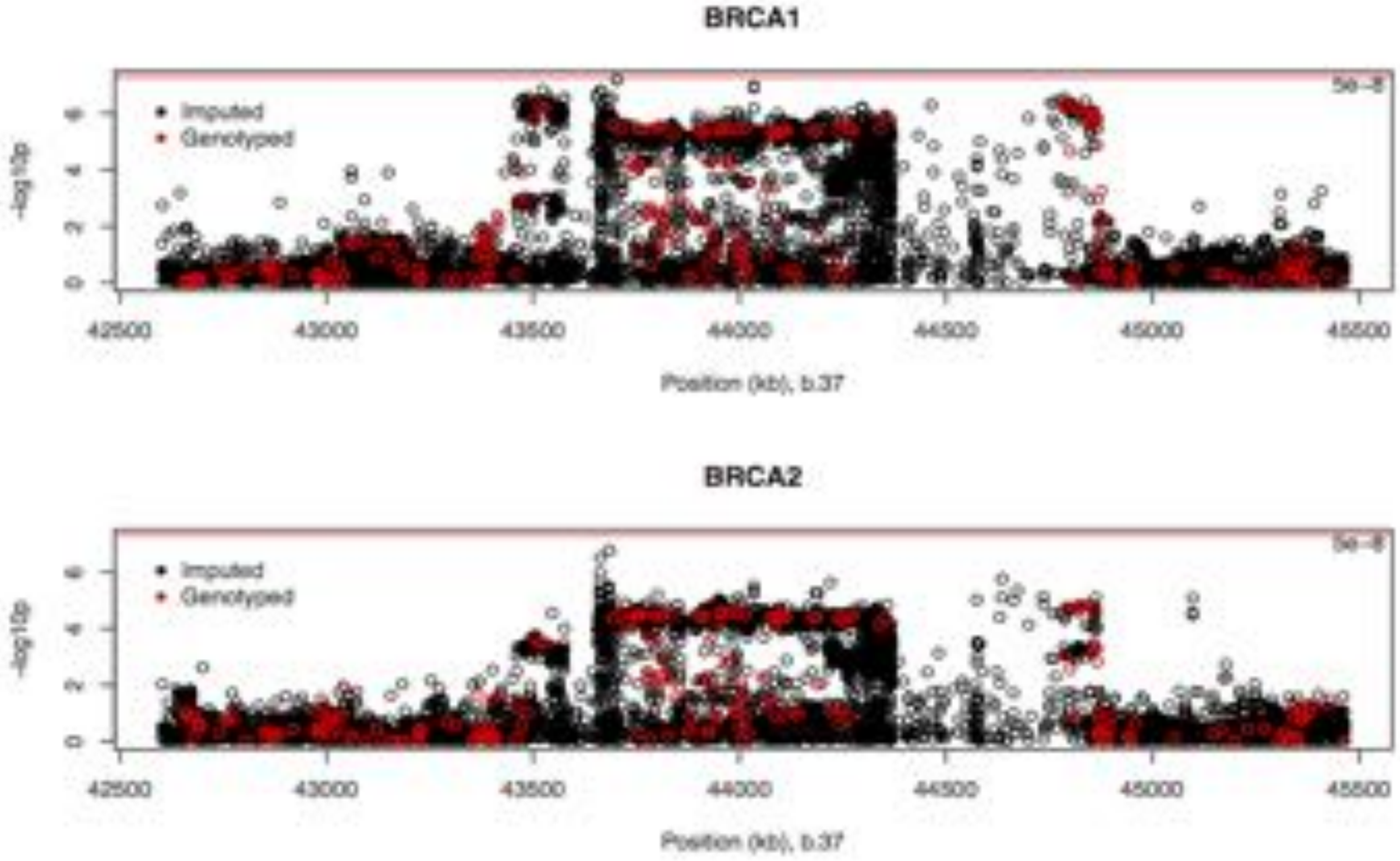
Nature genetics, 2013, *45*(12), 1452-1458.

Gene Mutations that Cause (?) Late-Onset Alzheimer's



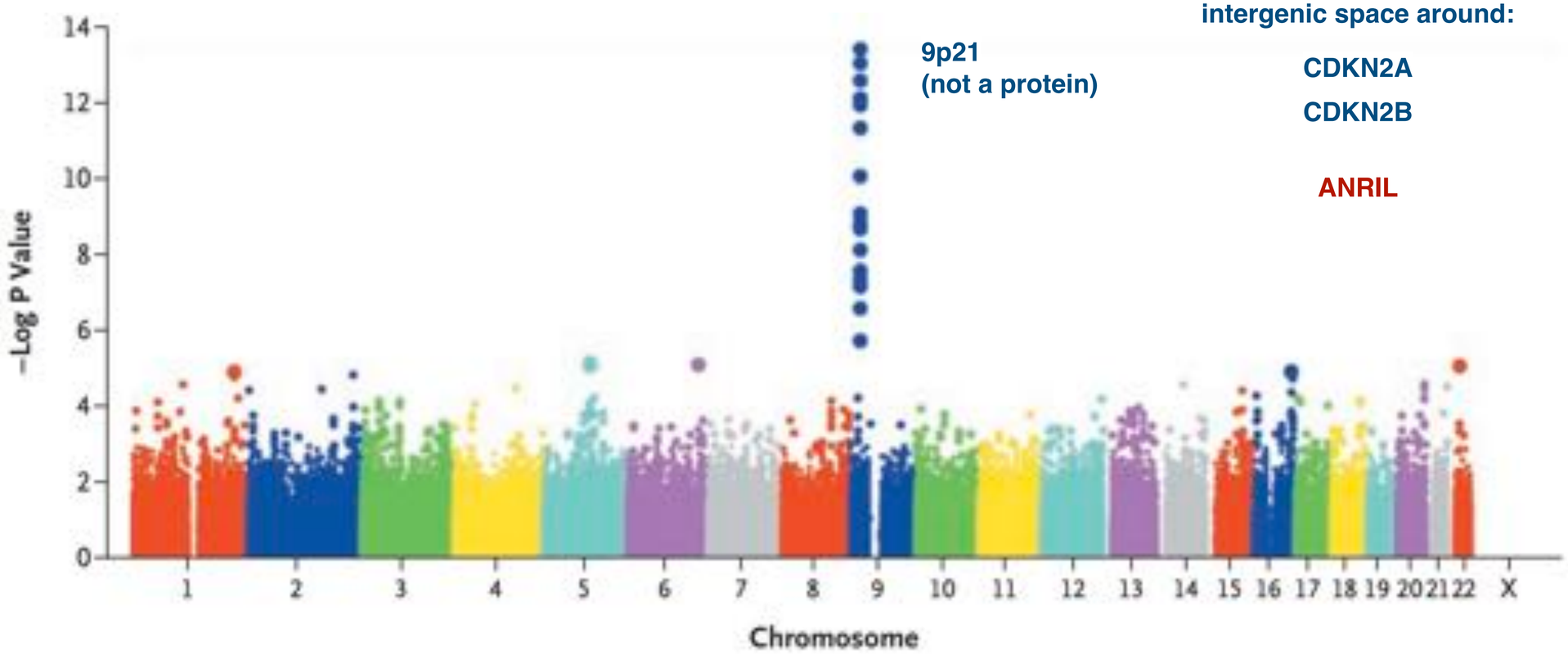
Nature genetics, 2013, *45*(12), 1452-1458.

Hot-spot Mutations in Known Cancer Drivers



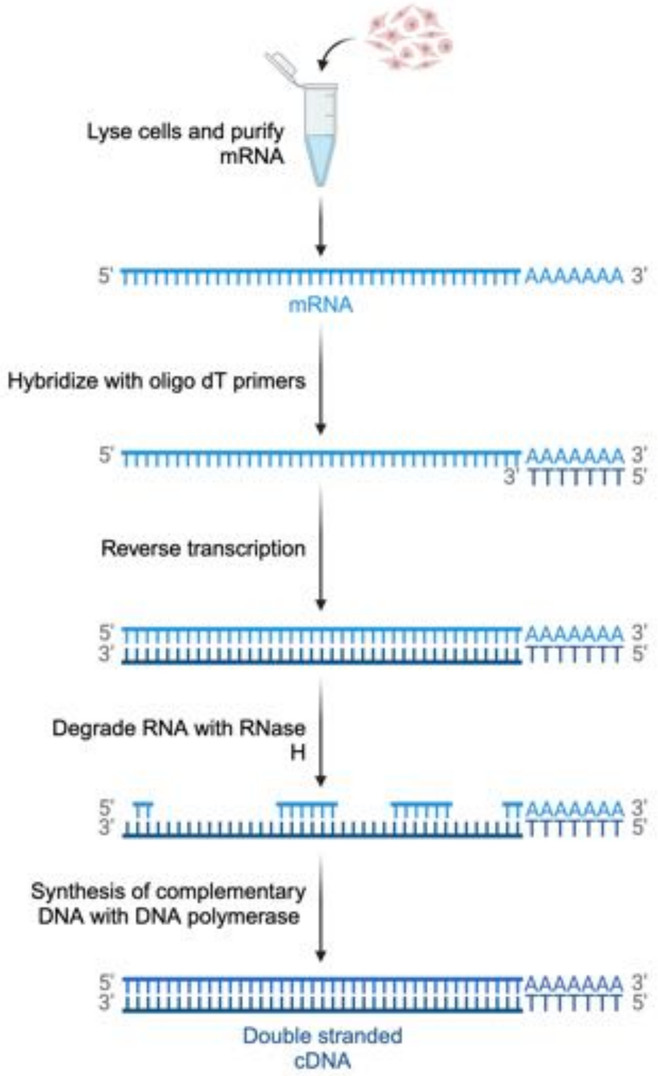
PLoS genetics, 2013, *9*(3), e1003212

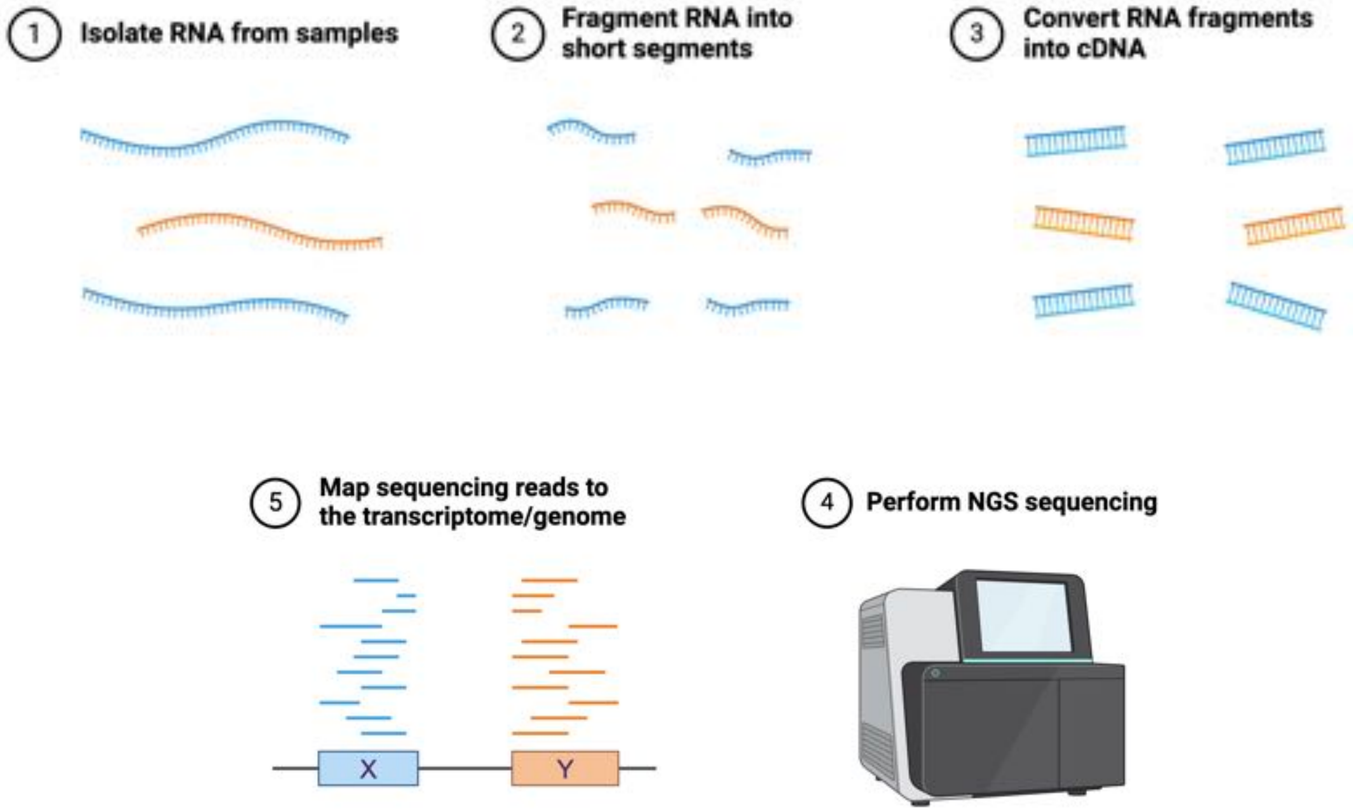
Gene Mutations that Correlate with Coronary Artery Disease



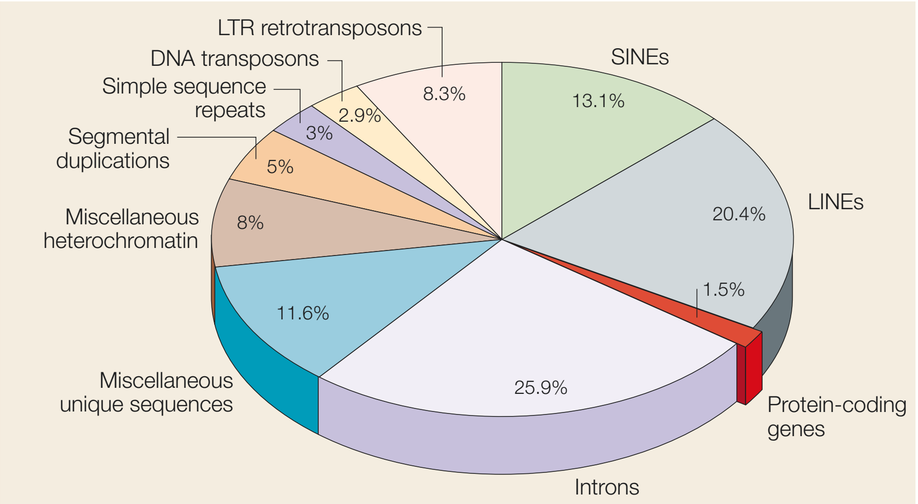
New England Journal of Medicine, 2007, *357*(5), 443-453.

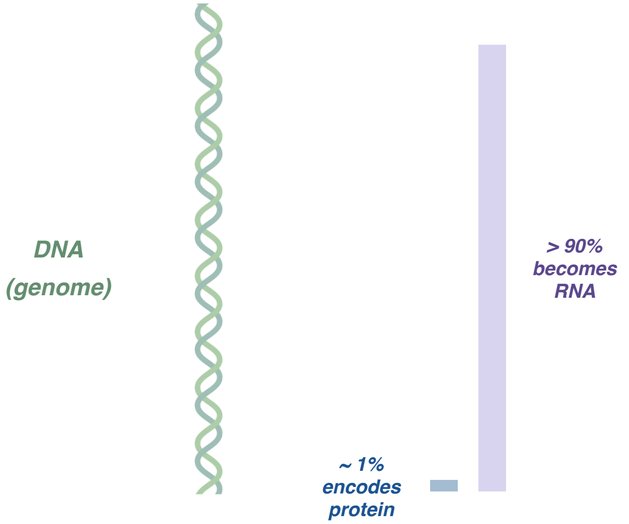
RNA-seq

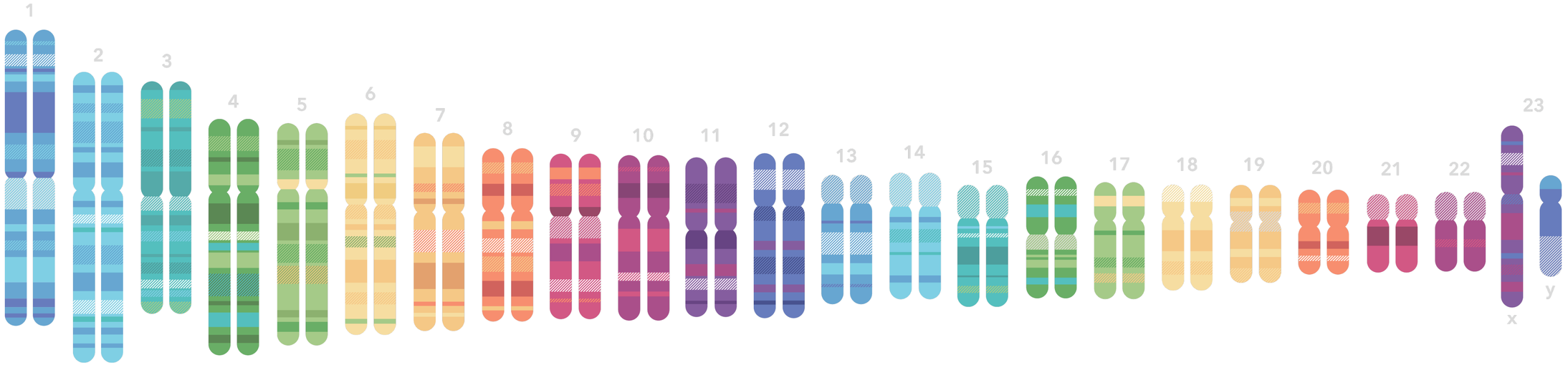

Lyse cells and purify mRNA
5' AAAAAAA 3'
mRNA
Hybridize with oligo dT primers
5' AAAAAAA 3'
3' 5'
Reverse transcription
5' AAAAAAA 3'
3' 5'
Degrade RNA with RNase H
5' AAAAAAA 3'
3' 5'
Synthesis of complementary DNA with DNA polymerase
5' AAAAAAA 3'
3' 5'
Double stranded cDNA


1 Isolate RNA from samples
2 Fragment RNA into short segments
3 Convert RNA fragments into cDNA
5 Map sequencing reads to the transcriptome/genome
X
Y
4 Perform NGS sequencing

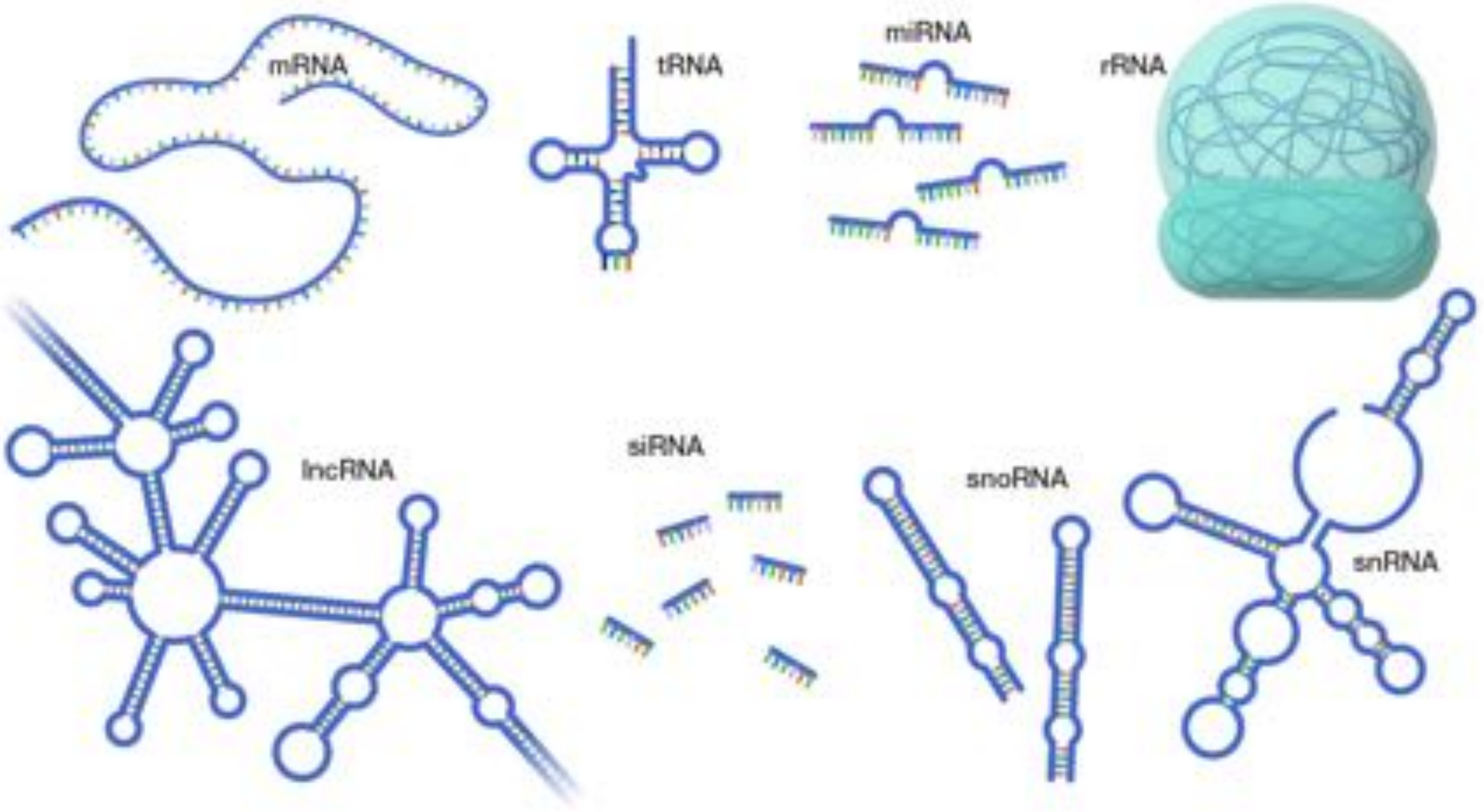
Surprise #1 - Most of the Genome is Transcribed





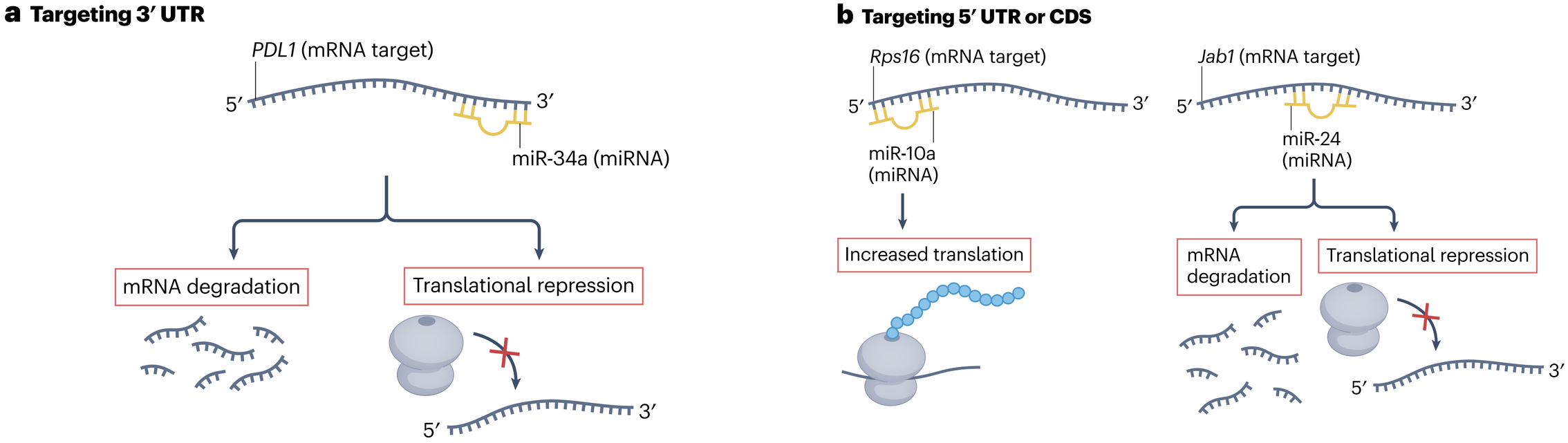


Non-Coding RNAs



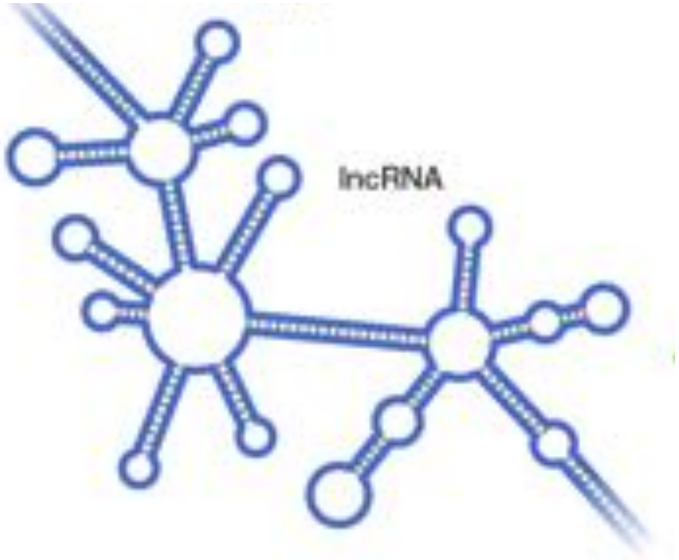
Nature Reviews Genetics, 2024, *25*(3), 211-232

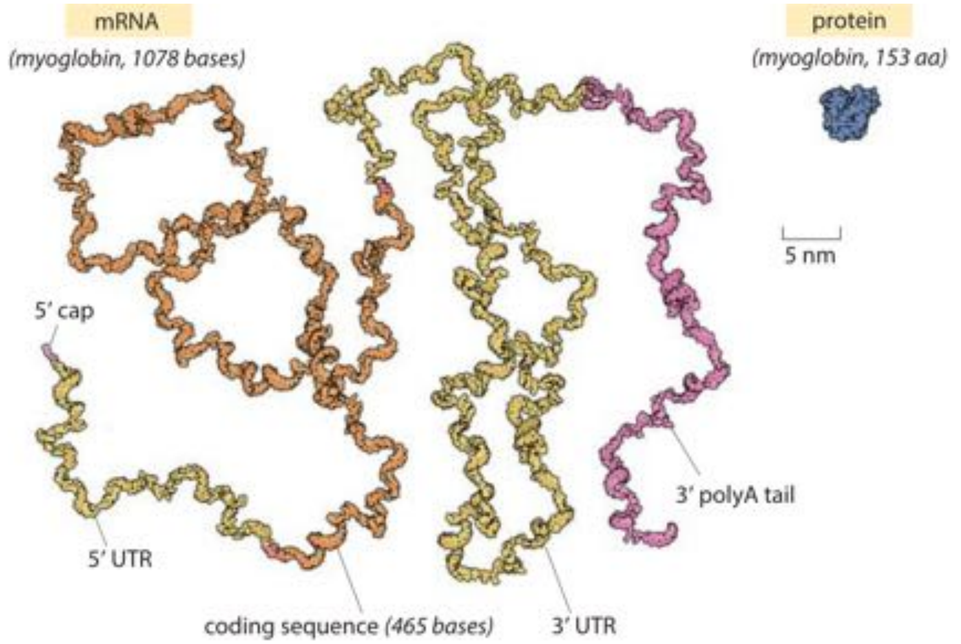
Non-Coding RNAs ….. are bioactive



Nature Reviews Genetics, 2024, *25*(3), 211-232

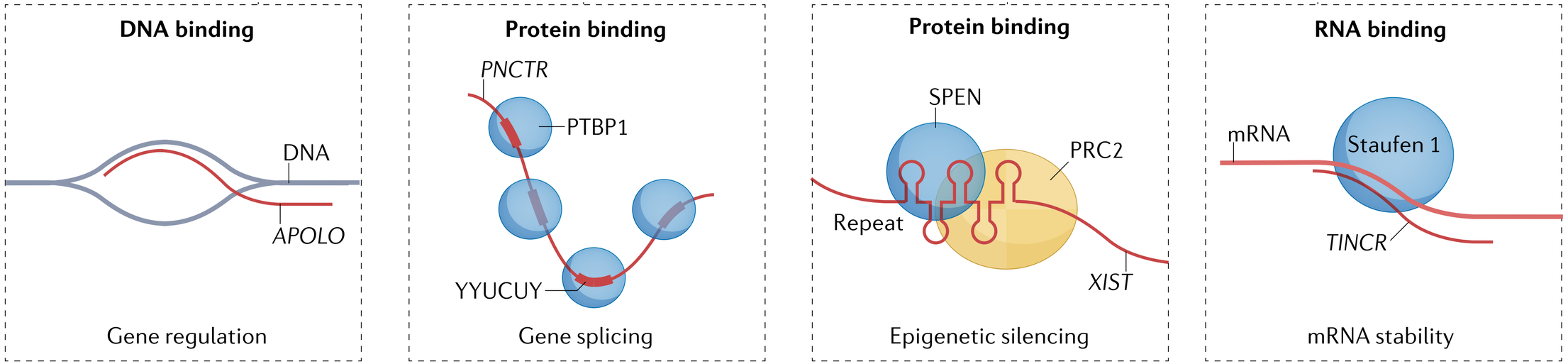
Long Non-Coding RNAs





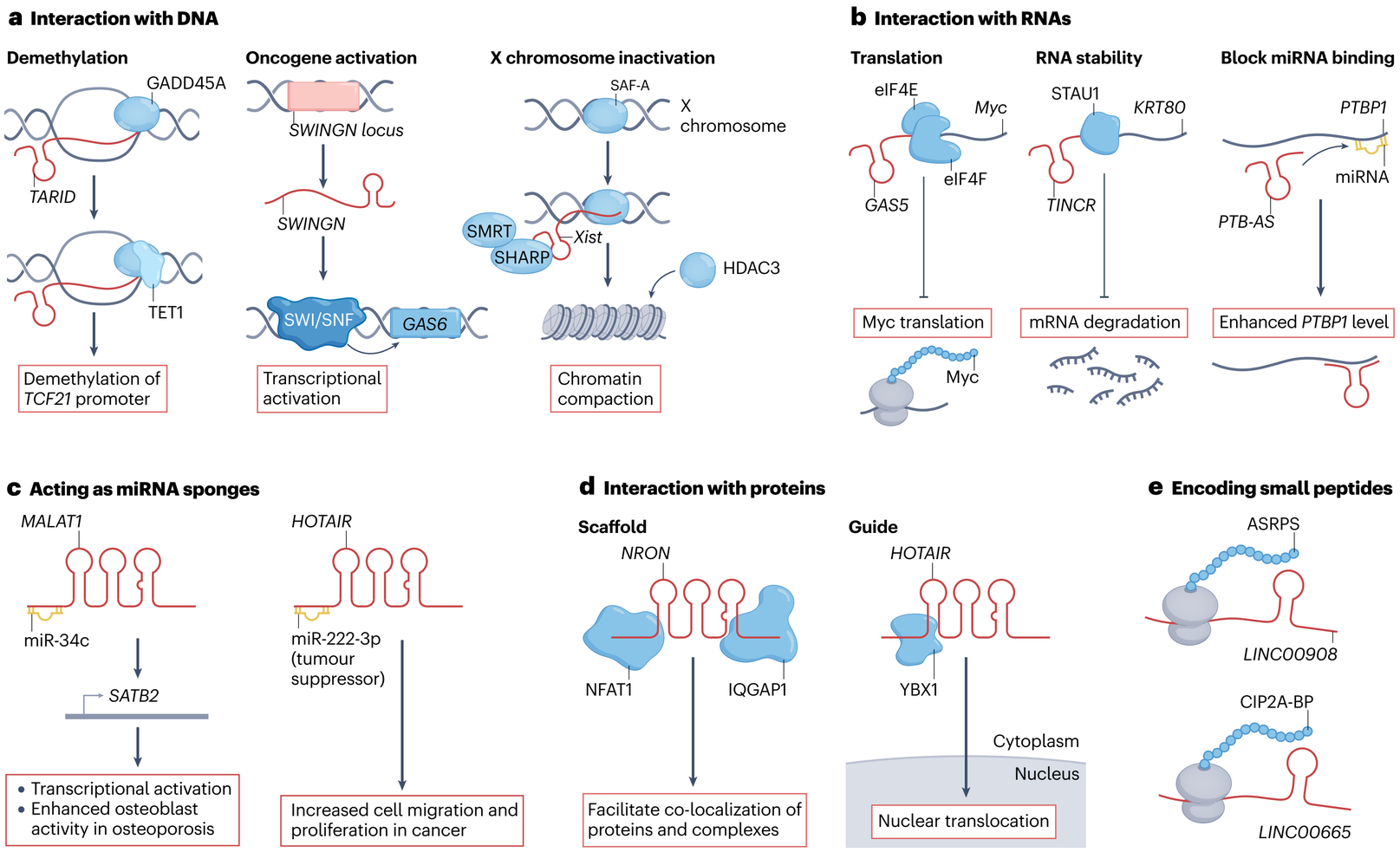
Nature Reviews Genetics, 2024, *25*(3), 211-232

Long Non-Coding RNAs



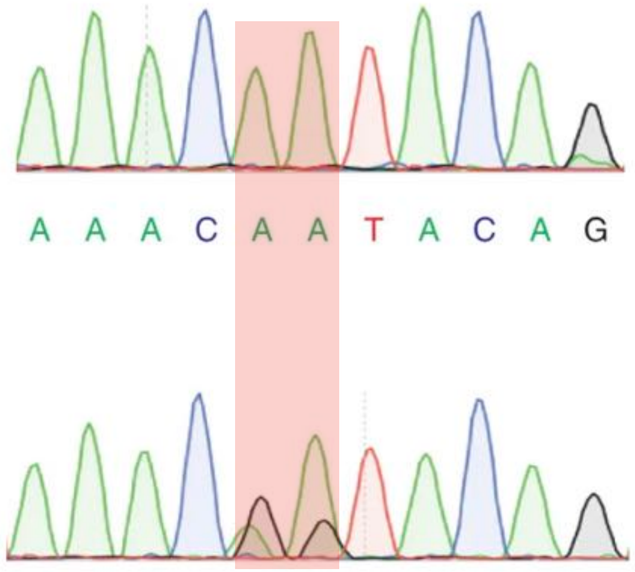
Nature Reviews Genetics, 2024, *25*(3), 211-232

Long Non-Coding RNAs



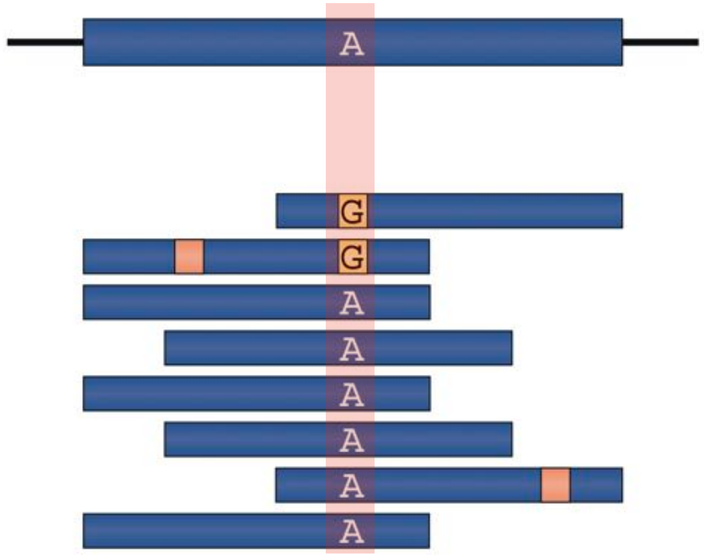
Nature Reviews Genetics, 2024, *25*(3), 211-232

Surprise #2 - RNA is Extensively “Edited” After Transcription

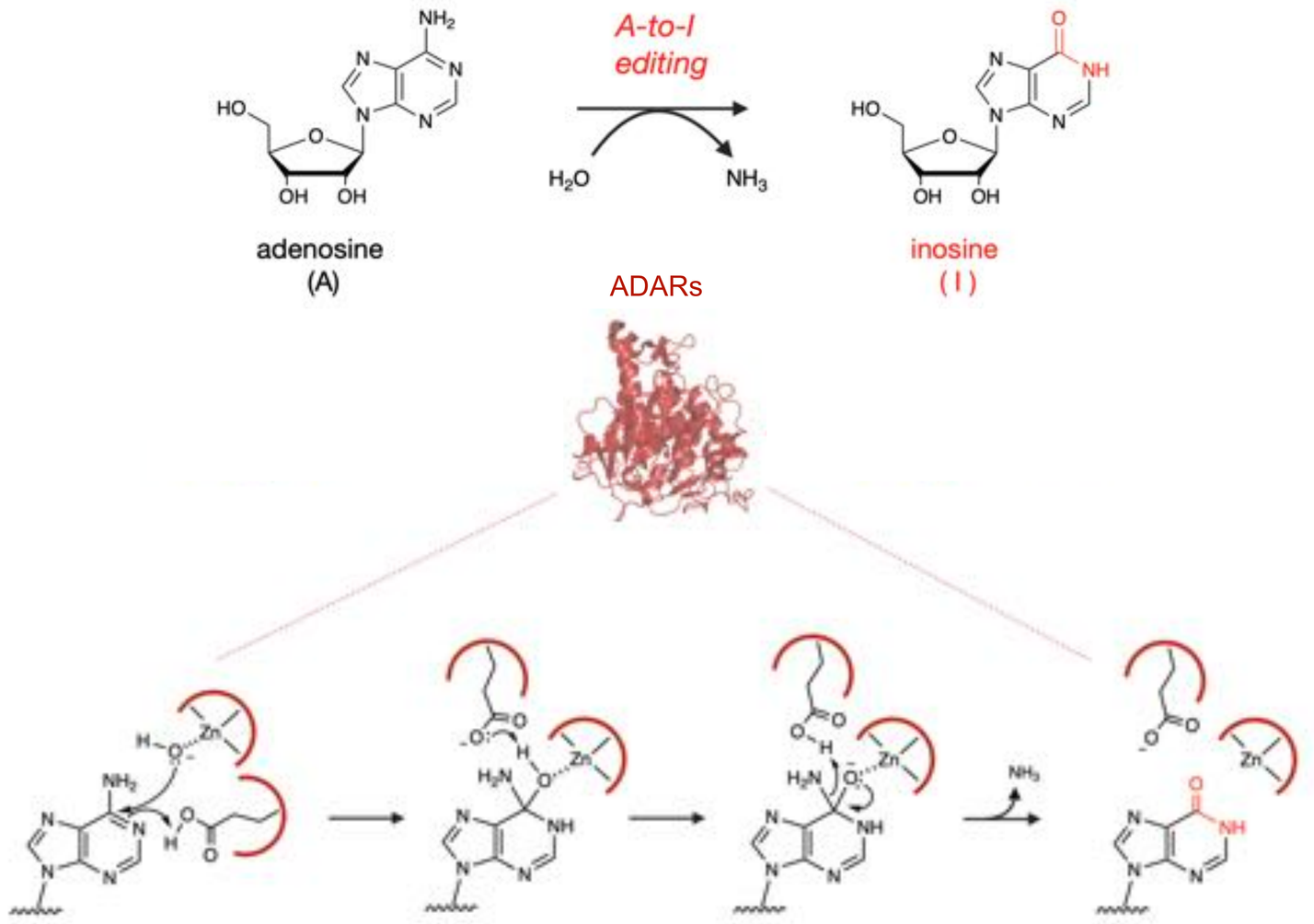


DNA sequence

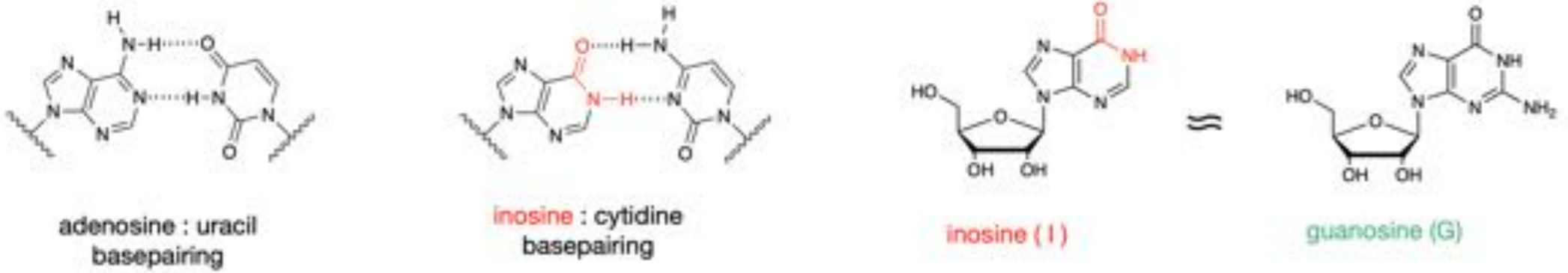
RNA sequence



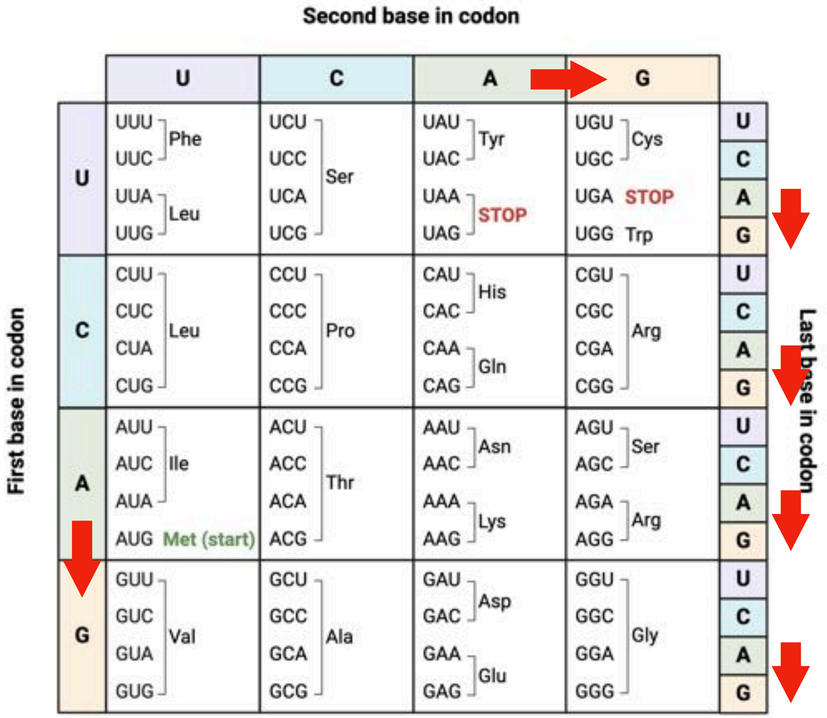
A-to-I RNA Editing

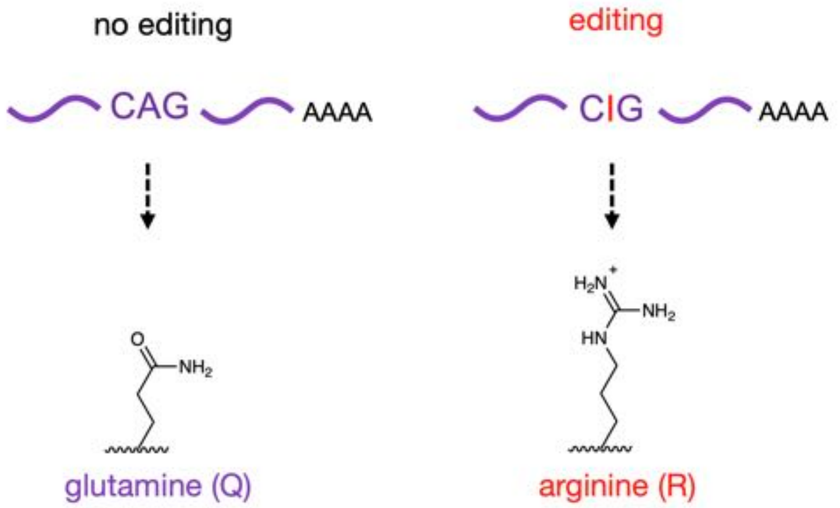


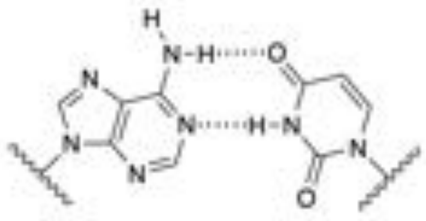
A-to-I RNA Editing



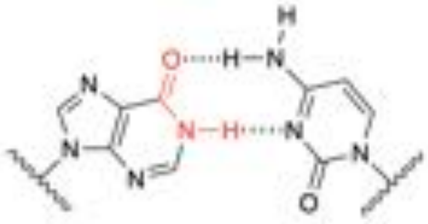
A-to-I RNA Editing



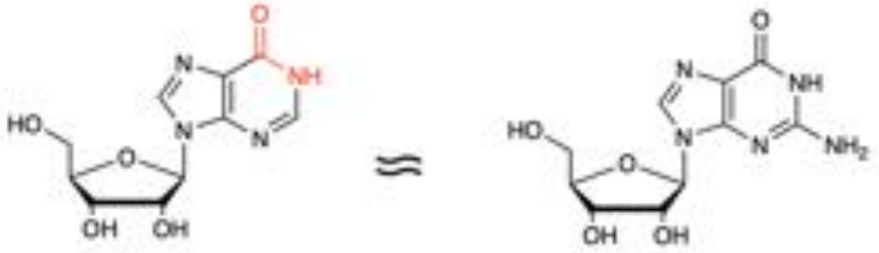




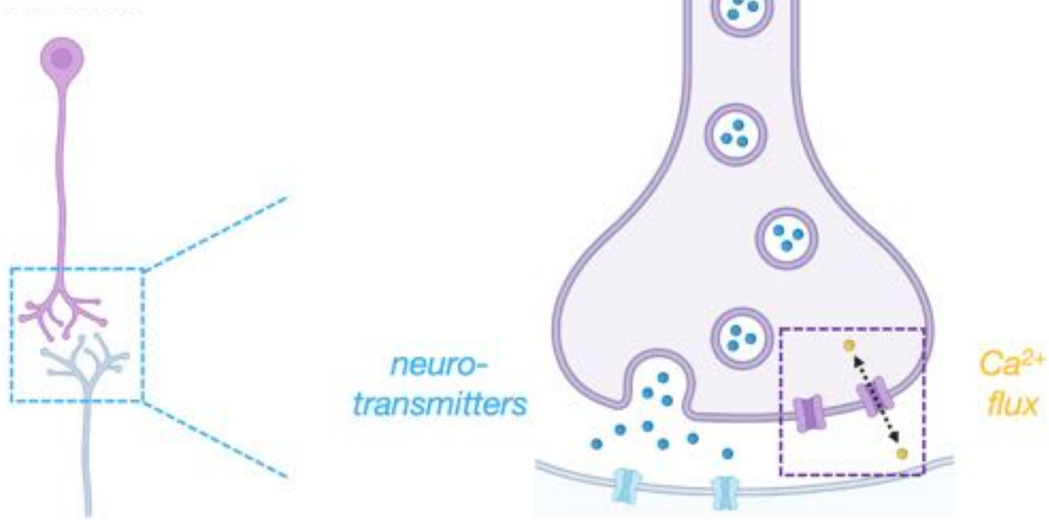
adenosine : uracil basepairing

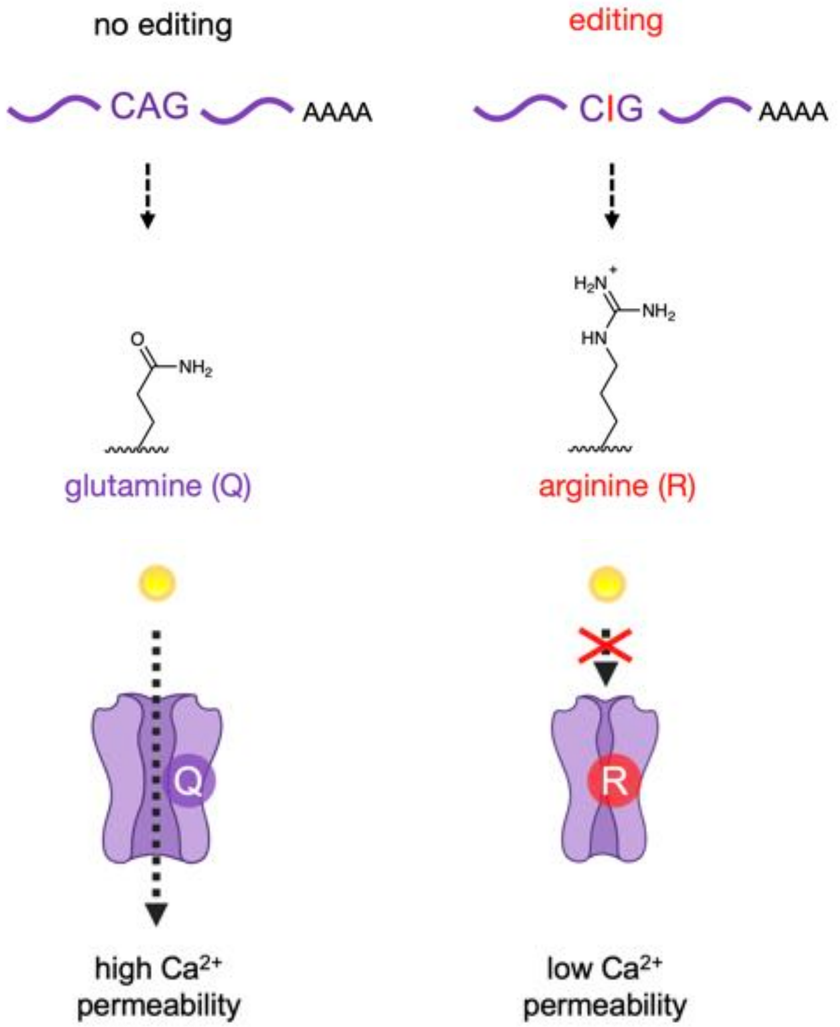


inosine : cytidine basepairing

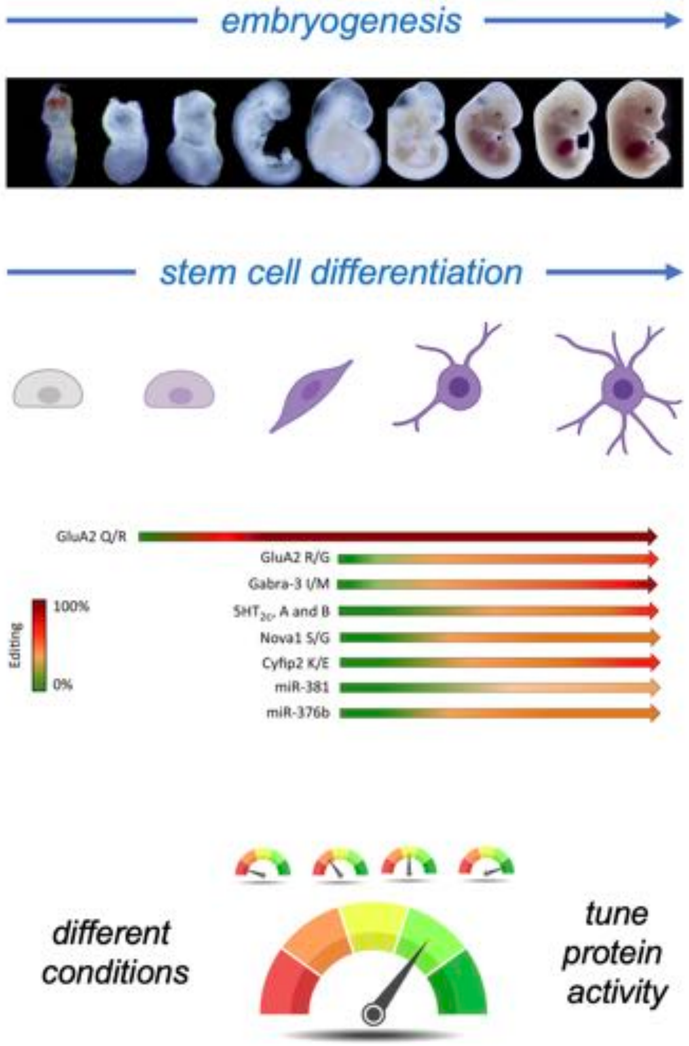


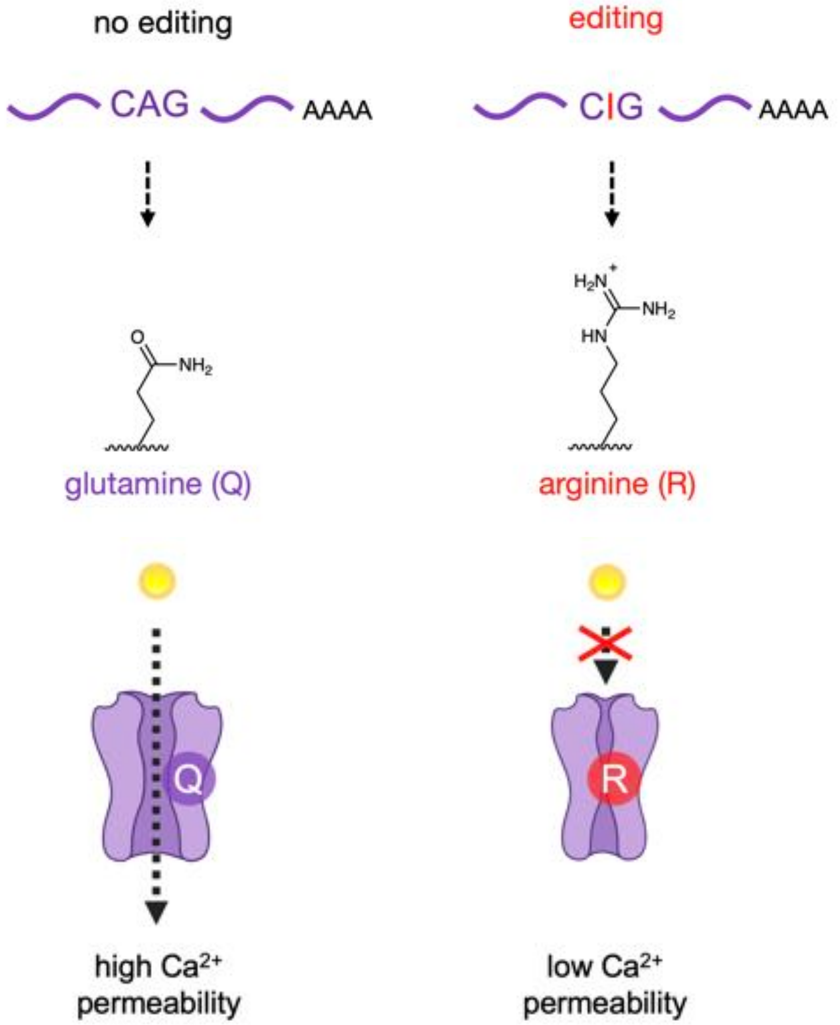
A-to-I RNA Editing



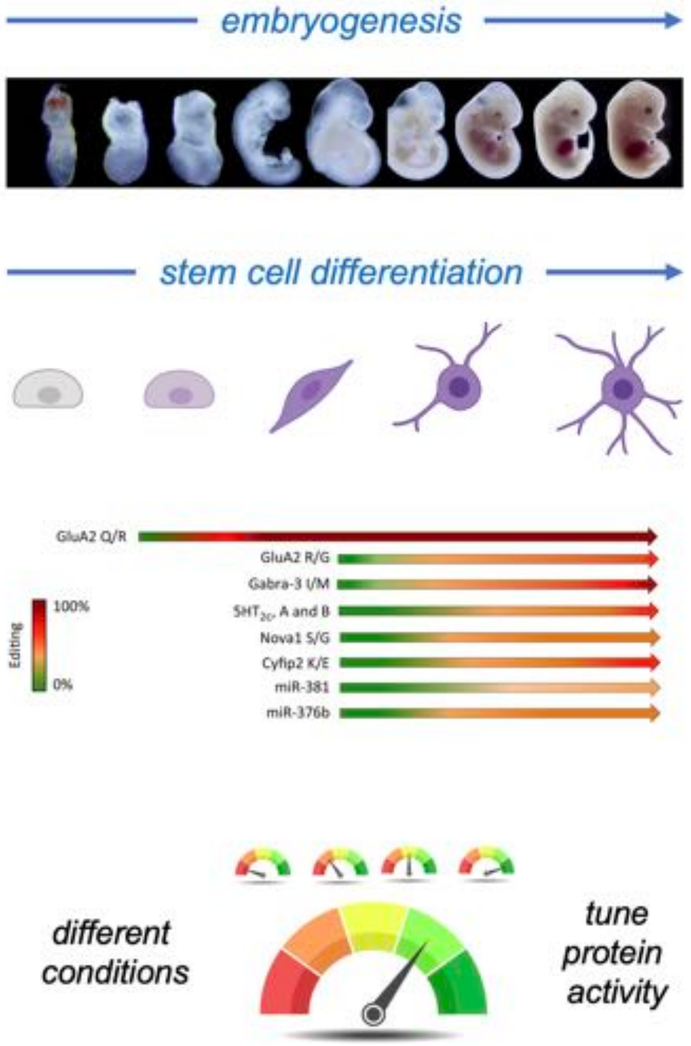


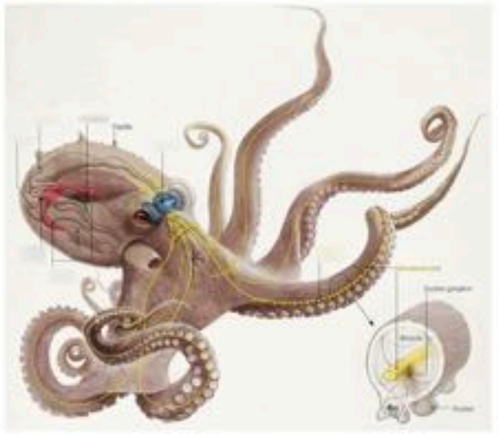
A-to-I RNA Editing

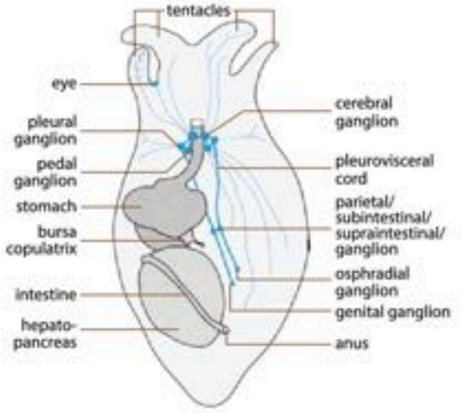


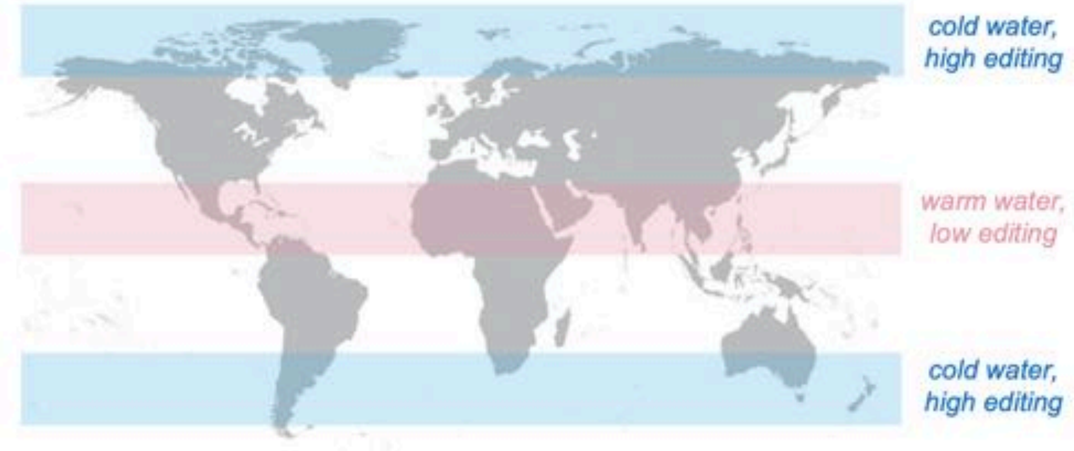


A-to-I RNA Editing







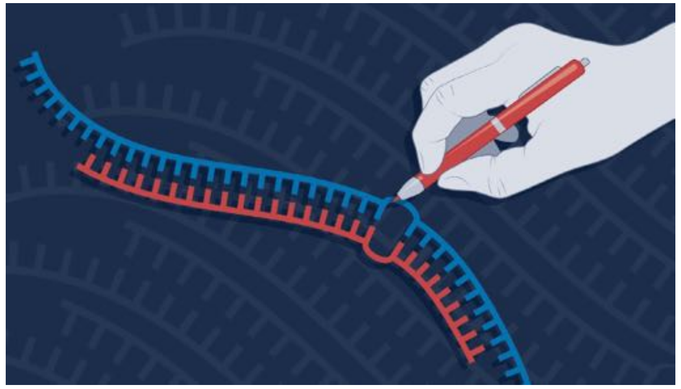


Site-directed A-to-I RNA Editing

RNA-editing drugs advance into clinical trials

By Asher Mullard

ADAR-based editors that can change the mRNA code offer new opportunities in both rare genetic diseases and common complex ones.



Oligonucleotide-based drugs already come in many flavours. The newest of these now aims to edit mRNA one base at a time, by harnessing endogenous enzymes called adenosine deaminases acting on RNA (ADAR).

Wave Life Sciences advanced the first ADAR-based RNA editor into healthy volunteers in 2023 for the hereditary disorder alpha-1 antitrypsin deficiency (AATD). The company is set to start dosing patients with the disease shortly. A growing list of biotechs are setting their sights on similar RNA-editing

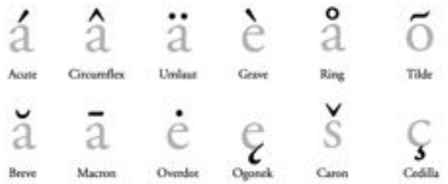
Table 1 | ADAR-based editors in and approaching the clinic

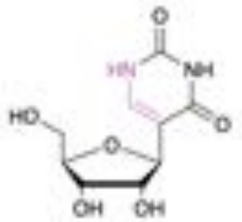
Drug	Sponsor	Properties	Lead indication	Status
WVE-006	Wave Life Sciences/GSK	*SERPINA1*/AAT mRNA editor	Alpha-1 antitrypsin deficiency	Phase I
AX-1412	ProQR	B4GALT1 mRNA editor	Cardiovascular disease	To start late 2024/early 2025
AX-0810	ProQR	NTCP mRNA editor	Cholestatic diseases	To start late 2024/early 2025
KRRO-110	Korro Bio	*SERPINA1*/AAT mRNA editor	Alpha-1 antitrypsin deficiency	IND in 2024
NA	ADARx	*SERPINA1*/AAT mRNA editor	Alpha-1 antitrypsin deficiency	Preclinical
NA	AIRNA	*SERPINA1*/AAT mRNA editor	Alpha-1 antitrypsin deficiency	Preclinical
NA	Vico Therapeutics	MECP2-R255X mRNA editor	Rett syndrome	Preclinical
NA	EdiGene	Undisclosed	Undisclosed	Undisclosed
NA	ShapeTx	Undisclosed	Undisclosed	Undisclosed

KORRO BIO

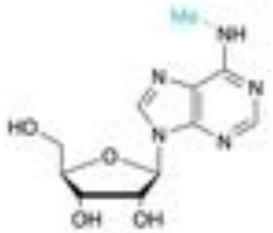
WAVE® LIFE SCIENCES

Epitranscriptomics

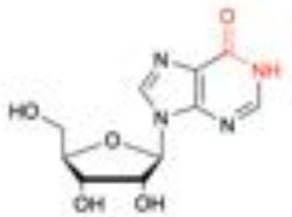




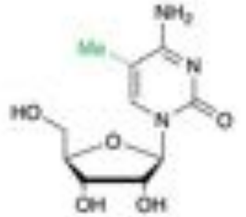
Pseudouridine
(Ψ)



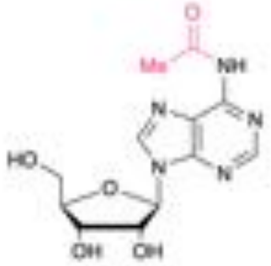
N^6-methyladenosine
(m^6A)



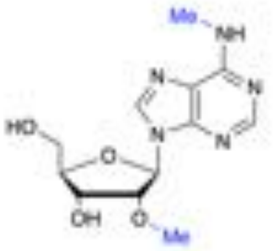
Inosine
(I)



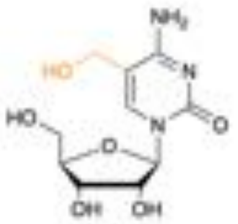
5-methylcytidine
(m^5C)



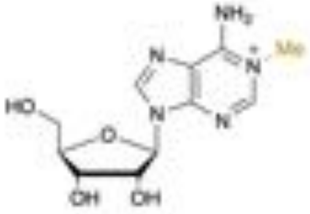
N^6-acetyladenosine
(ac^6A)



N^6, 2'-*O*-dimethyladenosine
(m^6Am)

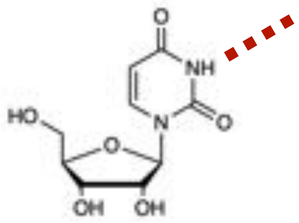


5-hydroxylmethylcytidine
(hm^5C)



N^1-methyladenosine
(m^1A)

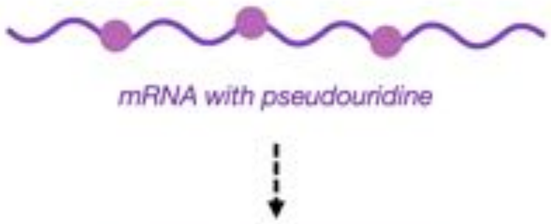
Epitranscriptomics



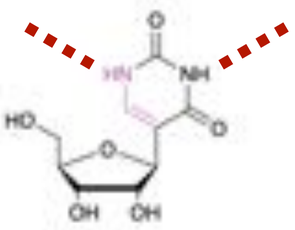
Uridine (U)

tRNA with pseudouridine

structural stability

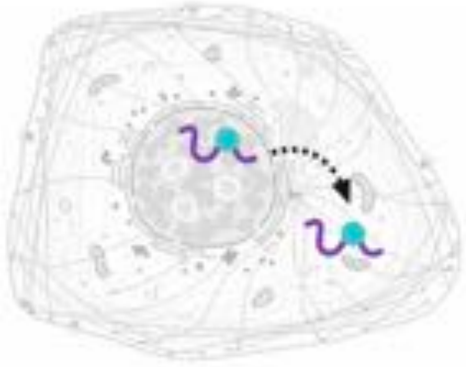


enhanced protein expression

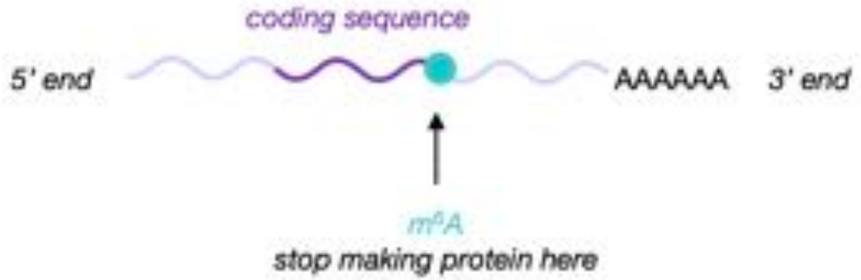


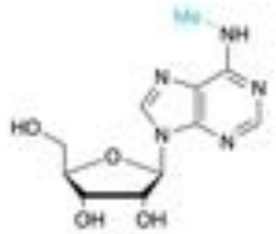
Pseudouridine
(Ψ)

Epitranscriptomics



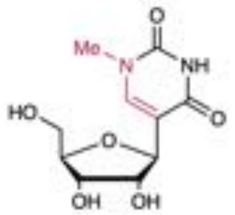
RNA export, enhanced stability, increased translation





N⁶-methyladenosine (m^6A)

Epitranscriptomics

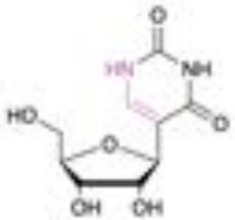


N^1-methylpseudouridine
($m^1\Psi$)

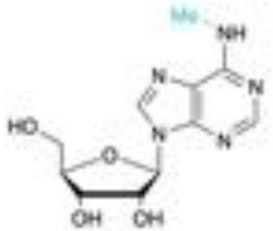
moderna

RNA encoding virus protein piece

enhanced mRNA stability,
increased protein production,
prevents adverse immune responses



Pseudouridine
(Ψ)



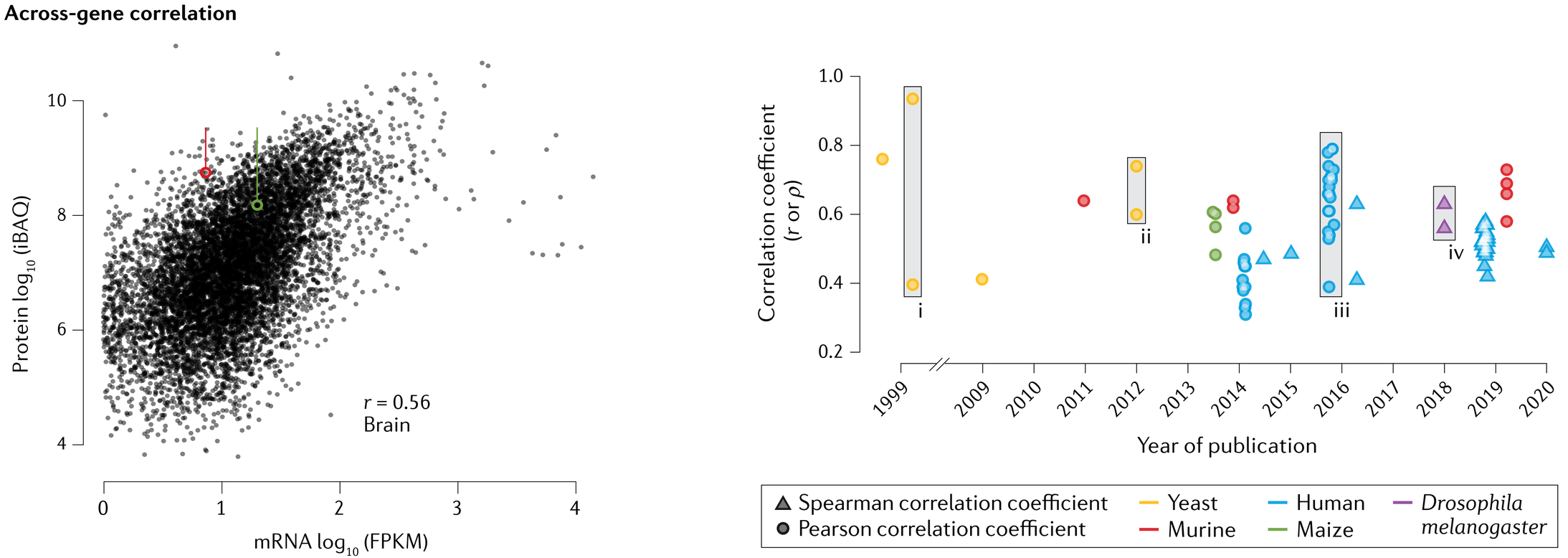
N^6-methyladenosine
(m^6A)

Inosine
(I)

5-methylcytidine
(m^5C)

Global transcriptomics + proteomics

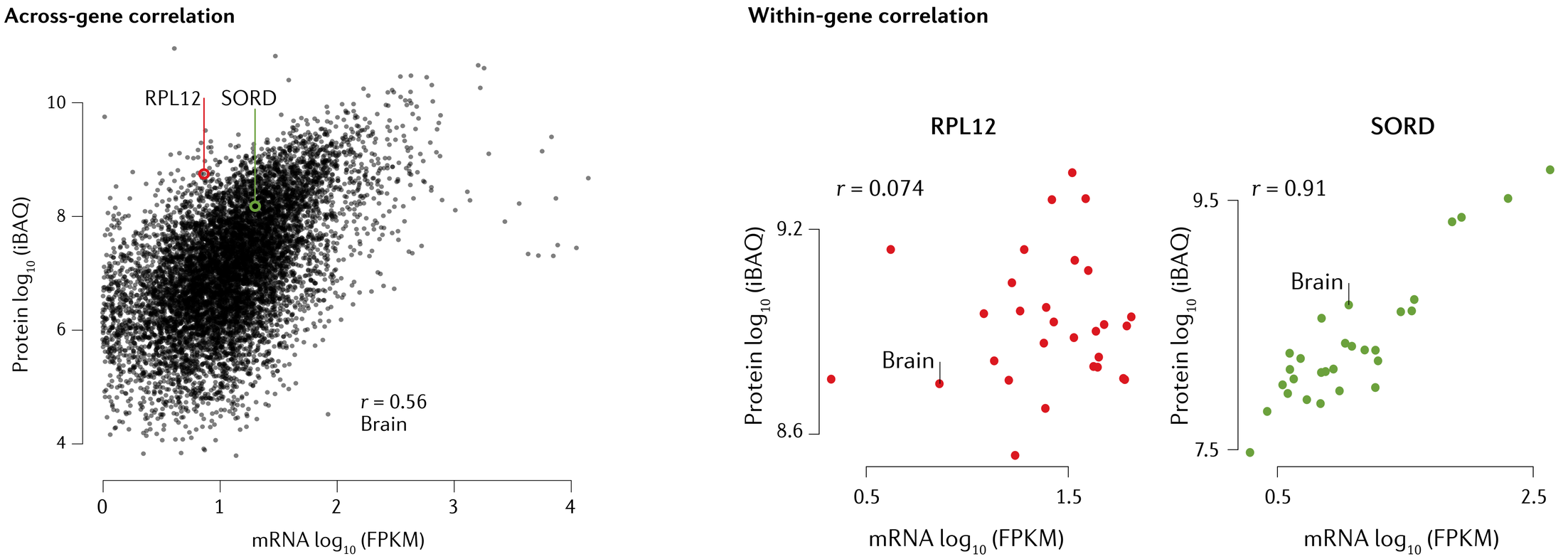
how well does mRNA correlate with protein?



Nature Reviews Genetics, 2024, *21*, 630-644

Global transcriptomics + proteomics

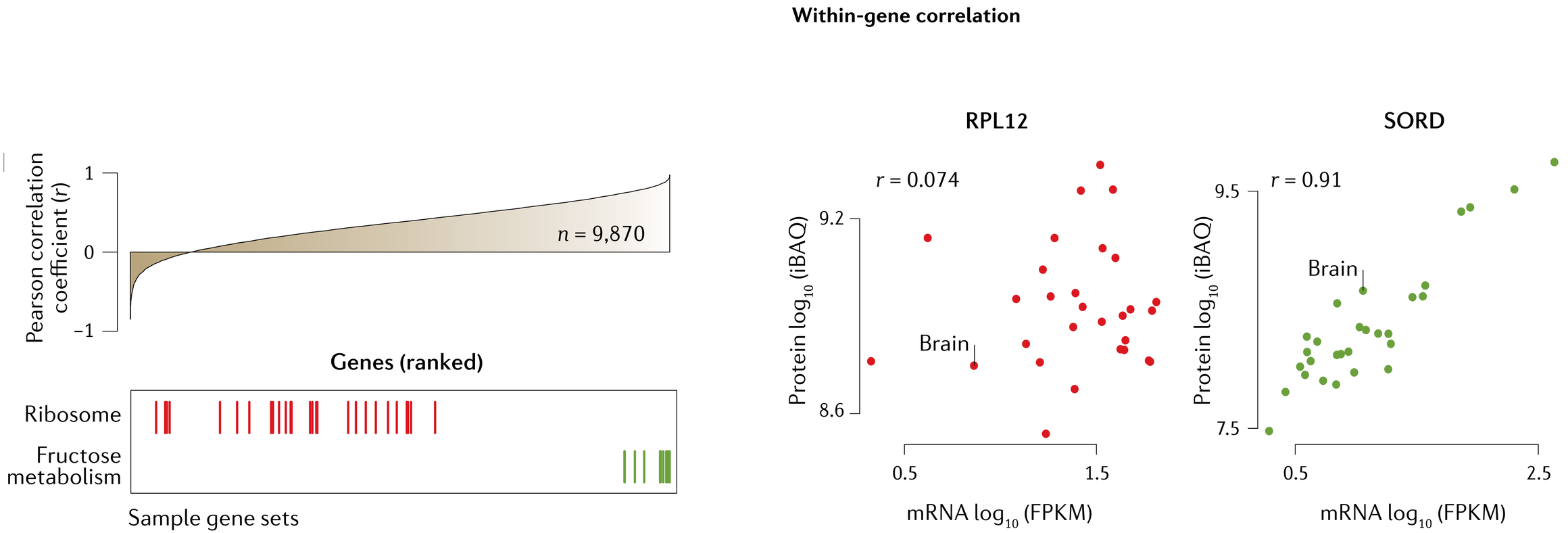
how well does mRNA correlate with protein?



Nature Reviews Genetics, 2024, *21*, 630-644

Global transcriptomics + proteomics

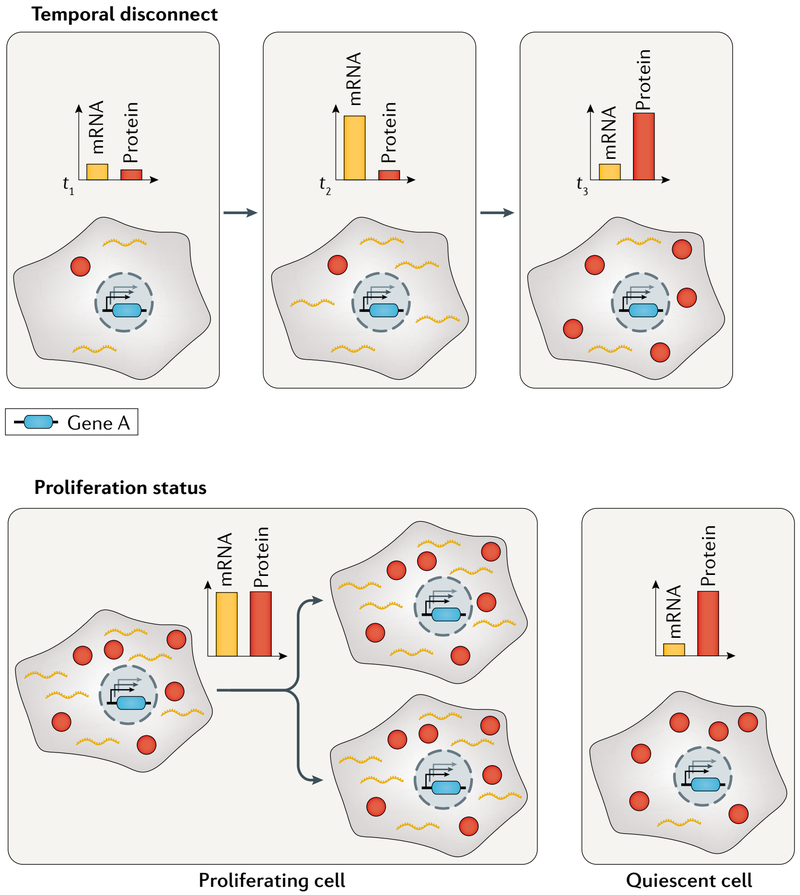
how well does mRNA correlate with protein?

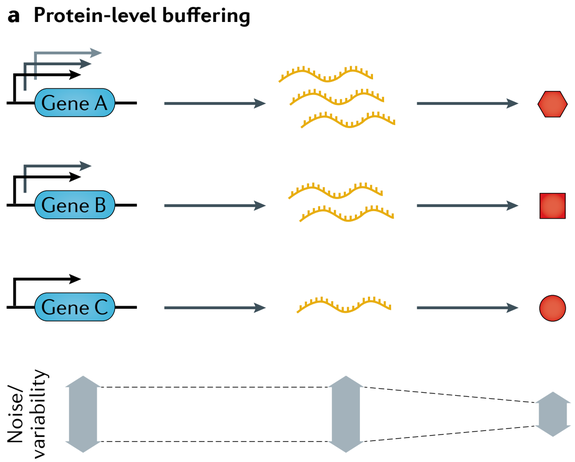


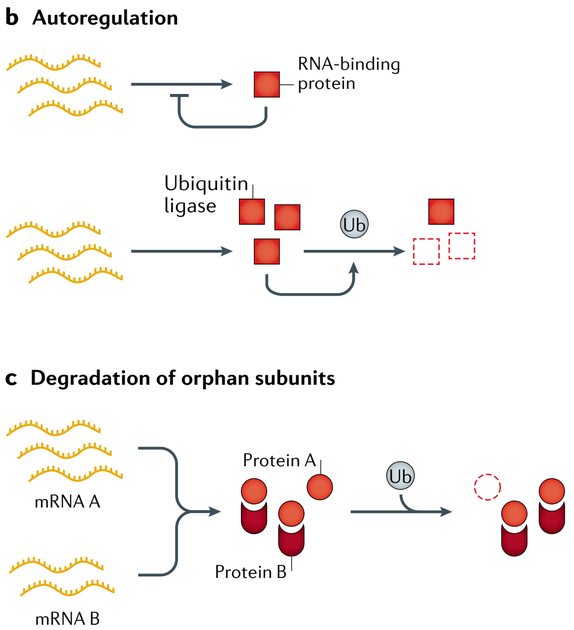
Nature Reviews Genetics, 2024, *21*, 630-644

Global transcriptomics + proteomics

how well does mRNA correlate with protein?

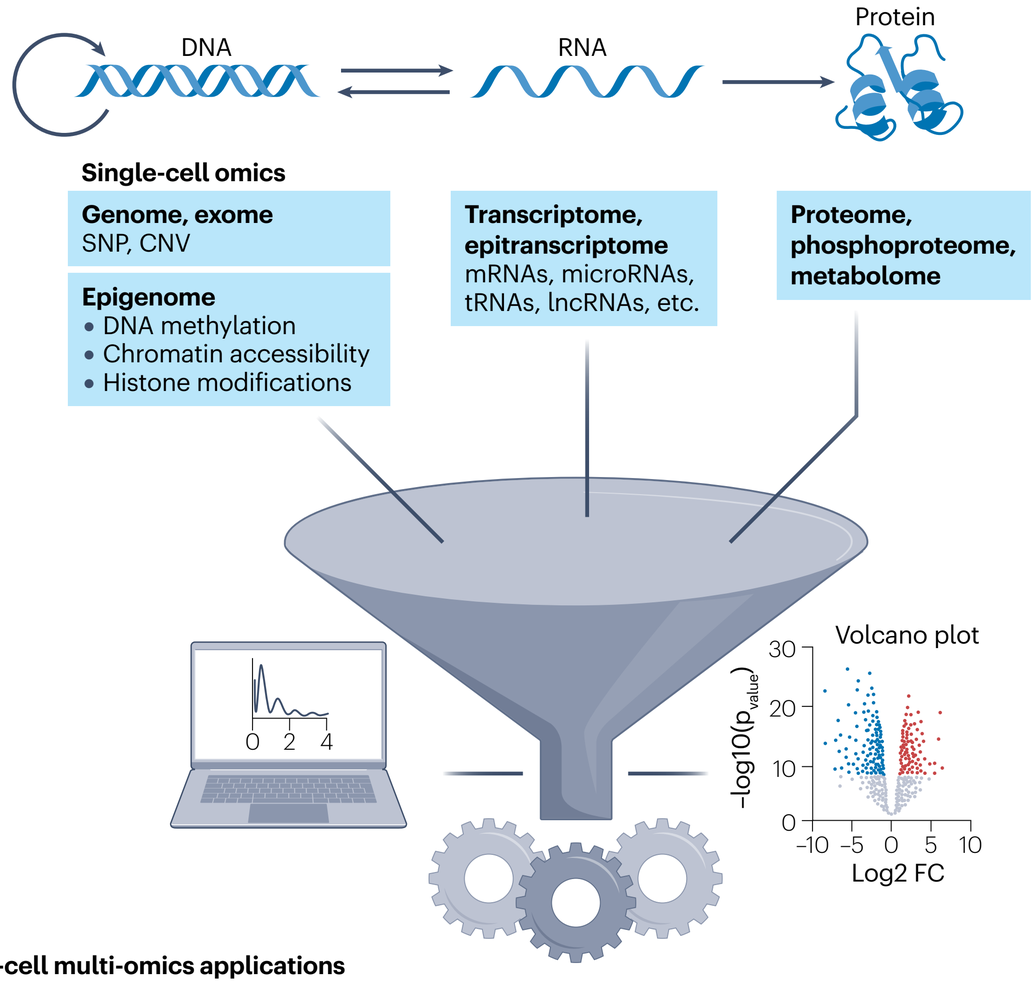






Nature Reviews Genetics, 2024, *21*, 630-644

MultiOMICs

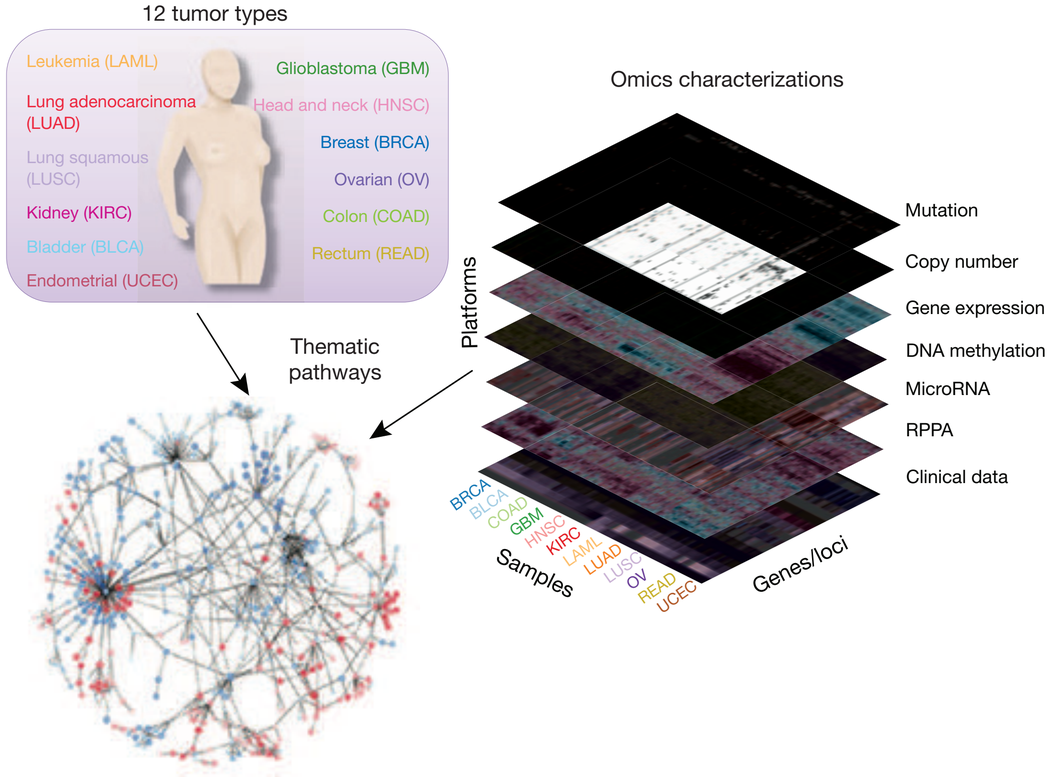


Single-cell multi-omics applications

Discovery of novel cell types	Tissue and tumour heterogeneity	Atlas generation	Biomarker discovery	Insights into complex diseases	Novel pathways and networks

MultiOMICs

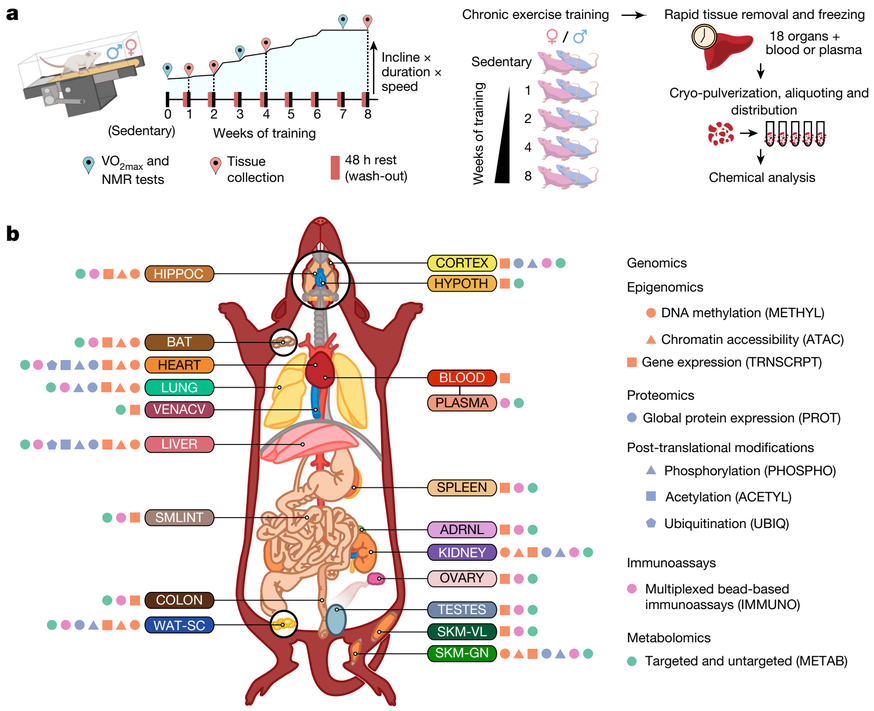
The Cancer Genome Atlas



- *HER2 amplification landscape in breast cancer*
- *KRAS and NRAS mutations*
- *microRNAs and long non-coding RNAs (lncRNAs) across cancers*
- *BRAF V600E mutations in melanoma and thyroid cancers*
- *MGMT promoter methylation in glioblastoma*
- *IDH1/2 mutations in low-grade gliomas*
- *oncometabolites like 2-hydroxyglutarate (2HG) in IDH-mutant gliomas*

MultiOMICs

Temporal dynamics of the multi-omic response to endurance exercise training



Nature, 2024

A multiomic atlas of the aging hippocampus reveals molecular changes in response to environmental enrichment

Nat. Comms., 2024

Nonlinear dynamics of multi-omics profiles during human aging

Nature Aging, 2024

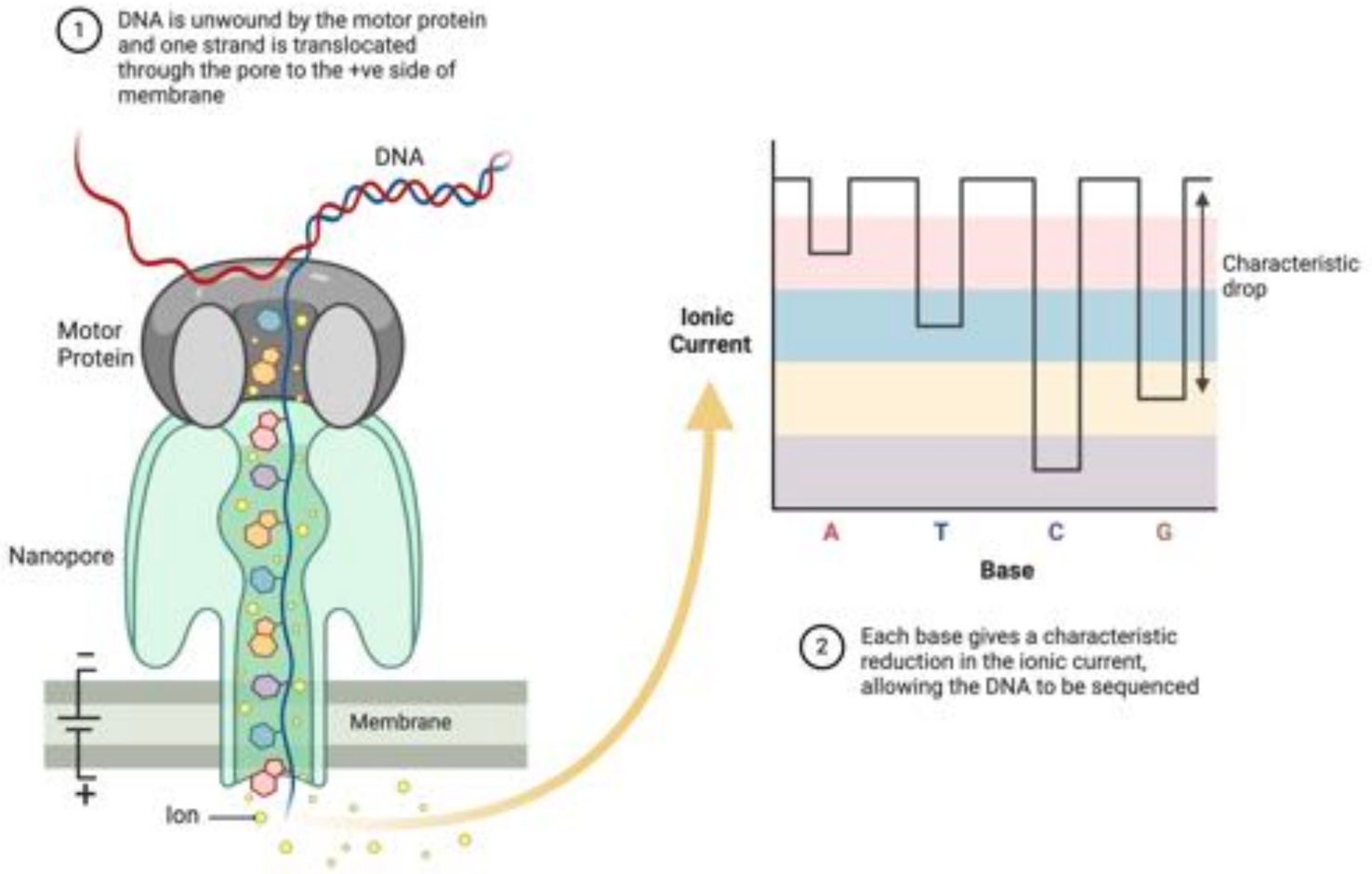
Host-microbe multiomic profiling reveals age-dependent immune dysregulation associated with COVID-19 immunopathology

Science, 2024

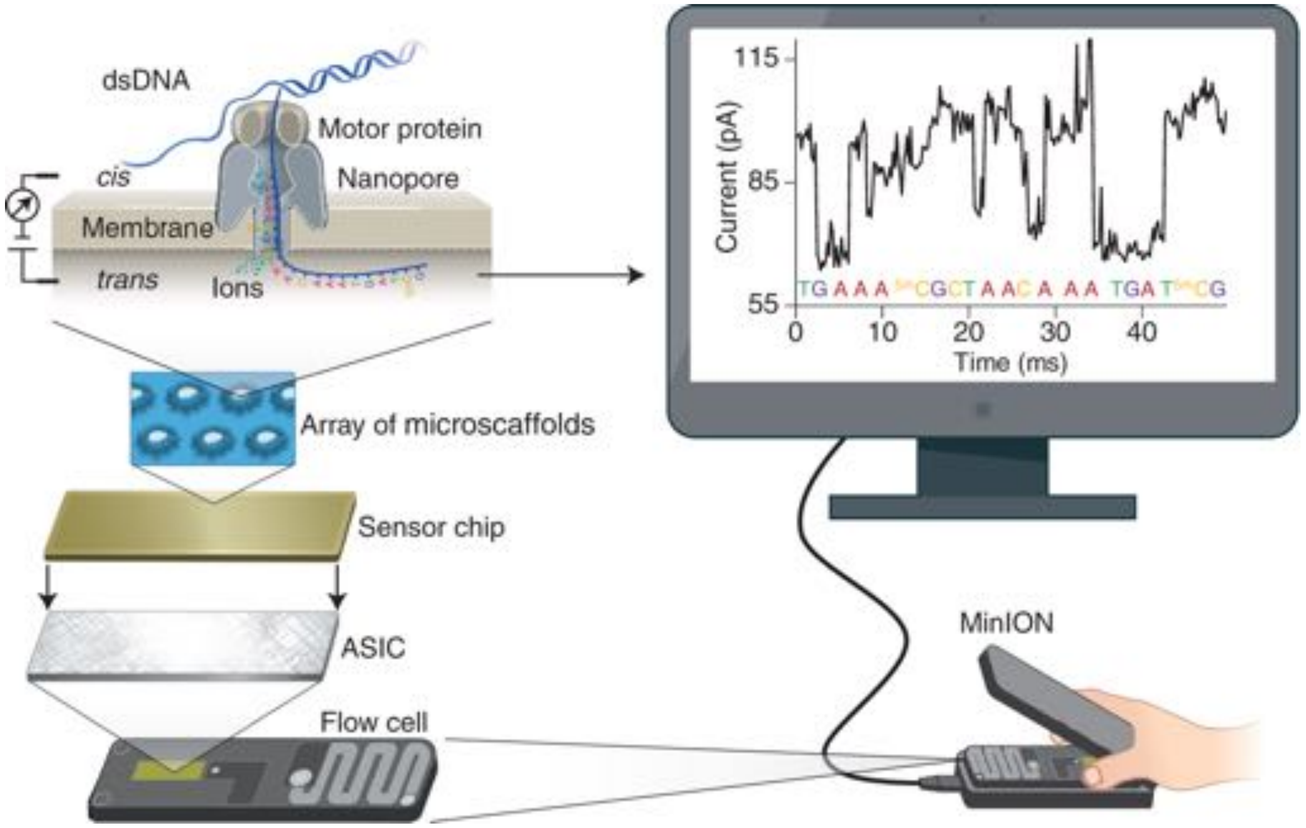
Multi-omic applications for understanding and enhancing tropical fruit flavour

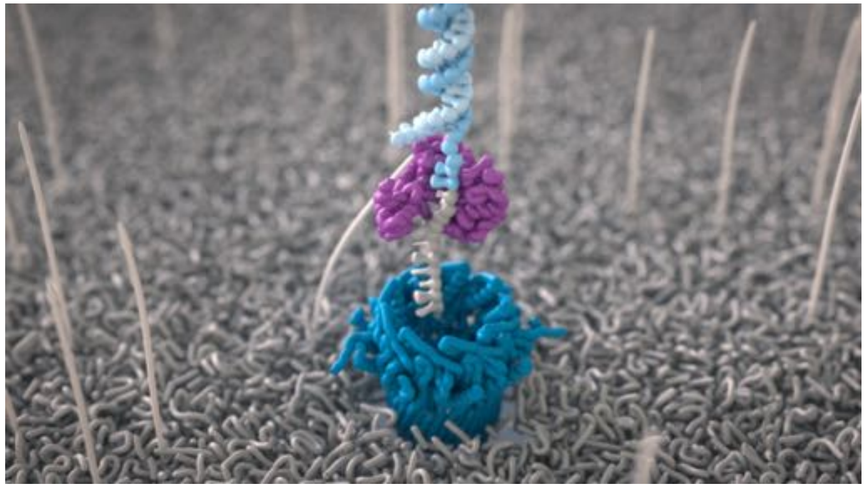
Plant Molecular Biology, 2024

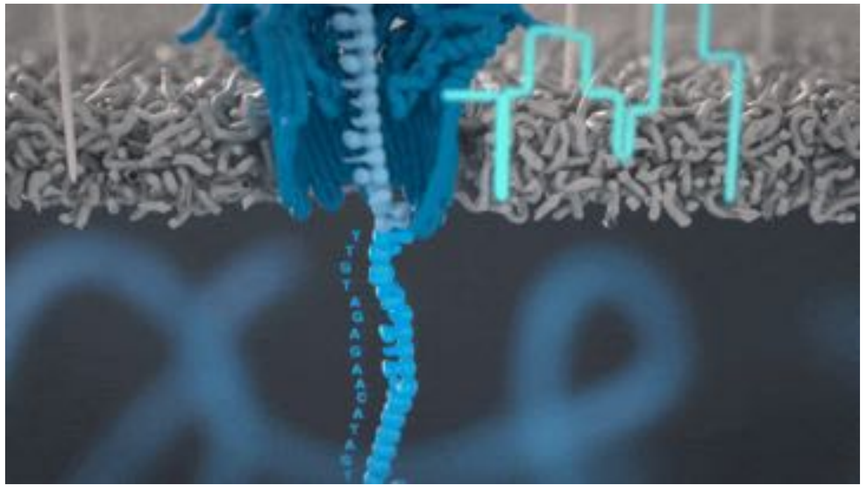
Third-generation sequencing



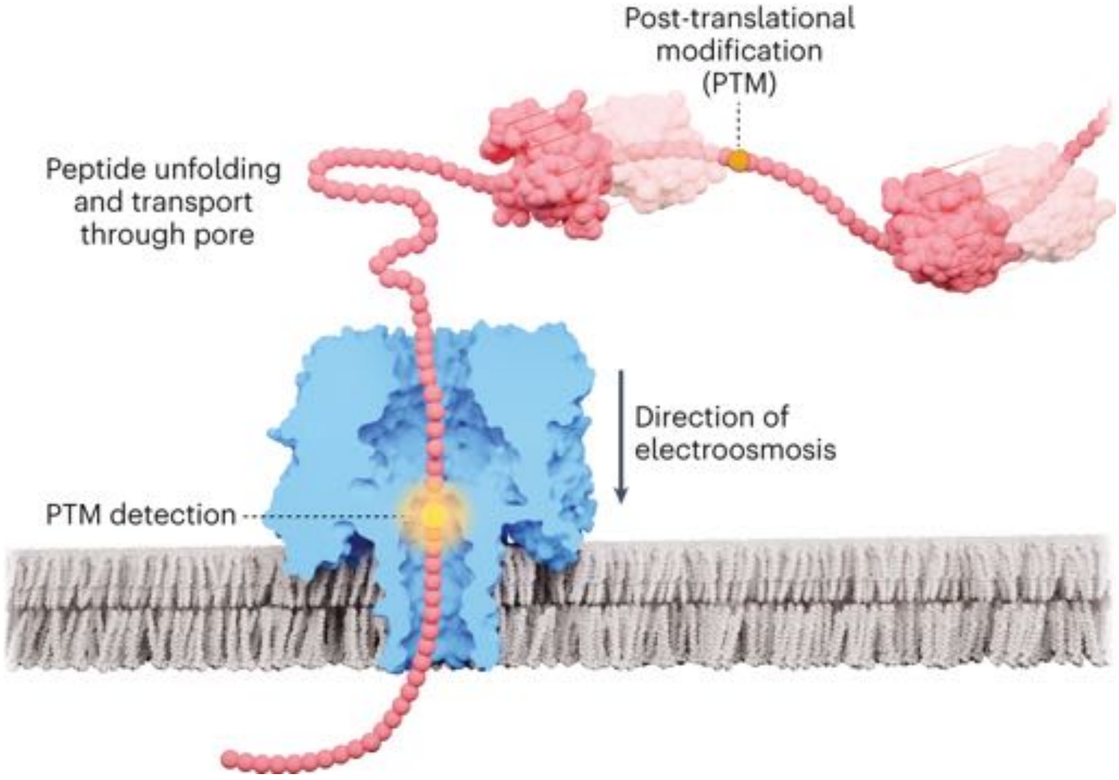
Third-generation sequencing







Third-generation sequencing

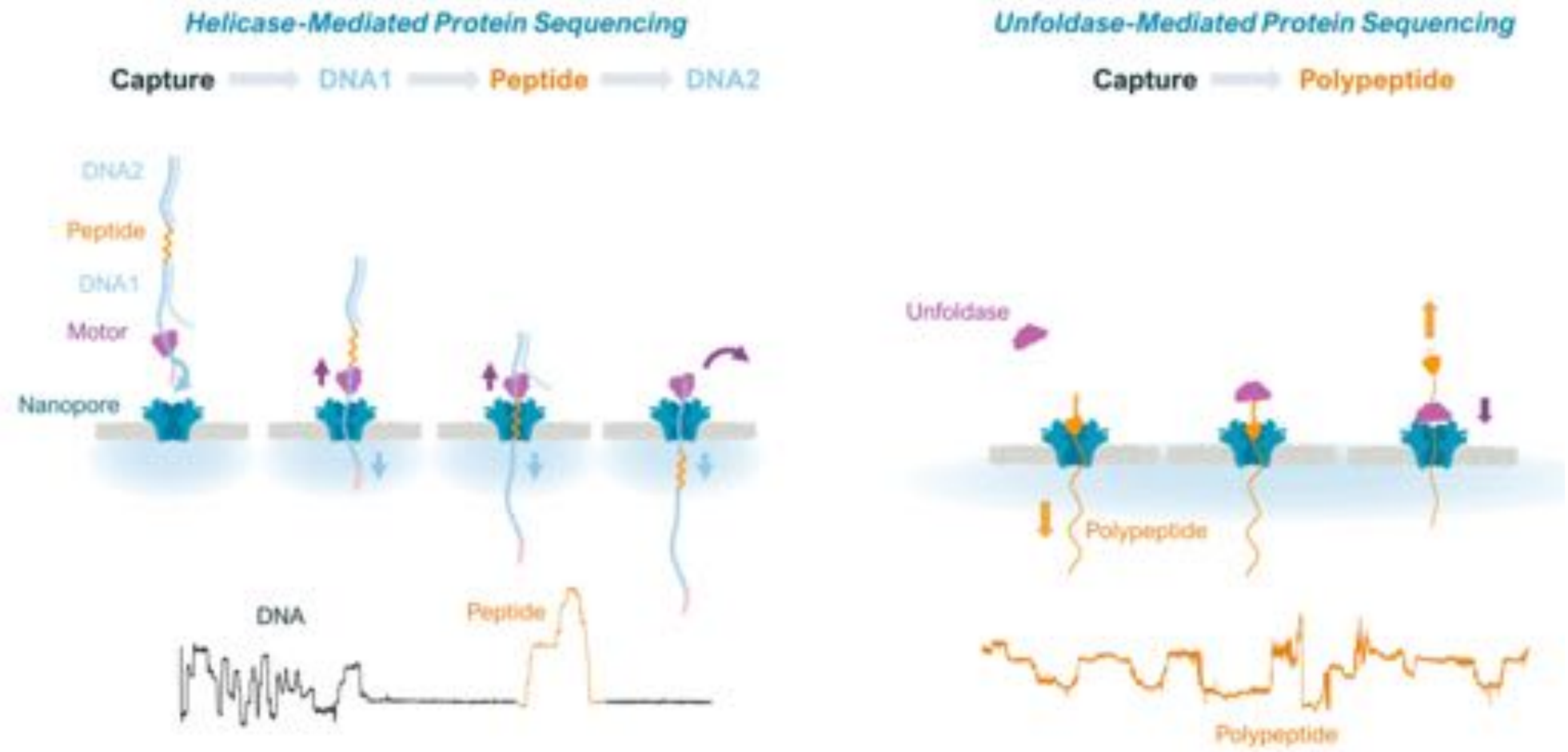


Unfolding the path to nanopore protein sequencing

Nanopore sequencing set to transform our understanding of proteins

Not if but when nanopore protein sequencing meets single-cell proteomics

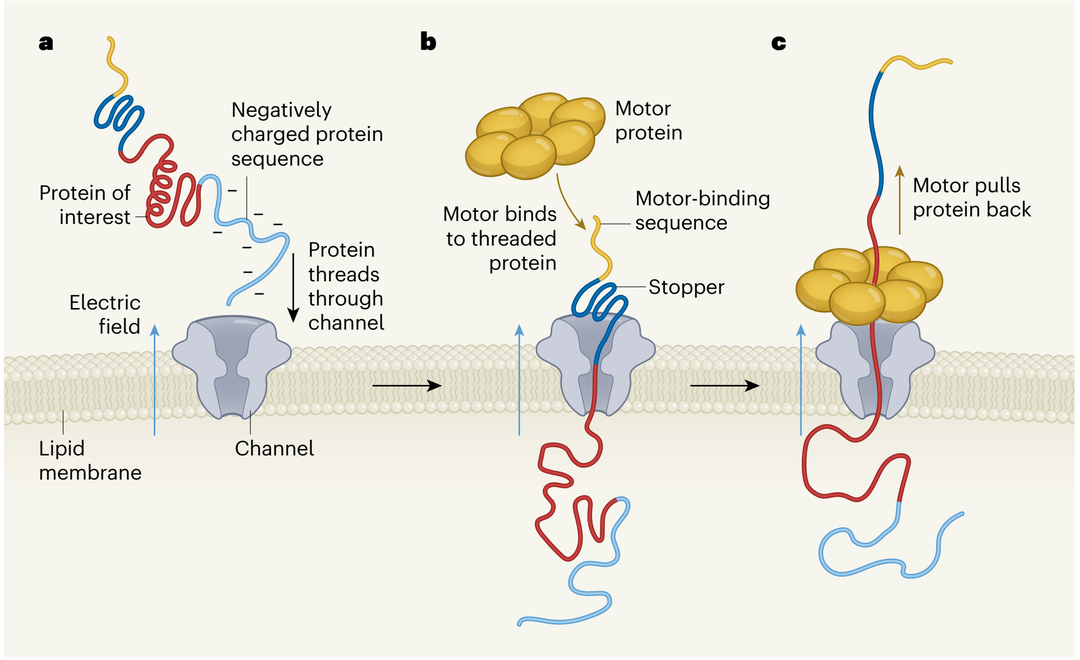
Third-generation sequencing

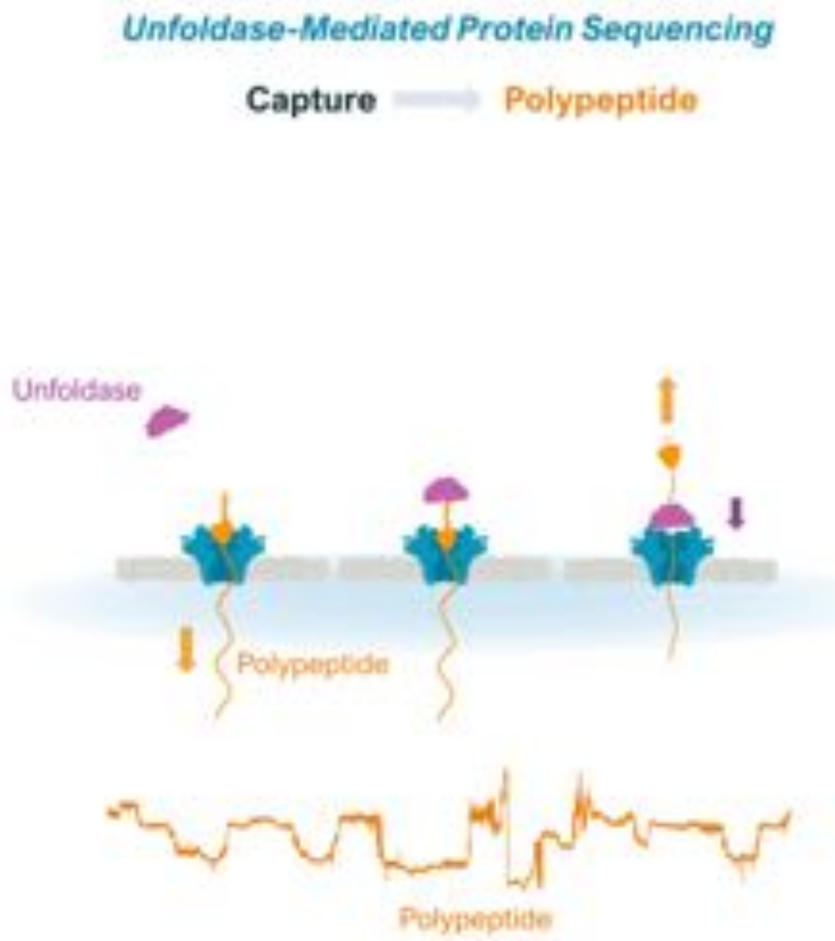


Third-generation sequencing

Multi-pass, single-molecule nanopore reading of long protein strands

Nature, 2024





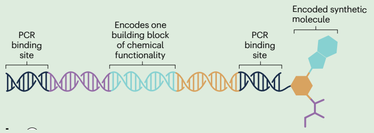
*The OMICs **outlook***

The MVP

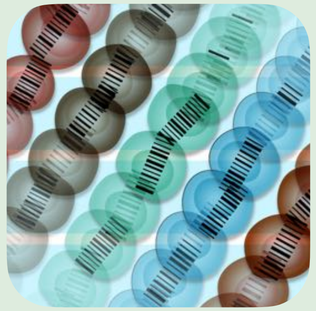
DNA sequencing/barcoding

depmap | Cancer Dependency Map

TCGA



PRISM
MULTIPLEXED CELL LINE PROFILING

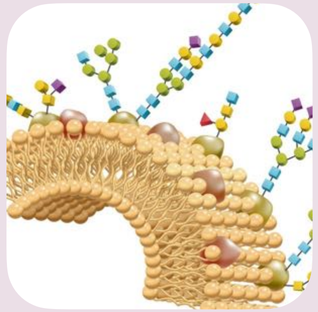


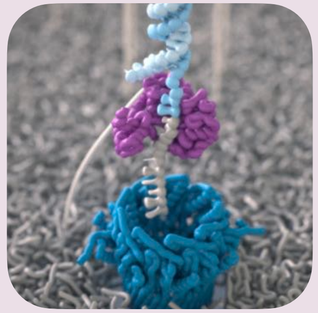
Technological Gaps

some, but not many remain

top needs:

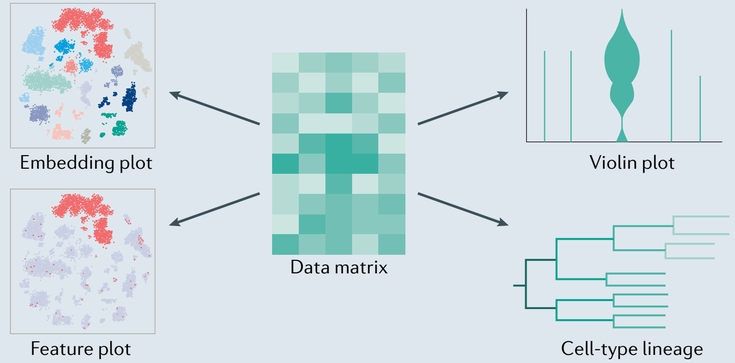
1. *direct protein sequencing*
2. *glycomics/lipidomics*
3. *accurate metabolite ID*

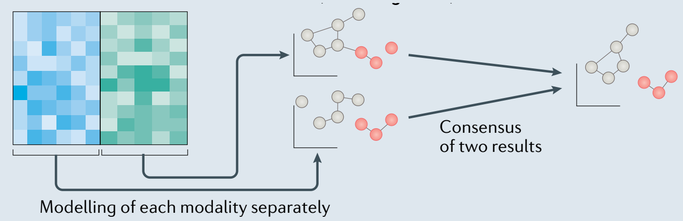




Data Analysis

we have “too much” data





Thank you

Thank you