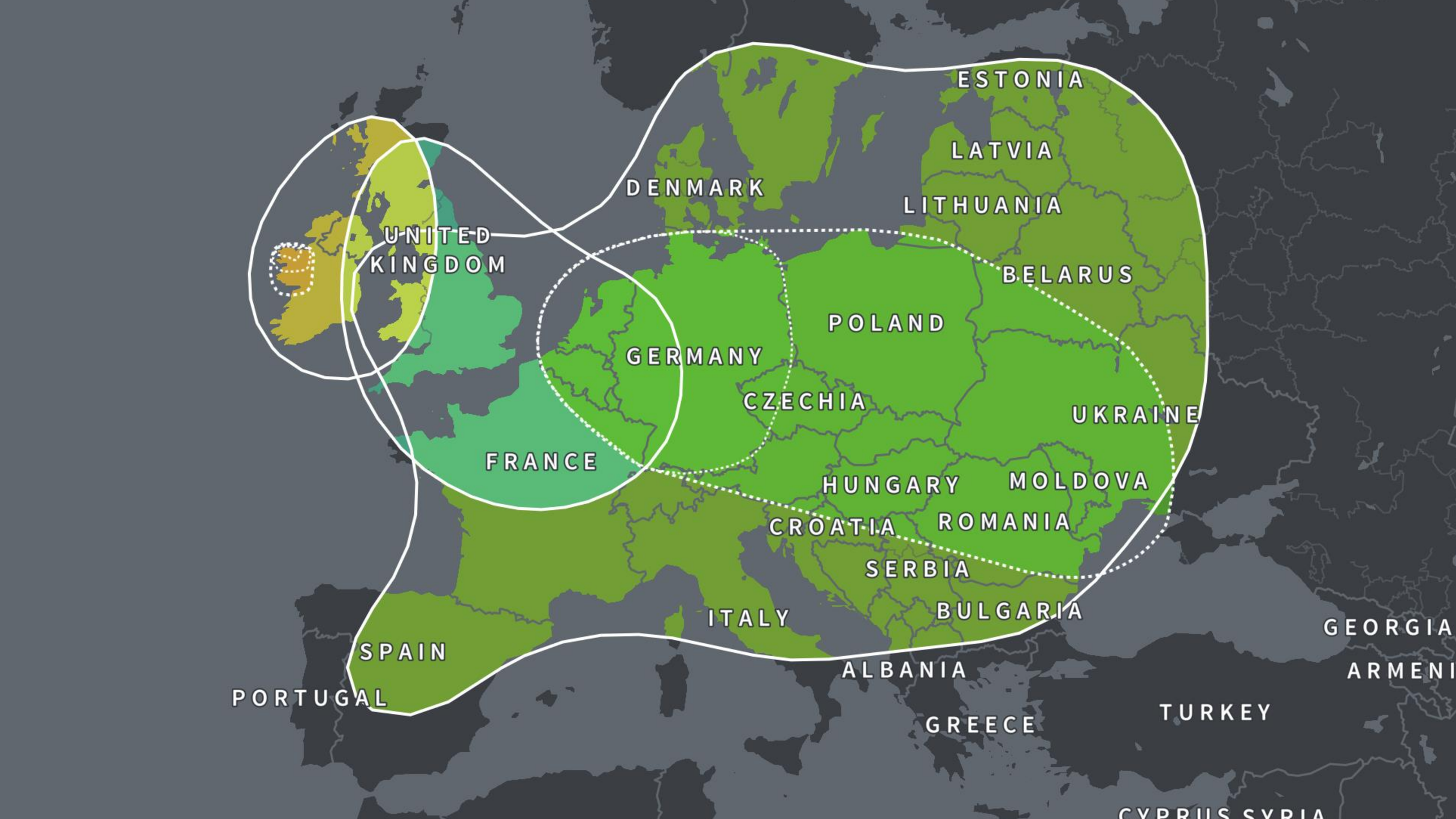
Welcome to The Age of Personal Genomes

Carl Zimmer
carlzimmer.com
March 8, 2023, Yale MBB 452


ESTONIA
LATVIA
DENMARK
LITHUANIA
UNITED KINGDOM
BELARUS
POLAND
GERMANY
CZECHIA
UKRAINE
FRANCE
HUNGARY
MOLDOVA
ROMANIA
CROATIA
SERBIA
BULGARIA
ITALY
GEORGIA
SPAIN
ALBANIA
ARMENI
PORTUGAL
TURKEY
GREECE
CYPRUS SYRIA

Our History

THE EARLY YEARS | GLOBAL GROWTH & GENOMES | THE TRANSFORMATIVE YEARS | **THE NEXT GENERATION**

2020

Ancestry reaches over 3.5 million subscribers and 27 billion records.

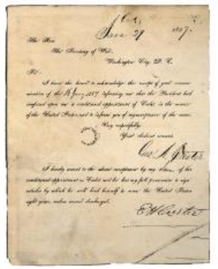
2021

AncestryDNA reaches a new milestone with over 20 million people in its DNA network.

ancestryDNA®

2021

Ancestry hits over 30 billion records in its online database.



2022

AncestryDNA expands to new global markets in North America, South America, Europe, Asia and Africa making AncestryDNA available in more than 80 markets on five continents.

https://www.ancestry.com/corporate/about-ancestry/our-story

Late-Onset Alzheimer's Disease

Alzheimer's disease is characterized by memory loss, cognitive decline, and personality changes. Late-onset Alzheimer's disease is the most common form of Alzheimer's disease, developing after age 65. Many factors, including genetics, can influence a person's chances of developing the condition. This test includes the most common genetic variant associated with late-onset Alzheimer's disease.

Jamie, you **do not have** the ε4 variant we tested.

Your risk for Alzheimer's disease also depends on other factors, including lifestyle, environment, and genetic variants not covered by this test.

Our Scale Enables Real-Time Genetics Health Research[1]

(numbers below represent the number of research participants with the condition indicated)

1,876,573 High cholesterol	**358,275** Type 2 Diabetes	**37,853** Type 1 Diabetes
1,785,456 Depression	**2,355,068** APOE e4 carriers (Alzheimer's risk)	**85,604** Epilepsy
1,113,057 Asthma	**667,019** Eczema	**250,764** Psoriasis
634,734 Irritable Bowel	**107,126** UC / Crohn's	**64,800** Barrett's Esophagus
534,696 Arrhythmia	**159,135** Coronary Artery	**42,836** Pulmonary Embolism
9,047 Systemic Sclerosis	**7,334** Sarcoidosis	**4,528** Idiopathic Pulmonary Fibrosis

[1] As of August 2, 2021. [2] As of September 2021. [3] 23andMe COVID-19 manuscript live on MedRXiv September 7, 2020.

https://investors.23andme.com/

1,287,060[2]
COVID-19 study participants

750K
Consumers participated in the COVID-19 study in the **first 90 days**

COVID-19 Research (2020)

- **March 16** Kicked Off Study
- **April 6** Launched Study
- **June 8** Preliminary Findings
- **Sept. 7** Posted Findings[3]

Re-contactable Customers Participate in Health Research

The Powers, Perversions, and Potential of Heredity

"Magisterial."
—THE ATLANTIC

"Extraordinary."
—THE NEW YORK TIMES

"Engrossing."
—WIRED

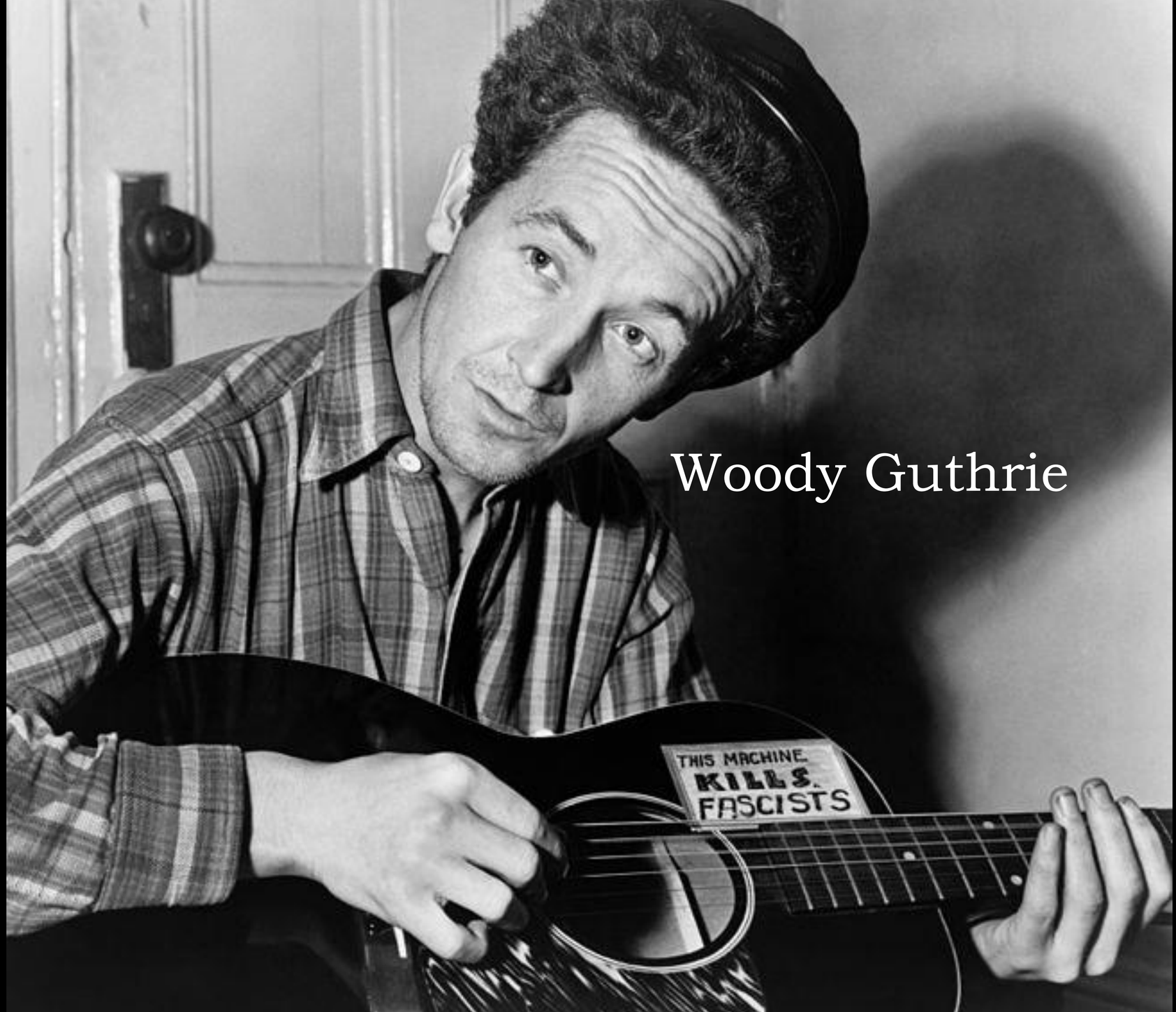
CARL ZIMMER

"SCIENCE BOOK OF THE YEAR."—THE GUARDIAN

Gregor Mendel
(1822-1884)

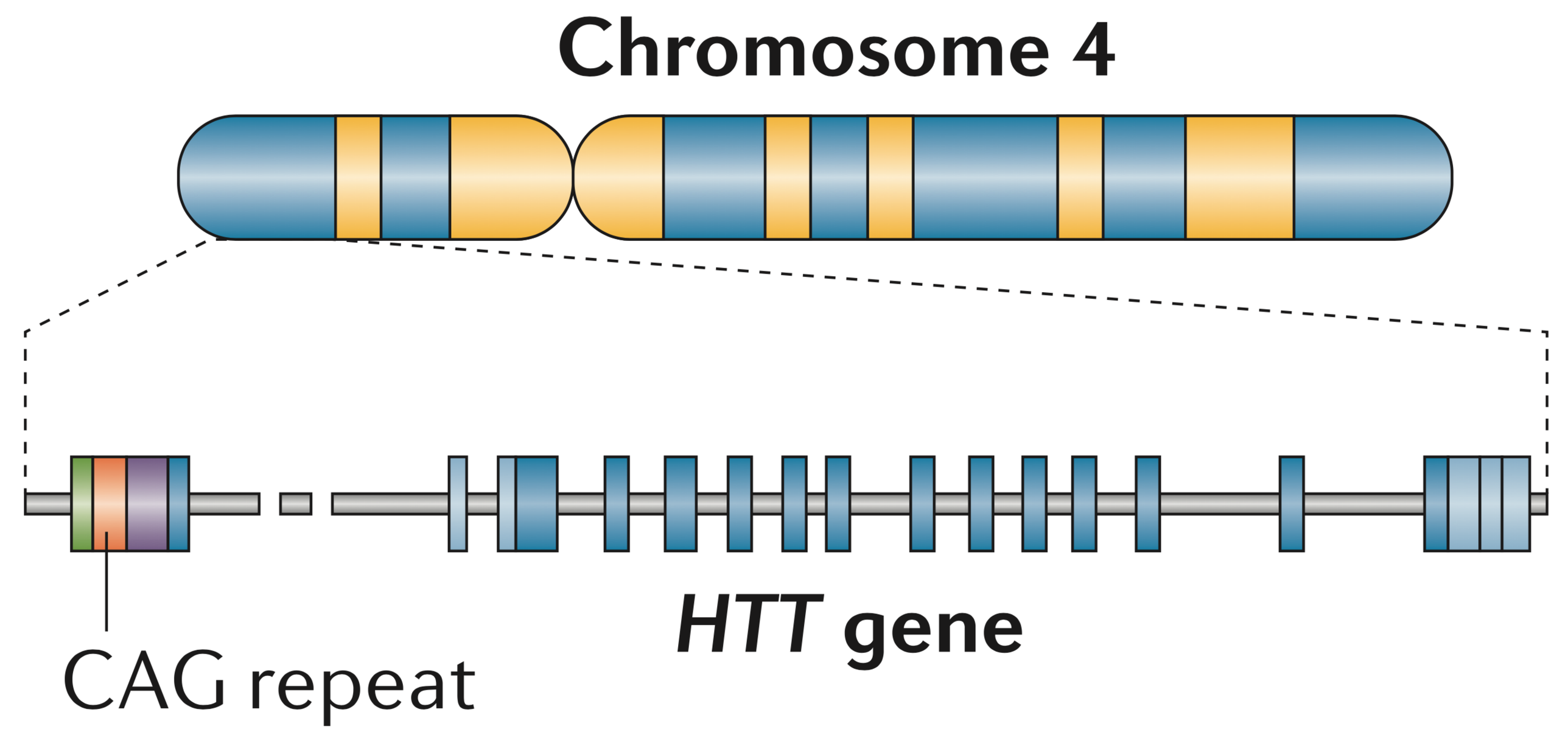
		pollen ♂	
		B	**b**
pistil ♀	**B**	**BB**	**Bb**
	b	**Bb**	**bb**

Wikipedia



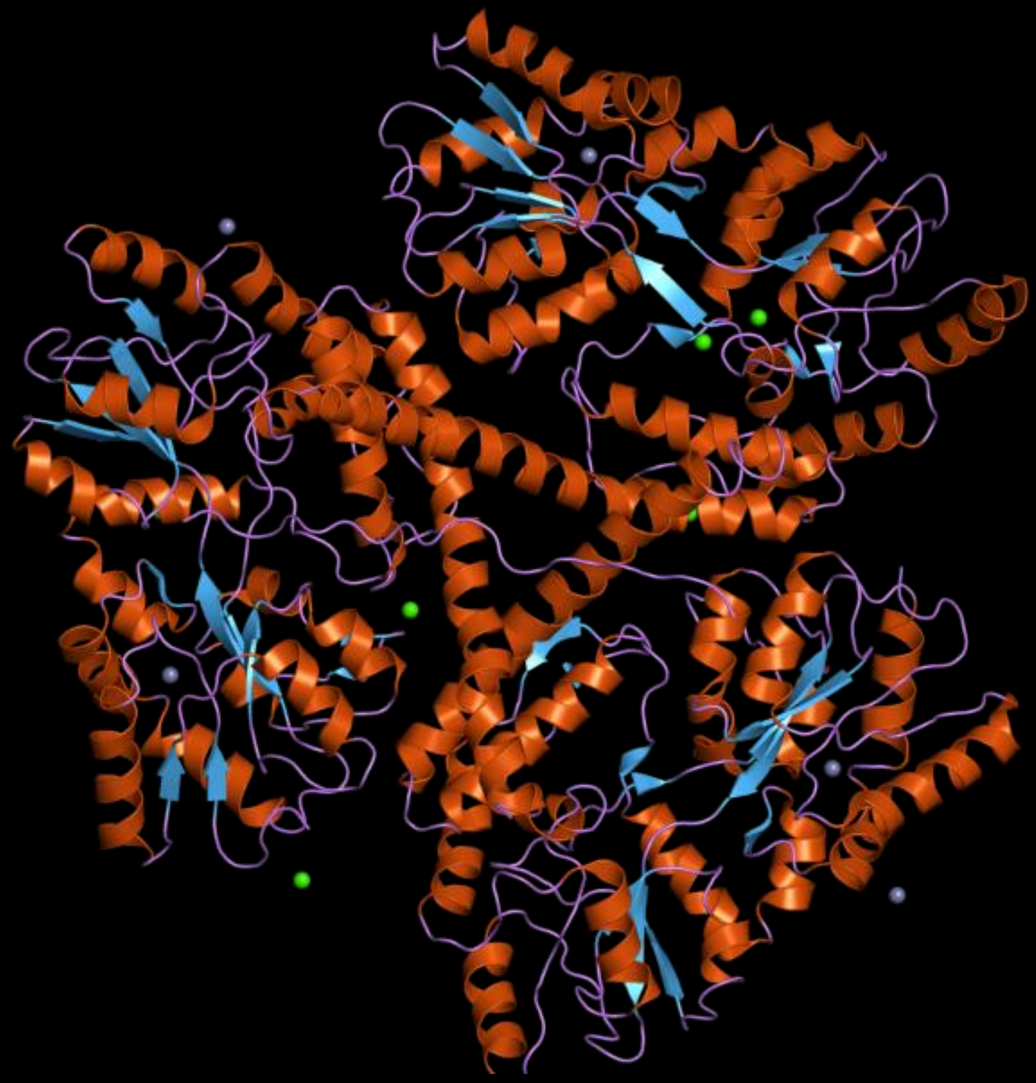
Woody Guthrie

Wikipedia

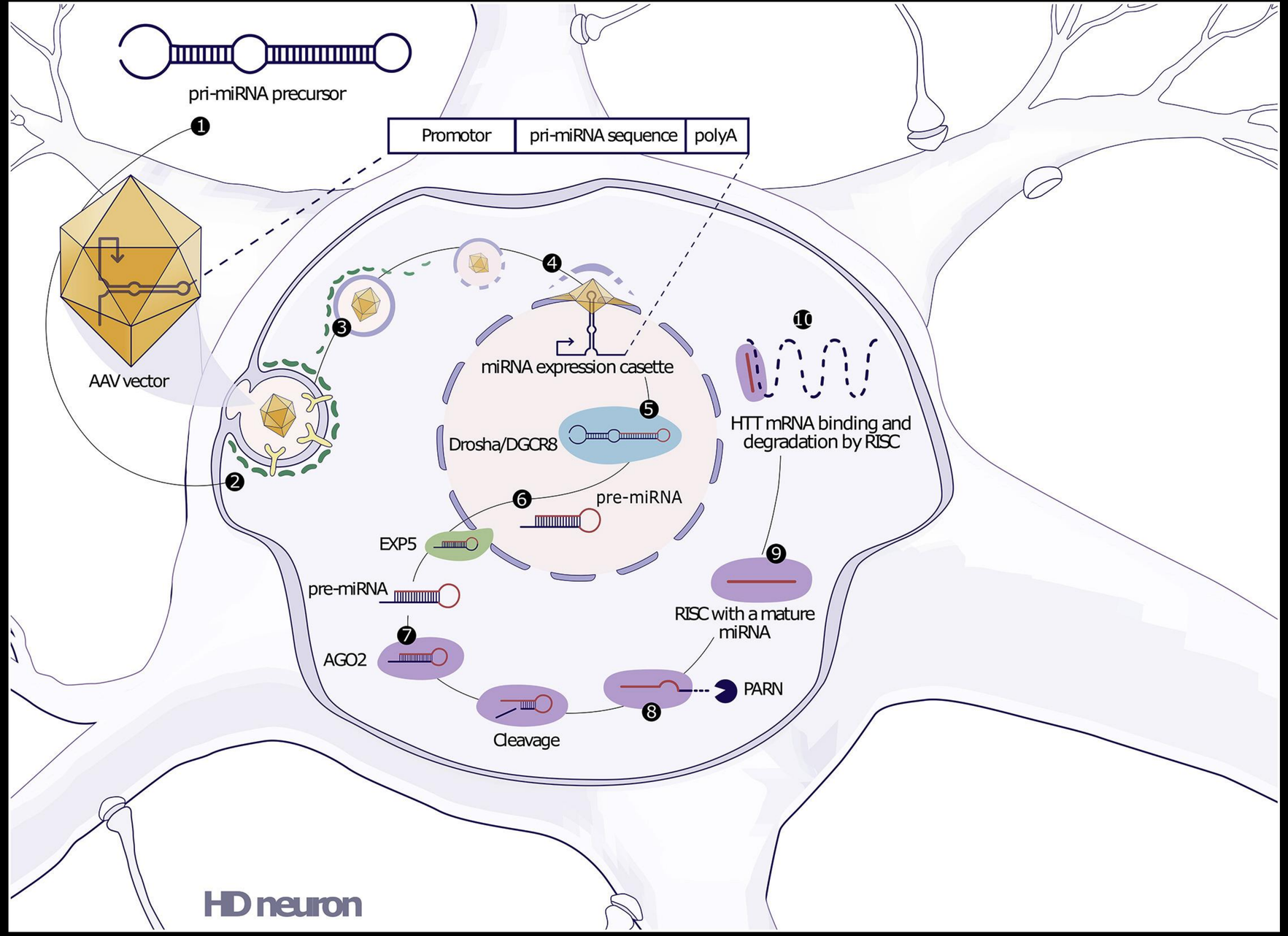


Bates et al, *Nature Reviews Disease Primers* 2015

Huntingtin Protein



Bates et al, *Nature Reviews Disease Primers* 2015



Miniarikova, J., Evers, M.M. and Konstantinova, P., 2018. Translation of microRNA-based huntingtin-lowering therapies from preclinical studies to the clinic. *Molecular Therapy*, *26*(4), pp.947-962.

Predictive Testing for HUNTINGTON DISEASE

Healthcare Provider Area Login

Search:

1. WHAT IS HD? | 2. TESTING FOR HD | 3. RESOURCES

How Is It Done? | Prenatal Testing | Deciding To Test | Alternatives | Interpreting Results | Undergo Testing

What is Predictive Testing?

This section is intended to help the individual considering testing for HD reflect on some of the issues involved in testing and in dealing with the test results.

Family, friends and professional support people may also find this material useful in supporting those considering testing.

In 1983, genetic markers closely linked to the Huntington disease (HD) gene were identified. This discovery, together with the identification of additional genetic markers, led to the development of predictive testing

HD Resources

There are many other online websites and resources which provide information regarding HD in general, support groups in your area, research updates and opportunities to be involved in clinical trials.

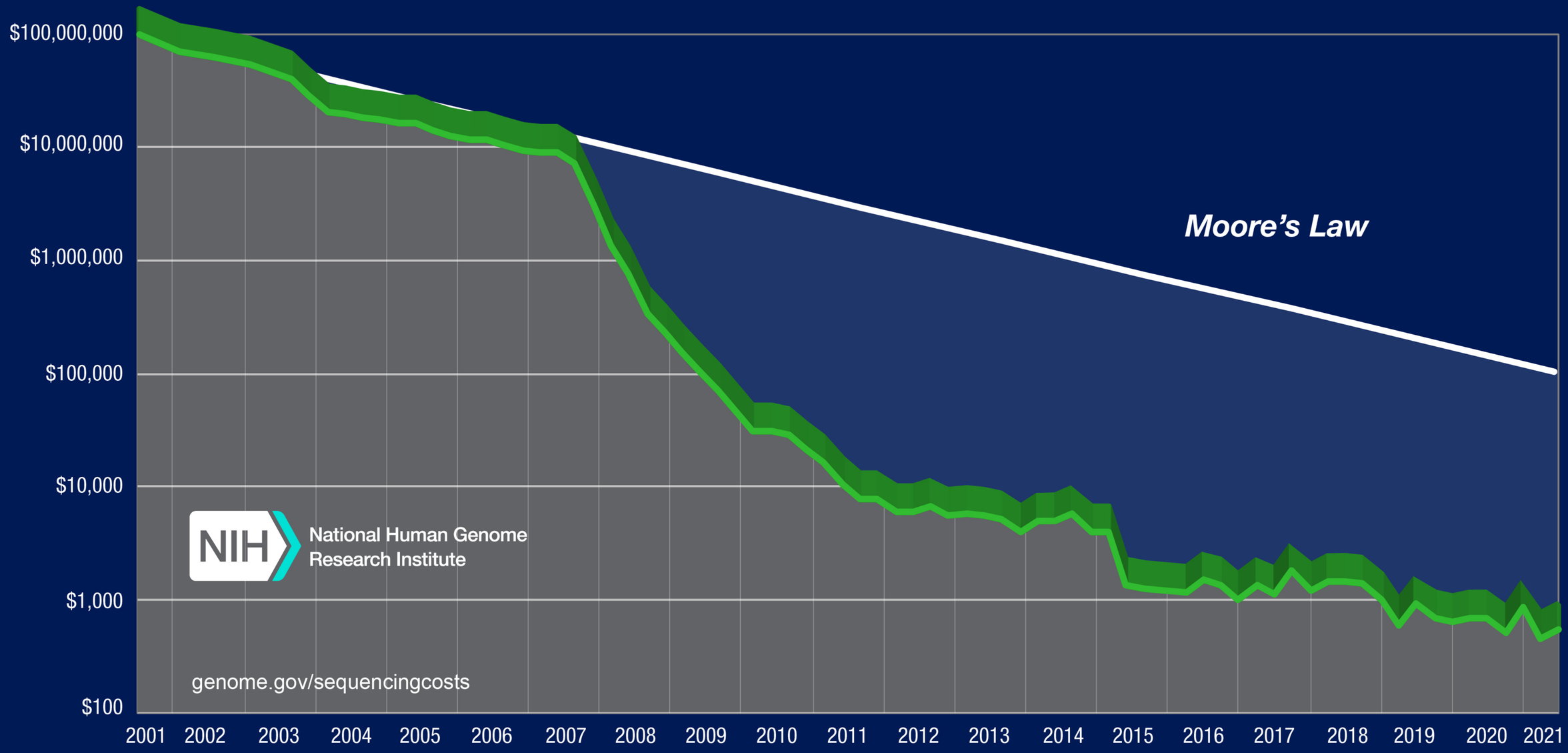
Find out more

Our Stories

We understand that learning that someone in your family has HD can be devastating. It can leave you with questions, concerns, and no idea where to turn next. Find about more about what others have done in your situation – you are not alone.

predictivetestingforhd.com/testing-for-hd/

Cost per Human Genome



UNDERSTAND YOUR GENOME®

IT STARTS WITH YOU

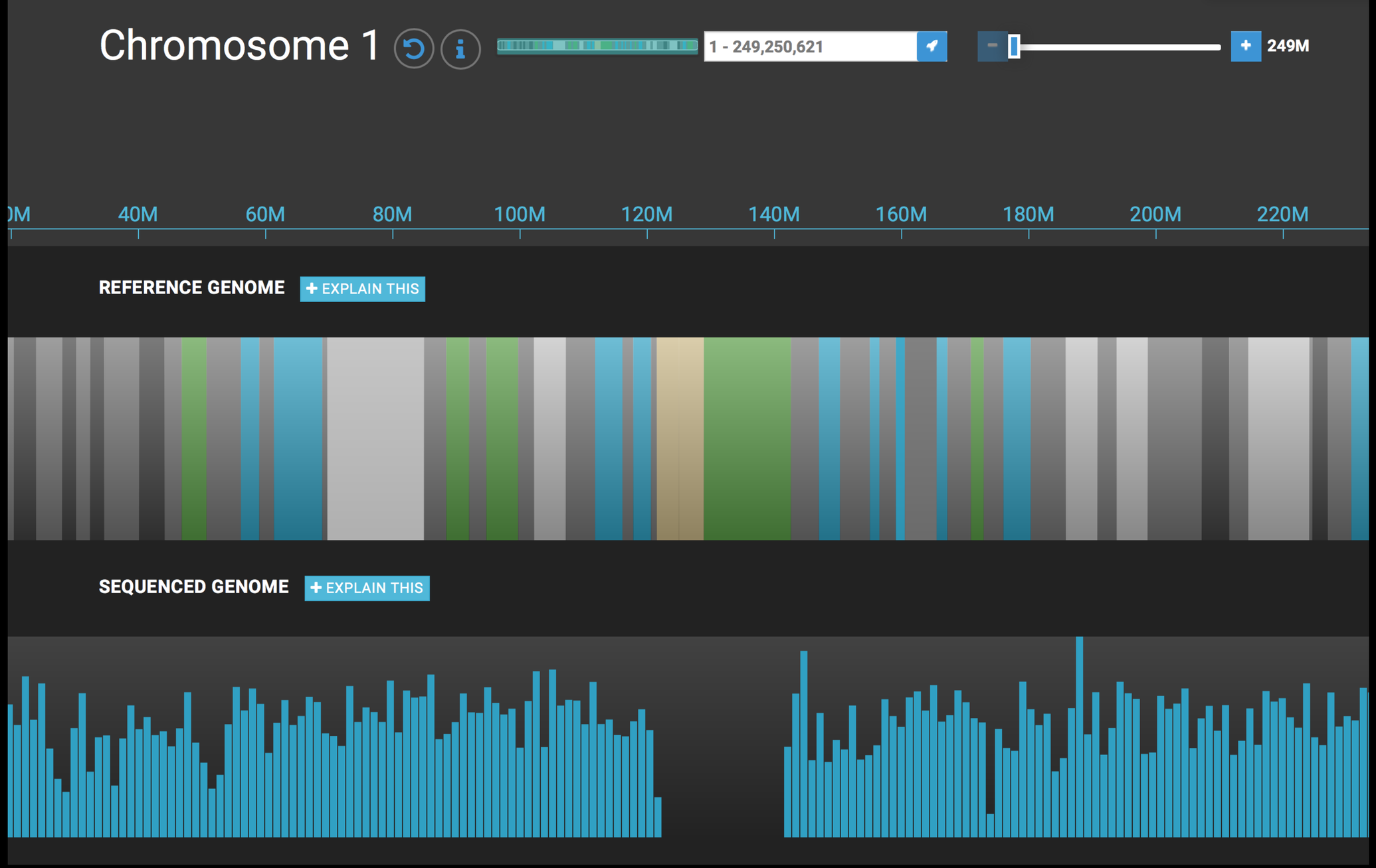
THINK ME. THINK WE. *think* **BIG**

illumina

TruGenome Predisposition Screen

Clinical Report

No pathogenic or likely pathogenic variants were found in the 1,691 genes evaluated that are expected to be clinically significant for the patient. However, this screen only detects single nucleotide substitutions and insertions and deletions of up to seven base pairs. Other types of genetic variants, including but not limited to larger insertions or deletions, copy number variants and trinucleotide repeats are not reported in this screening test. Further, the coverage of each gene is less than 100%. Therefore, clinically significant variants could exist in this genome that are not detected with this test. The coverage for each gene is provided in the Gene-Disease appendix.


Chromosome 1
1 - 249,250,621
249M
0M
40M
60M
80M
100M
120M
140M
160M
180M
200M
220M
REFERENCE GENOME
+ EXPLAIN THIS
SEQUENCED GENOME
+ EXPLAIN THIS

Here is what we see in your DNA...

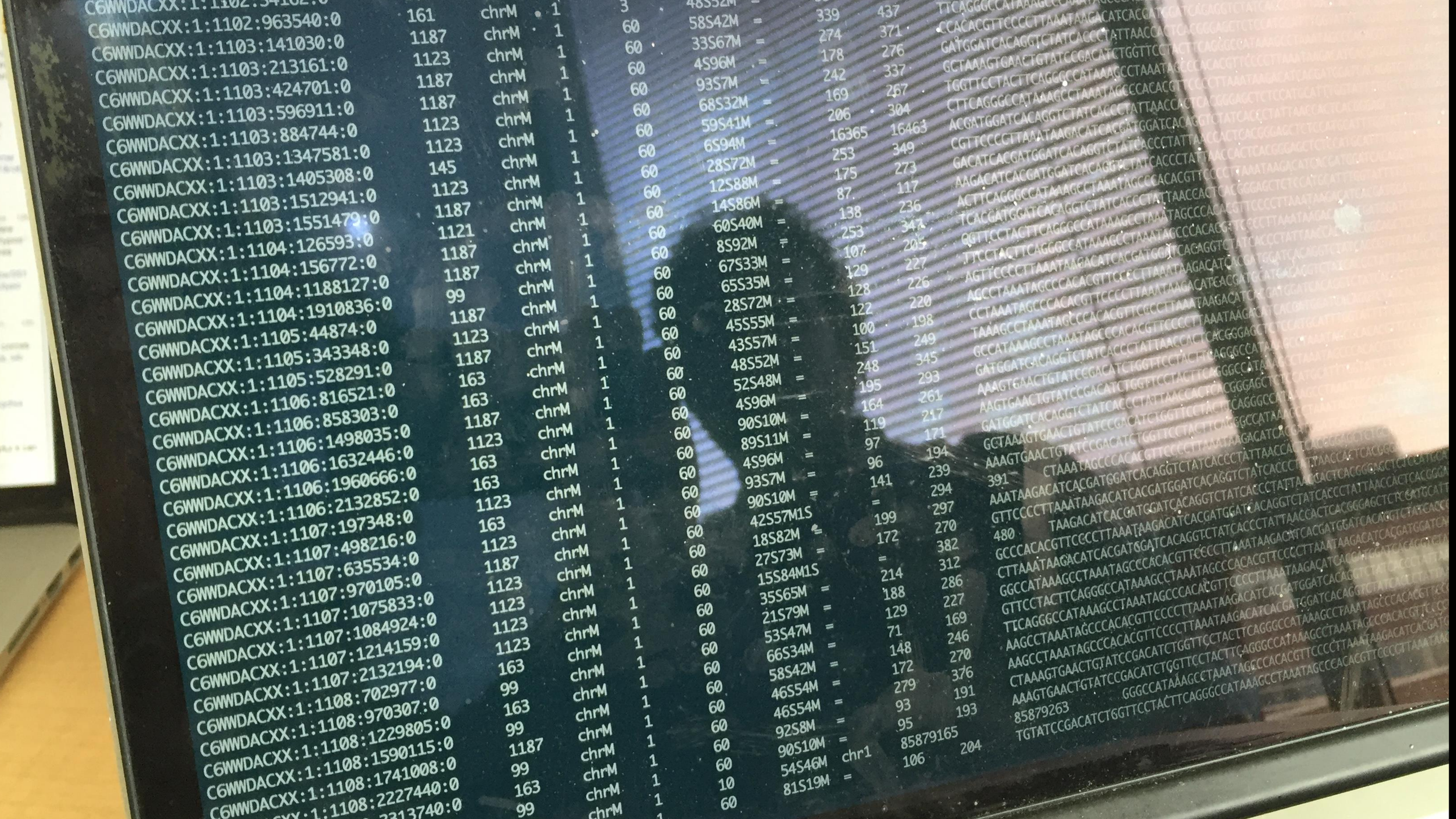
Your odds of developing male pattern baldness are increased if you are Caucasian.

You are less likely to have flush reaction if you drink alcohol.

Your muscle fibers are built for power.

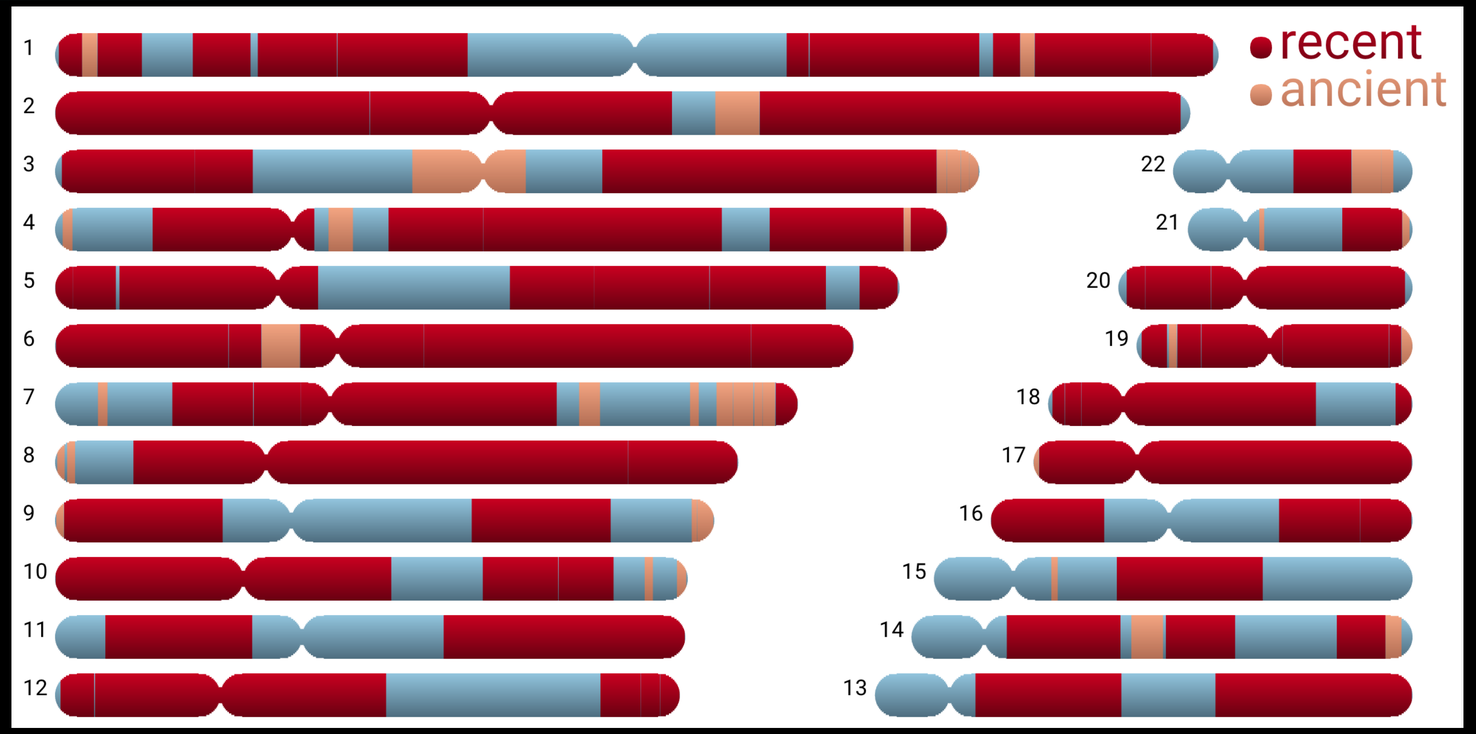
You are likely to perceive bitter tastes.


illumina®
Seagate



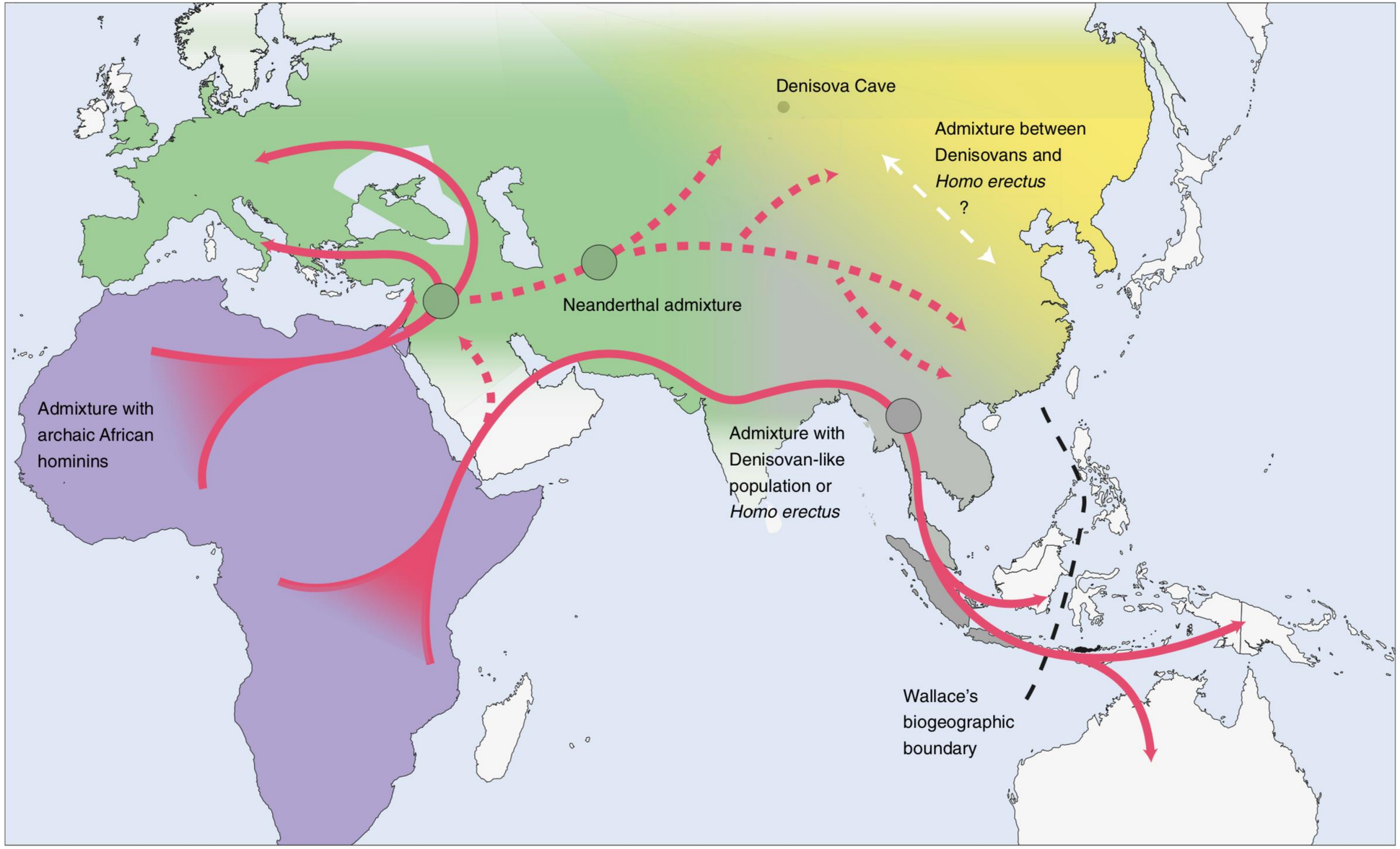
Ali Torkamani,
Scripps Translational Science Institute

My variant: rs11209026
Gene: IL23R

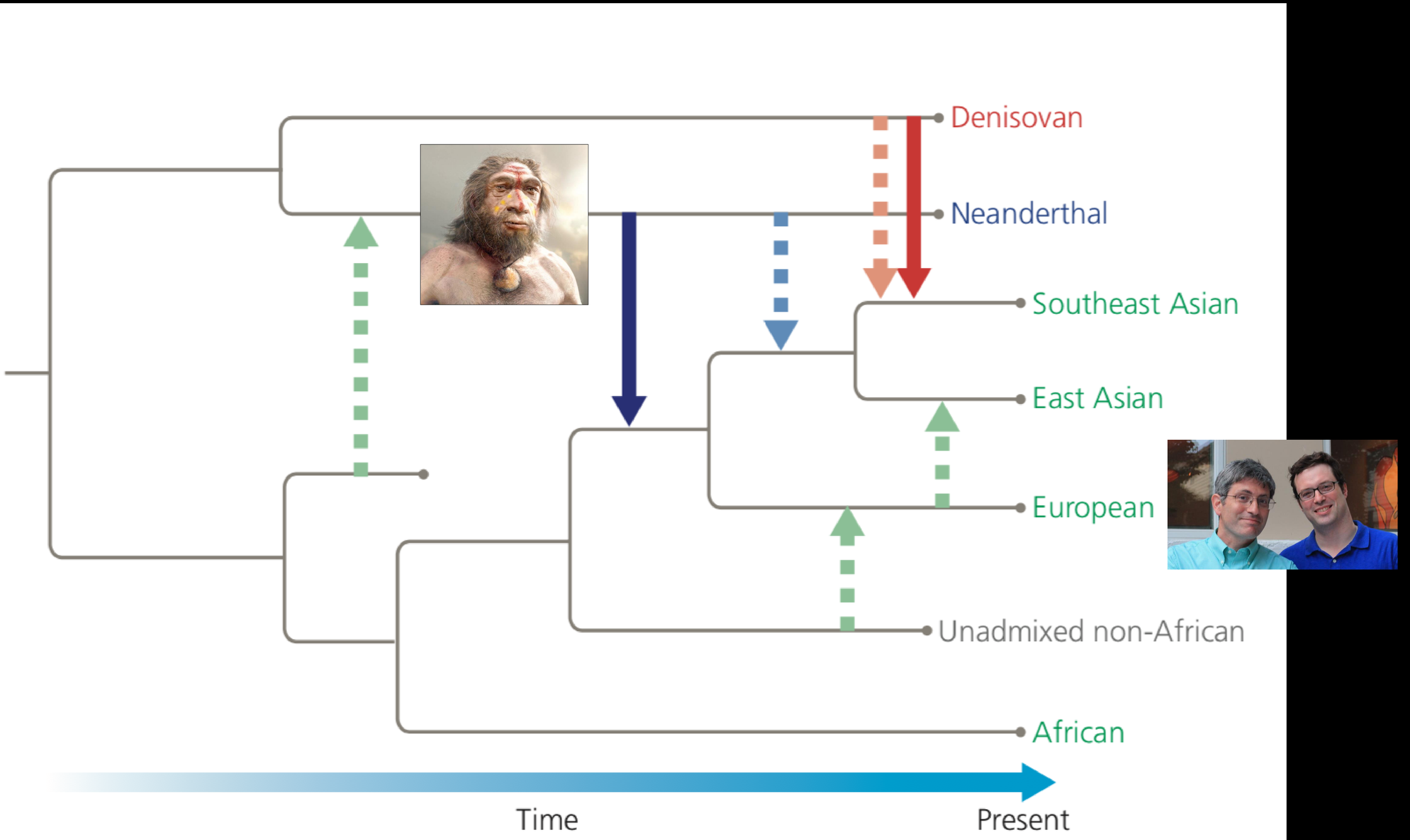


DNA.Land

Image: Viktor Deak



Roberts et al., *Nature Human Behaviour* (2018), 1



Emlen & Zimmer, *Evolution: Making Sense of Life* (Macmillan, 2019)

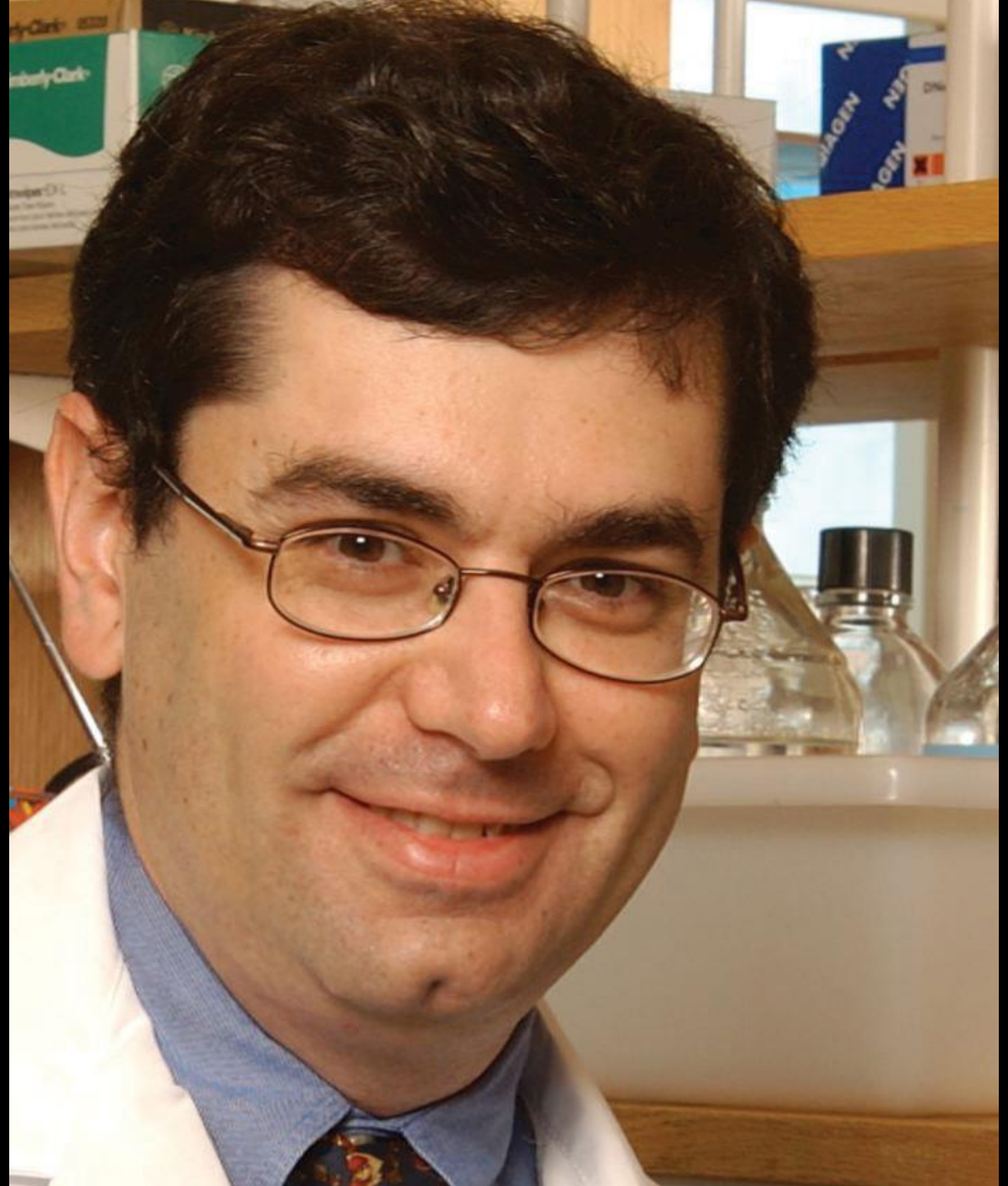
gene	chrom	gene_start	gene_end
MIR7846	chr1	12226999	12227095
MIR4632	chr1	12251769	12251830
TNFRSF1B	chr1	12227059	12269277
TNFRSF8	chr1	12123433	12204264
LRRC38	chr1	13801444	13840242
C1orf64	chr1	16330730	16333190
HSPB7	chr1	16340522	16345285
ZBTB17	chr1	16268363	16302627
CLCNKA	chr1	16348485	16360545
LDLRAD2	chr1	22138757	22151714
HSPG2	chr1	22148724	22263790
USP48	chr1	22004791	22109688
FGR	chr1	27938800	27961727
AKIRIN1	chr1	39456915	39471737
PABPC4	chr1	40026484	40042521
HEYL	chr1	40089102	40105348
OXCT2	chr1	40235196	40237020
PPIE	chr1	40204516	40229586
BMP8B	chr1	40223902	40254533
SMAP2	chr1	40839377	40888998
ZFP69B	chr1	40916336	40929390
C1orf168	chr1	57184476	57285369
LOC1019275	chr1	84041470	84326679
MIR548AP	chr1	84259597	84379059
LOC1019275	chr1	84267198	84326229
NTNG1	chr1	107682539	108027521
RPL31P11	chr1	161653494	161655042
FCGR2B	chr1	161632904	161648444
FCRLA	chr1	161676761	161684142
FCRLB	chr1	161691333	161697933
DUSP12	chr1	161719557	161726954
OLFML2B	chr1	161952981	161994255
ATF6	chr1	161736033	161933860
LINC00970	chr1	168873142	169056243
LINC01142	chr1	170240545	170253349
FAM163A	chr1	179712297	179785333
TOR1AIP1	chr1	179851176	179889212
TOR1AIP2	chr1	179809101	179846941
CEP350	chr1	179923907	180084015
FLJ23867	chr1	180167143	180169859
QSOX1	chr1	180123967	180167169

Analysis courtesy of Joshua Akey
& Selina Vattathil, Princeton

Charles Byrne (1761–1783)

“Mr. O'Brien the Irish Giant the Tallest Man in the Known World Being Near Nine Feet High”

whitney.med.yale.edu



Joel Hirschhorn
Harvard Medical School

2007

1 variant

4921 people

Weedon et al

2010

180 variants

183,727 people

Lango Allen et al

2014

697 variants

253,288 people

Wood et al

2018

3290 variants

693,529 people

Yengo et al

2020

9900 variants

4.1 million people

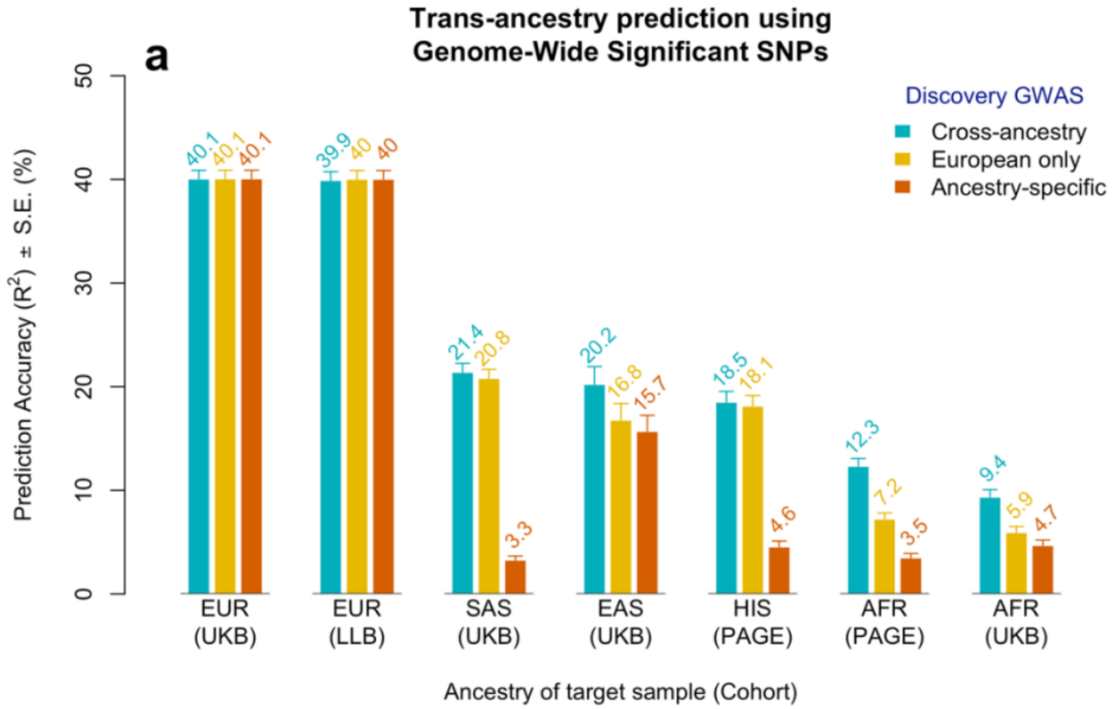
Yengo et al

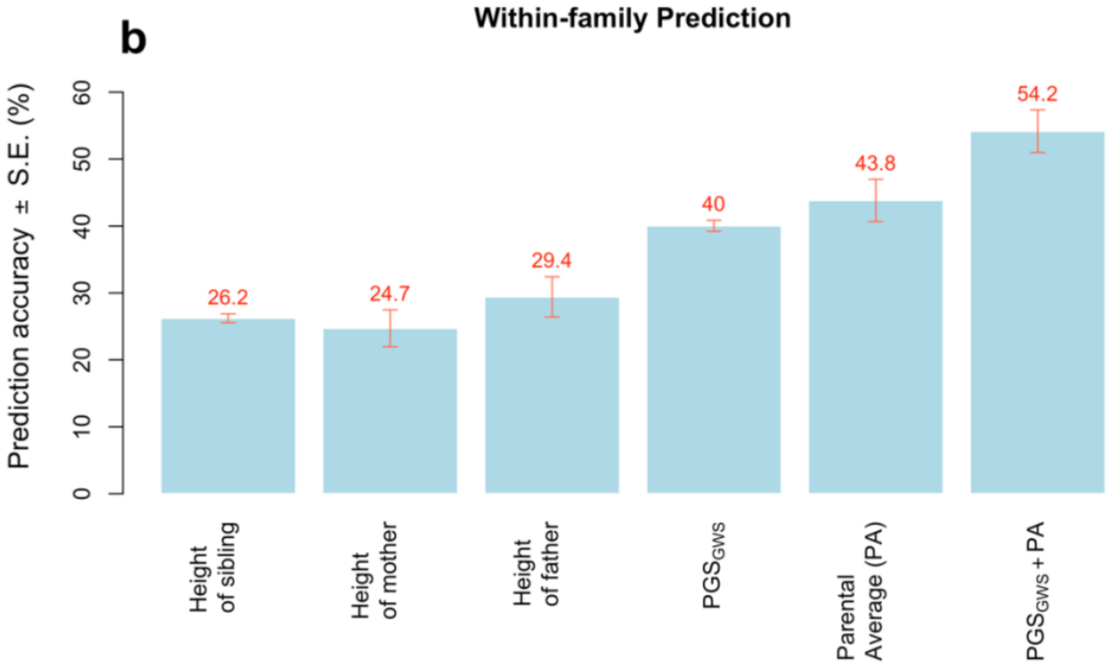
2022

12,111 variants

5.4 million people

Yengo et al





Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A.U., Jiang, Y., Raghavan, S. and Miao, J., 2022. A Saturated Map of Common Genetic Variants Associated with Human Height from 5.4 Million Individuals of Diverse Ancestries. *bioRxiv*.

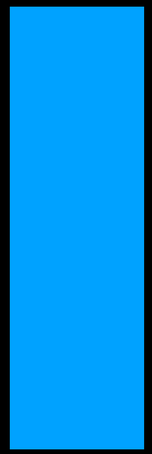
AVERAGE WOMEN'S HEIGHT (cm)

Canada

157.6

Barbados

152.1



http://ncdrisc.org/height-mean-ranking.html

AVERAGE WOMEN'S HEIGHT (cm)

Canada

157.6

Barbados

152.1

1896

1996

165.3

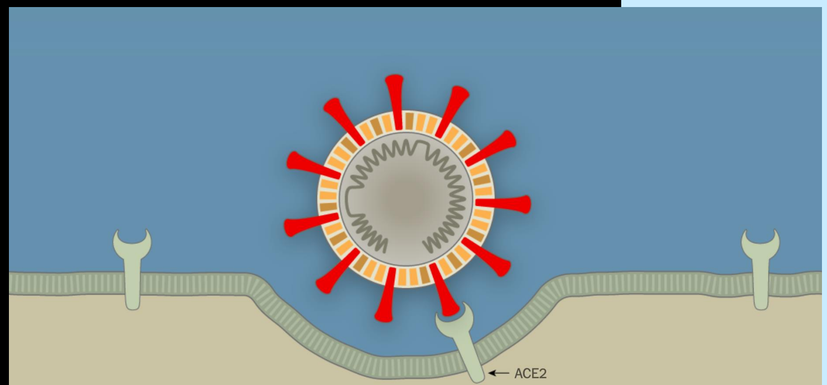
163.9

http://ncdrisc.org/height-mean-ranking.html


CNS TV

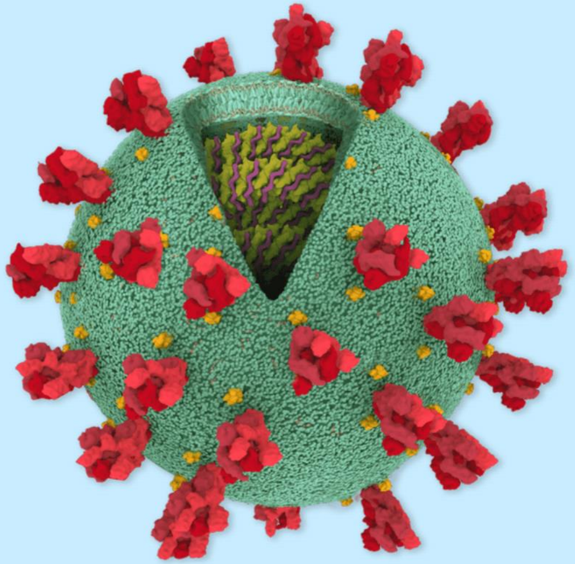
Bad News Wrapped in Protein: Inside the Coronavirus Genome

By Jonathan Corum and Carl Zimmer April 3, 2020

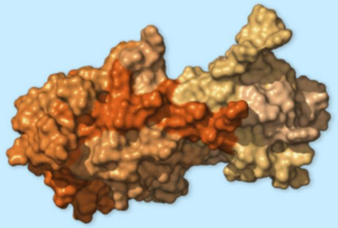


Entering a Vulnerable Cell

The virus enters the body through the nose, mouth or eyes, then attaches to cells in the airway that produce a protein called ACE2. The virus is believed to have originated in bats, where it may have attached to a similar protein.

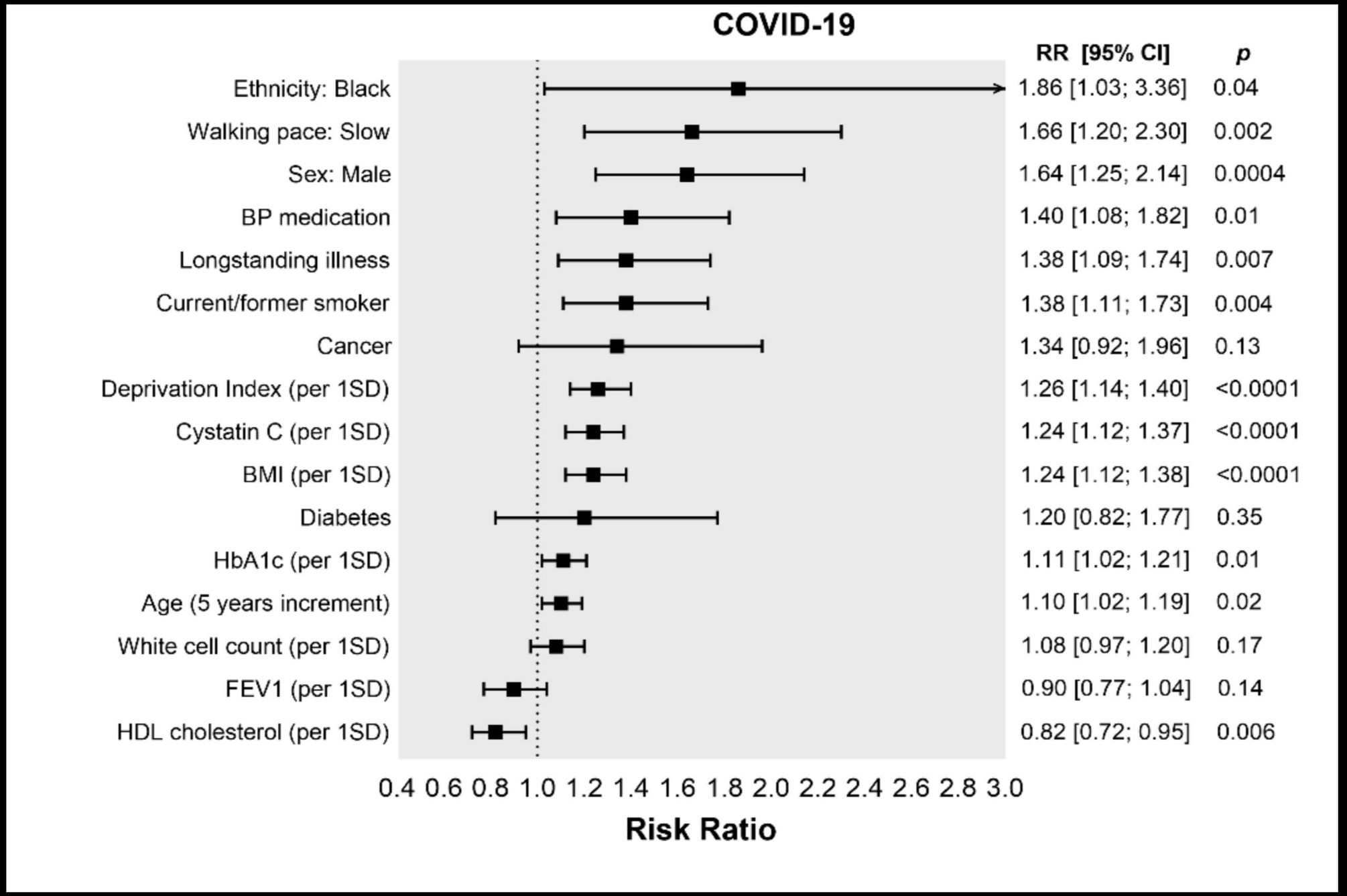


Protein Scissors · NSP5

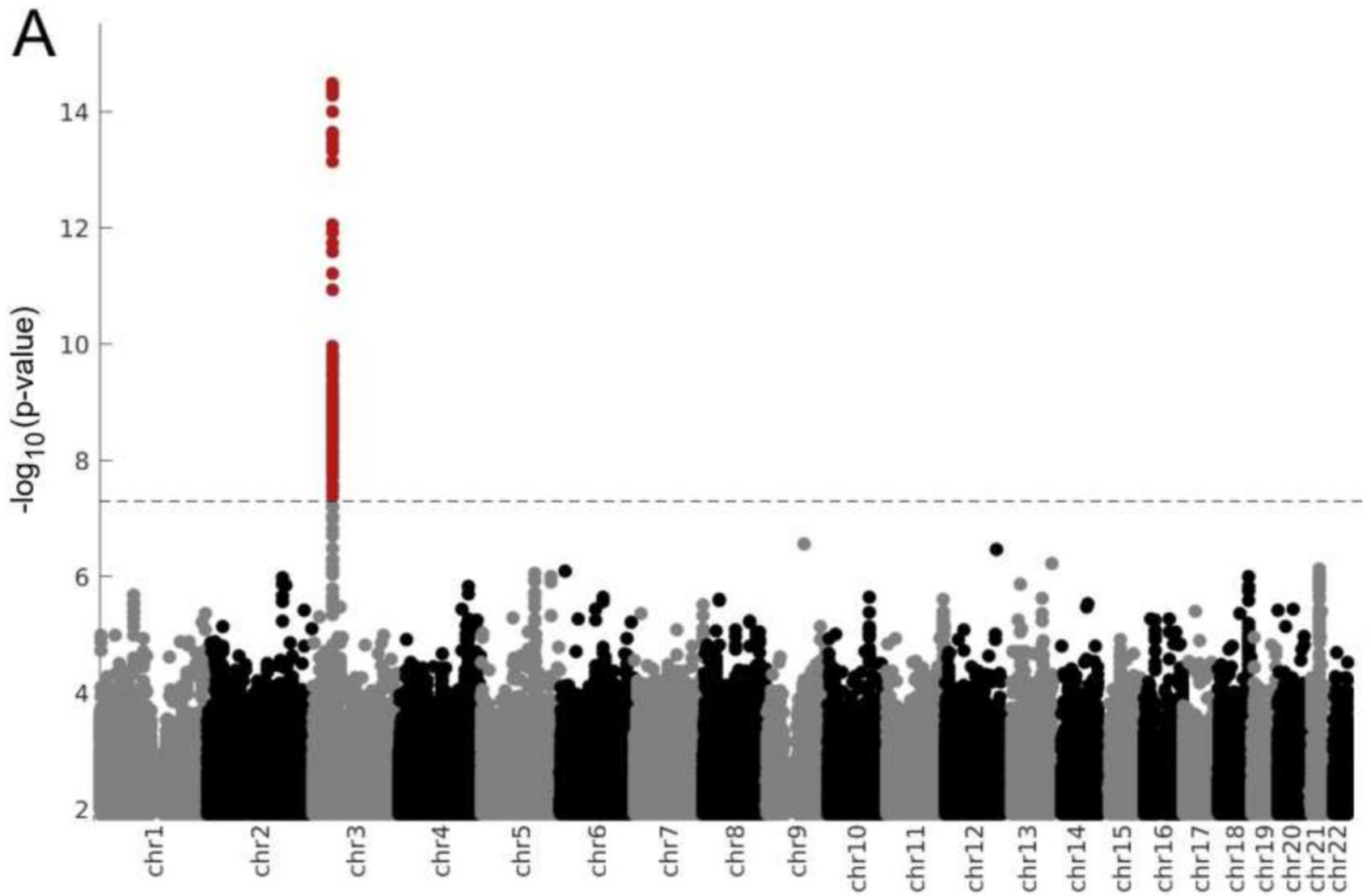


This protein makes most of the cuts that free other NSP proteins to carry out their own jobs.

https://www.nytimes.com/interactive/2020/03/11/science/how-coronavirus-hijacks-your-cells.html
https://www.nytimes.com/interactive/2020/04/03/science/coronavirus-genome-bad-news-wrapped-in-protein.html



medrxiv.org/content/10.1101/2020.04.28.20083295v1



biorxiv.org/content/10.1101/2020.07.03.186296v1

Article | Published: 30 September 2020

The major genetic risk factor for severe COVID-19 is inherited from Neanderthals

Hugo Zeberg ✉ & Svante Pääbo ✉

Nature **587**, 610–612(2020) | Cite this article

DNA Inherited From Neanderthals May Increase Risk of Covid-19

The stretch of six genes seems to increase the risk of severe illness from the coronavirus.

A researcher excavating a Neanderthal skeleton last year. Gailan Haji/EPA-EFE/REX

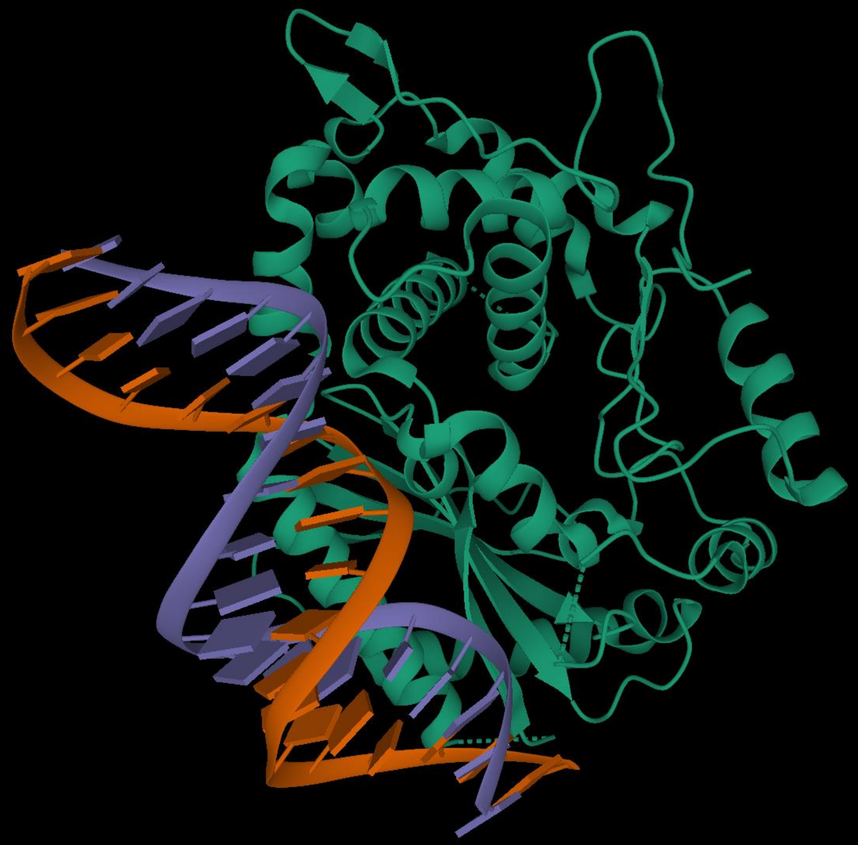
By Carl Zimmer

Published July 4, 2020 Updated July 8, 2020

7

A stretch of DNA linked to Covid-19 was passed down from Neanderthals 60,000 years ago, according to a new study.

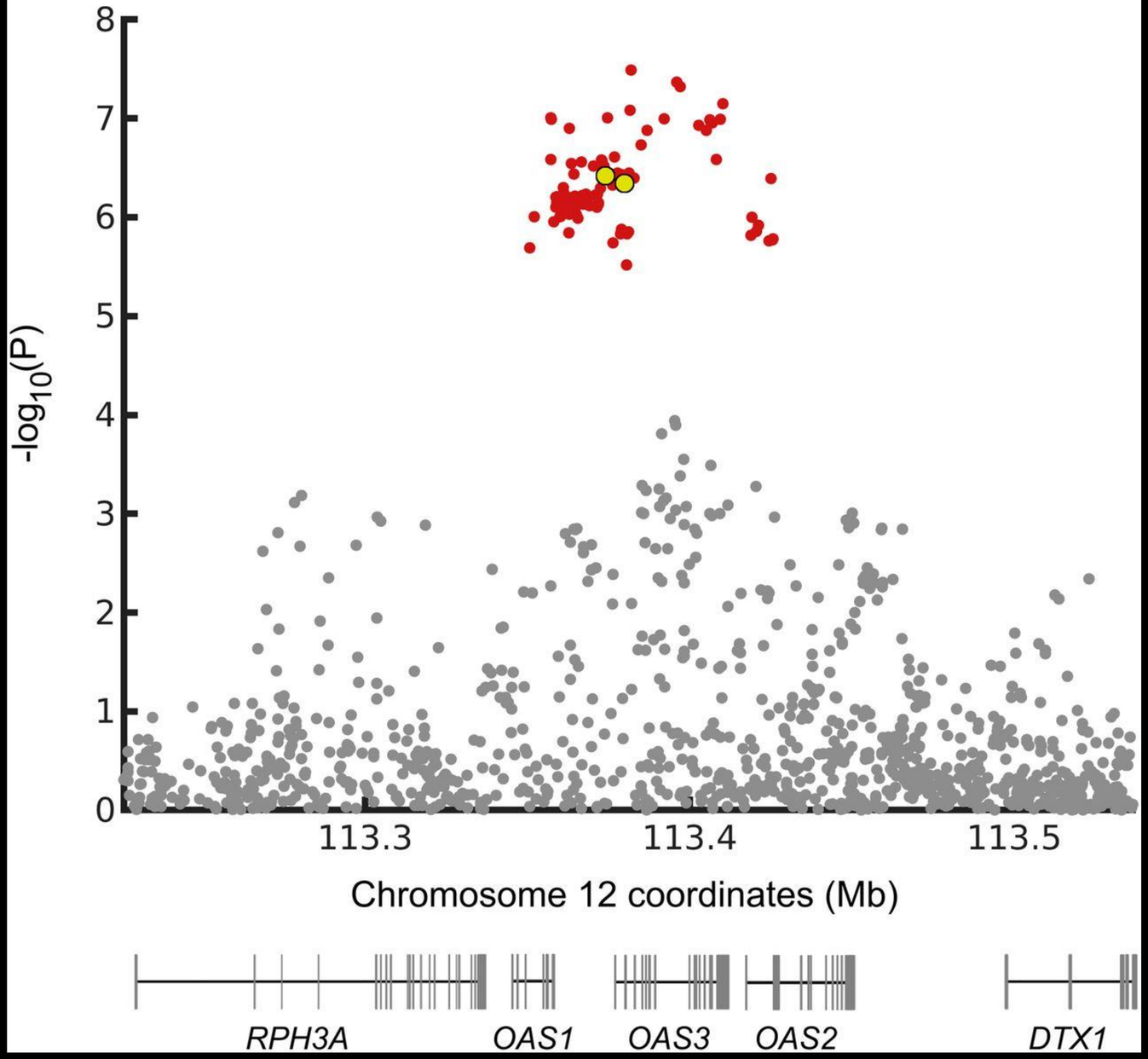
OAS3 (green)
Double-stranded RNA (orange and purple)



pnas.org/content/118/9/e2026309118
https://www.ebi.ac.uk/pdbe/

A genomic region associated with protection against severe COVID-19 is inherited from Neandertals

Hugo Zeberg[a,b,1] and Svante Pääbo[a,c,1]




illumina®
Seagate

Alleles for increased risk on chromosome 13: NONE

Alleles for decreased risk on chromosome 12: Two heterozygous SNPs

On balance, a good inheritance?

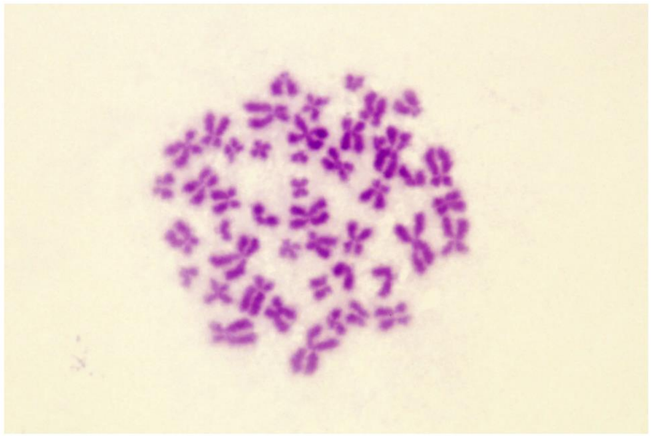
Courtesy of Jiahao Gao

The New York Times

MATTER

Scientists Finish the Human Genome at Last

The complete genome uncovered more than 100 new genes that are probably functional, and many new variants that may be linked to diseases.



A century ago, scientists knew that genes were spread across 23 pairs of chromosomes. But pinpointing any single gene and deciphering its sequence was a struggle that could have consumed a career. Michael Abbey/Science Source

By Carl Zimmer

Published July 23, 2021 Updated July 26, 2021

Two decades after the draft sequence of the human genome was unveiled to great fanfare, a team of 99 scientists has finally deciphered the entire thing. They have filled in vast gaps and corrected a long list of errors in previous versions, giving us a new view of our DNA.

master · 1 branch · 0 tags · Go to file · Code

aphillippy updating dotplot link to v1.1 assembly — 1b28da9 on Nov 18, 2021 — 76 commits

LICENSE.md	Add more data description, add first pass data links	3 years ago
README.md	updating dotplot link to v1.1 assembly	3 months ago
pbP6Cells	Create pbP6Cells	3 years ago

README.md

Telomere-to-telomere consortium

We have sequenced the CHM13hTERT human cell line with a number of technologies. Human genomic DNA was extracted from the cultured cell line. As the DNA is native, modified bases will be preserved. The data includes 30x PacBio HiFi, 120x coverage of Oxford Nanopore, 70x PacBio CLR, 50x 10X Genomics, as well as BioNano DLS and Arima Genomics HiC. Most raw data is available from this site, with the exception of the PacBio data which was generated by the University of Washington/PacBio and is available from NCBI SRA.

UCSC browsers are available for v1.0 and v1.1. An interactive dotplot visualization of all genomic repeats is also available from resgen.io. Known issues identified in the assembly are tracked at CHM13 issues.

About

The complete sequence of a human genome

Readme
View license
404 stars
49 watching
54 forks

Releases

No releases published

Packages

No packages published

Contributors 4

skoren Sergey Koren

github.com/marbl/CHM13

Thank you

For more information, please visit carlzimmer.com