Genome-Wide Association Studies

Part I: historical overview

International Statistical Genetics Workshop – 2026

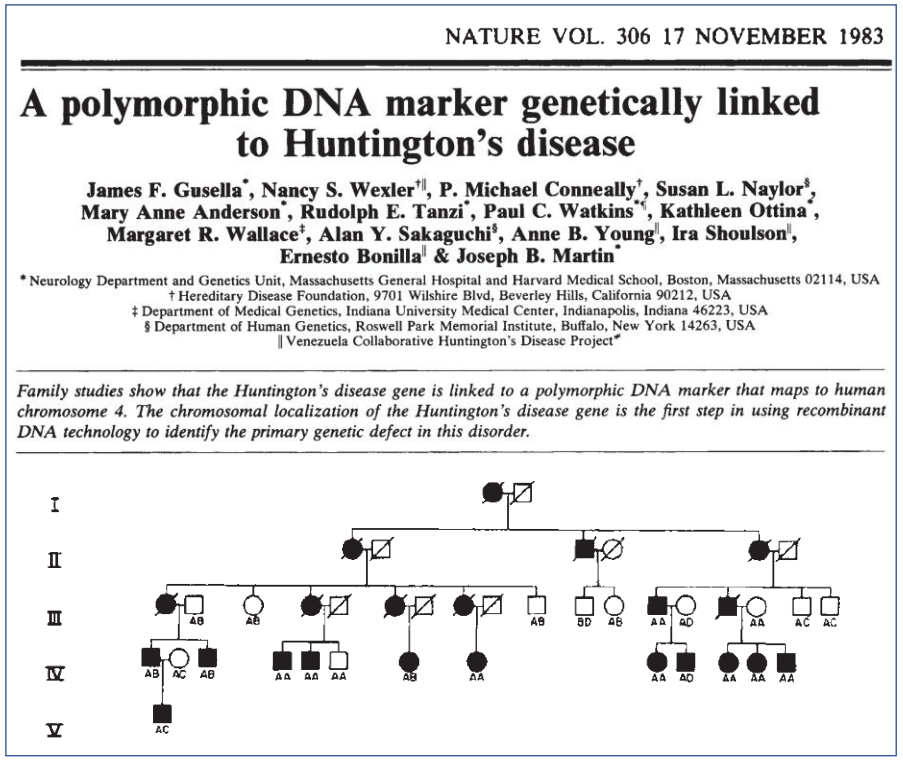
Abdel Abdellaoui

a.abdellaoui@amsterdamumc.nl


IN GENOME-WIDE ASSOCIATION STUDIES, OR GWAS, RESEARCHERS INTERROGATE SNPs THAT COMMONLY ARISE IN THE POPULATION... AND ASK WHETHER A VARIANT OF THAT SNP ASSOCIATES WITH DISEASE.

Before GWAS

Linkage studies



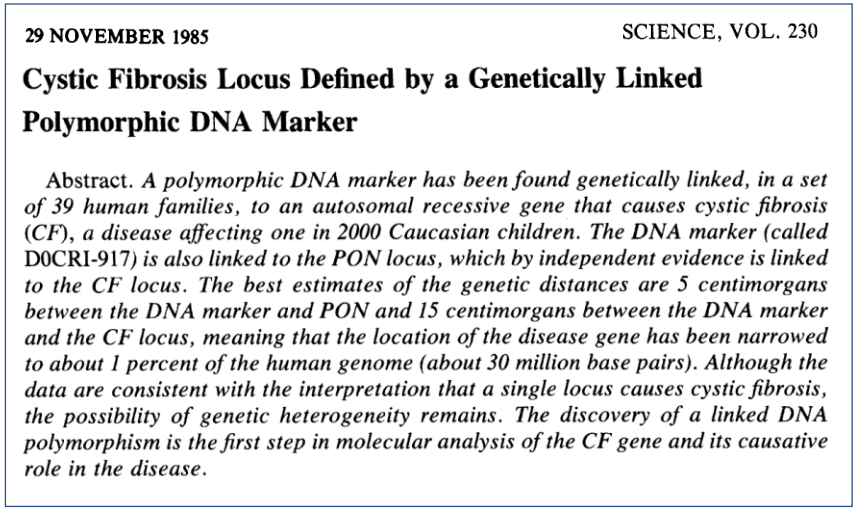
NATURE VOL. 306 17 NOVEMBER 1983

A polymorphic DNA marker genetically linked to Huntington's disease

James F. Gusella*, Nancy S. Wexler†‖, P. Michael Conneally†, Susan L. Naylor§, Mary Anne Anderson*, Rudolph E. Tanzi*, Paul C. Watkins*¶, Kathleen Ottina*, Margaret R. Wallace‡, Alan Y. Sakaguchi§, Anne B. Young‖, Ira Shoulson‖, Ernesto Bonilla‖ & Joseph B. Martin*

* Neurology Department and Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA
† Hereditary Disease Foundation, 9701 Wilshire Blvd, Beverley Hills, California 90212, USA
‡ Department of Medical Genetics, Indiana University Medical Center, Indianapolis, Indiana 46223, USA
§ Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14263, USA
‖ Venezuela Collaborative Huntington's Disease Project

Family studies show that the Huntington's disease gene is linked to a polymorphic DNA marker that maps to human chromosome 4. The chromosomal localization of the Huntington's disease gene is the first step in using recombinant DNA technology to identify the primary genetic defect in this disorder.



29 NOVEMBER 1985 SCIENCE, VOL. 230

Cystic Fibrosis Locus Defined by a Genetically Linked Polymorphic DNA Marker

Abstract. *A polymorphic DNA marker has been found genetically linked, in a set of 39 human families, to an autosomal recessive gene that causes cystic fibrosis (CF), a disease affecting one in 2000 Caucasian children. The DNA marker (called* D0CRI-917*) is also linked to the PON locus, which by independent evidence is linked to the CF locus. The best estimates of the genetic distances are 5 centimorgans between the DNA marker and PON and 15 centimorgans between the DNA marker and the CF locus, meaning that the location of the disease gene has been narrowed to about 1 percent of the human genome (about 30 million base pairs). Although the data are consistent with the interpretation that a single locus causes cystic fibrosis, the possibility of genetic heterogeneity remains. The discovery of a linked DNA polymorphism is the first step in molecular analysis of the CF gene and its causative role in the disease.*

Doesn't work for complex traits, unless the variant has a very large effect:

- Diabetes (Insulin gene)
- Breast cancer (BRCA1 & BRCA2 genes)
- Alzheimer (APOE genes)

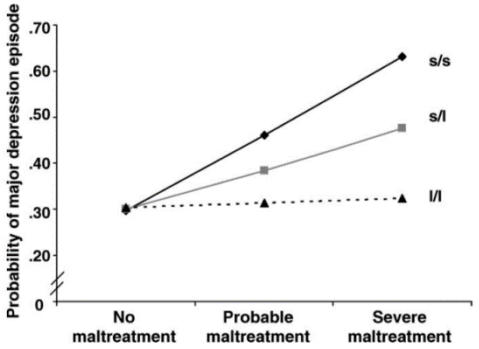
Before GWAS

Candidate gene studies

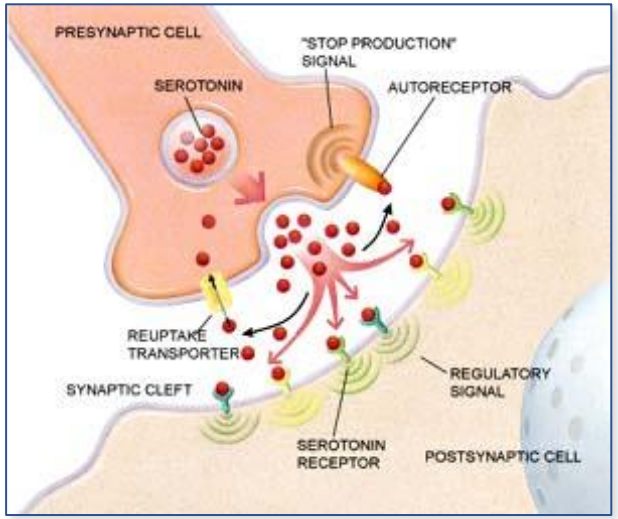
Serotonine & depression

Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene

Avshalom Caspi,[1,2] Karen Sugden,[1] Terrie E. Moffitt,[1,2]*
Alan Taylor,[1] Ian W. Craig,[1] HonaLee Harrington,[2]
Joseph McClay,[1] Jonathan Mill,[1] Judy Martin,[3]
Antony Braithwaite,[4] Richie Poulton[3]



18 JULY 2003 VOL 301 SCIENCE www.sciencemag.org



Molecular Psychiatry (2017) 00, 1–10

www.nature.com/mp

ORIGINAL ARTICLE

Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression

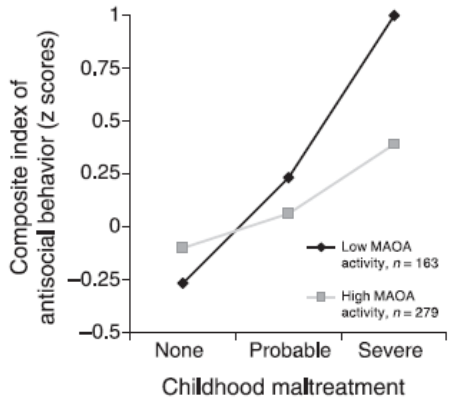
Before GWAS

Candidate gene studies

Mono-amine oxidase (MAO-A) & aggression

Role of Genotype in the Cycle of Violence in Maltreated Children

Avshalom Caspi,[1,2] Joseph McClay,[1] Terrie E. Moffitt,[1,2]* Jonathan Mill,[1] Judy Martin,[3] Ian W. Craig,[1] Alan Taylor,[1] Richie Poulton[3]



www.sciencemag.org SCIENCE VOL 297 2 AUGUST 2002

Published online 30 October 2009 | Nature | doi:10.1038/news.2009.1050

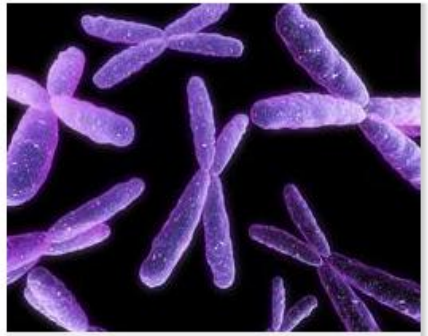
News

Lighter sentence for murderer with 'bad genes'

Italian court reduces jail term after tests identify genes linked to violent behaviour.

Emiliano Feresin

An Italian court has cut the sentence given to a convicted murderer by a year because he has genes linked to violent behaviour — the first time that behavioural genetics has affected a sentence passed by a European court. But researchers contacted by *Nature* have questioned whether the decision was based on sound science.



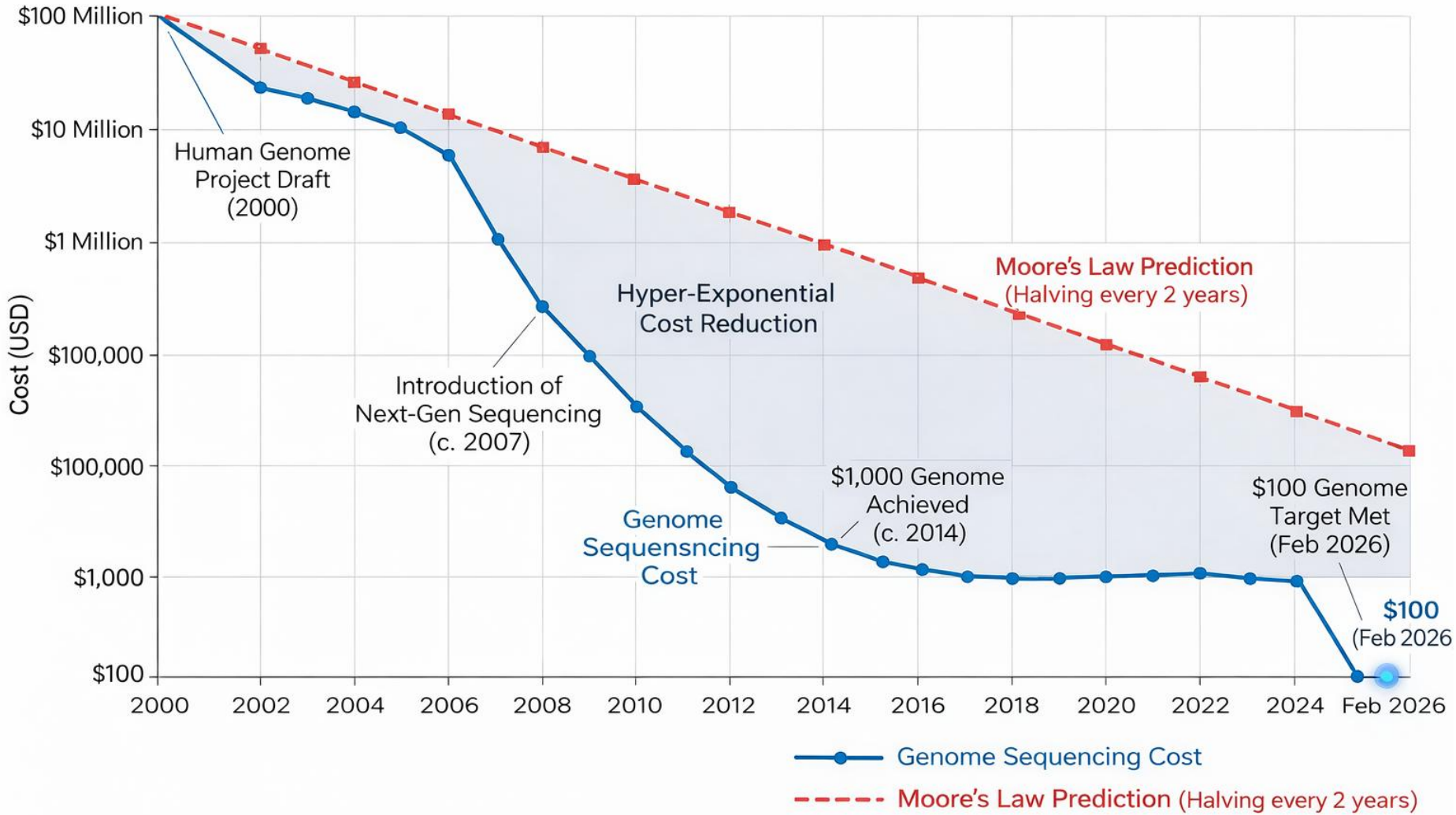
A court in Italy has cut a prisoner's jail term because he has genes associated with aggressive behaviour.

Ingram Publishing

Our a priori hypotheses are bad …

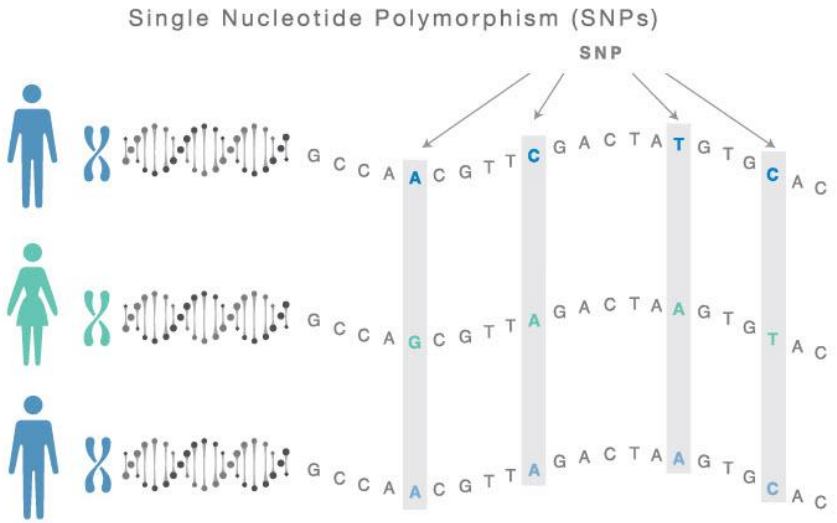
We need a hypothesis-free design → scan the whole genome?

Cost of Sequencing a Whole Human Genome (2000–2026) vs. Moore's Law



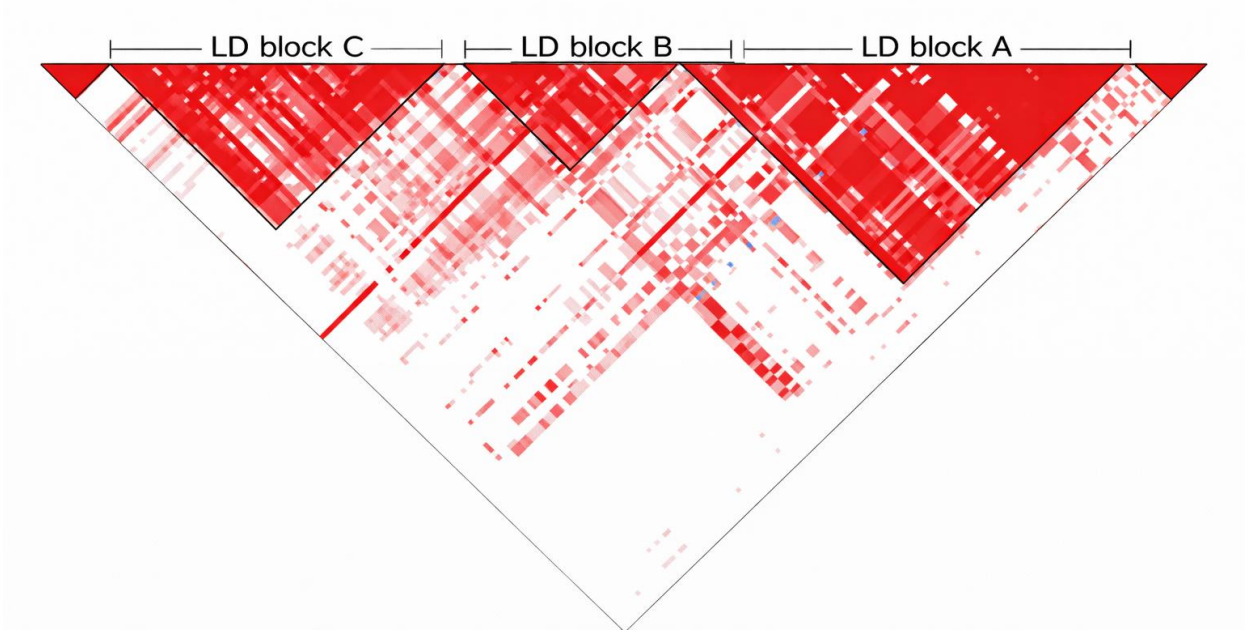
We don’t need to sequence the whole genome to capture most variation!

- **~3 billion** base pairs in the genome
- **~10-15 million** common SNPs across humans
- **~500,000–1,000,000** SNPs on a DNA microarray



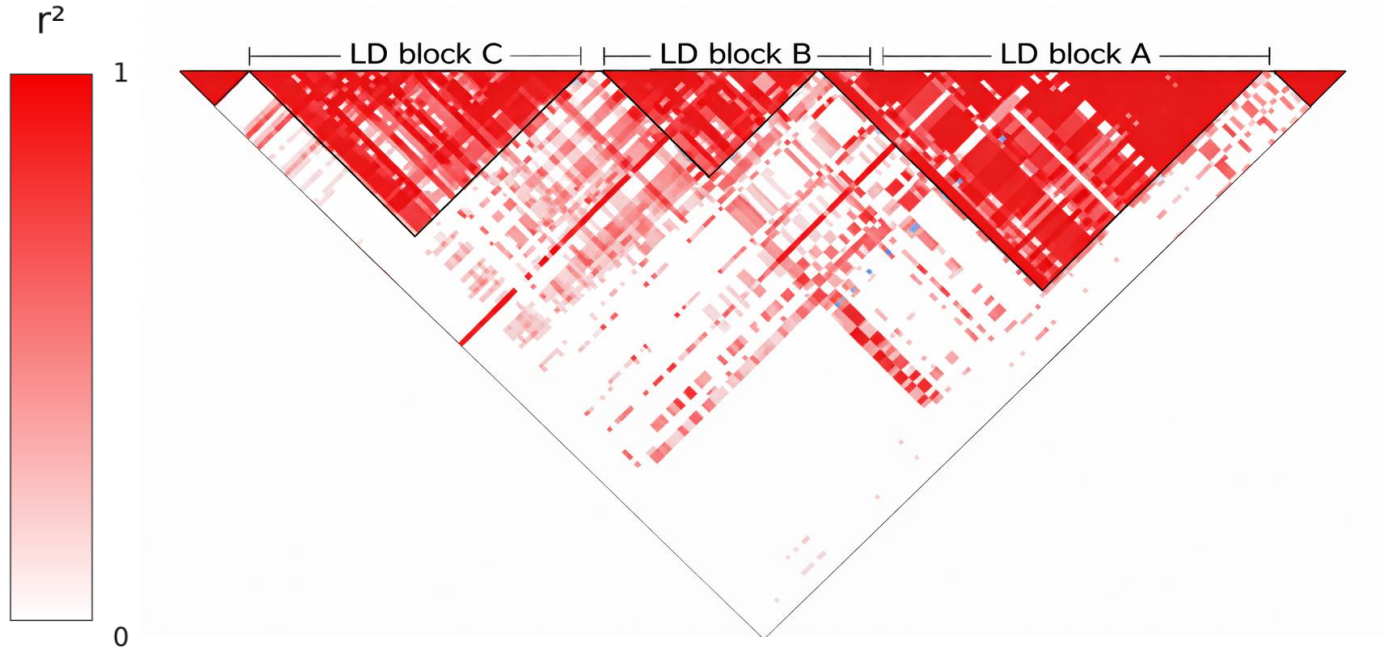
Linkage Disequilibrium (LD)

- Alleles are in LD when they're inherited together more frequently than expected (they're correlated).



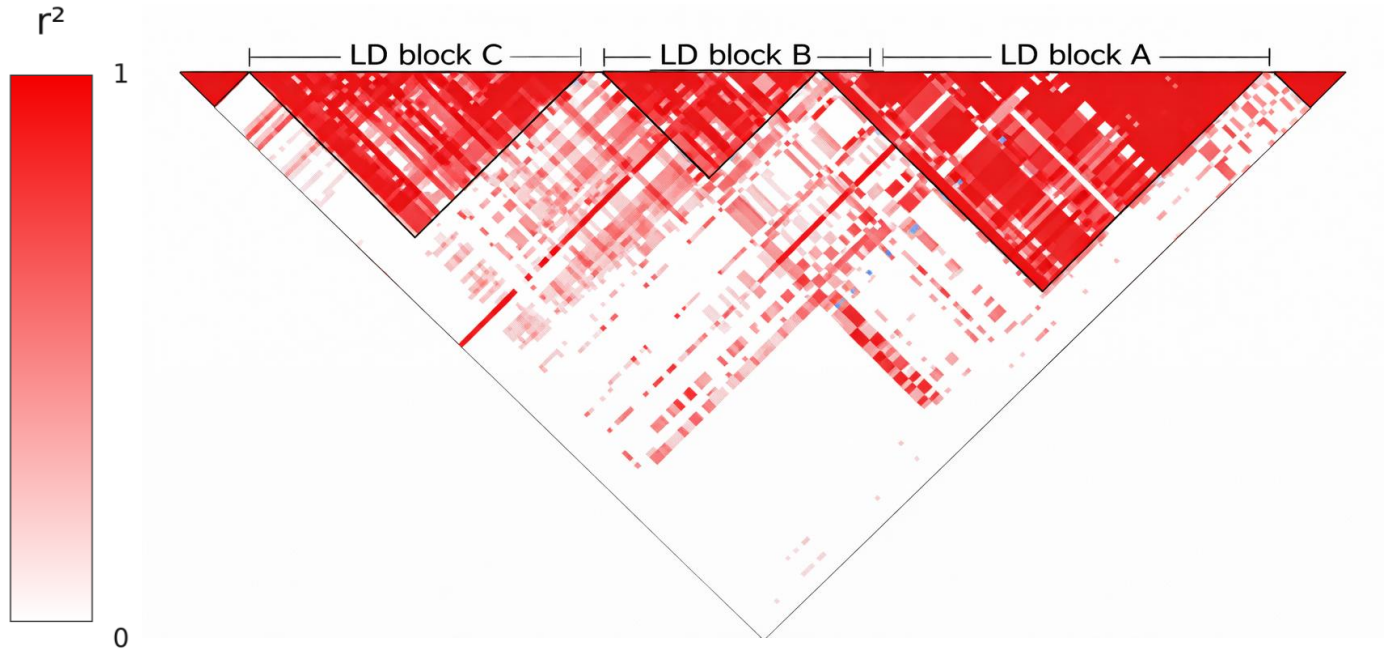
Linkage Disequilibrium (LD)

- Alleles are in LD when they're inherited together more frequently than expected (they're correlated).
- r^2 is between 0 (no LD) and 1 (perfect correlation).



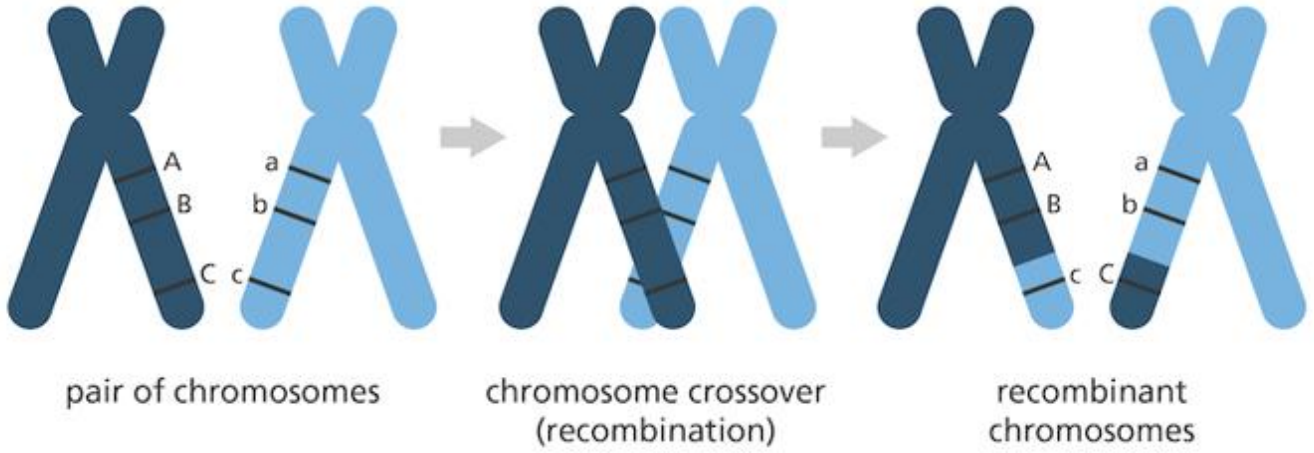
Linkage Disequilibrium (LD)

- Alleles are in LD when they're inherited together more frequently than expected (they're correlated).
- r^2 is between 0 (no LD) and 1 (perfect correlation).
- LD can be influenced by:
  - Recombination
  - Population history



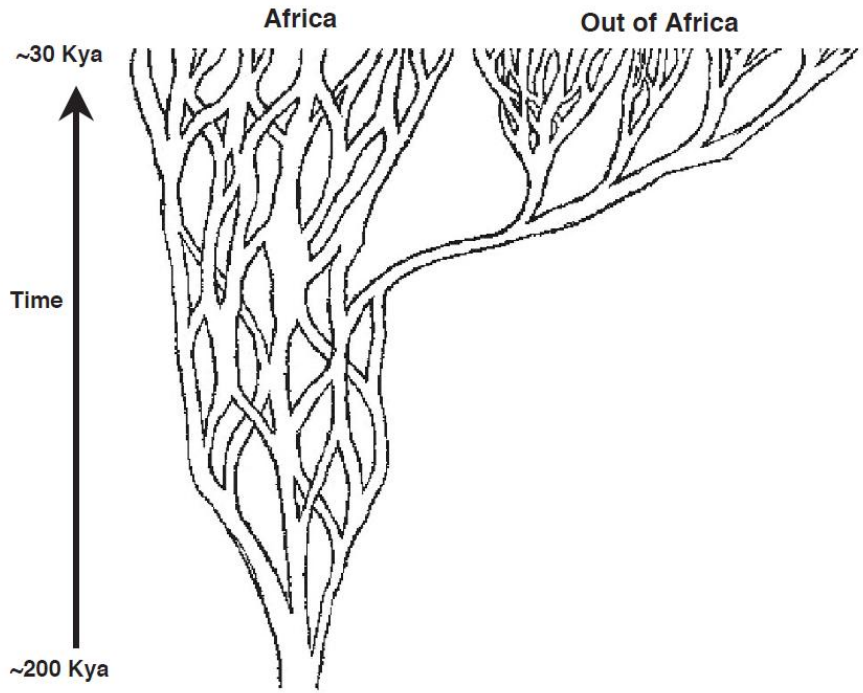
Linkage Disequilibrium (LD)

- Alleles are in LD when they're inherited together more frequently than expected (they're correlated).
- r^2 is between 0 (no LD) and 1 (perfect correlation).
- LD can be influenced by:
  - Recombination
  - Population history

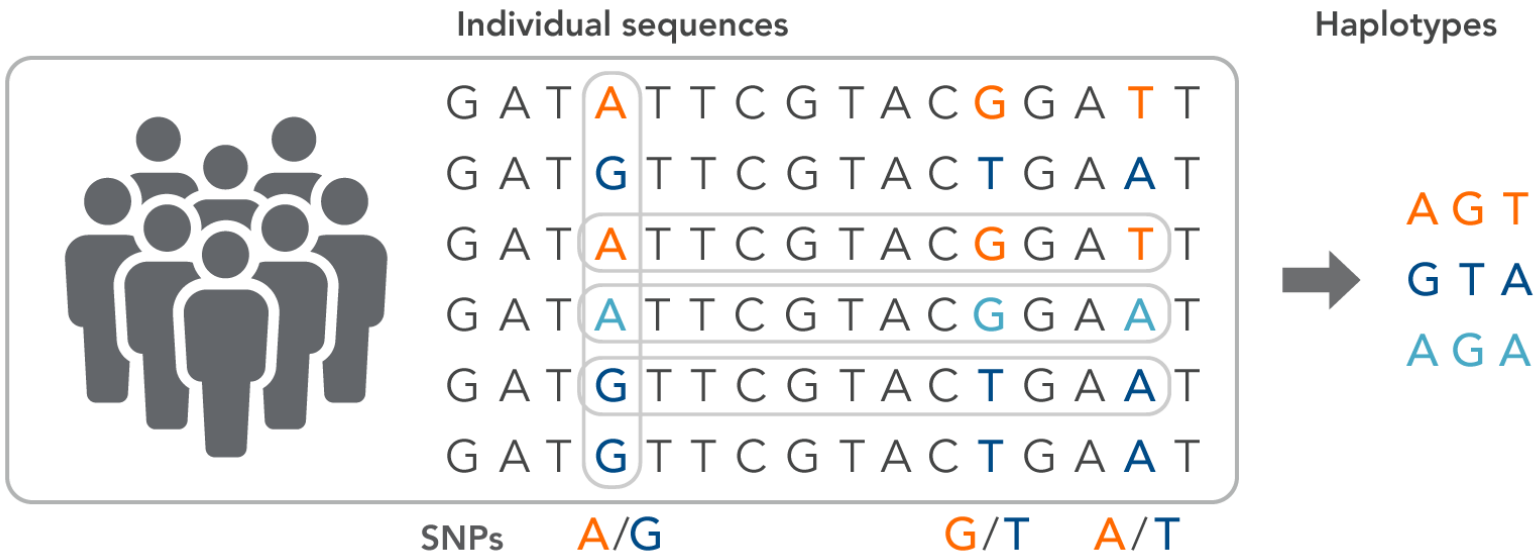


Linkage Disequilibrium (LD)

- Alleles are in LD when they're inherited together more frequently than expected (they're correlated).
- r^2 is between 0 (no LD) and 1 (perfect correlation).
- LD can be influenced by:
  - Recombination
  - Population history



Next step: create haplotype maps

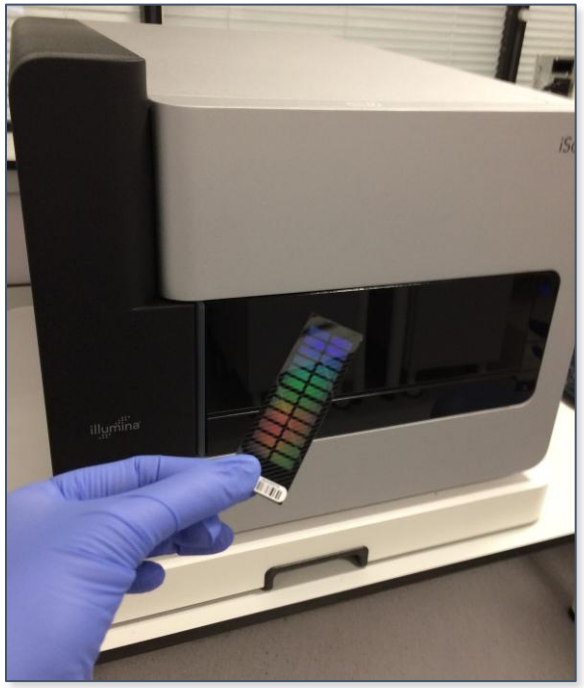


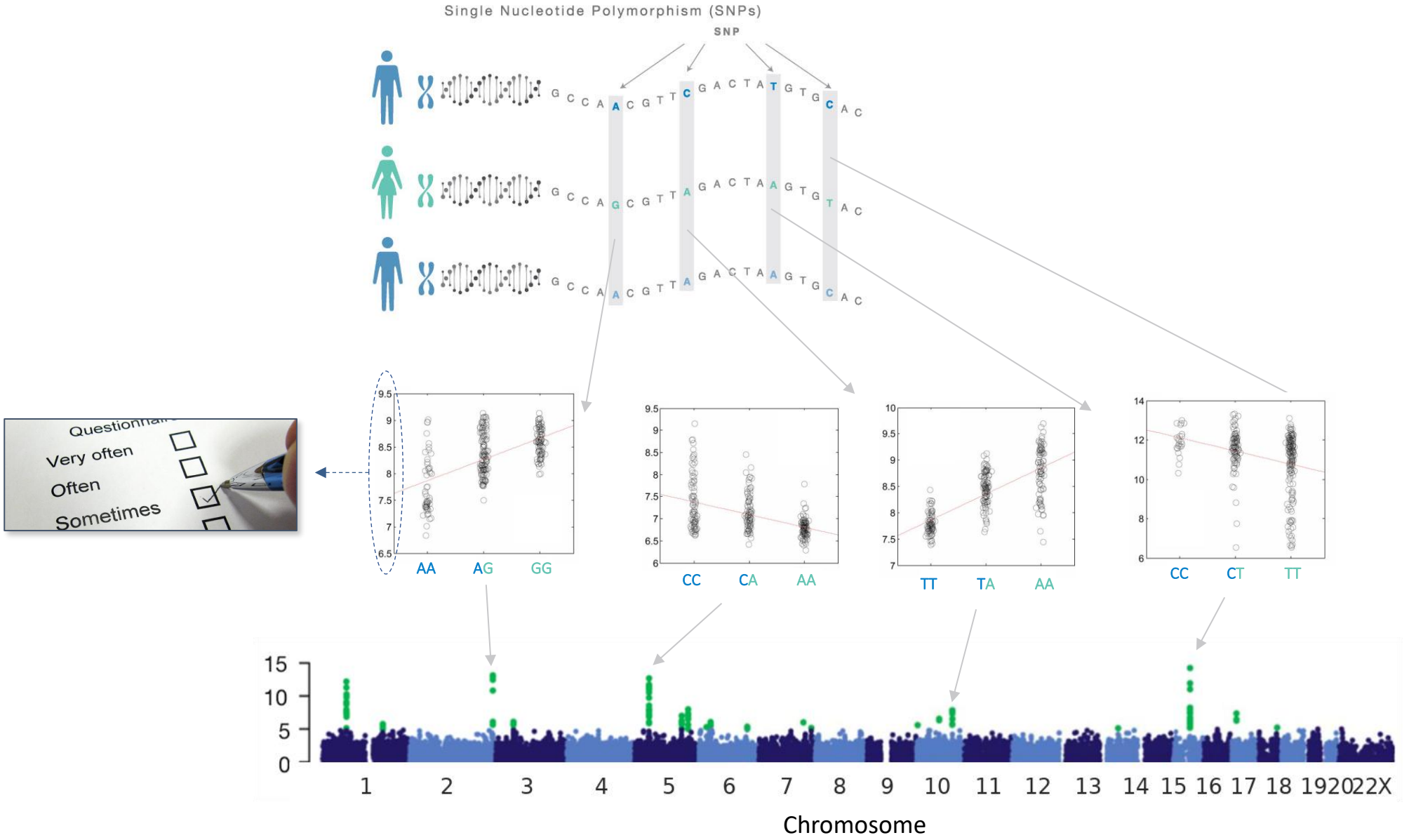
Next step: create haplotype maps by genotyping/sequencing populations

- 2005: 1st phase of the HapMap project: 269 individuals from 4 populations with several million well-defined genetic variants
- 2012: 1st phase of the 1000 Genomes project: 1,092 individuals from 14 populations were sequenced

Video:

DNA Microarray




Single Nucleotide Polymorphism (SNPs)
SNP
Very often
Often
Sometimes
AA
AG
GG
CC
CA
AA
TT
TA
AA
CC
CT
TT
Chromosome

When did GWAS begin?

SCIENCE • VOL. 273 • 13 SEPTEMBER 1996

The Future of Genetic Studies of Complex Human Diseases

Neil Risch and Kathleen Merikangas

Thus, the primary limitation of genome-wide association tests is not a statistical one but a technological one. A large number of genes (up to 100,000) and polymorphisms (preferentially ones that create alterations in derived proteins or their expression) must first be identified, and an extremely large number of such polymorphisms will need to be tested. Although testing such a large number of polymorphisms on several hundred, or even a thousand families, might currently seem implausible in scope, more efficient methods of screening a large number of polymorphisms (for example, sample pooling) may be possible.

		Linkage			Association			
					Singletons		Sib pairs	
Genotypic risk ratio (γ)	Frequency of disease allele A (p)	Probability of allele sharing (Y)	No. of families required (N)	Probability of transmitting disease allele A P(tr-A)	Proportion of heterozygous parents (Het)	(N)	(Het)	(N)
4.0	0.01	0.520	4260	0.800	0.048	1098	0.112	235
	0.10	0.597	185	0.800	0.346	150	0.537	48
	0.50	0.576	297	0.800	0.500	103	0.424	61
	0.80	0.529	2013	0.800	0.235	222	0.163	161
2.0	0.01	0.502	296,710	0.667	0.029	5823	0.043	1970
	0.10	0.518	5382	0.667	0.245	695	0.323	264
	0.50	0.526	2498	0.667	0.500	340	0.474	180
	0.80	0.512	11,917	0.667	0.267	640	0.217	394
1.5	0.01	0.501	4,620,807	0.600	0.025	19,320	0.031	7776
	0.10	0.505	67,816	0.600	0.197	2218	0.253	941
	0.50	0.510	17,997	0.600	0.500	949	0.490	484
	0.80	0.505	67,816	0.600	0.286	1663	0.253	941

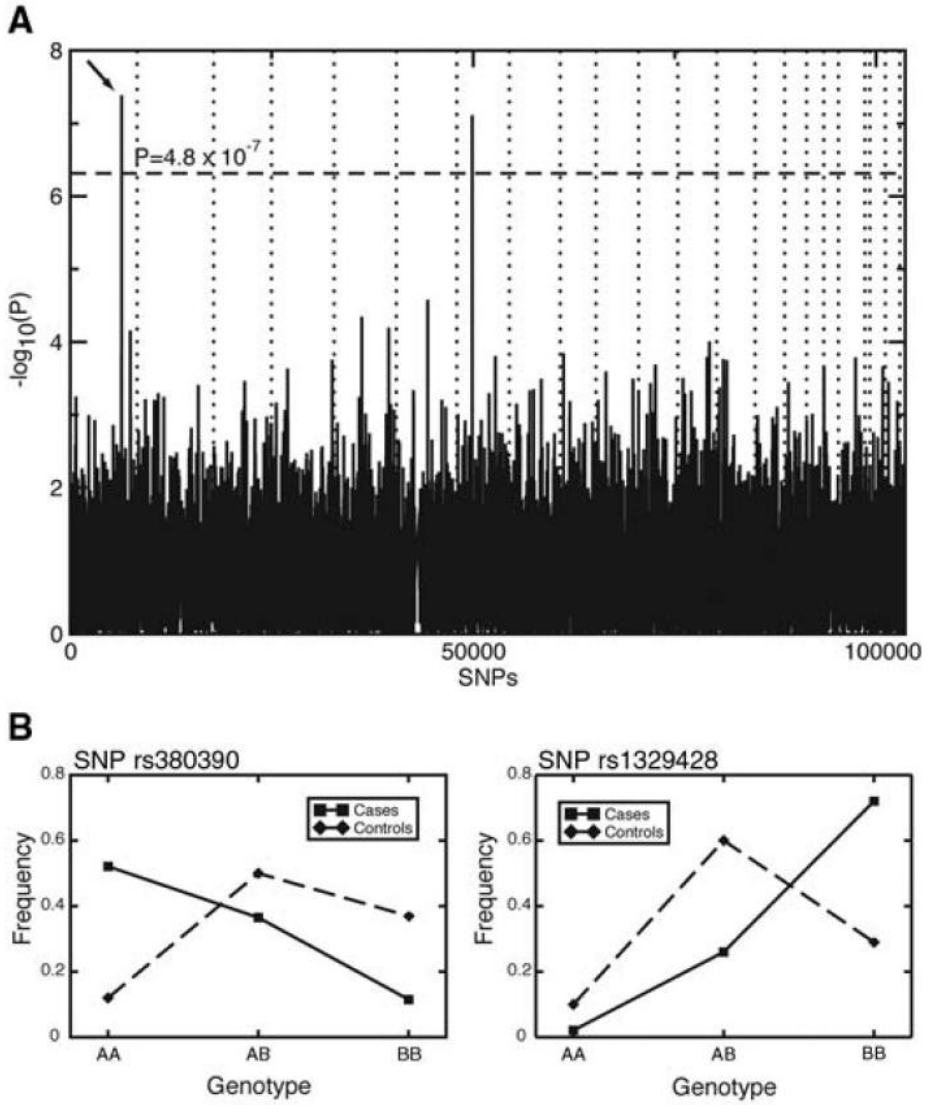
Comparison of linkage and association studies. Number of families needed for identification of a disease gene.

When did GWAS begin?

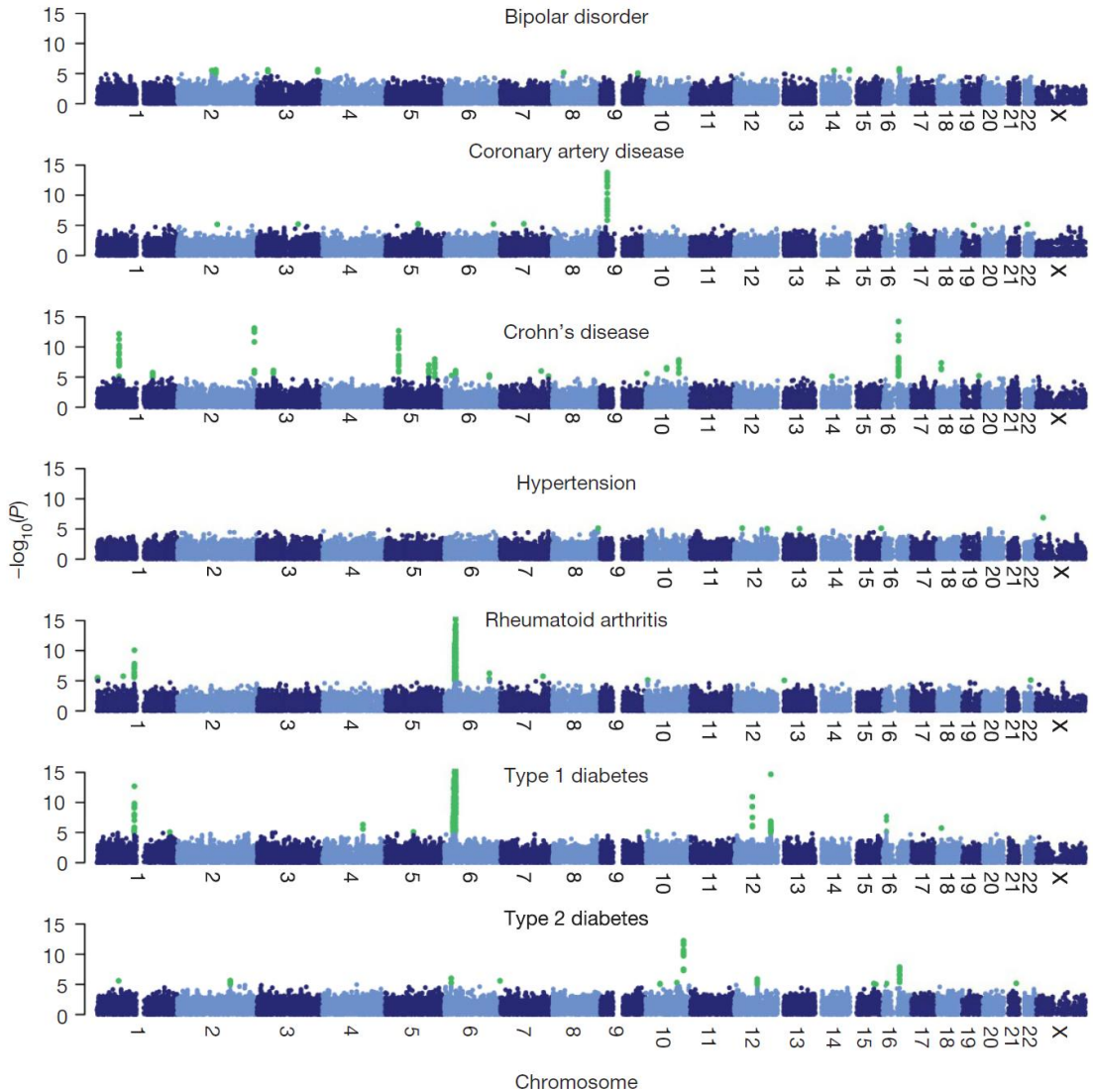
Complement Factor H Polymorphism in Age-Related Macular Degeneration

Robert J. Klein,[1] Caroline Zeiss,[2*] Emily Y. Chew,[3*]
Jen-Yue Tsai,[4*] Richard S. Sackler,[1] Chad Haynes,[1]
Alice K. Henning,[5] John Paul SanGiovanni,[3] Shrikant M. Mane,[6]
Susan T. Mayne,[7] Michael B. Bracken,[7] Frederick L. Ferris,[3]
Jurg Ott,[1] Colin Barnstable,[2] Josephine Hoh[7†]

www.sciencemag.org SCIENCE VOL 308 15 APRIL 2005



- 96 cases and 50 controls
- 116,204 SNPs

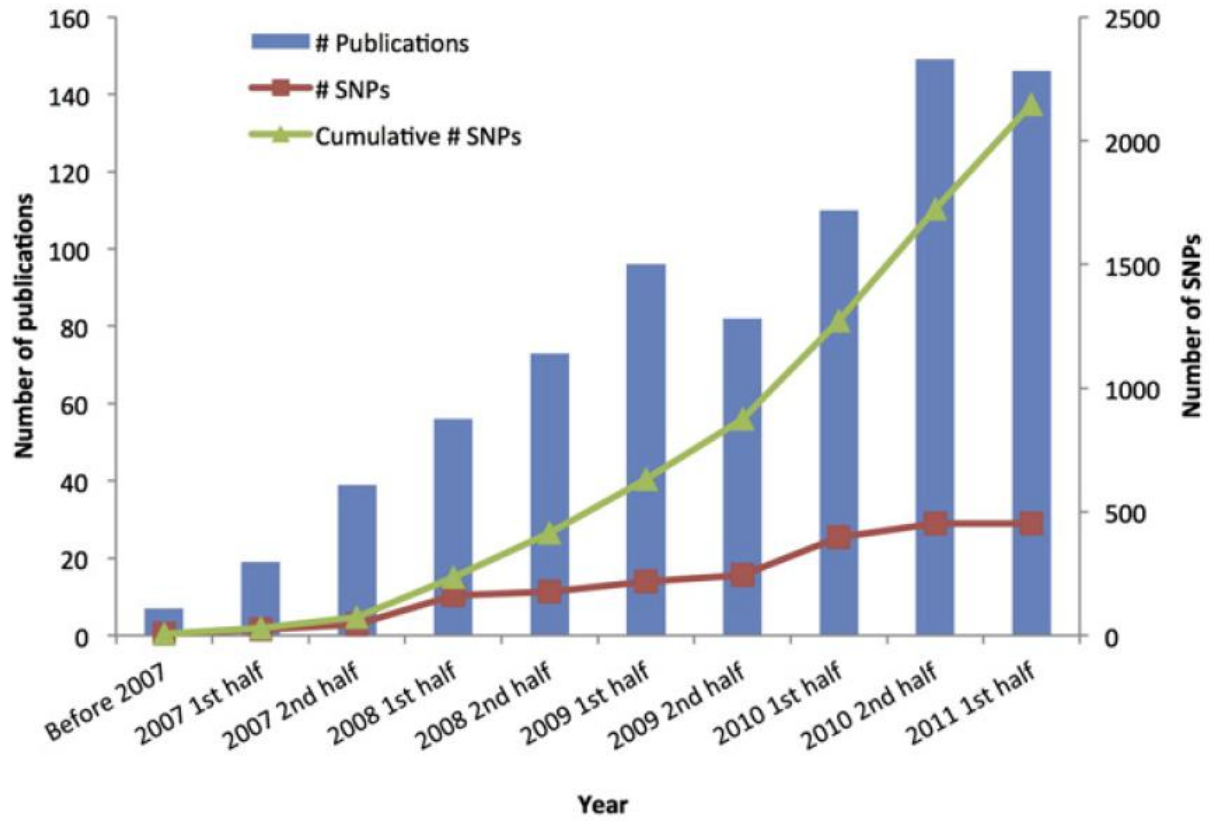


Vol 447|7 June 2007|doi:10.1038/nature05911 nature

ARTICLES

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

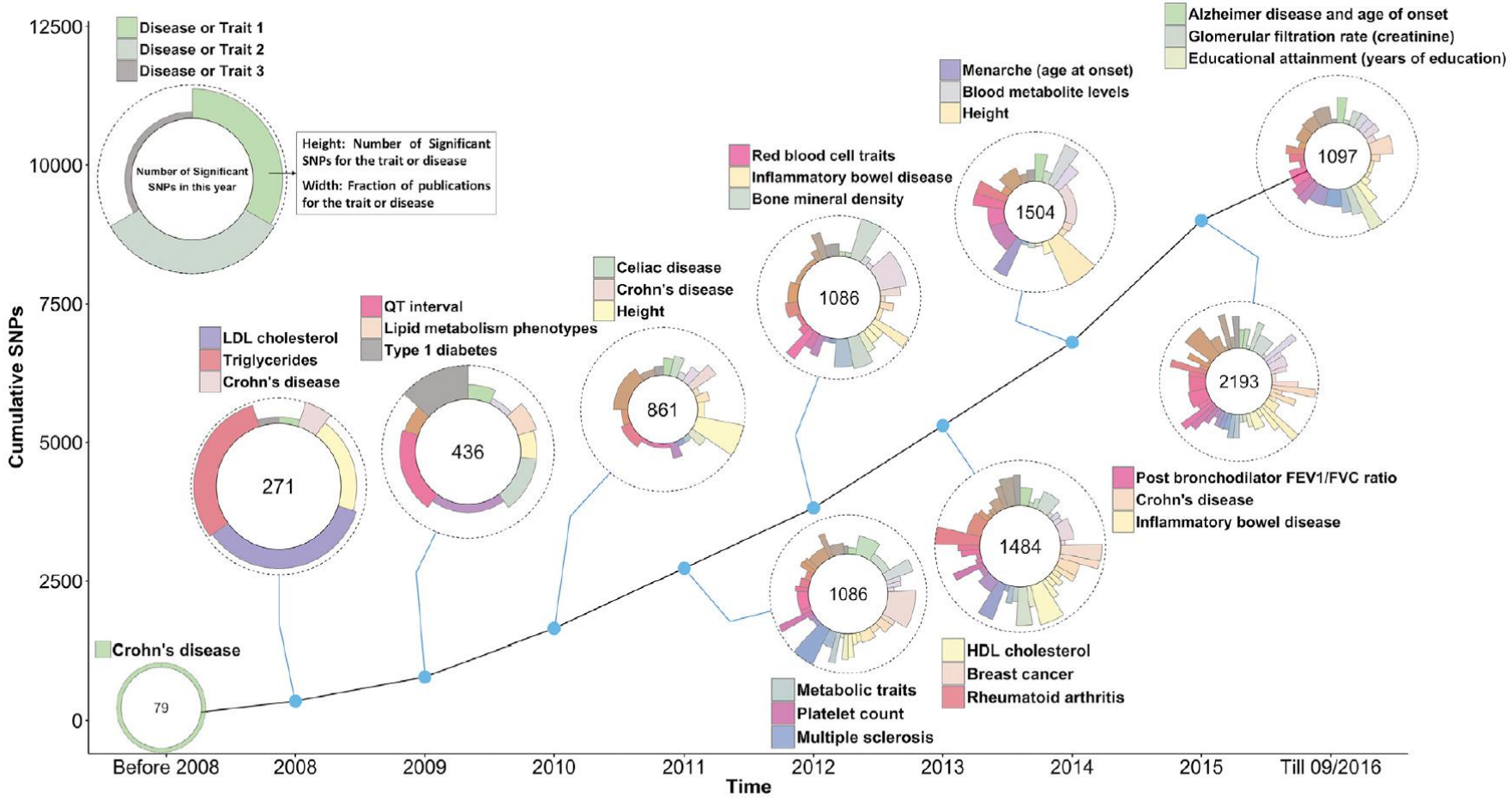
The Wellcome Trust Case Control Consortium*



The American Journal of Human Genetics REVIEW

Five Years of GWAS Discovery

Peter M. Visscher,[1,2,*] Matthew A. Brown,[1] Mark I. McCarthy,[3,4] and Jian Yang[5]

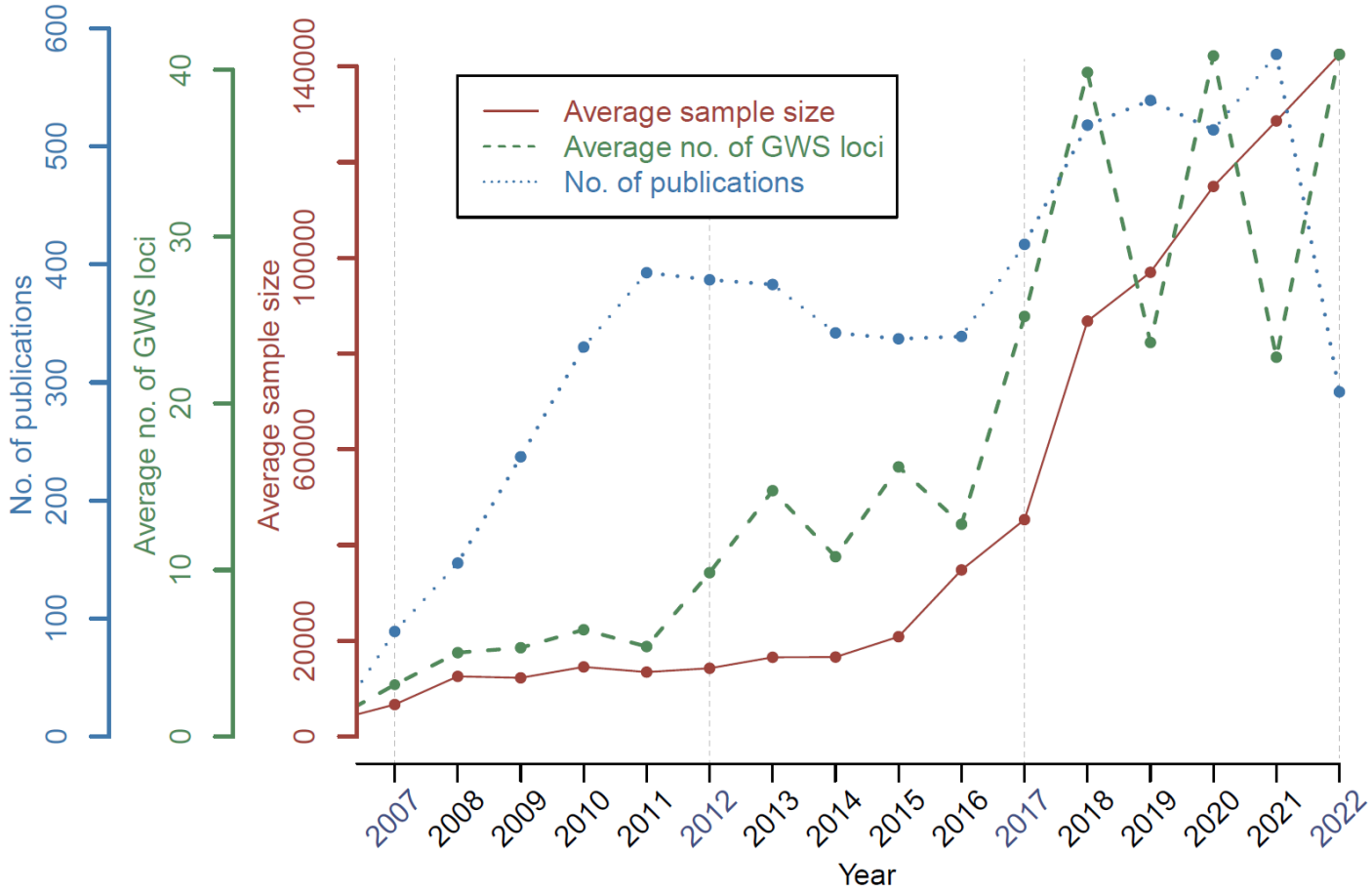


The American Journal of Human Genetics

REVIEW

10 Years of GWAS Discovery: Biology, Function, and Translation

Peter M. Visscher,[1,2,*] Naomi R. Wray,[1,2] Qian Zhang,[1] Pamela Sklar,[3] Mark I. McCarthy,[4,5,6] Matthew A. Brown,[7] and Jian Yang[1,2]



The American Journal of Human Genetics

REVIEW

15 years of GWAS discovery: Realizing the promise

Abdel Abdellaoui,[1,*] Loic Yengo,[2] Karin J.H. Verweij,[1] and Peter M. Visscher[2]

N = ~500,000

N = ~220,000

N = ~5,000,000

N = ~200,000

N = ~500,000

N = ~200,000

N = ~170,000

N = ~200,000

N = ~500,000

N = ~750,000

N = ~900,000

N = ~14,000,000

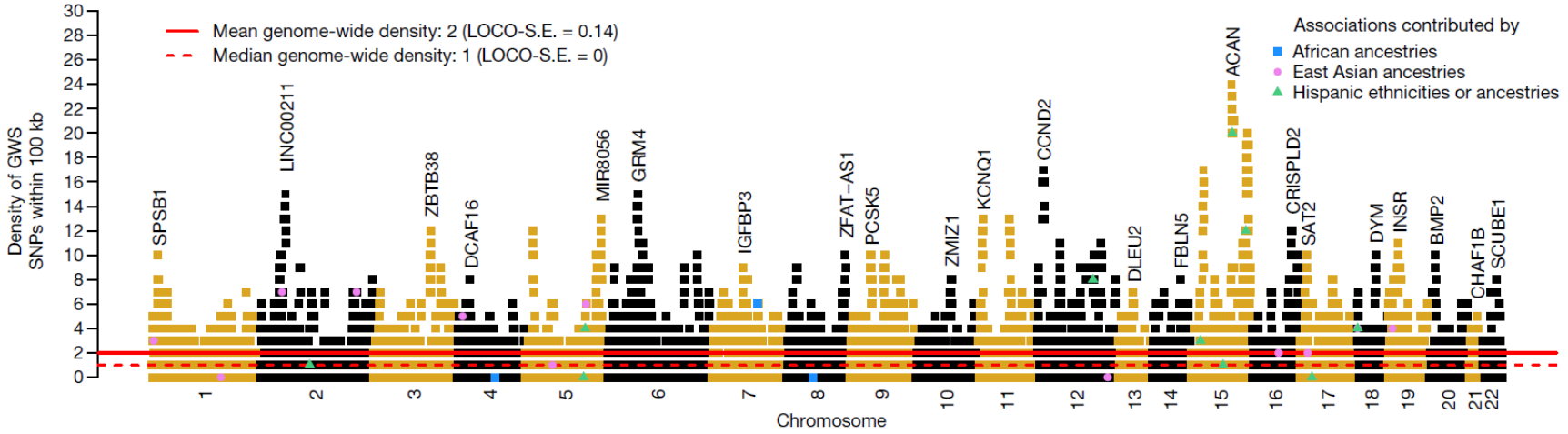
N = ~200,000

Article

A saturated map of common genetic variants associated with human height

Nature | Vol 610 | 27 October 2022

N = 5.4 million

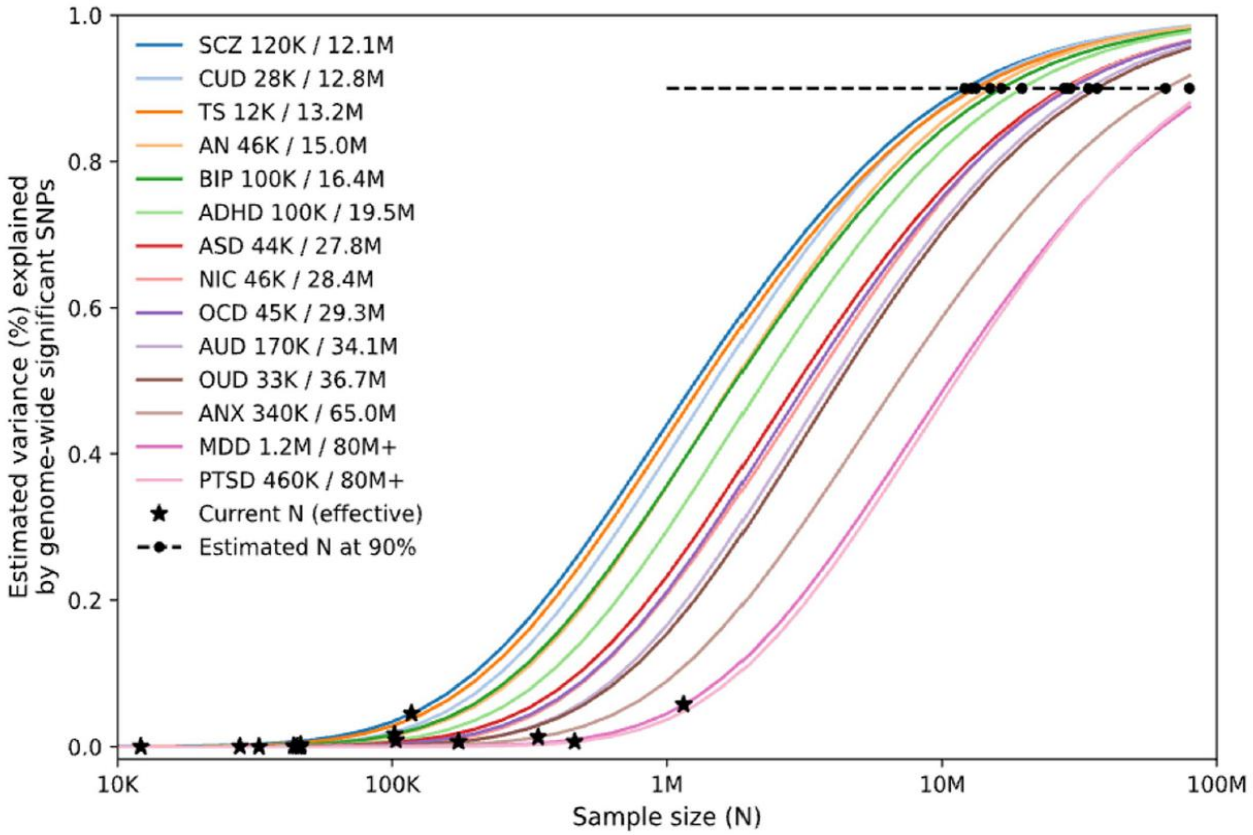


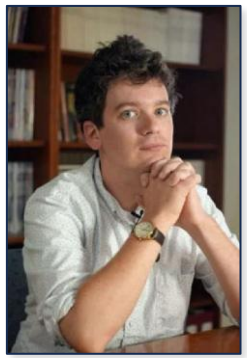
Loic Yengo et al, 2022

Article

Mapping the genetic landscape across 14 psychiatric disorders

Nature | Vol 649 | 8 January 2026





Andrew Grotzinger et al, 2026

Next video: signal versus noise