Summer Institutes of Statistical Genetics, 2022

Module 2: INTRODUCTION TO GENETICS AND GENOMICS
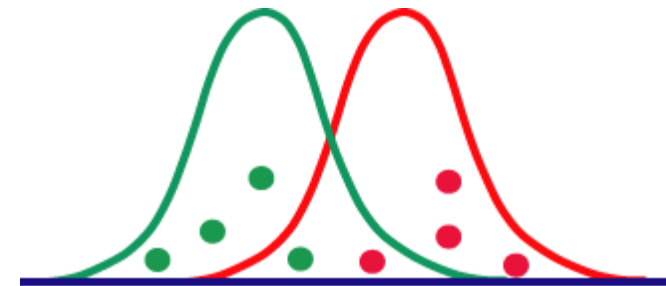
Greg Gibson and Joe Lachance

Georgia Institute of Technology

Lecture 5: GENOME-WIDE ASSOCIATION STUDIES

greg.gibson@biology.gatech.edu

http://www.cig.gatech.edu

Principle of Association Studies

| Individual	Site	Score
1	A T C **C** G A	**9**
2	A C T **C** G A	**8**
3	A C C A – G	3
4	T T C A G A	5
5	A T C A G A	2
6	A C C **C** – G	**7**
7	T C T A – G	4
8	A T C **C** G A	**8**

Are the phenotype scores associated with each class of SNP drawn from the same or different distributions ?

Linkage versus Association

Linkage examines recent recombination events in a pedigree:

- over just several generations

- large chromosomal regions detected

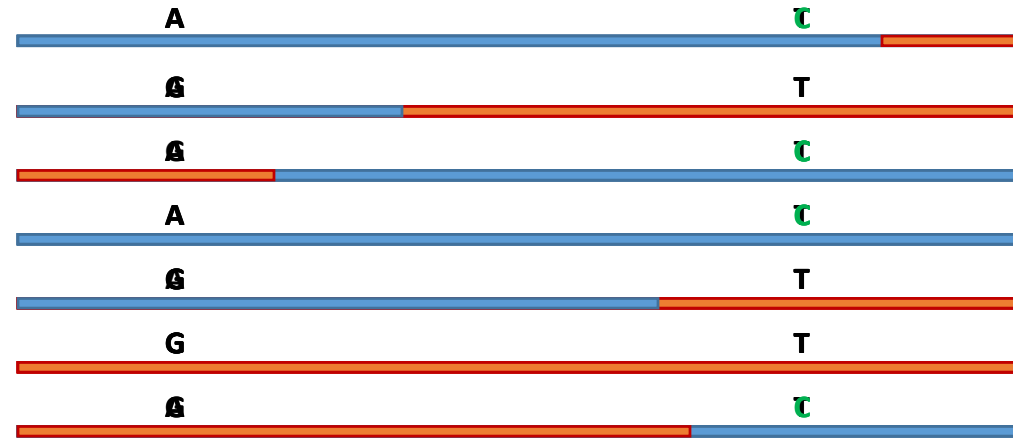- no information on allele frequency
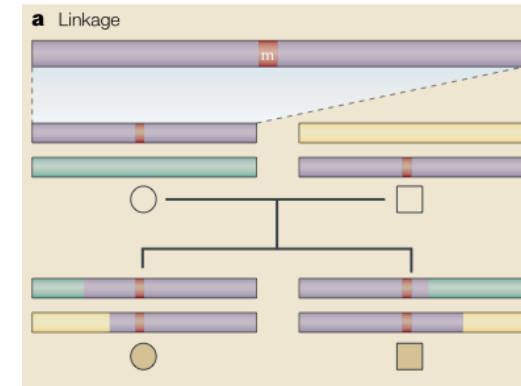
Association examines historical recombination events in a population:

- basically a 10,000 generation pedigree
- resolution to single genes
- estimates effect size and frequency

Cardon & Bell. 2003. *Nat. Rev. Genet*. **2**: 91
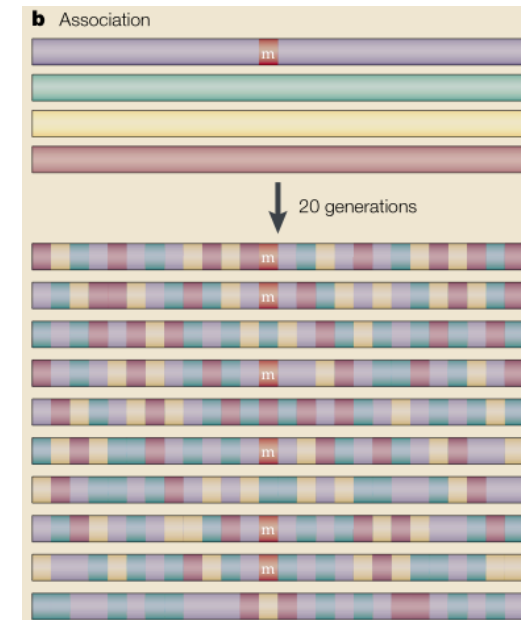
Why LD (Linkage Disequlibrium) happens

When a mutation occurs, by definition it is only on one chromosome and hence "associated" with the genotypes elsewhere on that chromosome.

Over time, the mutation increases in frequency and becomes a polymorphism. It remains in LD with the genotypes on the chromosome it appeared on.

Eventually recombination breaks up the LD, in proportion to genetic distance.

Measurement of LD

LD is the non-random association of genotypes.

Expected

		AA	AG	GG
		24	48	24
TT	24	6	12	6
TC	48	12	24	12
CC	24	6	12	6

Observed

		AA	AG	GG
		24	48	24
TT	24	24	0	0
TC	48	0	48	0
CC	24	0	0	24

LD can be quantified as a proportion of the maximal possible LD given the allele frequencies (D'),
Or as the squared correlation between allele frequencies (r^2).

Haplotypes and Tagging SNPs

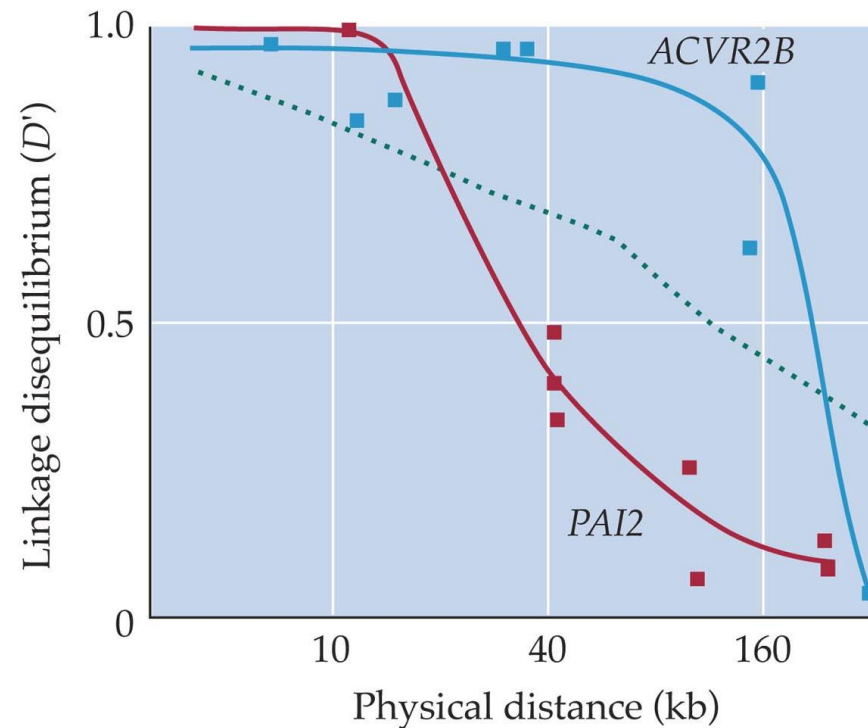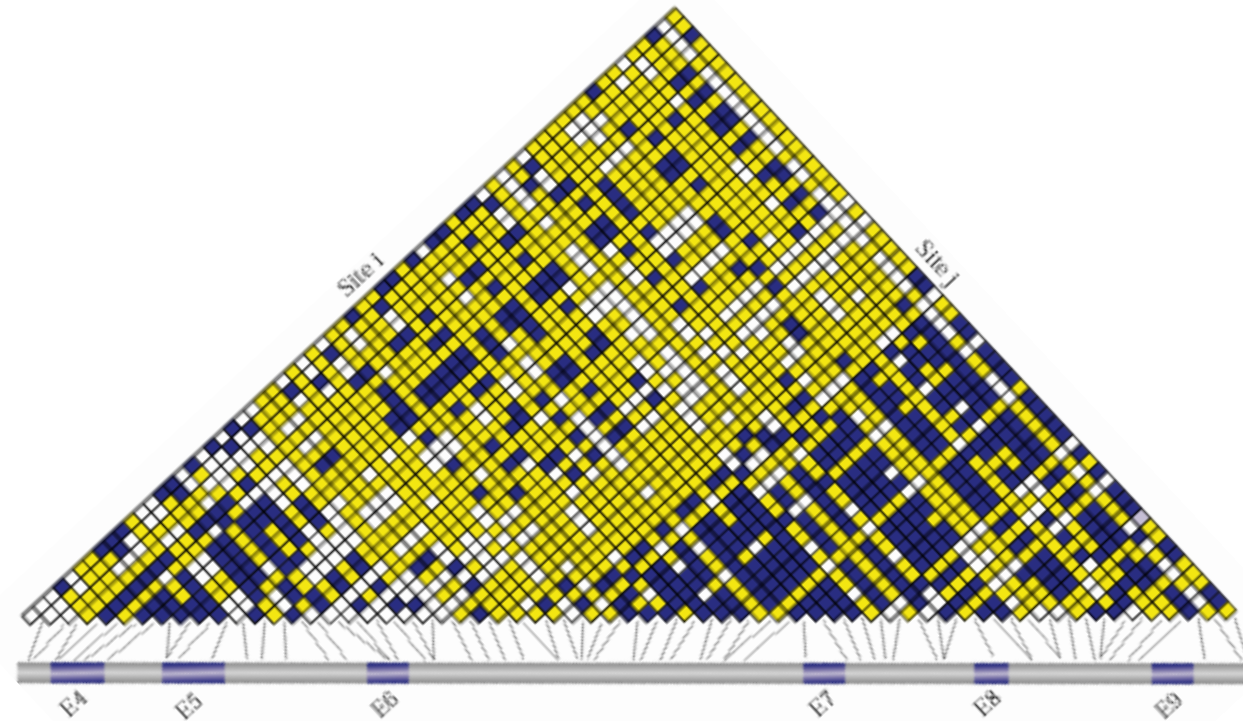# Visualizing LD: The LPL example
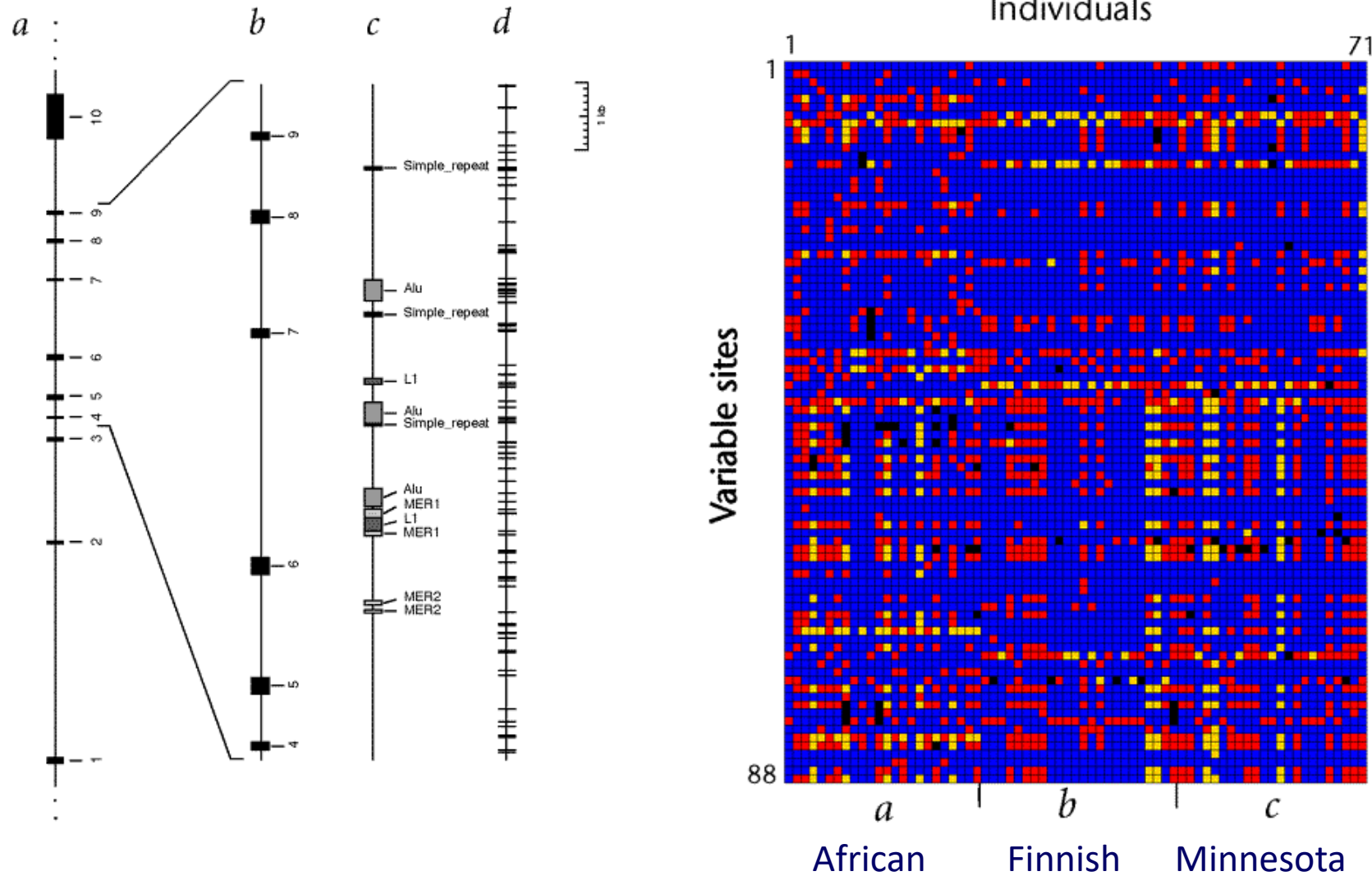
An LD Plot (for the LPL locus)

Case-Control and Family Designs
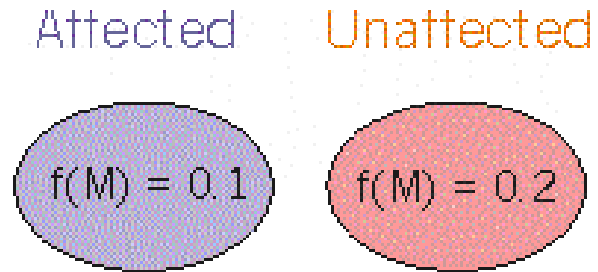
	Observed		Expected	
	Allele M	Allele m	Allele M	Allele m
Affected	34	278	61	265
Unaffected	69	256	62	269

Affected f(M) = 0.1

Unaffected f(M) = 0.2

$\chi^2 = 14.0$ P <0.001

$$\frac{M}{m} \; X \; \frac{M}{m}$$

f(M) = 0.63
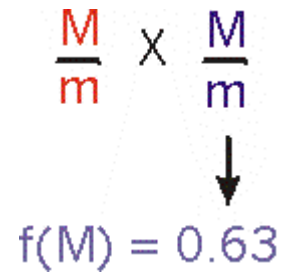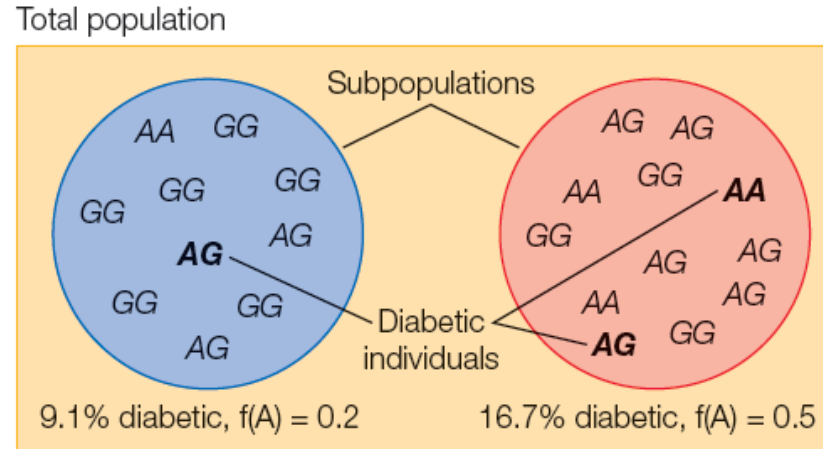
Transmitted Allele		
	M	m
Observed	78	46
Expected	62	62

$\chi^2 = 8.2$ P < 0.001

Repeatability and Forest Plots
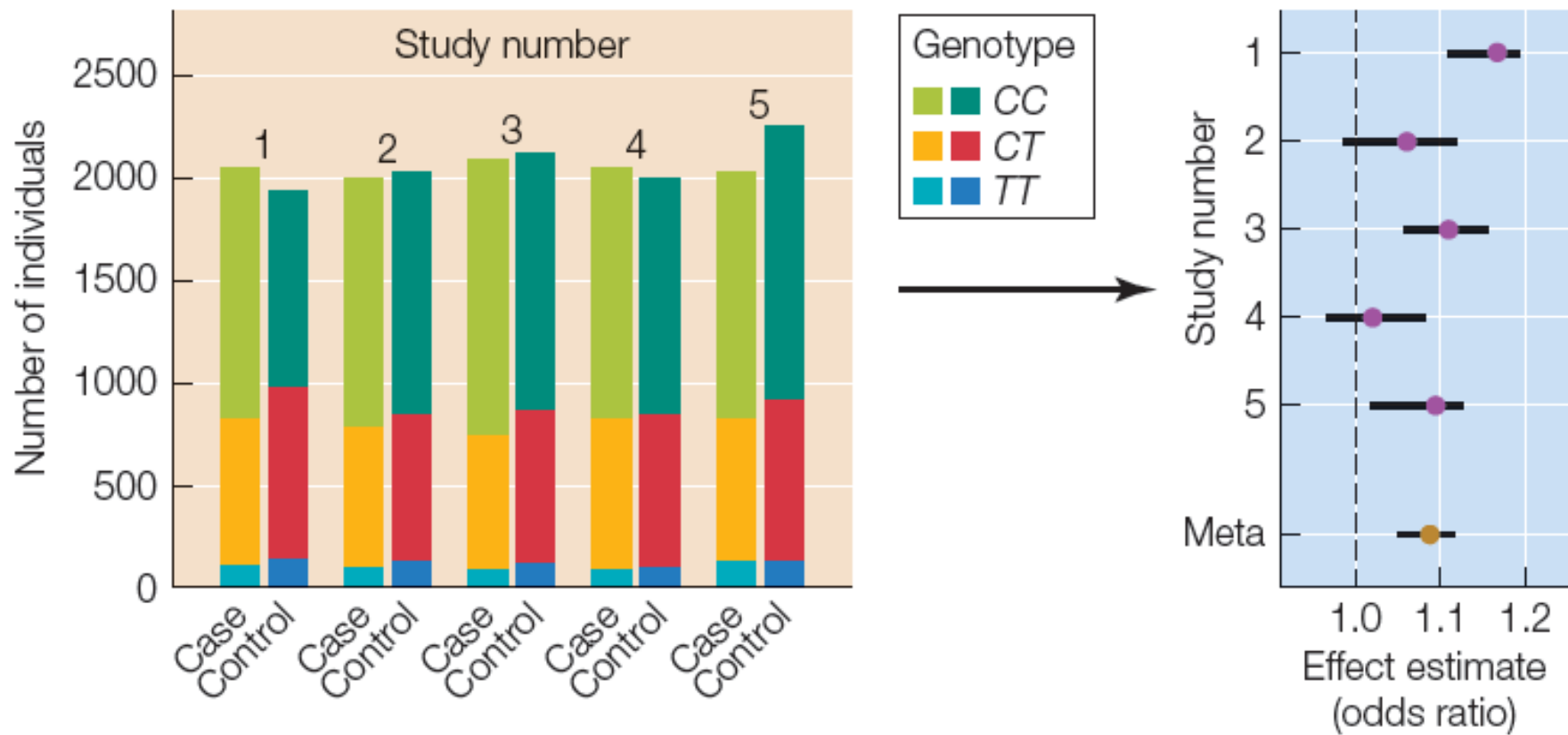
Population Structure

If the allele frequency AND the trait frequency vary among hidden sub-populations, false positives can arise

Total population

Subpopulations

AA GG
GG
GG GG
GG AG
AG
GG GG
AG

AG AG
GG
AA **AA**
GG
AG AG
AA AG
AG GG

Diabetic individuals

9.1% diabetic, f(A) = 0.2

16.7% diabetic, f(A) = 0.5

Blue subpopulation

	AA	AG	GG
Case	80	640	1280
Control	800	6400	12,800
Case/control	0.1	0.1	0.1

Red subpopulation

	AA	AG	GG
Case	200	400	200
Control	1000	2000	1000
Case/control	0.2	0.2	0.2

Total population

	AA	AG	GG
Case	280	1040	1480
Control	1800	8400	13,800
Case/control	0.155	0.124	0.107

Odds ratio (A:G) = 1.2 $p = 10^{-8}$
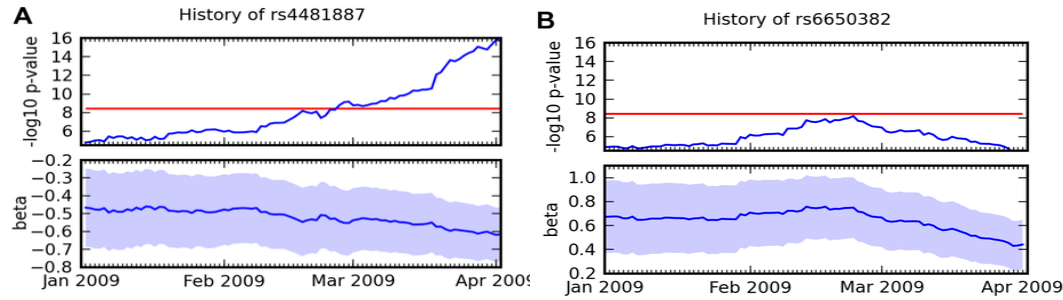
GWAS in 2009: The WTCCC
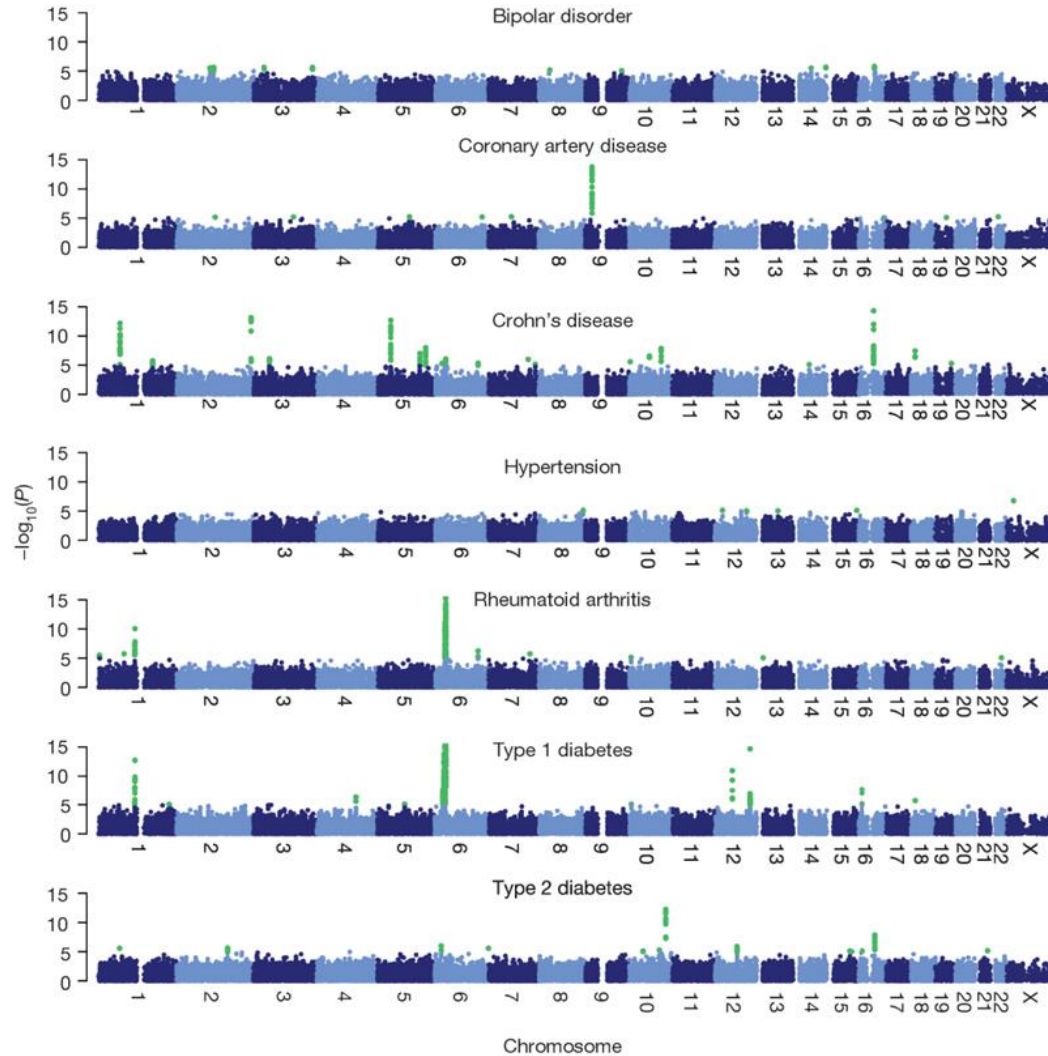
GWAS first appeared 10 years ago, now several new diseases each month

Inflammatory diseases show multiple associations, with some common variants (notably the MHC)

Depression and Hypertension show nothing: likely no variants with a relative risk greater than 1.5

Q-Q Plots in 23andme studies
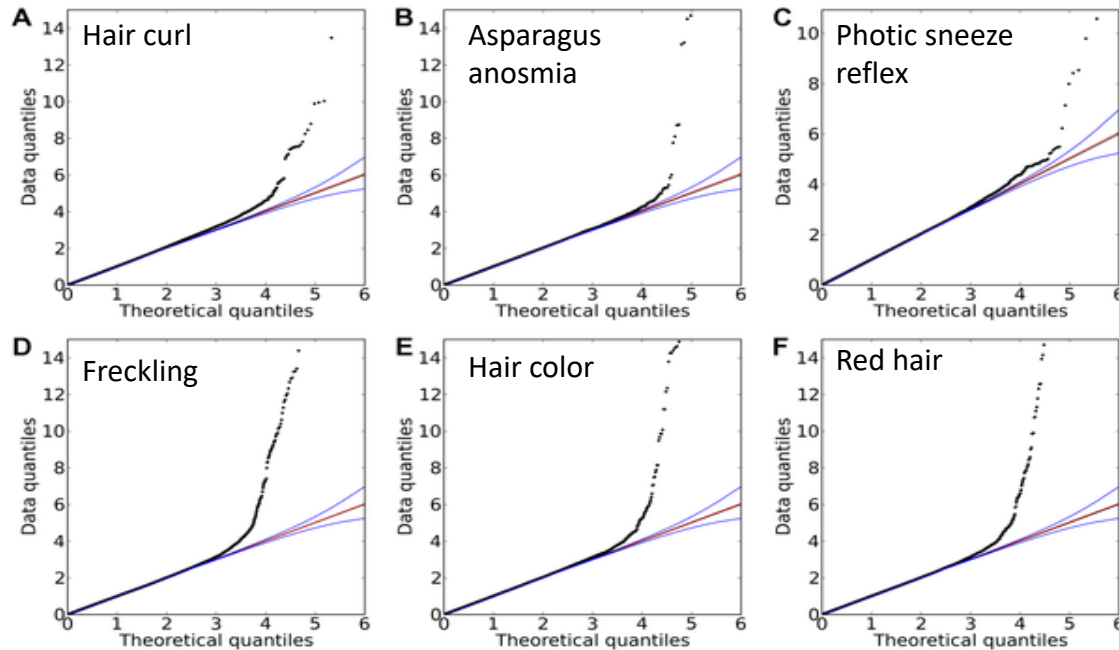

Other interesting traits:

Endurance Runner vs Sprinter (30% of people change their answer if they know their ACTN3)

Left vs Right Handedness (nothing striking)

Have you ever needed braces or wisdom teeth surgery?

Breast size (finds breast cancer risk loci)

Hand-clasp dominance ...

Eriksson et al, (2010) *PLoS Genetics* **6**: e1000993

Genetics of Obesity

R.J.F. Loos / Best Practice & Research Clinical Endocrinology & Metabolism 26 (2012) 211–226

Heritability of obesity ~ 60%

2/3 Americans BMI > 25

One gene, FTO, is repeatedly associated with BMI, hip circumference and weight, in most human populations

Homozygote classes differ in weight by up to 2 kg
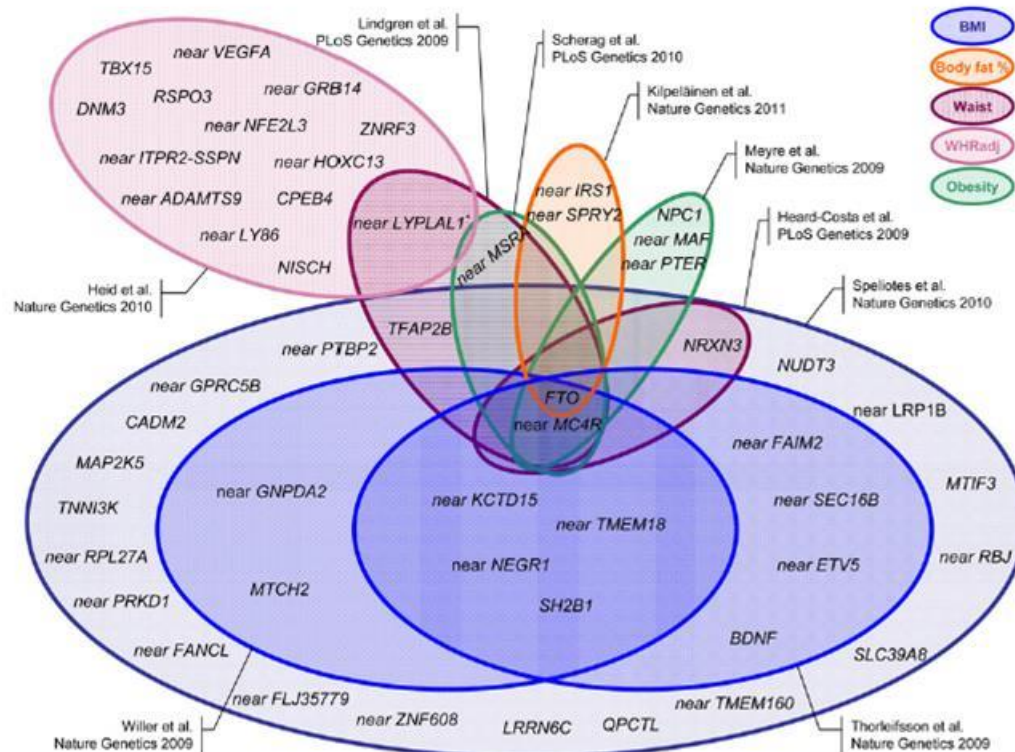
Study of 230,000 people →

49 loci for WHR, many linked to adipose, insulin biology   20 loci only in women
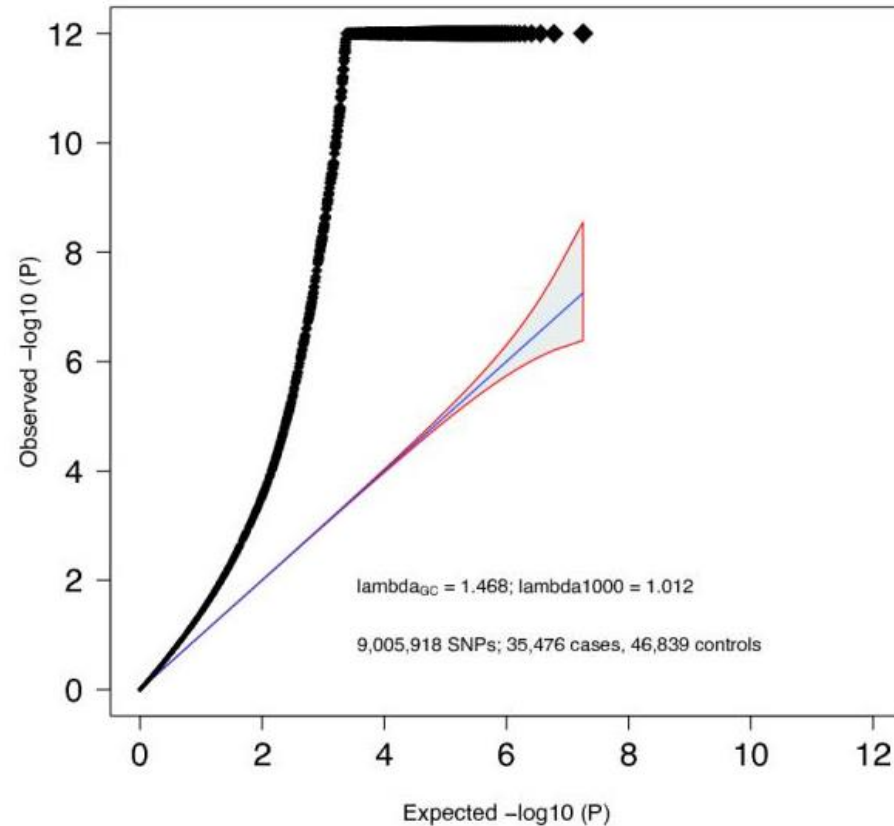
Study of 340,000 people →

97 loci for BMI, many linked to neuronal function

Little overlap with WHR

Genetics of Schizophrenia

128 independent SNP associations from GWAS of 37,000 cases

Strong enrichment in genes expressed in certain neuronal cell types or implicated in synaptic transmission

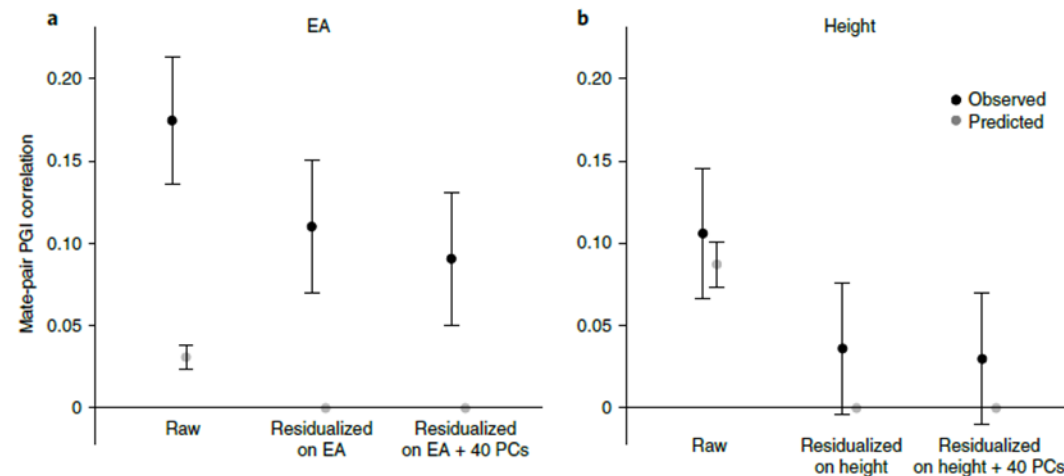But at least 5% of cases attributable to CNV: copy number variation

3 major chromosomal deletions of >100kb at frequency <1% are almost exclusively found in schizophrenics
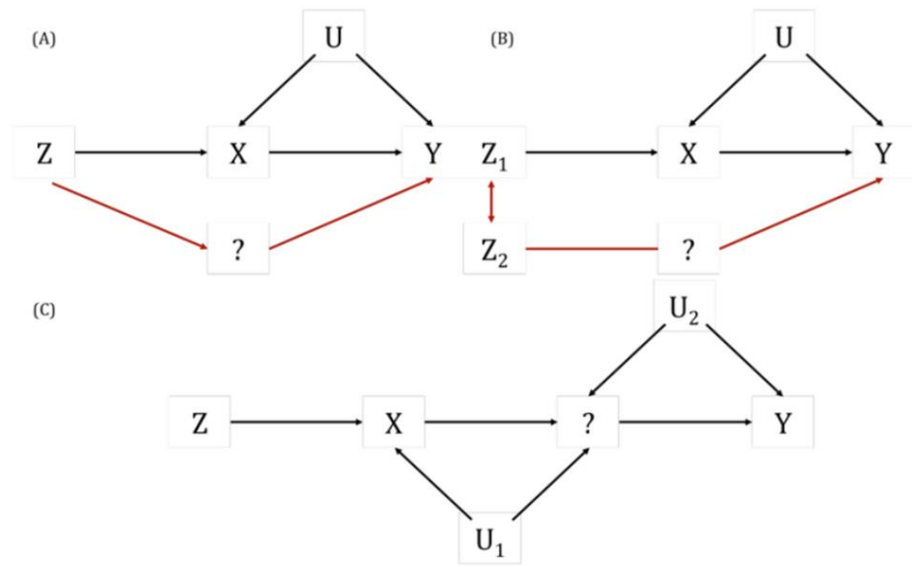
Genetics of Educational Attainment (on 3M people)

A PGI for Educational Attainment is also predictive of a wide rang of health outcomes
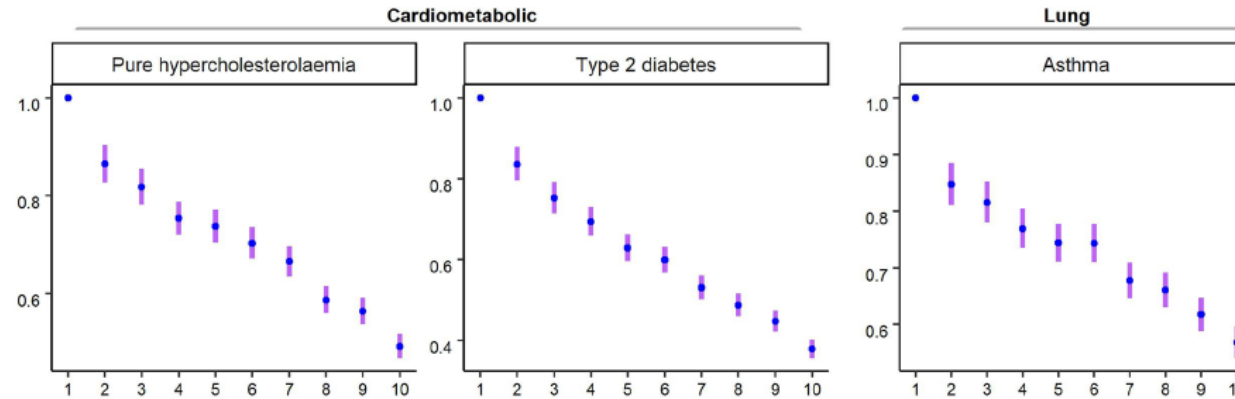
Couples are much more genetically similar for EA (but not height) than expected given their phenotypes

Okbay et al (2022) *Nature Genetics* **54**: 437-449

Mendelian Randomization establishes Causality

Hemani et al (2018) *eLife* **7**: e34408; Astore and Gibson, submitted