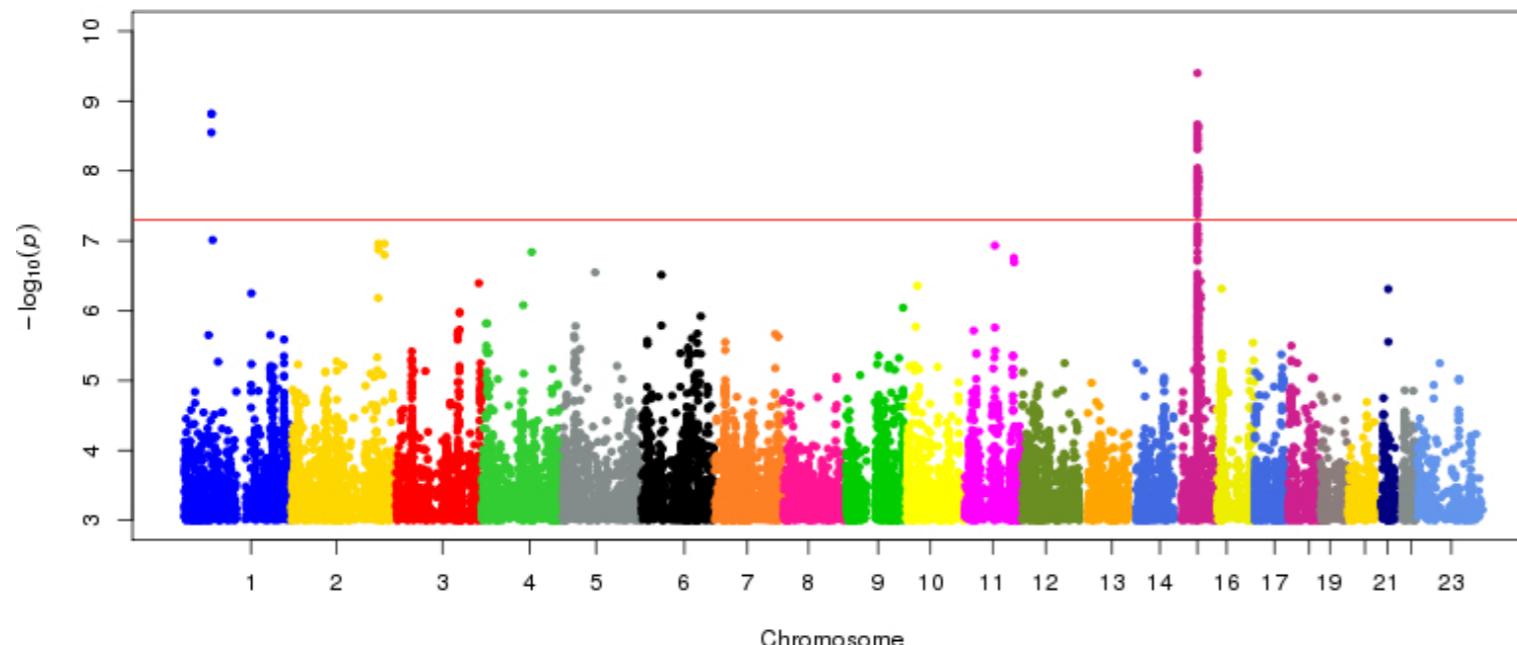


# Bio393: Genetic Analysis

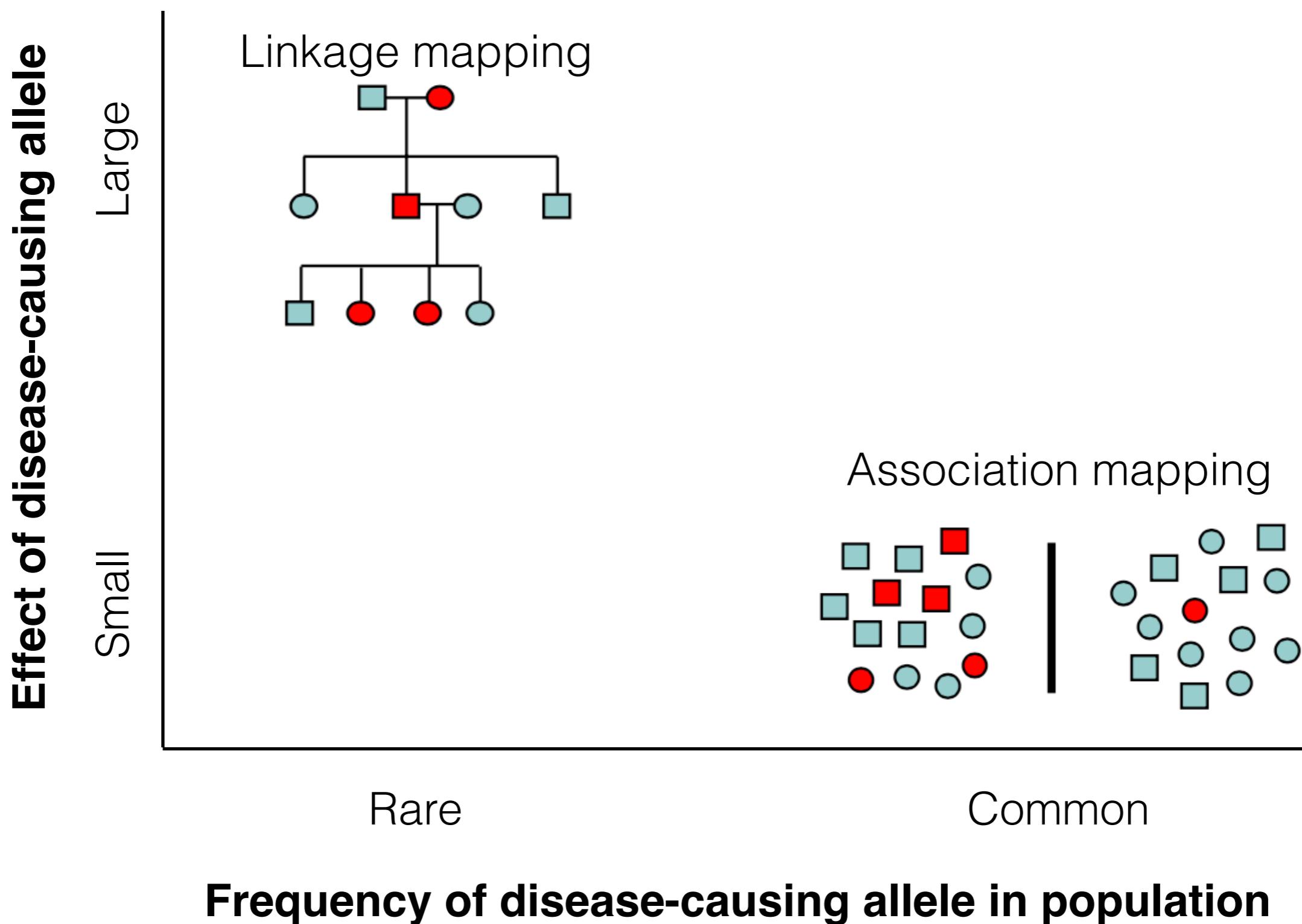
GWAS, relative risk, complex traits, and the future of genetic medicine



Styrkarsdottir *et al.* Nature 2014

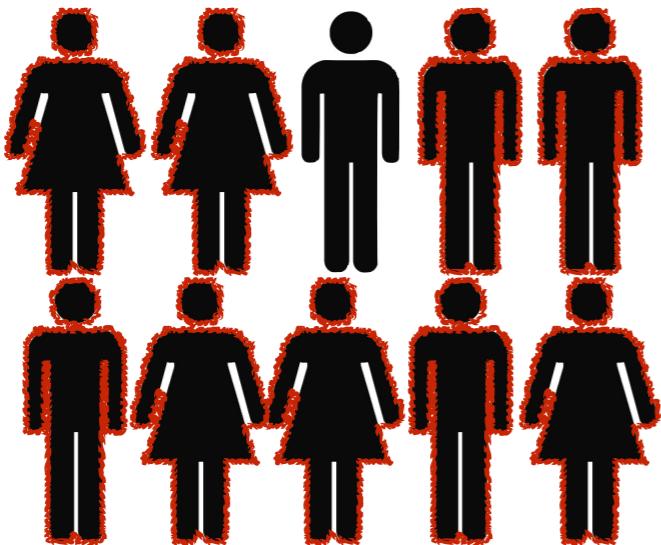
**Please send two review questions by Sunday night (8 PM)**

# Human gene mapping has two general flavors

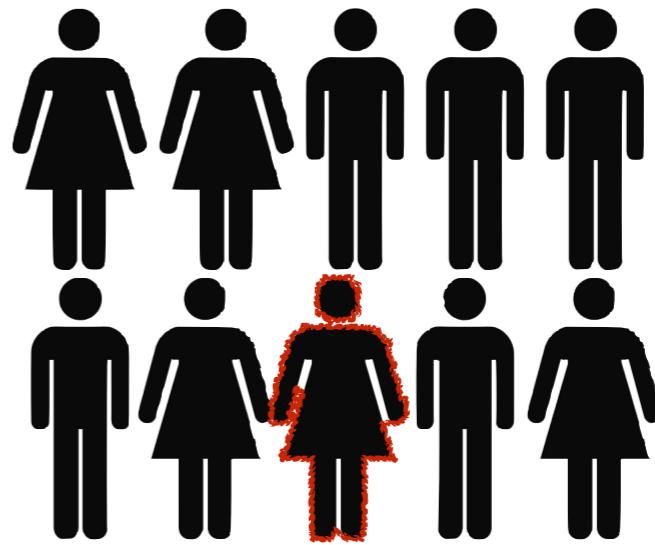


# The set up of a genome-wide association (GWA) mapping

Case-control study design



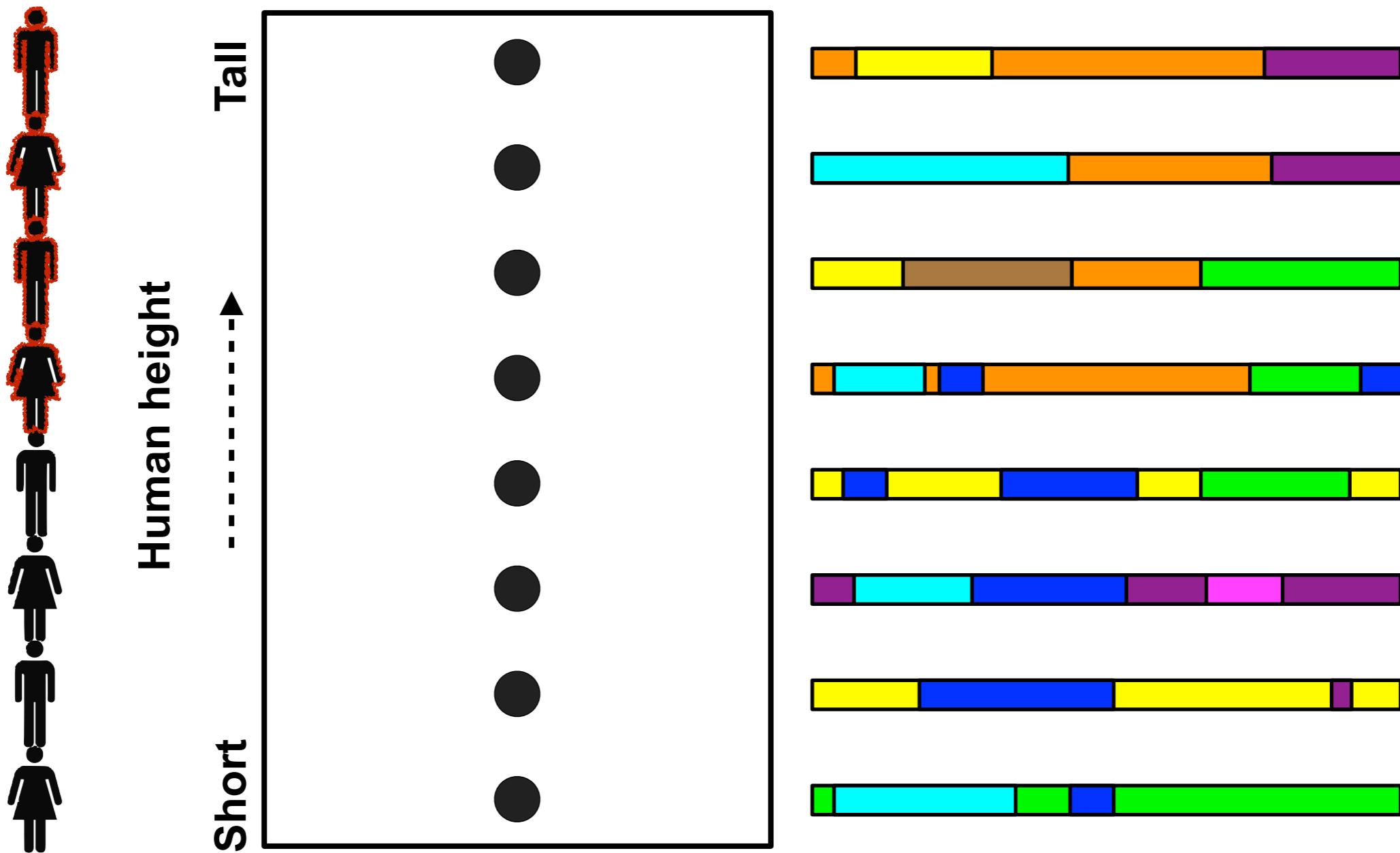
Cases  
(People with trait)



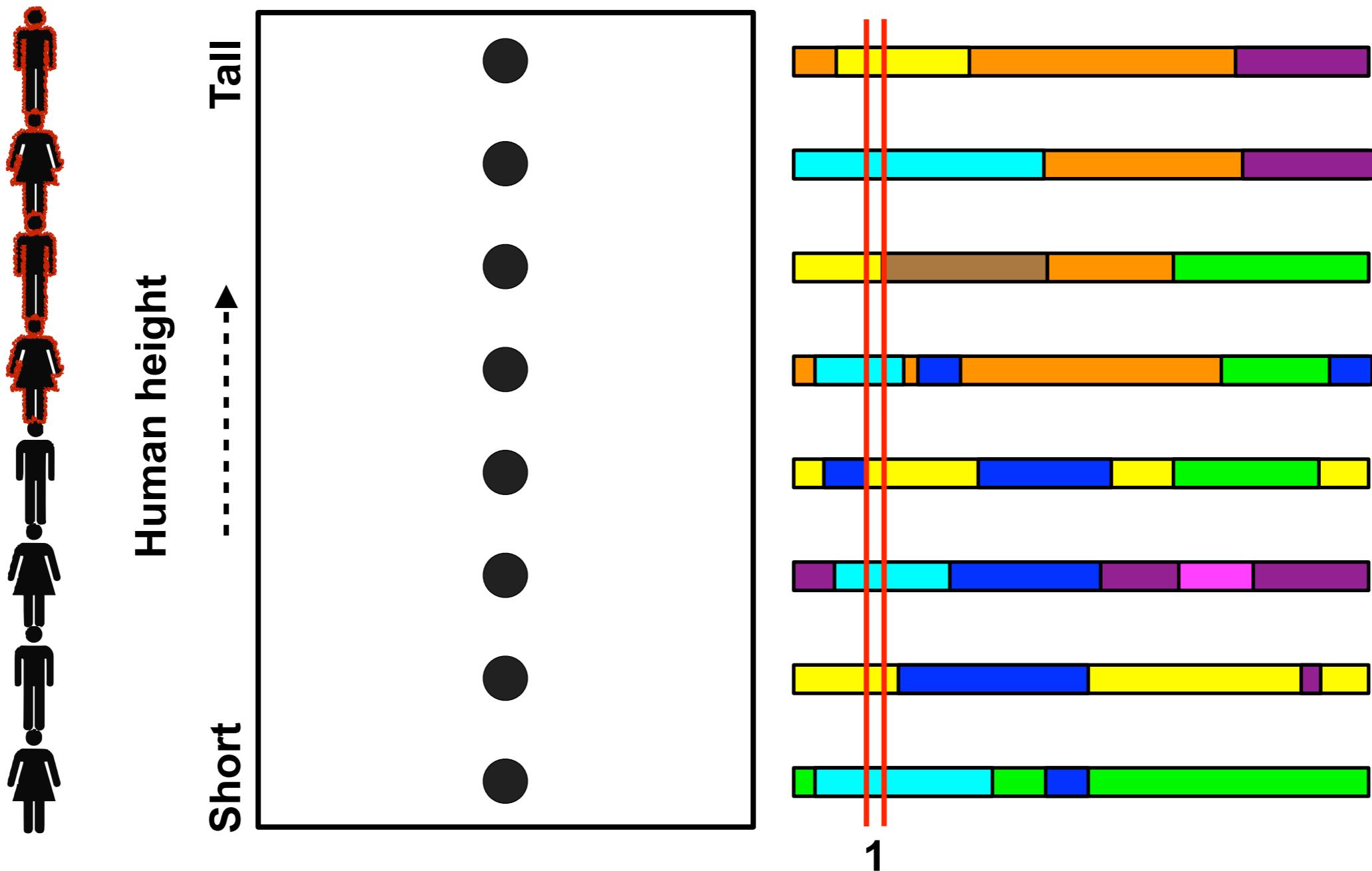
Controls  
(People without trait)

What alleles do the cases share that the controls lack?

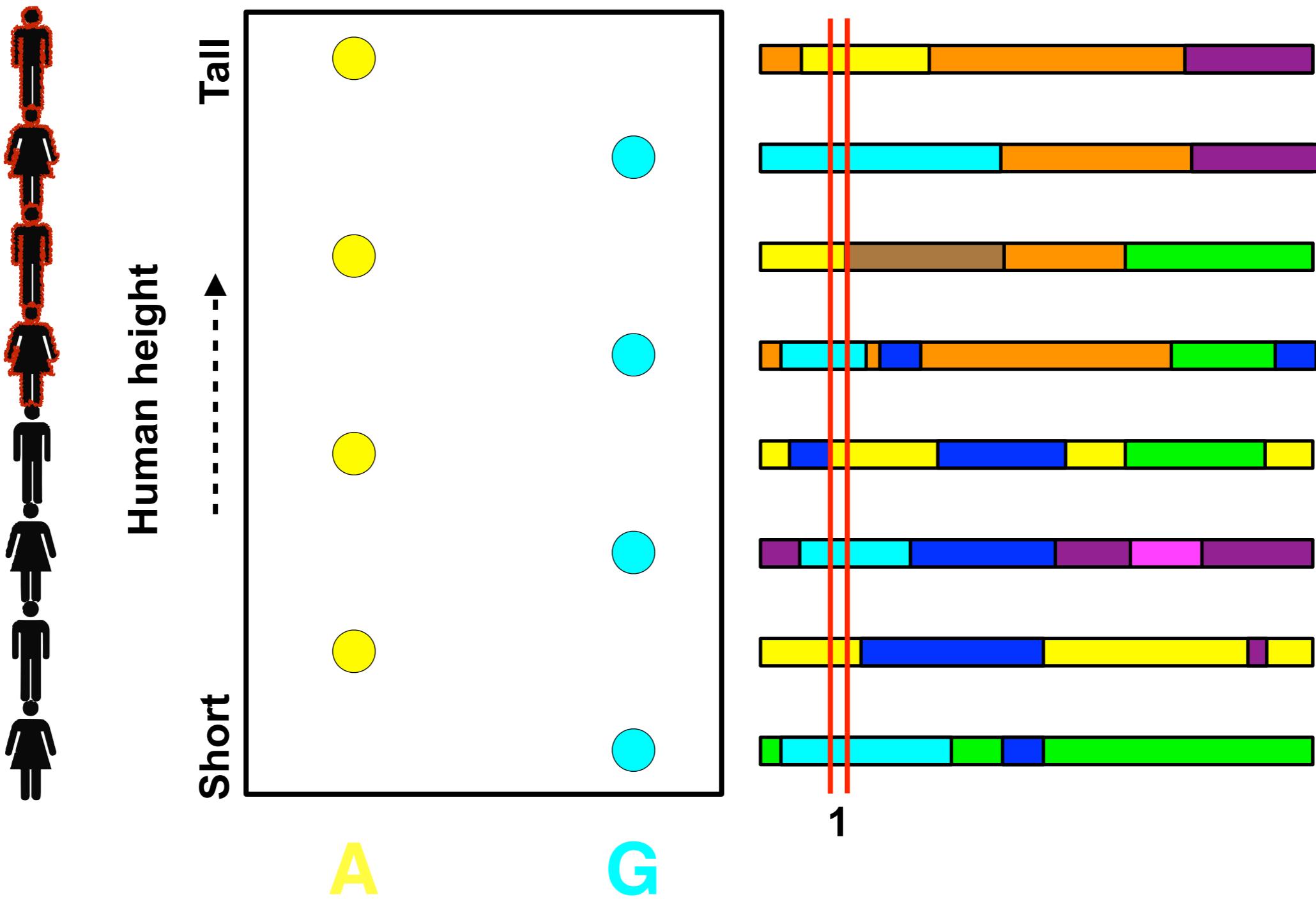
# Association mapping: Correlating genotype with phenotype



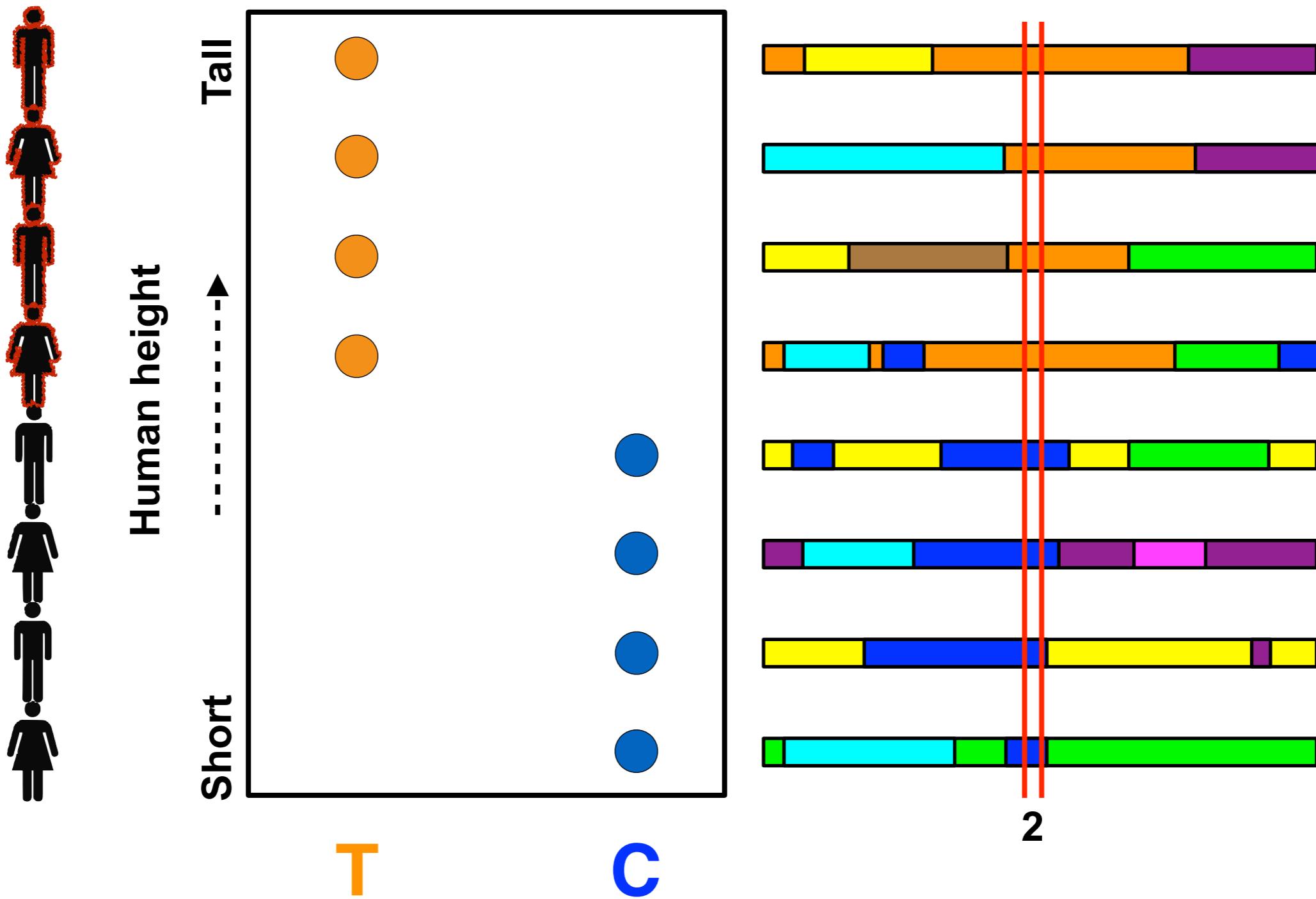
# Association mapping: Correlating genotype with phenotype



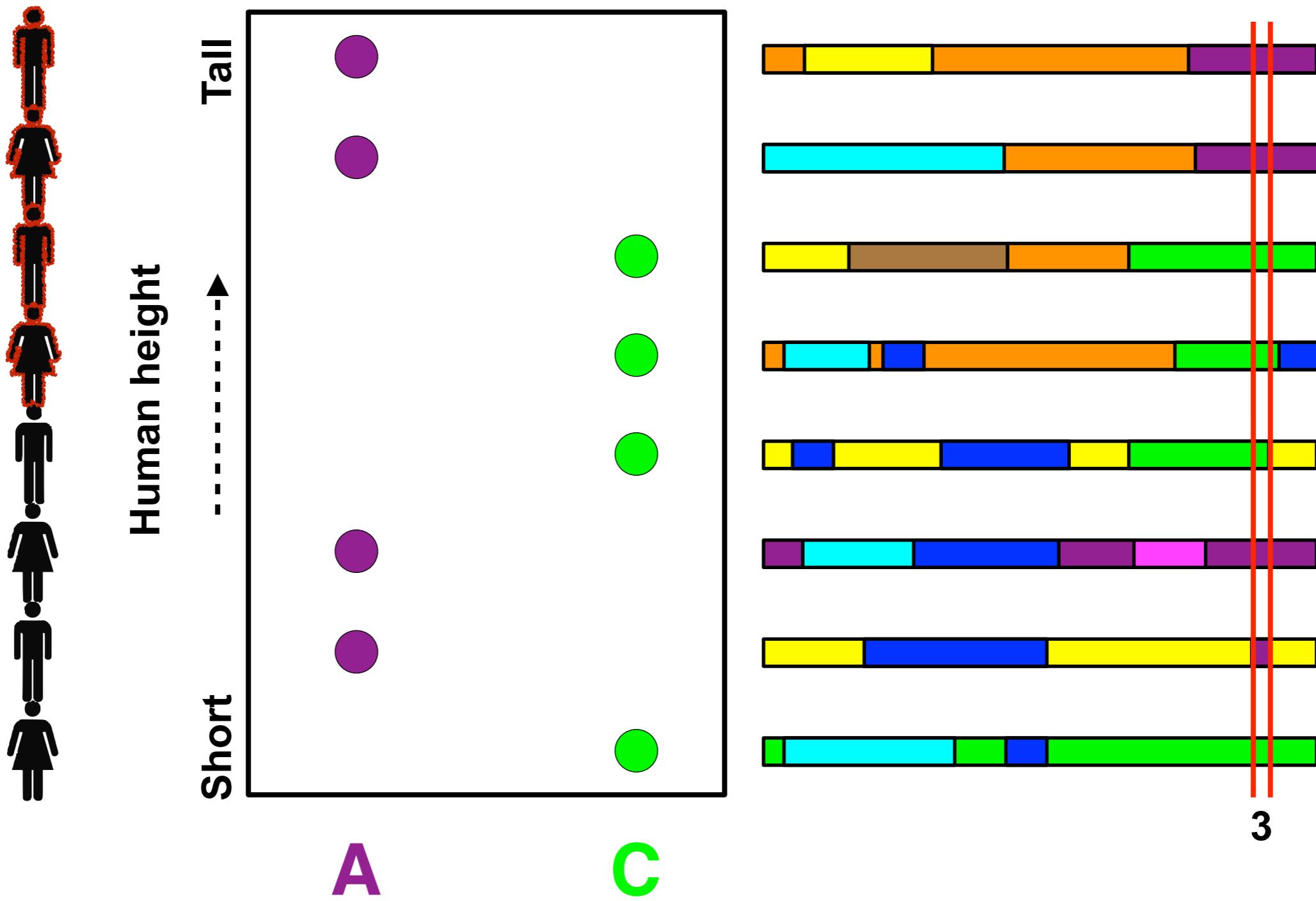
# Association mapping: Correlating genotype with phenotype



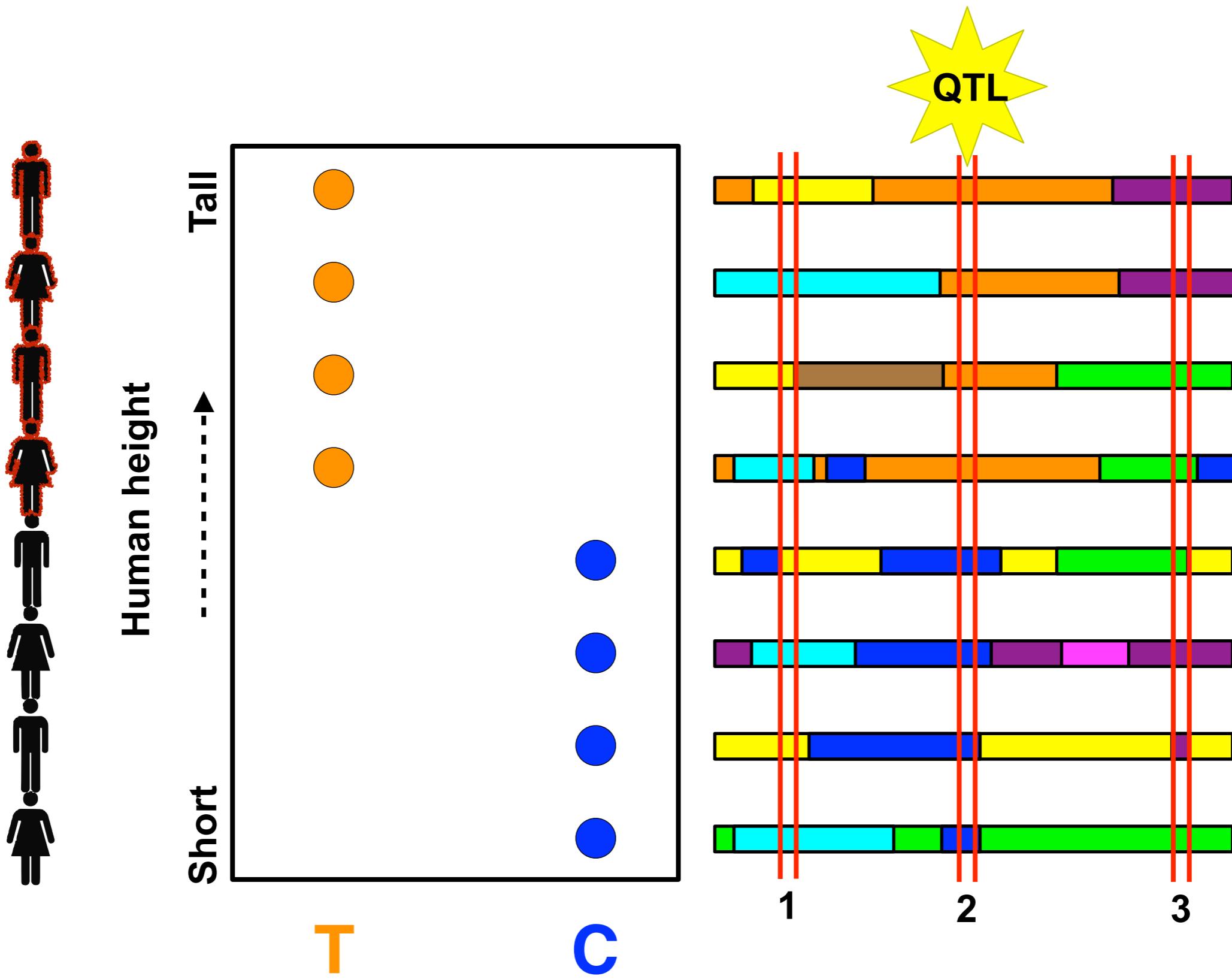
# Association mapping: Correlating genotype with phenotype



# Association mapping: Correlating genotype with phenotype

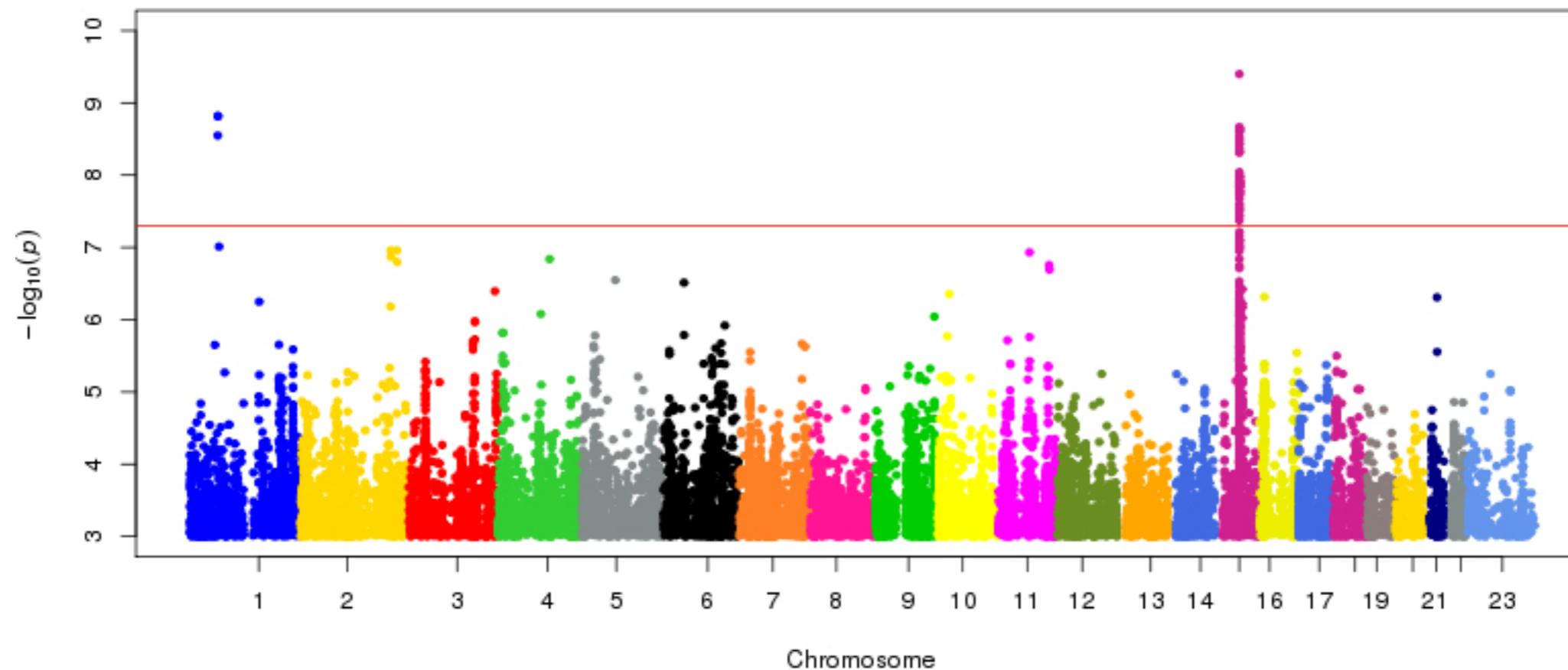


# Association mapping: Correlating genotype with phenotype



QTL=Quantitative Trait Locus

# An example Manhattan plot of GWA mapping results

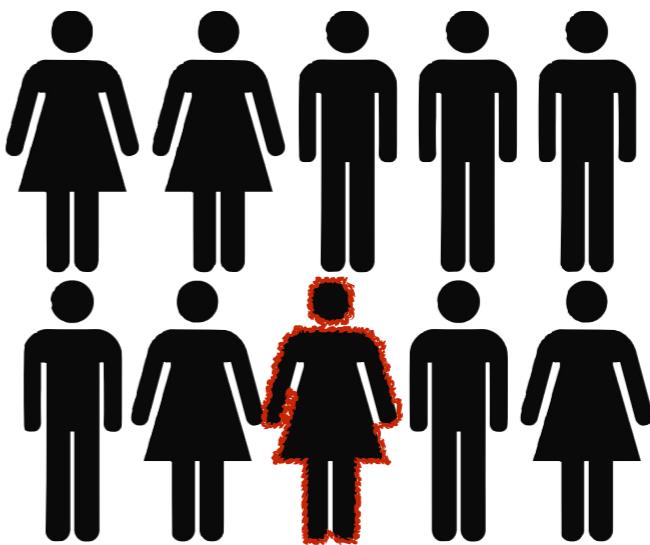


Styrkarsdottir *et al.* Nature 2014

# GWAS calculation



4000 Cases



6000 Controls

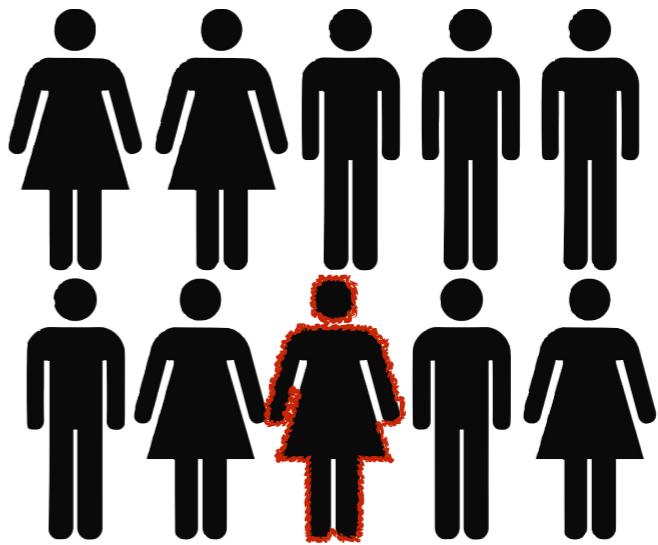
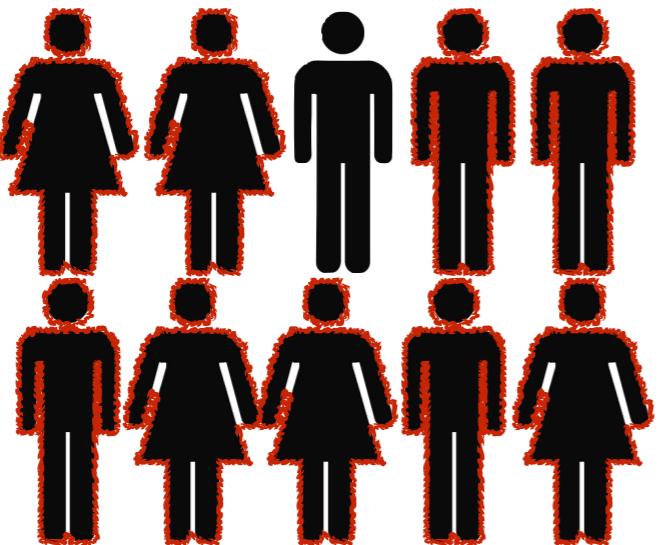
SNP1  
(G or A) 2000 of 4000 (50%)

2500 of 6000 (42%)

	Cases	Controls
G	2000	2500
A	2000	3500

Observed                      Expected

# GWAS calculation



SNP1  
(G or A) 2000 of 4000 (50%)

2500 of 6000 (42%)

	Cases	Controls
G	2000	2500
A	2000	3500

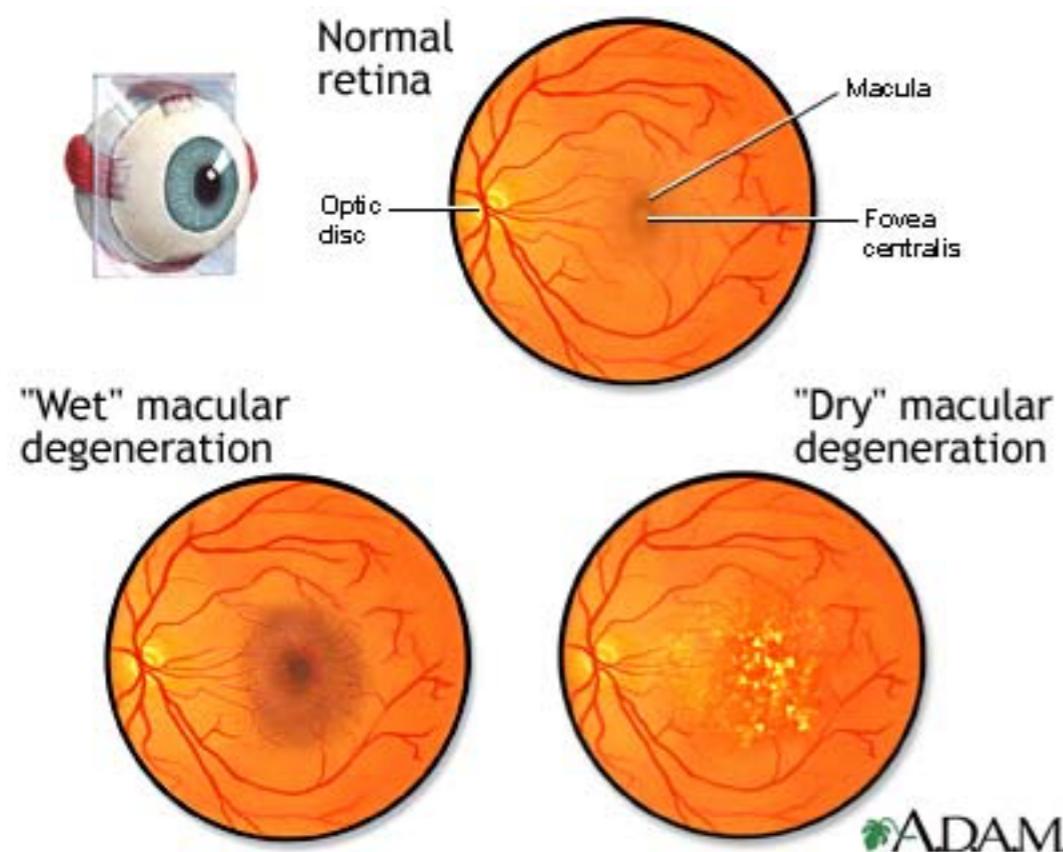
Observed

Expected

Pearson's chi-squared test  
with one degree of freedom

67.0038 or p-value of 2.71e-16

# Age-related macular degeneration: first (and best) GWAS



- 30-50 million people globally
- Age-related loss of vision
- Accumulation of extracellular material on the retina



# Age-related macular degeneration: first (and best) GWAS



96 Cases



50 Controls

Klein *et al.* Science 2005

103,611 SNPs genotyped (Bonferroni p-value  $0.05/103,611 = 4.8\text{E-}7$ )

# Age-related macular degeneration: first (and best) GWAS



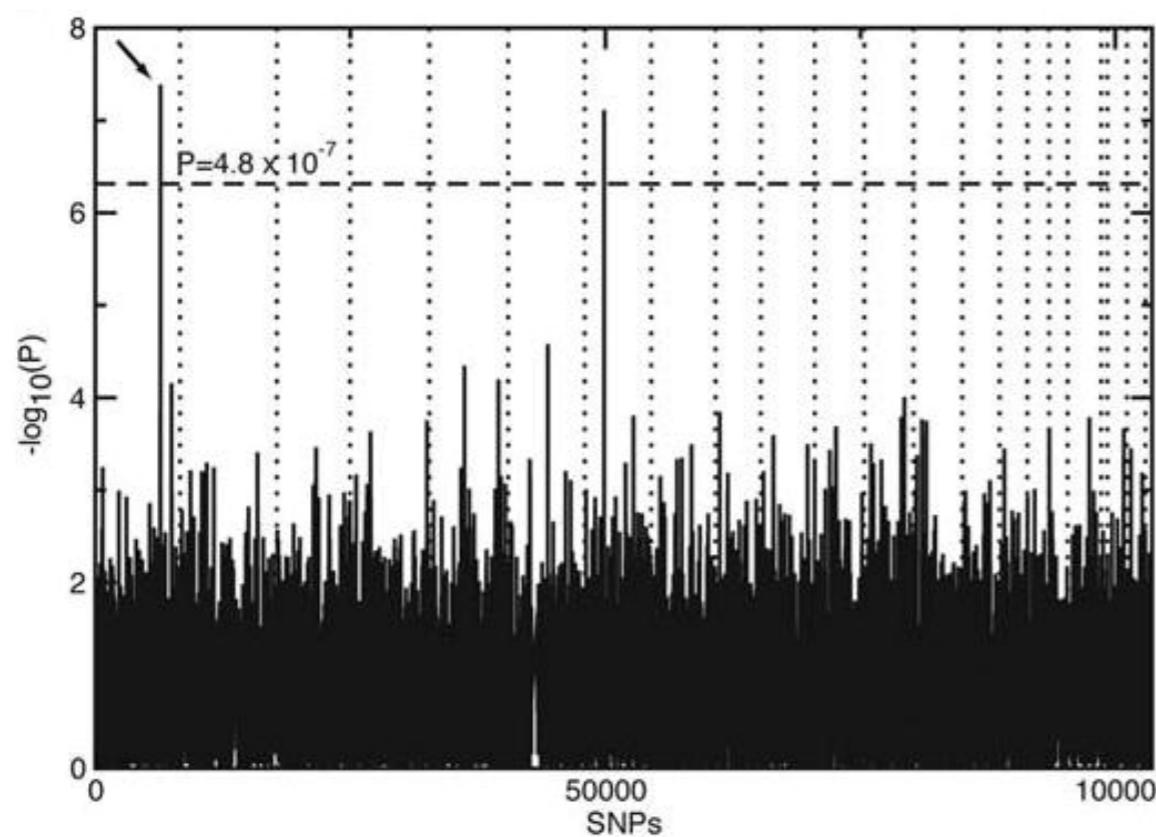
96 Cases



50 Controls

Klein *et al.* Science 2005

103,611 SNPs genotyped (Bonferroni p-value  $0.05/103,611 = 4.8E-7$ )



- Complement Factor H
- Haploinsufficient gene
- Inhibitor of inflammatory cascade
- Six family-based linkage studies found same region of genome

# Age-related macular degeneration: first (and best) GWAS



96 Cases



50 Controls

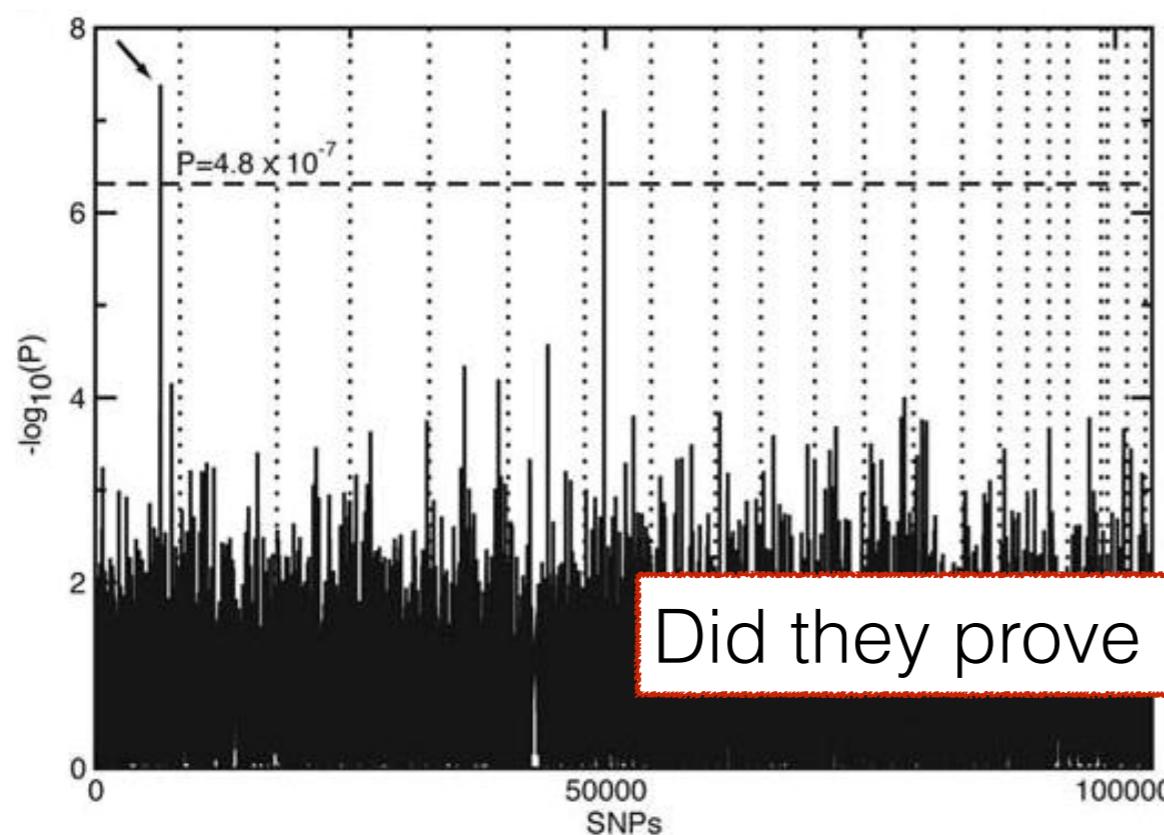
Klein *et al.* Science 2005

103,611 S

What does this success tell us

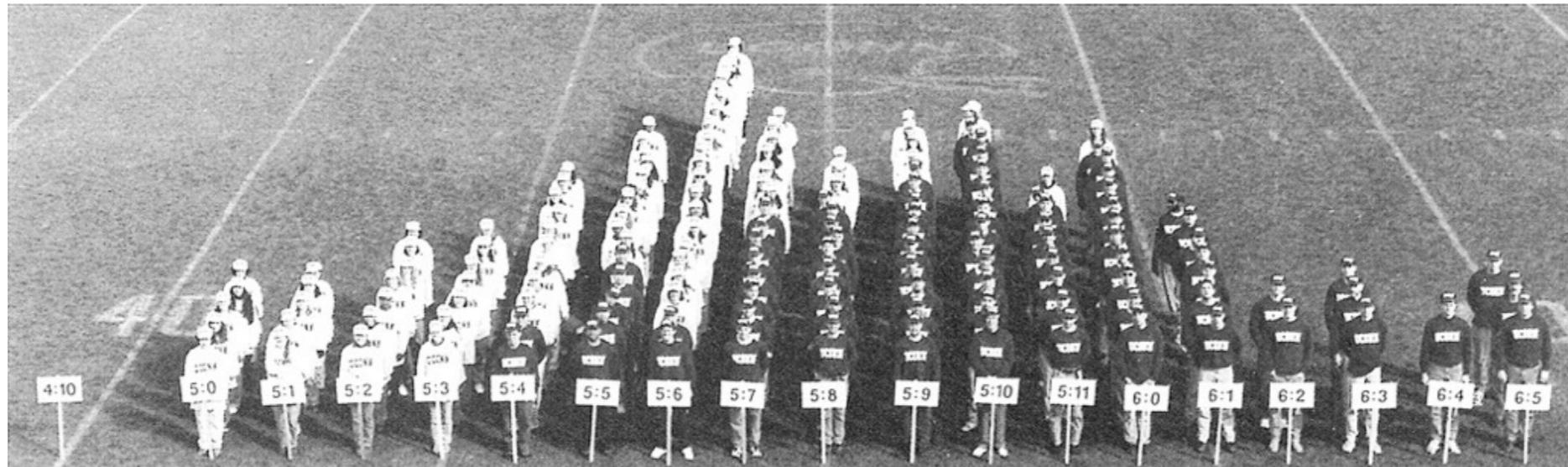
= 4.8E-7)

about the genetic underpinnings of AMD?



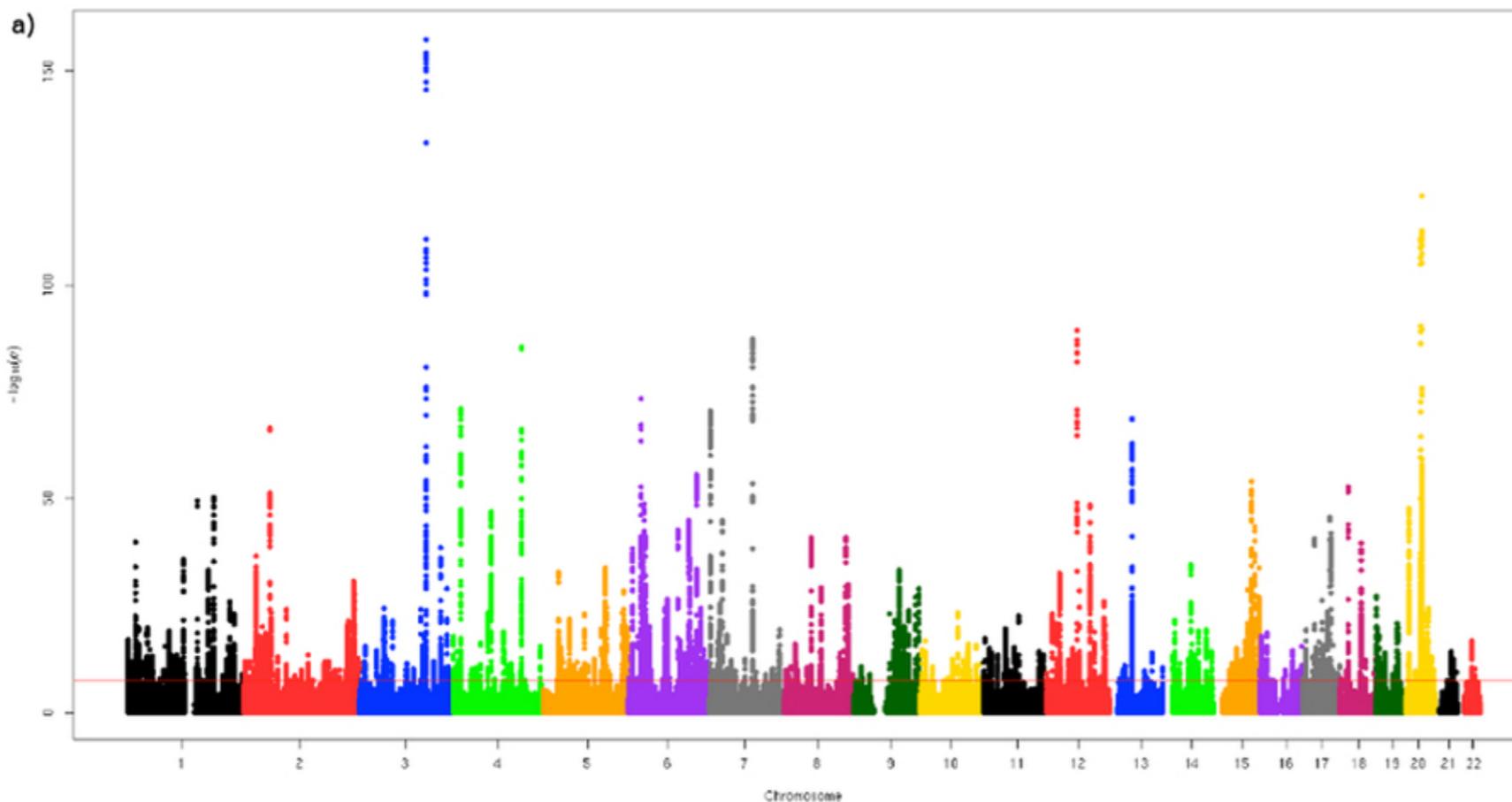
- Complement Factor H
- Haploinsufficient gene
- Inhibitor of inflammatory cascade
- Six family-based linkage studies found same region of genome

# Human height: the most powerful GWAS



University of Connecticut, 1997

253,288 individuals genotyped and phenotyped



- 697 loci reach significance
- Enriched for “growth” genes
- Each individually explains very little variation

# Lessons from the GWAS era

- Many traits are polygenic
- Effect sizes of common variants are very small
- Many associated SNPs are near genes
- Most functional variants *might* affect gene expression as opposed to protein function

**Do we have predictive ability?**

# What is your “risk” of having the trait given your genotype?

## Genotype relative risk (GRR)

For a diallelic locus (A or a), we have three genotypes: AA, Aa, aa

Choose one genotype as the reference (aa), and express GRR for the other two genotypes

$$GRR_{AA} = \frac{\text{Risk for AA genotype}}{\text{Risk for aa genotype}}$$

$$GRR_{Aa} = \frac{\text{Risk for Aa genotype}}{\text{Risk for aa genotype}}$$

Risks are estimated as odds ratios

Genotype	Cases	Controls	Case:Control Ratio
AA	D <sub>AA</sub>	H <sub>AA</sub>	D <sub>AA</sub> / H <sub>AA</sub>
Aa	D <sub>Aa</sub>	H <sub>Aa</sub>	D <sub>Aa</sub> / H <sub>Aa</sub>
aa	D <sub>aa</sub>	H <sub>aa</sub>	D <sub>aa</sub> / H <sub>aa</sub>

Case:control ratios are equivalent to the odds of disease given genotype in the population

Ratios of case:control ratios estimate relative risks in a population

# What is your “risk” of having the trait given your genotype?

## Genotype relative risk (GRR)

Risks are estimated as odds ratios

Genotype	Cases	Controls	Case:Control Ratio
AA	$D_{AA}$	$H_{AA}$	$D_{AA} / H_{AA}$
Aa	$D_{Aa}$	$H_{Aa}$	$D_{Aa} / H_{Aa}$
aa	$D_{aa}$	$H_{aa}$	$D_{aa} / H_{aa}$

Case:control ratios are equivalent to the odds of disease given genotype in the population

Ratios of case:control ratios estimate relative risks in a population

$$GRR_{AA} = \frac{\text{Risk for AA genotype}}{\text{Risk for aa genotype}}$$

$$GRR_{AA} = \frac{D_{AA} / H_{AA}}{D_{aa} / H_{aa}}$$

$$GRR_{Aa} = \frac{\text{Risk for Aa genotype}}{\text{Risk for aa genotype}}$$

$$GRR_{Aa} = \frac{D_{Aa} / H_{Aa}}{D_{aa} / H_{aa}}$$

# Genotype relative risk (GRR) - example

Risks are estimated as odds ratios

Genotype	Cases	Controls	Case:Control Ratio
AA	400	250	400 / 250
Aa	350	250	350 / 250
aa	400	300	400 / 300

$$GRR_{AA} = \frac{\text{Risk for AA genotype}}{\text{Risk for aa genotype}}$$

$$GRR_{AA} = \frac{400 / 250}{400 / 300} = 1.2$$

$$GRR_{Aa} = \frac{\text{Risk for Aa genotype}}{\text{Risk for aa genotype}}$$

$$GRR_{Aa} = \frac{350 / 250}{400 / 300} = 1.05$$

The AA genotype is 1.2x more likely than the aa genotype to have the disease

The Aa genotype is 1.05x more likely than the aa genotype to have the disease

# My genotype relative risks (GRR) for some traits

SHOW RESULTS FOR Erik Andersen

[SEE NEW AND RECENTLY UPDATED REPORTS »](#)

These reports provide information about your possible risk for developing certain health conditions based on genetics. Environmental and lifestyle factors also often play a large role in your risk for developing these conditions.

## Elevated Risk

NAME	CONFIDENCE	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Atrial Fibrillation	★★★★	33.9%	27.2%	1.25x 
Prostate Cancer ♂	★★★★	23.8%	17.8%	1.33x 
Gallstones	★★★★	11.1%	7.0%	1.58x 
Exfoliation Glaucoma	★★★★	2.2%	0.7%	2.90x 
Ulcerative Colitis	★★★★	1.00%	0.77%	1.30x 
Esophageal Squamous Cell Carcinoma (ESCC)	★★★★	0.43%	0.36%	1.21x 
Stomach Cancer (Gastric Cardia Adenocarcinoma)	★★★★	0.28%	0.23%	1.22x 
Abdominal Aortic Aneurysm	★★★			
Alopecia Areata	★★★			

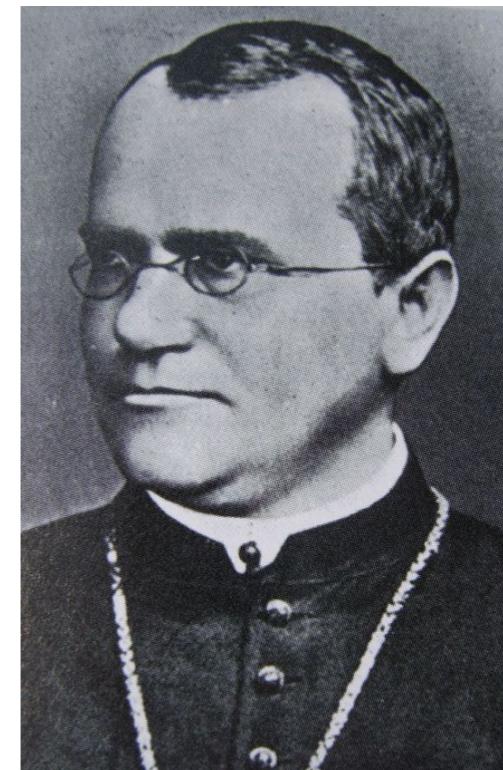
# My genotype relative risks (GRR) for some traits

## Decreased Risk ?

NAME	CONFIDENCE	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Gout	★★★★	17.1%	22.8%	0.75x 
Venous Thromboembolism	★★★★	9.0%	12.3%	0.73x 
Alzheimer's Disease	★★★★	4.3%	7.2%	0.60x 
Age-related Macular Degeneration	★★★★	3.1%	6.5%	0.48x 
Melanoma	★★★★	2.2%	2.9%	0.75x 
Rheumatoid Arthritis	★★★★	1.5%	2.4%	0.63x 
Restless Legs Syndrome	★★★★	1.5%	2.0%	0.74x 
Parkinson's Disease	★★★★	0.94%	1.61%	0.58x 
Multiple Sclerosis	★★★★	0.24%	0.34%	0.69x 
Crohn's Disease	★★★★	0.16%	0.53%	0.30x 
Type 1 Diabetes	★★★★	0.11%	1.02%	0.10x 
Celiac Disease	★★★★	0.06%	0.12%	0.54x 
Primary Biliary Cirrhosis	★★★★	0.04%	0.08%	0.48x 
Atopic Dermatitis	★★★			
Basal Cell Carcinoma	★★★			

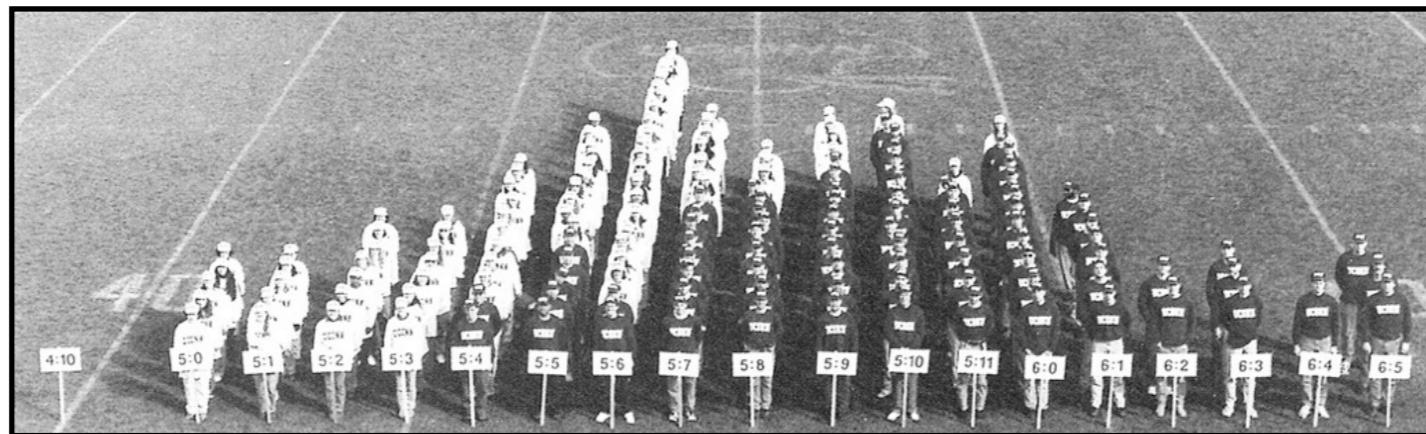
# Most well understood traits are determined by single genes with large phenotypic effects

Round vs. Wrinkled



# Complex traits are controlled by many genes and interactions with the environment

Height



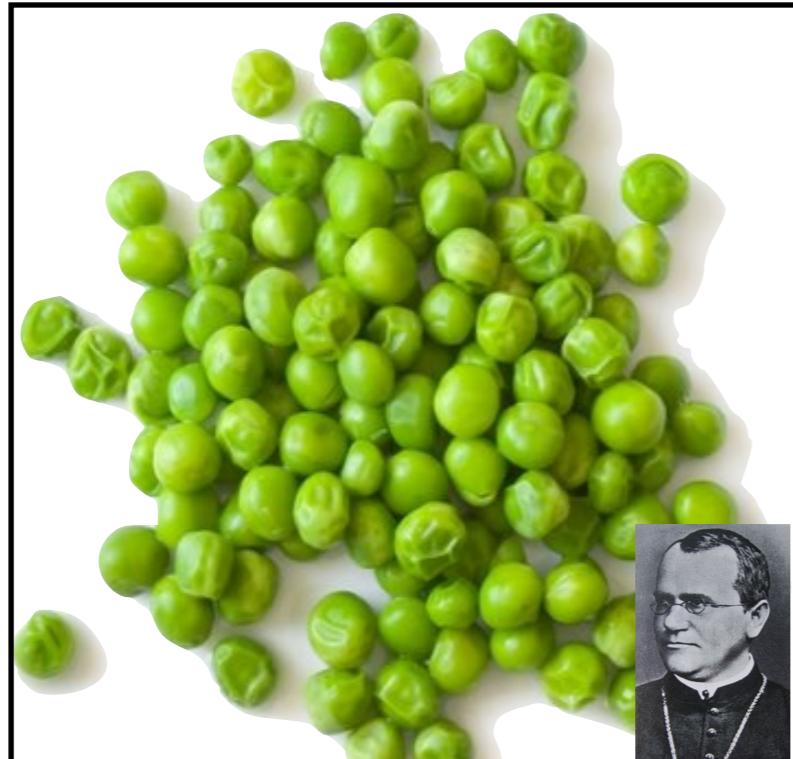
697 loci

(Wood *et al.* 2014)

20% phenotypic variance explained

# Complex traits are controlled by many genes and interactions with the environment

## Round vs. Wrinkled

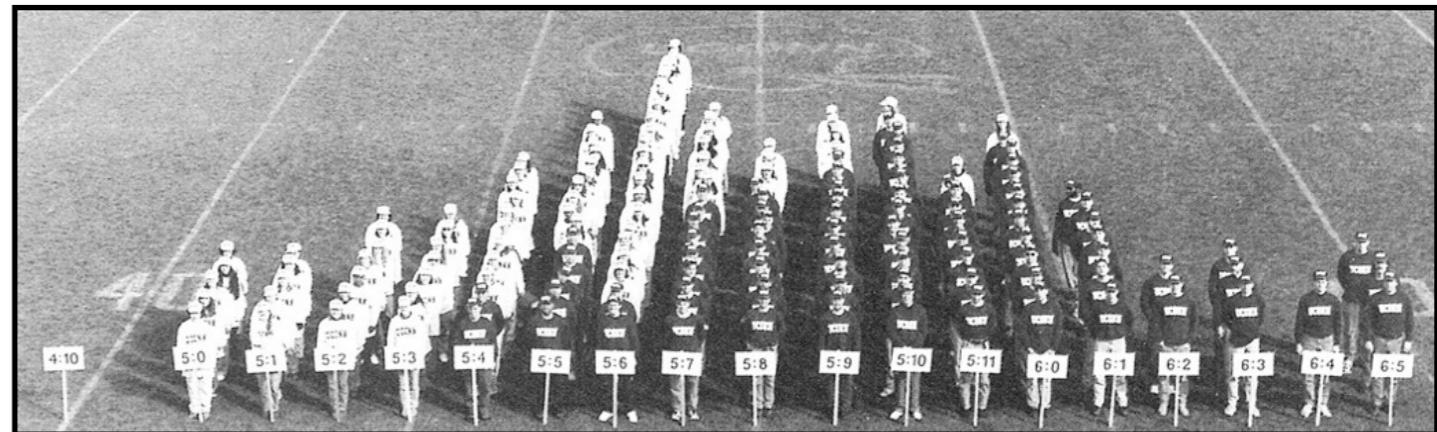


Starch branching enzyme 1  
(Bhattacharyya *et al.* 1990)

100% phenotypic variance explained

binary traits

## Height



697 loci

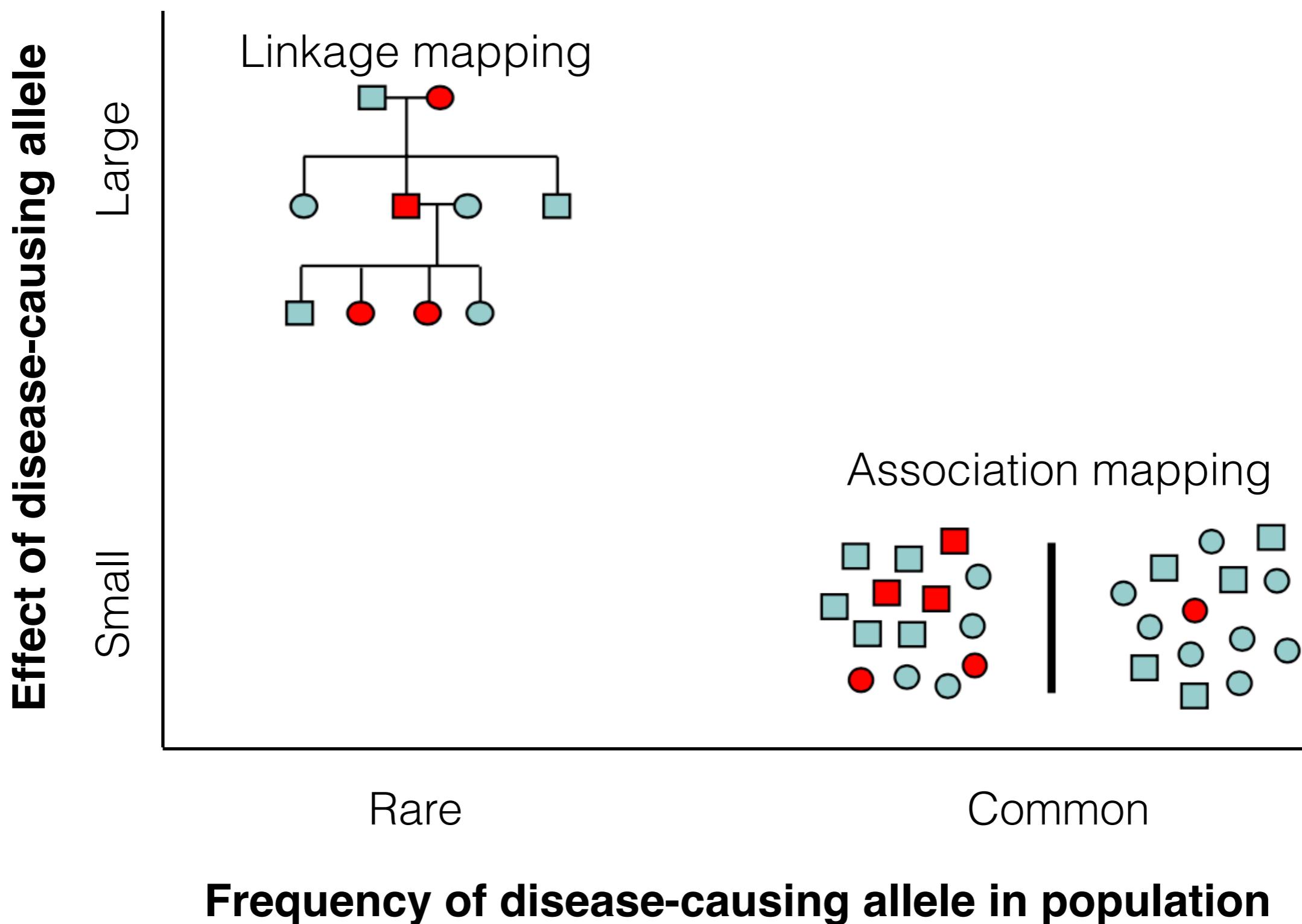
(Wood *et al.* 2014)



20% phenotypic variance explained

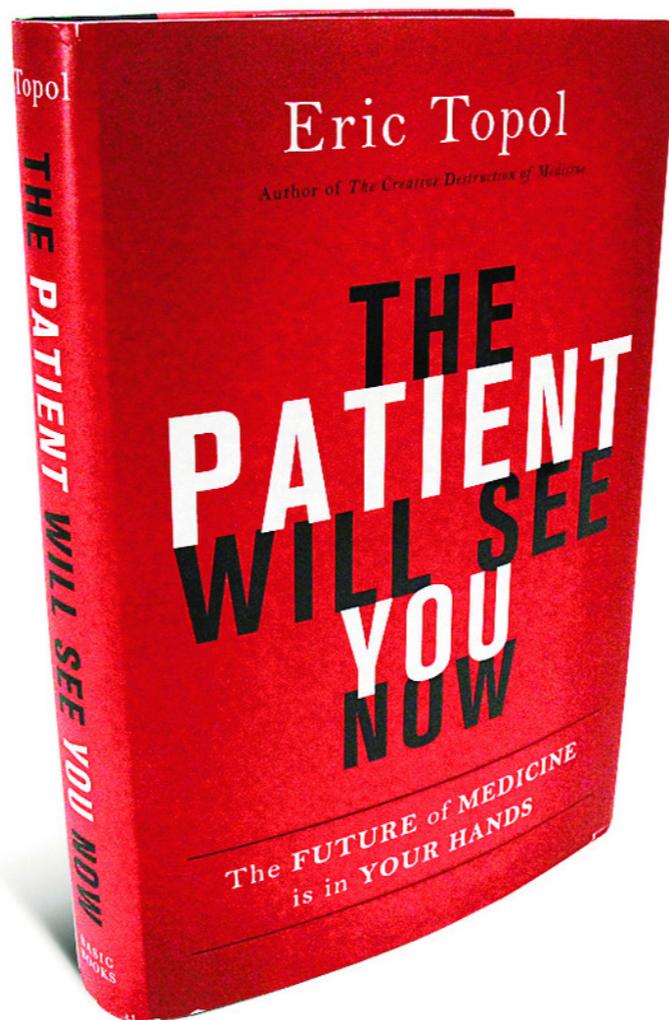
continuous traits

# Complex traits can be mapped using both techniques



# Present and future of genetic medicine (positives)

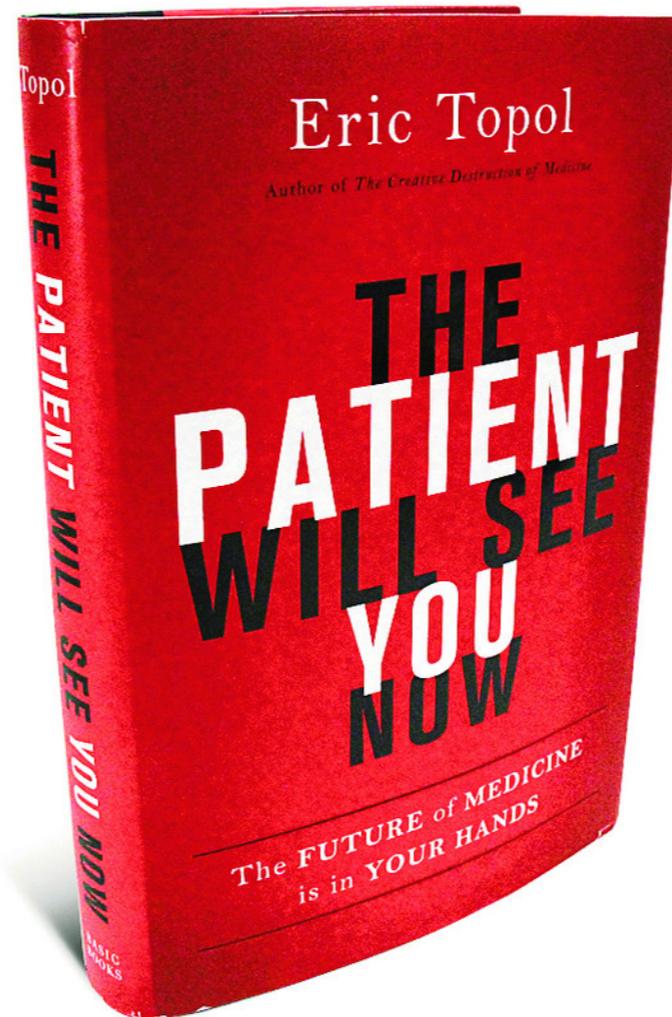
Sequencing  
will be  
cheap,  
accessible,  
and  
standard practice.



With sequencing,  
do we still need  
genetics?

- Rare disease causal genes found
- 12 drugs approved by FDA with genetic test since 2012
- More than 120 drugs have genotype on label - check before use
- Fetal sequencing is safer than amniocentesis
- Infectious disease ID, sepsis

# Present and future of genetic medicine (negatives)



- Lots of taxpayers' money spent on little valuable data
- Genotype data are being sold
- Genotype data are being evaluated by insurance companies
- Most common diseases are influenced more by behavior, diet, and environment than genes

**Be skeptical!**

**Lecture 17**

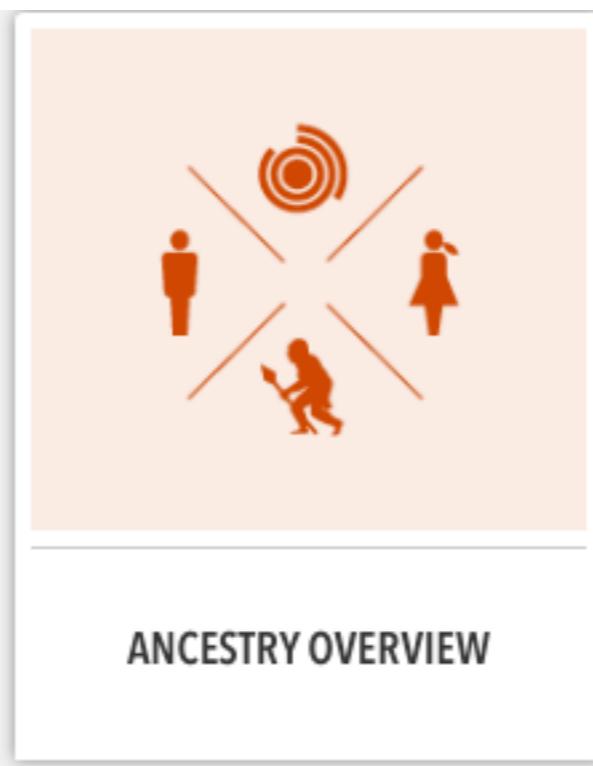
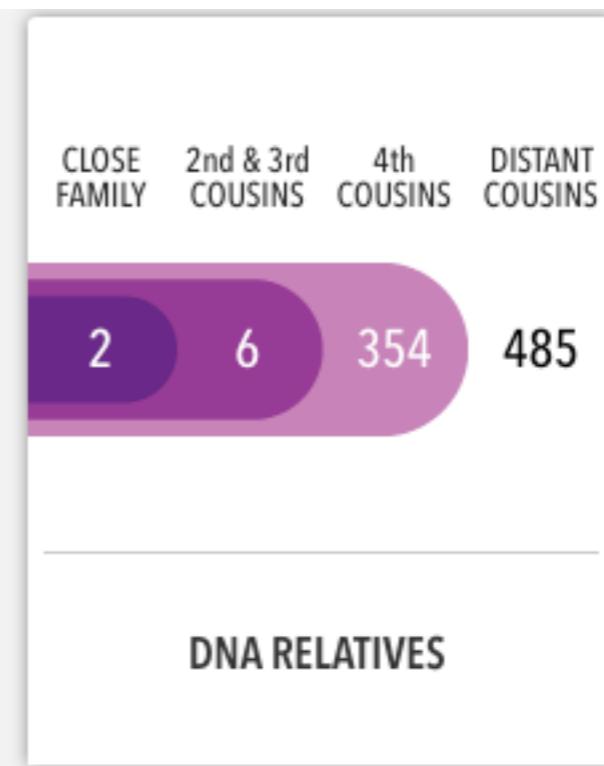
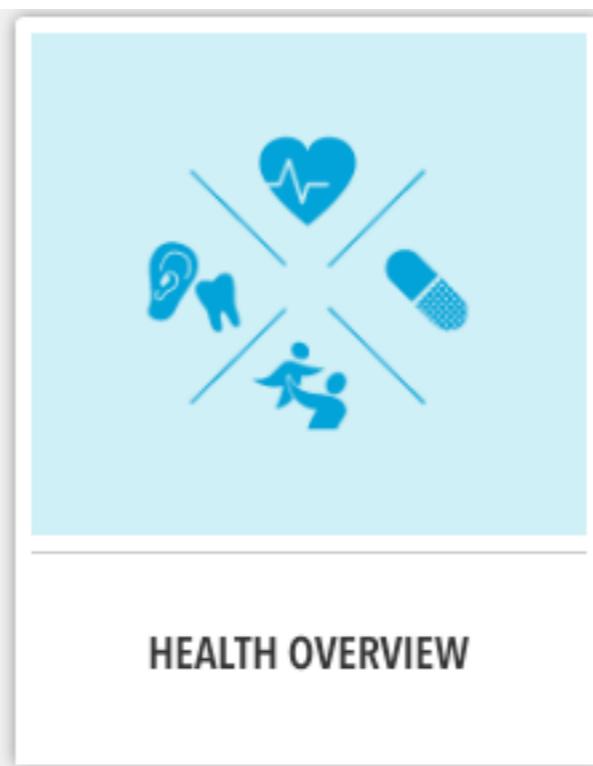
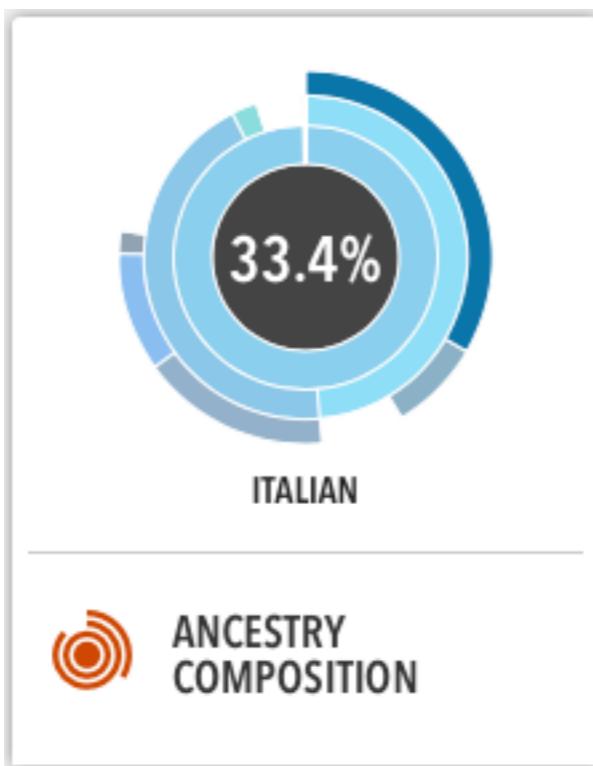


# 23andMe





# 23andMe



**A T  
C G**

**BROWSE RAW DATA**