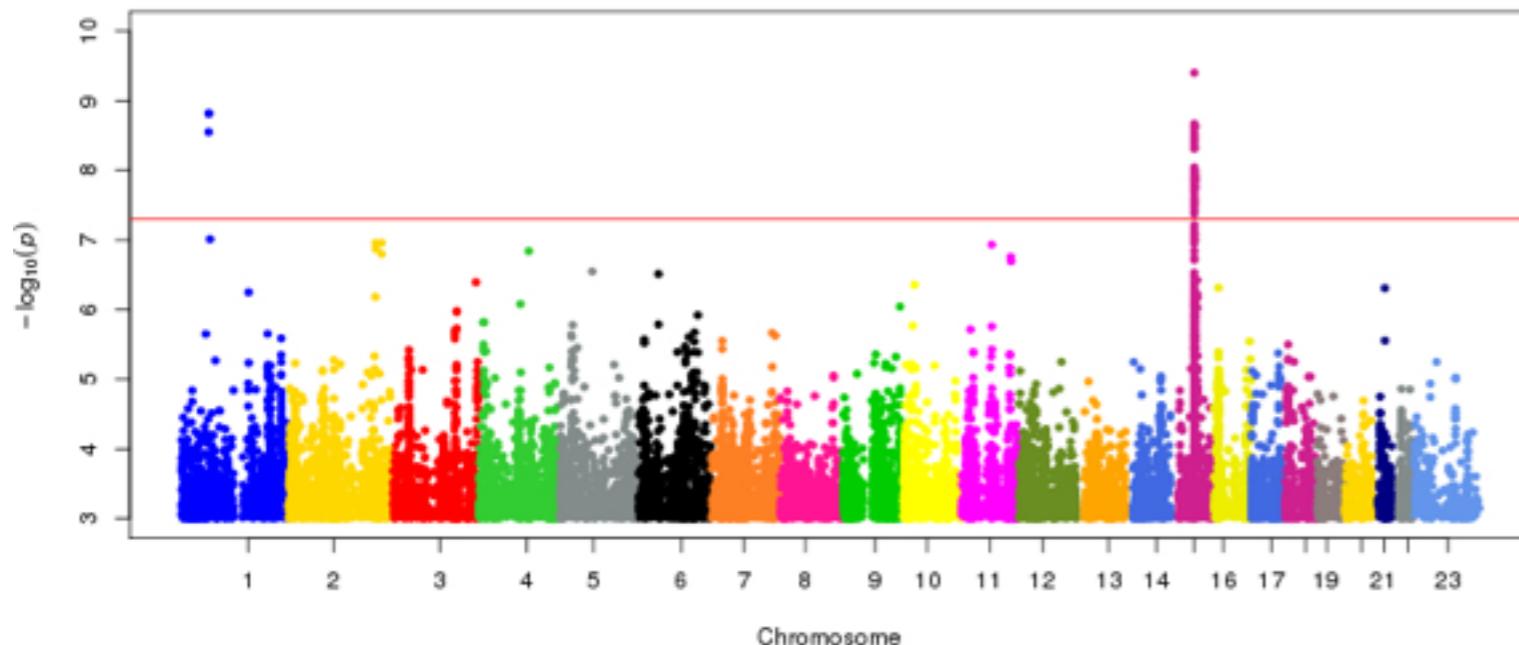


Bio393: Genetic Analysis

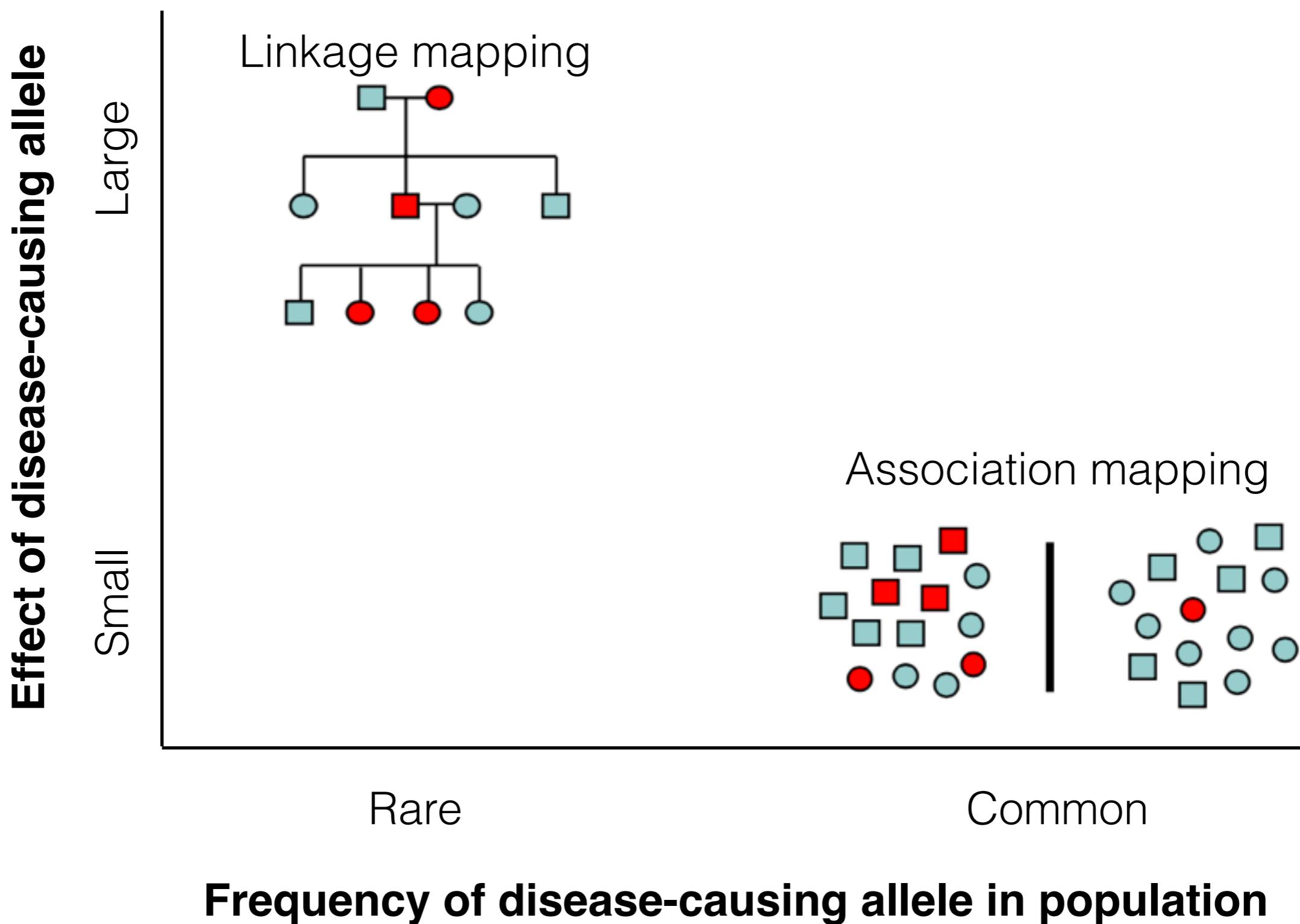
GWAS, relative risk, complex traits, and the future of genetic medicine



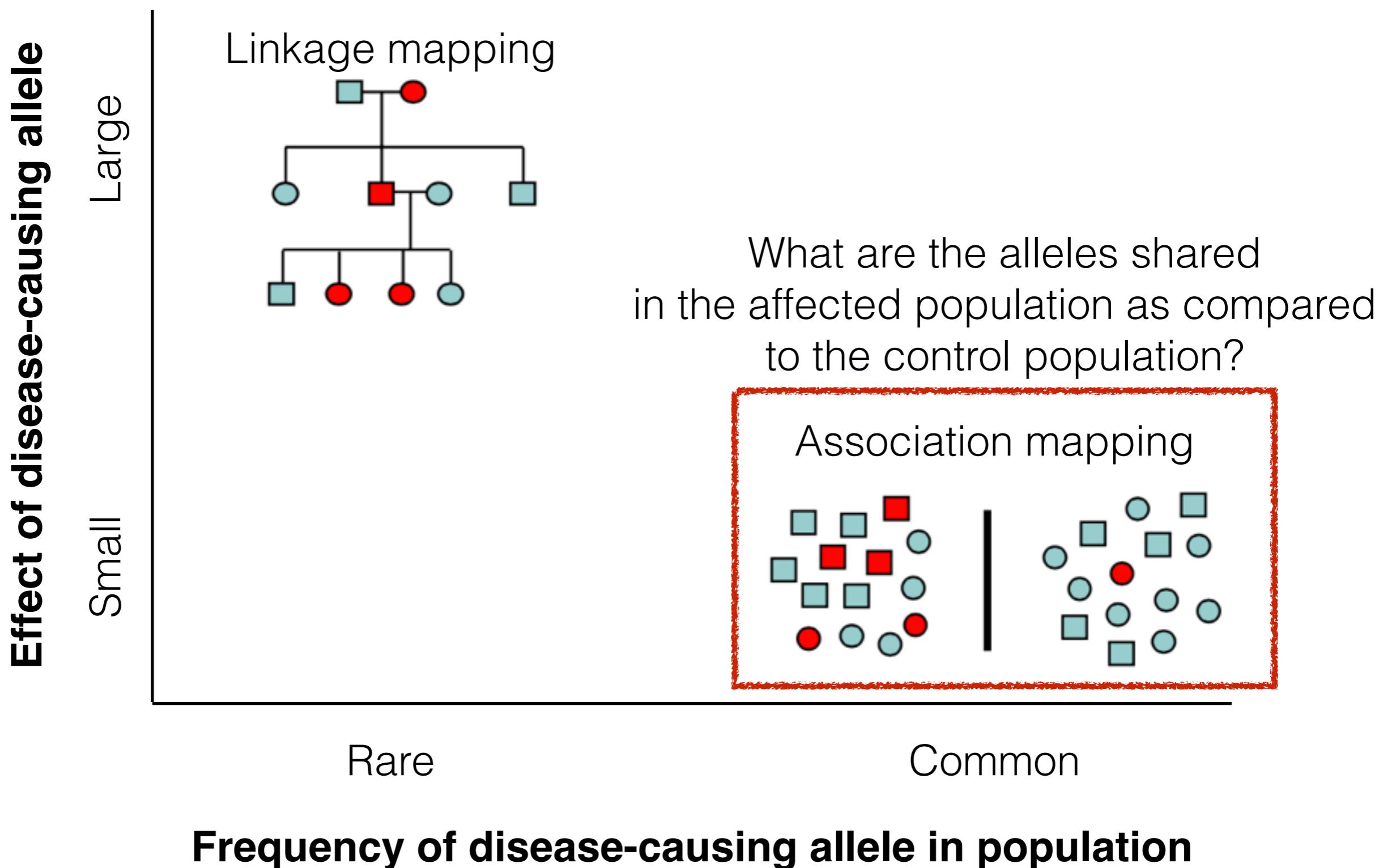
Styrkarsdottir *et al.* Nature 2014



Human gene mapping has two general flavors



Human gene mapping has two general flavors

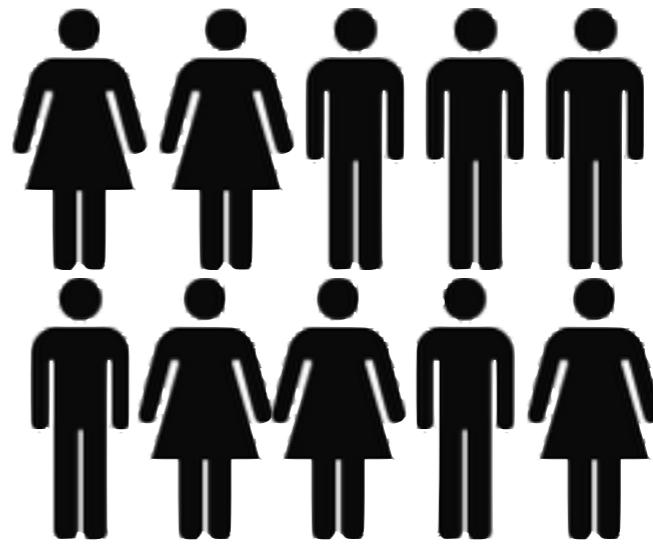


The set up of a genome-wide association (GWA) mapping

Case-control study design



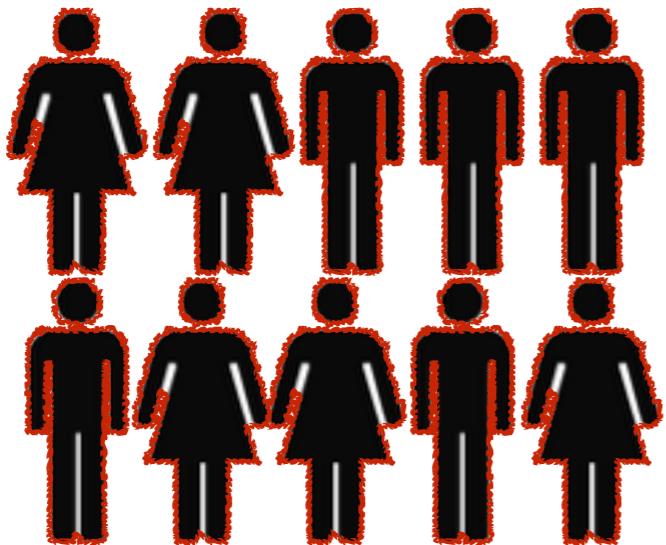
Cases
(People with trait)



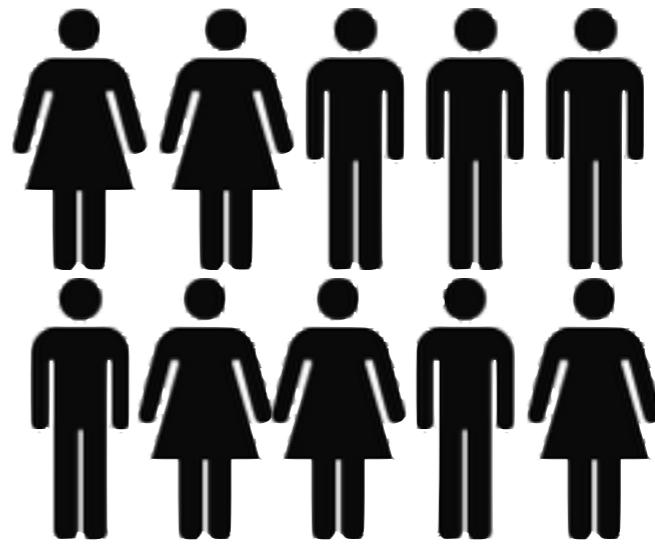
Controls
(People without trait)

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

Case-control study design



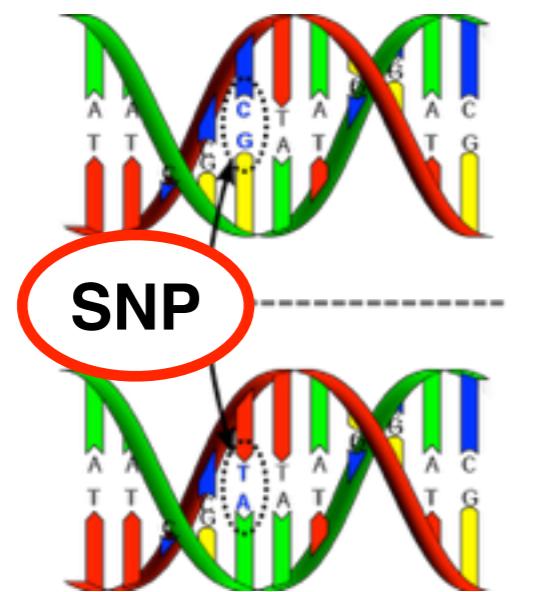
Cases
(People with trait)



Controls
(People without trait)

What alleles do the cases share that the controls lack?

Measure correlation between genetic variation and phenotypic variation in cases and compare to controls

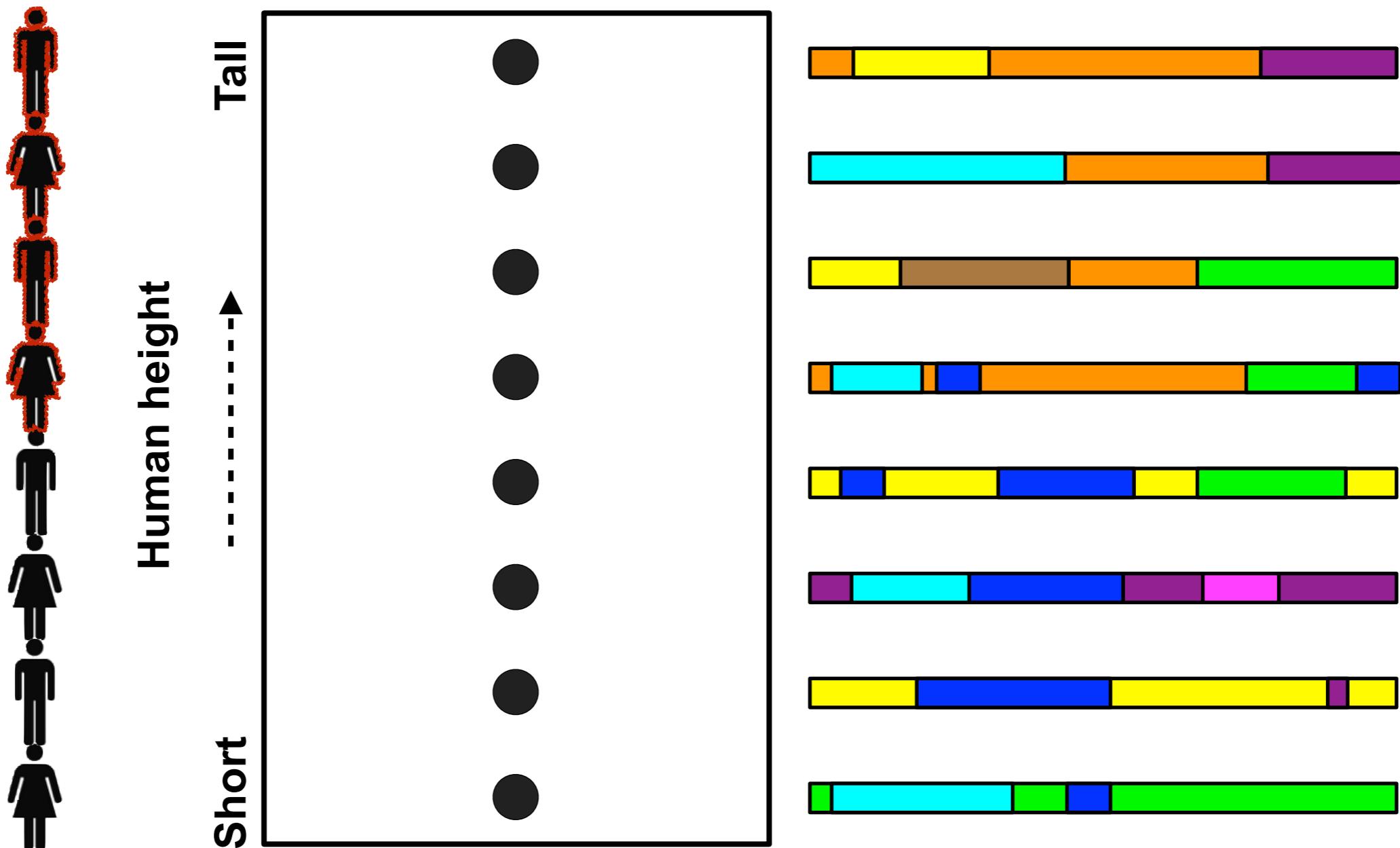


Genetic variation

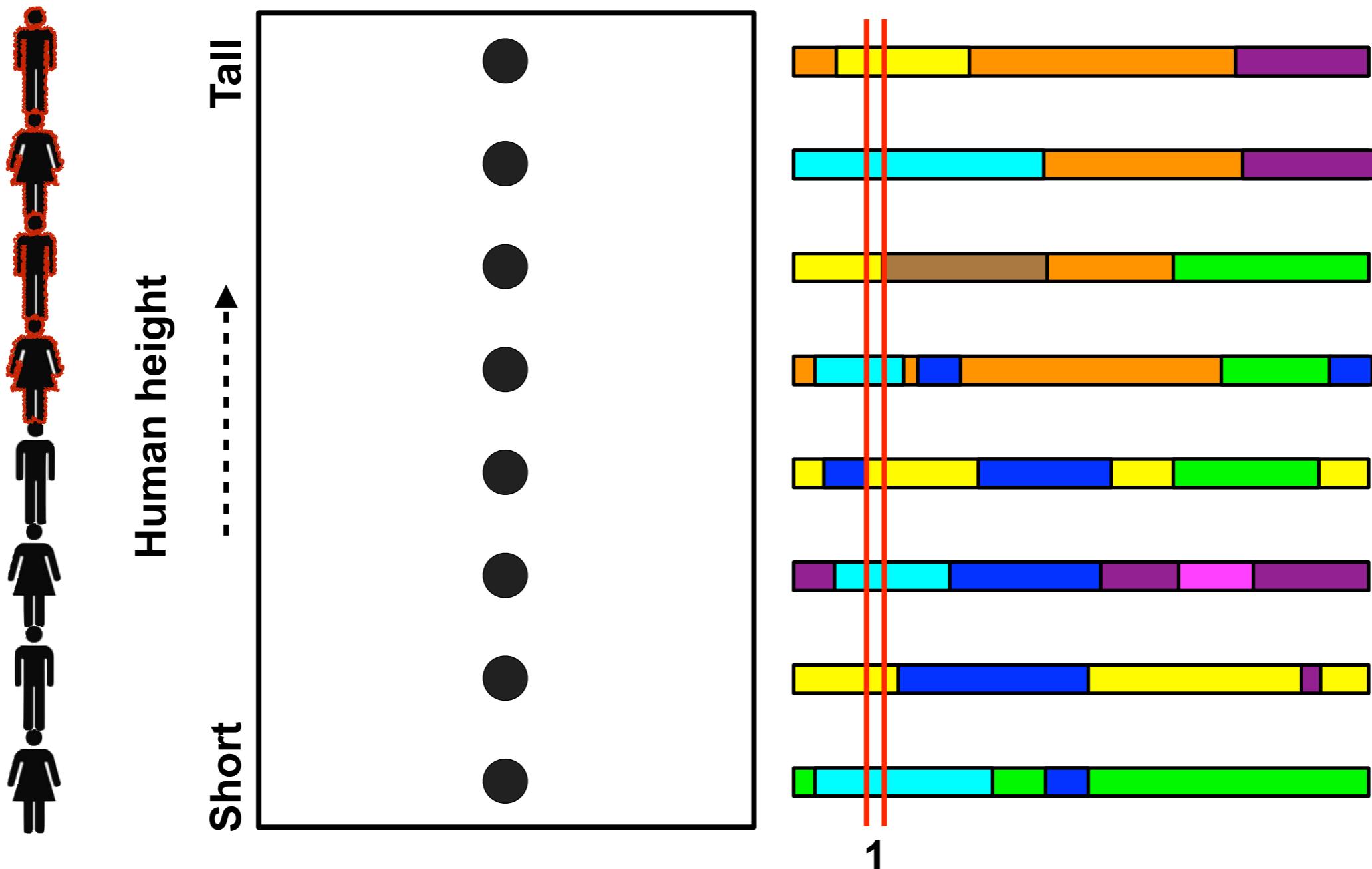


Phenotypic variation

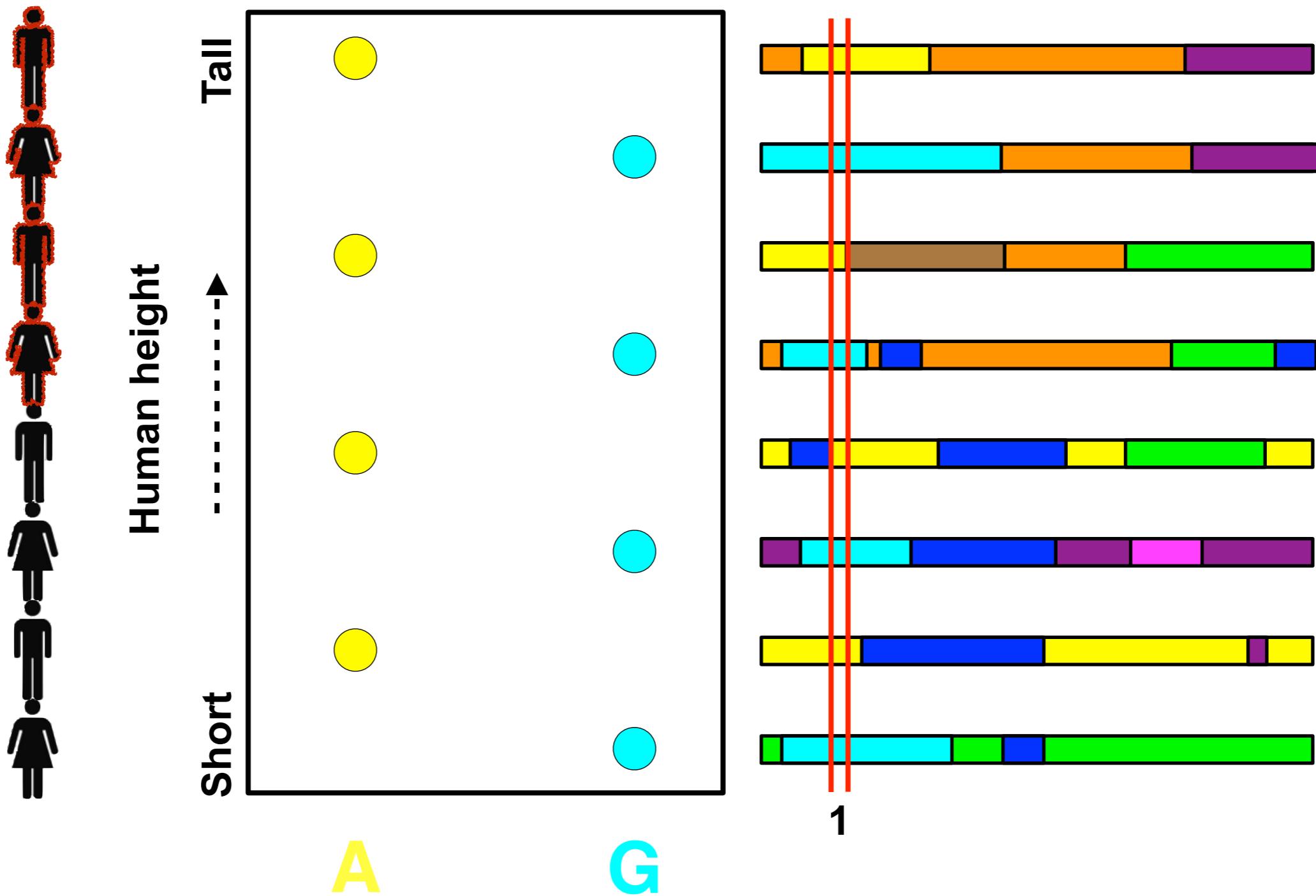
Association mapping: Correlating genotype with phenotype



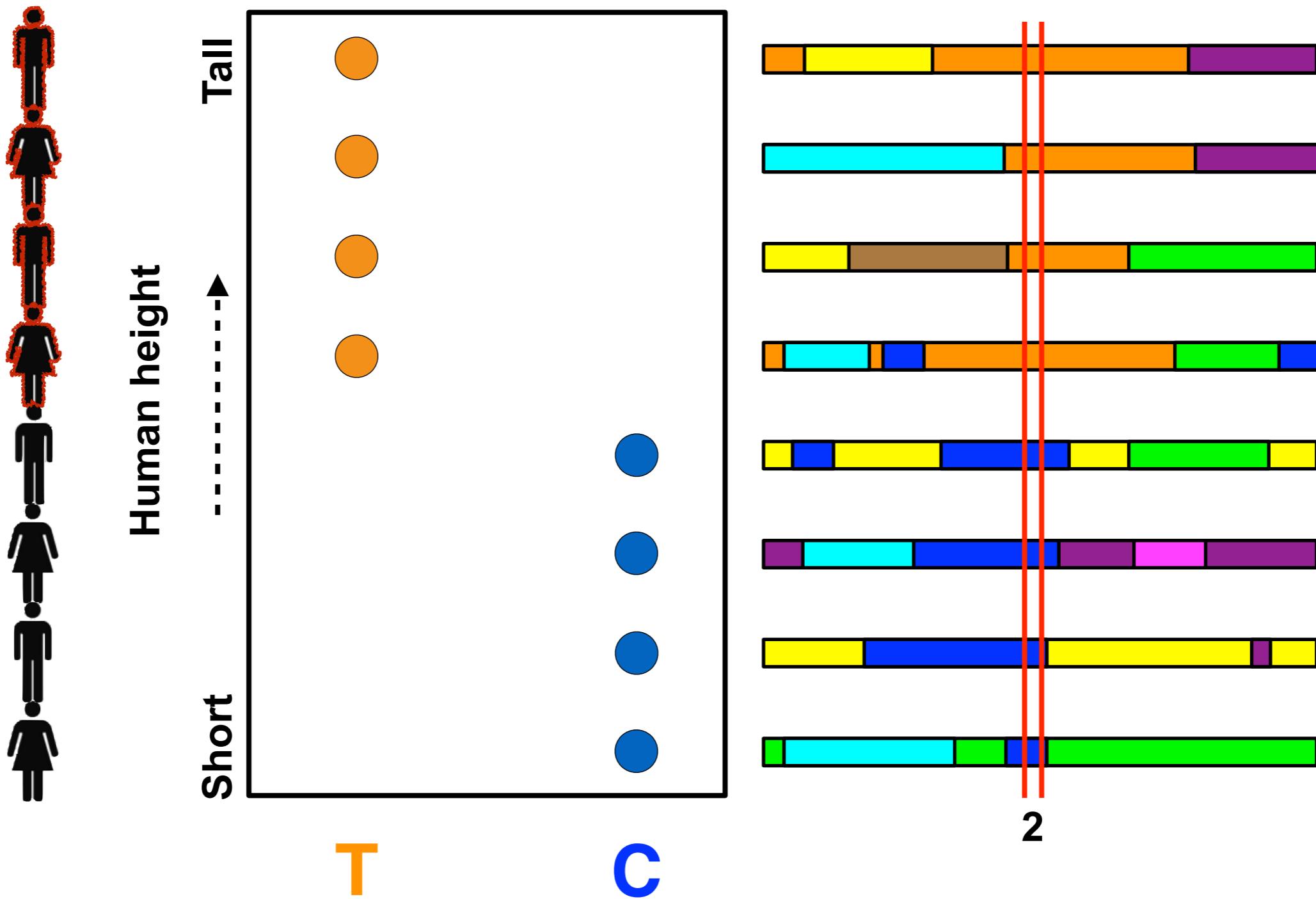
Association mapping: Correlating genotype with phenotype



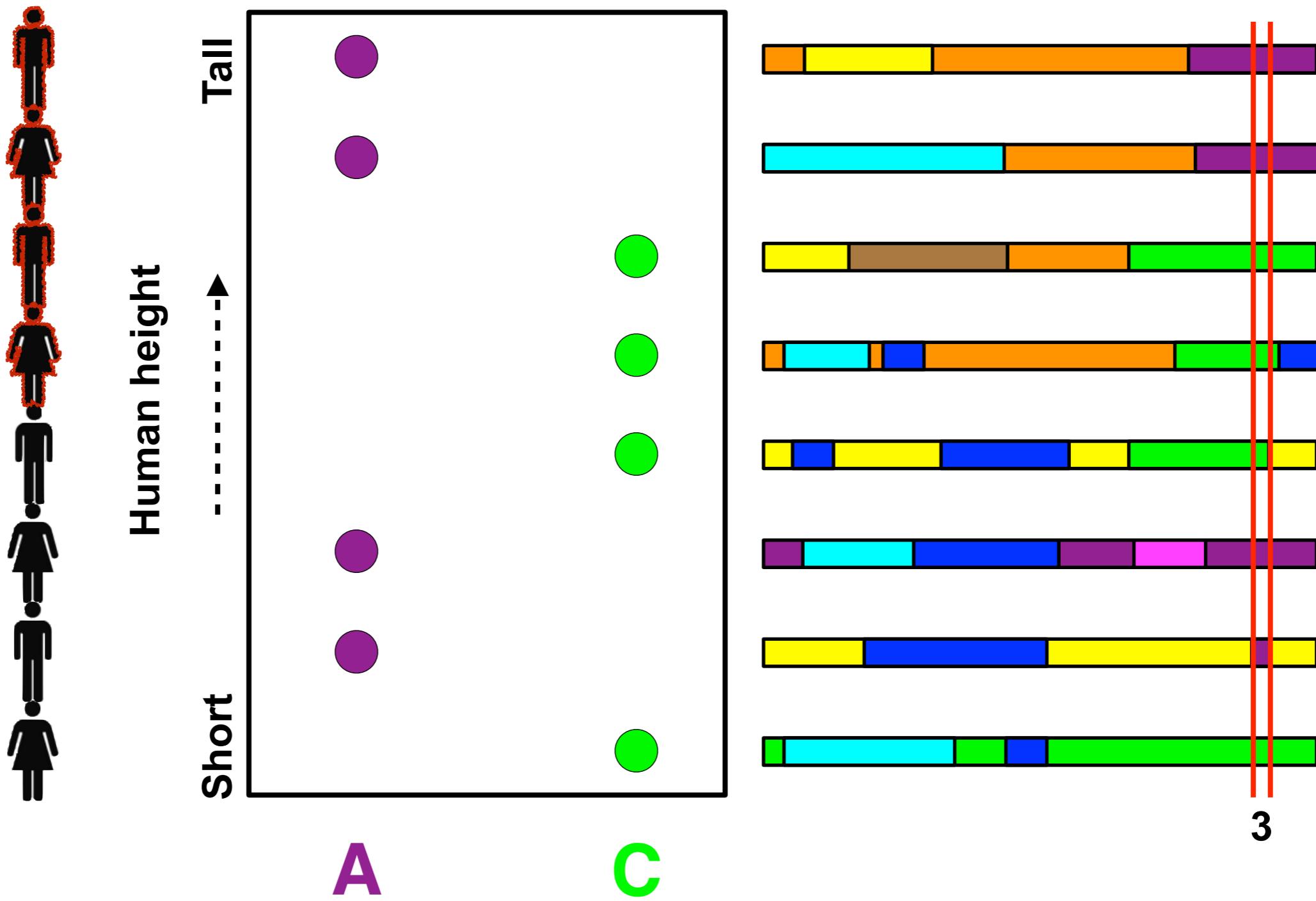
Association mapping: Correlating genotype with phenotype



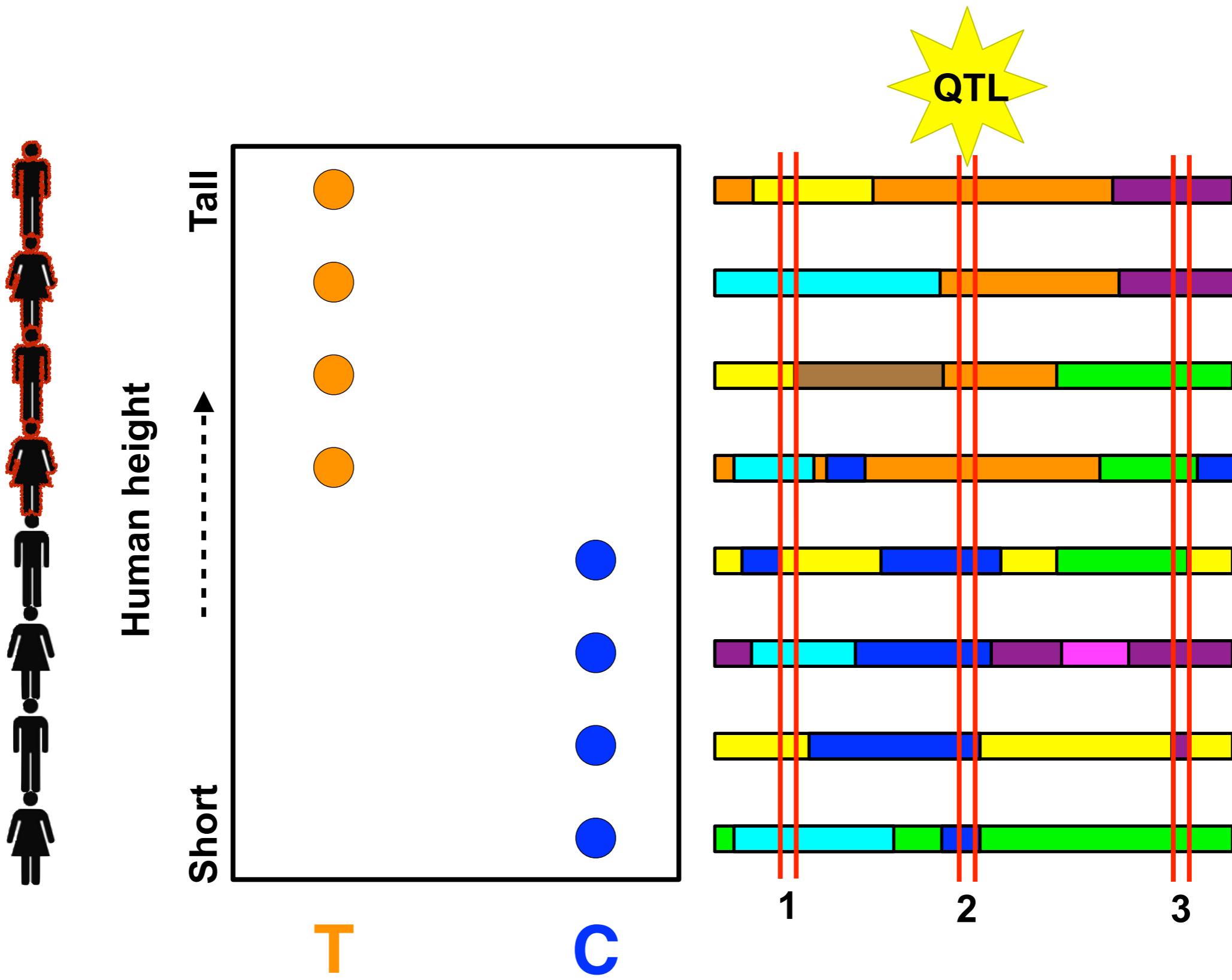
Association mapping: Correlating genotype with phenotype



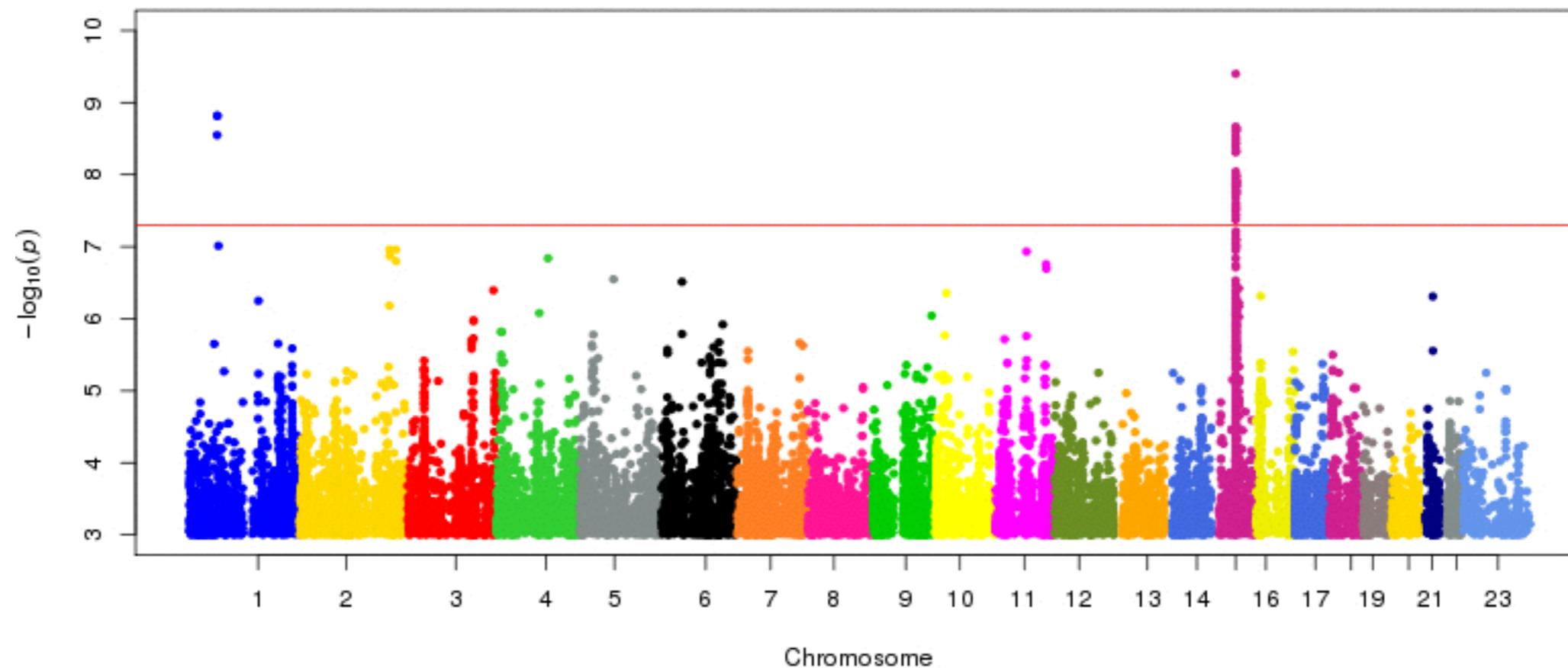
Association mapping: Correlating genotype with phenotype



Association mapping: Correlating genotype with phenotype

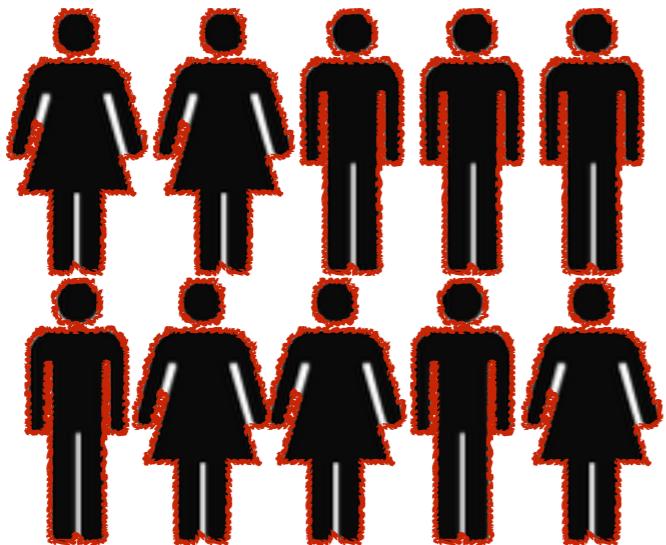


An example Manhattan plot of GWA mapping results



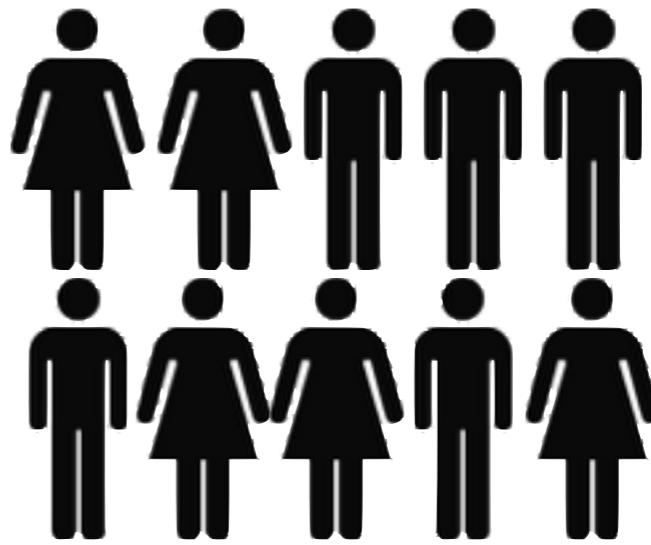
Styrkarsdottir *et al.* Nature 2014

GWAS calculation



4000 Cases

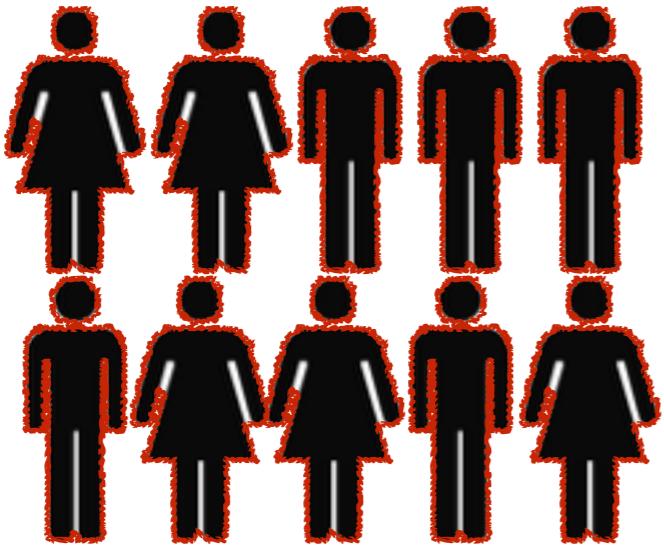
SNP1
(G or A) 2000 of 4000 (50%)



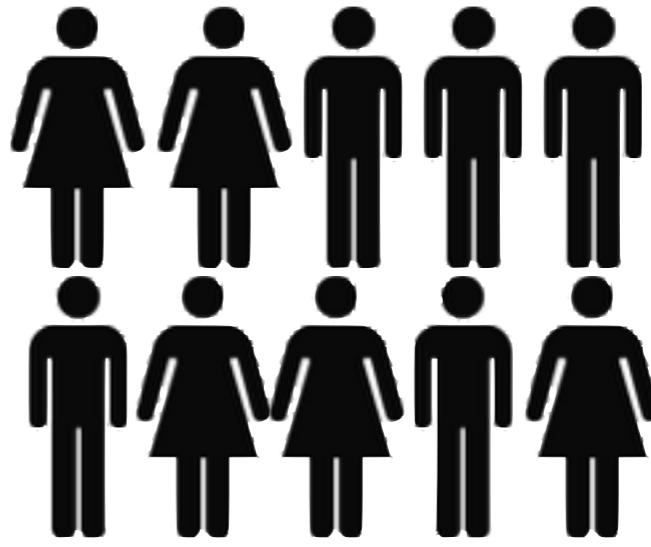
6000 Controls

2500 of 6000 (42%)

GWAS calculation



4000 Cases



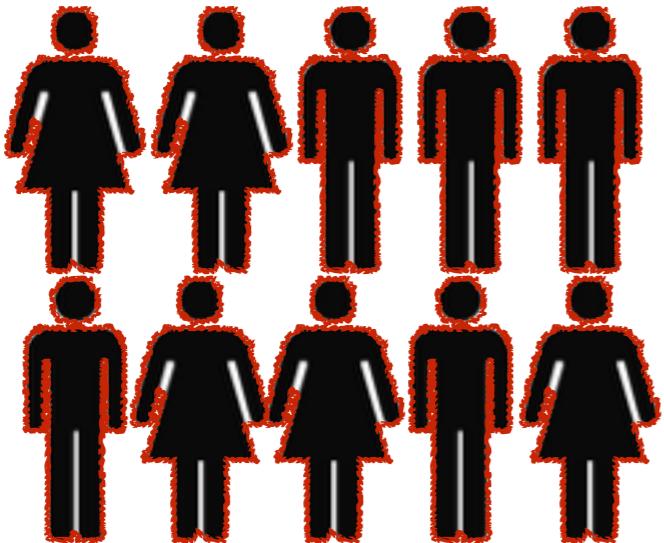
6000 Controls

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

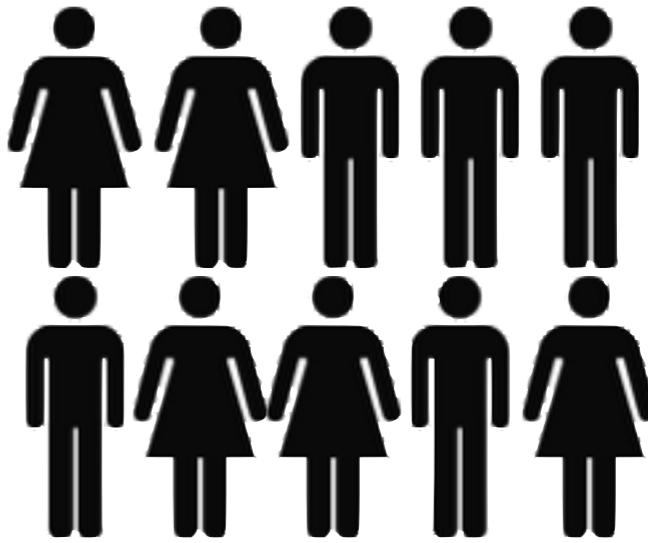
2500 of 6000 (42%)

	Cases	Controls
G	2000	2500
A	2000	3500

GWAS calculation



4000 Cases



6000 Controls

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

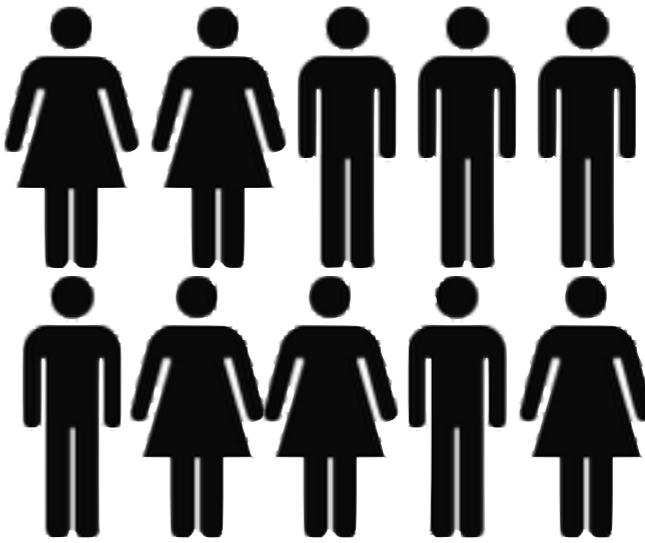
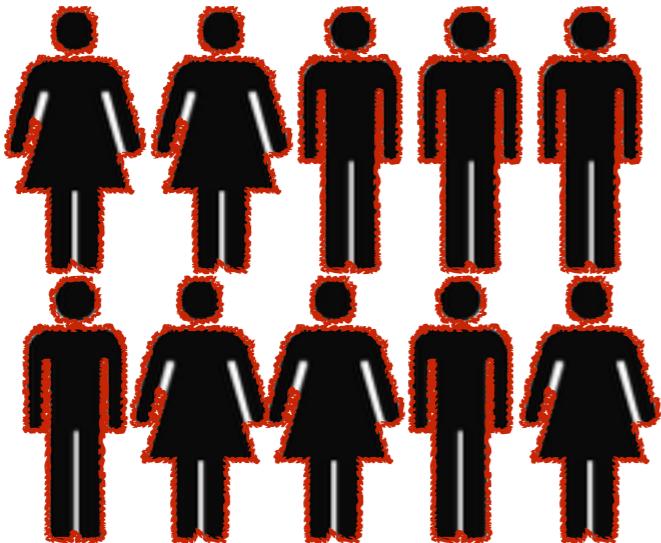
2500 of 6000 (42%)

	Cases	Controls
G	2000	2500
A	2000	3500

Observed

Expected

GWAS calculation



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

2500 of 6000 (42%)

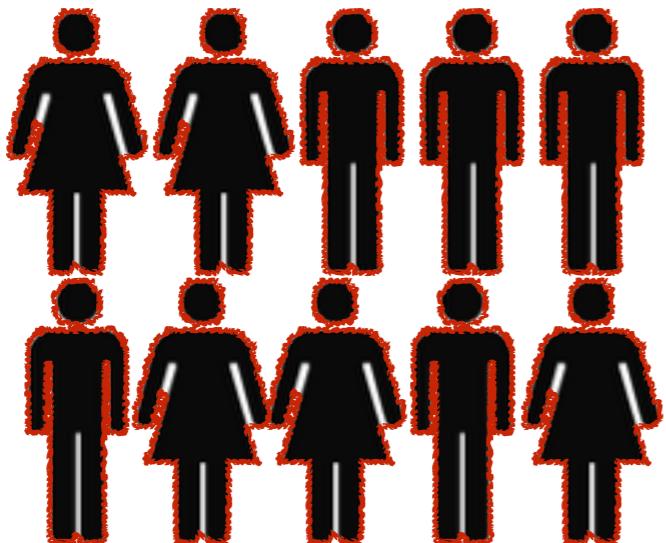
	Cases	Controls
G	2000	2500
A	2000	3500

Observed Expected

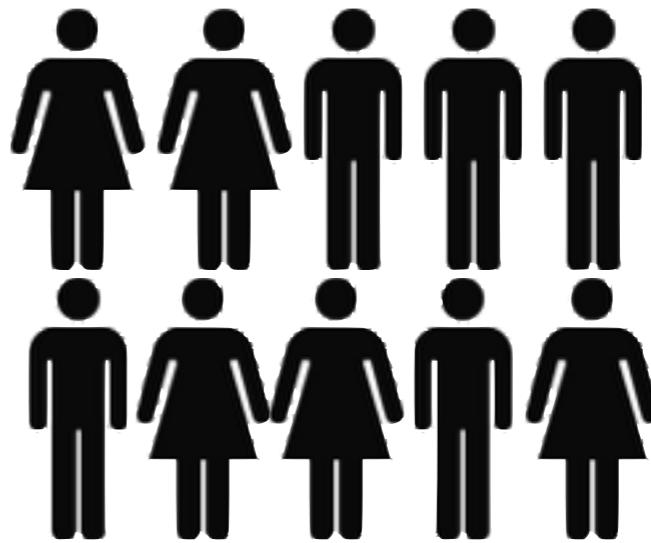
Pearson's chi-squared test
with one degree of freedom

67.0038 or p-value of 2.71e-16

GWAS calculation



4000 Cases



6000 Controls

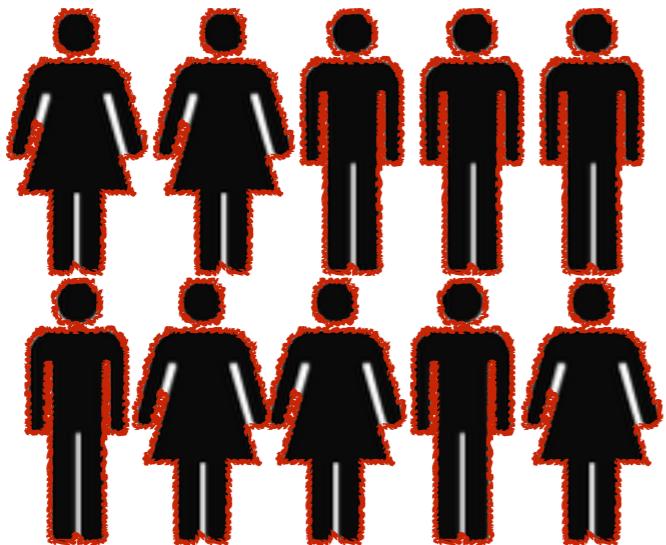
SNP1
(G or A) 2000 of 4000 (50%)

2500 of 6000 (42%)

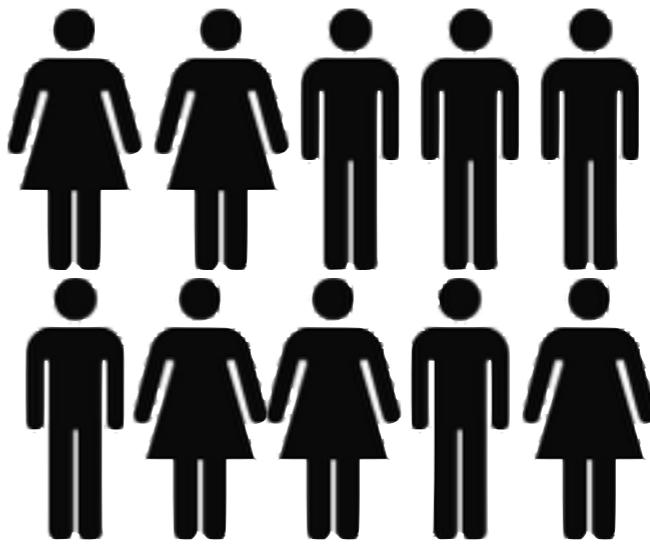
SNP2
(T or C) 1600 of 4000 (40%)

2300 of 6000 (38%)

GWAS calculation



4000 Cases



6000 Controls

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

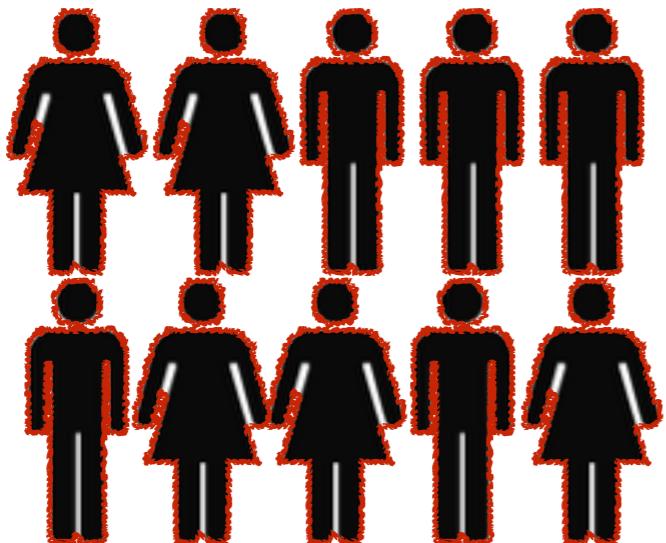
2500 of 6000 (42%)

SNP2
(T or C) 1600 of 4000 (40%)

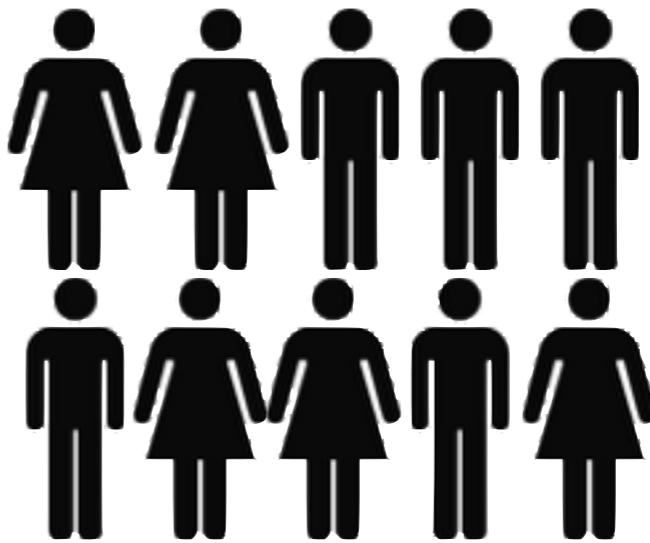
2300 of 6000 (38%)

	Cases	Controls
T	1600	2300
C	2400	3700

GWAS calculation



4000 Cases



6000 Controls

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

2500 of 6000 (42%)

SNP2
(T or C) 1600 of 4000 (40%)

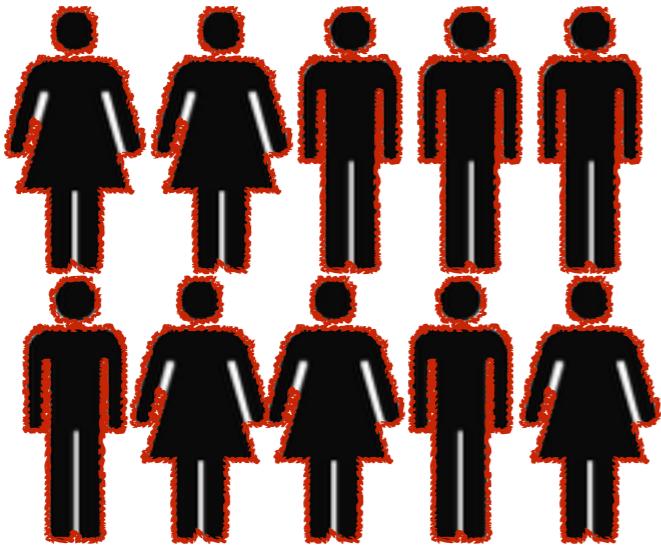
2300 of 6000 (38%)

	Cases	Controls
T	1600	2300
C	2400	3700

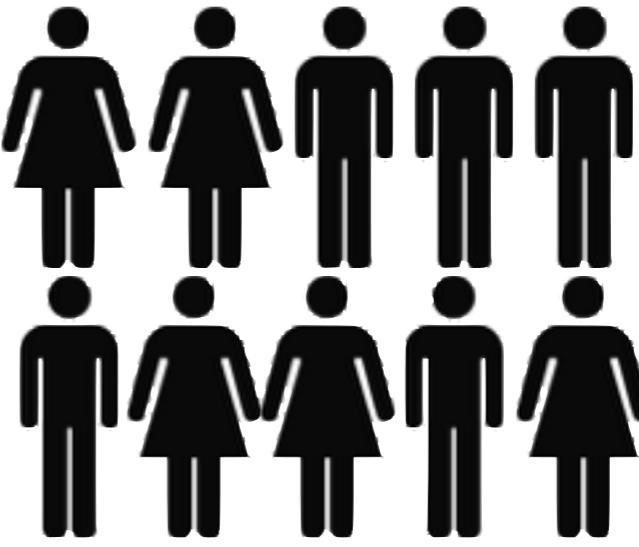
Observed

Expected

GWAS calculation



4000 Cases



6000 Controls

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

2500 of 6000 (42%)

SNP2
(T or C) 1600 of 4000 (40%)

2300 of 6000 (38%)

	Cases	Controls
T	1600	2300
C	2400	3700

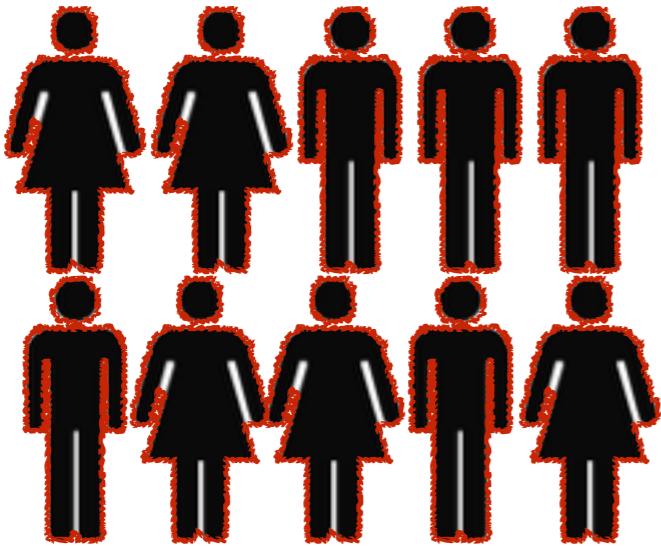
Observed

Expected

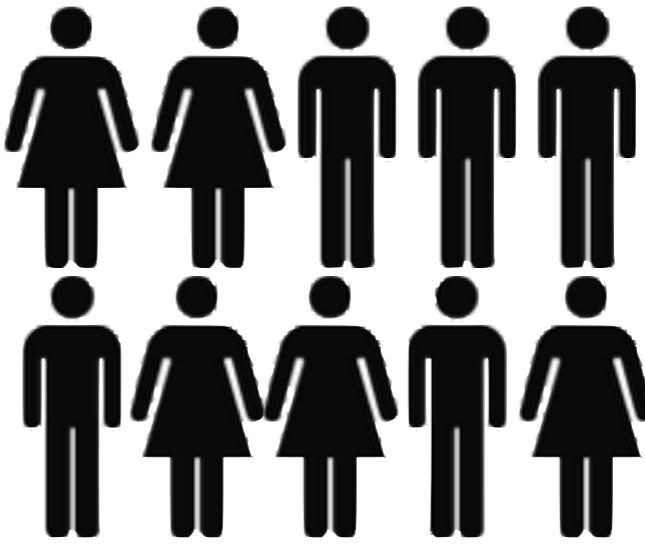
Pearson's chi-squared test
with one degree of freedom

2.7327 or p-value of 0.09831

GWAS calculation



4000 Cases



6000 Controls

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

2500 of 6000 (42%)

SNP2
(T or C) 1600 of 4000 (40%)

2300 of 6000 (38%)

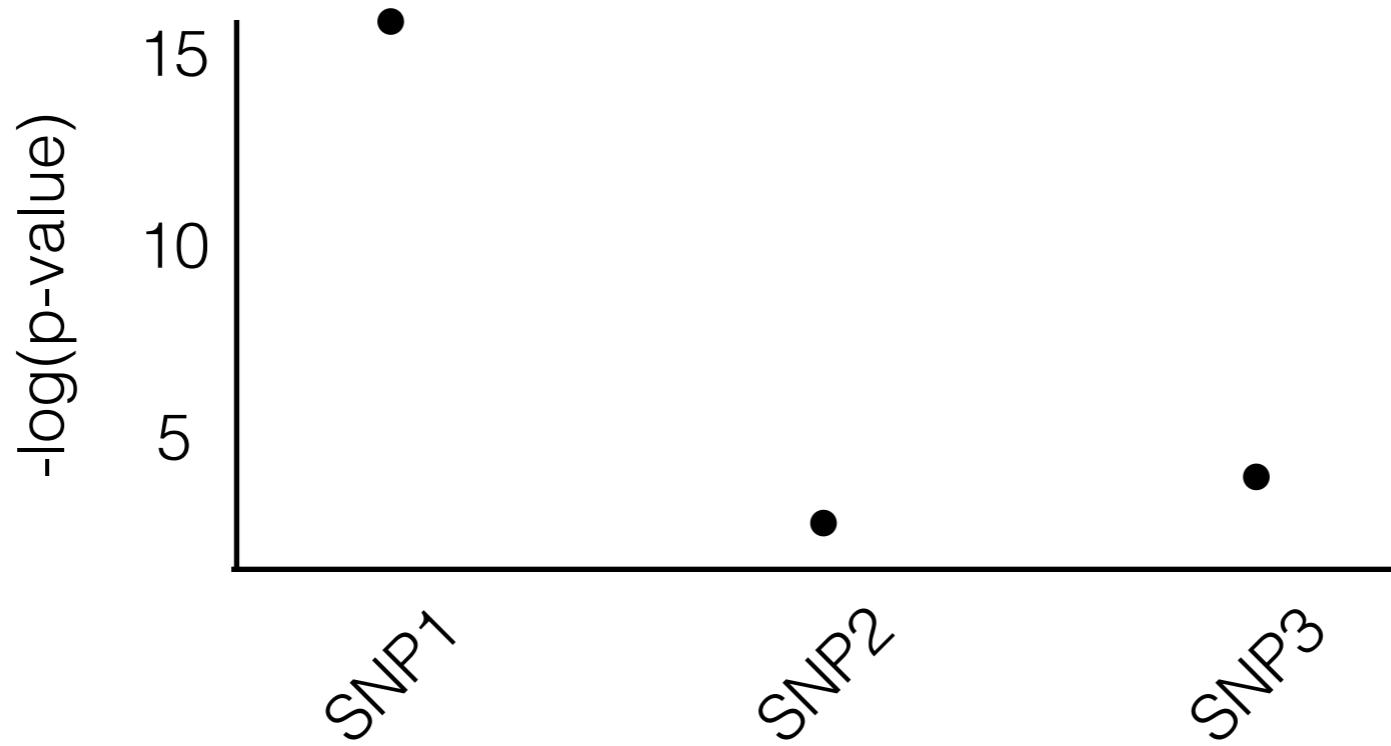
SNP3
(C or A) 1800 of 4000 (45%)

2500 of 6000 (40%)

	Cases	Controls
T	1800	2500
C	2200	3500

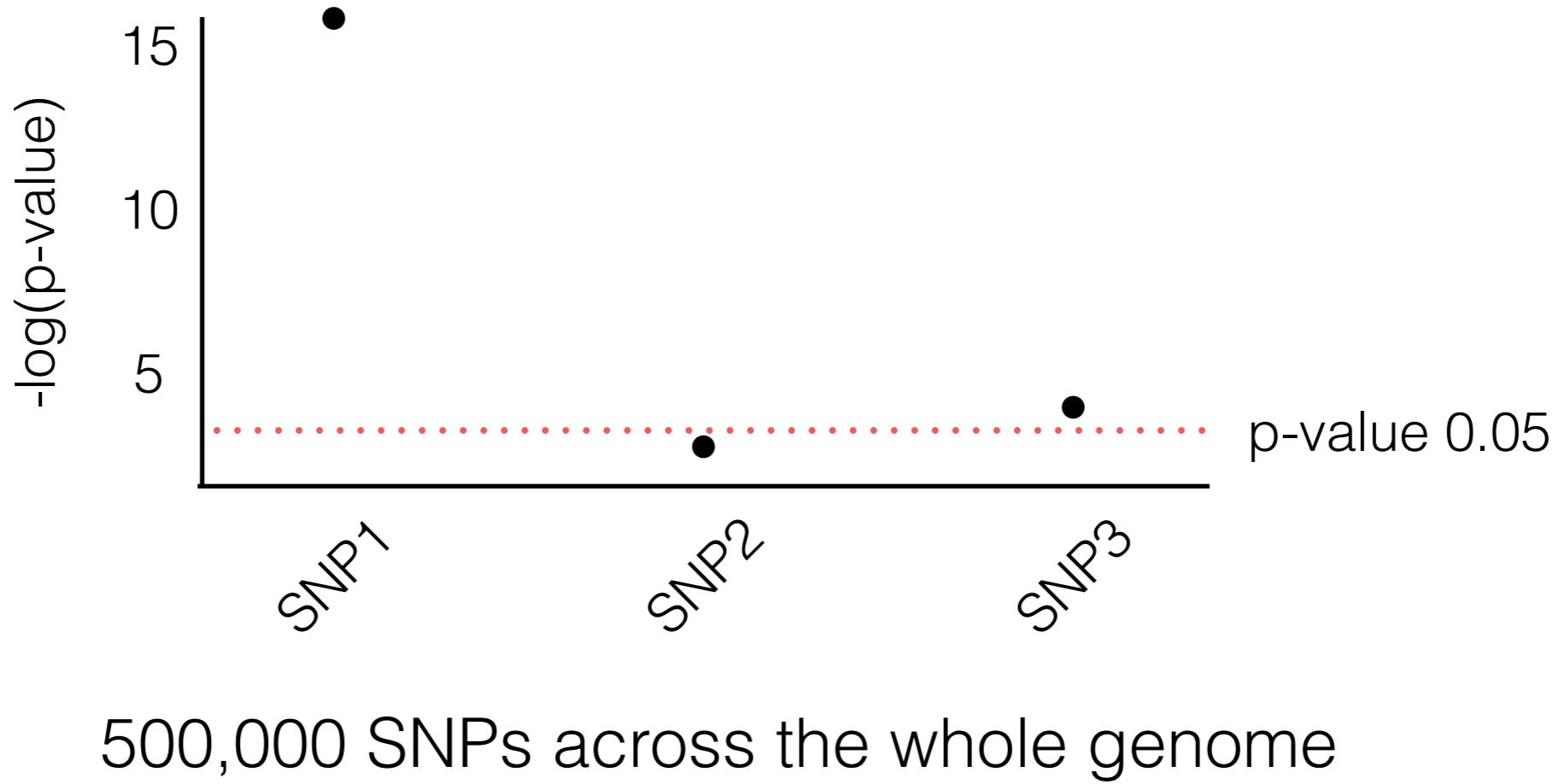
10.7443 or p-value of 0.001046

GWAS results

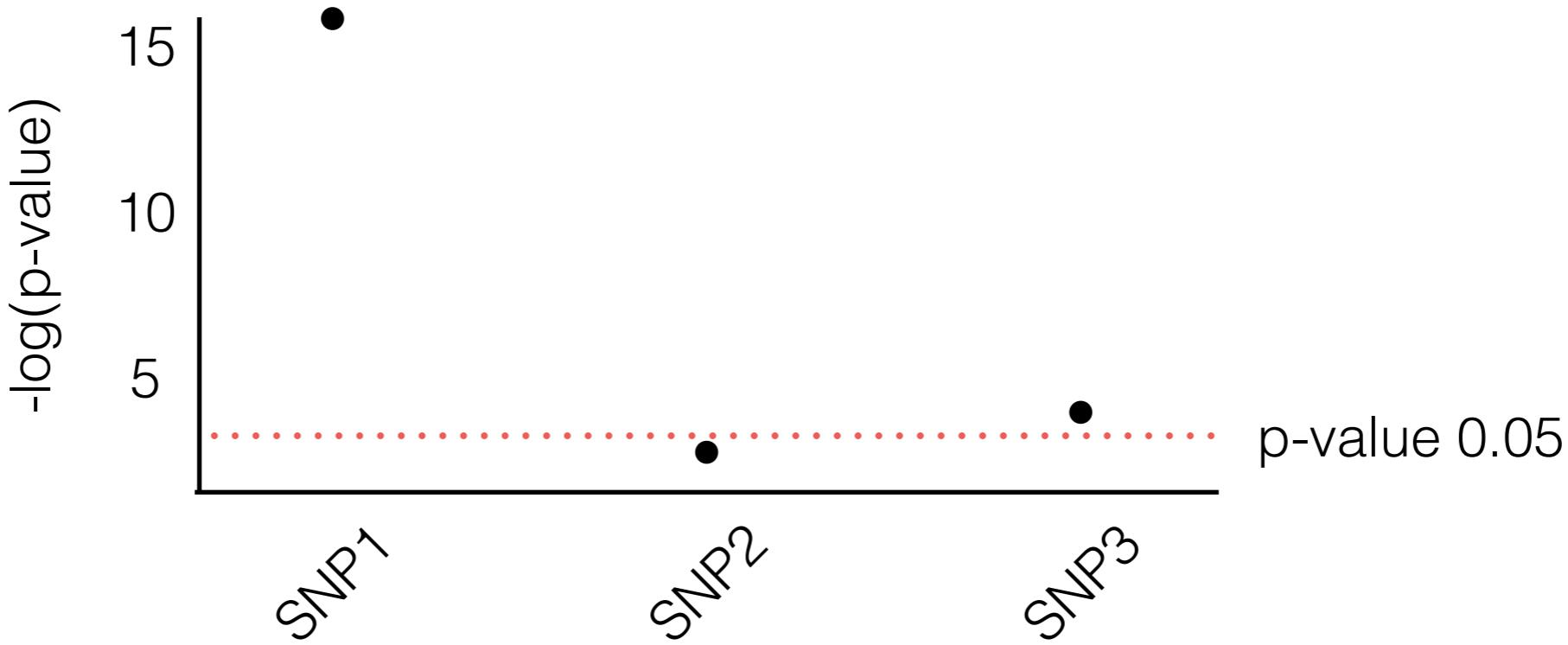


500,000 SNPs across the whole genome

GWAS results



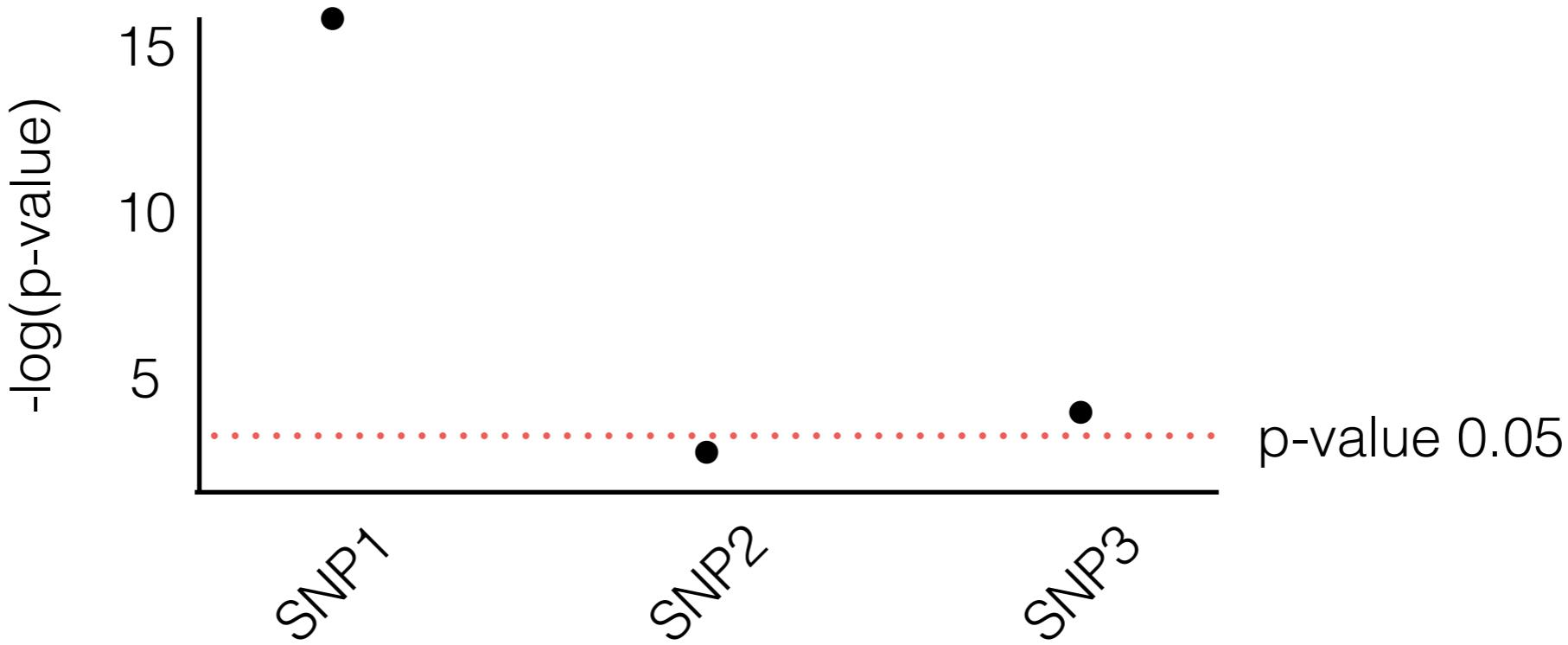
GWAS results



500,000 SNPs across the whole genome

500,000 tests with a p-value of 0.05 means
that we would reject the null hypothesis
for 25,000 SNPs by chance

GWAS results



500,000 SNPs across the whole genome

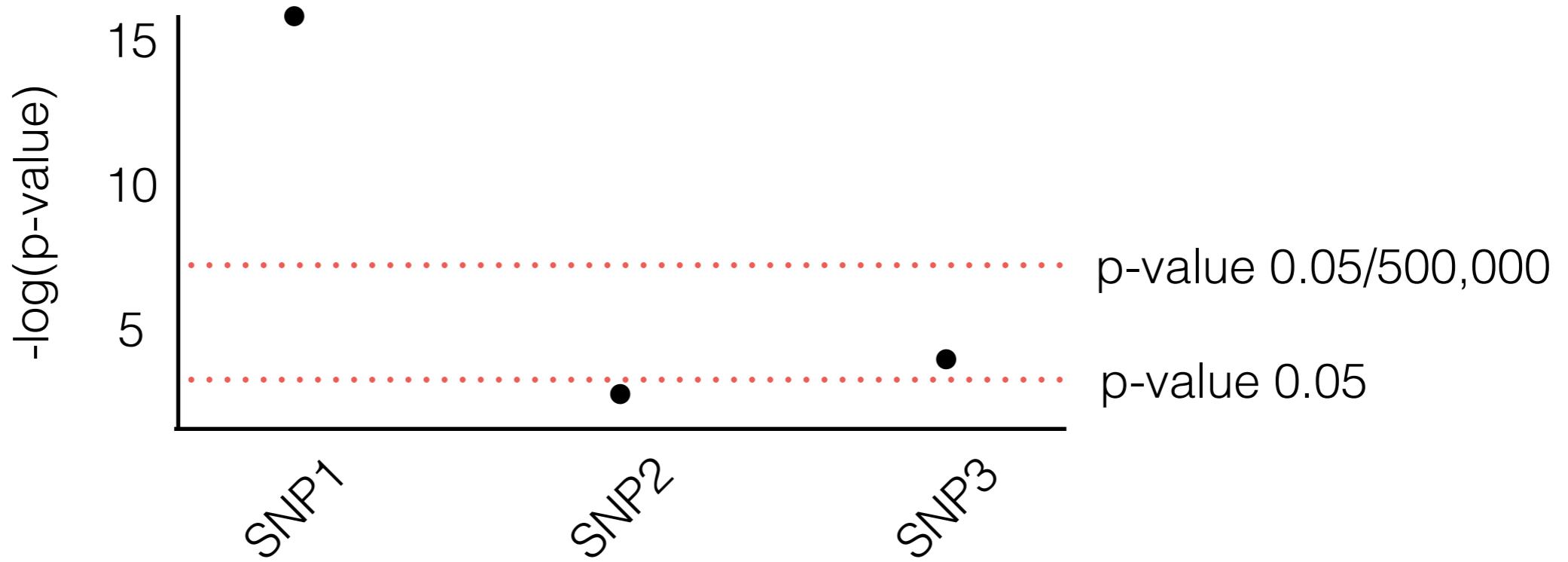
500,000 tests with a p-value of 0.05 means
that we would reject the null hypothesis
for 25,000 SNPs by chance



Carlo Bonferroni

Bonferroni correction $0.05 / 500,000$ or $1e-7$

GWAS results



500,000 SNPs across the whole genome

500,000 tests with a p-value of 0.05 means
that we would reject the null hypothesis
for 25,000 SNPs by chance



Carlo Bonferroni

Bonferroni correction $0.05 / 500,000$ or $1e-7$

Three possibilities for the results of any GWA mapping

Three possibilities for the results of any GWA mapping

1. Marker is the *functional variant*

Three possibilities for the results of any GWA mapping

1. Marker is the *functional variant*
2. Marker is in *linkage disequilibrium* with functional variant

Three possibilities for the results of any GWA mapping

1. Marker is the *functional variant*
2. Marker is in *linkage disequilibrium* with functional variant
3. Marker is associated because of *population relatedness*
(population structure)

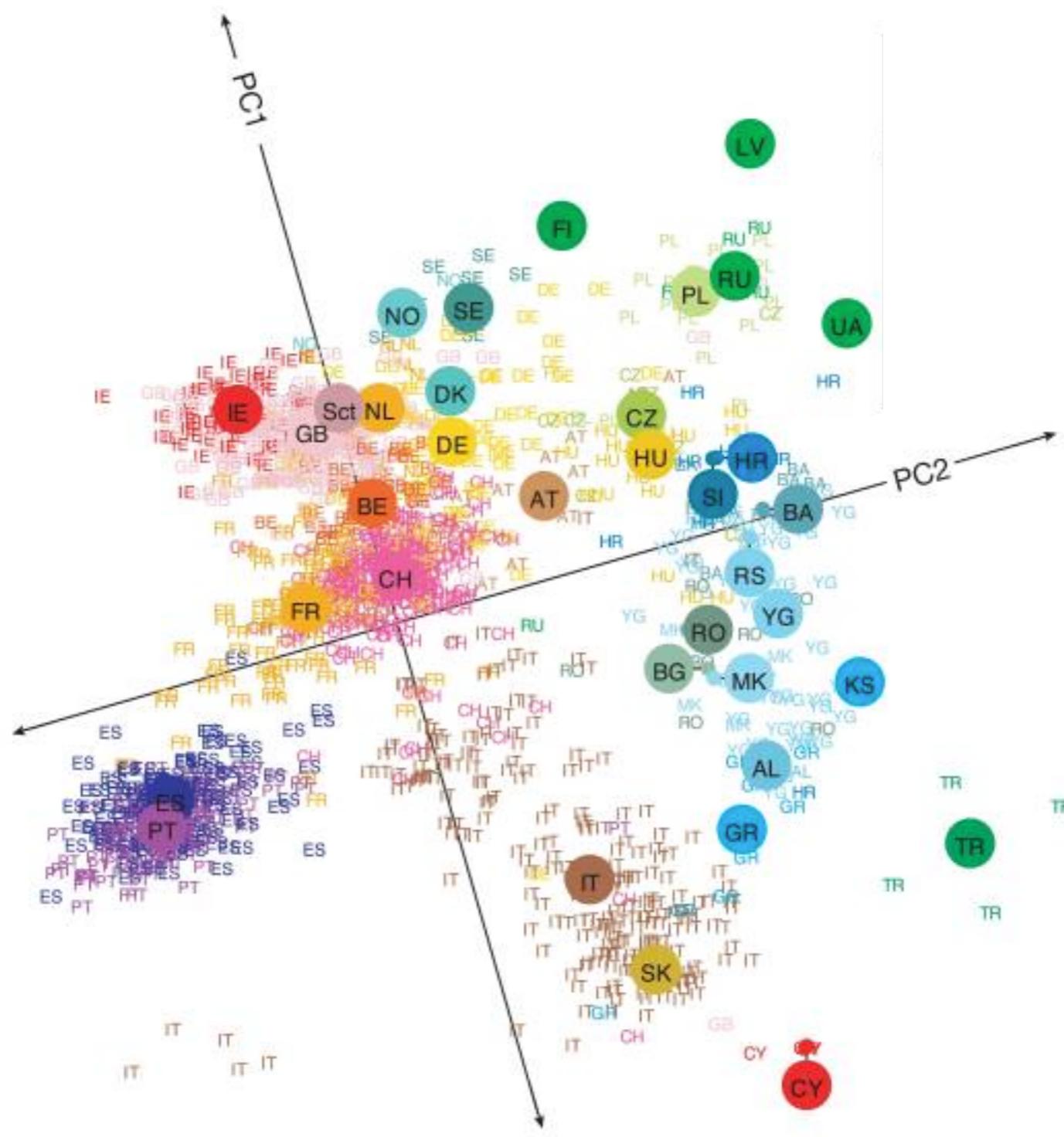
Population structure confounds human association mapping



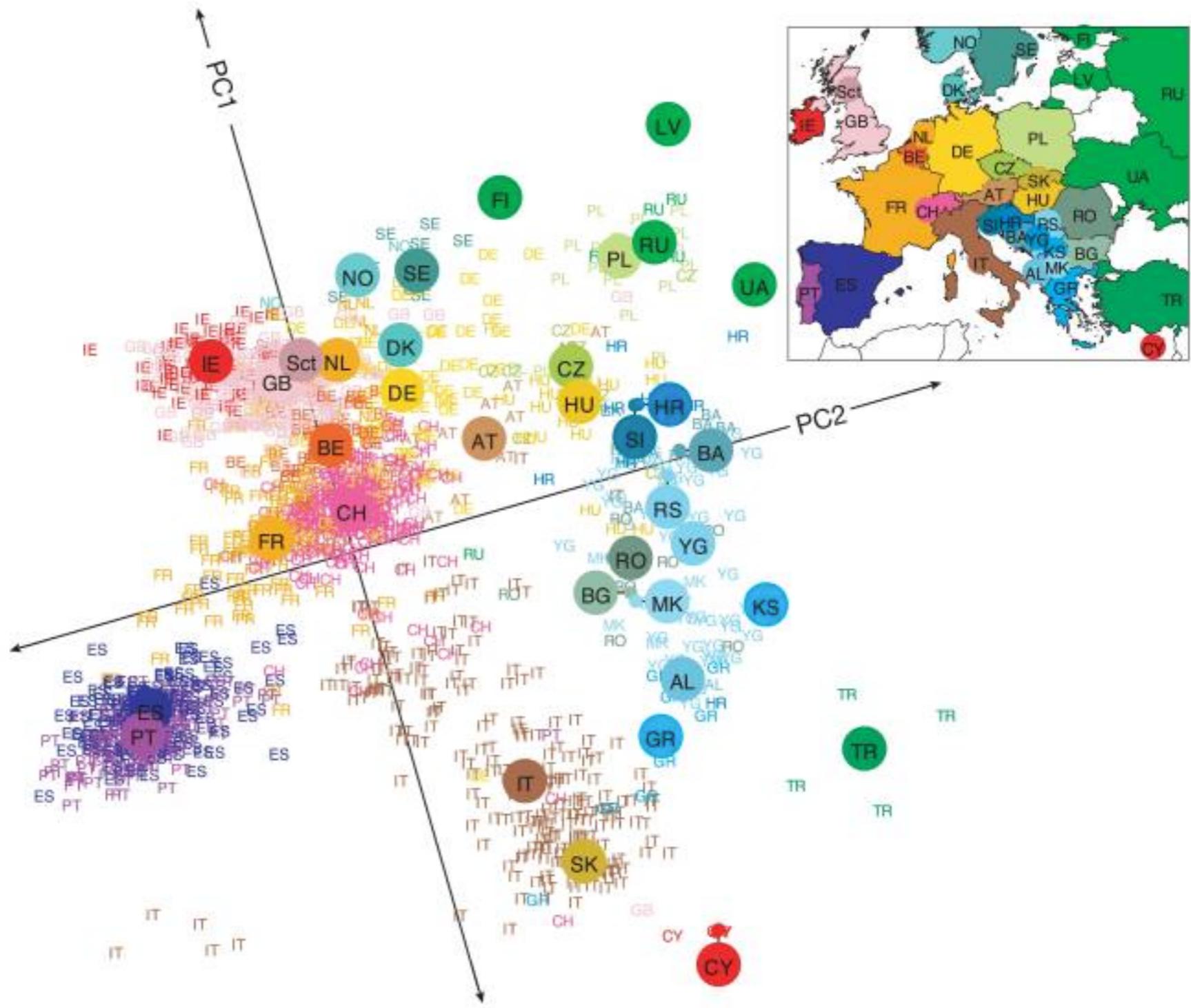
Relatedness of people caused by non-random mating
is called population structure (or stratification)

GWA mapping across populations might find signals of relatedness
if the disease is correlated.

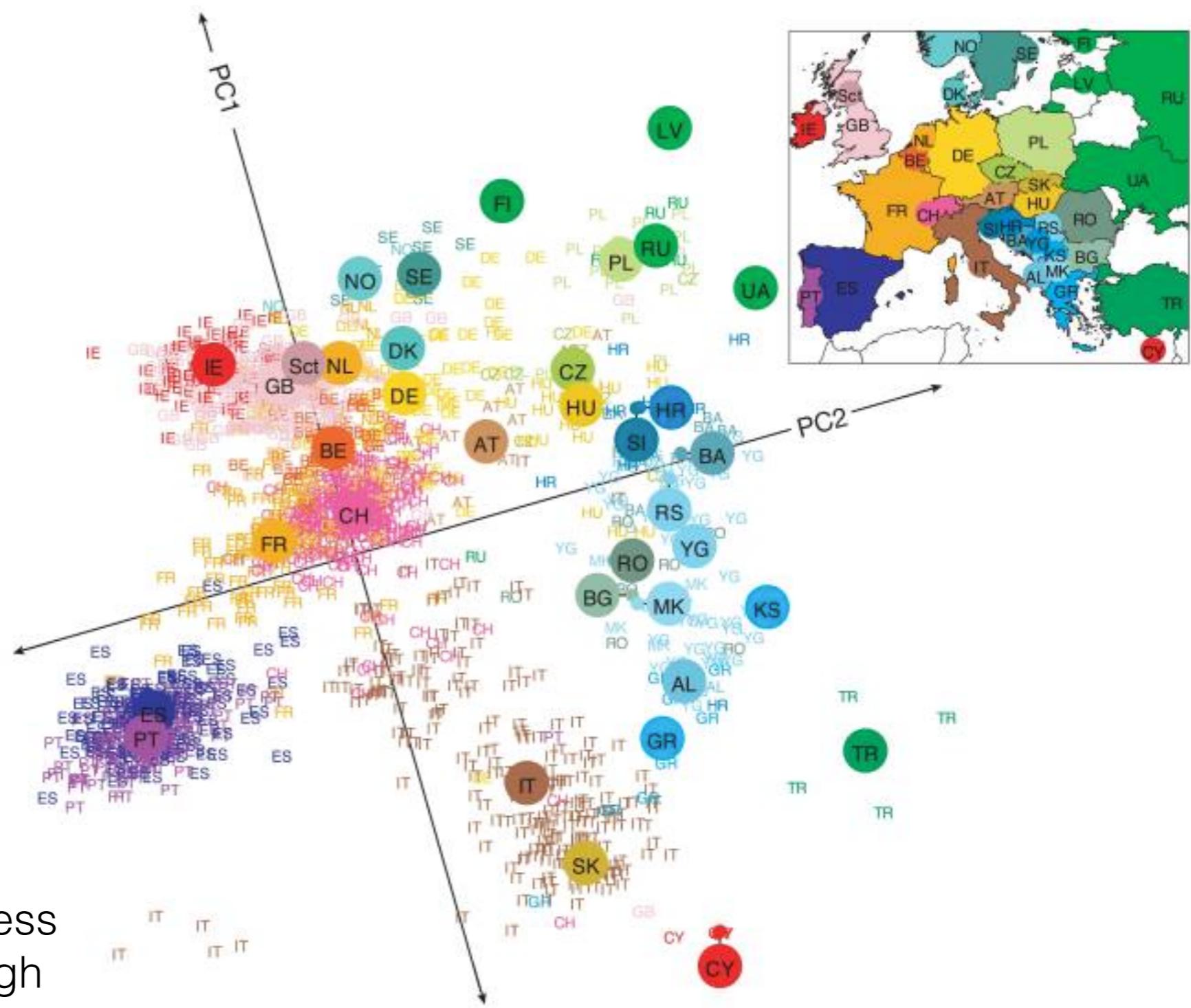
Population structure confounds human association mapping



Population structure confounds human association mapping



Population structure confounds human association mapping



The effects of relatedness can be reduced through regression.

GWA mapping within groups and replication



GWA mapping works best within a related population

The mapping *might* be replicated in different populations

GWA mapping within groups and replication



GWA mapping works best within a related population

The mapping *might* be replicated in different populations

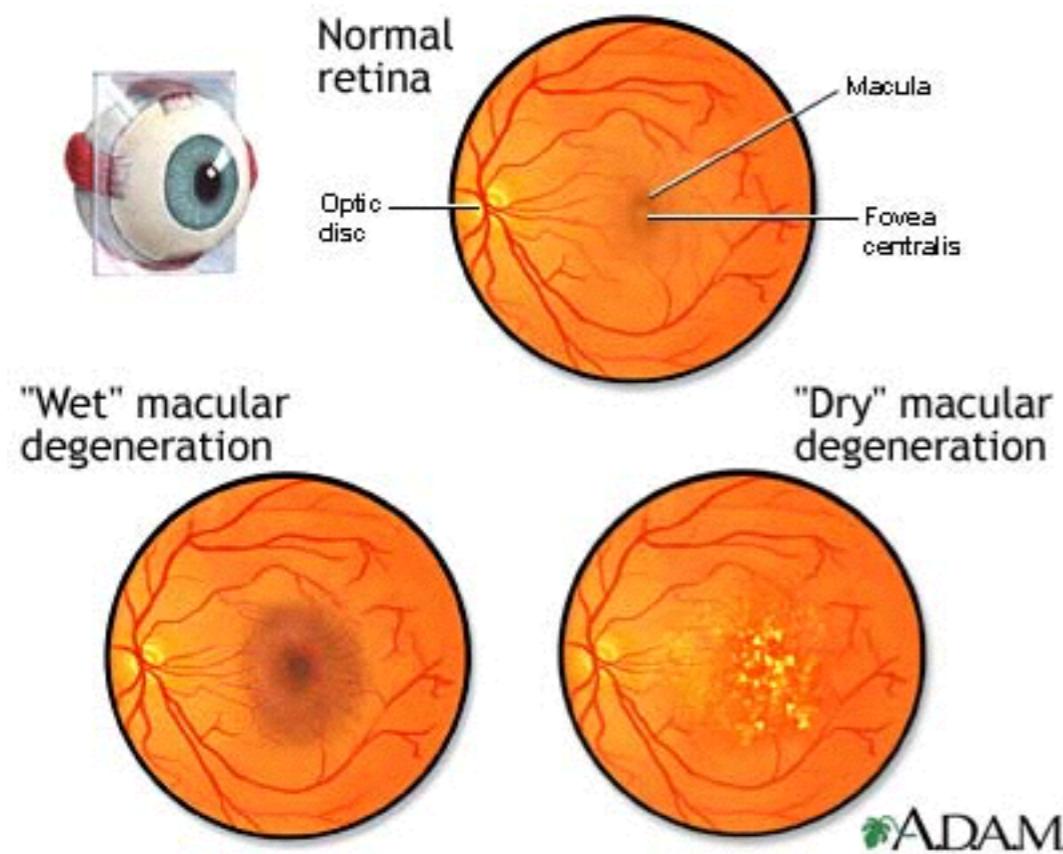
GWA mapping within groups and replication



GWA mapping works best within a related population

The mapping *might* be replicated in different populations

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



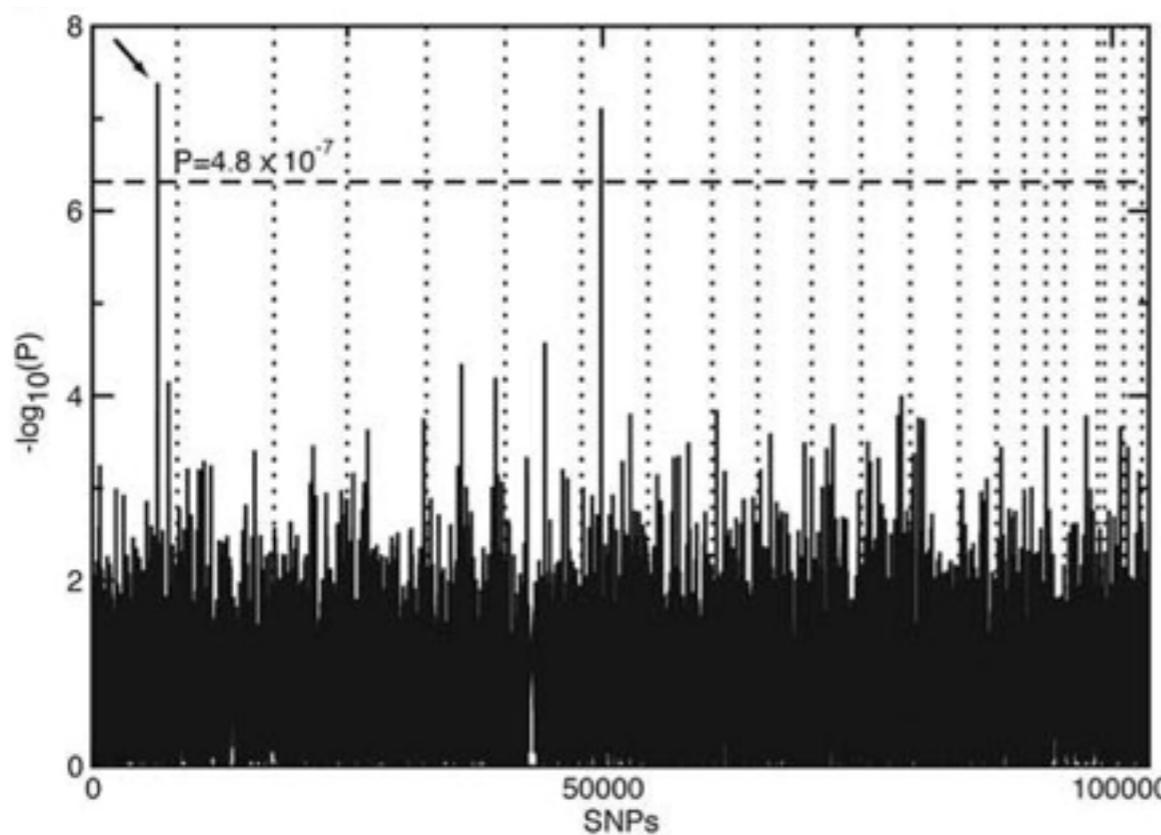
96 Cases



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



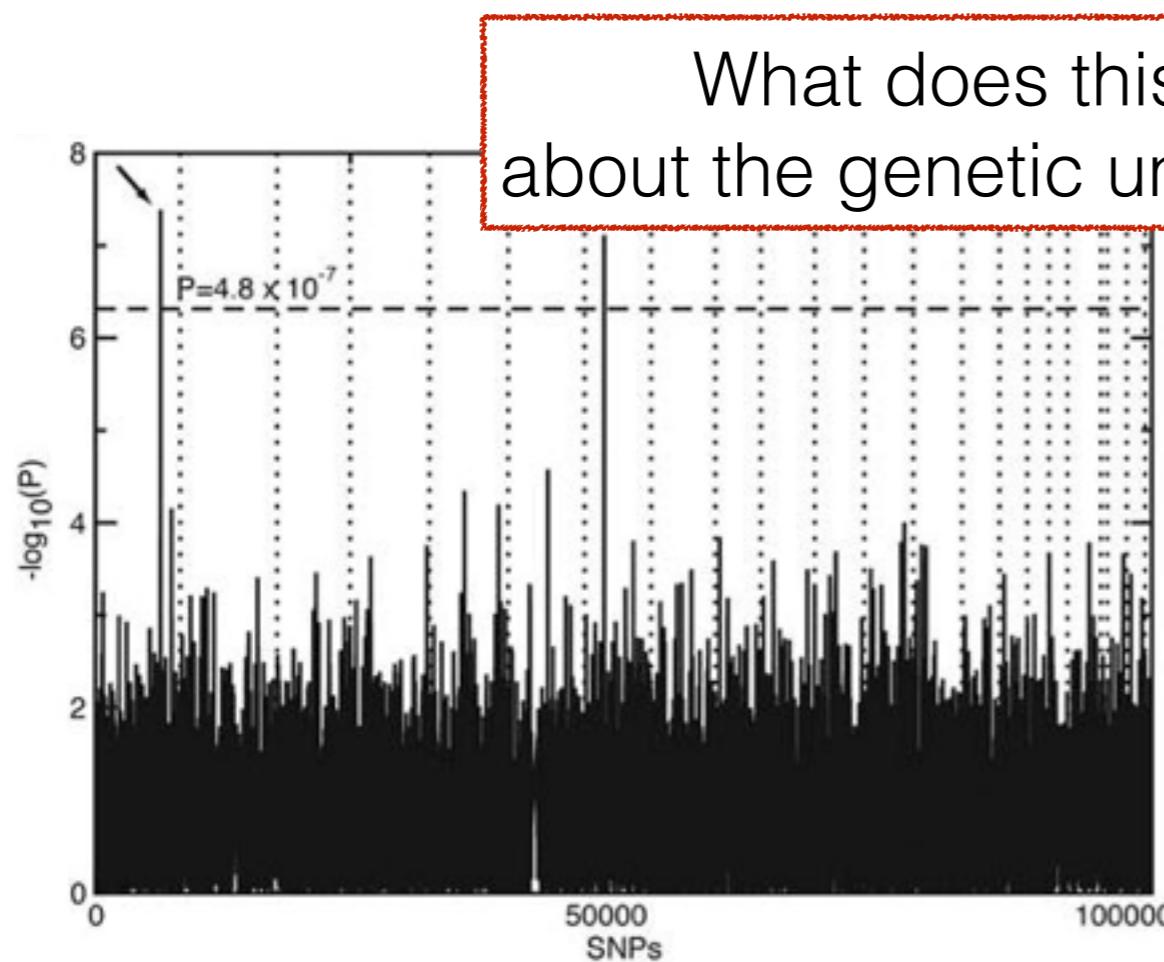
96 Cases



50 Controls

Klein *et al.* Science 2005

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



What does this success tell us
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

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



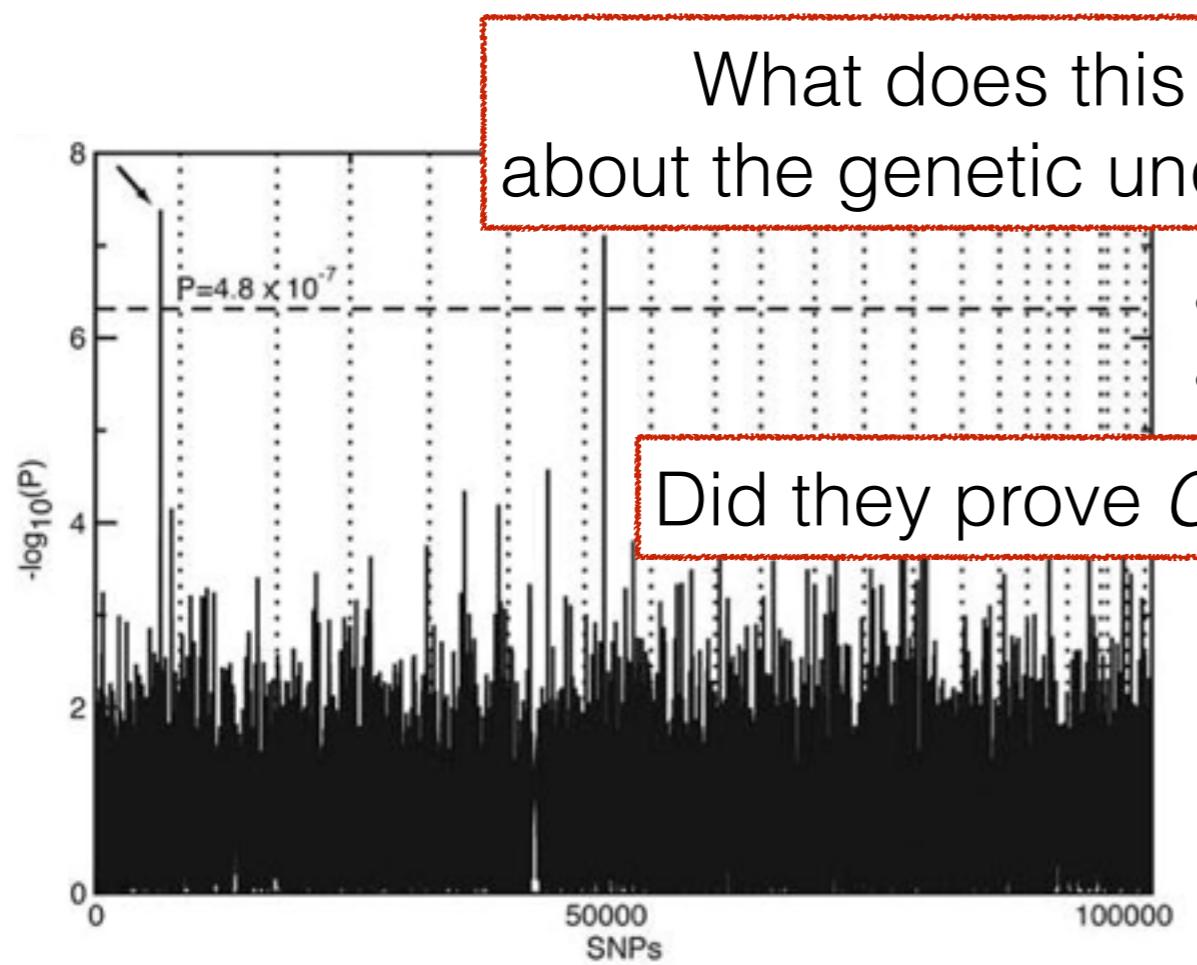
96 Cases



50 Controls

Klein *et al.* Science 2005

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

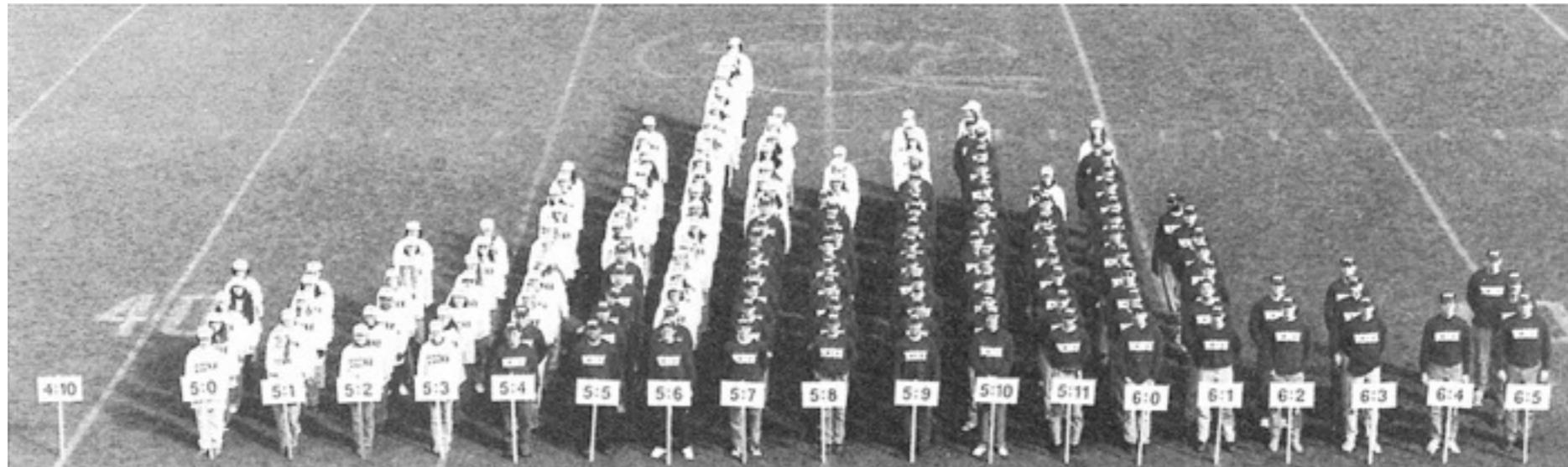


What does this success tell us
about the genetic underpinnings of AMD?

- Complement Factor H
- Haploinsufficient gene

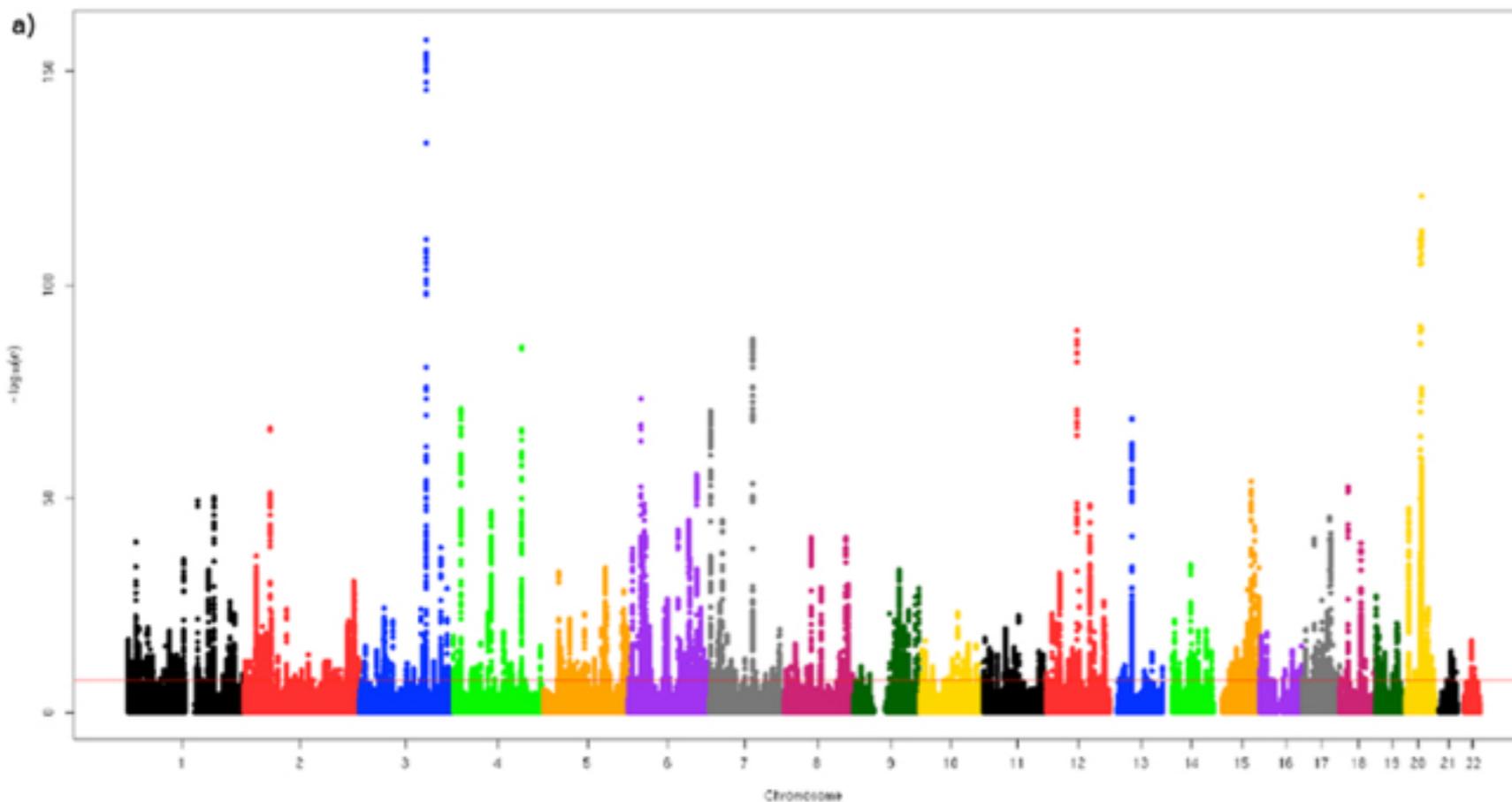
Did they prove *CFH* is the gene? matory cascade linkage studies found same region of genome

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

Lessons from the GWAS era

- Many traits are polygenic

Lessons from the GWAS era

- Many traits are polygenic
- Effect sizes of common variants are very small

Lessons from the GWAS era

- Many traits are polygenic
- Effect sizes of common variants are very small
- Many associated SNPs are near genes

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

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), there are three genotypes: AA, Aa, aa

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

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

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

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

Genotype relative risk (GRR)

For a diallelic locus (A or a), there are 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 _{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

Odds ratios estimate risk ratio, so ratios of case:control ratios estimate relative risks

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

Odds ratios estimate risk ratio, so ratios of case:control ratios estimate relative risks

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

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

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

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

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

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

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

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

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

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 20% more likely than the aa genotype to have the disease

The Aa genotype is 5% more likely than the aa genotype to have the disease

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

Round vs. Wrinkled



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

Round vs. Wrinkled



Starch branching enzyme 1

(Bhattacharyya *et al.* 1990)

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

Round vs. Wrinkled



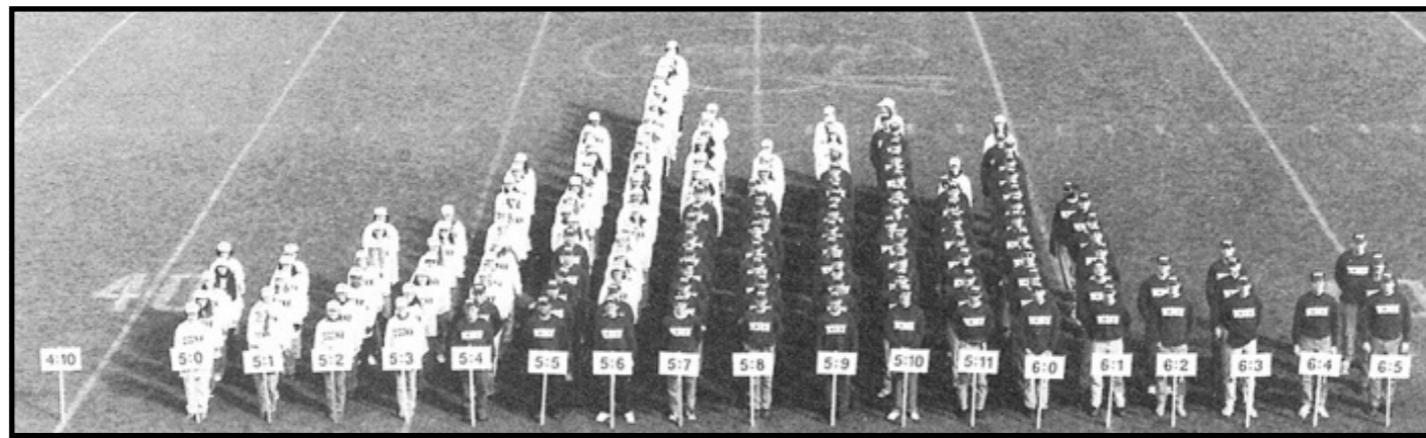
Starch branching enzyme 1

(Bhattacharyya *et al.* 1990)

100% phenotypic variance explained

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

Height



University of Connecticut, 1997



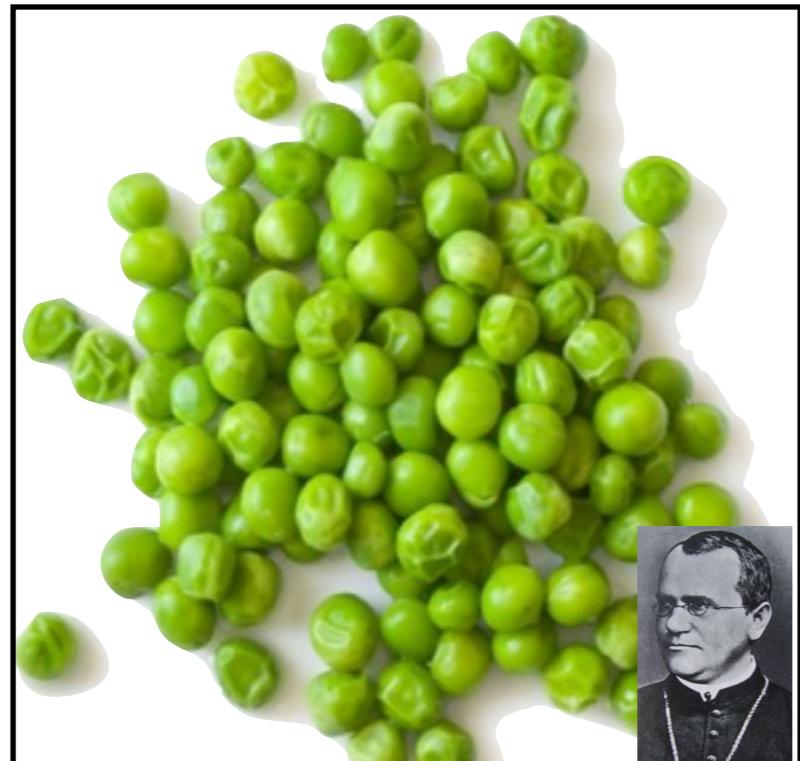
697 loci

(Wood *et al.* 2014)

20% phenotypic variance explained

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

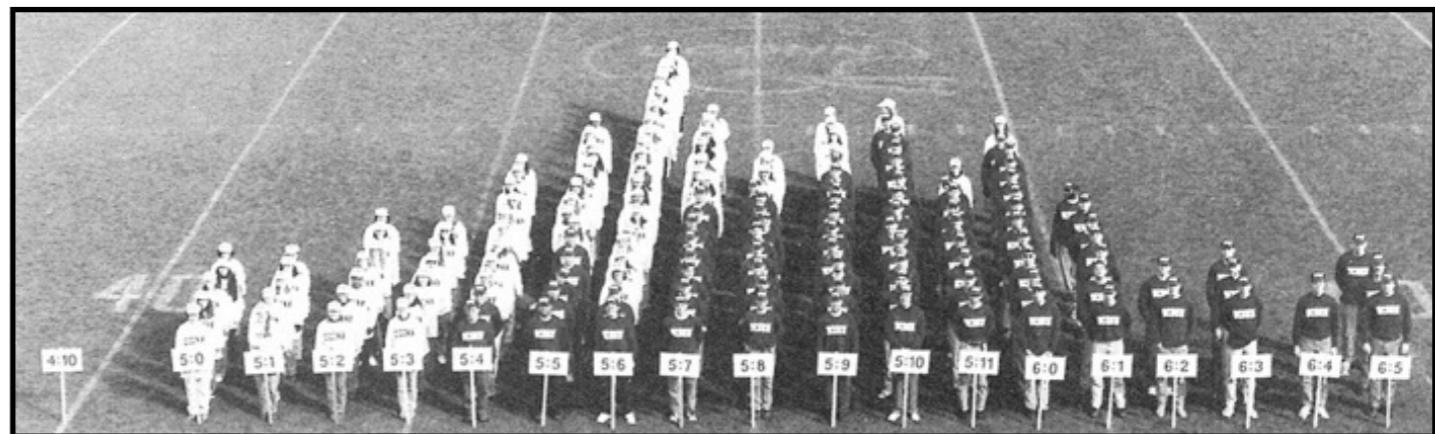
Round vs. Wrinkled



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

100% phenotypic variance explained

Height



University of Connecticut, 1997



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

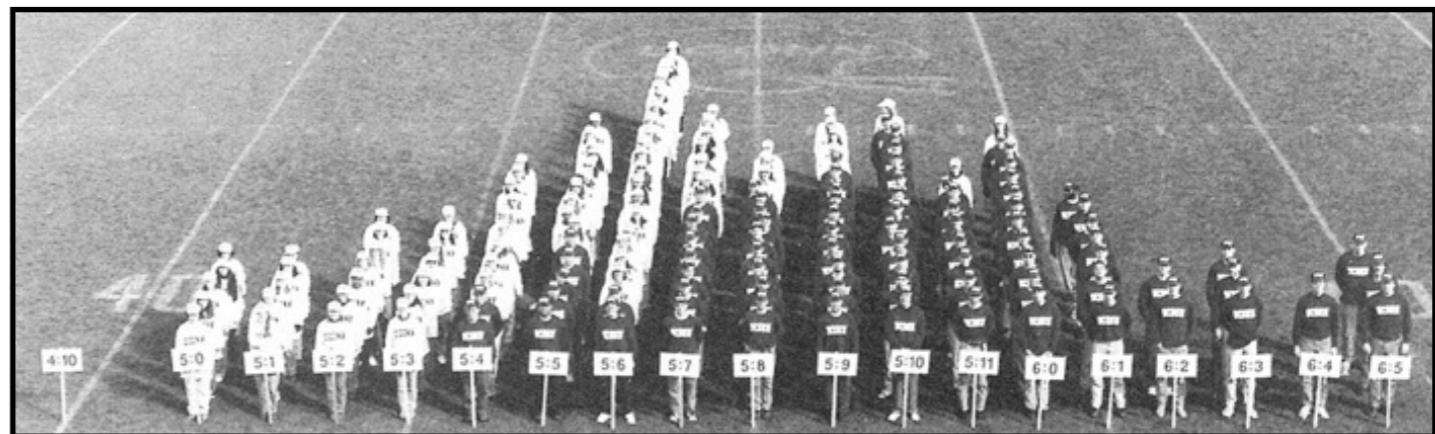


Starch branching enzyme 1
(Bhattacharyya *et al.* 1990)

100% phenotypic variance explained

discontinuous traits

Height



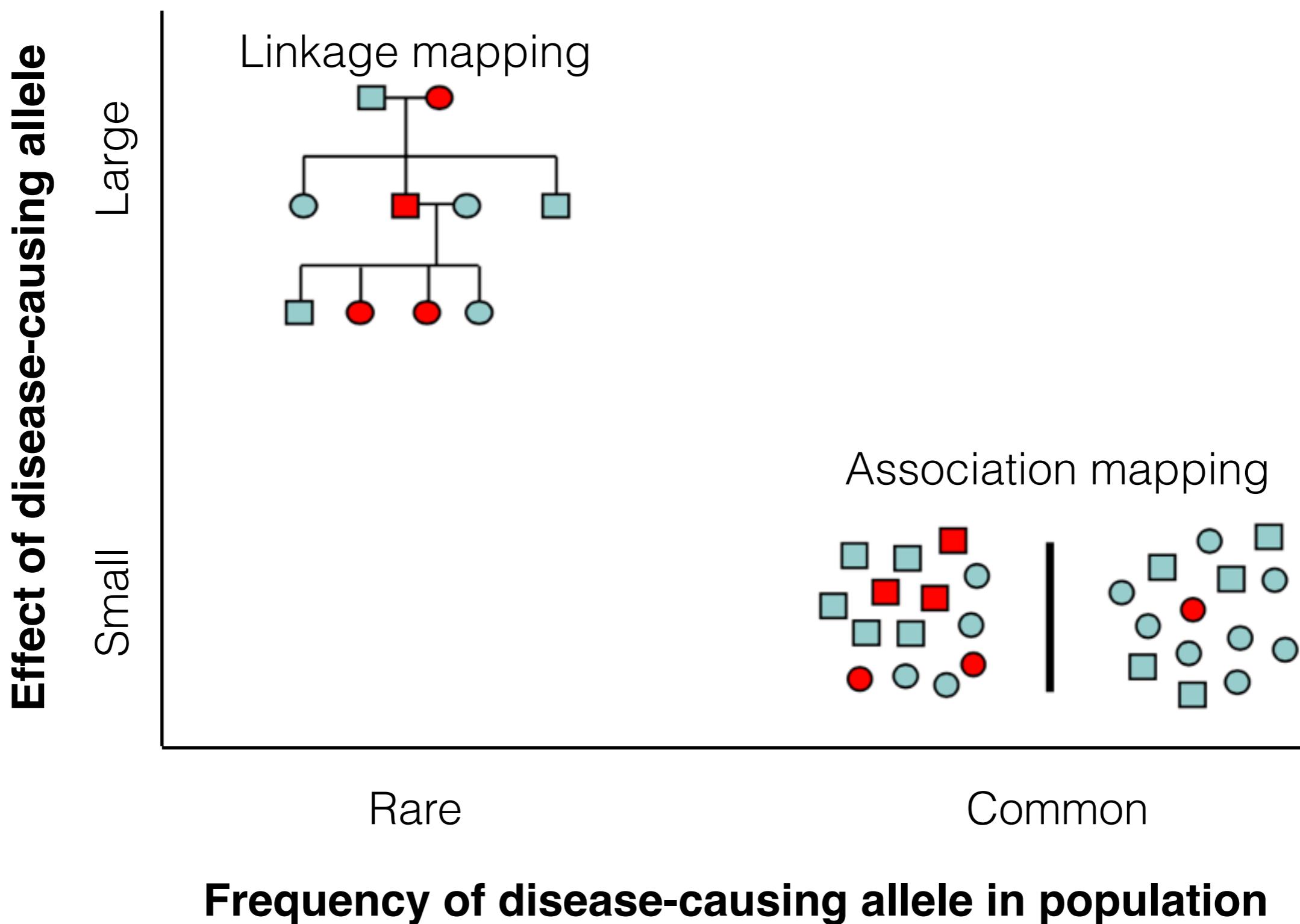
697 loci

(Wood *et al.* 2014)

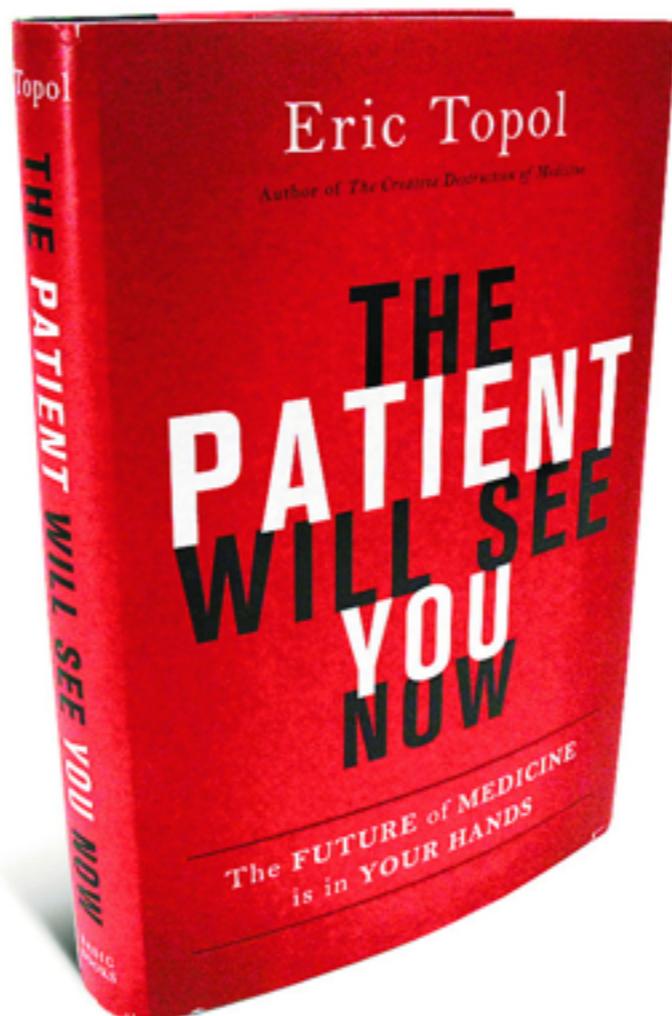
20% phenotypic variance explained

continuous traits

Complex traits can be mapped using both techniques

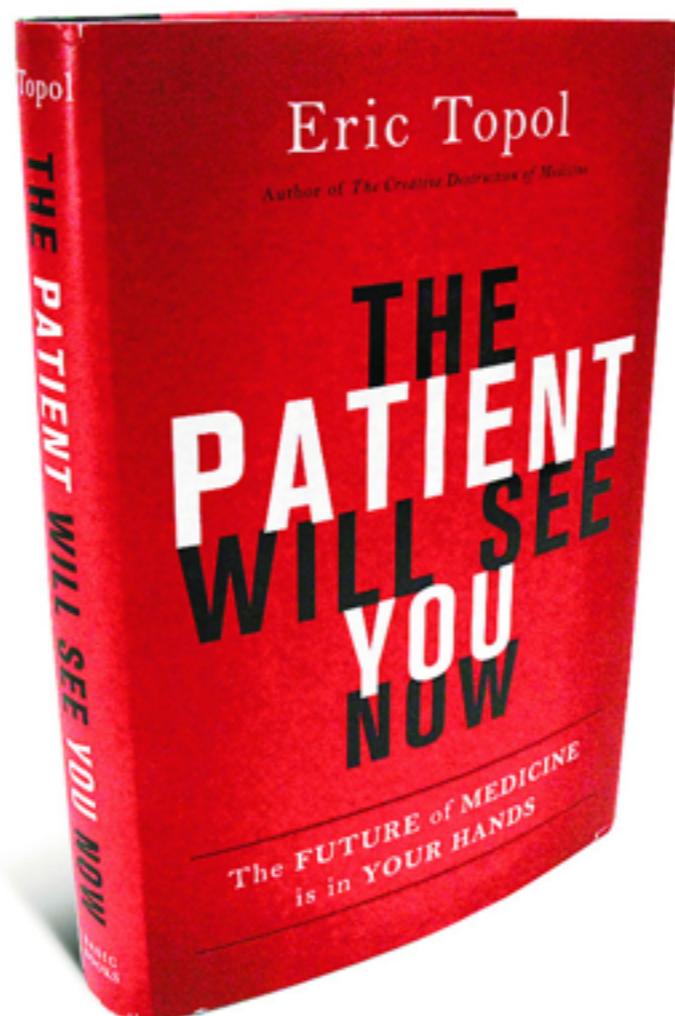


Present and future of genetic medicine (positives)



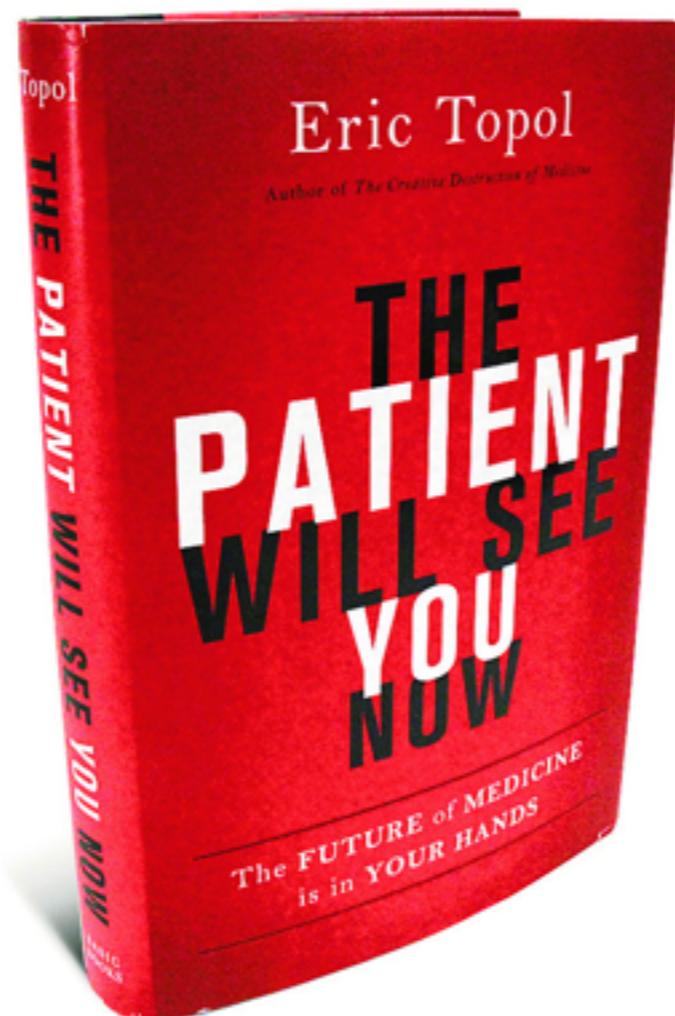
Present and future of genetic medicine (positives)

Sequencing
will be
cheap,
accessible,
and
standard practice.



Present and future of genetic medicine (positives)

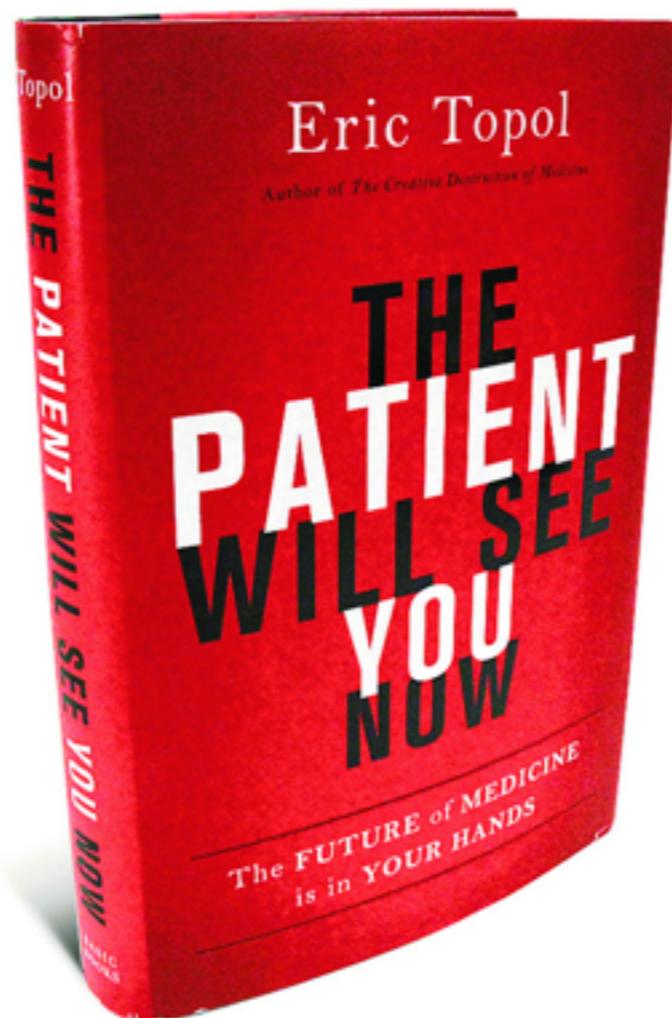
Sequencing
will be
cheap,
accessible,
and
standard practice.



With sequencing,
do we still need
genetics?

Present and future of genetic medicine (positives)

Sequencing
will be
cheap,
accessible,
and
standard practice.

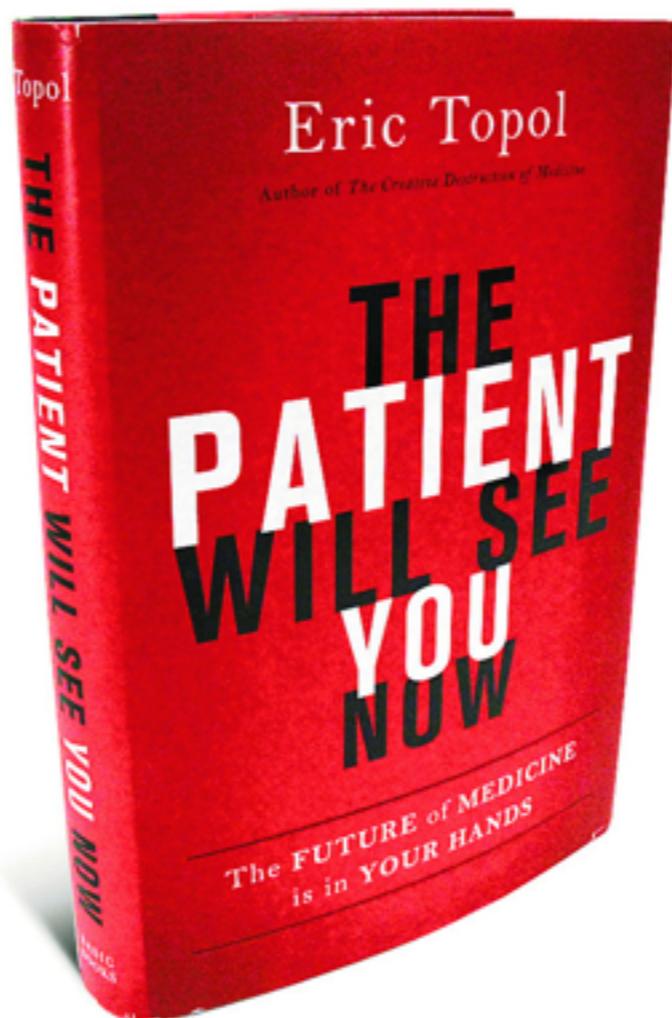


With sequencing,
do we still need
genetics?

- Rare disease causal genes found

Present and future of genetic medicine (positives)

Sequencing
will be
cheap,
accessible,
and
standard practice.

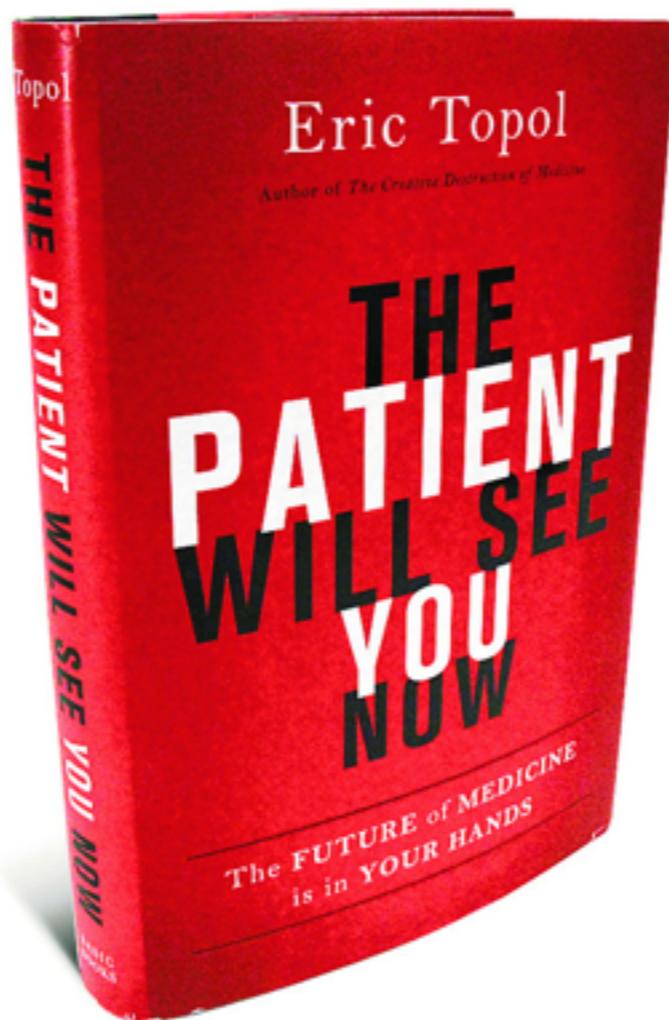


With sequencing,
do we still need
genetics?

- Rare disease causal genes found
- 12 drugs approved by FDA with genetic test since 2012

Present and future of genetic medicine (positives)

Sequencing
will be
cheap,
accessible,
and
standard practice.

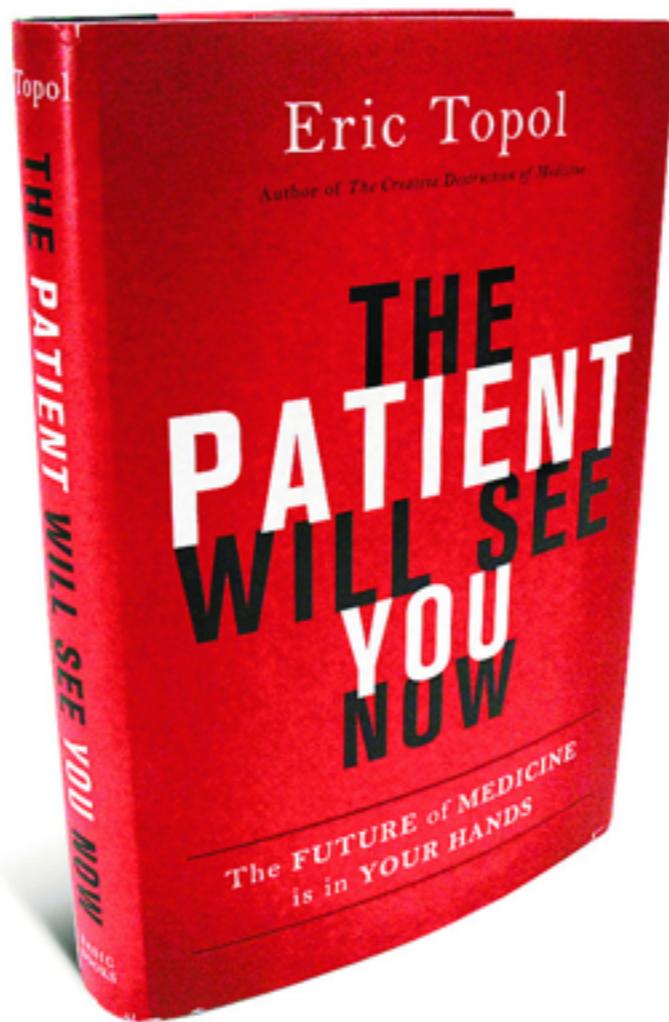


With sequencing,
do we still need
genetics?

- 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

Present and future of genetic medicine (positives)

Sequencing
will be
cheap,
accessible,
and
standard practice.

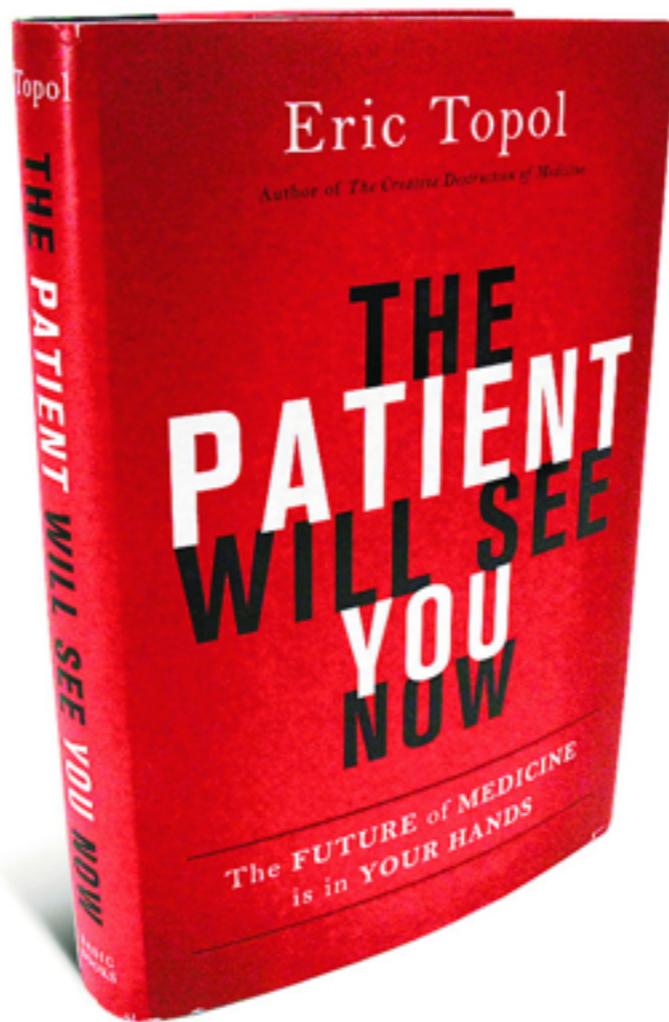


With sequencing,
do we still need
genetics?

- 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

Present and future of genetic medicine (positives)

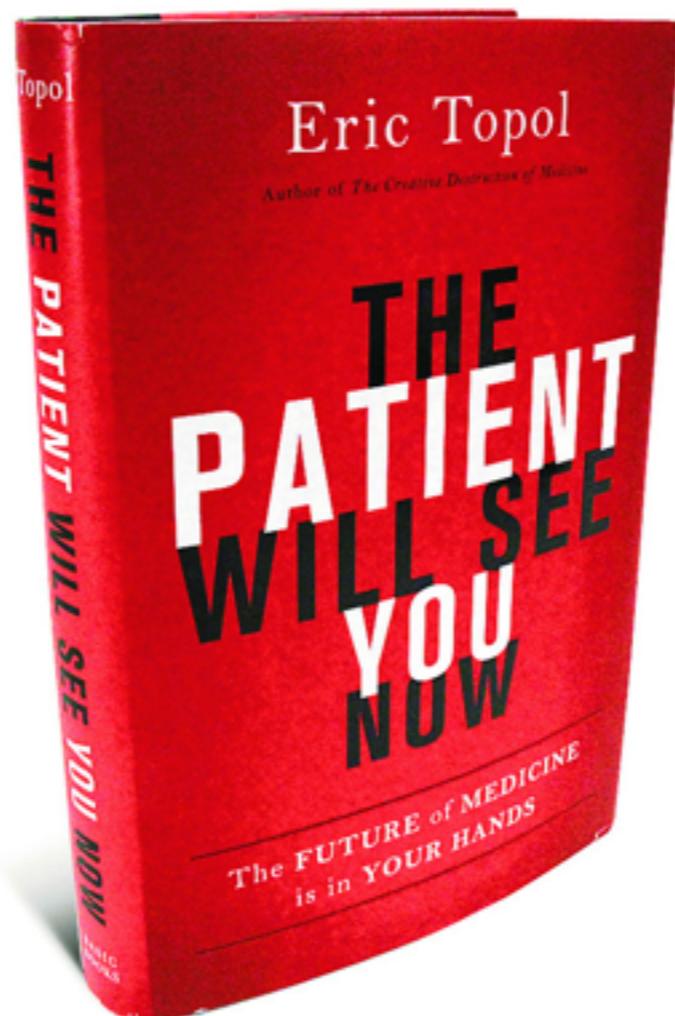
Sequencing
will be
cheap,
accessible,
and
standard practice.



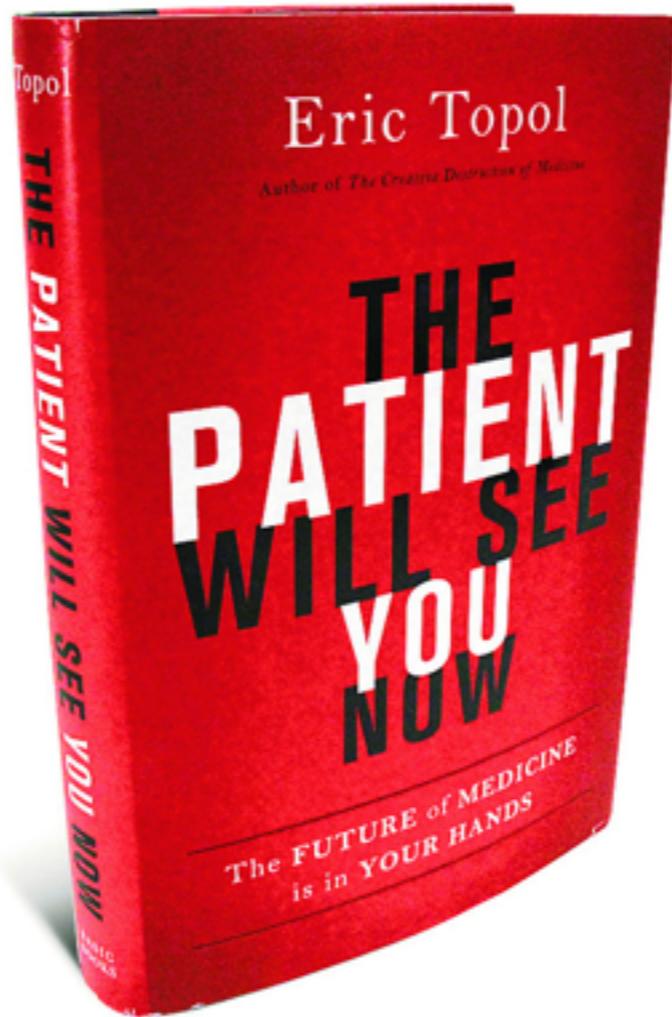
With sequencing,
do we still need
genetics?

- 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)

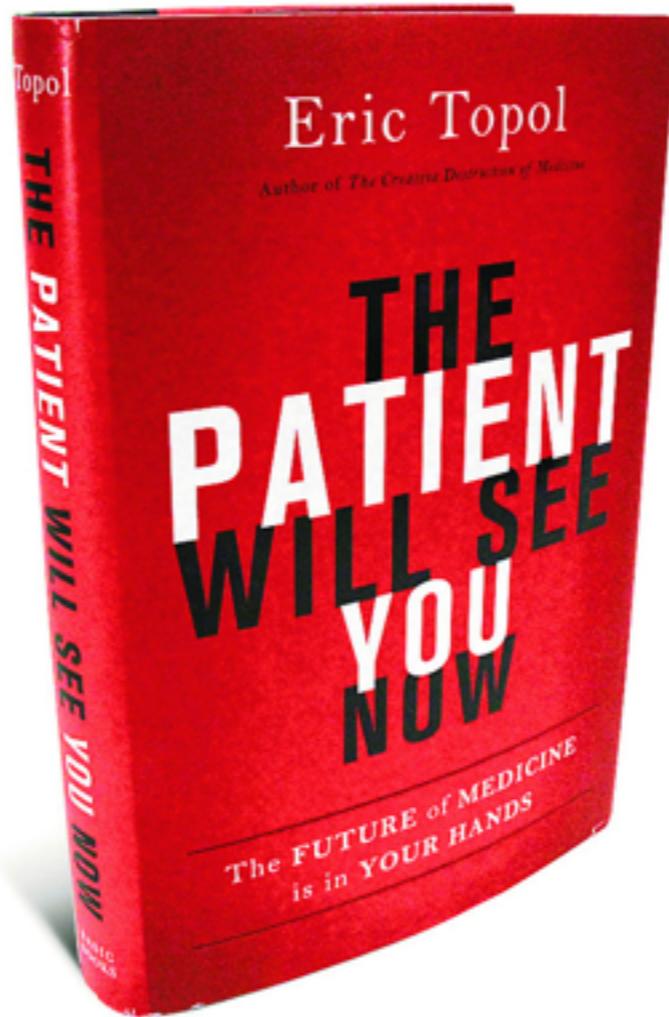


Present and future of genetic medicine (negatives)



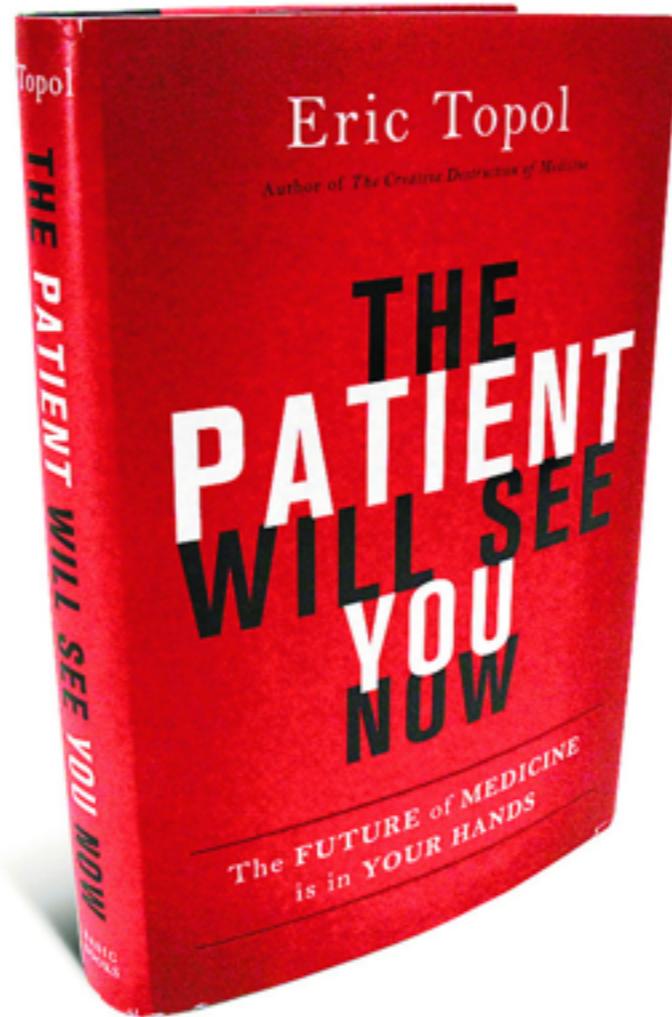
- Lots of taxpayers' money spent on little valuable data

Present and future of genetic medicine (negatives)



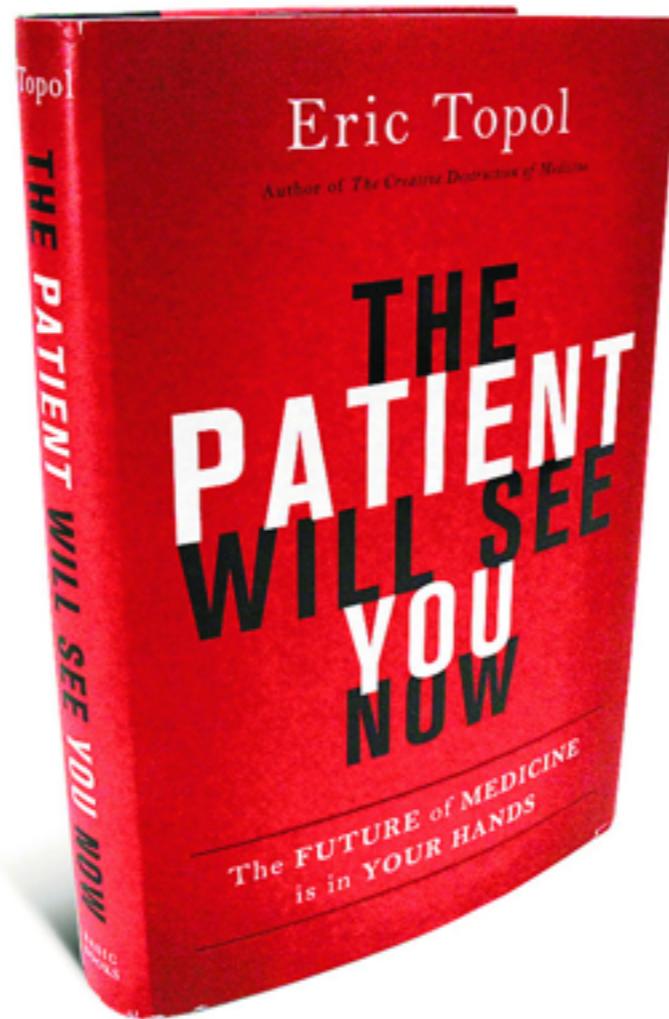
- Lots of taxpayers' money spent on little valuable data
- Genotype data are being sold

Present and future of genetic medicine (negatives)



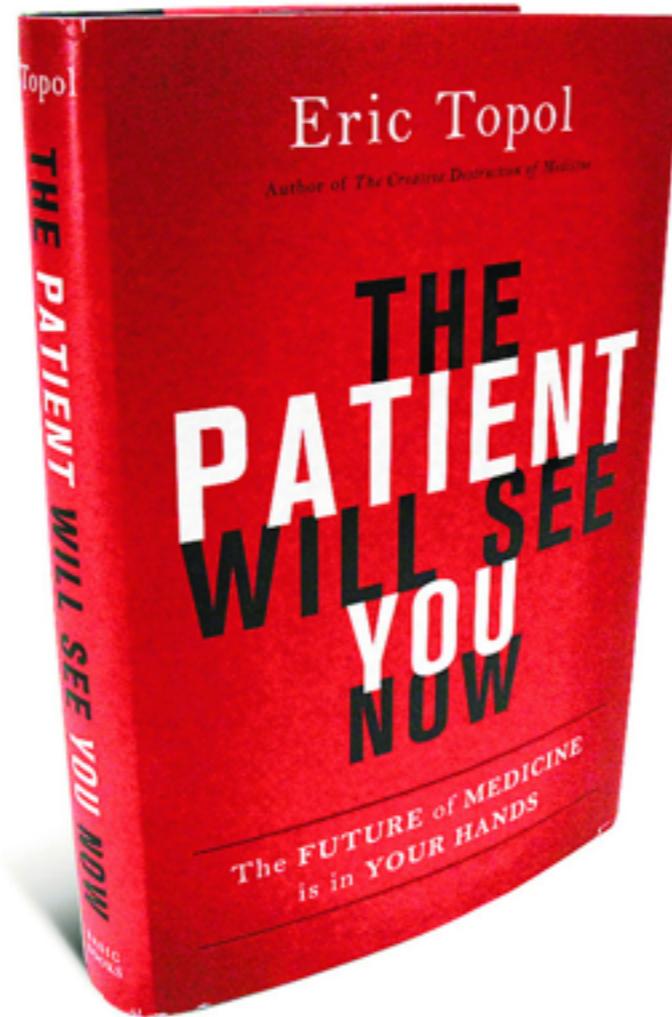
- Lots of taxpayers' money spent on little valuable data
- Genotype data are being sold
- Genotype data are being evaluated by insurance companies

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

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 18