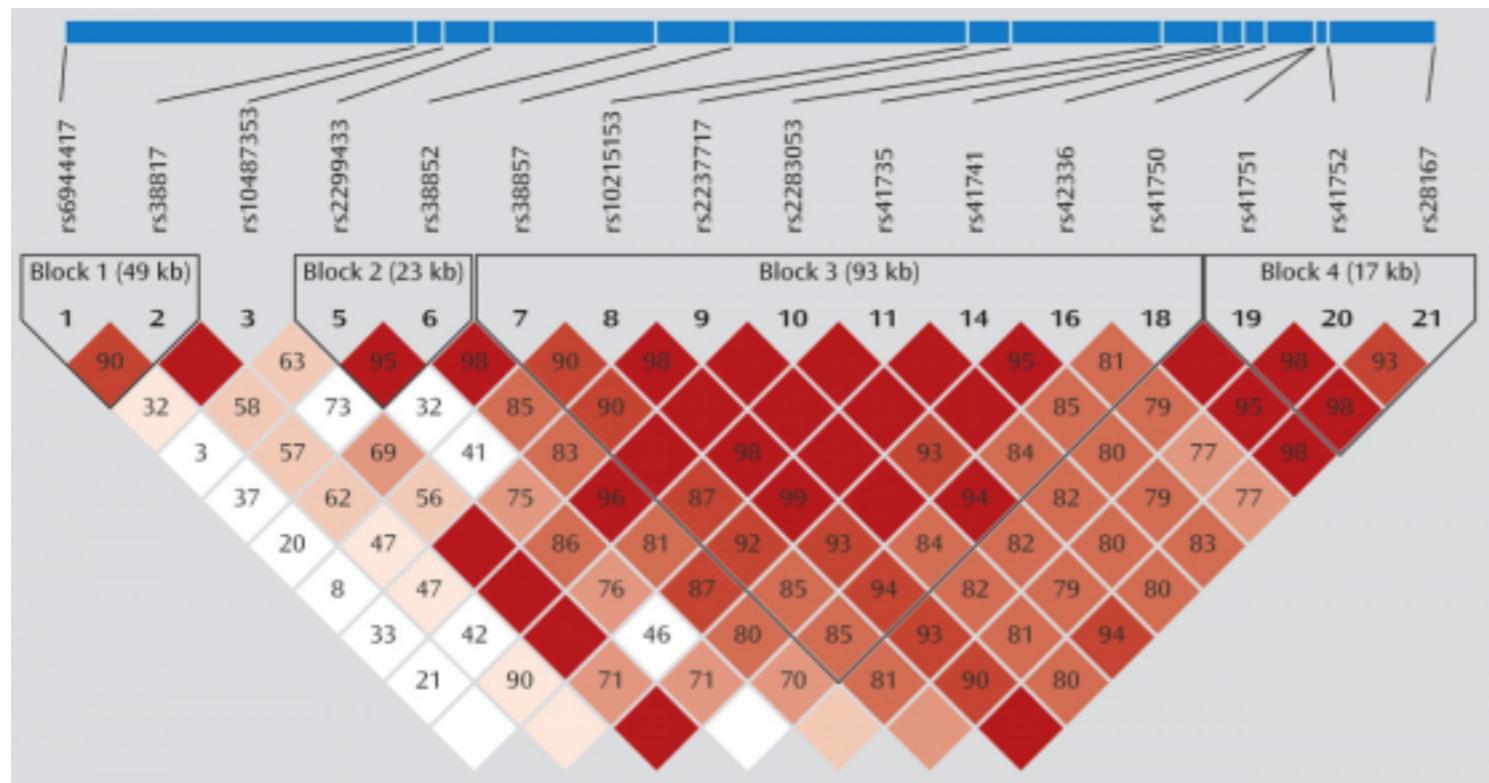
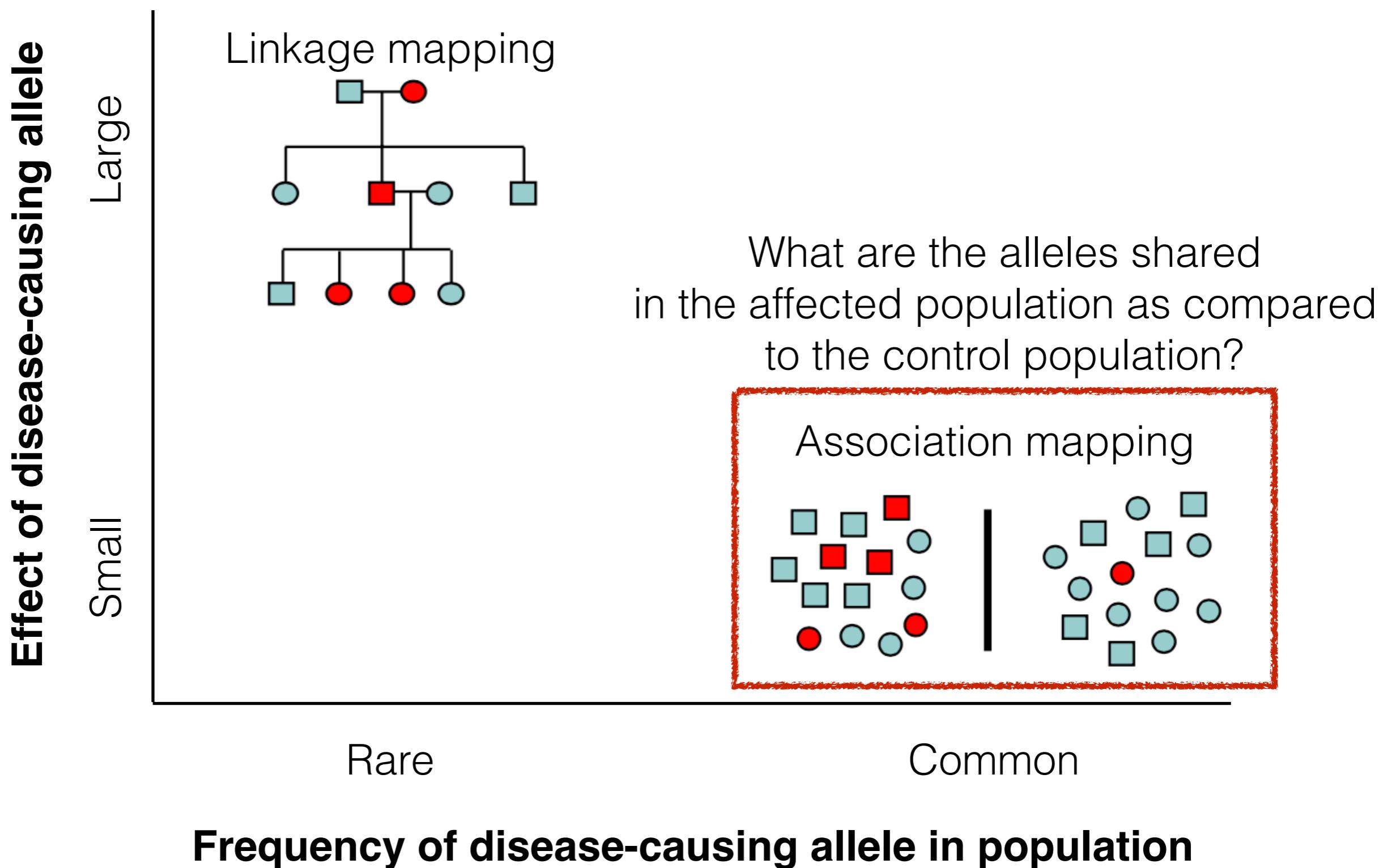


Linkage disequilibrium, haplotypes, and GWAS



Human gene mapping has two general flavors



Correlation between marker and disease-causing allele drastically affects how well mappings will work

Big haplotype blocks (long-range LD) = coarse mapping

Small haplotype blocks (little LD) = fine mapping



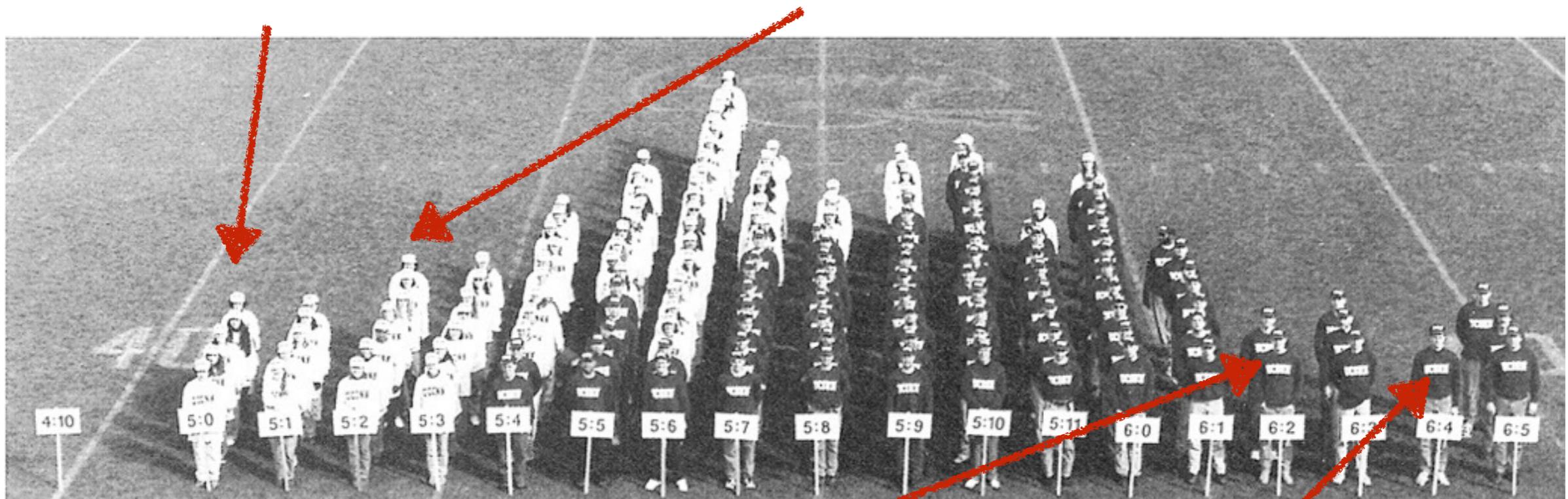
vs.



How many people need to be genotyped?

Genome-wide association studies measure correlation between “tag-SNV” and disease-causing allele

CAGCGATAGGCTTAATGTT	CAGCGATAGGCTTAATGTT
AGCCC GTTT <ins>T</ins> ATGACCAACG	AGCCC GTTT <ins>T</ins> ATGACCAACG
GGGTTCACAGTGAGCTGTGT	GGGTTCACAGTGAGCTGTGT



University of Connecticut, 1997

CAGCGATAGGCTTAATGTT
AGCCC GTTT <ins>G</ins> ATGACCAACG
GGGTTCACAGTGAGCTGTGT

CAGCGATAGGCTTAATGTT
AGCCC GTTT <ins>G</ins> ATGACCAACG
GGGTTCACAGTGAGCTGTGT

Common polymorphisms facilitate genome-wide association (GWA) mapping



The Human Haplotype Map (HapMap) identified
10 million common polymorphisms

LD blocks in humans are 20-100 kb

500,000 common variants gives us a SNV every 10 kb

2-10 SNV mark each LD block for the statistical test

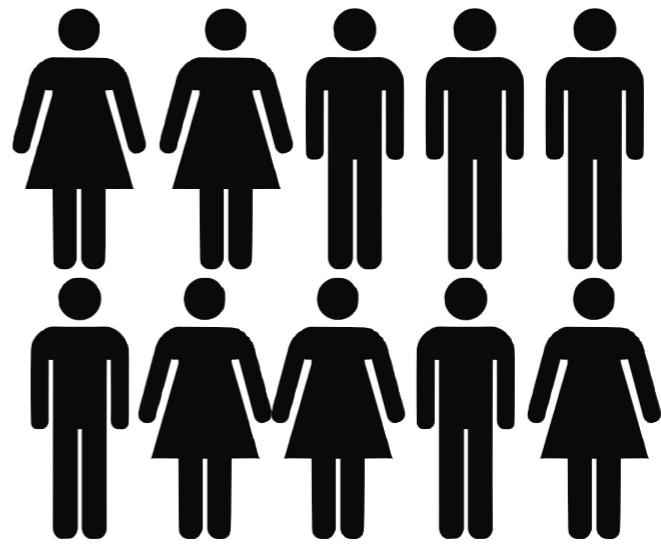
Why do 4.3M SNV tests on current arrays?

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

Case-control study design



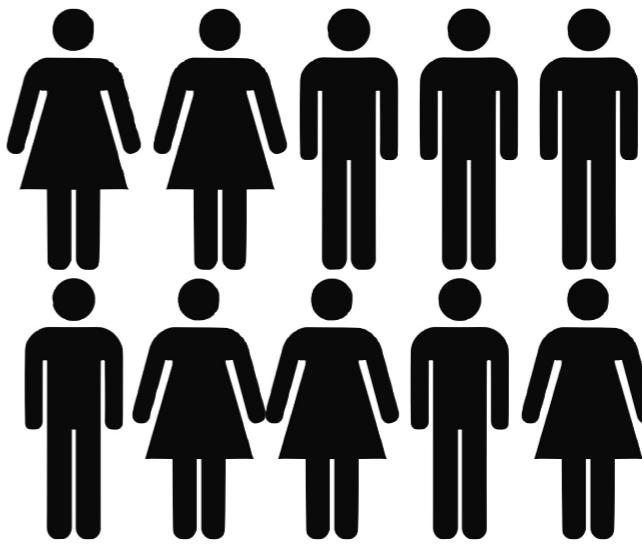
Cases
(People with trait)



Controls
(People without trait)

What alleles do the cases share that the controls lack?

Collect genotype and phenotype data for lots of people

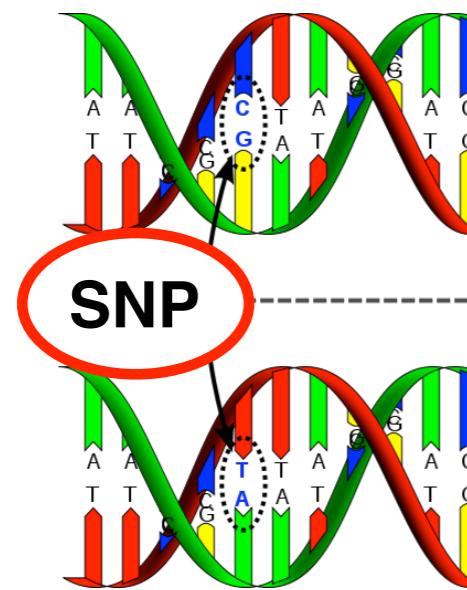


Genotype: SNV arrays (>500k) or sequencing

Phenotype: Measure quantitative values

\$250 million spent since 2006 on GWAS

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

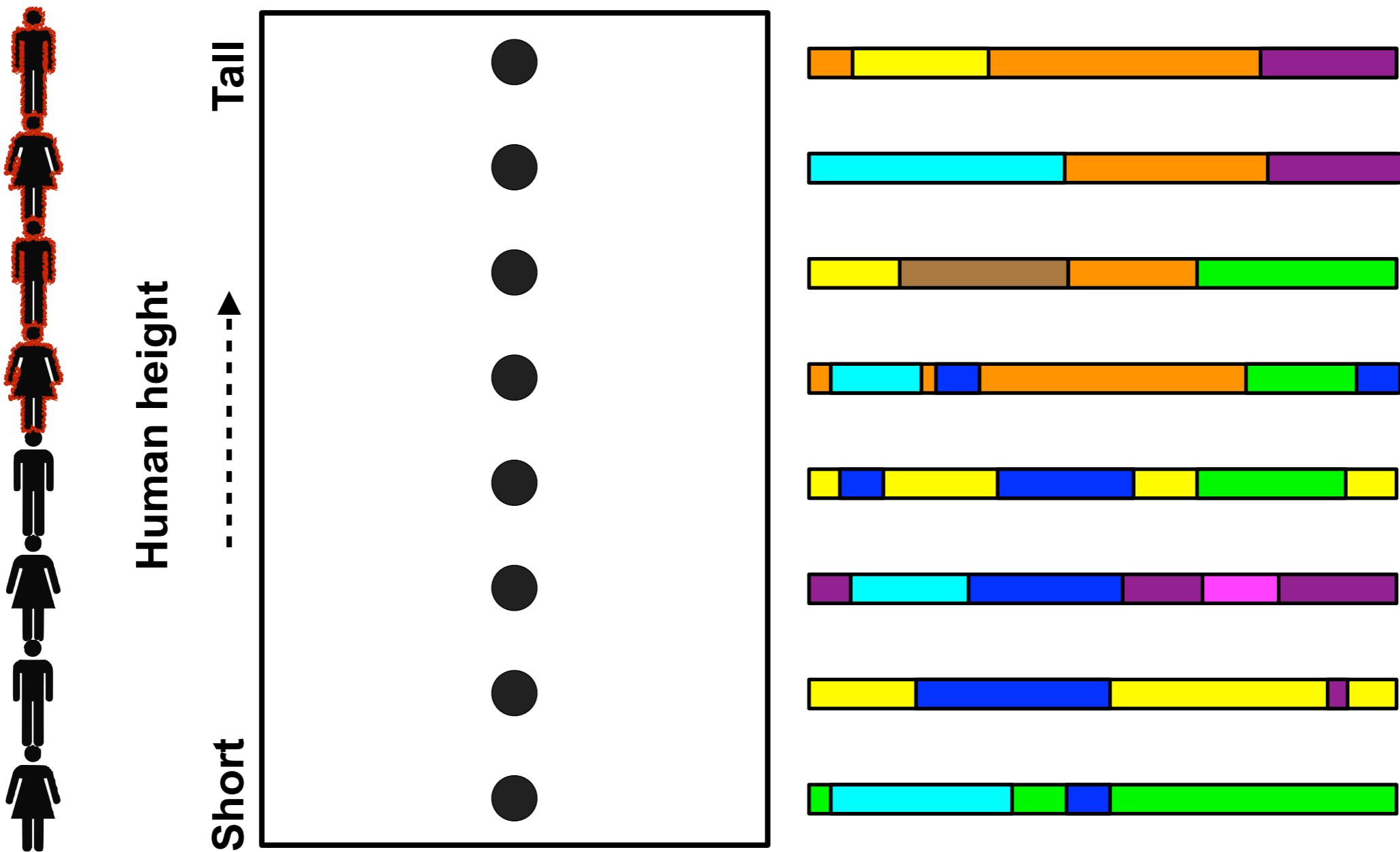


Genetic variation

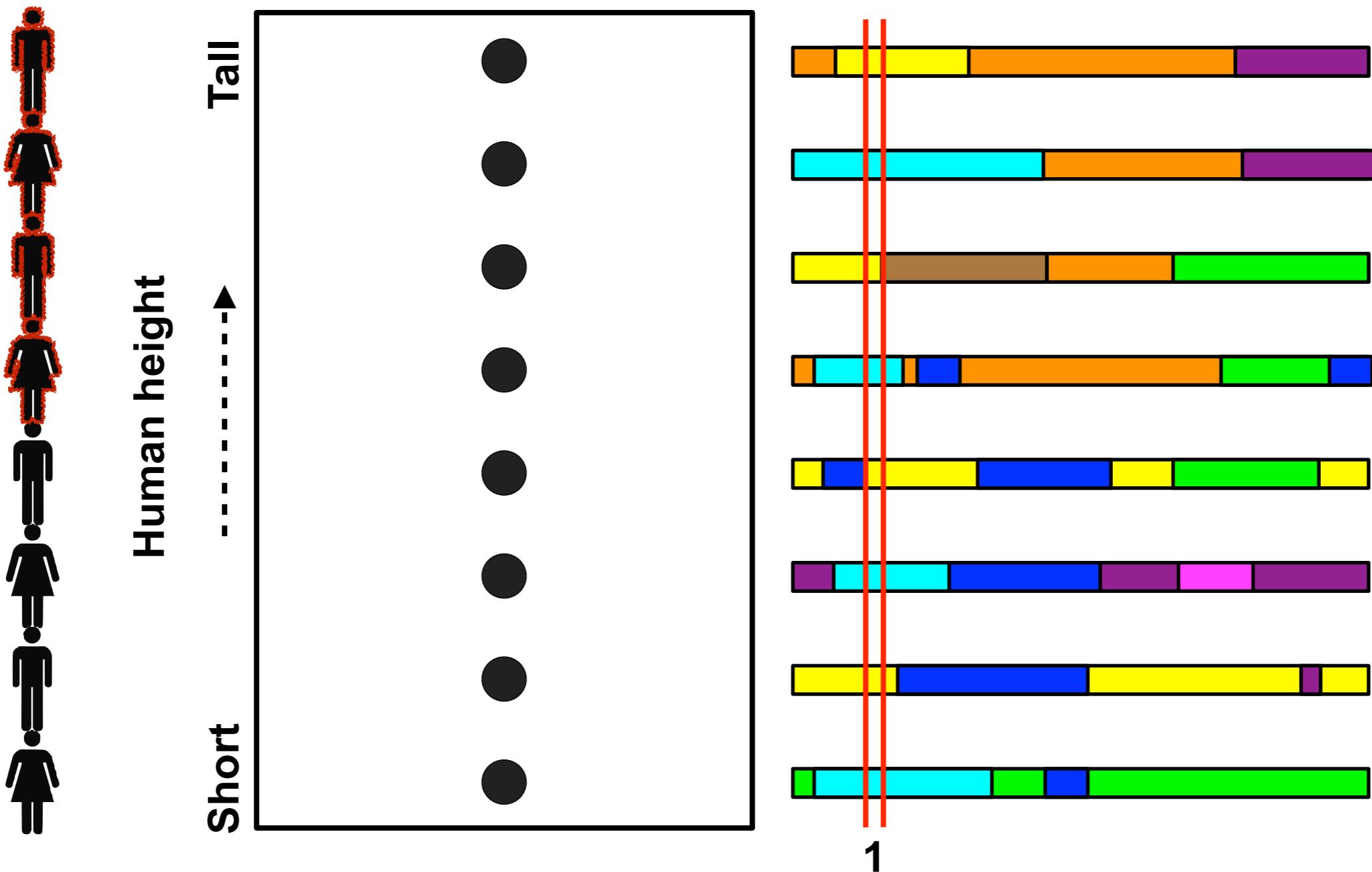


Phenotypic variation

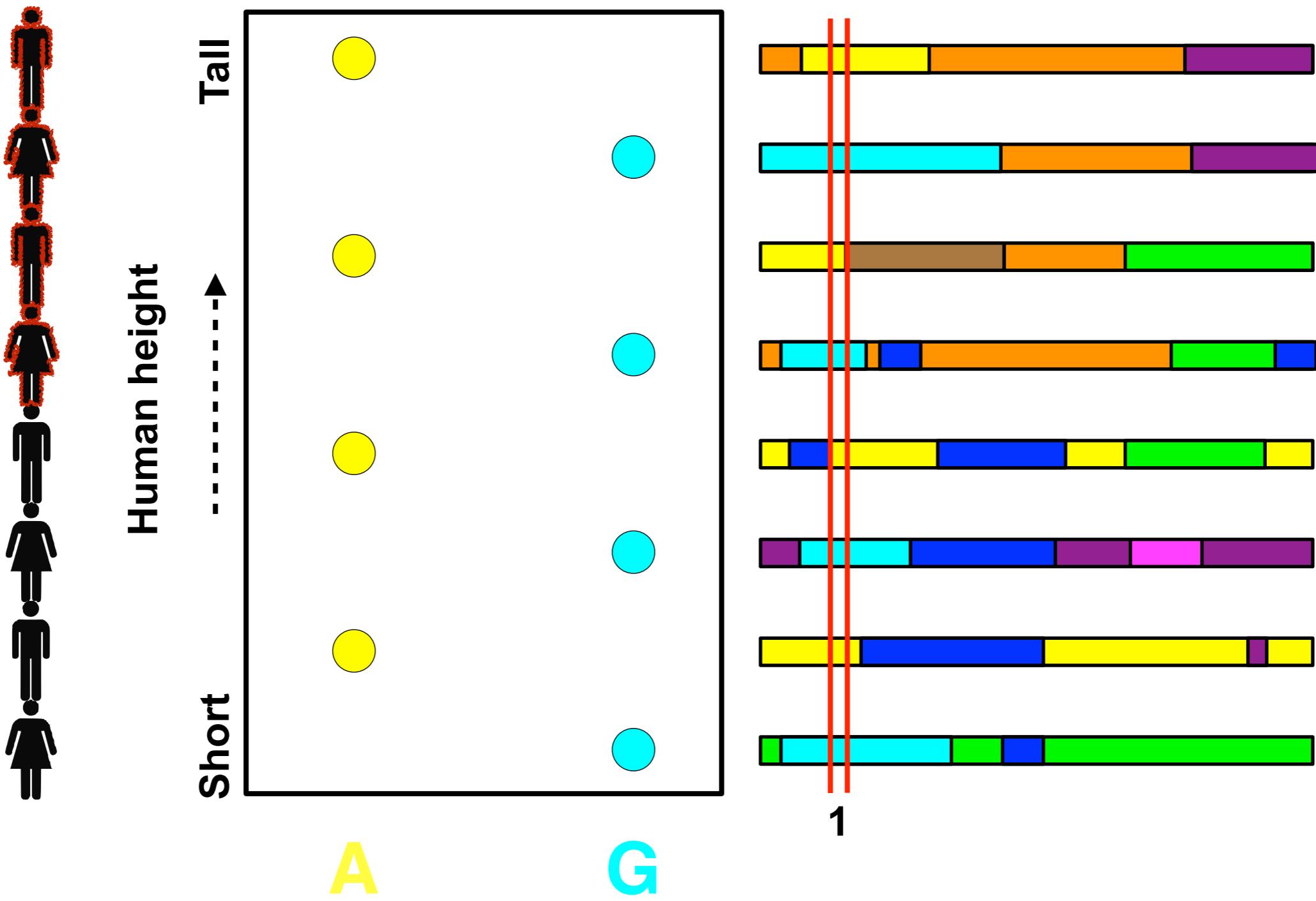
Association mapping: Correlating genotype with phenotype



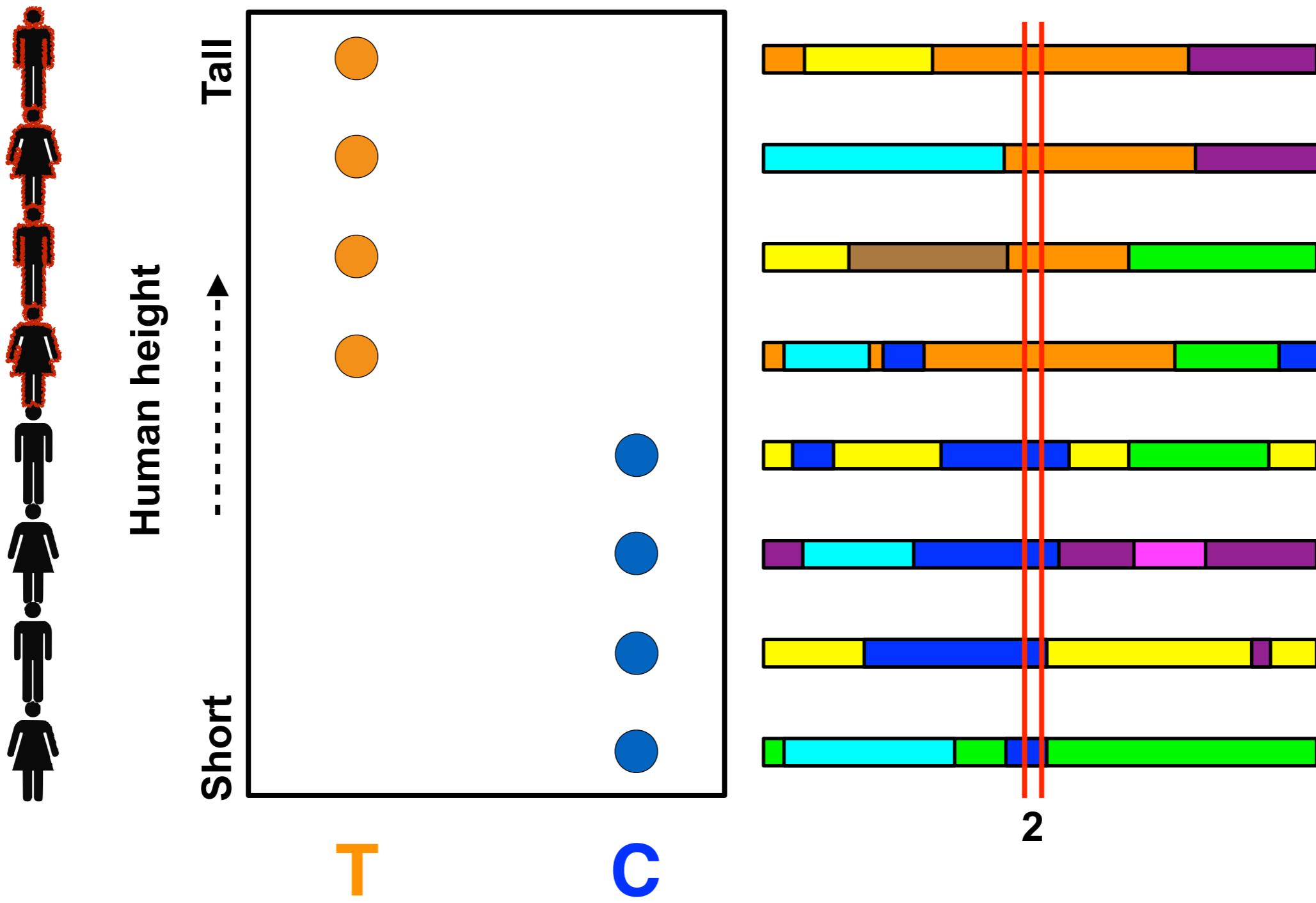
Association mapping: Correlating genotype with phenotype



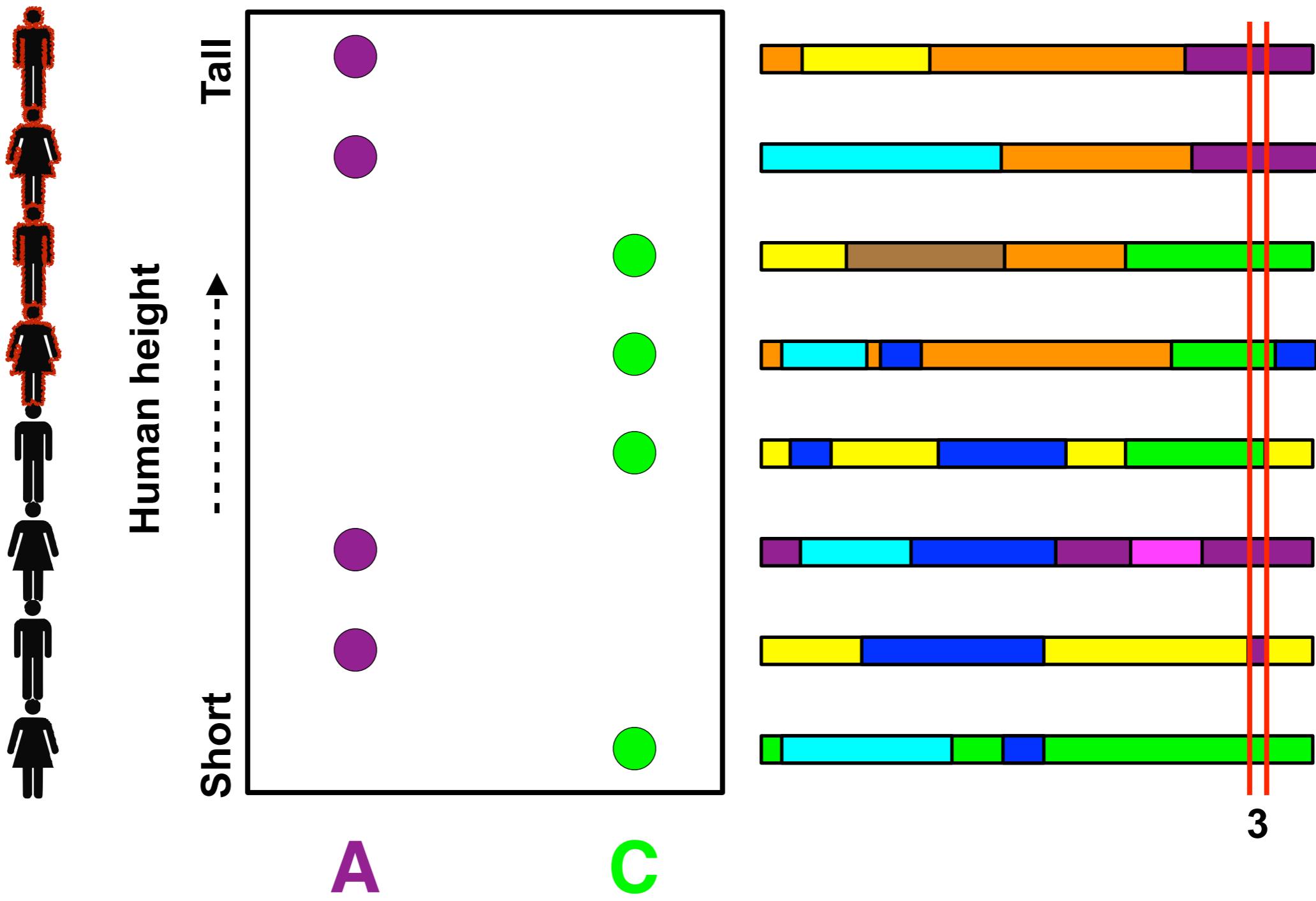
Association mapping: Correlating genotype with phenotype



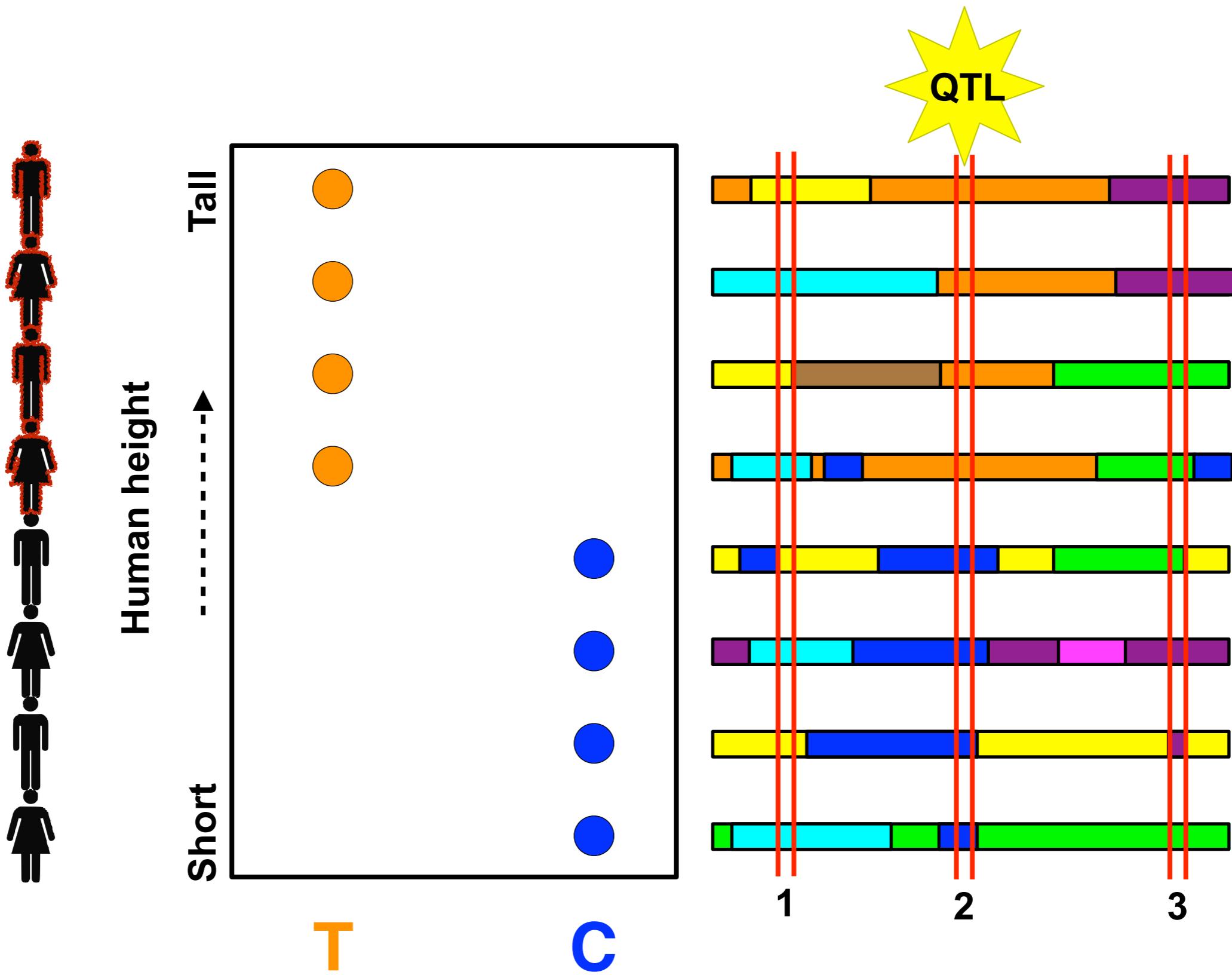
Association mapping: Correlating genotype with phenotype



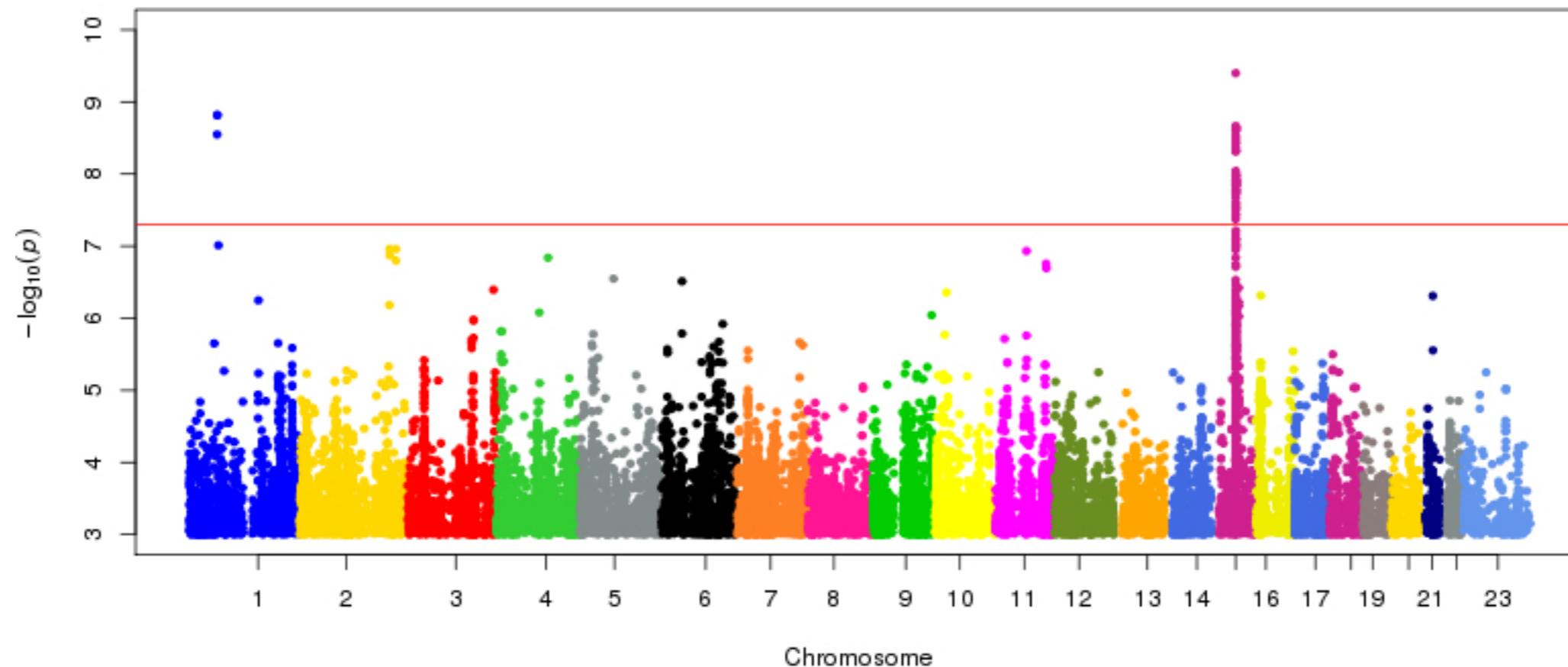
Association mapping: Correlating genotype with phenotype



Association mapping: Correlating genotype with phenotype

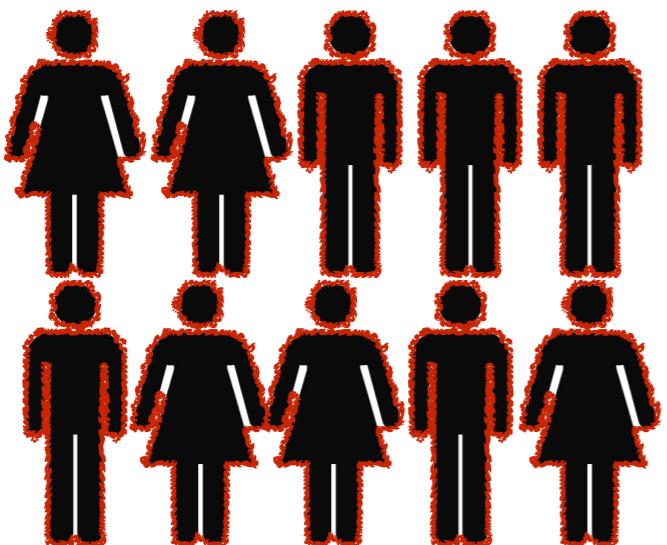


An example Manhattan plot of GWA mapping results



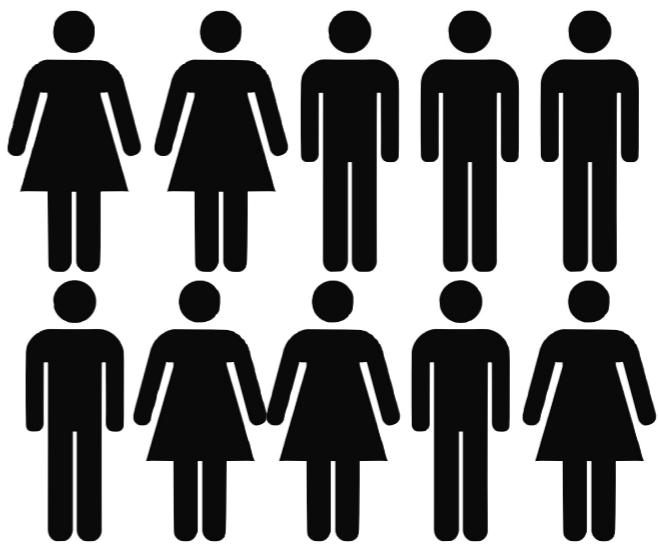
Styrkarsdottir *et al.* Nature 2014

GWAS calculation



4000 Cases

SNV1
(G or A) 4000 of 8000 (50% G)



6000 Controls

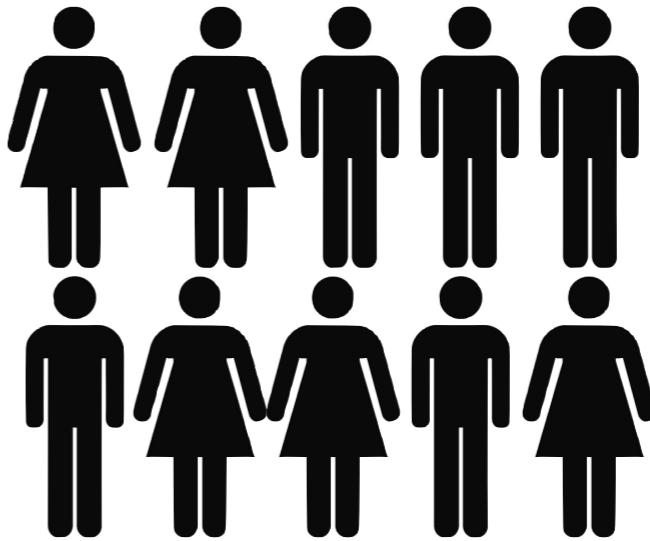
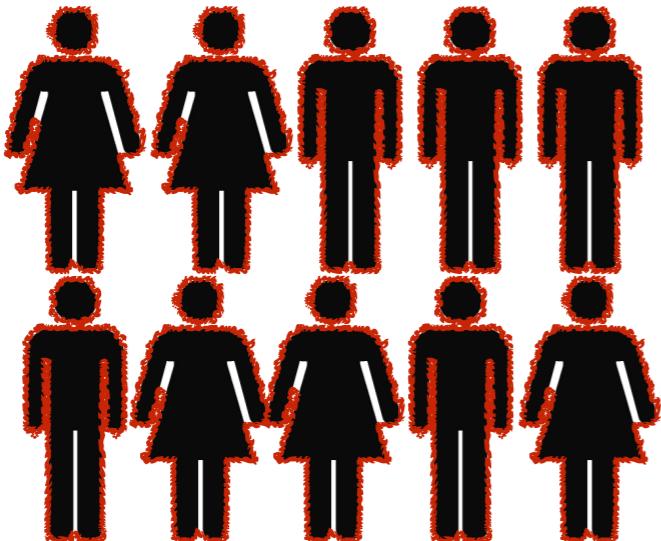
5000 of 12000 (42% G)

	Cases	Controls
G	4000	5000
A	4000	7000

Observed

Expected

GWAS calculation



SNV1
(G or A) 4000 of 8000 (50% G)

5000 of 12000 (42% G)

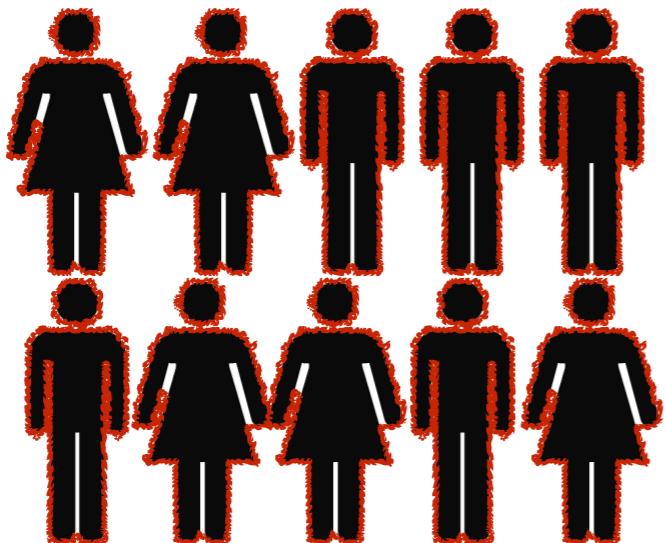
	Cases	Controls
G	4000	5000
A	4000	7000

Observed Expected

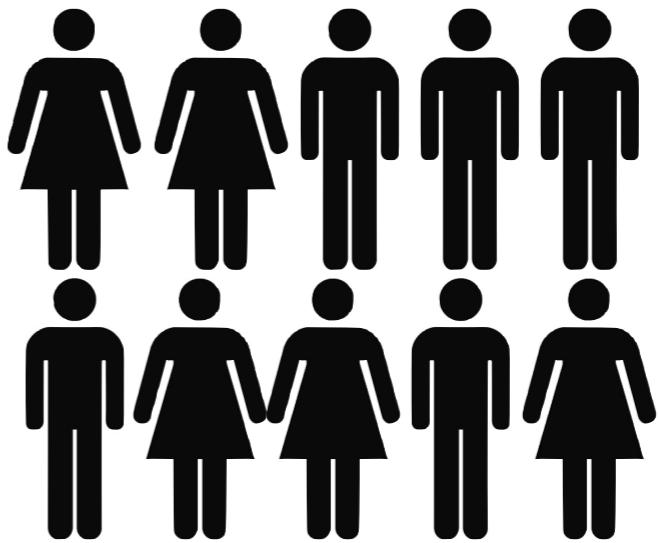
Pearson's chi-squared test
with one degree of freedom

67.0038 or p-value of 2.71e-16

GWAS calculation



4000 Cases



6000 Controls

SNV1
(G or A) 4000 of 8000 (50% G)

5000 of 12000 (42% G)

SNV2
(T or C) 3200 of 8000 (40% T)

4600 of 12000 (38% T)

	Cases	Controls
T	3200	4600
C	4800	7400

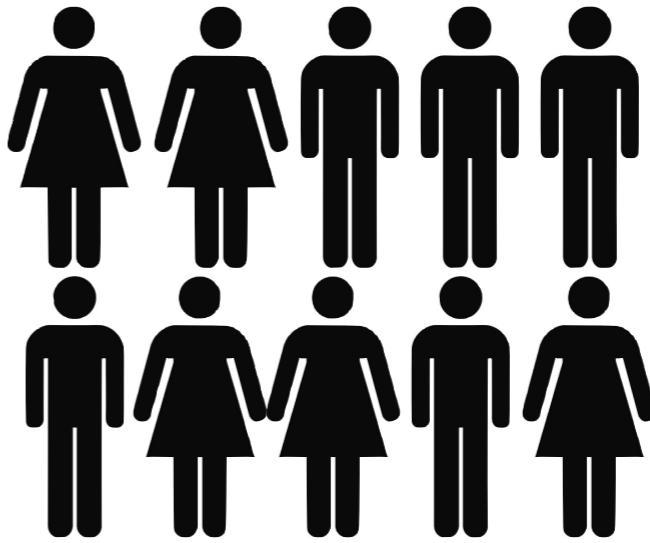
Observed

Expected

GWAS calculation



4000 Cases



6000 Controls

SNV1
(G or A) 4000 of 8000 (50% G)

5000 of 12000 (42% G)

SNV2
(T or C) 3200 of 8000 (40% T)

4600 of 12000 (38% T)

	Cases	Controls
T	3200	4600
C	4800	7400

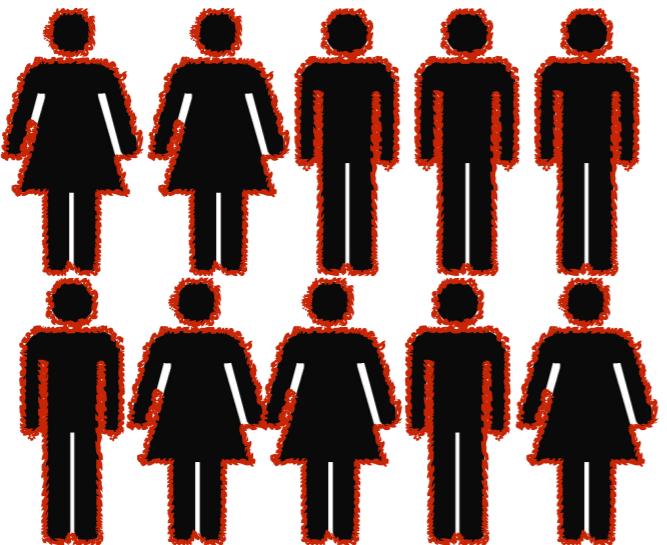
Observed

Expected

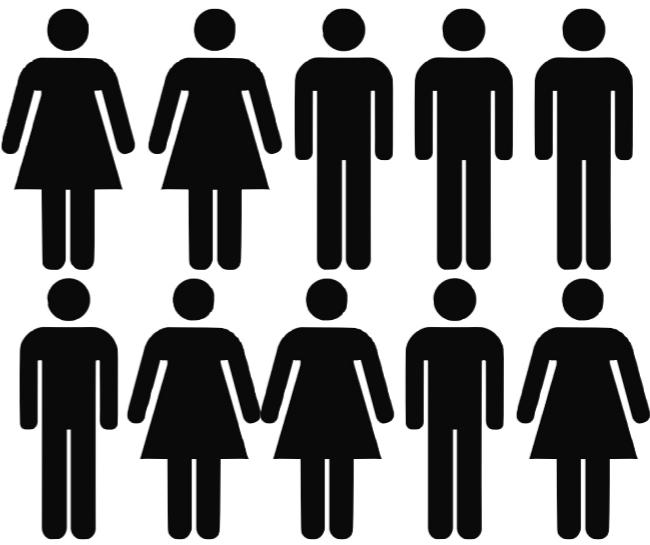
Pearson's chi-squared test
with one degree of freedom

2.7327 or p-value of 0.09831

GWAS calculation



4000 Cases



6000 Controls

SNV1
(G or A) 4000 of 8000 (50% G)

5000 of 12000 (42% G)

SNV2
(T or C) 3200 of 8000 (40% T)

4600 of 12000 (38% T)

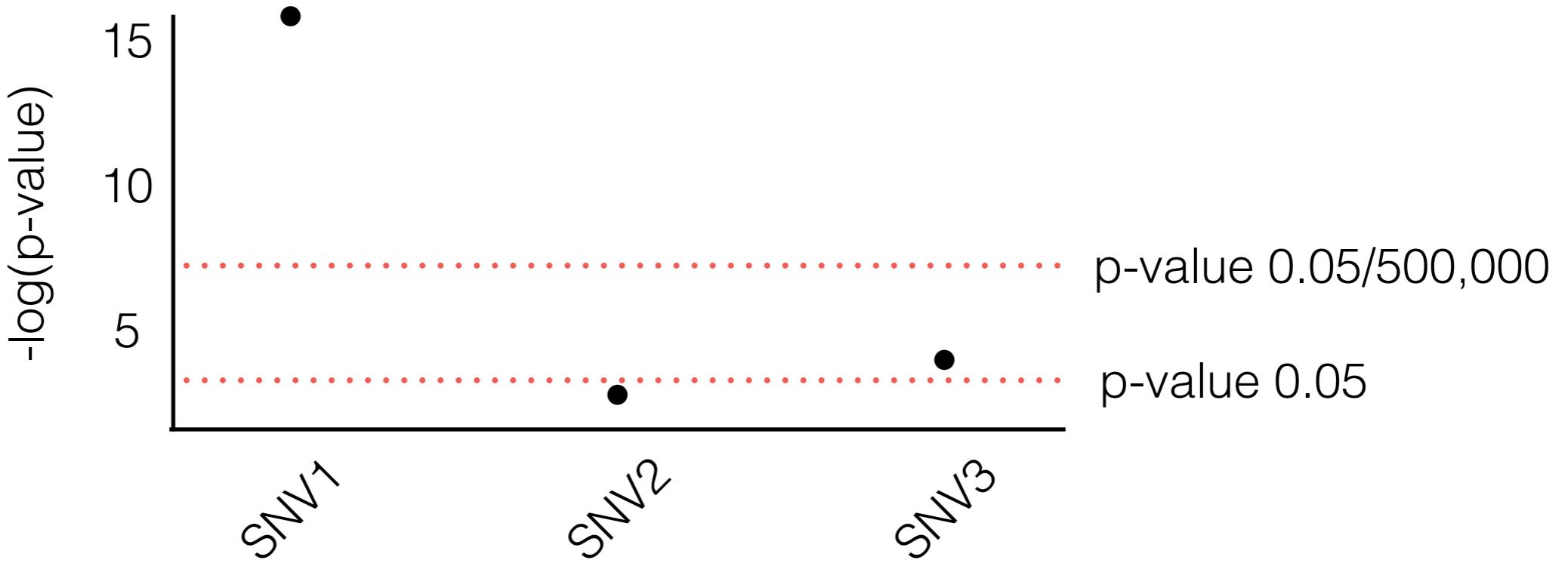
SNV3
(C or A) 3600 of 8000 (45% C)

5000 of 12000 (40% C)

	Cases	Controls
C	3600	5000
A	4400	7000

10.7443 or p-value of 0.001046

GWAS results



500,000 SNVs 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 SNVs by chance

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

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)

GWA mapping within groups and replication



GWA mapping works best within a related population

The mapping *might* be replicated in different populations