Pop Gen meets Quant Gen

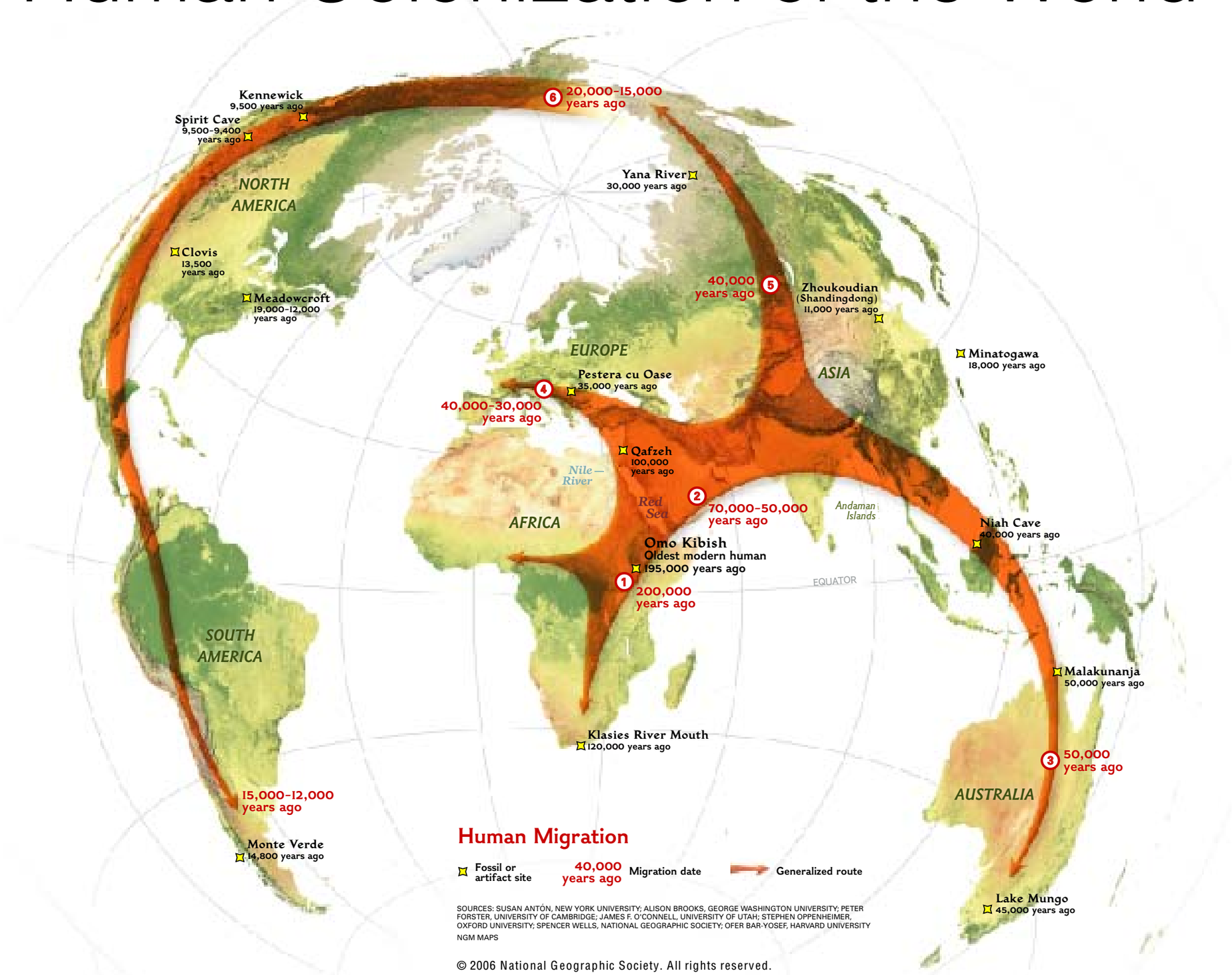
Ryan D. Hernandez

ryan.hernandez@ucsf.edu

Modern Human Genomics

http://www.finca.org

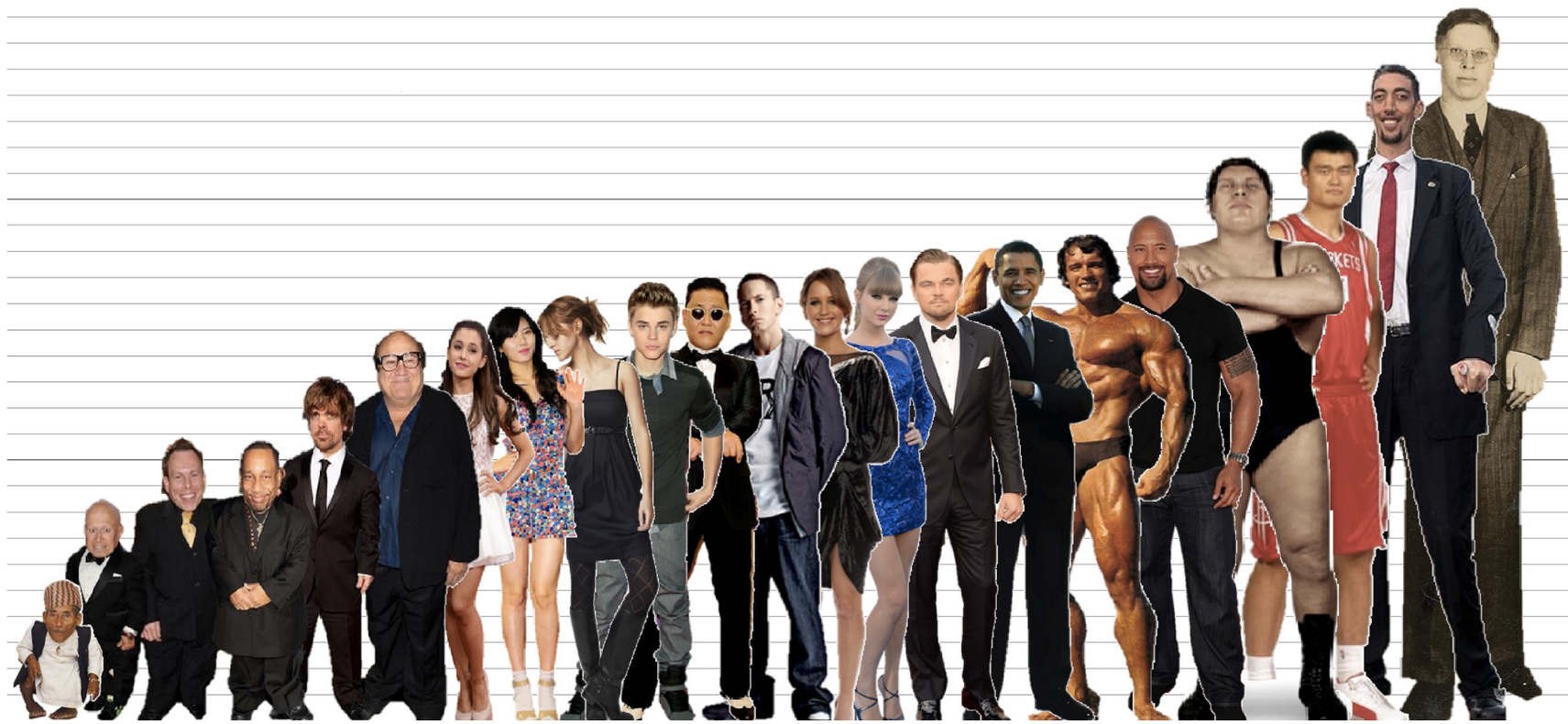
Human Colonization of the World



http://ngm.nationalgeographic.com

Heritability and Human Height

Studies of heritability ask questions such as how much genetic factors play a role in **differences in height between people**. This is not the same as asking **how much genetic factors influence height in any one person**.



https://en.wikipedia.org/wiki/Heritability

http://i.ytimg.com/vi/E0Aeks_id6c/maxresdefault.jpg

Heritability

- $V_P = V_G + V_E$
    - Variance in a phenotype = variance in genotypes + variance in environment

Heritability

- $V_P = V_A + V_D + V_I + V_E$
  - Genetic variance decomposed into several **variance components**.
- V_A is the additive component:
  - The sum of the average effects of all the genes an individual carries.
  - If an individual mated to a number of individuals taken at random from the population, then the additive effect is twice the mean deviation of the progeny from the population mean.

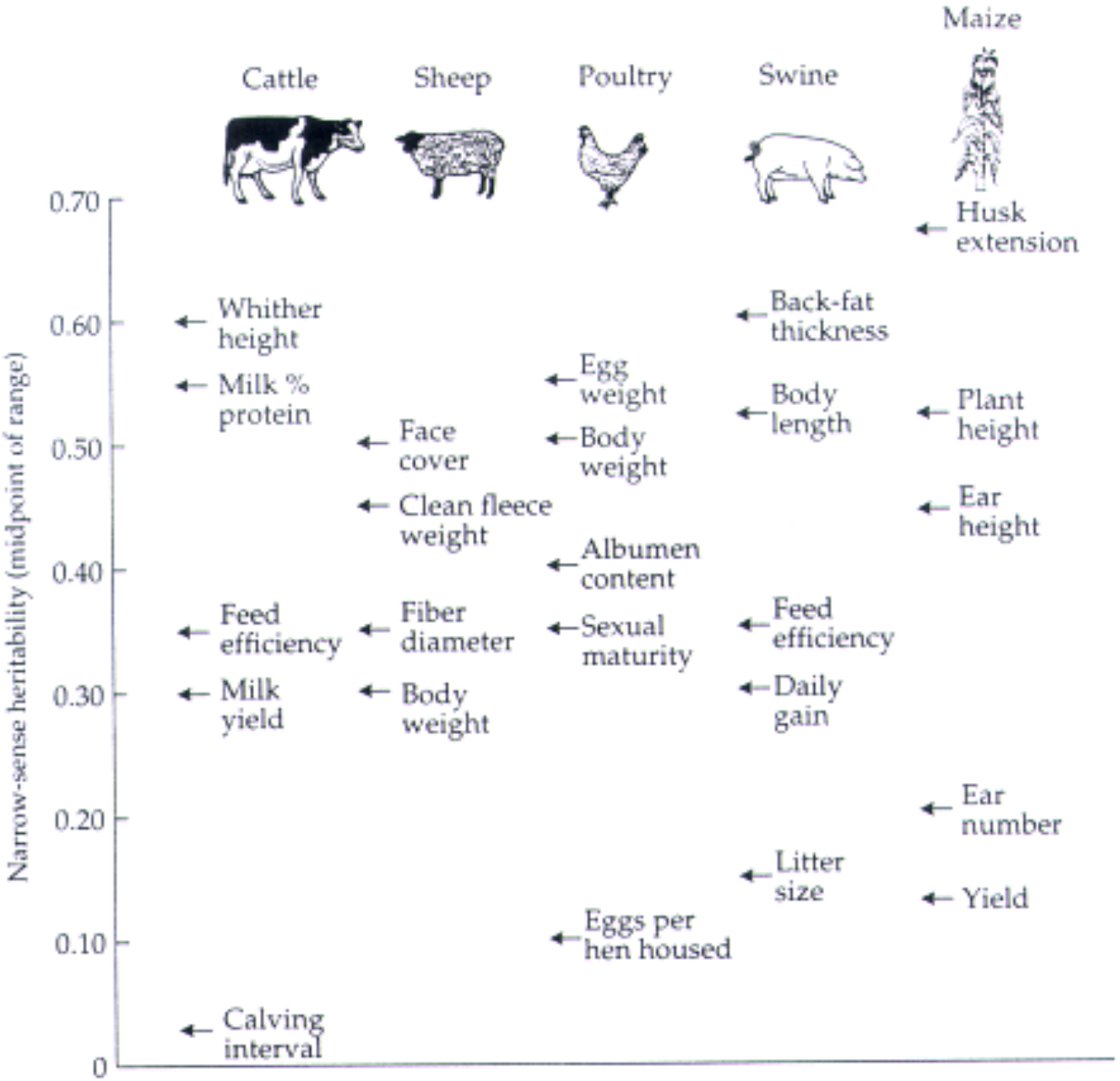


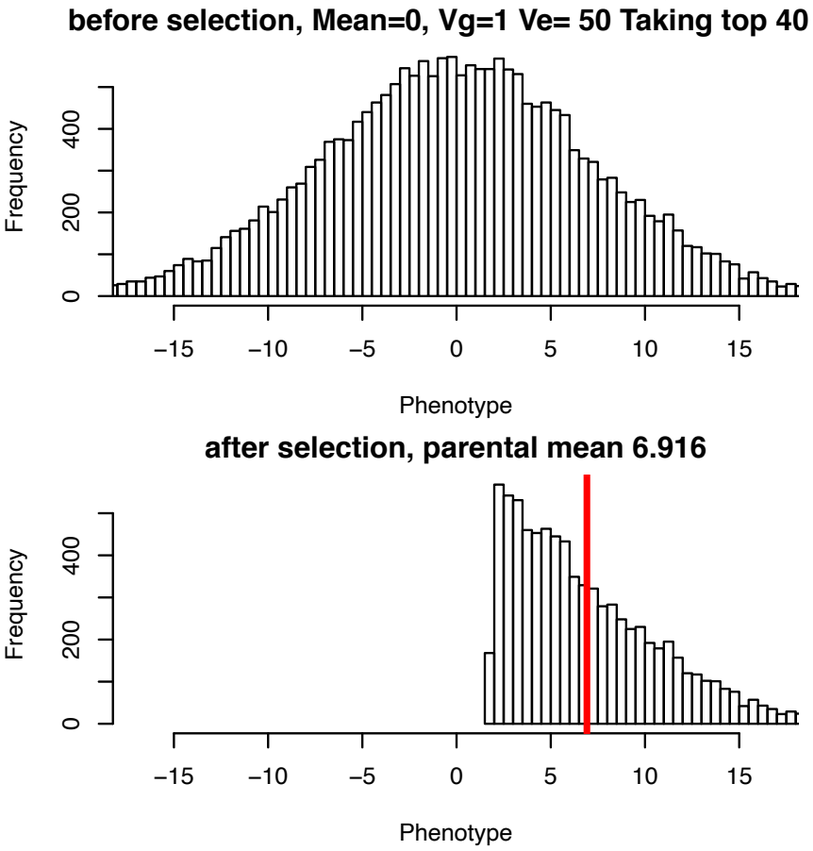
Figure 9.17 Narrow-sense heritabilities for representative traits in plants and animals. Traits closely related to fitness (calving interval, eggs per hen, litter size of swine, yield and ear number of corn) tend to have rather low heritabilities. (Animal data from Pirchner 1969, who gives the range of heritabilities in various studies. The midpoint of the range is plotted here. Corn data from Robinson et al. 1949.)

Truncation Selection

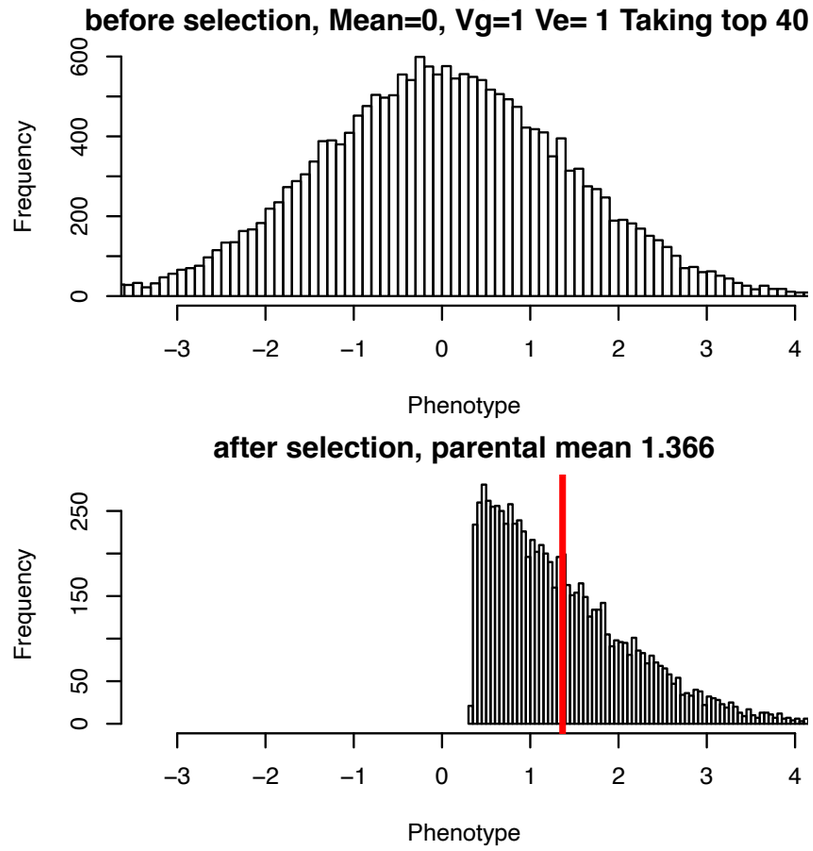
- Imagine a trait with 100 loci contributing.
- Suppose on the top X% of individuals are able to mate.
- How would the phenotype evolve in 1 generation?

Truncation Selection

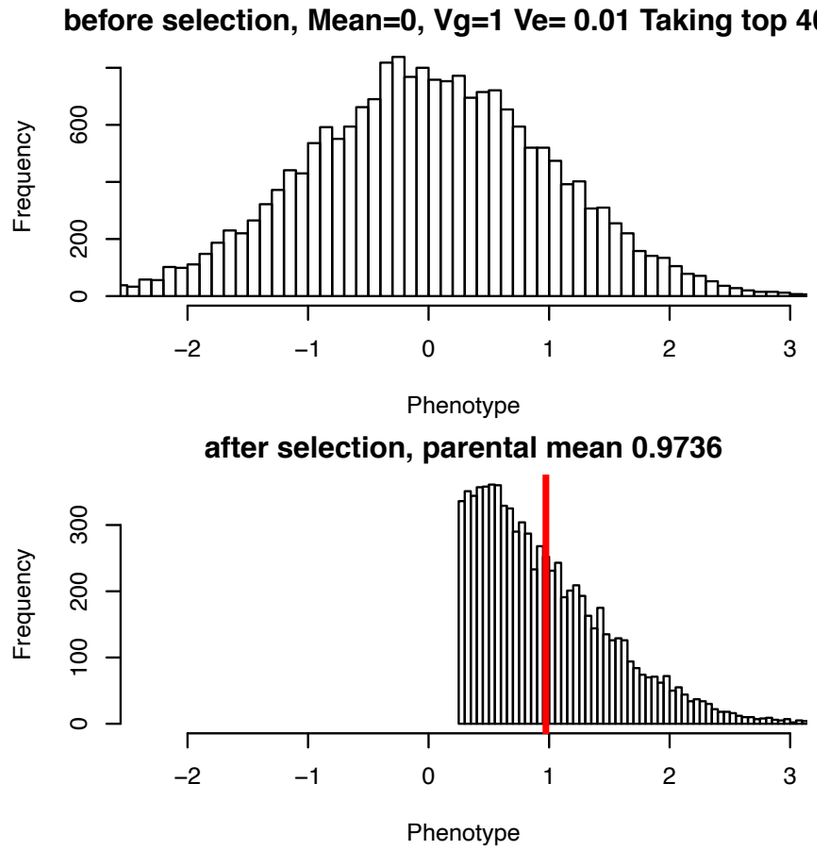
$h^2 \approx 2\,\%$



$h^2 = 50\,\%$

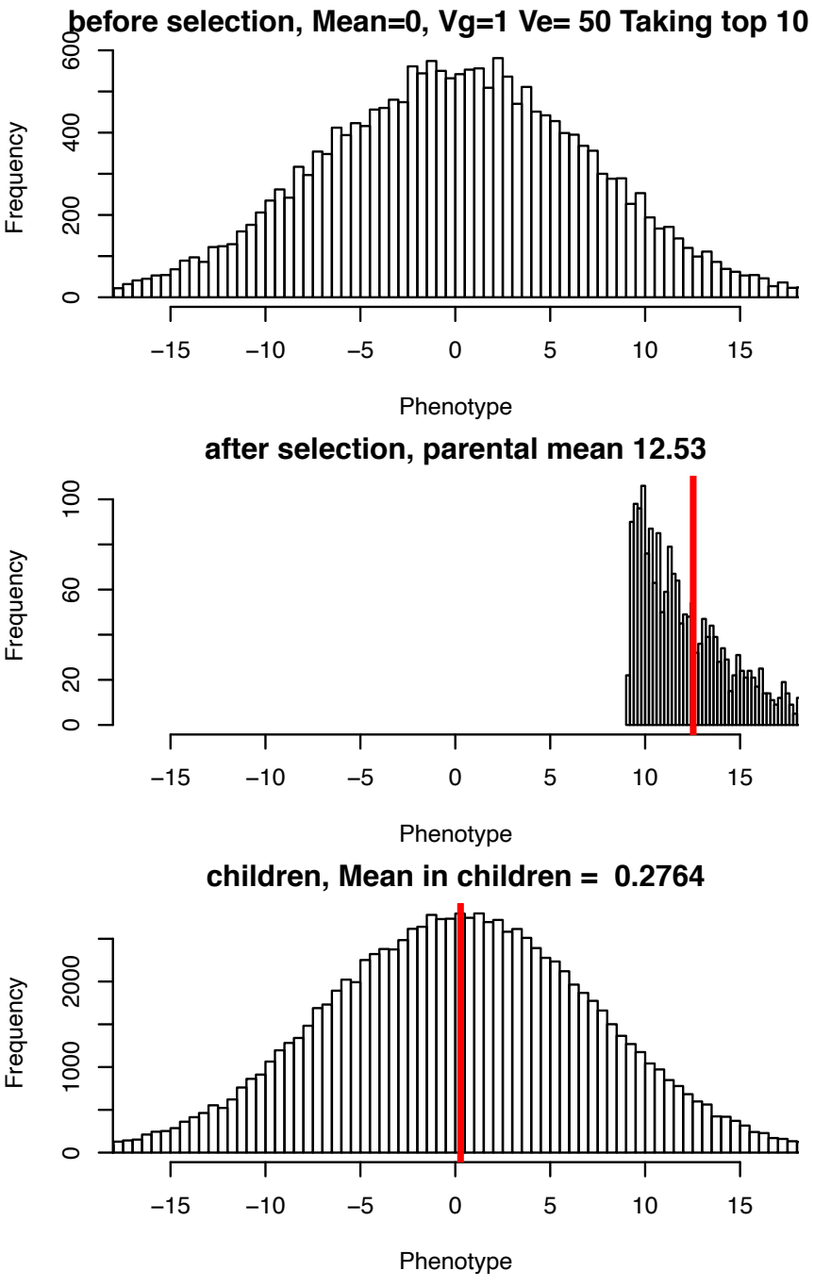


$h^2 \approx 99\,\%$

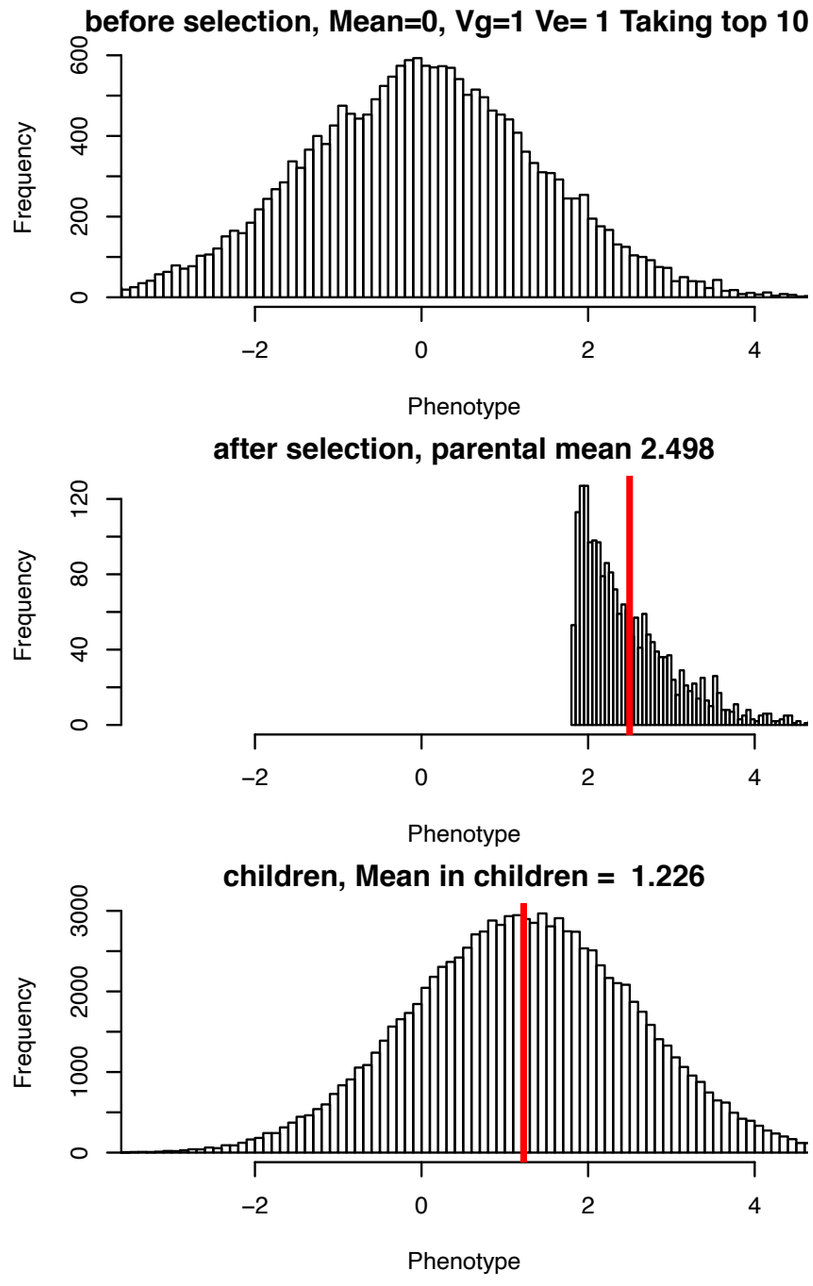


Truncation Selection

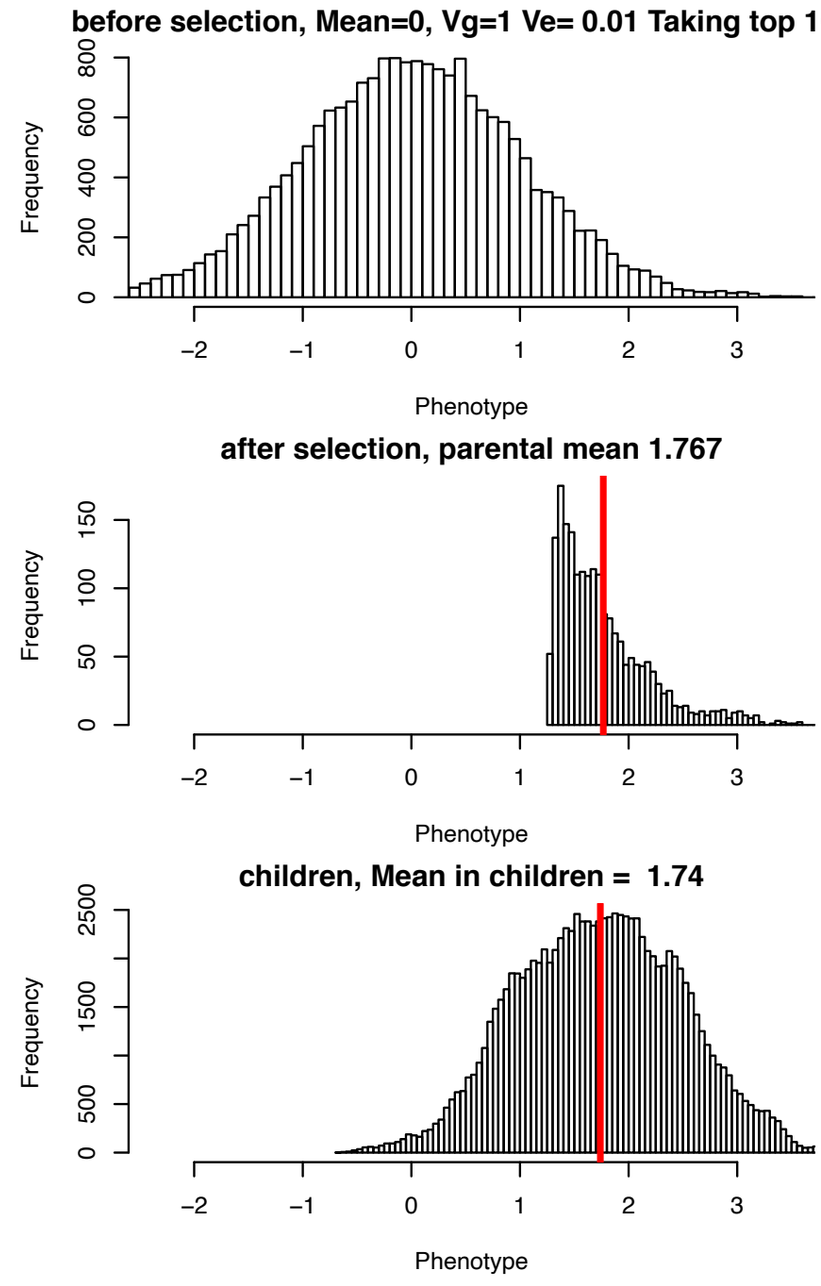
$h^2 \approx 2\,\%$



$h^2 = 50\,\%$

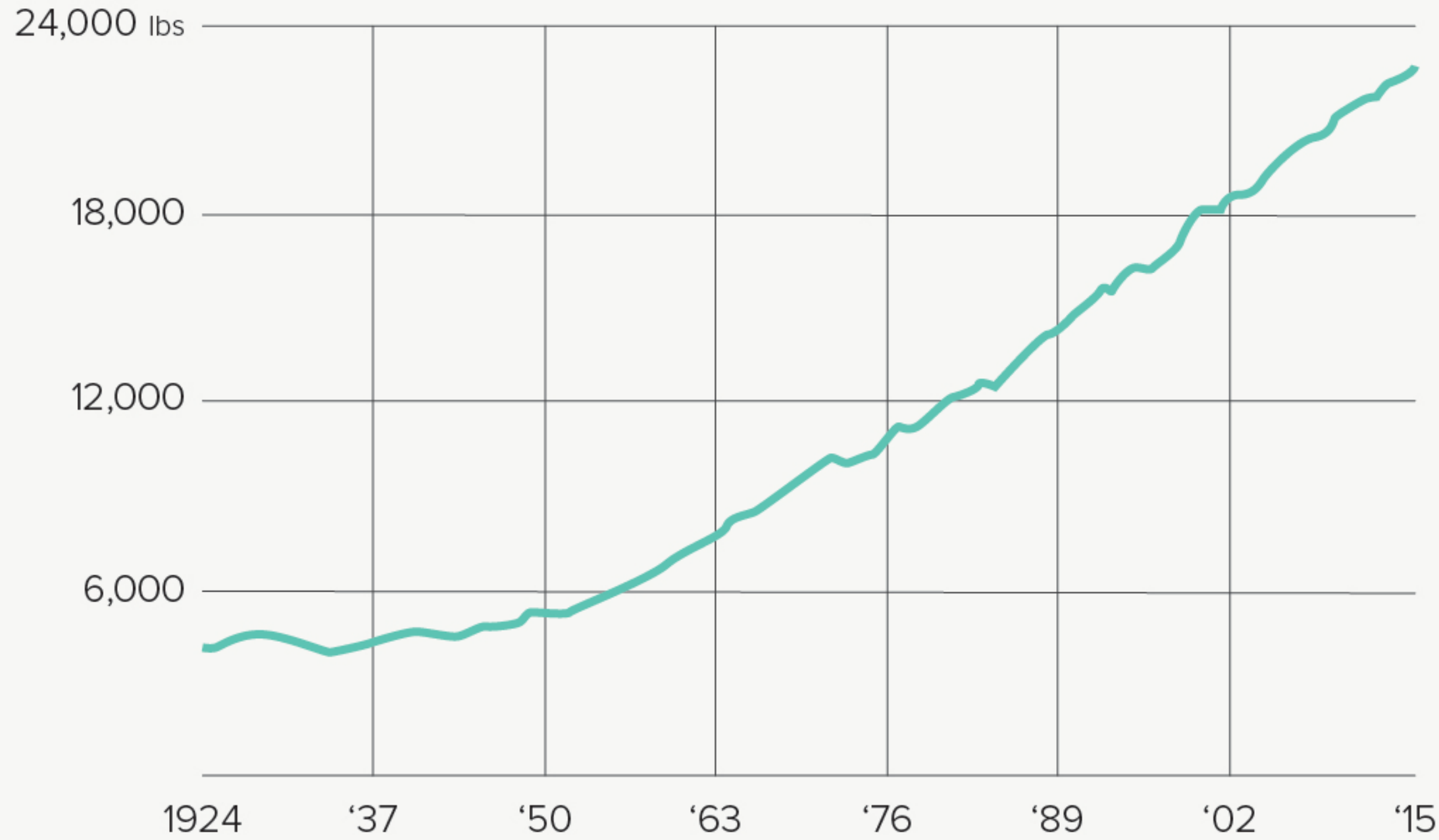


$h^2 \approx 99\,\%$



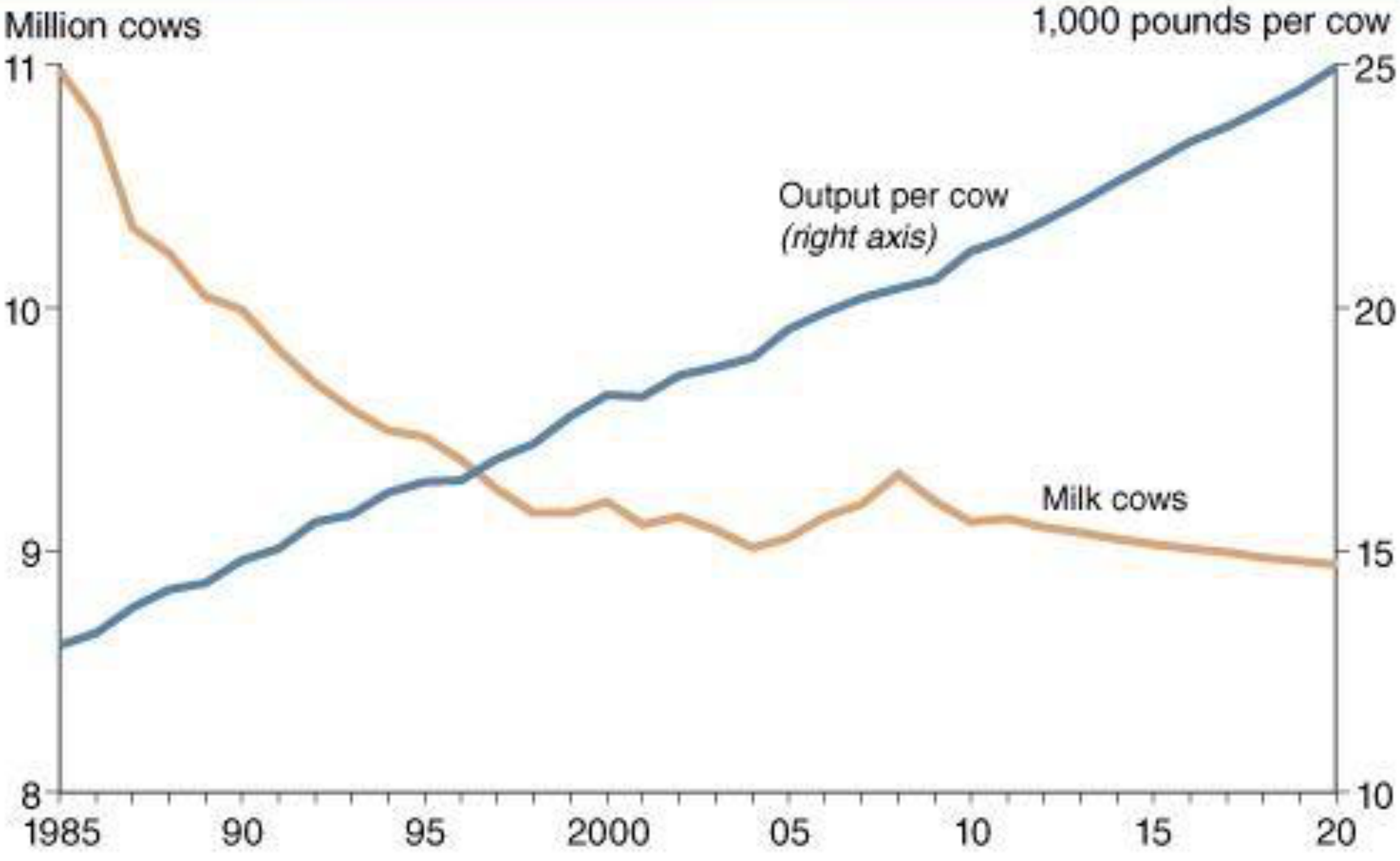
Milk Per Cow

$h^2 \approx 30\,\%$



Source: USDA-NASS

U.S. dairy herd and milk production per cow



Source: USDA, Economic Research Service using USDA Agricultural Projections to 2020.

One Sperm Donor, 150 Offspring

David Ahntholz for The New York Times

OFFSPRING Ryan Kramer, 20, of Pasadena, Calif., is the child of a donor.

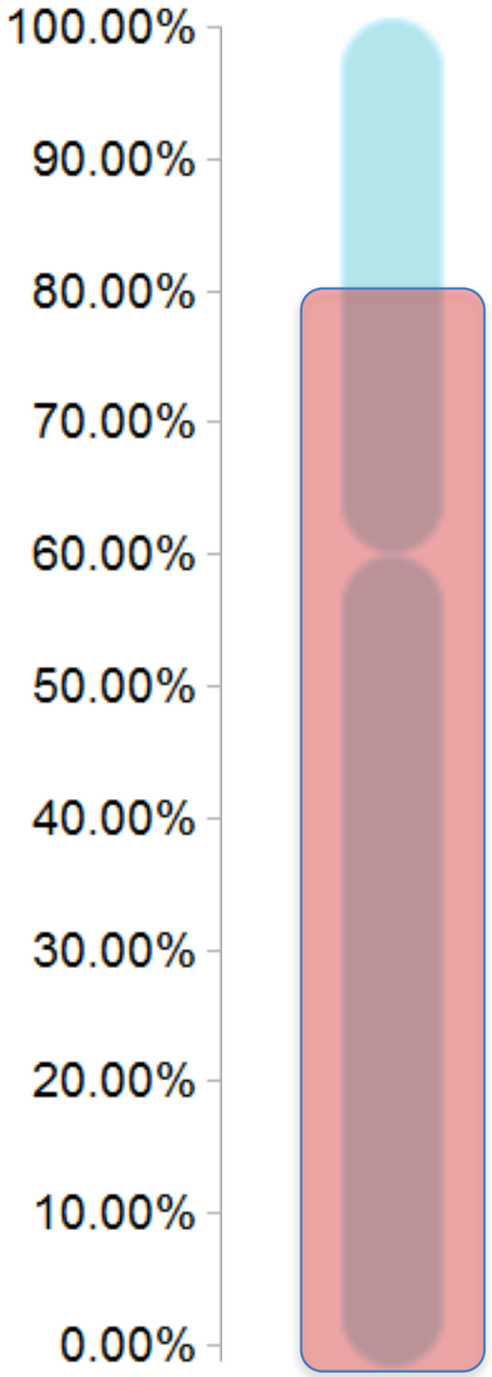
By JACQUELINE MROZ
Published: September 5, 2011

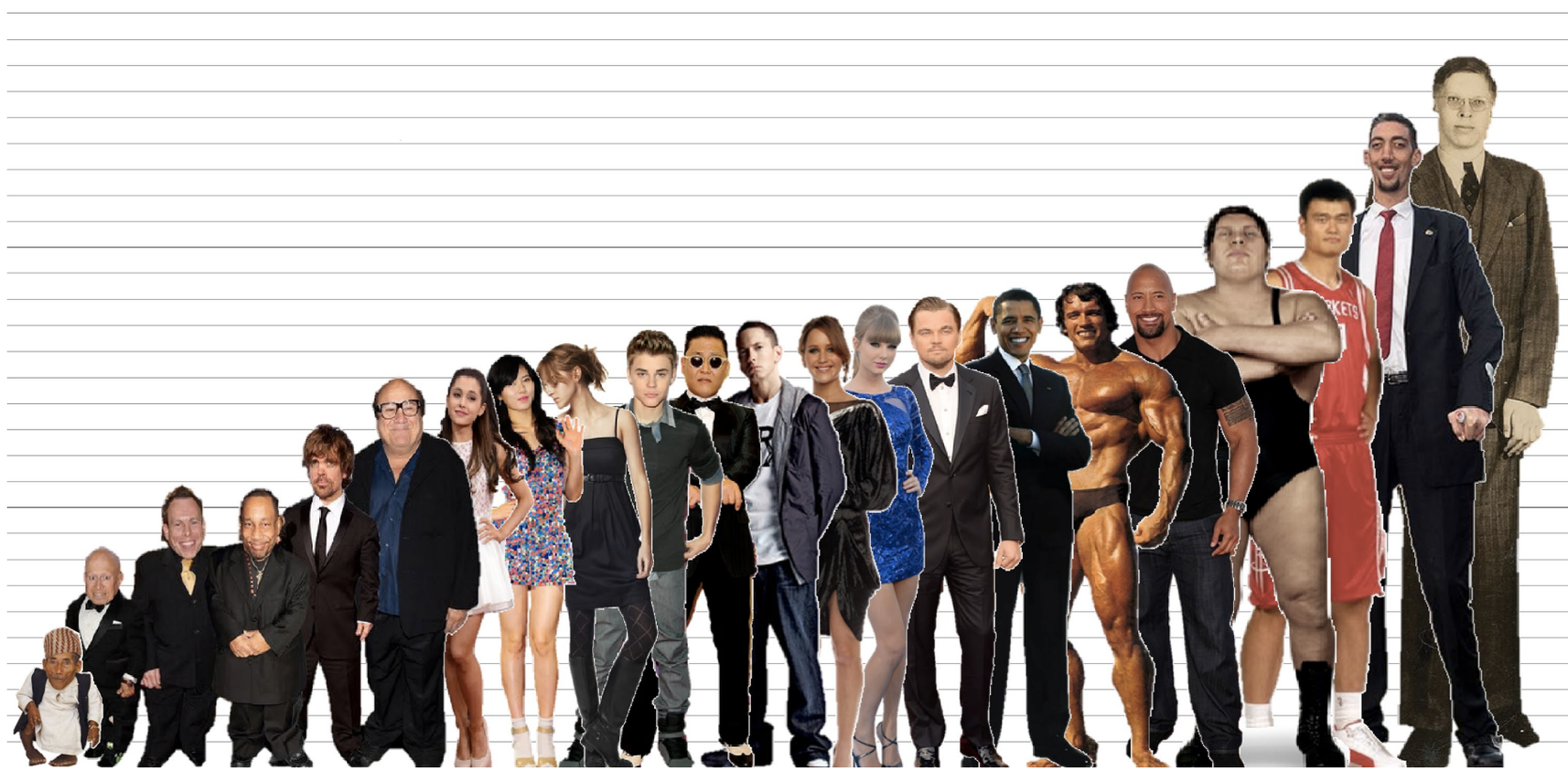
- As perfect a case of "random mating" as you will ever get!
- Looking at the offspring of major donors could illuminate effects of genetics vs environment.

- Today there are 150 children, all conceived with sperm from one donor, in this group of half siblings, and more are on the way. "It's wild when we see them all together — they all look alike," said Ms. Daily, 48, a social worker in the Washington area who sometimes vacations with other families in her son's group.

http://www.nytimes.com/2011/09/06/health/06donor.html

An estimated 80% of variation in height driven is driven by genetics



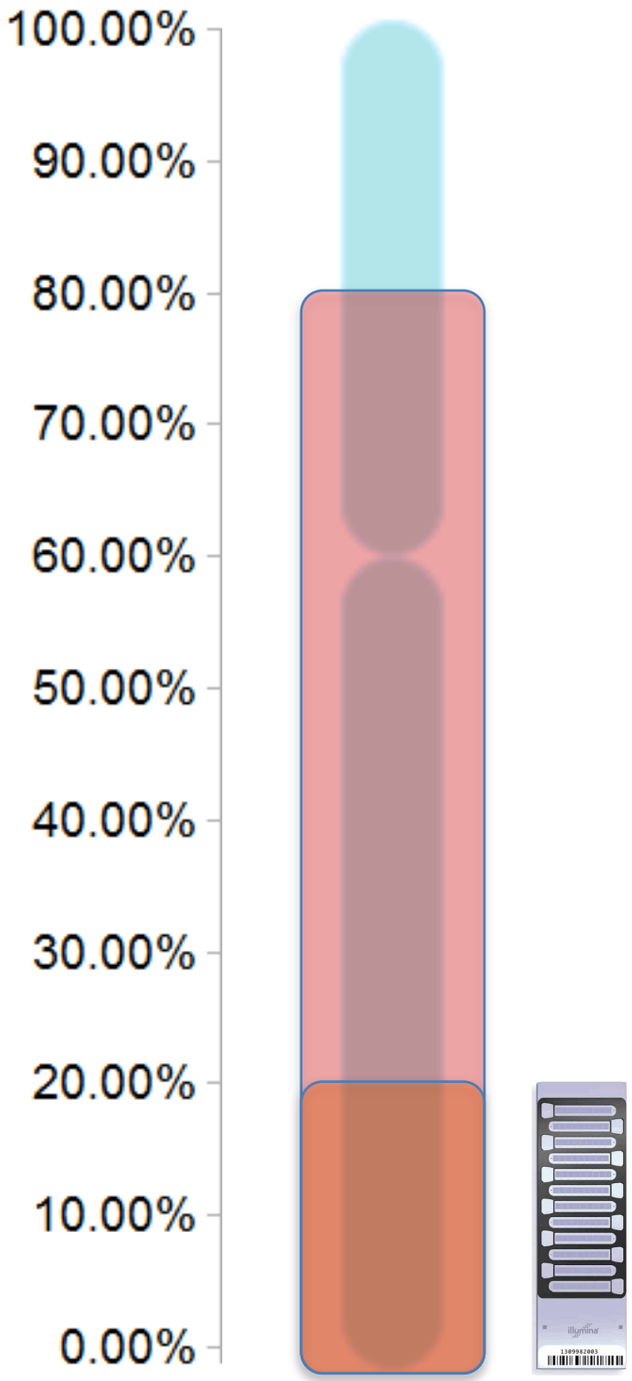


Large twin study

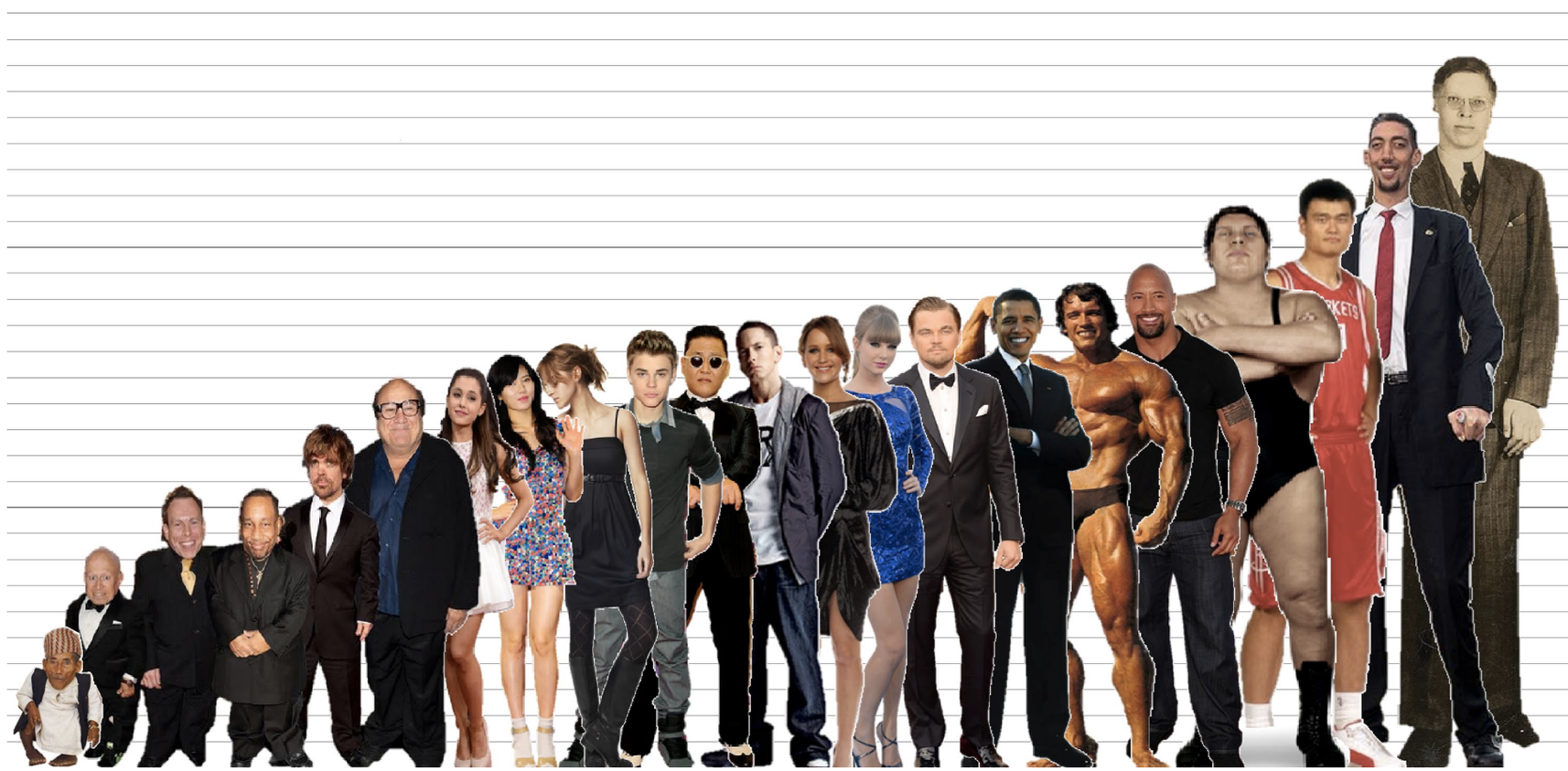
Silventoinen et al, 2003 Twin Research

http://i.ytimg.com/vi/E0Aeks_id6c/maxresdefault.jpg

But GWAS explain only 20% of the variation in height



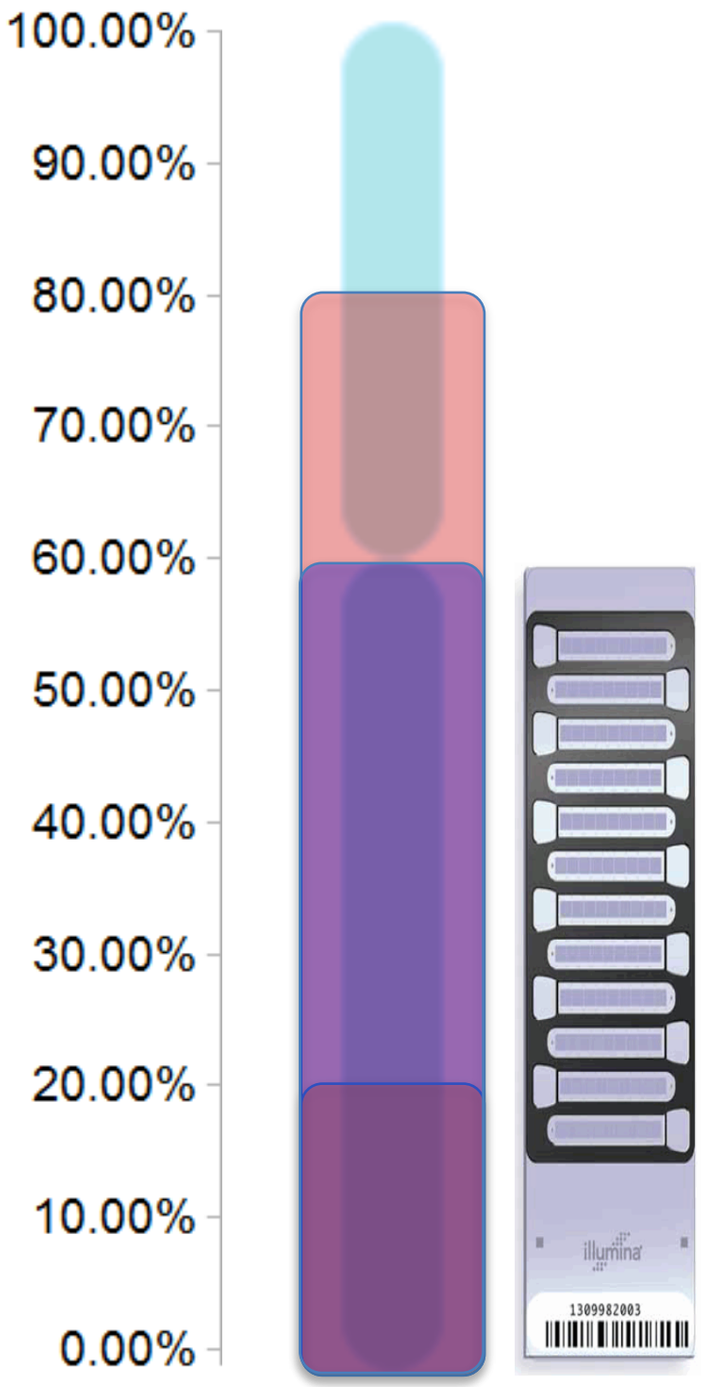
h^2_{GWAS}: The narrow-sense heritability explained by summing the effects of GWAS identified SNPs.



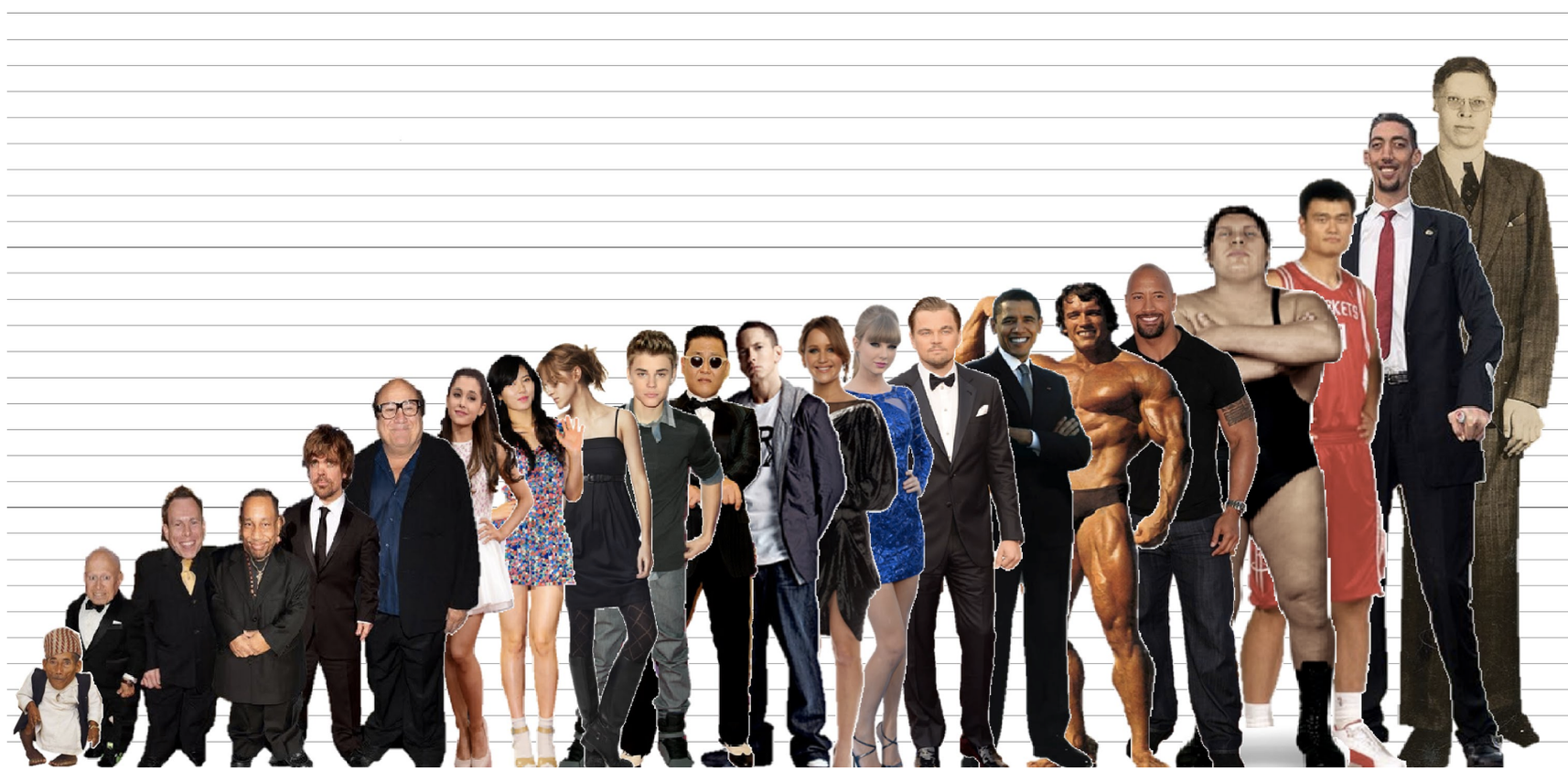
250,000 subjects

Wood et al, 2014 Nat. Genet.
i.ytimg.com/vi/E0Aeks_id6c/maxresdefault.jpg

GWAS have the potential to explain 60% of the variation in height



h_g^2 : The narrow-sense heritability explained by all genotyped SNPs.



250,000 subjects

Wood et al, 2014 Nat. Genet.
i.ytimg.com/vi/E0Aeks_id6c/maxresdefault.jpg

Challenges For Studying Complex Diseases

The case of the missing heritability Maher, *Nature* (2008).

Major Problem

- There are no complex traits in which we know:
    - The number of causal variants
    - The frequencies of all the causal variants
    - The effect sizes of all the causal variants
    - The fitness effect of all the causal variants
- We need a thorough simulation study where we can vary all of these parameters and see how they effect our answer!

Possible Origins Of Missing Heritability

Candidates
Common variants of weak effect
Incomplete linkage to causal alleles/multiple causal alleles in locus
GxG / GxE Interactions
Rare variants
Structural variation

From GWAS To Deep Sequencing

- Genome-wide association studies (GWAS) seek to identify common variants that contribute to common disease

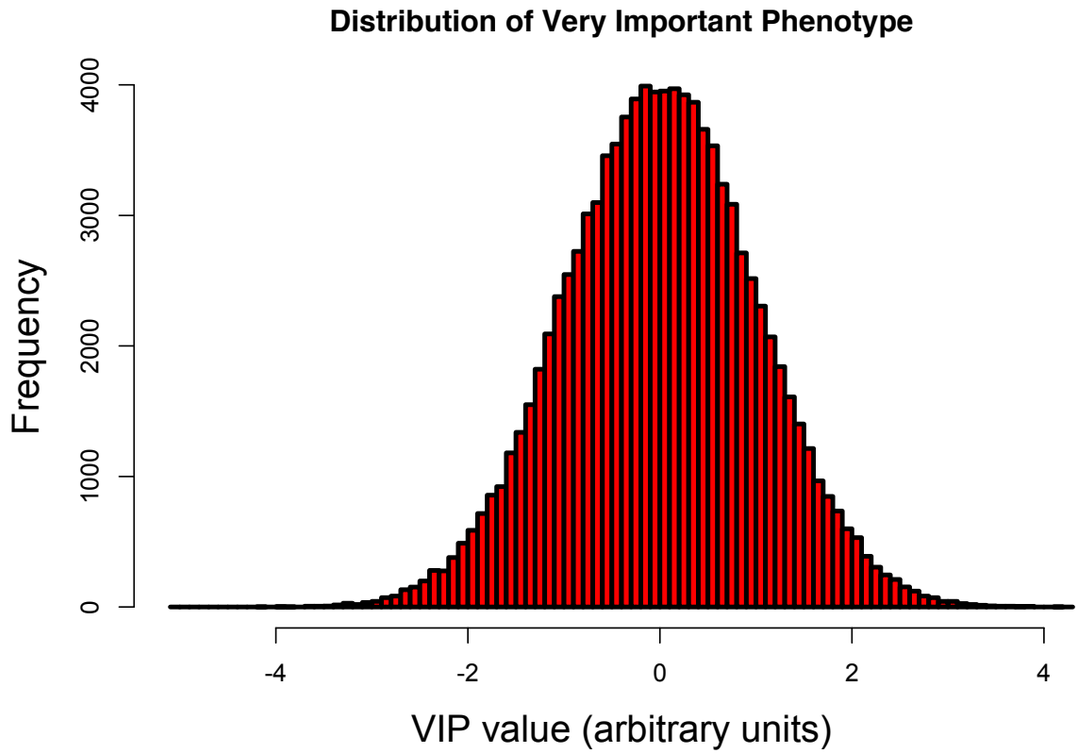
- Successfully identified many candidate disease-associated genes

- Challenges:
  - Generally have low relative risk
  - Explain only a small proportion of the phenotypic variance

  - Provides candidate loci, but causal variant is rarely typed

- Implication:
  - Predictive power of GWAS is minimal…

“Missing” heritability - calculating variance accounted for by GWAS



Suppose *k* variants are found to be associated with VIP…

Contribution from each SNP

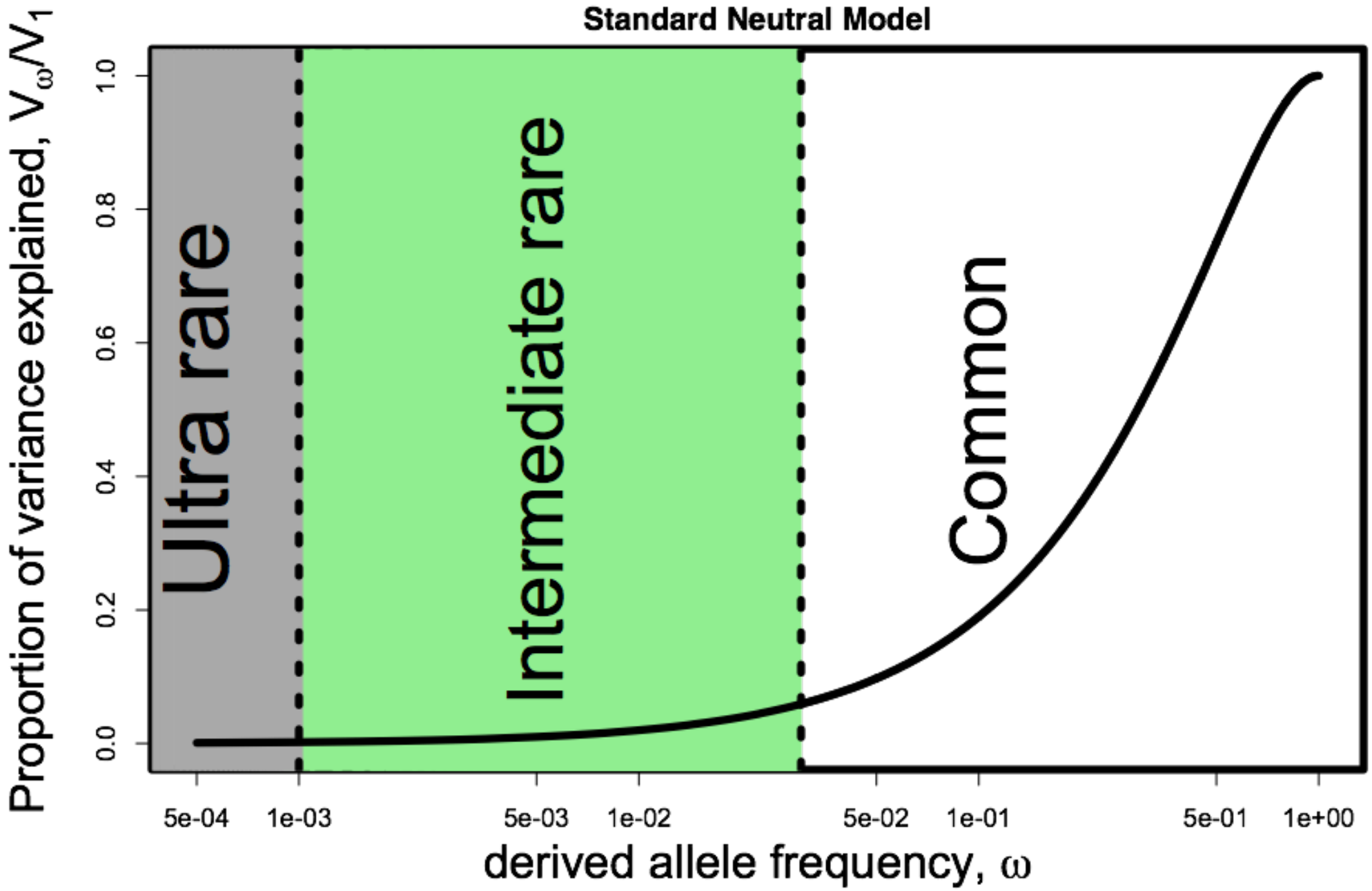
$$v = \frac{1}{2} z^2 x(1-x)$$

Total variance from GWAS

$$V_{\text{GWAS}}(P) = \sum_k v_k$$

Compare to GWAS $V_{\text{GWAS}}(P) < h^2 \times V(P)$

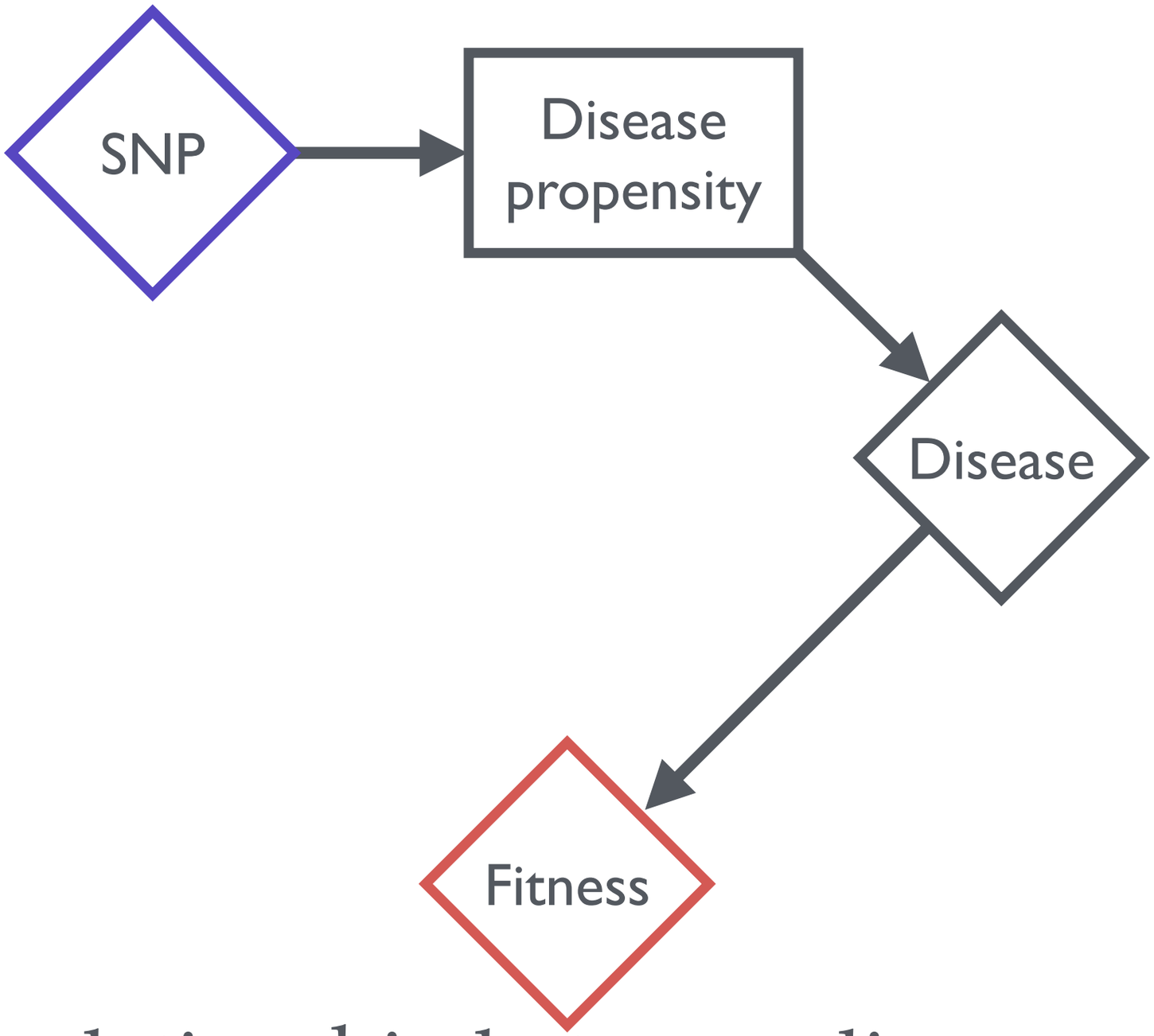
Where is the “missing” heritability?



Population Genetics

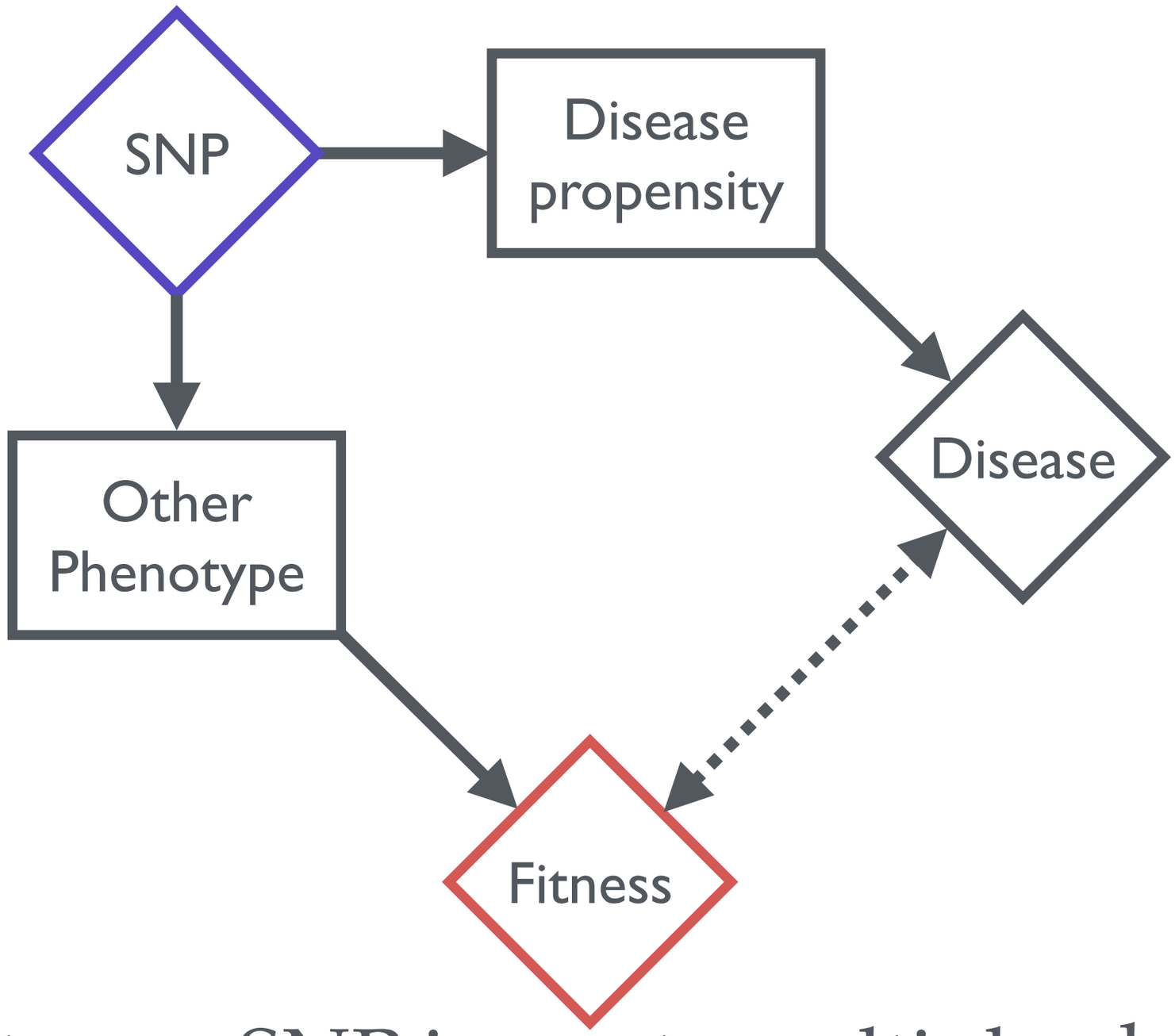
- Why would cases have an excess of **rare** non-synonymous variants in disease-associated genes?
    - Recent neutral mutations that have not had time to spread
    - Deleterious mutations restricted to low frequency
- Population genetic analyses are ideally suited to distinguish these cases.

Evolutionary Models Of Complex Disease



Direct relationship between disease and fitness

Evolutionary Models Of Complex Disease



Pleiotropy: SNP impacts multiple phenotypes

Uricchio et al., Genome Research (2016)

The Model Of Eyre-Walker (2010)

- The phenotypic effect size has a direct relationship to selection coefficient of causal mutations:

$$z = \delta S^{\tau}(1 + \epsilon)$$

- Where:
  - $\varepsilon \sim N(0, \sigma^2)$
  - δ = random sign (trait increasing / decreasing)
  - S = selection coefficient
  - τ = measures how the mean absolute effect of a mutation on the trait increases with the strength of selection

Eyre-Walker, PNAS (2010)

The Model Of Simons Et Al (2014)

- The phenotypic effect size **may** have a direct relationship to selection coefficient of causal mutations:

$$z_s \propto \begin{cases} s & \text{with probability } \rho \\ s_r & \text{with probability } (1-\rho) \end{cases}$$

- Where:
  - ρ = Probability that the trait effect is proportional to the selection coefficient: **Pleiotropy**!!
  - s = selection coefficient
  - s_r = random selection coefficient

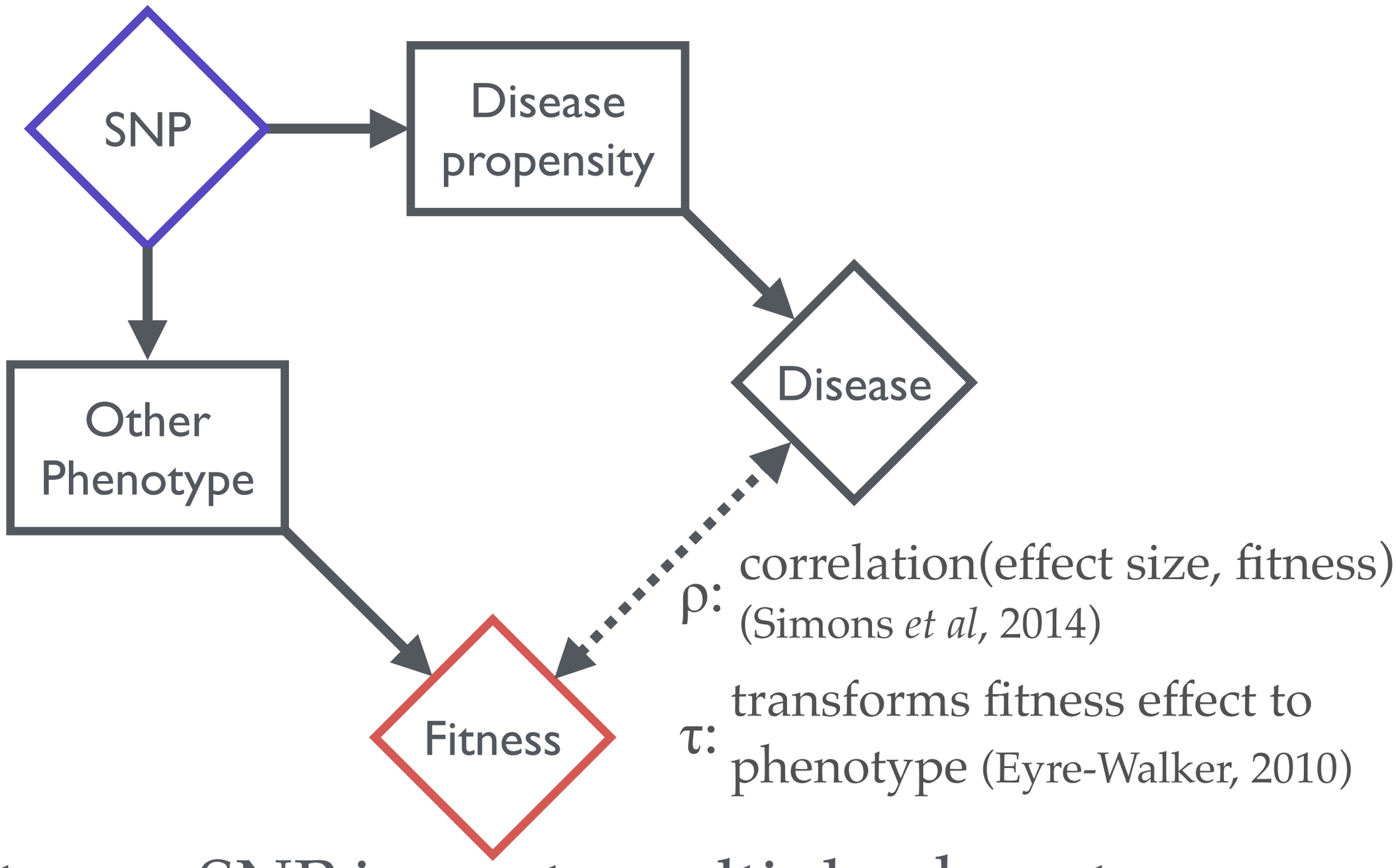
Simons et al, Nat Genet (2014)

The Model Of Uricchio Et Al (2016)

- A hybrid of the two: $z_s \propto \begin{cases} \delta|s|^{\tau} & \text{with probability } \rho \\ \delta|s_r|^{\tau} & \text{with probability } (1-\rho) \end{cases}$
- Where:
  - δ = random sign (trait increasing / decreasing)
  - τ = measures how the mean absolute effect of a mutation on the trait increases with the strength of selection
  - ρ = Probability that the trait effect is proportional to the selection coefficient: **Pleiotropy**!!
  - s = selection coefficient
  - s_r = random selection coefficient

Uricchio et al, Genome Research (2016)

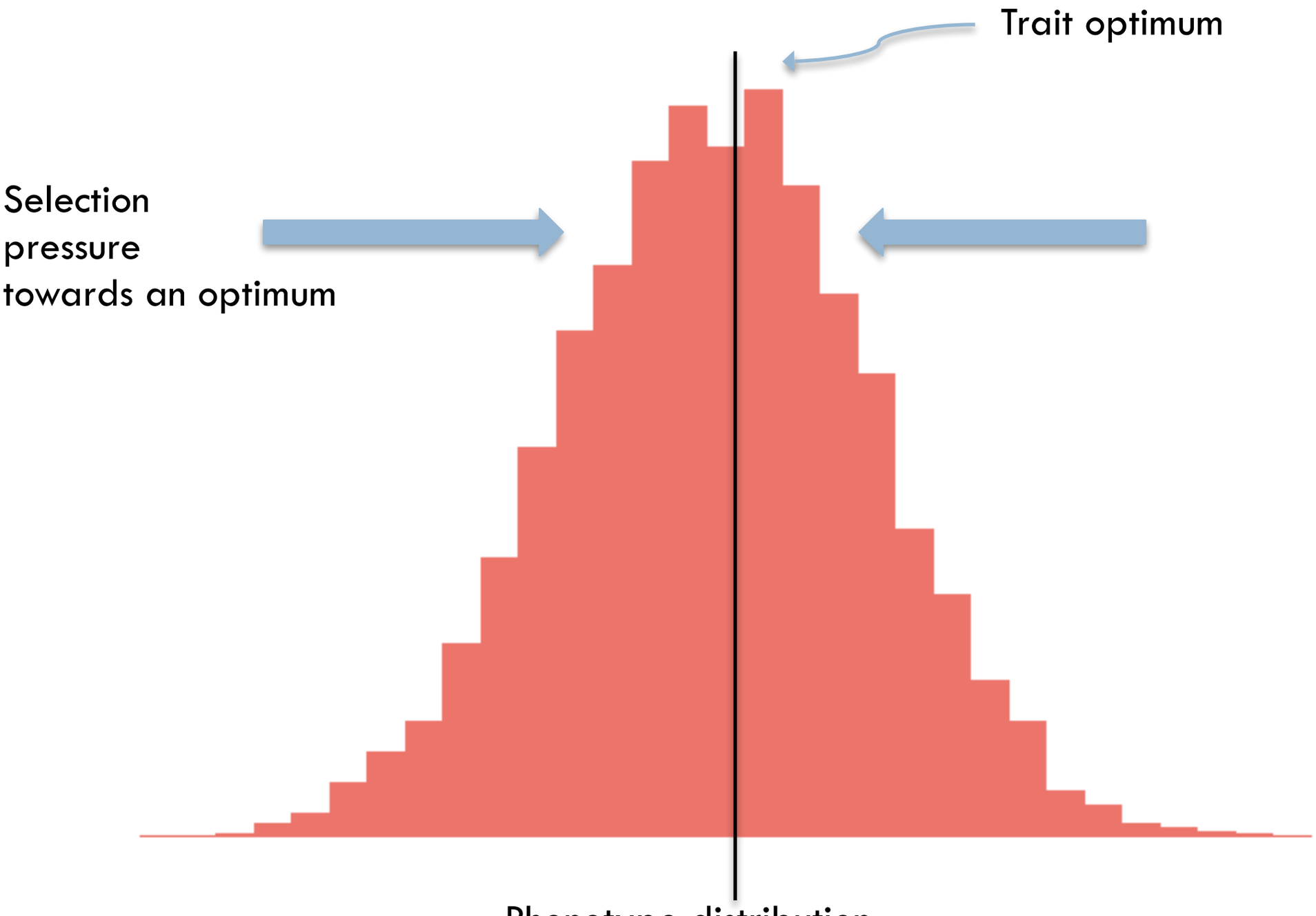
Evolutionary Models Of Complex Disease



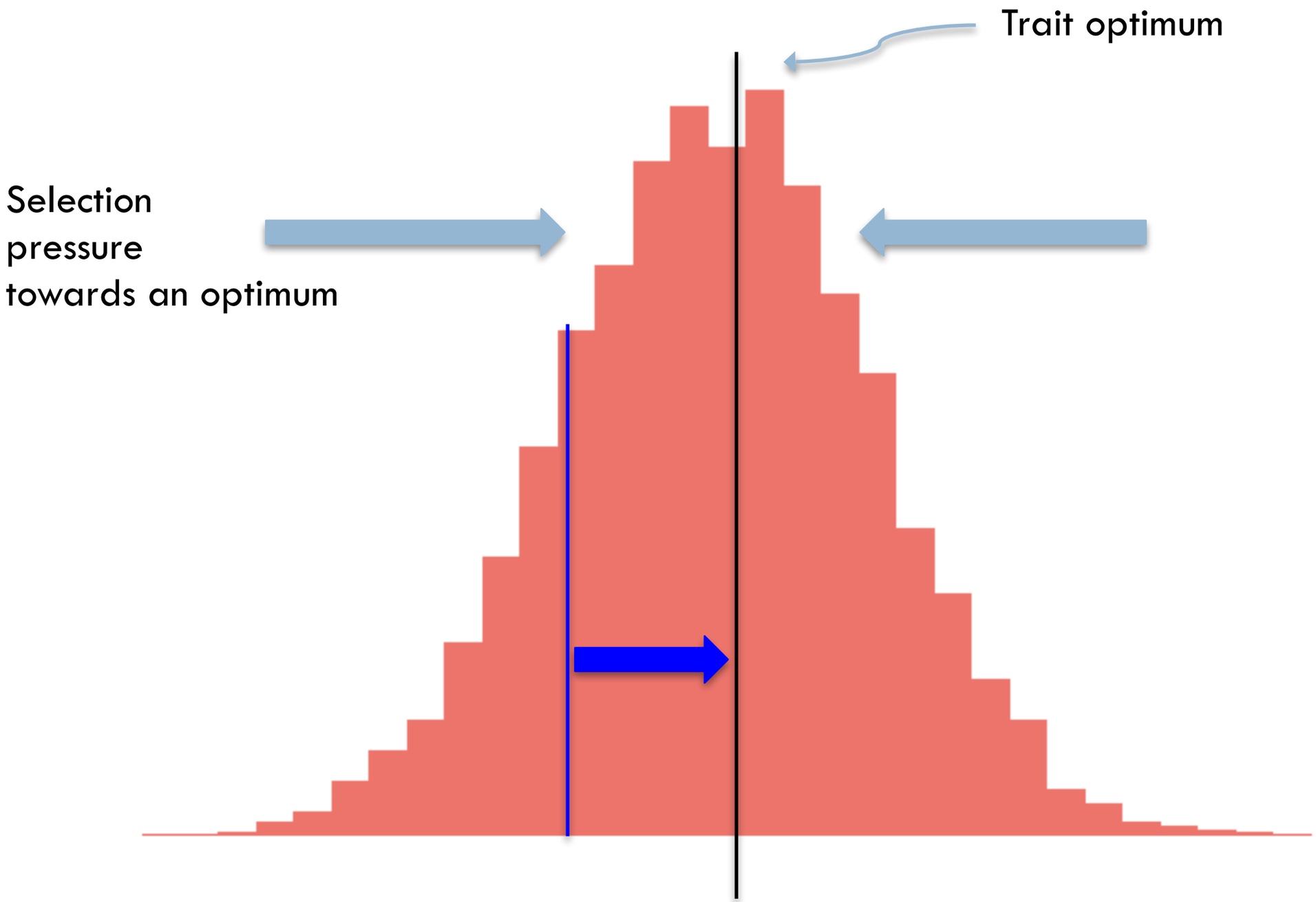
Pleiotropy: SNP impacts multiple phenotypes

Uricchio et al., Genome Research (2016)

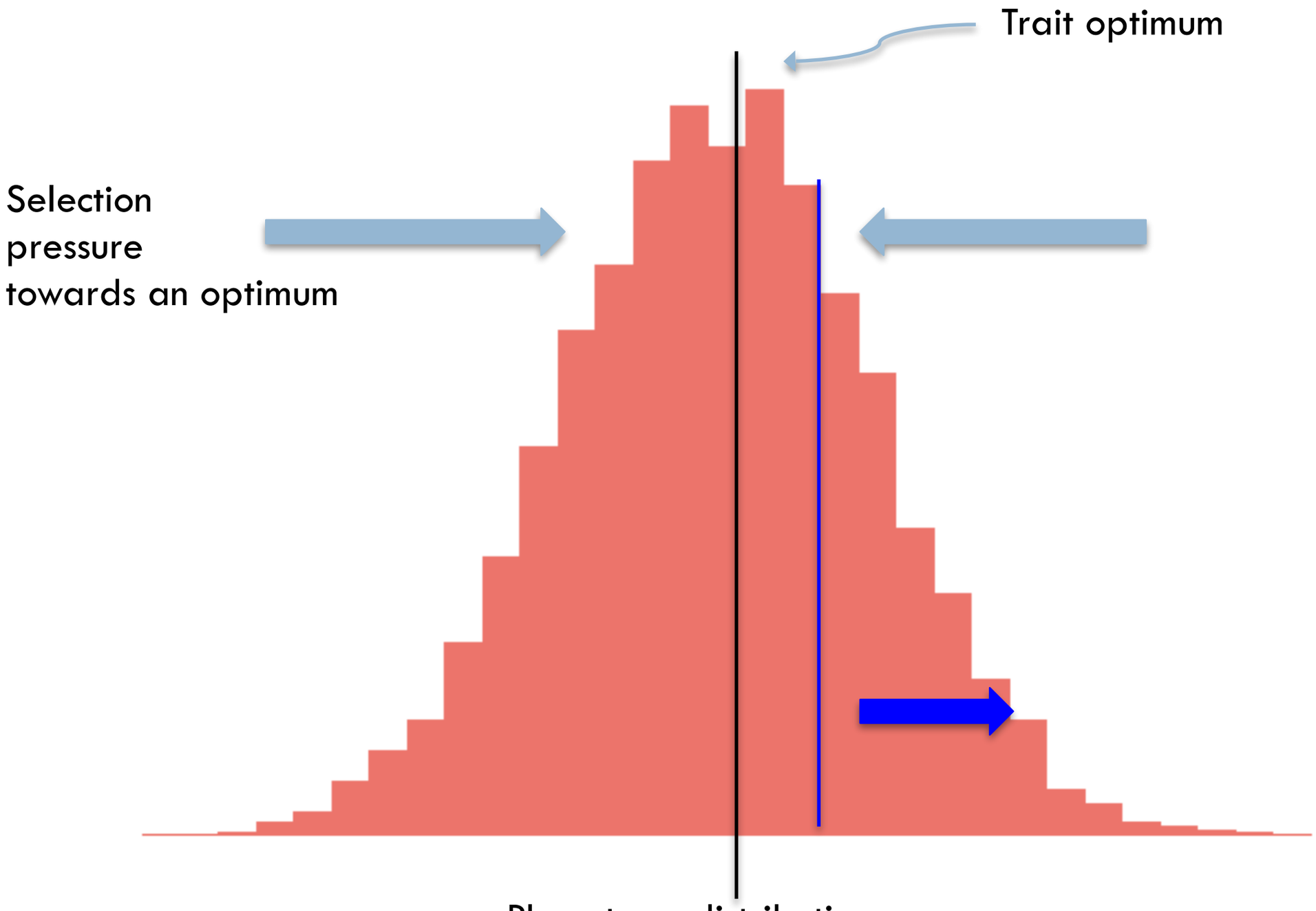
Why should we think about evolution?



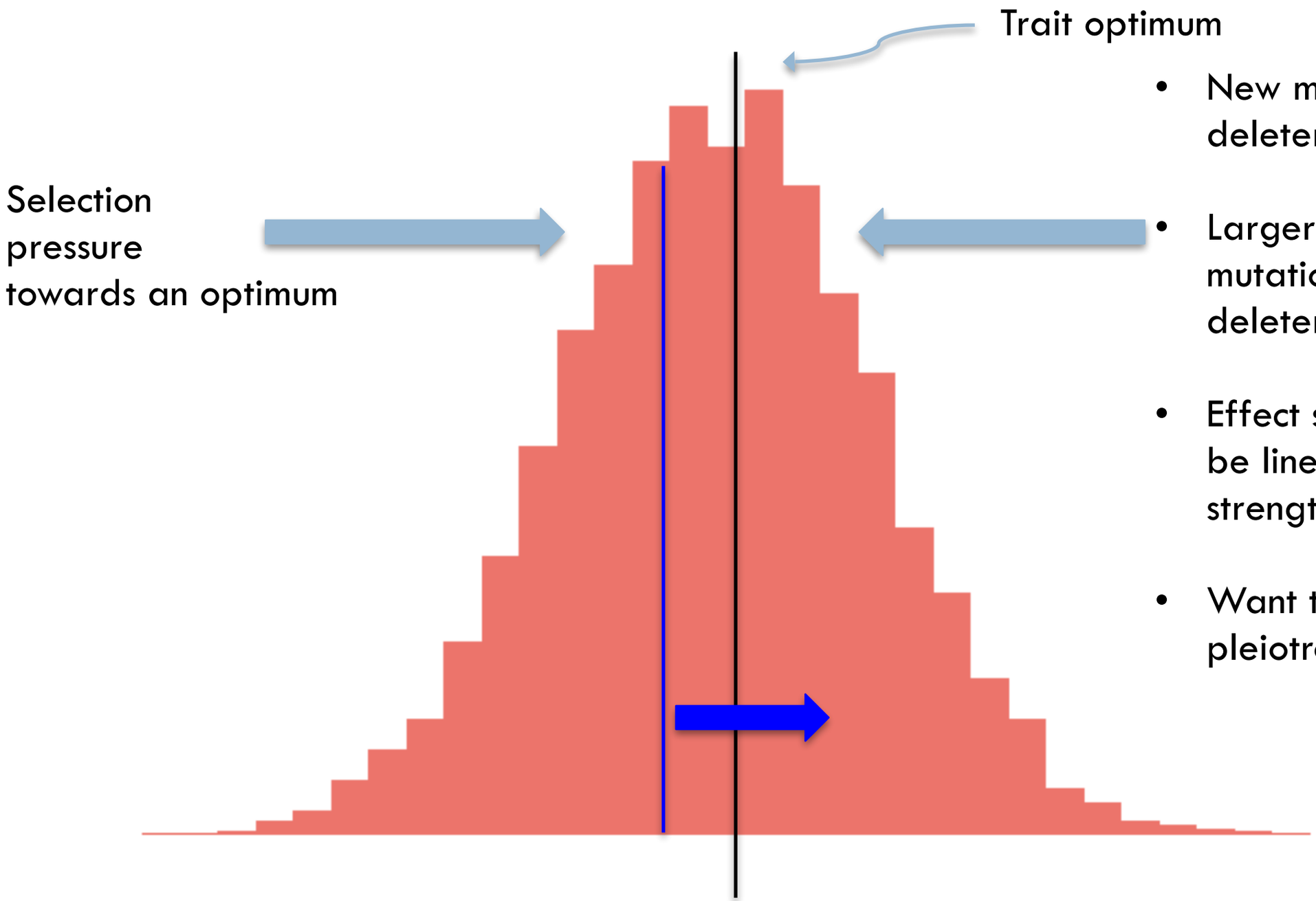
Stabilizing selection



Stabilizing selection

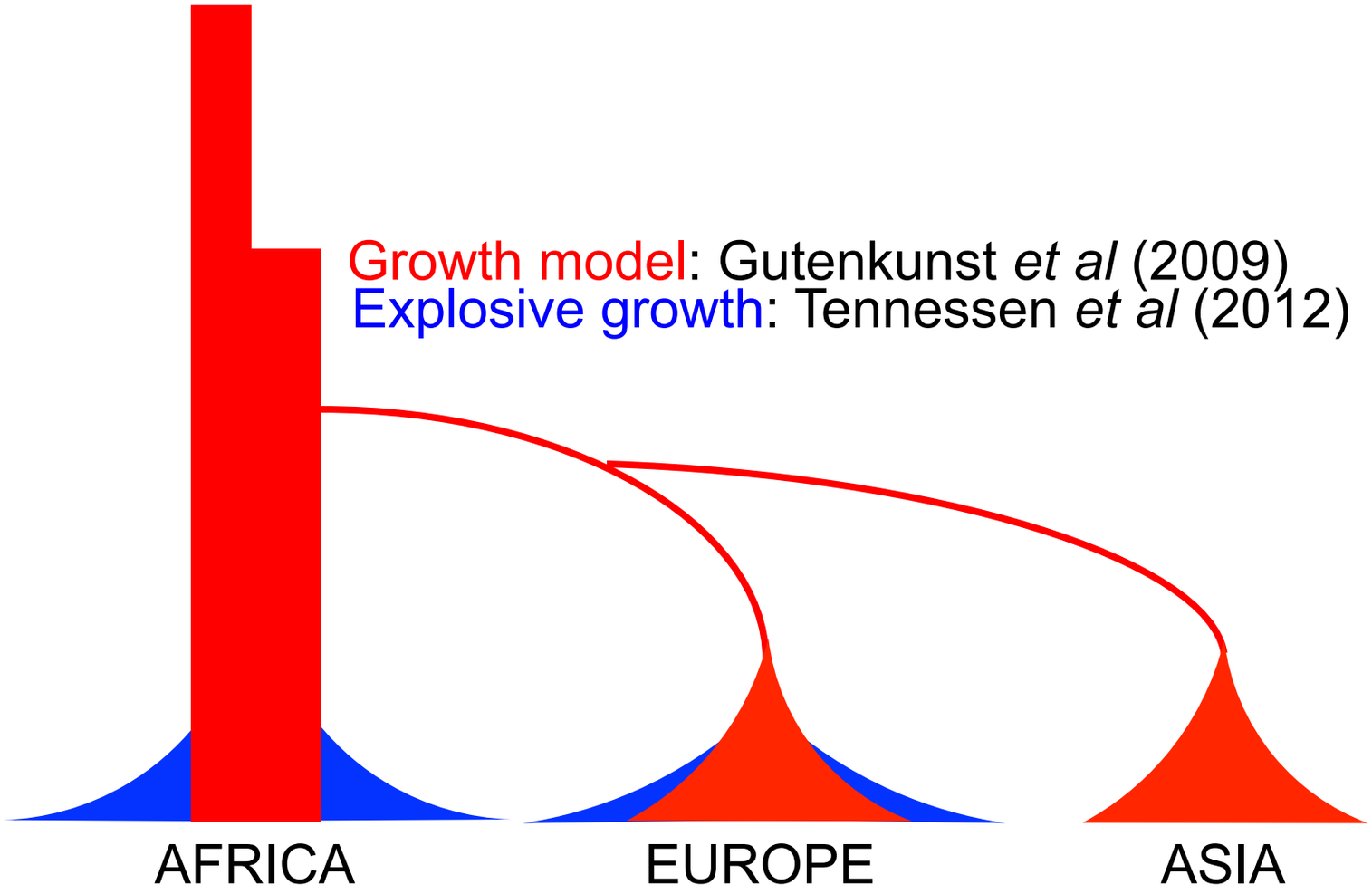


Stabilizing selection

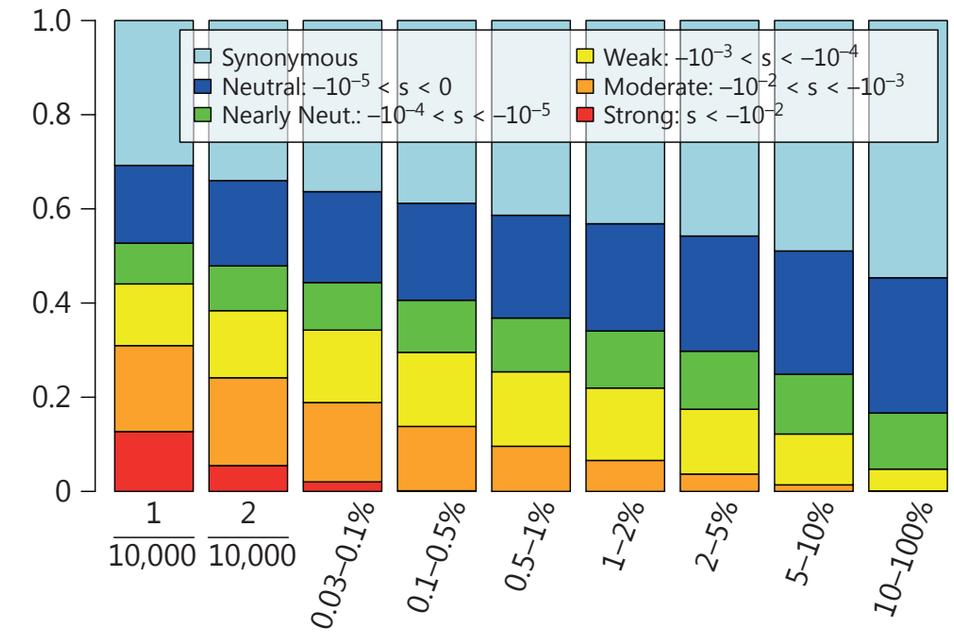


- New mutations deleterious
- Larger effect mutations are more deleterious
- Effect sizes may not be linear in selection strength
- Want to allow for pleiotropy

Human-specific demography and Selection



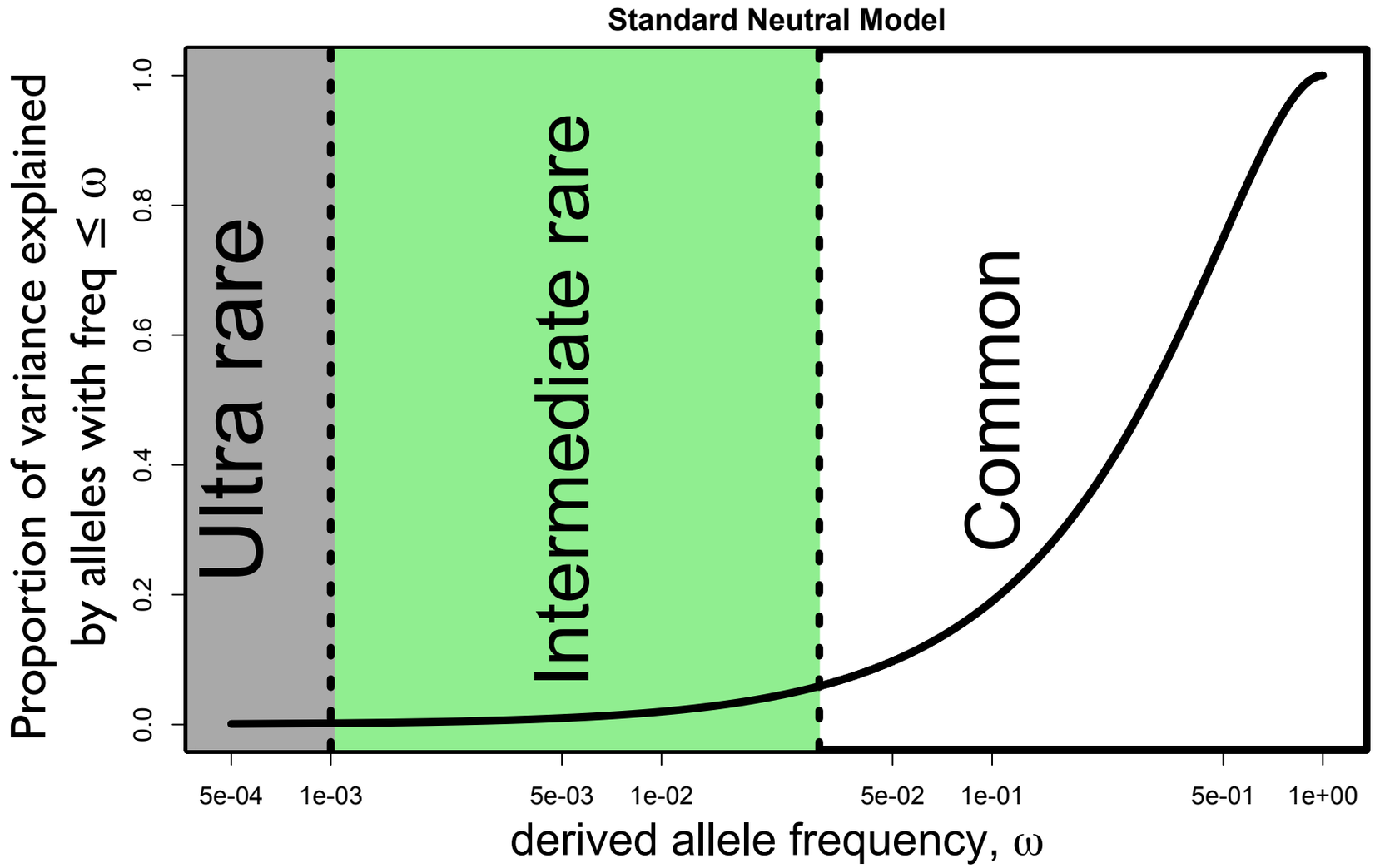
Fitness effects in non-coding DNA:
Torgerson *et al* (2009)



effect size = f(demography, natural selection)

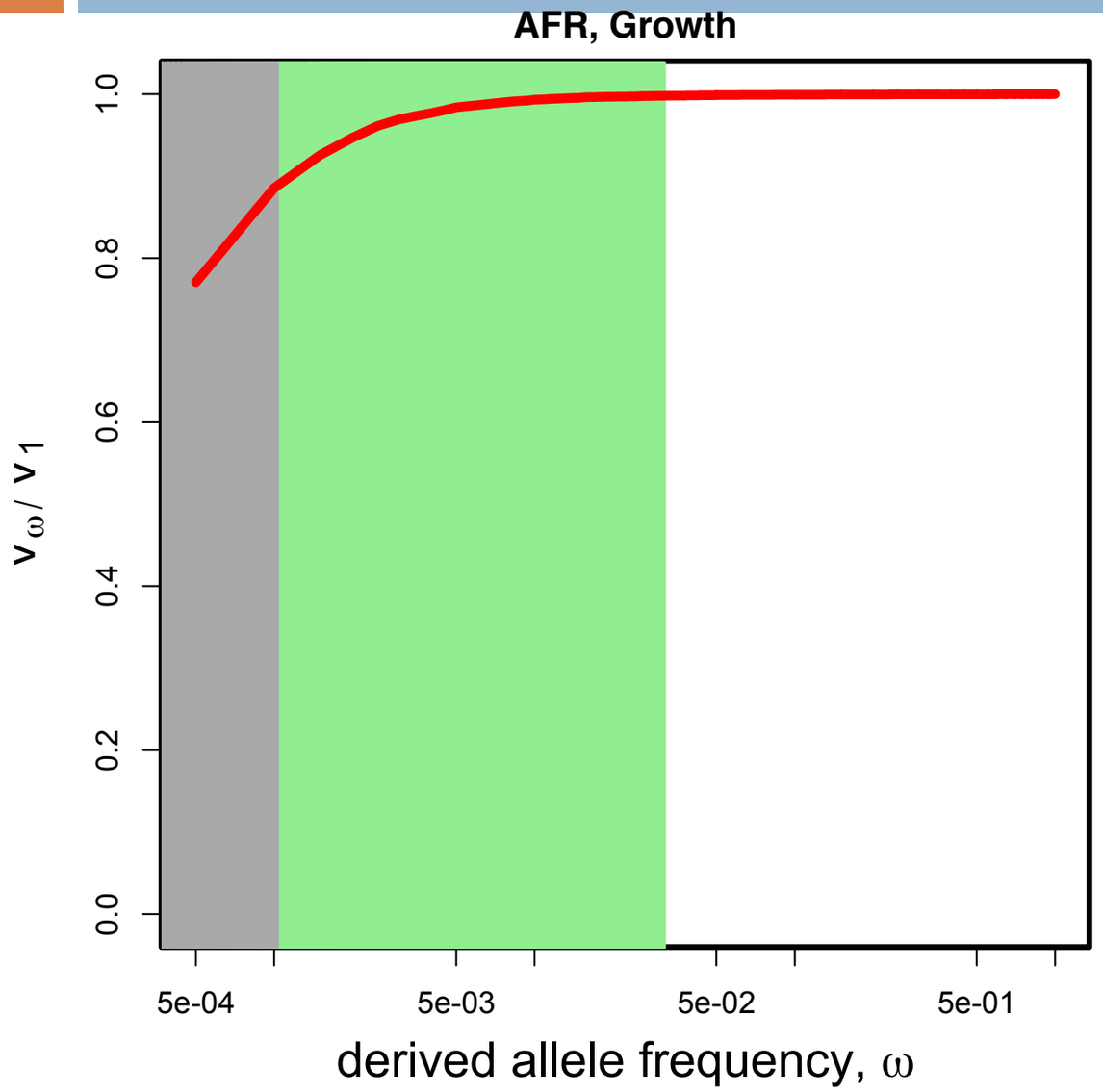
Uricchio, et al. *Genome Res* **26**, 863-873 (2016).

Neutral model: most variance explained by common alleles



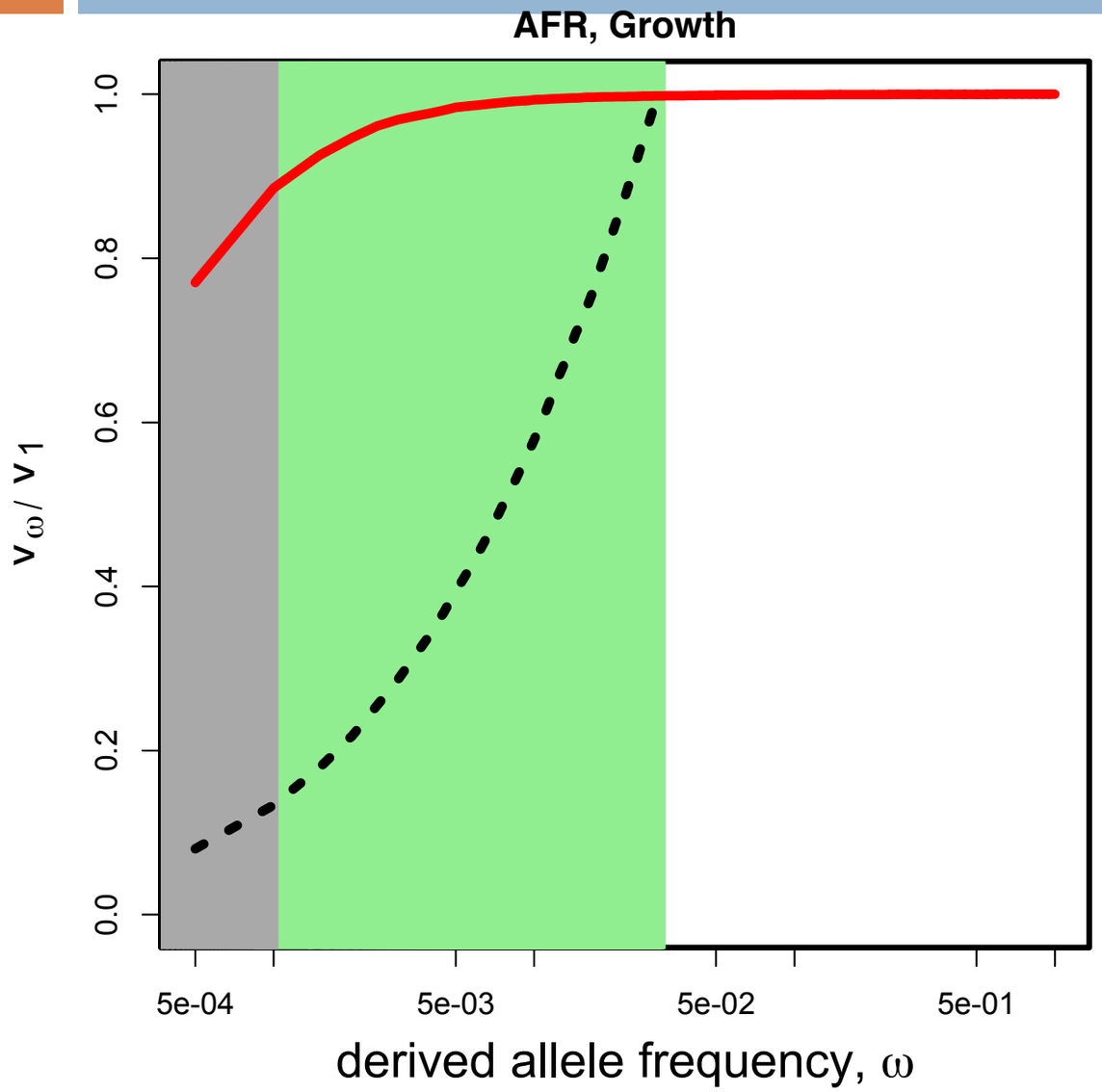
Uricchio, et al. *Genome Res* **26**, 863-873 (2016).

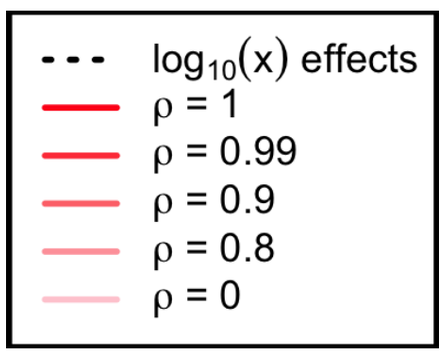
Genetic architecture is altered by selection and demography



Uricchio, et al. *Genome Res* **26**, 863-873 (2016).

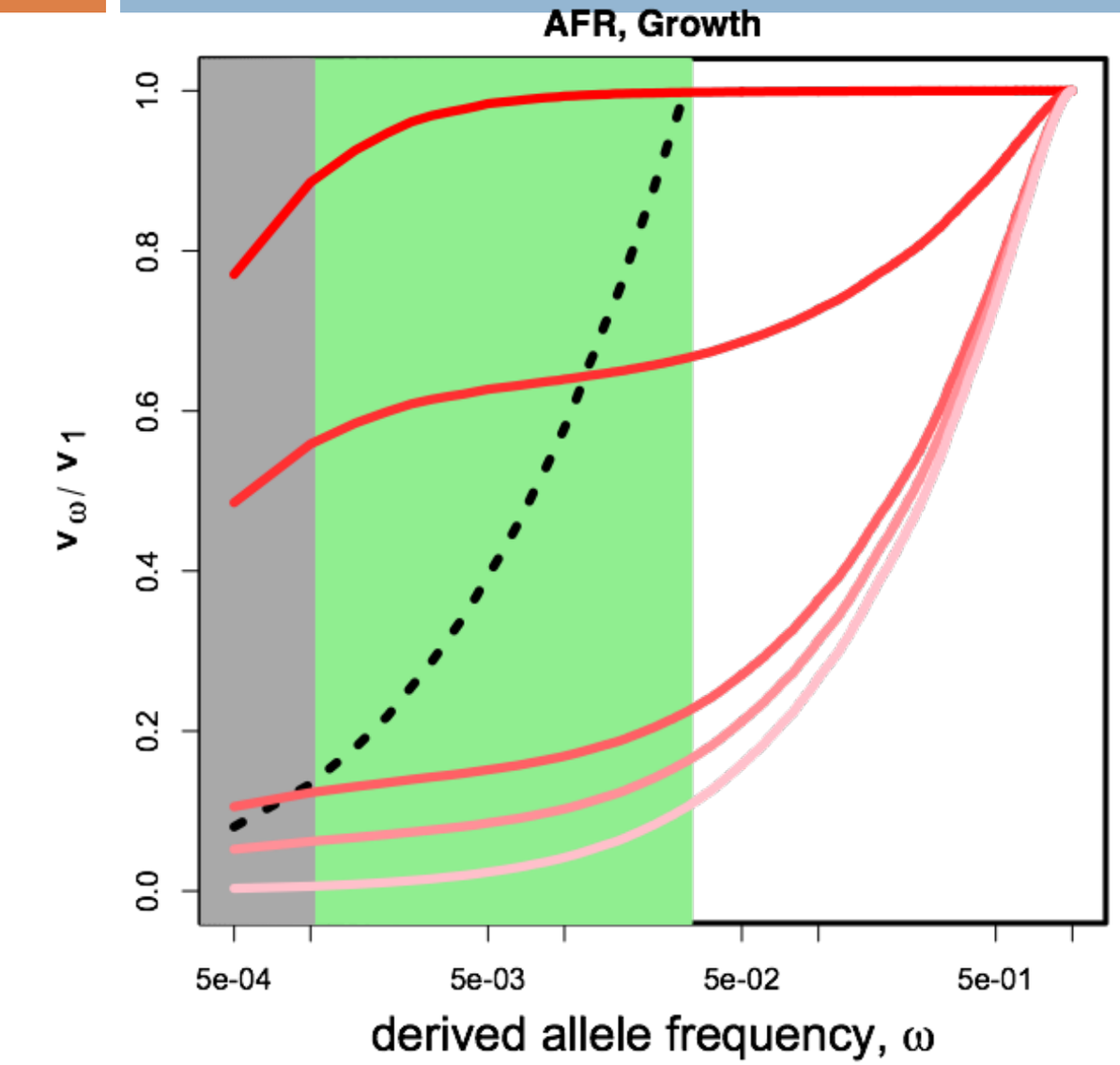
Genetic architecture is altered by selection and demography

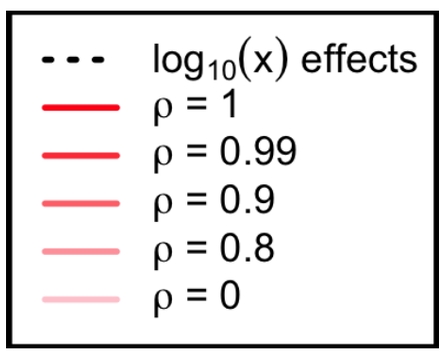




Uricchio, et al. *Genome Res* **26**, 863-873 (2016).

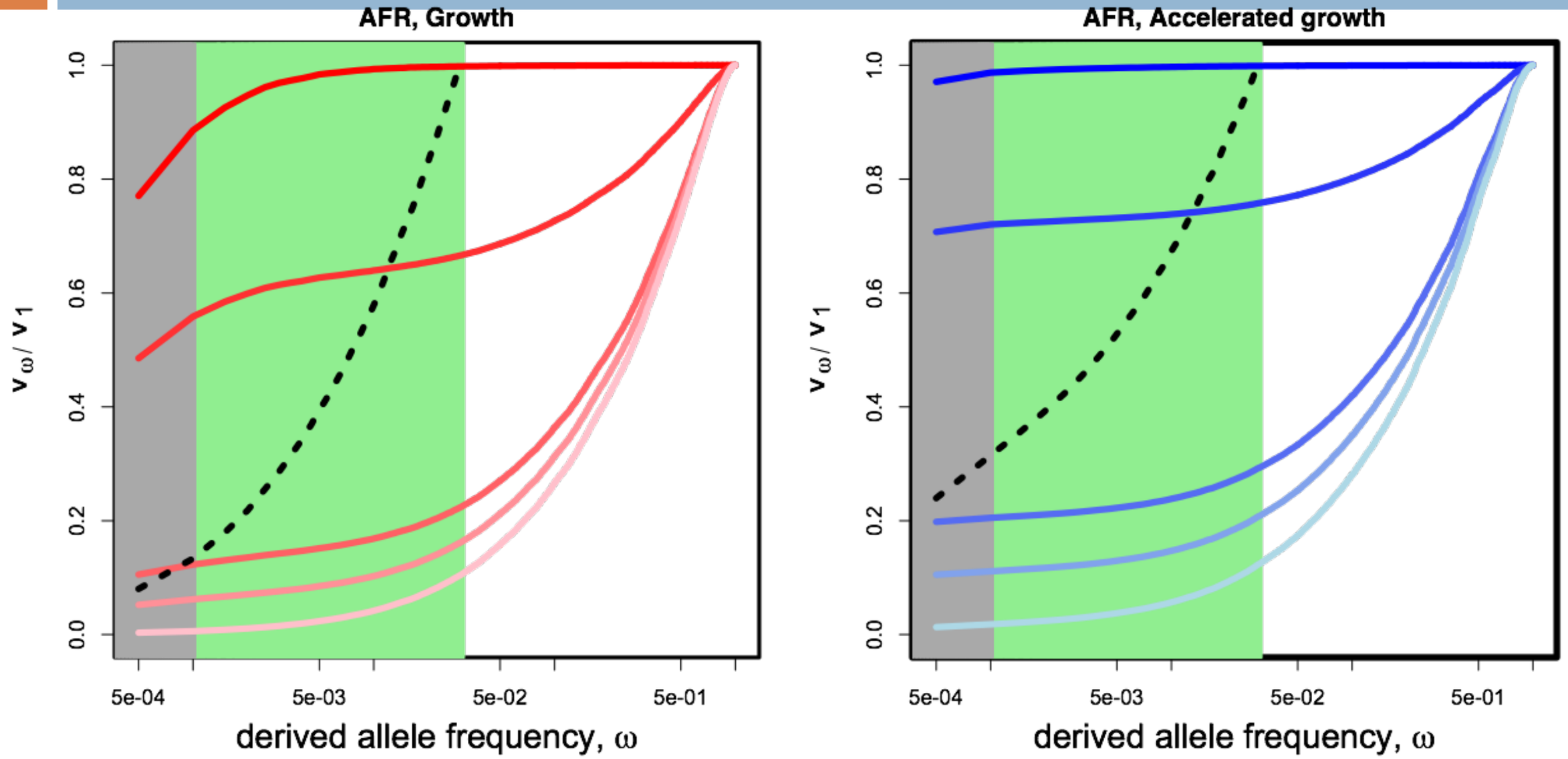
Genetic architecture is altered by selection and demography





Uricchio, et al. *Genome Res* **26**, 863-873 (2016).

Genetic architecture is altered by selection and demography

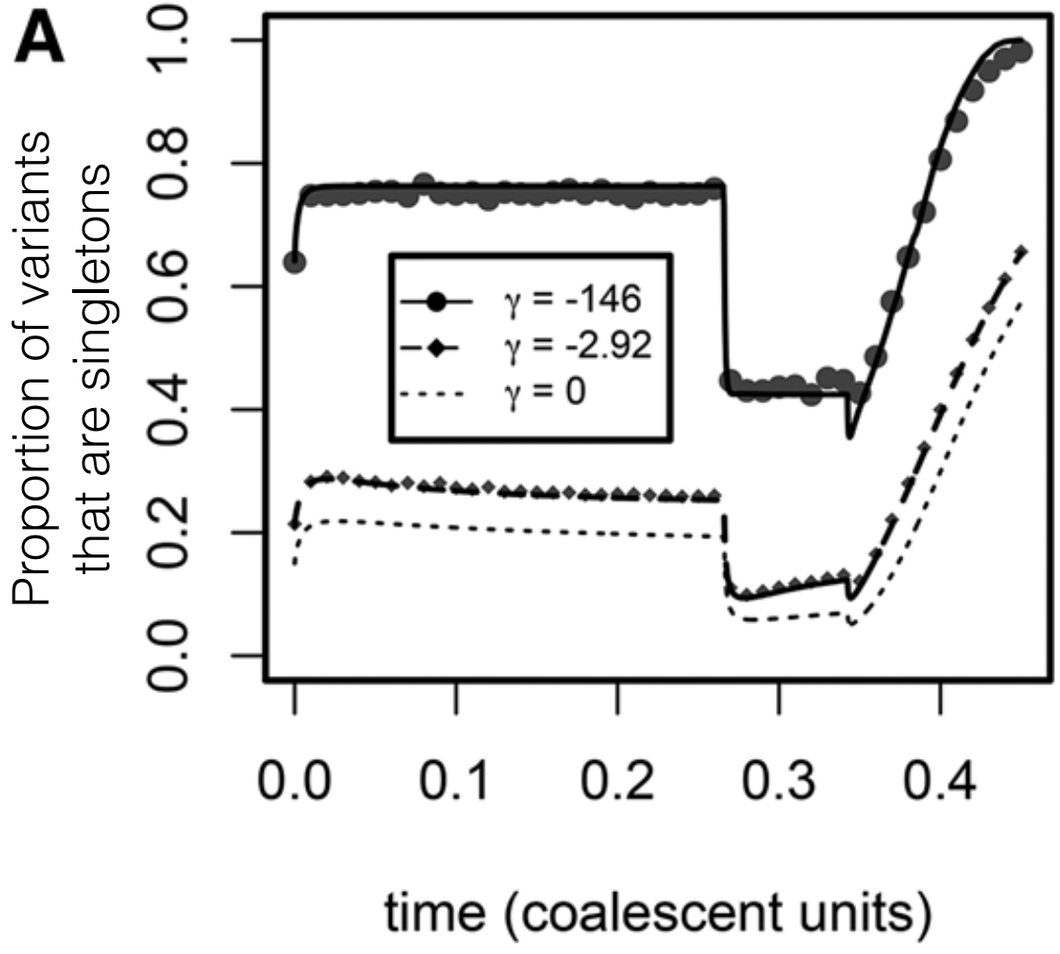


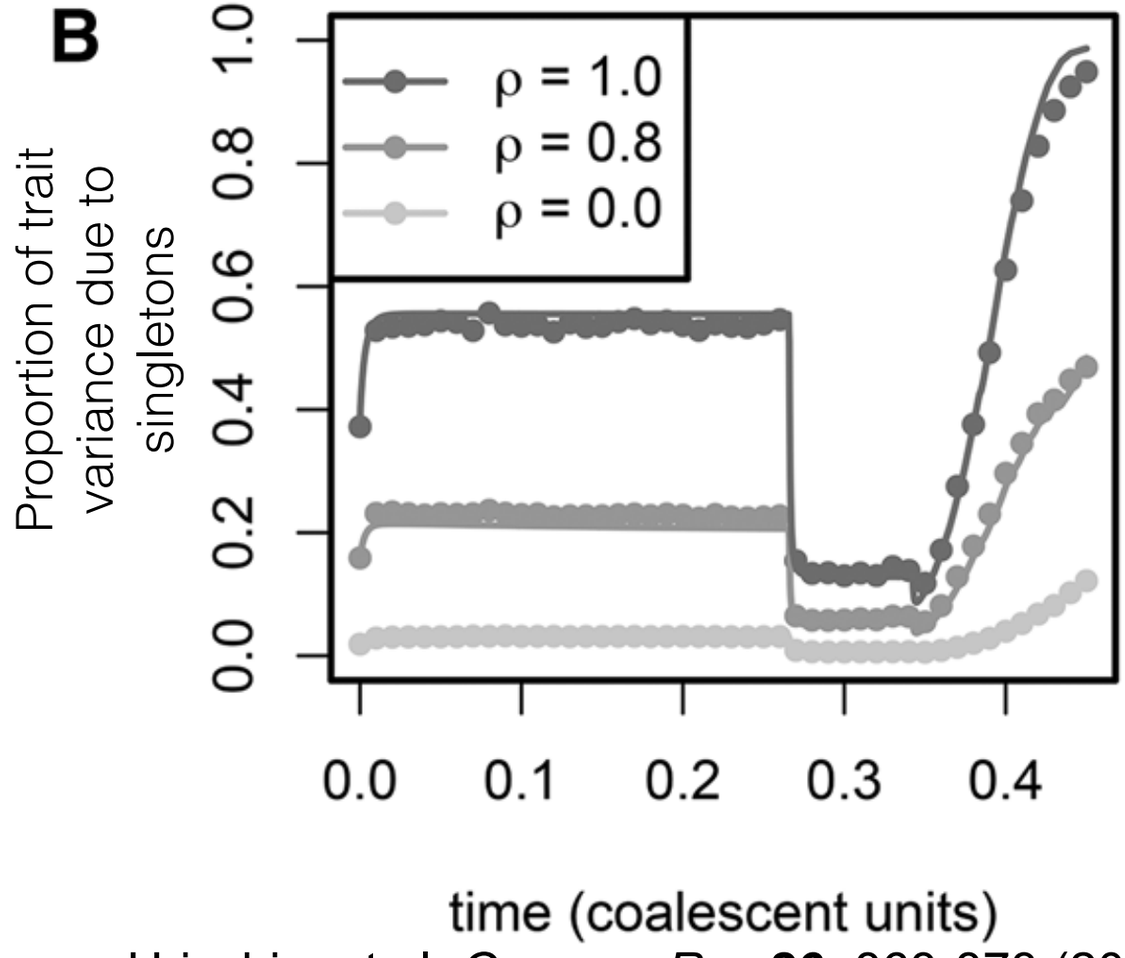
Implication: in some cases, largest effect alleles are very rare, so we may not detect them with GWAS!

Uricchio, et al. *Genome Res* **26**, 863-873 (2016).

Demography and selection matter!

- As populations expand and contract, or strength of selection changes, the frequency spectrum responds.
- This can **and should** impact the genetic architecture of traits!

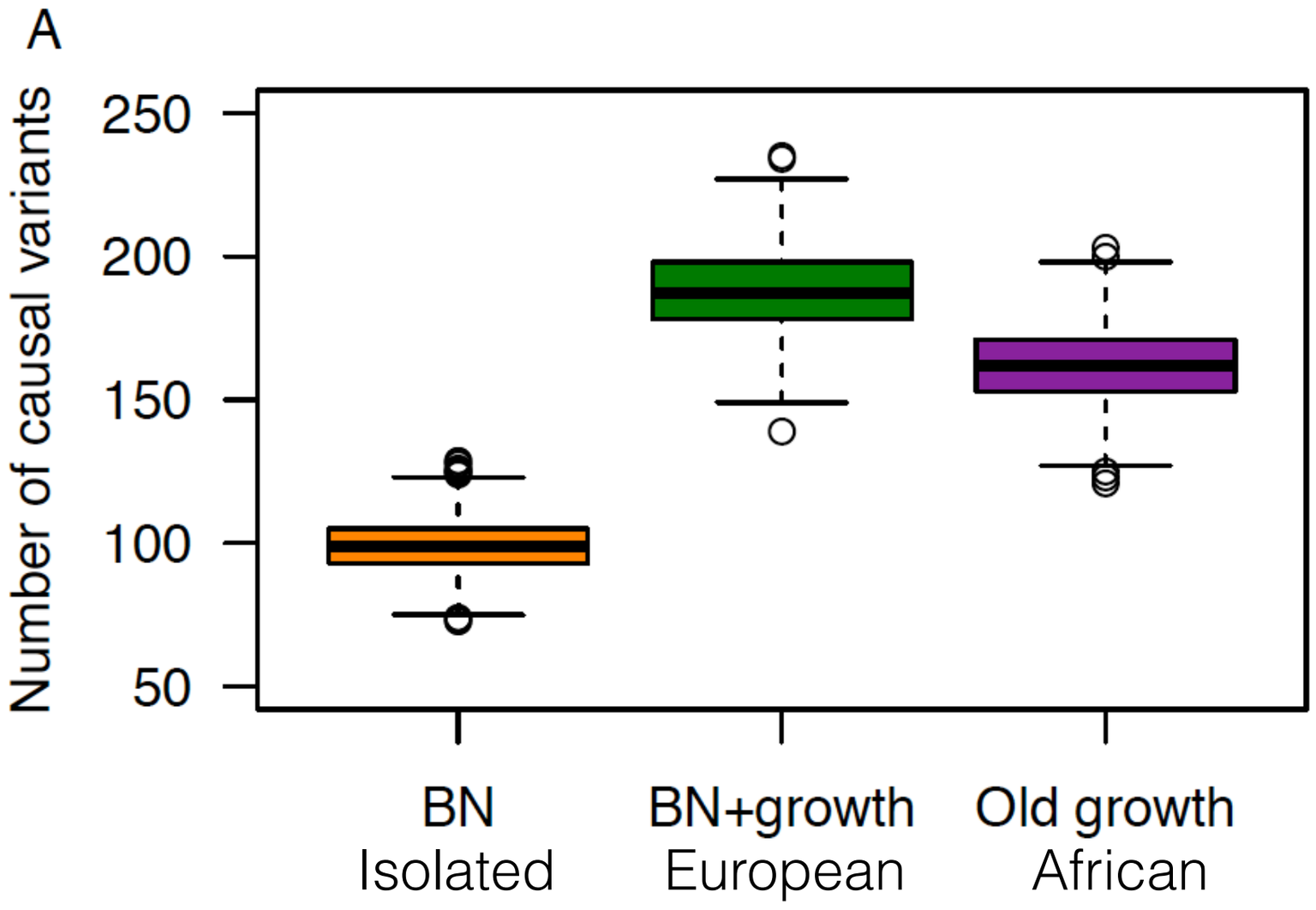




Uricchio, et al. *Genome Res* **26**, 863-873 (2016).

Demography and selection matter!

- Demography and selection also impacts the number of causal variants!



Lohmueller, *PLoS Genet* (2014).

Open Questions

- What does does the genetic architecture of a complex trait really look like?
- How many causal variants are there?
- Proportion of effects from rare/common alleles?
- Additive vs epistatic interactions?
- Pleiotropy?

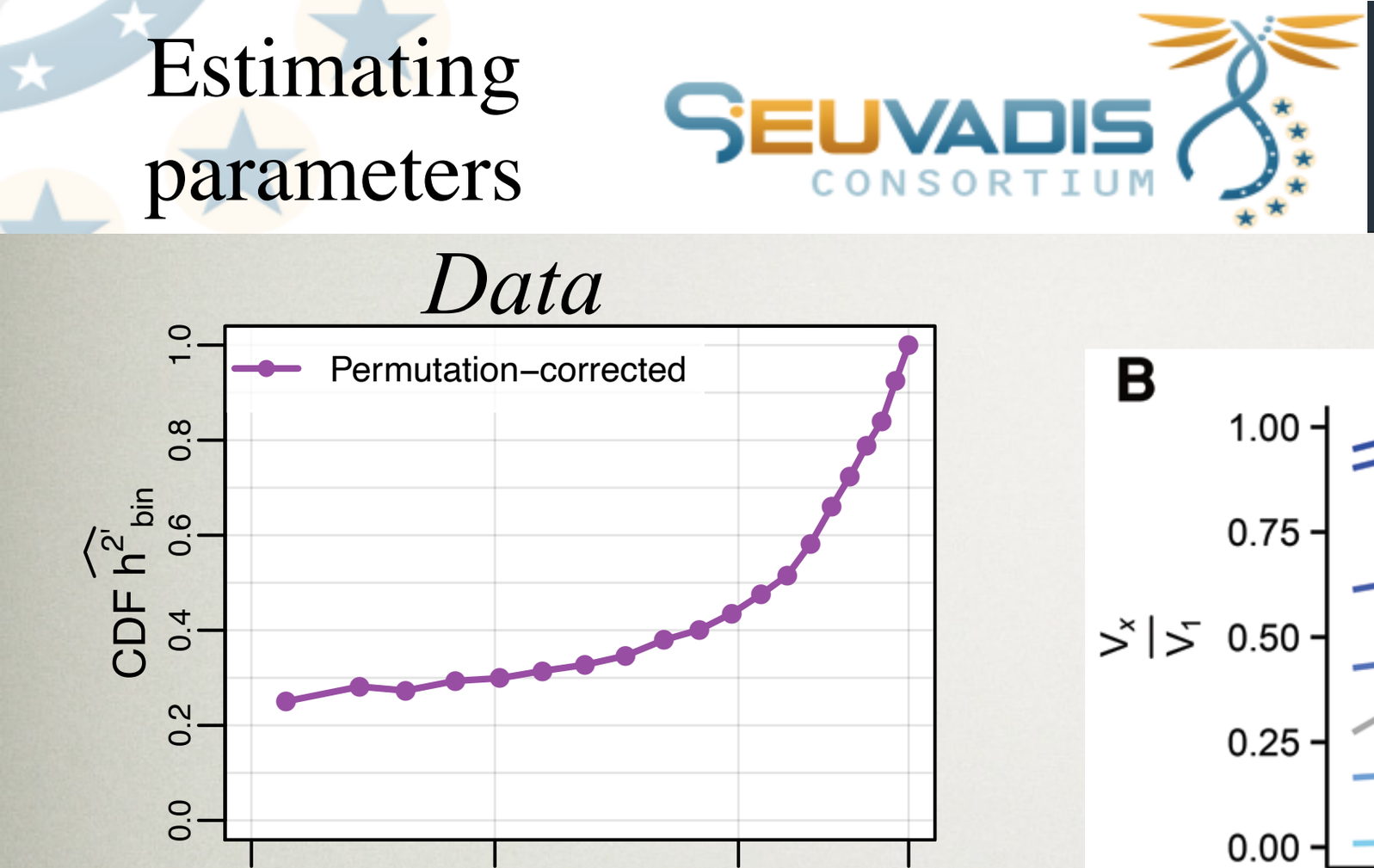
- Large-scale RNA sequencing + WGS
  - 4 European populations
  - 360 individuals
    - low coverage WGS + high coverage exome: Phase 3.
    - RNA-seq: median depth 58.3M reads
      - Gene expression:
        log2 transformed, median centered, and quantile normalized.
      - 10,077 unique genes.

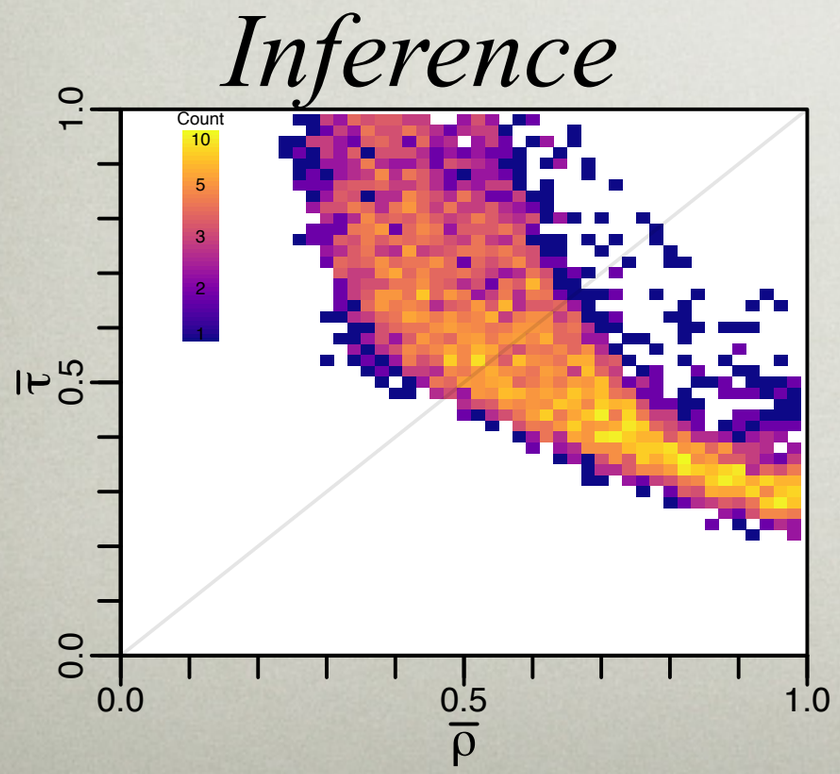
Hernandez, et al. (bioRxiv, 2019)

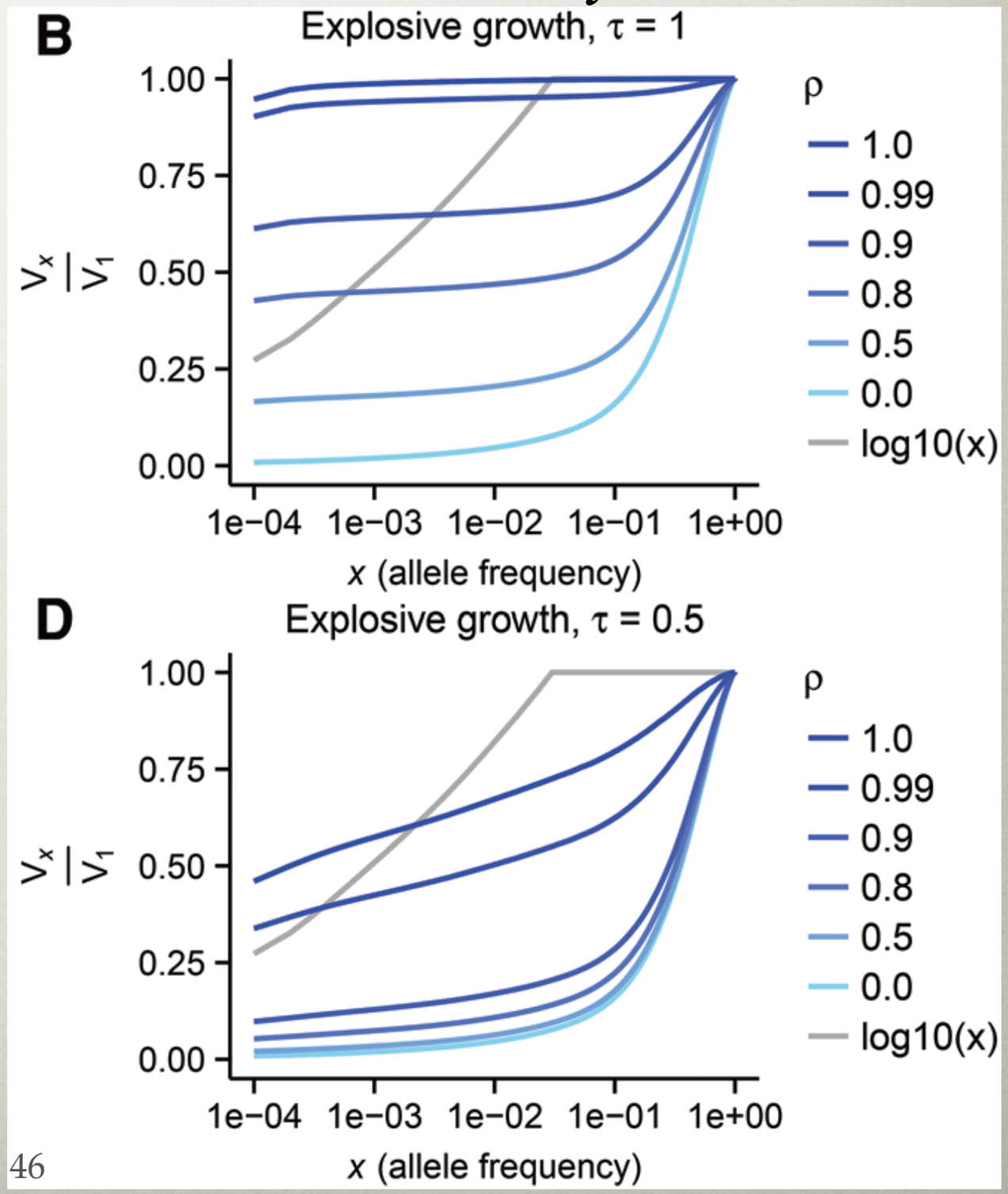
- Large-scale RNA sequencing + WGS
  - 4 European populations
  - 360 individuals
  - low coverage WGS + high coverage exome: Phase 3.
  - RNA-seq: median depth 58.3M reads
    - Gene expression:
      log2 transformed, median centered, and quantile normalized.
    - 10,077 unique genes.

- Our sample size is **small**, but can we learn anything about the **genetic basis of complex traits from these 10k genes**?
- Let's analyze heritability of gene expression due to *cis* variation (within 1Mb of gene)

Estimating parameters



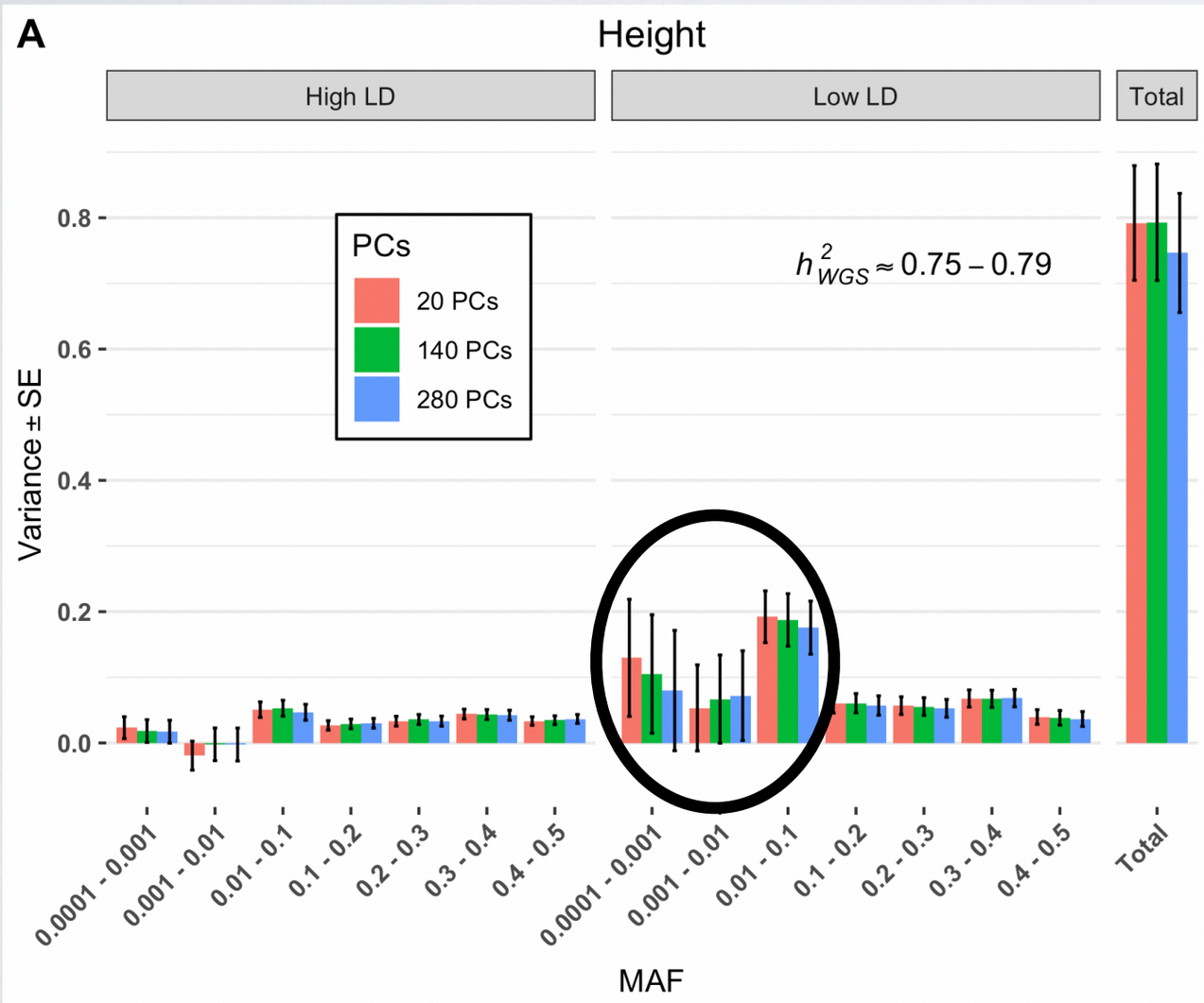


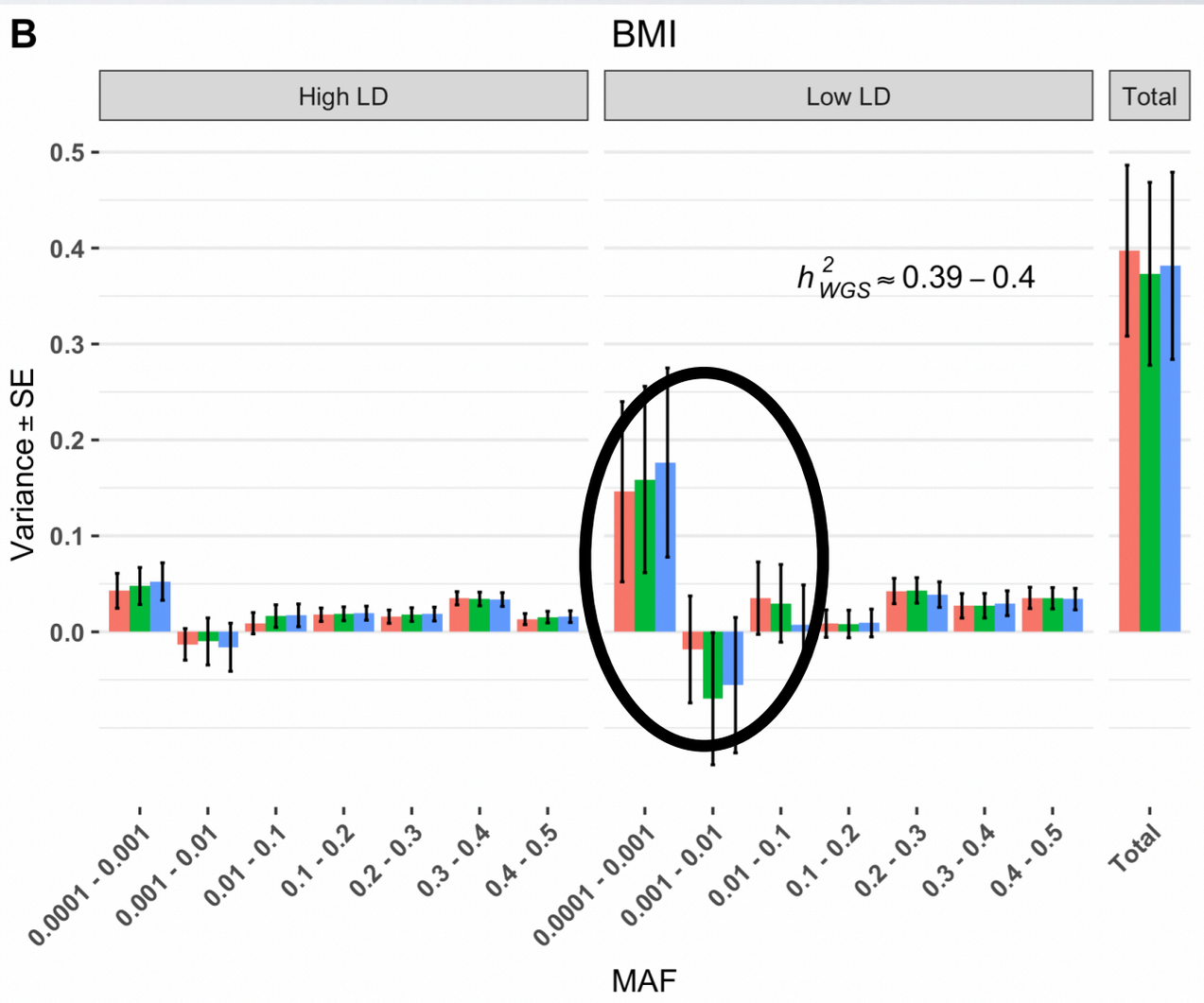


HUMAN HEIGHT AND BMI

n = 21,620 Individuals

Low MAF explains >50% of heritability





Wainschtein, et al. Recovery of trait heritability from whole genome sequence data. **bioRxiv**.

CONCLUSIONS

- Patterns of genetic variation within and between populations are shaped by their evolutionary history.
    - Demography: Growth/decline, migration/admixture
    - Natural selection
- These same evolutionary forces shape the genetic architecture of complex traits!
- Evolutionary forces that increase the incidence of rare variants, also increase the role of rare variants in complex traits!