Structure of heritability and its implications for mapping of rare variation

Yurii Aulchenko

yurii [dot] aulchenko [at] gmail [dot] com

slides available at

http://tinyurl.com/ge132013

Overview

- Missing heritability
- Expected composition of heritability
- Mapping rare variation
- Conclusions

Loci identified for complex traits

	# Loci			
	<2005	2008	2010	2012
Lipids	few	~30	95	+200
Height	0	~50	100+	+300

Loci identified for complex traits

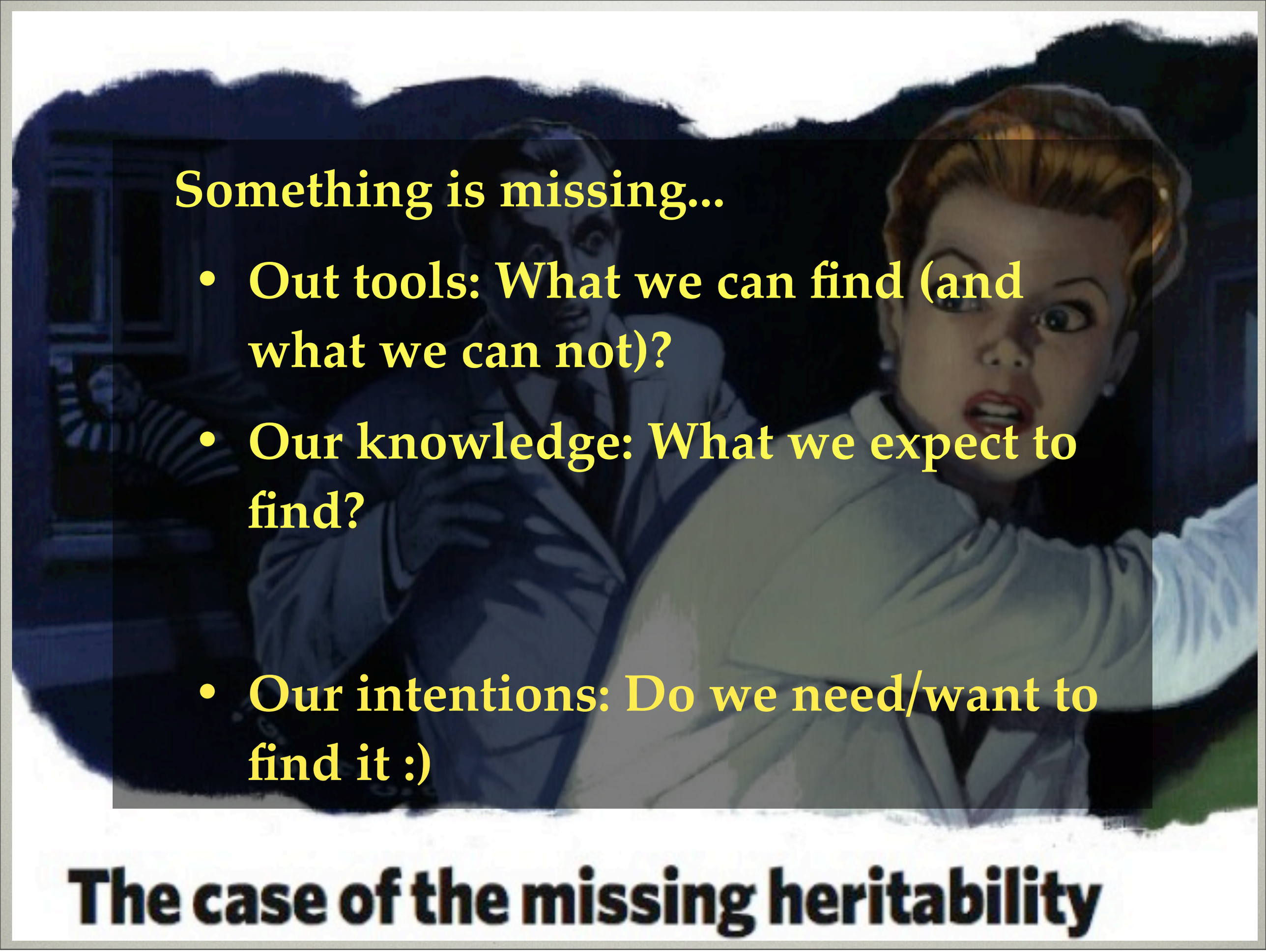
	# Loci			
	<2005	2008	2010	2012
Lipids	few	~30	95	+200
Height	0	~50	100+	+300

Total h2=0.3

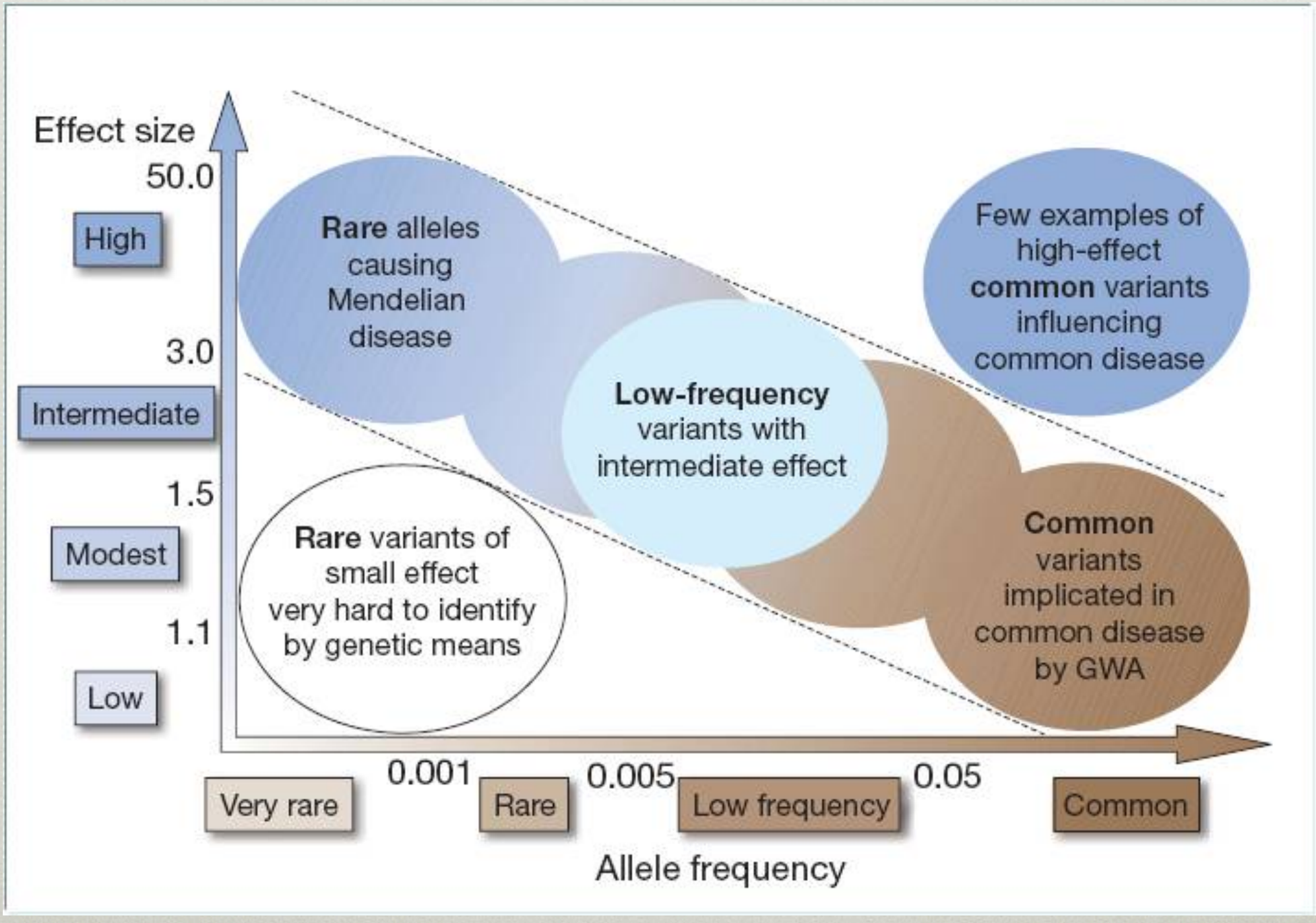
	%Var			
	<2005	2008	2010	2012
Lipids	~2%	5%	10%	+15%
Height	0	4%	8%	+10%

Total h2=0.9

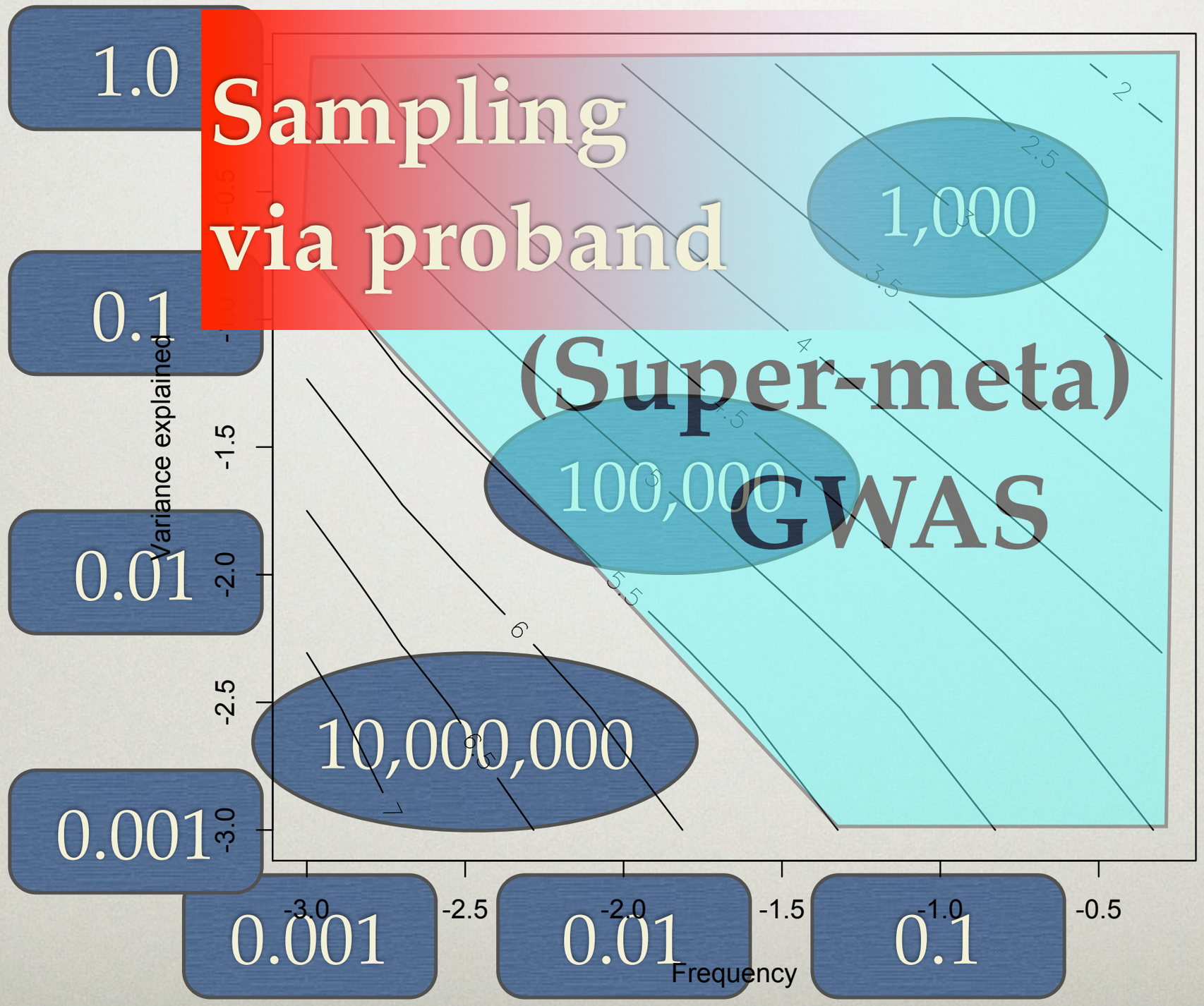
The case of the missing heritability


Something is missing...
• Out tools: What we can find (and what we can not)?
• Our knowledge: What we expect to find?
• Our intentions: Do we need/want to find it :)
The case of the missing heritability

What can we find? - statistically

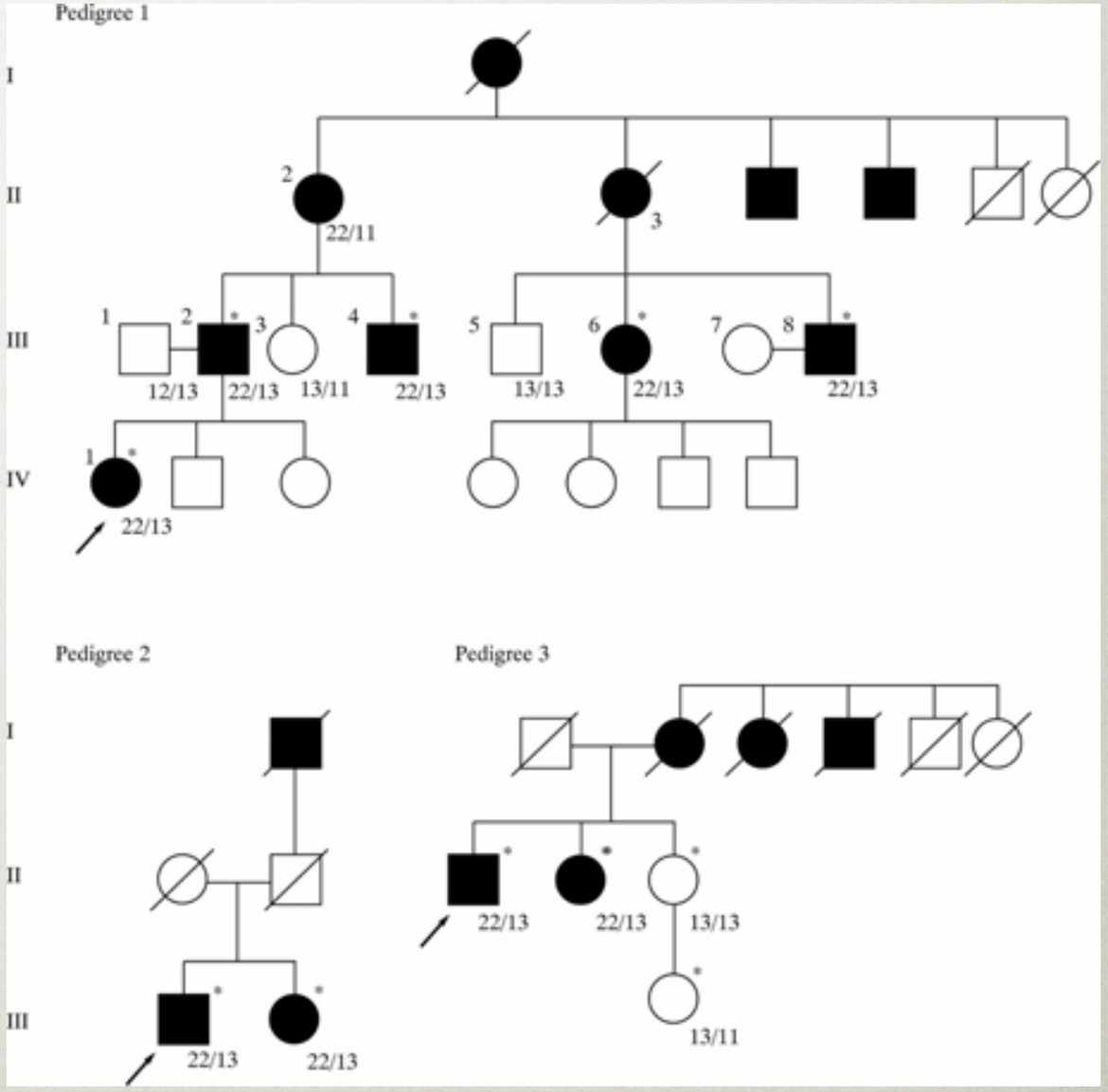


Why effect and frequency?



Enrichment design

- Effect is large
- Sampling via proband
- While mutation is rare in general population, it is prevalent in your study population



Summary for our tools

- Statistically, no way to solve variants from lower-left area
- Straightforward solutions are
  - Brute force: increase sample size and increase reachable area
  - Tricks to shift the problem to the right and/or up (by e.g. statistical or design means)

Overview

- Missing heritability
- **Expected composition of heritability**
- Mapping rare variation
- Conclusions

Composition of heritability

- Which alleles can reach high frequency?
- What proportion of heritability is explained by common/rare variation?

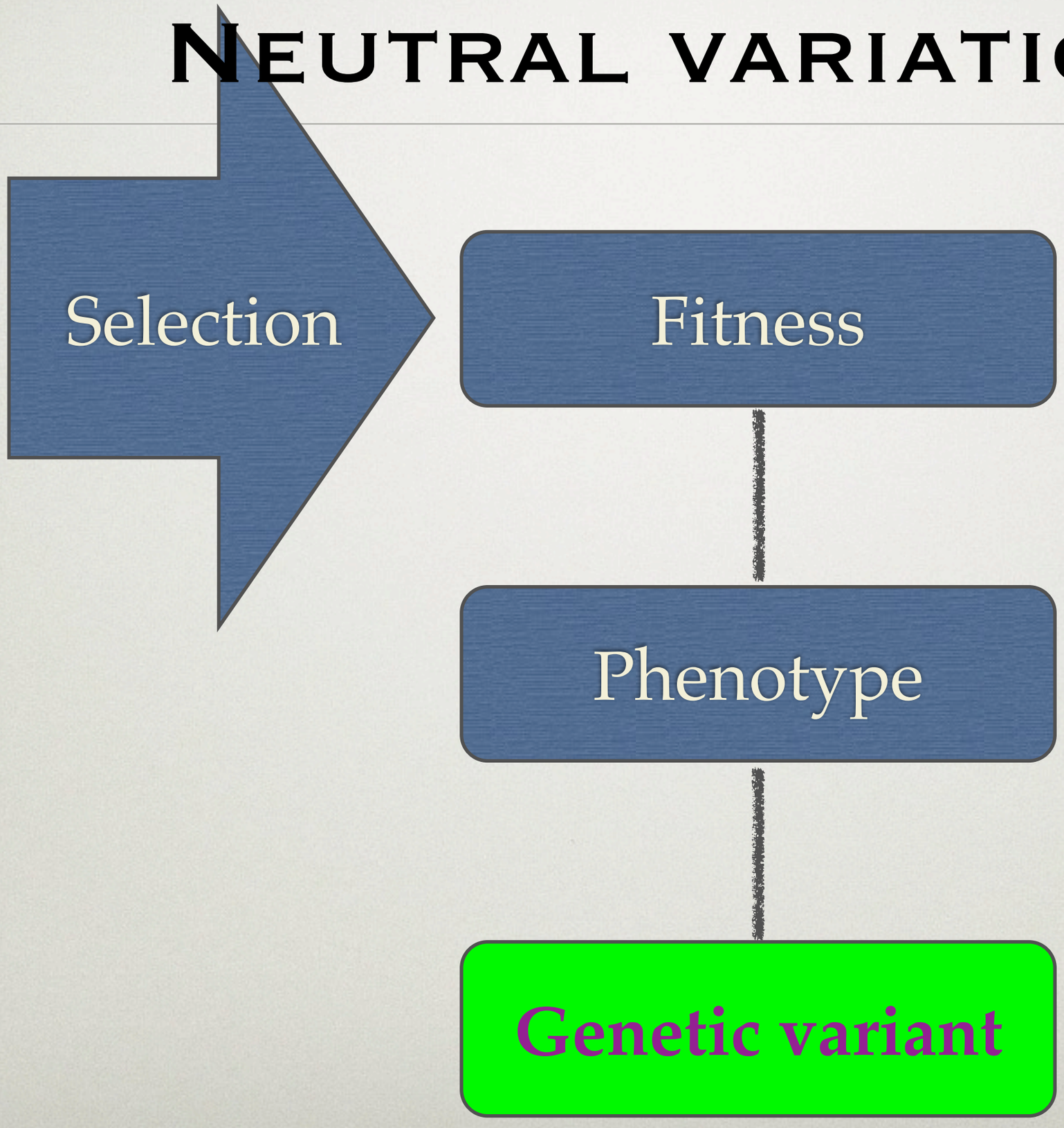
Selection

Fitness

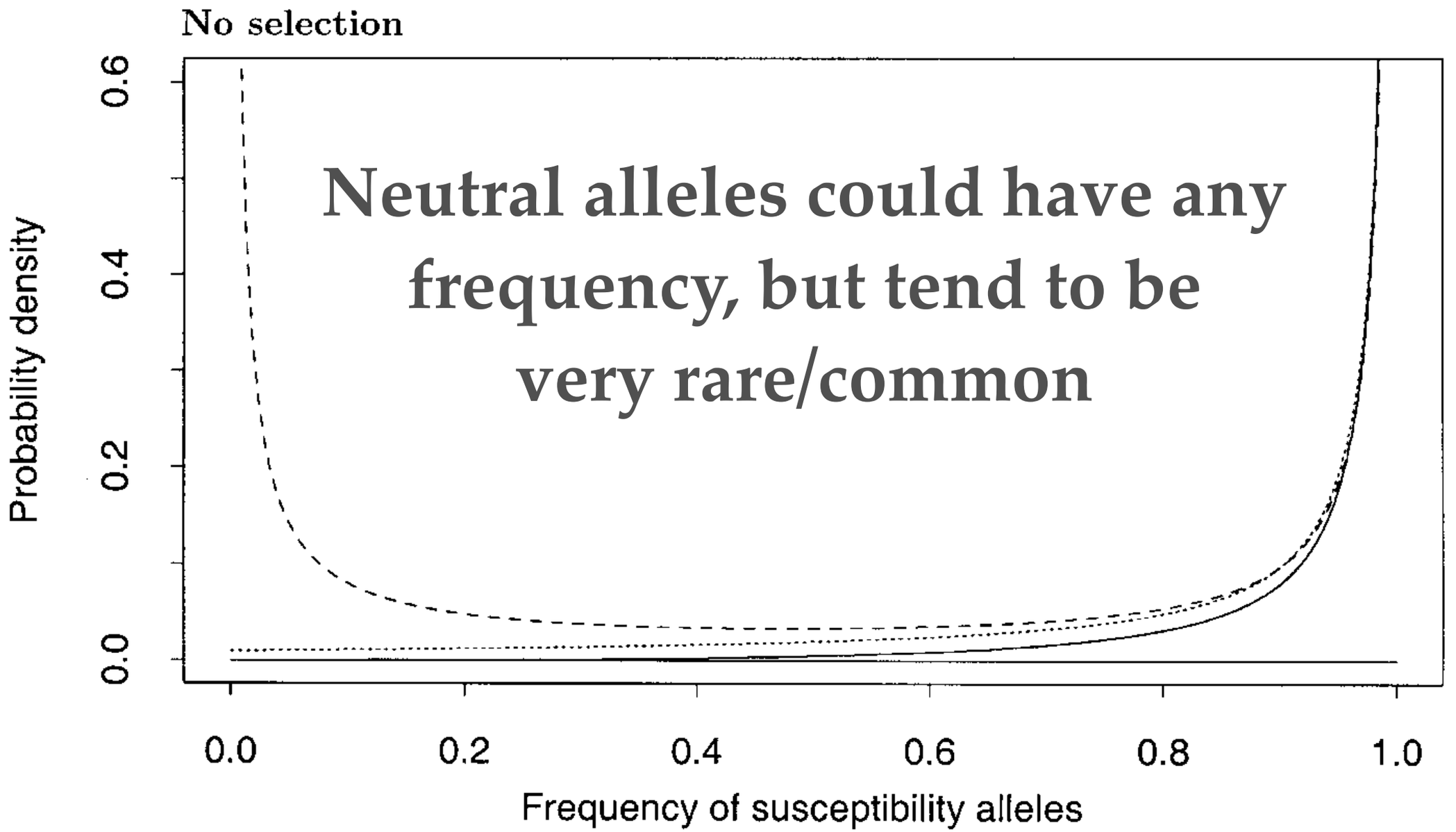
Phenotype

Genetic variant

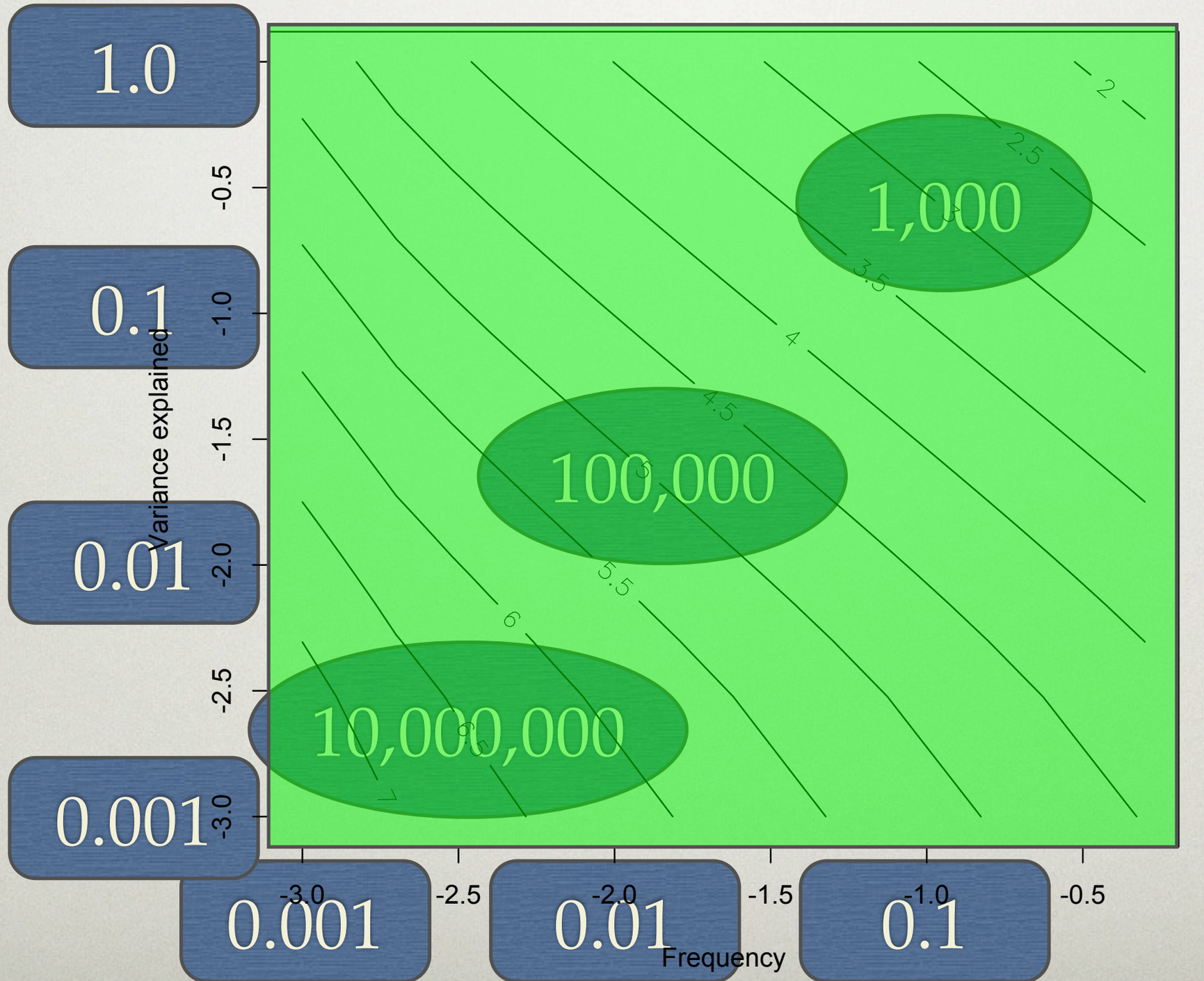
Neutral variation



Distribution of neutral allele frequency



Selectively neutral alleles



Selectively neutral alleles

Most of the (common and rare) variation we observe

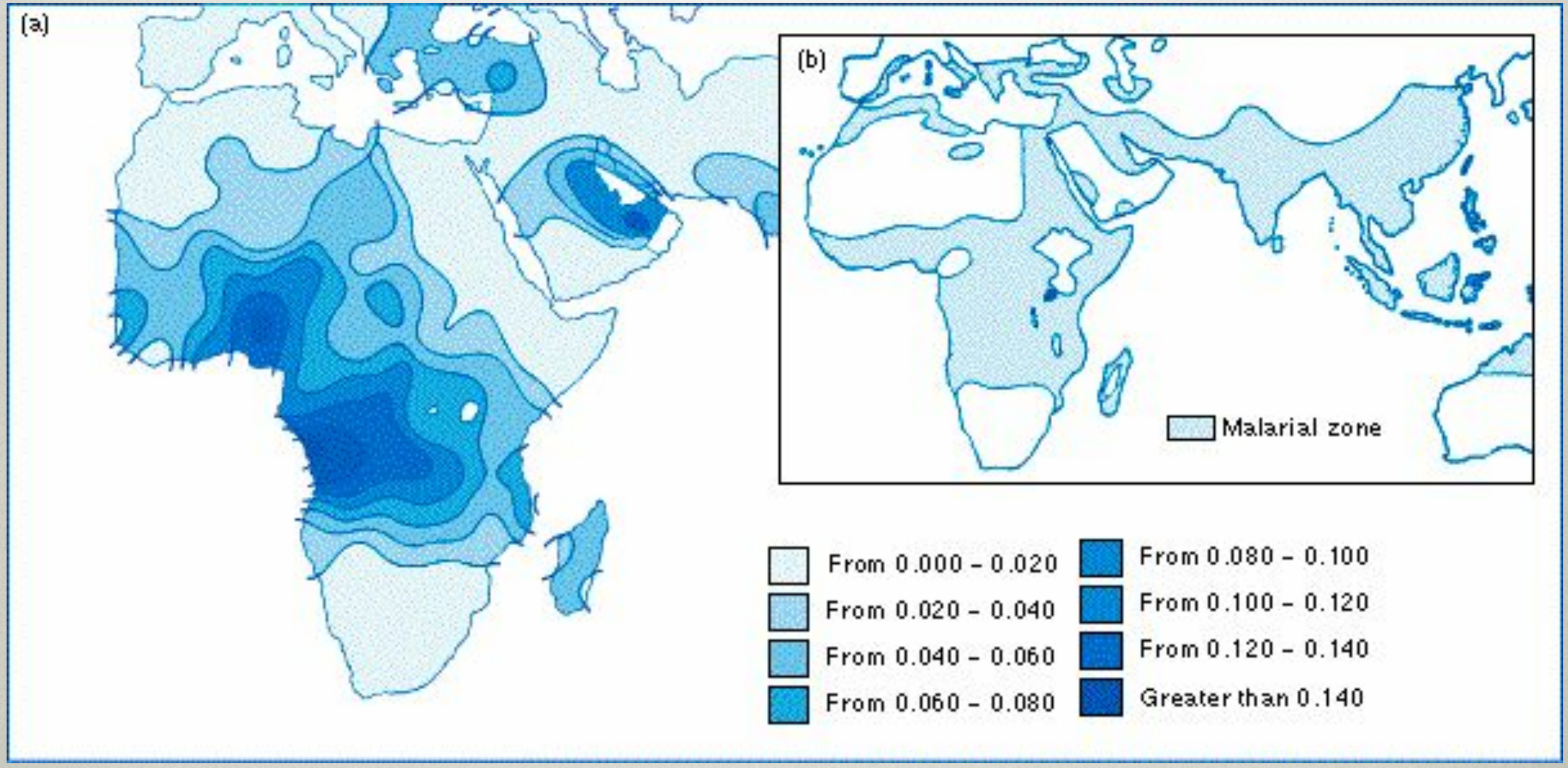
- Is selectively neutral
- Is not related to any phenotype

"Selectively neutral" (?) traits

- Eye color
- Late-onset diseases
  - Age-related macular degeneration (AMD)
  - e4 allele of APOE

Context-dependent selection

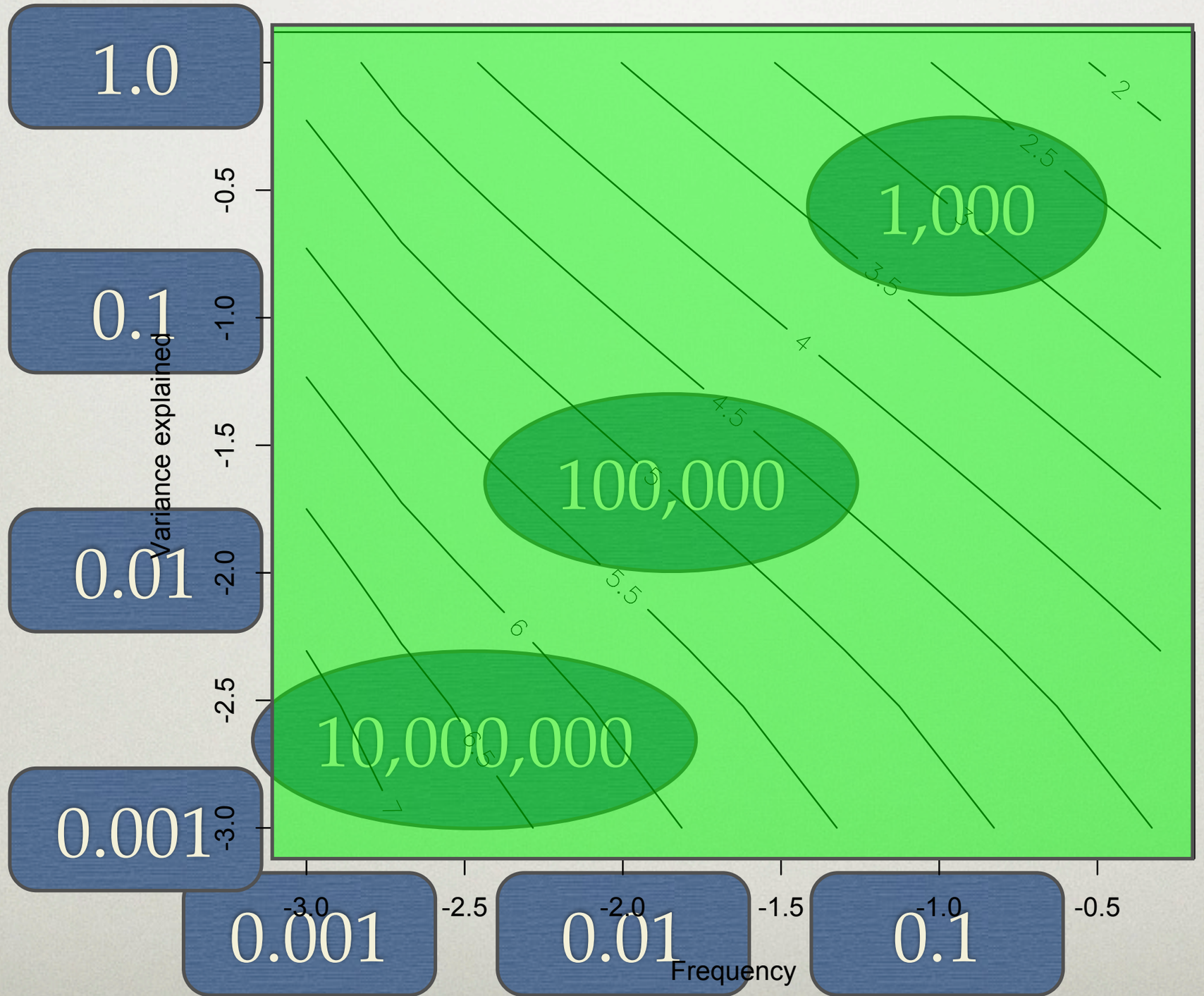
Heterozygote advantage



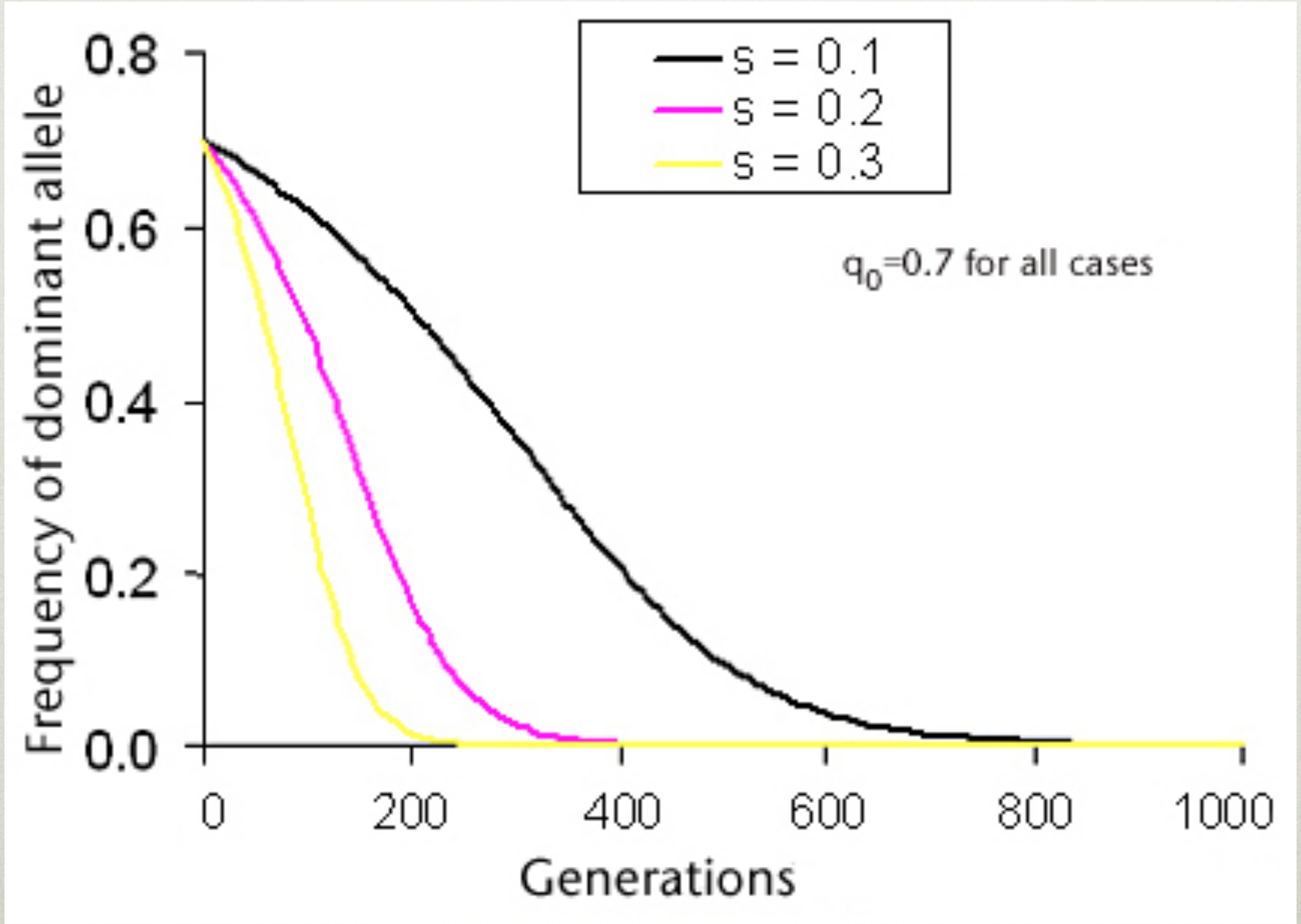
Antagonistic pleiotropy

- Positive effect early in life, negative later in life
- APOE e4 (?)

Context-dependently selected alleles

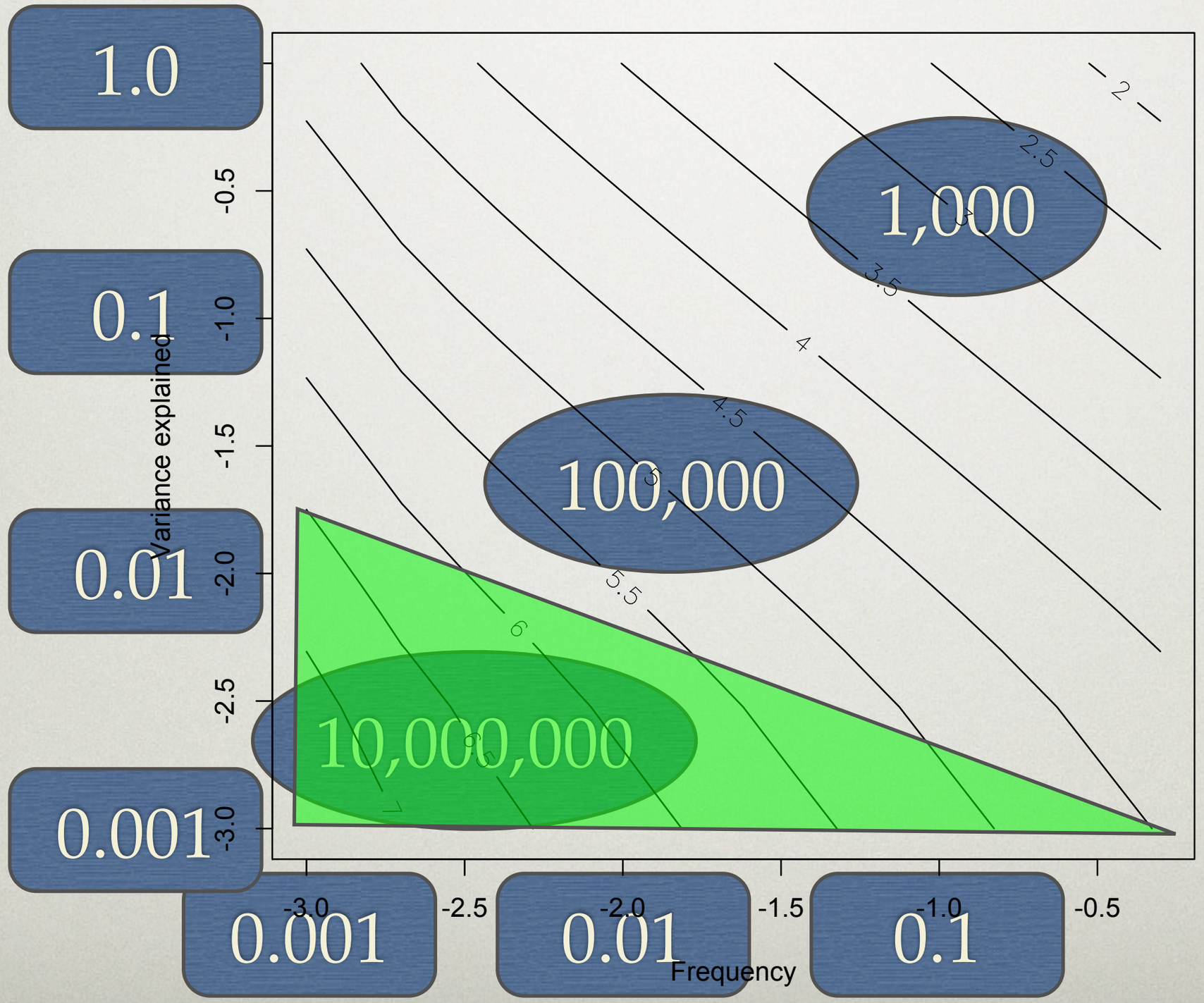


Selection of dominant allele

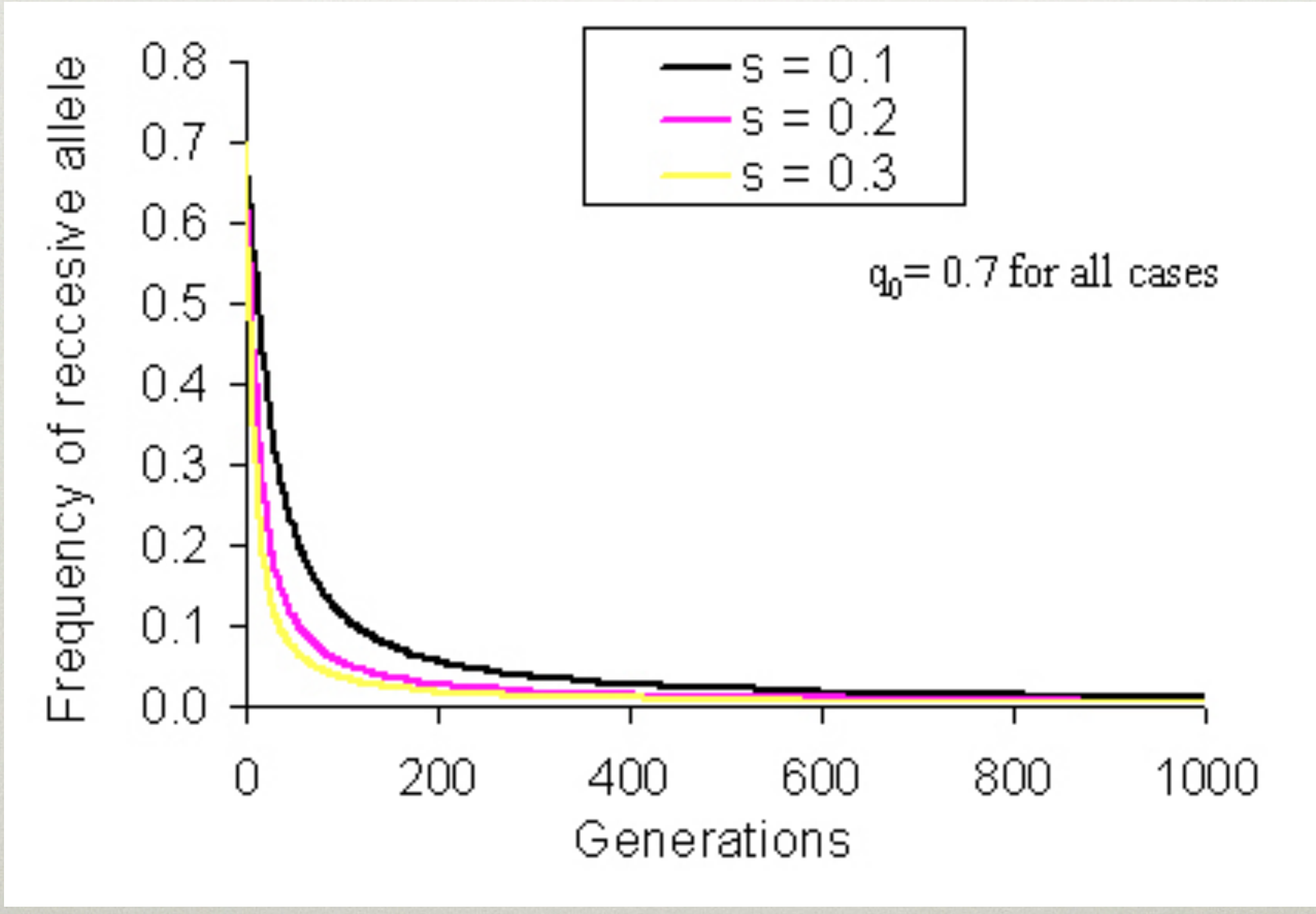


Source: http://www.apsnet.org/edcenter/advanced/topics/PopGenetics/Pages/NaturalSelection.aspx

Dominant/additive alleles of small effects

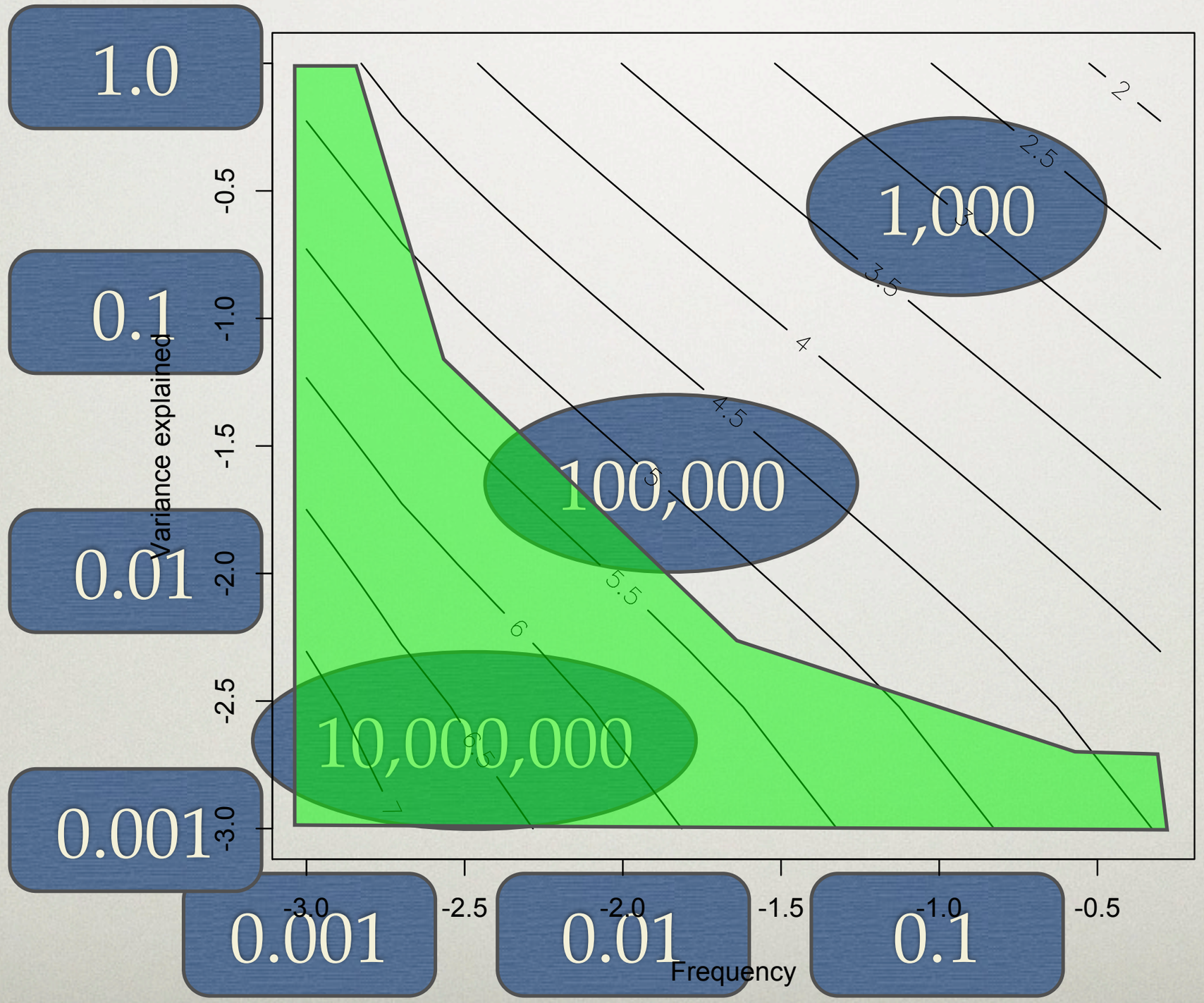


Selection of recessive allele

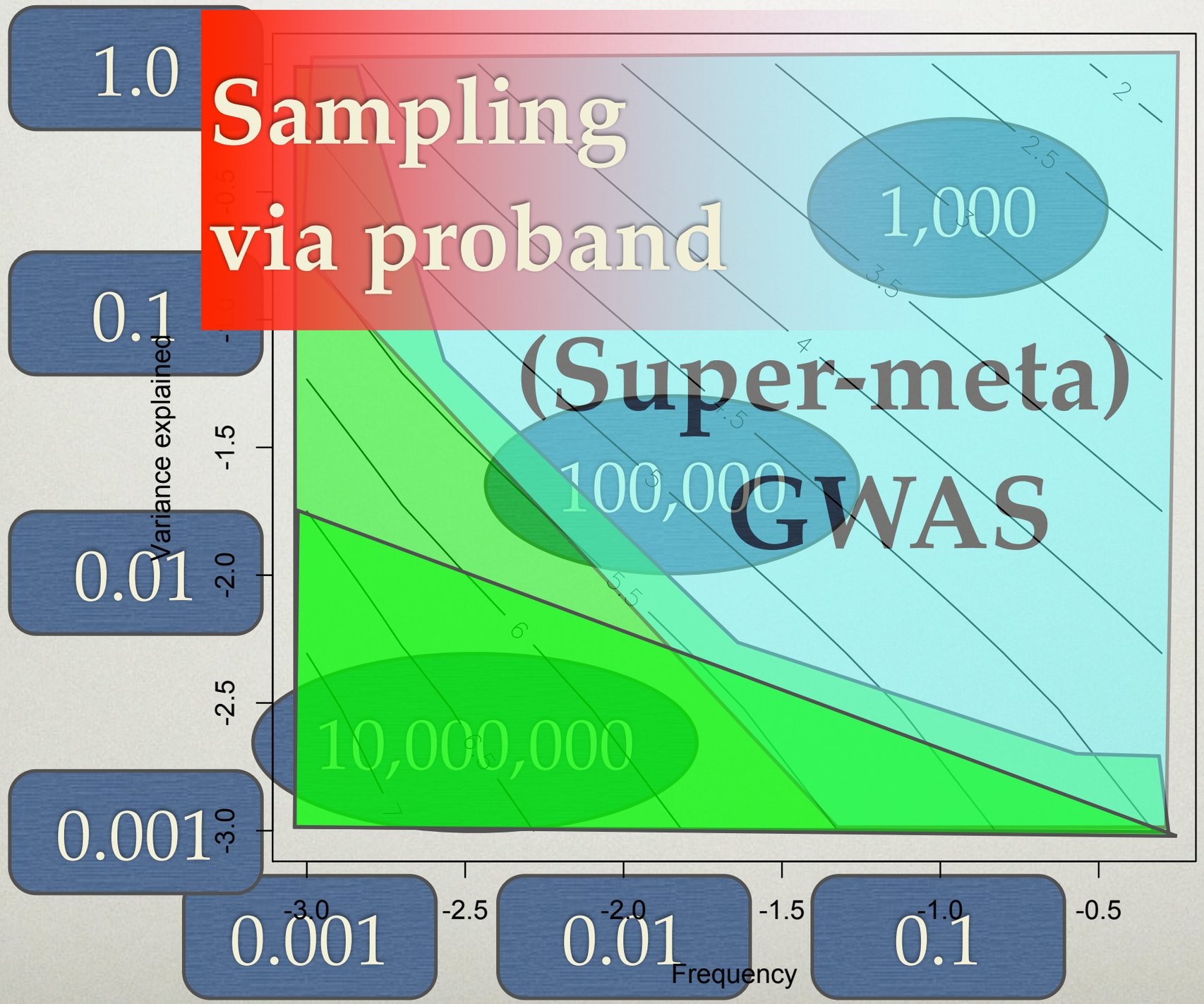


Source: http://www.apsnet.org/edcenter/advanced/topics/PopGenetics/Pages/NaturalSelection.aspx

Selected recessive alleles



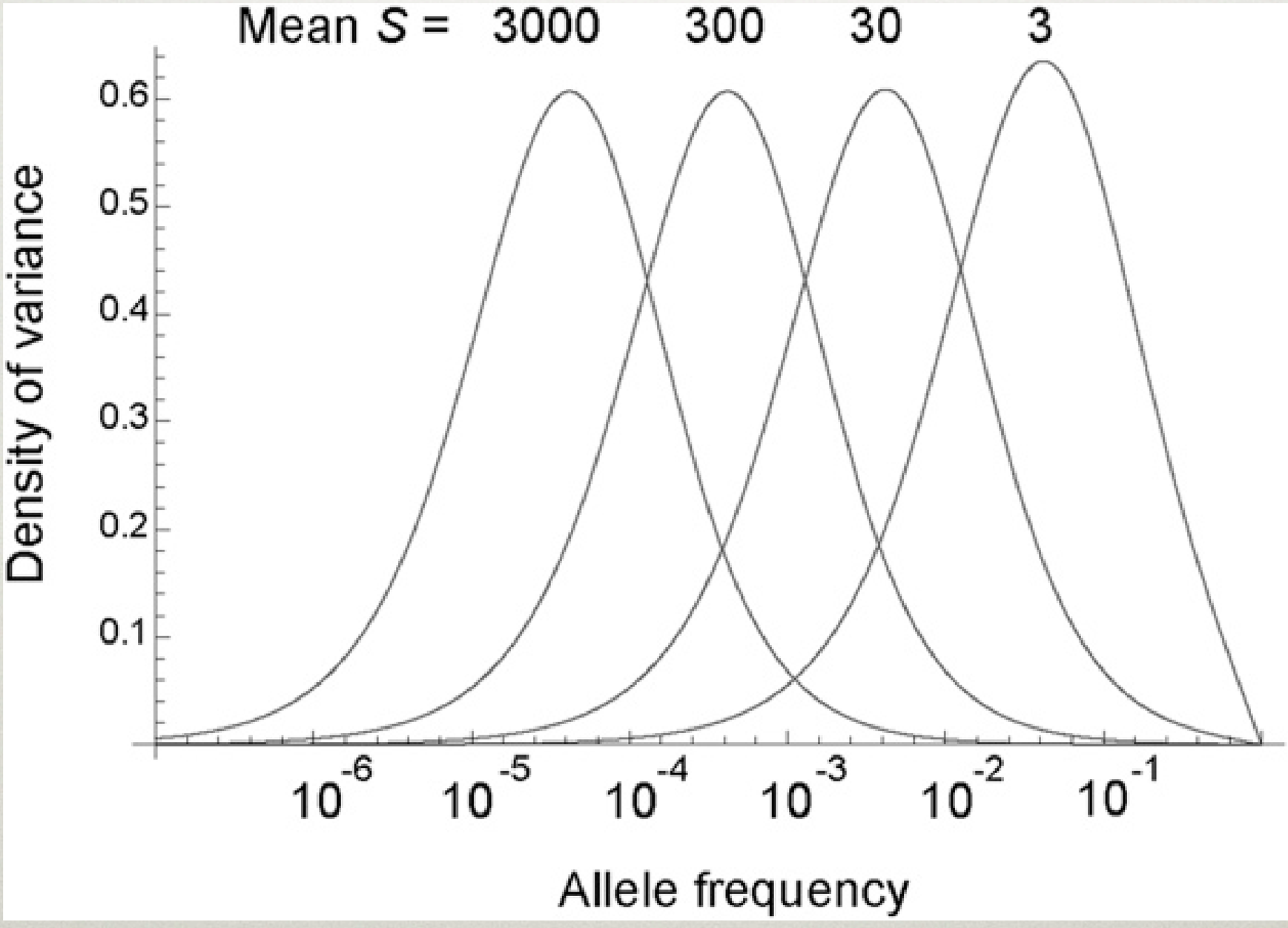
Selected recessive alleles



Expectation

- For "neutral" traits, or traits with context-dependent selection, alleles of large effect could reach high frequencies
- As soon as alleles are selected, they are expected to be rare. For traits directly related to fitness, you expect that
  - Alleles with large effect are rare
  - Common alleles have small effects

Composition of complex traits: common vs rare



Eyre-Walker, PNAS, 2010

Contribution of rare vs. common: estimates

Table 1 Estimates of the variance explained by all autosomal SNPs for height, BMI, vWF and QTi

		No PC[a]		10 PCs[b]			
Trait	*n*	h^2_G (s.e.)[c]	*P*	h^2_G (s.e.)	*P*	Heritability[d]	GWAS[e]
Height	11,576	0.448 (0.029)	4.5×10^{-69}	0.419 (0.030)	7.9×10^{-48}	80–90%[32]	~10%[23]
BMI	11,558	0.165 (0.029)	3.0×10^{-10}	0.159 (0.029)	5.3×10^{-9}	42–80%[25,26]	~1.5%[14]
vWF	6,641	0.252 (0.051)	1.6×10^{-7}	0.254 (0.051)	2.0×10^{-7}	66–75%[33,34]	~13%[15]
QTi	6,567	0.209 (0.050)	3.1×10^{-6}	0.168 (0.052)	5.0×10^{-4}	37–60%[35,36]	~7%[16]

The traits vWF and QTi were available in the ARIC cohort only.
[a]Without principal component adjustment. [b]Adjustment with the first 10 principal components from principal component analysis. [c]Estimate of variance explained by all autosomal SNPs. [d]Narrow sense heritability estimate from family or twin studies from the literature. [e]Variance explained by GWAS associated loci from the literature. PC, principal component; s.e., standard error.

Yang et al., Nat Genet, 2011

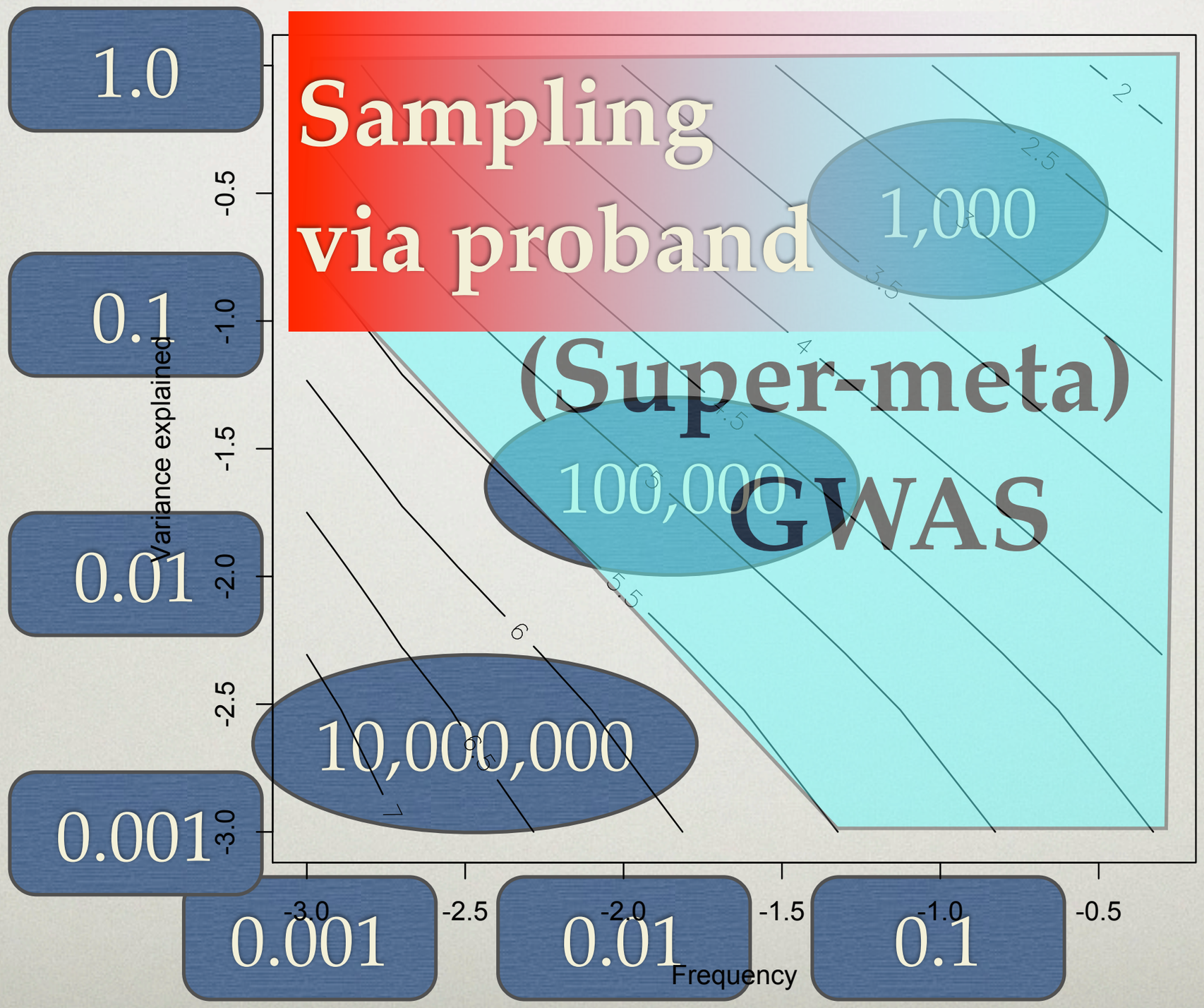
Conclusions

- Some (essential) part of genetic variance is explained by common variants acting in additive manner
- How big is this part, depends on evolutionary history of the trait
- Residual heritability may be explained by different mechanisms, most likely rare variation

Overview

- Missing heritability
- Expected composition of heritability
- **Mapping rare variation**
- Conclusions

Pushing the frequency up



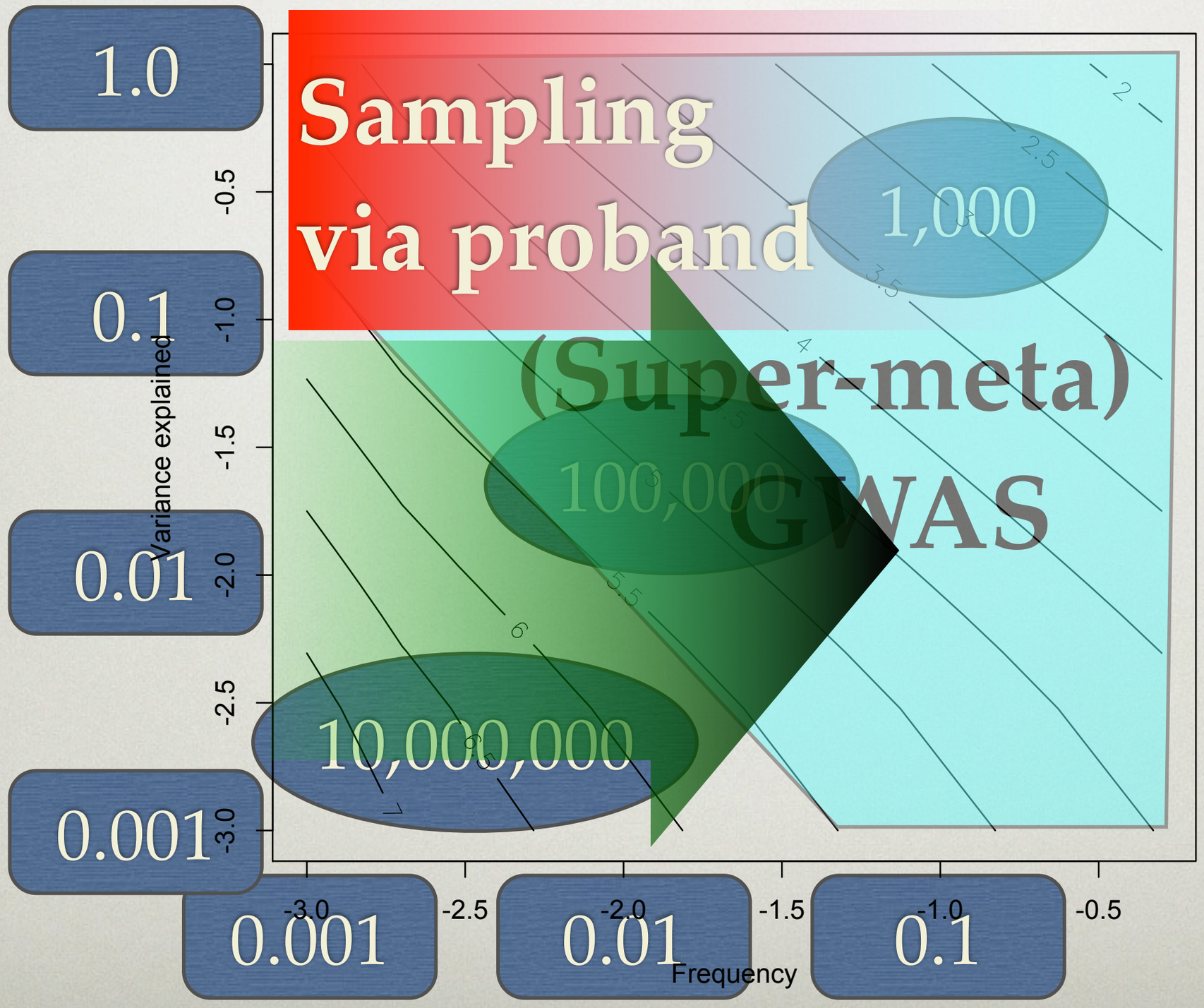
Summary for our tools

- Statistically, no way to solve variants from lower-left area
- Straightforward solutions are
  - Brute force: increase sample size and increase reachable area
  - **Tricks to shift the problem to the right and/or up (by e.g. statistical or design means)**

Rare variation tests

- Are 'global' tests of association between variation in a genomic region and a trait
- Why? In a way, they try to bundle different variants together in a single compound "alleles" with higher frequency

Pushing the frequency up



In order to combine effects

Rare variation tests make assumptions about

- Distribution (possibly conditional) of the effect
- Location of causative variants (region)
- Model of interaction between alleles of the same locus

A simple collapsing method

- Distribution of the effect: *Rare means deleterious*
- Location of causative variants (region): *All exomes of a gene: looking for strong effects*
- Model of interaction between alleles of the same locus: *Presence of one or more rare variant(s) leads to change of phenotype*

Questioning assumptions

- Distribution of the effect: *Rare means deleterious? More rare means bigger deleterious effect? What is the exact relation? What about quantitative traits? Can we include functional information?*
- Location of causative variants (region): *Why not introns and regulatory regions? What about enhancers? Expected effect different for different regions?*
- Model of interaction between alleles of the same locus: *One rare is enough? May be effects add up? Recessive / compound heterozygosity model?*

What is the best method to detect rare variants?

OPEN ACCESS Freely available online

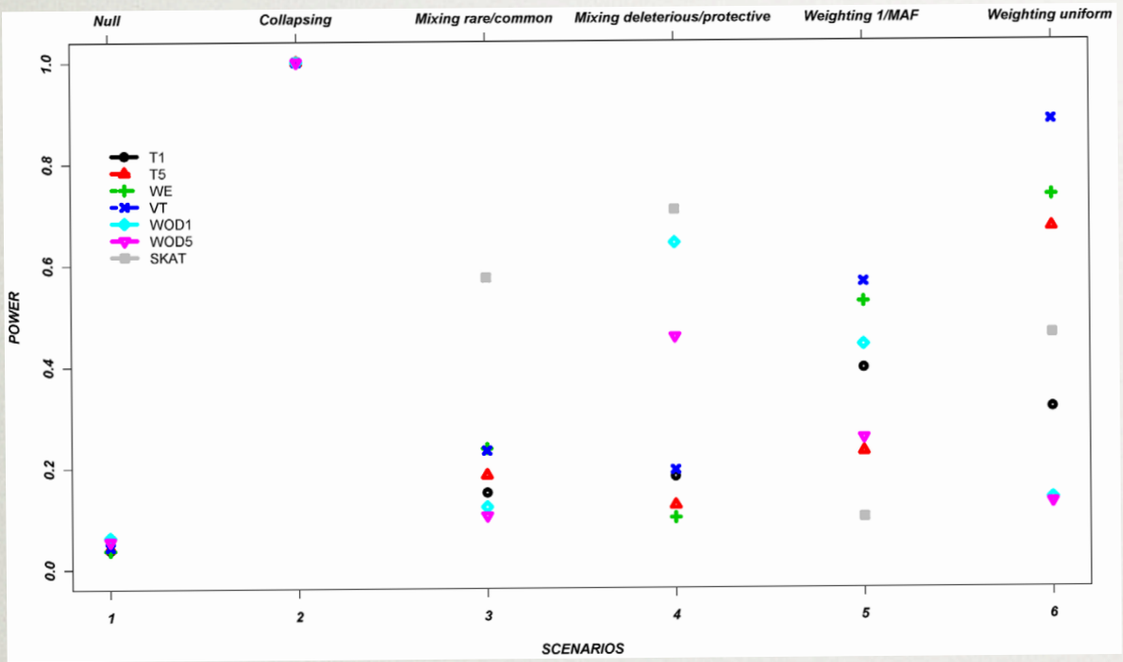
PLoS GENETICS

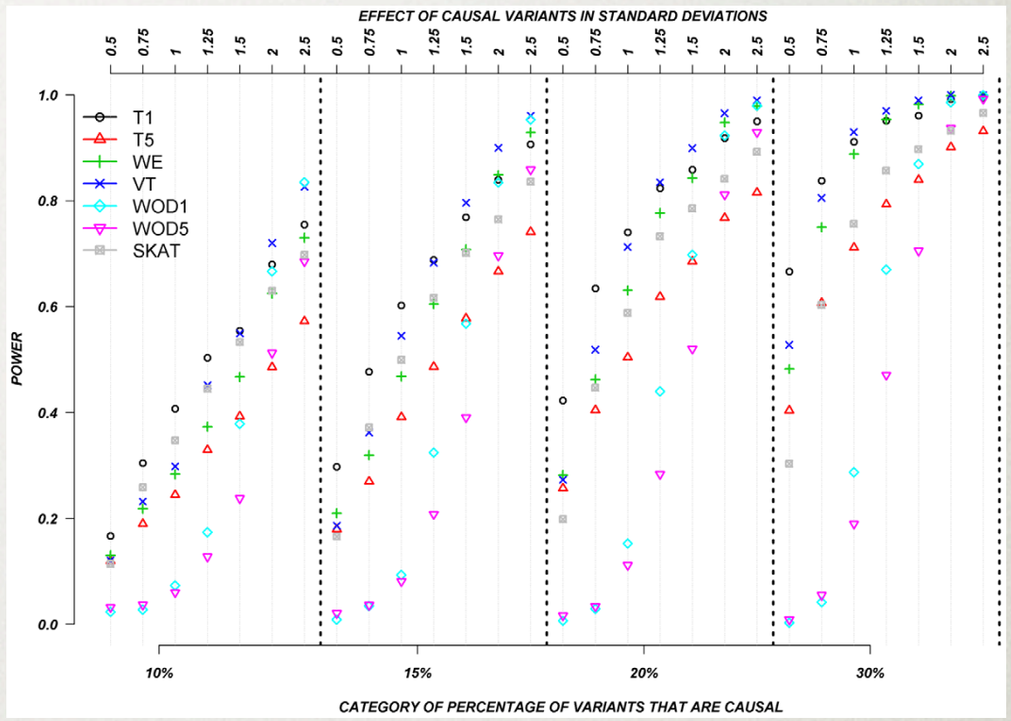
The Empirical Power of Rare Variant Association Methods: Results from Sanger Sequencing in 1,998 Individuals

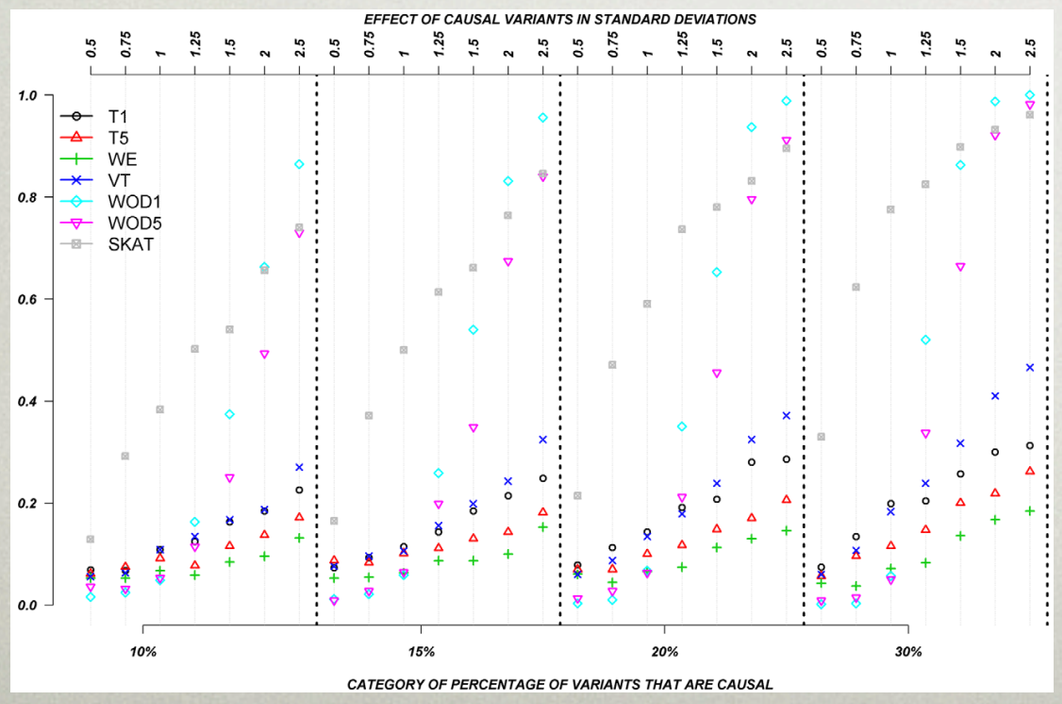
Martin Ladouceur[1,2], Zari Dastani[2,3], Yurii S. Aulchenko[4,5], Celia M. T. Greenwood[2,3,6], J. Brent Richards[1,2,7,8]*

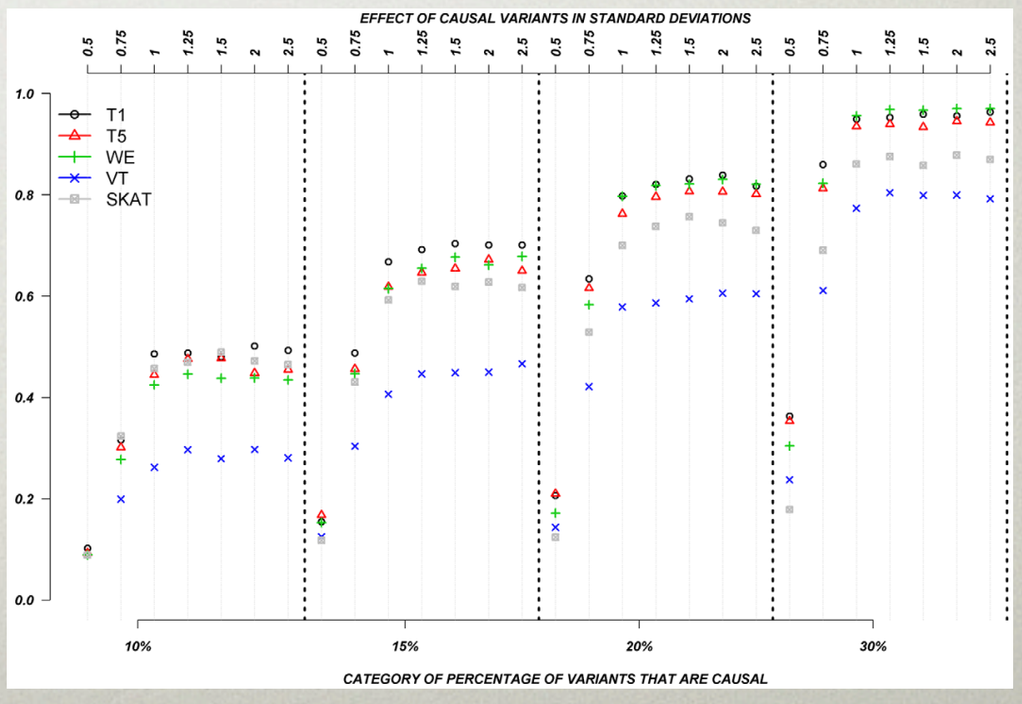
1 Department of Human Genetics, McGill University, Montreal, Canada, **2** Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Canada, **3** Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada, **4** Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands, **5** Institute of Cytology and Genetics SD RAS, Novosibirsk, Russia, **6** Department of Oncology, McGill University, Montreal, Canada, **7** Department of Medicine, Jewish General Hospital, McGill University, Montreal, Canada, **8** Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom

What is the best method to detect rare variants?

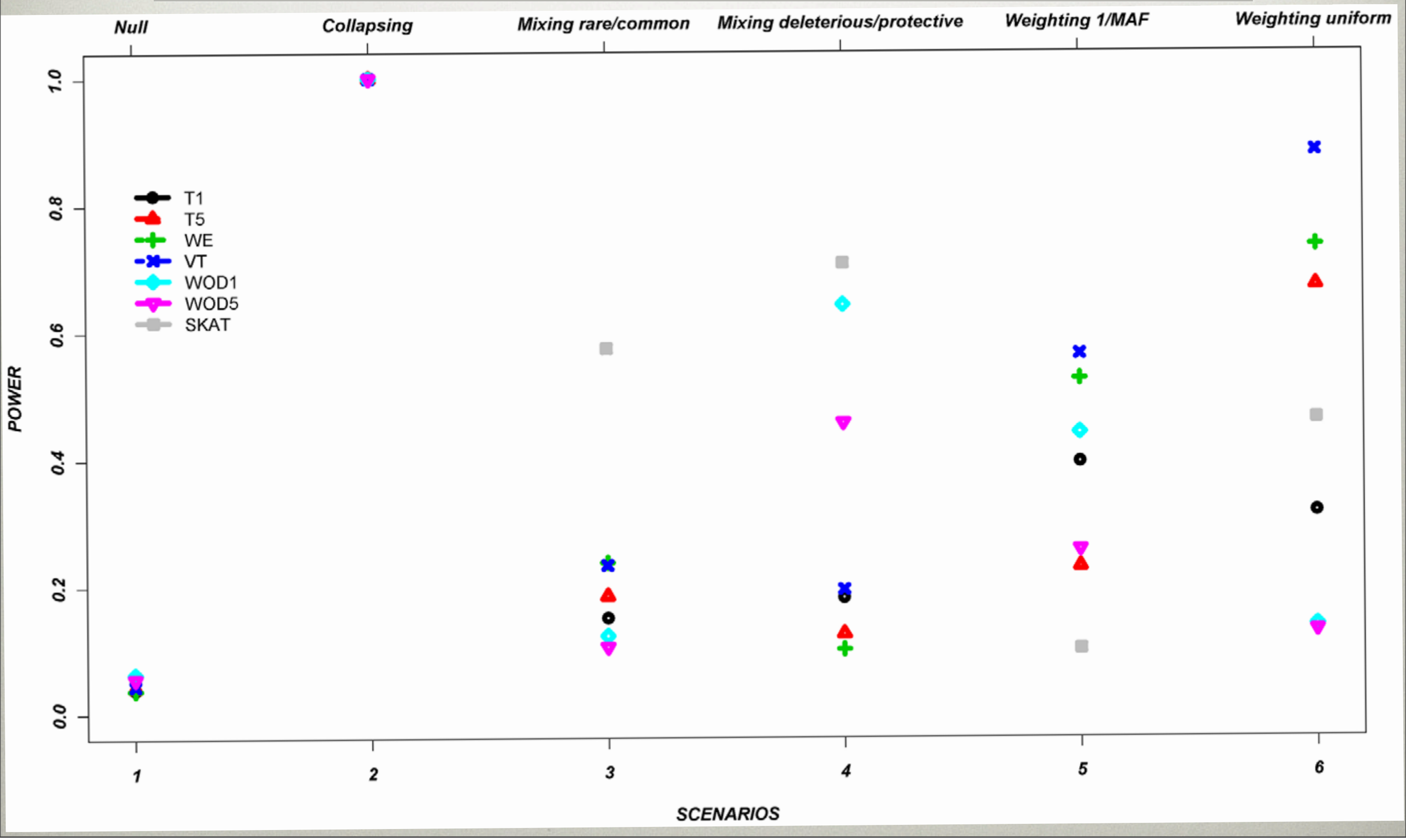




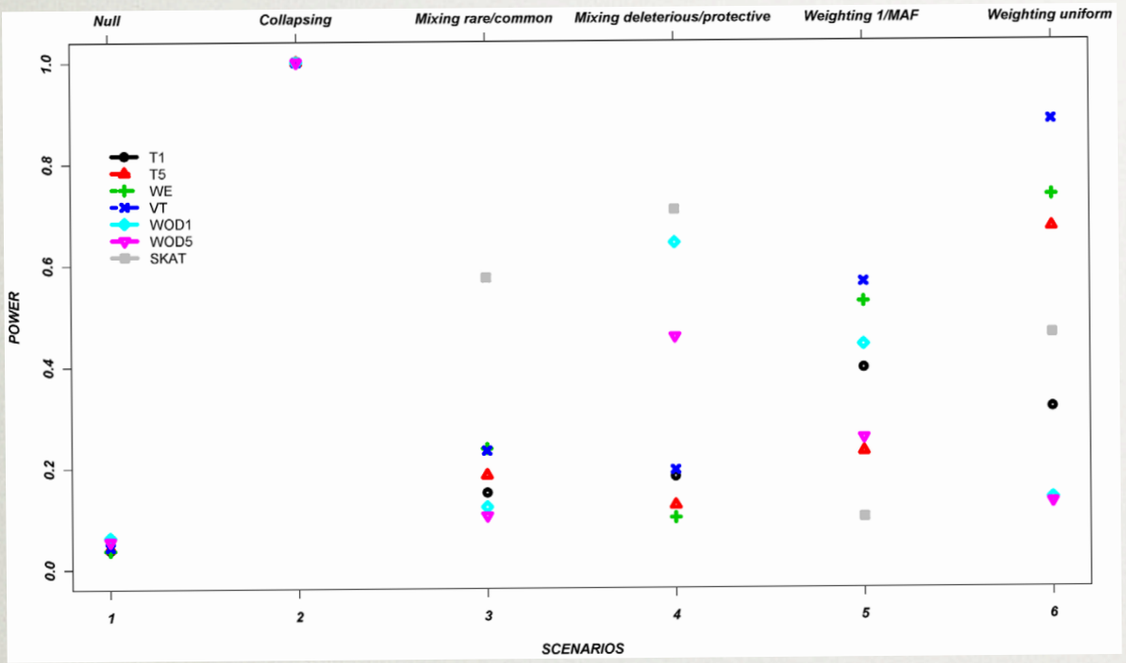


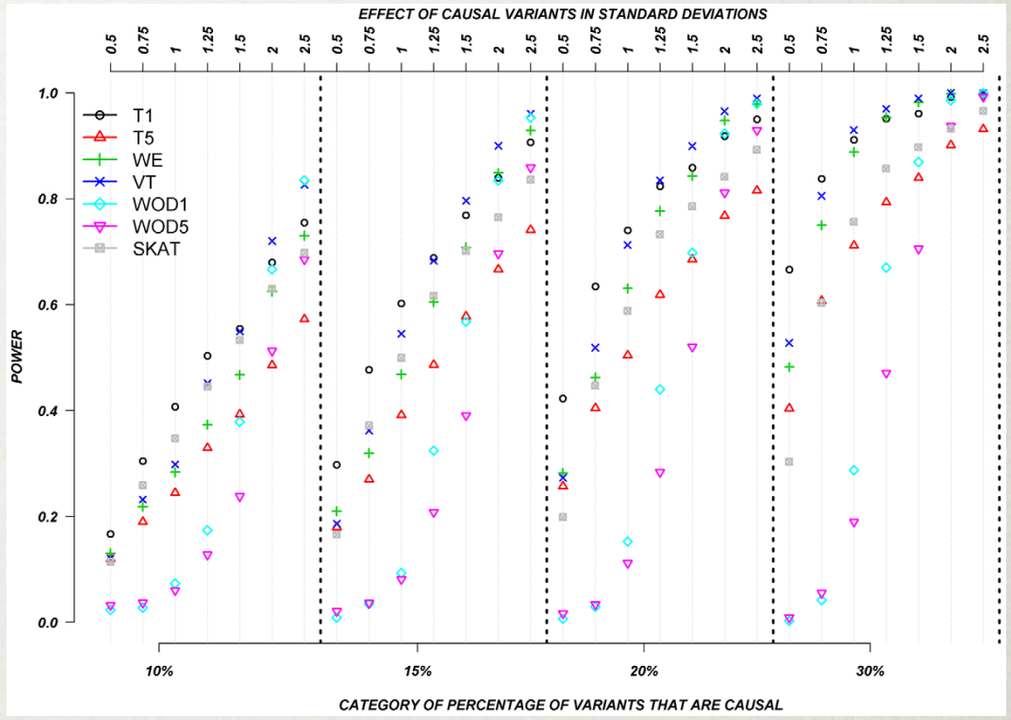


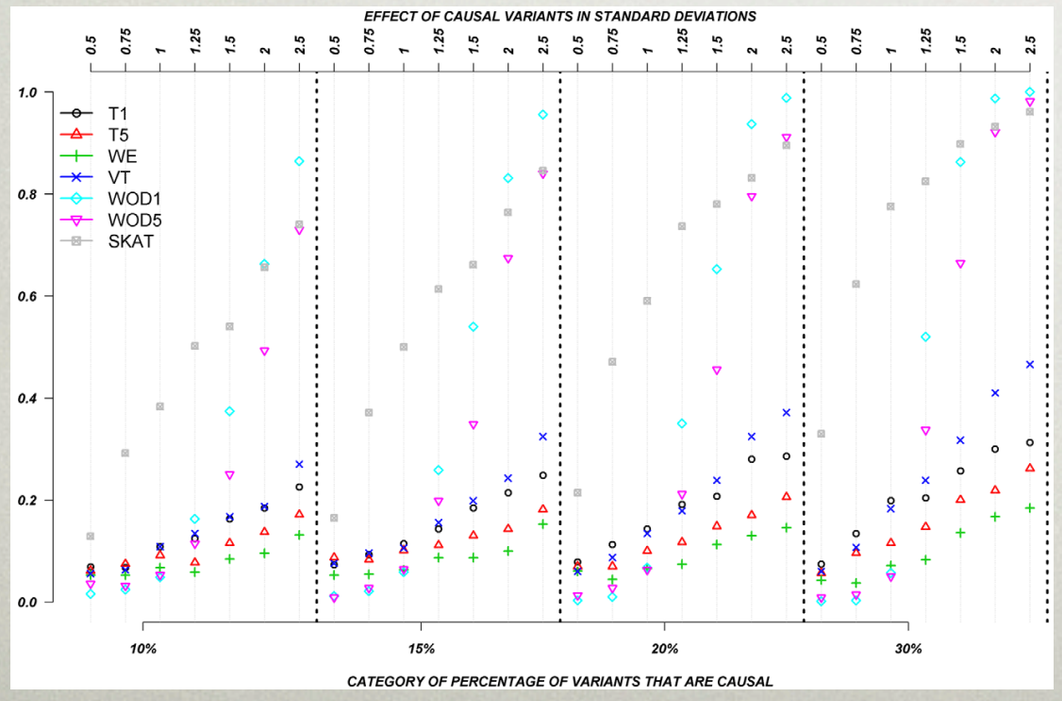
T WE VT* W S*

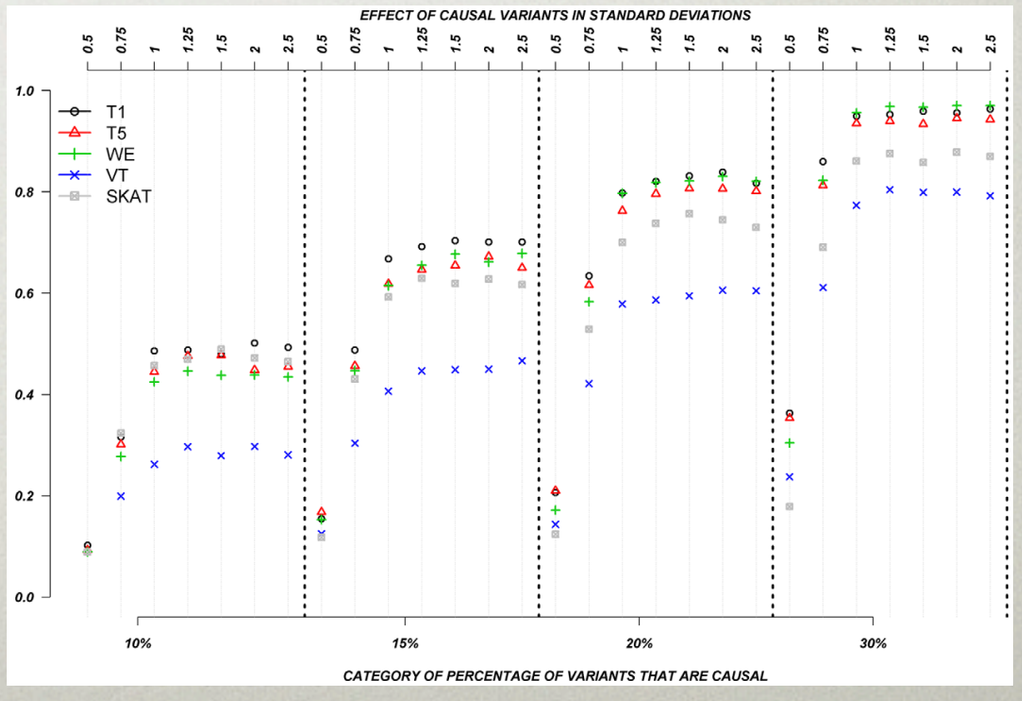

Null
Collapsing
Mixing rare/common
Mixing deleterious/protective
Weighting 1/MAF
Weighting uniform
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
SCENARIOS

T WE VT* W S*

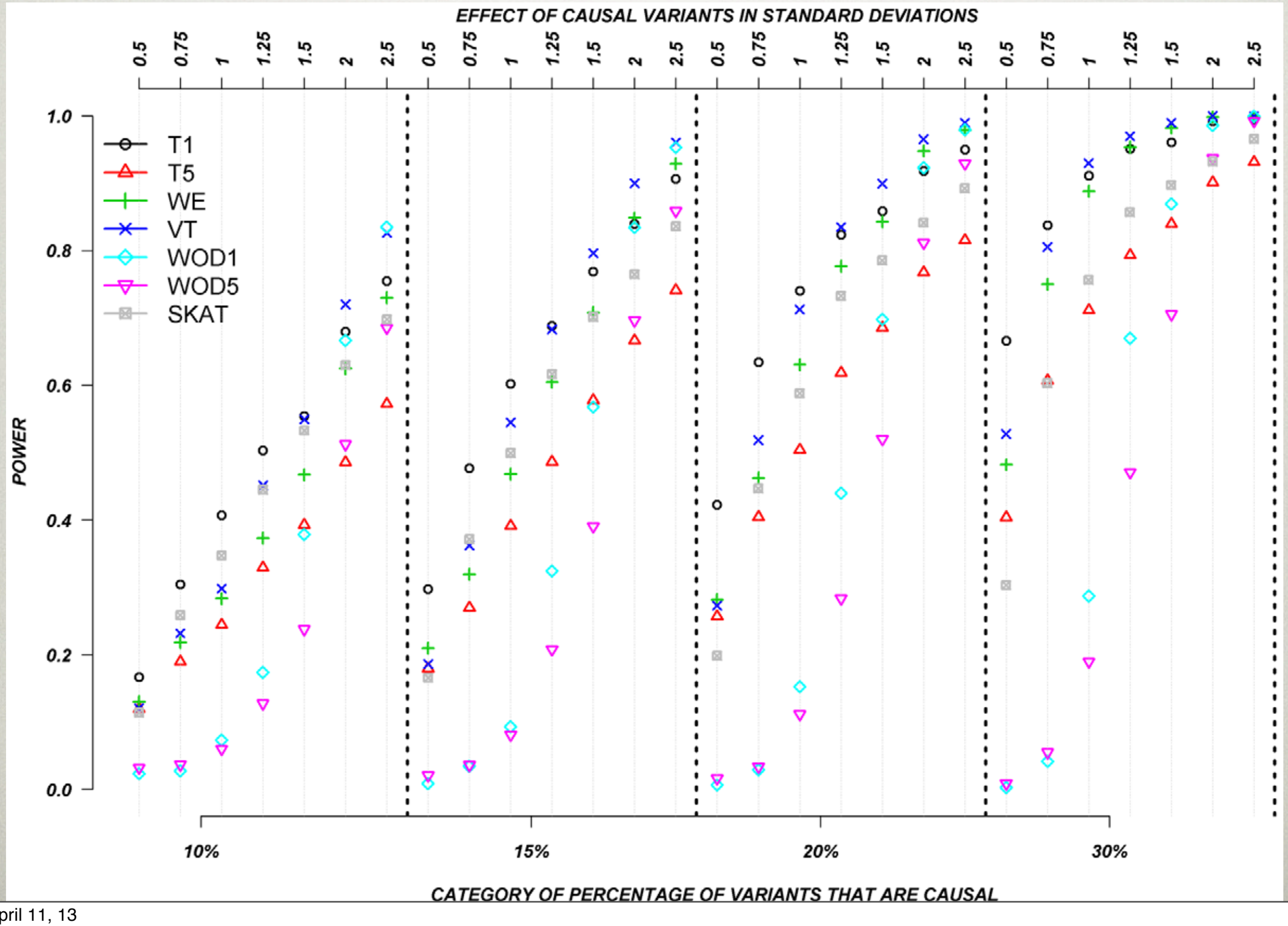

Null
Collapsing
Mixing rare/common
Mixing deleterious/protective
Weighting 1/MAF
Weighting uniform
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
SCENARIOS


EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5
0.75
1
1.25
1.5
2
2.5
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

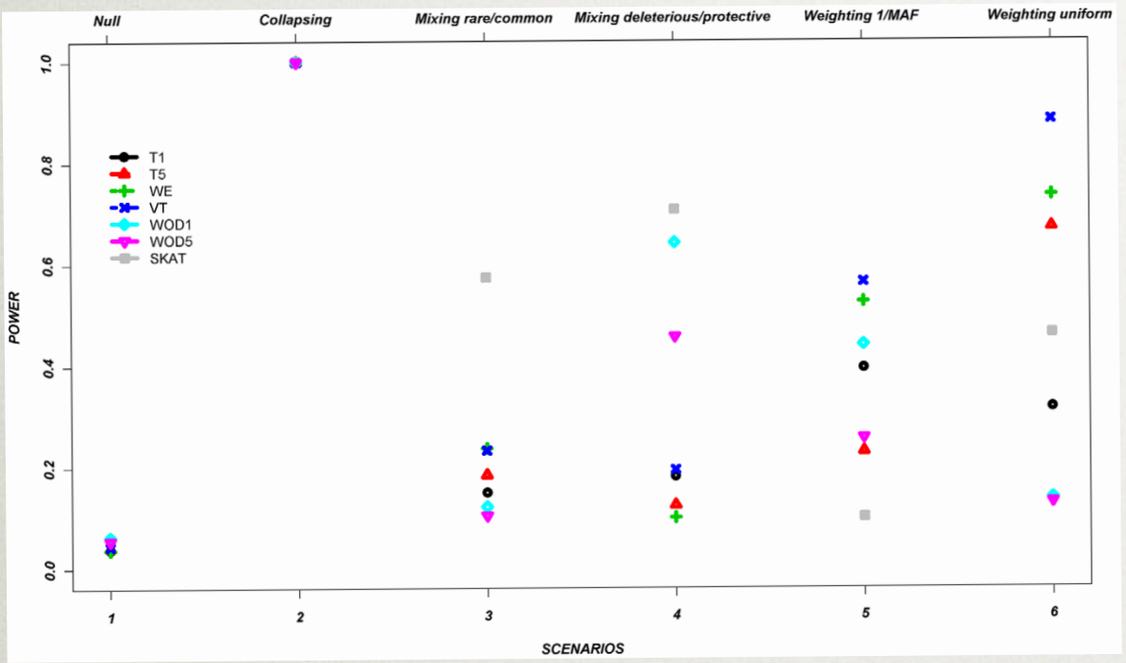

EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5
0.75
1
1.25
1.5
2
2.5
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

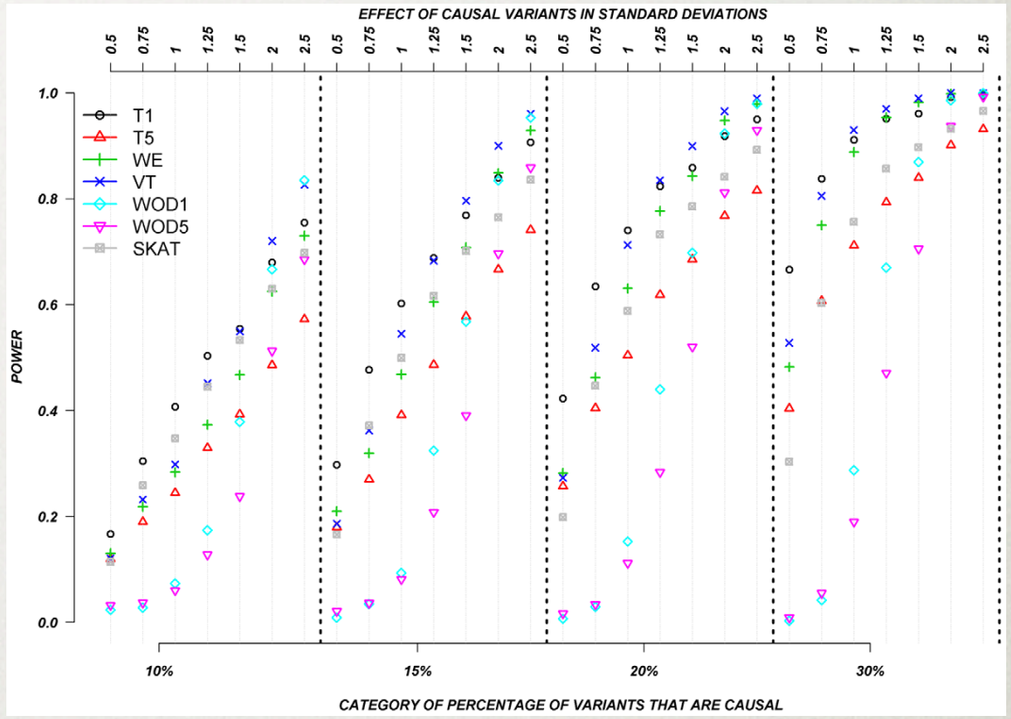

EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5
0.75
1
1.25
1.5
2
2.5
T1
T5
WE
VT
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

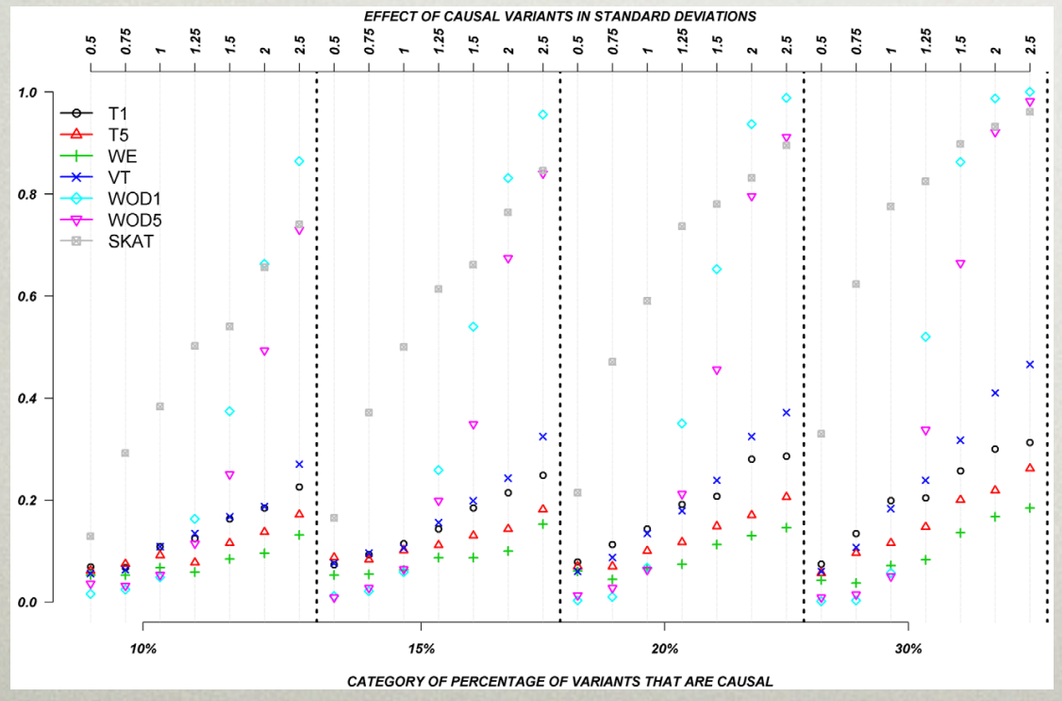
T* WE VT** W S**

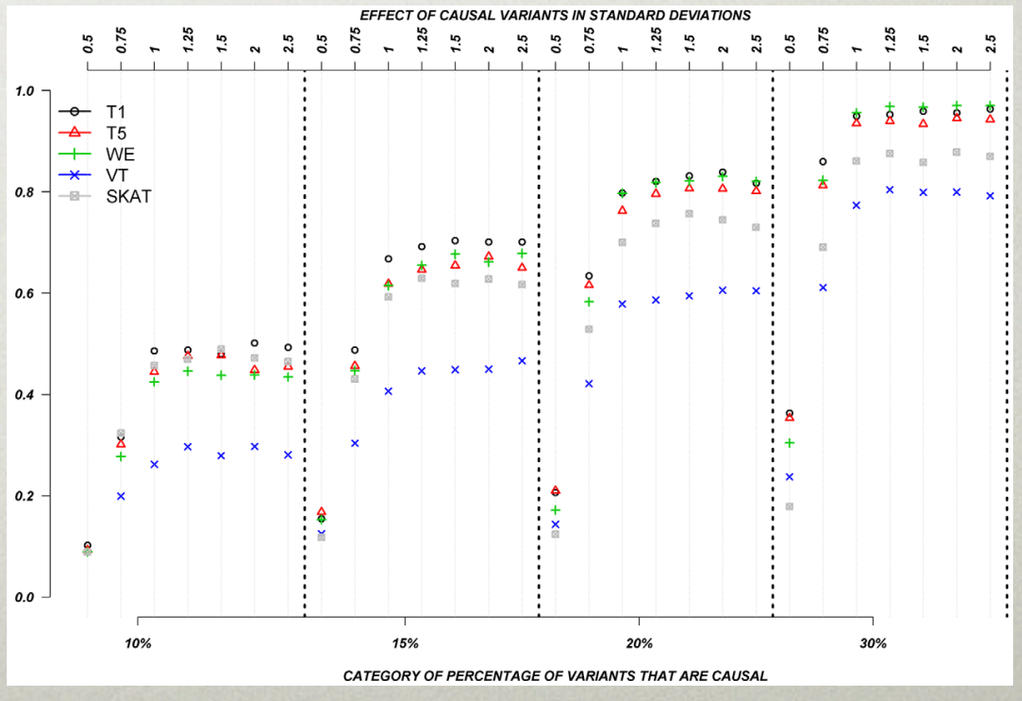

EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
1.0
0.8
0.6
0.4
0.2
0.0
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

T* WE VT** W S**

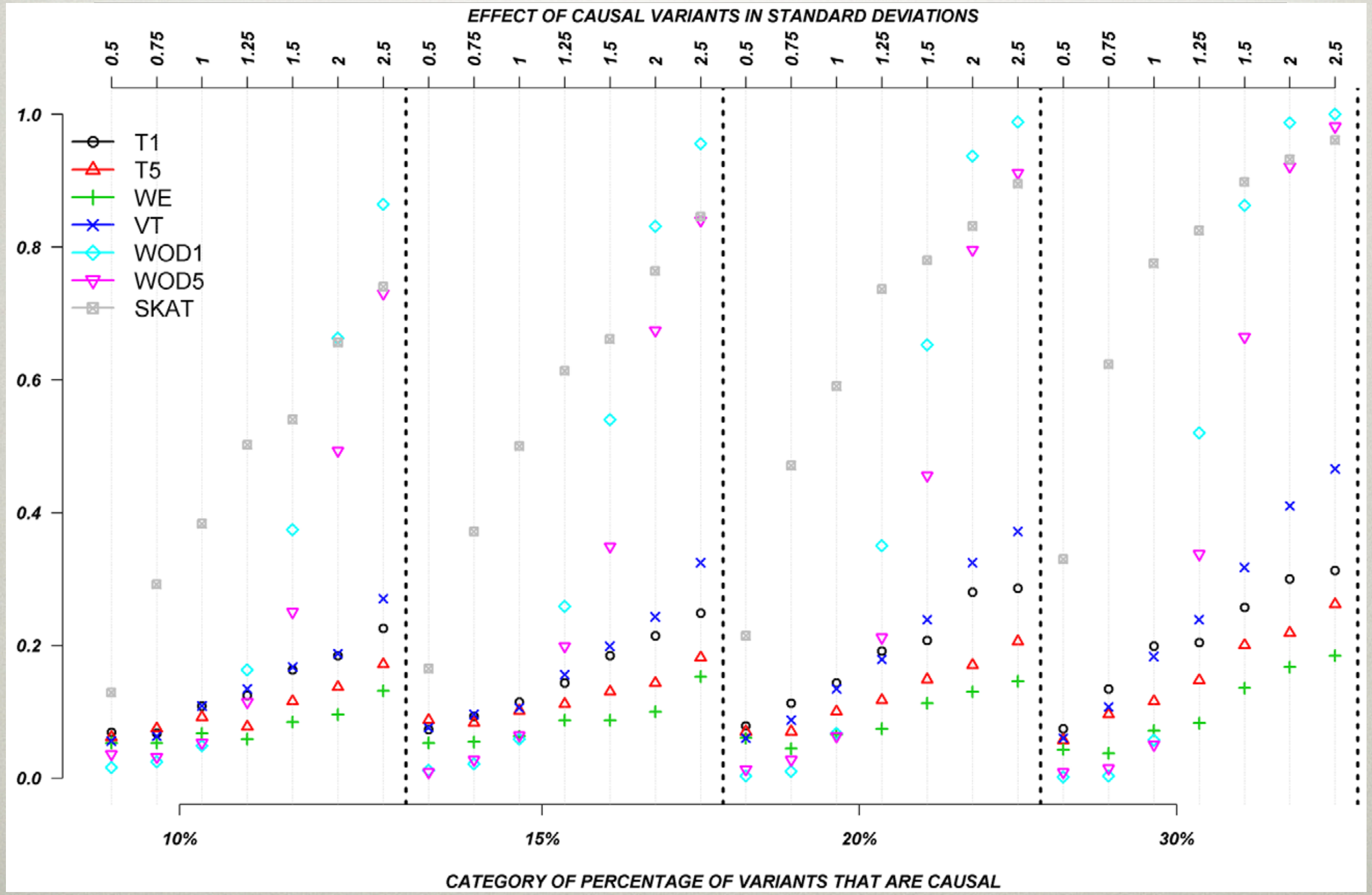

Null
Collapsing
Mixing rare/common
Mixing deleterious/protective
Weighting 1/MAF
Weighting uniform
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
SCENARIOS


EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

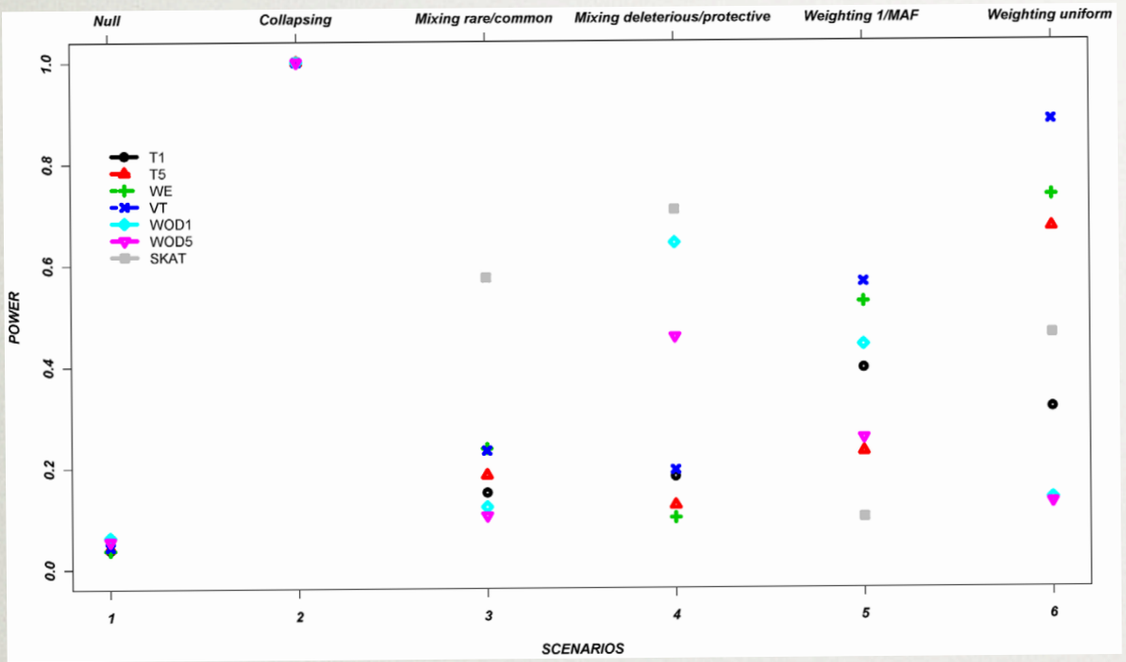

EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
T1
T5
WE
VT
WOD1
WOD5
SKAT
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

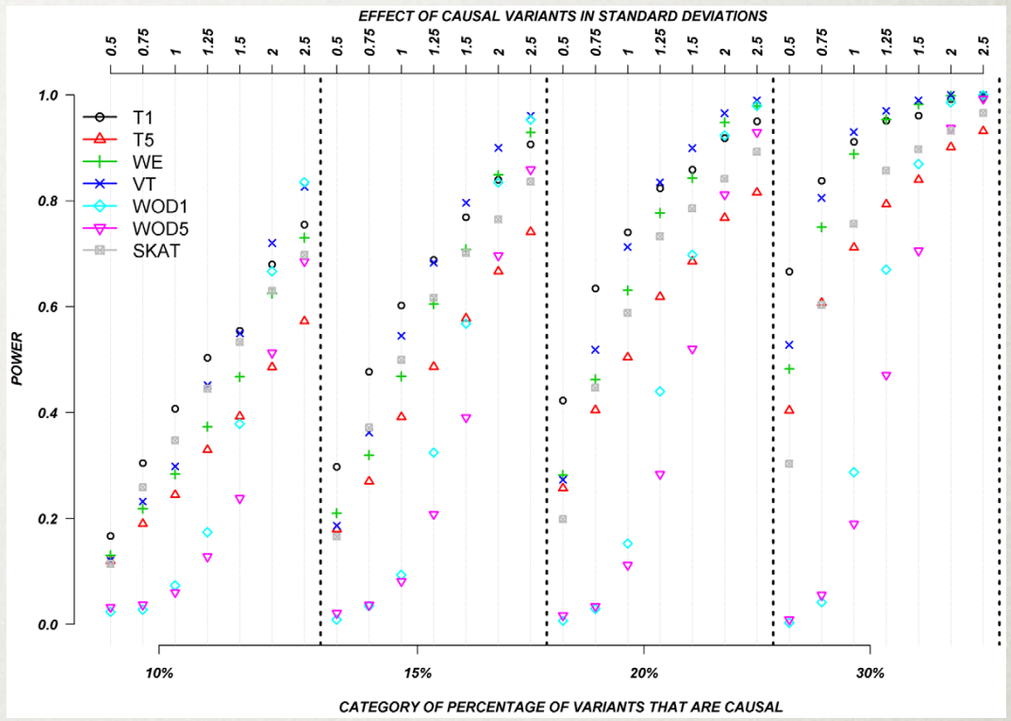

EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
T1
T5
WE
VT
SKAT
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

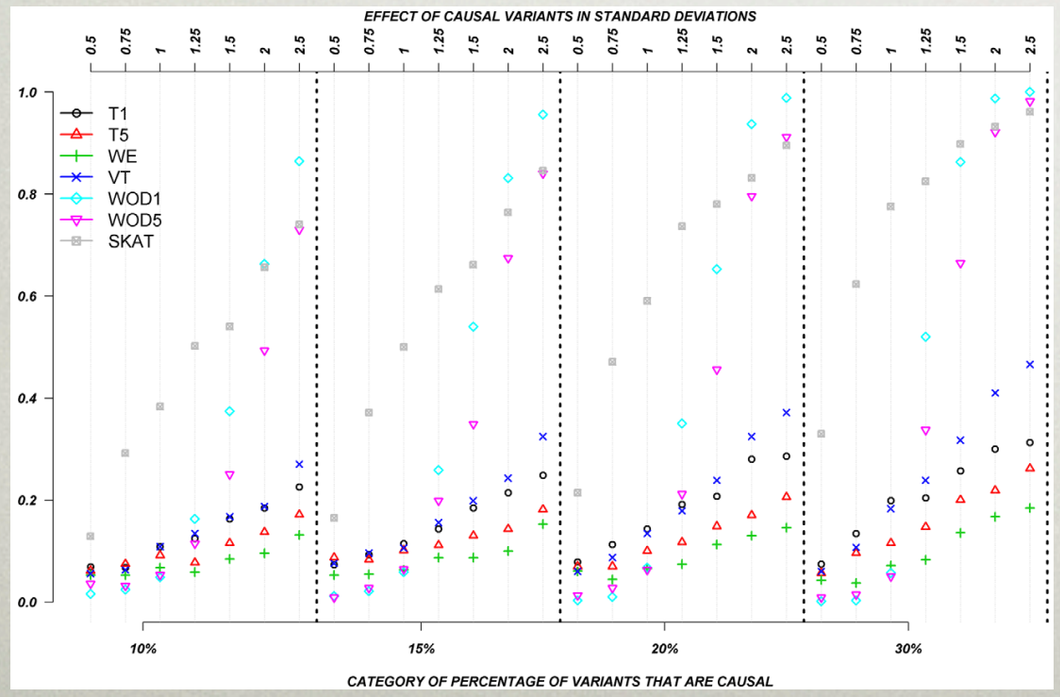
T* WE VT** W* S***

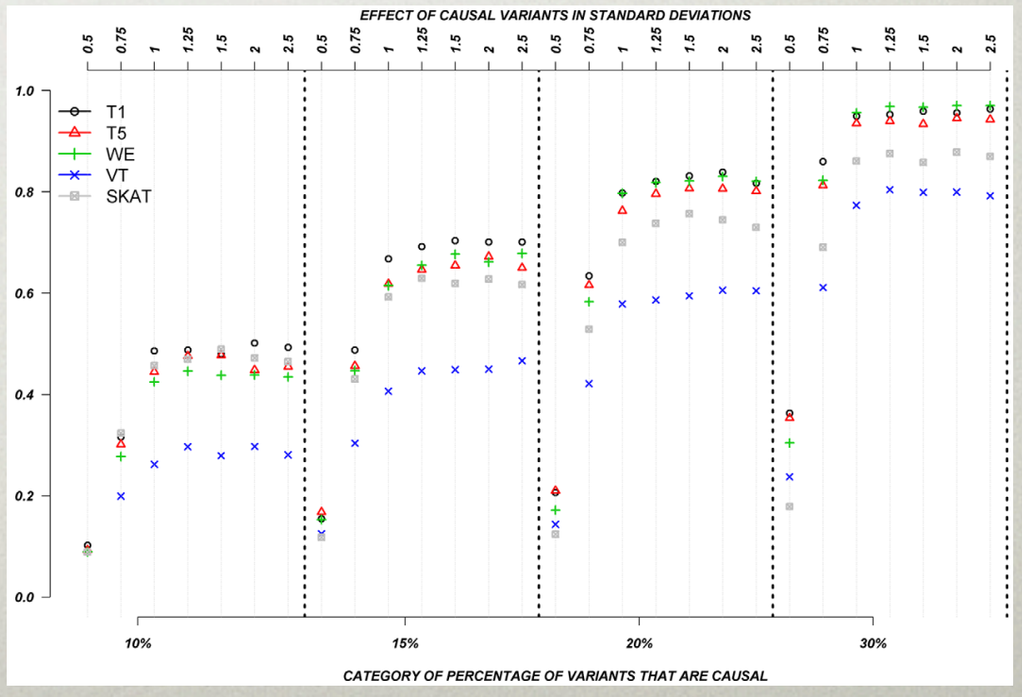

EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5
1.0
0.8
0.6
0.4
0.2
0.0
T1
T5
WE
VT
WOD1
WOD5
SKAT
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

T* WE VT** W* S***

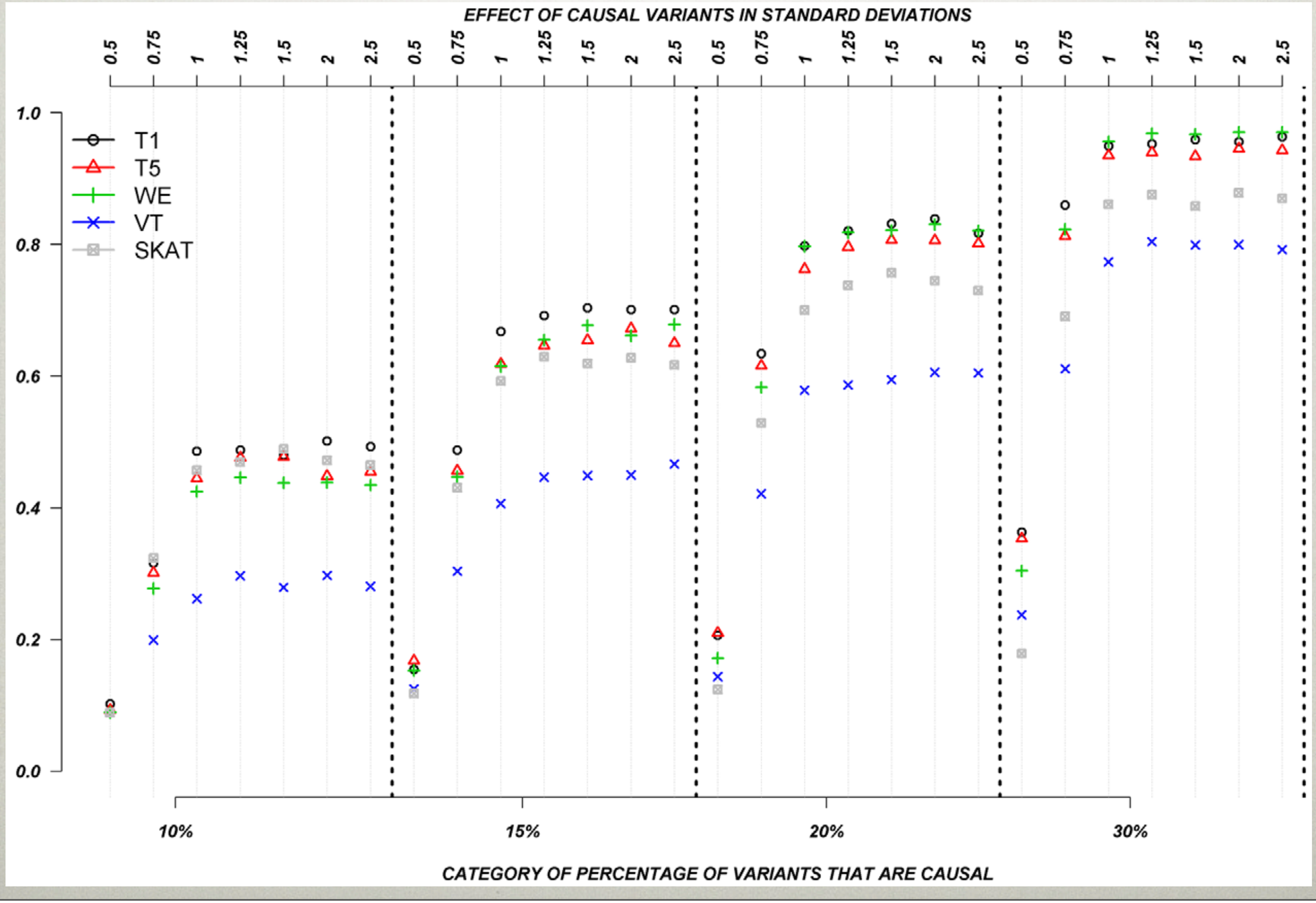

Null
Collapsing
Mixing rare/common
Mixing deleterious/protective
Weighting 1/MAF
Weighting uniform
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
SCENARIOS


EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5
0.75
1
1.25
1.5
2
2.5
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL


EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5
0.75
1
1.25
1.5
2
2.5
T1
T5
WE
VT
WOD1
WOD5
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL


EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5
0.75
1
1.25
1.5
2
2.5
T1
T5
WE
VT
SKAT
POWER
0.0
0.2
0.4
0.6
0.8
1.0
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

T** WE* VT** W* S***


EFFECT OF CAUSAL VARIANTS IN STANDARD DEVIATIONS
0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5 0.5 0.75 1 1.25 1.5 2 2.5
1.0
0.8
0.6
0.4
0.2
0.0
T1
T5
WE
VT
SKAT
10%
15%
20%
30%
CATEGORY OF PERCENTAGE OF VARIANTS THAT ARE CAUSAL

What is the best method to detect rare variants?

Our results demonstrate that the power of recently proposed statistical methods depend strongly on the underlying hypotheses.... No method demonstrates consistently acceptable power... Sensitivity analyses are therefore recommended.., and promising results should be replicated using the same method in an independent sample.

Meaning that...

- Under specific assumptions, we can build a method, and it will work brilliantly
- The same method may work miserably under other set of plausible assumptions
- And there are many genetically plausible scenarios!

Overview

- Missing heritability
- Expected composition of heritability
- Mapping rare variation
- **Conclusions**

Conclusions

- Rare alleles of small effect are hard to solve statistically (mind that "rare" and "small" is relative to sample size). Extreme example - private *de novo* mutations
- We need to figure out what of the multiple plausible scenarios are more prevalent in reality
- (?) Need methods combining knowledge from different domains (evolutionary, systems, and functional biology)