The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses

Armando Caballero, Albert Tenesa and Peter D. Keightley

Genetics, 2015. 201(4) 1601-1613. DOI: 10.1534/genetics.115.177220

Yeni Liliana Bernal Rubio. PhD.
QuantGen Group
Department of Epidemiology and Biostatistics
Michigan State University

Outline

Introduction - background

- Missing heritability – Arguments and concepts.

Objectives

- Determine genetic variation (%VA) and power to detect QTL under different methodologies.

Datasets and Methods

- Simulation
- Comparison of methodologies in regard of QTL detected and additive genetic variance explained

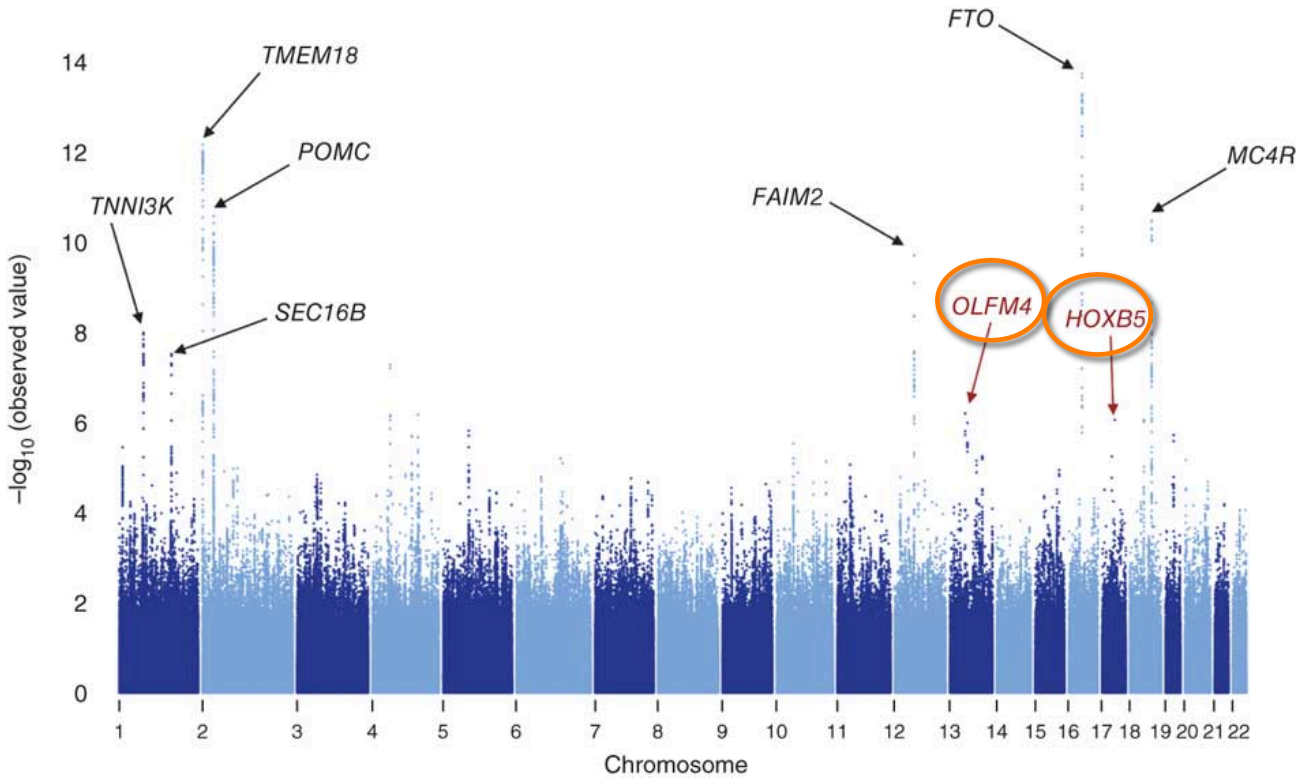
Results and Discussion

- Nature of quantitative trait variation for complex traits.
- Characterization of missing genetic variance at different methodologies.

Conclusions

Introduction

Genome-wide association studies (GWAS): Association of SNPs with variation of the trait.



Manhattan plot of the results from meta-analysis for childhood obesity.

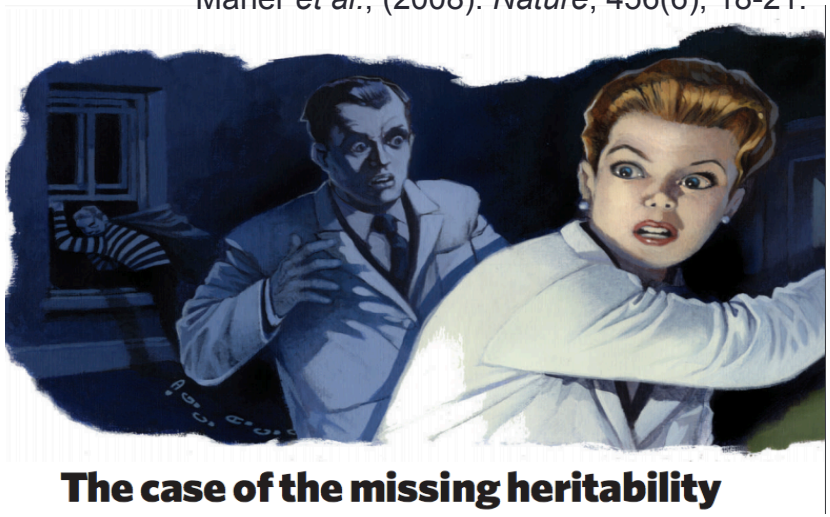
Bradfield *et al. (*2012*) Nature Genetics.* **44**, 526–531.

GWAS have allowed (*Visscher et al., 2012*):

→ To discover **many loci/variants** influencing complex traits.

↓ To reveal that the significantly associated SNPs account for only a **small proportion** of the trait's genetic variation.

Maher *et al.*, (2008). *Nature*, 456(6), 18-21.



Is the gap between the heritability explained by the loci identified in GWAS and the trait's heritability.

Finding the missing heritability of complex diseases

Teri A. Manolio[1], Francis S. Collins[2], Nancy J. Cox[3], David B. Goldstein[4], Lucia A. Hindorff[5], David J. Hunter[6], Mark I. McCarthy[7], Erin M. Ramos[5], Lon R. Cardon[8], Aravinda Chakravarti[9], Judy H. Cho[10], Alan E. Guttmacher[1], Augustine Kong[11], Leonid Kruglyak[12], Elaine Mardis[13], Charles N. Rotimi[14], Montgomery Slatkin[15], David Valle[9], Alice S. Whittemore[16], Michael Boehnke[17], Andrew G. Clark[18], Evan E. Eichler[19], Greg Gibson[20], Jonathan L. Haines[21], Trudy F. C. Mackay[22], Steven A. McCarroll[23] & Peter M. Visscher[24]

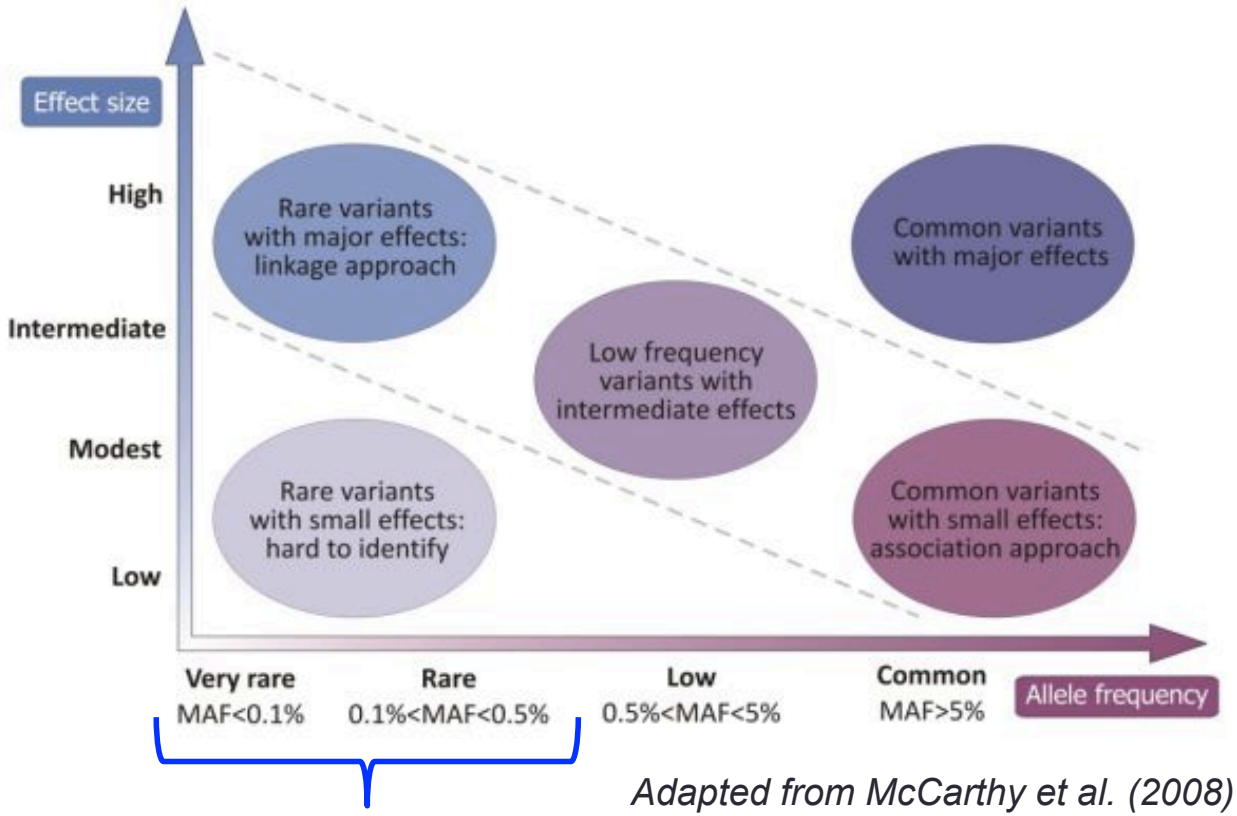
Trait	Number of Loci	% h^2 explained
Height (h^2=0.8)	40	5

Different arguments explaining missing heritability

- Rare variants; small effect variants
- Epistasis (Zuk et al. 2012; Bloom et al. 2013).
- Synthetic associations common feature of GWAS (Dickson et al. 2010; Wang et al. 2010)
- Imperfect LD between SNP and causal variants (Thornton et al. 2013).

Arguments explaining missing heritability

- Alleles causing variation for quantitative traits may be rare with moderate effects (Manolio et al., 2009)

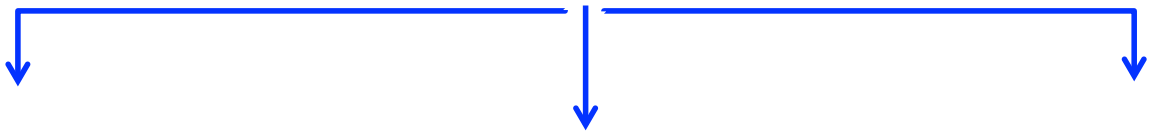


Adapted from McCarthy et al. (2008)

Deleterious fitness effects

Fitness traits

Physical traits and behavior that enable an organism to survive and reproduce

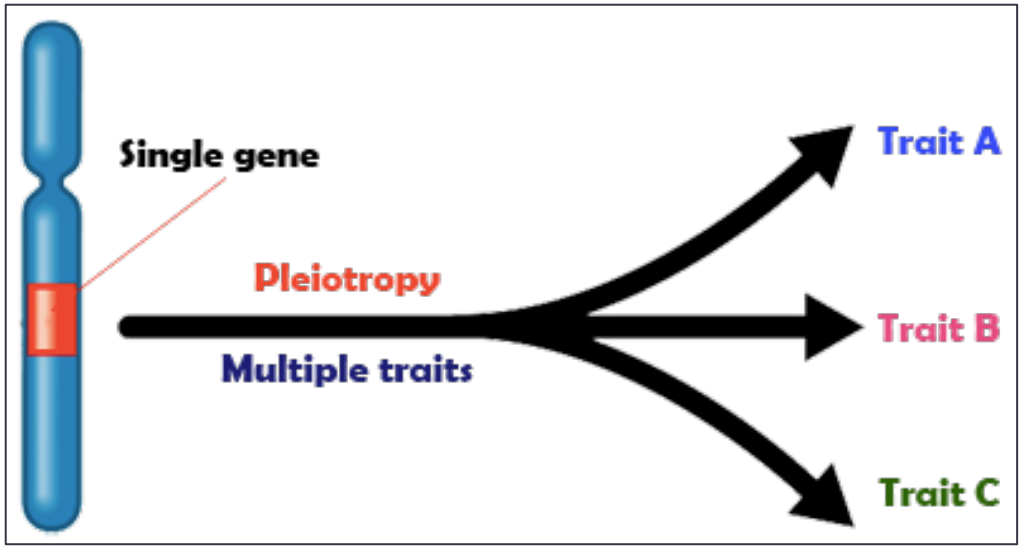


Survival

Number of offsprings per mating that survive

Mating success

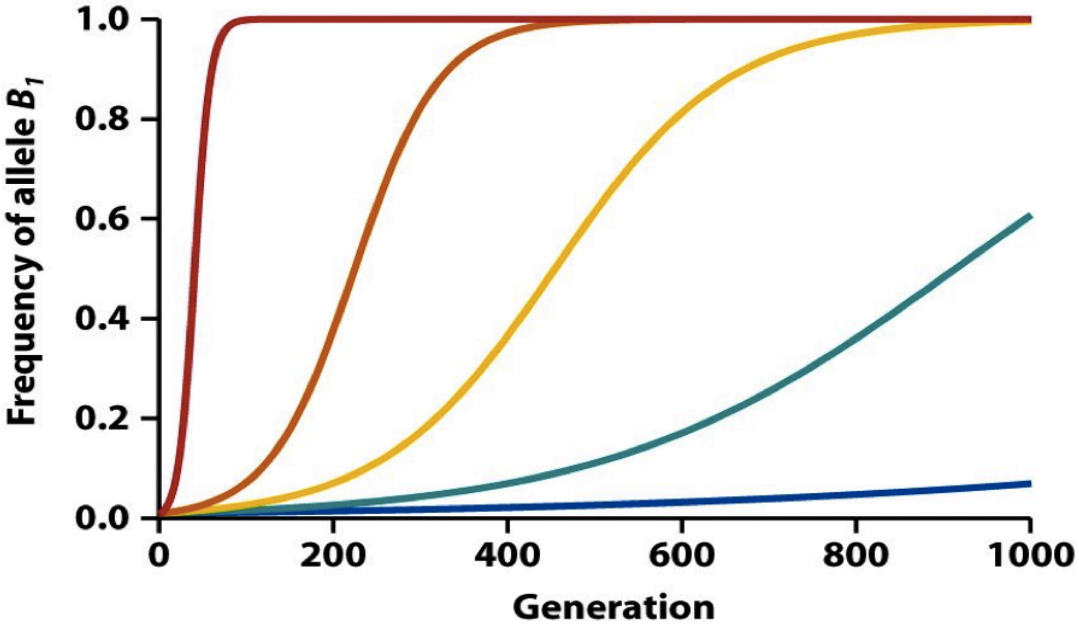
GWAS have provide evidences of **pleiotropy.**



Variants affecting a complex trait subject to selection because:

- Trait is fitness → * Selection against deleterious SNPs * Genetic architecture based on rare variants.
- Variants have pleiotropic effects on fitness → Low frequency of variants because of deleterious effect on fitness

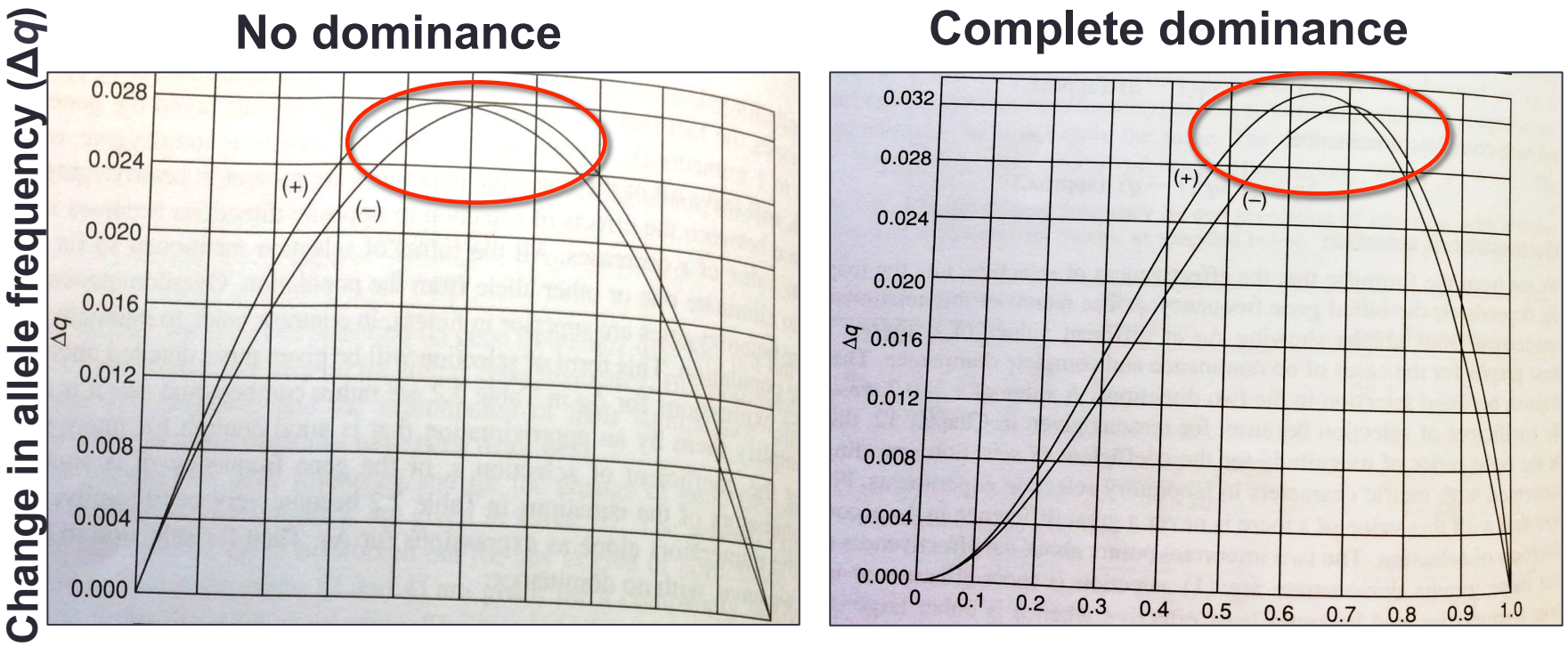
Effect of selection in allele frequency



Selection	Percent surviving B_1B_1	B_1B_2	B_2B_2
Strong	100	90.0	80.0
	100	98.0	96.0
	100	99.0	98.0

Selection	Percent surviving B_1B_1	B_1B_2	B_2B_2
	100	99.5	99.0
	100	99.8	99.6
Weak			

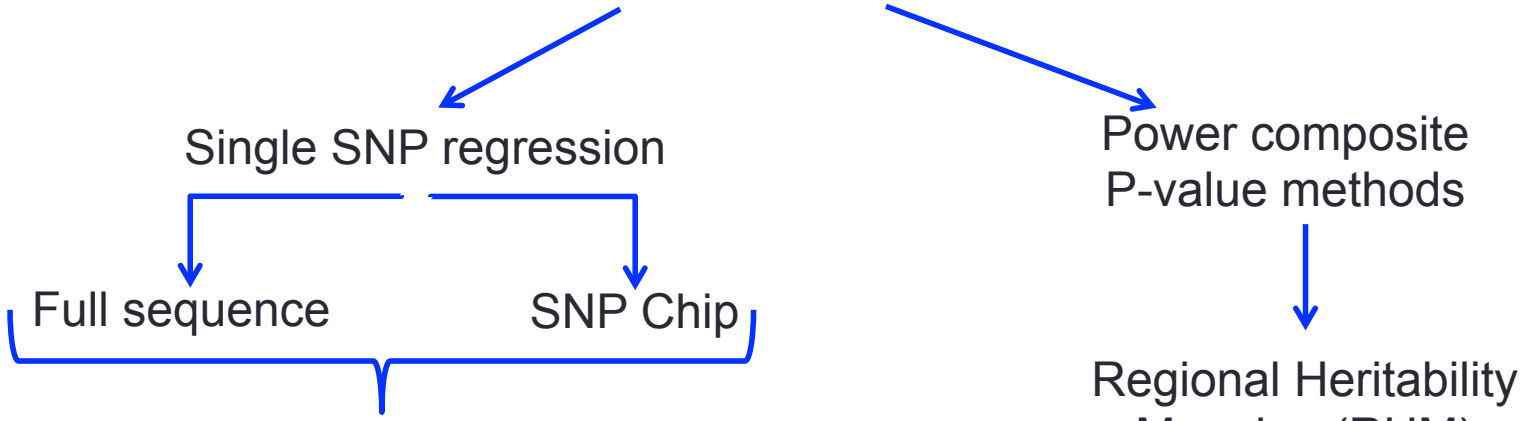
Effectiveness of selection (Falconer &MacKay, 1960)



(-) Slx against allele
(+) Slx in favour of allele

Missing h^2 and power of detection of genetic variants

Comparison of power of detection using



What is the extent of the genetic variation revealed?

Objective

Assuming complex traits with different proportion of rare variants we aim to

Compare results obtained from **GWAS or Regional Heritability Mapping (RHM) applied to full sequence data** in regard of:

1. Number of **quantitative trait variants** detected

2. **Additive genetic variance explained**.

Materials and Methods

Dataset

Simulation of a <u>complex trait correlated with fitness</u>.

Mus Musculus genome sequence of 10Mb (SSC19; 14Mb-24Mb)

- Constant unstructured population (N_e=1000;10,000 generations)

- Mutations with pleiotropic effects on a quantitative trait and on fitness.

Response variables (Y)

Fitness → All mutations assumed deleterious

Genetic correlation → **Five Quantitative traits**

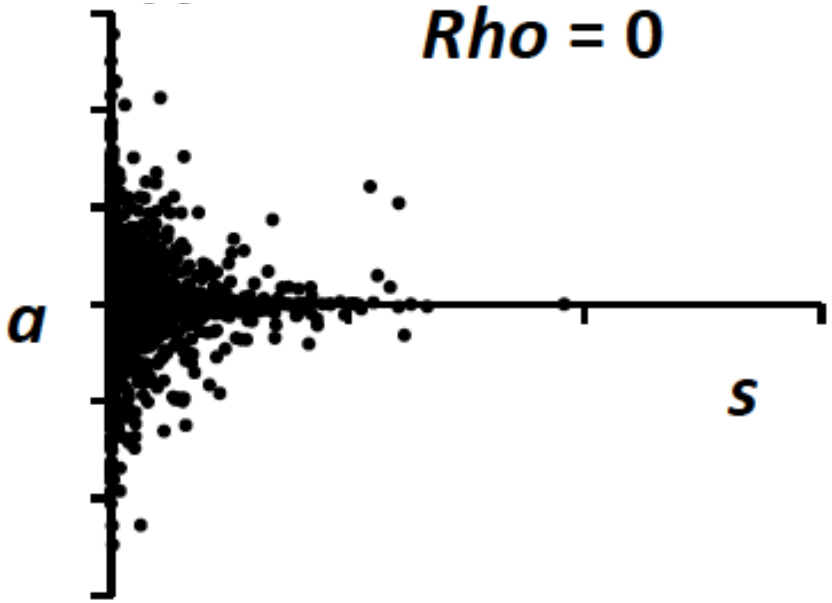
Between fitness and trait:
0, 0.25, 0.5, 0.75 and 1 → Quantitative trait is fitness

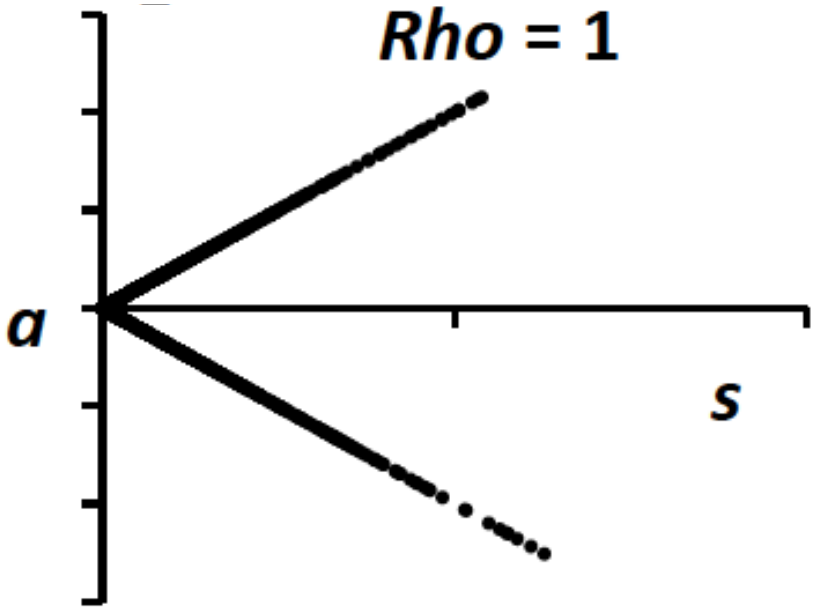
- **Quantitative traits**: Genotypic values = $\sum$ effects of mutations

  Residuals ~ N (0, σ_e^2)

  Heritability = 0.8

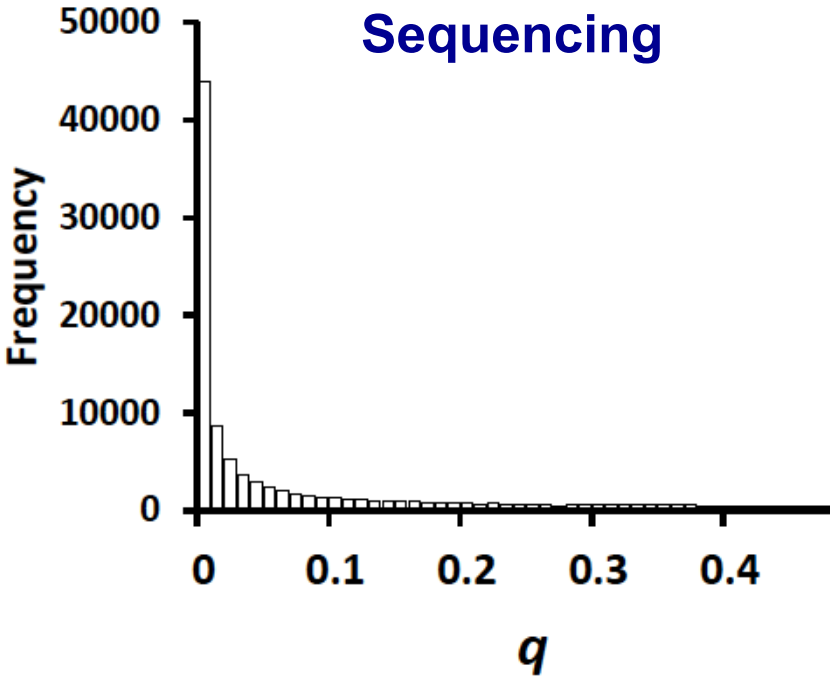
Joint distribution of effects on the quantitative trait (a) and fitness (s)

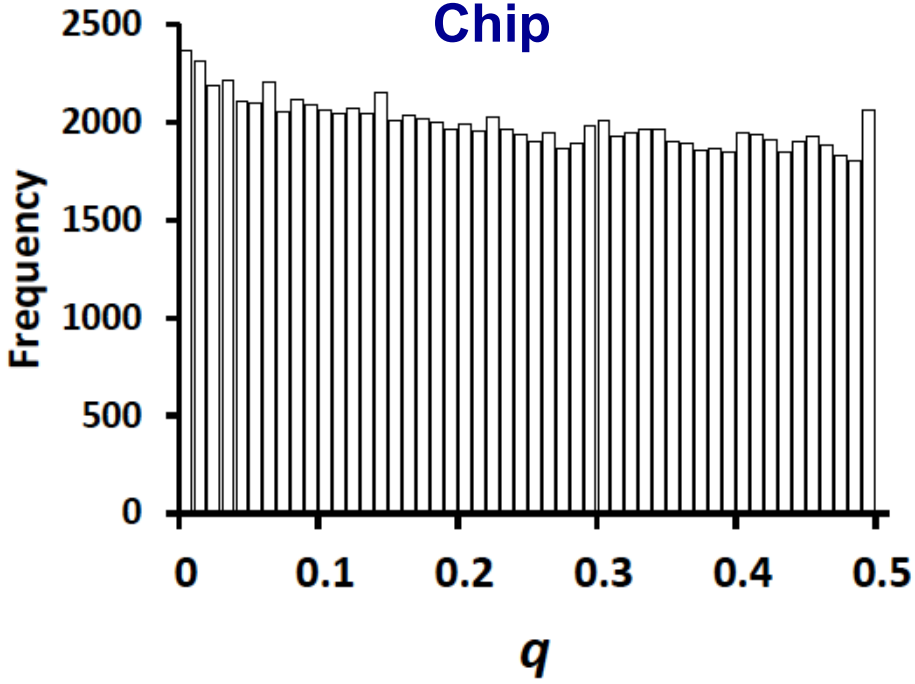




Predictors (X)

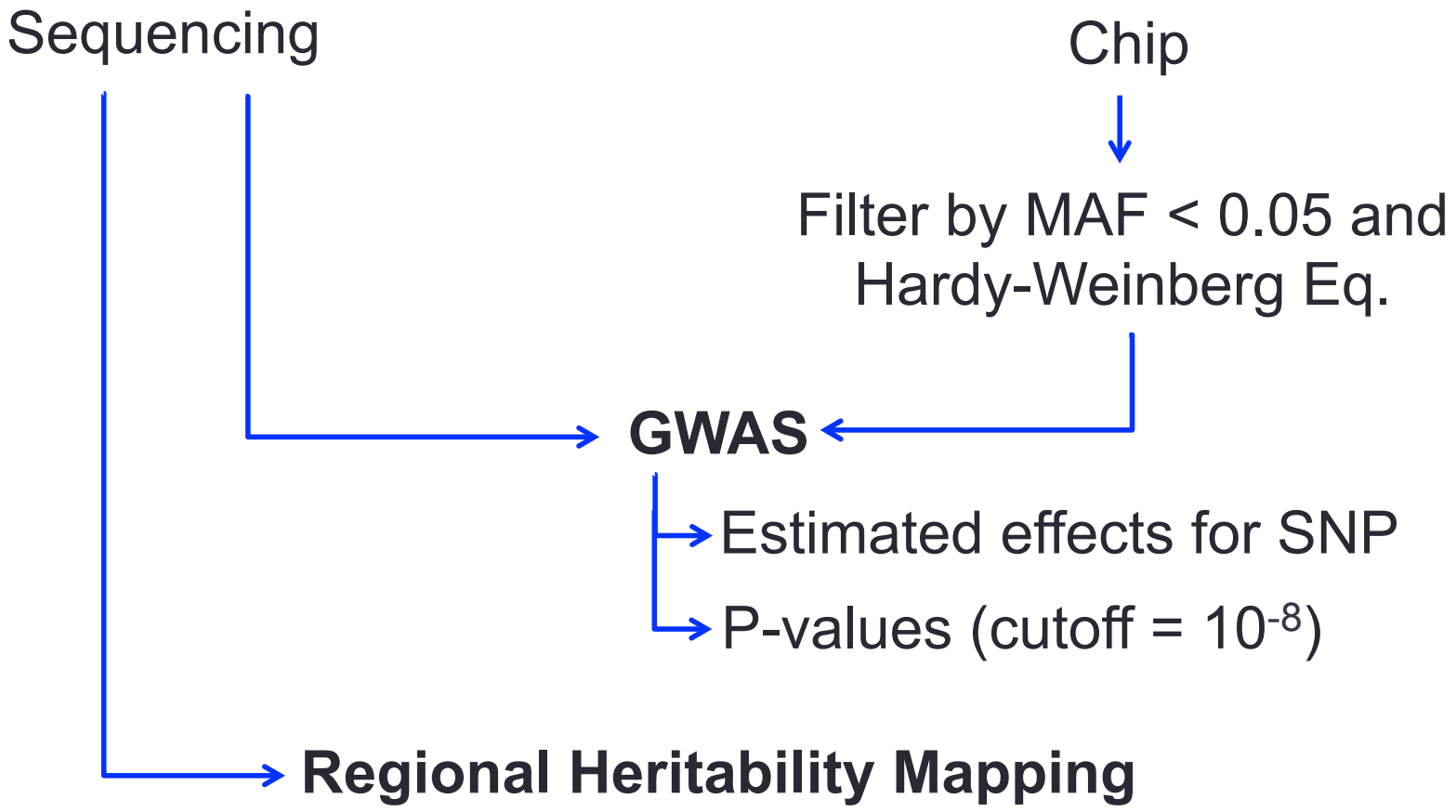
SNP simulated (~9.7 million per genome)





~9.7 million per genome → Sample for uniform distribution and LD $r^2 < 0.9$ → ~1.36 million of SNP

Methods (analysis)



Regional Heritability Mapping (RHM)(Nagamine et al., 2012)

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Zv} + \mathbf{e}$$

Whole genomic additive genetic effect

Regional genomic additive genetic effect

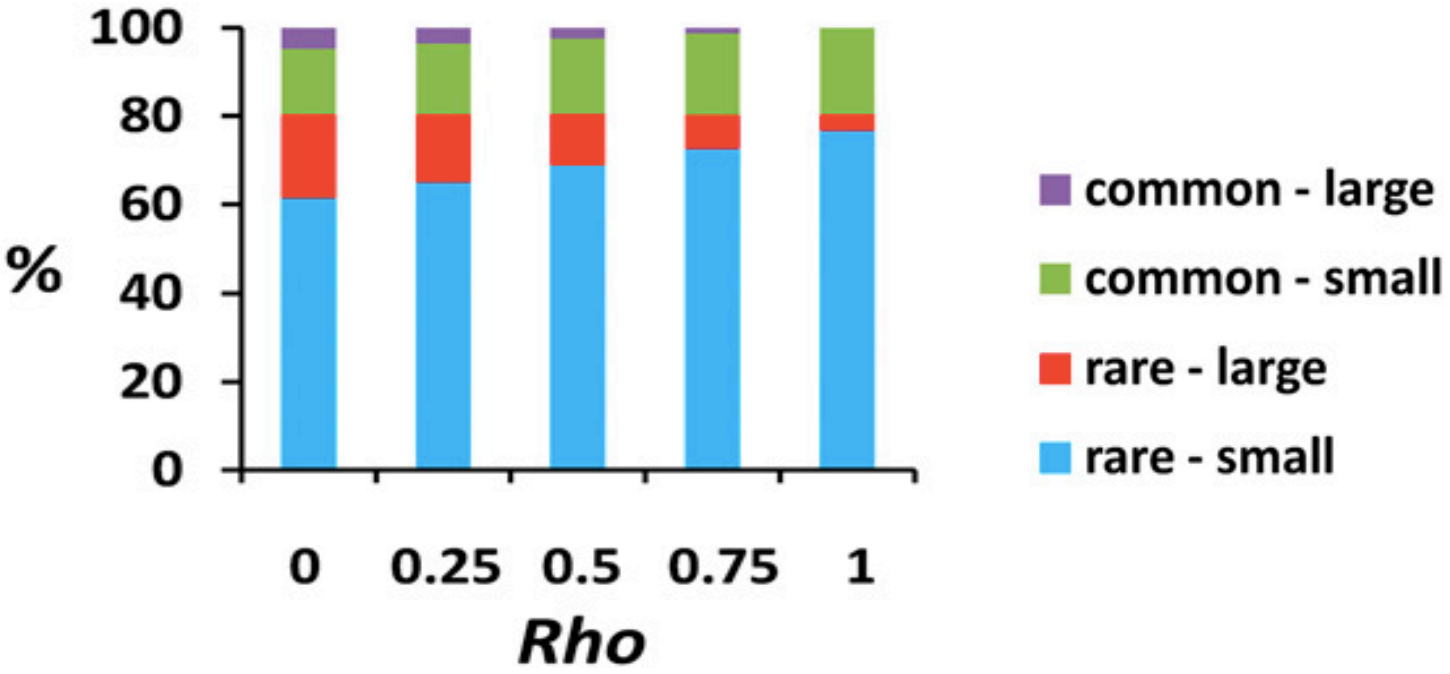
$$\mathrm{Var}(\mathbf{u}) = \mathbf{G}\sigma_u^2,\ \mathrm{Var}(\mathbf{v}) = \mathbf{Q}\sigma_v{}^2$$

For the sequencing:

- $\sigma_v{}^2$ for windows of 20 consecutive SNPs and an overlap of 10 SNPs.
- Estimates of $\sigma_v{}^2$ were significant if ($10^{-12} < p\text{-value} < 10^{-4}$).

Results and discussion

Number of QTL segregating in the population



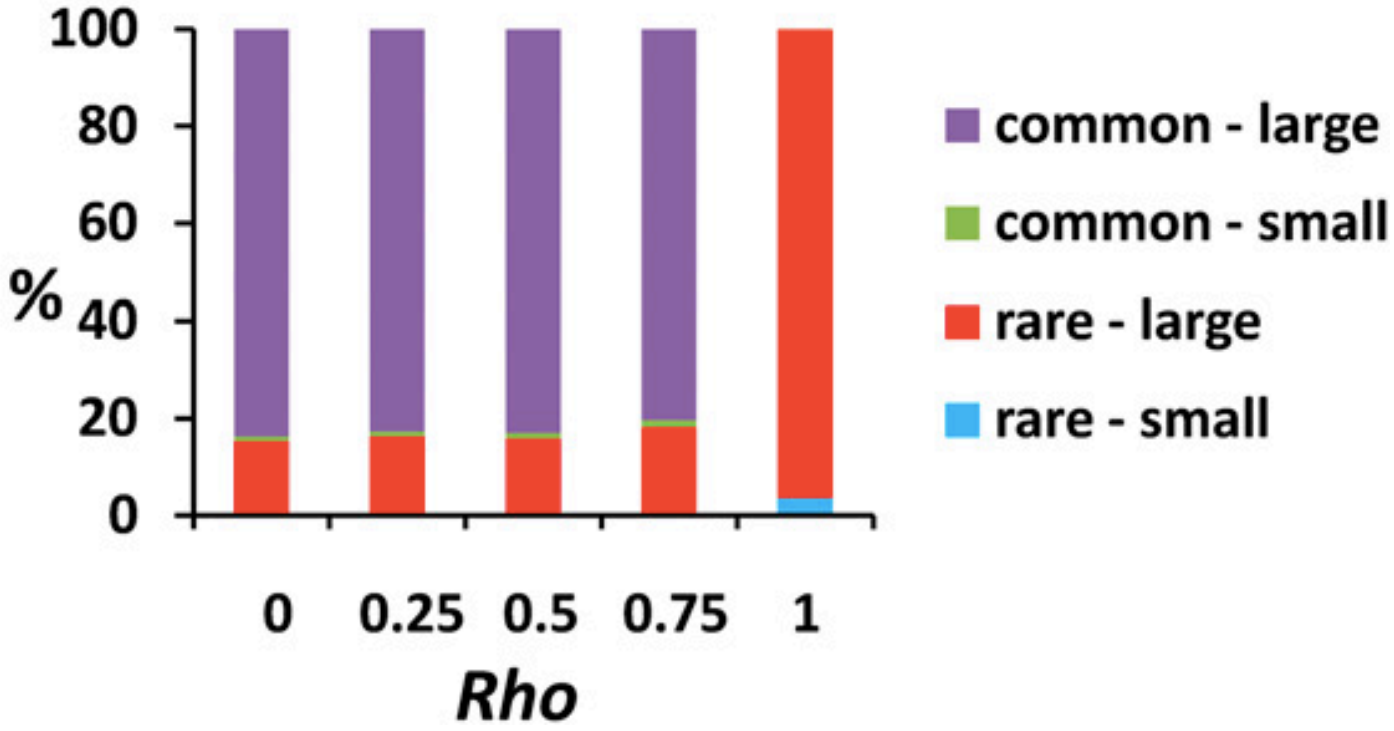
Number of QTLs is dominated by **rare variants.**

Rare variants with → **large effect** tend to disappear

→ **small effect** increase

When trait correlates more with fitness.

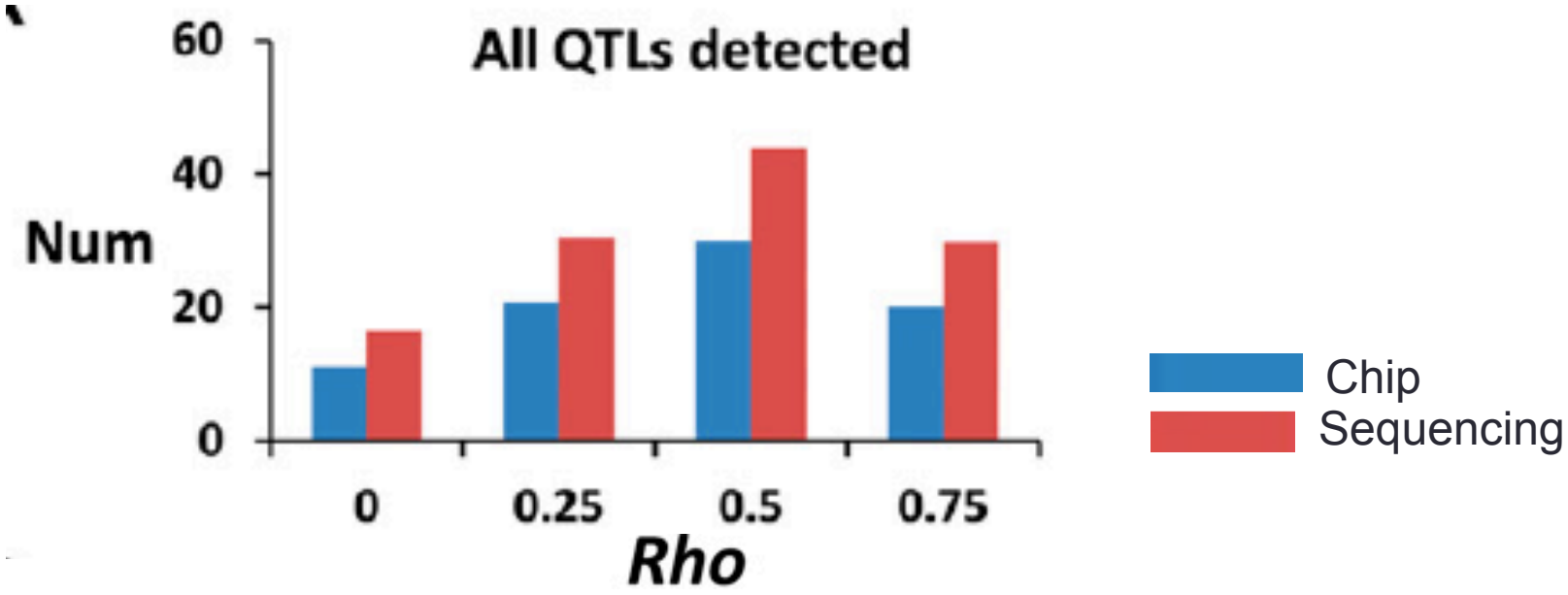
Contribution to the additive genetic variance



σ_a^2 is mostly explained by **common variants**.

When the trait is a fitness trait (ρ = 1), all σ_a^2 comes from rare variants.

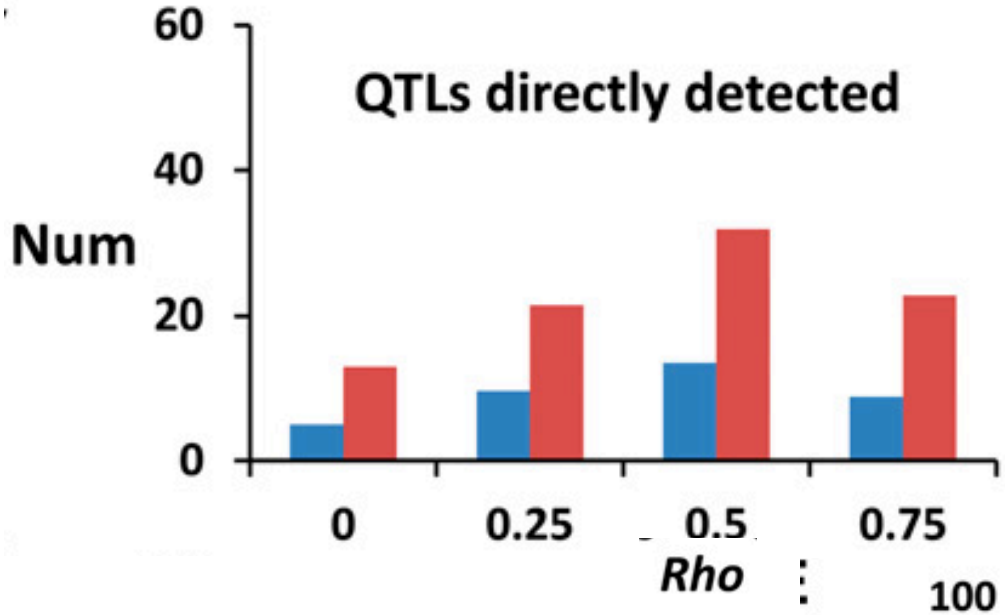
Number of QTL detected by GWAS



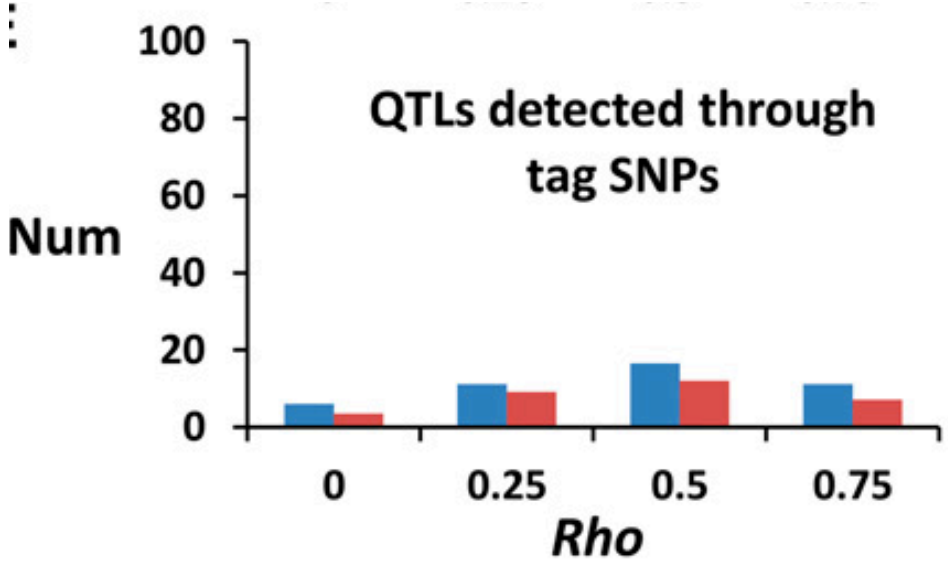
Number of QTLs detected was **larger** for

- Intermediate values of ρ
- Sequencing analysis

Number of QTL detected by GWAS



Sequencing detected QTL with higher precision



Number of QTL detected indirectly was higher in Chip

Number of QTL detected by GWAS vs. Regional Heritability Mapping (RHM) on SEQUENCING

		Number			
Frequency	Effect *a* (pSDs)	All QTL	GWAS	RHM	GWAS + RHM
Rare, $q \leq 0.05$	0–0.03	99,801	14	127	138
	0.03–0.14	13,530	0	18	18
	0.14–0.28	2,504	0	1	1
	0.28–0.56	758	2	1	3
	>0.56	71	4	3	4
Common, $q > 0.05$	0–0.03	24,637	52	38	64
	0.03–0.14	2,727	14	5	14
	0.14–0.28	555	73	26	76
	0.28–0.56	127	92	67	95
	>0.56	12	12	12	12
		144,722	263	298	425

Missing heritability can be explained by inability to detect QTL with moderate effects (0.03–0.28 pSD)

Multiple SNP or genomic region analyses are more powerful than GWAS

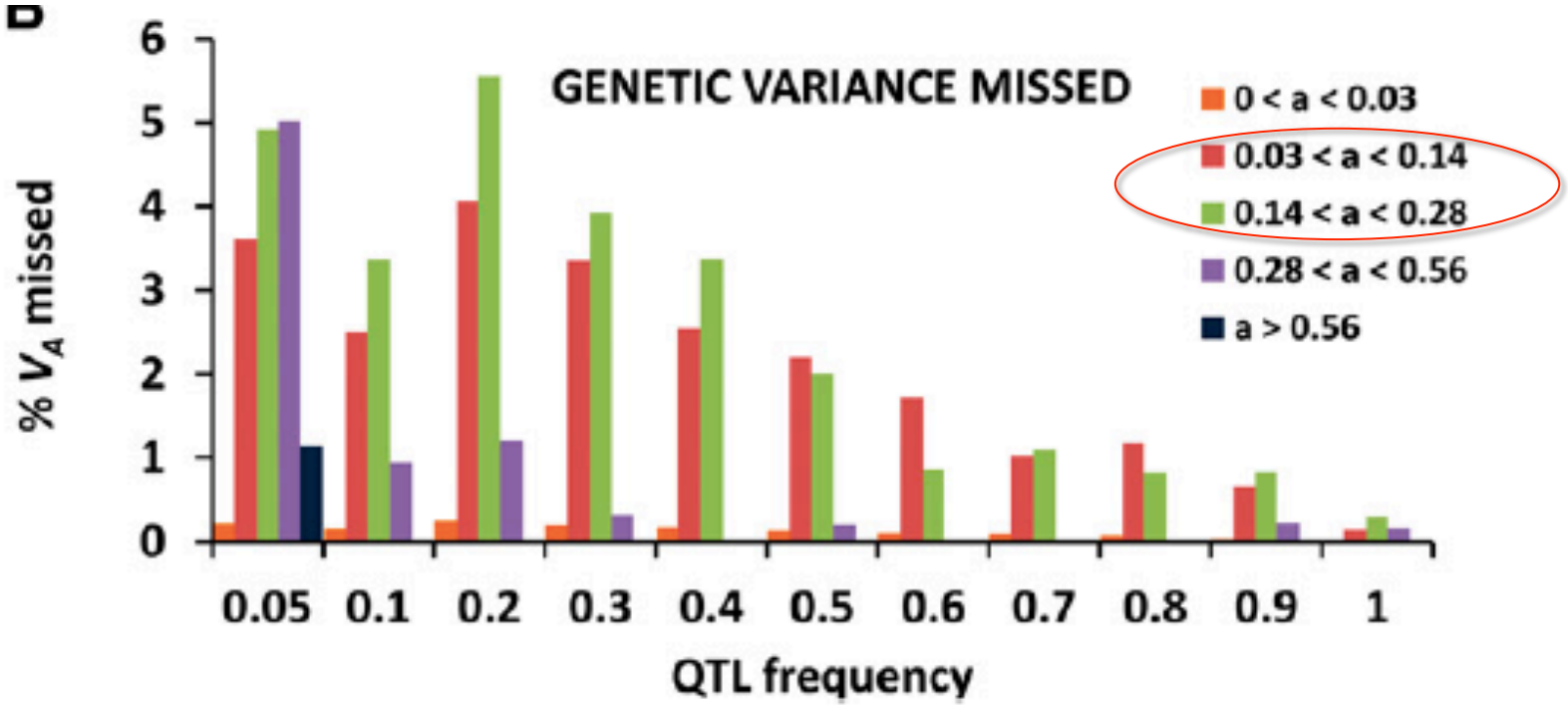
Additive genetic variance explained by GWAS vs. Regional Heritability Mapping (RHM) on SEQUENCING

	$\%V_A$			
Frequency	All QTL	GWAS	RHM	GWAS + RHM
Rare, $q \leq 0.05$	0.2	0.0	0.0	0.0
	3.6	0.0	0.0	0.0
	4.9	0.0	0.0	0.0
	5.1	0.1	0.0	0.2
	1.9	0.8	0.7	0.8
Common, $q > 0.05$	1.2	0.0	0.0	0.0
	19.5	0.2	0.1	0.2
	30.7	8.6	3.6	8.9
	24.9	21.8	18.8	22.7
	7.9	7.9	7.9	7.9
	100.0	39.4	31.1	40.6

Contribution to VA is larger from common variants

The combination of GWAS + RHM gave a marginal increase in %VA

Characterization of the missing genetic variance



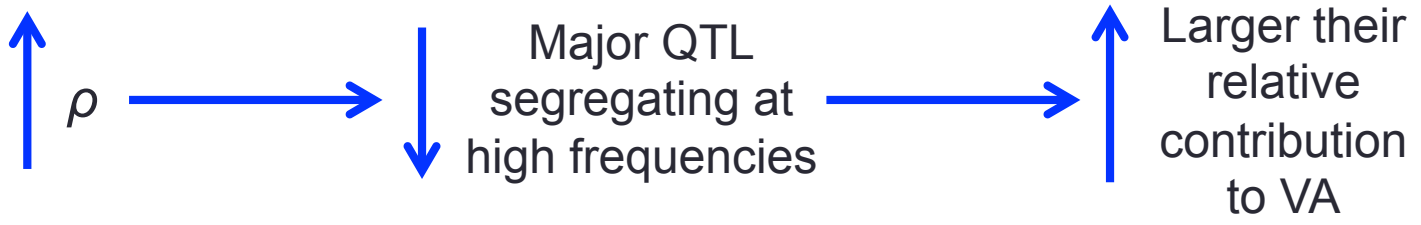
Missing VA is mostly due to QTL of minor to moderate effect segregating at intermediate frequencies

Conclusions

GWAS vs RHM applied to full sequence data and the number of **quantitative trait variants** detected and the **additive genetic variance explained**.

The use of full sequence data and RHM can improve the detection rate of variants.

The amount of variance explained increases with the value of ρ:



Missing heritability can be explained by the inability to detect QTL of moderate effect (0.03–0.28 pSD) segregating at substantial frequencies.

Most variation is due to common QTL of substantial effect

Except when the trait under study is fitness itself (most variation due to rare QTL).

For Regional Heritability Mapping:

Cebamanos, L., A. Gray, I. Stewart, and A. Tenesa, 2014. Regional heritability advanced complex trait analysis for GPU and traditional parallel architectures. Bioinformatics 30: 1177-1179.

For Simulation:

Messer, P. 2013 SLiM: Simulating Evolution with Selection and Linkage. Genetics 194: 1037-1039.