Negative selection in humans and fruit flies involves synergistic epistasis

Mashaal Sohail, Olga A. Vakhrusheva, Jae Hoon Sul, Sara L. Pulit, Laurent C. Francioli, Genome of the Netherlands Consortium, Alzheimer's Disease Neuroimaging Initiative, Leonard H. van den Berg, Jan H. Veldink, Paul I. W. de Bakker, Georgii A. Bazykin, Alexey S. Kondrashov, Shamil R. Sunyaev

Science, 5 May 2017

Thibault Latrille

RAGE meeting - January 17 2018

Outline of the presentation

I. Why are we talking about epistasis in a RAGE meeting?

III. Sohail *et al.* - *Theory*

III. Sohail *et al.* - *Results*

IV. Reproductibility of the study.

Part I.

Why are we talking about epistasis in a RAGE meeting?

"*[...] Sohail et al. found that deleterious loss-of-function mutations are further away from each other in the genome than expected by chance, which suggests that genetic interactions are driving selection. [...]* ***This explains why high levels of variation can be maintained and why sex and recombination are advantageous.***"

Science report, 2017

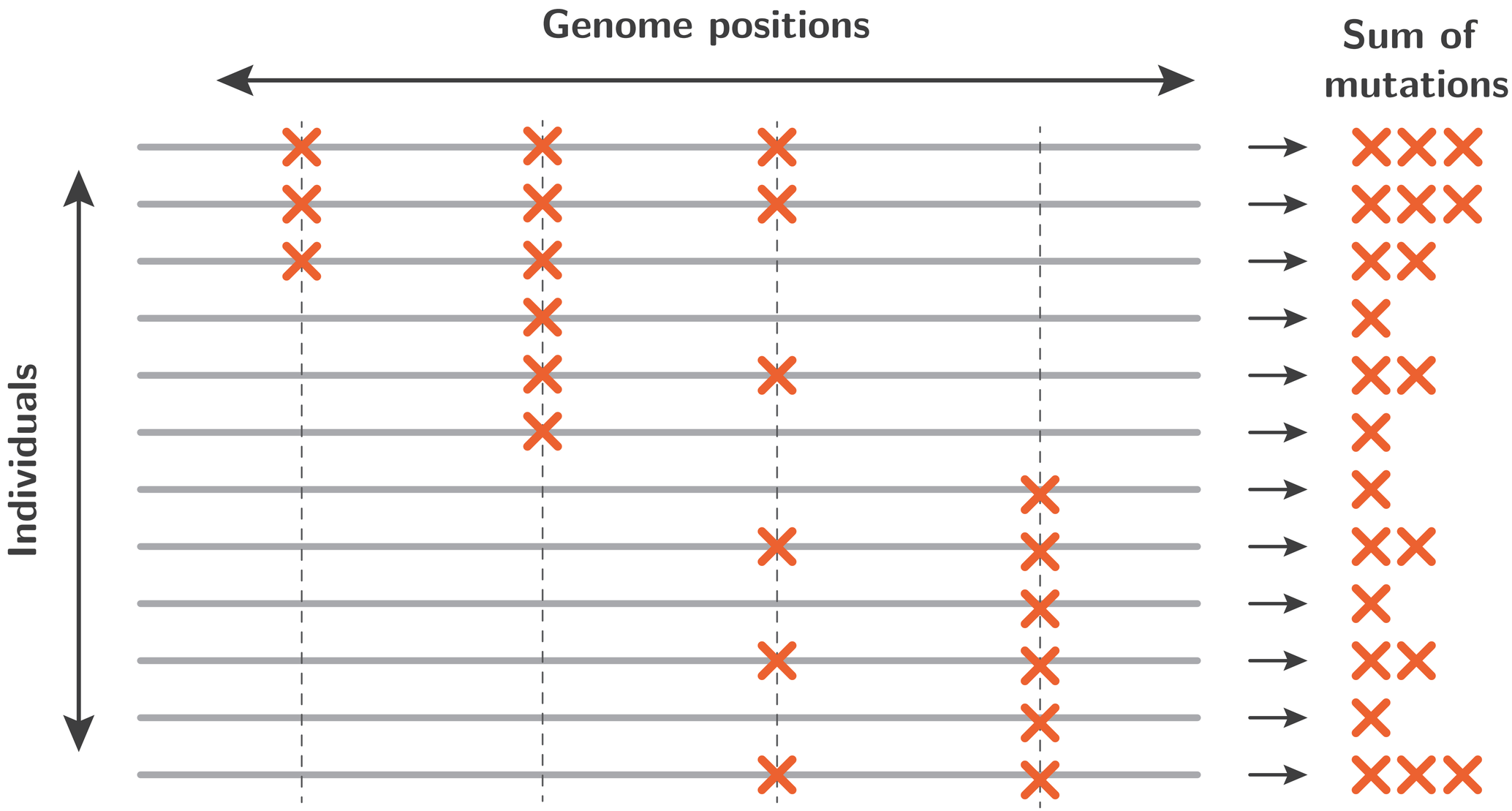
Is recombination advantageous ?

	Direct selection	Indirect selection		
Models:	**Stabilizing selection**	**Heterogeneity in selection**	**Negative epistasis**	**Hill-Robertson effects**
Assumptions:	An optimal number of crossovers per chromosome is required for sucessful meiosis.	I. Temporal fluctuations in the enviroment favour different combinations of alleles. II. Rare combination have an advantage via resistance to parasite.	I. Recurrent deleterious mutations II. Negative epistasis. *i.e.* the combined effects of deleterious mutations is stronger than expected.	I. Population is subject to positive and negative selection. II. Population is finite, thus generating genetic drift.
Linkage desequilibrium:	None.	The sign of LD changes.	Negative LD generated by epistatic selection.	Negative LD generated by drift.
Predictions:	Recombination rate is advantageous for an optimal value and fluctuates around an upper and a lower bound.	Increased recombination rate in environments with temporal variations or in populations exposed to to virulent parasites.	Negative LD generated by epistasis impedes the effect of selection. Thus breaking down LD is advantageous.	Advantageous mutations compete with each other. Weakly deleterious mutations can be associated with advantageous mutations.

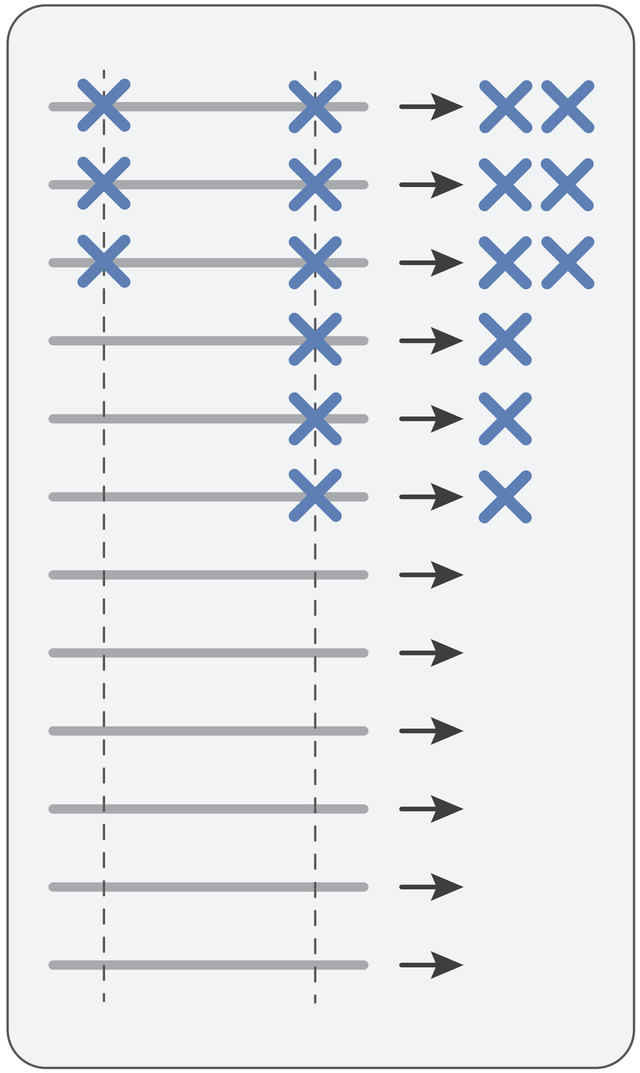
Part II.

Sohail *et al.* - Theory

How to measure linkage desequilibrium (LD) at the genome-wide level?

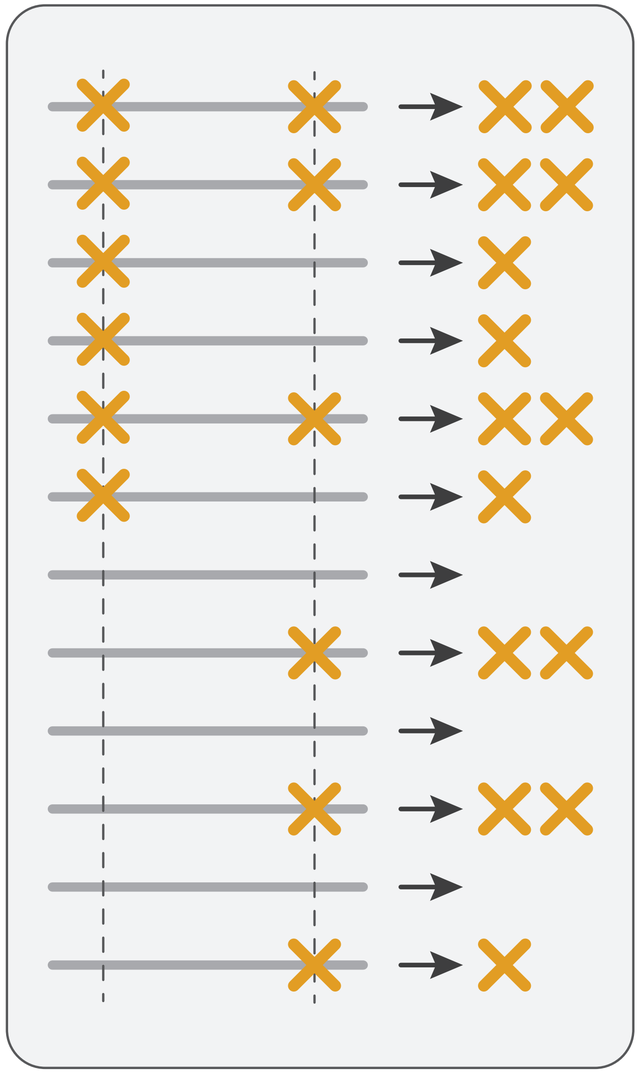


Pair-wise LD

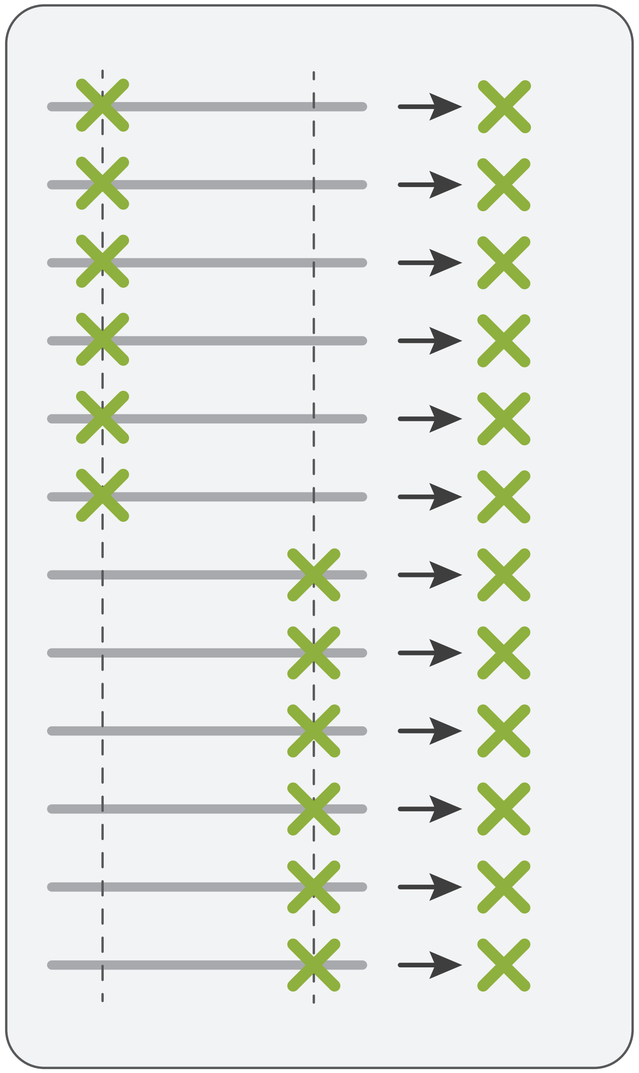


Association

Antagonistic epistasis



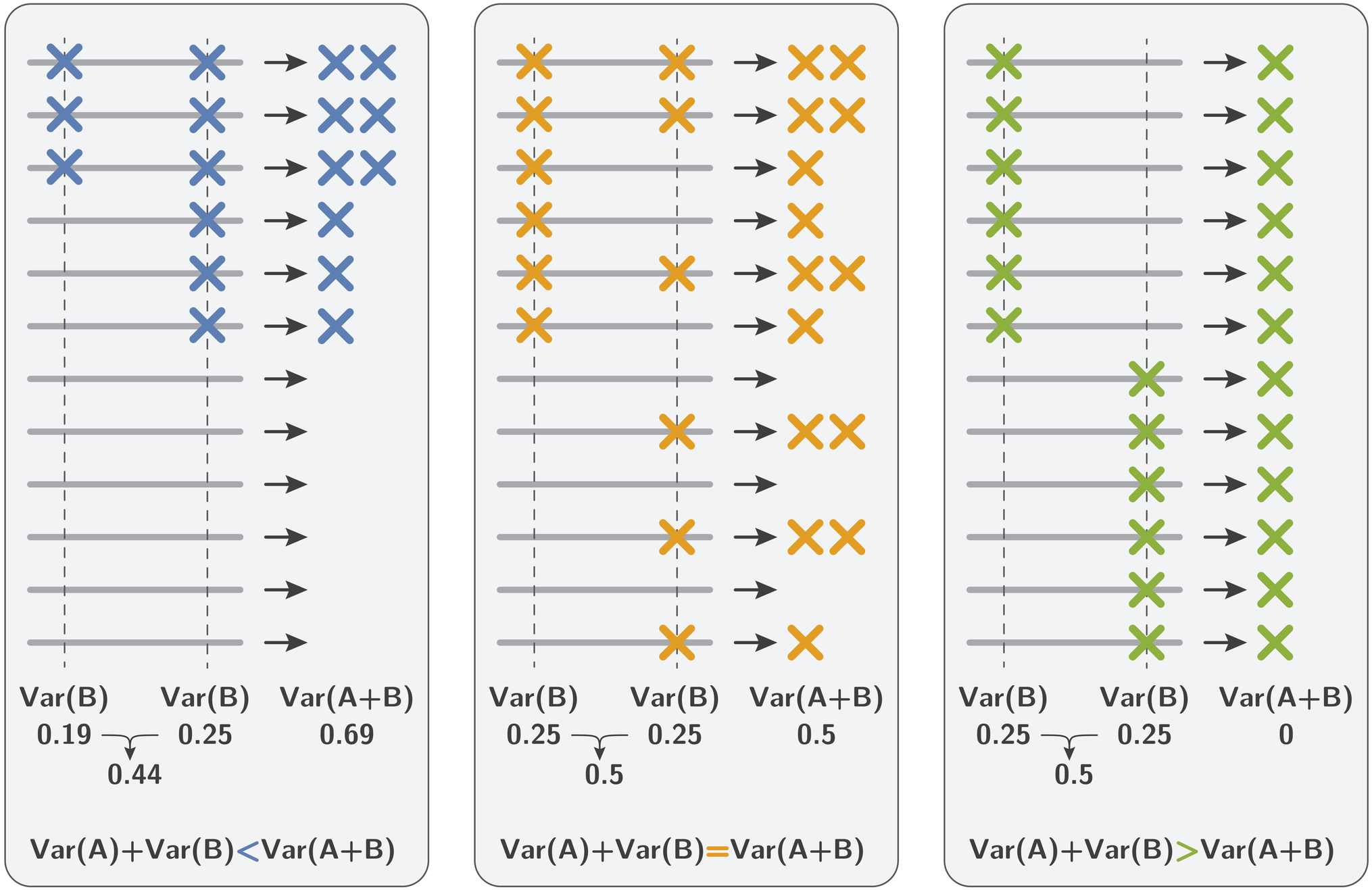
Independence



Repulsion

Synergistic epistasis

Measure of pair-wise LD

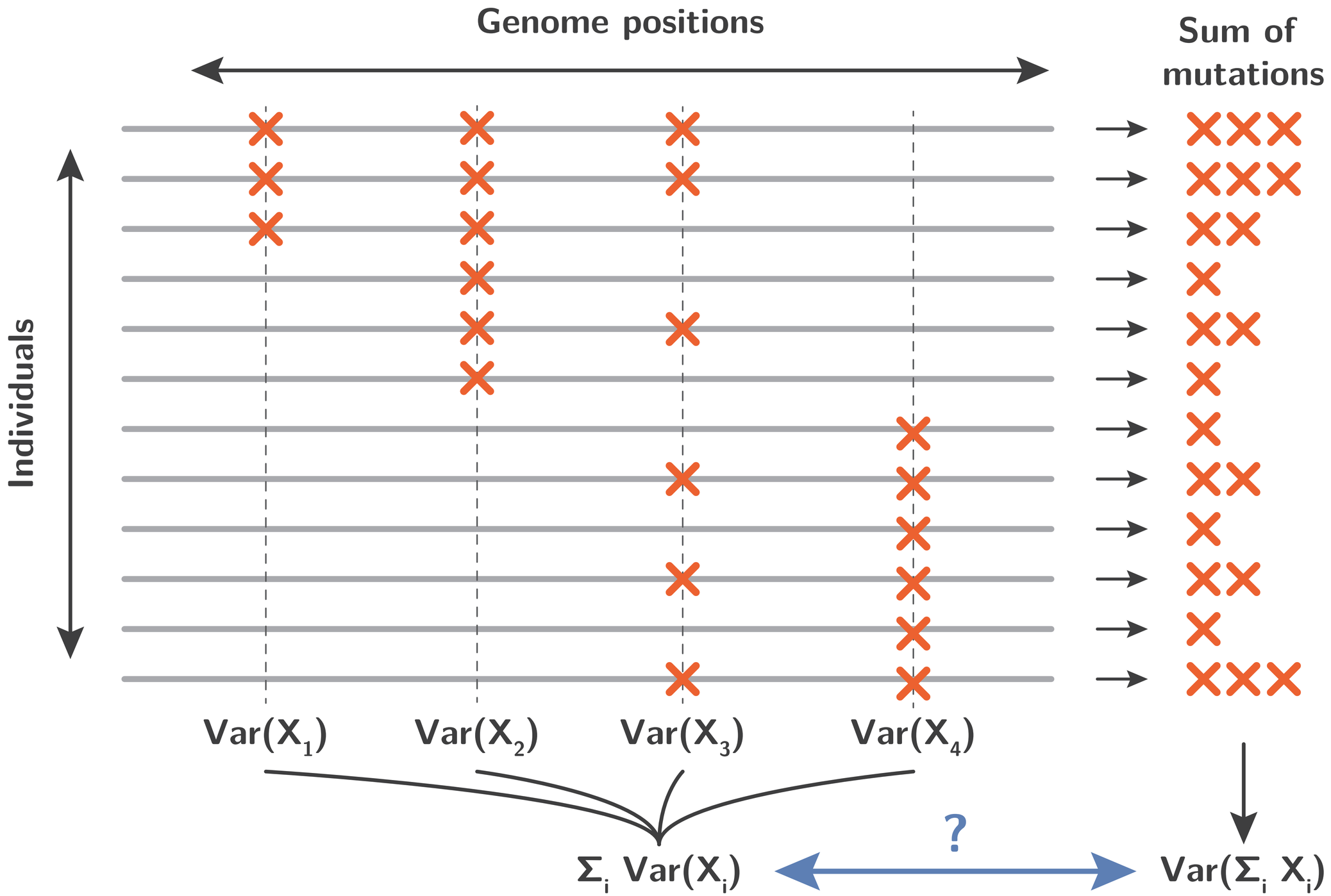


Association
Antagonistic epistasis

Independence

Repulsion
Synergistic epistasis

Measure of genome-wide LD



Part III.

Sohail *et al.* - *Results*

Loss-of-function (LOF) mutations compared to missense and synonymous mutations.

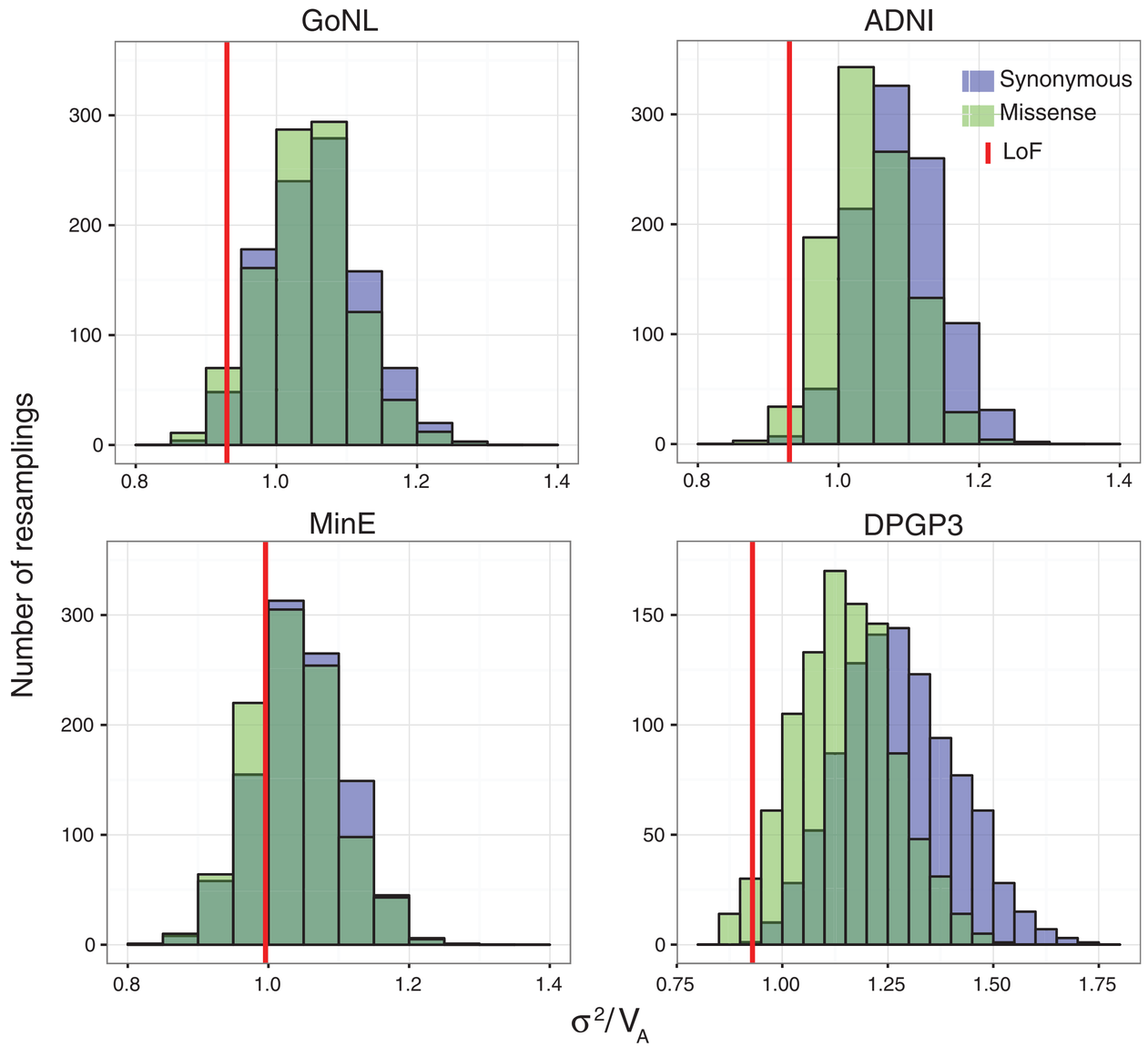


Fig. 3. Resampling distributions of σ^2/V_A for rare LoF mutation burden in humans and *D. melanogaster*. Synonymous (purple) and missense (green) alleles were resampled at the same allele frequency as LoF alleles to obtain empirical null distributions for σ^2/V_A in each data set. For humans, only singletons, and for flies, only alleles up to a minor allele count of 5, are included. A one-sided *P* value for σ^2/V_A of the rare LoF mutation burden (red) was obtained, and a joint *P* value for all three human data sets shown (GoNL, ADNI, MinE) was computed by meta-analysis using Stouffer's method (*11*) (*P* = 0.0003).

Genome-wide LD compared to pair-wise LD

Table 1. Negative linkage disequilibrium (LD) between rare LoF alleles in human and *D. melanogaster* genomes. For humans, only singletons, and for flies, only alleles up to a minor allele count of 5, are included (see tables S2 and S3 for other frequency cut-offs). Net LD is normalized per pair of alleles and per pair of loci (*11*). A one-sided *P* value was obtained for σ^2/V_A by permutation, and a joint *P* value for all three human data sets shown (GoNL, ADNI, MinE) was computed by meta-analysis using Stouffer's method (*11*) (coding synonymous $P = 0.999$, missense $P = 5.155 \times 10^{-4}$, LoF $P = 0.002$). The number of samples is given in parentheses for each data set.

Variant type	Mean	σ^2/V_A	Net LD	
			Per pair of derived alleles	Per pair of loci
Humans				
Genome of the Netherlands GoNL (495)				
Synonymous	30.26	1.675	0.022	4.554×10^{-8}
Missense	60.88	2.077	0.018	3.609×10^{-8}
Nonsense	1.67	0.929	−0.039	-8.013×10^{-8}
Splice	0.90	0.953	−0.049	-1.008×10^{-7}
LoF	2.58	0.930	−0.029	-5.848×10^{-8}
European ancestry ADNI (714)				
Synonymous	38.99	2.077	0.028	2.709×10^{-8}
Missense	77.98	2.008	0.013	1.268×10^{-8}
Nonsense	2.10	0.933	−0.032	-3.126×10^{-8}
Splice	1.16	0.878	−0.104	-1.020×10^{-7}
LoF	3.26	0.930	−0.022	-2.126×10^{-8}
Dutch MinE (601)				
Synonymous	42.93	1.749	0.017	2.414×10^{-8}
Missense	79.34	1.960	0.012	1.675×10^{-8}
Nonsense	1.89	1.057	0.028	3.898×10^{-8}
Splice	0.95	0.972	−0.033	-4.641×10^{-8}
LoF	2.83	0.996	−0.001	-1.727×10^{-9}
D. melanogaster				
Zambian DPGP3 (191)				
Synonymous	3577.06	57.473	0.016	1.658×10^{-6}
Missense	2051.52	18.536	0.008	6.710×10^{-7}
Nonsense	10.21	0.928	−0.007	-4.139×10^{-7}
Splice	2.60	0.948	−0.020	-1.308×10^{-6}
LoF	12.81	0.929	−0.005	-3.298×10^{-7}

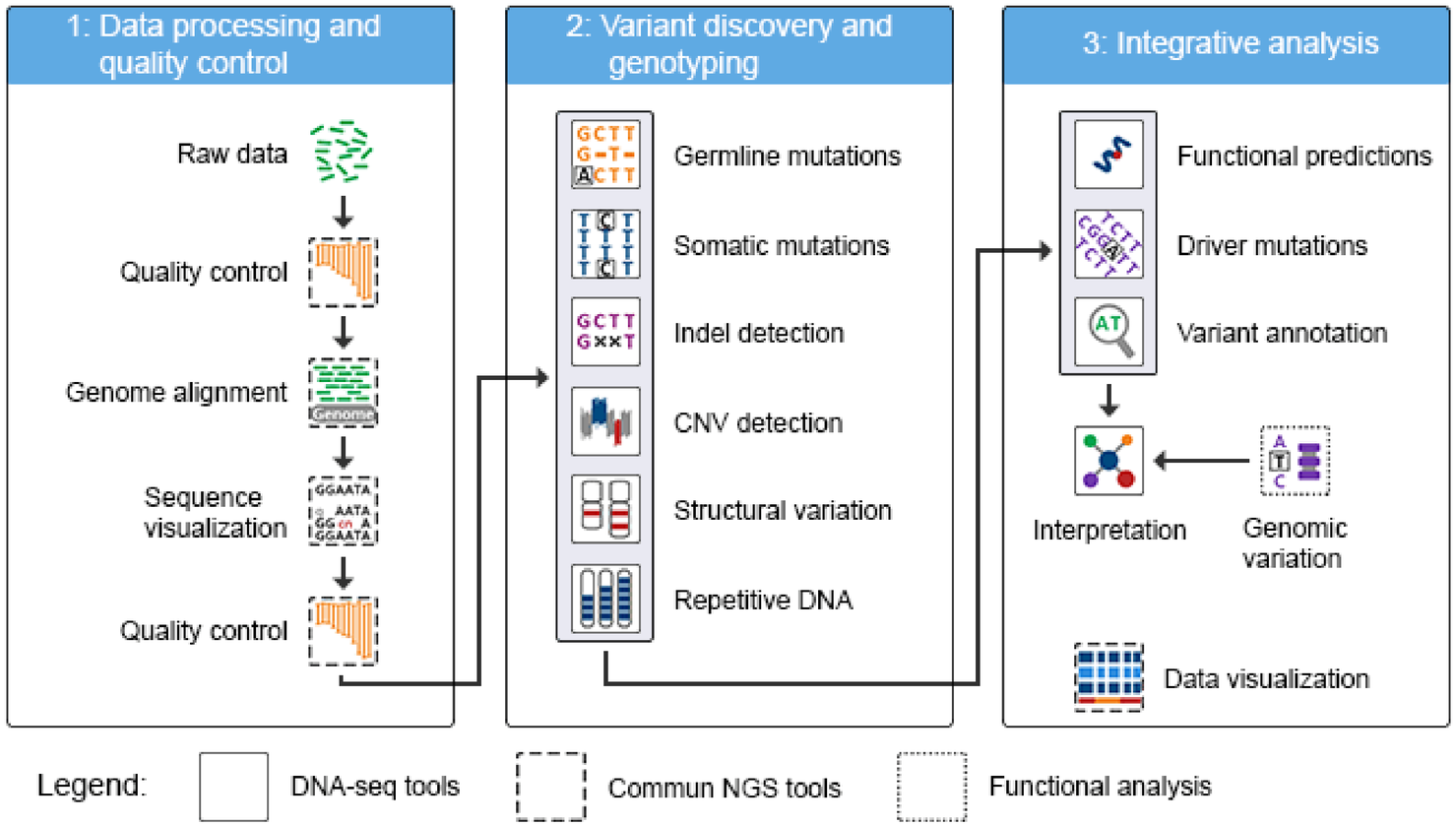
Part IV.

Reproductibility of the study

Next-Generation Sequencing

ENS de Lyon - Carine Rey & Marie Semon

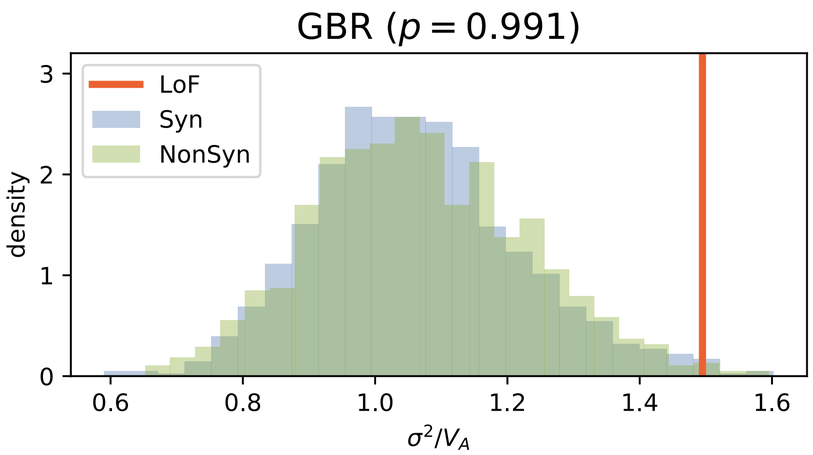
Part I (3 days): from raw data to variant calling in humans.

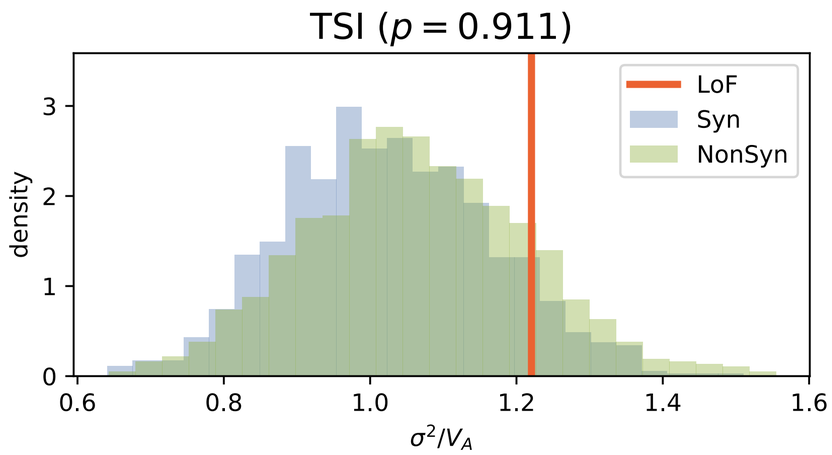


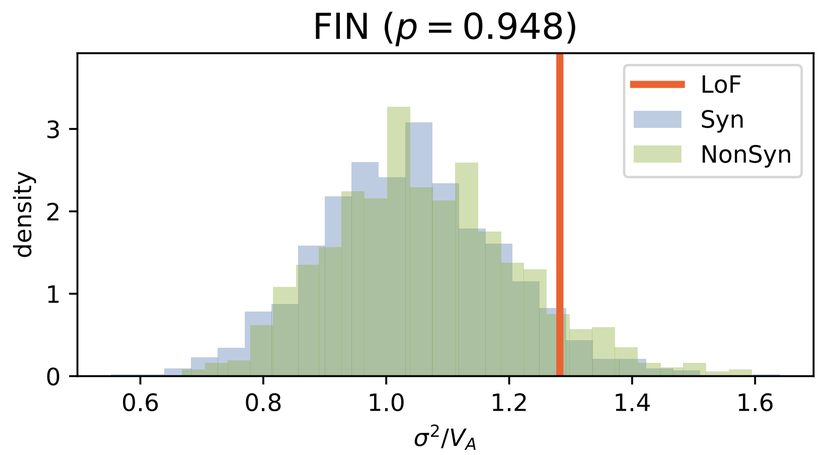
Part II (3 days): reproducing Sohail *et al* using 1000 genomes SNP dataset

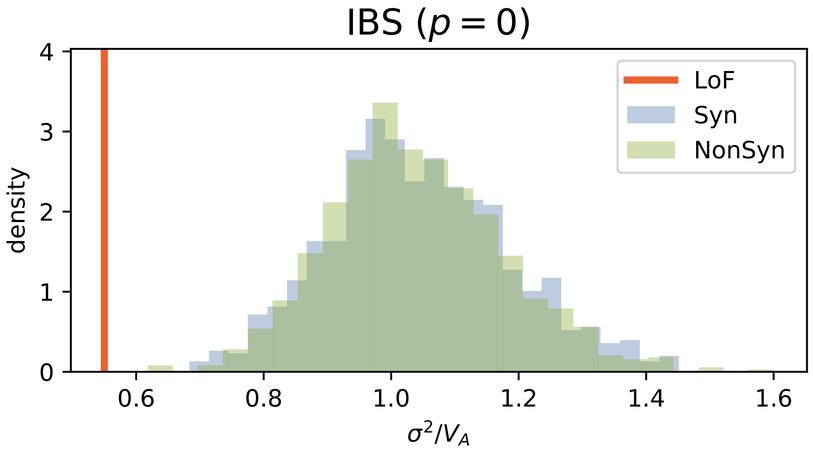
https://omictools.com

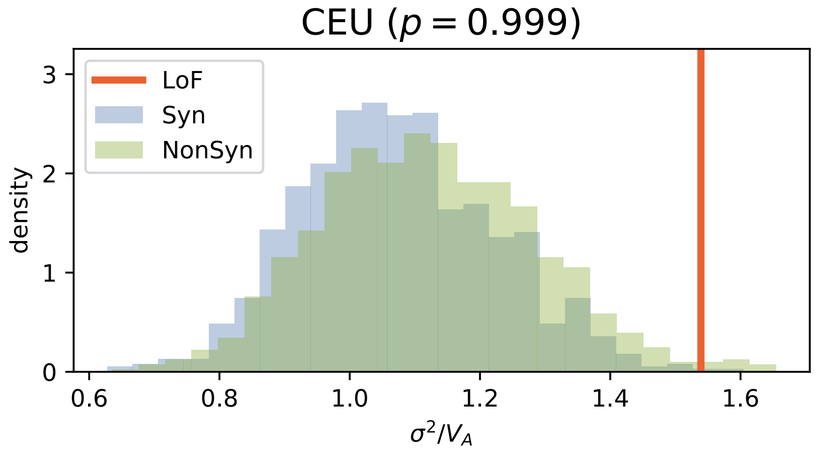
Reproducing Sohail *et al* on 1000G dataset (1/3)

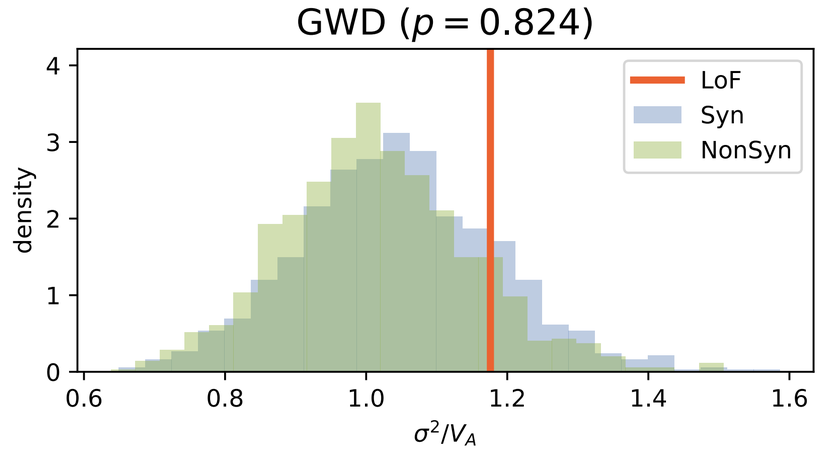


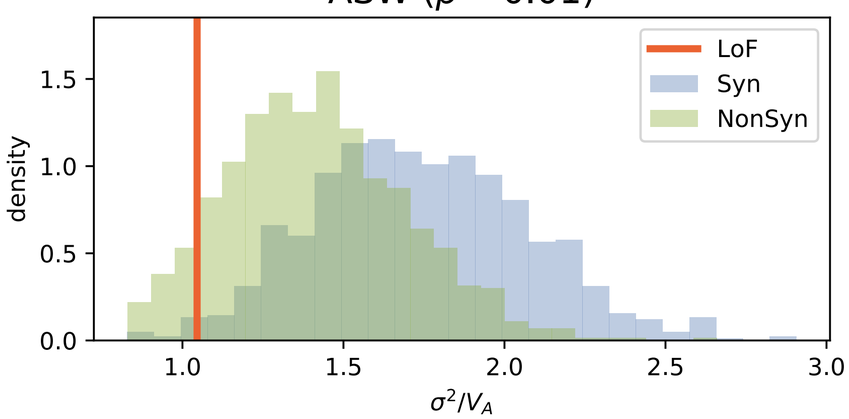


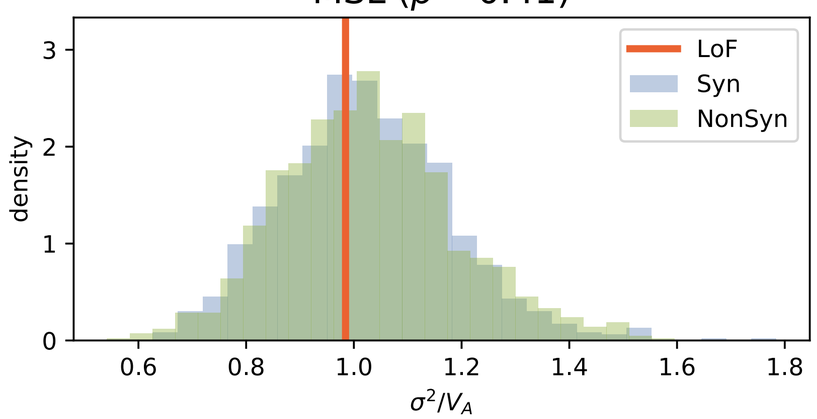


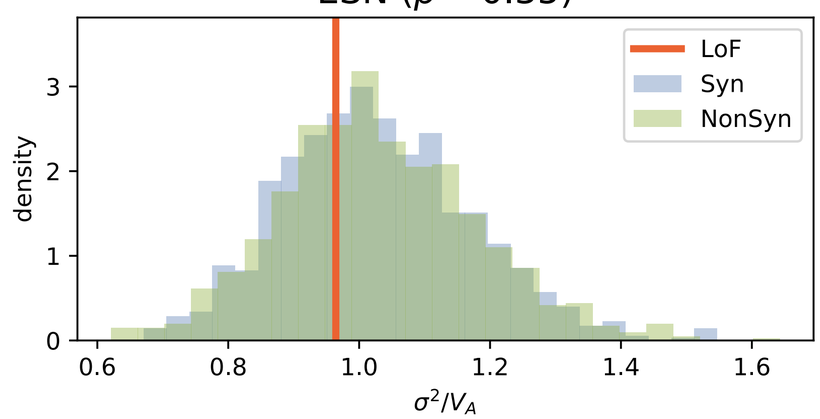




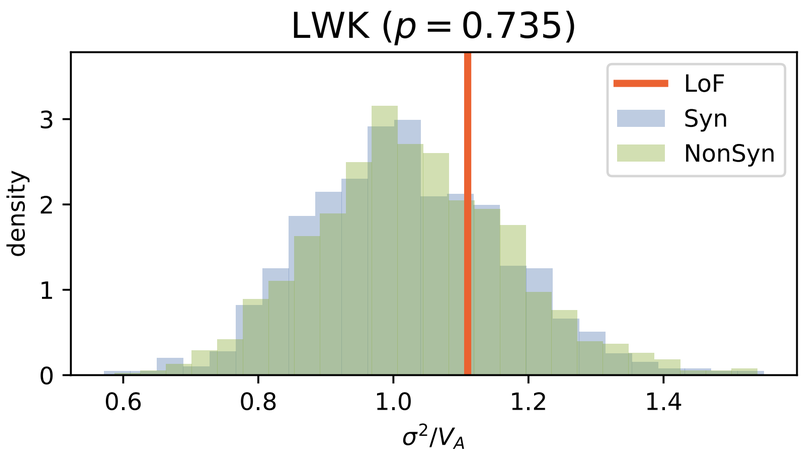


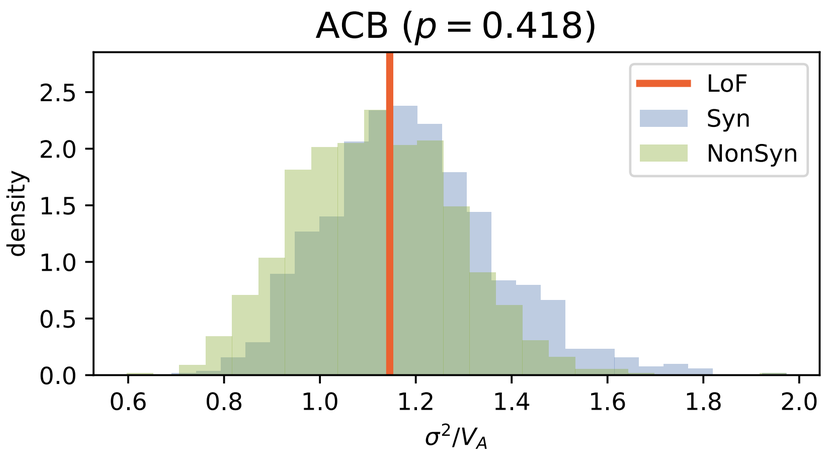


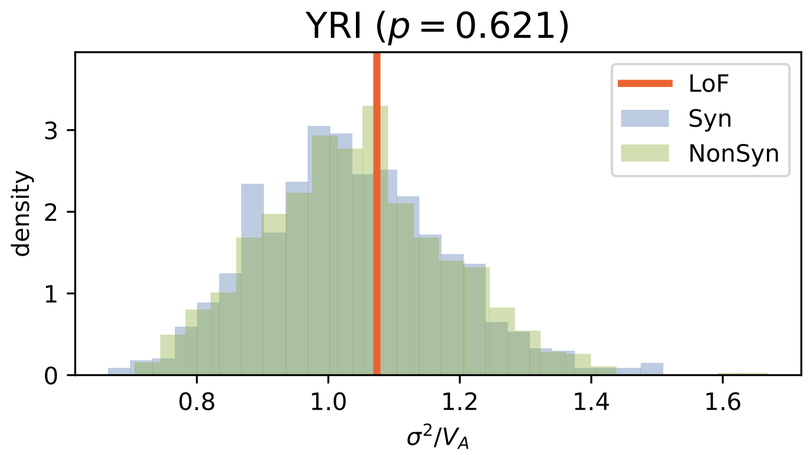


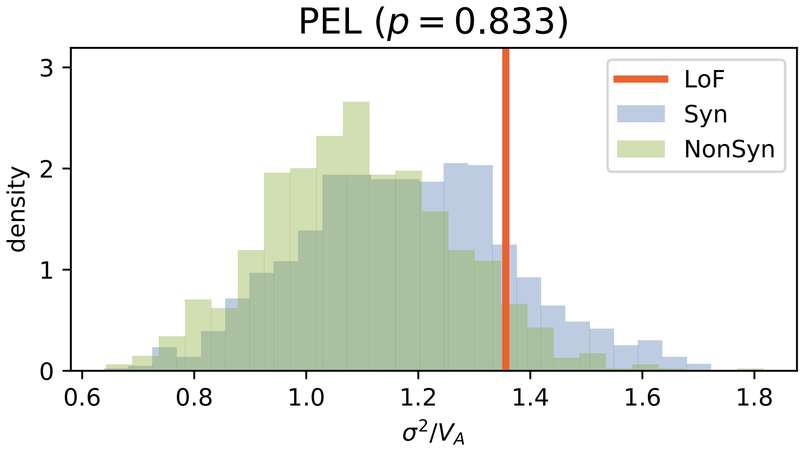


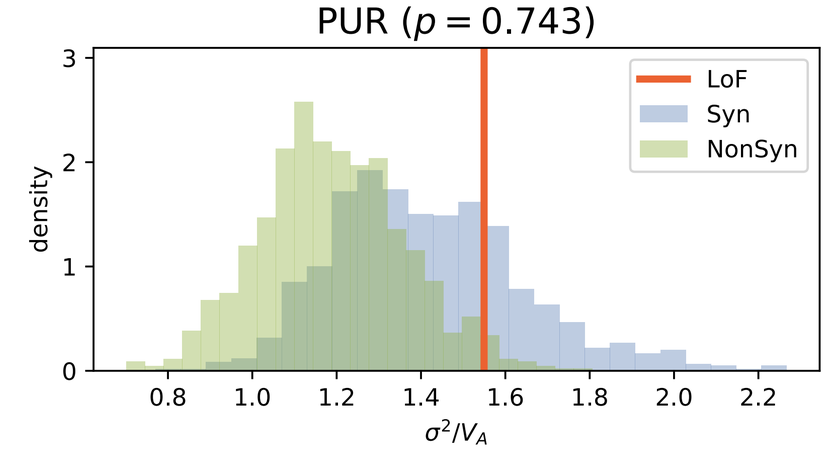
Reproducing Sohail *et al* on 1000G dataset (2/3)

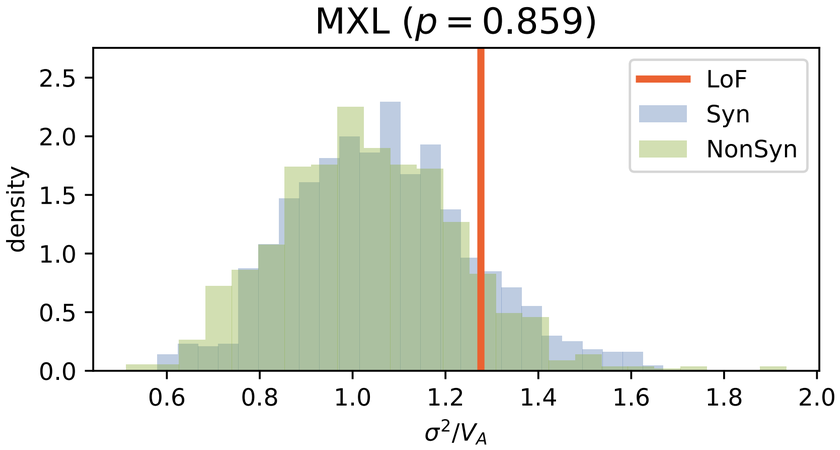


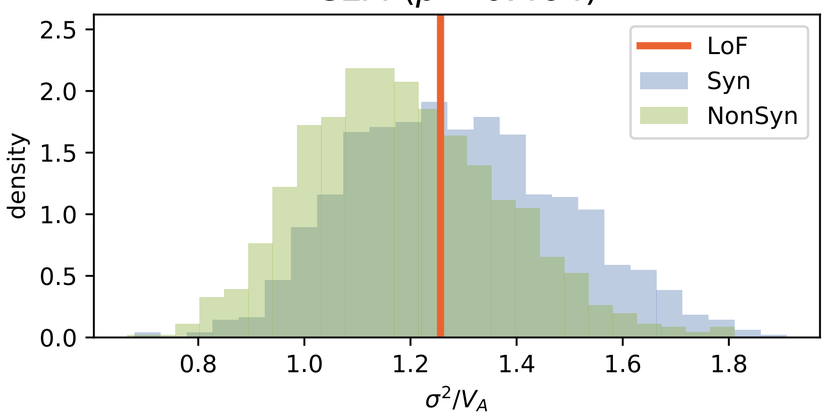


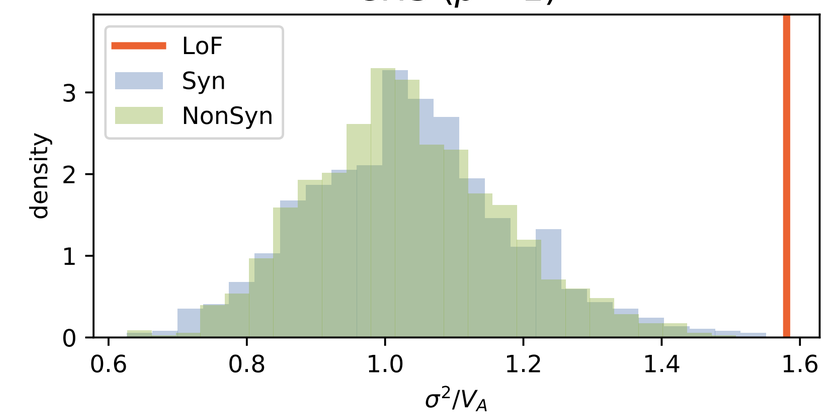


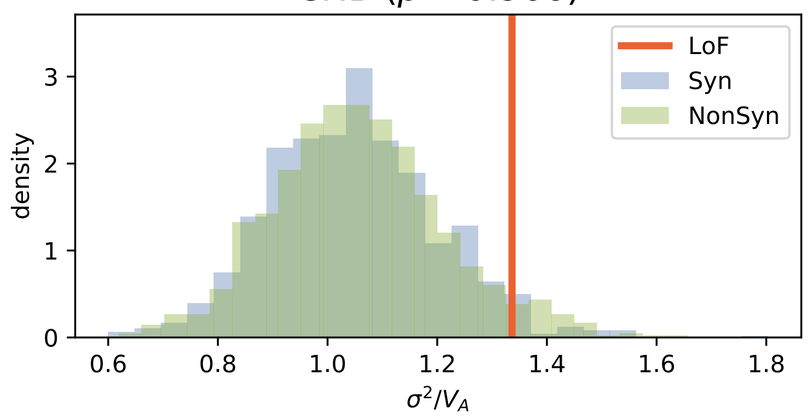




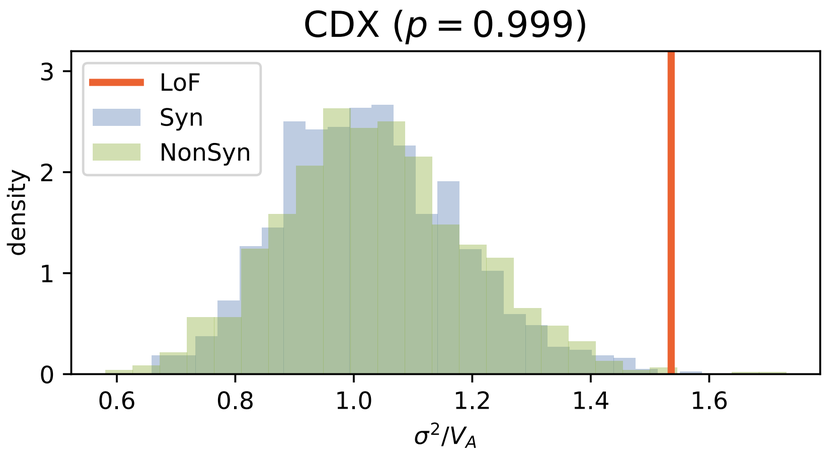


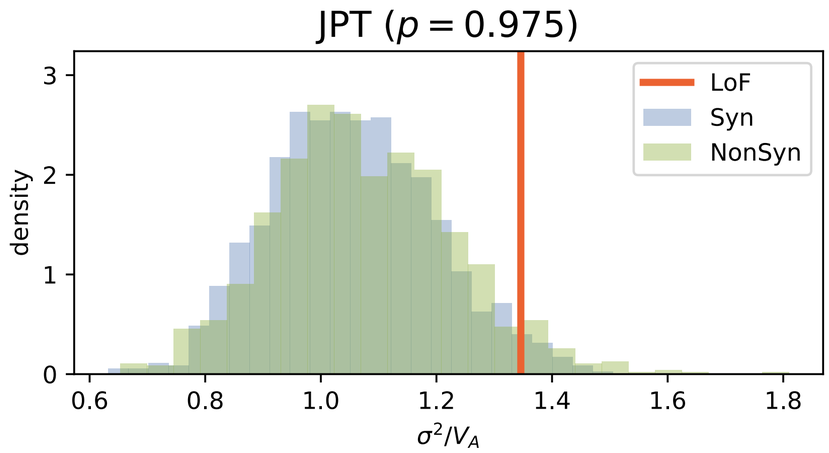


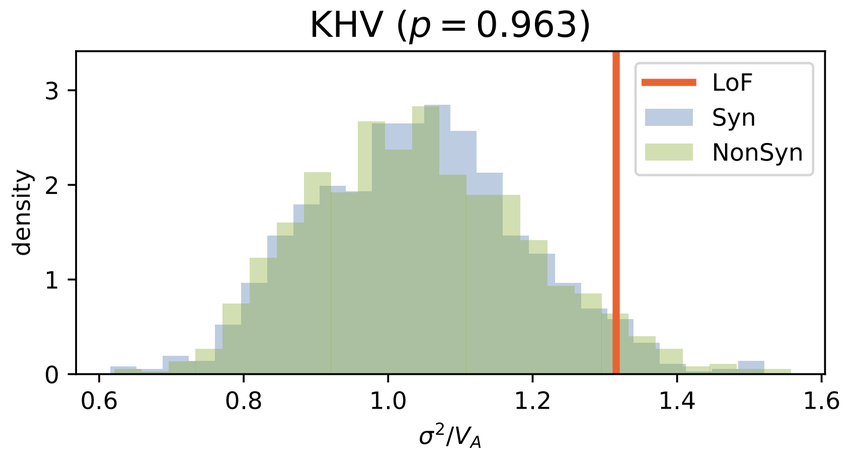


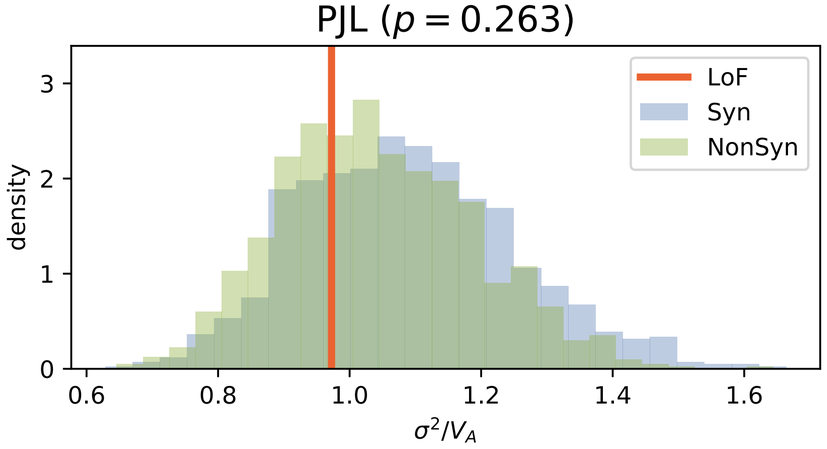


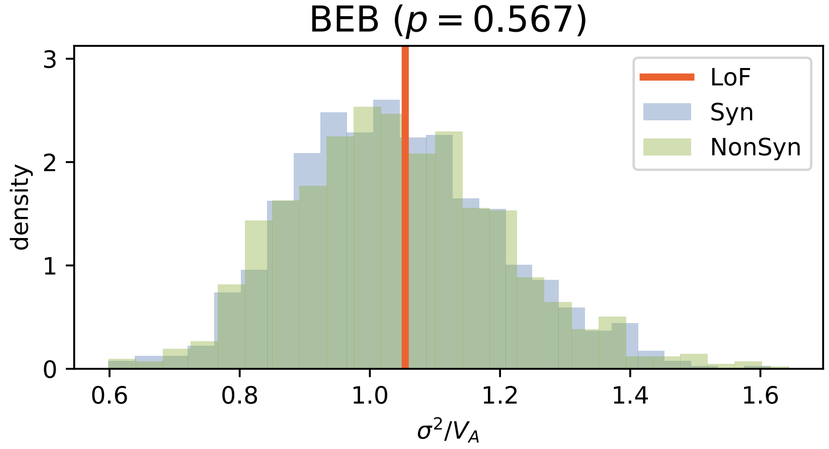
Reproducing Sohail *et al* on 1000G dataset (3/3)

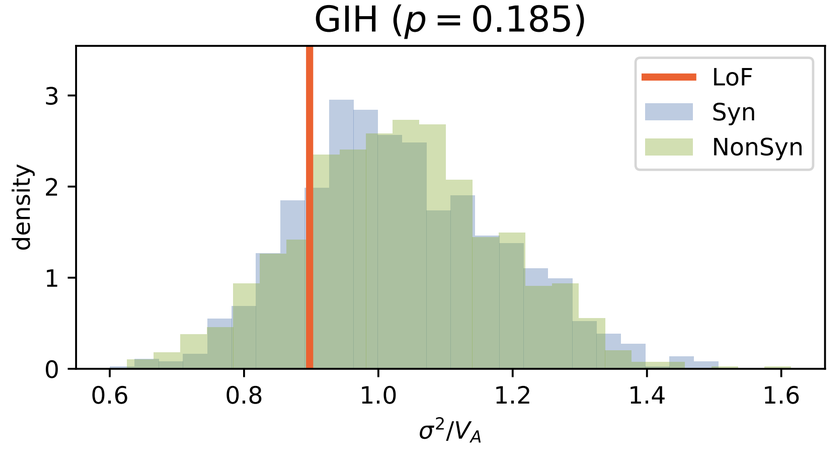


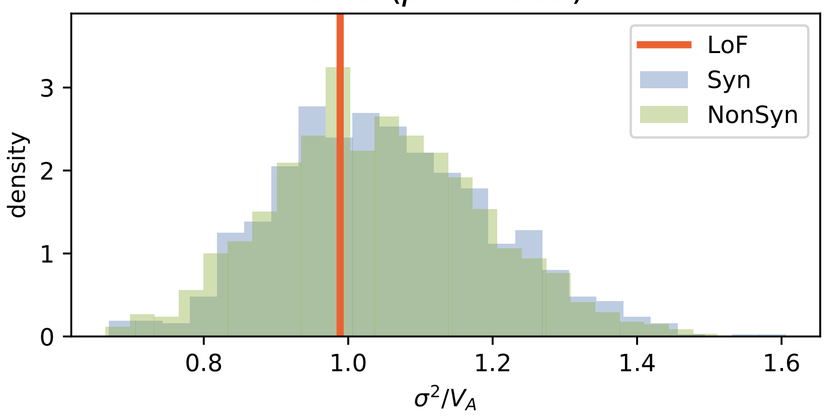


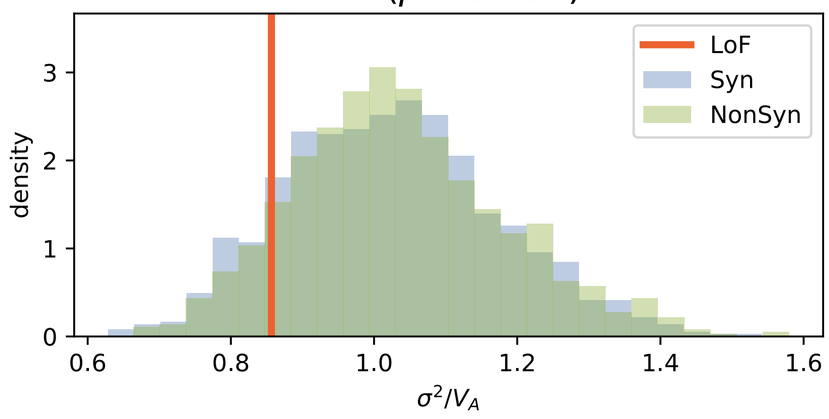












Conclusions

I. Not so sure that epistasis is negative genome-wide in humans as stated in Sohail *et al*.

II. The evolutionary advantage of recombination is still an open question. Negative epistasis is not a fondamentaly necessary assumption.

III. Discrepancy between theory and data in the field of recombination.

"Models that address the evolution of recombination rate were generated to explain the evolutionary advantage of recombination, rather than quantitative differences in rate among individuals."

Dapper & Payseur, 2017