# Population Genetic Attributes of Variants That Disrupt Protein-Protein Interactions

CORNELL UNIVERSITY · FOUNDED A.D. 1865

Cornell CALS
College of Agriculture and Life Sciences

*Mitchell G. Lokey[1,2], Robert Fragoza[1], Philipp W. Messer[2], Haiyuan Yu[1,2], Andrew G. Clark[1,2]*
[1]Department of Molecular Biology & Genetics, Cornell University, Ithaca, New York 14853, USA
[2]Department of Computational Biology, Cornell University, Ithaca, New York 14853, USA
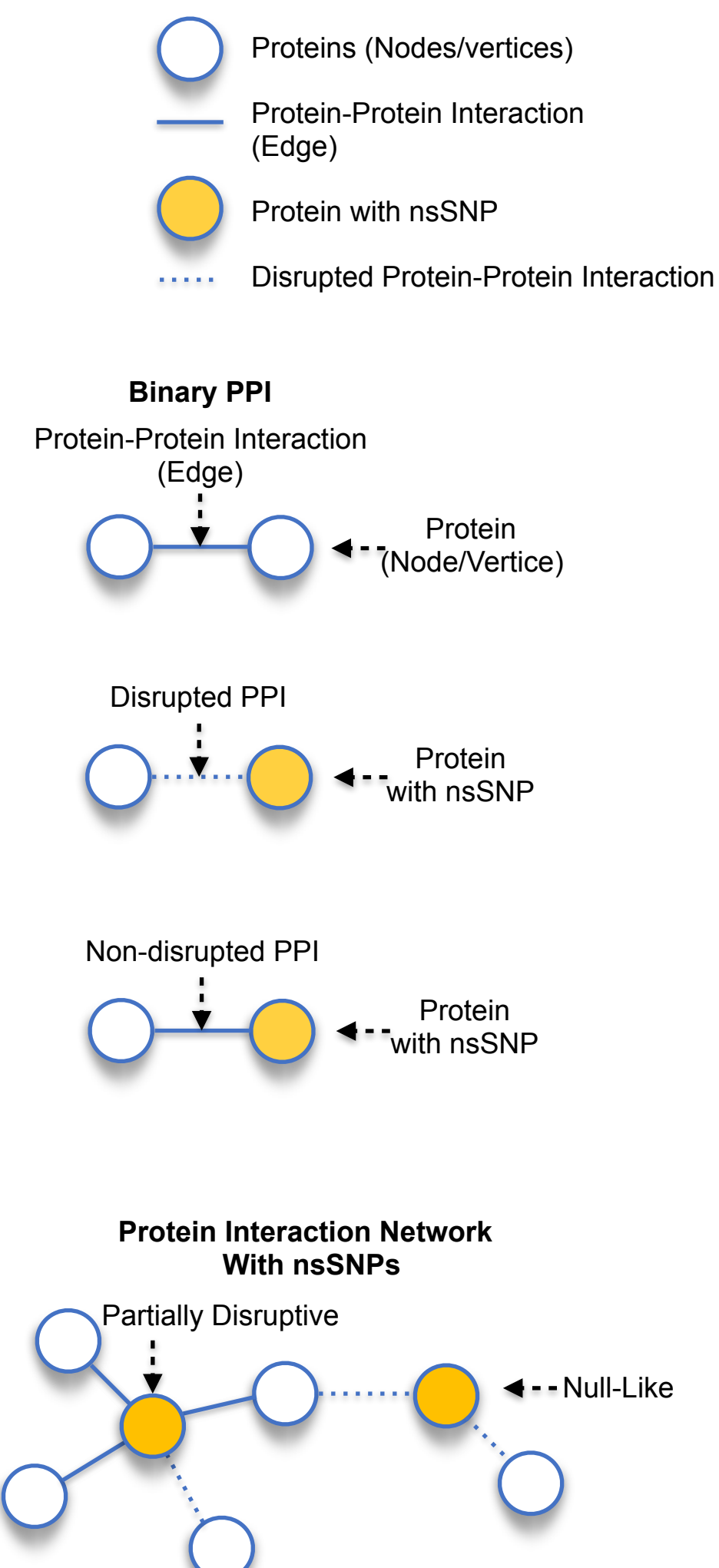Email: mgl77@cornell.edu  @mitch_lokey

## Background

- We seek improved methods to discover genotype to phenotype relationships. One approach to this problem focuses on the network of macromolecular interactions known as the interactome. Of great importance for cellular function, and our focus here, are networks of Protein-Protein Interactions (PPIs).[1]
- Yeast-2-Hybrid (Y2H) and similar assays have allowed mapping of the human protein interaction network by identifying large numbers of binary PPIs.[1]
- Studies of the effects of disease-associated variants on the known network of PPIs have shown that ~1/2 of tested Mendelian disease-associated variants disrupt PPIs.[2]
- Until recently, no work had looked at the extent to which naturally segregating variants in the human population affect PPIs. Fragoza et al (2019) sampled variants from the ExAC database and tested PPIs of >2,000 non-synonymous SNPs (nsSNPs) with the Y2H assay.[3]
- Fragoza et al (2019) showed that 10.5% of tested variants disrupted PPIs, a larger estimate of functional variation than previously reported. Surprisingly they also found that disruptive variants often segregate at relatively high allele frequencies.[3]
- As the links between cellular phenotypes, such as PPIs, and impacts on fitness are not straightforward we seek to illuminate population genetic attributes of PPI disruptive variants by leveraging large human databases and simulations.



Proteins (Nodes/vertices)
Protein-Protein Interaction (Edge)
Protein with nsSNP
Disrupted Protein-Protein Interaction

**Binary PPI**
Protein-Protein Interaction (Edge) — Protein (Node/Vertice)
Disrupted PPI — Protein with nsSNP
Non-disrupted PPI — Protein with nsSNP

**Protein Interaction Network With nsSNPs**
Partially Disruptive — Null-Like

## Naturally Segregating Variants & PPIs

- Fragoza et al investigated fitness impacts of PPI variants using PolyPhen-2 a machine learning tool that probabilistically classifies the deleteriousness of missense mutations. PolyPhen-2 scores show that PPI disruptive variants are more deleterious than non-disruptive and even more so for variants that disrupt all of their PPIs.
- Theory tells us that deleterious variants such as disease-associated variants, and likely PPI disruptive variants, will be kept at low allele frequencies due to purging selection. However, Fragoza et al found that although ~20% of very rare variants disrupt PPIs, ~9.6% of common variants (allele frequency > 10%) also disrupt PPIs.[3]
- To better understand the evolutionary history and impact of PPI disruptive variants we first looked at the ages of our disruptive and non-disruptive variants using Genealogical Estimation of Variant Age (GEVA).[4] If PPI disruptive variants are indeed deleterious they should be younger than non-disruptive variants due to purging.
- Inbreeding can increase the allele frequency of neutral variants, however excess homozygosity due to inbreeding can also expose and thus drive out recessive deleterious variants. To test for differences of PPI variants presence in long Runs of Homozygosity (ROH >1.5 Mb) we used genomic data from the UKBiobank.
- Recent human demography has led to an excess of rare variation and differences between populations. We used forward genetic simulations to test if this could lead to the observed allele frequency of PPI disruptive variants.
- We find PPI disruptive variants are younger and less likely to be in ROH than non-disruptive variants. Simulations also show that a small number of recessive moderately deleterious variants can reach common allele frequencies.


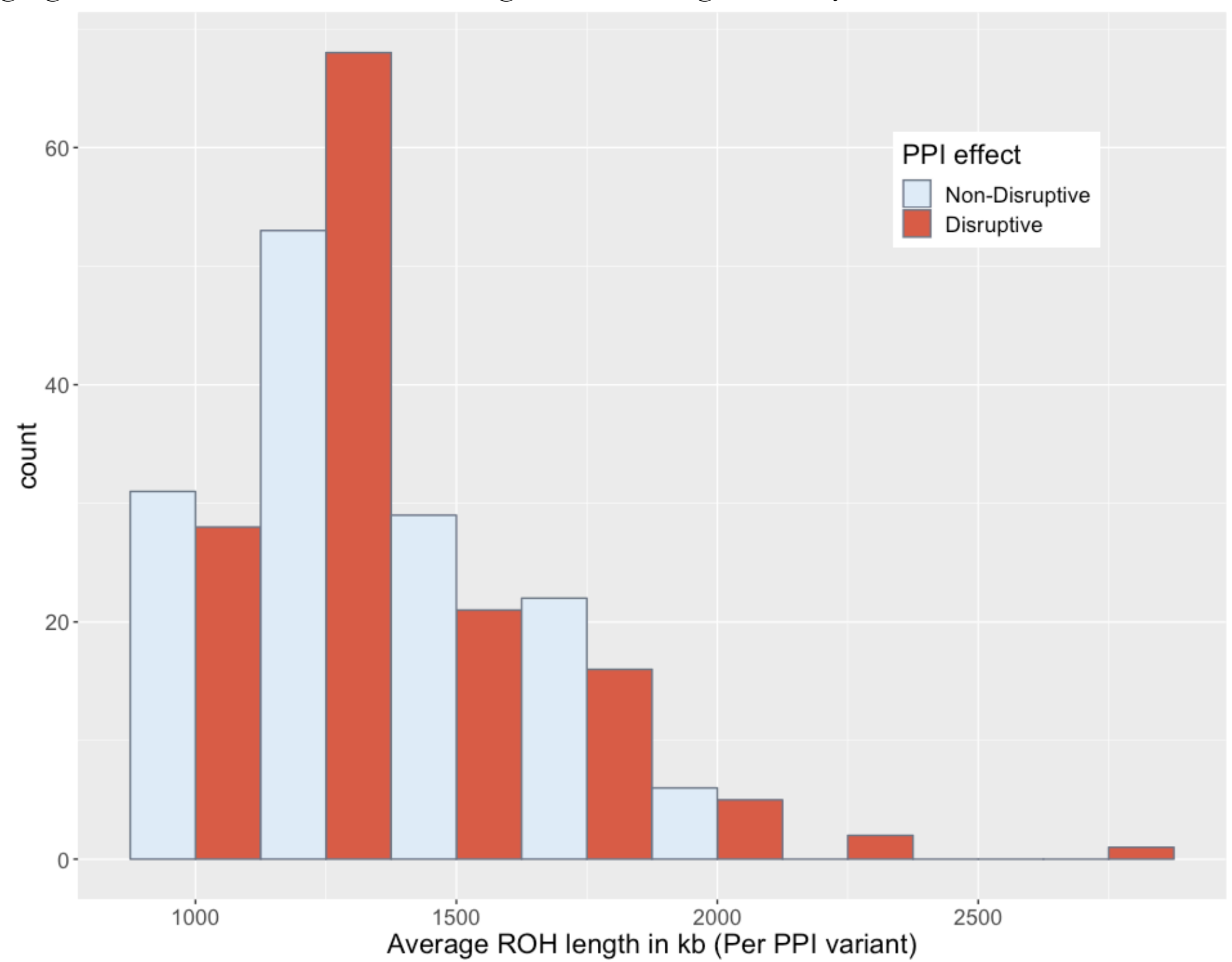
Figures from Fragoza et al 2019
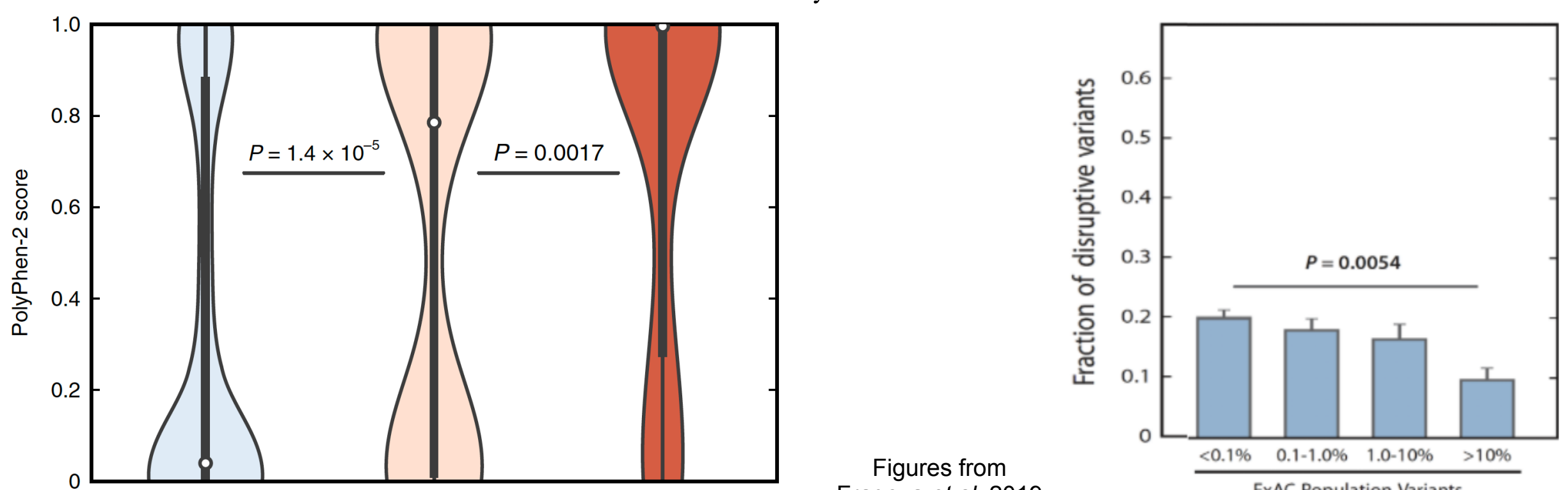
## Variant Age

- GEVA (Albers & McVean 2020) takes shared haplotype segment pairs, and infers the posterior density for time-to-most-recent-common-ancestor (TMRCA) using a joint molecular and recombination coalescent-based clock model.[4]
- We determined PPI variant ages using the 1000 Genomes and Simons Genome Diversity Projects, and see a significant (p<0.05) difference in the allelic ages for disruptive and non-disruptive variants using a Mann-Whitney U test.
- The younger age of PPI disruptive variants supports the idea that natural selection is removing these variants from the population, leaving only younger alleles to be counted.
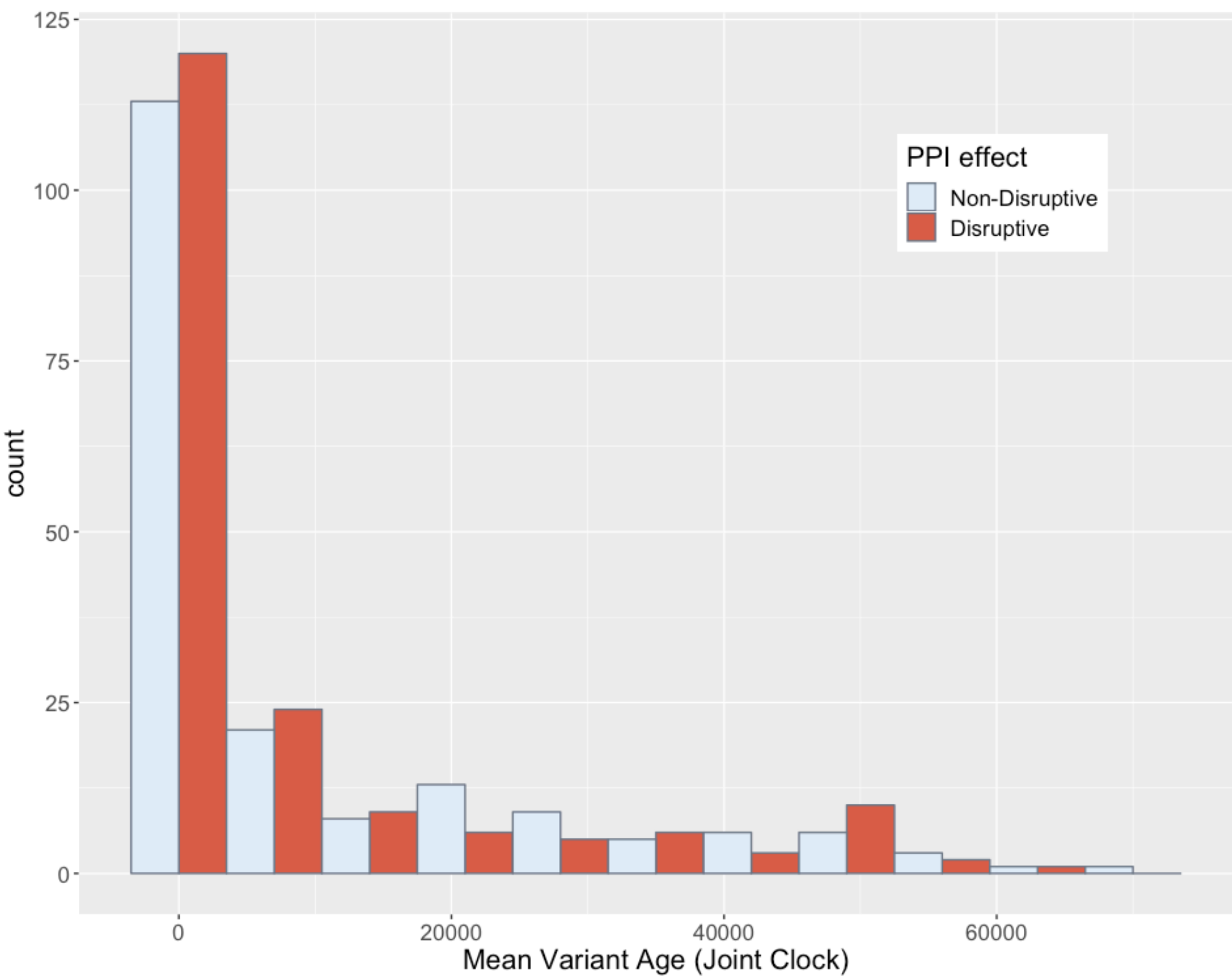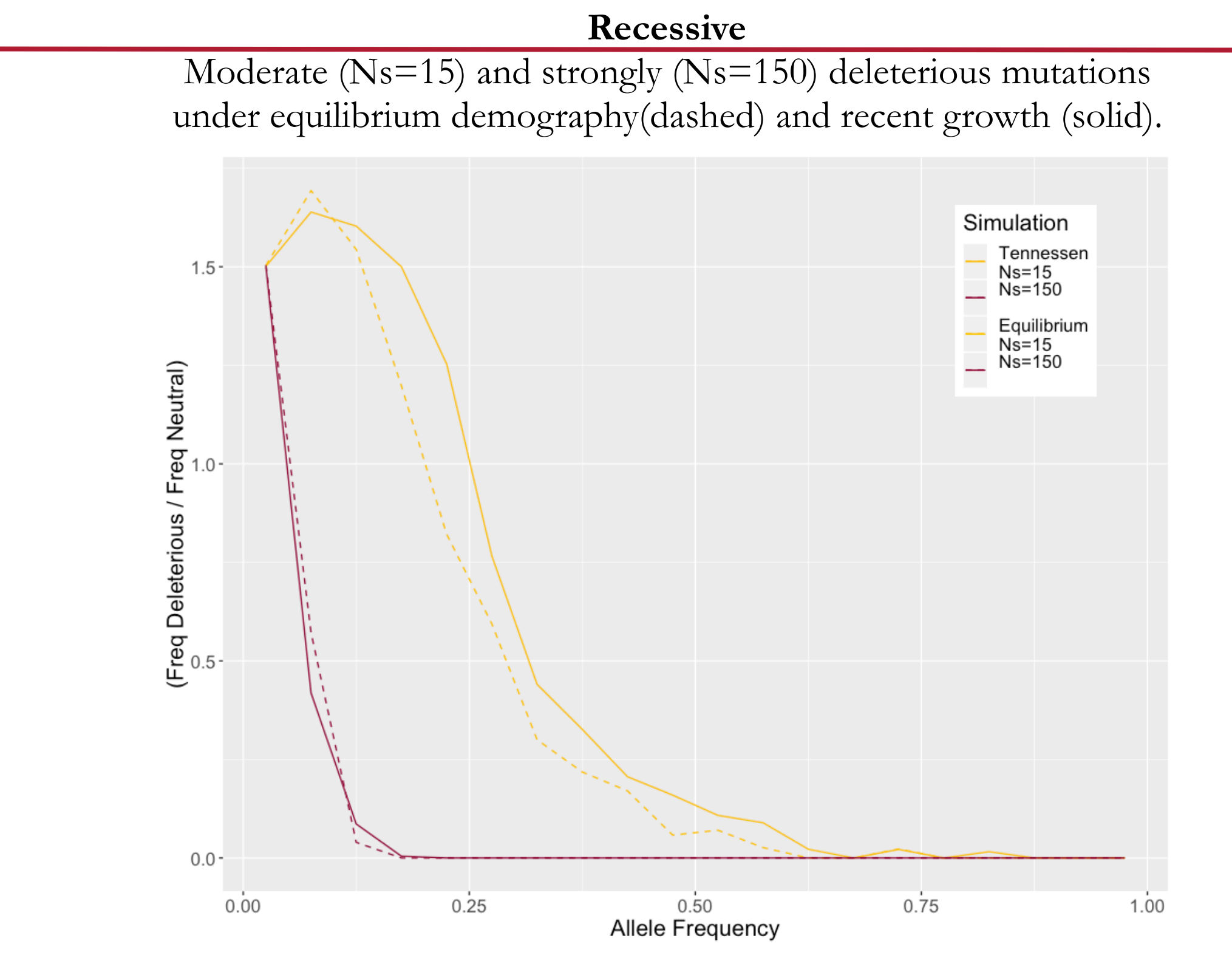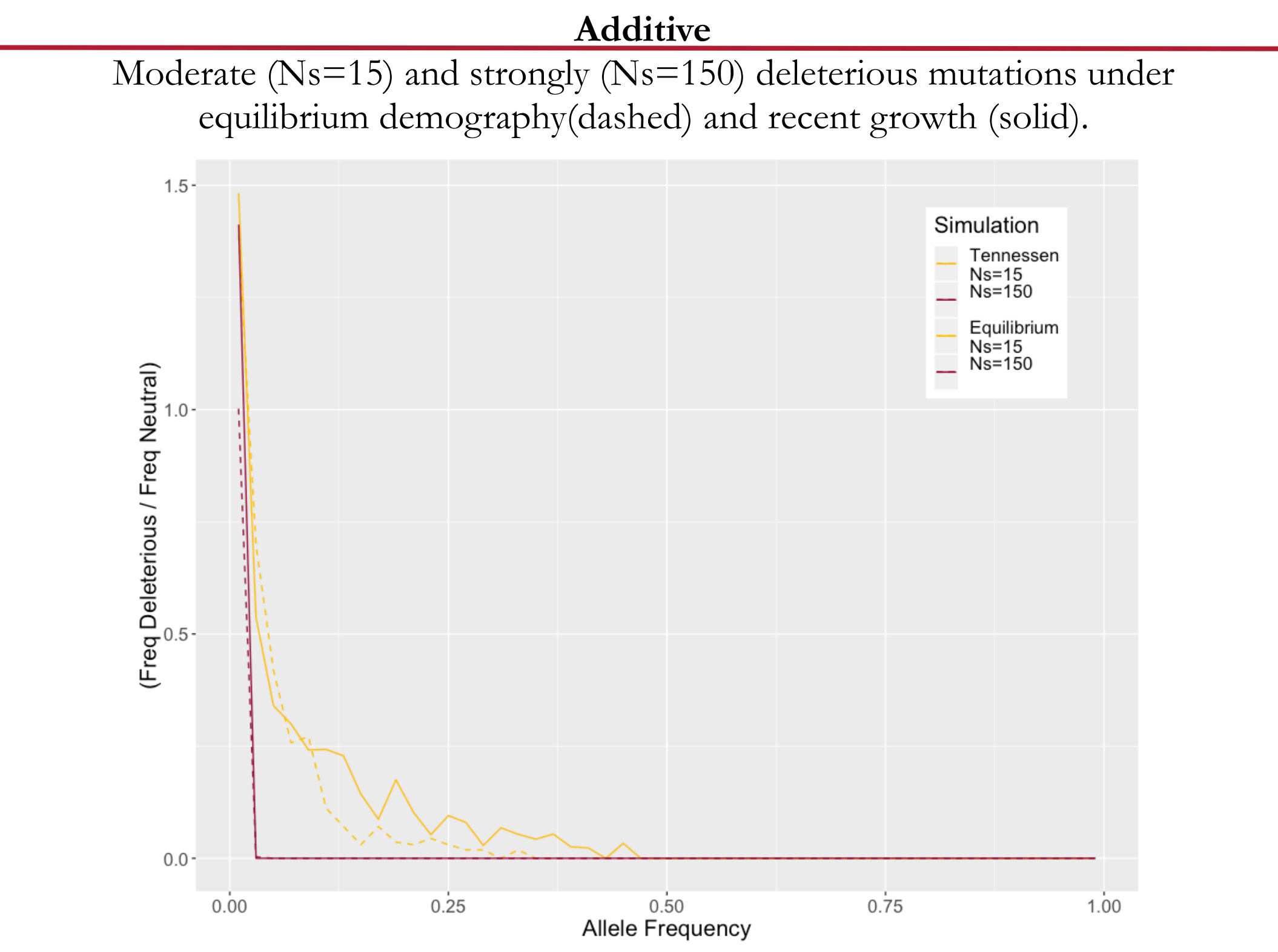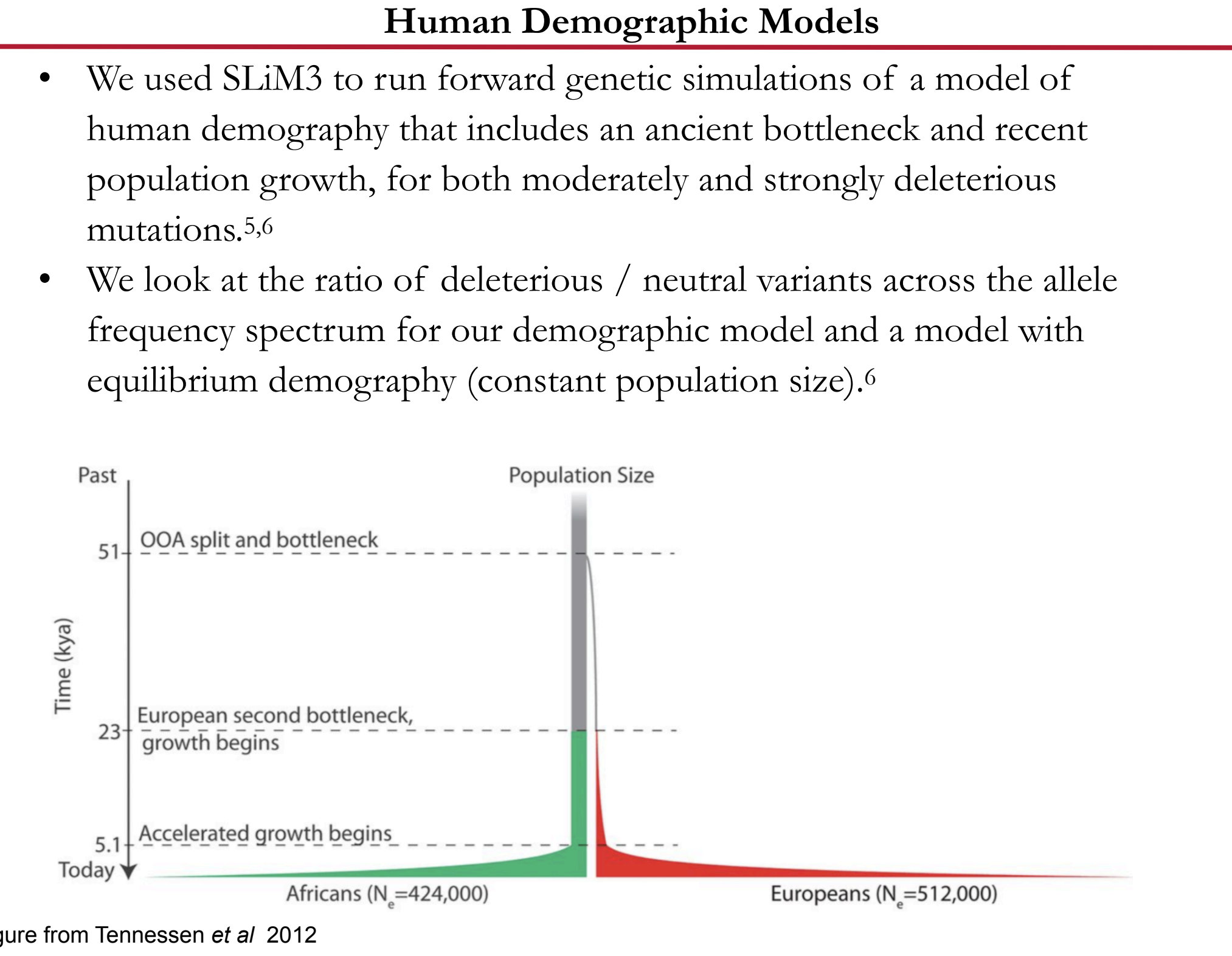


## Runs Of Homozygosity

- Long ROH are associated with non-random mating and inbreeding which can increase variant allele frequency. Deleterious effects of recessive PPI disruptive variants would be exposed in individuals with long ROH.
- We used the UKBiobank to determine which PPI variants were found in long ROH and calculated an Odds Ratio of 0.72 (Fisher & χ² p<0.005) suggesting that disruptive SNPs are less likely to be found in ROH. Likely because of purging selection. However, ROH lengths do not significantly differ.

|  | In ROH | Not in ROH |
|---|---|---|
| Disruptive | 248 | 171 |
| Non-disruptive | 1287 | 645 |



## Simulations of Human Demography

### Human Demographic Models

- We used SLiM3 to run forward genetic simulations of a model of human demography that includes an ancient bottleneck and recent population growth, for both moderately and strongly deleterious mutations.[5,6]
- We look at the ratio of deleterious / neutral variants across the allele frequency spectrum for our demographic model and a model with equilibrium demography (constant population size).[6]



Figure from Tennessen et al 2012

### Additive
Moderate (Ns=15) and strongly (Ns=150) deleterious mutations under equilibrium demography(dashed) and recent growth (solid).



### Recessive
Moderate (Ns=15) and strongly (Ns=150) deleterious mutations under equilibrium demography(dashed) and recent growth (solid).



## Conclusion

### Takeaways

- Taken together, the younger age and lower chance of lying in long ROH suggest that PPI disruptive variants are indeed deleterious.
- And simulations show that recent demography can push recessive deleterious mutations to higher frequencies as observed in PPI disruptive variants, although the vast majority of common polymorphisms are neutral.

### Future Directions

- Investigating the evolutionary signatures of PPI disruptive variants may help us to better identify which variants within a protein interaction network are important for disease.
- The results presented here should be tested against a control set of variants with similar allele frequencies.

### Acknowledgements

- This work is supported by NIH grant R01 GM125639 to H. Yu and A.G. Clark
- Clark Lab
- Messer Lab
- Yu Lab
- Sarah Kunz

Clark Lab
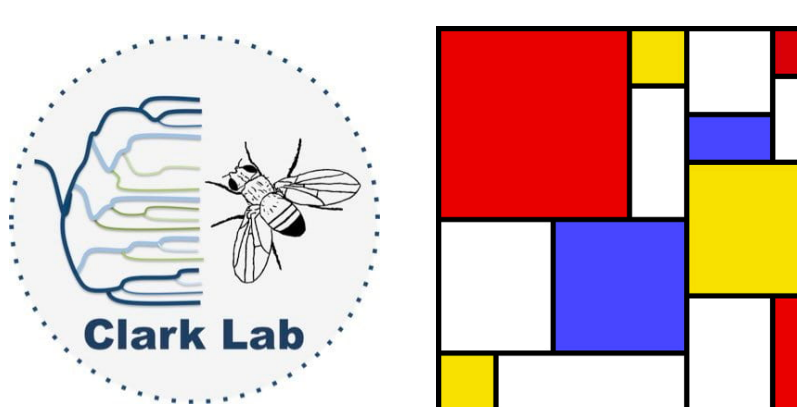
## Citations
1. Vidal, M., Cusick, M., & Barabasi, A.L., 2011. Interactome Networks and Human Disease. Cell. 986-998.
2. Sahni et al 2015. Macromolecular Interaction Perturbation Human Disorders. Cell. 161:3, 647-660.
3. Fragoza, R., et al 2019. Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations. Nat Comm.
4. Albers & McVean 2020. Dating genomic variants and shared ancestry in population-scale sequencing data. PLoS Biology. 18(1):e30000586.
5. Haller, B.C., Messer, P.W., 2019 SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. Molecular Biology and Evolution. 36:3, 632–637.
6. Tennessen et al 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 337:64-9.