

Improving Genome-wide Significance Thresholds in COVID19 Admixture Mapping Study

Sasha Yousefi^{1*}, Jessica Houghton^{1*}, Daniel Mas Montserrat², Alexander G. Ioannidis^{2,3}

*Contributed equally

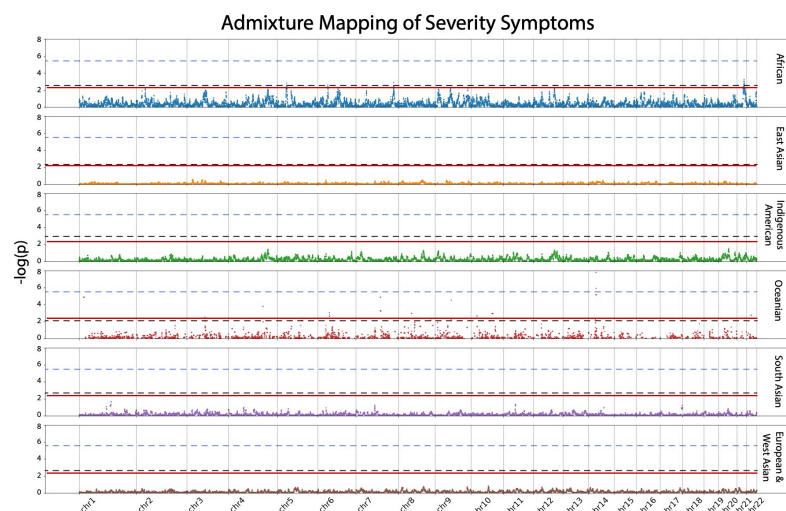
¹Department of Statistics, Stanford University

²Department of Biomedical Data Science, Stanford University

³Institute for Computational and Mathematical Engineering, Stanford University



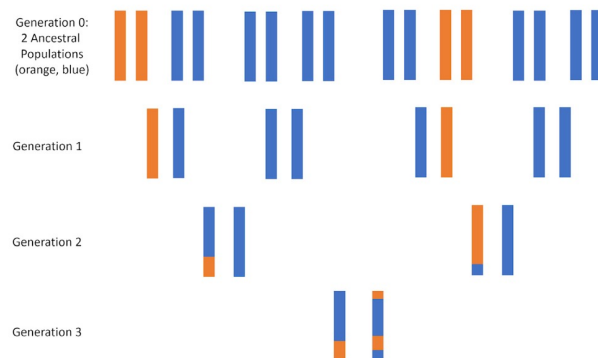
“Deconvoluting complex correlates of COVID19 severity with local ancestry inference and viral phylodynamics: Results of a multiomic pandemic tracking strategy” by Parikh et al.



- Black dashed line: less stringent ancestry-specific multiple test correction
- Blue dashed line: naïve Bonferroni correction (more conservative)
- Regions in chromosome 5 of African ancestry and chromosome 14 of Oceanian ancestry showed a significant association with high-severity COVID19
- Red line: estimated significance threshold from results

STEAM: Significance Threshold Estimation for Admixture Mapping by Grinde et al.

- Multiple hypothesis testing problem: the number of independent hypothesis tests depends on the amount of admixture



- Goal: obtain a more accurate and less conservative significance threshold
- STEAM: analytic and simulation based approach to obtain genome-wide significance thresholds
- Appropriate significance threshold depends on the number of ancestral populations, generations since admixture, and population structure of the sample.

Methods:

1. Calculate the observed correlation of local ancestries at pairs of loci for our samples
 1. Attained dataframe consisting of ancestry pairs, recombination fraction between loci, and correlation
2. Use non-linear least squares to estimate the number of generations since admixture:

$$\text{Corr} = a + b \times (1 - \theta)^g$$

3. Estimate the p-value threshold which will control the family-wise error rate at the 0.05 level.

Results:

- Estimated generations since admixture : 11 (11.357)
 - Using 30 years per generation, this translates to 330 years ago, which is around 1690.
 - This makes sense, as 1690 is the middle of the colonial period when the majority of our six ancestries were likely to admix!
- Naive Bonferroni correction: 2.57×10^{-6}
- Estimated adjusted p-value: 3.98×10^{-3}
 - 95% confidence interval: $[3.61 \times 10^{-3}, 4.39 \times 10^{-3}]$