Estimating cross-population LD decay and its effect on trans-ethnic polygenic scores

Davide Piffer
pifferdavide@gmail.com

Abstract

Differing patterns of linkage disequilibrium (LD) across populations (LD decay) potentially reduce the trans-ethnic validity of GWAS results. In this study, I compute a measure of LD decay to test the hypothesis that it mediates the difference in polygenic scores computed using GWAS summary statistics between Europeans and Africans.

There was a significant PGS difference between CEU and YRI (50.312% - 47.995%, $p=0.001$). In total, there were 1,680,781 SNPs in LD with 2,596 GWAS SNP within a 500 Kb window.

There was no correlation between a measure of LD decay and population differences in polygenic scores of educational attainment.

This study provides no evidence that LD decay affects the trans-ethnic predictive validity of polygenic scores.

Introduction

The trans-ethnic validity of polygenic scores (PGS) is compromised by LD decay. This is the decay in linkage disequilibrium with time, meaning that the older the causal polymorphism, the lower the level of linkage disequilibrium due to recombination events that occur with constant probability in every generation. As a consequence, linkage patterns can be different in different populations, especially those that separated a long time ago and that underwent population bottlenecks after separation (Piffer, 2019).

The effect of LD decay on comparison of risk alleles between populations is still unclear. Since most GWAS hits are actually tag SNPs, decay in LD implies that the causal SNPs will be less efficiently flagged by the tag SNPs until the tag SNPs will resemble a sample of random SNPs. With less significant associations, it is not only more likely that the distance between the GWAS hit and the causal polymorphism is larger and linkage is weaker in the European-origin populations that are represented in the GWASs, but it is also more likely that the linkage phase is different in different races.

Simulations found that the effect of LD decay on true causal variants was null to negligible (Zanetti et al., 2016).

LD decay is estimated using several methods (Vos et al, 2017), mostly based on the correlation between LD measure and genetic distance between markers.

In this study, I calculate LD decay score for GWAS alleles using a novel method, and I test the hypothesis that it mediates polygenic score differences in educational attainment between Europeans and Africans.

Methods

The SNPs that reached genome-wide significance ($P \leq 5 \times 10^{-8}$) in the largest GWAS of educational attainment (Lee et al., 2018) were selected. Those with minor allele frequency (MAF) < 0.01 in the 1000 Genomes database among Europeans (CEU) and Africans (YRI) were removed. There were 2,596 SNPs satisfying these criteria.

To explore the different LD patterns across populations, the SNPs were uploaded to LDlink, and all variants ± 500 Kb of the query variant with a pairwise R^2 value greater than 0.01 were downloaded using CEU and YRI as reference populations. There were 4,696,863 and 5,739,447 variants for CEU and YRI, respectively.

The two files were merged and 1,680,781 overlapping SNPs were retained.

For each query variant, the correlation between the pairwise R^2 values for CEU and YRI was computed. The correlation coefficient was used as an index of differential LD decay across Europeans and Africans relative to the query variant. That is, the higher the correlation between the CEU and YRI R^2 values, the lower the amount of trans-ethnic LD decay.

For each SNP, the PGS for CEU and YRI was computed.

Results

The average correlation coefficient between the CEU and YRI R^2 values across the 2596 SNPs was $r=0.608$. This shows the presence of a moderate amount of LD decay.

There was a significant PGS difference between CEU and YRI. Welch Two Sample t-test: means: 50.312% - 47.995% respectively, $t= 3.134$, 95% C.I 0.008 - 0.037, $p= 0.001$.

There was no correlation between LD decay and the CEU-YRI PGS difference ($r= 0.015$, $p= 0.451$).

Figure 1. Association between LD decay and European-African difference in polygenic score.

Discussion

The lack of a correlation between LD decay and population differences in polygenic scores suggests that low p value GWAS SNPs have a causal effect on the phenotype or they are tag SNPs in close LD with the causal variants.

References

Piffer, D. Evidence for Recent Polygenic Selection on Educational Attainment and Intelligence Inferred from Gwas Hits: A Replication of Previous Findings Using Recent Data. *Psych* 2019, 1, 55-75. <https://doi.org/10.3390/psych1010005>

Vos, P.G., Paulo, M.J., Voorrips, R.E. et al. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor Appl Genet* 130, 123–135 (2017). <https://doi.org/10.1007/s00122-016-2798-8>

Zanetti, D.; Weale, M.E. True causal effect size heterogeneity is not required to explain trans-ethnic differences in GWAS signals. *bioRxiv* 2016, 085092, doi:10.1101/085092.