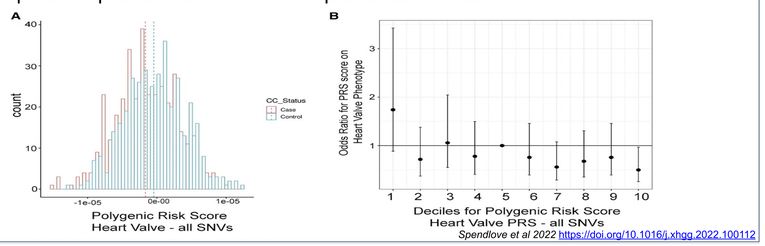
Identifying case versus control differences in polygenic risk scores for endo-phenotypes associated with congenital heart disease

DESIREE CERVANTES[1], Sarah Spendlove[2,3], Maria Palafox[3], Valerie Arboleda[3,4,5]
1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
2 Bioinformatics Interdepartmental Program, UCLA
3 Department of Human Genetics, David Geffen School of Medicine, UCLA
4 Pathology and Laboratory Medicine, University of California Los Angeles
5 Department of Computational Medicine, University of California Los Angeles

Background

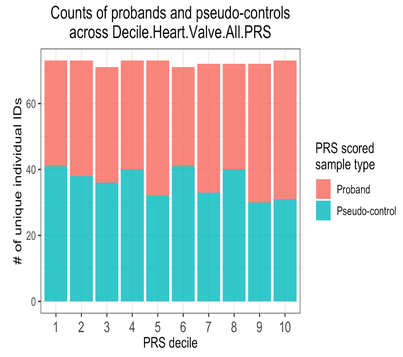
This investigation explores the distribution of polygenic risk scores (PRS) for congenital heart disease (CHD) across six different endo-phenotypes. Previous work showed a significant relationship between case/pseudo-control status and polygenic risk scores (PRS) but the directionality of the effect was opposite from what was expected, with individuals having higher PRS scores having a lower risk of CHD. PRS scores only consider common variants, so perhaps rare variants may be behind this unexpected result. The cases analyzed are probands, which are the offspring of two parents per family, and pseudo controls that are made with one half of the un-transmitted genetic material from parent 1 and the other from parent 2. The purpose of making a pseudo-control instead of using the parents individually was to utilize individuals with the least possible overlap in common allele variants to compare the pseudo-controls and the probands with each other.

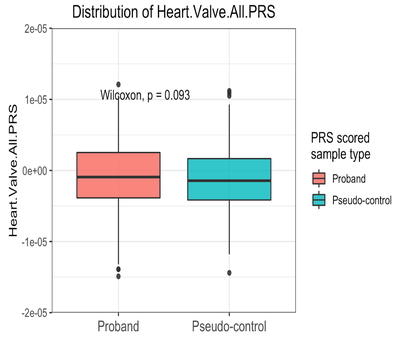


Spendlove et al 2022 https://doi.org/10.1016/j.xhgg.2022.100112

Introduction

To expand upon this study, the deciles of each endo-phenotype PRS were investigated to characterize finer resolution differences between probands and pseudo-controls. Although the previous paper showed that the heart valve PRS is significantly associated with the phenotype and demonstrates that the tenth decile is significantly different, a Wilcoxon test was conducted to determine the mean difference between the probands and pseudo-controls amongst all deciles and found that there is no significant difference between the two.





Computational Methods

Box plots of individual deciles were made to see finer resolution of differences between the first and tenth decile specifically. Once this was determined it was suspected that these interesting but miniscule differences could account for de novo variants, so singletons were used to investigate this hypothesis. Singletons were identified using VCFtools in order to start investigating the contribution of rare variants to the deciles. Singletons by definition are variants that are found in one individual in the entire dataset. Since we have the probands parents' genetic data, singletons are a good proxy for quantifying de novo variants, which are super-rare variants that are individual to a person and not inherited. The number of singletons for individuals in the top and bottom deciles for all heart valve SNVs PRS and all severity heart sounds SNVs PRS was calculated.

PRS Data Analysis

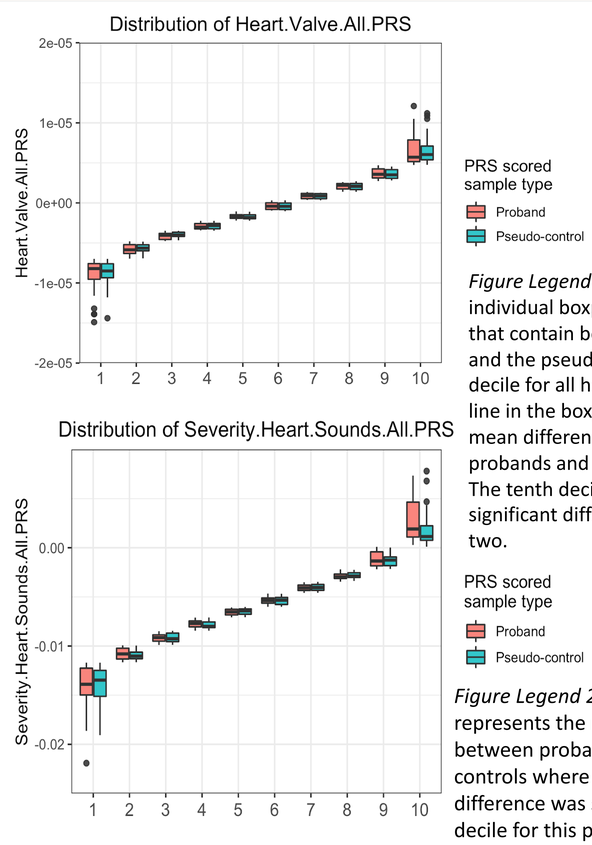


Figure Legend 1: Depicted are the individual boxplots for each decile that contain both the probands and the pseudo-controls in each decile for all heart valve PRS. Each line in the boxplot represents the mean difference between probands and pseudo-controls. The tenth decile demonstrates a significant difference between the two.

Figure Legend 2: This figure represents the mean difference between probands and pseudo-controls where an even greater difference was seen in the tenth decile for this phenotype compared to all heart valves.

Singleton Analysis for All Heart Valves PRS

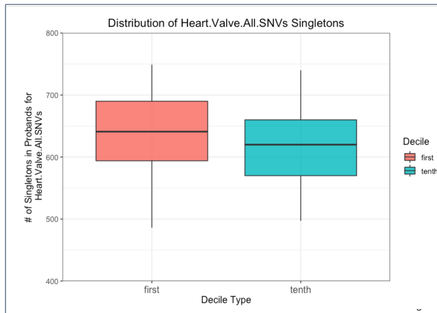
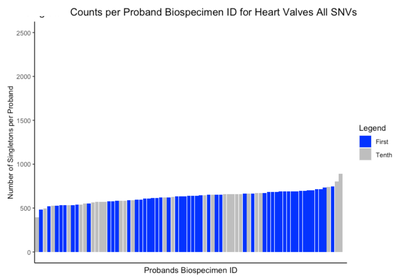
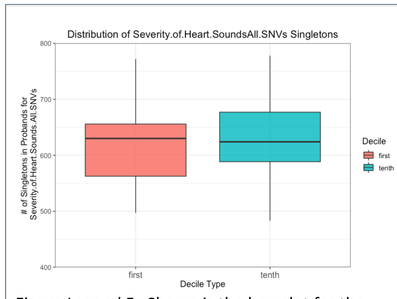


Figure Legend 3: This figure shows the number of singletons across all probands in the first and tenth decile. A Mann-Whitney U test was conducted and the p-value for the mean difference between the deciles was 0.1893 therefore showing no significant difference for all heart valves, this p-value was still more significant that the p-value calculated for the heart sounds PRS (below). Overall, the first decile has more singletons.

Figure Legend 4: Figure shows a bar plot with the count of singletons for each individual biospecimen that corresponds to the probands for the first and tenth deciles. There is a slight (albeit insignificant) of the first decile and this illustrates the distribution of specific probands in their respective deciles.



Singleton Analysis for Severity All Heart Sounds PRS



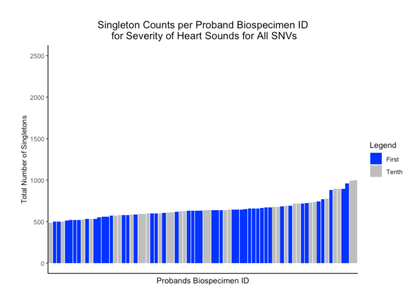


Figure Legend 5: Shown is the box plot for the number of singletons of probands in the heart sounds phenotype shows more in the tenth decile than the first, showing opposite effects from the heart valve phenotype. The p-value for the mean difference between the deciles was 0.453 supporting the idea that the overall difference is insignificant for this phenotype and could be due to random effects.

Figure Legend 6: Shown is the distribution of proband singleton counts in the heart sounds phenotype where again differences between the first and tenth deciles do not seem to be significant.

Conclusion

In brief, after closely looking at singleton analyses, heart valves for all SNVs PRS was found to have a nominally significant p-value. Thus, it was revealed that more rare variants was seen in the first decile over the tenth, but generally there is no significant difference between the two when examining probands only. To continue this investigation, more types of rare variants and even common variants can be determined in each decile to yield a more accurate understanding of the affected proband's genetic architecture. Genetic architecture underlying the endophenotypes themselves is rather more complex than expected. Thus, different PRS's might be better for different phenotypes.