Combining polygenic risk scores with gene-based burden scores

E. Hassanin[1,2], R. Aldisi[1], P. May[2,3], P. Krawitz[1], DR Bobbili[2], C. Maj[1]

[1] Institute of Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany
[2] MeGeno S.A, Esch-sur-Alzette, Luxembourg
[3] Bioinformatics Core, Luxembourg Centre for Systems Biomedicine (LCSB), Belvaux, Luxembourg

Introduction

1. In complex traits "common diseases", how a large number of genes associated variants contributes to influence development of the disease over a life course?.
2. Polygenic risk score (PRS) uses weighted sum of trait-associated alleles (**common variant**) to calculate an overall risk score of getting a particular common disease.
3. Gene-level scoring is based on the combined effects of multiple variants weighted according to functional annotations and variant frequency and deleteriousness[1] (**rare variants**).

Objectives

In this study, we explores whether the integration of PRS and Gene-level scoring can improve risk prediction or no.

Methods

Data Source

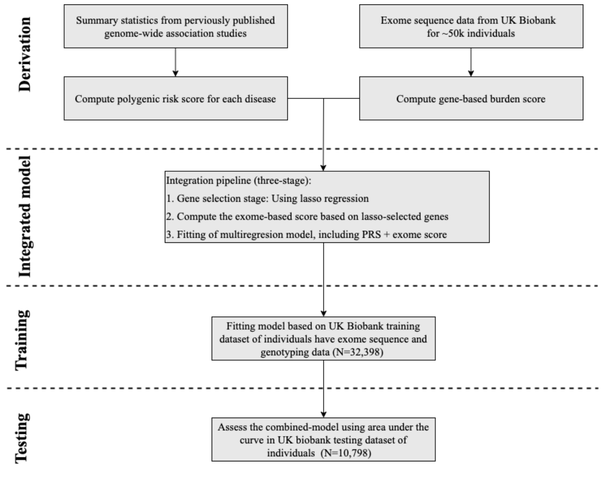
This study was performed using genotypic and phenotypic data from the Ukbiobank (UKB).

Study participants

We studied three common diseases: Breast Cancer, CAD, and type 2 diabetes. Diagnosis of the three disease was based self-reported, ICD-9, ICD-10, and OPCS-4 codes.[2]

Statistical model

we developed a combined model to integrated PRS, and gene-based score.



Results

For each disease, we calculated PRS based on previously reported GWAS, and gene-based scores based on trait-associated genes.

Table 1: PRS derivation and gene-based scores for three common, complex diseases.

Disease	Discovery GWAS (n)	Discovery genes (n)	Prevalence in training (cases/controls)	Prevalence in testing (cases/controls)
Breast Cancer		920	779/16751	259/5583
CAD		74	861/31537	315/10483
Type 2 diabetes		53	717/31681	238/10560

The derived genetic risk models were assessed on the basis of their ability to classify individuals diagnosed with breast cancer, CAD, and diabetes in the UKB testing dataset.

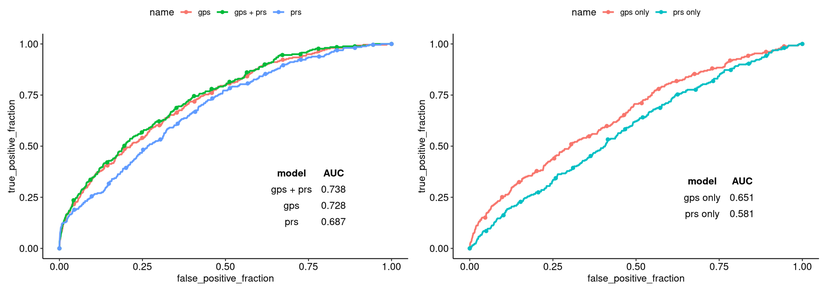


Figure 1: Breast cancer models. On left hand side, AUC of PRS, gene-based score (gps), and combined model was determined using a logistic regression model adjusted for age, sex, and the first four principal components of ancestry. On right hand side, AUC of PRS, gene-based score (gps) adjusted only for the first four principal components of ancestry.

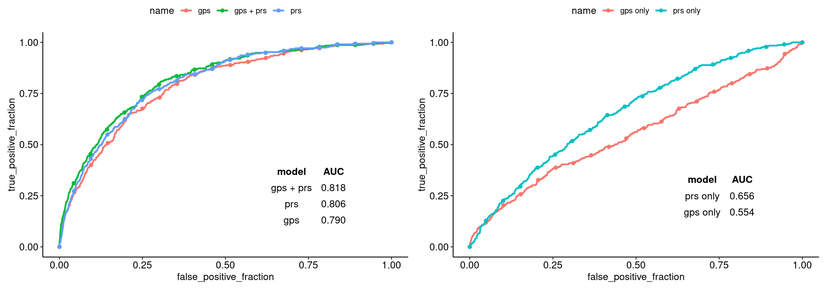


Figure 2: Coronary artery disease (CAD) models

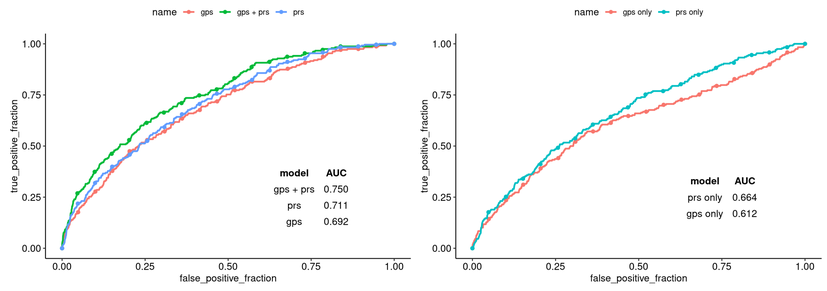


Figure 3: Type 2 diabetes models

Conclusion

1.The genetic contributions to human traits can be associated with both common variant and rare variants

2. A better evaluation of the genetic susceptibility can be obtained with a combined evaluation of PRS and gene-based burden scores

Next Steps

This study was performed on ~50,000 samples that have both genotyping and exome sequence data in the UKB. Next step is to reproduce the same work on recently releases 200,000 exome-sequence sample in the UKB

References

1. Ashton, J., Pengelly, R., Beattie, R. & MacArthur, B. GenePy – a score for estimating gene pathogenicity in individuals using next-generation sequencing data. *bioRxiv* 336701 (2018). doi:10.1101/336701

2. V.Khera, M.Chaffin, G. Aragam, E. Haas, C.Roselli, S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations Amit. *Physiology & behavior* **176,** 139–148 (2019).