Partitioned PGS for type 2 diabetes (T2D) and high blood pressure (BP) shared genetics and comorbidity

Background

This study explores the comorbidity between type 2 diabetes (T2D) and high blood pressure (BP) through the genetic relationships between single nucleotide variants (SNVs) associated with T2D, BP, or both conditions and respective condition/comorbidity risks. The associated SNVs were grouped to build partitioned polygenic scores (PGSs) according to the mechanistic subdivision of their effects on phenotypes into five distinct clusters based on their inferred pathogenetic processes, namely: **Inverse T2D-BP risk**, **Metabolic Syndrome**, **Higher adiposity**, **Vascular dysfunction**, and **Reduced beta-cell function**.

To validate these clusters, we constructed partitioned PGSs in the UK Biobank using the SNVs from each cluster.

GitHub repository of the code: https://github.com/VP-biostat/T2D-BP-Manuscript/

Step 1: Building partitioned PGSs

Objective

The aim of this analysis is to replicate the construction of the five PGSs on another dataset/cohort. These PGS are exploratory in nature and are not fine-tuned for predicting any single phenotype but rather to assess associations with mechanistically derived groups of SNVs related to pathophysiological processes within T2D-high BP comorbidity.

Base data

We defined five **unweighted PGSs (Beta = 1), assuming an additive model**. The effect allele (EA) is aligned to T2D risk allele. The list can be found in ***Supplementary Table 7***.

Target data

The target dataset used was the UK Biobank including all genotyped individuals. The variants should be imputed with an appropriate quality (**info ≥0.4 filter from imputation should be applied**).

PGS construction

We recommend using directly **PLINK v.1.9** "--score" function to build the PGS without further fine-tuning (e.g., PRSice, LDpred…). Below is an example command using the unweighted_t2d_beta_cluster_5.txt base data:

```
plink --dosage [your_file.dosage] \
    --fam [your_file.fam] \

    --score [unweighted_t2d_beta_cluster_5.txt] 2 3 5 'header' sum double-dosage include-cnt \

    --out [your_output]
```

This process should result in **five partitioned PGS ready for further association analyses**.

Step 2: Association analysis between partitioned PGS and T2D-BP comorbidity

Objective

The goal is to assess the association between the partitioned PGSs and the prevalence of T2D-high BP comorbidity, following the structure of the table below:

Table 1. T2D-high BP comorbidity in UK biobank for partitioned PGSs

Being top 10% of the partitioned PGS based on	**Frequency of T2D-BP comorbidity**	**Relative risk of T2D-BP comorbidity**
Inverse T2D-BP risk cluster	5.73%	1.04
Vascular dysfunction cluster	6.26%	1.14
Higher adiposity cluster	7.48%	1.36
Metabolic Syndrome cluster	7.88%	1.44
Reduced beta-cell function cluster	8.51%	1.55
Metabolic Syndrome & ***Reduced beta-cell function*** clusters	11.66%	2.13

Methodology

Percentile Calculation:

For each partitioned PGS and for each individual, calculate percentiles. Individuals in the **top 10% (90th–100th percentile)** are considered part of the respective PGS cluster.

Assignment Rules:

Assign individuals to the cluster where they rank highest if they belong to multiple clusters.

The final row from *Table 1* represents individuals in the top 10% of both the ***Metabolic Syndrome*** and ***Reduced Beta-cell Function*** PGS (calculated afterwards the first assignment).

Outcome Definition:

T2D-BP comorbidity is defined as having both T2D and primary hypertension, based on ICD-10 codes **E11*** (T2D) and **I10*** (primary hypertension).

Frequency calculation:

The frequency of T2D-BP comorbidity was defined by the proportion of cases divided by the total number of individuals in the UKB, $\mathbf{Freq} = \frac{\mathrm{N_{cases}}}{\mathrm{N_{total}}}$.

Relative risk calculation:

The relative risk of T2D-BP comorbidity is calculated as the ratio of the proportion of cases within a cluster subgroup to the proportion of cases in the overall population $\boldsymbol{RR} = \frac{\boldsymbol{Prev_{cases\ in\ cluster}}}{\boldsymbol{Prev_{cases\ overall}}}$.

Step 3: Phenome-Wide Association Study (PheWAS) with Partitioned PGS and Complications

Objective

The aim is to use *comorbidPGS* to collect the associations between partitioned PGSs and additional complications.

Analysis pipeline

ICD-10 outcome codes:

Association is done between partitioned PGSs with available ICD-10 codes in YOUR_STUDY. Subcategories of interest include *"Endocrine, nutritional, or metabolic", “Mental and behavioral disorders", "Nervous system", “Eyes, ears, nose, and throat", "Circulatory system", "Respiratory system", "Digestive system", “Skin", "Musculoskeletal system", "Genitourinary system"*. Therefore, the full set of ICD-10 codes is **filtered to start with specific letters**:

```
C("E", "F", "G", "H", "I", "J", "K", "L", "M", "N")
```

Association:

Association with complications were performed using ***comorbidPGS* on R version 4.2 or higher**. *comorbidPGS* uses binary logistic regressions and format the results.

The package is handling missing data and scaling the PGS automatically, it should also detect binary phenotypes if the ICD-10 values are either TRUE/FALSE or 1/0.

The association analysis can be computationally costly, and we recommend to parallelise (in-built function). Below an example using parallelisation and 6 cores.

```
install.packages(“comorbidPGS”)

library(comorbidPGS)

res = multiassoc(df, association_table, scale = T, covar_col, parallel = T, num_cores = 6)

#df is the dataset,

#association_table is a matrix with PGS colnames on first column and phenotype colnames on the second, perform one association analysis per row of association_table
```

Covariates:

Covariates for the logistic regression include **age, sex (genotypic), genetic array, and the first six principal components**:

```
covar_col = c("age", "sex", "array", "PC1", "PC2", "PC3", "PC4", "PC5", "PC6")
```

Multiple testing correction:

We use Bonferroni multiple testing correction, with the formula $\boldsymbol{P} = \frac{\mathbf{0.05}}{\boldsymbol{n}}$ with n the number of logistic regressions performed.

Data.frame results:

comorbidPGS multiassoc() function should return a `data.frame` with the following properties:

Table 2. Output `data.frame` column specifications from the multiassoc() function

Column name	Information
PGS	Name of the PGS column
Phenotype	Name of the Phenotype column
Phenotype_type	The type of the trait detected
Stat_method	Name of the regression used (depends on the Phentoype_type)
Covar	Name(s) of the covariate(s)
N_cases	If Phenotype is a cases/controls, the number of cases
N_controls	If Phenotype is a cases/controls, the number of controls
N	Number of individuals in the association
Effect	Beta for linear regression, Odds Ratio (OR) for logistic regressions
SE	Standard Error (only for linear regression)
lower_CI	95% lower Confidence Interval
upper_CI	95% upper Confidence Interval
P_value	95% upper Confidence Interval

Contact

Prof. Inga Prokopenko i.prokopenko@surrey.ac.uk

Dr. Vincent Pascat vincent.pascat20@imperial.ac.uk