



UKB PRS Release Version 2 - Readme and Changelog

May 2024

1 Overview

The UK Biobank PRS Release is a set of polygenic risk scores (Category 300) calculated for 28 diseases and 25 quantitative traits, either on all individuals in UK Biobank (the 'Standard PRS set') or on a subset of individuals in UK Biobank (the 'Enhanced PRS set') (Resource 5202).

Further details can be found in Thompson et al (2022) "UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits", submitted.

Version 1 of the UK Biobank PRS Release was released in May 2022.

Version 2 of the UK Biobank PRS Release is scheduled for release in May 2024, and incorporates feedback from users, and new data.

Version 2 of PRS scores is available on Showcase and is updated as of May 2024. Version 1 remains accessible on the Returns Catalogue for researchers.

For any questions or comments regarding these files, please contact Genomics plc at research@genomicsplc.com

2 Description of UKB Showcase (Resource 5202)

2.1 PRS values

Two types of polygenic risk score (PRS) values are provided. The 'Standard PRS set' (Category 301) is calculated on all individuals in UK Biobank, and is built from external GWAS training data only. The 'Enhanced PRS set' (Category 302) is calculated on a subset of individuals in UK Biobank (not used for training), and is built from a combination of external and within-UKB GWAS training data. Standard PRS values are provided for 28 diseases and 11 quantitative traits. Enhanced PRS values are provided for 28 diseases and 25 quantitative traits (the additional 14 traits are currently lacking well-powered external GWAS training sets, and so Standard PRS values cannot be calculated).

Version 2 of PRS scores are available on Showcase.

2.2 Other variables

Field 26200 "In UK Biobank PRS Release Testing subgroup" - 0/1 indicator of whether an individual is part of the Testing Subgroup defined in Thompson et al.

Field 26201 "Genetic principal components" - a set of 4 variables (PC1-PC4) describing an individual's position in 4D genetic ancestry space, and used to define genetic ancestry labels as described in Thompson et al.

3 Description of data bundle in Returns Catalogue

Values as described for the UKB Showcase are stored in the 'PRS_values' folder, with separate files for the Standard and the Enhanced PRS set. These correspond to Version 1 PRS scores.

4 Version 2 Release changelog

See also https://github.com/Genomicsplc/ukb-pret/blob/main/UK_Biobank_PRS_Release_Dataset_Notes.md for additional release notes (including reported issues and clarifications).

4.1 PRS values

4.1.1 Subgroup on whom Standard PRS scores are calculated.

Following new UKB withdrawals, the number of UKB individuals on whom Standard PRS scores are calculated has reduced.

4.1.2 Subgroup on whom Enhanced PRS scores are calculated.

Following feedback from users, the subgroup of UK Biobank for whom Enhanced PRS scores are calculated has been expanded to include all individuals in UK Biobank who are not members of White British Unrelated subgroup (the latter was used to generate GWAS training data for the Enhanced PRS set - see Thompson et al.). This includes 44,284 new participants for whom these traits were calculated.

4.1.3 Updated Standard and Enhanced PRSs - new data.

Following the ingestion of new external summary statistic GWAS data, updated or new Standard and Enhanced PRS scores have now been created and the Standard PRSs for resting heart rate, total triglycerides, and total cholesterol are new (Version 1 did not include Standard PRSs for these traits).