UK Biobank MTP Return of Results README

This Return-of-Results package relates to the following paper:
Thompson et al. (2022). UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. Submitted. The results data were generated under UK Biobank project number 9659 (under its extension to therapeutic applications).

Additional release notes about the PRS release (including reported issues and clarifications) can be found in: https://github.com/Genomicsplc/ukb-pret/blob/main/UK_Biobank_PRS_Release_Dataset_Notes.md

UK BIOBANK PRS RELEASE BRIEF DESCRIPTION

Our paper describes the UK Biobank PRS Release: a set of polygenic scores for 53 disease and quantitative traits, together with associated phenotypic and genetic information (see below for list of traits).

A separate tool (UKB-PRET) is provided to facilitate evaluation in a testing subset of the UK Biobank. See https://github.com/Genomicsplc/ukb-pret

PRS scores were generated using a Bayesian approach applied to meta-analysed (and, when possible, ancestry specific) summary statistics GWAS data, obtained either entirely from external GWAS data (the Standard PRS set) or from a combination of external and internal UK Biobank data (the Enhanced PRS set). A subsequent principal component(PC)-based ancestry centering step (Khera et al 2019, Circulation 139, 1593–1602) was applied to approximately centre the score distributions on zero across all ancestries, although it should be noted that residual deviations of mean(PRS) from zero were still observed for some traits in some ancestries. Score distributions were also standardised to have approximately unit variance within ancestry groups, as determined by a geometric inference in PC space. The Standard PRS set was calculated for all UK Biobank individuals; the Enhanced PRS set was calculated for the UK Biobank PRS Release Testing subgroup only. For further details, please refer to the paper.

UK BIOBANK PRS RELEASE DATA FILES

The Release is packaged as a series of data files, with fields as described below.

DATA

In UK Biobank PRS Release Testing subgroup – field 26200
PRS genetic principal components – field 26201
PRS fields:
One for each trait and PRS type (Category 301 - Standard PRS and Category 302 - Enhanced PRS)

COMMON DATA CONTENT DESCRIPTION

In UK Biobank PRS Release Testing subgroup: indicator of participants in the UK Biobank PRS Release Testing subgroup.

PRS genetic principal components: the first four PC axes obtained from the common genotype data in the 1,000 Genomes reference dataset.

PRS FILES (Standard Set) CONTENT DESCRIPTION
---
prs: polygenic score for the trait for all UK Biobank participants, as described in Thompson et al. (2022).

PRS FILES (Enhanced Set) CONTENT DESCRIPTION
---
prs: polygenic score for the trait for every participant in the UK Biobank PRS Release Testing subgroup, as described in Thompson et al. (2022).

PHENOTYPE CODES

The following is a list of traits and their phenotype codes (as used in file naming).

DISEASE TRAITS
Age-related macular degeneration AMD
Alzheimer's disease AD
Asthma AST
Atrial fibrillation AF
Bipolar disorder BD
Bowel cancer CRC
Breast cancer BC
Cardiovascular disease CVD
Coeliac disease CED
Coronary artery disease CAD
Crohn's disease CD
Epithelial ovarian cancer EOC
Hypertension HT
Ischaemic stroke ISS
Melanoma MEL
Multiple sclerosis MS
Osteoporosis OP
Prostate cancer PC
Parkinson's disease PD
Primary open angle glaucoma POAG
Psoriasis PSO
Rheumatoid arthritis RA
Schizophrenia SCZ
Systemic lupus erythematosus SLE
Type 1 diabetes T1D
Type 2 diabetes T2D
Ulcerative colitis UC
Venous thromboembolic disease VTE

QUANTITATIVE TRAITS
Age at menopause AAM
Apolipoprotein A1 APOEA
Apolipoprotein B APOEB
Body mass index BMI
Calcium CAL
Docosahexaenoic acid DOA
Estimated bone mineral density T-score EBMDT
Estimated glomerular filtration rate (creatinine based) EGCR
Estimated glomerular filtration rate (cystatin based) EGCY

Glycated haemoglobin HBA1C_DF
High density lipoprotein cholesterol HDL
Height HEIGHT
Intraocular pressure IOP
Low density lipoprotein cholesterol LDL_SF
Omega-6 fatty acids OSFA
Omega-3 fatty acids OTFA
Phosphatidylcholines PDCL
Phosphoglycerides PHG
Polyunsaturated fatty acids PFA
Resting heart rate RHR
Remnant cholesterol (Non-HDL, Non-LDL cholesterol) RMNC
Sphingomyelins SGM
Total cholesterol TCH
Total fatty acids TFA
Total triglycerides TTG

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