Instructions to calculate polygenic risk scores for COPD

1. Download files here:
   https://storage.googleapis.com/polygenicscores_copd/PolygenicScores_COPD.tar.gz

2. Unzip file, and there is a folder named "PolygenicScores_COPD". In this folder, you will see the following files:

Filename	Description
copd_riskscore.public.R	Main R script
pheno_manifest.public.txt	Phenotype manifest file, defining names of phenotype variables and covariates. This should match data fields in phenotype file
pheno.file.header.txt	This is the header from our phenotype file provided as an example. Actual data cannot be provided due to IRB restrictions, but phenotype data can use this format.
score.weights.RData	This file has lassosum.pipeline objects that contain all weights for PRSs, and can be used to calculate scores in the R script provided
prs_fev1_genkols_weights.csv	Variant names and weights (labelled fev1.weights) for direct download
prs_ratio_genokols_weights.csv	Variant names and weights (labelled ratio.weights) for direct download

   Files you will need to provide:
   -pheno.file – your phenotype file.
   -plink binary files (.bim, .bed, .fam) – this should be a file containing genetic data for your sample across all chromosomes (1-22).

3. Open copd_riskscore.public.R. Fill in your path and file names in the "START CONFIG" section.

4. Run R script. We would recommend submitting to a cluster or cloud, if possible. This script will output a file with your phenotype data and merged PRSs (for FEV_1, FEV_1/FVC, and a combined PRS -- the latter was the one used in the manuscript). The PRSs will be oriented to COPD risk (caCoStatus variable in the example phenotype file); this step can be skipped if desired.

   The number of variants in your reference panel will be recorded as well. Make sure there is adequate genotyping to ensure high quality scores.