Steps taken to make PGS for new GWAS

# Gwas preprocessing

GWAS is about birth length, calling it “Birth\_length”. Also contains year 2015.

Navigate to PRS folder -> scripts -> SumStats

Open Sumstats.R

Run source

Get manifest

No birthlengh included

Checking GWAS …. -> ChildhoodBirthlength\_2015\_18

Opened the GWAS:

-CHR BP RSID A1 A2 BETA SE P N

-N changes per SNP.

Decide to ADD new GWAS

GWAS n is requested, setting it to highest I can observe in GWAS file (in the order of 28322)

f\_addGWAStoManifest(short=c("Birth\_length"),

n=c(28322),

filename=c("D:/DATA\_STORAGE/GWAS/Childhood\_birth\_length\_2015/ChildhoodBirthlength\_2015\_18.txt"),

year=c("2015"),

trait=c("Childhood Birth length"),

DOI=c("?"),

genomeBuild = c("?"),

traitType = c("CONT"),

rawSNPs = c("?"),

finalModelSNPs = c("?"),

modelRunningTime = c("?"),

usedRefSet = c("?"),

processed=c(0),

FORCE = FALSE)

I get a message to confirm, yes

>> Press [y] if information correct, then press [ENTER] <<

y

The gwas is now in the manifest, but didn’t go through correct formatting

I remember the short (name) “Birth\_length”, and continue in sumstats.R

Because this is standardized, I add the Birth\_length name to the listofStandardizedGWASes (@RRR which should be generalized)

This way, it automatically processes the GWAS data into the format that we want.

I check the manifest again (Ref\_gwas\_manifest)

And Birth\_length is on the 7th row, therefore I set ‘i’ to 7. @RRR this needs to be made into function! NOW!

I run the internal part of the loop of sumstats.R, check if loading was successful by dim() message. (2201971 snps loaded.)

I run the “Future input method.” As this is the correct method.

I run the remainder of the script.

I check the manifest to confirm. There is updated process flag (processed =1 now) and rawsnps is set to 2201971.

At this point the gwas summary is set up properly.

Lets calculate the PGS model

# Calculate PGS

I run the dedicated function to check all traits “f\_getTraits()”. I want “Birth\_length” to be calculated.

Therefore, I run “f\_calcPGS\_LDAK(Trait="Birth\_length") “

Unable to run due to limited overlap of SNPs!! Only 1000 overlap?