Genome-wide Association Studies

Practical 2: Do the GWAS

Data Use Agreement

- To maximize your learning experience, we will be working with genuine human genetic data

- Access to this data requires agreement to the following in to comply with human genetic data ethics regulations

- Please email pctgadmin@imb.uq.edu.au to confirm that you agree with the following:

    - “I agree that access to data is provided for educational purposes only and that I will not make any copy of the data outside the provided computing accounts.”

Note – Creating Plots in R

- Some options...
- #1: Transfer **results files** and create plot on the computer in front of you
- #2: Save plots in R and transfer them using SFTP client
-
```
png()
plot(….)
dev.off()
```

- #3: Enable X11 forwarding (using “ssh -X” or clicking the relevant option in PuTTY) - may not work….

Data

- Data for this practical is found in the directory:
  - `/data/module1/gwas/part2/`
- Three files:
  - `gwas.bed` → binary file containing all genotypes
    `gwas.bim` → information about SNP markers
    `gwas.fam` → information about individuals
- Heaps of phenotypes → Choose your own adventure!
- Fasting glucose, fasting insulin, ferritin, height, neuroticism, sleep duration, smoking (pack years), systolic blood pressure, waist-to-hip ratio
- Covariates: age, sex, PC 1-5 (`covariates.cov`)

GWAS

- Command: `--assoc`
- `plink --bfile /data/module1/gwas/part2/gwas --assoc --pheno <file>`

```
[allan@analysis1 ~]$ plink --bfile /data/module1/gwas/part2/gwas --assoc --pheno
 /data/module1/gwas/part2/Fasting_Insulin_QC.phen
PLINK v1.90b6.26 64-bit (2 Apr 2022)           www.cog-genomics.org/plink/1.9/
(C) 2005-2022 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to plink.log.
Options in effect:
  --assoc
  --bfile /data/module1/gwas/part2/gwas
  --pheno /data/module1/gwas/part2/Fasting_Insulin_QC.phen

64141 MB RAM detected; reserving 32070 MB for main workspace.
277719 variants loaded from .bim file.
11780 people (5346 males, 6434 females) loaded from .fam.
11770 phenotype values present after --pheno.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 11780 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.995966.
277719 variants and 11780 people pass filters and QC.
Phenotype data is quantitative.
Writing QT --assoc report to plink.qassoc ... done.
```

Plotting Results

- Using R
- ```
  d = read.table(“plink.qassoc”, head=T)
  manhattan(d)
  qq(d)
  ```
- Do your plots look good?
- Any evidence for inflation? Calculate the genomic inflation factor:
- ```
  gif = qchisq(1-median(p),1)/qchisq(0.5,1)
  ```

Generate PCs

- **Takes a long time to run! Use the pregenerated PCs in the covariate file**
- Command: `--pca <n>` Calculate the first n PCs

Add covariates

- Command: `--linear --covar <file>`
- **SLOWER!**
- **A LOT SLOWER IF YOU INCLUDE PCS TOO!**

- Alternative: regress the phenotype against the covariates in R and create a new phenotype file with the residuals
- Results in some power loss

Binary Phenotype

- Command: `--assoc`
- Command: `--logistic --covar <file>`

Clumping

- How many “roughly independent” signals at your top hit?
- Command: `--clump <results file>`

- Options to control clumping: https://www.cog-genomics.org/plink/1.9/postproc#clump
- Example:
  - `--clump-p1 5e-8`
    `--clump-p2 5e-8`
    `--clump-r2 0.2`
    `--clump-kb 1000`