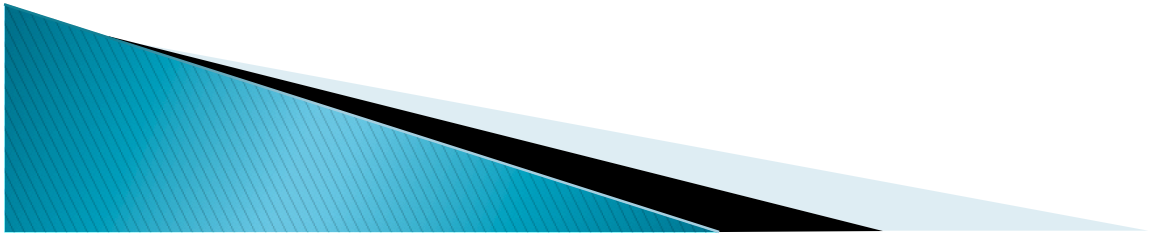
GWAS for quantitative traits:

using PLINK and R

PLINK software

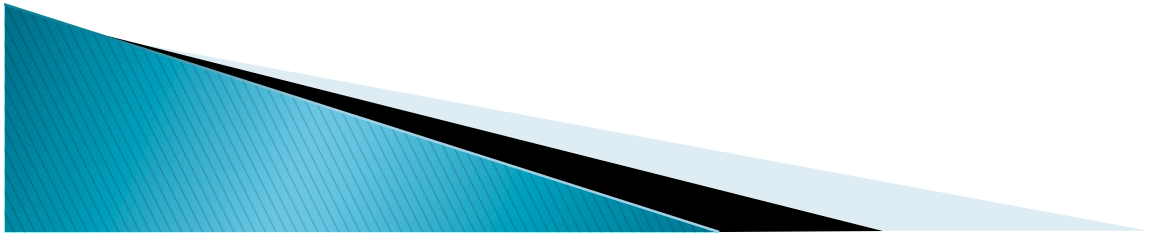
- Ultimate package for GWAS
  - Data management
  - Summary statistics
  - Population stratification
  - Association analysis
  - Meta analysis
  - IBD estimation
  - Gene based report
  - Simulation



Input Files

- Pedigree file `(.ped)`
- Map file `(.map)`

- Fam file `(.fam)`
- Bim file `(.bim)`
- Bed file `(.bed)`



Input Files

- Pedigree File – the first six columns are mandatory:
  - Family ID
  - Individual ID
  - Paternal ID
  - Maternal ID
  - Sex (1=male; 2=female; other=unknown)
  - Phenotype

Input Files

- MAP File – 4 columns:
  - chromosome (1-22, X, Y or 0 if unplaced)
  - rs# or snp identifier
  - Genetic distance (morgans)
  - Base-pair position (bp units)

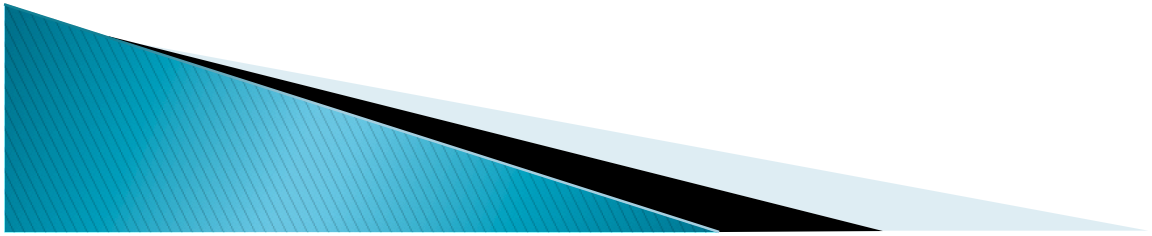
Input Files

- Binary files
  PLINK --file myfile --make-bed --out myfile
  generates

  ```
  Myfile.bed
  Myfile.bim
  Myfile.fam
  ```

Data management

- **Inclusion criteria**
  - --keep mylist.txt, --remove mylist.txt
  - --extract mysnps.txt, --exclude mysnps.txt
  - --chr 6, --from rs273744 --to rs89883
- **Other data management options**
  - --make-bed, --recode, -bmerge
- **Using phenotypes files**
  - --pheno, --all-pheno, --mpheno



QC

Summary statistics

- MAF (--freq)
- SNP missing rate (--missing)
- Individual missing rate (--missing)
- Hardy-Weinberg (--hardy)

Inclusion criteria

- MAF (--maf)
- SNP missing rate (--geno)
- Individual missing rate (--mind)
- Hardy-Weinberg (--hwe)

Association analysis

- Basic association test (--assoc)
- Stratified analysis (--within mycluster.dat)
- Covariates (--covar --mycov.dat)
- GxE interaction (--gxe --mycov.dat)

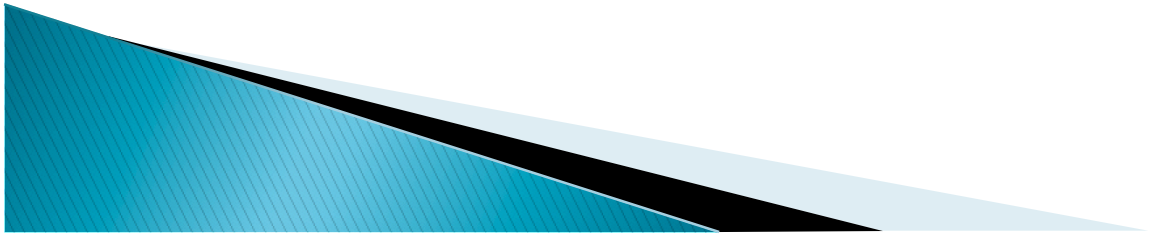
Final Exam Data – overview

- File Exam data is available in the Final Exam Dropbox folder
- Final Exam PLINK files are: Transferrin.bed, Transferrin.fam,Transferrin.bim
- GWAS for serum transferrin
- File with some R commands: Commands_Transferrin_Data.R

- HELP: http://pngu.mgh.harvard.edu/~purcell/plink/

Final Exam Data – File inspection

- Copy the files to a folder
- Use the R script to inspect files (not the .bed file)
- How many individuals are there? How many SNPs are there?



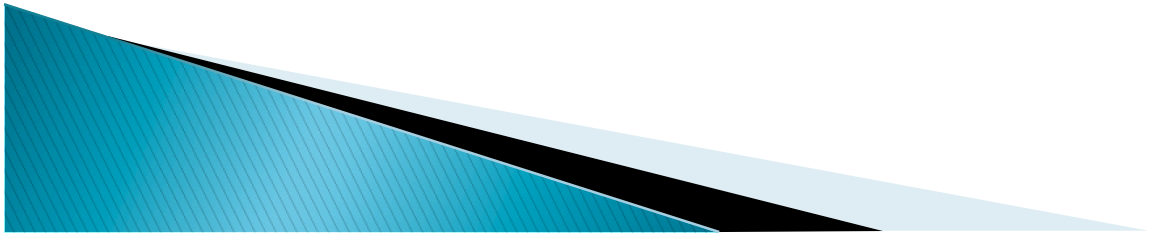
Final Exam Data – QC

- Can estimate allele frequency for the SNPs using plink
  - `./plink --bfile Transferrin --pheno Tr.pheno --freq --out Trans_freq`
- Can calculate SNP and individual missingness using plink with the following options
  - --missing --out Trans_missing
- What are p-values for HWE?
  - --hardy --out Trans_hardy

Final Exam Data: GWAS Analysis

- Run GWAS analysis for Transferrin data with PLINK.
- Command to use QC thresholds such as maf 0.05 / missing 0.01 / HWE 0.001 for the analysis with PLINK: the following commands

```
plink --bfile Transferrin --pheno Tr.pheno --maf
  0.05 --geno 0.01 --hwe 0.001 --assoc --out
  GWAS_T_add
```



Final Exam Data: Using a subset of SNPs

- Can do analyses on a subset of SNPs
- Create a file named "SNP_List.txt" that contains a list of SNPs of interest.
- PLINK command with "extract" to perform association testing on a subset of SNPs:

```
plink --bfile Transferrin --pheno Tr.pheno --extract
   SNP_List.txt  --assoc --out GWAS_T_add_Subset
```

- PLINK command to do r^2 LD calculations for all possible pairs of SNPs in the subset SNP file

```
plink --bfile Transferrin --extract SNP_List.txt --r2
   --out LD_T_Subset
```