

# 704 Assignment 1 - GWAS

*Peter Shen*

*Sys.date()*

## Background Questions

- 1) The diseases of interest in WTCCC is bipolar disorder (BD), coronary artery disease (CAD), Crohn's disease (CD), hypertension (HT), rheumatoid arthritis (RA), type 1 diabetes (T1D), and type 2 diabetes (T2D).

E + G -> P -> D

E: None examined G: 500K SNP Microarray Data P: Gene expression of the variant D: bipolar disorder (BD), coronary artery disease (CAD), Crohn's disease (CD), hypertension (HT), rheumatoid arthritis (RA), type 1 diabetes (T1D), and type 2 diabetes (T2D)

2)

- *BD*: Individuals have suffered one or more episodes of pathologically elevated mood. They were interviewed by psychiatrists and measured using a OPCRIT checklist and psychiatric diagnoses were assigned using Research Diagnostic Criteria.
- *CAD*: Hospital validated history of either myocardial infarction or coronary revascularization (coronary artery bypass surgery or percutaneous coronary angioplasty) before their 66th birthday.

- 3) 500k Affymetrix SNP Microarray Chip and then imputed against 2,193,483 HapMap SNPs not on the Affymetrix chip

Import sample data for chromosome 22.

```
# ct22 <- read.delim('sampledata//Affx_gt_58C_Chiamo_22.tped.gz', sep="\t", header=F)
# t1d22 <- read.delim('sampledata//Affx_gt_T1D_Chiamo_22.tped.gz', sep="\t", header=F)
# snps <- read.delim('sampledata/snps_22', header=F)
```

Background Questions:

- 4) \$ wc -l \*tped 500568 total SNPs were assayed for each individual