LDSC

# Environment

## Anaconda

Initialize conda:

source /home1/08005/cz5959/anaconda3/bin/activate

conda init

Activate conda and ldsc env:

conda activate /scratch1/08005/cz5959/ldsc/env

source activate ldsc

# Features

## LD Scores

* For European GWAS, no need to compute own LD scores and can instead use the pre-computed LD scores
  + This will create a eur\_w\_ld\_chr/ directory
  + Computed using 1000 Genomes European data

wget https://data.broadinstitute.org/alkesgroup/LDSCORE/eur\_w\_ld\_chr.tar.bz2

tar -jxvf eur\_w\_ld\_chr.tar.bz2

* Currently, bulik recommends using those same LD scores for both –w-ld-chr and –ref-ld-chr flags for non-partitioned ldsc

## Heritability and Genetic Correlation

<https://github.com/bulik/ldsc/wiki/Heritability-and-Genetic-Correlation>

#### Reformatting Summary Statistics

* Convert your results to .sumstats format using munge\_sumstats.py
* Requires following information: rsid; effect allele; non-effect allele; sample size; p-value; signed summary statistic (ex. beta)
* –merge-alleles makes sure that the alleles in your results match the alleles listed in the data used to estimate LD scores
* Check log file for errors with: grep ‘WARNING’ \*log



* Reformat Neale Lab to use munge\_sumstats.py with
  + Switch signs of betas is A1 != ALT; remove NA



#### Estimating Heritability and LDSC Intercept

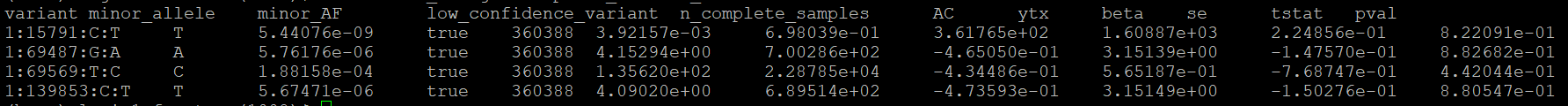
* LDSC – get heritability and intercept



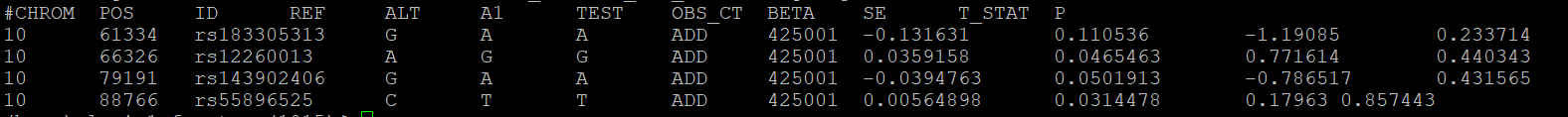
# Log

### 6/10/2021

* Set up environment by downloading anaconda for linux <https://github.com/bulik/ldsc#readme>
* Neale Lab Header



* + Neale doesn’t have rsid included; I can try to match with plink2 results
* Plink2 Header



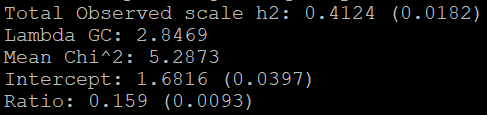
### 6/11/2021

* Download LD scores and alleles used to compute LD scores
* Create sumstats file format from results



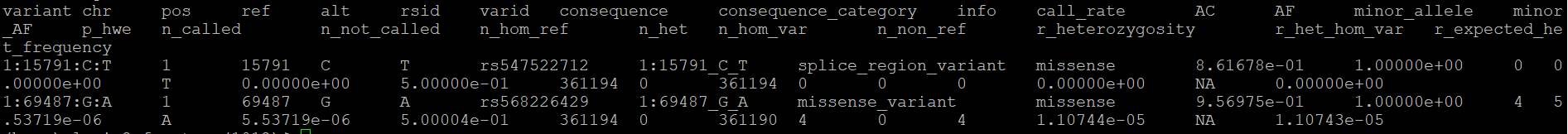
### 6/14/2021

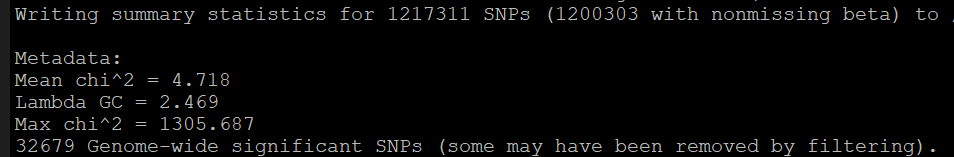
* Calculate heritability and ldsc intercept for height
* Error: IndexError while reading LD Scores
  + Forgot to add the “/” after eur\_w\_ld\_chr
* H2 and intercept for height
  + What is lamda and ratio
  + Lambda: percent variance in dependent variable (height) not explained by difference differences in levels of independent variable (genotype)
  + Ratio: Ratio is (intercept-1)/(mean(chi^2)-1), which measures the proportion of the inflation in the mean chi^2 that the LD Score regression intercept ascribes to causes other than polygenic heritability.



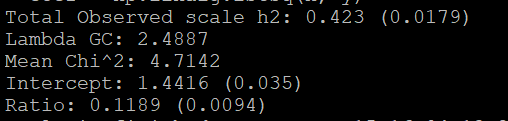
### 6/15/2021

* Format Neale lab to be suitable for munge\_py
  + Download variants file





* H2 and intercept for Height from Neale Results



### 6/16/2021

* GC bias: genomic region of a higher GC content tends to have more (or less) Illumina reads covering that region
* Intercept, mean chi square when tagging 0 casual SNPs
* GC bias
  + Recombination: get errors one sis as GC; other sis has AT; should both have GC
    - Heteroduplex gets repaired, repairs more GC because stronger bond
  + Will see more GC in areas of higher recombination rates
* Neale Lab for height: heritability = 0.485 ; intercept = 1.313