(ldsc) Irenes-MBP:ldsc irenecho$ python /Users/irenecho/Desktop/final/ldsc/munge\_sumstats.py \

> --N 132091 \

> --sumstats /Users/irenecho/Desktop/final/5/Smoke/SmokingInitiation.txt.gz \

> --snp RSID \

> --a1 REF \

> --a2 ALT \

> --p PValue \

> --out /Users/irenecho/Desktop/final/ldscstep/step1 \

> --merge-alleles /Users/irenecho/Desktop/BMI/HW4-selected/ldsc\_inputs/w\_hm3.snplist \

>

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* LD Score Regression (LDSC)

\* Version 1.0.1

\* (C) 2014-2019 Brendan Bulik-Sullivan and Hilary Finucane

\* Broad Institute of MIT and Harvard / MIT Department of Mathematics

\* GNU General Public License v3

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Call:

./munge\_sumstats.py \

--out /Users/irenecho/Desktop/final/ldscstep/step1 \

--merge-alleles /Users/irenecho/Desktop/BMI/HW4-selected/ldsc\_inputs/w\_hm3.snplist \

--N 132091.0 \

--a1 REF \

--a2 ALT \

--snp RSID \

--sumstats /Users/irenecho/Desktop/final/5/Smoke/SmokingInitiation.txt.gz \

--p PValue

Interpreting column names as follows:

BETA: [linear/logistic] regression coefficient (0 --> no effect; above 0 --> A1 is trait/risk increasing)

N: Sample size

RSID: Variant ID (e.g., rs number)

ALT: Allele 2, interpreted as non-ref allele for signed sumstat.

REF: Allele 1, interpreted as ref allele for signed sumstat.

PVALUE: p-Value

Reading list of SNPs for allele merge from /Users/irenecho/Desktop/BMI/HW4-selected/ldsc\_inputs/w\_hm3.snplist

Read 1217311 SNPs for allele merge.

Reading sumstats from /Users/irenecho/Desktop/final/5/Smoke/SmokingInitiation.txt.gz into memory 5000000 SNPs at a time.

... done

Read 11733344 SNPs from --sumstats file.

Removed 10537353 SNPs not in --merge-alleles.

Removed 10709 SNPs with missing values.

Removed 0 SNPs with INFO <= 0.9.

Removed 0 SNPs with MAF <= 0.01.

Removed 0 SNPs with out-of-bounds p-values.

Removed 0 variants that were not SNPs or were strand-ambiguous.

1185282 SNPs remain.

Removed 0 SNPs with duplicated rs numbers (1185282 SNPs remain).

Removed 15 SNPs with N < 421868.0 (1185267 SNPs remain).

Median value of BETA was 2.82e-05, which seems sensible.

Removed 0 SNPs whose alleles did not match --merge-alleles (1185267 SNPs remain).

Writing summary statistics for 1217311 SNPs (1185267 with nonmissing beta) to /Users/irenecho/Desktop/final/ldscstep/step1.sumstats.gz.

Metadata:

Mean chi^2 = 1.732

Lambda GC = 1.486

Max chi^2 = 144.538

1693 Genome-wide significant SNPs (some may have been removed by filtering).

Conversion finished at Tue May 3 15:42:04 2022

Total time elapsed: 1.0m:0.29s

(ldsc) Irenes-MBP:ldsc irenecho$