

# Lab Notebook

Joyce Wang

9/8/2021

## Contents

<b>scripts Folder</b>	<b>1</b>
00_make_directories.sh . . . . .	1
01_UKBB_genotypes_filtered.sh . . . . .	2
01a_get_ambiguous_indel_snps.py . . . . .	2
01b_remove_ambiguous_indel_snps.py . . . . .	2
01c_find_duplicates.py . . . . .	2
01d_import_1KG.sh . . . . .	2

## scripts Folder

### 00\_make\_directories.sh

This file creates some subdirectories under the same directory where this file is located, in the following structure:

- data
  - ambiguous\_indel\_snps
  - intersecting\_filtered
  - kgp\_filtered
  - kgp\_merged
  - kgp\_meta
  - ukb\_filtered
  - ukb\_merged
  - ukb\_meta
  - ukb\_populations
  - models
  - phenotypes
  - gwas\_results
  - prs
  - kgp\_populations
  - fst
  - LDpred
    - \* prs

- \* tmp-data
  - \* val\_prs
  - prs\_comparisons
  - theory
  - theor\_herit
  - theoretical
- img

For me, these directories are under \$WORK2/pgs\_portability

### 01\_UKBB\_genotypes\_filtered.sh

This file filters out the indels and ambiguous variants.

**Currently having an error here trying to load the .bim files:**

FileNotFoundError: [Errno 2] No such file or directory: '/corral-repl/utexas/Recombining-sex-chro/ukb/da

01a\_get\_ambiguous\_indel\_snps.py

01b\_remove\_ambiguous\_indel\_snps.py

01c\_find\_duplicates.py

01d\_import\_1KG.sh