```
##########################################################
##############  Generating PGS using PRSCS     #################
###    Daniel Gustavson - daniel.gustavson@colorado.edu - 4/1/25   ###
##########################################################


################################################
###         STEP 1: REFORMAT SUMMARY STATISTICS          ###
################################################

# Use an awk command to keep only the columns you need for prs-cs
# Replace the column numbers in the following command (i.e., $2, $3, etc.) with the columns from you sst file
# COLUMNS NEEDED: SNP A1 A2 BETA P
# NOTE: You want to use summary statistics in Build 37 format!

# EXAMPLE:
# Current columns:
# ROW SNP    A1    A2    EAF    beta   SE    p

# Needed columns:
# SNP A1 A2 BETA P

# awk -F"\t" 'NR==1{print "SNP","A1","A2","BETA","P";next}{print $2,$3,$4,$6,$8}' PHENOTYPE_summstats.txt >
PHENOTYPE_cs_format.txt


#######################################
###              STEP 2: RUN PRS-CS              ###
#######################################
```

```
# Submit a SLURM script to run PRS-CS.
# An example can be found here: /pl/active/IBG/dang/CATSLife_PGSs/PRScs/EA4/prs_cs_EA4_CATSLife

# Your script should look like this (this example is for the educational attainment [EA4] GWAS):

#!/bin/bash
#SBATCH --mail-type=ALL
#SBATCH --nodes=1
#SBATCH --ntasks=20
#SBATCH --time=0-15:00:00     # 1 day - 1 hrs
#SBATCH --output=CATSLife_PRScs_PHENOTYPE_Log.txt
#SBATCH --job-name=CATSLife_PRScs_PHENOTYPE.txt

#########################
####        User Input        ####
#########################
# Name of the phenotype
PHENO=EA4
# Your directory where this script is run
PGS_PATH=/pl/active/IBG/people/dang/CATSLife_PGSs/PRScs/EA4/
# N for this GWAS
N_GWAS=765283
# Path for Sumstat file
SUM_PATH=/pl/active/IBG/people/dang/CATSLife_PGSs/PRScs/EA4/EA4_cs_format.txt

#### Probably leave these as is ####
# path to directory for genetic data
BIM_PATH=/pl/active/IBG/data/CATSLife/derived/PGS/CATSLife_hrc_filtered_April_2025
# path to reference directory (EUR)
```

```
REF_PATH=/pl/active/IBG/people/dang/PRScs/ld_files/ldblk_1kg_eur

########################
####       Run PRScs       ####
########################

ml anaconda

python /pl/active/IBG/people/dang/PRScs/PRScs.py \
--ref_dir=$REF_PATH \
--bim_prefix=$BIM_PATH \
--n_gwas=$N_GWAS \
--sst_file=$SUM_PATH \
--out=$PGS_PATH/${PHENO}

cat $PGS_PATH/${PHENO}_pst_eff_a1_b0.5_phiauto_chr* > $PGS_PATH/${PHENO}_prs_cs_gwas.txt

plink --bfile $BIM_PATH \
--score $PGS_PATH/${PHENO}_prs_cs_gwas.txt 2 4 6 \
--out $PGS_PATH/${PHENO}_scores




#####################################################################
###                          STEP 3: MERGE WITH ANCESTRY & PC FILE                          ###
#####################################################################

## IF everything worked out, you should have a PHENOTYPE.profile file in your output directory

# Next, download this file which has the ancestry and PC information for CATSLife:
```

```
# /pl/active/IBG/data/CATSLife/derived/PGS/CATSLife_Ancestry_Sept2024.csv


# Here is some example R code


# Read files

PCs <- read.csv("CATSLife_Ancestry_April2025.csv",header=T)

EA4 <- read.table("EA4_scores.profile",header=T)


# Rename SCORE column and select just the IID and PGS for merging (you'll have to name the phenotype yourself)

PGS$PHENOTYPE_PGS <- PGS$SCORE

PGS <- select(PGS, IID, PHENOTYPE_PGS)

head(PGS)

length(PGS$IID)


# Merge and keep only IDs in the Ancestry file

PGS_PCS <- merge(PGS, PCs, all.y=T)

head(PGS_PCS)

length(PGS_PCS$IID)


## Now you should be able to merge this file with additional phenotypic information. Good luck!
```