A screenshot of a computer

Description automatically generated

cd /media/MPEdge16/MM137/geno/

## BED/BIM/FAM PLINK TO VCF

/media/MPEdge16/MM137/geno/plink2 --bfile /media/MPEdge16/MM137/geno/data/PLINKfiles/deliverable\_MM-10873-Mashayekhi\_BestPerformingSNPs --recode vcf

## REMOVE INDEL (ID)

cat /media/MPEdge16/MM137/geno/plink2.vcf | awk '$4 != "I" && $4 != "D"' > /media/MPEdge16/MM137/geno/plink2\_id\_rm.vcf

use ref from wget <https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz>;

## do

/home/simmonsns/jdk-17.0.9/bin/java -jar /media/MPEdge16/MM137/geno/picard.jar CreateSequenceDictionary R=/media/MPEdge16/hg38\_ref/hg38.fa

## this hg38 ref has chr before number, so do

awk 'BEGIN {FS = OFS = "\t"} !/^#/ {$1 = "chr"$1} 1 ' /media/MPEdge16/MM137/geno/plink2\_id\_rm.vcf > /media/MPEdge16/MM137/geno/plink2\_renamed.vcf

## then do

/home/simmonsns/jdk-17.0.9/bin/java -jar /media/MPEdge16/MM137/geno/picard.jar LiftoverVcf I=/media/MPEdge16/MM137/geno/plink2\_renamed.vcf O=/media/MPEdge16/MM137/geno/lifted\_over.vcf CHAIN=/media/MPEdge16/MM137/geno/hg19ToHg38.over.chain.gz REJECT=/media/MPEdge16/MM137/geno/rejected\_variants.vcf R=/media/MPEdge16/hg38\_ref/hg38.fa

## proceed with vcf correlation vs souporcell vcf