## GET REF

mkdir /media/MPEdge16/hg38\_kir\_ref

cd /media/MPEdge16/hg38\_kir\_ref/

wget <https://zenodo.org/records/7762015/files/cellranger-refdata-GRCh38-KIR.tar.gz?download=1>

mv cellranger-refdata-GRCh38-KIR.tar.gz?download=1 cellranger-refdata-GRCh38-KIR.tar.gz

tar -xvzf cellranger-refdata-GRCh38-KIR.tar.gz

/home/simmonsns/jdk-17.0.9/bin/java -jar /media/MPEdge16/MM137/geno/picard.jar CreateSequenceDictionary R=/media/MPEdge16/hg38\_kir\_ref/GRCh38/fasta/genome.fa

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

## PREPEND chr TO CHROMOSOME NUMBER

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

## CONVERT TO HG38 USING ALVES KIR REFERENCE

/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\_kir\_ref/GRCh38/fasta/genome.fa

ERROR: Encountered a contig, chr1 that is not part of the target reference

A black screen with many small lines

Description automatically generated with medium confidence

TROUBLESHOOT:

head -10000 /media/MPEdge16/hg38\_kir\_ref/GRCh38/fasta/genome.fa > /media/MPEdge16/MM137/geno/head10000\_ref.fa.txt

tail -10000 /media/MPEdge16/hg38\_kir\_ref/GRCh38/fasta/genome.fa > /media/MPEdge16/MM137/geno/tail10000\_ref.fa.txt

head -10000 /media/MPEdge16/MM137/geno/plink2\_renamed.vcf > /media/MPEdge16/MM137/geno/head10000\_renamed.vcf.txt

tail -10000 /media/MPEdge16/MM137/geno/plink2\_renamed.vcf > /media/MPEdge16/MM137/geno/tail10000\_renamed.vcf.txt

TRY FIX: do not add “chr” in front of chromosome number; instead:

cp /media/MPEdge16/MM137/geno/plink2\_id\_rm.vcf /media/MPEdge16/MM137/geno/plink2\_renamed.vcf

A computer screen shot of a computer screen

Description automatically generated

Runs without error but no variants are successfully lifted over.

TRY FIX: do not use KIR ref; 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

which runs fine

ALTERNATIVE: try keep indel using non-kir ref

## do

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

which fails, indels not supported