HS Rats Lc-WGS Genotyping Pipeline

Denghui Chen, Palmer Lab
June 07 2022

HS Rats Lc-WGS Genotyping Pipeline - Overview

Genotyping

Quality Control

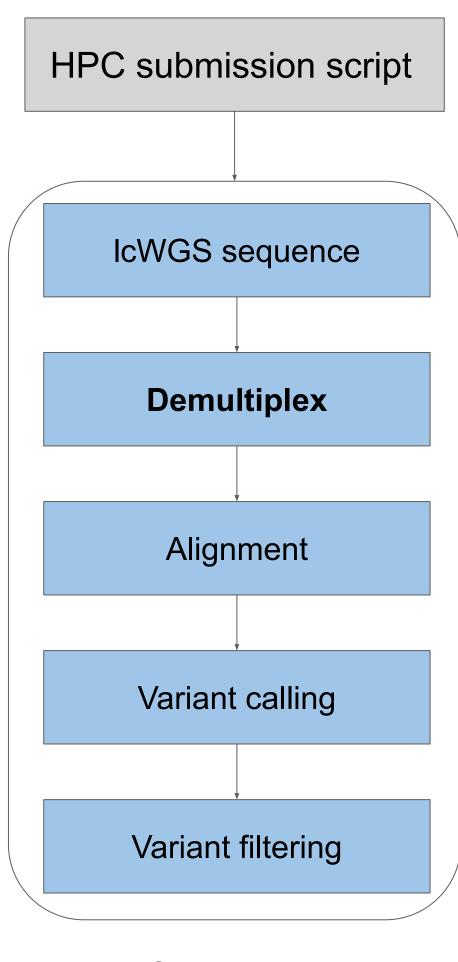**Genotyping log table**

Genotyping summary report

HS Rats Lc-WGS Genotyping Pipeline - Demultiplex

Demultiplex
Command:
java -jar ${fgbio} DemuxFastqs \
    --inputs ${pre_demux_fastq_R1} ${pre_demux_fastq_R2} \
    --metadata ${sample_sheet} \
    --read-structures 8B12M+T 8M+T \
    --output-type=Fastq \
    --threads ${ncpu} \
    --output ${demux_dir}/fastq \
    --metrics ${demux_dir}/metrics/${metrics_name}_demux_barcode_metrics.txt

- Fgbio: http://fulcrumgenomics.github.io/fgbio/

HPC submission script

lcWGS sequence

Demultiplex

Alignment

Variant calling

Variant filtering

Genotyping

HS Rats Lc-WGS Genotyping Pipeline - Alignment

Alignment
Command:
bwa mem -aM -R

- bwa: http://bio-bwa.sourceforge.net/bwa.shtml

- -R: read group header line such as
  "@RG\tID:${flowcell_name}.${lane}\tLB:${library_id}\tPL:${platform}\tSM:${sample}"

samtools sort -o ${prefix}_sorted.bam ${prefix}.sam

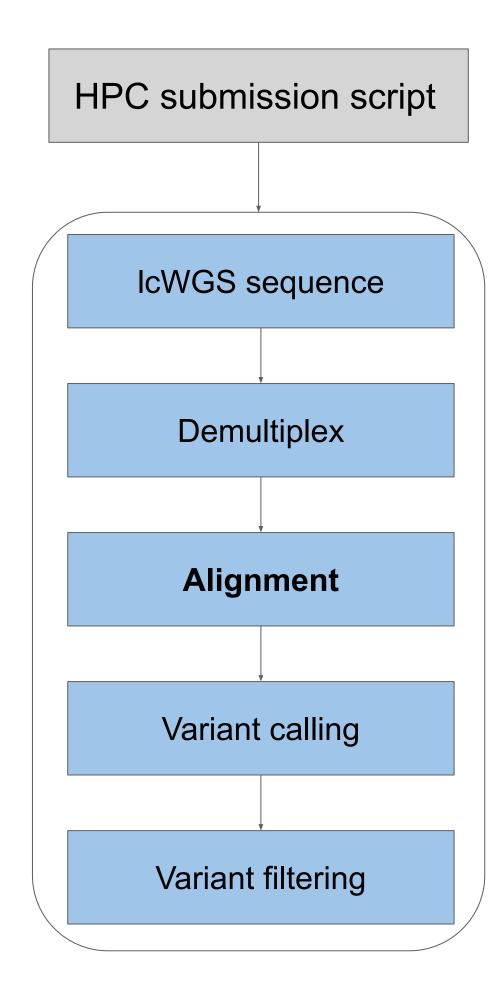- samtools: https://www.htslib.org/doc/samtools.html

java -jar ${picard} MarkDuplicates \
    --INPUT ${prefix}_sorted.bam \
    --REMOVE_DUPLICATES false \
    --ASSUME_SORTED true \
    --METRICS_FILE ${prefix}_sorted_mkDup_metrics.txt \
    --OUTPUT ${prefix}_sorted_mkDup.bam &

- picard: https://broadinstitute.github.io/picard/

samtools index ${prefix}_sorted_mkDup.bam ${prefix}_sorted_mkDup.bai
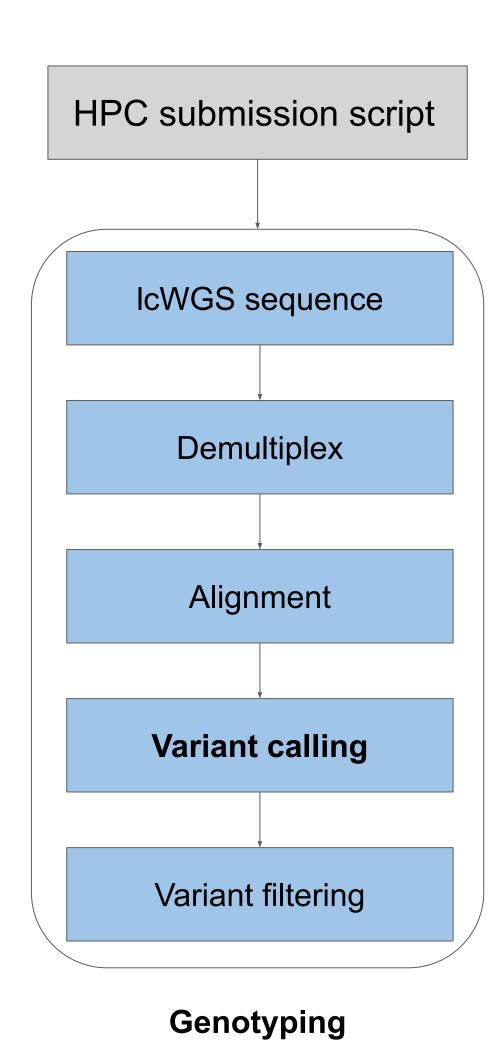
- samtools: https://www.htslib.org/doc/samtools.html

HPC submission script

lcWGS sequence

Demultiplex

Alignment

Variant calling

Variant filtering

Genotyping

HS Rats Lc-WGS Genotyping Pipeline - Variant Calling

Variant Calling
Command:
STITCH(

                    buffer = 1Mb,
                    method = "diploid",
                    posfile = posfile,
    reference_haplotype_file = refHap,
        reference_legend_file = refLgd,
                    K = 8,
                    niterations = 2,
                    nGen = 100,
    output_haplotype_dosages = TRUE
)

- STITCH: https://github.com/rwdavies/STITCH

HPC submission script

lcWGS sequence

Demultiplex

Alignment

Variant calling

Variant filtering

Genotyping

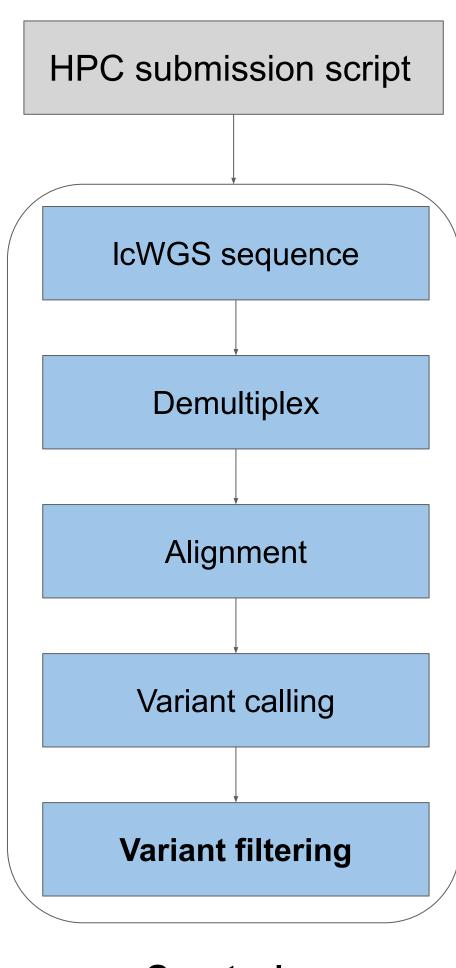# HS Rats Lc-WGS Genotyping Pipeline - Variant Filtering

Variant Filtering
Command:
```
bcftools view -e "INFO_SCORE<0.9" \
    -Oz -o ${stitch_path}/stitch_HD_INFO_0.9.vcf.gz \
    ${stitch_path}/stitch_HD.vcf.gz

bcftools view -T ^${remove_snps} \
  -Oz -o ${stitch_path}/stitch_HD_INFO_0.9_rm_discordant.vcf.gz \
  ${stitch_path}/stitch_HD_INFO_0.9.vcf.gz
```

- Bcftools: http://samtools.github.io/bcftools/bcftools.html

HPC submission script

lcWGS sequence

Demultiplex

Alignment

Variant calling

Variant filtering

Genotyping