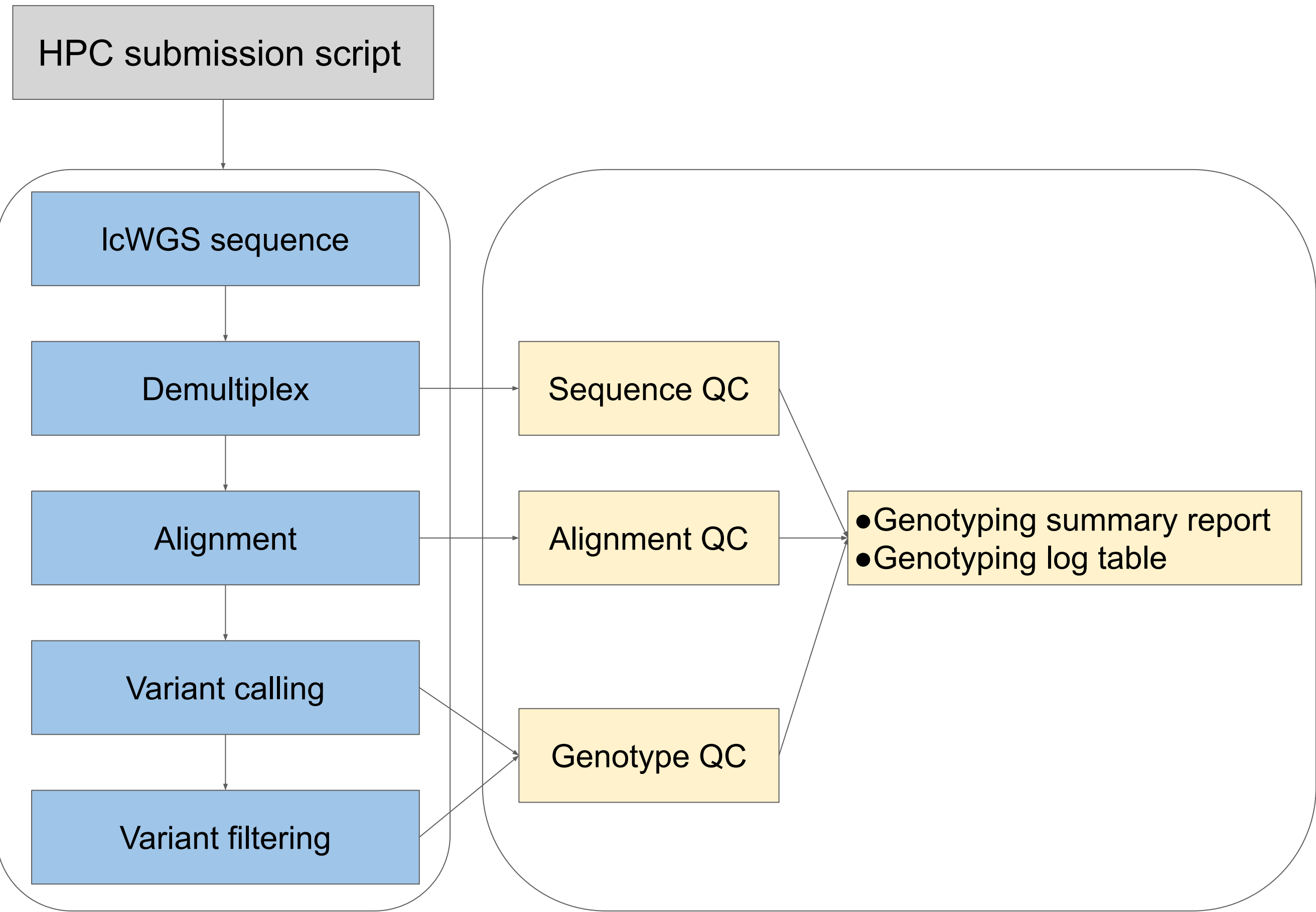


HS Rats Lc-WGS Genotyping Pipeline

HS Rats Lc-WGS Genotyping Pipeline - Overview



hs_rats_n1912_01202021_genotype_log																
rid	demux_reads	mapped_read_pairs	unmapped_read_ratio	duplication_ratio	uniquely_mapped_ratio	QC_reads	QC_sex	date	missing_rate	QC_missing	heterozygosity_rate	QC_heterozygosity	colorcode	QC_color	albino	QC_pass
0007E76F	404586	3963945	0.0129424877804669	0.094519	0.825935056711803	pass	pass	1202021	0.0632717	pass	0.311291206161811	pass	BROWN	pass	pass	
0007E7AC5	495575	4866408	0.0114243483812553	0.104095	0.80086888023501	pass	pass	1202021	0.0491388	pass	0.31555266332747	pass	ALBINO	pass	pass	
0007E7B77	448027	4410559	0.010871562706198	0.09046	0.79679862646055	pass	pass	1202021	0.0494957	pass	0.320133765403521	pass	BLACK	pass	pass	
0007E7C0F	223293	2208949	0.0119636046196201	0.091652	0.818670758879461	pass	pass	1202021	0.0610555	pass	0.328181814375031	pass	BROWN	pass	pass	
0007E7C0D	521645	5128248	0.0107400218021464	0.095589	0.807638886554516	pass	pass	1202021	0.0489411	pass	0.306327805338772	pass	BLACK	pass	pass	
0007E7B8B	2129639	2090405	0.011951766984343	0.0938	0.78673820691967	pass	pass	1202021	0.058943	pass	0.316202131945788	pass	BROWN	pass	pass	
0007E7B8B	482489	4743939	0.011578018388664	0.094234	0.817326567142137	pass	pass	1202021	0.0484717	pass	0.3144764823267071	pass	ALBINO	pass	pass	
0007E7B8B	3518738	3454891	0.0113591829345863	0.094777	0.8205853216019	pass	pass	1202021	0.0518971	pass	0.319006469459522	pass	ALBINO	pass	pass	
0007E7B40	2301232	2258517	0.0116510503403394	0.08325	0.8055740398443	pass	pass	1202021	0.060085	pass	0.31282608914947	pass	BLACK	pass	pass	
0007E7B40	4478238	4394181	0.0116752058312523	0.094365	0.800791623925252	pass	pass	1202021	0.0502188	pass	0.3022108881288	pass	BROWN	pass	pass	
0007E7B4C	3475460	3416295	0.0106802082120755	0.083984	0.787824607781991	pass	pass	1202021	0.0525314	pass	0.306878821488453	pass	BROWNHOOD	pass	pass	
0007E7B4F	2580796	2545405	0.0121744257153816	0.083325	0.822147176751659	pass	pass	1202021	0.0568432	pass	0.280742821638007	pass	BLACKHOOD	pass	pass	
0007E7B78	2963547	2901830	0.011298464675555	0.083275	0.78811918800862	pass	pass	1202021	0.0556884	pass	0.322841584343428	pass	BROWN	pass	pass	
0007E7B0E	9729231	9538678	0.0118044762538995	0.087257	0.824804372288116	pass	pass	1202021	0.0417072	pass	0.32508450595069	pass	BROWNHOOD	pass	pass	
0007E7B4D	3162990	3105865	0.0114281107433157	0.083341	0.804437889465348	pass	pass	1202021	0.035868	pass	0.312171474896151	pass	BLACKHOOD	pass	pass	
0007E7B42	3220934	3166543	0.01125575816566133	0.083789	0.80891188054006	pass	pass	1202021	0.0518414	pass	0.2841728283113	pass	BROWNHOOD	pass	pass	
0007E7B05	5820519	5814567	0.0115253283880126	0.095138	0.82218608211944	pass	pass	1202021	0.0448823	pass	0.29980037249794	pass	BLACKHOOD	pass	pass	
0007E7B0A	3325086	325818	0.0122729130814968	0.09112	0.82515315323039	pass	pass	1202021	0.0524737	pass	0.31448848703879	pass	ALBINO	pass	pass	
0007E7B05	2767965	2714207	0.012188092118789	0.091117	0.82515315323039	pass	pass	1202021	0.0557854	pass	0.3245546202023099	pass	BROWN	pass	pass	
0007E7B01	2146567	2108346	0.0118286778815007	0.091258	0.796892518762078	pass	pass	1202021	0.0587152	pass	0.28754588421099	pass	ALBINO	pass	pass	
0007E7B0E	4735187	4648178	0.011328804851644	0.094368	0.803880194784946	pass	pass	1202021	0.0487808	pass	0.290254388458734	pass	ALBINO	pass	pass	
0007E7B53	4888622	4811008	0.0107271467861977	0.094884	0.79565777467313	pass	pass	1202021	0.0477179	pass	0.32177894159271	pass	BLACK	pass	pass	
0007E7BAC	3048696	2988399	0.0123172233351702	0.083844	0.826874221108615	pass	pass	1202021	0.0542804	pass	0.28893538807051	pass	BROWNHOOD	pass	pass	
0007E7BFC	2649187	2595189	0.0131438551567038	0.082795	0.815432782579614	pass	pass	1202021	0.058693	pass	0.312780284168063	pass	BROWN	pass	pass	
0007E7B09	5511988	5404439	0.012233886219854	0.093861	0.81682862752025	pass	pass	1202021	0.0457838	pass	0.315588802837386	pass	ALBINO	pass	pass	
0007E7B08	4828242	4708266	0.0120711912118865	0.090337	0.815985393174548	pass	pass	1202021	0.0478658	pass	0.3086087678147	pass	ALBINO	pass	pass	
0007E7B06	2228703	218742	0.012523044548883	0.094489	0.807863557177795	pass	pass	1202021	0.0584239	pass	0.322887858153584	pass	BLACKHOOD	pass	pass	
0007E7B0A	2786688	2744817	0.0114475515373058	0.08236	0.80634848366516	pass	pass	1202021	0.0502678	pass	0.318018522760758	pass	BLACK	pass	pass	
0007E7B12	3585470	3508587	0.0130640587147571	0.094574	0.822288151901582	pass	pass	1202021	0.0518233	pass	0.31573518588558	pass	BLACK	pass	pass	
0007E7B1B	2782346	2732882	0.013254884440288	0.083886	0.814475882314205	pass	pass	1202021	0.0587882	pass	0.29888114867865	pass	BROWNHOOD	pass	pass	
0007E7B1A	4910558	4828891	0.01058148064884	0.094485	0.811728004142912	pass	pass	1202021	0.0482001	pass	0.322784231388348	pass	ALBINO	pass	pass	
0007E7B1E	2627967	2578227	0.0120451467048957	0.082363	0.78782604898909	pass	pass	1202021	0.0572396	pass	0.309587341488409	pass	BLACKHOOD	pass	pass	
0007E7B0D	2006740	1988175	0.0123713088897537	0.091884	0.785451087202428	pass	pass	1202021	0.0607825	pass	0.304182838391851	pass	BROWNHOOD	pass	pass	
0007E7B05	2483634	2419405	0.01362815453413	0.082382	0.81982541786652	pass	pass	1202021	0.058659	pass	0.3081528822037	pass	ALBINO	pass	pass	
0007E7B15	3185382	3138687	0.011872886022738	0.09008	0.80784583036799	pass	pass	1202021	0.0547235	pass	0.3084788233303	pass	BROWNHOOD	pass	pass	
0007E7B27	7046872	6928872	0.0101271919587138	0.083876	0.80008467913364	pass	pass	1202021	0.044212	pass	0.31034524781415	pass	BLACK	pass	pass	
0007E7B70	4012845	3944505	0.0109612238311637	0.094836	0.79081341088391	pass	pass	1202021	0.0513888	pass	0.28381818717703	pass	BROWNHOOD	pass	pass	
0007E7B91	2536877	2488958	0.0115370197280759	0.082138	0.797780498430313	pass	pass	1202021	0.0588307	pass	0.311728880205857	pass	BROWNHOOD	pass	pass	
0007E7B06	3082882	3023124	0.012167828682116	0.082385	0.8158612001827	pass	pass	1202021	0.0508642	pass	0.2801088818873	pass	BLACKHOOD	pass	pass	

Genotyping log table

Genotyping

Quality Control

Genotyping summary report

1 Pipeline Overview
2 Data
3 Demultiplex Results
3.1 Number of reads for each sample
3.2 Percentage of unmatched barcode reads on demultiplex
4 Alignment Results
5 MultiQC Summary for Demultiplex and Alignment Steps
6 Genotype Results
7 Outliers Report
8 Appendix A. Relevant Softwares
9 Appendix B. Preparation for Reference Data

3 Demultiplex Results

Basic statistics of demultiplexing with Fgbio are shown in figure 2, 3 and 4. These are extracted from the SampleBarcodeMetric of Fgbio DemuxFastqs function.

3.1 Number of reads for each sample

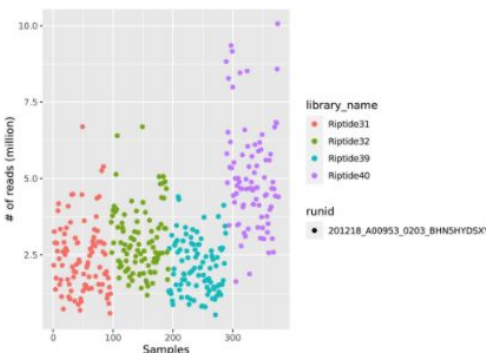


Figure 2: Number of reads for each sample (scatter plot)

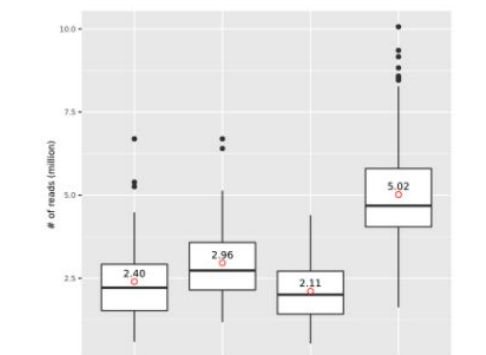


Figure 3: Number of reads for each sample (box plot)

3.2 Percentage of unmatched barcode reads on demultiplex

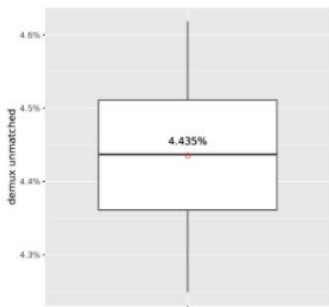


Figure 4: Percentage of unmatched barcode reads on demultiplex step

4 Alignment Results

Basic statistics of alignment with BWA mem are shown in figure 5 to figure 18. Figure 5 to figure 9 are extracted from the DuplicationMetrics of Picard MarkDuplicates function.

4.1 Number of mapped reads for each sample

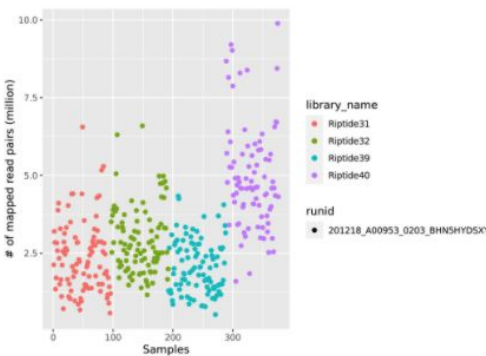


Figure 5: Number of mapped reads for each sample (scatter plot)

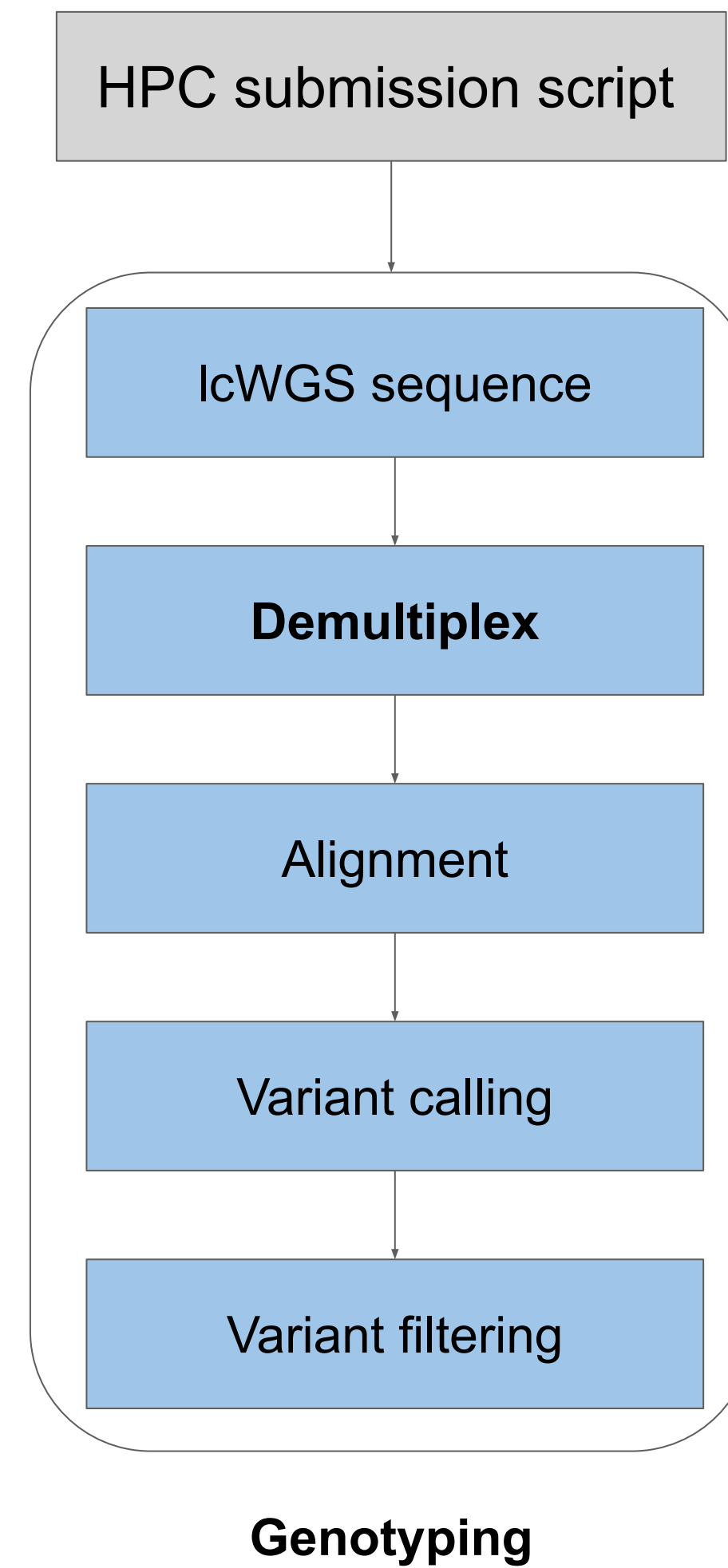
HS Rats Lc-WGS Genotyping Pipeline - Demultiplex

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/>



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

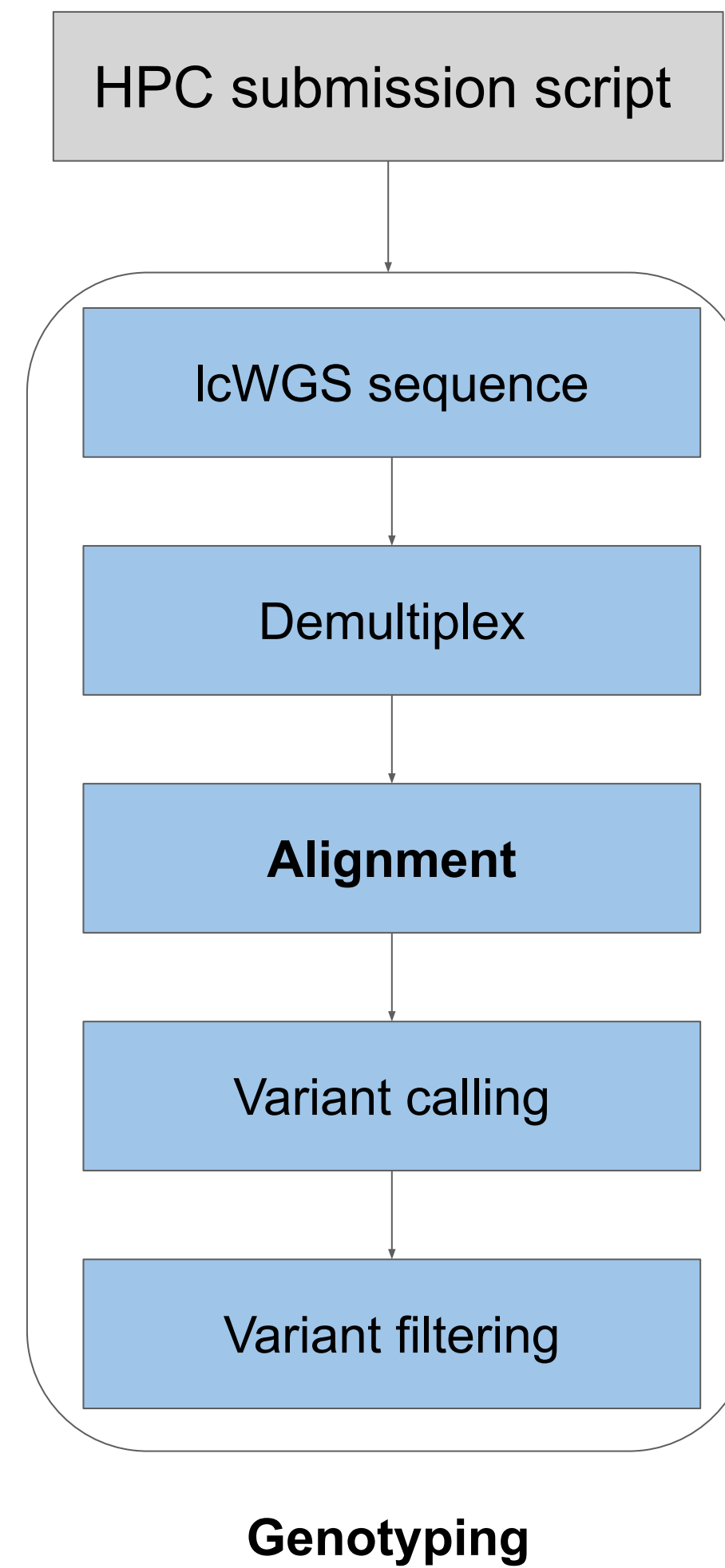
- samtools: <https://www.htslib.org/doc/samtools.html>

```
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>



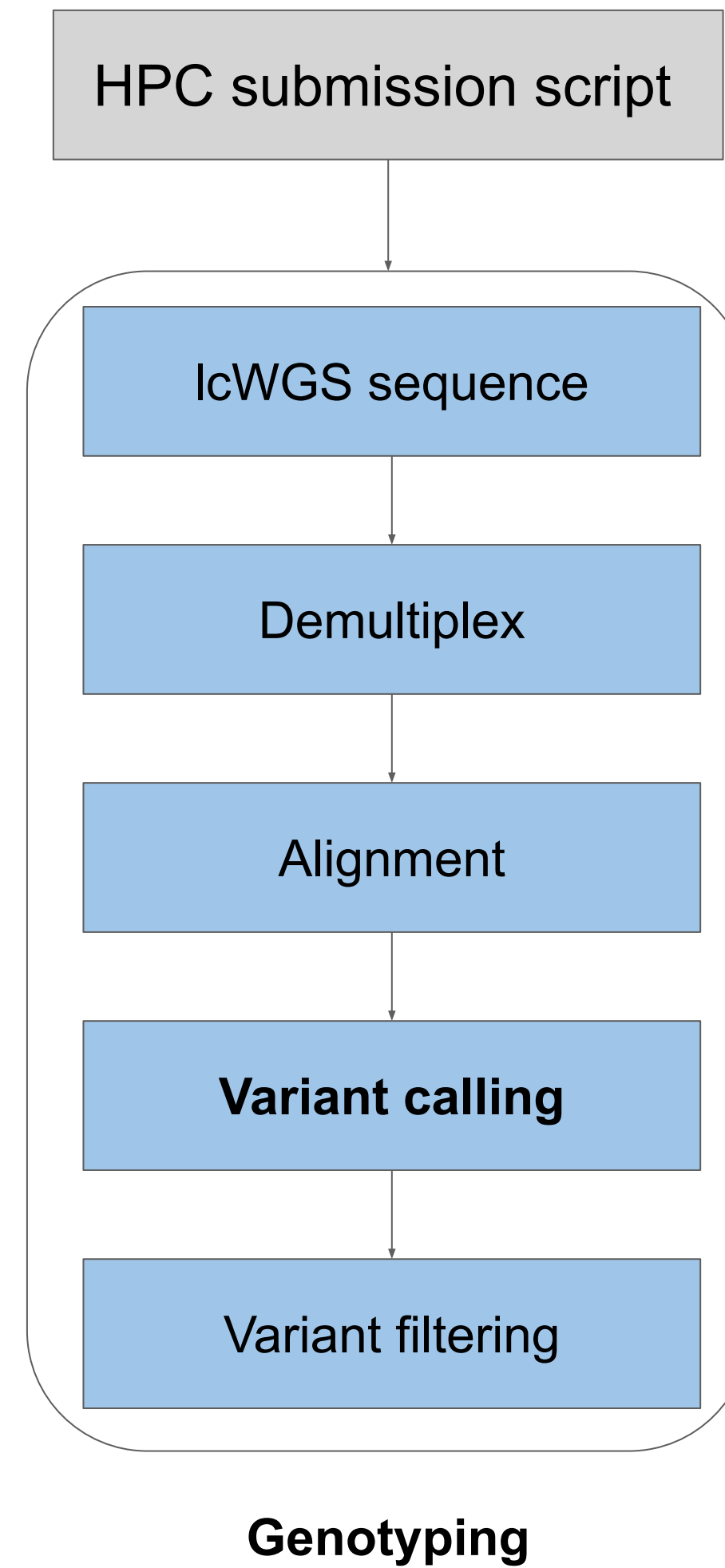
HS Rats Lc-WGS Genotyping Pipeline - Variant Calling

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>



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>

