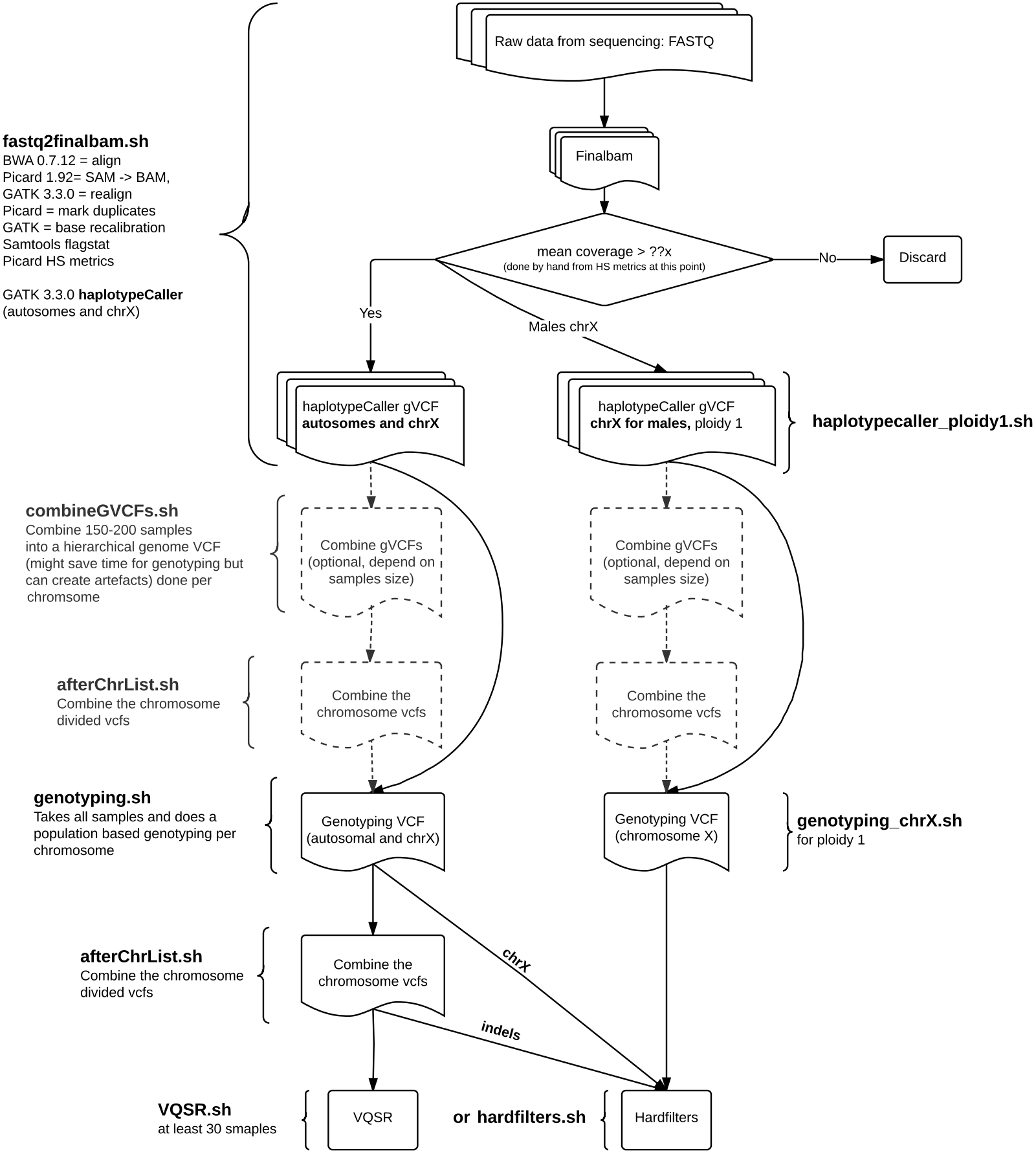
Raw data from sequencing: FASTQ
Finalbam
mean coverage > ??x
(done by hand from HS metrics at this point)
No
Discard
Yes
Males chrX
haplotypeCaller gVCF
autosomes and chrX
haplotypeCaller gVCF
chrX for males, ploidy 1
Combine gVCFs
(optional, depend on
samples size)
Combine gVCFs
(optional, depend on
samples size)
Combine the
chromosome vcfs
Combine the
chromosome vcfs
Genotyping VCF
(autosomal and chrX)
Genotyping VCF
(chromosome X)
Combine the
chromosome vcfs
chrX
indels
VQSR
or hardfilters.sh
Hardfilters
fastq2finalbam.sh
BWA 0.7.12 = align
Picard 1.92= SAM -> BAM,
GATK 3.3.0 = realign
Picard = mark duplicates
GATK = base recalibration
Samtools flagstat
Picard HS metrics
GATK 3.3.0 haplotypeCaller
(autosomes and chrX)
haplotypecaller_ploidy1.sh
combineGVCFs.sh
Combine 150-200 samples
into a hierarchical genome VCF
(might save time for genotyping but
can create artefacts) done per
chromsome
afterChrList.sh
Combine the chromosome
divided vcfs
genotyping.sh
Takes all samples and does a
population based genotyping per
chromosome
genotyping_chrX.sh
for ploidy 1
afterChrList.sh
Combine the chromosome
divided vcfs
VQSR.sh
at least 30 smaples