snakemake rule	program used	input	output	scatter-gather status (# jobs submitted)
processRef	1.) samtools faidx 2.) picard CreateSequenceDictionary	reference.fasta	reference.fasta.fai reference.dict	NA
bam2gvcf	gatk HaplotypeCaller	s1.bam { list1.list, list2.list, list3.list } s2.bam { list1.list, list2.list, list3.list } s3.bam { list1.list, list2.list, list3.list }	s1_L1.raw.g.vcf s1_L2.raw.g.vcf s1_L3.raw.g.vcf s2_L1.raw.g.vcf s2_L2.raw.g.vcf s2_L3.raw.g.vcf s3_L1.raw.g.vcf s3_L2.raw.g.vcf s3_L3.raw.g.vcf	scattered by sample scattered by list file # jobs = (# samples)*(# list files)
gvcf2DB	gatk GenomicsDBImport	$\sum_{i=1}^{3} s\boldsymbol{i}\_L1.\,raw.\,g.\,vcf$ $\sum_{i=1}^{3} s\boldsymbol{i}\_L2.\,raw.\,g.\,vcf$ $\sum_{i=1}^{3} s\boldsymbol{i}\_L3.\,raw.\,g.\,vcf$	DB_L1 DB_L2 DB_L3	gathered by sample scattered by list file # jobs = # list files
DB2vcf	gatk GenotypeGVCFs	DB_L1 DB_L2 DB_L3	L1.vcf L2.vcf L3.vcf	scattered by list file # jobs = # list files
gatherVcfs	1.) gatk GatherVcfs 2.) gatk VariantFiltration	L1.vcf L2.vcf L.3vcf	Combined.vcf Combined_hardFiltered.vcf	scattered by list file # jobs = # list files

Flow diagram (left column): processRef → bam2gvcf → gvcf2DB → DB2vcf → gatherVcfs

list1.list, list2.list, list3.list → bam2gvcf