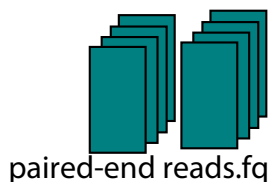


Cohort: 1x ancestor + n x MA line



reference genome

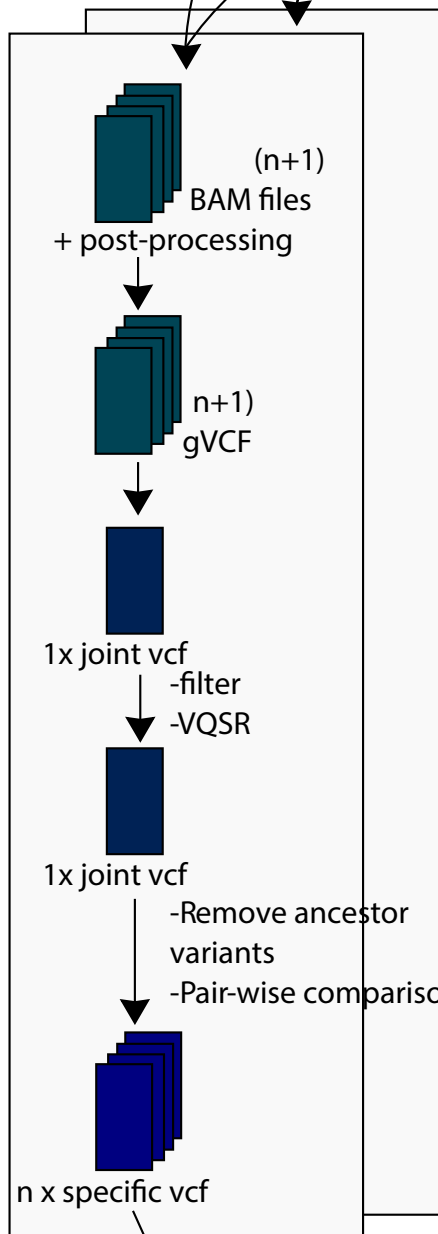
genome.fa

repeats  
masked

non-  
masked

(based on GATK)

script 'GATK\_fq-to-gVCF.sh'



CALLING SHORT VARIANTS

Subtract  
masking  
artefacts

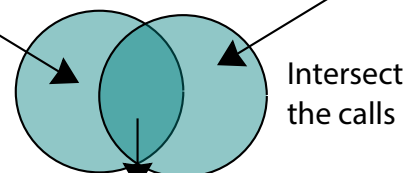
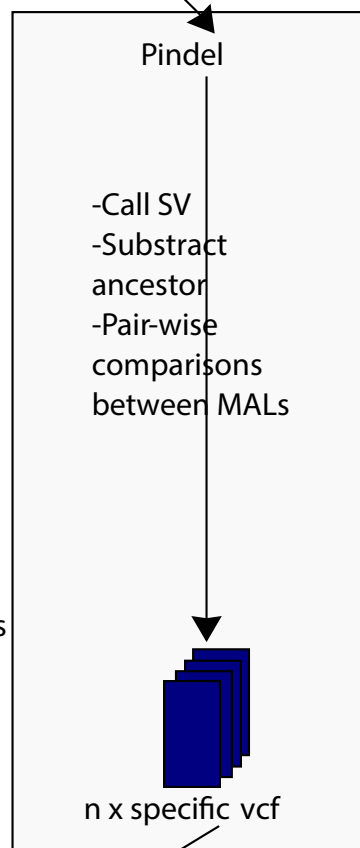
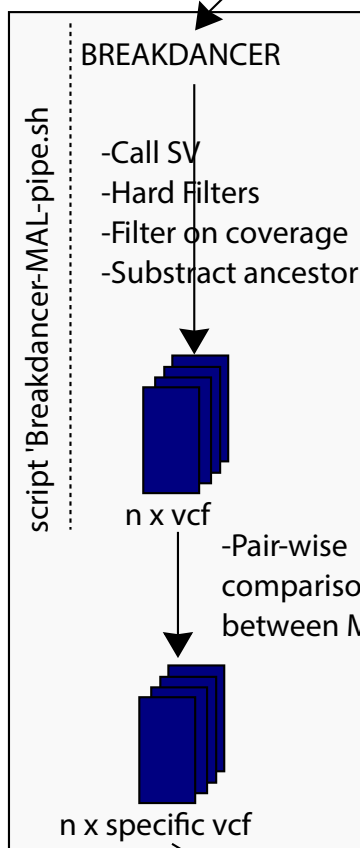
n x final vcf  
'small variants'

script 'fq-to-bam.sh'

(n+1) BAM files  
+ post-processing

filter unmapped  
read pairs

CALLING LONG VARIATIONS



Intersect  
the calls

Validation by  
inspecting read  
alignments

n x final vcf  
'long variants'

Pool small and long variants

n x final vcf  
'total variants'