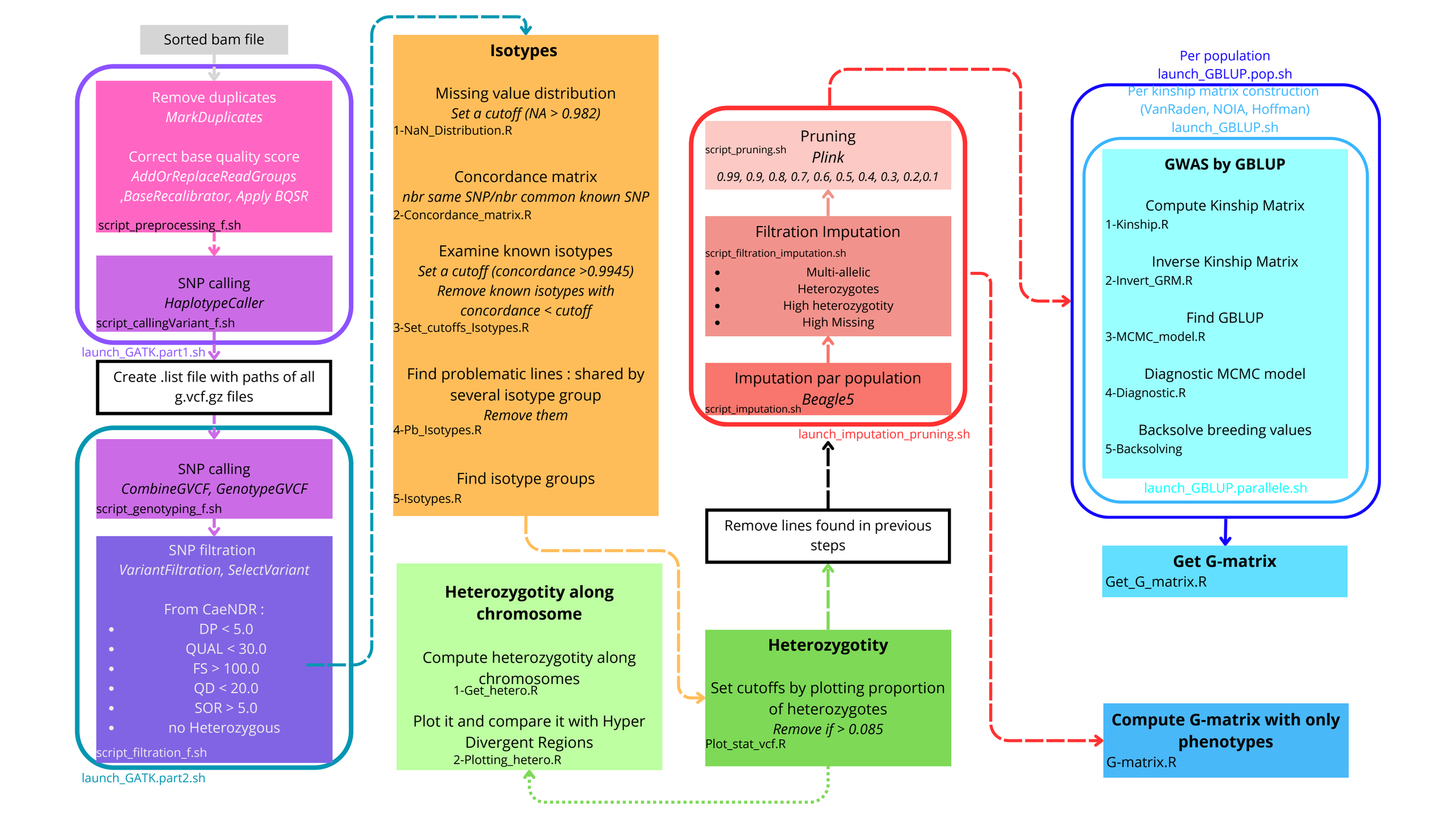
Sorted bam file
Remove duplicates
MarkDuplicates
Correct base quality score
AddOrReplaceReadGroups
,BaseRecalibrator, Apply BQSR
script_preprocessing_f.sh
SNP calling
HaplotypeCaller
script_callingVariant_f.sh
launch_GATK.part1.sh
Create .list file with paths of all g.vcf.gz files
SNP calling
CombineGVCF, GenotypeGVCF
script_genotyping_f.sh
SNP filtration
VariantFiltration, SelectVariant
From CaeNDR :
DP < 5.0
QUAL < 30.0
FS > 100.0
QD < 20.0
SOR > 5.0
no Heterozygous
script_filtration_f.sh
launch_GATK.part2.sh
Isotypes
Missing value distribution
Set a cutoff (NA > 0.982)
1-NaN_Distribution.R
Concordance matrix
nbr same SNP/nbr common known SNP
2-Concordance_matrix.R
Examine known isotypes
Set a cutoff (concordance >0.9945)
Remove known isotypes with concordance < cutoff
3-Set_cutoffs_Isotypes.R
Find problematic lines : shared by several isotype group
Remove them
4-Pb_Isotypes.R
Find isotype groups
5-Isotypes.R
Heterozygotity along chromosome
Compute heterozygotity along chromosomes
1-Get_hetero.R
Plot it and compare it with Hyper Divergent Regions
2-Plotting_hetero.R
Heterozygotity
Set cutoffs by plotting proportion of heterozygotes
Remove if > 0.085
Plot_stat_vcf.R
Remove lines found in previous steps
script_pruning.sh
Pruning
Plink
0.99, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2,0.1
Filtration Imputation
script_filtration_imputation.sh
Multi-allelic
Heterozygotes
High heterozygotity
High Missing
Imputation par population
Beagle5
script_imputation.sh
launch_imputation_pruning.sh
Per population
launch_GBLUP.pop.sh
Per kinship matrix construction
(VanRaden, NOIA, Hoffman)
launch_GBLUP.sh
GWAS by GBLUP
Compute Kinship Matrix
1-Kinship.R
Inverse Kinship Matrix
2-Invert_GRM.R
Find GBLUP
3-MCMC_model.R
Diagnostic MCMC model
4-Diagnostic.R
Backsolve breeding values
5-Backsolving
launch_GBLUP.parallele.sh
Get G-matrix
Get_G_matrix.R
Compute G-matrix with only phenotypes
G-matrix.R