**From VCF to HAPMAP for GWAS**

*Script name:* from\_vcf\_to\_hapmap\_for\_gwas.pl

What it does: filter VCF and get a hapmap for GWAS

*Synopsis:*

***perl from\_vcf\_to\_hapmap\_for\_gwas\_new.pl -r REF\_PATH -n number\_of\_indv -f minor\_allele\_frequency -c min\_inbreeding\_coeff -a allowed\_missing\_percent -i missingness -s readdepthlimitmin -t readdepthlimitmax -u readdepthmin\_allele -v maxindivwithmissingdata -p ploidy -x file\_extension***

*Parameters:*

-n (int): number of individuals

-f (float): minor allele frequency

-c (float): min inbreeding coefficient

-a (int): allowed missing percent per SNP

-i (int): Set the maximum missingness percentage allowed per individual

-s (int): min coverage per accession

-t (int): max coverage per accession

-u (int): min coverage per allele

-v (int): max number of accession with missing data

-p (int): ploidy

-x (string): file extension (vcf)

*All the steps in detail*

* SNP filtering (SnpCluster max 3 in 10bp and Qualitybydepth < 1.5) with ViariantFiltration and SelectVariants from GATK

<https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-0/org_broadinstitute_gatk_tools_walkers_variantutils_SelectVariants.php>

* Discad individuals based on their missingness with custom perl script (discard\_individuals\_based\_on\_missingness\_v3.pl)
* Filter VCF with maf, inbreeding coefficient and missing data with SelectVariants from GATK

<https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-0/org_broadinstitute_gatk_tools_walkers_variantutils_SelectVariants.php>

* Set to missing data all genotypes that do not satisfy to the coverage requirement (min depth, max depth and min depth allele) with a custome perl script (set\_missing\_genotypes\_with\_readdepth\_on\_vcf\_new.pl)
* Imputation of missing data and phasing with Beagle

<https://faculty.washington.edu/browning/beagle/beagle.html>

Convert phased VCF file in hapmap file with custom perl script (from\_beagle\_phased\_vcf\_file\_to\_hapmap\_file.pl)

Remarks: All the fixed parameters can be changed directly in the script.