

# transformation pipeline

H3Agwas : transformation of vcf in plink format

Sat Jul 9 19:45:07 SAST 2022

## **1 Introduction**

This report gives a brief overview of transformation vcf file from imputation in plink format.

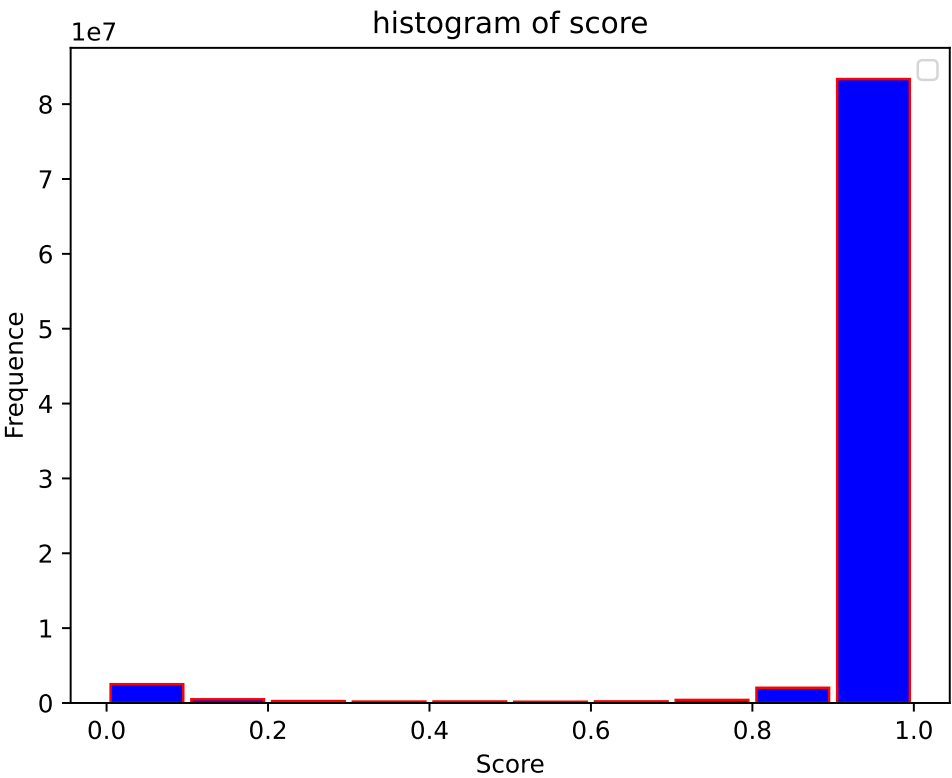
**Table 1** Table contains resume of score number ( $Sc$ ) higher than upper limit and frequencies ( $f$ ) not null by file

File	$f > 0$	$Sc > 0.6$	$Sc > 0.6 \text{ and } f > 0$	Total
22.pbwt_reference_impute.vcf.stat	830831	1111581	828177	1165064
21.pbwt_reference_impute.vcf.stat	850805	1135955	848775	1183101
19.pbwt_reference_impute.vcf.stat	1354029	1793624	1349398	1886100
14.pbwt_reference_impute.vcf.stat	2099453	2816929	2093317	2939523
20.pbwt_reference_impute.vcf.stat	1451051	1955950	1446747	2036437
17.pbwt_reference_impute.vcf.stat	1769010	2370050	1763417	2483649
15.pbwt_reference_impute.vcf.stat	1876787	2529600	1871442	2642129
18.pbwt_reference_impute.vcf.stat	1818318	2446804	1813937	2543884
16.pbwt_reference_impute.vcf.stat	2053373	2763399	2045247	2909149
13.pbwt_reference_impute.vcf.stat	2296092	3085243	2289323	3214114
12.pbwt_reference_impute.vcf.stat	3053375	4111389	3045163	4283159
9.pbwt_reference_impute.vcf.stat	2763078	3742474	2755629	3903677
10.pbwt_reference_impute.vcf.stat	3154340	4238523	3146387	4411996
11.pbwt_reference_impute.vcf.stat	3184906	4275739	3173067	4483216
6.pbwt_reference_impute.vcf.stat	3962862	5313190	3948421	5560553
7.pbwt_reference_impute.vcf.stat	3673810	4938142	3662614	5166355
8.pbwt_reference_impute.vcf.stat	3673908	4967200	3661800	5191483
5.pbwt_reference_impute.vcf.stat	4206969	5679381	4193906	5924338
4.pbwt_reference_impute.vcf.stat	4540013	6081660	4525781	6350604
3.pbwt_reference_impute.vcf.stat	4644141	6284902	4630203	6561776
1.pbwt_reference_impute.vcf.stat	5035197	6791356	5019430	7084113
2.pbwt_reference_impute.vcf.stat	5595851	7576057	5579562	7913667
Total	63888199	86009148	63691743	7913667

Resume by files format, SNPs number higher than the score higher than 0.6 and with frequencies more than 0

The score distribution plot can be found in Figure 1. The frequencies distribution total be found in Figure 2.

**Figure 1** score distribution



**Figure 2** Distribution of frequencies

