

QC-Report of leu-gda-eur-lt

Psychiatric GWAS Consortium

March 3, 2023

Abstract

This is an automatic output of the QC-Step of the ricopili-pipeline, created at MGH, April/Mai 2008. It is now in version 8.0.0. It is supposed to check and clean a GWAS dataset for technical problems and/or uncontrollable population stratification.

This script is still under construction, so please be patient with formatting problems. If you have some ideas, questions, requests, feel free to write me.

Stephan Ripke, MGH, Boston; sripke (at) broadinstitute org

Contents

1	General Info	2
1.1	Size of Sample	2
1.2	Exclusion overview	2
2	Manhattan	3
2.1	Manhattan-Plot - pre-QC (QQplot with MAF 0.02)	3
2.2	Manhattan-Plot - post-QC (QQplot with MAF 0.02)	4
2.3	Manhattan-Plot - post-QC, GC-corrected	5
3	Per Individual Characteristics Analysis	6
3.1	Missing Rates - pre-QC	6
3.2	Missing Rates - post-QC	7
3.3	Fhet - post QC	8
3.4	Fhet - zoomed	9

4	Per SNP Characteristics Analysis	10
4.1	pre-QC missing rate	10
4.2	Lambda-Plot to various variables (all post-QC)	11
5	Top-Plots for SNPs	21
5.1	Minor Allele Frequency	21
5.2	Missing Rate	22
5.3	Missing - Difference (case-control)	23
5.4	Significance of Missing - Difference (case-control)	24
5.5	Missing Rate - affected	25
5.6	Missing Rate - unaffected	26
5.7	Hardy Weinberg Equilibrium - unaffected (exact test)	27
5.8	"Minor" Allele Frequency - affected	28
5.9	"Minor" Allele Frequency - unaffected	29

1 General Info

1.1 Size of Sample

Test	pre QC	post QC	exclusion-N
Cases,Controls,Missing	193,427,0	110,245,0	83,182,0
Males,Females,Unspec	322,206,92	187,82,86	135,124,6
SNPs	1904599	1014902	889697 (46.7%)

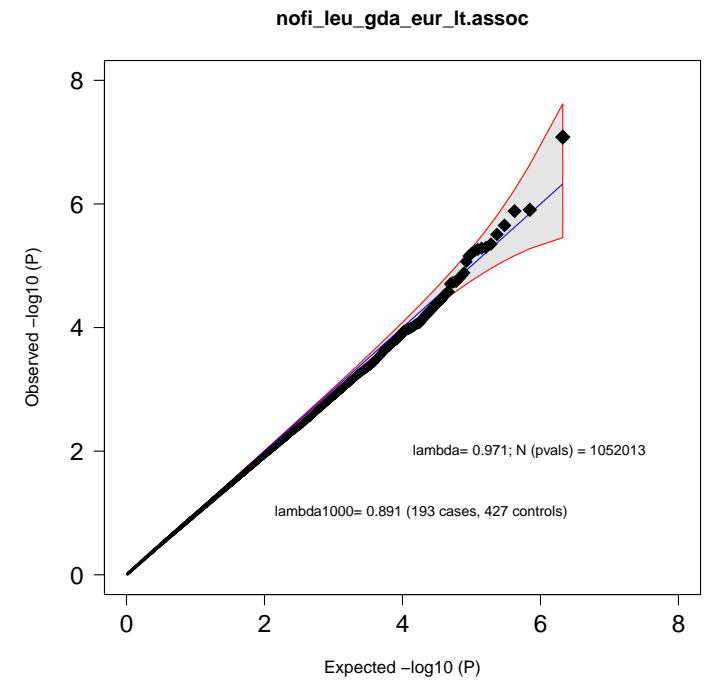
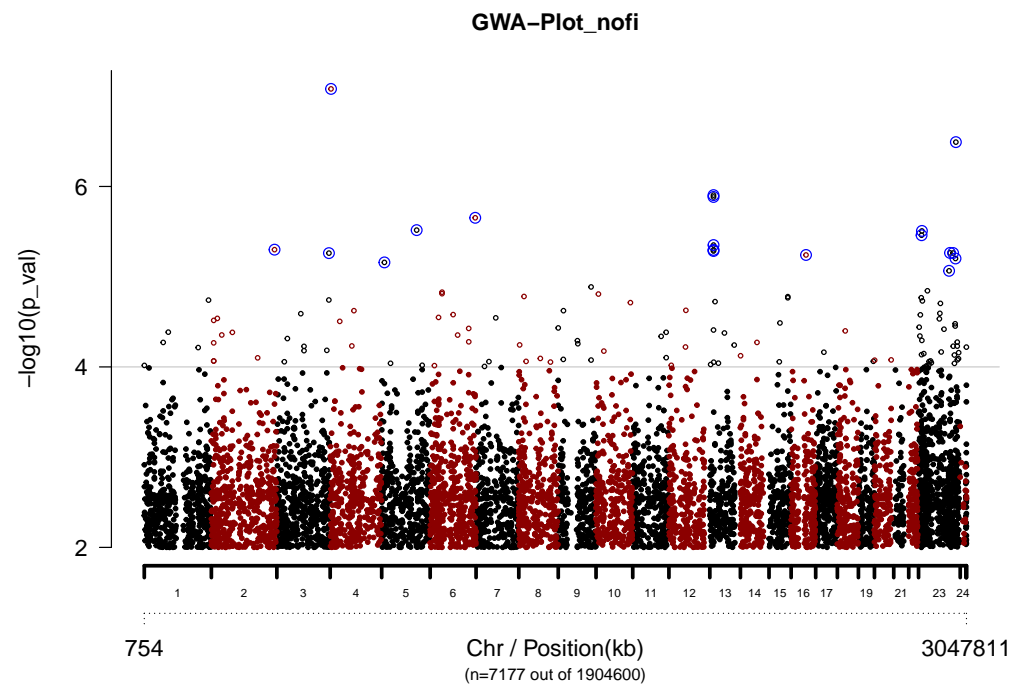
1.2 Exclusion overview

- would have excluded 362 individuals without pre-filter (SNP-Missing 0.05)

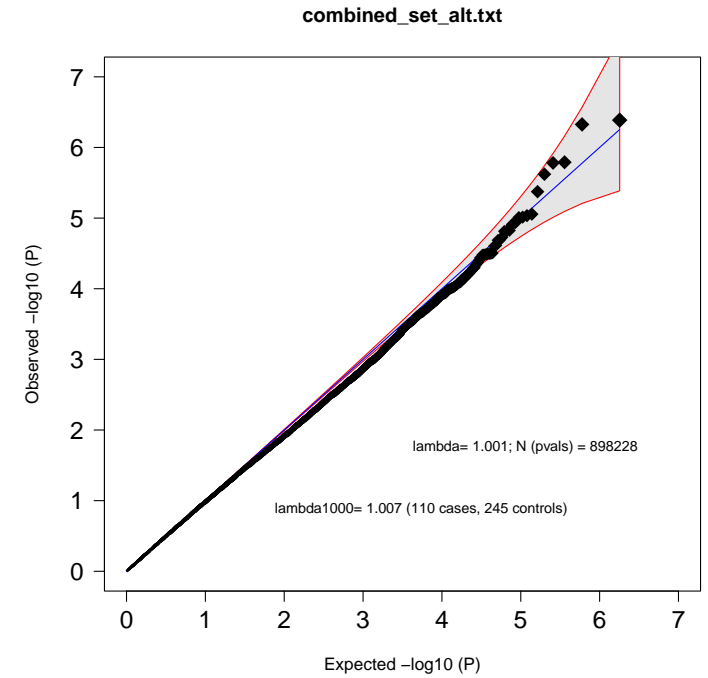
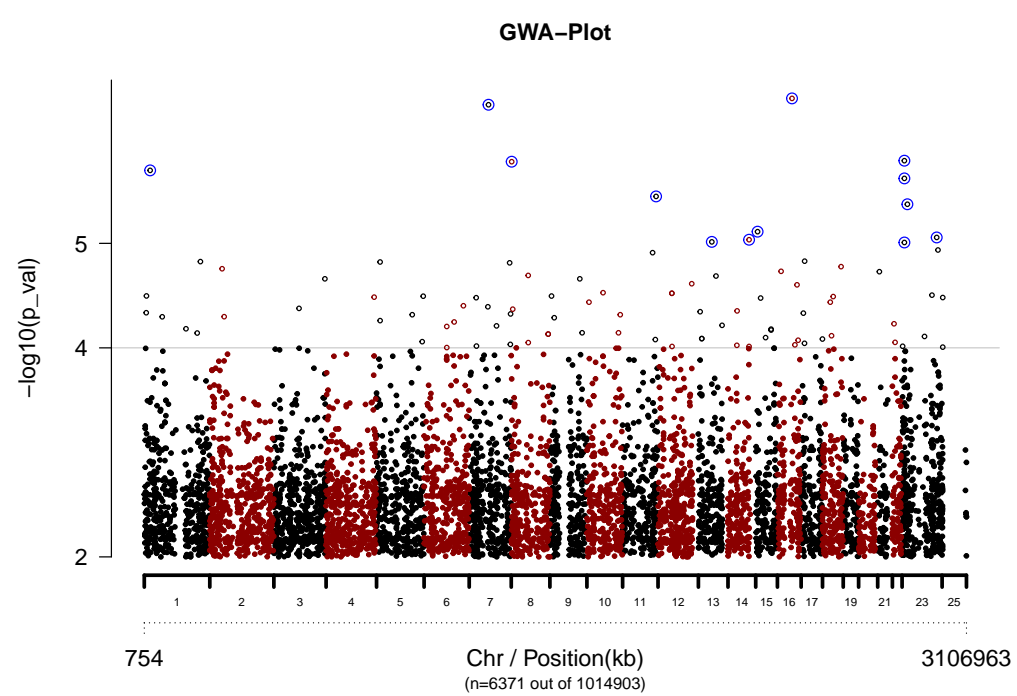
Filter	N
SNPs: call rate < 0.950 (pre - filter)	164727 (8.6%)
IDs: call rate (cases/controls) < 0.980	42 (10/32)
IDs: FHET outside +- 0.20 (cases/controls)	12 (2/10)
IDs: Sex violations -excluded- (N-tested)	235 (620)
IDs: Sex warnings (undefined phenotype / ambiguous genotypes)	96 (92/4)
SNPs: call rate < 0.980	50185 (2.6%)
SNPs: missing difference > 0.020	14641 (0.9%)
SNPs: without valid association p-value (invariant)	478434 (28.3%)
SNPs: with MAF < 0.010	663102 (39.2%)
SNPs: HWE-controls < -6	1174 (0.1%)
SNPs: HWE-cases < -10	4 (0.0%)
Warning: genomewide significant SNPs (autosomal/known)	0 (0/0)

2 Manhattan

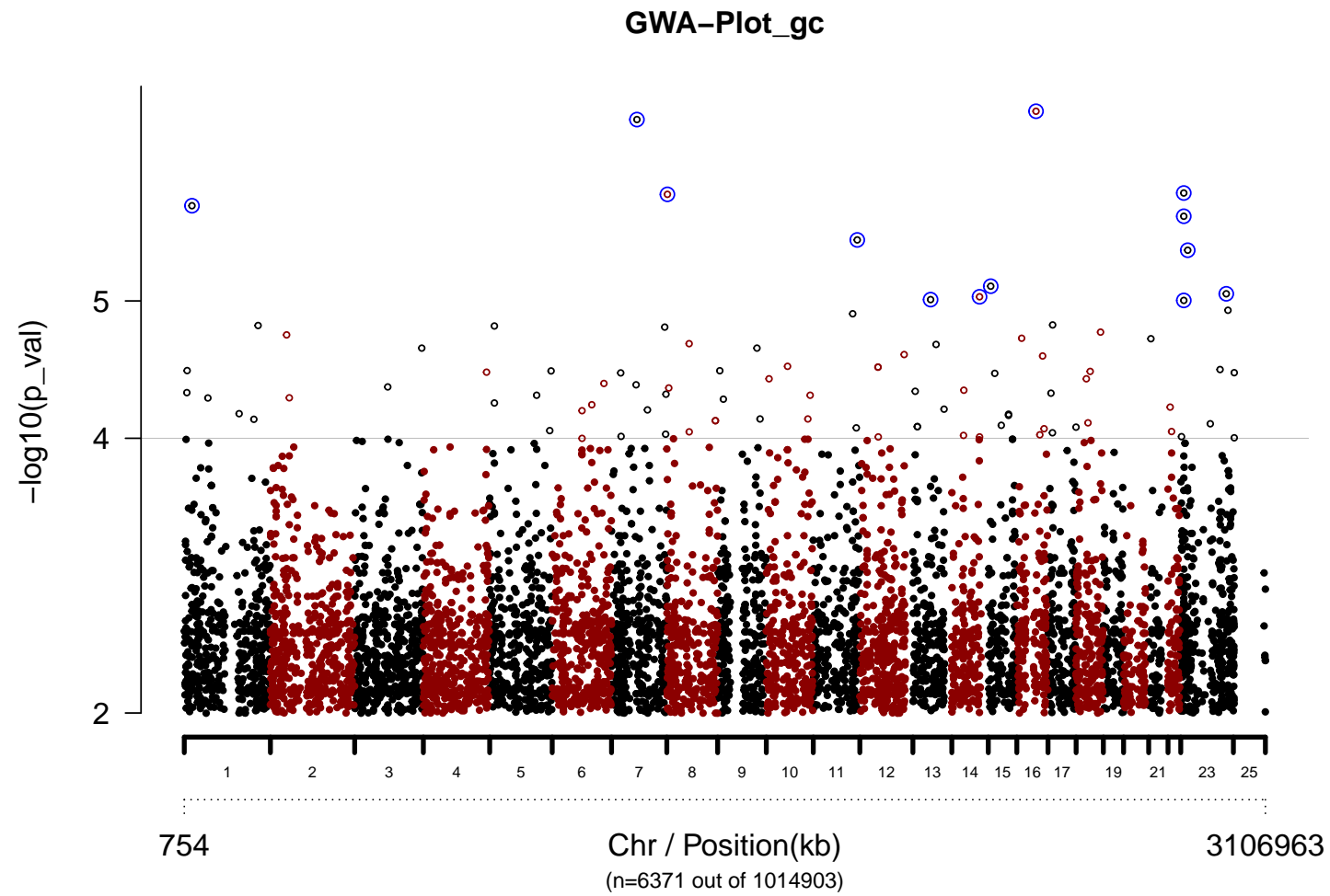
2.1 Manhattan-Plot - pre-QC (QQplot with MAF 0.02)



2.2 Manhattan-Plot - post-QC (QQplot with MAF 0.02)

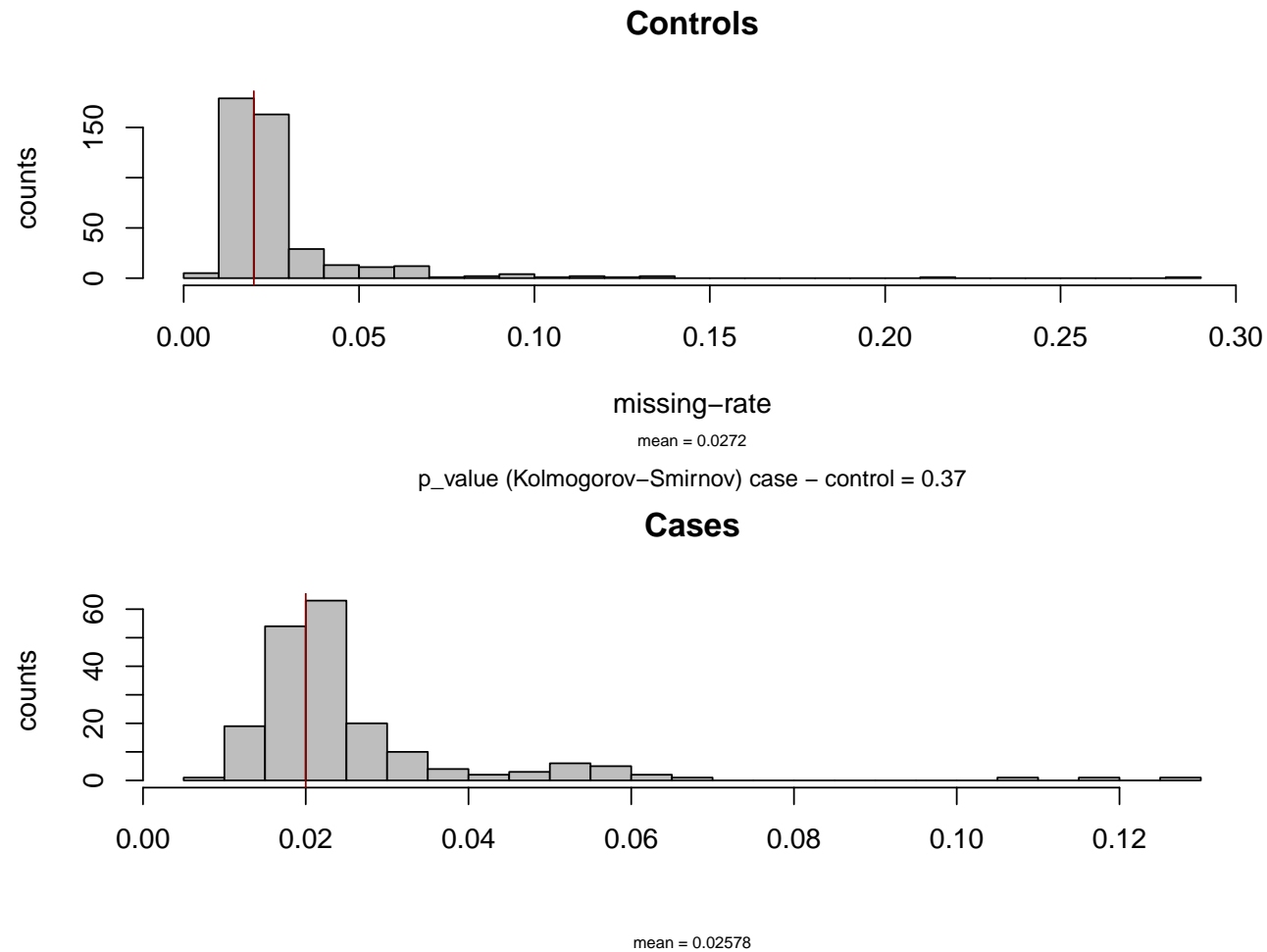


2.3 Manhattan-Plot - post-QC, GC-corrected

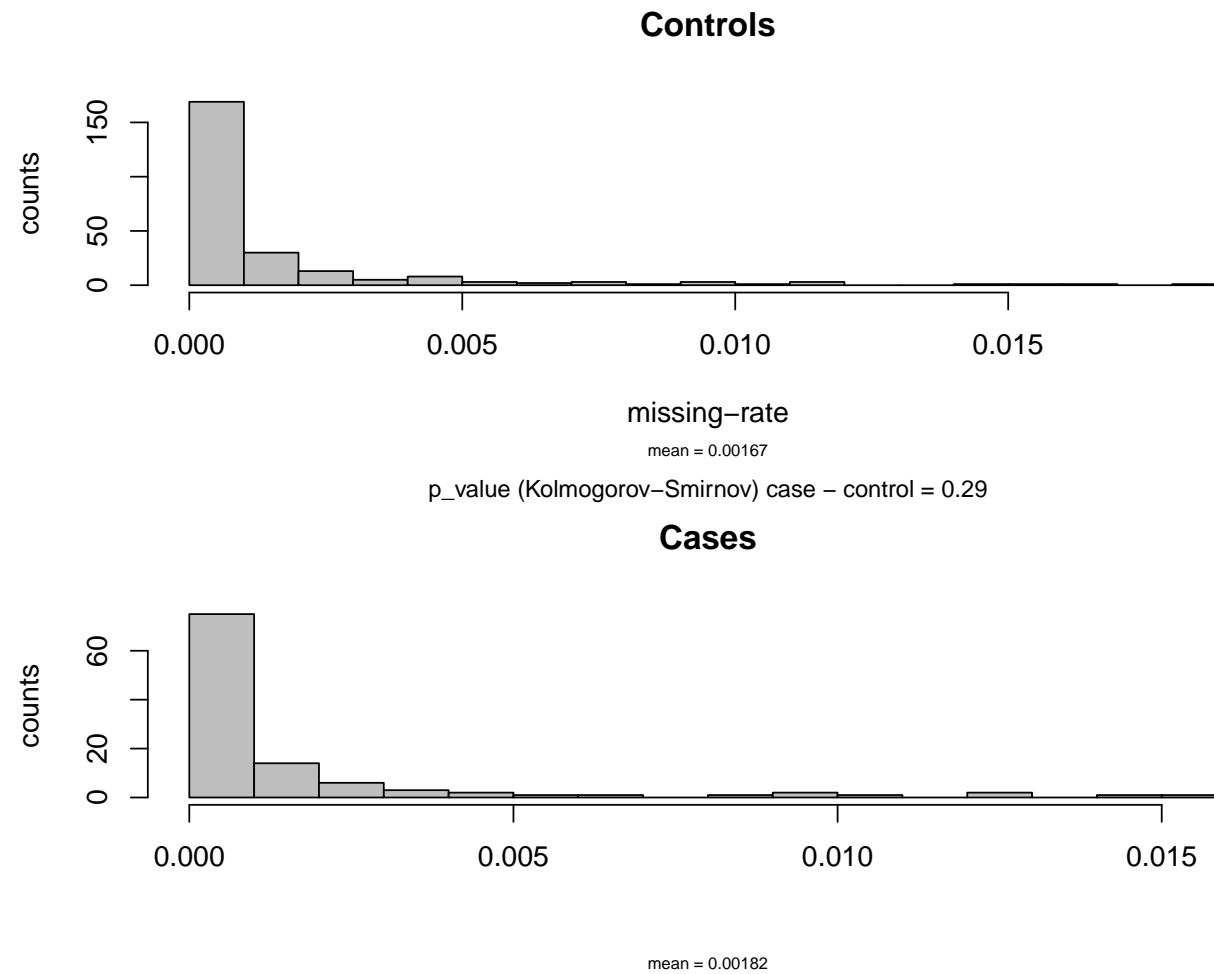


3 Per Individual Characteristics Analysis

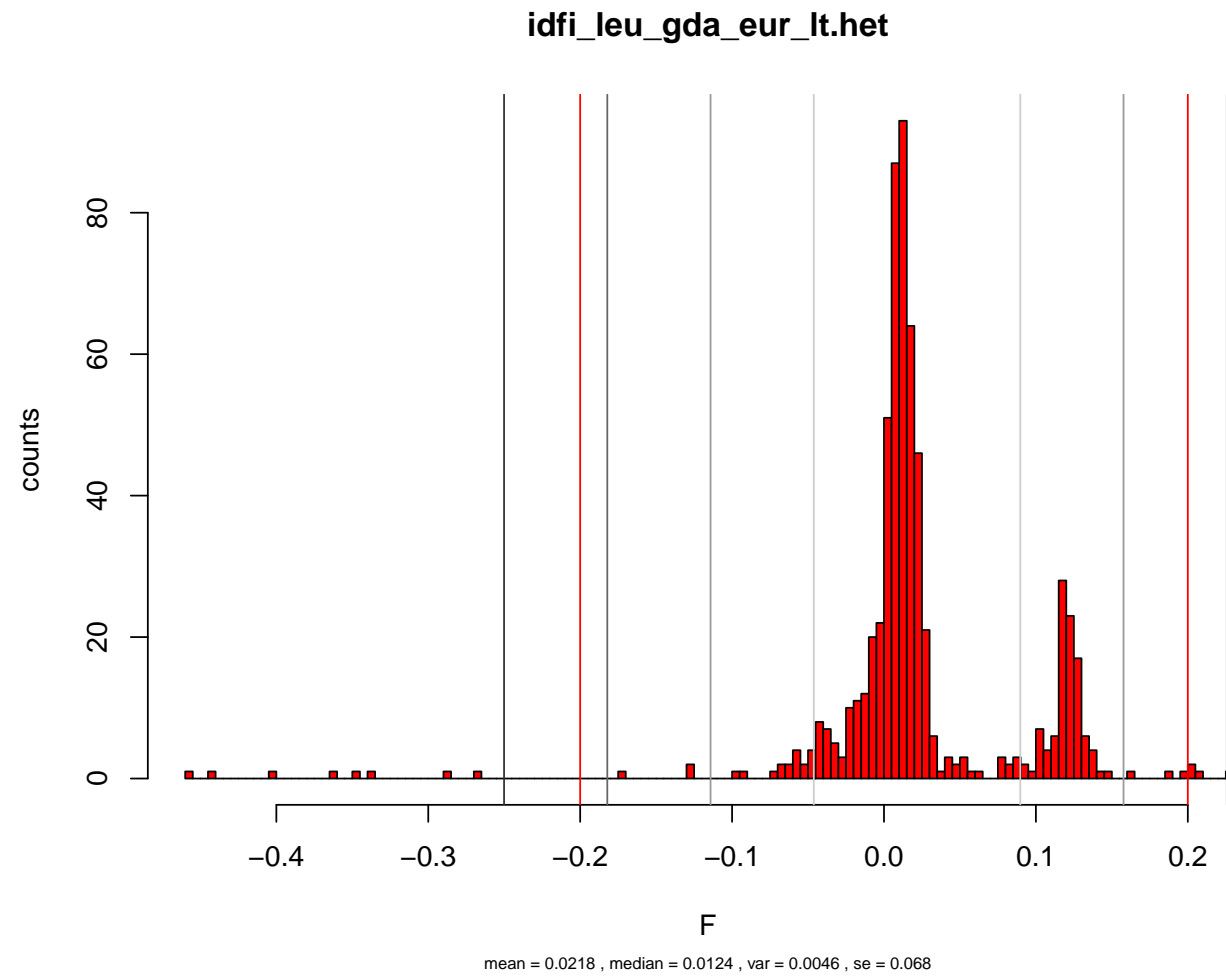
3.1 Missing Rates - pre-QC



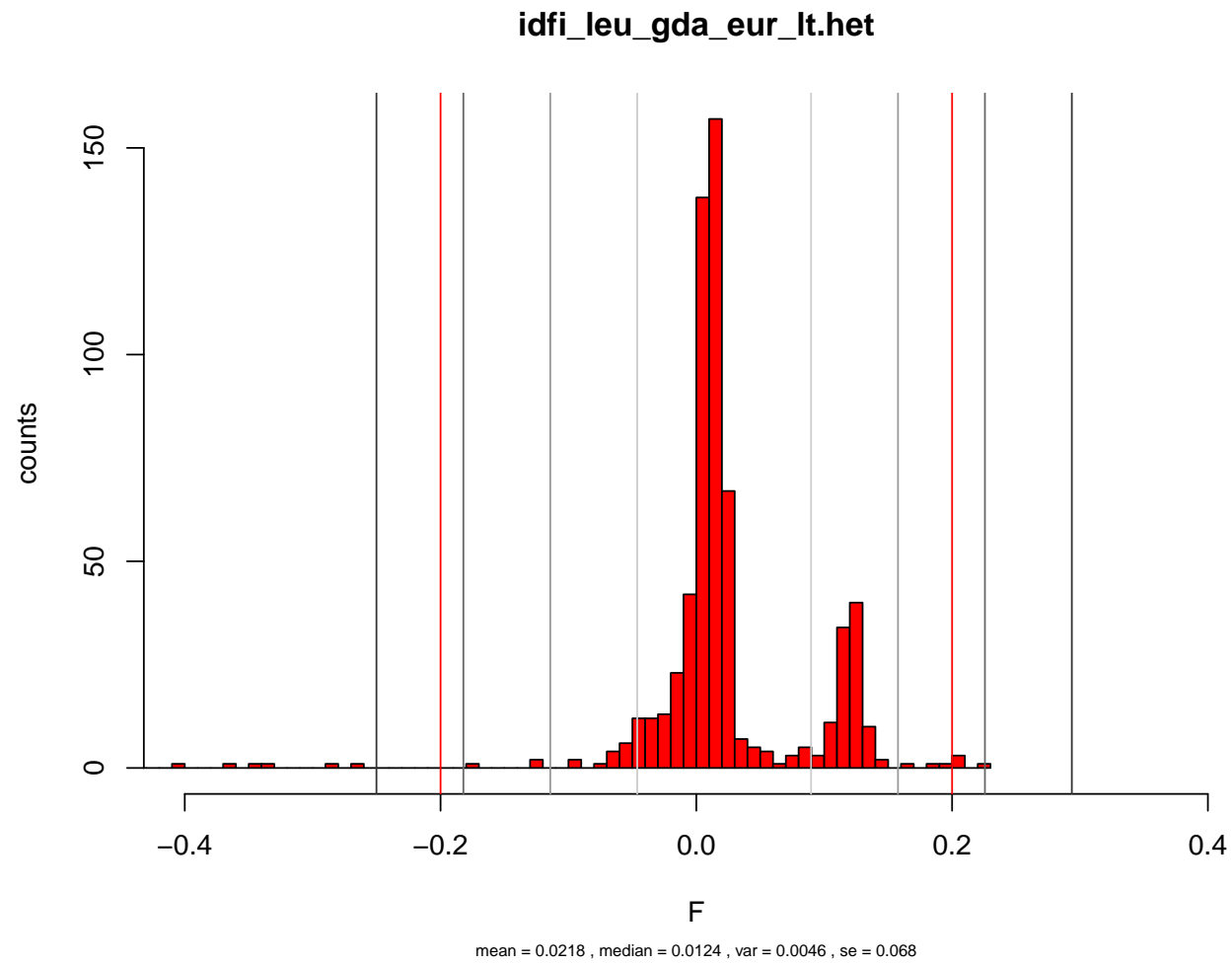
3.2 Missing Rates - post-QC



3.3 Fhet - post QC



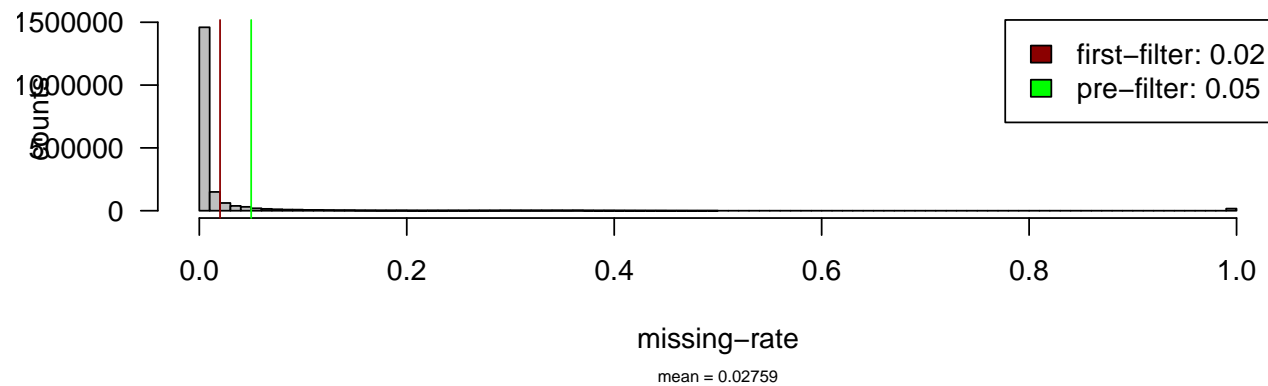
3.4 Fhet - zoomed



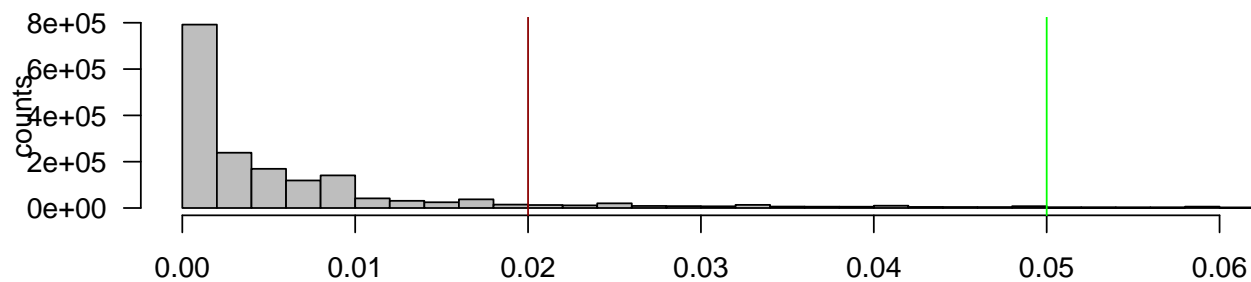
4 Per SNP Characteristics Analysis

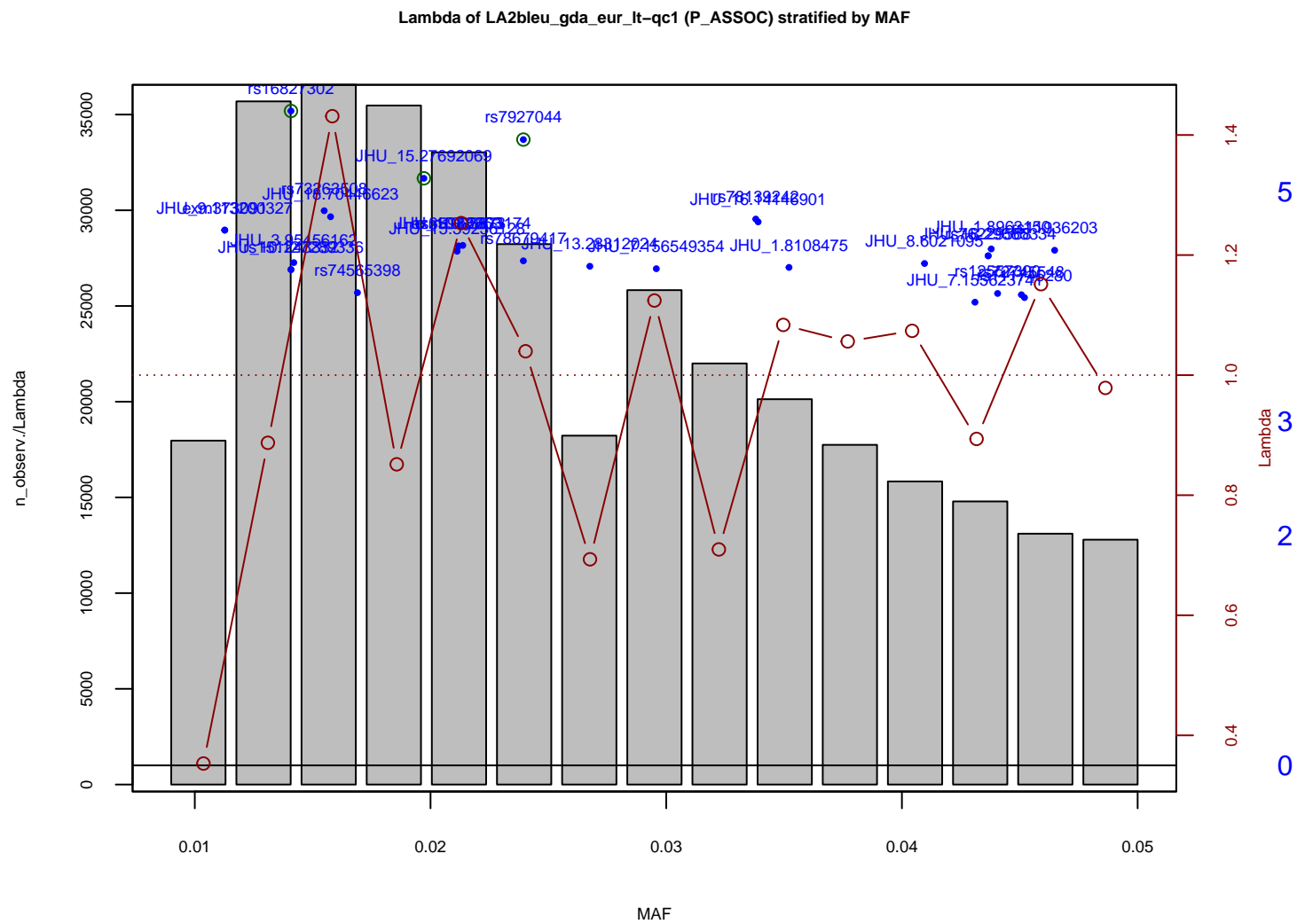
4.1 pre-QC missing rate

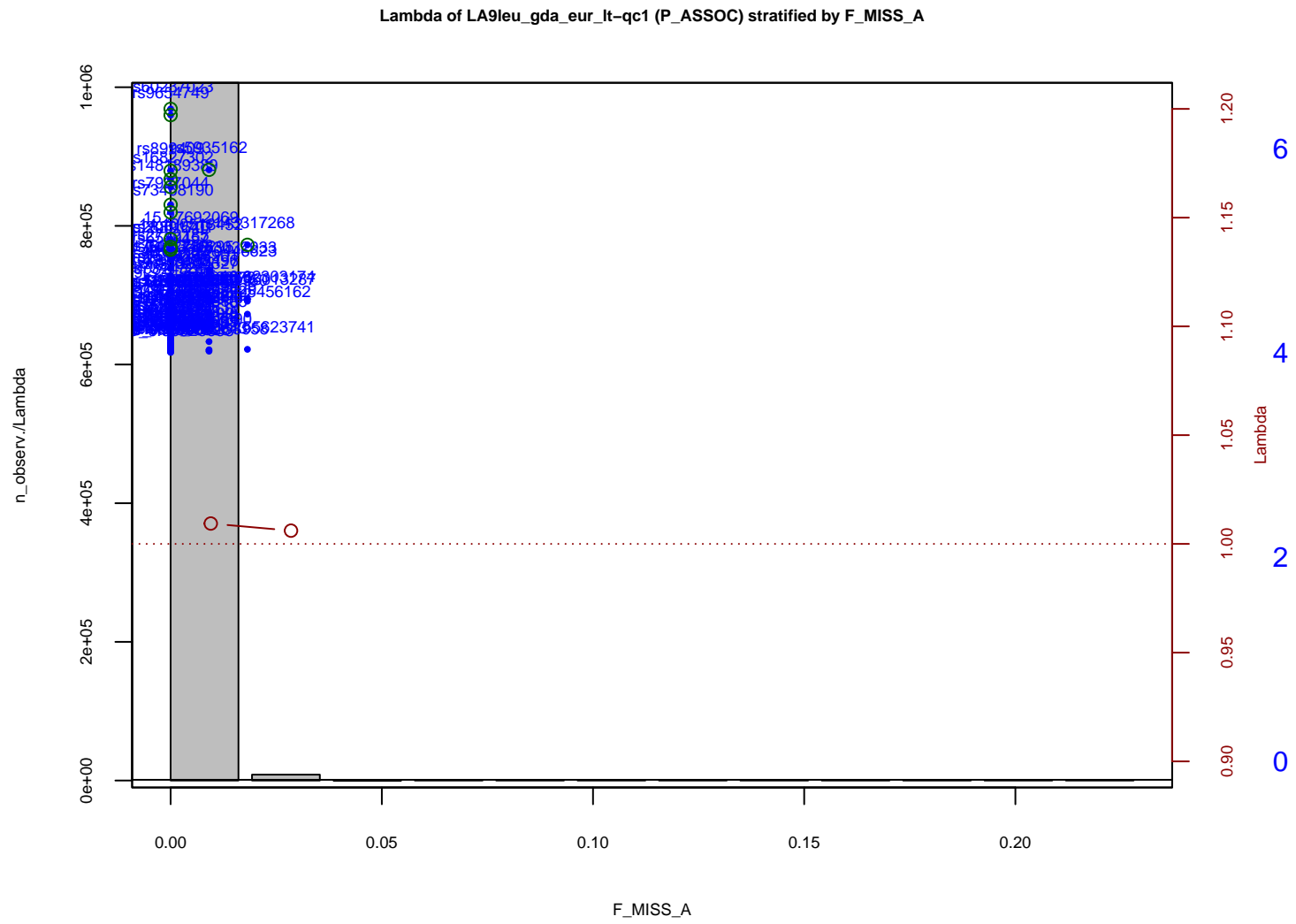
Histogram of SNP-Missing-Rate, all

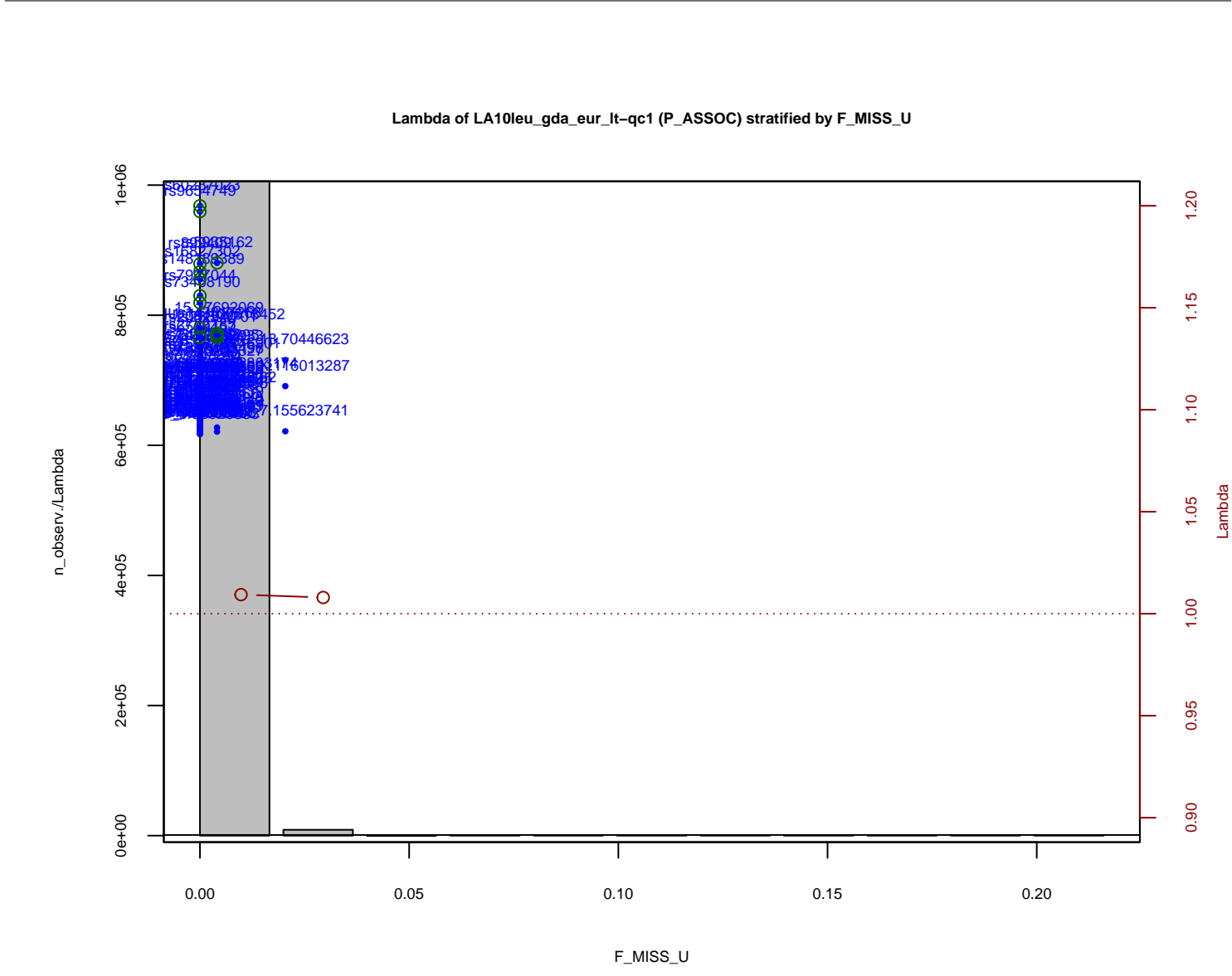


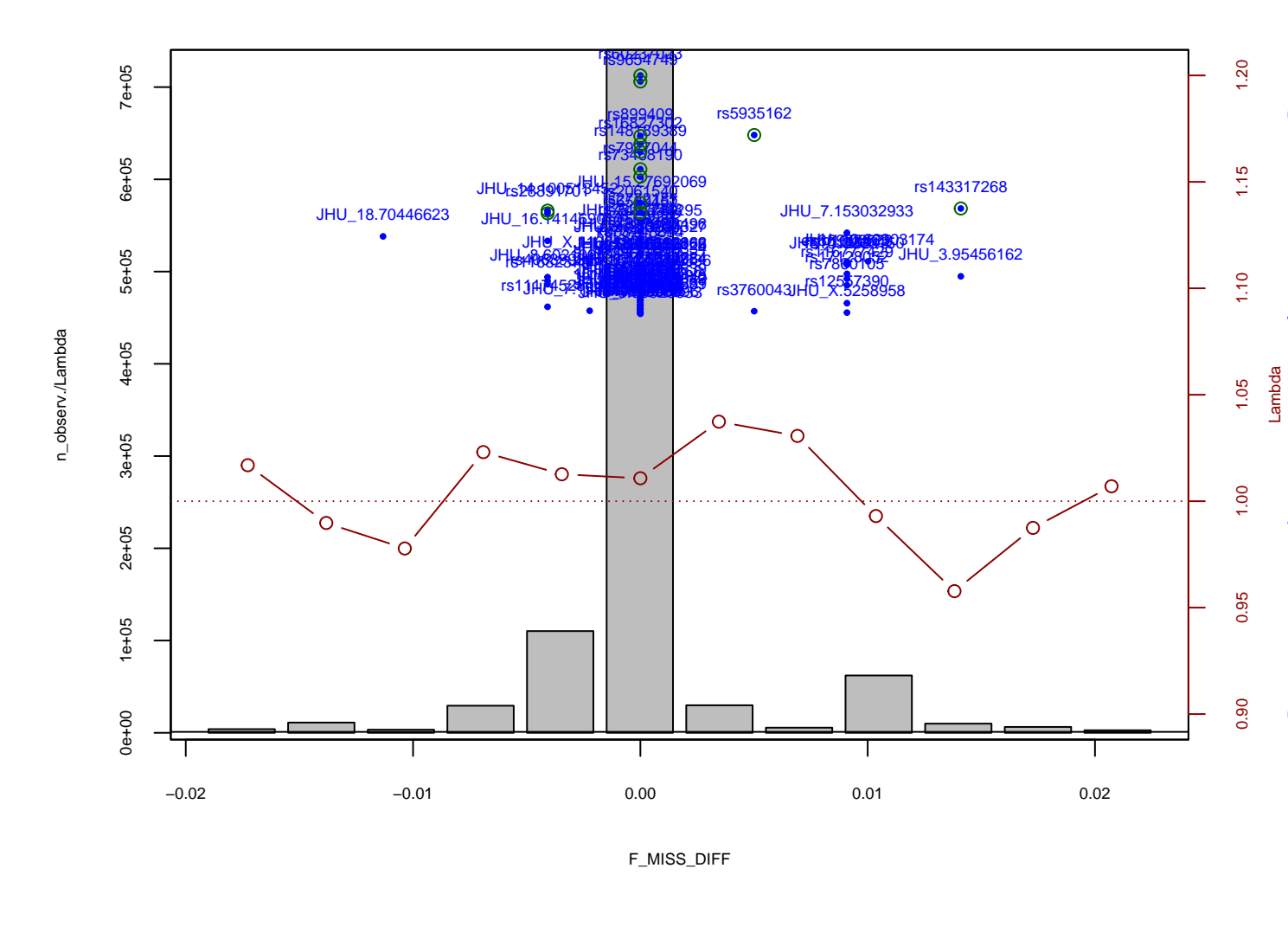
Histogram of SNP-Missing-Rate, zoomed

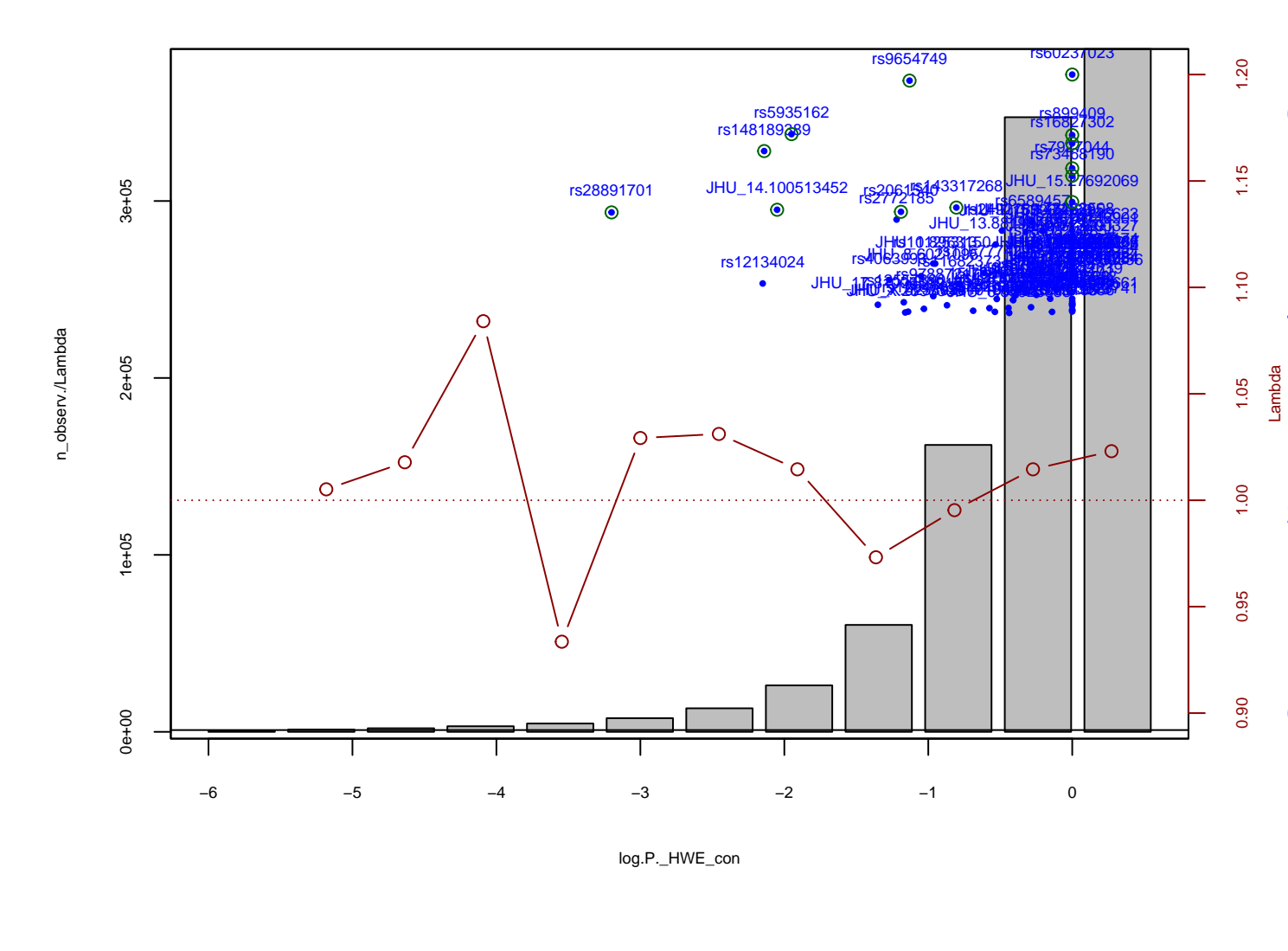


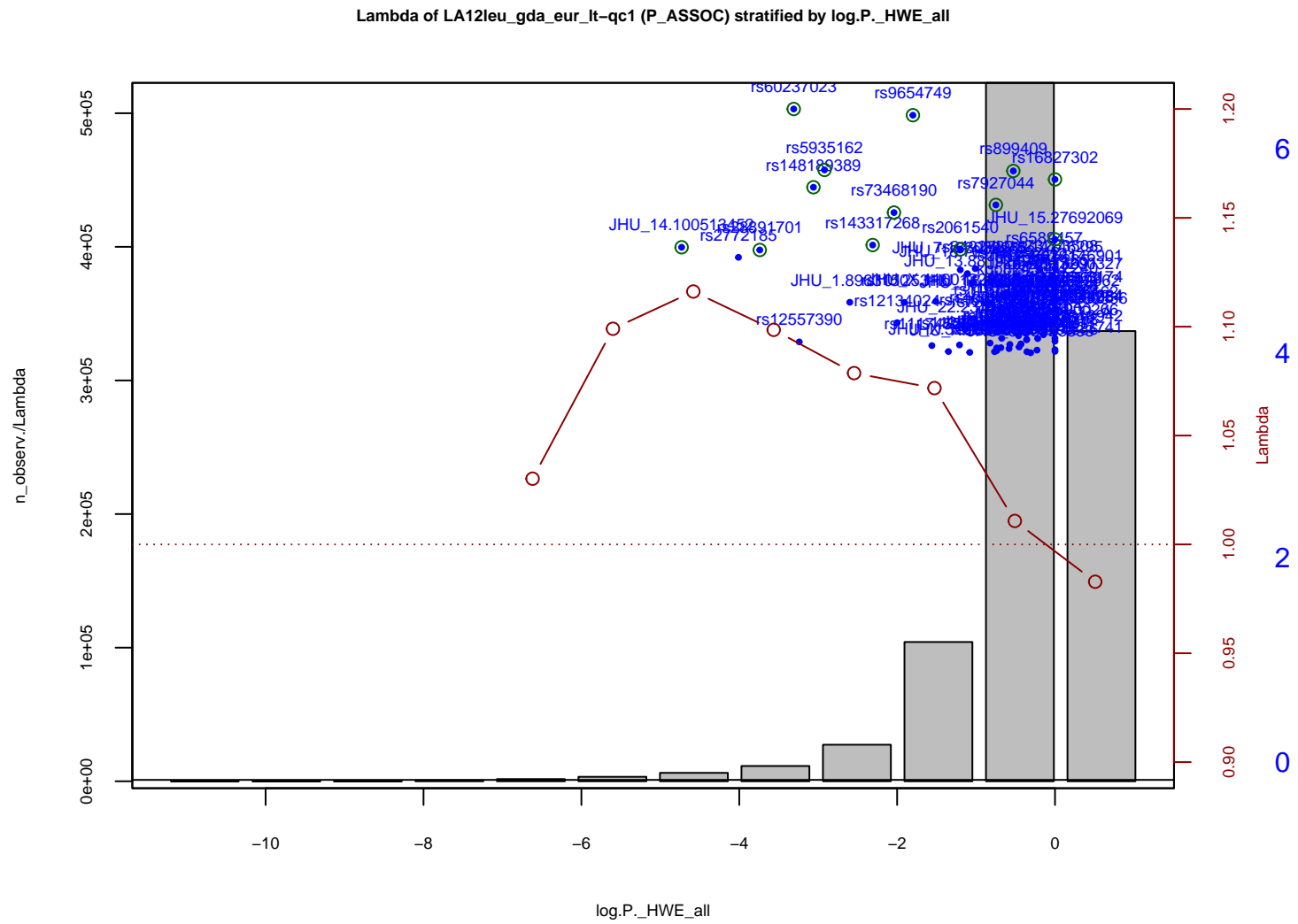










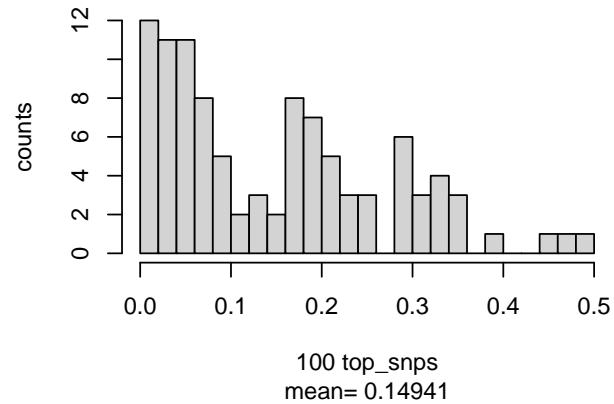




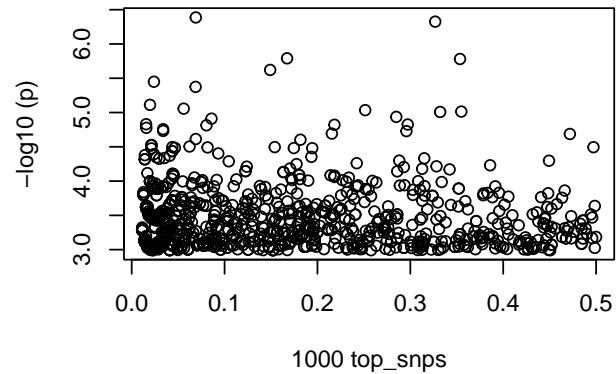
5 Top-Plots for SNPs

5.1 Minor Allele Frequency

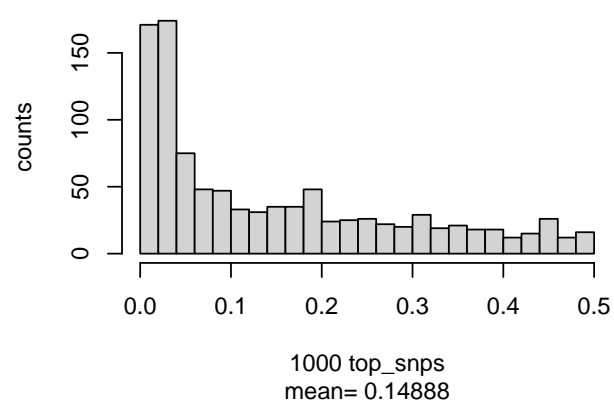
Histo of MAF – TOP100



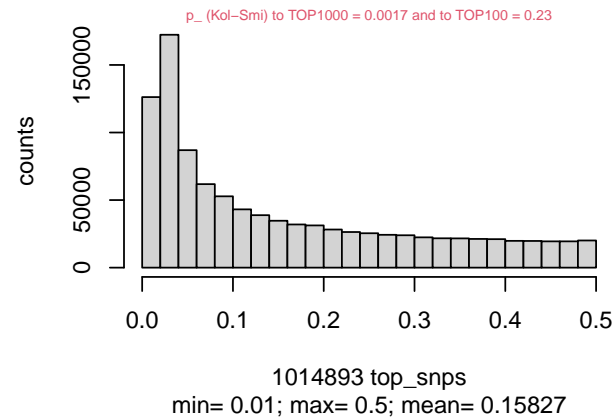
MAF – TOP1000



Histo of MAF – TOP1000

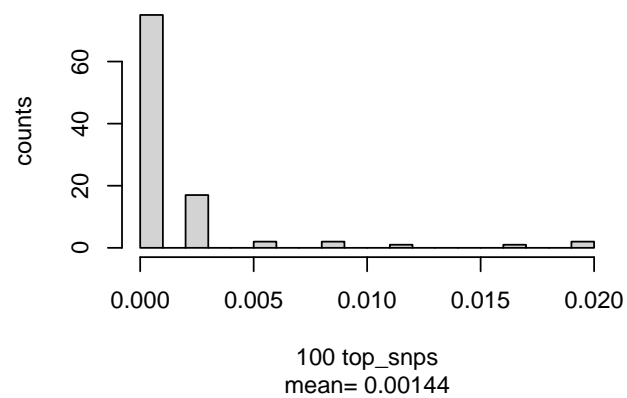


Histo of MAF – ALL

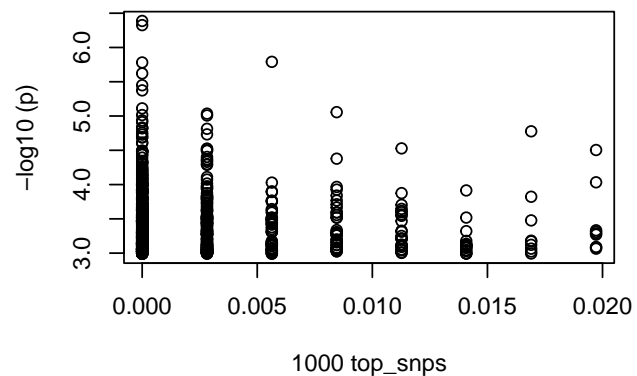


5.2 Missing Rate

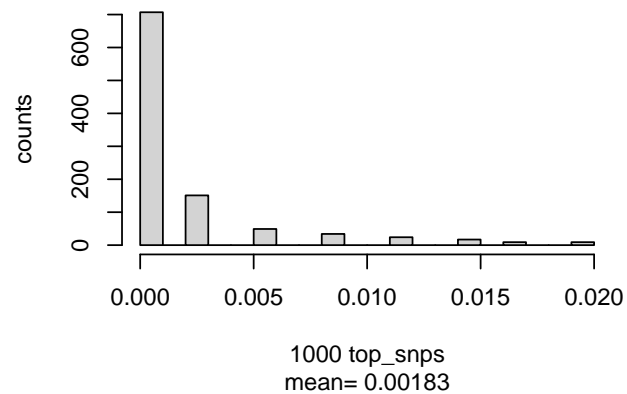
Histo of F_MISS – TOP100



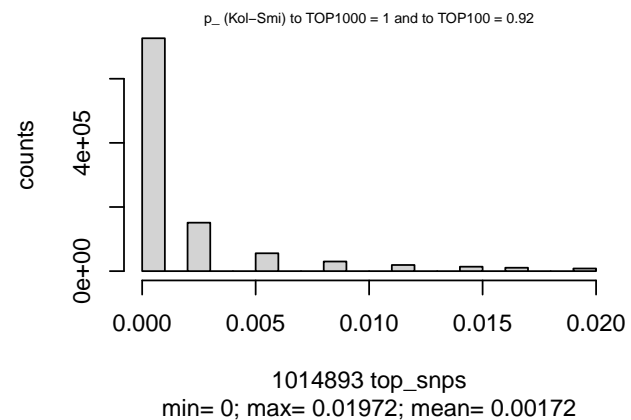
F_MISS – TOP1000



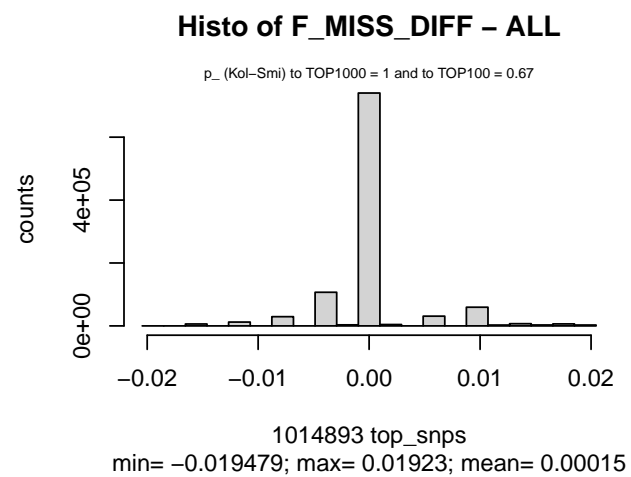
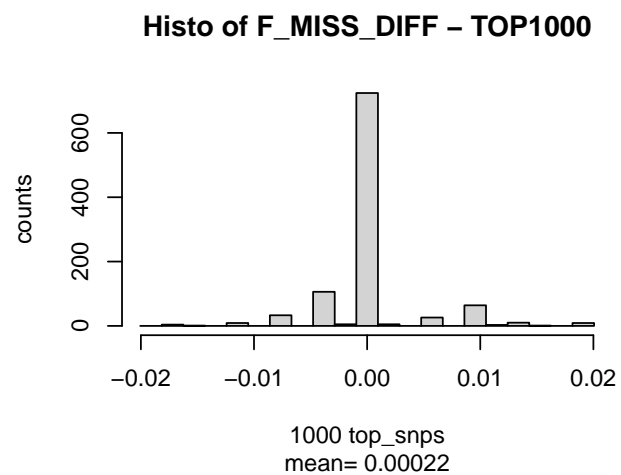
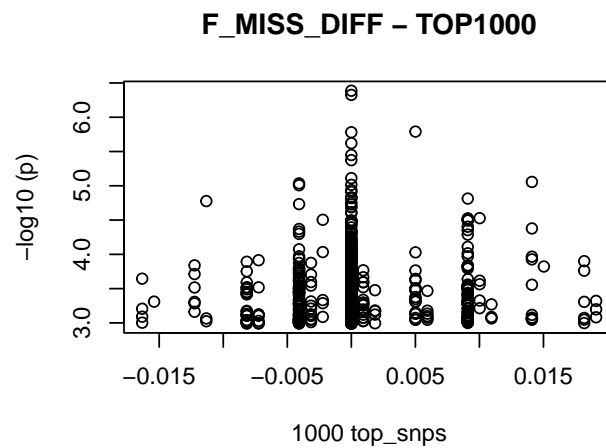
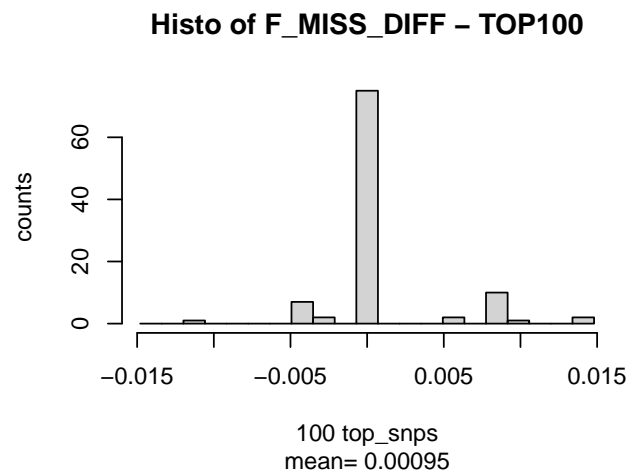
Histo of F_MISS – TOP1000



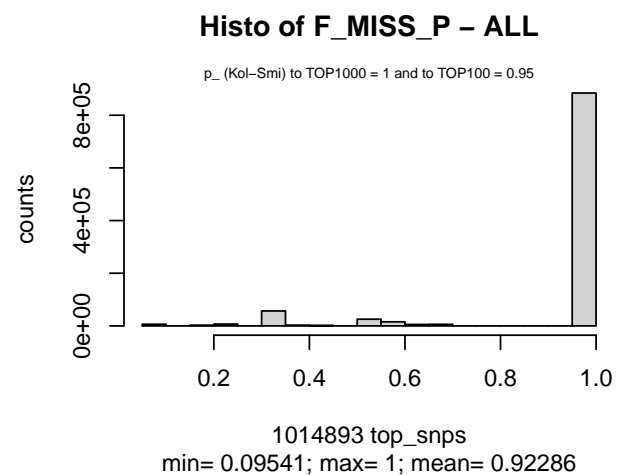
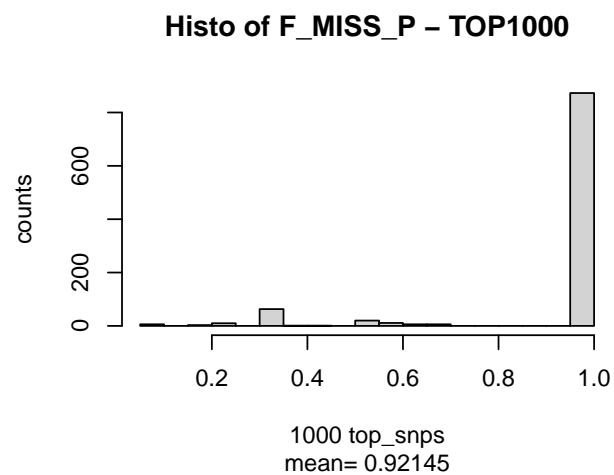
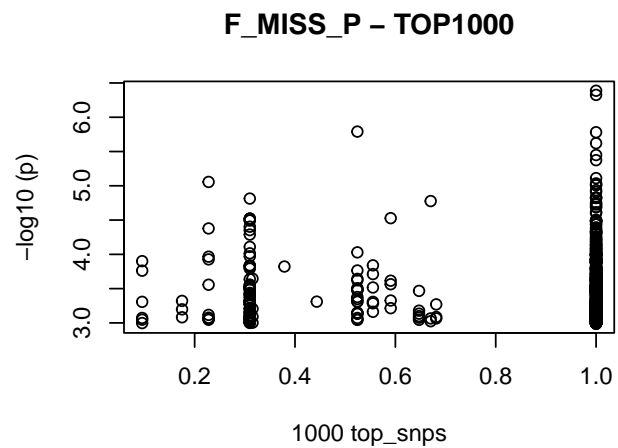
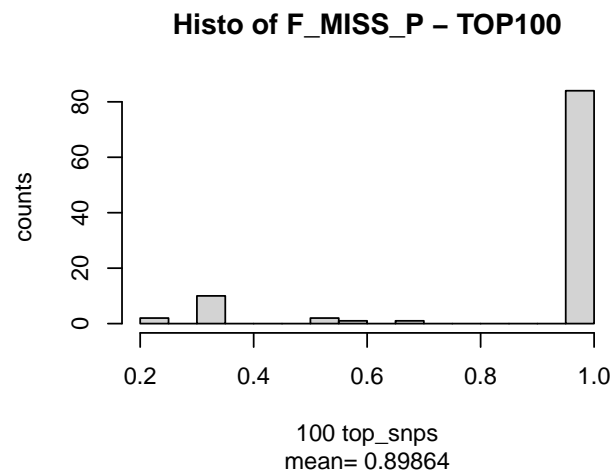
Histo of F_MISS – ALL



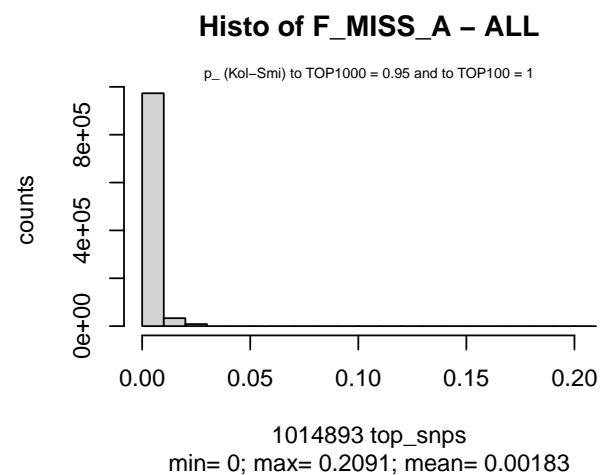
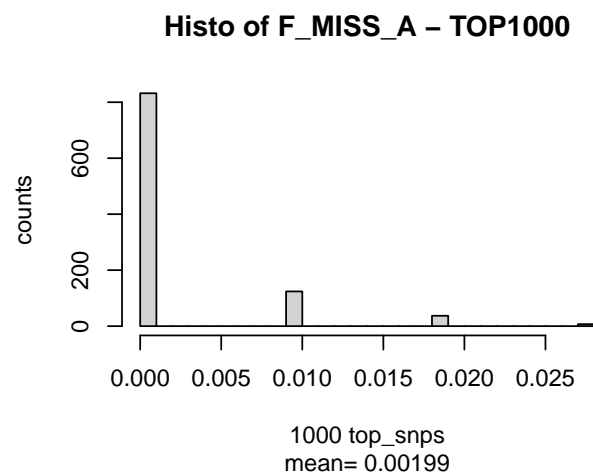
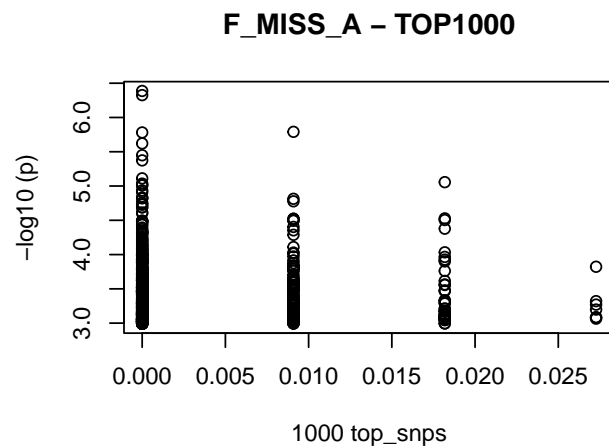
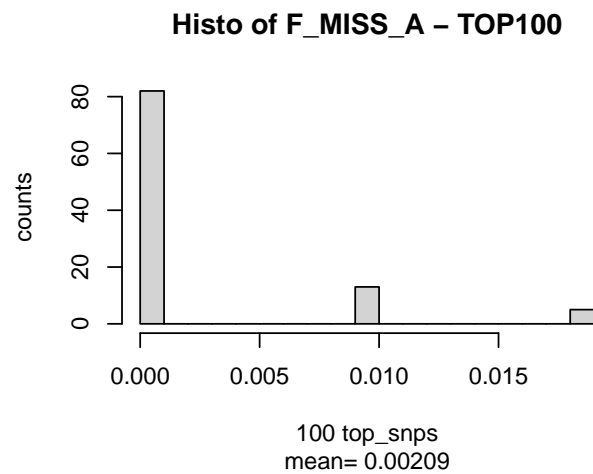
5.3 Missing - Difference (case-control)



5.4 Significance of Missing - Difference (case-control)

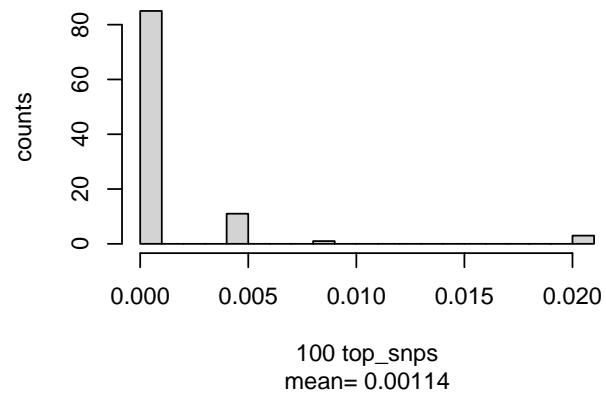


5.5 Missing Rate - affected

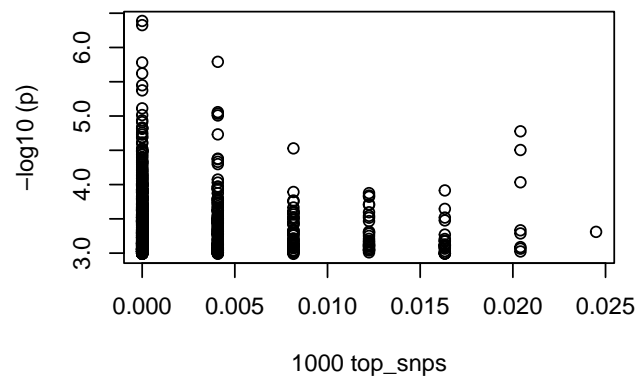


5.6 Missing Rate - unaffected

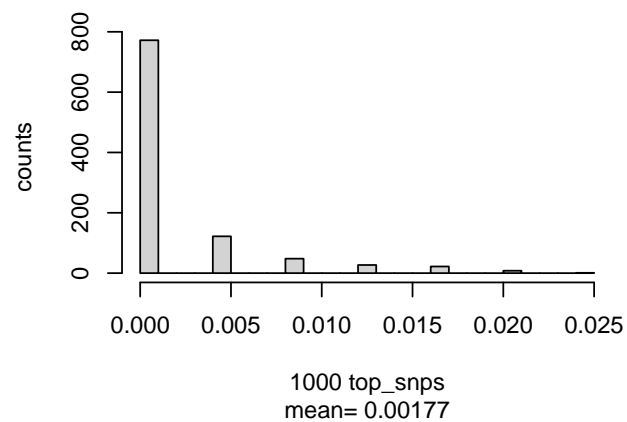
Histo of F_MISS_U – TOP100



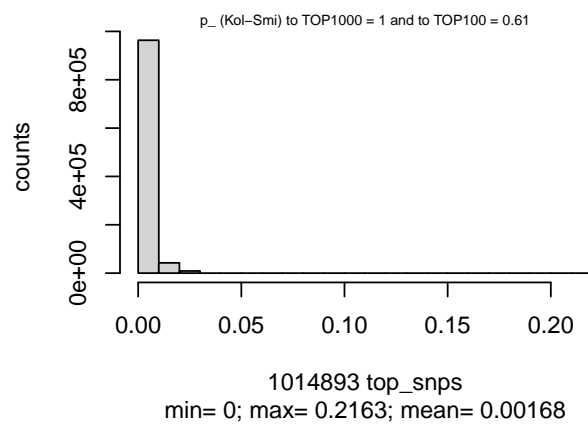
F_MISS_U – TOP1000



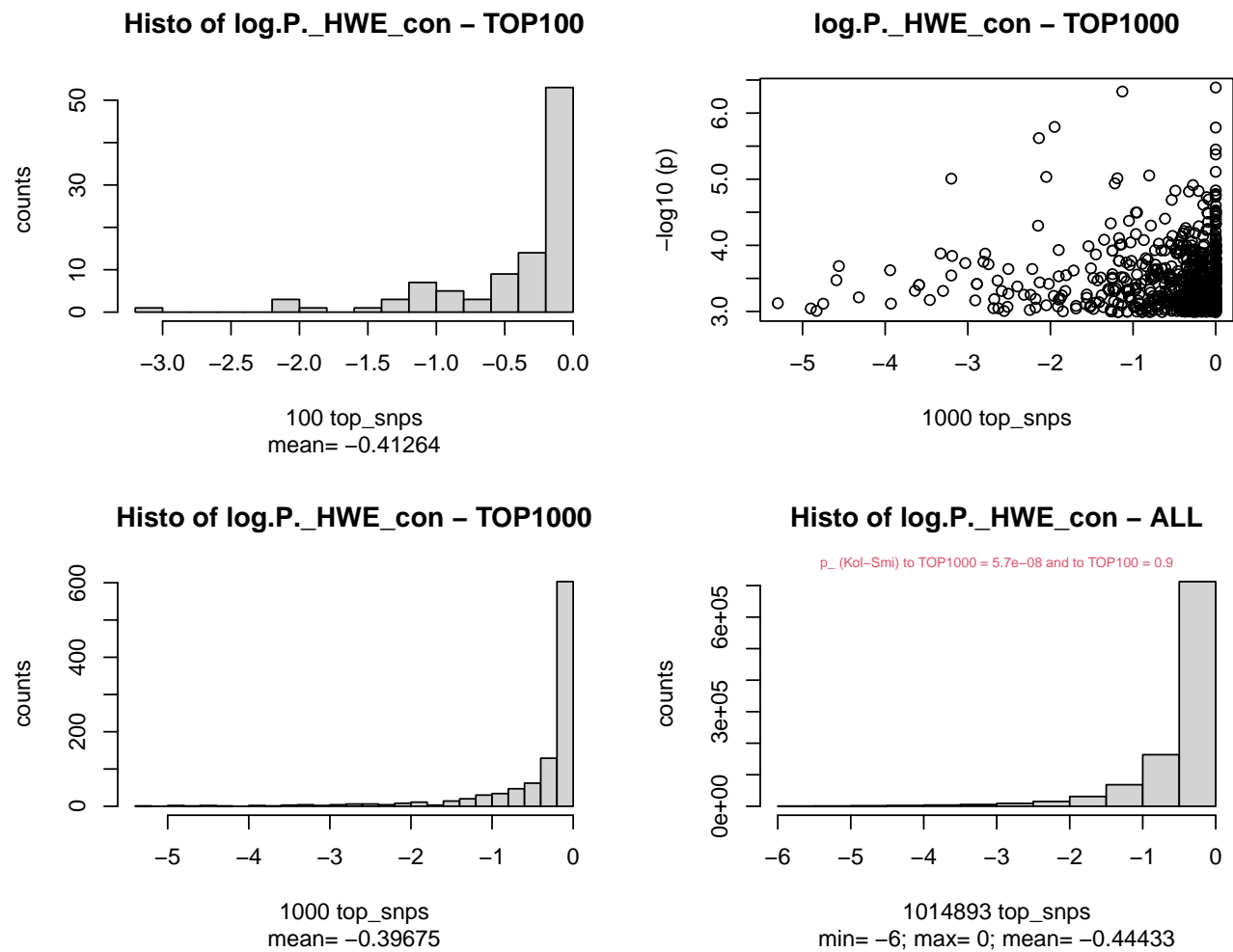
Histo of F_MISS_U – TOP1000



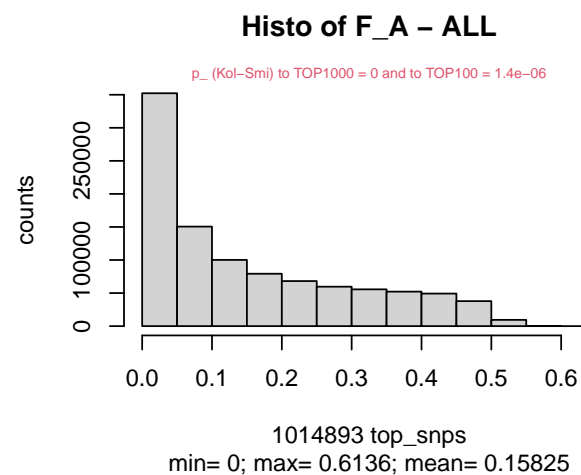
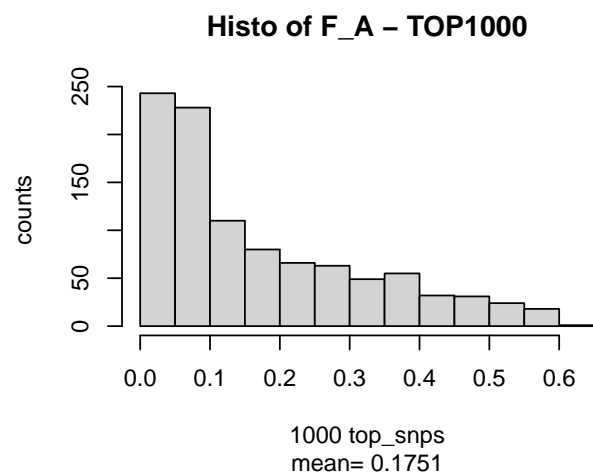
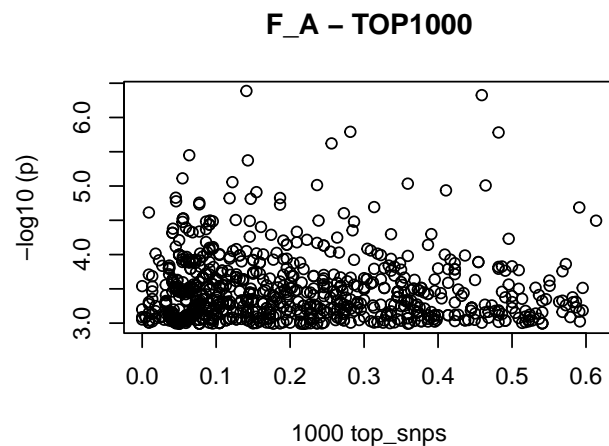
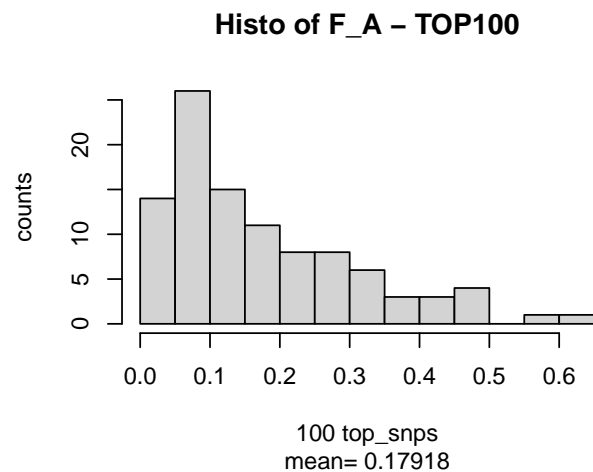
Histo of F_MISS_U – ALL



5.7 Hardy Weinberg Equilibrium - unaffected (exact test)

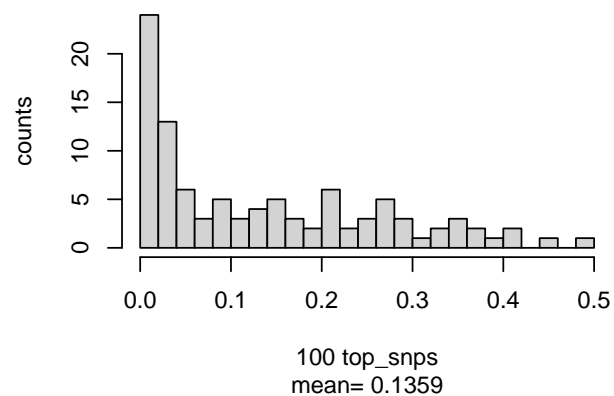


5.8 "Minor" Allele Frequency - affected

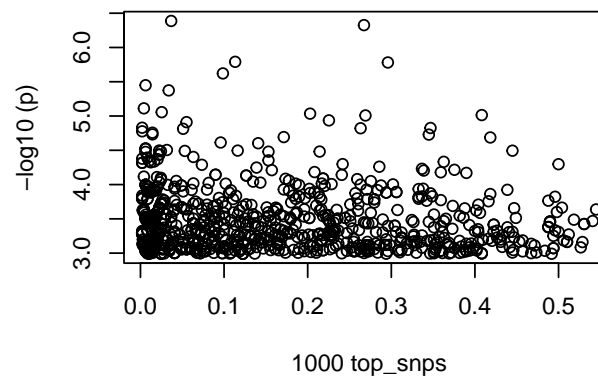


5.9 "Minor" Allele Frequency - unaffected

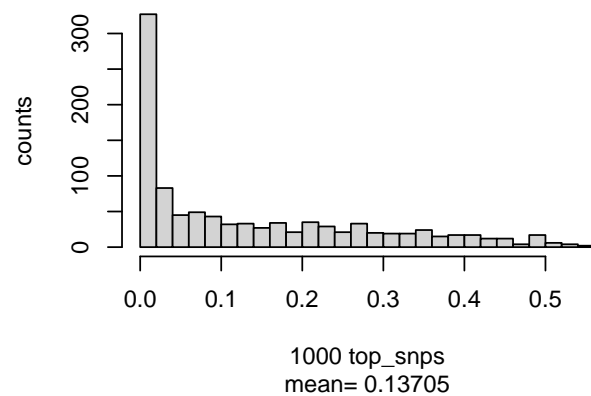
Histo of F_U – TOP100



F_U – TOP1000



Histo of F_U – TOP1000



Histo of F_U – ALL

