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	1.1	neral Info Size of Sample Exclusion overview	2 2 2					
2	Manhattan							
	2.1	Manhattan-Plot - pre-QC (QQplot with MAF 0.02)	3					
		Manhattan-Plot - post-QC (QQplot with MAF 0.02)						
	2.3	Manhattan-Plot - post-QC, GC-corrected	5					
3	Per Individual Characteristics Analysis							
	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					

leu-gda-eur-lt

4	4.1	SNP Characteristics Analysis pre-QC missing rate	
5	Тор	-Plots for SNPs	2
	5.1^{-}	Minor Allele Frequency	2
		Missing Rate	
	5.3	Missing - Difference (case-control)	23
	5.4	Significance of Missing - Difference (case-control)	24
		Missing Rate - affected	
	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	$\operatorname{pre} \operatorname{QC}$	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%)

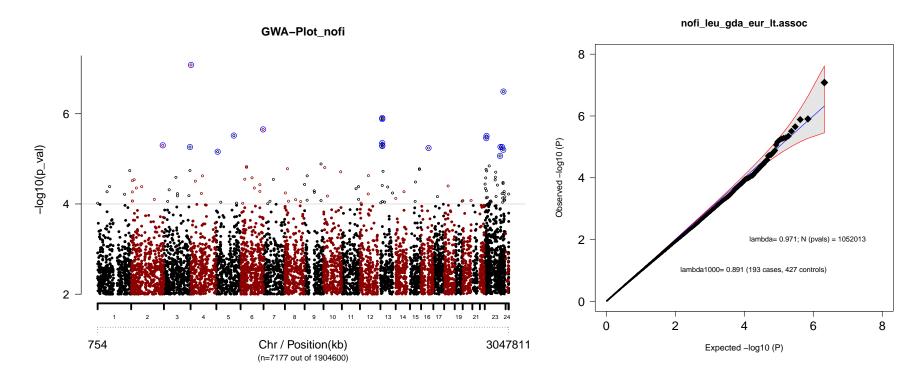
1.2 Exclusion overview

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

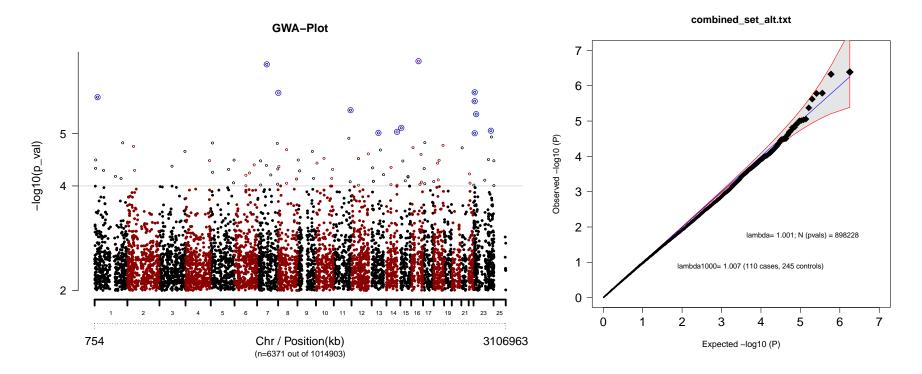
Filter	N
SNPs: call rate < 0.950 (pre - filter)	164727 (8.6%)
IDs: call rate (cases/controls) < 0.980	42 (10/32)
IDs: FHET outside +- 0.20 (cases/controls)	12 (2/10)
IDs: Sex violations -excluded- (N-tested)	235 (620)
IDs: Sex warnings (undefined phenotype / ambiguous genotypes)	96 (92/4)
SNPs: call 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)

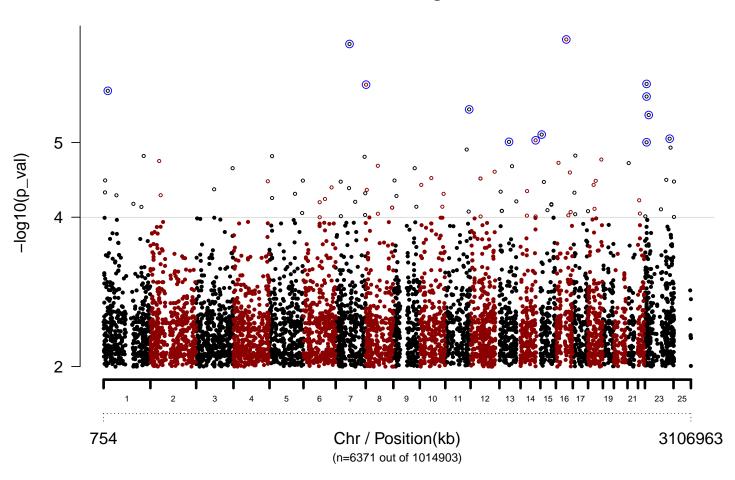


2.2 Manhattan-Plot - post-QC (QQplot with MAF 0.02)



${\bf 2.3}\quad {\bf Manhattan\text{-}Plot - post\text{-}QC, \, GC\text{-}corrected}$

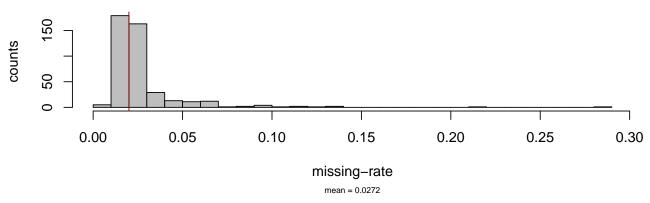




3 Per Individual Characteristics Analysis

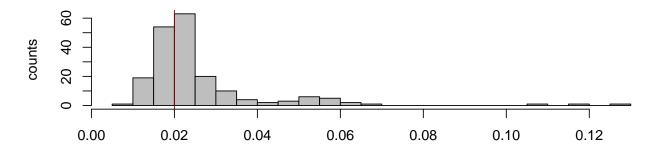
3.1 Missing Rates - pre-QC

Controls



p_value (Kolmogorov-Smirnov) case - control = 0.37

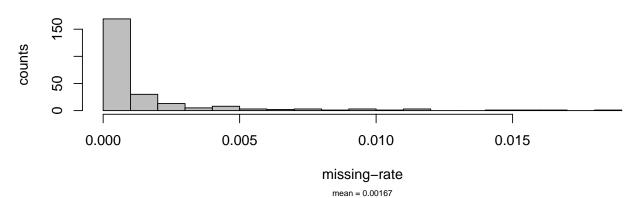
Cases



mean = 0.02578

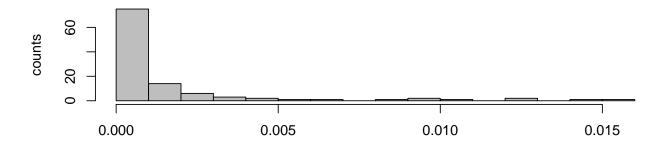
3.2 Missing Rates - post-QC

Controls



p_value (Kolmogorov–Smirnov) case - control = 0.29

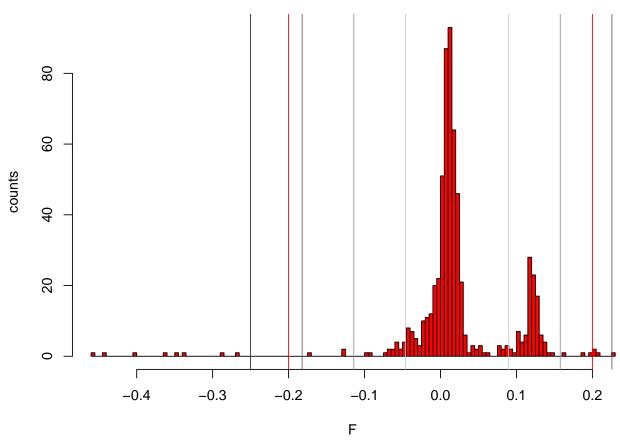
Cases



mean = 0.00182

3.3 Fhet - post QC

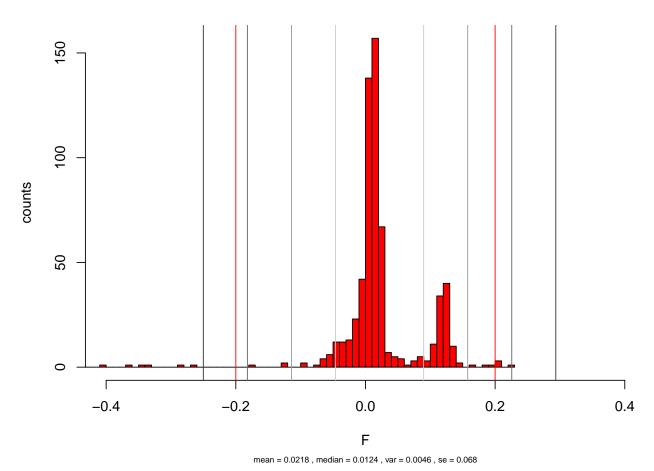
idfi_leu_gda_eur_lt.het



mean = 0.0218 , median = 0.0124 , var = 0.0046 , se = 0.068

3.4 Fhet - zoomed

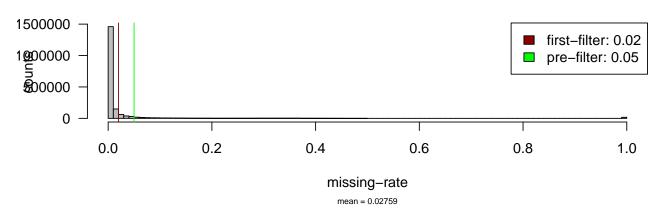




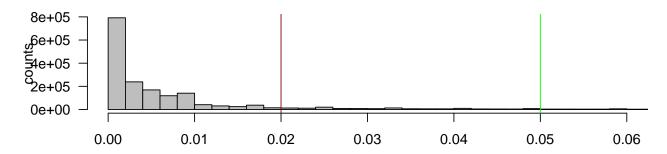
4 Per SNP Characteristics Analysis

4.1 pre-QC missing rate

Histogram of SNP-Missing-Rate, all

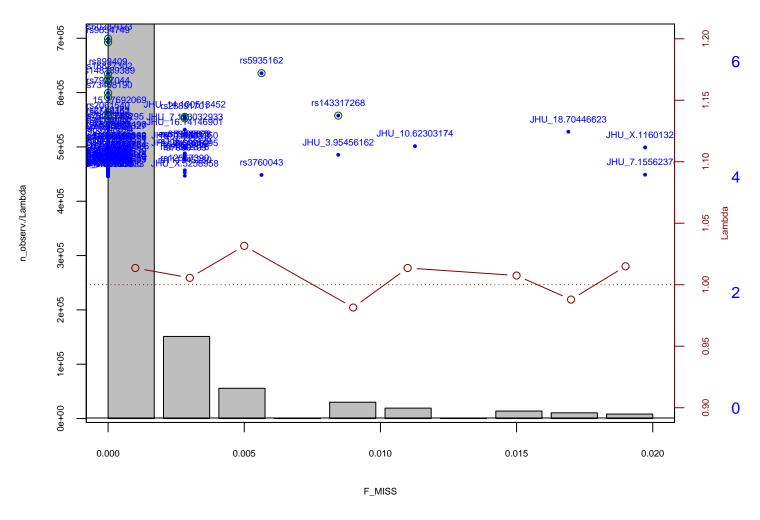


Histogram of SNP-Missing-Rate, zoomed

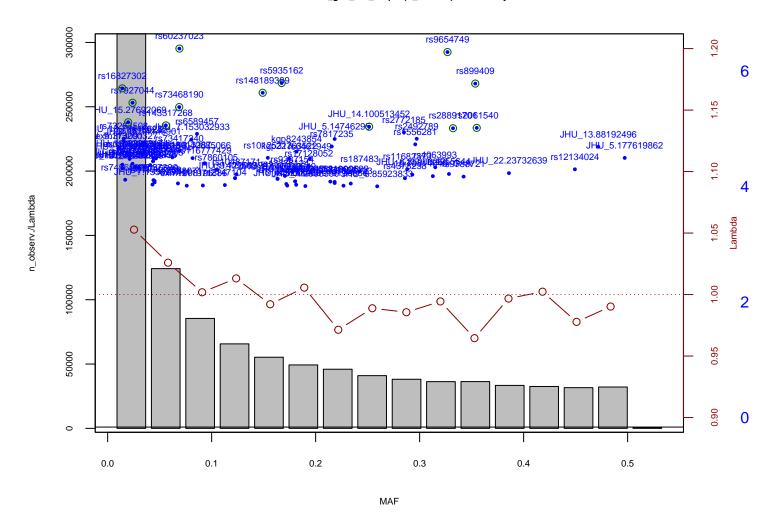


4.2 Lambda-Plot to various variables (all post-QC)

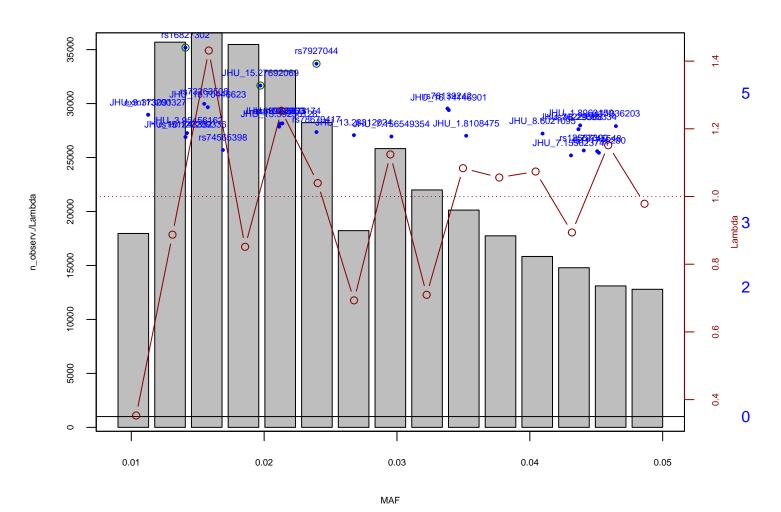
Lambda of LA1leu_gda_eur_lt-qc1 (P_ASSOC) stratified by F_MISS



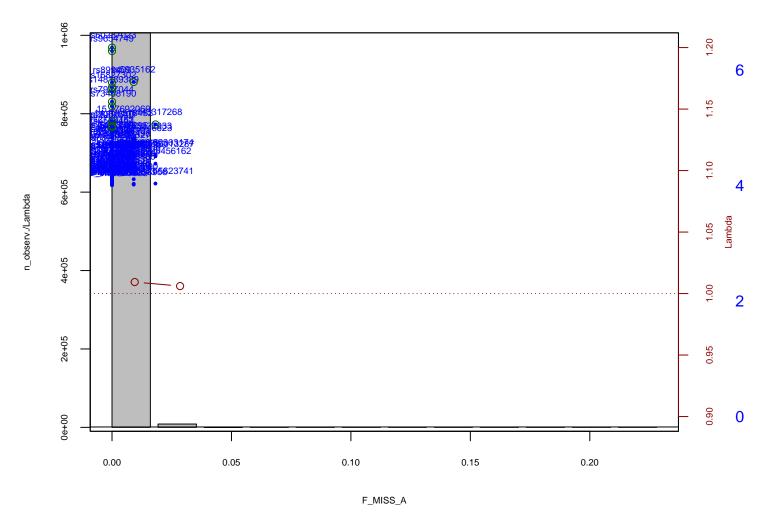
Lambda of LA2leu_gda_eur_lt-qc1 (P_ASSOC) stratified by MAF



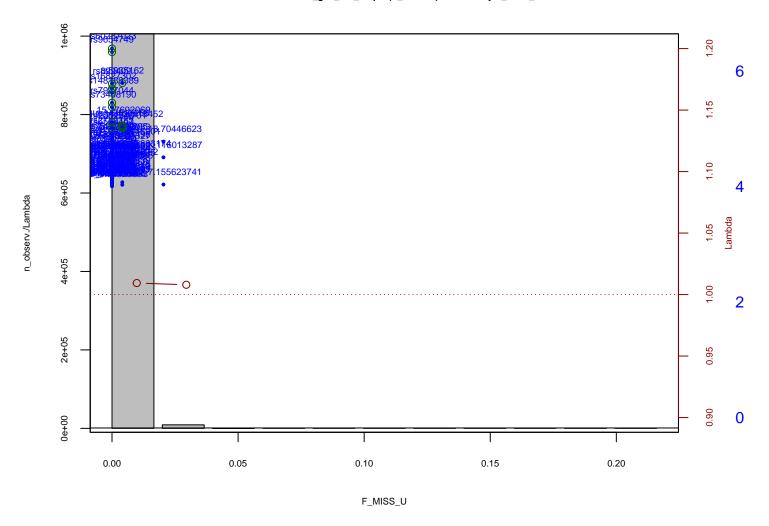
Lambda of LA2bleu_gda_eur_lt-qc1 (P_ASSOC) stratified by MAF



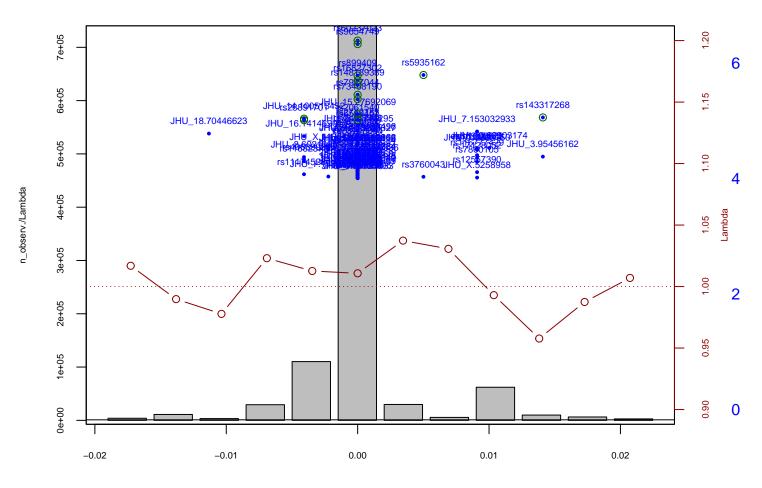
Lambda of LA9leu_gda_eur_lt-qc1 (P_ASSOC) stratified by F_MISS_A



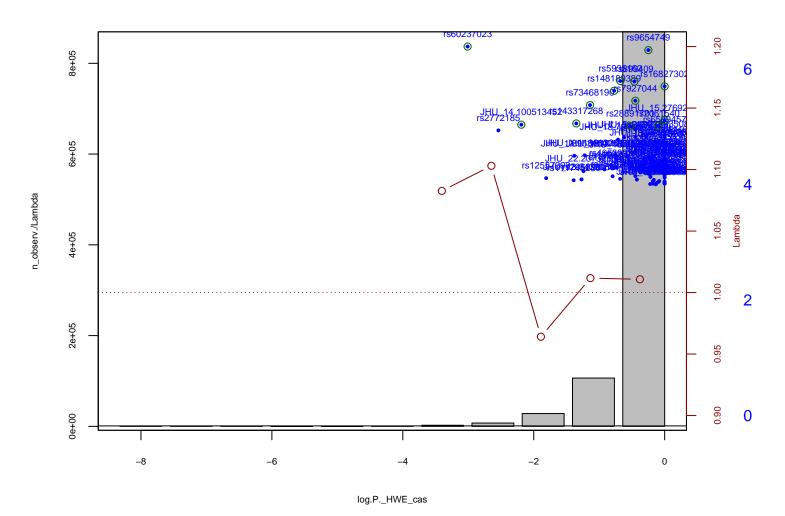
Lambda of LA10leu_gda_eur_lt-qc1 (P_ASSOC) stratified by F_MISS_U



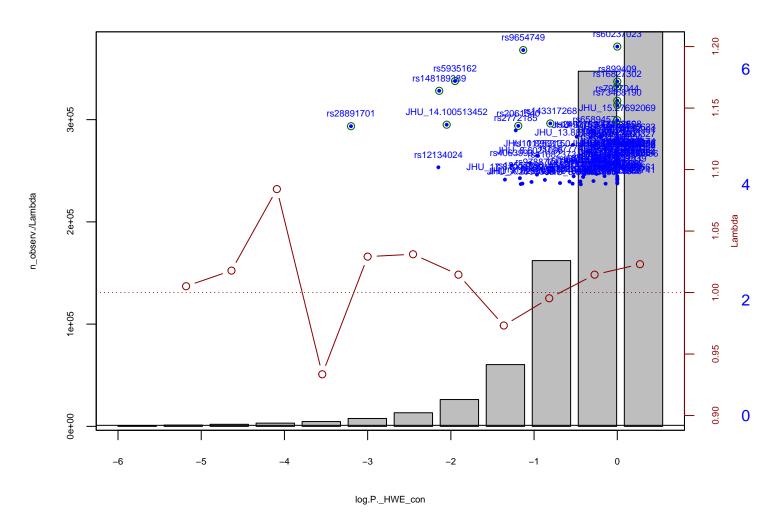
Lambda of LA3leu_gda_eur_lt-qc1 (P_ASSOC) stratified by F_MISS_DIFF



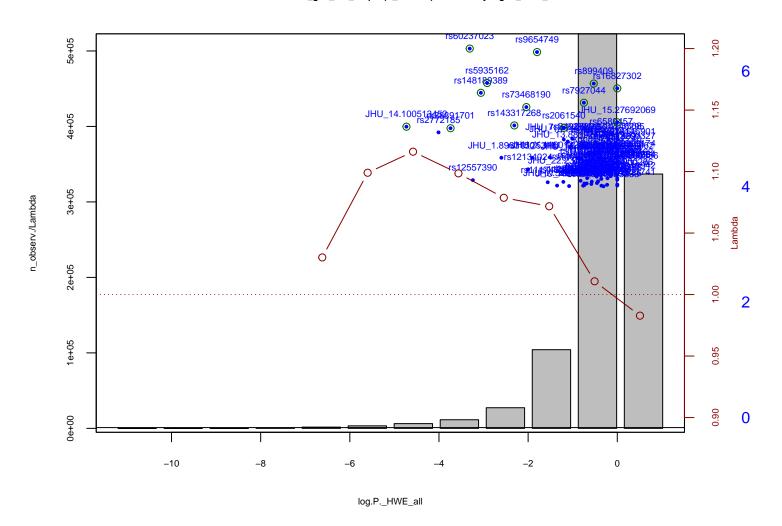
Lambda of LA6leu_gda_eur_lt-qc1 (P_ASSOC) stratified by log.P._HWE_cas



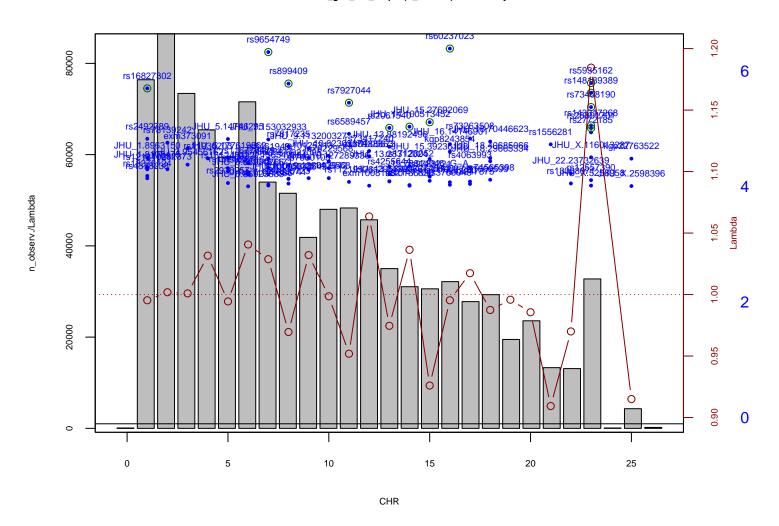
Lambda of LA5leu_gda_eur_lt-qc1 (P_ASSOC) stratified by log.P._HWE_con



Lambda of LA12leu_gda_eur_lt-qc1 (P_ASSOC) stratified by log.P._HWE_all



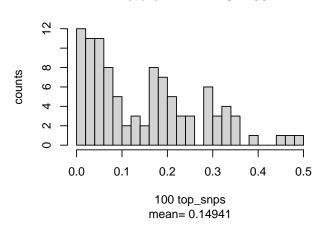
Lambda of LA4leu_gda_eur_lt-qc1 (P_ASSOC) stratified by CHR



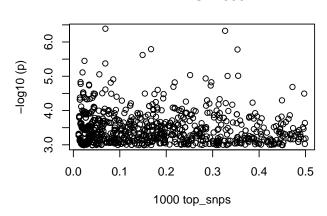
$5\quad \text{Top-Plots for SNPs}$

5.1 Minor Allele Frequency

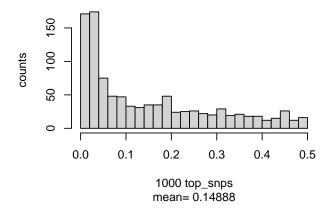
Histo of MAF - TOP100



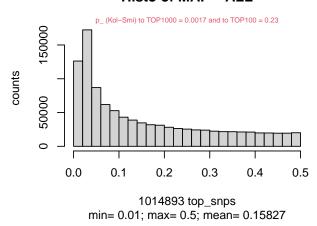
MAF - TOP1000



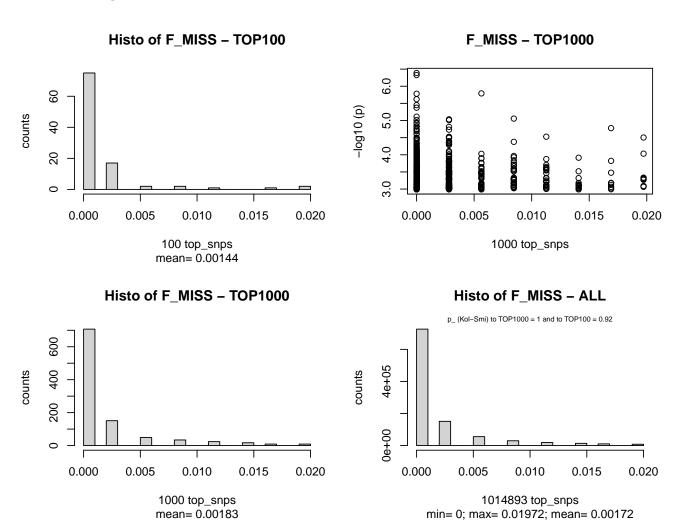
Histo of MAF - TOP1000



Histo of MAF - ALL

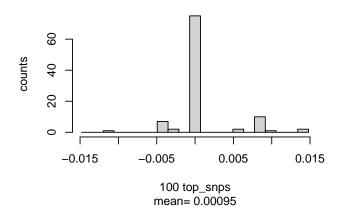


5.2 Missing Rate

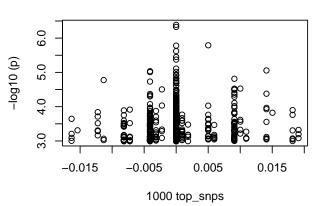


5.3 Missing - Difference (case-control)

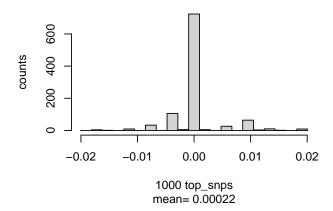
Histo of F_MISS_DIFF - TOP100



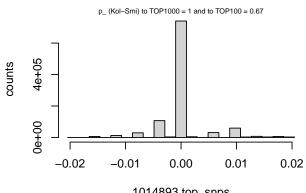
F_MISS_DIFF - TOP1000



Histo of F_MISS_DIFF - TOP1000

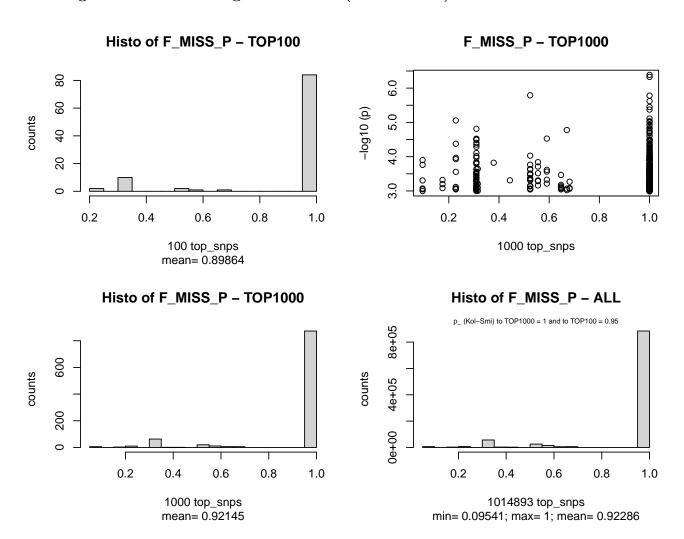


Histo of F_MISS_DIFF - ALL



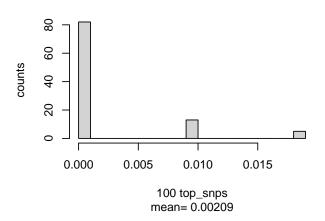
1014893 top_snps min= -0.019479; max= 0.01923; mean= 0.00015

5.4 Significance of Missing - Difference (case-control)

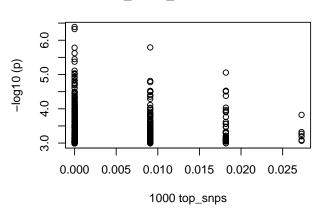


5.5 Missing Rate - affected

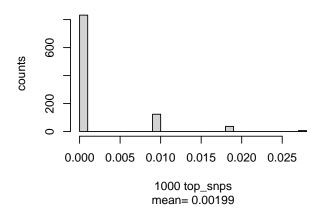
Histo of F_MISS_A - TOP100



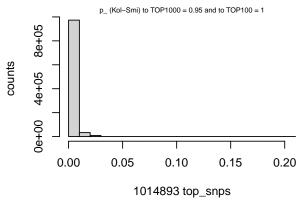
F_MISS_A - TOP1000



Histo of F_MISS_A - TOP1000

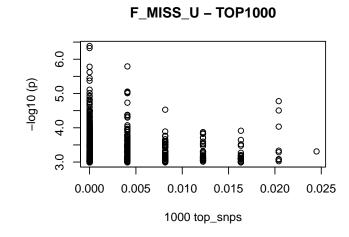


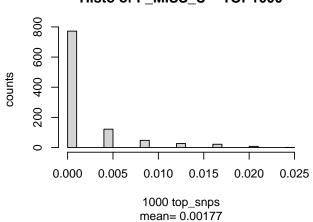
Histo of F_MISS_A - ALL

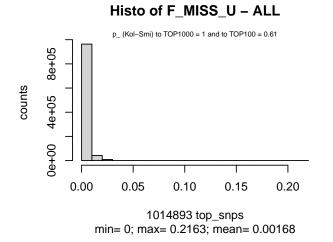


min= 0; max= 0.2091; mean= 0.00183

5.6 Missing Rate - unaffected

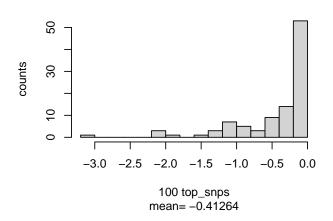




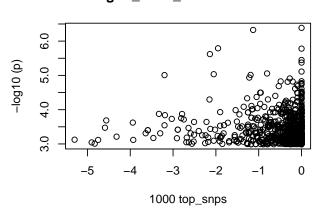


5.7 Hardy Weinberg Equilibrium - unaffected (exact test)

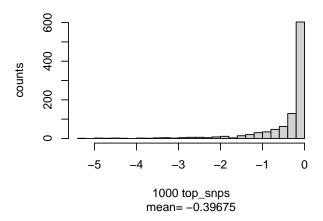
Histo of log.P._HWE_con - TOP100



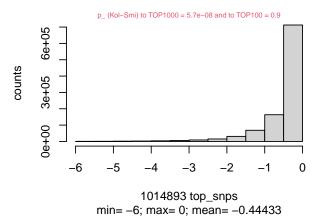
log.P._HWE_con - TOP1000



Histo of log.P._HWE_con - TOP1000

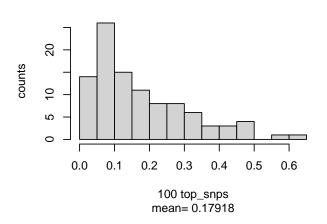


Histo of log.P._HWE_con - ALL

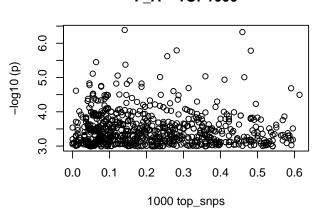


5.8 "Minor" Allele Frequency - affected

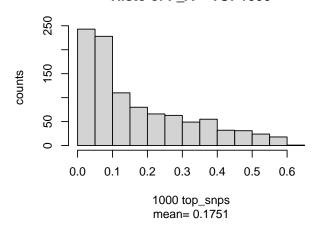
Histo of F_A - TOP100



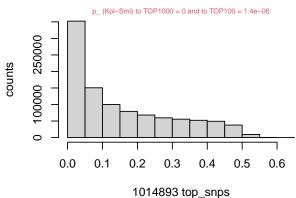
F_A - TOP1000



Histo of F_A - TOP1000



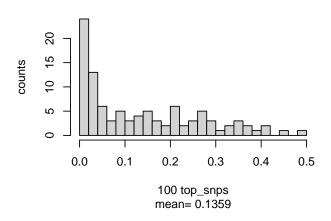
Histo of F_A - ALL



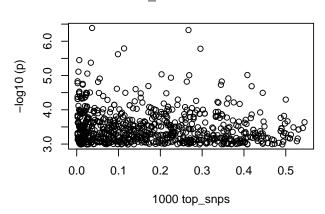
1014893 top_snps min= 0; max= 0.6136; mean= 0.15825

5.9 "Minor" Allele Frequency - unaffected

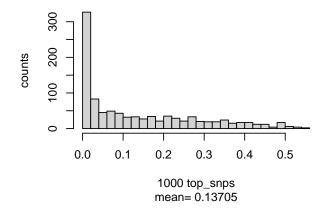
Histo of F_U - TOP100



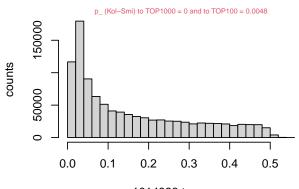
F_U - TOP1000



Histo of F_U - TOP1000



Histo of F_U - ALL



1014893 top_snps min= 0.002041; max= 0.5449; mean= 0.15827