

MDD MWAS - comparison between smoking variables

Miruna Barbu

13/05/2021

3 ways to predict smoking using EpiSmoker Source article: <https://doi.org/10.2217/epi-2019-0206>

There are 3 options to estimate smoking status using EpiSmoker, all derived from whole blood 450K array:

1. Smoking status

Uses 121 CpGs to predict smoking status of each individual (LASSO approach). Sex and intercept derived from study are also used in the calculation.

Bollepalli et al. developed EpiSmokEr (Epigenetic Smoking status Estimator), an easy-to-use R package to facilitate the prediction of smoking status. Given methylation data from microarray assays, it provides predicted smoking status as an output. EpiSmokEr can be run not only on whole datasets, but also on isolated individual samples. To be comprehensive, Bollepalli et al. also provide functions for computing SSc and MS.

N=86/121 of CpGs are available in GS.

2. Smoking score

Uses 187 smoking-associated CpGs identified in Zeilinger et al., 2013 (<https://doi.org/10.1371/journal.pone.0063812>) and calculated using method from Elliott et al., 2014 (<https://doi.org/10.1186/1868-7083-6-4>)

Elliot et al. have suggested calculating a weighted score referred to as ‘smoking score’ by adding up the methylation levels of the 187 CpG sites found to be significantly associated with smoking by Zellinger et al., after first multiplying each methylation value by its effect size in Zellinger et al.’s EWAS study.

Study population in Zeilinger et al.: KORA S4

N=165/187 of CpGs are available in GS.

3. Methylation score

Uses 4 smoking-associated CpGs identified in Zhang et al., 2016 (<https://doi.org/10.1016/j.envres.2016.01.026>)

Zhang et al. used a regression approach. They first determined significantly associated CpG sites in their cohort, and then employed stepwise logistic regression with forward selection. They used a significance-based stopping criterion and found that already after including the **four most significant loci**, the addition of the fifth locus no longer resulted in a significant reduction of the fit deviance. Therefore, their proposed ‘methylation score’ only requires **four CpG sites**.

N=2/4 of CpGs are available in GS.

Descriptive statistics

Table 1: Predicted smoking status, mean smoking score, and mean methylation score for current, former, and never smokers in GS.

GS smoke status	N	Pred smoke status	Smoke score mean	Methyl score mean
Current smoker	1620	Current Smoker	8.584973	2.913198
Former smoker (stopped < 12 months)	274	Current Smoker	6.914629	2.825632
Former smoker (stopped > 12 months)	2604	Current Smoker	5.117609	2.579194
Never smoker	4787	Current Smoker	4.060137	2.363031

Possible issues with calculating these in different cohorts

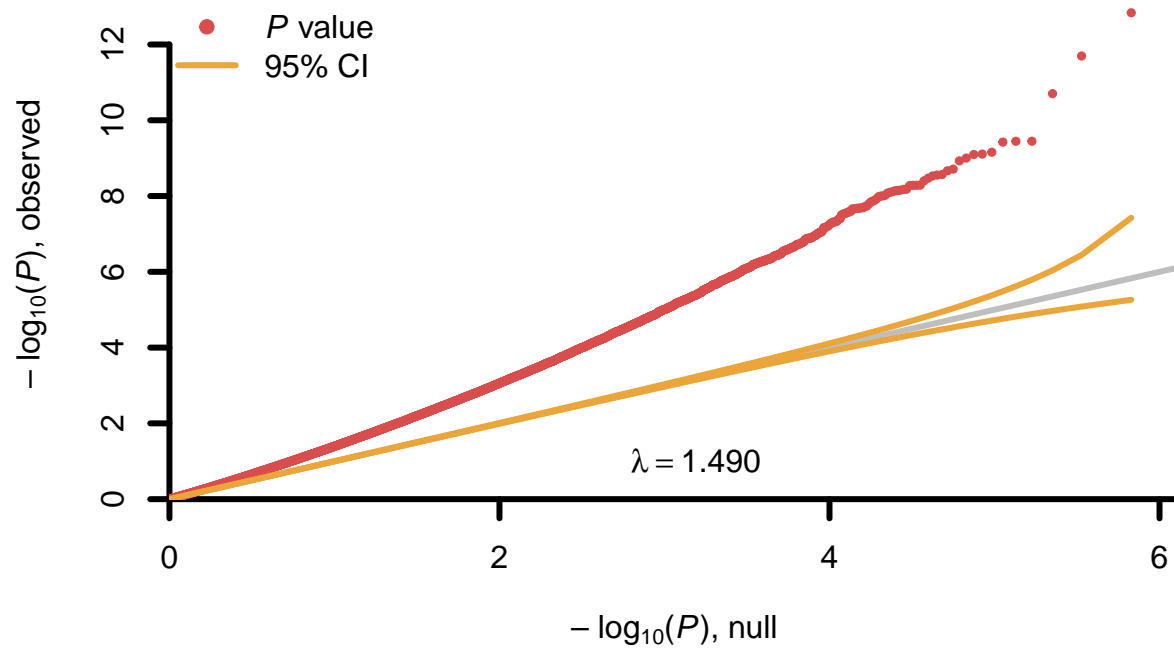
- M-values required for MWAS, but beta values required for EpiSmoker calculation
- Different cohorts will have different CpGs available for calculation
- Required to extract each set of CpGs for Epismoker

Comparing smoking signatures across MWAS - MDD 5 basic models were run, with technical and biological covariates accounted for. CpGs were included as outcomes and MDD diagnosis as the predictor variable, with:

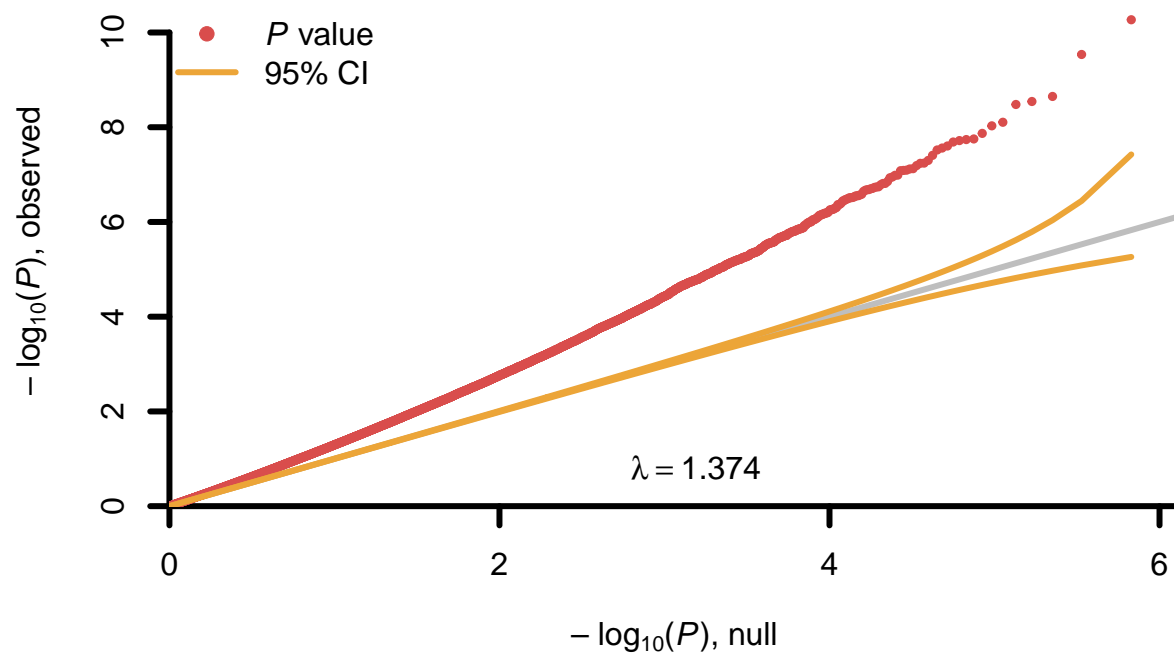
- (1) no smoking covariate
- (2) phenotypic smoking (current, former, never smokers)
- (3) AHRR probe (cg05575921)
- (4) EpiSmoker smoking-derived score (based on 165/187 CpGs)
- (5) EpiSmoker methylation-derived smoking score (based on 2/4 CpGs).

Comparing statistics across MWAS - MDD QQ-plots for indicating genomic inflation factor.

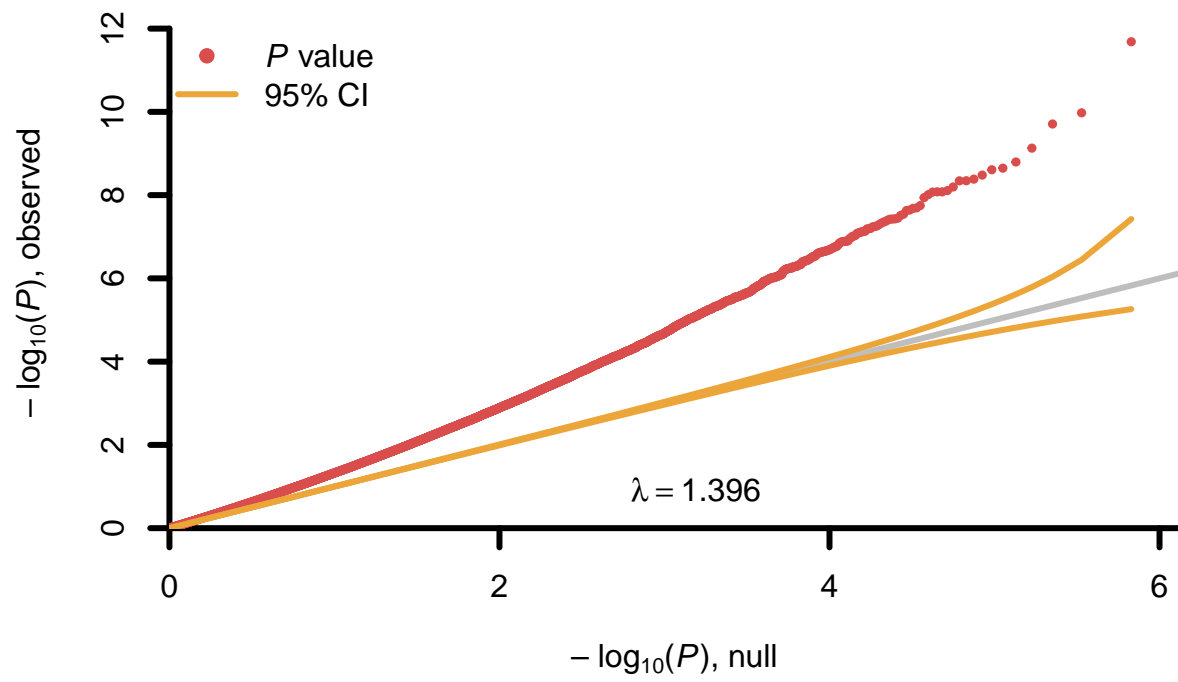
QQ-plot No smoking covariate



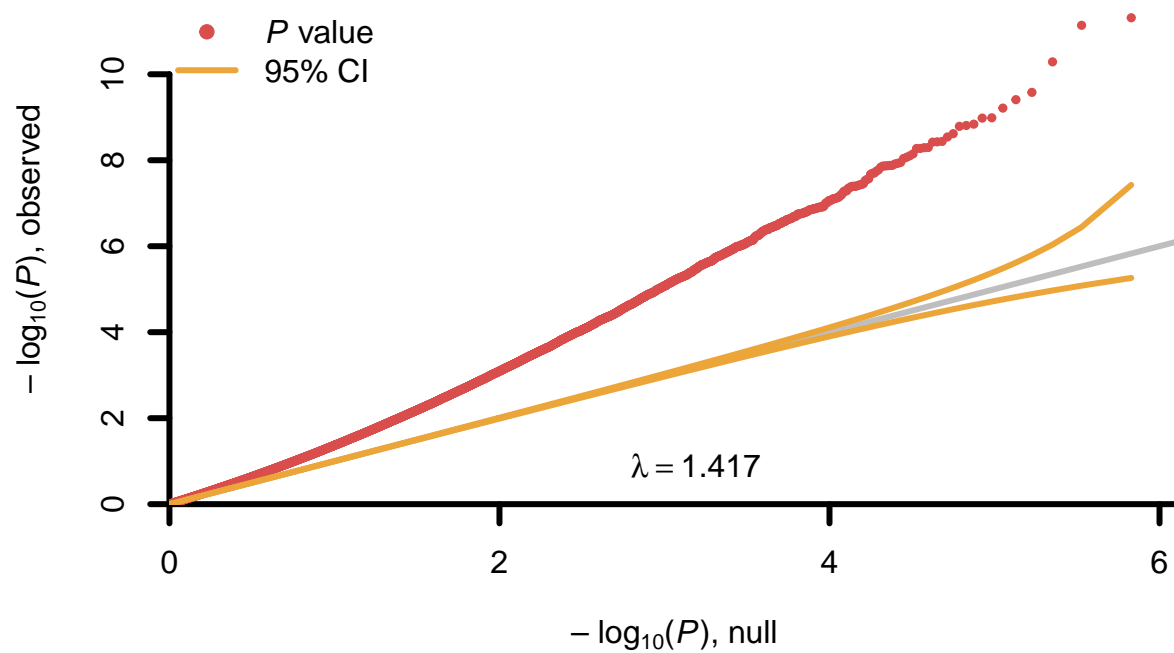
QQ-plot Phenotypic smoking covariate



QQ-plot
AHRR probe (cg05575921) covariate



QQ-plot EpiSmoker: smoke score covariate



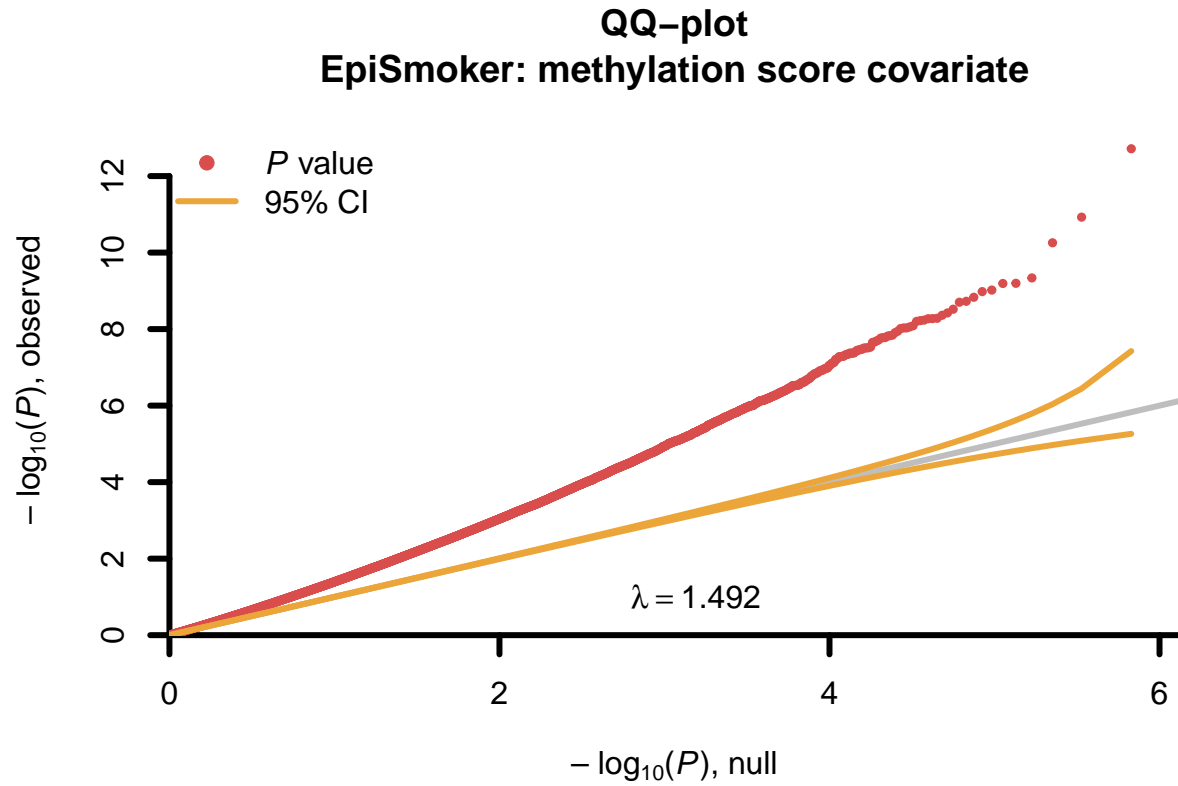
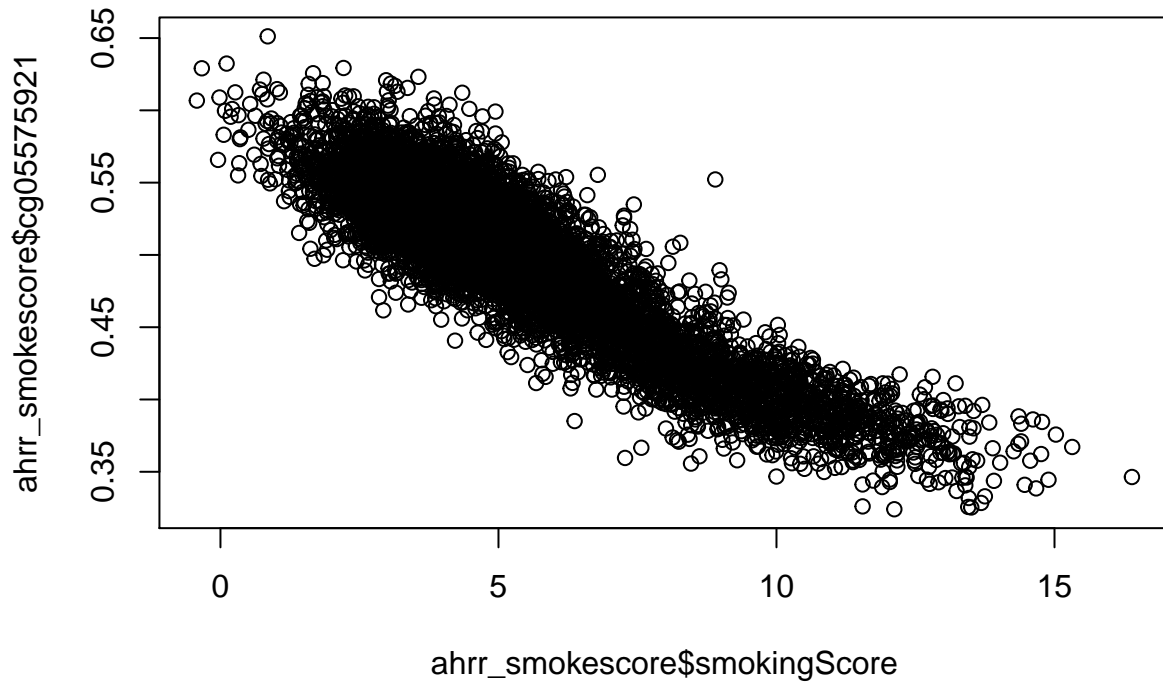


Table 2: Variance inflation factor, number of Bonferroni-corrected CpGs, minimum p-value, and minimum & maximum effect sizes for every MWAS based on smoking covariate included. All models are corrected for technical and biological covariates.

Covariate included in MWAS	lambda	N Bonferroni-corrected CpGs	Min p-value	Min effect size	Max effect size
Phenotypic smoking	1.374	15	5.38433595238757e-11	-0.040	0.050
AHRR probe (cg05575921)	1.396	26	2.07051685731157e-12	-0.041	0.051
EpiSmoker smoke score	1.417	42	4.85097353029382e-12	-0.040	0.055
No smoking covariate	1.49	57	1.45303929463449e-13	-0.050	0.055
EpiSmoker methylation score	1.492	46	1.9299926391313e-13	-0.050	0.054

Correlation between AHRR probe (cg05575921) and EpiSmoker smoking score

[1] -0.8983258



Check MDD MWAS results against Joehanes et al. FDR-corrected CpGs

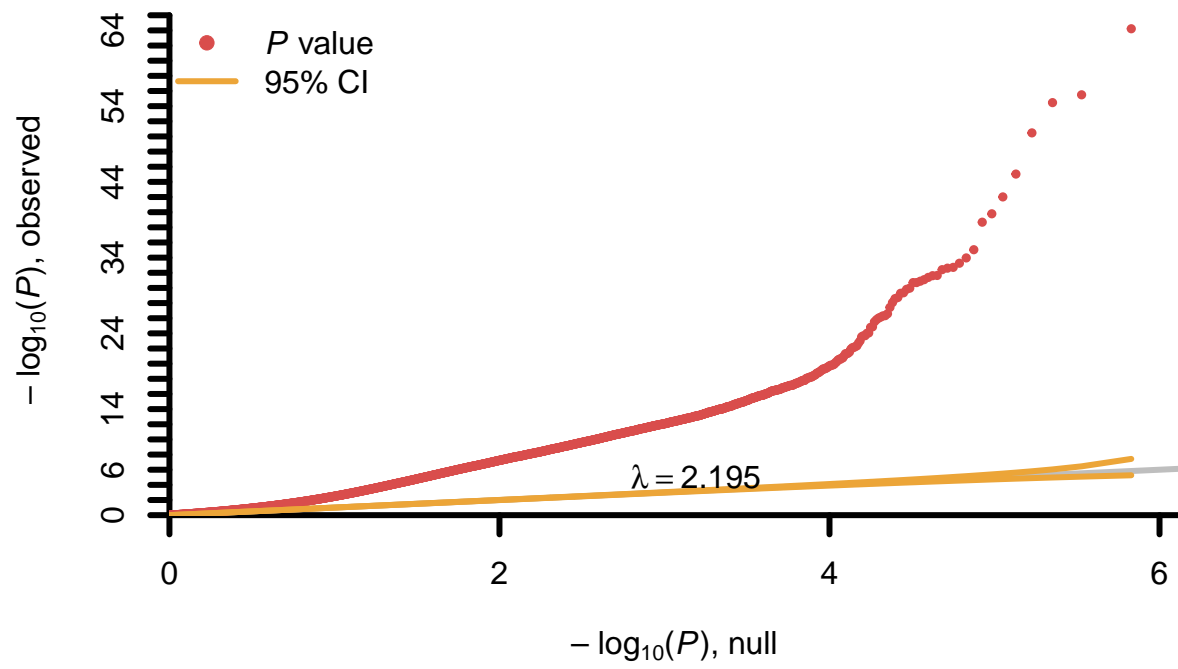
Table 3: Number of Bonferroni-corrected CpGs, minimum p-value, and overlapping top CpGs with Joehanes et al. FDR-corrected CpGs for every MDD MWAS based on smoking covariate included. All models are corrected for technical and biological covariates.

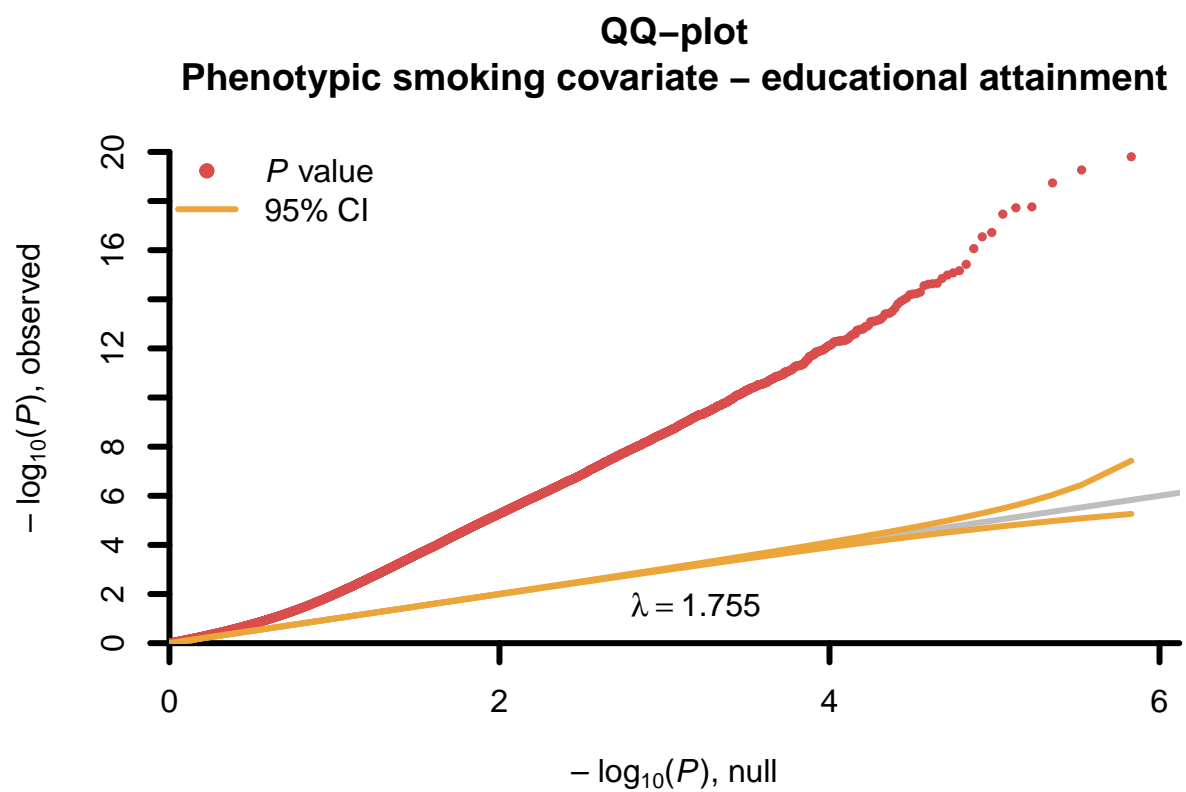
Covariate included in MWAS	N Bonferroni-corrected CpGs	Min p-value	Overlapping CpGs with Joehanes et al.
Phenotypic smoking	15	5.38433595238757e- 011	
AHRR probe (cg05575921)	26	2.07051685731157e- 012	
EpiSmoker smoke score	42	4.85097353029382e- 012	
EpiSmoker methylation score	46	1.9299926391313e- 913	
No smoking covariate	57	1.45303929463449e- 1513	

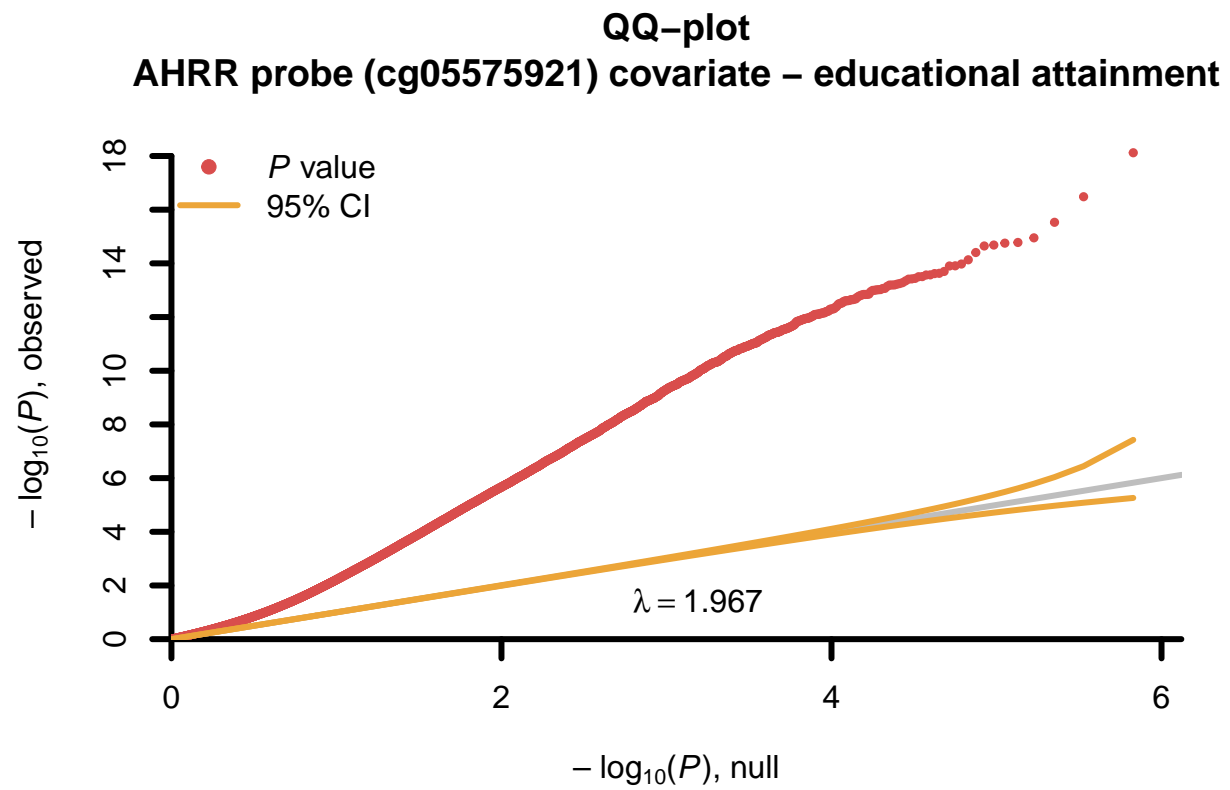
MWAS for trait associated with smoking - educational attainment

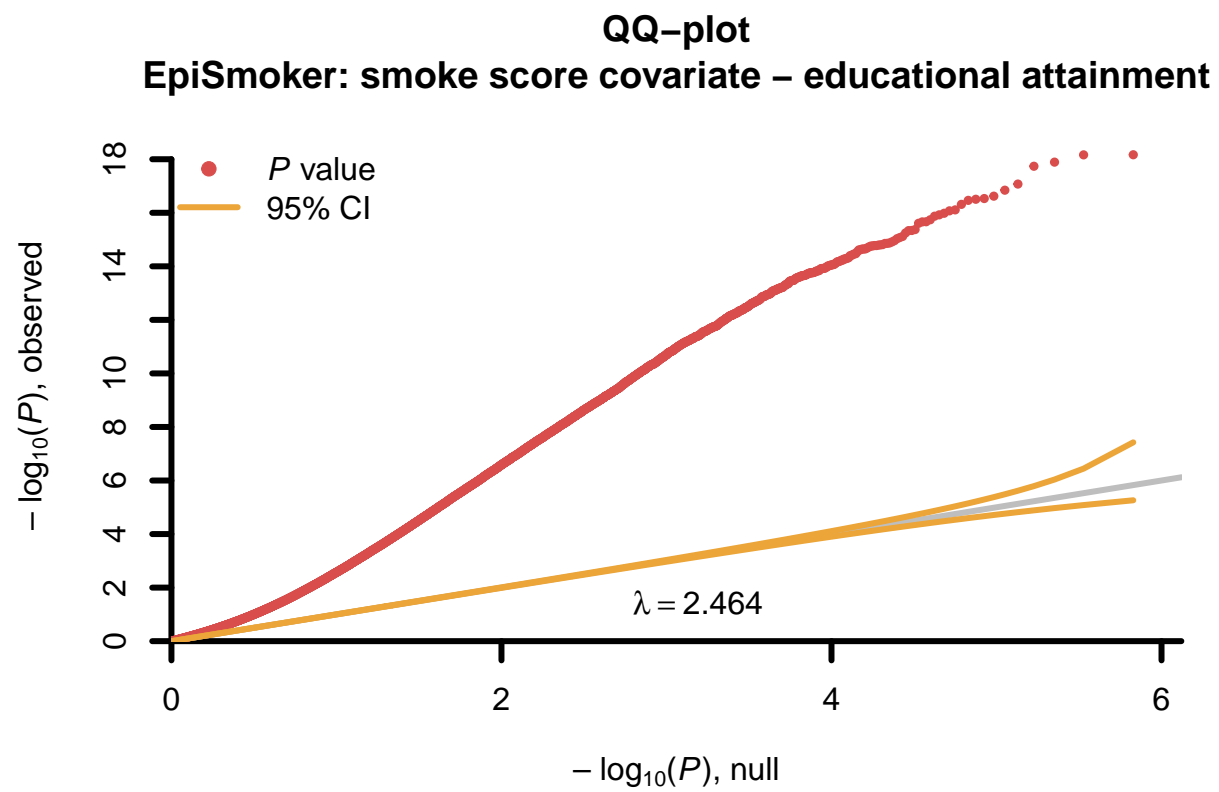
Comparing statistics across MWAS QQ-plots for indicating genomic inflation factor.

QQ-plot
No smoking covariate – educational attainment









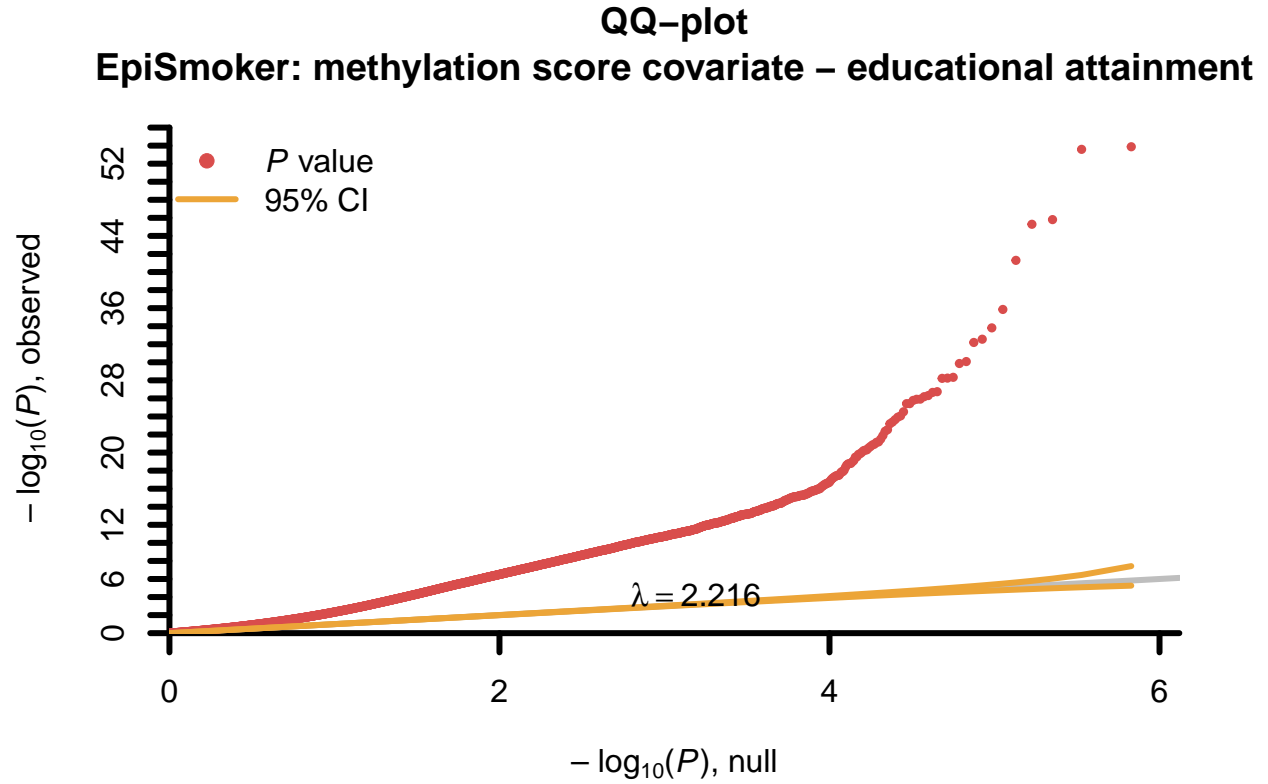


Table 4: Variance inflation factor, number of Bonferroni-corrected CpGs, minimum p-value, and minimum & maximum effect sizes for every Educational Attainment MWAS based on smoking covariate included. All models are corrected for technical and biological co-variates.

Covariate included in MWAS	lambda	N Bonferroni-corrected CpGs	Min p-value	Min effect size	Max effect size
Phenotypic smoking	1.755	1468	1.57247945788674e-20	-0.008	0.012
AHRR probe (cg05575921)	1.967	2129	7.56149718981094e-19	-0.007	0.012
No smoking covariate	2.195	6097	6.06704958142841e-65	-0.021	0.016
EpiSmoker methylation score	2.216	4042	1.33167029947138e-54	-0.017	0.015
EpiSmoker smoke score	2.464	4182	6.7966527419853e-19	-0.005	0.012