Supplementary Table 2

Blood EWAS summary statistics

Data location and access

The complete, machine-readable results for Supplementary Table 2 have been deposited in Figshare through the data-deposition service associated with Nature Mental Health manuscript NATMH-25-1573A.

Dataset title: Epigenome-wide association study in a large admixed cohort reveals immune and stress response epigenetic alterations in cocaine use disorder.

The deposit is private at this submission stage. Editorial and peer-review access is provided through the Figshare record linked to the manuscript submission.

The two compressed CSV files identified below contain the full results. This PDF is the accompanying documentation: it describes the models, defines each data column and provides checksums to identify the deposited files.

Models and file contents

Complete results from the corrected blood EWAS. Each compressed CSV contains 627,130 unique CpGs, sorted by increasing BACON-calibrated P value. Both files contain the same tested CpG universe.

File	Model	CUD	Controls	Total	Significant CpGs
Supplementary_Table_2_EWAS_smoking_adjusted.csv.gz	Primary smoking-adjusted model (panel B)	562	679	1,241	420
Supplementary_Table_2_EWAS_without_smoking_score.csv.gz	Without smoking score (panel A)	562	683	1,245	107

Significance means **BACON two-sided P < 9×10^{-8}**, with a strict inequality. The adjusted list has 336 negative and 84 positive coefficients; the model without smoking score has 90 negative and 17 positive coefficients. The lists share 72 CpGs, giving a union of 455 (348 adjusted-only and 35 without-score-only).

Both models include age, sex, CD8T, CD4T, NK, B-cell and monocyte proportions, 10 EPISTRUCTURE PCs and array position, with a Sentrix chip random intercept. The primary model additionally adjusts for the methylation-derived smoking score. Four controls lack that score. The input beta values were transformed to M values as log2(beta/(1−beta)), after replacing exact 0/1 with 0.0001/0.9999 in the recorded analysis. Coefficients are on the M-value scale, not beta-value differences or methylation percentages.

BACON calibration was performed separately for the two models with BACON 1.36.0 (5,000 iterations, 2,000 burn-in, globalSeed 42). The raw P uses the model t distribution; the calibrated P uses the standard normal distribution. All raw and BACON numerical fields are retained separately. BH adjustment uses the full set of 627,130 CpGs separately for each model and stage. The source genomic-control P and raw Holm flag are retained as diagnostics and are not used to define the significant lists.

Annotation is from IlluminaHumanMethylationEPICanno.ilm10b4.hg19 version 0.6.0. All tested CpGs have matching chromosome/position annotations. Original UCSC gene and gene-region fields are preserved, including repeated semicolon-separated transcript entries. Empty annotation fields mean no annotation in this package; they do not establish that a locus is intergenic.

Column definitions

Column	Meaning
Rank_BACON_P	Within-model ascending rank by BACON P; ties are ordered by CpG identifier.
CpG	Illumina EPIC CpG probe identifier.
Chr_hg19	Chromosome in the hg19 annotation, without the chr prefix.
Position_hg19	Genomic position from the hg19 annotation package.
UCSC_RefGene_Name	Unmodified UCSC RefGene name field; semicolon-separated transcript annotations may repeat a gene. Blank means not annotated in this package.
UCSC_RefGene_Group	Unmodified gene-region field, potentially containing repeated or multiple semicolon-separated annotations. Blank means not annotated.
Relation_to_Island	Relation to CpG island from the EPIC hg19 annotation.
Effect_raw_M	Uncalibrated CUD-minus-control model coefficient on the M-value scale.
SE_raw_M	Uncalibrated model standard error on the M-value scale.
t_raw	Raw coefficient divided by raw SE; uses the raw model degrees of freedom.
P_raw	Two-sided raw model P value from the t statistic and denominator degrees of freedom.
BH_FDR_raw	Benjamini-Hochberg adjusted raw P across all 627,130 tested CpGs within this model.
Effect_BACON_M	BACON-calibrated coefficient on the M-value scale. Negative denotes lower calibrated methylation in CUD.

Column definitions continued

Column	Meaning
SE_BACON_M	BACON-calibrated standard error; not an estimate of residual biological variance.
z_BACON	BACON-calibrated statistic (source field post_t); coefficient_BACON / SE_BACON, interpreted using the standard normal distribution.
P_BACON	Two-sided standard-normal P from z_BACON. This is the P value used to select EPIC-wide significant CpGs.
BH_FDR_BACON	Benjamini-Hochberg adjusted BACON P across all 627,130 tested CpGs within this model.
df_numerator_raw	Numerator degrees of freedom in the source model output (1).
df_denominator_raw	Denominator degrees of freedom in the raw model: 1061 adjusted and 1066 without smoking score.
Intercept_adjusted_M	Source field adj.intercept, retained as exported by CpGassoc; not the CUD effect.
Holm_significant_raw_0_05	Source Boolean flag for Holm-adjusted raw model P < 0.05 across all tested CpGs; not the BACON selection criterion.
P_genomic_control	Source genomic-control diagnostic P calculated from raw P and genomic inflation; not the BACON P and not used for selection.
Significant_BACON_P_lt_9e_8	TRUE if and only if P_BACON < 0.00000009 (strict inequality).

File integrity

Compressed files can be read directly by standard R/Python CSV readers. Blank annotation fields are empty strings; there are no missing numerical summary results. Scientific notation and full exported precision are retained.

File	SHA-256
Supplementary_Table_2_EWAS_smoking_adjusted.csv.gz	6602e52d99f9afccce3c287eb620e916eeba863c92320ede07f58f14997771a7
Supplementary_Table_2_EWAS_without_smoking_score.csv.gz	675edd0e6c35a23d0fdbae0c213b63001e0cfbaf01e91e891e047b1270f76936

The files are publication exports of the finalized corrected 2024 analysis; no new association model or BACON fit was run to produce these tables.