Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Epigenome-wide association study in a large admixture cohort reveals immune and stress response epigenetic alterations in cocaine use disorder

Rodrigo Grassi-Oliveira

rogo@clin.au.dk

Pontifical Catholic University of Rio Grande do Sul https://orcid.org/0000-0001-9911-5921

Gabriel Fries

McGovern Medical School

Bruno Kluwe-Schiavon

McGovern Medical School, Houston

Thiago Viola

Pontifical Catholic University of Rio Grande do Sul

Diego Rovaris

University of Sao Paulo https://orcid.org/0000-0002-8910-4927

João Antônio Leal

Pontifical Catholic University of Rio Grande do Sul

Eugenio Grevet

Universidade Federal do Rio Grande do Sul

Claiton Bau

Universidade Federal do Rio Grande do Sul

Joy Schimidt

McGovern Medical School

Consuelo Walss-Bass

McGovern Medical School, Houston

Article

Keywords: cocaine use disorder, epigenetics, DNA methylation, trauma, stress response, immune regulation

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9544880/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Title

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

Authors

Rodrigo Grassi-Oliveira[1,2]*, Gabriel R. Fries[3], Bruno Kluwe-Schiavon[1], Thiago Wendt Viola[1], Diego Luiz Rovaris[4], João Antônio Valadão Leal[1], Eugenio Horacio Grevet[5], Claiton Henrique Dotto Bau[5], Joy M. Schmitz[3], and Consuelo Walss-Bass[3]

Affiliations

[1] Nanopsychiatry Lab, Brain Institute and School of Medicine, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Brazil

[2] Translational Neuropsychiatry Unit, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

[3] Department of Psychiatry and Behavioral Sciences, The University of Texas Health Science Center at Houston (UTHealth), Houston, TX, USA

[4] Department of Physiology and Biophysics, Institute of Biomedical Sciences, University of São Paulo (USP), São Paulo, Brazil

[5] Department of Psychiatry, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

Corresponding author

*Rodrigo Grassi-Oliveira, Ph.D.

Translational Neuropsychiatry Unit, Aarhus University, Aarhus, Denmark

and Brain Institute, PUCRS, Porto Alegre, Brazil

Email: rodrigo.grassi@pucrs.br; rogo@clin.au.dk

Manuscript type

Article

Word count (Introduction, Results and Discussion)

3333

Figures/Tables

4 figures; 2 tables

Supplementary Information

3 supplementary figures; 9 supplementary tables

Conflict of interest statement

The authors declare no competing interests.

Data availability

Complete blood EWAS summary statistics are deposited in the Figshare record linked to this submission (Supplementary Table 2); secondary-analysis results are provided in the remaining supplementary tables. Individual-level clinical and methylation data are available through the corresponding author upon reasonable request and institutional approval, as described in the Data availability section.

Code availability

Code for the final EWAS models, BACON calibration, downstream analyses and figure generation, with documentation of required inputs and software versions, is publicly available in Zenodo under the MIT License (https://doi.org/10.5281/zenodo.22851173).

Abstract

Here, we aimed to characterize blood DNA methylation associated with cocaine use disorder (CUD), examine its relationship with substance-use histories and trauma, and explore concordance with brain methylation. In an epigenome-wide study of 1,241 Brazilian participants, adjustment for smoking and other covariates followed by BACON calibration identified 420 differentially methylated positions ($P < 9 \times 10^{-8}$), predominantly with lower methylation in CUD. Leading loci included *FKBP5* and *NLRC5*, while enrichment analyses connected the findings to immune and cytokine signaling. Within CUD, alcohol-use history was associated with two loci after correction for multiple testing, whereas associations with trauma and childhood maltreatment did not survive correction. Among 30 brain donors, no association survived correction across 388 blood-selected loci. Together, these findings characterize a predominantly hypomethylated peripheral profile associated with CUD and place its interpretation within the immune–stress and polysubstance context in which cocaine use occurs.

Keywords

cocaine use disorder; epigenetics; DNA methylation; trauma; stress response; immune regulation

Introduction

Cocaine use disorder (CUD), characterized by compulsive cocaine use despite adverse medical, psychological and behavioral consequences[1], remains a major global public health concern. The World Drug Report 2026 estimated that approximately 25.2 million people used cocaine worldwide in 2024, corresponding to 0.47% of the population aged 15–64 years[2]. Understanding the biology of CUD requires considering multiple levels of regulation, from neuroanatomical alterations[3,4] to genetic vulnerabilities[5-7] and epigenetic mechanisms. DNA methylation is of particular interest because it provides a molecular level at which to investigate the relationship between substance exposure and biological variation. Experimental and human postmortem studies have identified cocaine-associated methylation changes in brain tissue[8,9], while genome-wide studies have reported peripheral blood differences associated with cocaine dependence[10] and active injection drug use[11], as well as umbilical cord blood differences following prenatal cocaine exposure[12]. These studies provide a basis for investigating cocaine-associated methylation in humans, but differences in sample size, exposure definition and clinical characterization leave unresolved how the peripheral findings relate to the broader clinical context of CUD.

A central challenge is that CUD rarely occurs as an isolated exposure. Individuals with CUD commonly have histories of tobacco, alcohol and cannabis use, so their peripheral methylome may reflect multiple overlapping exposures. This pattern extends beyond treatment-seeking cohorts. In NESARC-III, a nationally representative survey of 36,309 US adults, 86.0% of adults with past-year CUD had at least one additional past-year substance use disorder[13]. Similarly, in the 2022–2023 NSDUH, comprising 92,233 adults, 90.5% of those reporting past-year cocaine use reported using at least four substances during that year[14]. Together with observations from Brazilian clinical cohorts[15], these findings place polysubstance involvement at the center of the clinical setting in which cocaine-associated methylation is studied. Tobacco exposure is especially relevant because it produces some of the most reproducible methylation differences in human blood, including at *AHRR* and *F2RL3*[16]. Accounting for smoking in the primary case–control model, while examining other substance-use histories within CUD, is therefore essential to the interpretation of the peripheral methylation profile.

Environmental adversity adds another dimension to this exposure history. Population studies have linked adverse childhood experiences to substance use and other health risks[17]. Childhood exposure to potentially traumatic events (PTEs) has been associated with an increased risk of cocaine use during adolescence[18], while cocaine dependence itself has been associated with higher rates of trauma exposure and comorbid post-traumatic stress disorder (PTSD)[19]. These observations make lifetime PTE exposure and childhood maltreatment relevant to the clinical interpretation of CUD and raise the question of whether they are associated with variation at the methylation loci that distinguish cases from controls. Examining adversity alongside substance-use histories allows this question to be addressed directly, rather than treating trauma solely as a background characteristic of the cohort.

Here, we hypothesized that blood methylation would be associated with CUD diagnosis and environmental exposures within CUD. We tested this in a large admixed Brazilian cohort using a smoking-adjusted EWAS and analyses of lifetime PTE exposure, childhood maltreatment, substance-use histories and income, and explored concordance with postmortem prefrontal cortex findings.

Results

Sample characteristics

The blood cohort comprised 562 participants with CUD and 683 controls; four controls lacked a smoking score, leaving 1,241 participants in the primary EWAS. Participants with CUD had lower income and educational attainment, and the groups differed in self-reported racial/ethnic composition (Table 1). Lifetime nicotine, cannabis and alcohol use disorders were recorded in 88.1%, 73.3% and 68.7% of participants with CUD, respectively.

Participants with CUD reported a greater burden of lifetime PTE exposure and childhood maltreatment. Life-threatening events were reported by 54.9% of cases versus 19.9% of controls, and rape by 29.0% versus 2.7%. Severe childhood physical abuse (25.2% versus 4.1%), emotional abuse (26.2% versus 6.7%) and sexual abuse (10.7% versus 2.9%) showed a similar pattern (Table 1). These differences establish the clinical context for examining exposure-related methylation variation within CUD.

DNA methylation differences

The primary EWAS tested 627,130 autosomal CpGs adjusting for technical factors, age, sex, estimated blood-cell composition, ancestry-related methylation structure and the methylation-derived smoking score. After BACON calibration, 420 differentially methylated positions (DMPs) reached $P < 9 \times 10^{-8}$: 336 showed lower and 84 higher methylation in CUD (Figure 1; Supplementary Tables 1–2). The model without the smoking score identified 107 DMPs, comprising 90 with lower and 17 with higher methylation. The models shared 72 loci; after smoking adjustment, 35 no longer met the threshold and 348 additional loci did.

Before BACON calibration, 10,533 loci met the threshold without smoking adjustment and 3,751 with it. Smoking adjustment increased raw standard errors (median paired ratio 1.131), but reduced the BACON empirical-null scale from 1.978 to 1.371, yielding lower calibrated standard errors (median ratio 0.784). Genomic inflation factors were 4.979 without and 2.821 with smoking adjustment before calibration, and 1.024 and 1.186 afterward (Supplementary Figure 1). The larger calibrated discovery set therefore accompanied changes in both fitted associations and empirical-null calibration.

FKBP5 cg03546163 and *NLRC5* cg07839457 were the strongest smoking-adjusted associations, with additional leading signals at *FBLN5*, *OSCAR* and *BST2*/*BISPR*/*MVB12A* (Table 2). *SOCS3*, *B2M* and *PARP9*/*DTX3L* were represented by three, five and three hypomethylated CpGs, respectively (Supplementary Table 1). *AHRR* cg05575921 remained significant, although its calibrated M-value coefficient attenuated from −1.1864 to −0.3772 after smoking adjustment. Complete results are available in the associated Figshare deposit (Supplementary Table 2).

Smoking-adjusted DMPs were depleted in CpG islands (OR = 0.084, BH-adjusted $P = 2.06 \times 10^{-23}$) and enriched in OpenSea regions (OR = 1.74, BH-adjusted $P = 3.34 \times 10^{-7}$). They were also depleted in promoter-proximal TSS200/TSS1500 annotations and enriched in 5′ untranslated regions (UTRs), gene bodies and 3′UTRs (BH-adjusted $P < 0.05$; Figure 2; Supplementary Table 3). CpG-island depletion and 5′UTR enrichment were also observed without smoking adjustment.

Gene-set enrichment

Gene-set analysis used BACON-calibrated *P* values for all 627,130 tested CpGs, combining methylation directions and accounting for the number of probes mapped to each gene. It identified 498 of 4,984 Gene Ontology (GO) terms and 39 of 322 KEGG pathways at FDR < 0.05 (Figure 3; Supplementary Table 4). Leading GO annotations included specific granule and leukocyte activation involved in immune response.

Herpes simplex virus 1 infection and osteoclast differentiation led the KEGG findings (both FDR = 1.15×10^{-4}), alongside TNF, chemokine, NF-κB, JAK–STAT, B cell receptor and NOD-like receptor signaling. Reactome identified 89 of 1,091 pathways at FDR < 0.05, prominently neutrophil degranulation, interleukin signaling and interferon alpha/beta signaling (Supplementary Table 4). Together, the databases highlighted related immune-cell and cytokine processes.

Exposure associations at CUD-associated CpGs

Within CUD, we examined alcohol, nicotine, cannabis and cocaine use histories, lifetime PTE exposure, childhood maltreatment and income at the 420 selected loci and 2,100 comparator CpGs matched on methylation mean and variability. Available samples included 562 participants for substance-use histories and income, 526 for lifetime PTE scores, and 513 for childhood-maltreatment scores.

Alcohol-use history, measured as years of regular use divided by age, was associated with 12 target CpGs after within-exposure BH correction. Two *CDC42BPB* sites, cg02003183 and cg04987734, also survived correction across all 2,940 tests spanning seven exposures ($q = 0.0109$ and 0.0494). *SOCS3* cg11047325 met only the within-alcohol criterion. No target CpG met the within-exposure criterion for lifetime PTE exposure, childhood maltreatment, income or the duration of regular nicotine, cannabis or cocaine/crack use relative to age. Nicotine associations were conditional on the methylation-derived smoking score included in all models (Supplementary Table 5; Supplementary Figure 2).

Nominal alcohol associations were more frequent among target than comparator CpGs (45/420 versus 120/2,100; OR = 1.98, 95% CI 1.35–2.87; BH-adjusted $P = 0.00235$). Lifetime PTE exposure showed no enrichment (10/420 versus 89/2,100; OR = 0.55, 95% CI 0.25–1.07; BH-adjusted $P = 0.104$). Cannabis, cocaine and childhood maltreatment had fewer nominal associations among targets; nicotine and income showed no corrected difference. These exploratory comparisons used BH correction across seven exposures (Supplementary Table 5).

Restricting CUD to 256 participants reporting no years of regular alcohol abuse, compared with the same 679 controls, retained the effect direction at all 420 loci. Uncalibrated subgroup and primary EWAS coefficients were strongly correlated ($r = 0.992$), with a median coefficient ratio of 0.958 (IQR 0.901–1.018; Supplementary Table 6).

In 498 participants, ten methylation principal components captured 80.3% of variance. None was associated with the seven-domain cumulative adversity index after BH correction, either with primary nuisance covariates or additional substance-use and income adjustment (minimum $q = 0.278$ and 0.194; Supplementary Table 5).

Exploring the link between peripheral blood and brain methylation

The Blood–Brain Epigenetic Concordance (BECon) resource[20] describes whether individuals with higher blood methylation also have higher or lower methylation in paired cortical samples from the same 16 reference donors. Of the 420 loci queried, 136 had available reference correlations. Median Spearman correlations were −0.001 in BA7, −0.028 in BA10 and −0.018 in BA20, ranging from −0.800 to 0.809 across regions. At *FKBP5* cg03546163, correlations were −0.050, −0.262 and 0.488, respectively (Supplementary Figure 3; Supplementary Table 9). These reference correlations describe covariation across individuals rather than CUD–control differences.

Separately, we tested 388 blood-selected loci in postmortem BA9 from 13 donors with CUD and 17 controls. Sixteen showed nominal $P < 0.05$, but none survived BH correction; additional race/ethnicity adjustment yielded 14 nominal associations and none at FDR < 0.05 (Supplementary Table 7). *ASAP1* cg10068788 had negative effects in both tissues, with brain $P = 0.0277$ and $q = 0.9367$; after race/ethnicity adjustment, $P = 0.0562$ and $q = 0.9543$ (Figure 4). These exploratory findings did not establish cross-tissue replication.

Discussion

The strongest smoking-adjusted associations in this large blood EWAS of CUD occurred at loci involved in glucocorticoid signaling and immune regulation. Within the predominantly hypomethylated profile, associations at *FKBP5* and *NLRC5* were accompanied by multiple sites at *SOCS3*, *B2M* and *PARP9*/*DTX3L*. Their biological functions suggest a connection between stress-response regulation, antigen presentation and the mechanisms controlling cytokine signaling. This interpretation was also supported by enrichment in immune-cell and granule-associated processes. Understanding how such a profile relates to CUD requires considering the physiological context of the disorder together with the overlapping exposures that characterize affected individuals.

Smoking adjustment was particularly informative in this respect. Interpreting peripheral correlates of epigenetic dysregulation in CUD[21] requires accounting for tobacco co-exposure, as illustrated by the attenuation of *AHRR* cg05575921, the strongest signal in our model without smoking adjustment and a well-established smoking-sensitive site[22-25]. The influence of tobacco is also evident in a lifetime-cannabis-use meta-analysis of 9,436 participants, in which adjustment reduced the number of FDR-significant CpGs from 608 to four[26]. In our study, the increase in calibrated discoveries accompanied lower empirical-null inflation despite higher raw standard errors. Smoking adjustment therefore changed both the fitted associations and the calibration of their test statistics, providing the analytical context for interpreting the loci that remained prominent.

With this adjustment, our study extends the earlier Brazilian blood EWAS of 23 cocaine/crack-dependent participants and 24 controls[10] through a substantially larger admixed cohort and detailed characterization of co-occurring exposures. Among the resulting associations, *FKBP5* cg03546163 is especially relevant to comparisons across substance-use phenotypes. The same site ranked among the ten strongest signals in a smoking-adjusted, BACON-calibrated alcohol use disorder meta-analysis of 3,775 participants[27] and among the leading, although not epigenome-wide significant, associations with active injection drug use in 401 predominantly African American participants[11]. Its recurrence across these settings draws attention to stress regulation as a biological feature that may be shared across different forms of substance use.

In our cohort, cg03546163, annotated to the 5′ untranslated region of *FKBP5*, showed lower methylation in participants with CUD. *FKBP5* encodes FKBP51, a co-chaperone of the HSP90 complex that reduces glucocorticoid-receptor sensitivity and participates in feedback regulation of glucocorticoid signaling[28]. Acute

cocaine administration increases cortisol secretion in humans[29], directly engaging the system in which FKBP51 operates. Repeated engagement of this system provides a plausible physiological context for the methylation association and connects it to the regulation of inflammatory responses. Yu et al. reported an inverse relationship between cg03546163 methylation and *FKBP5* mRNA in Behçet's disease[30]. In complementary experiments using healthy-donor peripheral blood mononuclear cells, global demethylating treatment increased *FKBP5* expression and LPS-stimulated IL-1β production. FKBP51 also potentiates NF-κB signaling by strengthening the interaction between NIK and IKKα[31]. These observations support a mechanistic hypothesis in which *FKBP5*-associated methylation variation accompanies altered coordination of glucocorticoid feedback and inflammatory signaling in CUD.

The immune component of the methylation profile extends from this connection with glucocorticoid signaling to the regulation of antigen presentation. *NLRC5* cg07839457, the second strongest association, has a promoter-proximal annotation and showed lower methylation in CUD. *NLRC5* coordinates transcription of MHC class I genes and associated components[32,33], whereas β2-microglobulin, encoded by *B2M*, forms part of the MHC-I complex. The presence of five hypomethylated CpGs at *B2M* brings another component of the same machinery into the observed profile. This convergence raises the possibility of altered regulation of communication between antigen-presenting cells and cytotoxic lymphocytes, a question given a complementary cellular basis by separate immune-phenotyping work describing altered T-cell subsets and differentiation in women with CUD[34].

Alongside antigen presentation, the associations at *SOCS3* suggest that the control of cytokine responses deserves attention. Three nearby CpGs, cg18181703, cg11047325 and cg13343932, showed lower methylation in CUD and remained significant in both blood models. *SOCS3* restrains cytokine signaling through the JAK–STAT pathway[35]. Issarapu et al. examined the region encompassing these same CpGs and demonstrated enhancer activity that was reduced by methylation in HepG2 cells[36]. Lower methylation could therefore accompany a compensatory change in *SOCS3* regulation during sustained cytokine stimulation, placing these associations within the feedback mechanisms that control the duration and intensity of immune responses. The association at *CISH* cg08996521 adds a related link to this interpretation, as experimental deletion of *Cish* increases IL-15 responsiveness and JAK–STAT signaling in NK cells[37].

Other associated loci broaden this cytokine-related interpretation. Three hypomethylated CpGs at *PARP9/DTX3L* connect the profile to STAT1-mediated interferon responses[38], while the shared annotation of cg07362849 to *BST2/BISPR/MVB12A* brings interferon-responsive *BISPR* regulation of *BST2*[39] into consideration. Two hypomethylated *OSCAR* sites also point toward myeloid signaling, given the receptor's ability to mediate surfactant-protein-D-induced TNF release in human monocytes[40]. Across these individual loci, the recurring involvement of immune regulation is consistent with the leukocyte, granule and cytokine processes identified by GO, KEGG and Reactome enrichment. Reports of altered circulating cytokines in CUD[41-43], inflammation and infection in cocaine addiction[44], and cocaine-associated pulmonary complications[45] provide clinical settings in which these processes may be engaged. Together, they frame a hypothesis of altered peripheral immune regulation in the context of repeated substance exposure, associated illness and stress.

The genomic distribution of the findings adds another dimension to this interpretation. Associations extended beyond promoters to OpenSea regions, UTRs and gene bodies, and included the leading hypermethylated locus *FBLN5* cg25132241, linked to extracellular-matrix biology through fibulin-5's role in elastic-fiber organization[46]. Rodent substance-use studies likewise describe methylation changes outside promoters and complex relationships with expression[47]. The predominantly hypomethylated profile thus

encompasses several regulatory contexts whose biological significance depends on the locus and its cellular setting.

Because the physiological processes implicated by these findings can be engaged by overlapping exposures, we also examined how the selected methylation sites related to exposure histories within CUD. Alcohol provided the clearest associations. The two nearby *CDC42BPB* sites surviving joint correction showed higher methylation with greater alcohol-use history and in the case–control EWAS, while the negative within-alcohol association at *SOCS3* cg11047325 connected this exposure to a cytokine-feedback locus already implicated in the diagnostic comparison. Alcohol associations were also enriched relative to matched CpGs. At the same time, the overall case–control pattern persisted among participants reporting no years of regular alcohol abuse, suggesting that the broader diagnostic profile includes variation beyond the alcohol associations detected at individual sites.

Trauma-related exposures were central to the clinical characterization of this cohort: both lifetime PTE exposure and childhood maltreatment were markedly more common among participants with CUD. Examining these histories alongside substance use allowed us to ask whether the burden of adversity tracked variation in the methylation profile associated with diagnosis. Neither exposure yielded a corrected individual-CpG association; lifetime PTE exposure was not enriched relative to matched comparators, and cumulative adversity was not associated with the retained methylation components after correction. These findings help specify the stress-related interpretation of the study. The mechanistic hypothesis involving *FKBP5* arises from its glucocorticoid and inflammatory functions and the physiological context of substance exposure. Relating that hypothesis specifically to trauma will require attention to when adversity occurred and how it preceded subsequent molecular and clinical changes.

The remaining exposure analyses further illustrate why a diagnostic methylation profile may not follow a simple gradient of reported use history. Cannabis, cocaine and nicotine histories, as well as income, lacked corrected CpG-level associations. Nicotine was examined conditional on the methylation-derived smoking score, while duration of regular cocaine use captures a different dimension of the clinical phenotype from CUD diagnosis. Considering these distinctions alongside the alcohol findings helps locate the observed methylation differences within the broader exposure histories of people with CUD.

The peripheral origin of these findings also raises the question of how they relate to methylation in brain tissue. We approached this through two analyses addressing different relationships. BECon describes inter-individual covariation in paired blood and cortical reference samples. Median correlations were close to zero, but the positive BA20 correlation at *FKBP5* cg03546163, alongside weaker or negative correlations in other regions, illustrates the regional variation in correspondence between peripheral and cortical methylation.

Our independent postmortem BA9 analysis then examined whether the blood-selected loci also showed CUD-associated group differences. None of the 388 loci survived FDR correction. Within this exploratory analysis, *ASAP1* cg10068788 showed a concordantly negative nominal association, presented in Figure 4. *ASAP1* regulates membrane and cytoskeletal processes, with functional evidence for dendritic-cell migration[48] and photoreceptor ciliary receptor trafficking[49]. These functions give the locus a biological context for further investigation while the distinction between reference covariation and diagnostic group differences remains essential to interpreting the two tissue comparisons.

The broader brain literature provides a useful setting for these observations. White et al.'s systematic review of 20 postmortem methylome-wide studies across substance use disorders identified recurring inflammatory, signaling and transcriptional processes, together with limited overlap of individual findings and

substantial regional and methodological heterogeneity[50]. Our peripheral immune findings and the exploratory *ASAP1* observation contribute biological processes and a candidate locus to this developing literature. They also make the relationship between peripheral immune regulation and tissue-specific methylation a focused question for subsequent studies of CUD.

Limitations

The cross-sectional design cannot distinguish vulnerability from exposure-related changes or genetic influences[51], and the blood EWAS lacks independent replication. Residual substance-use and cellular confounding remain possible despite adjustment. Secondary analyses selected loci within the discovery cohort and conditioned on CUD, with complete-case sampling and unmodeled CpG dependence in matched comparisons limiting inference. The alcohol subgroup reused discovery participants and controls and cannot quantify alcohol's contribution to the diagnostic contrast. Null exposure findings apply to the selected loci, available measures and retained components, leaving other genomic regions and developmental periods unexamined.

BECon covers only part of the discovery set. The 30-donor BA9 cohort differs in recruitment and population background, and recorded race/ethnicity does not measure genetic ancestry. Tissue specificity, residual ancestry and cellular differences, and limited power cannot be separated as explanations for the absence of corrected associations. The nominal *ASAP1* association weakened after race/ethnicity adjustment and does not establish replication.

Expression, protein abundance and pathway activity were not measured. Global demethylation did not isolate cg03546163, the *SOCS3* reporter assays did not establish individual-CpG effects in blood cells, and photoreceptor trafficking does not establish an *ASAP1* cortical mechanism. Shared annotations also leave transcript attribution unresolved at *BST2*/*BISPR*/*MVB12A*. Functional assessment in CUD is therefore needed: enrichment does not establish pathway activity, and lower methylation has no uniform transcriptional consequence.

In this admixed Brazilian cohort, CUD-associated peripheral methylation connects glucocorticoid regulation with antigen presentation and cytokine feedback. *FKBP5*, *NLRC5–B2M* and *SOCS3–CISH* give this interpretation biological specificity, while the alcohol findings show how exposure histories inform the diagnostic profile. Longitudinal and functional studies can now examine these relationships across vulnerability, exposure-related adaptation, disease course and treatment.

Methods

Sample

Participants were recruited in southern Brazil, with an eligibility age range of 16–65 years. Cases were invited from public inpatient and outpatient addiction treatment programs, and all were abstinent at the time of evaluation. The Structured Clinical Interview for DSM-5—Research Version (SCID-5-RV) was used to confirm CUD diagnosis and assess psychiatric disorders[52]. Controls were blood donors assessed using the non-patient overview of the SCID-5-RV and screened for current substance use disorder, except tobacco. Other psychiatric diagnoses were recorded and are reported in Table 1. Additional exclusion criteria for controls included severe or unstable medical conditions and neurological disorders. Informed consent was obtained from all participants in accordance with the Declaration of Helsinki and Brazilian regulations, with study approval by the Brazilian

National Committee of Ethics in Research (57370016.0.1001.5336/60633516.2.1001.5327) and the Office for Human Research Protections, United States (IRB 00000921).

The primary EWAS included 562 participants with CUD and 679 controls. As a sample-size benchmark, published EPIC-array calculations[53] estimated that 500 cases and 500 controls provide >80% power at 81% of probes to detect an absolute mean β-value difference of 0.02, or two percentage points, using $\alpha = 9.42 \times 10^{-8}$. This literature-based estimate provides context for the sample size; power in the present study depends on CpG-specific variability and the covariate-adjusted model.

DNA Methylation

For each participant, 10 mL of peripheral whole blood was collected by venipuncture into EDTA tubes (Becton Dickinson), and DNA was isolated with the DNeasy Blood & Tissue Kit (Qiagen). DNA was quantified with a Nanodrop spectrophotometer (Thermo). Bisulfite conversion of 450 ng of DNA was performed with the EZ DNA Methylation Kit (Zymo), followed by hybridization to the Infinium MethylationEPIC v1.0 BeadChip (Illumina), according to the manufacturer's instructions. Samples were systematically distributed across BeadChips to balance cases and controls within each array and minimize batch effects. Investigators were blinded to diagnostic status during preprocessing and normalization, with group labels revealed only after quality control. Chips were scanned in an iScan microarray reader (Illumina).

Association analyses were performed in R 4.5.2 with CpGassoc 2.70, using the preprocessed methylation matrix and updated clinical annotation. Quality control and filtering used minfi[54], retaining samples with mean detection $P < 0.05$ and probes with detection $P < 0.01$ in every retained sample. Probes with low bead count (<3 in >1% of samples), cross-reactive probes[55], probes associated with single nucleotide polymorphisms[55], and probes on sex chromosomes were excluded. Functional normalization included noob background correction, dye normalization and correction for the first two principal components on the array[56]. Preprocessing from IDAT files was not repeated for the revised association analyses. The final matrix contained 627,130 autosomal CpGs in 1,245 participants. Beta-values of zero and one were replaced by 0.0001 and 0.9999, respectively, before conversion to M-values as $\log_2[\text{beta}/(1 - \text{beta})]$ for statistical analysis[57].

Blood-cell proportions—CD8+ (CD8T) and CD4+ (CD4T) T-lymphocytes, natural killer cells (NK), B-lymphocytes (Bcell), monocytes (Mono) and neutrophils—were estimated using EpiDISH[58-63]. A methylation-derived estimate of smoking exposure was obtained using EpiSmokEr[64,65], and EPISTRUCTURE was used to estimate ancestry-related methylation structure for all participants[66]. These estimates informed the covariate adjustment described below.

Clinical and sociodemographic variables

Sociodemographic measures included age, biological sex, self-reported race/ethnicity, years of education, monthly gross income and occupational status. Income was recorded in original Brazilian reais (BRL), without inflation adjustment. Race/ethnicity was assessed using the five official categories defined by the Brazilian Institute of Geography and Statistics (IBGE): White (Branca), Black (Preta), Brown (Parda, referring to mixed ancestry), Yellow (Amarela, typically East Asian descent) and Indigenous (Indígena). Occupational categories were employed, unemployed, student, homemaker, retired or disabled; records with all category flags equal to zero were retained as having no listed category recorded. Marital status was not summarized because the available

CUD category indicators were internally inconsistent. Categorical variables are reported as frequency (%) in CUD/control order in Table 1.

Clinical variables, including substance use disorders and current mood disorders, were assessed using the SCID-5-RV. Substance use severity within the CUD group was evaluated with the Summary Scores for Recent Functioning (SS-Rs) derived from the Addiction Severity Index, 6th Edition (ASI-6), applying Nonparametric Item Response Theory models[67,68]. In the within-CUD methylation analyses, substance-use history was represented by years of regular use divided by age, calculated separately for alcohol, nicotine, cannabis and cocaine.

Exposure to PTEs was defined as having experienced at least one event involving risk of death, violent physical harm, rape or witnessing extreme violence. Controls completed the Life Events Checklist for DSM-5 (LEC-5), whereas participants with CUD were assessed using the ASI-6[67,68]. The ASI-6 items covered rape, other violent physical assaults such as beatings, life-threatening situations and witnessing extreme violence. For controls, LEC-5 items 6, 7, 8 and 11 captured physical aggression, assault with a weapon, rape and kidnapping, with information on whether events were witnessed. For within-CUD analyses, lifetime PTE exposure was defined as the count of four reported event domains (range 0–4), requiring complete binary responses.

Childhood maltreatment was assessed with the Childhood Trauma Questionnaire (CTQ). In participants with CUD, each domain was reconstructed from its five items, requiring valid responses from 1 to 5 and reverse-scoring the specified items. Severe-domain thresholds were ≥13 for physical abuse, ≥13 for sexual abuse, ≥16 for emotional abuse, ≥13 for physical neglect and ≥18 for emotional neglect. Table 1 reports domains for CUD participants with all five domain items valid; controls were classified using available integer domain totals within the valid range of 5–25 and the same severe thresholds. Raw item coverage was insufficient for equivalent item-level verification in controls.

For within-CUD methylation analyses, childhood maltreatment was analyzed separately from lifetime PTE exposure as the count of five severe CTQ domains (range 0–5). All 25 scored items were required, and missing or invalid responses were not assigned zero. The supplementary PCA used a distinct cumulative adversity index combining the four lifetime event domains with severe physical, sexual and emotional childhood abuse (range 0–7), requiring complete valid data for these seven domains.

Exploring the link between peripheral blood and brain methylation

The primary objective was to characterize peripheral blood methylation associated with CUD. For a descriptive reference comparison, we queried all 420 primary EWAS CpG identifiers in the BECon web application[20] on 19 September 2026. BECon reports Spearman correlations from paired blood and cortical BA7, BA10 and BA20 samples from 16 individuals. The query returned complete regional coefficients for 136 loci; the remaining 284 were recorded as unavailable, not assigned zero correlations. All coefficients and coverage of the queried set are reported in Supplementary Table 9. Supplementary Figure 3 shows the distributions across all available loci and, for illustration, the 12 highest-ranked blood EWAS loci with reference data, selected by blood EWAS *P* value rather than correlation magnitude. No inferential significance threshold was applied to these reference correlations. The independent postmortem case–control analysis described below addressed a separate question. For each CpG, these coefficients indicate whether individuals with higher blood methylation tend to have higher or lower methylation at the corresponding cortical site.

We then examined CUD-associated differences in frozen postmortem dorsolateral prefrontal cortex, Brodmann area 9 (BA9), from 13 donors with CUD and 17 non-psychiatric controls. Samples were obtained through the UTHealth Brain Collection for Research in Psychiatric Disorders in collaboration with the Harris County Institute of Forensic Science (Supplementary Table 7). Recorded race/ethnicity was classified as White or non-White: 11 of 13 donors with CUD (84.6%) and 9 of 17 controls (52.9%) were recorded as non-White. This classification did not measure genetic ancestry. Consent for brain donation and use in the protocol was obtained from the next-of-kin. Diagnosis was confirmed through a psychological autopsy interview with the next-of-kin[69], toxicology screening and review of relevant electronic medical records. Brain processing and BA9 dissection have been described previously[70]. DNA was isolated using the DNeasy Blood & Tissue Kit, and methylation was assessed using the same laboratory methods as for blood.

Statistical Analyses

All group comparisons used two-sided tests unless otherwise specified. Table 1 comparisons used Mann–Whitney tests for continuous variables and Pearson chi-square tests without continuity correction for categorical variables, with Fisher exact tests when any expected cell count was below five. Race/ethnicity and occupation were each evaluated with one global test. These descriptive *P* values were not adjusted for multiple comparisons.

DMPs were identified using linear mixed-effects models of M-values, with CUD status as the exposure of interest and Sentrix chip ID as a random intercept. Fixed covariates were age, sex, CD4T, CD8T, natural killer cell, B-cell and monocyte proportions, EPISTRUCTURE PCs 1–10, numeric array position and the methylation-derived smoking score. Array positions were coded as ordered numeric levels 1–8. The primary smoking-adjusted model included 1,241 participants (562 CUD and 679 controls); the comparison model omitting only the smoking score included 1,245 participants (562 CUD and 683 controls).

As a descriptive assessment of ancestry-related overlap, we compared the final DMP sets with the catalog of Galanter et al.[71], which identified 916 ethnicity-associated CpGs in an admixed Latino cohort, including subsets classified by their associations after adjustment for genetic ancestry. Of these, 708 were measured in our final dataset; none overlapped either the 420 smoking-adjusted DMPs or the 107 DMPs identified without smoking-score adjustment. This comparison does not exclude residual confounding by ancestry or correlated exposures (Supplementary Table 9).

We assessed multicollinearity among CUD status and the fixed-effect covariates using variance inflation factors (VIFs), calculated separately in each model's complete-case sample as $1/(1 - R^2)$. R^2 was obtained by regressing each predictor on all remaining fixed-effect predictors; the chip random intercept was not included as a fixed predictor in this diagnostic. Both design matrices were full rank. Maximum VIFs were 2.20 with smoking-score adjustment and 1.75 without it, and the smoking-score VIF was 1.92 (Supplementary Table 8).

EWAS scripts were adapted from those shared by the PGC PTSD EWAS Working Group[72]. For blood CpG discovery, we followed the empirical EPIC-array recommendation of Mansell et al.[53], applying $P < 9 \times 10^{-8}$ to two-sided BACON-calibrated *P* values in both models. This threshold is distinct from a Bonferroni calculation based on the 627,130 probes tested here. To assess and adjust for bias and inflation of the test statistics, we calculated the genomic inflation factor (λ), generated quantile–quantile (Q–Q) plots before and after calibration, and applied Bayesian adjustment with BACON[73]. BACON calibrates the empirical null for bias and inflation; it is separate from the criterion used to address multiple testing. Raw model and BACON-calibrated estimates and *P* values are distinguished in Supplementary Tables 1 and 2.

Secondary analyses used Benjamini–Hochberg FDR < 0.05 within the test families specified below. Nominal $P < 0.05$ was used only for descriptive comparisons or as input to exploratory set-level analyses and was not interpreted as multiplicity-adjusted significance.

Within-CUD exposure analyses tested each of the 420 CpGs identified at BACON-calibrated $P < 9 \times 10^{-8}$ in the primary EWAS. Seven exposures were analyzed separately: the four-domain lifetime PTE count, the five-domain severe-childhood-maltreatment count, four ratios of years of regular use to age (alcohol, nicotine, cannabis and cocaine), and monthly income in original BRL, scaled per BRL 1,000. M-values were analyzed with REML linear mixed models, using age, sex, CD8T, CD4T, natural killer cell, B-cell and monocyte proportions, EPISTRUCTURE PCs 1–10, numeric array position and methylation-derived smoking score as fixed covariates, and Sentrix chip ID as a random intercept. These were the nuisance covariates in the primary EWAS. Two-sided model *P* values underwent BH correction within each exposure across 420 loci and, separately, across all 2,940 target-CpG tests spanning the seven exposures. Results for these correction families were reported separately, with $q < 0.05$ defining significance. BACON was not re-estimated from the selected loci. Pointwise 95% confidence intervals were based on the model t distribution. These analyses were exploratory within the discovery cohort. Complete-case samples comprised 562 participants for substance-use histories and income, 526 for lifetime PTE exposure and 513 for childhood maltreatment. Nicotine associations were conditional on the methylation-derived smoking score.

To compare the frequency of exposure associations with that at other assayed loci, we selected 2,100 unique comparator CpGs at a ratio of five per target, without replacement. Matching used the mean and standard deviation of beta-values across the 562 CUD participants, excluding the union of 455 CpGs identified in the two case–control models. Selection did not use exposure-association results. Absolute calipers of 0.02 for the mean and 0.01 for the standard deviation were applied, followed by nearest distance in standardized moments. The same exposure models were fitted at the comparator CpGs, with BH correction across 2,100 loci per exposure and, separately, across 14,700 tests spanning all seven exposures.

For exploratory set-level comparisons, proportions of nominal associations at $P < 0.05$ were compared using two-sided Fisher exact tests, with conditional maximum-likelihood odds ratios, exact 95% confidence intervals and BH correction across seven comparisons. These comparisons do not account for correlation among CpGs and were treated as descriptive exploratory evidence. Matching on methylation mean and variability does not establish equivalent power or eliminate bias from selecting loci and conditioning on CUD.

The supplementary PCA examined cumulative adversity in relation to broader dimensions of variation across the selected loci. M-values at the 420 CpGs were centered and standardized in 498 participants with complete valid adversity and covariate data. Before association testing, we retained the smallest number of components explaining at least 80% of variance. Each component score was standardized to unit variance and regressed on the seven-domain adversity index using the same fixed nuisance covariates and chip random intercept. Sensitivity models added the four substance-use ratios and monthly income while retaining the same participants and PCA basis. BH correction was applied across the ten retained components separately for each specification; an additional correction across all 20 tests was also reported. Complete results are provided in Supplementary Table 5.

For the alcohol subgroup analysis, we restricted CUD to 256 participants reporting no years of regular alcohol abuse and retained the same 679 controls. We fitted the same smoking-adjusted mixed model at the 420 selected loci and compared uncalibrated coefficients using direction agreement, Pearson correlation and

subgroup/full-sample ratios. BACON was not re-estimated from these selected loci; raw two-sided t-test *P* values, pointwise 95% t intervals and BH adjustments across 420 CpGs are reported in Supplementary Table 6.

For the postmortem analysis, we tested the 388 CpGs available in the brain matrix among the 420 blood-selected loci; the remaining 32 were not tested. At each CpG, beta-values were transformed to normal scores using the shrunken empirical cumulative distribution (average rank divided by n + 1), followed by Gaussian quantiles and scaling to unit sample standard deviation across the 30 donors. Linear regression models included group membership, age, sex, current-smoking probability, postmortem interval (PMI) and brain tissue pH. A sensitivity model additionally included recorded binary race/ethnicity (White/non-White). BH correction was applied across all 388 CpGs, separately for the primary and sensitivity models, with significance defined as FDR < 0.05. Supplementary Table 7 reports nominal and BH-adjusted *P* values, effect estimates and 95% confidence intervals for every tested CpG. Brain coefficients represent CUD minus control differences on the normal-score scale and are not directly comparable in magnitude to blood M-value coefficients.

Bioinformatic Annotation and Functional Enrichment Analysis

All 627,130 tested CpGs were annotated using IlluminaHumanMethylationEPICanno.ilm10b4.hg19 (version 0.6.0). For each annotation category, a two-sided Fisher exact test compared membership among selected DMPs with membership among the remaining tested CpGs. Cross-product odds ratios and pointwise Woolf 95% confidence intervals were calculated from the corresponding 2 × 2 tables. BH correction was applied separately across six CpG-island contexts and eight gene-region categories within each EWAS model, with adjusted $P < 0.05$ defining enrichment or depletion. A CpG with multiple distinct gene-region annotations contributed to each category; those without a gene-region annotation were classified as unannotated.

GO, KEGG and complementary Reactome gene-set analyses used methylglm from methylGSA[74] (version 1.26.0) in R 4.5.2. We supplied BACON-calibrated *P* values for all 627,130 CpGs tested in the smoking-adjusted EWAS, without filtering by significance or separating methylation directions. Probe-to-gene mapping used the Infinium MethylationEPIC v1/hg19 annotation (IlluminaHumanMethylationEPICanno.ilm10b4.hg19, version 0.6.0) and methylGSA's default procedure, which retains probes with 10-character identifiers and non-empty gene annotations and assigns each probe to the first listed UCSC gene. This yielded 446,392 mapped CpGs representing 24,646 genes.

For each gene, the association statistic was the negative natural logarithm of the smallest *P* value among its mapped CpGs. Gene-set membership was modeled with quasibinomial regression using a logit link, with the association statistic and the natural logarithm of the number of analyzed CpGs mapped to the gene as predictors. Including probe coverage accounts for the greater opportunity to observe a small *P* value at genes represented by more CpGs. All 24,646 represented genes formed the analysis universe. Enrichment was evaluated with methylglm's default procedure, which assigns $P = 1$ to non-positive association-statistic coefficients.

GO annotations were obtained from org.Hs.eg.db and GO.db (both version 3.21.0). Human KEGG pathway definitions were retrieved on 18 September 2026, and gene identifiers were converted to symbols using org.Hs.eg.db. Reactome[75] version 97 human pathway definitions were downloaded from the official GMT release on 19 September 2026. Gene sets were deduplicated by gene symbol and restricted to 25–500 unique genes before intersection with the analysis universe; sets with no represented genes were excluded. The Reactome analysis reused the same CpG *P* values, mapping and gene universe as GO and KEGG. BH correction was applied across all 4,984 tested GO terms together, encompassing biological process, cellular component and molecular function,

and separately across all 322 KEGG and 1,091 Reactome pathways. Significance was defined as FDR < 0.05. Complete results for each database are provided in Supplementary Table 4.

Use of generative AI

Generative AI tools (OpenAI ChatGPT and Codex) assisted with code development and debugging, analytical auditing and documentation, and manuscript revision. The authors take responsibility for the methodological decisions, interpretation and reported results.

Data availability

Due to ethical and legal restrictions under Brazilian regulations governing the sharing of clinical and personally identifiable information, individual-level clinical data cannot be made publicly available. Access to de-identified participant-level data for scientifically valid research purposes can be granted upon reasonable request, subject to approval by the institutional ethics committee and the data governance board of the Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Requests should be directed to the corresponding author, who will facilitate the review and authorization process in compliance with all applicable ethical and legal requirements. Complete CpG-level EWAS summary statistics for both blood models are deposited in the Figshare record linked to this submission; Supplementary Table 2 provides the file descriptions and data dictionary. Selected-locus and secondary-analysis results are provided in the remaining supplementary tables. Individual-level methylation measurements are not included in these summary tables; access requests should follow the review and authorization process described above.

Acknowledgments

We are grateful to Associação Educadora São Carlos and its entire staff for providing the facilities and support that made data collection possible. We extend our special thanks to Julio Carlos Pezzi, whose psychiatric expertise and technical support were invaluable during data collection. We also thank all students involved in data collection over the years, especially Leonardo Mello Rothmann, Bruna Santos da Silva and Breno Sanvicente-Vieira for their contributions to the initial stages of database development.

Funding

This work was supported by the National Institute on Drug Abuse of the US National Institutes of Health (grant R01DA044859) and the Brazilian National Council for Scientific and Technological Development (CNPq; grants 306580/2022-5 and 466802/2014-5).

Author contributions

Rodrigo Grassi-Oliveira contributed to conceptualization, supervision, data collection, analyses, funding acquisition, and initial and final manuscript writing. Gabriel R. Fries and Bruno Kluwe-Schiavon contributed to analyses and initial manuscript writing. Thiago Wendt Viola contributed to analyses and initial and final manuscript writing. Diego Luiz Rovaris contributed to data collection and initial manuscript writing. João Antônio Valadão Leal contributed to analyses and initial manuscript writing. Eugenio Horacio Grevet and Claiton Henrique Dotto Bau contributed to data collection and initial manuscript writing. Joy M. Schmitz contributed to funding acquisition. Consuelo Walss-Bass contributed to supervision, analyses and initial manuscript writing.

References

1 Schwartz, E. K. C., Wolkowicz, N. R., De Aquino, J. P., MacLean, R. R. & Sofuoglu, M. Cocaine Use Disorder (CUD): Current Clinical Perspectives. *Substance Abuse and Rehabilitation* **13**, 25–46 (2022). https://doi.org/10.2147/sar.s337338

2 United Nations Office on Drugs and Crime. World Drug Report 2026: Drug Market Patterns and Trends. (United Nations Office on Drugs and Crime, 2026).

3 Tondo, L. P. *et al.* White matter deficits in cocaine use disorder: convergent evidence from in vivo diffusion tensor imaging and ex vivo proteomic analysis. *Translational Psychiatry* **11**, 252 (2021). https://doi.org/10.1038/s41398-021-01367-x

4 Rothmann, L. M. *et al.* The cortical thickness of tricenarian cocaine users assembles features of an octogenarian brain. *Journal of Neuroscience Research* **102**, e25287 (2024). https://doi.org/10.1002/jnr.25287

5 Sun, J., Kranzler, H. R., Gelernter, J. & Bi, J. A genome-wide association study of cocaine use disorder accounting for phenotypic heterogeneity and gene–environment interaction. *Journal of Psychiatry and Neuroscience* **45**, 34–44 (2020). https://doi.org/10.1503/jpn.180098

6 Cabana-Domínguez, J., Shivalikanjli, A., Fernàndez-Castillo, N. & Cormand, B. Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **94**, 109667 (2019). https://doi.org/10.1016/j.pnpbp.2019.109667

7 Fernàndez-Castillo, N., Cabana-Domínguez, J., Corominas, R. & Cormand, B. Molecular genetics of cocaine use disorders in humans. *Molecular Psychiatry* **27**, 624–639 (2022). https://doi.org/10.1038/s41380-021-01256-1

8 Massart, R. *et al.* Role of DNA Methylation in the Nucleus Accumbens in Incubation of Cocaine Craving. *The Journal of Neuroscience* **35**, 8042–8058 (2015). https://doi.org/10.1523/jneurosci.3053-14.2015

9 Zillich, E. *et al.* Multi-omics profiling of DNA methylation and gene expression alterations in human cocaine use disorder. *Translational Psychiatry* **14**, 428 (2024). https://doi.org/10.1038/s41398-024-03139-9

10 Camilo, C. *et al.* Genome-wide DNA methylation profile in the peripheral blood of cocaine and crack dependents. *Brazilian Journal of Psychiatry* **41**, 485–493 (2019). https://doi.org/10.1590/1516-4446-2018-0092

11 Shu, C. *et al.* Epigenome-wide association analyses of active injection drug use. *Drug and Alcohol Dependence* **235**, 109431 (2022). https://doi.org/10.1016/j.drugalcdep.2022.109431

12 Viola, T. W. *et al.* Prenatal cocaine exposure and its influence on pediatric epigenetic clocks and epigenetic scores in humans. *Scientific Reports* **14**, 1946 (2024). https://doi.org/10.1038/s41598-024-52433-5

13 McCabe, S. E., West, B. T., Jutkiewicz, E. M. & Boyd, C. J. Multiple DSM-5 substance use disorders: A national study of US adults. *Human Psychopharmacology: Clinical and Experimental* **32**, e2625 (2017). https://doi.org/10.1002/hup.2625

14 Han, B. *et al.* Polysubstance use disorders among US adults. *Molecular Psychiatry* **31**, 5303–5315 (2026). https://doi.org/10.1038/s41380-026-03618-z

15 Sanvicente-Vieira, B. *et al.* Sex-based differences in multidimensional clinical assessments of early-abstinence crack cocaine users. *PLOS ONE* **14**, e0218334 (2019). https://doi.org/10.1371/journal.pone.0218334

16 Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circulation: Cardiovascular Genetics* **9**, 436–447 (2016). https://doi.org/10.1161/circgenetics.116.001506

17 Hughes, K., Ford, K., Kadel, R., Sharp, C. A. & Bellis, M. A. Health and financial burden of adverse childhood experiences in England and Wales: a combined primary data study of five surveys. *BMJ Open* **10**, e036374 (2020). https://doi.org/10.1136/bmjopen-2019-036374

18 Carliner, H. *et al.* Childhood Trauma and Illicit Drug Use in Adolescence: A Population-Based National Comorbidity Survey Replication–Adolescent Supplement Study. *Journal of the American Academy of Child & Adolescent Psychiatry* **55**, 701–708 (2016). https://doi.org/10.1016/j.jaac.2016.05.010

19 Afful, S. E., Strickland, J. R., Cottler, L. & Bierut, L. J. Exposure to trauma: A comparison of cocaine-dependent cases and a community-matched sample. *Drug and Alcohol Dependence* **112**, 46–53 (2010). https://doi.org/10.1016/j.drugalcdep.2010.05.012

20 Edgar, R. D., Jones, M. J., Meaney, M. J., Turecki, G. & Kobor, M. S. BECon: a tool for interpreting DNA methylation findings from blood in the context of brain. *Translational Psychiatry* **7**, e1187 (2017). https://doi.org/10.1038/tp.2017.171

21 López, A. J., Siciliano, C. A. & Calipari, E. S. Activity-Dependent Epigenetic Remodeling in Cocaine Use Disorder. *Handbook of Experimental Pharmacology* **258**, 231–263 (2020). https://doi.org/10.1007/164_2019_257

22 Bravo-Gutiérrez, O. A. *et al.* Hypomethylation of AHRR (cg05575921) Is Related to Smoking Status in the Mexican Mestizo Population. *Genes* **12**, 1276 (2021). https://doi.org/10.3390/genes12081276

23 Grieshober, L. *et al.* AHRR methylation in heavy smokers: associations with smoking, lung cancer risk, and lung cancer mortality. *BMC Cancer* **20**, 905 (2020). https://doi.org/10.1186/s12885-020-07407-x

24 Skov-Jeppesen, S. M., Kobylecki, C. J., Jacobsen, K. K. & Bojesen, S. E. Changing Smoking Behavior and Epigenetics. *CHEST* **163**, 1565–1575 (2023). https://doi.org/10.1016/j.chest.2022.12.036

25 Bojesen, S. E., Timpson, N., Relton, C., Davey Smith, G. & Nordestgaard, B. G. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax* **72**, 646–653 (2017). https://doi.org/10.1136/thoraxjnl-2016-208789

26 Fang, F. *et al.* Trans-ancestry epigenome-wide association meta-analysis of DNA methylation with lifetime cannabis use. *Molecular Psychiatry* **29**, 124–133 (2024). https://doi.org/10.1038/s41380-023-02310-w

27 Zillich, L. *et al.* A large-scale meta-analysis of DNA methylation signatures of alcohol use disorder in the Psychiatric Genomics Consortium. *Nature Mental Health* **4**, 1088–1098 (2026). https://doi.org/10.1038/s44220-026-00666-w

28 Fries, G., Gassen, N. & Rein, T. The FKBP51 Glucocorticoid Receptor Co-Chaperone: Regulation, Function, and Implications in Health and Disease. *International Journal of Molecular Sciences* **18**, 2614 (2017). https://doi.org/10.3390/ijms18122614

29 Heesch, C. M. *et al.* Effects of Cocaine on Cortisol Secretion in Humans. *The American Journal of the Medical Sciences* **310**, 61–64 (1995). https://doi.org/10.1097/00000441-199508000-00004

30 Yu, H. *et al.* Epigenome-wide association study identifies Behçet's disease-associated methylation loci in Han Chinese. *Rheumatology* **58**, 1574–1584 (2019). https://doi.org/10.1093/rheumatology/kez043

31 Zannas, A. S. *et al.* Epigenetic upregulation of FKBP5 by aging and stress contributes to NF-κB–driven inflammation and cardiovascular risk. *Proceedings of the National Academy of Sciences* **116**, 11370–11379 (2019). https://doi.org/10.1073/pnas.1816847116

32 Kobayashi, K. S. & van den Elsen, P. J. NLRC5: a key regulator of MHC class I-dependent immune responses. *Nature Reviews Immunology* **12**, 813–820 (2012). https://doi.org/10.1038/nri3339

33 Meissner, T. B. *et al.* NLRC5 Cooperates with the RFX Transcription Factor Complex To Induce MHC Class I Gene Expression. *The Journal of Immunology* **188**, 4951–4958 (2012). https://doi.org/10.4049/jimmunol.1103160

34 Zaparte, A. *et al.* Cocaine Use Disorder Is Associated With Changes in Th1/Th2/Th17 Cytokines and Lymphocytes Subsets. *Frontiers in Immunology* **10**, 2435 (2019). https://doi.org/10.3389/fimmu.2019.02435

35 Carow, B. & Rottenberg, M. E. SOCS3, a Major Regulator of Infection and Inflammation. *Frontiers in Immunology* **5**, 58 (2014). https://doi.org/10.3389/fimmu.2014.00058

36 Issarapu, P. *et al.* DNA methylation at the suppressor of cytokine signaling 3 (SOCS3) gene influences height in childhood. *Nature Communications* **14**, 5200 (2023). https://doi.org/10.1038/s41467-023-40607-0

37 Delconte, R. B. *et al.* CIS is a potent checkpoint in NK cell–mediated tumor immunity. *Nature Immunology* **17**, 816–824 (2016). https://doi.org/10.1038/ni.3470

38 Zhang, Y. *et al.* PARP9-DTX3L ubiquitin ligase targets host histone H2BJ and viral 3C protease to enhance interferon signaling and control viral infection. *Nature Immunology* **16**, 1215–1227 (2015). https://doi.org/10.1038/ni.3279

39 Kambara, H. *et al.* Regulation of Interferon-Stimulated Gene BST2 by a lncRNA Transcribed from a Shared Bidirectional Promoter. *Frontiers in Immunology* **5**, 676 (2014). https://doi.org/10.3389/fimmu.2014.00676

40 Barrow, A. D. *et al.* OSCAR Is a Receptor for Surfactant Protein D That Activates TNF-α Release from Human CCR2+ Inflammatory Monocytes. *The Journal of Immunology* **194**, 3317–3326 (2015). https://doi.org/10.4049/jimmunol.1402289

41 Stamatovich, S. N. *et al.* Plasma pro- and anti-inflammatory cytokines may relate to cocaine use, cognitive functioning, and depressive symptoms in cocaine use disorder. *The American Journal of Drug and Alcohol Abuse* **47**, 52–64 (2021). https://doi.org/10.1080/00952990.2020.1828439

42 Moreira, F. P. *et al.* Cocaine abuse and effects in the serum levels of cytokines IL-6 and IL-10. *Drug and Alcohol Dependence* **158**, 181–185 (2016). https://doi.org/10.1016/j.drugalcdep.2015.11.024

43 Fox, H. C. *et al.* Immune system inflammation in cocaine dependent individuals: implications for medications development. *Human Psychopharmacology: Clinical and Experimental* **27**, 156–166 (2012). https://doi.org/10.1002/hup.1251

44 Ersche, K. D. & Döffinger, R. Inflammation and infection in human cocaine addiction. *Current Opinion in Behavioral Sciences* **13**, 203–209 (2017). https://doi.org/10.1016/j.cobeha.2016.12.007

45 Restrepo, C. S. *et al.* Pulmonary Complications from Cocaine and Cocaine-based Substances: Imaging Manifestations. *RadioGraphics* **27**, 941–956 (2007). https://doi.org/10.1148/rg.274065144

46 Yanagisawa, H. *et al.* Fibulin-5 is an elastin-binding protein essential for elastic fibre development in vivo. *Nature* **415**, 168–171 (2002). https://doi.org/10.1038/415168a
47 Diringer, M., Bruggeman, M. & Lutz, P.-E. The DNA methylation enzymatic machinery in substance use disorders: A systematic review. *Neurobiology of Disease* **218**, 107241 (2026). https://doi.org/10.1016/j.nbd.2025.107241
48 Curtis, J. *et al.* Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nature Genetics* **47**, 523–527 (2015). https://doi.org/10.1038/ng.3248
49 Wang, J. & Deretic, D. The Arf and Rab11 effector FIP3 acts synergistically with ASAP1 to direct Rabin8 in ciliary receptor targeting. *Journal of Cell Science* **128**, 1375–1385 (2015). https://doi.org/10.1242/jcs.162925
50 White, J. D. *et al.* A systematic review of substance use disorder methylome-wide association studies in postmortem human brain. *Translational Psychiatry* (2026). https://doi.org/10.1038/s41398-026-04375-x
51 Villicaña, S. & Bell, J. T. Genetic impacts on DNA methylation: research findings and future perspectives. *Genome Biology* **22**, 127 (2021). https://doi.org/10.1186/s13059-021-02347-6
52 First, M. B., Williams, J. B. W., Karg, R. S. & Spitzer, R. L. *Structured Clinical Interview for DSM-5—Research Version (SCID-5-RV).* (American Psychiatric Association, 2015).
53 Mansell, G. *et al.* Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics* **20**, 366 (2019). https://doi.org/10.1186/s12864-019-5761-7
54 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014). https://doi.org/10.1093/bioinformatics/btu049
55 McCartney, D. L. *et al.* Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data* **9**, 22–24 (2016). https://doi.org/10.1016/j.gdata.2016.05.012
56 Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology* **15**, 503 (2014). https://doi.org/10.1186/s13059-014-0503-2
57 Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010). https://doi.org/10.1186/1471-2105-11-587
58 Zheng, S. C. *et al.* A Novel Cell-Type Deconvolution Algorithm Reveals Substantial Contamination by Immune Cells in Saliva, Buccal and Cervix. *Epigenomics* **10**, 925–940 (2018). https://doi.org/10.2217/epi-2018-0037
59 Zheng, S. C., Breeze, C. E., Beck, S. & Teschendorff, A. E. Identification of differentially methylated cell types in epigenome-wide association studies. *Nature Methods* **15**, 1059–1066 (2018). https://doi.org/10.1038/s41592-018-0213-x
60 Teschendorff, A. E., Breeze, C. E., Zheng, S. C. & Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics* **18**, 105 (2017). https://doi.org/10.1186/s12859-017-1511-5
61 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012). https://doi.org/10.1186/1471-2105-13-86
62 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* **12**, 453–457 (2015). https://doi.org/10.1038/nmeth.3337
63 Teschendorff, A. E. & Zheng, S. C. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics* **9**, 757–768 (2017). https://doi.org/10.2217/epi-2016-0153
64 Bollepalli, S., Korhonen, T., Kaprio, J., Anders, S. & Ollikainen, M. Epismoker: A Robust Classifier to Determine Smoking Status from DNA Methylation Data. *Epigenomics* **11**, 1469–1486 (2019). https://doi.org/10.2217/epi-2019-0206
65 Elliott, H. R. *et al.* Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clinical Epigenetics* **6**, 4 (2014). https://doi.org/10.1186/1868-7083-6-4

66 Rahmani, E. *et al.* Genome-wide methylation data mirror ancestry information. *Epigenetics & Chromatin* **10**, 1 (2017). https://doi.org/10.1186/s13072-016-0108-y

67 Denis, C. M., Cacciola, J. S. & Alterman, A. I. Addiction Severity Index (ASI) summary scores: Comparison of the Recent Status Scores of the ASI-6 and the Composite Scores of the ASI-5. *Journal of Substance Abuse Treatment* **45**, 444–450 (2013). https://doi.org/10.1016/j.jsat.2013.06.003

68 Kessler, F. *et al.* Psychometric properties of the sixth version of the Addiction Severity Index (ASI-6) in Brazil. *Revista Brasileira de Psiquiatria* **34**, 24–33 (2012). https://doi.org/10.1590/s1516-44462012000100006

69 Meyer, T. D., Godfrey, C. J. & Walss-Bass, C. The UT health Psychological Autopsy Interview Schedule (UTH- PAIS) – Description and reliability of diagnoses and transdiagnostic personality measures. *Journal of Psychiatric Research* **156**, 221–227 (2022). https://doi.org/10.1016/j.jpsychires.2022.10.014

70 Cabrera-Mendoza, B. *et al.* Within subject cross-tissue analyzes of epigenetic clocks in substance use disorder postmortem brain and blood. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **192**, 13–27 (2023). https://doi.org/10.1002/ajmg.b.32920

71 Galanter, J. M. *et al.* Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *eLife* **6**, e20532 (2017). https://doi.org/10.7554/elife.20532

72 Katrinli, S. *et al.* Epigenome-wide association studies identify novel DNA methylation sites associated with PTSD: a meta-analysis of 23 military and civilian cohorts. *Genome Medicine* **16**, 147 (2024). https://doi.org/10.1186/s13073-024-01417-1

73 van Iterson, M., van Zwet, E. W., BIOS Consortium & Heijmans, B. T. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biology* **18**, 19 (2017). https://doi.org/10.1186/s13059-016-1131-9

74 Ren, X. & Kuan, P. F. methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinformatics* **35**, 1958–1959 (2019). https://doi.org/10.1093/bioinformatics/bty892

75 Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Research* **52**, D672–D678 (2024). https://doi.org/10.1093/nar/gkad1025

FIGURE LEGENDS

Figure 1. Directional blood EWAS Miami plots with and without smoking-score adjustment. (A) Model without the smoking score (n = 1,245; 107 DMPs: 17 hypermethylated and 90 hypomethylated). (B) Primary smoking-adjusted model (n = 1,241; 420 DMPs: 84 hypermethylated and 336 hypomethylated). Each panel displays all 627,130 autosomal CpGs, with positive BACON-calibrated CUD-minus-control coefficients above the chromosome axis and negative coefficients below it. Vertical distance from the center represents $-\log_{10}(P)$ for two-sided BACON-calibrated P values, using the same scale in both panels. Raw and calibrated coefficients agree in direction at all significant loci. Red dashed lines mark $P < 9 \times 10^{-8}$. Colors identify chromosomes 1–22. Chromosomes occupy equal-width blocks, with hg19 positions rescaled within each chromosome; horizontal distances are not comparable in base pairs across chromosomes. The eight significant CpGs with the smallest P values per direction in each model are labeled using gene annotations or, when unavailable, probe identifiers. Multiple gene annotations are separated by slashes. Both models retain the other covariates and chip random intercept. Full statistics are provided in Supplementary Table 2.

Figure 2. Genomic context of smoking-adjusted CUD-associated DMPs.

Enrichment and depletion among the 420 DMPs are shown for (A) CpG-island context and (B) gene region. Points represent $\log_2$ odds ratios and horizontal lines show pointwise Woolf 95% confidence intervals, comparing DMPs with the remaining tested CpGs in the 627,130-probe universe. The dashed vertical line indicates OR = 1. Colors identify enrichment or depletion at Benjamini–Hochberg FDR < 0.05, calculated separately within each panel; gray indicates FDR ≥ 0.05. Labels report DMP counts in each category. Distinct gene-region annotations for the same CpG are retained, so gene-region counts may overlap. DMPs were selected at BACON-calibrated $P < 9 \times 10^{-8}$. Complete results for both EWAS models are provided in Supplementary Table 3.

Figure 3. GO and KEGG gene-set enrichment in the smoking-adjusted EWAS.

The 20 Gene Ontology (GO) terms (a) and 20 KEGG pathways (b) with the lowest Benjamini-Hochberg false discovery rate (FDR) values are shown. Analyses used methylglm with BACON-corrected P values for all 627,130 tested CpGs, without a significance filter or separation by methylation direction, and accounted for the number of analyzed CpGs mapped to each gene. FDR correction was applied across all 4,984 tested GO terms and separately across all 322 tested KEGG pathways; all displayed sets had FDR < 0.05. Horizontal positions indicate -log10(FDR), with different horizontal-axis ranges in the two panels. Blue and orange denote GO and KEGG, respectively. Circle area is proportional to the number of genes in each set represented in the 24,646-gene analysis universe, using the same size scale in both panels. GO terms were not pruned for redundancy. BP, biological process; CC, cellular component. Complete results, gene-set identifiers and represented genes are provided in Supplementary Table 4.

Figure 4. Exploratory cross-tissue assessment of ASAP1 cg10068788. Points show adjusted CUD-minus-control effects with 95% confidence intervals; panels use different outcome scales. (a) Smoking-adjusted blood EWAS (679 controls and 562 CUD): BACON-calibrated M-value coefficient and normal-statistic *P* value. The approximate Wald interval (calibrated effect ± 1.96 × calibrated standard error) treats estimated BACON parameters as fixed. Blood BH correction covers all 627,130 tested CpGs. (b) Postmortem BA9 (17 controls and 13 CUD): beta values transformed to normal scores. The primary linear model adjusts for age, sex, current-smoking probability, postmortem interval and tissue pH; the sensitivity model adds recorded race/ethnicity (White/non-White), which does not measure genetic ancestry. Brain intervals and nominal *P* values use t statistics with 23 and 22 residual degrees of freedom, respectively. BH correction covers all 388 tested CpGs separately within each brain model. Neither brain model identified a CpG at FDR < 0.05. The ASAP1 estimates are exploratory and do not establish cross-tissue replication; their magnitudes cannot be directly compared between tissues.

Figures

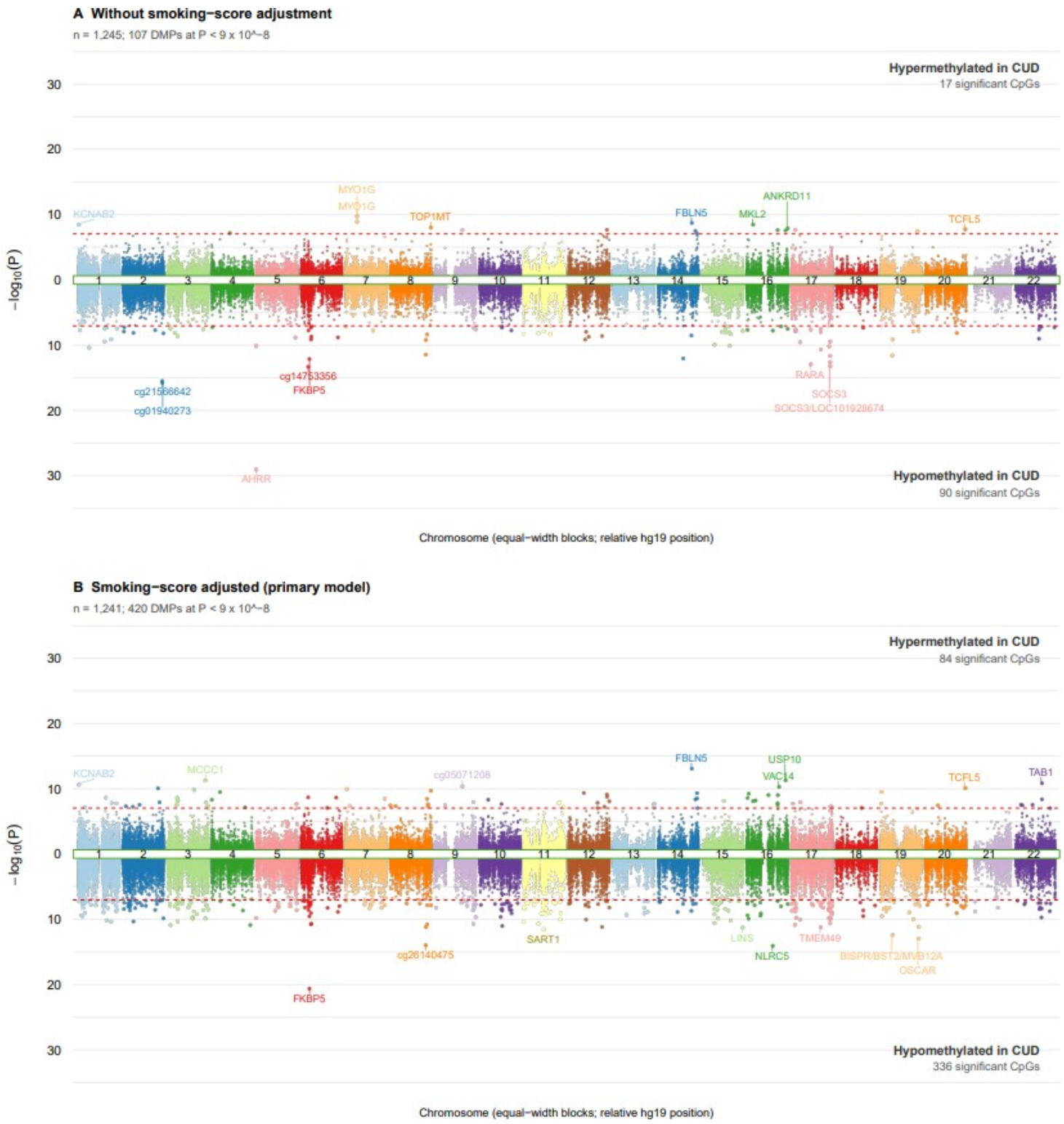


Figure 1

Directional blood EWAS Miami plots with and without smoking-score adjustment. (A) Model without the smoking score (n = 1,245; 107 DMPs: 17 hypermethylated and 90 hypomethylated). (B)

Primary smoking-adjusted model (n = 1,241; 420 DMPs: 84 hypermethylated and 336 hypomethylated). Each
panel displays all 627,130 autosomal CpGs, with positive BACON-calibrated CUD-minus-control coefficients
above the chromosome axis and negative coefficients below it. Vertical distance from the center represents
$-\log_{10}(P)$ for two-sided BACON-calibrated P values, using the same scale in both panels. Raw and calibrated
coefficients agree in direction at all significant loci. Red dashed lines mark $P < 9 \times 10^{-8}$. Colors identify chromosomes 1–22. Chromosomes occupy equal-width blocks, with hg19 positions rescaled within each
chromosome; horizontal distances are not comparable in base pairs across chromosomes. The eight significant
CpGs with the smallest P values per direction in each model are labeled using gene annotations or, when unavailable, probe identifiers. Multiple gene annotations are separated by slashes. Both models retain the other
covariates and chip random intercept. Full statistics are provided in Supplementary Table 2.

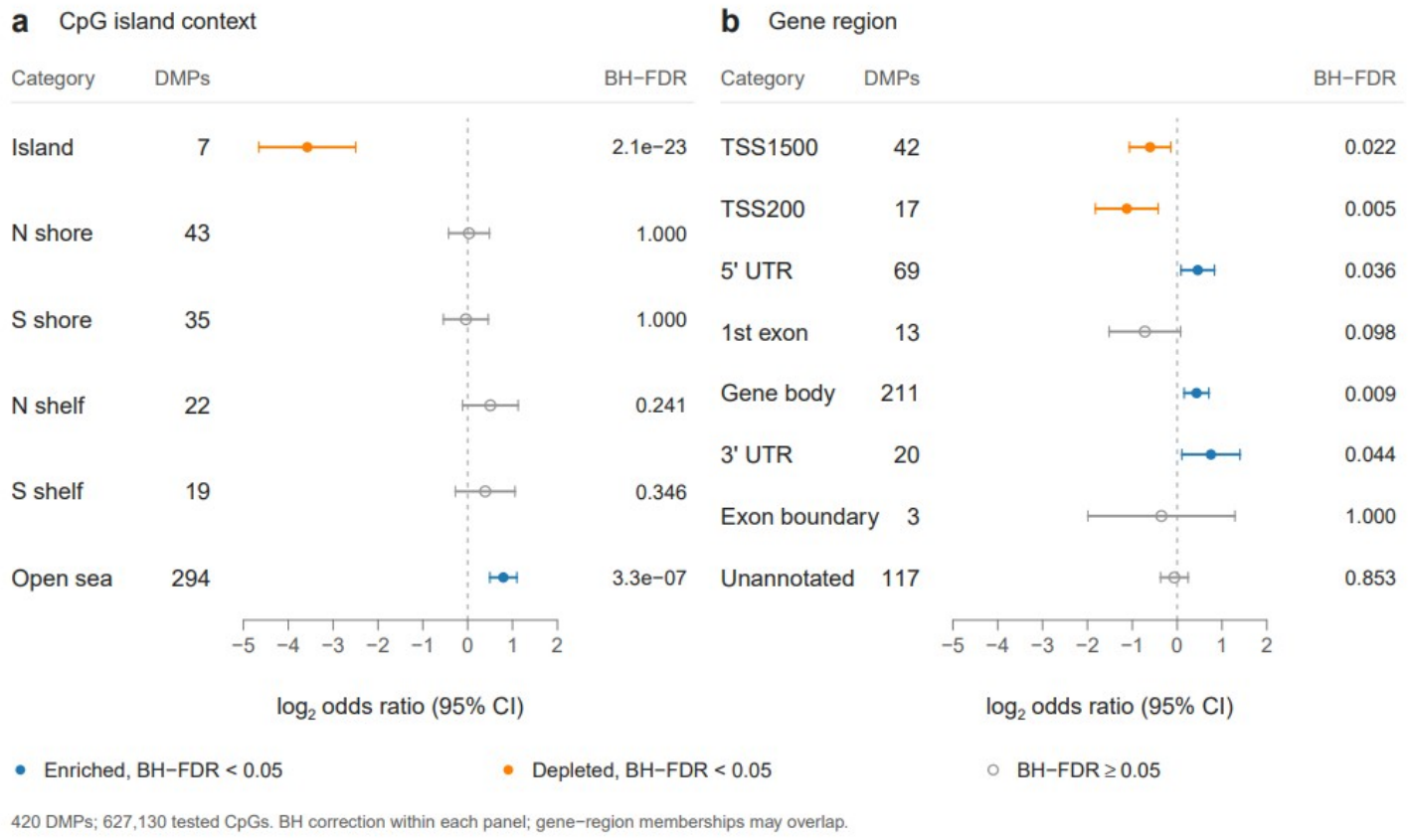


Figure 2

Genomic context of smoking-adjusted CUD-associated DMPs.
Enrichment and depletion among the 420 DMPs are shown for (A) CpG-island context and (B) gene region.
Points represent $\log_2$ odds ratios and horizontal lines show pointwise Woolf 95% confidence intervals, comparing
DMPs with the remaining tested CpGs in the 627,130-probe universe. The dashed vertical line indicates OR = 1.
Colors identify enrichment or depletion at Benjamini–Hochberg FDR < 0.05, calculated separately within each
panel; gray indicates FDR ≥ 0.05. Labels report DMP counts in each category. Distinct gene-region annotations
for the same CpG are retained, so gene-region counts may overlap. DMPs were selected at BACON-calibrated P
$< 9 \times 10^{-8}$. Complete results for both EWAS models are provided in Supplementary Table 3.

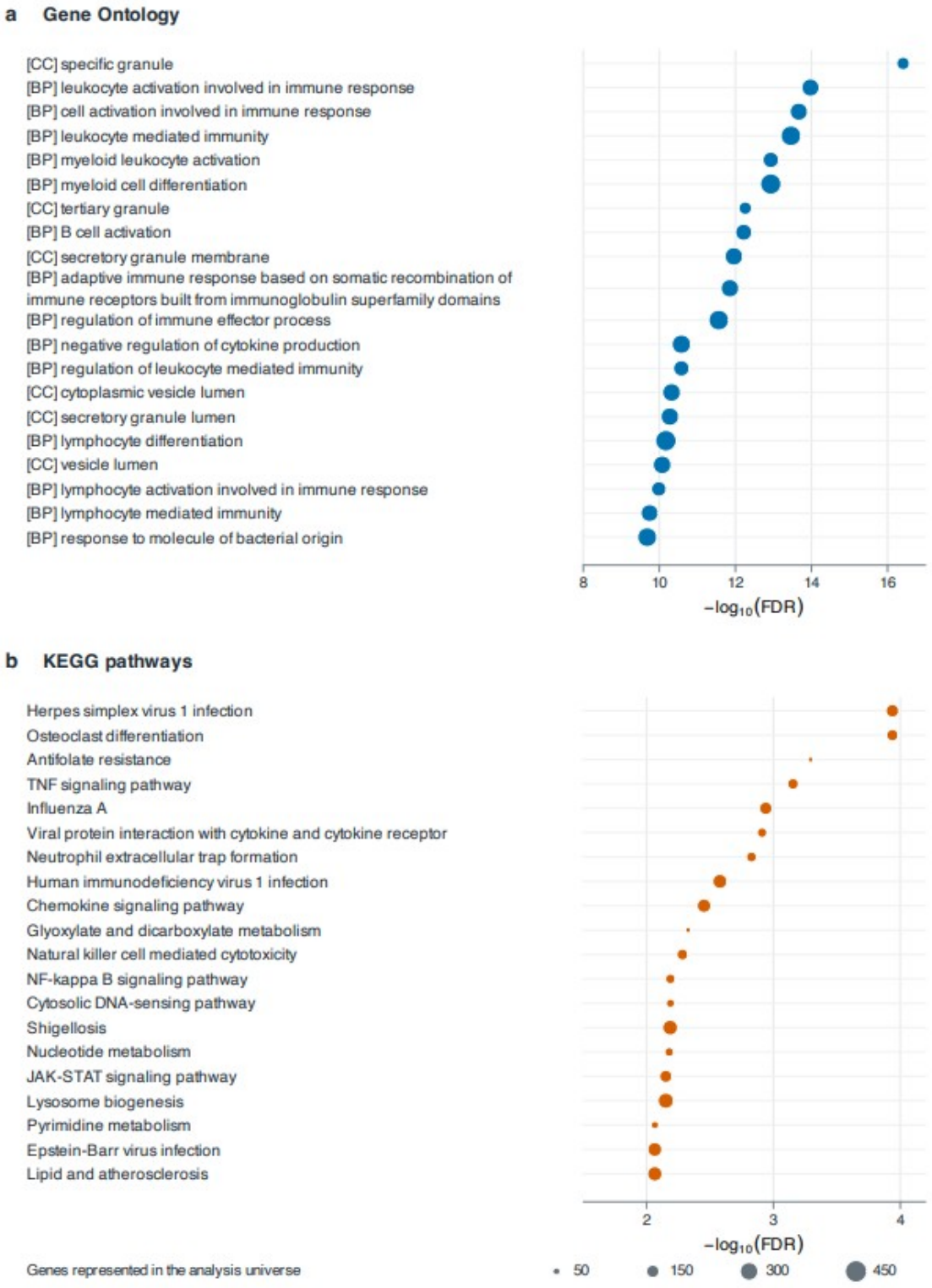


Figure 3

GO and KEGG gene-set enrichment in the smoking-adjusted EWAS.
The 20 Gene Ontology (GO) terms (a) and 20 KEGG pathways (b) with the lowest Benjamini-Hochberg false
discovery rate (FDR) values are shown. Analyses used methylglm with BACON-corrected P values for all 627,130 tested CpGs, without a significance filter or separation by methylation direction, and accounted

for the
number of analyzed CpGs mapped to each gene. FDR correction was applied across all 4,984 tested GO terms
and separately across all 322 tested KEGG pathways; all displayed sets had FDR < 0.05. Horizontal positions
indicate -log10(FDR), with different horizontal-axis ranges in the two panels. Blue and orange denote GO and
KEGG, respectively. Circle area is proportional to the number of genes in each set represented in the 24,646-
gene analysis universe, using the same size scale in both panels. GO terms were not pruned for redundancy. BP,
biological process; CC, *cellular* component. Complete results, gene-set identifiers and represented genes are
provided in Supplementary Table 4.

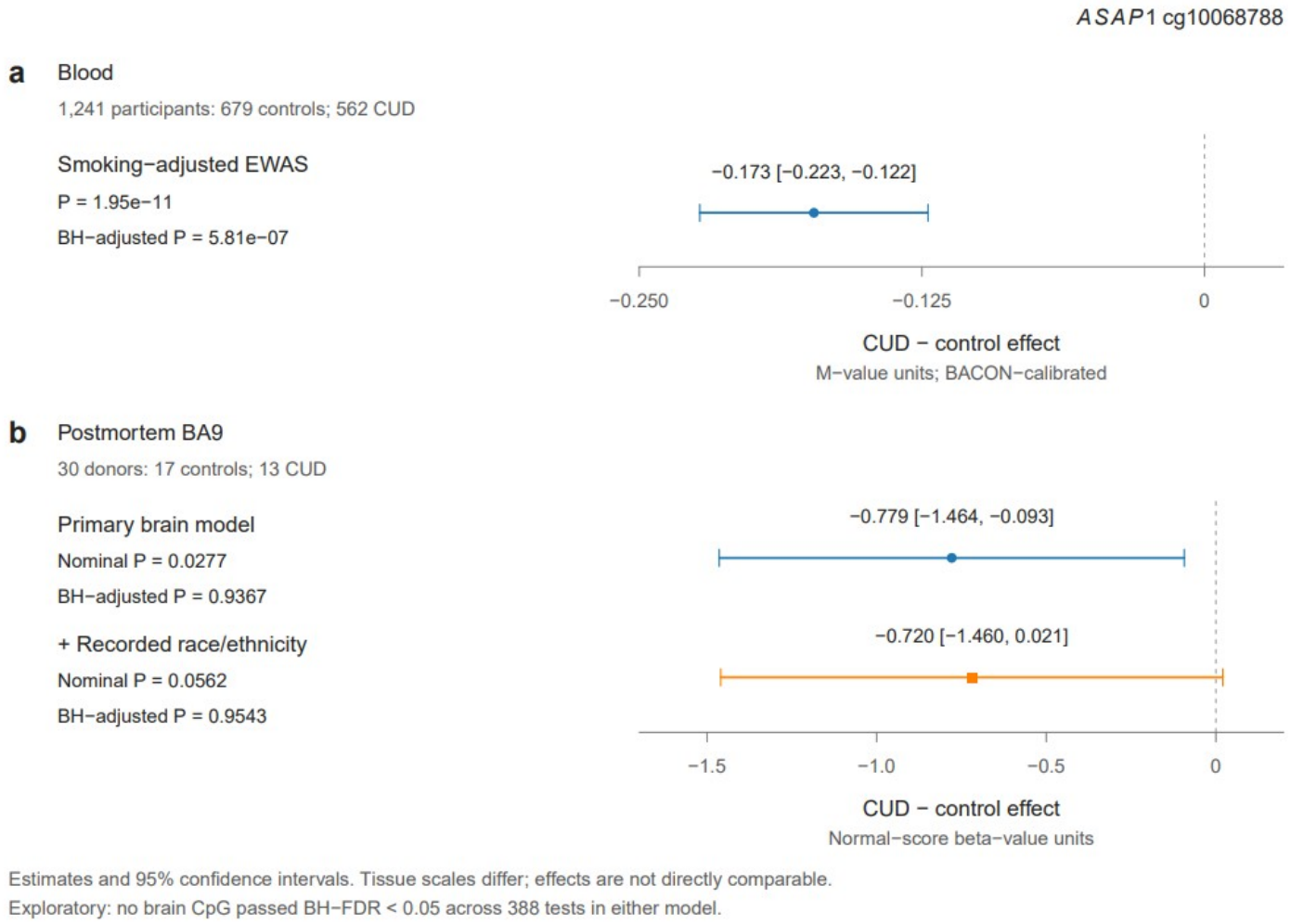


Figure 4

Exploratory cross-tissue assessment of ASAP1 cg10068788. Points show adjusted CUD-minus-control effects with 95% confidence intervals; panels use different outcome scales. (a) Smoking-

adjusted
blood EWAS (679 controls and 562 CUD): BACON-calibrated M-value coefficient and normal-statistic P value.
The approximate Wald interval (calibrated effect ± 1.96 × calibrated standard error) treats estimated BACON
parameters as fixed. Blood BH correction covers all 627,130 tested CpGs. (b) Postmortem BA9 (17 controls and
13 CUD): beta values transformed to normal scores. The primary linear model adjusts for age, sex, current-
smoking probability, postmortem interval and tissue pH; the sensitivity model adds recorded *race/ethnicity*
(White/non-White), which does not measure genetic ancestry. Brain intervals and nominal P values use t statistics
with 23 and 22 residual degrees of freedom, respectively. BH correction covers all 388 tested CpGs separately
within each brain model. Neither brain model identified a CpG at FDR < 0.05. The ASAP1 estimates are exploratory and do not establish cross-tissue replication; their magnitudes cannot be directly compared between
tissues.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- Table2.docx
- SupplementaryInformation.pdf
- SupplementaryTable1DMPs.xlsx
- SupplementaryTable2README.pdf
- SupplementaryTable3Genomiccontext.xlsx
- SupplementaryTable4GOKEGGReactome.xlsx
- SupplementaryTable5Exposures.xlsx
- SupplementaryTable6Alcoholsubgroup.xlsx
- SupplementaryTable7Brain.xlsx
- SupplementaryTable8VIF.xlsx
- SupplementaryTable9Referencecomparisons.xlsx