Trans-ancestry epigenome-wide association meta-analysis of DNA methylation with lifetime cannabis use

Fang Fang
RTI International
Laura J Bierut
Washington University School of Medicine in St. Louis
et al.

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4

Part of the Medicine and Health Sciences Commons

Please let us know how this document benefits you.

Recommended Citation
Fang, Fang; Bierut, Laura J; and et al., "Trans-ancestry epigenome-wide association meta-analysis of DNA methylation with lifetime cannabis use." Molecular Psychiatry. 29, 1. 124 - 133. (2024).
https://digitalcommons.wustl.edu/oa_4/3822

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Cannabis is widely used worldwide, yet its links to health outcomes are not fully understood. DNA methylation can serve as a mediator to link environmental exposures to health outcomes. We conducted an epigenome-wide association study (EWAS) of peripheral blood-based DNA methylation and lifetime cannabis use (ever vs. never) in a meta-analysis including 9436 participants (7795 European and 1641 African ancestry) from seven cohorts. Accounting for effects of cigarette smoking, our trans-ancestry EWAS meta-analysis revealed four CpG sites significantly associated with lifetime cannabis use at a false discovery rate of 0.05 ($p < 5.85 \times 10^{-5}$): cg22572071 near gene ADGRF1, cg15280358 in ADAM12, cg00813162 in ACTN1, and cg01101459 near LINC01132. Additionally, our EWAS analysis in participants who never smoked cigarettes identified another epigenome-wide significant CpG site, cg14237301 annotated to APOB. We used a leave-one-out approach to evaluate methylation scores constructed as a weighted sum of the significant CpGs. The best model can explain 3.79% of the variance in lifetime cannabis use. These findings unravel the DNA methylation changes associated with lifetime cannabis use that are independent of cigarette smoking and may serve as a starting point for further research on the mechanisms through which cannabis exposure impacts health outcomes.

Molecular Psychiatry (2024) 29:124–133; https://doi.org/10.1038/s41380-023-02310-w
In this study, we conducted the largest trans-ancestry EWAS (epidemiologic association study) for lifetime cannabis use (ever vs. never) in 9436 participants from seven cohorts. The initial model, which adjusted for sex, age at blood collection, blood cell proportions, and technical covariates, yielded 608 significant (False Discovery Rate (FDR) < 0.05) CpGs, among which 82% overlapped with prior EWAS findings for cigarette smoking, whose confounding effects have been found for many other phenotypes like educational attainment [25, 26], aggressive behavior [27] and coffee consumption [28]. We explored this finding further by first adjusting the analyses for cigarette smoking status and next conducting the EWAS in participants who never smoked cigarettes. These two analyses identified a total of five cigarette smoking-independent CpGs significantly associated with lifetime cannabis use. We evaluated these findings by constructing a methylation score, summarizing regional DNA methylation changes using differential methylation region (DMR) analysis, and integrating the DNA methylation findings with gene expression and genetic data.

**METHODS**

**Study cohorts**

This study included data from seven participating cohorts: the Sister Study [29], Gulf Long-Term Follow-Up Study (GuLF) [30], Netherlands Twin Register (NTR) [31], Veteran Aging Cohort Study (VACS) [32], Finnish Twin Cohort (FinnTwin) [33], Avon Longitudinal Study of Parents and Children (ALSPAC) [34], and UK Adult Twin Registry (TwinsUK) [35]. The final sample size consisted of 9436 participants, including 4190 individuals who reported ever using cannabis and 5246 who reported never using cannabis (Table 1). Detailed information for each cohort can be found in the Supplementary Methods. Informed consent was obtained from each participant, and each study was approved by their Institutional Review Boards.

**Cannabis assessment**

Our analyses focused on lifetime cannabis use based on self- or parent-report. Participants were classified as ever users if they reported using cannabis at least once prior to the blood sample collection used to generate DNA methylation data, and as never users if they reported never using cannabis prior to the blood draw. This definition of the phenotype aligns with the “Substances—Lifetime Use” variable in the PhenX Toolkit [36], making the results comparable and combinable.

**DNA methylation measurements**

DNA methylation was measured in peripheral blood using either the Illumina Infinium HumanMethylation450 BeadChip (450K array, 76%) or the Illumina Infinium Methylation EPIC BeadChip (EPIC array, 24%), as shown in Table 1. DNA methylation levels were calculated as β-values, which represent the percentage of DNA that is methylated at the interrogated CpG site and ranges from 0 to 1. Quality control and normalization procedures were implemented consistently across all cohorts, with considerations specific to each cohort (detailed in the Supplementary Methods). Cohort-specific methods are shown in Supplementary Table 1.

**EWAS for lifetime cannabis use**

In each cohort, the association between DNA methylation (DNAm) levels and lifetime cannabis use was tested under a linear model or a GEE model if participants were related. We stratified the EWAS analyses by genetic ancestry groups (EA and AA) and DNA methylation array types (450K and EPIC). For each CpG site, the DNA methylation β-value was considered as the outcome of interest, and two separate models were applied. In the basic model (Model 1), we included sex (except in cohorts with only one sex), age at blood collection, blood cell type estimation, and technical covariates. In Model 2, we additionally adjusted for cigarette smoking status defined as current, former, or never. Comparing with prior smoking EWAS [37], the results from Model 2 were considered as cigarette smoking-independent DNA methylation biomarkers for lifetime cannabis use. Additionally, we implemented EWAS within the subset of participants who never smoked cigarettes using Model 1 to minimize the possible confounding effect of cigarette smoking. More detailed information for models and covariates used in each cohort is provided in the Supplementary Methods.

### Table 1. Characteristics of the cohort participants.

<table>
<thead>
<tr>
<th>Study cohort</th>
<th>Total N</th>
<th>Ancestry</th>
<th>Countries</th>
<th>Ancestry</th>
<th>Total N</th>
<th>Cannabinoid ever use in never smokers %</th>
<th>Never smokers</th>
<th>DNA methylation array</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sister study</td>
<td>2563</td>
<td>US</td>
<td>US</td>
<td>100% EA</td>
<td>100%</td>
<td>450K</td>
<td>49%</td>
<td>49%</td>
</tr>
<tr>
<td>GuLF</td>
<td>1195</td>
<td>US</td>
<td>US</td>
<td>56% EA</td>
<td>56%</td>
<td>EPIC</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>NTR</td>
<td>2142</td>
<td>Netherland</td>
<td>US</td>
<td>70%</td>
<td>70%</td>
<td>450K</td>
<td>38%</td>
<td>38%</td>
</tr>
<tr>
<td>VACS</td>
<td>1116</td>
<td>US</td>
<td>US</td>
<td>44% AA</td>
<td>44%</td>
<td>EPIC</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>Finntwin</td>
<td>1368</td>
<td>Finland</td>
<td>US</td>
<td>50%</td>
<td>50%</td>
<td>450K</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>ALSPAC</td>
<td>922</td>
<td>UK</td>
<td>UK</td>
<td>50% EUR</td>
<td>50%</td>
<td>EPIC</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>TwinsUK</td>
<td>110</td>
<td>UK</td>
<td>UK</td>
<td>100%</td>
<td>100%</td>
<td>450K</td>
<td>40%</td>
<td>40%</td>
</tr>
</tbody>
</table>

aMean and standard deviation presented.

bThe majority of NTR have Dutch ancestry and individuals who are outliers on genetic PCs were removed.
Meta-analysis
We summarized cohort- and ancestry-specific EWAS results using inverse variance fixed effects meta-analysis implemented in the METAL software [38], with Model 1 (all participants), Model 2 (all participants), and Model 1 in participants who never smoked cigarettes, respectively. We reported overlapping CpGs between 450K and EPIC arrays, which included 452,453 CpG probes. Epigenome-wide significance was defined as FDR less than 5%. Manhattan and QQ plots were generated using the CMplot function within the R package mVPl [39]. Heterogeneity among the studies was assessed using the Cochran’s Q-test implemented in METAL.

Methylation scores
As the single largest cohort included in this study, the Sister Study was reserved as the testing dataset to evaluate DNAm scores. For each individual in the Sister Study, a methylation score was calculated as a weighted sum of CpGs significantly associated with lifetime cannabis use in the EWAS meta-analysis conducted without the Sister Study [40]. At a given CpG site i, the methylation beta value \( (\text{meth}_i) \) was multiplied by the effect size of the CpG in the meta-analysis \( (\text{eff}_i) \). Then a methylation score was obtained by summing over a selected CpG set:

\[
\text{methylation score} = \text{meth}_1 \cdot \text{eff}_1 + \text{meth}_2 \cdot \text{eff}_2 + \ldots + \text{meth}_n \cdot \text{eff}_n
\]

We applied different \( p \)-value thresholds to select the significant CpG sets for both Model 1 (\( p < 10^{-7}, 10^{-10}, 10^{-12}, 10^{-15}, 10^{-18} \)) and model 2 (\( p < 10^{-9}, 10^{-10}, 10^{-12}, 10^{-15} \)). To evaluate performance of the methylation scores, we quantified the percentage of variance \( (R^2) \) explained as proposed by Lee et al. [41] for binary responses.

Integrating EWAS results with gene expression
To investigate the potential relationship between DNAm and gene expression levels, we used the correlations between DNAm and expression data available in six tissues (brain, colon, kidney, liver, stomach, testis) in the EWAS Atlas [42]. Specifically, we analyzed the DNAm levels of the significant CpGs to determine if they were associated with the expression levels of nearby genes.

Correlation of DNAm between blood and brain tissues
An online database (https://epigenetics.essex.ac.uk/bloodbrain) [43] was used to examine the correlations of DNAm between whole blood and four brain regions (prefrontal cortex, entorhinal cortex, superior temporal gyrus, and cerebellum), respectively. For each CpG site, a boxplot was generated to display the distribution of DNAm levels across all five tissues, and the Pearson correlation was calculated between the DNAm level in whole blood and each of the four brain tissues.

Enrichment analysis against previous EWAS results
To determine the potential overlap of our top CpGs with previously reported EWAS results, we used the EWAS Atlas toolkit [42] to conduct enrichment analyses. To meet the minimum input requirement, we selected the top 20 CpGs from the EWAS meta-analysis using Model 2 and the top 20 CpGs in participants who never smoked cigarettes, and ran enrichment analyses on each group separately.

Methylation quantitative trait loci (meQTL)
To explore the potential genetic basis for DNAm changes identified in our study, we looked up genetic variants that are associated with DNAm levels of CpG sites within the Genetics of DNA Methylation Consortium (GoDMC) database [44], which includes both local (cis) and distal (trans) meQTLs. We examined the overlap between the meQTL variants for cannabis use–associated CpG sites and the variants identified in a GWAS for lifetime cannabis use [45].

DMR analyses
We used an R software tool, ipDMR [46], to identify DMRs in which a cluster of correlated CpGs showed evidence for association with lifetime cannabis use. ipDMR calculates an overall \( p \)-value for small intervals bordered by two adjacent CpGs based on the association \( p \)-values from an EWAS analysis. It then combines all nearby significant intervals (using a seed threshold) and calculates an FDR-adjusted \( p \)-value for the combined region. We applied ipDMR to summary statistics from our EWAS meta-analysis with the following parameters: seed \( p \)-value < 0.05, maximum distance to combine adjacent intervals 1000 bp and bin size 50 bp. To assess the biological significance of the identified DMRs, we then conducted gene set enrichment analysis using the tool provided by Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) [47] to test for enrichment of the genes that overlapped with the DMRs in predefined pathways.

Sensitivity analyses
To evaluate the potential confounding effects of alcohol use and BMI on the association between DNAm and lifetime cannabis use, we conducted sensitivity analyses. In these analyses, we compared the effect sizes and \( p \)-values of the significant CpGs identified in the main analysis with those obtained in the sensitivity analyses to assess the robustness of our results. Detailed definitions of alcohol use in each cohort can be found in the Supplementary Methods section.

Comparison of results across ancestry groups
Considering the differential DNAm patterns across different ancestry groups, we conducted a separate EWAS meta-analysis in EA \( (N = 7795) \) and AA \( (N = 1641) \) groups with Model 2. We compared the effect sizes and \( p \)-values of the top significant CpGs from the main EWAS results with the results in each ancestry group.

RESULTS
Sample characteristics
The demographic characteristics of 9436 study participants are summarized in Table 1. The sample consisted of 57% females, and the mean age at DNAm sampling ranged from 17.1 years in the ALSPAC cohort to 58.8 years in the TwinsUK cohort. A total of 44% of the participants reported having used cannabis at some point in their lives. Within the subset of participants who reported never smoking \( (N = 4146) \), 27% indicated having used cannabis at least once. DNAm sites were assessed using either the Illumina HumanMethylation450K BeadChip (76%) or the Illumina HumanMethylation EPIC (850K) BeadChip (24%). The sample included individuals of both European ancestry (EA, 83%) and African ancestry (AA, 17%).

EWAS meta-analysis results for lifetime cannabis use
We conducted EWAS meta-analysis on peripheral blood DNAm data from 9436 participants for lifetime cannabis use. In each cohort, we tested the association between lifetime cannabis use and DNAm at each CpG site using either a linear model (for unrelated participants) or a generalized estimation equation (GEE) model (for related participants) with the family ID as the clustering variable. To investigate the confounding effect of cigarette smoking, we compared two models. Model 1 included sex (in cohorts with more than one sex), age at blood collection, measured or estimated white blood cell proportions, and technical covariates. In addition to these covariates, Model 2 included cigarette smoking status defined as current, former, or never. The EWAS meta-analysis with Model 1 identified 608 CpGs significantly associated with lifetime cannabis use at an FDR threshold of 5%. Of these, 500 CpGs had been previously reported as being significantly associated with cigarette smoking [37, 48] (Supplementary Fig. 1). After adjusting for cigarette smoking, Model 2 EWAS meta-analysis identified four CpGs (Fig. 1; Table 2) significantly (FDR < 0.05) associated with lifetime cannabis use. None of these four cannabis use–associated CpGs had been reported as being significant in previous EWAS for cigarette smoking after accounting for multiple testing \( (p > 0.05/4) \) [48]. The quantile-quantile (QQ) plots from both models suggested minimal inflation \( (\lambda = 1.1) \). Although many CpGs did not reach epigenome-wide significance with Model 2, their effect sizes showed consistent directions of associations as in Model 1 (Supplementary Fig. 2), and their effect sizes were highly correlated \( (r = 0.85) \). The four cigarette smoking-independent CpGs identified with Model 2—cg01101459, cg22572071, cg15280538, and cg00813162—
were annotated to LINC01132, ADGRF1, ADAM12, and ACTN1, respectively, as the closest genes (Supplementary Fig. 3). The full list of top CpGs (p<0.001) from the Model 2 EWAS meta-analysis are summarized in Supplementary Table 2, and a forest plot showing the effect sizes of the four significant CpGs is in Supplementary Fig. 4.

Considering potential differences by ancestry, we performed EWAS meta-analyses stratified into EA (N = 7795) and AA (N = 1641) groups. The ancestry-specific results for the top CpGs reported from the Model 2 EWAS meta-analysis are shown in Supplementary Table 2. The effect sizes for the top CpGs were highly correlated between the EWAS results in the EA and AA groups (r = 0.77, Supplementary Fig. 5).

We additionally compared EWAS meta-analyses results within male-only cohorts (VACS and GuLF, N = 2311) and female-only cohorts (Sister Study and TwinsUK, N = 2693). More than 95% of the top CpGs demonstrated consistent effect directions between males and females (Supplementary Table 2). The correlation of the effect sizes of the top CpGs was 0.69 between males and females (Supplementary Fig. 6).

EWAS meta-analysis results for lifetime cannabis use in participants who never smoked cigarettes

To further explore the confounding effect of cigarette smoking on DNAm, we conducted an EWAS meta-analysis on the subset of participants who reported never having smoked cigarettes (N = 4146). The EWAS meta-analysis, using Model 1, identified one CpG significantly associated with lifetime cannabis use at FDR < 0.05 in this subset of participants (Fig. 2; Table 2). This CpG site, cg14237301, is annotated to the gene APOB4R, which has been reported to be significantly associated with lifetime cannabis use in a genome-wide association study (GWAS) [45] (Fig. 3). The full list of top CpGs (p<0.001) from the EWAS meta-analysis in participants who never smoked cigarettes is summarized in Supplementary Table 3. Their effect sizes were highly correlated with the EWAS meta-analysis results in all participants under either model (r>0.7, Supplementary Fig. 2).

Replication of previous reported DNAm changes associated with cannabis use

We compared our results with previously reported EWAS for cannabis use (Supplementary Table 4). All seven nominally significant CpGs identified in Markunas et al. [49] remained significant in our EWAS results using Model 2 after applying a Bonferroni correction (0.05/7 = 0.0071). Osborne et al. reported CpGs associated with cannabis use in individuals who exclusively used cannabis or used it in conjunction with tobacco [23]. We compared them with our findings in never smokers and EWAS model 1, separately. Out of a total of 30 CpGs examined, we successfully replicated 10 CpGs in our study with a Bonferroni correction. Additionally, a recent study examined the associations between DNAm and recent and cumulative cannabis use in middle-aged adults [50]. Our results replicated four out of 40 CpGs reported with a Bonferroni correction.

Methylation scores

To assess the ability of DNAm levels to predict lifetime cannabis use, we calculated a methylation score based on summary statistics from the EWAS meta-analyses without the Sister Study, the largest cohort in this project. We then compared the performance of multiple models with different p-value cutoffs by the variance in lifetime cannabis use explained by the methylation scores (Supplementary Table 5). The best-performing methylation score, based on 50 CpGs with p<10^-9 in the EWAS meta-analysis using Model 1 (Supplementary Table 6), explained 3.79% of the variance of lifetime cannabis use in the Sister Study (p = 1.71 x 10^-17). To further examine the confounding effect of cigarette smoking, we also evaluated the variance explained by the methylation score in participants who never smoked cigarettes. In this subset of participants, the methylation score based on the same 50 CpGs explained 0.91% of the variance of lifetime cannabis use (p = 1.9 x 10^-3). In contrast, the methylation score constructed based on EWAS meta-analysis results with Model 2 can explain 0.58% of the variance of lifetime cannabis use in all participants of the Sister Study, achieved with p-value cutoff at 10^-3.

Sensitivity analysis

We conducted additional analyses to adjust for other factors that influence DNAm as shown in previous studies: alcohol use [51] and BMI [52]. We examined if the EWAS results with Model 2 differed by expanding Model 2 to also include alcohol use and body mass index (BMI) as covariates. The effect sizes of the top cannabis use–associated CpGs (p<0.001) from Model 2 (Supplementary Table 2) were strongly correlated (r = 0.99) with the results of the sensitivity analyses (Supplementary Fig. 7).

Follow-up of CpGs significantly associated with lifetime cannabis use

We examined the five CpGs that were epigenome-wide significantly associated with lifetime cannabis use, four of which were identified using Model 2 in all participants (adjusted for cigarette smoking) and one CpG identified in participants who never smoked cigarettes. Using the EWAS Atlas [42] (https://
Differentially methylated regions

We used an R software tool, ipDMR [46], to identify DMRs based on EWAS meta-analysis results (Supplementary Table 11). Model 1 (no adjustment for cigarette smoking) yielded 514 significant DMRs (FDR < 0.05) with a minimum of two probes, while Model 2 (accounting for cigarette smoking) had 10 significant DMRs, showing that one of the four epigenome-wide significant CpG sites (cg00813162-\textit{ACTN1}) reside in a region where proximate CpGs were associated with lifetime cannabis use. Additionally, there were nine DMRs that included no single epigenome-wide significant CpG, and instead multiple correlated CpGs showed some evidence of association. The DMR analysis in participants who never smoked cigarettes showed significant overlap with CpGs previously associated with smoking ($p = 5.58 \times 10^{-8}$), smoking cessation, lung function, and lung carcinoma.

To explore whether the DNAm changes were genetic-driven, we looked up the methylation quantitative trait loci (meQTLs) for the five epigenome-wide significant CpGs in the Genetics of DNA Methylation Consortium (GoDMC) database [44] (http://mgmtdb.godmc.org/about), and overlapped them with the GWAS results for lifetime cannabis use [45] (Supplementary Table 10). None of the meQTLs were significantly associated with lifetime cannabis use ($p = 5 \times 10^{-8}$), indicating that our significantly associated CpGs were not directly driven by the genetic variants. The meQTLs for cg00813162-\textit{ACTN1} and cg14237301-\textit{APOBR} showed some degree of significance in the GWAS ($p<0.01$, Supplementary Table 10).

**Differentially methylated regions**

We used an R software tool, ipDMR [46], to identify DMRs based on EWAS meta-analysis results (Supplementary Table 11). Model 1 (no adjustment for cigarette smoking) yielded 514 significant DMRs (FDR < 0.05) with a minimum of two probes, while Model 2 (accounting for cigarette smoking) had 10 significant DMRs, showing that one of the four epigenome-wide significant CpG sites (cg00813162-\textit{ACTN1}) reside in a region where proximate CpGs were associated with lifetime cannabis use. Additionally, there were nine DMRs that included no single epigenome-wide significant CpG, and instead multiple correlated CpGs showed some evidence of association. The DMR analysis in participants who never smoked cigarettes showed significant overlap with CpGs previously associated with smoking ($p = 5.58 \times 10^{-8}$), smoking cessation, lung function, and lung carcinoma.

To explore whether the DNAm changes were genetic-driven, we looked up the methylation quantitative trait loci (meQTLs) for the five epigenome-wide significant CpGs in the Genetics of DNA Methylation Consortium (GoDMC) database [44] (http://mgmtdb.godmc.org/about), and overlapped them with the GWAS results for lifetime cannabis use [45] (Supplementary Table 10). None of the meQTLs were significantly associated with lifetime cannabis use ($p = 5 \times 10^{-8}$), indicating that our significantly associated CpGs were not directly driven by the genetic variants. The meQTLs for cg00813162-\textit{ACTN1} and cg14237301-\textit{APOBR} showed some degree of significance in the GWAS ($p<0.01$, Supplementary Table 10).

**Differentially methylated regions**

We used an R software tool, ipDMR [46], to identify DMRs based on EWAS meta-analysis results (Supplementary Table 11). Model 1 (no adjustment for cigarette smoking) yielded 514 significant DMRs (FDR < 0.05) with a minimum of two probes, while Model 2 (accounting for cigarette smoking) had 10 significant DMRs, showing that one of the four epigenome-wide significant CpG sites (cg00813162-\textit{ACTN1}) reside in a region where proximate CpGs were associated with lifetime cannabis use. Additionally, there were nine DMRs that included no single epigenome-wide significant CpG, and instead multiple correlated CpGs showed some evidence of association. The DMR analysis in participants who never smoked cigarettes showed significant overlap with CpGs previously associated with smoking ($p = 5.58 \times 10^{-8}$), smoking cessation, lung function, and lung carcinoma.

To explore whether the DNAm changes were genetic-driven, we looked up the methylation quantitative trait loci (meQTLs) for the five epigenome-wide significant CpGs in the Genetics of DNA Methylation Consortium (GoDMC) database [44] (http://mgmtdb.godmc.org/about), and overlapped them with the GWAS results for lifetime cannabis use [45] (Supplementary Table 10). None of the meQTLs were significantly associated with lifetime cannabis use ($p = 5 \times 10^{-8}$), indicating that our significantly associated CpGs were not directly driven by the genetic variants. The meQTLs for cg00813162-\textit{ACTN1} and cg14237301-\textit{APOBR} showed some degree of significance in the GWAS ($p<0.01$, Supplementary Table 10).

**Differentially methylated regions**

We used an R software tool, ipDMR [46], to identify DMRs based on EWAS meta-analysis results (Supplementary Table 11). Model 1 (no adjustment for cigarette smoking) yielded 514 significant DMRs (FDR < 0.05) with a minimum of two probes, while Model 2 (accounting for cigarette smoking) had 10 significant DMRs, showing that one of the four epigenome-wide significant CpG sites (cg00813162-\textit{ACTN1}) reside in a region where proximate CpGs were associated with lifetime cannabis use. Additionally, there were nine DMRs that included no single epigenome-wide significant CpG, and instead multiple correlated CpGs showed some evidence of association. The DMR analysis in participants who never smoked cigarettes showed significant overlap with CpGs previously associated with smoking ($p = 5.58 \times 10^{-8}$), smoking cessation, lung function, and lung carcinoma.

To explore whether the DNAm changes were genetic-driven, we looked up the methylation quantitative trait loci (meQTLs) for the five epigenome-wide significant CpGs in the Genetics of DNA Methylation Consortium (GoDMC) database [44] (http://mgmtdb.godmc.org/about), and overlapped them with the GWAS results for lifetime cannabis use [45] (Supplementary Table 10). None of the meQTLs were significantly associated with lifetime cannabis use ($p = 5 \times 10^{-8}$), indicating that our significantly associated CpGs were not directly driven by the genetic variants. The meQTLs for cg00813162-\textit{ACTN1} and cg14237301-\textit{APOBR} showed some degree of significance in the GWAS ($p<0.01$, Supplementary Table 10).

**Differentially methylated regions**

We used an R software tool, ipDMR [46], to identify DMRs based on EWAS meta-analysis results (Supplementary Table 11). Model 1 (no adjustment for cigarette smoking) yielded 514 significant DMRs (FDR < 0.05) with a minimum of two probes, while Model 2 (accounting for cigarette smoking) had 10 significant DMRs, showing that one of the four epigenome-wide significant CpG sites (cg00813162-\textit{ACTN1}) reside in a region where proximate CpGs were associated with lifetime cannabis use. Additionally, there were nine DMRs that included no single epigenome-wide significant CpG, and instead multiple correlated CpGs showed some evidence of association. The DMR analysis in participants who never smoked cigarettes showed significant overlap with CpGs previously associated with smoking ($p = 5.58 \times 10^{-8}$), smoking cessation, lung function, and lung carcinoma.

To explore whether the DNAm changes were genetic-driven, we looked up the methylation quantitative trait loci (meQTLs) for the five epigenome-wide significant CpGs in the Genetics of DNA Methylation Consortium (GoDMC) database [44] (http://mgmtdb.godmc.org/about), and overlapped them with the GWAS results for lifetime cannabis use [45] (Supplementary Table 10). None of the meQTLs were significantly associated with lifetime cannabis use ($p = 5 \times 10^{-8}$), indicating that our significantly associated CpGs were not directly driven by the genetic variants. The meQTLs for cg00813162-\textit{ACTN1} and cg14237301-\textit{APOBR} showed some degree of significance in the GWAS ($p<0.01$, Supplementary Table 10).

**Differentially methylated regions**

We used an R software tool, ipDMR [46], to identify DMRs based on EWAS meta-analysis results (Supplementary Table 11). Model 1 (no adjustment for cigarette smoking) yielded 514 significant DMRs (FDR < 0.05) with a minimum of two probes, while Model 2 (accounting for cigarette smoking) had 10 significant DMRs, showing that one of the four epigenome-wide significant CpG sites (cg00813162-\textit{ACTN1}) reside in a region where proximate CpGs were associated with lifetime cannabis use. Additionally, there were nine DMRs that included no single epigenome-wide significant CpG, and instead multiple correlated CpGs showed some evidence of association. The DMR analysis in participants who never smoked cigarettes showed significant overlap with CpGs previously associated with smoking ($p = 5.58 \times 10^{-8}$), smoking cessation, lung function, and lung carcinoma.

To explore whether the DNAm changes were genetic-driven, we looked up the methylation quantitative trait loci (meQTLs) for the five epigenome-wide significant CpGs in the Genetics of DNA Methylation Consortium (GoDMC) database [44] (http://mgmtdb.godmc.org/about), and overlapped them with the GWAS results for lifetime cannabis use [45] (Supplementary Table 10). None of the meQTLs were significantly associated with lifetime cannabis use ($p = 5 \times 10^{-8}$), indicating that our significantly associated CpGs were not directly driven by the genetic variants. The meQTLs for cg00813162-\textit{ACTN1} and cg14237301-\textit{APOBR} showed some degree of significance in the GWAS ($p<0.01$, Supplementary Table 10).
After additionally adjusting for smoking in the EWAS model, we found four CpG sites statistically independent of cigarette smoking that were significantly associated with lifetime cannabis use. Additionally, we conducted EWAS in participants who never smoked cigarettes to further eliminate the influence of the participants’ cigarette smoking. This analysis showed high consistency with the smoking-adjusted model in all participants (Model 2) and yielded one additional CpG significantly associated with lifetime cannabis use. The genes annotated to these five cannabis use–associated CpGs are relevant to a range of health outcomes [53–63].

The five epigenome-wide significant CpGs are annotated to the nearest genes: LINC01132, ADGRF1, ADAM12, ACTN1, and APOB. Of these, cg01101459-LINC01132, cg00813162-ACTN1, and cg14237301-APOB were inversely associated with cannabis use, meaning that individuals who had used cannabis had lower DNAm at these CpG sites compared to those who had never used cannabis. LINC01132 is a long noncoding RNA gene that has been reported to function as an oncogene that relates to the malignant behaviors of cancer cells in hepatocellular carcinoma (HCC) [53, 64, 65] and ovarian cancer [64, 65]. Interestingly, cannabis use has been linked to a reduced incidence rate of HCC [66] and the potential for treating ovarian cancer [67, 68]. ACTN1 (Alpha-Actinin-1) encodes a non-muscle cytoskeletal protein that binds actin to the cell membrane [54]. Genetic variants and differential expression of the ACTN1 have been reported in various diseases, including congenital macrothrombocytopenia, Angelman syndrome, Bowen disease, postmenopausal osteoporosis, lupus erythematosus, and COVID-19 [54–57, 69, 70]. There is evidence that cannabinoid receptors can attenuate seizures and EEG abnormalities in Angelman syndrome [71, 72], effective in treating individuals with osteoporosis [73], reduce cell proliferation and induce apoptosis in melanoma cells [74], and as well downregulate the immune response in lupus erythematosus [75]. Notably, ACTN1 was also reported as a differentially methylated gene in sperms associated with cannabis use [21]. APOB (Apolipoprotein B) receptor encodes a receptor protein that binds to dietary triglyceride-rich lipoproteins. Its genetic variants have been associated with obesity [58, 76], bladder cancer [61], pneumonia [60], allergy [59], and lifetime cannabis use [45]. Relatedly, cannabis use has been associated with a reduced obesity rate [77] but an increased risk for pneumonia [78]. Overall, DNAm may serve as a mediator between cannabis use and its impact on health outcomes.

On the other side, cg22572071-ADGRF1 and cg15280358-ADAM12 were positively associated with cannabis use. ADGRF1 is a receptor gene that is critical in neurodevelopment and neuroinflammation [62, 79]. The overexpression of ADGRF1 has been reported in breast cancer [80]. ADAM12 encodes protein that involving in cell–cell interaction, muscle development, and neurogenesis. The expression of ADAM12 has been reported to be upregulated in various tumor cells and is an emerging prognostic biomarker for cancer [81–85]. Genetic variants in ADAM12 have been associated with neurological diseases such as multiple sclerosis and Alzheimer’s disease. Taking together, cannabis use is associated with regulation of genes that function in neurogenesis, neurodevelopment, brain structures and oncogenesis. Further experimental studies are needed to determine the potential effects of changes in DNAm levels on the corresponding gene expressions.

The majority of significant CpGs in EWAS for cannabis use with the basic model (Model 1) overlapped with those identified in the EWAS for cigarette smoking [48]. The number of significant CpGs decreased dramatically after adjusting for smoking status in our Model 2. These findings suggest that cigarette smoking is a strong confounder for cannabis use. The overlap with EWAS results for cigarette smoking were also found for many other phenotypes like educational attainment [25, 26], aggressive behavior [27] and coffee consumption [28]. However, we found that DNAm scores calculated as weighted sums of the beta values of significant CpGs from Model 1 explained a greater amount of variance in lifetime cannabis use than the DNAm scores based on Model 2 (3.79% vs. 0.58%). Even in participants who never smoked cigarettes, the DNAm scores based on Model 1 could explain 0.91% of the variance in lifetime cannabis use. Additionally, the EWAS results in participants who never smoked cigarettes showed enrichment of CpGs associated with cigarette smoking. These results suggest that cannabis use and cigarette smoking may independently influence the DNAm levels of these CpGs, implying the potential influence of common combustible chemicals that are shared between cannabis use and cigarette smoking.

In previous studies, DNAm scores have been used to predict various outcomes with the variance explained ranging from 0.6% in low-density lipoprotein, 2.5% in educational attainment, 12.5% in alcohol use, to 60.9% in cigarette smoking [86]. Such predictors may provide more accurate measurements than self-reported phenotypes and could have clinical applications by relating to health outcomes. We assessed whether blood-based DNAm could predict cannabis use and found that the variance in lifetime cannabis use explained by DNAm scores was (3.79%) moderate and not as accurate as DNAm-based score for cigarette smoking. Future applications that integrate both polygenic risk scores and DNAm scores may improve prediction power [40, 86]. The accumulation of even larger scale GWAS and EWAS for lifetime cannabis use will be necessary to establish reliable biomarkers for clinical purposes.
As a complement to the main EWAS, the DMR analysis that combined nearby correlated CpGs identified additional regions associated with cannabis use. After adjusting for smoking, the EWAS for lifetime cannabis use identified fewer epigenome-wide significant hits, and the DMR analysis provides an improved power to detect correlated CpGs with small effects influenced by cannabis exposure. The gene set enrichment analysis revealed biological processes related to growth, response to stress, and assembly of actin filament bundle. The stress response pathways have been linked to the use of Δ9-tetrahydrocannabinol (THC), which has led to DNA damage and induced oxidative stress in both blood and brain cells [87, 88]. The actin filament bundle gene set was also reported in cellular remodeling events induced by cannabinoid that affect the brain architecture and wiring [89].

The large sample set included in this study provides a good representation for a wide range of populations across countries, sexes, ancestry groups, and age ranges, empowering more generalizable findings in identifying a common and robust DNAm signature for lifetime cannabis use. A recent genetic study [90] has shown that increased sample sizes of diverse ancestries improved detection power and generated more generalizable polygenic risk scores. However, one limitation of this approach is that associations with ancestry effects may have been attenuated when all data were combined, as heterogeneity across different cohorts would have reduced power to detect such specific associations in underrepresented populations. Such heterogeneity is also reflected by inconsistent directions of effects in meta-analysis. Future studies on more data from diverse populations may reveal ancestry-, sex- and age-specific DNAm associations. There may also be relevant confounders that we have not been able to adjust for, such as exposures and experiences that lead to cannabis use.

In this study, we analyzed DNAm profiles from blood samples. While DNAm profiles differ across tissues and cell types, our results based on blood samples may not be generalizable to other tissues that may be more biologically relevant to the addiction and other behavioral effects of cannabis use, such as the brain. It should also be noted that the current EWAS results may suffer from potential confounding effects caused by blood cell subtype heterogeneity that were not fully captured by reference-based deconvolution that we applied [91, 92].
The EWAS meta-analysis conducted in this study only reflects the association between lifetime cannabis use and DNAm levels, without any causal inference. The differentially methylated CpGs identified indicate persistent rather than transient associations with cannabis use. We expect that including ever users who had used cannabis long before blood drawn may bias results towards the null, attenuating effect sizes that may be driven by heavy users within the case group. To delve deeper into the persistent and transient effects of cannabis on DNAm alterations, more precise measurements of cannabis use regarding recency and frequency are needed. Taken together, we propose future studies to integrate cannabis use patterns, GWAS results, DNAm data from multiple tissues, and gene expression data, to infer causal links between cannabis exposure and DNAm levels.

In conclusion, our EWAS found that a large proportion of DNAm changes that are significantly associated with cannabis use overlap with those observed for cigarette smoking, even in participants who never smoked cigarettes, suggesting that cannabis use and cigarette smoking may independently influence the DNAm levels of shared CpGs. After adjusting for smoking status in the EWAS and conducting additional association testing in participants who never smoked cigarettes, we identified five cigarette smoking-independent CpGs that were significantly associated with lifetime cannabis use. The genes associated with these CpGs have been linked to various health outcomes, encompassing both disease risk and potential benefits related to cannabis use, highlighting the role of DNAm in the investigation of cannabis effects. These findings provide insights into DNAm profiles that are shared between smoking and cannabis use or specific to each substance, and suggest a substantial proportion of the variance in lifetime cannabis use are captured by DNAm. Follow-up studies are warranted to unravel the biological relevance of the differential DNAm to health outcomes.

DATA AVAILABILITY

The summary statistics from EWAS meta-analysis and analysis codes are available from the corresponding author upon reasonable request.

REFERENCES


ACKNOWLEDGEMENTS
This work was mainly supported by National Institute on Drug Abuse grant number R01DA048824 (PI: Fang). We also acknowledge the contributions of the staff and participants of all cohorts involved in this study, and cohort-specific funding and acknowledgement are included in Supplementary Section.

AUTHOR CONTRIBUTIONS
All authors reviewed the results and approved the manuscript. FF, BQ, DBH, and EOJ conceived and designed the study. FF, BQ, KGL, JvD, JAM, SL, ML, VVO, RC, LZ, MM, and YX conducted cohort-specific analyses. FF, JAM, and LZ performed meta-analysis. MO, JAT, JTB, JK, DIB, KX, and DPS oversaw the cohort research team and contributed to study design, analysis plan and interpretation of the results. JMV and LJB provided consulting to study design, analysis plan and interpretation of the results.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-023-02310-w.

Correspondence and requests for materials should be addressed to Fang Fang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023