

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

Fang Fang<sup>1\*</sup>, Bryan Quach<sup>1</sup>, Kaitlyn G. Lawrence<sup>2</sup>, Jenny van Dongen<sup>3</sup>, Jesse A. Marks<sup>1</sup>, Sara Lundgren<sup>4</sup>, Mingkuan Lin<sup>5</sup>, Veronika V. Odintsova<sup>3,11</sup>, Ricardo Costeira<sup>6</sup>, Zongli Xu<sup>2</sup>, Linran Zhou<sup>1</sup>, Meisha Mandal<sup>1</sup>, Yujing Xia<sup>6</sup>, Jacqueline M. Vink<sup>7</sup>, Laura J Bierut<sup>8</sup>, Miina Ollikainen<sup>4</sup>, Jack A. Taylor<sup>2</sup>, Jordana T. Bell<sup>6</sup>, Jaakko Kaprio<sup>4</sup>, Dorret I. Boomsma<sup>3</sup>, Ke Xu<sup>5,9</sup>, Dale P. Sandler<sup>2</sup>, Dana B. Hancock<sup>1</sup>, Eric O. Johnson<sup>1,10</sup>

1. GenOmics, Bioinformatics, and Translational Research Center, RTI International, Research Triangle Park, NC, USA.
2. Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA.
3. Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.
4. Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland.
5. Department of Psychiatry, Yale School of Medicine, West Haven, CT, USA.
6. Department of Twin Research & Genetic Epidemiology, King's College London, London, UK.
7. Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands.
8. Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, Missouri, USA.
9. VA Connecticut Healthcare System, West Haven, CT, USA.
10. Fellow Program, RTI International, Research Triangle Park, NC, USA.
11. Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

\*Corresponding Author

## Abstract

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 9,436 participants (7,795 European and 1,641 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 ( $p < 5.85 \times 10^{-7}$ ): 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 *APOBR*. 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.

## Introduction

Cannabis use is highly prevalent around the world.<sup>1</sup> In the United States, the legal use of cannabis has expanded across the states over time.<sup>2</sup> Despite the potential therapeutic benefits of medical use,<sup>3,4</sup> the widespread recreational use of cannabis has raised concerns because of its reported associations with many adverse health outcomes, including mental health (anxiety, depression, psychosis, schizophrenia, and mania),<sup>5-8</sup> cognitive deficits,<sup>9-11</sup> and addiction.<sup>12</sup> It is a pressing public health issue to better understand the full spectrum of the benefits and adverse consequences associated with cannabis use.

DNA methylation (DNAm), which involves the addition of a methyl group to the C5 position of cytosine in the context of CpG dinucleotides, has been extensively studied in relation to gene expression and can be influenced by the genome, the environment, and stochastic processes.<sup>13-15</sup> The DNAm changes induced by environmental exposure are sometimes persistent and long-lasting, while others are transient and reversible. For example, cigarette smoking has been shown to induce DNAm changes at CpGs throughout the genome. Some of these DNAm changes may revert after smoking cessation, while other DNAm changes may persist for years after cessation.<sup>16</sup>

In recent years, research toward understanding the effect of cannabis use on DNAm has grown.<sup>17</sup> Previous candidate gene studies identified DNAm changes of *CBI receptor*<sup>18</sup> and *DAT1*<sup>19</sup> in cannabis-dependent patients, *COMT* in adolescents defined as high-frequent cannabis users (> four times in the past 4 weeks),<sup>1</sup> and *DRD2* and *NCAM2* in moderate to heavy cannabis users (> 10 days in the last 30 days).<sup>20</sup> The first epigenome-wide association study (EWAS) of cannabis use, which compared DNAm between 12 cannabis users and 12 non-users in human sperm, found at least 10% DNAm differences at 3,979 CpG sites.<sup>21</sup> Our group performed the first blood-based EWAS of lifetime cannabis use (ever vs. never) in 2,583 women and found significant DNAm changes at cg15973234 (*CEMIP*).<sup>22</sup> Cannabis use-associated DNAm changes in blood have also been reported in heavy cannabis users (N = 96)<sup>23</sup> and adolescents (N = 525).<sup>24</sup> Taken together, these studies provide evidence that cannabis use impacts the epigenome, but knowledge of specific DNAm changes remain limited.

In this study, we conducted the largest trans-ancestry EWAS meta-analysis for lifetime cannabis use (ever vs. never) in 9,436 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. We explored this finding further by first adjusting the analyses for cigarette smoking status and next conducted 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 DNAm changes using differential methylation region (DMR) analysis, and integrating the DNAm findings with gene expression and genetic data.

## Results

***Sample characteristics.*** The demographic characteristics of 9,436 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. DNAm sites were assessed using either the Illumina HumanMethylation 450K Bead-Chip (76%) or the Illumina HumanMethylation EPIC (850K) Bead-Chip (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 9,436 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<sup>25,26</sup> (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 ( $p > 0.05/4$ ).<sup>26</sup> 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 1.

Considering potential differences by ancestry, we performed EWAS meta-analyses stratified into EA (N = 7,795) and AA (N = 1,641) groups. The ancestry-specific results for the top CpGs reported from the Model 2 EWAS meta-analysis are shown in Supplementary Table 1. 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. 4).

***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 = 4,146). 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 *APOBR*, which has been reported to be significantly associated with lifetime cannabis use in a genome-wide association study (GWAS)<sup>27</sup> (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 2. Their effect sizes were highly correlated with the EWAS meta-analysis results in all participants under either model ( $r > 0.7$ , Supplementary Fig. 2).

**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 3). The best-performing methylation score, based on 50 CpGs with  $p < 10^{-9}$  in the EWAS meta-analysis using Model 1, explained 3.79% of the variance of lifetime cannabis use in the Sister Study ( $p = 1.71 \times 10^{-17}$ ). 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 ( $p = 1.19 \times 10^{-3}$ ). 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  $10^{-5}$ .

**Sensitivity analysis.** We conducted additional analyses to adjust for other factors that influence DNA methylation as shown in previous studies: alcohol use<sup>28</sup> and BMI.<sup>29</sup> 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 1) were strongly correlated ( $r = 0.99$ ) with the results of the sensitivity analyses (Supplementary Fig. 5).

**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<sup>30</sup> (<https://ngdc.cncb.ac.cn/ewas/tools>), we found that three of the five CpGs were significantly correlated with nearby gene expression in brain (Supplementary Table 4,  $cor(\text{cg01101459}, LINC01132) = 0.121$ ,  $cor(\text{cg00813162}, ACTN1) = -0.135$ ,  $cor(\text{cg14237301}, APOBR) = -0.389$ )).

We also looked up the correlation of DNAm for the five CpGs between whole blood and brain regions (prefrontal cortex, entorhinal cortex, superior temporal gyrus, and cerebellum) through an existing database (<http://epigenetics.essex.ac.uk/bloodbrain/>).<sup>31</sup> Among the five CpGs, DNAm at cg22572071-*ADGRF1* and cg14237301-*APOBR* showed moderate correlations between whole blood and prefrontal cortex ( $r > 0.2$ ). At cg22572071-*ADGRF1* and cg15280358-*ADAM12*, DNAm was moderately correlated ( $r > 0.2$ ) between whole blood and cerebellum (Supplementary Fig. 6, Supplementary Table 5).

Enrichment analyses comparing the top cannabis use-associated CpGs from Model 2 to previously reported EWAS results in EWAS Atlas identified associations with other diseases and environment exposures (Supplementary Table 6). Among the enriched traits detected, Crohn's disease, alcohol consumption, BMI, and multiple sclerosis were significantly overlapped with our EWAS results. Interestingly, cannabis use-associated CpGs identified 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<sup>32</sup> (<http://mqtl.db.godmc.org.uk/about>), and overlapped them with the GWAS results for lifetime cannabis use<sup>27</sup> (Supplementary Table 7). 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-*ACTN1*

and cg14237301-*APOBR* showed some degree of significance in the GWAS ( $p < 0.01$ , Supplementary Table 7).

***Differentially methylated regions.*** We used an R software tool, ipDMR,<sup>33</sup> to identify DMRs based on EWAS meta-analysis results (Supplementary Table 8). 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 CpGs (cg00813162-*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 identified six DMRs with a minimum of two probes, none of which included the epigenome-wide significant CpG sites. The gene set enrichment analysis of the genes that overlap with the DMRs from Model 2 identified four Gene Ontology (GO) biological processes (Supplementary Fig. 7): growth, response to metal ion, actin filament bundle organization, and cellular response to zinc ion.

## **Discussion**

In this study, we performed the largest EWAS meta-analyses of lifetime cannabis use to date (9,436 multi-ancestry participants) using DNAm data from peripheral blood samples. The basic model showed that DNAm changes associated with lifetime cannabis use largely overlapped with cigarette smoking-associated DNAm sites. 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.<sup>34-44</sup>



The five epigenome-wide significant CpGs are annotated to the nearest genes: *LINC01132*, *ADGRF1*, *ADAM12*, *ACTN1*, and *APOBR*. Of these, cg01101459-*LINC01132*, cg00813162-*ACTN1*, and cg14237301-*APOBR* 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.<sup>34,45,46</sup> Lower expression of *LINC01132* has been associated with reduced oncogenicity. *ACTN1* (Alpha-Actinin-1) encodes a non-muscle cytoskeletal protein that binds actin to the cell membrane.<sup>35</sup> 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.<sup>35-38,47,48</sup> *APOBR* (Apolipoprotein B receptor) encodes a receptor protein that binds to dietary triglyceride-rich lipoproteins. Its genetic variants have been associated with obesity,<sup>39,49</sup> bladder cancer,<sup>42</sup> pneumonia,<sup>41</sup> allergy,<sup>40</sup> and lifetime cannabis use.<sup>27</sup> The results from our analyses using EWAS Atlas showed evidence that the expression of *LINC01132* in brain is positively correlated with the DNAm level of cg01101459, while the expression of *ACTN1* and *APOBR* in the brain are negatively correlated with the DNAm levels of the corresponding CpGs. However, both DNAm and gene expression patterns are tissue-specific, and their correlations merit further investigation in different tissues. Genetic variants in *APOBR* have been linked to lifetime cannabis use in GWAS.<sup>27</sup> However, we did not find overlap between the meQTLs and significant GWAS SNPs. Future EWAS integrating meQTLs is needed to confirm whether the association is genetic-driven.

On the other side, cg22572071-*ADGRF1* and cg15280358-*ADAM12* were positively associated with cannabis use. *ADGRF1* is a receptor gene that that is critical in neurodevelopment and neuroinflammation.<sup>43,50</sup> The overexpression of *ADGRF1* has been reported in breast cancer.<sup>51</sup> *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.<sup>52-56</sup> Genetic variants in *ADAM12* have been associated with neurological diseases such as multiple sclerosis and Alzheimer's disease. 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.<sup>26</sup> 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. 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.

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.<sup>57</sup> The variance in lifetime cannabis use explained by DNAm scores was (3.79%) is moderate. Future applications that integrate both polygenic risk scores and DNAm scores may improve prediction power.<sup>57,58</sup> 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 metal ion and zinc ion, and assembly of actin filament bundle.

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<sup>59</sup> 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. 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. Using data from an online database, we found moderate positive correlations between DNAm in whole blood and prefrontal cortex or cerebellum at three of the five cannabis use-associated CpG sites. Further experimental validation is needed to connect the DNAm changes in blood and other tissues that may relate to health effects associated with cannabis use. 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.<sup>60,61</sup>

The EWAS meta-analysis conducted in this study only reflects the association between lifetime cannabis use and DNAm levels, without any causal inference. More precise measurements of cannabis use regarding recency and frequency are needed to further investigate the persistent and transient effects on DNAm changes. 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. 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. Three of these CpGs showed significant correlations with the expression of nearby genes that have been linked to various health outcomes. 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.

#### **Methods** (less than 3,000 words)

***Study cohorts.*** This study included data from seven participating cohorts: the Sister Study,<sup>62</sup> Gulf Long-Term Follow-Up Study (GuLF),<sup>63</sup> Netherlands Twin Register (NTR),<sup>64</sup> Veteran Aging Cohort Study (VACS),<sup>65</sup> Finnish Twin Cohort (FinnTwin),<sup>66</sup> Avon Longitudinal Study of Parents and Children (ALSPAC),<sup>67</sup> and UK Adult Twin Registry (TwinsUK).<sup>68</sup> The final sample size consisted of 9,436 participants, including 4,190 individuals who reported ever using cannabis and 5,246 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 DNAm 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,<sup>69</sup> making the results comparable and combinable.

**DNA methylation measurements.** DNAm 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. DNAm levels were calculated as  $\beta$ -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).

**EWAS for lifetime cannabis use.** In each cohort, the association between 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 ancestry groups (EA and AA) and DNAm array types (450K and EPIC). For each CpG site, the DNAm beta value was considered as the outcome with lifetime cannabis use as the predictor 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,<sup>25</sup> the results from Model 2 were considered as cigarette smoking-independent DNAm 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.

**Meta-analysis.** We summarized cohort- and ancestry-specific EWAS results using inverse variance fixed effects meta-analysis implemented in the METAL software,<sup>70</sup> 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 rMVP.<sup>71</sup> 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.<sup>58</sup> At a given CpG site  $i$ , the methylation beta value ( $meth_i$ ) was multiplied by the effect size of the CpG in the meta-analysis ( $eff_i$ ). Then a methylation score was obtained by summing over a selected CpG set:

$$methylation\ score = meth_1 \cdot eff_1 + meth_2 \cdot eff_2 + \dots + meth_n \cdot eff_n$$

We applied different p-value thresholds to select the significant CpG sets for both Model 1 ( $p < 10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}, 10^{-11}, 10^{-13}$ ) and model 2 ( $p < 10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}$ ). To evaluate performance of the methylation scores, we quantified the percentage of variance ( $R^2$ ) explained as proposed by Lee et al.<sup>72</sup> 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.<sup>30</sup> 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/>)<sup>31</sup> 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<sup>30</sup> 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,<sup>32</sup> 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.<sup>27</sup>

***DMR analyses.*** We used an R software tool, ipDMR,<sup>33</sup> 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 1000bp and bin size 50bp. 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)<sup>73</sup> 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 DNA methylation 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 ethnic groups.*** Considering the differential DNAm patterns across different ethnicity groups, we conducted a separate EWAS meta-analysis in EA (N=7,795) and AA

(N=1,641) 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.

### **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. F.F., B.Q., D.B.H., and E.O.J. conceived and designed the study. F.F., B.Q., K.G.L., J.vD., J.A.M., S.L., M.L., V.V.O., R.C., Z.X., L.Z., M.M., and Y.X. conducted cohort-specific analyses. F.F., J.A.M., and L.Z. performed meta-analysis. M.O., J.A.T., J.T.B., J.K., D.I.B., K.X., and D.P.S. oversaw the cohort research team and contributed to study design, analysis plan and interpretation of the results. J.M.V. and L.J.B. provided consulting to study design, analysis plan and interpretation of the results.

### **Competing interests**

The authors declare no competing interests.



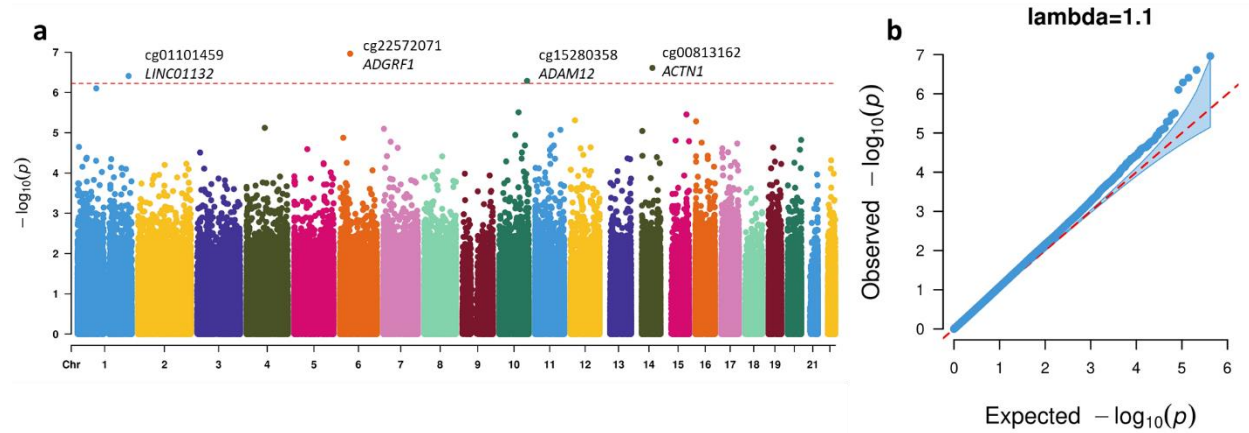


Figure 1. Manhattan (a) and QQ plots (b) from the EWAS meta-analysis for lifetime cannabis use with Model 2 adjusted for cigarette smoking. The dotted red line in (a) indicates the epigenome-wide significance cutoff at  $FDR < 0.05$  ( $P < 5.96 \times 10^{-7}$ ).

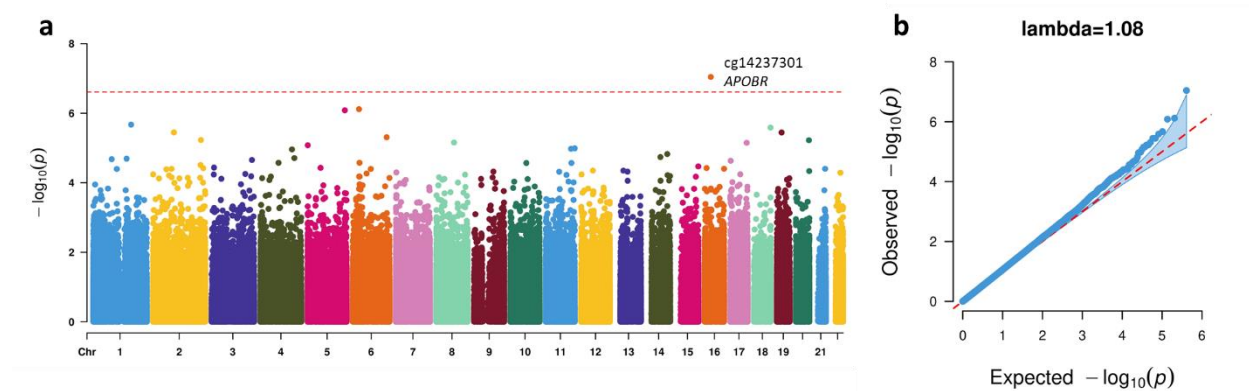


Figure 2. Manhattan (a) and QQ plots (b) from the EWAS meta-analysis for lifetime cannabis use in participants who never smoked cigarettes. The dotted red line in (a) indicates the epigenome-wide significance cutoff at  $FDR < 0.05$  ( $P < 2.45 \times 10^{-7}$ ).

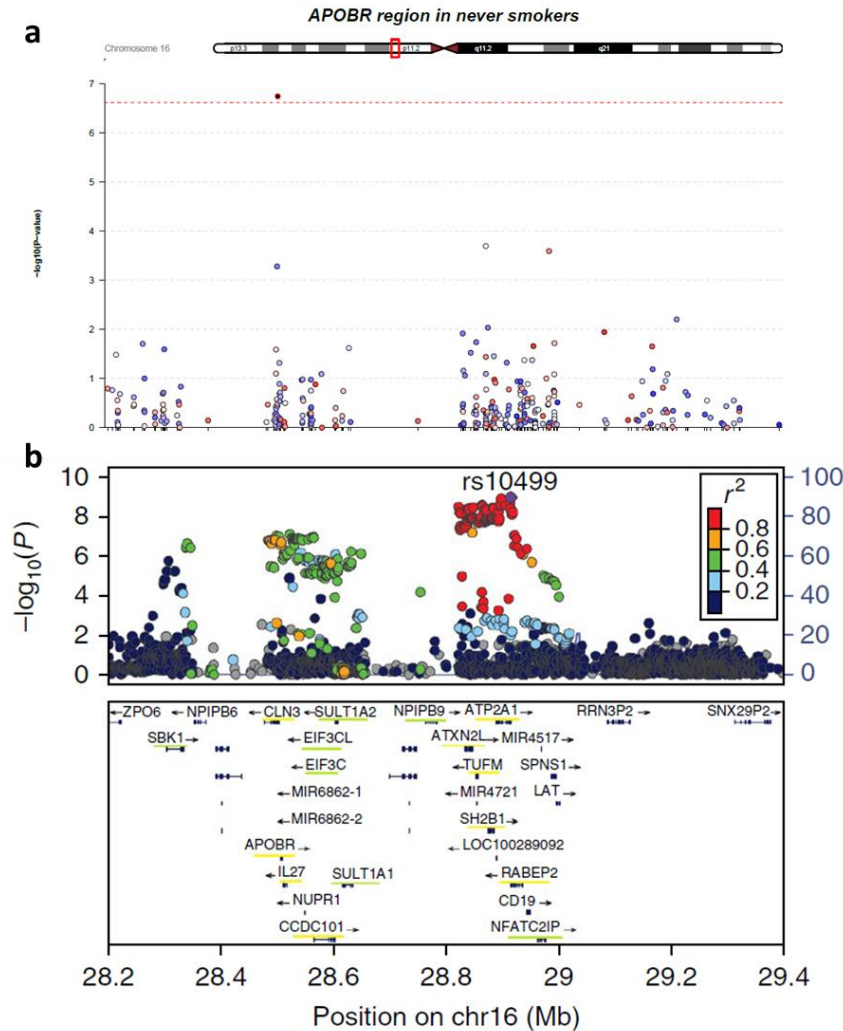


Figure 3. Regional plot for EWAS results in participants who never smoked cigarettes (a) and GWAS results (b) for lifetime cannabis use around the genes *APOBR/CLN3*. The x-axis shows the genomic position in base pair (bp) in hg19, while the y-axis shows the significance of associations ( $-\log_{10}$  p-values). (a) Each dot is a CpG probe, and the red dotted line indicates the epigenome-wide significance at  $FDR < 0.05$ . (b) Each dot is a SNP site, and the colors show different levels of LD in  $r^2$ . The box in the bottom includes the genes within the genomic region, and genes underlined in yellow were significant in the gene-based test while those underlined in green were identified in the S-PrediXcan analysis<sup>27</sup>.

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

Study Cohort	Total N	Countries	Ancestry	Female %	Age at blood draw*	DNAm array	Cannabis Ever users
Sister Study	2,583	US	100% EA <sup>§</sup>	100%	57.0±8.8	450K	49%
GuLF	1,195	US	56% EA 44% AA	0%	44.3±11.8	EPIC	56%
NTR	2,142	Netherland	100% EUR	70%	37.4±13.2	450K	26%
VACS	1,116	US	100% AA	0%	48.9±7.9	450K,EPIC	79%
FinnTwin	1,368	Finland	100% EUR	55%	23.5±1.9	450K,EPIC	27%
ALSPAC	922	UK	100% EUR	52%	17.1±1.0	450K	40%
TwinsUK	110	UK	100% EUR	100%	58.8±9.0	EPIC	27%
*Mean and standard deviation presented;							
§Abbreviations: EA: European American; EUR: European; AA: African American							

**Table 2. Epigenome-wide significant (FDR<0.05) CpGs identified in EWAS meta-analyses for lifetime cannabis use from Model 2, adjusted for cigarette smoking.**

CpG	Chr:Position (hg19)	Closest genes	Location <sup>#</sup>	Model 2 in all participants (N=9,436)			Model 1 in all participants (N=9,436)		Model 1 in participants who never smoked cigarettes (N=4,146)		EWAS meta-analysis of cigarette smoking <sup>26</sup> (current vs. non-current)		
				Direction	Effect*	P-value	Effect*	P-value	Effect*	P-value	Effect <sup>§</sup>	Direction	P-value
cg01101459	1:234871477	LINC01132	TFBS	-++++-	-0.0042	3.90E-07	-0.0030	2.27E-04	-0.0034	0.0064	0.18	++++	0.014
cg22572071	6:47074382	ADGRF1	Enhancer	+++++	0.0040	1.09E-07	0.0028	9.51E-05	0.0024	0.025	0.0074	++++	0.91
cg15280358	10:127904472	ADAM12	Gene body	+++?++++	0.0058	5.15E-07	0.0044	8.32E-05	0.0040	0.016	-0.11	+++	0.22
cg00813162	14:69443362	ACTN1	5'UTR	-+-----	-0.0028	2.46E-07	-0.0024	5.51E-06	-0.0025	0.0019	0.023	++?+	0.69
cg14237301	16:28506477	APOBR	1st Exon; 5' UTR	-----?-	-0.0040	1.76e-05	-0.0032	3.10E-4	-0.0071	9.05E-08	0.070	++++	0.37

<sup>#</sup>TFBS: Transcription Factor Binding Site; 5'UTR: 5' untranslated region.  
<sup>\*</sup>Positive effect size means higher methylation (hypermethylation) in ever cannabis users.  
<sup>§</sup>Positive effect size means higher methylation (hypermethylation) in current cigarette smokers.

## References

1. van der Knaap, L.J. *et al.* Catechol-O-methyltransferase gene methylation and substance use in adolescents: the TRAILS study. *Genes Brain Behav* **13**, 618-25 (2014).
2. Hasin, D. & Walsh, C. Trends over time in adult cannabis use: A review of recent findings. *Curr Opin Psychol* **38**, 80-85 (2021).
3. Wang, Y. *et al.* Low-Dose Delta-9-Tetrahydrocannabinol as Beneficial Treatment for Aged APP/PS1 Mice. *Int J Mol Sci* **23**(2022).
4. Bogale, K., Raup-Konsavage, W., Dalessio, S., Vrana, K. & Coates, M.D. Cannabis and Cannabis Derivatives for Abdominal Pain Management in Inflammatory Bowel Disease. *Med Cannabis Cannabinoids* **4**, 97-106 (2021).
5. Radhakrishnan, R., Wilkinson, S.T. & D'Souza, D.C. Gone to Pot - A Review of the Association between Cannabis and Psychosis. *Front Psychiatry* **5**, 54 (2014).
6. Gibbs, M. *et al.* Cannabis use and mania symptoms: a systematic review and meta-analysis. *J Affect Disord* **171**, 39-47 (2015).
7. Malone, D.T., Hill, M.N. & Rubino, T. Adolescent cannabis use and psychosis: epidemiology and neurodevelopmental models. *Br J Pharmacol* **160**, 511-22 (2010).
8. Leweke, F.M. & Koethe, D. Cannabis and psychiatric disorders: it is not only addiction. *Addict Biol* **13**, 264-75 (2008).
9. Crean, R.D., Crane, N.A. & Mason, B.J. An evidence based review of acute and long-term effects of cannabis use on executive cognitive functions. *J Addict Med* **5**, 1-8 (2011).
10. Gilman, J.M. *et al.* Identification of 9-tetrahydrocannabinol (THC) impairment using functional brain imaging. *Neuropsychopharmacology* **47**, 944-952 (2022).
11. Alegria, A.A. *et al.* Comorbidity of generalized anxiety disorder and substance use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry* **71**, 1187-95; quiz 1252-3 (2010).
12. Lopez-Quintero, C. *et al.* Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Drug Alcohol Depend* **115**, 120-30 (2011).
13. Lowdon, R.F., Jang, H.S. & Wang, T. Evolution of Epigenetic Regulation in Vertebrate Genomes. *Trends Genet* **32**, 269-283 (2016).
14. Straussman, R. *et al.* Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol* **16**, 564-71 (2009).
15. Moore, L.D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23-38 (2013).
16. Kaur, G., Begum, R., Thota, S. & Batra, S. A systematic review of smoking-related epigenetic alterations. *Arch Toxicol* **93**, 2715-2740 (2019).
17. Szutorisz, H. & Hurd, Y.L. Epigenetic Effects of Cannabis Exposure. *Biol Psychiatry* **79**, 586-94 (2016).
18. Rotter, A. *et al.* CB1 and CB2 receptor expression and promoter methylation in patients with cannabis dependence. *Eur Addict Res* **19**, 13-20 (2013).
19. Grzywacz, A. *et al.* Contribution of Dopamine Transporter Gene Methylation Status to Cannabis Dependency. *Brain Sci* **10**(2020).
20. Gerra, M.C. *et al.* Gene variants and educational attainment in cannabis use: mediating role of DNA methylation. *Transl Psychiatry* **8**, 23 (2018).
21. Murphy, S.K. *et al.* Cannabinoid exposure and altered DNA methylation in rat and human sperm. *Epigenetics* **13**, 1208-1221 (2018).

22. Markunas, C.A. *et al.* Epigenome-wide analysis uncovers a blood-based DNA methylation biomarker of lifetime cannabis use. *Am J Med Genet B Neuropsychiatr Genet* **186**, 173-182 (2021).
23. Osborne, A.J. *et al.* Genome-wide DNA methylation analysis of heavy cannabis exposure in a New Zealand longitudinal cohort. *Transl Psychiatry* **10**, 114 (2020).
24. Clark, S.L. *et al.* Methylomic Investigation of Problematic Adolescent Cannabis Use and Its Negative Mental Health Consequences. *J Am Acad Child Adolesc Psychiatry* **60**, 1524-1532 (2021).
25. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* **9**, 436-447 (2016).
26. Christiansen, C. *et al.* Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects. *Clin Epigenetics* **13**, 36 (2021).
27. Pasman, J.A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci* **21**, 1161-1170 (2018).
28. Lohoff, F.W. *et al.* Epigenome-wide association study of alcohol consumption in N = 8161 individuals and relevance to alcohol use disorder pathophysiology: identification of the cystine/glutamate transporter SLC7A11 as a top target. *Mol Psychiatry* **27**, 1754-1764 (2022).
29. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81-86 (2017).
30. Li, M. *et al.* EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res* **47**, D983-D988 (2019).
31. Hannon, E., Lunnon, K., Schalkwyk, L. & Mill, J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* **10**, 1024-32 (2015).
32. Min, J.L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* **53**, 1311-1321 (2021).
33. Xu, Z., Xie, C., Taylor, J.A. & Niu, L. ipDMR: identification of differentially methylated regions with interval P-values. *Bioinformatics* **37**, 711-713 (2021).
34. Zhang, J. *et al.* Long noncoding RNA LINC01132 enhances immunosuppression and therapy resistance via NRF1/DPP4 axis in hepatocellular carcinoma. *J Exp Clin Cancer Res* **41**, 270 (2022).
35. Kunishima, S. *et al.* ACTN1 mutations cause congenital macrothrombocytopenia. *Am J Hum Genet* **92**, 431-8 (2013).
36. Li, Y., Shu, J., Cheng, Y., Zhou, X. & Huang, T. Identification of key biomarkers in Angelman syndrome by a multi-cohort analysis. *Front Med (Lausanne)* **9**, 963883 (2022).
37. Xie, T. *et al.* Proteomics analysis of lysine crotonylation and 2-hydroxyisobutyrylation reveals significant features of systemic lupus erythematosus. *Clin Rheumatol* (2022).
38. Glessner, J.T. *et al.* COVID-19 in pediatrics: Genetic susceptibility. *Front Genet* **13**, 928466 (2022).
39. Voisin, S. *et al.* Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. *Genome Med* **7**, 103 (2015).
40. Ferreira, M.A.R. *et al.* Eleven loci with new reproducible genetic associations with allergic disease risk. *J Allergy Clin Immunol* **143**, 691-699 (2019).
41. Campos, A.I. *et al.* Genetic Susceptibility to Pneumonia: A GWAS Meta-Analysis Between the UK Biobank and FinnGen. *Twin Res Hum Genet* **24**, 145-154 (2021).
42. Hemissi, I. *et al.* Development of a custom next-generation sequencing panel for the determination of bladder cancer risk in a Tunisian cohort. *Mol Biol Rep* **49**, 1233-1258 (2022).

43. Lee, J.W. *et al.* Orphan GPR110 (ADGRF1) targeted by N-docosahexaenoylethanolamine in development of neurons and cognitive function. *Nat Commun* **7**, 13123 (2016).
44. Kveiborg, M., Albrechtsen, R., Couchman, J.R. & Wewer, U.M. Cellular roles of ADAM12 in health and disease. *Int J Biochem Cell Biol* **40**, 1685-702 (2008).
45. Meng, C., Zhou, J.Q. & Liao, Y.S. Autophagy-related long non-coding RNA signature for ovarian cancer. *J Int Med Res* **48**, 300060520970761 (2020).
46. Zhu, W., Xiao, X. & Chen, J. Silencing of the long noncoding RNA LINC01132 alleviates the oncogenicity of epithelial ovarian cancer by regulating the microRNA4315p/SOX9 axis. *Int J Mol Med* **48**(2021).
47. Biao, T. *et al.* From Bowen disease to cutaneous squamous cell carcinoma: eight markers were verified from transcriptomic and proteomic analyses. *J Transl Med* **20**, 416 (2022).
48. Zhang, L. *et al.* Bioinformatics analysis identification of AKT3 and RAC1 as key genes in postmenopausal osteoporosis. *Exp Ther Med* **24**, 656 (2022).
49. Liu, W., Xu, J., Dai, T., Shen, G. & Feng, J. An integrating strategy for serum metabolomics and microarray analysis to expand the understanding of diet-induced obesity. *Anal Methods* **13**, 3127-3135 (2021).
50. Park, T., Chen, H. & Kim, H.Y. GPR110 (ADGRF1) mediates anti-inflammatory effects of N-docosahexaenoylethanolamine. *J Neuroinflammation* **16**, 225 (2019).
51. Abdulkareem, N.M. *et al.* A novel role of ADGRF1 (GPR110) in promoting cellular quiescence and chemoresistance in human epidermal growth factor receptor 2-positive breast cancer. *FASEB J* **35**, e21719 (2021).
52. Wang, H., Wu, H. & Chen, Y. Serum ADAM metallopeptidase domain 12 as a promising diagnostic biomarker for breast and liver cancer. *Clin Transl Oncol* (2022).
53. Chung, M.W., Park, Y.L., Park, S.Y. & Joo, Y.E. Expression of ADAM12 in Gastric Cancer and its Relation to Tumor Cell Behavior and Prognosis. *In Vivo* **36**, 2194-2204 (2022).
54. Park, Y.L. *et al.* A Disintegrin and Metalloprotease 12 Promotes Tumor Progression by Inhibiting Apoptosis in Human Colorectal Cancer. *Cancers (Basel)* **13**(2021).
55. Shao, S. *et al.* ADAM-12 as a diagnostic marker for the proliferation, migration and invasion in patients with small cell lung cancer. *PLoS One* **9**, e85936 (2014).
56. Frohlich, C. *et al.* Molecular profiling of ADAM12 in human bladder cancer. *Clin Cancer Res* **12**, 7359-68 (2006).
57. McCartney, D.L. *et al.* Epigenetic prediction of complex traits and death. *Genome Biol* **19**, 136 (2018).
58. Odintsova, V.V. *et al.* DNA methylation in peripheral tissues and left-handedness. *Sci Rep* **12**, 5606 (2022).
59. Saunders, G.R.B. *et al.* Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720-724 (2022).
60. Michels, K.B. *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* **10**, 949-55 (2013).
61. Houseman, E.A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
62. Sandler, D.P. *et al.* The Sister Study Cohort: Baseline Methods and Participant Characteristics. *Environ Health Perspect* **125**, 127003 (2017).
63. Stewart, P.A. *et al.* Development of a total hydrocarbon ordinal job-exposure matrix for workers responding to the Deepwater Horizon disaster: The GuLF STUDY. *J Expo Sci Environ Epidemiol* **28**, 223-230 (2018).
64. Boomsma, D.I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* **9**, 849-57 (2006).
65. Justice, A.C. *et al.* Veterans Aging Cohort Study (VACS): Overview and description. *Med Care* **44**, S13-24 (2006).

66. Kaprio, J. The Finnish Twin Cohort Study: an update. *Twin Res Hum Genet* **16**, 157-62 (2013).
67. Relton, C.L. *et al.* Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int J Epidemiol* **44**, 1181-90 (2015).
68. Moayyeri, A., Hammond, C.J., Hart, D.J. & Spector, T.D. The UK Adult Twin Registry (TwinsUK Resource). *Twin Res Hum Genet* **16**, 144-9 (2013).
69. Conway, K.P. *et al.* Data compatibility in the addiction sciences: an examination of measure commonality. *Drug Alcohol Depend* **141**, 153-8 (2014).
70. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
71. Yin, L. *et al.* rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics Proteomics Bioinformatics* **19**, 619-628 (2021).
72. Lee, S.H., Goddard, M.E., Wray, N.R. & Visscher, P.M. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* **36**, 214-24 (2012).
73. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).