DNA methylation and substance use risk: A prospective, genome-wide study spanning gestation to adolescence

SELF-ARCHIVING VERSION

Charlotte A.M. Cecil, PhD¹; Esther Walton, PhD¹,²; Rebecca G. Smith, PhD³; Essi Viding, PhD⁴; Eamon J. McCrory, PhD⁴; Caroline L. Relton, PhD⁵; Matthew Suderman, PhD⁵; Jean-Baptiste Pingault, PhD⁴; Wendy McArdle, PhD⁶; Tom R. Gaunt, 12PhD⁵; Jonathan Mill, PhD³; and Edward D. Barker, PhD¹

1. Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK.
2. Department of Psychology, Georgia State University, USA.
3. Exeter Medical School, Exeter University, UK.
4. Division of Psychology and Language Sciences, University College London, UK.
5. Medical Research Council Integrative Epidemiology Unit, University of Bristol, UK.
6. School of Social and Community Medicine, University of Bristol, UK.

*Co-corresponding authors


Correspondence to: Edward D. Barker/Charlotte A.M. Cecil, Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, De Crespigny Park, London, SE5 8AF, UK. Tel: +44 (0)207 848 0992/ +44(0)207 848 0389, E-mail: ted.barker@kcl.ac.uk / charlotte.cecil@kcl.ac.uk.
Abstract
Epigenetic processes have been implicated in addiction; yet, it remains unclear whether these represent a risk factor and/or a consequence of substance use. Here, we conducted the first genome-wide, longitudinal study to investigate whether DNA methylation patterns in early life prospectively associate with substance use in adolescence. The sample comprised of 244 youth (51% female) from the Avon Longitudinal Study of Parents and Children (ALSPAC), with repeated assessments of DNA methylation (Illumina 450k array; cord blood at birth, whole blood at age 7) and substance use (tobacco, alcohol and cannabis use; age 14-18). We found that, at birth, epigenetic variation across a tightly interconnected genetic network (n = 65 loci; q<0.05) associated with greater levels of substance use during adolescence, as well as an earlier age of onset amongst users. Associations were specific to the neonatal period and not observed at age 7. Key annotated genes included PACSIN1, NEUROD4 and NTRK2, implicated in neurodevelopmental processes. Several of the identified loci were associated with known mQTLs, and consequently likely to be under significant genetic control. Collectively, these 65 loci were also found to partially mediate the effect of prenatal maternal tobacco smoking on adolescent substance use. Together, findings lend novel insights into epigenetic correlates of substance use, highlight birth as a potentially sensitive window of biological vulnerability and provide preliminary evidence of an indirect epigenetic pathway linking prenatal tobacco exposure and adolescent substance use.

Keywords
Substance use; DNA methylation; epigenome-wide; prenatal exposure; tobacco; mediation
DNA methylation and substance use risk

Introduction
Substance abuse is a major public health concern that incurs heavy costs to individuals, their families and wider society. Collectively, it is estimated that one in ten of all fatalities result from harmful use of alcohol, tobacco and illicit drugs, representing one of the leading preventable causes of death worldwide (1). In addition, substance abuse is associated with a range of negative outcomes that compromise quality of life and long-term productivity, including psychiatric illness, disability, criminality and unemployment (2, 3). Consequently, a key challenge for research is to identify factors that drive individual susceptibility to substance abuse, in order to inform effective prevention and early intervention strategies (4).

Like most complex phenotypes, substance abuse results from a dynamic interplay of genetic and environmental influences. Studies have shown that the heritability of substance use disorders is moderate to high (~49-70%; 5), and that this genetic vulnerability interacts with environmental risk exposure. Indeed, epidemiological studies have identified a number of pre- and postnatal factors associated with substance abuse risk, including substance exposure during pregnancy, parental psychopathology and criminality, low socioeconomic status, childhood maltreatment and affiliation with delinquent peers (6). However, the biological mechanisms through which these effects are mediated are poorly understood.

In recent years, epigenetic processes that regulate gene expression (7) have emerged as a potential mechanism of interest. One of these processes, DNA methylation (DNAm), has received increasing attention. DNAm modulates transcription via the addition of a methyl group to DNA base pairs, primarily in the context of cytosine-guanine (CpG) dinucleotides (7). Studies have shown that (i) DNAm is affected by genetic variability, as demonstrated by the discovery of a large number of methylation quantitative trait loci (mQTLs; 8, 9); (ii) DNAm is also sensitive to pre- and postnatal environmental influences, including nutritional, chemical, and psychosocial factors (e.g. prenatal tobacco exposure; 10, 11); and (iii) aberrant patterns of DNAm have been linked to a wide range of physical and psychiatric disorders, including addiction (12). For example, animal studies have shown that repeated drug administration (e.g. alcohol, cocaine) can lead to DNAm changes in reward-related regions of the brain (e.g. striatum; 13). In turn, these changes can influence the expression of genes involved in synaptic plasticity and memory consolidation, driving neuroadaptations that underlie the onset and persistence of addictive behaviors (14). Importantly, drug-induced epigenetic changes have been found to occur as early as gestation (12). For example, a recent study in mice reported an epigenetically-mediated effect of early nicotine exposure on pup’s neural structure and behavior, which persisted into adulthood (15). So far, studies in humans have provided initial support for animal findings, reporting methyloic differences between substance abusers and drug-free controls across several tissue types and substances (12, 16).

Despite these promising findings, current research in humans has been limited in four key ways (17). First, the vast majority of studies have examined adult samples already exposed to substances. As a result, it has not been possible to establish whether altered DNAm patterns are a risk factor for and/or consequence of substance use. Disentangling these associations is essential in order to better delineate the role of epigenetic mechanisms in addiction risk and to enable the identification of novel therapeutic targets. Second, because existing studies have typically included DNAm data at a single time point, it is unclear whether epigenetic effects may be observable across time or only during specific
DNA methylation and substance use risk
developmental periods. This is particularly relevant given that DNAm has been shown to be
highly dynamic across the lifespan, enabling cells to respond to changing internal and
external inputs (18). As such, clarifying how DNAm associates with substance use over time
may provide important insights into windows of biological vulnerability. Third, little is
known about what genetic and environmental factors may underlie variability in DNAm
patterns associated with substance use. Characterizing these potential influences may not only
offer valuable opportunities for preventative intervention, but also make it possible to test the
role of the epigenome as a potential mediator in the link between risk exposure and later
substance use. Finally, existing studies have primarily focused on one type of substance at a
time. Although substance-specific risk factors have been identified, evidence from both
genetically-informative and epidemiological studies indicate that substance use risk across
drug classes is largely accounted for by a common underlying liability dimension (5, 19).
Consequently, examining epigenetic markers common to multiple substances, in addition to
substance-specific markers, may help shed further light into the biological basis of substance
use liability.

To address these gaps in the literature, we conducted the first genome-wide, prospective
study to examine associations between DNAm in early life (i.e. collected at repeated time-
points pre-substance use initiation; birth, age 7) and substance use in adolescence (measured
as a latent factor spanning tobacco, cannabis and alcohol use). Our aim was to address the
following key questions:

1. Are DNAm patterns at birth prospectively associated with adolescent substance use?
2. Are these associations stable across early childhood (birth to age 7)?
3. Do the identified DNAm markers associate with genetic and environmental influences?

Materials and Methods

Participants
The Epigenetic Pathways to Conduct Problems Study consists of a subsample of youth
(n=339) drawn from the Avon Longitudinal Study of Parents and Children (ALSPAC) who
(i) have repeated measures of DNAm and (ii) follow previously established trajectories of
conduct problems (4-13 years; 20). Only youth who had complete substance use ratings (age
14-18) as well as epigenetic data at birth and age 7 (n = 244, 54% female) were included in
the present study. ALSPAC is an ongoing epidemiological study of children born between
1991-92 from 14,541 women residing in Avon, UK. Of these initial pregnancies, there was a
total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1
year of age (21). When compared to 1991 National Census Data, the ALSPAC sample was
found to be broadly similar to the UK population as a whole (22). Informed consent was
obtained from all ALSPAC participants and ethical approval was obtained from the ALSPAC
Law and Ethics Committee and the Local Research Ethics Committees. Please note that the
study website contains details of all the data that is available through a fully searchable data
dictionary: http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/.
**Measures**

**Adolescent substance use**

Substance use was assessed via self-report ratings of tobacco and cannabis use (age 14, 16 and 18 using frequency items ranging from ‘never’ to ‘daily’), as well as alcohol use (age 16 and 18 using the 10-item Alcohol Use Disorders Identification Test; 23). Confirmatory factor analysis was used to extract (i) three first-order factors of tobacco, cannabis and alcohol use, accounting for shared variance across time points for each of these substances; and (ii) a single second-order factor of substance use, accounting for shared variance between substances *and* across time. The factor model showed adequate fit: $\chi^2(18) = 49.55; p < .01$; comparative fit index (CFI) = .91; Tucker-Lewis index (TLI) = .86; root mean square error of approximation (RMSEA) = .08, 90% CIs = .05, .10; with standardized loadings ranging from .58 to .96 (see Supplementary Figure 1).

**DNA methylation data**

500ng high molecular weight genomic DNA from blood (cord at birth, whole at age 7) was bisulfite-converted using the EZ-DNA methylation kit (Zymo Research, Orange, CA, USA). DNAm was quantified using the Illumina HumanMethylation450 BeadChip (Illumina, USA) with arrays scanned using an Illumina iScan (software version 3.3.28). Initial data quality control was conducted using GenomeStudio (version 2011.1) to determine the status of staining, hybridization, target removal, bisulfite conversion, specificity, non-polymorphic and negative controls. Samples that survived this stage were quantile normalised using the dasen function within the watermelon 1.0.3 package (24) in R and batch corrected using the ComBat package (25). Probes were removed if they were cross-reactive, used for sample identification on the array, or had a SNP at the single base extension with a minor allele frequency larger than 5% (i.e. common polymorphisms), leaving a total of 413,510 probes (26, 27). DNAm levels are indexed by beta values (ratio of: methylated signal/methylated + unmethylated signal).

**Prenatal environmental risks**

We included prenatal risks that have been previously linked to adolescent substance use, including maternal prenatal smoking, alcohol use and exposure to stressful events (6).

Maternal smoking and alcohol use during the first trimester of pregnancy were measured via maternal ratings, using a yes/no binary variable for smoking (for biological validation see the results section), and a 4-point scale for alcohol use (‘never’ to ‘daily’). With regards to stress exposure, we included cumulative risk scores of prenatal (18 to 32 weeks) adversity covering the following four domains: (i) *life events* (e.g. death in family, accident); (ii) *contextual risks* (e.g. poor housing, financial problems); (iii) *parental risks* (e.g. psychopathology, criminal behaviour); and (iv) *interpersonal risks* (e.g. partner abuse, family conflict). These cumulative risk scores were estimated using confirmatory factor analysis based on maternal reports, as described elsewhere (28).
**Data Analysis**

Analyses were performed in R (version 3.0.1; 29) and Mplus (version 6.1.1; 30), adjusting for sex and cell-type proportions (CD8 T lymphocytes, CD4 T lymphocytes, natural killer cells, B lymphocytes, monocytes), estimated using the reference-based approach detailed in Houseman et al (31).

Step 1: Are DNAm patterns at birth associated with adolescent substance use?

Genome-wide association analyses between DNAm at birth and substance use were performed using the IMA package (32). Differentially methylated probes (DMPs) passing a False Discovery Rate (FDR) correction of q<0.05 were considered significant. These DMPs were then uploaded to the UCSC genome browser (GRCh37/hg19 assembly; 33) to explore their potential functional relevance, by comparing their genomic location to that of key regulatory elements recorded in the Encyclopedia of DNA Elements (ENCODE) database (http://genome.ucsc.edu/ENCODE/), including (i) transcription factor binding sites (data generated on 161 transcription factors in 91 cell types via ChIP-seq); (ii) DNase I hypersensitivity clusters (based on data from 125 cell types), and (iii) histone marks (only relevant cell lines examined, including blood [GM12878, K562] and umbilical vein endothelial [HUVEC] cells).

Genes to which DMPs were annotated were then examined to identify (i) underlying genetic networks, using the GeneMANIA bioinformatics software, which is based on known genetic and physical interactions, shared protein domains as well as co-expression data (http://www.genemania.org; see Supplementary Table 1); and (ii) enriched biological pathways, by using an optimized gene ontology method that controls for a range of potential confounds, including background probe distribution and gene size (Supplementary Table 1).

As a supplement to the probe-level analysis, we also used the Comb-p application within Python (34) with default settings (p threshold: 1.00E-04; sliding window size: 500bp), to identify wider differentially methylated regions based on spatially-correlated p-values.

Step 2: Are these associations stable across early childhood (i.e. birth to age 7)?

Given that DNAm is temporally dynamic (18) – particularly in early development (9) – markers identified at one time point may not necessary continue to be associated with substance use at other time points. To test this, we examined whether DMPs identified in Step 1 (i.e. birth) were also significantly associated with adolescent substance use at age 7 (i.e. follow-forward approach; FDR-corrected q<0.05).

Step 3: Do these markers relate to genetic and environmental influences?

As a last step, we investigated potential genetic and environmental factors that may influence DNAm levels of the identified DMPs. Given that our sample was underpowered to directly examine genetic polymorphisms (i.e. SNPs) affecting DNAm, we used the ALSPAC-derived mQTLdb resource (http://www.mqtldb.org/) to search for known methylation quantitative trait loci (mQTLs) associated with our DMPs (see Supplementary Table 1 for further details). Potential environmental influences were examined next by testing associations between prenatal exposures and DMPs. Because of the large number of DMPs identified, we grouped these into a single, cumulative DNA methylation risk score to minimize multiple testing burden. Specifically, we applied a method typically used for polygenic risk scores (35), where we
multiplied the methylation values of our DMPs by their respective standardized regression betas (i.e. weights), and then summed these together into a DNAm risk score. This approach enabled us to reduce the volume of our methylation data, while the use of weights ensured that DMPs maintained their relative predictive importance (i.e. as opposed to alternative approaches, e.g. averaging DNAm levels across DMPs). Once calculated, we examined associations between this DNAm risk score and prenatal exposures, using Pearson’s bivariate correlations. Significant prenatal risks (q<0.05) were then incorporated into a single path analytic model in Mplus (maximum likelihood estimation), together with the DNAm risk score and the substance use factor, in order to test for indirect effects. Associations in the model were considered significant if they survived bootstrapped confidence intervals (10,000 times; 36). Significant paths (i.e. prenatal risks $\rightarrow$ DNAm $\rightarrow$ substance use) were tested for an indirect effect using bootstrapped model constraint statements.

**Code availability**
Computer code used in our analyses is available from the authors upon request.

**Results**

**Epigenome-wide association analysis at birth**
At birth, 65 probes prospectively associated with adolescent substance use after genome-wide correction (q<0.05; Table 1; Figure 1A). Of these DMPs, 33 were ‘hypomethylated’ (i.e. lower DNAm associating with higher substance use), while the other 32 were ‘hypermethylated’ (i.e. higher DNAm associating with higher substance use). Overall, DMPs were most frequently located in the gene body (40%) or promoter region near the transcription start site (30%; see Supplementary Table 2). DNAm levels were significantly interrelated across the majority of DMPs (76% of correlations = q<0.05; $r_{max}$=0.58; $r_{min}$= -0.52; $r_{absolute\ average}$ =0.20; see Supplementary Table 3). The most significant probe, cg04941418 ($p = 1.10E-08; q = 0.005$, Figure 1B), is located in PACSIN1, a developmentally regulated gene that plays an important role in synaptic neurotransmission, axonal growth and dendritic branching (37, 38). Other annotated genes in the table include (i) SHC2 (cg02290110) and NTRK2 (cg01009697), both implicated in neuronal neurotrophin-activated Trk receptor signaling (39, 40), (ii) CLSTN1 (cg07395930), involved in calcium-mediated postsynaptic signals, and (iii) NEUROD4 (cg20056324), involved in neural differentiation. DMPs were then uploaded in Genome Browser for functional characterization, based on ENCODE data on regulatory elements. All DMPs overlapped with histone marks; 62% ($n = 40$) coincided with transcription factor binding sites; and 57% ($n = 37$) were located within DNAse I hypersensitive clusters. Overall, 48% ($n = 31$) of DMPs were mapped to all three regulatory elements (Supplementary Table 4).

DMPs were annotated to a total of 60 genes, which were examined further to identify underlying genetic networks and enriched biological pathways. Based on GeneMANIA analysis, 57 of the 60 genes were connected to form a compact cluster network (Figure 1C). Our gene ontology analysis also indicated that these genes are involved in a range of biological processes, including regulation of JAK-STAT cascade, vasoconstriction, cytokine-mediated signaling and axonogenesis (2.30E-18<$p<3.37E-03$; Figure 1D). Of note, enriched
cellular components included axon part, postsynaptic membrane and dendritic spine ($2.94 \times 10^{-04} < p < 3.09 \times 10^{-03}$; for the full list of GO terms, see Supplementary Table 5).

Results from the Comb-p analysis indicated that there were no significant differentially methylated regions after genome-wide correction.

Follow-forward at age 7

None of the DMPs identified at birth continued to prospectively associate with adolescent substance use by age 7, after multiple correction (i.e. q>0.05; Supplementary Table 6). Two DMPs showed nominal associations (cg02404636 [SFI1]: StdB =0.21, $p=0.001$; cg20056324 [NEUROD4]: StdB=0.13, $p=0.05$), both following the same direction of effects observed at birth. Given this lack of temporal stability, we proceeded to test, for each DMP, how much DNAm levels at birth correlated with those at age 7 (i.e. autocorrelation). We found that only 12 DMPs (18%) showed an autocorrelation significant at $p<0.05$, 11 of which were in the positive direction (across all DMPs: $r_{\text{max}}=0.67$; $r_{\text{min}}=-0.13$; $r_{\text{absolute average}}=0.07$). Interestingly, however, the pattern of intercorrelations across DMPs at age 7 resembled that observed at birth (Supplementary Figure 2). In other words, while DMPs typically did not correlate with themselves over time, the way in which they correlated with each-other within time points was very similar, potentially reflecting a similar underlying co-methylation network.

Genetic and environmental influences

The 65 DMPs identified at birth were carried forward to explore associations with potential genetic and environmental influences. Based on mQTLdb search, we found that five of the DMPs were associated with known mQTLs ($n_{\text{cis}}=4$; $n_{\text{trans}}=1$), suggesting that DNAm levels across these sites are likely to be under considerable genetic control (Supplementary Table 4). Of note, temporal stability of these DMPs was stronger ($r_{\text{average}}=0.25$) than the average across all DMPs noted above, consistent with what has previously been observed at the genome-wide level (9). With regards to environmental influences, we found that three prenatal exposures significantly correlated with DNAm (measured as a cumulative risk score comprising of all DMPs) – maternal tobacco smoking, maternal risks and contextual risks (Table 2). To test for indirect effects, we estimated a path analytic model (Figure 2A) that included these three prenatal exposures, the cumulative DNAm risk score, and the adolescent substance use outcome. Maternal smoking was the only prenatal factor to uniquely associate with higher cumulative DNAm risk (over and above other exposures), which in turn associated with higher substance use in adolescence (Figure 2B). Analysis of this pathway indicated a significant indirect effect of maternal smoking on substance use, via cumulative DNAm risk ($b=0.19$, $s.e.=0.07$, $p=0.01$, bootstrapped 95% CI = 0.05-0.37). To minimize the possibility that associations with prenatal exposures may simply reflect genetic confounding, we reran analyses using a cumulative DNAm risk score that did not include any of the DMPs associated with mQTLs (i.e. 5 probes removed). This score was highly correlated with the original score ($r=0.99$; $p=3.88\times10^{-252}$) and findings remained consistent.
Follow-up analyses

**PACSIN1: Relevance to the brain**

**PACSIN1**\textsubscript{cg04941418} emerged as the top DMP at birth to associate with adolescent substance use. Given that our DNAm data was extracted from peripheral blood, we used the Genotype-Tissue Expression project portal (GTex; http://www.gtexportal.org/home/; 41) and the EMBL-EBI Expression Atlas (https://www.ebi.ac.uk/gxa/home/; 42) to assess **PACSIN1** expression across tissues. **PACSIN1** was found to be most highly expressed in brain tissue, including regions implicated in drug-seeking behavior and addiction risk, such as the prefrontal cortex, nucleus accumbens, amygdala, and hippocampus (Supplementary Figure 3). We then used the BrainCloud tool (http://braincloud.jhmi.edu/plots/; 43), to trace the developmental course of **PACSIN1** expression across the lifespan (fetal – age 80), based on postmortem prefrontal cortex tissue from 269 healthy subjects. The resulting plot showed that the most dramatic change in expression levels occurs during the neonatal period, bridging lower expression levels during fetal development with a higher, stable trajectory of expression from around three months of age onward (Supplementary Figure 3).

Age of substance use onset

65 DMPs at birth prospectively associated with substance use severity in adolescence. As a sensitivity analysis, we additionally tested whether these DMPs also associated with age of onset amongst substance users. Based on three items that combined self-report data across age 16 and 18, we found that, within youth who endorsed using substances, higher cumulative DNAm risk correlated with lower reported age when first ‘smoked whole cigarette’ ($r = -0.19, p = 0.03, n\text{endorse} = 129$), ‘tried cannabis’ ($r = -0.36, p = 3.40\times10^{-04}, n\text{endorse} = 93$), and ‘had whole alcoholic drink’ ($r = -0.23, p = 0.001, n\text{endorse} = 195$), respectively. For data on frequencies, correlations and details about how the items were created, see Supplementary Table 7.

Indirect effects for specific substances

We found a significant indirect effect of prenatal tobacco smoking on adolescent substance use, via cumulative DNAm risk. Here, we wanted to clarify whether this indirect effect was observed across all substances or only specific ones (e.g. adolescent tobacco use). To this end, we reran the path analysis with the three first-order factors of tobacco, cannabis and alcohol use (see Supplementary Figure 4). Indirect effects were significant across all three substance types (tobacco: $b = 0.31, s.e. = 0.12, p = 0.01$, bootstrapped 95% CI = 0.09-0.59; cannabis: $b = 0.71, s.e. = 0.29, p = 0.01$, bootstrapped 95% CI = 0.22-1.34; alcohol: $b = 0.14, s.e. = 0.06, p = 0.03$, bootstrapped 95% CI = 0.04-0.30). Because the first-order factor of cannabis use contained one outlier (i.e. >3 s.d. from the mean), the analysis was also rerun with winsorized data for this score and results remained consistent.

Biological validation of prenatal maternal smoking

Finally, to ensure the validity of our measure of prenatal smoking – which was derived from a single yes ($n = 48$) /no ($n = 213$) item reported by mothers – we ran an epigenome-wide analysis with prenatal smoking predicting neonatal DNAm. As expected, the top differentially-methylated locus was cg05575921 (**AHRR**: $p = 6.96e^{-16}; q = 2.88E^{-10}$, see Supplementary Table 8), a well-established, sensitive and specific biomarker of tobacco
of note, there was no overlap between the maternal smoking and adolescent substance use DMPs.

Discussion
The aim of this study was to characterize DNA methylation patterns prospectively associated with substance use risk, using longitudinal data spanning gestation to adolescence. We highlight here three key findings: (i) epigenetic variation across 65 loci at birth associated with higher tobacco, cannabis and alcohol use in adolescence, as well as an earlier age of substance use onset; (ii) these effects were specific to the neonatal period and not observed in mid-childhood; and (iii) several of the identified loci were associated with known genetic mQTLs, and all, collectively, mediated the effect of prenatal tobacco smoking on adolescent substance use. These findings lend novel insights into epigenetic predictors of substance use, highlight birth as a potentially sensitive window of biological vulnerability and provide preliminary support for the role of DNAm as an indirect pathway linking prenatal exposures to adolescent behavioral outcomes.

Epigenetic variation at birth associates with substance use in adolescence
While the impact of substance use on DNAm has been repeatedly documented (12), less is known about the extent to which DNAm may confer risk for substance use, as existing studies have typically focused on adults already exposed to substances. To our knowledge, this is the first study to address this gap by examining DNA collected prior to substance use initiation. Furthermore, the use of a latent factor score comprising of tobacco, cannabis and alcohol use enabled us to examine the potential role of methylomic variation in broader substance use liability. Based on genome-wide analyses, we found that epigenetic variation across 65 loci at birth associated with higher substance use 14-18 years later, as well as an earlier age of onset amongst substance users. These loci were annotated to genes that, together, formed a compact underlying genetic network and were enriched for a range of biological pathways, including neural processes (e.g. axonogenesis, synaptic transport) and cellular components (e.g. axon, dendritic spine, post-synaptic membrane). The most differentially methylated locus was annotated to *PACSIN1*, a developmentally-modulated gene that plays an important role glutamate neurotransmission, axonal growth, dendritic branching and synaptic plasticity (45) and that is highly expressed in brain tissue (37, 38), including regions implicated in drug-seeking behavior and addiction risk (e.g. nucleus accumbens; frontal cortex, amygdala and hippocampus; 46). Other key annotated genes also implicated in early brain development included *NEUROD4*, involved in neuronal differentiation, and *NTRK2*, a Trk receptor for multiple neurodevelopmental genes, including Brain-derived Neurotrophic Factor, Neutrophin 4 and Nerve Growth Factor (40).

The neonatal period as a potential window of biological vulnerability
The inclusion of repeated DNAm measures enabled us to test the stability of epigenetic effects during childhood. We found that none of the loci identified at birth continued to predict substance use by age 7 (after multiple correction). This specificity of effects around birth is consistent with previous studies from our group examining longitudinal associations between DNAm and developmental outcomes (28, 47, 48). Findings are also consistent with
a recent study based on the ALSPAC sample that reported low genome-wide continuity in DNA methylation patterns over time (9), especially when comparing birth to other time points. A number of factors may drive the temporal differences observed. First, findings may reflect tissue-specific DNA methylation patterns, as data was extracted from two different blood sources (cord blood at birth vs whole blood at age 7). Second, differences may reflect the specific timing of environmental influences, whereby methylation patterns at birth may be a more reliable proxy for intra-uterine risk exposures and associated perturbations in fetal development (49), compared to age 7. Third, the neonatal period may represent a particularly sensitive window of biological vulnerability to future substance use. For example, epigenetic patterns at birth may trigger downstream developmental consequences resulting in enduring individual differences (e.g. in neural networks underlying drug-seeking behavior and addiction; 12, 15), without the epigenetic signature being maintained over time (18). Given that we still know little about the role of tissue differences, environmental influences and developmental processes on DNA methylation (50), the above explanations are inevitably speculative and will necessitate further investigation.

Genetic and environmental influences on DNA methylation patterns associated with substance use
The identification of neonatal DNA methylation patterns associated with adolescent substance use raises questions about what kind of factors may drive this methylocimic variation in the first place. Evidence suggests that DNA methylation patterns (8, 10) – like substance use liability (5, 6) – reflect the influence of both genetic and environmental factors. Based on data from mQTLdb (9), we found that five of our 65 DMPs were associated with known mQTLs, suggesting that they may be considerably influenced by genetic structure. While these associations point to potentially large genetic effects on a relatively small number of our DMPs, it is important to note that the heritability of DNA methylation patterns is greater than what can currently be explained using known mQTLs (9). As such, genetic effects on our other DMPs cannot be ruled out, especially the presence of polygenic effects. With regards to environmental influences, we found that three prenatal factors were associated with cumulative DNA methylation risk at birth (i.e. comprising all DMPs): maternal tobacco smoking (measured in the first trimester), maternal risks (e.g. psychopathology and criminal behaviour), and contextual risks (e.g. poor housing, financial problems). Associations remained consistent after removing any mQTL-related DMP from our DNA methylation risk score to minimize genetic confounding. These findings support the presence of both genetic and environmental influences on substance-use related DNA methylation patterns. It is important to note, however, that because associations were based on correlational analyses, they should be interpreted with caution and considered more as well-grounded hypotheses for further examination in larger longitudinal studies.

DNA methylation as an indirect pathway linking prenatal smoking to adolescent substance use
We found that one prenatal exposure – maternal tobacco smoking – uniquely associated with substance use over and above other exposures, and that this association was partially mediated by cumulative DNA methylation risk at birth. Importantly, this indirect effect was observed across all three substance types (i.e. not just tobacco use, but also cannabis and alcohol use) – pointing to a potential link between prenatal tobacco exposure and broader substance use liability. To our knowledge, this is the first example in humans of an indirect effect of
prenatal exposures on substance-related outcomes via DNAm, consistent with recent work reported in animals (15). However, due to the correlational nature of the analyses, such evidence should be considered preliminary and in need of rigorous assessment using advanced causal inference methods (e.g. two-step Mendelian randomization; 51, 52). In particular, further work will be needed to trace the specific biological pathways through which this indirect effect may be expressed. For example, experimental studies have shown that prenatal nicotine exposure causes neuromorphological changes (e.g. dendritic branching, axonal growth, spine density) in brain circuits underlying motivation, learning and reward-processing, which in turn confer latent vulnerability for substance use and other externalizing problems (e.g. hyperactivity, conduct problems; 15, 53, 54). As such, it will be interest to test whether the observed effect of prenatal nicotine exposure on substance use may be expressed via epigenetically-modulated changes in neural development, organization and structure. This will also require a more comprehensive investigation of DNAm in the context of other epigenetic processes, which have also been implicated in substance use and addiction (e.g. histone modifications and microRNAs, see 14 for a review).

**Limitations and future directions**

Findings should be interpreted in light of a number of limitations. First, the current study was based on a modestly sized population-based sample of youth. At present, ALSPAC is the only cohort, to our knowledge, that is prospective enough to enable the examination of neonatal DNAm patterns associated with adolescent substance use. Consequently, we were unable to replicate our results in an independent sample. In future, it will be important to test the robustness of findings using other epidemiological cohorts, as well as establishing the relevance of the identified markers in the development of more severe clinical phenotypes, including substance abuse and dependence. Second, findings were based on DNAm from peripheral samples; as such, more research will be needed to establish the relevance of the identified markers to brain function. Future studies incorporating imaging data will be important for establishing whether these markers associate with structural or functional alterations in addiction-relevant neural pathways (e.g. related to reward processing, impulse control, learning and memory), contributing to a more mechanistic understanding of the identified associations. Third, functional characterization of the DMPs was performed using ENCODE data, as we did not have access to RNA. Integration of transcriptomic data will mark an important step toward establishing the downstream effects of the observed DNAm changes on gene expression. Fourth, despite the fact that we identified prospective associations between DNAm and substance use (i.e. DNAm collected before initiation of substance use), it is not possible to establish causality. Finally, the study focused exclusively on DNA methylation, and other epigenetic processes (e.g. histone modifications, microRNAs) are likely to be important in mediating the onset and consequences of addiction (14).
Conclusions
The present findings lend novel insights into early epigenetic correlates of substance use, pinpointing specific markers for future interrogation. Evidence of temporally-specific effects points to birth as a potentially sensitive window of biological vulnerability, which may particularly benefit from intervention efforts. Findings also highlight prenatal smoking as an important prevention target, and contribute to a better understanding of the biological mechanisms through which tobacco exposure during pregnancy may increase risk for future substance use.
Acknowledgements: We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. With regard to the ALSPAC DNA methylation, we thank all involved, particularly the laboratory scientists and bioinformaticians who contributed considerable time and expertise to the data in this paper. This publication is the work of the authors who will serve as guarantors for the contents of this paper. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This work was funded by the National Institute of Child and Human Development grant (R01HD068437). CAMC was supported by the Economic and Social Research Council (ES/N001273/1). EW was supported by the German Research Foundation (Wa 3635/1-1). Open access for this article was funded by King’s College London.

Conflict of Interest: The authors declare no conflict of interest.

Supplementary information is available at Translational Psychiatry’s website.
References
DNA methylation and substance use risk


Figure legends

**Figure 1. Differentially methylated loci at birth associated with adolescent substance use.** A. Manhattan plot showing genome-wide associations between DNA methylation at birth and later substance use (age 14-18). The dotted line represents the FDR-correction threshold (i.e. loci above the line are considered significant). B. Prospective association between the top differentially methylated locus at birth and later substance use. The X axis shows substance use factor scores while the Y axis represents beta methylation values, adjusted for sex and cell-type proportions. C. Gene network analysis using GeneMANIA. Black circles represent genes (n =60) associated with the 65 probes found to be related to adolescent substance use in the genome-wide analysis at birth. Grey circles represent additional genes predicted by GeneMANIA based on genetic and physical interactions, shared protein domains as well as protein co-expression data. The gene network analysis demonstrates that, rather than being isolated, these genes clustered into a complex interconnected network. D. Significantly enriched biological processes (blue), molecular functions (purple) and cellular components (red), based on gene ontology (GO) analysis of 60 genes annotated to probes that predict substance use at birth (n = 65; q < 0.05). Circles represent GO terms that survive FDR correction and contain at least one gene. The X axis represents -log(10) p values. The opacity of the circles indicates level of significance (darker = more significant). The size of the circles indicates the percentage of genes in our results for a given pathway compared to the total number of genes in the same pathway (i.e. larger size = larger %).

**Figure 2. Indirect effect of prenatal smoking on adolescent substance use via neonatal DNA methylation.** A. Path analytic indirect effects model. Dotted arrowed lines indicate non-significant paths. Single arrowed lines indicate standardized path coefficients that survived bootstrap-corrected confidence intervals (i.e. significant path). Red arrows show significant indirect path. Population effect sizes are interpreted using the standardized estimates (Std. B) following Cohen’s guidelines: an effect of .10 is small effect, an effect of .24 is a medium effect, and an effect of .37 is a large effect. B. Graphical representation of the indirect effect, where prenatal smoking associates with higher cumulative DNA methylation risk at birth (top panel), which in turn associates with higher substance use in adolescence (bottom panel).
Table 1. DNA methylation loci at birth that prospectively associate with substance use in adolescence (n = 65, q<0.05)

<table>
<thead>
<tr>
<th>Probe</th>
<th>Gene</th>
<th>Chr</th>
<th>Genomic location</th>
<th>Position</th>
<th>Std B</th>
<th>P-value</th>
<th>FDR (q-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg04941418</td>
<td>PAX5N1</td>
<td>6</td>
<td>5'UTR</td>
<td>34493129</td>
<td>-0.34</td>
<td>1.10E-08</td>
<td>0.005</td>
</tr>
<tr>
<td>cg11074746</td>
<td>Memo1</td>
<td>2</td>
<td>TSS1500</td>
<td>32236343</td>
<td>0.31</td>
<td>1.77E-07</td>
<td>0.02</td>
</tr>
<tr>
<td>cg00335219</td>
<td>--</td>
<td>16</td>
<td>--</td>
<td>86012305</td>
<td>0.27</td>
<td>1.79E-07</td>
<td>0.02</td>
</tr>
<tr>
<td>cg07395930</td>
<td>CLSTN1</td>
<td>1</td>
<td>Body</td>
<td>9791419</td>
<td>-0.31</td>
<td>3.84E-07</td>
<td>0.02</td>
</tr>
<tr>
<td>cg01589998</td>
<td>FOXN4</td>
<td>12</td>
<td>Body</td>
<td>109729478</td>
<td>0.31</td>
<td>3.92E-07</td>
<td>0.02</td>
</tr>
<tr>
<td>cg13978601</td>
<td>PPP2R4</td>
<td>9</td>
<td>Body</td>
<td>131905041</td>
<td>-0.31</td>
<td>4.10E-07</td>
<td>0.02</td>
</tr>
<tr>
<td>cg23361356</td>
<td>SLC9A3</td>
<td>5</td>
<td>Body</td>
<td>508834</td>
<td>-0.31</td>
<td>4.17E-07</td>
<td>0.02</td>
</tr>
<tr>
<td>cg27020216</td>
<td>SGEF</td>
<td>3</td>
<td>Body</td>
<td>153540347</td>
<td>0.29</td>
<td>8.35E-07</td>
<td>0.03</td>
</tr>
<tr>
<td>cg14632140</td>
<td>LMO3</td>
<td>12</td>
<td>5'UTR</td>
<td>16758112</td>
<td>0.29</td>
<td>9.82E-07</td>
<td>0.03</td>
</tr>
<tr>
<td>cg08080985</td>
<td>C9orf95</td>
<td>9</td>
<td>Body</td>
<td>77703079</td>
<td>0.30</td>
<td>1.07E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg20685020</td>
<td>ATP6V0B</td>
<td>1</td>
<td>5'UTR</td>
<td>44440676</td>
<td>0.30</td>
<td>1.34E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg25229198</td>
<td>ADAMT56</td>
<td>5</td>
<td>Body</td>
<td>64660684</td>
<td>-0.30</td>
<td>1.38E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg05968179</td>
<td>USP6NL</td>
<td>10</td>
<td>Body</td>
<td>11505706</td>
<td>-0.29</td>
<td>1.42E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg01009697</td>
<td>NTRK2</td>
<td>9</td>
<td>TSS1500</td>
<td>87283470</td>
<td>0.29</td>
<td>1.56E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg19026817</td>
<td>--</td>
<td>20</td>
<td>--</td>
<td>56782259</td>
<td>-0.29</td>
<td>1.64E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg05116255</td>
<td>ANKRD30A</td>
<td>10</td>
<td>TSS1500</td>
<td>37413782</td>
<td>-0.29</td>
<td>1.85E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg04799664</td>
<td>NLRC5</td>
<td>16</td>
<td>5'UTR</td>
<td>57053580</td>
<td>-0.29</td>
<td>1.85E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg07746699</td>
<td>IFIT40</td>
<td>16</td>
<td>TSS200</td>
<td>1662305</td>
<td>0.30</td>
<td>1.87E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg02290110</td>
<td>SFC2</td>
<td>19</td>
<td>TSS1500</td>
<td>461808</td>
<td>0.29</td>
<td>1.91E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg06951646</td>
<td>TRRAP</td>
<td>7</td>
<td>Body</td>
<td>98586548</td>
<td>-0.29</td>
<td>1.92E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg02052845</td>
<td>HGS</td>
<td>17</td>
<td>Body</td>
<td>79658835</td>
<td>-0.29</td>
<td>2.00E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg05033322</td>
<td>ATM</td>
<td>11</td>
<td>TSS1500</td>
<td>108093245</td>
<td>0.29</td>
<td>2.08E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg02404636</td>
<td>SFI1</td>
<td>22</td>
<td>TSS1500</td>
<td>31891804</td>
<td>0.29</td>
<td>2.10E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg05620865</td>
<td>AP3B1</td>
<td>5</td>
<td>Body</td>
<td>77588408</td>
<td>-0.29</td>
<td>2.22E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg14291256</td>
<td>MMP21</td>
<td>10</td>
<td>Body</td>
<td>127461065</td>
<td>-0.29</td>
<td>2.25E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg12593849</td>
<td>STX12</td>
<td>1</td>
<td>3'UTR</td>
<td>28150710</td>
<td>-0.29</td>
<td>2.31E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg10844884</td>
<td>ZNF22</td>
<td>10</td>
<td>Body</td>
<td>45498904</td>
<td>-0.29</td>
<td>2.32E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg27242945</td>
<td>CAV1</td>
<td>7</td>
<td>1stExon</td>
<td>116165134</td>
<td>0.28</td>
<td>2.35E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg13562386</td>
<td>TAGLN2</td>
<td>1</td>
<td>TSS1500</td>
<td>159895724</td>
<td>0.29</td>
<td>2.36E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg25767859</td>
<td>PKD1L1</td>
<td>7</td>
<td>Body</td>
<td>47859324</td>
<td>-0.29</td>
<td>2.37E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg12202228</td>
<td>FND3C3B</td>
<td>3</td>
<td>Body</td>
<td>171871829</td>
<td>-0.29</td>
<td>2.52E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg20056324</td>
<td>NEUROD4</td>
<td>12</td>
<td>TSS200</td>
<td>55413610</td>
<td>0.29</td>
<td>2.60E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg27122536</td>
<td>FOXF1</td>
<td>16</td>
<td>1stExon</td>
<td>86544658</td>
<td>0.29</td>
<td>2.67E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg14661886</td>
<td>PRRT3</td>
<td>3</td>
<td>TSS200</td>
<td>9994197</td>
<td>0.29</td>
<td>2.75E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg27596068</td>
<td>SERPINH1</td>
<td>11</td>
<td>TSS1500</td>
<td>75272301</td>
<td>-0.28</td>
<td>2.77E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg23088142</td>
<td>--</td>
<td>3</td>
<td>--</td>
<td>137531265</td>
<td>0.29</td>
<td>2.99E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg08625693</td>
<td>DLG3</td>
<td>X</td>
<td>Body</td>
<td>69674126</td>
<td>-0.26</td>
<td>3.09E-06</td>
<td>0.03</td>
</tr>
<tr>
<td>cg05463325</td>
<td>RAB4A</td>
<td>1</td>
<td>TSS1500</td>
<td>229406534</td>
<td>0.29</td>
<td>3.23E-06</td>
<td>0.04</td>
</tr>
<tr>
<td>cg01894322</td>
<td>PPP1R1B</td>
<td>17</td>
<td>Body</td>
<td>37789575</td>
<td>-0.28</td>
<td>3.38E-06</td>
<td>0.04</td>
</tr>
<tr>
<td>cg13138952</td>
<td>RH3AL</td>
<td>17</td>
<td>Body</td>
<td>236013</td>
<td>-0.29</td>
<td>3.39E-06</td>
<td>0.04</td>
</tr>
<tr>
<td>cg01338630</td>
<td>RASD2</td>
<td>22</td>
<td>3'UTR</td>
<td>35948166</td>
<td>-0.27</td>
<td>3.77E-06</td>
<td>0.04</td>
</tr>
<tr>
<td>cg22409100</td>
<td>SLCO1A1</td>
<td>2</td>
<td>TSS1500</td>
<td>40658918</td>
<td>0.14</td>
<td>4.01E-06</td>
<td>0.04</td>
</tr>
<tr>
<td>cg24954864</td>
<td>CLIP1</td>
<td>12</td>
<td>TSS1500</td>
<td>122907641</td>
<td>0.27</td>
<td>4.08E-06</td>
<td>0.04</td>
</tr>
<tr>
<td>cg22467567</td>
<td>IGF2P5</td>
<td>2</td>
<td>5'UTR</td>
<td>217559885</td>
<td>0.28</td>
<td>4.39E-06</td>
<td>0.04</td>
</tr>
<tr>
<td>cg</td>
<td>Gene</td>
<td>Chrom</td>
<td>Type</td>
<td>Position</td>
<td>Beta</td>
<td>p-value</td>
<td>log2FDR</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>-------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>cg25179876</td>
<td>NRP1</td>
<td>10</td>
<td>Body</td>
<td>33483109</td>
<td>-0.19</td>
<td>4.47E-06</td>
<td></td>
</tr>
<tr>
<td>cg04388666</td>
<td>CCDC85C</td>
<td>14</td>
<td>TSS1500</td>
<td>100072073</td>
<td>0.28</td>
<td>4.61E-06</td>
<td></td>
</tr>
<tr>
<td>cg13244417</td>
<td>TMTC4</td>
<td>13</td>
<td>TSS200</td>
<td>101327186</td>
<td>0.28</td>
<td>4.75E-06</td>
<td></td>
</tr>
<tr>
<td>cg16661000</td>
<td>HAPLN3</td>
<td>15</td>
<td>5'UTR</td>
<td>89437710</td>
<td>0.27</td>
<td>4.81E-06</td>
<td></td>
</tr>
<tr>
<td>cg00276455</td>
<td>--</td>
<td>6</td>
<td>--</td>
<td>54904638</td>
<td>-0.28</td>
<td>4.82E-06</td>
<td></td>
</tr>
<tr>
<td>cg08261702</td>
<td>LOC728743</td>
<td>7</td>
<td>Body</td>
<td>150103112</td>
<td>0.25</td>
<td>5.02E-06</td>
<td></td>
</tr>
<tr>
<td>cg05168344</td>
<td>LZTR1</td>
<td>22</td>
<td>Body</td>
<td>21340160</td>
<td>-0.28</td>
<td>5.07E-06</td>
<td></td>
</tr>
<tr>
<td>cg02725795</td>
<td>FAM175A</td>
<td>4</td>
<td>Body</td>
<td>84405871</td>
<td>0.27</td>
<td>5.18E-06</td>
<td></td>
</tr>
<tr>
<td>cg20336014</td>
<td>MGEA5</td>
<td>10</td>
<td>TSS200</td>
<td>103578255</td>
<td>0.27</td>
<td>5.49E-06</td>
<td></td>
</tr>
<tr>
<td>cg26047334</td>
<td>TNS1</td>
<td>2</td>
<td>5'UTR</td>
<td>218785909</td>
<td>0.25</td>
<td>5.56E-06</td>
<td></td>
</tr>
<tr>
<td>cg18769584</td>
<td>LZTFL1</td>
<td>3</td>
<td>3'UTR</td>
<td>45866619</td>
<td>-0.27</td>
<td>5.64E-06</td>
<td></td>
</tr>
<tr>
<td>cg13083436</td>
<td>LUZP2</td>
<td>11</td>
<td>TSS200</td>
<td>24518414</td>
<td>0.28</td>
<td>5.74E-06</td>
<td></td>
</tr>
<tr>
<td>cg23530543</td>
<td>HEPN1</td>
<td>11</td>
<td>TSS1500</td>
<td>124788414</td>
<td>-0.28</td>
<td>5.98E-06</td>
<td></td>
</tr>
<tr>
<td>cg02957771</td>
<td>FBXO31</td>
<td>16</td>
<td>Body</td>
<td>87380349</td>
<td>-0.27</td>
<td>6.13E-06</td>
<td></td>
</tr>
<tr>
<td>cg17396676</td>
<td>EPS15L1</td>
<td>19</td>
<td>TSS1500</td>
<td>16583990</td>
<td>-0.28</td>
<td>6.56E-06</td>
<td></td>
</tr>
<tr>
<td>cg09278687</td>
<td>TBX6</td>
<td>16</td>
<td>Body</td>
<td>30100430</td>
<td>-0.28</td>
<td>6.91E-06</td>
<td></td>
</tr>
<tr>
<td>cg20643362</td>
<td>C19orf12</td>
<td>19</td>
<td>1stExon</td>
<td>30206369</td>
<td>0.27</td>
<td>7.37E-06</td>
<td></td>
</tr>
<tr>
<td>cg10395806</td>
<td>CTNN</td>
<td>11</td>
<td>Body</td>
<td>70280601</td>
<td>-0.28</td>
<td>7.43E-06</td>
<td></td>
</tr>
<tr>
<td>cg00877150</td>
<td>BCOR</td>
<td>X</td>
<td>5'UTR</td>
<td>39972039</td>
<td>-0.27</td>
<td>7.52E-06</td>
<td></td>
</tr>
<tr>
<td>cg20319698</td>
<td>LRRFIP1</td>
<td>2</td>
<td>Body</td>
<td>238644099</td>
<td>-0.27</td>
<td>7.73E-06</td>
<td></td>
</tr>
<tr>
<td>cg14712611</td>
<td>ANK2</td>
<td>4</td>
<td>5'UTR</td>
<td>113739379</td>
<td>0.27</td>
<td>7.74E-06</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Associations between prenatal exposures, cumulative DNAm risk at birth and adolescent substance use

<table>
<thead>
<tr>
<th>Prenatal exposures</th>
<th>Maternal smoking</th>
<th>Maternal alcohol use</th>
<th>Maternal risks</th>
<th>Family risks</th>
<th>Contextual risks</th>
<th>Life Events</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p</td>
<td>r</td>
<td>p</td>
<td>r</td>
<td>p</td>
</tr>
<tr>
<td>Cumulative DNAm risk (birth)</td>
<td>0.20</td>
<td>1.21E-03</td>
<td>-0.07</td>
<td>0.28</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>Substance use (age 14-18)</td>
<td>0.32</td>
<td>1.58E-07</td>
<td>-0.03</td>
<td>0.60</td>
<td>0.32</td>
<td>1.44E-07</td>
</tr>
</tbody>
</table>
Figure 1. Differentially methylated loci at birth associated with adolescent substance use.
Figure 2. Indirect effect of prenatal smoking on adolescent substance use via neonatal DNA methylation.