# *DRD4* methylation as a potential biomarker for physical aggression: An epigenome-wide, cross-tissue investigation

SELF-ARCHIVING VERSION

Charlotte A. M. Cecil[1,5], Esther Walton[2], Jean-Baptiste Pingault[3], Nadine Provencal[4], Irene Pappa[5], Frank Vitaro[6], Sylvana Côté[7], Moshe Szyf[8], Richard E. Tremblay[7], Henning Tiemeier[5,9], Essi Viding[3†] & Eamon J. McCrory[3†]

1. Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK.
2. Medical Research Council Integrative Epidemiology Unit, University of Bristol, UK.
3. Division of Psychology and Language Sciences, University College London, UK.
4. Faculty of Health Sciences, Simon Fraser University, Burnaby and BC Children's Hospital Research Institute, Canada.
5. Department of Child and Adolescent Psychiatry/Psychology, Erasmus Medical Center-Sophia Children's Hospital, Netherlands.
6. Department of Psychoeducation, Université de Montréal, Canada.
7. Departments of Pediatrics and Psychology, Université de Montréal, Canada.
8. Department of Pharmacology & Therapeutics, McGill University, Canada.
9. Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Harvard University, USA.

[†]Co-senior authors

**Correspondence to**: Charlotte Cecil, Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London, SE5 8AF, UK. Tel: +44 (0)207 848 0389, E-mail: charlotte.cecil@kcl.ac.uk.

**Short title**: *DRD4* methylation and physical aggression

**Citation:** Cecil CAM, Walton E, Pingault JB, Provencal N, Pappa I, Vitaro F, Cote S, Szyf M, Tremblay R, Tiemeier H, Viding E & McCrory EJ (In Press). *DRD4* methylation as a potential biomarker for physical aggression: An epigenome-wide, cross-tissue investigation. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics.

**Abstract**

Epigenetic processes that regulate gene expression, such as DNA methylation (DNAm), have been linked to individual differences in physical aggression. Yet, it is currently unclear whether: (i) DNAm patterns in humans associate with physical aggression independently of other co-occurring psychiatric and behavioral symptoms; (ii) whether these patterns are observable across multiple tissues; and (iii) whether they may function as a causal vs non-causal biomarker of physical aggression. Here, we used a multi-sample, cross-tissue design to address these questions. First, we examined genome-wide DNAm patterns (buccal swabs; Illumina 450k) associated with engagement in physical fights in a sample of high-risk youth (*n*=119; age=16-24 years; 53% female). We identified one differentially methylated region in *DRD4*, which survived genome-wide correction, associated with physical aggression above and beyond co-occurring symptomatology (e.g. ADHD, substance use), and showed strong cross-tissue concordance with both blood and brain. Second, we found that DNAm sites within this region were also differentially methylated in an independent sample of young adults, between individuals with a history of chronic-high vs low physical aggression (peripheral T cells; ages 26–28). Finally, we ran a Mendelian randomization analysis using GWAS data from the EAGLE consortium to test for a causal association of *DRD4* methylation with physical aggression. Only one genetic instrument was eligible for the analysis, and results provided no evidence for a causal association. Overall, our findings lend support for peripheral *DRD4* methylation as a potential biomarker of physically aggressive behavior, with no evidence yet of a causal relationship.

**Keywords**: DNA methylation; physical aggression; *DRD4*; externalizing problems; replication; Mendelian randomization

## Introduction

Physical aggression is a highly conserved trait across many species that serves adaptive functions for promoting fitness (Nelson & Trainor, 2007). In humans, however, *chronic* physical aggression (e.g. physically destructive, violent and injurious behavior) is a hallmark of several psychiatric disorders, and accounts for significant morbidity, mortality, and societal cost worldwide (Waters et al., 2004). Physical aggression features as a diagnostic criterion for conduct disorder (CD), intermittent explosive disorder (IED) and antisocial personality disorder (ASPD) based on the DSM-5 (APA, 2013). Furthermore, physically aggressive behavior is overrepresented in individuals affected by other psychiatric disorders, such as autism spectrum disorder, attention deficit hyperactivity disorder (ADHD), substance use disorders and schizophrenia (e.g. Hamshere et al., 2013; Krueger, Markon, Patrick, Benning, & Kramer, 2007; Manchia & Fanos, 2017). Importantly, longitudinal studies indicate that chronic physical aggression in childhood and adolescence predict a wide range of maladaptive outcomes in adulthood (Kim-Cohen et al., 2005). Together, this evidence points to physical aggression as a behavioral risk marker for broader psychopathology (Sanislow et al., 2010).

Research to date suggests that individual differences in physical aggression levels stem from the complex interplay of genetic and environmental influences. Twin and adoption studies have found that around half of the variance in physical aggression is attributable to genetic factors (Tuvblad & Baker, 2011) – with higher heritability estimates observed for physical aggression that is severe, chronic and pervasive across settings (Waltes, Chiocchetti, & Freitag, 2015). Consistent with these findings, molecular genetic studies have reported associations between common genetic polymorphisms and physically aggressive behavior, focusing mainly on candidate genes involved in dopaminergic and serotonergic neurotransmission (esp. *MAOA*, *COMT*, *5HTT*, and *DRD4*; Fernàndez-Castillo & Cormand, 2016; Vitaro et al., 2012) as well as hormonal regulation (e.g. *AR*, *OXTR*, *AVP*; Veroude et al., 2016). While only suggestive associations with individual single nucleotide polymorphisms (SNPs) have been identified by genome-wide studies, *total* SNP heritability has been found to account for substantial variance in aggressive behavior (measured globally), providing further evidence for genetic effects (10-54% across different population-based cohorts; Pappa et al., 2015). In addition to genetic predispositions, epidemiological and clinical studies have identified numerous environmental risk factors for pathological physical aggression, including childhood abuse and neglect, family dysfunction, low socioeconomic status, and violence exposure (Chistiakov & Chekhonin, 2017). Together, these factors are thought to co-act with genetic vulnerabilities to modulate propensity for physical aggression

(Tremblay, 2015). At present, however, little is known about the biological mechanisms through which these genetic and environmental inputs become translated into phenotypic variation.

In recent years, epigenetic processes that regulate gene expression, such as DNA methylation (DNAm), have emerged as a potential mechanism of interest. DNAm modulates transcriptional activity via the addition of a methyl group to DNA base pairs, primarily in the context of cytosine-guanine (CpG) dinucleotides (Jaenisch & Bird, 2003). Studies have shown that DNAm patterns are: (i) under significant genetic control – as evidenced by the discovery of a large number of methylation quantitative trait loci (mQTLs; Gaunt et al., 2016); and (ii) sensitive to environmental signals, including nutritional, chemical, physical and psychosocial exposures occurring in utero and beyond (Szyf & Bick, 2013). In turn, DNAm patterns have been shown to associate with a broad range of psychiatric, physical and behavioral phenotypes, including externalizing spectrum symptoms such as conduct problems, ADHD, and substance use (Barker, Walton, & Cecil, 2018; Klengel & Binder, 2015). Consequently, DNAm shows promise as a mechanism through which genetic and environmental factors interface, shaping trajectories of neurodevelopment, health and behavior across the lifespan (Tremblay & Szyf, 2010).

Thus far, only a handful of studies in humans have examined DNAm patterns associated with physical aggression. These studies were mainly based on data from young adults (DNA extracted from white blood cells at age 26-28yrs), who were recruited as part of the Montreal Longitudinal and Experimental Study (MLES; Boisjoli, Vitaro, Lacourse, Barker, & Tremblay, 2007) and the Quebec Longitudinal Study of Kindergarten Children (QLSKC; Rouquette et al., 2014) and who were classified as following either a chronic vs low trajectory of physical aggression during childhood (ages 6-15yrs; Broidy et al., 2003). Using both targeted gene (Provençal, Suderman, Caramaschi, et al., 2013; Dongsha Wang et al., 2012) and epigenome-wide (Guillemin et al., 2014; Provençal et al., 2014) approaches, these studies have reported group differences in DNAm across a large number of gene promoters. Of interest, several of these genes related to biological processes that have been previously implicated in physical aggression, including monoamine neurotransmission, neuroendocrine function and inflammatory response (Provençal, Booij, & Tremblay, 2015). Besides this sample, one other study (van Dongen et al., 2015) based on a general population sample of over 2000 adult twins examined epigenome-wide associations between DNAm (extracted from whole blood at mean age = 36.4yrs) and a global measure of aggressive behavior (i.e. not specifically related to physical aggression, but also including social and verbal

aggression). None of the DNAm sites were found to associate with aggressive behavior after multiple correction, which the authors suggested may have been due to low rates of more severe forms of aggression (i.e. physical aggression). However, a gene ontology analysis identified significant enrichment for genes involved in a range of central nervous system processes. As would be expected for brain-based phenotypes, epigenetic alterations in genes involved in neural function have also been observed in psychiatric disorders that feature aggression as a diagnostic criterion, such as childhood disruptive behavior disorders (Barker, et al., 2017; Cecil et al., 2017), intermittent explosive disorder (Montalvo-Ortiz, Zhang, Chen, Liu, & Coccaro, 2018) and antisocial personality disorder (Beach, Brody, Todorov, Gunter, & Philibert, 2011; Checknita et al., 2015).

Despite these promising findings, research on DNAm and physical aggression in humans has been limited in a number of important ways. One first limitation relates to *comorbidity*. Individuals with high physical aggression often present with elevated symptomatology across a range of psychiatric and behavioral domains, such as hyperactivity, callous-unemotional traits, emotional problems, and substance use, which have also been associated with altered DNAm patterns (Barker, Walton, & Cecil, 2018). Because these are not typically examined in the extant literature, however, it remains unclear whether the identified DNAm patterns may be unique to physical aggression, shared with – or confounded by – other symptom domains. Relatedly, little is known about the extent to which these epigenetic patterns may reflect environmental risk factors for physical aggression (rather than aggression per se), such as childhood maltreatment and deprivation. Examining these associations is crucial in order to characterize more precisely the epigenetic 'signature' of physically aggressive behavior, and its relationship to other common correlates of – and risk factors for – physical aggression.

A second limitation in the extant literature relates to *tissue specificity*. To date, studies on physical aggression in humans have been based entirely on DNA extracted from blood cells, as it is not possible to measure in vivo DNAm levels directly from the brain. This is problematic, given that DNAm patterns can vary substantially between tissues and cell-types (Smith et al., 2015). As such, it remains unclear to what extent DNAm patterns associated with physical aggression in blood may reflect those in the brain – likely the most relevant organ for the study of aggressive behavior (Bakulski, Halladay, Hu, Mill, & Fallin, 2016). Establishing peripheral-central nervous system (CNS) concordance in DNAm is all the more important given that existing findings in blood point to the involvement of neural-related genes. Furthermore, no study to date has examined associations between physical aggression

and DNAm in buccal cells, another easily accessible peripheral tissue that has recently been found to correlate more strongly with brain tissue than blood (Smith et al., 2015). To this end, cross-tissue designs are necessary in order to gauge the utility of peripheral DNAm as a potential marker of brain-based behavioral phenotypes, such as physical aggression (Bakulski et al., 2016).

Finally, there has been a paucity of investigation regarding the potential *mechanistic role* of peripheral DNAm in physical aggression. In other words, based on associations alone, it is not currently possible to establish whether DNAm acts as causal or non-causal marker of physical aggression. Addressing this question bears important implications for informing research and clinical practice. On the one hand, DNAm marks need not to exert a causal influence in order to show potential utility as *biomarkers* – so long as they are robustly and reliably associated with a phenotype of interest. This is well exemplified in oncology, where non-causal DNAm patterns are already being used in the clinic as biomarkers for cancer detection, prognosis and even treatment response (Ladd-Acosta & Fallin, 2015). In the case of physical aggression, non-causal biomarkers could show utility in a number of different ways. First, non-causal biomarkers that reliably index increased risk for developing pathological aggression may be used, for example, to build more accurate predictive models. In turn, such models may help early detection and prevention via allocation of resources towards improvement of risk trajectories (e.g. preventive interventions to enhance mental health, social communication, emotional and behavioral regulation, etc.). **Second,** non-causal biomarkers that are associated with response to treatment for pathological aggression (e.g. anger management, dialectical behavior therapy, etc), may help to inform decisions about treatment formulation. On the other hand, a causal effect is important in order to shed light on etiological pathways and to identify promising *intervention* targets. Indeed, interventions designed to modify DNAm marks (e.g. via pharmacological treatment) will be unlikely to have an effect on physical aggression unless these DNAm marks lie along the causal pathway (Szyf, 2015). Advanced inference methods, such as Mendelian randomization (MR), represent one important tool for testing causality. MR involves the use of genetic information to minimize confounding common in observational research and strengthen causal inference about the effect of an exposure on an outcome (Davey Smith & Hemani, 2014). While this method has been previously adapted to epigenetic markers (Relton & Davey Smith, 2012; Caramaschi et al., 2017; Richardson et al., 2018) and utilized to examine causality between peripheral biomarkers and a range of psychiatric outcomes (Prins et al., 2016; Pingault, Cecil,

Murray, Munafo, & Viding, 2016;), it has yet to be utilized to clarify the nature of the observed relationship between DNAm and physical aggression.

### The present study

In light of the aforementioned gaps in the literature, the present study had three main aims. First, we used a hypothesis-free, genome-wide approach to examine DNAm patterns (DNA extracted from buccal swabs) associated with physically aggressive behavior, measured as engagement in physical fights over the past year, based on data from a community sample of high-risk youth. Second, we sought to comprehensively characterize the properties of aggression-associated DNAm patterns, including: (i) unique vs shared associations with commonly co-occurring psychiatric and behavioral symptoms; (ii) relation to genetic and environmental factors; (iii) cross-tissue concordance with blood and brain using publicly available resources (Edgar, Jones, Meaney, Turecki, & Kobor, 2017; Hannon, Lunnon, Schalkwyk, & Mill, 2015; Smith et al., 2015); and, (iv) independent replication in adults with a chronic vs low history of physical aggression – the sample upon which most epigenetic studies of aggression to date have been based (Guillemin et al., 2014; Provençal et al., 2014). We considered DNAm loci to be potential biomarkers if they showed associations with physical aggression that were replicable and robust to multiple sensitivity analyses, based on the broad definition of a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathologic processes or biological responses to a therapeutic intervention" (Biomarkers Definition Working Group, 2001). Finally, we explored the causal nature of the identified associations using Mendelian randomization based on GWAS data from the Early Genetics and Lifecourse Epidemiology consortium (EAGLE; Pappa et al., 2015), in order to provide preliminary data as to the potential role of DNAm patterns as a causal vs non-causal biomarker of physical aggression. Given the lack of prior evidence in this area, we did not have a-priori hypotheses about causality in the association between peripheral DNA methylation and physical aggression. As such, our application of MR was guided by the epigenome-wide association analysis and should be considered as hypothesis-generating (Evans & Davey Smith, 2015).

## Materials and Methods

### Participants

The current sample was recruited as part of a larger cross-sectional study examining the influence of early adversity on youth outcomes ($n$ = 204, age range = 16-24 years). Youth

from deprived areas of London were recruited through multiple channels including inner-city colleges, internet websites, and a charity providing services to vulnerable youth, typically via self-referral. The study was carried out in accordance with the latest version of the Declaration of Helsinki. All procedures were approved by the UCL Research Ethics Committee (*ID No*: 2462/001) and participants provided informed consent prior to participation, after the nature of the procedures had been fully explained. Full details of the sample and study procedures are available elsewhere (Cecil et al., 2016; Cecil, Viding, Barker, Guiney, & McCrory, 2014)

The present analyses only included participants with complete data on DNAm and self-report engagement in physical fights ($n$ = 119). This subsample was 53% female, ethnically diverse (49% White, 33% Black, 12% Other, 6% Asian) and at high-risk of adversity, with 68% of youth reporting experiences of childhood maltreatment. This subsample was comparable to the full sample in terms of sociodemographic characteristics and exposure to adversity. In addition to self-report measures, informant reports were available for 80% of the subsample ($n$ = 95), whereby an external informant (*schools*: teacher; *charity*: key worker) who knew each participant well completed a questionnaire booklet about the young person's emotional and behavioral function.

**Measures**

**DNA methylation**. DNA was extracted from buccal swabs (representing a mixture of buccal epithelial cells and leukocytes) using procedures described in Freeman et al. (2003). 500ng of high molecular weight DNA was subjected to sodium bisulfite conversion using the EZ-DNA methylation kit (Zymo Research, Orange, CA, USA) using the manufacturers standard protocol. DNAm was quantified using the Illumina HumanMethylation450 BeadChip (Illumina, USA) with arrays scanned using an Illumina iScan (software version 3.3.28). Samples were run in a single batch (i.e. bisulfite treatment and processing run at the same time across two plates side by side). To account for potential chip and position effects, we randomized sample chip allocation and placement on the chip. Furthermore, surrogate variable analysis was applied as a sensitivity step to minimize the influence of unmeasured sources of confounding (see results section). Initial data quality control was conducted using GenomeStudio (version 2011.1) to determine the status of staining, extension, hybridization, target removal, bisulfite conversion, specificity, non-polymorphic and negative controls. Probes that survived this stage were checked for concordance between their reported and predicted sex and then quantile normalised using the *dasen* function within the *waterRmelon*

package (waterRmelon_1.0.3; Pidsley et al., 2013) in R. Probes were removed if they were cross-reactive, polymorphic, used for sample identification on the array, had a SNP at the single base extension with a minor allele frequency larger than 5% (i.e. common polymorphisms; Chen et al., 2013; Price et al., 2013), leaving a total of 413,510 probes in our sample. DNAm levels are indexed by beta values (ratio of methylated signal divided by the sum of the methylated and unmethylated signal, M/M+U).

**Engagement in physical fights**. Engagement in physical fights was assessed via a self-reported item adapted from the Youth Risk Behavior Survey (YRBS; Eaton et al., 2008). The item specifically asked about the frequency of being in physical fights over the past year ("*During the past 12 months, how many times were you in a physical fight?*"). The item was rated on a 4-point scale, with 0 = '*0 times'*, 1 = '*1 time'*, 2 = '*2-5 times'*, and 3 = *'5+ times*'.

**Environmental risks**. *Childhood maltreatment* was measured using the total score of the 28-item, self-report Childhood Trauma Questionnaire (CTQ; Bernstein & Fink, 1998; α = .84), which assesses experiences of abuse and neglect "while growing up". *Community violence exposure* (CVE) over the past year was measured using the self-report Children's Report of Exposure to Violence (CREV; Cooley, Turner, & Beidel, 1995). The CREV records frequency of exposure to different forms of violence, including fights, robberies, shootings, and killings. A total CVE score was obtained by summing the three subscales of hearing about, witnessing and directly experiencing community violence (α = .79–.89). *Neighborhood Deprivation* was measured using the postcode-derived Index of Multiple Deprivation (IMD; Noble, Wright, Smith, & Dibben, 2006) score, an aggregate measure based on population census data that encompasses multiple indicators of neighborhood deprivation, spanning: (i) income; (ii) employment; (iii) health and disability; (iv) education skills and training; (v) barriers to housing and services; (vi) crime; and, (vii) living environment. Higher values indicate more severe childhood maltreatment, CVE, and neighborhood deprivation.

**Psychiatric and behavioral correlates of physical aggression**. We assessed multiple domains of individual functioning that have previously been associated with physical aggression. Self-reported substance use was measured via the Alcohol Use Disorders Identification Test (AUDIT; Saunders, Aasland, Babor, de la Fuente, & Grant, 1993) and the Drug Use Disorders Identification Test (DUDIT; Berman, Bergman, Palmstierna, & Schlyter, 2003). The AUDIT and DUDIT include 10 (α = .82) and 11 (α = .90) items respectively, measuring substance use, harmful use and symptoms of dependence. External informants (i.e. teachers/key workers) completed the DSM-based Adolescent Symptom Inventory (ASI-4; Gadow & Sprafkin, 1998) to assess symptoms of emotional and behavioral disorders,

including generalized anxiety disorder (GAD), major depressive disorder (MDD), attention-deficit hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder (CD), and antisocial personality disorder (ASPD). Each scale contained between 7 and 9 items ($\alpha = .85 – .96$). In addition to examining each symptom domain separately, the GAD and MDD subscales were combined into an overall score of *internalizing problems* ($\alpha = .93$), whereas the ODD, CD, ASPD and ADHD subscales were combined into an overall score of *externalizing problems* ($\alpha = .97$; Cecil et al., 2014). External informants also rated *prosocial behavior* via the Strength and Difficulties Questionnaire prosocial behavior subscale (SDQ; Goodman, 1997; $\alpha = .76$) as well as *callous-unemotional (CU) traits* via the Inventory of Callous-Unemotional Traits (ICU; Frick, 2004; $\alpha = .79$).

**Data Analysis**

All analyses were performed in R (version 3.0.1), controlling for sex, age and self-reported ethnicity (coded as a dummy variable) to account for potential confounding effects (Liang & Cookson, 2014). The analysis proceeded in three steps:

**Step 1: Are DNAm patterns in buccal cells associated with engagement in physical fights in a sample of high-risk youth?** We addressed this question by using both probe- and region-level analyses. In probe-level analyses, we ran an epigenome-wide association analysis (EWAS) using the *IMA* package (Dan Wang et al., 2012) to identify differentially methylated probes (DMPs). DMPs were considered to be genome-wide significant if they passed a False Discovery Rate (FDR) correction of $q \leq .05$. To identify enriched biological pathways, DMPs with a $p \leq .001$ were analyzed further using an optimized gene ontology (GO) method that corrects for multiple potential confounds, including background probe distribution and gene size (for a full description, please refer to the **Supplementary Methods**). As a complement to the probe-level analysis, we also used two independent regional approaches with default settings – Comb-p (Pedersen, Schwartz, Yang, & Kechris, 2012) and DMRcate (Peters et al., 2015) – to locate wider differentially methylated regions (DMRs). DMRs were only considered significant if they survived multiple correction across both methods. Relevant DNAm markers were then characterized further by: (i) testing associations with environmental risks and domains of individual functioning known to correlate with physical aggression; and (ii) investigating cross-tissue concordance with both blood and brain, based on independent published data (Hannon et al., 2015; Smith et al., 2015). Further details regarding these cross-tissue resources are provided in the results section below.

**Step 2: Are these markers differentially methylated in blood samples drawn from young adults with a history of chronic vs low physical aggression?** Next, we examined whether the DNAm markers that met genome-wide significance in Step 1 were differentially methylated in peripheral leukocytes drawn from an independent sample of 38 young adults (Guillemin et al., 2014). Participants were identified as either following a chronic (CPA; *n*= 13; 38% female) or low (LPA; *n*= 25; 54% female) trajectory of physical aggression during childhood, based on prospective, longitudinal data (Broidy et al., 2003). DNAm from peripheral T cells was extracted at ages 26-28 years and quantified using the Illumina 450k array. Due to the low volume of available DNA for Illumina 450k quantification, the 38 samples were originally pooled into 12 samples. Each sample comprised the same amount of DNA from 2 to 4 individuals of the same sex creating 6 samples for the CPA group (3 pools for men CPA and 3 pools for women CPA) and 6 samples for the LPA group (3 pools for men LPA and 3 pools for women LPA). For full details about the sample as well as the processing of the DNAm data, see (Guillemin et al., 2014). All participants were of Caucasian ethnicity and both groups were matched for sex; consequently, ethnicity and sex were not included as covariates. In the analyses, group differences in DNAm were tested using linear regression models, correcting for age. All regressions were bootstrapped 10,000 times (*boot* package), and bias-corrected 95% confidence intervals (CI) were derived using the adjusted bootstrap percentile method (BCa; *boot.ci* function). Rather than evaluating replication based on FDR-corrected p-values – which can be unreliable in small samples due to bias in the estimates and standard errors – we used multiple other indicators that are less sensitive to sample size, namely the direction and magnitude of effects as well as the bias-corrected bootstrapped 95% confidence intervals. Bootstrapping is a nonparametric resampling simulation method that enables to address some of the issues of asymptotic inference by performing bias-adjustment and controlling for the proportion of Type I errors based on the empirical distribution of the data (Pace, 2012). Associations with bootstrapped confidence intervals that did not cross zero were considered significant.

**Step 3: Do these associations potentially reflect causal pathways?** As a final step, we tested the potential causal effect of the identified DNAm markers on aggressive behavior using Mendelian randomization. MR is a technique that uses genetic variants as robust instrumental variables, in order to strengthen causal inferences about the effect of an exposure on an outcome while minimizing potential confounding (Pingault et al., 2016). Full details of the MR analysis are provided in the **Supplementary Methods**. Briefly, the MR analysis involved (i) selecting genetic instruments (i.e. SNPs) for the identified DNAm markers using

the mQTLdb database (http://www.mqtldb.org; Gaunt et al., 2016); (ii) extracting summary statistics about their association with aggressive behavior, based on recently published GWAS data from the EAGLE consortium (Pappa et al., 2015); and (iii) running the MR analysis based on the results from the mQTLdb search (i.e. SNP → DNAm) and the GWAS data (i.e. SNP → physical aggression) using the TwoSampleMR package in R, available as part of the MR-Base platform (www.mrbase.org; Hemani et al., 2016).

## Results

Response frequencies for engagement in physical fights over the past year were as follows: $N_{0\ times} = 88$, $N_{1\ time} = 16$, $N_{2\text{-}5\ times} = 11$, $N_{5+\ times} = 4$. This variable did not correlate with age (Spearman's rho; $r = .03$, $p = .67$) or differ by sex ($F(1) = .64$, $p = .42$). However, it was associated with self-reported ethnicity ($F(3) = 4.62$, $p = .004$), with Black ethnicity associating with a higher number of fights.

### Step 1: Epigenome-wide associations between buccal cell DNAm and engagement in physical fights

**Probe-level analysis.** The results of the probe-level EWAS are depicted in **Table 1**. Overall, inspection of the Q-Q plot indicated little evidence of inflation **(Figure S1)**, which was confirmed using the Bayesian method in the *Bacon* R package (van Iterson et al., 2017; **Figure S2**). Four DMPs were found to meet a genome-wide threshold of significance ($q \leq .05$; bolded in **Table 1**; see **Figure S1** for the Manhattan plot and scatterplots for each DMP): (i) cg21217577, located in the body of the *TRH* gene, a key member of the hypothalamic–pituitary–thyroid axis involved in hormonal regulation (Yarbrough, Kamath, Winokur, & Prange, 2007); (ii) cg11107262 in the body of *GAK*, a gene involved in cellular function (Shimizu, Nagamori, Yabuta, & Nojima, 2009); (iii) cg27349081, annotated to the promoter region of *RIC3*, a molecular chaperone modulating the expression of nicotinic acetylcholine and serotoninergic 5-HT3 receptors (Castillo et al., 2005); and (iv) cg27331554 located in the body of *CYFIP2*, a component of the WAVE1 complex implicated in immune function and BDNF-NTRK2 signaling (Xu, Fu, Zhu, & Liu, 2016). Also listed in **Table 1** are suggestive DMPs that did not meet genome-wide correction ($q > .05$). Of interest, five of these were located in the gene body of the dopamine receptor gene *DRD4*, one of the most extensively studied candidate genes for physical aggression and related externalizing problems (Fernàndez-Castillo & Cormand, 2016). Furthermore, a number of additional DMPs were annotated to neurotransmitter and neuroendocrine genes, including other dopaminergic

(*COMT)*, serotonergic (*GPM6B*), glucocorticoid (*SGK1*) and thyroid hormone (*TXNRD2, TNX2, FDX1*) pathway genes. More broadly, our GO analysis indicated enrichment for a range of biological processes, including axon guidance ($p$ = 1.66E-09), hormone metabolic processes ($p$ = 1.13E-08), behavioral regulation ($p$ = 8,86E-07), dopamine receptor signaling ($p$ = 1.51E-06), cytokine secretion ($p$ = 6.73E-06), and drug response ($p$ = 1.17E-04; **Table S1**). As a control, we reran the GO analysis with a comparable number of randomly selected CpG sites and found only a chance-level overlap in identified terms (<5%).

**Region-level analysis.** The Comb-p analysis identified two genome-wide significant regions (DMRs), both of which were annotated to the *DRD4* gene. Consistent with the probe-level analysis above, these two regions were adjacent to one another (spanning a total of 603bp) and encompassed the five DMPs listed in **Table 1** – all of which were hypo-methylated in relation to physical aggression, following a dose-response gradient (see **Figure S3** for scatterplots). The first region included probes cg03909863, cg01616529, and cg05717871 (chromosomal coordinates: start = 638404, end = 638507; Sidak-corrected regional $p$ = 1.09E-07), whereas the second region included probes cg11335335 and cg07212818 (chromosomal coordinates: start = 637885, end = 638076; Sidak-corrected regional $p$ = 1.62E-04). Overall, these five probes showed moderate-to-high intercorrelations ($r$ = .47 – .79), with 74% of shared variance between them explained by a single principal component (eigenvalue = 3.72, *princomp* function; see **Figure S4**). Importantly, results from Comb-p were confirmed using the DMRcate package, in which the same five probes formed the only DMR to survive genome-wide correction (chromosomal coordinates: start = 637885, end = 638507; q = 6.61E-15, Stouffer-corrected $p$ = .01). Consequently, these five probes were examined as a single region in subsequent analyses.

********** **Table 1************

**Follow-up analyses: characterizing *DRD4* methylation.** In light of the above findings, we carried out a number of follow-up analyses to better characterize patterns of *DRD4* methylation in this sample.

***1. Associations between the DRD4 DMR and correlates of physical aggression.*** First, we examined associations with known environmental and phenotypic correlates of physical aggression. **Table 2** shows Spearman correlations for (i) the engagement in fights variable, and (ii) the five CpG probes contained in the DMR, as well as the principal component capturing shared variance between them. As expected, engagement in physical fights was significantly correlated with all variables, associating with greater exposure to adversity, more

severe psychiatric symptoms (especially externalizing problems), higher substance use and CU traits as well as lower prosocial behavior. All probes contained in the DMR were negatively associated with engagement in physical fights (i.e. individuals with lower DNAm in these probes engaged in a higher number of fights). Associations with environmental risks were generally weak, with lower DNAm levels associating significantly with greater community violence exposure and neighborhood deprivation, but not childhood maltreatment. With regards to co-occurring outcomes, lower DNAm across sites associated most strongly with higher self-reported drug use, informant-rated CU traits and low prosocial behavior, as well as externalizing (especially CD and ASPD) – but not internalizing – psychiatric symptoms. DNAm levels in this region remained significantly associated with engagement in physical fights even after adjusting for all domains of individual functioning (spearman correlation with DMR principal component using the *ppcor* package: (i) bivariate: $r$ = -.39, $p$ = 1.10E-05; (ii) partial: $r$ = -.28, $p$ = 0.01; partial correlations for individual CpGs are also shown in **Table 2**). While this resulted in a considerable drop in effect size, the difference between the adjusted and non-adjusted correlation was not significant (Z = 1.06, $p$ = .14).
********** **Table 2**********

**2.** ***Wider analysis of methylomic variation across the DRD4 gene***. Next, we examined whether any additional CpG sites annotated to this gene associated with engagement in physical fights at a nominal level. As can be seen in **Table 3A**, lower DNAm across 12 (i.e. 57%) out of the 21 probes annotated to *DRD4* were nominally associated with higher engagement in physical fights ($p$ < .05, bolded in table). Interestingly, these tended to show a high degree of variability, with most probes exceeding an $SD$ > .10 (**Table S2**). To put this in context, only 9% of probes on the Illumina 450k array in this sample meet this threshold of variability. The results are visualized as a coMET plot (Martin, Yet, Tsai, & Bell, 2015) in **Figure 1**. Based on ENCODE data extracted from the UCSC GenomeBrowser (GRCh37/hg19 assembly; Kent et al., 2002), we found that all *DRD4* sites overlapped with at least one key regulatory element, suggesting that they are located in genomic regions that are likely to play a role in transcriptional activity (**Table S2**). First, all sites overlapped with histone marks (only relevant cell lines examined, namely blood [GM12878] and umbilical vein endothelial [HUVEC] cells). Histone marks are chemical modifications that influence how tightly packaged the DNA is around a histone protein, regulating its accessibility. The strongest signal was found for H3K27me3 and H3K4me1, both of which are typically associated with active or potentially active enhancer regions (Zhu, et al., 2013). Second, we

found that 90% (*n* = 19) of sites coincided with transcription factor binding sites (data generated from 91 cell-types) - DNA regions where one or more specific proteins responsible for regulating transcription bind to. This overlap was identified across multiple tissues spanning all three germ layers (i.e. endoderm, mesoderm and ectoderm), with the strongest signal linked to the EZH2 transcription factor, a functional enzymatic component of the Polycomb complex PRC2 that is involved in transcriptional repression and early development (Margueron & Reinberg, 2011). Finally, 71% (*n* = 15) of sites were located within DNAse I hypersensitive clusters (data generated from 125 cell-types), which tend to indicate an open chromatin state that facilitates transcription. As with the transcription factor binding sites, we found that overlap with DNAse I hypersensitivity clusters was observed across multiple tissues and germ layers, with the strongest signal identified for epithelial tissue.

********** **Figure 1**********

********** **Table 3**********

***3. Sensitivity analyses.*** We carried out a set of sensitivity analyses to test the robustness of associations between *DRD4* methylation and our measure of physical aggression. First, we winsorized DNAm data to reduce the influence of potential outliers (> 3 *SD*) using the *DescTools* package with default settings. The winsorizing method uses censoring rather than exclusion, which is preferable with small sample sizes (Sheskin, 2003). Second, we ran a surrogate variable analysis (sva package in R; Leek, Johnson, Parker, Jaffe, & Storey, 2012) to identify unwanted sources of variation in the genome-wide DNA methylation data. No significant surrogate variables were detected after correction for the covariates already included in the statistical model. To ensure that the observed associations were not biased by unknown confounders of smaller effect, we re-ran associations between the *DRD4* sites and physical aggression additionally controlling for ten surrogate variables. Third, because buccal swabs can contain a mixture of different cells (mainly buccal epithelial cells and leukocytes) we tested the potential influence of cell-type heterogeneity on our findings, using two independent methods. The first derives a 'predicted tissue' based on DNA methylation patterns (Horvath, 2013), whereas the second uses two CpG sites to estimate the percentage of buccal epithelial cells for each individual (Eipel et al., 2016). Each of these estimates was then examined in relation to *DRD4* methylation and physical aggression (see the **Supplementary Methods** for full details and results). Finally, given the skewed distribution of the variable measuring engagement in fights, we tested whether *DRD4* sites

remain differentially methylated when collapsing the ordinal variable into a binary yes/no variable (i.e. 0 for '*no fights'* [*n* = 88] *vs* 1 for '*any fights'* [*n* = 31] over the past year).

While slight variations in effect sizes were observed, results remained consistent across all of the above sensitivity analyses, indicating that the associations are unlikely to be unduly influenced by unwanted sources of variation or extreme scores in either DNAm or the engagement in fights variable (Spearman correlation of effect sizes: (i) original vs winsorised analysis: $r$ = .92, $p$ = 5.16E-09; (ii) original vs SVA-corrected analysis: $r$ = .96, $p$ = 1.94E-11; (iii) original vs cell-type restricted analysis: $r$ = .91, $p$ = 1.03E-08; and (iv) original vs binary fights analysis: $r$ = .95, $p$ = 6.71E-11; see **Table S3** for full results).

***4. Saliva-blood concordance.*** Before proceeding to the replication step, we wanted to ensure that we could validly compare associations between *DRD4* methylation and physical aggression across different peripheral tissues. This is because in our discovery sample, DNAm was drawn from buccal swabs, whereas in the replication sample it was extracted from blood (T cells). As we did not have access to multiple tissues in our discovery sample, we examined instead published data characterizing saliva-blood concordance for all Illumina 450k probes, based on matched samples from 64 adults (Smith et al., 2015). Of note, the Smith et al. study examined saliva, not buccal swabs. However, these two DNA sources typically contain the same mixture of cell-types (i.e. buccal epithelial cells and leukocytes), although proportions of buccal epithelial cells tend to be higher in buccal swabs (Theda et al., 2018). Furthermore, our sensitivity analyses for cell-type heterogeneity further supported the comparability of *DRD4* DNAm patterns between buccal swabs and saliva (see **Supplementary Methods**). Based on the Smith et al resource, we found that, for 10 of the 12 *DRD4* sites associated with physical fights in our sample, DNAm levels in saliva significantly associated with DNAm levels in blood – including all sites contained in our DMR region ($p$ range: 7.00E-09 – 1.24E-14; **Table S2**). This contrasted with 33% of the remaining sites that did not associate with physical fights. More generally, across all *DRD4* probes, the strength of association with physical fights in our sample correlated positively with the strength of saliva-blood concordance ($r$ = 0.59, $p$ = 0.01; **Table S4**).

**Step 2. Replication in peripheral T cells in an independent sample**

Given the above evidence for cross-tissue concordance, we proceeded to replicate our findings on *DRD4* methylation in a different peripheral tissue (blood T cells) extracted from an independent sample of young adults following a low (LPA) vs chronic (CPA) trajectory of physical aggression from childhood to adolescence. Methylation values for one participant in

the LPA group were winsorized due to presence of outliers (>3 SD), even though findings were consistent when using the original values. Descriptive statistics are presented in **Table 3B**, including means and standard deviations for each *DRD4* site, percent methylation differences and associated effect sizes (Hedge's *g*, given the small sample size). To facilitate comparability with the replication sample – which used a case-control design – we include the same descriptive information for our discovery sample, split into two groups based on the binary fights variable (i.e. '*No fights'* vs '*Any fights'*).

The direction of associations was consistent across both samples, with *lower* DNAm associating with *higher* physical aggression. On average (i.e. across all 21 probes annotated to *DRD4*), there was a 3.85% difference in *DRD4* methylation between groups in our sample (range = 0.01-10.7%), compared to 3.06% in the replication sample (range = 0.20-8.14%). Effect sizes were typically modest in the discovery sample (mean $g$ = -0.41) and large in the replication sample (mean $g$ = -0.94). **Table 3B** also shows the results of the bootstrapped regression analyses in the replication sample. Overall, 8 of the 21 (38%) CpG sites in *DRD4* were significantly differentially methylated between groups, as evidenced by the bootstrapped CIs (bolded in table). Two of these were located in our genome-wide DMR (cg03909863, cg11335335), and three others were nominally significant in the discovery sample (cg06299284, cg06825142, cg23501406).

As a more stringent test of replication, we created a poly-epigenetic score (ePGS) of *DRD4* methylation, following a method commonly used in molecular genetic studies and recently extended to epigenetic data (Shah et al., 2015). Specifically, for each CpG site, we multiplied the methylation values from the replication sample by their respective standardized regression betas from the discovery sample (i.e. independent weights), and then summed these weighted DNAm values together into a single, cumulative risk score. This ePGS score was found to significantly associate with physical aggression in a bootstrapped regression model, showing lower methylation in the CPA vs LPA group (effect size $g$ = -0.90, $T$ = -1.73, SE = 0.84, bootstrapped 95% CI = -3.44 – -0.15).

**Follow-up analysis: relevance to the brain**

***1. Blood-CNS concordance in DRD4 methylation*.** The results from the replication analysis suggest that DNAm patterns in *DRD4* may reliably associate with physical aggression across distinct peripheral tissues (buccal vs blood T cells). Yet, the extent to which these markers may reflect DNAm patterns in the brain is unclear. To address this, we investigated peripheral-CNS concordance in *DRD4* methylation, using published Illumina

450k data from two independent sources. The first consisted of data from over 70 individuals for whom matched DNA samples were isolated from both premortem whole blood and postmortem CNS tissue across four brain regions (prefrontal cortex [PFC], entorhinal cortex [EC], superior temporal gyrus [STG] and cerebellum [CER]; Hannon et al., 2015; http://epigenetics.iop.kcl.ac.uk/bloodbrain). For descriptive purposes, we found that of the 12 *DRD4* sites nominally associated with physical fights in our sample, 10 (83%) showed significant blood-CNS correspondence with at least one brain region, including all five sites contained in our genome-wide DMR (for a table of associations with each brain region, see **Table S2**; for a visual example of associations with the top *DRD4* probe cg07212818, see **Figure S5**). In contrast, blood-CNS correspondence with at least one region was observed for only 1 of the 9 (11%) *DRD4* sites that were not associated with physical fights. In line with the findings on saliva-blood correspondence presented above, we found that, overall, the strength of associations with physical fights in our sample correlated significantly with the strength of blood-CNS associations in Hannon et al.'s sample. This was true across all four CNS tissues, with stronger associations observed for cortical (PFC: $r$ = .69, $p$ = 1.00E-03; SGT: $r$ = .77, $p$ = 5.40E-05; EC: $r$ = .77, $p$ = 4.90E-05) vs non-cortical tissue (cerebellum: $r$ = .49, $p$ = .02; **Table S4**). The difference in these correlations was significant when comparing the cerebellum to the SGT and EC (z-score = 2.67, $p$ = .01), but not the PFC (z-score = 1.70, $p$ = .09). Of note, one of the *DRD4* sites that associated with physical aggression across both the discovery and replication sample (cg06299284) was reported in Hannon et al's study to be one of only 5.39% of probes on the Illumina450k array where inter-individual variation predicted more of the variance in DNAm than tissue type.

Blood-brain concordance was further confirmed using the second data source, consisting of paired samples of whole blood and three brain regions from 16 individuals (Broadmann area [BA] 7: parietal cortex; BA10: anterior prefrontal cortex; and BA20: inferior temporal cortex; Edgar et al., 2017; BECon online tool: https://redgar598.shinyapps.io/BECon/). Of note, four of the five probes contained in our DMR were reported by Edgar et al. to be in the top 10% of sites showing the strongest blood-brain concordance in their sample (**Table S2**, which also includes information on all other *DRD4* probes). Overall, *DRD4* probes showed high variability and strong cross-tissue concordance in this sample. Regions that were most comparable across resources (e.g. Hannon's prefrontal cortex and Edgar's BA10) also showed the strongest convergence in findings, which adds validity to the results (**Table S4**). As with the other resources, the strength of associations with physical fights in our sample correlated with that of blood-brain

correspondence in Edgar et al.'s sample ($r$ = .55, $p$ = .01). A graphical representation of BECon results showing blood-brain variability and correspondence for *DRD4* sites contained in our DMR is provided in **Figure S5**.

***2. DRD4 expression in the brain*.** Given the above support for cross-tissue concordance in *DRD4* methylation across saliva, blood and brain, we investigated expression levels of *DRD4* in CNS vs peripheral tissues using the Genotype-Tissue Expression project portal (GTEx; http://www.gtexportal.org/home/; GTEx Consortium, 2015), which is based on data on 53 tissues from 544 donors. Expression levels across available brain regions as well as the three most relevant peripheral tissues to our study are graphically presented in **Figure S5**, including: (i) salivary gland, (ii) esophagus mucosa, which comprises of squamous epithelial cells that are part of the same cell-type family as buccal epithelial cells, and (iii) whole blood. Generally, *DRD4* expression was low across all tissues, with the cerebellum showing the highest levels of expression in the brain. Other brain regions of relevance to physical aggression, including the frontal cortex, amygdala and hippocampus showed similar levels of expression to the two peripheral tissues closest to buccal cells (i.e. salivary gland and esophagus mucosa), whereas whole blood showed the lowest expression levels.

**Step 3. Mendelian randomization analysis to test for causal pathways**

As a final step, we investigated whether peripheral *DRD4* methylation exerts a causal effect on physical aggression using Mendelian randomization (MR) analysis. Of the 12 *DRD4* sites nominally associated with physical aggression in our discovery sample, 6 (50%) were linked with at least one mQTL, suggesting that DNAm levels across these CpG sites are likely to be under considerable genetic control (**Table S2**). One of these mQTLs (*cis* SNP rs2740373) was eligible to be used as a genetic instrument in the MR analysis, as it was the only polymorphism which was investigated for an association with physical aggression, based on data from a recently published GWAS by the EAGLE Consortium (Pappa et al., 2015; n = 18,832, $p$ = 0.95). Of note, this mQTL was predictive of DNAm levels across three CpG sites, all of which were contained in the DMR (cg07212818, cg11335335, cg01616529; **Table S2**). Consequently, we ran three MR analyses with the same genetic instrument – one for each of these CpG sites. Across all CpG sites, MR results were non-significant, meaning that a causal effect of peripheral DNAm on engagement in physical fights was not supported, even though the direction of effects was consistent with the findings from our discovery and replication analyses (i.e. negative association between DNAm and physical aggression; for full results see **Figure S6**).

## Discussion

The overall aim of the present study was to characterize DNA methylation patterns associated with physical aggression, using a multi-sample, cross-tissue design. We highlight here three key findings. First, we conducted an epigenome-wide analysis of physical aggression – measured as engagement in physical fights – based on buccal cell DNA drawn from a sample of high-risk youth. We found that lower DNAm levels in one region spanning the *DRD4* gene survived genome-wide correction, was associated with higher physical aggression over and above co-occurring symptomatology, and showed strong cross-tissue concordance with both blood and brain tissue. Second, we replicated these findings in an independent sample of young adults based on DNA from a different peripheral tissue (blood T cells). Specifically, we found that *DRD4* methylation levels significantly differentiated between adults with a chronic-high vs low history of physical aggression in childhood and adolescence. Third, using published GWAS data from the EAGLE consortium, we ran a Mendelian randomization analysis to test for a causal effect of peripheral *DRD4* methylation on aggression. Based on the only eligible genetic instrument, we did not find evidence to support causal effects. Overall, our findings point to peripheral *DRD4* methylation as a potential biomarker of physical aggression, with no evidence yet supporting a causal association.

### The epigenetic signature of physical aggression in buccal cells

Recent data suggests that DNAm in buccal cells may be a useful tool for the investigation of brain-based phenotypes, as they are derived from the same ectodermal layer as brain cells during development, and have been shown to correlate with CNS methylation more strongly than other commonly examined peripheral tissues, such as blood (Smith et al., 2015). So far, published epigenetic studies on physical aggression in humans have all made use of blood samples (Guillemin et al., 2014; Provencal, Suderman, Caramaschi, et al., 2013; Provençal et al., 2014; Dongsha Wang et al., 2012). Here, we performed the first genome-wide study characterizing DNAm patterns associated with physical aggression in buccal cells, using both probe- and region-level approaches. Of note, regional analyses are a useful complement to probe-level analyses, as they draw on the intercorrelated nature of DNAm data to alleviate multiple-testing burden, increase power to detect effects, and reduce the likelihood of false positives (Robinson et al., 2014).

The key finding to emerge from our genome-wide analyses implicated epigenetic regulation of the Dopamine Receptor D4 (*DRD4*) gene – one of the most extensively studied candidate genes for aggression and related externalizing problems (Fernàndez-Castillo &

Cormand, 2016). Specifically, we identified one region in the gene body of *DRD4* that associated with engagement in physical fights after genome-wide correction (using two separate methods). More broadly, over half of all sites annotated to *DRD4* showed at least nominal associations, whereby individuals who showed *lower* DNAm levels in these sites reported *more frequent* engagement in physical fights. These associations were robust to sensitivity tests and not unduly driven by extreme scores in either the DNAm data or the physical fights outcome variable. Furthermore, gene ontology analysis indicated significant enrichment for '*dopamine receptor signaling*', supporting the probe- and region-level results. Together, these data converge on *DRD4* methylation as the most robust epigenetic signal associated with physical aggression.

Besides dopamine signaling, we identified enrichment of other biological processes relevant to aggression. One of these was *'hormone metabolic processes'*, with several high-ranking DMPs annotated to thyroid hormone pathway genes (*TRH, FDX1, TNX2, TXNRD2)*, consistent with prior evidence of thyroid hormone dysregulation in pathological aggression in humans (Evrensel, Unsalver & Ozsahin, 2016) and attack behavior in mice (Hrabovszky et al., 2005). Another set of enriched pathways related to inflammatory processes, including '*cytokine secretion'*. This is in line with reported alterations in DNAm and expression levels of cytokine genes in individuals with a chronic history of childhood physical aggression (Provençal, Suderman, Caramaschi, et al., 2013; Provencal, Suderman, Vitaro, Szyf, & Tremblay, 2013), as well as in individuals diagnosed with Intermittent Explosive Disorder (Montalvo-Ortiz et al., 2017). We also identified enrichment for multiple CNS-related processes, such as '*Axon guidance*'. Interestingly, this same GO pathway was found to be highly enriched in the EWAS study on global aggression by van Dongen et al. (2015) as well as a recent bioinformatic analysis combining findings from six GWAS studies on aggression-related phenotypes (Fernàndez-Castillo & Cormand, 2016). Finally, we observed enrichment for regulation of '*behavior'*, which was again reported in the EWAS by van Dongen et al (2015) in a general population of twins, but also in the EWAS by Guillemin et al (2014) in adults following a chronic vs low trajectory of physical aggression in childhood. In summary, our findings point to an 'epigenetic signature' of physical aggression in buccal cells that is enriched for numerous biological processes, spanning neural, hormonal, immune and behavioral domains.

**Findings support the involvement of *DRD4* in aggressive behavior**

Dopamine is widely considered to be a key player in the neurobiology of aggressive behavior (Hagenbeek et al., 2016). Its uptake from the synaptic cleft is modulated by the dopamine receptor family, encoded by genes *DRD1* through to *DRD5.* Of these, *DRD4* has received the most attention for its potential role in aggression and related phenotypes (Pappa et al., 2015). Specifically, genetic variants of *DRD4* that confer lower expression have been associated with higher aggression, delinquency and externalizing behavior in child and adult populations (Fernàndez-Castillo & Cormand, 2016) as well as elevated risk for psychiatric diagnoses that are often accompanied by high aggression (e.g. CD, ADHD, ASPD, schizophrenia, substance use disorders; Ptacek, Kuzelova, & Stefano, 2011). Furthermore, *DRD4* polymorphisms have been associated with temperamental and personality traits closely related to aggression, including sensation-seeking, risk-taking, impulsivity, anger and reward sensitivity (Munafò, Yalcin, Willis-Owen, & Flint, 2008). The genetic literature on *DRD4*, however, has been far from consistent, with several studies reporting null findings (see Pappa, Mileva-Seitz, Bakermans-Kranenburg, Tiemeier, & van IJzendoorn, 2015, for a review). It has been suggested that, in addition to methodological differences, these inconsistencies may reflect moderation by the environment, as the dopamine system is known to be responsive to environmental inputs across development (Vassos, Collier, & Fazel, 2014).

The present study supports and extends previous findings of a link between *DRD4* and aggression at multiple biological levels, from genetics to physiology. Specifically, it is the first to show that this link is also observable at the epigenetic level, with lower levels of *DRD4* methylation found to associate with greater physically aggression. Interestingly, we found that several sites in the DMR were strongly associated with known genetic mQTLs (Gaunt et al., 2016) as well as showing small associations with environmental exposures in our sample, including neighborhood disadvantage and community violence exposure. While preliminary, these data are consistent with the hypothesis that DNAm patterns may reflect (and potentially mediate) the influence of genetic and environmental inputs on behavioral phenotypes (Uher, 2014). Although no study has specifically examined *DRD4* methylation in the context of aggressive behavior, several studies have reported an association with related phenotypes – particularly risk of ADHD (e.g. Dadds, Schollar-Root, Lenroot, Moul, & Hawes, 2016; van Mil et al., 2014). The question therefore arises as to whether DNAm patterns in this gene may be specific to aggression, shared with – or even entirely confounded by – co-occurring symptomatology. To address this, we examined associations between the *DRD4* DMR and common psychiatric and behavioral correlates of physical aggression. We

found that, besides physical aggression, DNAm levels in this region were most strongly correlated with self-reported drug use, followed by informant-rated levels of (low) prosocial behavior, callous-unemotional traits and externalizing – but not internalizing – psychiatric symptoms (esp. CD and ASPD). Overall, these findings are consistent with genetic studies implicating *DRD4* in externalizing psychiatric symptoms and substance use (Ptacek et al., 2011), as well as callous-unemotional and psychopathic traits (Yildirim & Derksen, 2015). In part, these shared associations may reflect the role of dopamine signaling in processes that are common to multiple externalizing problems, such as sensation-seeking, impulsivity and reward processing (Beauchaine & McNulty, 2013). Importantly, however, the association between *DRD4* methylation and physical aggression was not entirely explained by shared variance with other functional domains, remaining significant even after adjusting for all co-occurring symptomatology. Consequently, our findings suggest that while *DRD4* methylation patterns may be shared across multiple forms of externalizing problems, they also show an independent association with physically aggressive behavior. In future, more work will be needed to identify what factors may be driving the shared vs unique associations observed between *DRD4* methylation, physical aggression and co-occurring externalizing symptoms. Furthermore, an important avenue for future research will be to test whether specificity of associations can be improved if *DRD4* methylation is combined with other known risk factors for physical aggression (e.g. across psychosocial, demographic and contextual domains), and conversely, whether the inclusion of *DRD4* methylation in predictive models for physical aggression can result in greater predictive power and classification accuracy above the use of traditional risk factors alone.

### *DRD4* methylation as a potential cross-tissue biomarker of physical aggression

Having established a robust link between *DRD4* methylation and physical aggression in buccal cells, we proceeded to test the degree to which these patterns may also be observed in other peripheral and central tissues. We found that for over 80% of *DRD4* sites that were at least nominally associated with physical aggression in our sample (including all sites in the DMR region), DNAm levels in saliva significantly associated levels in blood. Our findings are consistent with previous reports of saliva-blood concordance in *DRD4* methylation, including a recent study documenting comparable associations between DNAm levels and ADHD risk in these two peripheral tissues (Dadds et al., 2016).

Given evidence of strong concordance, our next aim was to replicate the *DRD4*-aggression findings in blood, using the same sample upon which most of the epigenetic

studies of physical aggression to date have been based (Guillemin et al., 2014; Provençal, Suderman, Caramaschi, et al., 2013; Provençal et al., 2014; Dongsha Wang et al., 2012). Both samples were found to show consistency in the direction of associations with physical aggression (i.e. *lower* DNAm = *higher* physical aggression) as well as in the average *DRD4* methylation difference between groups (> 3%). Furthermore, effect sizes were typically modest in the discovery sample (average $g$ = -0.41) and large in the replication sample (average $g$ = -0.94), which may reflect more precise phenotyping in the latter (i.e. *discovery sample:* self-reported item indexing engagement in fights in the past year; *replication samples*: trajectories of informant-rated physical aggression based on repeated longitudinal assessments). In the replication sample, we found that 38% of probes annotated to *DRD4* were differentially methylated between the CPA and LPA group (compared to 57% in the discovery sample), including two sites located in our genome-wide DMP. As a more stringent test of replication, we built a polyepigenetic score (ePGS), where *DRD4* methylation levels from the replication sample were weighted by the (independent) effect sizes from the discovery sample, and found that this ePGS score was also differentially methylated between groups.

The above findings suggest that DNAm patterns in *DRD4* are comparable between buccal cells and blood, and reliably associate with physical aggression across these distinct peripheral tissues. Yet, based on this data alone, it is not possible to establish whether cross-tissue concordance extends to the brain – likely the most relevant organ for the study of aggression. To test this, we examined peripheral-CNS correlations in *DRD4* methylation, based on two independent data sources of matched premortem whole blood and postmortem brain tissue across (Hannon et al., 2015; Edgar et al., 2017). Similar to the saliva-blood findings, we found that the vast majority of *DRD4* sites at least nominally associated with engagement in physical fights showed significant peripheral-CNS concordance across both resources, including all sites in our DMR. These findings are in line with data showing high concordance in *DRD4* methylation patterns between blood and brain tissue in humans (Docherty et al., 2012) as well as in animals selectively bred for high vs low exploratory behavior (Verhulst et al., 2016). Interestingly, we found that physical aggression-related *DRD4* sites showed stronger peripheral-CNS convergence in frontotemporal regions compared to the cerebellum. However, it is important to note that we were not able to test whether *DRD4* methylation levels *in the brain* also associate with physical aggression, which will require access to postmortem samples with recorded histories on this behavioral phenotype.

Overall, the strong cross-tissue concordance in *DRD4* methylation observed in our study suggests that methylation status of this gene in easily accessible peripheral tissues, such as buccal cells and blood, may have utility as a valid proxy of methylation status in the brain. Furthermore, the robustness and reproducibility of our findings across independent samples and tissues points to *DRD4* methylation as a potential biomarker for physical aggression, in line with the broad definition of a biomarker (Biomarkers Definitions Working Group, 2001). This, however, represents only the first step in a longer process of biomarker discovery and validation. In particular, it will be important to systematically evaluate the accuracy (i.e. sensitivity and specificity) of *DRD4* methylation as a biomarker for physical aggression, in order to inform its potential translational applications.

**Probing causal pathways using Mendelian randomization**

Up to this point, our findings support *DRD4* methylation as a potential cross-tissue biomarker for physical aggression, with strong concordance observed across buccal cells/saliva, blood and brain. Yet, this alone does not necessarily imply that *DRD4* methylation plays a *mechanistic* role in physically aggressive behavior. In other words, a biomarker can be reliably associated with a phenotype of interest, without being causally related to it. Differentiating between causal and non-causal biomarkers is crucial for informing which translational application may be most appropriate (e.g. *non-causal biomarker* = use in detection, management and monitoring of symptoms; *causal biomarker* = potential target for prevention and treatment strategies). Based on functional studies on *DRD4*, DNAm levels in the gene body (i.e. where our DMR was located) have been found to *positively* associate with *DRD4* gene expression (i.e. *higher* DNA methylation relating to *higher* gene expression; Cheng et al., 2014). Thus, the association between DNAm and physical aggression observed in the present study could be interpreted as indicating that: (i) *lower* gene body methylation is linked with *lower DRD4* expression; which in turn (ii) relates to a *lower* density of *DRD4* receptors and *decreased* dopamine binding potential; thereby (iii) associating with *higher* physically aggressive behavior. Consistent with this, one study found that children with ADHD had lower concentrations of DRD4-mRNA in whole blood compared to a control group (Taurines et al., 2011). Furthermore, lower binding potential of D2-like dopamine receptors has been associated with a range of negative outcomes, including substance use and addiction (Martinez et al., 2007).

Here, we tested for a causal effect of *DRD4* methylation on physical aggression using Mendelian randomization analysis (Relton & Davey Smith, 2012). We found that only one of

the SNPs identified as mQTLs for *DRD4* methylation in our sample had the necessary characteristics to be used as a genetic instrument in the MR analysis (i.e. the only one with available GWAS data on aggression; Pappa et al., 2015). Using this single SNP, we did not find evidence to support a causal association between peripheral *DRD4* methylation and physical aggression, even though the direction of effects was consistent with our data (i.e. lower DNAm = higher aggression). If confirmed in future studies, these findings would suggest that peripheral *DRD4* methylation patterns may have greater utility as a non-causal biomarker of physical aggression rather than an intervention (i.e. modifiable) target for strategies aiming to reduce aggressive behavior. However, it is important to note that, given the reliance on a single SNP, our findings do not preclude entirely the possibility of causal effects, and as such should not be interpreted as conclusive evidence for lack of causality. Furthermore, it is possible that, rather than exerting causal effects per se, peripheral *DRD4* methylation patterns reflect DNAm status in the brain, where causal effects on physical aggression are exerted. In order to evaluate this possibility, future MR studies should aim to select as genetic instruments only SNPs that reliably associate with DRD4 methylation patterns in both peripheral and brain tissue, which we were not able to ascertain in our study. Finally, the analysis had a number of additional limitations (see below), so that findings should be considered preliminary, until more-powered and comprehensive MR studies are conducted testing this association.

**Limitations and future directions**

The present findings should be interpreted in light of several limitations. First, our discovery and replication samples were relatively small. Although both samples were at high-risk of psychiatric symptomatology and showed variability in physically aggressive behavior – thus increasing power to detect effects – more work will be needed to replicate findings in larger populations. Of note, no sex differences in physical aggression were identified in our sample, which is consistent with prior research reporting less pronounced sex differences in aggression within high-risk youth (Connor, 2012). Second, our genome-wide analysis was based on one self-reported item indexing engagement in physical fights over the past year. From this item, it was not possible to ascertain whether participants had actually initiated the fight or not. However, the fact that this measure associated with informant-rated psychiatric and behavioral correlates of physical aggression in the expected direction, and that *DRD4* findings replicated in an independent sample of individuals with well-characterized histories of physical aggression adds confidence to the present results.

Third, although we identified significant convergence in aggression-related *DRD4* methylation patterns across peripheral and CNS tissues, associations with aggression were only verified in buccal cells and blood. Consequently, it will be important to replicate our findings in the brain, for example by using post-mortem samples (Bakulski et al., 2016) and imaging epigenetic strategies in live humans (Nikolova & Hariri, 2015; Walton et al., 2017). Bearing this in mind, the strong cross-tissue convergence in *DRD4* methylation levels observed across tissues from different germ layers (*ectoderm*: buccal epithelial cells and brain; *mesoderm*: blood) is notable. Future studies should investigate whether this concordance may be driven by genetic effects – as supported by the large number of mQTLs associated with our *DRD4* loci – and/or driven by stochastic and environmental effects occurring during early embryonic development, as in the case of metastable epialleles (Rakyan, Blewitt, Druker, Preis & Whitelaw, 2002). Indeed, when cross-checking our five DMR sites with a recently published list of potential metastable epialleles (Van Baak et al., 2018), we found that none of the mQTL-associated sites were included the list, but conversely that one of the sites not associated with mQTLs (cg03909863) was listed as a candidate metastable epiallele. On a broader level, more work will be needed to characterize the cross-tissue properties of mQTLs and expression patterns associated with our aggression-related *DRD4* loci.

Fourth, data on smoking status and medication use were not available in the present study. In addition, our measures of substance use did not enable us to capture important characteristics, such as type, history and quantity of substance use, all of which may influence DNA methylation patterns. While we applied multiple strategies to minimize the possibility of residual confounding (e.g. SVA, independent replication, etc), future studies are needed to confirm our findings controlling for these potential confounders. It is noteworthy, however, that none of the DNAm sites identified in the present study overlapped with those found to be robustly affected by smoking in a recent systematic review (Gao et al., 2015). Fifth, we only focused on *DRD4* methylation following our epigenome-wide analyses, given that this gene was identified using both probe- and region-level approaches, as well as converging with prior evidence from molecular genetic research implicating *DRD4* in aggression. Nevertheless, our epigenome-wide analysis identified a number of additional DNA methylation sites of interest, which deserve further investigation in future.

Finally, although MR findings did not support a causal association of peripheral *DRD4* methylation with physical aggression, it is important to note that analyses were based on the single, currently usable genetic instrument (rather than focusing on all SNPs implicated in our

analyses), and may have therefore been underpowered to detect what are likely to be small effects. Furthermore, GWAS data for the analysis was extracted from a study investigating global aggression (Pappa et al., 2015), as opposed to physical aggression per se, which denotes a particularly severe aggression phenotype (Tremblay, 2000). In future, studies should aim to use genetic instruments from larger samples and derived from multiple SNPs, which explain greater variance in DNAm. It will also be of interest to apply a two-way MR to examine the possibility of reverse causation – that is, whether greater physical aggression leads to **lower** *DRD4* methylation. This approach should be complemented with prospective, longitudinal studies featuring repeated measures of *DRD4* methylation and physical aggression, in order to further disentangle the directionality of associations within a transactional framework. In order to systematically evaluate causality in the relationship between DRD4 methylation and physical aggression, it will also be necessary to triangulate findings from MR with those of other causal inference approaches , such as genetically-informative twin studies and animal experimental models (Pingault, O'Reilly, Schoeler, Ploubidis, Rijsdijk & Dudbridge, 2018). Together, future work addressing these research gaps will mark an important step toward characterizing more precisely the potential translational utility of *DRD4* methylation in the detection, prevention and **management** of physically aggressive behavior.

**Acknowledgements**

We would like to thank the young people, teachers and key workers who have taken part in this project. We are grateful to the Urban Academy, Haberdashers' Aske's Hatcham College and Hertswood Academy for their valuable collaboration. We thank Jo Guiney, Yvonne Whelan and Kathryn Hubbard for their assistance with data collection. The MLES and the QLSKC were both funded by Canadian Institutes for Health Research, the Social Sciences and Humanities Research Council of Canada, the Quebec Research Fund for Health, the Quebec Research Fund for Society and Culture. Research reported in this publication was supported by the Waterloo Foundation (Award Number: R1233/1872; PI: E.V.) and Kids Company charity. C.C. is supported by the Economic and Social Research Council (grant ref: ES/N001273/1). E.V. is a Royal Society Wolfson Research Merit Award Holder. J.B.P. is a fellow of MQ: Transforming Mental Health (MQ16IP16).

**Conflicts of interest:** None.

**References**

American Psychiatric Association. (2013). DSM-V. Diagnostic and Statistical Manual of Mental Disorders.

Bakulski, K. M., Halladay, A., Hu, V. W., Mill, J., & Fallin, M. D. (2016). Epigenetic Research in Neuropsychiatric Disorders: the "Tissue Issue". *Curr Behav Neurosci Rep, 3*(3), 264-274.

Barker, E. D., Walton, E., & Cecil, C. A. M. (2018). Annual Research Review: DNA methylation as a mediator in the association between risk exposure and child and adolescent psychopathology. *J Child Psychol Psychiatry*.

Barker, E. D., Walton, E., Cecil, C. A. M., Rowe, R., Jaffee, S. R., Maughan, B., . . . Gaunt, T. R. (2017). A Methylome-Wide Association Study of Trajectories of Oppositional Defiant Behaviors and Biological Overlap With Attention Deficit Hyperactivity Disorder. *Child Dev*.

Beach, S. R., Brody, G. H., Todorov, A. A., Gunter, T. D., & Philibert, R. A. (2011). Methylation at 5HTT mediates the impact of child sex abuse on women's antisocial behavior: an examination of the Iowa adoptee sample. *Psychosomatic medicine, 73*(1), 83.

Beauchaine, T. P., & McNulty, T. (2013). Comorbidities and continuities as ontogenic processes: Toward a developmental spectrum model of externalizing psychopathology. *Development and psychopathology*, 25(4pt2), 1505-1528.

Berman, A., Bergman, H., Palmstierna, T., & Schlyter, F. (2003). DUDIT: the drug use disorders identification test manual. *Karolinska Institutet, Department of Clinical Neuroscience, Stockholm*.

Bernstein, D. P., & Fink, L. (1998). *Childhood trauma questionnaire: A retrospective self-report manual*. Retrieved from San Antonio:

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther;69:89–95.

Boisjoli, R., Vitaro, F., Lacourse, E., Barker, E. D., & Tremblay, R. E. (2007). Impact and clinical significance of a preventive intervention for disruptive boys. *The British Journal of Psychiatry, 191*(5), 415-419.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., . . . Laird, R. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: a six-site, cross-national study. *Developmental psychology, 39*(2), 222.

Caramaschi, D., Sharp, G. C., Nohr, E. A., Berryman, K., Lewis, S. J., Davey Smith, G., & Relton, C. L. (2017). Exploring a causal role of DNA methylation in the relationship between maternal vitamin B12 during pregnancy and child's IQ at age 8, cognitive performance and educational attainment: a two-step Mendelian randomization study. *Human molecular genetics*, 26(15), 3001-3013.

Castillo, M., Mulet, J., Gutierrez, L. M., Ortiz, J. A., Castelan, F., Gerber, S., . . . Criado, M. (2005). Dual role of the RIC-3 protein in trafficking of serotonin and nicotinic acetylcholine receptors. *J Biol Chem, 280*(29), 27062-27068.

Cecil, C. A.M., Smith, R. G., Walton, E., Mill, J., McCrory, E. J., & Viding, E. (2016). Epigenetic signatures of childhood abuse and neglect: Implications for psychiatric vulnerability. *J Psychiatr Res, 83*, 184-194.

Cecil, C. A.M., Viding, E., Barker, E. D., Guiney, J., & McCrory, E. J. (2014). Double disadvantage: The influence of childhood maltreatment and community violence exposure on adolescent mental health. *J Child Psychol Psychiatry*.

Cecil, C. A. M., Walton, E., Jaffee, S. R., O'connor, T., Maughan, B., Relton, C. L., . . . Barker, E. D. (2017). Neonatal DNA methylation and early-onset conduct problems: A genome-wide, prospective study. *Dev Psychopathol*.

Checknita, D., Maussion, G., Labonte, B., Comai, S., Tremblay, R. E., Vitaro, F., . . . Turecki, G. (2015). Monoamine oxidase A gene promoter methylation and transcriptional downregulation in an offender population with antisocial personality disorder. *Br J Psychiatry, 206*(3), 216-222.

Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., . . . Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics, 8*(2), 203-209. doi:10.4161/epi.23470

Cheng, J., Wang, Y., Zhou, K., Wang, L., Li, J., Zhuang, Q., . . . Duan, S. (2014). Male-Specific Association between Dopamine Receptor D4 Gene Methylation and Schizophrenia. *PLoS ONE, 9*(2), e89128.

Chistiakov, D. A., & Chekhonin, V. P. (2017). Early-life adversity-induced long-term epigenetic programming associated with early onset of chronic physical aggression: Studies in humans and animals. *The World Journal of Biological Psychiatry*, 1-20.

Connor, D. F. (2012). Aggression and antisocial behavior in children and adolescents: Research and treatment. Guilford Press.

Cooley, M. R., Turner, S. M., & Beidel, D. C. (1995). Assessing community violence: the children's report of exposure to violence. *J Am Acad Child Adolesc Psychiatry*, *34*(2), 201-208.

Dadds, M. R., Schollar-Root, O., Lenroot, R., Moul, C., & Hawes, D. J. (2016). Epigenetic regulation of the DRD4 gene and dimensions of attention-deficit/hyperactivity disorder in children. *Eur Child Adolesc Psychiatry*, 1-9.

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics*, *23*(R1), R89-R98.

Docherty, S. J., Davis, O. S. P., Haworth, C. M. A., Plomin, R., D'Souza, U., & Mill, J. (2012). A genetic association study of DNA methylation levels in the DRD4 gene region finds associations with nearby SNPs. *Behavioral and Brain Functions : BBF*, *8*, 31-31.

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., Hawkins, J., . . . Chyen, D. (2008). Youth risk behavior surveillance—United States, 2007. *MMWR Surveill Summ*, *57*(4), 1-131.

Edgar, R. D., Jones, M. J., Meaney, M. J., Turecki, G., & Kobor, M. S. (2017). BECon: A tool for interpreting DNA methylation findings from blood in the context of brain. *bioRxiv*, 111609.

Eipel, M., Mayer, F., Arent, T., Ferreira, M. R., Birkhofer, C., Gerstenmaier, U., ... & Wagner, W. (2016). Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. *Aging*, 8(5), 1034.

Evans, D. M., & Davey Smith, G. (2015). Mendelian randomization: new applications in the coming age of hypothesis-free causality. Annual review of genomics and human genetics, 16, 327-350.

Evrensel, A., ÜNsalver, B. Ö., & ÖZŞAhİN, A. (2016). The Relationship between Aggression and Serum Thyroid Hormone Level in Individuals Diagnosed with Antisocial Personality Disorder. *Nöro Psikiyatri Arşivi, 53*(2), 120-125.

Fernàndez-Castillo, N., & Cormand, B. (2016). Aggressive behavior in humans: genes and pathways identified through association studies. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *171*(5), 676-696.

Freeman, B., Smith, N., Curtis, C., Huckett, L., Mill, J., & Craig, I. W. (2003). DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability

and suitability for multiplex polymerase chain reaction genotyping. *Behav Genet, 33*(1), 67-72.

Frick, P. J. (Producer). (2004). The inventory of callous-unemotional traits. Retrieved from Available: http://psyc.uno.edu/Frick%20Lab/ICU/ICU-teacher.pdf

Gadow, K. D., & Sprafkin, J. N. (1998). *Adolescent Symptom Inventory-4*: Checkmate Plus.

Gaunt, T. R., Shihab, H. A., Hemani, G., Min, J. L., Woodward, G., Lyttleton, O., . . . Ho, K. (2016). Systematic identification of genetic influences on methylation across the human life course. *Genome biology, 17*(1), 1.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586.

Guillemin, C., Provencal, N., Suderman, M., Cote, S. M., Vitaro, F., Hallett, M., . . . Szyf, M. (2014). DNA methylation signature of childhood chronic physical aggression in T cells of both men and women. *PLoS ONE, 9*(1), e86822.

Hagenbeek, F. A., Kluft, C., Hankemeier, T., Bartels, M., Draisma, H. H., Middeldorp, C. M., . . . Boomsma, D. I. (2016). Discovery of biochemical biomarkers for aggression: A role for metabolomics in psychiatry. *Am J Med Genet B Neuropsychiatr Genet, 171*(5), 719-732.

Hamshere, M. L., Langley, K., Martin, J., Agha, S. S., Stergiakouli, E., Anney, R. J., . . . Thapar, A. (2013). High loading of polygenic risk for ADHD in children with comorbid aggression. *Am J Psychiatry, 170*(8), 909-916.

Hannon, E., Lunnon, K., Schalkwyk, L., & Mill, J. (2015). Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics, 10*(11), 1024-1032.

Hemani, G., Zheng, J., Wade, K. H., Laurin, C., Elsworth, B., Burgess, S., . . . Yarmolinsky, J. (2016). MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv*, 078972.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome biology*, 14(10), 3156.

Hrabovszky, E., Halasz, J., Meelis, W., Kruk, M. R., Liposits, Z., & Haller, J. (2005). Neurochemical characterization of hypothalamic neurons involved in attack behavior: glutamatergic dominance and co-expression of thyrotropin-releasing hormone in a subset of glutamatergic neurons. *Neuroscience, 133*(3), 657-666.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet, 33 Suppl*, 245-254.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome research, 12*(6), 996-1006.

Kim-Cohen,J., Arseneault, L., Caspi, A., Tomás, M.P., Taylor, A., Moffitt, T.E. Validity of DSM-IV Conduct Disorder in 4½–5-Year-Old Children: A Longitudinal Epidemiological Study. (2005). *American Journal of Psychiatry*, 162(6),1108-17.

Klengel, T., & Binder, E. B. (2015). Epigenetics of stress-related psychiatric disorders and gene× environment interactions. *Neuron*, 86(6), 1343-1357.

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of abnormal psychology*, *116*(4), 645.

Ladd-Acosta, C., & Fallin, M. D. (2015). The role of epigenetics in genetic and environmental epidemiology.

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, *28*(6), 882-883.

Liang, L., & Cookson, W. O. C. (2014). Grasping nettles: cellular heterogeneity and other confounders in epigenome-wide association studies. *Human molecular genetics*, *23*(R1), R83-R88.

Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, 469(7330), 343.

Manchia, M., & Fanos, V. (2017). Targeting aggression in severe mental illness: the predictive role of genetic, epigenetic, and metabolomic markers. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*.

Martin, T. C., Yet, I., Tsai, P. C., & Bell, J. T. (2015). coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC bioinformatics*, *16*, 131.

Martinez, D., Narendran, R., Foltin, R. W., Slifstein, M., Hwang, D. R., Broft, A., . . . Laruelle, M. (2007). Amphetamine-induced dopamine release: markedly blunted in cocaine dependence and predictive of the choice to self-administer cocaine. *Am J Psychiatry*, *164*(4), 622-629.

Montalvo-Ortiz, J. L., Zhang, H., Chen, C., Liu, C., & Coccaro, E. F. (2018). Genome-wide DNA Methylation Changes Associated with Intermittent Explosive Disorder: A Gene-

Based Functional Enrichment Analysis. *International Journal of Neuropsychopharmacology*, 21(1), 12-20.

Munafò, M. R., Yalcin, B., Willis-Owen, S. A., & Flint, J. (2008). Association of the dopamine D4 receptor (DRD4) gene and approach-related personality traits: meta-analysis and new data. *Biological psychiatry, 63*(2), 197-206.

Nelson, R. J., & Trainor, B. C. (2007). Neural mechanisms of aggression. *Nat Rev Neurosci, 8*(7), 536-546.

Nikolova, Y. S., & Hariri, A. R. (2015). Can we observe epigenetic effects on human brain function? *Trends Cogn Sci, 19*(7), 366-373.

Noble, M., Wright, G., Smith, G., & Dibben, C. (2006). Measuring multiple deprivation at the small-area level. *Environment and Planning A, 38*(1), 169-185.

Pace, L. (2012). The "New" Statistics: Resampling and Bootstrapping. In Beginning R (pp. 257-268). Apress, Berkeley, CA.

Pappa, I., Mileva-Seitz, V. R., Bakermans-Kranenburg, M. J., Tiemeier, H., & van IJzendoorn, M. H. (2015). The magnificent seven: A quantitative review of dopamine receptor d4 and its association with child behavior. *Neuroscience & Biobehavioral Reviews, 57*, 175-186.

Pappa, I., St Pourcain, B., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M. G., . . . Tiemeier, H. (2015). A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *Am J Med Genet B Neuropsychiatr Genet*.

Pedersen, B. S., Schwartz, D. A., Yang, I. V., & Kechris, K. J. (2012). Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics, 28*(22), 2986-2988.

Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., Lord, R. V., . . . Molloy, P. L. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenetics & chromatin, 8*(1), 6.

Pidsley, R., CC, Y. W., Volta, M., Lunnon, K., Mill, J., & Schalkwyk, L. C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics, 14*, 293.

Pingault, J.-B., Cecil, C.A.M., Murray, J., Munafo, M., & Viding, E. (2016). Causal inference in psychopathology: A systematic review of Mendelian randomisation studies aiming to identify environmental risk factors for psychopathology. *Psychopathology Review*, *4*(1), 4-25.

Pingault, J. B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, 1.

Price, M. E., Cotton, A. M., Lam, L. L., Farre, P., Emberly, E., Brown, C. J., . . . Kobor, M. S. (2013). Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin, 6*(1), 4.

Prins, B. P., Abbasi, A., Wong, A., Vaez, A., Nolte, I., Franceschini, N., ... & Valdes, A. M. (2016). Investigating the causal relationship of C-reactive protein with 32 complex somatic and psychiatric outcomes: a large-scale cross-consortium Mendelian randomization study. *PLoS medicine*, 13(6), e1001976.

Provençal, N., Booij, L., & Tremblay, R. E. (2015). The developmental origins of chronic physical aggression: biological pathways triggered by early life adversity. *Journal of Experimental Biology, 218*(1), 123-133.

Provençal, N., Suderman, M. J., Caramaschi, D., Wang, D., Hallett, M., Vitaro, F., . . . Szyf, M. (2013). Differential DNA methylation regions in cytokine and transcription factor genomic loci associate with childhood physical aggression. *PLoS ONE, 8*(8), e71691.

Provençal, N., Suderman, M. J., Guillemin, C., Vitaro, F., Côté, S. M., Hallett, M., . . . Szyf, M. (2014). Association of Childhood Chronic Physical Aggression with a DNA Methylation Signature in Adult Human T Cells. *PLoS ONE, 9*(4), e89839.

Provencal, N., Suderman, M. J., Vitaro, F., Szyf, M., & Tremblay, R. E. (2013). Childhood chronic physical aggression associates with adult cytokine levels in plasma. *PLoS ONE, 8*(7), e69481.

Ptacek, R., Kuzelova, H., & Stefano, G. B. (2011). Dopamine D4 receptor gene DRD4 and its association with psychiatric disorders. *Med Sci Monit, 17*(9), RA215-220.

Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I., & Whitelaw, E. (2002). Metastable epialleles in mammals. *Trends in Genetics*, 18(7), 348-351.

Relton, C. L., & Davey Smith, G. (2012). Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol, 41*(1), 161-176.

Richardson, T. G., Haycock, P. C., Zheng, J., Timpson, N. J., Gaunt, T. R., Davey Smith, G., ... & Hemani, G. (2018). Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. *Human molecular genetics*.

Robinson, M. D., Kahraman, A., Law, C. W., Lindsay, H., Nowicka, M., Weber, L. M., & Zhou, X. (2014). Statistical methods for detecting differentially methylated loci and regions. *Frontiers in Genetics, 5*, 324.

Rouquette, A., Côté, S. M., Pryor, L. E., Carbonneau, R., Vitaro, F., & Tremblay, R. E. (2014). Cohort profile: The Quebec longitudinal study of kindergarten children (QLSKC). *International journal of epidemiology, 43*(1), 23-33.

Sanislow, C. A., Pine, D. S., Quinn, K. J., Kozak, M. J., Garvey, M. A., Heinssen, R. K., . . . Cuthbert, B. N. (2010). Developing constructs for psychopathology research: research domain criteria. *Journal of abnormal psychology, 119*(4), 631.

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction, 88*(6), 791-804.

Shah, S., Bonder, M. J., Marioni, R. E., Zhu, Z., McRae, A. F., Zhernakova, A., . . . Mendelson, M. M. (2015). Improving phenotypic prediction by combining genetic and epigenetic associations. *The American Journal of Human Genetics, 97*(1), 75-85.

Sheskin, D. (2003). *The handbook of parametric and nonparametric statistical procedures.* Boca Raton, FL: CRC Press.

Shimizu, H., Nagamori, I., Yabuta, N., & Nojima, H. (2009). GAK, a regulator of clathrin-mediated membrane traffic, also controls centrosome integrity and chromosome congression. *J Cell Sci, 122*(Pt 17), 3145-3152.

Smith, A. K., Kilaru, V., Klengel, T., Mercer, K. B., Bradley, B., Conneely, K. N., . . . Binder, E. B. (2015). DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain. *Am J Med Genet B Neuropsychiatr Genet, 168B*(1), 36-44.

Szyf, M. (2015). Prospects for the development of epigenetic drugs for CNS conditions. *Nature Reviews Drug Discovery, 14*(7), 461-474.

Szyf, M., & Bick, J. (2013). DNA methylation: a mechanism for embedding early life experiences in the genome. *Child Dev, 84*(1), 49-57.

Taurines, R., Grunblatt, E., Schecklmann, M., Schwenck, C., Albantakis, L., Reefschlager, L., . . . Romanos, M. (2011). Altered mRNA expression of monoaminergic candidate genes in the blood of children with attention deficit hyperactivity disorder and autism spectrum disorder. *World J Biol Psychiatry, 12 Suppl 1*, 104-108.

The GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science, 348*(6235), 648-660.

Theda, C., Hwang, S. H., Czajko, A., Loke, Y. J., Leong, P., & Craig, J. M. (2018). Quantitation of the cellular content of saliva and buccal swab samples. *Scientific reports*, 8(1), 6944.

Tremblay, R. E. (2000). The development of aggressive behaviour during childhood: What have we learned in the past century? *International Journal of Behavioral Development, 24*(2), 129-141.

Tremblay, R. E. (2015). Developmental origins of chronic physical aggression: An international perspective on using singletons, twins and epigenetics. *European Journal of Criminology, 12*(5), 551-561.

Tremblay, R. E., & Szyf, M. (2010). Developmental origins of chronic physical aggression and epigenetics. *Epigenomics, 2*(4), 495-499.

Tuvblad, C., & Baker, L. A. (2011). Human aggression across the lifespan: genetic propensities and environmental moderators. *Advances in genetics, 75*, 171.

Uher, R. (2014). Gene–environment interactions in common mental disorders: an update and strategy for a genome-wide search. *Social Psychiatry and Psychiatric Epidemiology, 49*(1), 3-14.

Van Baak, T. E., Coarfa, C., Dugué, P. A., Fiorito, G., Laritsky, E., Baker, M. S., ... & Saffari, A. (2018). Epigenetic supersimilarity of monozygotic twin pairs. *Genome biology*, 19(1), 2.

van Dongen, J., Nivard, M. G., Baselmans, B. M., Zilhão, N. R., Ligthart, L., Heijmans, B. T., . . . Consortium, B. (2015). Epigenome-wide association study of aggressive behavior. *Twin research and human genetics, 18*(6), 686-698.

van Iterson, M., van Zwet, E. W., & Heijmans, B. T. (2017). Controlling bias and inflation in epigenome-and transcriptome-wide association studies using the empirical null distribution. *Genome biology*, 18(1), 19.

van Mil, N. H., Steegers-Theunissen, R. P., Bouwland-Both, M. I., Verbiest, M. M., Rijlaarsdam, J., Hofman, A., . . . Verhulst, F. C. (2014). DNA methylation profiles at birth and child ADHD symptoms. *J Psychiatr Res, 49*, 51-59.

Vassos, E., Collier, D. A., & Fazel, S. (2014). Systematic meta-analyses and field synopsis of genetic association studies of violence and aggression. *Mol Psychiatry*, *19*(4), 471-477.

Verhulst, E. C., Mateman, A. C., Zwier, M. V., Caro, S. P., Verhoeven, K. J., & van Oers, K. (2016). Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD4 DNA methylation. *Mol Ecol, 25*(8), 1801-1811.

Veroude, K., Zhang-James, Y., Fernàndez-Castillo, N., Bakker, M. J., Cormand, B., & Faraone, S. V. (2016). Genetics of aggressive behavior: An overview. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 171*(1), 3-43.

Vitaro, F., Turecki, G., Ouellet-Morin, I., Pingault, J., Booij, L., Castellanos-Ryan, N., . . . Côté, S. (2012). Age-dependent effect of the MAOA gene on childhood physical aggression. *Molecular psychiatry, 18*(11), 1151.

Waltes, R., Chiocchetti, A. G., & Freitag, C. M. (2015). The neurobiological basis of human aggression: A review on genetic and epigenetic mechanisms. *Am J Med Genet B Neuropsychiatr Genet*.

Walton, E., Cecil, C. A. M., Suderman, M., Liu, J., Turner, J. A., Calhoun, V., . . . Barker, E. D. (2017). Longitudinal epigenetic predictors of amygdala:hippocampus volume ratio. *J Child Psychol Psychiatry, 58*(12), 1341-1350.

Walton, E., Pingault, J.-B., Cecil, C.A.M., Gaunt, T., Relton, C., Mill, J., & Barker, E. (2016). Epigenetic profiling of ADHD symptoms trajectories: a prospective, methylome-wide study. *Molecular psychiatry*.

Wang, Donshga, Szyf, M., Benkelfat, C., Provençal, N., Turecki, G., Caramaschi, D., . . . Booij, L. (2012). Peripheral SLC6A4 DNA methylation is associated with in vivo measures of human brain serotonin synthesis and childhood physical aggression. *PLoS ONE, 7*(6), e39501.

Wang, Dan, Yan, L., Hu, Q., Sucheston, L. E., Higgins, M. J., Ambrosone, C. B., . . . Liu, S. (2012). IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics, 28*(5), 729-730.

Waters, H., Hyder, A. A., Rajkotia, Y., Basu, S., Rehwinkel, J. A., Butchart, A., & Organization, W. H. (2004). The economic dimensions of interpersonal violence.

Xu, C., Fu, X., Zhu, S., & Liu, J. J. (2016). Retrolinkin recruits the WAVE1 protein complex to facilitate BDNF-induced TrkB endocytosis and dendrite outgrowth. *Mol Biol Cell, 27*(21), 3342-3356.

Yarbrough, G. G., Kamath, J., Winokur, A., & Prange, A. J., Jr. (2007). Thyrotropin-releasing hormone (TRH) in the neuroaxis: therapeutic effects reflect physiological functions and molecular actions. *Med Hypotheses, 69*(6), 1249-1256.

Yildirim, B. O., & Derksen, J. J. L. (2015). Mesocorticolimbic dopamine functioning in primary psychopathy: A source of within-group heterogeneity. *Psychiatry research, 229*(3), 633-677.

Zhu, Y., Sun, L., Chen, Z., Whitaker, J. W., Wang, T., & Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. *Nucleic acids research,* 41(22), 10032-10043.

**Table 1.** Epigenome-wide analysis of engagement in physical fights: Top 30 DMPs

| CpG Probe | Gene | Description | Chr | Position | Genomic location | Proximity to CpG Island | Standardized Beta | P-value | FDR (q) |
|---|---|---|---|---|---|---|---|---|---|
| **cg21217577** | *TRH* | *Thyrotropin-Releasing Hormone* | 3 | 129694683 | Body | Island | -.49 | 2,07E-08 | .01 |
| **cg11107262** | *GAK* | *Cyclin G Associated Kinase* | 4 | 871294 | Body | Island | -.46 | 1,52E-07 | .03 |
| **cg27349081** | *RIC3* | *RIC3 Acetylcholine Receptor Chaperone* | 11 | 8190670 | TSS200 | Island | .44 | 4,80E-07 | .05 |
| **cg27331554** | *CYFIP2* | *Cytoplasmic FMR1 Interacting Protein 2* | 5 | 156763290 | Body | - | -.44 | 5,18E-07 | .05 |
| cg05082738 | *LINC00301* | *Long Intergenic Non-Protein Coding RNA 301* | 11 | 60414918 | Body | Island | -.43 | 1,03E-06 | .08 |
| cg15916166 | *FAIM* | *Fas Apoptotic Inhibitory Molecule* | 3 | 138327519 | TSS200 | Island | .42 | 1,79E-06 | .11 |
| cg07212818 | *DRD4* | *Dopamine Receptor D4* | 11 | 638076 | Body | Island | -.42 | 2,16E-06 | .11 |
| cg05717871 | *DRD4* | *Dopamine Receptor D4* | 11 | 638507 | Body | Island | -.42 | 2,23E-06 | .11 |
| cg03909863 | *DRD4* | *Dopamine Receptor D4* | 11 | 638404 | Body | Island | -.42 | 2,29E-06 | .11 |
| cg08455772 | *FLJ12825* | *Uncharacterized LOC440101* | 12 | 54452069 | Body | North Shelf | -.41 | 3,92E-06 | .16 |
| cg22175303 | *FAIM* | *Fas Apoptotic Inhibitory Molecule* | 3 | 138327450 | TSS200 | North Shore | .40 | 5,20E-06 | .19 |
| cg11335335 | *DRD4* | *Dopamine Receptor D4* | 11 | 637885 | Body | Island | -.40 | 5,63E-06 | .19 |
| cg13694680 | *FIBIN* | *Fin Bud Initiation Factor Homolog* | 11 | 27016671 | 1stExon | - | .40 | 7,44E-06 | .23 |
| cg05930207 | *GJD3* | *Gap Junction Protein, Delta 3, 31.9kDa* | 17 | 38518520 | 1stExon | North Shore | -.40 | 8,23E-06 | .23 |
| cg01616529 | *DRD4* | *Dopamine Receptor D4* | 11 | 638424 | Body | Island | -.40 | 8,43E-06 | .23 |
| cg11481687 | - | - | 14 | 105499998 | - | Island | .39 | 1,05E-05 | .27 |
| cg03205258 | *TXNRD2;COMT* | *Thioredoxin Reductase 2; Catechol-O-Methyltransferase* | 22 | 19929274 | 1stExon | Island | .39 | 1,22E-05 | .29 |
| cg02445292 | *FCGBP* | *Fc Fragment Of IgG Binding Protein* | 19 | 40417583 | Body | North Shelf | .39 | 1,25E-05 | .29 |
| cg04580929 | *PDE4D* | *Phosphodiesterase 4D, CAMP-Specific* | 5 | 59711378 | 5'UTR | - | -.39 | 1,39E-05 | .30 |
| cg00996790 | - | - | 8 | 1432751 | - | South Shore | -.38 | 1,63E-05 | .31 |
| cg10105971 | *SGK1* | *Serum/Glucocorticoid Regulated Kinase 1* | 6 | 134638867 | 1stExon | Island | -.38 | 1,68E-05 | .31 |
| cg11153328 | *TXN2* | *Thioredoxin 2* | 22 | 36877737 | TSS200 | - | .38 | 1,69E-05 | .31 |
| cg10096566 | - | - | 14 | 95231817 | - | North Shelf | -.38 | 1,74E-05 | .31 |
| cg14234104 | *ANO1* | *Anoctamin 1, Calcium Activated Chloride Channel* | 11 | 69931467 | TSS200 | North Shelf | .38 | 1,92E-05 | .33 |
| cg13822329 | - | - | 10 | 130855111 | - | - | .38 | 2,10E-05 | .35 |
| cg26312965 | *SDF4* | *Stromal Cell Derived Factor 4* | 1 | 1152547 | 3'UTR | Island | -.38 | 2,36E-05 | .37 |
| cg10734526 | *GPM6B* | *Glycoprotein M6B* | X | 13956917 | TSS200 | Island | .38 | 2,40E-05 | .37 |
| cg01836910 | *CDK18* | *Cyclin-Dependent Kinase 18* | 1 | 205496136 | Body | North Shore | -.38 | 2,47E-05 | .37 |
| cg06674932 | *FDX1* | *Ferredoxin 1* | 11 | 110299342 | TSS1500 | North Shore | .38 | 2,62E-05 | .37 |
| cg23720732 | *SLC12A5* | *Solute Carrier Family 12, Member 5 (Potassium/Chloride Transporter)* | 20 | 44650380 | 5'UTR | North Shore | .37 | 3,00E-05 | .41 |

**Table 2**. Associations between methylomic variation in the DRD4 DMR and known correlates of physical aggression

| | **N of fights in the past year**[a] | | **Environmental risks**[b] | | | **Individual functioning** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Internalizing symptoms[c] | | | Externalizing symptoms[c] | | | | | Socio-emotional function[c] | | | Substance use[b] | |
| | *Bivariate r* | *Partial r* | *Childhood maltreatment* | *CVE* | *IMD* | *GAD* | *MDD* | *Internalizing total* | *ADHD* | *ODD* | *CD* | *ASPD* | *Externalizing total* | *Prosocial behavior* | *CU traits* | | *Alcohol use* | *Drug use* |
| **N of fights in the past year** | - | - | .26** | .39*** | .26** | .23* | .21* | .23* | .41*** | .31*** | .44*** | .52*** | .45*** | -.21* | .29** | | .20* | .49*** |
| **_DRD4_ Methylation (DMR)** | | | | | | | | | | | | | | | | | | |
| Individual CpGs | | | | | | | | | | | | | | | | | | |
| *cg07212818* | -.36*** | -.26** | -.17 | -.21* | -.11 | -.15 | -.19 | -.17 | -.24* | -.13 | -.22* | -.30*** | -.24* | .30** | -.34*** | | .17 | -.37*** |
| *cg05717871* | -.37*** | -.23** | -.15 | -.20* | -.23* | -.12 | -.19 | -.15 | -.21* | -.15 | -.25* | -.26* | -.23* | .33*** | -.35*** | | .03 | -.46*** |
| *cg03909863* | -.32*** | -.16 | -.16 | -.15 | -.23* | -0,1 | -.20 | -.14 | -.15 | -.07 | -.20 | -.25* | -.15 | .23* | -.30** | | .01 | -.43*** |
| *cg11335335* | -.38*** | -.30*** | -.11 | -.27** | -.20* | -.01 | -.08 | -.05 | -.12 | -.05 | -.23* | -.20* | -.13 | .29** | -.25* | | -.07 | -.33*** |
| *cg01616529* | -.34*** | -.22** | -.10 | -.12 | -.15 | -.04 | -.12 | -.08 | -.18 | -.09 | -.32** | -.33*** | -.20 | .29** | -.35*** | | .03 | -.41*** |
| DMR Principal component | -.39*** | -.28** | -.16 | -.13 | -.21* | -.12 | -.19 | -.15 | -.23* | -.14 | -.29** | -.33*** | -.24* | .35*** | -.35*** | | .07 | -.45*** |

Spearman correlations; * $p < .05$, ** $p < .01$, ***$p < .001$.

*Abbreviations.* CVE, community violence exposure; IMD, index of multiple deprivation; GAD, generalized anxiety disorder; MDD, major depressive disorder; ADHD, attention-deficit hyperactivity disorder; ODD, oppositional defiant disorder; CD conduct disorder; ASPD, antisocial personality disorder; CU, callous-unemotional.

[a] Correlations between *DRD4* methylation and engagement in physical fights, showing bivariate (unadjusted) spearman rho coefficients on the left, and partial rho coefficients on the right, adjusting for all domains of individual functioning.

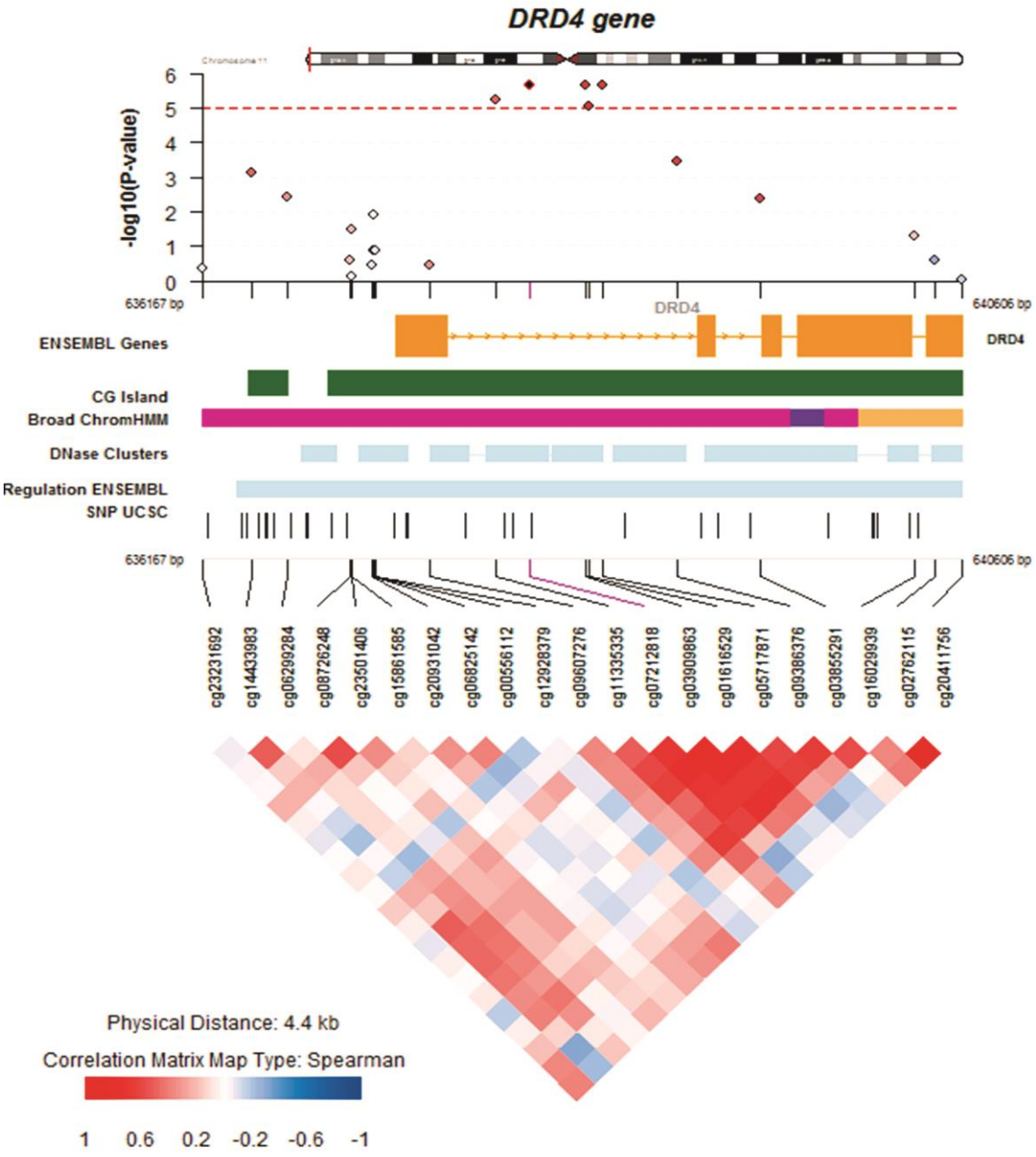
[b] Self-report; $n = 119$

[c] External-report, $n = 95$

**Table 3**. Methylomic variation across the *DRD4* gene and replication in an independent sample

| **A. Association between *DRD4* probes and engagement in physical fights** | | | | | | **B. Replication analysis** | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CpG | Genomic location | Proximity to CpG island | Position | *Standardized B* | P-value | **Discovery sample, split into two groups for comparability** (buccal epithelial cells) | | | | | | | | **Replication sample** (peripheral T cells; pooled DNA from 38 individuals) | | | | | | | | | | | |
| | | | | | | Overall sample (n = 119) | | No fights group (n = 88) | | Fights group (n = 31) | | % Diff | Effect size *g* | Overall sample (n = 12) | | LPA group (n = 6) | | CPA group (n = 6) | | % Diff | Effect size *g* | Bootrapped estimates | | | |
| | | | | | | *M* | *Sd* | *M* | *Sd* | *M* | *Sd* | | | *M* | *Sd* | *M* | *Sd* | *M* | *Sd* | | | *beta* | *SE* | *LCI* | *UCI* |
| *cg07212818* | Body | Island | 638076 | -0,42 | **2,16E-06** | 0,41 | 0,08 | 0,43 | 0,07 | 0,37 | 0,09 | -6,3% | -0,83 | 0,59 | 0,04 | 0,60 | 0,03 | 0,58 | 0,06 | -2,0% | -0,40 | -0,02 | 0,03 | -0,08 | 0,03 |
| *cg05717871* | Body | Island | 638507 | -0,42 | **2,23E-06** | 0,59 | 0,13 | 0,62 | 0,12 | 0,52 | 0,13 | -10,7% | -0,85 | 0,79 | 0,06 | 0,82 | 0,02 | 0,77 | 0,08 | -4,6% | -0,72 | -0,05 | 0,04 | -0,12 | 0,03 |
| *cg03909863* | Body | Island | 638404 | -0,42 | **2,29E-06** | 0,87 | 0,11 | 0,89 | 0,08 | 0,80 | 0,15 | -9,4% | -0,90 | 0,97 | 0,02 | 0,98 | 0,00 | 0,96 | 0,02 | -1,5% | -1,04 | **-0,02** | **0,01** | **-0,04** | **-0,001** |
| *cg11335335* | Body | Island | 637885 | -0,40 | **5,63E-06** | 0,44 | 0,12 | 0,47 | 0,11 | 0,37 | 0,12 | -10,0% | -0,86 | 0,35 | 0,07 | 0,39 | 0,04 | 0,31 | 0,07 | -7,9% | -1,23 | **-0,08** | **0,03** | **-0,18** | **-0,04** |
| *cg01616529* | Body | Island | 638424 | -0,40 | **8,43E-06** | 0,52 | 0,13 | 0,54 | 0,12 | 0,45 | 0,15 | -9,4% | -0,75 | 0,74 | 0,06 | 0,77 | 0,02 | 0,72 | 0,08 | -4,6% | -0,71 | -0,05 | 0,03 | -0,11 | 0,01 |
| *cg09386376* | Body | Island | 638939 | -0,32 | **3,57E-04** | 0,65 | 0,11 | 0,67 | 0,10 | 0,59 | 0,12 | -7,3% | -0,68 | 0,75 | 0,04 | 0,76 | 0,02 | 0,74 | 0,05 | -2,2% | -0,52 | -0,02 | 0,03 | -0,07 | 0,03 |
| *cg14433983* | TSS1500 | Island | 636460 | -0,30 | **7,72E-04** | 0,71 | 0,08 | 0,72 | 0,08 | 0,68 | 0,08 | -4,7% | -0,56 | 0,88 | 0,02 | 0,89 | 0,02 | 0,87 | 0,03 | -2,3% | -0,98 | -0,02 | 0,01 | -0,06 | 0,00 |
| *cg06299284* | TSS1500 | Island | 636659 | -0,26 | **3,70E-03** | 0,39 | 0,12 | 0,41 | 0,11 | 0,34 | 0,11 | -6,2% | -0,54 | 0,64 | 0,06 | 0,68 | 0,04 | 0,59 | 0,05 | -8,1% | -1,68 | **-0,09** | **0,03** | **-0,13** | **-0,01** |
| *cg03855291* | Body | Island | 639423 | -0,26 | **4,13E-03** | 0,48 | 0,12 | 0,49 | 0,11 | 0,43 | 0,14 | -5,7% | -0,48 | 0,54 | 0,03 | 0,55 | 0,03 | 0,53 | 0,02 | -2,3% | -0,79 | -0,02 | -0,02 | -0,06 | 0,02 |
| *cg06825142* | TSS200 | Island | 637170 | -0,23 | **0,01** | 0,09 | 0,02 | 0,09 | 0,02 | 0,08 | 0,01 | -1,1% | -0,60 | 0,05 | 0,01 | 0,05 | 0,01 | 0,04 | 0,01 | -1,2% | -1,45 | **-0,01** | **0,00** | **-0,02** | **-0,003** |
| *cg23501406* | TSS1500 | Island | 637035 | -0,20 | **0,03** | 0,06 | 0,02 | 0,06 | 0,02 | 0,06 | 0,01 | -0,6% | -0,39 | 0,11 | 0,04 | 0,13 | 0,03 | 0,09 | 0,04 | -4,1% | -1,20 | **-0,04** | **0,02** | **-0,08** | **-0,01** |
| *cg16029939* | Body | Island | 640328 | -0,18 | **0,05** | 0,40 | 0,12 | 0,41 | 0,12 | 0,36 | 0,11 | -5,7% | -0,49 | 0,10 | 0,04 | 0,12 | 0,04 | 0,09 | 0,01 | -3,4% | -0,97 | -0,04 | 0,03 | -0,07 | 0,02 |
| *cg12928379* | TSS200 | Island | 637175 | 0,14 | 0,13 | 0,07 | 0,01 | 0,06 | 0,01 | 0,07 | 0,01 | 0,4% | 0,36 | 0,04 | 0,01 | 0,05 | 0,01 | 0,04 | 0,01 | -1,1% | -1,35 | **-0,01** | **0,01** | **-0,02** | **-0,002** |
| *cg00556112* | TSS200 | Island | 637173 | -0,14 | 0,14 | 0,12 | 0,02 | 0,12 | 0,02 | 0,11 | 0,02 | -1,0% | -0,48 | 0,06 | 0,01 | 0,06 | 0,01 | 0,06 | 0,01 | -0,5% | -0,41 | -0,01 | 0,01 | -0,02 | 0,01 |
| *cg02762115* | Body | Island | 640446 | 0,11 | 0,26 | 0,55 | 0,11 | 0,55 | 0,11 | 0,56 | 0,13 | 0,3% | 0,03 | 0,25 | 0,05 | 0,27 | 0,05 | 0,23 | 0,04 | -4,6% | -0,95 | -0,05 | 0,03 | -0,10 | 0,01 |
| *cg08726248* | TSS1500 | Island | 637032 | -0,10 | 0,27 | 0,06 | 0,01 | 0,06 | 0,01 | 0,06 | 0,01 | -0,1% | -0,10 | 0,08 | 0,03 | 0,09 | 0,03 | 0,07 | 0,03 | -1,8% | -0,58 | -0,02 | 0,02 | -0,05 | 0,01 |
| *cg20931042* | TSS200 | Island | 637162 | -0,09 | 0,34 | 0,08 | 0,01 | 0,08 | 0,02 | 0,08 | 0,01 | -0,1% | -0,07 | 0,06 | 0,02 | 0,07 | 0,01 | 0,05 | 0,01 | -2,1% | -1,71 | **-0,02** | **0,01** | **-0,03** | **-0,01** |
| *cg09607276* | 1stExon | Island | 637491 | -0,08 | 0,37 | 0,07 | 0,03 | 0,07 | 0,03 | 0,07 | 0,02 | -0,3% | -0,12 | 0,08 | 0,01 | 0,08 | 0,01 | 0,08 | 0,01 | -0,8% | -0,72 | -0,01 | 0,01 | -0,02 | 0,00 |
| *cg23231692* | TSS1500 | N_Shore | 636167 | 0,07 | 0,46 | 0,84 | 0,03 | 0,84 | 0,03 | 0,84 | 0,03 | 0,0% | 0,00 | 0,97 | 0,01 | 0,97 | 0,01 | 0,96 | 0,01 | -0,2% | -0,28 | 0,00 | 0,00 | -0,02 | 0,01 |
| *cg15861585* | TSS1500 | Island | 637038 | -0,03 | 0,78 | 0,07 | 0,02 | 0,07 | 0,02 | 0,07 | 0,02 | -0,2% | -0,10 | 0,13 | 0,04 | 0,16 | 0,04 | 0,10 | 0,03 | -5,4% | -1,42 | **-0,06** | **0,02** | **-0,10** | **-0,02** |
| *cg20411756* | 3'UTR | Island | 640606 | 0,00 | 0,98 | 0,72 | 0,13 | 0,73 | 0,12 | 0,70 | 0,16 | -2,6% | -0,20 | 0,15 | 0,05 | 0,17 | 0,05 | 0,13 | 0,04 | -3,3% | -0,68 | -0,03 | 0,03 | -0,10 | 0,00 |

**Figure 1.** coMET plot of DRD4 methylation




N.b. from top to bottom: (i) the genomic location of all CpG sites annotated to DRD4; (ii) their strength of association with physical fights; (iii) their overlap with key regulatory elements; and (iv) their pattern of intercorrelations.

**Supplementary Material**
***DRD4 methylation as a potential biomarker for physical aggression: An epigenome-wide, cross-tissue investigation***

**List of supplementary material:**

**Supplementary Methods**

Optimized gene ontology method

A logistic regression approach was used to test if genes in the test list predicted pathway membership while controlling for the number probes annotated to each gene. Pathways were downloaded from the Gene Ontology website and all genes annotated to parent terms were also included. Illumina UCSC gene annotation was used to create a test gene list from probes that were associated with engagement in physical fights ($p < 0.001$). All genes with at least one methylation probe annotated and annotated to at least one GO pathway were considered. Pathways were filtered to those with between 10 and 2000 genes in. After applying this method to all pathways, significant pathways ($p < 0.05$) were taken and grouped where overlapping genes explained the signal. This was achieved by taking the most significant pathway, and retesting all remaining significant pathways while controlling additionally for the best term. If genes in the test list no longer predicted the pathway, the term was said to be explained by the most significant pathway, and hence these pathways were grouped together. This algorithm was repeated, taking the next most significant term, until all pathways had been considered as the most significant or found to be explained by a more significant term. GO terms were interpreted exclusively if they if they contained at least 2 genes and passed FDR-correction ($q < 0.05$), based on the final list of independent (non-redundant) terms.

Mendelian Randomization (MR) analysis

To run the MR analysis, we first selected genetic instruments (i.e. SNPs) for the identified DNAm markers using the mQTLdb database (http://www.mqtldb.org). The mQTLdb database contains the results of a study comprising 1000 mother-child pairs drawn from the Accessible Resource for Integrated Epigenomics Studies (ARIES; Relton et al., 2015), characterizing genome-wide significant SNP effects on DNAm levels for all Illumina 450k probes across five different life stages - birth, childhood, adolescence, pregnancy and middle age (Gaunt et al., 2016). SNP effects were examined both in *cis* (i.e. SNP within ±1000 base pairs of the DNAm site) and in *trans* (i.e. ±1 million base pairs). Here, we searched for the most representative, independent SNPs associated with our identified DNAm markers using the 'Genome-wide Complex Trait Analysis' (GCTA) setting, which was used to account for linkage disequilibium. To increase specificity with respect to the developmental period under investigation (i.e. adolescence), we selected the adolescence-specific mQTL weights and associated standard errors for use in the MR analysis, which were available under the 'Plink' setting. Once the appropriate SNPs were identified from the mQTLdb search, we proceeded to extract summary statistics about their association with physically aggressive behavior, based on recently published GWAS data from the EAGLE consortium encompassing 18,988 children and adolescents across nine cohorts (Pappa et al., 2015). Finally, the results from the mQTLdb search (i.e. SNP -> DNAm) and the GWAS data (i.e. SNP -> physical aggression) were included in the MR analysis, which was carried out using the TwoSampleMR package in R, available as part of the MR-Base platform (www.mrbase.org; Hemani et al., 2016).

Evaluating the influence of cell-type heterogeneity on our findings

While the use of buccal swabs typically results in greater cell-type homogeneity than other oral sample alternatives (e.g. saliva collection via passive drool; Theda et al 2018), these samples can still contain a mixture of different cell-types, namely buccal epithelial cells and leukocytes. To date, few methods have been developed for estimating cell-type heterogeneity in buccal swab samples, and these lack extensive validation (Langie et al., 2017). Bearing this in mind, we used two available methods that await full validation to examine the influence of cell-type heterogeneity on our findings:

First, we used the *Horvath online age calculator* (Horvath, 2013) to derive 'predicted tissue'. Although the goal of this calculator is to estimate epigenetic 'age' based on DNA methylation data, it also provides additional output, including an estimate of predicted tissue (i.e. most likely tissue that each sample is derived from). The results from this analysis indicated that 94 out of the 119 participants (79%) were correctly classified as '*buccal'*, and the remaining 25 participants were classified as '*saliva*'. No other tissue was predicted. As a sensitivity analysis, we reran regression models between the *DRD4* sites and physical aggression only in the subsample of participants with a *buccal* predicted tissue. The findings were highly consistent, with all sites that were at least nominally associated with physical aggression in the full sample (including probes in our DMR) remaining significant in the subsample (concordance between original and buccal-restricted analysis: $r$ = .91, p = 1.03E-08; results showin in Table S3). Of note, associations also remained significant in the subsample predicted as *saliva*, which supports our use of the data published by Smith et al (2014), to examine cross-tissue concordance between saliva and blood.

Second, we used the "*Buccal-Cell-Signature*" formula developed by Eipel et al (2016) to estimate the percentage of buccal epithelial cells, based on DNA methylation values from two CpG sites: *Buccal-Cell-Signature formula = (99.8(β-value of cg07380416) + 1.92) / 2 + (−98.12 (β-value of cg20837735) + 88.54) / 2.*The predicted fraction of buccal cells derived from our sample ranged between 22-84%, similar to the range reported by Eipel et al in their study (24-91%). We found that this variable did not correlate with either *DRD4* methylation levels (e.g. correlation with the *DMR* PCA: $r$ = -0.15; $p$ = 0.10) or with engagement in physical fights ($r$ = 0.03; $p$ = 0.75), making buccal cell fraction an unlikely confounder in the association between *DRD4* and aggression. This was confirmed using a Spearman partial correlation, which showed that this association was virtually unchanged after controlling for buccal cell fraction (*bivariate*: rho = -0.39, $p$ = 1.06E-05; *partial*: rho = -0.40, $p$ = 9.00E-06). Given that both methods await further validation, these findings should be interpreted with caution. Reassuringly, however, the two methods showed good convergence with one another (r = 0.73, p = 7.14E-21). Together, the above findings suggest that cell-type heterogeneity is unlikely to have affected our results.

References

Eipel, M., Mayer, F., Arent, T., Ferreira, M. R., Birkhofer, C., Gerstenmaier, U., ... & Wagner, W. (2016). Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. Aging (Albany NY), 8(5), 1034.

Hemani, G., Zheng, J., Wade, K. H., Laurin, C., Elsworth, B., Burgess, S., . . . Yarmolinsky, J. (2016). MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. bioRxiv, 078972.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome biology, 14(10), 3156.

Langie, S. A., Moisse, M., Declerck, K., Koppen, G., Godderis, L., Vanden Berghe, W., ... & De Boever, P. (2017). Salivary DNA methylation profiling: aspects to consider for biomarker identification. Basic & clinical pharmacology & toxicology, 121(S3), 93-101.

Pappa, I., St Pourcain, B., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M. G., . . . Tiemeier, H. (2015). A genome-wide approach to children's aggressive behavior: The EAGLE consortium. Am J Med Genet B Neuropsychiatr Genet.

Relton, C. L., Gaunt, T., McArdle, W., Ho, K., Duggirala, A., Shihab, H., ... & Paul, Y. L. (2015). Data resource profile: accessible resource for integrated epigenomic studies (ARIES). International journal of epidemiology, 44(4), 1181-1190.

Theda, C., Hwang, S. H., Czajko, A., Loke, Y. J., Leong, P., & Craig, J. M. (2018). Quantitation of the cellular content of saliva and buccal swab samples. Scientific reports, 8(1), 6944.

**Table S1.** Gene ontology (GO) analysis results

| GO Pathway | *n*genes in input list | Total *n*in pathway | % | *p*-value | *GO ID* | Genes |
|---|---|---|---|---|---|---|
| **Biological Processes** | | | | | | |
| negative regulation of receptor-mediated endocytosis | 3 | 10 | 30% | 2.94E-10 | GO:0048261 | *RAC1;PICALM;ARF6* |
| positive regulation of phosphatidylinositol 3-kinase activity | 5 | 25 | 20% | 4.66E-10 | GO:0043552 | *EPHA8;PDGFRB;RAC1;PTK2;VAV2* |
| positive regulation of protein kinase B signaling | 9 | 71 | 13% | 1.18E-09 | GO:0051897 | *SEMA5A;EGFR;PTK2;TCF7L2;GATA3;NOX4;IGF1R;C1QTNF1;CCL3* |
| **negative regulation of axon extension involved in axon guidance** | 3 | 10 | 30% | 1.66E-09 | GO:0048843 | *SEMA3F;SEMA5A;SLIT1* |
| positive regulation of keratinocyte differentiation | 3 | 12 | 25% | 1.05E-08 | GO:0045618 | *NOTCH1;PRKCH;NCOA3* |
| **regulation of hormone metabolic process** | 5 | 27 | 19% | 1.13E-08 | GO:0032350 | *POR;TCF7L2;GATA3;IGF1R;CACNA1A* |
| establishment of nucleus localization | 3 | 11 | 27% | 1.23E-08 | GO:0040023 | *SYNE1;PTK2;SLIT1* |
| coronary vasculature development | 3 | 13 | 23% | 1.76E-07 | GO:0060976 | *PDGFRB;NOTCH1;TBX1* |
| cochlea morphogenesis | 4 | 22 | 18% | 2.11E-07 | GO:0090103 | *GLI2;RAC1;PAX2;TBX1* |
| **regulation of interleukin-5 production** | 3 | 13 | 23% | 7.50E-07 | GO:0032674 | *PRKCZ;PDE4D;GATA3* |
| tongue development | 4 | 19 | 21% | 8.72E-07 | GO:0043586 | *PRDM16;WNT10A;EGFR;TBX1* |
| **negative regulation of behavior** | 5 | 37 | 14% | 8.86E-07 | GO:0048521 | *SEMA3F;TRH;SEMA5A;NOTCH1;SLIT1* |
| **dopamine receptor signaling pathway** | 4 | 24 | 17% | 1.51E-06 | GO:0007212 | *CALY;DRD4;GNAO1;GNAL* |
| negative regulation of axonogenesis | 6 | 46 | 13% | 2.21E-06 | GO:0050771 | *SEMA3F;SEMA5A;SYNGAP1;PTK2;SLIT1;ARHGEF1* |
| regulation of glial cell proliferation | 3 | 15 | 20% | 3.42E-06 | GO:0060251 | *SKI;NOTCH1;PRKCH* |
| circadian rhythm | 8 | 89 | 9% | 3.47E-06 | GO:0007623 | *BTBD9;EGFR;ADK;DRD4;RBM4B;TPH2;CRY1;RORA* |
| pharyngeal system development | 3 | 16 | 19% | 4.21E-06 | GO:0060037 | *PLXNA2;GATA3;TBX1* |
| circadian regulation of gene expression | 3 | 18 | 17% | 6.20E-06 | GO:0032922 | *ADK;RBM4B;CRY1* |
| **cytokine secretion** | 4 | 29 | 14% | 6.73E-06 | GO:0050663 | *AIM2;TRIM27;NOTCH1;GATA3* |
| neuron projection morphogenesis | 30 | 514 | 6% | 7.43E-06 | GO:0048812 | *PLXNB3;NTRK1;EPHA8;PLXNA2;GLI2;CREB1;SEMA3F;SEMA5A;EGFR;RAC1;FEZF1;BAI1;ANK1;PTK2;NOTCH1;VAV2;TLN1;COL5A1;ANK3;GATA3;SLIT1;PAX2;PICALM;WEE1;ATP8A2;IGF1R;PTPRM;MYL12B;CACNA1A;BRSK1* |
| **response to ethanol** | 9 | 109 | 8% | 8.16E-06 | GO:0045471 | *NTRK1;SDF4;TRH;GSN;MGMT;GATA3;HTR3A;DRD4;AACS* |
| regulation of protein targeting to membrane | 3 | 18 | 17% | 8.28E-06 | GO:0090313 | *FIS1;ANK3;KCNE1* |
| semaphorin-plexin signaling pathway | 3 | 17 | 18% | 1.17E-05 | GO:0071526 | *PLXNA2;SEMA3F;RAC1* |
| establishment of cell polarity | 7 | 65 | 11% | 2.19E-05 | GO:0030010 | *PRKCZ;PTK2;WEE1;ARF6;IGF1R;NDE1;BRSK1* |
| regulation of cellular component organization | 58 | 1468 | 4% | 3.10E-05 | GO:0051128 | *PLXNB3;GPM6B;NTRK1;PRKCZ;FMN2;AIM2;SHOX2;UBE2E1;SEMA3F;RUFY1;NEUROG1;SEMA5A;PDGFRB;SYNGAP1;BTBD9;SGK1;FOXK1;MAD1L1;RAC1;FIS1;ANK1;PTK2;DLC1;PPP3CC;TERF1;NOTCH1;ARPC5L;CREB3;GSN;VCP;COL5A1;ANK3;MTG1;TCF7L2;GATA3;CALY;SLIT1;NRG3;PAX2;PIC* |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | *ALM;NOX4;PRKCH;ARF6;SNW1;ISLR2;BUB1B;TBCD;CDC42EP4;ACCN1;SEZ6;PSMC5;CCL3;MYL12B;CACNA1A;ARHGEF1;TIAM1;C21orf2;KCNE1* |
| digestive tract morphogenesis | 6 | 54 | 11% | 3.33E-05 | GO:0048546 | *GLI2;SHOX2;FOXP4;EGFR;NOTCH1;TCF7L2* |
| white fat cell differentiation | 3 | 12 | 25% | 3.60E-05 | GO:0050872 | *PRDM16;NCOR2;AACS* |
| negative regulation of ion transport | 6 | 67 | 9% | 6.12E-05 | GO:0043271 | *GPM6B;PPOX;TRH;TRIM27;GNAO1;ERG* |
| regulation of actin filament-based process | 15 | 248 | 6% | 8.19E-05 | GO:0032970 | *GPM6B;FMN2;PDE4D;SEMA5A;PDGFRB;RAC1;DLC1;ARPC5L;GSN;NOX4;FRMD6;ARF6;DSC3;KCNE1;ERG* |
| artery morphogenesis | 5 | 48 | 10% | 9.07E-05 | GO:0048844 | *PDGFRB;CHD7;NOTCH1;TBX2;TBX1* |
| vesicle coating | 4 | 39 | 10% | 1.14E-04 | GO:0006901 | *COPA;SEC31A;CALY;PICALM* |
| **response to drug** | 20 | 395 | 5% | 1.17E-04 | GO:0042493 | *NTRK1;PPOX;AIM2;ABCG5;HDAC4;CREB1;EGFR;POR;TERF1;MGMT;GATA3;AACS;CYP1A2;GNAO1;RPH3AL;PTPRM;SLC12A5;ERG;COMT;TXN2* |
| regulation of system process | 28 | 584 | 5% | 1.24E-04 | GO:0044057 | *APLN;GPM6B;NTRK1;PRKCZ;HSPB7;SKI;PPOX;ABCG5;HDAC4;KNG1;PDE4D;SYNGAP1;BTBD9;SGK1;EGFR;MTG1;TCF7L2;ADK;DRD4;SHANK2;GNAO1;ACCN1;CCL3;PTPRM;DSC3;CACNA1A;KCNE1;ERG* |
| regulation of protein localization | 21 | 450 | 5% | 1.25E-04 | GO:0032880 | *GPM6B;PRKCZ;AIM2;SEMA5A;TRIM27;EGFR;SLC26A4;FIS1;DLC1;PPP3CC;CREB3;ANK3;TCF7L2;GATA3;PICALM;DRD4;ARF6;RPH3AL;CCL3;KCNE1;RANGAP1* |
| positive regulation of DNA replication | 5 | 53 | 9% | 1.36E-04 | GO:0045740 | *GLI2;EGFR;RAC1;IGF1R;MAP2K4* |
| regulation of cellular component movement | 26 | 554 | 5% | 1.46E-04 | GO:0051270 | *PLXNB3;PLXNA2;SEMA3F;PDE4D;SEMA5A;PDGFRB;SGK1;BMPER;EGFR;RAC1;CDK6;PTK2;DLC1;NOTCH1;CREB3;GATA3;SLIT1;NOX4;ROBO4;ONECUT1;IGF1R;CCL3;PTPRM;DSC3;KCNE1;ERG* |
| glucosamine-containing compound metabolic process | 3 | 25 | 12% | 1.47E-04 | GO:1901071 | *NAGK;PGM3;GNE* |
| **nervous system process** | 58 | 1616 | 4% | 1.67E-04 | GO:0050877 | *TBL1X;CLCN5;APLN;NTRK1;CNIH3;PRKCZ;PPOX;COPA;ABCG5;HDAC4;KCNJ3;CREB1;VIPR1;CPA3;KNG1;PDE4D;COL11A2;SYNGAP1;BTBD9;SGK1;EGFR;SLC26A4;PTK2;CHD7;DFNB31;TLN1;ANK3;GATA3;SLIT1;PAX2;HTR3A;DRD4;SHANK2;RIC3;TNNI2;KCNA6;PRPH;ATP8A2;SLC24A4;APBA2;TRPM1;A2BP1;GNAO1;CRYM;ACCN1;SEZ6;MC2R;GNAL;MYL12B;CACNA1A;AP3D1;BRSK1;SLC12A5;KCNE1;ERG;COMT;TBX1;PI4KA* |
| sensory perception of mechanical stimulus | 10 | 138 | 7% | 2.07E-04 | GO:0050954 | *TBL1X;NTRK1;COL11A2;SLC26A4;CHD7;DFNB31;CRYM;ACCN1;KCNE1;TBX1* |
| regulation of heart morphogenesis | 3 | 23 | 13% | 2.66E-04 | GO:2000826 | *NOTCH1;TBX2;TBX1* |

| | | | | | | |
|---|---|---|---|---|---|---|
| ventricular cardiac muscle cell action potential | 3 | 21 | 14% | 2.68E-04 | GO:0086005 | *DSC3;KCNE1;ERG* |
| response to organic cyclic compound | 27 | 647 | 4% | 2.78E-04 | GO:0014070 | *NTRK1;NFIA;GSTM3;LIN28;GLI2;TRH;PDE4D;PDGFRB;EGFR;GSN;MGMT;GATA3;HTR3A;DRD4;NOX4;NCOR2;AACS;TPH2;IGF1R;CYP1A2;GNAO1;CCL3;GNAL;KCNE1;HLCS;COMT;TXN2* |
| inositol lipid-mediated signaling | 10 | 162 | 6% | 3.52E-04 | GO:0048017 | *NTRK1;CREB1;PDGFRB;EGFR;GSN;GATA3;IGF1R;BUB1B;TNRC6C;PI4KA* |
| vesicle-mediated transport | 35 | 903 | 4% | 3.61E-04 | GO:0016192 | *CLCN5;RER1;FMN2;SDF4;COPA;TUBA4A;KNG1;SEC31A;RUFY1;CYFIP2;EXOC2;KIAA1244;RAC1;ANK1;PTK2;ATP6V1H;VAV2;TLN1;GSN;VCP;ANK3;CALY;LDLRAD3;PICALM;RIN1;CD63;ARF6;COG4;RPH3AL;CCL3;CACNA1A;AP3D1;ICAM5;TRAPPC10;SNAP29* |
| positive regulation of ion transport | 9 | 143 | 6% | 3.81E-04 | GO:0043270 | *PPOX;TRH;PDGFRB;CREB3;ANK3;DRD4;CCL3;KCNE1;ERG* |
| positive regulation of membrane potential | 3 | 21 | 14% | 3.82E-04 | GO:0045838 | *PRKCZ;ANK3;DRD4* |
| eating behavior | 3 | 26 | 12% | 3.88E-04 | GO:0042755 | *PPOX;TRH;ATP8A2* |
| mitotic spindle assembly checkpoint | 4 | 33 | 12% | 3.98E-04 | GO:0007094 | *UBE2E1;MAD1L1;TERF1;BUB1B* |
| negative regulation of BMP signaling pathway | 4 | 37 | 11% | 5.18E-04 | GO:0030514 | *SKI;BMPER;NOTCH1;TCF7L2* |
| gland development | 16 | 285 | 6% | 5.48E-04 | GO:0048732 | *APLN;GLI2;WNT10A;CREB1;PDGFRB;EGFR;NOTCH1;MGMT;TCF7L2;GATA3;NRG3;NCOR2;IGF1R;TBX2;NCOA3;TBX1* |
| positive regulation of vasodilation | 3 | 27 | 11% | 5.56E-04 | GO:0045909 | *APLN;EGFR;PTPRM* |
| retina development in camera-type eye | 8 | 113 | 7% | 5.85E-04 | GO:0060041 | *SKI;PDGFRB;CHD7;PAX2;DRD4;MDM1;PRPH;PTPRM* |
| positive regulation of ion transmembrane transport | 3 | 24 | 13% | 6.40E-04 | GO:0034767 | *PDGFRB;KCNE1;ERG* |
| **sympathetic nervous system development** | 3 | 24 | 13% | 6.68E-04 | GO:0048485 | *NTRK1;SEMA3F;GATA3* |
| multicellular organismal movement | 3 | 27 | 11% | 7.38E-04 | GO:0050879 | *TNNI2;ATP8A2;CACNA1A* |
| regulation of reactive oxygen species metabolic process | 5 | 64 | 8% | 8.79E-04 | GO:2000377 | *PDGFRB;EGFR;RAC1;PAX2;NOX4* |
| **regulation of interleukin-2 production** | 4 | 45 | 9% | 8.88E-04 | GO:0032663 | *PDE4D;TRIM27;CARD11;GATA3* |
| negative regulation of osteoblast differentiation | 4 | 34 | 12% | 8.97E-04 | GO:0045668 | *SKI;HDAC4;CDK6;NOTCH1* |
| homeostatic process | 38 | 1087 | 3% | 9.89E-04 | GO:0042592 | *PPOX;NCSTN;ABCG5;KNG1;TERT;PDGFRB;COL11A2;TAP2;BTBD9;POLD2;EGFR;RAC1;SLC26A4;FIS1;ANK1;ATP6V1H;TERF1;DFNB31;TCF7L2;GATA3;PAX2;DRD4;NOX4;RIC3;NCOR2;AACS;SLC24A4;TRPM1;FAM82A2;NARFL;SLC9A5;RPH3AL;CCL3;CACNA1A;SLC12A5;ERG;TXNRD2;TXN2* |
| head development | 5 | 53 | 9% | 1.21E-03 | GO:0060322 | *SKI;GLI2;CHD7;TCF7L2;TBX1* |
| pancreas development | 6 | 80 | 8% | 1.21E-03 | GO:0031016 | *CDK6;TCF7L2;PAX2;ONECUT1;IGF1R;MEIS2* |
| membrane depolarization | 6 | 73 | 8% | 1.29E-03 | GO:0051899 | *PRKCZ;PPOX;DRD4;SEZ6;CACNA1A;ERG* |
| epithelial cell morphogenesis | 4 | 39 | 10% | 1.41E-03 | GO:0003382 | *RAC1;GATA3;PAX2;FRMD6* |
| regulation of calcium ion-dependent exocytosis | 3 | 27 | 11% | 1.47E-03 | GO:0017158 | *NOTCH1;RPH3AL;CACNA1A* |

| | | | | | | |
|---|---|---|---|---|---|---|
| prostate gland epithelium morphogenesis | 3 | 26 | 12% | 1.55E-03 | GO:0060740 | *GLI2;NOTCH1;IGF1R* |
| anatomical structure maturation | 4 | 46 | 9% | 1.58E-03 | GO:0071695 | *NOTCH1;GATA3;PAX2;CACNA1A* |
| chloride transmembrane transport | 4 | 48 | 8% | 1.91E-03 | GO:1902476 | *CLCN5;SLC26A4;ANO1;SLC12A5* |
| lung epithelial cell differentiation | 3 | 27 | 11% | 2.28E-03 | GO:0060487 | *CREB1;FOXP4;NCOR2* |
| heart valve development | 3 | 27 | 11% | 2.40E-03 | GO:0003170 | *SHOX2;NOTCH1;GATA3* |
| **long-term memory** | 3 | 27 | 11% | 2.44E-03 | GO:0007616 | *PRKCZ;BTBD9;SGK1* |
| positive regulation of multicellular organism growth | 3 | 32 | 9% | 2.61E-03 | GO:0040018 | *CREB1;CHD7;ATP8A2* |
| protein N-linked glycosylation | 6 | 103 | 6% | 2.66E-03 | GO:0006487 | *SEC31A;PGM3;VCP;MGAT5B;KCNE1;ALG12* |
| protein trimerization | 3 | 32 | 9% | 3.15E-03 | GO:0070206 | *SKI;TRIM27;C1QTNF1* |
| adult locomotory behavior | 6 | 86 | 7% | 3.31E-03 | GO:0008344 | *TRH;BTBD9;CHD7;DRD4;SEZ6;CACNA1A* |
| regulation of skeletal muscle tissue development | 5 | 61 | 8% | 3.46E-03 | GO:0048641 | *HDAC4;SHOX2;NOTCH1;TCF7L2;TBX1* |
| cell fate determination | 4 | 43 | 9% | 3.56E-03 | GO:0001709 | *LBX1;GATA3;PAX2;TBX2* |
| cellular protein complex assembly | 11 | 235 | 5% | 4.03E-03 | GO:0043623 | *TUBA4A;TUBB;RAC1;PTK2;GSN;CALY;NRG3;PICALM;RIC3;COG4;CCT6B* |
| porphyrin-containing compound metabolic process | 3 | 41 | 7% | 7.56E-03 | GO:0006778 | *PPOX;ANK1;CYP1A2* |
| G2/M transition of mitotic cell cycle | 7 | 133 | 5% | 9.13E-03 | GO:0000086 | *TUBA4A;TUBB;DLC1;TERF1;DCTN3;WEE1;NDE1* |
| **Cellular components** | | | | | | |
| vesicle coat | 4 | 40 | 10% | 1.80E-04 | GO:0030120 | *COPA;SEC31A;EGFR;PICALM* |
| dendritic shaft | 4 | 35 | 11% | 2.91E-04 | GO:0043198 | *CNIH3;SYNGAP1;SEZ6;SLC12A5* |
| apical part of cell | 14 | 313 | 4% | 2.17E-03 | GO:0045177 | *CLCN5;PRKCZ;ABCG5;AMOTL2;PDE4D;PDGFRB;EGFR;SLC26A4;PTK2;EPB41L4B;SHANK2;NOX4;MYL12B;KCNE1* |
| chromosomal part | 23 | 600 | 4% | 2.62E-03 | GO:0044427 | *DNMT3A;CREB1;TERT;C6orf167;ORC3L;HUS1B;MAD1L1;POLD2;DLC1;TERF1;VCP;DCTN3;TCF7L2;GATA3;RNF169;NCOR2;PDS5B;SNW1;USP3;BUB1B;NDE1;HLCS;RANGAP1* |
| nuclear matrix | 6 | 95 | 6% | 8.78E-03 | GO:0016363 | *PRKCZ;DNMT3A;NCOR2;SNW1;HLCS;PHF5A* |
| **Molecular functions** | | | | | | |
| transmembrane receptor protein tyrosine kinase activity | 8 | 67 | 12% | 3.34E-07 | GO:0004714 | *NTRK1;INSRR;ROR1;EPHA8;PDGFRB;TRIM27;EGFR;IGF1R* |
| insulin receptor substrate binding | 3 | 13 | 23% | 1.31E-06 | GO:0043560 | *INSRR;PRKCZ;IGF1R* |
| peptide hormone receptor binding | 3 | 16 | 19% | 1.48E-06 | GO:0051428 | *GNAO1;RUNDC3A;PSMC5* |
| neuropeptide receptor binding | 3 | 23 | 13% | 8.42E-05 | GO:0071855 | *APLN;PPOX;GNAO1* |
| proteoglycan binding | 3 | 21 | 14% | 1.82E-04 | GO:0043394 | *SEMA5A;COL5A1;SLIT1* |
| structural constituent of cytoskeleton | 7 | 91 | 8% | 1.98E-04 | GO:0005200 | *TUBA4A;TUBA4B;TUBB;ANK1;TLN1;EPB41L4B;ANK3* |
| protein kinase C activity | 3 | 20 | 15% | 2.03E-04 | GO:0004697 | *PRKCZ;PRKCH;CCL3* |
| retrograde vesicle-mediated transport | 3 | 24 | 13% | 3.48E-04 | GO:0006890 | *RER1;COPA;COG4* |
| potassium channel regulator activity | 4 | 37 | 11% | 3.62E-04 | GO:0015459 | *PRKCZ;SGK1;DRD4;KCNE1* |
| phosphatidylinositol 3-kinase binding | 3 | 23 | 13% | 4.68E-04 | GO:0043548 | *INSRR;PDGFRB;IGF1R* |
| iron-sulfur cluster binding | 5 | 64 | 8% | 4.74E-04 | GO:0051536 | *RFESD;TYW1B;FDX1;NARFL;ACO2* |
| metal cluster binding | 5 | 64 | 8% | 4.74E-04 | GO:0051540 | *RFESD;TYW1B;FDX1;NARFL;ACO2* |
| carboxy-lyase activity | 3 | 35 | 9% | 2.57E-03 | GO:0016831 | *UXS1;ME3;PPCDC* |

**Table S2.** Functional characterization of *DRD4* CpG loci

| | CpG information | | | | | | | ENCODE Regulatory Elements[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CpG | Genomic location | Proximity to CpG island | Position | *Std B* | P-value | Mean | *SD* | TF | Histone | DNAase I |
| cg07212818 | Body | Island | 638076 | *-0.42* | 2.16E-06 | 0.41 | 0.08 | Y | Y | Y |
| cg05717871 | Body | Island | 638507 | *-0.42* | 2.23E-06 | 0.59 | 0.13 | Y | Y | Y |
| cg03909863 | Body | Island | 638404 | *-0.42* | 2.29E-06 | 0.87 | 0.11 | Y | Y | Y |
| cg11335335 | Body | Island | 637885 | *-0.40* | 5.63E-06 | 0.44 | 0.12 | Y | Y | Y |
| cg01616529 | Body | Island | 638424 | *-0.40* | 8.43E-06 | 0.52 | 0.13 | Y | Y | Y |
| cg09386376 | Body | Island | 638939 | *-0.32* | 3.57E-04 | 0.65 | 0.11 | Y | Y | Y |
| cg14433983 | TSS1500 | Island | 636460 | *-0.30* | 7.72E-04 | 0.71 | 0.08 | N | Y | N |
| cg06299284 | TSS1500 | Island | 636659 | *-0.26* | 3.70E-03 | 0.39 | 0.12 | Y | Y | N |
| cg03855291 | Body | Island | 639423 | *-0.26* | 4.13E-03 | 0.48 | 0.12 | Y | Y | Y |
| cg06825142 | TSS200 | Island | 637170 | *-0.23* | 0.01 | 0.09 | 0.02 | Y | Y | Y |
| cg23501406 | TSS1500 | Island | 637035 | *-0.20* | 0.03 | 0.06 | 0.02 | Y | Y | N |
| cg16029939 | Body | Island | 640328 | *-0.18* | 0.05 | 0.40 | 0.12 | Y | Y | Y |
| cg12928379 | TSS200 | Island | 637175 | *0.14* | 0.13 | 0.07 | 0.01 | Y | Y | Y |
| cg00556112 | TSS200 | Island | 637173 | *-0.14* | 0.14 | 0.12 | 0.02 | Y | Y | Y |
| cg02762115 | Body | Island | 640446 | *0.11* | 0.26 | 0.55 | 0.11 | Y | Y | Y |
| cg08726248 | TSS1500 | Island | 637032 | *-0.10* | 0.27 | 0.06 | 0.01 | Y | Y | Y |
| cg20931042 | TSS200 | Island | 637162 | *-0.09* | 0.34 | 0.08 | 0.01 | Y | Y | Y |
| cg09607276 | 1stExon | Island | 637491 | *-0.08* | 0.37 | 0.07 | 0.03 | Y | Y | N |
| cg23231692 | TSS1500 | N_Shore | 636167 | *0.07* | 0.46 | 0.84 | 0.03 | N | Y | N |
| cg15861585 | TSS1500 | Island | 637038 | *-0.03* | 0.78 | 0.07 | 0.02 | Y | Y | N |
| cg20411756 | 3'UTR | Island | 640606 | *0.00* | 0.98 | 0.72 | 0.13 | Y | Y | Y |

[a] Overlap between each CpG site annotated to the DRD4 gene and regulatory elements based on ENCODE data, including (i) transcription factors (data generated on 161 transcription factors in 91 cell types via ChIP-seq), (ii) histone marks (only relevant cell lines examined, including blood [GM12878], and umbilical vein endothelial [HUVEC] cells), and (iii) DNase I hypersensitivity clusters (based on data from 125 cell types). Y indicates that the CpG site coincides with the regulatory element, whereas N indicates no overlap.

**Table S2.** Cont'd

| CpG | Saliva-blood concordance (Smith et al., 2014)[b] | | Blood-CNS concordance 1 (Hannon et al., 2015)[c] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *T* | P-value | correlation >1 region | PFC | | STG | | EC | | CER | |
| | | | | *r* | P-value | *r* | P-value | *r* | P-value | *r* | P-value |
| cg07212818 | 10.38 | 2.07E-13 | Y | *0.69* | 6.93E-12 | *0.71* | 1.03E-12 | *0.60* | 2.53E-08 | *0.39* | 8.06E-04 |
| cg05717871 | 11.44 | 1.24E-14 | Y | *0.49* | 8.35E-06 | *0.49* | 7.25E-06 | *0.52* | 4.08E-06 | *0.16* | 0.19 |
| cg03909863 | 9.70 | 1.69E-12 | Y | *0.54* | 7.82E-07 | *0.60* | 1.37E-08 | *0.51* | 5.97E-06 | *0.13* | 0.30 |
| cg11335335 | 7.15 | 7.00E-09 | Y | *0.68* | 2.92E-11 | *0.46* | 5.36E-05 | *0.62* | 2.82E-09 | *0.44* | 1.38E-04 |
| cg01616529 | 9.10 | 1.13E-11 | Y | *0.64* | 1.08E-09 | *0.65* | 3.85E-10 | *0.70* | 3.88E-11 | *0.21* | 0.08 |
| cg09386376 | 7.34 | 3.58E-09 | Y | *0.53* | 1.41E-06 | *0.57* | 9.26E-08 | *0.47* | 3.36E-05 | *0.36* | 0.03 |
| cg14433983 | 14.44 | 2.62E-18 | Y | *0.49* | 7.79E-06 | *0.68* | 1.89E-11 | *0.67* | 1.64E-10 | *0.17* | 0.16 |
| cg06299284 | 16.72 | 1.10E-20 | Y | *0.40* | 4.05E-04 | *0.54* | 4.72E-07 | *0.55* | 9.16E-07 | *0.16* | 0.17 |
| cg03855291 | 7.01 | 1.12E-08 | Y | *0.28* | 0.02 | *0.48* | 1.30E-05 | *0.42* | 2.42E-04 | *0.34* | 3.96E-03 |
| cg06825142 | 0.37 | 0.71 | N | *-0.16* | 0.17 | *0.10* | 0.39 | *-0.08* | 0.48 | *0.15* | 0.20 |
| cg23501406 | 3.00 | 4.40E-03 | N | *-0.02* | 0.85 | *0.15* | 0.19 | *0.15* | 0.21 | *0.02* | 0.84 |
| cg16029939 | 1.56 | 0.13 | Y | *0.39* | 6.01E-04 | *0.40* | 4.46E-04 | *0.34* | 3.60E-03 | *0.37* | 1.75E-03 |
| cg12928379 | 0.32 | 0.75 | Y | *0.20* | 0.10 | *0.31* | 0.01 | *0.08* | 0.52 | *0.12* | 0.33 |
| cg00556112 | 1.41 | 0.16 | N | *0.00* | 0.99 | *-0.01* | 0.93 | *-0.18* | 0.14 | *-0.14* | 0.26 |
| cg02762115 | -0.01 | 0.99 | N | *0.14* | 0.22 | *0.11* | 0.36 | *-0.03* | 0.84 | *0.06* | 0.61 |
| cg08726248 | 2.67 | 0.01 | N | *-0.06* | 0.60 | *0.18* | 0.13 | *0.26* | 0.03 | *0.07* | 0.59 |
| cg20931042 | 1.47 | 0.15 | N | *-0.15* | 0.17 | *0.03* | 0.80 | *0.01* | 0.92 | *0.30* | 0.01 |
| cg09607276 | 3.82 | 4.15E-04 | N | *0.19* | 0.11 | *0.21* | 0.07 | *0.09* | 0.45 | *0.07* | 0.54 |
| cg23231692 | 11.32 | 1.29E-14 | N | *0.09* | 0.44 | *-0.01* | 0.93 | *0.11* | 0.36 | *-0.05* | 0.67 |
| cg15861585 | 1.95 | 0.06 | N | *0.05* | 0.68 | *0.12* | 0.30 | *0.06* | 0.61 | *0.19* | 0.12 |
| cg20411756 | 1.32 | 0.20 | N | *0.17* | 0.14 | *0.10* | 0.38 | *0.17* | 0.15 | *0.17* | 0.17 |

[b] Results extracted from Supplementary Table S1 of Smith et al (2014). T-values and associated p-values represent the output of regression models that were run for each CpG site in the Illumina 450k array, with DNAm levels in saliva entered as the predictor variable and DNAm levels in whole blood as the outcome variable, adjusting for technical covariates and proportion of epithelial cells in each sample.

[c] Results extracted from Hannon et al (2015), showing for each CpG site the correlation between DNAm levels in peripheral blood and that of CNS tissue across four brain regions (PFC: prefrontal cortex; STG: superior temporal gyrus; EC: entorhinal cortex; CER: cerebellum). For legibility, we also indicate whether each CpG shows significant blood-CNS concordence with at least one brain region, whereby Y indicates yes, and N indicates no.

**Table S2.** Cont'd

| CpG | Blood-CNS concordance 2 (Edgar et al., 2017)[d] | | | | | Association with mQTL (Gaunt et al., 2016)[e] | | |
|---|---|---|---|---|---|---|---|---|
| | BA10 | BA20 | BA7 | Total Brain | Total Brain | Cis/ Trans | N | SNP ID |
| | *r* | *r* | *r* | *r* | *%* | | | |
| cg07212818 | *0.36* | *0.43* | *0.44* | *0.41* | 90% | Cis | 11 | **rs150403728, rs35482931, rs61877858, rs79915070**, rs138188226, rs202011316, rs2740373, rs6421975, rs72841224, rs72844713, rs72844728 |
| cg05717871 | *0.58* | *0.61* | *0.60* | *0.60* | 90% | - | | - |
| cg03909863 | *0.45* | *0.60* | *0.41* | *0.49* | 90% | - | | - |
| cg11335335 | *0.13* | *0.38* | *-0.06* | *0.15* | 50-75% | Cis | 2 | **rs150403728,** rs2740373 |
| cg01616529 | *0.62* | *0.71* | *0.80* | *0.71* | 90% | Cis | 6 | **rs150403728, rs112888889, rs72844713, rs61877858,** rs2740373, rs61876662 |
| cg09386376 | *0.39* | *0.29* | *0.26* | *0.31* | 75-90% | Cis | 7 | **rs138188226, rs150403728, rs150403728, rs11246235, rs936472**, rs112888889, rs72844710 |
| cg14433983 | *0.26* | *0.29* | *0.11* | *0.22* | 50-75% | - | | - |
| cg06299284 | *0.30* | *0.51* | *0.52* | *0.44* | 90% | Cis | 5 | **rs2740375, rs7116535**, rs10902196, rs7928305, rs936472 |
| cg03855291 | *0.60* | *-0.01* | *0.35* | *0.31* | 75-90% | Cis | 4 | **rs150403728, rs11246235, rs7124601, rs936472** |
| cg06825142 | *0.39* | *0.10* | *0.12* | *0.20* | 50-75% | - | | - |
| cg23501406 | *-0.04* | *0.08* | *0.10* | *0.05* | <50% | - | | - |
| cg16029939 | *0.35* | *0.07* | *0.01* | *0.14* | 50-75% | - | | - |
| cg12928379 | *0.19* | *0.35* | *0.37* | *0.30* | 75-90% | - | | - |
| cg00556112 | *0.15* | *0.25* | *0.24* | *0.21* | 50-75% | - | | - |
| cg02762115 | *0.13* | *-0.35* | *-0.10* | *-0.11* | - | Cis | 4 | **rs150403728, rs117429558, rs936472**, rs55771159 |
| cg08726248 | *0.24* | *0.12* | *0.16* | *0.17* | 50-75% | | | |
| cg20931042 | *-0.18* | *-0.12* | *0.14* | *-0.05* | - | - | | - |
| cg09607276 | *0.07* | *0.01* | *-0.11* | *-0.01* | - | - | | - |
| cg23231692 | *-0.34* | *0.11* | *0.04* | *-0.06* | - | - | | - |
| cg15861585 | *0.45* | *0.11* | *-0.10* | *0.15* | 50-75% | - | | - |
| cg20411756 | *0.09* | *0.25* | *-0.13* | *0.07* | <50% | Cis | 1 | **rs117429558** |

[d] Results extracted from Edgar et al (2017), showing for each CpG site the correlation between DNAm levels in peripheral blood and that of CNS tissue across three brain regions (BA10: anterior-most portion of the prefrontal cortex; BA20: inferior temporal area; BA7: parietal cortex), as well as total brain (mean of these three regions). Effect sizes were extracted from the BECon online tool. Due to small sample size (n = 15), no p-values are provided. Instead, the ranking of each CpG site in terms of strength of blood-brain concordance is reported (as a percentile).

[e] Association between each CpG site and genetic mQTLs, based on GCTA analysis results extracted from the mQTLdb resource (Gaunt et al., 2016). Information includes whether the mQTL is cis (i.e. SNP within ±1000 base pairs of the DNAm site) or trans (i.e. ±1 million base pairs), the number of mQTLs associated with the CpG site and their SNP IDs (bolded if shown to associate with the CpG site at > 1 time point). The underlined SNP ID indicates the mQTL that was used in the Mendelian randomisation analysis (rs2740373).

**Table S3.** Sensitivity analyses of *DRD4* CpG loci

| | CpG information | | | Original analysis[a] | | Sensitivity analysis 1[b]: Winsorized data | | Sensitivity analysis 2[c]: SVA correction | | Sensitivity analysis 3[d]: Subsample with buccal cells as predicted tissue | | Sensitivity analysis 4[e]: Binary fights variable | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CpG ID | Genomic location | Proximity to CpG island | Position | *T* | P-value | *T* | P-value | *T* | P-value | *T* | P-value | *T* | P-value |
| cg07212818 | Body | Island | 638076 | *-4.99* | 2.16E-06 | *-4.99* | 2.16E-06 | *-4.32* | 3.53E-05 | -4,53 | 1,78E-05 | *-4.00* | 1.11E-04 |
| cg05717871 | Body | Island | 638507 | *-4.98* | 2.23E-06 | *-4.98* | 2.23E-06 | *-5.02* | 2.04E-06 | -4,27 | 4,68E-05 | *-4.13* | 6.72E-05 |
| cg03909863 | Body | Island | 638404 | *-4.97* | 2.29E-06 | *-4.80* | 4.79E-06 | *-4.49* | 1.81E-05 | -4,45 | 2,38E-05 | *-4.36* | 2.77E-05 |
| cg11335335 | Body | Island | 637885 | *-4.76* | 5.63E-06 | *-4.76* | 5.63E-06 | *-4.09* | 8.51E-05 | -4,33 | 3,85E-05 | *-4.18* | 5.74E-05 |
| cg01616529 | Body | Island | 638424 | *-4.66* | 8.43E-06 | *-4.66* | 8.43E-06 | *-4.29* | 3.96E-05 | -4,17 | 6,78E-05 | *-3.60* | 4.63E-04 |
| cg09386376 | Body | Island | 638939 | *-3.68* | 3.57E-04 | *-3.68* | 3.57E-04 | *-3.79* | 2.53E-04 | -3,09 | 2,63E-03 | *-3.30* | 1.27E-03 |
| cg14433983 | TSS1500 | Island | 636460 | *-3.45* | 7.72E-04 | *-3.46* | 7.59E-04 | *-3.41* | 9.16E-04 | -3,59 | 5,25E-04 | *-2.73* | 7.32E-03 |
| cg06299284 | TSS1500 | Island | 636659 | *-2.96* | 3.70E-03 | *-2.96* | 3.70E-03 | *-2.16* | 0.03 | -2,92 | 4,34E-03 | *-2.61* | 0.01 |
| cg03855291 | Body | Island | 639423 | *-2.93* | 4.13E-03 | *-2.93* | 4.13E-03 | *-2.87* | 4.90E-03 | -2,49 | 0,01 | *-2.31* | 0.02 |
| cg06825142 | TSS200 | Island | 637170 | *-2.53* | 0.01 | *-2.68* | 0.01 | *-1.19* | 0.24 | -2,30 | 0,02 | *-2.91* | 4.31E-03 |
| cg23501406 | TSS1500 | Island | 637035 | *-2.18* | 0.03 | *-2.22* | 0.03 | *-1.49* | 0.14 | -2,62 | 0,01 | *-1.88* | 0.06 |
| cg16029939 | Body | Island | 640328 | *-1.99* | 0.05 | *-1.99* | 0.05 | *-1.84* | 0.07 | -1,95 | 0,05 | *-2.37* | 0.02 |
| cg12928379 | TSS200 | Island | 637175 | *1.53* | 0.13 | *-1.52* | 0.13 | *0.15* | 0.88 | 1,66 | 0,10 | *1.73* | 0.09 |
| cg00556112 | TSS200 | Island | 637173 | *-1.49* | 0.14 | *1.51* | 0.13 | *-0.05* | 0.96 | -1,19 | 0,24 | *-2.33* | 0.02 |
| cg02762115 | Body | Island | 640446 | *1.14* | 0.26 | *1.14* | 0.26 | *2.19* | 0.03 | 1,39 | 0,17 | *0.13* | 0.90 |
| cg08726248 | TSS1500 | Island | 637032 | *-1.12* | 0.27 | *-1.10* | 0.27 | *-1.10* | 0.27 | -1,57 | 0,12 | *-0.51* | 0.61 |
| cg20931042 | TSS200 | Island | 637162 | *-0.96* | 0.34 | *-0.90* | 0.37 | *0.50* | 0.62 | -0,99 | 0,33 | *-0.34* | 0.74 |
| cg09607276 | 1stExon | Island | 637491 | *-0.90* | 0.37 | *-0.81* | 0.42 | *-0.74* | 0.46 | -0,88 | 0,38 | *-0.57* | 0.57 |
| cg23231692 | TSS1500 | N_Shore | 636167 | *0.74* | 0.46 | *0.68* | 0.50 | *0.09* | 0.93 | 0,99 | 0,32 | *0.01* | 0.99 |
| cg15861585 | TSS1500 | Island | 637038 | *-0.28* | 0.78 | *-0.12* | 0.91 | *-0.53* | 0.60 | -0,16 | 0,87 | *-0.49* | 0.62 |
| cg20411756 | 3'UTR | Island | 640606 | *0.03* | 0.98 | *0.03* | 0.98 | *0.64* | 0.53 | 0,16 | 0,87 | *-0.96* | 0.34 |

N.b. all analyses control for age, sex and ethnicity

[a] Results from the original analysis, associating non-winsorised DNA methylation values and engagement in physical fights (4-level ordinal variable, spanning 0-5+ times in the past year)

[b] Results from the first sensitivity analysis, associating winsorized DNA methylation values (threshold: 3*IQR) and engagement in physical fights.

[c] Results from the second sensitivity analysis, associating non-winsorized DNA methylation values and engagements in physical fights, additionally controlling for ten surrogate variables to minimize the potential influence of unwanted confounders.

[d] Results from the third sensitivity analysis, associating non-winsorized DNA methylation values and engagements in physical fights only in the subsample of participants with 'buccal' as predicted tissue, based on the Horvath algorithm (Horvath et al., 2013).

[e] Results from the fourth sensitivity analysis, associating non-winsorized DNA methylation values with a collapsed variable of engagements in physical fights (2-level binary variable, 0 = no fights, 1 = any number of fights)

**Table S4.** Cross-tissue correspondence in *DRD4* methylation

| | | **Strength of association with fights** (discovery sample) | **Saliva-Blood correspondence** (Smith et al., 2014) | **Blood-CNS correspondence 1** (Hannon et al., 2015) | | | | **Blood-CNS correspondence 2** (Edgar et al., 2017) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | PFC | STG | EC | CER | BA10 | BA20 | BA7 |
| **Strength of association with fights** | **r** | - | | | | | | | | |
| | *p* | | | | | | | | | |
| **Saliva-Blood correspondence** | **r** | **0.59** | - | | | | | | | |
| | *p* | *0.01* | | | | | | | | |
| **Blood-CNS correspondence 1:** | | | | | | | | | | |
| PFC | **r** | **0.69** | **0.65** | - | | | | | | |
| | *p* | *1.00E-03* | *0.001* | | | | | | | |
| STG | **r** | **0.77** | **0.69** | **0.90** | | | | | | |
| | *p* | *5.40E-05* | *0.001* | *3.61E-08* | - | | | | | |
| EC | **r** | **0.77** | **0.77** | **0.89** | **0.93** | | | | | |
| | *p* | *4.90E-05* | *4.40E-05* | *8.13E-08* | *2.11E-09* | - | | | | |
| CER | **r** | **0.49** | **0.15** | **0.55** | **0.55** | **0.57** | - | | | |
| | *p* | *0.02* | *0.52* | *0.01* | *0.01* | *0.01* | | | | |
| **Blood-CNS correspondence 2:** | | | | | | | | | | |
| BA10 | **r** | **0.34** | **0.34** | **0.53** | **0.47** | **0.34** | **-0.01** | - | | |
| | *p* | *0.13* | *0.13* | 0.01 | *0.03* | *0.13* | *0.96* | | | |
| BA20 | **r** | **0.54** | **0.17** | **0.26** | **0.44** | **0.28** | **0.02** | **0.49** | - | |
| | *p* | *0.01* | *0.47* | *0.25* | *0.05* | *0.21* | *0.93* | *0.03* | | |
| BA7 | **r** | **0.54** | **0.25** | **0.31** | **0.38** | **0.25** | **0.03** | **0.73** | **0.56** | **-** |
| | *p* | *0.01* | *0.27* | *0.17* | *0.09* | *0.27* | *0.89* | *1.85E-04* | *0.01* | |
| Brain (Mean) | **r** | **0.55** | **0.30** | **0.44** | **0.50** | **0.35** | **0.02** | **0.87** | **0.79** | **0.90** |
| | *p* | *0.01* | *0.18* | *0.05* | *0.02* | *0.13* | *0.94* | *2.77E-07* | *2.05E-05* | *3.22E-08* |

This table shows how indices of correspondence extracted from the three cross-tissue resources used in our study (Smith et al., 2014; Hannon et al, 2015; Edgar et al., 2017) relate to our findings and to one another, based on data from all DRD4 methylation sites (21 CpGs). From this table, two key findings emerge. First, there is a significant correlation between how strongly DRD4 sites associate with physical aggression in our sample and how much cross-tissue concordance they show across all three resources. In other words, DRD4 sites that showed stronger associations with physical aggression, also showed stronger cross-tissue concordance in saliva, blood and brain. Second, brain regions that are anatomically most comparable across resources (e.g. Hannon's prefrontal cortex and Edgar's BA10) also show the strongest convergence with one another.

**Figure S1.** Visualization of probe-level EWAS results

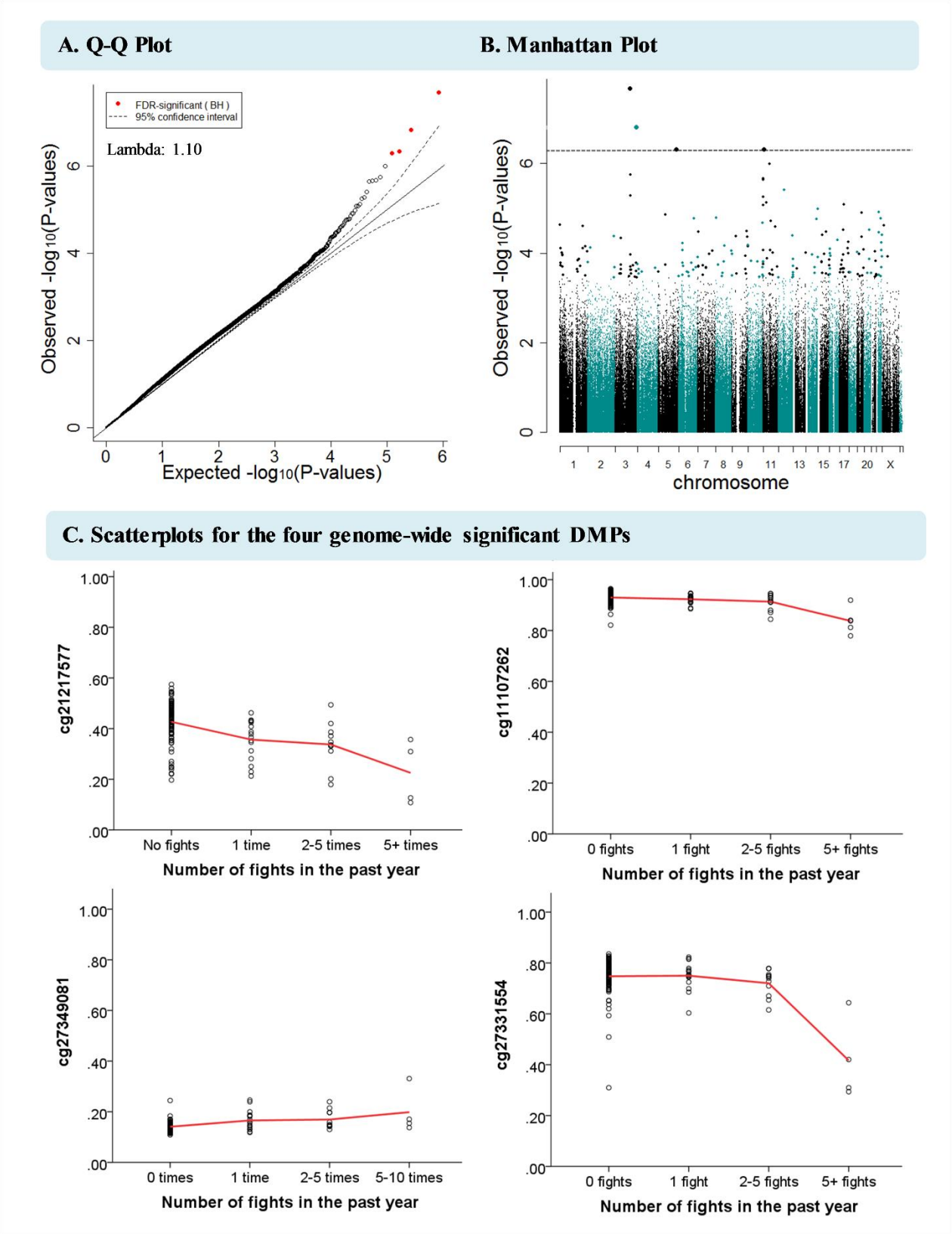
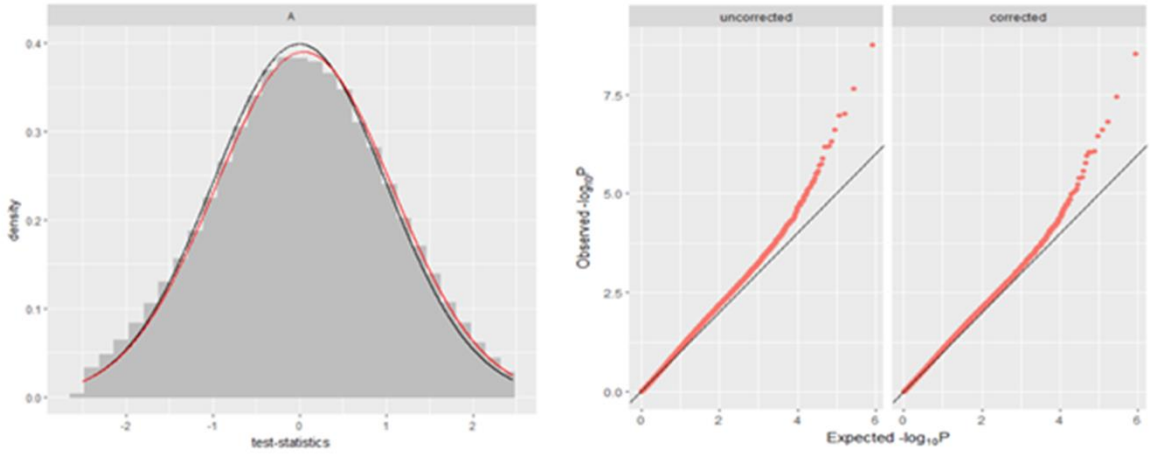


**Figure S2.** Bias- and inflation-corrected EWAS results using the Bacon R package

**A. Test statistic histogram** **B. Uncorrected vs corrected Q-Q plot**




**C. Bacon-corrected p-values for top 30 EWAS DMPs**

| CpG | Gene | Standardized Beta | p-value: Original | p-value: Bacon-corrected |
|---|---|---|---|---|
| cg21217577 | *TRH* | *-.49* | 2.07E-08 | 2.99E-09 |
| cg11107262 | *GAK* | *-.46* | 1.52E-07 | 3.57E-08 |
| cg27349081 | *RIC3* | *.44* | 4.80E-07 | 2.43E-07 |
| cg27331554 | *CYFIP2* | *-.44* | 5.18E-07 | 1.57E-07 |
| cg05082738 | *LINC00301* | *-.43* | 1.03E-06 | 3.55E-07 |
| cg15916166 | *FAIM* | *.42* | 1.79E-06 | 1.13E-06 |
| **cg07212818** | ***DRD4*** | ***-.42*** | **2.16E-06** | **8.56E-07** |
| **cg05717871** | ***DRD4*** | ***-.42*** | **2.23E-06** | **8.87E-07** |
| **cg03909863** | ***DRD4*** | ***-.42*** | **2.29E-06** | **9.15E-07** |
| cg08455772 | *FLJ12825* | *-.41* | 3.92E-06 | 1.71E-06 |
| cg22175303 | *FAIM* | *.40* | 5.20E-06 | 3.82E-06 |
| **cg11335335** | ***DRD4*** | ***-.40*** | **5.63E-06** | **2.60E-06** |
| cg13694680 | *FIBIN* | *.40* | 7.44E-06 | 5.73E-06 |
| cg05930207 | *GJD3* | *-.40* | 8.23E-06 | 4.03E-06 |
| **cg01616529** | ***DRD4*** | ***-.40*** | **8.43E-06** | **4.14E-06** |
| cg11481687 | - | *.39* | 1.05E-05 | 8.44E-06 |
| cg03205258 | *TXNRD2;COMT* | *.39* | 1.22E-05 | 9.95E-06 |
| cg02445292 | *FCGBP* | *.39* | 1.25E-05 | 1.02E-05 |
| cg04580929 | *PDE4D* | *-.39* | 1.39E-05 | 7.36E-06 |
| cg00996790 | - | *-.38* | 1.63E-05 | 8.76E-06 |
| cg10105971 | *SGK1* | *-.38* | 1.68E-05 | 9.11E-06 |
| cg11153328 | *TXN2* | *.38* | 1.69E-05 | 1.43E-05 |
| cg10096566 | - | *-.38* | 1.74E-05 | 9.46E-06 |
| cg14234104 | *ANO1* | *.38* | 1.92E-05 | 1.66E-05 |
| cg13822329 | - | *.38* | 2.10E-05 | 1.83E-05 |
| cg26312965 | *SDF4* | *-.38* | 2.36E-05 | 1.33E-05 |
| cg10734526 | *GPM6B* | *.38* | 2.40E-05 | 2.12E-05 |
| cg01836910 | *CDK18* | *-.38* | 2.47E-05 | 1.41E-05 |
| cg06674932 | *FDX1* | *.38* | 2.62E-05 | 2.33E-05 |
| cg23720732 | *SLC12A5* | *.37* | 3.00E-05 | 2.71E-05 |

The results from the Bacon analysis indicated minimal bias and inflation in out probe-specific EWAS (Bayesian estimates of bias = .049; Bayesian inflation factor = 1.023). This is also visually represented in the histogram **(A)**, which indicates only a very small bias towards the right (red line) as well as the QQ plot **(B)**, which shows minimal change after correction. **(C)** shows the top 30 CpG loci from our EWAS, with the addition of bacon-corrected p-values. All associations became stronger after correction (including all sites in our DRD4 DMR, showed in bold), indicating that potentially confounding effects biased our initial effects towards the null.

**Figure S3.** Scatterplots of the five *DRD* 4CpG sites contained in the genome-wide significant DMR

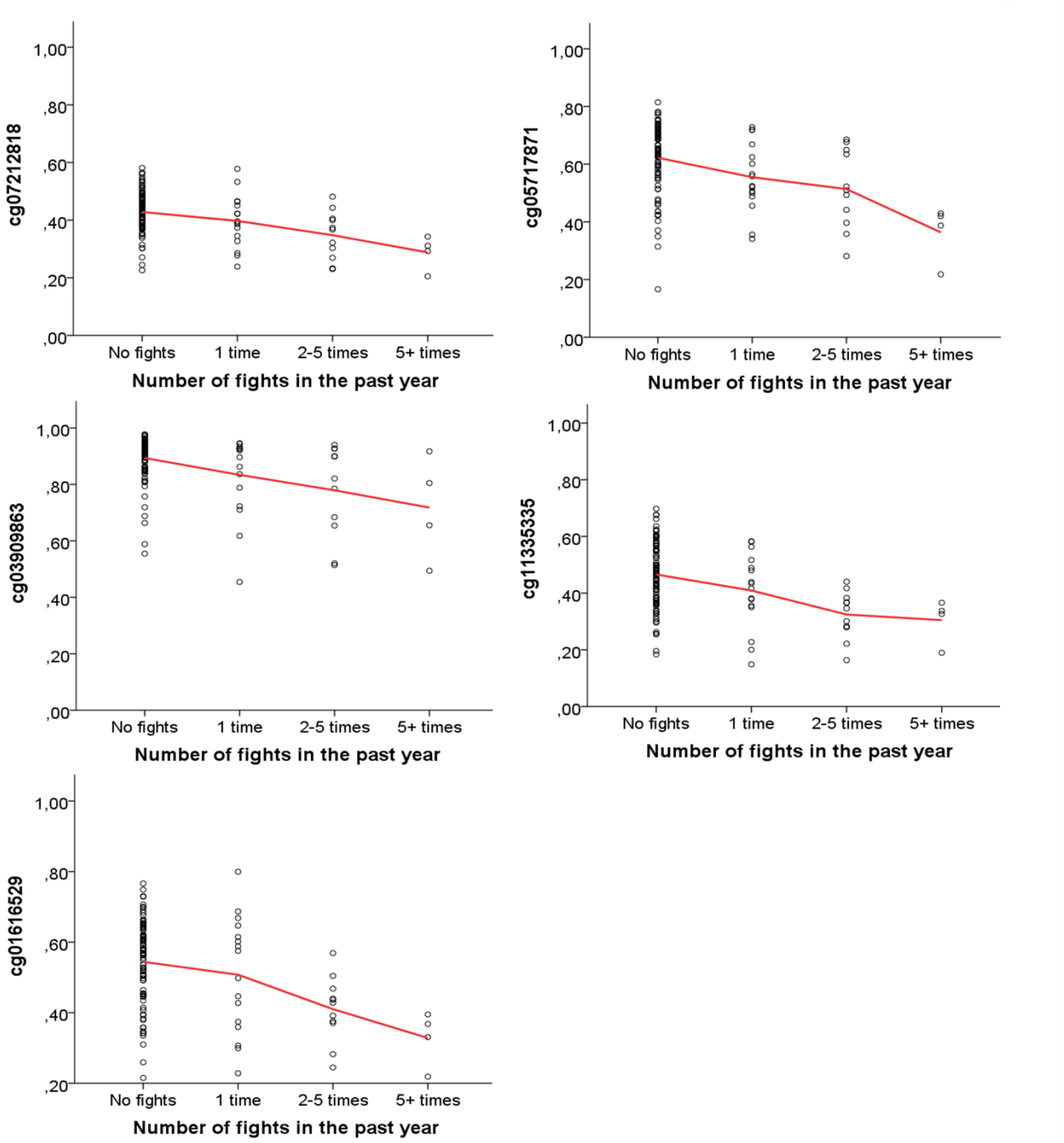


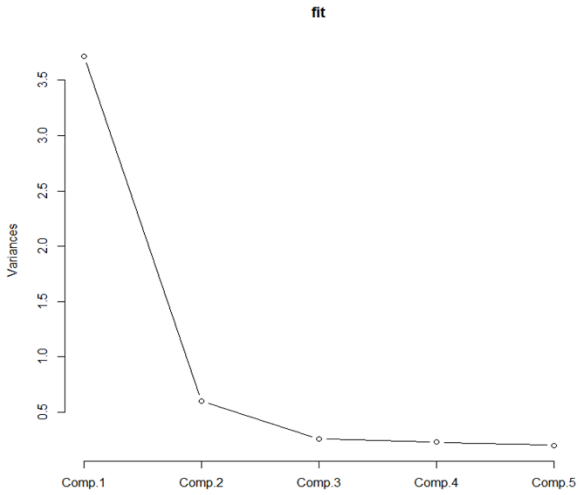
**Figure S4.** PCA analysis of the 5 CpG sites contained in the *DRD4* DMR (R *princomp* package)

**Importance of components:**

| | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|---|---|---|---|---|---|
| Eigenvalue | 3.72 | 0.60 | 0.26 | 0.23 | 0.20 |
| Standard deviation | 1.93 | 0.77 | 0.51 | 0.48 | 0.45 |
| Proportion of variance (%) | 74% | 12% | 5% | 5% | 4% |
| Cumulative % | 74% | 86% | 91% | 96% | 100% |

**Loadings:**

| CpG sites | Comp.1 |
|---|---|
| *cg07212818* | 0.89 |
| *cg05717871* | 0.87 |
| *cg03909863* | 0.90 |
| *cg11335335* | 0.72 |
| *cg01616529* | 0.91 |

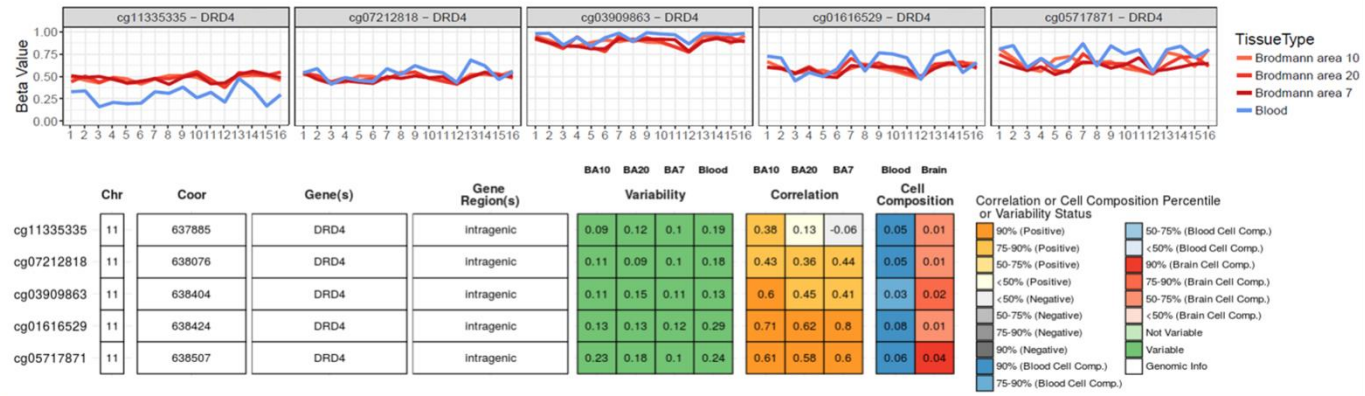



**Spearman correlations with engagement in physical fights**

| Components | Number of fights |
|---|---|
| *Comp.1* | -0.39 |
| *Comp.2* | -0.08 |
| *Comp.3* | 0.04 |
| *Comp.4* | -0.01 |
| *Comp.5* | -0.10 |

**Figure S5.** *DRD4* methylation and expression across peripheral and central nervous system (CNS) tissue

**A.** Variability and blood-CNS concordance across the five *DRD4* sites contained in the DMR based on data from Edgar et al (2017)

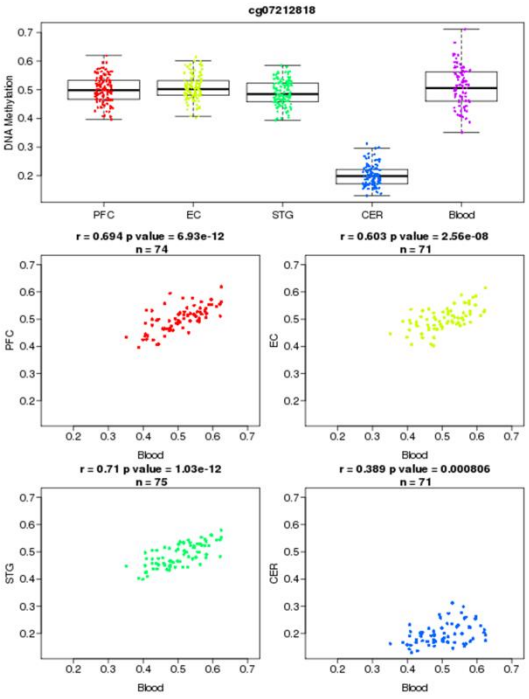



| | Chr | Coor | Gene(s) | Gene Region(s) | Variability BA10 | Variability BA20 | Variability BA7 | Variability Blood | Correlation BA10 | Correlation BA20 | Correlation BA7 | Cell Composition Blood | Cell Composition Brain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cg11335335 | 11 | 637885 | DRD4 | intragenic | 0.09 | 0.12 | 0.1 | 0.19 | 0.38 | 0.13 | -0.06 | 0.05 | 0.01 |
| cg07212818 | 11 | 638076 | DRD4 | intragenic | 0.11 | 0.09 | 0.1 | 0.18 | 0.43 | 0.36 | 0.44 | 0.05 | 0.01 |
| cg03909863 | 11 | 638404 | DRD4 | intragenic | 0.11 | 0.15 | 0.11 | 0.13 | 0.6 | 0.45 | 0.41 | 0.03 | 0.02 |
| cg01616529 | 11 | 638424 | DRD4 | intragenic | 0.13 | 0.13 | 0.12 | 0.29 | 0.71 | 0.62 | 0.8 | 0.08 | 0.01 |
| cg05717871 | 11 | 638507 | DRD4 | intragenic | 0.23 | 0.18 | 0.1 | 0.24 | 0.61 | 0.58 | 0.6 | 0.06 | 0.04 |

Correlation or Cell Composition Percentile or Variability Status

90% (Positive)
75-90% (Positive)
50-75% (Positive)
<50% (Positive)
<50% (Negative)
50-75% (Negative)
75-90% (Negative)
90% (Negative)
90% (Blood Cell Comp.)
75-90% (Blood Cell Comp.)
50-75% (Blood Cell Comp.)
<50% (Blood Cell Comp.)
90% (Brain Cell Comp.)
75-90% (Brain Cell Comp.)
50-75% (Brain Cell Comp.)
<50% (Brain Cell Comp.)
Not Variable
Variable
Genomic Info

**B.** Blood-CNS concordance for *DRD4* probe cg07212818 based on data from Hannon et al (2015)




**C.** *DRD4* expression levels across peripheral and central tissues based on GTEx data

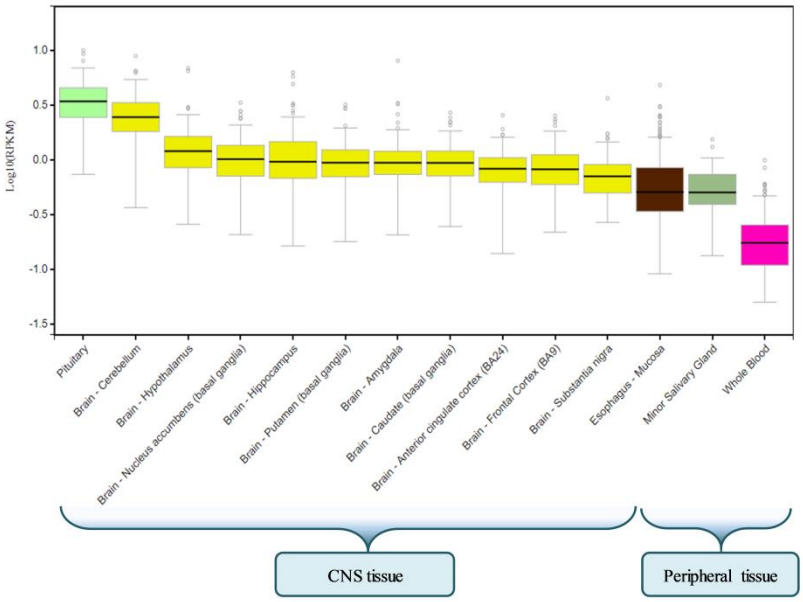
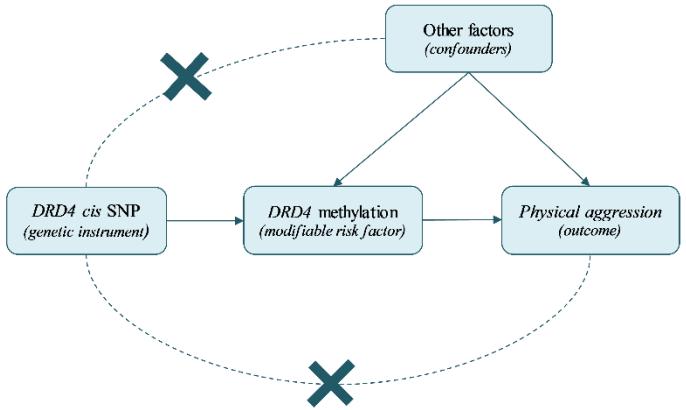


**Figure S6.** Mendelian randomization analysis

**A.** Schematic representation of the MR assumptions




**B.** MR analysis results

**Step 1.** Identifying a genetic instrument for *DRD4* methylation using the mQTLdb database (http://www.mqtldb.org)

| | | | | | | | | output | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **SNP Chr** | **SNP Pos** | **A1** | **A2** | **MAF** | **Trans** | **CpG** | **CpG Chr** | **CpG Pos** | **Timepoint** | **beta** | **t-stat** | **effect size** | **p-value** | **se** |
| rs2740373 | 11 | 621634 | A | G | 0.12 | N | cg07212818 | 11 | 638076 | Adolescence | -0.732 | -11.630 | 0.020 | 4.31E-29 | 0.063 |
| | | | | | | | cg11335335 | 11 | 637885 | Adolescence | -0.539 | -8.334 | 0.005 | 3.19E-16 | 0.065 |
| | | | | | | | cg01616529 | 11 | 638424 | Adolescence | -0.662 | -10.540 | 0.113 | 1.88E-24 | 0.063 |

**Step 2.** Extracting summary statistics for the genetic instrument, based on published GWAS data (Pappa et al., 2015)

| **Marker Name** | **Allele1** | **Allele2** | **Freq1** | **Zscore** | **se** | **p-value** | **Sample Size** |
|---|---|---|---|---|---|---|---|
| rs2740373 | a | g | 0.1074 | 0.065 | 0.44375 | 0.9478 | 18832 |

**Step 3.** Running the MR analysis using the TwoSample MR package in R

| **Outcome** | **Exposure** | **Method** | **nsnp** | **b** | **se** | **95% CI** | | **p-value** |
|---|---|---|---|---|---|---|---|---|
| Aggression | cg07212818 | Wald ratio | 1 | -0.089 | 0.606 | -1.277 | 1.100 | 0.884 |
| | cg01616529 | Wald ratio | 1 | -0.098 | 0.671 | -1.413 | 1.216 | 0.884 |
| | cg11335335 | Wald ratio | 1 | -0.121 | 0.824 | -1.736 | 1.494 | 0.884 |

**A.** Schematic representation of the MR variable assumptions, including that (i) the genetic instrument (i.e. *DRD4* SNP) should have an effect on methylation levels (solid line), and (ii) that the genetic instrument should not be associated with the outcome (only through the risk factor, i.e. DNAm) or other correlated factors (i.e. confounders; dotted line)

**B.** B. Results of the MR analysis, including (i) the mQTLdb output identifying a genetic instrument for *DRD4* methylation within our genome-wide significant DMR; (ii) the summary statistics for the genetic instrument, based on the EAGLE Consortium GWAS on physical aggression; and (iii) the results of the MR analysis, which was run using the TwoSampleMR package in R, available as part of the MR-Base platform.