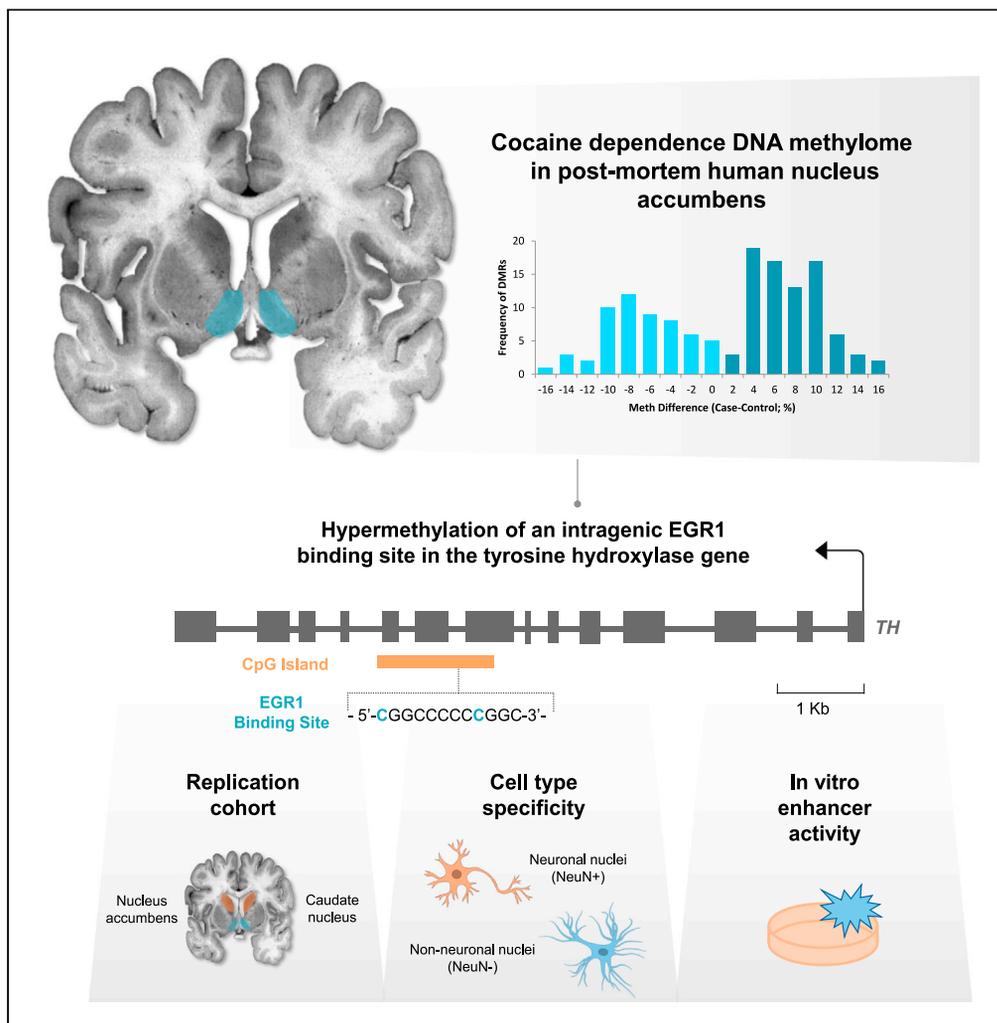


Article

# Methylation of the tyrosine hydroxylase gene is dysregulated by cocaine dependence in the human striatum



Kathryn Vaillancourt, Gang G. Chen, Laura Fiori, ..., Naguib Mechawar, Deborah C. Mash, Gustavo Turecki

gustavo.turecki@mcgill.ca

Highlights

Chronic cocaine dependence alters DNA methylation in human nucleus accumbens

The *TH* gene contains a binding site for EGR1, a cocaine-induced DNA binding protein

The EGR1 binding site is hypermethylated after chronic cocaine in striatal neurons

This region has enhancer activity that is responsive to EGR1 and methylation levels

Vaillancourt et al., iScience 24, 103169  
October 22, 2021 © 2021 The Authors.  
<https://doi.org/10.1016/j.isci.2021.103169>



## Article

## Methylation of the tyrosine hydroxylase gene is dysregulated by cocaine dependence in the human striatum

Kathryn Vaillancourt,<sup>1,2</sup> Gang G. Chen,<sup>1</sup> Laura Fiori,<sup>1</sup> Gilles Maussion,<sup>3</sup> Volodymyr Yerko,<sup>1</sup> Jean-François Thérault,<sup>1</sup> Carl Ernst,<sup>7</sup> Benoit Labonté,<sup>4</sup> Erin Calipari,<sup>5</sup> Eric J. Nestler,<sup>6</sup> Corina Nagy,<sup>1,7</sup> Naguib Mechawar,<sup>1,7</sup> Deborah C. Mash,<sup>8</sup> and Gustavo Turecki<sup>1,2,7,9,\*</sup>

## SUMMARY

**Cocaine dependence is a chronic, relapsing disorder caused by lasting changes in the brain. Animal studies have identified cocaine-related alterations in striatal DNA methylation; however, it is unclear how methylation is related to cocaine dependence in humans. We generated methylomic profiles of the nucleus accumbens using human postmortem brains from a cohort of individuals with cocaine dependence and healthy controls (n = 25 per group). We found hypermethylation in a cluster of CpGs within the gene body of tyrosine hydroxylase (TH), containing a putative binding site for the early growth response 1 (EGR1) transcription factor, which is hypermethylated in the caudate nucleus of cocaine-dependent individuals. We replicated this finding and found it to be specific to striatal neuronal nuclei. Furthermore, this locus demonstrates enhancer activity which is attenuated by methylation and enhanced by EGR1 overexpression. These results suggest that cocaine dependence alters the epigenetic regulation of dopaminergic signaling genes.**

## INTRODUCTION

Chronic cocaine exposure — and the behavioral symptoms that accompany the cocaine dependence phenotype — are associated with long lasting changes in brain biology. Long-term dysregulation of signaling in downstream targets of midbrain dopamine projections accompanies compulsive drug seeking behavior and withdrawal (Nestler and Lüscher, 2019; Tomasi et al., 2010; Volkow et al., 2006). Accordingly, regions of the mesocorticolimbic neurocircuit, including the dorsal (caudate nucleus) and ventral (nucleus accumbens) striatum, have been actively studied in humans and animal models. In humans, serial connectivity between the caudate and accumbens increases as patients transition to an addicted state, and widespread transcriptomic changes are associated with chronic exposure paradigms in rodents (Albertson et al., 2004; Belin and Everitt, 2008; Walker et al., 2018). Epigenetic mechanisms, including histone post-translational modifications and cytosine methylation, have been implicated as mediators of the impact of these relationships, and DNA methylation in particular has been shown to be necessary for the development of drug seeking behaviors (LaPlant et al., 2010).

Recently, we have shown that chronic cocaine dependence in humans is associated with genome-wide alterations in DNA methylation within the caudate nucleus (Vaillancourt et al., 2020). In the present study, we complement these findings with methylome-wide profiles of cocaine dependence in human postmortem nucleus accumbens and provide evidence for epigenetic regulation of the tyrosine hydroxylase (TH) gene. We have identified a region within TH, which contains portions of exons 8 and 9, as well as the intron between them, which is hypermethylated in both striatal subregions and is modulated by changes in methylation and transcription factor availability in vitro. This is of particular neurobiological interest because although TH-expressing cell bodies are thought to reside predominantly in the midbrain, recent evidence suggests that up to 3% of cell bodies within the anterior striatum express the gene (Savell et al., 2020).

<sup>1</sup>McGill Group for Suicide Studies, Douglas Hospital Research Center, Verdun, QC, Canada

<sup>2</sup>Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada

<sup>3</sup>Department of Neurology and Neurosurgery, Montreal Neurological Institute, Montreal, QC, Canada

<sup>4</sup>Centre de Recherche Cervo, Université Laval, Québec, QC, Canada

<sup>5</sup>Departments of Pharmacology, Molecular Physiology and Biophysics, Psychiatry and Behavioral Sciences; Vanderbilt Center for Addiction Research; Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA

<sup>6</sup>Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>7</sup>Department of Psychiatry, McGill University, Montreal, QC, Canada

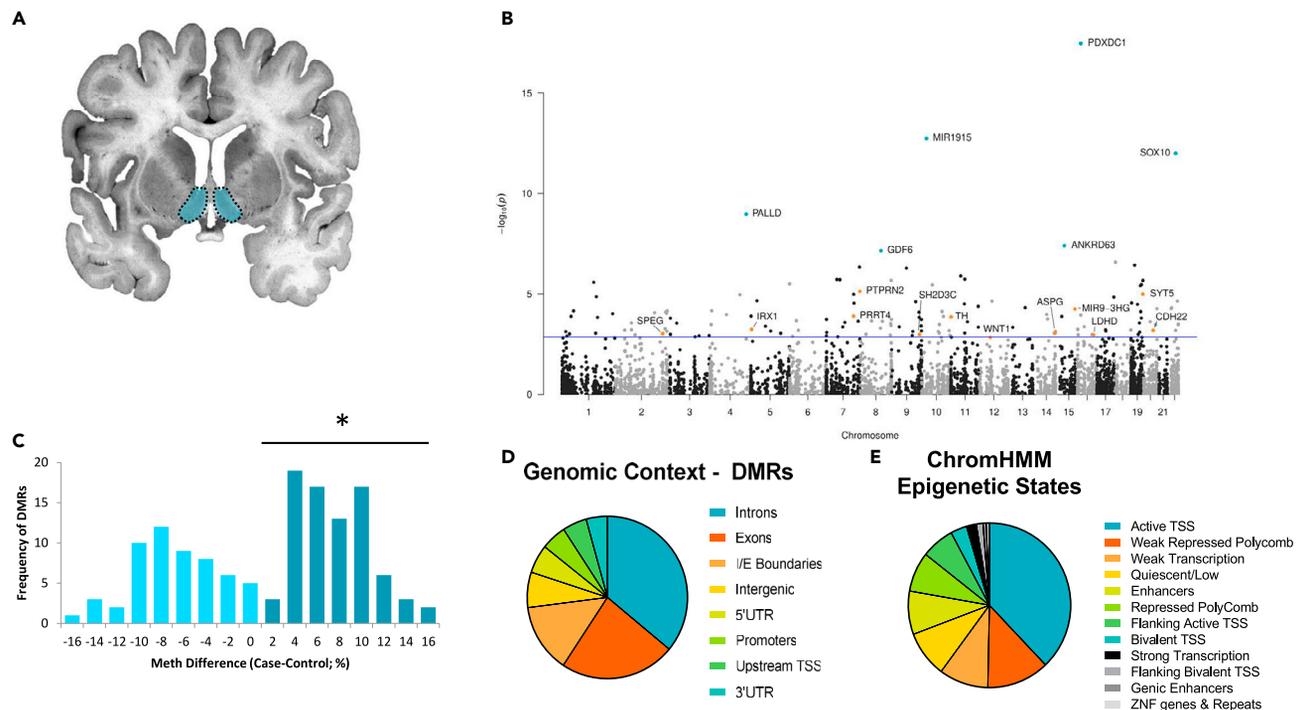
<sup>8</sup>Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA

<sup>9</sup>Lead contact

\*Correspondence:

gustavo.turecki@mcgill.ca  
<https://doi.org/10.1016/j.isci.2021.103169>





**Figure 1. Cocaine dependence is associated with widespread alterations in DNA methylation in the human nucleus accumbens**  
(A) Nucleus accumbens tissue from 25 individuals with a history of cocaine dependence and 25 drug-free controls was used to generate RRBS methylation data.

(B) Differentially methylated regions (DMRs) were distributed across autosomes, with the most significantly different regions indicated by blue dots, and those which overlapped with data from the caudate nucleus in orange.

(C) DMRs are significantly skewed towards hypermethylation. chi-squared Goodness of Fit  $*p < 0.05$

(D) are found primarily within gene bodies.

(E) and are enriched in loci that overlap with active TSSs. TSS = transcription start site, UTR = untranslated region, I/E = intron/exon, ZNF = zinc finger,

## RESULTS

### DNA methylation profile of cocaine dependence in human nucleus accumbens

We investigated nucleus accumbens tissue (Figure 1A) dissected from 25 individuals with chronic cocaine dependence and 25 psychiatrically healthy, drug-naïve controls (Table S1), and performed reduced representation bisulfite sequencing (RRBS) to identify methylome-wide alterations associated with case status (Chen et al., 2014). This strategy resulted in similar sequencing depth, bisulfite conversion rates, and CpG coverage between groups (Table S2). Our analysis identified 4814 clusters of CpGs that passed filtering, including 145 differentially methylated regions (DMRs), with a Benjamini-Hochberg corrected q-value  $< 0.05$  when comparing cases and controls and after controlling for covariates such as ethnicity (Table S3). The majority of DMRs were located within 5 kb of an annotated gene (113; 77.9%). Although DMRs were distributed across all autosomes (Figure 1B) and varied in cluster size and strength of differential methylation, there were significantly more hypermethylated DMRs than hypomethylated (chi-squared Goodness of Fit = 4.235;  $p < 0.05$ ; Figure 1C and Table 1), which is analogous to our previous findings in the dorsolateral caudate nucleus (Vaillancourt et al., 2020) and is consistent with animal studies of cocaine self-administration (Baker-Andresen et al., 2015). Next, we overlapped our list of DMRs with data sets containing known genomic and epigenomic features and found that all identified clusters, including those which were differentially methylated, overlapped primarily with CpG islands (data not shown) and intragenic regions, including introns, exons, and intron-exon boundaries (Figures S2A; 1D). Interestingly, when we compared our dataset to ChromHMM predicted epigenome states, we found a significant enrichment of regions overlapping with active transcription start sites in the DMRs compared to the list of all clusters (FDR  $< 0.05$ ; Figures 1E and S2B). Finally, we searched for potential functional commonality between the genes nearest to the DMRs using PANTHER gene ontology analysis (Thomas et al., 2003). Although it is

**Table 1. Top hypermethylated region in the nucleus accumbens**

Chromosome	From	To	Nearest gene	Methylation difference (%)
10	21789249	21789295	MIR1915	15.93
1	147718180	147718225	NBPF8	15.22
11	3244384	3244429	MRGPRE	14.33
19	39798143	39798241	LRFN1	13.41
19	55944580	55944625	SHISA7	13.17
19	44038525	44038544	ZNF575	12.28
13	79183633	79183659	RNF219	11.85
19	55685105	55685146	SYT5	11.14
9	138148440	138148475	LOC401557	10.88
7	149489530	149489549	SSPO	10.50
2	240241219	240241263	HDAC4	10.40
10	91294442	91294456	SLC16A12	10.11
2	74875158	74875201	M1AP	9.82
7	128556152	128556192	KCP	9.71
4	169798988	169799036	PALLD	9.46
5	28927814	28927856	LSP1P3	9.26
11	2188129	2188165	TH	9.17
15	89922099	89922217	MIR9-3HG	9.16
11	65352989	65353032	EHBP1L1	8.79
7	2660963	2661007	IQCE	8.77

understood that the nearest gene to a regulatory site is not always its primary target (Schoenfelder and Fraser, 2019), we still found the genes nearest to the DMRs to be enriched for DNA binding proteins and transcription factors (FE = 5.74; FDR <0.05).

Since chronic cocaine seeking in animals is associated with changes in the expression of DNA methyltransferase genes (LaPlant et al., 2010; Tian et al., 2012; Wright et al., 2015), we first investigated the expression of these genes. We found that *DNMT3A* was significantly increased in the nucleus accumbens of the cocaine group compared with controls ( $t = 2.96$ ,  $df = 41$ ,  $p = 0.005$ ; Figure S1A), as was the case in the caudate nucleus of the same individuals (Vaillancourt et al., 2020). The expression of the other two methyltransferase genes, *DNMT3B* and *DNMT1*, was not associated with case status ( $p_s > 0.05$ ; Figures S1B and S1C).

Notably, a cluster of CpGs within the body of the *TH* gene was among the top most hypermethylated loci in our analysis (9.17%, Table 1). Since *TH* is a necessary component of dopamine synthesis and has a known relationship with reward-driven behaviors (Kaminer et al., 2019; Logan et al., 2019) we were immediately struck by its potential importance. Interestingly, *TH* was also identified among the short list of DMRs that were shared between the current study and our previous work in the caudate nucleus (orange data points Figures 1B and Table 2), where the exact same cluster of CpGs was differentially methylated in the same direction, and similar to a magnitude (9.09%). Notably, up to 3% of GABAergic interneurons within the striatum express *TH*, and although these cells do not appear to be dopaminergic, they have been shown to play a role in motivational behavior in rodents (Ibáñez-Sandoval et al., 2010; Kaminer et al., 2019; Savell et al., 2020). The *TH* DMR is comprised of 3 CpGs within a CpG island in the eighth and ninth exon of *TH*, and we chose to further investigate this locus in the context of human cocaine neurobiology.

### ***TH* hypermethylation is related to cocaine dependence in the caudate and accumbens**

In order to validate and extend our findings, we performed deep bisulfite amplicon sequencing (Chen et al., 2017) on separate aliquots of genomic DNA from a larger sample set including the discovery cohort, and again found increased methylation to be associated with cocaine dependence in the caudate

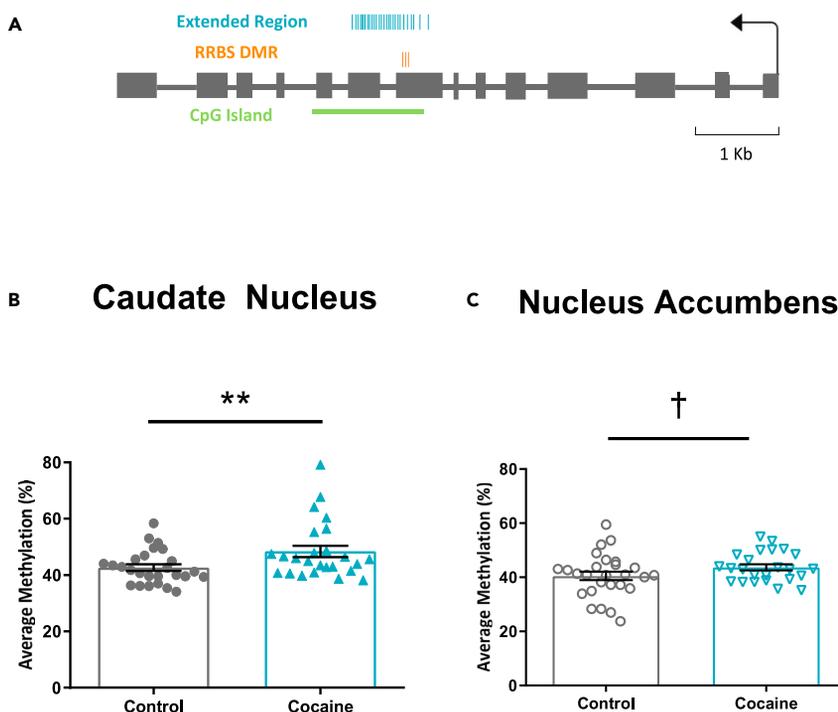
**Table 2. Differentially methylated regions and genes in common between the nucleus accumbens and the caudate nucleus**

Chromosome	From	To	Gene	Nucleus accumbens methylation difference (%)	Caudate nucleus methylation difference (%)
<b>Common DMRs</b>					
2	220313302	220313338	SPEG	-11.5	8.91
11	2188129	2188165	TH	9.17	9.09
16	15083748	15083904	PDXDC1	6.02	3.07
16	75148495	75148543	LDHD	4.32	-6.65
19	55685105	55685146	SYT5	11.1	3.59
<b>Common genes</b>					
5	3599609	3599704	IRX1		10.7
	3605955	3605982		8.56	
7	157484891	157484931	PTPRN2		-3.61
	157478003	157478040		-8.40	
7	127991684	127991788	PRRT4		-1.72
	127991615	127991788		4.75	
9	130517323	130517347	SH2D3C		5.99
	130517429	130517456		4.82	
12	49373486	49373526	WNT1		6.85
	49374220	49374248		-10.01	
14	104561244	104561286	ASPG		6.29
	104552290	104552330		-9.54	
15	89921805	89921837	MIR9-3HG		9.22
	89922099	89922217		9.16	
20	44879878	44879903	CDH22		9.00
	44803094	44803290		2.84	

(Mann-Whitney test,  $p < 0.05$ , Figure S3A), with no significant difference between groups in the nucleus accumbens ( $p > 0.05$ , Figure S3B). Furthermore, since methylation of individual CpGs within an island tends to be related (Barrera and Peinado, 2012), we decided to extend the amplicon beyond the original 3 CpGs, and interrogated the methylation status of an extended group of 27 cytosines within the intragenic CpG island (Figure 2A). Here, we found 5.8% more methylation across all CpGs in the caudate nucleus in the cocaine group (Mann-Whitney test,  $p < 0.01$ ; Figure 2B), and a similar trend in the nucleus accumbens with a 3.2% increase in methylation (Mann-Whitney test,  $p = 0.058$ ; Figure 2C). We generated RRBS data from the striatum of mice who were trained to self-administer cocaine and used it to compare the methylation status of a homologous cluster of 11 CpGs across the seventh exon, and the seventh and eighth exon of *Th* (87.9% homology; Figure S4A). We found that chronic cocaine exposure and seeking was related to 4.8% more methylation at this region within the nucleus accumbens ( $t = 4.115$ ,  $df = 16$ ,  $p < 0.001$ , Figure S4B), and no significant increase in the caudate-putamen ( $p > 0.05$ , Figure S4C).

### **TH hypermethylated region includes an EGR1 binding motif, and is specific to neuronal nuclei**

Next, to investigate the potential functional implications of increased methylation at our DMR, we used published chromatin immunoprecipitation data (ENCODE, (Bernstein et al., 2012)) to search for DNA binding proteins. One such factor, early growth response protein 1 (EGR1), was found to bind within the *TH* gene in multiple cell models and has a putative consensus motif (5'-CGGCCCCCCGGC-3', Figures S5; 3A, (Zandarashvili et al., 2015)) within the cocaine-associated DMR. Importantly, EGR1 has been heavily implicated in cocaine-related behaviors in animals, including cocaine conditioned place preference and context-related drug seeking (Hearing et al., 2008; Valjent et al., 2006). The potential binding site within *TH* contains CpGs at both ends (at hg 19\_ch11:2188079 and 2188087), which demonstrate the expected



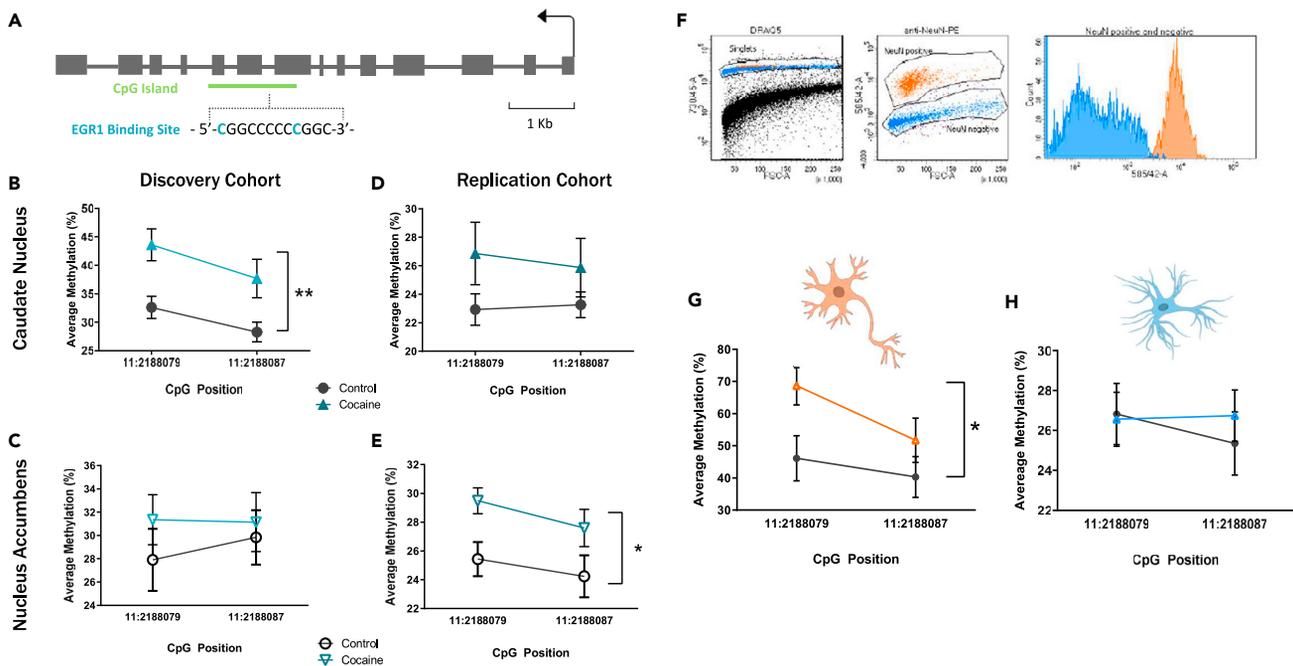
**Figure 2. Exon 8/9 of TH contains a CpG island that is hypermethylated in the cocaine group**

(A) Schematic diagram of the tyrosine hydroxylase gene containing the position of CpGs identified in the methylome wide analysis (orange, RRBS DMR), as well as those which were sequenced in follow-up experiments (blue, Extended Region). (B and C) Average methylation of the extended region is significantly higher in the cocaine group ( $n = 25$ ) compared to controls ( $n = 26$ ) in the caudate nucleus, and (C) there is a trend towards higher methylation in the cocaine group ( $n = 24$ ) compared to controls ( $n = 27$ ) in the nucleus accumbens. Data represented as mean  $\pm$  SEM, Mann-Whitney U or Student's t-tests \*\* $p < 0.01$ , † $p = 0.058$

hypermethylation in the caudate nucleus (two-way repeated measures ANOVA, main effect of Group,  $F(1,49) = 8.914$ , Sidak's multiple comparisons  $p < 0.01$ , Figure 3B), but not in the nucleus accumbens (two-way repeated measures ANOVA,  $F(1,49) = 0.65$ ,  $p > 0.05$ , Figure 3C).

We then sought to strengthen our findings through replication and investigated tissue from an independent cohort of psychiatrically healthy controls and individuals who died with chronic cocaine dependence ( $n = 18$  per group, Table S4). Using an identical bisulfite amplicon sequencing strategy, we found no differences in the caudate nucleus and significantly higher methylation across the putative EGR1 site in the nucleus accumbens in this cohort (two-way repeated ANOVAs,  $F_s(1,33) = 2.292, 5.745$   $p_s = 0.1399, 0.0224$ , respectively, Figures 3D and E). Although the variability within each dataset reduced our power to detect significant differences, there is a trend toward increased methylation of both CpGs within the EGR1 binding site between two brain regions and across both cohorts.

DNA methylation patterns are cell type-specific and are related to distinct cellular functions (Kozlenkov et al., 2015; Lister et al., 2013; Luo et al., 2017), so we used fluorescence-activated nuclei sorting (FANS) to separate neuronal from non-neuronal nuclei in new dissections of caudate and accumbens tissue from the discovery cohort. We separated intact nuclei from cellular debris using the DRAQ5 DNA stain, and neuronal from non-neuronal nuclei based on fluorescence with the NeuN neuronal marker (Figure 3F). We hypothesized that DNA methylation within the flanking CpGs of the EGR1 binding site would be significantly higher in neuronal nuclei from the cocaine group than in controls, and that this effect would be absent in non-neuronal nuclei. Indeed, in the caudate nucleus, we observed increased methylation in neuronal, but not non-neuronal, nuclei from the cocaine group (Two-way repeated measures ANOVAs,  $F_s(1,40) = 0.75, 0.08$  and  $p_s > 0.05$  Figures 3G and 3H). We found no effect of group on DNA methylation in either nuclear fraction in the nucleus accumbens (Two-way repeated measures ANOVAs,  $F_s(1,40) = 7.410, 0.1581$  and  $p_s = 0.210, 0.762$ ; Figure S6).



**Figure 3. Striatal hypermethylation of CpGs within an EGR1 binding site is specific to neurons in the caudate nucleus**

(A–E) The intragenic CpG island in *TH* contains a potential EGR1 binding site containing 2 flanking CpGs that are (B) hypermethylated within the caudate nucleus of the discovery cohort and (E) the nucleus accumbens of the replication cohort ( $n = 17$  cases and 18 controls). (C) There is a non-significant trend towards hypermethylation within the nucleus accumbens of the discovery cohort and (D) caudate nucleus of the replication cohort ( $n = 16$  cases and 18 controls).

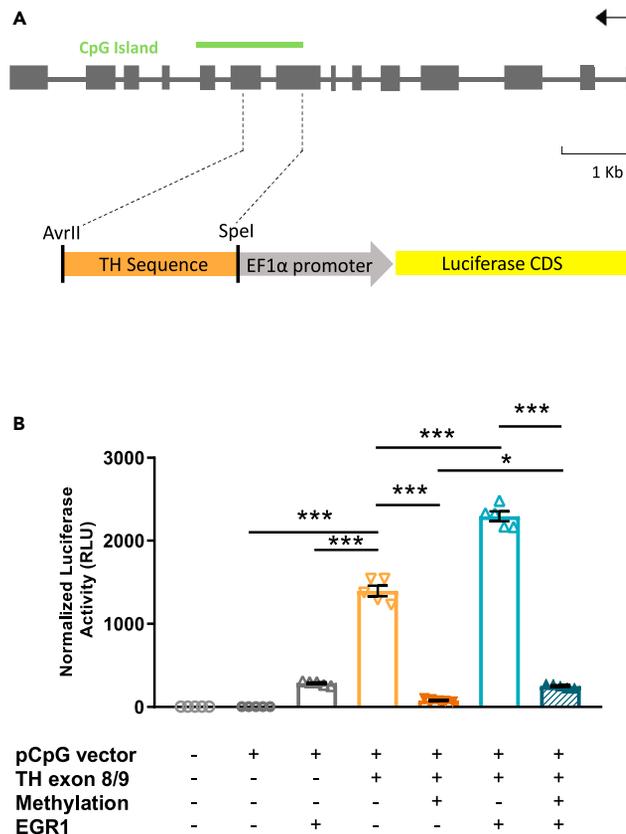
(F) Single nuclei from discovery accumbens and caudate were separated using FACS, based on DRAQ5 immunofluorescence, and neuronal nuclei were separated based on NeuN.

(G and H) Hypermethylation of the CpGs within the EGR1 binding site is specific to neuronal nuclei ( $n = 22$  cases and 20 controls) and not (H) non-neuronal nuclei. Data represented as mean  $\pm$  SEM, Two-way repeated measures ANOVA  $**p < 0.01$ ,  $*p < 0.05$

### Exon 8/9 of *TH* has enhancer activity that is regulated by DNA methylation and EGR1 expression

Since the regulatory potential of this genomic region has yet to be thoroughly investigated, we asked whether the segment containing part of exon 8 and 9, along with the intron between them, can act as an enhancer for *TH* gene expression. We performed an enhancer luciferase assay to assess the regulatory potential of a 398 bp fragment of our DMR with *TH*. We cloned the fragment into a pCpG-free vector upstream of the *EF1 $\alpha$*  promoter and transfected methylated or unmethylated constructs into HEK293 cells (Figure 4A). After 24 hours, we quantified luciferase activity and found a highly significant effect of group (One-way ANOVA,  $F = 169.95$ ;  $df = 4$ ;  $p = 1.37 \cdot 10^{-14}$ , Figure 4B). The inclusion of the *TH* fragment resulted in a 7.5-fold increase in luciferase activity compared with the empty vector (Empty vector vs. *TH*, Tukey's multiple comparisons;  $p < 0.0001$ , Figure 4B). Furthermore, when the fragment was fully methylated prior to transfection, the enhancer activity of the region was effectively lost, where luciferase expression was not significantly different from the empty vector group (Empty vector vs. Methylated *TH*, Tukey's multiple comparisons;  $p > 0.05$ , Figure 4B). This strongly suggests that the cocaine-related DMR in the *TH* gene has regulatory potential, which is attenuated by DNA methylation.

We further hypothesized that the enhancer activity of this locus would be increased in the presence of EGR1, so we co-transfected cells with one of the *TH* enhancer vectors described above, and either an EGR1 overexpression vector or a negative control. We found that excess EGR1 greatly increased enhancer activity (*TH* vs *TH* + EGR1, Tukey's multiple comparisons;  $p < 0.0001$ , Figure 4B). Importantly, when we co-transfected the methylated *TH* plasmid with the EGR1 overexpression vector, we found a significant decrease in luciferase activity compared with the unmethylated *TH* plasmid, either with or without EGR1 overexpression (Methylated *TH* + EGR1 vs *TH* + EGR1 or vs *TH*, Tukey's multiple comparisons;  $p < 0.0001$ , Figure 4B). Together, these results suggest that the intragenic region within *TH* may serve as



**Figure 4. The exon 8/9 region of *TH* functions as an enhancer in vitro and is modulated by methylation and *EGR1* overexpression**

(A) The region containing the cocaine DMR was cloned into a Luciferase enhancer plasmid and transfected into HEK293 cells for functional analyses.

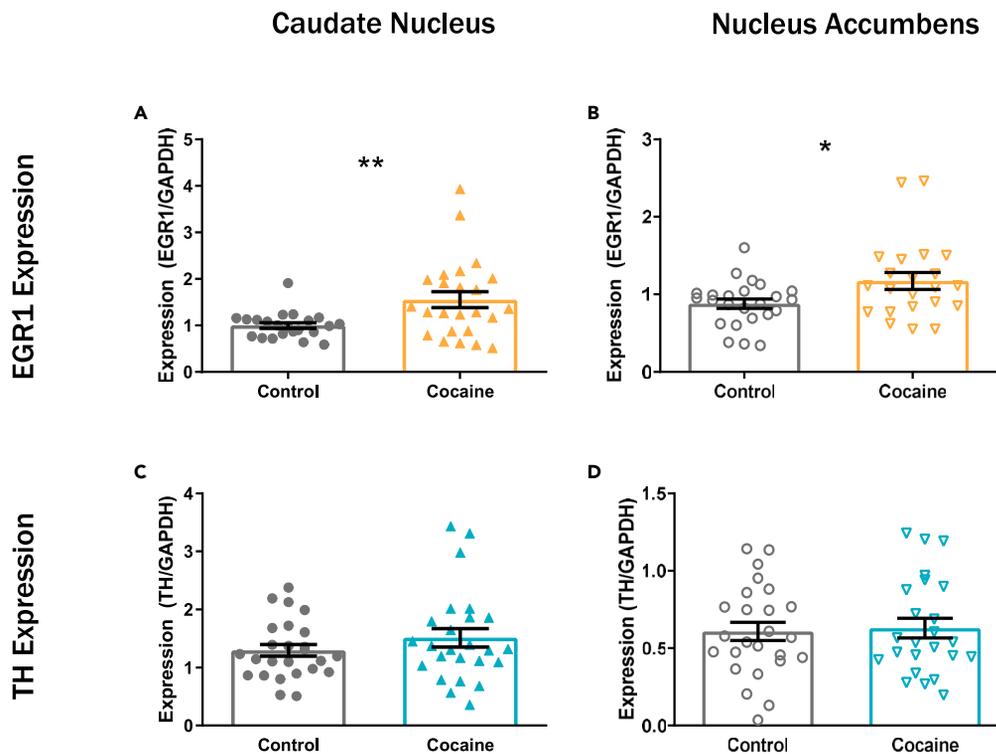
(B) The target region acts as an enhancer to significantly increase luciferase fluorescence, compared to mock transfected and empty vector controls. Enhancer activity is abolished by DNA methylation and enhanced by co-transfection with an *EGR1* overexpression vector ( $n = 5$  biological replicates per condition). Data represented as mean  $\pm$  SEM, One-way ANOVA with Tukey's multiple corrections \*\*\* $p < 0.005$ , \* $p < 0.05$

a regulatory element to nearby promoters and that it is responsive to upregulation by *EGR1* and downregulation by DNA methylation.

Finally, given the apparent relationship between DNA methylation and *EGR1* on the regulatory potential of this genomic region, we asked whether *TH* or *EGR1* mRNA expression was dysregulated with cocaine dependence in the human striatum. We found significantly more *EGR1* expression in both the caudate nucleus (Mann Whitney  $U = 144$ ,  $p = 0.008$ , Figure 5A) and the nucleus accumbens ( $t = 2.38$ ,  $df = 32.9$ ,  $p = 0.023$ , Figure 5B) in individuals who had chronic cocaine dependence compared with controls, which is consistent to what has been found in animal models of psychostimulant exposure (Moratalla et al., 1992). Conversely, we found no significant differences in *TH* expression in either brain region ( $ps > 0.1$ , Figures 5C and 5D). Potential explanations for these data are offered in the next section.

## DISCUSSION

This study has generated cocaine dependence methylome data in the human nucleus accumbens and fills a vital gap in our knowledge of how chronic cocaine exposure and addiction impact brain cell biology. We have identified over 100 DMRs that are associated with cocaine dependence in humans, including an enrichment for those overlapping with intragenic regions and active transcription start sites. This mirrors our recent findings in the human caudate nucleus (Vaillancourt et al., 2020) and fits well with the knowledge that intragenic regions, including enhancers, are important to brain health. For example, methylation at



**Figure 5. EGR1 expression is increased in the cocaine group, while TH expression remains unchanged**

(A) There is significantly more EGR1 mRNA in the caudate nucleus and (B) nucleus accumbens in the cocaine group (n = 24–25) compared to controls (n = 25).

(C) There was no significant difference in TH mRNA expression in the caudate nucleus or (D) the nucleus accumbens between cases and controls (n = 24–25 cases and 25 controls). Data represented as mean  $\pm$  SEM, Two-tailed Student's t-test \*\*p < 0.01, \*p < 0.05.

enhancers plays an important role in neuronal cell identity, and genetic polymorphisms in enhancer regions confer increased risk for multiple psychiatric disorders (Hannon et al., 2019; Kozlenkov et al., 2015). Additionally, gene body methylation has been shown to regulate gene expression through the regulation of alternative promoters and mRNA splice variants, which have both been implicated in psychiatric disorders including cocaine dependence (Feng et al., 2014; Gandal et al., 2018; Maunakea et al., 2010, 2013). Although we did not find evidence of cocaine-related alternative splicing of *TH*, the potential for differential expression of *TH* isoforms should be further explored. It stands to reason that epigenetic modifications, including DNA methylation, within this and other cocaine-sensitive genomic regions might play an important role in disease-related brain functioning as well.

In addition, we found an over-representation of hypermethylated loci within the striatum, which adds to decades of work in animal models using multiple drug delivery paradigms and time-courses (Vaillancourt et al., 2017). For example, cocaine self-administration in rodents results in more hypermethylation than hypomethylation in some downstream targets of midbrain dopamine projections (Baker-Andresen et al., 2015; Fonteneau et al., 2017). In addition, expression of the *DNMT3a* de novo methyltransferase is induced by cocaine exposure and can be maintained after extended periods of withdrawal (LaPlant et al., 2010). Furthermore, these phenomena combined, in the nucleus accumbens, appear to be necessary for the maintenance of compulsive drug seeking behaviors in rats (Massart et al., 2015). In human postmortem tissue, we found *DNMT3A* to be increased after chronic cocaine dependence in the caudate nucleus of humans (Vaillancourt et al., 2020), and the present study recapitulates this finding in the nucleus accumbens. Increased expression of the experience-dependent methyltransferase, along with the over representation of hypermethylated DMRs in both striatal subregions, suggests that chronic cocaine dependence is associated with genome-wide alterations in DNA methylation in humans.

When comparing DMRs between brain regions, we found the exact same cluster of CpGs within exon 8 of *TH* to be hypermethylated in the caudate nucleus and nucleus accumbens of human patients; our expanded analysis included 27 CpGs that spanned a portion of both exons 8 and 9, as well as the intronic region that separates them. Although we reliably detected low levels of *TH* RNA in both regions, we found no significant difference in *TH* mRNA expression in either the caudate nucleus or the nucleus accumbens in humans. There are several possible explanations for these negative data, in particular, the lack of cellular resolution. *TH* gene expression dominates in midbrain dopamine neurons that innervate striatal regions, while relatively low levels of *TH* mRNA are found in these downstream targets. It is assumed that most of this mRNA is present in dopaminergic nerve terminals that innervate striatum (Gervasi et al., 2016), although a single-nucleus transcriptome study in rats and an electrophysiology study in mice identified sparse cell clusters expressing *Th* in the nucleus accumbens; such transcript levels were unchanged following cocaine exposure (Ibáñez-Sandoval et al., 2010; Savell et al., 2020). Importantly, *Th* was detected in up to 4% of cell bodies in the anterior striatum, and although there were no changes in expression following cocaine exposure, this study used a single exposure to cocaine and therefore does not directly compare with the prolonged behavioral changes associated with a dependent phenotype (Savell et al., 2020). Additionally, since both *EGR1* expression and DNA methylation within *TH* are increased in human subjects, *TH* transcript levels may be kept in check through a combination of opposing mechanisms. In any event, future studies using single-cell technologies and dependence-like behavioral models are likely required to identify changes in mRNA expression, protein levels or protein activity in the striatal regions studied here. Most importantly, however, our discovery of a novel enhancer role for the designated region of *TH* raises the possibility that its methylation and *EGR1* binding controls the expression of other genes, which are spatially connected to this locus.

*TH* codes for tyrosine hydroxylase, the rate-limiting step in the biosynthesis of dopamine within the terminals of presynaptic cells. It has been well documented that as patients transition from recreational cocaine use to physiological dependence and addiction, they undergo lasting changes in dopamine signaling within the mesocorticolimbic circuitry (Letchworth et al., 2001; Volkow et al., 1999). Furthermore, although transcription and protein immunoreactivity of TH is known to increase after extended cocaine exposure in the rodent midbrain (Beitner-Johnson et al., 1991; Logan et al., 2019; Masserano et al., 1996), attempts to measure the effects of cocaine on TH in downstream targets have yielded mixed results. TH immunoreactivity is upregulated in the central amygdala after 45 days of withdrawal from cocaine self-administration in rats; however there appears to be no change in protein levels in the prefrontal or orbitofrontal cortex after cocaine conditioned place preference (Grimm et al., 2002; Hámor et al., 2020). In the dorsal striatum, chronic cocaine injection increases TH protein levels, while the number of TH positive varicosities in the nucleus accumbens has been reported to both increase and decrease after cocaine exposure, depending on tissue sub-sectioning and length of withdrawal (Balda et al., 2009; Schmidt et al., 2001; Todtenkopf et al., 2000). Interestingly, the decreased immunoreactivity that is found in the nucleus accumbens shell in rats after 12 days of self-administration can be reversed by extinction training and appears to be specific to cocaine and not other appetitive reward stimuli (Schmidt et al., 2001). This suggests that levels of TH protein abundance and activity in the nucleus accumbens have an important role in cocaine-seeking behavior, however, it is still unclear whether the protein in these studies is located within afferent dopaminergic projections or within striatal cell bodies themselves.

In this study, we found a variable increase in methylation within the gene body of *TH* in the human striatum that may be specific to neuronal nuclei in the caudate nucleus. The variability within our data, along with our lack of sensitivity to detect transcriptional dysregulation is highly suggestive of these findings occurring in a small proportion of cells. Indeed, in a recent single-nucleus transcriptomic survey of the rat nucleus accumbens, *Th* transcripts were detected in less than 30% of nuclei, and mouse data show relatively few instances of *Th* in single-cell transcriptome experiments (Saunders et al., 2018; Savell et al., 2020). Canonically, the mammalian striatum is composed primarily of D1- and D2-expressing medium spiny projection neurons (MSNs, 75–95%), as well as at least 4 subtypes of inhibitory interneurons (Lobo et al., 2006). Among them, are a rare group of TH-expressing interneurons (THINs, approximately 0.04%) that have distinct electrophysiological properties (Ibáñez-Sandoval et al., 2010). THINs do not appear to release dopamine, and do not co-localize with other proteins involved in monoamine synthesis (Xenias et al., 2015). Instead, these cells exhibit strong, GABAergic inhibition of MSNs in response to thalamic input, and may have a distinct role in striatum-related behaviors; selective ablation of THINs results in deficits in specific goal-directed behaviors (Kaminer et al., 2019; Xenias et al., 2015). Although the present study represents a step toward identifying cell-type specific methylation patterns associated with chronic cocaine dependence in humans,

the findings with respect to *TH* should be further examined in specifically relevant datasets including those targeting THINs.

Additionally, we found increased *EGR1* mRNA expression following cocaine dependence in humans; *EGR1* is a transcription factor that is known to have an important role in the neurobiology of cocaine in rodents. It is an immediate-early gene which is induced in both the dorsal and ventral striatum via dopaminergic signaling upon initial exposure to cocaine, but it is also necessary for context-related drug cues (Bhat et al., 1992; Fritz et al., 2011; Moratalla et al., 1992). *EGR1* is necessary for the development of cocaine conditioned place preference and its expression is re-induced upon exposure to drug-related environments, even after prolonged abstinence (Hearing et al., 2008; Valjent et al., 2006). Our data support the notion that *EGR1* is involved in cocaine-induced neuroadaptation, and we extend its role to include activation of a novel regulatory element within *TH*, thereby connecting two previously distinct pathways of dysregulation.

This study has begun to characterize the regulatory role of an intragenic element within *TH* that contains a putative *EGR1* binding site and is differentially methylated in the striatum of individuals with cocaine dependence. Future research should explore the molecular dynamics of *EGR1* binding at this locus, as well as investigate its regulatory potential in diverse conditions compared to well-known genomic enhancers. Although in vitro results from the current study are supportive evidence of enhancer activity of this locus, a downstream section of the gene beginning with exon 9 (part of which, is included within our region of interest) has been identified as a bivalent promoter, and displays a local enrichment of the enhancer-associated histone marks H3K27ac and H3K4me1 in the Roadmap Epigenome Consortium dataset (<http://epigenomegateway.wustl.edu>, (Kundaje et al., 2015)). Interestingly, these features are annotated within the samples from the human striatum, but not the substantia nigra, which is a midbrain source of dopamine production. Although more experiments are needed, this phenomenon in addition to the findings presented here, suggests a specific role of this locus in cocaine-related neurobiology in the nucleus accumbens and caudate nucleus.

### Limitations of the study

The use of postmortem human tissues is invaluable in understanding fundamentally human disorders, and although our hypotheses are supported by in vitro studies, this work is not without limitations. Given the difficulty in tissue dissection, we are unable to disentangle the effects the core versus the shell of the nucleus accumbens in our data, although each subregion is known to have distinct roles in drug seeking behavior in animals (Dumitriu et al., 2012). Furthermore, we have shown that the hypermethylation of the *EGR1* binding site is specific to neuronal nuclei in the caudate nucleus; however given the magnitude and variability of this effect, we suspect this signal to be coming from a small proportion of cells. As single-nucleus DNA methylation technologies advance so will our understanding of the importance of specific, TH-expressing cell types in cocaine neurobiology.

Our study provides evidence of DNA methylation dysregulation in the *TH* gene that is associated with cocaine dependence; however, given the nature of postmortem research, we are unable to definitively conclude whether this dysregulation is a direct consequence of the addicted phenotype. It is possible that altered *TH* methylation prior to drug exposure could predispose some individuals to develop cocaine dependency, but we are unable to test this hypothesis in human subjects. Future studies should make use of in vitro epigenome editing, using targeted DNA methylation in the striatum of living animals, to investigate the behavioral outcomes of *Th* hypermethylation.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS

- Reduced representation bisulfite sequencing (RRBS)
- RNA sequencing
- Cocaine self-administration in mice
- Targeted bisulfite amplicon sequencing
- Fluorescence activated cell sorting
- Luciferase enhancer assay
- qPCR gene expression assays
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- Differential methylation analysis of RRBS data
- Enrichment analyses
- Methyltransferase gene expression from RNA sequencing
- RRBS in cocaine self-administering mice
- Targeted bisulfite amplicon sequencing
- Luciferase enhancer assay
- qPCR gene expression

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.103169>.

### ACKNOWLEDGMENTS

We are deeply grateful to the families of the subjects used in this study. We would also like to acknowledge the teams of technicians at the Miami Brain Endowment Bank™ and the past and current members of the McGill Group for Suicide Studies for their assistance and support. This work was supported by a Canadian Institute of Health Research Doctoral Fellowship awarded to K.V., National Institute of Drug Abuse Grants DA033684 awarded to G.T and D.M., and P01 DA047233 to E.N.

### AUTHOR CONTRIBUTIONS

Manuscript preparation: K.V., experimental design and data collection: K.V., G.G.C., G.M., and L.F., data analysis: K.V., C.E., A.B., and J-F.T., animal experiments: B.L. and E.C., resources and support: E.N., C.N., N.M., D.C.M., and G.T.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 15, 2021

Revised: June 15, 2021

Accepted: September 21, 2021

Published: October 22, 2021

### REFERENCES

- Albertson, D.N., Pruetz, B., Schmidt, C.J., Kuhn, D.M., Kapatos, G., and Bannon, M.J. (2004). Gene expression profile of the nucleus accumbens of human cocaine abusers: evidence for dysregulation of myelin. *J. Neurochem.* *88*, 1211.
- Baker-Andresen, D., Zhao, Q., Li, X., Jupp, B., Chesworth, R., Lawrence, A.J., and Bredy, T.W. (2015). Persistent variations in neuronal dna methylation following cocaine self-administration and protracted abstinence in mice. *Neuroepigenetics* *4*, 1–11.
- Balda, M.A., Anderson, K.L., and Itzhak, Y. (2009). The neuronal nitric oxide synthase (NOS) gene contributes to the regulation of tyrosine hydroxylase (TH) by cocaine. *Neurosci. Lett.* *457*, 120–124.
- Barrera, V., and Peinado, M.A. (2012). Evaluation of single CpG sites as proxies of CpG island methylation states at the genome scale. *Nucl. Acids Res.* *40*, 11490–11498.
- Beitner-Johnson, D., Guitart, X., and Nestler, E.J. (1991). Dopaminergic brain reward regions of Lewis and Fischer rats display different levels of tyrosine hydroxylase and other morphine- and cocaine-regulated phosphoproteins. *Brain Res.* *561*, 147–150.
- Belin, D., and Everitt, B.J. (2008). Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. *Neuron* *57*, 432–441.
- Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74.
- Bhat, R.V., Cole, A.J., and Baraban, J.M. (1992). Role of monoamine systems in activation of Zif268 by cocaine. *J. Psychiatry Neurosci.* *17*, 94–102.
- Cavalcante, R.G., and Sartor, M.A. (2017). Annotatr: genomic regions in context. *Bioinformatics* *33*, 2381–2383.
- Chen, G.G., Diallo, A.B., Poujol, R., Nagy, C., Staffa, A., Vaillancourt, K., Lutz, P.-E., Ota, V.K., Mash, D.C., Turecki, G., and Ernst, C. (2014). BisQC: an operational pipeline for multiplexed bisulfite sequencing. *BMC Genomics* *15*, 290.
- Chen, G.G., Gross, J.A., Lutz, P.-E., Vaillancourt, K., Maussion, G., Bramouille, A., Thérout, J.-F., Gardini, E.S., Ehler, U., Bourret, G., et al. (2017). Medium throughput bisulfite sequencing for accurate detection of 5-methylcytosine and 5-hydroxymethylcytosine. *BMC Genomics* *18*, 96.

- Dumitriu, D., Laplant, Q., Grossman, Y.S., Dias, C., Janssen, W.G., Russo, S.J., Morrison, J.H., and Nestler, E.J. (2012). Subregional, dendritic compartment, and spine subtype specificity in cocaine regulation of dendritic spines in the nucleus accumbens. *J. Neurosci. Off. J. Soc. Neurosci.* **32**, 6957–6966.
- Feng, J., Wilkinson, M., Liu, X., Purushothaman, I., Ferguson, D., Vialou, V., Maze, I., Shao, N., Kennedy, P., Koo, J., et al. (2014). Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol.* **15**, R65.
- Fonteneau, M., Filliol, D., Anglard, P., Befort, K., Romieu, P., and Zwiller, J. (2017). Inhibition of DNA methyltransferases regulates cocaine self-administration by rats: a genome-wide DNA methylation study. *Genes Brain Behav.* **16**, 313–327.
- Fritz, M., El Rawas, R., Salti, A., Klement, S., Bardo, M.T., Kemmler, G., Dechant, G., Saria, A., and Zernig, G. (2011). Reversal of cocaine-conditioned place preference and mesocorticolimbic Zif268 expression by social interaction in rats. *Addict. Biol.* **16**, 273–284.
- Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., Van Bakel, H., Varghese, M., Wang, Y., and Shieh, A.W. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127.
- Gervasi, N.M., Scott, S.S., Aschrafi, A., Gale, J., Vohra, S.N., MacGibeny, M.A., Kar, A.N., Gioio, A.E., and Kaplan, B.B. (2016). The local expression and trafficking of tyrosine hydroxylase mRNA in the axons of sympathetic neurons. *RNA* **22**, 883–895.
- Grimm, J.W., Shaham, Y., and Hope, B.T. (2002). Effect of cocaine and sucrose withdrawal period on extinction behavior, cue-induced reinstatement, and protein levels of the dopamine transporter and tyrosine hydroxylase in limbic and cortical areas in rats. *Behav. Pharmacol.* **13**, 379–388.
- GTEX Consortium (2015). Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N.Y.)* **348**, 648–660.
- Hámor, P.U., Edelmann, M.J., Gobin, C., and Schwendt, M. (2020). Long-term changes in the central amygdala proteome in rats with a history of chronic cocaine self-administration. *Neuroscience* **443**, 93–109.
- Hannon, E., Marzi, S.J., Schalkwyk, L.S., and Mill, J. (2019). Genetic risk variants for brain disorders are enriched in cortical H3K27ac domains. *Mol. Brain* **12**, 7.
- Hearing, M.C., See, R.E., and McGinty, J.F. (2008). Relapse to Cocaine-seeking increases activity-regulated gene expression differentially in the striatum and cerebral cortex of rats following short or long periods of abstinence. *Brain Struct. Funct.* **213**, 215–227.
- Ibáñez-Sandoval, O., Tecuapetla, F., Unal, B., Shah, F., Koós, T., and Tepper, J.M. (2010). Electrophysiological and morphological characteristics and synaptic connectivity of tyrosine hydroxylase-expressing neurons in adult mouse striatum. *J. Neurosci. Off. J. Soc. Neurosci.* **30**, 6999–7016.
- Kaminer, J., Espinoza, D., Bhimani, S., Tepper, J.M., Koos, T., and Shiflett, M.W. (2019). Loss of striatal tyrosine-hydroxylase interneurons impairs instrumental goal-directed behavior. *Eur. J. Neurosci.* **50**, 2653–2662.
- Kozlenkov, A., Wang, M., Roussos, P., Rudchenko, S., Barbu, M., Bibikova, M., Klotzle, B., Dwork, A.J., Zhang, B., Hurd, Y.L., et al. (2015). Substantial DNA methylation differences between two major neuronal subtypes in human brain. *Nucl. Acids Res.* **44**, 2593–2612.
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330.
- LaPlant, Q., Vialou, V., Covington, H.E., Dumitriu, D., Feng, J., Warren, B.L., Maze, I., Dietz, D.M., Watts, E.L., Iñiguez, S.D., et al. (2010). Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. *Nat. Neurosci.* **13**, 1137–1143.
- Letchworth, S.R., Nader, M.A., Smith, H.R., Friedman, D.P., and Porrino, L.J. (2001). Progression of changes in dopamine transporter binding site density as a result of cocaine self-administration in rhesus monkeys. *J. Neurosci. Off. J. Soc. Neurosci.* **21**, 2799–2807.
- Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science (New York, N.Y.)* **341**, 1237905.
- Lobo, M.K., Karsten, S.L., Gray, M., Geschwind, D.H., and Yang, X.W. (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nat. Neurosci.* **9**, 443–452.
- Logan, R.W., Parekh, P.K., Kaplan, G.N., Becker-Krail, D.D., Williams, W.P., Yamaguchi, S., Yoshino, J., Shelton, M.A., Zhu, X., Zhang, H., et al. (2019). NAD<sup>+</sup> cellular redox and SIRT1 regulate the diurnal rhythms of tyrosine hydroxylase and conditioned cocaine reward. *Mol. Psychiatry* **24**, 1668–1684.
- Luo, C., Keown, C.L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J.R., Sandoval, J.P., et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604.
- Massart, R., Barnea, R., Dikshtein, Y., Suderman, M., Meir, O., Hallett, M., Kennedy, P., Nestler, E.J., Szyf, M., and Yadid, G. (2015). Role of DNA methylation in the nucleus accumbens in incubation of cocaine craving. *J. Neurosci. Off. J. Soc. Neurosci.* **35**, 8042–8058.
- Masserano, J.M., Baker, I., Natsukari, N., and Wyatt, R.J. (1996). Chronic cocaine administration increases tyrosine hydroxylase activity in the ventral tegmental area through glutamatergic and dopaminergic D2-receptor mechanisms. *Neurosci. Lett.* **217**, 73–76.
- Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23**, 1256–1269.
- Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D’Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257.
- Moratalla, R., Robertson, H.A., and Graybiel, A.M. (1992). Dynamic regulation of NGFI-A (Zif268, Egr1) gene expression in the striatum. *J. Neurosci.* **12**, 2609–2622.
- Nestler, E.J., and Lüscher, C. (2019). The molecular basis of drug addiction: linking epigenetic to synaptic and circuit mechanisms. *Neuron* **102**, 48–59.
- Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e16.
- Savell, K.E., Tuscher, J.J., Zipperly, M.E., Duke, C.G., Phillips, R.A., Bauman, A.J., Thukral, S., Sultan, F.A., Goska, N.A., Ianov, L., and Day, J.J. (2020). A dopamine-induced gene expression signature regulates neuronal function and cocaine response. *Sci. Adv.* **6**, eaba4221.
- Schmidt, E.F., Sutton, M.A., Schad, C.A., Karanian, D.A., Brodtkin, E.S., and Self, D.W. (2001). Extinction training regulates tyrosine hydroxylase during withdrawal from cocaine self-administration. *J. Neurosci. Off. J. Soc. Neurosci.* **21**, 137.
- Schoenfelder, S., and Fraser, P. (2019). “Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455.
- Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and bioconductor. *Bioinformatics* **32**, 587–589.
- Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141.
- Thomsen, M., and Caine, S.B. (2007). Intravenous drug self-administration in mice: practical considerations. *Behav. Genet.* **37**, 101–118.
- Tian, W., Zhao, M., Li, M., Song, T., Zhang, M., Quan, L., Li, S., and Sun, Z.S. (2012). Reversal of cocaine-conditioned place preference through methyl supplementation in mice: altering global DNA methylation in the prefrontal cortex. *PLoS One* **7**, e33435.
- Todtenkopf, M.S., De Leon, K.R., and Stellar, J.R. (2000). Repeated cocaine treatment alters tyrosine hydroxylase in the rat nucleus accumbens. *Brain Res. Bull.* **52**, 407–411.
- Tomasi, D., Volkow, N.D., Wang, R., Carrillo, J.H., Maloney, T., Alia-Klein, N., Woicik, P.A., Telang, F., and Goldstein, R.Z. (2010). Disrupted

functional connectivity with dopaminergic midbrain in cocaine abusers. *PLoS One* 5, e10815.

Vaillancourt, K., Ernst, C., Mash, D., and Turecki, G. (2017). DNA methylation dynamics and cocaine in the brain: progress and prospects. *Genes* 8, 138.

Vaillancourt, K., Yang, J., Chen, G.G., Yerko, V., François Thérout, J., Aouabed, Z., Lopez, A., Thibeault, K.C., Calipari, E.S., Labonté, B., et al. (2020). Cocaine-related DNA methylation in caudate neurons alters 3D chromatin structure of the IRXA gene cluster. *Mol. Psychiatry*. Epub ahead of print. <https://doi.org/10.1038/s41380-020-00909-x>.

Valjent, E., Aubier, B., Gaëlle Corbillé, A., Brami-Cherrier, K., Topilko, P., Girault, J.A., and Hervé, D. (2006). Plasticity-associated gene *Krox24*

*Zif268* is required for long-lasting behavioral effects of cocaine. *J. Neurosci.* 26, 4956–4960.

Volkow, N.D., Wang, G.J., Fowler, J.S., Logan, J., Gatley, S.J., Gifford, A., Hitzemann, R., Ding, Y.S., and Pappas, N. (1999). Prediction of reinforcing responses to psychostimulants in humans by brain dopamine D2 receptor levels. *Am. J. Psychiatry* 156, 1440–1443.

Volkow, N.D., Wang, G.J., Telang, F., Fowler, J.S., Logan, J., Childress, A.R., Jayne, M., Ma, Y., and Wong, C. (2006). Cocaine cues and dopamine in dorsal striatum: mechanism of craving in cocaine addiction. *J. Neurosci.* 26, 6583–6588.

Walker, D.M., Cates, H.M., Loh, Y.H.E., Purushothaman, I., Ramakrishnan, A., Cahill, K.M., Lardner, C.K., Godino, A., Kronman, H.G., Rabkin, J., et al. (2018). "Cocaine self-administration alters transcriptome-wide responses in the

brain's reward circuitry. *Biol. Psychiatry* 84, 867–880.

Wright, K.N., Hollis, F., Duclot, F., Dossat, A.M., Strong, C.E., Francis, T.C., Mercer, R., Feng, J., Dietz, D.M., Lobo, M.K., et al. (2015). Methyl supplementation attenuates cocaine-seeking behaviors and cocaine-induced c-fos activation in a DNA methylation-dependent manner. *J. Neurosci. Off. J. Soc. Neurosci.* 35, 8948–8958.

Xenias, H.S., Ibáñez-Sandoval, O., Koós, T., and Tepper, J.M. (2015). Are striatal tyrosine hydroxylase interneurons dopaminergic? *J. Neurosci. Off. J. Soc. Neurosci.* 35, 6584–6599.

Zandarashvili, L., White, M.A., Esadze, A., and Iwahara, J. (2015). Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor *Egr-1*. *FEBS Lett.* 589, 1748–1753.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Milli-Mark anti-NeuN-PE antibody, clone A60	Millipore, Darmstadt, Germany	Cat no. FCMAB317PE; RRID:AB_11212465
<b>Bacterial and virus strains</b>		
One Shot™ TOP10 Chemically Competent <i>E. coli</i>	Invitrogen (ThermoFisher)	Cat no. C404003
<b>Biological samples</b>		
Postmortem human brain (dorsolateral caudate nucleus and nucleus accumbens)	University of Miami, Miller School of Medicine, Brain Endowment Bank	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
<i>MspI</i> Restriction Enzyme	New England Biolabs, Ipswich, Massachusetts	Cat no. R0106S
<i>AvrII</i> Restriction Enzyme	New England Biolabs, Ipswich, Massachusetts	Cat no. R0174S
<i>SpeI</i> Restriction Enzyme	New England Biolabs, Ipswich, Massachusetts	Cat no. R0133S
T4 Ligase	New England Biolabs, Ipswich, Massachusetts	Cat no. M0202M
<i>SssI</i> Methylase	New England Biolabs, Ipswich, Massachusetts	Cat no. M0226L
<i>HpaII</i> Restriction Enzyme	New England Biolabs, Ipswich, Massachusetts	Cat no. R0171S
<b>Critical commercial assays</b>		
QIAamp DNA Mini Kit	Qiagen, Germantown, Maryland	Cat no.51304
EpiTect Fast Bisulfite Conversion Kit	Qiagen, Germantown, Maryland	Cat no. 59824
TruSeq DNA PCR-Free High Throughput Library Prep Kit	Illumina, San Diego, California	Cat no. 20015963
TruSeq DNA Single Indexes Set A	Illumina, San Diego, California	Cat no. 20015960
RNeasy Lipid Tissue Mini Kit	Qiagen, Germantown, Maryland	Cat no. 74804
MiSeq Reagent Kit v3 (600-cycle)	Illumina, San Diego, California	Cat no. MS102-3003
Quant-iT PicoGreen dsDNA assays	ThermoFisher, Waltham, Massachusetts	Cat no. P7589
QIAquick PCR Purification Kit	Qiagen, Germantown, Maryland	Cat no. 28104
QIAprep Spin Miniprep Kit	Qiagen, Germantown, Maryland	Cat no. 27104
QIAquick Gel Extraction Kit	Qiagen, Germantown, Maryland	Cat no. 28706X4
QIAGEN Plasmid Midi Kit	Qiagen, Germantown, Maryland	Cat no. 12143
jetPRIME Transfection kit	Polyplus, Strasbourg, France	Cat no. 114
Dual-Luciferase® Reporter Assay System	Promega, Madison, Wisconsin	Cat no. E1910
TaqMan Assay <i>TH</i> (Hs00165941_m1)	ThermoFisher, Waltham, Massachusetts	Cat no. 4331182
TaqMan Assay <i>EGR1</i> (Hs00152928_m1)	ThermoFisher, Waltham, Massachusetts	Cat no. 4331182
TaqMan Assay <i>GAPDH</i> (Hs02758991_g1)	ThermoFisher, Waltham, Massachusetts	Cat no. 4331182
<b>Deposited data</b>		
Reduced representation sequencing data from human dorsolateral caudate and nucleus accumbens	This paper	GEO: GSE182585
<b>Experimental Models: Cell Lines</b>		
HEK293 cell line (human embryonic kidney)	Lab of Dr. Carl Ernst	N/A

(Continued on next page)

<i>Continued</i>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: organisms/strains		
Mouse:C57/B6 mice	Lab of Dr. Eric Nestler	N/A
Oligonucleotides		
Methylated C adaptor: mC-PE1	Chen et al. 2014	ACACTCTTTCCCTACACGACGCT CTTCCGATCsT-OH
Methylated C adaptor: mC-PE2	Chen et al. 2014	p-GATCGGAAGAGCGGTTCAG CAGGAATGCCGAG-OH
PCR primer: IndPEPCR_F	Chen et al. 2014	AATGATACGGCGACCACCGA GATCTACACTCTTCCCTACAC GACGCTCTCCGATCsT
PCR primer: IndPEPCR_R	Chen et al. 2014	GTGACTGGAGTTCAGACGTGTGC TCTCCGATCsT
Index_1R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGAGAT CGTGATGTGACTGGAGTTC-OH
Index_2R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGAGATAC ATCGGTGACTGGAGTTC-OH
Index_3R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGAGAT GCCTAAGTACTGGAGTTC-OH
Index_4R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGA GATTGGTCACTGGAGTTC-OH
Index_5R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGAGA TCACTGTGTGACTGGAGTTC-OH
Index_6R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGA GATATTGGCGTACTGGAGTTC-OH
Index_7R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGA GATGATCTGGTACTGGAGTTC-OH
Index_8R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGA GATCAAGTGTGACTGGAGTTC-OH
Index_9R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGAGA TCTGATCGTACTGGAGTTC-OH
Index_10R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGA GATAAGCTAGTACTGGAGTTC-OH
Index_11R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGA GATGTAGCCGTACTGGAGTTC-OH
Index_12R primer	Chen et al. 2014	CAAGCAGAAGACGGCATAACGA GATGAACATGTACTGGAGTTC-OH
Bisulfite amplicon primer TH_1_F	Integrated DNA Technologies, Coralville, Iowa	5' AAG TTY GTG YGT TTT GTA AGG 3'
Bisulfite amplicon primer TH_1_R	Integrated DNA Technologies, Coralville, Iowa	5' TCT ACA CCA CRC TAA AAA ACC TC 3'
Bisulfite amplicon primer TH_2_F	Integrated DNA Technologies, Coralville, Iowa	5' GTT GTT YGT AGG AAG GAG GT 3'
Bisulfite amplicon primer TH_2_R	Integrated DNA Technologies, Coralville, Iowa	5' AAA CRC TTA ACT AAC CAT CCC 3'
Luciferase primer THup	Integrated DNA Technologies, Coralville, Iowa	5' AGG CAT TAG AGG GCC CTG AGC CTG G 3'
Luciferase primer THdw	Integrated DNA Technologies, Coralville, Iowa	5' ATA TAC TGG GTG CAC TGG AAC ACG C 3'
Luciferase primer THAvrllup	Integrated DNA Technologies, Coralville, Iowa	5' TAA TCC TAG GGG CAT TAG AGG GCC CTG AGC CTG G 3'
Luciferase primer THSpe1dw	Integrated DNA Technologies, Coralville, Iowa	5' ATT AAC TAG TAT ATA CTG GGT GCA CTG GAA CAC GC 3'

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Luciferase primer pCPGfree-promMCSFw	Integrated DNA Technologies, Coralville, Iowa	5' CAC ACA CAT GTG TGC ATT CAT AAA TAT ATA C 3'
Luciferase primer pCPGfree-promMCSrev	Integrated DNA Technologies, Coralville, Iowa	5' TTC TCA GGG ACT GTG GGC CAT GT 3'
<b>Recombinant DNA</b>		
pCpGfree-Lucia cloning plasmid	Invivogen, San Diego, California	Cat code - pcpgf-lucia
<b>Software and algorithms</b>		
RStudio	<a href="https://www.rstudio.com/">https://www.rstudio.com/</a>	bumphunter 3.5; annotatr 1.10.0;
PANTHER	<a href="http://www.pantherdb.org/">http://www.pantherdb.org/</a>	N/A
Methyl Primer Express 1.0	Applied Biosystems, CA	N/A
BD FACSDIVA	BD Biosciences, San Jose, CA	N/A
PrimerQuest	Integrated DNA Technologies, Coralville, Iowa; <a href="http://www.idtdna.com/primerquest/home/index">http://www.idtdna.com/primerquest/home/index</a>	N/A
NCBI PrimerBlast	<a href="http://www.ncbi.nlm.nih.gov/tools/primer-blast/">http://www.ncbi.nlm.nih.gov/tools/primer-blast/</a>	N/A
USCS Genome Browser	<a href="https://genome.ucsc.edu/">https://genome.ucsc.edu/</a>	N/A
SPSS 20	IBM	N/A
GraphPad Prism 6	<a href="http://www.graphpad.com">www.graphpad.com</a>	N/A

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be made directly to, and will be fulfilled by, the Lead Contact, Gustavo Turecki ([gustavo.turecki@mcgill.ca](mailto:gustavo.turecki@mcgill.ca)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- The RRBS data generated in this study have been deposited at GEO and are publicly available as of the date of publication. The accession number can be found in the [key resources table](#).
- This study did not generate original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

All the methods in this study were approved by the Douglas Hospital Research Ethics Board, and human postmortem tissues were acquired from the Brain Endowment Bank at the University of Miami Miller School of Medicine, where autopsy and tissue handling was performed in accordance to the established standards. For the discovery cohort, nucleus accumbens and dorsolateral caudate nucleus tissue was dissected from 25 adult male subjects who had long term histories of cocaine dependence (as determined by licensed clinicians), and who died from cocaine-related complications as determined by forensic pathology and tissue toxicology. These subjects ("cases") were selected based on the absence of illicit drug toxicology other than cocaine, and the absence of other psychiatric diagnoses as determined by medical records and reports from next of kin. Additionally, 25 drug-naïve, psychiatrically healthy adult male subjects who died from accidental or natural causes, were selected as "controls". The replication cohort was selected from independent "cases" and "controls" using the same selection criteria (n=18 per group), and nucleus accumbens and caudate nucleus tissue was dissected as described above. All subjects in this cohort were male, which is reflective of the opportunistic composition of the brain samples available

at autopsy, and no *a priori* power analyses were performed in relation to sample sizes, due to the rarity of these samples. This experimental design precluded the study of the associations between sex, gender, or both on the outcome variables.

## METHOD DETAILS

### Reduced representation bisulfite sequencing (RRBS)

**Library preparation.** We used 20 mg of frozen nucleus accumbens tissue from each subject in the discovery cohort to extract genomic DNA (gDNA), using Qiagen DNA MiniKits as per manufacturer's instructions (Cat no. 51304, Qiagen, Germantown, Maryland). Next, we digested 1 µg of gDNA overnight with the *MspI* restriction enzyme (Cat no. R0106S, New England Biolabs, Ipswich, Massachusetts) and completed RRBS library preparation as follows (Chen et al., 2014). The ends of the fragmented DNA was repaired using the and dA-tails were added according to manufacturer instructions (NEBNext modules; Cat: E6050 and E6053; New England Biosystems). After phenol/chloroform purification, pre-annealed, fully methylated adaptors were added via T4 ligase annealing (Chen et al., 2014; primer sequences available in the [key resources table](#)). Final libraries were purified with QiaQuick PCR Purification (Qiagen, Cat# 28104) and converted with the EpiTect fast bisulfite conversion kit (Qiagen, Cat# 59824) according to the standard protocol. Converted libraries were indexed through PCR amplification before sequencing.

**Sequencing and bioinformatic processing.** Libraries were prepped (TruSeq DNA PCR-Free Kit; Illumina), indexed (TruSeq DNA Single Indexes, Illumina), and sequenced on the Illumina HiSeq 2000 platform at the Genome Quebec Innovation Center (Montreal, Canada) with 50bp single end sequencing. All bioinformatic processing and quality control was performed in-house, as described (Chen et al., 2014). We calculated the mean bisulfite conversion efficiency using the ratio of reads containing T to C at the unmethylated cytosine position that was added during the end-repair step of library construction for each library, and then averaged across group (case and control).

### RNA sequencing

We used RNeasy Lipid Tissue Kits (Cat no. 74804, Qiagen) to extract RNA from 100mg of nucleus accumbens tissues, according to the manufacturer's standards. The RNA integrity number (RIN) for controls was, on average,  $8.1 \pm 0.6$  and  $8.1 \pm 1.2$  for cases, as measured with RNA bioanalyzer chips (Agilent). Aliquots of 100ng/ul of RNA were sent for library preparation and sequencing at the Broad Institute (Cambridge, MA) and libraries were prepared using a standard non-strand specific protocol (Illumina TruSeq), including poly-A selection, and multiplexed for 50bp paired end sequencing on the Illumina HiSeq 2000 platform, and alignment was performed and reads per kilobase per million (RPKM) were calculated as previously described (GTEx Consortium, 2015).

### Cocaine self-administration in mice

**Self-administration.** Adult male C57/B6 mice (n= 10 per group) were anesthetized with ketamine (100 mg/kg) and xylazine (10 mg/kg) implanted with chronic indwelling jugular catheters and trained for i.v. self-administration as outlined below (Thomsen and Caine, 2007). The sterilized catheter tubing was passed subcutaneously from the back to the jugular vein and 1.2 cm of tubing was inserted into the vein and secured with silk suture. Following surgery, animals were singly housed, and allowed to recover for 48-72 hours. Each animal was maintained on a reversed light cycle (7:00am lights off; 7:00 pm lights on) and all self-administration procedures occurred during the active/dark cycle. Self-administration training sessions were two hours in length and animals self-administered cocaine (0.5 mg/kg/inj over 3 sec) on a fixed-ratio 1 schedule of administration. At the beginning of each session a house light was illuminated signifying the availability of drug. After recovery from surgery, the animals were placed in a drug self-administration operant chamber where they were connected to a drug-intake line. The operant boxes are equipped with two retractable levers; depression of the "active" lever resulted in a drug infusion, whereas no infusion occurred by pressing the "inactive" lever. Concurrent with the start of each injection, the lever retracted, the house light was turned off, and a stimulus light was activated for 5 seconds to signal a time-out period. Under these conditions, animals acquired a stable pattern of intake within 1 to 5 days. For self-administering animals, acquisition (Day 1) was counted when the animal reached 70% responding on the active lever and 10 or more responses. Following acquisition, the animals were given access to a cocaine-paired lever for 120 minutes per day for a period of 10 days. Control animals underwent the same experimental procedures but had access to a saline-paired lever.

**Reduced representation bisulfite sequencing in mice.** Animals were sacrificed immediately after their last self-administration session and brain tissue was removed and flash frozen at  $-80^{\circ}\text{C}$  until micro-dissection. Caudate-putamen and nucleus accumbens tissue was dissected from frozen tissue, and DNA was extracted using Qiagen DNA Minikits as per manufacturer's protocol. RRBS libraries were prepared as above, using 2ng of genomic DNA from each animal. Each library was individually barcoded, and pooled for 50 bp SE sequencing, 4-5 libraries per lane, on the Illumina HiSeq 2000 platform at the Genome Quebec Innovation Center.

### Targeted bisulfite amplicon sequencing

**Validation cohort.** We chose to validate our findings in as many of the original cases and controls as possible. Due to variable tissue availability and sequencing quality, our validation analyses were conducted using 23-25 cases and 26-27 controls (exact sample sizes are provided within figure legends).

**Library preparation.** We used Methyl Primer Express Software v1.0 (Applied Biosystems, CA, USA) to design redundant pairs of bisulfite specific primers against the target region with *TH* with optimal melting temperatures of  $60 \pm 2^{\circ}\text{C}$  for multiplexed reactions (Table S5). We amplified each sample using 10ul reactions with 5X multiplexed primers (10uM), 3X bisulfite converted DNA and 2X KAPA HiFi HotStart Uracil+ReadyMix (Kapa Biosystems, MA, USA). Each strand was amplified separately, and after two rounds of paramagnetic bead purification at 0.8X, amplicons from both strands were combined and amplified for 10 additional cycles to add custom primer sequences in 20 ul reactions consisting of 2.5X sample, 5X combined CS1 and CS2 primers (10uM) and 2X KAPA HiFi HotStart ReadyMix (Kapa Biosystems, MA, USA). Samples were indexed for 10 cycles in a 20ul reaction consisting of 2.5X amplicons, 5X indexing primers (10uM) and 2X KAPA HiFi HotStart ReadyMix (Kapa Biosystems, MA, USA). Each indexed library went through two rounds of double ended bead purification (final ratio 0.8X) for amplicon size selection (400-700bp). Final library concentrations and quality control was performed on the Agilent 2200 TapeStation (Agilent Technologies, CA, USA) before samples were pooled and sequenced.

**MiSeq sequencing.** We pooled libraries to a final concentration of 2nM and included a 5-10% PhiX spike-in control before sequencing on the Illumina MiSeq platform (Illumina, San Diego, CA) using customized 300bp paired end sequencing as described in (Chen et al., 2017) (MiSeq Reagent Kit v3 600-cycle; Illumina, Cat no. MS102-3003). All quality control and read alignment, without removing duplicates, were performed in-house and methylation was calculated as the percent of reads containing cytosine rather than thymidine at each position (methylation = [# of cytosine reads/ # total reads] x 100).

### Fluorescence activated cell sorting

**Nuclear extraction and labeling.** For cell-type specific experiments, 20 controls and 22 cases from the discovery cohort with adequate caudate tissue supply remaining, and 19 cases and 23 controls from the discovery cohort with adequate nucleus accumbens tissue supply remaining, were used. In order to liberate intact nuclei, we homogenized 50mg of frozen tissue in nuclei buffer containing 10mM PIPES (pH 7.4), 10mM KCl, 2mM  $\text{MgCl}_2$ , 1mM DTT, 0.1% TritonX-100 and 10X Protease Inhibitor Cocktail (Sigma Aldrich, Darmstadt, Germany). Homogenates were passed through a 30% sucrose gradient in nuclei buffer in order to separate nuclei from cellular debris, then after a wash with nuclei buffer, nuclei pellets were resuspended in blocking buffer containing 0.5% bovine serum albumin in 10X normal goat serum. Each sample was co-incubated with the DNA labeling dye DRAQ5 (1:300) (ThermoFisher, Waltham, MA) and an anti-NeuN-PE antibody (1:300) (cat no. FCMAB317PE, Millipore, Darmstadt, Germany) for 60 min at room temperature, then passed through 40uM filter caps before sorting.

**Nuclei sorting.** Labeled nuclear extracts were processed on our in-house BDFACSaria III platform (BD Biosciences, San Jose, CA) according to technical specifications provided by the company. We used BD FACSDIVA software (BD Biosciences, San Jose, CA) to first isolate single, intact nuclei based on DRAQ5 fluorescence at the 730/45-A filter (DRAQ5), then to sort neuronal from non-neuronal nuclei based on fluorescence detected by the 585/42 filter (PE). Sorted nuclear fractions were stored at  $-20^{\circ}\text{C}$  until DNA extraction.

**Nuclear DNA extraction.** We incubated nuclear fractions with 50X protease (Qiagen, Montreal, Canada) at  $56^{\circ}\text{C}$  for at least 12 hours to ensure thorough digestion of the nuclear membranes. Liberated DNA was

precipitated onto 0.2X Agencourt AMPure XP (Beckman Coulter, Brea, CA) beads after adding 20% PEG-8000 2.5M NaCl to a final PEG concentration of 10%. The beads were washed twice in a magnetic stand with 70% EthOH, then DNA was eluted in MilliQ H<sub>2</sub>O. We measured the concentration of each DNA sample using Quant-iT PicoGreen dsDNA assays (ThermoFisher, Waltham, MA) according to manufacturer specifications. 500ng of DNA from each fraction was bisulfite converted using the EpiTect fast bisulfite conversion kit (QIAGEN, Cat# 59824) according to the standard protocol, and targeted methylation amplicon libraries were constructed and sequenced as described above.

### Luciferase enhancer assay

**Primer design and PCR.** Primers were designed, using the PrimerQuest tool from IDT (<http://www.idtdna.com/primerquest/home/index>) to amplify the DNA region Chr11 : 2187899-2188296 (hg19). Primers were chosen based on their melting temperature and their ability not to generate primer dimers. Their specificities were assessed by using the NCBI primer blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and the UCSC genome browser (<https://genome.ucsc.edu/>). The PCR products were generated by two successive PCRs, each followed by purification with standard columns (QIAquick PCR Purification Kit, Qiagen). A 398bp fragment was obtained using the primer a and b (Table S6). Using the first purified PCR product as template, secondary PCR products containing restriction sites were generated with the primers c and d (Table S6).

**Cloning.** The PCR products were cloned into the TOPO 2.1 vector (Life Technologies) and TOP10 bacteria were transformed with the ligation products. Positive clones were selected based on the absence of Beta Galactosidase expression. Plasmids were extracted using the Qiagen miniprep kit according to standard protocols (QIAprep Spin Miniprep Kit, cat no. 27104) and sanger sequenced at the Genome Quebec Innovation Center using the M13 primers.

The inserts identified in the topo TA cloning vector were released from the vector with a double digestion using *AvrII* (R0174S) and *SpeI* (R0133S-New England Biolabs) enzymes and gel purification (QIAquick Gel Extraction Kit, Qiagen, cat no. 28706X4). pCpGfree promoter Lucia vectors (Cat code: pcpgf-lucia, Invivogen, San Diego, California) were digested under the same conditions and were dephosphorylated using Antarctic Phosphatase (M0289S-New England Biolabs) followed by column purification (Qiagen). The inserts and pCpGfree vector were ligated overnight at 16°C using T4 Ligase (M0202M-New England Biolabs). GT115 competent bacteria were transformed with the ligation products and plated on LB agar medium with Zeocin (Invivogen). Isolated colonies were then regrown in liquid medium and plasmids were extracted using the Qiagen miniprep kit (Qiagen). Plasmids were verified by PCR using primer designed on the pCpGfree vector (primers e and f) (Table S6) and sanger sequenced with the same primers at the Genome Quebec Innovation Center (Montreal, Canada). Clones containing the fragments of interest were regrown in 200ml of LB medium with Zeocin (Invitrogen). Plasmids were re-extracted using a fast mid-prep kit (QIAGEN Plasmid Midi Kit, Qiagen, cat no. 12143). Finally, the final construct was verified by sequencing.

**In vitro methylation.** The constructs in pCpGfree-promoter-Lucia vector were methylated in vitro by treatment with the *Sss1* methylase (M0226L-New England Biolabs) and S-Adenin-Methionine for 6 hours and then purified. The efficiency of the in vitro methylation was assessed by *HpaII* digestion (R0171S-New England Biolabs) of native and methylated vectors followed by agarose gel electrophoresis.

**Cell culture.** HEK293 cells were seeded in 6 well plates for expansion in DMEM medium (InvitroGen) supplemented with 10% FBS and penicillin/streptomycin (Life Technologies). Cells were then split in 24-well plates 24 hours before transfection. The cells were transfected at 50 - 60% of confluence with 70ng of pCpGfree vector and 70ng of pGI3 control vector using the jetPRIME transfection kit (Polyplus, cat no. 114). Each condition was treated in five biological replicates.

**Luciferase assay.** The cell culture medium and the cellular extracts were collected 24 hours after transfection. The luciferase activities were assessed in both fractions using a Dual Luciferase reporter assay Kit (Promega, cat no.E1910) and a Berthold luminometer. The firefly bioluminescence data was normalized to *Renilla* bioluminescence data which was co-transfected in order to control for differences in transfection efficiency between wells, as per manufacturer's protocol. Data were acquired using the Simplicity 4.2 Software.

### qPCR gene expression assays

For expression analysis, we used extracted RNA from our RNA sequencing experiments. Nucleic acid concentration was determined via nanodrop and RNA integrity numbers (RIN) were as follows (mean  $\pm$  s.d.):  $8.4 \pm 0.78$  for controls and  $7.9 \pm 1.3$  for cases from the caudate, and  $8.1 \pm 0.6$  for controls and  $8.1 \pm 1.2$  for cases in the nucleus accumbens. Aliquots of 25ng/ul were used to generate standard curves and diluted 1 in 5 for expression analysis.

The expression of *TH* and *EGR1* was determined using quantitative reverse transcription PCR (RT-qPCR). We used pre-designed probe-based assays for *TH* (Hs00165941\_m1; cat no. 4331182) and *EGR1* (Hs00152928\_m1, cat no. 4331182) with FAM-TAMRA dyes (Taqman assays, ThermoFischer), and ran 10ul assays on an Applied Biosystems QuantStudio 6 instrument under default cycling conditions (ThermoFischer).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Differential methylation analysis of RRBS data

We filtered out CpGs that were not present in at least 25 subjects from both cases and controls and had less than 5X coverage across libraries. Next, we grouped CpGs within 50 bp of each other into clusters using the *bumphunter* 3.5 package for R, and removed those that had a standard deviation of less than 5% methylation across all subjects (i.e., irrespective of status) to reduce unnecessary comparisons between stably methylated regions. Each cluster contained at least 2 CpGs, without an upper limit to cluster size, resulting in 32 535 clusters moving forward. We analyzed each cluster using a general linear model with status (case or control) as a fixed factor and covariates including ethnicity, age, smoking status, and ethanol toxicology. We treated CpGs independently in each cluster but investigated only those clusters that had a Bejamaini-Hochberg FDR corrected p-value  $<0.05$  and which were  $<0.05$  when calculating a single mean from all CpGs per individual.

### Enrichment analyses

We used the *annotatr* 1.10.0 package in R to annotate all CpG clusters and DMRs to their genomic context, CpG island proximity, and predicted ChromHMM chromatin state using (Cavalcante and Sartor, 2017). We used the 15-core marks from ChromHMM that were generated using human striatum tissue and calculated enrichment q values for DMRs against all CpG clusters using the LOLA algorithm (Sheffield and Bock, 2016) (cite ChromHMM).

For gene ontology, we matched each DMR to its nearest Refseq gene and used over representation tests in the gene list analysis functions of the PANTHER classification system (<http://www.pantherdb.org/>). We compared the full DMR list to all human genes with respect to molecular function, biological processes and cellular components and p-values were calculated using Fisher's Exact tests with FDR correction.

### Methyltransferase gene expression from RNA sequencing

Data were extracted for the three methyltransferase genes (DNMT1, DNMT3A, and DNMT3B), and RPKM values were averaged across groups, after removing statistical outliers (ROUT Q=1%). Cases and controls were compared using two-tailed student's t-tests. No genome-wide corrections were performed as these data were analyzed with a hypothesis-driven approach.

### RRBS in cocaine self-administering mice

Bioinformatic processing was done as above, except reads were aligned to the mouse genome (GRCm38/mm10). Rather than methylome-wide analysis, we searched the computed clusters for any that aligned with the *Th* gene (chr7:142,892,752-142,900,014) and found a single cluster within exon 8 that contained 11 CpGs, and was detected in at least 80% of sequenced libraries. After removing statistical outliers using ROUT tests (Q=1%), we performed two-tailed Student's t-tests between control and cocaine animals, separately for the dorsal and ventral striatum.

### Targeted bisulfite amplicon sequencing

After removing statistical outliers (ROUT test; Q=1%), mean methylation between cases and controls was compared using one-tailed Mann-Whitney U or Student's t-tests based on Shapiro Wilks normality which

was determined separately for each comparison. Methylation of the putative EGR1 binding site was analyzed using Two-way, repeated measures ANOVAs with Sidak's multiplicity correction.

#### **Luciferase enhancer assay**

A one-way ANOVA analysis was performed to assess inter group differences for luciferase activities. Post-hoc T test were performed and corrected using Tukey's method to analyze group differences. All the statistical analyses were done using SPSS 20 software (IBM).

#### **qPCR gene expression**

Expression was calculated based on standard curves and normalized against *GAPDH* (Hs02758991\_g1, cat no. 4331182) using QuantStudio Real-Time PCR Software. Normalized expression was compared between cases and controls using two-tailed Student's t-tests in GraphPad Prism 6 (<http://www.graphpad.com/>). The average Ct value for TH in these experiments was 28.7 and the average Ct value for EGR1 was 27.3.