HAL
open science

Cocaine-related DNA methylation in caudate neurons alters 3D chromatin structure of the IRXA gene cluster

Kathryn Vaillancourt, Jennie Yang, Gary Chen, Volodymyr Yerko, Jean-François Théroux, Zahia Aouabed, Alberto Lopez, Kimberly Thibeault, Erin Calipari, Benoit Labonté, et al.

To cite this version:

Kathryn Vaillancourt, Jennie Yang, Gary Chen, Volodymyr Yerko, Jean-François Théroux, et al.. Cocaine-related DNA methylation in caudate neurons alters 3D chromatin structure of the IRXA gene cluster. Molecular Psychiatry, 2021, 26, pp.3134-3151. 10.1038/s41380-020-00909-x . hal-02989971

HAL Id: hal-02989971

https://hal.science/hal-02989971v1

Submitted on 5 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cocaine-related DNA methylation in caudate neurons alters 3D chromatin structure of the *IRXA* gene cluster.

Kathryn Vaillancourt[1, 2], Jennie Yang[1], Gary G. Chen[1], Volodymyr Yerko[1], Jean-François Théroux[1], Zahia Aouabed[1], Alberto Lopez[7], Kimberly C. Thibeault[7], Erin S. Calipari[4,7], Benoit Labonté[4], Naguib Mechawar[1,2,3], Carl Ernst[3] , Corina Nagy[1,2,],Thierry Forne[5], Eric J. Nestler[4], Deborah C. Mash[6] and Gustavo Turecki*[1, 2 ,3]

[1]McGill Group for Suicide Studies, Douglas Hospital Research Center; [2]Integrated Program in Neuroscience, McGill University; [3]Department of Psychiatry, McGill University, Montreal, Quebec, CA; [4]Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY USA; [5]Institute de Génétique Moléculaire de Montpellier, CNRS, Université de Montpellier, FR; [6]Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA;[7]Department of Pharmacology, Department of Molecular Physiology and Biophysics, Department of Psychiatry and Behavioral Sciences, Vanderbilt Center for Addiction Research; Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA.

Keywords: cocaine, addiction, epigenetics, DNA methylation, 3C, 3D chromatin structure, psychiatry, post-mortem brain

Main Figures: 5

Main Tables: 1

Supplementary Figures: 11

Supplementary Tables: 11

Word count: 4189

Abstract

Epigenetic mechanisms, like those involving DNA methylation, are thought to mediate the relationship between chronic cocaine dependence and molecular changes in addiction-related neurocircuitry but have been understudied in human brain. We initially used reduced representation bisulfite sequencing (RRBS) to generate a methylome-wide profile of cocaine dependence in human post-mortem caudate tissue. We focused on the Iroquois Homeobox A (*IRXA)* gene cluster, where hypomethylation in exon 3 of *IRX2* in neuronal nuclei was associated with cocaine dependence. We replicated this finding in an independent cohort and found similar results in dorsal striatum from cocaine self-administering mice. Using epigenome editing and 3C assays, we demonstrated a causal relationship between methylation within the *IRX2* gene body, CTCF protein binding, 3D chromatin interaction, and gene expression. Together, these findings suggest that cocaine-related hypomethylation of *IRX2* contributes to the development and maintenance of cocaine dependence through alterations in 3D chromatin structure in the caudate nucleus.

55

56

57 ## Introduction

58

59 Like other drug use disorders, cocaine dependence is characterized by cycles of binging,
60 preoccupation and compulsive drug seeking behaviors despite negative outcomes[1] . The
61 development and maintenance of dependence-related behaviors, in both humans and animal
62 models, is accompanied by profound alterations in gene expression and lasting changes in
63 cellular plasticity in the mesolimbic dopamine neurocircuitry [2–7]. Accordingly, multiple targets of
64 midbrain dopamine projections display widespread epigenetic alterations, particularly histone
65 post translational modifications, that may mark distinct phases of dependence and withdrawal[7].
66 DNA methylation may act to stabilize dependence-related gene expression programs and it has
67 become an active area of research in cocaine addiction neurobiology.

68 Cocaine-related changes in DNA methylation have primarily been measured at the level of
69 individual gene promoters, although methylome-wide studies have begun to appear in animal
70 models (for review, see [8]). Furthermore, research is just beginning to investigate how changes in
71 methylation are likely to contribute to maladaptive behavioral phenotypes [9,10]. However, the
72 relationship between DNA methylation and cocaine dependence has been understudied in the
73 human brain. Of particular interest is the caudate nucleus, as it appears to be necessary for the
74 development of addiction related drug cravings, and has been implicated in the transition from
75 recreational drug use to dependence [11–14]. Here, we report on the findings of the first
76 methylome-wide study of cocaine dependence in the human caudate nucleus, using post-
77 mortem tissue samples. We also present supporting evidence for a role of cocaine-related gene
78 body methylation of the *IRX2*, a gene located in a region containing the largest cluster of

79 differentially methylated CpGs from our methylome-wide analysis, in regulating local chromatin
80 architecture and expression of two genes in the *IRXA* neurodevelopmental gene cluster.

81 Results

82 ***Chronic cocaine dependence in humans is associated with genome-wide changes in DNA***
83 ***methylation in the caudate nucleus***

84 To understand how DNA methylation patterns in chronic cocaine users may differ from
85 unaffected non-cocaine users in the caudate nucleus (Figure 1a), we performed reduced
86 representation bisulfite sequencing (RRBS) from 25 cases who died from cocaine intoxication
87 that had a lifetime history of cocaine dependence, but no other diagnosed psychopathology,
88 and 25 psychiatrically healthy drug-free controls who died suddenly. RRBS allowed us to
89 interrogate the methylation status of genome-wide loci, while enriching for CpG islands[15]. The
90 groups were matched for commonly confounding factors such as age, post-mortem interval, and
91 tissue pH, with small effect sizes on comparison (ps > 0.1, d < 0.3; Supplemental Table 1), and
92 sequencing statistics were not different between the groups (ps > 0.1; Supplemental Table 2).
93 Since DNA methylation across a region of CpGs is more likely to be biologically relevant than
94 methylation at a single nucleotide, we combined all CpGs within 50bp of another into functional
95 regions. We detected 6712 CpG regions containing at least 2 CpGs (Supplemental Figure 1), and
96 173 clusters were differentially methylated between groups when correcting for ethnicity, age,
97 smoker status and ethanol toxicology (FDR corrected q<0.05; Figure 1b; Supplemental Table 3).

98 Although we detected differentially methylated regions (DMRs) that were both hyper- and
99 hypomethylated in the cocaine group, there were significantly more hypermethylated regions
100 than hypomethylated (Chi Squared Goodness of Fit Tests, χ^2= 26.575; p < 0.05, Figure 1c). Using

101 RNA sequencing data from the same subjects and brain nuclei, we found that transcription of
102 the *de novo* methyltransferase DNMT3a is increased in the cocaine group (t=2.628, df=42,
103 p=0.0120, Supplemental Figure 2a). This finding is in-line with animal studies that have shown
104 that cocaine exposure can induce de novo methyltransferase expression in the striatum[10]. We
105 found no differences in expression of the two other DNA methyltransferase genes, DNMT1 and
106 DNMT3b (ps > 0.05, Supplemental Figure 2b, c). To determine if the corrected DMRs were
107 functionally related, we performed PANTHER gene ontology analysis of genes that either
108 overlapped with, or were in the closest proximity to, the differential methylation signal [16].
109 Although we found no enrichment for genes belonging to any particular cellular component or
110 biological process, which is likely due to unknown long-range target genes, we found a
111 significant enrichment for genes involved in regulatory sequence-specific DNA binding and
112 transcriptional activation (fold enrichment=4.38, FDR=0.0253).

113 Most of the regions, including those which were differentially methylated, mapped onto known
114 CpG islands (Figure 1d; Supplemental Figure 3a), and although they were highly present within
115 gene bodies (introns, exons and intron-exon boundaries), they were significantly enriched for
116 intergenic regions (Figure 1e; Supplemental Figure 3b, Fisher's exact test $q=7.10 \times 10^{-4}$). We used
117 data from the 15-state core model of chromatin states from the Roadmaps Epigenome
118 Consortium[17], which was generated from human caudate nucleus tissues, to annotate the
119 hypothetical chromatin status of our CpG clusters. We found that DMRs were significantly
120 enriched for enhancers (Fisher's exact test $q=1.76 \times 10^{-10}$), regions flanking active transcription
121 start sites (Fisher's exact test $q=2.69 \times 10^{-2}$), and weakly transcribed and quiescent regions
122 (Fisher's exact test $q= 5.26 \times 10^{-6}$ and 1.76×10^{-7} respectively) when compared against the list of
123 all CpG clusters (Supplemental Figure 3c, d). When we assessed the hyper- and hypomethylated

124 DMRs separately, we found that this effect was driven by the hypermethylated loci which were
125 enriched in the same context as the overall list (qs = 1.30 5 x 10^{-4} - 2.46 x 10^{-8}).

126 ***DNA methylation related to gene expression in cis.***

127 To determine whether differentially methylated regions were related to transcriptomic changes
128 in *cis*, we generated RNA-seq data from caudate tissue from the same subjects. We calculated
129 the fold change of all genes within 5kb of a DMR, and took a liberal approach to generate a list
130 of putative DMR-gene pairs with a nominally significant expression difference (uncorrected $p <$
131 0.1) and statistically significant differences in methylation (q value < 0.05). This analysis
132 identified 23 DMR-gene pairs (Table 1).

133 Given our analysis strategy, we rationalized that regions with more CpGs would be most likely to
134 represent strong, biologically meaningful signals. Interestingly, the largest and third largest
135 DMRs overlapped with members of the Iroquois Homeobox (*IRXA*) gene family that are grouped
136 in a highly conserved cluster within vertebrate genomes and were both upregulated according
137 to RNA-seq (Table 1). Iroquois Homeobox 2 (*IRX2*) and Iroquois Homeobox 1 (*IRX1*) are head-to-
138 head neighbours on chromosome 5, and code for transcription factors that are involved in
139 embryonic patterning during neural development [18]. Our RRBS analysis identified 21 CpGs within
140 the third exon of *IRX2* that were 3% less methylated (Figure 2a, Table 1), and 9 CpGs within the
141 second exon of *IRX1* that were 11% more methylated in the cocaine group (Supplemental
142 Figure4a; Table 1).

143 ***Decreased gene body methylation of IRX2 is associated with chronic cocaine dependence***

144 We turned to an independent cohort of dorsal caudate tissue samples from individuals with
145 cocaine dependence (who died by causes other than cocaine overdose) and unaffected controls

146 to replicate the methylation findings (Supplementary Table 4). Using bisulfite amplicon
147 sequencing, we found a significant decrease in methylation of the same region within exon 3 of
148 *IRX2* in the cocaine group compared to controls (5.5%, t=1.908, df=31.83, p=0.033, d=0.631;
149 Figure 2b), but no differences in methylation of the *IRX1* region (t=1.149, df=33, p=0.13,
150 d=0.397; Supplemental Figure 4b). Based on this study, we chose to focus on *IRX2* for further
151 analyses.

152 It is well known that DNA methylation patterns are cell-type specific, and are particularly
153 important in the distinct functions of neuronal and non-neuronal cell types within the central
154 nervous system [19,20]. Thus, we investigated our findings in distinct populations of nuclei,
155 separated from samples of the human caudate from our discovery cohort, using fluorescence
156 activated nuclei sorting (FANS). We separated intact nuclei based on DRAQ5 DNA stain
157 fluorescence, and neuronal nuclei (including D1- and D2-medium spiny neurons, as well as
158 GABAergic and cholinergic interneurons) from non-neuronal nuclei (glial and epithelial cells)
159 based on the nuclear marker NeuN (Supplemental Figure 4a-e). We found the cocaine-
160 associated decrease in *IRX2* methylation to be specific to neuronal nuclei (t=1.923, df=46,
161 p=0.03; Figure 2c). There was no group-wise difference in methylation in non-neuronal nuclei
162 (t=0.3254, df=48, p=0.37; Supplemental Figure 6a).

163 ***Hypomethylation of a putative CTCF binding site***

164 Expression of *IRX2*, as well as other members of the *IRXA* family, is known to be regulated by a
165 large evolutionarily conserved group of enhancers that form a three-dimensional chromatin
166 loop in animals[21]. We hypothesized that cocaine-related methylation might interfere with this
167 regulatory framework, perhaps by impairing the binding of CCCTC-binding factor (CTCF) —a
168 well-studied transcriptional repressor involved in anchoring three dimensional chromatin

169 structures, that can be inhibited by methylation at its binding site [22]. We searched our target
170 sequence for putative CTCF binding sites using its consensus sequence[23], as well as the most
171 highly enriched motifs from chromatin immunoprecipitation data generated by the ENCODE
172 consortium (accessed through FactorBook[24]). We found a known peak motif (3'-AGGGGGCG-5')
173 96 base pairs upstream of our DMR and a putative CTCF consensus sequence (3'-
174 CCGCGGGGCGCGG-5') spanning 4 CpGs within the DMR itself. When considering the
175 methylation state of the consensus sequence separately from the overall region, we found a
176 main effect of cocaine status in both whole tissue homogenates from the replication cohort ($F(1,$
177 $35)=4.333$, $p=0.045$, Figure 2d)) and in neuronal nuclei ($F(1,46)=6.284$, $p=0.016$, Figure 2e)), but
178 not in non-neuronal nuclei from the discovery cohort ($F(1,48)=0.072$, $p=0.78$, Supplemental
179 Figure 6b)). Post hoc comparisons showed the fourth CpG to be significantly less methylated in
180 the cocaine group in the tissue ($t=2.55$, $df=140$, $p=0.012$, Figure 2d) and in neuronal nuclei
181 ($t=3.73$, $df=184$, $p < 0.001$, Figure 2e), but not in the non-neuronal fraction (Supplemental Figure
182 6b).

183 Next, we sought to corroborate our findings in a well-studied mouse model of self
184 administration (Supplemental Figure 7). After removing outliers using ROUT tests (Q=2%), we
185 found a significant effect of group ($W(2,15.7)=6.64$, $p=0.008$, Supplemental Figure 8) where the
186 cocaine group ($n=6$) was significantly less methylated than the saccharin group ($n=9$, $p=0.029$)
187 and nominally significantly hypomethylated compared to controls ($n=14$. $p=0.054$). Notably, this
188 sequence contains the only occurrence of the canonical CTCF binding site (5'-
189 CCGCGCCGCGCGGTGG-3') in the entire 5kB *Irx2* gene, and when we examined the methylation
190 status of the upstream-most CpG, we again found methylation to be lower in the cocaine group
191 compared to both control and saccharin animals ($W(2, 13.52)=10.06$, $p= 0.002$, cocaine vs.
192 control $p=0.032$, cocaine vs. saccharin $p=0.016$, Figure 2f). Importantly, in both analyses,

193 methylation in the control animals did not differ from those who were-trained to self-administer
194 saccharin, which suggests that the cocaine-related hypomethylation is not generalizable to all
195 reward-driven behaviors.

196 We were able to replicate our initial genome-wide significant finding of decreased methylation
197 within *IRX2,* across sample cohorts, tissue types, and species, which is suggestive of a conserved
198 and functionally relevant genomic response. As such, we decided to explore the relationship
199 between *IRX2* methylation and *IRXA* cluster gene expression in our sample set.

200

201 ***Intragenic IRX2 methylation is negatively associated with IRXA gene cluster gene expression***

202 Our genome-wide analyses suggested that gene expression might be disrupted in the *IRXA* gene
203 cluster in relation to cocaine dependence, since the expression of both *IRX1* and *IRX2* was
204 increased according to RNA-seq (Table 1). In order to validate these findings, we used
205 nanoString technology to count the number of *IRX1* and *IRX2* transcripts in RNA extracted from
206 samples in our discovery cohort (n=21 cases and n=23 controls). We found significantly higher
207 expression of *IRX2* in the cocaine group (Mann-Whitney U = 170; p= 0.019; Figure 3a), and
208 although the increased expression of *IRX1* was not statistically significant (t=1.057, df=42;
209 p=0.148; Figure 3b), we found the expression of the two genes to be highly positively correlated
210 in our samples overall (r=0.622; p < 0.0001; Figure3c).

211 We next turned to *in vitro* modelling to explore the relationship between *IRX2* DNA methylation
212 and gene expression of the *IRXA* gene cluster because homogeneous groups of cells allow better
213 resolution than can be obtained through brain tissue homogenates. We measured endogenous
214 methylation and expression levels in two distinct human cell lines; HEK293 kidney epithelial cells

215 (ATCC, Virginia, US) and RENcell immortalized fetal midbrain cells (Millipore, Burlington, US).
216 Since *IRX1* and *IRX2* are neurodevelopmental transcription factors, we hypothesized that their
217 expression would be higher in RENcell neural progenitor cells (NPCs) compared to epithelial
218 cells. We found this to be the case, with RENcells expressing both transcripts, whereas neither
219 transcript was detectable in HEK293 samples (Figure 3d). There are likely multiple epigenetically
220 relevant regulatory elements that contribute to the striking dichotomy in gene expression
221 (alternative promoters and enhancers, for example); however, if the region within exon 3 has
222 regulatory potential, we hypothesized that its endogenous methylation level would differ
223 between the cell types. Indeed, HEK293 cells were on average, 40% more methylated within this
224 region than NPCs ($t=36.76$, $df=4$, $p < 0.0001$; Figure 3e).

225 To determine whether DNA methylation has any causal impact on changes in gene expression,
226 we designed a CRISPR/Cas-9-based epigenome editing experiment in NPCs. We designed three
227 guide RNAs (gRNAs) targeting the region within exon 3, and used a deactivated Cas9 (dCas9)
228 enzyme fused with the active domain of a DNA methyltransferase[25] to experimentally increase
229 methylation and study its regulatory influence on gene expression dynamics. Cells that were
230 transfected with the active construct were on average 4.9% more methylated than wild-type
231 and 7.1% more methylated than the cells that were transfected with a dCas9-DNMT3a plasmid
232 with a mutated methyltransferase domain (inactive) ($F(2,4)=16.9$, $p = 0.011$; Active vs WT $p =$
233 0.021, Active vs Inactive $p = 0.008$), Figure 3f). Importantly, methylation was increased across
234 the CpGs within the CTCF binding site (6.7-8.4%, Supplemental Figure 9a) and unchanged within
235 an amplicon in *IRX1* that was used as a control for off-target methylation ($F(2,6)=3.054$, $p=0.122$,
236 Supplemental Figure 9b). Increased methylation of *IRX2* exon 3 resulted in a significant decrease
237 in *IRX2* ($F(2,6)= 7.928$, $p=0.021$, Active vs. WT $p = 0.018$, Active vs. Inactive $p=0.037$)and *IRX1*

238 (F(2,6)= 8.417, p=0.018, Active vs. WT p = 0.020, Active vs. Inactive p=0.023) gene expression
239 (Figure 3g).

240

241 ***The three-dimensional chromatin structure of IRXA is associated with gene expression, and is***
242 ***altered by methylation of IRX2 exon 3***

243 It has been shown in animal models that *Irx2* and *Irx1* share enhancer elements located within
244 the intergenic region between them, and that the two genes are brought closer together to
245 access these enhancers during transcription[21]; but to date, no such regulatory loop has been
246 identified in humans. We designed a 3C assay in human cells to detect the frequency of physical
247 proximity between the promoter of *IRX1* (viewpoint, Figure 4a), and the genomic region
248 encompassing *IRX2* (test primers, Figure 4a), which are separated by over 850kB of linear
249 genome. We found that, in both neural-progenitor and in kidney epithelial cells, the two genes
250 are in close physical proximity more often than would be expected by chance (dotted grey line,
251 Figure 4a). Interestingly, fragments 2 and 3, which encompass the first two-thirds of *IRX2*, are
252 physically close to the 5' end of *IRX1* more often in NPCs, where both genes are expressed, than
253 in cells where the genes are not expressed (t=4.41-6.54, df=7, ps < 0.01, Figure 4a). The cocaine-
254 associated hypomethylation that we observed, including the putative CTCF binding site, is
255 located within fragment 2, which is further evidence for a relationship between exon 3
256 methylation and 3D chromatin structure.

257 We next sought to understand whether methylation of the CpGs within this fragment could
258 directly cause changes in chromatin architecture, and we again turned to epigenome-editing,
259 this time in HEK293 cells which allowed us to transfect and grow the higher amount of cells
260 necessary for 3C (Supplemental Figure 10). The active plasmid increased the methylation of this

261 region by 4.37% compared to the inactive plasmid and by 9.87% compared to untreated cells
262 ($F(2,6)=15.28$, $p = 0.004$, WT vs Active $p = 0.003$, Active vs Inactive $p = 0.087$, Supplemental
263 Figure 11a). We also found significantly more methylation in the actively transfected cells,
264 compared to WT, when averaging across the entire CTCF binding site (9.91%, $F(2,6)=8.78$,
265 $p=0.017$; Active vs WT $p = 0.012$, Supplemental Figure 11b). There were no significant effects of
266 group on percent methylation of the off-target control (Supplemental Figure 11c).

267 In order to investigate whether these findings could be translated into alterations in long range
268 chromatin structures, we assayed the local chromatin architecture in cells; particularly the
269 frequency with which restriction fragments 2 and 3 interacted with the viewpoint in the *IRX1*
270 gene. Strikingly, we found that methylation of *IRX2* exon 3 brought the interaction frequency of
271 fragment 2, which contains the putative CTCF binding site, down to levels near those expected
272 by chance ($F(2,6)=71.03$, $p <0.0001$, Active vs WT $p <0.0001$, Active vs Inactive $p = 0.003$, Figure
273 4b). We found no significant effect of methylation on the interaction frequency of fragment 3,
274 which does not contain the putative CTCF binding site, with *IRX1* (Figure 4b).

275 ***CTCF binds to IRX2 exon 3, and is disrupted by DNA methylation***

276 Finally, to assess whether methylation of exon 3 could alter CTCF protein binding, we performed
277 anti-CTCF ChIP-qPCR on wildtype HEK293 cells and cultures transfected with either the active or
278 inactive dCas9-DNMT3A construct. Importantly, CTCF binding to the fragment of exon 3
279 containing the putative binding site was experimentally validated to be significantly higher than
280 a non-specific IgG control ($F(1,5)=8.329$, $p = 0.034$; Supplemental Figure 11d). Furthermore,
281 methylating the same sequence decreased CTCF binding compared to wildtype cells ($F(2,5) =$
282 13.19, $p=0.010$; Active vs. WT $p = 0.007$, Figure 4c). These data suggest that exon 3 may indeed

283 contain a functional CTCF binding site that is sensitive to modest changes in cytosine
284 methylation, such as those observed in caudate neurons after chronic cocaine dependence.

285 Based on the cumulation of data from human, mouse, and cell line experiments, we suggest that
286 exon 3 of IRX2 contains a methylation sensitive CTCF binding site that is disrupted following
287 long-term cocaine exposure and dependence (Figure 5).

288 Discussion

289 Our experiments show that chronic cocaine dependence in humans is associated with decreased
290 methylation of an intragenic region of CpGs in the *IRX2* gene, which overlaps with a novel
291 regulatory site for local gene expression and three-dimensional chromatin structure (Figure 5).
292 This region is one of over 100 DMRs that we have identified in the human caudate nucleus.
293 Although this is the first methylome-wide study of cocaine use disorders using brain tissue from
294 human patients, our work is well-aligned with two decades of studies in animals that have
295 identified regions of both hyper- and hypomethylation in addiction relevant neural circuitry[8].

296 The caudate nucleus is increasingly implicated in the pathogenesis of drug use disorders as
297 individuals transition from recreational use to compulsive drug seeking behaviors [13]. The
298 neurons within the human caudate are mostly GABAergic medium spiny projection neurons
299 (MSNs) surrounded by at least 4 distinct types of inhibitory interneurons [26,27]. Striatal MSNs can
300 be classified into two major subgroups, D1 and D2 dopamine receptor expressing cells, which
301 have opposing effects on drug-related behaviors, with D1-MSNs enhancing drug seeking while
302 D2-MSNs inhibit these behaviors in animals[28]. While technical limitations prevent us from
303 discerning the contributions of individual neuronal subtypes, the separation of neuronal from
304 non-neuronal methylation profiles presented here represents the first step towards a human

305 cell-specific cocaine methylome. Future progress in single-cell methylome technologies will
306 undoubtedly guide deconvolution efforts on datasets such as those presented here.

307 Most of the work in the field, to date, has focused on methylation at specific gene promoters,
308 but intragenic methylation and methylation at distal regulatory elements may have relevance to
309 tissue-specific disease etiology. Indeed, the majority of DMRs identified here do not fall within
310 annotated promoters and may disrupt other regulatory processes that contribute to addiction
311 neurobiology. Additionally, although the DMRs in this study are not enriched for any one
312 particular cellular component or biological process, the effects of human chronic cocaine
313 dependence may not impact all cellular pathways equally; further research into epigenetic
314 alterations of specific processes will be a welcome addition to this work.

315 Non-promoter elements are enriched for neuropsychiatric heritability factors, and levels of DNA
316 methylation and chromatin accessibility in these regions has been shown to have brain-region
317 specific effects on disease[20,29,30]. Thus, although DNA methylation is perturbed in multiple
318 addiction-related brain regions, the exact DMRs are likely to differ between brain nuclei.
319 Furthermore, DNA methylation within gene bodies may directly promote gene expression, direct
320 the use of alternative promoters or regulate alternative splicing events [31–33]. In recent years, it
321 has become clear that these biological processes have important implications for psychiatric
322 phenotypes overall [34], and for cocaine dependence where it has been shown that, in the nucleus
323 accumbens, repeated cocaine exposure can induce genome-wide alternative splicing events that
324 are related to drug seeking behaviors in rodents[35,36].

325 Animal work has also identified distinct alterations in methylation and associated machinery
326 that are related to different administration paradigms and exposure time courses. For example,
327 although DNMT3A is initially decreased in the nucleus accumbens during cocaine withdrawal in

328 mice, levels of the *de novo* methyltransferase becomes significantly increased after 28 days[10].
329 Indeed, we found increased DNMT3A expression in the caudate nucleus of our cocaine-
330 dependent samples. Similarly, distinct patterns of differential methylation emerge in studies
331 using passive cocaine injection *versus* self-administration, which are related to long term
332 behavioral changes[37]. Although we are unable to separate the effects of acute and chronic
333 cocaine in our discovery cohort due to positive cocaine toxicology at the time of death, our
334 replication cohort was negative for cocaine metabolites, and suggests that the findings with
335 respect to *IRX2* are more likely linked to long-term dependence than to an acute
336 pharmacological effect.

337 *IRX2* is a transcriptional repressor that is highly expressed during neural development and may
338 be related to social behavior in animals[38,39]. Its expression is known to be regulated by three-
339 dimensional chromatin architecture, which in turn is regulated by the CTCF architectural
340 protein[21,40]. Although this study is the first to report on the relationship between *IRX2* and
341 cocaine dependence, it has already been shown that dependence-related behaviors rely on
342 long-lasting alterations in the expression of transcription factors genes [41,42] and that genes
343 involved in transcription and chromatin regulation are dysregulated in brain tissue from human
344 patients[43]. Cocaine-related expression of transcription factor genes can be regulated by DNA
345 methylation mechanics[44], and are likely cell-type specific[45], which is in line with what we have
346 shown at this locus, where *IRX2* is more highly expressed in the cocaine group. Although the
347 downstream targets of IRX2 regulation have yet to be experimentally identified, target
348 prediction algorithms suggest that it may impact the expression of genes including *ADAM10,* a
349 metalloprotease that has been linked to multiple psychiatric diseases, and may be involved in
350 the cognitive impairments that can accompany long term psychostimulant use[46–48] . Additionally,
351 little is known about the dynamics of epigenetic regulation in the *IRXA* gene cluster during

352 neuronal development — future work in animals should identify how the relationship between
353 methylation, expression, and three-dimensional chromatin structures changes throughout
354 development.d

355 We have shown that cocaine-related methylation of *IRX2* exon 3 is negatively associated with
356 gene expression through decreased frequency of three-dimensional chromatin structure. This is
357 in line with evidence that suggests that DNA methylation can compete with CTCF binding,
358 especially at specific CpGs at key regulatory sites[22,49]. Moreover, repeated cocaine
359 administration has been shown to increase DNA methylation and decrease CTCF-mediated
360 chromatin looping at the *Auts2-Caln* locus in mice[9].

361 Like other work using post-mortem samples, this study presents limitations that need to be
362 considered when interpreting the findings. First, the molecular profiles gathered from these
363 tissues highlight the epigenetic landscape immediately prior to death, and although every effort
364 is made to characterize the demographic information of the donors, we are unable to account
365 for corollary factors such as lifestyle, and lifetime history of non-dependent drug exposure that
366 could influence DNA methylation. Complementary evidence from animal models, such as has
367 been presented here, can begin to account for the effect of extraneous factors. Similarly,
368 although we were able to distinguish between broad categories of cell-types (neurons vs. non-
369 neurons), the magnitude of methylation differences observed in our study suggest that the
370 signal is coming from a relatively rare cell type and being masked by cellular diversity.
371 Nonetheless, small changes in methylation have been shown to have physiologically relevant
372 effects on transcription factor binding, and RNA transcription[50,51], and have previously been
373 associated with cellular and molecular alterations in post mortem psychiatric research[52–54].

374 Future work on the *IRX2* locus, as well as other DMRs identified in this study, should incorporate
375 information about additional levels of epigenetic regulation, including histone modifications as
376 well as DNA modifications outside of the canonical CpG methylation context. For example, non-
377 CpG methylation (CpH), N6-Methyladenosine (m6A) and hydroxymethylation are epigenetic
378 regulators that are highly abundant in the brain and are likely be important mechanisms to drug
379 dependence[55–57]. Additionally, direct manipulation of *Irx2* in animals will allow important insight
380 into the behavioral consequences of cocaine-related epigenetic changes. Furthermore, although
381 studies suggest that dependence to other psychostimulants, including amphetamine, associates
382 with DNA methylation changes, direct comparisons between drugs of abuse, and between
383 addiction-related brain regions, will add specificity to epigenome-wide studies. Additionally, as
384 data from single-cell epigenomic experiments continue to become available, researchers will be
385 able to detect differences in rare cell types that are currently masked by bulk and near-bulk
386 tissue experiments[58].

387 # Methods

388 **Subjects**

389 All methods used in this study were approved by the Douglas Hospital Research Ethics Board,
390 and written informed consent was obtained from the next-of-kin for each subject. Autopsy and
391 tissue sampling were performed in accordance to the established standards of the University of
392 Miami Miller School of Medicine, or the Douglas-Bell Canada Brain Bank, depending on cohort
393 source location.

394 Post-mortem caudate nucleus tissues from our discovery cohort were obtained from the Brain
395 Endowment Bank at the University of Miami Miller School of Medicine (Supplemental Table 1).
396 Samples were dissected from the dorsolateral sector of the caudate from 25 subjects who had

397 long term histories of cocaine dependence as determined by licenced clinicians, and who died
398 from cocaine related complications as determined by forensic pathology and brain and blood
399 toxicology. These subjects were selected based on the absence of toxicology for illicit drugs
400 other than cocaine and were determined to have no other psychiatric diagnoses based on
401 medical records and the reports of next-of kin. Drug-naïve, psychiatrically healthy control
402 subjects (n=25) were selected from accidental or natural deaths. All subjects in this cohort were
403 male, which is reflective of the opportunistic composition of the brain samples available at
404 autopsy.

405 Caudate tissue from the replication cohort of 15 cases and 21 controls was obtained from the
406 Douglas-Bell Canada Brain Bank (www.douglasbrainbank.ca). Subjects underwent a medical
407 chart review and proxy-based interviews that were used in the characterization of substance
408 use, which was determined through psychological autopsy by the clinical staff of the brain bank.
409 Case status was determined based on these results, as well as toxicology at the time of death.
410 Age, PMI and pH did not significantly differ between groups (t=0.79-1.73, df=34, p=0.09-0.44,
411 Supplemental Table 4). Grey matter was dissected from the left hemisphere of all samples and
412 stored at -80°C until further processing.

413 **Reduced representation bisulfite sequencing (RRBS)**

414 *Tissue and Library Preparation*

415 We extracted DNA from 20mg of frozen tissue of all cases and controls from our initial cohort
416 using Qiagen DNA MiniKits as per manufacturer's instructions. To prepare RRBS libraries, we
417 digested 1ug of genomic DNA with *MspI* restriction enzyme, repaired the fragment ends and
418 ligated Illumina adapters as described in elsewhere[59]. Purified libraries were treated with

419 EpiTect fast bisulfite conversion kit (QIAGEN, Cat# 59824) according to the standard protocol
420 and indexed through PCR amplification.

421 *Sequencing and Bioinformatic Processing*

422 Final libraries were sequenced on the Illumina HiSeq 2000 platform at the Genome Quebec
423 Innovation Center (Montreal, Canada) using 50bp single end sequencing, and bioinformatics
424 processing was performed in-house, as described[59]. Bisulfite conversion efficiency was
425 determined by the ratio of T to C at the unmethlyated cytosine position added during the end-
426 repair step of library construction.

427 *Differential Methylation Analysis*

428 We defined methylation region as any CpG within 50 bp of another CpG, with no limit on the
429 number of CpGs in a given region, but with a minimum of at least 2 CpGs using the bumphunter
430 3.5 package for R. For CpGs to be included in the analysis, they must have been present in at
431 least 25 subjects from both cases and controls and have $\geq$5X coverage, which resulted in 270191
432 CpGs that went into clustering analysis. CpG regionsthat had a standard deviation <5%
433 methylation across all subjects (*i.e.*, irrespective of status) were removed to avoid comparisons
434 between stable methylation sites. For each cluster, we performed differential analysis using a
435 general linear model with status (cocaine or control) as a fixed factor, and age, ethnicity,
436 smoking status and ethanol toxicology as covariates. We treated CpGs independently in a given
437 regions and used only those regions that had a Benjamini-Hochberg FDR corrected *p*-value <0.05
438 and which were <0.05 when calculating a single mean from all CpGs per individual.

439 *Cluster Annotation and Enrichment Analyses*

440 All CpG regions were annotated relative to their genomic context, their CpG island proximity,
441 and their predicted ChromHMM chromatin state using the *annotatr 1.10.0* package in R[60]. We
442 calculated enrichment q values for DMRs against all CpG clusters using the *LOLA* algorithm[61].

443 *Gene Ontological Analysis*

444 We annotated each DMR to its nearest Refseq gene. Gene ontologies were examined with over
445 representation tests in the gene list analysis functions of the PANTHER classification system
446 (www.pantherdb.org). We compared the full DMR list to all human genes with respect to
447 molecular function, biological processes and cellular components and *p*-values were calculated
448 using Fisher's Exact tests with FDR correction.

449

450 **RNA sequencing**

451 RNA was extracted from 100mg dissections of dorsolateral caudate nucleus using RNeasy Lipid
452 Tissue Kits (Qiagen) according to standard procedure. The RNA integrity number (RIN) for the
453 cases was 7.9 ± 1.28 and controls was 8.4 ± 0.78 (mean ± s.d.), and aliquots of 100ng/ul of RNA
454 were sent for library preparation and sequencing at the Broad Institute (Cambridge, MA).
455 Libraries were prepared using a standard non-strand specific protocol (Illumina TruSeq),
456 including poly-A selection, and multiplexed for 50bp paired end sequencing on the Illumina
457 HiSeq 2000 platform. Sequencing data was processed as previously described[62], and fold
458 changes and *p*-values were generated by student's t-tests. Uncorrected *p*-values of <0.1 were
459 used to create DMR-gene pairs for follow-up investigation.

460 **nanoString gene expression validation**

461 For count based RNA quantifications, custom 100bp probe sequences were designed to uniquely
462 capture the majority of transcript variants of the genes indicated in Supplemental Table 5, by
463 nanoString Technologies (Seattle, Washington). Each probe was associated with a unique
464 fluorescent barcode, and 20ng/ul of total RNA, from the same extraction that was used in
465 sequencing, was run on the nCounter system under the high field of view setting at the Lady
466 Davis Institute (Montreal, Quebec). All normalization and statistical analyses were performed
467 with the nSolver software from nanoString technologies. Raw probe counts were normalized to
468 4 negative control probes and then compared between groups, using unpaired student's *t*-tests
469 or their non-parametric equivalent when necessary. Two samples were removed from each
470 group due to technical failure, and statistical outliers were removed after ROUT analysis (Q=1%)
471 resulting in 23 controls compared to 21 cases in the *IRX2* analysis and 22 cases and controls in
472 the *IRX1* comparison.

473

474 **Fluorescence Activated Nuclei Sorting**

475 *Nuclear extraction and labelling*

476 In order to liberate intact nuclei from the caudate nucleus tissue samples , we homogenized
477 50mg of frozen tissue in nuclei buffer containing 10mM PIPES (pH 7.4), 10mM KCl, 2mM $MgCl_2$,
478 1mM DTT, 0.1% TritonX-100 and 10X Protease Inhibitor Cocktail (Sigma Aldrich, Darmstadt,
479 Germany). Homogenates were passed through a 30% sucrose gradient in nuclei buffer in order
480 to separate nuclei from cellular debris, then after a wash with nuclei buffer, nuclei pellets were
481 resuspended in blocking buffer containing 0.5% bovine serum albumin in 10X normal goat
482 serum. Each sample was co-incubated with the DNA labelling dye DRAQ5 (1:300) (ThermoFisher,
483 Waltham, MA) and an anti-NeuN-PE antibody (1:300) (cat no. FCMAB317PE, Millipore,

484 Darmstadt, Germany) for 60 min at room temperature, then passed through 40uM filter caps to
485 remove any remaining cellular debris before sorting.

486 *Nuclei Sorting*

487 Labelled nuclear extracts were processed the BDFACSAria III platform (BD Biosciences, San Jose,
488 CA) according to technical specifications provided by the company. We used BD FACSDIVA
489 software (BD Biosciences, San Jose, CA) to first isolate single, intact nuclei based on DRAQ5
490 fluorescence at the 730/45-A filter (DRAQ5), then to sort neuronal from non-neuronal nuclei
491 based on fluorescence detected by the 585/42 filter (PE). Sorted nuclear fractions were stored
492 at -20°C in sheath fluid (1X PBS) until DNA extraction. On average, we isolated 180 000 NeuN+
493 nuclei and 492 500 NeuN- nuclei from 50mg of tissue, with roughly 37% NeuN+ in each sample.
494 There were no differences between cases and controls in terms of total nuclei in either fraction
495 or in the ratio of neuronal to non-neuronal nuclei captured per dissection (ps > 0.1,
496 Supplemental Figure 3b-d).

497 *Nuclear DNA extraction and processing*

498 We incubated nuclear fractions with 50X protease (Qiagen, Montreal, Canada) at 56°C for at
499 least 12 hours to ensure thorough digestion of the nuclear membranes. Liberated DNA was
500 precipitated onto 0.2X Agencourt AMPure XP beads (Beckman Coulter, Brea, CA) after adding
501 20% PEG-8000 2.5M NaCl to increase the final PEG concentration to 10%. The beads were
502 washed twice in a magnetic stand with 70% EthOH, and then DNA was eluted in 50ul MilliQ H_2O.
503 We measured the concentration of each DNA sample using Quant-iT PicoGreen dsDNA assays
504 (ThermoFisher, Waltham, MA) according to manufacturer specifications.

505 **<u>Bisulfite amplicon sequencing</u>**

506 *DNA extraction and conversion*

507 We obtained genomic DNA from post-mortem, homogenate tissue samples from both the
508 discovery and replication cohorts using QIAmp DNA Mini Kits (Qiagen, Montreal, Canada) as per
509 manufacturer specifications. Notably, DNA for the discovery cohort was extracted from the
510 initial dissections that were used for RRBS library construction. Genomic DNA from both cohorts
511 (2ug/sample), as well as from all sorted nuclear fractions (>100ng/sample) was converted using
512 EpiTect Fast 96 Bisulfite Conversion Kits (Qiagen, Montreal, Canada), diluted to 150ul using
513 MilliQ water, and stored at -20 °C.

514 *Library Preparation*

515 To optimize our ability to cover the desired CpGs within each amplicon, and to increase
516 amplicon diversity for sequencing, we designed three pairs of bisulfite specific primers per DNA
517 strand using Methyl Primer Express Software v1.0 (Applied Biosystems, CA, USA). Redundant
518 primers were designed to be non-overlapping, to have optimal melting temperatures of 60± 2°C
519 and to be between 18 and 24bp long in order to optimize amplification in a multiplexed reaction
520 (primer sequences and PCR conditions in Supplementary Tables 6 and 7). We amplified each
521 sample using 10ul reactions consisting of 5X combined primers (10uM), 3X bisulfite converted
522 DNA and 2X KAPA HiFi HotStart Uracil+ ReadyMix (Kapa Biosystems, MA, USA). Each strand was
523 amplified separately, and after two rounds of paramagnetic bead purification at 0.8X, amplicons
524 from both strands were combined and amplified for 10 additional cycles to add custom primer
525 sequences in 20 ul reactions consisting of 2.5X sample, 5X combined CS1 and CS2 primers
526 (10uM) and 2X KAPA HiFi HotStart ReadyMix (Kapa Biosystems, MA, USA). After an additional
527 round of 0.8X bead purification, we indexed each sample for 10 cycles in a 20ul reaction
528 consisting of 2.5X amplicons, 5X indexing primers (10uM) and 2X KAPA HiFi HotStart ReadyMix

529 (Kapa Biosystems, MA, USA). Each indexed library went through two rounds of double ended
530 bead purification (final ratio 0.8X) to select only those fragments in the predicted range of our
531 amplicons (400-700bp). Final library concentrations and quality control was performed on the
532 Agilent 2200 TapeStation (Agilent Technologies, CA, USA) before samples were pooled and
533 sequenced.

534 *Next Generation Sequencing*

535 We pooled libraries to a final concentration of 2nM and included a 5-10% PhiX spike-in control
536 for each sequencing run. Final libraries were run on the Illumina MiSeq platform (Illumina, San
537 Diego, CA) using customized 300bp paired end sequencing as described elsewhere[63]. All quality
538 control and read alignment, without removing duplicates, were performed in-house and
539 methylation was calculated as the percent of reads containing cytosine rather than thymine at
540 each position.

541 *Statistical Analysis*

542 Samples were removed for poor sequencing (less than 5X coverage of CpGs or less than 80% of
543 CpGs covered within an amplicon) and the number of samples used in each analysis is reflected
544 by the respective degrees of freedom. Methylation at each position was defined as the number
545 of reads called as cytosine, over the number total number of reads. Percent methylation was
546 averaged across all CpGs within an amplicon and compared between groups using unpaired *t*-
547 tests or Mann-Whitney U tests where groups were unlikely to be normally distributed as per
548 Shapiro-Wilks test. Two-tailed significance tests were used except for analyses with sorted
549 nuclei fractions and mouse samples, where a priori hypotheses allowed for one-tailed testing.
550 Statistical outliers were removed according to ROUT analysis (Q=1%), and final sample sizes are
551 reported in the corresponding figure legends.

552 For analysis of methylation at individual CpGs within CTCF binding sites, ordinary two-way
553 ANOVAs were performed with Group and CpG as factors. Main effects of group were dissected
554 using t tests and the Holm-Sidak method for multiple testing corrections.

555 **Mouse reward self-administration**

556 *Jugular catheter implantation:*

557 Mice were anesthetized with ketamine (100mg/kg) and xylazine (10mg/kg), implanted with
558 chronic indwelling jugular catheters, and trained for i.v. self-administration as previously
559 described[64]. The catheter tubing was passed subcutaneously from the back to the jugular vein
560 and 1.0 cm of tubing was inserted into the vein and secured with silk suture. Catheters were
561 flushed daily with ampicillin (0.5mg/kg) and heparin (10U/mL) solution in sterile saline (0.9%
562 NaCl). Mice recovered >3d before commencing behavioral training. All animals were maintained
563 on a reverse light cycle (7:00am lights off; 7:00 pm lights on) and behavioral training was
564 conducted during the animal's dark cycle.

565 *Cocaine self-administration*

566 Mice were trained to self-administer cocaine as previously described. Briefly, mice were
567 maintained at ~90% of their free-feeding weight and trained in standard mouse operant
568 chambers (Med Associates, St Albans, USA) equipped with a white noise generator and two
569 illuminated nose-pokes. Start of daily sessions (2h) were signaled by white noise. For each task,
570 one nose-poke was designated as the “active poke” and the other designated as the “inactive
571 poke”. Active nose pokes resulted in a cocaine infusion (0.5mg/kg/inj, 3 sec) with a concurrent
572 presentation of the Active nose poke light for 5 seconds. Inactive pokes resulted in no
573 programmed consequences. Mice were trained to self-administer cocaine for 10 consecutive

574 days under a fixed-ratio (FR1) schedule of reinforcement. For cocaine self-administering animals,
575 acquisition (Day 1) was counted when the animal reached 70% responding on the active lever
576 and 10 or more responses. Control animals underwent the same experimental procedures but
577 had access to a saline-paired nose poke. For both the cocaine and saccharin self-administration
578 groups, most mice reached acquisition criteria on the first day, and since the experimental
579 animals were trained for 10 days, all saline animals underwent 10 days of saline self-
580 administration as controls.

581 *Saccharin Self-Administration*

582 8-week old c57BL/6J mice were ordered from the Jackson Laboratory and were housed in a 12-
583 hour 6:00/6:00 reverse dark/light cycle. Saccharin self-administration was run during the
584 animal's dark cycle and mice were food restricted to ~90% of free-feeding weight with water
585 provided ad libitum. Mice were trained to self-administer saccharin (0.1% solution in water,
586 ~80uL/infusion) or water for 1 h over 10 consecutive days. Briefly, activation of white noise
587 signaled the initiation of the (1h) daily session. In each task, one nose-poke was designated the
588 "active" while the other was designated the "inactive". Active responses initiated a 1 sec
589 saccharin (or water) delivery into an accessible dipper with a concurrent 5 sec presentation of
590 both the nose poke light and dipper light under a Fixed-Ratio 1 (FR1) schedule of reinforcement.
591 Inactive nose pokes had no programmed consequences but were recorded throughout all
592 behavioral sessions.

593 *Tissue preparation*

594 After the 10th test session, animals were euthanized, and brain tissue was removed and flash
595 frozen at -80°C before the caudate-putamen was dissected.

596 *Library Preparation*

597 After standard extraction using, 500ng of genomic DNA from each sample was bisulfite
598 converted using the EZ DNA Methylation Gold kit (Zymo Research, Irvine, California). Bisulfite
599 specific primers were designed as above, using genomic regions homologous to the hg19
600 coordinates in the mouse genome (mm10) (Supplemental Table 6). Bisulfite DNA was subjected
601 to the same amplification (Supplemental Table 7) and purification methods as described above,
602 and the final purified libraries were spiked into a 300bp paired end sequencing run containing
603 customized sequencing primers as above. After de-multiplexing and adapter trimming,
604 sequencing reads were aligned to the mm10 mouse genome, and percent methylation at each
605 position was determined as above.**Cell Culture**

606 Human embryonic kidney cells (HEK293) were maintained in eagle's minimum essential medium
607 (EMEM; ATCC, Virginia, US) supplemented with 10% fetal bovine serum (FBS; Gibco
608 Laboratories, Gaithersburg, MD), 100 I.U./mL penicillin and 100ug/mL streptomycin. Cells were
609 cultured at 37°C in a humidified incubator with 5% CO_2.

610 Human neural progenitor cells (ReNcell; Millipore, Burlington, US) were maintained in STEMdiff
611 neural progenitor medium (STEMCELL Technologies, Vancouver, Canada) supplemented with
612 100 I.U./mL penicillin and 100ug/mL streptomycin.

613 **Quantitative PCR**

614 The expression of *IRX1* and *IRX2* in all *in vitro* experiments was determined using quantitative
615 reverse transcription PCR (RT-qPCR). We used pre-designed probe based assays for *IRX1*
616 (Hs.PT.58.2400) and *IRX2* (Hs.PT.58.24473971) with FAM-TAMRA dyes (PrimeTime, IDT,
617 Hampton, New Hampshire, Supplemental Table 8), and ran 10ul assays on an Applied

618 Biosystems QuantStudio 6 instrument under default cycling conditions (ThermoFisher,
619 Hampton, New Hampshire). Amplification curves were normalized using QuantStudio Real-Time
620 PCR Software, and the average expression between treatment conditions was compared with
621 two-tailed t tests in GraphPad Prism 6 (www.graphpad.com).

622 **dCas9 epigenome editing**

623 *Plasmid preparation*

624

625 The pdCas9-DNMT3A-PuroR and pdCas9-DNMT3A-Puro (ANV) plasmids were a gift from Vlatka
626 Zoldoš (Addgene plasmids #71667 and #716840). Guide RNA (sgRNA) sequences targeting
627 human *IRX2* exon 3 were designed using an online tool (CRISPR gRNA Design Tool,
628 www.ATUM.bio) and purchased from Integrated DNA Technologies (IDT, Iowa, US). All 3 sgRNAs
629 were annealed and cloned into the expression plasmids through the *BbsI* restriction site. All
630 cloned constructs were confirmed by Sanger sequencing using the U6 sequencing primer
631 (Supplemental Table 9).

632

633 *Transfection*

634 HEK293 cells were seeded in 10cm culture dishes and transfected the next day, at over 80%
635 confluence, using TransIT-293 transfection reagent (Mirus, Brampton, ON). Transfections were
636 done with a pool of 15ug dCas9-DNMT3A plasmids carrying all three sgRNA targeting *IRX2*. The
637 experiment was performed in triplicate and constructs carrying inactive DNMT3A were used as
638 negative controls. 48h after transfection, cells were selected with 1.6ug/mL puromycin (Gibco
639 Laboratories, Gaithersburg, MD) for another 48h. Cells were harvested after seven days

640 transfection for DNA and RNA extraction (see above) as well as the nuclei preparation (see
641 below).

642 RENcells were tranfected by electroporation using the Neon transfection system (Invitrogen,
643 Carlsbad, CA), according to the manufacturer's standard procedure. Briefly, cells were washed in
644 PBS, detached from the culture vessel using accutase, pelleted by centrifugation and
645 resuspended in Resuspension Buffer R at a final density of 13 x 10^6 cells/mL. Cells were
646 immediately electroporated three times using 100uL neon tips at voltage 1300, width 20, and
647 pulse 3, giving final 15ug DNA and 4 x 10^6 cells per 10cm culture dish. After electroporation, cells
648 were immediately transferred into the prepared 10cm culture dish containing prewarmed
649 medium, but without antibiotics. Cells were selected with 0.3ug/mL puromycin with the same
650 time period as in HEK293 and harvested at Day 7. Percent methylation was assessed with
651 bisulfite amplicon sequencing, and gene expression was assessed by qPCR as described above.
652

653 *Statistical Analyses*

654 Percent methylation, gene expression and interaction frequency were compared between
655 conditions using one-way ANOVAs with Tukey's post-hoc comparisons.

656 **Chromatin Immunoprecipitation (ChIP-qPCR)**

657 *Chromatin Preparation*

658 HEK293 cells were grown, transfected, and nuclei were harvested as above. Aliquots of 5 million
659 nuclei each were prepared from wildtype cells, cells transfected with the active dCas9-DNMT3A
660 + gRNA construct (active group), and cells that were transfected with the mutant dCas9-
661 DNMT3A construct (inactive group). Pelleted nuclei were resuspended with 1mL filtered 1X PBS
662 and 200X Protease Inhibitor Cocktail, and cross-linked with 1% formaldehyde for 7 minutes, at

663 room temperature, with gentle rotation. After the fixation reaction was stopped with 10X
664 glycine, nuclei were pelleted and washed twice with 1XPBS+PIC and resuspended in Sonication
665 Nuclear Lysis Buffer, and chromatin was prepared as per the manufacturer's protocol
666 (SimpleChIP Plus Sonication Chromatin IP Kit, Cell Signalling Technologies, Massachusetts).
667 Chromatin was sheared on a S220 focused ultrasonicator (Covaris, Massachusetts) under the
668 following conditions: 150 peak power, 200 cycles/burst, duty factor 10, for 15 minutes.

669 *Immunoprecipitation*

670 Diluted "input" chromatin was removed before immunoprecipitation, and anti-IgG was added to
671 chromatin from 1 x106 nuclei as per manufacturer protocol (SimpleChIP Plus Sonication
672 Chromatin IP Kit, Cell Signalling Technologies, Massachusetts). Anti-CTCF antibody (1:25 dilution,
673 cat no. 2899, Cell Signalling Technologies, Massachusetts) was added to the sonicated chromatin
674 of 4 x 10^6 nuclei and both anti-IgG and anti-CTCF preparations were precipitated overnight at
675 4°C with rotation. Chromatin-antibody complexes were separated with magnetic protein G
676 beads, eluted, and de-crosslinked by proteinase K digestion at 65°C overnight (16 hours).

677 *qPCR*

678 DNA from anti-CTCF, anti-IgG and input fractions was purified per manufacturer's instructions
679 (SimpleChIP® DNA Purification Buffers and Spin Columns Kit, Cell Signalling Technologies,
680 Massachusetts), 10ul SYBRGreen qPCR reactions were prepared using the primers provided in
681 Supplementary Table 10, and 3 technical replicates were amplified using an Applied Biosystems
682 QuantStudio 6 instrument under default cycling conditions (ThermoFisher, Hampton, New
683 Hampshire).

684 *Data normalization and analysis*

685 Ct values were averaged across technical replicates and the amplicon enrichment was calculated
686 with the following formula: $E^{\wedge(Ct_{input}-Ct_{IP})}$ where E= log(efficiency) of each primer pair. These data
687 were then multiplied by the input dilution factor to obtain "percent of input". For each sample,
688 the percent input for the test amplicon was normalized to the percent input generated from the
689 amplification of a control amplicon, located upstream of the exon 3 target, by making a ratio.
690 These values were then analysed using Kruskal-Walis test (one outlier was removed from the
691 wildtype group for having a 10-fold difference in normalized percent input) followed by Dunn's
692 multiple comparisons test, comparing the wildtype and inactive groups to the actively
693 methylated group.

694 **Chromatin Conformation Capture (3C-qPCR)**

695 *3C library preparation*

696 In order to assess three dimensional chromatin structure, we used chromatin conformation
697 capture, followed by quantitative PCR (3C-qPCR), using the protocol from Ea and colleagues[65].
698 Nuclei were extracted from 5-10 million cells using a sucrose gradient, crosslinked in 1%
699 formaldehyde for 10 minutes, and quenched in 125mM glycine before centrifugation and
700 resuspension in a restriction enzyme compatible 3C buffer. Samples were digested overnight at
701 37°C with 450U of high concentration EcoRI (Promega, Wisconsin) shaking at 200rpm, and then
702 diluted in 4ml of ligation buffer to promote intramolecular ligation and prevent chromatin
703 tangles. We ligated the samples with 195U high concentration T4 ligase (Promega, Wisconsin)
704 overnight at 16°C, and then extracted the ligation products using standard proteinase K
705 digestion, phenol chloroform extraction, and ethanol precipitation with the addition of 1ul
706 glycogen. In order to prevent circularization or coiling of ligation products, we performed a

707 complementary digestion using 100U of BglI (ThermoFisher, Waltham, MA), followed by phenol-
708 chloroform extraction and ethanol precipitation of the final 3C libraries.

709 *3C library quality control*

710 Digestion efficiency can have a crucial impact on the outcome of 3C-based assays and as such,
711 we assessed the digestion efficiency of each cut site within our experiment by comparison with
712 an undigested control (UND) taken from cross linked chromatin, and a digested control (DIG)
713 taken after EcoRI digestion but before ligation. We purified the UND and DIG control samples
714 using proteinase K digestion, phenol chloroform extraction and ethanol precipitation, and then
715 performed a 2-hour BglI digestion at 37°C. Using primers designed to span across each
716 restriction site (R; Supplemental Table 11), we performed SybrGreen qPCR on the UND and DIG
717 fractions for each sample (PowerUp SYBR Green Master Mix, ThermoFisher, Waltham, MA). To
718 control for differences in the amount of starting material, we also amplified each fraction using
719 primers designed for region within *GAPDH* that does not contain an EcoRI, nor a BglI cut site (C,
720 Supplemental Table 11). We calculated the restriction digestion efficiency for each restriction
721 site using the following formula: % Efficiency = $100 – 100/2^{((Ct_R-Ct_C)_{DIG}-(Ct_R-Ct_C)_{UND})}$; we
722 excluded all samples whose efficiency, averaged across all restriction sites, was less than 70%.

723 *PCR Control Template Library*

724 In order to determine the minimum concentration of 3C library needed for each qPCR reaction,
725 and to compare relative interaction frequencies between primer pairs, we generated a control
726 template library containing all possible ligation fragments, across our region of interest, in
727 equimolar concentrations. To do this, we obtained two human BACS (RP1182M24 and
728 RP11596I24; ThermoFisher, Waltham, MA) and combined them in equimolar concentration. We
729 digested the BAC pool with EcoRI for two and a half hours at 37°C, and then ligated with T4

730 ligase overnight at 16°C. After phenol chloroform extraction, we performed BglI digestion at
731 37°C for two hours and then purified with phenol chloroform and ethanol precipitation for a
732 final time. We made five-fold serial dilutions of this template library in 25ng/ul BglI digested
733 gDNA, starting at a concentration of 25ng/ul to mimic the behavior of the 3C libraries. For each
734 primer pair in our assay, we performed qPCR using these serial dilutions (see below) and
735 obtained a standard curve with a slope (b) and intercept (a) that were used to normalize the Ct
736 values of our samples.

737 *3C-qPCR*

738 Prior to measuring the amount of each ligation product in our 3C libraries, we measured the
739 concentration of each library by SybrGreen PCR using primers for the non-digested site within
740 *GAPDH*. We then adjusted the concentration of our libraries to 25ng/ul and re-measured the
741 concentration to ensure accuracy. The final concentration values were used as loading control
742 values during normalization.

743 We designed our qPCR assay to cover the ~1Mb region encompassing both *IRX2* and *IRX1* (chr5:
744 2744845-2752662, primers in Supplemental Table 10). We designed a constant reverse primer
745 on the 220th EcoRI digestion fragment, which overlaps the promoter region and TSS of *IRX1*, as
746 determined by in silico digestion. This fragment also bound to a custom PrimeTime Probe
747 (Integrated DNA Technologies, Iowa) that was positioned between the constant primer and the
748 EcoRI cut site and had a 5' FAM fluorescent dye and a 3' TAMRA quencher. We designed 12 test
749 primers, adjacent to the cut sites of fragments concentrated around both *IRX* genes
750 (Supplemental Table 11), such that the resulting amplicons would be between 100bp and 150bp
751 when paired with the constant reverse primer. We determined the Ct value for each ligation
752 product in 10ul reactions run in triplicate, using 2X TaqMan Master Mix, on a QuantStudio 6

753 instrument using QuantStudio Real-Time PCR Software (ThermoFisher, Hampton, New
754 Hampshire).

755 *Data Normalization and Analysis*

756 To account for differences in efficiency between primer pairs, we first normalized our
757 experimental Ct values to the standard curve obtained for each reaction using our control
758 template library as follows: Normalized Ct_1 = 10^((Ct-b)/a). In order to account for variation in
759 the amount of template, we normalized each Ct_1 to the concentration of the input library (Ct_2 =
760 Ct_1/loading control), and these values were used as the relative interaction frequency between
761 each fragment and the constant. Finally, given that the ends of each hybrid ligation fragment
762 originate from the same DNA molecule, we calculated the basal interaction level (BIL) using the
763 procedure as previously described[66] and defined it as the relative frequency of interaction that
764 would be expected by chance ("random collisions"). Using the BIL for each library, and the
765 standard error of the mean of these values, we determined the "noise band" where any
766 observed interaction within this range would be attributed to chance and not a biologically
767 meaningful effect.

768 We used student's t tests to compare the average relative interaction frequencies between
769 groups for each fragment and used the Holm-Sidak method to correct for multiple comparisons
770 across each 3C experiment.

771

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 Douglas Bell Canada Brain Bank, and the bioinformaticians who have worked on these data (A.D., R.P., and A.B.).

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., R00 DA04211 to ESC, and P01 DA047233 to E.N., and grants from the Whitehall Foundation, the Edward Mallinckrodt Jr, Foundation, and the Brain and Behavior Research Foundation to ESC.

Author Contributions

Manuscript preparation: K.V., Experimental design and data collection: K.V., J.Y., G.G.C., Data Analysis: K.V., C.E., T.F., J-F.T., Z.A., Animal experiments: A.L, K.C.T., B.L, Resources and support: E.N., E.C., C.N.,D.C.M., G.T.

Competing Interests

The authors declare no competing interests.

References

1. American Psychiatric Association (APA). *Diagnostic and Statistical Manual of Mental Disorders*. (American Psychiatric Association (APA), 2000).

2. Freeman, W. M. *et al.* Gene expression changes in the medial prefrontal cortex and nucleus accumbens following abstinence from cocaine self-administration. *BMC Neurosci.* **11**, 29 (2010).

3. Albertson, D. *et al.* Gene expression profile of the nucleus accumbens of human cocaine

794 abusers: evidence for dysregulation of myelin. *J. Neurochem.* **88**, 1211–1219 (2004).

795 4. Bannon, M. J. *et al.* Identification of long noncoding RNAs dysregulated in the midbrain of
796 human cocaine abusers. *J. Neurochem.* **135**, 50–59 (2015).

797 5. Mash, D. *et al.* Gene expression in human hippocampus from cocaine abusers identifies
798 genes which regulate extracellular matrix remodeling. *{PloS} one* **2**, (2007).

799 6. Zhou, Z., Yuan, Q., Mash, D. C. & Goldman, D. Substance-specific and shared transcription
800 and epigenetic changes in the human hippocampus chronically exposed to cocaine and
801 alcohol. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6626–31 (2011).

802 7. Nestler, E. J. Epigenetic mechanisms of drug addiction. *Neuropharmacology* **76 Pt B**, 259–
803 68 (2014).

804 8. Vaillancourt, K., Ernst, C., Mash, D. & Turecki, G. DNA Methylation Dynamics and Cocaine
805 in the Brain: Progress and Prospects. *Genes (Basel).* **8**, 138 (2017).

806 9. Engmann, O. *et al.* Cocaine-Induced Chromatin Modifications Associate With Increased
807 Expression and Three-Dimensional Looping of Auts2. *Biol. Psychiatry* **82**, 794–805 (2017).

808 10. LaPlant, Q. *et al.* Dnmt3a regulates emotional behavior and spine plasticity in the nucleus
809 accumbens. *Nat. Neurosci.* **13**, 1137–43 (2010).

810 11. Garavan, H. *et al.* Cue-Induced Cocaine Craving: Neuroanatomical Specificity for Drug
811 Users and Drug Stimuli. *Am. J. Psychiatry* **157**, 1789–1798 (2000).

812 12. Volkow, N. D. *et al.* Cocaine Cues and Dopamine in Dorsal Striatum: Mechanism of
813 Craving in Cocaine Addiction. *J. Neurosci.* **26**, 6583–6588 (2006).

814 13. Belin, D. & Everitt, B. Cocaine seeking habits depend upon dopamine-dependent serial
815 connectivity linking the ventral with the dorsal striatum. *Neuron* **57**, 432–441 (2008).

816 14. Everitt, B. J. & Robbins, T. W. From the ventral to the dorsal striatum: devolving views of
817 their roles in drug addiction. *Neurosci. Biobehav. Rev.* **37**, 1946–54 (2013).

818 15. Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for
819 genome-scale DNA methylation profiling. *Nat. Protoc.* **6**, 468–81 (2011).

820 16. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by
821 function. *Genome Res.* **13**, 2129–41 (2003).

822 17. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**,
823 317–330 (2015).

824 18. Matsumoto, K. *et al.* The prepattern transcription factor Irx2, a target of the FGF8/MAP
825 kinase cascade, is involved in cerebellum formation. *Nat. Neurosci.* **7**, 605–612 (2004).

826 19. Kozlenkov, A. *et al.* Differences in DNA methylation between human neuronal and glial
827 cells are concentrated in enhancers and non-CpG sites. *Nucleic Acids Res.* **42**, 109–27
828 (2014).

829 20. Rizzardi, L. F. *et al.* Neuronal brain-region-specific DNA methylation and chromatin
830 accessibility are associated with neuropsychiatric trait heritability. *Nat. Neurosci.* **22**,
831 307–316 (2019).

832 21. Tena, J. J. *et al.* An evolutionarily conserved three-dimensional structure in the
833 vertebrate Irx clusters facilitates enhancer sharing and coregulation. *Nat. Commun.* **2**,
834 310 (2011).

835 22. Hashimoto, H. *et al.* Structural Basis for the Versatile and Methylation-Dependent
836 Binding of CTCF to DNA. *Mol. Cell* **66**, 711-720.e3 (2017).

837 23. Kim, T. H. *et al.* Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the
838 Human Genome. *Cell* **128**, 1231–1245 (2007).

839 24. Wang, J. *et al.* Sequence features and chromatin structure around the genomic regions
840 bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812 (2012).

841 25. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation.
842 *Nucleic Acids Res.* **44**, 5615–5628 (2016).

843 26. Bernácer, J., Prensa, L. & Giménez-Amaya, J. M. Distribution of GABAergic interneurons
844 and dopaminergic cells in the functional territories of the human striatum. *PLoS One* **7**,
845 (2012).

846 27. Tepper, J. M., Tecuapetla, F., Koós, T. & Ibáñez-Sandoval, O. Heterogeneity and diversity
847 of striatal GABAergic interneurons. *Front. Neuroanat.* **4**, 150 (2010).

848 28. Lobo, M. K. & Nestler, E. J. The striatal balancing act in drug addiction: Distinct roles of
849 direct and indirect pathway medium spiny neurons. *Frontiers in Neuroanatomy* **5**, (2011).

850 29. Yao, P. *et al.* Coexpression networks identify brain region–specific enhancer RNAs in the
851 human brain. *Nat. Neurosci.* **18**, 1168–1174 (2015).

852 30. Hannon, E., Marzi, S. J., Schalkwyk, L. S. & Mill, J. Genetic risk variants for brain disorders
853 are enriched in cortical H3K27ac domains. *Mol. Brain* **12**, 7 (2019).

854 31. Maunakea, A. K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates
855 alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23**, 1256–
856 1269 (2013).

857 32. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating
858 alternative promoters. *Nature* **466**, 253–257 (2010).

859 33. Yang, X. *et al.* Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic
860 Target in Cancer. *Cancer Cell* **26**, 577–590 (2014).

861 34. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD,
862 schizophrenia, and bipolar disorder. *Science (80-.).* **362**, eaat8127 (2018).

863 35. Feng, J. *et al.* Chronic cocaine-regulated epigenomic changes in mouse nucleus
864 accumbens. *Genome Biol.* **15**, R65 (2014).

865 36. Cates, H. M. *et al.* Transcription Factor E2F3a in Nucleus Accumbens Affects Cocaine
866 Action via Transcription and Alternative Splicing. *Biol. Psychiatry* **84**, 167–179 (2018).

867 37. Baker-Andresen, D. *et al.* Persistent variations in neuronal DNA methylation following
868 cocaine self-administration and protracted abstinence in mice. *Neuroepigenetics* **4**, 1–11
869 (2015).

870 38. Ahn, J.-I. *et al.* Comprehensive transcriptome analysis of differentiation of embryonic
871 stem cells into midbrain and hindbrain neurons. *Dev. Biol.* **265**, 491–501 (2004).

872 39. Kasper, C., Hebert, F. O., Aubin-Horth, N. & Taborsky, B. Divergent brain gene expression
873 profiles between alternative behavioural helper types in a cooperative breeder. *Mol.*
874 *Ecol.* **27**, 4136–4151 (2018).

875 40. Gomez-Velazquez, M. *et al.* CTCF counter-regulates cardiomyocyte development and
876 maturation programs in the embryonic heart. *PLOS Genet.* **13**, e1006985 (2017).

877 41. Kelz, M. B. *et al.* Expression of the transcription factor ΔFosB in the brain controls
878 sensitivity to cocaine. *Nature* **401**, 272–276 (1999).

879 42. Nestler, E. J. The neurobiology of cocaine addiction. *Sci. Pract. Perspect.* **3**, 4–10 (2005).

880 43. Bannon, M. J. *et al.* A Molecular Profile of Cocaine Abuse Includes the Differential
881 Expression of Genes that Regulate Transcription, Chromatin, and Dopamine Cell
882 Phenotype. *Neuropsychopharmacology* **39**, 1–9 (2014).

883 44. Cannella, N. *et al.* Dnmt3a2 in the Nucleus Accumbens Shell Is Required for
884 Reinstatement of Cocaine Seeking. *J. Neurosci.* **38**, 7516–7528 (2018).

885 45. Chandra, R. *et al.* Opposing Role for Egr3 in Nucleus Accumbens Cell Subtypes in Cocaine
886 Action. *J. Neurosci.* **35**, 7927–7937 (2015).

887 46. Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to
888 serve and mine knowledge about genes and proteins. *Database* **2016**, baw100 (2016).

889 47. Saftig, P. & Lichtenthaler, S. F. The alpha secretase ADAM10: A metalloprotease with
890 multiple functions in the brain. *Prog. Neurobiol.* **135**, 1–20 (2015).

891 48. Shukla, M., Maitra, S., Hernandez, J.-F., Govitrapong, P. & Vincent, B. Methamphetamine
892 regulates βAPP processing in human neuroblastoma cells. *Neurosci. Lett.* **701**, 20–25
893 (2019).

894 49. Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor
895 Occupancy. *Cell Rep.* **12**, 1184–1195 (2015).

896 50. Zhang, T. Y. *et al.* Maternal care and DNA methylation of a glutamic acid decarboxylase 1
897 promoter in rat hippocampus. *J. Neurosci.* **30**, 13130–13137 (2010).

898 51. Zhang, T. Y. *et al.* Environmental enrichment increases transcriptional and epigenetic
899 differentiation between mouse dorsal and ventral dentate gyrus. *Nat. Commun.* **9**, 1–11
900 (2018).

901 52. Gross, J. A., Fiori, L. M., Labonté, B., Lopez, J. P. & Turecki, G. Effects of promoter
902 methylation on increased expression of polyamine biosynthetic genes in suicide. *J.*
903 *Psychiatr. Res.* **47**, 513–519 (2013).

904 53. Iwata, A. *et al.* Altered CpG methylation in sporadic Alzheimer's disease is associated
905 with APP and MAPT dysregulation. *Hum. Mol. Genet.* **23**, 648–56 (2014).

906 54. Lutz, P. E. *et al.* Association of a history of child abuse with impaired myelination in the
907 anterior cingulate cortex: Convergent epigenetic, transcriptional, and morphological
908 evidence. *Am. J. Psychiatry* **174**, 1185–1194 (2017).

909 55. Guo, J. U. *et al.* Distribution, recognition and regulation of non-CpG methylation in the
910 adult mammalian brain. *Nat. Neurosci.* **17**, 215–22 (2013).

911 56. Li, X. *et al.* The DNA modification N6-methyl-2'-deoxyadenosine (m6dA) drives activity-
912 induced gene expression and is required for fear extinction. *Nat. Neurosci.* **22**, 534–544
913 (2019).

914 57. Feng, J. *et al.* Role of Tet1 and 5-hydroxymethylcytosine in cocaine action. *Nat. Neurosci.*
915 **18**, 536–544 (2015).

916 58. Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements
917 in mammalian cortex. *Science (80-.).* **357**, 600–604 (2017).

918 59. Chen, G. G. *et al.* BisQC: an operational pipeline for multiplexed bisulfite sequencing.
919 *BMC Genomics* **15**, 290 (2014).

920 60. Cavalcante, R. G. & Sartor, M. A. Annotatr: Genomic regions in context. *Bioinformatics*
921 **33**, 2381–2383 (2017).

922 61. Sheffield, N. C. & Bock, C. LOLA: Enrichment analysis for genomic region sets and
923 regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016).

924 62. GTEx Consortium, T. Gte. Human genomics. The Genotype-Tissue Expression (GTEx) pilot
925 analysis: multitissue gene regulation in humans. *Science* **348**, 648–60 (2015).

926 63. Chen, G. G. *et al.* Medium throughput bisulfite sequencing for accurate detection of 5-
927 methylcytosine and 5-hydroxymethylcytosine. *BMC Genomics* **18**, 96 (2017).

928 64. Johnson, A. R. *et al.* Cues play a critical role in estrous cycle-dependent enhancement of
929 cocaine reinforcement. *Neuropsychopharmacology* (2019). doi:10.1038/s41386-019-
930 0320-0

931 65. Ea, V., Court, F. & Forne, T. Quantitative Analysis of Intra-chromosomal Contacts: The 3C-
932 qPCR Method. *Methods Mol. Biol.* **1589**, 75–88 (2017).

933 66. Braem, C. *et al.* Genomic matrix attachment region and chromosome conformation
934 capture quantitative real time PCR assays identify novel putative regulatory elements at
935 the imprinted Dlk1/Gtl2 locus. *J. Biol. Chem.* **283**, 18612–20 (2008).

936

937

938 ## Figure Legends

939 **Figure 1. Widespread changes in DNA methylation associated with chronic cocaine**
940 **dependence in the human caudate nucleus**. a) Dissections from dorsolateral caudate nucleus
941 were used for reduced representation bisulfite sequencing (n=25 per group, boxed). b)
942 Manhattan plot showing the chromosomal location of all significantly differentially methylated
943 CpG regions (DMRs); blue line represents FDR q < 0.05. c) Although hyper- and hypomethylated
944 DMCs were identified, there was a significant bias towards clusters with increased methylation
945 in the cocaine group χ^2= 26.575; * p < 0.05. d)Most of the DMCs overlap with known CpG islands
946 and e) annotated introns, exons and intron-exon boundaries.

947 **Figure 2. *IRX2* is hypomethylated in the caudate nucleus.** a) RRBS analysis identified a cluster of
948 21 CpGs within the third exon of *IRX2* that were less methylated in the cocaine group of the
949 discovery cohort (n=25 per group). b) This data was replicated in an independent cohort of
950 caudate samples (n=15 cases and n=20 controls). c) Hypomethylation was specific to neuronal
951 (NeuN+) nuclei (n=24 cases and n=24 controls; discovery cohort). d) The 5' CpG within the CTCF
952 binding site (exon 3) was hypomethylated in the cocaine group in caudate tissue homogenate
953 (n= 20 controls and 17 cases, replication cohort) and e) neuronal nuclei (n=24 controls and n=23
954 cases). f) The 5' most CpG in the mouse CTCF site was significantly less methylated after cocaine
955 self-administration (n=6) , compared to non-drug reward self-administration (n=9) or controls
956 (n=14). Box plots indicate mean and range of data. Bar data represented as mean ± s.e.m. * p <
957 0.05; ** q val < 0.02

958 **Figure 3. *IRX2* expression is increased in cocaine use disorder and is related to exon 3**
959 **methylation in cells.** a) *IRX2* expression was significantly increased in the caudate nucleus of
960 cocaine dependent subjects (n=21 cases and n=23 controls). b) while no significant increase in
961 *IRX1* was detected (n=22 cases and n=22 controls), c) the expression of both *IRX1* and *IRX2*
962 transcripts was highly correlated (n=36). d) Human neural progenitor cells (RENcells)
963 endogenously expressed *IRX1* and *IRX2* while kidney epithelial cells (HEK293) did not (n=3 per
964 group). e) Endogenous methylation of *IRX2* exon 3 was higher in kidney epithelial cells than in
965 neural progenitor cells (n=3 per group). f) Transfection of an active dCas9-DNMT3A construct,
966 along with a pool of 3 guide RNA constructs significantly increased methylation of *IRX2* exon 3 in
967 RenCells, compared to transfection with an inactive construct or wildtype controls. g) Active
968 methylation of *IRX2* decreased transcription of both *IRX2* and *IRX1* compared to inactive or
969 wildtype cells. (n=2-3 replicates per group)WT=wildtype. Data represented as mean ± s.e.m
970 **** p <0.0001, * p < 0.05

971 **Figure 4. Long range chromatin structure of the *IRXA* gene cluster is impacted by methylation.**
972 a) Chromatin conformation capture (3C) experimental design contained a standard viewpoint
973 within the first intron of *IRX1* (orange arrowhead), and test primers (open arrowheads) tiled
974 across IRX2 and the intergenic region between the genes. The genomic fragments containing
975 *IRX2* functionally interacted with the 5' end *IRX1* in two human cell types, with significantly

976 higher rates of interaction observed in RenCells. The noise band, where interaction frequencies
977 would be expected by chance ("random collisions"), is indicated by the horizontal dashed lines.
978 Vertical dashed lines pair restriction fragment with its corresponding data points. N=3 replicates
979 per group. b) Active methylation of HEK293 cells decreased long range interaction between
980 restriction fragment 2 and the *IRX1* gene (n=3 replicates of 10 x 10^6 cells per group). c) The
981 dCas9-DNMT3A transfection significantly (n=3 replicates of 5 x 10^6 cells per group) decreased
982 CTCF binding to IRX2 exon 3 compared to wildtype cells. WT=wildtype. Data represented as
983 mean ± s.e.m. ** $p < 0.01$; *** $p < 0.0005$

984 **Figure 5**. **A model for cocaine-sensitivity of 3D chromatin organization at the *IRXA* gene**
985 **cluster.** Cocaine dependence is associated with decreased intragenic methylation of *IRX2*, which
986 may increase *IRXA* gene expression through CTCF-mediated chromatin architecture.

987 **Table 1**. Differentially methylated regions with nominally differential gene expression.

988

989

990

991

992

993

994

995

996

997

998

999

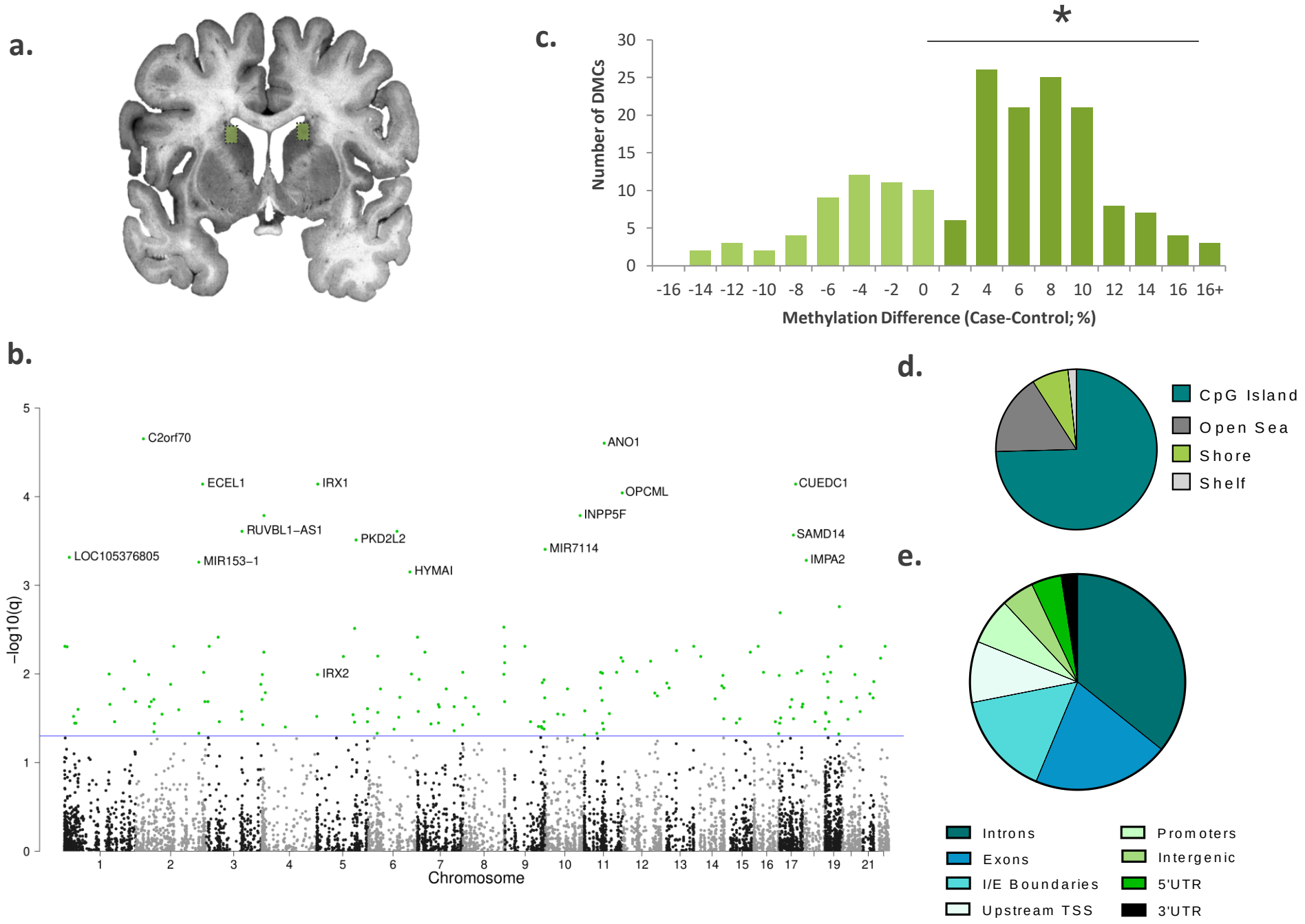
1000

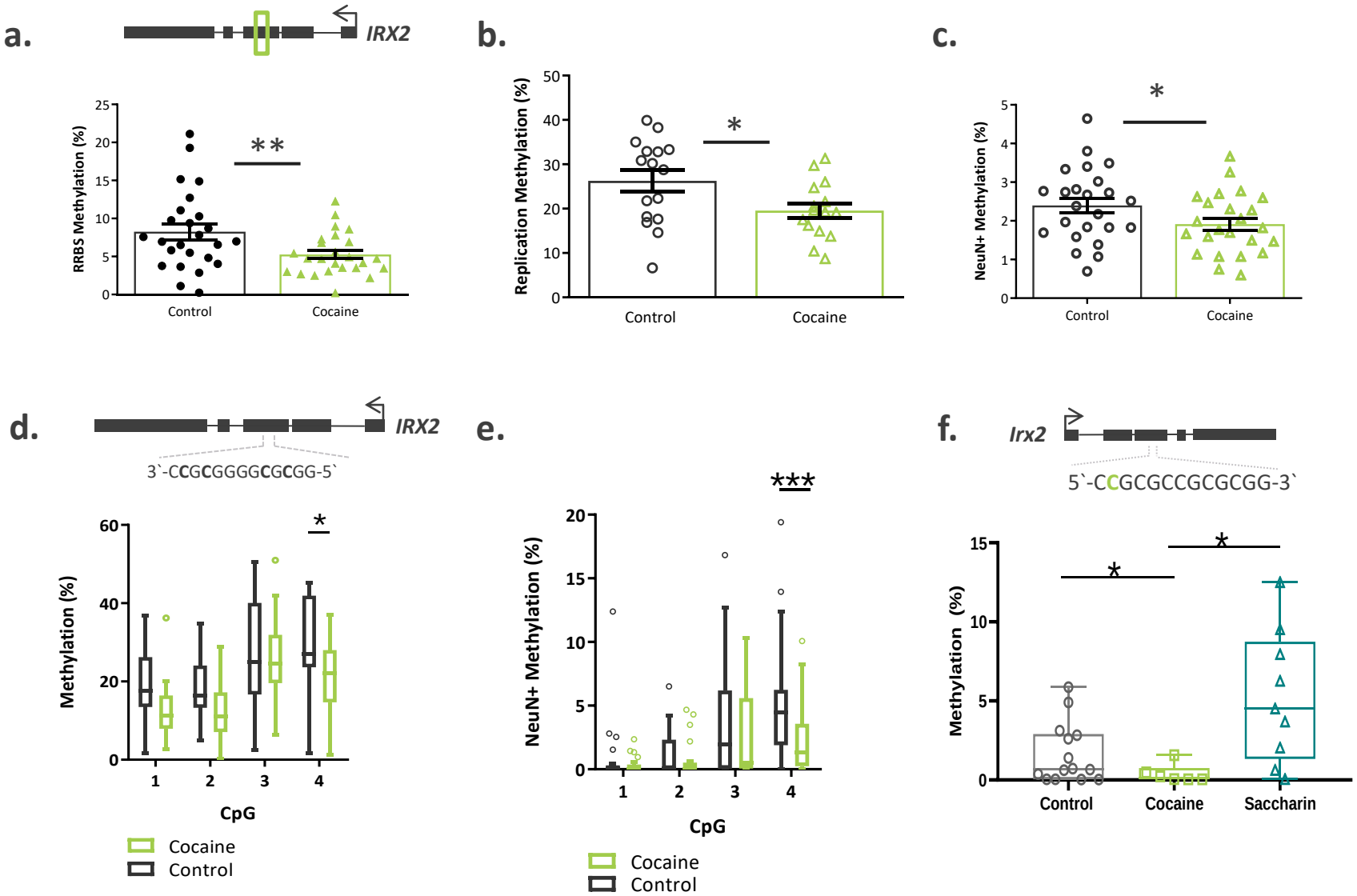
1001 **Tables**

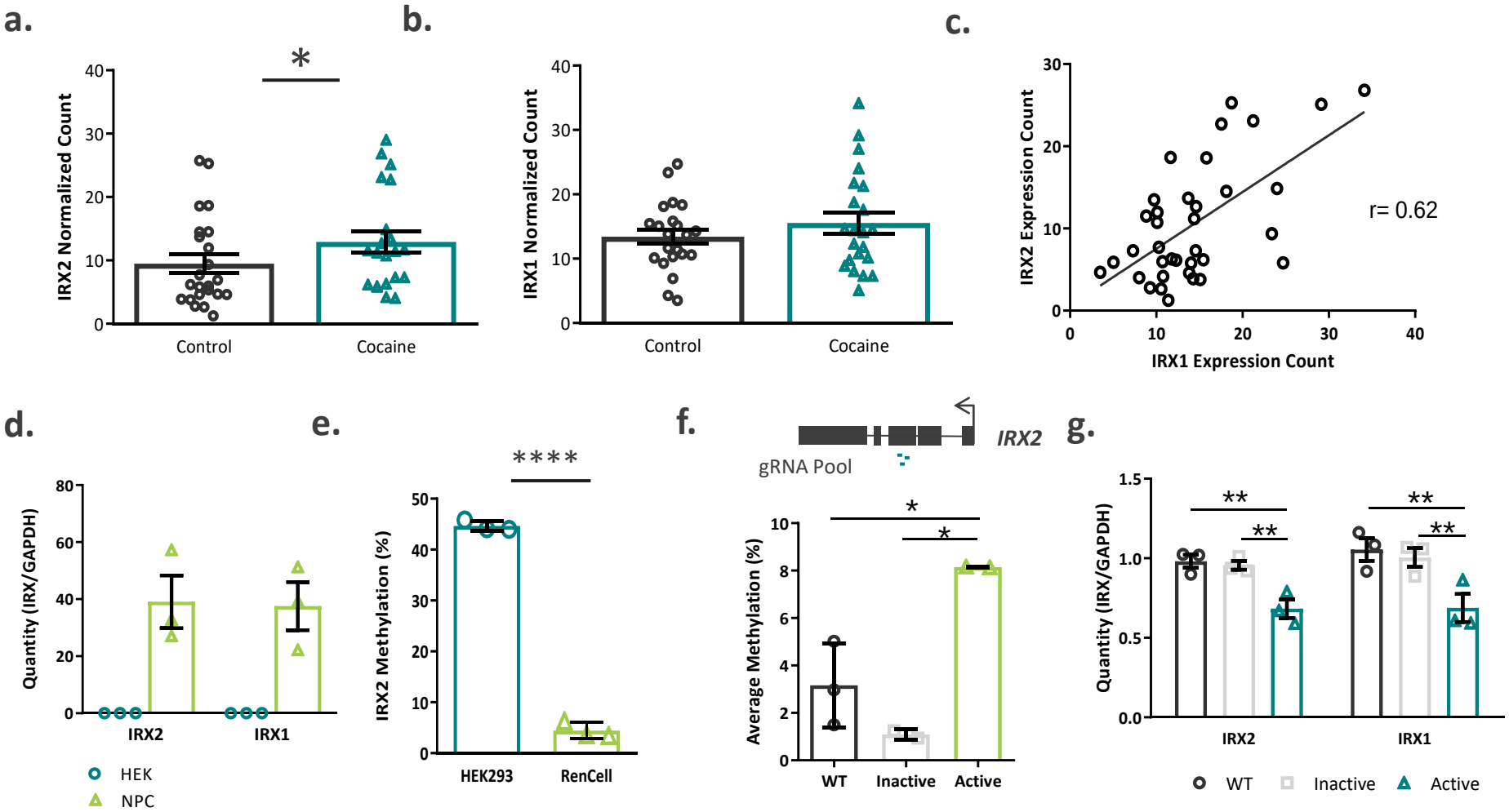
1002 **Table 1**. Differentially methylated regions with nominally differential gene expression

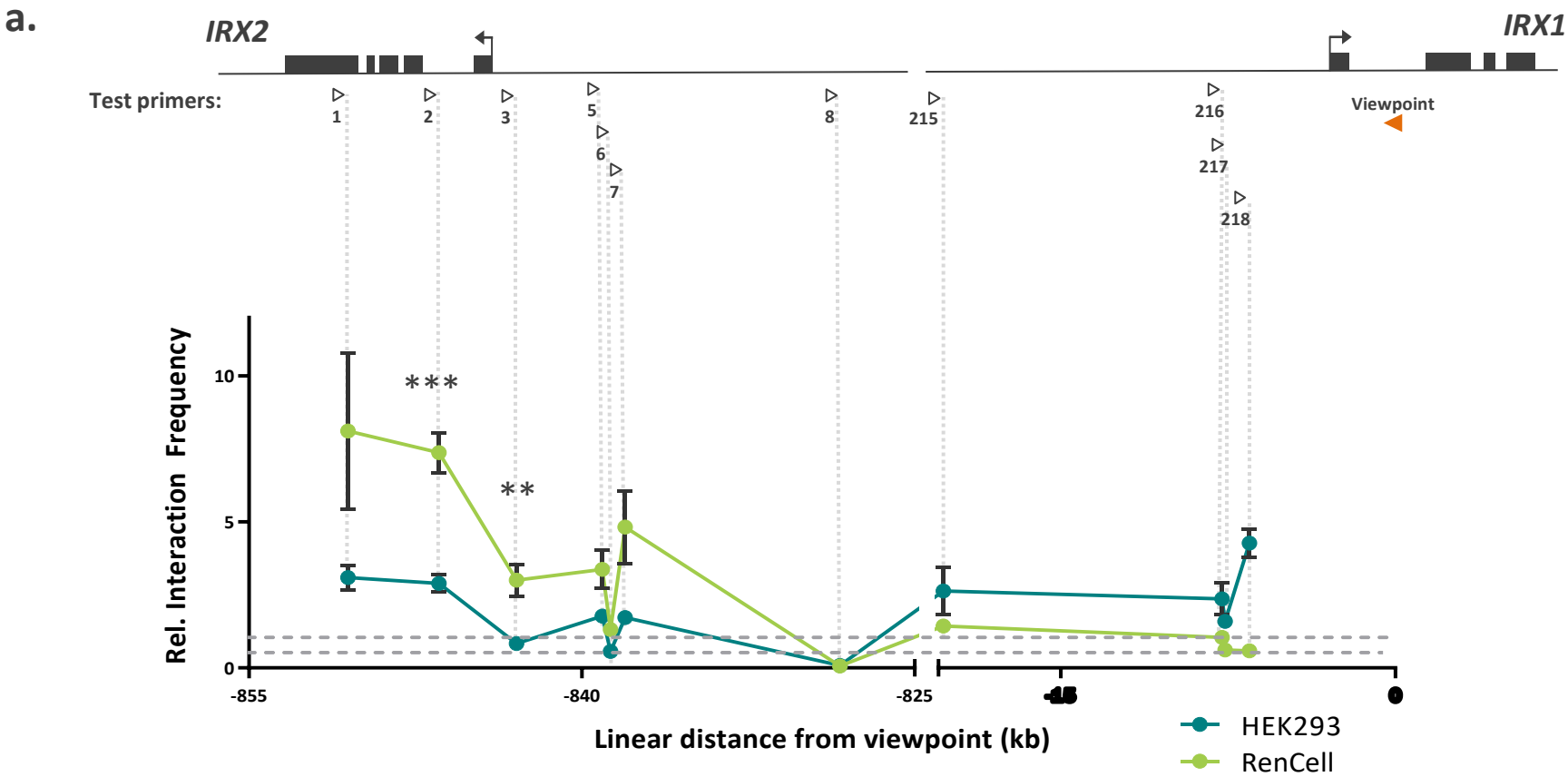
Chr	From	To	Gene	Number of CpGs	Methylation Difference	RNAseq Fold Change	RNAseq p-value
5	2748781	2748955	IRX2	21	-3.06%	1.20	8.17E-02
9	126776177	126776282	LHX2	12	-8.22%	1.22	1.46E-02
5	3599609	3599704	IRX1	9	11.05%	1.39	1.21E-03
4	53474	53566	ZNF595	9	7.67%	0.79	1.08E-02
17	1960987	1961029	HIC1	8	-11.51%	1.25	7.09E-03
2	26785211	26785290	C2orf70	7	7.44%	1.29	1.31E-02
1	17215449	17215492	CROCC	7	-5.45%	1.24	1.85E-03
11	132812684	132812729	OPCML	7	15.93%	0.86	7.03E-02
2	63274825	63274901	OTX1	6	9.92%	1.25	2.53E-02
18	77918229	77918253	PARD6G	6	-10.83%	1.51	3.98E-05
5	131607235	131607278	PDLIM4	6	8.28%	1.36	2.78E-03
7	150002	150037	AC093627.10	5	8.95%	0.84	2.14E-02
12	120654707	120654747	PXN	5	10.21%	1.17	5.55E-02
19	18980163	18980188	UPF1	5	10.03%	1.07	2.35E-02
2	63275003	63275040	OTX1	4	7.49%	1.25	2.53E-02
17	76172805	76172850	TK1	4	7.10%	0.87	8.77E-02
8	37556087	37556121	ZNF703	4	12.14%	1.20	1.41E-02
6	32165134	32165176	NOTCH4	3	8.58%	1.14	9.81E-02
3	8799985	8800008	OXTR	3	12.55%	1.29	1.35E-02
19	47220817	47220856	PRKD2	3	5.76%	1.33	3.70E-04
5	176877611	176877636	PRR7	3	11.64%	1.53	4.52E-11
4	6273547	6273577	WFS1	3	21.42%	1.12	7.12E-02
9	136654410	136654426	VAV2	2	12.46%	1.17	3.08E-02

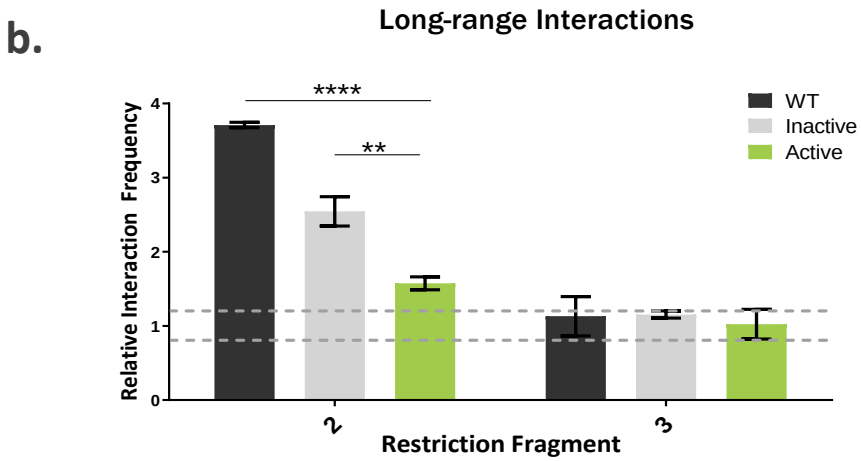
1003

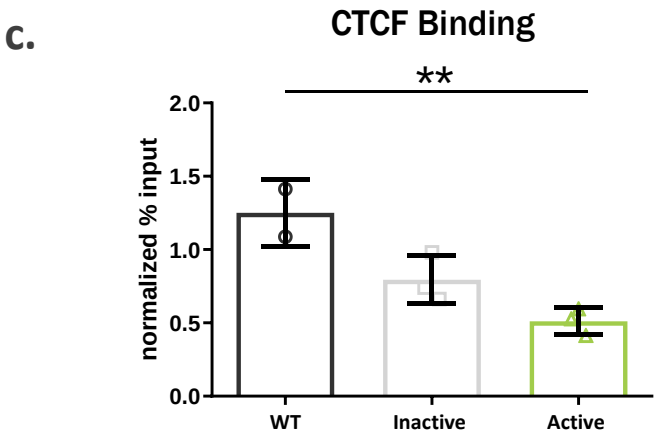

a.
b.
c.
d.
e.
*
Number of DMCs
30
25
20
15
10
5
0
-16 -14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16 16+
Methylation Difference (Case-Control; %)
C2orf70
ECEL1
IRX1
RUVBL1–AS1
PKD2L2
LOC105376805
MIR153–1
HYMAI
IRX2
ANO1
OPCML
INPP5F
MIR7114
CUEDC1
SAMD14
IMPA2
−log10(q)
Chromosome
CpG Island
Open Sea
Shore
Shelf
Introns
Exons
I/E Boundaries
Upstream TSS
Promoters
Intergenic
5'UTR
3'UTR

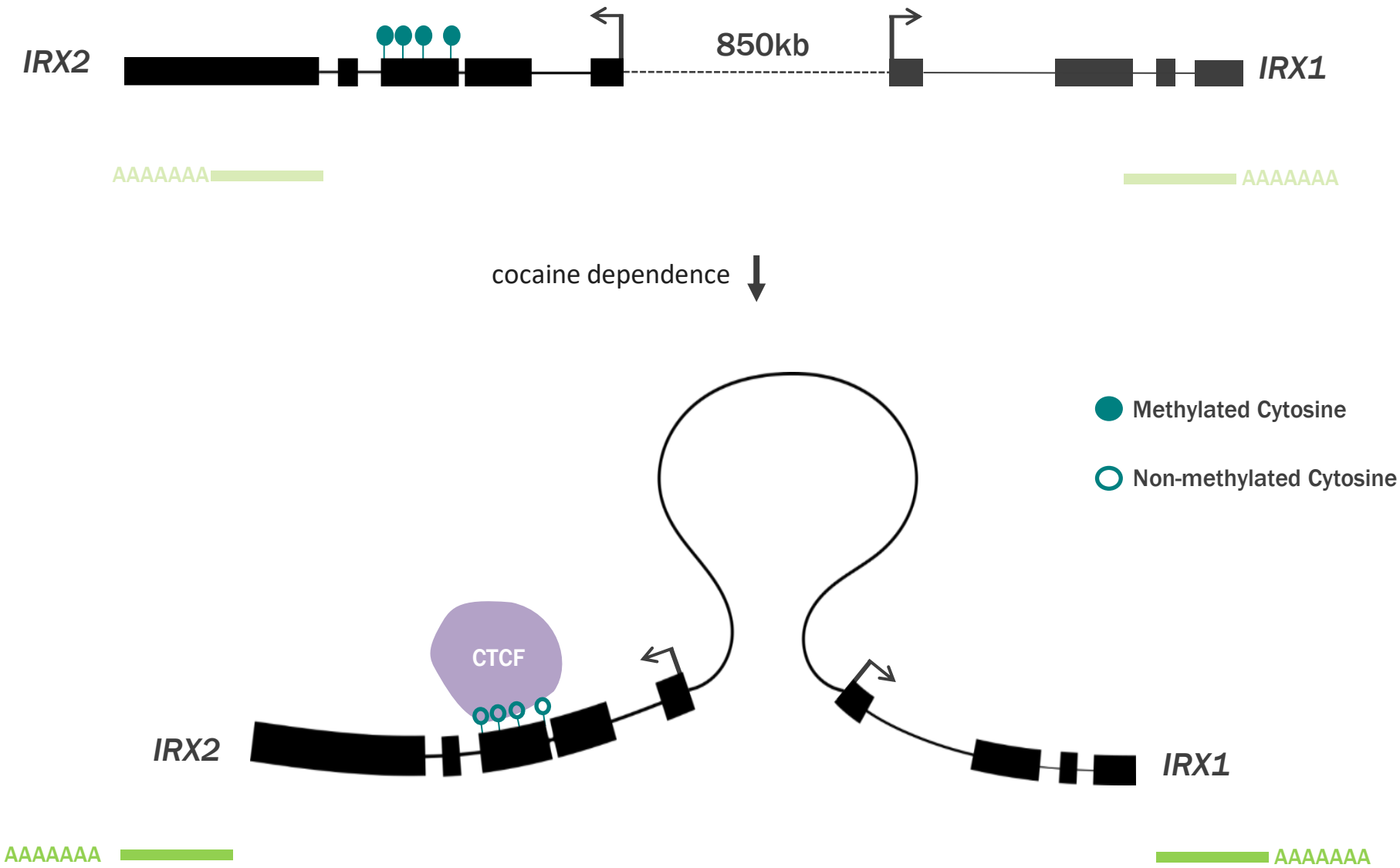

a.
IRX2
RRBS Methylation (%)
25
20
15
10
5
0
**
Control
Cocaine
b.
Replication Methylation (%)
50
40
30
20
10
0
*
Control
Cocaine
c.
NeuN+ Methylation (%)
5
4
3
2
1
0
*
Control
Cocaine
d.
IRX2
3`-CCGCGGGGCGCGG-5`
Methylation (%)
60
40
20
0
*
1
2
3
4
CpG
Cocaine
Control
e.
NeuN+ Methylation (%)
20
15
10
5
0
***
1
2
3
4
CpG
Cocaine
Control
f.
Irx2
5`-CCGCGCCGCGCGG-3`
Methylation (%)
15
10
5
0
*
*
Control
Cocaine
Saccharin


a.
b.
c.
IRX2 Normalized Count
IRX1 Normalized Count
Control
Cocaine
*
IRX2 Expression Count
IRX1 Expression Count
r= 0.62
d.
e.
f.
g.
Quantity (IRX/GAPDH)
IRX2
IRX1
HEK
NPC
IRX2 Methylation (%)
****
HEK293
RenCell
gRNA Pool
IRX2
Average Methylation (%)
WT
Inactive
Active
**


a.
IRX2
IRX1
Test primers:
1
2
3
5
6
7
8
215
216
217
218
Viewpoint
Rel. Interaction Frequency
10
5
0
***
**
-855
-840
-825
-15
0
Linear distance from viewpoint (kb)
HEK293
RenCell


b.
Long-range Interactions
****
**
WT
Inactive
Active
Relative Interaction Frequency
4
3
2
1
0
2
3
Restriction Fragment


c.
CTCF Binding
**
normalized % input
2.0
1.5
1.0
0.5
0.0
WT
Inactive
Active


IRX2
850kb
IRX1
AAAAAAA
AAAAAAA
cocaine dependence
Methylated Cytosine
Non-methylated Cytosine
CTCF
IRX2
IRX1
AAAAAAA
AAAAAAA