


DNA methylation and hydroxymethylation characterize the identity of D1 and D2 striatal projection neurons

Lucile Marion-Poll ^{1,2,3,4,8}, Jean-Pierre Roussarie ^{5,9,13}, Lieng Taing^{1,2,3,10,13}, Cloelia Dard-Dascot^{6,13}, Nicolas Servant ^{7,13}, Yan Jaszczyszyn⁶, Emmanuelle Jordi^{1,2,3,11}, Eskeatnaf Mulugeta ^{4,12}, Denis Hervé^{1,2,3}, Déborah Bourc'his ⁴, Paul Greengard⁵, Claude Thermes ⁶ & Jean-Antoine Girault ^{1,2,3}

Neuronal DNA modifications differ from those in other cells, including methylation outside CpG context and abundant 5-hydroxymethylation whose relevance for neuronal identities are unclear. Striatal projection neurons expressing D1 or D2 dopamine receptors allow addressing this question, as they share many characteristics but differ in their gene expression profiles, connections, and functional roles. We compare translating mRNAs and DNA modifications in these two populations. DNA methylation differences occur predominantly in large genomic clusters including differentially expressed genes, potentially important for D1 and D2 neurons. Decreased gene body methylation is associated with higher gene expression. Hydroxymethylation differences are more scattered and affect transcription factor binding sites, which can influence gene expression. We also find a strong genome-wide hydroxymethylation asymmetry between the two DNA strands, particularly pronounced at expressed genes and retrotransposons. These results identify novel properties of neuronal DNA modifications and unveil epigenetic characteristics of striatal projection neurons heterogeneity.

¹INSERM UMR-S1270, Paris 75005, France. ²Sorbonne Université, Faculty of Sciences and Engineering, Paris 75005, France. ³Institut du Fer à Moulin, Paris 75005, France. ⁴Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, Paris 75005, France. ⁵Laboratory of Molecular and Cellular Neuroscience, The Rockefeller University, New York, NY 10065, USA. ⁶Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette 91198, France. ⁷Institut Curie, INSERM U900, CBIO-Centre for Computational Biology, Mines Paris Tech, PSL-Research University, Paris 75005, France. ⁸Present address: Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, Geneva 1211, Switzerland. ⁹Present address: Department of Anatomy & Neurobiology, Boston University Chobanian & Avedisian School of Medicine, Boston, MA 02118, USA. ¹⁰Present address: UMR1166 Inserm and Sorbonne Université, Faculty of Medicine, Paris 75013, France. ¹¹Present address: Coave Therapeutics, Paris 75014, France. ¹²Present address: Erasmus University Medical Center (Erasmus MC), Department of Cell Biology, Rotterdam 3000 CA, The Netherlands. ¹³These authors contributed equally: Jean-Pierre Roussarie, Lieng Taing, Cloelia Dard-Dascot, Nicolas Servant. ✉email: lucile.marion-poll@unige.ch; jean-antoine.girault@inserm.fr

Epigenetic marks, including DNA modifications, play a key role in defining cell identities. An intriguing area concerns the role of epigenetic modifications in controlling specific properties of neurons which are long-lived post-mitotic cells, dynamically regulated by external stimuli, including environment and life experiences. In mammalian cells, cytosines can be methylated at the fifth carbon, generally when followed by a guanine (CG), a context where both strands can be symmetrically methylated¹, whereas in neurons, up to half of the DNA methylation (5mC) is found in a non-CG (i.e., CH) context². 5mC can be oxidized to 5-hydroxymethylated-cytosine by the TET (ten-eleven translocation) enzymes, and this modification is especially abundant in neurons³. DNA hydroxymethylation (5hmC) is an intermediate for demethylation during development^{4,5} and may play specific roles in the adult brain, through the recruitment of 5hmC-specific binding proteins⁶.

DNA modification in neurons contributes to neuronal development and function. Mutations in DNA methyltransferase or chromatin modifier genes are associated with brain developmental disorders^{7,8}. The methylcytosine binding protein 2 (MECP2), which binds to 5mC and 5hmC, is particularly abundant in neurons^{9,10} and is mutated in Rett syndrome, the second cause of mental disability in girls. 5mC and 5hmC can be modified following stimulation¹¹ and may contribute to the formation and stabilization of long-term memory¹².

The striatum has essential roles in movement control, action selection, and reinforcement learning. This is achieved through the existence of two main types of GABAergic striatal projection neurons (SPNs, a.k.a. medium-size spiny neurons or MSNs), which have opposite but complementary roles^{13,14}. SPNs that project to the substantia nigra pars reticulata form the direct striatonigral pathway and express dopamine D1 receptors, while SPNs that project to the external globus pallidus participate in the indirect pathway and express D2 receptors¹⁵, with very few neurons expressing both types of receptors¹⁶. D1 receptors are activated by phasic increases in extracellular levels of dopamine, triggered by unexpected rewards¹⁷ and other salient stimuli¹⁸. D1- and D2-SPNs have opposite effects on locomotion¹⁹ but shape behavior in an integrated manner²⁰. These neurons are involved in major pathological conditions, including Parkinson's disease and addiction^{21,22}. Histone and DNA modifications in the striatum are altered by the treatment of Parkinson's disease, levodopa²³, and by drugs of abuse, while epigenetic modifiers or epigenetic editing can affect addictive behaviors²⁴.

D1- and D2-SPNs are very closely related²⁵, but differentially express several hundreds of genes^{16,26}. During development, both populations originate from the lateral ganglionic eminence and migrate to the future striatum, where they are intermingled. The generation of SPNs starts around E10.5 in mice and continues until birth²⁷. The precise mechanisms determining their differential projections are not known, but key transcription factors (TFs) for their identities have been identified, including *Ebf1*²⁸ or *Isl1*²⁹ for D1-SPNs and *Sp9*³⁰ or *Six3*³¹ for D2-SPNs. The expression of specific markers, notably D1- and D2-receptors, starts early on but keeps increasing after birth^{32,33}. Previous studies have explored the genome-wide distribution of DNA modifications in very different neuronal types, or in mixed neurons compared to other cell types^{2,9,34–38}. In contrast, comparing D1- and D2-SPNs provides an excellent model to explore the specificities of DNA modification established during terminal neuronal differentiation. Here, we characterize the methylomes and hydroxymethylomes of the two types of SPNs in relation to their transcriptome. Our aim was to investigate D1/D2 differentially modified regions and to compare them with gene expression differences rather than establishing quantitative profiles of these DNA modifications in each cell

type at a single base resolution. We, therefore, chose enrichment-based methods over bisulfite-based methods for this study, because they allow direct readouts of 5mC and 5hmC, and have a good sensitivity for sparse modifications such as 5hmC³⁹. Antibody-based approaches can show some biases⁴⁰, but these would be identical between SPNs.

We find that most 5mC differences map to 15 structural domains of hundreds of kilobases, which include D1/D2-specific genes, likely to be of key importance for SPNs specificities. We show at the single gene level that decreased 5mC at promoters or gene bodies is correlated with increased expression of differentially expressed genes. 5hmC differences preferentially affect transcription factor binding sites (TFBSs) and are also associated with differential gene expression. We find strand asymmetry of both DNA modifications, but much more pronounced for 5hmC. Hydroxymethylation is asymmetric all over the genome, with nested asymmetric regions with a bimodal size distribution of up to 1 Mb. Asymmetry is high in retrotransposons and highly transcribed genes, which depends on the direction of transcription. Our study highlights specific features of neuronal DNA modifications in relation to gene expression and characterizes their differences between the two main populations of striatal dopamine-sensitive neurons.

Results

Translating mRNA and DNA modifications patterns in D1 and D2 neurons. To assess epigenomic and transcriptomic patterns in D1 and D2 neurons, we used mice carrying a bacterial artificial chromosome (BAC) that express the ribosomal protein L10a (RPL10A) fused to EGFP under the control of either *Drd1* (D1R) or *Drd2* (D2R) promoters^{25,26} (Supplementary Fig. 1a). These BAC-TRAP (translating ribosome affinity purification) mice allow immunopurification and sequencing of cell-type specific translating mRNAs (TRAP-seq). As RPL10A is also abundant in the nucleoli where ribosomes are assembled (Supplementary Fig. 1b), the GFP-positive nuclei from these mice can be sorted by fluorescence-activated nuclear sorting³ (FANS) with very high purity (>97%, see Supplementary Fig. 2a and Methods for the gating strategy). Because the nuclear tagging is not part of the chromatin, it avoids the critical pitfall of disturbing chromatin organization and epigenomic profiles⁴¹. We combined D1/D2-TRAP-Seq with 5mC- and 5hmC-immunoprecipitation (MeDIP and hMeDIP, respectively) followed by DNA strand-specific sequencing (Fig. 1a, b, and Methods).

The results were highly reproducible between the three replicates (each comprised of bilateral striata from 4 mice), and the two neuronal types showed very close proximity (Fig. 2a). The TRAP-seq was clearly enriched for striatal neuronal markers and contained very low levels of markers from other cell types (as identified by Gokce et al.¹⁶, Supplementary Fig. 2b) notably cholinergic interneurons (*Chat*, *Scl18a3*, Supplementary Table 1). In addition, it showed very high specificity in all samples, as illustrated by the expression of *Drd1* and *Drd2* genes (Supplementary Fig. 2c). Genes known to be specifically expressed in D1 (e.g., *Pdyn*, *Tac1*, and *Slc35d3*) or D2 neurons (e.g., *Adora2a* and *Penk*)^{26,42} were, as expected, highly differentially expressed, along with 1963 other genes for an adjusted Pvalue (Padj) of 0.05 (Fig. 2b, Supplementary Fig. 2d, Supplementary Table 1), mostly related to neuronal functions (Fig. 2c). Our results were in accordance with previous results obtained with various methods^{16,26,42–44} (Supplementary Fig. 2e). The DNA methyltransferases (DNMT) genes were not differentially expressed, while two of the TET enzymes showed a higher expression in D2-SPNs (*Tet1*, +39%, Padj = 1.7×10^{-2} , and *Tet2*, +44%, Padj = 3.8×10^{-2} , Supplementary Table 1).

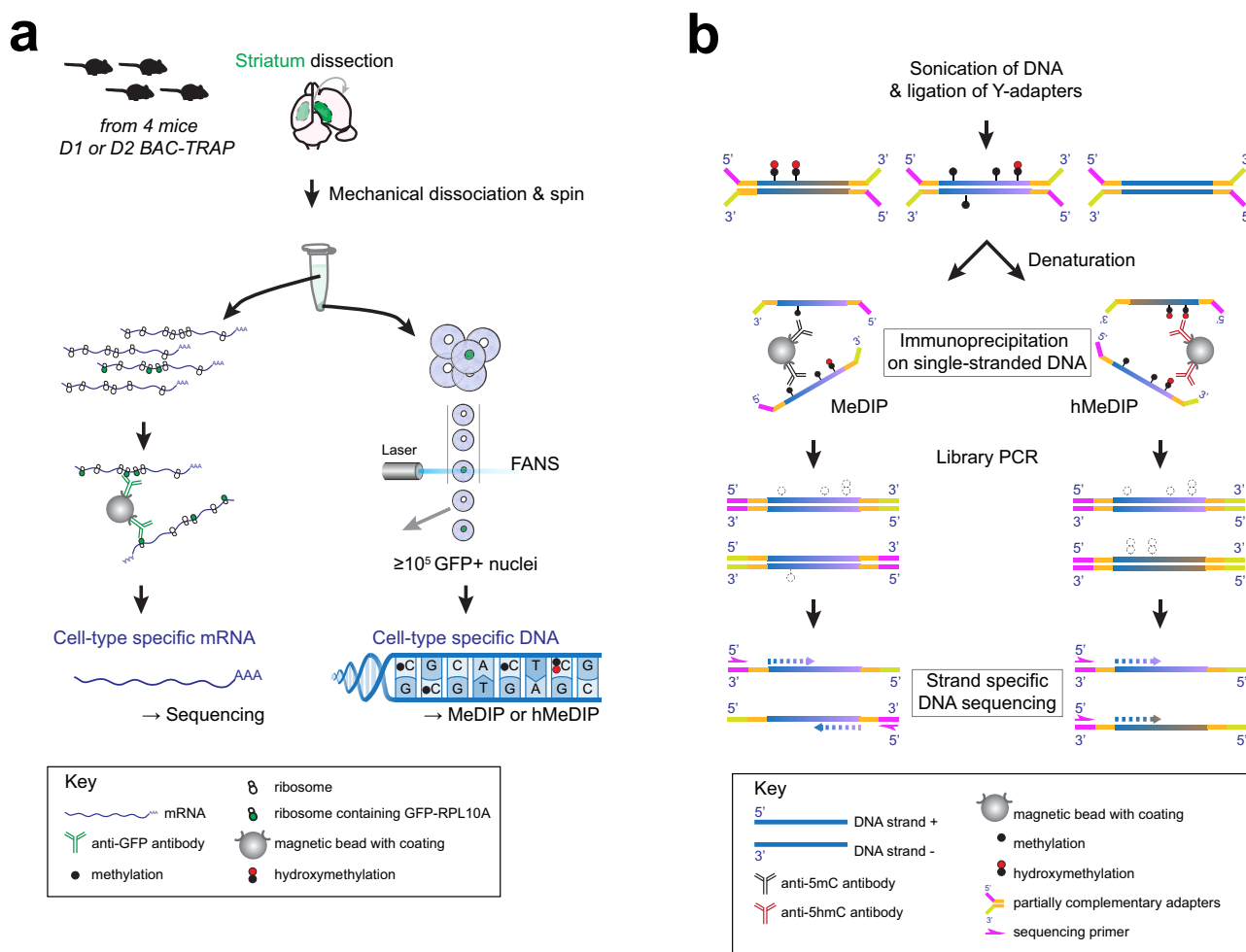


Fig. 1 Experimental strategy. **a** Cell population-specific mRNA and DNA preparation from D1-BAC-TRAP and D2-BAC-TRAP mice. D1- or D2-specific mRNA was recovered by immunoprecipitation of ribosomes containing EGFP-RPL10A with anti-GFP antibodies. D1- or D2-specific DNA was obtained by FANS using the fluorescence of nucleoli containing EGFP-RPL10A during ribosomal assembly, followed by DNA extraction. **b** Strand-specificity of the MeDIP-seq and hMeDIP-seq procedures.

The methylomes and hydroxymethylomes were remarkably similar between D1- and D2-SPNs (Fig. 2d), as expected because of their close similarity. As 5hmC is an intermediate for demethylation, the two profiles showed an overall similarity. However, we observed some large regions already visible at the megabase scale, where D1 and D2 5mC, but not 5hmC, appeared different (highlighted in Fig. 2d). We, therefore, analyzed how the D1/D2 differences in DNA modification are organized at a finer scale.

5mC and 5hmC form cell type-specific clusters. To further analyze differentially methylated or hydroxymethylated regions between D1 and D2 cells, we first compared them using adjacent 1-kb windows over the genome (Fig. 3a), excluding the BAC transgene sequences. We identified 1403 differentially methylated and 1386 differentially hydroxymethylated 1-kb windows (Supplementary Tables 2 and 3). The majority of the differential 5mC windows were grouped in specific regions of the genome, which were already visible at the megabase scale (Figs. 3a, 2d). A quarter of them was also differentially hydroxymethylated, in a concordant manner (i.e., more hydroxymethylated and more methylated in the same cell type, Fig. 3b). To estimate the size of the differentially modified regions, we clustered significantly differentially modified 1-kb windows which were in close proximity (see Methods for details). Among 1403 differentially

methylated 1-kb windows, 84% could be grouped (Fig. 3c) into 92 regions of 2 kb or more (Fig. 3d, Supplementary Table 4). We found 15 large clusters >100 kb, as illustrated in Fig. 3e for cluster #10, which spans 237 kb and includes *Adora2a* and other genes (Supplementary Table 5). All the main 5mC clusters also contained some 5hmC differences (Fig. 3a). However, only 55% of the differentially hydroxymethylated 1-kb windows could be grouped (Fig. 3c), into 161 clusters (Fig. 3d, Supplementary Table 6), including five large clusters >100 kb (Fig. 3d, Supplementary Table 7), four of which overlapped large 5mC clusters. Thus, the D1/D2 5mC differences were more clustered across the genome than the 5hmC differences, which tended to be more scattered. We, therefore, investigated whether the clusters of 5mC differences correspond to chromatin regions relevant for D1 and D2 identities.

D1/D2 5mC clusters contain differentially expressed genes. We first assessed the relation between differential 5mC clusters and known chromatin structural domains. We used high-resolution chromosome conformation capture (Hi-C) maps of neural progenitor cells since topologically associating domains (TADs) are relatively conserved between cell types, despite some genome reorganization occurring during differentiation⁴⁵. All the main differential 5mC clusters were located within TADs identified by

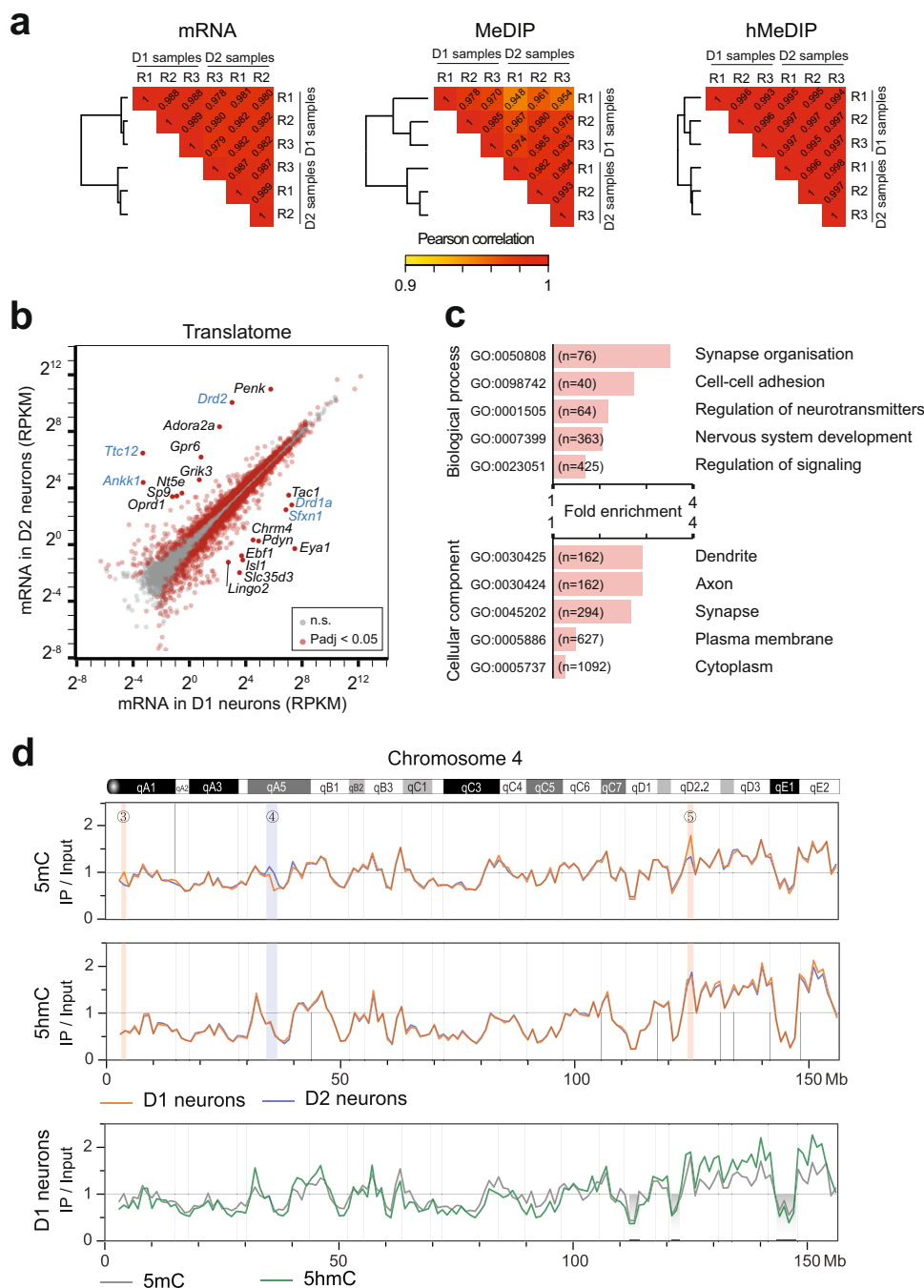


Fig. 2 Similarities and differences of striatal D1 and D2 translomes, methylomes, and hydroxymethylomes. **a** Hierarchical clustering and Pearson correlations between D1 and D2 samples. Rn, replicate #n. **b** Scatter plot of gene expression levels in D1 and D2 neurons. The most different gene names are indicated, with those included in the BAC in blue (excluded from further analysis). **c** Enriched gene ontology terms (FDR < 10^{-4} for biological processes and FDR < 10^{-21} for cellular components) identified with the differentially expressed genes ($n = 1963$). **d** MeDIP enrichment of 5mC and hMeDIP 5hmC (expressed as a number of reads in IP/number of reads in input per 1 megabase bin) throughout chromosome 4 (shown as an example), for D1 and D2 neurons. The blue and orange vertical shaded areas highlight the main D1/D2 differences in 5mC (numbering corresponds to Fig. 3a). Previously described 5mC deserts² are shaded gray. The chromosome regions and bands are indicated at the top.

Bonev et al.⁴⁵, but were smaller, and appeared to coincide with sub-TADs (Fig. 3f, Supplementary Fig. 3a–f). Importantly, these clusters contained D1/D2 differentially expressed genes, and the 5mC was higher in the cell type with lower expression of these genes. These observations indicate that the main differences in 5mC are found within clusters that map within structural domains and contain differentially expressed genes of potentially key importance for D1/D2 differences and striatal function.

5mC and 5hmC D1/D2 differences are mostly found in coding gene bodies. We then examined the location of the DNA modification differences. The observation that most of the differentially methylated clusters include several genes that show differential expression led us to examine in more detail whether the 5mC and 5hmC differences might correspond to specific genes, their bodies, or their regulatory elements. We first assessed whether the 1-kb windows of D1/D2 differences overlapped

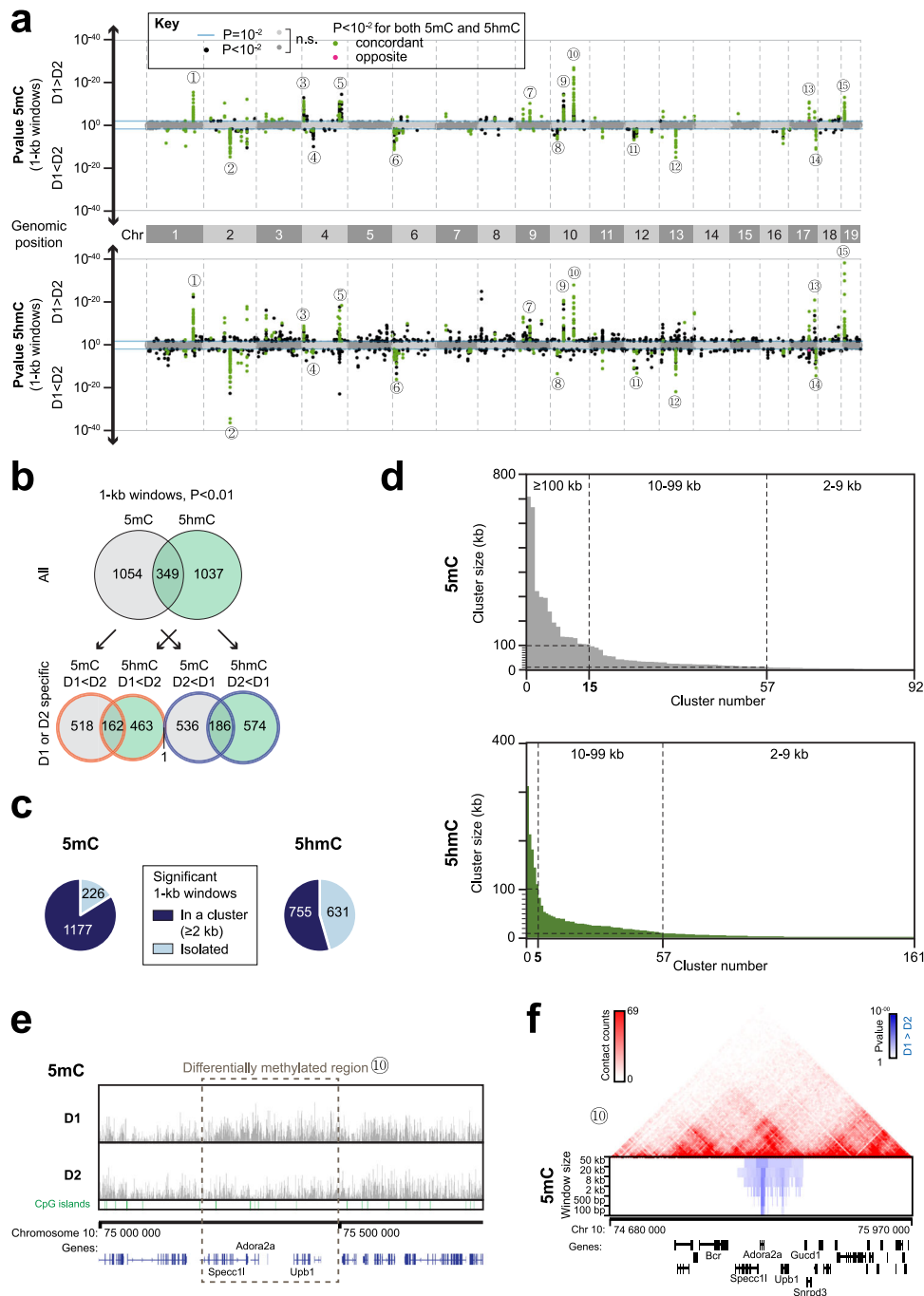


Fig. 3 Large clusters of D1/D2 5mC differences and more scattered 5hmC differences. **a** Manhattan plots of 5mC and 5hmC differences along the autosomes. Dots represent 1-kb windows. The P value is plotted upwards when the modification is enriched in D1 vs. D2 neurons, and downwards otherwise. Circled numbers represent the biggest clusters of windows, numbered from left to right, which are referred to in the text and in other panels of this figure. P, adjusted P value. **b** Venn diagrams of the significant 1-kb windows. **c** Clustering of the significant 1-kb windows. The pie charts show the proportions of 1-kb windows that are in a cluster (as defined in Methods). **d** Distribution of the cluster sizes for 5mC and 5hmC. Clusters are ordered from the largest to the smallest. **e** Genome browser representation of the differentially methylated cluster number 10 (as numbered in panel (a)). **f** Hi-C interaction frequencies from the study of Bonev et al.⁴⁵ displayed as a two-dimensional heatmap, superimposed with differentially methylated regions of cluster 10. The differentially methylated regions were analyzed with multiple window sizes as indicated (see Methods).

specific genomic features (Fig. 4a). Remarkably, even though the genic regions represent only a small proportion of the genome, most of the 5mC and 5hmC differences (72% and 62%, respectively) were located within coding genes, overlapping 185 and 365 genes, respectively (Supplementary Tables 8, 9). The vast majority of the differences were found within gene bodies (introns and exons). We also used the *OREGanno*⁴⁶ and *Cistrome*⁴⁷ databases

to assess whether differentially modified regions encompassed TFBSs. We observed that differential 5hmC windows overlapped TFBSs more often than the differential 5mC windows (Fig. 4a, b, Supplementary Fig. 4a). Interestingly, differential 5hmC windows were also enriched in TET1 sites (Fig. 4b), an enzyme that catalyzes the oxidation of 5mC into 5hmC. It should be noted that we probably underestimate the percentage of targeted TFBS since

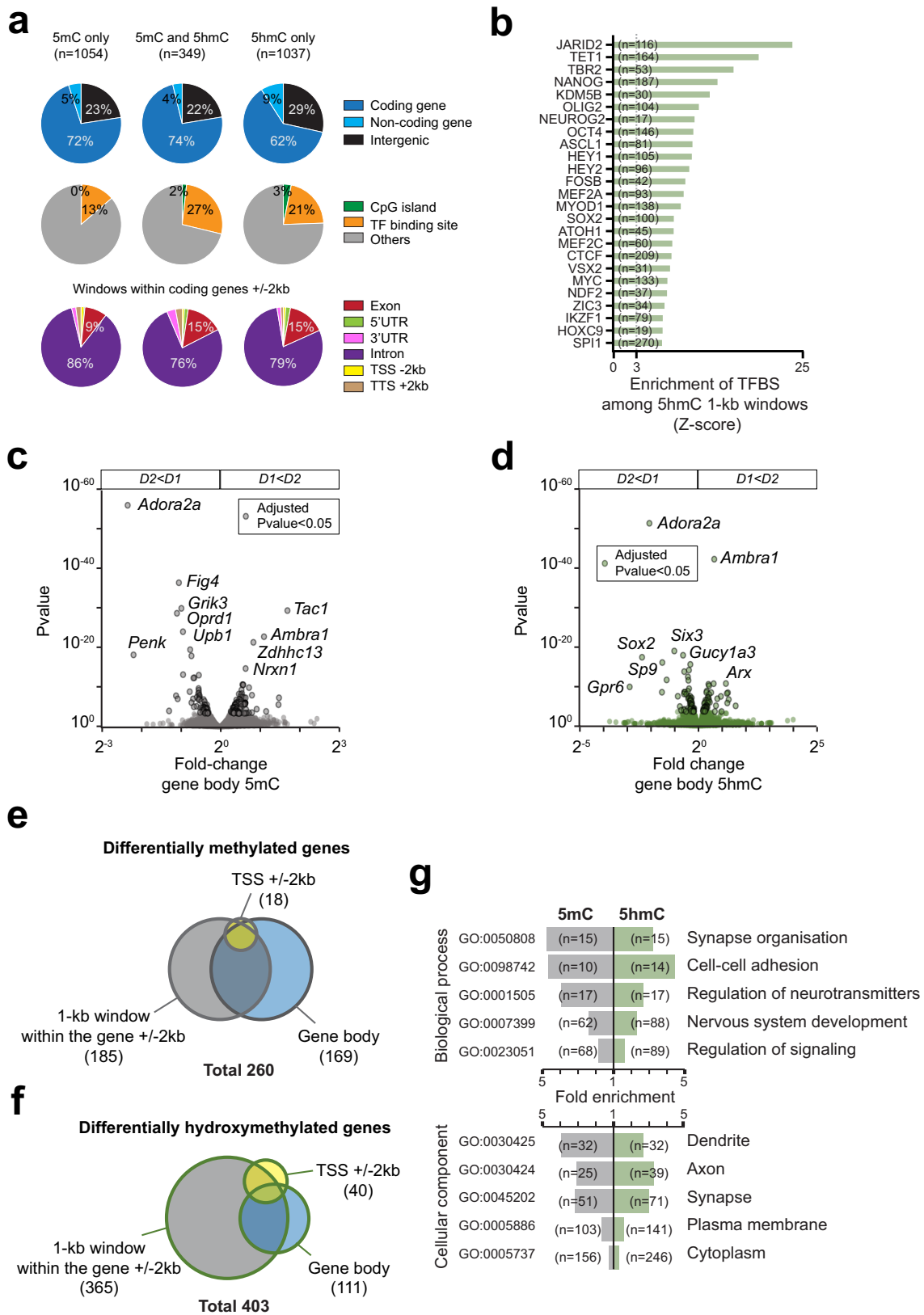


Fig. 4 D1/D2 differences in 5mC and 5hmC are mostly found in genes. **a** Overlap of D1/D2 5mC and 5hmC significantly different 1-kb windows with genomic features. **b** Top 25 TFBSs enriched among the significant 1-kb 5hmC windows. **c** Volcano plot of gene body 5mC differences. **d** Volcano plot of gene body 5hmC differences. **e** Venn diagram of genes with 5mC differences found by the TSS, gene body, or 1-kb windows analyses. **f** Same as in **(e)** for 5hmC. **g** Enriched gene ontology terms (FDR < 3×10^{-2} for biological processes and FDR < 2×10^{-2} for cellular components) for differentially methylated ($n = 258$) or hydroxymethylated ($n = 380$) genes, identified with the 1-kb window, the TSS, or the gene body analyses.

important ones are missing in databases (e.g., SP9 important for D2-SPNs differentiation³⁰).

As most of the differences for both DNA modifications were localized within gene bodies, we identified the differentially modified genes (from transcription start site [TSS] to transcription termination site [TTS], see Methods). We found 169 genes that differ for global gene body 5mC content between D1 and D2, (Fig. 4c, Supplementary Table 10), including genes characterized for their functions in D1-SPNs (e.g., *Tac1*⁴⁸) or D2-SPNs (e.g., *Adora2a*⁴⁹ and *Penk*⁵⁰). For 5hmC, we found 111 genes with differential gene body modification (Fig. 4d, Supplementary Table 11), including 53 of the differentially methylated genes (e.g., *Adora2a*, *Ambra1*) that were generally more methylated and hydroxymethylated in the same cell type (Supplementary Fig. 4b). Differentially hydroxymethylated genes included some genes encoding TFs necessary for striatal development, such as *Sp9*⁵¹ and *Six3*³¹, and other TFs of unexplored function to date in SPNs (e.g., *Sox2*, *Arx*). We also analyzed 5mC and 5hmC differences at the TSSs (± 2 kb) of genes and found 18 significantly different genes for 5mC (Supplementary Fig. 4c, Supplementary Table 12) and 40 for 5hmC (Supplementary Fig. 4d, Supplementary Table 13).

Overall, a total of 260 genes displayed a significant D1/D2 difference in 5mC, either at the TSS, at the gene body level, and/or more locally in a 1 kb window within the gene body ± 2 kb (Supplementary Table 14). Comparison of these various analyses showed that most of the 5mC differences between D1 and D2 were detected by total gene body comparison or 1-kb windows analysis (Fig. 4e). Hydroxymethylation was significantly different in 403 genes either at the TSS, at the gene body level, and/or more locally in a 1 kb-window within the gene body ± 2 kb (Supplementary Table 14). These 5hmC differences were more local than 5mC differences, mostly detected by the 1-kb windows analysis (Fig. 4f), in accordance with their enrichment at TFBSs as reported in other cell types^{52,53}. The differentially modified genes were predominantly related to neuronal-specific functions (Fig. 4g), similar to what we observed for gene expression. Thus, analyses of the D1/D2 differences show that differentially methylated and hydroxymethylated regions are highly enriched in genic regions. Hydroxymethylation frequently overlaps with TFBS, indicating a possible link with differential gene expression.

Differential 5mC and gene expression are inversely related. We first looked for a global relationship between gene expression and DNA modifications in D1- and D2-SPNs. We compared the 5mC at the TSS and gene bodies with the transcriptome ranked as quintiles of gene expression (Supplementary Fig. 5a) and found a global trend of higher gene expression associated with lower 5mC in both D1- (Fig. 5a) and D2-SPNs (Supplementary Fig. 5b). This pattern was similar to that in other neurons but contrasted with non-neuronal cell types^{2,9}. We observed a relative decrease in 5mC upstream of the TSS for all expressed genes, whereas the 5mC profile of silent protein-coding or non-coding genes was flat, and low for the latter, in agreement with previous studies^{2,9}. Despite these general trends, 5mC levels at the TSS and gene bodies were variable between genes and not predictive of expression levels on a gene per gene basis (Fig. 5b, Supplementary Fig. 5c). When we did the same analyses for 5hmC, the most expressed genes displayed a drop in 5hmC at the level of the TSS, which was less pronounced in less expressed genes (Fig. 5c, Supplementary Fig. 5b). In contrast, no systematic differences among expressed genes was observed in gene body regions, which showed consistently high levels of 5hmC, as observed previously in some neuronal types⁹. This illustrates the differences between 5mC and 5hmC distribution patterns, and indicates that 5hmC is

less associated with decreased gene expression than 5mC. No change in 5hmC was observed along the TSS or gene body regions for silent coding or non-coding genes. The 5hmC levels were also very variable between individual expressed genes and were not predictive of mRNA expression levels (Fig. 5d, Supplementary Fig. 5c).

We then assessed the relationship between the differences in DNA modifications and gene expression between D1 and D2 samples. Overall, 11% of the differentially expressed genes carried significant differences in DNA modifications (Supplementary Fig. 5d). Among the top 100 genes more expressed in either D1- or D2-SPNs (excluding the genes in the BACs), 60 carried significant differences in either DNA modifications, with 46 differentially methylated and 53 differentially hydroxymethylated (Fig. 5e). Hence, both DNA modifications appear to be highly relevant to SPNs identities. Among the genes differentially methylated between D1 and D2 neurons, there was a strong negative correlation between the 5mC fold-changes at the TSS or gene body, and mRNA changes (Fig. 5f). This indicated that in the D1/D2 comparison, a decrease in DNA 5mC at the TSS and/or gene body is a good predictor of increased expression. The results suggest an involvement of 5mC at both the TSS and gene body in regulating differential gene expression between D1/D2 neurons.

Local 5hmC is associated with differential expression in either direction. We then focused on D1/D2 5hmC differences and investigated their relation with translating mRNA levels. In genes differentially hydroxymethylated in D1 and D2 neurons, we observed an inverse correlation of TSS or gene body 5hmC with mRNA levels (Supplementary Fig. 5e). However, these correlations were less tight than for 5mC. D1/D2 differences in DNA 5mC and 5hmC in gene bodies were generally correlated (Fig. 5g), with three exceptions (*Cntnap3*, *Ebf1*, and *Ptprm*) in which 5hmC levels were higher in the cell type in which 5mC was lower and expression higher. Because most of the 5hmC differences were local and found through the 1-kb window analysis (see above, Fig. 4f), we assessed whether the genes showing these local differences were also differentially expressed. Among the genes including at least one 1 kb-window significant for 5hmC D1/D2 difference, 53% were also differentially expressed between D1 and D2 neurons (Fig. 5h). However, these differences went in either direction, with 37% less expressed in the cell population in which they were more hydroxymethylated, 9% more expressed, and the remaining 7% containing 5hmC changes in both directions. Because we had found that 5hmC correlates with 5mC, and to exclude the possibility that associated 5mC could explain the association of 5hmC with gene repression observed here, we excluded the genes also differentially methylated at the gene body and/or TSS and found similar results. Thus, local 5hmC differences can be associated with either up- or down-regulation of gene expression, in line with its possible role in regulating TFBS suggested above.

Whole-genome asymmetry of DNA 5mC and 5hmC. Our MeDIP and hMeDIP data allowed us to distinguish the 5mC/5hmC levels on each DNA strand. When we examined the two strands separately, we observed an asymmetry of DNA modifications that was more pronounced for 5hmC than for 5mC throughout the genome in both D1- and D2-SPNs (e.g., Fig. 6a, Supplementary Fig. 6a). The asymmetry was higher at smaller scales, although it was still present in intervals larger than 100 kb (Fig. 6b). In order to estimate the size of the asymmetric regions, we tested whether pairs of adjacent windows of a given size would have an asymmetry bias for the same strand more frequently than by chance. If so, that indicates that the tested windows are smaller

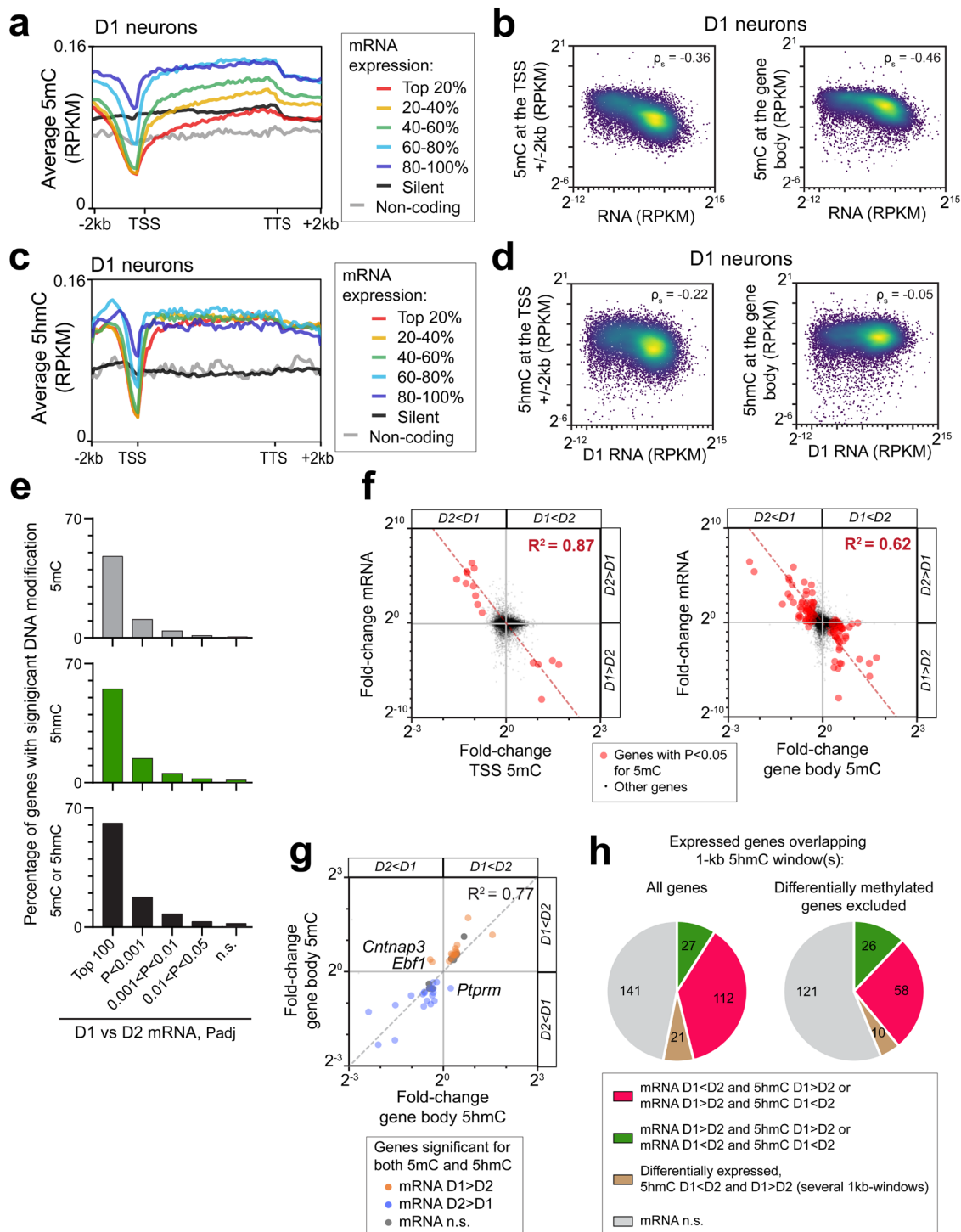


Fig. 5 Relation between gene 5mC and 5hmC, and mRNA levels. **a** Metagenes profiles of 5mC according to gene expression levels in D1 neurons. In these plots the gene body lengths are normalized while the preceding and following 2-kb are linear. **b** Density plots of 5mC levels at TSS or gene body, as a function of mRNA levels for expressed coding genes in D1 neurons. ρ_s , Spearman correlation coefficient. **c, d** Same as in **(a, b)** but for 5hmC. **e** Percentages of D1/D2 differentially expressed genes showing 5mC, 5hmC, or both differences (found by the TSS, gene body, or 1-kb windows analyses), depending on the significance of mRNA differences. **f** Scatter plots of the D1/D2 differences in mRNA levels as a function of the changes in 5mC at TSS or gene body. The linear regressions are shown. **g** Scatter plot of 5hmC changes compared to 5mC changes for genes with significant D1/D2 differences for both. **h** Expression changes of genes overlapping at least one 1-kb window significantly different for 5hmC between D1 and D2 neurons.

than the typical size of the asymmetric regions. We observed that both 5mC and 5hmC displayed the same pattern, with a bimodal distribution indicating the existence of asymmetrically modified regions with nested sizes typically smaller than 5 kb and bigger ones up to 1 Mb (Fig. 6c).

We assessed the asymmetry distribution with respect to identified regions, including genes and retrotransposons, which represent >40% of the genome. All retrotransposons were particularly asymmetric, and different families of transposons displayed different strand asymmetry (Fig. 6d, Supplementary

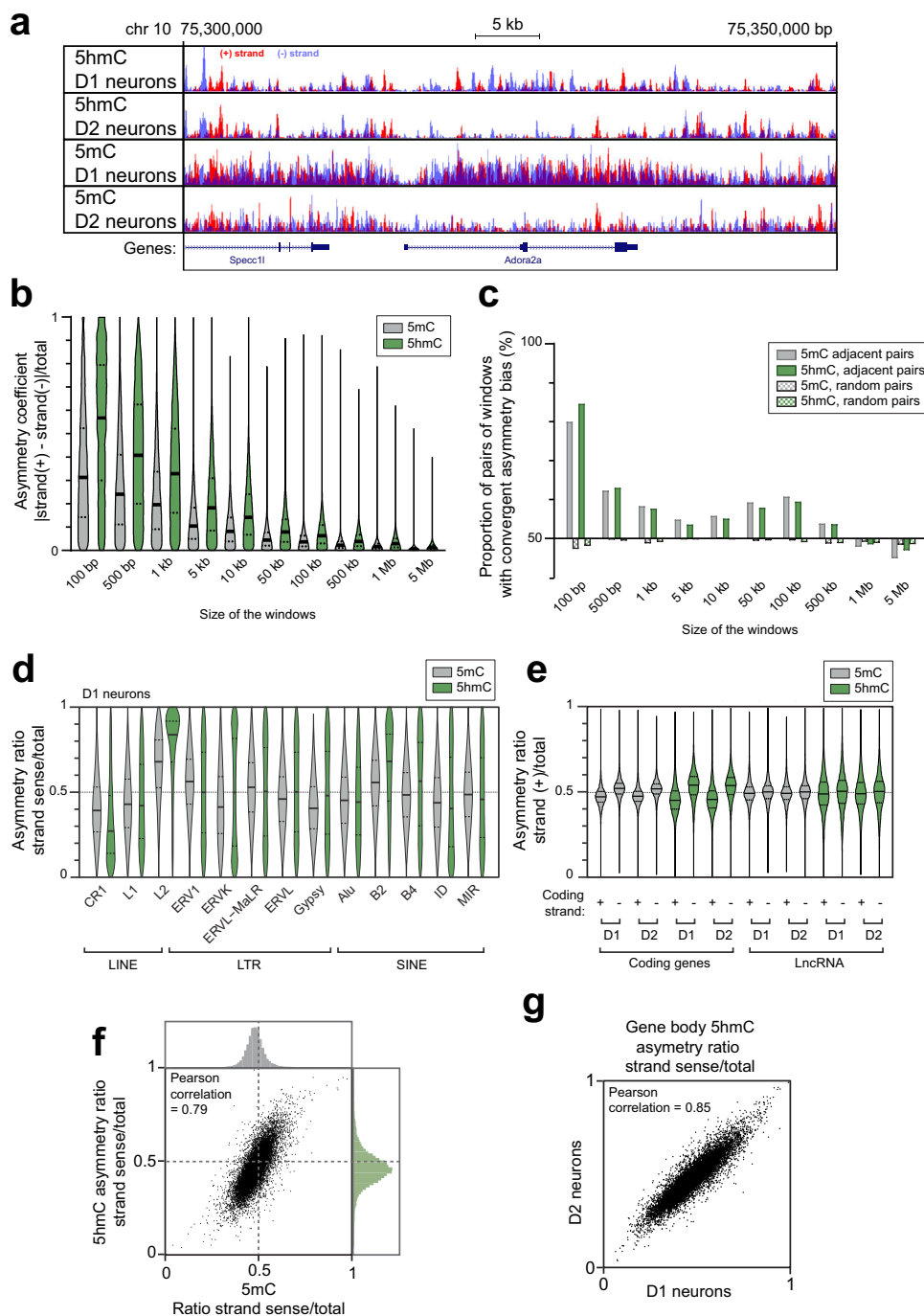


Fig. 6 Strand asymmetry of DNA modifications. **a** Example of genome browser view of 5hmC and 5mC strand distribution in D1 and D2 neurons. Peak sizes correspond to the number of reads on the (+) strand and (–) strand. **b** Asymmetry coefficient of 5mC and 5hmC depending on the size of the windows considered. Random sampling of 10,000 windows for each size (see Methods). **c** Evaluation of the size of the asymmetric regions of 5mC and 5hmC. For each window size, a random sampling of 10,000 windows was chosen, and compared to the adjacent window or to a random window on the genome. Pairs are considered convergent when they both have more reads on the same strand [(+) or (–)]. **d** Genome-wide distribution of strand asymmetry of 5mC and 5hmC in repetitive transposable elements in D1 neurons. **e** Genome-wide distribution of strand asymmetry of 5mC and 5hmC in coding genes and lncRNA genes. **f** Correlation of 5hmC and 5mC asymmetry ratios in gene bodies. **g** Correlation of 5hmC asymmetry ratios in gene bodies between D1 and D2 neurons. In **b**, **d**, **e**, horizontal solid lines are medians and dotted line quartiles.

Fig. 6b). Repetitive transposable elements were particularly asymmetric, including LINE-1 elements, which are known to be active in neurons⁵⁴, but also inactive elements (e.g., LINE-2).

We also observed that coding genes had more 5hmC on the template (i.e., non-coding) strand (Fig. 6e), but not in neighboring intergenic regions (Supplementary Fig. 6c). In contrast, long

non-coding RNAs (lncRNAs) had globally similar levels of DNA modifications on both strands (Fig. 6e). Although strand asymmetry was less pronounced for 5mC than for 5hmC, the asymmetries of 5mC and 5hmC were highly correlated (Fig. 6f). The asymmetry on the genes was more consistent between replicates for 5hmC than for 5mC (Supplementary Fig. 6d) and

was highly correlated between D1 and D2 neurons (Fig. 6g), with no significant difference between the two populations. DNA modification asymmetry did not depend on GC content or gene size (Supplementary Fig. 6e, f), but showed some degree of correlation with the strand ratio of CH (Supplementary Fig. 7a). Our data thus revealed a profound strand asymmetry of DNA modifications, in retrotransposons and coding genes, predominantly for 5hmC, and with a consistent organization between the two types of SPNs.

Asymmetric gene body 5hmC is associated with higher gene expression. Since modification asymmetry was high in coding genes, we asked whether the degree of 5hmC asymmetry was related to gene expression levels by comparing the strand asymmetry ratios for genes ranked by expression levels deciles (Fig. 7a, Supplementary Fig. 7b). Non-expressed genes displayed almost no asymmetry (median ratio 0.51 for the template/total), whereas the median ratio increased with expression and reached a plateau for the most expressed deciles. We checked the asymmetry of strand composition in relation to transcription levels and observed a similar bias in CH for the template strand, although less pronounced than 5hmC (Supplementary Fig. 7c). 5hmC asymmetry increases on average with the degree of gene expression, it is highly variable from gene to gene and therefore not predictive of expression on a gene per gene basis. The excess of 5hmC on the template strand was detected on gene bodies of highly expressed genes, but not at the TSS, where 5hmC on both strands dropped to the same low level (Fig. 7b). We did not detect differences in strand asymmetry for D1/D2 differentially expressed genes.

We then examined which genes were most represented among those with a high degree of 5hmC asymmetry (Fig. 7c) and identified highly expressed genes, such as those coding for ribosomal proteins. These genes were overrepresented among the 100 genes with the highest asymmetry ratio, and the whole family of ribosomal protein genes had a very high strand asymmetry (0.63 template/total, Fig. 7d). In contrast, olfactory receptor genes, which are virtually not expressed in the striatum, had particularly low ratios (0.48, Fig. 7c, d). Overall, our data reveal a link between DNA 5mC and 5hmC strand asymmetry and the degree of gene expression, which is much stronger in the case of 5hmC.

Discussion

In this study, we compared the methylomes, hydroxymethylomes, and translomes of two types of very closely related neurons, the D1- and D2-SPNs, characterized by the expression of dopamine D1 and D2 receptors, respectively. Previous reports on methylomes compared non-neuronal cells and neurons, or distant neuronal types^{2,9,34–38}. Here, the comparison between these two homogeneous and transcriptionally similar populations sheds light on the regulation and influence of cytosine modifications in the last steps of differentiation towards specialized neurons. We found that the two cytosine modifications display similarities and differences in their localization, relation to transcription, and distribution between the two DNA strands.

Methylation differences between D1- and D2-SPNs are grouped in clusters with lower 5mC in one or the other population, including fifteen large regions in the genome (>100 kb). These clusters include genes differentially expressed between D1 and D2 neurons, and more expressed in the population in which they are less methylated. These differential 5mC clusters contain differentially expressed genes characteristic of D1- or D2-SPNs identities or important for their respective functions (e.g., *Adora2a*⁴⁹, *Chrm4*⁵⁵, *Penk*⁵⁰, *Slc35d3*⁵⁶, *Gpr6*⁵⁷, *Gpr52*⁵⁸, and

*Nrxn1*⁵⁹). They also contain other genes that have been linked with striatal-related pathologies (e.g., *Lingo2*⁶⁰, *Nrxn2*⁶¹, *Mdk*⁶², *Grik3*⁶³, *Fig4*⁶⁴), as well as genes not previously identified as characterizing the two SPN types (e.g., *Dgkz*, *Ptprm*, and *Sntg2*). We suggest that these latter genes may also have important cell type-specific roles in striatal development and/or physiology. The extent of large 5mC clusters coincided with structural sub-domains within previously identified TADs. It is possible that 5mC is removed (or deposited) locally at some anchor points with the contribution of specific TFs⁶⁵ and that the modification spreads within a chromatin sub-domain. Alternatively, the transcriptional activity within the sub-domain during neuronal development could prevent DNA methyltransferase DNMT3A activity⁶⁶ which has been reported to play a role in depositing DNA 5mC de novo⁶⁷ and is essential for neuronal maturation⁶⁸. The existence of 5mC differences spanning several genes and intergenic regions (e.g., Fig. 3f) supports the first hypothesis, although, the second possibility is more in line with the preferential localization of 5mC differences to gene bodies (Fig. 4a). It is possible that the two scenarios are combined in various proportions.

DNA 5mC at gene TSSs is largely recognized as a repressive mark, whereas its role in gene bodies is more elusive. Gene body 5mC is positively correlated with transcription during development and has been proposed to facilitate transcription elongation⁸. In contrast, studies of various types of neurons and Bergmann glia have reported lower 5mC across the most highly expressed genes^{2,9,34,36}. In SPNs, we similarly observed an inverse correlation between gene body 5mC and mRNA expression, although the levels of 5mC were very variable on a gene-per-gene basis, as in other cell types. In contrast, in the D1/D2 comparison at the single gene level, a decrease in gene body 5mC in one population was actually a predictor of increased expression in this population. Hence, our data confirm and refine the association between both TSS and gene body 5mC with gene repression in SPNs and, presumably, in other fully differentiated neurons. The fact that gene body 5mC correlates with repression in neurons, unlike other cell types, has been suggested to originate from the presence of CH 5mC on gene bodies in neurons, but it can also be linked to the specific high abundance of the methylated DNA-binding protein MECP2 in neurons^{9,69}, which participates in transcriptional repression⁷⁰.

We found that DNA 5hmC differences between D1 and D2 neurons were partly correlated to 5mC differences, but they were more scattered, in clusters of smaller size, and frequently overlapped TFBSs. Even though 5hmC is an intermediate for demethylation^{4,5}, we observed that the 5mC and 5hmC profiles are distinct, and not related to transcription in the same manner, in agreement with previous observations. An inverse correlation was consistently found between 5mC and transcription, while this does not hold true for 5hmC. For example, an overall positive correlation between gene body 5hmC and gene expression was found for some brain cell types⁹. Our data show that differential 5hmC mostly affects TFBSs and that these localized 5hmC differences are associated with gene expression differences. In line with this, 5hmC has been shown to influence TF binding⁷¹. Thus, 5hmC has the potential to fine-tune the expression of particular genes, by influencing the binding of TFs.

Interestingly, we observed an asymmetry of methylation and 5hmC between the two strands of DNA genome-wide. Some degree of asymmetry was previously observed using bisulfite-based methods^{72–74}, but enrichment-based single-strand methods MeDIP and hMeDIP allowed us to further characterize and probe the extent of the asymmetries. Our data reveal a widespread asymmetry over the genome, of small-size regions (5 kb or less) nested in bigger regions up to 1 Mb. The asymmetry of 5hmC is

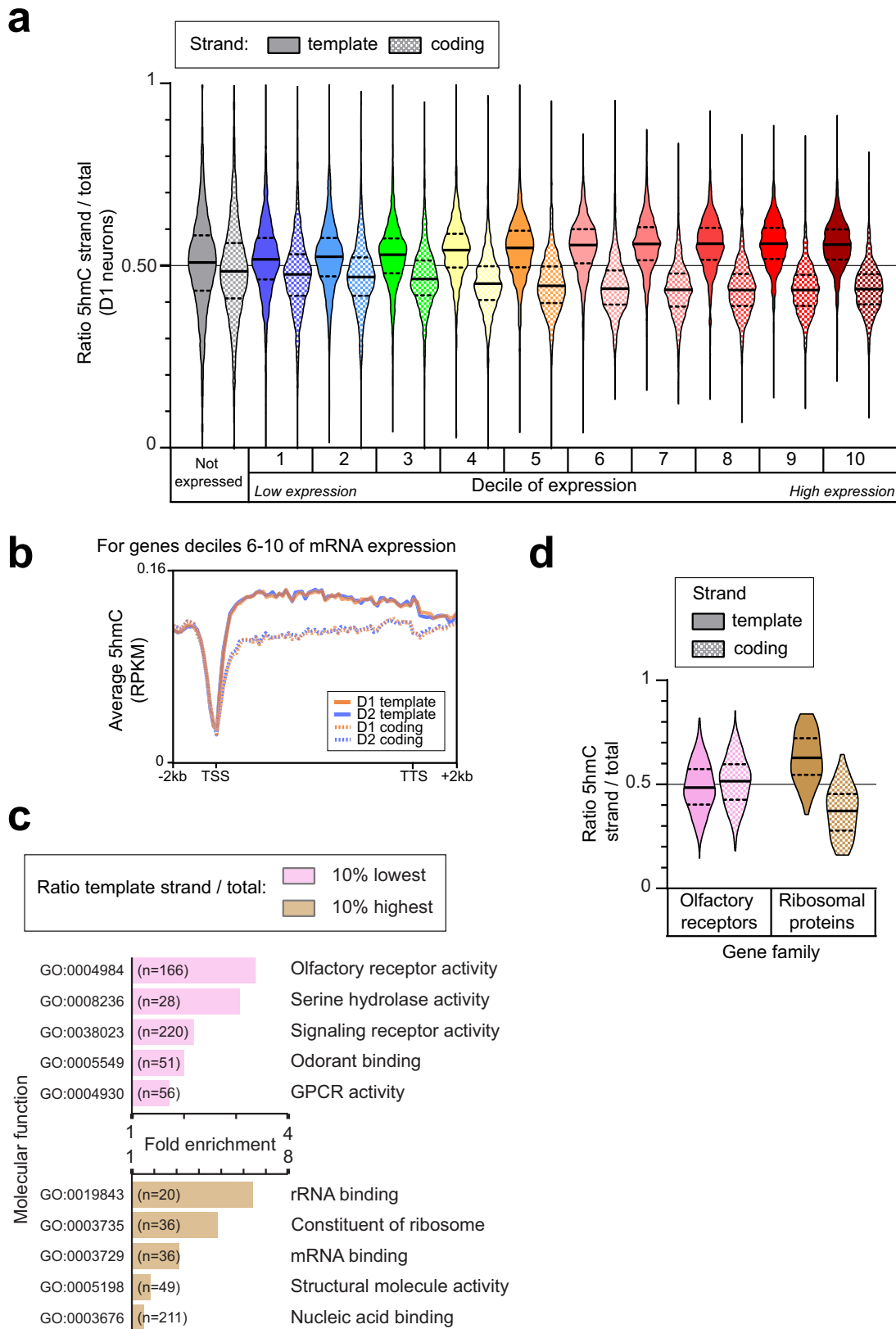


Fig. 7 5hmC asymmetry is associated with gene expression. **a** Violin plots of the 5hmC asymmetry ratios in D1 neurons on the template and coding strands of coding genes with different expression levels. **b** Metagene profiles of 5hmC on the template and coding strands, for the highly expressed genes (top 50% of expressed coding genes). **c** Enriched gene ontology terms (molecular function, $FDR < 2 \times 10^{-2}$) for genes with the 10% lowest or 10% highest 5hmC ratio of the template strand. **e** Violin plots of the 5hmC asymmetry ratios on template and coding strands, for olfactory receptors genes and ribosomal proteins. In **a** and **d**, horizontal solid lines are medians, and dotted lines are quartiles.

more pronounced than that of 5mC and is strongly associated with expression levels. Indeed, highly expressed genes such as ribosomal protein genes typically have more 5hmC on the template strand, whereas silent or lowly expressed genes have rather balanced 5hmC.

The asymmetry of 5hmC could be simply explained if 5hmC would occur in the CH context, where only one strand can be hydroxymethylated, rather than CG, where cytosines in both strands can be hydroxymethylated⁷⁵. However, previous bisulfite-based studies have shown that 5hmC is found mostly in a CG context, although not exclusively^{2,76}, and this would suggest that the 5hmC asymmetry is not directly dependent on base composition. Still, at the level of coding genes, we observed a correlation between CH and 5hmC asymmetries. Whether or not the CH content would be sufficient to explain the 5hmC asymmetry remains an open question, and would have to be investigated at the single-base level using bisulfite-based methods with extensive coverage.

Hydroxymethylation asymmetry in neurons could have functional consequences, notably for chromatin structure. Studies on synthetic (symmetric) hydroxymethylated DNA have shown that 5hmC can either enhance or decrease strand separation and affect the conformation of the double helix⁷⁷. The 5hmC hydroxyl group can participate in hydrogen bonding, causing higher solvation energy⁷⁸ and stabilizing DNA-protein interaction⁷⁹. Therefore, the distribution of 5hmC on DNA has the potential to participate in chromatin conformation and/or stabilization in neurons where 5hmC is particularly abundant. Transposable elements are known to be enriched in DNA modifications⁸⁰, including 5hmC⁸¹, and contain strand-specific methylation in non-CpG context⁷⁵. We found marked strand asymmetry in retrotransposons, for 5mC and even more for 5hmC. Transposons play a role in chromatin organization⁸², and the consequences of asymmetrical 5hmC on DNA conformation remain to be investigated. In transcribed genes, 5hmC could favor strand separation and facilitate transcription. Our data may suggest that asymmetry in transcribed genes is not a direct consequence of transcription, since differentially expressed genes between D1 and D2 do not show asymmetry differences.

Our work uncovers specific features of DNA modifications in neurons and characterizes the DNA 5mC and 5hmC differences between the two major populations of SPNs. It highlights the importance of DNA modifications in distinguishing the most differentially expressed genes in the two major populations of dopamine target cells in the striatum. Some of the key TFs for SPNs differentiation are differentially methylated or hydroxymethylated, such as *Ebf1* and *Sp9*. Differentially modified regions include yet uncharacterized genes that have the potential to be highly relevant for the identities or function of D1 and D2 neurons. The characterization of epigenetic differences between D1 and D2 SPNs also provides the necessary background for assessing their possible alterations in physiological and pathological conditions that differentially affect these two populations, including reinforcement learning, drug addiction, Huntington's disease, and Parkinson's disease. Genes identified in this study might be used as particular targets and contribute to the development of new therapeutic strategies in striatum-related pathologies.

Methods

Animals. BAC transgenic mice that express enhanced green fluorescent protein fused to the N-terminus of the large subunit ribosomal protein L10a under the control of dopamine D1a or D2 receptor promoter (*Drd1-EGFP-L10a* or *Drd2-EGFP-L10a*), generated as described²⁶, were maintained as heterozygotes on a C57Bl/6J background. Experiments with both lines were run in parallel. All the experiments were in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by Rockefeller University's

Institutional Animal Care and Use Committee. For all experiments, male and female mice were 2–3 month-old, generated by in vitro fertilization (Transgenic and Reproductive Technology Center, Rockefeller University). Animals were housed on a 12-h light-dark cycle, in stable conditions of temperature, with food and water *ad libitum*. Four mice were pooled for each sample, males and females were mixed, and the sex ratio was counterbalanced between groups. All conditions were run in triplicates.

Tissue preparation. Each sample consisted of whole striata from 4 mice. Each mouse was slightly anesthetized with CO₂ before being decapitated. Striata from both hemispheres were dissected and placed into ice-cold Hank's Balanced Salt Solution 1× (Invitrogen 10× solution: 12.6 mM CaCl₂, 4.92 mM MgCl₂, 4.07 mM MgSO₄, 53.3 mM KCl, 4.41 mM KH₂PO₄, 1380 mM NaCl, 3.36 mM Na₂HPO₄, 55.6 mM D-glucose) containing 2.5 mM HEPES-KOH pH 7.4, 35 mM glucose, 4 mM NaHCO₃ and 100 µg.mL⁻¹ cycloheximide. When the striata from the 4 mice of the same sample had been collected, they were placed in a 2-mL Dounce homogenizer (Dominique Dutscher, Brumath, France) containing 1 mL of homogenization buffer (20 mM HEPES-KOH pH 7.4, 5 mM MgCl₂, 150 mM KCl, 0.5 mM DL-dithiothreitol, 100 µg.mL⁻¹ cycloheximide, EDTA-free protease inhibitors (Roche), 400 U.mL⁻¹ Supersin (Life Technologies), 200 U.mL⁻¹ RNasin (Promega). Ten strokes of pestle A (clearance 76–127 µm) followed by 10 strokes of pestle B (clearance 12–63 µm) were applied gently to avoid damage to nuclei. The homogenate was then centrifuged at 2000 × g 10 min 4°C. The pellet and supernatant were separated to proceed with nuclei sorting and mRNA recovery, respectively.

Cell-type-specific mRNA recovery and sequencing. The supernatant was complemented with NP-40 (final concentration 1% vol/vol) and 1,2-dihexanoyl-sn-glycero-3-phosphocholine (final concentration 30 mM) and incubated for 5 min. It was then centrifuged at 20,000 × g for 10 min 4°C and the pellet was discarded. Immunoprecipitation was performed according to previous protocols^{26,83}. First, magnetic beads coated with anti-GFP antibody were prepared as follows: 300 µL of Streptavidin MyOne T1 Dynabeads (Invitrogen) per sample were washed in phosphate-buffered saline (PBS), incubated 35 min at room temperature (RT) with 120 µg of biotinylated protein L in PBS, washed 5 times with bovine serum albumin (BSA) 30 g.L⁻¹ in PBS, incubated 1 h at RT with 100 µg of monoclonal anti-GFP antibodies (50 µg clone 19F7 + 50 µg clone 19C8, Memorial Sloan-Kettering Monoclonal Antibody Facility, New York) in the homogenization buffer containing 1% (vol/vol) NP-40, washed 3 times and finally resuspended in 200 µL of homogenization buffer complemented with 1% (vol/vol) NP-40.

Magnetic beads coated with anti-GFP antibodies were added to the homogenates. After the addition of Supersin (final concentration 200 U.mL⁻¹, Life Technologies) and RNasin (final concentration 400 U.mL⁻¹, Promega), the samples were incubated for 16 h at 4°C under gentle end-over-end rotation. After 4 washes with homogenization buffer complemented with 1% (vol/vol) NP-40 and 200 mM KCl (total concentration KCl 350 mM), the RNA was eluted with RLT Plus buffer from the RNeasy Plus Micro kit (Qiagen) and 10 µL.mL⁻¹ β-mercaptoethanol (10-min incubation at RT and vortex). Then the RNA was purified according to the manufacturer's instructions, with an on-column DNase-I digestion step. The quantity of RNA was determined by fluorimetry using the Quant-iT Ribogreen, and its integrity was checked using the Bio-Analyzer Pico RNA kit before library preparation.

Ten nanograms of RNA were used for reverse transcription, performed with the Nugen Ovation RNaseq v2 kit. cDNAs were quantified by fluorometry, using the Quant-iT Picogreen reagent, and ultra-sonicated using a Covaris S2 sonicator with the following parameters: duty cycle 10%, intensity 5, 100 cycles/burst, 5 minutes. Two hundred nanograms of sonicated cDNA were then used for library construction using the Illumina TruSeq RNA sample prep kit, starting at the End-Repair step, and following the manufacturer's instructions. The libraries were quantified with the Bio-Analyzer High-sensitivity DNA kit, multiplexed, and sequenced on an Illumina HiSeq 2500 instrument. We obtained more than 40 million 50 bp paired-end reads per sample.

Cell-type specific DNA recovery. The pellet containing the nuclei (obtained after the 2000 × g centrifugation) was resuspended in a density solution with 29% iodixanol, prepared as follows: five volumes of Optiprep™ (Sigma Aldrich) containing 60% iodixanol were mixed with one volume of 150 mM KCl, 30 mM MgCl₂ and 120 mM Tris, pH 7.4. This 50% iodixanol solution was then further diluted to make a solution containing 29% iodixanol, using 250 mM sucrose, 25 mM KCl, 5 mM MgCl₂, and 20 mM Tris, pH 7.4, as a diluent. The pellet was dissociated by gently pipetting up and down 15 times. The homogenate in the 29% iodixanol solution was centrifuged at 10,000 × g 30 min 4°C (swinging buckets, TLS55 rotor, Beckman TL-100 ultracentrifuge). The nuclear pellet was resuspended in 250 mM sucrose, 25 mM KCl, 5 mM MgCl₂, 20 mM Tricine-KOH pH 7.4, 1% (vol/vol) donkey serum, 10 µM DyeCycle Ruby (Invitrogen). The rest of the sorting procedure was performed as in our previous study⁸⁴. The preparation was sorted with a FACSAria (BD) cell sorter equipped with 640 and 488 nm excitation lasers and an 85 µm nozzle. Nuclei were gated by two criteria: the signal from DyeCycle Ruby corresponding to single nuclei and a GFP signal above background fluorescence (as

assessed by comparison with nuclei from a wild-type littermate mouse). At least 100,000 nuclei were collected for each sample, with a purity >97%. Sorted nuclei in PBS were snap-frozen in liquid nitrogen.

DNA fragmentation and ligation. DNA extraction was performed after proteinase K and RNase A treatment, using a phenol/chloroform standard protocol. Glyco-blue™ (Life Technologies) was added at the precipitation step to avoid subsequent pellet loss. DNA was then resuspended in 130 μ L Tris pH 8.0. DNA was fragmented on an S2 Focused-ultrasonicator (COVARIS). The size of the fragments was assessed on a Bioanalyzer (Agilent Technologies) using a High Sensitivity DNA Kit. The mean size of the fragments was 150 bp for the MeDIP experiment and 250 bp for the hMeDIP experiment. Non-methylated TruSeq DNA adapters (synthesized by Sigma), with different indexes (for sample multiplexing before sequencing), were ligated using a SPRIworks Fragment Library System 1 kit (Beckman) on an SPRI-TE instrument, according to the Illumina TruSeq DNA sample prep kit protocol.

Hydroxymethylated/methylated DNA immunoprecipitation. The immunoprecipitation (IP) protocol was performed as follows, it was adapted from Weber et al.⁸⁵, with some modifications, including the ligation step with Y-adapters, performed prior to IP, to increase the yield and allow strand-specificity. It was miniaturized, and for this purpose, 200- μ L tubes were used. IP conditions were optimized beforehand using DNA from a similar number of non-GFP nuclei, which underwent all the previous steps, assessing the best conditions using a DNA 5mC control package (Diagenode). The obtained fragmented DNA with adapters in 80 μ L water was denatured at 95 °C 10 min and quickly cooled on ice. All the immunoprecipitation steps were then performed at 4 °C to keep the DNA in single-strand conformation. The IP buffer was 10 mM sodium phosphate buffer pH 7.0, 150 mM NaCl, 0.5% (vol/vol) Triton X-100. For each sample, 1 μ g of 5-methylcytosine monoclonal mouse antibody clone 33D3 (Diagenode Mab-081) was incubated for 2 h under gentle end-over-end rotation, then 10 μ L of anti-mouse IgG-coated magnetic beads (which had been previously washed with BSA 10 g.L⁻¹ in PBS and IP buffer) were added for overnight incubation (total volume 110 μ L). After 3 washes with the IP buffer, the DNA fraction bound to the beads was eluted by a 3-h incubation at 37 °C in a solution containing 10 mM EDTA, 50 mM Tris pH 8.0, 0.5% (vol/vol) sodium dodecyl sulfate, and 250 μ g.mL⁻¹ proteinase K, with shaking. The immunoprecipitated DNA was then purified using a standard phenol/chloroform extraction.

For hMeDIP, the IP was performed as for the MeDIP, also in a single-strand conformation, except that only 5 μ L of magnetic beads were used per sample, with 0.1 μ g 5-hydroxymethylcytosine monoclonal mouse antibody (Diagenode Mab-31HMC).

Both 5mC and 5hmC antibodies were previously validated. Datasheets showing the specificity of the antibodies for either 5mC or 5hmC are available on the manufacturer's website. These antibodies are not expected to show a preference for the CG or the CH context, as they were both generated using only a modified C as hapten. The IP conditions were optimized using the “5-hmC, 5-mC, and cytosine DNA standard pack for hMeDIP” (Diagenode, AF-107-0040). This kit contains hydroxymethylated, methylated, and unmethylated DNA standards, to ensure optimum conditions for the recovery of either methylated or hydroxymethylated DNA.

Library preparation of immunoprecipitated DNA and sequencing. The immunoprecipitated DNA was amplified for 12 cycles and purified with AMPureXP magnetic beads (Beckman Coulter Genomics) to remove fragments smaller than 100 bp. After quantification using Qubit and quality assessment with a Bioanalyzer, libraries were mixed in equimolar proportions and sequenced on an Illumina HiSeq 1000 instrument, running a single read 50 bp protocol using the P5 primer. We obtained more than 100 million reads per sample for the MeDIP experiment and 50 million reads for the hMeDIP experiment.

Strand specificity. The MeDIP and hMeDIP protocols included several specific adaptations: ligation of Y-unmethylated Illumina adapters right after fragmentation by sonication and before denaturation of the DNA and IP of single-stranded DNA. The choice to ligate the adapters prior to denaturation was originally done to improve the efficiency and use a limited number of sorted nuclei (100,000). In both MeDIP and hMeDIP experiments, DNA was denatured before immunoprecipitation of single strands, allowing us to distinguish the modification levels of each strand individually.

Read alignment. The sequencing quality was checked with FASTQC software (0.10.1)⁸⁶. The read library manipulations were performed using the FASTX-toolkit software suite (0.0.13). The reads were clipped according to their respective adapter sequence and trimmed according to their per base sequence quality for each library. The reads were aligned to the mm10 mouse downloaded from the UCSC genome using BWA (0.7.5)⁸⁷ for MeDIP and hMeDIP libraries and Tophat (2.0.10) for the RNA libraries. The aligned libraries were filtered for mapping quality “-q 30” using Samtools (0.1.19). After these quality steps, there were, on average, 108 \times 10⁶ reads per library for the MeDIP, 49.10⁶ reads per library for the

hMeDIP, and 105 \times 10⁶ reads for the RNA. Genomic views of read coverage were generated using Integrated Genomics Viewer tools and browsers (IGV 2.0; <http://www.broadinstitute.org/igv/>).

Statistics and reproducibility. *P* values were computed using R packages, as detailed below for each specific analysis. Pearson and Spearman correlation coefficients were computed using the cor() function in R. Pearson correlation coefficients were computed to assess the reproducibility between replicates.

Differential expression. Reads were assigned to protein-coding genes using Rsubread (v1.28.0). Differentially expressed genes were obtained using the Limma package with the voom function (3.36.5). A cutoff of adjusted *P*_{value} < 0.05 was used for differentially expressed genes. Detailed results from this analysis are supplied in Supplementary Table 1 and include *P* values and fold changes for the protein-coding genes.

MeDIP and hMeDIP windows analyses. The differentially methylated and hydroxymethylated regions were assessed using the R Bioconductor package MEDIPS (v1.16.0)⁸⁸, with several window sizes tested (100 bp, 500 bp, 1 kb, 2 kb, 8 kb, 20 kb, 50 kb). Results from the 1-kb analyses are supplied in Supplementary Tables 2 and 3. The complete lists of genes overlapping 1-kb windows with *p* < 0.01 are provided in Supplementary Tables 8 and 9.

Clustering. To estimate the size of differentially modified DNA regions between D1- and D2-SPNs, we first grouped 1-kb windows, which were significantly differentially modified and were less than 5 kb from each other. We then included neighboring significant 1-kb windows as long as the density of significant windows in the cluster was above 1/8, to obtain the final clusters. We only grouped windows that were significant in the same direction (e.g., both more methylated in D2). Detailed results from these analyses are supplied in Supplementary Tables 4 and 6. The lists of the genes overlapping the largest clusters are provided in Supplementary Tables 5 and 7.

Feature analysis. Differentially methylated or hydroxymethylated windows were assigned to the genomic features using Bedtools (2.29.2). The TFBSs used as a reference originated from the ORegAnno database⁴⁶.

Transcription factor analysis. Mouse TFBSs were downloaded from the Cistrome database⁴⁷. For each TF, the overlap with significant 1-kb windows was quantified using Bedtools (2.29.2). In order to compute a z-score, the average random overlap and standard deviation were obtained by shuffling the binding sites intervals 100 times on the mouse genome.

MeDIP and hMeDIP, TSS, or gene body analyses. Reads were assigned to genes or TSSs using Rsubread (v1.28.0). The differentially methylated or hydroxymethylated genes were obtained with the R Bioconductor package DESeq2 (v1.27.32)⁸⁹ without independent filtering. A cutoff of adjusted *P*_{value} < 0.05 was used for differentially modified genes or TSSs. Detailed results from these analyses are supplied in Supplementary Tables 10–13 and include *P* values and fold changes for all the genes.

Gene ontologies. Gene ontologies enrichment analysis was performed with the GO Consortium online tool (<http://geneontology.org/>)^{90,91}, using all the *Mus musculus* genes in the database as the common reference list for all the analyses.

Metagene. Metagene plots were done using the R package Metagene (v2.4.3).

Venn diagrams. The Venn diagrams were built using the <https://www.stefanijol.nl/venny> resource.

Strand-specific visualization. Bam files were split into Forward and Reverse reads, and replicates were merged using Samtools (0.1.19). Genomic views of read coverage were generated using IGV 2.0.

Asymmetry coefficient of DNA modifications. To assess the distribution of the asymmetry coefficient at different window sizes, 10,000 random windows of each size were selected using Bedtools shuffle. The asymmetry coefficient was computed as follows: |number of reads on the (+) strand—number of reads on the (–) strand|/total number of reads. The windows with less than 20 reads total were excluded.

Size of the asymmetric regions. To estimate the size of the asymmetric regions, we first selected 10,000 random windows of each size (from 100 bp to 5 Mb) using Bedtools shuffle. For each window, we determined which strand had more reads. We then assessed whether the downstream neighboring window of the same size

had a bias towards the same strand or not. As a control, the windows were also compared to random windows on the genome. If adjacent pairs of windows of a given size tend to have a concordant bias compared to chance, it implies that the typical asymmetric regions are bigger than the size of the windows considered. The windows with less than 20 reads total were excluded.

Asymmetry coefficient of the CH content. The R package Rtracklayer (v1.50.0) was used to import the bed file of the regions of interest. The full genome sequences were obtained from the package BSgenome.Mmusculus.UCSC.mm10 (3.15). We manipulated the genomic intervals with GenomicRanges (v3.15) and analyzed the DNA content with BSgenome (v1.58.0).

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequencing data (FASTQ files) generated in this study have been deposited in the GEO database under accession code GSE186572. All other data that support this study are in the Supplementary Tables and available from the corresponding authors upon reasonable request.

Received: 10 January 2022; Accepted: 16 November 2022;

Published online: 01 December 2022

References

1. Schubeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
2. Lister, R. et al. Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
3. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
4. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
5. He, Y. F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
6. Spruijt, C. G. et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
7. Iwase, S. et al. Epigenetic etiology of intellectual disability. *J. Neurosci.* **37**, 10773–10782 (2017).
8. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
9. Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012).
10. Gabel, H. W. et al. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89–93 (2015).
11. Cholewa-Waclaw, J. et al. The role of epigenetic mechanisms in the regulation of gene expression in the nervous system. *J. Neurosci.* **36**, 11427–11434 (2016).
12. Guan, J. S., Xie, H. & Ding, X. The role of epigenetic regulation in learning and memory. *Exp. Neurol.* **268**, 30–36 (2015).
13. Mink, J. W. The Basal Ganglia and involuntary movements: impaired inhibition of competing motor patterns. *Arch. Neurol.* **60**, 1365–1368 (2003).
14. Grillner, S., Robertson, B. & Stephenson-Jones, M. The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *J. Physiol.* **591**, 5425–5431 (2013).
15. Gerfen, C. R. & Young, W. S. 3rd Distribution of striatonigral and striatopallidal peptidergic neurons in both patch and matrix compartments: an in situ hybridization histochemistry and fluorescent retrograde tracing study. *Brain Res.* **460**, 161–167 (1988).
16. Gokce, O. et al. Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-seq. *Cell Rep.* **16**, 1126–1137 (2016).
17. Schultz, W. Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *J. Neurophysiol.* **56**, 1439–1461 (1986).
18. Bromberg-Martin, E. S., Matsumoto, M. & Hikosaka, O. Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. *Neuron* **67**, 144–155 (2010).
19. Kravitz, A. V., Tye, L. D. & Kreitzer, A. C. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat. Neurosci.* **15**, 816–818 (2012).
20. Tecuapetla, F., Jin, X., Lima, S. Q. & Costa, R. M. Complementary contributions of striatal projection pathways to action initiation and execution. *Cell* **166**, 703–715 (2016).
21. Gibb, W. R. Functional neuropathology in Parkinson's disease. *Eur. Neurol.* **38**, 21–25 (1997).
22. Fasano, S. & Brambilla, R. Cellular mechanisms of striatum-dependent behavioral plasticity and drug addiction. *Curr. Mol. Med.* **2**, 649–665 (2002).
23. Figge, D. A., Eskow Jaunarajs, K. L. & Standaert, D. G. Dynamic DNA methylation regulates levodopa-induced dyskinesia. *J. Neurosci.* **36**, 6514–6524 (2016).
24. Hamilton, P. J. & Nestler, E. J. Epigenetics and addiction. *Curr. Opin. Neurobiol.* **59**, 128–136 (2019).
25. Doyle, J. P. et al. Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).
26. Heiman, M. et al. A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
27. Knowles, R., Dehorter, N. & Ellender, T. From progenitors to progeny: shaping striatal circuit development and function. *J. Neurosci.* **41**, 9483–9502 (2021).
28. Lobo, M. K., Yeh, C. & Yang, X. W. Pivotal role of early B-cell factor 1 in development of striatonigral medium spiny neurons in the matrix compartment. *J. Neurosci. Res.* **86**, 2134–2146 (2008).
29. Lu, K. M., Evans, S. M., Hirano, S. & Liu, F. C. Dual role for Islet-1 in promoting striatonigral and repressing striatopallidal genetic programs to specify striatonigral cell identity. *Proc. Natl Acad. Sci. USA* **111**, E168–E177 (2014).
30. Zhang, Q. et al. The zinc finger transcription factor Sp9 is required for the development of striatopallidal projection neurons. *Cell Rep.* **16**, 1431–1444 (2016).
31. Song, X. et al. Homeobox gene Six3 is required for the differentiation of D2-type medium spiny neurons. *Neurosci. Bull.* **37**, 985–998 (2021).
32. Schambra, U. B. et al. Ontogeny of D1A and D2 dopamine receptor subtypes in rat brain using in situ hybridization and receptor binding. *Neuroscience* **62**, 65–85 (1994).
33. Araki, K. Y., Sims, J. R. & Bhide, P. G. Dopamine receptor mRNA and protein expression in the mouse corpus striatum and cerebral cortex during pre- and postnatal development. *Brain Res.* **1156**, 31–45 (2007).
34. Mo, A. et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015).
35. Sharma, A., Klein, S. S., Barboza, L., Lohdi, N. & Toth, M. Principles governing DNA methylation during neuronal lineage and subtype specification. *J. Neurosci.* **36**, 1711–1722 (2016).
36. Kozlenkov, A. et al. A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Sci. Adv.* **4**, eaau6190 (2018).
37. Rizzardi, L. F. et al. Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. *Nat. Neurosci.* **22**, 307–316 (2019).
38. Noack, F. et al. Assessment and site-specific manipulation of DNA (hydroxy-) methylation during mouse corticogenesis. *Life Sci. Alliance* **2**, e201900331 (2019).
39. Shen, L. & Zhang, Y. 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Curr. Opin. Cell Biol.* **25**, 289–296 (2013).
40. Nair, S. S. et al. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics* **6**, 34–44 (2011).
41. Ito, S. et al. Loss of neuronal 3D chromatin organization causes transcriptional and behavioural deficits related to serotonergic dysfunction. *Nat. Commun.* **5**, 4450 (2014).
42. Lobo, M. K., Karsten, S. L., Gray, M., Geschwind, D. H. & Yang, X. W. FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nat. Neurosci.* **9**, 443–452 (2006).
43. Kronman, H. et al. Biology and bias in cell type-specific RNAseq of nucleus accumbens medium spiny neurons. *Sci. Rep.* **9**, 8350 (2019).
44. Montalban, E. et al. Translational profiling of mouse dopaminergic neurons reveals region-specific gene expression, exon usage, and striatal prostaglandin E2 modulatory effects. *Mol. Psychiatry* **27**, 2068–2079 (2022).
45. Bonev, B. et al. Multiscale 3D genome wiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
46. Lesurf, R. et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res.* **44**, D126–D132 (2016).
47. Vorontsov, I. E. et al. Genome-wide map of human and mouse transcription factor binding sites aggregated from ChIP-Seq data. *BMC Res. Notes* **11**, 756 (2018).
48. He, Z. X. et al. Nucleus accumbens Tac1-expressing neurons mediate stress-induced anhedonia-like behavior in mice. *Cell Rep.* **33**, 108343 (2020).
49. Li, Y. et al. Optogenetic activation of adenosine A2A receptor signaling in the dorsomedial striatopallidal neurons suppresses goal-directed behavior. *Neuropsychopharmacology* **41**, 1003–1013 (2016).

50. Steiner, H. & Gerfen, C. R. Enkephalin regulates acute D2 dopamine receptor antagonist-induced immediate-early gene expression in striatal neurons. *Neuroscience* **88**, 795–810 (1999).
51. Xu, Z. et al. SP8 and SP9 coordinately promote D2-type medium spiny neuron production by activating Six3 expression. *Development* **145**, dev165456 (2018).
52. Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* **12**, R54 (2011).
53. Tsagaratou, A. et al. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. *Proc. Natl Acad. Sci. USA* **111**, E3306–E3315 (2014).
54. Ferrari, R., Grandi, N., Tramontano, E. & Dieci, G. Retrotransposons as drivers of mammalian brain evolution. *Life (Basel)* **11**, 376 (2021).
55. Jeon, J. et al. A subpopulation of neuronal M4 muscarinic acetylcholine receptors plays a critical role in modulating dopamine-dependent behaviors. *J. Neurosci.* **30**, 2396–2405 (2010).
56. Zhang, Z. et al. Mutation of SLC35D3 causes metabolic syndrome by impairing dopamine signaling in striatal D1 neurons. *PLoS Genet.* **10**, e1004124 (2014).
57. Lobo, M. K., Cui, Y., Ostlund, S. B., Balleine, B. W. & Yang, X. W. Genetic control of instrumental conditioning by striatopallidal neuron-specific S1P receptor Gpr6. *Nat. Neurosci.* **10**, 1395–1397 (2007).
58. Nishiyama, K., Suzuki, H., Maruyama, M., Yoshihara, T. & Ohta, H. Genetic deletion of GPR52 enhances the locomotor-stimulating effect of an adenosine A2A receptor antagonist in mice: A potential role of GPR52 in the function of striatopallidal neurons. *Brain Res.* **1670**, 24–31 (2017).
59. Alabi, O. O. et al. Disruption of Nrxn1alpha within excitatory forebrain circuits drives value-based dysfunction. *Elife* **9**, e54838 (2020).
60. Wu, Y. W. et al. Lingo2 variants associated with essential tremor and Parkinson's disease. *Hum. Genet.* **129**, 611–615 (2011).
61. Sebate, B. et al. Prioritization of candidate genes for a South African family with Parkinson's disease using in-silico tools. *PLoS ONE* **16**, e0249324 (2021).
62. Gramage, E. et al. Regulation of extinction of cocaine-induced place preference by midkine is related to a differential phosphorylation of peroxiredoxin 6 in dorsal striatum. *Behav. Brain Res.* **253**, 223–231 (2013).
63. Preuss, U. W. et al. Ionotropic glutamate receptor gene GRIK3 SER310ALA functional polymorphism is related to delirium tremens in alcoholics. *Pharmacogenomics J.* **6**, 34–41 (2006).
64. Chow, C. Y. et al. Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J. *Nature* **448**, 68–72 (2007).
65. Luo, C., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: in the right place at the right time. *Science* **361**, 1336–1340 (2018).
66. Stroud, H. et al. Early-life gene expression in neurons modulates lasting epigenetic states. *Cell* **171**, 1151–1164.e16 (2017).
67. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
68. Zocher, S. et al. De novo DNA methylation controls neuronal maturation during adult hippocampal neurogenesis. *EMBO J.* **40**, e107100 (2021).
69. Meehan, R. R., Lewis, J. D. & Bird, A. P. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.* **20**, 5085–5092 (1992).
70. Kinde, B., Gabel, H. W., Gilbert, C. S., Griffith, E. C. & Greenberg, M. E. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc. Natl Acad. Sci. USA* **112**, 6800–6806 (2015).
71. Lercher, L. et al. Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* **50**, 1794–1796 (2014).
72. Yu, M. et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
73. Ficiz, G. et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
74. Wen, L. et al. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol.* **15**, R49 (2014).
75. Guo, W., Chung, W. Y., Qian, M., Pellegrini, M. & Zhang, M. Q. Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells. *Nucleic Acids Res.* **42**, 3009–3016 (2014).
76. Mellen, M., Ayata, P. & Heintz, N. 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. *Proc. Natl Acad. Sci. USA* **114**, E7812–E7821 (2017).
77. Severin, P. M., Zou, X., Schulten, K. & Gaub, H. E. Effects of cytosine hydroxymethylation on DNA strand separation. *Biophys. J.* **104**, 208–215 (2013).
78. Renciuik, D., Blacque, O., Vorlickova, M. & Spingler, B. Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.* **41**, 9891–9900 (2013).
79. Frauer, C. et al. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306 (2011).
80. Deniz, O., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **20**, 417–431 (2019).
81. de la Rica, L. et al. TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.* **17**, 234 (2016).
82. Zhang, D., Lam, J. & Blobel, G. A. Engineering three-dimensional genome folding. *Nat. Genet.* **53**, 602–611 (2021).
83. Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P. & Heintz, N. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protoc.* **9**, 1282–1291 (2014).
84. Jordi, E. et al. Differential effects of cocaine on histone posttranslational modifications in identified populations of striatal neurons. *Proc. Natl Acad. Sci. USA* **110**, 9511–9516 (2013).
85. Weber, M. et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853–862 (2005).
86. Wingett, S. W. & Andrews, S. FastQ Screen: a tool for multi-genome mapping and quality control. *F1000Res* **7**, 1338 (2018).
87. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
88. Lienhard, M., Grimm, C., Morkel, M., Herwig, R. & Chavez, L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30**, 284–286 (2014).
89. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
90. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
91. The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).

Acknowledgements

The manuscript is dedicated to the memory of Paul Greengard, who passed away on April 13th, 2019, before the work was completed. We thank Samuel Collombet, Delphine Naquin, Yuvia Alheli Perez Rico, Anna Ramisch, Maud Silvain, and Aurélie Teissandier for bioinformatics advice, Sarah Meller for her assistance with the animals, Sophie Ajjan for advice concerning the MeDIP, members of the Girault & Hervé lab for stimulating discussions, Edith Heard and Stéphane Jamain for their critical reading of the manuscript. We are grateful to Songyan Han, Svetlana Mazel, Stanka Semova and Selamawit Tedese for assistance with the Flow Cytometry, which was carried out at the Flow Cytometry Resource Center (FCRC, Rockefeller University). High-throughput sequencing of the mRNA was performed at the Genomics Resource Center (Rockefeller University, New York), while the MeDIP and hMeDIP were sequenced at the NGS Core Facility of I2BC (Institut de Biologie Intégrative de la Cellule, CNRS, Gif-sur-Yvette, France). This work was supported by Inserm and Sorbonne Université, and grants from European Research Council (ERC, AIG-250349), Biology for Psychiatry Laboratory of Excellence (Labex Bio-Psy, Investissements d'avenir, ANR-11-IDEX-0004-02), Fondation pour la Recherche Médicale (FRM # DPA20140629798 & FDT201805005390), and ANR Epitraces (Project-ANR-16-CE16-0018) to J.A.G.

Author contributions

J.A.G. conceived and supervised the project. D.H., J.A.G., J.P.R., and L.M.P. designed the experiments. E.J., J.P.R., L.M.P., and Y.J. performed experiments. C.D.D., E.M., L.M.P., L.T., and N.S. analyzed data. C.T., D.B., D.H., J.A.G., L.M.P., L.T., and P.G. interpreted and discussed the data. J.A.G. and L.M.P. wrote the paper, with critical revision of the article by C.D.D., D.B., D.H., E.M., J.P.R., and N.S. All the authors approved the final version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-04269-w>.

Correspondence and requests for materials should be addressed to Lucile Marion-Poll or Jean-Antoine Girault.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Alex Nord and Manuel Breuer.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022