Epigenetic mechanisms governing cell type specific somatic expansion and toxicity in Huntington's disease

Matthew Baffuto[1,3], Kert Mätlik[1,3], Isaac Ilyashov[1], Maria Esterlita Siantoputri[1], Hasnahana Chetia[1], Yurie Maeda[1], Erika Sipos[1], Paul Darnell[1], Laura Kus[1], Thomas S. Carroll[2], Douglas Barrows[2], Christina Pressl[1], Nicholas Didkovsky[1], Nathaniel Heintz[1*]

[1]Laboratory of Molecular Biology, The Rockefeller University, New York, NY, United States of America
[2]Bioinformatics Resource Center, The Rockefeller University, New York, NY, United States of America
[3]These authors contributed equally
*Corresponding Author

Summary:

Huntington's disease (HD) is characterized by neuronal dysfunction and degeneration that varies markedly by brain region and cell type. We previously showed that CAG repeat expansion in exon 1 of the *mHTT* gene correlates with increased expression of the mismatch repair genes *MSH2* and *MSH3* in striatal medium spiny neurons[1], and demonstrated that, in the striatum and cerebral cortex of individuals with HD, hundreds of genes are dysregulated in neuronal cell types carrying somatically expanded CAG repeat in *mHTT*[1,2]. Here we employ comprehensive epigenetic profiling in specific neuronal and glial cell types from the human striatum, cerebral cortex, hippocampus and cerebellum of control and HD donor samples to identify cell type- and species-specific transcriptional control mechanisms in the mismatch repair genes *MSH2*, *MSH3* and *FAN1* that can explain the specificity of somatic CAG expansion in the first stage of HD. In the second, toxic phase of HD we identify two distinct epigenetic mechanisms that disrupt regulation of hundreds of genes in the majority of HD MSNs, including several that cause haploinsufficient neurological disorders. Our data support a mechanistic model of HD pathogenesis in which regulation of mismatch repair gene transcription determines the selectivity of somatic expansion, and DNA methylation stabilizes the toxic effect of mutant huntingtin on HD-modifying proteins MED15 and TCERG1, which regulate enhancer function and transcription elongation.

Introduction:

Huntington's disease (HD) is a fatal late-onset neurodegenerative disease caused by inheritance of an abnormally long CAG repeat in exon 1 of the Huntingtin gene (*HTT*) [3]. Studies of human tissue samples have established that somatic expansion of the mutant *HTT* (*mHTT*) CAG repeat occurs in specific brain regions [4] and cell types [1,2,5]. Genetic studies in humans and mouse models have demonstrated that genes involved in DNA mismatch repair modulate the age at onset of HD [6-8] and the capacity for somatic CAG expansion of the *mHTT* allele in the mouse central nervous system (CNS) [reviewed in [9]]. In the human striatum, somatic CAG expansion in MSNs coincides with high levels of nuclear MutS homologues 2 and 3 (MSH2, MSH3), which together comprise MutSβ [1]. *In vitro*, MutSβ inhibits the nucleolytic excision of DNA molecules with CAG slip-outs by FANCD2 and FANCI Associated Nuclease 1 (FAN1), thus promoting somatic *mHTT* CAG expansion [1]. Although DNA mismatch repair enzyme-dependent somatic CAG expansion is considered to be a necessary initial step in HD pathogenesis [10], the mechanisms that govern CNS cell type-specificity of the somatic CAG expansion process have not been determined.

Cell types carrying *mHTT* allele with somatically expanded CAG repeat have severe HD-associated transcript level changes [1,2]. These data support a model in which a second step of HD pathogenesis is caused by disruption of transcriptional regulation, leading to cell loss [11]. The identification of genes encoding transcriptional regulators MED15 and TCERG1 as HD progression-modifying genes in humans [7,8], and the demonstration that these genes do not alter somatic expansion in HD mouse models [12,13], support this hypothesis. Studies in HD mouse models have demonstrated that transcriptional dysregulation is accompanied by a variety of epigenetic changes, including altered epigenetic aging, measured based on selected methylated cytosines [14], downregulation of cell identity genes [15,16], changes in post-translational histone modifications (PTMs) [17,18], reversal of polycomb-dependent developmental repression [19], and disease locus specific alterations of higher order chromatin structure [20]. Although these studies suggest that altered epigenetic control of gene expression may occur in HD, there is no evidence available identifying cell type specific epigenetic mechanisms and their contributions to transcriptional disturbances in response to somatic expansion of *mHTT* in the human brain.

Here we have used high resolution epigenetic analysis of cell types in the striatum, cerebral cortex, hippocampus, and cerebellum to understand the specificity of somatic expansion in the HD brain, and to identify mechanisms of toxicity in cell types carrying somatically expanded *mHTT* CAG repeats. The finding that MutSb is elevated in human MSNs [1], coupled with the data we present here demonstrating cell type-specific enhancers in *MSH2*, *MSH3*, *FAN1* and other genes involved in somatic CAG expansion, suggests selective somatic expansion of *mHTT* CAG repeat in specific cell types is driven by transcriptional mechanisms controlling the balance of expression of these critical DNA repair genes in each cell type. We demonstrate also that chromatin accessibility changes in HD brain are cell type specific, and that their magnitude and number differ between cell types where *mHTT* CAG repeat is expanding vs stable. In striatal MSNs, we find that transcriptional enhancers associated with genes that are repressed during HD progression lose accessibility and gain cytosine methylation (5mC) and hydroxymethylation (5hmC) at CpG dinucleotides. In contrast, genes that are de-repressed in HD MSNs gain chromatin accessibility and lose 5mCG, 5mCH and 5hmCH in their gene body. These data support a model of toxicity in HD that is driven by disruption of transcriptional mechanisms by mutant huntingtin polyQ

domains, and stabilization of the disrupted pattern of transcriptional events by altered DNA methylation and hydroxymethylation. The magnitude, scale and cell-specificity of epigenetic changes in expanding HD neurons, and the observation that genes responsible for haploinsufficient neurological or neurodegenerative diseases are strongly repressed in striatal MSNs, suggest that intrinsic mechanisms drive cell type- and circuit-specific toxicity by affecting the majority of HD MSNs.

Results:

Epigenetic profiling and cell type-specific transcriptional control in the human brain.

We previously established that select neuronal populations undergo somatic CAG expansion within exon 1 of the mutant Huntingtin (*mHTT*) gene in the human brain [1,2]. In striatal medium spiny neurons (MSNs), somatic expansion of the *mHTT* CAG repeat correlates with high level of nuclear *MSH2* and *MSH3* transcripts and their encoded proteins [1]. To test the hypothesis that cell type-specific transcriptional regulation of genes involved in somatic expansion contributes to its cell type-specificity, we first identified candidate cis-regulatory elements (cCREs) in neuronal and glial populations exhibiting varying degrees of CAG expansion. The cell types studied include: (1) neurons with significant *mHTT* CAG expansion in Huntington's disease (HD), namely striatal direct-pathway MSNs (dMSNs), indirect-pathway MSNs (iMSNs), and cortical L5a, L6a, and L6b pyramidal neurons from primary motor cortex (BA4); (2) neurons with minimal or moderate CAG expansion, including hippocampal CA1, CA2/3 neurons, ADARB2-expressing interneurons, dentate granule neurons, and cerebellar granule cells; and (3) glial populations (oligodendrocytes, microglia, and astrocytes) from striatum exhibiting minimal expansion (Fig 1a).

Although chromatin accessibility alone can indicate potential cCREs, definitive identification of transcriptional promoters, enhancers, and architectural elements requires comprehensive gene expression data integrated with high-resolution mapping of multiple characteristic epigenetic marks [21]. Given the absence of these datasets for human brain cell types, we used fluorescence-activated nuclei sorting sequencing (FANS-seq) to generate genome-wide profiles of chromatin accessibility (ATAC-seq), histone modifications (CUT&RUN-seq), and DNA methylation and hydroxymethylation at single-nucleotide resolution (oxBS-seq) from brain tissue samples of control and HD donors (Fig 1a). Consistent with previous studies of murine brain [22], UMAP dimension reduction of genome-wide consensus peak sets, defined from our cell type-specific ATAC-seq datasets via iterative peak-calling strategy [23], clearly distinguished neurons from glia and segregated specific neuronal and glial subtypes (Fig 1b). Correlation matrix clustering further validated these distinctions, demonstrating high within-cell-type correlations and, as expected, similarities among closely related neuronal populations (Fig S1b).

Accurate regulatory network model construction requires cell type-specific peak-to-gene annotation. We therefore employed cell type-specific FANS-seq datasets and iterative ATAC-seq consensus peaks to develop a correlation-based model to systematically link accessible regions to target genes (Fig S1a), enabling functional annotation of over 80,000 accessible regions per cell type as either (+)cCREs (predicted transcriptional enhancers) or (-)cCREs (predicted repressors) (Table S1).

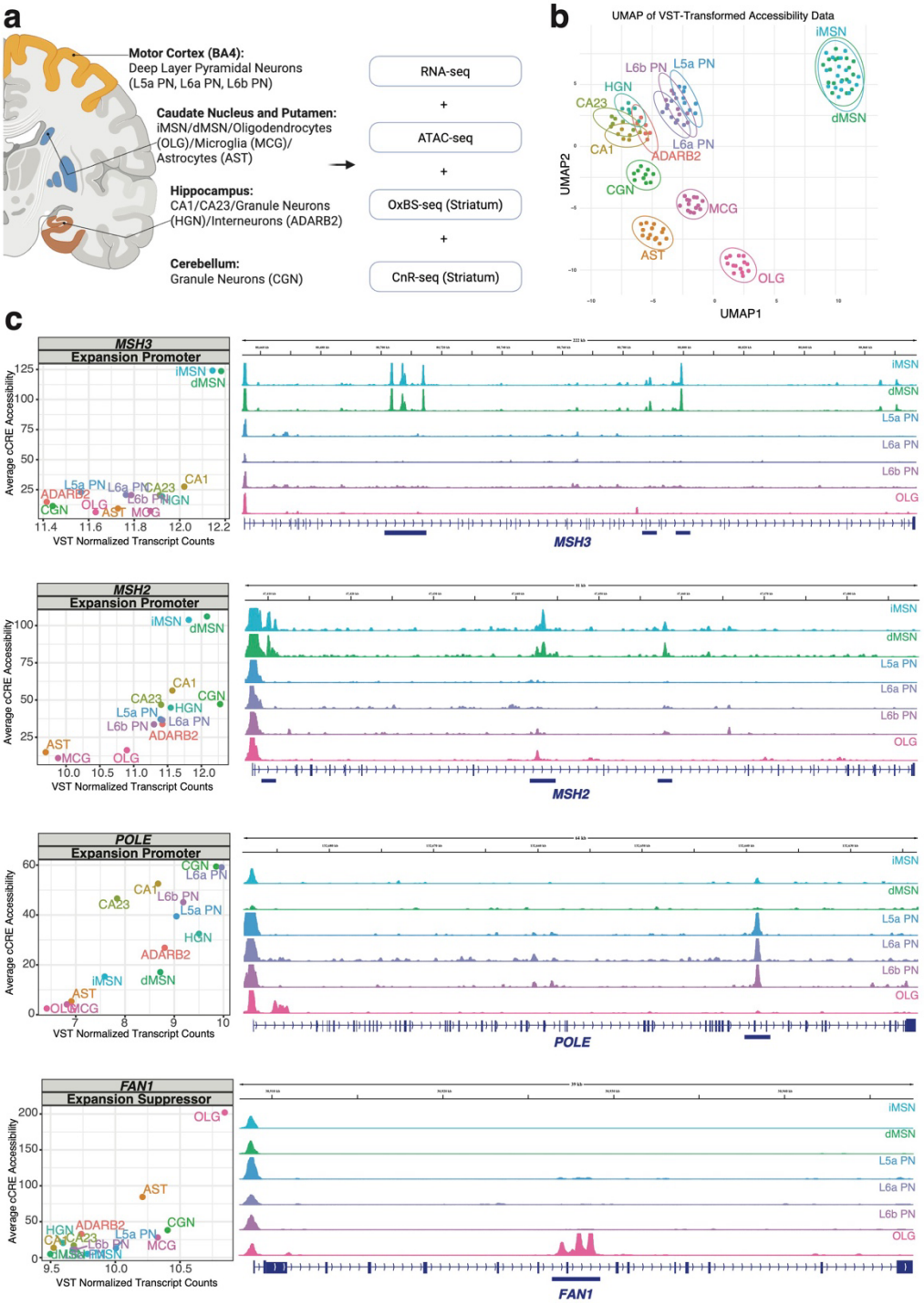


Fig 1 | Cell type–specific enhancer elements in genes modifying CAG repeat expansion

a, Schematic of the experimental workflow. Nuclei from the following cell types were collected for RNA-seq and ATAC-seq profiling: L5a, L6a, and L6b projection neurons (motor cortex, BA4); direct and indirect medium spiny neurons (dMSNs and iMSNs), oligodendrocytes, microglia, and astrocytes (caudate and putamen); CA1 and CA2/3 projection neurons, granule neurons, and ADARB2^{+} interneurons (hippocampus); and granule neurons (cerebellum). In-depth profiling of striatal MSNs and glia included CUT&RUN-seq for histone modifications, as well as OXBS-seq analysis of CpG and CpH methylation and hydroxymethylation. Created with BioRender.com. **b,** UMAP dimensional reduction of top 5000 variable accessible regions across all cell types reveals clustering of consensus ATAC-seq datasets by both cell type and brain region. **c,** Cell type-specific (+)cCREs accessibility in genes modifying of HD onset or progression (*MSH3, FAN1)* or somatic CAG repeat expansion in mouse models *(MSH2, MSH3, POLE, FAN1).* (Left) Normalized ATAC-seq read counts in (+)cCREs were averaged across control donor samples by cell type and plotted against the average VST-transformed read counts of FANS-seq datasets collected for the respective cell types from control donors' samples. (Right) Representative IGV tracks showing selected (+)cCREs (blue bars) linked to *MSH3, MSH2, POLE, and FAN1.*

Genetic modifiers of somatic expansion contain cell type-specific (+)cCREs

Applying our peak-to-gene linkage model, we discovered cell-type specific (+)cCREs for genes identified in genome-wide association studies (GWAS) as potential modifiers of HD onset or progression [6-8], for genes encoding components of the mismatch repair complex or linked to somatic expansion of *mHTT* CAG repeat [12], as well as for *HTT* itself (Fig 1c; Fig S2a; Table S1). For example, prominent cell type specific (+)cCREs are evident in intron 7 of *MSH2* and in *MSH3* between exons 8 and 9 in medium spiny neurons (MSNs), between exons 40 and 41 of *POLE* in deep layer cortical neurons and between exons 7 and 9 of *FAN1* in astrocytes and oligodendrocytes (Fig 1c, S2b, Table S1). Notably, based on our own and previously published chromatin accessibility datasets [22], these (+)cCREs in *MSH3* are not found in MSNs in mouse brain and therefore are species-specific regulatory elements (Fig S2b,c). When comparing striatal oligodendrocytes to MSNs, most of the cell type-specific (+)cCREs coincided with cell type-specific presence of H3K27ac (Fig S3a,b) and demethylation of CpGs (Fig S3a,c), features that identify these (+)cCREs as transcriptional enhancers. In contrast, the promoter regions of *MSH2*, *MSH3* and *FAN1* do not exhibit cell type-specific accessibility. Thus, our data suggest that cell type-specific transcriptional enhancers may be critical for establishing the expression levels of genes determining cell type-specific expansion of the *mHTT* CAG repeat in the human brain.

Altered enhancer accessibility in neuronal cell types undergoing somatic CAG expansion drives pathogenic HD-associated transcriptional responses

Studies in mouse model systems [15,16] and humans [1,2] have established that severe transcriptional alterations are evident in cell types that undergo somatic CAG expansion. We next investigated whether human cell types exhibit chromatin changes that can explain transcriptional dysregulation. Notably, the severity of chromatin dysregulation positively correlated with the mean somatic length gain of the *mHTT* CAG tract (Fig 2a; Extended Data Figure 1a,b), consistent with our findings that the number of HD-associated gene expression changes was greater in cell types with more somatic expansion of the *mHTT* CAG repeat (Extended data Fig 1c) [1]. Cell types with low and high *mHTT* CAG instability differed especially strongly in the number of chromatin regions that become less accessible (Fig 2b). It is noteworthy that in each cell type that showed an HD associated chromatin accessibility signature, these changes are largely cell type-specific (Fig S4a,b). For example, even closely related deep layer neurons from the same cortical region and the two classes of striatal MSNs differ by HD-associated accessibility changes at thousands of regions. In line with this finding, for many genes repression in HD MSNs can be attributed to reduced accessibility of linked (+)cCREs (Fig 2c). Notably, this subset of repressed genes includes those known to cause neurological disease when mutated (Fig 2c).

The widespread gene repression and reduction of enhancer accessibility in cells with somatically expanded *mHTT* CAG repeat point to the possibility that mHTT is disrupting the function of the Mediator complex, as components of the latter have been shown to interact with mHTT [24]. As Mediator complex has previously been shown to be enriched at broadly accessible and H3K27-acetylated “super-enhancer” regions in embryonic stem cells [25], we asked whether human MSN genes with these properties are especially prone to repression in the presence of somatically expanded mHTT. Indeed, genes with super-enhancer properties, high gene-body accessibility and

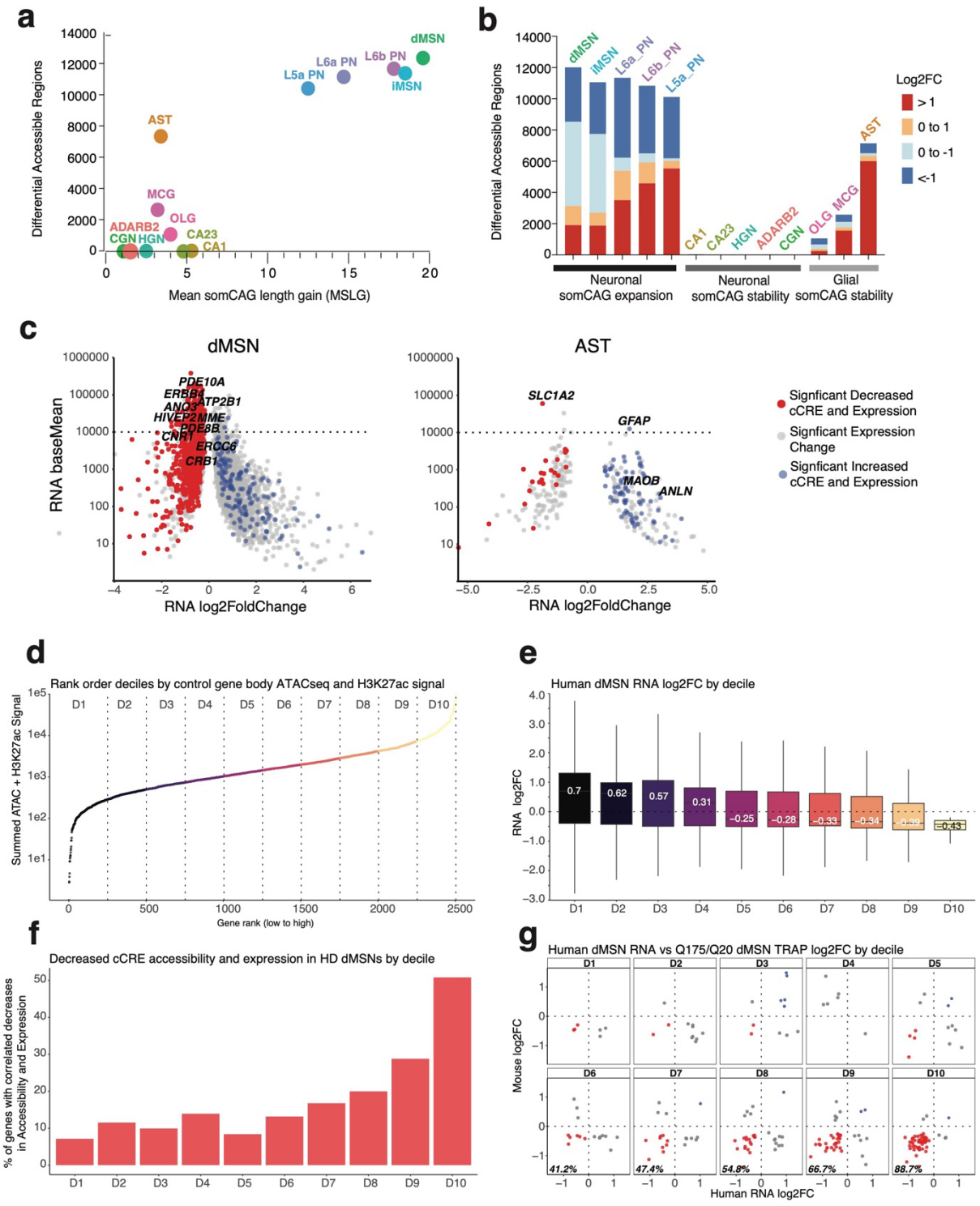
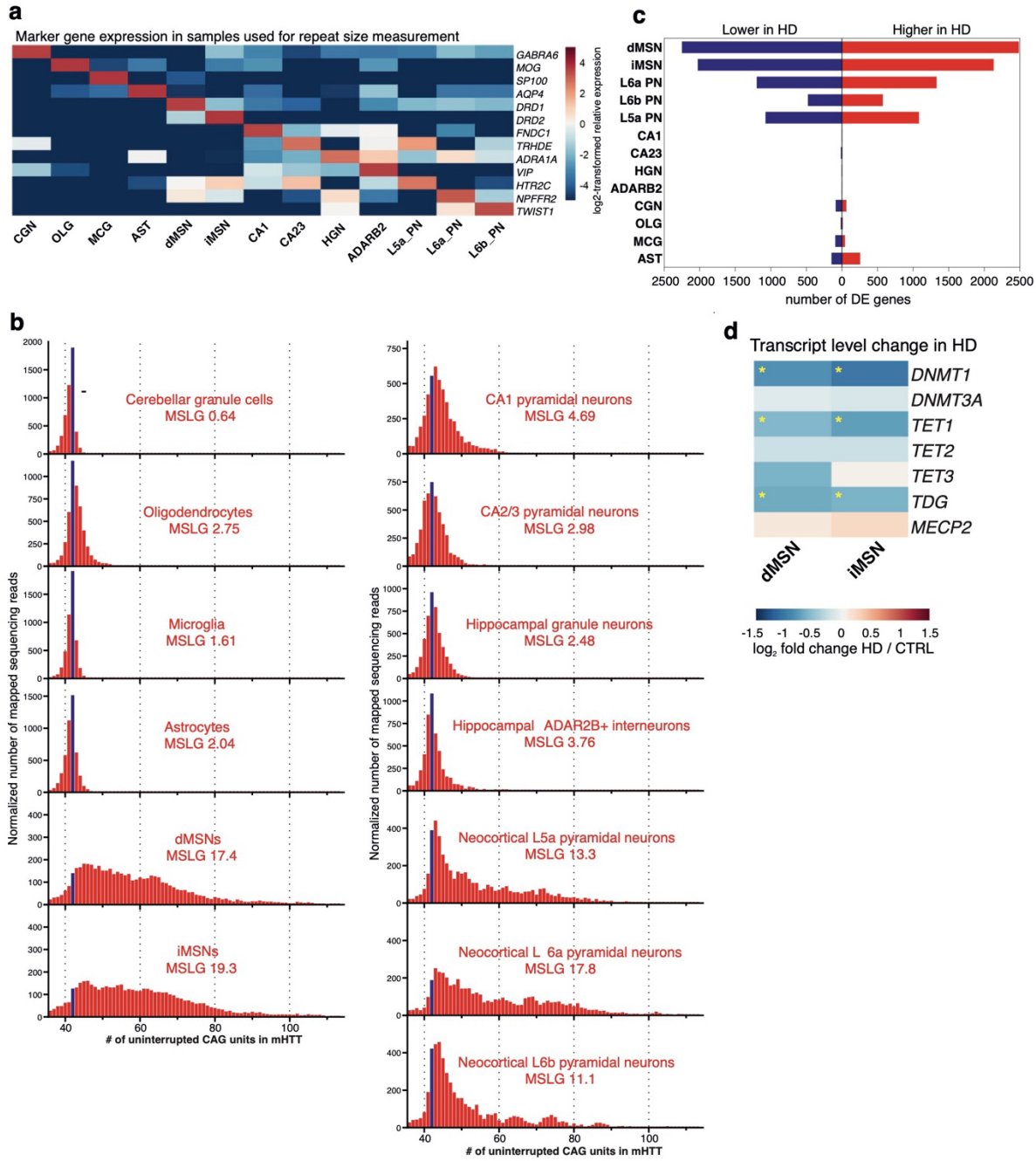


Fig 2 | Somatic CAG expansion correlates with cell type specific reduced chromatin accessibility in vulnerable neurons in Huntington's disease.

a, The number of differentially accessible chromatin regions (*p adj* < 0.05) and somatic *mHTT* CAG instability (Mean Somatic Length Gain) in each cell type analyzed. **b,** The number of differentially accessible chromatin regions (*p adj* < 0.05) in the cell types analyzed. Log2FoldChange < 0 and Log2FoldChange > 0 values indicate reduced and increased chromatin accessibility in samples from individuals with HD, respectively. **c,** Scatter plots of all significant (padj < 0.05) differentially expressed genes in HD dMSNs and astrocytes, marking genes for which repression or

upregulation correlated with reduced (red) or increased (blue) (+)cCRE accessibility, respectively. Genes that are known to cause haploinsufficient human disease (excl. *CNR1*) are labelled for dMSNs and upregulated genes associated with astrogliosis are labelled for astrocytes. **d,** Rank ordering of dMSN genes with HD-associated expression change based on the sum of H3K27ac and accessibility signal, from least to most acetylated and accessible. Dotted lines mark division into bins of equal size (deciles; D1-D10) **e,** Distribution of HD/Control dMSN RNAseq log2FoldChange values in deciles of differentially expressed genes, as defined in panel d. Box-plot elements are defined (e.g. center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers) **f,** Bar plot indicating the fraction of genes, in deciles D1-D10, exhibiting correlated decreases in expression level and (+)cCRE accessibility in HD dMSNs.**g,** Scatter plots display those human dMSN genes for which the expression of mouse orthologue was also changed significantly in dMSNs of zQ175 mice (compared to Q20 mice), based on TRAP data [16]. The extent of overlap of genes downregulated in both human dMSNs and in the zQ175 mouse model is indicated for deciles D6-D10 (as % in quadrant III).



Extended Data Fig 1 | Somatic expansion of *mHTT* and transcriptional dysregulation in the cell types analyzed.
a, Relative expression level of marker genes in the transcriptome of nuclei collected for *mHTT* CAG repeat length analysis. Heatmap depicts log2-transformed relative expression in each sample (calculated based on DESeq2-normalized counts). **b**, Length distribution of the *mHTT* CAG repeat in different cell types in two 69-year-old male donors that carried a tract of 42 uninterrupted CAG units. Data from striatal glia is derived from donor 1, the rest of the cell types from donor 2. In cell types analyzed from both donors, there were no notable inter-donor differences (not shown). Blue bar marks sequencing reads derived from the initial unexpanded CAG tract. y axes denote normalized number of sequencing reads mapped to reference sequences with different CAG tract lengths (normalized by scaling to 5,000 reads). Reads derived from the normal *HTT* allele are not shown. Median somatic length gain (MSLG) is calculated as in Mätlik et al, *Nat Genet.*, 2024 [1]. **c,** Number of differentially expressed (DE) genes (Padj < 0.05 by DESeq2, adjusted for multiple comparisons) in the comparison of HD and control donor FANS-seq datasets from dMSNs and iMSNs (n = 7 and n = 8 HD and control datasets, respectively), neocortical L6a PNs (n = 19, n = 25), L6b (n = 22, n = 19) and L5a PNs (n = 17, n = 25), hippocampal CA1 (n = 6, n = 7), CA2/3 (n = 5, n = 6), granule neurons (n = 6, n = 7) and ADARB2+ INs (n = 6, n = 4), cerebellar GNs (n = 9, n = 9), and striatal oligodendrocytes (n = 7, n = 7), microglia (n = 7, n = 7) and astrocytes (n = 7, n = 7). **d,** Heatmap depicting disease-associated changes in transcript levels of selected genes regulating DNA methylation. Statistically significant differences are marked with an asterisk (Padj < 0.05 by DESeq2, adjusted for multiple comparisons). H3K27ac levels, were far more likely to be repressed than upregulated in HD MSNs (Fig 2d,e), and, for a large fraction of these genes, repression correlated with a decrease in (+)cCRE

accessibility (Fig 2f). Interestingly, while our previous work showed only modest concordance between overall gene expression changes in MSNs in HD mouse models and human MSN FANS-seq datasets [1], the repression of the most highly accessible and H3K27-acetylated genes *(D10)* is recapitulated in dMSNs of the zQ175 mouse model of HD (Fig 2g, Supp. Table S4) [16], and several of these have been identified previously as "super-enhancer" regulated genes that are repressed in HD mouse models [17].

The epigenetic impact of HD is minimal in most neuronal and glial cell types where the *mHTT* CAG repeat is relatively stable (oligodendrocytes, microglia and the hippocampal and cerebellar neuron types we have analyzed here) (Fig 2a). However, an obvious exception is astrocytes in the striatum which do not experience significant somatic CAG expansion (Extended data Fig 1b). In this case, the data reveal a strong HD-associated signature of increased chromatin accessibility that correlates with increased expression of known markers of astrogliosis (*GFAP, MAOB, ANLN*) (Fig 2c) [26] and reflects predominantly increased transcription (Extended data Fig 1c). It should be noted that while astrocytic responses in HD mouse models are very well characterized [27], it is difficult to compare the human and mouse model system data because a pre-expanded *mHTT* allele is present in all cell types, including astrocytes, in the mouse models.

Taken together, the fact that chromatin accessibility is more dramatically affected, and accessibility loss more prevalent, in cell types that carry longer *mHTT* CAG repeats, and that the majority of these changes are cell type specific, provides direct evidence that mHTT causes epigenetic changes that result in pervasive transcriptional repression in cell types where the CAG repeat is expanding versus those where extensive expansion does not occur.

Increased 5mCpG and 5hmCpG at methylation-sensitive enhancers stabilize chromatin accessibility loss

We have demonstrated here that in each cell type carrying a somatically expanded CAG repeat allele of *HTT*, transcriptional repression is accompanied by accessibility decreases in specific chromatin regions predicted to be transcriptional enhancers. In both peripheral cell types [28] and neuronal progenitors [29,30], a subset of enhancers requires active DNA demethylation to remain active. This is because DNA methylation and hydroxymethylation inhibit binding of most TFs to their cognate binding sites [31] thereby reducing the process of enhanceosome assembly [32] that can maintain enhancer function and accessibility. For each region with an HD-associated accessibility change, we computed the combined methylation difference (meth.diff) for significantly altered CpGs (q*value* < 0.05) in MSNs and oligodendrocytes. This reflects the HD-associated change in the modification level of individual cytosines in a CpG context. Broadly, the number and magnitude of HD-associated accessibility changes and cytosine modifications within these regions are larger in MSNs compared to oligodendrocytes, particularly for regions that decrease in accessibility and gain DNA methylation (Fig 3a).

In MSNs more than 9,500 CpGs in accessible domains exhibit a positive change in cytosine modification level, compared to less than 1,500 exhibiting a negative change in modification of cytosines (Fig 3a). This equates to ~46x and 14.7x higher likelihood of decreases in accessibility being associated with higher CpG modification levels in enhancers in iMSNs and dMSNs, respectively (Fig 3b). The magnitude of HD-associated loss or gain of accessibility also correlates with the magnitude of accumulation or loss of CpG modifications in HD MSNs (Fig 3a).

Furthermore, the majority (~ 6,000 peaks, ~70%) of regions that decrease in accessibility are H3K27ac positive enhancers whereas there are only ~1,300 that increase in accessibility and are H3K27ac-positive (Fig 3c, Fig S4c). In contrast to oligodendrocytes, MSNs also display higher average levels of H3K27ac in regions that become less accessible in HD (Fig 3c). Moreover, combined methylation difference values >50% indicate strong regulation of these enhancer elements in most MSNs (Fig 3a). For example, the *PDE10A*, *PDE8B* and *CNR1* genes, which are severely repressed in HD MSNs [1], have enhancers that lose accessibility and contain CpG dinucleotides carrying HD-associated modification in >50 % of MSNs, and for some enhancers this number can exceed 80% of MSNs (Fig 3a,d). Although any disruption to enhancer demethylation equilibrium may stabilize de novo methylation and lead to repression, MSN-specific HD-associated downregulation of *TET1* and *TDG,* which encode enzymes that carry out different steps of cytosine demethylation pathway [33,34], could further exacerbate this effect (Extended Data Fig 1d).

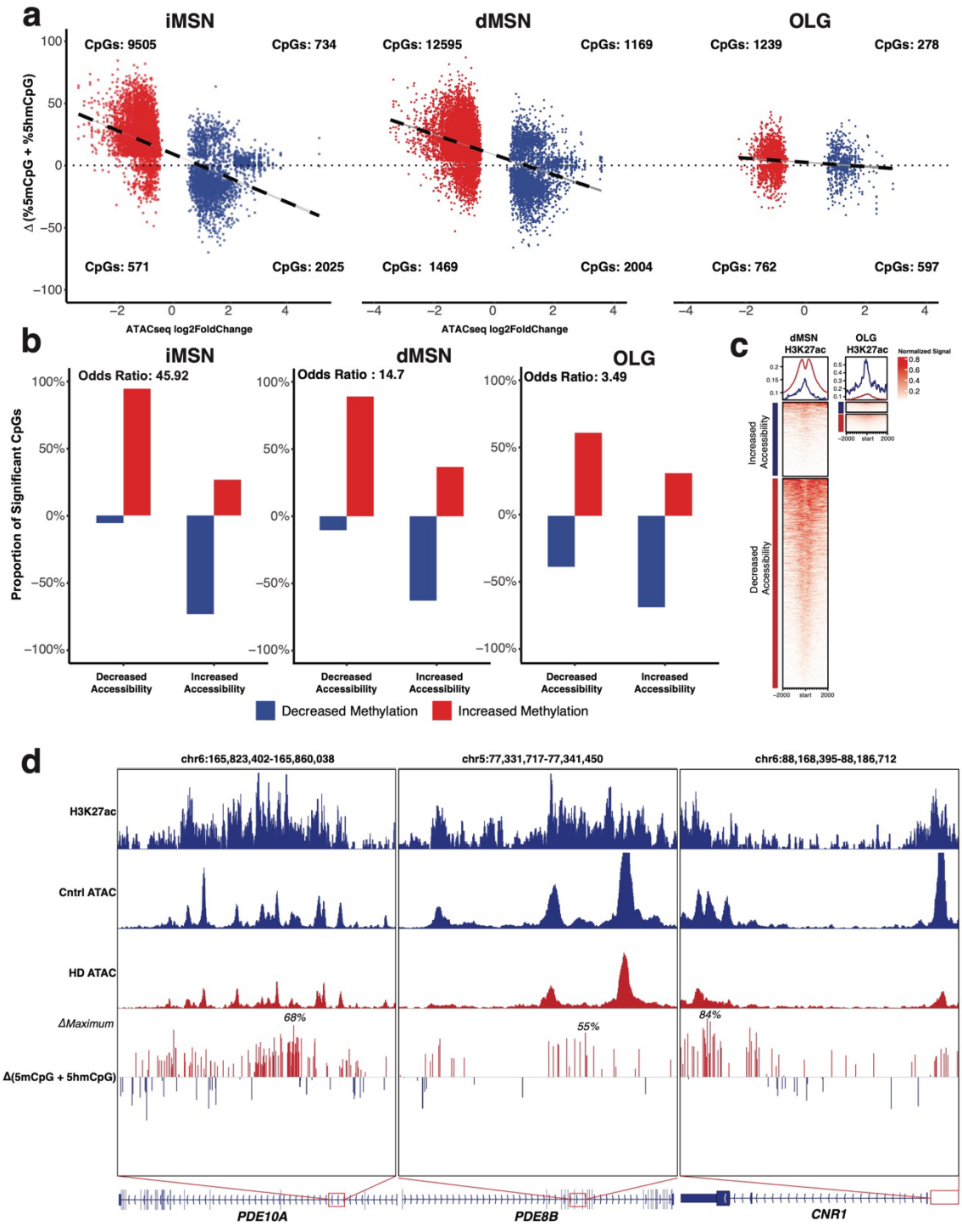


Fig 3 | HD-associated chromatin accessibility loss in MSNs occurs at methylation sensitive enhancer elements

a, HD-associated accessibility change (ATAC-seq log2FoldChange of HD/Control) in dMSNs, iMSNs and oligodendrocytes was plotted against the combined methylation change Δ(%5mCpG+%5hmCpG) at individual CpG dinucleotides overlapping with these differentially accessible regions. The number of significant CpGs in each quadrant are displayed (n=5 HD and n=5 control donors).**b,** Proportion of significant CpG's that have a positive combined change in (hydroxy)methylation or negative combined change in (hydroxy)methylation, in regions that increase or decrease in accessibility in HD iMSNs, dMSNs and Oligodendrocytes. The odd's ratio for the probability of (hydroxy)methylation increase over (hydroxy)methylation decrease is shown for regions with HD-associated accessibility decrease. **c,** Control MSN or Oligodendrocyte H3K27ac signal over regions with HD-associated accessibility change in dMSNs or Oligodendrocytes, respectively **d,** Representative IGV tracks showing decreased accessibility over H3K27ac-positive enhancers in *PDE10A*, *PDE8B*, and *CNR1*, with differential track displaying the Δ(%5mCpG+%5hmCpG) at individual CpG dinucleotides. The magnitude of modification change for most highly altered CpGs is shown.

The regulation of gene expression is disturbed in most HD MSNs

The epigenetic profiles presented here represent chromatin accessibility and DNA methylation measurements in thousands of individual nuclei from dMSN and iMSNs in each donor. Thus, we can infer the proportion of cells in which transcriptional control is disrupted in HD. For example, in Figure 3, marked declines in enhancer accessibility coupled with > 50 % increases in CpG methylation demonstrate that these regulatory elements are altered in at least half of both MSN subtypes. To illustrate further gene-level correlates of this dysregulation, we examined *PDE10A* and *PCP4* (Fig 4a,b), whose significant HD-associated reductions in transcript abundance and chromatin accessibility indicate repression in the majority of MSNs (Fig 4c,d,e,f). This idea is strongly supported by analysis of DNA methylation, showing that in these genes a large fraction of CpGs that are normally demethylated in MSNs gain methylation in HD (Fig 4a,b), and there are cytosines that become modified on more than 50% of the DNA molecules analyzed (Fig S5). As each nucleus can contribute only two copies of a gene to DNA methylation analysis, the data show that biallelic co-occurring (i.e. on the same molecule) modification of these CpGs takes place in a minimum of 50% of MSNs in HD. Alternatively, mono-allelic stochastic loss of demethylated status at these cytosines could occur in the great majority of HD MSNs.

To determine if the documented epigenetic and transcriptional changes take place only in a subset MSNs in HD, we used RNAScope to measure changes in mRNA abundance for *PDE10A* and *PCP4* in individual MSNs, relative to the abundance of an MSN-specific mRNA whose expression does not change (*COCH*) [1]. In sections from control donor samples, *PDE10A* (Fig 4g) and *PCP4* (Fig 4h) mRNAs were detectable in the majority of *COCH*-positive MSNs in each control donor analyzed. As expected, although the number of MSN nuclei were substantially reduced in HD donor samples, *COCH* mRNA was detected at comparable levels, whereas little or no *PDE10A* and *PCP4* mRNA was detected. Quantitative image analysis showed that *PDE10A* and *PCP4* transcript levels were reduced in the vast majority of HD MSNs (Fig. 4i, j): only a small subset of HD MSNs for PCP4, and virtually none for PDE10A, exhibited PCP4/COCH or PDE10A/COCH signal ratios comparable to those in control MSNs. Together, FANS-seq, ATAC-seq, OxBS-seq, and mRNA in situ hybridization demonstrate that most striatal MSNs in HD patients undergo marked transcriptional dysregulation.

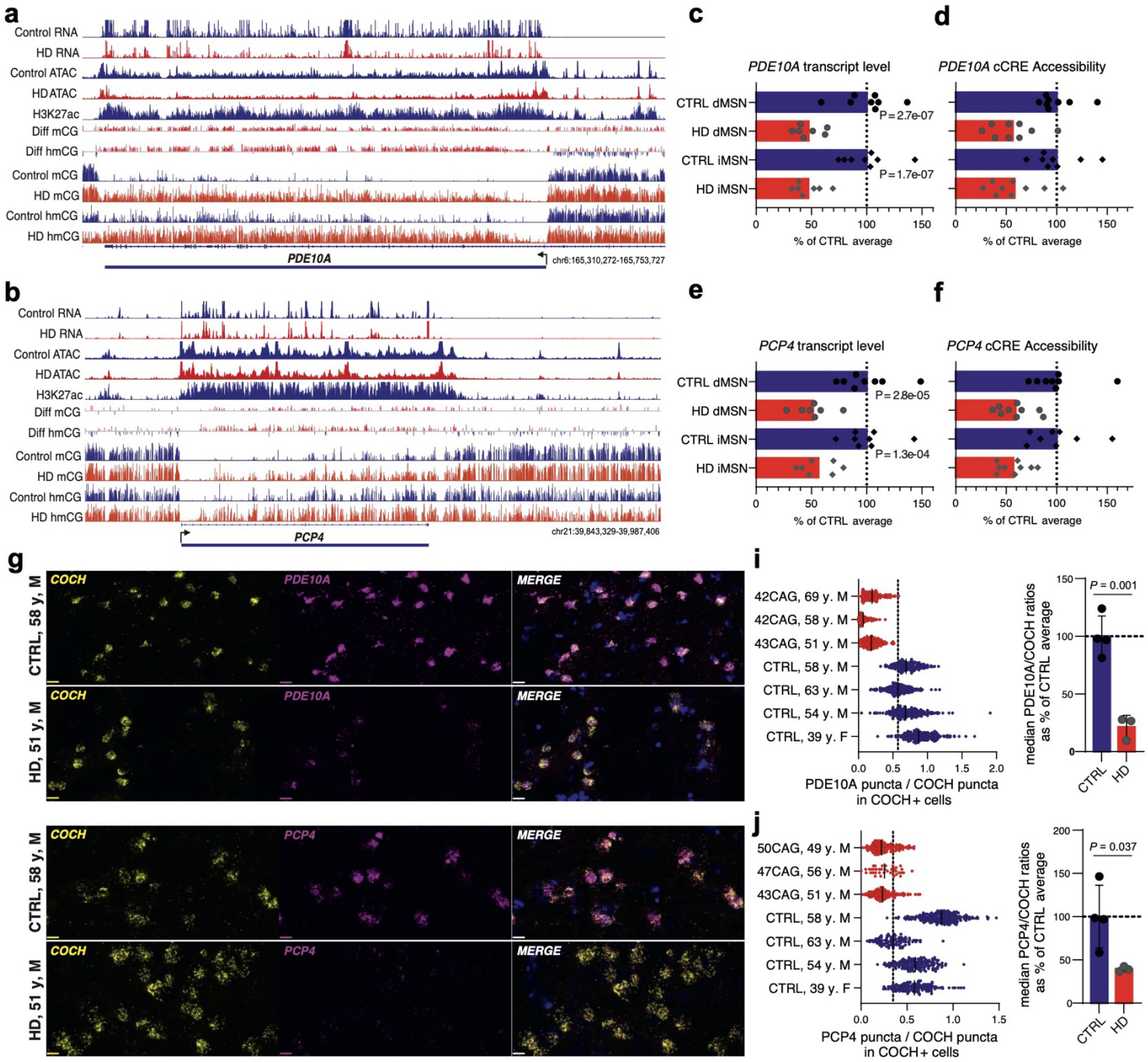


Fig 4 | Gene expression is disrupted in the majority of MSNs in HD

a,b, Distribution of FANS-seq and ATAC-seq reads, and the position of modified cytosines in dMSN datasets from a control (blue) and HD donor (red), showing HD-associated changes in nuclear transcript level, chromatin accessibility, cytosine methylation and hydroxymethylation across *PDE10A* and *PCP4*. *Differential* tracks show the position of CpG dinucleotides with statistically significant HD-associated (hydroxy)methylation gain (red) or loss (blue). **c,d,e,f,** *PDE10A* and *PCP4* transcript level (c, e) and bCRE accessibility (d, f) in HD MSNs represented as % of their transcript level and accessibility in control donor MSNs (mean plotted), based on (c, e) FANS-seq reads (n = 7 HD, n = 8 CTRL donors) mapped to *PDE10A* and *PCP4*, and (d, f) ATAC-seq reads (n = 9 HD, n = 8 CTRL donors) mapped to differentially accessible cCREs linked to *PDE10A* (average of 29 cCREs) and *PCP4* (average of 51 cCREs). P values were calculated with DESeq2 and are adjusted for multiple comparisons. P values for each of the cCREs can be found in Table S3. **g,h,** Representative sections of putamen stained with RNAScope for the quantification of *COCH* (yellow) and (d) *PDE10A* or (e) *PCP4* (magenta) transcripts (cell nuclei stained with DAPI, blue). **i,j,** Ratio of PDE10A+ puncta to COCH+ puncta (f) and PCP4+ puncta to COCH+ puncta (g) was calculated for each COCH+ cell (median of the ratio is plotted for cells from each tissue donor analyzed). Vertical dashed line indicates the lowest median PDE10A/COCH ratio and PCP4/COCH ratio out of the four control donors analyzed. The average of median PDE10A/COCH and PCP4/COCH ratios in HD donors was significantly lower than in control donors (P = 0.001 and P = 0.0368 for *PDE10A* and *PCP4*, respectively, using unpaired two-tailed T-test, average +/- SD is shown).

Modeling MeCP2–Dependent gene derepression predicts up-regulation in HD MSNs

Despite cytosine methylation at a subset of enhancers predicting decreased chromatin accessibility and gene repression in HD MSNs, the mechanistic basis for why certain genes are up-regulated in HD remains unresolved. To address this, we first examined how HD-associated loss of DNA methylation modifications within gene bodies correlates with increased expression of specific genes in HD dMSNs. Using gene expression and cytosine methylation data from dMSNs, we constructed a logistic regression model where the predictor was the proportion of significantly differential cytosines in gene bodies that lost methylation in HD (Fig. 5b, left). Individually, decreases in 5mCpG, 5mCpH, and hmCpH each showed a monotonic increase in the probability of gene up-regulation, while 5hmCpG decrease showed only a poor correlation with the probability of upregulation (Fig. 5b, left panel). Since 5mCpG, 5mCpH, and 5hmCpH are known to be high affinity binding sites of transcriptional repressor methyl-CpG binding protein 2 (MeCP2) (Fig 5a), we then combined all three contexts into a single predictor, the fraction of MeCP2 sites lost per gene, to model increased expression for genes with methylation changes biased towards loss of these sites (Figure 5b right panel, dashed line). We then applied the model to methylation data from iMSNs, to see if it could accurately predict de-repression of genes in HD iMSNs based on this independent dataset [35,36]. For iMSN genes where the proportion of decreased MeCP2 binding sites exceeded 0.5, the model predicted HD-associated increased expression with >80% accuracy (Fig. 5b, right panel, blue line), confirming that extensive loss of high-affinity MeCP2 binding sites predicts transcriptional de-repression in HD MSNs. In contrast, a high proportion of increased MeCP2 binding sites (i.e. proportion of decreased MeCP2 binding sites <0.5) was a poor predictor of lower expression in HD iMSNs, consistent with our prior finding that enhancer inactivation, rather than gene-body methylation, drives the repression of gene expression in HD.

To visualize per-gene methylation dynamics in HD MSNs, we plotted the mean and distribution of single-Cp differentials (HD/Control) for each context in four representative de-repressed genes (*CADM1, GABRA5, GDA, ABLIM1*; Fig. 5c, d), illustrating how loci with the greatest loss of MeCP2 sites display concomitant increases in transcription and chromatin accessibility. These results demonstrate that loss of high-affinity MeCP2 binding sites within gene bodies is a robust predictor of increased gene expression in HD MSNs, thus providing a quantitative framework to distinguish derepressed targets from those repressed via enhancer loss.

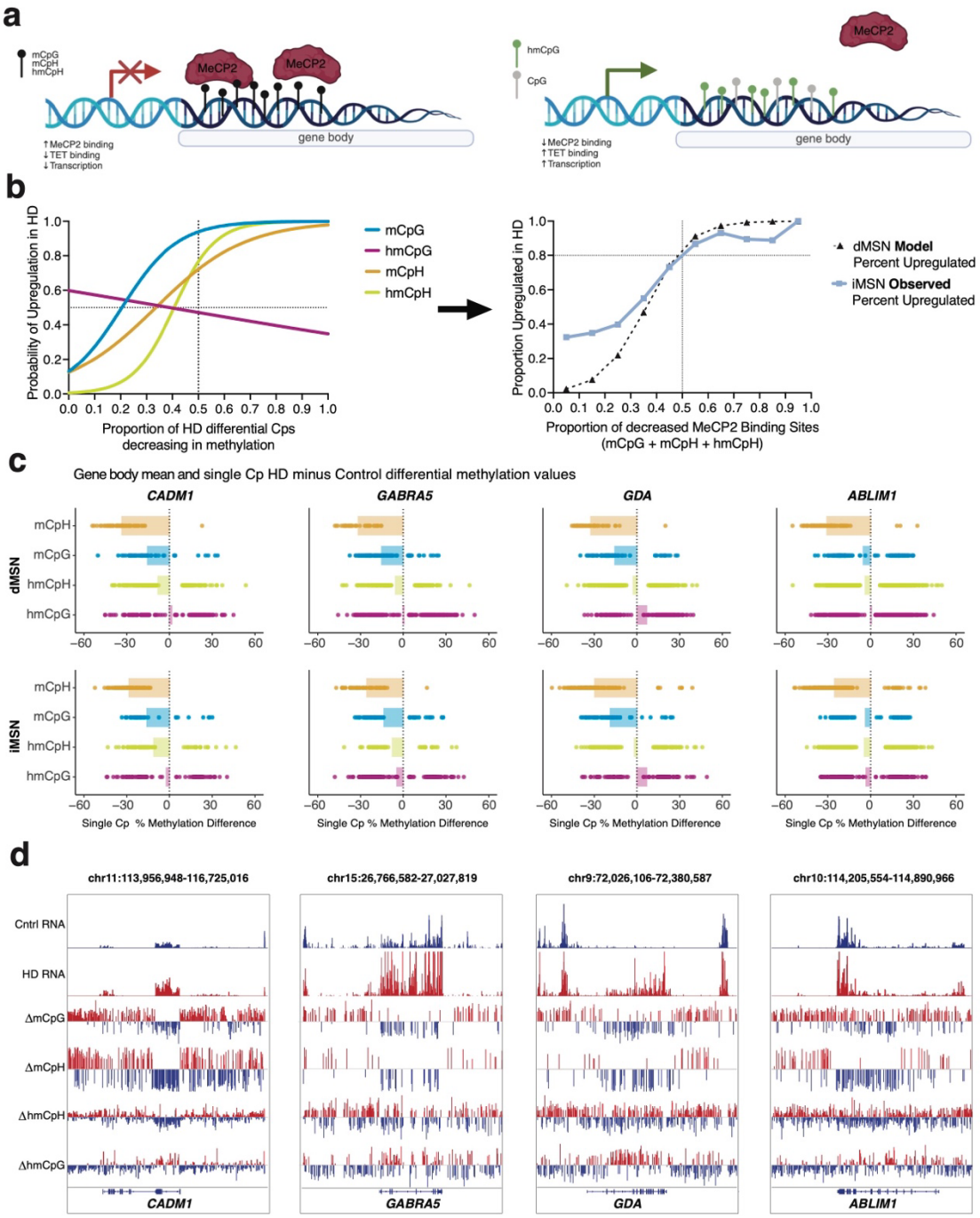


Fig 5 | Loss of preferred MeCP2 binding sites in gene bodies stabilize increased expression in HD.
a, Graphical model indicating the relationship between cytosine modifications 5mCpH, 5hmCpH, 5mCpG and 5hmCpG, their ability to act as binding sites for the neuronal transcriptional repressor MeCP2, and their impact on gene expression. **b(left),** Logistic model predicting the fraction of genes upregulated in HD dMSNs as a consequence of the proportion of gene body cytosines with decreased levels of either 5mCpG, 5hmCpG, 5mCpH or 5hmCpH. **b(right),** Aggregated logistic model, based on dMSN data, showing the relationship between the reduction of high-affinity MeCP2 binding sites (gene body 5mCpG, 5mCpH, 5hmCpH combined) and gene up-regulation (dashed line). The model was used to predict the direction of HD-associated gene expression change based on HD-associated change in the level of 5mCpG, 5mCpH, 5hmCpH modifications in iMSNs (blue line). Dotted vertical line indicates the threshold of >50% of high affinity MeCP2 sites losing methylation in HD, and dotted horizontal line marks the threshold of 80% of genes predicted to be upregulated. **c,** Distribution of HD-associated modification change values for individual gene body cytosines (Cp) in dMSN and iMSN (mean change in methylation indicated with horizontal bar). **d,** Distribution of FANS-seq and ATAC-seq reads, and the position of modified cytosines in dMSN datasets from control (blue) and HD donors (red), showing HD-associated changes in nuclear transcript level, chromatin accessibility, cytosine methylation and hydroxymethylation across gene bodies. *Differential* tracks (Δ) show the position of CpG dinucleotides with statistically significant HD-associated (hydroxy)methylation gain (red) or loss (blue).

Discussion

We present here comprehensive epigenetic profiling, genome-wide DNA methylation and hydroxymethylation mapping of specific neuronal and glial cell types from the striatum, cerebral cortex and cerebellum of control and HD donor samples, as well RNAScope validation studies, to discover mechanisms responsible for the abundant transcriptional disturbances we have previously reported in human brain cell types in Huntington's disease [1,2]. In contrast to single nucleus-based analyses of specific cell types, FANS-purified populations of nuclei were analyzed because it enables the generation of genome-wide chromatin accessibility profiles and single nucleotide resolution DNA (hydroxy)methylation profiles to uncover correlated cell type-specific disease-associated epigenetic events at promoters, enhancers, and gene bodies that are known to regulate gene expression. These data establish that the cell types we have analyzed in the human striatum, cerebral cortex and cerebellum have tens of thousands of accessible chromatin regions that are positively correlated with gene expression (Supp. Table 1). As expected [37], UMAP clustering of the ATAC-seq data reveals that developmentally and functionally related cell types have more similar genome-wide chromatin accessibility profiles whereas cell types from different lineages display largely distinctive accessibility and DNA methylation signatures (Fig 1b, S1c,e,f). The depth and quality of these data have allowed us to discover several important epigenetic mechanisms associated with both phases of HD pathophysiology.

Transcriptional control of cell type-specific somatic CAG expansion in the human brain.

In the first stage of HD pathogenesis [10], somatic expansion of the *mHTT* exon 1 CAG repeat occurs in select neuronal cell types in the striatum [1] and cerebral cortex [2]. We show here that the expression of DNA repair and handling genes that modify HD onset or progression in humans, or modulate somatic CAG expansion in model systems [38-40], correlates with their associated enhancer(s) accessibility in specific human CNS cell types (Fig 1c; S2a,b). As the balance of different mismatch repair proteins in each cell type must determine their propensity for somatic CAG expansion, our data suggest that enhancer-mediated transcriptional control of MMR proteins may plays a major role in the cell type-specificity of somatic CAG expansion. For example, the elevated expression and chromatin accessibility of the *MSH2* and *MSH3* genes in human MSNs (Fig 1c), coupled with the presence of cell type- and species-specific transcriptional enhancers (Fig 1c; S2b,c), suggest that enhancer-mediated transcriptional mechanisms contribute to the high levels of MutSβ and CAG repeat instability in human MSN nuclei [1]. Our data indicate also that suppression of expansion in oligodendrocytes and astrocytes may result from high levels of *FAN1* expression, dictated by a cell type-specific enhancer cluster in this important gene (Fig 1c; S2b). Although data from human postmortem studies is not sufficient to test these hypotheses, studies of accelerated evolution of human specific CREs have shown that the functions of many predicted human-specific enhancer elements are expressed accurately when the elements are inserted into the mouse genome [41].

Compromised function at DNA demethylation-dependent enhancers drives cell type-specific transcriptional repression in HD.

The data we present here demonstrate that HD-associated disruptions in chromatin accessibility depend on both cell autonomous responses to expanded mHTT and regional degeneration. In cell types undergoing somatic CAG expansion, iMSNs, dMSNs, and L5a, L6a, and L6b pyramidal

neurons, chromatin accessibility is predominantly repressed. In contrast, oligodendrocytes, astrocytes and microglia in the striatum exhibit fewer changes and primarily increased accessibility. In hippocampal CA1 and CA2/3 pyramidal cells, interneurons, dentate granule cells, and cerebellar granule neurons, where there is neither somatic CAG expansion (Extended Data Fig 1b) nor evidence of local degeneration [42], chromatin accessibility changes are minimal or absent. Consistent with transcript-level alterations, ~70% of regions losing accessibility in MSNs overlap H3K27ac-marked enhancers. Because ATAC-seq provides a single accessibility measurement per chromosome, these cell-type-specific declines in enhancer accessibility demonstrate that dysregulation of enhancer function in response to CAG repeat expansion occurs in a large fraction of iMSNs, dMSNs, and deep-layer pyramidal cells.

Our finding that HD-associated reduced enhancer accessibility coincides with increased DNA methylation and hydroxymethylation in both iMSNs and dMSNs is consistent with the requirement of DNA demethylation for enhancer activity, as shown for peripheral cell types [28] and neural progenitors [29,30]. Since these cytosine modifications inhibit the binding of most transcription factors [31], loss of demethylation at these regions is expected to contribute to formation of inaccessible chromatin, enhancer inactivation and repression of transcription. Thus, progressive accumulation of both 5mC and 5hmC, substrates of demethylating TET enzymes, at a subset of enhancers in HD MSNs is a manifestation of altered DNA methylation/demethylation equilibrium caused by the convergence of two pathogenic events: the reduced occupancy of enhancers due to compromised Mediator complex formation [24], and a reduction in *TET1* and *TDG* expression (Extended Data Fig 1d) leading to impaired DNA demethylation efficiency and stochastic accumulation of 5mC and 5hmC. In support of this proposed mechanism, the MED15 gene, which encodes an important component of the Mediator complex, has been identified as an HD modifying gene [8]. Furthermore, HD-associated increases in methylation and subsequent repression in MSNs is especially prominent in highly accessible and demethylated H3K27ac-positive 'super-enhancer' regions (Fig 2d-f) which have been shown to be the regions heavily enriched for mediator complex occupancy in embryonic stem cells [25].

Disturbed DNA methylation maintenance in gene bodies is responsible for loss of MeCP2 biding sites and de-repression of genes in HD MSNs.

Analysis of gene expression in neocortical and striatal cell types that undergo CAG repeat expansion has also demonstrated that hundreds of genes are de-repressed during HD progression [1,2]. Analysis of chromatin accessibility and DNA methylation demonstrate that these genes display distinct HD-associated epigenetic changes. In contrast to repressed genes that accumulate repressive cytosine methylation, a class of genes whose expression increases in HD MSNs primarily demethylate cytosines within their respective gene body. In this case, the DNA methylation context of these genes is highly informative (Fig 5). These genes exhibit decreased levels of 5mCpG, 5mCpH and 5hmCpH, which are all high affinity binding sites for MECP2, and gain 5hmCG, which is not a high affinity binding site for MECP2 [35,36]. The net result of the described HD-associated changes in DNA methylation, therefore, is loss of high affinity binding sites for this critical, neuron-specific transcriptional repressor. Notably, recent computational studies have revealed that DNA methylation is "the strongest epigenomic correlate of local elongation rate" [43], pointing to the possibility that genetic variation in *TCERG1* could be modifying HD onset through its effect on de-repression of this specific class of genes (GeM-HD Consortium 2015, GeM-HD Consortium 2019, Genetic Modifiers of Huntington's Disease et al 2024).

Transcription-mediated toxicity in Huntington's disease.

Based on human genetic studies, single nucleotide polymorphisms (SNPs) in a variety of individual genes can have significant impacts on age at onset of HD and other disorders [6-8,44]. Furthermore, the finding that expression of the *PDE10A*, *PDE8B*, *ANO3*, *FOXG1*, *ATP2B1* and *HIVEP2* genes are all strongly repressed in HD [1] and the fact that each can by itself cause human CNS disorders that include striatal degeneration [45-47], dyskinesia [45,48,49], dystonia [50], and intellectual disability [51,52] suggests that repressed expression of these genes together in the majority of human MSNs compromises their function. Moreover, significant reduction in expression of several genes in the cAMP signal transduction pathway (e.g. *ADCY1*, *PPP1R12A*, *PPP1R2*, Supp. Table 3) in MSNs suggests that their combined actions would severely compromise cAMP signaling. Given the central role of cAMP in both direct and indirect pathway MSNs [53], and severe deficiencies in dopamine-dependent cAMP signaling in presymptomatic HD mice [54], we believe these and other changes we have documented cause major functional deficiencies in the human basal ganglia circuitry.

In mouse models of Rett syndrome, an X-linked disorder in which the transcriptional repressor methyl CpG binding protein 2 (MeCP2) is depleted in only 50% of cells in the brain, widespread transcriptional dysregulation in the absence of cell loss causes severe behavioral phenotypes [55,56]. These phenotypes can be rescued in adult animals by reversing the mutation to restore wild-type transcription. Furthermore, deletion of MECP2 in adult mice has demonstrated that transcriptional and chromatin changes precede overt phenotypic changes, contributing to reduced neuronal function in the absence of developmental effects [57,58]. In SCA1 mouse models, transcriptional dysregulation in several cell types as well as altered electrophysiological and behavioral phenotypes occur well before widespread degeneration of cerebellar Purkinje cells [59]. These data further support the general conclusion that disrupted gene expression is sufficient to alter cellular function and behavior in the absence of cell loss.

A mechanistic model of HD pathogenesis

The epigenetic data we have presented here provides mechanistic insight into the two stages of HD pathogenesis [10]. In the first stage, cell type specific somatic expansion of the *mHTT* CAG repeat occurs in select cell types in several brain regions. Our data contribute to understanding the selectivity of somatic CAG repeat expansion in striatal and neocortical neurons in the human brain [1,2] by identifying cell type-specific transcriptional enhancers in genes known to promote somatic CAG expansion (*MSH2, MSH3, POLD2, POLD4*) [12] (Fig 6). In contrast, astrocytes and oligodendrocytes, in which very limited expansion occurs, have strong enhancer signatures for *FAN1*, a strong suppressor of somatic expansion. In response to somatic expansion, a second, toxic phase of HD pathogenesis then ensues as a result of somatic expansion of the *mHTT* CAG repeat. The epigenetic and transcriptional profiling data we have presented demonstrate clearly that HD-associated disruption in the regulation of genes known to cause haploinsufficient human disease occurs in cell types in the human brain that have undergone somatic CAG expansion. Thus, our data support a mechanistic model of HD pathogenesis in which the toxic effect of mutant huntingtin on MED15-mediated enhancer function and TCERG1-dependent transcription elongation is stabilized by altered local DNA-methylation/demethylation dynamics, and neuronal dysfunction arises due to the consequent impact on cellular biochemistry and connectivity (Fig 6). These mechanisms are responsible for disrupting the expression of hundreds of genes in the

majority of striatal MSNs, as demonstrated by the large-magnitude of HD-associated changes we have documented. Since >90% of human MSNs in adult-onset HD carry *mHTT* alleles with *less* than 150 CAG repeats [5], our data prove that severely dysregulated gene expression occurs in neurons at a wide range *mHTT* CAG repeat lengths.

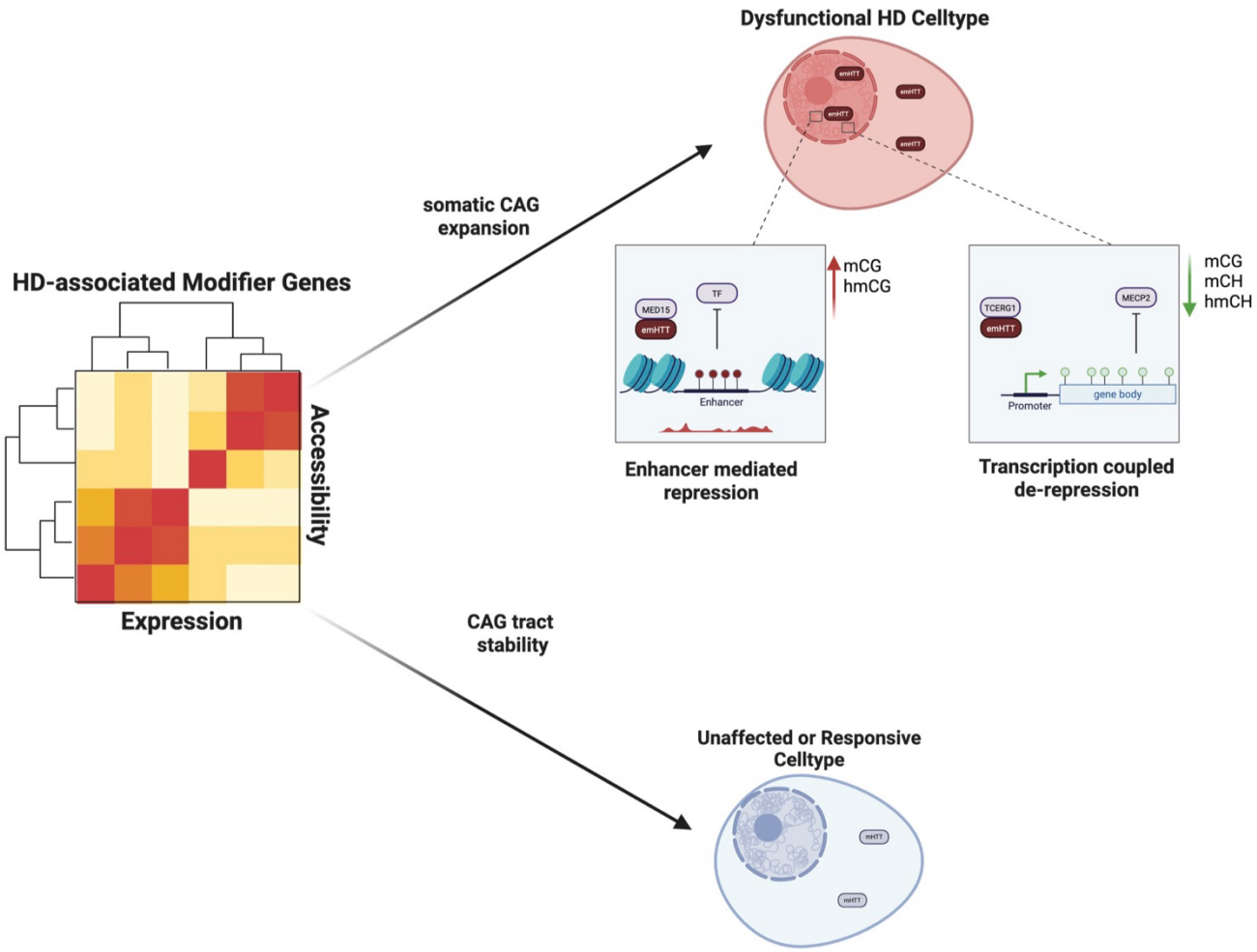


Fig 6 | Model of epigenetic states that govern the two stages of cell type specific pathology in Huntington's Disease
Stage 1: Cell type-specific accessibility establishes expression of genes affecting the stability of *mHTT* exon 1 CAG repeat.
Stage 2: Cell that undergo somatic expansion of the *mHTT* CAG repeat accumulate expanded mHTT protein, leading to two distinct mechanisms that drive disease progression:

1. The interaction of mHTT protein with MED15 [24] perturbs enhancer function, leading to DNA methylation and repression of genes, especially those dependent on 'super-enhancers'
2. Interaction with mHTT reduces the inhibitory effect TCERG1 has on transcriptional elongation [24,80,81], leading to transcription-coupled gene body DNA demethylation. This can inhibit MeCP2 binding and repressor complex formation, resulting in aberrant upregulation of genes with low expression.

Therapeutic implications of the epigenetic contribution to the two-stages of HD pathogenesis

So far, studies of genetic HD-modifiers have focused primarily on how these genes affect CAG repeat stability and as well as their patterns of expression across cell types and tissues [1,2,9,12,60]. Despite these critical insights, the ubiquitous and necessary expression of these repair genes makes systemic inhibition a possible barrier to therapeutic implementation [61]. Our data expand on this by showing that the expression patterns of these genes are established via cell type specific enhancer-mediated mechanisms in the human brain. As such, further studies of how enhancers control MMR genes, either by promoting or inhibiting somatic CAG expansion, may reveal strategies to stabilize the *mHTT* CAG repeat in HD in a highly cell-specific manner, thereby avoiding possible unwanted effects as a consequence of systemic inhibition of DNA repair.

In patients with manifest HD, both enhancer-mediated repression and aberrant de-repression of normally modestly expressed genes are observed in the majority of MSNs. Although altered dynamics of DNA methylation and hydroxymethylation are not primary drivers of toxicity, we believe these changes may be critical to the progressive nature of HD pathogenesis and the temporal window of symptomatic onset since the accumulation of these modifications stabilize the initial *mHTT*-induced defects in mediator engagement and transcription complex occupancy. Thus, our data argue strongly that inhibition of somatic CAG expansion in HD patients that are already symptomatic will be insufficient to reverse the epigenetically stabilized disruptions in gene expression. The data also raise the interesting possibility that TET agonists may restore enhancer activity and re-activate repressed genes, thereby alleviating deleterious impacts of expanded mutant huntingtin and restoring neuronal function.

Concluding Remarks

The debate raised by this and other recent studies of human HD pathophysiology centers on an important issue: are cells experiencing massive disruption of gene expression functional or dysfunctional in the human brain? The nature and precision of the epigenetic data allow the conclusion that >80% of HD MSNs are experiencing epigenetic changes that alter the functions of biochemical pathways shown to be critical for striatal function, as well as strongly suppressing expression of individual genes that by themselves can cause human neurological and neurodegenerative disease. Our data demonstrate that these gene expression changes in HD MSNs result from the impact of expanded mutant huntingtin on intrinsic epigenetic mechanisms regulating transcription, and provide mechanistic models for the findings that variants in *MED15* and *TCERG1* impact HD pathogenesis in humans. When taken together with human genetic studies and results from cell culture and animal models of HD, our findings support a model for HD pathogenesis that includes a protracted stage of cellular dysfunction prior to loss of neurons from the HD brain. They argue against a recent computational model of HD concluding "that in HD, at any one time, most neurons have an innocuous but unstable HTT gene" [5].

Acknowledgements

This study was supported by funding from the CHDI Foundation, and we are thankful to T.F. Vogt, V. Beaumont and J. Chen for helpful discussions throughout the project. We thank Karen Liu and C. Wang for technical assistance. We are also grateful to the Rockefeller University Genomics Resource Center for advice and support. We thank J.P.G. Vonsattel and A.F. Teich from Columbia University Alzheimer's Disease Research Center (funded by NIH grant P30AG066462), D. Keene from University of Washington BioRepository and Integrated Neuropathology Laboratory (supported by Alzheimer's Disease Research Center grants AG066567 and AG066509), S. Berretta and the Harvard Brain Tissue Resource Center, The University of Michigan Brain Bank (P30AG053760/ P30AG072931 University of Michigan Alzheimer's Disease Core Center), University of Maryland Brain and Tissue Bank, Netherlands Brain Bank, and UCLA Human Brain & Spinal Fluid Resource Center (supported by the NIH and the US Department of Veterans Affairs) for assisting and providing post-mortem brain tissues.

Methods

Human samples

Deidentified tissue samples analyzed in this study were determined to be exempt from Institutional Review Board review according to 45 CFR 46.102 (f). For this work, fresh frozen brain samples were obtained from Miami's Brain Endowment Bank, University of Washington BioRepository and Integrated Neuropathology Laboratory, Columbia University Alzheimer's Disease Research Center, University of Maryland, Science Care, and Netherlands Brain Bank or through the National Institutes of Health (NIH) NeuroBioBank and sourced from either the Harvard Brain Tissue Resource Center, The University of Michigan Brain Bank or the NIH Brain & Tissue Repository-California, Human Brain & Spinal Fluid Resource Center, VA West LA Medical Center (Los Angeles, CA). Drug addiction and schizophrenia as well as clinical evidence of brain cancers were reasons for sample exclusion, whereas samples from donors with a history of other non-brain cancers and diabetes were accepted. Caudate nucleus, putamen BA4, hippocampus and cerebellar vermis were used for isolation of nuclei. The brain regions used from each donor and their age, race, sex and post-mortem interval are noted in Supplementary Table 1.

Sorting nuclei of striatal MSNs and glia, cerebellar granule neurons, hippocampal neurons and cortical pyramidal neurons

Nuclei were isolated from fresh frozen brain tissue samples as previously described in ***Basic Protocol 1*** [62]. Nuclei were labelled and sorted with FANS as previously described in ***Basic Protocol 2*** (neocortical cell types), ***Basic Protocol 3*** (striatal MSNs) [62], and in [1](cerebellar granule neurons and glial cell types of the striatum). The nuclei of hippocampal neuron types were labeled with anti-NeuN (1:500, Millipore Sigma ABN91), anti-SATB2 (1:100, Santa Cruz sc-81376), and anti-ELAVL2 (1:100, Sigma-Aldrich HPA063001) antibodies and with donkey anti-chicken Alexa Fluor™ 647 (1:1000, Thermo Fisher A78952), donkey anti-mouse Alexa Fluor™ 488 (1:1000, Thermo Fisher A21202), and donkey anti-rabbit Alexa Fluor™ 594 (1:1000, Thermo Fisher A21207) secondary antibodies. The nuclei of ADARB2-expressing interneurons (NeuN+^{lo}), CA1 pyramidal neurons (NeuN+^{hi}, SATB2+ ELAVL2-), CA2/3 pyramidal neurons (NeuN+^{hi}, SATB2-, ELAVL2+) and granule neurons of the dentate gyrus (NeuN+^{hi}, SATB2- ELAVL2-) were collected as described in **Supplementary Note 1**.

Isolation of RNA, gDNA and Tn5-Treated DNA

RNA and gDNA from sorted nuclei, and gDNA from Tn5-treated nuclei were isolated and turned into sequencing libraries as previously described in ***Basic Protocol 4, Support Protocol 1*** and ***Support Protocol 2*** [62]. FANS-seq and ATAC-seq sequencing reads were processed and quality-controlled according to ***Basic Protocol 5*** [62].

Differential gene expression analysis

Differential gene expression analysis was performed using 'Genebody.Counts' (FANS-seq reads mapped to full gene bodies) and DESeq2 [63,64] (version 1.40.2). Differential gene expression analysis results were filtered to exclude genes for which none of their annotated TSS positions in NCBI Refseq hg38 (version 109.20211119) overlapped with ATAC-seq consensus peaks defined for the cell types based on presence in presence in at least half of control samples analyzed. For neocortical neuron types and MSNs, previously published lists of differentially expressed genes

were used [1,2]. Pheatmap R package (version 1.0.12) was used for visualizing gene expression levels on heatmaps.

oxBS-seq Library Preparation and Processing

OxBS-seq libraries were prepared according to manufacturer's instructions in the Ultralow Methyl-Seq with TrueMethyl® oxBS (Tecan Genomics, #0541-32). Briefly, 100-300 ng of gDNA (~50,000-150,000 nuclei) with 1% unmethylated lamda phage DNA spike-in was sonicated and purified for 200bp fragments. Purified gDNA fragments were then split into paired samples such that 30 ng was used for BS-seq (mock-oxidation reaction) and the remaining gDNA was used for oxBS-seq (with oxidation reaction). Libraries were indexed, PCR amplified and sequenced to >300million reads per sample on NovaSeq6000 (2 × 100 bp).

The raw OxBS-seq fastq reads were trimmed with the parameters "--paired --fastqc --cores 6 –stringency 3 --clip_R1 5 --clip_R2 10" using trim-galore (v 0.6.6), cutadapt (v3.4) and fastqc (0.11.9) [65,66] (https://github.com/s-andrews/FastQC). Comprehensive quality control reports were generated using multiqc (v1.28) [67].

The bisulfite reference genome for the human genome hg38 (GRCh38.primary) was prepared using the command "bismark_genome_preparation" in Bismark (v0.22.3) [68]. The trimmed reads were then mapped to the bisulfite genome using the parameters "--bowtie2 --parallel 16 --bam '$HG38_BISMARK' -1 '$[FASTQ_R1_TRIMMED]' -2 '$[FASTQ_R2_TRIMMED]'" (bowtie2 v2.4.4; samtools v1.12) [69,70]. The bam file was processed with deduplicate_bismark with default parameters to remove duplicate reads. The deduplicated bam file was used to calculate methylation level at each cytosine site (CpG and Non-CpG) using Bismark_methylation_extractor with the parameters "-p --gzip --comprehensive --multicore 5 --merge_non_CpG --bedGraph --counts --buffer_size 200G --cytosine_report --genome_folder '$HG38_BISMARK' --report". The resulting genomewide Cytosine Reports were used as inputs for differential methylation analysis with methylKit [71].

For visualization of methylation levels in IGV, .meth files were generated from the deduplicated bam files using methcounts and fasta file for HG38 with default parameters (Methpipe package (v5.0.1) [72,73]. These .meth files were used as inputs for the mlml tool (within the methpipe package) with the parameters "-v -u '$BS' -m '$OXBS'" to estimate methylation and hydroxymethylation levels at each Cytosine position across the genome (separately for CpG and Non-CpG contexts) [74]. A custom R script was developed that dynamically split the MLML outputs, specifically, the mC and hmC data for each Cytosine and saved two separate bedGraph files. These bedGraphs were sorted by chromosome positions using bedSort (ucsc-bedsort-469) which were finally converted to bigwigs using bedgraphtobigwig (ucsc-bedgraphtobigwig-472) and visualized with IGV.

ATACseq Consensus peak calling and differential analysis

Consensus peak generation and filtering

NarrowPeak files from each sample were imported as *GRanges* objects via *rtracklayer::import()* (*rtracklayer).* To focus on summitcentered regions, each peak was resized to a 501 bp window (±250 bp) around its midpoint using *GenomicRanges::resize()* (*GenomicRanges*). Within each sample, we then applied an iterative clustering algorithm:

1. All peaks were merged into preliminary clusters via GenomicRanges::reduce().

2. Cluster membership was assigned by *findOverlaps()*, and within each cluster the single peak with the highest CPMnormalized score (computed from the original narrowPeak pvalue field using *edgeR::cpm())* was retained.
3. Retained peaks were removed from the pool, and steps 1–2 were repeated until no peaks remained.

The resulting samplespecific "winner" peaks were concatenated across all samples into a pansample GRangesList, unlisted, and sorted. We then built a peak × sample score matrix by reoverlapping each peak to every sample's peak set (via findOverlaps()) and extracting CPM scores. Finally, peaks with CPM > 5 in at least two samples (i.e. rowSums(scoreMat > 5) ≥ 2) were selected to produce the final iterative consensus peak file filtered for 2 samples and at least 5CPM.

Read counting and differential analysis

The filtered consensus peaks were converted to a *SummarizedExperiment* object by counting overlaps between the iterative consensus peaks against all BAMs using *GenomicAlignments::summarizeOverlaps*. A DESeq2 dataset was constructed with *DESeqDataSetFromMatrix()* and tested for differential accessibility between experimental groups *(DESeq2::DESeq(); DESeq2)*, incorporating sample metadata from the design matrix. Variancestabilized counts were obtained via vst(), and principalcomponent plots were generated with plotPCA(). Differential peaks were called as padj < 0.05.

Peak Binarization and Cell type Assisgnment

To identify highly reproducible peaks selectively active in each cell type, we began with our final ATACseq SummarizedExperiment and sample metadata grouping samples by the Celltype column. Promoter regions were defined as ±150 bp around transcription start sites using GenomicRanges::promoters() on TxDb.Hsapiens.UCSC.hg38.knownGene Peaks overlapping these promoters were separated from nonpromoter peaks via findOverlaps(). Nonpromoter peaks were converted into a DESeq2 dataset and variance-stabilized counts computed via DESeq2::vst() yielding a matrix (vst_nonpromoter) in which rows are peaks and columns are samples. To facilitate comparisons across cell types, we reindexed rows sequentially and ensured row names matched the vst output. For each cell type with at least two replicates, we calculated the perpeak mean and standard deviation of vst counts across samples of that cell type using matrixStats::rowMeans2() and matrixStats::rowSds() The results were stored in matrices meansATAC and sdsATAC, with columns corresponding to cell types in a consistent order. To call binary presence/absence across cell types, we first sorted each peak's means in ascending order (maintaining the matching cell type labels). For each cell type *i* (in rank order after the first), a peak was called "present" in *i* if

$$\text{Mean}_i > \text{mean}_{i-1} + 2 \times \text{sd}_{i-1}$$

This threshold guards against calling peaks present in a cell type unless their signal exceeds the previous cell type's mean by two standard deviations. Peaks called "present" for cell type *i* were removed from further consideration in subsequent rounds, and their binary membership vectors (indicating which of the remaining cell types they would also "pass" had they not been removed) were recorded. Combining these membership vectors across all iterations produces a perpeak, percelltype logical matrix. A concatenated label string (e.g. "Astrocyte_Oligo") summarizes each peak's cell type-specificity in a new clusterfull column. We merged this binary membership information back into the rowData of the nonpromoter DESeq2 object, aligning by peak index.

The enriched rowData now includes, for each peak, logical flags for presence in each cell type and the composite clusterfull label.

software:
– rtracklayer v1.58.0; GenomicRanges v1.50.1; GenomicAlignments v1.32.0; edgeR v3.38.4; SummarizedExperiment v1.30.0; DESeq2 v1.38.1; TxDb.Hsapiens.UCSC.hg38.knownGene v3.17.0; writexl v1.5.2; plus tidywrangling via stringr v1.5.0 and dplyr v1.1.2, matrixStats v0.62.0

cCRE mapping

In order to identify candidate cis-regulatory elements (cCREs) whose chromatin accessibility is linked to local gene expression, we developed a custom R function utilizing *rowCorr*, that integrates paired RNAseq and ATACseq data originating from the same tissue donor and brain region. Briefly, both assays are represented as *SummarizedExperiment* objects (SE_RNA and SE_ATAC), with rows corresponding to genes and peaks, respectively, and columns to paired biological samples. For each gene in SE_RNA, we define a genomic search window of ±1 Mb (configurable via the max.dist argument) around its transcription start site by resizing the gene ranges, and we pair these to ATAC-seq peaks whose midpoints fall within that window. The resulting overlaps are recorded in a data frame *o* that includes the gene index, peak index and linear distance between each gene–peak pair.

Within *getCorr*, we then compute Pearson correlations of normalized accessibility and expression values for all gene–peak pairs using our C++-backed *rowCorCpp* routine, yielding a vector of correlation coefficients per pair. Test statistics are converted from correlation to t-statistics accounting for the number of samples, and two-sided P values are derived via the Student's t-distribution. False-discovery rates (FDR) are controlled by the Benjamini–Hochberg procedure to produce FDR-adjusted P values.

To assess background correlation levels, the auxiliary function *.getNullCorrelations* constructs a chromosome-stratified null distribution: for each chromosome, peaks on all other chromosomes are randomly sampled (up to 1,000 per chromosome) and correlated against genes on the focal chromosome to generate a pool of "trans-" correlations. The empirical mean and standard deviation of this null pool are then used to compute empirical P values for the observed cis correlations, with subsequent FDR correction. Finally, gene–peak pairs with both nominal and empirical FDR below predefined thresholds can be labeled as cCREs for downstream analysis.

Methylation modelling

We analyzed two independent datasets, dMSN and iMSN, each comprising per-gene summaries of single-site methylation changes and corresponding RNA-seq differential expression. For each gene i, we extracted counts of significantly decreased mCpG ($n\downarrow_1$), mCpH ($n\downarrow_2$), and hmCpH ($n\downarrow_3$) sites, as well as the counts of increased sites ($n\uparrow_1$, $n\uparrow_2$, $n\uparrow_3$). Genes lacking any changing sites in a context were assigned a pseudo-count of one to avoid division by zero.

Aggregated methylation-loss metric

$$D_i = n\downarrow_1 + n\downarrow_2 + n\downarrow_3$$

$$N_i = (n\uparrow_1 + n\downarrow_1) + (n\uparrow_2 + n\downarrow_2) + (n\uparrow_3 + n\downarrow_3)$$

$$x_i = D_i / \max(N_i, 1), \quad 0 \le x_i \le 1$$

Response variable

We defined a binary outcome:

$$y_i = 1 \text{ (if } \log_2(\text{FoldChange}) > 0) \text{ else } 0.$$

Model fitting

$$logit(P(y_i=1)) = \beta_0 + \beta_1 x_i$$

where $logit(p) = ln[p/(1-p)]$.

Prediction and calibration

$$\hat{p}_i = 1 / [1 + exp(-(\hat{\beta}_0 + \hat{\beta}_1 x_i))]$$

iMSN genes were binned by x_i (width=0.10); for each bin we computed observed up-fraction and mean predicted probability, then plotted them against bin centers for calibration.

FANS-C&R-seq Nuclear Staining, Library Preparation and Processing

Nuclei were isolated from fresh frozen brain tissue samples as previously described in ***Basic Protocol 1*** [62], but with modified fixation conditions to ensure sensitive detection of C&R-seq targets. Once isolated, bulk nuclei were lightly fixed with 0.5% methanol-free formaldehyde for 1 minute rotating at RT, quenched with 125mM glycine for 5 minutes rotating, washed, then resuspended in 1ml FANS wash buffer (1X PBS, 0.5% BSA, 1X Roche cOmplete EDTA-free Protease Inhibitor) with 0.01% digitonin to block and permeabilize for 30 minutes rotating at RT. 10µl of nuclei were removed for yield quantification via Thermo Fisher Countess II, then sample was partitioned into 5 pools of nuclei for 5 separate histone PTM target conditions – euchromatic marks H3K4me3 (Active Motif 39159) and H3K27ac (Active Motif 39133), and rabbit IgG negative control (Epicypher 13-0042). FANS-C&R antibody buffer (FANS wash buffer, 0.01% digitonin, 0.5mM Spermidine, 2mM EDTA) was then prepared containing appropriate FANS labelling antibody dilutions as described in ***Basic Protocol 3*** (striatal MSNs) [62] and in [1] (cerebellar granule neurons and glial cell types of the striatum).

Partitioned bulk nuclei conditions were pelleted (4 minutes, 800xg). Due to yield variability between donor aliquots, the following guidelines for bulk nuclei resuspension were established to obtain reproducible target signals, each target condition must contain: 1) 3-4x10^3 nuclei/µl of FANS-C&R-antibody buffer, 2) a minimum of 50µl FANS-C&R-antibody buffer per target condition, 3) 0.01 ug/µl histone PTM target antibody. Once resuspended, each condition was incubated rotating 4C overnight. Samples were then processed for secondary antibody labeling and sorted into neuronal and glial cell types for each histone PTM target condition as described in ***Basic Protocol 3*** [62] and in [75]. Sorted nuclei counts for each histone PTM condition and cell type sample were kept within the range of 1x10^4 – 3x10^5, with maximum yield prioritized due to sample scarcity.

C&R-seq was carried out same-day using a modified protocol from (Meers et al. 2019) and (Epicypher C&R Manual v5.1) with reagents from CUTANA™ ChIC / CUT&RUN kit (Epicypher 14-1048). Sorted samples were centrifuged (4 minutes, 800xg), resuspended in C&R-wash buffer (20mM HEPES pH 7.5, 150mM NaCl, 0.5mM Spermidine, 1x Roche cOmplete EDTA-free Protease Inhibitor), centrifuged again, resuspended in 100µl C&R-wash buffer and transferred to labeled 8-strip tubes. 10.5µl of activated ConA beads (see Epicypher C&R Manual 5.1 for activation protocol) were added to each 100µl reaction and incubated for 15 minutes RT on a gentle lateral nutator. Sample 8-strip tubes were placed on a 96 tube magnetic rack (Thermo Fisher AM10050) for 3 minutes, supernatant was removed and replaced with 200µl cold C&R-wash buffer while still on magnet. Samples were washed twice more in cold C&R-DIG buffer (C&R-wash buffer, 0.01% digitonin), then resuspended in 50µl cold C&R-DIG buffer with 1X pA/G-MNase (Epicypher, 15-1016) and incubated 10 minutes RT on a gentle lateral nutator. After returning to magnetic rack on ice and washing three times in cold C&R-DIG buffer, samples were resuspended in 50µl cold C&R-DIG buffer containing 2mM Cacl2 and incubated rotating at 45 ° angle for 2 hours at 4C to cleave pA/G-Mnase-bound chromatin targets. 34µl of RT C&R-Stop buffer (340mM NaCl, 20mM EDTA, 4 mM EGTA, 50µg/mL RNase A, 50ug/mL Glycogen) containing 1pg E.Coli K12 spike-in DNA (Epicypher 18-1401) was added to samples, which were then incubated in thermocycler for 10 min at 37C to release cleaved chromatin into supernatant. Nuclei were quick-centrifuged and sample chromatin supernatants were removed after a 3 minute incubation on magnet. 1µl 8% SDS and 2µl 10µg/µl Proteinase K (Qiagen, 19131) were added to all extracted sample chromatin supernatants, which were then mixed and incubated overnight at 55C via thermocycler.

The following day, samples were transferred to low-bind microcentrifuge tubes, brought up to 250µl with water, then vortexed with 250µl phenol:chloroform:isoamyl alcohol 25:24:1 pH 8 (Sigma-Aldrich P2069) and incubated 10 mins RT. Samples were quick-centrifuged, then transferred to pre-centrifuged phase lock gel heavy tubes (Qiagen, 129056) and centrifuged (5 min, 16000xg). 250µl chloroform (Sigma-Aldrich C2432) was mixed with sample aqueous phases, and tubes were centrifuged again. Sample aqueous phases were transferred to new pre-centrifuged phase lock gel heavy tubes, mixed with 400µl chloroform and centrifuged (3 minutes, 12000xg). Sample aqueous phases were transferred to new microcentrifuge tubes and brought up to 250µl with water. Samples were vortexed with 25µl 3M NaAc (Thermo Fisher R1181), 550µl 100% EtOH, and 15µg GlycoBlue™ Coprecipitant (Thermo Fisher AM9515) and placed at -20C for 1-4 days to precipitate cleaved DNA. Samples were then centrifuged down into DNA pellets (1 hour, 20000xg 4C), washed twice in 1ml 75% EtOH (5 minutes, 20000xg), air dried for 5-10 minutes, then DNA was resuspended in 25µl 0.1X TE buffer and frozen.

Libraries were prepared from purified sample DNA using the CUTANA™ CUT&RUN Library Prep Kit (Epicypher 14-1001, 14-1002) as described by (Epicypher C&R library prep manual v1.5) with the following modifications: 1) Illumina adapter was reduced to 0.75 pmol per sample, 2) index amplification was raised to 16 cycles, 3) sample libraries were brought up to 50µl with 0.1X TE before the second bead cleanup, 4) the final elution volume was raised to 14µl 0.1X TE with 13µl sample library extracted. Due to an increased likelihood of adapter-dimer formation caused by low sample input material, fragment size distribution was quantified on the Agilent 4200 TapeStation system using 1µl of each sample library. The molar proportion of dimer fragments (0-170bp) to signal fragments (170-1000bp) was used to pool all samples on a weighted basis, producing a pool containing equimolar signal fragments for each sample. Pool underwent 1X SPRI

bead clean up (Beckman Coulter) and was sequenced 50-80million reads per sample on NovaSeq6000 (2 × 100 bp).

Sample paired-end FastQs were aligned in R using rsubread::align() with fragment length range of 10bp-600bp, first aligned against human genome GRCh38.p13 and then aligned against E.Coli MG1655 for downstream spike-in normalization. Peaks were called in python using MACS2 callpeak function with flags -f BAMPE -p 1e-5 --keep-dup=all -g hs -c <appropriate IgG control>.

Software: Rsubread 2.16.0, macs2 2.2.9.1

In situ hybridization (ISH) and image analysis

ISH was performed using RNAscope™ Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics, 323100) according to the manufacturer's instructions for fresh-frozen samples with the following modifications: sections were fixed for 15 minutes in chilled 4% formaldehyde in PBS, then rinsed in PBS, dehydrated, and photobleached in 100% ethanol for 40-45 hours using a photobleaching device [76]. The following probes were used: Hs COCH (Advanced Cell Diagnostics, 1104401-C1), Hs PDE10A (Advanced Cell Diagnostics, 466151-C2), Hs PCP4 (Advanced Cell Diagnostics, 446111-C2), TSA Vivid™ fluorophore 570 (Advanced Cell Diagnostics, #323272, used for COCH detection) and TSA Vivid™ fluorophore 650 (Advanced Cell Diagnostics, #323273, used to detect PDE10A and PCP4 probes). Sections were counterstained with DAPI and then cover-slipped with ProLong™ Diamond Antifade Mountant (ThermoFisher, P36970). Images were acquired on a Zeiss Confocal Microscope LSM 710 confocal microscope using a Plan-Apochromat 20x, 0.8 NA objective lens.

Regions of interest (ROIs), each containing an MSN, were defined in ImageJ (1.54f) based solely on the signals from DAPI and the probe for *COCH*, an MSN marker gene whose expression level does not change in HD based on our previously published data [1]. For each ROI, the number of *COCH*+ puncta and *PDE10A*+ puncta (84-305 ROIs per donor) or *PCP*+ puncta (38-224 ROIs per donor) was counted from each ROI using a previously published ImageJ script (https://zenodo.org/records/14506480) [77]. The '*PDE10A* puncta'/'*COCH* puncta' or '*PCP4* puncta'/'*COCH* puncta' ratios were calculated and plotted separately for each donor. Each donor's median value of these ratios was used for statistical analysis of the difference between HD and control donors.

Author contributions

Conceptualization: M.B., K.M., N.H. Formal analysis: M.B., K.M., I.I., M.E.S., H.C., T.S.C., D.B., C.P., N.D. Investigation: M.B., K.M., I.I., M.E.S., Y.H., E.S., P.D., C.P. Resources: L.K., N.D. Data Curation: M.B., L.K., N.D. Visualization: M.B., K.M. Writing of original draft: M.B., K.M., N.H. Writing, review and editing: M.E.S., I.I., Supervision: N.H. Funding acquisition: N.H.

Data availability

All sequencing datasets generated as part of this study will be publicly available in NCBI GEO and will be available for the review process upon request. Further information and requests for resources and reagents should be directed to the lead contact, N. Heintz (heintz@rockefeller.edu). TRAP data from zQ175 and R6/2 mice (GEO dataset GSE152058, 10.1016/j.neuron.2020.06.021) have been published before. The publicly available resources

used (reference genome sequence and annotation, and computational analysis tools) are described in the Methods section.

Code availability

Only publicly available tools were used in data analysis. The analysis parameters used have been described in Methods.

Competing interests

The authors declare no competing interests.

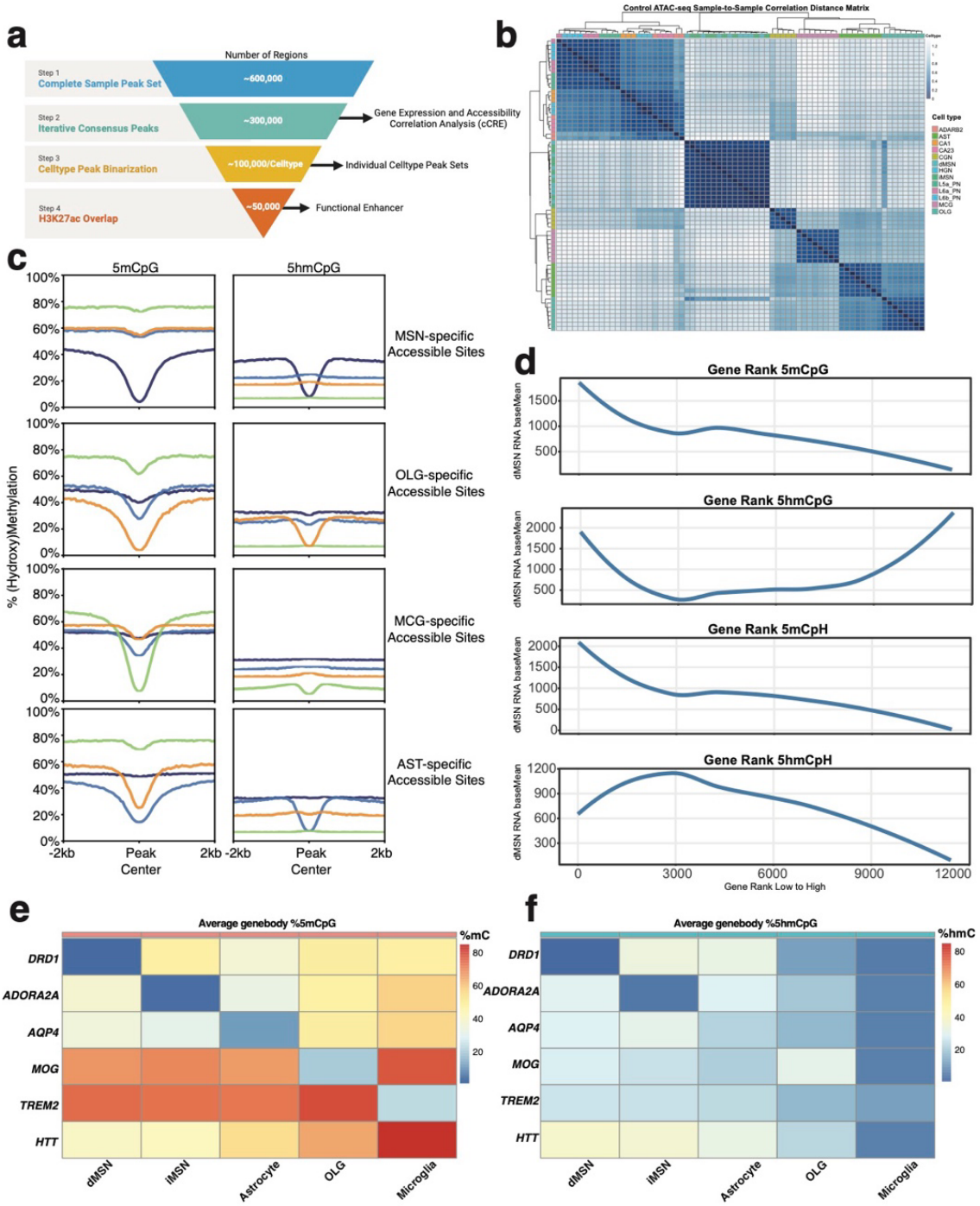


Fig S1

a, The analysis of chromatin accessibility data included (step 1) peak calling, (step 2) identification of robust and reproducible consensus peaks and, as a subset, cCREs that correlate positively or negatively with transcript levels in matched FANS-seq datasets, (step 3) peak binarization-based definition of peak sets for each of the cell types, and their use in the analysis of HD-associated differential accessibility and (step 4) integration with H3K27ac CUT&RUN-seq datasets to identify accessible chromatin regions that are active enhancers. **b,** Distance clustering matrix of chromatin accessibility profiles of individual cell type-specific ATAC-seq datasets from control donors, based on normalized counts over iterative consensus peaks. **c,** Average level of CpG (hydroxy)methylation over cell type–specific accessible regions in striatal MSNs, Oligodendrocytes, Astrocytes and Microglia. **d,** Relationship between expression level (baseMean in FANS-seq data) and gene body average cytosine modification level (shown separately for 5mCG, 5hmCG, 5mCH, 5hmCH) in dMSN samples from control donors. Genes have been ranked along the x-axes according to average cytosine modification level.

e, f, Average level of (f) CpG methylation and (g) hydroxymethylation over gene bodies of cell type-specific marker genes of striatal dMSNs (*DRD1*), iMSNs (*ADORA2A*), oligodendrocytes (*MOG*), astrocytes (*AQP4*) and microglia (*TREM2*).

Fig S2 | Cell type-specific accessible regions in *MSH3*, *FAN1* and *HTT* genes.
a,b, Representative distribution of sequencing reads over (a) *HTT*, (b) *MSH3* and *FAN1* genes in ATAC-seq datasets generated from nuclei of human brain cell types. The position of cCREs linked to these genes is shown.
c, Distribution of sequencing reads over *Msh3* and *Pcp4* genes in ATAC-seq datasets generated from anti-NeuN-positive and -negative nuclei isolated from mouse striatum. Accessibility at neuronal marker genes (like *Pcp4*) was used to assess purity of isolated populations.

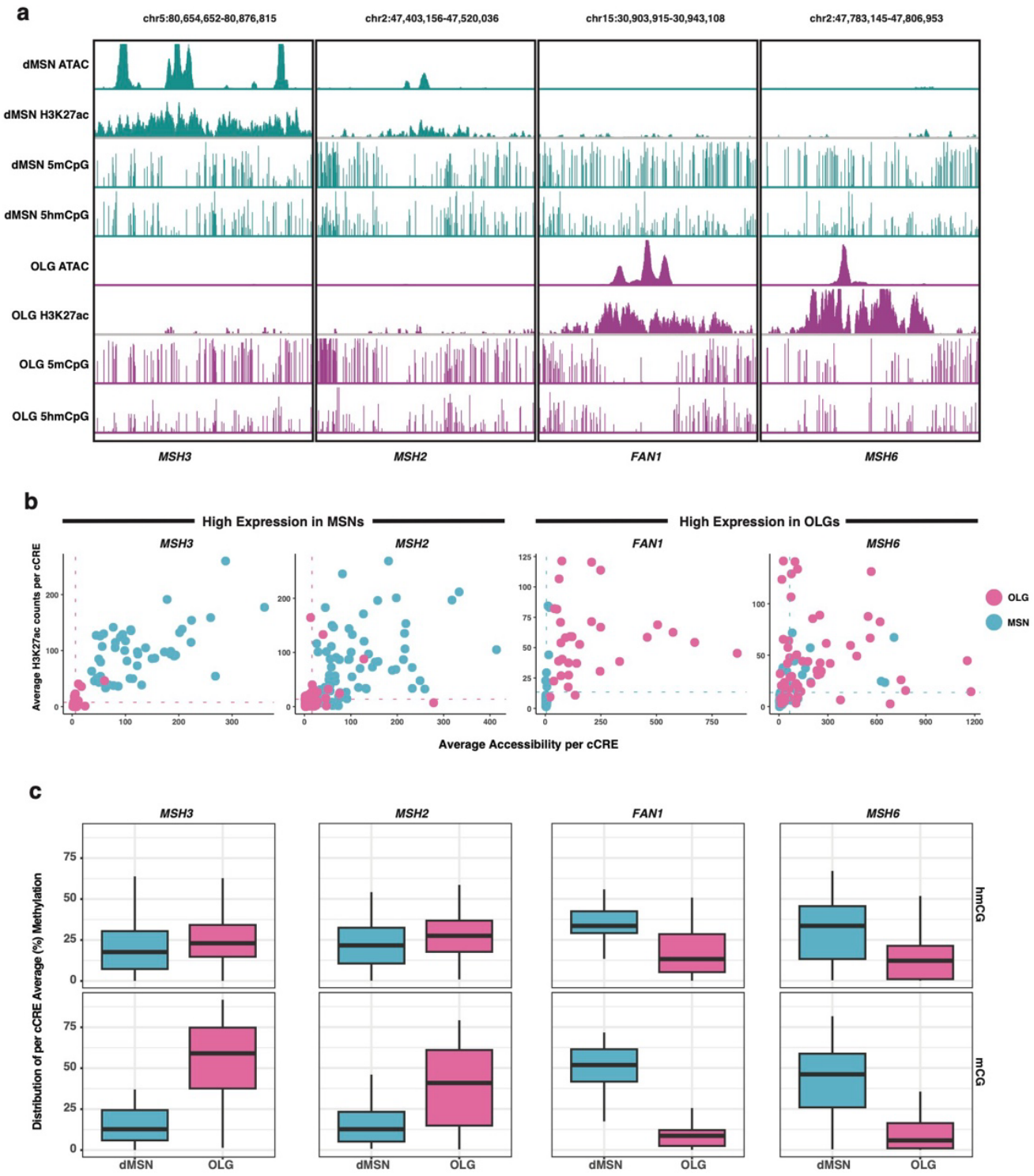


Fig S3

a, Representative IGV tracks showing selected (+)cCREs linked to *MSH3, MSH2, FAN1 and MSH6,* and their accessibility, H3K27ac signal and level of 5mCpG and 5hmCpG in dMSNs and oligodendrocytes. **b,** Normalized counts of ATAC-seq and CUT&RUN-seq reads from dMSN and oligodendrocyte samples were plotted to compare the levels of accessibility and H3K27ac at individual (+)cCREs linked to *MSH2* and *MSH3* (higher expression in MSNs) or to *FAN1* and *MSH6* (higher expression in oligodendrocytes). Dotted lines represent the mean ATAC-seq signal (vertical) or mean H3K27ac signal (horizontal) over all (+)cCREs in the cell type with lower expression of the gene. **c,** Level of CpG hydroxymethylation and methylation over (+)cCREs linked to genes with higher expression in MSNs (*MSH2, MSH3*) or in oligodendrocytes (*FAN1, MSH6*). The mean, Q1 and Q3 of all (+)cCREs are shown.

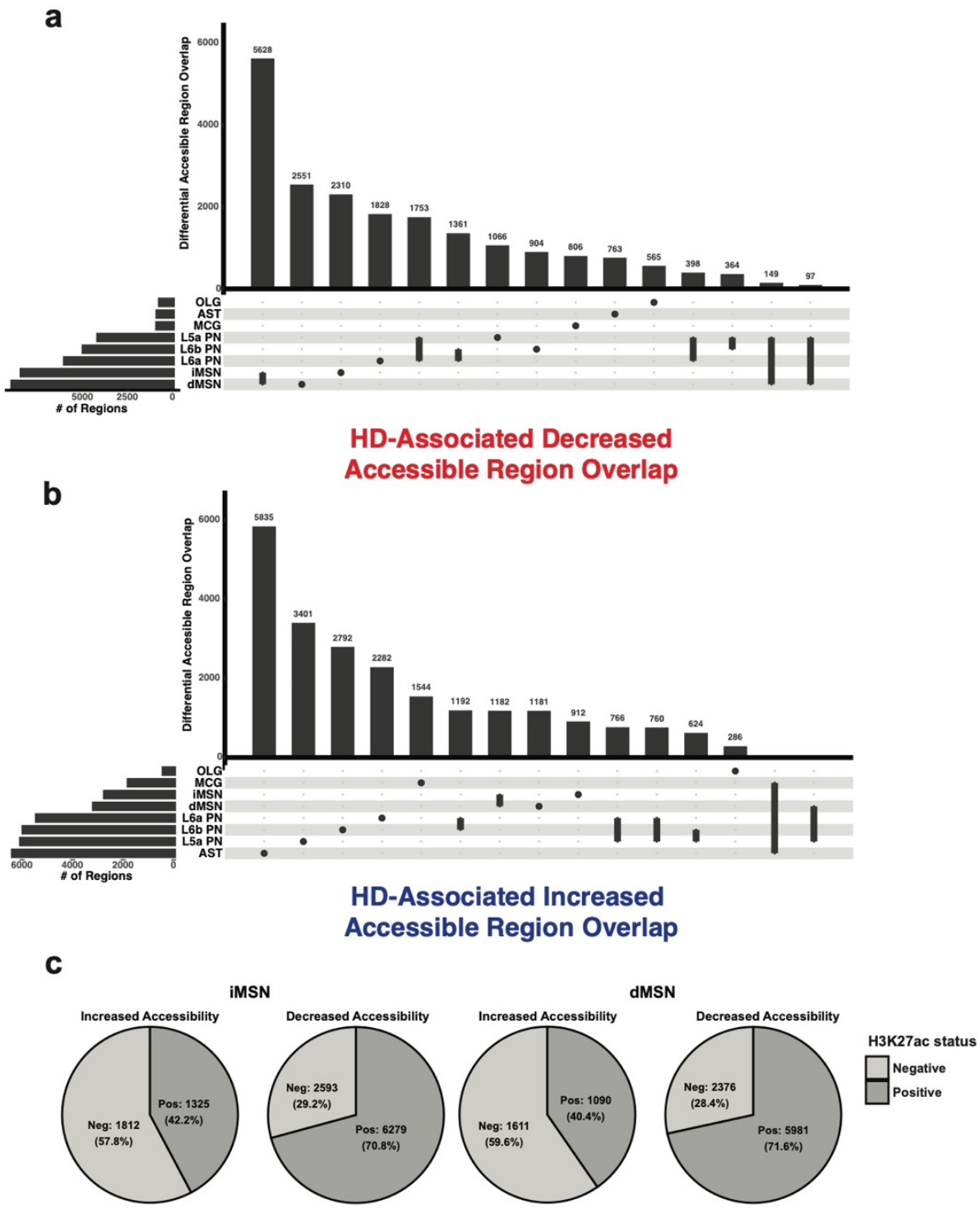
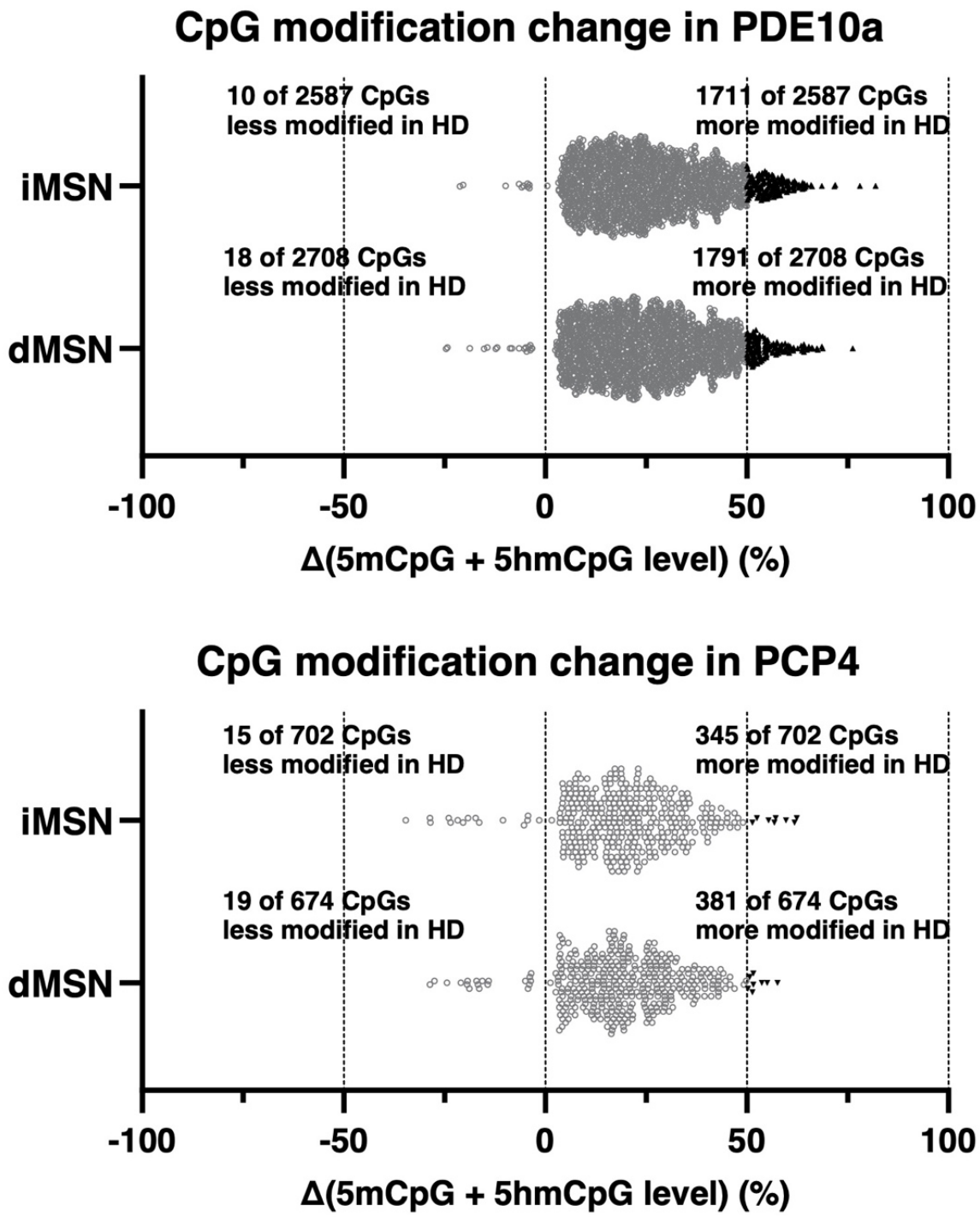


Fig S4

a,b, Upset Plot indicating the number of HD-associated increased and decreased acessible chromatin regions unique to or shared by the cell types analyzed. Note the minimal overlap between differentially accessible regions in cortical PNs and MSNs, indicating cell type–specific chromatin remodeling in response to somatic CAG expansion.

c, Overlap of HD-associated differential accessible regions separately for iMSNs and dMSNs with control MSN H3K27ac peak sets, the percentage of peaks that overlap is displayed (dark grey, pos)

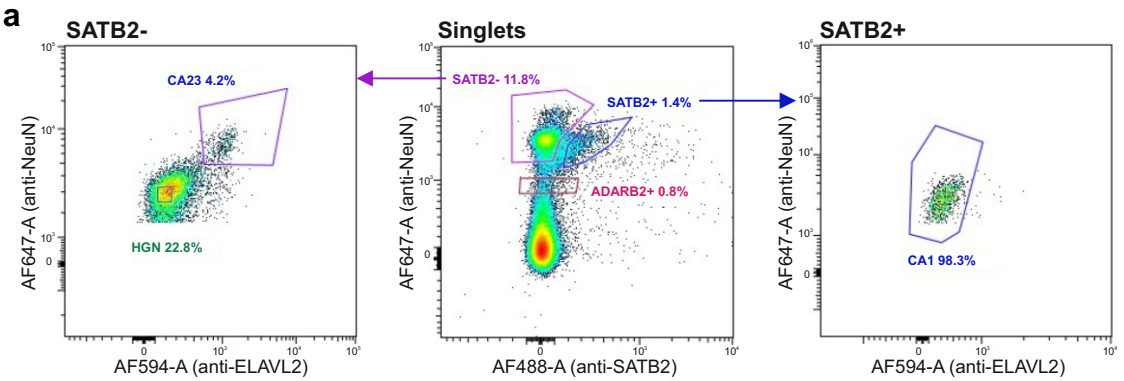


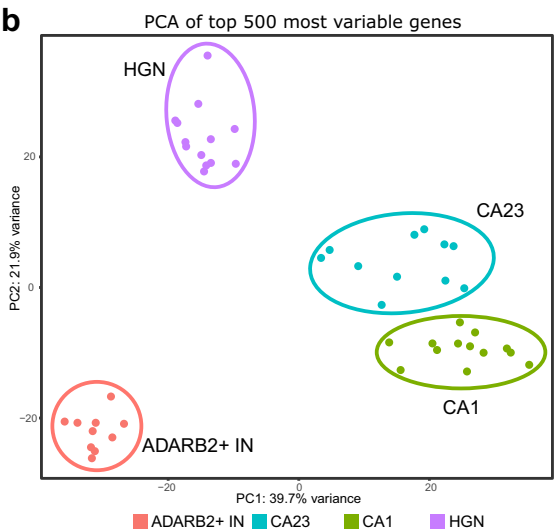
878 **Fig S5**

a, HD-associated change in the modification level of individual CpGs in *PDE10A* and *PCP4* gene bodies (n = 5 control and n = 5 HD donors)

879

Supplementary Note 1 - Labeling and sorting of the nuclei of hippocampal neuronal types





b, Principal component analysis (PCA) of FANS-seq data from hippocampal neuronal cell types.

Staining procedure: Nuclei were resuspended in 100 µl of wash buffer and incubated overnight with primary antibodies anti-NeuN (1:500, Millipore Sigma ABN91), anti-SATB2 (1:100, Santa Cruz sc-81376), and anti-ELAVL2 (1:100, Sigma-Aldrich HPA063001). The nuclei were washed three times by pelleting (3 minutes, 1000xg) and resuspension in wash buffer (1x PBS, 0.2% bovine serum albumin). They were resuspended in 500 µl wash buffer and incubated for 30 minutes with secondary antibodies donkey anti-chicken Alexa Fluor™ 647 (1:1000, Thermo Fisher A78952), donkey anti-mouse Alexa Fluor™ 488 (1:1000, Thermo Fisher A21202), and donkey anti-rabbit Alexa Fluor™ 594 (1:1000, Thermo Fisher A21207). After three more washes, the nuclei were incubated for 10 minutes with DAPI (1:10,000) before sorting on the same day.

Sorting procedure: Sorting was conducted on Sony MA900. Intensity of DAPI signal was used to select single nuclei and to exclude aggregates of nuclei. Nuclei of ADARB2 interneurons were sorted from the lower end of the NeuN+ population (**a**, middle panel). The NeuN+ population was separated into SATB2+, which contains nuclei of CA1 pyramidal neurons, and SATB2- which contains nuclei of CA2/3 pyramidal neurons and granule neurons of the dentate gyrus. The SATB2+ CA1 neurons were further purified by selecting the ELAVL2- population (**a**, right panel). From the SATB2- population, ELAVL2+ CA2/3 neuron nuclei and ELAVL2- granule neuron nuclei were collected (**a**, left panel).

References

1. Matlik, K. *et al.* Cell-type-specific CAG repeat expansions and toxicity of mutant Huntingtin in human striatum and cerebellum. *Nat Genet* **56**, 383-394 (2024).
2. Pressl, C. *et al.* Selective vulnerability of layer 5a corticostriatal neurons in Huntington's disease. *Neuron* **112**, 924-941 e10 (2024).
3. Gusella, J.F., Lee, J.M. & MacDonald, M.E. Huntington's disease: nearly four decades of human molecular genetics. *Hum Mol Genet* **30**, R254-R263 (2021).
4. Kennedy, L. *et al.* Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet* **12**, 3359-67 (2003).
5. Handsaker, R.E. *et al.* Long somatic DNA-repeat expansion drives neurodegeneration in Huntington's disease. *Cell* **188**, 623-639 e19 (2025).
6. GeM-HD Consortium, C. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* **162**, 516-26 (2015).
7. GeM-HD Consortium, C. CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell* **178**, 887-900 e14 (2019).
8. Genetic Modifiers of Huntington's Disease, C. *et al.* Genetic modifiers of somatic expansion and clinical phenotypes in Huntington's disease reveal shared and tissue-specific effects. *bioRxiv* (2024).
9. Wheeler, V.C. & Dion, V. Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models. *J Huntingtons Dis* **10**, 123-148 (2021).
10. Hong, E.P. *et al.* Huntington's Disease Pathogenesis: Two Sequential Components. *J Huntingtons Dis* **10**, 35-51 (2021).
11. Gu, X. *et al.* Uninterrupted CAG repeat drives striatum-selective transcriptionopathy and nuclear pathogenesis in human Huntingtin BAC mice. *Neuron* **110**, 1173-1192 e7 (2022).
12. Mouro Pinto, R. *et al.* In vivo CRISPR-Cas9 genome editing in mice identifies genetic modifiers of somatic CAG repeat instability in Huntington's disease. *Nat Genet* **57**, 314-322 (2025).
13. Wang, N. *et al.* Distinct mismatch-repair complex genes set neuronal CAG-repeat expansion rate to drive selective pathogenesis in HD mice. *Cell* (2025).
14. Horvath, S. *et al.* Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging (Albany NY)* **8**, 1485-512 (2016).
15. Langfelder, P. *et al.* Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat Neurosci* **19**, 623-33 (2016).
16. Lee, H. *et al.* Cell Type-Specific Transcriptomics Reveals that Mutant Huntingtin Leads to Mitochondrial RNA Release and Neuronal Innate Immune Activation. *Neuron* **107**, 891-908 e8 (2020).
17. Achour, M. *et al.* Neuronal identity genes regulated by super-enhancers are preferentially down-regulated in the striatum of Huntington's disease mice. *Hum Mol Genet* **24**, 3481-96 (2015).

18. Merienne, N. *et al.* Cell-Type-Specific Gene Expression Profiling in Adult Mouse Brain Reveals Normal and Disease-State Signatures. *Cell Rep* **26**, 2477-2493 e9 (2019).
19. Brule, B. *et al.* Accelerated epigenetic aging in Huntington's disease involves polycomb repressive complex 1. *Nat Commun* **16**, 1550 (2025).
20. Alcala-Vida, R. *et al.* Age-related and disease locus-specific mechanisms contribute to early remodelling of chromatin structure in Huntington's disease mice. *Nat Commun* **12**, 364 (2021).
21. Agbleke, A.A. *et al.* Advances in Chromatin and Chromosome Research: Perspectives from Multiple Fields. *Mol Cell* **79**, 881-901 (2020).
22. Li, Y.E. *et al.* An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129-136 (2021).
23. Granja, J.M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021).
24. Justice, J.L. *et al.* Multi-epitope immunocapture of huntingtin reveals striatum-selective molecular signatures. *Mol Syst Biol* (2025).
25. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19 (2013).
26. Escartin, C. *et al.* Reactive astrocyte nomenclature, definitions, and future directions. *Nat Neurosci* **24**, 312-325 (2021).
27. Khakh, B.S. & Goldman, S.A. Astrocytic contributions to Huntington's disease pathophysiology. *Ann N Y Acad Sci* **1522**, 42-59 (2023).
28. Wang, D. *et al.* Active DNA demethylation promotes cell fate specification and the DNA damage response. *Science* **378**, 983-989 (2022).
29. Hon, G.C. *et al.* 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell* **56**, 286-297 (2014).
30. MacArthur, I.C. *et al.* Developmental DNA demethylation is a determinant of neural stem cell identity and gliogenic competence. *Sci Adv* **10**, eado5424 (2024).
31. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**(2017).
32. Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274-1285 (2014).
33. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-5 (2009).
34. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-7 (2011).
35. Sperlazza, M.J., Bilinovich, S.M., Sinanan, L.M., Javier, F.R. & Williams, D.C., Jr. Structural Basis of MeCP2 Distribution on Non-CpG Methylated and Hydroxymethylated DNA. *J Mol Biol* **429**, 1581-1594 (2017).
36. Gabel, H.W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89-93 (2015).
37. Brain Initiative Cell Census Network. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86-102 (2021).

38. Dragileva, E. *et al.* Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol Dis* **33**, 37-47 (2009).
39. Kovalenko, M. *et al.* Msh2 acts in medium-spiny striatal neurons as an enhancer of CAG instability and mutant huntingtin phenotypes in Huntington's disease knock-in mice. *PLoS One* **7**, e44273 (2012).
40. Loupe, J.M. *et al.* Promotion of somatic CAG repeat expansion by Fan1 knock-out in Huntington's disease knock-in mice is blocked by Mlh1 knock-out. *Hum Mol Genet* **29**, 3044-3053 (2020).
41. Cui, X. *et al.* Comparative characterization of human accelerated regions in neurons. *Nature* **640**, 991-999 (2025).
42. Wilkes, F.A. *et al.* Hippocampal morphology in Huntington's disease, implications for plasticity and pathogenesis: The IMAGE-HD study. *Psychiatry Res Neuroimaging* **335**, 111694 (2023).
43. Liu, L. *et al.* Probabilistic and machine-learning methods for predicting local rates of transcription elongation from nascent RNA sequencing data. *Nucleic Acids Res* **53**(2025).
44. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* **54**, 412-436 (2022).
45. Diggle, C.P. *et al.* Biallelic Mutations in PDE10A Lead to Loss of Striatal PDE10A and a Hyperkinetic Movement Disorder with Onset in Infancy. *Am J Hum Genet* **98**, 735-43 (2016).
46. Appenzeller, S. *et al.* Autosomal-dominant striatal degeneration is caused by a mutation in the phosphodiesterase 8B gene. *Am J Hum Genet* **86**, 83-7 (2010).
47. Azuma, R. *et al.* A novel mutation of PDE8B Gene in a Japanese family with autosomal-dominant striatal degeneration. *Mov Disord* **30**, 1964-7 (2015).
48. Kumakura, A., Takahashi, S., Okajima, K. & Hata, D. A haploinsufficiency of FOXG1 identified in a boy with congenital variant of Rett syndrome. *Brain Dev* **36**, 725-9 (2014).
49. Mitter, D. *et al.* FOXG1 syndrome: genotype-phenotype association in 83 patients with FOXG1 variants. *Genet Med* **20**, 98-108 (2018).
50. Charlesworth, G. *et al.* Mutations in ANO3 cause dominant craniocervical dystonia: ion channel implicated in pathogenesis. *Am J Hum Genet* **91**, 1041-50 (2012).
51. Rahimi, M.J. *et al.* De novo variants in ATP2B1 lead to neurodevelopmental delay. *Am J Hum Genet* **109**, 944-952 (2022).
52. Srivastava, S. *et al.* Loss-of-function variants in HIVEP2 are a cause of intellectual disability. *Eur J Hum Genet* **24**, 556-61 (2016).
53. Calabresi, P., Picconi, B., Tozzi, A., Ghiglieri, V. & Di Filippo, M. Direct and indirect pathways of basal ganglia: a critical reappraisal. *Nat Neurosci* **17**, 1022-30 (2014).
54. Bibb, J.A. *et al.* Severe deficiencies in dopamine signaling in presymptomatic Huntington's disease mice. *Proc Natl Acad Sci U S A* **97**, 6809-14 (2000).
55. Johnson, B.S. *et al.* Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. *Nat Med* **23**, 1203-1214 (2017).

56. Amir, R.E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**, 185-8 (1999).
57. He, L.J. *et al.* Conditional deletion of Mecp2 in parvalbumin-expressing GABAergic cells results in the absence of critical period plasticity. *Nat Commun* **5**, 5036 (2014).
58. Ito-Ishida, A., Ure, K., Chen, H., Swann, J.W. & Zoghbi, H.Y. Loss of MeCP2 in Parvalbumin-and Somatostatin-Expressing Neurons in Mice Leads to Distinct Rett Syndrome-like Phenotypes. *Neuron* **88**, 651-8 (2015).
59. Tejwani, L. & Lim, J. Pathogenic mechanisms underlying spinocerebellar ataxia type 1. *Cell Mol Life Sci* **77**, 4015-4029 (2020).
60. Sjostedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**(2020).
61. Ferguson, R. & Tabrizi, S.J. Can MSH3 lowering stop HTT repeat expansion in its CAG tract? *Mol Ther* **31**, 1509-1511 (2023).
62. Pressl, C. *et al.* Isolation and Molecular Profiling of Nuclei of Specific Neuronal Types from Human Cerebral Cortex and Striatum. *Curr Protoc* **4**, e70067 (2024).
63. Love, M.I., Hogenesch, J.B. & Irizarry, R.A. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nat Biotechnol* **34**, 1287-1291 (2016).
64. Love, M.I., Soneson, C. & Patro, R. Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. *F1000Res* **7**, 952 (2018).
65. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).
66. Krueger, F. Trim galore. *A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files* **516**, 517 (2015).
67. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-8 (2016).
68. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2 (2011).
69. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
70. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
71. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* **13**, R87 (2012).
72. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
73. Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* **8**, e81148 (2013).
74. Qu, J., Zhou, M., Song, Q., Hong, E.E. & Smith, A.D. MLML: consistent simultaneous estimates of DNA methylation and hydroxymethylation. *Bioinformatics* **29**, 2645-6 (2013).

75. Matlik, K., Abo-Ramadan, U., Harvey, B.K., Arumae, U. & Airavaara, M. AAV-mediated targeting of gene expression to the peri-infarct region in rat cortical stroke model. *J Neurosci Methods* **236**, 107-13 (2014).
76. Murakami, T.C. *et al.* Open-source Photobleacher for Fluorescent Imaging of Large Pigment-Rich Tissues. *bioRxiv*, 2025.02.24.639965 (2025).
77. Zou, Y. *et al.* ATG8 delipidation is not universally critical for autophagy in plants. *Nat Commun* **16**, 403 (2025).
78. Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600-604 (2017).
79. Liu, H. *et al.* DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598**, 120-128 (2021).
80. Goldstrohm, A.C., Albrecht, T.R., Sune, C., Bedford, M.T. & Garcia-Blanco, M.A. The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol Cell Biol* **21**, 7617-28 (2001).
81. Holbert, S. *et al.* The Gln-Ala repeat transcriptional activator CA150 interacts with huntingtin: neuropathologic and genetic evidence for a role in Huntington's disease pathogenesis. *Proc Natl Acad Sci U S A* **98**, 1811-6 (2001).