Extensive changes in DNA methylation are associated with expression of mutant huntingtin

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Extensive changes in DNA methylation are associated with expression of mutant huntingtin

Christopher W. Ng*, Ferah Yildirim*, Yoon Sing Yap*, Simona Dalin*, Bryan J. Matthews*, Patricio J. Velez*, Adam Labadorf*, David E. Housman*,1 and Ernest Fraenkel*.1

Departments of *Biological Engineering and 1Biology, Massachusetts Institute of Technology, Cambridge, MA 02142

Contributed by David E. Housman, December 6, 2012 (sent for review November 16, 2012)

The earliest stages of Huntington disease are marked by changes in gene expression that are caused in an indirect and poorly understood manner by polyglutamine expansions in the huntingtin (HTT) protein. To explore the hypothesis that DNA methylation may be altered in cells expressing mutated HTT, we use reduced representation bisulphite sequencing (RRBS) to map sites of DNA methylation in cells carrying either wild-type or mutant HTT. We find that a large fraction of the genes that change in expression in the presence of mutant huntingtin demonstrate significant changes in DNA methylation. Regions with low CpG content, which have previously been shown to undergo methylation changes in response to neuronal activity, are disproportionately affected. On the basis of the sequence of regions that change in methylation, we identify AP-1 and SOX2 as transcriptional regulators associated with DNA methylation changes, and we confirm these hypotheses using genome-wide chromatin immunoprecipitation sequencing (ChIP-Seq). Our findings suggest new mechanisms for the effects of polyglutamine-expanded HTT. These results also raise important questions about the potential effects of changes in DNA methylation on neurogenesis and cognitive decline in patients with Huntington disease.

Huntington disease (HD) is a fatal, neurodegenerative disorder caused by an expanded CAG repeat in the huntingtin (HTT) gene, which encodes an abnormally long polyglutamine repeat in the HTT protein. In the early stages of the disease, patients are largely asymptomatic, although they may suffer from mild cognitive impairment and behavioral changes. With time, they develop severe motor dysfunction as well as more pronounced cognitive and psychiatric symptoms. Transcriptional dysregulation is a major component of the early stages of HD, before significant neuronal death. Changes in transcription have been detected in human postmortem tissue (1), mouse models (2–4), and cell culture models (5). Genes reproducibly shown to be affected in HD include NRSF/REST, CBP, PGC1α, Spl1, BCL11b, p53, LXRα, polycomb-group proteins, SIN3A, and NCO1 (6). Any effect of HTT on the subcellular localization, activity, or concentration of these proteins would be likely to directly change gene expression. Two lines of evidence suggest that alterations in HTT could also influence DNA methylation. First, histone marks are altered in HD (7–11) and SETD2, a SET domain regulator of H3K36me3, has been previously reported to interact with mutant HTT (12). Due to biochemical interactions between DNA methyltransferases and SET domain histone methyltransferases (13), changes in repressive histone marks could result in changes in DNA methylation. Second, recent findings show that DNA-bound transcription factors can alter DNA methylation patterns (14, 15), suggesting that the effects of HTT on DNA-binding proteins could also directly influence DNA methylation. For example, NRSF/REST (16) is a neuronal repressor implicated in HD (7) that recruits repressive complexes containing HDACs and the DNA methylation reader MECP2 (17), and changes to NRSF/REST binding could affect local DNA methylation levels. Because DNA methylation in the brain influences a wide range of cognitive functions (18–20), these epigenetic modifications could have long-term consequences on HD patients.

For our analysis, we used cell lines derived from mouse striatal neurons expressing full-length HTT with either the wild-type or expanded polyglutamine repeat (5). Derived from mouse embryos, these cells serve as a highly reproducible model for the earliest stages of the disease. Using reduced representation bisulphite sequencing (RRBS), we identified extensive changes in DNA methylation. Regions with low CpG content are overrepresented among those that change in methylation. These regions are largely intergenic and transition from fully methylated in wild-type cells to partially methylated in cells expressing mutant HTT. As partially methylated regions have been shown to be occupied by transcription factors that influence methylation (14), we analyzed the sequences of these sites to identify several potential DNA-binding proteins. Using genome-wide chromatin immunoprecipitation sequencing (ChIP-Seq), we confirmed that the binding of SOX2 and the AP-1 proteins FRA-2 and JUND are associated with changes in DNA methylation resulting from expression of mutant HTT.

For many years, DNA methylation was believed to be an extremely stable epigenetic mark relevant mostly to developmental processes and, by extension, to tumor development. However, it is becoming increasingly clear that DNA methylation has a role in a much wider range of biological processes, including neuronal activity, learning, and memory (18–22). Our analysis reveals that polyglutamine-expanded HTT causes changes in DNA methylation at both promoter proximal and distal regulatory regions. These results shed new light on the etiology of HD and suggest a possible explanation for mild cognitive impairment that occurs at the early stages of the disease, before significant loss of neurons.

Results

Expression of Mutant huntingtin Is Associated with Large-Scale Changes in DNA Methylation. On the basis of the importance of changes in DNA methylation to development (23–26) and neuronal activity (22), we sought to determine whether mutant HTT might perturb this epigenetic mark. To examine this hypothesis, we used RRBS to map DNA methylation at base pair resolution in striatal cells carrying polyglutamine-expanded HTT (STHdhQ111/Q111) and wild-type cells (STHdhQ7/Q7) (see Dataset S1 for full list of datasets in this study). We measured the degree of methylation


The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE43438).

1To whom correspondence may be addressed. E-mail: fraenkel-admin@mit.edu or dhousman@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221292110/-/DCSupplemental.
Changes in DNA Methylation. We used ChIP-Seq experiments (Dataset S1) to test whether FRA-2, JUND, and SOX2 were, in fact, bound to the sites where DNA methylation is altered between wild-type and mutant cells. We had previously noted in our mRNA-Seq data that SOX2 mRNA is expressed at much higher levels in the wild-type cells. Additional analysis of methylated regions (LMRs) indicated that these regions are enriched in regulatory sites (14). We therefore focused our analysis on these regions. We used three separate methods for sequence analysis to identify the highest-confidence motifs (details in Materials and Methods). These three methods identified a small set of common sequence motifs (Table S1 and Dataset S3), associated with the CREB/ATF, AP-1, SOX, and ETS families. By examining the expression of members of these families in each cell line, we selected FRA-2, JUND, and SOX2 for further analysis (Fig. S3A and Dataset S4).

ChIP-Seq Confirms Binding of FRA-2, JUND, and SOX2 to Sites with Changes in DNA Methylation. We used ChIP-Seq experiments (Dataset S1) to test whether FRA-2, JUND, and SOX2 were, in fact, bound to the sites where DNA methylation is altered between wild-type and mutant cells. We had previously noted in our mRNA-Seq data that SOX2 mRNA is expressed at much higher levels in the wild-type cells. Additional analysis of methylated regions (LMRs) indicated that these regions are enriched in regulatory sites (14). We therefore focused our analysis on these regions. We used three separate methods for sequence analysis to identify the highest-confidence motifs (details in Materials and Methods). These three methods identified a small set of common sequence motifs (Table S1 and Dataset S3), associated with the CREB/ATF, AP-1, SOX, and ETS families. By examining the expression of members of these families in each cell line, we selected FRA-2, JUND, and SOX2 for further analysis (Fig. S3A and Dataset S4).
levels in STHdhQ7 cells, and Western blotting demonstrated that the SOX2 protein is almost undetectable in STHdhQ111 cells (Fig. S3B). Using ChIP-Seq in STHdhQ7 cells, we found that SOX2 bound 6,476 genes, which were associated with a wide range of biological processes, including regulation of cell proliferation, metabolic processes, cell death, signaling, and transcriptional regulation (Dataset S2). The SOX motif was highly enriched within the bound sites (P < 1e-100; Dataset S3). Sites bound by FRA-2 and JUND, which form the AP-1 complex, were largely the same (83% in STHdhQ7 and 79% in STHdhQ111) and enriched for the AP-1 motif (P < 1e-200; Dataset S3). FRA-2- and JUND-bound genes were associated with a range of biological processes, including regulation of cell death, signal transduction, developmental process, transcription, phosphorylation, and cell migration in both cell lines. Although there was FRA-2 and JUND binding in both cell lines, we observed many sites significantly changing in their degree of binding between the cell lines (FRA-2, 10,596 sites more bound in STHdhQ7 and 14,079 more bound in STHdhQ111; JUND, 8,393 sites more bound in STHdhQ7 and 7,344 more bound in STHdhQ111). For both FRA-2 and JUND, genes more bound in STHdhQ7 were enriched for the following Gene Ontology (GO) terms: regulation of neurogenesis, neuron differentiation, and G-protein–coupled receptor signaling. Genes more bound in STHdhQ111 were enriched for the following GO terms: innate immune response–activating cell surface receptor signaling pathway, regulation of defense response, and G-protein–coupled receptor signaling.

Notably, a majority of the FRA-2–, JUND–, and SOX2-bound sites in STHdh cells were distal regulatory regions (greater than 50% were 20 kb or more from a transcriptional start site (TSS) and greater than 80% were 2 kb or more from a TSS for all individual datasets) and CpG-poor regions (greater than 70% of binding sites for all individual datasets). To test our hypothesis that the binding of each factor was associated with changes in methylation at these CpG-poor regions, we examined the extent of methylation changes at regions bound by each protein (Fig. 3). SOX2 is detectable only in STHdhQ7 cells, and the regions bound in these cells had higher methylation than in STHdhQ111 cells (P < 1e-14). For example, regions surrounding the Pax6 and Nes genes are bound by SOX2 and unmethylated or partially methylated in wild-type cells. However, in STHdhQ111 cells, where SOX2 is not expressed, the same regions are partially methylated or fully methylated (Fig. 4). Interestingly, we find that SOX2-bound sites are enriched for the AP-1 motif (P < 1e-37; Dataset S4) and overlap with FRA-2- and JUND-bound sites (Fig. S3C). In addition, regions bound by SOX2 and bound more by FRA-2 and JUND in STHdhQ7 had higher methylation than in STHdhQ111 cells (P < 1e-16; Fig. 3), including the aforementioned Pax6 and Nes regulatory regions (Fig. 4).

FRA-2 and JUND bind to DNA in both cell types and changes in their binding are strongly associated with changes in DNA methylation (P < 1e-20; Fig. 3). Regions that show increased binding by these two proteins in one cell type also show decreased methylation in those cells. For example, both proteins bind more strongly to sites near the microRNA-143/microRNA-145 (miR-143/miR-145) cluster in STHdhQ111 cells and the nearest methylation site goes from being 85% methylated in wild-type cells to only 3% methylated in STHdhQ111 cells. Similarly, STHdhQ111-specific binding of these proteins near Fbn1 in STHdhQ111 causes a drop from 84% methylation to 55% methylation at an adjacent region (Fig. S4).

**Discussion**

Our study demonstrates significant changes in DNA methylation in cells expressing polyglutamine-expanded HTT. We showed that...
these changes are gene specific, with some sites increasing in methylation and others decreasing. Our results fit with a growing literature indicating that DNA methylation is not, as was previously thought, a stable epigenetic mark. Rather, DNA methylation changes dynamically in response to cellular processes, with potential long-term consequences for cellular behavior.

**Regulation of Changes in DNA Methylation.** What causes the widespread changes in DNA methylation? The DNA methyltransferase Dnmt1 is highly expressed in both cell types, as is the Gadd45 family of genes, which is responsible for DNA demethylation. Interestingly, we find that Dnmt1 expression decreases twofold and Gadd45g expression increases more than fivefold in the presence of mutant HTT (Dataset S4). In addition, the presence of sequence motifs near CpG-poor regions changing in DNA methylation suggests that the gain or loss of DNA-binding proteins may lead to changes in DNA methylation at specific loci. These binding changes could result from the presence of mutant HTT, the absence of wild-type HTT, or both. We also note that mutant HTT has been shown to bind directly to DNA (28), raising the possibility that HTT could directly influence the recruitment of epigenetic modifiers to the genome. The most significant motifs belong to four families: CREB/ATF, AP-1, SOX, and ETS. Before this study, global methylation changes have not been linked to these motifs.

Fig. 4. *Pax6* and *Nes* regulatory regions are bound less by SOX2, FRA-2, and JUND and increase in DNA methylation and the genes decrease in expression in STHdhQ111 relative to STHdhQ7. (A and B) DNA methylation (STHdhQ7 and STHdhQ111), SOX2 ChIP-Seq (STHdhQ7 only), FRA-2 ChIP-Seq (STHdhQ7 and STHdhQ111), JUND ChIP-Seq (STHdhQ7 and STHdhQ111). mRNA-Seq (STHdhQ7 and STHdhQ111), RefSeq gene annotation, and University of California, Santa Cruz (UCSC) CpG island tracks around the *Pax6* (A) and *Nes* (B) loci. Data for DNA methylation are shown at the level of base pairs and regions (see Materials and Methods for definition of regions). The ball-and-stick plots illustrate RRBS read depth (y axis) and methylation fraction (marker color scale) at base pair resolution, whereas regions are depicted by colored boxes. For ChIP-Seq and mRNA-Seq tracks, the y axis indicates the number of reads.
proteins besides the recent association of ETS-binding sites and dynamic methylation in hematopoiesis (15). Using ChIP-Seq experiments, we are able to demonstrate that changes in binding of the AP-1 family members FRA-2 and JUND and the SOX2 protein are strongly associated with changes in DNA methylation. In particular, loss of binding is associated with increased DNA methylation.

Some members of these families have been previously associated with neurodegeneration. In particular, mice heterozygous for a Sox2 knockout exhibited epileptic and neurodegenerative phenotypes (29, 30) and SOX2 mutations in humans are associated with neurological phenotypes including seizures (31). In addition, huntingtin knockin mice show altered levels of SOX2 protein and impaired striatal neurogenesis during brain maturation (32). In PC12 cells, AP-1 transcription factors were shown to govern the choice between neuronal differentiation and apoptosis (33). Long-term induction of AP-1 and in particular, FRA-2, occurs in the hippocampus and cortex in several mouse and rat models of brain injury (34, 35) and NMDA activation. In addition, neuronal expression of AP-1 proteins has been linked to excitotoxic neurodegenerative disorders (36).

Along with colocalized changes in DNA methylation, we observe loss of SOX2 and diminished binding of FRA-2 and JUND to neurogenesis and neuronal differentiation genes such as Pax6 and Nes in this early-stage model of HD. Conversely, binding of FRA-2 and JUND increases adjacent to genes associated with stress response, implicating a pathogenic switch in AP-1 regulation in cells with mutant HTT. In addition, FRA-2 and JUND binding increases at a regulatory region with methylation changes near the miR-143/miR-145 cluster. MiR-145 has been previously shown to be elevated in STHdhQ111 relative to STHdhQ7 (37) and to repress pluripotency in a human embryonic stem cell model by targeting OCT4, SOX2, and KLF4 (38). Many of the dysregulated genes we report including Sox2, Pax6, and Fru2 have also been recently reported to be altered between induced pluripotent stem cell (iPSC) lines generated from patients with HD and controls (39).

Mutant huntingtin’s Effects on Neurogenesis. Genes associated with neurogenesis and neuronal differentiation such as Sox2, Pax6, and Nes were expressed at lower levels and show increased DNA methylation in cells expressing polyQ-expanded HTT. This finding is in line with the recent literature that perturbed neurogenesis is involved in the pathogenesis of HD (40). Increased neurogenesis in the SVZ has been observed in postmortem HD brain and mouse models (41–44). These findings may suggest a delayed and insufficient endogenous reaction of brain tissue to counter the severe neuropathology. Apart from these studies, there is a strong body of evidence demonstrating impaired hippocampal neurogenesis in a variety of well-established animal models of HD (45–49). Crucially, several agents that potentiated hippocampal or SVZ neurogenesis proved highly beneficial in HD mouse models (48, 50–53), suggesting neurogenesis enhancement as a promising therapeutic candidate against HD progression.

DNA Methylation and Neuronal Activity. Recent studies have shown that DNA methylation has roles in both disease and normal neural differentiation. Once considered to be a stable epigenetic mark that was primarily altered in development and cancer, DNA methylation has now been linked to a wide variety of diseases, including insulin resistance, cardiovascular disease, and neurological disorders (54–56). In fact, even normal neuronal activity causes rapid changes to DNA methylation. For instance, the Bdnf gene undergoes differential DNA methylation in response to a wide variety of behavioral stimuli, including early life stress and contextual fear conditioning (57, 58). Although the full mechanistic details of behavior-induced methylation changes remain to be discovered, activity-induced changes in DNA methylation at individual loci have been shown to depend on DNA methyltransferases (59) and Cadd45a (60).

Our observation that expression of mutant HTT and/or loss of wild-type HTT are associated with DNA methylation changes in a striatal cell line raises the possibility that HTT might also be associated with physiological changes in DNA methylation in neurons. By impairing the normal regulation of DNA methylation, HTT could cause neuronal dysfunction long before any signs of neuronal death. In fact, recent studies show that mild cognitive impairment can be detected in 40% of premanifest patients with Huntington disease (61). It will therefore be important to determine whether changes in DNA methylation cause changes in gene expression changes or the reverse. Such studies could reveal whether pharmacological treatments targeting the potential effects of HTT on DNA methylation might have the potential to slow or halt the progression of the earliest stages of HD.

Materials and Methods

Cell Culture. STHdhQ7 and STHdhQ111 cell lines were cultured as described previously (5). To stop cell division and mitigate cell-cycle differences, the culture is maintained at 33 °C and raised to 39 °C for 2 d before each experiment.

RRBS. RRBS libraries were constructed according to a previously published protocol (62). Bismark (63) was used for read alignment and methylation calls within the short reads. CpGs with sufficient coverage in both cell lines (≥50 reads in each cell line) across the genome were assembled into methylation regions with a greedy approach that combined all adjacent CpGs with nearest neighbors less than 100 bp apart. Methylated and nonmethylated counts across all CpGs within each region were summed for each cell line. For each methylation region, statistical significance of differential methylation was calculated using a Fisher’s exact test on a 2 × 2 contingency table of methylated and nonmethylated counts in the two cell lines. Multiple-hypothesis correction was applied using the Benjamini–Hochberg procedure. DMRs for downstream analysis were identified using a final cutoff of P < 1e-5. Additionally, we required DMRs to have at least 15% differential methylation. Differential methylation was quantified as the difference in overall methylation fraction between cell lines for each methylation region. Detailed methods are provided in SI Materials and Methods.

ChIP-Seq Assay. Cells were cross-linked with a 1% formaldehyde solution for 10 min and the cross-linking was quenched by adding glycine to a final concentration of 0.125 M. The cross-linked material was then rinsed with 1× PBS, pelleted, and frozen in liquid nitrogen for later use. ChIP-Seq assays were performed as previously described (67). Details of the ChIP-Seq experiments and its computational analysis are provided in SI Materials and Methods.

Motif Scanning. We used three separate methods to identify motifs associated with changes in DNA methylation in CpG-poor regions: (i) overrepresentation of motifs using the Mann–Whitney–Wilcoxon ranked sum test, (ii) overrepresentation of motifs using THEMGE (68), and (iii) prediction of differential methylation using motif scores in a regression-based approach. In each of these methods we began with a set of motifs derived from all vertebrate position-specific scoring matrices (PSSMs) from TRANSFAC, filtered for sufficient information content [information content (IC) > 8 total bits]. As many of these motifs are redundant, we clustered them on the basis of pairwise distance by KL-divergence of the PSSMs, using Affinity Propagation. The TAMO programming environment (69) was used to store the PSSMs and score sequences. Details of these approaches are provided in SI Materials and Methods.

ACKNOWLEDGMENTS. We acknowledge Marcy MacDonald (Massachusetts General Hospital, Boston) and Coriell Institute (Camden, NJ) for the striatal cell lines and Candace Chouinard [Massachusetts Institute of Technology (MIT)] and the MIT BioMicro Center for assistance. This work was supported by grants from the National Institutes of Health (U54CA112967 and R01GM089903) and the Hereditary Disease Foundation. Computing resources funded by the National Science Foundation under Award DB1-D821391 were used for this work.


