

RESEARCH ARTICLE

The Role of H3K4me3 in Transcriptional Regulation Is Altered in Huntington's Disease

Xianjun Dong^{1☯✉}, Junko Tsuji^{1☯}, Adam Labadorf^{2,3}, Panos Roussos^{4,5}, Jiang-Fan Chen^{2‡}, Richard H. Myers^{2,3,6‡}, Schahram Akbarian^{4‡}, Zhiping Weng^{1‡*}

1 Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA, United States of America, **2** Department of Neurology, Boston University School of Medicine, Boston, MA, United States of America, **3** Bioinformatics Program, Boston University, Boston, MA, United States of America, **4** Friedman Brain Institute, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, United States of America, **5** Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY, United States of America, **6** Genome Science Institute, Boston University School of Medicine, Boston, MA, United States of America

☯ These authors contributed equally to this work.

✉ Current address: Laboratory for Neurogenomics, Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States of America

‡ These authors also contributed equally to this work.

* zhiping.weng@umassmed.edu



OPEN ACCESS

Citation: Dong X, Tsuji J, Labadorf A, Roussos P, Chen J-F, Myers RH, et al. (2015) The Role of H3K4me3 in Transcriptional Regulation Is Altered in Huntington's Disease. PLoS ONE 10(12): e0144398. doi:10.1371/journal.pone.0144398

Editor: Axel Imhof, Ludwig-Maximilians-Universität München, GERMANY

Received: July 17, 2015

Accepted: November 17, 2015

Published: December 4, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The ChIP-seq data for six Huntington's disease patients and six controls from this study have been submitted to GEO (<http://www.ncbi.nlm.nih.gov/geo/>) under the accession number of GSE68952.

Funding: This work was supported by the Jerry McDonald HD Research Fund and Public Health Service, National Institutes of Health grants, National Institute of Neurological disorders and stroke (R01 NS073947; Epigenetic Markers in Huntington's disease Brain) National Institute of Mental Health (R01 MH106056; Higher Order Chromatin and Genetic Risk for Schizophrenia). The authors would

Abstract

Huntington's disease (HD) is an autosomal-dominant neurodegenerative disorder resulting from expansion of CAG repeats in the Huntingtin (*HTT*) gene. Previous studies have shown mutant *HTT* can alter expression of genes associated with dysregulated epigenetic modifications. One of the most widely studied chromatin modifications is trimethylated lysine 4 of histone 3 (H3K4me3). Here, we conducted the first comprehensive study of H3K4me3 ChIP-sequencing in neuronal chromatin from the prefrontal cortex of six HD cases and six non-neurologic controls, and its association with gene expression measured by RNA-sequencing. We detected 2,830 differentially enriched H3K4me3 peaks between HD and controls, with 55% of them down-regulated in HD. Although H3K4me3 signals are expected to be associated with mRNA levels, we found an unexpected discordance between altered H3K4me3 peaks and mRNA levels. Gene ontology (GO) term enrichment analysis of the genes with differential H3K4me3 peaks, revealed statistically significantly enriched GO terms only in the genes with down-regulated signals in HD. The most frequently implicated biological process terms are organ morphogenesis and positive regulation of gene expression. More than 9,000 H3K4me3 peaks were located not near any recognized transcription start sites and approximately 36% of these "distal" peaks co-localized to known enhancer sites. Six transcription factors and chromatin remodelers are differentially enriched in HD H3K4me3 distal peaks, including EZH2 and SUZ12, two core subunits of the polycomb repressive complex 2 (PRC2). Moreover, PRC2 repressive state was significantly depleted in HD-enriched peaks, suggesting the epigenetic role of PRC2 inhibition associated with up-regulated H3K4me3 in Huntington's disease. In summary, our study provides new insights into transcriptional dysregulation of Huntington's disease by analyzing the differentiation of H3K4me3 enrichment.

like to acknowledge the Harvard Brain Tissue Resource Center McLean Hospital, Belmont, Massachusetts for providing the HD tissue. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Huntington's disease (HD) is an autosomal-dominant neurodegenerative disorder characterized by abnormal involuntary choreiform movements, cognitive impairment, and psychiatric dysfunction [1,2]. Neuropathologically, HD patients exhibit neuronal cell loss and gliosis, primarily in the striatum, but also involving the cerebral cortex and other brain regions [3,4]. Although there are currently no effective treatments for persons afflicted with HD, recent trials of creatine monohydrate among asymptomatic HD gene carriers suggest it may have potential to delay the disease onset [5]. The cause of HD is an expansion in the number of CAG trinucleotide repeats in the coding region of exon 1 of the Huntingtin (*HTT*) gene [6]. In comparison to the normal *HTT* allele, which ranges from 8 to 35 triplet repeats, mutant *HTT* (*mHTT*) contains 36 or more repeats resulting in an expanded polyglutamine tract [6]. Repeats between 36 and 39 units show reduced penetrance [7,8].

The toxic *mHTT* expansion induces numerous and widespread aberrant molecular effects. Transcriptional dysregulation has been proposed to be a central component of HD pathogenesis [9,10]. Altered gene expression has been reported [11] and several studies support alterations at one or more of the stages of RNA processing, translation, protein post-translational modification and trafficking [12–14]. Previous studies show that mHTT protein physically binds CREB-binding protein (CBP) and blocks both CBP's transcription coactivator function in human and mice [15] and its histone acetyltransferase activity in *Drosophila* [16]. Studies in mouse models of HD further illustrate that transcriptional dysfunction is associated with histone hypoacetylation [17,18]. Elevated expression of ERG-associated protein with SET domain (ESET) in HD patients and in the R6/2 transgenic mouse model increases abnormal trimethylation of histone H3 lysine 9 (H3K9me3) [19]. The decreased expression of acetylcholine receptor 1 (CHRM1) in HD also affects H3K9me3-mediated chromatin remodeling [20]. Interestingly, the gene expression changes in HD are usually associated with dysregulated epigenetic modifications [21–23].

One of the most widely studied histone modifications is trimethylated lysine 4 of histone 3 (H3K4me3) [24,25]. H3K4me3 marks active transcription start sites (TSSs) [26–28] and the strength of H3K4me3 signal at the promoter is strongly correlated with the expression of the gene [29,30]. Recently, studies in the R6/2 HD transgenic mouse model revealed that H3K4me3 may play a critical role in the pathway leading to transcriptional dysregulation in HD [31]. Our earlier study found that epigenetic changes in the *HES4* gene are associated with reduced H3K4me3 and increased DNA methylation in its promoter, and showed that these changes were closely correlated with degeneration in the striatum of HD patients [32]. These previous findings motivated us to investigate the genome-wide pattern of H3K4me3 in HD. In the current study, we conducted the first chromatin immunoprecipitation using anti-H3K4me3 antibody followed by sequencing (ChIP-seq) experiments on NeuN-selected neuronal cell nuclei from post-mortem prefrontal cortical samples for six HD cases and six non-neurologic controls.

Results

H3K4me3 modification enrichment characteristics

We analyzed H3K4me3 enrichment in FACS sorted neuronal nuclei derived from postmortem prefrontal cortex (BA9) of six HD cases and six non-neurological controls (Table 1). A total of 28,608 H3K4me3 peaks were defined when the filtered peaks called from HD and control samples were combined (S1 Fig). We found that most peaks (59.2%) were called in all twelve samples (S2A Fig) and on average 63% of total H3K4me3 ChIP-seq reads were mapped to TSS-

Table 1. Characteristics of human brain samples in the study.

Diagnosis	ID*	Sex	CAG Repeat	Onset age	Age at death	PMI*	Antibody	Total reads	Uniquely mapped reads	Mappability (%)	Peaks
HD	H0001	M	45	44	55	37.25	H3K4me3	6,134,681	5,206,490	84.90%	27,661
HD	H0003	M	43	52	71	20.5	H3K4me3	3,292,319	2,930,762	89.00%	24,955
HD	H0008	M	49	28	43	21.3	H3K4me3	6,647,419	5,979,404	90.00%	29,710
HD	H0009	M	42	45	68	3.73	H3K4me3	2,817,246	2,540,805	90.20%	29,640
HD	H0681	M	42	50	69	19.06	H3K4me3	5,567,462	4,858,701	87.30%	27,732
HD	H0601	F	45	unknown	56	19.0	H3K4me3	5,943,156	5,282,199	88.90%	30,647
Control	C1	M	N/A	N/A	55	17	H3K4me3	6,028,542	4,927,721	81.70%	40,150
Control	C2	M	N/A	N/A	63	24.5	H3K4me3	8,356,841	7,519,896	90.00%	30,434
Control	C3	M	N/A	N/A	64	28.21	H3K4me3	13,238,359	11,132,414	84.10%	27,419
Control	C4	M	N/A	N/A	74	12	H3K4me3	3,091,666	2,581,032	83.50%	39,323
Control	C5	M	N/A	N/A	81	8	H3K4me3	10,313,831	7,955,599	77.10%	32,907
Control	C6	F	N/A	N/A	68	7	H3K4me3	8,287,623	5,993,036	72.30%	38,970
Control	C7	F	N/A	N/A	69	40.15	Input	15,875,168	12,246,790	77.10%	1,616

*IDs match those for the HD samples that we published previously [12,36] (Labadorf et al. submitted)

*PMI: postmortem interval.

doi:10.1371/journal.pone.0144398.t001

proximal peaks. These are consistent with our prior data showing that H3K4me3 is enriched in promoters and similar across different individuals [33–35]. The H3K4me3 signal tends to be stronger when the peak is called in more samples (S2B Fig).

Most H3K4me3 peaks (n = 18,836, 65.8%) were located proximally (within 1 kb) to a TSS and most (93.5%) were shared between HD and control, with 6.5% peaks (n = 1,850) unique to HD or control (Fig 1A). Proximal peaks were positioned with a slight upstream position bias relative to TSSs, in both control and HD samples (Fig 1B), which was seen for both common and uniquely defined proximal peaks. The 9,772 distal peaks (>1 kb from TSS) consist of 5,671 (58.0%) are from intragenic regions of annotated genes and 4,102 (42.0%) from intergenic regions. Both proximal and distal peaks were confidently mapped with strong read depth (Fig 1D).

Among proximal peaks, the distribution of peak widths was bimodal with the two modes being 350 bp and 1700 bp (Fig 1C). The proximal peaks unique to HD are slightly larger (t-test p-value = 3.28×10^{-14}) than those unique to controls (400bp versus 300bp respectively). The mode for distal peaks is 350bp, similar to the mode for the smaller proximal peaks.

Differential H3K4me3 between HD and control

To investigate the transcript specific differences in H3K4me3 signal, we extracted H3K4me3 reads within a 2000 bp window centered around every annotated TSS (referred to as promoters hereafter) regardless of whether a peak was called within that region or not. This strategy permits an assessment of the relationship of H3K4me3 signal to mRNA-seq abundance (see Method). As a result, we detected 720 genes with significantly different H3K4me3 level in promoters between HD and control, including 104 genes with significantly higher H3K4me3 signal in HD and 616 genes with significantly higher H3K4me3 in controls (S3 Table). We also ran the differential analysis for all 9,772 distal peaks, and detected 209 and 51 distal peaks with significantly higher H3K4me3 signal in HD and control respectively (S4 Table).

Table 2 lists the 22 genes that have the most differential H3K4me3 signal at the proximal promoter regions between HD and control (q-value < 1×10^{-4}). Among them 15 genes have

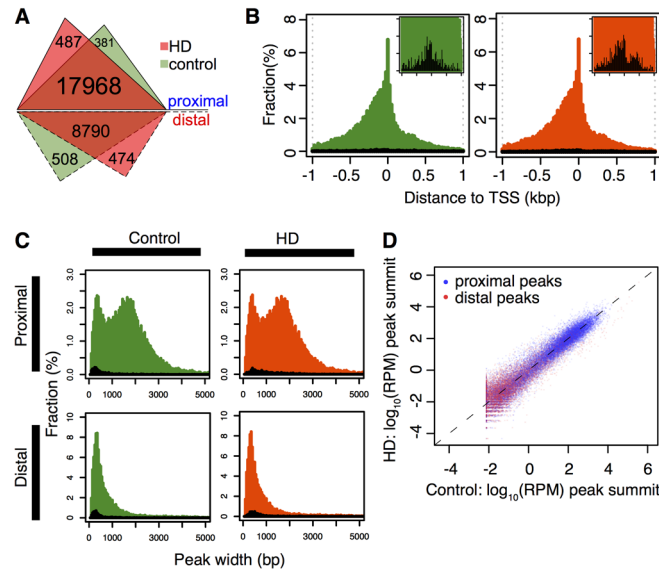


Fig 1. H3K4me3 proximal and distal peaks in HD and control. (A) The number of H3K4me3 peaks in HD and control. 487 proximal peaks were unique to HD samples while 381 were unique to controls. 474 distal peaks were unique to HD samples and 508 were unique to controls. (B) Histograms of the position of proximal peaks relative to the TSS in HD and control show a slight enrichment upstream the TSS. Small right panels show zoomed-in histograms for unique proximal peaks (highlighted in black). (C) Histograms of the distribution of peak widths for proximal and distal peaks were similar in control and HD samples. Proximal peaks show a bimodal distribution in peak width not seen in distal peaks. The width distributions of unique peaks are in black. (D) Peak intensities of proximal (blue) and distal (red) peaks are plotted for HD versus control.

doi:10.1371/journal.pone.0144398.g001

lower H3K4me3 in HD cases than in controls. *HES4*, which we previously reported [32], is the fourth protein-coding gene in this list. The top gene *ALG14* catalyzes asparagine (N)-glycosylation, which is an essential modification that regulates protein folding and stability.

Proximal H3K4me3 signal is correlated with gene expression

We calculated the correlation between the signal at proximal H3K4me3 peaks and the expression level of the corresponding genes. Even though we performed both H3K4me3 ChIP-seq and RNA-seq experiments on brain tissue from the BA9 region of the prefrontal cortex, ChIP-seq was performed on sorted neuronal nuclei while RNA-seq was performed on brain tissue homogenate because it is not feasible to sort entire cells. Previously we showed that H3K4me3 levels at promoters were highly correlated with transcription levels of genes across multiple cell types [37]. We confirmed this pattern in both control and HD cases (Spearman's correlation coefficient $\rho = 0.57$, p -value $< 2.2 \times 10^{-16}$, see Fig 2A). In addition, because of the bimodality seen for both the mRNA expression (S3A Fig) and H3K4me3 level (S3B Fig) for both HD and control, we subdivided genes into high and low groups by gene expression and H3K4me3 level respectively. Applying a χ^2 test, we found that genes with high and low H3K4me3 levels strongly associate with high and low expression respectively (χ^2 p -value $< 2.2 \times 10^{-16}$; see Fig 2B for all samples). This relationship was seen in both controls and HD cases (S3C Fig).

We evaluated whether differentially expressed genes (DEGs) between HD and control also exhibit differential H3K4me3 levels at their promoters. Most genes are neither differentially expressed nor have differential H3K4me3 signal at the promoters. Among the 961 DEGs and the 720 genes with differential H3K4me3 signals, 58 genes were found to have both differential expression and H3K4me3 (blue dots in Fig 2C). No correlation was found between the fold

Table 2. Genes that correspond to the top 22 differential proximal H3K4me3 peaks.

Gene symbol	Gene type	log ₂ (HD/control)	q-value	Gene / Function
<i>ALG14</i>	protein coding	1.028	3.63E-09	UDP-GlcNAc transferase: regulates protein folding and stability
<i>SERTAD4-AS1</i>	Antisense	-1.075	1.02E-07	Antisense of SERTAD4: no known function
<i>LRRTM2</i>	protein coding	1.110	1.08E-06	leucine rich repeat transmembrane neuronal 2: development of excitatory synapse in CNS
<i>CTD-3064M3.4</i>	processed transcript	-1.091	1.96E-06	N/A
<i>FAM184B</i>	protein coding	-1.067	2.28E-06	No known function
<i>HES4</i>	protein coding	-1.506	1.30E-05	Hairy and enhancer of split 4: protein dimerization & transcription factor binding.
<i>CTB-31O20.8</i>	antisense	-1.347	1.43E-05	N/A
<i>TMEM61</i>	protein coding	-1.958	1.53E-05	Transmembrane Protein 61: no known function
<i>SNORD118</i>	snoRNA	1.464	2.19E-05	Small Nucleolar RNA, C/D Box 118, predicted to serve as guide RNAs in ribose methylation of rRNA.
<i>COX7B</i>	protein coding	1.294	2.24E-05	Cytochrome C oxidase subunit 7B: Mitochondrial, terminal component of respiratory chain
<i>RP11-783K16.10</i>	processed transcript	-1.007	2.43E-05	N/A
<i>MIR5188</i>	miRNA	-1.072	2.84E-05	N/A
<i>RAMP3</i>	protein coding	-1.198	3.78E-05	Receptor activity modifying protein 3: transport calcitonin-receptor-like-receptor to plasma membrane
<i>AC092171.1</i>	lincRNA	-1.075	3.88E-05	N/A
<i>RP11-982M15.5</i>	antisense	-1.738	3.88E-05	N/A
<i>RP11-44N21.4</i>	processed transcript	-1.067	3.95E-05	N/A
<i>RP5-1120P11.3</i>	lincRNA	-1.126	4.65E-05	N/A
<i>SNORD50B</i>	snoRNA	1.310	4.99E-05	Small nucleolar RNA, C/D box 50B, predicted to guide 2'-O-methylation of 28S rRNA
<i>DSG2</i>	protein coding	-1.592	6.23E-05	Desmoglein 2, cadherin family member 5: calcium ion binding
<i>SNORD13</i>	snoRNA	1.102	7.15E-05	Small nucleolar RNA, C/D Box 13: no predicted function
<i>PALD1</i>	protein coding	-1.190	8.66E-05	Phosphatase domain containing paladin 1: no known function
<i>FLRT3</i>	protein coding	1.033	9.09E-05	Fibronectin leucine rich transmembrane protein 3: receptor signaling protein activity and protein binding, bridging

Note: N/A means either the protein's function is uncharacterized or there is no literature found on the function of the gene.

doi:10.1371/journal.pone.0144398.t002

change of differential expression and that of differential H3K4me3 (Spearman $\rho = 0.04$, Fig 2C). Out of the 58 genes, 20 showed the same direction of alteration; 18 cases of both decreased expression and H3K4me3 levels in HD and only two cases of both increased genes and H3K4me3 in HD. The trend that the majority of the dysregulated genes are down-regulated is consistent with a previous study in the HD mouse R6/2 model [31]. It is worth noting that the remaining 38 genes are in the opposite direction; increased gene expression but decreased H3K4me3 in HD (as listed in Fig 2D). These include calcium-signaling genes (e.g. *DSG2*, *ANXA2*, and *ANXA3*), as well as protein kinase or kinase receptors (e.g. *WEE1*, *OSMR*, and *IL17RB*). Interestingly, epigenetic regulators are also found, such as *PRDM6* (histone methyltransferase) and *UHRF1* (which can recruit the DNA methyltransferase *DNMT1* to regulate chromatin structure and gene expression). Several genes involved in neuronal function are seen. For instance, *EGFLAM* (gene for the Pikachurin protein) is known to play important role in the retinal photoreceptor ribbon synapse formation [38], and *EPHA2*, a kinase receptor, plays multiple roles in neural development [39,40].

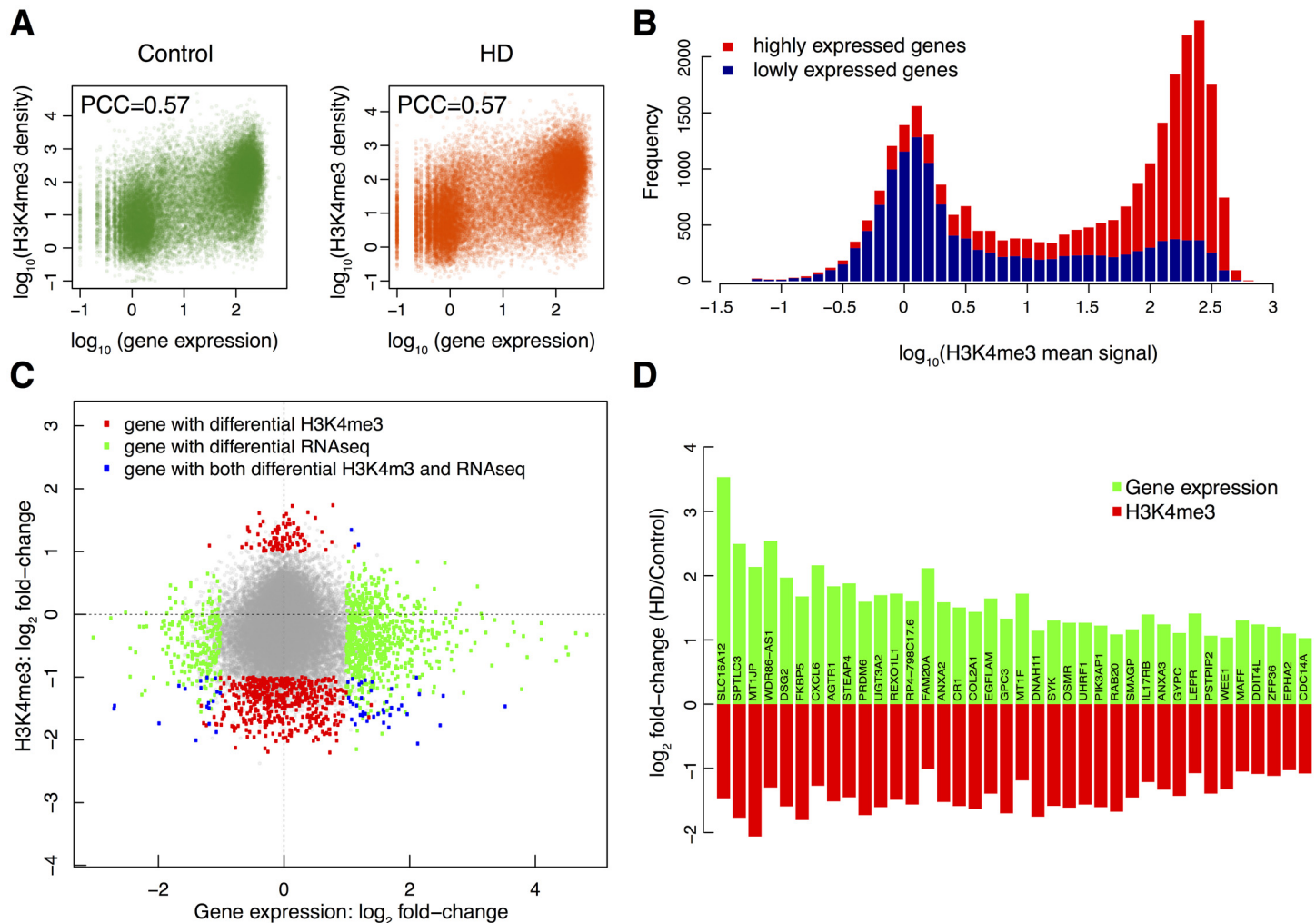


Fig 2. Gene expression and H3K4me3 signal at promoters. (A) Correlation of H3K4me3 signal at promoters and gene expression in control and HD. X-axis shows the logarithmic base mean value of normalized expression values CPKM (RNA-seq reads count per kilobase of exon length per millions of total mapped reads), and Y-axis shows the logarithmic signal density of ChIP-seq reads at proximal promoters (i.e. [-1k, 1k] of TSS). It shows that gene expression levels are positively correlated with the density of H3K4me3 signal at promoters in both controls and HD, with Spearman correlation $\rho = 0.57$. (B) Histogram of normalized H3K4me3 signal in all samples. Bars are colored according to the frequency of genes with high (red) and low (blue) expression. (C) Scatter plot of \log_2 (fold change between HD and control) of H3K4me3 signal at promoter versus gene expression. Genes with differential expression but not differential H3K4me3 are in green; genes with differential H3K4me3 but not differential expression are in red, and genes with both differential expression and H3K4me3 are in blue. In total, 58 genes show both differential expression and H3K4me3, where only 20 of them show the same direction of differentiation between control and HD cases. The remainder (38 of 58) shows a lack of concordance between differential expression and differential H3K4me3 at promoter. (D) Fold changes of expression (in green) and H3K4me3 (in red) between HD and control for the 38 genes with lack of concordance.

doi:10.1371/journal.pone.0144398.g002

GO terms enriched in the genes with proximal H3K4me3 peaks

We computed the enriched GO terms of the genes with 720 proximal differential H3K4me3 signals; 104 up-regulated and 616 down-regulated in HD. We only detected GO terms significantly enriched in the 616 genes with lower H3K4me3 signals in HD than in control. The non-redundant GO terms are shown in Fig 3A. Although the biological process GO terms were diverse, organ morphogenesis and positive regulation of gene expression were the most frequently implicated terms. The molecular function and cellular component GO terms were related to transcription and the extracellular matrix.

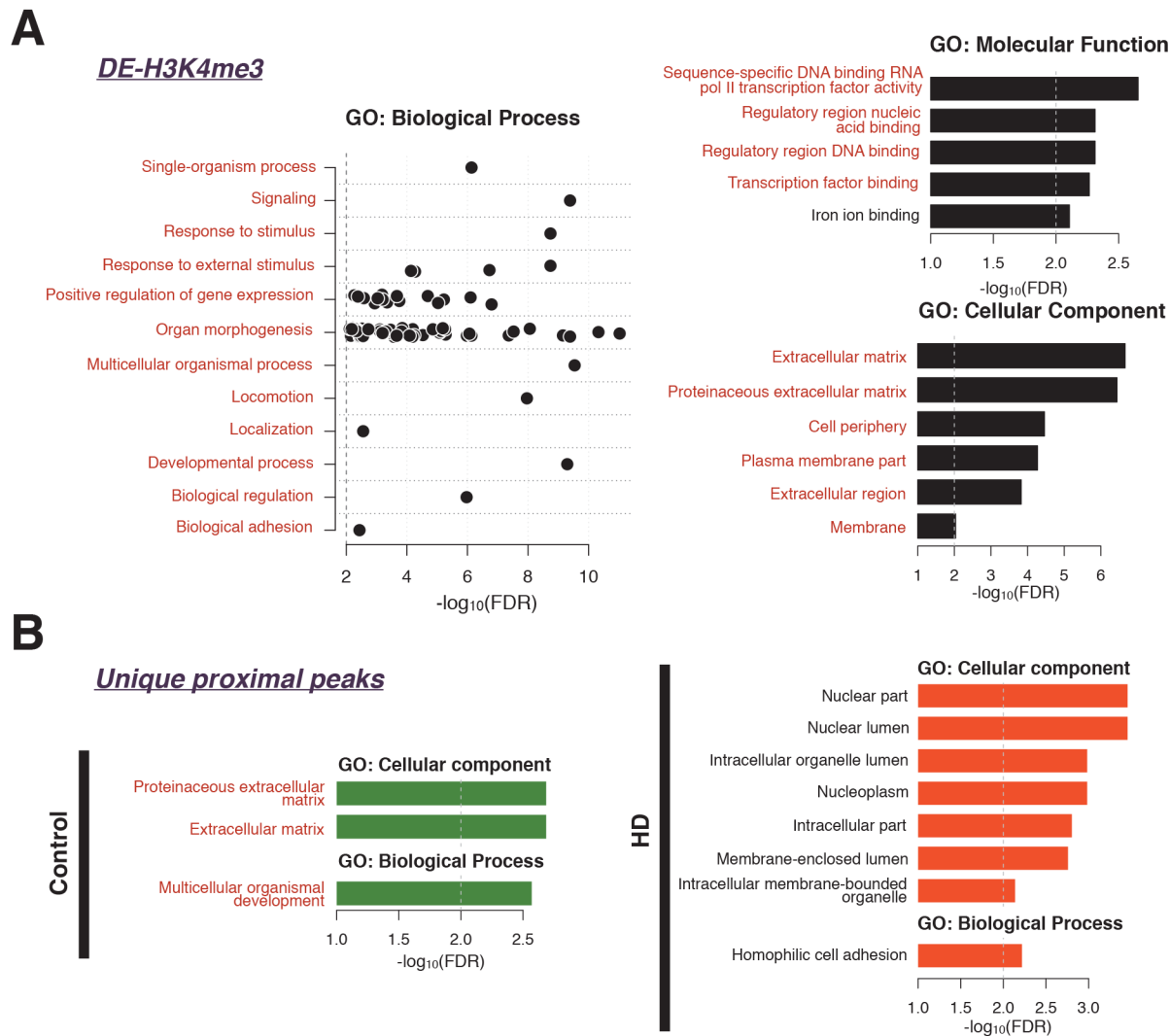


Fig 3. GO terms enriched with proximal peaks in control and HD. (A) GO terms enriched with genes marked with down-regulated H3K4me3 signals in HD. The GO terms overlapped with those derived from differentially expressed genes are highlighted in red text, with iron ion binding (in black) representing the only term not seen in the differentially expressed genes. X-axis shows the negative logarithm FDR values. **(B)** GO terms enriched with genes marked with unique proximal peaks in HD (right; orange) and in control (left; green). These terms also overlap with those seen in HD differentially down-regulated genes.

doi:10.1371/journal.pone.0144398.g003

We also calculated enriched GO terms of the 961 DEGs; 718 up-regulated and 243 down-regulated in HD. Contrary to the differential H3K4me3 results, all detected 664 GO terms were those enriched with up-regulated genes; 576 terms in biological processes, 51 terms in molecular functions and 37 terms in cellular components (S4 Fig).

Interestingly, the GO terms enriched with the genes marked with down-regulated H3K4me3 signals resembled or overlapped with the GO terms enriched with up-regulated genes as parent-child relationships in the GO hierarchical structure (Fig 3A, S4 Fig). Except for iron ion binding (GO:0005506) in molecular function terms, all of the GO terms enriched with the genes with down-regulated H3K4me3 signals were covered by those enriched with up-regulated genes.

GO terms enriched with the genes with unique proximal peaks are presented; 381 and 487 peaks in control and HD respectively. As seen in Fig 3B, all GO terms enriched with genes

marked with unique proximal peaks in control overlapped with those for genes marked with down-regulated peaks in HD (multicellular organismal development (GO:0007275) in [Fig 3B](#) is included in multicellular organismal process (GO:0032501) in [Fig 3A](#)). In HD, interestingly GO terms related to lumen (nuclear lumen (GO:0031981), intracellular organelle lumen (GO:0070013), and membrane-enclosed lumen (GO:0031974)), were detected. The genes for which unique HD proximal peaks are found at their TSS, tend to represent nuclear, organelle, and intracellular components. Among the GO terms for biological processes, homophilic cell adhesion (GO:0007156) was detected for genes with unique H3K4me3 proximal peaks.

Networks of the genes with differential proximal H3K4me3 peaks

We conducted pathway analysis with the SPIA tool [\[41\]](#) and mapped the genes with differential H3K4me3 on biological pathways. We detected eleven significant pathways (see [S5 Table](#)), and five of these are signaling pathways (Rap1, PI3K-Akt, cAMP, Hippo, and Calcium). Consistent with our GO analysis above indicating that many genes with down-regulated H3K4me3 in HD are enriched in extracellular or membrane GO terms, our pathway analysis found that these genes tend to be clustered at the cellular membrane and located in the upper ranks in the signaling pathways and the extracellular matrix (ECM)-receptor interaction pathway.

We also calculated the enriched pathways for DEGs. We detected 25 significant pathways and found that most genes in these pathways were up-regulated in HD (see [S6 Table](#)). The PI3K-Akt and Hippo signaling pathways were also detected to be enriched. As reported by Labadorf et al. (submitted), many of the pathways were related to immune function.

To investigate the relationships between the genes with differential H3K4me3 and DEGs, we extracted the gene interaction sub-networks of the detected pathways in [S5](#) and [S6](#) Tables. The gene interaction annotation used in this analysis is also based on the annotation in KEGG database. All genes in significant pathways are summarized in [S7 Table](#). As we observed in the above pathway gene enrichment analysis, the two gene sub-networks of DEGs were extracted as the immune system module ([Fig 4](#)). Each sub-network was composed of a specific gene family. One of the two immune system sub-networks was formed by chemokine receptors (CCR1, 5, and 9; CXCR1, 2, and 4) and ligands (CCL2, and 26; CXCL1, 5, 6, and 7), and the other was formed by complement subcomponents (C1QA, C1QB, C1QC, C1R, and C4B).

On the other hand, the sub-network of the genes with differential H3K4me3 was mainly detected as the signaling pathway module ([Fig 4](#)) and was composed of various signaling genes listed in [S7 Table](#). Those genes in the signaling pathway module straddled across the cell membrane to the inside of the cell and preserved cascade structures in the pathways detected in the above pathway gene enrichment analysis. Although several DEGs were also included in the signaling pathway module, those DEGs were in a specific gene family (e.g. immunoglobulin-like receptor gene family in the upstream of the signaling pathway module) and were not the hubs in the module. That is, even if we remove the DEGs from the signaling pathway module, still the hierarchical structure composed of the genes with differential H3K4me3 remains in the module. We also found the ECM-receptor pathway module composed of 12 extracellular component genes such as collagen and laminin.

Evidence for regulatory role of distal H3K4me3 peaks

One third of H3K4me3 peaks were localized at more than 1 kb away from any known TSS. To investigate their potential regulatory functions, we first assessed whether they overlapped with genomic regions with other regulatory signals. Recent progress in functional genomics has amassed many biochemical data for regulatory elements, including the co-occurring histone marks of H3K4me1 and K3K27ac [\[26\]](#), bidirectional CAGE signal at promoters [\[42\]](#), hotspot

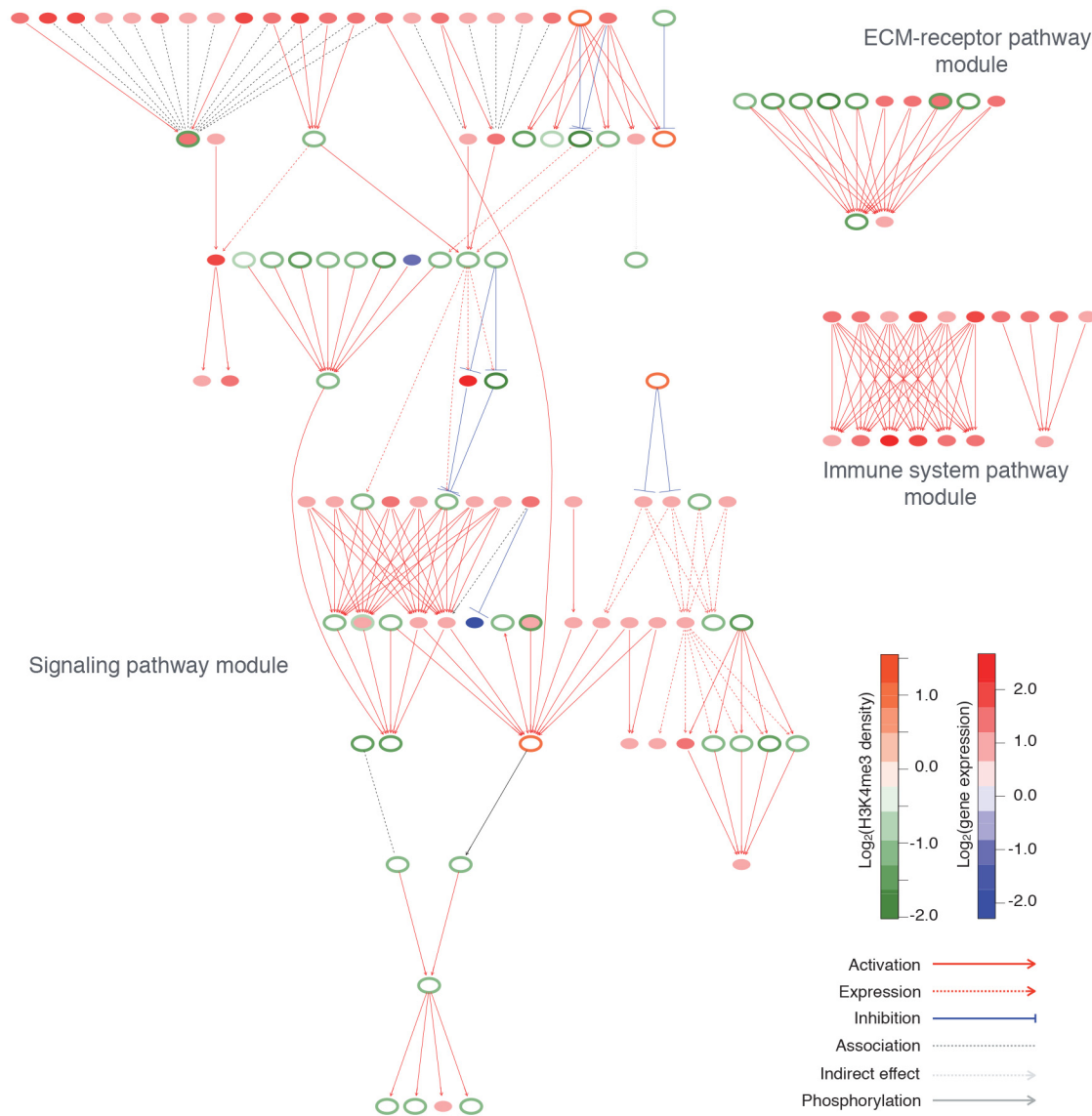


Fig 4. Sub-networks of DEGs and genes with differential H3K4me3. Each node represents a gene; the fill color of each node indicates the fold-change of gene expression between HD and control, and the outline color of each node indicates the fold-change of H3K4me3 between HD and control. Immune system and ECM-receptor pathway modules were composed of specific gene families. On the other hand, signaling pathway modules were composed of various signaling genes forming hierarchical structure. The directions of all the arrows were drawn toward the bottom of figure.

doi:10.1371/journal.pone.0144398.g004

of transcription factor binding sites [43], clustered DNaseI hypersensitive sites [44], and enriched eQTLs [45]. By aligning these signals from prefrontal cortex and other cortical related brain regions with our H3K4me3 data, we found interestingly many of the distal H3K4me3 peaks co-localize with regulatory elements defined by these signals. For example, although the distal peak in Fig 5A was called in both HD and control samples, it is much stronger in HD. It harbors eleven SNPs that were shown to be eQTLs in brain (black vertical lines in Fig 5A). Notably, this region is also centered on ChIP-seq peak of enhancer marks of H3K27ac and H3K4me1 on human frontal cortex from the Roadmap Epigenomics project. Active enhancers often correspond to open chromatin regions, which facilitate binding of transcription factors. Indeed, this region has open chromatin (the DNase track in Fig 5A) and is bound by more

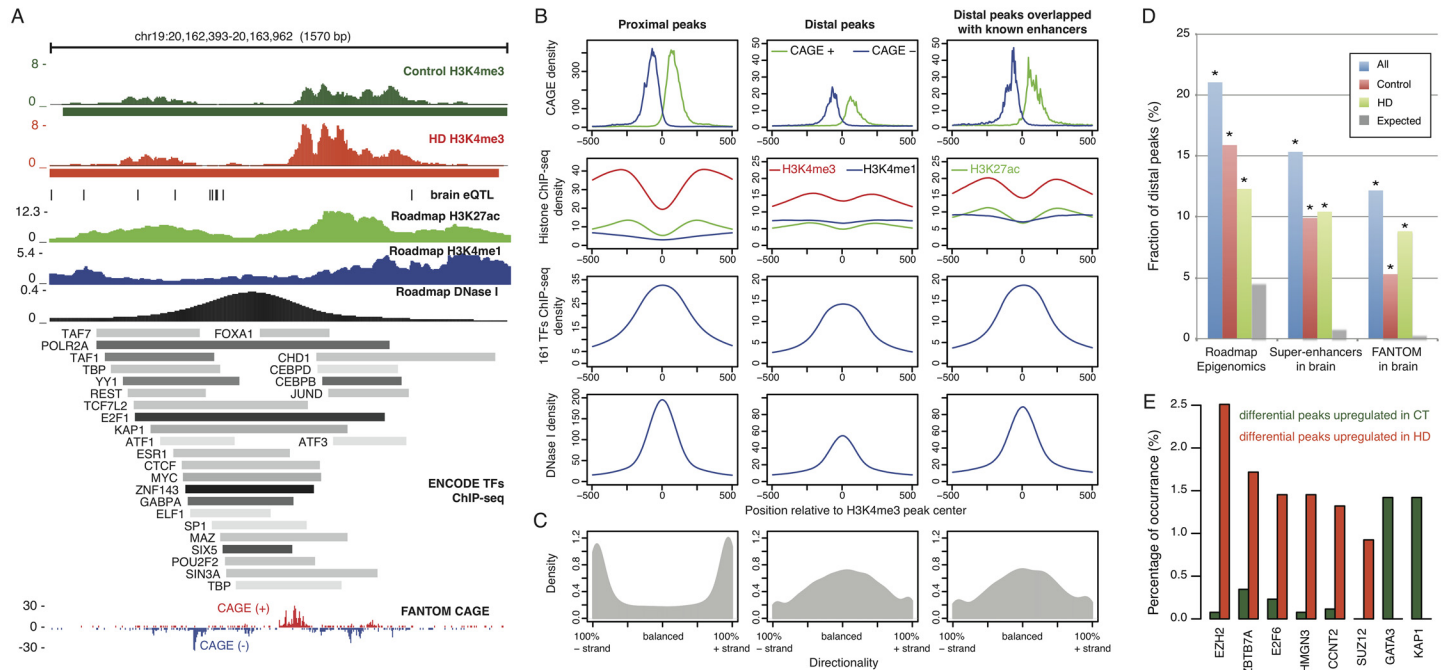


Fig 5. Overlapped regulatory signals and differentially enriched TF binding on distal peaks. (A) UCSC genome browser of a genomic locus (coordinate in hg19 shown on the top) where a distal H3K4me3 peak, called in both control and HD cases, overlaps with enhancer regulatory features including enriched H3K4me1 (“Roadmap H3K4me1”), H3K27ac (“Roadmap H3K27ac”), DNase I (“Roadmap DNase I”), TFs binding hotspot (“ENCODE TFs ChIP-seq”), and bidirectional CAGE signals (“FANTOM CAGE”). (B) Aggregation plots of several types of regulatory signals, including CAGE (1st row), histone marks (2nd row), TF binding (3rd row), and DNase I hypersensitivity (4th row), centered on the middle of H3K4me3 peaks. (C) Directionality of H3K4me3 peaks represented by CAGE tags shows that, unlike the strong bias towards sense strand for proximal peaks, distal peaks show similar level of CAGE tags in both directions. (D) Fraction of distal peaks overlapping with different sources of known enhancers occurred at rates greater than expected (All = blue, control = red, HD = green, Expected = grey). (E) Transcription factors binding in differentially enriched distal peaks.

doi:10.1371/journal.pone.0144398.g005

than 20 transcription factors (ENCODE TFs ChIP-Seq track in Fig 5A) particularly those known key factors for facilitating the 3D structure of chromatin, such as CTCF [46] and SP1 [47,48]. Finally, CAGE, a technique for capturing the 5' end of mature transcripts, was recently found useful for identifying active enhancers based on a bi-directional signal [42]. It showed clear bi-directional signals at this region and an enhancer was predicted by FANTOM5 consortium (CAGE track in Fig 5A). Furthermore, we also found that the role of distal peaks as enhancers is also supported by the aggregated signal of CAGE tags, H3K4me3, H3K9ac, and DNase I on all distal peaks (Fig 5B). Transcription directionality based on the CAGE data from the FANTOM5 project further supports this view (Fig 5C). Thus the distal differential H3K4me3 peaks may represent differential enrichment at enhancer sites.

By computing the percentage of distal peaks overlapping with three datasets of known enhancers (see Methods), we found that 15–22% of distal peaks significantly overlapped known enhancers (p-value < 2.2×10^{-16} ; Fig 5D). Altogether, an estimated 36% of distal peaks overlapped with at least one of three enhancer datasets (S4 Table). The aggregation signals of regulatory marks were even more enriched on this subset of distal peaks than the total set of distal peaks, as seen from the third column of Fig 5B. From the set of distal peaks that overlap with at least one of the three known enhancer sets, 74 enhancer-like distal peaks are differentially enriched between HD and control (S4 Table).

To investigate the potential functional roles of distal H3K4me3 peaks, we analyzed the TFBS enrichment at these peaks with the most recent TF ChIP-seq data from the ENCODE

project (see [Methods](#)). We identified those TFs specifically binding to distal peaks that were differentially enriched between HD and control. By comparing the TF binding occurrences between 209 and 51 differential distal peaks that were up-regulated in HD and control respectively, we found that six transcription factors (ZBTB7A, E2F6, CCNT2, HMGN3, EZH2, and SUZ12) had enriched binding sites in HD-specific distal peaks and two TFs (GATA3 and KAP1) had enriched binding sites in control-specific distal peaks ([Fig 5E](#)). Among these, EZH2, a histone methyltransferase, had the most significant difference (adjusted p-value = 2.87×10^{-10}).

Both EZH2 and SUZ12 are subunits of polycomb repressive complex 2 (PRC2), which can primarily methylate histone 3 on lysine 27 (H3K27me3) and therefore repress transcription. In this study, we further looked into the enrichment of Polycomb-repressed regions (indicated by the 'ReprPC' and 'ReprPCWk' ChromHMM states in human frontal cortex of the Roadmap Epigenomics project [[49](#)]) in HD-specific peaks. We found that the repressed Polycomb state, representing the H3K27me3 signal, is depleted in the distal peak regions with significantly enriched H3K4me3 signal in HD (hypergeometric test p-value = 4.66×10^{-5}), but not in control-enriched distal peaks (hypergeometric test p-value = 0.39). Previous study had shown that H3K4me3 can inhibit the activity of PRC2 in an allosteric fashion assisted by the Su(z)12 C terminus [[50](#)]. Our result of PRC2 depletion in the regions with increased H3K4me3 in HD might suggest that the epigenetic role of PRC2 inhibition is caused by up-regulated H3K4me3 in Huntington's disease.

Discussion

While patterns of H3K4me3 have been previously studied in the R6/2 mouse model of HD [[31](#)], we believe this is the first genome wide study of this mark in human HD versus control brains. We had previously performed mRNA-sequencing in a series of 20 HD and 49 non-neurological control brains, and therefore, we were able to compare the patterns of H3K4me3 enrichment with 961 differentially expressed mRNAs (FDR $q < 0.01$) in HD and control brains (Labadorf et al. submitted). Because 50% to 90% of neurons have degenerated in the HD striatum, even among low-grade HD cases [[3](#)], we chose to analyze the prefrontal cortex (Brodmann area 9). While the hallmarks of HD pathology (e.g. intranuclear inclusion bodies) are present in the prefrontal cortex, the extent of neuronal degeneration is considerably less severe than that seen in the striatum [[3,51](#)]. Some of the HD samples studied here have been previously studied by microRNA-sequencing and mRNA-sequencing of the same region of the prefrontal cortex [[12,36](#)] (Labadorf et al. submitted).

In this study, we sought to minimize the biases introduced by simple neurodegeneration in two ways. First, the H3K4me3 ChIP-sequencing was performed in NeuN+ sorted neuronal nuclei [[35,52](#)]. Thus, the contrast is specific to neurons in both HD and controls and accounts for possible differences in the densities of specific cell types in the HD brain due to degeneration or reactive gliosis. Second, the prefrontal cortex (BA9), a region with mild neuropathological involvement in HD [[3,4](#)], was selected for analysis. A total of 28,608 H3K4me3 peaks were defined in the combined HD and control samples. While most peaks (93.5%) were called in both HD and control brains and a majority (59.2%) were called in all twelve samples studied, 961 were uniquely called in HD brain and 889 were uniquely called in controls ([Fig 1, S2A Fig](#)). The H3K4me3 mark is known to mark active transcription start sites (TSS), and 65.8% of the called peaks were located within 1,000 bp of an annotated TSS (designated "proximal peaks"). Both the proximal peaks and distal peaks (positioned at greater than 1000 bp of an annotated TSS) were robustly called with confident read depths ([Fig 1D](#)).

In the differential analysis of H3K4me3 peaks, 720 proximal peaks were differentially enriched, and these were primarily down-regulated among HD cases relative to controls (104 up-regulated and 616 down-regulated in HD). In contrast, 260 distal peaks were differentially enriched, and these were primarily up-regulated in HD relative to controls (209 up-regulated and 51 down-regulated in HD). These findings suggest that proximal and distal H3K4me3 peaks perform different functions and respond differently to the pathology initiated by the mutant HTT protein.

The H3K4me3 mark is known to associate with active gene transcription, and this relationship is confirmed in this study, where high versus low H3K4me3 proximal peak density and was strongly related to high versus low mRNA expression (Fig 2B, χ^2 p-value $< 2.2 \times 10^{-16}$). It should be noted however, that for the 720+961 “differential” genes, the rho for both HD and controls is reduced, although more in HD (Spearman rho = 0.33) than in controls (rho = 0.4). In contrast however, differentially enriched H3K4me3 proximal peaks were not associated with increased expression of the implicated genes. Few genes with differential enrichment of H3K4me3 density showed differential mRNA levels. The disruption of the regulatory influence of the H3K4me3 enrichment on gene expression in HD brains provides strong evidence for transcriptional dysregulation as a major component in disease expression in HD (see Fig 2B and 2C), as has been previously proposed [9,10], or it may suggest that additional epigenetic marks or mechanisms are required to better explain the full range of transcriptional dysregulation associated with the disease [53].

The genes with the most differentially enriched H3K4me3 marks at their TSS are *ALG14* (Asparagine (N) glycosylation) which regulates protein folding and stability; *LRRTM2* (leucine rich repeat transmembrane neuronal 2), involved in development of excitatory synapses; *HES4* (Hairy And Enhancer Of Split 4) which is involved in protein dimerization and transcription factor binding; and *COX7B* (Cytochrome C Oxidase Subunit 7B) which is a terminal component of the mitochondrial respiratory chain. Of these only HES4 is differentially expressed, and we have previously described *HES4* as having differential H3K4me3 enrichment and being related to the extent of pathological involvement in the striatum [32].

Other genes shown to have a reduced H3K4me3 signal in HD or corresponding mouse model from a previous study [31] are also revealed in this study. For instance, the down-regulated *Bdnf* gene has reduced H3K4me3 signal at one of its promoters (exon II) in HD and R6/2 mouse. We also observed the same trend (fold change = 0.78, q-value = 0.027). Other genes shown have significant decreases of H3K4me3 in Vashishtha et al. also had decreased H3K4me3 levels in our results, such as *PENK1* (fold change = 0.34, but not significant), and *DRD2* (fold change = 0.43, q-value = 0.029).

Differential H3K4me3 enrichment is not associated with differential gene expression

While a genome wide analysis of mRNA levels from our sequencing studies showed a strong correspondence with H3K4me3 enrichment (χ^2 p-value $< 2.2 \times 10^{-16}$; see Fig 2B), this relationship was notably absent for differentially enriched H3K4me3 peaks and differentially expressed genes (DEGs) (see Fig 2C). Notably, 718 (74.7%) of the 961 DEGs are up regulated in HD, suggesting that one might expect to see an increase in differential H3K4me3 enrichment at the TSS of these DEGs. However, only 58 of the 720 differential H3K4me3 proximal peaks are associated with DEGs, and only 20 of these (2.8% of the 720) are in the expected regulatory direction (18 down-regulated DEGs have reduced differential H3K4me3, and 2 up-regulated DEGs have increased differential H3K4me3 enrichment), suggesting that differential H3K4me3 is not responsible for the DEGs. This result holds with a previous study [31] where

genome-wide analysis of H3K4me3 and RNAseq in the 12-week cortex of R6/2 mouse revealed a high degree of overlap between genes with decreased expression and decreased H3K4me3 levels (e.g. 293 in down:down, 24 in up:up). Although the mouse data may not be directly comparable, by simply checking the orthologs of the 293 down:down genes in our human data we found that most of these genes also have a down:down relationship (i.e. decreased H3K4me3 levels and decreased expression in HD). While both studies found consistent results for the selected genes and there is a large overlap between genes with same direction change in expression and H3K4me3 levels, nonetheless, a proportion of genes were found not infrequently and unexpectedly in opposite directions in both studies. This discordant pattern holds up in an examination of the concordance of the GO enriched terms for down-regulated DEGs overlap with increased differential H3K4me3 signals described below.

Analysis of GO term enrichment for genes at enriched H3K4me3 peaks

In an analysis of GO term enrichment only the genes associated with down-regulated peaks in HD showed significant GO terms, with a total of 166 enriched GO terms. We compared the GO terms for differential H3K4me3 with those for the 961 differentially expressed mRNAs ($q < 0.01$). In GO and pathway analysis, we observed similar GO term enrichment in the up-regulated mRNA genes in HD and the down-regulated H3K4me3 target genes in HD.

Although the biological processes of these terms were diverse, organ morphogenesis and positive regulation of gene expression were the most frequently implicated terms. Common molecular function and cellular component GO terms relating to transcription (e.g. Sequence specific DNA binding and RNA pol II activity) and the extracellular matrix (membrane and cell periphery) were observed (see [Fig 3](#)). The unexpected pattern of opposite trend of regulation (e.g. increased mRNA levels with decreased H3K4me3) presenting overlapping GO term enrichment, suggests the increased levels of mRNA are likely attributed to regulatory mechanisms other than H3K4me3 control of transcriptional initiation.

Transcriptional dysregulation has a crucial role in HD pathology. Interactions between mutant HTT and important TFs including SP1 [[54–56](#)], REST [[54,57,58](#)], TAFII130 [[56](#)] and others have been observed and may contribute to the observed transcriptional dysregulation [[9,59](#)]. Transcriptional dysregulation may also be influenced by the accumulation of the mutant HTT protein. Two recent studies with in vivo and HD Drosophila models showed that the mutant HTT protein facilitates the chromatin regulator PRC2 and alters H3K4me3 states at transcriptionally active genomic loci [[60,61](#)].

Several mechanisms might be responsible for the disparity between H3K4me3 enrichment and mRNA levels. Our findings of decreased histone modifications may represent an attempt by the neuron to rescue or reverse via epigenetic means, the aberrant increase in transcription that we see in mRNA levels by RNA-sequencing. Alternatively, decreased H3K4me3 signal has been associated with apoptosis [[62](#)] which has been strongly implicated in HD and thus the differences that we observe may represent part of that neurodegenerative process in HD.

Role of distal H3K4me3 peaks

We subdivided the peaks into proximal (defined as within 1000bp of a transcription start site), and distal peaks (located greater than 1000bp of a transcription start site). One third of H3K4me3 peaks (34.2%) were located distally from transcription start sites, and 260 of them were differentially enriched between HD and control ([S4 Table](#)) with 209 up-regulated and 51 down-regulated in HD.

H3K4me3 is usually considered as a promoter mark for actively transcribed genes [[26,63–65](#)]. However, several studies support that H3K4me3 can mark enhancers as well [[66,67](#)]. In

this study we projected the commonly known regulatory marks on H3K4me3 peaks and found that the distal peaks show largely similar patterns of aggregation as proximal peaks, as seen in Fig 5B. One possibility is that some distal peaks are promoters of unannotated genes or non-coding RNAs [68,69]. But this explanation does not appear to account for most distal peaks for the following reasons: First, if the distal H3K4me3 signal is an active promoter for an unannotated gene, one may expect a corresponding increase RNA-seq reads derived from the transcription at that position. However, most distal peaks have very low RNA-seq read coverage (0 or 1 read, see S5 Fig). Second, in our CAGE tag directionality analysis, distal peaks show a clear pattern of bidirectionality, unlike the uni-directionality for proximal peaks. This finding is consistent with the defining feature of transcribed enhancers (or eRNAs) recently characterized by the FANTOM consortium [42]. By intersecting distal peaks with three known sets of enhancers, an estimated 36% of the distal H3K4me3 peaks appear to localize to enhancer sites.

It is worth noting that the distal peaks are generally less intensive than proximal peaks (Fig 1D). Although we had previously checked the specificity of the anti-H3K4me3 antibody we used, for absence of detectable immunoreactivity for the related epitopes of H3K4me2 and H3K4me1 [70], future studies are needed to explore whether the antibody detects other types of histone modification at distal peaks.

TF differentially bound in distal peaks

We compared the TFs bound at the proximal and distal peaks, and found that the TFs associated with chromatin structuring are enriched in distal peaks. Chromatin looping allows far-distant enhancers to be brought into the proximity of promoters of protein-coding genes. Enhancers are thought to interact with protein complexes necessary for chromosome looping and to facilitate RNA Polymerase II recruitment of the promoter of the target gene. This result is consistent with the finding from a recent study using ChIA-PET interaction data that cohesin, CTCF, and ZNF143 are the key components of 3D chromatin structure [71]. Studies in peripheral tissues found that many of the ectopic H3K4me3 sites corresponded to enhancers bound by CTCF and cohesin [66,67]. Recently the link between distal H3K4me3 peaks and promoter-enhancer loopings has also been reported for human cerebral cortex [72,73].

When looking at the enriched TFs in differentially enriched H3K4me3 peaks between HD and control, we found six transcription factors have enriched binding sites in HD-specific distal peaks, including EZH2 and SUZ12. Previous studies also found the role of EZH2 for DNA hypermethylation in the epigenetic inhibition of the dopamine receptor D4 (DRD4), which is important for neurodegenerative disease such as HD and Parkinson's disease [74]. These findings may implicate the enrichment of EZH2 in the HD-specific distal H3K4me3 peaks.

EZH2 and SUZ12 are also core subunits of the polycomb repressive complex 2 (PRC2). Given that PRC2 should recognize the mark that EZH2 generates, the finding that PRC2 is depleted while EZH2 is enriched suggests that the H3K4me3 mark may appear preferentially on those EZH2 target regions that are not regular targets of the PRC2 complex.

H3K4me3 was recently found to inhibit PRC2 assisted by the Su(z)12 C terminus [50]. Our findings of depleted PRC2 in HD-enriched H3K4me3 peak regions may indicate that increased H3K4me3 in HD is associated with PRC2 inhibition.

In summary, we examined the differential enrichment of H3K4me3 signals and the relationship of this histone mark to RNA levels in both HD and control prefrontal cortex. Genes with differential H3K4me3 signal at their promoters were rarely found to be differentially expressed. These findings support the hypothesis that transcriptional dysregulation plays an important role in the pathogenesis of HD. GO term analysis of the genes with differential H3K4me3 enrichment identified terms only among the set of genes with down-regulated H3K4me3 in

HD, and most frequently implicated biological process terms with organ morphogenesis and positive regulation of gene expression. More than 9000 H3K4me3 peaks were located not near to any recognized transcription start sites and approximately 36% of these “distal” peaks co-localized to known enhancer sites. Transcription factors enriched in distal peaks, implicated the polycomb repressive complex 2 (PRC2), suggesting that H3K4me3 histone hypermethylation is linked to dysregulated polycomb activity in the HD neuronal epigenome.

Materials and Methods

H3K4me3 ChIP-seq sample preparation

Postmortem brain tissues from the prefrontal cortex Brodmann area 9 (BA9) were obtained from the Harvard Brain and Tissue Resource Center (HBTRC), McLean Hospital, Belmont MA. Six human HD post-mortem samples were selected for ChIP-seq. HD cases were confirmed to have expanded CAG repeats longer than 40 (mean = 44.3), a mean age at death of 60.3 years, and a mean postmortem interval of 20.1 hours. Six non-neurological controls of similar age (mean = 67.7), sex, and postmortem interval (mean = 19.6) were selected from samples studied previously in the Akbarian Lab [33–35] for H3K4me3 ChIP-sequencing (Table 1). This study has been designated exempt (Protocol # H-28974) by the Boston University School of Medicine Institutional Review Board, as no human subjects were studied and all data are derived from post-mortem human brain specimens.

Neuronal nuclei from prefrontal cortex BA9 were NeuN+ labeled and fluorescence activated sorted (FACS) for subsequent H3K4me3 ChIP-seq. Briefly, approximately 300 mg to 500 mg of -80°C frozen grey matter dissected from BA9 was homogenized in 5 ml lysis buffer and ultracentrifuged over sucrose solution at 24,400 rpm for two hours at 4°C. The resulting pellet was suspended in 550 µl PBS, and immunolabeled with NeuN for 60 minutes while rotated in darkness. NeuN+ and NeuN- fractions were separated and collected by FACS [52].

ChIP sequencing was performed for the NeuN+ fraction. Briefly, genomic regions with H3K4me3 modification were extracted by chromatin immunoprecipitation by incubating digested nuclei with anti-H3K4me3 (Upstate; cat # 07–473) at 4°C overnight and further prepared into ChIP-seq library for sequencing, as described previously [75]. For each brain sample, 1 µg of DNA was used to construct sequencing libraries using Illumina TruSeq Sample Prep Kit, according to the manufacturer's protocol (Illumina, San Diego, CA), and sequenced using 51nt single-end reads on Illumina's GA II. The ChIP-seq data for six Huntington's disease patients and six controls from this study have been submitted to GEO (<http://www.ncbi.nlm.nih.gov/geo/>) under the accession number of GSE68952.

H3K4me3 ChIP-seq read mapping and peak detection

We mapped all reads in H3K4me3 ChIP-seq libraries to the human reference genome (hg19) with Bowtie (version 1.0.0) [76] allowing 1 mismatch (see Table 1 for mapping statistics). Mapped reads were used for the input of the MACS algorithm (version 1.4.0rc2) [77] to identify regions in the genome that are enriched in H3K4me3 (called H3K4me3 peaks or peaks in short). We extracted H3K4me3 peaks with false discovery rate (FDR) ≤ 0.01 . Peaks that were called in fewer than half the HD samples or fewer than half the controls were removed from that group respectively. We then merged the remaining peaks from each group to create a union set of peaks for all subsequent analyses. Thus some peaks were called in only the HD group, while others were called only among controls.

The union set of peaks were designated as “proximal” if the distance between the central points of peaks to the nearest transcription start site (TSS) of a gene is less than or equal to

1 kb, otherwise they were designated as “distal” peaks. We further annotated the peaks as either “common” or “unique” based upon the fold change of H3K4me3 signal at the peak summit between HD and control: common peaks have less than or equal to 2 fold change, and unique peaks have greater than 2 fold change. We defined the H3K4me3 signal at a peak by first taking the average of the RPM (reads per million) at the summit position over samples in HD and control, then normalizing the peak signals across the HD and control groups based on the “invariant set” method [78] (see [S1 Table](#) for summary of detected peaks).

RNA-seq data

Analyses reporting differential expression of mRNA in HD versus controls are reported elsewhere (Labadorf et al. submitted). DESeq2 [79] was used to identify differentially expressed genes (DEGs) between HD and control, adjusting for age at death binned into intervals 0–45, 46–60, 61–75, and 90+ and a categorical RNA Integrity Number ($RIN > 7$ and $RIN \leq 7$) as covariates. Differential expression analysis identified 961 out of 27,907 confidently expressed genes with significantly altered expression at q -values ≤ 0.01 and $\text{abs}[\log_2(\text{fold change})] \geq 1$ in 20 HD vs. 49 control samples from the BA9 region of the prefrontal cortex. The list of all 961 DEGs is in [S2 Table](#). Sample information is provided in [Table 1](#) and the methods to define DEGs are found in (Labadorf et al. submitted).

To correlate gene expression with H3K4me3 signal at promoters, we extracted the normalized counts from DESeq2 (e.g. `counts(dds, normalized = TRUE)`) and further normalized the values by the total length of meta-exons of annotated genes. Meta-exon were defined by merging all overlapping exons of a gene.

Detection of differentially enriched H3K4me3 peaks

We extracted ChIP-seq reads mapped to the promoter regions (defined as ± 1000 bp of TSS) for all annotated transcripts (Gencode v17) [80] using the `coverageBed` tool in `bedtools` (version 2.19.0) [81]. We then applied the DESeq2 algorithm (version 1.4.5) [79] to detect transcripts with differential H3K4me3 signals at promoters between HD and control (q -value ≤ 0.01 and $\text{abs}[\log_2(\text{fold change})] \geq 1$). For genes with multiple expressed transcripts, we selected the transcript with the lowest q -value from the DESeq2 result for subsequent analyses, in order to match with the differentially expressed genes from RNA-seq. A list of 720 genes with differential H3K4me3 proximal peaks is in [S3 Table](#).

For distal peaks, we extracted the total reads mapped to each peak across all samples and used DESeq2 to compute enrichment of H3K4me3 signal between HD and control. Using the same cutoff as for promoters, we defined 260 distal peaks with significant enrichment and this list is in [S4 Table](#).

Gene ontology analysis

Using GOseq (version 1.16.2) [82], we identified significantly enriched gene ontology (GO) terms for the following sets of genes: (1) the genes that correspond to the proximal differential H3K4me3 peaks and (2) the nearest genes to the unique proximal peaks. To compare these results with differentially expressed genes between HD and control, we also computed the GO term enrichment (q -value ≤ 0.01) for the 961 DEGs (Labadorf et al. submitted). We used REVIGO web server [83] to summarize and reduce the complexity of the significant GO terms (*simRel* score cutoff = 0.7).

Pathway analysis

We identified significantly enriched pathways with SPIA (version 2.16.0) [41] for the genes that corresponded to the proximal differential H3K4me3 peaks. We then extracted the sub-pathways containing the genes with KEGGgraph [84]. We also applied the same analysis to the set of 961 DEGs (Labadorf et al. submitted).

Transcription factor enrichment in H3K4me3 peaks

Transcription factor (TF) binding regions were extracted from the TFBS cluster file (wgEncodeRegTfbsClusteredV3) downloaded from the UCSC genome browser [85], which correspond to peaks called from CHIP-seq data for 161 TFs in 91 cell types performed by the ENCODE consortium [86,87]. We then intersected the H3K4me3 peaks with the TF binding regions and determined the number of TFs found in each category of H3K4me3 peak. Each overlapping TF was counted only once, regardless of the number of different cell types in which it was found. We then compared the number of occurrences of TFs between the two categories of H3K4me3 peak by Fisher's exact test, and corrected the resulting p-values for multiple testing using the Benjamini-Hochberg procedure. Odds ratios (OR) were estimated from Fisher's exact test. TFs with q-value < 0.01 and odds difference > 30% (i.e. OR > 1.3 or OR < 0.7) were deemed significant.

Enrichment of other regulatory signals with H3K4me3 peaks

In addition to TF binding data from the ENCODE consortium, we also compared our H3K4me3 peaks with the regulatory signals from the following three sources: (1) CAGE signal in the frontal lobe adult brain from the FANTOM5 project [42], (2) H3K4me1, H3K4me3, H3K27ac, and H3K9ac CHIP-seq signal in the middle frontal lobe of brain, and (3) DNase I signal in human fetal brain downloaded from the Roadmap Epigenomics project [49]. For each H3K4me3 peak, we first extended to 1000bp in length centered at the maximum value of overlapped DNase narrowPeak (if any) or the midpoint of the H3K4me3 peak. Then we subdivided each peak into 200 bins (i.e. 5bp per bin) and computed the mean values of the signal in each bin, which were then used to create the aggregation plots in Fig 5B. The transcription directionality (in Fig 5C) of each peak was computed using the same definition as that used in the FANTOM5 project [42]. Briefly, the normalized CAGE signals from first half bins (i.e. 1–100 bins) and second half bins (101–200 bins) were used to calculate a peak-set wide directionality score, D , for each peak over aggregated normalized reverse, R , and forward, F , strand CAGE signals across all peaks, $D = (F-R) / (F+R)$.

To investigate whether our distal H3K4me3 peaks may function as enhancers, we compared them with three sets of enhancer annotations: (1) the 'strong', 'weak', or 'bivalent' enhancers based on genome-wide chromatin states of the brain dorsolateral prefrontal cortex tissue generated by the Roadmap Epigenomics consortium [49], (2) the super enhancers in brain mid-frontal lobe defined by occupancy of the master transcription factors and mediators [88], and (3) the transcribed enhancers defined based on the bidirectional CAGE signal from the FANTOM5 project [42].

Supporting Information

S1 Fig. Flowchart of H3K4me3 peak calling. This figure depicts the process by which the CHIP-seq data were aligned and processed to identify proximal and distal peaks, as well as differentially expressed common and unique peaks. (TIF)

S2 Fig. (A) The number of HD and control samples supporting a called H3K4me3 peak. Most peaks (59.2%) were called in all six HD and control brains. The area of the bars corresponds to the number of peaks. Only peaks with at least 3 samples supported in either of the groups are included in this plot and analyzed in this study. (B) The number of samples supporting a H3K4me3 peak and its normalized signal in each group. Peaks called in all six HD and all six controls had higher peak densities than did peaks called in fewer than six.
(TIF)

S3 Fig. (A) mRNA levels are bimodal for both HD and control, and consequently we subdivided them into “high” and “low” groups (dotted line). (B) The distributions of both HD and control H3K4me3 peaks are bimodal and consequently we subdivided them into “high” and “low” groups. (C) Histogram of H3K4me3 mean normalized signals in HD cases (left) and controls (right). Bars are colored according to the high (red) or low (blue) expression level. Both HD and control cases show a similar level of correspondence between expression and H3K4me3 level.
(TIF)

S4 Fig. GO terms enriched with genes up-regulated in HD. The GO terms overlapped with those of genes with down-regulated H3K4me3 peaks are highlighted in red.
(TIF)

S5 Fig. The distribution of expression level of distal peaks, measured as the median of RNA-seq reads summit in the distal peaks across the samples in HD and controls, shows that most distal peaks are covered with no (0) or low (e.g. <10) RNA-seq reads.
(TIF)

S1 Table. Detected H3K4me3 peaks in control and HD. CT: Control. Distance to TSS: Distance is measured from the center of each peak to a TSS of the nearest gene. If the peak center is located upstream (5') of TSS, the distance is minus; otherwise, plus. Location: If a peak overlaps with any gene in GENCODE v17, it is categorized as intragenic; otherwise, intergenic. Note that a peak may be located intragenic of a gene, but it may refer to a different gene with the nearest distance to TSS.
(XLSX)

S2 Table. List of 961 differentially expressed genes. SE: Standard error.
(XLSX)

S3 Table. List of 720 genes with differential H3K4me3 signal in the promoter regions. SE: Standard error.
(XLSX)

S4 Table. List of 9772 distal peaks. 260 distal peaks with differential H3K4me3 signal are marked as “Yes” in the DE-H3K4me3 column. 3560 distal peaks overlapping at least one enhancer sets are marked as “Yes” in the “putative enhancer peak” column. SE: Standard error.
(XLSX)

S5 Table. Pathways enriched with genes with differential H3K4me3 signals. In the link to the detected KEGG pathways, genes with both up-regulated and down-regulated differential H3K4me3 signals in HD are highlighted in red. #Gene: Statistics of detected genes (left) and total genes (right) in a pathway. Nested detected pathways: Pathways detected in this analysis and embedded in other detected pathways.
(XLSX)

S6 Table. Pathways enriched with differentially expressed genes. In the link to the detected KEGG pathways, up-regulated and down-regulated genes in HD are highlighted in red. #Gene: Statistics of detected genes (left) and total genes (right) in a pathway. Nested detected pathways: Pathways detected in this analysis and embedded in other detected pathways. (XLSX)

S7 Table. List of genes enriched in detected pathways. (XLSX)

Acknowledgments

We thank Dr. Anshul Kundaje of Stanford University and the NIH Roadmap Epigenomics Consortium (nihroadmap.nih.gov/epigenomics/) for providing chromatin state data. We thank Iris Cheung for conducting ChIP-seq assays and the UMASS flow cytometry and sequencing cores for help with the experiments. This work was supported by the Jerry McDonald HD Research Fund and Public Health Service, National Institutes of Health grants, National Institute of Neurological disorders and stroke (R01 NS073947; Epigenetic Markers in Huntington's disease Brain) National Institute of Mental Health (R01 MH106056; Higher Order Chromatin and Genetic Risk for Schizophrenia), NSF-EARly-concept Grants for Exploratory Research (EAGER) (PHY-1444389), National Institutes of Health (NIH/NIA R01AG050986, Roussos) and the Veterans Affairs (Merit grant BX002395, Roussos). We would like to acknowledge the Harvard Brain Tissue Resource Center McLean Hospital (Belmont, Massachusetts), which is supported in part by PHS contract, HHSN-271-2013-00030C for providing the HD tissue. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: ZW RHM JFC SA. Performed the experiments: JFC SA RHM. Analyzed the data: XD JT AL. Contributed reagents/materials/analysis tools: PR SA. Wrote the paper: XD JT RHM ZW.

References

1. Huntington G. On Chorea. *Med Surg Report*. 1872; 26: 317–321.
2. Myers RH. Huntington's disease genetics. *NeuroRx*. 2004; 1: 255–62. doi: [10.1602/neurorx.1.2.255](https://doi.org/10.1602/neurorx.1.2.255) PMID: [15717026](https://pubmed.ncbi.nlm.nih.gov/15717026/)
3. Vonsattel JP, Myers RH, Stevens TJ, Ferrante RJ, Bird ED, Richardson EP. Neuropathological classification of Huntington's disease. *J Neuropathol Exp Neurol*. 1985; 44: 559–577. doi: [10.1097/00005072-198511000-00003](https://doi.org/10.1097/00005072-198511000-00003) PMID: [2932539](https://pubmed.ncbi.nlm.nih.gov/2932539/)
4. Sotrel A, Paskevich PA, Kiely DK, Bird ED, Williams RS, Myers RH. Morphometric analysis of the prefrontal cortex in Huntington's disease. *Neurology*. 1991; 41: 1117–1123. doi: [10.1212/WNL.41.7.1117](https://doi.org/10.1212/WNL.41.7.1117) PMID: [1829794](https://pubmed.ncbi.nlm.nih.gov/1829794/)
5. Rosas HD, Doros G, Gevorkian S, Malarick K, Reuter M, Couto JP, et al. PRECREST: A phase II prevention and biomarker trial of creatine in at-risk huntington disease. *Neurology*. 2014; 82: 850–857. doi: [10.1212/WNL.000000000000187](https://doi.org/10.1212/WNL.000000000000187) PMID: [24510496](https://pubmed.ncbi.nlm.nih.gov/24510496/)
6. Macdonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, et al. A Novel Gene Containing A Trinucleotide Repeat That Is Expanded and Unstable on Huntingtons-Disease Chromosomes. *Cell*. 1993; 72: 971–983. PMID: [8458085](https://pubmed.ncbi.nlm.nih.gov/8458085/)
7. Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, et al. Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet*. 1993; 4: 387–392. doi: [10.1038/ng0893-387](https://doi.org/10.1038/ng0893-387) PMID: [8401587](https://pubmed.ncbi.nlm.nih.gov/8401587/)

8. Myers RH, MacDonald ME, Koroshetz WJ, Duyao MP, Ambrose CM, Taylor SA, et al. De novo expansion of a (CAG)_n repeat in sporadic Huntington's disease. *Nat Genet.* 1993; 5: 168–173. doi: [10.1038/ng1093-168](https://doi.org/10.1038/ng1093-168) PMID: [8252042](https://pubmed.ncbi.nlm.nih.gov/8252042/)
9. Cha JH. Transcriptional dysregulation in Huntington's disease. *Trends Neurosci.* 2000; 23: 387–392. S0166-2236(00)01609-X [pii] PMID: [10941183](https://pubmed.ncbi.nlm.nih.gov/10941183/)
10. Cha JHJ. Transcriptional signatures in Huntington's disease. *Progress in Neurobiology.* 2007. pp. 228–248. doi: [10.1016/j.pneurobio.2007.03.004](https://doi.org/10.1016/j.pneurobio.2007.03.004) PMID: [17467140](https://pubmed.ncbi.nlm.nih.gov/17467140/)
11. Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, et al. Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet.* 2006; 15: 965–977. doi: [10.1093/hmg/ddl013](https://doi.org/10.1093/hmg/ddl013) PMID: [16467349](https://pubmed.ncbi.nlm.nih.gov/16467349/)
12. Hoss AG, Kartha VK, Dong X, Latourelle JC, Dumitriu A, Hadzi TC, et al. MicroRNAs Located in the Hox Gene Clusters Are Implicated in Huntington's Disease Pathogenesis. *PLoS Genet. Public Library of Science;* 2014; 10: e1004188. doi: [10.1371/journal.pgen.1004188](https://doi.org/10.1371/journal.pgen.1004188) PMID: [24586208](https://pubmed.ncbi.nlm.nih.gov/24586208/)
13. Trushina E, Dyer RB, Badger JD, Ure D, Eide L, Tran DD, et al. Mutant huntingtin impairs axonal trafficking in mammalian neurons in vivo and in vitro. *Mol Cell Biol.* 2004; 24: 8195–8209. doi: [10.1128/MCB.24.18.8195-8209.2004](https://doi.org/10.1128/MCB.24.18.8195-8209.2004) PMID: [15340079](https://pubmed.ncbi.nlm.nih.gov/15340079/)
14. Cattaneo E, Zuccato C, Tartari M. Normal huntingtin function: an alternative approach to Huntington's disease. *Nat Rev Neurosci.* 2005; 6: 919–930. doi: [10.1038/nrn1806](https://doi.org/10.1038/nrn1806) PMID: [16288298](https://pubmed.ncbi.nlm.nih.gov/16288298/)
15. Steffan JS, Bodai L, Pallos J, Poelman M, McCampbell A, Apostol BL, et al. Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in Drosophila. *Nature.* 2001; 413: 739–743. doi: [10.1038/35099568](https://doi.org/10.1038/35099568) PMID: [11607033](https://pubmed.ncbi.nlm.nih.gov/11607033/)
16. Nucifora FC, Sasaki M, Peters MF, Huang H, Cooper JK, Yamada M, et al. Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science (80-).* 2001; 291: 2423–2428. doi: [10.1126/science.1056784](https://doi.org/10.1126/science.1056784) PMID: [11264541](https://pubmed.ncbi.nlm.nih.gov/11264541/)
17. McFarland KN, Das S, Sun TT, Leyfer D, Xia E, Sangrey GR, et al. Genome-wide histone acetylation is altered in a transgenic mouse model of Huntington's disease. *PLoS One.* 2012; 7. doi: [10.1371/journal.pone.0041423](https://doi.org/10.1371/journal.pone.0041423)
18. Sadri-Vakili G, Bouzou B, Benn CL, Kim MO, Chawla P, Overland RP, et al. Histones associated with downregulated genes are hypo-acetylated in Huntington's disease models. *Hum Mol Genet.* 2007; 16: 1293–1306. doi: [10.1093/hmg/ddm078](https://doi.org/10.1093/hmg/ddm078) PMID: [17409194](https://pubmed.ncbi.nlm.nih.gov/17409194/)
19. Ryu H, Lee J, Hagerty SW, Soh BY, McAlpin SE, Cormier KA, et al. ESET/SETDB1 gene expression and histone H3 (K9) trimethylation in Huntington's disease. *Proc Natl Acad Sci U S A.* 2006; 103: 19176–19181. doi: [10.1073/pnas.0606373103](https://doi.org/10.1073/pnas.0606373103) PMID: [17142323](https://pubmed.ncbi.nlm.nih.gov/17142323/)
20. Lee J, Hwang YJ, Shin JY, Lee WC, Wie J, Kim KY, et al. Epigenetic regulation of cholinergic receptor M1 (CHRM1) by histone H3K9me3 impairs Ca²⁺ signaling in Huntington's disease. *Acta Neuropathol.* 2013; 125: 727–739. doi: [10.1007/s00401-013-1103-z](https://doi.org/10.1007/s00401-013-1103-z) PMID: [23455440](https://pubmed.ncbi.nlm.nih.gov/23455440/)
21. Ng CW, Yildirim F, Yap YS, Dalin S, Matthews BJ, Velez PJ, et al. Extensive changes in DNA methylation are associated with expression of mutant huntingtin. *Proc Natl Acad Sci U S A.* 2013; 110: 2354–2359. doi: [10.1073/pnas.1221292110](https://doi.org/10.1073/pnas.1221292110) PMID: [23341638](https://pubmed.ncbi.nlm.nih.gov/23341638/)
22. Lee J, Hwang YJ, Kim KY, Kowall NW, Ryu H. Epigenetic Mechanisms of Neurodegeneration in Huntington's Disease. *Neurotherapeutics.* 2013. pp. 664–676. doi: [10.1007/s13311-013-0206-5](https://doi.org/10.1007/s13311-013-0206-5) PMID: [24006238](https://pubmed.ncbi.nlm.nih.gov/24006238/)
23. Villar-Menéndez I, Blanch M, Tyebji S, Pereira-Veiga T, Albasanz JL, Martín M, et al. Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease. *NeuroMolecular Med.* 2013; 15: 295–309. doi: [10.1007/s12017-013-8219-0](https://doi.org/10.1007/s12017-013-8219-0) PMID: [23385980](https://pubmed.ncbi.nlm.nih.gov/23385980/)
24. Ruthenburg AJ, Allis CD, Wysocka J. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell.* 2007. pp. 15–30. doi: [10.1016/j.molcel.2006.12.014](https://doi.org/10.1016/j.molcel.2006.12.014)
25. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell.* 2007; 129: 823–837. doi: [10.1016/j.cell.2007.05.009](https://doi.org/10.1016/j.cell.2007.05.009) PMID: [17512414](https://pubmed.ncbi.nlm.nih.gov/17512414/)
26. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007; 39: 311–318. doi: [10.1038/ng1966](https://doi.org/10.1038/ng1966) PMID: [17277777](https://pubmed.ncbi.nlm.nih.gov/17277777/)
27. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell.* 2005; 120: 169–181. doi: [10.1016/j.cell.2005.01.001](https://doi.org/10.1016/j.cell.2005.01.001) PMID: [15680324](https://pubmed.ncbi.nlm.nih.gov/15680324/)

28. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, et al. Active genes are tri-methylated at K4 of histone H3. *Nature*. 2002; 419: 407–411. doi: [10.1038/nature01080](https://doi.org/10.1038/nature01080) PMID: [12353038](https://pubmed.ncbi.nlm.nih.gov/12353038/)
29. Okitsu CY, Hsieh JCF, Hsieh C-L. Transcriptional activity affects the H3K4me3 level and distribution in the coding region. *Mol Cell Biol*. 2010; 30: 2933–2946. doi: [10.1128/MCB.01478-09](https://doi.org/10.1128/MCB.01478-09) PMID: [20404096](https://pubmed.ncbi.nlm.nih.gov/20404096/)
30. Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, et al. H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell*. 2014; 158: 673–688. doi: [10.1016/j.cell.2014.06.027](https://doi.org/10.1016/j.cell.2014.06.027) PMID: [25083876](https://pubmed.ncbi.nlm.nih.gov/25083876/)
31. Vashishtha M, Ng CW, Yildirim F, Gipson TA, Kratter IH, Bodai L, et al. Targeting H3K4 trimethylation in Huntington disease. *Proc Natl Acad Sci U S A*. 2013; 110: E3027–E3036. doi: [10.1073/pnas.1311323110](https://doi.org/10.1073/pnas.1311323110) PMID: [23872847](https://pubmed.ncbi.nlm.nih.gov/23872847/)
32. Bai G, Cheung I, Shulha HP, Coelho JE, Li P, Dong X, et al. Epigenetic dysregulation of hairy and enhancer of split 4 (HES4) is associated with striatal degeneration in postmortem Huntington brains. *Hum Mol Genet*. 2015; 24: 1441–56. doi: [10.1093/hmg/ddu561](https://doi.org/10.1093/hmg/ddu561) PMID: [25480889](https://pubmed.ncbi.nlm.nih.gov/25480889/)
33. Shulha HP, Cheung I, Guo Y, Akbarian S, Weng Z. Coordinated Cell Type-Specific Epigenetic Remodeling in Prefrontal Cortex Begins before Birth and Continues into Early Adulthood. *PLoS Genet*. 2013; 9. doi: [10.1371/journal.pgen.1003433](https://doi.org/10.1371/journal.pgen.1003433)
34. Shulha HP, Crisci JL, Reshetov D, Tushir JS, Cheung I, Bharadwaj R, et al. Human-Specific Histone Methylation Signatures at Transcription Start Sites in Prefrontal Neurons. *PLoS Biol*. 2012; 10. doi: [10.1371/journal.pbio.1001427](https://doi.org/10.1371/journal.pbio.1001427)
35. Cheung I, Shulha HP, Jiang Y, Matevossian A, Wang J, Weng Z, et al. Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex. *Proc Natl Acad Sci U S A*. 2010; 107: 8824–8829. doi: [10.1073/pnas.1001702107](https://doi.org/10.1073/pnas.1001702107) PMID: [20421462](https://pubmed.ncbi.nlm.nih.gov/20421462/)
36. Hoss AG, Labadorf A, Latourelle JC, Kartha VK, Hadzi TC, Gusella JF, et al. miR-10b-5p expression in Huntington's disease brain relates to age of onset and the extent of striatal involvement. *BMC Med Genomics*. 2015; 8: 10. doi: [10.1186/s12920-015-0083-3](https://doi.org/10.1186/s12920-015-0083-3) PMID: [25889241](https://pubmed.ncbi.nlm.nih.gov/25889241/)
37. Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, et al. Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol*. 2012; 13: R53. doi: [10.1186/gb-2012-13-9-r53](https://doi.org/10.1186/gb-2012-13-9-r53) PMID: [22950368](https://pubmed.ncbi.nlm.nih.gov/22950368/)
38. Sato S, Omori Y, Katoh K, Kondo M, Kanagawa M, Miyata K, et al. Pikachurin, a dystroglycan ligand, is essential for photoreceptor ribbon synapse formation. *Nat Neurosci*. 2008; 11: 923–931. doi: [10.1038/nn.2160](https://doi.org/10.1038/nn.2160) PMID: [18641643](https://pubmed.ncbi.nlm.nih.gov/18641643/)
39. Flanagan JG, Vanderhaeghen P. The ephrins and Eph receptors in neural development. *Annu Rev Neurosci*. 1998; 21: 309–345. doi: [10.1146/annurev.neuro.21.1.309](https://doi.org/10.1146/annurev.neuro.21.1.309) PMID: [9530499](https://pubmed.ncbi.nlm.nih.gov/9530499/)
40. Wilkinson DG. Multiple roles of EPH receptors and ephrins in neural development. *Nat Rev Neurosci*. 2001; 2: 155–64. PMID: [11256076](https://pubmed.ncbi.nlm.nih.gov/11256076/)
41. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, et al. A novel signaling pathway impact analysis. *Bioinformatics*. 2009; 25: 75–82. doi: [10.1093/bioinformatics/btn577](https://doi.org/10.1093/bioinformatics/btn577) PMID: [18990722](https://pubmed.ncbi.nlm.nih.gov/18990722/)
42. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014; 507: 455–61. doi: [10.1038/nature12787](https://doi.org/10.1038/nature12787) PMID: [24670763](https://pubmed.ncbi.nlm.nih.gov/24670763/)
43. Kvon EZ, Stampfel G, Omar Yáñez-Cuna J, Dickson BJ, Stark A. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev*. 2012; 26: 908–913. doi: [10.1101/gad.188052.112](https://doi.org/10.1101/gad.188052.112) PMID: [22499593](https://pubmed.ncbi.nlm.nih.gov/22499593/)
44. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012. pp. 75–82. doi: [10.1038/nature11232](https://doi.org/10.1038/nature11232)
45. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013; 45: 1238–1243. doi: [10.1038/ng.2756](https://doi.org/10.1038/ng.2756) PMID: [24013639](https://pubmed.ncbi.nlm.nih.gov/24013639/)
46. Phillips JE, Corces VG. CTCF: Master Weaver of the Genome. *Cell*. 2009. pp. 1194–1211. doi: [10.1016/j.cell.2009.06.001](https://doi.org/10.1016/j.cell.2009.06.001) PMID: [19563753](https://pubmed.ncbi.nlm.nih.gov/19563753/)
47. Milavetz BI. SP1 and AP-1 elements direct chromatin remodeling in SV40 chromosomes during the first 6 hours of infection. *Virology*. 2002; 294: 170–179. doi: [10.1006/viro.2001.1308](https://doi.org/10.1006/viro.2001.1308) PMID: [11886275](https://pubmed.ncbi.nlm.nih.gov/11886275/)
48. Tan NY, Khachigian LM. Sp1 phosphorylation and its regulation of gene transcription. *Mol Cell Biol*. 2009; 29: 2483–8. doi: [10.1128/MCB.01828-08](https://doi.org/10.1128/MCB.01828-08) PMID: [19273606](https://pubmed.ncbi.nlm.nih.gov/19273606/)
49. Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2015; 518: 317–330. doi: [10.1038/nature14248](https://doi.org/10.1038/nature14248) PMID: [25693563](https://pubmed.ncbi.nlm.nih.gov/25693563/)

50. Schmitges FW, Prusty AB, Faty M, Stützer A, Lingaraju GM, Aiwazian J, et al. Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol Cell*. 2011; 42: 330–341. doi: [10.1016/j.molcel.2011.03.025](https://doi.org/10.1016/j.molcel.2011.03.025) PMID: [21549310](https://pubmed.ncbi.nlm.nih.gov/21549310/)
51. Hadzi TC, Hendricks AE, Latourelle JC, Lunetta KL, Cupples LA, Gillis T, et al. Assessment of cortical and striatal involvement in 523 Huntington disease brains. *Neurology*. 2012; 79: 1708–1715. doi: [10.1212/WNL.0b013e31826e9a5d](https://doi.org/10.1212/WNL.0b013e31826e9a5d) PMID: [23035064](https://pubmed.ncbi.nlm.nih.gov/23035064/)
52. Jiang Y, Matevosian A, Huang H-S, Straubhaar J, Akbarian S. Isolation of neuronal chromatin from brain tissue. *BMC Neurosci*. 2008; 9: 42. doi: [10.1186/1471-2202-9-42](https://doi.org/10.1186/1471-2202-9-42) PMID: [18442397](https://pubmed.ncbi.nlm.nih.gov/18442397/)
53. Valor LM, Guiretti D, Lopez-Atalaya JP, Barco A. Genomic landscape of transcriptional and epigenetic dysregulation in early onset polyglutamine disease. *J Neurosci*. 2013; 33: 10471–82. doi: [10.1523/JNEUROSCI.0670-13.2013](https://doi.org/10.1523/JNEUROSCI.0670-13.2013) PMID: [23785159](https://pubmed.ncbi.nlm.nih.gov/23785159/)
54. Ravache M, Weber C, Mérienne K, Trottier Y. Transcriptional activation of REST by Sp1 in huntington's disease models. *PLoS One*. 2010; 5. doi: [10.1371/journal.pone.0014311](https://doi.org/10.1371/journal.pone.0014311)
55. Qiu Z, Norflus F, Singh B, Swindell MK, Buzescu R, Bejarano M, et al. Sp1 is up-regulated in cellular and transgenic models of huntington disease, and its reduction is neuroprotective. *J Biol Chem*. 2006; 281: 16672–16680. doi: [10.1074/jbc.M511648200](https://doi.org/10.1074/jbc.M511648200) PMID: [16595660](https://pubmed.ncbi.nlm.nih.gov/16595660/)
56. Dunah AW, Jeong H, Griffin A, Kim Y-M, Standaert DG, Hersch SM, et al. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science*. 2002; 296: 2238–2243. doi: [10.1126/science.1072613](https://doi.org/10.1126/science.1072613) PMID: [11988536](https://pubmed.ncbi.nlm.nih.gov/11988536/)
57. Zuccato C, Belyaev N, Conforti P, Ooi L, Tartari M, Papadimou E, et al. Widespread disruption of repressor element-1 silencing transcription factor/neuron-restrictive silencer factor occupancy at its target genes in Huntington's disease. *J Neurosci*. 2007; 27: 6972–6983. doi: [10.1523/JNEUROSCI.4278-06.2007](https://doi.org/10.1523/JNEUROSCI.4278-06.2007) PMID: [17596446](https://pubmed.ncbi.nlm.nih.gov/17596446/)
58. Zuccato C, Tartari M, Crotti A, Goffredo D, Valenza M, Conti L, et al. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nat Genet*. 2003; 35: 76–83. doi: [10.1038/ng1219](https://doi.org/10.1038/ng1219) PMID: [12881722](https://pubmed.ncbi.nlm.nih.gov/12881722/)
59. Sugars KL, Rubinsztein DC. Transcriptional abnormalities in Huntington disease. *Trends in Genetics*. 2003. pp. 233–238. doi: [10.1016/S0168-9525\(03\)00074-X](https://doi.org/10.1016/S0168-9525(03)00074-X) PMID: [12711212](https://pubmed.ncbi.nlm.nih.gov/12711212/)
60. Biagioli M, Ferrari F, Mendenhall EM, Zhang Y, Erdin S, Vijayvargia R, et al. Htt CAG repeat expansion confers pleiotropic gains of mutant huntingtin function in chromatin regulation. *Hum Mol Genet*. 2015; 24: 2442–57. doi: [10.1093/hmg/ddv006](https://doi.org/10.1093/hmg/ddv006) PMID: [25574027](https://pubmed.ncbi.nlm.nih.gov/25574027/)
61. Seong IS, Woda JM, Song J-J, Lloret A, Abeyathne PD, Woo CJ, et al. Huntingtin facilitates polycomb repressive complex 2. *Hum Mol Genet*. 2010; 19: 573–83. doi: [10.1093/hmg/ddp524](https://doi.org/10.1093/hmg/ddp524) PMID: [19933700](https://pubmed.ncbi.nlm.nih.gov/19933700/)
62. Peña P V., Hom RA, Hung T, Lin H, Kuo AJ, Wong RPC, et al. Histone H3K4me3 Binding Is Required for the DNA Repair and Apoptotic Activities of ING1 Tumor Suppressor. *J Mol Biol*. 2008; 380: 303–312. doi: [10.1016/j.jmb.2008.04.061](https://doi.org/10.1016/j.jmb.2008.04.061) PMID: [18533182](https://pubmed.ncbi.nlm.nih.gov/18533182/)
63. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, et al. A high-resolution map of active promoters in the human genome. *Nature*. 2005; 436: 876–880. doi: [10.1038/nature03877](https://doi.org/10.1038/nature03877) PMID: [15988478](https://pubmed.ncbi.nlm.nih.gov/15988478/)
64. Koch CM, Andrews RM, Flicek P, Dillon SC, Karaoz U, Clelland GK, et al. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Research*. 2007. pp. 691–707. doi: [10.1101/gr.5704207](https://doi.org/10.1101/gr.5704207) PMID: [17567990](https://pubmed.ncbi.nlm.nih.gov/17567990/)
65. Tae HK, Barrera LO, Qu C, Van Calcar S, Trinklein ND, Cooper SJ, et al. Direct isolation and identification of promoters in the human genome. *Genome Res*. 2005; 15: 830–839. doi: [10.1101/gr.3430605](https://doi.org/10.1101/gr.3430605) PMID: [15899964](https://pubmed.ncbi.nlm.nih.gov/15899964/)
66. Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, et al. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev*. 2012; 26: 1714–1728. doi: [10.1101/gad.194209.112](https://doi.org/10.1101/gad.194209.112) PMID: [22855832](https://pubmed.ncbi.nlm.nih.gov/22855832/)
67. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, et al. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO Journal*. 2011. pp. 4198–4210. doi: [10.1038/emboj.2011.295](https://doi.org/10.1038/emboj.2011.295) PMID: [21847099](https://pubmed.ncbi.nlm.nih.gov/21847099/)
68. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007; 448: 553–560. doi: [10.1038/nature06008](https://doi.org/10.1038/nature06008) PMID: [17603471](https://pubmed.ncbi.nlm.nih.gov/17603471/)
69. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009; 458: 223–227. doi: [10.1038/nature07672](https://doi.org/10.1038/nature07672) PMID: [19182780](https://pubmed.ncbi.nlm.nih.gov/19182780/)

70. Connor C, Cheung I, Simon A, Jakovcevski M, Weng Z, Akbarian S. A simple method for improving the specificity of anti-methyl histone antibodies. *Epigenetics*. 2010; 5: 392–395. doi: [10.4161/epi.5.5.11874](https://doi.org/10.4161/epi.5.5.11874) PMID: [20458167](https://pubmed.ncbi.nlm.nih.gov/20458167/)
71. Heidari N, Phanstiel DH, He C, Grubert F, Jahanbanian F, Kasowski M, et al. Genome-wide map of regulatory interactions in the human genome. *Genome Res*. 2014; gr.176586.114–. doi: [10.1101/gr.176586.114](https://doi.org/10.1101/gr.176586.114)
72. Bharadwaj R, Jiang Y, Mao W, Jakovcevski M, Dincer A, Krueger W, et al. Conserved Chromosome 2q31 Conformations Are Associated with Transcriptional Regulation of {GAD1} {GABA} Synthesis Enzyme and Altered in Prefrontal Cortex of Subjects with Schizophrenia. *J Neurosci*. 2013; 33: 11839–11851. doi: [10.1523/JNEUROSCI.1252-13.2013](https://doi.org/10.1523/JNEUROSCI.1252-13.2013) PMID: [23864674](https://pubmed.ncbi.nlm.nih.gov/23864674/)
73. Bharadwaj R, Peter CJ, Jiang Y, Roussos P, Vogel-Ciernia A, Shen EY, et al. Conserved higher-order chromatin regulates NMDA receptor gene expression and cognition. *Neuron*. 2014; 84: 997–1008. doi: [10.1016/j.neuron.2014.10.032](https://doi.org/10.1016/j.neuron.2014.10.032) PMID: [25467983](https://pubmed.ncbi.nlm.nih.gov/25467983/)
74. Unland R, Kerl K, Schlosser S, Farwick N, Plagemann T, Lechtape B, et al. Epigenetic repression of the dopamine receptor D4 in pediatric tumors of the central nervous system. *J Neurooncol*. 2014; 116: 237–249. doi: [10.1007/s11060-013-1313-1](https://doi.org/10.1007/s11060-013-1313-1) PMID: [24264533](https://pubmed.ncbi.nlm.nih.gov/24264533/)
75. Shulha HP, Cheung I, Whittle C, Wang J, Virgil D, Lin CL, et al. Epigenetic Signatures of Autism: Trimethylated H3K4 Landscapes in Prefrontal Neurons. *Archives of General Psychiatry*. 2012. pp. 314–324. doi: [10.1001/archgenpsychiatry.2011.151](https://doi.org/10.1001/archgenpsychiatry.2011.151) PMID: [22065254](https://pubmed.ncbi.nlm.nih.gov/22065254/)
76. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinforma*. 2010; doi: [10.1002/0471250953.bi1107s32](https://doi.org/10.1002/0471250953.bi1107s32)
77. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008; 9: R137. doi: [10.1186/gb-2008-9-9-r137](https://doi.org/10.1186/gb-2008-9-9-r137) PMID: [18798982](https://pubmed.ncbi.nlm.nih.gov/18798982/)
78. Li C, Wong W, Hung Wong W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol*. 2001; 2: RESEARCH0032. Available: <http://www.biomedcentral.com/content/pdf/gb-2001-2-8-research0032.pdf> PMID: [11532216](https://pubmed.ncbi.nlm.nih.gov/11532216/)
79. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15: 550. doi: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8) PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
80. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res*. 2012; 22: 1760–1774. doi: [10.1101/gr.135350.111](https://doi.org/10.1101/gr.135350.111) PMID: [22955987](https://pubmed.ncbi.nlm.nih.gov/22955987/)
81. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842. doi: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/)
82. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. 2010; 11: R14. doi: [10.1186/gb-2010-11-2-r14](https://doi.org/10.1186/gb-2010-11-2-r14) PMID: [20132535](https://pubmed.ncbi.nlm.nih.gov/20132535/)
83. Supek F, Bošnjak M, Škunca N, Šmuc T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011; 6. doi: [10.1371/journal.pone.0021800](https://doi.org/10.1371/journal.pone.0021800)
84. Zhang JD, Wiemann S. KEGGgraph: A graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics*. 2009; 25: 1470–1471. doi: [10.1093/bioinformatics/btp167](https://doi.org/10.1093/bioinformatics/btp167) PMID: [19307239](https://pubmed.ncbi.nlm.nih.gov/19307239/)
85. James Kent W, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002; 12: 996–1006. doi: [10.1101/gr.229102](https://doi.org/10.1101/gr.229102) Article published online before print in May 2002 PMID: [12045153](https://pubmed.ncbi.nlm.nih.gov/12045153/)
86. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res*. Cold Spring Harbor Lab; 2012; 22: 1798–1812. doi: [10.1101/gr.139105.112](https://doi.org/10.1101/gr.139105.112) PMID: [22955990](https://pubmed.ncbi.nlm.nih.gov/22955990/)
87. Wang J, Zhuang J, Iyer S, Lin X-Y, Greven MC, Kim B-H, et al. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res*. Oxford Univ Press; 2012;
88. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013; 153: 307–319. doi: [10.1016/j.cell.2013.03.035](https://doi.org/10.1016/j.cell.2013.03.035) PMID: [23582322](https://pubmed.ncbi.nlm.nih.gov/23582322/)