PNAS
www.pnas.org

Supplementary Information for

Early epigenomic and transcriptional changes reveal Elk-1 transcription factor as a therapeutic target in Huntington's disease

Ferah Yildirim, Christopher W. Ng, Vincent Kappes, Tobias Ehrenberger, Siobhan K Rigby, Victoria Stivanello, Theresa A. Gipson, Anthony R. Soltis, Peter Vanhoutte, Jocelyne Caboche, David E. Housman and Ernest Fraenkel

Corresponding Authors:

Ferah Yildirim

ferah.yildirim@charite.de

David E. Housman

dhousman@mit.edu

Ernest Fraenkel

fraenkel@mit.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1908113116

Supplementary Figure Legends

Supplementary Figure 1. Assessment of body weight and clasping behavior in the R6/1 and CHL2 mice, and overlap of striatal transcriptional changes in HD mice and HD patients (related to Figure 1). (A and B) Venn Diagrams showing overlaps of differentially-expressed genes between the caudate nucleus of patient HD (Kuhn et al., 2007) and the striata of 8- weeks old R6/1 (A) and 1 year old CHL2 mice (B) (Hypergeometric test). (C) Body weight and (D) clasping reflex assessments in the R6/1 mice, (E) Body weight assessment in the CHL2 mice. n = 6-8 mice/genotype/group. Data values presented as mean +/- SD. *p; Significant effect of genotype, *p < 0.05, **p < 0.01, ***p < 0.001, ****p<0.0001. Student's t test.

Supplementary Figure 2. Histone H3K27ac occupancy at distal and proximal regulatory sites and distribution of upregulated genes in R6/1 striatum across five classes (related to Figure 2). (A) Pie chart showing the distribution of histone H3K27ac sites in WT striatum. (B) The numbers of genes with increased expression expected and observed in each class are shown for 8 week old R6/1 striatum.

Supplementary Figure 3. Elk-1 expression in R6/1 (at 4 and 8 weeks of age) and CHL2 (at 1 year of age) mice striata and distribution of Elk-1 ChIP-seq peaks in WT and differential Elk-1 ChIP-seq peaks R6/1 striata (related to Figure 3). (A) Bar graph showing Elk-1 mRNA expression levels (fpkm) in 4- and 8-weeks old R6/1 and 1 year old CHL2 striata. (B) Western blotting for Elk-1 showing Elk-1 protein levels in the striata of 8-weeks old R6/1 and WT mice. n=3-4 mice. (C) Pie chart showing distribution of Elk-1 ChIP-seq peaks in WT mouse striatum as well as differential Elk-1 ChIP-seq peaks in R6/1 striatum compared to WT mice. (D) UCSC Browser image representing normalized Elk-1

and H3K27ac ChIP-Seq read density from 8 week old R6/1 and respective WT mice mapped at *Egr1* example gene locus for proximal and distal regulatory site binding by Elk-1.

Supplementary Figure 4. AAV-mediated Elk-1 overexpression in R6/1 striatum and its impact on key dysregulated transcripts in HD (related to Figure 5). 5-6 weeks old R6/1 and wild-type (WT) mice received unilateral striatal injections of either Elk-1 AAV or mCherry control AAV and were sacrificed at 10 weeks of age followed by tissue collection for gene expression analyses. (A) Quantitative RT-PCR confirmation of the expression of several key dysregulated genes (*Drd1a, Drd2, Egr-1, Arc, Ppp1r1b, Lhx9, Hap1, Agt, Polr2a, Rbm28, Aqp1*) in R6/1 and WT mice as well as of *Elk-1* overexpression by Elk-1-expressing AAV (n = 3-5 mice/genotype/group). (B) Quantitative RT-PCR for *mCherry* (AAV-mediated gene expression) and *Iba1* (inflammatory marker) in the ipsilateral and contralateral striata of Elk-1 AAV and control AAV-treated R6/1 and WT mice. Data presented as mean ± SEM. (C) Western blotting for Elk-1 showing Elk-1 AAV-mediated Elk-1 overexpression in the ipsilateral (injected) striatum. n=2 mice. (D) Fluorescent microscopy images of coronal brain sections showing AAV-mediated gene expression (detected by mCherry signal) in mouse dorsal striatum.

Supplementary Datasets

Dataset S1. Number of differentially expressed genes in the striata of R6/1 and CHL2 mice.

Dataset S2. Enrichment of Gene ontology terms among the differentially expressed genes in the striata of R6/1 and CHL2 mice.

Dataset S3. Overlap of differentially expressed genes between the striata of R6/1 and CHL2 mice with differentially expressed genes in the caudate nuclei of postmortem brains from HD patients.

Dataset S4. Gene ontology enrichments for genes with highest H3K27acetylation signal (top 1000 genes by tags) in striata of WT mice.

Dataset S5. Gene ontology enrichments for genes with differential H3K27acetylation in the striata of R6/1.

Dataset S6. Enrichment of Gene ontology terms among genes with Class 1 H3K27ac profiles.

Dataset S7. Motifs associated with differential H3K27acetylation in R6/1 striatum.

Dataset S8. Gene ontology enrichments for the top genes bound by Elk-1 in the striata of WT mice.

Dataset S9. Gene ontology enrichments for genes with differential Elk-1 binding in the striata of R6/1 mice.

Dataset S10. List of differentially expressed genes in the striata of 8 weeks old R6/1 mice, that are differentially-bound by Elk-1.

Dataset S11. Motifs associated with the top genes bound by Elk-1 in the striata of WT mice.

Dataset S12. Motifs associated with the differential Elk-1 binding in the striata of R6/1 mice.

Dataset S13. List of differentially expressed genes and their expression levels in R6/1 striatum upon Elk-1 AAV treatment of animals.

Dataset S14. Gene set enrichment analysis results for the differentially expressed genes in R6/1 striatum upon Elk-1 AAV treatment of animals.

Supplementary Materials and Methods

Animals. All animal procedures were performed according to protocols approved by the Institutional Animal Care and Use Committee at Massachusetts Institute of Technology (approved animal experiment proposal number; 0212-014-15). The first breeders of both the R6/1 and CHL2 colonies were obtained from the Jackson Lab. Hemizygous R6/1 mice were bred by crossing R6/1 males to C57BL/6 females. CHL2 mice were bred by crossing CHL2 males to C57BL/6 females. The mice were maintained on a 12-hr light/dark cycle with food and water available ad libitum. All animal procedures including sacrifice for tissue harvest for RNA-seq and ChIP-seq experiments, stereotaxic surgeries and behavioral testing were performed during the animals' light cycle (between the hours of 7 a.m. and 7 p.m.). All efforts were made to minimize the number and suffering of animals used. For RNA-seq and ChIP-seq experiments, R6/1, CHL2 and their wildtypes were killed using

CO2 asphyxiation, which was followed immediately by postmortem dissection of the striatum. Flash-frozen tissues for RNA-seq and cross-linked tissues for ChIP-seq were stored at −80 °C until use.

RNA Sequencing (RNA-seq). Total RNA was extracted from brain tissues using RNeasy kit (Qiagen) and mRNA was isolated from DNA-free total RNA using an Illumina-supplied mRNA Purification Kit. Briefly, oligo(dT) conjugated beads were used to pull-down poly-A containing mRNA and mRNA fragmentation was done by using divalent cations. First and second strand cDNA was synthesized using random primers, end-repaired, adenylated and ligated with paired-end adapters (Illumina, catalog no. PE102-1004). The cDNA library was size-selected to 200-400bp centering on 300bp via gel electrophoresis, followed by PCR amplification. The final RNA-seq libraries were sequenced on an Illumina platform. After sequencing, the raw paired-end reads of the cDNA fragments were aligned to the mouse (RefSeq, mm9) transcriptome using a spliced junction discovery tool, TopHat [1], and short-read alignment tool, Bowtie [2]. Differential expression across the biological replicates in each condition were called using the Cufflinks algorithm [3] with bias and multi-read correction options. Reads mapping to rRNA and mitochondrial transcripts were masked. Genes meeting a minimum threshold of 0.1 FPKM in at least one of the conditions were tested for differential expression and used as a background for functional enrichment of differentially expressed genes in Gorilla [4]. Pairwise log2 fold changes in expression were calculated from the FPKM levels at every gene after pseudocounts of 0.05 FPKM were added to every gene to mitigate inflated fold changes at poorly expressed genes. Genes were called differentially expressed if the p-value was less

than 0.05 after multiple hypothesis correction and the absolute log fold change was greater than 0.5.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq). Crosslinked tissues were fragmented to the size range of 100–500 bp using a Bioruptor (Bioruptor Next Gen; Diagenode). Antibodies that specifically recognizes H3K27 acetylation (Abcam, ab4729) and Elk-1 protein (Santa Cruz, sc-355X) along with nonspecific rabbit IgG (Millipore, 17-614) were incubated with beads for 6 h before incubating with sonicated chromatin overnight. Resulting immunoprecipitated DNA and nonspecific IgG-bound DNA were prepared for high-throughput sequencing according to Illumina's instructions and sequenced on an Illumina platform following the manufacturer's standard protocol. Raw ChIP-seq data were processed using the Illumina software pipeline. ChIP-seq reads were aligned to the reference mouse genome (mm9; UCSC) using Bowtie [2]. Binding events were identified using the GPS (genome positioning system) [5] and MACS (model-based analysis of ChIP-seq) [6] algorithms. IgG-bound DNA libraries were used as the control. We used a calculated alignable genome size of 2.107 Gbp, a standard expected ChIP-seq read distribution, a multiple hypothesis corrected enrichment q-value cutoff of 1e-2, and a minimum α-value of 30. Genes associated with binding events inferred from ChIPseq were identified using annotations from the refFlat (RefSeq database) file from the UCSC mm9 tables. For visualization of read density at specific loci, ChIP-seq aligned reads were shifted according to a peak shift model built by MACS [6] and uploaded to the UCSC genome browser.

Gene Ontology. Functional enrichments in gene ontology biological processes were calculated using the two unranked list approach in GOrilla [4]. For enrichment within

differentially expressed genes, all expressed genes [above fragments per kilobase of exon per million fragments mapped (FPKM) > 0.1] were used as background. For enrichment within differentially H3K27acetylated and classes of H3K27ac genes based on TSS profile, all H3K27ac genes (above 50 tags in the −2/+3-kb TSS window) and for differentially Elk-1-bound genes, all Elk-1-bound genes (above 50 tags in the −/+10-kb TSS window) were used as background.

GSEA. To determine gene sets associated with preferentially up- or down-regulated genes, we used an in-house implementation of single sample Gene Set Enrichment Analysis (ssGSEA) [7, 8] on the log2-based fold change values between HD conditions and wild-types on the full data sets (containing all genes with measurements across all groups). The following parameter settings were used for this analysis: rank-based sample scoring ("sample.norm.type": ranks), area under curve statistic to calculate scores ("statistic": area.under.RES), default weight (0.75), 10k permutations, z-score pre-processing ("correl.type": z.score), and normalized enrichment score output ("NES"). We used the gene sets available in the Molecular Signature Database (MSigDB v6.1) [9, 10], considering canonical pathways (C2CP), GO terms (C5), and motif gene sets (C3). Prior to running ssGSEA, mouse gene symbols were converted to human gene symbols using the OrthoRetriever web tool (http://lighthouse.ucsf.edu/orthoretriever/) to match the gene set definitions.

Motif Analysis. Sequences mapping to a ±100-bp window around each GPS binding event was used as input for analysis of sequence motifs. We used a hypothesis-based approach to identify known protein–DNA recognition elements enriched in each dataset. The set of hypotheses is derived position-specific scoring matrices (PSSMs) from TRANSFAC [11]

filtered for sufficient information content (>8 total bits). Because many of these motifs are very similar to each other, they were clustered based on pairwise distance by KL divergence of the PSSMs using Affinity Propagation. When presenting the results of the motif analysis, we show the motif within each cluster that had the most significant P value. The TAMO programming environment [12] was used to store the PSSMs and calculate the max motif score for each sequence (across all k-mers in the sequence for a motif of width k). Overrepresentation of motifs in a foreground set of sequences was assessed against a background set of sequences using the Mann–Whitney Wilcoxon ranked sum test. For each independent motif test, sequences were ranked by the maximum motif score in each sequence (across all k-mers in the sequence for a motif of width k). This ranked list was used to compute the U statistic from which we computed a P value. The background sequences were selected to match the GC content, CpG content, and distance to the TSS of each foreground set. We also used a regression procedure to identify motifs significantly associated with gene expression, as performed previously [13]. Briefly, we scanned H3K27ac sites near the transcription start sites of differentially expressed genes for motif matches and, for each motif against each gene, computed a transcription factor affinity (TFA) score as a distance-weighted sum of motif match scores. For each motif, we regressed the gene TFA scores against their corresponding expression levels, standardizing (i.e. z-scoring) both the predictor and response vectors prior to the analysis. We used the t-distribution to compute the statistical significance of each regression coefficient and corrected raw p-values for multiple hypotheses using the Benjamini-Hochberg FDR procedure.

Cell culture and transfection with wildtype and mutant *Htt* exon 1 and *Elk-1* plasmids. Primary striatal neuron cultures were prepared as previously described [14]. Briefly, embryos were removed at day 14 from timed-pregnant Swiss mice (Janvier, Le Genest-St-Isle, France). Striata were dissected and mechanically dissociated by gently pipetting in neurobasal medium (GIBCO/BRL). After decantation for 5 min at room temperature to eliminate tissue debris, cells were collected by centrifugation at 1,000´ g for 5 min. Cell pellets were resuspended in neurobasal medium containing B27 supplement (Invitrogen), 500 mM L-glutamine (Invitrogen), 100 units/ml penicillin-streptomycin (Invitrogen) and 25 mM 2-mercaptoethanol (Sigma). Cells were then plated at a density of 1,000 cells per mm2 into Nunc multiple-well plates coated with 50 mg/ml poly-D-lysine (Sigma). Cells were cultured in the complete neurobasal medium at 37°C with 5% CO2. After 3 days, cell culture medium was changed, and all experiments were performed 7 days after dissection, when most of the cells were of neuronal phenotype and postmitotic. cDNAs for huntingtin (*Htt*) and expanded *Htt* (exp*Htt*) were provided by the Huntington's Disease Foundation Resource Bank (University of California, Los Angeles). pcDNA3.1 expression plasmids containing the entire exon 1 sequence for the human huntingtin gene (IT15), with either 25 or 103 continuous CAA or CAG repeats, were engineered behind a CMV promoter. A sequence encoding an EGFP was inserted in frame at the C terminus of each construct. In parallel, the coding sequence of the EGFP gene was expressed under the control of the CMV promoter (referred to as GFP). Lipofectamine 2000 reagent was used to transfect Striatal neurons with cDNA plasmids coding human exon1-*HTT* with 25 (*Htt*) or 103 polyglutamine (Exp-*HTT*) stretch, according to previously established protocols in the lab [14, 15]. Co-transfection was performed using cDNA constructs encoding for HA-tagged

Elk-1 either in its wild-type (Elk-1 wt) or mutated (3R-Asp–Elk-1 or Elk-1 mut) versions. Thus, Ser383 and Ser389 were mutated into constitutively active sites for phosphorylation (Ser to Asp383/389) [16] and dominant negative for sumoylation (Lys to Arg230/249/254) [17].

RNA extraction and quantitative RT-PCR. Total RNA was isolated by RNeasy kit (Qiagen) from striatum from R6/1 mice brain tissues. 500 ng of total RNA was used for cDNA synthesis using MMLV reverse transcriptase (Promega), and quantitative PCR was performed with SYBR Select Select Master Mix (Life Technologies) detected by QuantStudio3 PCR System (Applied Biosystems). Each sample was run in duplicate. Relative gene expression was determined by using the ddCT method and normalizing to Glyceraldehyde 3-phosphate dehydrogenase (Gapdh) and β-Actin mRNA levels as two independent housekeeping genes. Primers for qPCR are listed below.

Drd1a_F ATGGCTCCTAACACTTCTACCA, Drd1a_R GGGTATTCCCTAAGAGAGTGGAC; Drd2_F AAGCGTCGGAAGCGGGTCAAC, Drd2_R TCGGCGGGCAGCATCCATTCT; Ppp1r1b_F CCAGAAACCCACTCTGTCCC, Ppp1r1b_R GGCTTCAGCCAAAGCAAACA; Egr1_F TATGAGCACCTGACCACAGAGTCC, Egr1_R CGAGTCGTTTGGCTGGGATAAC, Arc_F TACCGTTAGCCCCTATGCCATC, Arc_R TGATATTGCTGAGCCTCAACTG, Elk-1_F CAGGAATGACAGGCCAAGGT, Elk-1_R GGTGGGGTTAGGATAACCTGC.

Statistics. For the next generation DNA sequencing assays (RNA-seq, ChIP-seq) statistics were conducted as described in respective sections. For the rest of the studies, all data are expressed as mean ± SEM. For studies that used two groups, the Student t test was used

for statistical comparison. For comparisons of more than two groups, one-way analysis of variance (ANOVA) was used followed by the Tukey or Sidak post-hoc test (Prism; GraphPad Software, San Diego, CA) as indicated in the respective figure legends. $P < 0.05$ was considered a statistically significant difference.

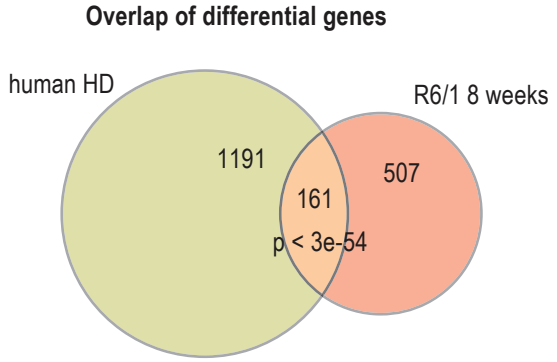
References

1. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
2. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
3. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.
4. Eden, E., et al., *GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.* BMC Bioinformatics, 2009. **10**: p. 48.
5. Guo, Y., et al., *Discovering homotypic binding events at high spatial resolution.* Bioinformatics, 2010. **26**(24): p. 3028-34.
6. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
7. Barbie, D.A., et al., *Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1.* Nature, 2009. **462**(7269): p. 108-12.
8. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
9. Liberzon, A., *A description of the Molecular Signatures Database (MSigDB) Web site.* Methods Mol Biol, 2014. **1150**: p. 153-60.
10. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.
11. Wingender, E., et al., *TRANSFAC: a database on transcription factors and their DNA binding sites.* Nucleic Acids Res, 1996. **24**(1): p. 238-41.
12. Gordon, D.B., et al., *TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs.* Bioinformatics, 2005. **21**(14): p. 3164-5.
13. Soltis, A.R., et al., *Hepatic Dysfunction Caused by Consumption of a High-Fat Diet.* Cell Rep, 2017. **21**(11): p. 3317-3328.
14. Charvin, D., et al., *Unraveling a role for dopamine in Huntington's disease: the dual role of reactive oxygen species and D2 receptor stimulation.* Proc Natl Acad Sci U S A, 2005. **102**(34): p. 12218-23.
15. Garcia, M., et al., *The mitochondrial toxin 3-nitropropionic acid induces striatal neurodegeneration via a c-Jun N-terminal kinase/c-Jun module.* J Neurosci, 2002. **22**(6): p. 2174-84.
16. Vanhoutte, P., et al., *Opposing roles of Elk-1 and its brain-specific isoform, short Elk-1, in nerve growth factor-induced PC12 differentiation.* J Biol Chem, 2001. **276**(7): p. 5189-96.

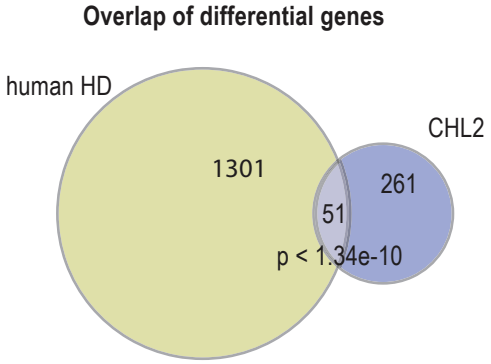
17. Salinas, S., et al., *SUMOylation regulates nucleo-cytoplasmic shuttling of Elk-1.* J Cell Biol, 2004. **165**(6): p. 767-73.

Supplementary figure 1

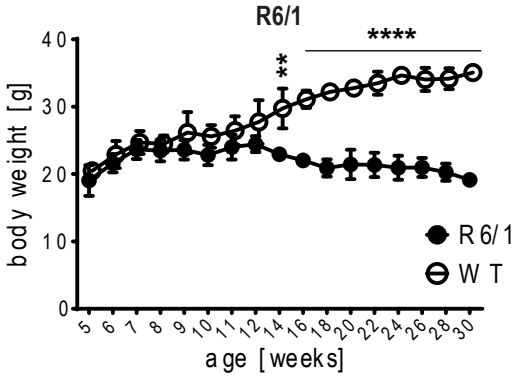
A



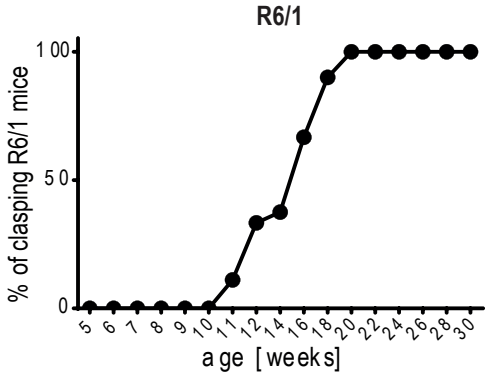
B



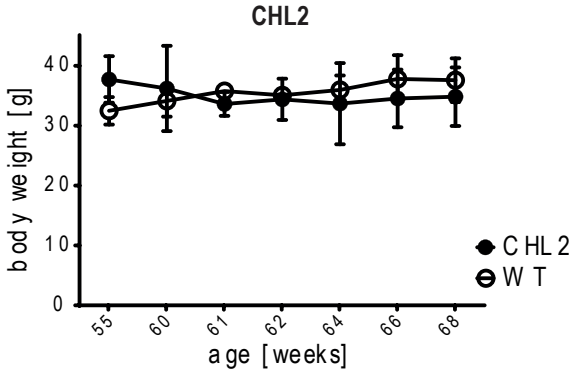
C



D

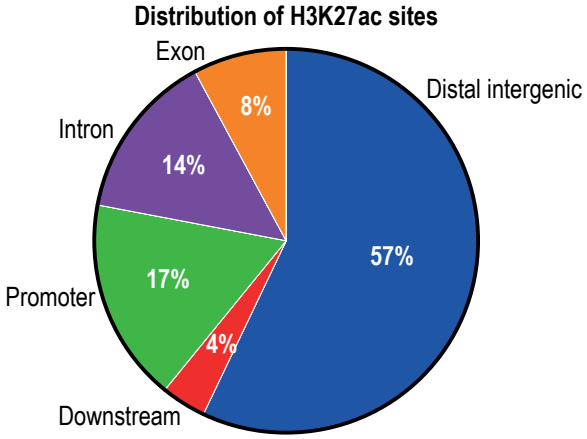


E

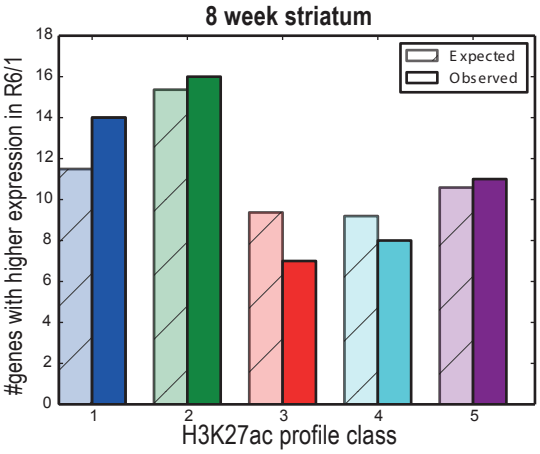


Supplementary figure 2

A

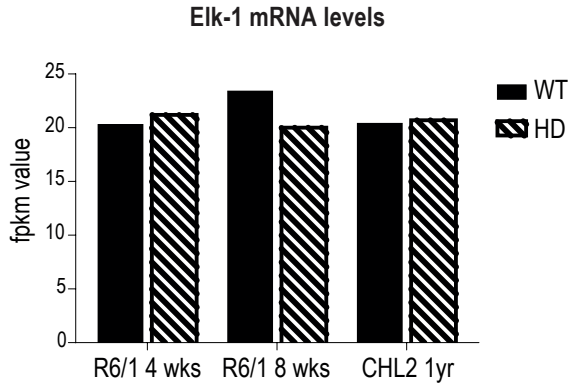


B

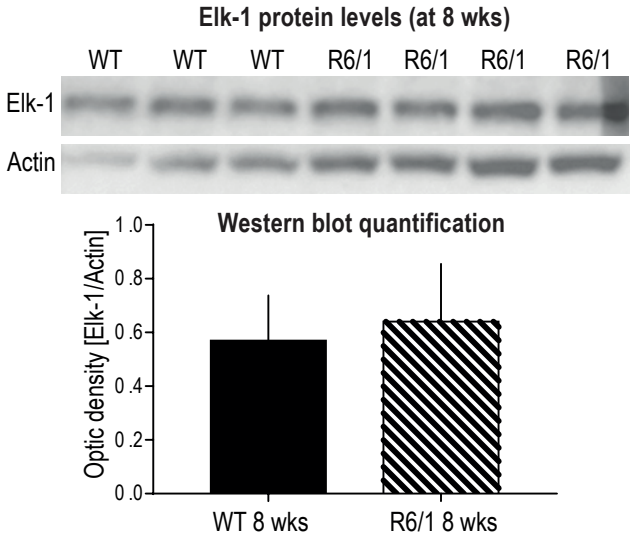


Supplementary figure 3

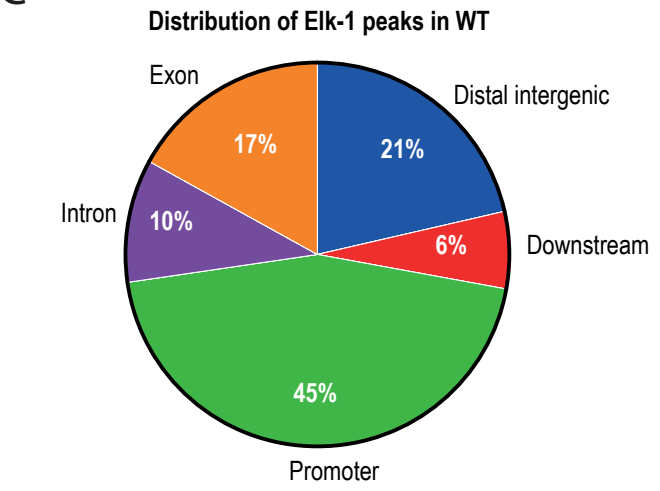
A

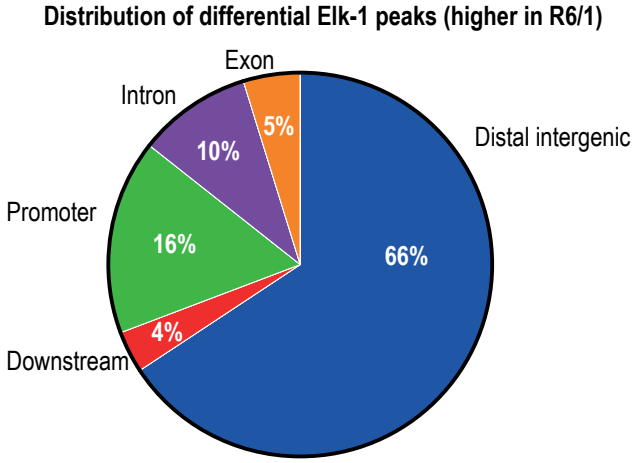


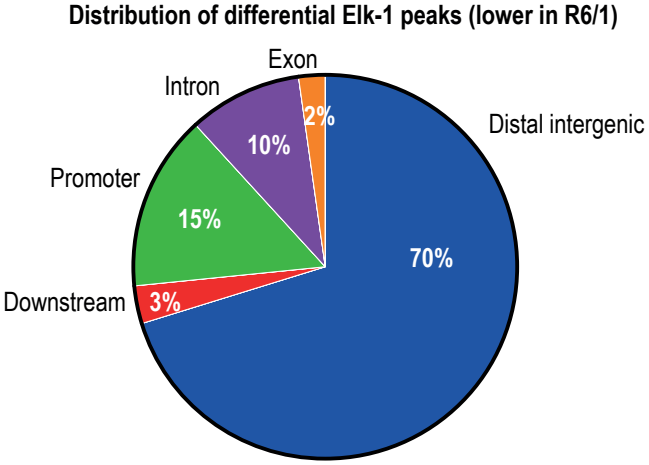
B



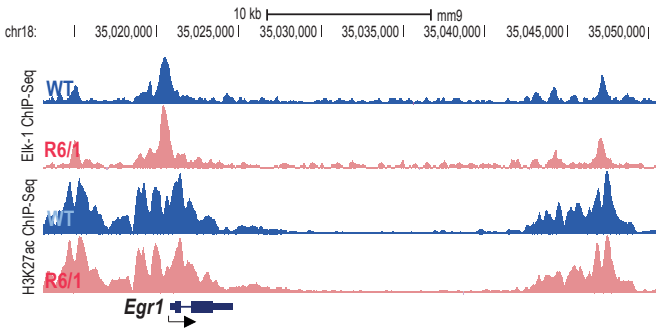
C







D



Supplementary figure 4