

ORIGINAL ARTICLE

High resolution time-course mapping of early transcriptomic, molecular and cellular phenotypes in Huntington's disease CAG knock-in mice across multiple genetic backgrounds

Seth A. Ament^{1,2,†}, Jocelynn R. Pearl^{1,3,†}, Andrea Grindeland^{4,†}, Jason St. Claire⁵, John C. Earls^{1,6}, Marina Kovalenko⁵, Tammy Gillis⁵, Jayalakshmi Mysore⁵, James F. Gusella⁵, Jong-Min Lee⁵, Seung Kwak⁷, David Howland⁷, Min Young Lee¹, David Baxter¹, Kelsey Scherler¹, Kai Wang¹, Donald Geman⁸, Jeffrey B. Carroll⁹, Marcy E. MacDonald⁵, George Carlson⁴, Vanessa C. Wheeler⁵, Nathan D. Price¹ and Leroy E. Hood^{1,*}

¹Institute for Systems Biology, Seattle, WA, USA, ²Institute for Genome Sciences and Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA, ³Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, WA, USA, ⁴McLaughlin Research Institute, Great Falls, MT, USA, ⁵Center for Human Genetic Research, Massachusetts General Hospital, Department of Neurology, Harvard Medical School, Boston, MA, USA, ⁶Department of Computer Science, University of Washington, Seattle, WA, USA, ⁷CHDI Management/CHDI Foundation, Princeton, NJ, USA, ⁸Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA and ⁹Behavioral Neuroscience Program, Department of Psychology, Western Washington University, Bellingham, WA, USA

*To whom correspondence should be addressed at: Institute for Systems Biology, 401 Terry Ave N, Seattle, WA 98109, USA. Tel: +1 2067320202; Fax: +1 2067320260; Email: lhood@systemsbiology.org

Abstract

Huntington's disease is a dominantly inherited neurodegenerative disease caused by the expansion of a CAG repeat in the *HTT* gene. In addition to the length of the CAG expansion, factors such as genetic background have been shown to contribute to the age at onset of neurological symptoms. A central challenge in understanding the disease progression that leads from the HD mutation to massive cell death in the striatum is the ability to characterize the subtle and early functional consequences of the CAG expansion longitudinally. We used dense time course sampling between 4 and 20 postnatal weeks to characterize early transcriptomic, molecular and cellular phenotypes in the striatum of six distinct knock-in mouse models of the HD mutation. We studied the effects of the *Htt*^{Q111} allele on the C57BL/6J, CD-1, FVB/NCr1, and 129S2/SvPasCr1 genetic backgrounds, and of two additional alleles, *Htt*^{Q92} and *Htt*^{Q50}, on the C57BL/6J background. We describe the emergence of a transcriptomic signature in *Htt*^{Q111/+} mice involving hundreds of differentially expressed genes and changes in diverse

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Received: November 4, 2016. Revised: December 9, 2016. Accepted: January 3, 2017

© The Author 2017. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

molecular pathways. We also show that this time course spanned the onset of mutant huntingtin nuclear localization phenotypes and somatic CAG-length instability in the striatum. Genetic background strongly influenced the magnitude and age at onset of these effects. This work provides a foundation for understanding the earliest transcriptional and molecular changes contributing to HD pathogenesis.

Introduction

Huntington's disease (HD) is a dominantly inherited neurodegenerative disorder that is characterized by progressive motor impairments, accompanied by cognitive impairment and psychiatric symptoms, leading to premature death (1). HD is caused by an expanded CAG trinucleotide repeat in the *HTT* gene, encoding an elongated polyglutamine (polyQ) tract in the huntingtin protein (2).

The age at onset of HD symptoms ranges from rare juvenile cases to more typical onset in mid- to late-adulthood. Rare, longer CAG lengths (>60) lead to juvenile onset (3–5) whereas 40–50 CAG repeats is typical of adult onset disease. Approximately 66% of variation in age at onset for adult CAG ranges (40–53 CAGs) is explained by an inverse relationship with the length of the expanded CAG tract (6). Genetic modifiers of age at onset at other genomic loci explain some of the remaining variance, with loci on chromosomes 8 and 15 reaching genome-wide significance in a recent genome-wide association study (7). Environmental and lifestyle factors may also contribute to variation in the age at onset.

Neuropathological studies of HD have revealed specific loss of GABAergic medium spiny projection neurons in the striatum (8) that occurs earlier and is more profound than cell death in other brain regions. It is thought that this loss of medium spiny neurons explains many of the symptoms of HD. Mouse models have revealed a wide range of molecular and cellular pathology in the striatum preceding cell death, including synaptic dysfunction (9), neuroinflammation (10), altered cholesterol and energy metabolism (11,12), and changes in chromatin structure (13). Many of these phenotypes were originally discovered through gene expression profiling, which revealed thousands of differentially expressed genes in both post-mortem striatal tissue from HD patients and in mouse models of HD (14–16).

Here, we describe large-scale gene expression profiling and molecular phenotyping of striatal tissue from six distinct knock-in mouse models of HD during a time course from 4 to 20 postnatal weeks. Our study included the *Htt*^{Q111} allele on the C57BL/6J, CD-1, FVB/NCr1, and 129S2/SvPasCr1 genetic backgrounds and two additional shorter CAG repeat alleles (*Htt*^{Q92} and *Htt*^{Q50}) on the C57BL/6J background. Some behavioral changes have been observed in *Htt*^{Q111} mice at 11–12 weeks of age as well as memory deficits at 16 weeks of age (17,18). But motor phenotypes were not seen in *Htt*^{Q111} mice until 9 months of age (19). Overall, our longitudinal study likely encompasses some behavioral and cognitive changes, but no clear motor deficits.

The goals of our study were two-fold. First, we aimed to identify very early events in the progression of molecular and cellular phenotypes underlying HD. Identifying early events in the progression of the disease is critical, since these changes are likely to be enriched for causal mechanisms. By contrast, most previous gene expression profiling experiments and similar studies in HD mouse models have focused on the later stages of the disease progression or sampled at low density time points that do not capture early inflection points (15,16,20).

Second, we aimed to characterize similarities and differences in the effects of *Htt* CAG expansion when occurring on

different genetic backgrounds. Given the now strong evidence for genetic modifiers in the human disease, modeling the effect of genetic variation in mice is an increasingly high priority. In addition, identifying changes that are robust across multiple genetic backgrounds is a proven strategy to improve signal-to-noise and eliminate false positive results (21).

Here, we find a cascade of gene expression changes associated with *Htt* CAG expansion beginning in mice as young as 10–16-week old. We show that this time course spanned the onset of two known early molecular phenotypes in HD mouse models: nuclear localization of the mutant huntingtin protein and striatal CAG-length genomic instability. Genetic background strongly influenced the rate at which molecular and transcriptional phenotypes emerged, suggesting the presence of multiple genetic modifiers, and emphasizing the importance of studying human disease in model systems that account for genetic variation.

Results

Experimental design

In view of the dominant nature of HD, we used heterozygous knock-in mouse lines in which expanded CAG repeats of various lengths have been inserted into one of the mouse's endogenous *Htt* alleles (Fig. 1A). The *Htt*^{Q111} allele (previously *Hdh*^{Q111}; 'Q111') has been backcrossed onto four different genetic strain backgrounds: C57BL/6J ('B6'), CD-1, FVB/NCr1 ('FVB'), and 129S2/SvPasCr1 ('129') (22–24). On the B6 background, we also characterized knock-in mice carrying shorter alleles, *Htt*^{Q50} ('Q50') and *Htt*^{Q92} ('Q92') (16,24,25). Note that Q111, Q92 and Q50 are the nominal allele names, but due to intergenerational instability in Q92 and Q111 that is background strain-dependent (23,24), mice from these lines vary in their actual CAG repeat length (Fig. 1B). Siblings of these knock-in mice that did not carry pathogenic alleles ('WT') were used as controls. Three mice of each genotype were collected each week between 4 and 20 postnatal weeks, and an additional ten mice were collected every fourth week for the B6.Q50, B6.Q111 and CD-1.Q111 lines. 2–10 mice per week from each genotype were used for gene expression profiling and for cellular and molecular phenotyping (Fig. 1A; Supplementary Material, Table S1).

We measured the body mass and brain mass of each mouse. Age and strain influenced these traits, but there were no statistically significant differences in body mass or brain mass associated with *Htt* alleles (Supplementary Materials, Tables S2 and S3).

Dynamic huntingtin nuclear localization patterns across strain and over time

Nuclear localization of mutant huntingtin protein (mHTT) is one of the earliest known molecular phenotypes in HD knock-in mouse models (26). We measured nuclear huntingtin immunostaining phenotypes in the striata of 84 Q111 mice, using a conformation-specific antibody that recognizes forms of mHTT in the nuclei of striatal medium spiny neurons (Fig. 1C and D)

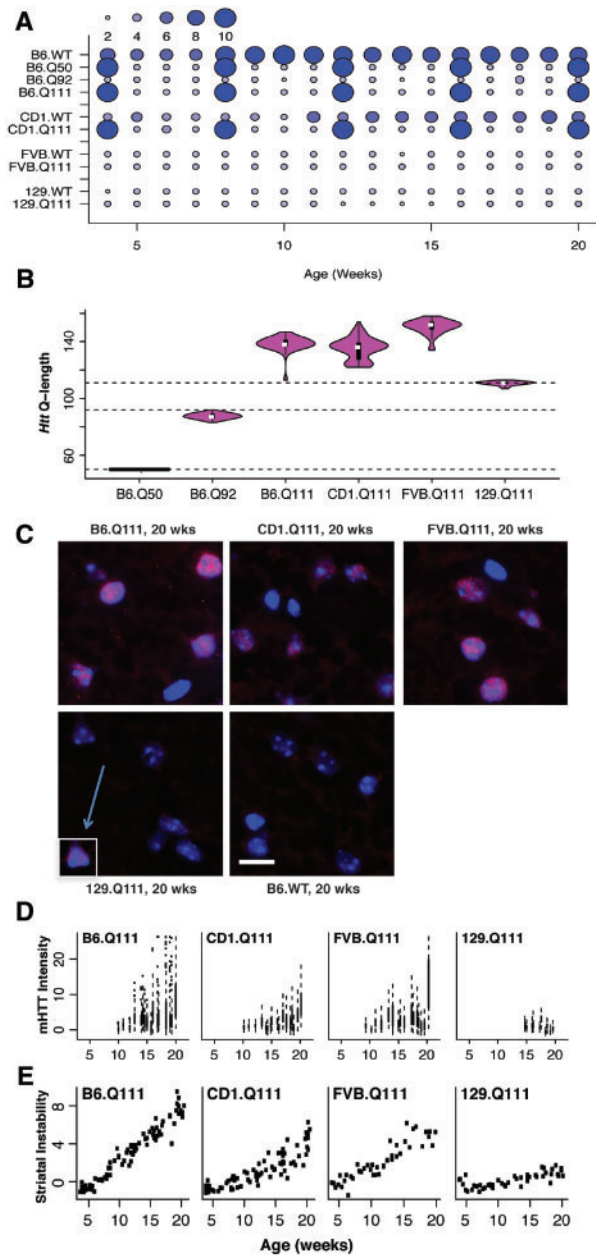


Figure 1. Time course of mHTT nuclear localization and of striatal Htt CAG length expansion in 4- to 20-week-old HD knock-in mice. (A) Striatal tissue was collected from a total of 731 4- to 20-week-old mice in order to compare mice with heterozygous mutant *Htt* alleles (Q111, Q92, Q50) to mice with wild-type *Htt* alleles (WT) on four genetic backgrounds. (B) Inherited CAG repeat length of *Htt* alleles in *Htt*^{Q111} mice varies with genetic background. The CAG repeat length in each mouse with a nominal *Htt*^{Q111} allele was measured in a tail snip sample at weaning. (C) Nuclear mutant huntingtin immunofluorescence using mAb5374 (red) and DAPI (blue) counterstain in 20-week-old B6.Q111, CD1.Q111, FVB.Q111, 129.Q111 mice, along with negative control 20 week-old B6.WT. All images were acquired at the same sensitivity and displayed with equal adjustment of the levels to enhance the red signal relative to background and nuclear staining. To illustrate the pattern of the weak staining in 129.Q111 the cell in the inset was further enhanced; arrows indicate positive cells. The white scale bar represents 10 μ m. (D) Quantitation of mHTT immunofluorescence intensity in 9- to 20-week-old Q111 mice from each background ($n = 84$ total mice, $n = 20$ –25 mice per strain). Each point indicates the immunofluorescence intensity in a single cell. (E) Age- and background strain-dependent expansion of *Htt* CAG tracts in the striata of 4- to 20-week-old Q111 mice. Each point represents the ‘instability index’ for a single mouse, a quantitative metric representing the mean CAG length change in the population, relative to the modal allele.

(26,27). Diffuse nuclear mHTT immunostaining was visible in a fraction of striatal cells in heterozygous B6.Q111, CD-1.Q111, and FVB.Q111 mice as young as 9- to 10-week old (Fig. 1D). In these strains, the intensity of nuclear staining increased over time (ANOVA, $P_{\text{age}} = 1.5e-5$; Fig. 1D). By 20 weeks, diffuse nuclear mHTT was prominent in striatal neurons in these strains (Fig. 1C). Punctate nuclear mHTT inclusions were visible in these strains beginning at 17 weeks. In contrast to the other three strains, diffuse nuclear huntingtin was not visible in the 129.Q111 mice until they were at least 15 weeks old and was not as intense as in the other three strains (ANOVA: $P_{\text{strain}} = 1.1e-3$). Punctate nuclear inclusions were not observed in 129.Q111 mice up to 20 weeks of age. These strain differences are consistent with the relative timing of nuclear localization and inclusion formation in previous observations in these strains (23,26,28).

Somatic CAG length expansion in striatum across strain and over time

We profiled the distribution of *Htt* CAG lengths in striatal tissue from Q111 mice ($n = 250$). Previous studies showed that genetic differences between B6.Q111, 129.Q111 and FVB.Q111 mice influence the rate of striatal CAG length expansion (23,29). While repeat expansion dynamics have been analyzed in CD-1.Q111 mice over a broad (2–16 months) time window (22), very early repeat expansion dynamics and the relationship to molecular phenotypes are not well established. We observed age- and strain-dependent effects on the extent of striatal CAG length instability (Fig. 1E). The ‘instability index’ summarizes the mean change in CAG repeat length observed in each mouse’s striatum. Instability index increased as a linear function of age in all four strains, but the rate of instability differed. Consistent with previous results (23,29), CAG tracts expanded more extensively in B6.Q111 and FVB.Q111 mice than in 129.Q111 heterozygous mice. The rates of CAG tract expansion in CD-1.Q111 mice were roughly similar to those in FVB.Q111 mice, but CD-1.Q111 exhibited greater variation in instability indices, consistent with this being an outbred strain.

Differentially expressed genes in 4- to 20-week-old knock-in mouse models of the HD mutation

We used linear modeling to characterize gene expression changes in each HD mouse model. We considered 18 conditions, defined by a mouse’s genotype and age. For each of the six genetically distinct HD mouse models—B6.Q111, CD-1.Q111, FVB.Q111, 129.Q111, B6.Q92, and B6.Q50—we considered three non-overlapping age windows with equal numbers of mice: ‘early’, 4–9 postnatal weeks; ‘middle’, 9–16 postnatal weeks; and ‘late’, 16–20 postnatal weeks. We compared these mice to age-matched wildtype mice of the same strain. In the early age window, we detected fewer than 25 differentially expressed genes (DEGs, $q < 0.05$) in all mouse models (Fig. 2A). In the middle age window, we detected 279 DEGs in B6.Q111 mice and < 25 DEGs in the other mouse models. In the late age window, we detected 1,513 DEGs in B6.Q111 mice, 222 DEGs in CD-1.Q111 mice, and < 25 DEGs in the other mouse models. Therefore, counts of DEGs increased progressively with age in B6.Q111 and CD-1.Q111 mice, while few DEGs could be detected in the other mouse models.

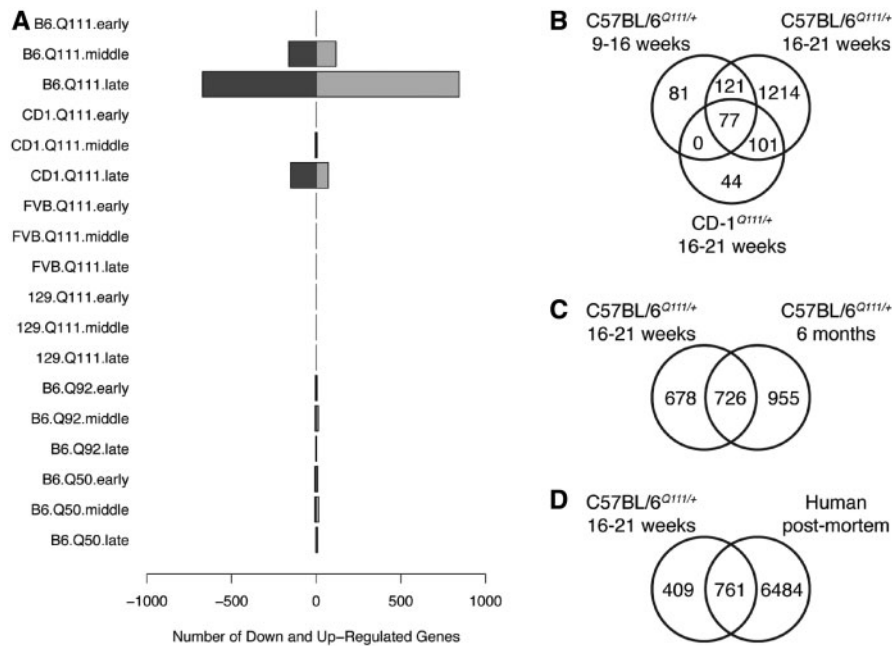


Figure 2. Age- and strain-dependent effects of *Htt* alleles on gene expression. (A) Counts of up- and down-regulated genes (FDR $q < 0.05$) in each genotype at three time windows: ‘early’ (4- to 9-week-old), ‘middle’ (9- to 16-week-old), and ‘late’ (16- to 20-week-old). (B) Overlap between differentially expressed genes (DEGs) detected in the three conditions with >25 DEGs. (C) Overlap between DEGs in 16- to 20-week-old B6.Q111 mice (this study) and DEGs in 6-month-old B6.Q111 mice from Langfelder et al. (2016). (D) Overlap between DEGs in 16- to 20-week-old B6.Q111 mice (this study) and DEGs in post-mortem caudate nucleus from HD cases vs. controls (Hodges et al. 2006).

Replication of DEGs across conditions and in independent datasets

There was extensive overlap between the DEGs identified in different age x genotype conditions (Fig. 2B). 178 of the 222 DEGs in 16- to 20-week-old CD-1.Q111 mice (80%) were also differentially expressed in 16- to 20-week-old B6.Q111 mice (Fisher’s exact test: $P = 1.4e-173$). 198 of the 279 DEGs in 9- to 16-week-old B6.Q111 mice were also differentially expressed in 16- to 20-week-old B6.Q111 mice ($P = 1.0e-174$). 77 DEGs were shared among all three conditions with > 25 DEGs.

There was also extensive overlap between the DEGs identified in our study and striatal gene expression changes in independent datasets from HD knock-in mouse models and human postmortem brain. 726 of the DEGs in 16- to 20-week-old B6.Q111 mice were differentially expressed in striata of 6-month-old B6.Q111 mice profiled by RNA-seq (16) ($P < 2.2e-308$; Fig. 2C). 761 of the DEGs in 16- to 20-week-old B6.Q111 mice were differentially expressed in post-mortem striatal tissue from (human) HD cases vs. controls (Hodges et al. 2006; $P = 3.1e-14$; Fig. 2D). Notably, there were far fewer DEGs in the <20 -week-old mice from our study than in the older mice profiled by Langfelder et al. (16) and in post-mortem human striatum. These results suggest that the DEGs identified in our study represent a robust transcriptomic signature for the early effects of the HD mutation in the striatum.

Differentially expressed genes are involved in diverse neuronal functions

We used Gene Ontology (30) and Cell-type Specific Expression Analysis (31,32) to characterize biological functions enriched for DEGs in 16- to 20-week-old B6.Q111 vs. WT mice and the dominant cell types in which they are expressed. Down-regulated

genes were enriched for genes that are expressed specifically in *Drd1*- and *Drd2*-positive medium spiny neurons ($P = 7.4e-53$, $P = 9.6e-44$, respectively; Supplementary Material, Table S4). Down-regulated genes were enriched for 389 Gene Ontology terms (FDR < 0.05 ; Supplementary Material, Table S5). Many of the most strongly enriched pathways relate to synaptic activity, such as ‘dopamine receptor signaling pathway’ ($P = 1.0e-14$), ‘regulation of membrane potential’ ($P = 1.9e-19$), ‘regulation of ion transport’ ($5.0e-19$), ‘neuron projection morphogenesis’ ($P = 3.4e-18$), and ‘ionotropic glutamate receptor binding’ ($P = 1.4e-12$). Down-regulated genes were also enriched for components of intracellular signaling pathways known for their roles in coupling the activity of transmembrane receptors to downstream cellular processes. These pathways included ‘regulation of G-protein-coupled receptor protein signaling pathway’ ($P = 2.9e-22$), ‘calmodulin binding’ ($P = 1.3e-25$), ‘GTPase activator activity’ ($P = 1.3e-20$), ‘response to organic cyclic compound’ ($P = 3.4e-14$), and ‘response to insulin’ ($P = 2.7e-11$).

Up-regulated genes in 16- to 20-week-old B6.Q111 mice were not strongly enriched for cell type-specific genes, suggesting that these expression changes are distributed across multiple cell types (Supplementary Material, Table S4). These genes were enriched for 188 pathways (FDR < 0.05 ; Supplementary Material, Table S6). Like down-regulated genes, up-regulated genes were enriched for neuronal functions, such as ‘neuron projection morphogenesis’ ($P = 3.2e-13$). In addition, up-regulated genes were enriched for more general cellular functions, such as ‘translation’ ($P = 5.3e-13$), ‘ubiquitin protein-ligase binding’ ($P = 8.2e-9$), and ‘regulation of homeostatic process’ ($P = 4.7e-10$). These results indicate that the *Htt* CAG expansion influences the expression of genes with diverse functions, even early in the pathogenic process.

Notably, we found few expression changes in genes related to neuroinflammation, astrogliosis, and apoptosis (Supplementary

Materials, Tables S5 and S6). All of these processes are upregulated in striatal tissue from HD mouse models sampled later in life and in post-mortem tissue from HD patients (14,16). Therefore, our results suggest that neuroinflammation-related transcriptomic changes are secondary to the 'neuronal' signature detected in the young mice from our study.

Gene expression changes in 4- to 16-week-old HD knock-in mice

The results above suggest that robust HD-related gene expression changes occur in 16- to 20-week-old B6.Q111 and CD-1.Q111 mice. We next asked whether we could detect early signs of this HD molecular signature in even younger mice and in other mouse strains. Although we detected few DEGs in these conditions, we hypothesized that younger mice and/or those with lower CAG lengths might show subtle changes in expression that would be correlated with the expression changes detected at later stages of the phenotypic progression. To test this hypothesis, we focused on the expression patterns of the top 100 genes with the lowest P-values in 16- to 20-week-old B6.Q111 mice (Fig. 3). Many of these genes—e.g. *Pde10a*, *Penk*, *Cnr1*, and *Gpx6*—are differentially expressed in striatal tissue from HD patients (14) and in previous studies of HD knock-in mouse models (16).

We fit a linear model to test whether the fold changes of these top 100 genes predicted their fold changes in other conditions. That is, we asked to what extent the direction and magnitude of the fold changes were consistent across mouse models and time points. In this model, the correlation coefficient (r) describes the strength of the linear relationship between the fold changes of the top 100 genes in each pair of conditions, while the regression coefficient (b_1) describes the relative sizes of the fold changes.

In 16- to 20-week-old mice, the fold changes of the top 100 genes were positively correlated between B6.Q111 mice and each of the other genetic models (Fig. 4; $r = 0.38$ – 0.97). The fold changes varied between strains, with the largest fold changes in B6.Q111, followed by CD-1.Q111 ($b_1 = 0.62$, relative to 16- to 20-week-old B6.Q111 mice), FVB.Q111 ($b_1 = 0.43$), B6.Q92 ($b_1 = 0.25$), 129.Q111 ($b_1 = 0.20$) and B6.Q50 ($b_1 = 0.07$).

In 9- to 16-week-old mice, the fold changes of the top 100 genes were positively correlated in B6.Q111, CD-1.Q111, FVB.Q111, and B6.Q92 mice ($r = 0.34$ – 0.96) but were not correlated in 129.Q111 or B6.Q50 mice. B6.Q111 mice had the largest fold changes ($b_1 = 0.58$), followed by CD-1.Q111 ($b_1 = 0.38$), FVB.Q111 ($b_1 = 0.18$), B6.Q92 ($b_1 = 0.06$).

In 4- to 9-week-old mice, fold changes were more variable. Fold changes were positively correlated in B6.Q111 ($r = 0.72$, $b_1 = 0.17$), FVB.Q111 ($r = 0.44$, $b_1 = 0.08$), and 129.Q111 ($r = 0.57$, $b_1 = 0.15$), compared to 16- to 20-week-old B6.Q111 mice. However, fold changes were negatively correlated in B6.Q92 mice of this age ($r = -0.41$, $b_1 = -0.11$), and were uncorrelated in CD-1.Q111 and B6.Q50 mice. These results suggest that HD mutations influence many of the same genes across multiple mouse models of the HD mutation, but genetic background and CAG repeat length influence the rate at which these genes become differentially expressed.

Multiple molecular and genetic mechanisms may underlie strain differences in response to HD mutations

Our results show that both CAG repeat length and strain genetic background modify the rate at which molecular and cellular

phenotypes progress in mice with *Htt* CAG expansions. Phenotypic variation may be driven at least in part by the different inherited CAG repeat lengths in each of the four Q111 strains (Fig. 1B), but is not explained by variation in *Htt* expression (Supplementary Material, Fig. S1). As the strains also exhibit differences in the rate of somatic CAG expansion (Fig. 1E) we hypothesized that these strain differences involve a known genetic polymorphism that influences somatic instability of the CAG repeat in B6.Q111 vs. 129.Q111 mice. A previously reported genetic mapping study in C57BL/6N.Q111 x 129.Q111 mice showed that variants at the *Mlh1* locus were linked to slower *Htt* CAG somatic expansion in the striata of 129.Q111 vs. B6N.Q111 mice (29). These variants were associated with lower expression of the *Mlh1* gene, which encodes a mismatch repair enzyme, MutL homolog 1. We examined *Mlh1* expression in mice from each of the four background strains in our study to replicate the previously reported expression polymorphism in B6N or B6J vs. 129 mice (29) and to assess differences in CD-1 and FVB mice. Wildtype 129 mice had 2.8-fold lower expression of *Mlh1*, compared to wildtype B6 mice, consistent with previous findings (Fig. 5), while B6, CD-1 and FVB mice all had similar high expression of *Mlh1*. The intermediate rates of somatic CAG expansion in FVB and CD-1 strains are consistent with high *Mlh1* expression in these strains, but suggest additional modifier genes(s) that might reduce expansion relative to B6. Nuclear mutant huntingtin immunostaining phenotypes correlate reasonably well with the rates of somatic CAG expansion across strain background, indicating that genetic modifiers of these *Htt* CAG-dependent events may be shared. Notably, however, the very different transcriptional responses to the Q111 allele in B6, FVB and CD-1 strains are unlikely to be explained by genetically encoded changes in *Mlh1* expression. Overall, these results indicate that multiple genetic factors may modify responses to the HD mutation.

Discussion

The cascade of early molecular, cellular and transcriptional changes that occurs as a result of CAG expansion in HD mouse models has not previously been studied systematically across a week-by-week timescale. We generated transcriptomic data in a dense time-series from the striatum of mice with *Htt* CAG repeat expansions in parallel with molecular and cellular phenotyping to elucidate the timing and magnitude of CAG length- and age-dependent changes across four genetic background strains. We observed robust striatal CAG length expansion, mHTT nuclear localization, and gene expression changes in B6.Q111 mice as young as 9- to 16-week old. We find evidence that subtle changes in these same phenotypes begin to manifest even earlier in life and with shorter CAG repeat tracts. While the Q111 allele induced similar molecular changes on the other background strains studied, the rate of progression differed across these strains. Our study provides a dense and dynamic view of very early phenotypic and transcriptomic changes in the striatum of mice with *Htt* CAG repeat expansions.

Our results reveal very early gene expression changes in the striatum of Q111 mice. Most of these expression changes were very small, typically less than 1.5-fold differences in expression in 16- to 20-week-old B6.Q111 vs. WT mice. These effect sizes in our analysis of bulk striatal tissue are consistent either with subtle changes occurring in many striatal cells or with more dramatic changes occurring in a small subset of striatal cells.

Many of the differentially expressed genes are involved in synaptic functions and are predominantly expressed in neurons

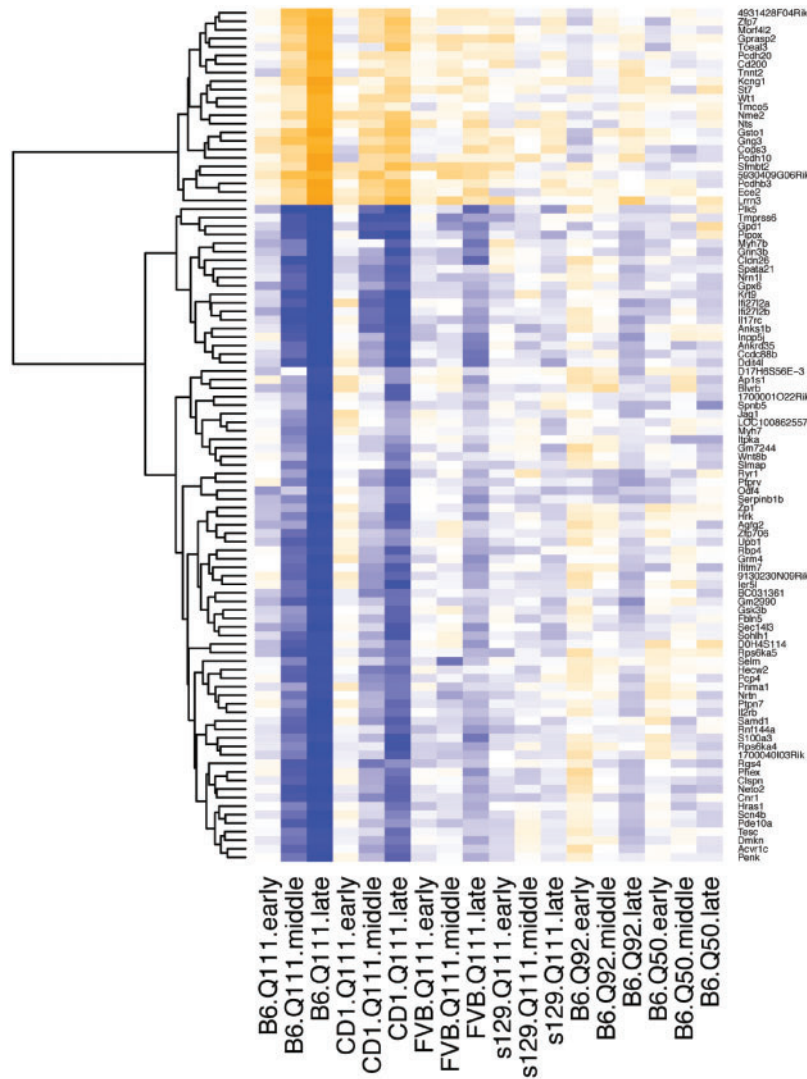


Figure 3. Expression patterns of the top 100 differentially expressed genes. The top 100 genes were defined by their P -values in 16- to 20-week-old B6.Q111 mice. The heatmap indicates the $-\log_{10}(P\text{-values})$ of these genes in 18 conditions defined by genotype and age window. Orange gradient = up-regulated in Q111; blue gradient = down-regulated in Q111; color range, $-7 < z\text{-score} < 7$.

rather than glia. Most of these same pathways are also differentially expressed in older mice with HD mutations and in post-mortem striatal tissue from HD cases vs. controls. Notably, however, the biological processes perturbed in young Q111 mice represent only a small fraction of processes that are perturbed later in the disease. For instance, in contrast to studies of 6- and 10-month-old B6.Q111 mice, we find a few changes in genes associated with neuroinflammation and activated gliosis (16). Therefore, our results suggest that the sequence of molecular changes associated with HD mutations begins with a relatively small number of changes in neurons, before cascading onward to broader pathology in both neurons and in glia.

Our results suggest that multiple genetic modifiers influence the effects of the Q111 allele on molecular phenotypes. B6 mice consistently presented the most rapid phenotypic progression while 129 mice were the slowest. These phenotypic differences appear to be primarily quantitative and not qualitative: i.e. we observed correlated gene expression responses in all strains, but the rate of change differed. Our study was not specifically designed to detect potential modifier effects of strain

background, but rather to identify robust gene expression changes across multiple strains. Thus, the relatively low CAG repeat length in 129.Q111 mice could, at least in part, contribute to the slower phenotypic progression in this strain. However, the slow rates of the transcriptional responses in CD-1 and FVB compared to B6, despite having similar or longer CAG repeat lengths than B6, indicates the presence of modifier genes. Further, an expression QTL at the *Mlh1* locus that is linked to differences in *Htt* CAG length expansion in B6 vs. 129 mice do not appear to explain strain differences in CD-1 and FVB relative to B6, suggesting that other genetic modifiers of the transcriptional response are present in these strains. Further unbiased identification of the genetic modifiers of different phenotypes in *Htt* CAG knock-in mice, e.g. by crossing mice with HD mutations onto a mapping population such as the Collaborative Cross or Mouse Diversity outbred collection (33,34), is a promising future direction that could reveal multiple modifier loci. Identifying these genetic modifiers could point to novel therapeutic targets.

The proximal mechanisms by which HD mutations cause early gene expression changes remain unclear. Several non-

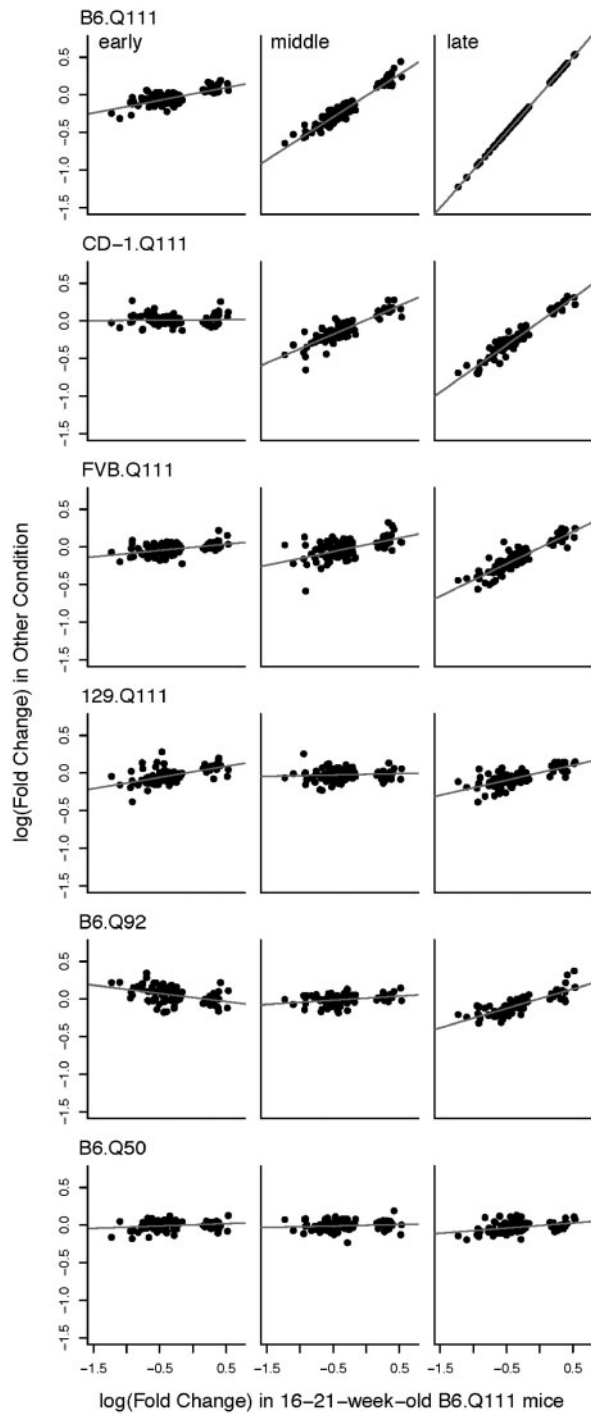


Figure 4. Age, CAG length and genetic background influences the magnitude but not the direction of gene expression changes in HD knock-in mice. Each scatterplot compares the fold changes of the top 100 genes in 16- to 20-week-old B6.Q111 mice (x-axis) to the fold changes of these genes in another condition, defined by age and genotype (y-axis). Each point on the scatterplot indicates the fold change estimate of a single gene.

mutually-exclusive mechanisms have been proposed (35–38): (i) Both wildtype and mutant HTT form protein complexes with transcription factors, and these interactions could change the ability of these TFs to regulate their target genes; (ii) direct interaction of huntingtin protein with chromatin modifying enzymes and with genomic DNA influences gene expression

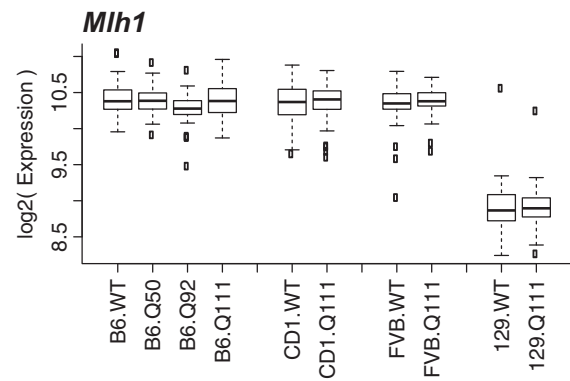


Figure 5. Effects of genetic background and of HD mutations on *Mlh1* expression. Box plots represent mean expression of *Mlh1* in each strain and genotype group.

through changes in chromatin or regulatory states (39–41); (iii) gene expression changes are secondary or occur indirectly as a result of another mechanism. In our study, we found that the rates of somatic Htt CAG length expansion, mHTT nuclear localization, and transcriptional changes were all correlated with age. These correlations suggest that these phenotypes are connected but make it difficult to distinguish causal from correlative relationships. Unbiased genetic studies to determine whether common or different modifiers underlie these phenotypes will help to disentangle these relationships. More detailed biochemical experiments will be required to identify a subset of the observed changes that are caused by direct interactions with huntingtin protein.

Since the 1993 discovery of the HD mutation, research into the causal mechanism underlying neurodegeneration has generally focused on late stage disease. However, a wide range of pathophysiology is present in late stages of HD, much of which may be only peripherally related to pathogenesis. Our approach focuses instead on studying earlier, more subtle perturbations, enabling us to identify a smaller number of candidate mechanisms. By measuring transcriptomic, molecular, and cellular changes with many samples across a dense time series we were able to precisely time inflection points such as gene expression changes and nuclear mHTT phenotypes. Equally critical is our observation of striking differences in the magnitude and timing of these effects across strains. This study contributes to the burgeoning field of ‘pre-manifest’ HD research, clarifying molecular changes that occur well before the onset of diagnosable clinical symptoms with the potential of proffering new therapeutic targets.

Methods

Breeding

Heterozygous males of six different strains (129S2/SvPasCrl.Htt^{Q111/+}, FVB/Ncrl.Htt^{Q111/+}, CD-1.Htt^{Q111/+}, C57BL/6J.Htt^{Q92/+}, C57BL/6J.Htt^{Q50/+}, C57BL/6J.Htt^{Q111/+}) were crossed with wild type females of corresponding background strains in order to generate cohorts of age-matched heterozygous pups and WT controls for the time-course assigned tissue harvests. We also generated through the same breeding scheme mice heterozygous for the CD-1. Htt^{neoQ50/+} allele and their wildtype littermates. Transcriptomic data that were generated from these mice are available in GEO (GSE88920) but were not analyzed as part of this study as the presence of the neo-resistance gene renders this allele hypomorphic (25). A tail snip was

acquired at pre-weaning (2–4 weeks) to determine the CAG repeat length of each mouse.

One heterozygous animal and one age-matched wild type control animal from each strain were harvested three times weekly from 4 to 20 weeks of age. The sex of mice harvested at these times was randomized. An additional 5 female and 5 male heterozygous animals from the B6.Q50, B6.Q111 and CD1.Q111 strains were harvested at 4, 8, 12, 16, and 20 weeks of age. Time of day of harvest was randomized within the hours of light in the mouse room, typically ≥ 3 hours after lights came on and ≥ 3 hours before lights went off.

Euthanasia was performed using live decapitation. Trunk blood was collected and plasma was frozen and stored at -80°C . The right hemisphere of each animal's brain was micro-dissected into thalamus, cortex, striatum, hippocampus, and remaining brain regions. These tissues were snap frozen and stored at -80°C . The left hemisphere was embedded in OCT embedding medium, snap frozen, and stored at -80°C for sectioning. Snap-frozen striata from the right hemisphere of each animal were used for isolation of RNA and DNA for transcriptional profiling and analyses of CAG instability, and sectioned tissue from the left hemisphere was used for immunostaining with mAb5374.

mAb5374 immunofluorescence

Nuclear mutant huntingtin localization was assessed by immunofluorescence imaging of frozen $7\mu\text{m}$ coronal sections with the mAb5374 antibody (Chemicon), as previously described (42). The mAb5374 antibody is a conformationally sensitive antibody that detects diffusely immunostaining nuclear mutant huntingtin and intranuclear inclusions. We used Cell Profiler image analysis software (43) to quantify the intensity of nuclear mAb5374 staining in each cell. The max intensity in each cell was normalized to intensity levels in wild type animals to calculate a normalized score.

Determination of Htt CAG repeat length and striatal instability

Assays for knock-in mouse genotyping and determining Htt CAG repeat length were as described previously (42). The Htt CAG repeat was amplified using a human-specific PCR assay that amplifies the repeat from the knock-in allele but does not amplify the mouse sequence. The forward primer was fluorescently labeled with 6-FAM (Applied Biosystems) and products were resolved using the ABI 3730xl DNA analyzer (Applied Biosystems) with GeneScan 500 LIZ as internal size standard (Applied Biosystems). GeneMapper v3.7 (Applied Biosystems) was used to generate CAG repeat size distribution traces. Repeat size was determined from the peak with the greatest intensity in the GeneMapper trace. An instability index was quantified from the GeneMapper CAG repeat distributions in striatal DNA as previously described (44). Briefly, the highest peak in each trace was used to determine a relative threshold of 10% and peaks falling below this threshold were excluded from analysis. Peak heights normalized to the sum of all peak heights were multiplied by the change in CAG length of each peak relative to the highest peak. These values were summed to generate an instability index, which represents the mean CAG repeat length change in the population of cells being analyzed. Note that the tail CAG length at pre-weaning does not vary appreciably from the main CAG allele in striatum; rather the instability in

striatum is reflected in a broader distribution of alleles, strongly biased to longer CAG lengths.

Microarray gene expression profiling

RNA was extracted from the striata of individual mice using Trizol reagent (DNA was isolated for instability analyses in the same procedure). We profiled gene expression on SurePrint G3 8x60k Mouse Microarrays (Agilent, Santa Clara, CA). We performed quantile normalization and log₂-transformation. We used ComBat (45) to remove batch effects related to the date on which microarrays were hybridized. Linear models were fit using the limma R package (46). Age was treated as a categorical variable with three classes: 'early', 4- to 9-week-old; 'middle', 9- to 16-week-old; 'late', 16- to 20-week-old. We considered the nominal Htt allele as a categorical variable. We fit a linear model to predict each gene's expression, considering the effects of background strain, age, sex, and Htt allele. We used post-hoc contrasts to evaluate the effects of Htt allele in each background strain x age condition. A false discovery rate for these P-values was calculated using the Benjamini-Hochberg method (47). We note that multiple widely accepted methods are available to analyze microarray data, and lists of differentially expressed genes would inevitably differ to some extent depending on the choice of method. This dataset has been deposited in GEO, accession GSE88920. Scripts and processed data have also been deposited in a GitHub repository located at <https://github.com/seth-ament/hd-time-series>.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank Nathan Goodman for his role in initiating this project and for performing microarray data pre-processing.

Conflict of Interest statement. None declared.

Funding

This work was supported by the University of Luxembourg-Institute for Systems Biology Strategic Partnership, a contract from the CHDI Foundation, a National Science Foundation Graduate Student Research Fellowship to JRP, and NIH grant NS049206 to VCV.

References

1. Bates J, Tabrizi S. and Jones L. eds. (2014) Huntington's Disease Oxford University Press, USA.
2. The Huntington's Disease Collaborative Research Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, **72**, 971–983.
3. Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M., Abbott, M., et al. (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat. Genet.*, **4**, 387–392.
4. Andrew, S.E.S., Goldberg, Y., Kremer, B., Paul Goldberg, Y., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., et al. (1993) The relationship between

- trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat. Genet.*, **4**, 398–403.
5. Langbehn, D.R., Brinkman, R.R., Falush, D., Paulsen, J.S. and Hayden, M.R. (2004) A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin. Genet.*, **65**, 267–277.
 6. Lee, J.M., Ramos, E.M., Lee, J.H., Gillis, T., Mysore, J.S., Hayden, M.R., Warby, S.C., Morrison, P., Nance, M., Ross, C.A., et al. (2012) CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology*, **78**, 690–695.
 7. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2015) Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell*, **162**, 516–526.
 8. Vonsattel, J.P., Myers, R.H., Stevens, T.J., Ferrante, R.J., Bird, E.D. and Richardson, E.P. (1985) Neuropathological classification of Huntington's disease. *J. Neuropathol. Exp. Neurol.*, **44**, 559–577.
 9. Smith, R., Brundin, P. and Li, J.Y. (2005) Synaptic dysfunction in Huntington's disease: a new perspective. *Cell. Mol. Life Sci.*, **62**, 1901–1912.
 10. Möller, T. (2010) Neuroinflammation in Huntington's disease. *J. Neural Transm.*, **117**, 1001–1008.
 11. Valenza, M., Rigamonti, D., Goffredo, D., Zuccato, C., Fenu, S., Jamot, L., Strand, A., Tarditi, A., Woodman, B., Racchi, M., et al. (2005) Dysfunction of the cholesterol biosynthetic pathway in Huntington's disease. *J. Neurosci.*, **25**, 9932–9939.
 12. Koroshetz, W.W.J., Jenkins, B.B.G., Rosen, B.R. and Beal, M.F. (1997) Energy metabolism defects in Huntington's disease and effects of coenzyme Q10. *Ann. Neurol.*, **41**, 160–165.
 13. Zuccato, C., Valenza, M. and Cattaneo, E. (2010) Molecular mechanisms and potential therapeutic targets in Huntington's disease. *Physiol. Rev.*, **90**, 905–981.
 14. Hodges, A., Strand, A.D., Aragaki, A.K., Kuhn, A., Sengstag, T., Hughes, G., Elliston, L.A., Hartog, C., Goldstein, D.R., Thu, D., et al. (2006) Regional and cellular gene expression changes in human Huntington's disease brain. *Hum. Mol. Genet.*, **15**, 965–977.
 15. Seredenina, T. and Luthi-Carter, R. (2012) What have we learned from gene expression profiles in Huntington's disease? *Neurobiol. Dis.*, **45**, 83–98.
 16. Langfelder, P., Cantle, J.P., Chatzopoulou, D., Wang, N., Gao, F., Al-Ramahi, I., Lu, X.H., Ramos, E.M., El-Zein, K., Zhao, Y., et al. (2016) Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat. Neurosci.*, **19**, 623–633.
 17. Giral, A., Puigdemívol, M., Carretón, O., Paoletti, P., Valero, J., Parra-Damas, A., Saura, C.A., Alberch, J. and Ginés, S. (2012) Long-term memory deficits in Huntington's disease are associated with reduced CBP histone acetylase activity. *Hum. Mol. Genet.*, **21**, 1203–1216.
 18. Hölter, S.M., Stromberg, M., Kovalenko, M., Garrett, L., Glasl, L., Lopez, E., Guide, J., Götz, A., Hans, W., Becker, L., et al. (2013) A broad phenotypic screen identifies novel phenotypes driven by a single mutant allele in Huntington's disease CAG knock-in mice. *PLoS One*, **8**, e80923.
 19. Yhnell, E., Dunnett, S.B. and Brooks, S.P. (2016) A Longitudinal Motor Characterisation of the HdhQ111 Mouse Model of Huntington's Disease. *J. Huntingtons. Dis.*, **5**, 149–161.
 20. Kuhn, A., Goldstein, D.R., Hodges, A., Strand, A.D., Sengstag, T., Kooperberg, C., Becanovic, K., Pouladi, M.A., Sathasivam, K., Cha, J.H.J., et al. (2007) Mutant huntingtin's effects on striatal gene expression in mice recapitulate changes observed in human Huntington's disease brain and do not differ with mutant huntingtin length or wild-type huntingtin dosage. *Hum. Mol. Genet.*, **16**, 1845–1861.
 21. Hwang, D., Lee, I.Y., Yoo, H., Gehlenborg, N., Cho, J.H., Petritis, B., Baxter, D., Pitstick, R., Young, R., Spicer, D., et al. (2009) A systems approach to prion disease. *Mol. Syst. Biol.*, **5**, 252.
 22. Lee, J.M., Pinto, R.M., Gillis, T., St Claire, J.C. and Wheeler, V.C. (2011) Quantification of age-dependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals different expansion dynamics in striatum and liver. *PLoS One*, **6**, e23647.
 23. Lloret, A., Dragileva, E., Teed, A., Espinola, J., Fossale, E., Gillis, T., Lopez, E., Myers, R.H., MacDonald, M.E. and Wheeler, V.C. (2006) Genetic background modifies nuclear mutant huntingtin accumulation and HD CAG repeat instability in Huntington's disease knock-in mice. *Hum. Mol. Genet.*, **15**, 2015–2024.
 24. Wheeler, V.C., Auerbach, W., White, J.K., Srinidhi, J., Auerbach, A., Ryan, A., Duyao, M.P., Vrbanc, V., Weaver, M., Gusella, J.F., et al. (1999) Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum. Mol. Genet.*, **8**, 115–122.
 25. White, J.K., Auerbach, W., Duyao, M.P., Vonsattel, J.P., Gusella, J.F., Joyner, A.L. and MacDonald, M.E. (1997) Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion. *Nat. Genet.*, **17**, 404–410.
 26. Wheeler, V.C., White, J.K., Gutekunst, C.A., Vrbanc, V., Weaver, M., Li, X.J., Li, S.H., Yi, H., Vonsattel, J.P., Gusella, J.F., et al. (2000) Long glutamine tracts cause nuclear localization of a novel form of huntingtin in medium spiny striatal neurons in HdhQ92 and HdhQ111 knock-in mice. *Hum. Mol. Genet.*, **9**, 503–513.
 27. Gutekunst, C.A., Li, S.H., Yi, H., Mulroy, J.S., Kuemmerle, S., Jones, R., Rye, D., Ferrante, R.J., Hersch, S.M. and Li, X.J. (1999) Nuclear and neuropil aggregates in Huntington's disease: relationship to neuropathology. *J. Neurosci.*, **19**, 2522–2534.
 28. Wheeler, V.C., Gutekunst, C.A., Vrbanc, V., Lebel, L.A., Schilling, G., Hersch, S., Friedlander, R.M., Gusella, J.F., Vonsattel, J.P., Borchelt, D.R., et al. (2002) Early phenotypes that presage late-onset neurodegenerative disease allow testing of modifiers in Hdh CAG knock-in mice. *Hum. Mol. Genet.*, **11**, 633–640.
 29. Pinto, R.M., Dragileva, E., Kirby, A., Lloret, A., Lopez, E., St Claire, J., Panigrahi, G.B., Hou, C., Holloway, K., Gillis, T., et al. (2013) Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. *PLoS Genet.*, **9**, e1003930.
 30. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
 31. Dougherty, J.D., Schmidt, E.F., Nakajima, M. and Heintz, N. (2010) Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.*, **38**, 4218–4230.
 32. Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al. (2008) Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell*, **135**, 749–762.

33. Collaborative Cross Consortium C.C., Aylor, D.L., Valdar, W., Foulds-Mathes, W., Buus, R.J., Verdugo, R.A., Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., et al. (2012) The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics*, **190**, 389–401.
34. Svenson, K., Gatti, D., Valdar, W. and Welsh, C. (2012) High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics*, **190**, 437–447.
35. Cha, J.H. (2000) Transcriptional dysregulation in Huntington's disease. *Trends Neurosci.*, **23**, 387–392.
36. Sugars, K.L., Brown, R., Cook, L.J., Swartz, J. and Rubinsztein, D.C. (2004) Decreased cAMP response element-mediated transcription: an early event in exon 1 and full-length cell models of Huntington's disease that contributes to polyglutamine pathogenesis. *J. Biol. Chem.*, **279**, 4988–4999.
37. Thomas, E.A. (2006) Striatal specificity of gene expression dysregulation in Huntington's disease. *J. Neurosci. Res.*, **84**, 1151–1164.
38. Benn, C.L., Sun, T., Sadri-Vakili, G., McFarland, K.N., DiRocco, D.P., Yohrling, G.J., Clark, T.W., Bouzou, B. and Cha, J.H.J. (2008) Huntingtin Modulates Transcription, Occupies Gene Promoters In Vivo, and Binds Directly to DNA in a Polyglutamine-Dependent Manner. *J. Neurosci.*, **28**, 10720–10733.
39. Freiman, R.N. and Tjian, R. (2002) Neurodegeneration. A glutamine-rich trail leads to transcription factors. *Science*, **296**, 2149–2150.
40. Zhai, W., Jeong, H., Cui, L., Krainc, D. and Tjian, R. (2005) In vitro analysis of huntingtin-mediated transcriptional repression reveals multiple transcription factor targets. *Cell*, **123**, 1241–1253.
41. Seong, I.S., Woda, J.M., Song, J.J., Lloret, A., Abeyrathne, P.D., Woo, C.J., Gregory, G., Lee, J.M., Wheeler, V.C., Walz, T., et al. (2010) Huntingtin facilitates polycomb repressive complex 2. *Hum. Mol. Genet.*, **19**, 573–583.
42. Kovalenko, M., Dragileva, E., St. Claire, J., Gillis, T., Guide, J.R., New, J., Dong, H., Kucherlapati, R., Kucherlapati, M.H., Ehrlich, M.E., et al. (2012) Msh2 Acts in Medium-Spiny Striatal Neurons as an Enhancer of CAG Instability and Mutant Huntingtin Phenotypes in Huntington's Disease Knock-In Mice. *PLoS One*, **7**, e44273.
43. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I., Friman, O., Guertin, D.A., Chang, J., Lindquist, R.A., Moffat, J., et al. (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.*, **7**, R100.
44. Lee, J.M., Zhang, J., Su, A.I., Walker, J.R., Wiltshire, T., Kang, K., Dragileva, E., Gillis, T., Lopez, E.T., Boily, M.J., et al. (2010) A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Syst. Biol.*, **4**, 29.
45. Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.
46. Smyth, G.K.G. (2005) Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer-Verlag, New York, pp. 397–420.
47. Benjamini, Y., Drai, D. and Elmer, G. (2001) Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.*, **125**, 279–284.