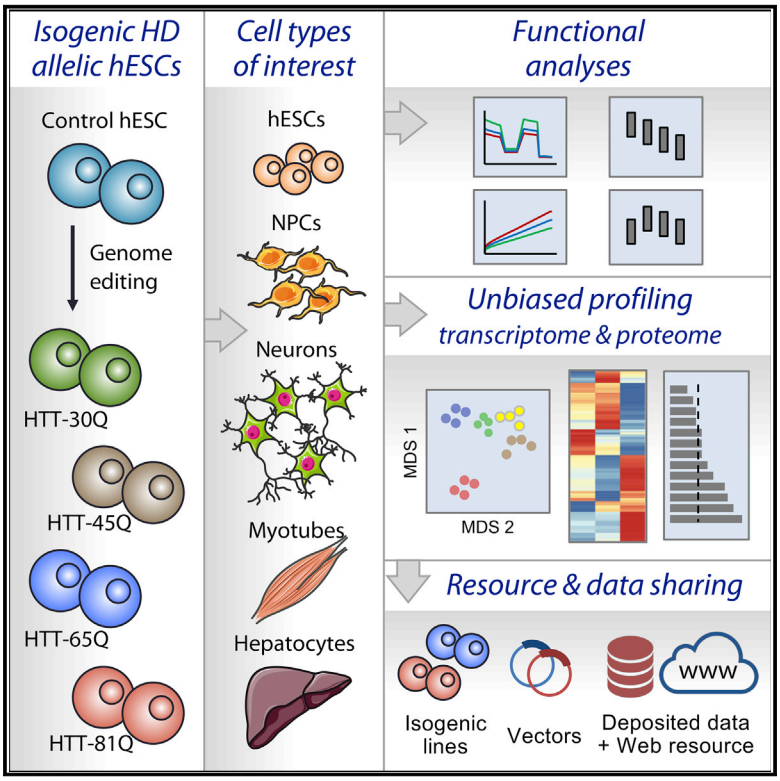
Resource

Cell Reports

Unbiased Profiling of Isogenic Huntington Disease hPSC-Derived CNS and Peripheral Cells Reveals Strong Cell-Type Specificity of CAG Length Effects

Graphical Abstract



HIGHLIGHTS

- Generation of isogenic HD (IsoHD) hESCs to enable the study of CAG length effects
- IsoHD hESCs and neural cells exhibit CAG-dependent phenotypic abnormalities
- CNS and peripheral cell types of interest can be differentiated from IsoHD hESCs
- Large-scale, unbiased -omics analyses reveal cell-type-specific, CAG-dependent effects

Authors

Jolene Ooi, Sarah R. Langley, Xiaohong Xu, ..., Florent Ginhoux, Enrico Petretto, Mahmoud A. Pouladi

Correspondence

map@pouladilab.org

In Brief

Ooi et al. use genome engineering to establish an allelic panel of isogenic Huntington disease (IsoHD) hESCs. Using unbiased analyses on neural progenitors, neurons, hepatocytes, and skeletal myotubes derived from the IsoHD hESCs, the authors illustrate how this approach can identify cell-type-specific, CAG-dependent effects of relevance to HD etiology.

Ooi et al., 2019, Cell Reports *26*, 2494–2508
February 26, 2019
https://doi.org/10.1016/j.celrep.2019.02.008

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

Unbiased Profiling of Isogenic Huntington Disease hPSC-Derived CNS and Peripheral Cells Reveals Strong Cell-Type Specificity of CAG Length Effects

Jolene Ooi,[1,11] Sarah R. Langley,[2,3,11] Xiaohong Xu,[1,4,5,11] Kagistia H. Utami,[1,11] Bernice Sim,[1] Yihui Huang,[1] Nathan P. Harmston,[2] Yi Lin Tay,[1] Amin Ziaei,[1] Ruizhu Zeng,[1] Donovan Low,[6] Folefac Aminkeng,[1] Radoslaw M. Sobota,[7,8] Florent Ginhoux,[6] Enrico Petretto,[2] and Mahmoud A. Pouladi[1,9,10,12,*]

[1]Translational Laboratory in Genetic Medicine (TLGM), Agency for Science, Technology and Research (A*STAR), Singapore 138648, Singapore
[2]Duke-NUS Medical School, 8 College Road, Singapore 169857, Singapore
[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 636921, Singapore
[4]Department of Neurology and Stroke Center, The First Affiliated Hospital, Jinan University, 613 Huangpu Avenue West, Guangzhou, Guangdong 510632, China
[5]Clinical Neuroscience Institute of Jinan University, 613 Huangpu Avenue West, Guangzhou, Guangdong 510632, China
[6]Singapore Immunology Network (SIgN), A*STAR, Singapore 138648, Singapore
[7]Functional Proteomics Laboratory, Institute of Molecular and Cell Biology (IMCB), A*STAR, Singapore 138648, Singapore
[8]Institute of Medical Biology (IMB), A*STAR, Singapore 138648, Singapore
[9]Department of Medicine, National University of Singapore (NUS), Singapore 117597, Singapore
[10]Department of Physiology, National University of Singapore (NUS), Singapore 117597, Singapore
[11]These authors contributed equally
[12]Lead Contact
*Correspondence: map@pouladilab.org
https://doi.org/10.1016/j.celrep.2019.02.008

SUMMARY

In Huntington disease (HD), the analysis of tissue-specific CAG repeat length effects has been challenging, given the difficulty in obtaining relevant patient tissues with a broad range of CAG repeat lengths. We used genome editing to generate an allelic panel of isogenic HD (IsoHD) human embryonic stem cell (hESC) lines carrying varying CAG repeat lengths in the first exon of *HTT*. Functional analyses in differentiated neural cells revealed CAG repeat length-related abnormalities in mitochondrial respiration and oxidative stress and enhanced susceptibility to DNA damage. To explore tissue-specific effects in HD, we differentiated the IsoHD panel into neural progenitor cells, neurons, hepatocytes, and muscle cells. Transcriptomic and proteomic analyses of the resultant cell types identified CAG repeat length-dependent and cell-type-specific molecular phenotypes. We anticipate that the IsoHD panel and transcriptomic and proteomic data will serve as a versatile, open-access platform to dissect the molecular factors contributing to HD pathogenesis.

INTRODUCTION

Huntington disease (HD) is an inherited, autosomal-dominant, progressive, neurodegenerative disorder that is clinically characterized by a triad of motor, cognitive, and psychiatric symptoms (Walker, 2007). The disease is caused by a CAG trinucleotide expansion encoding an elongated polyglutamine (polyQ) stretch in the first exon of huntingtin (*HTT*). *HTT* alleles harboring 36 or more CAG repeats cause HD, with repeat lengths greater than 40 being fully penetrant (Walker, 2007). Despite concerted efforts, there are no effective therapies, and patients die from disease complications 10–15 years after onset.

There are two striking relationships observed in the pathogenesis of HD. First, there is an inverse correlation between the length of the CAG tract and the age of neurological onset (Langbehn et al., 2004), suggesting that the length of the expanded CAG repeats influences the rate of pathogenic events leading to disease onset. Second, although *HTT* is expressed ubiquitously, the pathological manifestations of HD exhibit regional differences with, for example, striatal and cortical tissues being more prominently affected than other brain regions (Han et al., 2010). Great efforts have been undertaken to understand the basis of the CAG length- and tissue-dependent effects. Although a number of studies have examined the relationship between CAG repeat length and various cellular processes, the mechanisms involved and their relative contribution to the pathogenesis of HD remain poorly understood. Furthermore, with rare exceptions (HD iPSC Consortium, 2017), such studies have largely relied on animal models, which may not accurately reflect disease pathogenesis in a human context.

Here, we use genome editing techniques to generate human embryonic stem cell (hESC) lines that carry a broad range of CAG repeat lengths (30, 45, 65, and 81 CAGs). The CAG lengths chosen correspond to the age of neurological onset in HD. Patients with CAG lengths greater than 60 experience disease onset before the age of 20 (adolescence), and those with CAG lengths greater than 80 experience onset before the age of 10 (juvenile); together, these patients account for 5%–7% of HD cases (Nance and Myers, 2001). Adult onset HD patients have CAG lengths

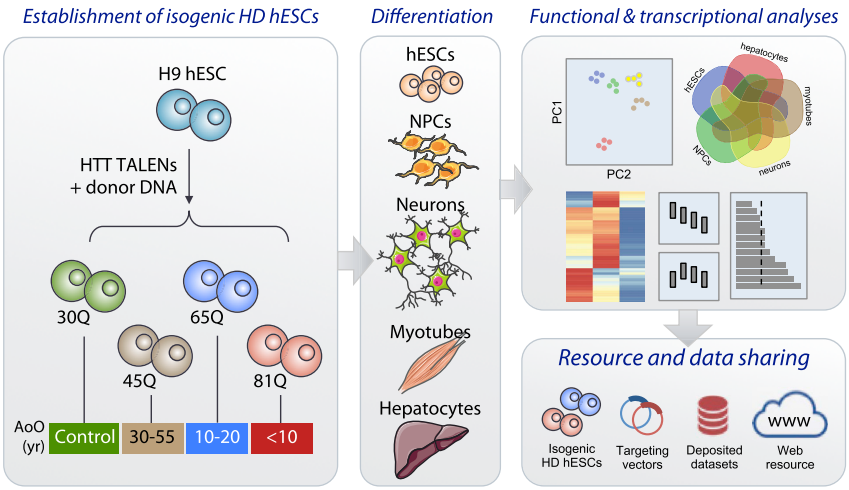


Figure 1. Study Overview and Workflow

Schematic diagram showing an overview of the study, where H9 human embryonic stem cells (hESCs) were genetically engineered to carry 30, 45, 65, and 81 CAG repeats at exon 1 of *HTT* (i.e., 30Q, 45Q, 65Q, and 81Q). The spectrum of CAG sizes in the resulting panel of hESCs cover a range in the ages of onset of HD. The CAG allelic panel and the parental unmodified H9 hESCs were differentiated into various cell lineages. Transcriptomic profiles of all the cell types and CAG lengths were obtained to derive signatures that provide molecular insights into CAG dependence and tissue specificity in the pathogenesis of HD. These tools provide a data-sharing resource for the scientific community in a wider effort to uncover molecular events that contribute to HD etiology. HTT, huntingtin; AoO, ages of onset; yr, years; Q, glutamine; NPCs, neural progenitor cells; PC, principal component.

greater than 40 and constitute most HD cases. By creating a selection cassette-free, isogenic HD (IsoHD) hESC allelic series, we ensure that all cells within the panel contain the same genetic background and that differences in functional and molecular measurements can be directly attributed to CAG length variations in a human cellular context. Functional analyses in differentiated neural cells revealed CAG repeat length-related abnormalities in mitochondrial respiration and oxidative stress, as well as elevated markers of DNA damage and enhanced susceptibility to drug-induced DNA damage, among other molecular aberrations.

As a proof of concept to explore tissue-specific effects in HD (Figure 1), we differentiated the IsoHD panel into cell types that exhibit different degrees of susceptibility to mutant *HTT*. Although HD is mainly viewed as a disease of the central nervous system, pathological changes in peripheral tissues, such as muscle wasting and metabolic abnormalities of hepatic origin, have been described (Carroll et al., 2015). Thus, we chose to analyze neural progenitor cells (NPCs), neurons, hepatocytes, and skeletal muscle myotubes derived from the IsoHD panel. We performed whole-transcriptome analysis of the resultant cell types as well as whole-proteome analysis of hESCs and NPCs, demonstrating the potential of the IsoHD experimental system to investigate the differential susceptibility of tissues and cell-types to HD pathology. Within this data set, we identified CAG repeat length-dependent and cell-type-specific molecular phenotypes that were enriched for specific functional categories and comment on the comparison of IsoHD panel with other transcriptomic HD resources. Finally, we show how this IsoHD panel serves as a versatile platform to dissect the effect of *HTT* CAG length on cellular and molecular factors that may contribute to the pathogenesis of HD in the context of human physiology. We have developed a web portal to provide query-based access to the transcriptional and proteomic data (http://isohd.pouladilab.org).

RESULTS

Generation of an IsoHD ESC Allelic Panel

H9 hESCs were chosen as the parental cell type to incorporate increasing CAG repeat lengths in exon 1 of *HTT* because they have been extensively studied, with a well-characterized propensity to differentiate into different cell types (Kobold et al., 2015). To facilitate gene modifications in the CAG tract of *HTT*, two pairs of transcription activator-like effector nucleases (TALENs), which cleave either upstream (USPQ) or downstream (DSPQ) of the CAG repeat tract (Figure 2A), were generated. DSPQ TALENs were more active than USPQ TALENs were (Figure 2B).

Donor DNA was constructed using a previously described strategy (Xu et al., 2017) and comprised homology arms carrying 30, 45, 65, or 81 CAG repeats in exon 1 of *HTT*. The homology arms flank a piggyBac selection cassette that contains genes encoding EGFP, puromycin resistance ($puro^R$), and thymidine kinase (TK; Figure 2C). DSPQ TALENs and donor DNA were transfected into H9 hESCs, and triple-antibiotic selection was employed to select for successful transformants (Figure 2D). For each of the donor DNAs carrying one of the chosen CAG repeat lengths, we obtained at least four hESC lines. We confirmed monoallelic knockin of the expanded CAG tract in all 16 clones (four clones for each CAG length) by PCR (Figure 2E). To eliminate potential confounding effects of the selection cassettes, targeted hESCs were transfected with excision-only PiggyBac transposase plasmid, and removal of the selection cassette was validated by PCR (Figure 2F). Quantification of HTT protein levels demonstrated a decrease in expression with increasing CAG repeat length (Figure 2G), which is consistent with published findings (Persichetti et al., 1996; Pouladi et al., 2010) and may relate to an influence of HTT pollyQ length on antibody affinities. The mean efficiencies were 4% for genomic targeting (Table S1) and 32% for selection-cassette excision (32 successful excisions from 100 clones screened), respectively. Overall, the resultant allelic panel contains heterozygous hESCs with a high CAG allele of 30, 45, 65, and 81 repeats in exon 1 of *HTT* (encoding HTT protein with a corresponding number of polyQs), henceforth, referred to as "IsoHD cells."

Characterization of the IsoHD hESC Panel

Using the PluriTest and G-banding analysis, we first confirmed that the IsoHD clones retained pluripotency and maintained a

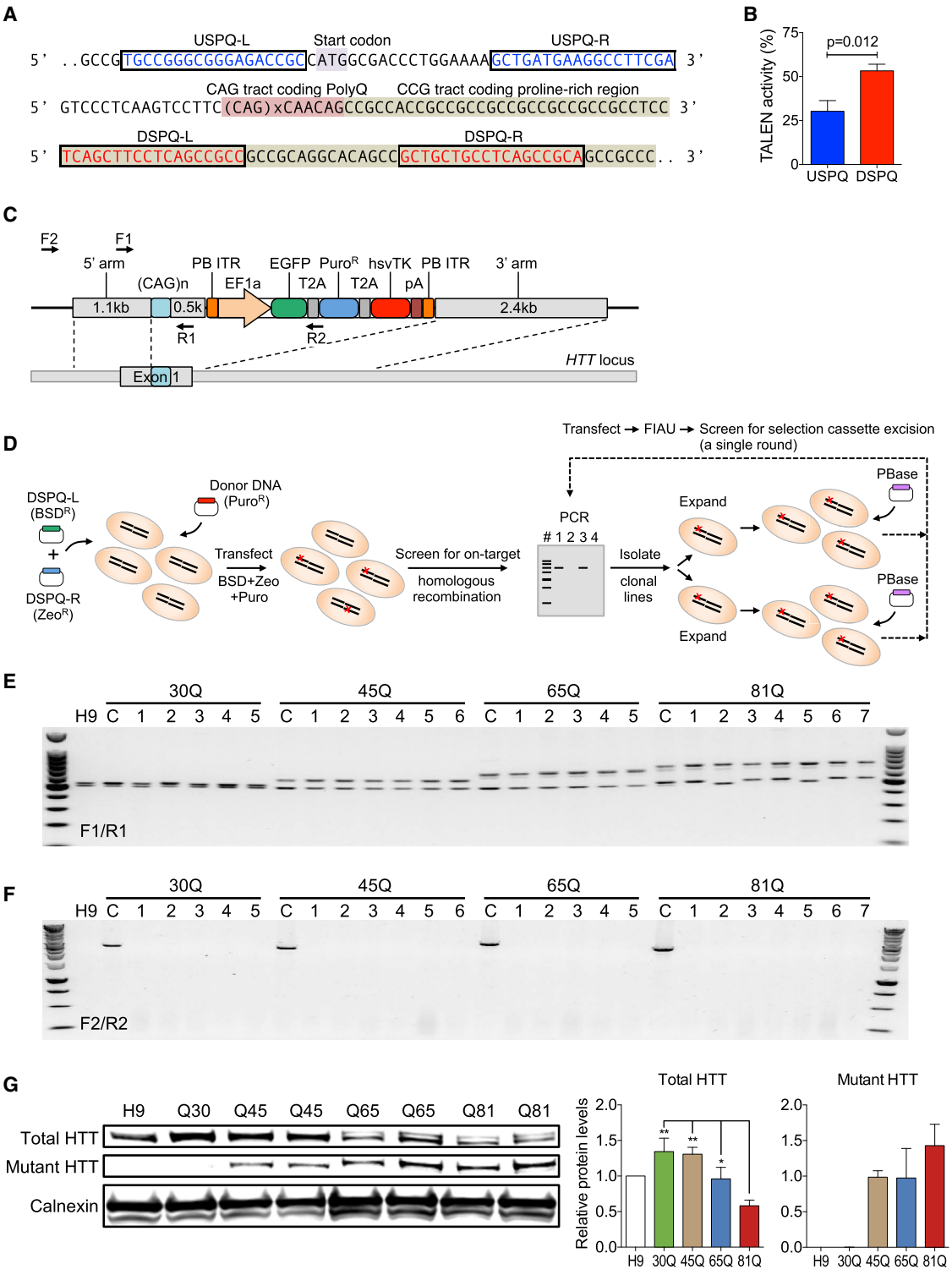


Figure 2. Gene Targeting by TALEN-Mediated, Homologous Recombination in H9 hESCs

(A) Schematic diagram showing the position of the TALEN recognition sites. Two pairs of TALEN recognition sites were designed upstream and downstream of the polyQ site, denoted as USPQ and DSPQ, respectively. L refers to the left TALEN, and R refers to the right TALEN of a given pair.

(B) TALEN activity as measured by fluorescence emitted from a surrogate reporter. n = 4 independent replicates/group; unpaired t test, p = 0.012.

(C) Schematic diagram of the donor DNA used to target the CAG tract in exon 1 of the *HTT* locus. The 5′ and 3′ homology arms of the plasmid are 1.7 kb and 2.4 kb, respectively. The selection cassette to detect targeted events was designed to reside in intron 1 of *HTT*. It is flanked by piggyBac transposon arms and comprises

(legend continued on next page)

normal 46,XX karyotype (Figures S1A and S1B). Fragment sizing analysis of the 30Q and 81Q lines revealed that the CAG repeat tracts in the IsoHD hESCs are largely stable across 15 passages (Figure S1C). As one advantage of our IsoHD panel is the isogenic background, we interrogated the IsoHD panel for copy number variants (CNVs) acquired as a result of the cellular passaging involved in targeting process (Soldner et al., 2011). We performed a genome-wide CNV analysis using an Illumina Infinium Global Screening Array and did not find a high incidence of CNVs. We identified an average of four CNVs per sample (Table S2) with four regions that were present in more than one sample (Figure S2).

We then investigated the IsoHD panel for the ability to recapitulate typical HD-related cellular phenotypes. Mitochondrial dysfunction and increased reactive oxygen species (ROS) have been previously described in HD NPCs and other cell types (Seong et al., 2005; Xu et al., 2017). We found that the IsoHD-derived NPCs (Figures S3A and S3B) showed deficits in mitochondrial function (Figure 3A) as reflected by reduced basal respiration, ATP production, and maximal respiration rates (Figure 3B). The deficit in mitochondrial respiration was more apparent in 81Q NPCs, which also exhibited lower mitochondrial membrane potential (Figure 3C). We next assessed the ROS levels in the mitochondrial and cytosolic compartments of IsoHD NPCs. Both 65Q and 81Q NPCs displayed increased mitochondrial (Figure 3D) and cytosolic ROS (Figure 3E), consistent with previous reports describing increased ROS levels in cells expressing mutant *HTT* (Mason et al., 2013).

Given that IsoHD NPCs display increased ROS levels, we examined whether that was associated with increased DNA damage by measuring γH2A.X levels, an indicator of double-stranded DNA breaks. Whole-cell protein levels of γH2A.X were significantly increased in 45Q and 81Q NPCs, with 65Q NPCs exhibiting a similar trend (Figure 4A). This observation is supported by the increased number of γH2A.X foci per cell in 81Q versus 30Q NPCs, as determined by immunostaining (Figure 4B). To reveal additional phenotypes, two DNA-damaging agents, tert-Butyl hydroperoxide (TBHP) and doxorubicin (DOX), were used to induce DNA damage. Both agents led to an increase in the percentage of cells that exhibited detectable γH2A.X foci (Figure 4C). Treated 81Q NPCs were most sensitive to induced DNA damage, showing the greatest increase in cells with detectable γH2A.X foci, which was significantly greater than that observed for treated 30Q and 45Q NPCs (Figure 4C). A similar trend was observed when cells were stained for 53BP1, an additional DNA damage marker (Figures S3C and S3D). However, we noted that the changes in 53BP1 foci were more striking in cells treated with DOX, a more potent DNA-damaging agent (Kurz et al., 2004; Figure S3D) compared with TBHP (Figure S3C). This may be due to low endogenous 53BP1 levels, which limit its detection by microscopy (Banáth et al., 2009). To ascertain that the increased sensitivity to induced DNA damage observed with increased CAG repeats was not limited to NPCs, hESCs were treated with TBHP, resulting in a similar trend (Figure S3E).

To determine whether the increased amount of γH2A.X levels in the presence of cell stressors led to increased cell death, YOYO-1, a fluorescent DNA staining dye, was incorporated as a surrogate marker for apoptotic cells. Tracking YOYO-1 uptake across 40 h demonstrated that the fold change in the number of apoptotic 81Q and 65Q NPCs was significantly elevated as compared with 30Q and 45Q NPCs (Figure 4D).

p53 and other mitotic factors are known to be dysregulated in HD mutant cells (Steffan et al., 2000); therefore, we also measured cell replication in IsoHD NPCs. Although there was no clear difference between the different mitotic phases across the IsoHD panel (data not shown), 81Q NPCs exhibited significantly less labeling of Ki67 compared with 30Q NPCs at the same cell-passage number, indicative of fewer proliferating cells (Figure S3F). A similar trend was observed in hESCs (Figure S3G).

In summary, we demonstrate that previously observed HD-specific cellular phenotypes, such as mitochondrial dysfunction and dysregulated ROS levels, can be recapitulated in the IsoHD panel. We also identify less well-characterized features of HD mutant cells, such as alterations in DNA damage and cell proliferation.

Derivation of Central Nervous System and Peripheral Cell Types from the IsoHD Panel

We leveraged the differentiation potential of the IsoHD panel to assess tissue-specific molecular signatures. To provide a comprehensive analysis of tissue-specific molecular phenotypes, a combination of CNS and peripheral cell types with differential vulnerabilities to mutant *HTT* were chosen. To that end, IsoHD hESCs were differentiated into NPCs, neurons, hepatocytes, and myotubes. Published protocols or commercially available differentiation protocols were used (see STAR Methods for details). Differentiation efficiency was assessed by the expression of cell-type-specific transcripts and protein markers.

First, IsoHD hESCs were examined to ensure that site-directed modification of *HTT* did not compromise the pluripotent nature of the cells. Consistent with the PluriTest results (Figure S1A), IsoHD hESCs exhibited uniform expression of OCT4 and SSEA-4 protein markers, similar to the parental H9 hESCs, indicating no deviation from their original pluripotent nature (Figure 5A). OCT4, LIN28, and NANOG transcript levels were

puromycin resistance ($Puro^R$) and EGFP under the EF1α promoter. F1, F2, R1, and R2 indicate the position of PCR primers used for (E) and (F). PB ITR, piggyBac inverted terminal repeats; T2A, 2A self-cleaving linker peptide; hsvTK, herpes simplex virus thymidine kinase type 1 negative selection marker.

(D) Workflow to generate the IsoHD panel.

(E) Gel image showing amplicons that correspond to the CAG tract of the targeted lines. H9 denotes the parental untargeted hESC line. Control lines (C) for each CAG size denotes a representative IsoHD line before the selection cassette was excised.

(F) Gel image of genomic PCR showing amplicons that correspond to the junction of the donor DNA and host genome. Cells that lack the selection cassette should not produce the amplicon. The control lines (C) have the selection cassette intact and act as a positive PCR control.

(G) IsoHD lines express the mutant protein. Four clones from each CAG size were chosen for further characterization. (Left) western blot images showing expression of HTT (MAB2166) and mutant HTT (MW1). (Right) Band intensities were quantified and values normalized to the calnexin housekeeping protein. n = 4 biological replicates/IsoHD line. Values shown as means ± SEM; *p < 0.05, **p < 0.01, as determined by one-way ANOVA with Fisher's LSD.

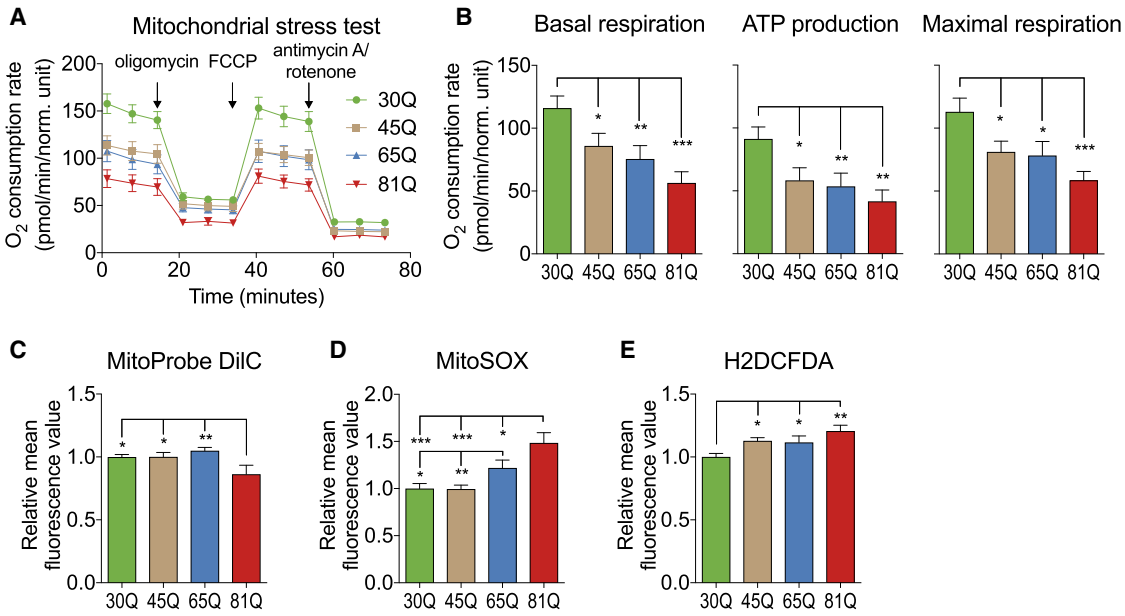


Figure 3. IsoHD Neural Cells Exhibit Mitochondrial Deficits and Increased Levels of Reactive Oxygen Species

(A and B) Measures of mitochondrial respiration in IsoHD neural progenitor cells (NPCs).

(A) Oxygen consumption rates (OCR) of IsoHD NPCs under a mitochondrial stress test. Seahorse experiments were performed with NPCs derived from IsoHD hESCs.

(B) CAG-dependent effects on mitochondrial function, as demonstrated by graphs showing OCR of the IsoHD NPC panel during the mitochondrial stress test to reflect basal respiration, ATP production, and maximal respiration.

For (A) and (B), n = 37, 37, 42, and 28 replicates for 30Q, 45Q, 65Q, and 81Q, respectively. Values shown as means ± SEM; *p < 0.05, **p < 0.01, ***p < 0.001, as determined by one-way ANOVA with Fisher's LSD.

(C) IsoHD 81Q NPCs show compromised mitochondrial membrane potential. MitoProbe DilC was used to measure mitochondrial membrane potential in NPCs derived from IsoHD hESCs. n = 12, 12, 15, and 9 biological replicates for 30Q, 45Q, 65Q, and 81Q, respectively. Values shown as means ± SEM; *p < 0.05, **p < 0.01, as determined by one-way ANOVA with Fisher's LSD.

(D) IsoHD 65Q and 81Q NPCs have increased mitochondrial reactive oxygen species (ROS) levels using mitoSOX. n = 16, 16, 20, and 12 biological replicates for 30Q, 45Q, 65Q, and 81Q, respectively. Values shown as means ± SEM; *p < 0.05, **p < 0.01, ***p < 0.001, as determined by one-way ANOVA with Fisher's LSD.

(E) NPCs with increasing CAG lengths show increasing cytosolic ROS levels using H2DCFDA. n = 12, 12, 15, and 9 biological replicates for 30Q, 45Q, 65Q, and 81Q, respectively. Values shown as means ± SEM; *p < 0.05, **p < 0.01, as determined by one-way ANOVA with Fisher's LSD.

also maintained across the IsoHD lines (Figure 5A, bottom). CAG repeat instability is a well-established phenomena in HD and has been shown to be both age- and tissue-dependent (Wheeler et al., 1999). There were no marked changes in the size of the CAG repeat tract in the four cell lineages differentiated from the IsoHD hESCs (Figure S4).

To obtain neuronal populations, we adapted a previously described protocol (Xu et al., 2017). At day 11 of differentiation, we obtained NPCs expressing PAX6, NESTIN, and FOXG1 (Figure 5B). These cells were then replated and provided with a cocktail of growth factors to terminally differentiate them into post-mitotic neurons. The matured cells expressed the neuronal markers MAP2 and DARPP32 (Figure 5C).

Hepatocytes were derived using a modified published protocol (Szkolnicka et al., 2014). Cuboidal cells resembling hepatocytes were obtained after 20 days. The cells were assessed at various points during differentiation, and appropriate cell-specific markers were expressed at the transcript and protein levels (Figure 5D). Moreover, the cells were able to secrete albumin (data not shown), indicative of functional hepatocytes.

Differentiation of the IsoHD panel into myotubes was performed with a commercially available platform (Caron et al., 2016). Undergoing three stages of differentiation, the cells were assessed at each point to ensure that early and late markers of myotube differentiation were expressed. At the end of 28 days, the cells displayed tubular structures, resembling myotubes (Figure 5E), and there were no differences in the transcript levels of the maturation markers *MYH3* and *MYH8* (Schiaffino et al., 2015) between the groups, indicating comparable level of maturation (data not shown). Therefore, our IsoHD hESCs were found to be stable and capable of differentiating into a range of relevant cell types.

Transcriptional Profiling of the Derived CNS and Peripheral Cell Types

To investigate CAG length- and cell-type-specific transcriptional changes, we performed genome-wide RNA sequencing on four lines per CAG length for each of the five cell types. Subsequent analysis of the RNA sequencing revealed transcriptional differences between both CAG lengths and cell types. Transcriptional profiles separated according to cell type, with myotubes and hepatocytes exhibiting more variability than hESCs, NPCs, and neurons (Figures 6A and 6B). Differential gene expression was assessed in each of the cell types and across the CAG lengths

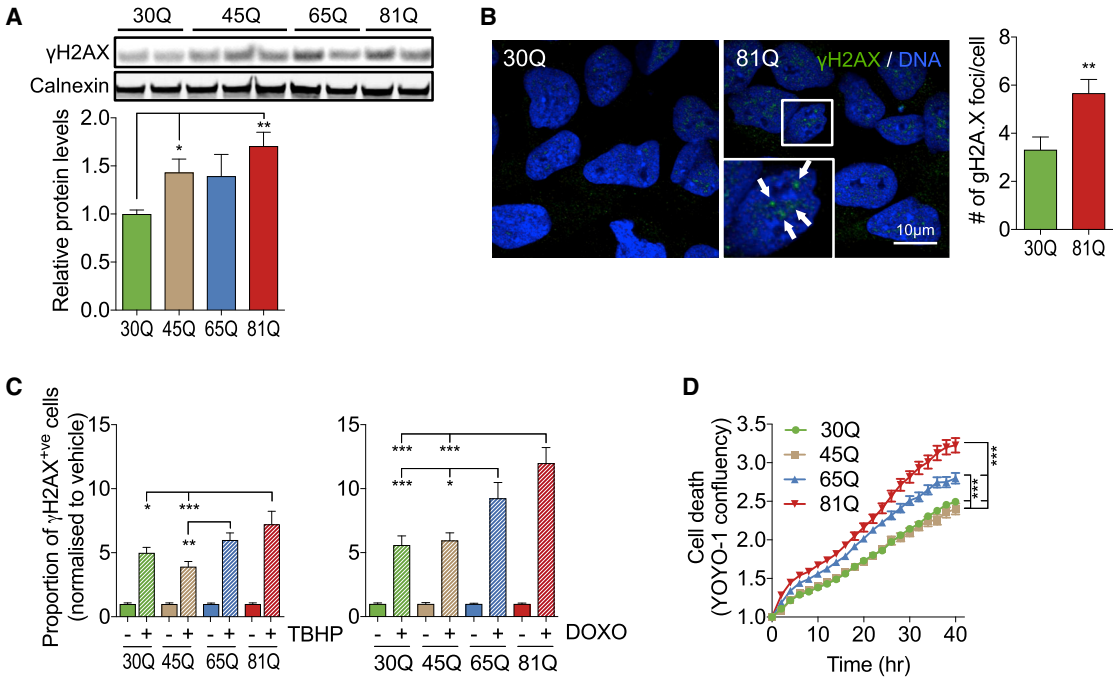


Figure 4. Increased Susceptibility to DNA Damage in IsoHD Neural Cells

(A) DNA damage marker γH2A.X is increased in mutant IsoHD NPCs. Whole-cell lysates from IsoHD NPCs were analyzed for γH2A.X protein levels by western blotting. The band intensities were quantified. n = 4, 3, 2, and 3 biological replicates for 30Q, 45Q, 65Q, and 81Q, respectively. Values shown as means ± SEM; *p < 0.05, **p < 0.01, as determined by one-way ANOVA with Fisher's LSD.

(B) IsoHD 81Q NPCs have an increased number of γH2A.X foci per cell. IsoHD NPCs were grown to confluency and stained with a γH2A.X antibody. Images of the cells were taken at 100× magnification, and the images were processed using ImageJ software to count the number of foci per cell that exceeded the background threshold. n = 47 cells for 30Q and 61 cells for 81Q. Values shown as means ± SEM; **p < 0.01, as determined by an unpaired t test.

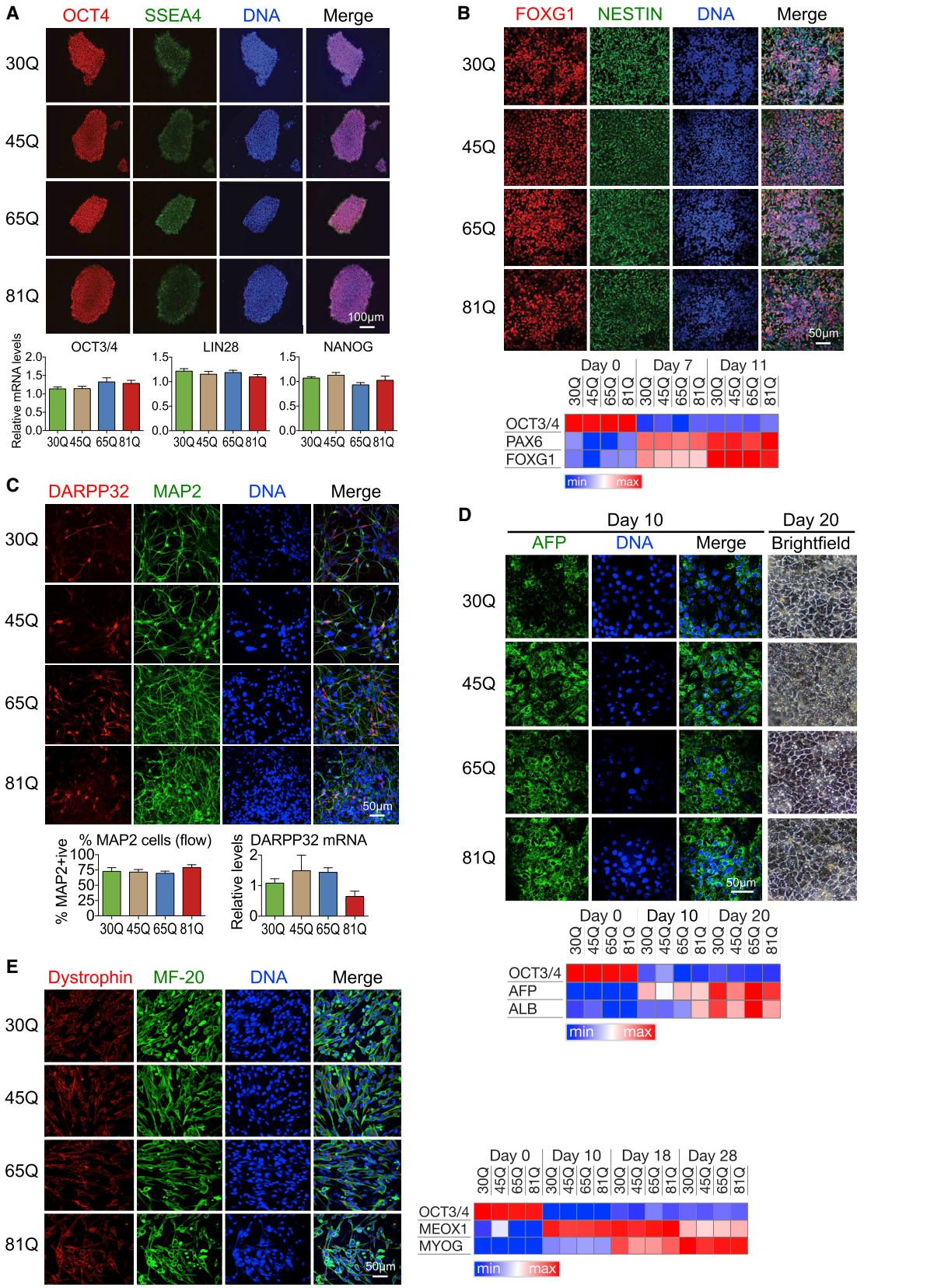
(C) IsoHD 81Q NPCs are more susceptible to DNA damage in the presence of DNA-damaging agents. Tert-butyl hydrogen peroxide (TBHP) and doxorubicin (DOXO) were added to IsoHD NPCs in separate experiments to induce DNA damage. Number of γH2A.X foci was determined as described in (B). For TBHP, n = 24, 24, 30, and 18 cells for 30Q, 45Q, 65Q, and 81Q, respectively; for DOXO, n = 50, 48, 61, and 38 cells for 30Q, 45Q, 65Q, and 81Q, respectively. Values shown as means ± SEM; *p < 0.05, **p < 0.01, ***p < 0.001, as determined by one-way ANOVA with Fisher's LSD. −, vehicle treatment; +, treatment with DNA-damaging agent (TBHP or DOXO).

(D) IsoHD 81Q NPCs show increased cell death over time. The IsoHD NPC panel was tracked with IncuCyte Zoom, with the addition of YOYO-1, which stains for dead cells to emit a green fluorescence. Number of cells emitting green fluorescence was calculated over 40 h. n = 72, 71, 90, and 54 images for 30Q, 45Q, 65Q, and 81Q, respectively, which were pooled from four biological replicates/genotype. Values shown as means ± SEM; ***p < 0.001, as determined by two-way ANOVA with Fisher LSD *post hoc* test. hr, hours.

(false discovery rate [FDR] = 10%). Significantly differentially expressed genes showed strong cell type specificity (Figure 6C; Table S3), with only eight genes differentially expressed in all cell types: *ZNF785* (zinc finger protein 785), *SHROOM2* (shroom family member 2), *TP53* (tumor protein 53), *EPHB2* (*EPH* receptor B2), *SYT11* (synaptotagmin), *CHCHD2* (coiled-coil-helix-coiled-coil-helix domain containing 2), *CDK6* (cyclin dependent kinase 6), and CD99L2 (CD99 molecule like 2). The number of significantly differentially expressed genes (DEGs) ranged from 352 (2.25% of expressed genes) in NPCs to 6,419 (35.7% of expressed genes) in hepatocytes (Figures 6D–6H).

The DEG in the NPC and hESC populations were not significantly enriched for functional terms (g:Profiler, FDR < 5%), whereas DEG in neurons were enriched in gene ontology terms relating to the collagen metabolic process, myotubes were enriched in terms relating to cardiovascular system development and vasculature development, and DEG in hepatocytes were enriched in terms relating to the carboxylic acid metabolic process and protein-lipid complex subunit organization (g:Profiler, FDR < 5%; Table S4).

To identify transcriptional signatures for CAG lengths, fold changes were calculated for each DEG with respect to 30Q; those fold changes were scaled and clustered within each cell type, resulting in 12–18 clusters across the cell types (Figures 6D–6H). Each cluster was analyzed for significant functional enrichments showing functional specialization (see Table S4). Of note, in the hESCs, the magenta cluster (clusters are indicated by colors in the bar; Figure 6I) was significantly enriched for ontology terms relating to RNA binding and RNA processing (g:Profiler, FDR < 5%; Table S4). For neurons, the midnight blue cluster (Figure 6J) was enriched for biological process terms relating to cell cycle progression (g:Profiler, FDR < 5%; Table S4). For hepatocytes, the turquoise cluster (Figure 6K) was enriched for immune response terms (g:Profiler, FDR < 5%; Table S4). For myotubes, the black cluster (Figure 6L) was enriched for mitochondrial and NADH dehydrogenase ontology



(legend on next page)

terms (g:Profiler, FDR < 5%; Table S4). Other clusters not mentioned were also significantly enriched for functional terms and pathways (Table S4); however, there were no significant functional enrichments for the clusters in NPCs, likely because of the small number of DEG as compared with the DEG detected in other cell types. Furthermore, the patterns of clustering highlight the non-monotonic relationship of CAG length with many of the DEGs.

Comparison with Published Transcriptomic Data

Several models of HD, both patient-derived, induced pluripotent stem cells (iPSCs) and animal based, have been transcriptionally profiled to inform on the underlying molecular basis of the disease. We have reanalyzed those transcriptional data to compare with our IsoHD transcriptional profiling.

The HD iPSC Consortium (2017) generated and profiled human neuronal cultures from iPSC from control (21–33 CAG repeats) and HD patients (60–109 CAG repeats; referred to here as "HDC iPSC"). We downloaded the publicly available RNA-sequencing data and processed it (Figures S5A–S5F) using the same pipeline as our IsoHD RNA-sequencing data (see STAR Methods). A multidimensional scaling (MDS) of the transcriptome reveals a separation in transcriptional profiles between the control and HD samples (Figure S5G). This separation is reflected in the differential expression analysis between the control and HD iPSC neurons because there were 2,980 DEGs (Table S5; likelihood ration test [LRT]; FDR < 10%). *HTT* was expressed at comparable levels in both the control and HD samples (adjusted p value = 0.26; Figure S5H). The gene with the strongest evidence for differential expression was *XIST*, X-inactive specific transcript (Figure S5I; adjusted p value = 2.53×10^{-28}). However, the differences in expression in *XIST* were driven by the differences in gender between the control and HD samples (control, male; HD, female) and not as a result of disease pathology. Hierarchical clustering of the DEGs show two groups, one upregulated in HD and one downregulated (Figure S5J). Within those clusters, there was variability in expression levels within the groups. This may reflect the differing CAG lengths in the control and the HD samples and the shared genetic background of two of the HD samples. The DEGs from the HDC neurons showed an overlap of 471 genes with the DEGs from IsoHD neurons across the CAG lengths (Figure S5K). The overlap between the HDC DEGs and the IsoHD neurons between the 30Q and 81Q was 16 (Figure S5L). Neither of the overlapping gene sets were significantly enriched for functional pathways and terms.

A large mouse HD resource was generated and made publicly available by Langfelder et al. (2016), comprising transcriptional data for multiple tissues from a knockin allelic series (referred to here as the "HDinHD" series). To compare with our IsoHD data, we have downloaded and analyzed RNA-sequencing data in the striatum and liver for the expansion series of 20, 50, and 80 CAG lengths. In the transcriptional profile of the mouse striatum, the MDS does not show a clear separation by CAG length (Figure S6A) or between male and female samples (Figure S6B). There were 2,801 DEGs (Table S6; FDR < 10%, adjusted for gender), and when clustered, the expression levels both showed an effect of the CAG lengths (Figure S6C). *Htt* was expressed at comparable levels across the CAG lengths (adjusted p value = 0.82; Figure S6D). The gene with the strongest evidence for differential expression was *Gm6410*, a long non-coding RNA (Figure S6E; adjusted p value = 6.24×10^{-11}); *Gm6410* does not have a human ortholog. The DEGs from the HDinHD striatum showed an overlap of 464 genes with the DEGs from the IsoHD neurons across the CAG lengths (Figure S6F). The overlap between the pairwise DEGs at 20Q and 91Q and between the 30Q and 81Q was 1 (Figure S6G).

In the transcriptional profile of the mouse liver, the MDS does not show a clear separation by CAG length (Figure S6H) but does show a strong separation between male and female samples (Figure S6I). There were 239 DEGs (Table S6; FDR < 10%, adjusted for gender), and when clustered, the expression levels both showed an effect of the CAG lengths but also of the gender for the sample (Figure S6J). *Htt* was expressed at significantly decreasing levels with longer CAG lengths (adjusted p value = 0.09; Figure S6K). The gene with the strongest evidence for differential expression was *Igha* (immunoglobulin heavy constant alpha; Figure S6L; adjusted p value = 3.20×10^{-07}). *Igha* is an

Figure 5. Differentiation of the IsoHD Panel into Various Cell Lineages

(A) The IsoHD hESC lines retain pluripotency markers. (Top) Immunofluorescence images showing expression of OCT4 and SSEA-4, both are proteins that are highly expressed in pluripotent stem cells. (Bottom) Graphs showing the transcript levels of OCT3/4, LIN28, and NANOG, as assessed by quantitative PCR. n = 12 biological replicates per genotype. Values, shown as means ± SEM, have been normalized to the transcript levels of beta-actin, which acts as an internal control.

(B) Differentiation of the IsoHD panel into neural progenitor cells. (Top) Immunofluorescence images showing expression of FOXG1 and NESTIN. NESTIN is an early neuronal marker, whereas FOXG1 is an early forebrain marker. (Bottom) Heat map showing relative transcript levels of OCT3/4, PAX6, and FOXG1, as assessed by quantitative PCR. PAX6 and FOXG1 are expressed as early as day 7, whereas OCT3/4 is rapidly downregulated upon differentiation.

(C) Differentiation of the IsoHD panel into post-mitotic neurons. (Top) Immunofluorescence images showing expression of MAP2 and DARPP-32. MAP2 is a mature neuronal marker, whereas DARPP-32 is a marker for medium spiny neurons, the affected neuronal population in HD. (Bottom, left) Graph showing the purity of mature neurons after spontaneous differentiation. The differentiated culture was dissociated and analyzed by flow cytometry after MAP2 labeling. Each individual point represents an IsoHD clone that comprises the corresponding CAG repeat length. (Bottom, right) Graph showing the transcript levels of DARPP-32, as assessed by quantitative PCR. n = 3, 4, 3, and 4 biological replicates for 30Q, 45Q, 65Q, and 81Q, respectively. Values, shown as means ± SEM, have been normalized to the transcript levels of beta-actin, which acts as an internal control.

(D) Differentiation of the IsoHD panel into hepatocytes. (Top) Immunofluorescence images showing expression of the endodermal marker, alpha fetoprotein (AFP). The brightfield images reflect the organization of the hepatocytes into a globular arrangement, similar to that *in vivo*. (Bottom) Heat map showing relative transcript levels of OCT3/4, AFP, and albumin (ALB), as assessed by quantitative PCR.

(E) Differentiation of the IsoHD panel into myotubes. (Left) Immunofluorescence images showing expression of mature myotube markers, dystrophin, and myosin. (Right) Heat map showing relative transcript levels of OCT3/4, MEOX1, and MYOG, as assessed by qPCR. OCT3/4 is rapidly downregulated upon differentiation, whereas expression of MEOX1, an early muscle marker, was upregulated on days 10 and 18, but quickly downregulated by day 28. That is in contrast to MYOG, a mature muscle marker, expression of which is only upregulated from day 18, and markedly increased by day 28.

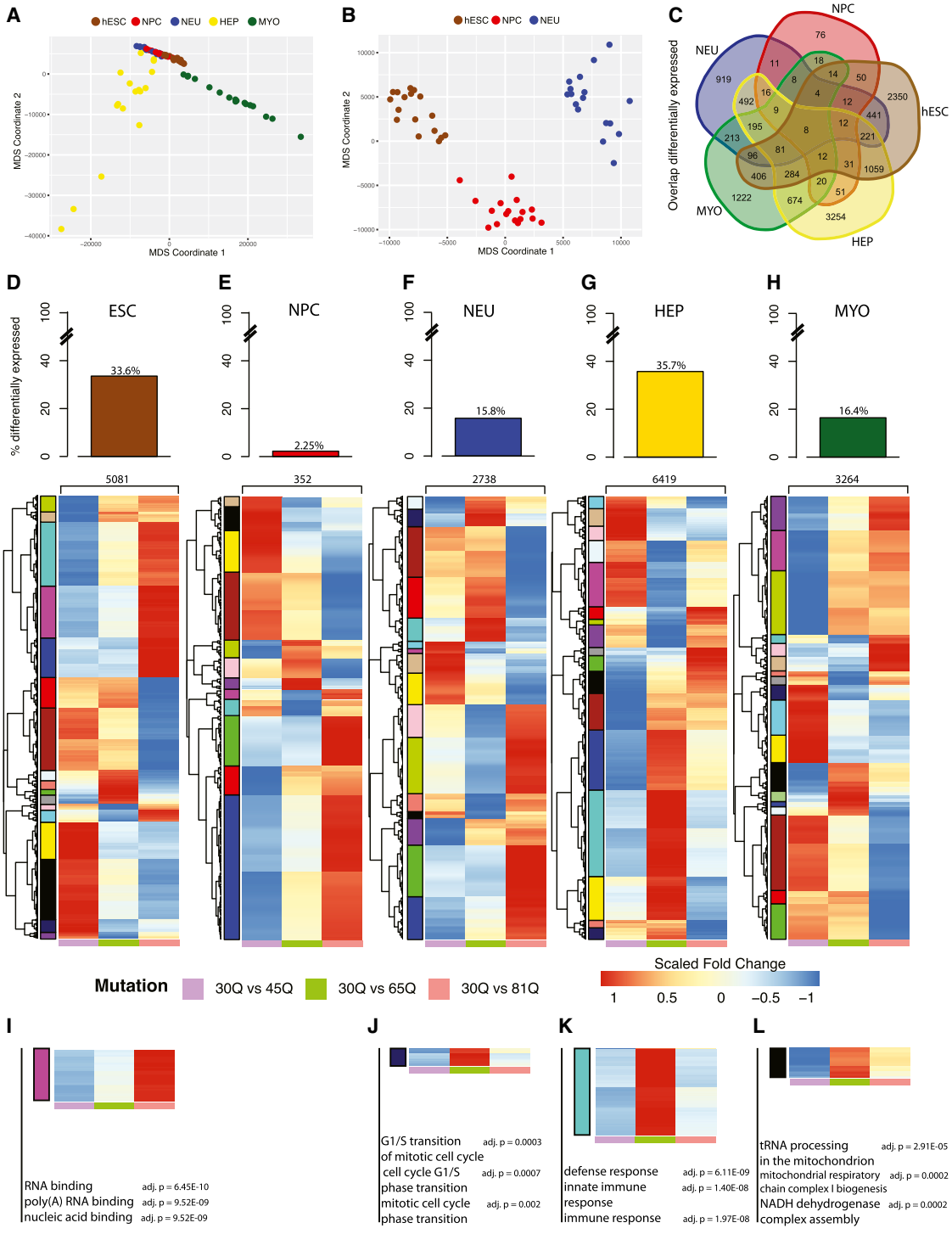


Figure 6. Transcriptional Landscape across CAG Lengths in IsoHD hESC-Derived CNS and Peripheral Cells

(A) Multi-dimensional scaling (MDS) of gene expression across hESC (brown), hepatocyte (HEP; yellow), myotube (MYO; green), NPC (red), and neuron (NEU; blue), of the first MDS component (x axis) and second MDS component (y axis).

(B) MDS of gene expression across hESC (brown), NPC (red), and NEU (blue), of the first MDS component (x axis) and second MDS component (y axis).

(C) Venn diagram showing the overlap in genes that were significantly differentially expressed by CAG length in each of the cell types (LRT, FDR = 10%; n = 4 biological replicates/genotype/cell type).

(D–H) The percentage of expressed genes within each cell type that were significantly differentially expressed across CAG length (FDR = 10%), (D) hESC, (E) NPC, (F) NEU, (G) HEP, and (H) MYO. (Bottom) Heat maps and hierarchical clustering based on fold changes with respect to 30Q. For each cell type, the significantly

(legend continued on next page)

ortholog to human immunoglobulin heavy constant alpha 1 (*IGHA1*) and immunoglobulin heavy constant alpha 2 (*IGHA2*). The DEGs from the HDinHD liver showed no overlap with the DEGs from the IsoHD hepatocytes across the CAG lengths (Figure S6M). The overlap between the pairwise DEGs at 20Q and 91Q and the IsoHD hepatocytes between the 30Q and 81Q was 72 (Figure S6N). The minimal overlap may be due to the effect of gender in the HDinHD liver samples and the murine versus human comparison.

Proteomic Analysis of IsoHD hESCs and NPCs

TMT-10plex proteomics analysis was performed in both hESCs and NPCs on each CAG length of the four lines. Multidimensional scaling clearly separated the hESCs and NPCs, with NPCs exhibiting greater variability than the hESCs (Figure 7A). Differential expression analysis at the protein level revealed 37 proteins that were differentially expressed in the hESCs (0.5%; FDR < 10%, moderated ANOVA) and 370 proteins that were significantly differentially expressed in the NPCs (5.4%; FDR < 10%, moderated ANOVA; Figure 7B; Table S7). Thirteen proteins were identified as differentially expressed in both cell types, including SAP30 (Sin3A associated protein 30), TBL1X (transducin beta like 1X-linked), ZGPAT (zinc finger CCCH-type and G-patch domain containing), and ZNF3 (zinc finger protein 3). When we compared the results from the transcriptomic and proteomic analysis, we could identify genes and proteins that exhibited the same expression patterns at both the gene and protein level, e.g., TBL1X (Figure 7C). We also identified genes that were significantly differentially expressed at the transcript level but not at the protein level, such as INPP5F (Figure 7D). In addition, we identified proteins that were significantly differentially expressed but did not show differential expression in the corresponding gene, such as *PJA2* and *EIF2S2* (Figures 7E and 7F).

In summary, we observed a complex interdependence of transcript and protein abundance across hESCs and NPCs. This complexity advocates the study of CAG-dependent, cell-autonomous effects, which should be investigated at both the transcript and protein levels. We, therefore, provide the reader with an easy-to-use and interactive platform (http://isohd.pouladilab.org) to enable analyses of CAG repeat length at the gene or protein level across cell types.

DISCUSSION

Here, we describe the generation and characterization of an isogenic allelic hESC panel to explore CAG-dependent and tissue-specific effects that may contribute to the pathogenesis of HD. We took advantage of advances in genome editing to create an allelic panel carrying different CAG sizes (30, 45, 65, and 81 repeats) within exon 1 of *HTT* (Figure 1).

To determine whether our IsoHD cells were able to mimic hallmarks of HD pathology, we assessed a number of cellular phenotypes previously associated with HD. Mitochondrial abnormalities have been strongly implicated in the pathogenesis of HD and are suggested to cause energy impairments that lead to metabolic and neurological deficits in patients with HD (Costa and Scorrano, 2012). Consistent with previous findings (Xu et al., 2017), IsoHD mutant cells demonstrated impaired basal and maximal respiration and ATP production. In our transcriptomic profiling of the peripheral IsoHD cell types, we also saw clusters of DEGs in myotubes and hepatocytes that were significantly enriched for mitochondrial, respiration, and ATP-related functional terms. Mitochondrial dysfunction in skeletal muscle and liver cells has been observed in HD patients, indicating that the effects of mutant *HTT* on mitochondrial function may not be limited to neuronal cell types (Stüwe et al., 2013; Zielonka et al., 2014).

Although mutant HTT aggregation is a pathological hallmark of HD, we (Xu et al., 2017) and other groups (An et al., 2012; HD iPSC Consortium, 2012) have previously shown that HD neurons differentiated from hPSCs do not show reactivity to EM48, an antibody typically used to detect mutant HTT aggregates. This may be due to the relatively early time point at which the cells are assessed after differentiation (6–8 weeks). Indeed, in studies in which the differentiated neurons were kept for prolonged periods of time *in vitro* (~24 weeks; Nekrasov et al., 2016) or as transplants *in vivo* (~33 weeks; Jeon et al., 2012), mutant HTT aggregates have been observed.

There is growing evidence implicating the DNA damage response system in the pathogenesis of HD. This includes evidence of elevated ATM signaling (Lu et al., 2014), as well as human genome-wide association studies (GWAS), which identified polymorphisms in DNA handling factors as potential modifiers of the age at neurological onset in HD (GeM-HD Consortium, 2015; Moss et al., 2017). In our IsoHD panel, we noted an increased number of foci expressing DNA damage markers (e.g., γH2A.X) with longer CAG repeat lengths, and that was associated with reduced cellular proliferation, possibly as a result of increasing DNA integrity signaling. Interestingly, there was minimal instability in the CAG tracts of the IsoHD panel and a low incidence of copy number variation across 15 passages or upon differentiation. This is distinct from the age-dependent CAG repeat instability observed in the striatum and liver of mutant *HTT* knockin mice (Lee et al., 2011). This discrepancy is likely due to the high fidelity of the DNA repair machinery in hESCs and the relatively young age of the differentiated cells examined.

HD pathology exhibits differential tissue vulnerability with striking atrophy of basal ganglia and cortical structures (Aylward et al., 2004) but is less pronounced and has poorly characterized abnormalities in other tissues, such as the liver (Sathasivam et al., 1999). This is despite broad expression of *HTT* in many parts of the body. To explore this phenomenon, we differentiated the IsoHD hESC panel into cell types with varying perceived

differentially expressed genes (the number indicated at the top of the heat map) were clustered into a hierarchical tree using Euclidean distance and cut at a height of 0.75 to form clusters of genes given by the color bar to the left. The fold changes corresponding to 30Q versus 45Q, 30Q versus 65Q, and 30Q versus 81Q are indicated by the colored bars at the bottom: purple, green, and pink, respectively. The colors of the heat map indicate the direction of the fold change, where red is upregulated with respect to 30Q, and blue is downregulated with respect to 30Q.
(I–L) Top three significant functional enrichment terms (FDR < 5%) within select differentially expressed gene clusters for hESC (I), NEU (J), HEP (K), and MYO (L).

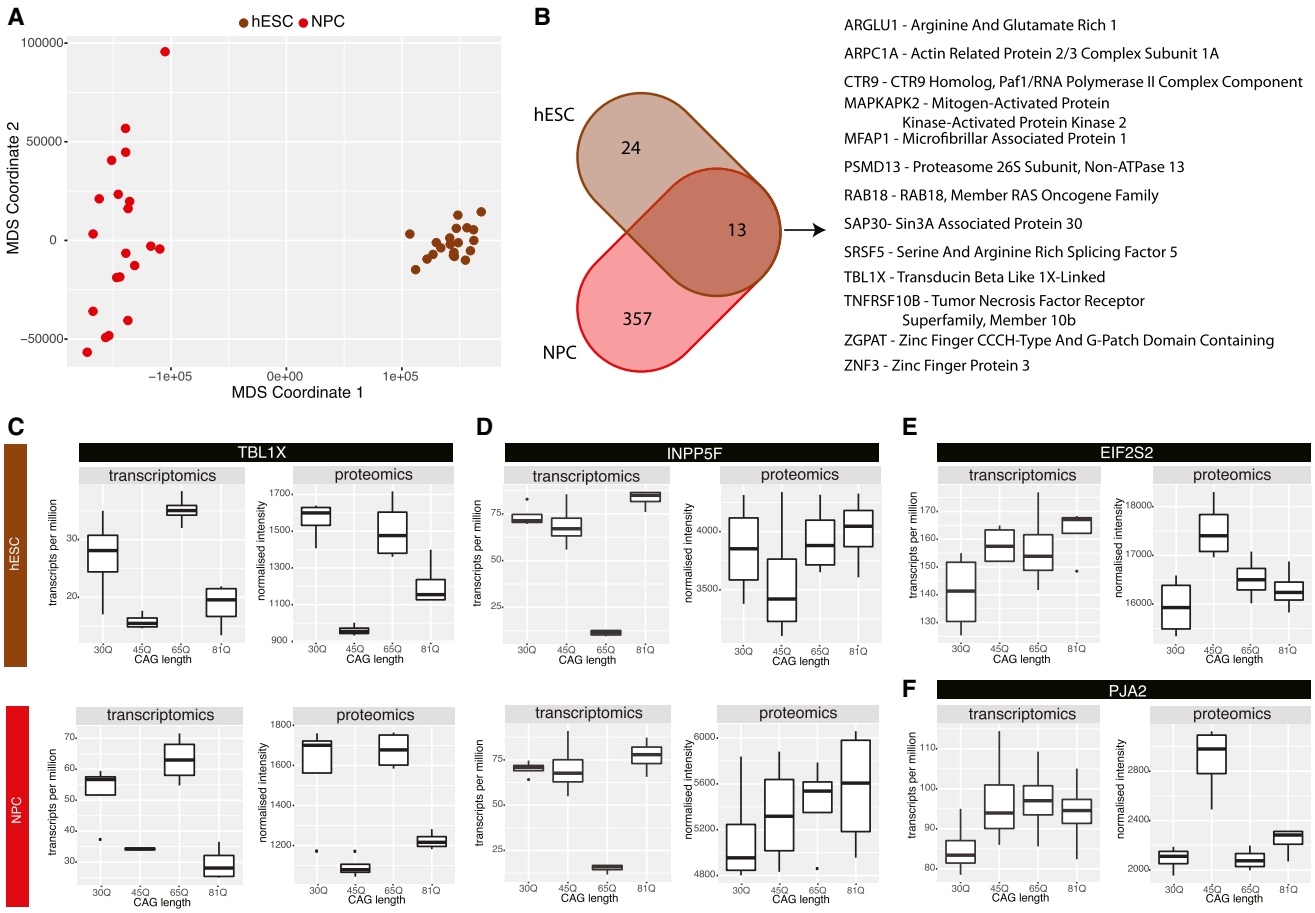


Figure 7. Whole Proteome Analysis in IsoHD hESCs and Derived Neural Progenitor Cells

(A) Multi-dimensional scaling (MDS) of protein expression across hESC and NPCs of the first MDS component (x axis) and second MDS component (y axis).

(B) Venn diagram showing the overlap in proteins that were significantly differentially abundant by CAG length in hESCs and NPCs (moderated ANOVA, FDR = 10%; hESC, n = 16; NPC, n = 16).

(C–F) Comparison of mRNA transcript and protein levels for (C) TBL1X (hESC transcriptomic FDR = $2.79e^{-07}$; hESC proteomic FDR = 0.069; NPC transcriptomic FDR = 0.0001; NPC proteomic FDR = 0.042), (D) INPP5F (hESC transcriptomic FDR = 1.71e-79; hESC proteomic FDR = 0.28; NPC transcriptomic FDR = 2.41e-34; NPC proteomic FDR = 0.40), (E) EIF2S2 (hESC transcriptomic FDR = 0.21; hESC proteomic FDR = 0.069), and (F) PJA2 (NPC transcriptomic FDR = 0.59; NPC proteomic FDR = 0.006). Differential gene expression was calculated using an LRT with a 10% FDR significance level and differential protein expression was calculated using a moderated ANOVA with a 10% FDR significance level.

susceptibilities to the *HTT* mutation to examine the interaction between tissue- and CAG size-related effects in HD. A differential expression analysis against CAG length revealed strong cell type specificity, with only eight genes differentially expressed in all cell types. Those eight genes included *TP53*, a known interaction partner with mutant HTT in neuronal cell types that affects mitochondrial and respiratory chain functions (Bae et al., 2005; Matoba et al., 2006), and *CHCHD2*, a factor linked to mitochondrial function, which we (Xu et al., 2017) and others (Feyeux et al., 2012) have shown to be upregulated in HD. These findings parallel the deficits in mitochondrial function that we observed in IsoHD cells.

Surprisingly, hepatocytes, which are considered to be relatively spared from HD pathology, were found to have the largest transcriptional response, with significant differential expression across CAG lengths in ~36% of the expressed genes. Within those DEGs, we identified a cluster of 1,253 of genes that was strongly enriched for immune response genes. In our analysis of the HDinHD mouse liver, the gene with the strongest evidence for differential expression was *IGHA*, an immunoglobulin gene that exhibited a similar expression pattern, with respect to CAG length, to that of the gene cluster in hepatocytes. Hepatocytes secrete immune proteins into circulation (Zhou et al., 2016), and HD patients at the pre-symptomatic and symptomatic stages of the disease exhibit altered levels of inflammatory mediators (Björkqvist et al., 2008) in circulation. These immune proteins may cross into the CNS directly or activate downstream immune responses, which can then cross the blood-brain barrier (Teeling and Perry, 2009), highlighting a putative role for hepatic dysfunction in the pathogenesis of HD.

When taken together, the DEGs were not significantly enriched for functional terms and pathways, but when the genes were

clustered by cell type expression profiles, significant and specific functional enrichments, such as the immune response cluster in the hepatocytes, emerged. There are several of these functional categories that appear only in mature neuronal cells and may point to processes that cause cell loss at greater CAG lengths in neurons compared with other cell types. For example, we observed a cluster of 108 genes, which were significantly enriched for cell cycle and the transition from G1 to S phase. Cell cycle re-entry has been proposed as a mechanism for cell death and neurodegeneration in Alzheimer's disease (Bonda et al., 2010) and HD (Liu et al., 2015). Neurons, although normally quiescent or remaining at G0, can be induced to G1/S phase by oxidative stress (Klein and Ackerman, 2003) or DNA damage (Kruman et al., 2004), subsequently leading to apoptosis (Kruman et al., 2004). Given the well-established relationship between mitochondrial dysregulation and oxidative stress (Lin and Beal, 2006), it is conceivable that mitochondrial dysfunction might lead to neuronal loss through aberrant cell cycle re-entry.

Another observation derived from our transcriptomic profiling is the complex relationship between molecular measurements and the different CAG repeat lengths. Despite the strong inverse linear correlation between CAG size and age at motor onset of HD patients (Andrew et al., 1993), changes in DEGs across different CAG sizes did not necessarily follow a linear relationship. Although we acknowledge that this observation may be influenced by mutant *HTT*-dependent effects on differentiation, this relationship has been described previously in studies aiming to delineate CAG-dependent molecular signatures (Langfelder et al., 2016) and was present in the HDinHD mouse tissues. This CAG-length dependent transcriptional response may be the reason why we did not see high levels of overlap between the pairwise comparisons in the two data sets because the CAG lengths were similar but not exactly the same between the two. The CAG length transcriptional dependencies may also manifest differently in the two systems, e.g., the expression pattern at a given CAG length in a mouse model may not correlate with the expression pattern at the same CAG length in a human model. A separate study modeling the relationship between CAG dependence and gene expression levels in 107 HD lymphoblastoid cells found that CAG size only accounts for 21% of the variance (Lee et al., 2013). Here, we noted the complex relationship between CAG size, transcriptional changes, and functional outcomes of IsoHD NPCs. Indeed, whereas most transcriptional changes did not correlate with CAG size, changes in cellular phenotypes and function largely did. The phenotypic abnormalities showed a correlation with CAG size for many, but not all, measures and were most readily observed in IsoHD 81Q NPCs, followed by IsoHD 65Q NPCs, and finally IsoHD 45Q NPCs. Research that expands the range of CAG sizes studied may uncover more linear and complex CAG-dependent relationships.

In the HDinHD liver transcriptome, the large differences due to gender reduce the variability in the data and can restrict the ability to detect disease-specific expression changes. By using isogenic, single-gender lines, we avoided that confounding factor in the hepatocytes. However, we appreciate that it may be worthwhile to further investigate that gender difference in liver with respect to disease status.

In addition to transcriptomic profiling, untargeted proteomics was performed on IsoHD hESCs and NPCs. The significant differences in protein levels observed across the CAG lengths corroborated the linear relationship seen at the gene expression level but also revealed potential translational differences in expression. A comparison of significantly differentially expressed proteins in hESCs and NPCs identified 13 common proteins. They appear to have a regulatory function, either as transcription factors or transcriptional regulators. One of those 13 proteins, TBL1X, was also differentially expressed at the transcript level, with matching patterns of transcript and protein expression.

Additionally, we observed significant differences at the transcript level, which have then been potentially buffered at the translational protein level and, conversely, a lack of significant transcript differences enhanced at the translational protein level. *INPP5F* was differentially expressed at the mRNA level, but the effect was attenuated at the protein level, whereas *TBLX1* showed a consistent pattern in both NPCs and ESCs. In the opposite direction, no differences at the transcription level of *PJA2* and *EIFS2S* were detectable, but we observed significant differences across the CAG lengths at the protein level between NPCs and ESCs. EIFS2S is the beta subunit of eukaryotic initiation factor 2 (EIF2), which is required for the initiation of translation. The alpha component of EIF2, EIF2S1, has recently been a target of drug repurposing efforts in neurodegeneration (Halliday et al., 2017), highlighting the role of protein translation in disease progression.

Although neuron differentiation from the IsoHD lines did not show mutant HTT aggregates in the nucleus, that finding does not exclude the possibility that the transcriptional dysregulation we describe is the result of mutant HTT effects in the nucleus. Indeed, numerous studies have convincingly shown that soluble forms of mutant HTT interfere with transcription by affecting the transcriptional machinery directly rather than indirectly by sequestering transcription factors into aggregates (Kegel et al., 2002; Kim et al., 2016; Li et al., 2002).

A caveat of the transcriptomic and proteomics data we present is the possibility that the changes observed may not represent purely CAG repeat-related effects but also cellular heterogeneity, as our analyses were performed on bulk cellular populations and not purified cells expressing a defined set of markers. In this regard, future studies may leverage recent advances in single cell-sequencing and analysis technologies to address that possibility directly.

The analyses and data reported here highlight the utility of our IsoHD hESC panel, which can be used to derive new hypotheses into the pathogenesis of mutant *HTT* as well as the neural and systemic effects of the CAG repeat expansion. By developing the IsoHD hESC panel along with a web portal to provide query-based access to the transcriptional and proteomic data from the IsoHD lines and differentiated cell types (http://isohd.pouladilab.org), we hope to facilitate further research into therapeutically relevant molecular factors and key etiological pathways contributing to the pathogenesis of HD.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://doi.org/10.1016/j.celrep.2019.02.008.

ACKNOWLEDGMENTS

We thank Weiping Yu (Biological Resources Centre, A*STAR) and Ray Dunn (Institute of Molecular Biology, A*STAR) for reagents, Huck Hui Ng and Winston Chan (Genome Institute of Singapore, A*STAR) for technical advice, Zahra Khan (TLGM, A*STAR) for technical assistance, and members of the Pouladi lab for helpful comments. The work was partly funded by a Strategic Positioning Fund for Genetic Orphan Diseases (SPF2012/005) and a Joint Council Office Project grant (1431AFG122) from the Agency for Science, Technology and Research (Singapore) to M.A.P. R.M.S. was supported by Core funding from IMCB and IMB Strategic Positioning Fund (SPF, BMRC, A*STAR), a Young Investigator Grant (YIG 2015; BMRC, A*STAR), and an NMRC MS-CETSA platform grant (MOHIAFCAT2/004/2015). A.Z. was supported by the A*STAR Singapore International Graduate Award (SINGA).

AUTHOR CONTRIBUTIONS

M.A.P. conceived and was responsible for the overall supervision of the study. J.O. and M.A.P. designed the research. J.O., X.X., K.H.U., B.S., Y.H., A.Z., D.L., and R.M.S. carried out the experiments. J.O., S.R.L., Y.L.T., and M.A.P. performed analysis of the data. N.P.H. created the web interface. F.G. and E.P. provided intellectual guidance. M.A.P., J.O., and S.R.L. wrote the paper with feedback and input from the other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 7, 2017
Revised: December 17, 2018
Accepted: February 1, 2019
Published: February 26, 2019

REFERENCES

An, M.C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., Melov, S., and Ellerby, L.M. (2012). Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. Cell Stem Cell *11*, 253–263.

Andrew, S.E., Goldberg, Y.P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., Lin, B., Kalchman, M.A., et al. (1993). The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nat. Genet. *4*, 398–403.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aylward, E.H., Sparks, B.F., Field, K.M., Yallapragada, V., Shpritz, B.D., Rosenblatt, A., Brandt, J., Gourley, L.M., Liang, K., Zhou, H., et al. (2004). Onset and rate of striatal atrophy in preclinical Huntington disease. Neurology *63*, 66–72.

Bae, B.-I., Xu, H., Igarashi, S., Fujimuro, M., Agrawal, N., Taya, Y., Hayward, S.D., Moran, T.H., Montell, C., Ross, C.A., et al. (2005). p53 mediates cellular dysfunction and behavioral abnormalities in Huntington's disease. Neuron *47*, 29–41.

Banáth, J.P., Bañuelos, C.A., Klokov, D., MacPhail, S.M., Lansdorp, P.M., and Olive, P.L. (2009). Explanation for excessive DNA single-strand breaks and endogenous repair foci in pluripotent mouse embryonic stem cells. Exp. Cell Res. *315*, 1505–1520.

Björkqvist, M., Wild, E.J., Thiele, J., Silvestroni, A., Andre, R., Lahiri, N., Raibon, E., Lee, R.V., Benn, C.L., Soulet, D., et al. (2008). A novel pathogenic pathway of immune activation detectable before clinical onset in Huntington's disease. J. Exp. Med. *205*, 1869–1877.

Bonda, D.J., Lee, H.-P., Kudo, W., Zhu, X., Smith, M.A., and Lee, H.-G. (2010). Pathological implications of cell cycle re-entry in Alzheimer disease. Expert Rev. Mol. Med. *12*, e19.

Caron, L., Kher, D., Lee, K.L., McKernan, R., Dumevska, B., Hidalgo, A., Li, J., Yang, H., Main, H., Ferri, G., et al. (2016). A human pluripotent stem cell model of facioscapulohumeral muscular dystrophy-affected skeletal muscles. Stem Cells Transl. Med. *5*, 1145–1161.

Carroll, J.B., Bates, G.P., Steffan, J., Saft, C., and Tabrizi, S.J. (2015). Treating the whole body in Huntington's disease. Lancet Neurol. *14*, 1135–1142.

Costa, V., and Scorrano, L. (2012). Shaping the role of mitochondria in the pathogenesis of Huntington's disease. EMBO J. *31*, 1853–1864.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Feyeux, M., Bourgois-Rocha, F., Redfern, A., Giles, P., Lefort, N., Aubert, S., Bonnefond, C., Bugi, A., Ruiz, M., Déglon, N., et al. (2012). Early transcriptional changes linked to naturally occurring Huntington's disease mutations in neural derivatives of human embryonic stem cells. Hum. Mol. Genet. *21*, 3883–3895.

GeM-HD Consortium (2015). Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell *162*, 516–526.

Halliday, M., Radford, H., Zents, K.A.M., Molloy, C., Moreno, J.A., Verity, N.C., Smith, E., Ortori, C.A., Barrett, D.A., Bushell, M., and Mallucci, G.R. (2017). Repurposed drugs targeting eIF2&α;-P-mediated translational repression prevent neurodegeneration in mice. Brain *140*, 1768–1783.

Han, I., You, Y., Kordower, J.H., Brady, S.T., and Morfini, G.A. (2010). Differential vulnerability of neurons in Huntington's disease: the role of cell type-specific features. J. Neurochem. *113*, 1073–1091.

HD iPSC Consortium. (2012). Induced pluripotent stem cells from patients with Huntington's disease show CAG-repeat-expansion-associated phenotypes. Cell Stem Cell *11*, 264–278.

HD iPSC Consortium. (2017). Developmental alterations in Huntington's disease neural cells and pharmacological rescue in cells and mice. Nat. Neurosci. *20*, 648–660.

Jeon, I., Lee, N., Li, J.-Y., Park, I.-H., Park, K.S., Moon, J., Shim, S.H., Choi, C., Chang, D.-J., Kwon, J., et al. (2012). Neuronal properties, in vivo effects, and pathology of a Huntington's disease patient-derived induced pluripotent stem cells. Stem Cells *30*, 2054–2062.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Kegel, K.B., Meloni, A.R., Yi, Y., Kim, Y.J., Doyle, E., Cuiffo, B.G., Sapp, E., Wang, Y., Qin, Z.-H., Chen, J.D., et al. (2002). Huntingtin is present in the nucleus, interacts with the transcriptional corepressor C-terminal binding protein, and represses transcription. J. Biol. Chem. *277*, 7466–7476.

Kennedy, L., Evans, E., Chen, C.-M., Craven, L., Detloff, P.J., Ennis, M., and Shelbourne, P.F. (2003). Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. Hum. Mol. Genet. *12*, 3359–3367.

Kim, Y.E., Hosp, F., Frottin, F., Ge, H., Mann, M., Hayer-Hartl, M., and Hartl, F.U. (2016). Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors. Mol. Cell *63*, 951–964.

Klein, J.A., and Ackerman, S.L. (2003). Oxidative stress, cell cycle, and neurodegeneration. J. Clin. Invest. *111*, 785–793.

Kobold, S., Guhr, A., Kurtz, A., and Löser, P. (2015). Human embryonic and induced pluripotent stem cell research trends: complementation and diversification of the field. Stem Cell Reports *4*, 914–925.

Kruman, I.I., Wersto, R.P., Cardozo-Pelaez, F., Smilenov, L., Chan, S.L., Chrest, F.J., Emokpae, R., Jr., Gorospe, M., and Mattson, M.P. (2004). Cell cycle activation linked to neuronal cell death initiated by DNA damage. Neuron *41*, 549–561.

Kurz, E.U., Douglas, P., and Lees-Miller, S.P. (2004). Doxorubicin activates ATM-dependent phosphorylation of multiple downstream targets in part through the generation of reactive oxygen species. J. Biol. Chem. *279*, 53272–53281.

Langbehn, D.R., Brinkman, R.R., Falush, D., Paulsen, J.S., and Hayden, M.R.; International Huntington's Disease Collaborative Group (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.

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.

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.

Lee, J.-M., Galkina, E.I., Levantovsky, R.M., Fossale, E., Anne Anderson, M., Gillis, T., Srinidhi Mysore, J., Coser, K.R., Shioda, T., Zhang, B., et al. (2013). Dominant effects of the Huntington's disease HTT CAG repeat length are captured in gene-expression data sets by a continuous analysis mathematical modeling strategy. Hum. Mol. Genet. *22*, 3227–3238.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, S.-H., Cheng, A.L., Zhou, H., Lam, S., Rao, M., Li, H., and Li, X.-J. (2002). Interaction of Huntington disease protein with transcriptional activator Sp1. Mol. Cell. Biol. *22*, 1277–1287.

Li, W., Sun, W., Zhang, Y., Wei, W., Ambasudhan, R., Xia, P., Talantova, M., Lin, T., Kim, J., Wang, X., et al. (2011). Rapid induction and long-term self-renewal of primitive neural precursors from human embryonic stem cells by small molecule inhibitors. Proc. Natl. Acad. Sci. USA *108*, 8299–8304.

Lin, M.T., and Beal, M.F. (2006). Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature *443*, 787–795.

Liu, K.-Y., Shyu, Y.-C., Barbaro, B.A., Lin, Y.-T., Chern, Y., Thompson, L.M., James Shen, C.-K., and Marsh, J.L. (2015). Disruption of the nuclear membrane by perinuclear inclusions of mutant huntingtin causes cell-cycle re-entry and striatal cell death in mouse and cell models of Huntington's disease. Hum. Mol. Genet. *24*, 1602–1616.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, X.-H., Mattis, V.B., Wang, N., Al-Ramahi, I., van den Berg, N., Fratantoni, S.A., Waldvogel, H., Greiner, E., Osmand, A., Elzein, K., et al. (2014). Targeting ATM ameliorates mutant Huntingtin toxicity in cell and animal models of Huntington's disease. Sci. Transl. Med. *6*, 268ra178.

Mason, R.P., Casu, M., Butler, N., Breda, C., Campesan, S., Clapp, J., Green, E.W., Dhulkhed, D., Kyriacou, C.P., and Giorgini, F. (2013). Glutathione peroxidase activity is neuroprotective in models of Huntington's disease. Nat. Genet. *45*, 1249–1254.

Matoba, S., Kang, J.-G., Patino, W.D., Wragg, A., Boehm, M., Gavrilova, O., Hurley, P.J., Bunz, F., and Hwang, P.M. (2006). p53 regulates mitochondrial respiration. Science *312*, 1650–1653.

Moss, D.J.H., Pardiñas, A.F., Langbehn, D., Lo, K., Leavitt, B.R., Roos, R., Durr, A., Mead, S., et al.; TRACK-HD investigators; REGISTRY investigators (2017). Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet. Neurol *16*, 701–711.

Nance, M.A., and Myers, R.H. (2001). Juvenile onset Huntington's disease—clinical and research perspectives. Ment. Retard. Dev. Disabil. Res. Rev. *7*, 153–157.

Nekrasov, E.D., Vigont, V.A., Klyushnikov, S.A., Lebedeva, O.S., Vassina, E.M., Bogomazova, A.N., Chestkov, I.V., Semashko, T.A., Kiseleva, E., Suldina, L.A., et al. (2016). Manifestation of Huntington's disease pathology in human induced pluripotent stem cell-derived neurons. Mol. Neurodegener. *11*, 27.

Persichetti, F., Carlee, L., Faber, P.W., McNeil, S.M., Ambrose, C.M., Srinidhi, J., Anderson, M., Barnes, G.T., Gusella, J.F., and MacDonald, M.E. (1996). Differential expression of normal and mutant Huntington's disease gene alleles. Neurobiol. Dis. *3*, 183–190.

Pouladi, M.A., Xie, Y., Skotte, N.H., Ehrnhoefer, D.E., Graham, R.K., Kim, J.E., Bissada, N., Yang, X.W., Paganetti, P., Friedlander, R.M., et al. (2010). Full-length huntingtin levels modulate body weight by influencing insulin-like growth factor 1 expression. Hum. Mol. Genet. *19*, 1528–1538.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. *44* (*W1*), W83-9.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Sathasivam, K., Hobbs, C., Turmaine, M., Mangiarini, L., Mahal, A., Bertaux, F., Wanker, E.E., Doherty, P., Davies, S.W., and Bates, G.P. (1999). Formation of polyglutamine inclusions in non-CNS tissue. Hum. Mol. Genet. *8*, 813–822.

Schiaffino, S., Rossi, A.C., Smerdu, V., Leinwand, L.A., and Reggiani, C. (2015). Developmental myosins: expression patterns and functional significance. Skelet. Muscle *5*, 22.

Seong, I.-S., Ivanova, E., Lee, J.-M., Choo, Y.S., Fossale, E., Anderson, M., Gusella, J.F., Laramie, J.M., Myers, R.H., Lesort, M., and MacDonald, M.E. (2005). HD CAG repeat implicates a dominant property of huntingtin in mitochondrial energy metabolism. Hum. Mol. Genet. *14*, 2871–2880.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318–331.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.

Steffan, J.S., Kazantsev, A., Spasic-Boskovic, O., Greenwald, M., Zhu, Y.Z., Gohler, H., Wanker, E.E., Bates, G.P., Housman, D.E., and Thompson, L.M. (2000). The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. Proc. Natl. Acad. Sci. USA *97*, 6763–6768.

Stüwe, S.H., Goetze, O., Lukas, C., Klotz, P., Hoffmann, R., Banasch, M., Orth, M., Schmidt, W.E., Gold, R., and Saft, C. (2013). Hepatic mitochondrial dysfunction in manifest and premanifest Huntington disease. Neurology *80*, 743–746.

Szkolnicka, D., Farnworth, S.L., Lucendo-Villarin, B., and Hay, D.C. (2014). Deriving functional hepatocytes from pluripotent stem cells. Curr. Protoc. Stem. Cell Biol. *30*, 1G.5.1–1G.5.12.

Teeling, J.L., and Perry, V.H. (2009). Systemic infection and inflammation in acute CNS injury and chronic neurodegeneration: underlying mechanisms. Neuroscience *158*, 1062–1073.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Walker, F.O. (2007). Huntington's disease. Lancet *369*, 218–228.

Wheeler, V.C., Auerbach, W., White, J.K., Srinidhi, J., Auerbach, A., Ryan, A., Duyao, M.P., Vrbanac, 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.

Xu, X., Tay, Y., Sim, B., Yoon, S.-I., Huang, Y., Ooi, J., Utami, K.H., Ziaei, A., Ng, B., Radulescu, C., et al. (2017). Reversal of phenotypic abnormalities by CRISPR/Cas9-mediated gene correction in Huntington disease patient-derived induced pluripotent stem cells. Stem Cell Reports *8*, 619–633.

Zhou, Z., Xu, M.-J., and Gao, B. (2016). Hepatocytes: a key cell type for innate immunity. Cell. Mol. Immunol. *13*, 301–315.

Zielonka, D., Piotrowska, I., Marcinkowski, J.T., and Mielcarek, M. (2014). Skeletal muscle pathology in Huntington's disease. Front. Physiol. *5*, 380.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-Huntingtin, a.a.181-810 (1HU-4C8)	Millipore	Cat# MAB2166; RRID:AB_11213141
Mouse anti-polyglutamine-expansion (5TF1-1C2)	Millipore	Cat# MAB1574; RRID:AB_94263
Mouse anti-mutant HTT	Developmental Studies Hybridoma Bank	Cat# mw1; RRID:AB_528290
Rabbit anti-Human Calnexin	Sigma Aldrich	Cat# C4731; RRID:AB_476845
Rabbit anti-gamma H2A.X (pS139)	Santa Cruz Biotechnology	Cat# sc-101696; RRID:AB_2114997
Rabbit anti-gamma H2A.X (pS139)	Abcam	Cat# ab2893; RRID:AB_303388
Rabbit anti-53BP1 (H-300); *discontinued in 2016*	Santa Cruz Biotechnology	Cat# sc-22760; RRID:AB_2256326
Mouse anti-Ki-67 (Ki-S5)	Millipore	Cat# MAB4190; RRID:AB_95092
Rabbit anti-Oct-3/4 (H-134); *discontinued in 2016*	Santa Cruz Biotechnology	Cat# sc-9081; RRID:AB_2167703
Mouse anti-SSEA-4 (MC-813-70)	STEMCELL Technologies	Cat# 60062; RRID:AB_2721031
Rabbit anti-Pax-6	Covance Research Products Inc	Cat# PRB-278P; RRID:AB_291612
Mouse anti-Nestin (10C2)	Millipore	Cat# MAB5326; RRID:AB_2251134
Rabbit anti-FOXG1	Abcam	Cat# ab18259; RRID:AB_732415
Mouse anti-MAP2 (AP20)	Millipore	Cat# MAB3418; RRID:AB_94856
Rabbit anti-DARPP-32 (H-62)	Santa Cruz Biotechnology	Cat# sc-11365; RRID:AB_639000
Mouse anti-AFP (1G7)	Millipore	Cat# ST1673; RRID:AB_10693988
EM48	Millipore	Cat# MAB5374; RRID:AB_2307353
Rabbit anti-Dystrophin	Abcam	Cat# ab15277; RRID:AB_301813
Mouse anti-Human Myosin Heavy Chain (MF20)	R&D Systems	Cat# MAB4470; RRID:AB_1293549
Bacterial and Virus Strains		
One Shot TOP10 Chemically competent *E. coli*	Thermo Fisher	Cat# C404006
Chemicals, Peptides, and Recombinant Proteins		
Fialuridine (FIAU)	Sigma Aldrich	Cat# SML0632; Cat# 69123-98-4
Tert-Butyl hydroperoxide (TBHP)	Sigma Aldrich	Cat# 416665; Cat# 75-91-2
Doxorubicin (DOX)	Sigma Aldrich	Cat# D1515; Cat# 25316-40-9
YOYO-1 Iodide (491/509) - 1 mM in DMSO	Thermo Fisher	Cat# Y3601
bFGF Recombinant Human Protein	Thermo Fisher	Cat# 13256029
CHIR99021 (GSK-3 inhibitor)	Axon Medchem	Cat# 1386; Cat# 252917-06-9
SB 431542 hydrate	Sigma Aldrich	Cat# S4317; Cat# 301836-41-9
Recombinant Human LIF	STEMCELL Technologies	Cat# 78149
Y-27632 (Rock inhibitor)	AdooQ Biosciences	Cat# A11001
Poly-L-ornithine hydrochloride	Sigma Aldrich	Cat# P2533; Cat# 26982-21-8
Laminin	Thermo Fisher	Cat# 23017015
LDN193189	Stemgent	Cat# 04-0074
XAV939	Stemgent	Cat# 04-0046
Recombinant Human Sonic Hedgehog/Shh (C24II) N terminus	R&D Systems	Cat# 1845-SH
Recombinant Human BDNF	R&D Systems	Cat# 248-BD
Recombinant Human GDNF	R&D Systems	Cat# 212-GD
Recombinant Human FGF basic (146 aa)	R&D Systems	Cat# 233-FB
Adenosine 3′,5′-cyclic monophosphate (cAMP)	Sigma Aldrich	Cat# A9501; Cat# 60-92-4
Ascorbic acid	STEMCELL Technologies	Cat# 72132; Cat# 50-81-7

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant Mouse Wnt-3a	R&D Systems	Cat# 1324-WN/CF
Recombinant Human/Murine/Rat Activin A (E.coli derived)	Peprotech	Cat# 120-14E
Hydrocortisone 21-hemisuccinate sodium salt	Sigma Aldrich	Cat# H4881; Cat# 125-04-2
Recombinant Human HGF (Insect derived)	Peprotech	Cat# 100-39
Recombinant Human Oncostatin M (227 a.a.)	Peprotech	Cat# 300-10
Biocoat Matrigel Matrix HESC-Qualified Mouse	BioLaboratories Pte Ltd	Cat# 354277
Collagen I, Rat Tail	Corning	Cat# 354236
MitoSOX Red Mitochondrial Superoxide Indicator	Thermo Fisher	Cat# M36008
MitoProbe DiIC1(5) Assay Kit - For Flow Cytometry	Thermo Fisher	Cat# M34151
Carboxy-H2DCFDA	Life Technologies	Cat# C400
Lysyl Endopeptidase, Mass Spectrometry Grade	Wako, Japan	Cat# 125-05061
TMT10plex Isobaric Mass Tagging Kit	Thermo Fisher	Cat# 90113
Critical Commercial Assays		
HumanHT-12 v4 Expression BeadChip Kit	Illumina	BD-103-0204
Quant-iT PicoGreen dsDNA Assay Kit	Thermo Fisher	P11496
YOYO-1 Cell Viability Assay	Thermo Fisher	Y3601
MitoSox	Thermo Fisher	M36008
MitoProbe DiIC	Thermo Fisher	M34151
H2DCFDA	Thermo Fisher	D399
Bradford Bio-Rad Protein Assay Dye Reagent Concentrate	Bio-Rad	5000006
TruSeq Stranded mRNA sample preparation Kit	Illumina	20020594
Seahorse XF Cell Mito Stress Test Kit	Agilent	103015-100
Superscript II Reverse Transcription Kit	Thermo Fisher	18064014
Deposited Data		
RNA-Seq data of IsoHD panel	This paper	SRA: SRP107072
Proteomic data of IsoHD panel	This paper	JPOST: JPST000243
Experimental Models: Cell Lines		
Human: Human Embryonic Stem Cells H9 hESC	WiCell	WA09
Human: Human Embryonic Kidney Line HEK293T	ATCC	CRL-2316
Oligonucleotides		
5′ HA donor L 5′ GTCACACTTGGGGTCCTCAG 3′	This study	N/A
5′ HA donor R 5′ AAGAACCCCCGCCCTGGT 3′	This study	N/A
3′ HA donor L 5′ CTGCGTTGTGAAGAGAA 3′	This study	N/A
3′ HA donor R 5′ CAGGCTGGTCTCAAACTCCT 3′	This study	N/A
5′ HR junction_F: 5′ TCTCCGTTTGCCCCTCTACCAC 3′	This study	N/A
5′ HR junction_R: 5′ CTTCTCGATGCGGGTGTTGGTG 3′	This study	N/A
3′ HR junction_F: 5′ ACTTACCGCATTGACAAGCACG 3′	This study	N/A
3′ HR junction_R: 5′ CCACAGTTCCACACCAAAGAGC 3′	This study	N/A
TTAAseq-F: 5′ CCTGTCCTGAATTCACCGAGGG 3′	This study	N/A
TTAAseq-R: 5′ CCTGCAGACCAACTTAGGCTTAGA 3′	This study	N/A
Primers used for CAG sizing, and FAM	Kennedy et al., 2003	N/A
qPCR primers	This study; Table S8	N/A
Recombinant DNA		
USPQ-HTT-TALEN1	This study	Addgene Plasmid # 92243
USPQ-HTT-TALEN2	This study	Addgene Plasmid # 92244
Surrogate reporter for USPQ L and R	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DSPQ-HTT-TALEN1 ZEO	This study (backbone from Ray Dunn, IMB)	Addgene Plasmid # 92245
DSPQ-HTT-TALEN2 BSD	This study (backbone from Ray Dunn, IMB)	Addgene Plasmid # 92246
Surrogate reporter for DSPQ L and R	This study	N/A
PiggyBac Gene Editing HR Targeting Vector (MCS1-5′PB TR-EF1α-GFP-T2A-Puro-T2A-hsvTK-pA-3′ PB TR-MCS2)	System Biosciences	Cat# PBHR100A-1
Donor DNA (pJOP-HTT-HR30Q) with 30 CAG repeats	This study	Addgene Plasmid # 92247
Donor DNA (pJOP-HTT-HR45Q) with 45 CAG repeats	This study	Addgene Plasmid # 92248
Donor DNA (pJOP-HTT-HR65Q) with 65 CAG repeats	This study	Addgene Plasmid # 92249
Donor DNA (pJOP-HTT-HR81Q) with 81 CAG repeats	This study	Addgene Plasmid # 92250
Excision-only PiggyBac transposase plasmid	Systems Biosciences	Cat# PB210PA-1
Software and Algorithms		
Odyssey V3.0	Li-Cor Biosciences	N/A
ImageJ software	Imagej.net	N/A
GenomeStudio 2.0.3	Illumina	N/A
PennCNV	http://wglab.org/software/13-penncnv	N/A
Seahorse Wave software	Agilent	N/A
IncuCyte Zoom system	Essen Bioscience	N/A
StepOne Software V2.3/QuantStudio Real-Time PCR Software	Applied Biosystems	N/A
FastQC	Andrews, 2010	N/A
STAR 2.5.2a	Dobin et al., 2013	N/A
RSEM 1.2.30	Li and Dewey, 2011	N/A
DESeq2	Love et al., 2014	N/A
tximport	Soneson et al., 2015	N/A
Proteome Discoverer 2.1 software	Thermo Fisher	N/A
Sequest HT	Thermo Fisher	N/A
Mascot 2.5.1	Matrix Science	N/A
Combat	Johnson et al., 2007	N/A
Limma with eBayes	Ritchie et al., 2015	N/A
g:Profiler	Reimand et al., 2016	N/A
Prism v7	GraphPad	RRID:SCR_002798
Other		
RNA-Seq data of HDC patient derived iPSC neurons	HD iPSC Consortium, 2017	SRA: PRJNA376704
RNA-Seq data of HDinHD Mouse Liver	Langfelder et al., 2016	GEO: GSE78273
RNA-Seq data of HDinHD Mouse Striatum	Langfelder et al., 2016	GEO: GSE78274
Resource website for the RNA-seq and proteomics results	This paper	http://isohd.pouladilab.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mahmoud A. Pouladi (map@pouladilab.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human ESC lines

The H9 female hESC line was previously reported (Thomson et al., 1998). The isogenic IsoHD hESC lines were generated by TALEN editing the H9 line. The lines were grown and expanded in feeder-free conditions using mTeSR1 medium (STEMCELL Technologies), at 37°C with 5% CO2. Medium was replaced daily.

METHOD DETAILS

Cloning of TALEN plasmids

USPQ and DSPQ TALENs were constructed by a single step of the 4 × 4 Golden Gate assembly protocol using a premade tetramer TALE repeat library. Subsequently, the DSPQ TALENs were cloned into an antibiotic resistance cassette containing plasmids by amplifying the TALENs using primers with BamHI and EcoRI overhangs. The amplicons of DSPQ1 and DSPQ2 were ligated into pCAG-IRES-Zeocin and pCAG-IRES-Blasticidin (gift from Dr. Ray Dunn, Institute of Medical Biology, A*STAR), respectively. The TALENs have been deposited into Addgene (USPQ-HTT-TALEN1, Addgene, no. 92243; USPQ-HTT-TALEN2, Addgene, no. 92244; DSPQ-HTT-TALEN1 ZEO, Addgene, no. 92245; DSPQ-HTT-TALEN2 BSD, Addgene, no. 92246).

Cloning of donor DNA plasmids

The 2.4-kb 3′ homology arm and the 1.7-kb 5′ homology arms, containing human HTT exon 1 with varying (30, 45, 65, and 81) CAG repeat lengths, were cloned into PBHR100A-1 (System Biosciences). The donor DNA constructs have been deposited into Addgene (pJOP-HTT-HR30Q, Addgene, no. 92247; pJOP-HTT-HR45Q, Addgene, no. 92248; pJOP-HTT-HR65Q, Addgene, no. 92249; pJOP-HTT-HR81Q, Addgene, no. 92250).

Cell Culture

H9 cells (WiCell) and the derivative IsoHD hESC panel were maintained on matrigel-coated plates (BD Biosciences) in mTeSR1 (STEMCELL Technologies). hESCs were passaged using dispase and titurated once using a plugged glass Pasteur pipette before seeding at a ratio of 1:6. Media were changed daily. Neural precursor cells were maintained as described by Li et al. (2011).

Generation of IsoHD allelic panel

To transfect the hESCs, 70%–80% confluent cells were treated with 10 μM Y-27632 overnight before dissociation using Accutase (STEMCELL Technologies) at 37°C for 5 min. Dissociated cells were pelleted at 201 x *g rcf.* for 5 min. One million cells were transfected with 5 μg donor DNA and 1 μg of each TALEN plasmid using the NEON Transfection System (Thermo Fisher) at 1400 V, 10 ms, 3 pulses. Transfected cells were immediately transferred into pre-warmed mTesR1 with 10 μM Y-27632. Fresh mTeSR-1 was replaced daily. Targeted cells were selected using 5 μg/ml blasticidin and 2.5 ng/ml zeocin for 48 h from day 1, and 1 μg/ml puromycin for 48 h from day 5. Surviving colonies at two weeks post-transfection were manually picked and expanded for culture and screening via junction PCR.

Screening of transfected clones

Junction PCR was performed using genomic DNA extracted from expanded colonies that survived triple (blasticidin, zeocin, and puromycin) selection. Two sets of primers that amplify junctions at the 5′ and 3′ homology arms were used: 5′ HR junction_F: 5′ TCTCCGTTTGCCCCTCTACCAC 3′; 5′ HR junction_R: 5′ CTTCTCGATGCGGGTGTTGGTG 3′; 3′ HR junction_F: 5′ ACTTACCGCA TTGACAAGCACG 3′; 3′ HR junction_R: 5′ CCACAGTTCCACACCAAAGAGC 3′.

Excision of piggyBac cassette

IsoHD hESCs were grown in syringe-filtered conditioned media [DMEM/F-12 (GIBCO), 1 × NEAA (GIBCO), 1 × Glutamine (GIBCO), 0.1 mM beta-mercaptoethanol (Sigma Aldrich), and 6 ng/ml b-FGF (Thermo Fisher) from overnight culture on mitomycin-C treated CF-1 fibroblasts] for at least one passage. One million cells were transfected with 2 μg of Excision-only PiggyBac transposase (Systems Biosciences) using the NEON Transfection System (Thermo Fisher) at 1400 V, 10 ms, 3 pulses. Targeted cells were selected using 0.2 μM FIAU for six days from day 3. Media was switched from conditioned media to mTeSR1 at day 11. Surviving colonies at two weeks post-transfection were manually picked and expanded for culture and screening via junction PCR.

CNV Analysis

Genomic DNA was extracted cell lysates and quantified by PicoGreen assay (Thermo Fisher). The extracted DNA was whole genome genotyped using the Infinium Global Screening Array-24 v1.0 (GSA) for 642,824 markers, according to the manufacturer's instructions (Illumina, San Diego, CA, USA). The B Allele frequency and Log R Ratio were exported from GenomeStudio 2.0.3 and used with PennCNV to call regions with copy number variations in hg19. The CNV regions were filtered to include those greater than 50kb and with at least 10 SNPs in the region.

Immunoblotting

Cells were lysed with RIPA buffer (Sigma Aldrich) containing 1 mM PMSF, 5 μM z-VAD, 1 mM sodium orthovanadate, and cOmplete Protease Inhibitor Cocktail Tablets (Roche), and protein concentrations were measured using the Bradford Assay (Bio-Rad). Samples were denatured at 70°C for 10 min in 4 × NuPAGE sample buffer and 10 × NuPAGE reducing agent (Thermo Fisher). A total of 30 μg of protein per sample was separated on NuPAGE 4%–12% Bis-Tris gradient gels in NuPAGE MES running buffer at 100 V for 2 h followed by transfer to nitrocellulose membrane at 120 V for 1.5 h at room temperature. The following primary antibodies were used for detection: anti-total HTT (Millipore, MAB2166), anti mutant HTT [1C2] (Millipore, MAB1574), anti mutant HTT [MW1]

(Developmental Studies Hybridoma Bank), anti-calnexin (Sigma, C4731), and anti-gH2A.X (Santa Cruz, sc101696). Alexa Fluor 680 goat anti-mouse and Alexa Fluor 790 goat anti-rabbit (Thermo Fisher) were used as secondary antibodies. Membranes were imaged using Li-Cor Odyssey infrared imaging system and quantified by ImageJ software.

Immunostaining

hESCs and NPCs were fixed with 4% formaldehyde in PBS at room temperature for 30 min. The coverslips were washed three times for 5 min each using wash buffer (PBS with 0.3% Triton X-100 and 2% goat serum) and blocked with blocking buffer (PBS with 0.3% Triton X-100 and 3% goat serum) at room temperature for 1 h. For staining, the following antibodies and concentrations were used: anti-gH2A.X (ab2893; Abcam) 1:200; anti-53BP1 (H-300; Santa Cruz) 1:1000; anti-OCT4(sc-9081; Santa Cruz), anti SSEA4 (MC-813-70; StemCell Technologies); anti-FOXG1 (ab18259; Abcam); anti-NESTIN (MAB5326; Millipore) 1:500; anti-PAX6 (PRB-278P; Covance) 1:250; anti-MAP2 (MAB3418; Millipore) 1:2000; anti-DARPP32 (sc-11365; Santa Cruz) 1:1000; anti-AFP (ST1673; Millipore) 1:100; anti-Ki67 (MAB4190; Millipore) 1:200; anti-Dystrophin (AB15277; Abcam) 1:100; and anti-MF-20 (MAB4470; R&D Systems) 1:50. Antibodies were diluted in wash buffer and incubated at 4°C overnight. The cells were then washed three times for 5 min each using wash buffer and stained with secondary antibodies (AlexaFluor, Thermo Fisher) at a dilution of 1:200. After three 5-min washes, the cells were stained with DAPI and mounted onto microscope slides using ProLong® Gold Antifade Mountant (Thermo Fisher).

Mitochondrial Respiration Analysis

Mitochondrial respiration analysis was carried out on NPCs ($1.5x10^5$ cells per well) using the Seahorse XF96 Extracellular Flux Analyzer following the instructions of Seahorse XF Cell Mito Stress Test Kit (Seahorse Bioscience). Readings were normalized to total protein measured by Bradford assay, and data were analyzed using the Seahorse Wave software. Results were representative of at least three independent experiments (5-well repetition for each group in each experiment).

MitoSox, MitoProbe DilC, and H2DCFDA

One million neural precursor cells dissociated using Accutase were used for each analysis, as indicated by the manufacturer: MitoSox (Thermo Fisher), MitoProbe DilC (Thermo Fisher), and H2DCFDA (Thermo Fisher). The cells were washed and the mean fluorescence value was obtained using FACS Aria (BD Biosciences). Results are representative of at least three independent experiments (at least 3 biological replicates for each CAG length).

YOYO-1 Cell Viability Assay

One million neural precursor cells dissociated using Accutase were seeded onto matrigel-coated well in media supplemented with 5 μM Y-27632 overnight. Cells were first incubated for 2 h with 0.2 mM TBHP, before wash and replacement with fresh growth media containing 0.1 μM YOYO-1 (Thermo Fisher Scientific). Imaging was performed for 40 h using the IncuCyte Zoom System (Essen Bioscience). Results are representative of 3 independent experiments (6 technical replicates for at least 3 biological replicates for each CAG size).

Forebrain neuronal differentiation

hESCs were differentiated into forebrain neurons according to our published protocol (Xu et al., 2017). In brief, hESCs were induced into neural progenitor cells (NPCs) in N2B27 medium supplemented with 100 nM LDN193189 (Stemgent),10 μM SB-431542 (Sigma Aldrich), 2 μM XAV939 (Stemgent) and 200 ng/mL SHH (R&D). For forebrain neuron differentiation, NPCs were cultured in N2B27 medium supplemented with 20 ng/mL BDNF and 20 ng/mL GDNF, 0.5 mM cAMP (Sigma Aldrich) and 0.2 mM ascorbic acid (STEMCELL Technologies). The differentiated cells on Day 11 (NPCs) and Day 45 (neurons) were used for RNA-seq analysis.

Hepatocyte differentiation

hESCs were maintained in mTESR medium on feeder-free Matrigel-coated plates prior to differentiation. hESCs were differentiated into hepatocytes using a previously described protocol (Szkolnicka et al., 2014). Briefly, hESC differentiation to endoderm was driven by plating cells at a density of 200,000 cells/cm^2 in RPMI/B27 supplemented with 50 ng/mL Wnt3a and 100 ng/mL Activin A for three days. To generate hepatoblastic populations, cells were maintained in 20% Knockout Serum Replacement (KOSR)/1% DMSO in Knockout DMEM for five days. Hepatocytes were subsequently specified in HepatoZyme-SFM supplemented with 10 μm Hydrocortisone hemi-succinate, 10 ng/ml HGF, and 20 ng/ml Onco-Statin M for a further 10–12 days.

Muscle differentiation

The Skeletal Muscle Platform (Genea Biocells) was used to generate myotubes. In brief, hESCs were dissociated using 0.05% Trypsin-EDTA for 5 min at 37°C, before seeding in Skeletal Muscle Induction Media onto a collagen I coated 6-well (Corning). Cells were maintained in the Skeletal Muscle Induction Media for 10 days. After 10 days, cells were dissociated using 0.05% Trypsin-EDTA for 5 min at 37°C, before seeding in Skeletal Myoblast Medium onto a collagen I-coated 6-well plate. Media change was performed every 2–3 days over 7 days before switching to Myotube Medium for 10 days before harvest or characterization.

RNA Isolation and Quantitative PCR

RNA from cells was extracted using an RNeasy mini kit (QIAGEN) according to the manufacturer's instructions. To generate cDNA, the Superscript® II Reverse Transcription Kit (Thermo Fisher) was used. A total of 20 μl of cDNA was generated per 1 μg of RNA. To perform quantitative PCR (qPCR), cDNA was diluted ten-fold and 2 μl was used per qPCR reaction. To complete the reaction volume, 0.67 mM primers and SYBR® Select Master Mix (Thermo Fisher) were added. The primers used are listed in Table S8.

Fragment Sizing

To determine the length of the CAG repeats, PCR of the CAG repeats was done using KOD Xtreme Hot Start DNA Polymerase (EMD Millipore) and the following primers described in (Kennedy et al., 2003): F1: 5′ CCCATTCATTGCCCCGGTGCTG 3′ and R1: 5′ TGGGTTGCTGGGTCACTCTGTC 3′, in a 50ul reaction. 5ul of the PCR mix was used to visualize the amplicons on an agarose gel. The remaining PCR mixture was use for fragment sizing in HiDi formamide (Applied Biosystems) and run on Applied Biosystems 3130XL, using LIZ600 (Thermo Fisher) as the size standard.

TTAA sequencing

To determine the sequence around the piggyBac excision site, PCR of the region was performed using KOD Xtreme Hot Start DNA Polymerase (EMD Millipore) and the following primers: TTAAseq-F: 5′ CCTGTCCTGAATTCACCGAGGG 3′ and TTAAseq-R: 5′ CCTGCAGACCAACTTAGGCTTAGA 3′, in a 50-μl reaction. Five microliters of the PCR mix was used to visualize the amplicons on an agarose gel. The remaining PCR mixture was purified using a PCR purification kit (QIAGEN) and sequenced using capillary sequencing.

IsoHD RNA-seq

RNA was extracted from cells using an RNeasy mini kit (QIAGEN) according to the manufacturer's instructions, with slight modifications. In brief, cells grown on tissue culture dishes were treated with 1 mL TRIzol, scraped, and collected into an Eppendorf tube. Chloroform (0.2 ml) was added and samples were vortexed vigorously for 15 s and incubated at room temperature for 2 min. Samples were centrifuged at 12,000 × *g* for 15 min at 4°C to allow for separation of the upper aqueous phase. The upper aqueous phase (350 μl) was transferred to a new tube before one volume of 70% ethanol was added and mixed well. The mixture was transferred to an RNeasy spin column and RNA was isolated following the manufacturer's instructions.

The RNA were further purified using ethanol before quality assessments for RNA-seq. Subsequent library preparation (TruSeq® Stranded mRNA sample preparation kit; Illumina) and paired end 2 × 100-bp sequencing using HiSeq were performed by Exiqon, Denmark. Fastq files from both runs were merged and data quality assessed using FastQC (Andrews, 2010). The reads were aligned to the hg38 genome using STAR 2.5.2a (Dobin et al., 2013), and quantified using RSEM 1.2.30 (Li and Dewey, 2011). Differential expression was calculated with DESeq2 (Love et al., 2014) using tximport (Soneson et al., 2015). Mitochondrial and ribosomal genes were removed from further analysis to avoid confounding technical factors (e.g., differences in poly-A capture/Ribo-Zero rRNA removal efficiencies and the number of mitochondria between samples). Significant differential expression across the CAG mutation lengths within each cell type was calculated using the likelihood ratio test (LRT), and differential expression between the 45Q, 65Q, 81Q, and 30Q was calculated using the Wald test; RIN was included as a covariate in both analyses. Differences in gene expression were significant at a BH-corrected 10% false-discovery rate (FDR) level. Functional enrichment was calculated using gprofileR at a 5% FDR, excluding Inferred from Electronic Annotation (IEA) terms (Reimand et al., 2016) with the background set to all expressed genes within a given cell type.

For genes significantly differentially expressed across all CAG lengths, fold changes were calculated for 45Q, 65Q, and 81Q with respect to 30Q. The fold changes were scaled and clustered using hierarchical clustering with Euclidean distance. The hierarchical tree was cut using a height of 0.75, resulting in 14–18 clusters of genes per cell type. Functional enrichment was calculated using gprofileR at a 5% FDR, excluding IEA terms (Reimand et al., 2016) with the background set to all significantly DEGs within each cell type. A summary of the clusters was visualized by scaling and averaging the transcripts per million (TPM) of the genes within each cluster.

Comparison with HDC patient iPSC neurons

HD (n = 4) and control (n = 3) data from patient derived iPS neurons were downloaded from SRA (PRJNA376704) and converted to fastq files using *fastq-dump*. Data quality was assessed using FastQC. One sample, HD1, failed quality control measures on GC content and enrichment for repeated sequences, indicating a higher ribosomal content. The reads were aligned to the hg38 genome using STAR 2.5.2a and quantified using RSEM 1.2.30. Sample HD1 had a lower number of reads aligning (~7 Million) as compared to the other HD and control samples (> 10 Million) and was removed from further analyses. Mitochondrial and ribosomal genes were removed from further analysis to avoid confounding technical factors (e.g., differences in poly-A capture/Ribo-Zero rRNA removal efficiencies and the number of mitochondria between samples). Differential expression was calculated with DESeq2 using tximport. Significant differential expression was calculated between control and HD using the Wald test with significance set at a BH-corrected 10% FDR level. The significantly DEGs were hierarchically clustered with Euclidean distance.

Comparison with HDinHD Mouse Liver & Striatum

RNA-seq counts for liver and striatum at 6 months were downloaded from GEO (GSE78273 and GSE78274 - 50Q expansion series). Mitochondrial and ribosomal genes were removed from further analysis to avoid confounding technical factors (e.g., differences in poly-A capture/Ribo-Zero rRNA removal efficiencies and the number of mitochondria between samples). Differential expression was calculated with DESeq2. Within each tissue type, significant differential expression was calculated across CAG lengths (20Q, 50Q, 92Q) using the likelihood ratio test (LRT) and pairwise using the Wald test; gender was included as a covariate in each test. Differences in gene expression were significant at a BH-corrected 10% FDR level. The significantly DEGs were hierarchically clustered with Euclidean distance. Human orthologs were downloaded from Ensembl.

Proteomics

Proteins extracted from the IsoHD hESC lines, and derived NPCs, were digested in solution. Briefly, cell pellets were re-suspended in 8 M urea/ 50 mM Tris-HCl pH 8.5 and subsequently sonicated followed by reduction in 20 mM TCEP (20 min, 25°C) and alkylation with 55 mM CAA (20 min, 25°C, protected from light). Two-step digestion after 100 mM TEAB dilution (urea < 1 M) was performed using LysC enzyme (Wako, Japan) for 4 h, followed by trypsin (Promega, USA) treatment for 18 h at 25°C (1:100, enzyme: protein ratio). After acidification with 1% TFA, peptides were centrifuged, desalted using C-18 SepPak columns (Waters), and vacuum dried. Equal amount of peptides were taken for TMT10 isobaric tag (Thermo Fisher) labeling. Following labeling, samples were combined, desalted, and vacuum dried. Samples were subsequently re-suspended in 10 mM ammonia/ 5% acetonitrile and separated on C18 high pH reverse phase chromatography using AKTA Micro (GE) and Zorbax C-18 Extend (Agilent) columns (10 mM ammonia formate and 90% acetonitrile in 10 mM ammonia formate pH 10.5). Combined fractions were vacuum dried and subsequently analyzed on an Easy nLC1000 (Thermo Fisher) chromatography system coupled with Orbitrap Fusion (Thermo Fisher). Each sample was separated on a 120-min gradient (0.1% formic acid in water and 99.9% acetonitrile with 0.1% formic acid) using a 50 cm x 75 μm ID Easy-Spray column (C-18, 2-μm particles, Thermo Fisher). Data dependent mode was used in a speed mode −3 s cycle using Orbitrap analyzer with ion targets and resolution (OT-MS 4 × E5 ions, resolution 60K, OT-MS/MS 6E4 ions, resolution 60k). Peak lists were generated with Proteome Discoverer 2.1 software (Thermo Fisher), and searches were performed with Sequest HT (Thermo Fisher) and Mascot 2.5.1 (Matrix Science) against a concatenated forward/decoy Human-HHV4 Uniprot database (88559 entries) with the following parameters: precursor mass tolerance (MS) 30 ppm, OT-MS/MS 0.06 Da, 3 missed cleavages. Static modifications: Carboamidomethyl (C), TMT10plex. Variable modifications: Oxidation (M), Deamidated (NQ), Phospho (STY), Acetyl N-terminal protein. Forward/decoy searches were used for false discovery rate estimation (FDR 1%).

MS2 intensities were normalized for total reporter abundance; proteins that were not identified in one or more batches were removed from further analyses. Batch correction was performed with Combat (Johnson et al., 2007). Differential expression across CAG lengths was performed using limma with eBayes (Ritchie et al., 2015). Significance was set to a BH-corrected 10% FDR level. Functional enrichment was calculated using gprofileR at a 5% FDR, excluding IEA terms (Reimand et al., 2016) with the background set to all identified proteins within the cell types.

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless otherwise stated, represent the mean ± standard error of the mean (SEM). Comparisons between groups were assessed using a one-way ANOVA with Fisher's LSD post hoc analysis. Where indicated, pairwise comparisons were assessed with Student's t test. GraphPad Prism v7 software was used to analyze data for statistical significance. Differences were considered statistically significant when $p < 0.05$.

DATA AVAILABILITY

The RNA-seq and proteomics data reported in this paper have been deposited in the SRA under accession number SRA: SRP107072 and in JPOST under accession number JPST000243, respectively. The RNA-seq data of HDC patient derived iPSC neurons is available in SRA under accession number SRA: PRJNA376704. The RNA-seq data of HDinHD mouse tissue is available in GEO under accession number GEO: GSE78273 for liver and GEO: GSE78274 for striatum.

ADDITIONAL RESOURCES

We have developed a web portal to provide query-based access to the transcriptional and proteomic data reported in this paper (http://isohd.pouladilab.org).