Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease


Huntington disease (HD) is a devastating, late-onset, inherited neurodegenerative disorder that manifests with personality changes, movement disorders, and cognitive decline. It is caused by a CAG repeat expansion in exon 1 of the HTT gene that translates to a polyglutamine tract in the huntingtin protein (HTT). The formation of HTT fragments has been implicated as an essential step in the molecular pathogenesis of HD and several proteases that cleave HTT have been identified. However, the importance of smaller N-terminal fragments has been highlighted by their presence in HD postmortem brains and by the fact that nuclear inclusions only detected by antibodies to the N terminus of HTT. Despite an intense research effort, the precise length of these fragments and the mechanism by which they are generated remains unknown. Here we show that CAG repeat length-dependent aberrant splicing of exon 1 HTT results in a short polyadenylated mRNA that is translated into an exon 1 HTT protein. Given that mutant exon 1 HTT proteins have consistently been shown to be highly pathogenic in HD mouse models, the aberrant splicing of HTT mRNA provides a mechanistic basis for the molecular pathogenesis of HD. RNA-targeted therapeutic strategies designed to lower the levels of HTT are under development. Many of these approaches would not prevent the production of exon 1 HTT and should be reviewed in light of our findings.

Results

Aberrant Splicing of Mouse Htt Exon 1 to Exon 2 Results in a Small Polyadenylated mRNA. The HdhQ150 knock-in mouse model was generated by the insertion of ∼150 CAGs into exon 1 of the mouse Htt gene (14). To determine whether exon 1 of Htt mRNA had spliced to exon 2, we used a series of RT-PCR assays (Fig. 1A) on cDNA prepared from the brains of homozygous (HdhQ150/Q150), heterozygous (HdhQ150/G150) and WT animals at 2 mo of age (Fig. 1B). Nonquantitative end-point RT-PCR products were comparable between all three genotypes for assays that amplified exon 2 and spanned the exon 1–exon 2 junction. In contrast, the levels of the RT-PCR products obtained for the exon 1–intron 1 boundary and for intron 1 sequences upstream of ∼1.2 kb were more intense in cDNA prepared from the brains of HdhQ150/Q150 compared with HdhQ150/G150 animals, which were in turn more intense than those obtained from their WT littermates (Fig. 1B). Examination of the genomic sequence for Htt intron 1 identified a cryptic polyadenylation signal at position 677–685 bp and 3′RACE showed the presence of a polyadenylated short mRNA in HdhQ150/Q150 and HdhQ150/G150 but not WT brains (Fig. 1C). This unspliced exon 1–intron 1 transcript was present throughout the brain and in a range of peripheral tissues (Figs. S1 and S2). To independently verify the presence of the exon 1–intron 1 transcript, we used total RNA sequencing (RNA-Seq) on RNA prepared from the cortex of HdhQ150/Q150 and WT mice. Considerable read density mapping to the first 1.2 kb of intron 1 were observed in HdhQ150/Q150, but not in WT littermate samples (Fig. 1D and Fig. S3). To approximate the percentage of nonspliced transcripts, we compared the 3′UTR read densities by the Mixture-of-Isosforms (MISO) software and found a 20% reduction in the full-length mRNA produced in the HdhQ150/Q150 samples (Fig. S3). Finally, we failed to detect any evidence for higher intronic reads in homozygotes compared with WT at any other position along the mouse Htt gene (Fig. S3C).

Aberrant Splicing Occurs in All HD Knock-In Mouse Models and Is CAG Repeat Length–Dependent. As stated previously, individuals with mutations of (CAG)n and above will develop HD within a normal lifespan, whereas repeats above (CAG)70 will invariably


The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

1K.S. and A.N. contributed equally to this work.

2To whom correspondence may be addressed. E-mail: gillian.bates@kdl.ac.uk or dhousman@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221891110/-/DCSupplemental.
cause childhood onset (1). Therefore, we used a series of knock-in mouse lines: HdhQ20, Q50, Q80, Q100, Q150, and zQ175 (Figs. 1A and 2A) to determine whether aberrant splicing occurred in the context of a wide range of CAG repeats (Table S1 shows CAG repeats lengths). We analyzed RNA at 2 mo of age for each of the knock-in lines and were able to detect the same 3’RACE product in all cases except for HdhQ20 (Fig. 2B). Quantitative RT-PCR (Fig. 2C) indicated that the spliced full-length Htt mRNA levels, as determined by amplifying the exon 1–exon 2 junction and exon 2 sequences, were largely comparable between genotypes for Q20, Q50, Q80, and Q100. However, the full-length Htt transcripts were underrepresented in the Q150 and zQ175 mice (Fig. 2C and D), consistent with the RNA-Seq data (Fig. S3B) and a previous report (15). It was not possible to normalize the levels of any intronic sequences to WT because in these mice, intron 1 sequences in processed mRNA have been removed by splicing and any signal represents noise (Fig. 2E). However, quantitative RT-PCR showed that transcripts containing early intron 1 sequences were highly increased in a gene–dose-dependent manner for all lines except HdhQ20 (Fig. 2E). The level of these transcripts also increased in a polynucleotide length–dependent manner when comparing the Q50, Q100, and Q150 lines (mouse Htt sequences only) or the Q80 and Q175 lines (human-mouse chimeric Htt) (Fig. 2A and E).

### Aberrant Splicing Occurs in the Context of Human Htt

To extrapolate these findings to human Htt, we performed in silico analysis and identified nine predicted polyadenylation signals in intron 1 (Fig. S4). 3’RACE on brain RNA from 2-mo-old YAC128 (16) and BACHD (17) mice, both of which are transgenic for human Htt, showed that the signal with the highest predictive score (∼7.3 kb) results in a polyadenylated transcript (Fig. 2F). A second polyadenylated transcript, resulting from a signal at ∼2.7 kb, was also found in the YAC128 mice (Fig. S5).

Next we sought to determine whether the aberrantly spliced mRNA was translated. We performed RT-PCR on RNA isolated from polysome gradients prepared from homozygous zQ175 and WT brains at 2 mo of age (Fig. 3A and Fig. S6). RT-PCR with primers to Hsf1 and Atp5b demonstrated that the polysomes isolated from both genotypes were intact (Fig. S6 A–C). RNA isolated from the polysome gradients from WT mice showed the expected profiles: Htt exonic sequences, but not early intron 1 sequences were associated with polysomes (fractions 12–18) (Fig. 3A and Fig. S6D). In contrast, early intron 1 sequences from zQ175 mice were present in the polysome fractions (Fig. 3A), which is consistent with the nonspliced RNA being translated.

### Splicing Factor SRSF6 Binds to Expanded CAG Repeats in HD Transcripts

To investigate the underlying mechanism, we used bioinformatics to predict regulatory motifs in exon 1 of Htt (Fig. 4A) and mapped the binding site of SRSF6 to a CAG or CAGCAA repeat (Fig. 4B). RNA immunoprecipitation with an antibody to SRSF6 captured Htt 5’UTR and early intron 1 sequences from homozygous zQ175 mice with a much higher efficiency than those from WT brain lysates (Fig. 4C). Exon 2 transcripts were not immunoprecipitated from either zQ175 or WT lysates (Fig. 4C), which is consistent with SRSF6 binding to the expanded CAG repeat in the zQ175 mice and inhibiting the splicing of exon 1 to exon 2.

### The Aberrantly Spliced Transcript Is Present in HD Patient Tissues

To determine whether the aberrant splicing also occurs in humans, we performed quantitative RT-PCR (indicated that the spliced full-length Htt transcript is present in HD patient tissues (Fig. S6 A–C).
we performed RT-PCR on RNA from two juvenile HD and two control fibroblast lines, peripheral blood mononuclear cells from HD and control subjects, postmortem sensory motor cortex from two HD and two control subjects, and cortical tissue from two individuals with juvenile HD (Table S2). In the fibroblast lines, the signals were relatively comparable for exon 1, the exon 1–exon 2 junction, and exon 2 transcripts. In contrast, the early intron 1 transcripts were elevated in the HD fibroblast compared with the control lines (Fig. 4D). However, the pattern was more difficult to interpret in the postmortem brain samples, probably reflecting differences in the extent to which the RNA had degraded in these tissues. We performed 3′ RACE for the same cryptic polyadenylation signal as used in the YAC128 and BACHD mice. The polyadenylated transcript was clearly apparent in the two HD broblast lines, peripheral blood mononuclear cells from two HD and two control subjects, and cortical tissue from two individuals with juvenile HD (Table S2). In the fibroblast lines, the signals were relatively comparable for exon 1, the exon 1–exon 2 junction, and exon 2 transcripts. In contrast, the early intron 1 transcripts were elevated in the HD fibroblast compared with the control lines (Fig. 4D). However, the pattern was more difficult to interpret in the postmortem brain samples, probably reflecting differences in the extent to which the RNA had degraded in these tissues. We performed 3′ RACE for the same cryptic polyadenylation signal as used in the YAC128 and BACHD mice. The polyadenylated transcript was clearly apparent in the two fibroblast lines as well as in the sensory motor cortex of an HD patient with 42 CAGs and the cortex of a juvenile HD individual with 72 CAGs (Fig. 4E). We were only able to detect a 3′ RACE product in one of eight peripheral blood mononuclear cells samples, suggesting that the missplicing events are very rare in these cells.

**Discussion**

We have identified a small exon 1–intron 1 polyadenylated mRNA transcript in the brains of all HD mouse models expressing mutant Htt (mouse) or HTT (human). Furthermore, we have shown that the same transcript is also present in fibroblast lines derived from HD patients and in postmortem HD brains. We have shown that the SRSF6 splicing factor binds to the 5′ end of the Htt gene with an expanded CAG repeat, consistent with its known recognition motif (18). SRSF6 regulates splicing and facilitates translation of partially spliced transcripts (19, 20). SR proteins might also sequester U1 small nuclear ribonucleoprotein complexes (21), a phenomenon that promotes polyadenylation from cryptic polyA signals within introns (22). Therefore, an increased association of SRSF6 with expanded CAG repeats could account for the CAG repeat length–dependent production of the exon 1–intron 1 transcript. Translation of this transcript produces an exon 1 HTT protein.

The pathological consequences of the expression of this aberrantly spliced product in mice have been demonstrated in multiple experiments. The R6/2 mouse lines are transgenic for a genomic fragment that spans the 5′ end of the Htt gene, exon 1, and a portion of intron 1 (13). Therefore, the R6 lines can be considered to be models of aberrant splicing in HD and demonstrate that an exon 1 HTT protein is highly pathogenic, a result that was recapitulated in a set of independent experiments (23). Surprisingly, the late-stage phenotypes of R6/2 mice closely
resemble those present in the HdhQ150 knock-in mice (24–27), the main difference between these two models being the age of phenotype onset and rate of disease progression. This suggests that the pathogenic process in HdhQ150 mice could be driven by the same exon 1 HTT fragment differing only in its abundance and accumulation over time.

Our discovery that exon 1 HTT is generated by aberrant splicing will provide the opportunity to test the extent to which an exon 1 protein contributes to disease pathogenesis in the knock-in lines and the extent to which full-length HTT and/or other N-terminal HTT fragments, generated by proteolytic cleavage, are also pathogenic. RNA-targeted therapeutic approaches designed to

Fig. 3. The aberrantly spliced Htt transcript is translated and produces an exon 1 HTT protein. (A) Polysome gradients showing the relative distribution of early Htt intron 1 (135f/200r) transcripts (n = 2). Data are mean ± SEM. (B) Schematic shows the position of the HTT antibody epitopes. (C–F) HTT proteins were immunoprecipitated with 3B5H10-coupled magnetic beads from WT and (C) zQ175, (D) HdhQ100, (E) HdhQ80, and (F) YAC128 brain lysates, and Western immunoblots were immunoprobed with S830, MW8, and 1H6 antibodies. Dotted lines indicate the gel migration of the exon 1 HTT proteins, which are retarded by the polyglutamine tract and do not migrate as would be predicted by their molecular weight. To resolve exon HTT proteins with different polyQ tracts, the gel depicted in (C) was run for a longer time than those in (D–F). On the HdhQ80 blots, exon 1 HTT comigrates with a nonspecific band (*). 3B5 = 3B5H10.

Fig. 4. The splicing factor SRSF6 binds to expanded CAG repeats in Htt transcripts and the aberrantly spliced transcript is present in HD patient tissues. (A) RegRNA predicts a cluster of R0987 regulatory motifs (CTGN) in the expanded CAG repeat of Htt. Using ESEfinder 3.0, this motif was mapped to the binding site of SRSF6. (B) The RNA recognition motif of SRSF6 is YRCRKM, closely resembling a CAG or CAGCAA repeat. Nucleotide abbreviations: Y, T or C; R, A or G; K, G or T; M, A or C. (C) RNA-IP of SRSF6 from zQ175 brain resulted in the coprecipitation of higher levels of early intronic (135f-200r) and exon 1 (5′UTR 1f-5′UTR 1r) transcripts compared with WT. In contrast, the coprecipitated levels of zQ175 and WT exon 2 transcripts were not significantly different. n = 6/genotype. Horizontal bar, IgG control immunoprecipitation. Data are mean ± SEM; **P < 0.01. (D) RT-PCR of human samples. Human fibroblast lines can be found in Table S2. (E) 3′ RACE of human samples. ⦿, the expected RACE product size of about 260 bp; Ctr, control subject; Ctx, cortex; HD, adult onset HD; ju, juvenile-onset HD; M, low-molecular-weight markers (New England Biolabs); SMC, sensory motor cortex; tg, transgenic; W, water.
lower the levels of HTT through the use of antisense oligonucleotides, RNAi, or small hairpin RNAs are under development (28). Many of these approaches would not prevent the production of exon 1 HTT and should be reviewed in the light of our findings.

**Materials and Methods**

**Mouse Maintenance and Breeding.** The HdhQ50 and HdhQ100 lines were generated by selective breeding for alterations in germ-line repeat size starting with a C7BL/6 congenic of the HdhQ150 lines (14). The Hdh<sup>Q20</sup>, Hdh<sup>Q200</sup> (29, 30), and zQ175 (31) knock-in mice were supplied from CHDI colonies maintained at The Jackson Laboratory. The HdhQ20, HdhQ50, HdhQ80, HdhQ100, and zQ175 lines were maintained by backcrossing to C7BL/6J (Charles River). All experimental procedures were approved by the King's College London Ethical Review Committee and performed in accordance with United Kingdom Home Office regulations.

**RNA Sequencing.** Samples were prepared using a modified, strand-specific version of the Illumina Tru-Seq protocol. Final PCR amplification was performed with KAPA HiFi polymerase and GC buffer (Kapa Biosystems). The paired-end, strand-specific cDNA libraries were multiplexed onto the Illumina HiSeq (40-bp reads). Real data were mapped to the mm9 build with the Bowtie alignment program using the “best” setting. For quality control statistics of the RNA-Seq libraries, see Table S4. Differential expression of the Htt gene was analyzed with the R DEseq package. Gene expression was calculated as reads per kilobase of exon per million mapped reads (RPKMx). Splicing was analyzed using the Python/C version of MISO.

**Polyosome Gradients.** A 10–40% (wt/vol) sucrose gradient was prepared in 50 mM Tris Cl (pH 6.6), 140 mM NaCl, 12 mM magnesium chloride, 200 μg/mL cycloheximide, and 1 mM DTT. Mouse brain tissue was lysed in freshly prepared polysome buffer [10 mM Tris Cl (pH 7.4), 140 mM NaCl, 12 mM magnesium chloride, 1% (v/vol) Triton X-100, 1 mM DTT, 200 μg/mL cycloheximide, 0.5 U/μL RNaseA, and 10 mM ribonucleoside vanadyl complex]. A volume corresponding to 250 μg absorbance at 260 nm was layered on the 40–100% sucrose gradients. The gradients were centrifuged at 39,000 rpm at 4°C for 1 h 40 min in a SW41-Ti swing out rotor. Fractions were collected and RNA was extracted.

**SR56 RNA Coimmunoprecipitation.** Mouse brain tissue was lysed in freshly prepared Triton buffer [50 mM HEPES/NaOH pH 7.6, 140 mM NaCl, 7 mM magnesium chloride, 3 mM calcium chloride, 5 mM potassium chloride, 1% (v/vol) Triton X-100, 1 mM phenylmethylsulfonyl fluoride, 0.5 U/μL RNaseA and complete protease and phosphatase inhibitor mixture]. Two milligrams of total protein was immunoprecipitated for 5 h on a rotating wheel at 4°C with 9 μL of Dynabeads protein G and 3 μg of anti-SR56 antibody (LSB5712; LifeSpan Biosciences). The beads were washed four times with 0.5 mL Triton buffer followed by RNA extraction.

**Acknowledgments.** We thank Jenny Morton for providing YAC128 mice, Yvon Trottier for the HTs antibody, the Massachusetts Institute of Technology BioMicro Center for RNA-Seq preparation and sequencing, and the Neurological Foundation of New Zealand for Huntington disease post-mortem brain tissue. This work was supported by Grant G0801314 from the Medical Research Council and a grant from the CHDI Foundation.