Journal of Huntington's Disease 4 (2015) 161–171
DOI 10.3233/JHD-150151
IOS Press

Research Report

Huntingtin Exists as Multiple Splice Forms in Human Brain

Matthew Mort[a,1], Francesca A. Carlisle[b], Adrian J. Waite[b], Lyn Elliston[b], Nicholas D. Allen[c], Lesley Jones[b] and Alis C. Hughes[b,1,*]
[a]*Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff, UK*
[b]*Institute of Psychological Medicine and Clinical Neuroscience, MRC Centre for Neuropsychiatric Genetics and Genomics, Hadyn Ellis Building, Cardiff University, UK*
[c]*Cardiff School of Biosciences, Sir Martin Evans Building, Museum Avenue, Cardiff, UK*

Abstract.
Background: A CAG repeat expansion in *HTT* has been known to cause Huntington's disease for over 20 years. The genomic sequence of the 67 exon *HTT* is clear but few reports have detailed alternative splicing or alternative transcripts. Most eukaryotic genes with multiple exons show alternative splicing that increases the diversity of the transcriptome and proteome: it would be surprising if a gene with 67 known exons in its two major transcripts did not present some alternative transcripts.
Objective: To investigate the presence of alternative transcripts directly in human *HTT*.
Methods: An overlapping RT-PCR based approach was used to determine novel *HTT* splice variants in human brain from HD patients and controls and 3D protein homology modelling employed to investigate their significance on the function of the HTT protein.
Results: Here we show multiple previously unreported novel transcripts of *HTT*. Of the 22 splice variants found, eight were in-frame with the potential to encode novel HTT protein isoforms. Two splice variants were selected for further study; *HTTΔex4,5,6* which results in the skipping of exons 4, 5 and 6 and *HTTex41b* which includes a novel exon created via partial retention of intron 41. 3D protein homology modelling showed that both splice variants are of potential functional significance leading to the loss of a karyopherin nuclear localisation signal and alterations to sites of posttranslational modification.
Conclusions: The identification of novel *HTT* transcripts has implications for HTT protein isoform expression and function. Understanding the functional significance of *HTT* alternative splicing would be critical to guide the design of potential therapeutics in HD that aim to reduce the toxic *HTT* transcript or protein product including RNA silencing and correction of mis-splicing in disease.

Keywords: Huntington's disease, alternative splicing, RNA species, protein isoforms, structural modelling

INTRODUCTION

Huntington's disease (HD) is an hereditary neurodegenerative disorder presenting in midlife with motor, cognitive and behavioural symptoms [1]. The Huntingtin (*HTT*) gene discovered over 20 years ago [2] contains the CAG repeat that when expanded causes this disorder. Studies of HD pathogenesis have largely focussed on the mutant HTT protein, or N-terminal fragments that contain the expanded polyQ, and their toxic role in cellular processes. However, the direct effect of the mutant *HTT* transcript has also been considered more recently [3]. RNA toxicity mechanisms are well described for some triplet-repeat diseases [4]. In CAG repeat disorders, the expanded CAG tract can lie either within the untranslated region of a gene or

[1]These authors contributed equally to this work.

*Correspondence to: Alis C. Hughes, Institute of Psychological Medicine and Clinical Neuroscience, MRC Centre for Neuropsychiatric Genetics and Genomics, Hadyn Ellis Building, Cardiff University, CF24 4HQ. Tel.: +44 2920 688 472; Fax: +44 2920 687 068; E-mail: HughesAC@cardiff.ac.uk.

ISSN 1879-6397/15/$35.00

within the coding region; the latter case resulting in a polyQ disease [5]. In both cases the expanded CAG tract can alter the conformation of RNA leading to the formation of hairpin structures [4]. It is possible that these can mediate pathogenic effects through RNA-toxicity mechanisms including sequestration of proteins such as muscleblind-like 1 (MBNL1) and alterations to alternative splicing, effects on gene expression, mislocalisation of the transcript and repeat-associated non-ATG translation (RAN) [5]. In HD, RNA foci which sequester MBNL1 have been shown in human HD fibroblasts [6]. This could mediate effects on the splicing of transcripts in HD, including that of *HTT* itself.

Splicing is the process whereby introns are removed from pre-mRNA to yield a protein-encoding RNA transcript [7] and alternative splicing is a mechanism which expands the transcriptional, proteomal and phenotypic diversity of the cell and organism [8]. Most human genes are alternatively spliced [9]. Until recently, little was known about the alternative transcripts of the *HTT* gene and whether its encoded protein HTT exists simply as a single, or multiple protein isoforms. Sathasivam et al, 2013 reported a short *HTT* transcript encoding a pathogenic truncated HTT protein [10] in HD and our screen of the murine *Htt* gene revealed two novel splice variants either lacking exon 29 or partial retention of intron 28 [11]. Any alternative transcript of *HTT* has the potential to encode a novel HTT protein isoform, the function of which could differ from that of full length HTT. Here, we have extended our previous study with HD mouse models and identified novel human alternatively spliced *HTT* transcripts in brain material. The human *HTT* gene in HD and control brain regions was screened using an RT-PCR approach of overlapping amplicons spanning the entire gene. We report the existence of multiple *HTT* splice variants. Further *in silico* investigation of two splice forms not predicted to activate nonsense mediated decay (NMD) and thus having the potential to be translated, demonstrated potential functional significance for these proteins. These findings could be important for HTT biology, and the significance of the novel variants for HD pathogenesis requires further investigation.

MATERIALS AND METHODS

Cell culture

HeLa cells (human cervix epitheloid carcinoma cells; ECACC) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) containing 4.5 g/L glucose and L-glutamine (Life Technologies) supplemented with 50 U/ml penicillin (Life Technologies), 50 μg/ml streptomycin (Life Technologies) and 10% fetal bovine serum (FBS) (PAA). Cells were grown in a humid environment at 37°C with 5% CO_2 and subcultured by trypsinisation in TryplE solution (Life Technologies) for 5 mins at 37°C prior to replating in fresh medium.

Human Samples

HD patient and control cerebellum, caudate nucleus, prefrontal cortex (Brodmann area 9) and motor cortex (Brodmann area 4) samples were obtained from the New Zealand Neurological Brain Bank in accordance with the ethical principles stated in the Declaration of Helsinki (World Medical Association, 2008). Details of all samples used in this study (11 HD cases and 10 controls) are shown in Table S1. To initially screen the *HTT* gene for splice variants, RNA from 4 HD patients and 4 controls from each of the four brain regions noted above was used. For the qPCR assay, RNA from 9 HD patients and 9 controls from cerebellum was used.

RNA extraction and Reverse transcription

RNA was extracted from HeLa cells and cerebellum, caudate nucleus, Brodmann area 4 and Brodmann area 9 of HD and control human brain tissue, and subject to quality to control as described in Hughes et al, 2014 [11]. Fastprep parameters were set at speed 4.5 for 25 secs for the brain material. Reverse transcription was carried out using oligo$(dT)_{12–18}$ primers as described previously [11].

Primer design

Intron/exon boundaries of the human *HTT* mRNA sequence (**HTT-001 ENST00000355072**) were obtained from Ensembl [12] (Table S2). Fragments of mRNA sequence were inserted into Primer 3 [13, 14] in order to design suitable amplicons 400–700 bp in size. Twenty eight sets of overlapping primers were selected (Table S3).

PCR

PCR was carried out with BioTaq DNA polymerase according to standard protocol (Bioline) and thermocycling consisting of 30 cycles of 95°C for 2 mins, 95°C for 30 sec, 52–67°C for 30 sec (see Table S2 for annealing temperatures), 72°C for 45 sec and a final

extension at 72°C for 10 mins. PCR products were visualised using gel electrophoresis (1-2% agarose) with GelRed (Cambridge Bioscience) and a UVP documentation system. Any bands in addition to the predicted amplicon on the gel were isolated using the band stab method of Bjourson and Cooper, 1992 [15]. Re-amplification of these isolated sequences was performed by PCR as described above. DNA sequencing of isolated PCR products was performed as described previously [11].

Taqman assays for splice variant-specific quantitative PCR

Custom Taqman assays were designed by Life Technologies for splice variants 3 and 17 (Assay ID: AJT96B4 (ex3-7); AJRR9Z0 (ex41b)) and assay Hs00918153_m1 was used to detect the exon 41–42 boundary, representative of the canonical *HTT* transcript. Thermocycling was performed on the 7500 real time PCR system (Life Technologies) as follows: 50°C for 2 mins, 95°C for 10 mins then 40 cycles of 95°C for 15 secs, 60°C for 1 min. For each transcript (*HTT*Δex4,5,6; *HTTex41b*; *HTTex41–42* (canonical)), expression levels were compared between HD patient (n = 9) and control (n = 9) cerebellum by means of a *t* test using a Bonferroni corrected significance level of 0.05. *HTT* expression levels were normalised to that of β-actin using Taqman assay ID: Hs99999903_m1 (Life Technologies). Data was analysed using the ddCt method [16]. Taqman assays were validated on cerebellar cDNA using standard curve.

Three-dimensional homology modelling

The 3D protein structural homology modelling was carried out using the I-TASSER server [17] for canonical wild-type HTT (ENST00000355072), in addition to SV3 (*HTTΔex4,5,6*) and SV17 (*HTTex41b*). I-TASSER was also employed to predict binding sites residing within the modelled structures. Models were annotated with evolutionary sequence conservation data using ConSurf server (http://consurf.tau.ac.il/) to identify functionally important regions. For predicted binding sites only sites with a BS-score>1 were considered as this indicates a high confidence match between the template binding site and the predicted binding site. Phosphorylation sites were predicted using the NetPhos 2.0 Server [18]. All modelled structures were visualised and annotated using Chimera [19].

Analysis of exonic variants which may modulate alternative splicing of HTT.

Using the Exome Variant Server [http://evs.gs.washington.edu/EVS/] 289 exonic variants located in the *HTT* gene were identified. It should be noted that only single-base substitutions within exonic regions were considered in this analysis. MutPred Splice [20] was then used to predict if any of the 289 variants were splicing sensitive and the underlying nature of any splicing mechanism disrupted (e.g. exon skipping via loss of exonic splicing enhancers or loss of natural splice site etc). High confidence splicing sensitive exonic variants were cross referenced against the reported novel alternative splicing events identified in this study, to identify exonic variants which may modulate expression levels of the alternative splicing events identified here.

RESULTS

Alternative transcripts of HTT in human brain

HTT was amplified in 28 overlapping segments spanning all *HTT* exons (P1–P27, Table S3) from cDNA derived from human cerebellum, caudate nucleus, Brodmann area 4 and Brodmann area 9 of four HD patients and four controls (For demographic information see Table S1). This revealed 22 alternative transcripts of *HTT* (Table 1). PCR conditions for each primer set were firstly optimised using cDNA generated from HeLa cells. This was used as it provided a readily available and abundant source of human cDNA. The process showing how each variant was discovered is depicted in Fig. 1 using the examples of SV3 (*HTTΔex4,5,6*) and SV17 (*HTTex41b*) as these were selected for further follow up due to their potential impact on HTT function. Supplementary Figure S1 shows each amplicon on agarose gels with the corresponding novel variants amplified from cerebellar or cortical cDNA. Cerebellum and Brodmann area 9 were used here as they showed the most abundant levels of splice variants when compared with the other brain regions and thus provided the best chance of isolating the novel variant from the canonical transcript by the band-stab method employed here (data not shown). All of the variants were present at lower levels than the canonical transcript (Figure S1). The structure of HTT including known domains, motifs and post-translational modifications is illustrated in Fig. 2 and the predicted

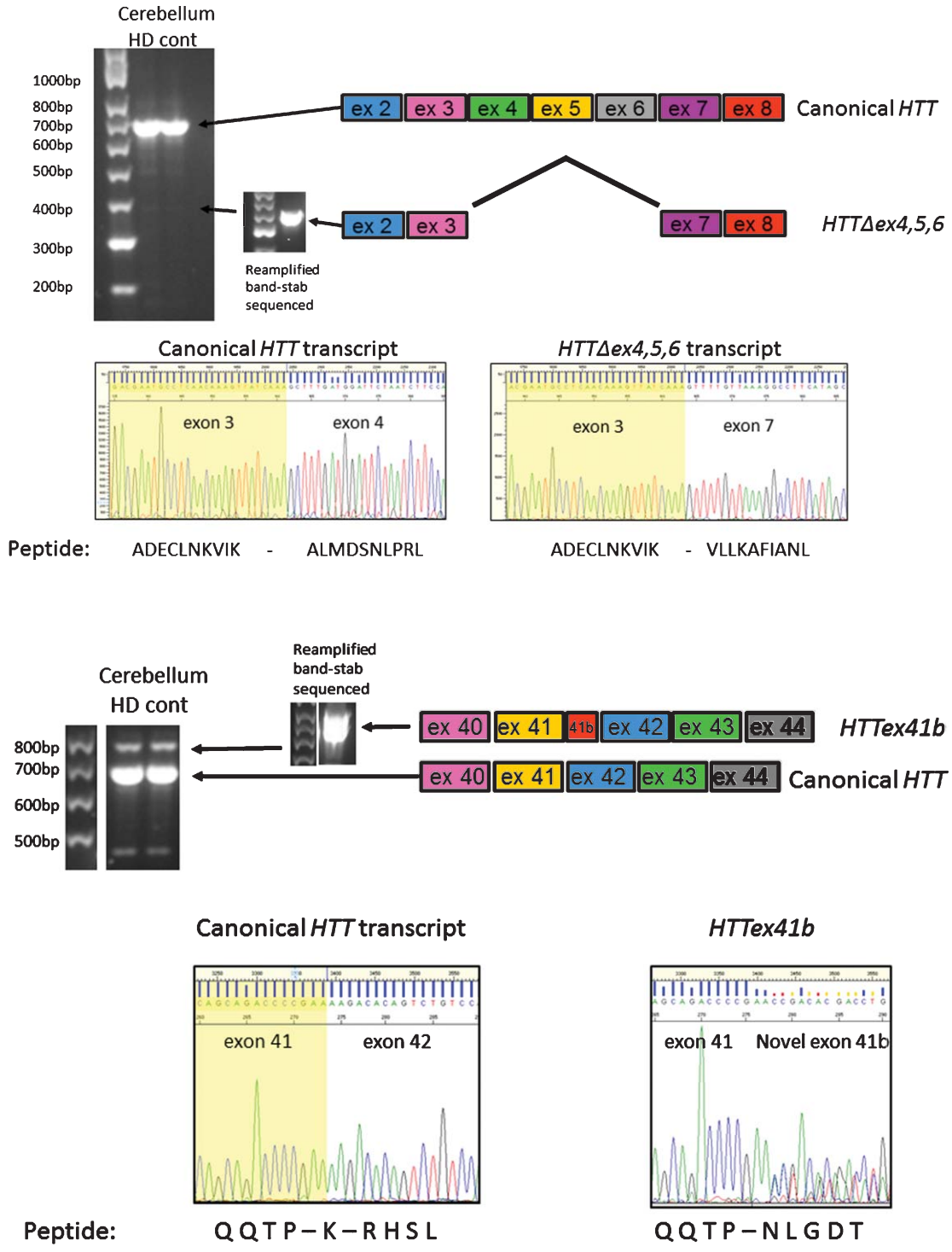

Fig. 1. Subset of novel alternatively spliced transcripts identified in this study. *HTT* was amplified from human cerebellum cDNA and amplicons were isolated by band-stabbing, reamplified then sequenced to confirm identity. Two of the novel splice variants found (SV3 and SV17) are depicted here. SV3 (*HTTΔex4,5,6*) was detected using primers within exon 2 (forward) and exon 8 (reverse) (primer set P2). SV17 (*HTTex41b*) was detected using primers within exon 40 (forward) and exon 44 (reverse) (primer set P16). The main amplicons of interest are shown. Primer sequences (P2 and P16) are detailed in Supplementary Table S3.

Table 1
A list of the 22 human *HTT* splice variants found in brain. Amino acid numbering is according to Ensembl: ENST00000355072. Splice variants which represent in-frame changes and therefore not likely to be subject to nonsense-mediated decay (NMD) are highlighted in grey

SV	Splice variant name	Nature of variant	Change to RNA	Effect on translation	Amino acids affected
SV 1	*HTTΔex3*	Exon 3 skipping	−121bp	Frameshift and PTC	aa116 – Leu – STOP
SV 2	*HTTΔex4*	Exon 4 skipping	−60bp	Frame maintained; loss of 20aa	Δaa157 – 176 inclusive (20aa)
SV 3	*HTTΔex4,5,6*	Exon 4,5 & 6 skipping	−279bp	Frame maintained; loss of 93aa	Δaa157 – 249 inclusive (93aa)
SV 4	*HTTΔex12*	Exon 12 skipping	−341bp	Frameshift; PTC	aa467 – 19aa novel peptide – STOP
SV 5	$HTT\Delta ex13^{trunc}$	Exon 13 3' truncation (48bp)	−48bp	Frame maintained; loss of 16aa	Δaa607 – 622 inclusive (16aa)
SV 6	*HTTin13*	Retention of intron 13	+257bp	Frameshift; PTC	aa622 – 37aa novel peptide – STOP
SV 7	*HTTΔex22*	Exon 22 skipping	−147bp	Frame maintained; loss of 49aa	Δaa934 – 982 inclusive (49aa)
SV 8	$HTT^{+89}in26$	Partial retention of Intron 26 (89bp)	+89bp	Frame maintained; PTC	aa1166 – 35aa novel peptide – STOP
SV 9	*HTTΔex27,28*	Exon 27 and 28 skipping	−255bp	Frame maintained; loss of 85aa	Δaa1167 – 1251 inclusive (85aa)
SV 10	*HTTΔex27*	Exon 27 skipping	−127bp	Frameshift; PTC	aa1166 – 34aa novel peptide – STOP
SV 11	*HTTΔex28*	Exon 28 skipping	−128bp	Frameshift; PTC	aa1208 – 39aa novel peptide – STOP
SV 12	$HTT\Delta ex30\text{-}31^{trunc}$	Combined truncation of exons 30 and 31	−99bp	Frame maintained; loss of 33aa and mutation of 1aa	aa1292 E to D – Δaa1293 – 1325 inclusive (33aa)
SV 13	*HTTin30*	Intron 30 retention	+510bp	Frame maintained; PTC	aa1314 – 30aa novel peptide – STOP
SV 14	$HTT^{+87}in31$	Partial retention of Intron 31	+87bp	Frameshift; PTC	aa1388 – G – STOP
SV 15	*HTTΔex38*	Exon 38 skipping	−123bp	Frame maintained; loss of 41aa	Δaa1623 – 1663 inclusive (41aa)
SV 16	*HTTΔex41*	Exon 41 skipping	−208bp	Frameshift; PTC	aa1789 – novel 34aa peptide – STOP
SV 17	*HTTex41b*	Novel exon 41b	+90bp	Frame maintained; extra 30aa peptide	Insertion after aa1858; aa1859 changed K to R
SV 18	*HTTin43*	Intron 43 retention	+131bp	Frame maintained; PTC	aa1966 – novel 16aa peptide – STOP
SV 19	*HTTin44*	Intron 44 retention	+297bp	Frame maintained; PTC	aa2025 – novel 53aa peptide – STOP
SV 20	*HTTin43+44*	Intron 43 and 44 retention	+428bp	Frame maintained; PTC	aa1966 – novel 16aa peptide – STOP
SV 21	$HTT^{+109}in45$	Partial intron 45 retention	+109bp	Frameshift; PTC	aa2050 – R STOP
SV 22	*HTTin58*	Intron 58 retention	+134bp	Frameshift; PTC	aa2660 – novel 27aa peptide – STOP

effects of each potentially coding novel variant on this structure are shown. Fourteen of the splice variants would result in premature STOP codons thus rendering these transcripts susceptible to NMD and therefore less likely to result in a novel truncated HTT protein isoform. However, 8 of the 22 splice variants (*HTTΔex4*, *HTTΔex4,5,6*, $HTT\Delta ex13^{trunc}$, *HTTΔex22*, *HTTΔex27,28*, $HTT\Delta ex30\text{-}31^{trunc}$, *HTTΔex38*, *HTTex41b*) contain in-frame changes which could translate into alternative HTT protein isoforms, through either the inclusion or exclusion of specific peptide sequences. As far as we are aware none of the variants would encode a full length HTT protein isoform devoid of any of the known proteolytic cleavage sites as the exon skipping events reported here do not involve these regions. Some noteworthy predictions however, are isoforms that lose the karyopherin B type nuclear localisation signal (NLS) encoded by amino acids 174–207 [21] from splice variants *HTTΔex4* and *HTTΔex4,5,6*. Three sites of posttranslational modification of importance (acetylation at Lys 178 and Lys 236 [22]; palmitoylation at Cys 214 [23] would also be lost from the *HTTΔex4,5,6* variant which could modify the regulation and function of HTT. Another reported HTT NLS encoded by amino acids 1182 – 1190 [24] would be lost from splice variant *HTTΔex27,28*. This variant would also lose phosphorylation sites Ser1181 and Ser1201 phosphorylated by cdk5 [25] and known to influence HD pathogenesis and

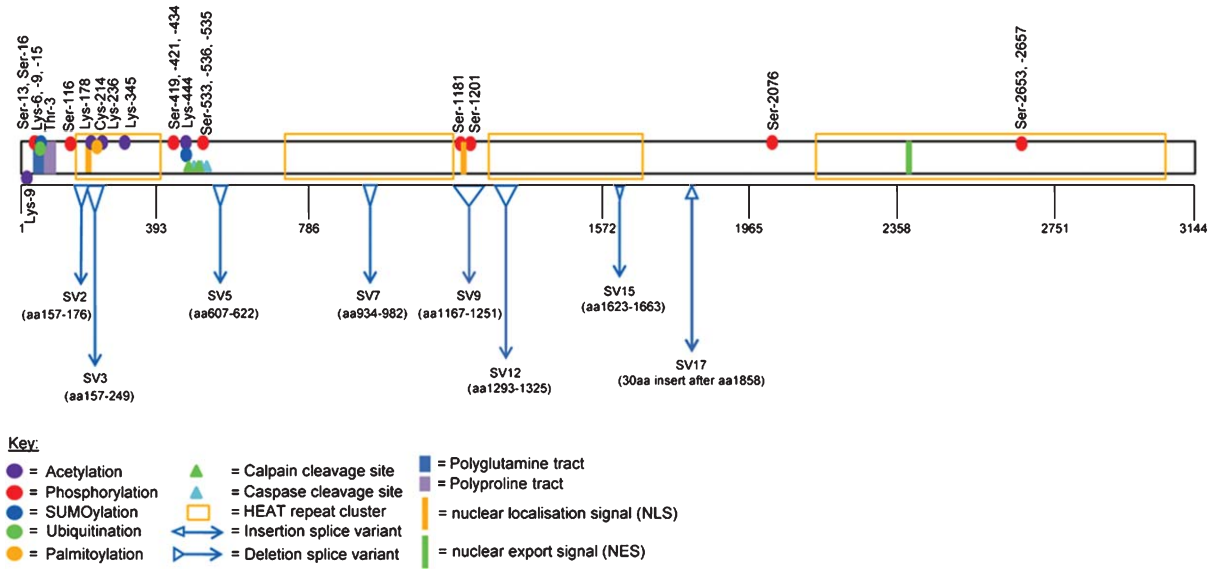

Fig. 2. Annotation of the HTT protein relative to its known domains and post-translational modifications. Huntingtin (HTT) is a large protein with few known domains and various sites of post-translational modifications. These are depicted in the diagram here. Alterations to the *HTT* transcript have the potential to impact these regions of the protein if protein-coding. Eight in-frame splice variants found in this study and predicted to encode novel HTT protein isoforms are mapped on to show their potential impact on the HTT protein.

phenotype [26] (Fig. 2). It is therefore feasible that any protein encoded by these RNA variants could exhibit different features to that of full length canonical HTT.

Determining expression levels of HTT alternative transcripts in HD and control brain

Of the 22 variants, the expression levels of two variants (SV3 and SV17) were assessed by quantitative PCR (qPCR). SV3 was selected for further study given the potential impact on the HTT protein of the removal of a NLS and various sites of post-translational modification of importance. SV17 was selected as it was the only variant that would introduce a novel peptide sequence into the HTT protein. Taqman probes designed selectively to amplify SV3 (*HTTΔex4,5,6*) (probe spanning exons 3 to 7 and thus not detecting the canonical *HTT* transcript) and SV17 (*HTTex41b*) (probe binding the novel exon 41b sequence and thus not detecting the canonical *HTT* transcript) showed that these alternative transcripts were expressed in HD brain, however there was no statistically significant difference in the levels of the individual novel splice variants between HD cases and controls (SV3 $p=0.11$; SV17 $p=0.09$; Table 2).

3D structural homology modelling of novel splice variants

To identify potential functional consequences of *HTTΔex4,5,6* (SV3) and *HTTex41b* (SV17) we employed 3D structural homology modelling. These splice variants were selected for the study of expression levels and 3D modelling given their potential to impact the HTT protein as noted above. Structural homology modelling of wild type HTT encompassing exons 4, 5 and 6 (absent in *HTTΔex4,5,6*) is shown in Fig. 3. Exons 4, 5 and 6 (aa157–249) are predicted to encode a protein binding site (Fig. 3A) with 9 residues involved as protein-protein interaction sites (aa207, aa210, aa211, aa214, aa215, aa246, aa247, aa250, aa251). Seven of these residues occur within exons 4, 5 and 6. The predicted binding site in exons 4, 5 and 6 demonstrates a high level of homology with a binding site found in Importin subunit beta-1, a protein involved in nuclear import, which binds to five FxFG nucleoporin repeats from Nsp1p (PDB ID: 1F59; [27]). Further supporting evidence for the predicted significance of exons 4, 5 and 6 is highlighted by the high evolutionary sequence conservation of this region (Fig. 3B), that it is known to encode a NLS between residues aa174–207 [21] and that in its absence, post-translational modification sites of significance would be ablated (Fig. 2). For the other splice variant

Table 2
Assessment of expression levels of *HTT* alternative transcripts in HD ($n = 9$) and control ($n = 9$) brain by quantitative PCR. Taqman assays were used in order to amplify the canonical *HTT* transcript (assay Hs00918153_m1) and to exclusively amplify each unique splice variant (assays AJRR920 and AJT96B4)

Taqman Assay	Transcript amplified	$2^{-\Delta\Delta CT}$ (Controls)	$2^{-\Delta\Delta CT}$ (HD)	p value
Hs00918153_m1	exon 41–42	1	0.85	0.468
AJRR920	exon 41b	1	0.73	0.091
AJT96B4	exon 3–7	1	0.75	0.110

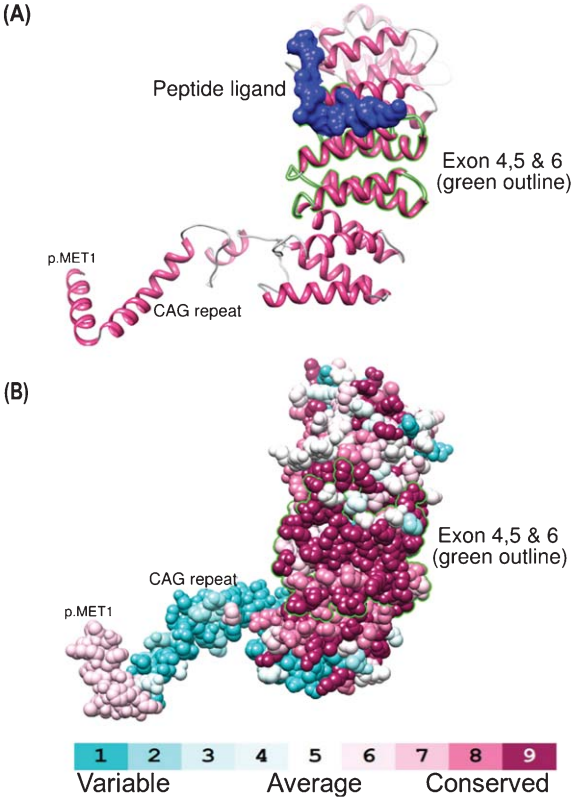

Fig. 3. 3D structural model of wild-type HTT (residues 1–400). This model of N-terminal HTT is annotated with a predicted ligand binding site **(A)** and evolutionary sequence conservation data **(B)**. The region corresponding to exons 4, 5 & 6 (aa157–249), which is absent in the novel *HTT* splice variant named SV3, is shown with a green outline. The ligand (shown in blue) is predicted to bind at 9 residues (aa207, aa210, aa211, aa214, aa215, aa246, aa247, aa250, aa251) of which 7 residues are within exons 4, 5 & 6. The high sequence conservation of exons 4, 5 & 6 and predicted ligand binding, indicates that this region is of potential functional significance. The absence of this region (exons 4, 5 & 6) in SV3, indicates that SV3 encodes a potential loss of function protein variant.

HTTex41b (SV17), which results in the inclusion of a novel exon (exon 41b) and thus extending the size of full length HTT, it is predicted that the novel peptide encoded by exon 41b would gain two phosphorylation sites at aaT1863 and aaS1883 (Fig. 4).

Identification of exonic variants which may modulate expression levels of naturally occurring alternative splicing events in HTT

Outside of the splice sites, additional auxiliary *cis*-acting elements play a crucial role in exon recognition [28]. In some cases genetic variation, both exonic and intronic, can serve to weaken (or strengthen) exon definition resulting in aberrant splicing of pre-mRNA or alternatively serve to alter expression levels of naturally occurring alternative splicing (such as exon skipping). Our focus here was exclusively on exonic variants as exonic cis-acting elements such as exonic splicing enhancers (ESE) and exonic splicing silencers (ESS) are currently much better characterized than their intronic counterparts. *In silico* analysis of 289 exonic variants (126 same-sense, 163 missense; MAF% 0.0078 to 25.3) within the *HTT* gene identified 21 variants predicted to either modulate natural alternative splicing or result in aberrant pre-mRNA splicing of *HTT* (Table 3). Of these, the majority were rare variants (MAF <1%), with only one common variant found (rs2276881G>A; EA MAF% = 7.1%, AA MAF% = 1.3%). Based on the exon location of these genetic variants, it is possible that within an individual, expression levels of naturally occurring alternatively spliced transcripts could in some cases be either up or down regulated for individuals harbouring the appropriate exonic variant. Five exonic variants (rs368590997C>T, rs368866386C>T, rs372467345C>T, rs369912429A>G, rs371907257A>G; Table 3) were predicted *in silico* to alter the expression level of six different novel alternative splicing events identified here; *HTTΔex3* (SV1), $HTT\Delta ex13^{trunc}$ (SV5), *HTTin13* (SV6), $HTT^{+89}in26$ (SV8), *HTTΔex41* (SV16) and *HTTex41b* (SV17).

DISCUSSION

Since the discovery of the *HTT* gene in 1993 [2] there have been few reports supporting the existence of transcripts other than the full length canonical transcript [10, 11, 29, 30, 31]. Here, we present evidence for mul-

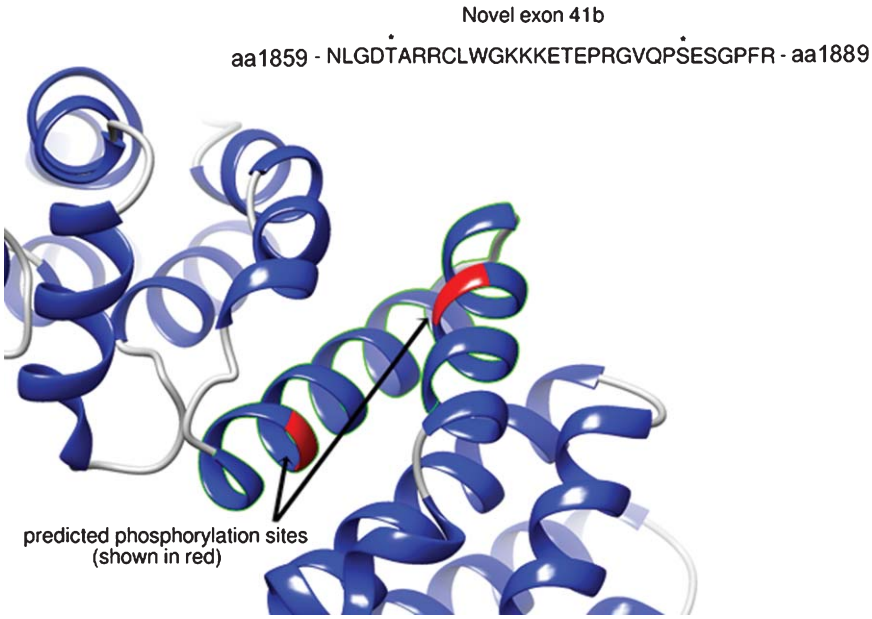

Fig. 4. 3D structural model of HTTex41b (residues 1201–2340). This region of HTT has an additional in-frame insertion of a novel peptide sequence encoded by a novel exon termed exon 41b (shown with green outline). Novel exon 41b is annotated with two predicted phosphorylation sites (shown in red and by asterisk on exon 41b protein sequence). The gain of two phosphorylation sites within exon 41b indicates that this splice variant is of potential functional significance.

tiple alternative splice forms of *HTT* in human brain. Of the 22 splice variants discovered, eight were in-frame changes which were not predicted to be subject to NMD and as such could result in translated HTT protein isoforms. However, the small size difference between these splicing events could preclude the detection of the translated proteins by western blotting alone where separation from the canonical 350kDa HTT protein can prove difficult. In studies where the full length wildtype and mutant HTT has been well resolved into its 2 alleles by western blot [32] there is no evidence of these alternative proteins, which may simply reflect their relatively low abundance. It was not possible to accurately determine the abundance of these splice variants by the standard PCR approach used here, however it was noted that they were present at low levels compared to the canonical transcript (Fig. 1 and Figure S1). This finding was recapitulated for the two splice variants selected for follow up by qPCR (SV3 and SV17) given their high Ct values when compared to the canonical transcript.

Close inspection of the structure of HTT and its post-translational modifications (Fig. 2) highlights those variants likely to exert the most significant functional effect on HTT. Splicing can affect domain architecture [8] and here, two of the most interesting variants were modelled by 3D protein homology modelling: 1) SV3, where HTT would lose its NLS, various post-translational modification sites and a predicted ligand binding site and 2) SV17, where HTT gains a novel in-frame peptide sequence introducing two novel phosphorylation sites. For SV3, loss of a palmitoylation site in HTT could influence its normal trafficking to the Golgi and thus affect the toxicity of the HTT protein [23]. Protein phosphorylation is involved in a wide range of cellular activities including the regulation of protein function [33] and in HD, it is well established that the phosphorylation of HTT at various sites can alter its cellular toxicity [34]. Therefore for *HTTex41b* (SV17), the gain of two phosphorylation sites within the novel exon 41b could have important functional consequences for this putative HTT protein isoform.

Our previous study in mouse showed significantly reduced splicing in HdhQ150 mouse cerebellum compared with wild type cerebellum [11]. Investigations of *HTT* splicing and its significance to HD pathology are important as splicing can occur either co-transcriptionally or post-transcriptionally with the former thought to be the dominant method [35]. It is therefore possible that transcription of the mutant *HTT* transcript could influence its splicing and also the splicing of *HTT* could regulate its transcriptional activity.

Table 3
In silico identification of 21 (out of 289) exonic single base substitutions in *HTT* identified in 6503 individuals from NHLBI GO Exome Sequencing Project which may modulate the expression level of naturally occurring alternative splicing events or result in aberrant pre-mRNA splicing. The majority of predicted splicing sensitive variants listed here are rare (MAF <1%) with one variant named rs2276881 occurring at polymorphic frequencies (MAF = 5%). MAF = Minor allele frequency in percent, EA = European Americans, AA = African Americans, NA = not applicable

dbSNP	HGVS (cdna)	HGVS (Protein)	Splice variant predicted to be affected (Exon location of genetic variant)	Confident Splicing Hypothesis for exonic variant	MAF % (EA/AA/All)
rs368590997	c.405C>T	p.(C135 =)	Increased expression of SV1 (Exon 3)	Exon skipping via ESS Gain ($P = 0.034846$), ESE Loss ($P = 0.000173$)	0.0119/0.0/0.008
rs370043648	c.1275T>A	p.(A425 =)	NA (Exon 10)	Loss of natural 3' SS ($P < 0.000001$)	0.0/0.0272/0.0084
rs368866386	c.1752C>T	p.(D584 =)	SV5 & SV6 (Exon 13)		0.0/0.0262/0.0083
rs372467345	c.1758C>T	p.(T586 =)	SV5 & SV6 (Exon 13)	Cryptic 5' SS ($P = 0.016038$)	0.0121/0.0/0.0083
rs374614171	c.2241A>G	p.(E747 =)	NA (Exon 17)	Cryptic 5' SS ($P = 0.019091$)	0.0/0.0258/0.0082
rs376638534	c.3136C>T	p.(H1046Y)	NA (Exon 24)		0.0/0.0266/0.0084
rs200289562	c.3127T>G	p.(L1043V)	NA (Exon 24)		0.0122/0.0/0.0083
rs369197307	c.3072A>G	p.(G1024 =)	NA (Exon 24)	Exon skipping via ESS Gain ($P = 0.000732$), ESE Loss ($P = 0.046525$)	0.0122/0.0/0.0083
rs369912429	c.3470A>G	p.(D1157G)	Increased expression of SV8 (Exon 26)	Exon skipping via ESS Gain ($P = 0.034846$), ESE Loss ($P = 0.004064$)	0.0121/0.0/0.0082
rs375899948	c.4255C>A	p.(H1419N)	NA (Exon 33)		0.0123/0.0/0.0085
rs372640158	c.4250C>G	p.(A1417G)	NA (Exon 33)		0.0/0.0276/0.0085
rs369637161	c.4455C>T	p.(G1485 =)	NA (Exon 34)		0.0/0.0279/0.0085
rs377474866	c.4586C>T	p.(A1529V)	NA (Exon 35)	Cryptic 5' SS ($P = 0.010735$)	0.0/0.024/0.0079
rs371907257	c.5371A>G	p.(M1791V)	SV16 & SV17 (Exon 41)		0.0/0.0494/0.0161
rs370970140	c.5667C>T	p.(C1889 =)	NA (Exon 42)		0.0122/0.0/0.0083
rs199941479	c.6163C>T	p.(L2055F)	NA (Exon 46)		0.0122/0.0/0.0083
rs376635317	c.6951C>T	p.(A2317 =)	NA (Exon 50)		0.0119/0.0/0.008
rs368077431	c.6976C>T	p.(L2326F)	NA (Exon 51)	Exon skipping via ESS Gain ($P = 0.000003$)	0.012/0.0/0.0082
rs2276881	c.8157G>A	p.(L2719 =)	NA (Exon 60)	Exon skipping via ESS Gain ($P = 0.000732$), Cryptic 5' SS ($P = 0.016129$)	7.0495/1.3272/5.1167
rs368288911	c.8792A>G	p.(E2931G)	NA (Exon 64)	Exon skipping via ESE Loss ($P = 0.000173$)	0.0/0.0255/0.0082
rs369633799	c.9213G>A	p.(A3071 =)	NA (Exon 66)		0.0/0.0241/0.008

Moreover, RNA toxicity mechanisms have the potential to affect splicing in HD [6]. However, we did not detect any significant changes in levels of the novel splice forms between human HD and control cerebellum in this study. Given the genetic variation in humans, it is likely that substantially more samples would be needed to detect such differences.

When compared with our murine study [11] it was apparent that there were many more splice variants of *HTT* in human brain compared with mouse brain. Even though high levels of conservation for alternative splicing between human and mouse have been reported [36], there are more alternatively spliced transcripts in human than mouse [9, 37]. This could in part explain the differences between our studies. Non-conserved isoforms have been shown to be functional [38] so the human specific nature of these variants does not necessarily preclude their functional importance, and could reflect evolution of HTT function [39]. Even though measures were taken to ensure detection of *bone fide* splice variants through the generation of cDNA using polyd(T) primers to detect only polyadenylated mature transcripts, we cannot rule out the possibility that aberrant splicing may occur and so, as more RNA sequencing data is generated, it will be interesting to screen these datasets to further investigate their presence, abundance, tissue specificity and pathological significance. The 22 *HTT* splice variants reported here (Table 1) may not be an exhaustive list as the RT-PCR approach employed to screen for them may not have been all inclusive. Moreover, this approach would not have detected the novel variant discovered by Sathasivam et al, 2013 as it is generated by the use of a cryptic polyadenylation signal and does not involve any exon

skipping events [10]. Further studies are required to confirm and validate these variants.

From the exploratory *in silico* analysis of exonic variants (rare mutations and common polymorphisms), it is unclear what the impact of changes to the expression levels of naturally occurring splice variants would be.

In conclusion, we have discovered multiple novel alternative splice forms of *HTT* in human brain with the potential to encode HTT protein isoforms which may differ in function to the full length HTT protein. It remains to be shown whether the individual splicing events occur together in one transcript or whether several different alternative transcripts comprising combinations of these events exist. A comprehensive knowledge of all *HTT* RNA species is important on a number of levels. Firstly, RNA silencing therapies in HD are currently of great interest. Some approaches target specific regions of *HTT* RNA [40], therefore consideration of all variants may be important in this strategy in order to silence all *HTT* transcripts that could encode HTT protein isoforms and interfere with phenotype. Secondly, even though a high proportion of alternative splice events result in an altered protein [41], alternative splicing does not solely operate to generate distinct protein isoforms, but may also influence gene regulation through translation control, mRNA stability and localisation [42]. RNA variants could therefore play a role in the regulation of the *HTT* gene. Thirdly, these variants (or combination of in-frame variants) may point to the existence of as yet unidentified HTT protein isoforms. These could contribute to a better understanding of the function of the wild type HTT protein which is yet to be fully elucidated. Of interest is the finding that splicing events can be forced to alter the structure of the HTT protein which has been shown to reduce the toxicity of the mutant HTT protein [43]. It is therefore important that the full repertoire of *HTT* transcripts and protein isoforms is delineated.

ACKNOWLEDGMENTS

We thank Richard Faull and the New Zealand Neurological Brain Bank for tissue. We are grateful to Niels Henning-Skotte and Chris Kay at the University of British Columbia for helpful discussions and critical reading of the manuscript. The technical assistance of Claudia Consoli, Keith Hart and Barrie Francis of Cardiff University's Central Biotechnology Services is gratefully acknowledged. This project was funded by the E J Norman Fund (AH), EHDN (AH) and the MRC Centre for Neuropsychiatric Genetics and Genomics (Grant code: G0801418) (FC).

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

SUPPLEMENTARY MATERIAL

The supplementary table and figure are available in the electronic version of this article: http://dx.doi.org/10.3233/JHD-150151.

REFERENCES

[1] Ross CA, Aylward EH, Wild EJ, Langbehn DR, Long JD, Warner JH, Scahill RI, Leavitt BR, Stout JC, Paulsen JS, Reilmann R, Unschuld PG, Wexler A, Margolis RL, Tabrizi SJ. Huntington disease: Natural history, biomarkers and prospects for therapeutics. Nat Rev Neurol. 2014;10(4):204-16.

[2] HDCRG. A novel gene containing a trinucleotde repeat that is expanded and unstable on Huntington's disease chromosomes. Cell. 1993;72(6):971-83.

[3] Fiszer A, Krzyzosiak WJ. RNA toxicity in polyglutamine disorders: Concepts, models, and progress of research. J Mol Med (Berl). 2013;91(6):683-91.

[4] Krzyzosiak WJ, Sobczak K, Wojciechowska M, Fiszer A, Mykowska A, Kozlowski P. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. Nucleic Acids Res. 2012;40(1):11-26.

[5] Nalavade R, Griesche N, Ryan DP, Hildebrand S, Krauss S. Mechanisms of RNA-induced toxicity in CAG repeat disorders. Cell Death Dis. 2013;4:e752.

[6] de Mezer M, Wojciechowska M, Napierala M, Sobczak K, Krzyzosiak WJ. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. Nucleic Acids Res. 2011;39(9):3852-63.

[7] Havens MA, Duelli DM, Hastings ML. Targeting RNA splicing for disease therapy. Wiley Interdiscip Rev RNA. 2013;4(3):247-66.

[8] Light S, Elofsson A. The impact of splicing on protein domain architecture. Curr Opin Struct Biol. 2013;23(3):451-8.

[9] Mudge JM, Frankish A, Fernandez-Banet J, Alioto T, Derrien T, Howald C, Reymond A, Guigo R, Hubbard T, Harrow J. The origins, evolution, and functional potential of alternative splicing in vertebrates. Mol Biol Evol. 2011;28(10):2949-59.

[10] Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, Smith DL, Faull RL, Roos RA, Howland D, Detloff PJ, Housman DE, Bates GP. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proc Natl Acad Sci USA. 2013;110(6):2366-70.

[11] Hughes AC, Mort M, Elliston L, Thomas RM, Brooks SP, Dunnett SB, Jones L. Identification of novel alternative splicing events in the huntingtin gene and assessment of the functional consequences using structural protein homology modelling. J Mol Biol. 2014;426(7):1428-38.

[12] Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt SE, Janacek SH,

Johnson N, Juettemann T, Kahari AK, Keenan S, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Overduin B, Parker A, Patricio M, Perry E, Pignatelli M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Aken BL, Birney E, Harrow J, Kinsella R, Muffato M, Ruffier M, Searle SM, Spudich G, Trevanion SJ, Yates A, Zerbino DR, Flicek P. Ensembl 2015. Nucleic Acids Res. 2015;43(Database issue):D662-9.

[13] Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007;23(10):1289-91.

[14] Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3–new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115.

[15] Bjourson AJ, Cooper JE. Band-stab PCR: A simple technique for the purification of individual PCR products. Nucleic Acids Res. 1992;20(17):4675.

[16] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.

[17] Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5(4):725-38.

[18] Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol. 1999;294(5):1351-62.

[19] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-12.

[20] Mort M, Sterne-Weiler T, Li B, Ball EV, Cooper DN, Radivojac P, Sanford JR, Mooney SD. MutPred Splice: Machine learning-based prediction of exonic variants that disrupt splicing. Genome Biol. 2014;15(1):R19.

[21] Desmond CR, Atwal RS, Xia J, Truant R. Identification of a karyopherin beta1/beta2 proline-tyrosine nuclear localization signal in huntingtin protein. J Biol Chem. 2012;287(47):39626-33.

[22] Cong X, Held JM, DeGiacomo F, Bonner A, Chen JM, Schilling B, Czerwieniec GA, Gibson BW, Ellerby LM. Mass spectrometric identification of novel lysine acetylation sites in huntingtin. Mol Cell Proteomics. 2011;10(10):M111 009829.

[23] Yanai A, Huang K, Kang R, Singaraja RR, Arstikaitis P, Gan L, Orban PC, Mullard A, Cowan CM, Raymond LA, Drisdel RC, Green WN, Ravikumar B, Rubinsztein DC, El-Husseini A, Hayden MR. Palmitoylation of huntingtin by HIP14 is essential for its trafficking and function. Nat Neurosci. 2006;9(6):824-31.

[24] Bessert DA, Gutridge KL, Dunbar JC, Carlock LR. The identification of a functional nuclear localization signal in the Huntington disease protein. Brain Res Mol Brain Res. 1995;33(1):165-73.

[25] Anne SL, Saudou F, Humbert S. Phosphorylation of huntingtin by cyclin-dependent kinase 5 is induced by DNA damage and regulates wild-type and mutant huntingtin toxicity in neurons. J Neurosci. 2007;27(27):7318-28.

[26] Ben M'Barek K, Pla P, Orvoen S, Benstaali C, Godin JD, Gardier AM, Saudou F, David DJ, Humbert S. Huntingtin mediates anxiety/depression-related behaviors and hippocampal neurogenesis. J Neurosci. 2013;33(20):8608-20.

[27] Bayliss R, Littlewood T, Stewart M. Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking. Cell. 2000;102(1):99-108.

[28] Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: Diversification, exon definition and function. Nat Rev Genet. 2010;11(5):345-55.

[29] Lin B, Nasir J, MacDonald H, Hutchinson G, Graham RK, Rommens JM, Hayden MR. Sequence of the murine Huntington disease gene: Evidence for conservation, alternate splicing and polymorphism in a triplet (CCG) repeat [corrected]. Hum Mol Genet. 1994;3(1):85-92.

[30] Matsuyama N, Hadano S, Onoe K, Osuga H, Showguchi-Miyata J, Gondo Y, Ikeda JE. Identification and characterization of the miniature pig Huntington's disease gene homolog: Evidence for conservation and polymorphism in the CAG triplet repeat. Genomics. 2000;69(1):72-85.

[31] Gissi C, Pesole G, Cattaneo E, Tartari M. Huntingtin gene evolution in Chordata and its peculiar features in the ascidian Ciona genus. BMC Genomics. 2006;7:288.

[32] Southwell AL, Skotte NH, Kordasiewicz HB, Ostergaard ME, Watt AT, Carroll JB, Doty CN, Villanueva EB, Petoukhov E, Vaid K, Xie Y, Freier SM, Swayze EE, Seth PP, Bennett CF, Hayden MR. *In vivo* evaluation of candidate allele-specific mutant huntingtin gene silencing antisense oligonucleotides. Mol Ther. 2014;22(12):2093-106.

[33] Barford D, Das AK, Egloff MP. The structure and mechanism of protein phosphatases: Insights into catalysis and regulation. Annu Rev Biophys Biomol Struct. 1998;27:133-64.

[34] Ehrnhoefer DE, Sutton L, Hayden MR. Small changes, big impact: Posttranslational modifications and function of huntingtin in Huntington disease. Neuroscientist. 2011;17(5):475-92.

[35] Brugiolo M, Herzel L, Neugebauer KM. Counting on co-transcriptional splicing. F1000Prime Rep. 2013;5:9.

[36] Thanaraj TA, Clark F, Muilu J. Conservation of human alternative splice events in mouse. Nucleic Acids Res. 2003;31(10):2544-52.

[37] Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ, Blencowe BJ. The evolutionary landscape of alternative splicing in vertebrate species. Science. 2012;338(6114):1587-93.

[38] Merkin J, Russell C, Chen P, Burge CB. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science. 2012;338(6114):1593-9.

[39] Lo Sardo V, Zuccato C, Gaudenzi G, Vitali B, Ramos C, Tartari M, Myre MA, Walker JA, Pistocchi A, Conti L, Valenza M, Drung B, Schmidt B, Gusella J, Zeitlin S, Cotelli F, Cattaneo E. An evolutionary recent neuroepithelial cell adhesion function of huntingtin implicates ADAM10-Ncadherin. Nat Neurosci. 2012;15(5):713-21.

[40] van Bilsen PH, Jaspers L, Lombardi MS, Odekerken JC, Burright EN, Kaemmerer WF. Identification and allele-specific silencing of the mutant huntingtin allele in Huntington's disease patient-derived fibroblasts. Hum Gene Ther. 2008;19(7):710-9.

[41] Modrek B, Lee C. A genomic view of alternative splicing. Nat Genet. 2002;30(1):13-9.

[42] Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. Genes Dev. 2003;17(4):419-37.

[43] Evers MM, Tran HD, Zalachoras I, Meijer OC, den Dunnen JT, van Ommen GJ, Aartsma-Rus A, van Roon-Mom WM. Preventing formation of toxic N-terminal huntingtin fragments through antisense oligonucleotide-mediated protein modification. Nucleic Acid Ther. 2014;24(1):4-12.