




# Small molecule splicing modifiers with systemic HTT-lowering activity

Anuradha Bhattacharyya<sup>1</sup>, Christopher R. Trotta <sup>1</sup>, Jana Narasimhan<sup>1</sup>, Kari J. Wiedinger<sup>1</sup>, Wencheng Li<sup>1</sup>, Kerstin A. Effenberger<sup>1</sup>, Matthew G. Woll<sup>1</sup>, Minakshi B. Jani<sup>1</sup>, Nicole Risher<sup>1</sup>, Shirley Yeh<sup>1</sup>, Yaofeng Cheng<sup>1</sup>, Nadiya Sydorenko<sup>1</sup>, Young-Choon Moon<sup>1</sup>, Gary M. Karp<sup>1</sup>, Marla Weetall<sup>1</sup>, Amal Dakka<sup>1</sup>, Vijayalakshmi Gabbeta<sup>1</sup>, Nikolai A. Naryshkin<sup>1</sup>, Jason D. Graci<sup>1</sup>, Thomas Tripodi Jr.<sup>1</sup>, Amber Southwell<sup>2</sup>, Michael Hayden<sup>3</sup>, Joseph M. Colacino<sup>1</sup> & Stuart W. Peltz <sup>1</sup> 

Huntington's disease (HD) is a hereditary neurodegenerative disorder caused by expansion of cytosine-adenine-guanine (CAG) trinucleotide repeats in the huntingtin (*HTT*) gene. Consequently, the mutant protein is ubiquitously expressed and drives pathogenesis of HD through a toxic gain-of-function mechanism. Animal models of HD have demonstrated that reducing huntingtin (HTT) protein levels alleviates motor and neuropathological abnormalities. Investigational drugs aim to reduce HTT levels by repressing *HTT* transcription, stability or translation. These drugs require invasive procedures to reach the central nervous system (CNS) and do not achieve broad CNS distribution. Here, we describe the identification of orally bioavailable small molecules with broad distribution throughout the CNS, which lower *HTT* expression consistently throughout the CNS and periphery through selective modulation of pre-messenger RNA splicing. These compounds act by promoting the inclusion of a pseudoexon containing a premature termination codon (stop-codon psiExon), leading to *HTT* mRNA degradation and reduction of HTT levels.

<sup>1</sup> PTC Therapeutics, Inc. 100 Corporate Court, South Plainfield, NJ, USA. <sup>2</sup> Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, Orlando, FL, USA. <sup>3</sup> Centre for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada. ✉email: [speltz@ptcbio.com](mailto:speltz@ptcbio.com)

Huntington's disease (HD) is an autosomal dominant progressive neurodegenerative disorder. HD is caused by cytosine–adenine–guanine (CAG) repeat expansions in the huntingtin (*HTT*) gene resulting in the production of a ubiquitously expressed pathogenic mutant HTT (mHTT) protein<sup>1–3</sup>. HD is characterised by progressive atrophy of the striatum, cortex and other areas of the brain that causes motor, cognitive and psychiatric symptoms. Neurodegeneration progresses to the thalamus, substantia nigra pars reticulata and subthalamic nucleus in advanced disease<sup>4–6</sup>.

Currently, there are no approved disease-modifying treatments for HD. However, significant advancements are being made in identifying huntingtin (HTT) protein-lowering therapies using multiple approaches, including ribonucleic acid (RNA) interference using short interfering RNAs, short-hairpin RNAs, or microRNAs and antisense oligonucleotides causing translational repression or messenger RNA (mRNA) degradation. Transcriptional repression approaches using zinc finger proteins or clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) have also been employed<sup>1,2</sup>.

The key challenge for these HTT-lowering therapies is optimal delivery and distribution throughout the central nervous system (CNS). These agents do not cross the blood–brain barrier, they require intrathecal or intraparenchymal delivery, and result in uneven CNS distribution and inconsistent HTT lowering throughout the CNS. Peripheral morbidities such as weight loss and skeletal muscle atrophy can occur and may contribute to disease severity<sup>7,8</sup>. Therefore, an orally bioavailable, systemically distributed, brain-penetrating small molecule that reduces the toxic burden of mHTT protein uniformly throughout the brain and peripheral tissues would be highly beneficial for HD management.

We have developed a drug discovery splicing platform to identify compounds that modulate splicing. This platform can be used to develop target-selective small molecules, such as the first-ever orally bioavailable modifier of survival of motor neuron 2 (*SMN2*) splicing for the treatment of spinal muscular atrophy (SMA)<sup>9</sup>. Evrysdi™ (risdiplam) was discovered using this platform and has recently been approved by the Food and Drug Administration and several additional regulatory authorities<sup>10</sup>. Here, we describe the discovery of a class of small molecule splicing modifiers that were specifically synthesised to promote selective splicing of an inducible pseudoexon containing a premature termination codon (stop-codon psiExon), reducing huntingtin mRNA and protein levels in cells and animal models. These results extend the application of our splicing platform to include the identification of orally bioavailable compounds to treat diseases by decreasing gene expression.

## Results

**Identification of HTT-lowering compounds.** HTT-lowering compounds were identified by screening ~300,000 diverse molecules from a proprietary chemical library. The screening was performed using a highly sensitive and robust HTT protein detection assay in fibroblasts isolated from HD patients (HD fibroblasts). Multiple classes of active compounds (hits) were identified, including heat shock protein 90 inhibitors (HTT-A) previously shown to reduce mHTT levels<sup>11</sup> (Supplementary Fig. 1a–c) and splicing modifiers HTT-C1 and HTT-D1 (Fig. 1a), two compounds similar to molecules identified in the SMA drug development programme<sup>12,13</sup>.

HD fibroblasts were treated with the splicing modifiers, and mHTT mRNA and protein levels were determined using reverse transcription-quantitative polymerase chain reaction (RT-qPCR), electrochemiluminescence (ECL) and western blot analysis. Both

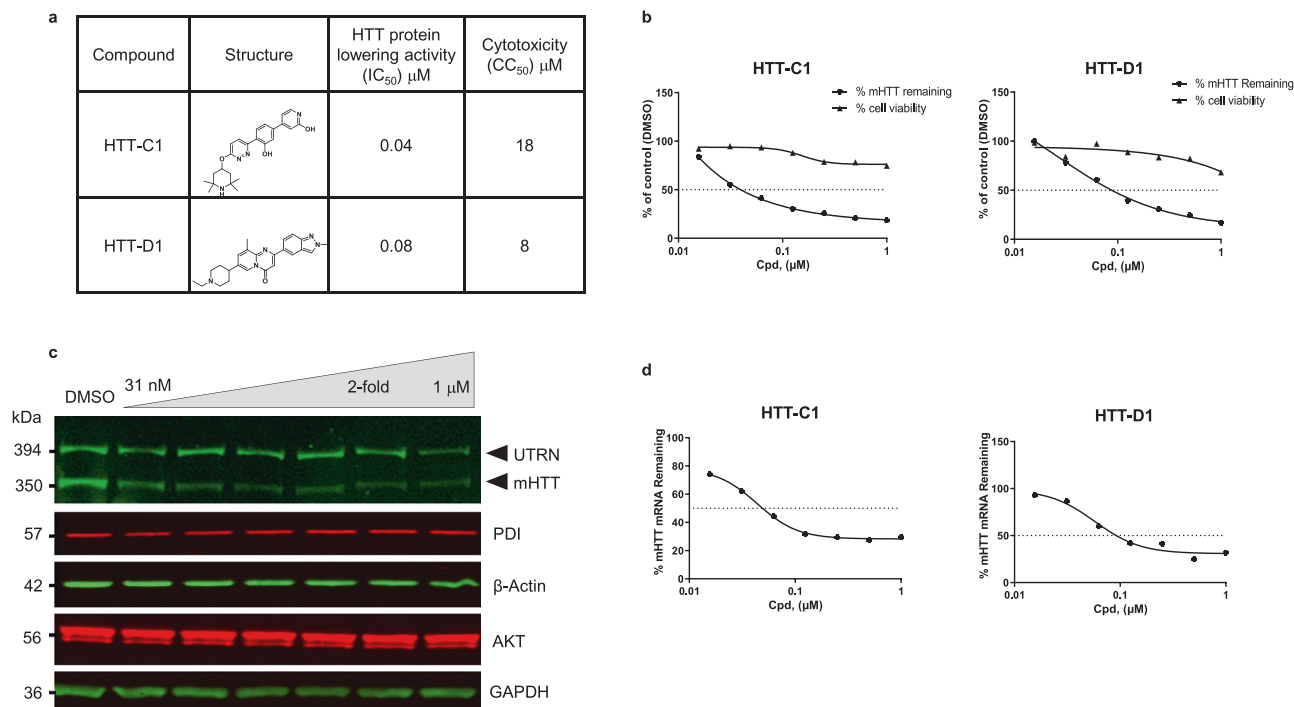
HTT-C1 and HTT-D1 dose-dependently decreased mHTT mRNA and protein levels (Fig. 1b–d; Fig. 2a; Supplementary Fig. 1d). Since most HD patients carry both wild-type (wt) and mHTT alleles, we quantified wt HTT protein levels and demonstrated similar decreases in the control fibroblasts (Supplementary Fig. 1e).

**Compounds affect *HTT* pre-mRNA splicing.** *HTT* pre-mRNA splice junctions were characterised using primer walking. Differential splicing was observed between exons 49 and 54 (Fig. 2b, Supplementary Fig. 2a, b). Splice junctions from these samples were also analysed using targeted next-generation sequencing (NGS; AmpliSeq™). The combined results demonstrated that the compounds induced inclusion of an exon derived from a sequence inside intron 49 of the *HTT* pre-mRNA (Fig. 2c–e; junction expression index (JEI) reduction of >25%; Supplementary Fig. 2c). This intronic sequence represents a compound-inducible pseudoexon (psiExon) which, when included, alters the coding sequence and introduces a premature translation termination codon, leading to the degradation of *HTT* mRNA through nonsense-mediated mRNA decay (NMD) (Fig. 2f), and reduction in HTT protein levels.

Analysis of this stop-codon psiExon shows that it has weak 5' and 3' splice sites (ss) (Supplementary Fig. 3). The 5'ss has a noncanonical GA dinucleotide at the –2 and –1 positions, differing from the canonical AG sequence. Splicing of exons with a GAGU 5'ss is inefficient (as is the case with *SMN2* exon 7<sup>14</sup>) and can be enhanced using splicing modifiers such as risdiplam<sup>9,15,16</sup> and other investigational compounds<sup>13,17,18</sup>. However, the inclusion of psiExons by these molecules had not been previously reported<sup>12,13,17</sup>.

**Global effects of splicing modification.** To determine the full extent of gene expression (mRNA-level) and splicing (exon-level) changes by HTT-lowering splicing modifiers, we performed RNA-Seq analysis in human SH-SY5Y cells treated with HTT-C2 (branaplam; a more potent, but structurally similar analogue of HTT-C1), or control (dimethyl sulphoxide; DMSO) (Supplementary Fig. 4a). At 24 nM (twice the inhibitory concentration<sub>50</sub> [IC<sub>50</sub>] of 12 nM), HTT-C2 downregulated the expression of wt *HTT* and a few other genes (Fig. 3a). After treatment with 100 nM HTT-C2, a concentration eightfold the IC<sub>50</sub>, additional mRNAs were downregulated, indicating a dose-dependent effect and a favourable selectivity window for wt *HTT* over other genes at the lower concentration (24 nM) (Supplementary data 1).

Splicing analysis demonstrated that, in addition to the *HTT* stop-codon psiExon49a inclusion, 165 and 215 splicing events were altered when treated with 24 nM and 100 nM HTT-C2, respectively (Supplementary data 2). The majority of splicing events resulted in inclusion (Inc) or skipping (Skp) of cassette exons (CE; exons either included or skipped to create distinct protein isoforms<sup>19</sup>) (Fig. 3b, c). Inc events were more common (1.4×–3.3×) than Skp events (Fig. 3d). Similar to *HTT* stop-codon psiExon49a, many of the Inc CEs (22% and 44% in 24 nM and 100 nM HTT-C2, respectively), had no annotations for at least one ss (Fig. 3e; Supplementary Fig. 4b), and had significant enrichment of –2, –1 GA, and –3 A, +3 A sequence at the 5'ss (represented by AGAguag) of exons activated by HTT-C2 (Fig. 4a, b, Supplementary data 2). These 31 psiExons are minimally included at basal level (Fig. 3f; median percent-spliced-in [PSI] index = 0.7%), poorly conserved (Fig. 3g), shorter (median size = 64 bps), and with significantly weaker 5'ss than annotated exons (Supplementary Fig. 5). Inclusion of these psiExons containing a stop codon or causing frameshift (NMD-psiExons) were correlated with downregulation of their host



**Fig. 1** Huntingtin (HTT)-lowering activity in vitro. **a** Chemical structures of HTT-C1 and HTT-D1. **b** Electrochemiluminescence (ECL) analysis of mutant HTT protein from fibroblasts isolated from a homozygous patient with Huntington's disease (HD) (GM04857) after 96 h of continuous treatment with HTT-C1 and HTT-D1 (0.01–1.0 μM). Representative graphs show percent HTT remaining relative to the dimethyl sulphoxide (DMSO) control. Cell viability assays were performed in parallel. Data represent mean of two ( $n = 2$ ) biologically independent samples per data point from one dose-response experiment. **c** Western blot of HTT protein and housekeeping proteins, oxidoreductase-protein disulphide isomerase (PDI), beta-actin, alpha serine/threonine-protein kinase (AKT) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in HD fibroblasts after 96 h of continuous treatment with HTT-C1 (0.015–1.0 μM). Utrrophin (UTRN) was also used as a loading control. The western blot data used a representative splicing modifier (tested at multiple concentrations) from a single experiment. Multiple splicing modifiers from the same class were tested and evaluated by western blot analyses. **d** Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis of *HTT* mRNA in patient fibroblasts after 24 h of treatment with HTT-C1 and HTT-D1 (0.01–1.0 μM). Representative graphs show percent *HTT* mRNA remaining relative to DMSO control; normalised to housekeeping gene, TATA-box binding protein. Data represent mean of two ( $n = 2$ ) biologically independent samples per data point from one dose-response experiment.

genes ( $P < 0.05$ , Fig. 3h, Supplementary data 1), suggesting an NMD mechanism similar to *HTT*.

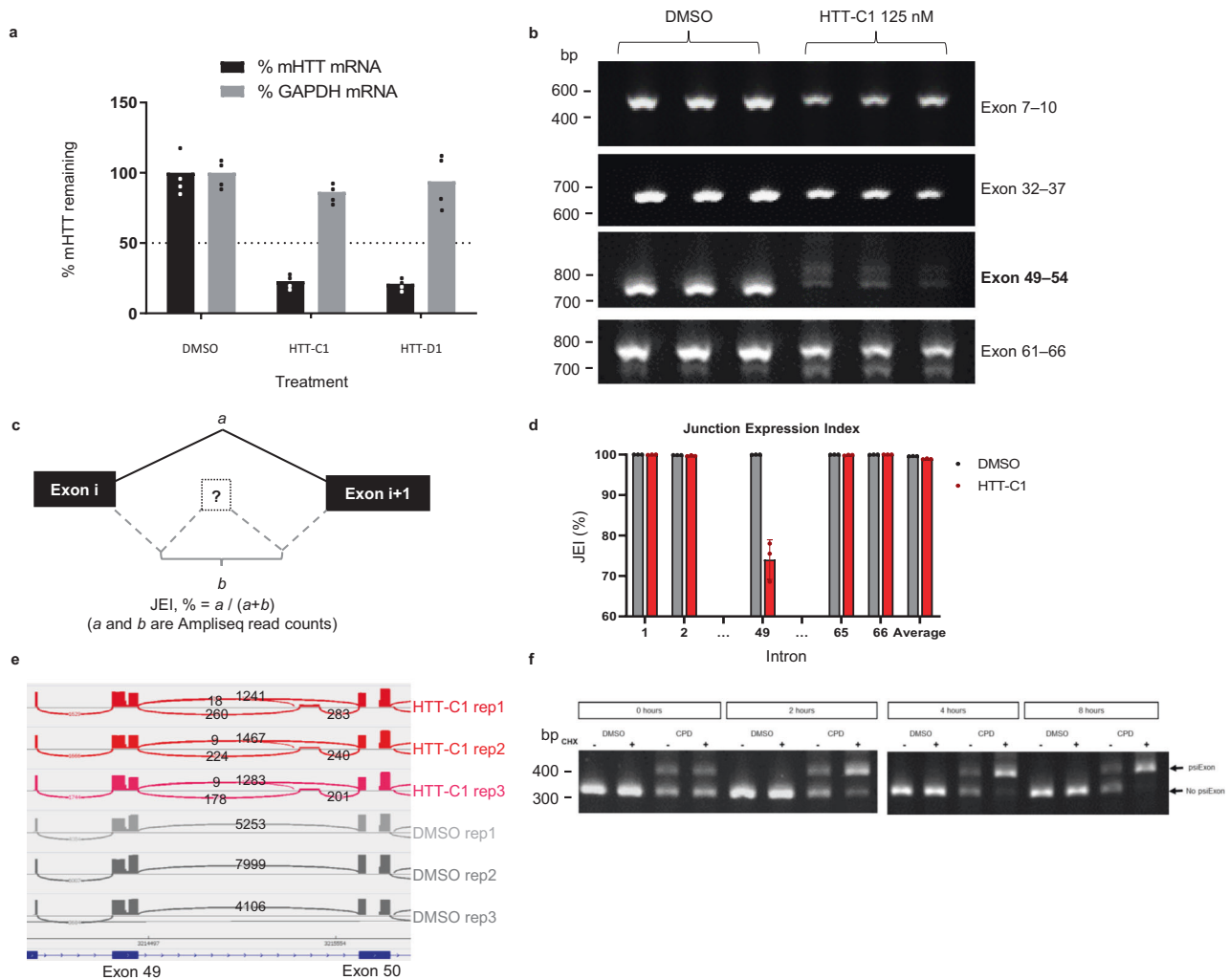
The strong preference for AGAguag 5' ss for HTT-lowering splicing modifiers prompted us to compare the effect of the *SMN* splicing modifier SMN-C3 (a close analogue of risdiplam) on mHTT lowering and global splicing changes. We observed that SMN-C3 demonstrated a striking 170-fold decrement in potency for mHTT as compared to HTT-C2, whereas the molecules show similar SMN EC<sub>1.5X</sub> (Supplementary Fig. 6). When splicing effects were analysed globally by examination of previously published RNA-Seq data<sup>12</sup>, treatment with SMN-C3 led to the generation of 33 psiExons with the predominant consensus of ADGAguaag (D = U, A or G) (Fig. 4a, b, Supplementary data 2), consistent with the inclusion of *SMN* exon 7 with an AGGAguaag 5' ss. Together, these data demonstrate that HTT-C2 and SMN-C3 represent two distinct classes of splicing modifiers that target 5' ss with the noncanonical GA dinucleotide at position -2, -1 of the 5' ss, with the distinction being a preference for adenosine at either -3 for *HTT*-selective molecules or -4 for *SMN*-selective molecules.

The strong structure-based sequence selectivity between the two classes of splicing modifiers for *HTT* and *SMN* splicing supports previous data proposing that these molecules bind a specific RNA interface created by the noncanonical GAgU 5' ss in complex with U1<sup>13,18</sup>. To further validate that the stabilisation of the U1/pre-mRNA complex leads to the induced inclusion of the psiExons identified through our small molecule treatment, we transfected human embryonic kidney 293 (HEK293) cells with a

variant U1 small nuclear RNA (snRNA) (U1-GA variant). This variant is re-programmed to perfectly bind to a noncanonical GAgU 5' ss and was shown to activate the wt *HTT* psiExon49a and 23 other psiExons activated by 100 nM HTT-C2 (Supplementary Fig. 7a, b). Interestingly, 582 additional psiExons were activated only by the U1-GA variant (Supplementary Fig. 7b, Supplementary data 3) and were enriched for GA at the -2 to -1 position of 5' ss, with no preference for adenosine at the -4 or -3 positions (Supplementary Fig. 7a). These results further support the hypothesis that HTT-C2 functions to stabilise U1-5' ss interaction of a highly selectively subset (AGAguag 5' ss) of potential psiExons located throughout the human genome, although further experiments are needed to prove a direct interaction.

**Elements required for psiExon inclusion.** The importance of the AGAguag 5' ss sequence of *HTT* stop-codon psiExon49a and the species-specific activity of the compounds was confirmed in a series of minigene experiments using human and mouse wt *HTT* sequences, where alterations to the -2 and -1 sequences demonstrated requirements for GA and -2, -1 position (Fig. 4c, d, Supplementary Fig. 8a–d).

Although the RNA-Seq data identified a limited number of psiExon inclusion events occurring by either direct or compound-induced U1 recruitment ( $n = 609$ ; Supplementary Fig. 7b), we developed a bioinformatic algorithm (see Methods) and identified >58,000 potential psiExons with an AGAguag 5' ss throughout the human genome, including four within introns 1, 8 and 40 of



**Fig. 2** Splicing of human huntingtin (*HTT*) pre-mRNA resulting in lowering of *HTT* mRNA. **a** Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis of *HTT* mRNA in B-lymphocytes from the same patient (GM04856 cells) after 24 h of treatment with HTT-C1 and HTT-D1 (0.25  $\mu$ M). Representative graphs show percent *HTT* mRNA remaining relative to dimethyl sulphoxide (DMSO) control; normalised to housekeeping gene, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*). Data represent the mean of two ( $n = 2$ ) biologically independent samples per data point. **b** Reverse transcription-polymerase chain reaction (RT-PCR) analysis of *HTT* mRNA after 24-hour treatment with 125 nM HTT-C1 or DMSO in patient-derived B-lymphocytes (GM04856). The data are from a single experiment with three biologically independent samples per data point. The comprehensive data set is provided in Supplementary Fig. 2. **c** Diagram illustrating how the junction expression index (JEI) was calculated for each of the 66 introns in the *HTT* gene. **d** JEI of intron 49 and a selection of other introns. The JEI of intron 49 was significantly reduced, indicating a splicing event (>25% reduction;  $P < 0.05$ ). Average JEIs are shown as bars. Error bars represent standard deviation. Data were based on three biological replicates of next-generation sequencing (NGS) data. **e** Sashimi plot of alternative splicing (AS) within intron 49 of the *HTT* pre-mRNA using NGS data. A minimum threshold of five reads was used to visualise these data in the Integrative Genomics Viewer. **f** Endpoint polymerase chain reaction (EP-PCR) analysis of GM04856 cells treated with DMSO or 250 nM HTT-C1. After 18 hrs, cells were treated with 10  $\mu$ M cycloheximide or DMSO. Total RNA was isolated at 0, 2, 4 and 8 hrs. The data are from a single experiment with two biologically independent samples per data point.

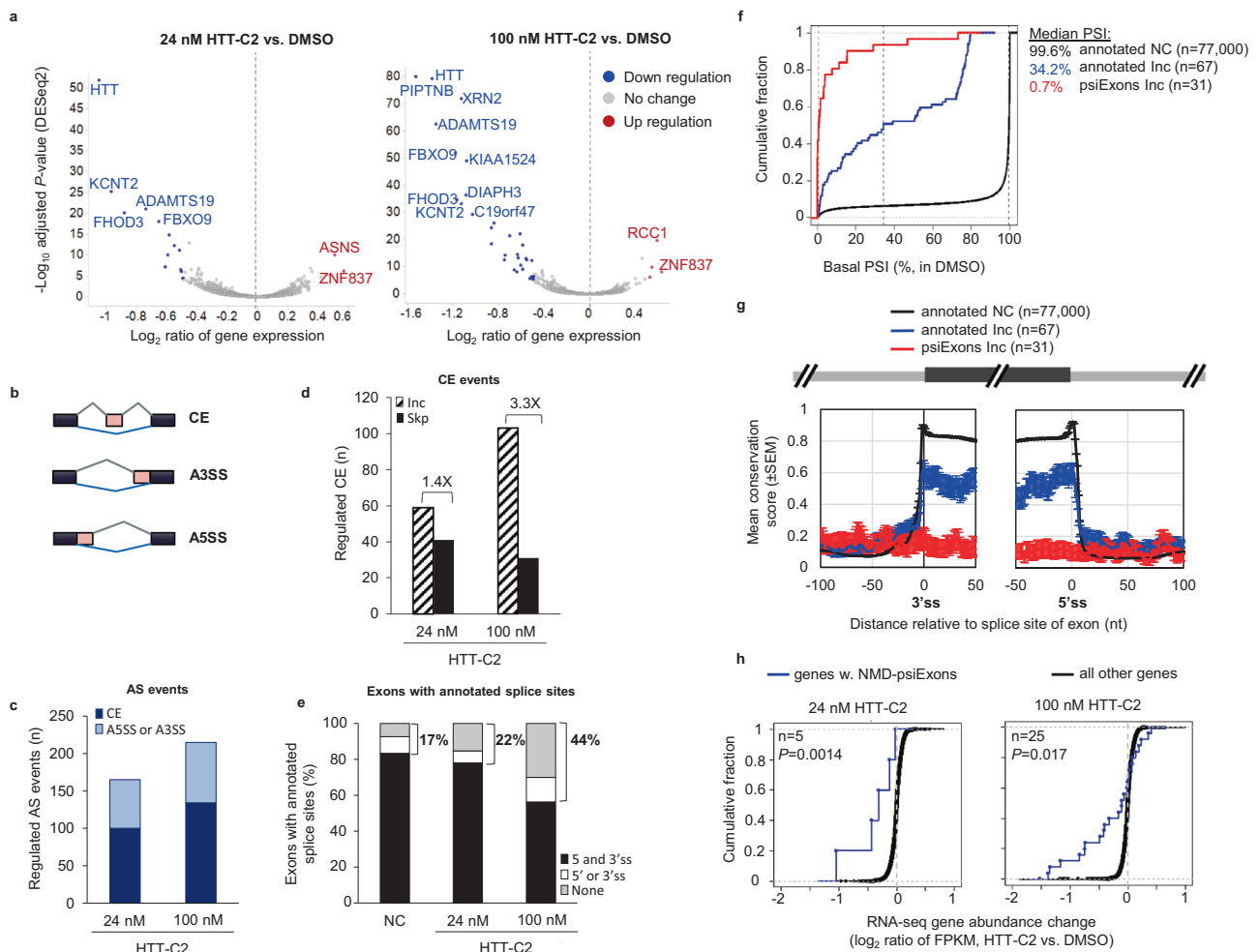
the *HTT* gene (Supplementary Fig. 8c). The lack of response of these potential psiExons to either HTT-C2 or the U1-GA variant suggests that additional sequence elements are required to promote psiExon inclusion. This is supported by the examination of splicing of minigenes in which the mutation of  $-2$ ,  $-1$  positions to canonical AG of psiExon49a resulted in its constitutive inclusion, while the same mutation in other potential psiExons had no effect (Supplementary Fig. 8d).

Bioinformatic analysis of *HTT* stop-codon psiExon49a identified a number of potential sequences that could function as exonic splicing enhancers (ESE) upstream of the 5'ss (Supplementary Fig. 8e). Deletion of 20 nucleotides ( $-38$  to  $-19$ ), and more precise partial deletion or mutation of nucleotides CAGGA at the  $-38$  to  $-34$  positions resulted in the total loss or reduced

compound-induced splicing (Fig. 4e, Supplementary Fig. 8f, g), suggesting the functional importance of this region.

**HTT-C2 elicits mHTT lowering in HD mice.** As our splicing modifiers require the presence of the human *HTT* intron 49, we studied the pharmacodynamic effects of HTT-C2 in the BACHD mouse model that expresses a full-length human mHTT gene<sup>20</sup>. Compound HTT-C2 was evaluated due to its superior exposure relative to HTT-C1 (Fig. 5a). Daily oral HTT-C2 reduced mHTT levels within brain tissue of mice in a dose-dependent manner (Fig. 5b). Maximal reductions of mHTT levels were achieved by Day 21 of treatment, with no further reduction observed (Fig. 5c). These effects were reversible, as protein expression levels returned





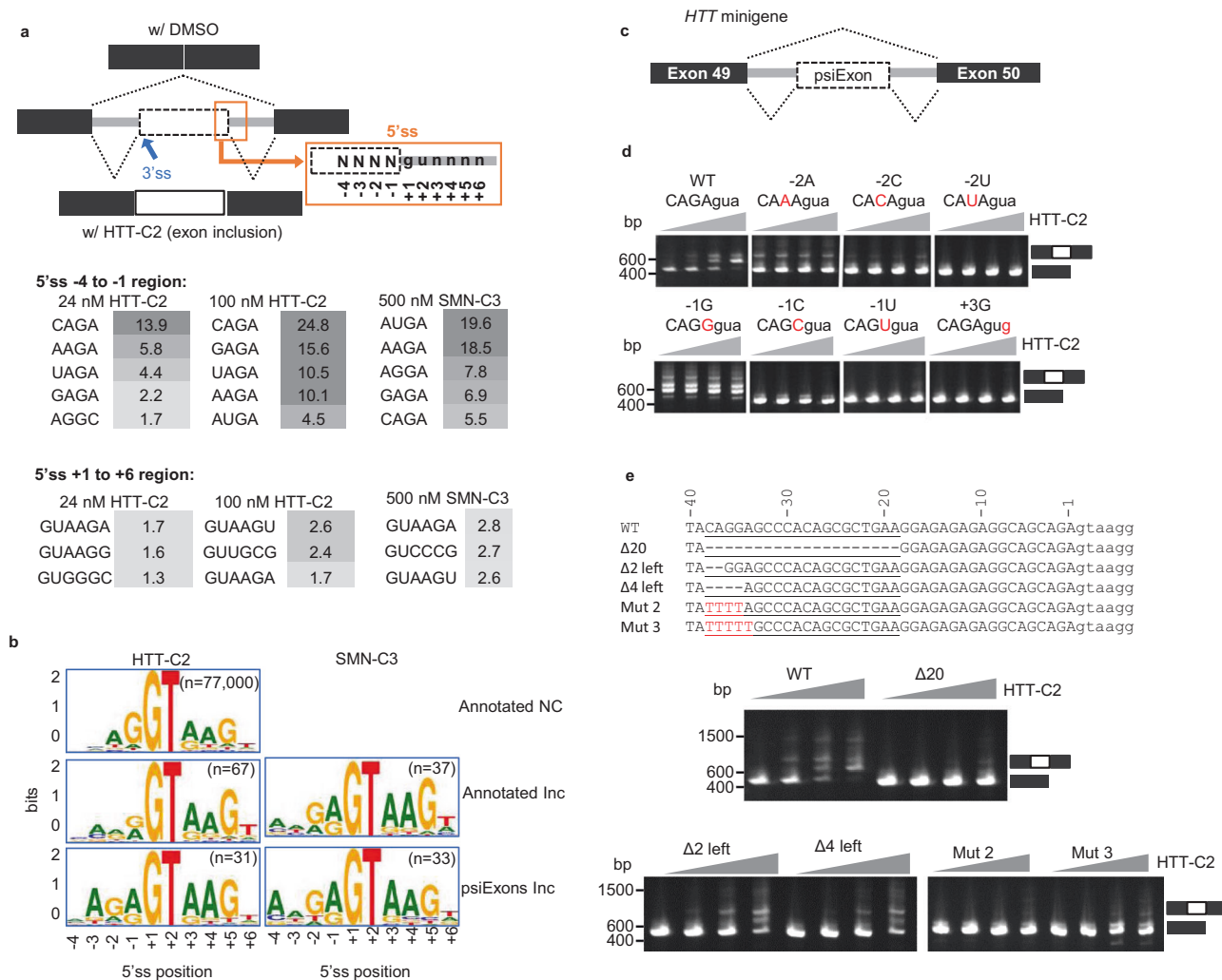
**Fig. 3 Selectivity of compound-induced splicing.** **a** Volcano plot of RNA-Seq analysis comparing gene expression in SH-SY5Y cells treated with either 24 nM or 100 nM of HTT-C2 with dimethyl sulphoxide (DMSO) treatment. mRNAs with significant changes in expression (>1.5-fold, false discovery rate (FDR) < 5%) are shown as blue and red dots for down- and upregulation, respectively. **b** A schematic of alternative splicing (AS) events. CE, cassette exon; A3SS, alternative 3' splice sites (ss); A5SS, alternative 5'ss. **c** Number of regulated AS events in SH-SY5Y RNA-Seq data following treatment with 24 nM and 100 nM HTT-C2. **d** Number of CEs inclusion (Inc) or skipping (Skp) after HTT-C2 treatment; ratio of Inc/Skp are shown in text. **e** Percentage of exons with 3' and 5'ss annotated by public databases (Refseq, Ensembl, or UCSC Known Genes) for NC (no change) or Inc exons. **f** Cumulative distribution function (CDF) curves of basal percent-spliced-in (PSI) index (average PSI in DMSO samples). The graph shows data for exons separated into three groups; Inc is based on  $\Delta$ PSI > 20% and two-sided Fisher's Exact Test  $P < 0.001$  in any one of the two conditions (24 nM or 100 nM HTT-C2 vs. DMSO). Median values are shown as dashed vertical lines for each group. **g** Sequence conservation of 3'ss and 5'ss region. Conservation is based on phastCons score for 46-way placental mammals. Mean (standard error of mean [SEM]) conservation scores are shown. **h** CDF curves of RNA-Seq mRNA abundance change for genes with predicted nonsense-mediated mRNA decay (NMD)-psiExons. NMD-psiExons are psiExons whose inclusion in mRNA introduces a premature termination codon or causes frameshift or both, and are included (Inc) following HTT-C2 treatment. Number of genes (n) and  $P$  value are indicated.  $P$  value is based on comparison with "all other genes" group using Wilcoxon rank-sum test (two-sided).

to control levels within 10 days of treatment cessation (Fig. 5d). Uniform >50% mHTT protein-level lowering was achieved throughout the brain following HTT-C2 treatment, most importantly within the striatum and cortex (Fig. 5e).

**Optimisation of HTT splicing modifiers.** Our splicing modifiers target products of both *HTT* alleles by promoting *HTT* stop-codon psiExon49a inclusion that leads to similar reduction of wt and mHTT levels (Supplementary Fig. 1e). Research suggests that 50% reduction of mHTT would be beneficial to HD patients, and a 50% global reduction of wt HTT would be well tolerated<sup>21,22</sup>, which highlights the need for therapies that lower HTT in the brain while not depleting HTT in the periphery. However, HTT-C2 treatment had a far greater mHTT-lowering effect in peripheral tissues (>90%) than in the brain (Supplementary Fig. 9a).

As P-glycoprotein (P-gp) is one major transport protein expressed on BBB, which limits the entry of various drugs into the CNS<sup>23</sup>, HTT-C2 and other compounds were tested in an in vitro assay using Madin-Darby canine kidney cells over-expressing human P-gp (Supplementary Fig. 9c). HTT-C2 was determined to be a P-gp substrate. Subsequently, a lower unbound concentration of HTT-C2 was confirmed in the mouse brain compared with plasma by measuring the unbound brain partition coefficient ( $K_{p,uu}$ ) (Fig. 5h). This suggests that reduced mHTT lowering in the brain versus the periphery was most likely attributed to P-gp efflux.

To achieve similar lowering between the periphery and brain, chemical optimisation led to HTT-D3, an equally potent compound with reduced P-gp efflux (Fig. 5h). Oral administration of HTT-D3 resulted in dose-dependent and approximately



**Fig. 4 Regulation of compound-induced splicing for HTT-C2 and SMN-C3. a** 5' splice sites (ss) sequence at regions (−4 to −1 and +1 to +6) were studied for the enrichment of Inc (included) vs. NC (no change) exons. Significance scores are shown. Wide boxes represent exons. **b** Schematic of 5'ss sequence logo in the three exon groups: annotated NC, annotated (inclusion) Inc and psiExons Inc. **c** Diagram illustrating the design of huntingtin (*HTT*) minigene constructs for studying compound-induced splicing. **d** Polymerase chain reaction (PCR) analysis of RNA extracts from HEK293 cells transfected with wild-type (wt) human *HTT* minigene or constructs with point mutations in the −2 to +3 region of the 5'ss; cells were treated with dimethyl sulphoxide (DMSO) or HTT-C2 (0.010–1 μM). The data are from a single transfection experiment with multiple concentrations tested for a given construct; the “WT” control construct has been used multiple times ( $n > 3$ ). **e** Sequence of the 20-nucleotide region upstream of the 5'ss of *HTT* stop-codon psiExon49a showing partial deletions and mutations performed on this region, and their effects on HTT-C2 induced splicing (lower panel). Through partial deletion or mutation of the nucleotides CAGGA at positions −38 to −34, this region was shown to be important in regulating splicing events. The data are from a single transfection experiment with multiple concentrations tested for a given construct; the “WT” control construct has been used multiple times ( $n > 3$ ).

equivalent mHTT protein lowering in both brain and peripheral tissues in two mouse models carrying human mHTT transgene, BACHD<sup>20</sup> and Hu97/18<sup>24</sup> (Fig. 5f). In BACHD mice, we also assessed HTT-D3 activity on mHTT splicing. BACHD mice were given either 1 dose (single dosing) or 21 consecutive daily doses (multiple dosing) of 10 mg/kg HTT-D3. We observed a similar response in mHTT mRNA and protein levels in BACHD brain after compound treatment (Supplementary Fig. 9d). In the Hu97/18 model, we observed uniform mHTT protein reduction in two critical brain sections, the striatum and cortex (Fig. 5f). Brain and cerebrospinal fluid (CSF) mHTT levels were significantly correlated (Fig. 5g). A similar correlation was observed between plasma and CSF mHTT levels upon HTT-D3 treatment (Fig. 5g). This underscores the advantage of advancing molecules with reduced efflux, such as HTT-D3, as potential HTT-lowering therapeutics for HD, where uniform HTT lowering is observed throughout the body.

## Discussion

HD is a devastating disorder with no approved treatment. Several preclinical studies support the hypothesis that targeting the expression of mHTT may prevent and/or slow disease progression<sup>25–29</sup>. To our knowledge, this work represents the first example of the identification and optimisation of orally bioavailable splicing modifiers that penetrate all tissues (including every cell type throughout the brain and periphery) and exert their action on mHTT splicing and protein lowering evenly throughout the body. A molecule from our HD drug discovery programme has recently entered a phase 1 clinical trial.

We initially discovered molecules such as HTT-C2 that lowered mHTT levels in the CNS. However, 50% lowering in the brain resulted in near depletion of mHTT (>90%) in peripheral tissues. Ablation of wt HTT in mice has been reported to cause acute pancreatitis due to degeneration of pancreatic acinar cells<sup>22</sup>. We undertook a chemical optimisation campaign that led to the



in gene expression. Together with our results highlight the critical recognition step in pre-mRNA splicing and how it can be manipulated by small molecules. The requirement for a predicted ESE sequence within *HTT* stop-codon psiExon49a demonstrates the importance of exon definition in the activation of potential compound-induced psiExons. As has been shown for exonisation of Alu elements, mutation to strong 5' and 3' ss is not sufficient for exon recognition by the spliceosome; additional factors are required, such as serine/arginine-rich protein-binding sites and ESEs established over thousands to millions of years<sup>31</sup>. The small number of psiExon inclusion events, identified by U1-GA variant and compound treatment demonstrates this requirement for exon definition and highlights the high specificity of our *HTT*-selective compounds, which activate only 31 of >58,000 potential psiExons.

Finally, through comparison of RNA-Seq from SMN-C3 to that of *HTT*-C2, we identify a strong preference for A at the -4 and -3 positions of the 5' ss, respectively, defining two distinct classes of splicing modifiers that target the GAgu noncanonical 5' ss. Since ~45% of known human exons contain a noncanonical 5' ss in the -2, -1 position (non-AG; Supplementary Fig. 10), including GA (1.3% of total Refseq defined exons), our results suggest that additional noncanonical 5' ss could potentially be targeted by small-molecule splicing modifiers that remain to be discovered. The insights from our work can lead to the discovery of therapeutics for diseases of unmet medical need.

## Methods

**Cell culture.** Human B-lymphocytes and fibroblasts derived from the same homozygous patient with Huntington's disease (HD) (GM04856/GM04857) and a healthy donor (GM07492/GM07491) (Coriell Institute for Medical Research), human neuroblastoma (SH-SY5Y) cells (ATCC<sup>®</sup>), human embryonic kidney 293 (HEK293) cells (ATCC); Madin-Darby Canine Kidney (MDCK) cells (ATCC<sup>®</sup>); MDCK cells expressing multidrug-resistance mutation 1 (MDCK-MDR1) (Absorption Systems); mouse CT26 cells (ATCC) were all grown at 37 °C in a humidified 5% CO<sub>2</sub> atmosphere. Fibroblasts were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with 10% (v/v) fetal bovine serum (FBS) (Thermo Fisher Scientific) and 1% antibiotic cocktail (penicillin-streptomycin/Thermo Fisher Scientific). All cell lines tested negative for mycoplasma contamination. Following the purchase of the cell lines, none of the cell lines was further authenticated. No commonly misidentified cell lines were used in the study.

**High-throughput screening for the identification of *HTT*-lowering molecules.** Human fibroblasts derived from a homozygous patient with HD (GM4857) were grown for 96 h in the presence of test compounds (in 0.5% DMSO) or controls at 37 °C in 5% CO<sub>2</sub>. After 96 h, cells were lysed and frozen. Huntingtin protein (*HTT*) levels were measured in lysates as described below. Compounds that decreased *HTT* levels relative to DMSO control were further tested in full dose-response.

**Quantification of *HTT* protein in cells.** For analysis in the ECL assays, test compounds were serially diluted with a threefold step in 100% DMSO (Sigma<sup>®</sup>) to generate a seven-point concentration-response curve. A solution of test compound (500 nL, 200× in DMSO) was added to each test well; the final concentration of DMSO was 0.5%. Fibroblasts were seeded in 96-well flat-bottomed plates (Thermo Fisher Scientific) at 4 × 10<sup>3</sup> cells/well in 100 μL of culture medium containing the test compound or DMSO vehicle control and incubated for 96 h (37 °C, 5% CO<sub>2</sub>, 100% relative humidity). After removal of the supernatant, cells were lysed in 50 μL of 1× Lysis Buffer11 (LB11) extraction buffer (50 mM Tris [pH 7.4], 300 mM NaCl, 10% [w/v] glycerol, 3 mM ethylenediaminetetraacetic acid (EDTA), 1 mM MgCl<sub>2</sub>, 20 mM glycerophosphate, 25 mM NaF, 1% Triton X-100), containing a Complete<sup>™</sup> protease inhibitor cocktail (Roche Diagnostics<sup>™</sup>) with shaking at 4 °C for 30 min; the plates were then stored at -20 °C.

For western blot analysis, fibroblasts were plated at 5 × 10<sup>4</sup> cells/well in 1 mL 10% FBS/DMEM with GlutaMAX<sup>™</sup> supplement (Thermo Fisher Scientific) in 24-well plates (Thermo Fisher Scientific) and incubated for 3–4 h (37 °C, 5% CO<sub>2</sub>, 100% relative humidity). Cells were then treated with *HTT*-C1 at different concentrations (0.5% DMSO) in triplicate wells for 96 h. Cells were lysed in 75 μL Laemmli buffer (Bio-Rad Laboratories, Inc.).

**ECL protein assay.** MESO SCALE DISCOVERY<sup>®</sup> 96-well plates (MSD<sup>®</sup>) were coated overnight at 4 °C with primary antibodies in phosphate-buffered saline (PBS; 30 μL/well). The plates were washed three times with 0.05% Tween-20 in 1× PBS (PBS-T; 200 μL/well) then blocked (100 μL/well; 5% bovine serum albumin (BSA) in PBS-T) for 5–6 h at room temperature with shaking. Plates were then

washed three times with PBS-T. Cell lysates were transferred to the antibody-coated plates (25 μL/well) and incubated with shaking overnight at 4 °C. After removal of the lysates, the plates were washed three times with PBS-T, and 25 μL of detection antibody in 1% BSA, PBS-T was added to each well and incubated with shaking for several hours at room temperature. After three washes with PBS-T, 25 μL of SULFO-TAG secondary antibody (MSD; 0.25 μg/ml in 1% BSA, PBS-T) was added to each well and incubated with shaking for 1 h at room temperature. After washing three times with PBS-T, 150 μL of read buffer T with surfactant (MSD) was added to each empty well and the plate was imaged on the SI 6000 imager (MSD) according to manufacturers' instructions for 96-well plates. Primary capture antibodies included: anti-polyglutamine-expanded *HTT* mouse monoclonal antibody (mAb) at 1 μg/mL (clone MW1; Developmental Studies Hybridoma Bank); anti-*HTT* mAb at 1 μg/mL (clone 1HU-4C8; Millipore; catalogue # MAB2166); anti-Kirsten rat sarcoma viral oncogene homologue (KRAS) rabbit polyclonal antibody at 1 μg/mL (Abcam; catalogue # ab137739). Detection antibodies included: Huntingtin (D7F7) XP<sup>®</sup> Rabbit mAb at 0.25 μg/mL (Cell Signalling Technology<sup>®</sup>; catalogue # 5656); anti-human-KRAS mAb at 0.25 μg/mL (clone 2C1; LSBio; catalogue # LS-C175665-100).

**RT-qPCR quantification of *HTT* mRNA in cells.** Test compounds were serially diluted threefold in 100% DMSO to generate a seven-point concentration curve. A solution of test compound (500 nL, 200× in DMSO) was added to each test well. Fibroblasts were seeded in 96-well flat-bottomed plates (Thermo Fisher Scientific) at 1 × 10<sup>4</sup> cells/well in 100 μL of culture medium containing the test compound or DMSO vehicle control and incubated for 24 h (37 °C, 5% CO<sub>2</sub>, 100% relative humidity). After removal of the supernatant, cells were lysed in RNA lysis buffer (1 M Tris-HCl pH 7.4, 5 M NaCl, 10% IGEPAL<sup>™</sup>CA-630; 50 μL/well) for 1 min at room temperature, before 50 μL of chilled nuclease-free water was added to each well; plates were then transferred immediately onto the ice before storing at -80 °C overnight.

The mRNA levels of the huntingtin gene (*HTT*) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) were quantified using TaqMan-based RT-qPCR primers and probes (Thermo Fisher Scientific; Supplementary Table 1) and the AgPath-ID<sup>™</sup> one-step RT-PCR Kit (Thermo Fisher Scientific). RNA samples were transferred (2 μL/well) to the Armadillo 384-well PCR plate (Thermo Fisher Scientific) containing 8 μL/well of the AgPath-ID<sup>™</sup> one-step RT-PCR reaction mixture (Thermo Fisher Scientific) in a final volume of 20 μL. The plate was then sealed with MicroAmp<sup>™</sup> Optical Adhesive Film (Thermo Fisher Scientific) and placed in the CFX384 Touch<sup>™</sup> Real-Time PCR thermocycler (Bio-Rad Laboratories, Inc.). RT-qPCR was carried out at the following temperatures for indicated times: Step 1: 48 °C (30 min); Step 2: 95 °C (10 min); Step 3: 95 °C (15 s); Step 4: 60 °C (1 min); then Steps 3 and 4 were repeated for a total of 40 cycles.

**Western blot analysis.** HD fibroblasts (GM04857) were treated with compounds and lysed in 70 μL sample buffer (24-well plates/Thermo Fisher Scientific). Medium was aspirated, rinsed once with PBS, 70 μL Invitrogen<sup>™</sup> NuPAGE<sup>™</sup> Sample Buffer was added, rocked at room temperature for 10 min, and lysates frozen. Cell lysates were boiled for 10 min—loaded 45 μL/well. Cell lysates (45 μL/lane) were separated using pre-cast 3–8% Tris-Acetate gels (NuPAGE<sup>™</sup> 3–8%, Tris-Acetate, 1.0 mm, Midi Protein Gel, 12 + 2-well; Thermo Fisher Scientific) for 5–6 h at 130 V. After electrophoresis, proteins were transferred to nitrocellulose membranes (0.45 μM nitrocellulose/Bio-Rad) at 150 mA in NuPAGE Transfer Buffer (20× diluted to 1×; Thermo Fisher Scientific) for 90 min at 4 °C. Membranes were blocked overnight in blocking buffer (Li-Cor blocking buffer + 0.1% Tween) at 4 °C, washed four times (for 5 min) in PBS + 0.1% Tween (PBS-T; Thermo Fisher Scientific), and incubated overnight at 4 °C with primary antibodies (in Li-Cor blocking buffer + 0.1% Tween). Blots were washed with PBS-T, probed with secondary antibodies (in Li-Cor blocking buffer + 0.1% Tween) at room temperature for 1 h and washed again with PBS-T. Bound antibodies were visualised using the Odyssey<sup>®</sup> imaging system (LI-COR) according to the manufacturer's instructions. The following primary antibodies were used: anti-*HTT* at 1:1000 (clone 1HU-4C8; Millipore; catalogue # MAB2166), anti-utrophin (UTRN) at 1:250 (clone DRP3/20C5; Vector Laboratories; catalogue # VP-U579), anti-oxidoreductase-protein disulphide isomerase (PDI) at 1:10,000 (Santa Cruz; catalogue # SC20132), anti-β-actin at 1:10,000 (clone AC-74; Sigma; catalogue # A2228), anti-GAPDH at 1:1000 (Thermo Fisher Scientific; catalogue # PA1-987), anti-alpha serine/threonine-protein kinase (AKT) at 1:1000 (Cell Signalling; catalogue # 9272). Secondary antibodies included: Alexa Fluor<sup>®</sup> 680 goat anti-mouse immunoglobulin G (IgG) at 1:10,000 (Thermo Fisher Scientific; catalogue # A28183), IRDye<sup>®</sup> 800CW donkey anti-mouse IgG at 1:10,000 (Li-Cor; catalogue # 926-32212), and IRDye<sup>®</sup> 800CW donkey anti-rabbit IgG at 1:10,000 (Li-Cor; catalogue # 925-32213).

## Primer walking assay, endpoint RT-PCR and AmpliSeq analysis.

B-lymphocytes (GM04856 cells) were plated in six-well plates (ThermoFisher) at 5 × 10<sup>5</sup> cells/well in 2 mL of 10% FBS, DMEM and incubated for 6 h (37 °C, 5% CO<sub>2</sub>, 100% relative humidity). Cells were then treated with *HTT*-C1 at 125 nM (in 0.5% DMSO) in triplicate for 24 h. RNA was purified with the RNeasy Mini Kit (Qiagen) using the manufacturer's protocol. Samples were prepared for RT-PCR



(described previously in Methods) using 0.04  $\mu\text{L}$  of each primer (at 100  $\mu\text{M}$ ). For reverse transcription and PCR, the following steps were performed: reverse transcription step: 48 °C (15 min); PCR steps: Step 1: 95 °C (10 min), Step 2: 95 °C (30 s), Step 3: 55 °C (30 s), Step 4: 68 °C (1 min); Steps 2–4 were repeated for 34 cycles, then held at 4 °C. PCR products were separated on 2% pre-cast agarose E-gels (Invitrogen), stained with ethidium bromide and visualised using a UVP gel imager (Thermo Fisher Scientific). Primer sets used for primer walking are provided in Supplementary Table 2.

Ion AmpliSeq analysis of *HTT* pre-mRNA splicing GM04856 cells were plated in six-well plates at  $5 \times 10^5$  cells/well in 2 mL of 10% FBS, DMEM and incubated for 6 h (37 °C, 5%  $\text{CO}_2$ , 100% relative humidity) before treatment with 125 nM HTT-C1 (in 0.5% DMSO) in triplicate for 24 h. Cellular RNA was then extracted and purified (described previously in Methods). The Ion AmpliSeq technology (Thermo Fisher Scientific), which is a PCR-based target enrichment and next-generation sequencing platform, was used as a targeted measure of exons across the entire *HTT* transcript, with the goal of monitoring the presence of any (and all) novel splice variants in the presence of compound. PCR enrichment of *HTT* exon targets was accomplished by applying a custom *HTT* AmpliSeq panel (PTC proprietary primer set). The panel consisted of two separate PCR primer pools, each producing 33 amplicons. The complete *HTT* assay has 66 amplicons (mean size, 135 bp) covering all 67 exons. The AmpliSeq workflow (Supplementary Fig. 11) included: (a) RNA reverse transcription, (b) target amplification, (c) partial primer digestion, (d) adapter ligation, (e) library amplification, (f) sequencing reaction, (g) sequencing data analysis. For data analysis, AmpliSeq reads (Fastq format) were mapped to human genome (hg19) using TopHat<sup>23</sup> which allows identification of both known and novel splice junctions. For each one of the 66 introns of the *HTT* gene, we calculated a JEI (Fig. 2c) using the percent of read supporting the splicing of the exact annotated intron among all reads supporting the splicing isoforms using either the 5' splice or/and 3' splice of that intron. A JEI value of 100% indicates full splicing of the intron. A JEI value <100% indicates alternative splicing paths (eg. inclusion of a cryptic exon or use of alternative 5' or 3' splice). Biological triplicates were performed for each treatment group. We compared the difference of JEIs between compound and DMSO treated samples using the Student's *t* test.

**Splice-site score.** The 5' and 3' splice MAXENT scores were calculated using MaxEntScan<sup>33</sup> ([http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan\\_scoreseq.html](http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html)) representing the strength of splice sites.

**Analysis of nonsense-mediated decay.** GM04856 cells were treated with DMSO or 250 nM HTT-C1. After 18 h, cells were treated with 10  $\mu\text{M}$  cycloheximide (Sigma) or DMSO. Total RNA was isolated after 2, 4 and 8 h and analysed by endpoint PCR (described previously in Methods).

**Transfection.** Wild-type and mHTT and U1 minigene constructs were designed at PTC and synthesised at GenScript<sup>®</sup>. For U1 constructs  $5 \times 10^5$  HEK293 cells were transfected with 2  $\mu\text{g}$  of plasmid deoxyribonucleic acid (DNA) or mock control in six-well plates, using 6  $\mu\text{L}$  Fugene6<sup>®</sup> (Promega) according to the manufacturer's instructions; after incubating for 24 h (37 °C, 5%  $\text{CO}_2$ , 100% relative humidity), cells were treated with either 1  $\mu\text{M}$  HTT-C1 or 0.5% DMSO control and incubated for 48 h. For *HTT* constructs,  $5 \times 10^5$  HEK293 cells were transfected with 50 ng of plasmid DNA in 24-well plates, using 6  $\mu\text{L}$  Fugene6<sup>®</sup> according to the manufacturer's instructions. After incubating overnight (37 °C, 5%  $\text{CO}_2$ , 100% relative humidity), cells were treated with varying concentrations of compounds in a final concentration 0.05% DMSO and incubated for 24 h.

#### RNA-Seq library preparation from SHY5Y and U1 transfected HEK293.

SHY5Y cells were seeded in six-well plates at  $6 \times 10^5$  cells/well in 2 mL 10% FBS, DMEM and incubated for 4 h. Cells were then treated with two biological replicates of HTT-C1 at 24 nM or 100 nM (in 0.1% DMSO), or four biological replicates of vehicle control (DMSO) for 24 h (37 °C, 5%  $\text{CO}_2$ , 100% relative humidity). HEK293 cells were transfected with U1-GA variant minigene construct or mock for 48 h (37 °C, 5%  $\text{CO}_2$ , 100% relative humidity).

Total RNA was extracted using the RNeasy Plus Mini Kit. RNA concentration and quality were assessed using a NanoDrop spectrophotometer (ThermoFisher Scientific). For library preparation and sequencing, mRNA was enriched from  $\sim 3 \mu\text{g}$  of total RNA using oligo(dT) beads. The mRNA was fragmented randomly using fragmentation buffer followed by complementary DNA (cDNA) synthesis using an mRNA template and random hexamers primer. Second-strand synthesis buffer (Illumina), deoxynucleotides, ribonuclease H and DNA polymerase I were added to initiate second-strand synthesis. After a series of terminal repair, A-ligation and sequencing adaptor ligation, the double-stranded cDNA library was completed through size selection and PCR enrichment. RNA libraries were sequenced in a HiSeq sequencer (Illumina).

**RNA-Seq analysis of pre-mRNA splicing.** RNA sequencing reads were mapped to human genome (hg19) using STAR (version 2.5)<sup>34</sup>; only uniquely mapped reads (with MAPQ > 10) with <5nt/100nt mismatches and properly paired reads were used. For gene expression analysis, the number of reads in the coding sequence

(CDS) region of protein-coding genes and exonic region of non-coding genes were counted and analysed using DESeq2<sup>35</sup> (Bioconductor). For splicing analysis, all junction reads (read with a gap in alignment indicating splicing) were used, including the ones mapped to unannotated splice sites. Reads were counted for different exons (for cassette exon [CE]) or exonic regions (for A5'ss or A3'ss). For each splicing event, a percent-spliced-in (PSI) value was calculated using the percent of average read number supporting the inclusion among all reads supporting either the inclusion or the exclusion. A minimum of 20 for the denominator of PSI calculation was required. Otherwise, a 'NA' value would be generated. PSI values for biological replicates were averaged, and the PSI difference between the two treatment groups was calculated. For a statistical test, a  $2 \times 2$  read counts table was made for each event, with rows for reads supporting inclusion or exclusion, and columns for the two comparing sample groups (biological replicates were combined). Fisher's exact test was used for the statistical tests. PSI change of >20% (or < -20%) and  $P < 0.001$  was used to select splicing events being regulated by the treatment. PSI calculation, Fisher's exact test, k-mer analysis and statistical analysis were performed using R (3.5.1).

**K-mer analysis.** For comparing sequence difference of a particular region for two groups of exons (e.g., Inc vs. NC) we compared the k-mer ( $k = 4-6$ ) frequencies of the two groups by Fisher's Exact Test (one k-mer vs. all other k-mers, Group 1 vs. Group 2). The resulting *P* value was converted to a significance score ( $[\text{SS}] = -5^{\text{P}} \log_{10} P$  value), in which *S* is the sign indicating enrichment (1) or depletion (-1) of the k-mer in Group 1.

**Sequence logo.** Sequence logos were generated using WebLogo<sup>36</sup> (University of California, Berkeley).

**Genome-wide identification of putative GA-psiExons.** We searched the human genome with "AGAgtaag" sequences (potential 5' splice sequence responding to the compound) within introns of ReSeq annotated genes. Then 3' splice sequences were scanned upstreaming of the "AGAgtaag" sequence using MaxEntScan<sup>33</sup>. A putative GA-psiExon is defined as an unannotated exon with length between 6 and 200nt and 3' splice score >2.3.

**Minigene constructs.** HEK293 cells transfected with minigene constructs were treated with varying concentrations of test compounds in a final concentration 0.05% DMSO and incubated for 24 h. Total RNA was isolated from the cells using the RNeasy Plus Mini Kit (Qiagen) and RNA concentration and quality were assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific). To determine splicing changes, cDNA was synthesised using the iScript<sup>™</sup> cDNA synthesis kit (Bio-Rad Laboratories) according to the manufacturer's instructions. Endpoint PCRs were set up using Platinum<sup>™</sup> PCR SuperMix High Fidelity (Invitrogen) and the resulting PCR products were separated on 2% E-gels (Invitrogen). Primers (designed by PTC Therapeutics, supplied by Invitrogen) were directed against common sequences in the minigene constructs: T7 Forward: 5'-TAATA CGACTCACTATAGGG-3'; BGH Reverse, 5'-TAGAAGGCACAGTCGAGG-3'.

**Animal studies.** All in-life animal procedures were performed in a laboratory certified by the American Association for the Accreditation of Laboratory Animal Care (AAALAC) with approval from the Institutional Care and Animal Use Committee. BACHD<sup>20</sup> and Hu97/18<sup>24</sup> mice were used, and the genotype of each BACHD animal was confirmed by an in-house PCR assay prior to enrolment in the study.

**Quantification of HTT protein in animal tissues.** Test mice were euthanised, and brain, muscle (quadriceps), other peripheral tissues and blood samples were harvested 2 h after the last dose on Day 20. Prior to analysis, crude total protein from brain and peripheral tissue samples were prepared by sample lysis in MSD assay buffer 1 (MSD) with Complete<sup>™</sup> Protease Inhibitor Cocktail added (Roche Diagnostics). Tissues were then homogenised using TissueLysor II (Qiagen), plus a 5 mm stainless steel bead. The lysate was clarified by centrifugation at  $16,000 \times g$  for 20 min at 4 °C, and the total protein concentration quantified with the Pierce<sup>™</sup> BCA Protein Assay Kit (Thermo Scientific), according to the manufacturer's instructions. Whole blood was collected by cardiac puncture into EDTA collection tubes. An aliquot (100–200  $\mu\text{L}$ ) was added to 1.5 mL of eBioscience<sup>™</sup> 1v Red Blood Cell Lysis Buffer (Thermo Fisher) and mixed well for 5 min before collecting the white blood cells (WBC) by centrifugation at  $400 \times g$ . The supernatant was discarded, and the pellet of WBCs was frozen in liquid nitrogen and stored at  $-70 \text{ }^\circ\text{C}$ . Brain and peripheral sample lysates were analysed for hHTT and KRAS protein expression using the ECL protein assay (described previously in Methods); using the same method WBC samples were also analysed for hHTT expression but not KRAS.

Each tissue sample was tested in duplicate using the ECL protein assay and the average hHTT and KRAS readouts were calculated. The ratio of the mean hHTT signal to the mean KRAS signal ( $\times 1000$ ) was determined for each test animal. The hHTT/KRAS ratio grand mean for the vehicle group of five test animals was calculated, and the fold change relative to the vehicle grand mean was determined

for each test animal in each group. Percent hHTT (%hHTT) lowering normalised to KRAS was determined for each test animal by subtracting the fold change from one and multiplying the difference by 100. Each experiment was performed twice yielding ten %hHTT-lowering values for each treatment group. For each treatment group, the mean %hHTT lowering plus the standard error of the mean was plotted as a bar graph. The %hHTT lowering in WBC samples was determined without KRAS using the grand hHTT vehicle mean, instead of the grand hHTT/KRAS ratio vehicle mean.

**In vivo pharmacokinetic studies.** Oral pharmacokinetics (PK) of compounds were evaluated in wt littermates from the BACHD colony (FVB background). Mice were treated with test compounds (10 mg/kg) by oral gavage in 0.5% hydroxypropylmethyl cellulose (HPMC) with 0.1% Tween 80. Blood was collected by terminal cardiac puncture at specified time points (three mice per time point) and centrifuged to generate plasma. Brain tissue was collected at the time of blood collection and homogenised in water. Protein was precipitated from plasma and brain homogenates with acetonitrile, methanol mixture (5:1, v/v) containing an internal standard that is a close analogue of the test compounds. The mixture was filtered through an EMD Millipore MultiScreen™ Solvint Filter Plate (MSRLN04, Millipore, Burlington, MA). Calibration standards were prepared in the same matrix and processed with the testing samples. Filtrates were analysed using an Acquity ultra performance liquid chromatography (UPLC) system (Waters Corporation) in tandem with Xevo TQ-S Spectrometer (Waters Corporation). Samples were injected on to a Waters UPLC Acquity BEH C18 Column (2.1 × 50 mm, 1.7 μm) maintained at 50 °C. The injection volume was 3 μL and the mobile phase flow rate was 0.45 mL/min. The mobile phase consisted of two solvents: a) 0.1% formic acid in water and b) 0.1% formic acid in acetonitrile. The initial mobile phase started with 5% solvent B for 0.4 min, which was changed to 98% solvent B over 0.8 min with linear gradients, and then maintained at 95% solvent B for another 0.4 min. The drug concentrations were acquired and processed with MassLynx 4.1 software. PK parameters were estimated using the non-compartment method within Phoenix® WinNonlin® Build 8.1 (Certara USA, Inc., Princeton, NJ).

**In vivo pharmacodynamic studies.** BACHD: pharmacodynamic (PD) evaluations were performed in BACHD mice aged 6–10 weeks. Compound or vehicle (HPMC/0.1% Tween 80) was administered to BACHD mice (five female mice per group) once daily for 21 doses (QD×21) by oral gavage; dosing volumes were 10 mL/kg. Each animal was regularly observed for mortality or signs of pain, distress or overt toxicity, and findings were recorded. Body weights were recorded at the start, and at least once a week, during the course of the study. Tissue samples were obtained and prepared for ECL protein assay analysis (described previously in Methods) from each animal.

Hu97/18: both sexes of 2–4-month-old Hu97/18 mice were used. Mice were maintained under a 12-hour light:12-hour dark cycle in a clean facility with free access to food and water. Experiments were performed with the approval of the Institute Animal Care and Use Committee of the University of Central Florida. Mice were treated with vehicle control or 2, 6 or 12 mg/kg of compound daily by oral gavage for 21 consecutive days. Mice were weighed 3x weekly and observed daily for general health and neurological signs, including gait, head tilt and circling. No adverse events were observed, and no mice were removed from the study.

**Hu97/18 terminal tissue and sample collection.** Mice were anaesthetised with Avertin (2,2,2-tribromoethanol, Sigma Aldrich, catalogue # T48402) and secured in a stereotaxic frame (Stoelting). The ear bars were raised and the nose piece used to position the mice in a manner that would allow for a near 90° tilt of the head to access the cisterna magna. A 1 cm<sup>2</sup> section of dorsal neck skin was removed, and muscle layers were completely dissected away to expose the cisterna magna, which was then cleaned with PBS and 70% ethanol and dried using compressed air. A 50cc Hamilton® syringe with point style 2 and a 12o bevel was then lowered carefully into the cisterna magna. CSF was slowly withdrawn at a rate of 10 μL/min using an UltraMicroPump with a Micro4 controller (World Precision Instruments). CSF samples were collected in pre-chilled tubes, centrifuged, then flash frozen in liquid nitrogen prior to storage at –80 °C.

Whole blood was then collected by cardiac puncture into EDTA-coated tubes and divided into three aliquots. One was immediately snap frozen, while plasma was isolated from another and crude peripheral blood mononuclear cells from the third. Mice were then decapitated, and the brain removed and placed in ice for ~1 min to increase tissue rigidity. During this interval, liver, heart and quadriceps muscle were isolated and snap frozen. Brains were then micro-dissected into cortex, hippocampus, striatum, cerebellum, and midbrain/brain stem.

**Immunoprecipitation and flow cytometry mtHTT quantification.** Approximately 10,000 5-μm carboxylate-modified latex beads (Invitrogen, catalogue # C37255) were coupled with capture antibody, HDB4E10 anti-HTT, in 50 μL of NP40 lysis buffer (150 mM NaCl, 50 mM Tris pH 7.4, Halt phosphatase (Thermo Scientific, catalogue # 78420) and Halt protease inhibitor cocktails (Thermo Scientific, catalogue # 78429), 2 mM sodium orthovanadate, 10 mM sodium fluoride NaF, 10 mM Iodoacetamide, and 1% NP40). Capture antibody coupled beads were then combined with 10 μL of CSF, or 20 μL of plasma in triplicate in a 96-well V-

bottom plate (Thermo Scientific, catalogue # 249944), brought to a total volume of 50 μL in NP40 lysis buffer, mixed well, and incubated overnight at 4 °C. The next day, the plate was spun down for 1 min at 650 RCF and supernatant was removed. Beads were washed three times in immunoprecipitation and flow cytometry (IP-FCM) wash buffer (100 mM NaCl, 50 mM Tris pH 7.4, 1% BSA, 0.01% sodium azide). MW1 anti-expanded polyglutamine probe antibody was biotinylated using EZ-Link Sulfo-NHS-Biotin (Thermo Scientific, catalogue # 21217), and 50 μL of the diluted antibody was incubated with the HDB4E10 beads bound to mtHTT for 2 h at 4 °C. Beads were washed three times with 200 μL of IP-FCM wash buffer. Streptavidin-phycoerythrin (PE) (BD Biosciences, catalogue # 554061) was prepared at 1:200 and 50 μL added to each well and incubated at RT, protected from light, for 30 min. Beads were washed three times with 200 μL of IP-FCM buffer, resuspended in 200 μL of IP-FCM wash buffer, and fluorescence intensity of ~2000 beads per sample, HDB4E10/MW1 mtHTT bead complexes, was measured using an Acuri C6 flow cytometer (BD Biosciences). Median fluorescent intensity of PE was measured for each sample to determine relative mtHTT protein levels.

**MDCK-MDR1 efflux assay.** The MDR1 efflux assay was conducted at Absorption System LLC (Exton, PA). In brief, MDCK-MDR1 and MDCK-wt cell monolayers were grown to confluence on collagen-coated, microporous membranes in 12-well assay plates (ThermoFisher). Compound solutions (10 μM) in permeability assay buffer (Hanks' balanced salt solution [HBSS], 10 mM HEPES, 15 mM glucose; pH of 7.4) were placed in the donor chamber. The receiver chamber was filled with assay buffer plus 1% BSA. Cell monolayers were dosed on the apical side (A-to-B) or basolateral side (B-to-A) and incubated at 37 °C (5% CO<sub>2</sub>, 100% relative humidity). Sampling from the donor chambers was performed at 0 and 1 hr; and from the receiver chambers at 1 hr. Each determination was performed in duplicate. The flux of lucifer yellow was also measured post-experimentally for each monolayer to ensure no damage was inflicted to the cell monolayers during the flux period. All samples were assayed by liquid chromatography-tandem mass spectrometry using electrospray ionisation. The apparent permeability ( $P_{app}$ ) and percent recovery was determined using the following equation:

$$P_{app} = (dC_r/dt) * V_r / (A * C_0)$$

$dC_r/dt$  represents the slope of the cumulative receiver concentration vs. time in μM/s;  $V_r$  is the volume of the receiver compartment (cm<sup>3</sup>);  $V_d$  is the volume of the donor compartment in (cm<sup>3</sup>);  $A$  is the area of the insert (1.13 cm<sup>2</sup> for 12-well);  $C_0$  is the average measured concentration of the donor chamber at time zero in μM; net efflux ratio is defined as  $P_{app(B-to-A)} - P_{app(A-to-B)}$ .

**Unbound brain partition coefficient ( $K_{p,uu}$ ).** The unbound brain partition coefficient ( $K_{p,uu}$ ) is defined as the ratio between unbound brain-free drug concentration and unbound plasma concentration. It was calculated using the following equation:

$$K_{p,uu} = C_{brain} * f_{u,b} / (C_{plasma} * f_{u,p})$$

$C_{brain}$  and  $C_{plasma}$  represent the compound concentrations in brain and plasma, respectively.  $f_{u,b}$  and  $f_{u,p}$  are the unbound fraction of each testing article in brain and plasma, respectively. Both  $f_{u,b}$  and  $f_{u,p}$  were determined in vitro using the Pierce device for rapid equilibrium dialysis at Absorption System LLC (Exton, PA).

$K_{p,uu}$  was calculated individually for each animal from multiple mouse PK studies and the average values are reported here.

**Quantification of human HTT mRNA in animal brain.** One group of transgenic BACHD mice in the study was orally administered a single dose of HTT-D3 at 10 mg/kg. One group (three mice) was administered vehicle alone on the same schedule. Dosing volumes were 10 mL/kg based on individual mouse weights. Dosing solutions were prepared once as the free base HTT-D3 dissolved in a vehicle comprising 0.5% HPMC and 0.1% TWEEN® 80 and stored at ambient temperature. Samples of brain tissues were obtained at 2, 4, and 8 h post HTT-D3 dosing. Samples were taken from the vehicle control mice at 2-h post vehicle dosing. Brain samples for each time point were obtained from three mice per group and total RNAs were prepared for analysis. Total RNAs from brain tissues were prepared following sample homogenisation and lysis in QIAzol Lysis Reagent using TissueLyser and the RNeasy Lipid Tissue Mini Kit (Qiagen #74804, Germantown, MD) according to the instructions provided in the manufacturer's kit.

Quantitative RT-qPCR was performed on ~50 ng of total RNA using AgPath-ID™ One-Step RT-qPCR Reagents (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions using the custom TaqMan gene expression assays detailed in Supplementary Table 3.

An RT-qPCR reaction mixture of primers and probe sets for *HTT* and *GAPDH* was prepared according to Supplementary Table 4.

A volume of 2 μL of the RNA preps was transferred from each well to the Armadillo 384-well PCR plate containing 8 μL/well of the RT-qPCR reaction mixture that was prepared as detailed in Supplementary Table 4. The plate was then sealed with MicroAmp™ Optical Adhesive Film and placed in the C1000 thermocycler. The RT-qPCR was carried out at the following temperatures for the

indicated times: Step 1: 48 °C (30 min); Step 2: 95 °C (10 min); Step 3: 95 °C (15 s); Step 4: 60 °C (1 min); then, repeated Steps 3 and 4 for a total of 40 cycles.

The amplification efficiency was calculated from the slope of the PCR amplification curve for *HTT* and *GAPDH* individually. The abundances of *HTT* mRNA and *GAPDH* mRNA were then calculated as  $(1 + E)^{-C_t}$ , where Cycle threshold ( $C_t$ ) is the threshold value for each amplicon. The abundance of *HTT* mRNA was normalised to *GAPDH* abundance. The normalised *HTT* mRNA level was then used to calculate the percent splicing in the HTT-D3-dosed group compared to the vehicle-treated group using a modified  $2^{-\Delta\Delta C_t}$  method<sup>37</sup>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The data supporting the findings of this study are available from the corresponding authors upon reasonable request. The AmpliSeq and RNA-Seq data generated in this study have been deposited in the GEO database under accession code [GSE162814](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE162814). The SMN-C3 RNA-Seq data was downloaded from the GEO database under accession code [GSE62540](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62540). The code for the RNA-Seq data analysis can be found in GitHub (<https://github.com/liwc01/DEDSeq>). The following databases were also used: Refseq (<https://www.ncbi.nlm.nih.gov/refseq/>), Ensembl (<https://useast.ensembl.org/index.html>) and UCSC Known Genes (<http://genome.ucsc.edu/>). Source data are provided with this paper.

Received: 17 March 2021; Accepted: 28 October 2021;

Published online: 15 December 2021

## References

- Wild, E. J. & Tabrizi, S. J. Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol.* **16**, 837–847 (2017).
- Tabrizi, S. J., Ghosh, R. & Leavitt, B. R. Huntingtin lowering strategies for disease modification in Huntington's disease. *Neuron* **101**, 801–819 (2019).
- Nopoulos, P. C. Huntington disease: a single-gene degenerative disorder of the striatum. *Dialogues Clin. Neurosci.* **18**, 91–98 (2016).
- Saudou, F. & Humbert, S. The biology of Huntingtin. *Neuron* **89**, 910–926 (2016).
- Rüb, U. et al. Huntington's disease (HD): the neuropathology of a multisystem neurodegenerative disorder of the human brain. *Brain Pathol.* **26**, 726–740 (2016).
- Reiner, A., Dragatsis, I. & Dietrich, P. Genetics and neuropathology of Huntington's disease. *Int Rev. Neurobiol.* **98**, 325–372 (2011).
- Carroll, J. B., Bates, G. P., Steffan, J., Saft, C. & Tabrizi, S. J. Treating the whole body in Huntington's disease. *Lancet Neurol.* **14**, 1135–1142 (2015).
- Bates, G. P., Tabrizi, S. J. & Jones, L. Huntington's Disease, Fourth Edition. (Oxford; Oxford University Press, 2014).
- Ratni, H. et al. Discovery of risdiplam, a selective survival of motor neuron-2 (SMN2) gene splicing modifier for the treatment of spinal muscular atrophy (SMA). *J. Med. Chem.* **61**, 6501–6517 (2018).
- Dhillon, S. Risdiplam: first approval. *Drugs* **80**, 1853–1858 (2020).
- Baldo, B. et al. A screen for enhancers of clearance identifies huntingtin as a heat shock protein 90 (Hsp90) client protein. *J. Biol. Chem.* **287**, 1406–1414 (2012).
- Naryshkin, N. A. et al. Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* **345**, 688–693 (2014).
- Palacino, J. et al. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* **11**, 511–517 (2015).
- Ratni, H., Mueller, L. & Ebeling, M. Rewriting the (tran)script: application to spinal muscular atrophy. *Prog. Med. Chem.* **58**, 119–156 (2019).
- Ratni, H. et al. Specific correction of alternative survival motor neuron 2 splicing by small molecules: discovery of a potential novel medicine to treat spinal muscular atrophy. *J. Med. Chem.* **59**, 6086–6100 (2016).
- Food and Drug Administration. EVRYSDI® (risdiplam) Prescribing information. [https://www.gene.com/download/pdf/evrydsi\\_prescribing.pdf](https://www.gene.com/download/pdf/evrydsi_prescribing.pdf) (2021).
- Sivaramkrishnan, M. et al. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat. Commun.* **8**, 1476 (2017).
- Campagne, S. et al. Structural basis of a small molecule targeting RNA for a specific splicing correction. *Nat. Chem. Biol.* **15**, 1191–1198 (2019).
- Cui, Y., Cai, M. & Stanley, H. E. Comparative analysis and classification of cassette exons and constitutive exons. *BioMed. Res. Int.* **2017**, 7323508 (2017).

- Gray, M. et al. Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. *J. Neurosci.* **28**, 6182–6195 (2008).
- Leavitt, B. R., Kordasiewicz, H. B. & Schobel, S. A. Huntingtin-lowering therapies for huntington disease: a review of the evidence of potential benefits and risks. *JAMA Neurol.* **77**, 764–772 (2020).
- Wang, G., Liu, X., Gaertig, M. A., Li, S. & Li, X.-J. Ablation of huntingtin in adult neurons is nondeleterious but its depletion in young mice causes acute pancreatitis. *Proc. Natl. Acad. Sci. USA* **113**, 3359–3364 (2016).
- Giacomini, K. M., Huang, S.-M. & Tweedie, D. J. Membrane transporters in drug development. *Nat. Rev. Drug Discov.* **9**, 215–236 (2010).
- Southwell, A. L. et al. A fully humanized transgenic mouse model of Huntington disease. *Hum. Mol. Genet.* **22**, 18–34 (2013).
- Kordasiewicz, H. B. et al. Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron* **74**, 1031–1044 (2012).
- Stanek, L. M. et al. Silencing mutant huntingtin by adeno-associated virus-mediated RNA interference ameliorates disease manifestations in the YAC128 mouse model of Huntington's disease. *Hum. Gene Ther.* **25**, 461–474 (2014).
- Caron, N. S. et al. Potent and sustained huntingtin lowering via AAV5 encoding miRNA preserves striatal volume and cognitive function in a humanized mouse model of Huntington disease. *Nucleic Acids Res.* **48**, 36–54 (2020).
- Southwell, A. L. et al. Huntingtin suppression restores cognitive function in a mouse model of Huntington's disease. *Sci. Transl. Med.* **10**, <https://doi.org/10.1126/scitranslmed.aar3959> (2018).
- Zeitler, B. et al. Allele-selective transcriptional repression of mutant HTT for the treatment of Huntington's disease. *Nat. Med.* **25**, 1131–1142 (2019).
- Ando, S. et al. Discovery of a CNS penetrant small molecule SMN2 splicing modulator with improved tolerability for spinal muscular atrophy. *Sci. Rep.* **10**, 17472 (2020).
- Lei, H., Day, I. N. M. & Vořechovský, I. Exonization of AluYa5 in the human ACE gene requires mutations in both 3' and 5' splice sites and is facilitated by a conserved splicing enhancer. *Nucleic Acids Res.* **33**, 3897–3906 (2005).
- Kim, D. et al. TopHat2: accurate alignment of transcriptsomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
- Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_t}$  method. *Methods* **25**, 402–408 (2001).

## Acknowledgements

Writing and editorial assistance was provided to the authors by Nicholas Black and Georgina Collett, PhD, on behalf of Lighthouse Medical Communications US LLC, New York, NY, USA, and funded by PTC Therapeutics, South Plainfield, NJ, USA. All authors met the ICMJE authorship criteria. Neither honoraria nor payments were made for authorship.

## Author contributions

A.B. conceived and designed the study. A.B., M.J., K.W., T.T., A.D. and V.G. developed the assays. A.B., M.J. and T.T. performed the high-throughput screening. A.B., N.S., Y.C.M., G.K. and M.W. analysed the hit selection/structure-activity relationship/chemistry. A.B., M.J. and K.W. performed the in vitro HTT protein and mRNA quantification and western blot analysis. A.B. and M.J. performed the primer walking assay. A.B., M.J., K.W., C.T., W.L., K.E. and J.G. performed the targeted next-generation sequencing. J.N., N.R., A.B., K.W., S.Y., Y.C., A.S. and M.H. performed the in vivo animal studies, drug metabolism and PK. W.L. and C.T. performed the bioinformatic analysis. W.L., C.T., A.B., K.E., N.N. and A.D. performed the statistical analyses and data interpretation. A.B., C.T., J.N., K.E., M.Wo, Y.C., W.L. and M.We drafted the manuscript. J.C., N.N., M.We, Y.C.M. and G.K. provided additional project support. S.P. provided full project support.

## Competing interests

Anuradha Bhattacharyya, Amal Dakka, Kerstin A. Effenberger, Vijayalakshmi Gabbeta, Minakshi B. Jani, Wencheng Li, Nikolai Naryshkin, Christopher R. Trotta and Kari J. Wiedinger are inventors in International Application Number PCT/US2018/037412, assigned to PTC Therapeutics, Inc., entitled 'Methods for Modifying RNA Splicing', relating to the use of HTT-C1. Anuradha Bhattacharyya, Minakshi B. Jani, Young-Choon Moon and Nadiya Sydorenko are inventors in U.S. Patent 10,874,672, assigned to PTC



Therapeutics, Inc., entitled 'Methods for Treating Huntington's Disease', relating to the use of HTT-C1. Anuradha Bhattacharyya, Minakshi B. Jani, Nadiya Sydorenko and Matthew G. Woll are inventors in International Application Number PCT/US2018/039775, assigned to PTC Therapeutics, Inc., entitled 'Methods for Treating Huntington's Disease', relating to the use of HTT-D1. Nikolai Naryshkin is an inventor in U.S. Patent 10,195,202, assigned to PTC Therapeutics, Inc., entitled 'Methods for Modulating the Amount of RNA Transcripts', relating to the use of HTT-D1. Anuradha Bhattacharyya, Christopher R. Trotta, Jana Narasimhan, Wencheng Li, Kerstin A. Effenberger, Matthew G. Woll, Minakshi Jani, Nicole Risher, Shirley Yeh, Yaofeng Cheng, Nadiya Sydorenko, Young-Choon Moon, Gary M. Karp, Marla Weetall, Amal Dakka, Vijayalakshmi Gabbeta, Jason D. Graci, Thomas Tripodi, Jr., Joseph M. Colacino and Stuart W. Peltz are present employees of PTC Therapeutics, Inc., a biotechnology company. In connection with such employment, the authors received salary, benefits and stock-based compensation, including stock options, restricted stock, other stock-related grants and the right to purchase discounted stock through PTC's employee stock purchase plan. Kari Wiedinger and Nikolai Naryshkin are former employees of PTC Therapeutics, Inc. and hold stock in the company. Amber Southwell and Michael Hayden declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27157-z>.

**Correspondence** and requests for materials should be addressed to Stuart W. Peltz.

**Peer review information** *Nature Communications* thanks James Palacino and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021