# Modulation of prion protein expression through cryptic splice site manipulation

Juliana E. Gentile[1,2], Taylor L. Corridon[1,2], Meredith A. Mortberg[1,2], Elston Neil D'Souza[3], Nicola Whiffin[3,4], Eric Vallabh Minikel[1,2], Sonia M. Vallabh[1,2]

1. McCance Center for Brain Health and Department of Neurology, Massachusetts General Hospital, Boston, MA 02114
2. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142
3. Big Data Institute and Centre for Human Genetics, University of Oxford, Oxford OX3 7LF, UK
4. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA

†Correspondence to svallabh@broadinstitute.org

## Abstract

Lowering expression of prion protein (PrP) is a well-validated therapeutic strategy in prion disease, but additional modalities are urgently needed. In other diseases, small molecules have proven capable of modulating pre-mRNA splicing, sometimes by forcing inclusion of cryptic exons that reduce gene expression. Here, we characterize a cryptic exon located in human *PRNP*'s sole intron and evaluate its potential to reduce PrP expression through incorporation into the 5' untranslated region (5'UTR). This exon is homologous to exon 2 in non-primate species, but contains a start codon that would yield an upstream open reading frame (uORF) with a stop codon prior to a splice site if included in *PRNP* mRNA, potentially downregulating PrP expression through translational repression or nonsense-mediated decay. We establish a minigene transfection system and test a panel of splice site alterations, identifying mutants that reduce PrP expression by as much as 78%. Our findings nominate a new therapeutic target for lowering PrP.

## Introduction

Prion disease is a rapidly fatal neurodegenerative disease caused by the templated misfolding of the prion protein, PrP, encoded by the prion protein gene (*PRNP* in humans)[1]. Prion disease naturally afflicts a range of mammals and has long been modeled in laboratory rodents, in which the full disease process can be induced. Both genetic[2] and pharmacological[3,4] experiments in such models have demonstrated that reducing the amount of PrP in the brain is protective against prion disease, inspiring hope that a PrP-lowering therapy could be used to effectively treat, delay, and prevent disease in patients and individuals at risk[5]. An RNAse H1 antisense oligonucleotide targeting *PRNP* RNA for degradation is now in preclinical development[3,4,6,7], but additional therapeutic candidates are urgently needed.

Recently, the FDA-approved drug risdiplam[8–11] and clinical candidates kinetin and branaplam[12–15] have highlighted small molecule modulation of pre-mRNA splicing as another tool for

therapeutic tuning of gene expression. Branaplam causes incorporation of a piece of intronic sequence — variously called a non-annotated exon, cryptic exon, or poison exon — into mature *HTT* mRNA, causing a frameshift and nonsense-mediated decay[15]. Inspired by this work, we were led to inquire whether the architecture of *PRNP* would lend itself to disruption via splice site manipulation. *PRNP*'s coding sequence is located entirely within a single exon, precluding frameshift strategies. We hypothesized, however, that inclusion of a novel upstream open reading frame (uORF) in the *PRNP* 5'UTR could decrease PrP expression. It is known that uORFs can have dramatic effects on gene expression[16,17] either through reduced abundance of ribosomes on the canonical ORF, or possibly through nonsense-mediated decay (NMD) triggered by the presence of a stop codon prior to the final splice junction, though the latter mechanism is debated[18]. The existence of Mendelian diseases caused by variants introducing uORFs[19], the evolutionary constraint of genetic variants that cause or extend uORFs in dosage sensitive genes[20], as well as work with uORF-targeting antisense oligonucleotides[21], underscore the potential functional impact of uORFs.

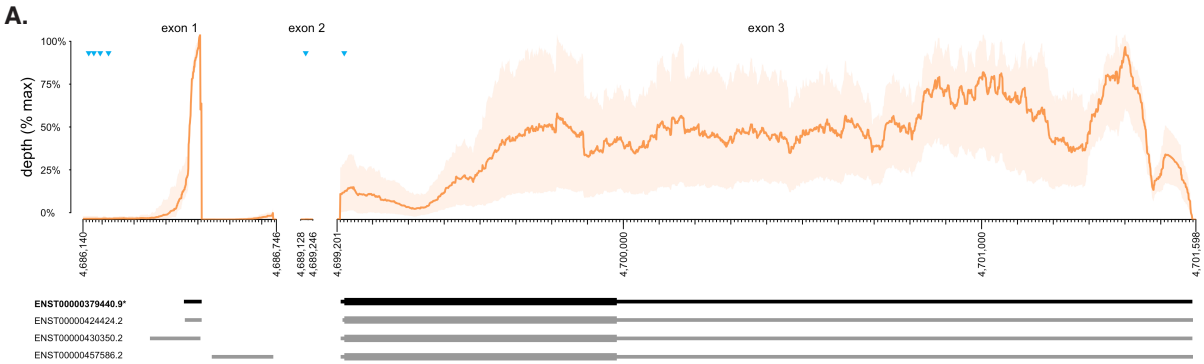
Here, we identified a potential uORF within a cryptic exon located in *PRNP*'s sole intron, homologous to exon 2 in many non-primate species. By genetically strengthening the splice sites surrounding the cryptic exon located in *PRNP*'s 5' UTR, we show that the mutations yielding the most robust inclusion of exon 2 reduced PrP expression by up to 78% in human cells. Certain other mutants reduced *PRNP* transcript levels and PrP protein expression without yielding cryptic exon inclusion detectable by qPCR, suggesting multiple mechanisms may be operative. These efforts nominate a novel strategy for lowering PrP.

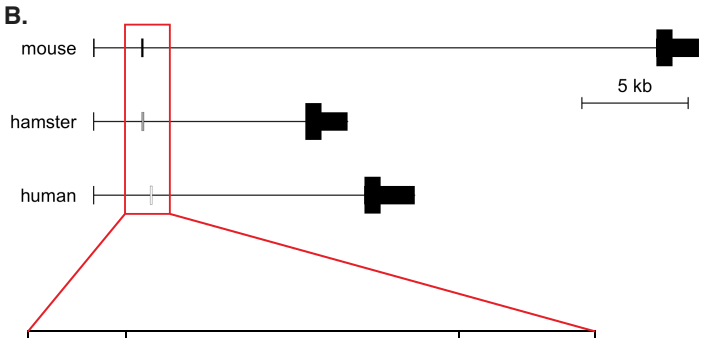
## Results

*PRNP* is a small gene of roughly 15 kilobases (kb) in humans (Figure 1A). In all mammals, the entire coding sequence is contained in the final exon of the gene, while the 5'UTR is divided across exons, however, the number of exons differs. In mouse and most other preclinical species of interest, there are 3 constitutive exons[22,23], with introns 1 and 2 dividing the 5'UTR (Figure 1B). In Syrian hamsters, exon 2 is subject to variable splicing and is included in ~27% of transcripts[24] (Figure 1B). In humans and several closely related primate species, *PRNP* has only two annotated exons, the equivalent of exons 1 and 3 from other mammals; exon 2 remains as a cryptic exon within the sole intron[25]. For clarity, herein we will refer to human *PRNP* exons 1, 2, and 3, and introns 1 and 2, even though the naturally occurring *PRNP* transcript contains only 2 exons and 1 intron.

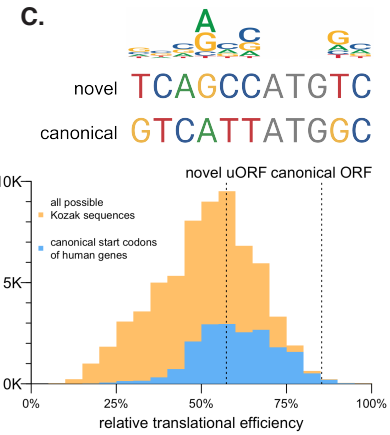
Although essential splice sites — AG at the A-1 and A-2 and GT at the D+1 and D+2 positions — are conserved in human exon 2, we hypothesized that other nearby base pair substitutions may contribute to exclusion of this exon, particularly the loss of the G at the highly constrained D+5 position[26] (Figure 1B). Human *PRNP* exon 2 contains an ATG in a moderately strong Kozak context (Figure 1C), estimated to yield 57% maximal translational efficiency, near the median of canonical ORFs of all other human protein-coding genes[27] (Figure 1D). Human *PRNP* was previously reported[28] to already contain 4 uORFs in exon 1, however, RNA-seq data from human brain tissue[29] provide no support for transcription initiation beginning this far upstream: mean RNA-seq coverage at these uORFs is <0.5% of the peak coverage within exon

1 (Figure 1A). Thus, if exon 2 were included, its ATG would yield a new, sole uORF upstream of *PRNP*'s canonical start codon (Figure 1E) with the potential to downregulate PrP expression through its impact on ribosomal activity[16,30]. Its stop codon also occurs 22 bp prior to the exon 2/3 splice junction (Figure 1E), creating a possible opportunity to trigger nonsense-mediated decay (NMD; see Discussion). Alignment of *PRNP* exon 2 sequences across all available mammalian species (Figure 1F and Figure S1) reveals that exon 2 ATGs are present only in species with exon 2 splice site variants known or predicted to exclude exon 2 from mature mRNA, consistent with the possibility of exon 2 uORF having a strong negative effect on PrP expression. We thereby hypothesized that acting through either of these mechanisms, inclusion of exon 2 and thus the uORF of interest in *PRNP* mRNA would reduce PrP expression.




| species | exon 2 splice sites acceptor | donor | Inclusion? |
|---|---|---|---|
| **mouse** | TTAAAG \| GACTCC…CAGACT \| GTGAGT | | 100% |
| **hamster** | TTGAAG \| GACTCC…CAGATC \| GTAAGT | | 27% |
| **human** | TTTAAG \| GACTCC…CAGATT \| GTACAT | | 0% |




E.

GCCAGTCGCTGACAGCCGCGGCGCCGCGAGCTTCTCCTCTC
CTCACGACCGAGGCAGGACTCCTGAATATTTTTCAAAACTG
AACAATTTCAGCCATGTCTGAGCTTTCCGTCTTCCTGGAGG
CACAAATCTAGTTTAGCTGAACCACAACAGATTAGCAGTCA
TTATGGCGAACCTTGGCTGC
M A N L G C

Key: Exon1 **Exon2** Exon3 **Start** **Stop** ORF

F.

```
          Mouse  ttaaag|gactcct-gag--tatatttcagaactgaaccatttcaaccgagct---gaagcattctg---ccttcctag-tggtaccagtccaattt-aggag-agcca-agcagact|gtgagt
 Golden hamster  ttgaag|gactcct-gaa--tatattccaaaactgaacaatttcaactgagct---gaagtactctg---tttttctag-aggtaccagttcagttt-aggag-agtcacagcagatc|gtaagt
          Sheep  tttaag|gacttct-gaa--tatatttgaaaactgaacagtttcaaccaagct---gaagca-tctg---tcttcccag-agacacagatccaacttgagctg-aatcacagcagat-|gtaggt

          Human  tttaag|gactcct-gaa--tatttttcaaaactgaacaatttcagccatgtc---tgagctttccg---tcttcctgg-aggcacaaatctagttt-agctg-aaccacaacagatt|gtacat
          Chimp  tttaag|gactcct-gaa--tatttttcaaaactgaacaatttcagccatgtc---tgagctttctg---tcttcctgg-aggcacaaatctagttt-agctg-aaccgcaacagatt|gtacat
        Gorilla  tttaag|gactcct-gaa--tatttttcaaaactgagtaatttcagccatgtc---taagctttctg---tcttcctgg-aggcacaaatctagttt-agctg-aaccacaacagatt|gtacat
      Orangutan  tttaag|gactcct-gaa--tatttttcaaaactgaacaatttcagccatgtc---taagctttctg---tcttcctgg-aggcacaaatctagtta-agctg-aaccacaacagatt|gtacat
         Gibbon  tttaag|gactcct-gaa--tatttttcagaactgaacaatttcagccatgtc---taagctttctg---tcttcccgg-aggcacaaatctagttt-agctg-aaccacaacagatt|gtacat
         Rhesus  tttaag|gactcct-gaa--tatttttcaaaactgaacaatttcagccatgtc---taagctttctg---tcttcctgg-aggcacaaatctagttt-agctg-agccacaacagatt|gtacat
Crab-eating macaque  tttaag|gactcct-gaa--tatttttcaaaactgaacaatttcagccatgtc---taagctttctg---tcttcctgg-aggcacaaatctagttt-agctg-agccacaacagatt|gtacat
         Baboon  tttaag|gactcct-gaa--tgttttcaaaactgaacaattgcagccatgtc---taagctttctg---tcttcctgg-aggcacaaatctagttt-agctg-agccacaacagatt|gtacat
    Green monkey  tttaag|gactcct-gaa--tatttttcaaaactgaacaatttcagccatgtc---taagctttctg---tcttcctgg-agacacaaatctagttt-agctg-agccacaacagatt|gtacat
       Marmoset  tttaag|ggctcct-gaa--actttttcaaaactgaacaatttcagccatgtc---taagctttctg---tcttcttgg-aggcacaaatccagttt-agctg-agccacaacagatt|gtacat
  Squirrel monkey  -ttaag|gactcct-gaa--actttttcaaaactgaacaatttcagccatgtc---taagctttctg---tcttcttgg-aggcacaaatccagttt-agctg-agccacaacagatt|gtacat
        Bushbaby  ttgaag|gactcct-gaa--tatgtgtcaaaactgaacagtttcagccatgct---gaagcattctg---tcttcctgg-aggcacaaatccaacct-aactg-agctgcaacagatg|gcatgt
Chinese tree shrew  tttaag|aactcct-gaa--tacctttaaaatggaaccatttcggc---------taggcattctg---tcttactgt-aggcactggtccagttt-agctg-agctacaacagatt|gtacct
            Pika  ttt caa|gactcct-gaa--tgtatttggaagc--aacgttttcagcctagcc---------ttgcg---tctt---------ccccaatccagtgt-agctc-aactatagcagatt|----gt
```

***Figure 1. A cryptic exon in human PRNP. A)*** *Human PRNP transcript structure in human brain. Top panels show GTEx[29] v8 bulk RNA-seq coverage — mean (orange lines) and range (orange shaded area) across 13 brain regions. Coverage depth for exons 1 and 2 is normalized to the max for exon 1; depth for exon 3 is normalized to the max for exon 3. ATGs representing candidate upstream open reading frames and the canonical open reading frame are shown as blue triangles. Ensembl GRCh38.p14 annotated transcripts are shown below, canonical in black, alternatives in gray.* ***B)*** *Comparison of orthologous exon 2 sequence in mouse, hamster, and human. Hamster inclusion percentage from ref [24].* ***C)*** *Comparison of PRNP canonical and exon 2 novel ATG Kozak contexts with a sequence logo of human initiation sites (see Methods).* ***D)*** *Relative strength of canonical and novel PRNP ORFs in context. Shown for comparison are histograms of translational efficiency of all 65,536 (4^8) possible Kozak contexts (yellow) and of all 18,784 actual human canonical ORF Kozak contexts (blue), expressed as a percentage of the translation of the most efficient Kozak context, TTCATCATGCA, according to data from Noderer et al[27].* ***E)*** *Annotated sequence of the PRNP 5'UTR if exon 2 were included. Frame is relative to the canonical ORF, and percentile indicates strength of the Kozak context as a percentile of all possible Kozak sequences, using rankings from ref [27].* ***F)*** *Multiple alignment of PRNP exon 2 sequences known to be constitutively or variably included in mRNA from mouse[22], hamster[24], and sheep[23] versus all orthologous sequences in eutherian mammals that contain ATGs. ATGs are shown in blue, splice site variants absent from mouse, hamster, or sheep are shown in orange. A full alignment including all eutherian mammals is shown in Figure S1.*

To test this hypothesis, we first sought to generate a *PRNP* minigene system to support facile splice site manipulation, transfection, and screening in cell culture. A 4.8 kb minigene lacking most of intron 1 yielded no detectable PrP expression in HEK293 cells by Western blot (Figure S2). A 6.5 kb construct retaining all of intron 1 and only the first and last 500 bp of intron 2 (Figure 2A) expressed robustly, and was used for all subsequent experiments. Codon optimization of exon 3 allowed for qPCR primer/probe pairs to discriminate minigene *PRNP* RNA from endogenous *PRNP* RNA (Figure 2B).

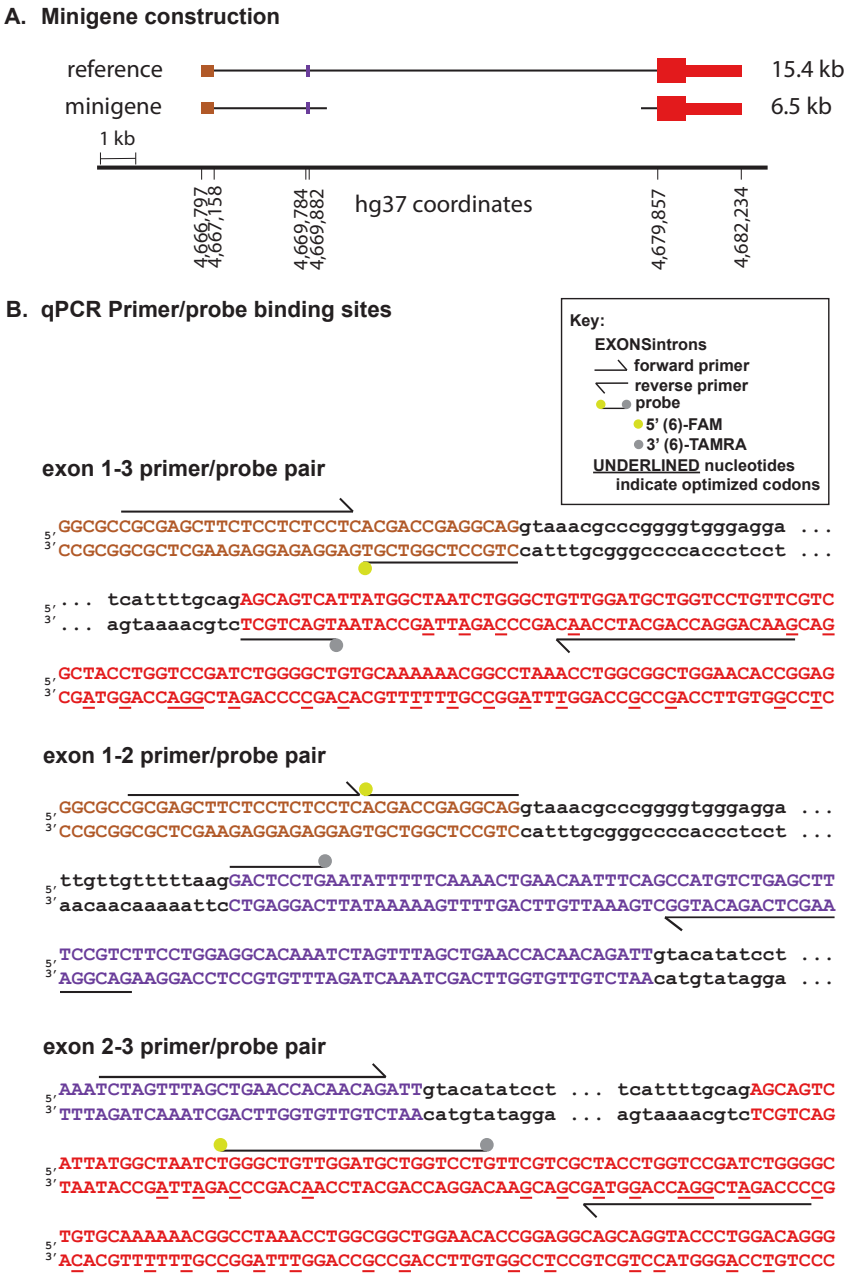



***Figure 2. Design of PRNP minigene and primer pairs. A)*** *Diagram of minigene versus human reference sequence. Intronic sequence upstream of exon 2 and 500 bp on either end of the intronic sequence downstream of exon 2 are included.* ***B)*** *Design of primer/probe pairs used to interrogate splicing of the minigene. Note that codon optimization in exon 3 (underlines) enables these pairs to discriminate the minigene from endogenous PRNP in HEK293 cells.*

Using this 6.5 kb minigene as a template, we designed a panel of splice site modifications that we hypothesized would strengthen exon 2 inclusion in the context of human *PRNP* (Figure 3A).

These included 1) installation of the consensus strongest[31] human splice donor and acceptor ("canonical ss"; 6 nucleotide changes required), 2) installation of the mouse *Prnp* exon 2 splice sites ("mouse ss"; 5 changes required), 3) conversion of the donor +5 site from A to G, as this site shows the strongest nucleotide preference of any extended splice site position[26] (D+5, A>G); 4) conversion of the acceptor -3 site from A to C, to assess whether this single change could mimic the effect of installing the consensus human splice site (A-3, A>C); and 5) conversion of the acceptor -4 site from T to A, to assess whether this single change could mimic the effect of installing the mouse *Prnp* splice site (A-4, T>A).

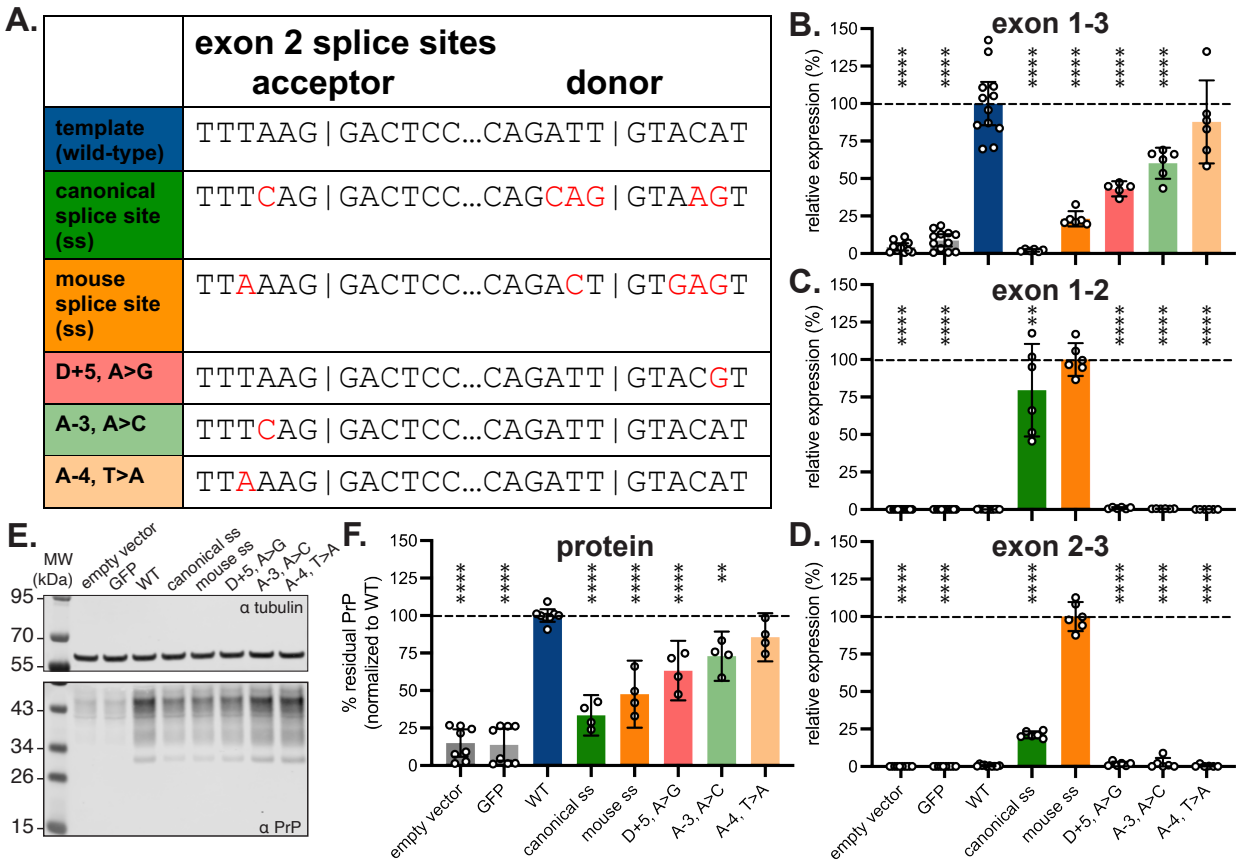



***Figure 3. Inclusion of exon 2 lowers PrP expression. A)*** *Sequence variants of minigene tested in HEK293 cells.* ***B-D)*** *Expression of (B) exon 1-3 (n = 5-12 transfected wells/variant), (C) exon 1-2 (n = 6-12 transfected wells/variant), and (D) exon 2-3 (n = 6-10 transfected wells/variant) junctions in minigene mRNA for each variant transfected into HEK293 cells. Normalized to the template minigene for exons 1-3 and normalized to the highest-expressing variant for exons 1-2 and 2-3. Note that codon optimization in exon 3 enables discrimination from endogenous PRNP.* ***E)*** *Immunoblot (POM2 primary antibody[32] of PrP expression in HEK293 cells transfected with each variant.* ***F)*** *Quantification of PrP expression from ≥4 immunoblots per construct, n = 4-8 transfected wells/variant.*

Each mutant was separately transfected into HEK293 cells alongside the parent minigene construct and empty vector and GFP transfection controls and analyzed by qPCR. Each primer/probe set (Figure 2B) was designed to amplify only if the targeted exons are adjacent. In keeping with these expectations, the empty vector and GFP controls yielded negligible signal for all primer pairs; trace amplification of exon 1-3 may reflect imperfect allele specificity, as only 2 bases differ from endogenous *PRNP* in the exon 3 codon-optimized primer. The parent minigene yielded PCR product for exon 1-3 but not for exon 1-2 or 2-3, reflecting the baseline exclusion of cryptic exon 2 in a human system.

All 5 splice site mutants appeared to reduce the amount of normally spliced *PRNP* RNA, as measured by the exon 1-3 primer pair, with the change being significant for 4 mutants (Figure 3B). The two mutants yielding the greatest reduction — the canonical ss and mouse ss mutants — showed a corresponding increase in the presence exon 1-2 and 2-3 junctions (Figure 3C-D). For all other constructs, exon 2 remained undetectable, or nearly so, by these primer/probe sets. Note that the results for exons 1-2 and 2-3 are normalized to the highest value obtained for any mutant; 100% does not necessarily mean 100% exon 2 inclusion.

Immunoblots on cell lysates revealed apparent reductions in PrP for all mutants tested (Figure 3E-F). The canonical ss mutant yielded 33% and the mouse ss mutant 48% of the PrP expression level of the parent minigene (Figure 3F). HEK293 cells express endogenous PrP, however, at ~15% the level achieved by transfection of the parent minigene (Figure 3E-F); adjusting for this floor yielded residual PrP expression of 22% and 38% for the canonical ss and mouse ss mutants, respectively. Across all mutants, PrP levels tracked closely with exon 1-3 qPCR results, with significant reductions for mutants D+5, A>G and A-3, A>C despite the lack of detectable exon 1-2 and 2-3 junctions (Figure 3F, 3C, 3D).

## Discussion

We find that splice site manipulation can modulate the level of PrP in a human cell system, reducing the levels of this disease-causing protein by 78% in the strongest condition tested. For the strongest mutants, which incorporated 5-6 nucleotide changes across the splice donor and acceptor sites, this reduction in protein level was observed in tandem with exon 2 inclusion at the mRNA level. This would be consistent with uORF-mediated translational repression, however, we cannot rule out that nonsense-mediated decay (NMD) may be at work, with the exon 1-2 and 2-3 qPCR simply picking up the small fraction of exon 2-including mRNA that has not yet been degraded. NMD was long held to require 50 bp of distance between the stop codon and the splice donor[33], versus only 22 bp here, but data from protein-truncating variants in human tissues show this is not a hard-and-fast rule, and that distance from the splice donor is but one of many imperfect predictors of NMD[34]. Still, the evidence for NMD caused by uORFs in human genes is equivocal[18,35]. The single point mutants tested here reduced PrP and normal exon 1-3 splicing without yielding detectable exon 1-2 and 2-3 splicing. Thus, additional mechanisms not foreseen by our initial hypothesis could be operative. One possibility is that these mutants cause inclusion of exon 2 but also retention of a portion of intronic sequence,

causing the exon 1-2 and 2-3 qPCR to not amplify, while still resulting in NMD and/or uORF-mediated translational repression.

Our study has several limitations. The battery of splice manipulations that we tested was limited, leaving open the possibility that other splice site changes could yield more dramatic results. As our experiments were limited to human cell culture, in vivo relevance was not demonstrated. Most importantly, the genetic engineering used to establish this proof of concept does not offer a direct path to therapeutic application.

In principle, several therapeutic modalities could be deployed to modulate *PRNP* splicing[36,37]. Antisense oligonucleotides (ASOs) are a well-established modality capable of causing exon inclusion[38], but may be unlikely to be deployed towards this end: given the desired mechanism of reducing *PRNP* expression, RNAse H1 ASOs are likely to yield greater target suppression than splice-modulating ASOs. Adenine base editors have been successfully deployed to disrupt splice sites[39,40], however, the single point mutants identified here had relatively modest effects on PrP expression. Instead, small molecule modulation of *PRNP* splicing is the most enticing possibility suggested by our results. PrP-lowering small molecules could have desirable pharmacologic properties, particularly in terms of distribution to deep brain structures less well-reached by oligonucleotides[41]. Attempts to discover small molecules to bind PrP have been unsuccessful[42], so splicing could offer a new mechanism for small molecule therapies in prion disease. Because *PRNP* does not share the preferred splice site motifs of any known splice-modulating small molecule series[10,12,14], discovery of a modulator would require a new screening effort.

Despite these limitations, we are encouraged to discover a novel mechanism by which PrP expression can be influenced. PrP's role in prion disease is uniquely pivotal, as it serves as protein-only pathogen, amplification substrate, and mediator of neuronal neurotoxicity. The therapeutic benefit of PrP lowering has been shown across multiple prion strains[4], both through genetic reduction and by use of antisense oligonucleotides, and evidence for tolerability is provided by multiple nonhuman species as well as human genetics[43–48]. Given this clarity, PrP and its precursors are disease targets worthy of ongoingly creative angles of attack.

## Methods

### Kozak sequences

Files were retrieved from the Matched Annotation from NCBI and EMBL-EBI (MANE) database (version 1.0) https://ftp.ncbi.nlm.nih.gov/refseq/MANE/MANE_human/release_1.0/ cDNA Transcript sequences from the MANE.GRCh38.v1.0.refseq_rna.fna.gz file were then filtered to only MANE Select transcripts and an 11-bp context surrounding the CDS start were extracted, excluding transcripts where a 11bp CDS context could not be retrieved, as in the case for a leaderless mRNA. To generate a sequence logo, these 11-bp sequences were then super-imposed to align with each other and plotted using ggplot2 and the R package ggseqlogo using the bits method. To generate a histogram, the relative translational efficiencies of each

sequence were taken from Noderer et al[27] and normalized to the most efficient Kozak sequence.

**Comparative genomic analyses**

*PRNP* sequences, and multiple alignments thereof, were obtained from UCSC Genome Browser[49] (accessed September 6, 2023). Kozak sequence strength percentiles were obtained from the rank order among all possible Kozak sequences reported by ref [27]. GTEx[29] RNA-seq coverage data were obtained from UCSC Table Browser (accessed November 14, 2023). Exon 1 and 2 in diagrams correspond to the canonical Ensembl transcript ENST00000379440.9.

**Cell culture and transfections**

HEK293 cells were maintained in DMEM/F-12 (Gibco, cat no. 11320033) supplemented with 1% Penicillin-streptomycin (Gibco, cat no. 15140163) and 10% FBS (Gibco, cat no. 16000044). For transfection, cells were plated in a 12-well or 96-well plate for protein or RNA analysis, respectively, and were allowed to adhere for 18 hours. Cells were then transfected using Lipofectamine 3000 transfection reagent (Invitrogen, cat no. L3000015) according to the manufacturer protocol. In short, lipofectamine 3000 reagent was diluted in Opti-MEM I reduced serum media (Gibco, cat no. 31985088) for a final mixture containing 3% lipofectamine 3000. In a separate tube, 1 µg (12- well plate) or 0.1 µg (96- well plate) DNA was mixed with 4% P3000 reagent in Opti-MEM. The two tubes were slowly mixed then allowed to incubate at room temperature for 10 minutes before applying the mixture to the cell media. Transfection was incubated on cells for 48 hr before lysing cells.

**Plasmids**

Plasmid cloning was performed by Genscript using a modified version of pcDNA3.1(+). The CMV promoter was cloned out of the pcDNA3.1(+) backbone (addgene V790-20) by digesting the vector with NruI and NheI. The human PGK promoter (addgene 82579) was cloned into the backbone, creating pcDNA3.1(+)-hPGK. The minigene was synthesized with codon optimized exon 3, then was ligated into pcDNA3.1(+)-hPGK between NheI and EcoRI.

**Western blot analysis**

Following the 48 hr transfection, cells were washed thoroughly with ice- cold PBS then were lysed in 0.2% CHAPS containing cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Sigma, cat no. 4693159001). Protein concentration was determined using a DC protein assay kit (Bio-rad, cat no. 5000112). NuPAGE 4-12%, Bis-Tris, mini protein gels (Invitrogen, cat no. NP0323BOX) were loaded with 10 µg total protein for each sample and run at 180 V in 1x MES buffer (Thermo, cat no. NP0002). Gels were transferred to PVDF membranes using an iBlot 2 device (iBlot™ 2 Transfer Stacks, PVDF, mini, Thermo, cat no. IB24002), 20 V, 7 minutes. Membrane was then cut right under 55 kDa band before blocking with LICOR TBS blocking buffer (LICOR, cat no. 927-60001), 1 hr at room temperature. Primary antibodies were diluted in LICOR TBS blocking buffer + 0.2% Tween-20 (Teknova, cat no. T0710) and incubated at 4°C overnight: α-Tubulin (Invitrogen, cat no. A11126), final 100 ng/µL; POM2 (Millipore, cat no. MABN2298), final 50 ng/µL; 6D11 (BioLegend, cat no. 808001), final 2 µg/µL. Membranes were washed in 1x TBST then incubated in secondary antibody (IRDye® 800CW Goat anti-Mouse

IgG, LICOR, cat no. 926-32210) diluted in LICOR TBS blocking buffer + 0.2% Tween-20 and incubated at room temperature for 1 hr. Membranes were again washed with 1x TBST then scanned on a LICOR Odyssey CLx Infrared Imaging System. Blots were analyzed in Fiji[50].

**qPCR**

Following the 48 hr transfection, cells were lysed using the Cells-to-CT 1-step Taqman Kit (Invitrogen, cat no. A25602) using the manufacturer protocol. In short, media was aspirated, each well was washed with 200 µL of ice cold 1x PBS then wash was completely aspirated. Room temperature DNase/Lysis solution (0.5 µL: 50 µL) was added to the cells then plate was put on a shaker for 5 minutes. Finally, 5 µL of room temperature stop solution was added to the cells then plate was put back on the shaker for 2 minutes before moving the plate to ice. RT-PCR samples were prepared using Taqman 1-Step qRT-PCR master mix and Taqman gene expression assays for human *TBP* (Invitrogen, cat no. Hs00427620_m1). Custom primers and probes were ordered from Genscript to quantify the different splice variants (see Table S1 for sequences and Figure 2 for alignment on the minigene sequence). Samples were run on a QuantStudio 7 Flex system (Applied Biosystems) using the following cycling conditions: Reverse transcription (RT) 50°C, 5 min; RT inactivation/initial denaturation 95°C, 20 sec; Amplification 95°C, 3 sec, 60°C, 30 sec, 40 cycles. Each biological sample was run in duplicate and the level of all targets were determined by ΔΔCt whereby results were first normalized to the housekeeping gene *TBP* and then to the wild-type template (exon 1-3) or the mouse ss (exon 1-2 and 2-3), depending on the primer pair used.

**Experimental design and statistical analysis**

All data was generated from at least 3 independent transfections, N are as indicated in figure legends. Throughout, all error bars in figures represent 95% confidence intervals. All data were compared with an ordinary one-way ANOVA and Dunnett's multiple comparison test, with a single pooled variance. P values less than 0.05 were considered nominally significant. In plots, **, $p < 0.01$ and ****, $p < 0.0001$. Intron/exon diagrams were plotted in R, qPCR analysis was performed in Google Sheets, and barplots and statistical analyses were performed in Graphpad Prism. Raw data, Prism files, and source code will be made available at https://github.com/ericminikel/cryptic_exon

## FUNDING

This study was supported by the Broad Institute (Chemical Biology and Therapeutic Science program funds) and National Institutes of Health (R01 NS125255). NW is supported by a Sir Henry Dale Fellowship jointly funded by Wellcome and the Royal Society (220134/Z/20/Z) and research grant funding from the Rosetrees Trust (PGL19-2/10025)

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

## AUTHOR CONTRIBUTIONS

Dr. Vallabh had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: SMV, EVM, JEG. Sample analysis: JEG, TLC, MAM. Statistical analysis: JEG. Genomic analysis: EVM, NW, END. Drafting of the manuscript: EVM and SMV. Critical review of the manuscript: all authors. Obtained funding: SMV, NW.

## DISCLOSURES

SMV acknowledges speaking fees from Ultragenyx, Illumina, Biogen, Eli Lilly; consulting fees from Invitae and Alnylam; research support from Ionis, Gate, Sangamo. EVM acknowledges speaking fees from Eli Lilly; consulting fees from Deerfield and Alnylam; research support from Ionis, Gate, Sangamo.

# REFERENCES

1. Prusiner SB. Prions. Proc Natl Acad Sci USA. 1998 Nov 10;95(23):13363–13383. PMCID: PMC33918

2. Büeler H, Raeber A, Sailer A, Fischer M, Aguzzi A, Weissmann C. High prion and PrPSc levels but delayed onset of disease in scrapie-inoculated mice heterozygous for a disrupted PrP gene. Mol Med. 1994 Nov;1(1):19–30. PMCID: PMC2229922

3. Raymond GJ, Zhao HT, Race B, Raymond LD, Williams K, Swayze EE, Graffam S, Le J, Caron T, Stathopoulos J, O'Keefe R, Lubke LL, Reidenbach AG, Kraus A, Schreiber SL, Mazur C, Cabin DE, Carroll JB, Minikel EV, Kordasiewicz H, Caughey B, Vallabh SM. Antisense oligonucleotides extend survival of prion-infected mice. JCI Insight. 2019 30;5. PMID: 31361599

4. Minikel EV, Zhao HT, Le J, O'Moore J, Pitstick R, Graffam S, Carlson GA, Kavanaugh MP, Kriz J, Kim JB, Ma J, Wille H, Aiken J, McKenzie D, Doh-Ura K, Beck M, O'Keefe R, Stathopoulos J, Caron T, Schreiber SL, Carroll JB, Kordasiewicz HB, Cabin DE, Vallabh SM. Prion protein lowering is a disease-modifying therapy across prion disease stages, strains and endpoints. Nucleic Acids Res. 2020 Aug 10; PMID: 32776089

5. Vallabh SM, Minikel EV, Schreiber SL, Lander ES. Towards a treatment for genetic prion disease: trials and biomarkers. Lancet Neurol. 2020 Apr;19(4):361–368. PMID: 32199098

6. Mortberg MA, Zhao HT, Reidenbach AG, Gentile JE, Kuhn E, O'Moore J, Dooley PM, Connors TR, Mazur C, Allen SW, Trombetta BA, McManus A, Moore MR, Liu J, Cabin DE, Kordasiewicz HB, Mathews J, Arnold SE, Vallabh SM, Minikel EV. Regional variability and genotypic and pharmacodynamic effects on PrP concentration in the CNS. JCI Insight. 2022 Mar 22;7(6):e156532. PMCID: PMC8986079

7. Mortberg MA, Gentile JE, Nadaf NM, Vanderburg C, Simmons S, Dubinsky D, Slamin A, Maldonado S, Petersen CL, Jones N, Kordasiewicz HB, Zhao HT, Vallabh SM, Minikel EV. A single-cell map of antisense oligonucleotide activity in the brain. Nucleic Acids Res. 2023 May 16;gkad371. PMID: 37188501

8. Naryshkin NA, Weetall M, Dakka A, Narasimhan J, Zhao X, Feng Z, Ling KKY, Karp GM, Qi H, Woll MG, Chen G, Zhang N, Gabbeta V, Vazirani P, Bhattacharyya A, Furia B, Risher N, Sheedy J, Kong R, Ma J, Turpoff A, Lee CS, Zhang X, Moon YC, Trifillis P, Welch EM, Colacino JM, Babiak J, Almstead NG, Peltz SW, Eng LA, Chen KS, Mull JL, Lynes MS, Rubin LL, Fontoura P, Santarelli L, Haehnke D, McCarthy KD, Schmucki R, Ebeling M, Sivaramakrishnan M, Ko CP, Paushkin SV, Ratni H, Gerlach I, Ghosh A, Metzger F. Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. Science. 2014 Aug 8;345(6197):688–693. PMID: 25104390

9. Ratni H, Ebeling M, Baird J, Bendels S, Bylund J, Chen KS, Denk N, Feng Z, Green L, Guerard M, Jablonski P, Jacobsen B, Khwaja O, Kletzl H, Ko CP, Kustermann S, Marquet A, Metzger F, Mueller B, Naryshkin NA, Paushkin SV, Pinard E, Poirier A, Reutlinger M, Weetall M, Zeller A, Zhao X, Mueller L. 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. 2018 Aug 9;61(15):6501–6517. PMID: 30044619

10. Campagne S, Boigner S, Rüdisser S, Moursy A, Gillioz L, Knörlein A, Hall J, Ratni H, Cléry A, Allain FHT. Structural basis of a small molecule targeting RNA for a specific splicing correction. Nat Chem Biol. 2019 Dec;15(12):1191–1198. PMID: 31636429

11. Baranello G, Darras BT, Day JW, Deconinck N, Klein A, Masson R, Mercuri E, Rose K, El-Khairi M, Gerber M, Gorni K, Khwaja O, Kletzl H, Scalco RS, Seabrook T, Fontoura P, Servais L, FIREFISH Working Group. Risdiplam in Type 1 Spinal Muscular Atrophy. N Engl J Med. 2021 Mar 11;384(10):915–923. PMID: 33626251

12. Hims MM, Ibrahim EC, Leyne M, Mull J, Liu L, Lazaro C, Shetty RS, Gill S, Gusella JF, Reed R, Slaugenhaupt SA. Therapeutic potential and mechanism of kinetin as a treatment for the human splicing disease familial dysautonomia. J Mol Med. 2007 Feb;85(2):149–161. PMID: 17206408

13. Axelrod FB, Liebes L, Gold-Von Simson G, Mendoza S, Mull J, Leyne M, Norcliffe-Kaufmann L, Kaufmann H, Slaugenhaupt SA. Kinetin improves IKBKAP mRNA splicing in patients with familial dysautonomia. Pediatr Res. 2011 Nov;70(5):480–483. PMCID: PMC3189334

14. Palacino J, Swalley SE, Song C, Cheung AK, Shu L, Zhang X, Van Hoosear M, Shin Y, Chin DN, Keller CG, Beibel M, Renaud NA, Smith TM, Salcius M, Shi X, Hild M, Servais R, Jain M, Deng L, Bullock C, McLellan M, Schuierer S, Murphy L, Blommers MJJ, Blaustein C, Berenshteyn F, Lacoste A, Thomas JR, Roma G, Michaud GA, Tseng BS, Porter JA, Myer VE, Tallarico JA, Hamann LG, Curtis D, Fishman MC, Dietrich WF, Dales NA, Sivasankaran R. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. Nat Chem Biol. 2015 Jul;11(7):511–517. PMID: 26030728

15. Krach F, Stemick J, Boerstler T, Weiss A, Lingos I, Reischl S, Meixner H, Ploetz S, Farrell M, Hehr U, Kohl Z, Winner B, Winkler J. An alternative splicing modulator decreases mutant HTT and improves the molecular fingerprint in Huntington's disease patient neurons. Nat Commun. 2022 Nov 10;13(1):6797. PMCID: PMC9649613

16. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci USA. 2009 May 5;106(18):7507–7512. PMCID: PMC2669787

17. Johnstone TG, Bazzini AA, Giraldez AJ. Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J. 2016 Apr 1;35(7):706–723. PMCID: PMC4818764

18. Lee DSM, Park J, Kromer A, Baras A, Rader DJ, Ritchie MD, Ghanem LR, Barash Y. Disrupting upstream translation in mRNAs is associated with human disease. Nat Commun. 2021 Mar 9;12(1):1515. PMCID: PMC7943595

19. Wright CF, Quaife NM, Ramos-Hernández L, Danecek P, Ferla MP, Samocha KE, Kaplanis J, Gardner EJ, Eberhardt RY, Chao KR, Karczewski KJ, Morales J, Gallone G, Balasubramanian M, Banka S, Gompertz L, Kerr B, Kirby A, Lynch SA, Morton JEV, Pinz H, Sansbury FH, Stewart H, Zuccarelli BD, Genomics England Research Consortium, Cook SA, Taylor JC, Juusola J, Retterer K, Firth HV, Hurles ME, Lara-Pezzi E, Barton PJR, Whiffin N. Non-coding region variants upstream of MEF2C cause severe developmental disorder through three distinct loss-of-function mechanisms. Am J Hum Genet. 2021 Jun 3;108(6):1083–1094. PMCID: PMC8206381

20. Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, Evans DG, Roberts AM, Quaife NM, Schafer S, Rackham O, Alföldi J, O'Donnell-Luria AH, Francioli LC, Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Cook SA, Barton PJR, MacArthur DG, Ware JS. Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. Nat Commun. 2020 May 27;11(1):2523. PMCID: PMC7253449

21. Liang XH, Shen W, Sun H, Migawa MT, Vickers TA, Crooke ST. Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. Nat Biotechnol. 2016;34(8):875–880. PMID: 27398791

22. Westaway D, Cooper C, Turner S, Da Costa M, Carlson GA, Prusiner SB. Structure and polymorphism of the mouse prion protein gene. Proc Natl Acad Sci U S A. 1994 Jul 5;91(14):6418–6422. PMCID: PMC44213

23. Westaway D, Zuliani V, Cooper CM, Da Costa M, Neuman S, Jenny AL, Detwiler L, Prusiner SB. Homozygosity for prion protein alleles encoding glutamine-171 renders sheep susceptible to natural scrapie. Genes Dev. 1994 Apr 15;8(8):959–969. PMID: 7926780

24. Li G, Bolton DC. A novel hamster prion protein mRNA contains an extra exon: increased expression in scrapie. Brain Res. 1997 Mar 21;751(2):265–274. PMID: 9099814

25. Lee IY, Westaway D, Smit AF, Wang K, Seto J, Chen L, Acharya C, Ankener M, Baskin D, Cooper C, Yao H, Prusiner SB, Hood LE. Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res. 1998 Oct;8(10):1022–1037. PMID: 9799790

26. Zhang S, Samocha KE, Rivas MA, Karczewski KJ, Daly E, Schmandt B, Neale BM, MacArthur DG, Daly MJ. Base-specific mutational intolerance near splice sites clarifies the role of nonessential splice nucleotides. Genome Res. 2018;28(7):968–974. PMCID: PMC6028136

27. Noderer WL, Flockhart RJ, Bhaduri A, Diaz de Arce AJ, Zhang J, Khavari PA, Wang CL. Quantitative analysis of mammalian translation initiation sites by FACS-seq. Mol Syst Biol. 2014 Aug 28;10(8):748. PMCID: PMC4299517

28. Moreno JA, Radford H, Peretti D, Steinert JR, Verity N, Martin MG, Halliday M, Morgan J, Dinsdale D, Ortori CA, Barrett DA, Tsaytler P, Bertolotti A, Willis AE, Bushell M, Mallucci GR. Sustained translational repression by eIF2α-P mediates prion neurodegeneration. Nature. 2012 May 24;485(7399):507–511. PMCID: PMC3378208

29. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020 Sep 11;369(6509):1318–1330. PMCID: PMC7737656

30. Whiffin N. Characterising the loss-of-function impact of 5' untranslated region variants in 15,708whole-genomes. In preparation.

31. Blakes AJM, Wai HA, Davies I, Moledina HE, Ruiz A, Thomas T, Bunyan D, Thomas NS, Burren CP, Greenhalgh L, Lees M, Pichini A, Smithson SF, Taylor Tavares AL, O'Donovan P, Douglas AGL, Genomics England Research Consortium, Splicing and Disease Working Group, Whiffin N, Baralle D, Lord J. A systematic analysis of splicing variants identifies

new diagnoses in the 100,000 Genomes Project. Genome Med. 2022 Jul 26;14(1):79. PMCID: PMC9327385

32. Polymenidou M, Moos R, Scott M, Sigurdson C, Shi YZ, Yajima B, Hafner-Bratkovic I, Jerala R, Hornemann S, Wuthrich K, Bellon A, Vey M, Garen G, James MNG, Kav N, Aguzzi A. The POM monoclonals: a comprehensive set of antibodies to non-overlapping prion protein epitopes. PLoS ONE. 2008;3(12):e3872. PMCID: PMC2592702

33. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci. 1998 Jun;23(6):198–199. PMID: 9644970

34. Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ, Maller JB, Kukurba KR, DeLuca DS, Fromer M, Ferreira PG, Smith KS, Zhang R, Zhao F, Banks E, Poplin R, Ruderfer DM, Purcell SM, Tukiainen T, Minikel EV, Stenson PD, Cooper DN, Huang KH, Sullivan TJ, Nedzel J, GTEx Consortium, Geuvadis Consortium, Bustamante CD, Li JB, Daly MJ, Guigo R, Donnelly P, Ardlie K, Sammeth M, Dermitzakis ET, McCarthy MI, Montgomery SB, Lappalainen T, MacArthur DG. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. Science. 2015 May 8;348(6235):666–669. PMCID: PMC4537935

35. Aliouat A, Hatin I, Bertin P, François P, Stierlé V, Namy O, Salhi S, Jean-Jean O. Divergent effects of translation termination factor eRF3A and nonsense-mediated mRNA decay factor UPF1 on the expression of uORF carrying mRNAs and ribosome protein genes. RNA Biol. 2020 Feb;17(2):227–239. PMCID: PMC6973328

36. Rogalska ME, Vivori C, Valcárcel J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. Nat Rev Genet. 2023 Apr;24(4):251–269. PMID: 36526860

37. Childs-Disney JL, Yang X, Gibaut QMR, Tong Y, Batey RT, Disney MD. Targeting RNA structures with small molecules. Nat Rev Drug Discov. 2022 Oct;21(10):736–762. PMCID: PMC9360655

38. Bennett CF, Kordasiewicz HB, Cleveland DW. Antisense Drugs Make Sense for Neurological Diseases. Annu Rev Pharmacol Toxicol. 2021 Jan 6;61:831–852. PMCID: PMC8682074

39. Musunuru K, Chadwick AC, Mizoguchi T, Garcia SP, DeNizio JE, Reiss CW, Wang K, Iyer S, Dutta C, Clendaniel V, Amaonye M, Beach A, Berth K, Biswas S, Braun MC, Chen HM, Colace TV, Ganey JD, Gangopadhyay SA, Garrity R, Kasiewicz LN, Lavoie J, Madsen JA, Matsumoto Y, Mazzola AM, Nasrullah YS, Nneji J, Ren H, Sanjeev A, Shay M, Stahley MR, Fan SHY, Tam YK, Gaudelli NM, Ciaramella G, Stolz LE, Malyala P, Cheng CJ, Rajeev KG, Rohde E, Bellinger AM, Kathiresan S. In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. Nature. 2021 May;593(7859):429–434. PMID: 34012082

40. Davis JR, Wang X, Witte IP, Huang TP, Levy JM, Raguram A, Banskota S, Seidah NG, Musunuru K, Liu DR. Efficient in vivo base editing via single adeno-associated viruses with size-optimized genomes encoding compact adenine base editors. Nat Biomed Eng. 2022 Nov;6(11):1272–1283. PMCID: PMC9652153

41. Jafar-Nejad P, Powers B, Soriano A, Zhao H, Norris DA, Matson J, DeBrosse-Serra B, Watson J, Narayanan P, Chun SJ, Mazur C, Kordasiewicz H, Swayze EE, Rigo F. The atlas of RNase H antisense oligonucleotide distribution and activity in the CNS of rodents and non-human primates following central administration. Nucleic Acids Res. 2021 Jan 25;49(2):657–673. PMCID: PMC7826274

42. Reidenbach AG, Mesleh MF, Casalena D, Vallabh SM, Dahlin JL, Leed AJ, Chan AI, Usanov DL, Yehl JB, Lemke CT, Campbell AJ, Shah RN, Shrestha OK, Sacher JR, Rangel VL, Moroco JA, Sathappa M, Nonato MC, Nguyen KT, Wright SK, Liu DR, Wagner FF, Kaushik VK, Auld DS, Schreiber SL, Minikel EV. Multimodal small-molecule screening for human prion protein binders. J Biol Chem. 2020 Sep 25;295(39):13516–13531. PMCID: PMC7521658

43. Büeler H, Fischer M, Lang Y, Bluethmann H, Lipp HP, DeArmond SJ, Prusiner SB, Aguet M, Weissmann C. Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. Nature. 1992 Apr 16;356(6370):577–582. PMID: 1373228

44. Bremer J, Baumann F, Tiberi C, Wessig C, Fischer H, Schwarz P, Steele AD, Toyka KV, Nave KA, Weis J, Aguzzi A. Axonal prion protein is required for peripheral myelin maintenance. Nat Neurosci. 2010 Mar;13(3):310–318. PMID: 20098419

45. Benestad SL, Austbø L, Tranulis MA, Espenes A, Olsaker I. Healthy goats naturally devoid of prion protein. Vet Res. 2012;43:87. PMCID: PMC3542104

46. Richt JA, Kasinathan P, Hamir AN, Castilla J, Sathiyaseelan T, Vargas F, Sathiyaseelan J, Wu H, Matsushita H, Koster J, Kato S, Ishida I, Soto C, Robl JM, Kuroiwa Y. Production of cattle lacking prion protein. Nat Biotechnol. 2007 Jan;25(1):132–138. PMCID: PMC2813193

47. Minikel EV, Vallabh SM, Lek M, Estrada K, Samocha KE, Sathirapongsasuti JF, McLean CY, Tung JY, Yu LPC, Gambetti P, Blevins J, Zhang S, Cohen Y, Chen W, Yamada M, Hamaguchi T, Sanjo N, Mizusawa H, Nakamura Y, Kitamoto T, Collins SJ, Boyd A, Will RG, Knight R, Ponto C, Zerr I, Kraus TFJ, Eigenbrod S, Giese A, Calero M, de Pedro-Cuesta J, Haïk S, Laplanche JL, Bouaziz-Amar E, Brandel JP, Capellari S, Parchi P, Poleggi A, Ladogana A, O'Donnell-Luria AH, Karczewski KJ, Marshall JL, Boehnke M, Laakso M, Mohlke KL, Kähler A, Chambert K, McCarroll S, Sullivan PF, Hultman CM, Purcell SM, Sklar P, van der Lee SJ, Rozemuller A, Jansen C, Hofman A, Kraaij R, van Rooij JGJ, Ikram MA, Uitterlinden AG, van Duijn CM, Exome Aggregation Consortium (ExAC), Daly MJ, MacArthur DG. Quantifying prion disease penetrance using large population control cohorts. Sci Transl Med. 2016 Jan 20;8(322):322ra9. PMCID: PMC4774245

48. Minikel EV, Karczewski KJ, Martin HC, Cummings BB, Whiffin N, Rhodes D, Alföldi J, Trembath RC, van Heel DA, Daly MJ, Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Schreiber SL, MacArthur DG. Evaluating drug targets through human loss-of-function genetic variation. Nature. 2020 May;581(7809):459–464. PMCID: PMC7272226

49. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996–1006. PMCID: PMC186604

50. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012 Jun 28;9(7):676–682. PMCID: PMC3855844

## SUPPLEMENT

Supplementary tables 1-6 provided as a separate Excel file.

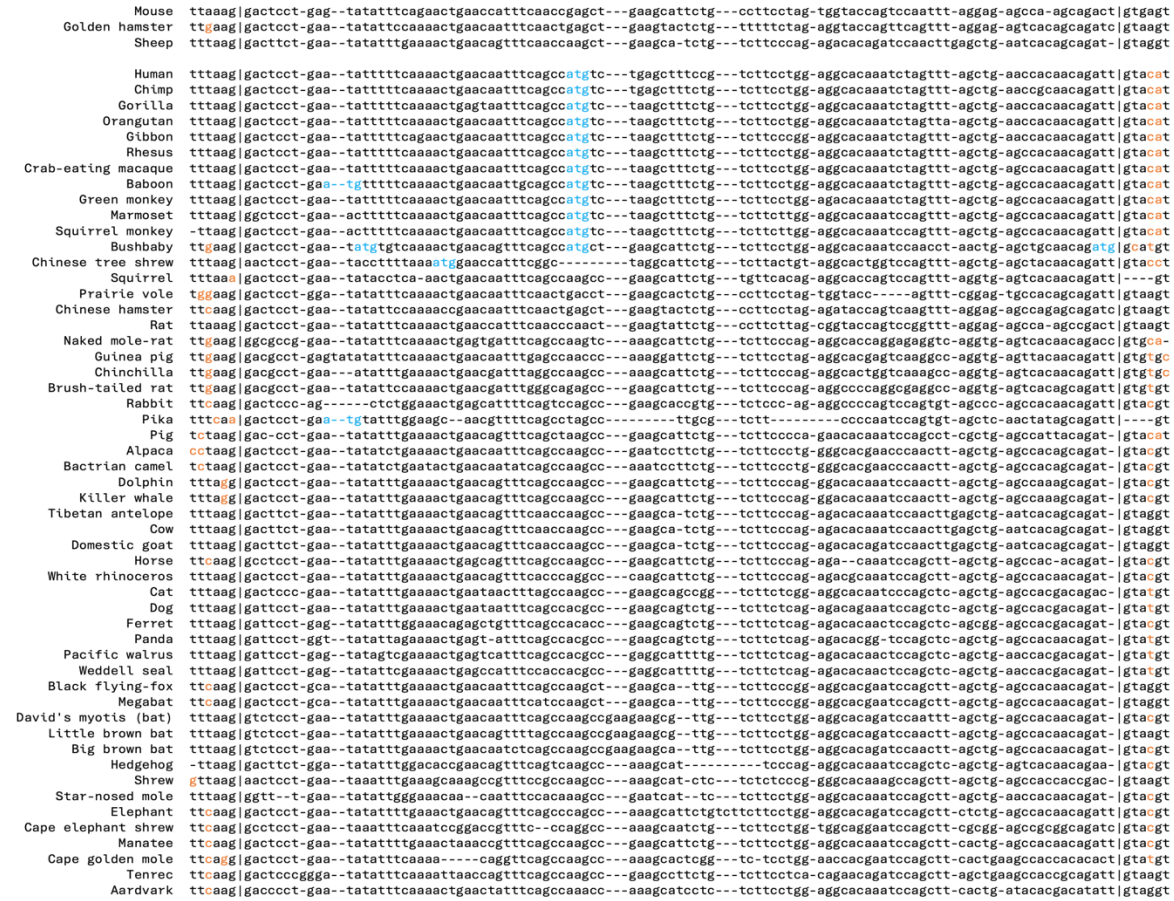



***Figure S1. Multiple alignment of PRNP exon 2 orthologous sequence for all available eutherian mammals.*** *As in Figure 1D, but including eutherian mammals without ATGs in exon 2. Lesser Egyptian jerboa is excluded because orthologous sequence was identified for only part of exon 2.*

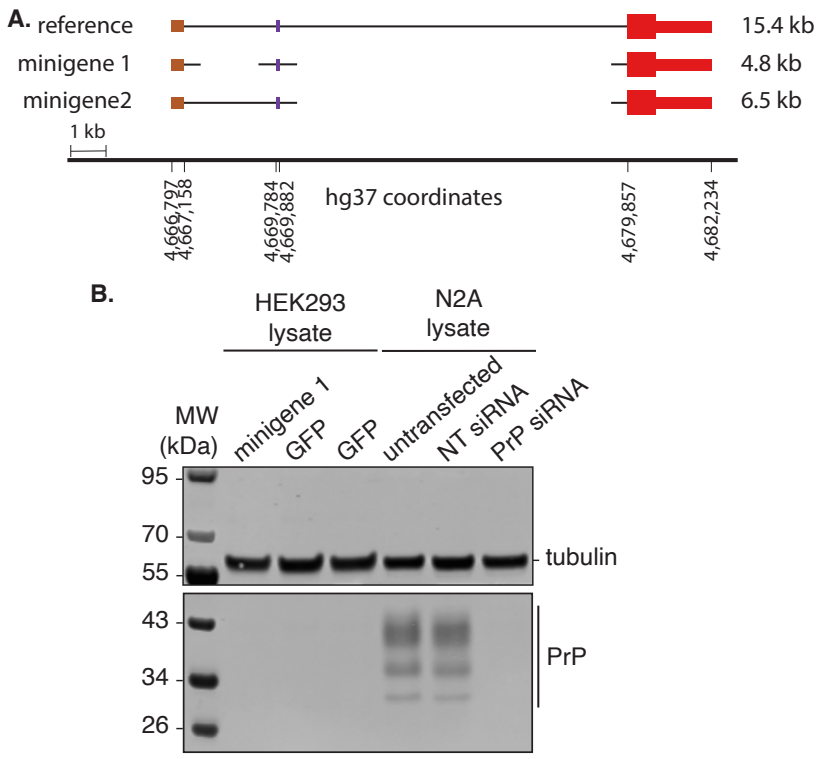



***Figure S2. Alternative minigene construct tested in cells. A)*** *Comparison of human reference sequence with an alternative "minigene 1" containing only 500 bp at either end of intron 1, and the "minigene 2" used throughout the main text of this manuscript.* ***B)*** *Immunoblot failing to detect any expression of minigene 1 in transfected HEK293 cells. Primary antibody: 6D11, see Methods.*