NEURODEGENERATION

Creation of de novo cryptic splicing for ALS and FTD precision medicine

Oscar G. Wilkins[1,2]*, Max Z. Y. J. Chien[1,2]†, Josette J. Wlaschin[3,4]†, Simone Barattucci[1], Peter Harley[1], Francesca Mattedi[1], Puja R. Mehta[1], Maria Pisliakova[1,2], Eugeni Ryadnov[1], Matthew J. Keuss[1], David Thompson[5], Holly Digby[2,6], Lea Knez[1,2], Rebecca L. Simkin[1], Juan Antinao Diaz[7], Matteo Zanovello[1,2], Anna-Leigh Brown[1], Annalucia Darbey[1], Rajvinder Karda[7], Elizabeth M. C. Fisher[1], Thomas J. Cunningham[5,8], Claire E. Le Pichon[3], Jernej Ule[2,6], Pietro Fratta[1,2]*

Loss of function of the RNA-binding protein TDP-43 (TDP-LOF) is a hallmark of amyotrophic lateral sclerosis (ALS) and other neurodegenerative disorders. Here we describe TDP-REG, which exploits the specificity of cryptic splicing induced by TDP-LOF to drive protein expression when and where the disease process occurs. The SpliceNouveau algorithm combines deep learning with rational design to generate customizable cryptic splicing events within protein-coding sequences. We demonstrate that expression of TDP-REG reporters is tightly coupled to TDP-LOF in vitro and in vivo. TDP-REG enables genomic prime editing to ablate the UNC13A cryptic donor splice site specifically upon TDP-LOF. Finally, we design TDP-REG vectors encoding a TDP-43/Raver1 fusion protein that rescues key pathological cryptic splicing events, paving the way for the development of precision therapies for TDP43-related disorders.

Amyotrophic lateral sclerosis (ALS) is a devastating and incurable neurodegenerative disease. In 97% of cases, there is pronounced formation of neuronal cytoplasmic aggregates of the RNA-binding protein TDP-43 (*1*). Further, such pathology is found in ~45% of frontotemporal dementia (FTD) cases and is also observed in limbic-predominant age-related TDP-43 encephalopathy (LATE) and Alzheimer's disease, suggesting that TDP-43–mediated RNA dysregulation is a critical element in many neurodegenerative diseases (*2*, *3*).

TDP-43 is a key regulator of splicing, protecting the transcriptome from toxic "cryptic exons" (CEs), which become prominently expressed upon TDP-43 loss of function (TDP-LOF); CEs typically introduce premature termination codons (PTCs) into transcripts, preventing the expression of crucial proteins including STMN2 and UNC13A (*4*–*9*). We and others previously showed that genetic modulation of even a single CE can influence disease progression, meaning that therapeutics active in the stage in which CEs are expressed could still influence disease course (*7*, *8*). Numerous preclinical studies aim to reduce TDP-43–associated toxicity, for example, through blocking CE inclusion using antisense oligonucleotides (ASOs) or transgenes, or by targeting the aggregation process itself (*4*, *10*–*12*).

A major barrier to bringing gene therapy approaches to the clinic is the lack of methods to tightly regulate transgene expression. Only a small fraction of motor neurons, let alone all neurons, display clear pathology at any given time in patients (*13*). Therefore, a conventional targeting approach [for example, combining an adeno-associated virus (AAV) serotype with central nervous system (CNS) tropism with a neuronal promoter] would result in expression of the therapeutic transgene in a vast number of nondegenerating cells. Such expression is not merely unnecessary but could worsen prognosis by interfering with the homeostasis of otherwise healthy cells, especially given the permanent nature of these approaches (*14*). Such concerns recently led to the development of therapeutics with activity-dependent transcriptional promoters for epilepsy (*15*).

Here, we leverage our understanding of TDP-43's regulation of splicing together with deep-learning–based splicing prediction tools to generate vectors featuring novel cryptic splice sites. Crucially, whereas pathological cryptic splicing blocks protein expression, our vectors explicitly require the use of the cryptic splice site(s) to express the encoded transgene. This approach limits expression of transgenes to cells with TDP-LOF, which co-occurs with TDP-43 aggregation in patients, thus reducing the risk of off-target side effects. In addition to demonstrating highly effective TDP-LOF–dependent expression in vitro and in vivo, we apply this approach to two candidate ALS and FTD gene therapies, paving the way toward safer and more efficacious treatments for these devastating diseases.

Modification of ALS and FTD CEs for TDP-43–regulated expression vectors

Numerous TDP-43–regulated CEs have been reproducibly detected in cellular and animal models, showing high specificity for TDP-LOF; furthermore, a number of these, including CEs in *STMN2*, *UNC13A*, *HDGFL2*, and *AARS1*, have been found specifically within neurons with TDP-43 pathology in postmortem ALS and FTD samples (*4*–*8*, *16*). We reasoned that a CE with these features could be used to regulate expression of a therapeutic or reporter transgene (Fig. 1A). We identified the CE in the *AARS1* gene (hg38: chr16:70272796-70272882) as the most suitable candidate because of its reproducible detection, its lack of stop codons in at least one frame, and its short, defined TDP-43 binding motif (Fig. 1B and fig. S1). We added an extra residue to the CE to cause frameshifting and removed sections of the flanking introns that were likely unnecessary for regulation by TDP-43 (Fig. 1B; see materials and methods). We then placed this sequence between an upstream start codon and downstream transgene such that only when the CE was included would the start codon be in-frame with the transgene (Fig. 1B). We included a P2A "self-cleavage" site so that the upstream peptide encoded by the CE is released separately from the main protein (Fig. 1B). Furthermore, we added a constitutively spliced intron derived from *RPS24* to promote nonsense-mediated decay (NMD) if a PTC is encountered. We refer to this type of construct as TDP-REGv1.

We used mCherry as the transgene to visualize whether this system restricted expression to cells with TDP-LOF. This reporter construct, along with a positive control in which the CE sequence is constitutively expressed, was transfected into SK-N-BE(2) cells (a human neuroblastoma cell line) with doxycycline-inducible TDP-43 knockdown. A >16-fold increase in mCherry expression was detected in cells with TDP-43 knockdown when using the cryptic construct (Fig. 1, C and D). In agreement with this, Nanopore sequencing revealed a >43-fold increase in CE inclusion upon TDP-43 knockdown (Fig. 1E). We noted that in cells with normal TDP-43, mild leaky expression of mCherry protein was detectable, even though CE inclusion was less than 0.5%, suggesting a mechanism of protein translation that circumvents the presence of the upstream frameshift. Using Western blotting, we observed low-degree leaky expression even in a vector in which the CE is deleted, potentially due to use of alternative

[1]UCL Queen Square Motor Neuron Disease Centre, Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, London WC1N 3BG, UK. [2]The Francis Crick Institute, London NW1 1AT, UK. [3]*Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. [4]Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA. [5]Mammalian Genetics Unit, MRC Harwell Institute, Oxfordshire OX11 0RD, UK. [6]UK Dementia Research Institute at King's College London, London SE5 9RX, UK. [7]EGA-Institute for Women's Health, University College London, London WC1E 6HX, UK. [8]MRC Prion Unit at UCL and UCL Institute of Prion Diseases, London W1W 7FF, UK.
***Corresponding author. Email: p.fratta@ucl.ac.uk (P.F.); oscar.wilkins.18@ucl.ac.uk (O.G.W.)**
†These authors contributed equally to this work.

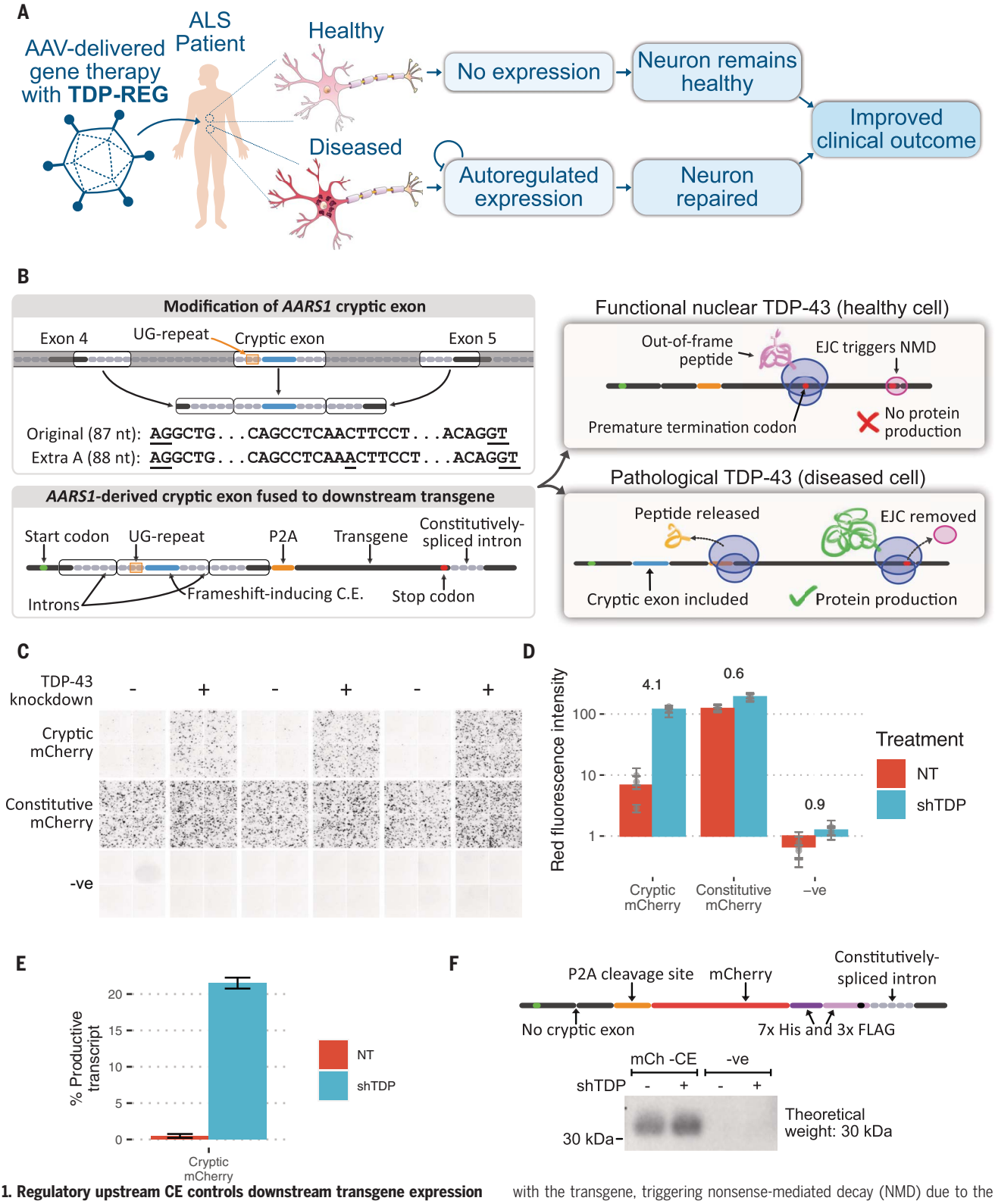

Fig. 1. Regulatory upstream CE controls downstream transgene expression (TDP-REGv1). (**A**) Schematic showing the intended function and purpose of this work. (**B**) Schematic of TDP-REGv1. (Top left) The genomic locus surrounding the human *AARS1* CE region was modified, with the middle parts of the introns removed to reduce size and an extra adenosine added to the CE sequence to enable frameshifting. The AG and GT splice sites of the CE are underlined. (Bottom left) The modified sequence from (A) was incorporated into a "minigene." (Top right) When the CE is excluded, the start codon is out-of-frame with the transgene, triggering nonsense-mediated decay (NMD) due to the downstream exon junction complex (EJC). (Bottom right) When the CE is included, the transgene is in-frame with the start codon, resulting in protein expression. nt, nucleotide. (**C**) Fluorescence microscopy images (red channel) showing SK-N-BE(2) cells with (shTDP) or without (NT, "not treated") TDP-43 knockdown, transfected with a vector fusing the upstream *AARS1*-derived sequence to mCherry ("cryptic mCherry") or a constitutive vector. (**D**) Quantification of the images in (C); numbers show $\log_2$ fold change in TDP-43 knockdown cells;

each dot shows the average of one well (three wells per condition), with error bars showing the standard deviation within each well (four per well). (**E**) Summary of Nanopore sequencing results for the cells in (C); error bars show the standard deviation across three replicates. (**F**) (Top) Schematic of mCh -CE vector, without the CE but with a downstream His/Tri-FLAG dual tag to enable sensitive detection. (Bottom) Western blot of cells transduced with the above vector ("mCh -CE") or a completely different vector ("-ve" denotes a PE vector). Samples were enriched with a His-tag pulldown, then blotted with an anti-FLAG antibody.

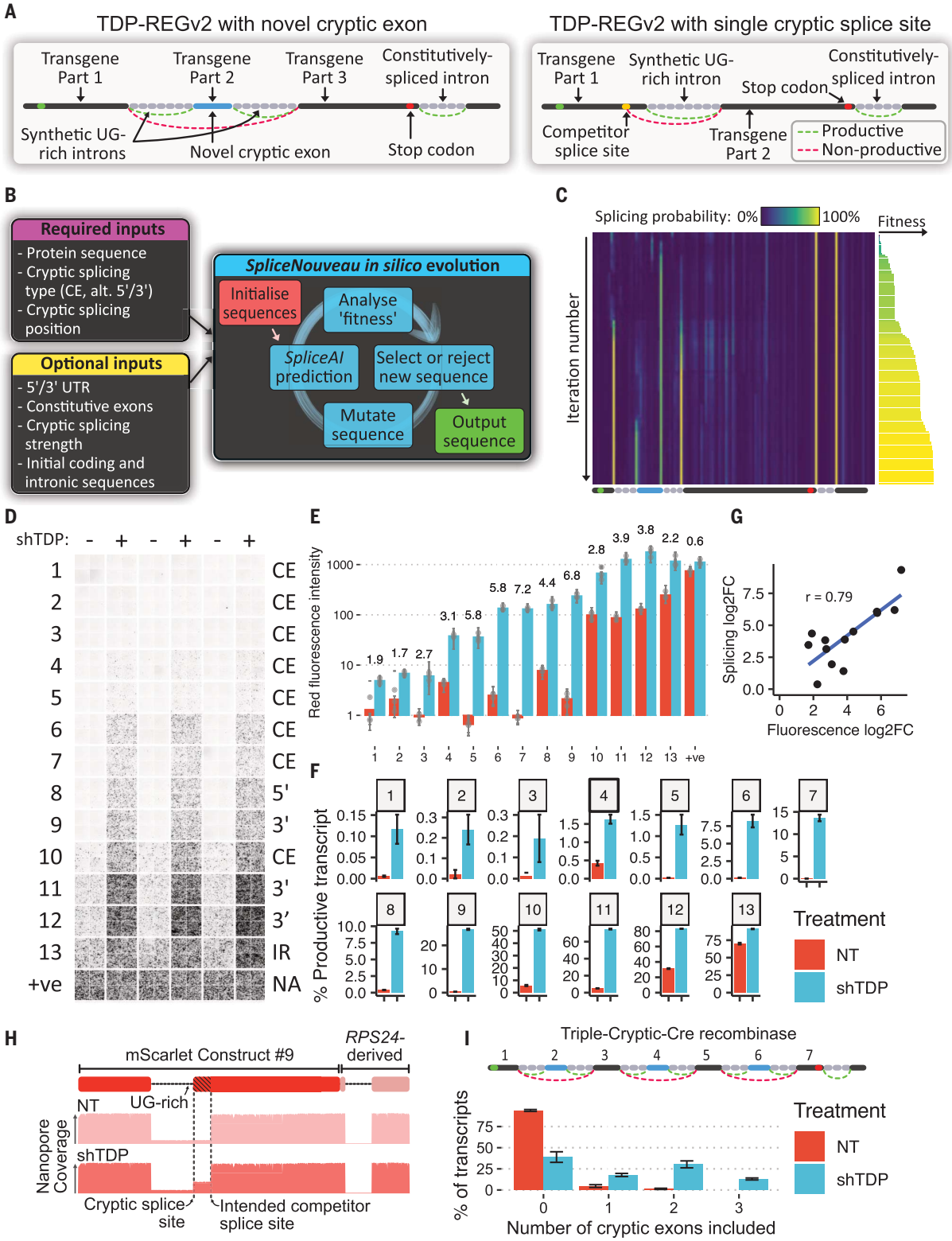

Fig. 2. Deep-learning-guided design of novel cryptic splicing events (TDP-REGv2). (**A**) Diagrams of internal CE (left) and single intron (right) designs; for the single intron design, an alternative 5′ splicing design is shown. UTR, untranslated region. (**B**) Schematic of the SpliceNouveau algorithm for

designing new cryptic splicing-dependent expression vectors. (**C**) Heatmap showing the in silico evolution trajectory for an example internal CE design (below), with associated "fitness" of each (right), which was initialized with a constitutively spliced intron from *RPS24* at the 3′ end. As the iteration number increases, splice sites are "evolved" at the desired positions, and off-target splice sites are depleted. The splice sites flanking the CE are weaker than the constitutive splice sites, as specified by the user. (**D**) Fluorescence microscopy images showing mScarlet expression for 13 constructs generated by SpliceNouveau and a positive control. Each replicate consists of four images taken from different parts of the well. (**E**) Quantification of (D); each dot represents the average of the four images for each replicate; error bars show the standard deviation of images in each well; numbers refer to $\log_2$ fold change in signal intensity compared with control. (**F**) Percentage of productive transcripts (transcripts that are predicted to produce full-length mScarlet), as determined by Nanopore sequencing; error bars show standard deviation across three replicates. (**G**) Correlation of fold changes (FC) in fluorescence and productive isoform fraction; Pearson correlation shown. (**H**) Diagram of mScarlet construct 9 (top); representative Nanopore pileup with and without TDP-43 knockdown (bottom). (**I**) Schematic of construct encoding Cre recombinase, split across seven exons; exons 2, 4, and 6 are flanked by UG-rich regions (top); number of CEs included in each transcript without and with shTDP, assessed by Nanopore sequencing (bottom); error bars show standard deviation across three replicates.

transcriptional start sites or leaky ribosomal scanning (Fig. 1F).

Creation of de novo cryptic splicing sequences

As a next step, we reasoned that if new CEs could be created within the transgene-encoding region of the vector, this would remove any risk of leaky expression when the CE is spliced out because the full, uninterrupted transgene coding sequence would not be present in the mature mRNA. Further, this approach would avoid expression of unwanted upstream peptides and reduce the vector size. In pilot tests, we found that the *AARS1*-derived CE could be replaced with other sequences and that the CE strength of these sequences correlated with SpliceAI splicing predictions (fig. S2, A to D) (*17*). Encouraged by this, we built an in silico directed evolution algorithm, SpliceNouveau, which combines rational design principles with SpliceAI predictions, to design de novo TDP-43–regulated cryptic cassette exons or single cryptic splice sites (Fig. 2, A to C). TDP-43 preferentially binds to stretches of UG repeats (*18*, *19*). We therefore programmed SpliceNouveau to incorporate different types of UG-rich regions into our designs, including upstream UG-repeats (as in *AARS1*), downstream UG repeats (as in *Sars*), upstream and downstream repeats (as in *Smg5*) or UG-rich regions without extended UG repeats (as in *UNC13A*) (fig. S3A).

To create single-intron vectors, which can be smaller than those based on CEs flanked by two introns, we modified our algorithm to design candidate alternative 5′ or alternative 3′ splice sites, which compete with the designed cryptic splice site (Fig. 2A). In each case, the single intron was heavily enriched for TDP-43 binding sites; for alternative 3′ splice sites, the algorithm optimized part of the coding sequence to produce a polypyrimidine-rich region. Using SpliceNouveau, we also attempted to design variants exhibiting TDP-43–regulated intron retention by specifying weaker splice sites with lower SpliceAI scores combined with a single UG-rich intron. We refer to constructs generated using this approach as TDP-REGv2.

Of the 27 designs that were tested in SK-N-BE(2) cells, 13 resulted in increased expression upon TDP-43 knockdown (Fig. 2, D to H). Many constructs achieved greater dynamic range than our initial *AARS1*-based design, with two constructs achieving >100-fold increases in expression. The maximal expression varied greatly across constructs, from >100-fold less to greater than the positive control, thus demonstrating that expression can be fine-tuned (Fig. 2E). Using targeted Nanopore sequencing of 56 vectors, we confirmed that those with higher optimization were more likely to be spliced in the expected manner (fig. S4). Overall, only a modest number of vectors need to be experimentally screened to identify those with the desired splicing properties.

Seven of the successful designs were further tested in an SH-SY5Y cell line (human neuroblastoma cell line) with inducible TDP-43 knockdown (*7*), which yielded very similar results (fig. S3, B and C). To demonstrate that the splicing regulation was due specifically to TDP-43 depletion, we cotransfected a selection of the above constructs into SK-N-BE (2) cells with TDP-43 knockdown, together with either functional TDP-43/Raver1 fusion protein, which has been shown to rescue TDP-LOF, or an inactive TDP-43/Raver1 fusion protein with impaired binding to RNA ("2FL") (*4*, *18*). Cells cotransfected with the inactive 2FL mutant exhibited greatly increased fluorescence compared to cells cotransfected with functional TDP-43/Raver1 (fig. S5A). Furthermore, whereas short hairpin RNAs (shRNAs) against TDP-43 substantially increased fluorescence, shRNAs against other key splicing regulators did not (fig. S5, B and C).

Multiple CEs further increase specificity

Next, we examined whether multiple CEs could be included within the same construct, further reducing the risk of leaky expression. We designed a construct encoding Cre recombinase enzyme split across seven exons, three of which were cryptic (Fig. 2I). Nanopore sequencing revealed that in untreated cells, <0.05% of transcripts featured all three CEs, whereas shTDP cells expressed all three CEs in ~10% of transcripts (Fig. 2I). Furthermore, lower leaky expression was detected for the construct with three CEs versus constructs with only one or two CEs (fig. S6).

TDP-REG is activated by TDP-LOF in vivo

We then examined whether TDP-REG constructs are also functional in vivo in mammalian spinal cord, the main target of ALS therapeutic approaches. We injected AAVs (PHP.eB serotype with hSynapsin promoter) containing TDP-REGv1 mCherry or TDP-REGv2 mScarlet (construct 7) into TDP-43 conditional knockout (cKO) mice (TDP-43$^{Fl/Fl}$;Chat-Cre$^{+/wt}$), where TDP-43 loss was driven by a choline acetyltransferase (ChAT)–Cre driver line and therefore directed, within the spinal cord, to motor neurons and other spinal cholinergic cells.

We observed pronounced expression of mCherry or mScarlet in the motor neurons of these mice, with at least 50% of identified motor neurons having clear expression in all but one case for each construct (Fig. 3, A, C, and E, and fig. S7A). By contrast, control mice (TDP-43$^{Fl/wt}$;Chat-Cre$^{+/wt}$) displayed very little expression of mCherry, with between 0 and 2% of identified motor neurons having detectable mCherry or mScarlet expression (Fig. 3, B, D, and E, and fig. S7A). A positive control, mScarlet AAV without TDP-REG, showed no such specificity when injected into control mice, instead strongly expressing mScarlet in the majority of TDP-43–positive motor neurons (Fig. 3E and fig. S7B).

TDP-43 aggregation activates TDP-REG expression

We then tested whether TDP-REG vectors could be activated by TDP-43 cytoplasmic aggregation, which would more accurately simulate the disease process than TDP-43 knockdown. We cotransfected TDP-REGv2: mScarlet 7 (and an mGreenLantern transfection control) into human embryonic kidney 293T (HEK293T) cells with an aggregation-prone version of TDP-43 in which the Q/N-rich

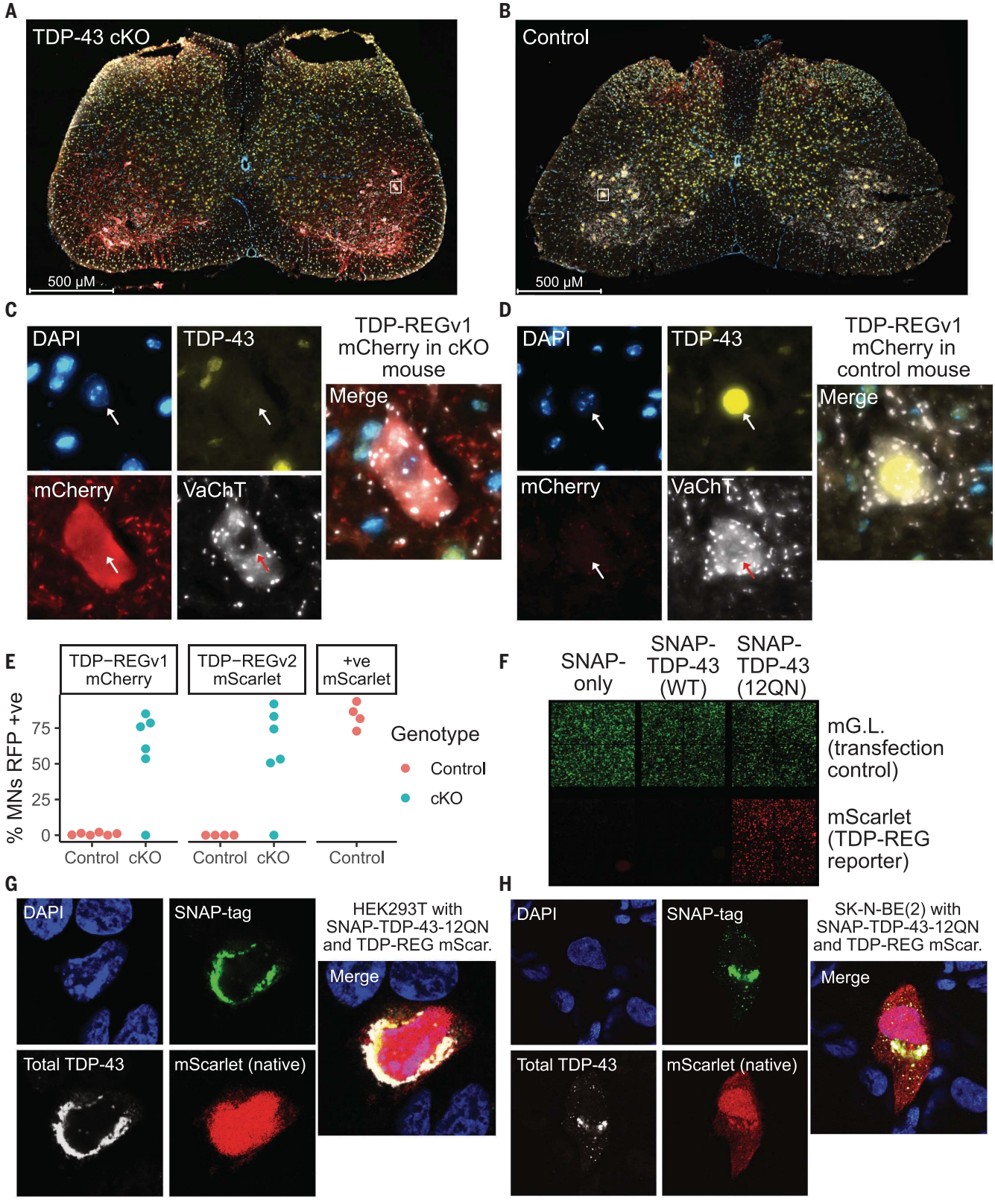

Fig. 3. Functionality of TDP-REG in different biological contexts. (**A**) Spinal cord of a TDP-43 cKO mouse injected with TDP-REGv1 mCherry AAV [blue, DAPI (4′,6-diamidino-2-phenylindole); yellow, TDP-43; red, mCherry; white = VAChT (vesicular acetylcholine transporter)]. (**B**) Equivalent to (A), but with a control mouse. (**C** and **D**) Magnified representative motor neurons from figures in (A) and (B), respectively; white boxes show regions of magnification. (**E**) Quantification of the percentage of motor neurons (MNs) with clear mCherry (TDP-REGv1 mCherry), mScarlet (TDP-REGv2 mScarlet no. 7, see fig. S7A), or

positive control mScarlet for cKO and control mice (only control mice for positive control AAV; see fig. S8B); N = 4 to 6 mice per condition. (**F**) Representative fluorescence microscopy images of HEK293T cells transduced with TDP-REGv2:mScarlet no. 7, a constitutive mGreenLantern vector, and expression vectors for SNAP-tag only, SNAP-tag-TDP-43(wild-type), or SNAP-tag-TDP-43(12QN mutant); see also fig. S7C for N = 6 replicates. (**G** and **H**) Confocal microscopy of HEK293T and SK-N-BE(2) cells, respectively, cotransfected with TDP-REGv2:mScarlet reporter (no. 7) and SNAP-TDP-43-12QN.

domain is repeated 12 times (SNAP-TDP-43-12QN) (*20*). We observed strong expression of mScarlet induced by SNAP-TDP-43-12QN but not by SNAP-TDP-43 (wild-type) or SNAP-tag alone (Fig. 3F and fig S7C). Using confocal microscopy, we confirmed that, in a subset of cells, SNAP-TDP-43-12QN was highly expressed and heavily enriched in the cytoplasm, activating TDP-REG mScarlet (construct 7) expression (Fig. 3, G and H). Therefore, TDP-REG is functional in vivo and is activated by TDP-43 aggregation, strongly suggesting this approach will function within ALS and FTD patients.

Application of TDP-REG to biomarker expression

Although TDP-LOF is an established key feature of ALS progression, there is a lack of tools to detect it in cell and animal models of disease, thus limiting preclinical studies. In addition to the fluorescent reporters described above, we generated a sensitive luminescent biomarker that could potentially be adapted to monitor TDP-LOF in organoid and animal models. We fused the Gaussia Princeps luciferase (Gluc) sequence downstream of the *AARS1*-based minigene described above (TDP-REGv1) and also used SpliceNouveau to create five vectors encoding Gluc with an internal, synthetic CE (TDP-REGv2). All of these vectors featured increased expression of productively spliced transcripts upon TDP-43 knockdown (fig. S8A). Both the TDP-REGv1 vector and the best TDP-REGv2 vectors featured high CE inclusion upon TDP-43 knockdown with minimal leakiness, but the best-performing TDP-REGv2 vector had better dynamic range (Fig. 4, A and B): It led to a >200-fold increase in Gluc expression upon TDP-43 knockdown in SK-N-BE(2) cells, with negligible leaky expression (Fig. 4C). These vectors could enable sensitive detection of TDP-LOF in disease models and high-throughput screens.

TDP-REG tightly regulates genome editing

We explored the possibility of performing genome editing to remove cryptic splice sites. The prime editing (PE) approach, which avoids the need for highly mutagenic double-stranded breaks, can be adapted for high-efficiency in vivo editing of the CNS (*21*, *22*). However, even PE can lead to unwanted editing events that could potentially be toxic (*23*). Using SpliceNouveau, we designed a TDP-REGv2 vector encoding an optimized PE construct ("PE-Max") in which part of the Cas9 sequence was encoded by a TDP-REG CE (Fig. 4D) (*24*). Using reverse transcription–quantitative polymerase chain reaction (RT-PCR) and Western blotting, we confirmed that the CE had near-undetectable leaky expression (Fig. 4, E and F).

We then designed a prime editing guide RNA (pegRNA) to edit the donor splice site of the *UNC13A* CE (Fig. 4G) (*7*, *8*), combined with a nicking single guide RNA to improve editing efficiency (*21*). We tested editing efficiency with or without TDP-43 knockdown in SK-N-BE(2) cells. The constitutive PE-Max vector led to editing regardless of TDP-43 knockdown, whereas the CE-containing vector tightly restricted editing to cells with TDP-43 knockdown (Fig. 4H). Thus, TDP-REG can ensure that genome editing occurs only in cells where it is beneficial, minimizing the potential for off-target effects when delivered to all cells.

TDP-REG can enable autoregulated splicing rescue

Finally, we used our system to directly rescue TDP-LOF. Fusions of TDP-43 to a less aggregation-prone splicing repressor domain (such as from *RAVER1*) can partially rescue TDP-43 splicing function (*4*). However, even small imbalances in TDP-43 expression can cause substantial toxicity, and therefore despite its lesser aggregation, generalized expression of TDP-43/Raver1 fusions is undesirable (*25*). We used SpliceNouveau to design TDP-REGv2 constructs encoding TDP-43 fused to the *Raver1* splicing repressor domain (*TDP-REGv2:Raver1*) featuring an internal CE (Fig. 4I). To ensure that expression of the synthetic CE occurred upon even a mild decrease in TDP-43 function, thus enabling rescue at the first signs of TDP-43 pathology, we specified high SpliceAI scores (60 to 100%) for the CE splice sites and used short UG repeats to limit the binding affinity of the introns to TDP-43.

TDP-43/Raver1 fusion protein mimics TDP-43's splicing repressor function and would therefore be expected to autoregulate by inhibiting expression of the synthetic CEs in these constructs. Although this is a great advantage therapeutically, as it would prevent overexpression of TDP-43/Raver1, it could also mask the responsiveness of our synthetic CEs to TDP-LOF during these initial tests. We therefore used an inactive 2FL mutant during initial RT-PCR screens (*18*). All but one construct displayed detectable CE expression upon TDP-43 knockdown, with six showing high expression; presumably owing to the use of high SpliceAI scores and relatively short UG-rich regions, some vectors had detectable leaky CE expression (Fig. 4J).

Using the piggyBac system (see materials and methods), we generated polyclonal SK-N-BE(2) lines expressing TDP-43/Raver1 fusion in triplicate, either as a constitutive version or with an internal CE from constructs 6 or 9, or mScarlet. In this case, the wild-type TDP-43 sequence was used to enable both repression of cryptic splicing and autoregulation. Whereas the constitutive vector expressed TDP-43/Raver1 regardless of endogenous TDP-43 amounts, vectors featuring an internal CE only expressed detectable TDP-43/Raver1 upon knockdown of endogenous TDP-43 (Fig. 4K and fig. S8B). We then generated an AAV featuring the most promising vector (no. 9) and, using Nanopore sequencing, assessed its behavior in the spinal cord of wild-type and TDP-43 cKO mice. We observed a substantial increase in expression of the TDP-43–encoding CE in cKO mice compared with WT mice (Fig. 4L).

Next, we assessed whether these vectors rescue key human cryptic splicing events. *TDP-REGv2:Raver1* vectors were able to repress the *UNC13A* and *AARS1* CEs by >85%, and the *STMN2* CE by ~40% (Fig. 4, M and N). Furthermore, whereas knockdown of TDP-43 in cells expressing mScarlet featured essentially undetectable STMN2 protein, cells with TDP-43/Raver1 vectors displayed a rescue consistent with that observed for splicing (Fig. 4K and fig. S8C).

The *UNC13A* CE leads to the loss of a crucial presynaptic protein. We therefore assessed its rescue by these constructs in induced pluripotent stem cell (iPSC)–derived cortical neurons, which form synapses. *TDP-REGv2:Raver1* vectors were able to rescue UNC13A cryptic splicing in iPSC-derived neurons and also rescued *UNC13A* protein at the synapses (Fig. 4O and fig. S9, A to C).

Higher levels of TDP-43/Raver1 expression, and more efficient repression of endogenous CEs, were observed upon TDP-43 knockdown for cryptic vectors (especially no. 9) than for the constitutive vector (Fig. 4, K and M to N). We propose that this is due to toxicity of TDP-43/Raver1 when expressed in cells with functional endogenous TDP-43, leading to a selection pressure against cells with high constitutive expression of TDP-43/Raver1. Indeed, a strong growth advantage was observed for cells expressing TDP-43/Raver1 regulated by TDP-REG compared with those expressing constitutive TDP-43/Raver1, which severely affected growth in a dose-dependent manner

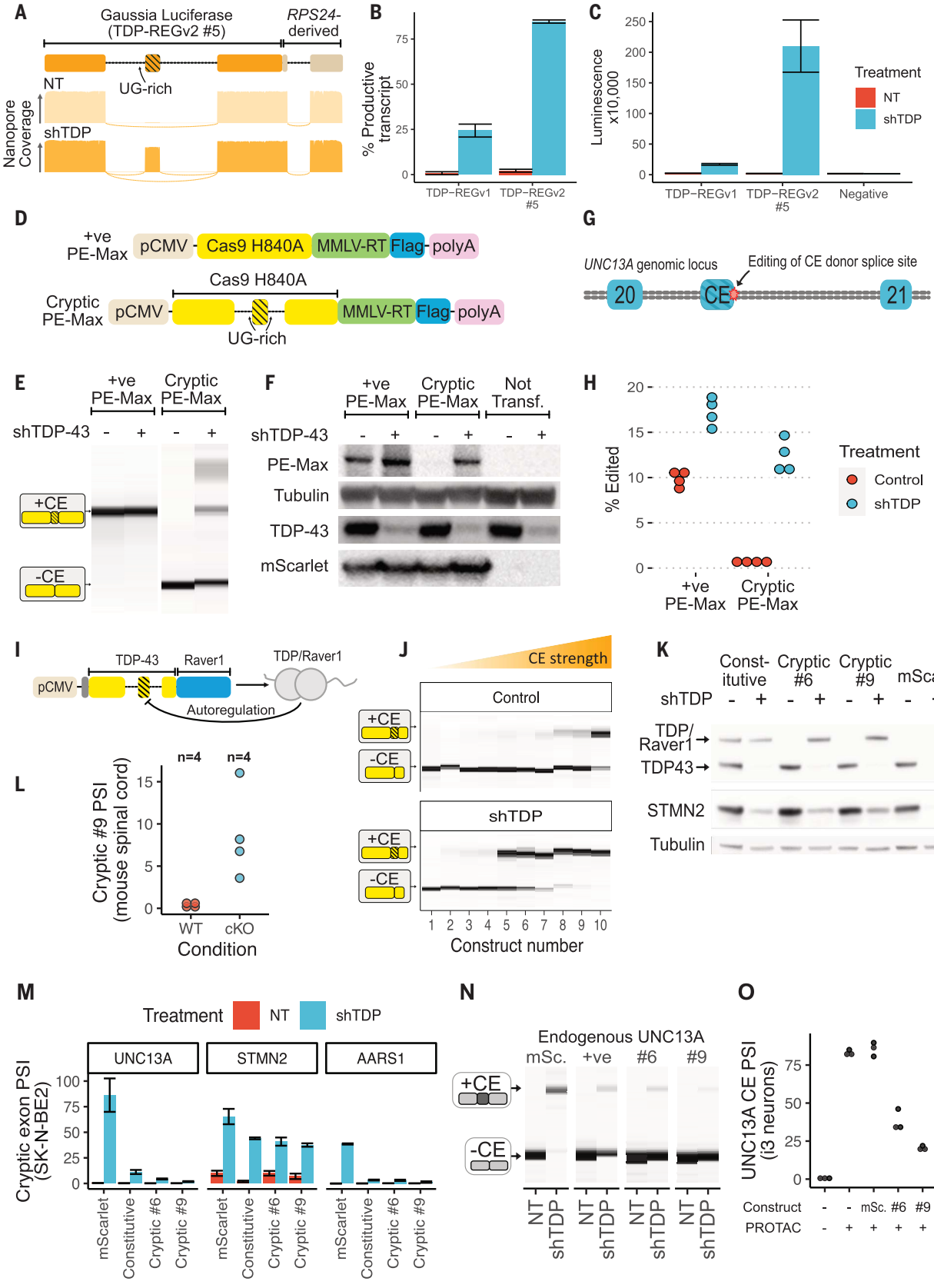

Fig. 4. Biosensors, genome editing, and splicing rescue. (**A**) Schematic of TDP-REGv2 Gluc vector #5 (top); representative Nanopore pileups with (shTDP) and without (NT, "not treated" with doxycycline) TDP-43 knockdown (bottom). (**B**) Quantification of Nanopore data for Gluc TDP-REGv1 and TDP-REGv2 #5 for SK-N-BE(2) cells with and without TDP-43 knockdown; error bars show standard deviation across four replicates. (**C**) Quantification of luminescence from

supernatant of cells used for (B); error bars show standard deviation across four replicates. (**D**) Schematic of constitutive and cryptic (TDP-REGv2) PE vector (based on Addgene PE-Max vector). pCMV, cytomegalovirus promoter; MMLV-RT, Moloney murine leukemia virus reverse transcriptase. (**E**) Capillary electrophoresis results from RT-PCR of cryptic or constitutive PE-Max vector transfected into SK-N-BE(2) cells with or without TDP-43 knockdown. (**F**) Western blot of FLAG-tagged PE-Max vectors (cotransfected with FLAG-tagged mScarlet) in SK-N-BE(2) cells with or without TDP-43 knockdown; representative blot of $N = 3$ blots. (**G**) Diagram of *UNC13A* genomic locus surrounding the *UNC13A* CE between exons 20 and 21; the position of genome editing is shown. (**H**) Quantification of intended editing of *UNC13A* in SK-N-BE(2) cells with either vector, +/− TDP-43 knockdown, assessed by targeted Nanopore sequencing. (**I**) Schematic of TDP-43/Raver1 with an internal CE. The design features an additional dual N-terminal nuclear localization signal; the constitutive vector encodes the same amino acid sequence, but with the introns removed. (**J**) RT-PCR analysis of 10 constructs designed [as in (A)], using the 2FL mutation to block autorepression of the CE; constructs were transfected into SK-N-BE(2) cells without (NT) or with (shTDP) doxycycline-induced TDP-43 knockdown; representative traces from $N = 3$ traces. (**K**) Western blot of SK-N-BE(2) cells stably expressing constitutive or cryptic (vectors 6 and 9) TDP-43/Raver1, or mScarlet, with or without endogenous TDP-43 knockdown. Endogenous TDP-43 and TDP-43/Raver1 are labeled in the anti-TDP-43 blot; STMN2 expression is visible in the middle blot; tubulin loading control is shown below. All lines are polyclonal piggyBac lines; three polyclonal lines were made per construct with consistent results; representative blot of $N = 3$ blots. (**L**) Nanopore-based quantification of splicing of vector 9 when expressed in the spinal cord (via AAV injection) of wild-type or TDP-43 cKO mice. (**M**) Quantification of UNC13A, STMN2, and AARS1 cryptic splicing, using all three polyclonal replicates; error bars show standard deviation across three replicates (one replicate for mScarlet control "NT" was excluded because the RNA failed quality control). PSI, percent spliced in. (**N**) RT-PCR analysis of endogenous UNC13A cryptic splicing for the same lines as in (K), showing results from a single replicate. (**O**) Quantification of RT-PCRs against UNC13A CE in polyclonal cortical i3 neuron lines, derived from a clonal line in which TDP-43 is fused to HaloTag enabling PROTAC-mediated inducible degradation, expressing mScarlet or TDP/Raver1 vectors 6 or 9.

(fig. S10). Thus, *TDP-REGv2:Raver1* vectors can rescue endogenous CEs while avoiding toxicity due to constitutive TDP-43/Raver1 protein expression.

Discussion

We have developed TDP-REG, which enables disease-specific spatial and temporal regulation of gene therapy expression by exploiting the predictable splicing changes induced by TDP-LOF. Given that TDP-LOF is a hallmark of ALS, FTD, and other common neurodegenerative diseases including Alzheimer's and LATE, TDP-REG could have broad utility in the treatment of these disorders. Disease-induced activation of gene therapies at single-cell resolution could help mitigate the potential risks of permanent transgene expression in patients (*14*). Furthermore, in at-risk individuals carrying causal genetic variants of ALS, the spatial and temporal specificity of TDP-REG could allow the therapeutics to be delivered at the presymptomatic stage, lying dormant until the very first stages of TDP-43 pathology are detected. Additionally, TDP-REG can be used during the preclinical phase of drug development as a real-time readout for TDP-43 pathology in cells or even live animals.

TDP-REGv1 involves the fusion of a transgene to an upstream regulatory module, featuring a modified CE sequence derived from an existing human gene. This approach has several benefits: As a modular design, different transgenes can be controlled by the same upstream regulatory module. Furthermore, TDP-REGv1 is based on a well-validated endogenous CE that is detected in patients in a disease-specific manner.

By contrast, TDP-REGv2 uses synthetic splicing sensors that are embedded within the transgene sequence itself. This addresses several limitations with TDP-REGv1 and previous approaches (*26*): It removes the need for a lengthy upstream regulatory region, thus aiding packaging of large transgenes into AAV vectors; it prevents an upstream, unwanted peptide being expressed and released into the cell; and it is highly tunable, meaning that splicing characteristics can be optimized for the specific transgene being delivered.

Splicing regulation is highly complex and difficult to predict with conventional algorithms. During the development of TDP-REGv2, we created an algorithm, SpliceNouveau, which leverages the power of deep-learning splicing prediction and rational design principles to optimize vectors, helping ensure that they are spliced in the desired manner. Although SpliceAI was not directly trained to predict cryptic splicing, we found that its prediction scores correlate with cryptic splice site usage. Compared to using high-throughput screens to find sequences with desired splicing characteristics (*27*), use of SpliceNouveau means that only a few sequences need to be experimentally validated. Using SpliceNouveau, we successfully generated cryptic cassette exons, and also cryptic alternative splice sites, by computationally designing competitor alternative splice sites that are used preferentially when TDP-43 is present. The flexibility and high success rate of SpliceNouveau-designed vectors therefore represents a major step forward in splicing-regulated vector technology (*26*, *27*).

Overall, this study demonstrates how cryptic splicing, ostensibly a driver of ALS and FTD progression, can be reverse-engineered and exploited to enable exceptionally targeted therapeutic protein expression. This approach, which is compatible with conventional AAVs that have previously been approved for gene therapy, can readily be adapted for different transgenes, including tuning of their maximal expression and sensitivities to TDP-43 depletion. This will minimize the risk of toxic side effects and thus improve the chance of obtaining measurable improvements during clinical trials, reducing danger for patients and increasing incentives for drug developers.

REFERENCES AND NOTES

1. E. L. Scotter, H.-J. Chen, C. E. Shaw, *Neurotherapeutics* **12**, 352–363 (2015).
2. K. E. McAleese *et al*., *Brain Pathol.* **27**, 472–479 (2017).
3. P. T. Nelson *et al*., *Brain* **142**, 1503–1527 (2019).
4. J. P. Ling, O. Pletnikova, J. C. Troncoso, P. C. Wong, *Science* **349**, 650–655 (2015).
5. J. R. Klim *et al*., *Nat. Neurosci.* **22**, 167–179 (2019).
6. Z. Melamed *et al*., *Nat. Neurosci.* **22**, 180–190 (2019).
7. A.-L. Brown *et al*., *Nature* **603**, 131–137 (2022).
8. X. R. Ma *et al*., *Nature* **603**, 124–130 (2022).
9. P. R. Mehta, A.-L. Brown, M. E. Ward, P. Fratta, *Mol. Neurodegener.* **18**, 16 (2023).
10. S. Lu *et al*., *Nat. Cell Biol.* **24**, 1378–1393 (2022).
11. M. W. Baughn *et al*., *Science* **379**, 1140–1149 (2023).
12. H. Riemenschneider *et al*., *Acta Neuropathol. Commun.* **11**, 112 (2023).
13. E. Y. Liu *et al*., *Cell Rep.* **27**, 1409–1421.e6 (2019).
14. M. Van Alstyne *et al*., *Nat. Neurosci.* **24**, 930–940 (2021).
15. Y. Qiu *et al*., *Science* **378**, 523–532 (2022).
16. S. Seddighi *et al*., *Sci. Transl. Med.* **16**, eadg7162 (2024).
17. K. Jaganathan *et al*., *Cell* **176**, 535–548.e24 (2019).
18. E. Buratti, F. E. Baralle, *J. Biol. Chem.* **276**, 36337–36343 (2001).
19. P. J. Lukavsky *et al*., *Nat. Struct. Mol. Biol.* **20**, 1443–1449 (2013).
20. M. Budini, V. Romano, Z. Quadri, E. Buratti, F. E. Baralle, *Hum. Mol. Genet.* **24**, 9–20 (2015).
21. A. V. Anzalone *et al*., *Nature* **576**, 149–157 (2019).
22. J. R. Davis *et al*., *Nat. Biotechnol.* **42**, 253–264 (2023).
23. J. Lee *et al*., *Nat. Commun.* **14**, 1786 (2023).
24. P. J. Chen *et al*., *Cell* **184**, 5635–5652.e29 (2021).
25. C. Yang *et al*., *PLOS ONE* **17**, e0255710 (2022).
26. J. P. Ling *et al*., *Nat. Commun.* **13**, 5773 (2022).
27. K. North *et al*., *Nat. Biotechnol.* **40**, 1103–1113 (2022).
28. O. Wilkins, Delayed-Gitification/TDP-REG-Paper: For_resubmission, Zenodo (2024); https://doi.org/10.5281/zenodo.11576269.

ACKNOWLEDGMENTS

We thank M. Strom and A. Cunha of the Crick Institute viral vector core for producing the AAVs; R. Faraway for useful discussions on prime editing; F. C. Y. Lee, C. Capitanchik, and W. C. Lee for feedback on the manuscript; S. Barmada for suggestions on validation methods; and K. Georgievskaya/SciDraw for a cartoon image of a neuron (10.5281/zenodo.4717443). **Funding:** This work was supported by a Wellcome Trust PhD Studentship (to O.G.W.); a UK Medical Research Council Senior Clinical Fellowship and Motor Neurone Disease Association Lady Edith Wolfson Fellowship (MR/S006508/1; to P.F.); this project is part of the UCL Neurogenetic Therapies Programme (to O.G.W. and P.F.), funded by The Sigrid Rausing Trust; Target ALS (P.F.); NIH U54NS123743 (P.F.); a Wellcome Trust Clinical Training Fellowship (102186/B/13/Z; to P.R.M.); the UK Medical Research Council (MC_EX_MR/N501931/1 to T.J.C. and E.M.C.F.); a Wellcome Trust Investigator Award

(215593/A/19/Z to J.U.); LifeArc (P2020-0008 to R.K.) and Great Ormond Street Hospital Children's Charity (V4720 to R.K.); the National Institutes of Health NICHD Intramural Research Program (ZIA-HD008966 to C.L.P.); the Francis Crick Institute, which receives its core funding from Cancer Research UK (CC0102), the UK Medical Research Council (CC0102), and the Wellcome Trust (CC0102); and by the UK Dementia Research Institute (award no. UK DRI-RE13553) through UK DRI Ltd, principally funded by the UK Medical Research Council. **Author contributions:** Conceptualization: P.F., O.G.W. Software: O.G.W. Formal analysis: O.G.W., M.Z.Y.J.C., J.J.W., A.-L.B., P.H., P.R.M. Resources: R.K., E.M.C.F., T.J.C., J.U., C.L.P., P.F. Supervision: O.G.W., R.K., E.M.C.F., T.J.C., J.U., C.L.P., P.F. Funding acquisition: E.M.C.F., R.K., E.F., T.J.C., J.U., C.L.P., P.F. Investigation: O.G.W., M.Z.Y.J.C., J.J.W., M.P., D.T., H.D., R.L.S., J.A.D., M.J.K., M.Z., A.-L.B., P.R.M., P.H., F.M., S.B., E.R., L.K., A.D. Visualization: O.G.W., J.J.W., A.-L.B. Methodology: O.G.W., J.J.W., P.H., C.L.P., P.F. Writing - original draft: O.G.W., P.F. Writing - review and editing: M.Z.Y.J.C., J.U., C.L.P., J.J.W., E.M.C.F., M.Z. **Competing interests:** P.F. and O.G.W. have filed a patent application, "WO2023198347A1: A construct, vector, and system and uses thereof," relating to TDP-REG technology. P.F. consults for, holds shares in, and is academic founder of Trace Neuroscience. P.F. and O.G.W. have acted as consultants for VectorY. **Data and materials availability:** RNA-Seq Data for i3Neurons, SH-SY5Y and SK-N-BE(2) are available through the European Nucleotide Archive (ENA) under accession PRJEB42763. Public data were obtained from Gene Expression Omnibus (GEO): iPSC MNs (*5*)-GSE121569, Appocher SK-N-BE(2)-GSE97262, and HeLa Ferguson-GSE136366. NYGC ALS Consortium RNA-seq: RNA-Seq data generated through the NYGC ALS Consortium in this study can be accessed via the NCBI's GEO database (GEO GSE137810, GSE124439, GSE116622, and GSE153960). All RNA-Seq data generated by the NYGC ALS Consortium are made immediately available to all members of the Consortium and with other consortia with whom we have a reciprocal sharing arrangement. To request immediate access to new and ongoing data generated by the NYGC ALS Consortium and for samples provided through the Target ALS Postmortem Core, complete a genetic data request form at ALSData@nygenome.org. Targeted sequencing datasets and accompanying Snakemake scripts for data processing are available at https://doi.org/10.5061/dryad.cjsxksnfr. A full R markdown containing code for all data visualization, including small amounts of downstream processing, is available at Zenodo (*28*). The full sequences of all vectors used in this study are available in the supplementary materials. A selection of TDP-REG vectors has been made available on Addgene.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk2539
Materials and Methods
Figs. S1 to S10
Tables S1 to S4
References (*29*–*47*)

Submitted 1 September 2023; accepted 15 August 2024
10.1126/science.adk2539