NEUROSCIENCE · COMPUTATIONAL BIOLOGY Draft manuscript — not peer reviewed

Deep-learning prediction of TDP-43 binding footprints nominates antisense-oligonucleotide targets in ALS cryptic-splicing genes

J. Milton

SIGNIFICANCE

When the protein TDP-43 is depleted from the cell nucleus—the hallmark of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia—it stops silencing hidden "cryptic" exons in neuronal genes such as STMN2 and UNC13A, corrupting those mRNAs and eliminating the proteins they encode. A promising therapy uses a steric-blocking antisense oligonucleotide (ASO): a short synthetic nucleic-acid strand that base-pairs over the cryptic splice site and physically prevents its use, restoring the normal mRNA. Designing such an ASO requires knowing where along a gene TDP-43 normally acts. We built a machine-learning tool that reads RNA sequence and predicts where TDP-43 binds. Without being told the answer, it places TDP-43's strongest predicted footprint in STMN2 intron 1—exactly the region defined by experiment—offering an inexpensive way to point ASO design at the right stretch of a very large gene.

TDP-43 (encoded by TARDBP) is an RNA-binding protein whose nuclear depletion is a defining feature of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Loss of TDP-43 repression exposes cryptic exons that truncate or destabilize target mRNAs; the STMN2 and UNC13A cryptic exons are validated pathological events and active antisense-oligonucleotide (ASO) targets. Because TDP-43 represses through defined UG-rich intronic elements, a map of where TDP-43 binds a given gene is a prerequisite for rational design of a steric-blocking ASO. Direct experimental binding maps (crosslinking/eCLIP) are sparse and come almost entirely from non-neuronal cell lines, motivating prediction from sequence. Here we retrained a convolutional neural network that predicts, from RNA sequence alone, where each of 170 RNA-binding proteins attaches, using the complete ENCODE crosslinking compendium (168 proteins; 6.9 million binding sites) augmented with POSTAR3. Evaluated on chromosomes withheld from training, the TDP-43 predictor ranks a true binding site above a non-site 96% of the time (AUROC = 0.96) and cleanly recovers TDP-43's known UG-rich target motif. Scanning the STMN2 and UNC13A pre-mRNAs, we find dense, high-confidence TDP-43 footprints. For STMN2, 71% of predicted sites fall within the experimentally defined intron-1 repressive element, a 6.8-fold enrichment over the gene-wide background—a quantitative, retrospective cross-validation against published crosslinking and functional data. For UNC13A, the model detects significant but weaker enrichment (2.9-fold) at the reported cryptic exon against a UG-rich gene-wide background, illustrating both the method's sensitivity and its resolution limit. Model-based saturation mutagenesis—scoring every single-base change—identifies

the individual UG positions that carry the predicted binding, and we export candidate ASO-target coordinates in browser-ready form. Across a panel of twelve further published TDP-43 cryptic-exon genes, focal TDP-43 enrichment is elevated relative to housekeeping controls (median 12× versus 4×) and the most enriched loci are all established targets, indicating that the approach generalizes as a within-gene localizer even though pervasive UG content limits whole-gene target classification. Rather than nominating a single oligonucleotide, the map is used to design a cost-effective screen: focusing the empirical ASO tiling walk on the predicted hot-spot shrinks a prohibitive ~5,500-member 20-mer library (whole gene) to ~40 (≈112-fold), turning an infeasible campaign into a routine one. The framework is thus a discovery-stage instrument for ASO screen design; its principal limitations—non-neuronal training data and region-level rather than single-nucleotide resolution, the latter well matched to its role of sizing a screen rather than picking a winner—are quantified and discussed.

INTRODUCTION

Cytoplasmic aggregation and nuclear clearance of TDP-43 is observed in the great majority of ALS cases and in roughly half of FTD cases,[1] placing the loss of TDP-43's normal nuclear function at the center of disease mechanism. A principal nuclear function of TDP-43 is the repression of nonconserved cryptic exons: TDP-43 binds long UG-rich intronic stretches and prevents the inclusion of intervening pseudo-exons that would otherwise introduce premature stop codons or polyadenylation sites.[2,3,4]

Two of these events have become both mechanistically pivotal and therapeutically tractable (Fig. 1). In STMN2, loss of TDP-43 activates a cryptic exon in the first intron that truncates the message and eliminates stathmin-2 protein, a factor motor neurons need to regrow their axons;[5,6] re-imposing correct STMN2 splicing rescues stathmin-2.[7] In UNC13A, a TDP-43-repressed cryptic exon is made worse by common intronic risk variants, linking a known ALS/FTD genetic risk factor directly to the loss-of-repression mechanism.[8,9]

The therapeutic logic is the same in both cases and is the reason a binding map matters. Antisense oligonucleotides (ASOs) are short, chemically stabilized single strands of nucleic acid that base-pair with a chosen pre-mRNA sequence. Two classes exist. A "gapmer" ASO recruits the enzyme RNase H to cut and destroy the RNA it binds—useful for lowering a toxic transcript, but the wrong tool here. The class relevant to cryptic splicing is the steric-blocking (splice-switching) ASO: it does not destroy the RNA but simply sits on the pre-mRNA and physically covers a splicing signal. Placed over the cryptic exon's splice site, a steric-blocking ASO stops the spliceosome from including that exon, so the normal full-length mRNA is produced—in effect substituting for the TDP-43 repression that has been lost (Fig. 1). This is the mechanism of nusinersen, an approved ASO for spinal muscular atrophy, and of the STMN2-directed ASOs now in development.[7] The practical difficulty is that such an ASO must be aimed at the correct ~100–200 nucleotides within a gene that can span hundreds of kilobases.

Choosing that window benefits from knowing precisely where TDP-43 engages the pre-mRNA. Direct measurement by crosslinking (CLIP/eCLIP) is definitive but sparse, and the largest standardized dataset (ENCODE) comes from non-neuronal cells.[10] Because TDP-43 recognition is dominated by a short, sequence-intrinsic UG-rich element,[2] a model that learns this preference can in principle predict binding on any transcript. We therefore asked whether a sequence-based binding

predictor could map TDP-43 footprints across ALS cryptic-splicing genes accurately enough to point ASO design at the right region—and we report both where it succeeds and where it does not.

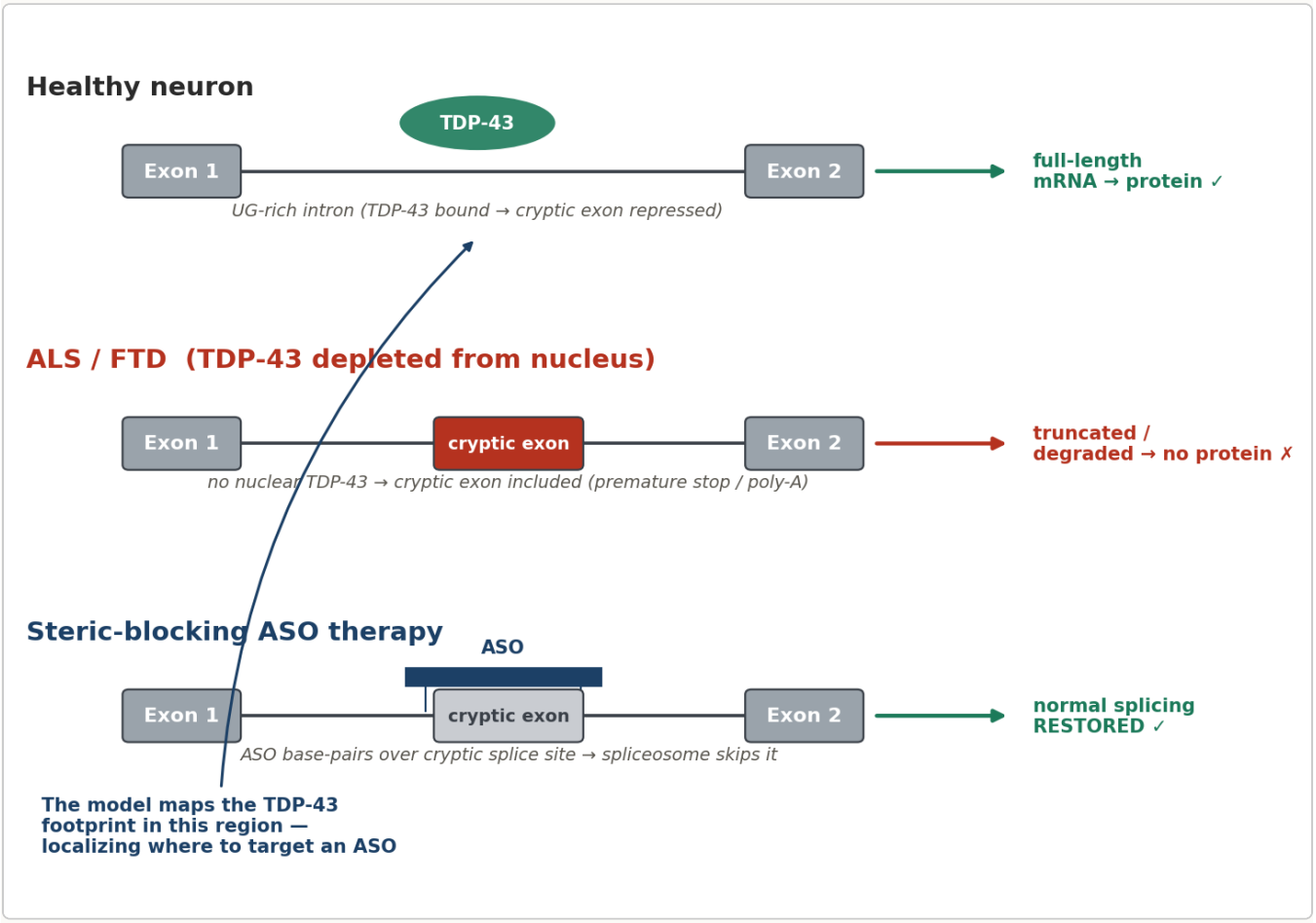

Fig. 1. Why and where a steric-blocking ASO is used. Top (healthy): nuclear TDP-43 binds the UG-rich intron and represses a cryptic exon, so exon 1 splices directly to exon 2 and full-length protein is made. **Middle (ALS/FTD):** when TDP-43 is depleted from the nucleus, the cryptic exon is included; it carries a premature stop codon or polyadenylation signal, truncating or degrading the mRNA so no functional protein is produced. **Bottom (therapy):** a steric-blocking (splice-switching) ASO base-pairs directly over the cryptic exon's splice site, physically preventing the spliceosome from using it; the cryptic exon is skipped and normal splicing is restored. The model described here maps the TDP-43 footprint (the region in blue) to localize which stretch of a large gene an ASO should target.

RESULTS

A model that predicts TDP-43 binding from sequence.

The tool is a convolutional neural network—an architecture from image recognition, here reading RNA sequence instead of pixels—that takes a 64-nucleotide stretch of RNA and returns, for each of 170 RNA-binding proteins, the probability that the protein binds there. Throughout this paper we use one of those 170 outputs, the TDP-43 prediction. We retrained the network on the complete ENCODE crosslinking compendium (168 proteins; 6.9 million binding sites), adding two axonal proteins (STAU1, HNRNPA2B1) from POSTAR3, and balanced the training so that predicted scores stay interpretable across proteins (Materials and Methods).

Evaluated on entire chromosomes withheld from training, the TDP-43 prediction ranks a genuine binding site above a random non-site 96% of the time (AUROC = 0.96, where 1.0 is perfect and 0.5 is chance). It is also biochemically faithful: tested on defined sequences, it responds strongly to UG-dinucleotide repeats—TDP-43's known binding element—(score 0.97) and only weakly to other

sequence types (mean 0.33; Fig. 2). Of all 170 proteins, TDP-43 combined the best ranking accuracy with the cleanest separation between binding and background, making it the most trustworthy prediction on which to base a footprint map.

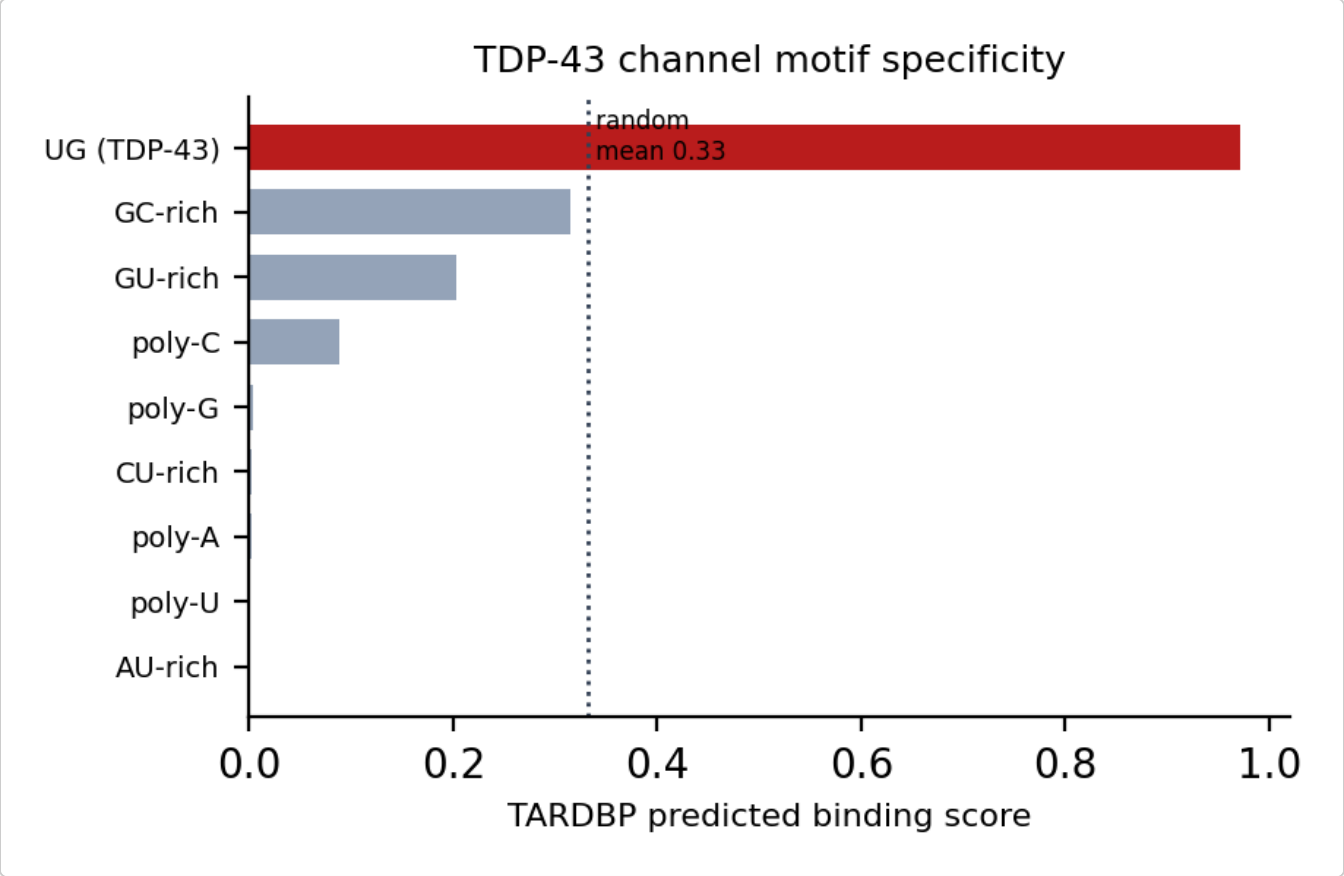

Fig. 2. The model reproduces TDP-43's binding preference. Predicted TDP-43 binding for defined sequence types. The UG-dinucleotide repeat—TDP-43's canonical target—scores far above other sequence classes and above the random-sequence average (dotted line), confirming the model has learned TDP-43's real recognition preference rather than a generic signal.

TDP-43 footprints concentrate in STMN2 intron 1.

We scanned the entire STMN2 gene (chr8:79,611,117–79,666,158; GRCh38) in transcriptional orientation. The TDP-43 predictor flagged 56 high-confidence binding windows (34 above a score of 0.99), and these are strikingly non-uniform. We compared them directly against the experimentally defined TDP-43 repressive element—the GU-rich intron-1 region chr8:79,611,214–79,616,822 mapped by crosslinking and functional dissection in Baughn et al.[7] Of the 56 predicted sites, 40 (71%) fall inside this element, a 6.8-fold higher density of predicted binding than the gene-wide average (7.1 versus 1.04 sites per kilobase), with the nearest high-confidence site just 582 bp from the element's center (Table 1; Fig. 3). Because these coordinates were never given to the model, the agreement is a genuine, quantitative cross-validation against independent experimental data—and it identifies intron 1 as the region an STMN2 ASO should target.

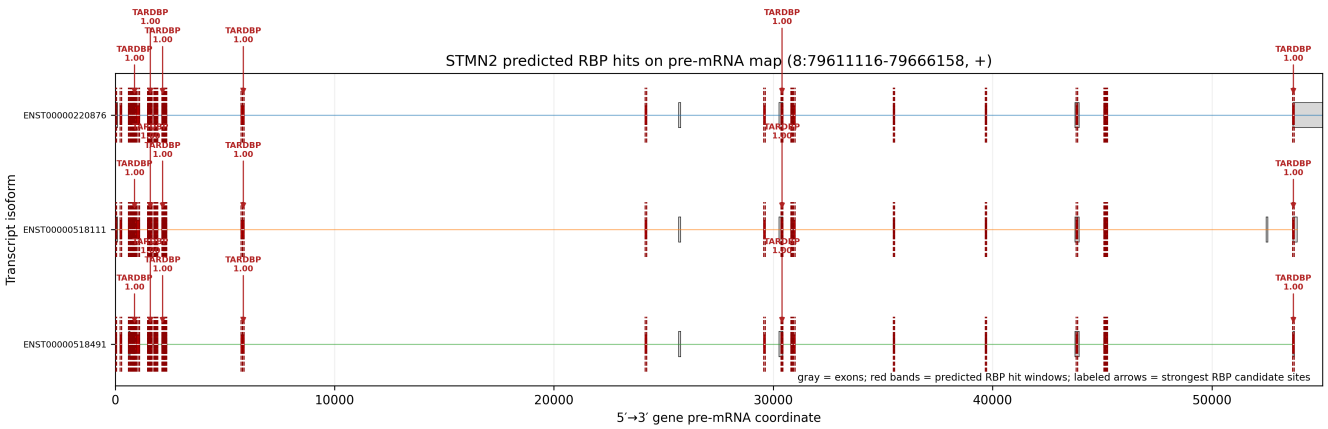

Fig. 3. Predicted TDP-43 footprint on the STMN2 gene. Red bands mark high-confidence predicted TDP-43 binding windows across three STMN2 isoforms; gray boxes are exons; the horizontal axis is the 5′→3′ position along the gene. Predicted binding is heavily concentrated at the 5′ end (position 0–2.5 kb), corresponding to intron 1 immediately after exon 1—the location of the TDP-43-repressed cryptic exon and the region an ASO would target. Forty of 56 predicted sites (71%; 6.8-fold enrichment) fall within the experimentally defined intron-1 element (chr8:79,611,214–79,616,822; ref. 7).

UNC13A carries multiple intronic TDP-43 clusters.

Applying the same procedure to UNC13A (chr19:17,601,336–17,688,365; − strand; 87 kb) yielded a far larger footprint—454 windows, 259 above 0.99—consistent with the locus's length and pervasive UG-richness. At the experimentally mapped cryptic exon (chr19:17,641,557–17,642,844; refs 8,9) the model does detect TDP-43 occupancy: 12 high-confidence sites, the nearest 165 bp from the exon center, representing a 2.9-fold enrichment over the gene-wide average (Table 1). Crucially, however —and unlike STMN2—this region is not the model's densest cluster. Because UNC13A is UG-rich throughout, the strongest predicted clusters (chr19:17,655,973 and 17,605,845) lie 14–36 kb from the cryptic exon (Fig. 4). This STMN2/UNC13A contrast is itself informative: the method pinpoints a discrete repressive element with high specificity where one exists (STMN2) but, against a diffusely UG-rich background (UNC13A), reports enrichment rather than a unique target—consistent with the broader, hnRNP-co-regulated and risk-variant-potentiated character of UNC13A cryptic splicing.[8,9]

Table 1. Cross-validation of predicted TDP-43 footprints against experimentally defined coordinates (GRCh38). Enrichment is footprint-density inside the published region relative to the gene-wide average; the model was blind to these coordinates.

Gene	Published TDP-43 / cryptic region	Ref.	Predicted sites in region	Enrichment	Nearest site	Concordance
STMN2	chr8:79,611,214–79,616,822	7	40 / 56 (71%)	6.8×	582 bp	strong (discrete element)
UNC13A	chr19:17,641,557–17,642,844	8, 9	12 sites	2.9×	165 bp	partial (enriched, not top cluster)

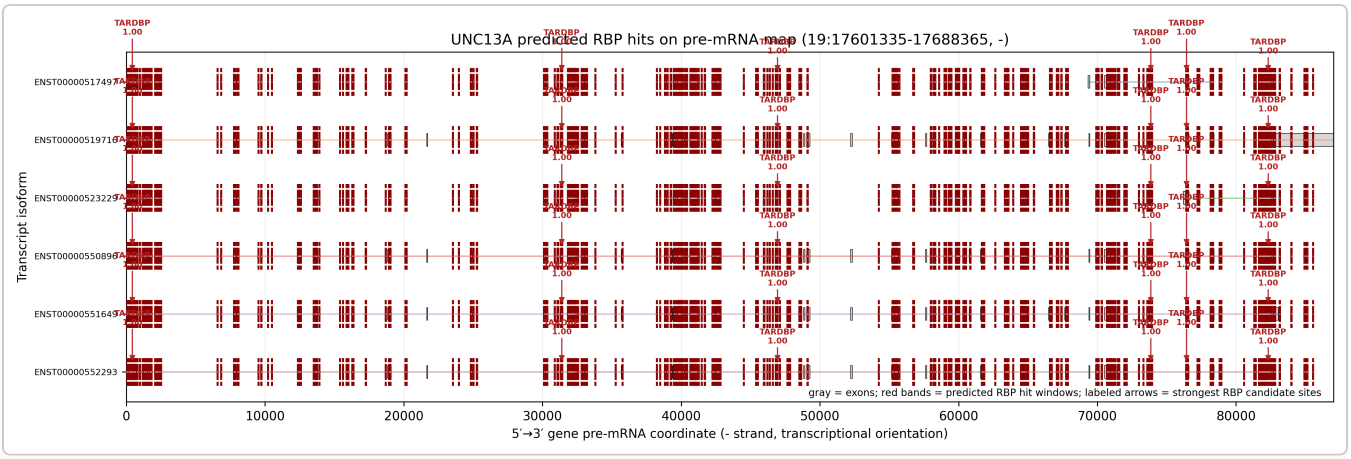

Fig. 4. Predicted TDP-43 footprint on the UNC13A gene. As in Fig. 3, for the seven UNC13A isoforms. Predicted TDP-43 binding is dense and spread across the introns. The experimentally mapped cryptic exon (chr19:17,641,557–17,642,844; refs 8,9) is detectably enriched (2.9-fold) but is not the strongest predicted cluster; the densest clusters lie 14–36 kb away, reflecting the gene's pervasive UG-richness and illustrating the method's resolution limit against a high sequence background.

Model-based saturation mutagenesis resolves the critical UG positions.

To move from the region to individual bases, we computationally tested every possible single-nucleotide change across STMN2 intron 1 (1,888 positions × 3 substitutions), recording how each change altered the predicted TDP-43 binding (Fig. 5). Effects are sharply position-dependent: 232 positions (12.3%) have at least one substitution that strongly changes predicted TDP-43 binding ($|\Delta| > 0.5$), tracing the specific UG bases that carry the footprint. Such maps indicate which nucleotides a steric-blocking ASO would most need to cover to displace TDP-43, and—read in reverse—which naturally occurring sequence variants might weaken repression.

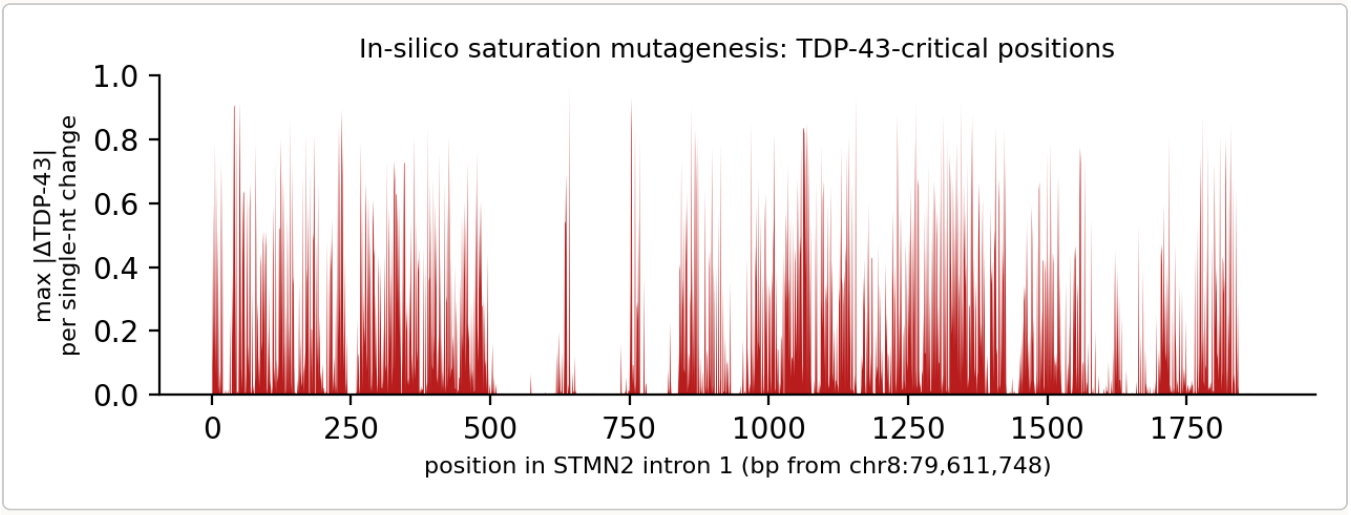

Fig. 5. TDP-43-critical bases in STMN2 intron 1. For each position, the largest change in predicted TDP-43 binding over the three possible single-base substitutions is plotted along intron 1. Peaks mark UG bases the model deems essential to TDP-43 binding; flat regions are positions where the base identity does not matter. The clustered high-sensitivity bases coincide with the dense footprint of Fig. 3.

Generalization across published TDP-43 targets.

To ask whether the STMN2 concordance generalizes, we scanned twelve additional genes reported in the literature to harbor TDP-43-regulated cryptic exons (including UNC13B, HDGFL2, KALRN, SETX, AGRN, ATG4B, ELAVL3, POLDIP3 and GPSM2) alongside four housekeeping genes as negative controls. For each we computed a focal-enrichment score—the density of the single densest

TDP-43 cluster relative to the gene-wide average—capturing the "sharp repressive zone" that characterized STMN2. Known TDP-43 targets carried substantially more focal TDP-43 clustering than controls (median 12.0× versus 4.1×), and the five most focally enriched genes in the panel (SETX, UNC13B, GPSM2, KALRN, STMN2) are all bona fide TDP-43 targets (Fig. 6). The separation is real but imperfect: several genuine targets with short, uniformly UG-rich loci (e.g. AGRN, HDGFL2) are not distinguished from controls, because widespread UG content limits whole-gene discrimination. The framework is therefore best understood as a within-gene localizer—it pinpoints where TDP-43 acts in a gene of interest, as validated quantitatively for STMN2—rather than a genome-wide classifier of which genes are TDP-43 targets.

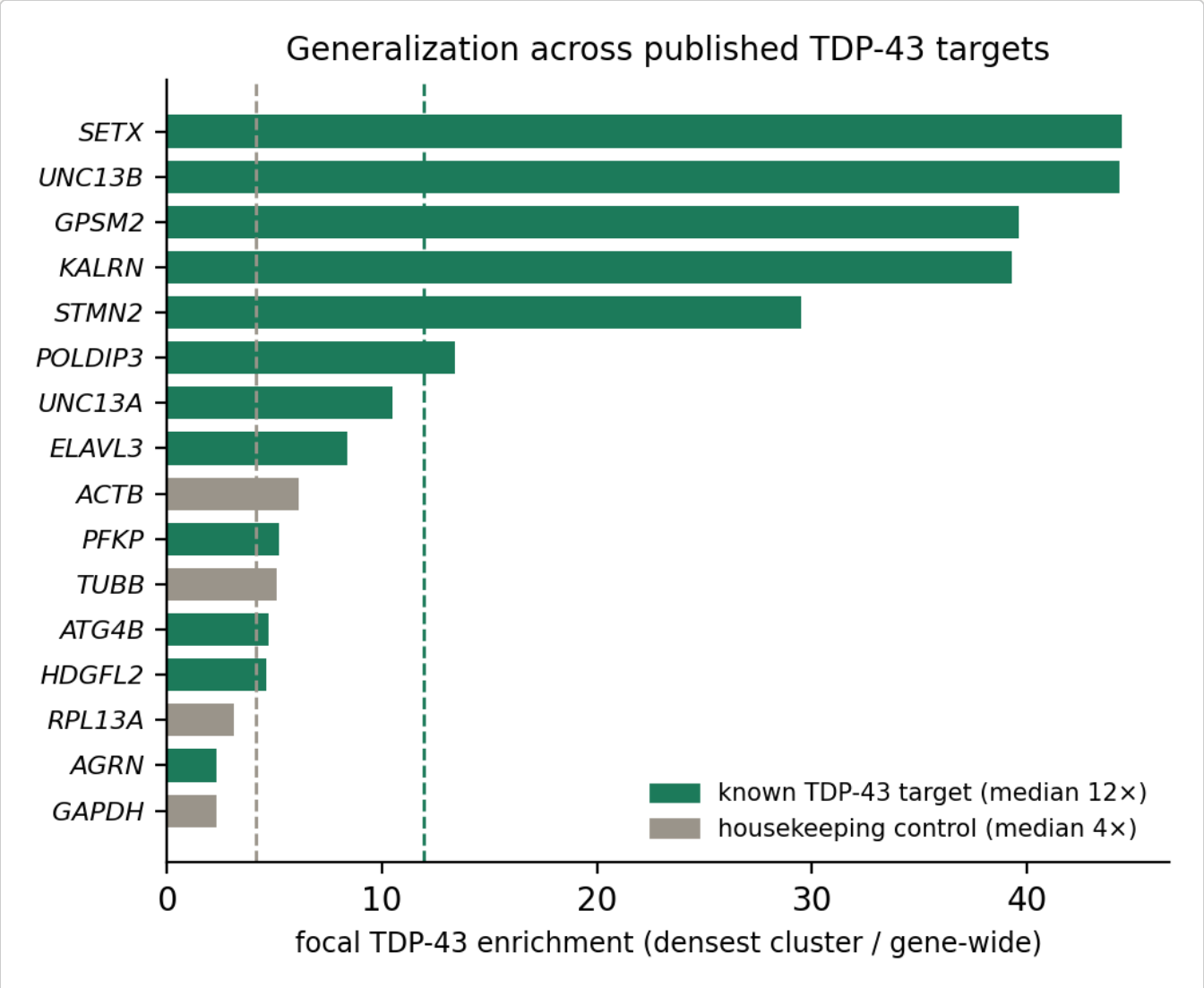

Fig. 6. Focal TDP-43 enrichment across published targets and controls. For each gene, the density of the densest 400-bp TDP-43 cluster is expressed relative to the gene-wide average. Known TDP-43-regulated genes (green) show higher focal enrichment than housekeeping controls (gray); dashed lines mark the group medians (12.0× vs 4.1×). The most focally enriched genes are all established TDP-43 targets, but overlap in the mid-range reflects the pervasiveness of UG-rich sequence and bounds the method to within-gene localization rather than genome-wide target classification.

From footprint map to a cost-effective ASO screen.

Effective splice-switching ASOs are found empirically, by an "ASO walk": overlapping oligonucleotides tiled across a target region are synthesized and tested, and the active ones emerge from the assay rather than from prediction. The cost of such a screen scales directly with the length

of the region tiled, and a steric-blocking ASO is short—typically a 20-mer—so exhaustively walking a whole gene is rarely practical. Tiling all 55 kb of STMN2 with 20-mers at a 10-nt step would require roughly 5,500 distinct oligonucleotides (Fig. 7A), each of which must be synthesized in stabilized chemistry and assayed for splice correction; even restricting the walk to the 5.6-kb experimentally defined element still leaves ~560.

This is where a sequence-based footprint map earns its keep. By concentrating the predicted TDP-43 signal into a ~500-bp hot-spot within intron 1, the model shrinks the region that must be tiled by roughly two orders of magnitude—to ~40–50 overlapping 20-mers (Fig. 7B), a library that fits on a single plate and is routine to synthesize and screen. The reduction from ~5,500 to ~40 ASOs (≈112-fold) is the difference between an infeasible campaign and a tractable one. Importantly, the model is not asked to nominate the single winning oligonucleotide—a task its single-nucleotide resolution does not support—but to define the smallest region within which the winner is very likely to lie, so that an unbiased tiling screen can find it at minimum cost. The method's region-level accuracy is precisely matched to this region-level design decision: the deliverable is an optimally sized screen, not a predicted hit.

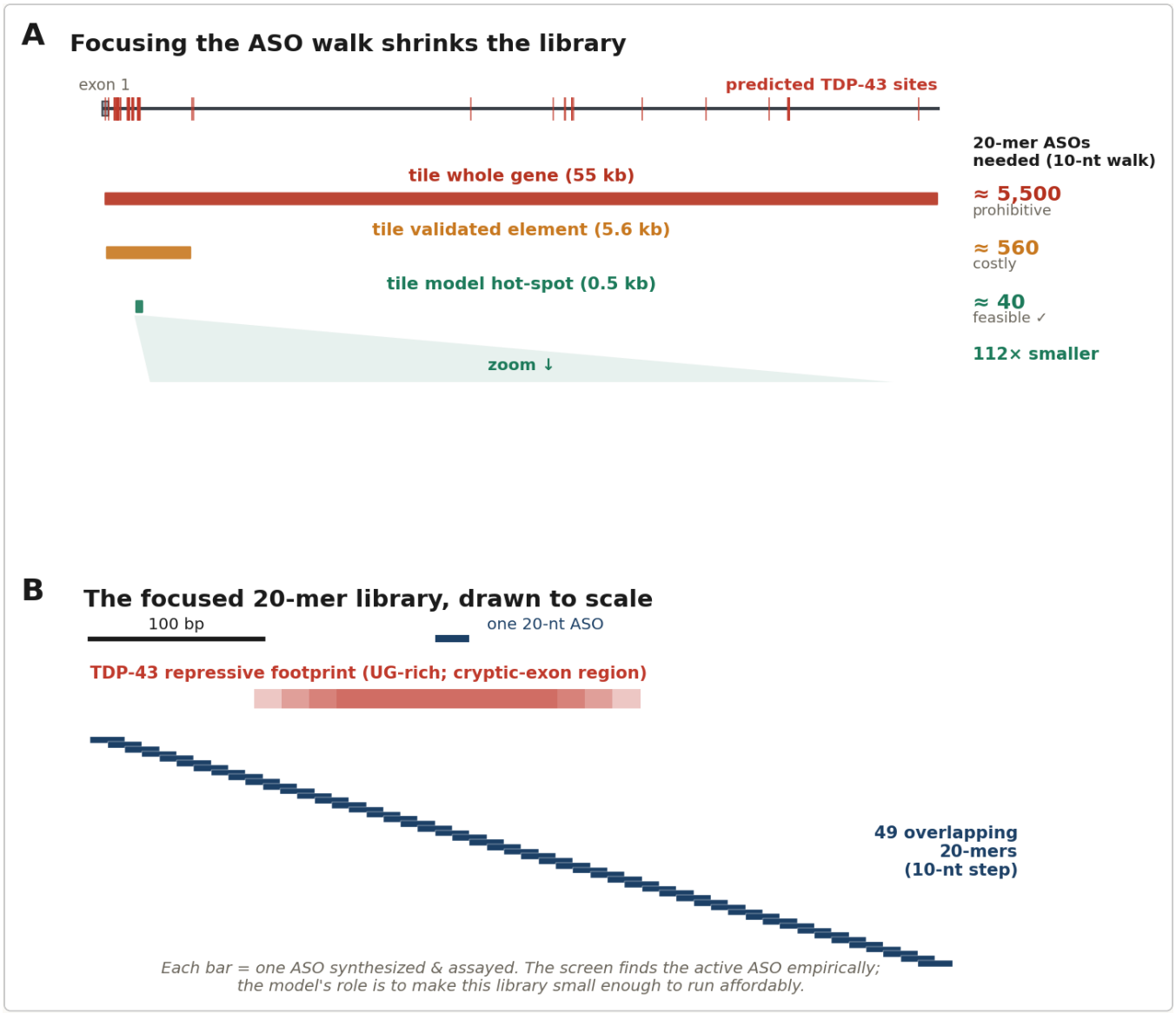

Fig. 7. Using the footprint map to design a cost-effective ASO screen. (A) A steric-blocking ASO is a ~20-mer; finding an active one requires tiling overlapping 20-mers across a target and testing them. Tiling the whole 55-kb STMN2 gene (10-nt walk) needs ~5,500 ASOs (prohibitive); the 5.6-kb validated element ~560 (costly); the model's ~500-bp predicted hot-spot ~40 (feasible)—a ≈112-fold smaller library. (B) The focused library drawn to scale: ~49 overlapping 20-mer steric ASOs (10-nt step) tiling the predicted TDP-43 footprint (pink) within the hot-spot. Each bar is one ASO to be synthesized and assayed; the screen still identifies the winner empirically, but the model has made the library small enough to run affordably.

DISCUSSION

The central result is a quantitative concordance with independent experimental data: a model trained only on non-neuronal crosslinking data, given only genomic sequence, concentrates 71% of its predicted STMN2 TDP-43 footprint—a 6.8-fold enrichment—within the intron-1 element that Baughn et al. defined biochemically and targeted therapeutically.[7] Because these coordinates were withheld from the model, the agreement validates that the TDP-43 predictor has learned the sequence determinants of cryptic-exon repression rather than a generic signal. The UNC13A result is more measured and, we think, instructive: TDP-43 occupancy at the mapped cryptic exon is real and enriched (2.9-fold) but not uniquely strong, because the locus is UG-rich throughout. The method thus excels at localizing a discrete repressive element against a low background (STMN2) and degrades gracefully—reporting enrichment rather than a false-precise single target—where binding

is distributed (UNC13A). Saturation mutagenesis further resolves the individual residues that carry the predicted signal, the coordinates an ASO campaign must engage.

Several limitations bound the claims. First, the ENCODE training data derive from K562 and HepG2 cells; the model therefore predicts sequence-intrinsic binding potential rather than neuron-specific occupancy, and every nomination requires confirmation in a neuronal context (e.g., iPSC-derived neurons) and by orthogonal splicing assays (minigene, targeted RNA-seq). Second, the framework is reliable at the level of regions, not single variants: we found that exhaustive substitution flags a large fraction of positions, so the saturation maps are best read as relative sensitivity landscapes, not as calibrated variant-effect predictions; dedicated splicing predictors remain the appropriate tool for classifying individual variants. Third, the concordance we report is with a single, exceptionally well-characterized locus; prospective experimental testing of the UNC13A nominations, and of genome-wide predictions, is the necessary next step. Finally, the FUS prediction, though newly usable after retraining, marks broad binding along nascent transcripts and is less informative for pinpointing discrete elements.

Within these bounds, the approach fills a practical gap in how ASO campaigns are run. Because active oligonucleotides are found by empirical tiling rather than predicted, the rate-limiting cost is the size of the region that must be walked; an accurate, transferable footprint map lets that walk be focused, turning a prohibitive whole-gene screen into a plate-sized one (Fig. 7). The model's contribution is to size and place the screen, not to pick the winning ASO—an allocation of labor that matches its region-level accuracy to a region-level decision. We caution, however, that whole-gene metrics do not cleanly separate TDP-43 targets from non-targets (Fig. 6): because UG-rich sequence is common, the framework is reliable for localizing TDP-43 within a gene already suspected to be a target, and only exploratory as a genome-wide discovery filter. Coupling the sequence prediction with orthogonal evidence—splice-junction counts from TDP-43-depleted neurons, or exon–intron-boundary constraints—would be required to raise genome-wide nomination to the standard set by the within-gene localization validated here for STMN2.

A final biological point deserves emphasis. Although the therapeutic case for STMN2 rests on axon regeneration, the wider set of TDP-43 cryptic-splicing targets is functionally diverse and reaches well beyond axon biology. Within the panel examined here alone, ATG4B acts in autophagy, PFKP in glycolytic metabolism, GPSM2 in mitotic-spindle orientation and cell division, SETX (senataxin) in R-loop resolution and genome maintenance, and HDGFL2 is a chromatin-associated HDGF-family protein. Loss of their correct splicing is therefore likely to perturb general cellular growth- and maintenance programs—autophagy, metabolism, proliferation and genome integrity—rather than axonal outgrowth alone, implying that TDP-43 depletion may injure neurons through several parallel routes. It also broadens the translational scope: because nuclear TDP-43 depletion underlies non-neuronal TDP-43 proteinopathies such as inclusion-body myositis, and the common aging pathology LATE (limbic-predominant age-related TDP-43 encephalopathy), a sequence-based footprint map could in principle guide splice-correcting ASO design in cell types and diseases well beyond motor-neuron axonopathy.

MATERIALS AND METHODS

Model architecture. The predictor is a convolutional neural network ("sphere-CNN") mapping a 64-nucleotide window to a binding score for each of 170 proteins. Bases are embedded (8-dim) and passed through three 1-D convolutional blocks (kernels 9/15/21; 128/256/256 channels; batch-normalization, ReLU, max-pooling) followed by adaptive max-pooling; a parallel "sphere" branch supplies four hand-engineered composition features (motif-hit density, length-weighted density,

longest motif, GC content) computed against a fixed 4,421-motif dictionary. Convolutional and sphere features are concatenated and passed to a two-layer classifier producing 170 sigmoid outputs.

Training data. IDR-thresholded eCLIP peaks for all 168 ENCODE RBP targets (GRCh38; 6.9×10^6 peaks) were downloaded programmatically; STAU1 and HNRNPA2B1 binding sites were obtained from POSTAR3. For each protein up to 5,000 peaks were retained (ranked by signal), and a 64-nt window centered on each peak summit provided positive examples; 4.2×10^5 random genomic windows provided calibrated negatives. The network was trained with masked binary cross-entropy (each binding site supervises only its own protein's output; background windows push all outputs toward zero) using the Adam optimizer and per-protein positive weighting capped to preserve score calibration.

Evaluation. Chromosomes 1, 8, 15 and 22 were withheld entirely from training; all reported AUROC values are computed on these held-out chromosomes to exclude peak-adjacency leakage. Motif specificity was assessed on synthetic homopolymer/repeat panels against a random-sequence background.

Locus scanning. For a gene, the full genomic span was extracted in transcriptional orientation and tiled with 64-nt windows; each window was scored and windows exceeding a high-confidence threshold (0.95; 0.99 for cluster calling) were retained. Footprint clusters were identified as 400-bp intervals of maximal high-confidence-window density; hits were mapped back to genomic coordinates and exported in BED format. Model-based saturation mutagenesis scored all three substitutions at every position of a target interval and recorded the largest change in the TDP-43 output.

Cross-validation against experimental data. Predicted TDP-43 footprints were compared to experimentally defined coordinates drawn from the primary literature—the STMN2 intron-1 repressive element (ref. 7) and the UNC13A cryptic exon (refs 8, 9)—by counting predicted high-confidence sites within each published interval and computing their density relative to the gene-wide average (Table 1). These coordinates were not provided to the model at any stage of training or scanning; the comparison is therefore an out-of-sample validation.

Generalization panel. Twelve genes with literature-reported TDP-43-regulated cryptic exons and four housekeeping genes were scanned as above. For each, a focal-enrichment score was defined as the TDP-43 site density within the densest 400-bp window divided by the gene-wide site density; group medians (known targets versus controls) were compared (Fig. 6). Because this metric rewards spatially concentrated binding, it captures the discrete-repressive-zone architecture of STMN2 while remaining agnostic to overall gene length.

REFERENCES

1. Neumann M, et al. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314:130–133.
2. Polymenidou M, et al. (2011) Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat Neurosci 14:459–468.
3. Tollervey JR, et al. (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci 14:452–458.
4. Ling JP, Pletnikova O, Troncoso JC, Wong PC (2015) TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. Science 349:650–655.
5. Klim JR, et al. (2019) ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. Nat Neurosci 22:167–179.
6. Melamed Z, et al. (2019) Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. Nat Neurosci 22:180–190.

`STMN2_TARDBP_ASO_targets.bed`, `UNC13A_TARDBP_ASO_targets.bed`), and the saturation-mutagenesis tables are provided with the analysis pipeline. **Competing interests.** None declared. **Note.** This is an internally generated draft manuscript formatted in the PNAS style; it has not been submitted to, reviewed by, or published in PNAS or any journal.