



Published in final edited form as:

Nature. 2021 August ; 596(7871): 291–295. doi:10.1038/s41586-021-03770-2.

## Regulated control of gene therapies by drug-induced splicing

Alex Mas Monteys<sup>1,2,✉</sup>, Amiel A. Hundley<sup>1</sup>, Paul T. Ranum<sup>1</sup>, Luis Tecedor<sup>1</sup>, Amy Muehlmann<sup>1</sup>, Euyun Lim<sup>1</sup>, Dmitriy Lukashov<sup>3</sup>, Rajeev Sivasankaran<sup>3</sup>, Beverly L. Davidson<sup>1,2,✉</sup>

<sup>1</sup>Raymond G. Perelman Center for Cellular and Molecular Therapeutics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA.

<sup>2</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA.

<sup>3</sup>Neuroscience Disease Area, Novartis Institutes for BioMedical Research (NIBR), Cambridge, MA, USA.

### Abstract

So far, gene therapies have relied on complex constructs that cannot be finely controlled<sup>1,2</sup>. Here we report a universal switch element that enables precise control of gene replacement or gene editing after exposure to a small molecule. The small-molecule inducers are currently in human use, are orally bioavailable when given to animals or humans and can reach both peripheral tissues and the brain. Moreover, the switch system, which we denote  $X^{on}$ , does not require the co-expression of any regulatory proteins. Using  $X^{on}$ , the translation of the desired elements for controlled gene replacement or gene editing machinery occurs after a single oral dose of the inducer, and the robustness of expression can be controlled by the drug dose, protein stability and redosing. The ability of  $X^{on}$  to provide temporal control of protein expression can be adapted for cell-biology applications and animal studies. Additionally, owing to the oral bioavailability and safety of the drugs used, the  $X^{on}$  switch system provides an unprecedented opportunity to refine and tailor the application of gene therapies in humans.

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

✉ Correspondence and requests for materials should be addressed to A.M.M. or B.L.D. [monteysam@chop.edu](mailto:monteysam@chop.edu); [davidsonbl@chop.edu](mailto:davidsonbl@chop.edu).

**Author contributions** A.M.M. designed the research, performed experiments, analysed data and wrote the manuscript. A.A.H., E.L. and D.L. performed experiments. P.T.R. performed all bioinformatic analyses and wrote the manuscript. L.T. performed all histology analysis and contributed to statistical analyses. A.M. contributed to all in vivo studies. R.S. and D.L. assisted with progranulin studies. B.L.D. designed and supervised the research, analysed data and wrote the manuscript.

**Competing interests** B.L.D. serves on the advisory boards and/or has sponsored research from Homology Medicines, Intellia Therapeutics, Spirovant Sciences, Resilience Bio, Patch Bio, Saliogen, Moment Bio, Triplet Biosciences, Panorama Medicines, Roche and NBIR. A.M.M. has sponsored research from NBIR.

Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03770-2>.

**Peer review information** Nature thanks Alessandra Biffi, Mikhail Shapiro and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Although viral and non-viral approaches for gene therapies have undergone substantial advancement over the past twenty years, the major focus has been on the cargo-delivery system. For example, viral capsid evolution and engineering has improved the cell and tissue targeting of adeno-associated viruses (AAVs), and the landscape of cell targeting for lentiviruses has been expanded by using different envelopes in viral production. Additionally, lipid nanoparticles have been refined for improved uptake. However, the cargo itself—and more importantly the elements that control the expression from that cargo—have not received the same attention, although engineered promoters, riboswitches or other 3′ regulatory elements that restrict expression to certain cell types have advanced<sup>1–4</sup>.

To address this gap, here we develop a method to finely control protein translation via a drug-inducible switch. The X<sup>on</sup> system takes advantage of alternative RNA splicing<sup>5</sup>, a mechanism that provides for RNA and protein diversity through the inclusion or exclusion of different protein-coding exons, parts of exons, and different 5′ and 3′ noncoding exons. By controlling alternative splicing, we can regulate which exons are spliced in (or out). Notably, the X<sup>on</sup> switch does not require any foreign elements for regulation, but rather takes advantage of drugs that are orally bioavailable and in human use to induce the rapid inclusion of an exon containing a start codon. As such, the X<sup>on</sup> system can be applied to any genetic element of interest in cells or in animals.

Initially, alternative splicing for the regulated control of protein translation was achieved using drugs developed for the treatment of spinal muscular atrophy, a disorder that is caused by reduced levels of SMN1. The rationale is as follows: *SMN2*, a pseudogene of *SMN1*, is correctly spliced only around 10% of the time, resulting in an SMN2 protein that is functionally equivalent to SMN1<sup>6</sup>. Two drugs, LMI070 and RG7800/RG7619, can improve the proportion of correct *SMN2* splicing, and both show efficacy in animal models of spinal muscular atrophy<sup>7,8</sup>. Notably, these drugs are in later-stage clinical testing (LMI070)<sup>9</sup> or have been approved for use in humans (RG7800/RG7619)<sup>10</sup>.

Our first derivation of a splicing switch for controlled translation used a modified *SMN2* minigene, such that exon 7 (e7) exclusion and premature termination would occur in the absence of drug, but e7 inclusion and in-frame gene expression would occur in the presence of splice modifiers (Fig. 1a). For this, the original minigene was modified at the e7 acceptor splice site to reduce the approximately 10% background splicing in the native condition (indSMN2; Extended Data Fig. 1a, b). IndSMN2 was further evaluated in response to various doses of LMI070 or RG7800. Both drugs induced e7 splicing, with a complete splicing switch evident at concentrations greater than 1 μM (Extended Data Fig. 1c). For indSMN2 plus LMI070, there was an approximately 20-fold induction (Fig. 1b, Extended Data Fig. 1d). LMI070 showed greater induction activity than did RG7800 (Extended Data Fig. 1c), possibly reflecting their different mechanisms. RG7800/RG7619 binds to the SMN2 pre-mRNA sequence and promotes a conformational change that increases e7 splicing activator affinity, whereas LMI070 stabilizes U1 small nuclear RNA interactions at the e7 5′ splice site<sup>8,11</sup>.

Because the fully spliced-in switch required drug doses that are unlikely to translate to the clinic, we searched for exons for which splicing was responsive to therapeutically relevant

conditions. HEK293 cells were treated with 25 nM LMI070 for 12 h, and RNA sequencing (RNA-seq) was performed to determine alternatively spliced exons that were present only under drug-treated conditions. Reads were aligned to the genome and splicing events that were exclusive to the LMI070-treated samples were identified. After LMI070 treatment, 45 drug-induced splicing events were above our threshold (an average of more than five novel intron-splicing events in LMI070-treated samples; Extended Data Tables 1, 2). Among them, 23 were exclusive to LMI070-treated samples and 22 were identified in all treated samples and were below our threshold criteria in control samples. To assess exclusivity, we queried the 45 events in a resource containing all exon–exon junctions from 21,504 human RNA-seq datasets (Methods). Using *SF3B3* as an example, the canonical exon–exon junction in *SF3B3* was observed in 12,872 datasets at an average frequency of 64 counts, whereas the LMI070-induced exon was observed in 10 and 1 dataset(s), respectively, for the 5′ and 3′ exon–exon junctions (Fig. 1c, Extended Data Tables 1, 2). The average counts per dataset for the 5′ exon–exon junction was 1.3, whereas the 3′ exon junction was observed only once. Thus, the LMI070-induced *SF3B3* exon almost never occurs naturally.

The candidate exons from LMI070-treated samples share a strong 3′ AGAGUA motif that is consistent with the identified LMI070-targeted U1 RNA binding site<sup>8</sup> (Extended Data Fig. 2a). The top five novel LMI070-induced splicing events were validated by PCR, which amplified the novel exons exclusive to cDNA samples from LMI070-treated cells (Fig. 1d). The global effect of LMI070 treatment on gene expression was also evaluated by differential expression analysis. There were 6 upregulated and 24 downregulated genes that passed the threshold for significance ( $P < 0.05$ ; Benjamini–Hochberg multiple testing correction). After filtering out genes with low fold-change values (less than 0.1-fold) five upregulated and nine downregulated genes were identified (Fig. 1e).

We next used our custom pipeline to re-analyse a previous RNA-seq dataset<sup>8</sup> obtained from human cells treated with 100 nM LMI070. We observed strong concordance with our results, detecting 70% of our candidate genes in this previous dataset. The *SF3B3* and *BECN1* exons were the most significantly alternatively spliced in both datasets (Extended Data Fig. 2b, c). In addition, only 7 of the 42 reference splice events<sup>8</sup> induced by LMI070 were significantly differentially expressed in our study (Benjamini–Hochberg adjusted  $P$  value ( $P_{\text{adj}} < 0.05$ ; Extended Data Table 3), probably because of the low LMI070 concentration used here.

We developed switch-on luciferase cassettes using the minimal intronic intervening sequences to recapitulate splicing of the drug-responsive pseudo exons in *SF3B3*, *BECN1*, *C12orf4* and *PDXDC2*. To limit translation such that it would occur only in response to the drug, the novel exon was engineered to contain the Kozak sequence and AUG start codon (Fig. 1f). All possible downstream start codons were also removed to ensure that an AUG start would be included in response to LMI070 binding only. Increased luciferase expression was observed for each candidate cassette transfected into HEK293 cells in response to LMI070, with the *SF3B3*-on switch exhibiting a greater than 100-fold induction (Extended Data Fig. 3a, b), five times that of the 20-fold induction afforded by the indSMN2 minigene (Fig. 1b).

To assess background splicing from the engineered switch cassette, we performed two PCR assays. The LMI070-spliced-in exon was not detected when using primer pairs that bound the flanking exons, whereas a faint signal was observed when using primers binding within the novel exon sequence (Extended Data Fig. 3c). Thus, in the absence of LMI070, the alternative exon may be included in a small fraction of the transcripts, mirroring what was found in the Intropolis dataset. Overall, these assays showed marginal baseline luciferase activity in the absence of LMI070 for all candidate cassettes, which may reflect in-frame non-AUG start codon activity<sup>12</sup> (Extended Data Figs. 3b, 4). The SF3B3-on switch, which showed the least background splicing, was used in all subsequent studies and is hereafter referred to as X<sup>on</sup>.

We first compared X<sup>on</sup> induction from the Rous sarcoma virus (RSV), the phosphoglycerate kinase (PGK) and the minimal cytomegalovirus (mCMV) promoters. All promoters drove dose-responsive inducible expression that was mirrored by splicing (Fig. 1g, Extended Data Fig. 5). The induction strength was of the order PGK  $\approx$  RSV > mCMV, thereby providing tools for a gradient of induction.

To assess X<sup>on</sup> in vivo, eGFP was cloned into the cassette, validated in vitro (Extended Data Fig. 6a) and the construct packaged into AAV9. AAV9-X<sup>on</sup>-eGFP was administered intravenously to mice, and 4 weeks later the mice were given a single oral dose of either vehicle or LMI070 and eGFP expression assessed 24 h later in the liver, heart and skeletal muscle (Fig. 2a). There was notable eGFP fluorescence in liver and heart sections (Fig. 2b, Extended Data Fig. 6b), and the dose response of the eGFP signal that was observed by microscopy was confirmed by western blot (Fig. 2c, Extended Data Fig. 6c), RT-qPCR and splicing assay (Fig. 2d, Extended Data Fig. 6d-f). Notably, eGFP protein levels and novel exon splicing correlated directly to the dose of LMI070.

We next tested the responsiveness of the X<sup>on</sup> cassette to a second LMI070 dose. Mice were treated intravenously with AAV9-X<sup>on</sup>-eGFP, and 4 weeks later were given one oral dose of vehicle or of LMI070. One day later, some of the mice were euthanized, and eGFP induction was assessed by histology, splicing and RT-qPCR assays. The remaining mice underwent one week of drug washout, after which they were given LMI070 or vehicle (Fig. 2e). Induction in the liver was evident after the first and the second dose by histology, western blot, splicing and RT-qPCR assays (Fig. 2f-h, Extended Data Fig. 6g, h). Induction in the skeletal muscle and the heart was also observed (Extended Data Fig. 6h, i), the relative levels of which were greatest for the latter (around 400-fold and 3,000-fold, respectively). Notably, the fold induction for each tissue was similar after each dose (Fig. 2h, Extended Data Fig. 6i). Cumulatively, these data show that eGFP expressed from an AAV-encoded X<sup>on</sup> cassette remains undetectable in the uninduced state, and that LMI070 robustly induces expression.

We next assessed the efficacy of X<sup>on</sup> for controlling erythropoietin (Epo), which is used to treat anaemia associated with chronic kidney disease<sup>13</sup>. The X<sup>on</sup> cassette expressing mouse Epo (mEpo) was packaged in AAV8 (AAV8-X<sup>on</sup>-Epo) and the construct was delivered intravenously to mice, which were subsequently treated with LMI070 or vehicle (Fig. 3a). mEpo plasma levels increased 25-fold to 62-fold in response to 2.5 or 10 mg kg<sup>-1</sup>

of LMI070, respectively, and returned to baseline 5 days after induction. One day after induction, Epo levels in AAV-eGFP or AAV-mEpo vehicle-treated mice were  $173.6 \pm 26$  pg ml<sup>-1</sup> and  $446 \pm 40$  pg ml<sup>-1</sup>, respectively. For LMI070 treatment, mEpo levels increased in AAV8-X<sup>on</sup>-mEpo mice to  $4,333 \pm 797$  pg ml<sup>-1</sup> or  $10,870 \pm 1,383$  pg ml<sup>-1</sup> for doses of 2.5 mg kg<sup>-1</sup> or 10 mg kg<sup>-1</sup>, respectively (Fig. 3b, left). The elevation of mEpo induced an increase in haematocrit to 60–70% above baseline levels depending on dose, after which a slow return to basal levels was observed (Fig. 3b, right).

We next tested the re-inducibility of AAV-X<sup>on</sup>-mEpo in response to LMI070, as re-administrations are required for prolonged therapy in patients with chronic kidney disease (Fig. 3c). The re-induction of Epo generated similar results to the initial experiments, showing the practical utility of the X<sup>on</sup> system for the regulation of therapeutic transgenes in response to repeat oral dosing (Fig. 3d, e).

To assess the applicability of our system for brain-targeted gene therapies, the X<sup>on</sup>-eGFP cassette was packaged into AAVPHPeB<sup>14</sup> and delivered intravenously to mice (Fig. 4a). LMI070 induced robust eGFP expression throughout the brain, and all regions evaluated showed a clear dose response according to histological assessment (Fig. 4b, Extended Data Fig. 7a), western blot (Fig. 4c), RT-qPCR and splicing assays (Fig. 4d, Extended Data Fig. 7b).

Some gene-replacement therapies have only a narrow therapeutic expression window, including the expression of progranulin (PGRN) in the treatment of *PGRN*-deficient frontotemporal dementia and neuronal ceroid lipofuscinosis<sup>15</sup>. For X<sup>on</sup>-regulated delivery of PGRN, AAVPHPeB-X<sup>on</sup>-PGRN vectors were generated and administered intravenously to mice (Fig. 4e). Mice injected with AAVPHPeB vectors that constitutively express PGRN were used as a control. Although all groups showed similar *PGRN* mRNA levels after AAV injection (Extended Data Fig. 7c), PGRN protein was induced to levels of  $1,135 \pm 183$  pg ml<sup>-1</sup> and  $11,393 \pm 1,153$  pg ml<sup>-1</sup>, respectively, after LMI070 doses of 5 or 50 mg kg<sup>-1</sup> (Fig. 4f). Splicing data were consistent with these results (Extended Data Fig. 7d).

It would be desirable to be able to regulate the expression of gene-editing proteins delivered by viral vectors. As such, we tested X<sup>on</sup> for regulated *Staphylococcus aureus* Cas9 (SaCas9) editing in liver using Ai14 mice<sup>16</sup>. Because the X<sup>on</sup>-SaCas9 cassette exceeds current AAV packaging limits (5.6 kb), we generated a smaller version that we denote miniX<sup>on</sup> (Extended Data Fig. 8a). Induction and splicing of miniX<sup>on</sup> was maintained and its overall amplitude was only modestly reduced relative to SF3B3-X<sup>on</sup>; at 100 nM LMI070, induction of the miniX<sup>on</sup> was 132-fold compared with 152-fold for X<sup>on</sup> (Extended Data Fig. 8b, c). AAV8-miniX<sup>on</sup>-SaCas9 plus single-guide RNAs targeting the *loxP*-STOP sequence in Ai14 mice were administered intravenously; after several weeks, LMI070 was administered and gene editing was assessed 7 days later (Extended Data Fig. 8d). The expression of tdTomato was detected by the fluorescence-activated cell sorting analysis of hepatocytes and histology on liver tissue sections (Extended Data Fig. 8e,f). Additionally, a genomic DNA PCR assay followed by Sanger sequencing confirmed gene editing (Extended Data Fig. 8g, h). Cumulatively, our results expand the utility of X<sup>on</sup> to control Cas9 protein translation for gene-editing applications.

Substantial improvements have been made towards the development of small, potent, ubiquitous and tissue-specific promoters. However, the lack of efficient switches with which to fine-tune protein expression limits the broader application of gene therapy for disorders in which either loss of function or overexpression is consequential. Indeed, there are many clinical scenarios in which a narrow therapeutic window exists<sup>15,17</sup>. Moreover, toxicity can emerge as a result of sustained expression at supraphysiological levels. In AAV9 therapy for spinal muscular atrophy, the initial benefits of SMN overexpression were reversed by later toxicities caused by a splicing-machinery imbalance resulting from high levels of recombinant SMN in mice<sup>18</sup>. Although our work used AAV X<sup>on</sup>, it could also be engineered for chimeric antigen receptor (CAR) T cell therapies and applied to the setting of CAR-T cell exhaustion<sup>19</sup> to pause CAR expression. Similarly, CRISPR-based treatments will benefit from using tunable switches, because refined control to provide a short burst of expression of CRISPR effector proteins would reduce unwanted off-target editing<sup>20</sup>, and—because the proteins would be expressed for hours rather than for months or years—immune responses to nuclease-expressing cells would be limited<sup>21,22</sup>. Additionally, the transient exposure of cells to active nucleases should reduce the genomic integration of free AAVs that is seen upon prolonged expression of editing machinery<sup>23</sup>.

Our switch system was originally designed for use with RG7800 or LMI070, but the refined system, X<sup>on</sup>, was built around LMI070. Here we show that normally unspliced sequences that are rarely found in the transcriptome are spliced in in a dose-dependent manner. Although other splicing events induced by LMI070 represent potential drawbacks, we anticipate that most applications would entail infrequent dosing, providing substantial periods of washout that would enable the re-expression of gene products that are silenced by the inclusion of suicide exons. Moreover, the X<sup>on</sup> platform was developed using exons that were spliced in at very low drug exposures, further minimizing unwanted splicing events.

Most strategies for the control of gene expression have focused on controlling RNA levels by including *cis*-regulatory elements, although an approach to control protein translation—as is achieved with X<sup>on</sup>—has been reported<sup>24</sup>. In this system, L7Ae controls translation through binding to a specific RNA motif<sup>24</sup>; however, immune responses could compromise clinical translation. For RNA regulation, ribozyme-based switches have been engineered to self-cleave mRNA sequences in which they were embedded<sup>1</sup>. Recently, optimization of a type-III hammerhead ribozyme has been reported, with higher self-cleavage activity and tighter regulation using morpholino-blocking oligomers for in vivo application<sup>25</sup>. However, morpholinos have a slow off-rate, which renders fine control problematic. Additionally, morpholinos do not reach the brain when delivered systemically. A tetracycline-responsive ribozyme switch that combines the hammerhead ribozyme N79 from *Schistosoma mansoni* with a tetracycline-binding aptamer showed tetracycline induction<sup>26</sup>. However, background expression was high. Another approach uses microRNA-binding sites within the 3′ untranslated region of exogenous mRNAs to reduce transcript stability in undesired cell types by co-opting the endogenous RNA interference machinery<sup>27</sup>. However, although this controls an off state, it would have to be merged with a switch such as X<sup>on</sup> for precision in the on state.



The X<sup>on</sup> switch demonstrates robust control of protein expression, with the extent of induction managed by dose and by promoter strength. In this system, decreasing protein-expression levels are entirely dependent on the half-life of the protein (more than 1 week for eGFP (Fig. 2g, Extended Data Fig. 6g) and less than 5 days for mEpo (Fig. 3b, d)). Merging X<sup>on</sup> with protein domains targeted by proteolysis-targeting chimera molecules would provide another method to manage proteins with long half-lives<sup>28</sup>. Other possible improvements include adding an endoprotease site to the N-terminal region for fused miniexon sequence cleavage. Finally, we demonstrated control of CRISPR-effector protein induction using the X<sup>on</sup> switch, and in combination with regulated intein splicing<sup>29</sup> or drug-dependent nuclear import<sup>30</sup>, enhanced control of CRISPR activity could be further managed.

In summary, we present a simple, highly adaptable tool for regulated protein expression. We anticipate that this system could be adapted to any biological question or application in which the fine control of expression is desired.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-03770-2>.

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. The in vitro and in vivo experiments were not randomized and the investigators performing molecular and biological studies were not blinded to treatment group. Individuals involved in tissue collection and histology were blinded to treatment group.

### Cell culture, transfection and LMI070/RG7800 treatment

Human embryonic kidney (HEK293) cells (obtained from CHOP Research Vector Core stock) were maintained in DMEM media containing 10% fetal bovine serum (FBS), 1% L-glutamine and 1% penicillin/streptomycin at 37 °C with 5% CO<sub>2</sub>. Cells were cultured in 24-well plates and transfected at 80–90% confluence using Lipofectamine 2000 transfection reagent, according to the manufacturer's protocol. For all experiments, 4 h after plasmid transfection, cells were treated with LMI070 (MedChemExpress, HY-19620, suspended in DMSO) or RG7800 (MedChemExpress, HY-101792A, suspended in H<sub>2</sub>O) at the indicated concentrations. Cells were tested for mycoplasma by Research Vector Core. None of the lines used here are listed in the ICLAC database of commonly misidentified cell lines.

### Plasmids, primers and custom-made TaqMan gene-expression assays

All plasmids will be made available upon request and will be deposited at Addgene plasmid repository. All primer sequences and custom Taqman gene expression assays to determine

*SF3B3* novel exon inclusion are available upon request. Primers and custom Taqman gene expression assays were obtained from Integrated DNA Technologies (IDT).

### In vitro luciferase assays

HEK293 cells were cultured in DMEM (10% FBS (v/v), 1% penicillin/streptomycin (v/v) and 1% l-glutamine (v/v) in 24-well plates. At 70–80% confluence, cells were co-transfected with the  $X^{on}$ . Firefly luciferase cassettes (0.3  $\mu$ g per well) and a SV40p-*Renilla* luciferase cassette as transfection control (0.02  $\mu$ g per well). Four hours after transfection cells were treated with LMI070 or RG7800 at the indicated concentrations. At 24 hours post transfection, cells were rinsed with ice-cold PBS and *Renilla* and firefly luciferase activities were assessed using the Dual-Luciferase Reporter Assay System (Promega) as per the manufacturer's instructions. Luminescent readouts were obtained with a Monolight 3010 luminometer (Pharmigen). Relative light units (RLUs) were calculated as the quotient of *Renilla*/Firefly RLUs and results expressed relative to mock-treated control cells.

### Mice and histology

Animal protocols were approved by The Children's Hospital of Philadelphia Institutional Animal Care and Use Committee. Five to six-week-old male C57BL/6J mice used in  $X^{on}$  studies to control translation of eGFP, mouse erythropoietin and human progranulin were obtained from Jackson Laboratories. Five to six-week-old male and female Ai14 mice used in gene-editing studies were obtained from an in-house-bred Ai14 colony derived from mice obtained from Jackson Laboratories (stock no 007914). AAV vectors were generated at the CHOP Research Vector Core, and were administered by retro-orbital injection at 7–8 weeks old in a volume of 150–200  $\mu$ l. For the dose–response studies, several weeks after AAV administration, a single dose of LMI070 at 2.5, 5, 10 or 50 mg kg<sup>-1</sup> (MedChemExpress, HY-19620) or vehicle solution was administered by oral gavage. For the re-dose studies, LMI070 was re-administered at the indicated times and concentrations. At euthanasia, mice used for non-histology work were perfused with 0.9% cold saline mixed with 2-ml RNAlater (Ambion). Brains and liver samples were collected, flash-frozen in liquid nitrogen, and stored at –80 °C until use. For histological studies, mice were perfused with 15 ml ice-cold 0.1 M PBS followed by 15 ml 4% paraformaldehyde. Sections were analysed using a DM6000B Leica microscope equipped with a L5 ET filter cube (excitation and emission wavelengths of 470  $\pm$  20 nm and 525  $\pm$  15 nm, respectively; dichroic 495 nm), a 10X HC PLAPO (numerical aperture (NA) 0.40), and a 40X HC PLAPO CORR (NA 0.85) lens connected to a Sola Light Engine LED light source (Lumencor). Images were collected with a Hamamatsu Orca flash 4.0 monochrome camera controlled, or Leica SP8 confocal microscope equipped with a white-light laser, HyD sensors, and 40X HC PL APO CS2 (NA 0.75) lens. The excitation and emission wavelengths were set at 493 and 497–535 nm or 554 and 563–730 nm for the green and red channels. The Leica LAS X (v.3.0.3) software was used for both controlling microscopes and obtaining images.

### RNA extraction, RT–PCR and splicing assays

Total RNA was extracted using TRIzol (Life Technologies) according to the manufacturer's protocol, with the exception of 1  $\mu$ l Glycoblue (Life Technologies) in addition to the aqueous phase in the isopropanol precipitation step and a single wash with cold 70%



ethanol. To determine splicing of the SMN2-on and X<sup>on</sup> switches, 2 µg of total RNA from HEK293 cells or tissue samples was treated with DNaseI Free kit (Thermo Fisher) followed by cDNA generation using the High capacity cDNA kit (Thermo Fisher). Splicing was determined by PCR using the Phusion HighFidelity polymerase (Thermo Fisher) and PCR products separated on a 2.5% agarose gel pre-stained with EtBr and spliced-in and spliced-out band densitometry performed using with the ChemiDoc Imaging System (BioRad) and Image Lab analysis software. Splicing induction from mouse tissues was determined using two custom TaqMan assays designed to determine total or LMI070-spliced in mRNA transcripts. The percentage of induction was determined by dividing the average  $C_t$  novel exon and average  $C_t$  total, relative to control mice injected with AAV virus plus vehicle.

### Genomic DNA extraction and genome-editing analysis

Genomic DNA from liver samples was extracted using a Monarch genomic DNA purification kit (NEB) according to the manufacturer's instructions. Deletion of the *loxP*-STOP-*loxP* region of the transgenic *ROSA26* locus in the Ai14 mice was confirmed by PCR using primers binding outside the intervening segment cleaved by the sgRNA/SaCas9 complex pair (forward: 5'-GCT GGT TAT TGT GCT GTC TCA TC-3' and reverse: 5'-CAT GAA CTC TTT GAT GAC CTC CTC-3'). PCR products were cloned into TOPO plasmids using the TOPO TA Cloning Kit and subsequently transformed into DH5α competent cells. Individual colonies were analysed using Sanger sequencing to determine editing of the *loxP*-STOP cassette.

### Western blots

Tissues were homogenized in RIPA buffer (final concentration: 50 mM Tris, 150 mM NaCl, 1% Triton-X100, 0.1% SDS, 0.5% sodium deoxycholate, with Complete protease inhibitors (Roche)) and samples were incubated for 1 h rotating at 4 °C then clarified by centrifugation at 10,000g for 10 min. Total protein concentration was determined by DC protein assay (BioRad) and 30 µg was loaded onto 4–12% NuPAGE Bis-tris gels in MES buffer (Novex Life Technologies) to determine eGFP and β-catenin levels. After electrophoresis, proteins were transferred to 0.2 µm PVDF (Bio-Rad). Membranes were blocked with 5% milk in PBS-T and then blotted with rabbit anti-β-catenin (Ab2365, dilution 1:4,000; Abcam), rabbit anti-GFP (A11122, dilution 1:5,000; Invitrogen) followed by horseradish-peroxidase-coupled antibodies (goat anti-rabbit: 111-035-144, dilution 1:50,000; Jackson ImmunoResearch). Blots were developed with ECL Plus reagents (Amersham Pharmacia) and imaged on the ChemiDoc Imaging System (BioRad). eGFP and β-catenin protein levels were determined using the same PVDF membrane (whole blot). Restore western blot stripping buffer (21063, Thermo Fisher Scientific) was used to remove primary eGFP and secondary antibodies before β-catenin detection.

### Haematocrit and plasma epo assays

At each indicated time point, 70 µl of blood was collected from the eye into heparin-treated microhaematocrit capillary tubes (Fisher Scientific, 22-362-566) and centrifuged using a micro-haematocrit centrifuge (International Equipment Company model MB, 341118722). Haematocrit counts were then measured using a micro-haematocrit card reader (LW scientific, ZCP-EZRD-HEM7). Plasma samples in the capillary tubes were collected and

stored at  $-80^{\circ}\text{C}$ . Two- or fourfold-diluted (control and vehicle treated mice, respectively) or tenfold-diluted plasma samples (LMI070 treated mice) were used for measuring plasma Epo protein concentration using a Mouse Erythropoietin Quantikine ELISA Kit (R&D), and ELISA samples measured with a SpectraMax M5 plate reader (Molecular Devices). The normal haematocrit range was obtained from ref. <sup>31</sup>.

### Quantification of human progranulin by time-resolved FRET immunoassay

Brain samples were homogenized with Precellys in 400  $\mu\text{l}$  of STEN buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.6, 2 mM EDTA, 1% NP-40) with Complete Protease Inhibitor (Thermo, A32955) and PhosSTOP phosphatase inhibitor (Thermo, A32957). Brain extracts were then clarified by centrifugation for 20 min at 15,000 rpm at  $4^{\circ}\text{C}$ . Supernatant was collected, and total protein concentration was quantified by BioRAD DC protein assay kit (Bio-Rad, 5000116). Samples were aliquoted and frozen at  $-80^{\circ}\text{C}$  for future use as needed. Human progranulin antibodies (R&D Systems AF2420 and R&D Systems MAG2420) were labelled with donor Tb-fluorophore and acceptor d2-fluorophore, respectively (Cisbio). After optimization of antibody titres and incubation conditions, quantification of human progranulin levels was performed in a 384-well low-volume ProxiPlate (Perkin Elmer, 6008238). Then, 16  $\mu\text{l}$  of diluted sample was loaded in duplicate followed by the addition of 2  $\mu\text{l}$  antibody mix. A standard curve was prepared using Recombinant Human Progranulin protein standard (R&D Systems, 2420-PG-050). All sample and standard dilutions were performed in  $1\times$  PBS with 0.1% bovine serum albumin. For brain tissue, the standard was spiked with STEN buffer homogenized mouse brain extract to account for the matrix effects. Plates were then incubated at room temperature for 1 h. Time-resolved fluorescence resonance energy transfer readout was performed with an EnVision Reader (PerkinElmer). After the excitation of the donor fluorophore Tb at 320 nm and a time delay of 100  $\mu\text{s}$ , the resulting Tb and d2 emission signals were read at 615 nm and 665 nm, respectively.

### Hepatocyte isolation and flow cytometry

To isolate hepatocytes from Ai14 mice, whole livers were perfused (perfusion solution: cold 0.5 mM EDTA in  $\text{Ca}^{2+}$ -,  $\text{Mg}^{2+}$ -free HBSS), excised and disrupted with forceps. They were then minced into very small pieces using a razor blade and enzymatically digested for 2 h at  $37^{\circ}\text{C}$  with shaking (digestion solution: 3 mM  $\text{CaCl}_2$ , 200 units per ml collagenase IV, and 1.5% BSA in HBSS). The final hepatocyte suspension was passed through a 70- $\mu\text{m}$  nylon mesh cell strainer and rinsed to finish straining cells through the filter (washing solution: 1.5% BSA in HBSS ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ -, no phenol red)). Cells were pelleted by centrifugation and washed, before fixation at room temperature for 15 min (fixing solution: 4% PFA in HBSS). Fixed cells were pelleted, resuspended in FACS buffer (0.5% BSA, EDTA (2 mM), in PBS), and counted using Trypan Blue (1:10 dilution) to determine the total count of live cells. For flow cytometry, cells were aliquoted and stained with primary (ASGR1, Proteintech, 11739-1-AP) and secondary (Alexa647, Thermo Fisher, A32733) antibodies, all diluted at 1:100 in 0.1% PBS/BSA for 30 min at room temperature. Flow cytometry data was acquired using a BD FACSAria Fusion (BD Biosciences). ASGR1/A647 fluorescence was detected using a 640-nm laser and a 670/30 filter. tdTomato fluorescence was detected using a 561-nm laser and a 582/15 filter. Data were analysed using FlowJo 10 software (FlowJo LLC).

## RNA-seq

RNA-seq libraries were constructed using the Illumina TruSeq Stranded Total RNA with Ribozero gold protocol at the Iowa institute of Human Genomics. Data from four LMI070-treated (25 nM) and four DMSO-treated HEK293 cell groups were obtained after sequencing across two lanes run on an Illumina Hi-Seq 4000 150PE. The resulting fastq files were aligned to the GRCh38 human genome obtained from Ensembl using the STAR aligner<sup>32</sup>. Splice junction output by STAR were quantified using a custom R script designed to identify splicing events unique to LMI070 treatment. Top ranking LMI070-exclusive splice events were manually assessed for their applicability to function as a splicing switch. A primary requirement of this evaluation was that two splice events (donor and acceptor) were identified as exclusive or enriched in LMI070-treated cells creating inclusion of a pseudoexon of reasonable size. Candidate splice events were visually evaluated using the Sashimi plot function available in IGV<sup>33–35</sup>. Previously published RNA-seq results from ref. <sup>8</sup> (NVS-SM1 100nM, and DMSO) were obtained as fastq files from NCBI Gene Expression Omnibus. Splicing analysis was performed on these datasets using our analytical pipeline (as described above). NVS-SM1, also known as LMI070, induced splicing events from ref. <sup>8</sup> were compared against induced events obtained from our study. To assess the exclusivity of LMI070-induced splice sites to LMI070 treatment we evaluated the frequency with which candidate splice junctions were previously identified in diverse human RNA-seq datasets deposited in the Sequence Read Archive (SRA). This analysis was performed using Intropolis, a database of exon–exon junctions from 21,504 human RNA-seq samples in the SRA archive. The Intropolis database is indexed by GRCh37 genomic position so we first converted our GRCh38 positions to GRCh37 using the LiftOver tool from the UCSC genome browser<sup>36</sup>. We then queried LMI070-induced splice sites against the Intropolis database using a custom Python script. The results for each LMI070 candidate splice event are summarized in Extended Data Tables 1, 2. Differential gene expression analysis was performed using DESeq2 to compare samples from the LMI070 and DMSO conditions<sup>37</sup>. To visualize the abundance of meaningfully differentially expressed genes we generated a volcano plot with a Benjamini–Hochberg-adjusted *P* value threshold of 0.05 and a 0.1 log fold-change threshold.

## Statistical analysis

Statistical analyses were performed using GraphPad Prism v.7 software. Outlier samples were detected using the Grubb's test ( $\alpha = 0.05$ ). Normal distribution of the samples was determined using the D'Agostino and Pearson normality test. Data were analysed using one-way ANOVA or two-way ANOVA followed by a Bonferroni's post hoc test. Statistical significance was considered when  $P < 0.05$ . All results are shown as mean  $\pm$  s.e.m.

## Reporting summary

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

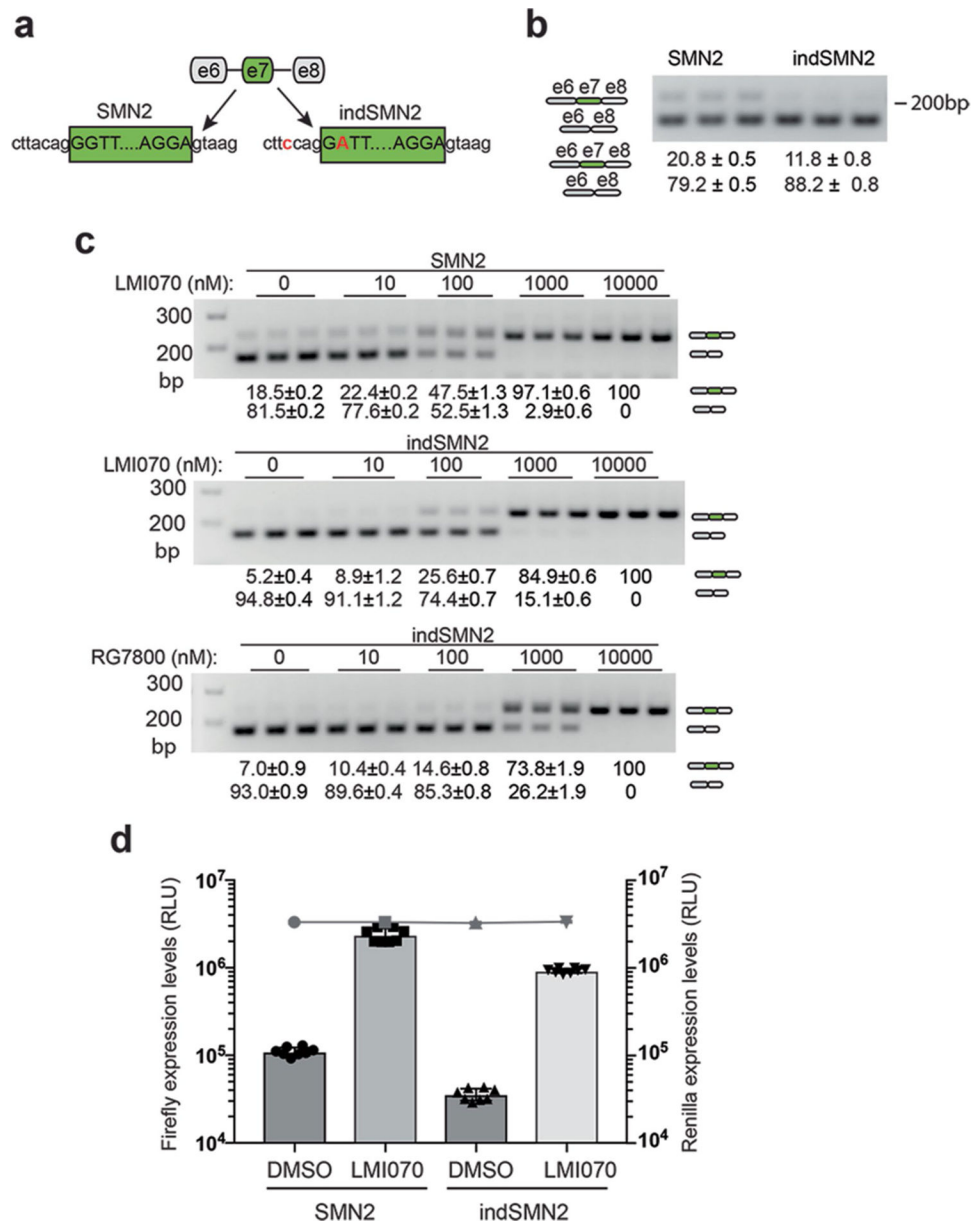
**Data availability**

RNA-seq datasets are archived in the NCBI Gene Expression Omnibus under accession number GSE176527.

**Code availability**

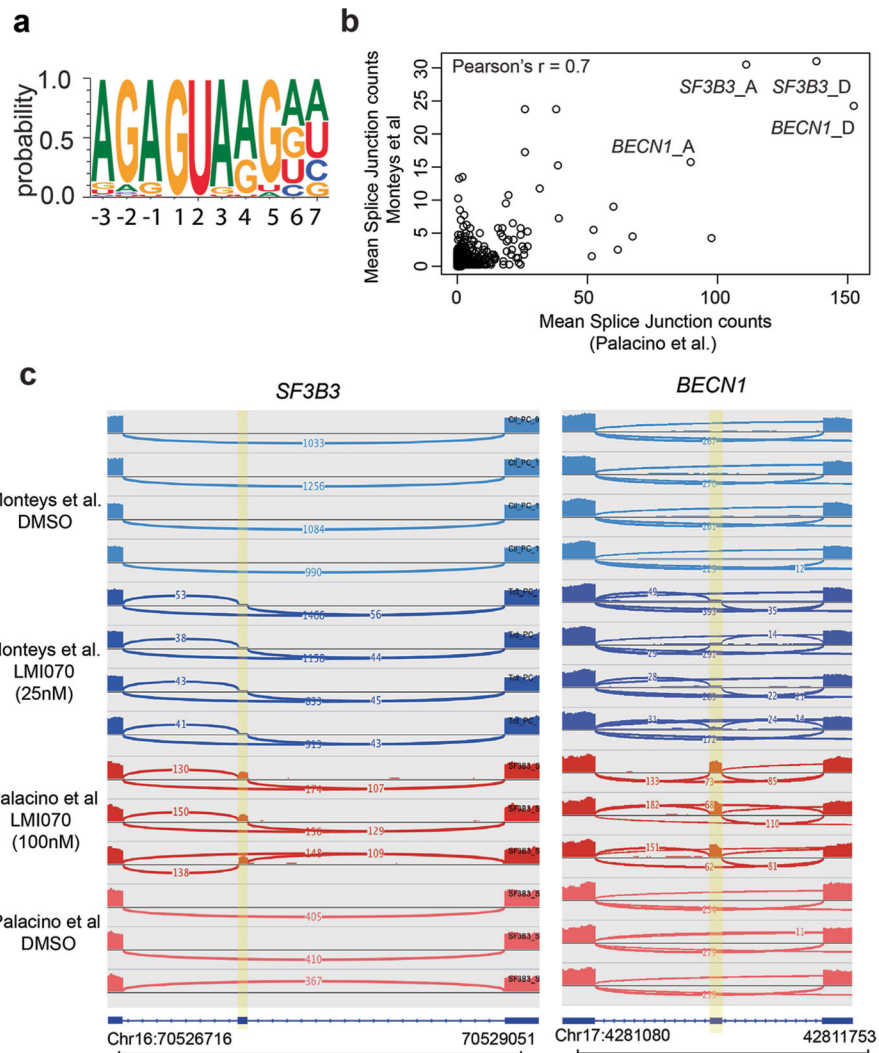
The R code to characterize novel LMI070-induced splicing events from RNA-seq data is available at <https://github.com/DavidsonLabCHOP/Xon>.

**Extended Data**



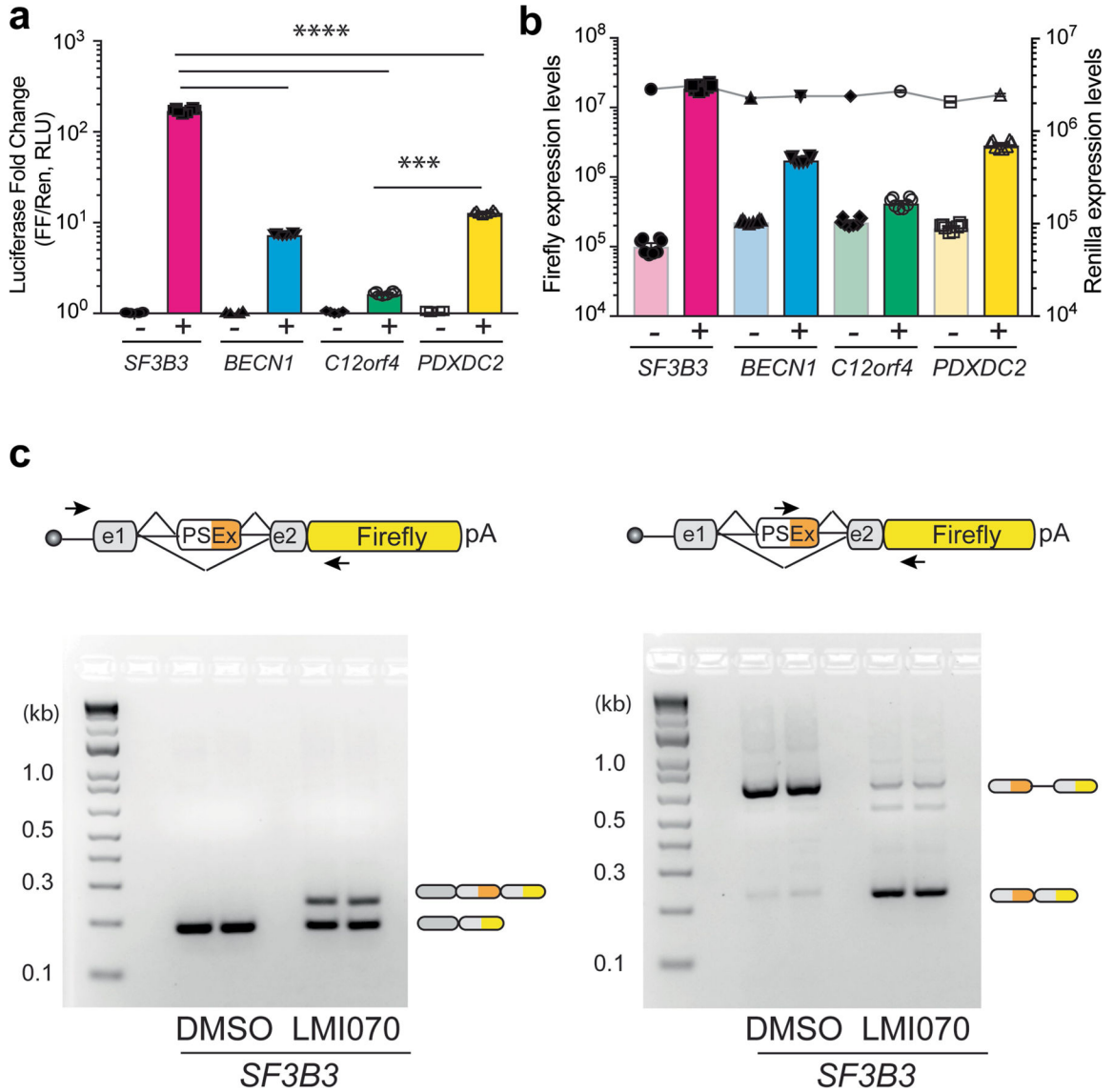
Extended Data Fig. 1 | In vitro assessment of SMN2-on cassettes.

**a**, Cartoon depicting *SMN2* exon 7 in its native sequence or with splice-site modifications introduced to reduce background levels of exon 7 inclusion (3' acceptor splice site, indSMN2). **b**, Representative RT-PCR reaction showing exon 7 inclusion with the SMN2-on cassettes in the absence of LMI070. The quantification of the exon 7 spliced-in or -out transcripts is depicted as the mean  $\pm$  s.e.m. of 6 biological replicates. **c**, Exon 7 splicing of the SMN2 and indSMN2 cassette in response to LMI070 or RG7800. Representative RT-PCR reaction showing exon 7 inclusion as a function of LMI070 or RG7800 dose. The quantification of the exon 7 spliced-in or spliced-out transcripts are the relative transcript levels presented as the mean  $\pm$  s.e.m. of 8 biological replicates. **d**, Luciferase activity of the SMN2 and indSMN2 cassettes in response to LMI070. The graph shows relative expression of firefly luciferase expressed from the SMN2-on or indSMN2-on cassettes in cells treated with DMSO or LMI070 (100 nM). The activity of the transfection control *Renilla* luciferase cassette is represented as a line above the bar graph. Data are the mean  $\pm$  s.e.m. of 8 biological replicates.



**Extended Data Fig. 2 | Comparison between induced splice junctions in a previous study and in this work.**

See ref. <sup>8</sup>. **a**, Sequence logo of U1 RNA site targeted by LMI070 from 45 spliced-in exons identified by RNA-seq. **b**, To identify splice junctions with the highest induction across both studies we plotted the mean splice junction counts from our study against the mean splice junction counts from the dataset in ref. <sup>8</sup> reprocessed using our pipeline. The datasets correlate with Pearson's  $r = 0.7$ , indicating that induced junctions performed similarly by rank across both datasets. Additionally, the top ranked splice junctions associated with *SF3B3* and *BECN1* were consistent across both datasets. **c**, Sashimi plots depicting novel LMI070-spliced in exons for the *SF3B3* and *BECN1* genes identified in our study and in ref. <sup>8</sup> by RNA-seq. Genomic location, position of the LMI070 spliced in exon, and intronic counts observed are indicated.

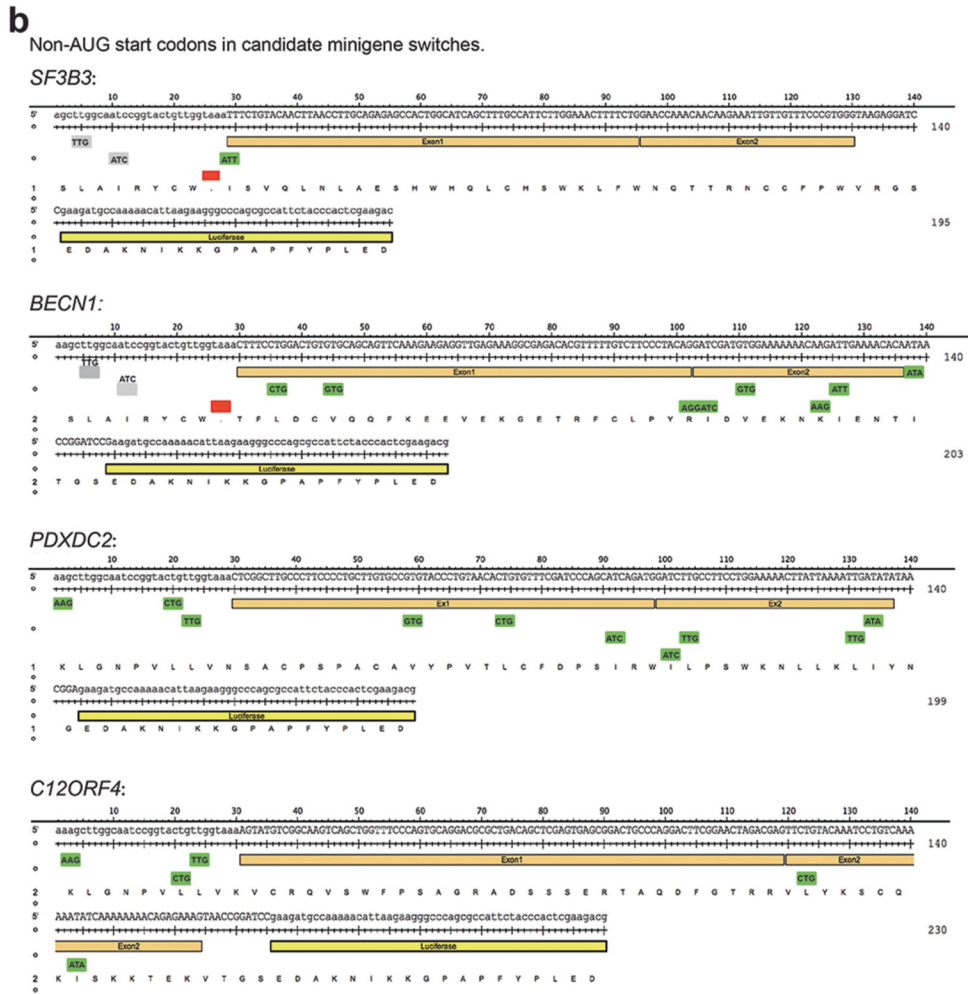


Extended Data Fig. 3 | Candidate minigene cassette responsiveness to LMI070 and splicing response to LMI070 of the SF3B3-X<sup>on</sup> cassette.



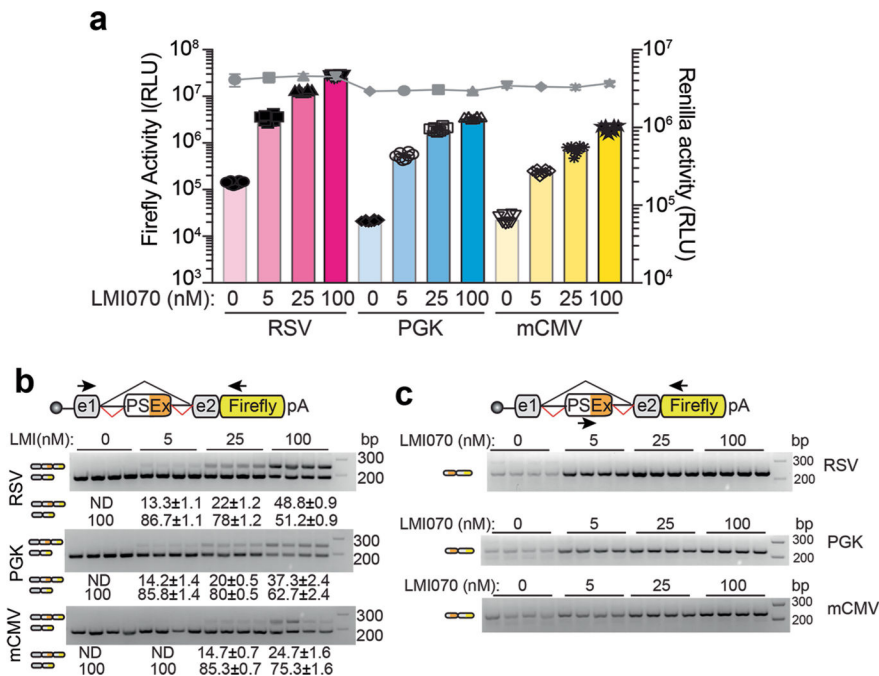
**a**, Luciferase induction of the minigene cassettes for *SF3B3*, *BECN1*, *C12ORF4* and *PDXDC2*. The fold-change luciferase activity in LMI070-treated samples (depicted as +) is relative to DMSO-treated (depicted as -) cells, with data normalized to *Renilla* luciferase expression. Data are the mean ± s.e.m. of 8 biological replicates. \*\*\*\**P* < 0.0001 for *SF3B3* versus other candidate exons, \*\*\**P* < 0.001 *C12orf4* versus *PDXDC2*, one-way ANOVA with Bonferroni's post hoc test. **b**, Luciferase activity of the minigene cassettes for *SF3B3*, *BECN1*, *C12ORF4* and *PDXDC2*. Data show expression of firefly luciferase from the minigenes in response to DMSO (-) or LMI070 (+) treatment relative to *Renilla* luciferase activity. Data are mean ± s.e.m. of 8 biological replicates. **c**, Splicing analysis of the *SF3B3*-X<sup>on</sup> cassette. Representative RT-PCR splicing assay (6 biological replicates) showing inclusion of the LMI070-induced *SF3B3* exon in response to DMSO or LMI070 treatment. Inclusion of the LMI070 spliced-in exon was detected using primers binding the exons flanking the LMI070-induced exon (left), or using primers binding within the novel exon sequence (right).

**a**  
 Codons: AUG>>CUG>>GUG>ACG>AUU=AUA>UUG=AUC>AAG=AGG  
 Frequency: 100 19 9 7 3 3 2 2 0.5 0.5



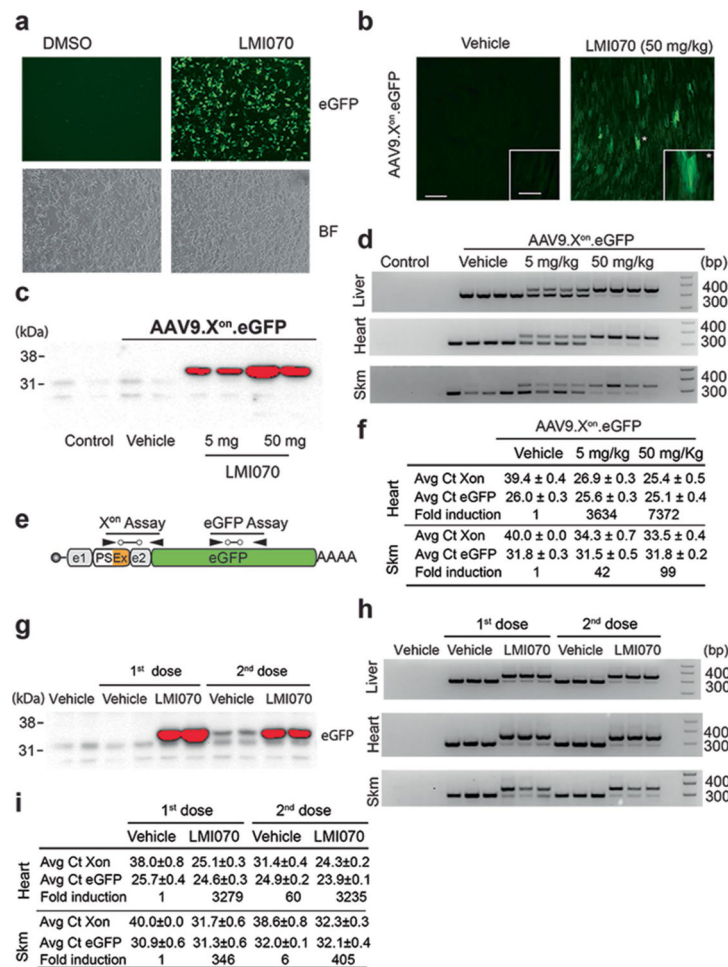
**Extended Data Fig. 4 |. Depiction of the frequency that non-AUG start codons are used determined by ribosome profiling.**

**a**, Translation frequency from AUG and non-AUG start codons determined from ribosome profiling<sup>39</sup>. **b**, Depiction of non-AUG start codons in frame luciferase transcripts expressed under control of the *SF3B3*, *BECN1*, *C12orf4* or *PDXDC2* minigenes.



**Extended Data Fig. 5 |. Analysis of SF3B3-X<sup>on</sup> expressed from different promoters.**

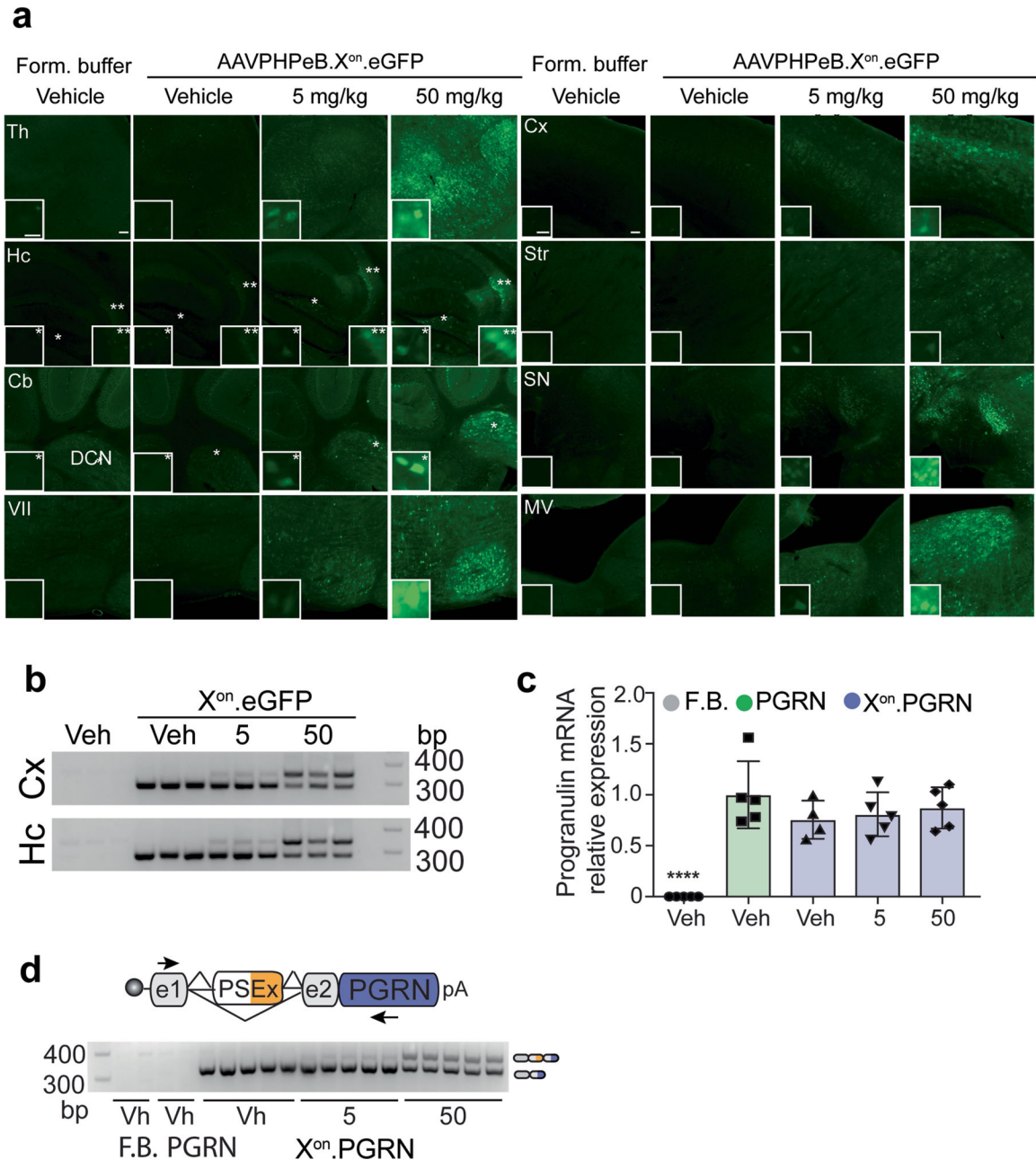
**a**, Firefly luciferase of the X<sup>on</sup> cassettes in response to varying doses of LMI070 relative to *Renilla* luciferase (grey line). The data are the mean ± s.e.m. of 8 biological replicates. **b**, Representative RT-PCR for assessment of LMI070-induced pseudo exon expression. Exon inclusion in the X<sup>on</sup> cassette was detected using primers flanking the pseudo exon. Data are the mean ± s.e.m. of 8 biological replicates. **c**, Representative gels from RT-PCR analysis for assessment of the LMI070-induced pseudo exons expressed from the noted promoters in response to varying doses of LMI070. Pseudo exon inclusion was detected using primers binding within the LMI070-induced pseudo exon and the downstream exon. Splicing was quantified and transcript levels presented as the mean ± s.e.m. of 8 biological replicates.



**Extended Data Fig. 6 | Assessment of the SF3B3-X<sup>on</sup> cassette to control translation of eGFP in vitro and in vivo.**

**a**, eGFP expression in HEK293 cells transfected with the SF3B3 minigene cassette (SF3B3-X<sup>on</sup>-eGFP) and treated 24 h later with DMSO (left) or LMI070 (right). Bright-field panels are shown below (4 technical replicates). **b**, Representative photomicrograph of heart tissue sections showing eGFP in heart 24 h after treatment with LMI070 at 50 mg kg<sup>-1</sup> (**n** = 2 mice). Scale bar, 200 μm. Inset, eGFP expression in cardiomyocytes at a higher magnification. Scale bar, 50 μm. **c**, Extended exposure of the western blot from Fig. 2c (4 mice per group). **d**, PCR assay demonstrates splicing activity in liver, heart and skeletal muscle in response to LMI070 (4 mice per group). **e**, Cartoon depiction of the X<sup>on</sup> assays designed to quantify the LMI070-induced transcripts and eGFP expression levels from the X<sup>on</sup> cassette after AAV9-X<sup>on</sup>-eGFP gene transfer. **f**, Data show average C<sub>t</sub> values for eGFP or LMI070-induced expression in heart and skeletal muscle. Fold change of the spliced-in expression cassette is shown relative to basal levels in mice injected with AAV9-X<sup>on</sup>-eGFP and treated with vehicle (4 mice per group). **g**, Extended exposure of the western blot from Fig. 2h (3 mice per group). **h**, PCR assay demonstrates splicing activity in liver, heart and skeletal muscle after each LMI070 dose (3 mice per group). **i**, Data show average C<sub>t</sub> values for eGFP or LMI070-induced expression in heart and skeletal muscle after each dose. Fold

change of the spliced expression cassette is shown relative to basal levels in mice injected with AAV9- $X^{on}$ -eGFP and treated with vehicle (3 mice per group).

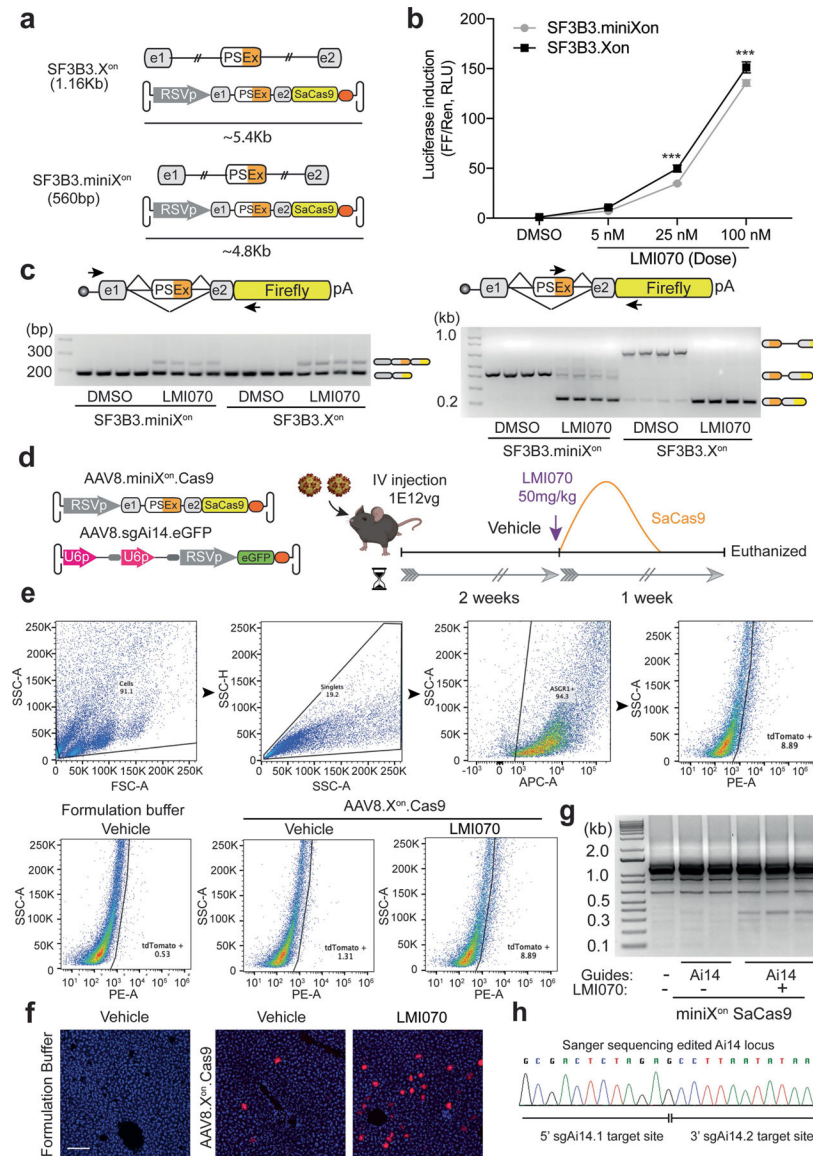


**Extended Data Fig. 7 | In vivo activity of  $X^{on}$  in brain.**

**a**, Representative photomicrographs (5 mice per group) showing eGFP expression from mice treated intravenously 4 weeks earlier with AAVPHPeB- $X^{on}$ -eGFP, and 24 h after treatment with vehicle or LMI070 at 5 or 50 mg kg<sup>-1</sup>. Thalamus (Th), hippocampus (Hc), cerebellum (Cb) and facial motor nucleus (VII), cortex (Cx), striatum (Str), substantia nigra (SN), and medial vestibular nucleus (MV) are shown. Scale bar, 100  $\mu$ m; inset scale bar, 25  $\mu$ m. In the hippocampus, \* and \*\* denote the polymorphic and CA3 areas, respectively. In the



cerebellum, \* denotes the deep cerebellar nuclei. **b**, Splicing assays for exon inclusion in the cortex and the hippocampus of mice injected with AAVPHPeB-X<sup>on</sup>-eGFP (3 mice per group). **c**, RT-qPCR of human progranulin expression. Data are mean ± s.e.m. of 5 mice per group, \*\*\*\**P* < 0.0001 vehicle versus AAV-treated groups, one-way ANOVA followed by Bonferroni's post-hoc test. **d**, Splicing assays for exon inclusion in cortex samples of mice injected with AAVPHPeB-X<sup>on</sup>-PGRN (3 mice per group).



**Extended Data Fig. 8 | Generation of the miniX<sup>on</sup> cassette and assessment of miniX<sup>on</sup> control of SaCas9 for in vivo gene editing in liver.**

**a**, Cartoon depicting the AAV genome size with the SF3B3-X<sup>on</sup> and SF3B3-miniX<sup>on</sup> cassettes. **b**, Luciferase induction in HEK293 cells transfected with SF3B3-miniX<sup>on</sup>-luciferase or SF3B3-X<sup>on</sup>-luciferase in response to varying doses of LMI070. All samples are normalized to *Renilla* luciferase activity and are relative to DMSO treated cells. Data are the mean ± s.e.m. of 8 biological replicates (\*\*\**P* < 0.001 versus SF3B3.X<sup>on</sup>, two-way ANOVA

followed by Bonferroni's post hoc test). **c.** Splicing inclusion assays of the LMI070-induced exon at 100 nM LMI070. Pseudo exon inclusion in the X<sup>on</sup> cassette was detected using primers flanking the pseudoexon (left) or by priming within the novel exon sequence (right; 4 technical replicates). **d.** Experimental design. Mice were injected with AAV8-miniX<sup>on</sup>-SaCas9 plus AAV8-sgAi14-eGFP (1 × 10<sup>12</sup> viral genomes, 1:1 ratio) and 2 weeks later dosed with vehicle or LMI070 at 50 mg kg<sup>-1</sup> to induce SaCas9 expression and editing of the *loxP*-STOP cassette (guides: sgA14\_1: 5'-CTCTAGAGTCGCAGATCCTC-3', sgA14\_2: 5'-ACGAAGTTATATTAAGGTT-3'). One week later, mice were euthanized, and livers processed to assess gene editing by genomic DNA PCR, histology and FACS of isolated hepatocytes. **e.** Representative FACS analysis of hepatocytes obtained from Ai14 mice after LMI070 or vehicle treatment. The gating/sorting strategy (above), and the percentage of tdTomato expressing cells for each condition (below) is shown (4 biological replicates). **f.** Representative photomicrographs of liver sections obtained from AAV injected Ai14 mice 1-week after LMI070 treatment. tdTomato expression (red) is evident in LMI070 treated mice (5 mice per group). Scale bars, 100 μm. **g.** SaCas9 mediated editing of the *loxP*-STOP cassette in Ai14 mice as detected by PCR assay of liver genomic DNA (3 of 5 mice with guides plus LMI, 2 of 4 mice with guides plus vehicle, 1 of 2 untreated mice are shown). A PCR product of 355 bp size corresponding to the edited Ai14 ROSA Locus was observed in the LMI070-treated mice. **h.** Sanger sequencing of the 355 bp PCR product confirmed targeted deletion of the *loxP*-STOP cassette and DNA repair of the Ai14 reporter locus.

**Extended Data Table 1 |**

The genomic locations of candidate LMI070-induced exons and the frequency of events from RNA-seq datasets

	Gene ID	GRCh38				GRCh37 (hg19)	
		Genomic Minigene region	Pseudo exon position	DMSO Observed Avg Counts (J1, J2)	LMI070 Observed Avg Counts (J1, J2)	Genomic Minigene region	Pseudo exon position
Exclusive	<i>SF3B3</i>	chr16:70,526,657–70,529,199	chr16:70,527,376–70,527,429	0, 0	31, 30.5	chr16:70560560–70563102	chr16:70561279–70561332
	<i>BENC1</i>	chr17:42,810,759–42,811,797	chr17:42,811,292–42,811,330	0, 0	24.45, 15.75	chr17:40962777–40963815	chr17:40963310–40963348
	<i>GXYLT1</i>	chr12:42,087,786–42,097,614	chr12:42,095,151–42,095,214	0, 0	10.75, 23.75	chr12:42481588–42491416	chr12:42488953–42489016
	<i>SKPI</i>	chr5:134,173,809–134,177,053	chr5:134,175,284–134,175,385	0, 0	5.75, 23.75	chr5:133509500–133512744	chr5:133510975–133511076
	<i>SKPI</i>	chr5:134,173,809–134,177,053	chr5:134,175,284–134,175,423	0, 0	15.25, 11.75	chr5:133509500–133512744	chr5:133510975–133511114
	<i>C12orf4</i>	chr12:4,536,017–4,538,508	chr12:4,537,380–4,537,514	0, 0	17.5, 19	chr12:4645183–4646764	chr12:4646546–4646680
	<i>SSBP1</i>	chr7:141,739,167–141,742,229	chr7:141,741,310–141,741,459	0, 0	17.25, 2.25	chr7:141438967–141442029	chr7:141441110–141441259
	<i>RARS</i>	chr5:168,517,815–168,519,190	chr5:168,518,369–168,518,523	0, 0	13.5, 1.5	chr5:167944820–167946195	chr5:167945374–167945528
	<i>RARS</i>	chr5:168,517,815–168,519,190	chr5:168,518,469–168,518,523	0, 0	13.6, 1.75	chr5:167944820–167946195	chr5:167945474–167945528



	Gene ID	GRCh38				GRCh37 (hg19)		
		Genomic Minigene region	Pseudo exon position	DMSO Observed Avg Counts (J1, J2)	LMI070 Observed Avg Counts (J1, J2)	Genomic Minigene region	Pseudo exon position	
	<i>PDXDC2P</i>	chr16:70,030,988–70,031,968	chr16:70,031,186–70,031,248	0, 0	13.25, 10.25	chr16:70064891–70065871	chr16:70065089–70065151	A
	<i>STRADB</i>	chr2:201,469,953–201,473,076	chr2:201,470,907–201,471,111	0, 0	9.5, 5.25	chr2:202334676–202337799	chr2:202335630–202335834	A
	<i>WNK1</i>	chr12:894,562–896,732	chr12:895,161–895,196	0, 0	9, 5.5	chr12:1003728–1005898	chr12:1004327–1004362	A
	<i>WDR27</i>	chr6:169,660,663–169,662,424	chr6:169,661,703–169,661,750	0, 0	8.5, 7.25	chr6:170060759–170062520	chr6:170061799–170061846	A
	<i>CIP2A</i>	chr3:108,565,355–108,566,638	chr3:108,565,898–108,565,931	0, 0	7.75, 5	chr3:108284202–108285485	chr3:108284745–108284778	A
	<i>ITF57</i>	chr3:108,191,521–108,206,696	chr3:108,192,476–108,192,526	0, 0	7.25, 5.25	chr3:107910368–107925543	chr3:107911323–107911373	A
	<i>HTT</i>	chr4:3,212,555–3,214,145	chr4:3213622–3213736	0, 0	7, 2.25	chr4:3214282–3215872	chr4:3215349–3215463	A
	<i>SKA2</i>	chr17:59,112,228–59,119,514	chr17:59119395–59119495	0, 0	6.75, 1	chr17:57189589–57196875	chr17:57196756–57196856	A
	<i>EVC</i>	chr4:5,733,318–5,741,822	chr4:5741334–5741441	0, 0	6.5, 3	chr4:5735045–5743549	chr4:5743061–5743168	A
	<i>DYRK1A</i>	chr21:37,420,144–37,473,056	chr21:37422581–37422652	0, 0	6.25, 6	chr21:38792446–38845358	chr21:38794883–38794954	A
	<i>GNAQ</i>	chr9:77,814,652–77,923,557	chr9:77920648–77920703	0, 0	6, 1	chr9:80429568–80538473	chr9:80535564–80535619	A
	<i>ZMYM6</i>	chr1:35,019,257–35,020,472	chr1:35020261–35020279	0, 0	5.75, 4	chr1:35484858–35486073	chr1:35485862–35485880	A
	<i>CYB5B</i>	chr16:69,448,031–69,459,260	chr16:69448605–69448753	0, 0	5.75, 1.25	chr16:69481934–69493163	chr16:69482508–69482656	T
	<i>MMS22L</i>	chr6:97,186,342–97,229,533	chr6:97201362–97201465	0, 0	5.75, 2	chr6:97634218–97677409	chr6:97649238–97649341	G
	<i>MEMO1</i>	chr2:31,883,262–31,892,301	chr2:31,887,035–31887087	0, 0	5, 2.25	chr2:32108331–32117370	chr2:32112104–32112156	A
<i>PNISR</i>	chr6:99,416,278–99,425,413	chr6:99420523–99420584	0, 0	5, 4	chr6:99864154–99873289	chr6:99868399–99868460	A	
<b>Enriched</b>	<i>CACNA2D1</i>	chr7:82,066,406–82,084,958	chr7:82,076,016–82,076,122	0.25, 0.75	18.5, 1.5	chr7:81695722–81714274	chr7:81705332–81705438	C
	<i>SSBPI</i>	chr7:141,739,083–141,742,248	chr7:141741310–141741459	0.25, 0	16.75, 2.25	chr7:141438883–141442048	chr7:141441110–141441259	A
	<i>DDX42</i>	chr17:63,805,048–63,806,672	chr17:63,806,151–63,805,994	0.25, 0	13.75, 4.5	chr17:61882408–61884032	Unable to lift over	A
	<i>ASAP1</i>	chr8:130,159,817–130,167,688	chr8:130,160,785–130,160,793	0.25, 0.25	13, 11.5	chr8:131172063–131179934	chr8:131173031–131173039	A
	<i>DUXAP10</i>	chr14:19,294,564–19,307,199	chr14:19,305,354–19,305,469	0.25, 0.75	9.5, 1.25	chr14:19882243–19894878	chr14:19893035–19893150	A
	<i>AVL9</i>	chr7:32,558,783–32,570,372	chr7:32,562,558–32,562,913	0.25, 0.5	7.5, 1.5	chr7:32598395–32609984	chr7:32602170–32602525	A
	<i>DYRK1A</i>	chr21:37,419,920–37,472,960	chr21:37,422,582–37,422,652	0.25, 0	6.25, 6	chr21:38792222–38845262	chr21:38794884–38794954	A

	Gene ID	GRCh38				GRCh37 (hg19)	
		Genomic Minigene region	Pseudo exon position	DMSO Observed Avg Counts (J1, J2)	LMI070 Observed Avg Counts (J1, J2)	Genomic Minigene region	Pseudo exon position
	<i>FAM3A</i>	chrX:154,512,311–154,512,939	chrX:154,512,568–154,512,706	0.25, 0	5.75, 1	chrX:153740635–153741263	chrX:153740892–153741030
	<i>FHOD3</i>	chr18:36,740,620–36,742,886	chr18:36,742,377–36,742,468	0.5, 0.25	15.25, 3.5	chr18:34320583–34322849	chr18:34322340–34322431
	<i>TBCA*</i>	chr5:77,707,994–77,777,000	chr5:77,774,217	0.5	14	chr5:77003819–77072824	chr5:77070041
	<i>MZT1</i>	chr13:72,718,939–72,727,611	chr13:72,725,642–72,725,778	0.5, 1	13.25, 1.25	chr13:73293077–73301749	chr13:73299780–73299916
	<i>LINC01296</i>	chr14:19,092,877–19,096,652	chr14:19094556–19094671	0.5,	8.25,	chr14:19680556–19684333	chr14:19682237–19682352
	<i>SF3B3</i>	chr16:70,541,627–70,544,553	chr16:70544169–70544249	0.5, 0	8.25, 3	chr16:70575530–70578456	chr16:70578072–70578152
	<i>SAFB</i>	chr19:5,654,060–5,654,457	chr19:5,654,140–5,654,368	0.5, 0	6.25, 2	chr19:5654071–5654468	chr19:5654151–5654379
	<i>GCFC2</i>	chr2:75,702,163–75,706,652	chr2:75,702,691–75,702,807	0.5, 0	6.25, 2	chr2:75929289–75933778	chr2:75929817–75929933
	<i>MRPL45</i>	chr17:38,306,450–38,319,088	chr17:38,312,587–38,312,661	0.5, 0	5.75, 1.25	chr17:36462417–36474972	chr17:36468550–36468624
	<i>SPIDR</i>	chr8:47,260,788–47,280,196	chr8:47,273,337–47,273,450	0.5, 0	5.5, 1.75	chr8:48173380–48192784	chr8:48185929–48186042
	<i>DUXAP8</i>	chr22:15,815,315–15,828,713	chr22:15,817,119–15,817,234	0.75, 0.25	13.75, 1.25	chr14:19,680,685–19,691,354	chr14:19682237–19682352
	<i>PDXDC1</i>	chr16:15,008,772–15,009,763	chr16:15,009,499–15,009,561	2, 1	16.75, 1.5	chr16:15102629–15103620	chr16:15103356–15103418
	<i>MANIA2</i>	chr1:117,442,104–117,461,030	chr1:117,456,085–117,456,206	0.75, 1	8, 5.25	chr1:117984726–118003652	chr1:117998707–117998828
	<i>RAF1</i>	chr3:12,600,376–12,604,350	chr3:12,603,478–12,603,537	1, 0.25	16.5, 1	chr3:12641875–12645849	chr3:12644977–12645036
	<i>ERGIC3</i>	chr20:35,548,787–35,554,452	chr20:35,549,163–35,549,207	1,0	7.5, 2.5	chr20:34136540–34142223	chr20:34136917–34136961

Summarized in this table are the differentially expressed candidate LMI070-induced splicing positions as identified by RNA-seq of HEK293 cells treated with either DMSO or LMI070 (25 nM). All candidates shown were manually selected from a bioinformatically generated list of top hits on the basis of their suitability for construction of an exon switching genomic minigene, their exclusivity to the LMI070 condition, and their minimal to undetectable levels in DMSO-treated cells. The top 25 rows (shaded green) indicate hits observed exclusively upon LMI070 exposure. The following 22 rows (shaded yellow) indicate candidates for which splicing was enriched but not totally exclusive to LMI070 treatment. Columns (from left to right) indicate: 1. Exclusivity to LMI070 induction. 2. The gene ID containing the splicing events of interest. 3. The GRCh38 genomic positions used to create the splice event-containing genomic minigene. 4. The GRCh38 genomic positions of the pseudo exon created by LMI070-induced splicing. 5. The average number of exon–exon junction spanning reads observed with DMSO treatment. 6. The average number of exon–exon junctions spanning reads observed with LMI070 treatment. To assess the frequency with which LMI070-induced events occur we queried Intropolis<sup>38</sup>, a database containing the frequency of splicing events observed across 21,504 human RNA-seq samples, representing a diverse set of human tissues and conditions. The reference genome used for the Intropolis database is GRCh37 so the LiftOver feature from the UCSC genome browser was used to convert the GRCh38 coordinates to GRCh37. 7. The GRCh37 genomic position of the minigene. 8. The GRCh37 genomic position of the pseudo exon. 9. The DNA sequence of the LMI070 binding sequence in the pseudo exon.

**Extended Data Table 2 |**

The genomic locations of candidate LMI070-induced exons and the frequency of events from RNA-seq datasets—continued

		GRCh37 (hg19)								
		Canonical Junction Metrics			LMI070 Junction1 Metrics					J2
	Gene ID	CJ exon-exon junction	CJ Intropolis Datasets	CJ_Total Counts	J1 Position	J1 Num Datasets	J1 % Data Sets	J1 Total Counts	J1 % Total counts	J2
Exclusive	<i>SF3B3</i>	chr16:70560630–70562775	12873	825337	chr16:70560630–70561278	10	0.08	13	0.002	chr16:70562
	<i>BENC1</i>	chr17:40962947–40963672	13635	910210	chr17:40962947–40963309	301	2.21	552	0.061	chr17:40963
	<i>GXYLT1</i>	chr12:42481750–42491243	7430	43383	chr12:42481750–42488952	8	0.11	9	0.021	chr12:42491
	<i>SKP1</i>	chr5:133509714–133512545	13277	2756168	chr5:133509714–133510974	15	0.11	16	0.001	chr5:13351
	<i>SKP1</i>	chr5:133509714–133512545	13277	2756168	chr5:133509714–133510974	15	0.11	16	0.001	chr5:13351
	<i>C12orf4</i>	chr12:4645386–4647574	9856	102163	chr12:4645386–4646545	354	3.59	563	0.551	chr12:46475
	<i>SSBP1</i>	chr7:141438991–141441968	14501	2473164	chr7:141438991–141441109	38	0.26	71	0.003	chr7:14144
	<i>RARS</i>	chr5:167945068–167946085	13602	826669	chr5:167945068–167945373	20	0.15	21	0.003	chr5:16794
	<i>RARS</i>	chr5:167945068–167946085	13602	826669	chr5:167945068–167945473	2	0.01	2	0.000	chr5:16794
	<i>PDXDC2P</i>	chr16:70064970–70065802	9586	117767	chr16:70064970–70065088	284	2.96	587	0.498	chr16:70065
	<i>STRADB</i>	chr2:202334776–202337677	9420	155599	chr2:202334776–202335629	150	1.59	189	0.121	chr2:20233
	<i>WNK1</i>	chr12:1003802–1005236	12935	800572	chr12:1003802–1004326	30	0.23	39	0.005	chr12:10052
	<i>WDR27</i>	chr6:170060863–170062399	9937	128203	chr6:170060863–170061798	1322	13.30	2407	1.877	chr6:17006
	<i>CIP2A</i>	chr3:108284302–108285343	9210	124822	chr3:108284302–108284744	126	1.37	181	0.145	Not Fo
	<i>ITF57</i>	chr3:107910491–107925474	12662	583056	chr3:107910491–107911322	64	0.51	150	0.026	chr3:10792
	<i>HTT</i>	chr4:3214437–3215684	11146	243427	chr4:3214437–3215348	452	4.06	599	0.246	chr4:32156
	<i>SKA2</i>	chr17:57189707–57196679	10377	297301	chr17:57189707–57196756	373	3.59	605	0.203	Not Fo
	<i>EVC</i>	chr4:5735163–5743442	6716	74348	chr4:5735163–5743060	86	1.28	120	0.161	chr4:57434
	<i>DYRK1A</i>	chr21:38792687–38844985	10821	152058	chr21:38792687–38794883	29	0.27	51	0.034	chr21:38844
<i>GNAQ</i>	chr9:80430687–80537076	10041	224120	chr9:80430687–80535563	23	0.23	28	0.012	chr9:80537	
<i>ZMYM6</i>	chr1:35485204–35485983	11692	199379	chr1:35485204–35485861	312	2.67	1041	0.522	chr1:35485	

		GRCh37 (hg19)								
		Canonical Junction Metrics			LMI070 Junction1 Metrics					
	Gene ID	CJ exon-exon junction	CJ Intropolis Datasets	CJ_Total Counts	J1 Position	J1 Num Datasets	J1 % Data Sets	J1 Total Counts	J1 % Total counts	J2
	<i>CYB5B</i>	chr16:69482048–69492995	13834	1150919	chr16:69482048–69482509	191	1.38	286	0.025	chr16:69492995
	<i>MMS22L</i>	chr6:97634567–97676769	8959	91164	chr6:97634567–97649237	41	0.46	58	0.064	chr6:97676769
	<i>MEMO1</i>	chr2:32108532–32117060	8678	99228	chr2:32108532–32112103	167	1.92	230	0.232	chr2:32117060
	<i>PNISR</i>	chr6:99864305–99873090	13184	888372	chr6:99864305–99868398	1	0.01	1	0.000	chr6:99873090
<b>Enriched</b>	<i>CACNA2D1</i>	chr7:81695841–81714084	6256	96528	chr7:81695841–81705331	1	0.02	1	0.001	chr7:81714084
	<i>SSBP1</i>	chr7:141438991–141441968	14501	2473164	chr7:141438991–141441109	38	0.26	71	0.003	chr7:141441968
	<i>DDX42</i>	chr17:61882536–61883894	12718	722896	NA	NA	NA	NA	NA	NA
	<i>ASAP1</i>	chr8:131172211–131179781	11448	396871	chr8:131172211–131173030	2429	21.22	16818	4.238	chr8:131179781
	<i>DUXAPI0</i>	chr14:19884030–19894699	638	1502	Not Found	Not Found	0.00	Not Found	0.000	chr14:19894699
	<i>AVL9</i>	chr7:32599077–32609631	11339	199392	chr7:32599077–32602169	2728	24.06	6204	3.111	chr7:32609631
	<i>DYRK1A</i>	chr21:38792687–38844985	10821	152058	chr21:38792687–38794883	29	0.27	51	0.034	chr21:38844985
	<i>FAM3A</i>	chrX:153740736–153741146	9459	98652	chrX:153740736–153740891	351	3.71	481	0.488	chrX:153741146
	<i>FHOD3</i>	chr18:34320802–34322699	7279	113308	chr18:34320802–34322339	210	2.89	295	0.260	chr18:34322699
	<i>TBCA*</i>	chr5:77004173–77072028	14180	942768	NA	NA	NA	NA	NA	chr5:77072028
	<i>MZT1</i>	chr13:73293236–73301661	12409	516948	chr13:73293236–73299779	51	0.41	103	0.020	chr13:73301661
	<i>LINC01296</i>	chr14:19680686–19683691	239	315	chr14:19680686–19682236	20	8.37	22	6.984	chr14:19683691
	<i>SF3B3</i>	chr16:70575738–70578340	12928	880623	chr16:70575738–70578071	233	1.80	467	0.053	chr16:70578340
	<i>SAFB</i>	chr19:5654212–5654378	13168	1083231	Not Found	Not Found	0.00	Not Found	0.000	Not Found
	<i>GCFC2</i>	chr2:75929550–75933648	11225	168121	chr2:75929550–75929816	124	1.10	195	0.116	chr2:75933648
	<i>MRPL45</i>	chr17:36462599–36474585	12718	608565	chr17:36462599–36468549	778	6.12	1191	0.196	chr17:36474585
	<i>SPIDR</i>	chr8:48173584–48192449	10729	158491	chr8:48173584–48185928	414	3.86	568	0.358	chr8:48192449
	<i>DUXAP8</i>	chr14:19680686–19683691	239	315	chr14:19680686–19682236	20	8.37	22	6.984	chr14:19683691
	<i>PDXDC1</i>	chr16:15102705–15103357	12939	767189	chr16:15102705–15103355	1472	11.38	3778	0.492	chr16:15103357

GRCh37 (hg19)										
	Canonical Junction Metrics				LMI070 Junction1 Metrics					
	Gene ID	CJ exon-exon junction	CJ Intropolis Datasets	CJ_Total Counts	J1 Position	J1 Num Datasets	J1 % Data Sets	J1 Total Counts	J1 % Total counts	J2
	<i>MANIA2</i>	chr1:117984948–118003110	11582	229559	chr1:117984948–117998706	427	3.69	914	0.398	chr1:118003110
	<i>RAF1</i>	chr3:12641915–12645634	12808	832362	chr3:12641915–12644976	1841	14.37	4800	0.577	chr3:12645634
	<i>ERGIC3</i>	chr20:34136619–34142142	4423	84287	chr20:34136619–34136916	2778	62.81	7045	8.358	chr20:34142142

Columns (from left to right) indicate: 1. Exclusivity to LMI070 induction. 2. The gene ID containing the splicing events of interest. 3. The position of canonical splice junction (CJ). 4. The number of Intropolis RNA-seq datasets in which each canonical splice event was observed. 5. The total number of observations identified for each canonical splice site. 6. The position of junction 1 (J1) and the first LMI070-induced exon–exon junction (sorted by genomic position) connecting a canonical exon to a LMI070-induced pseudo exon. Columns 7 and 8 indicate the number and percentage of Intropolis datasets in which each J1 splice event was observed. Columns 9 and 10 indicate the number and percentage of total counts in which each J1 splice event was observed. 11. The position of junction 2 (J2) the second LMI070-induced exon–exon junction (sorted by genomic position) connecting a LMI070-induced pseudo exon to a canonical exon. Columns 12 and 13 list the number and percentage of Intropolis datasets in which the LMI070-induced splicing event was observed. Columns 14 and 15 indicate the total number and percentage of reads containing each junction in the Intropolis dataset.

**Extended Data Table 3 |**

Comparison between disclosed hits in a previous study and in this work

gene	GRCh38	Splice Junction Found	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ABHD10	chr3:111990700–111990751	3:111987052–111990699	9.809816136	0.614834912	0.466585174	1.317733495	0.187592878	0.99999
ADAM12	chr10:126041344–126042235	Not Found						
AKT1	chr14:104793127–104793127	14:104792723–104793126	41.80684788	–0.26101857	0.231973571	–1.125208222	0.26050081	0.99999
ANXA11	chr10:80180941–80181136	10:80176156–80180940	10.82449812	–0.218718834	0.42387425	–0.515999343	0.605854868	0.99999
APLP2	chr11:130123612–130123779	11:130122514–130123611	490.449954	0.356584851	0.084290488	4.230428123	2.33E-05	0.00720
APPL2	chr12:105233091–105233405	Not Found						
ARMCX6	chrX:101617283–101617370	X:101616767–101617282 X:101616767–101617291	1.554115827 1.979014083	4.038853198 4.34277536	1.327731995 1.312975149	3.041919011 3.307583822	0.002350751 0.000941045	0.06748 0.04363
ATG5	chr6:106308364–106308491	6:106293107–106308363	172.1443391	–0.141439917	0.116742213	–1.211557606	0.225681771	0.99999
AXIN1	chr16:291190–291297	16:289608–291189	12.88615371	0.567337279	0.398949658	1.422077369	0.155003792	0.98622

gene	GRCh38	Splice Junction Found	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
BALAP2	chr17:81110914–81110959	17:81108510–81110913	4.125817311	1.37672654	0.718313789	1.916608815	0.055287647	0.6362
CCNB1IP1	chr14:20317794–20317974	Not Found						
CCT7	chr2:73235550–73235645	2:73234385–73235549	5.106947198	1.017231089	0.677438559	1.501584277	0.133204506	0.94952
CEP57	chr11:95795516–95795559	11:95790744–95795515	14.58205551	-0.09994828	0.374087616	-0.267178799	0.789331499	0.9999
CSF1	chr1:109926063–109926743	1:109925203–109926062	0.512925465	-2.527906075	1.798973282	-1.405193785	0.159963662	NA
<b>DLGAP4</b>	<b>chr20:36499242–36499321</b>	<b>20:36497067–36499241</b>	<b>2.321757283</b>	<b>4.629312735</b>	<b>1.234758303</b>	<b>3.7491651</b>	<b>0.000177424</b>	<b>0.01970</b>
EPN1	chr19:55677625–55677218	19:55677219–55677549	6.792614573	0.137794309	0.548684186	0.251135923	0.801709023	0.9999
EPN1	chr19:55677550–55677734	Not Found						
<b>ERGIC3</b>	<b>chr20:35554372–35554386</b>	<b>20:35548866–35554371</b>	<b>28.83410289</b>	<b>4.036660975</b>	<b>0.450049004</b>	<b>8.969380978</b>	<b>2.98E-19</b>	<b>1.3E-19</b>
		20:35549208–35554371	1.211236736	3.671748012	1.402133721	2.618686047	0.008826914	NA
FOXMI	chr12:2861299–2861412	12:2859664–2861298	10.02538278	1.393165926	0.478256758	2.913008346	0.00357965	0.08340
GGCT	chr7:30497741–30497896	Not Found						
GRAMD3	chr5:126466294–126466336	5:119456369–119471635	11.26191388	0.47467429	0.423301943	1.121360999	0.26213423	0.9999
HSD17B4	chr5:119471636–119471700	5:119457369–119471635	0.494108183	2.368999879	1.817387698	1.303519267	0.192397543	NA
<b>LARP7</b>	<b>chr4:112637644–112637787</b>	<b>4:112637240–112637643</b>	<b>27.19269642</b>	<b>3.451696147</b>	<b>0.399919423</b>	<b>8.630979006</b>	<b>6.08E-18</b>	<b>1.5E-18</b>
LARP7	chr4:112644425–112644488	4:112638674–112644424	0.728760441	2.934691979	1.612411811	1.820063558	0.068749326	NA
LRRC42	chr1:53947788–53947821	1:53946550–53947787	29.09604567	-0.664904577	0.265177179	-2.507397428	0.012162388	0.19036
<b>MADD</b>	<b>chr11:47292543–47292596</b>	<b>11:47290817–47292542</b>	<b>2.731362466</b>	<b>4.872013188</b>	<b>1.210216484</b>	<b>4.025736926</b>	<b>5.68E-05</b>	<b>0.01152</b>
MAN1B1	chr9:137088268–137088398	Not Found						
MRPL39	chr21:25587701–25587789	Not Found						



gene	GRCh38	Splice Junction Found	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
PCBP4	chr3:51962810–51962892	3:51961305–51962809	1.182311825	3.691518129	1.422245816	2.595555625	0.009443812	NA
		3:51962089–51962809	1.216301921	3.681908004	1.418185239	2.59621092	0.009425818	
PPHLN1	chr12:42351885–42352049	12:42335975–42351884	123.8298057	-0.046390515	0.135503904	-0.342355562	0.732083327	0.99999
PRKACB	chr1:84144269–84144548	1:111619935–111627469	2.568291852	1.215471674	0.941010031	1.291667074	0.196472465	0.99999
RAB23	chr6:57221319–57221459	Not Found						
RAP1A	chr1:111627470–111627563	Not Found						
RCC1	chr1:28518233–28518308	Not Found						
SREK1	chr5:66156065–66157740	5:66153597–66156064	33.02107777	-0.458074764	0.247829652	-1.848345265	0.064552411	0.69777
SREK1	chr5:66158809–66158932	5:66153597–66158808	10.54393267	0.145262673	0.426009401	0.340984665	0.733115126	0.99999
		5:66158231–66158808	7.253095127	-0.095361539	0.528821088	-0.180328548	0.856894645	0.99999
<b>STRN3</b>	<b>chr14:30929201–30929311</b>	<b>14:30913658–30929200</b>	<b>50.93805874</b>	<b>0.932636289</b>	<b>0.216274269</b>	<b>4.312285011</b>	<b>1.62E-05</b>	<b>0.00597</b>
STRN3	chr14:30918966–30919106	14:30919107–30929200	46.51202979	0.551658841	0.208079453	2.651193256	0.008020793	0.13040
		14:30913658–30918965	55.45310576	0.329108298	0.20344735	1.617658317	0.105736254	0.88134
TNRC6A	chr16:24758339–24758360	16:24750814–24758338	91.74639209	0.019177308	0.149159372	0.128569248	0.897698501	0.99999

See ref. <sup>8</sup>. Column identity from left to right; Column A, Gene ID of genes containing a significantly induced exon as identified by Palacino et al.<sup>8</sup>; Column B, GRCh38 position of the induced exon; Column C, the associated splice junction of the identified induced exons, if found in our dataset 'splice junction found'; Columns D–I, Differential expression status of each splice junction in our dataset, DMSO ( $n = 4$ ) versus 25 nM LMI070 ( $n = 4$ ). Differential expression was performed using DESeq2. Column D, The mean of normalized counts for all samples 'baseMean'; Column E, the log<sub>2</sub> fold change (maximum likelihood estimate) 'Log2FoldChange'; Column F, the associated standard error 'lfcSE'; Column G, the Wald statistic 'stat', Column H, the Wald test two-tailed  $P$  value 'p value'; Column I, the Benjamini-Hochberg (BH) multiple testing corrected  $P$  value "p adj". Genes observed to be significant ( $P_{adj} < 0.05$ ) in our dataset are bolded.

## Acknowledgements

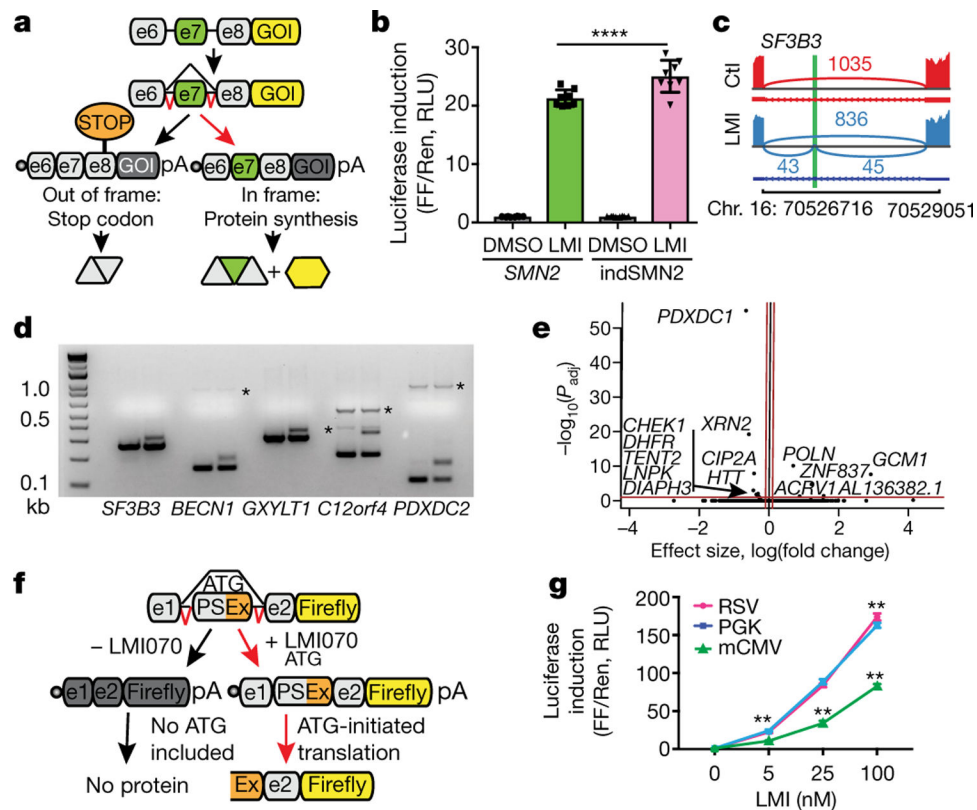
We thank E. Lysenko, Z. Hilfinger and J. Weber of the Davidson laboratory, as well as N. Hoque of the Novartis Institute for Biomedical Research for technical assistance. This work was supported in part by the Hereditary Disease Foundation (A.M.M. and P.T.R.), T32 HG009495 (National Institutes of Health) (P.T.R.), NBIR (R.S., D.L., A.M.M. and B.L.D.) and the Children's Hospital of Philadelphia Research Institute.

## References

1. Yen L et al. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* 431, 471–476 (2004). [PubMed: 15386015]

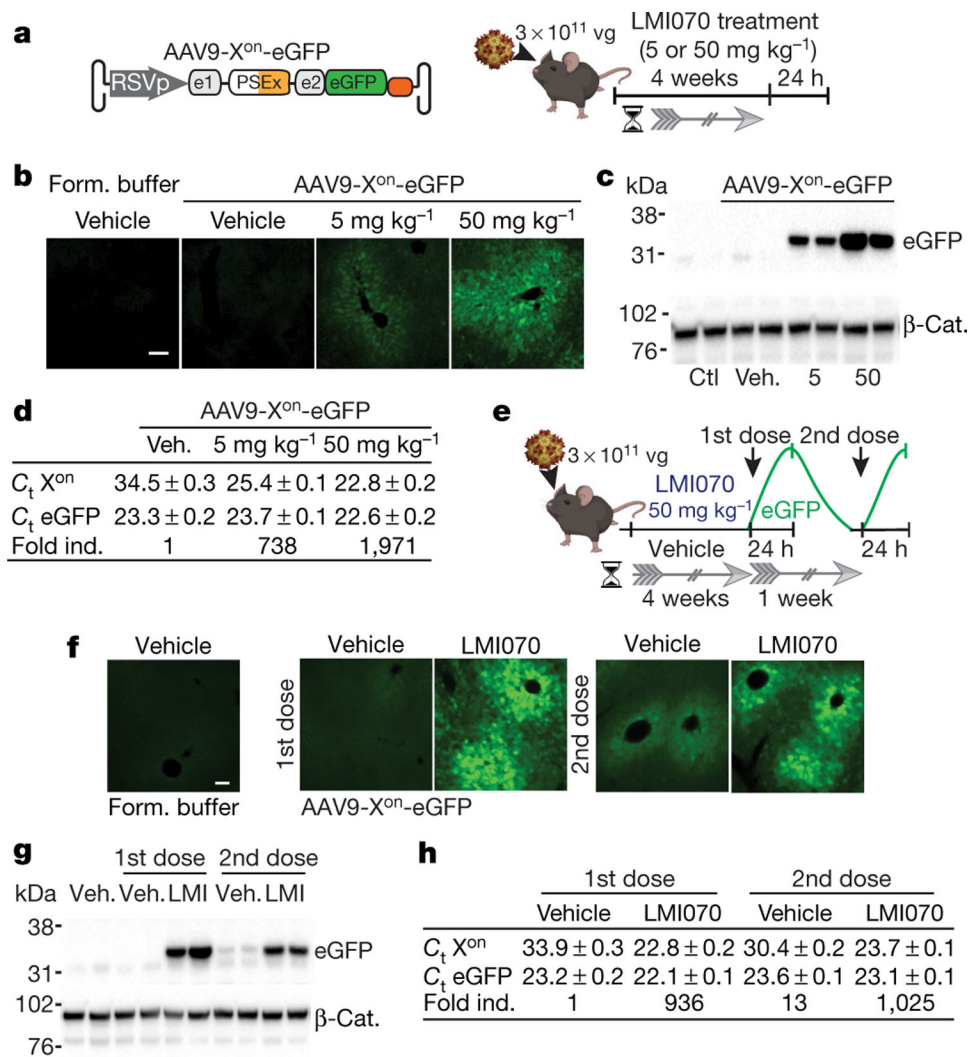
2. Dow LE et al. Inducible in vivo genome editing with CRISPR–Cas9. *Nat. Biotechnol.* 33, 390–394 (2015). [PubMed: 25690852]
3. Brown BD, Venneri MA, Zingale A, Sergi Sergi L & Naldini L Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer. *Nat. Med.* 12, 585–591 (2006). [PubMed: 16633348]
4. Domenger C & Grimm D Next-generation AAV vectors—do not judge a virus (only) by its cover. *Hum. Mol. Genet.* 28 (R1), R3–R14 (2019). [PubMed: 31261383]
5. Berget SM, Moore C & Sharp PA Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl Acad. Sci. USA* 74, 3171–3175 (1977). [PubMed: 269380]
6. Cartegni L & Krainer AR Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* 30, 377–384 (2002). [PubMed: 11925564]
7. Naryshkin NA et al. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* 345, 688–693 (2014). [PubMed: 25104390]
8. Palacino J et al. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* 11, 511–517 (2015). [PubMed: 26030728]
9. An Open Label Study Of LMI070 (Branaplam) in Type 1 Spinal Muscular Atrophy (SMA); <https://clinicaltrials.gov/ct2/show/NCT02268552>
10. FDA Approves Oral Treatment for Spinal Muscular Atrophy (7 August 2020); <https://www.fda.gov/news-events/press-announcements/fda-approves-oral-treatment-spinal-muscular-atrophy>
11. Wang J, Schultz PG & Johnson KA Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl Acad. Sci. USA* 115, E4604–E4612 (2018). [PubMed: 29712837]
12. Ingolia NT, Ghaemmaghami S, Newman JR & Weissman JS Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223 (2009). [PubMed: 19213877]
13. Eschbach JW, Kelly MR, Haley NR, Abels RI & Adamson JW Treatment of the anemia of progressive renal failure with recombinant human erythropoietin. *N. Engl. J. Med.* 321, 158–163 (1989). [PubMed: 2747747]
14. Chan KY et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* 20, 1172–1179 (2017). [PubMed: 28671695]
15. Amado DA et al. AAV-mediated progranulin delivery to a mouse model of progranulin deficiency causes T cell-mediated toxicity. *Mol. Ther.* 27, 465–478 (2019). [PubMed: 30559071]
16. Madisen L et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* 13, 133–140 (2010). [PubMed: 20023653]
17. Collins AL et al. Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* 13, 2679–2689 (2004). [PubMed: 15351775]
18. Van Alstyne M et al. Gain of toxic function by long-term AAV9-mediated SMN overexpression in the sensorimotor circuit. *Nat. Neurosci.* 24, 930–940 (2021). [PubMed: 33795885]
19. Weber EW et al. Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling. *Science* 372, eaba1786 (2021). [PubMed: 33795428]
20. Fu Y et al. High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826 (2013). [PubMed: 23792628]
21. Charlesworth CT et al. Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* 25, 249–254 (2019). [PubMed: 30692695]
22. Wagner DL et al. High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nat. Med.* 25, 242–248 (2019). [PubMed: 30374197]
23. Hanlon KS et al. High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat. Commun.* 10, 4439 (2019). [PubMed: 31570731]
24. Saito H et al. Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nat. Chem. Biol.* 6, 71–78 (2010). [PubMed: 20016495]
25. Zhong G et al. A reversible RNA on-switch that controls gene expression of AAV-delivered therapeutics in vivo. *Nat. Biotechnol.* 38, 169–175 (2020). [PubMed: 31873216]

26. Strobel B et al. A small-molecule-responsive riboswitch enables conditional induction of viral vector-mediated gene expression in mice. *ACS Synth. Biol.* 9, 1292–1305 (2020). [PubMed: 32427483]
27. Brown BD et al. Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nat. Biotechnol.* 25, 1457–1467 (2007). [PubMed: 18026085]
28. Winter GE et al. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* 348, 1376–1381 (2015). [PubMed: 25999370]
29. Buskirk AR, Ong YC, Gartner ZJ & Liu DR Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc. Natl Acad. Sci. USA* 101, 10505–10510 (2004). [PubMed: 15247421]
30. Roper J et al. In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. *Nat. Biotechnol.* 35, 569–576 (2017). [PubMed: 28459449]
31. O’Connell KE et al. Practical murine hematopathology: a comparative review and implications for research. *Comp. Med.* 65, 96–113 (2015). [PubMed: 25926395]
32. Dobin A & Gingeras TR Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics* 51, 11.14.1–11.14.19 (2015). [PubMed: 26334920]
33. Thorvaldsdóttir H, Robinson JT & Mesirov JP Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192 (2013). [PubMed: 22517427]
34. Robinson JT et al. Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26 (2011). [PubMed: 21221095]
35. Katz Y et al. Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics* 31, 2400–2402 (2015). [PubMed: 25617416]
36. Kent WJ et al. The human genome browser at UCSC. *Genome Res.* 12, 996–1006 (2002). [PubMed: 12045153]
37. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550 (2014). [PubMed: 25516281]
38. Nellore A et al. Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the Sequence Read Archive. *Genome Biol.* 17, 266 (2016). [PubMed: 28038678]
39. Kearse MG & Wilusz JE Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* 31, 1717–1731 (2017). [PubMed: 28982758]



**Fig. 1 | Generation and testing of X<sup>on</sup>.**

**a**, Schematic depicting the SMN2-on cassette. In the absence of exon 7 (e7), a premature stop codon blocks translation of the gene of interest (GOI). Inclusion of e7 permits translation. **b**, Change in firefly (FF) luciferase activity in LMI070-treated (100 nM) compared with DMSO-treated cells, normalized to *Renilla* (Ren) luciferase activity. RLU, relative light units. Data are mean  $\pm$  s.e.m. of 8 biological replicates, \*\*\*\* $P$  < 0.0001 SMN2 versus indSMN2, two-way ANOVA, Bonferroni's post-hoc test. **c**, *SF3B3* splicing without (red; control (Ctl)) or with (blue) 25 nM LMI070 treatment. The genomic location of the LMI070 spliced-in *SF3B3* exon (green bar) and intron counts are indicated. **d**, cDNAs for *SF3B3*, *BECN1*, *GXYLT1*, *C12orf4* and *PDXDC2* amplified from cells treated with DMSO or LMI070 (25 nM) (one representative plot from 4 RNA-seq samples). Novel splicing was confirmed by Sanger sequencing. The asterisks represent non-specific bands. **e**, Volcano plot illustrating differentially expressed genes between DMSO-treated and LMI070-treated cells. The red horizontal bar represents 0.05 significance on a  $-\log_{10}$  scale. The red vertical bars indicate thresholds of  $-0.1$  and  $0.1$  fold change. Genes meeting the significance and minimum fold-change thresholds are labelled (4 samples per group). **f**, Candidate minigenes for translation control; e1 and e2 are the flanking exons of the novel *SF3B3* exon (Extended Data Tables 1, 2). **g**, Induction of luciferase activity in cells transfected with the *SF3B3*-based X<sup>on</sup> cassette. Samples are normalized to *Renilla* luciferase and are relative to DMSO. Data are mean  $\pm$  s.e.m. of 8 biological replicates, \*\* $P$  < 0.01 mCMV versus RSV or PGK; \*\* $P$  < 0.01 RSV versus PGK, two-way ANOVA followed by Bonferroni's post hoc test.



**Fig. 2 |. Activity of X<sup>on</sup> in liver.**

**a**, Schematic of experiments using AAV9-X<sup>on</sup>-eGFP. Mice were injected intravenously with AAV9-X<sup>on</sup>-eGFP, after 4 weeks they were orally dosed once with LMI070, and tissues were collected 24 h later to assess mRNA (Extended Data Fig. 6) and protein expression. vg, viral genomes. **b**, Representative photomicrograph showing eGFP in the liver 24 h after dosing.  $n = 3$  mice per group. Scale bar, 100  $\mu$ m. **c**, Representative eGFP western blot with  $\beta$ -catenin ( $\beta$ -Cat.) as the loading control.  $n = 4$  mice per group. Form. buffer, vector formulation buffer; Veh., vehicle. **d**, Average (mean  $\pm$  s.e.m.) cycle threshold ( $C_t$ ) values for eGFP or LMI070-induced expression using TaqMan assays (Extended Data Fig. 6e). The fold change stated for the spliced product is relative to vehicle-treated, AAV9-X<sup>on</sup>-eGFP injected mice.  $n = 4$  mice per group. **e**, Schematic of the re-dosing experiment. Mice were injected intravenously and 4 weeks later were dosed with LMI070 or vehicle. LMI070-treated mice were subjected to drug washout for one week and then redosed with drug or vehicle ( $n = 5$  mice per group). Tissues were collected 24 h after each dose. **f**, Representative photomicrograph of the liver showing eGFP 24 h after each dose.  $n = 2$  mice per group. Scale bar, 100  $\mu$ m. **g**, Representative eGFP western blot 24 h after dosing, with  $\beta$ -catenin as

loading control.  $n = 3$  mice per group. **h**, Average (mean  $\pm$  s.e.m.)  $C_t$  values for eGFP and fold change relative to AAV9- $X^{on}$ -eGFP-injected mice treated with vehicle ( $n = 3$  mice per group).

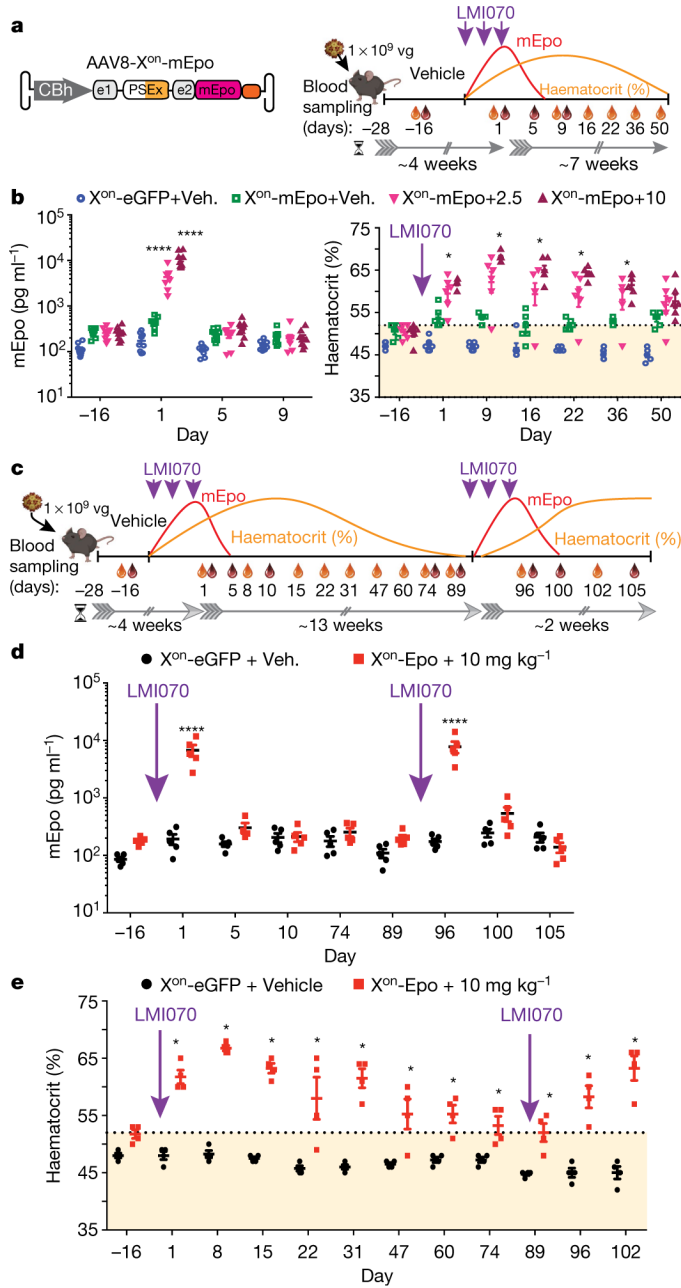
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

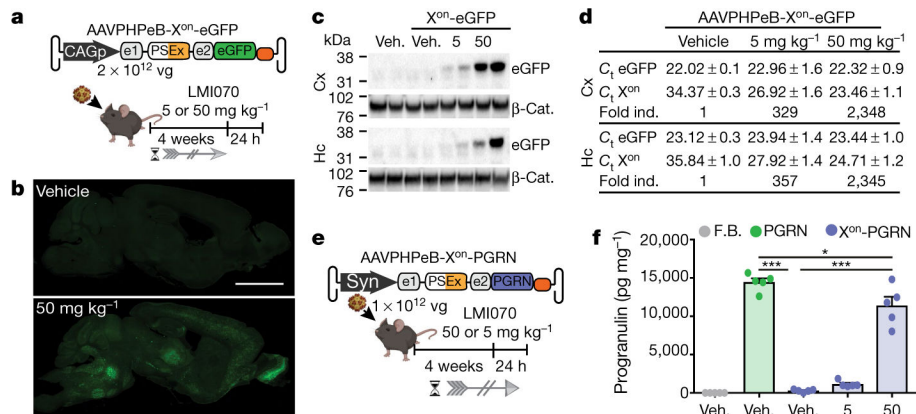




**Fig. 3 | Drug-induced regulation of Epo.**

**a**, Experimental design for mEpo induction and the effect on haematocrit levels. Three weeks after AAV or buffer delivery, mice were dosed with vehicle or LMI070 at 2.5 or 10 mg kg<sup>-1</sup> every other day for a total of 3 doses. Blood was collected for mEpo (red drop) and haematocrit (orange drop) assays as indicated. **b**, Left, mEpo levels 2 weeks after AAV injection, and 1, 5 and 9 days after the last dose. \*\*\*\**P* < 0.0001, LMI070-treated versus vehicle-treated mice; \*\*\*\**P* < 0.0001, 10 versus 2.5 mg kg<sup>-1</sup> LMI070-treated mice. Right, haematocrit levels two weeks after AAV injection, and for 50 days after dosing. The orange shaded area shows the haematocrit range of wild-type mice. \**P* < 0.05 mEpo injected, LMI070-treated mice versus vehicle. **c**, Schematic of the re-inducibility experiment. X<sup>on</sup>-

mEpo-AAVs were delivered intravenously and mice given vehicle or LMI070 orally every other day for a total of 3 doses. After haematocrit returned to baseline levels, LMI070 was re-administered and mice were monitored as before. **d**, mEpo levels from 2 weeks after AAV delivery. \*\*\*\* $P < 0.0001$  AAV-X<sup>on</sup>-mEpo-injected, LMI070-treated mice versus AAV-eGFP-injected mice. **e**, Haematocrit levels assessed over the time course of the experiment. \* $P < 0.05$  mEpo-injected mice treated with LMI070 versus eGFP-injected mice. Data are mean  $\pm$  s.e.m. of 9, 6 (**b**), 5 (**d**) or 4 (**e**) mice per group, significance was determined by two-way ANOVA followed by Bonferroni's post hoc test.



**Fig. 4 |. In vivo activity of X<sup>on</sup> in the brain.**

**a**, Schematic of the AAVPHPeB-X<sup>on</sup>-eGFP experiment. AAVs were injected intravenously, then after 4 weeks mice were given an oral dose of LMI070, and brains were collected 24 h later to assess splicing (Extended Data Fig. 7b), transcript levels and protein expression. **b**, Representative compiled photomicrographs (5 mice per group) from 40-μm-thick sagittal sections. Scale bar, 3 mm. **c**, Representative eGFP western blots of the cortex (Cx) and the hippocampus (Hc), with β-catenin as the loading control. *n* = 3 mice per group. **d**, Average (mean ± s.e.m.) C<sub>t</sub> values for eGFP and fold change relative to AAV-X<sup>on</sup>-eGFP-injected mice treated with vehicle. **e**, Schematic depicting the assessment of AAV-X<sup>on</sup>-PGRN for regulation of progranulin expression in the brain. Vectors were administered intravenously, then after 4 weeks mice were treated with a single oral dose of LMI070 or vehicle, and tissues were collected 24 h later. **f**, Progranulin protein levels in the cortex. Data are mean ± s.e.m. of 5 mice per group. \*\*\**P* < 0.001 for AAV-PGRN or AAV-X<sup>on</sup>-PGRN + 50 mg kg<sup>-1</sup> versus mice injected with AAV-X<sup>on</sup>-PGRN + vehicle, \**P* < 0.05 for AAV-X<sup>on</sup>-PGRN + 50 mg kg<sup>-1</sup> versus AAV-PGRN, one-way ANOVA followed by Bonferroni's post hoc test. F.B., vector formulation buffer.