nature genetics

ARTICLES

https://doi.org/10.1038/s41588-019-0555-z

RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons

James D. Thomas[1,2], Jacob T. Polaski[1,2,7], Qing Feng[1,2,7], Emma J. De Neef[1,2,3], Emma R. Hoppe[1,2,3], Maria V. McSharry[1,4], Joseph Pangallo[1,2], Austin M. Gabel[1,2,3,5], Andrea E. Belleville[5], Jacqueline Watson[6], Naomi T. Nkinsi[1,4], Alice H. Berger[1,3,4] and Robert K. Bradley[1,2,3]*

While RNA-seq has enabled comprehensive quantification of alternative splicing, no correspondingly high-throughput assay exists for functionally interrogating individual isoforms. We describe pgFARM (paired guide RNAs for alternative exon removal), a CRISPR–Cas9-based method to manipulate isoforms independent of gene inactivation. This approach enabled rapid suppression of exon recognition in polyclonal settings to identify functional roles for individual exons, such as an *SMNDC1* cassette exon that regulates pan-cancer intron retention. We generalized this method to a pooled screen to measure the functional relevance of 'poison' cassette exons, which disrupt their host genes' reading frames yet are frequently ultraconserved. Many poison exons were essential for the growth of both cultured cells and lung adenocarcinoma xenografts, while a subset had clinically relevant tumor-suppressor activity. The essentiality and cancer relevance of poison exons are likely to contribute to their unusually high conservation and contrast with the dispensability of other ultraconserved elements for viability.

Most biological processes are characterized by alternative splicing[1–3], which is correspondingly dysregulated in many diseases[4,5]. Mapping individual mis-spliced isoforms to specific molecular pathologies can enable the rational design of splicing-targeted therapeutics[6,7]. However, the vast majority of disease-associated RNA isoforms have not been functionally studied, hindering such therapeutic development.

This disparity between identification and functional characterization of isoforms arises from technological limitations. Antisense oligonucleotides are low throughput[8,9], while RNA-mediated interference does not alter alternative splicing. CRISPR–Cas9 has been used to knock out *DMD* isoforms or long non-coding RNAs by targeting splice sites[10,11], but has not been applied in a multiplexed fashion for studying alternative isoforms.

'Poison exons' provide a striking example of alternative splicing that is likely critical for organismal function, yet challenging to study. The human genome contains 481 'ultraconserved elements' that are perfectly conserved in the mouse and rat genomes[12]. Many ultraconserved and highly conserved elements overlap poison exons, defined as alternative exons that interrupt their host genes' reading frames[13,14] and trigger nonsense-mediated RNA decay (NMD)[15]. Although poison exons do not contribute to the protein-coding capacity of their host genes, a subset are known to play critical cellular roles. For example, poison exons within splicing factors can mediate gene expression autoregulation[13,14]. However, the vast majority of poison exons have not been functionally interrogated, and their hypothesized essentiality has never been tested.

Results

pgFARM enforces the production of exon exclusion isoforms. Simultaneously delivering two guide RNAs (paired guide RNA, or pgRNA) into cells can induce deletion of the intervening DNA sequence[16–19]. We therefore hypothesized that pgRNA delivery could manipulate isoform expression by deleting exons, splice sites and/or other *cis*-regulatory splicing elements. We termed this approach pgFARM (paired guide RNAs for alternative exon removal).

As a proof of principle, we designed pgRNAs that used distinct targeting strategies to remove a constitutive coding exon (exon two) of *HPRT1*, a non-essential gene whose inactivation permits resistance to 6-thioguanine (6TG; Fig. 1a). We cloned each pgRNA into the lentiGuide-Puro backbone[18] and introduced each construct into HeLa cells with doxycycline-inducible Cas9 (HeLa/iCas9[20]; Fig. 1b,c). pgRNA delivery induced rapid and effective skipping of *HPRT1* exon two (Fig. 1d).

We confirmed that exon skipping arose from on-target genomic DNA (gDNA) editing by sequencing individual *HPRT1* alleles. We detected pgRNA/Cas9-dependent edits at 91% of alleles. Complete gDNA excision was the most common editing event (40% of edited alleles), followed by diverse short insertions/deletions (indels; Fig. 1e, Extended Data Fig. 1a and Supplementary Table 1). Although pgRNAs can cause gDNA inversion in addition to excision[21], we detected no inversion events.

A recent study reported that Cas9-induced DNA breaks can result in rare large deletions[22], which could potentially cause unwanted gene disruptions. Although we did not observe any excision events > 350 base pairs (bp) by Sanger sequencing—far shorter than most

---

[1]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [2]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [3]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [4]Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [5]Medical Scientist Training Program, University of Washington, Seattle, WA, USA. [6]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [7]These authors contributed equally: Jacob T. Polaski, Qing Feng. *e-mail: rbradley@fredhutch.org

introns—this assay might not detect extremely large deletions. We therefore used long-range gDNA PCR to test whether pgRNA delivery caused large deletions. Consistent with the reported rarity of large deletions (3–7% of events[22]), we readily detected our positive control (deletion of ~600 bp) but no other large deletions (Fig. 1e). Large deletions therefore occur at sufficiently low rates to not significantly influence phenotypes in our polyclonal assays.

As gDNA excision disrupts gene structures, pgRNA delivery could potentially result in abnormal mis-splicing in addition to targeted exon skipping. We therefore used long-range PCR with reverse transcription (RT–PCR) to confirm that all pgRNAs caused skipping of the targeted *HPRT1* exon, but not production of unwanted additional isoforms (Extended Data Fig. 1b).

Inducing *HPRT1* exon skipping drove the expected 6TG resistance. Both HeLa/iCas9 and Cas9-expressing 293T cells treated with *HPRT1* exon two-targeting, but not non-targeting, pgRNAs formed 6TG-resistant outgrowths that exhibited *HPRT1* exon two skipping and loss of HPRT1 protein (Fig. 1f and Extended Data Fig. 1c–f). We confirmed pgFARM's generalizability by targeting another constitutively included exon. pgRNA delivery drove rapid skipping of *MET* exon 14 without inducing detectable cryptic splicing (Extended Data Fig. 1g,h).

We next used pgFARM to manipulate alternative splicing by targeting an *MBNL1* ultraconserved coding exon (exon five; Fig. 1g). We detected exon skipping 2 d after pgRNA delivery, with near-complete exon skipping for some pgRNAs after 7 d (Fig. 1h). Complete gDNA excision was the most common editing event (91%). We observed no unexpectedly large gDNA deletions, gDNA inversion or unwanted cryptic isoforms (Fig. 1i, Extended Data Fig. 2a,b and Supplementary Table 1). pgRNA delivery similarly induced *MBNL1* or *Mbnl1* exon skipping in Cas9-expressing untransformed human fibroblasts (IMR90), untransformed mouse melanocytes (Melan-a) and mouse melanoma cells (B16-F10), as well as on-target gDNA editing and splice site disruption (Extended Data Fig. 2c–e and Supplementary Table 1).

Induction of *MBNL1* exon skipping drove expected functional consequences. Nuclear levels of total MBNL1 were quantitatively lower following delivery of each pgRNA that induced appreciable exon skipping (Fig. 1j,k), as expected[12,23,24]. MBNL1 protein encoded by the exon five-containing messenger RNA was ablated in pgRNA-edited cell lines, while MBNL1 protein encoded by the exclusion isoform remained (Fig. 1l and Extended Data Fig. 2f). Induction of *MBNL1* exon five skipping caused quantitatively correlated differential splicing of *MBNL2*, whose own exon five is regulated by nuclear MBNL1[24,25] (Extended Data Fig. 2g). Together, these data demonstrate that pgFARM can suppress a specific RNA isoform independent of total gene disruption or induction of unwanted cryptic isoforms.

An *SMNDC1* poison exon regulates intron retention in cancer. We next used pgFARM to identify cellular roles for a highly conserved but less well-studied poison exon in *SMNDC1*, which is included at high levels in HeLa and lung adenocarcinoma (PC9) cells (Fig. 2a,b). As SMNDC1 is required for splicing catalysis in vitro[26], we hypothesized that its poison exon might influence the widespread intron retention that characterizes most cancers[27,28].

The *SMNDC1* poison exon enables splicing-dependent autoregulation via NMD in cell culture[29]. We therefore tested whether the same occurred in primary cancers profiled by The Cancer Genome Atlas (TCGA). Cancer samples exhibiting high *SMNDC1* poison exon inclusion relative to patient-matched peritumoral normal samples exhibited low *SMNDC1* gene expression, and vice versa (Fig. 2c and Extended Data Fig. 3a). *SMNDC1* poison exon inclusion was significantly dysregulated in cancer relative to patient-matched normal samples in 9 of the 14 cohorts with sufficient data for analysis, with reduced poison exon inclusion in most cancer types (Fig. 2d). Low *SMNDC1* poison exon inclusion and high gene expression were both associated with significantly poorer survival (Extended Data Fig. 3b,c).

We modeled cancer-associated *SMNDC1* poison exon skipping by delivering a pgRNA targeting the poison exon's 3′ splice site. We targeted the 3′ splice site to maximize the chance of exon skipping even if only one guide RNA (gRNA) induced cutting[30]. This strategy also allowed us to minimize the deleted region to reduce the chance of inadvertently affecting other functional elements. pgRNA delivery resulted in editing at 82% of sequenced *SMNDC1* alleles, with complete gDNA excision being the most common editing event (33%; Extended Data Fig. 4a and Supplementary Table 1). Almost all edited alleles exhibited dramatically reduced 3′ splice site strengths[31], even when only one cut occurred (Fig. 2e).

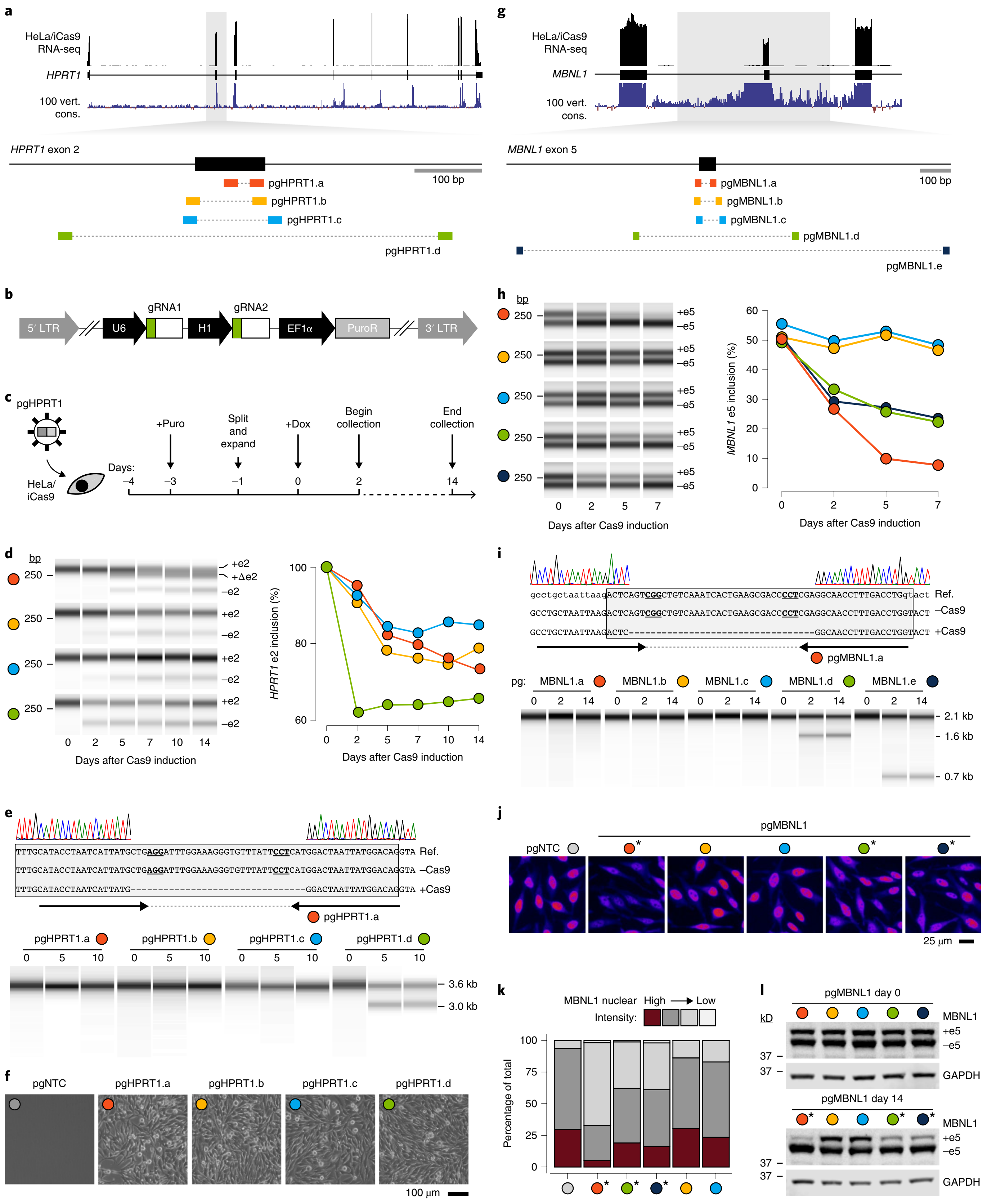
We next confirmed that individual editing events resulted in poison exon skipping. We generated Cas9-expressing PC9 lung adenocarcinoma cells (Extended Data Fig. 4b,c), delivered *SMNDC1*-targeting or control pgRNAs, and isolated monoclonal cell lines. Ninety percent of the *SMNDC1*-targeted clones carried 3′ splice site-disrupting edits (Extended Data Fig. 4d,e). We analyzed ten clones to find that all poison exon-targeted clones exhibited complete loss of *SMNDC1* poison exon inclusion, while no control clones did (Fig. 2f).

We functionally characterized the *SMNDC1* poison exon by delivering *SMNDC1*-targeting or control pgRNAs to HeLa/iCas9 cells and quantifying splicing with RNA sequencing (RNA-seq). *SMNDC1* poison exon-targeting pgRNA delivery eliminated poison exon inclusion without detectable induction of any cryptic splicing (Fig. 2g). Consistent with our hypothesis that *SMNDC1* regulates splicing efficiency, 221 genes exhibited significantly decreased intron retention following delivery of the poison exon-targeting pgRNA relative to an AAVS1-targeting control pgRNA, such as introns in *STK36* and *CENPT* (Fig. 2h).

We tested whether variable *SMNDC1* poison exon inclusion contributed to frequent intron retention in cancers[27,28,32]. We grouped the 512 lung adenocarcinoma samples with RNA-seq data[33] into terciles based on *SMNDC1* poison exon inclusion and quantified intron retention across each tercile[27]. Low *SMNDC1* poison exon inclusion was associated with notably widespread reductions in intron retention: 59% of constitutive introns exhibiting any retention

Fig. 1 | pgFARM facilitates rapid, programmable exon skipping. a, Top, RNA-seq read coverage and sequence conservation across *HPRT1* in HeLa/iCas9 cells. Bottom, pgRNAs targeting *HPRT1* exon two. **b**, A schematic of the pgRNA-expressing vector. **c**, A schematic of the pgRNA delivery strategy. Puro, puromycin; Dox, doxycycline. **d**, Left, RT–PCR analysis of *HPRT1* exon two (e2) inclusion. Right, RT–PCR quantification. **e**, Top, representative Sanger sequencing of pgFARM-edited *HPRT1* exon two (gray box). Bottom, PCR analysis of the *HPRT1* exon two genomic locus. pgHPRT1.a–c create gDNA excision events that are too small to resolve. **f**, A phase-contrast image of HeLa/iCas9 cells expressing a non-targeting control (pgNTC) or *HPRT1* exon two-targeting pgRNA after selection with 6-thioguanine. Representative images from $n = 3$ independent experiments are shown. **g**, As in (**a**), but for *MBNL1* exon five. **h**, As in (**d**), but for *MBNL1* exon five (e5) inclusion. **i**, As in (**e**), but for *MBNL1* exon five. **j**, Immunofluorescence images comparing nuclear MBNL1 abundance (orange, high intensity; blue, low intensity) in HeLa/iCas9 cells expressing non-targeting or *MBNL1* exon five-targeting pgRNAs. The asterisks mark pgRNAs that induced the greatest exon exclusion. **k**, Quantification of the data in (**j**). **l**, Western blot for MBNL1 and GAPDH from HeLa/iCas9 cells expressing the indicated pgRNAs before (top) and after (bottom) Cas9 induction. The colors are as in (**j**). Unless otherwise indicated, all data are representative results from $n = 2$ independent experiments. Uncropped gels are available as source data.

were spliced significantly more efficiently in samples with low poison exon inclusion (Fig. 2i,j). This signal persisted after restricting the analysis to cases where intron retention is not predicted to induce NMD (Extended Data Fig. 4f), and was equally strong but opposite on stratifying by *SMNDC1* gene expression (Fig. 2k and Extended Data Fig. 4g). We extended this analysis to find that almost all profiled cancer types exhibited significantly reduced intron retention in samples with low *SMNDC1* poison exon inclusion



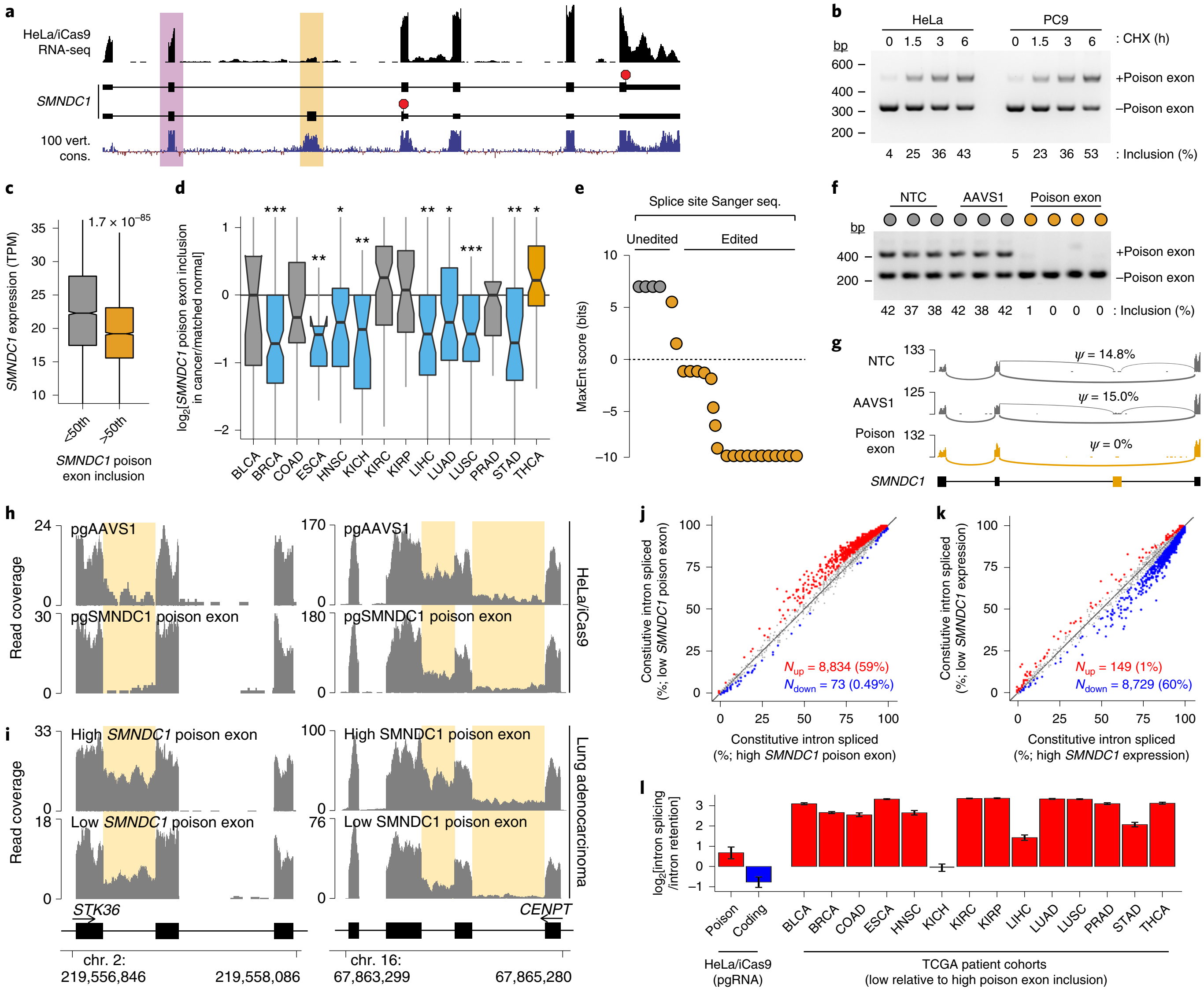


Fig. 2 | An *SMNDC1* poison exon modulates intron retention. a, pgRNAs were designed to disrupt inclusion of an *SMNDC1* constitutive coding (purple) or poison exon (yellow). **b**, *SMNDC1* poison exon inclusion following cycloheximide (CHX) treatment to inhibit NMD. $n = 4$ biologically independent time points. **c**, *SMNDC1* expression in cancers with *SMNDC1* poison exon inclusion greater (>50th) or lower (<50th) than the median. TPM, transcripts per million. The *P* value was computed with a two-sided Mann–Whitney *U* test. $n = 8{,}361$ cancers. **d**, Relative *SMNDC1* poison exon inclusion in cancers versus patient-matched peritumoral normal samples. The *P* values were computed with a two-sided Mann–Whitney *U* test. $^{*}P \leq 5 \times 10^{-2}$; $^{**}P \leq 5 \times 10^{-3}$; $^{***}P \leq 5 \times 10^{-5}$. **e**, MaxEnt[31] 3′ splice site scores for pgFARM-edited *SMNDC1* alleles. **f**, *SMNDC1* poison exon inclusion in cycloheximide-treated PC9-Cas9 clones expressing control (NTC, AAVS1) or *SMNDC1* poison exon-targeting pgRNAs. $n = 10$ biologically independent clones. **g**, RNA-seq coverage across the *SMNDC1* poison exon locus in HeLa/iCas9 cells treated with the indicated pgRNAs. $n = 1$ per pgRNA. ψ, poison exon inclusion. **h**, RNA-seq coverage across representative differentially retained introns in HeLa/iCas9 cells treated with the indicated pgRNAs. $n = 1$ per pgRNA. **i**, As in (**h**), but for lung adenocarcinoma samples with the highest or lowest *SMNDC1* poison exon inclusion. $n = 5$ per group. **j**, Constitutive intron splicing in lung adenocarcinomas with low (bottom tercile) or high (top tercile) *SMNDC1* poison exon inclusion. Red/blue, significantly increased/decreased splicing. **k**, As in (**j**), but samples stratified by *SMNDC1* expression. **l**, Constitutive intron splicing efficiency. Error bars, 5th/95th percentiles estimated by bootstrapping. The abbreviations, sample sizes and box plot elements are defined in the Methods. Uncropped gels are available as source data.

(Fig. 2l). Experimentally targeting the *SMNDC1* poison exon in HeLa/iCas9 cells similarly resulted in significantly decreased intron retention, while targeting the *SMNDC1* upstream exon resulted in significantly increased intron retention affecting 240 genes (Fig. 2l). These data suggest that the *SMNDC1* poison exon controls *SMNDC1* expression to modulate intron retention.

pgRNA library targeting highly conserved poison exons. We designed a pgRNA library targeting poison exons to perform a highly multiplexed screen (Fig. 3a). We identified 12,653 human poison exons that are predicted to induce NMD[15] and computed each exon's sequence conservation across 46 species[34], yielding 520 poison exons with high conservation at their 5′ and 3′ splice sites (Extended Data Fig. 5a–e). In contrast to frame-preserving cassette exons, highly conserved poison exons were uniquely enriched in genes encoding RNA-binding proteins (Fig. 3b,c and Extended Data Fig. 5f), in agreement with previous studies[13,14,35].

We selected 465 and 91 poison exons exhibiting high and low conservation to target with our library, with a preference for highly conserved poison exons given their presumed functional importance.

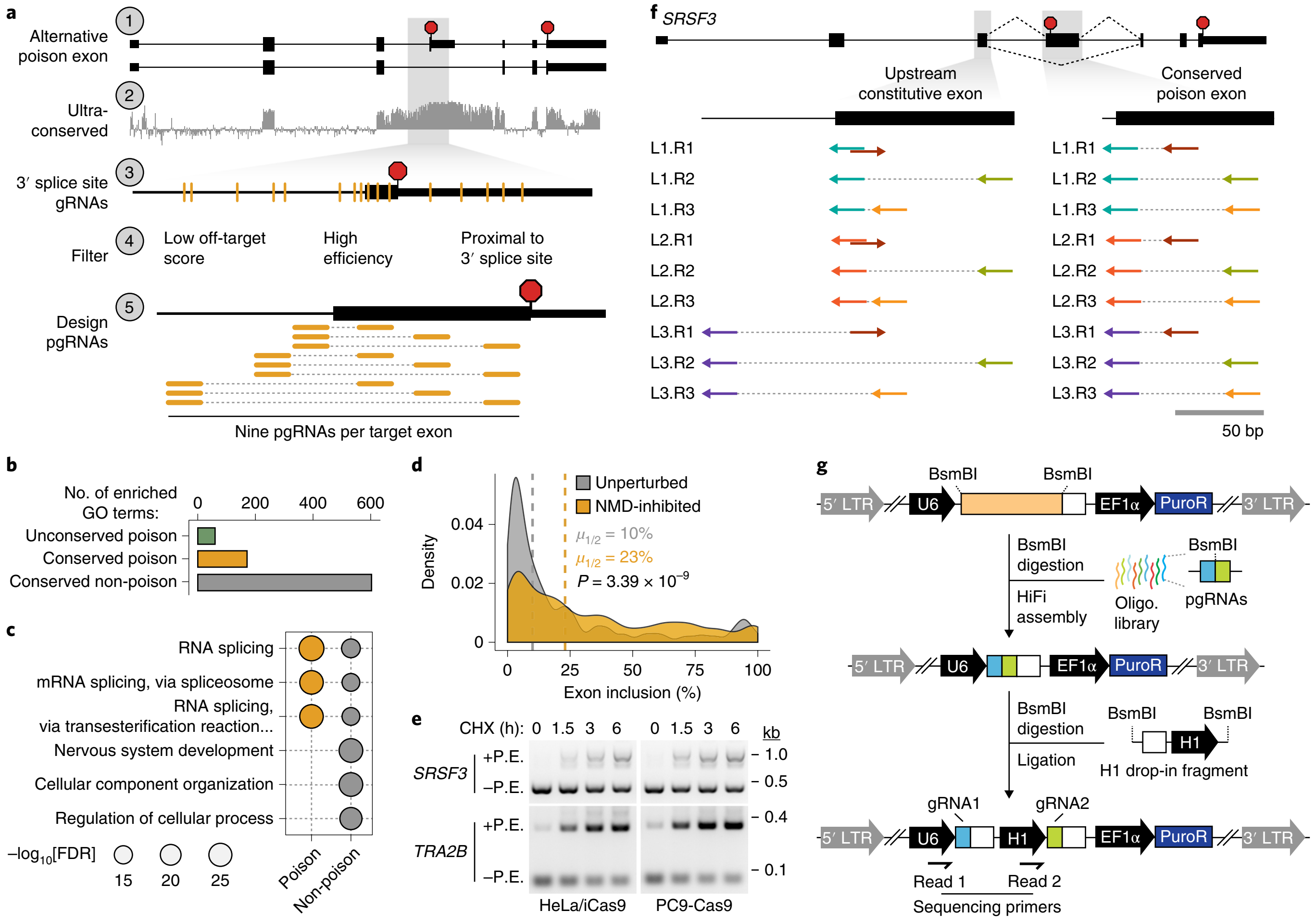


Fig. 3 | Design and construction of a poison exon loss-of-function library. **a**, A schematic of the selection criteria for poison exons targeted in this study as well as gRNA filtering criteria. **b**, A bar graph illustrating the numbers of significantly enriched (FDR ≤ 0.01) biological processes associated with the genes containing each of the indicated classes of alternative exons ($n = 2,363$, 352 and 888 for unconserved poison, conserved poison and conserved non-poison, respectively). Non-poison exons do not introduce premature termination codons. **c**, A bubble chart of the FDRs for the three most enriched biological processes that were associated with the sets of genes containing either highly conserved poison exons (left; $n = 352$) or highly conserved non-poison exons (right; $n = 888$). For (**b**) and (**c**), the FDRs were computed using the Wallenius method and corrected using the Benjamini–Hochberg method. **d**, A histogram illustrating the exon inclusion levels in unperturbed and NMD-inhibited HeLa cells[36] for the conserved poison exons ($n = 337$) targeted in our pgRNA library. The P value was computed by a two-sided Mann-Whitney U test. $\mu_{1/2}$, median. **e**, Inclusion of representative poison exons (P.E.) from (**d**) following NMD inhibition (Methods). A representative image from $n = 2$ independent experiments is shown. **f**, An illustration of the pgRNA targeting strategy for exemplary 3′ splice sites of an ultraconserved poison exon and the corresponding upstream constitutive exon in *SRSF3*. **g**, A schematic of the pgRNA library cloning strategy. Uncropped gels are available as source data.

We analyzed a published dataset[36] to find that the inclusion of those selected poison exons increased dramatically following *SMG6* and *SMG7* knockdown in HeLa cells, confirming that they induce NMD (Fig. 3d). Seventy-eight percent of targeted poison exons exhibited inclusion ≥5% in NMD-inhibited HeLa cells. We confirmed that representative poison exons were included at high levels and induced NMD in both HeLa/iCas9 and PC9-Cas9 cells (Fig. 3e).

We designed pgRNAs targeting the 3′ splice sites of each poison exon and the corresponding upstream constitutive coding exon (Fig. 3f). This design permitted us to compare the relative consequences of constitutive coding exon loss, which is typically equivalent to gene knockout, to poison exon loss. Our library targeted 556 poison and 407 upstream constitutive exons with an average of 9 pgRNAs per exon, and additionally included 1,000 non-targeting pgRNAs (Extended Data Fig. 5g–i and Supplementary Table 2).

We synthesized the pgRNA library with an oligonucleotide array and cloned the library at >1,000-fold coverage using a cloning strategy similar to those from previous pgRNA studies[17,18] (Fig. 3g). Sanger sequencing of individual bacterial colonies showed that ~98% of sequenced pgRNAs were properly paired after library construction, consistent with low (~7.5%) mis-pairing rates reported in other studies[17].

pgFARM enables isoform-resolution functional screens. We first performed a pilot cell viability screen in HeLa/iCas9 cells (Fig. 4a). We delivered the pgRNA library at a low multiplicity of infection of 0.2, collected gDNA 0, 8 and 14 d after Cas9 induction, and profiled pgRNA abundance by sequencing both gRNAs (Extended Data Fig. 6a). We sequenced each time point to ~400× coverage per pgRNA and computed the numbers of properly paired reads supporting each pgRNA. Non-targeting control pgRNAs were progressively enriched relative to targeting pgRNAs throughout the time course, as expected (Extended Data Fig. 6b).

We confirmed that the pgRNA library functioned in the context of a dropout screen with two metrics. First, we estimated gene expression in HeLa/iCas9 cells with RNA-seq to find that pgRNAs

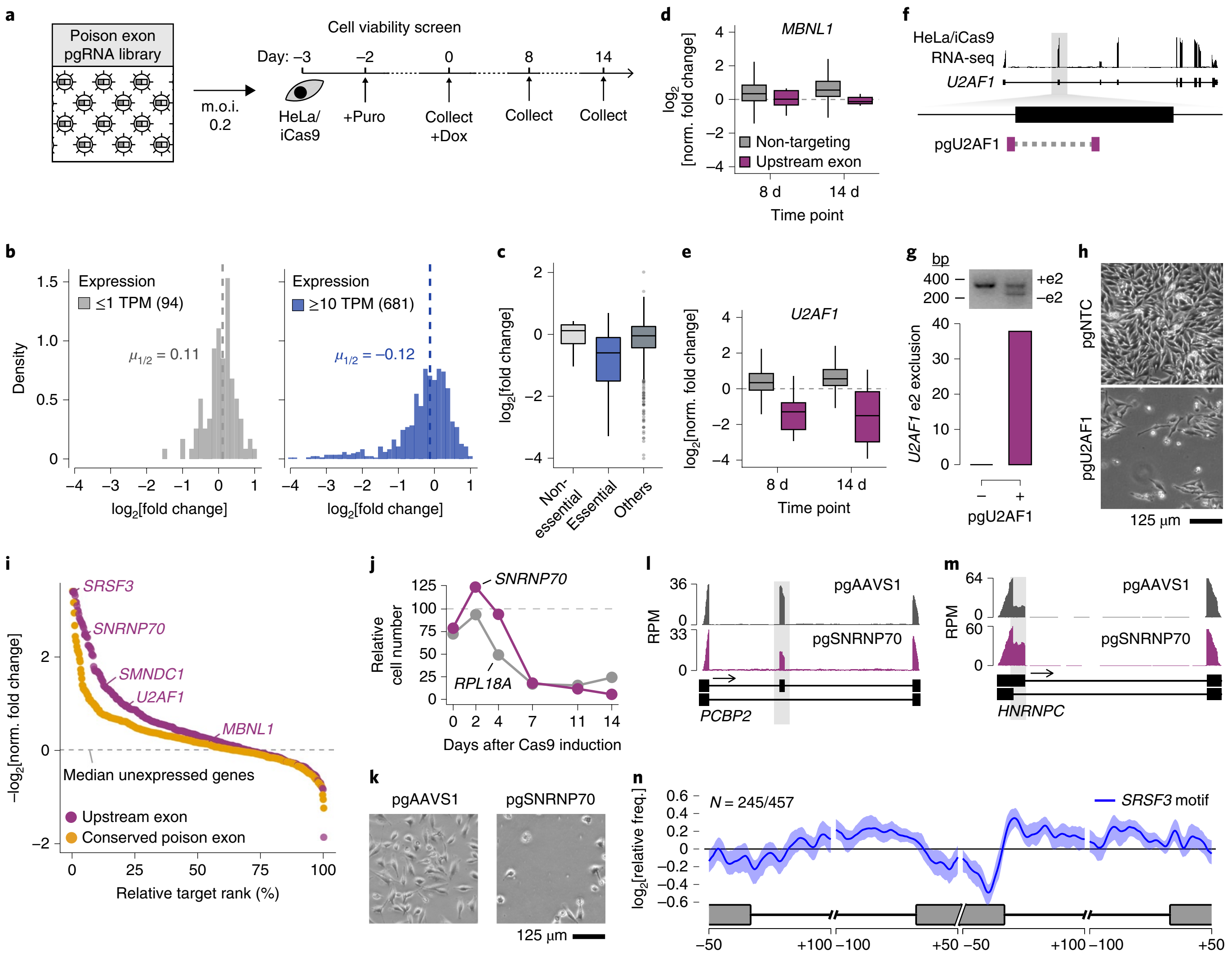


Fig. 4 | Unbiased detection of essential exons with pgFARM. a, A schematic of the dropout screen. m.o.i., multiplicity of infection. **b**, A histogram illustrating unnormalized fold changes associated with each targeted exon in unexpressed (left) or expressed (right) genes. $\mu_{1/2}$, median. **c**, Unnormalized fold changes associated with targeted exons in 'core essential' ($n = 51$), 'core non-essential' ($n = 12$)[37] or other genes ($n = 900$). **d**, Normalized fold changes for non-targeting (gray; $n = 1{,}000$) and *MBNL1* constitutive upstream exon-targeting (purple; $n = 9$) pgRNAs. **e**, As in (**d**), but for a *U2AF1* constitutive exon (purple; $n = 9$ pgRNAs). **f**, A schematic of *U2AF1* exon two-targeting pgRNA. **g**, *U2AF1* exon two exclusion in cells treated with pgRNA from (**f**). $n = 1$ independent experiment. **h**, Representative phase-contrast images of HeLa/iCas9 cells expressing the indicated pgRNAs. $n = 3$ independent experiments. **i**, A rank plot of normalized fold changes for conserved poison and upstream constitutive exons. *SRSF3*, *SNRNP70*, *SMNDC1* and *U2AF1* are essential genes; *MBNL1* is not. **j**, The viability of HeLa/iCas9 cells expressing the indicated pgRNAs relative to an AAVS1-targeting pgRNA. *RPL18A* is an essential gene. $n = 3$ biologically independent experiments. **k**, Representative phase-contrast images from (**j**). **l**, RNA-seq coverage illustrating differential cassette exon inclusion following treatment with an *SNRNP70* constitutive exon-targeting pgRNA. RPM, reads per million. **m**, As in (**l**), but illustrating differential 5′ splice site usage. **n**, A metagene plot illustrating relative *SRSF3* binding motif[44] occurrence in cassette exons exhibiting increased ($n = 245$) versus decreased ($n = 457$) inclusion following treatment with an *SRSF3* constitutive exon-targeting pgRNA. Exons exhibiting increased/decreased inclusion were depleted/enriched for the motif. Shading, 95% confidence interval. The box plot elements are defined in the Methods. Uncropped gels are available as source data.

targeting unexpressed and expressed genes were respectively enriched and depleted, as expected (Fig. 4b). Second, we confirmed that pgRNAs targeting a published set of 'core essential' genes[37] were depleted relative to pgRNAs targeting 'core non-essential' genes (Fig. 4c–e). We validated the on-target activity of a pgRNA targeting a constitutive exon within the essential gene *U2AF1* to find that it induced exon skipping and cell death (Fig. 4f–h), as well as differential requirements for the *SMNDC1* poison versus constitutive exons for cell growth (Extended Data Fig. 6c).

CRISPR–Cas9-induced DNA breaks can reduce cell fitness in a gene copy number-dependent manner[38–41]. We computed the copy number of each targeted unexpressed gene in the HeLa genome[42] and compared fold changes between different loci. While this analysis showed no correlation between the copy number and pgRNA depletion, we observed a modest depletion of exon-targeting pgRNAs relative to non-targeting pgRNAs (Extended Data Fig. 6d). We concluded that decreased cell viability caused by DNA breaks contributed to pgRNA depletion, although not in a copy number-dependent manner. We therefore normalized all fold changes relative to the median fold change for pgRNAs targeting unexpressed genes (Supplementary Table 3).

We next functionally validated additional constitutive exons that were identified as essential in our dropout screen. We ranked each exon according to the geometric mean of fold changes for all targeting pgRNAs (Fig. 4i and Supplementary Table 4) and selected a constitutive exon in *SNRNP70*, which encodes a core splicing factor[43], for detailed study. Treating cells with an *SNRNP70* constitutive exon-targeting pgRNA caused dramatic fitness defects that were rescued by overexpressing an SNRNP70-encoding cDNA (Fig. 4j,k and Extended Data Fig. 6e,f). We sequenced individual *SNRNP70* alleles 4 d after Cas9 induction to find that 79% of alleles exhibited 3′ splice site-disrupting edits, with ~40% exhibiting complete gDNA excision (Extended Data Fig. 6g and Supplementary Table 1).

We next performed RNA-seq to validate on-target exon skipping, which introduces a frameshift. Consistent with efficient NMD, we observed low levels of the exon exclusion isoform (versus none in control pgRNA-treated cells) with concomitant downregulation (greater than fourfold) of *SNRNP70* mRNA levels and reduced inclusion of *SNRNP70*'s poison exon (approximately fivefold; Extended Data Fig. 6h,i), consistent with the autoregulatory role of this poison exon[29]. We observed no RNA-seq reads indicative of unwanted cryptic isoforms.

We then tested the functional consequences of pgRNA-induced exon skipping. Consistent with SNRNP70's key role in 5′ splice site recognition[43], induction of *SNRNP70* constitutive exon skipping caused transcriptome-wide exon skipping and a shift towards intron-proximal 5′ splice site usage (Fig. 4l,m and Extended Data Fig. 6j,k). We extended these functional assays to *SRSF3*, which encodes a sequence-specific splicing factor[44]. We delivered a pgRNA targeting an *SRSF3* constitutive exon, confirmed on-target gDNA editing and performed RNA-seq (Fig. 4i, Extended Data Fig. 7a and Supplementary Table 1). pgRNA delivery caused *SRSF3* constitutive exon skipping and reduced inclusion of *SRSF3*'s poison exon (Extended Data Fig. 7b,c), consistent with its autoregulatory role[45]. Cassette exons that were repressed following *SRSF3*-targeting pgRNA delivery were enriched for SRSF3's RNA-binding motif (Fig. 4n and Extended Data Fig. 7d). In contrast to *SNRNP70* and *SRSF3* pgRNA-expressing cells, treatment with an AAVS1-targeting pgRNA resulted in little differential splicing relative to treatment with a non-targeting pgRNA (Extended Data Fig. 7e). No unwanted cryptic *SNRNP70* or *SRSF3* isoforms were detectable in any condition (Extended Data Fig. 7f,g). We conclude that pgFARM enables on-target induction of exon skipping in a high-content screen.

Many conserved poison exons are essential for cell growth. Having established the robustness of our method, we next tested the hypothesis that poison exons are important for viability. We performed a second dropout screen in HeLa/iCas9 and PC9-Cas9 cells with a re-cloned pgRNA library in biological quadruplicate (Extended Data Fig. 8a). Biological replicates segregated according to the day of collection and cell line following unsupervised hierarchical clustering (Fig. 5a). Per-pgRNA fold changes estimated for HeLa/iCas9 cells in our pilot and second screens had Pearson correlations of 0.88–0.93 (Extended Data Fig. 8b), highlighting our method's reproducibility. We therefore pooled data across biological replicates for subsequent analyses to maximize statistical power (Supplementary Table 5). pgRNAs targeting expressed versus unexpressed genes and essential versus non-essential genes were consistently depleted in both cell lines (Fig. 5b and Extended Data Fig. 8c).

As for our pilot screen, we normalized fold changes such that the median fold change for pgRNAs targeting unexpressed genes was equal to 1 for each cell line, replicate and time point. We computed a *P* value and empirical false discovery rate (FDR) for each exon by comparing the distribution of fold changes for all pgRNAs targeting that exon relative to the fold changes for all pgRNAs targeting unexpressed genes (Supplementary Tables 4 and 5). Gene copy number effects were not a confounding factor (Extended Data Fig. 8d).

We next tested whether poison exons are important for cell fitness. We enumerated exons that exhibited a significant depletion or enrichment (absolute fold change $\geq$ 25% with FDR $\leq$ 0.01 at day 14). Forty-three percent (169) and 10% (38) of targeted poison exons in expressed genes were depleted and enriched in HeLa/iCas9 cells, versus 58% (170) and 11% (32) of upstream constitutive exons—only a modest increase relative to poison exons. Poison exons that were frequently included in mRNA were preferentially depleted relative to exons that were typically excluded (Fig. 5c; $P = 0.004$). In PC9-Cas9 cells, 13% (51) and 6% (23) of targeted poison exons in expressed genes exhibited depletion and enrichment, versus 35% (101) and 5% (13) of upstream constitutive exons. Although constitutive Cas9 expression reduced the dynamic range of the PC9-Cas9 screen, skipping of both poison and upstream constitutive exons resulted in highly concordant fitness costs in the two cell lines (Fig. 5d and Extended Data Fig. 8e,f).

We validated our screens' estimates of cell viability by delivering individual pgRNAs targeting poison exons in *CPSF4* and *SMG1* and confirming that these exons are important for cell growth (Fig. 5e,f). We sequenced individual *CPSF4* and *SMG1* alleles to find that 96% of *CPSF4* alleles were subject to 3′ splice site-disrupting editing, including 58% with complete gDNA excision, while 75% of *SMG1* alleles contained indels that likely compromised exon recognition (Extended Data Fig. 9a and Supplementary Table 1). In neither case did targeting pgRNA delivery induce unwanted cryptic isoforms (Extended Data Fig. 9b,c).

Poison exon skipping leaves a gene's protein-coding capacity intact, while constitutive exon skipping typically does not. Nonetheless, pgRNA-induced skipping of many highly conserved and even some poorly conserved poison exons was associated with only modestly lower fitness costs than was loss of many constitutive exons (Fig. 5g and Extended Data Fig. 9d). These results support the intuitive, but untested, hypothesis that the high conservation of many poison exons is explained by purifying selection arising from those exons' contributions to cell fitness.

A subset of poison exons exhibit tumor-suppressor activity. We extended our approach to the context of lung adenocarcinoma xenografts to test two distinct hypotheses. First, we hypothesized that many poison exons would prove essential in vivo, just as in cell culture. Second, because of the difficulty of identifying positive selection in cultured transformed cells[46], we hypothesized that the stringency of growth in vivo might identify poison exons whose loss promoted tumor growth. We utilized PC9 cells, a common preclinical model of lung adenocarcinoma[47–49].

We transduced PC9-Cas9 cells with the poison exon pgRNA library using the same conditions as for our previous screens. After selection in cell culture for 4 d, we subcutaneously injected 3×10^7 cells (~3,000-fold pgRNA representation) into the flanks of immunocompromised (NU/J) mice (Fig. 6a and Supplementary Table 6). We observed similar growth rates for pgRNA library-transduced PC9-Cas9 xenografts and control parental PC9 (lacking Cas9) xenografts (Extended Data Fig. 10a,b). We collected gDNA from four and ten xenografts at early (~3 weeks) and late (~6 weeks) time points and measured the pgRNA abundance in the input plasmid pool, pre-injected cells, early tumors and late tumors with ~2,500-fold pgRNA coverage (Extended Data Fig. 10c).

All samples grouped according to biological condition and time of collection following unsupervised hierarchical clustering (Extended Data Fig. 10d). Late xenografts exhibited lower inter-tumor correlations than did early xenografts, consistent with prior reports[50]. We therefore used data from all replicates for statistical analyses to ensure that our results were robust with respect to high biological variability during tumorigenesis (Supplementary Table 5).

Few pgRNAs had no representation in early xenografts, while thousands were absent from late xenografts (Fig. 6b). Exon-targeting pgRNAs were preferentially lost relative to non-targeting pgRNAs. Therefore, almost all pgRNAs were compatible with engraftment, but negative selection led to subsequent loss of many exon-targeting pgRNAs.

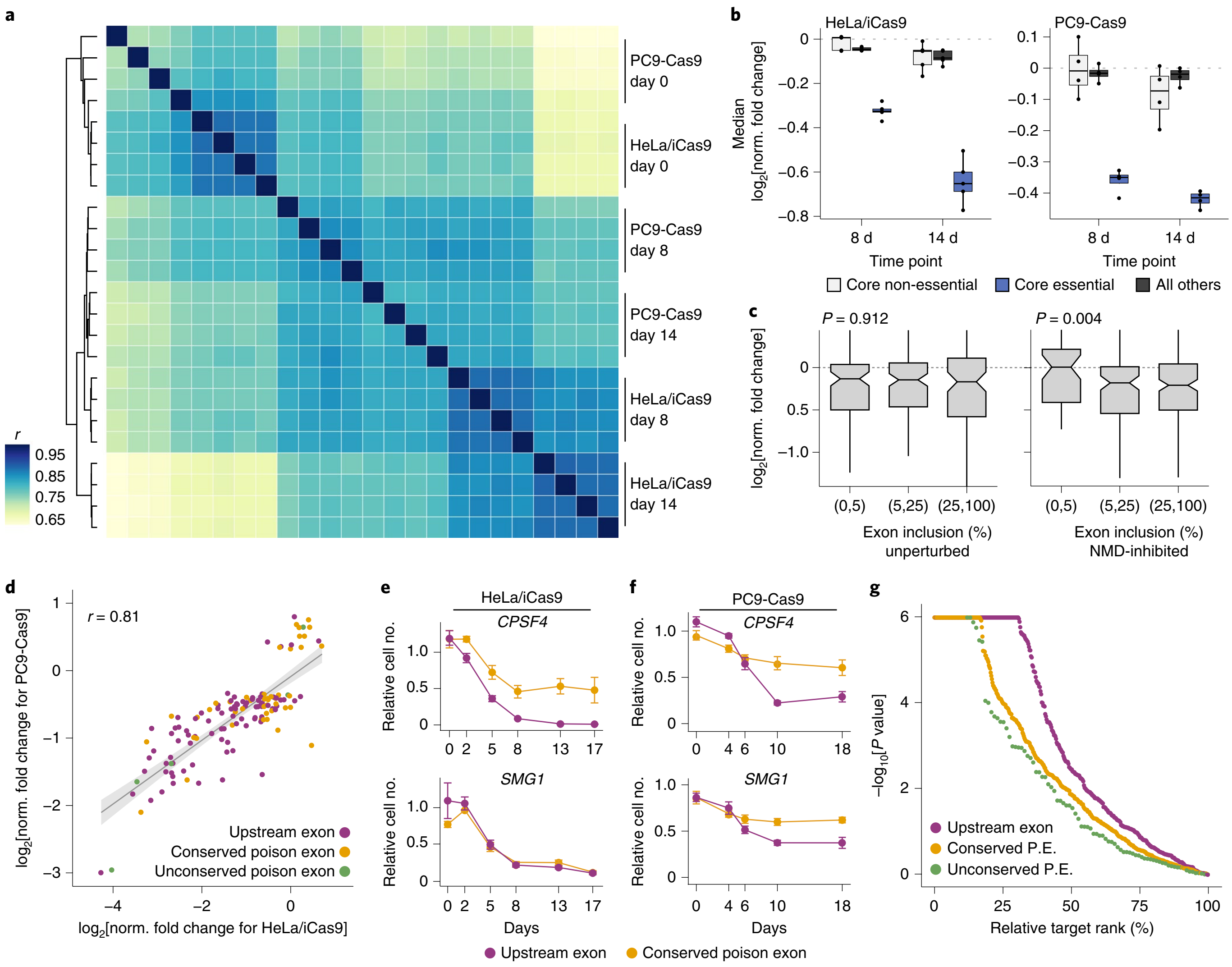


Fig. 5 | Many conserved poison exons are essential for cell fitness. a, A heat map illustrating Pearson correlations between raw counts supporting each pgRNA for all samples. Dendrogram, unsupervised clustering of raw counts by the complete-linkage method. $n = 9{,}508$ pgRNAs per sample. **b**, Normalized fold changes for targeted exons within 'core essential', 'core non-essential'[37] or all other genes. Each point illustrates the median over targeted exons within the indicated gene sets for a single screen replicate of the screen. $n = 5$ and 4 screens for HeLa/iCas9 and PC9-Cas9 cells. **c**, Normalized fold changes for targeted poison exons, stratified according to their inclusion in unperturbed or NMD-inhibited HeLa cells[36]. NMD inhibition decouples splicing and transcript degradation. The *P* values were computed by a two-sided Mann–Whitney *U* test. $n = 154/91/31$ (left) and 44/103/129 (right). **d**, A scatter plot comparing normalized fold changes for exons in HeLa/iCas9 versus PC9-Cas9 cells. As a result of the reduced dynamic range of the PC9 screen, the plot is restricted to exons with an absolute log fold change ≥ 1.25 and FDR ≤ 0.01 in PC9 cells and within genes with expression ≥ 10 TPM in both cell lines. *r*, Pearson correlation. $n = 86$, 46 and 5 for upstream, conserved poison and unconserved poison exons, respectively. **e**, The relative proliferation of HeLa/iCas9 cells treated with the indicated pgRNAs relative to cells treated with control (non-essential gene *CSPG4*-targeting) pgRNAs. The data are presented as mean ± s.d. $n = 3$ biologically independent experiments. **f**, As in (**e**), but for PC9-Cas9 cells. **g**, A rank plot of the *P* values for each targeted exon in the HeLa/iCas9 screen. The box plot elements are defined in the Methods.

We quantified exon essentiality by computing fold changes in pgRNA abundance in each tumor versus pre-injected cells and normalized the data as described above. One hundred and twelve upstream constitutive and 77 poison exons were significantly depleted in late xenografts. Consistent with our results, parent genes of these 112 constitutive exons were all previously reported as essential for lung cancer xenograft growth[50]. Most upstream constitutive and poison exons that exhibited significant depletion in the late xenografts were also depleted in our PC9-based cell culture screens, although a subset exhibited divergent behavior (Fig. 6c).

Although many poison exons are essential for cell growth, we hypothesized that a subset might have anti-tumorigenic effects. Splicing factors are frequently overexpressed in cancers[51], although pro-tumorigenic roles have been demonstrated only for a few factors[52–54]. We therefore tested whether modulating exon inclusion within genes encoding splicing factors influenced tumorigenesis. Skipping of constitutive exons within SR and hnRNP genes, many of which are essential[37,50], was strongly selected against (Fig. 6d and Extended Data Fig. 10e). In contrast, most targeted poison exons within SR and hnRNP genes exhibited enrichment in late xenografts (Fig. 6d). These data suggest that many RNA splicing factors are proto-oncoproteins whose pro-tumorigenic effects are constrained by poison exons.

The anti-tumorigenic effects of poison exons extend beyond splicing factors, with 61 poison exons enriched in late xenografts. Poison exon loss was more frequently associated with pro- relative

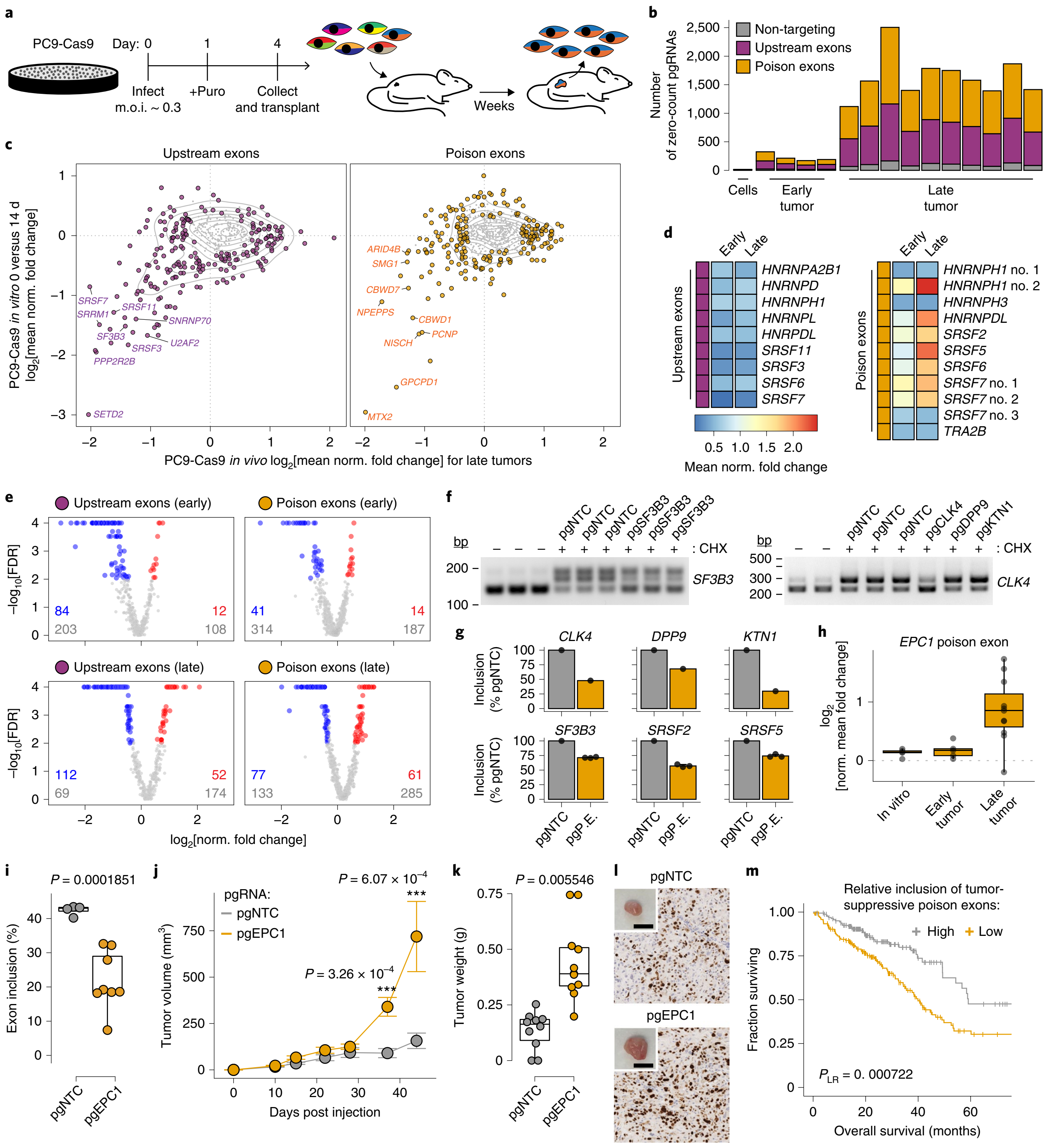


Fig. 6 | pgFARM uncovers modifiers of in vivo tumorigenesis. a, A schematic of the screens. **b**, The numbers of pgRNAs with zero counts. **c**, Normalized fold changes for exons measured in vivo and in vitro. **d**, Normalized fold changes for exons in SR and hnRNP genes. *HNRNPH1* and *SRSF7* contain multiple poison exons; *SRSF7* has a poison exon with competing 3′ splice sites. **e**, The numbers of significantly depleted (blue) and enriched (red) targets. **f**, *SF3B3* (left) or *CLK4* (right) poison exon inclusion in PC9-Cas9 cells expressing the indicated pgRNAs. **g**, Poison exon inclusion in the indicated genes in PC9-Cas9 cells expressing the indicated pgRNAs. pgP.E., pgRNA targeting the indicated poison exon. The data are presented as mean ± s.d. **h**, Normalized fold changes for the *EPC1* poison exon. **i**, *EPC1* poison exon inclusion in PC9-Cas9 clones expressing the indicated pgRNAs. The *P* value was computed with a two-sided Student's *t*-test. **j**, Tumor volumes for xenografts established from PC9-Cas9 cells expressing the indicated pgRNAs ($n = 10$ per group). The data are presented as mean ± s.e.m. The *P* values were computed with a two-sided Mann–Whitney *U* test. **k**, The tumor weights at the endpoint. The *P* value was computed with a two-sided Student's *t*-test. **l**, Representative Ki-67 immunohistochemistry images ($n = 17$ total histological analyses; for dissected tumor images; scale bar, 1 cm). **m**, The survival of lung adenocarcinoma patients stratified by inclusion of tumor-suppressive poison exons. The *P* values were computed with a two-sided logrank test. The sample sizes and box plot elements are defined in the Methods.

to anti-tumorigenic effects compared to constitutive exon loss ($P = 0.017$ by the one-sided binomial proportion test; Fig. 6e) We confirmed that enrichment was due to on-target activity by validating poison exon skipping for several pgRNAs (Fig. 6f,g).

We selected a poison exon within *EPC1* for further study due to its notable enrichment, previous reports of tumorigenic roles for *EPC1*[55,56] and inclusion at high rates (>40%) in NMD-inhibited cells (Fig. 6h,i). We confirmed on-target induction of exon skipping following pgRNA delivery in monoclonal cell lines (Fig. 6i) as well as a modest fitness advantage in cell culture (Extended Data Fig. 10f). We therefore extended these studies to in vivo tumorigenesis. Tumors derived from engraftment of polyclonal *EPC1* poison exon-targeted PC9-Cas9 cells were significantly larger and exhibited increased Ki-67 staining relative to control tumors (Fig. 6j–l).

We next tested whether poison exons with tumor-suppressor capacity in xenografts were clinically relevant. We stratified lung adenocarcinoma patients[33] according to their inclusion of essential (depleted) and tumor-suppressive (enriched) poison exons. Low inclusion of tumor-suppressive poison exons was associated with significantly worse progression-free and overall survival relative to high inclusion (Extended Data Fig. 10g,h; $P = 0.012$ and 0.0187). Further restricting our analysis to tumor-suppressive poison exons that exhibited high splicing variability across tumors yielded even more significant effects (Fig. 6m; $P = 0.013$ and 0.00072). Inclusion of essential poison exons was associated with no significant survival difference (Extended Data Fig. 10i,j), as expected. We conclude that many poison exons act as clinically relevant tumor suppressors.

Discussion

The ongoing discovery of new DNA- and RNA-targeting CRISPR–Cas systems will enable the development of diverse tool kits for manipulating isoform expression. Single gRNA delivery[10,57] and base editing[58,59] can alter exon recognition, while RNA-targeting CRISPR–Cas systems can enable direct manipulation of alternative splicing[60,61]. Each of these techniques is potentially amenable to a screening format.

As a result of their extraordinary sequence conservation, ultraconserved elements were initially assumed to be essential for life[12]. However, deletion of many ultraconserved enhancers has no effects on mouse organismal or cell viability[62–65]. Although poison exons are similar to enhancers with respect to their gene regulatory activities, we found that many poison exons exert robust effects on cell viability. Most unexpectedly, some poison exons have clinically relevant tumor-suppressive effects.

We focused on cassette exons to address the outstanding mystery of poison exons' high conservation. However, pgFARM can potentially be applied to many other kinds of alternative RNA processing[66–68]. We expect pgFARM to enable rapid and unbiased functional interrogation of specific RNA isoforms associated with diverse biological processes or disease states.

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/s41588-019-0555-z.

Received: 6 December 2018; Accepted: 21 November 2019;
Published online: 07 January 2020

References

1. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
2. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
3. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
4. Dvinge, H., Kim, E., Abdel-Wahab, O. & Bradley, R. K. RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer* **16**, 413–430 (2016).
5. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).
6. Stein, C. A. & Castanotto, D. FDA-approved oligonucleotide therapies in 2017. *Mol. Ther.* **25**, 1069–1075 (2017).
7. Inoue, D. et al. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* **574**, 432–436 (2019).
8. Cartegni, L. & Krainer, A. R. Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nat. Struct. Biol.* **10**, 120–125 (2003).
9. Taylor, J. K., Zhang, Q. Q., Wyatt, J. R. & Dean, N. M. Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides. *Nat. Biotechnol.* **17**, 1097–1100 (1999).
10. Long, C. et al. Correction of diverse muscular dystrophy mutations in human engineered heart muscle by single-site genome editing. *Sci. Adv.* **4**, eaap9004 (2018).
11. Liu, Y. et al. Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat. Biotechnol.* **36**, 1203–1210 (2018).
12. Bejerano, G. et al. Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
13. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926–929 (2007).
14. Ni, J. Z. et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
15. Kurosaki, T., Popp, M. W. & Maquat, L. E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* **20**, 406–420 (2019).
16. Zheng, Q. et al. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *Biotechniques* **57**, 115–124 (2014).
17. Zhu, S. et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nat. Biotechnol.* **34**, 1279–1286 (2016).
18. Gasperini, M. et al. CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *Am. J. Hum. Genet.* **101**, 192–205 (2017).
19. Diao, Y. et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* **14**, 629–635 (2017).
20. Cao, J. et al. An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Res.* **44**, e149 (2016).
21. Li, Y. et al. A versatile reporter system for CRISPR-mediated chromosomal rearrangements. *Genome Biol.* **16**, 111 (2015).
22. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
23. Lin, X. et al. Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **15**, 2087–2097 (2006).
24. Kino, Y. et al. Nuclear localization of MBNL1: splicing-mediated autoregulation and repression of repeat-derived aberrant proteins. *Hum. Mol. Genet.* **24**, 740–756 (2015).
25. Charizanis, K. et al. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75**, 437–450 (2012).
26. Rappsilber, J., Ajuh, P., Lamond, A. I. & Mann, M. SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome. *J. Biol. Chem.* **276**, 31142–31150 (2001).
27. Dvinge, H. & Bradley, R. K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **7**, 45 (2015).
28. Jung, H. et al. Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* **47**, 1242–1248 (2015).
29. Saltzman, A. L. et al. Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol. Cell Biol.* **28**, 4320–4330 (2008).
30. Amoasii, L. et al. Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. *Sci. Transl. Med.* **9**, eaan8081 (2017).
31. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).

32. Sowalsky, A. G. et al. Whole transcriptome sequencing reveals extensive unspliced mRNA in metastatic castration-resistant prostate cancer. *Mol. Cancer Res.* **13**, 98–106 (2015).
33. The Cancer Genome Atlas Research Network Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
34. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
35. Yan, Q. et al. Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. *Proc. Natl Acad. Sci. USA* **112**, 3445–3450 (2015).
36. Colombo, M., Karousis, E. D., Bourquin, J., Bruggmann, R. & Muhlemann, O. Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA* **23**, 189–201 (2017).
37. Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
38. Aguirre, A. J. et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* **6**, 914–929 (2016).
39. Munoz, D. M. et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* **6**, 900–913 (2016).
40. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
41. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
42. Adey, A. et al. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* **500**, 207–211 (2013).
43. Kohtz, J. D. et al. Protein–protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature* **368**, 119–124 (1994).
44. Anko, M. L. et al. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* **13**, R17 (2012).
45. Jumaa, H. & Nielsen, P. J. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.* **16**, 5077–5085 (1997).
46. Doench, J. G. Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* **19**, 67–80 (2018).
47. Sharma, S. V. et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
48. Shah, K. N. et al. Aurora kinase A drives the evolution of resistance to third-generation EGFR inhibitors in lung cancer. *Nat. Med.* **25**, 111–118 (2019).
49. Chmielecki, J. et al. Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. *Sci. Transl. Med.* **3**, 90ra59 (2011).
50. Chen, S. et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
51. Urbanski, L. M., Leclair, N. & Anczukow, O. Alternative-splicing defects in cancer: splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *WIREs RNA* **9**, e1476 (2018).
52. Karni, R. et al. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* **14**, 185–193 (2007).
53. Anczukow, O. et al. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.* **19**, 220–228 (2012).
54. Golan-Gerstl, R. et al. Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res.* **71**, 4464–4472 (2011).
55. Huang, X. et al. Enhancers of Polycomb EPC1 and EPC2 sustain the oncogenic potential of MLL leukemia stem cells. *Leukemia* **28**, 1081–1091 (2014).
56. Wang, Y. et al. Epigenetic factor EPC1 is a master regulator of DNA damage response by interacting with E2F1 to silence death and activate metastasis-related gene signatures. *Nucleic Acids Res.* **44**, 117–133 (2016).
57. Mou, H. et al. CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol.* **18**, 108 (2017).
58. Yuan, J. et al. Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *Mol. Cell* **72**, 380–394.e7 (2018).
59. Gapinske, M. et al. CRISPR-SKIP: programmable gene splicing with single base editors. *Genome Biol.* **19**, 107 (2018).
60. Konermann, S. et al. Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* **173**, 665–676.e14 (2018).
61. Jillette, N. & Cheng, A. W. CRISPR artificial splicing factors. Preprint at *bioRxiv* https://doi.org/10.1101/431064 (2018).
62. Ahituv, N. et al. Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* **5**, e234 (2007).
63. Nolte, M. J. et al. Functional analysis of limb transcriptional enhancers in the mouse. *Evol. Dev.* **16**, 207–223 (2014).
64. Dickel, D. E. et al. Ultraconserved enhancers are required for normal development. *Cell* **172**, 491–499.e15 (2018).
65. Schneider, A., Hiller, M. & Buchholz, F. Large-scale dissection suggests that ultraconserved elements are dispensable for mouse embryonic stem cell survival and fitness. Preprint at *bioRxiv* https://doi.org/10.1101/683565 (2019).
66. Alsafadi, S. et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.* **7**, 10615 (2016).
67. Mayr, C. & Bartel, D. P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673–684 (2009).
68. Pineda, J. M. B. & Bradley, R. K. Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes Dev.* **32**, 577–591 (2018).

Methods

pgRNA design, plasmids and cloning. For pgRNA optimization (Fig. 1), candidate gRNAs located near the targeted exon were identified and then paired on the basis of being located within the coding sequence or proximal/distal to splice sites. Both NAG and NGG protospacer adjacent motifs were utilized. pgRNAs were cloned following published methods[18] (Fig. 3g). Oligonucleotides containing both pgRNA spacer sequences were synthesized as DNA ultramers, amplified (primers RKB1169 and RKB1170; Supplementary Table 7) using NEBNext High Fidelity 2× Ready Mix (New England Biolabs) and purified with a 1.8× Ampure XP SPRI bead (Beckman Coulter) clean-up. This insert was cloned into BsmBI (FastDigestEsp3I, Thermo Fisher Scientific)-linearized lentiGuide-Puro (Addgene no. 52963) backbone using the NEBuilder HiFi (New England Biolabs) assembly system and transformed into NEB Stable competent *Escherichia coli* cells (New England Biolabs) to generate the pLGP-2×Spacer vector. Propagated plasmid was purified using the ZymoPURE Plasmid MiniPrep Kit (Zymogen) and linearized with BsmBI. An H1 drop-in gBlock (Integrated DNA Technologies) containing the second Pol III promoter and gRNA backbone was digested with BsmBI, purified using a 1.8× SPRI bead clean-up and ligated into the linearized pLGP-2×Spacer backbone using NEB Quick Ligase (New England Biolabs). This reaction was transformed into NEB Stable cells to propagate the plasmid and generate final pLGP-pgRNA vectors. All plasmids were sequence verified using Sanger sequencing (RKB1148 primer). pgRNAs used for validation studies are listed in Supplementary Table 8.

Cas9-expressing cell generation. PC9-Cas9 cells were generated by transducing PC9 cells (M. Meyerson) with pXPR_111 lentivirus and selecting with blasticidin for 5–7 d. Cas9 protein was detected with an anti-Cas9 antibody (Cell Signaling no. 14697) and anti-ACTB antibody (Cell Signaling no. 4970). Cas9-expressing B16-F10 (ATCC CRL-6475), Melan-a (D. Bennett) and HEK293T cells were generated by transducing cells with lentiCas9-Blast (Addgene 52962) lentivirus followed by blasticidin selection.

Cell culture. HeLa/iCas9 and Cas9-expressing HEK293T, IMR90 and B16-F10 cells were grown at 37 °C and 5% atmospheric CO_2 in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin–streptomycin (Gibco). The same conditions were used for PC9-Cas9 and Cas9-expressing Melan-a cells except that Roswell Park Memorial Institute 1640 medium was used instead of DMEM. Cas9-expressing Melan-A cell media was supplemented with 200 nM TPA (Sigma-Aldrich). All cell lines were periodically tested for mycoplasma contamination. For 6TG resistance assays, we treated cells with 15 μM 6-thioguanine (Sigma-Aldrich) for one week.

Lentivirus production and titration. For large-scale production, HEK293T cells were seeded in T225 flasks such that each flask would be ~80% confluent at the time of transfection. After overnight incubation, pCMV-VSV-G (Addgene no. 8454), psPAX2 (Addgene no. 12260) and pLGP-pgRNA transfer vectors were introduced into cells using PEI Max (Polysciences) transfection. Lentivirus-containing medium was collected 48 h later, filtered and stored as 1 ml aliquots at −80 °C until use. For small-scale production, HEK293T cells were seeded into individual wells of a six-well plate and all reagents were proportionally scaled. To determine lentiviral titers, HeLa/iCas9 or PC9-Cas9 cells were seeded in individual wells of a 12-well plate in medium supplemented with 8 μg ml^{-1} polybrene (EMD Millipore) and incubated at 37 °C for 2 h. Next, serial dilution of the lentivirus preparation was added to individual wells and incubated for 24 h at 37 °C. The next day, cells from individual wells of the 12-well plate were re-seeded into 8 wells of a 96-well plate. Cells in four of these wells were grown in culture medium supplemented with 1 μg ml^{-1} puromycin and the other four contained no puromycin. After all cells in the no-infection control wells were dead (typically 2–3 d), cell viability was quantified using a CellTiter-Glo (Promega) assay according to the manufacturer's instructions. Multiplicity of infection was determined by calculating the ratio of cells in the puromycin-treated compared to no puromycin treatment groups.

pgRNA vector delivery and sample collection. For testing individual pgRNA constructs, HeLa/iCas9 or PC9-Cas9 cells were seeded into individual wells of a multi-well plate and treated with viral supernatant to deliver pgRNA vectors. The next day, virus-containing medium was exchanged for standard growth medium supplemented with 1 μg ml^{-1} puromycin to select for stable integration. After selection, 1 μg ml^{-1} of doxycycline was added to HeLa/Cas9 cells to induce Cas9 expression. This was defined as day 0 for each experiment. As the PC9-Cas9 cells constitutively express Cas9, day 0 was defined as the time when all cells in a no-infection control plate died after puromycin selection. Cells in all treatment groups were passaged for 2–3 weeks. During this time, cell confluency and morphology were routinely analyzed using a Cytation 5 Imaging Reader (BioTek), cell number was measured using a CellTiter-Glo assay, and aliquots of cells were collected for molecular assays.

gDNA PCR, TOPO cloning and Sanger sequencing. gDNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol. Regions of interest were amplified by PCR using gene-specific primers (Supplementary Table 7) and analyzed using a 4200 TapeStation System (Agilent Genomics). For TOPO cloning and Sanger sequencing, purified amplicons were ligated into vectors for sequencing using the Zero Blunt TOPO PCR Cloning Kit (Thermo Fisher Scientific) following the manufacturer's protocol. Ligation reactions were transformed into One Shot TOP10 Chemically Competent *E. coli* (Thermo Fisher Scientific) using the manufacturer's protocol, plated onto LB agar supplemented with 50 μg ml^{-1} kanamycin and grown overnight at 37 °C. Sequences corresponding to each region of interest were generated by Direct Colony Sanger Sequencing (GENEWIZ). Sequence alignments were performed using MAFFT[69].

RT–PCR. Total RNA was extracted using the Direct-zol RNA MiniPrep (Zymo Research). cDNA was synthesized using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) following the manufacturer's protocol. RT–PCR was performed using gene-specific primers (Supplementary Table 7) using Q5 High-Fidelity DNA Polymerase (New England Biolabs) and amplicons were analyzed and quantified using either a 4200 TapeStation System (Agilent Genomics) or agarose gel electrophoresis followed by quantification of band intensity using FIJI/ImageJ. To detect poison exon-containing RNA isoforms, cells were treated with 50 μg ml^{-1} cycloheximide for up to 6 h to inhibit NMD.

Immunofluorescence. Cells grown on glass coverslips were washed with PBS, followed by fixation in 10% phosphate-buffered formalin (Fisher Scientific) for 10 min at room temperature and permeabilization with PBST (PBS, 0.2% Triton X-100) for 10 min at room temperature. Non-specific binding was blocked by incubating cells in PBS + 1% BSA (Fisher Scientific) for 1 h at room temperature followed by overnight incubation with primary antibody (Mb1a DSHB, 1:1,000) for 1 h at room temperature. Cells were washed three times with PBST for 10 min at room temperature and then incubated with secondary antibodies (goat anti-mouse DyLight 594, Thermo Fisher Scientific) for 1 h at room temperature. Cells were then washed three times with PBST for 10 min at room temperature and mounted with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Labs). Images were captured using an Aperio ScanScope FL (Leica Biosystems) and quantified using the HALO image analysis software (Indica Labs).

Immunohistochemistry. Xenograft tissue processing, embedding and staining was performed by the Fred Hutchinson Experimental Histopathology core. Human Ki-67 was detected using a mouse monoclonal antibody (Dako MIB-1). To mitigate background staining, mouse-on-mouse blocking was performed as previously described[70]. Staining was performed using a BOND RX autostainer (Leica Biosystems) and images were acquired using an Aperio ImageScope (Leica Biosystems).

Western blotting. Total protein lysates were prepared in 1× RIPA buffer (Cell Signaling) and quantified using the Pierce 660 nm Protein Assay Reagent. Total protein lysates were electrophoretically separated and transferred to nitrocellulose membranes using the NuPAGE system (Thermo Fisher Scientific). Membranes were blocked with Odyssey Blocking Buffer (LI-COR Biosciences) for 1 h at room temperature followed by overnight incubation at 4 °C with primary antibodies diluted in blocking buffer. HPRT1 (Abcam ab10479, 1:1,000) and GAPDH (Bethyl a300-639a, 1:5,000) were used as primary antibodies. IRDye (LI-COR Biosciences) secondary antibodies were used for detection and imaged using the Odyssey CLx Imager (LI-COR Biosciences).

pgRNA library design and construction. Poison exons were identified using transcript annotations from MISO v2.0[71] and pgRNAs targeting the 3′ splice sites of poison exons were designed using the methodology described in Fig. 3. The library cloning method followed previously published strategies[17,18] and was similar to cloning individual pgRNA vectors except for two adaptations. First, pgRNA oligonucleotides were synthesized using a DNA oligonucleotide array (Twist Bioscience) and used as input for the first PCR step. Second, for each step, multiple molecular reactions and bacterial transformations were performed such that each pgRNA was maintained at >1,000-fold coverage to prevent bottlenecking of the library diversity. Sanger sequencing of individual bacterial colonies was used to confirm proper gRNA pairing throughout the cloning procedure. The pgRNA library is available to the academic community (https://www.addgene.org/Robert_Bradley).

Cell viability screens. HeLa/iCas9 or PC9-Cas9 cells were seeded in 15-cm plates at a density of 5×10^6 cells per plate in complete medium supplemented with 8 μg ml^{-1} polybrene. A volume of the pgRNA library virus was added such that only 20–30% of cells were predicted to survive after selection with puromycin. The medium was changed 24 h later and replaced with complete medium supplemented with 1 μg ml^{-1} puromycin. After no cells remained in uninfected control plates, we collected the day 0 cell pellets and then added 1 μg ml^{-1} doxycycline to HeLa/iCas9 cells. At this point, cells were passaged every 2 to 3 d at a sufficient seeding density to maintain library diversity and cell pellets were collected on days 8 and 14 for gDNA extraction.

pgRNA deep sequencing library preparation and sequencing. Cell pellets were digested in lysis buffer (50 mM Tris, 50 mM EDTA, 1% SDS, 100 μg ml^{-1} proteinase K) overnight at 55 °C and gDNA was isolated using isopropanol precipitation. To build sequencing libraries, three PCR steps were performed as outlined in Extended Data Fig. 6a. First, 1 μg gDNA was used as input for amplification with NEBNext High Fidelity 2× Ready Mix using primers RKB2713/RKB2714 followed by Ampure XP SPRI bead clean-up. Second, 10 ng of amplicon from PCR no. 1 was used as input for amplification with primers RKB2715/RKB2716 followed by Ampure XP SPRI bead clean-up. Third, 10 ng of amplicon from PCR no. 2 was used as input for amplification with a common forward primer, RKB2717, and a sample-specific barcoding primer to accommodate multiplexing. For each PCR, multiple reactions were performed for each sample to maintain >1,000-fold coverage of each pgRNA in the library. Finally, purified libraries were combined in equimolar proportions and sequenced using an Illumina sequencer.

Animal use. All animal procedures were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals and approved by the Institutional Animal Care and Use Committees at Fred Hutchinson Cancer Research Center. NU/J (stock no. 002019) mice were obtained from the Jackson Laboratory.

Xenograft screen. PC9-Cas9 cells were grown in multiple 15-cm plates and treated with pgRNA lentiviral libraries at a multiplicity of infection of ~0.3. Infected cells were propagated in cell culture for ~4 d to select (1 μg ml^{-1} puromycin) stable cell lines and grow enough cells for transplantation. For injections, adult NU/J mice were anesthetized with isoflurane and 3×10^7 cells were injected subcutaneously into both flanks. Cohorts of mice were euthanized ~3 and ~6 weeks post injection, corresponding to the early and late time points, respectively (Supplementary Table 6), and tumors were dissected and stored at −80 °C. For gDNA extraction, 100 mg of tissue from each tumor was digested in lysis buffer (50 mM Tris, 50 mM EDTA, 1% SDS, 100 μg ml^{-1} proteinase K) overnight at 55 °C and gDNA was isolated using isopropanol precipitation. pgRNA libraries were constructed using the same methods as for the in vitro screens.

Validation xenograft studies. PC9-Cas9 cells were grown using standard conditions, transduced with lentivirus containing pgRNA expression vectors, and selected with 1 μg ml^{-1} puromycin. Before implantation, cells were grown for at least one week post-selection. For injections, adult NU/J mice were anesthetized with isoflurane and 2×10^6 cells were subcutaneously injected into both flanks. Tumor dimensions were measured using calipers throughout the time course. For histology, dissected tumors were fixed in 10% formalin solution at room temperature for 3 d before processing and paraffin embedding.

pgRNA deep sequencing data analysis. The first and second reads were separately mapped to a database of pgRNA sequences using Bowtie[72]. Correct pairings, for which both the first and second reads mapped to a given pgRNA, were kept; incorrect pairings were discarded. If a given first and second read had more than one possible correct pairing, then all correct pairings were kept but the degenerate pairings were down-weighted by 1/the number of possible pairings when counts of reads supporting each pgRNA were computed. A per-pgRNA pseudocount was computed as follows. For each pgRNA, 'reference' and 'comparison' pseudocounts were computed as max (5, 0.05 × (counts in the reference time point)) and max (5, (reference pseudocount) × (total counts for all pgRNAs in the comparison sample/total counts for all pgRNAs in the reference sample)). The reference and comparison pseudocounts were added to the actual counts for the reference and comparison time points when computing fold changes for each pgRNA. This procedure regularized fold-change computations in a manner proportional to the relative representation of each pgRNA within the library.

Fold changes were then normalized to account for the effects of DNA damage as described in the main text. The median fold change for all pgRNAs targeting unexpressed genes was computed for each time point relative to day 0 and each fold change was then divided by this number. After applying this normalization procedure, the median fold change for pgRNAs targeting unexpressed genes for a given cell type was equal to 1.

Statistical analyses of normalized fold changes were performed as follows at a per-target level. For a given targeted exon at a given time point, a *P* value for differential enrichment relative to day 0 was computed by performing a two-sided Mann–Whitney test between the fold changes for all pgRNAs targeting that exon relative to the fold changes for all pgRNAs targeting unexpressed genes. FDRs were computed by estimating a distribution of *P* values associated with the above procedure for fake targets derived by subsampling groups of nine pgRNAs from all pgRNAs targeting unexpressed genes. A *P* value was computed for each group. We performed this procedure 10,000 times to estimate an empirical distribution of *P* values derived from fake targets and then estimated FDRs for real targets via the cumulative distribution function of the fake *P*-value distribution. Unless otherwise specified, normalized fold changes associated with a given target exon were computed as the geometric mean over all targeting pgRNAs. These statistical procedures ensured that fold changes < 1 corresponded to decreased viability due to on-target effects, independent of DNA breaks, and permitted us to assess the statistical significance of depletion or enrichment of each targeted exon.

All statistical analyses were performed in the R programming environment with Bioconductor[73]. All plots and figures were generated with the dplyr[74] and ggplot2[75] packages.

RNA-seq library preparation. RNA was extracted from cell pellets using the Direct-zol RNA MiniPrep (Zymo Research) kit. Poly(A)-selected, unstranded Illumina libraries were prepared using the TruSeq protocol per the manufacturer's instructions. Libraries were analyzed using a 4200 TapeStation System to confirm proper size distribution before sequencing on an Illumina HiSeq. Libraries were sequenced as 2 × 50 bp to obtain ~40 million reads per sample.

RNA-seq data analysis. RNA-seq data were analyzed as previously described[76]. Briefly, reads were mapped to a transcriptome annotation created by merging the Ensembl 71[77], UCSC knownGene[78] and MISO v2.0[71] annotations using RSEM version 1.2.4[79] (modified to call Bowtie[72] with option '-v 2'). Unaligned reads were mapped to the genome (hg19/GRCh37 assembly) and a database consisting of all possible pairings between 5′ and 3′ splice sites for a given gene present in our merged transcriptome annotation with TopHat version 2.0.8b[80]. Mapped reads were merged and used as input to MISO v2.0. For TCGA studies, we analyzed the 5,718 available samples from the 14 cancer types with at least 10 patient-matched cancer and normal samples.

Survival analyses. Survival analyses and corresponding statistical tests were performed with the Kaplan–Meier estimator and logrank test (R package survival[81]). Patients were stratified as follows for Fig. 6m. For each cancer sample, we computed the following statistic: (number of tumor-suppressive poison exons for which exon inclusion ≤ 25th percentile of exon inclusion over the entire cohort)/(number of tumor-suppressive poison exons for which exon inclusion ≥ 75th percentile of exon inclusion over the entire cohort). The statistic was computed using the set of tumor-suppressive poison exons with defined exon inclusion for ≥90% of patients and high splicing variability (median exon inclusion level ≥ 10% with a standard deviation of inclusion across patients ≥ 25% of the median inclusion). Sixteen depleted and sixteen enriched poison exons met those criteria. Patients were stratified identically for Extended Data Fig. 10g–j using the sets of essential or tumor-suppressive poison exons described in the main text (as for Fig. 6m, but without filtering based on splicing variability, yielding a total of 62 depleted and 47 enriched poison exons).

Statistics and reproducibility. For Fig. 2d, sample sizes are *n* = 19; 111; 38; 12; 40; 25; 71; 30; 46; 57; 50; 52; 30; 59 (left to right). For Fig. 2l, sample sizes are *n* = 105/121; 326/484; 112/210; 54/66; 14/26; 17/22; 136/201; 68/104; 87/142; 132/237; 120/179; 135/171; 9/14; 88/151 (left to right, formatted as low/high terciles). Cancer type abbreviations follow TCGA standards (https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations). For Fig. 6e, sample sizes are *n* = 4/10 (top/bottom) biologically independent experiments. For Fig. 6f, sample sizes are *n* = 3 (pgNTC/pgSF3B3) and 1 (pgCLK4/pgDPP9/pgKTN1) technically independent experiments. For Fig. 6g, sample sizes are *n* = 1 (*CLK4*/*DPP9*/*KTN1*) and 3 (*SF3B3*/*SRSF2*/*SRSF5*) technically independent experiments. For Fig. 6h, sample sizes are *n* = 4 (in vitro/early tumor) and 10 (late tumor) biologically independent experiments. For Fig. 6i, sample sizes are *n* = 4 (pgNTC) and 8 (pgEPC1) biologically independent clones. For Fig. 6j,k, sample sizes are *n* = 10 tumors per group. For Fig. 6l, sample sizes are *n* = 17 histological analyses. For Fig. 6m, sample sizes are *n* = 171/170 samples for low/high categories.

For all box plots, the middle line, hinges, notches and whiskers indicate the median, 25th/75th percentiles, 95% confidence interval and most extreme data points within 1.5× the interquartile range from the hinge.

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

Data availability

RNA-seq data generated as part of this study have been deposited in the Gene Expression Omnibus (accession number GSE120703). RNA-seq data generated by TCGA were downloaded from the Cancer Genomics Hub (CGHub) and Genomic Data Commons (GDC). Other data that support this study's findings are available from the authors upon reasonable request. Source data for Figs. 1–4 and Extended Data Figs. 1, 2, 4, 6 and 10 are presented with the paper.

References

69. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
70. Goodpaster, T. & Randolph-Habecker, J. A flexible mouse-on-mouse immunohistochemical staining technique adaptable to biotin-free reagents, immunofluorescence, and multiple antibody staining. *J. Histochem. Cytochem.* **62**, 197–204 (2014).
71. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).

72. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
73. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
74. Wickham, H., François, R., Henry, L. & Müller, K. dplyr: A grammar of data manipulation. R package version 0.7.6. (2018).
75. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
76. Dvinge, H. et al. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc. Natl Acad. Sci. USA* **111**, 16802–16807 (2014).
77. Flicek, P. et al. Ensembl 2013. *Nucleic Acids Res.* **41**, D48–D55 (2013).
78. Meyer, L. R. et al. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**, D64–D69 (2013).
79. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
80. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
81. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model* (Springer, 2000).

Acknowledgements

We thank M. Gasperini, G. Findlay and J. Shendure for technical assistance and sharing pgRNA constructs, Q. Yan for sharing HeLa/iCas9 cells, A. Geballe for sharing Cas9-expressing IMR90 cells and D. Bennett for sharing Melan-a cells. J.D.T. is a Washington Research Foundation Postdoctoral Fellow. R.K.B. is a Scholar of The Leukemia and Lymphoma Society (1344-18). This research was supported in part by the Edward P. Evans Foundation, NIH/NIDDK (R01 DK103854), NIH/NHLBI (R01 HL128239), NIH/NINDS (P01 NS069539) and the Experimental Histopathology and Genomics Shared Resources of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704). The results published here are based in part on data generated by The Cancer Genome Atlas Research Network (http://cancergenome.nih.gov).

Author contributions

J.D.T., Q.F. and R.K.B. designed the study. J.D.T., J.T.P., Q.F., E.J.D.N., E.R.H., M.V.M., J.P., A.M.G., A.E.B., J.W., N.T.N. and A.H.B. performed experiments. J.D.T., J.T.P. and R.K.B. analysed data. J.D.T. and R.K.B. wrote the paper.

Competing interests

The authors declare no competing interests.

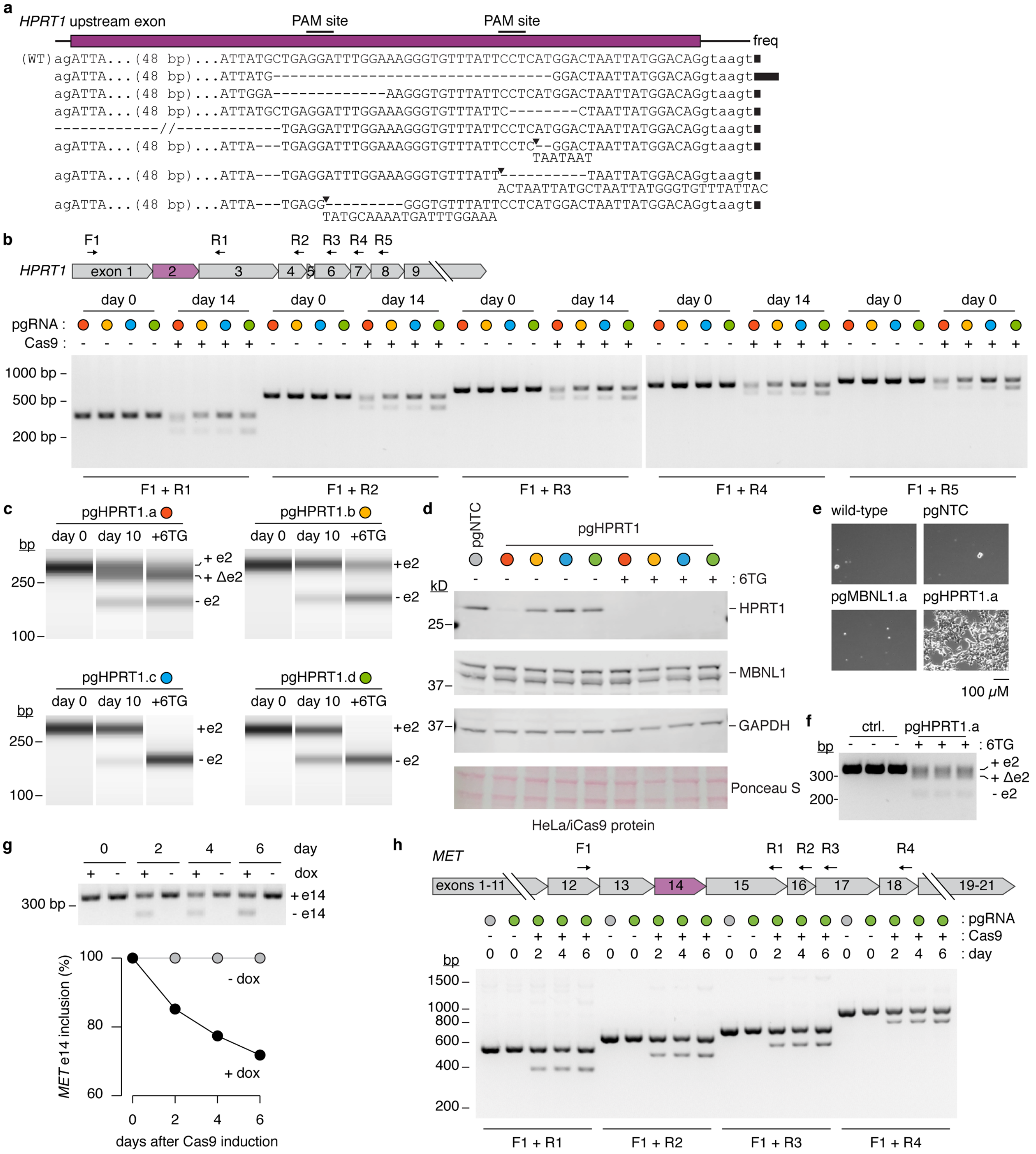
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-019-0555-z.

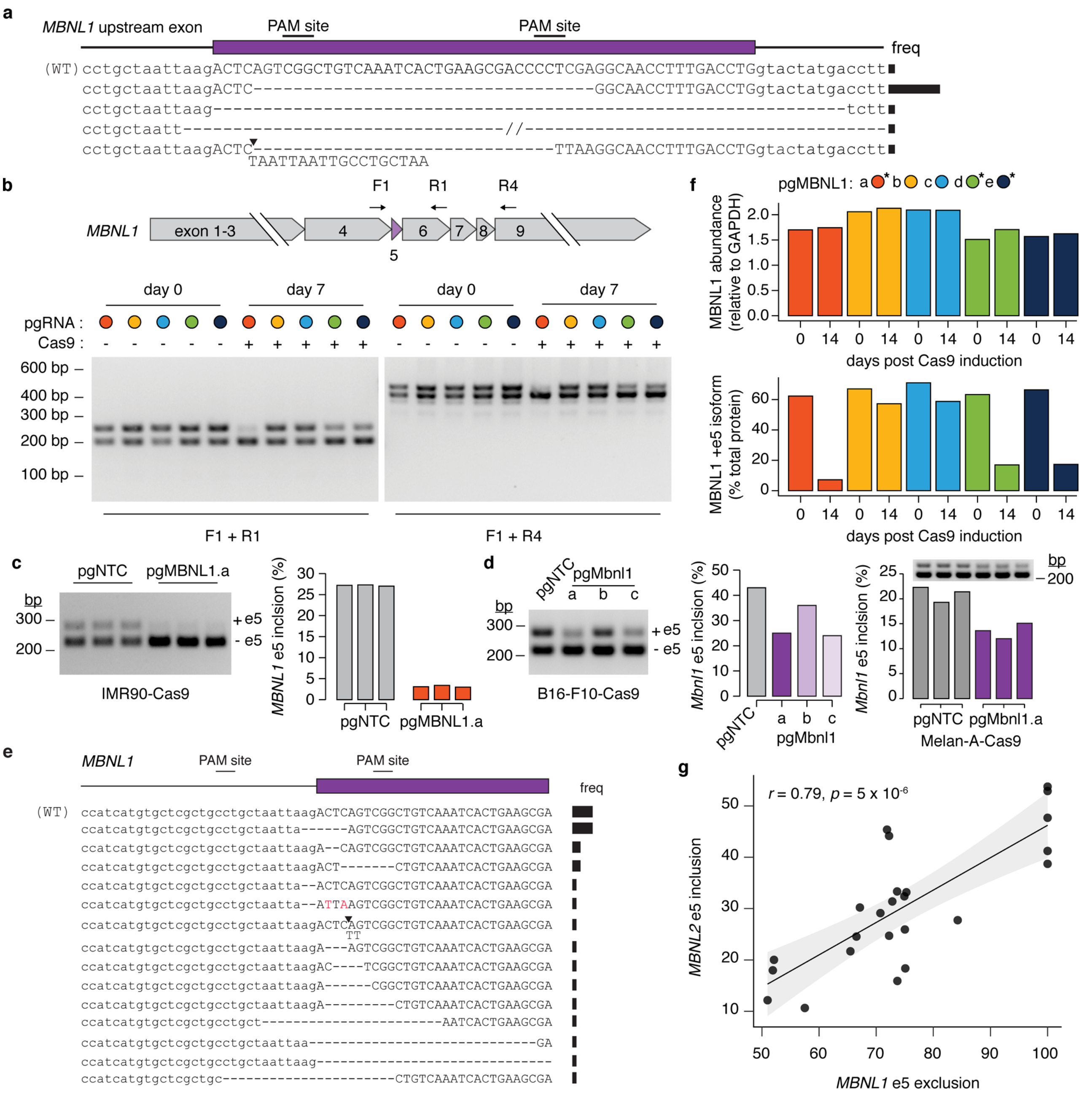
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0555-z.

Correspondence and requests for materials should be addressed to R.K.B.

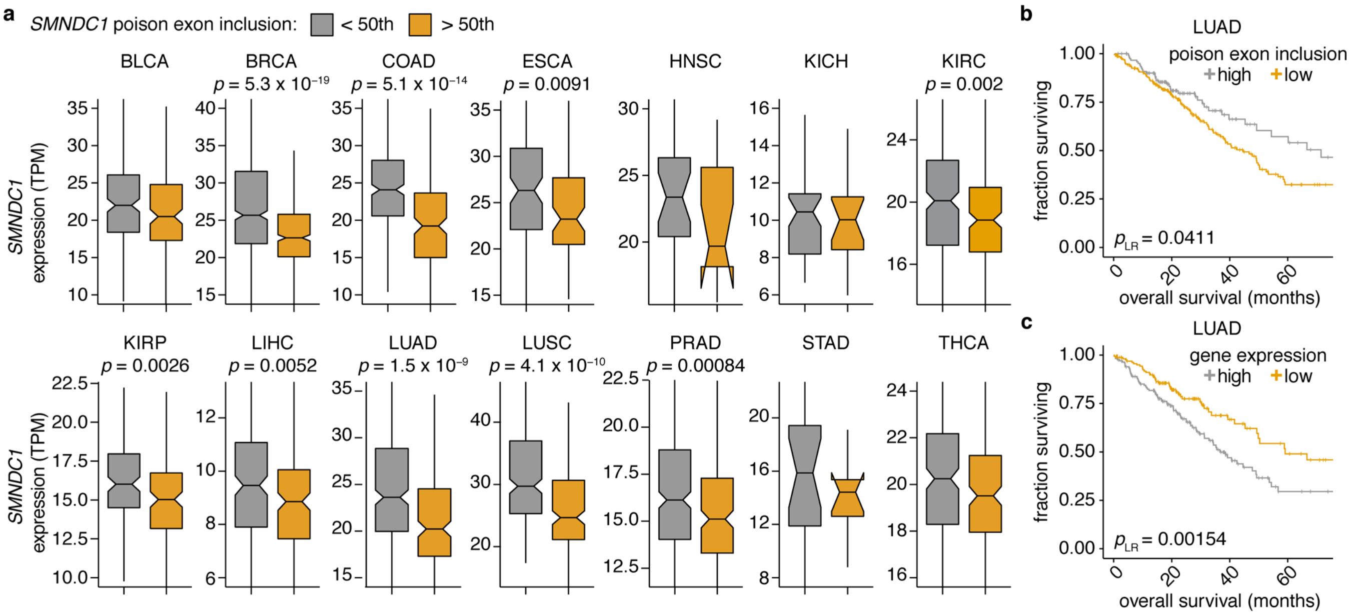
Reprints and permissions information is available at www.nature.com/reprints.



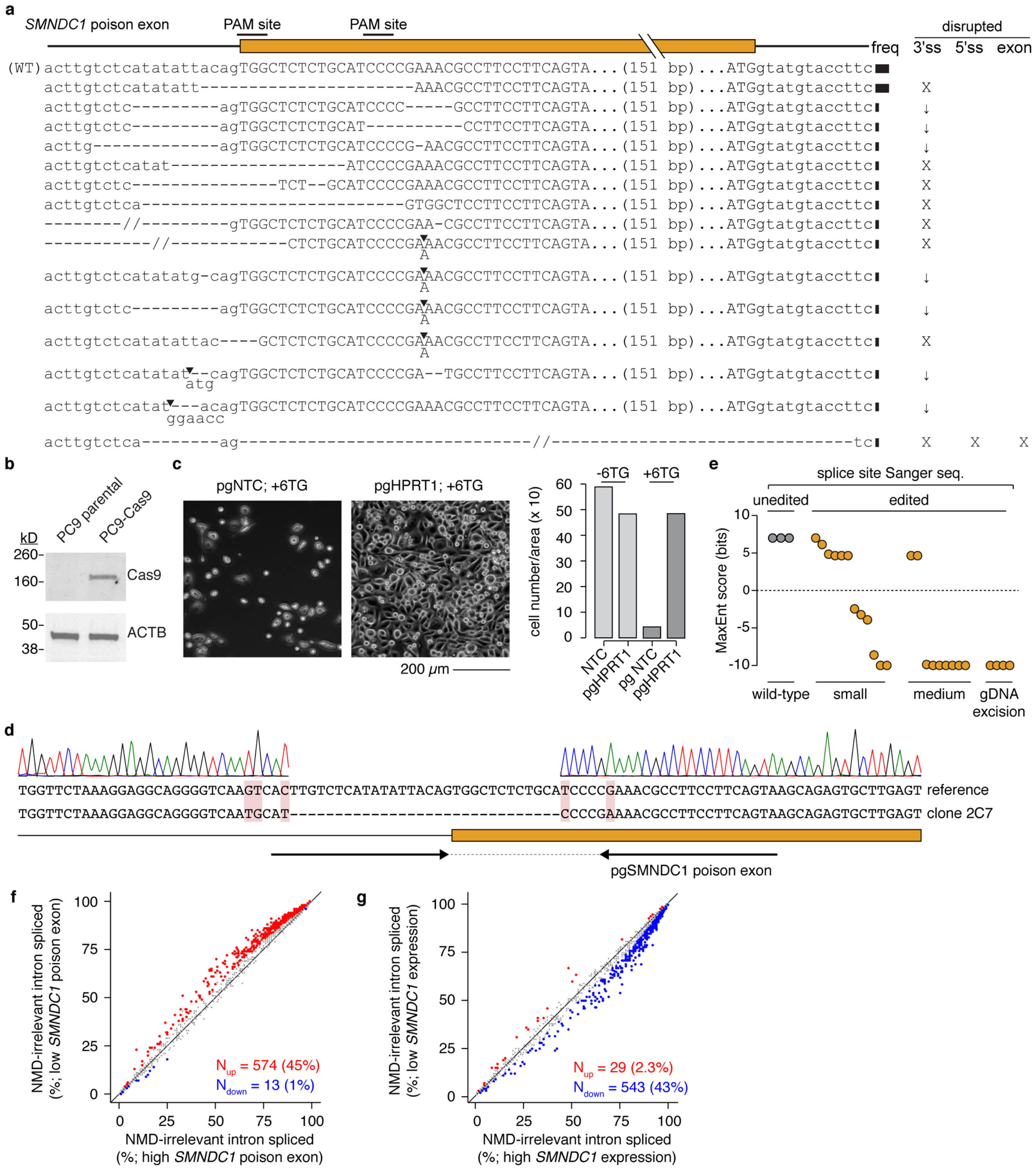
Extended Data Fig. 1 | pgFARM-induced exclusion of *HPRT1* exon two and *MET* exon 14. a, Sanger sequencing of pgFARM-edited *HPRT1* exon two in HeLa/iCas9 cells. **b**, Long range RT-PCR analysis of *HPRT1* exon two skipping. **c**, RT-PCR analysis of *HPRT1* exon two (e2) inclusion before/after Cas9 induction (day 0/day 10) and one week treatment with 6-thioguanine (+6TG). **d**, HPRT1 western blot analysis (n=1 independent experiments) before (-) and after (+) one week treatment with 6TG. **e**, Cas9-expressing HEK293T cells (n=3 biological replicates) that were untreated (wild-type) or expressing the indicated pgRNAs followed by one week treatment with 6TG. **f**, RT-PCR analysis of *HPRT1* exon two (e2) inclusion in Cas9-expressing HEK293T cells (n=3 biological replicates). **g**, Top, RT-PCR analysis of *MET* exon 14 (e14) inclusion with (+) or without (-) Cas9 expression. Bottom, quantification. (n=1 independent experiments). **h**, As for (**b**), but for *MET* exon 14. Gray, non-targeting pgRNA; green, pgRNA targeting *MET* exon 14. See Source Data for uncropped gels.



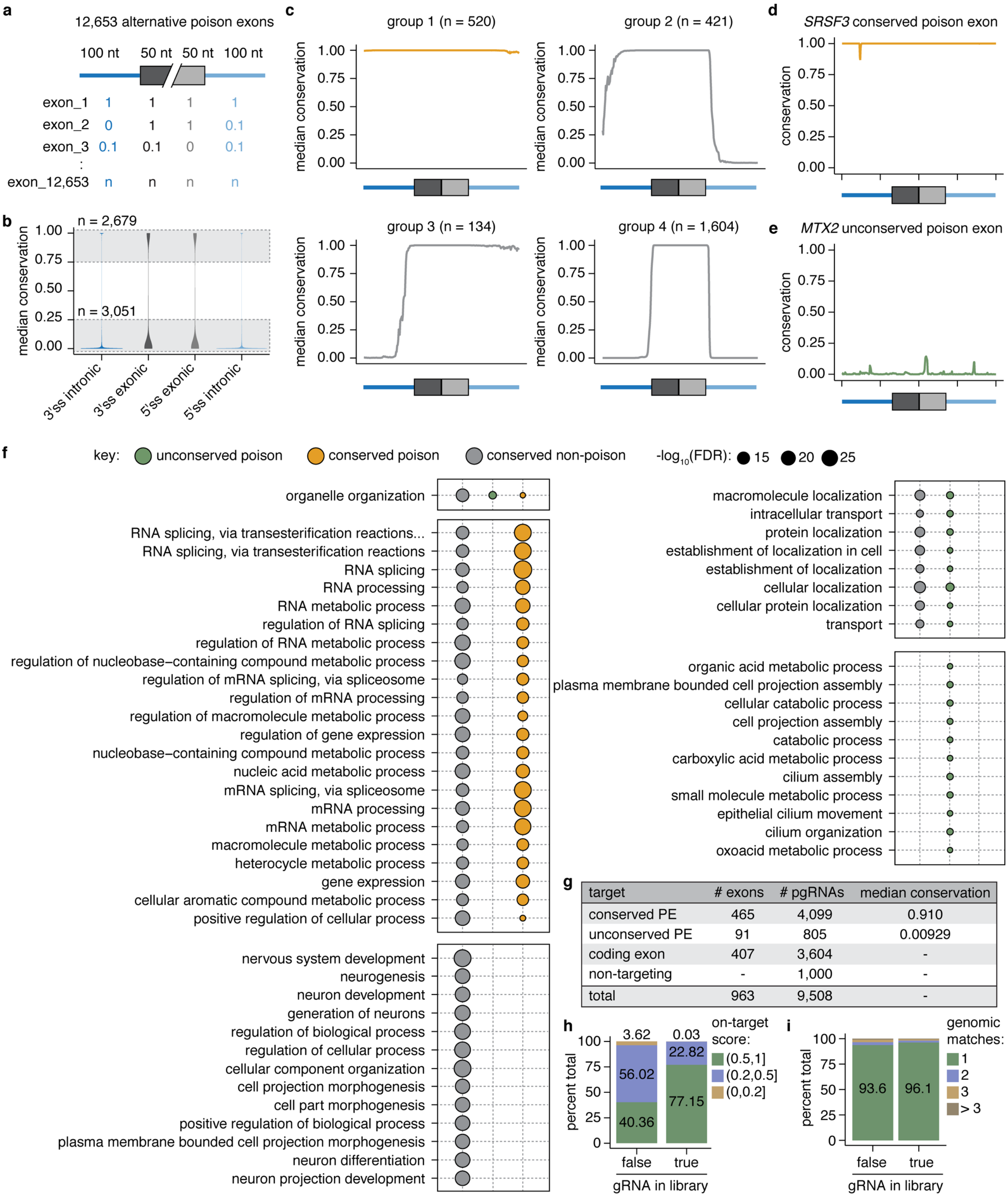
Extended Data Fig. 2 | pgFARM-induced exclusion of *MBNL1* exon five in multiple cell lines. a, Sanger sequencing of pgFARM-edited *MBNL1* exon two in HeLa/iCas9 cells. **b**, Long range RT-PCR analysis of *MBNL1* exon two skipping (n = 1 independent experiments). **c**, Left, RT-PCR analysis (n = 3 biological replicates per group) of *MBNL1* exon five (e5) inclusion in Cas9-expressing IMR90 cells expressing a non-targeting pgRNA (pgNTC) or pgMBNL1.a. Right, quantification of *MBNL1* exon 5 inclusion. **d**, Left and center, RT-PCR analysis and associated quantification of *Mbnl1* exon five (e5) inclusion in Cas9-expressing B16-F10 cells expressing the indicated pgRNA. Right, RT-PCR analysis (n = 3 biological replicates per group) and associated quantification of *Mbnl1* exon (e5) inclusion in Cas9-expressing Melan-A cells expressing the indicated pgRNA. **e**, Individual *Mbnl1* alleles that were cloned from gDNA of Cas9-expressing B16-F10 cells following delivery of a *Mbnl1* exon five-targeting pgRNA and subjected to Sanger sequencing. **f**, Quantification of total MBNL1 protein levels (top) and MBNL1 protein encoded by the exon five-including isoform (bottom) before (day 0) and after (day 14) Cas9 induction in HeLa/iCas9 cells expressing the indicated pgRNA, measured by immunoblot in Fig. 1l. *, pgRNAs that induced the greatest *MBNL1* exon five exclusion. Data are representative of n = 2 independent experiments. **g**, Scatter plot comparing pgRNA-mediated exclusion of *MBNL1* exon five (e5) and inclusion of *MBNL2* exon five (e5), a paralogous exon that is regulated by nuclear MBNL1. Datapoints (n = 24) are from HeLa/iCas9 cells treated with pgMBNL1.a, pgMBNL1.d, or pgMBNL1.e pgRNAs for two weeks. *r*, Pearson correlation; *p*, associated *p*-value computed using a two-sided Student's *t*-test; shaded region, 95% confidence interval. See Source Data for uncropped gels.



Extended Data Fig. 3 | *SMNDC1* poison exon inclusion in cancer. a, As Fig. 2c, but for all TCGA cohorts analyzed in Fig. 2d. *p* computed with two-sided Mann-Whitney *U* test. Hinges, notches, and whiskers indicate 25th/75th percentiles, 95% confidence interval, and most extreme datapoints within 1.5X interquartile range from hinge. Sample sizes are BLCA: n = 338; BRCA: n = 1089; COAD: n = 451; ESCA: n = 180; HNSC: n = 40; KICH: n = 62; KIRC: n = 430; KIRP: n = 262; LIHC: n = 350; LUAD: n = 502; LUSC: n = 447; PRAD: n = 481; STAD: n = 30; THCA: n = 362. **b**, Overall survival of lung adenocarcinoma (LUAD) patients, where patients were stratified according to the relative inclusion of the *SMNDC1* poison exon. High poison exon, top tercile of samples; low poison exon, bottom tercile of samples. *p* computed with a two-sided logrank test. n = 237 (low) and 132 (high) samples. The uneven sample allocation arises from edge effects at the boundaries of terciles (MISO only estimates exon inclusion to two significant digits). **c**, As (**b**), but for *SMNDC1* gene expression. High expression, top tercile of samples; low expression, bottom tercile of samples. *p* computed with a two-sided logrank test. n = 169 (low) and 174 (high) samples.

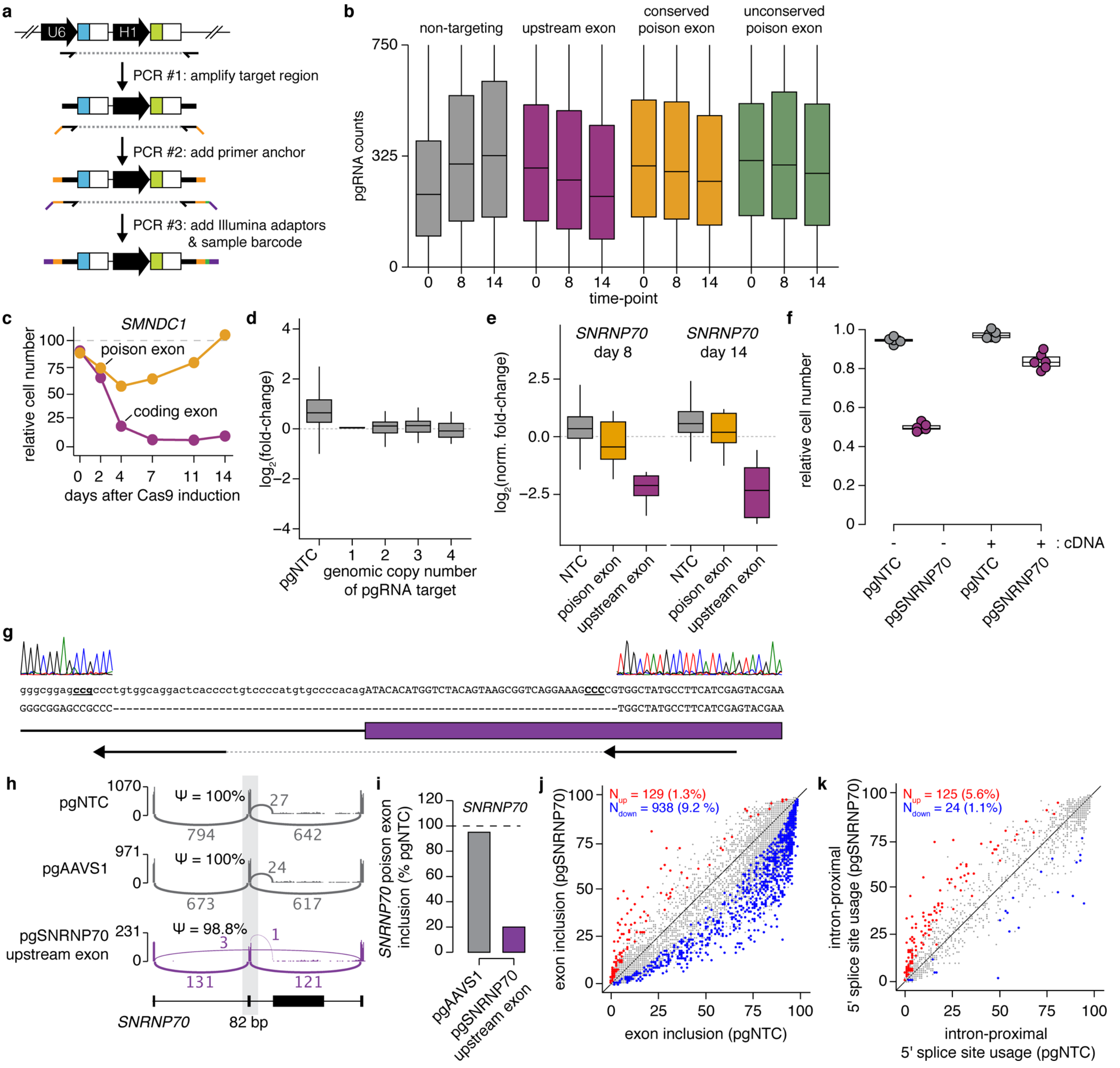


Extended Data Fig. 4 | pgFARM-induced exclusion of *SMNDC1*'s poison exon. **a**, Sanger sequencing of pgFARM-edited *SMNDC1* poison exon in HeLa/iCas9 cells. Annotations of eliminated (X) or disrupted (↓) sequence elements are indicated. **b**, Western blot for Cas9 and ACTB in parental PC9 and PC9-Cas9 (n = 3 biological replicates) transgenic cell lines. **c**, Left, PC9-Cas9 cells expressing the indicated pgRNAs following treatment with 6TG for one week. Right, quantification of cell survival. **d**, Representative *SMNDC1* allele (n = 25 total sequenced alleles) of a PC9-Cas9 clonal cell line isolated following delivery of an *SMNDC1* poison exon-targeting pgRNA. **e**, MaxEnt 3′ splice site scores for unedited (wild-type) or edited *SMNDC1* alleles from individual PC9-Cas9 clones. "small" and "medium" indicate alleles containing indels of length ~1–10 bp and > 10 bp without intervening gDNA excision; "gDNA excision" indicates alleles with complete excision of intervening gDNA. Each class of editing event can effectively reduce 3′ splice site strength. **f**, As Fig. 2j, but restricted to introns that are not NMD-targets (NMD-irrelevant). **g**, As Fig. 2k, but restricted to introns that are not NMD-targets (NMD-irrelevant). See Source Data for uncropped gels.

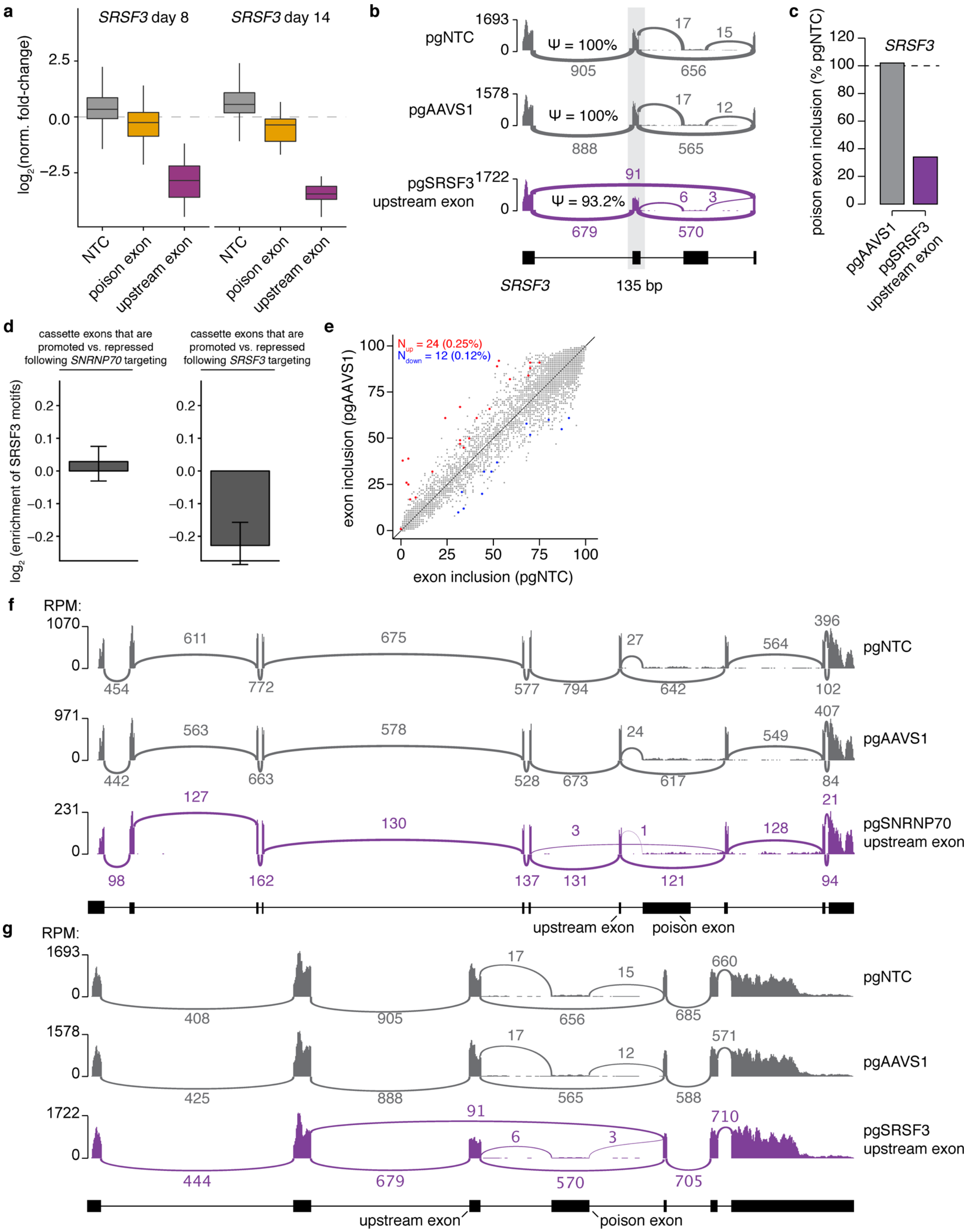


target	# exons	# pgRNAs	median conservation
conserved PE	465	4,099	0.910
unconserved PE	91	805	0.00929
coding exon	407	3,604	-
non-targeting	-	1,000	-
total	963	9,508	-

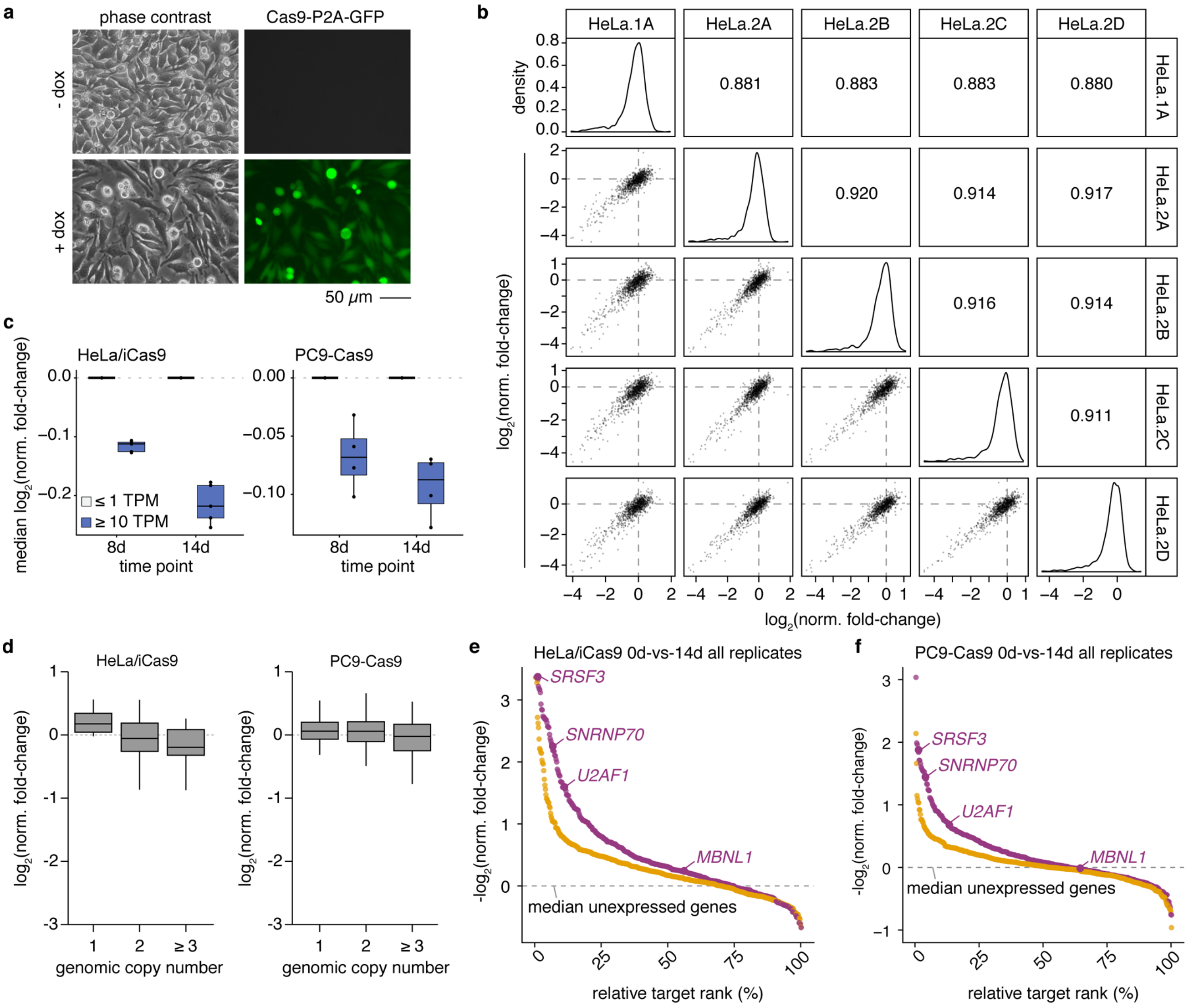
Extended Data Fig. 5 | pgRNA library design. a, Regions used to classify each poison exon (n = 12,653) according to its sequence conservation. **b**, Median conservation scores for each indicated region (violin plot width represents probability density of data distribution). **c**, Median per-nucleotide sequence conservation for exon groups described in the text. **d**, Per-nucleotide sequence conservation for an *SRSF3* ultraconserved poison exon. **e**, As (**d**), but for an *MTX2* poorly conserved poison exon. **f**, The most significant biological processes associated with genes containing unconserved poison exons (n = 2,363), conserved poison exons (n = 352), or conserved non-poison exons (n = 888) (related to Fig. 3c). FDR computed using the Wallenius method and corrected using the Benjamini-Hochberg method. **g**, pgRNA library summary. **h**, On-target scores (MIT score) for all gRNAs targeting 3′ splice sites analyzed in our study ("false") and those included in the final library ("true"). **i**, As (**h**), but for off-target scores identified using Cas-OFFinder.



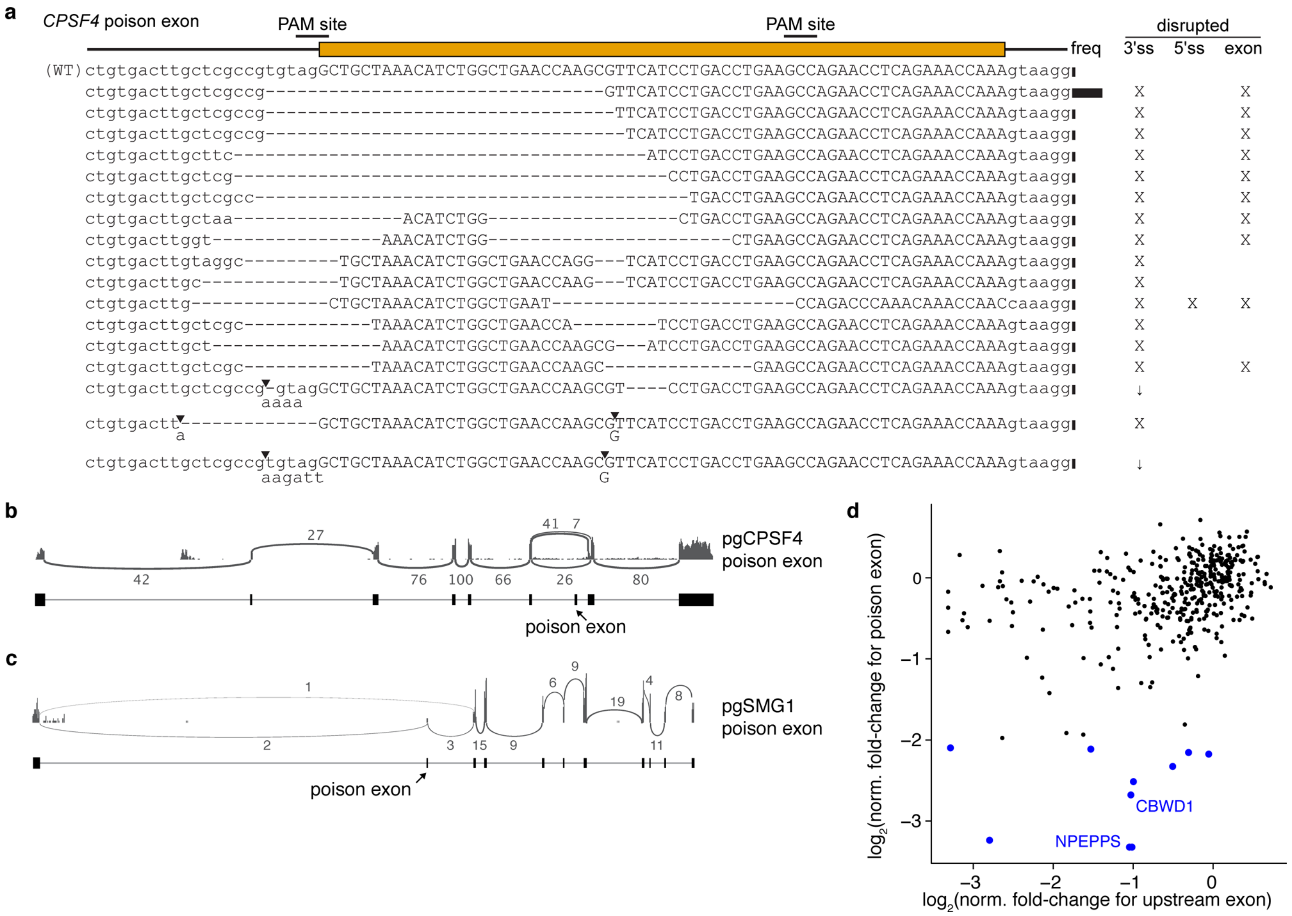
Extended Data Fig. 6 | Analysis of pilot pgFARM screen. a, pgRNA library generation for Illumina sequencing. **b**, pgRNA counts throughout the time course (n = 1,000; 3,604; 4,099; 805 for groups, left to right). **c**, Relative proliferation of HeLa/iCas9 cells expressing an *SMNDC1* upstream constitutive exon-targeting pgRNA relative to control pgRNA (non-essential gene *CSPG4*; n = 2 independent experiments). **d**, Unnormalized fold-changes for non-targeting pgRNAs (n = 1,000) and pgRNAs targeting unexpressed (< 1 transcripts per million, TPM) genes, located in genomic regions with the indicated copy numbers (n = 2, 38, 45, and 11, left to right). **e**, Normalized fold-changes for all non-targeting pgRNAs (NTC; n = 1,000) and pgRNAs targeting the indicated exons (n = 9 pgRNA per exon) in *SNRNP70*. **f**, Relative proliferation of HeLa/iCas9 cells expressing a *SNRNP70* upstream constitutive exon-targeting pgRNA without (-) or with (+) simultaneous overexpression of a SNRNP70-encoding cDNA (n = 6 replicates per condition). **g**, Representative Sanger sequencing of a pgFARM-edited *SNRNP70* upstream exon in HeLa/iCas9 cells (n = 19 total sequenced alleles). **h**, RNA-seq read coverage across the *SNRNP70* locus containing the targeted upstream constitutive exon (gray box) from HeLa/iCas9 cells expressing the indicated pgRNA (n = 1 per pgRNA). **Ψ**, percent spliced in. **i**, *SNRNP70* poison exon inclusion for HeLa/iCas9 cells expressing the indicated pgRNA relative to a non-targeting pgRNA (n = 1 per pgRNA). **j**, Scatter plot comparing cassette exon inclusion in HeLa/iCas9 cells treated with a non-targeting control pgRNA (pgNTC) or *SNRNP70* upstream constitutive exon-targeting pgRNA (pgSNRNP70). Points are shaded by statistical significance (two-sided Mann-Whitney test). **k**, As (**j**), but comparing alternative 5′ splice site usage. For box plots, the line, hinges, and whiskers represent median, 25th and 75th percentiles, and most extreme datapoints within 1.5X interquartile range from hinge. See Source Data for uncropped gels.



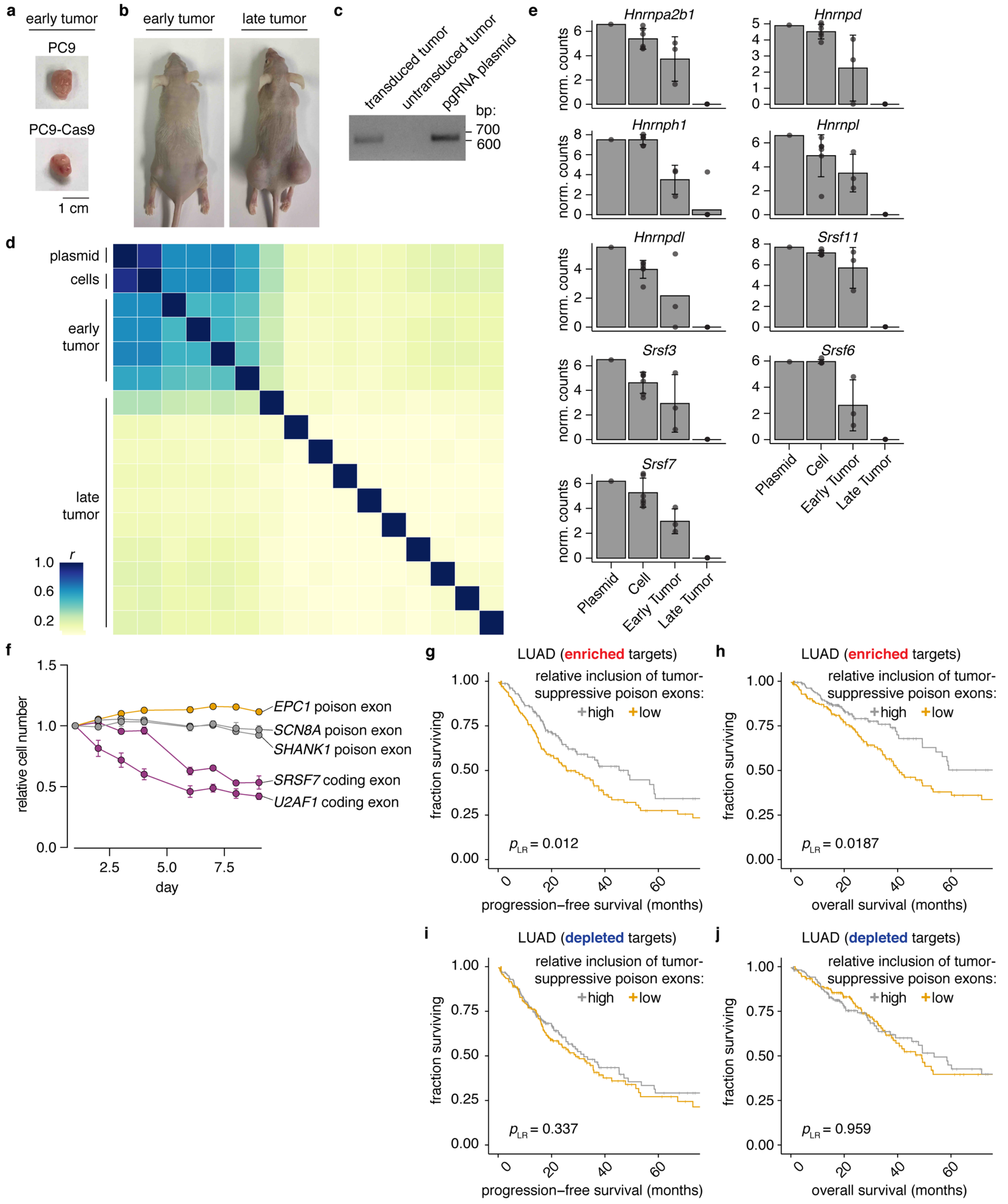
Extended Data Fig. 7 | Analysis of pilot pgFARM screen, continued. **a**, Normalized pgRNA fold-changes (n = 1,000 and 9 for non- and exon-targeting pgRNAs, respectively). The center line, hinges, and whiskers represent median, 25th and 75th percentiles, and most extreme datapoints within 1.5X interquartile range from hinge. **b**, RNA-seq read coverage across the *SRSF3* locus containing the targeted upstream constitutive exon (gray box) from HeLa/iCas9 cells expressing the indicated pgRNA (n = 1 per pgRNA). Ψ, percent spliced in. **c**, *SRSF3* poison exon inclusion for HeLa/iCas9 cells expressing the indicated pgRNA relative to a non-targeting pgRNA (n = 1 per pgRNA). **d**, *SRSF3* RNA binding motif enrichment in differentially spliced exons (n = 2,046 left; 727 right) in HeLa/iCas9 cells expressing the indicated pgRNA. Data presented as mean ± 95% confidence interval computed by bootstrapping. **e**, Scatter plot comparing cassette exon inclusion in HeLa/iCas9 cells treated with a non-targeting control pgRNA (pgNTC) or AAVS1-targeting control pgRNA (pgAAVS1). Points are shaded by statistical significance (two-sided Mann-Whitney *U* test). **f**, RNA-seq read coverage across the entire *SNRNP70* locus in HeLa/iCas9 cells expressing the indicated pgRNA (n = 1 per pgRNA). **g**, As (**f**), but for *SRSF3* (n = 1 per pgRNA).



Extended Data Fig. 8 | Analysis of large-scale pgFARM screens. a, HeLa/iCas9 cells (n = 4 biological replicates) treated with the poison exon pgRNA library and grown in the presence (+dox) or absence (- dox) of active Cas9. **b**, Scatter plots comparing normalized fold-changes (day 14 vs. day 0; n = 963 targeted exons) estimated with each replicate of the cell viability screen in HeLa/iCas9 cells. Pearson correlations for individual replicate comparisons are indicated. **c**, Normalized fold-changes for pgRNAs targeting exons in unexpressed (TPM ≤ 1; n = 96 for HeLa/iCas9 and 128 for PC9-Cas9) or highly expressed (TPM ≥ 10; n = 681 for HeLa/iCas9 and 661 for PC9-Cas9) genes. Each dot represents the median fold-change computed over all pgRNAs targeting exons in the indicated groups for a representative replicate from the screens in HeLa/iCas9 (left; n = 5) and PC9-Cas9 (right; n = 4) cells. TPM, transcripts per million. **d**, Normalized fold-changes for pgRNAs targeting lowly expressed genes (TPM < 5) located in genomic regions with the indicated copy numbers (n = 6, 165, and 14 per group, left to right, for HeLa/iCas9; n = 60, 107, and 45 per group, left to right, for PC9-Cas9). **e**, Rank plot of mean normalized fold-changes for conserved poison (orange) or upstream constitutive exons (purple) based on all replicates of the HeLa/iCas9 viability screen. **f**, As (**e**), but for all replicates of the PC9-Cas9 viability screen. For box plots, the center line, hinges, and whiskers represent median, 25th and 75th percentiles, and most extreme datapoints within 1.5X interquartile range from hinges, respectively.



Extended Data Fig. 9 | pgFARM-induced exclusion of *CPSF4* and *SMG1* poison exons. a, Sanger sequencing of pgFARM-edited *CPSF4* poison exon in HeLa/iCas9 cells. Annotations of eliminated (X) or disrupted (↓) sequence elements are indicated. **b**, RNA-seq read coverage across the entire *CPSF4* locus in HeLa/iCas9 cells expressing a *CPSF4* poison exon-targeting pgRNA (pgCPSF4; n = 1). We observed no read coverage indicative of cryptic splicing in pgCPSF4-treated cells. The two sets of splice junction reads downstream of the *CPSF4* poison exon correspond to usage of endogenous (naturally occurring in unedited cells) competing 3′ splice sites. **c**, As (**b**), but for an *SMG1* poison exon-targeting pgRNA (pgSMG1; n = 1). **d**, Scatter plot comparing normalized fold-changes for pgRNAs targeting a poison exon compared to matched upstream coding exon within the same gene.



Extended Data Fig. 10 | Analysis of xenograft screens. a, Tumors derived from parental PC9 or PC9-Cas9 cells (n = 4 per group). **b**, Mice from early and late tumor time points (n = 4 and 10 tumors, respectively). **c**, pgRNA Illumina libraries. **d**, Pearson correlation (*r*) matrix for xenograft screen samples. Unsupervised clustering of library depth-normalized pgRNA counts by the complete-linkage method. **e**, Normalized counts (mean ± S.D.) for gRNAs targeting coding exons in the indicated genes. Data from Chen et al, 2015 (n = 1, 6, 3, and 9 for groups, left to right). **f**, Relative cell number (mean ± S.D.) for PC9-Cas9 cells expressing a pgRNA targeting the indicating exons (n = 3 per group). **g**, Progression-free survival of lung adenocarcinoma patients (n = 167/171 for low/high categories), where patients were stratified by inclusion of tumor-suppressive poison exons. **h**, As (**g**), but for overall survival. **i**, As (**g**), but for essential poison exons (n = 166/169 for low/high categories). **j**, As (**i**), but for overall survival. See Source Data for uncropped gels.

nature research

Corresponding author(s): Robert K. Bradley

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	Publicly available software was used in this study. Specific programs are RSEM (v1.2.4), Bowtie (v1.0.0), TopHat (v2.0.8b), MISO (v2.0), and Bioconductor (v3.7) within the R (v3.5.1) programming environment. FIJI/ImageJ (v2.0.0) and HALO (v2.0, Indica Labs) were used for image analysis. MAFFT (v7.0) was used for Sanger sequence alignments.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-seq data generated as part of this study has been deposited in the Gene Expression Omnibus (accession number GSE120703). RNA-seq data generated by The Cancer Genome Atlas (TCGA) was downloaded from the Cancer Genomics Hub (CGHub) and Genomic Data Commons (GDC).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples sizes for pooled screens were based on results from our pilot screen as well as information garnered from published literature (Doench 2018, Chen et al 2015).
Data exclusions	No data were excluded.
Replication	All attempts at replication with both technical and biological replicates were successful.
Randomization	This is not relevant to our study, which relied on pooled screening. For experiments involving specific targets, all experiments involved control treatments (non-targeting, AAVS1-targeting, or CSPG4-targeting pgRNAs) or perturbation treatments (pgRNAs targeting exons).
Blinding	Blinding was not relevant to our study because of the pooled nature of the screen.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All unique materials are readily available from the authors.

Antibodies

Antibodies used	Antibodies against Cas9 (Cell Signaling #14697, 1:1000), ACTB (Cell Signaling (13E5) Rabbit mAb #4970, 1:5000), HPRT1 (Abcam ab10479, 1:1000), and GAPDH (Bethyl a300-639a, 1:5000) were used as primary antibodies for western blots. For immunofluorescence, MBNL1 (DSHB MB1a(4A8)) was used at a dilution of 1:500.

Validation	Cas9 antibody specificity was confirmed using 293T cells mock transfected or transfected with a construct expressing Cas9 (https://www.cellsignal.com/products/primary-antibodies/cas9-7a9-3a3-mouse-mab/14697). ACTB antibody specificity was confirmed by western blot analysis which revealed a band of the expected size (https://www.cellsignal.com/products/primary-antibodies/b-actin-13e5-rabbit-mab/4970). GAPDH antibody specificity was confirmed by western blot analysis which revealed a band of the expected size (https://www.bethyl.com/product/A300-639A/GAPDH+Antibody#). MBNL1 antibody specificity has been previously described (e.g., PMID: 19095965) and confirmed in our CRISPR-edited cell lines (e.g., Fig. 1l).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HeLa/iCas9 cells were obtained from Qin Yan (Yale School of Medicine). PC9 cells were obtained from M. Meyerson (Broad Institute/Dana-Farber Cancer Institute). Cas9-expressing IMR90 cells were obtained from Adam Geballe (FHCRC). Melan-a cells were obtained from Dorothy Bennett (St George's University of London). B16-F10 cells were obtained from ATCC (ATCC CRL-6475). HEK293T cells (PMID: 28335006) were obtained from Douglas Fowler (University of Washington).
Authentication	Cell lines were authenticated via RNA-seq expression profiling.
Mycoplasma contamination	Cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	None of the cell lines are commonly misidentified cell lines.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Species: Mus musculus. Strain: NU/J. Sex: male. Age: 8 to 12 weeks.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.