HOKKAIDO UNIVERSITY

Title	4.5SH RNA counteracts deleterious exonization of SINE B1 in mice
Author(s)	Yoshimoto, Rei; Nakayama, Yuta; Nomura, Ikuko et al.
Citation	Molecular cell, 83(24), 4479-4493 https://doi.org/10.1016/j.molcel.2023.11.019
Issue Date	2023-12-21
Doc URL	https://hdl.handle.net/2115/94052
Type	journal article
File Information	MOLECULAR-CELL-D-23-00692_R5_compresse.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Manuscript Click here to view linked References

4.5SH RNA counteracts deleterious exonization of *SINE B1* in mice

Rei Yoshimoto[1]*, Yuta Nakayama[2], Ikuko Nomura[2], Ikuko Yamamoto[2], Yumeka Nakagawa[1], Shigeyuki Tanaka[1], Misuzu Kurihara[2], Yu Suzuki[3], Takehiko Kobayashi[3], Hiroko Kozuka-Hata[4], Masaaki Oyama[4], Mari Mito[5], Shintaro Iwasaki[5,6], Tomohiro Yamazaki[7], Tetsuro Hirose[7,8], Kimi Araki[9, 10], and Shinichi Nakagawa[2]*†

[1]Department of Applied Biological Sciences, Faculty of Agriculture, Setsunan University, 45-1 Nagaotoge-cho, Hirakata City, Osaka, 573-0101

[2]RNA Biology Laboratory, Faculty of Pharmaceutical Sciences, Hokkaido University, Kita 12-jo Nishi 6-chome, Kita-ku, Sapporo 060-0812, Japan

[3]Laboratory of Genome Regeneration, Institute for Quantitative Biosciences, University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan.

[4]Medical Proteomics Laboratory, The Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

[5]RNA Systems Biochemistry Laboratory, RIKEN Cluster for Pioneering Research, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

[6]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba 277-8561, Japan

[7]RNA Biofunction Laboratory, Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan

[8]Institute for Open and Transdisciplinary Research Initiatives (OTRI), Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan

[9]Institute of Resource Development and Analysis, Kumamoto University, Kumamoto 860-0811, Japan.

[10]Center for Metabolic Regulation of Healthy Aging, Kumamoto University, 1-1-1, Honjo, Kumamoto 860-8556, Japan.

*Corresponding should be addressed to

Shinichi Nakagawa

nakagawas@pharm.hokudai.ac.jp

and

Rei Yoshimoto

rei.yoshimoto@setsunan.ac.jp

†Lead contact

Summary

4.5SH RNA is a highly abundant, small rodent-specific noncoding RNA that localizes to nuclear speckles enriched in pre-mRNA splicing regulators. To investigate physiological functions of *4.5SH* RNA, we have created mutant mice that lack the expression of *4.5SH* RNA. The mutant mice exhibited embryonic lethality, suggesting that *4.5SH* RNA is an essential species-specific noncoding RNA in mice. RNA-sequencing analyses revealed that *4.5SH* RNA protects the transcriptome from abnormal exonizations of the antisense insertions of the retrotransposon *SINE B1* (*asB1*), which would otherwise introduce deleterious premature stop codons or frameshift mutations. Mechanistically, *4.5SH* RNA base-pairs with complementary *asB1*-containing exons via the target recognition region, and recruits effector proteins including Hnrnpm via its 5' stem loop region. The modular organization of *4.5SH* RNA allows us to engineer a programmable splicing regulator to induce the skipping of target exons of interest. Our results also suggest the general existence of splicing regulatory non-coding RNAs.

Key words: *4.5SH* RNA, noncoding RNA, exon skipping, alternative splicing, retrotransposon, *SINE B1*, Hnrnpm

Introduction

Retrotransposons constitute a significant portion of the genomes of higher eukaryotes. Notably, the 7SL-derived retrotransposons, *SINE B1* in rodents and *Alu* in primates, represent the most prevalent mobile elements within these species [1]. When inserted in the antisense direction, *SINE B1* family retrotransposons provide all the critical sequence elements required for exon definition, except for a branch point sequence, including a polypyrimidine tract, a nearly optimal 3' splice site, and suboptimal 5' donor sites (Sorek, 2007). Thus, the antisense insertion of *SINE B1* family retrotransposons in introns serves as a potent source of deleterious exon that introduces premature stop codons or frameshift mutations, when inserted downstream of a branch point consensus sequence followed by mutations that optimize the splice sites [2,3]. Human cells have been shown to counteract these toxic exonizations of *Alu* elements, human counterparts of *SINE B1* family retrotransposons, using HNRNPC that competes with U2AF2 for polypyrimidine tract sequences of *Alu* [4]. It remains unknown whether this mechanism is conserved between species that harbor abundant *SINE B1* family retrotransposons in the genomes, such as rodents.

A huge number of noncoding RNAs are transcribed from the genomes of higher eukaryotes, many of which are species-specific and localized in the nucleus [5,6]. *4.5SH* RNA is a classic 90-nucleotide nuclear noncoding RNA that was identified in the 1970s [7-9], and is found specifically in a group of small and short-lived rodents, including mice and rats [10]. The nucleotide sequence of the *4.5SH* RNA is highly homologous to *SINE B1*, but *4.5SH* lacks transposing activity due to the absence of an A-rich tail required for the retrotransposition of *SINE*s [1]. The *4.5SH* gene is a multicopy gene and forms a large cluster of tandem repeats that produces highly abundant transcripts, reaching over 10,000 molecules per cell [11]. *4.5SH* RNA is ubiquitously expressed in various tissues and cell types, and the transcripts are localized to nuclear speckles enriched in splicing factors [12]. The depletion of *4.5SH* RNA in murine cell lines suppresses cell proliferation [12], however, both the molecular mechanism and its physiological relevance have long been unknown.

In this study, we created mutant mice that lack the expression of *4.5SH* RNA, and found that the mutant mice exhibited early embryonic lethality. Notably, hundreds of novel exons were abnormally included in the transcripts of mutant cells, the majority of which contained the antisense insertions of *SINE B1*. Minigene analyses revealed that *4.5SH* RNA base paired with

SINE B1-containing deleterious exons, and blocked their inclusion by inhibiting the use of the 3' splice sites. Base-pairing of *4.5SH* RNA to the target sites was not sufficient for the exon skipping activity, which required a recruitment of effector proteins, including Hnrnpm that bound to the 5' located stem loop structure of *4.5SH* RNA. The modular organization of *4.5SH* RNA allowed us to generate a programmable splicing regulator that can induce skipping of exons of interest, by replacing the target recognition module with sequences complementary to the target exons. Our results also suggest the general existence of noncoding RNAs that associate with Hnrnpm and regulate alternative splicing of target exons that contain complementary sequences.

Results

4.5SH RNA is essential for early mouse development and ES cell proliferation

To explore the physiological function of *4.5SH* RNA, we created mutant mice for this multicopy gene. We first performed pulse-field gel electrophoresis to precisely estimate size and copy number of the *4.5SH* gene cluster, which had remained ambiguous due to a large gap in this genomic region even in the latest mouse genome assembly (Figure 1A). Treatment of genomic DNA with a restriction enzyme that cuts within the repeat unit containing *4.5SH* yielded a band of uniform length (4.2 kb) on the Southern blot, as previously described [11] (Figure 1B). In contrast, enzymes that cut outside of the gene cluster yielded a single ~900 kb band (Figure 1B), suggesting that >200 copies of *4.5SH* formed a large tandem cluster at this single locus, which contains a vast majority, if not all, of the *4.5SH* genes. To delete the large megabase-sized *4.5SH* cluster, we utilized CRISPR-Cas9 to cut proximal regions of the arms of the gene-targeting vector to enhance homologous recombination efficiency (Figure 1A, Supplemental figure S1 A-D). We successfully obtained embryonic stem (ES) cell clones lacking the entire *4.5SH* cluster, which were subsequently used to generate *4.5SH* knockout (KO) mice (Supplemental figure S1E-H). KO embryos at pre-implantation stages, such as E2.5–3.5, exhibited no overt abnormality (Figure 1C, D). However, we could not obtain the homozygotes later than E9.5 (Figure 1C), and only abnormal cellular aggregates were recovered at E6.5 (Figure 1E). To further investigate cellular and molecular functions of *4.5SH* RNA, we established embryonic stem (ES) cells, pluripotent cells derived from the inner cell mass of blastocyst-stage embryos, from wild-type (WT) and KO embryos (Figure 1F). While the expression of pluripotent markers, such as alkaline phosphatase and Nanog, was maintained in *4.5SH* KO ES cells (Figure 1G), decreased cellular proliferation was observed compared to WT ES cells (Figure 1H). Exogenous *4.5SH* expression rescued the decrease in cell proliferation (Figure 1F, H), suggesting that transcribed *4.5SH* RNA, but not cis-DNA elements in the cluster, regulates the proliferation of ES cells.

Deleterious cryptic exons containing the antisense insertions of *SINE B1* were abnormally included in *4.5SH* KO ES cells

As previously described [12], *4.5SH* RNA was localized to nuclear speckles identified by the

expression of Srsf1 in ES cells (Figure 2A). Because nuclear speckles are enriched in splicing factors [13], we expected that *4.5SH* RNA regulates the pre-mRNA splicing. To investigate splicing events altered in the *4.5SH* KO ES cells, we performed rMATS analyses, which can detect changes in alternative splicing events from short-read deep sequencing data [14]. Using RNA-sequencing data obtained from WT and *4.5SH* KO ES cells, we detected 160 cassette exons differentially spliced (FDR <0.001) between WT and KO cells, among which 100 (63%) exons were preferentially found in KO cells (Figure 2B, Supplemental Table T1). Notably, 75% of the KO-enriched exons were RefSeq-unannotated cryptic exons (Figure 2B, Supplemental Table T1), which were not found in the normal transcriptome of ES cells.

We then sought to determine whether there are any characteristic sequences common to these KO-enriched exons. Although these exons displayed features reminiscent of typical exons when compared to RefSeq exons (Supplemental figure S2A-E), motif analyses with MEME [15] revealed three consensus sequences highly homologous to the antisense insertions of retrotransposon *SINE B1* (*asB1*), two of which (motif 1 and 3) were complementary to *4.5SH* (Figure 2C, D, Supplemental figure S2F). The preferential appearance of *asB1* in KO-enriched exons was strand-specific, and other retrotransposons, such as *SINE B2*, were not enriched in these exons (Figure 2E). Indeed, *asB1*s innately contain sequence motifs required for exon definition, including the 5' located polypyrimidine tract, the 3' splice site sequences AG flanked by pyrimidine and G, and the 5' splice site containing GT, although the latter sequences are not optimal (Supplemental figure S2F), as has previously been reported [3]. Accordingly, insertions of *asB1* downstream of branch point sequences are potentially recognized as exons, which seemed to be blocked by *4.5SH* RNA. Of the *asB1*-containing cryptic exons, 81% introduced PTC or frameshift mutations in the host gene (Supplemental figure S2G), implicating the inclusion of *asB1*-containing deleterious exons as the primary cause of embryonic lethality in *4.5SH* KO mice.

To further confirm the KO-specific inclusions of *asB1*-containing exons, we performed RT-PCR for the representative host genes, including *Slc25a40*, *Smchd1*, and *Smg6* using RNAs derived from WT and 4.*5SH* KO ES cells (Figure 2F-H). Importantly, the *asB1*-containing cryptic exons from these genes were largely skipped and were undetectable in WT cells (Figure 2G, H). In KO cells, however, transcripts containing these abnormal exons became predominant, which were no longer observed in the rescued cells expressing exogenous *4.5SH* RNA (Fig. 2G, H). These exons of the representative genes contained premature stop codons, and the decrease of Smchd1

and Smg6 full-length protein products was confirmed on western blots (Figure 2I). The decreased protein expression of representative host genes were rescued upon exogenous introduction of *4.5SH* (Figure 2I). The inclusion of *asB1*-containing exons from the representative genes in *4.5SH* KO ES cells was consistently observed in both nuclear and cytoplasmic RNAs (Supplemental figure S3), suggesting that the elevated *asB1*-inclusion was not due to differential subcellular localization of the transcripts. Additionally, these findings were consistent regardless of the choice of primers—be it oligo-dT or random primed—for the RT reactions (Supplemental figure S3), indicating that the *asB1*-containing exons in *4.5SH* KO ES cells are present in mature mRNAs.

We then performed analyses of differentially expressed genes and found that 1014 of Ensembl-annotated genes exhibited significantly altered expression patterns (absolute log2 fold change > 1, FDR < 0.01) in the *4.5SH* KO ES cells (Supplemental figure S4A). These include genes associated with ribosome biogenesis, fatty acid degradation, spliceosome components, and more (Supplemental figure S4B, C). These findings imply that *asB1* exon inclusion perturbs the functions of numerous host genes, leading to a cascade of effects on various cellular processes and resulting in decreased cellular proliferation in the *4.5SH* KO ES cells. Notably, genes neighboring the *4.5SH* gene cluster, including *Cul1*, *Ezh2*, *Pdia4*, and *Zfp786*, all exhibited a subtle upregulation (log2 fold change = 0.32~0.72) (green dots in Supplemental figure S4D, Figure S4E). This suggests that the deletion of the *4.5SH* cluster might have influenced the higher-order chromatin organization of this locus and affected their gene expression. We also noticed that genes with KO-specific inclusions of *asB1*-containing exons tend to be downregulated (magenta dots in Supplemental figure S4D). This downregulation is likely due to the introduction of premature stop codons and/or frameshift mutations from the *asB1*-containing exon inclusions, leading to the degradation of these aberrant transcripts through the nonsense-mediated decay pathway [16].

4.5SH RNA recognizes the target exon via base-pairing and inhibits the use of the 3' splice site

To further explore the molecular mechanisms of *4.5SH* RNA, we constructed a minigene by inserting the cryptic exon of *Slc25a40* containing *asB1* (*asB1-Slc25a40*) together with the flanking sequences into the intron of a well-characterized splicing reporter derived from adenovirus [17] (Figure 3A). When this minigene was introduced into human-derived HEK293 cells

that do not express *4.5SH* RNA, *asB1-Slc25a40* was almost completely spliced in (Figure 3B), suggesting that the cryptic exon was recognized as a *bona fide* exon in cells lacking the *4.5SH* expression. In contrast, when the minigene was introduced into murine cell line Neuro2A, which endogenously expresses *4.5SH* RNA, 55% of *asB1-Slc25a40* was skipped to produce the shorter product (Figure 3B). The skipped product was also observed when exogenous *4.5SH* RNA was expressed under the control of U6 promoter in HEK293 cells (Figure 3B), indicating that we could successfully recapitulate the *4.5SH* RNA-dependent exon skipping of *asB1* using the minigene system.

We then tested whether *4.5SH* RNA binds to the target exons via base-pairing using complementary mutations (Figure 3C). When mutations that disrupt base pair formation were introduced in either *asB1-Slc25a40* or *4.5SH* RNA, target exon skipping was decreased (Figure 3D). The lower skipping activity was completely recovered when reciprocal mutations to rescue base-pairing were simultaneously introduced (Figure 3D), suggesting that *4.5SH* RNA hybridizes to target exons to inhibit their inclusions. Using split minigenes, we also found that *4.5SH* RNA inhibits the use of the 3' splice site, but not the 5' splice site of *asB1-Slc25a40* (Figure 3E, F).

SL1-4.5SH is required for the exon skipping activity of *4.5SH* RNA

We then attempted to identify sequence elements required for the function of *4.5SH* RNA using a series of deletion mutants that harbor non-overlapping 4 nucleotides deletions (Figure 4A, B). Deletions that lead to marked reduction of *asB1-Slc25a40* skipping were mapped to one of the three stem loop regions (SL1, SL2, and SL3) predicted by RNAfold2 [18], whereas linker regions were rather resistant to the mutations (Figure 4C, D). Northern blot analyses revealed that the deletions in the latter half of SL2 or SL3 reduced the expression of the mutant *4.5SH* RNAs, while deletions in SL1 specifically affected the exon skipping activity without affecting the amount of the exogenously expressed RNA (Figure 4C, D), suggesting that SL1 of *4.5SH* RNA (SL1-4.5SH) was essential for the exon skipping activity of *4.5SH* RNA. We then noticed that SL1-4.5SH does not base-pair with the target exons of certain host genes including *Ppil4* and *Eya3*, which lack the motif3 that would base-pair with SL1-4.5SH (Figure 4B, E). Furthermore, MEME-identified motif3 was found in only 41% of the KO-enriched, *asB1*-containing exons (Supplemental figure S5A, B). These observations suggest that SL1-4.5SH has a unique function other than

hybridization-based target recognition. To test this possibility, we replaced SL1-4.5SH with a corresponding region of *SINE B1* (SL1-B1) that has homologous sequences yet exhibits a distinct predicted secondary structure (Figure 4F). Unlike wild-type *4.5SH* RNA, the chimeric molecule with SL1-B1 could not induce efficient exon skipping of *asB1-Slc25a40* (Figure 4G, H) even though it is predicted to form a more stable secondary structure with target sequences (Figure 4I). These observations suggest that base-pairing to target exons is not sufficient to induce exon skipping by itself. Instead, *4.5SH* RNA was predicted to consist of two modules, a target-recognition module complementary to *asB1* and a potential effector-binding module, SL1-4.5SH (Figure 4B). This idea was consistent with the aforementioned result where reciprocal mutations in the target recognition module restored the exon skipping of mutated *asB1-Slc25a40* (Figure 3C, 4B). We were then interested in whether we could improve the exon skipping efficiency by increasing the complementarity between *4.5SH* RNA and the target exon. We thus modified the target recognition module of *4.5SH* RNA to the sequences that perfectly base-pair with *asB1-Slc25a40*. Unexpectedly, the exon skipping activity of the perfect-match *4.5SH* RNA was rather decreased compared to the wild type *4.5SH* RNA (Figure 4J), further supporting the idea that simple hybridization is not sufficient for *4.5SH* RNA-dependent exon skipping.

SL1 of *4.5SH* RNA associates with effector proteins to induce exon skipping

Given that, we reasoned that SL1-4.5SH recruits protein(s) that induce exon skipping to the target asB1-containing cryptic exons. To identify the SL1-4.5SH-associating effector proteins, we performed RNA-pull down experiments using biotinylated SL1-4.5SH and nuclear lysates prepared from murine cell line Neuro2A (Figure 5A). Multiple proteins preferentially associated with SL1-4.5SH over SL1-B1, and a 75-kDa protein remained bound even after a stringent high-salt washing with 1M NaCl (Figure 5B, C). Mass spectrometry and subsequent western blot analysis revealed that at least three RNA binding proteins, Hnrnpm, Sfpq, and Nono, were preferentially associated with SL1-4.5SH (Figure 5D, E). Among these RNA binding proteins, Hnrnpm remained bound to SL1-4.5SH in the presence of 1M NaCl, which was consistent with the presence of the most representative binding motif GUGGUGG identified in a previous study [19], located in the distal region of SL1-4.5SH (Figure 5F). To further confirm the binding of these RNA binding proteins to *4.5SH* RNA in the cells, we performed CLIP analyses using HEK293

cells expressing *4.5SH* RNA and control cells expressing nonfunctional SL1-B1-*4.5SH* RNA. Detection of the immunoprecipitated RNAs by northern blot revealed that all three proteins bound to *4.5SH* RNA, but not to the non-functional chimeric molecule (Figure 5G).

We then examined the subnuclear distribution of these RNA binding proteins and compared their signals with that of *4.5SH* RNA using ES cells (Figure 5H). Hnrnpm formed one or two large foci in this cell type (Figure 5H, white arrowhead), which did not overlap with the signals of *4.5SH* RNA. However, many of the peak signals of *4.5SH* RNA overlapped with local peak signals of Hnrnpm in the other regions in the nucleoplasm (Figure 5H, black arrowheads). The signals of Sfpq and Nono were largely overlapped with that of *4.5SH* RNA (Figure 5H, black arrowheads). Sfpq and Nono are essential components of nuclear body paraspeckle, and are enriched in the nuclear bodies in cells expressing the architectural noncoding RNA *Neat1_2* [20]. Because ES cells do not express *Neat1_2* and thus lack the formation of paraspeckles [21], we also examined the expression of *4.5SH* RNA and other associating proteins in NIH3T3 cells that have prominent paraspeckles. As in ES cells, the signals of *4.5SH* RNA mostly overlapped with those of Srsf1 (Figure 5H, arrowheads), indicating that the RNA is also mainly localized to nuclear speckles in NIH3T3 cells. The signals of Hnrnpm were also detected in close proximity to those of *4.5SH* RNA, but the frequency of the overlap was reduced compared to Srsf1 (Figure 5H, arrowheads). Sfpq and Nono were mainly localized to the paraspeckles, and only weak signals were detected in the nucleoplasm under the experimental conditions we used for the simultaneous detection with RNA (Figure 5H). However, some of the peak signals of *4.5SH* RNA overlapped with local peak signals of Sfpq and Nono in the nucleoplasm (Figure 5H, arrowheads). These observations suggest that the association of *4.5SH* RNA with these effector proteins is not strictly stable, but is dynamically regulated within the nucleus.

To investigate the functional involvement of the effector proteins, we performed knockdown experiments. Introduction of the minigene into HNRNPM- and SFPQ-depleted HEK293 cells resulted in decreased skipping of *asB1-Slc25a40* (Figure 6A, Supplemental figure S6A), suggesting that these two proteins form a functional complex with *4.5SH* RNA to induce skipping of target exons. We also observed slight decrease in skipping activity in the cells depleted with NONO, although the effect was not statistically significant (Figure 6A). This may be due to the inefficient knockdown of NONO (Supplemental figure S6A), or due to a redundant function of Nono that are highly homologous to Sfpq [22]. We then examined a dominant effect of the effector

proteins by recruiting them to the target exon using the MS2 system [23] (Figure 6B, Supplemental figure S6B). As expected, forced recruitment of HNRNPM, SFPQ, and NONO induced skipping of the *asB1*-containing exon in the absence of *4.5SH* RNA, whereas recruitment of MS2 coat protein had little effect (Figure 6B). The dominant effects of the effector proteins were not observed when the proteins were recruited to the downstream intron (Figure 6C). These observations suggest that *4.5SH* RNA base-pairs with the target *asB1*-containning exons via the target-recognition module, recruits effector protein complex containing Hnrnpm and other associating proteins, and induces skipping of the target exon.

HNRNPC has previously been shown to prevent aberrant exonizations of antisense insertions of *Alu* elements, *SINE B1* family retrotransposons in humans, by competing with U2AF2 [4]. To investigate whether HNRNPC is also required for the exon skipping activity of *4.5SH* RNA, we introduced the splicing reporter minigene and the *4.5SH*-expression vector into HNRNPC-depleted HEK293 cells. As previously reported [4], *asAlu*-containing cryptic exons of *HELLS* and *KIAA1432* were abnormally spliced-in upon HNRNPC knockdown (Figure 6D). However, *4.5SH* RNA-dependent exon skipping of *asB1-Slc25a40* was unaffected even in the absence of HNRNPC (Figure 6D). The skipping of *asB1*-containing exons, such as *asB1-Slc25a40* and *asB1-Smg6*, was also unaffected in Neuro2A cells depleted with Hnrnpc (Figure 6E), whereas the Hnrnpc-mediated skipping of *asAlu*-containing exon from *CD55* minigene was evident in the murine cells (Figure 6E). These findings indicate that small rodents and primates have independently developed strategies to mitigate the adverse effect of exonizations from the antisense insertions of *SINE B1* family retrotransposon, utilizing noncoding RNA and RNA-binding proteins, respectively.

Programmable splicing regulator can be generated by utilizing the modular structure of *4.5SH* RNA

Given the modular organization of *4.5SH* RNA, we tested whether we could engineer a programmable splicing regulator consisting of SL1-4.5SH and designed complementary sequences to target exons. We prepared a minigene containing the well-characterized alternative exon (ex) 6 of *FAS* (*FAS*-ex6) [24] and flanking exons, and tested the activity of a chimeric molecule consisting of SL1-4.5SH fused to the sequences complementary to *FAS*-ex6 (SL1-4.5SH-*asFAS*) (Figure 7A, B). As expected, SL1-4.5SH-*asFAS* induced the skipping of *FAS*-ex6, whereas *4.5SH* RNA alone

had no effect on the splicing of this minigene (Figure 7A, B). To further confirm the broad applicability of the chimeric molecule, we also generated minigenes containing exon 10 of *MAPT* [25], or exon 53 of human *DMD* (Figure 7A), which are the target exons for antisense oligonucleotide-mediated exon skipping therapies [26,27]. While efficient exon skipping was also induced with the *MAPT* chimera (SL1-4.5SH-*asMAPT*) (Figure 7C), none of the *DMD* chimera could induce skipping of the target exon when used alone, presumably due to the larger size of the exon or the strength of the 3' splice site. We thus tried different combinations of *DMD* chimeras and the skipped product was detected when we simultaneously co-introduced SL1-4.5SH-*asDMD2* and SL1-4.5SH-*asDMD4*, although the PCR cycles used were too high to quantitatively detect abundant non-skipped products (Figure 7E). We noticed that the region targeted by SL1-4.5SH-*asDMD2* was partially overlapped with the target of Viltolarsen (vltr), an antisense morpholinos oligonucleotide therapeutic used to treat *DMD* patients [28]. Therefore, we prepared another chimera, SL1-4.5SH-*asDMDvltr*, which covered the entire region targeted by Viltolarsen. The replacement of *asDMD2* with *asDMDvltr* further improved the exon skipping activity of the combinational transfection, and the highest efficiency was obtained when we used a chimera with a hybrid target recognition module consisting of *asDMDvltr* and *asDMD4* (SL1-4.5SH-*asDMDvltr4,* Figure 7F). The same strategy was successfully used to generate a chimera RNA (SL1-4.5SH-*asDmd*) that efficiently induces the skipping of exon 23 of mouse *Dmd* (Figure 7A, F).

To explore whether the recruitment of the effector complex through SL1-4.5SH is critical for exon skipping in these chimera molecules, we substituted SL1-4.5SH with SL1-B1 (SL1-B1-*asFAS*, SL1-B1-*asMAPT*, SL1-B1-*asDmd*, and SL1-B1-*asDMDvltr4*). A pronounced reduction in exon skipping activity was observed for *FAS*, *Dmd*, and *DMD* by the replacement (Figure 7B, D, F), while a more modest decline was noted for *MAPT* (Figure 7C). These results indicate that the exon skipping of these synthetic chimera RNAs operates through diverse mechanisms, depending on the target exon. For certain exons, simple base-pairing with the chimera RNAs doesn't suffice, necessitating the engagement of splicing inhibitors recruited through SL1 4.5SH. Conversely, some exons are effectively skipped solely by base-pairing, possibly due to the masking of exonic splicing enhancers that work in tandem with weaker 5′ and 3′ splice sites.

Finally, we compared the exon skipping activity of SL1-4.5SH-*asDMDvltr4* with Viltolarsen using the reporter system. For this experiment, we used HeLa cells, which show

efficient transfection efficiency for both plasmid and morpholinos oligonucleotides. The exon skipping activity of this hybrid chimera molecule was similar to 10 μM Viltolarsen, a concentration used for the gene therapy (Figure 7G).

Discussion

In this study, we have demonstrated that *4.5SH* RNA acts as a "molecular antidote" to toxic exonizations of *asB1*, protecting the integrity of the transcriptome from invasion of *SINE B1* during the evolution of small rodents. HNRNPC has previously been shown to inhibit the exonizations of the *Alu* elements, a human homolog of *SINE B1*, by competing with U2AF2 for polypyrimidine (poly-U) tract sequences [4]. Since the depletion of HNRNPC did not affect the exon-skipping activity of *4.5SH* RNA exogenously expressed in human cells, rodent and primate species seem to have evolved unique systems to protect their transcriptome from toxic antisense insertions of *SINE B1* family retrotransposons, mediated by RNA and protein, respectively. It should be noted that, in primate genomes, the *Alu*-derived *BC200* gene is abundantly expressed in the brain [29]. Yet, *BC200* transcripts predominantly reside in the cytoplasm, inherently precluding their role in exon-skipping. While many *SINE*-derived noncoding RNAs are abundantly expressed, the acquisition of nuclear localization may have been a crucial step for the emergence of *4.5SH* RNA as a splicing regulator within the lineage of small rodents.

HNRNPC recognizes bipartite poly-U sequences of the antisense *Alu* elements, which may not work efficiently in mice because *asB1*s only have a single poly-U stretch. We propose a model that the introduction of point mutations in the SL1 region of the 3' truncated *SINE* B1 lacking the A-stretch led to the emergence of non-transposable, Hnrnpm-binding nuclear *4.5SH* RNA, which counteracted toxic exonizations of *asB1*s. Subsequently, genomic fragments containing this advantageous noncoding RNA were amplified to form a large tandem gene cluster, and abundant *4.5SH* RNA transcribed from this cluster efficiently blocked inclusion of toxic *asB1*-containing exons, which became indispensable in small rodents such as mice and rats. It should be noted that point mutations in the poly-U stretch of *Alu*-containing cryptic exons result in the inclusions of normally skipped exons [4]. Considering that the target recognition module of *4.5SH* RNA can tolerate mismatch mutations, *4.5SH* RNA-dependent skipping of *asB1*-containing exons seems to be more robust than HNRNPC-dependent skipping of *Alu*-containing exons. In other words, the antisense insertions of *Alu* are more likely to be exonized if they carry mutations in the poly-U stretch to escape the recognition by HNRNPC. Consistently, *Alu*-containing exons in humans are more abundant in RefSeq annotated genes than *SINE B1* containing exons in mice, even when normalized by the number of insertions [3]. We speculate that, during their evolution, primates have

increased the diversity of protein products by utilizing *Alu*-containing exons, a system that did not work in small rodents due to the presence of *4.5SH* RNA that tolerates mutations in target *asB1*-containing exons.

The target recognition module of *4.5SH* RNA is programmable, and chimeric molecules consisting of SL1-4.5SH and designed target recognition sequences can induce skipping of exons of interest. Modified Antisense oligonucleotides (ASOs) are emerging as promising therapeutics to treat diseases with mutations that can be functionally rescued by inducing exon skipping [30-32]. The exon skipping activity of the *4.5SH* chimeric RNA was comparable to that of ASO, at least when assayed using our reporter system. A limitation of ASO-mediated therapies is the repeated administration of high doses of oligonucleotides, which can be a significant burden for patients. To overcome this, Adeno-Associated Virus vector-based expression of antisense sequences fused to modified U7 has been successfully used in cellular and animal models to restore DMD protein expression [33]. However, it has been reported that overexpression of the U7 chimeric molecule results in aberrant processing of U7 snRNA by interfering with endogenous processes [34]. Considering that *4.5SH* RNA is a small rodent-specific noncoding RNA, and thus the expression of *4.5SH* chimeric RNA does not directly interfere with endogenous pathways, our method may provide an alternative approach for vector-based gene therapy.

While therapeutic ASOs base-pair with regulatory sequences to induce exon skipping, two observations suggest that *4.5SH* RNA inhibits exon inclusion via different molecular mechanisms. First, replacement of SL1-4.5SH by SL1-B1 markedly reduced exon skipping activity, despite increased complementarity to the target sequences. Second, the introduction of perfect-match mutations in the target recognition modules reduced the exon skipping activity of *4.5SH* RNA. Thus, simple base-pairing is not sufficient for *4.5SH* RNA-mediated exon skipping, and additional effector proteins are required for its full activity; this mechanism is reminiscent of the artificial bifunctional antisense oligonucleotides containing cis-elements that recruit trans-acting splicing regulators [35,36]. We found that at least three effector RNA binding proteins, Hnrnpm, Sfpq, and Nono, associate specifically with SL1-4.5SH but not with non-functional SL1-B1. Hnrnpm has previously been shown to associate with Sfpq/Nono and induce skipping of certain exons [37]. In addition, genome-wide analyses have revealed that Hnrnpm-associated exons are preferentially included upon depletion of this protein, functioning as a negative regulator of splicing [19,38]. Considering that the most representative Hnrnpm-binding motif GUGGUGG [19] is located in the

distal region of SL1-4.5SH, Hnrnpm may be a primary effector protein that binds to *4.5SH* RNA, which subsequently recruits additional effector complexes including Sfpq and Nono to induce exon skipping. Given that both Sfpq and Nono bind intronic regions of pre-mRNAs [39,40], formation of the *4.5SH* RNA-protein complex may recruit cryptic *asB1*-containing exons to intron-enriched nuclear compartments [41], which are enriched in Sfpq and Nono. Alternatively, the formation of a large ribonucleoprotein complex on target exons physically interferes with 3' splice site recognition by steric obstruction. To clarify the precise molecular mechanisms, it would be intriguing to perform cryo-EM observation of *in vitro* pre-mRNA processing in the presence or absence of *4.5SH* RNA, an approach that has been successfully used to reveal the stepwise formations and reactions of spliceosome [42].

Hnrnpm is a highly conserved RNA binding protein [43] and associates with various types of cellular RNA including long noncoding RNAs (lncRNAs) [38]. A huge number of lncRNAs are transcribed from the genomes of higher eukaryotes, and a previous bioinformatic approach has identified a number of potential lncRNA-mRNA interactions [44]. As such, it is possible that there exist more endogenous lncRNAs that recruit Hnrnpm to target exons and inhibit their inclusion. Consistent with this hypothesis, a lncRNA called *PLANE* was recently shown to associate with HNRNPM and regulate alternative splicing of *NCOR* [45]. Since *PLANE* binds to the downstream intronic region of the target exon, it may use a different molecular mechanism distinct from *4.5SH* RNA that recognizes the exon sequence and inhibits the use of upstream 3' splice site. Whichever the case is, control of the alternative splicing could be one of the major functions of lncRNAs, providing another layer of regulation to increase the complexity of the transcriptome.

Limitation of the study

While our study provides novel insights into the physiological functions of *4.5SH* RNA and its role as an molecular antidote for toxic exonization of *SINE B1*, there are several limitations. Our study predominantly focuses on small rodents, specifically mice, and the results may not be directly applicable to other species, as demonstrated for human *Alu*. In addition, the mechanistic understanding of *4.5SH* RNA's role in exon skipping remains incomplete. While we propose a model for *4.5SH* RNA's activity based on its interactions with Hnrnpm, Sfpq, and Nono, the precise molecular mechanisms are yet to be elucidated fully. Future studies, such as

cryo-EM observation, might provide more comprehensive insights. While the target recognition module of 4.5SH RNA has been shown to be programmable, the broader applications, especially in therapeutic contexts, need rigorous validation in more diverse biological systems and under various physiological conditions. Lastly, while our study hints at the potential existence of other lncRNAs that might regulate splicing similarly to 4.5SH RNA, direct evidence is lacking. Comprehensive screening and functional assays would be necessary to uncover other lncRNAs with similar regulatory roles.

Acknowledgments

We thank Dr. Takashi Sado for the anti-SMCHD1 antibody, Dr. Archa Fox for the anti-Nono antibody, Drs. Akila Mayeda and Ryo Inoue for careful reading of the manuscript, Editage (www.editage.com) for English language editing, as well as QB3 Genomics at UC Berkeley for RNA-Seq. This work was supported by JSPS KAKENHI Grant Number JP 16H06276 (AdAMS), 21H05274 and 21K19246 granted to SN, 22K05565 granted to RY, 21H00253 and 22H02545 granted to TY, as well as JP20H05784 granted to SI, and AMED Grant Number JP23ym0126801 granted to SN. This work was partly supported by Global Facility Center (GFC), Hokkaido University.

Author contributions

SN and RY designed and performed experiments, MK contributed quantification of *4.5SH* and ES cultures, YN and IY performed experiments using the ES cells and mice, IN performed experiments using deletion mutants and Viltolarsen KA created *4.5SH* mutant animals and established the ES cells, YN optimized Dmd chimera RNA, ST contributed to western and northern blot analyses, YS and TK contributed to pulse-field gel electrophoresis experiments, HKH and MO contributed to mass-spectrometry analyses, MM and SI contributed preparation of RNA sequencing library, TY and TH contributed preparation of MS2CP-tagged RBPs, SN wrote the manuscript, all the authors edited the manuscript. Correspondence should be addressed to SN for *4.5SH* KO mice and other materials except for splicing minigenes and chimeric molecules, which should be requested to RY.

Declaration of interests

Patent applications (JP21663/2022-076674 and PCT/JP2023/015828) have been submitted by RY and SN based on these results.

Inclusion and Diversity

We support inclusive, diverse, and equitable conduct of research.

Figure legends

Figure 1. *4.5SH* KO mice exhibit embryonic lethality

(A) Genomic organization of the *4.5SH* cluster on chromosome 6 and a schematic of the targeting strategy for the generation of *4.5SH* KO mice. (B) Southern blot analysis of *4.5SH* using conventional (left) and pulse-field gel electrophoresis (right). (C) Genotypes of offspring obtained by crossing heterozygous (+/-) mice. We could not explicitly distinguish heterozygotes from wild type embryos at E2.5-3.5 because genotyping was judged by the expression of *4.5SH* RNA detected with *in situ* hybridization. (D) Nomarski DIC image of WT and *4.5SH* KO at the 8-cell stage embryos overlayed with *4.5SH* RNA signals detected with the LNA probe (green). (E) Transverse sections of E6.5 embryos of WT and *4.5SH* KO mice visualized with *4.5SH* probe (green) and DAPI (magenta) at lower (top panels) and higher (bottom panels) magnifications. The dotted lines in the upper panels show the contours of the embryonic tissues. The dotted lines in the lower panels indicate the cell nuclei in the embryonic tissues. Asterisks indicate non-specific autofluorescence signals in the cytoplasm, as judged by the signals detected by multiple non-excitation channels. (F) RT-qPCR analyses of *4.5SH* expression in WT, KO, and rescued (res) ES cells. (G) Alkaline phosphatase activity (AP) and expression of Nanog in WT, *4.5SH* KO, and rescued (res) ES cells. (H) Proliferation of WT, *4.5SH* KO, and rescued (res) ES cells. Error bars represent standard deviations from biological duplicates. Scale Bar, 100 μm and 20 μm in upper and lower panels in (E), and 200 μm and 10 μm in upper and lower panels in (G), respectively.

Figure 2. Deleterious exons are abnormally included in *4.5SH* KO ES cells

(A) Simultaneous detection of *4.5SH* RNA (green) and nuclear speckle marker Srsf1 (magenta) in ES cells. (B) Volcano plot of differentially spliced RefSeq-annotated (blue) and unannotated novel (magenta) exons, and schematic overview of the KO-enriched exons. Exons with FDR=0 are shown as outliers on the same infinite (∞) Y-value in the volcano plot. (C) Motifs enriched in KO-specific cryptic exons revealed via MEME analysis. (D) Relative positions of consensus motifs to *asB1* and *4.5SH* RNA. (E) Volcano plots illustrating the inclusion differences for specific exons. Magenta dots indicate exons with retrotransposon insertions: *sB1* and *asB1* for

sense and antisense insertions of *SINE B1*, respectively; *sB2* and *asB2* for sense and antisense insertions of *SINE B2*, respectively. Exons with FDR=0 are shown as outliers on the same infinite (∞) Y-value in the volcano plot. (F) Genome browser view of mapped sequence reads and gene annotations. Magenta boxes indicate KO-specific RefSeq-unannotated cryptic exons, and blue bars indicate the positions of *asB1*. (G) Gel-like view of capillary electrophoresis of RT-PCR products using two independent WT, KO, and rescued (res) ES cells. (H) Quantification of data shown in (G). Each dot represents the percentage of exon-skipped transcripts in each ES cell line. (I) Western blot analyses of Smg6 and Smchd1 in WT, KO, and rescued (res) ES cells. The numbers indicate quantification of the relative intensity of the bands. The same loading control images of Gapdh was used for Smg6 and Smchd1. Scale Bar, 5 μm.

Figure 3. *4.5SH* RNA base-pairs with target *asB1*-containing exons and inhibits the use of the 3' splice site

(A) Schematic overview of the minigene including the cryptic exon of *Slc25a40* (*asB1-Slc25a40*) and a *4.5SH* expression vector driven by U6 promoter (U6 pro). (B) Gel-like view of capillary electrophoresis of *asB1-Slc25a40* alternative splicing, northern blot analysis of *4.5SH* RNA expression, and quantification of the percentage of skipped products. Asterisk represents the position of the heteroduplex bands that appear on the Bioanalyzer. (C) Reciprocal mutations used in D. (D) The alternative splicing of *asB1-Slc25a40* in minigene-transfected HEK293 cells expressing wild-type or mutated *4.5SH*/*asB1-Slc25a40* and quantification of skipped products. Asterisk represents the position of the heteroduplex bands that appear on the Bioanalyzer. (E) Schematic overview of split minigenes containing either the 3' or 5' splice sites (SS) of *asB1-Slc25a40*. (F) The splicing of *asB1-Slc25a40* in HEK293 cells transfected with the split minigenes upon expression of *4.5SH* RNA. Error bars in B, D, and F represent standard deviations from biological triplicates.

Figure 4. Modular organization of *4.5SH* RNA consisting of the effector binding and the

target recognition modules

(A) Predicted secondary structure of 4.5SH RNA. (B) Schematic of the modular structure of *4.5SH* RNA and regions of the deletion mutants. Magenta bars indicate the regions complementary to MEME-identified motif1 and motif3 of the target *asB1*-containing exons. Green bars indicate the regions of the stem loop structures. The cyan bar indicates the region where nucleotides are substituted in the reciprocal mutants shown in Figure 3C. (C) Northern blot analyses of the expression of *4.5SH* deletion mutants and the effects of the deletion mutation on the exon skipping activity of *4.5SH* RNA, examined by minigene transfected cells. (D) Summary of the results shown in C. Yellow circles represent the residues of *4.5SH* RNA, of which deletion resulted in reduced (<50%) exon skipping activity. Magenta characters represents the residues, of which deletion resulted in downregulation of the expression. (E) Predicted secondary structures formed between *4.5SH* RNA and the motif3-lacking cryptic exons of *Ppil4* and *Eya2*. (F) Sequences and predicted secondary structures of the 5' stem-loop of *4.5SH* (SL1-4.5SH) and *SINE B1* (SL1-B1). (G) Schematic of the minigene, including the cryptic exon of *Slc25a40* (*asB1-Slc25a40*) and U6-driven *4.5SH* RNA and chimeric RNA containing SL1-B1. (H) The alternative splicing of *asB1-Slc25a40* in minigene-transfected HEK293 cells expressing *4.5SH* RNA or SL1-B1-*4.5SH*. Asterisk represents the position of the heteroduplex bands that appear on the Bioanalyzer. (I) Predicted secondary structures formed between the *asB1-Slc25a40* and *4.5SH* RNA/SL1-B1-*4.5SH*. Note that SL1-B1-*4.5SH* exhibits higher complementarity compared to the wildtype *4.5SH* RNA, in spite of the reduced exon skipping activity. (J) The alternative splicing of *asB1-Slc25a40* in minigene-transfected HEK293 cells expressing the wildtype or perfect-match (pm) *4.5SH* RNA. Asterisk represents the position of the heteroduplex bands that appear on the Bioanalyzer. Error bars represent standard deviations from biological duplicates in C, and from biological triplicates in H and J.

Figure 5. The 5' stem loop of *4.5SH* RNA recruits effector proteins to inhibit the inclusion of target exons

(A) Schematic of the experimental design to identify effector proteins. (B, C) Silver staining of SDS-PAGE-separated proteins pulled down with the biotinylated 5' stem-loop 1 of SINE B1 (SL1-B1) and the 5' stem-loop 1 of 4.5SH (SL1-4.5SH) under normal salt concentration (150 mM NaCl)

(B) and after the high-salt washing with 1M NaCl (C). (D) Scatter plot of the results obtained by LC-MS analyses of proteins purified with SL1-B1 and SL1-4.5SH (n =1 for each). Each dot represents identified protein. PSM: peptide-spectrum match score. (E) Preferential association of Hnrnpm, Sfpq, and Nono with SL1-4.5SH detected via western blot. Note that 75-kDa Hnrnpm remains bound to SL1-4.5SH after stringent washing with 1M NaCl. (F) The presence of previously identified consensus Hnrnpm-binding sequences in SL1-4.5SH. Coral color represents the putative Hnrnpm-binding site identified in a previous study [19]. (G) CLIP analyses of *4.5SH* RNA. FLAG-tagged Hnrnpm, Nono, and Sfpq were co-expressed with wild-type *4.5SH* RNA or chimeric SL1-B1-*4.5SH* in HEK293 cells and immunoprecipitated using anti-FLAG antibody. Immunoprecipitated *4.5SH* was detected via northern blot analysis. Note that these RNA-binding proteins are associated with *4.5SH* RNA, but not with SL1-B1-*4.5SH*, in a UV-irradiation dependent manner. (H) Simultaneous detection of *4.5SH* RNA (green) and effector proteins (magenta) Hnrnpm, Sfpq, and Nono in NIH3T3 cells and the color profile plots. Yellow lines represent the regions shown in the color profile plots. Arrowheads represent the positions of overlapping peaks. Scale Bars, 5 μm.

Figure 6. Functional involvement of SL1-4.5SH-associated effector proteins in *4.5SH* RNA-dependent exon skipping

(A) Alternative splicing of *asB1-Slc25a40* in minigene-transfected HEK293 cells depleted for HNRNPM, SFPQ, and NONO. Knockdown of HNRNPM and SFPQ resulted in reduced exon skipping. Large asterisks indicate statistically significant changes compared to the control ($p < 0.016$, Welch''s t-test after Bonferroni correction), and small asterisk represents the position of the heteroduplex bands that appear on the Bioanalyzer. (B, C) Schematic of RNA-binding protein recruitment to the exon (B) and the downstream intron (C) of the minigene using the MS2 system, and the alternative splicing of *asB1-Slc25a40* in HEK293 cells expressing RNA-binding proteins fused to the MS2 coat protein (MS2CP). Asterisks indicate statistically significant changes compared to the control ($p < 0.0125$, Welch's t-test after Bonferroni correction). (D) Alternative splicing of *asB1-Slc25a40* in HNRNPC-depleted HEK293 cells transfected with the minigene and Western blot analyses confirming the reduction of HNRNPC upon knockdown (KD). Notably, *4.5SH* RNA-mediated exon-skipping of *asB1-Slc25a40* persisted even in the reduction of

HNRNPC, whereas the HNRNPC depletion led to the inclusions of *asAlu*-containing endogenous cryptic exons from *HELLS* and *KIAA1432*, as delineated in previous studies [4]. The asterisks pinpoint heteroduplex band locations evident on the Bioanalyzer. (E) The alternative splicing of endogenous *asB1*-containing exons in murine Neuro2A cells depleted with Hnrnpc and Western blot analyses confirming the reduction of Hnrnpc upon knockdown (KD). Note that the splicing of *asB1*-containing exons were not affected, whereas skipping of *asAlu*-containing exon from *CD55* minigene was reduced upon depletion of Hnrnpc. The error bars in A-E depict the standard deviation from biological triplicates.

Figure 7. Chimeric *4.5SH* RNAs function as designable splicing regulators

(A) Schematic showing the regions used to make chimeric molecules and the strategy used to evaluate the exon skipping activity. Cyan bars indicate the region targeted by the chimeric *4.5SH* RNAs consisting of SL1-4.5SH and complementary target recognition sequences. Minigenes containing the target exons and flanking intron/exon sequences were co-introduced into HEK293 cells together with vectors expressing the chimeric molecules under the control of U6 promoter. (B, C, F) Alternative splicing of *FAS-ex6* (B), *MAPT-ex10* (C), and *Dmd-ex23* (F) in HEK293 cells transfected with splicing minigenes and vectors expressing *4.5SH*-chimera RNAs. The expression of the chimera RNAs were verified via Northern blot. Note that the exon skipping efficiencies substantially decreased when SL1-4.5SH was replaced with SL1-B1 for *FAS* and *Dmd*. In contrast, only a modest reduction was observed for *MAPT*. Asterisk represents the position of the heteroduplex bands that appear on the Bioanalyzer. (D) Alternative splicing of *DMD-ex53* in HEK293 cells transfected with the splicing minigene and vectors expressing different combinations of *4.5SH-DMD* chimeras. The PCR condition used in this experiment was too large to quantitatively detect the amount of non-skipped product. (E) Quantification of the alternative splicing of *DMD-ex53* in HEK293 cells transfected with the splicing minigene and vectors expressing various *4.5SH-DMD* chimeras. Note that the hybrid chimera SL1-4.5SH-*DMDvltr*-4 induced efficient exon skipping. The expression of the chimera RNAs were verified via Northern blot. (G) Quantification of the alternative splicing of *DMD-ex53* in HeLa cells transfected with the splicing minigene and vectors expressing SL1-4.5SH-*DMDvltr*-4 and different concentrations of Viltolarsen. Note that the skipping efficiency of the chimeric RNA is comparable to that of the

highest concentration of Viltolarsen. The expression of SL1-4.5SH-*DMDvltr*-4 was confirmed via Northern blot. Error bars represent standard deviations from biological triplicates in B-D, and F, and from biological duplicates in G.

STAR Methods

RESOURCE AVAILABILITY

Lead contact

Requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Shinichi Nakagawa (nakagawas@pharm.hokudai.ac.jp).

Materials availability

All the materials generated in this study are accessible upon request.

Data and Code Availability

- RNA-seq raw data is available at PRJDB13752. Raw image data is available at DOI: 10.17632/265cd268pz.1
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We utilized the 4.5SH KO mouse model developed in this study. C57BL/6N mice served as the background strain for extensive backcrossing. All procedures were conducted in accordance with the guidelines approved by the Safety Division of Hokkaido University (Approval No. 2015-079). For anesthesia, medetomidine-midazolam-butorphanol [46] was intraperitoneally injected at a volume of 10 µl/g of body weight. Mice are extensively backcrossed with C57BL6/N. For Mendelian ratio assessments, we included both male and female adult mice aged over 8 weeks. Embryos were collected for analysis; however, their sex was not determined.

METHOD DETAILS

Gel electrophoresis and southern blot analysis

Pulsed-field gel electrophoresis was performed as previously described [47]. Mouse embryonic fibroblasts (MEFs) were harvested via trypsinization, washed twice with PBS, and 2×10^6 cells were suspended in 100 µl of PBS and pre-warmed at 48ºC. The cell suspensions were mixed with equal volumes of prewarmed 1.2% low melting agarose in PBS, and 100 µl of the mixture was loaded on a well to make a plug. Each plug was incubated in NDSK buffer (0.5M EDTA, 1% N-

laurylsarcosine, and 1 mg/ml Proteinase K (Thermo, #25530-015), pH 9.5) overnight at 50ºC and further incubated in fresh NDSK buffer for 10 h. After washing in TE containing 0.1 mM phenylmethanesulfonyl fluoride, one-third of each plug was equilibrated in 1× Cutsmart buffer (NEB) and treated with restriction enzymes overnight at 37ºC. The digested DNAs were run on a 1% agarose gel in TAE using CHEF-Mapper XA system (Bio-Rad) at 2V/cm with a pulse time of 5 min to 40 min for 92 h at 14°C (Angle: 106º) [48]. For conventional gel electrophoresis, 5 μg of DNAs were digested with restriction enzymes and separated on 0.7% agarose gel in TAE. After electrophoresis, gels were treated with 0.25M HCl for 30 min, denatured in 0.5M NaOH with 1.5M NaCl, neutralized in 0.5M Tris (pH 7.2) with 1.5M NaCl, and equilibrated in 20× SSC. DNAs were capillary blotted on Hybond-N+ (Sigma, #RPN203B) using a standard protocol and hybridized with LNA probes in DIG easy Hyb (Merck, #11603558001) at 60ºC following the manufacturer's instructions. After washing in 2× SSC followed by stringent washing in 0.2× SSC, hybridized probes were detected with an alkaline phosphatase-conjugated anti-DIG antibody (Merck, #11093274910) and CDP-Star (Merck, #11685627001) following the manufacturer's instructions. The chemiluminescent signals were detected using the Chemi Doc Touch imaging system (Biorad).

Construction of targeting vectors and gRNA-expressing vector

To generate the targeting vector, homologous arm sequences (left arm: chr6:47,648,963-47,650,909, right arm: chr6:47,778,356-47,780,356) together with loxP and restriction sites were synthesized and cloned into a plasmid by a company (GenScript), and the DNA fragments were subsequently cloned into the PGK-NeO-pA plasmid (Supplemental figure S1A, B). To enhance homologous recombination efficiency, gRNA sequences were designed against the terminal regions of the homologous arms (Supplemental figure S1A, B). To generate the gRNA-expressing plasmid, two sets of oligonucleotides shown in Supplemental Table T2 were annealed and subcloned into the BbsI and BsaI sites of pX330-B/B, as previously described [49].

Generation of *4.5SH* KO mice

6NK7 [50] is a feeder-dependent ES cell line established from C57BL/6N (Nippon Clea) and was grown as described previously [51]. 6NK7 ES cells (6×10^6 cells) were suspended in 0.8 ml of phosphate-buffered saline (PBS), co-electroporated with circular forms of 30 μg of the targeting

vector, 13 μg each of pX330 vectors, as well as 1.5 μg of pPGK-puro plasmid using Gene Pulser XcellTM (Bio-Rad) and then plated into three 10 cm plates. After 24 to 48 h, the cells were selected with 1.7 μg/ml puromycin and thereafter with 130 μg/ml of G418 for 1 week. On day 8, colonies were picked and stocked. G418-resistant ES clones were screened for homologous recombination with the *4.5SH* cluster and the absence of random integrations via PCR using primers indicated in Supplemental Table T2. Chimeric mice were produced through the aggregation of ES cells with eight-cell embryos of ICR mice (Nippon Clea), which were then mated with C57BL/6N females (Nippon Clea) to obtain F1 heterozygous mice. Germ line transmission was validated via PCR, and homologous recombination was further validated via Southern blot analyses using genomic DNA prepared from the tails of F1 animals. The PGK-NeO cassette was removed by electroporating Cre-recombinase to *in vitro*-fertilized eggs, which were transferred to recipient mice to obtain floxed animals. The primers used for genotyping *4.5SH* KO mice are indicated in Supplemental Table T2. The following PCR conditions were used for genotyping: one cycle at 94ºC for 2 min; 35 cycles at 94ºC for 30 sec, 57ºC for 30 sec, and 68ºC for 1 min; followed by 68ºC for 5 min.

In situ hybridization

To prepare tissue sections from E6.5 embryos, pregnant mice were perfused with 4% paraformaldehyde in HCMF (10 mM HEPES [pH 7.4], 137 mM NaCl, 5.4 mM KCl, 0.34 mM Na_2HPO_4, 5.6 mM glucose), and dissected decidua were further fixed overnight at 4ºC. After washing in PBS two times, the samples were cryo-protected in 30% sucrose in PBS and embedded in Tissue-Tek OCT compound (Sakura). For *in situ* hybridization of cultured cells, cells were cultured on a washed coverslip coated with PLL and then fixed in 4% paraformaldehyde in HCMF overnight at 4ºC. Probe preparation and *in situ* hybridization were performed as previously described [52]. Briefly, fixed tissue sections or cells were treated with HCl and Proteinase K, fixed in 4% PFA, and acetylated with acetic anhydride in triethanolamine-HCl buffer. The fixation by perfusion and the proteinase K treatment was essential for the detection of *4.5SH* RNA signals.

Establishment of *4.5SH* KO ES cells and cell culture

The establishment of ES cells was performed as previously described [53]. Fertilized eggs were obtained via *in vitro* fertilization using heterozygous females and a male. The embryos were

cultured for 4 days until expanded blastocysts in KSOM [54] medium. Blastocysts were plated individually on a 48-well plate coated with 0.1% gelatin in KSR-GMEM medium and allowed to hatch and grow. After 10 days in culture, the inner cell mass (ICM) outgrowth was plated onto a 24-well plate with a feeder layer, then expanded and stocked. Neuro2A, HEK293, and HEK293T cells were cultured in a 1:1 mixture of DMEM/HamF12 supplemented with 10% fetal bovine serum and penicillin/streptomycin.

Immunohistochemistry

Cells were cultured on PLL-coated coverslips, fixed in 4% PFA in HCMF for 10 min at room temperature, and permeabilized in 100% methanol for 5 min at -20ºC. Non-specific binding was blocked with 4% skim milk in TBST (TBS containing 0.05% Tween-20), followed by incubation with primary and secondary antibodies. The antibodies used were indicated in Supplemental Table T2. The fluorescent and DIC images were obtained using an epifluorescent microscope (Olympus, BX51) equipped with a CCD camera (DP74).

Construction of *4.5SH*-expressing lentiviruses

To generate 10× tandem *4.5SH* expression lentivirus, a gene fragment containing the *4.5SH* gene and flanking sequences required for the expression of this gene [55] were synthesized, harboring restriction sites required for subcloning (Supplemental Table T2). The gene fragment was digested with XhoI and BglII, then subcloned into the XhoI/BamHI sites of pGEX-4T. The resultant plasmid vector was digested with XhoI and BamHI, into which the second fragments were subcloned. This procedure was repeated until 10 tandem repeats were subcloned into the plasmid, resulting in the generation of pGEX4T-10×*4.5SH*. The DNA fragments containing 10× *4.5SH* were prepared by digesting pGEX-4T-10×*4.5SH* with SwaI and XhoI, which was subcloned into a CS-CDF-CG-PRE-derived vector [56] digested with MluI (blunted) and XhoI, resulting in the generation of pLenti-10×-4.5*SH*. Lentiviruses were prepared as per a previously described protocol [56]. Briefly, pLenti-10×*4.5SH* or pLenti-CA-EGFP were mixed with packaging plasmids and introduced into HEK293T cells using FuGENE HD Transfection Reagents (Promega) following the manufacturer's instructions. After 24 h, 10 μM of forskolin was added to the culture medium and further incubated for 48 h. Culture supernatants were collected, filtered through a 0.45 μm filter, and centrifuged at 20,000 g for 4 h. The pellets were re-suspended in 1/100 the volume of DMEM

containing 10% FBS, and aliquots were frozen at -80ºC. For infections, viruses were diluted in culture medium at 1× concentration and incubated with ES cells at 37ºC overnight.

RT-PCR for the quantitation of *4.5SH* expression

cDNA of *4.5SH* was synthesized from 0.5 μg of total RNA and 0.1 μg of nuclear RNA using ReverTra Ace (100U/μL) (TOYOBO, #TRT-101) according to the manufacturer's instructions and using primers indicated in Supplemental Table T2. Synthesized cDNAs were diluted with 4 volumes of distilled water and used as templates for subsequent qPCR. The qPCR reaction was performed on Bio-Rad CFX Connect using a THUNDERBIRD Next SYBR qPCR Mix (TOYOBO, #QPX-201) as per the manufacturer's instructions.

Cellular fractionation and RNA purification

The nuclear and cytoplasmic fractionation was performed as previously described [12]. Briefly, cells were washed twice with ice-cold HEPES-buffered saline (HBSS: 10 mM HEPES [pH 7.4], 137 mM NaCl, 5.4 mM KCl, 0.34 mM Na_2HPO_4, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 5.6 mM glucose) and suspended in CSKT (10 mM PIPES [pH 6.8], 100 mM NaCl, 300 mM Sucrose, 1 mM EGTA, 1mM $MgCl_2$, 1 mM DTT, 0.1% Triton-X100, 1× protease inhibitor (#03969-21, Nacalai)) for 5 min. The cell suspensions were centrifuged at 700 g for 5 min, followed by further centrifugation at 12,000 g for 5 min to collect supernatants, which were lysed in 2 volumes of Trizol-LS (Thermo Fisher Scientific, #10296010). The nuclear pellets were washed once with CSKT, centrifuged at 700 g for 1 min, and lysed in Trizol (Thermo Fisher Scientific, #15596026). To collect total RNAs, cells were washed once with HBSS and lysed in Trizol. The Trizol-lysed samples were heated at 50ºC for 5 min before the addition of chloroform to solubilize semi-extractive RNAs as previously described [57].

RNA-sequencing and data analysis

RNA-seq data is available from PRJDB13752. RNA-seq libraries were prepared with TruSeq Stranded Total RNA Library Prep Gold (Illumina, #20020598), according to the manufacturer's instructions and sequenced using HiSeq4000 (Illumina), generating single-end reads of 50 bp. Obtained fastq files were mapped on the mm10 genome using HISAT2 (v.2.2.1) [58]. Transcript assembly based on mapped sequence reads was performed using Stringtie (v2.1.7) [59]. Using the

obtained Stringtie GTF files as a reference, the BAM files were analyzed using rMATS (v4.4.1) [14] to examine the changes in alternative splicing isoforms. Significant changes in splicing events were defined as when the false discovery rate (FDR) was calculated at less than 0.001. Strengths of the 5′ and 3′ splice sites were calculated using MaxEntScan [60]. Branch point strength and PPT length were calculated by svm-bpfinder [61]. To identify enriched motifs, a fasta file containing the sequence information of exons preferentially included in KO cells was analyzed via MEME [15], with the following parameters: site distribution = any number of repetitions (anr), number of motifs = 6. Exons longer than 4 kb were removed from this analysis to avoid false discovery of repeat sequences. RNA secondary structures were predicted using RNAfold2 RNA-Fold [18].

Alternative splicing analysis via RT-PCR

Total RNA (100~500 ng) was reverse-transcribed using ReverTra Ace® qPCR RT Master Mix with gDNA Remover (Toyobo, #FSQ-301) according to the manufacturers' instructions. Obtained cDNAs were amplified using GoTaq Green Master Mix (Promega, #M712B) for 20~23 cycles and were analyzed using Bioanalyzer (DNA1000, Agilent). "Skip %" was calculated by dividing the concentration (nmol/l) of the skipped isoform by the overall (skipped and included) isoforms. All the primers used for RT-PCR were listed in Supplemental Table T2.

Western blot analysis

Cells or purified proteins were lysed in 1× Laemmli's SDS sample buffer and separated on 5–20% gradient gel (Funakoshi, #270M) using a standard SDS-PAGE protocol. Proteins were transferred to a PVDF membrane using Trans-Blot Turbo System (BioRad) and detected with ECL-prime (GE Healthcare, #RPN2232) following the manufacturers' instructions. The primary and secondary antibodies used to detect the proteins are indicated in Supplemental Table T2. The chemiluminescence signals were detected using Chemi-Doc touch (Biorad). The polyclonal antibody against SMCHD1 was a kind gift from Takashi Sado [62].

Construction of splicing minigene

DNA fragments derived either from *Slc25a40*, *CD55*, mouse *Dmd*, or human *DMD* were inserted into the Sac II site of the Adenovirus Major-late(AdV) minigene [63] using an Infusion HD cloning kit (Takara, #639648). Split minigenes derived from AdV-*Slc25a40* were constructed by inserting

PCR fragments into BamHI and XhoI sites of FLAG-pcDNA3 [63]. FAS minigene was constructed by inserting PCR fragment containing exon 5-7 into BamHI and XhoI sites of pcDNA3 using an Infusion HD cloning kit. *MAPT* minigene was constructed by inserting synthetic DNA fragment containing exon 9-11 [25] into BamHI and XhoI sites of pcDNA3 using an Infusion HD cloning kit. The sequences were listed in Supplemental Table T2. 293, 293T, or HeLa cells were co-transfected with the splicing minigene and expression vectors using FuGENE HD Transfection Reagents (Promega) following the manufacturer's instructions.

Northern blot analysis

Total RNAs (~0.5 μg) were separated with denaturing PAGE containing 8M Urea. RNAs were then transferred to Hybond-N+ (Sigma, #RPN203B) using semi-dry transfer platform with 300 mA constant current for 40 min. The membrane was cross-linked and hybridized with DIG-labelled DNA or RNA probes in PerfectHyb Plus (Merck, #H7033) at 60ºC (RNA probe) or 42ºC (DNA probe), following the manufacturer's instructions. After washing in 2× SSC followed by stringent washing in 0.2× SSC, hybridized probes were detected with an alkaline phosphatase-conjugated anti-DIG antibody (Merck, #11093274910) and CDP-Star (Merck, #11685627001) following the manufacturer's instructions. The chemiluminescent signals were detected using ImageQuant 800 (Cytiva).

siRNA knockdown

HEK293 cells were cultured in a 12-well plate and transfected with 50 pmol of each siRNA using Lipofectamine RNAiMAX (Thermo, #13778) according to the manufacturer's instructions. At 72 h post-transfection, total RNA was extracted and subjected to RT-PCR analysis.

RNA pull-down

Biotinylated RNA probes used for the RNA pull-down experiments are indicated in Supplemental Table T2. RNA probes were synthesized by IDT and resuspended in duplex buffer (IDT) at a concentration of 1 mM. Probes of 3 μl were mixed with 7 μl of the duplex buffer, denatured at 94ºC for 90 s, and refolded for 2 s at 85ºC for 60 cycles. Thirty μl of Dynabeads MyOne Streptoavidin T1 (Invitrogen, #65601) were washed twice with PBST (PBS, 0.1% Triton X100), resuspended in 20 μl of PBST, and incubated with the refolded probe for 2 h at 4ºC. The probe-

conjugated beads were washed twice with PBST and incubated with nuclear lysates for 3 h at 4ºC. To prepare the nuclear lysates, Neuro2A cells (~1×10^8) cultured in a 10 cm culture dish were washed twice with HBSS at 4ºC and treated with 1 ml of ice-cold CSKT for 5 min. Cell nuclei were collected with a scraper, transferred to 1.5 ml of low-protein bound tube (Sigma-Aldrich, #Z666548-250EA), centrifuged at 700 *g* for 1 min, and washed once with CSKT. The nuclei pellets were re-suspended in 1 ml of binding buffer (PBST, 1 mM DTT, 1× protease inhibitor), and sonicated 5 times for 5 s, with 10-s intervals at 4ºC using Handy Sonic (TOMY, #UR-21P). The lysates were centrifuged at 12,000 *g* for 10 min, and supernatants were transferred to a new tube. The lysates were pre-cleared with 30 µl of Dynabeads MyOne Streptoavidin T1 and 100 µl of Streptoavidin-conjugated Sepharose (Merck, #17-5113-01) for 2 h at 4ºC. Cleared lysates were mixed with the probe-conjugated beads after the addition of 5 µl RNase inhibitor (Takara, #2313A). Purified proteins on beads were then washed once with PBST with a protease inhibitor and 1 mM DTT as well as twice with PBST and 1 mM DTT prior to western blotting. For some experiments, beads were washed with PBST containing 1M NaCl. To prepare samples for LC-MS analyses, beads were further washed twice with PBS with 1 mM DTT.

CLIP analysis

HEK293 cells were washed twice with ice-cold PBS and immediately irradiated with 254 nm UV-light (Funakoshi, #FS-1500) at 120 mJ/cm^2 and a distance of ~5 cm. The cells were suspended with 200 µl of SDS buffer (50 mM Tris-HCl [pH 8.0], 1 mM EDTA, 150 mM NaCl, 1 mM DTT, 1% SDS, 1% Triton X-100) and sonicated (QSONICA, #Q125) at 20% amplitude for 5 s three times, with a 30-s incubation on ice between bursts, followed by the addition of 800 µl of dilution buffer (50 mM Tris-HCl [pH 8.0], 1 mM EDTA, 150 mM NaCl, 1 mM DTT, 1% Triton X-100, supplemented with 1× protease inhibitor cocktail (Merck, #P8340) and RNase inhibitor (Takara, #2313). Cells were then centrifuged at 13,500 rpm for 20 min at 4°C. For immunoprecipitation, 15 µl of anti-FLAG M2-Agarose beads (Merck, #A2220) were added to the extract and incubated for 2 h at 4°C with continuous rotation. The beads were washed six times with wash buffer (20 mM Tris-HCl [pH 8.0] 1 mM EDTA, 150 mM NaCl, 1 mM DTT, 0.1% SDS, 1% Triton X-100). The beads were suspended in Homomix (50 mM Tris-HCl [pH 7.4], 5 mM EDTA, 1.5% SDS, 300 mM NaCl, 1.5 mg/ml proteinase K) at 50ºC for 60 min and subjected to subsequent northern blot analyses.

LC-MS analysis

The affinity-purified proteins were reduced with 1 mM dithiothreitol for 90 min, alkylated with 5.5 mM iodoacetamide for 30 min, and digested with trypsin overnight at 37°C. The fragmented peptides were then desalted using ZipTip C18 (Millipore) and centrifuged in a vacuum concentrator. Shotgun proteomic analyses were performed on an Orbitrap Eclipse Tribrid mass spectrometer with FAIMS Pro interface (Thermo Fisher Scientific), which was connected to Ultimate 3000 RSLCnano pump (Thermo Fisher Scientific). The peptide samples were separated using a linear gradient of 2–24% mobile phase (0.1% formic acid in acetonitrile) at 300 nl/min. Full-scan MS spectra were acquired at an automatic gain control (AGC) target value of 4×10^5 with a resolution of 120,000. Subsequent MS/MS scans were performed in the ion trap mode using higher energy collision-induced dissociation (HCD) fragmentation with a normalized collision energy of 35% at an AGC target value of 1×10^4 with 10 ms maximum injection time. Protein identification was conducted by searching against the UniProt mouse reference proteome database (UP000000589) using Sequest HT algorithm in Proteome Discoverer Software (version 2.5) (Thermo Fisher Scientific). Cysteine carbamidomethylation was set as a fixed modification, whereas methionine oxidation and protein N-terminal acetylation were set as variable modifications. A maximum of two missed cleavages was allowed in our database search, while the mass tolerance was set to 10 ppm for peptide masses and 0.6 Da for MS/MS peaks, respectively. In the process of peptide identification, we applied a filter to satisfy a false positive rate <1%.

Constructions of MS2CP-tagged RBP expression vectors

To construct the MS2CP fusion protein expression vector (pLVSIN-CMV_Pur_FLAG-MS2CP), the 3xFLAG-MS2CP-$(GGGGS)_3$ linker fragment from pcDNA5/FRT/TO/i/FLAG/MCP [49] was inserted into the EcoRI site of pLVSIN-CMV_Pur vector (Takara). To generate FLAG-MS2CP-NONO, FLAG-MS2CP-SFPQ, and FLAG-MS2CP-HNRNPM expression vectors, the human NONO and SFPQ open reading frames were inserted using the XhoI site of pLVSIN-CMV_Pur_FLAG-MS2CP, and the human HNRNPM open reading frame was inserted using the XbaI site of pLVSIN-CMV_Pur_FLAG-MS2CP.

Transfection of Viltolarsen and splicing reporter minigene

Because the splicing reporter minigene and morpholinos oligonucleotides cannot be co-introduced at the same time, they were sequentially transfected. First, HeLa cells were transfected with a mixture of the splicing reporter minigene using FuGENE HD Transfection Reagents (Promega) following the manufacturer's instructions. After 24 hours, the cells were harvested and electroporated with Viltolarsen using Nucleofector. 5.0×10^5 HeLa cells were suspended in 100 µl Opti-MEM containing 1 µM, 2 µM, or 10 µM of Viltolarsen, and electroporation was performed using Nucleofector □ device with program I-013.

QUANTIFICATION AND STATISTICAL ANALYSIS

Error bars presented in the figures represent standard deviations derived from biological replicates. The specific number of replicates (duplicates or triplicates) for each experiment is detailed in the corresponding figure legends. For statistical evaluations, we performed Welch's t-tests and applied Bonferroni correction to adjust for multiple comparisons where applicable.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal antibody to Nono	A king gift from A. Fox	PMID: 20881053
Mouse monoclonal antibody to Sfpq	Abcam	#ab11825
Mouse monoclonal antibody to Hnrnpm	Santa Cruz	#sc-134360
Mouse monoclonal antibody to Srsf1	Invitrogen	#32-4600
Mouse monoclonal antibody to FLAG	FUJI FILM	#014-22383
Mouse monoclonal antibody to Nanog	Reprocell	#14-5761-80
Rabbit polyclonal antibody to Smchd1	A kind gift from T. Sado	PMID: 30126901
Rabbit polyclonal antibody to Smg6	Abcam	#ab87539
Mouse monoclonal antibody to hnRNP-C1/C2	Merck	#R5028
Mouse monoclonal antibody to Anti-alpha Tubulin	Abcam	#ab11304
Bacterial and virus strains		
Lentivirus expressing *4.5SH* RNA	This study	NA
Biological samples		
Mouse embryos from 4.5SH KO mouse	This study	maintained under C57BL6/N background
Chemicals, peptides, and recombinant proteins		
medetomidine	ZENOAQ	NA
midazolam	SANDOZ	NA
butorphanol	MEIJI	NA
Proteinase K	Thermo Fisher Scientific	#25530-015
Hybond®-N+ hybridization membranes	MERCK	GERPN203B
PerfectHyb Plus	MERCK	#H7033
DIG easy Hyb	MERCK	#11603558001
alkaline phosphatase-conjugated anti-DIG antibody	MERCK	#11093274910
CDP-Star	MERCK	#11685627001
CARD KSOM	KYUDO	NA
KnockOut™ Serum Replacement	Thermo Fisher Scientific	#10828010
GMEM	MERCK	#G5154
D-MEM / Ham's F-12	FUJI FILM	048-29785
ReverTra Ace	TOYOBO	#TRT-101
Protease Inhibitor Cocktail	NACALAI	#03969-21
Triton-X100	NACALAI	#9002-93-1
Tween-20	NACALAI	#9005-64-5

Trozol	Thermo Fisher Scientific	#15596026
Trizol-LS	Thermo Fisher Scientific	#10296010
Dynabeads MyOne Streptoavidin T1	Thermo Fisher Scientific	#65601
RNase inhibitor	TAKARA	#2313A
DTT	NACALAI	#14128-91
anti-FLAG M2-Agarose beads	MERCK	#A2220
Critical commercial assays		
THUNDERBIRD Next SYBR qPCR Mix	TOYOBO	#QPX-201
TruSeq Stranded Total RNA Library Prep Gold	ILLUMINA	#20020598
ReverTra Ace® qPCR RT Master Mix with gDNA Remover	TOYOBO	#FSQ-301
GoTaq Green Master Mix	PROMEGA	#M712B
DNA1000 Kit	Agilent	NA
ECL-prime	GE Healthcare	#RPN2232
Deposited data		
RNA-Seq data for ES cells derived from WT and *4.5SH* KO mice	This study	PRJDB13752
Original images	This study, Mendeley Data	10.17632/265cd268pz.1
Experimental models: Cell lines		
HEK293	ATCC	CRL-1573
NIH3T3	ATCC	CRL-1658
ES cells derived from WT embryos	This study	ES cells derived from WT embryos
ES cells derived from *4.5SH* KO mice	This study	ES cells derived from *4.5SH* KO mice
Neuro2A	ATCC	CCL-131
Experimental models: Organisms/strains		
Mouse embryos from 4.5SH KO mouse	This study	maintained under C57BL6/N background
Oligonucleotides		
Primers for qRT-PCR and vector constructions, as well as probes for RNA-FISH, see Supplemental Table T2	This study	NA
Recombinant DNA		

Software and algorithms		
HISAT2	http://daehwankimlab.github.io/hisat2/	v.2.2.1
Stringtie	https://ccb.jhu.edu/software/stringtie/	v2.1.7
rMATS	https://rnaseq-mats.sourceforge.net	v4.4.1
svm-bpfinder	https://github.com/ComputationalRNABiology/svm-bpfinder	NA
MaxEntScan	http://hollywood.mit.edu/burgelab/maxent/download/	NA
MEME	https://meme-suite.org/meme/	V5.4.1
RNAfold2	https://anaconda.org/bioconda/viennarna	V2.5.1
Other		

Supplemental table S1 rMATS output and annotation of the exons enriched in *4.5SH* KO ES cells (related to Figure 2)

Supplemental table S2 Primers and antibodies used in this study (related to STAR methods)

References

1. Kramerov, D.A., and Vassetzky, N.S. (2011). SINEs. Wiley Interdiscip Rev RNA *2*, 772-786. 10.1002/wrna.91.
2. Sorek, R. (2007). The birth of new exons: mechanisms and evolutionary consequences. RNA *13*, 1603-1608. 10.1261/rna.682507.
3. Sela, N., Mersch, B., Gal-Mark, N., Lev-Maor, G., Hotz-Wagenblatt, A., and Ast, G. (2007). Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. Genome Biol *8*, R127. 10.1186/gb-2007-8-6-r127.
4. Zarnack, K., Konig, J., Tajnik, M., Martincorena, I., Eustermann, S., Stevant, I., Reyes, A., Anders, S., Luscombe, N.M., and Ule, J. (2013). Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Cell *152*, 453-466. 10.1016/j.cell.2012.12.023.
5. Mattick, J.S. (2018). The State of Long Non-Coding RNA Biology. Noncoding RNA *4*. 10.3390/ncrna4030017.
6. Rinn, J.L., and Chang, H.Y. (2020). Long Noncoding RNAs: Molecular Modalities to Organismal Functions. Annu Rev Biochem *89*, 283-308. 10.1146/annurev-biochem-062917-012708.
7. Harada, F., and Kato, N. (1980). Nucleotide sequences of 4.5S RNAs associated with poly(A)-containing RNAs of mouse and hamster cells. Nucleic Acids Res *8*, 1273-1285.
8. Harada, F., Kato, N., and Hoshino, H. (1979). Series of 4.5S RNAs associated with poly(A)-containing RNAs of rodent cells. Nucleic Acids Res *7*, 909-917. 10.1093/nar/7.4.909.
9. Jelinek, W., and Leinwand, L. (1978). Low molecular weight RNAs hydrogen-bonded to nuclear and cytoplasmic poly(A)-terminated RNA from cultured Chinese hamster ovary cells. Cell *15*, 205-214.
10. Gogolevskaya, I.K., Koval, A.P., and Kramerov, D.A. (2005). Evolutionary history of 4.5SH RNA. Mol Biol Evol *22*, 1546-1554. 10.1093/molbev/msi140.
11. Schoeniger, L.O., and Jelinek, W.R. (1986). 4.5S RNA is encoded by hundreds of tandemly linked genes, has a short half-life, and is hydrogen bonded in vivo to poly(A)-terminated RNAs in the cytoplasm of cultured mouse cells. Mol Cell Biol *6*, 1508-1519.
12. Ishida, K., Miyauchi, K., Kimura, Y., Mito, M., Okada, S., Suzuki, T., and Nakagawa, S. (2015). Regulation of gene expression via retrotransposon insertions and the noncoding RNA 4.5S RNAH. Genes Cells *20*, 887-901. 10.1111/gtc.12280.
13. Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harb Perspect Biol *3*. 10.1101/cshperspect.a000646.
14. Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A *111*, E5593-5601. 10.1073/pnas.1419161111.
15. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res *43*, W39-49. 10.1093/nar/gkv416.

16. Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat Rev Mol Cell Biol *20*, 406-420. 10.1038/s41580-019-0126-2.
17. Zapp, M.L., and Berget, S.M. (1989). Evidence for nuclear factors involved in recognition of 5' splice sites. Nucleic Acids Res *17*, 2655-2674. 10.1093/nar/17.7.2655.
18. Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol *6*, 26. 10.1186/1748-7188-6-26.
19. Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep *1*, 167-178. 10.1016/j.celrep.2012.02.001.
20. Fox, A.H., Nakagawa, S., Hirose, T., and Bond, C.S. (2018). Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. Trends Biochem Sci *43*, 124-135. 10.1016/j.tibs.2017.12.001.
21. Chen, L.L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell *35*, 467-478. 10.1016/j.molcel.2009.06.027.
22. Passon, D.M., Lee, M., Rackham, O., Stanley, W.A., Sadowska, A., Filipovska, A., Fox, A.H., and Bond, C.S. (2012). Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. Proc Natl Acad Sci U S A *109*, 4846-4850. 10.1073/pnas.1120792109.
23. Bardwell, V.J., and Wickens, M. (1990). Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. Nucleic Acids Res *18*, 6587-6594. 10.1093/nar/18.22.6587.
24. Paronetto, M.P., Bernardis, I., Volpe, E., Bechara, E., Sebestyen, E., Eyras, E., and Valcarcel, J. (2014). Regulation of FAS exon definition and apoptosis by the Ewing sarcoma protein. Cell Rep *7*, 1211-1226. 10.1016/j.celrep.2014.03.077.
25. Yu, Q., Guo, J., and Zhou, J. (2004). A minimal length between tau exon 10 and 11 is required for correct splicing of exon 10. J Neurochem *90*, 164-172. 10.1111/j.1471-4159.2004.02477.x.
26. Sud, R., Geller, E.T., and Schellenberg, G.D. (2014). Antisense-mediated Exon Skipping Decreases Tau Protein Expression: A Potential Therapy For Tauopathies. Mol Ther Nucleic Acids *3*, e180. 10.1038/mtna.2014.30.
27. Komaki, H., Nagata, T., Saito, T., Masuda, S., Takeshita, E., Sasaki, M., Tachimori, H., Nakamura, H., Aoki, Y., and Takeda, S. (2018). Systemic administration of the antisense oligonucleotide NS-065/NCNP-01 for skipping of exon 53 in patients with Duchenne muscular dystrophy. Sci Transl Med *10*. 10.1126/scitranslmed.aan0713.
28. Roshmi, R.R., and Yokota, T. (2019). Viltolarsen for the treatment of Duchenne muscular dystrophy. Drugs Today (Barc) *55*, 627-639. 10.1358/dot.2019.55.10.3045038.
29. Tiedge, H., Chen, W., and Brosius, J. (1993). Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. J Neurosci *13*, 2382-2390. 10.1523/JNEUROSCI.13-06-02382.1993.
30. Matsuo, M. (2021). Antisense Oligonucleotide-Mediated Exon-skipping Therapies: Precision Medicine Spreading from Duchenne Muscular Dystrophy. JMA J *4*, 232-240. 10.31662/jmaj.2021-0019.

31. Li, D., Mastaglia, F.L., Fletcher, S., and Wilton, S.D. (2018). Precision Medicine through Antisense Oligonucleotide-Mediated Exon Skipping. Trends Pharmacol Sci *39*, 982-994. 10.1016/j.tips.2018.09.001.
32. Siva, K., Covello, G., and Denti, M.A. (2014). Exon-skipping antisense oligonucleotides to correct missplicing in neurogenetic diseases. Nucleic Acid Ther *24*, 69-86. 10.1089/nat.2013.0461.
33. Goyenvalle, A., Vulin, A., Fougerousse, F., Leturcq, F., Kaplan, J.C., Garcia, L., and Danos, O. (2004). Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. Science *306*, 1796-1799. 10.1126/science.1104297.
34. Eckenfelder, A., Tordo, J., Babbs, A., Davies, K.E., Goyenvalle, A., and Danos, O. (2012). The Cellular Processing Capacity Limits the Amounts of Chimeric U7 snRNA Available for Antisense Delivery. Mol Ther Nucleic Acids *1*, e31. 10.1038/mtna.2012.24.
35. Skordis, L.A., Dunckley, M.G., Yue, B., Eperon, I.C., and Muntoni, F. (2003). Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. Proc Natl Acad Sci U S A *100*, 4114-4119. 10.1073/pnas.0633863100.
36. Brosseau, J.P., Lucier, J.F., Lamarche, A.A., Shkreta, L., Gendron, D., Lapointe, E., Thibault, P., Paquet, E., Perreault, J.P., Abou Elela, S., and Chabot, B. (2014). Redirecting splicing with bifunctional oligonucleotides. Nucleic Acids Res *42*, e40. 10.1093/nar/gkt1287.
37. Marko, M., Leichter, M., Patrinou-Georgoula, M., and Guialis, A. (2010). hnRNP M interacts with PSF and p54(nrb) and co-localizes within defined nuclear structures. Exp Cell Res *316*, 390-400. 10.1016/j.yexcr.2009.10.021.
38. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature *583*, 711-719. 10.1038/s41586-020-2077-3.
39. Zhang, S., Cooper, J.A., Chong, Y.S., Naveed, A., Mayoh, C., Jayatilleke, N., Liu, T., Amos, S., Kobelke, S., Marshall, A.C., et al. (2022). NONO enhances mRNA processing of super-enhancer-associated GATA2 and HAND2 genes in neuroblastoma. EMBO Rep, e54977. 10.15252/embr.202254977.
40. Iida, K., Hagiwara, M., and Takeuchi, A. (2020). Multilateral Bioinformatics Analyses Reveal the Function-Oriented Target Specificities and Recognition of the RNA-Binding Protein SFPQ. iScience *23*, 101325. 10.1016/j.isci.2020.101325.
41. Liao, S.E., and Regev, O. (2021). Splicing at the phase-separated nuclear speckle interface: a model. Nucleic Acids Res *49*, 636-645. 10.1093/nar/gkaa1209.
42. Fica, S.M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. Nat Struct Mol Biol *24*, 791-799. 10.1038/nsmb.3463.
43. Datar, K.V., Dreyfuss, G., and Swanson, M.S. (1993). The human hnRNP M proteins: identification of a methionine/arginine-rich repeat motif in ribonucleoproteins. Nucleic Acids Res *21*, 439-446. 10.1093/nar/21.3.439.
44. Fukunaga, T., Iwakiri, J., Ono, Y., and Hamada, M. (2019). LncRRIsearch: A Web Server for lncRNA-RNA Interaction Prediction Integrated With Tissue-Specific Expression and Subcellular Localization Data. Front Genet *10*, 462. 10.3389/fgene.2019.00462.
45. Teng, L., Feng, Y.C., Guo, S.T., Wang, P.L., Qi, T.F., Yue, Y.M., Wang, S.X., Zhang, S.N., Tang, C.X., La, T., et al. (2021). The pan-cancer lncRNA PLANE regulates an alternative

splicing program to promote cancer pathogenesis. Nat Commun *12*, 3734. 10.1038/s41467-021-24099-4.

46. Kawai, S., Takagi, Y., Kaneko, S., and Kurosawa, T. (2011). Effect of three types of mixed anesthetic agents alternate to ketamine in mice. Exp Anim *60*, 481-487. 10.1538/expanim.60.481.
47. Kobayashi, T., Nomura, M., and Horiuchi, T. (2001). Identification of DNA cis elements essential for expansion of ribosomal DNA repeats in Saccharomyces cerevisiae. Mol Cell Biol *21*, 136-147. 10.1128/MCB.21.1.136-147.2001.
48. Stults, D.M., Killen, M.W., Pierce, H.H., and Pierce, A.J. (2008). Genomic architecture and inheritance of human ribosomal RNA gene clusters. Genome Res *18*, 13-18. 10.1101/gr.6858507.
49. Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G., and Hirose, T. (2018). Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. Mol Cell *70*, 1038-1053 e1037. 10.1016/j.molcel.2018.05.019.
50. Honda, I., Araki, K., Honda, S., Soeda, F., Shin, M.C., Misumi, S., Yamamura, K.I., and Takahama, K. (2018). Deletion of GIRK2 subunit containing GIRK channels of neurons expressing dopamine transporter decrease immobility time on forced swimming in mice. Neurosci Lett *665*, 140-146. 10.1016/j.neulet.2017.11.028.
51. Nakahara, M., Tateyama, H., Araki, M., Nakagata, N., Yamamura, K., and Araki, K. (2013). Gene-trap mutagenesis using Mol/MSM-1 embryonic stem cells from MSM/Ms mice. Mamm Genome *24*, 228-239. 10.1007/s00335-013-9452-4.
52. Mito, M., Kawaguchi, T., Hirose, T., and Nakagawa, S. (2016). Simultaneous multicolor detection of RNA and proteins using super-resolution microscopy. Methods *98*, 158-165. 10.1016/j.ymeth.2015.11.007.
53. Araki, K., Takeda, N., Yoshiki, A., Obata, Y., Nakagata, N., Shiroishi, T., Moriwaki, K., and Yamamura, K. (2009). Establishment of germline-competent embryonic stem cell lines from the MSM/Ms strain. Mamm Genome *20*, 14-20. 10.1007/s00335-008-9160-7.
54. Lawitts, J.A., and Biggers, J.D. (1993). Culture of preimplantation embryos. Methods Enzymol *225*, 153-164. 10.1016/0076-6879(93)25012-q.
55. Koval, A.P., and Kramerov, D.A. (2009). 5'-flanking sequences can dramatically influence 4.5SH RNA gene transcription by RNA-polymerase III. Gene *446*, 75-80. 10.1016/j.gene.2009.07.005.
56. Miyoshi, H., Blomer, U., Takahashi, M., Gage, F.H., and Verma, I.M. (1998). Development of a self-inactivating lentivirus vector. J Virol *72*, 8150-8157. 10.1128/JVI.72.10.8150-8157.1998.
57. Chujo, T., Yamazaki, T., Kawaguchi, T., Kurosaka, S., Takumi, T., Nakagawa, S., and Hirose, T. (2017). Unusual semi-extractability as a hallmark of nuclear body-associated architectural noncoding RNAs. EMBO J *36*, 1447-1462. 10.15252/embj.201695848.
58. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907-915. 10.1038/s41587-019-0201-4.
59. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc *11*, 1650-1667. 10.1038/nprot.2016.095.

60. Shamsani, J., Kazakoff, S.H., Armean, I.M., McLaren, W., Parsons, M.T., Thompson, B.A., O'Mara, T.A., Hunt, S.E., Waddell, N., and Spurdle, A.B. (2019). A plugin for the Ensembl Variant Effect Predictor that uses MaxEntScan to predict variant spliceogenicity. Bioinformatics *35*, 2315-2317. 10.1093/bioinformatics/bty960.
61. Corvelo, A., Hallegger, M., Smith, C.W., and Eyras, E. (2010). Genome-wide association between branch point properties and alternative splicing. PLoS Comput Biol *6*, e1001016. 10.1371/journal.pcbi.1001016.
62. Sakakibara, Y., Nagao, K., Blewitt, M., Sasaki, H., Obuse, C., and Sado, T. (2018). Role of SmcHD1 in establishment of epigenetic states required for the maintenance of the X-inactivated state in mice. Development *145*. 10.1242/dev.166462.
63. Yoshimoto, R., Kataoka, N., Okawa, K., and Ohno, M. (2009). Isolation and characterization of post-splicing lariat-intron complexes. Nucleic Acids Res *37*, 891-902. 10.1093/nar/gkn1002.

Figure 1

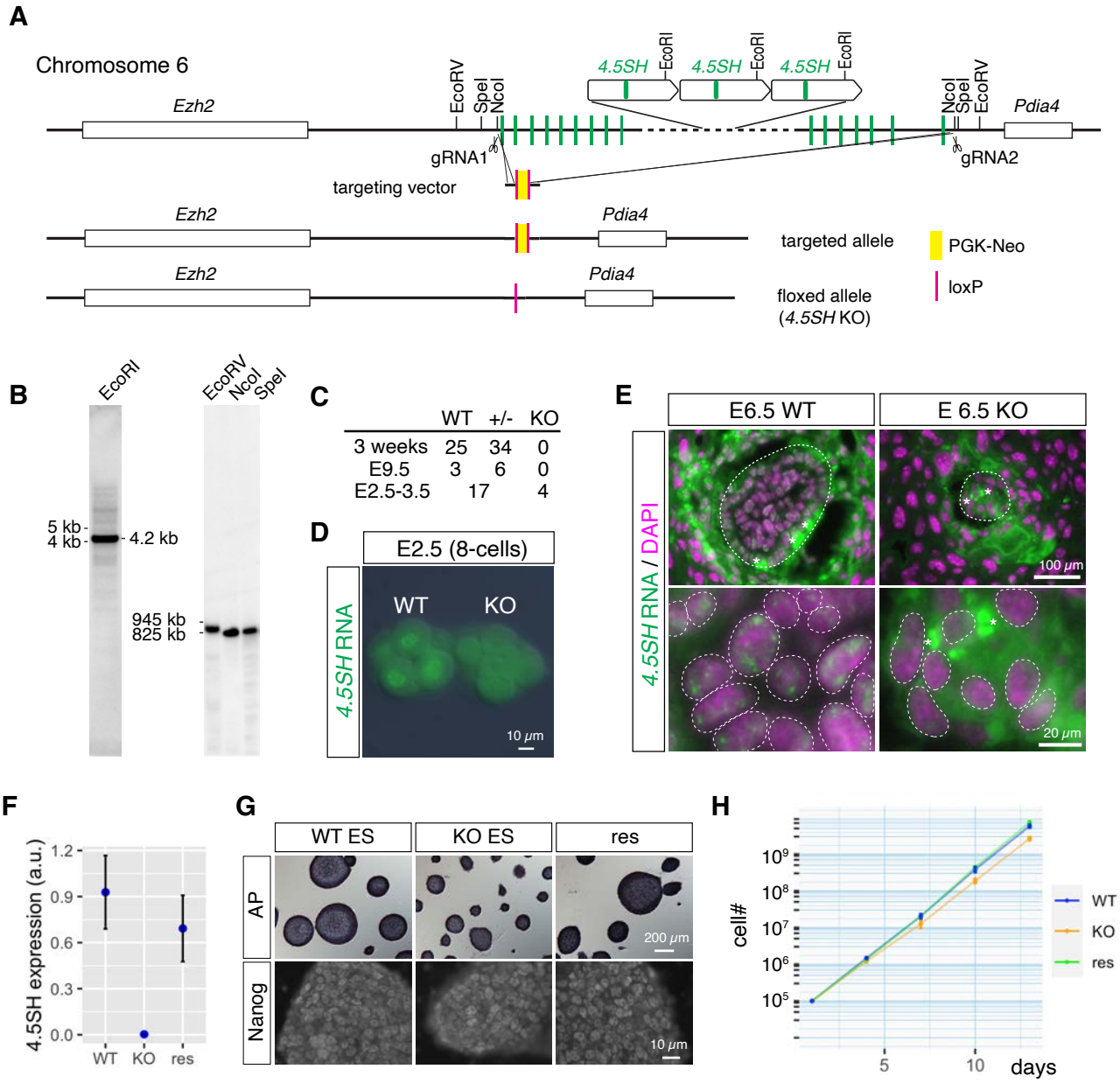

A
Chromosome 6
Ezh2
Pdia4
EcoRV
SpeI
NcoI
4.5SH
EcoRI
gRNA1
gRNA2
targeting vector
targeted allele
floxed allele
(4.5SH KO)
PGK-Neo
loxP
B
EcoRI
EcoRV
NcoI
SpeI
5 kb
4 kb
4.2 kb
945 kb
825 kb
C
WT +/- KO
3 weeks 25 34 0
E9.5 3 6 0
E2.5-3.5 17 4
D
E2.5 (8-cells)
4.5SH RNA
WT
KO
10 μm
E
E6.5 WT
E 6.5 KO
4.5SH RNA / DAPI
100 μm
20 μm
F
4.5SH expression (a.u.)
1.2
0.9
0.6
0.3
0.0
WT KO res
G
WT ES
KO ES
res
AP
Nanog
200 μm
10 μm
H
cell#
10^9
10^8
10^7
10^6
10^5
5
10
days
WT
KO
res

Figure 2

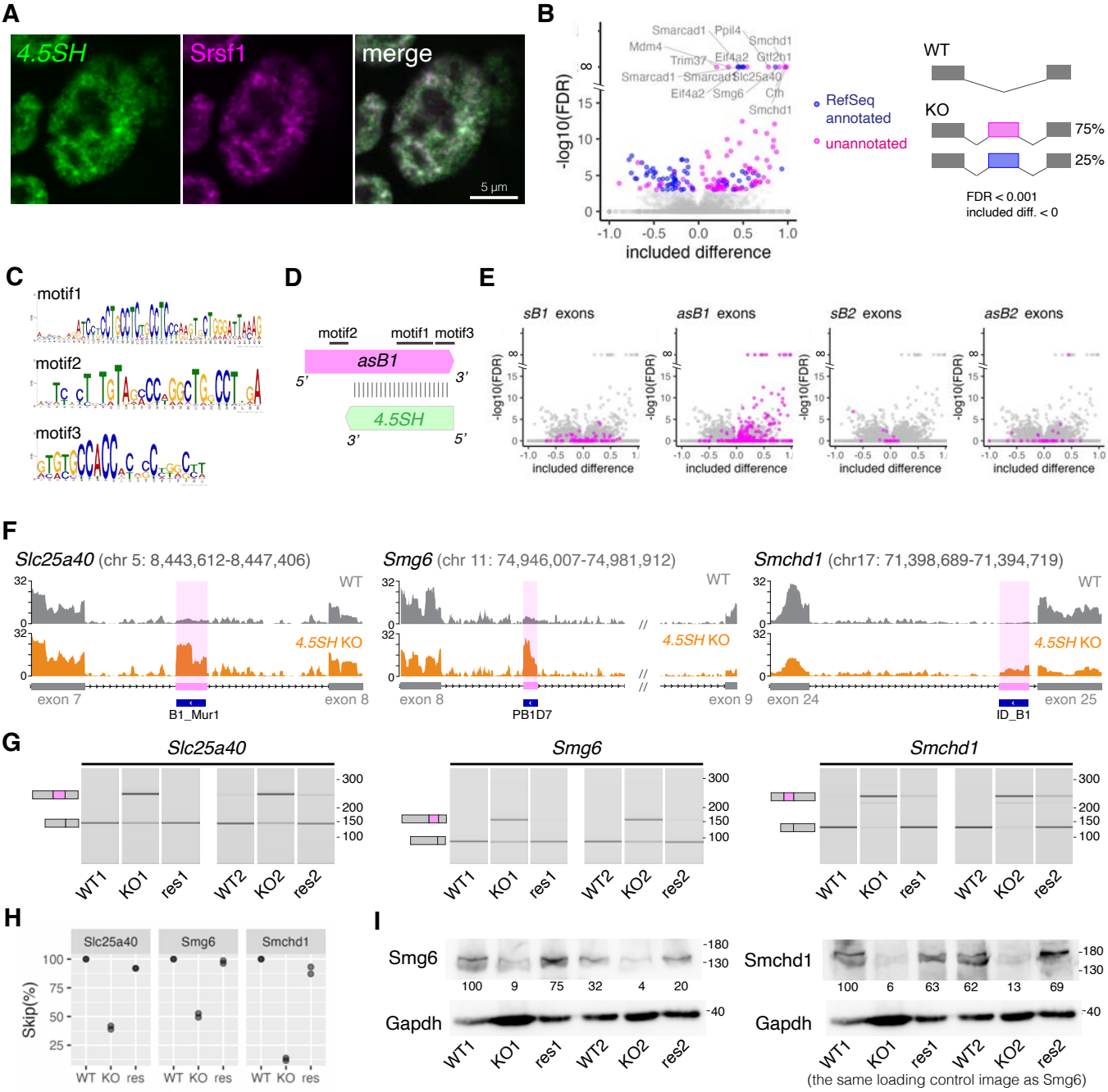

A
4.5SH
Srsf1
merge
5 μm
B
-log10(FDR)
included difference
RefSeq annotated
unannotated
WT
KO
75%
25%
FDR < 0.001
included diff. < 0
C
motif1
motif2
motif3
D
motif2
motif1 motif3
asB1
4.5SH
5'
3'
E
sB1 exons
asB1 exons
sB2 exons
asB2 exons
F
Slc25a40 (chr 5: 8,443,612-8,447,406)
Smg6 (chr 11: 74,946,007-74,981,912)
Smchd1 (chr17: 71,398,689-71,394,719)
WT
4.5SH KO
exon 7
exon 8
exon 9
exon 24
exon 25
B1_Mur1
PB1D7
ID_B1
G
Slc25a40
Smg6
Smchd1
WT1
KO1
res1
WT2
KO2
res2
H
Slc25a40
Smg6
Smchd1
Skip(%)
WT KO res
I
Smg6
Smchd1
Gapdh
100 9 75 32 4 20
100 6 63 62 13 69
(the same loading control image as Smg6)

Figure 3

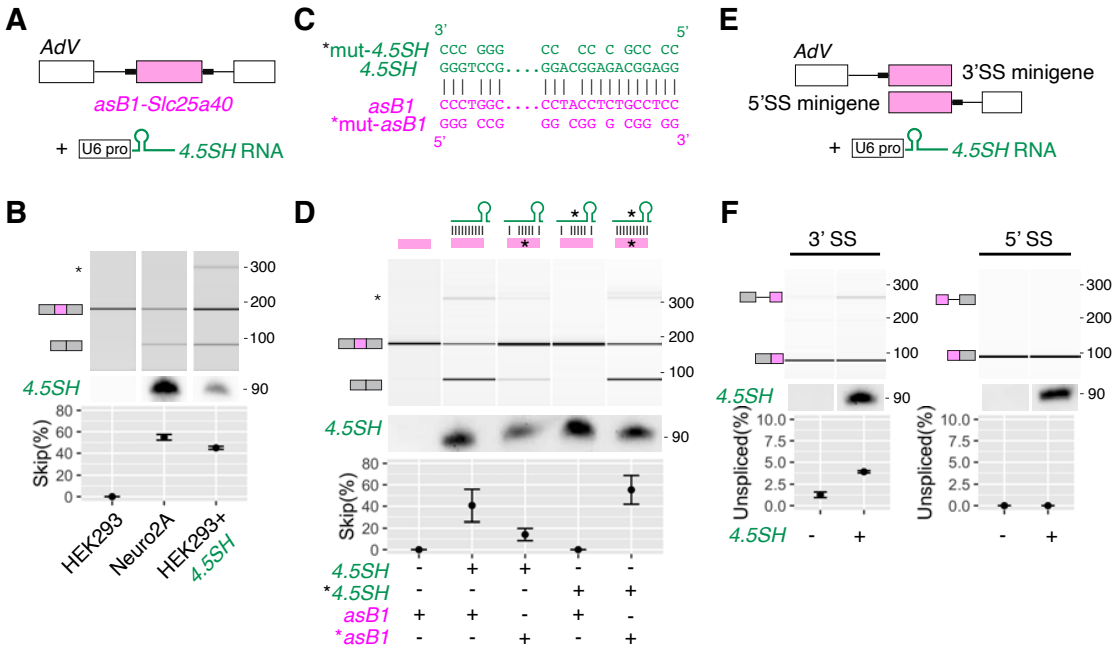

A
AdV
asB1-Slc25a40
+ U6 pro
4.5SH RNA
C
3'
5'
*mut-4.5SH CCC GGG CC CC C GCC CC
4.5SH GGGTCCG....GGACGGAGACGGAGG
asB1 CCCTGGC....CCTACCTCTGCCTCC
*mut-asB1 GGG CCG GG CGG G CGG GG
5'
3'
E
AdV
3'SS minigene
5'SS minigene
+ U6 pro
4.5SH RNA
B
300
200
100
90
4.5SH
Skip(%)
80
60
40
20
0
HEK293
Neuro2A
HEK293+ 4.5SH
D
4.5SH
*4.5SH
asB1
*asB1
F
3' SS
5' SS
Unspliced(%)
10.0
7.5
5.0
2.5
0.0

Figure 4

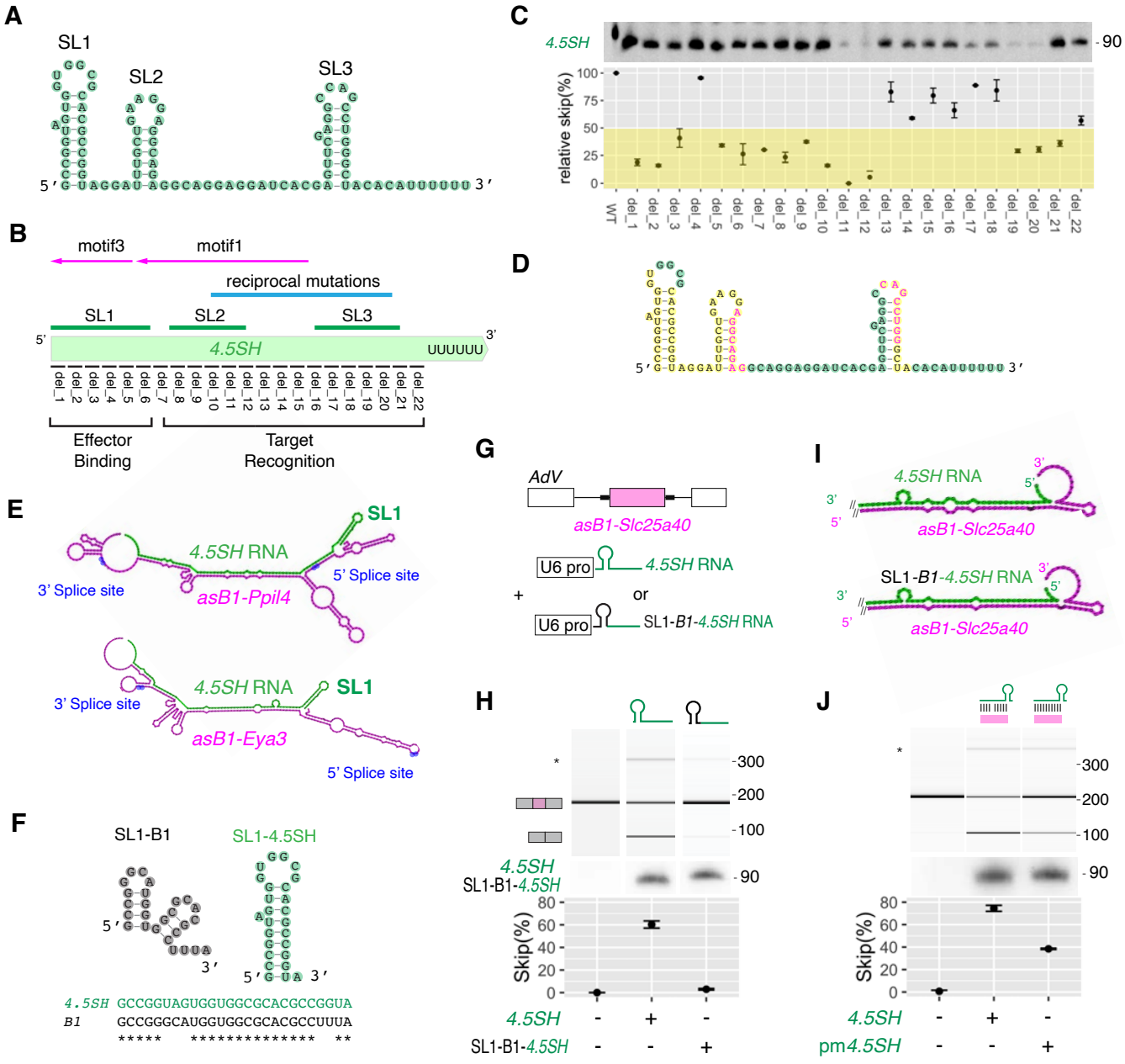

A
SL1
SL2
SL3
5′ G UAGGAU AGGCAGGAGGAUCACGA UACACAUUUUUU 3′
B
motif3
motif1
reciprocal mutations
SL1
SL2
SL3
5′
4.5SH
UUUUUU
3′
del_1 del_2 del_3 del_4 del_5 del_6 del_7 del_8 del_9 del_10 del_11 del_12 del_13 del_14 del_15 del_16 del_17 del_18 del_19 del_20 del_21 del_22
Effector Binding
Target Recognition
C
4.5SH
90
relative skip(%)
100
75
50
25
0
WT del_1 del_2 del_3 del_4 del_5 del_6 del_7 del_8 del_9 del_10 del_11 del_12 del_13 del_14 del_15 del_16 del_17 del_18 del_19 del_20 del_21 del_22
D
5′ G UAGGAU AGGCAGGAGGAUCACGA UACACAUUUUUU 3′
E
SL1
4.5SH RNA
3′ Splice site
5′ Splice site
asB1-Ppil4
4.5SH RNA
SL1
3′ Splice site
asB1-Eya3
5′ Splice site
F
SL1-B1
SL1-4.5SH
5′ 3′
5′ G UA 3′
4.5SH GCCGGUAGUGGUGGCGCACGCCGGUA
B1 GCCGGGCAUGGUGGCGCACGCCUUUA
***** ********** **
G
AdV
asB1-Slc25a40
U6 pro
4.5SH RNA
+
or
U6 pro
SL1-B1-4.5SH RNA
H
300
200
100
90
4.5SH
SL1-B1-4.5SH
Skip(%)
80
60
40
20
0
4.5SH - + -
SL1-B1-4.5SH - - +
I
4.5SH RNA
3′
5′
asB1-Slc25a40
SL1-B1-4.5SH RNA
3′
5′
asB1-Slc25a40
J
300
200
100
90
Skip(%)
80
60
40
20
0
4.5SH - + -
pm4.5SH - - +

Figure 5

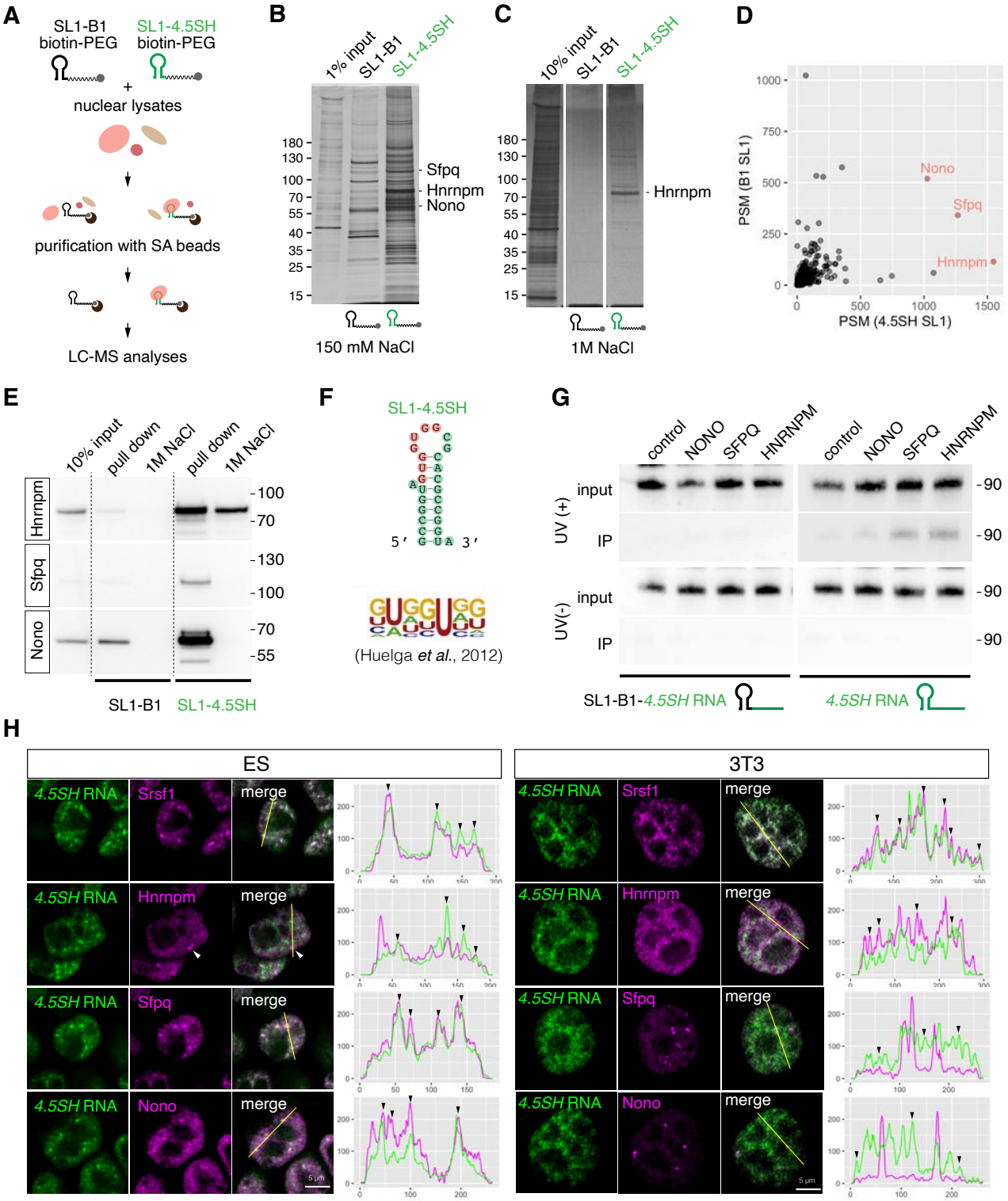

A
SL1-B1 biotin-PEG
SL1-4.5SH biotin-PEG
nuclear lysates
purification with SA beads
LC-MS analyses
B
1% input
SL1-B1
SL1-4.5SH
Sfpq
Hnrnpm
Nono
150 mM NaCl
C
10% input
SL1-B1
SL1-4.5SH
Hnrnpm
1M NaCl
D
PSM (B1 SL1)
PSM (4.5SH SL1)
Nono
Sfpq
Hnrnpm
E
10% input
pull down
1M NaCl
pull down
1M NaCl
Hnrnpm
Sfpq
Nono
SL1-B1
SL1-4.5SH
F
SL1-4.5SH
5′
3′
(Huelga et al., 2012)
G
control
NONO
SFPQ
HNRNPM
UV (+)
UV(-)
input
IP
SL1-B1-4.5SH RNA
4.5SH RNA
H
ES
3T3
4.5SH RNA
Srsf1
Hnrnpm
Sfpq
Nono
merge

Figure 6
Yoshimoto et al. Figure 6

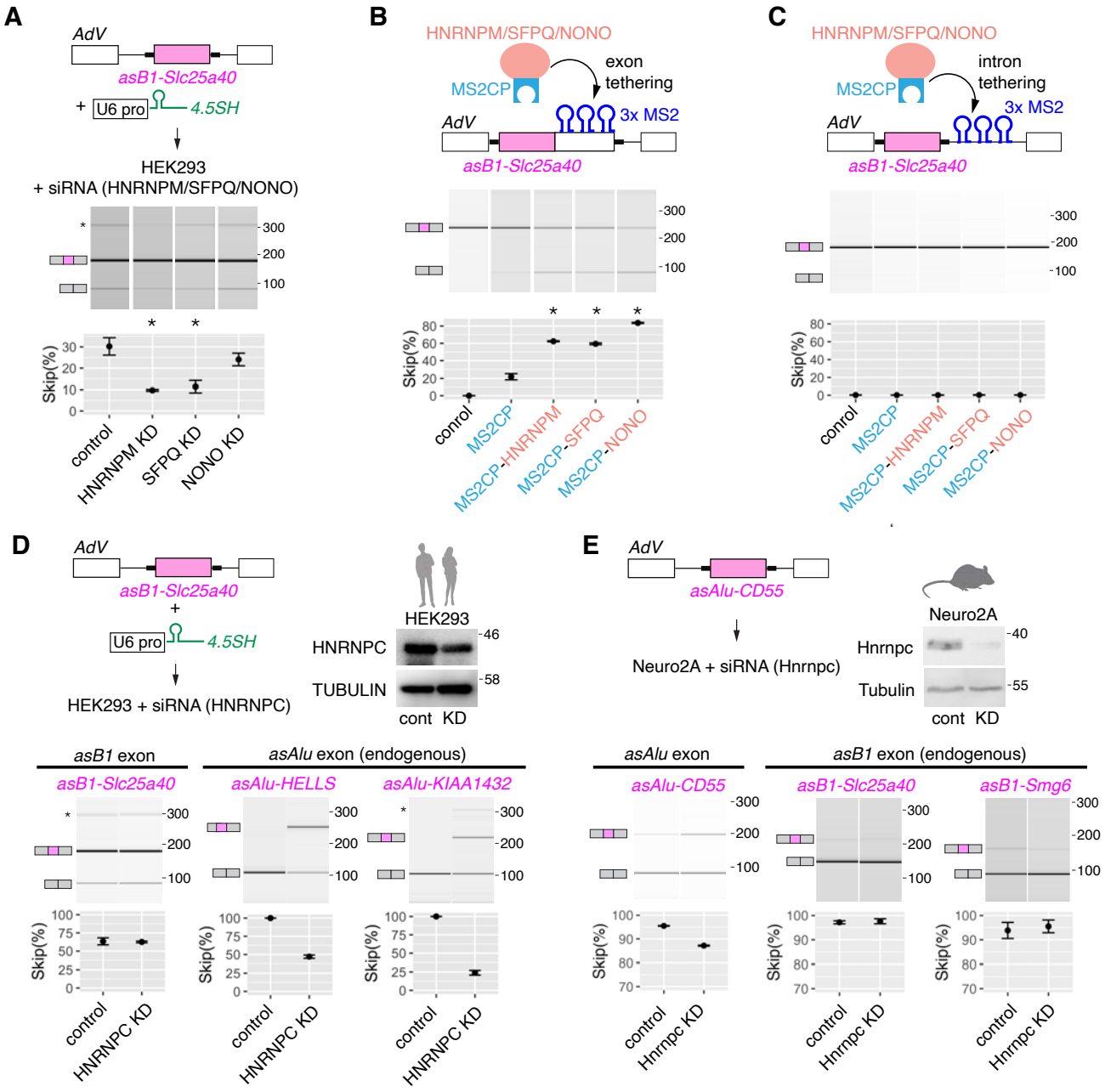
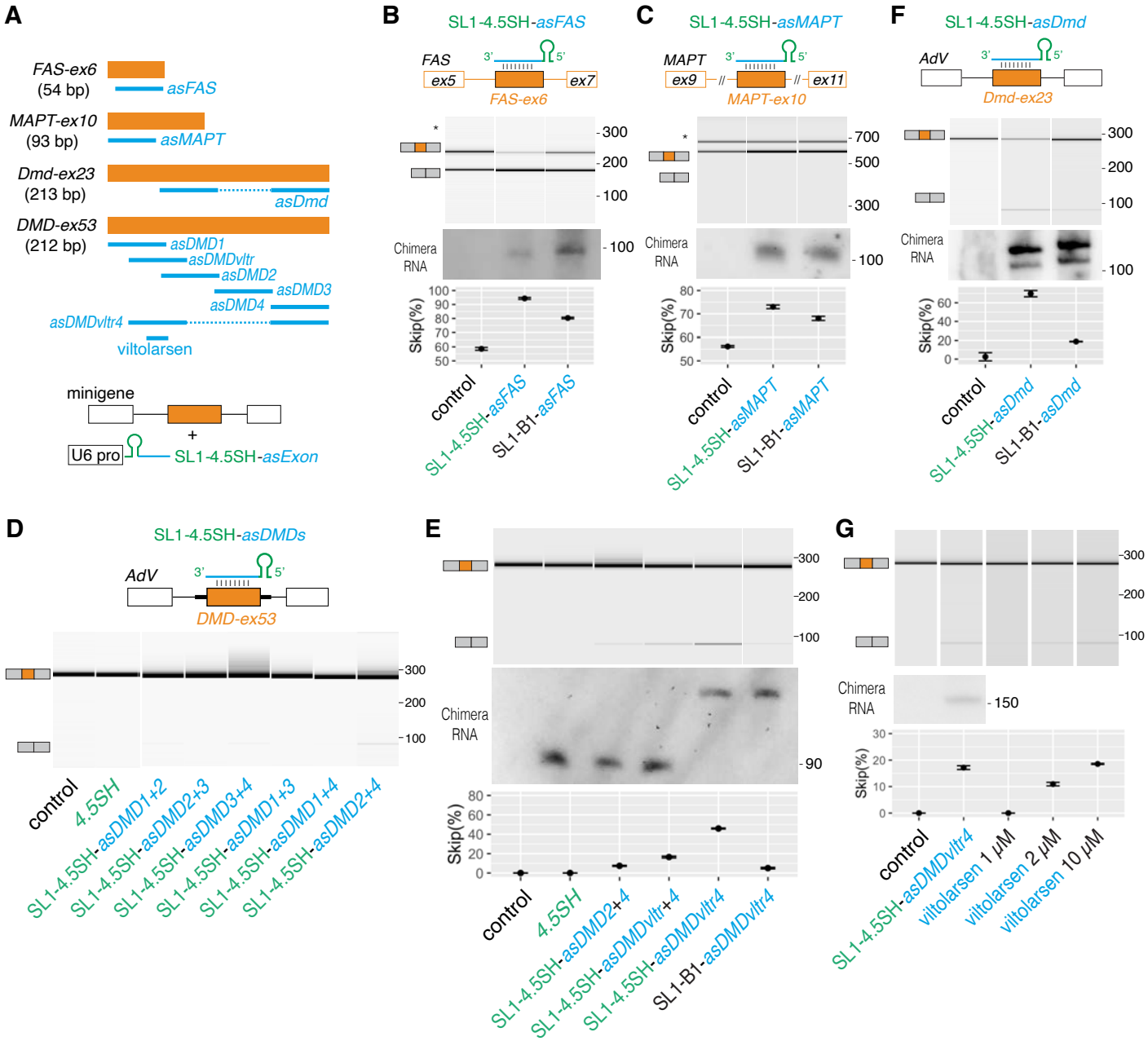
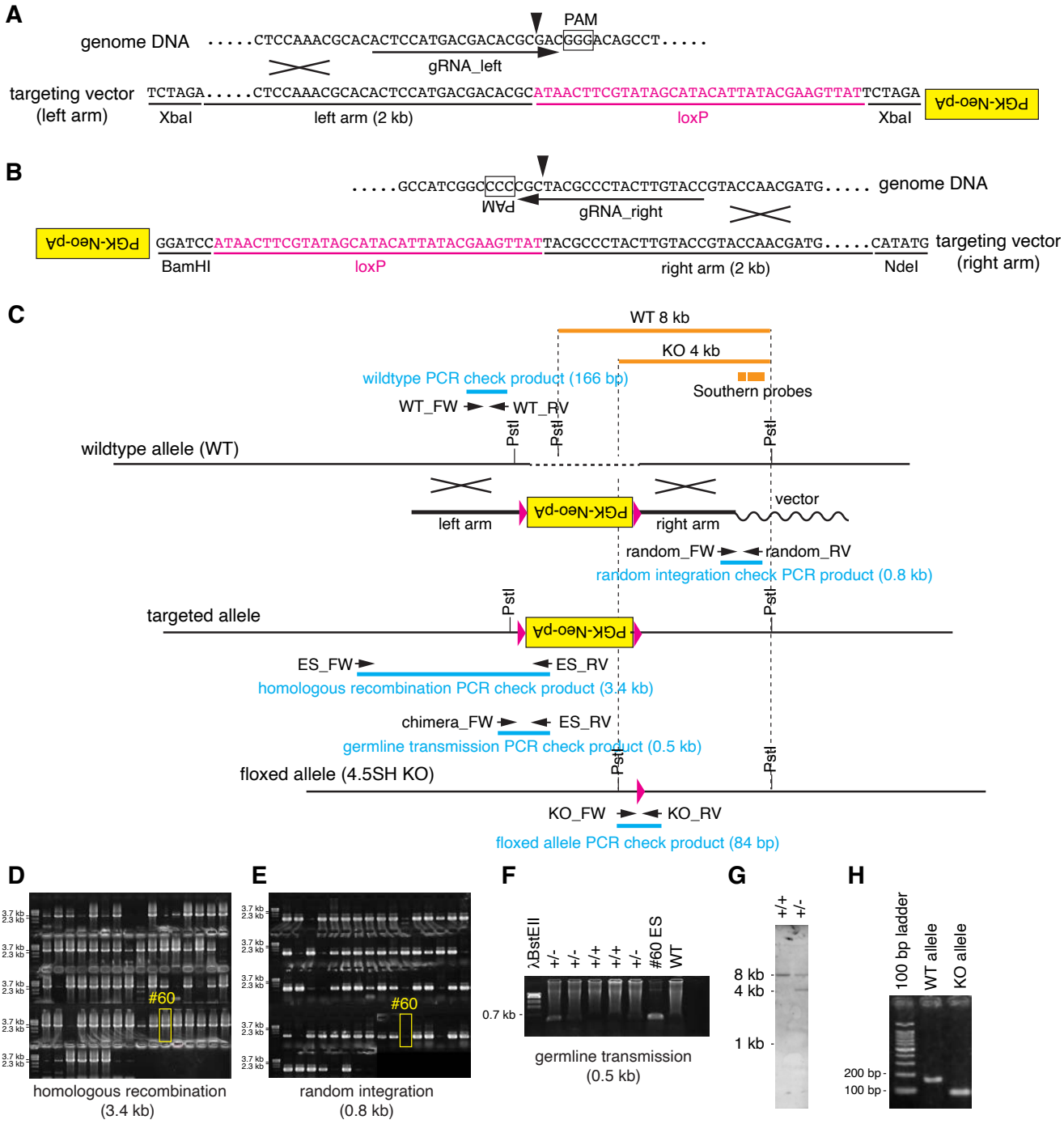


Figure 7

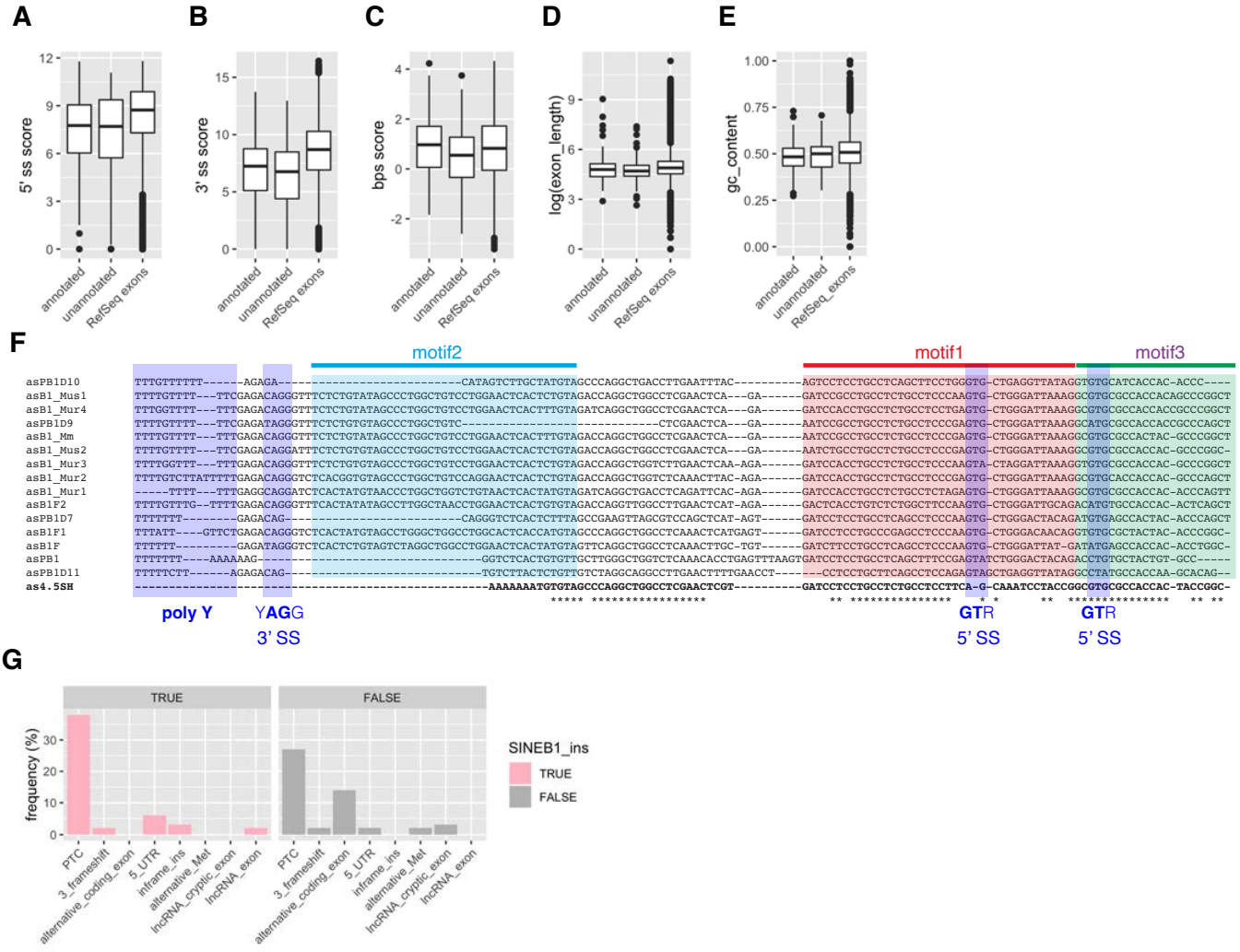
Yoshimoto et al. Figure 7



Supplemental Text and Figures

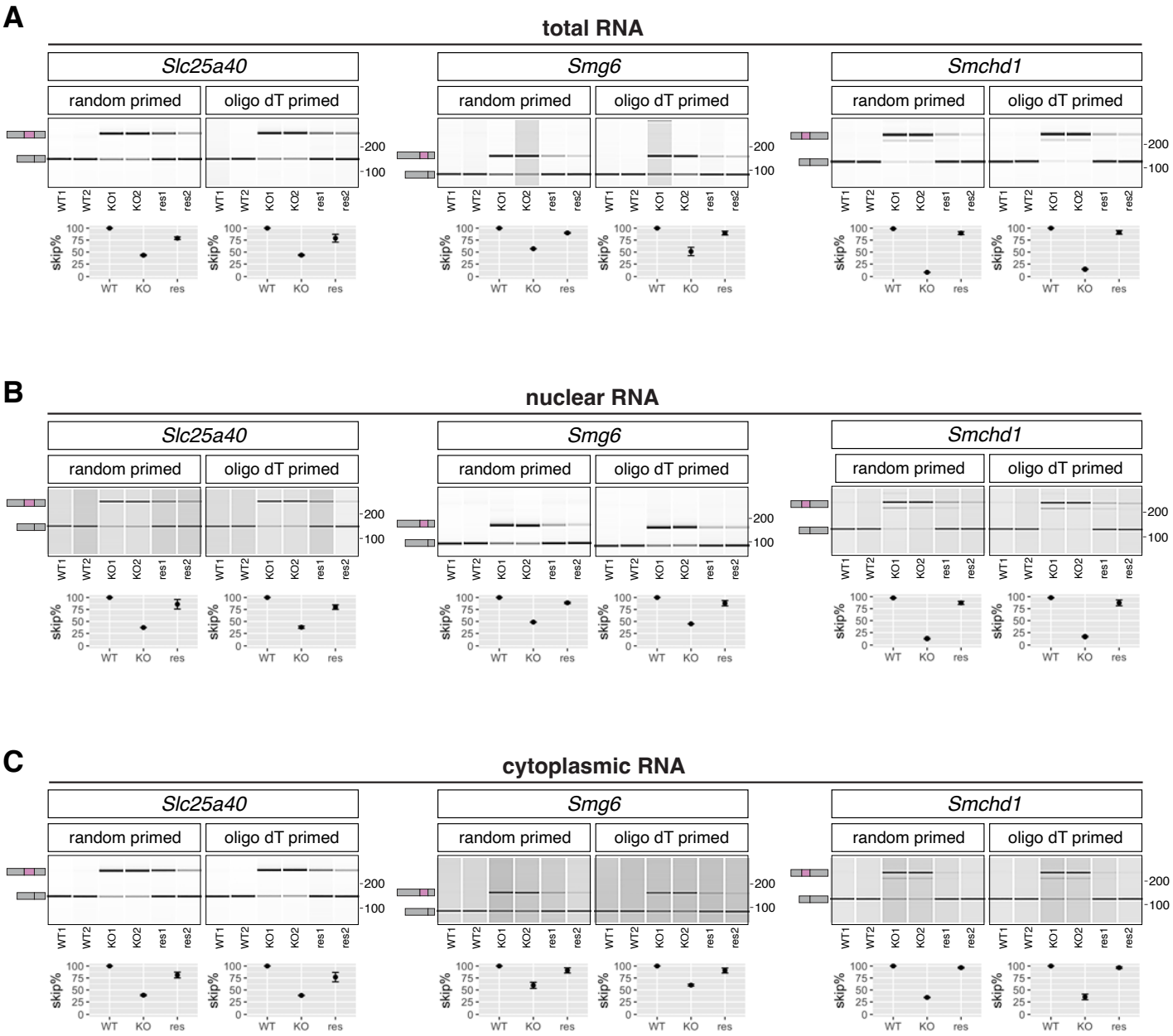


Supplemental figure S1. Generation of 4.5SH knockout mice (related to Figure 1)
(A, B) Sequences of the targeting vector at the vicinity of the junction between the left (A) and right (B) homologous arm and the loxP/PGK-Neo-pA cassette. The arrow and box indicate the sequence of the gRNA and the PAM sequences, respectively. The arrowhead indicates the putative cleavage site of Cas9. (C) A schematic of the targeting strategies and positions of the primers and probes used to validate the homologous recombination of targeted ES cells. (D) Validation of the homologous recombination using primers shown in D (ES_FW and ES_RV). 3.4 kb bands were observed in 87.5% of the ES clones. (E) Validation of random integration using primers shown in D (random_FW and random_RV). (F) Validation of germ line transmission using primers shown in C (chimera_FW and chimera_RV). (G) Southern blot analyses of genomic DNA obtained from wild-type (+/+) and heterozygous (-/-) animals. The bands were observed at the expected size. (H) A gel image of PCR products amplified by the genotyping primers shown in C (WT_FW and WT_RV, KO_FW and KO_RV).



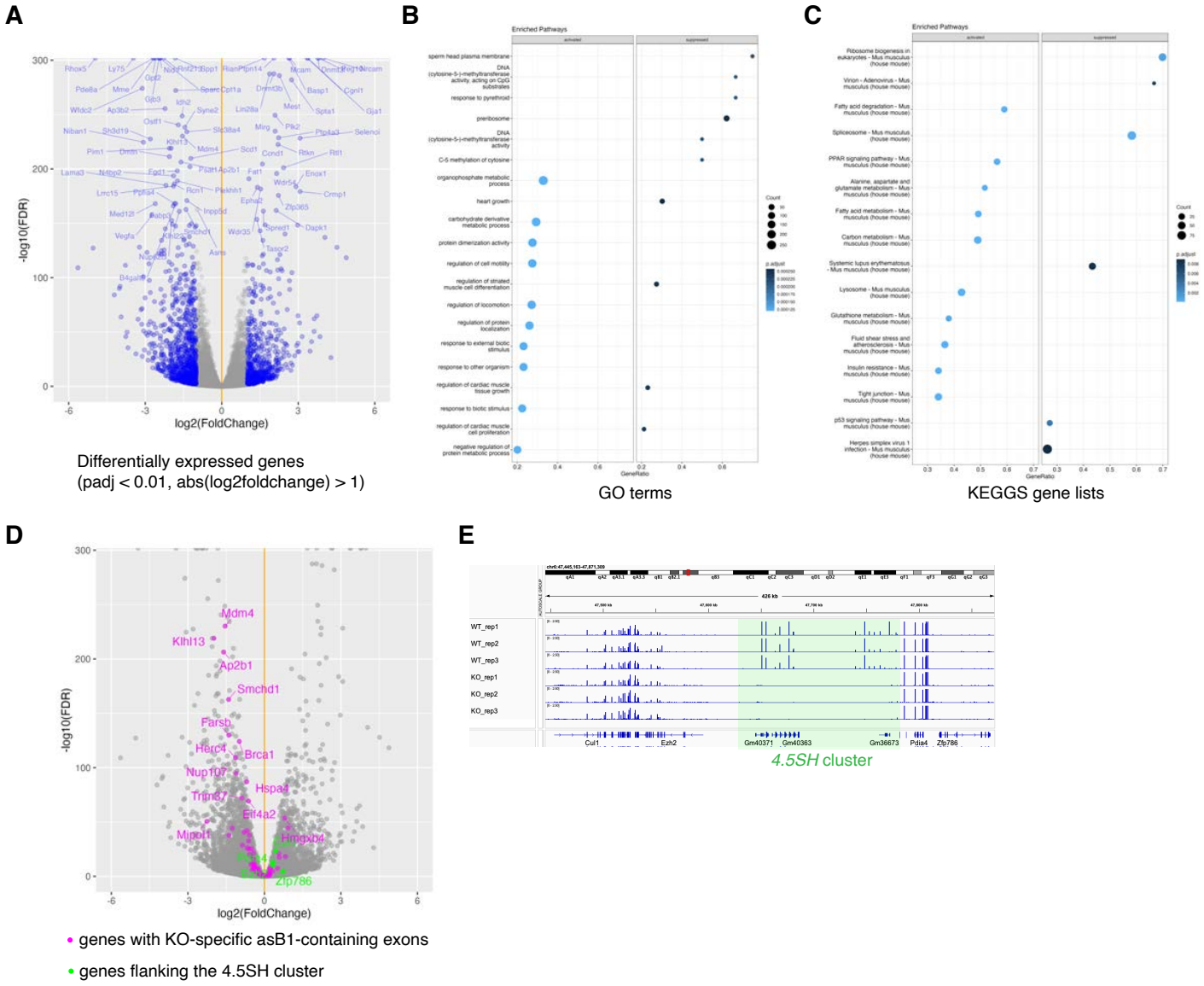
Supplemental figure S2. Characteristics of exons preferentially included in *4.5SH* KO ES cell (related to Figure 2)

(A-E) 5' (A) and 3' (B) splice site strength calculated with Maxentscan, branch point strength (bps) calculated with svm-bpfinder (C), lengths (D), and GC content (E) of Ref-Seq annotated KO-enriched exons (annotated), RefSeq-unannotated KO-enriched exons (annotated), and RefSeq exons. (F) Multiple sequence alignment of *SINE B1* antisense sequences and that of *4.5SH* (*as4.5SH*). Motifs shown in Figure 2F are indicated in red (motif1), cyan (motif2), and purple (motif4) lines and boxes. Asterisks indicate residues conserved between *as4.5SH* and *asB1*. Regions of splicing consensus sequences are shown in blue boxes. Optimal consensus sequences for the 5' and the 3' splice sites (SS) are shown in blue characters, and nucleotides in bold represent sequences conserved in the antisense sequences of the representative *SINE B1*, *asB1_Mm*. (G) Frequencies of exons enriched in *4.5SH* KO cells, by the presence (TRUE: pink) or absence (FALSE: gray) of *SINE B1* insertions: premature stop codon (PTC)-containing exons, frameshift mutation-causing exons (3_frameshift), RefSeq-annotated alternatively spliced coding exons (alternative_coding_exon), 5' untranslated region-inserted exons (5_UTR), in-frame insertion-encoding exons (inframe_ins), alternative methionine-encoding exons (alternative_Met), cryptic lncRNA exons (lncRNA_cryptic_exon), and RefSeq-annotated lncRNA exons (lncRNA_exon). Note that most *SINE B1*-containing exons contained premature stop codons.

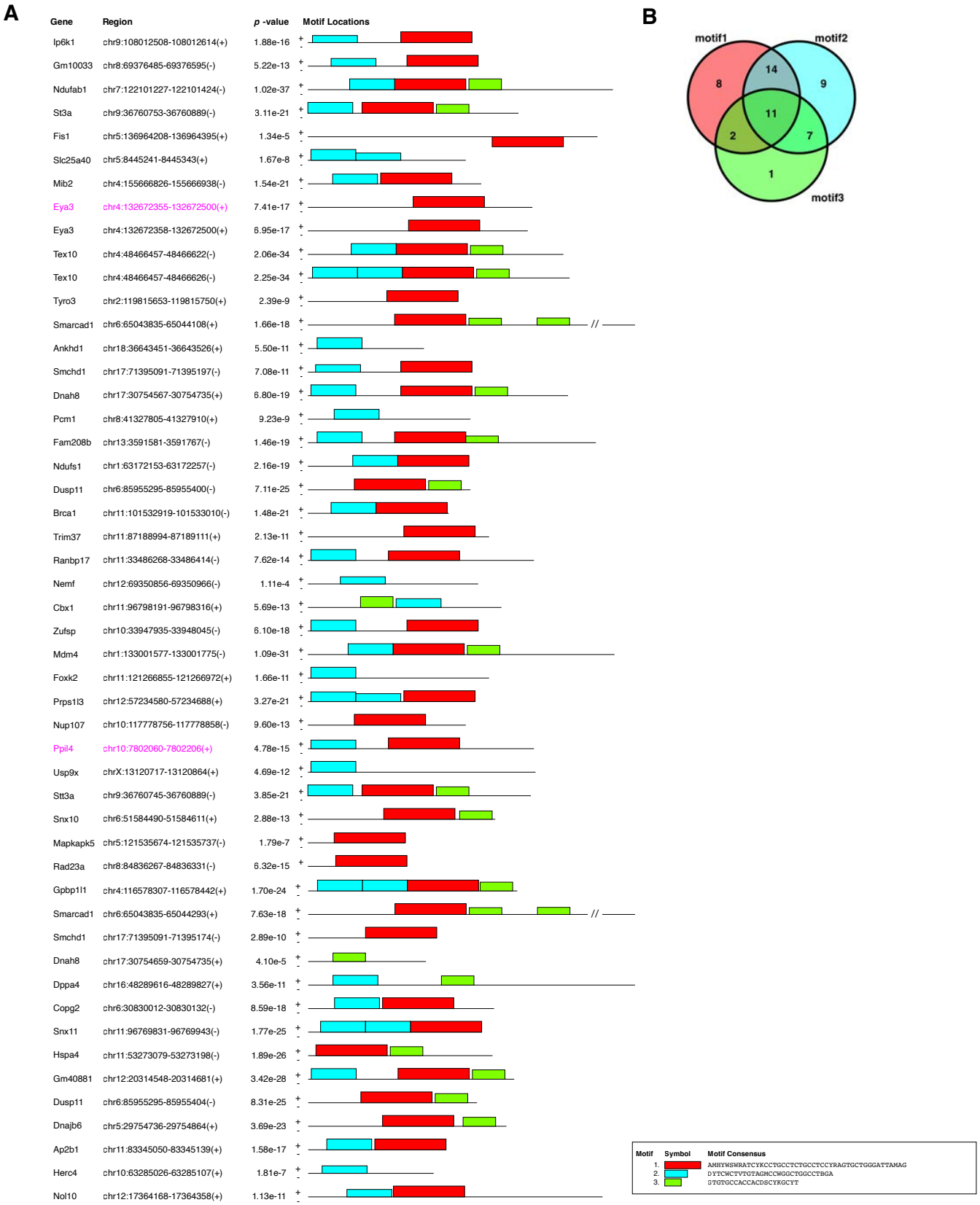


Supplemental figure S3. Confirmation of the inclusions of *asB1*-containing exons of representative genes in the *4.5SH*-KO ES cells by RT-PCR (related to Figure 2)

(A-C) The inclusion of asB1-containing exons from *Slc25a40*, *Smg6*, and *Smchd1* was examined using RT-PCR on RNAs derived from total (A), nuclear (B), and cytoplasmic (C) fractions. These RNAs were reverse-transcribed using either oligo-dT or random primers. The appearance of *asB1*-containing products was specifically observed in *4.5SH* KO ES cells as seen in the gel-like visualization of capillary electrophoresis of RT-PCR products from two independent set of WT, KO, and rescued (res) ES cells. The gel-like images, sourced from exported TIFF files ranging between 50 bp and 300 bp, were resized to uniformity in Photoshop, with constant image contrast maintained.

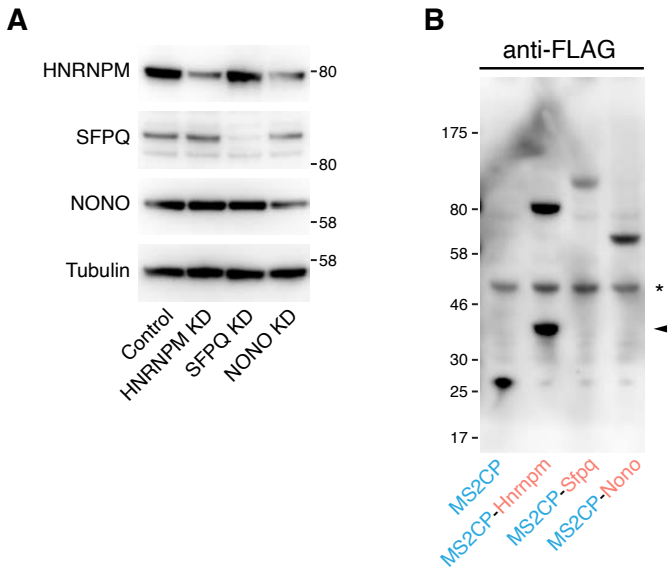


Supplemental figure S4. Differential Gene Expression Analysis in *4.5SH* KO ES Cells. (related to Figure 2) (A) A volcano plot illustrating the differentially expressed genes in *4.5SH* KO ES cells. Genes with significant changes (FDR < 0.01, absolute log2 foldchange > 1) are highlighted in blue. (B, C) Gene ontology analysis based on GO terms (B) and enrichment analysis using the KEGG pathway (C). (D) A volcano plot emphasizing the general decrease in expression of host genes with *4.5SH* KO-specific inclusions of asB1-containing exons (magenta). The Cul1, Ezh2, Pdia4, and Zfp786, which flank the 4.5SH cluster, are denoted in green. (E) Genome browser (IGV) view of RNA-seq reads surrounding the *4.5SH* cluster.



Supplemental figure S5. Positions of MEME-identified consensus sequence motifs in target exons regulated by *4.5SH* RNA (related to Figure 2)

(A) MEME output showing the positions of motif1 (red), motif2 (cyan), and motif3 (purple) in *asB1*-containing exons preferentially included in *4.5SH* KO ES cells. (B) Venn diagram showing the number of exons shared between the MEME-identified motifs.



Supplemental figure S6. Verification of the protein expression during knockdown and tethering experiments (related to Figure 6)

(A) Western blot analyses of the expression of HNRNPM, SFPQ, and NONO upon knockdown in the HEK293 cells used in Figure 6A. (B) Western blot analyses of the expression of proteins fused to the FLAG-tagged MS2 coat protein (MS2CP) in cells from Figure 6B and C. The asterisk shows non-specific bands detected by the anti-FLAG antibody, and the arrowhead indicates a short uncharacterized product of Hnrnpm.

Extended Data Table1: rMATS output and annotation of the exons enriched in 4.5SH KO ES cells

num_ID	GeneID	ID	name	chr	strand	exonStart_0base	exonEnd	upstreamES	upstreamEE	downstreamE	downstreamE	IJC_SAMPLE_1	SJC_SAMPLE_1	IJC_SAMPLE_2	SJC_SAMPLE_2	IncFormLen	SkipFormLen	PValue	FDR	IncLevel1	IncLevel2	IncLevelDifference	exon length	SINEB1	RefSeq_annotated_exon	type	chr_location
1	MSTRG.12269	4714	Cth	chr3	-	157903719	157903800	157897191	157897330	157905144	157905197	0,0,1	6,2,4	2,1,3	0,0,0	100	50	0	0	0.0,0.0,0.111	1.0,1.0,1.0	-0.963	81	FALSE	FALSE	PTC	chr3:157903719-157903800
2	MSTRG.1234	8249	Ppil4	chr10	+	7802060	7802206	7801844	7801969	7804955	7805022	1,2,0	17,10,9	14,18,14	2,2,1	100	50	0	0	0.029,0.091,0.0	0.778,0.818,0.875	-0.784	146	B1_Mur1	TRUE	PTC	chr10:7802060-7802206
3	MSTRG.13657	1574	Slc25a40	chr5	+	8445241	8445343	8443611	8443785	8447297	8447407	0,2,3	17,30,31	29,41,27	7,16,15	100	50	0	0	0.0,0.032,0.046	0.674,0.562,0.474	-0.544	102	B1_Mur1	FALSE	PTC	chr5:8445241-8445343
4	MSTRG.15261	9927	Smarcad1	chr6	+	65043835	65044293	65042648	65043445	65052619	65052797	5,6,2	41,55,31	26,39,52	21,20,18	100	50	0	0	0.057,0.052,0.031	0.382,0.494,0.591	-0.442	458	FALSE	FALSE	PTC	chr6:65043835-65044293
5	MSTRG.15261	3102	Smarcad1	chr6	+	65043835	65044108	65043205	65043445	65052619	65052797	5,6,2	41,55,31	28,49,58	21,20,18	100	50	0	0	0.057,0.052,0.031	0.4,0.551,0.617	-0.476	273	FALSE	FALSE	PTC	chr6:65043835-65044108
6	MSTRG.15261	3103	Smarcad1	chr6	+	65043835	65052247	65042648	65043445	65052619	65052797	5,6,3	41,55,31	27,40,54	21,20,18	100	50	0	0	0.057,0.052,0.046	0.391,0.5,0.6	-0.445	8412	B1_Mus1/B1	FALSE	PTC	chr6:65043835-65052247
7	MSTRG.16542	4263	Gtf2h1	chr7	+	46797396	46797465	46795867	46796288	46801671	46801840	3,3,1	14,11,2	11,12,16	0,0,0	100	50	0	0	0.097,0.12,0.2	1.0,1.0,1.0	-0.861	69	PB1D9	TRUE	5_UTR	chr7:46797396-46797465
8	MSTRG.2853	6589	Smg6	chr11	+	74948363	74948439	74946006	74946219	74981851	74981913	0,0,0	15,4,7	25,18,18	3,0,1	100	50	0	0	0.0,0.0,0.0	0.806,1.0,0.9	-0.902	76	PB1D7	FALSE	PTC	chr11:74948363-74948439
9	MSTRG.3042	6443	Trim37	chr11	+	87188994	87189111	87186378	87186515	87190073	87190159	0,0,0	59,47,68	22,8,15	31,33,24	100	50	0	0	0.0,0.0,0.0	0.262,0.108,0.238	-0.203	117	ID_B1	FALSE	PTC	chr11:87188994-87189111
10	MSTRG.7203	8017	Eif4a2	chr16	+	23112350	23112457	23111816	23111896	23113166	23113899	96,71,85	144,147,166	83,104,104	17,16,24	100	50	0	0	0.25,0.195,0.204	0.709,0.765,0.684	-0.503	107	FALSE	TRUE	PTC	chr16:23112350-23112457
11	MSTRG.7203	8018	Eif4a2	chr16	+	23112350	23112461	23111336	23111896	23113166	23113899	127,130,138	144,147,166	130,180,157	17,16,24	100	50	0	0	0.306,0.307,0.294	0.793,0.849,0.766	-0.5	111	FALSE	TRUE	PTC	chr16:23112350-23112461
12	MSTRG.730	7650	Mdm4	chr1	-	133001577	133001775	133001282	133001443	133003790	133003890	0,1,0	65,77,77	9,14,5	7,13,7	100	50	0	0	0.0,0.006,0.0	0.391,0.35,0.263	-0.333	198	B1_Mus2	FALSE	PTC	chr1:133001577-133001775
13	MSTRG.8463	10081	Smchd1	chr17	-	71395091	71395174	71394716	71394944	71398551	71398689	3,0,0	26,39,22	26,9,11	0,0,0	100	50	0	0	0.055,0.0,0.0	1.0,1.0,1.0	-0.982	83	ID_B1	FALSE	PTC	chr17:71395091-71395174
14	MSTRG.8463	3466	Smchd1	chr17	-	71395091	71395197	71394716	71394944	71398551	71398689	3,0,0	26,39,22	28,13,15	0,0,0	100	50	0	0	0.055,0.0,0.0	1.0,1.0,1.0	-0.982	106	ID_B1	FALSE	PTC	chr17:71395091-71395197
15	MSTRG.13142	1820	Eya3	chr4	+	132672355	132672500	132670188	132670255	132672895	132673032	0,0,0	22,23,20	11,11,7	8,3,5	100	50	4.44E-16	3.57E-13	0.0,0.0,0.0	0.407,0.647,0.412	-0.489	145	FALSE	FALSE	PTC	chr4:132672355-132672500
16	MSTRG.13783	11356	Dnajb6	chr5	+	29754736	29754864	29753541	29754591	29756233	29764530	0,0,0	7,7,13	16,2,5	0,0,2	100	50	1.11E-15	8.37E-13	0.0,0.0,0.0	1.0,1.0,0.556	-0.852	128	B1_Mus2	FALSE	PTC	chr5:29754736-29754864
17	MSTRG.444	6429	Farsb	chr1	-	78465010	78465086	78463271	78463333	78466966	78468032	0,1,1	11,23,22	8,15,6	3,1,3	100	50	1.29E-14	9.13E-12	0.0,0.021,0.022	0.571,0.882,0.5	-0.637	76	PB1D7	TRUE	3_frameshift	chr1:78465010-78465086
18	MSTRG.12875	9372	Gpbp1l1	chr4	+	116578307	116578442	116574336	116574614	116579246	116579319	0,1,0	14,18,17	13,5,8	1,3,0	100	50	2.51E-14	1.68E-11	0.0,0.027,0.0	0.867,0.455,1.0	-0.765	135	B1_Mus2	FALSE	PTC	chr4:116578307-116578442
19	MSTRG.2392	7097	Ranbp17	chr11	-	33486268	33486414	33485888	33486083	33487642	33487747	0,0,0	27,64,42	16,7,5	8,7,17	100	50	1.75E-13	1.11E-10	0.0,0.0,0.0	0.5,0.333,0.128	-0.32	146	B1_Mur3	FALSE	PTC	chr11:33486268-33486414
20	MSTRG.1348	7580	Zufsp	chr10	-	33947935	33948045	33943649	33943760	33948928	33949498	0,1,0	3,1,2	4,12,8	0,1,0	100	50	2.57E-13	1.55E-10	0.0,0.333,0.0	1.0,0.857,1.0	-0.841	110	B1_Mur1	TRUE	PTC	chr10:33947935-33948045
21	MSTRG.13594	1602	Mib2	chr4	-	155666826	155666938	155661105	155661374	155667013	155667169	1,2,0	22,21,19	23,7,12	7,9,5	100	50	6.55E-13	3.76E-10	0.022,0.045,0.0	0.622,0.28,0.545	-0.46	112	B1_Mur1	FALSE	5_UTR	chr4:155666826-155666938
22	MSTRG.1552	11450	Herc4	chr10	+	63285026	63285107	63283180	63283341	63285651	63285728	1,0,0	14,14,5	4,7,6	0,3,0	100	50	2.04E-12	1.12E-09	0.034,0.0,0.0	1.0,0.538,1.0	-0.835	81	PB1D7	FALSE	PTC	chr10:63285026-63285107
23	MSTRG.13142	1821	Eya3	chr4	+	132672358	132672500	132670188	132670255	132672895	132673032	0,0,0	22,23,20	14,1,9	8,3,5	100	50	7.83E-12	4.10E-09	0.0,0.0,0.0	0.467,0.143,0.474	-0.361	142	FALSE	FALSE	PTC	chr4:132672358-132672500
24	MSTRG.4077	7160	Nemf	chr12	-	69350856	69350966	69348850	69348937	69352306	69352374	1,0,2	9,9,7	8,1,2	1,0,0	100	50	1.07E-11	5.39E-09	0.053,0.0,0.125	0.8,1.0,1.0	-0.874	110	FALSE	FALSE	PTC	chr12:69350856-69350966
25	MSTRG.3150	7495	Cbx1	chr11	+	96798191	96798316	96789246	96789353	96800770	96800947	0,2,0	79,60,92	11,10,15	45,60,58	100	50	1.16E-11	5.62E-09	0.0,0.016,0.0	0.109,0.077,0.115	-0.095	125	FALSE	FALSE	5_UTR	chr11:96798191-96798316
26	MSTRG.19140	236	Ip6k1	chr9	+	108012508	108012614	108002720	108002822	108024102	108024450	0,0,0	26,22,28	4,3,10	7,4,6	100	50	3.77E-11	1.75E-08	0.0,0.0,0.0	0.222,0.273,0.455	-0.317	106	ID_B1	FALSE	5_UTR	chr9:108012508-108012614
27	MSTRG.18519	915	Stt3a	chr9	-	36760753	36760889	36759202	36759324	36763336	36763397	0,0,1	71,81,64	5,11,8	26,31,32	100	50	5.56E-11	2.39E-08	0.0,0.0,0.008	0.088,0.151,0.111	-0.114	136	B1_Mur2	FALSE	PTC	chr9:36760753-36760889
28	MSTRG.6103	10585	Rbm26	chr14	-	105139944	105140016	105139090	105139162	105140315	105140447	3,9,7	4,7,6	4,9,3	0,0,0	100	50	9.53E-11	3.96E-08	0.273,0.391,0.368	1.0,1.0,1.0	-0.656	72	FALSE	FALSE	PTC	chr14:105139944-105140016
29	MSTRG.15382	6269	Dusp11	chr6	-	85955295	85955400	85953370	85953494	85958583	85958715	4,3,1	8,4,6	11,3,5	1,0,0	100	50	2.11E-10	7.73E-08	0.2,0.273,0.077	0.846,1.0,1.0	-0.765	105	PB1D10	FALSE	PTC	chr6:85955295-85955400
30	MSTRG.14471	9090	Mapkapk5	chr5	-	121535674	121535737	121534334	121534430	121535853	121537220	0,2,0	8,14,8	7,6,3	3,3,1	100	50	1.31E-09	4.25E-07	0.0,0.067,0.0	0.538,0.5,0.6	-0.524	63	ID_B1	FALSE	PTC	chr5:121535674-121535737
31	MSTRG.15382	11324	Dusp11	chr6	-	85955295	85955404	85953370	85953494	85958583	85958715	2,0,0	8,4,6	2,2,3	1,0,0	100	50	2.23E-09	6.85E-07	0.111,0.0,0.0	0.5,1.0,1.0	-0.796	109	PB1D10	FALSE	PTC	chr6:85955295-85955404
32	MSTRG.15899	6457	Rps9	chr7	+	3706007	3706181	3704104	3705919	3706650	3707508	17,24,30	496,550,583	23,30,41	200,244,278	100	50	2.54E-09	7.46E-07	0.017,0.021,0.025	0.054,0.058,0.069	-0.039	174	FALSE	TRUE	PTC	chr7:3706007-3706181
33	MSTRG.498	3204	Dis3l2	chr1	+	86812511	86812589	86791415	86791457	86821221	86821450	2,4,5	4,2,2	9,6,7	0,0,0	100	50	3.53E-09	1.01E-06	0.2,0.5,0.556	1.0,1.0,1.0	-0.581	78	FALSE	FALSE	PTC	chr1:86812511-86812589
34	MSTRG.1980	8233	Nup107	chr10	-	117778756	117778858	117777592	117777664	117780677	117781448	0,0,0	19,17,21	2,8,7	2,0,6	100	50	3.94E-09	1.08E-06	0.0,0.0,0.0	0.333,1.0,0.368	-0.567	102	PB1D10	FALSE	PTC	chr10:117778756-117778858
35	MSTRG.12530	2039	Tex10	chr4	-	48466457	48466626	48462768	48462881	48467663	48467904	0,0,0	14,5,8	7,10,8	0,8,8	100	50	4.47E-09	1.17E-06	0.0,0.0,0.0	1.0,0.385,0.333	-0.573	169	B1_Mus1	FALSE	PTC	chr4:48466457-48466626
36	MSTRG.17639	550	Gm10033	chr8	-	69376485	69376595	69375305	69375432	69395440	69395487	0,0,1	1,1,2	1,1,1	0,0,0	100	50	4.82E-09	1.24E-06	0.0,0.0,0.2	1.0,1.0,1.0	-0.933	110	B1_Mur3	TRUE	lncRNA_exon	chr8:69376485-69376595
37	MSTRG.8795	9571	Brd8	chr18	-	34600606	34600848	34599834	34600241	34602202	34602247	0,0,1	10,8,7	10,5,5	8,0,2	100	50	1.02E-08	2.45E-06	0.0,0.0,0.067	0.385,1.0,0.556	-0.625	242	FALSE	FALSE	PTC	chr18:34600606-34600848
38	MSTRG.17786	8661	Hmgxb4	chr8	+	75016246	75016406	75000649	75001584	75020243	75020325	0,0,0	15,11,9	14,6,2	1,4,5	100	50	1.16E-08	2.70E-06	0.0,0.0,0.0	0.875,0.429,0.167	-0.49	160	FALSE	FALSE	PTC	chr8:75016246-75016406
39	MSTRG.12530	2038	Tex10	chr4	-	48466457	48466622	48462630	48462881	48467663	48467904	0,0,0	14,5,8	7,9,7	0,8,8	100	50	1.31E-08	2.97E-06	0.0,0.0,0.0	1.0,0.36,0.304	-0.555	165	B1_Mus1	FALSE	PTC	chr4:48466457-48466622
40	MSTRG.10622	2603	Tyro3	chr2	+	119815653	119815750	119813162	119813299	119816643	119818076	0,0,0	19,26,22	5,3,5	7,7,7	100	50	2.07E-08	4.38E-06	0.0,0.0,0.0	0.263,0.176,0.263	-0.234	97	B1F	FALSE	PTC	chr2:119815653-119815750
41	MSTRG.11586	5203	Rnf13	chr3	+	57747890	57748034	57736152	57736200	57764256	57764382	2,0,1	5,3,0	2,4,4	0,0,0	100	50	2.16E-08	4.41E-06	0.167,0.0,1.0	1.0,1.0,1.0	-0.611	144	FALSE	TRUE	PTC	chr3:57747890-57748034
42	MSTRG.5560	10872	Anxa11	chr14	+	25842706	25842784	25842230	25842295	25868863	25868926	0,2,1	3,0,6	1,1,2	0,0,0	100	50	2.28E-08	4.51E-06	0.0,1.0,0.077	1.0,1.0,1.0	-0.641	78	FALSE	TRUE	5_UTR	chr14:25842706-25842784
43	MSTRG.328	5511	Ndufs1	chr1	-	63172153	63172257	63170848	63170914	63176427	63176800	8,5,6	10,4,6	16,6,10	0,1,1	100	50	2.55E-08	4.96E-06	0.286,0.385,0.333	1.0,0.75,0.833	-0.526	104	B1_Mus2	TRUE	PTC	chr1:63172153-63172257
44	MSTRG.19454	8506	Usp9x	chrX	+	13120717	13120864	13116623	13116757	13121432	13121520	1,0,0	184,155,191	6,9,15	104,110,97	100	50	3.07E-08	5.79E-06	0.003,0.0,0.0	0.028,0.039,0.072	-0.045	147	B1_Mur1	FALSE	PTC	chrX:13120717-13120864
45	MSTRG.19478	169	Klhl13	chrX	-	23294214	23294280	23261434	23261573	23315129	23315201	9,7,12	5,1,4	2,2,3	0,0,0	100	50	4.77E-08	8.72E-06	0.474,0.778,0.6	1.0,1.0,1.0	-0.383	66	FALSE	TRUE	alternative_coding_exon	chrX:23294214-23294280
46	MSTRG.17870	9158	Rad23a	chr8	-	84836267	84836331	84835481	84835919	84837459	84837593	0,0,1	8,13,5	2,2,3	0,1,1	100	50	5.80E-08	1.03E-05	0.0,0.0,0.091	1.0,0.5,0.6	-0.67	64	B1F	FALSE	PTC	chr8:84836267-84836331
47	MSTRG.14078	1455	Aasdh	chr5	-	76894289	76894362	76891602	76891795	76896376	76896690	0,1,0	3,3,2	3,3,5	3,0,0	100	50	1.07E-07	1.82E-05	0.0,0.143,0.0	0.333,1.0,1.0	-0.73	73	FALSE	TRUE	alternative_coding_exon	chr5:76894289-76894362
48	MSTRG.5213	5071	Vcan	chr13	-	89702917	89705797	89688284	89693501	89712204	89712498	2,1,3	1,3,2	3,5,6	0,0,1	100	50	1.47E-07	2.46E-05	0.5,0.143,0.429	1.0,1.0,0.75	-0.559	2880	FALSE	TRUE	alternative_coding_exon	chr13:89702917-89705797
49	MSTRG.8867	3252	Ankhd1	chr18	+	36643451	36643526	36642925	36643094	36644260	36644714	1,0,0	22,29,17	11,5,5	11,20,27	100	50	2.23E-07	3.37E-05	0.022,0.0,0.0	0.333,0.111,0.085	-0.169	75	PB1D9	FALSE	inframe_ins	chr18:36643451-36643526
50	MSTRG.10683	5187	Ctdspl2	chr2	+	121995920	121996006	121993345	121993432	122003900	122003963	0,2,0	17,12,18	4,8,2	4,5,5	100	50	2.31E-07	3.44E-05	0.0,0.077,0.0	0.333,0.444,0.167	-0.289	86	FALSE	TRUE	PTC	chr2:121995920-121996006
51	MSTRG.15903	8899	Rdh13	chr7	-	4443736	4443771	4442565	4442749	4444202	4444321	0,0,0	15,6,9	4,6,3	8,7,8	85	50	2.45E-07	3.60E-05	0.0,0.0,0.0	0.227,0.335,0.181	-0.248	35	FALSE	TRUE	3_frameshift	chr7:4443736-4443771
52	MSTRG.18519	8879	Stt3a	chr9	-	36760745	36760889	36759202	36759324	36763336	36763397	0,0,1	71,81,64	2,8,7	26,31,32	100	50	2.57E-07	3.74E-05	0.0,0.0,0.008	0.037,0.114,0.099	-0.081	144	B1_Mur2	FALSE	PTC	chr9:36760745-36760889
53	MSTRG.12719	6859	Usp1	chr4	+	98926454	98926674	98923662	98924243	98928233	98928354	23,25,27	6,7,2	8,17,7	0,0,0	100	50	3.38E-07	4.74E-05	0.657,0.641,0.871	1.0,1.0,1.0	-0.277	220	FALSE	TRUE	alternative_Met	chr4:98926454-98926674
54	MSTRG.15756	983	Foxm1	chr6	+	128371967	128372012	128370956	128371085	128372530	128376155	6,4,4	4,2,1	2,3,8	0,0,0	95	50	3.93E-07	5.42E-05	0.441,0.513,0.678	1.0,1.0,1.0	-0.456	45	FALSE	TRUE	alternative_coding_exon	chr6:128371967-128372012
55	MSTRG.3799	11195	Gm40881	chr12	+	20314548	20314681	20312277	20312640	20319604	20324199	0,0,0	2,5,6	2,8,2	4,5,3	100	50	3.96E-07	5.42E-05	0.0,0.0,0.0	0.2,0.444,0.25	-0.298	133	B1_Mus2	FALSE	lncRNA_exon	chr12:20314548-20314681
56	MSTRG.4536	5455	Fam208b	chr13	-	3591581	3591767	3589940	3590648	3594130	3594193	0,0,1	8,8,6	11,8,7	19,12,10	100	50	4.01E-07	5.44E-05	0.0,0.0,0.077	0.224,0.25,0.259	-0.219	186	B1_Mur4	FALSE	PTC	chr13:3591581-3591767
57	MSTRG.14686	1259	Fis1	chr5	+	136964208	136964395	136962998	136963131	136965092	136965169	0,0,0	4,11,10	3,5,3	5,6,3	100	50	4.07E-07	5.45E-05	0.0,0.0,0.0	0.231,0.294,0.333	-0.286	187	FALSE	FALSE	PTC	chr5:136964208-136964395
58	MSTRG.12719	6858	Usp1	chr4	+	98926435	98926674	98923662	98924243	98928233	98928354	26,25,25	6,7,2	7,11,10	0,0,0	100	50	4.41E-07	5.84E-05	0.684,0.641,0.862	1.0,1.0,1.0	-0.271	239	FALSE	TRUE	alternative_Met	chr4:98926435-98926674
59	MSTRG.10195	7191	Scai	chr2	-	39082709	39082852	39080076	39080547	39083537	39083610	0,1,0	4,3,2	4,4,4	0,1,3	100	50	4.81E-07	6.17E-05	0.0,0.143,0.0	1.0,0.667,0.4	-0.641	143	FALSE	TRUE	PTC	chr2:39082709-39082852
60	MSTRG.8006	3681	Akap8l	chr17	-	32333905	32334045	32333002	32333144	32334587	32334696	2,3,2	14,10,11	8,7,10	1,5,0	100	50	4.79E-07	6.17E-05	0.067,0.13,0.083	0.8,0.412,1.0	-0.644	140	FALSE	FALSE	PTC	chr17:32333905-32334045
61	MSTRG.15052	10687	Copg2	chr6	-	30830012	30830132	30826137	30826271	30858737	30858895	0,0,0	3,5,9	1,1,3	0,1,0	100	50	5.50E-07	6.99E-05	0.0,0.0,0.0	1.0,0.333,1.0	-0.778	120	B1_Mus2	FALSE	inframe_ins	chr6:30830012-30830132
62	MSTRG.4033	7743	Prps1l3	chr12	+	57234580	57234688	57230453	57230611	57238362	57242236	3,4,0	4,6,3	4,6,9	0,2,0	100	50	5.76E-07	7.23E-05	0.273,0.25,0.0	1.0,0.6,1.0	-0.692	108	B1_Mur3	FALSE	5_UTR	chr12:57234580-57234688
63	MSTRG.4033	7743	Mipol1	chr12	+	57234580	57234688	57230453	57230611	57238362	57242236	3,4,0	4,6,3	4,6,9	0,2,0	100	50	5.76E-07	7.23E-05	0.273,0.25,0.0	1.0,0.6,1.0	-0.692	108	B1_Mur3	FALSE	5_UTR	chr12:57234580-57234688
64	MSTRG.7073	4142	Dnm1l	chr16	-	16318444	16318477	16316601	16316778	16318763	16318841	25,15,18	6,4,6	5,7,4	0,0,0	83	50	6.20E-07	7.70E-05	0.715,0.693,0.644	1.0,1.0,1.0	-0.316	33	FALSE	TRUE	alternative_coding_exon	chr16:16318444-16318477
65	MSTRG.7538	3714	Ttc3	chr16	+	94384345	94384388	94383911	94384066	94384783	94384934	29,25,30	4,7,4	26,29,31	1,0,0	93	50	6.37E-07	7.84E-05	0.796,0.658,0.801	0.933,1.0,1.0	-0.226	43	FALSE	TRUE	alternative_coding_exon	chr16:94384345-94384388
66	MSTRG.7986	3696	Dnah8	chr17	+	30754567	30754735	30752821	30752944	30754973	30755187	0,0,0	7,3,6	7,7,1	5,5,0	100	50	7.21E-07	8.70E-05	0.0,0.0,0.0	0.412,0.412,1.0	-0.608	168	B1_Mus1	FALSE	PTC	chr17:30754567-30754735
67	MSTRG.2978	11411	Ap2b1	chr11	+	83345050	83345139	83342599	83342698	83346647	83346907	0,0,0	15,13,24	5,1,2	2,0,1	100	50	7.78E-07	9.29E-05	0.0,0.0,0.0	0.556,1.0,0.5	-0.685	89	PB1D9	FALSE	PTC	chr11:83345050-83345139
68	MSTRG.19625	67	Gm35612	chrX	-	50560808	50561006	50557576	50558384	50562679	50562954	12,18,10	2,2,5	1,7,7	0,0,0	100	50	1.21E-06	0.00013995	0.75,0.818,0.5	1.0,1.0,1.0	-0.311	198	FALSE	FALSE	lncRNA_cryptic_exon	chrX:50560808-50561006
69	MSTRG.3298	6277	Brca1	chr11	-	101532919	101533010	101532019	101532159	101533938	101534027	0,1,2	24,18,11	7,6,2	1,9,0	100	50	1.22E-06	0.00014006	0.0,0.027,0.083	0.778,0.25,1.0	-0.639	91	B1_Mus1	FALSE	PTC	chr11:101532919-101533010
70	MSTRG.1229	6751	Katna1	chr10	+	7744026	7744166	7741407	7743736	7748775	7748897	0,1,2	17,16,16	10,7,4	5,6,17	100	50	1.24E-06	0.00014076	0.0,0.03,0.059	0.5,0.368,0.105	-0.295	140	FALSE	TRUE	PTC	chr10:7744026-7744166
71	MSTRG.13059	9527	Yars	chr4	+	129202779	129202845	129200725	129200855	129206128	129206209	0,3,2	28,26,28	6,5,7	13,8,10	100	50	1.54E-06	0.00016678	0.0,0.055,0.034	0.188,0.238,0.259	-0.199	66	FALSE	FALSE	PTC	chr4:129202779-129202845
72	MSTRG.7986	10182	Dnah8	chr17	+	30754659	30754735	30752821	30752944	30754973	30755187	0,0,0	7,3,6	7,6,1	5,5,0	100	50	1.58E-06	0.000169	0.0,0.0,0.0	0.412,0.375,1.0	-0.596	76	B1_Mus1	FALSE	PTC	chr17:30754659-30754735
73	MSTRG.18906	1391	Myo5a	chr9	+	75192383	75192464	75189899	75190000	75193938	75194137	4,0,5	3,2,1	10,11,10	0,0,0	100	50	1.94E-06	0.00020339	0.4,0.0,0.714	1.0,1.0,1.0	-0.629	81	FALSE	TRUE	alternative_coding_exon	chr9:75192383-75192464
74	MSTRG.10407	7907	Ctnnd1	chr2	-	84605183	84605246	84603648	84603752	84608328	84608416	30,23,20	8,12,5	20,27,22	1,0,2	100	50	2.37E-06	0.00024063	0.652,0.489,0.667	0.909,1.0,0.846	-0.316	63	FALSE	TRUE	alternative_coding_exon	chr2:84605183-84605246
75	MSTRG.17005	901	Ndufab1	chr7	-	122101227	122101424	122096608	122096731	122101654	122101810	0,0,0	20,27,25	10,1,2	4,2,14	100	50	2.59E-06	0.00025799	0.0,0.0,0.0	0.556,0.2,0.067	-0.274	197	B1_Mm	FALSE	PTC	chr7:122101227-122101424
76	MSTRG.19423	192	Bcor	chrX	-	12055638	12055692	12054968	12055155	12056343	12059190	21,20,22	1,3,6	19,10,17	0,0,0	100	50	2.63E-06	0.00025966	0.913,0.769,0.647	1.0,1.0,1.0	-0.224	54	FALSE	TRUE	alternative_coding_exon	chrX:12055638-12055692
77	MSTRG.1392	4556	Cep57l1	chr10	-	41728627	41728714	41723860	41723938	41729330	41729408	6,3,4	12,4,4	3,3,2	1,0,0	100	50	2.84E-06	0.00027582	0.2,0.273,0.333	0.6,1.0,1.0	-0.598	87	FALSE	TRUE	alternative_coding_exon	chr10:41728627-41728714
78	MSTRG.17559	4535	Pcm1	chr8	+	41327805	41327910	41326623	41327599	41328768	41329025	0,0,0	20,22,34	2,4,3	6,5,4	100	50	3.44E-06	0.00032955	0.0,0.0,0.0	0.143,0.286,0.273	-0.234	105	ID_B1	FALSE	PTC	chr8:41327805-41327910
79	MSTRG.14257	6431	Ccdc18	chr5	+	108197136	108197229	108195995	108196159	108197756	108197970	0,0,0	1,6,2	3,2,3	1,2,4	100	50	3.49E-06	0.00033149	0.0,0.0,0.0	0.6,0.333,0.273	-0.402	93	ID_B1	FALSE	PTC	chr5:108197136-108197229
80	MSTRG.2499	11058	Hspa4	chr11	-	53273079	53273198	53272726	53272878	53275142	53276505	0,0,1	87,75,98	2,5,5	16,18,17	100	50	3.75E-06	0.00035021	0.0,0.0,0.005	0.059,0.122,0.128	-0.101	119	B1_Mus1	FALSE	3_frameshift	chr11:53273079-53273198
81	MSTRG.3659	7701	Foxk2	chr11	+	121266855	121266972	121260264	121260522	121285421	121285616	0,0,1	9,4,2	10,2,1	7,2,2	100	50	3.72E-06	0.00035021	0.0,0.0,0.2	0.417,0.333,0.2	-0.25	117	B1_Mur2	TRUE	inframe_ins	chr11:121266855-121266972
82	MSTRG.15184	9013	Snx10	chr6	+	51584490	51584611	51580301	51580400	51588215	51588428	0,1,0	16,7,11	1,5,6	8,5,4	100	50	4.19E-06	0.00038907	0.0,0.067,0.0	0.059,0.333,0.429	-0.251	121	ID_B1	TRUE	PTC	chr6:51584490-51584611
83	MSTRG.4542	12247	Pfkp	chr13	-	6614472	6614565	6605635	6605728	6614703	6614799	24,14,11	161,134,103	19,18,11	47,38,39	100	50	4.36E-06	0.00039822	0.069,0.05,0.051	0.168,0.191,0.124	-0.104	93	FALSE	FALSE	PTC	chr13:6614472-6614565
84	MSTRG.3269	5784	Atp6v0a1	chr11	+	101049458	101049476	101048417	101048525	101055025	101055143	1,2,4	3,1,3	2,4,1	0,1,0	68	50	5.19E-06	0.00047033	0.197,0.595,0.495	1.0,0.746,1.0	-0.486	18	FALSE	TRUE	alternative_coding_exon	chr11:101049458-101049476
85	MSTRG.19454	177	Usp9x	chrX	+	13072214	13072325	13053234	13053350	13093923	13094177	21,32,19	4,1,5	5,10,16	0,0,0	100	50	6.00E-06	0.00052625	0.724,0.941,0.655	1.0,1.0,1.0	-0.227	111	B1_Mur1	FALSE	PTC	chrX:13072214-13072325
86	MSTRG.19140	237	Ip6k1	chr9	+	108012600	108012614	108002720	108002822	108024102	108024450	0,0,0	26,22,28	2,3,3	7,4,6	64	50	6.32E-06	0.00054823	0.0,0.0,0.0	0.182,0.369,0.281	-0.277	14	ID_B1	FALSE	5_UTR	chr9:108012600-108012614
87	MSTRG.13190	9304	Rsrp1	chr4	+	134925776	134926150	134925292	134925432	134926734	134926818	11,15,6	12,13,21	20,15,17	2,4,5	100	50	7.49E-06	0.00064042	0.314,0.366,0.125	0.833,0.652,0.63	-0.437	374	FALSE	FALSE	PTC	chr4:134925776-134926150
88	MSTRG.7355	10326	Dppa4	chr16	+	48289616	48289827	48289327	48289378	48291013	48291583	1,0,1	29,23,22	4,1,5	6,5,7	100	50	7.54E-06	0.00064042	0.017,0.0,0.022	0.25,0.091,0.263	-0.188	211	B1_Mus2	FALSE	PTC	chr16:48289616-48289827
89	MSTRG.11084	9831	Gm11457	chr2	-	164762543	164762881	164759228	164759340	164767712	164768263	0,4,3	7,0,3	7,4,6	0,0,1	100	50	8.64E-06	0.00072872	0.0,1.0,0.333	1.0,1.0,0.75	-0.472	338	FALSE	FALSE	lncRNA_cryptic_exon	chr2:164762543-164762881
90	MSTRG.13190	9302	Rsrp1	chr4	+	134925776	134925821	134925292	134925432	134926025	134926818	13,18,9	11,15,7	23,13,22	2,4,0	95	50	8.93E-06	0.00074771	0.383,0.387,0.404	0.858,0.631,1.0	-0.438	45	FALSE	FALSE	PTC	chr4:134925776-134925821
91	MSTRG.10208	9729	Orc4	chr2	-	48921713	48921803	48917164	48917338	48927095	48927247	0,0,1	5,1,15	3,5,1	2,0,2	100	50	9.31E-06	0.0007692	0.0,0.0,0.032	0.429,1.0,0.2	-0.532	90	FALSE	FALSE	PTC	chr2:48921713-48921803
92	MSTRG.11277	2277	Gm14403	chr2	+	177501742	177501797	177497312	177498317	177506908	177507035	0,1,2	3,0,2	2,2,6	0,1,0	100	50	9.50E-06	0.00077909	0.0,1.0,0.333	1.0,0.5,1.0	-0.389	55	FALSE	FALSE	lncRNA_cryptic_exon	chr2:177501742-177501797
93	MSTRG.331	2708	Nbeal1	chr1	+	60211366	60211434	60206333	60206416	60222066	60222373	0,0,0	6,8,9	0,4,7	6,5,4	100	50	9.59E-06	0.00078173	0.0,0.0,0.0	0.0,0.286,0.467	-0.251	68	FALSE	FALSE	3_frameshift	chr1:60211366-60211434
94	MSTRG.4216	5673	Numb	chr12	-	83797196	83797343	83794060	83796154	83799507	83799657	3,4,13	1,5,1	1,3,1	0,0,0	100	50	9.91E-06	0.00079983	0.6,0.286,0.867	1.0,1.0,1.0	-0.416	147	FALSE	TRUE	alternative_coding_exon	chr12:83797196-83797343
95	MSTRG.6048	2875	Naa16	chr14	-	79350688	79350830	79344894	79345108	79351372	79351501	5,5,5	4,1,5	2,4,8	0,1,0	100	50	1.00E-05	0.00079983	0.385,0.714,0.333	1.0,0.667,1.0	-0.412	142	FALSE	TRUE	PTC	chr14:79350688-79350830
96	MSTRG.12555	9969	Ctnnal1	chr4	-	56815802	56815991	56813829	56813886	56816180	56816229	0,0,0	10,15,4	2,10,1	10,9,9	100	50	1.04E-05	0.00082209	0.0,0.0,0.0	0.091,0.357,0.053	-0.167	189	FALSE	FALSE	PTC	chr4:56815802-56815991
97	MSTRG.3148	10709	Snx11	chr11	-	96769831	96769943	96766614	96769322	96770638	96770851	1,0,0	3,8,12	4,3,2	4,2,3	100	50	1.09E-05	0.0008414	0.143,0.0,0.0	0.333,0.429,0.25	-0.29	112	B1_Mur4	FALSE	PTC	chr11:96769831-96769943
98	MSTRG.10324	2693	Ubr3	chr2	+	69947168	69947304	69945013	69945147	69948423	69949534	0,0,1	14,19,10	6,5,1	4,12,8	100	50	1.12E-05	0.00086234	0.0,0.0,0.048	0.429,0.172,0.059	-0.204	136	FALSE	FALSE	PTC	chr2:69947168-69947304
99	MSTRG.3755	11768	Nol10	chr12	+	17364168	17364358	17362594	17362680	17368308	17368387	0,0,0	10,19,16	1,5,3	2,6,9	100	50	1.15E-05	0.00087566	0.0,0.0,0.0	0.2,0.294,0.143	-0.212	190	B1_Mur1	FALSE	PTC	chr12:17364168-17364358
100	MSTRG.10016	7890	Sptan1	chr2	+	30001099	30001159	29999910	30000062	30002230	30002429	49,48,35	67,73,70	88,94,98	48,89,68	100	50	1.22E-05	0.00092462	0.268,0.247,0.2	0.478,0.346,0.419	-0.176	60	FALSE	TRUE	alternative_coding_exon	chr2:30001099-30001159
101	MSTRG.10683	5186	Ctdspl2	chr2	+	121995920	121996006	121993345	121993432	121997431	121998202	0,1,0	3,5,3	4,8,2	1,5,6	100	50	1.27E-05	0.0009541	0.0,0.091,0.0	0.667,0.444,0.143	-0.388	86	FALSE	TRUE	alternative_coding_exon	chr2:121995920-121996006

Supplemental Table T3: primers and antibodies used in this study

Construction of targeting and gRNA-expressing vector

left arm (SbaI&AscI*mutSfiI&I)

right arm (NotI&mlu*SfiI&I)

gRNA_left_FW
gRNA_left_RV
gRNA_right_FW
gRNA_right_RV

Primers used to check recombinant ES cells

5S_FW
5S_RV
random_FW
random_RV
chrome+_FW

Primers used for genotyping of 4.03H KO mice/ES cells

4.03H_WT_FW
4.03H_WT_RV
4.03H_KO_FW
4.03H_KO_RV

Primers used to prepare probes for Southern blotting (KO check)

3' probe (1) FW
3' probe (1) RV
3' probe (2) FW
3' probe (2) RV

LNA probe to detect 4.03H DNA probe to detect SL1 sites(1)

4.03H tandem-expression vector insert (EcoRI-SbaI-SalI-HindIII-BsamHI-XL4.03H gene)-BglII-BsaRI)

Primers used for quantification of 4.03H by qPCR

4.03H FP
4.03H RP
4.03H PP

Primers used to detect alternative splicing of target genes

Stx2/sa-b2 FW
Stx2/sa-b2 RV
Stxg8 FW
Stxg8 RV
Stxbp1 FW
Stxbp1 RV
Stxbp1 FW
Stxbp1 RV
Ap1 mu1a FW
Ap1 mu1a RV
Ap1 mu1b FW
Ap1 mu1b RV

Biotinylated probes for RNA-pulldown experiments

SL1-4.03H
SL1-WT

Antibodies used for western blot and immunohistochemical analyses

Rbfox3
Hu-mAb
Hu-pAb
Stxbp1
Stxg8
Nanog

[Fig.1A] A Minigene by inserting the cryptic exon of Stx2/sa-b2 containing exB1 (ex-B1-Stx2/sa-b2)

Stx2/sa-b2 exonB.5 MUT

4.03H WT

4.03H MUT

[Fig. 2B] Primers for spli minigene

[Fig.2C] Mutated 4.03H/exB1 -Stx2/sa-b2 (Mutations in lower letters)

4.03H WT

4.03H MUT

[Fig.4B]. deletion series of 4.03H RNA

4.03H RNA del_1

4.03H RNA del_2

4.03H RNA del_3

4.03H RNA del_4

4.03H RNA del_5

4.03H RNA del_6

4.03H RNA del_7

4.03H RNA del_8

4.03H RNA del_9

4.03H RNA del_10

4.03H RNA del_11

4.03H RNA del_12

4.03H RNA del_13

4.03H RNA del_14

4.03H RNA del_15

4.03H RNA del_16

4.03H RNA del_17

4.03H RNA del_18

4.03H RNA del_19

4.03H RNA del_20

4.03H RNA del_21

4.03H RNA del_22

[Fig. 4H] SL1-WT-4.03H

SL1-WT-4.03H

[Fig. 6B&C]: A Minigene by inserting the cryptic exon of Stx2/sa-b2 & 2 copies of SE2 binding sites (exB1-Stx2/sa-b2-SE2)

exon tethering

intron tethering

[Fig. 6E] A Minigene by inserting the cryptic exon of CD55 (exSx-CD55)

[Fig. 7B]: A Minigene derived from PAS exon 6-7', SL1-4.03H-exPAS, and SL1-WT-exPAS

PAS exon 6-6-7

SL1-4.03H-exPAS

SL1-WT-exPAS

[Fig.7C]: A Minigene derived from MAPT exon 9+10+11, SL1-4.03H-exMAPT, and SL1-WT-exMAPT

MAPT_minigene

SL1-4.03H-exMAPT

SL1-WT-exMAPT

[Fig.7E-G]: A Minigene by inserting the exon S3 of mouse Dnajb9 gene (Ad1 Dbl0-ex03), SL1-4.03H-exDnaj, and SL1-WT-exDnaj

Ad1 Dbl0-ex03

SL1-4.03H-exDnaj03

SL1-WT-exDnaj03

[Fig.7E-G]. A Minigene by inserting the exon S3 of human DNB-gene (Ad1 DBD-ex03), SL1-4.03H-exDNB, and SL1-WT-exDNB

Ad1 DBD-ex03

SL1-4.03H-exDNB03

SL1-WT-exDNB03

SL1-4.03H-exDNB04

SL1-4.03H-exDNBv3

SL1-4.03H-exDNBv3-rc4

SL1-WT-exDNBv3-rc4

[Fig.4M] scRNAs for knockdown experiments.

Control
luciferase
BPTF
NONO