STRAP regulates alternative splicing fidelity during lineage commitment of mouse embryonic stem cells

Lin Jin1,2, Yunjia Chen3, David K. Crossman3, Arunima Datta1,2, Trung Vu1,2, James A. Mobley4, Malay Kumar Basu5, Mariangela Scarduzio6, Hengbin Wang7, Chenbei Chang8 & Pran K. Datta1,2

Alternative splicing (AS) is involved in cell fate decisions and embryonic development. However, regulation of these processes is poorly understood. Here, we have identified the serine threonine kinase receptor-associated protein (STRAP) as a putative spliceosome-associated factor. Upon Strap deletion, there are numerous AS events observed in mouse embryoid bodies (EBs) undergoing a neuroectoderm-like state. Global mapping of STRAP-RNA binding in mouse embryos by enhanced-CLIP sequencing (eCLIP-seq) reveals that STRAP preferably targets transcripts for nervous system development and regulates AS through preferred binding positions, as demonstrated for two neuronal-specific genes, Nnat and Mark3. We have found that STRAP involves in the assembly of 17S U2 snRNP proteins. Moreover, in Xenopus, loss of Strap leads to impeded lineage differentiation in embryos, delayed neural tube closure, and altered exon skipping. Collectively, our findings reveal a previously unknown function of STRAP in mediating the splicing networks of lineage commitment, alteration of which may be involved in early embryonic lethality in mice.
E
day mammalian embryogenesis is characterized by a series
of cell fate decisions that initiate pluripotent cells to transit
to cellular context-dependent lineage segregation, followed
by progressive assembly of early organs. At gastrulation of the
mouse embryo (E6.5–7.5), migrating epiblast cells give rise to the
body pattern, and stem/progenitor cells are committed to three
embryonic germ layers. The embryo undergoes organogenesis at
early-somite stages (E8.0–8.5) and further establishes a more
distinct embryonic configuration with a head, heart, limbs, and
spinal cord as early as E8.75/9.01.

Using single-cell RNA sequencing (sc-RNA seq), researchers
have profiled the transcriptome of mouse embryos at various
stages of early development from one-cell stage to mid-
gastrulation2–4. Recently, genome-wide epigenetic studies pro-
vide insight into patterns of epigenetic modulation and diver-
gence in mouse preimplantation embryos5,6. However, in the
early mouse embryo, the regulatory mechanisms at the post-
transcriptional level for orchestrating germ layer determination
and morphogenesis remain largely elusive.

For RNA transcripts, AS, a post-transcriptional event, results in
substantial proteomic expansion and is conserved in a function-
specific manner across vertebrate species7. The pre-mRNA splic-
ing process requires the dynamic assembly of RNA-binding
proteins (RBPs) into spliceosome machinery, a highly organized
intra-nuclear structure consisting of RBPs and small nuclear RNA
complexes (snRNAs), as well as other splicing factors, together
called small nuclear ribonucleoproteins (snRNPs)8,9. The U1, U2,
U4/U6, and U5 snRNPs are the main dynamic components of the
major spliceosome, which is responsible for removing most pre-
mRNA introns10,11.

We determine that the WD40 domain-containing protein
STRAP (also known as UNRIP12) promotes tumorigenicity13 and
maintains cancer stem-like cells14. STRAP is involved in intra-
cellular distribution of the survival motor neuron (SMN)
complex15,16 and in regulation of cap-independent translation of
viral mRNAs12. Besides these, to our knowledge, nothing is
known about the function of STRAP in AS during development.
Here, we show that STRAP interacts with components of
U2 snRNP and that its deficiency affects the splicing fidelity in
neuroectoderm-like cells. eCLIP-seq reveals that STRAP associ-
ates with a broad set of transcripts involved in nervous system
development. By use of Xenopus embryos, we have found that AS
sites recognized by STRAP are at highly conserved nucleotide
sequences throughout evolution. Thus, our study deciphers the
role of STRAP in modulating splicing programs associated with
lineage-specific commitment.

**Results**

**Substantial AS events occur during mouse early organogenesis.**

To delineate molecular features of mouse embryos from post-
gastrulation (E8.0) to early organogenesis (E9.0), we profiled
global transcripts expression using whole mouse embryos at these
two stages. At the E9.0 stage, there were substantial changes in
transcripts, including signature genes involved in the formation of
specialized organ systems (Supplementary Fig. 1a). In contrast,
key regulators of early development were enriched at the
E8.0 stage (Supplementary Fig. 1a). We thus identified groups of
differentially expressed targets at the onset of embryonic organ
formation.

Moreover, we identified 896 genes differentially expressed at
the isofrom level (Supplementary Data 1), as exemplified by Eil4,
Nin, and Lmna (Supplementary Fig. 1b), suggesting that variants
of genes contribute to transcriptional diversity in a developmen-
tally regulated manner. To understand AS patterns in the transitional stage (from E8.0 to E9.0), we performed percent
spliced-in (PSI) analyses of AS using the rMATS tool17. We
obtained 1264 AS events for 1035 protein-coding genes, including
spliced exon (SE), mutually exclusive exon (MXE), alternative 5’
splice site (AS’S), alternative 3’ splice site (AS’S), and retained
intron (RI) categories (Fig. 1a, b and Supplementary Data 2). The
regulated AS events mainly distributed in SE and MXE (38% and
41.4%, respectively) (Fig. 1b). Further, their PSI values had a
uniform distribution between enhanced and repressed splice
junctions (Fig. 1c). Gene Ontology (GO) function of AS genes
revealed diverse enriched ontologies, including actin binding, cell
division, brain and heart development, as well as other house-
keeping cellular processes (Fig. 1d). We also performed RT-PCR
to confirm several AS changes (Fig. 1e), finding a positive correlation with PSI values derived from RNA-seq (Supplemen-
tary Fig. 1c). Collectively, we uncover previously uncharacterized
AS signatures during mouse early organogenesis and established a
reference dataset for mammals.

STRAP has a regulatory role in mouse early organogenesis. We
recently employed zinc finger nuclease (ZFN) technology to
generate *Strap* deficient mice18. Although all *Strap*+/- male
and female mice did not display any phenotypic abnormalities, their
intercrosses yielded no homozygous offspring. We then collected
embryos at selected days of gestational development to determine
the stage at which *Strap*’/- embryos die. At E7.5 and E8.5, there
were no obvious differences in embryonic morphology between
wild type (WT) and mutant embryos (Fig. 2a). However, compar-
ed to WT littermates, E9.5 *Strap*’/- embryos were smaller with
delayed development, namely dilated heart cavities and no body
turning with truncated frontalasal regions (Fig. 2a), consistent
with a previous report of gene trap mutagenesis19. However, that
study reported only morphological changes; no mechanism was
provided for embryonic lethality at cellular or molecular levels.
Although E10.5 *Strap*’/- embryos were incompletely resorbed,
E11.5 or older *Strap*’/- embryos were not obtained (Fig. 2b),
indicating that one functional *Strap* allele is essential for mouse
embryonic development and that, after E9.5, a double mutant
leads to embryonic lethality. To study the spatial expression
pattern of *Strap* in mouse embryos, a whole-mount in situ
hybridization (WISH) assay was performed using E9.25–9.5
mouse embryos. Although STRAP was expressed throughout
the body, its expression was stronger in the developing brain, eyes,
limb buds, and neural tube (Supplementary Fig. 2a).

Next, we delineated the molecular cause of the *Strap*’/- lethal
phenotype. At E7.5, the expressions of primary germ layer
markers, including *Sox1*, *Brachury T*, and *Cer*, were not altered
upon *Strap* deletion (Supplementary Fig. 2b). *Strap*’/-mutants,
however, failed to express several early forebrain and midbrain
developmental markers (i.e, *Fgf8*, *Gsc*, *Otx2*, and *Shh*20,21) at E8.5,
(Fig. 2c). Other early brain markers (i.e, *En1* and *Hoxb1* for early
hindbrain22,23 and *Six3* for the rostral forebrain fate24), and early
cardiac and endoderm lineage markers (*Gata4* and *EpCAM*)
were not affected by *Strap* ablation (Supplementary Fig. 2c, d). In E9.5
*Strap*’/- embryos, expression of the later germ layer25–27 markers
was lower as compared to those in either *Strap*+/- or *Strap*’/-
counterparts (Fig. 2d). Together, these results indicate that *Strap*
deficiency initially has a negative impact on mouse embryo
forebrain and midbrain development at E8.5 and subsequently
causes a uniform and constant developmental delay from E9.5.
Immunohistochemical analysis also revealed that NESTIN was
largely absent from primitive neuroepithelia tubules in *Strap*’/-
teratoma compared to that in WT, characterizing an immature
neural tissue lacking neural progenitors (Fig. 2e). However, the
expression of other two germ layer markers was comparable
between groups (Fig. 2e).
STRAP interacts with components of spliceosomal machinery. STRAP is linked to the cytosolic and nuclear SMN complex, but its other functions in the nucleus have not been elucidated. To determine this, we performed co-immunoprecipitation (co-IP) experiments with antibodies against STRAP and IgG using nuclear extracts (NEs) from mouse ESCs (Supplementary Fig. 3a). Co-IP samples were then visualized on SDS-PAGE and identified by LC-MS/MS (Supplementary Fig. 3b and Supplementary Data 3). Multiple sub-complexes of the spliceosome were identified, including components of the major snRNPs (U1, U2, U4/U6, and U5) and some splicing factors (Fig. 3a). The ranked scores assigned by the COMPLEAT tool showed that 29.6% of hits (45 of 152) participated in various predicted networks between splicing protein interactions (Fig. 3b and Supplementary Fig. 3c). 

---

**Fig. 3a** Alternative splicing event types

- **SE**: Constitutive exon
- **MXE**: Alternative exon or region
- **A5’S**
- **A3’S**: Retained intron

---

**Fig. 3b** GO function for AS events

- Actinin binding
- Metal ion binding
- DNA binding
- Protein binding
- Actin binding
- Heart morphogenesis
- Cell division
- Chromatin binding
- Ubi-relative catabolic process
- DNA repair
- Metal ion binding
- Spindle pole
- Forebrain development
- Cell morphogenesis
- DNA binding
- Neuronal cell body

---

**Fig. 3c** GO function for AS events

- Actinin binding
- Metal ion binding
- DNA binding
- Protein binding
- Actin binding
- Heart morphogenesis
- Cell division
- Chromatin binding
- Ubi-relative catabolic process
- DNA repair
- Metal ion binding
- Spindle pole
- Forebrain development
- Cell morphogenesis
- DNA binding
- Neuronal cell body

---

**Fig. 3d** GO function for AS events

- Actinin binding
- Metal ion binding
- DNA binding
- Protein binding
- Actin binding
- Heart morphogenesis
- Cell division
- Chromatin binding
- Ubi-relative catabolic process
- DNA repair
- Metal ion binding
- Spindle pole
- Forebrain development
- Cell morphogenesis
- DNA binding
- Neuronal cell body
Fig. 1 Mouse embryo early organogenesis is associated with transcript isoform diversity. a Schematic display of five AS types. SE, skipped exon; MXE, mutually exclusive spliced exon; A5′SS/A3′SS, alternative 5′/3′ splicing site; RI, retained intron. b Violin plots representing distributions of statistically significant ΔPSI (percent spliced-in) values (ΔPSI = PSI(E9.0) - PSI(E8.0); |ΔPSI| > 0.1, FDR < 0.05) for different classes of AS events. Kernel density is shown as a symmetric curve. The lower and upper bounds of the embedded box represent the 25th and 75th percentile of the distribution, respectively. The horizontal line in the box represents the median. The lower and upper whiskers show minima and maxima, respectively. The numbers of events are shown below each plot. c Volcano plot showing the difference of AS events between E8.0 and E9.0 mouse embryos. ΔPSI is plotted against the -log10(FDR) value. The color points outside the two dashed gray lines in the plot represent the differentially expressed AS with statistical significance (|ΔPSI| > 0.1, FDR < 0.05). Genes with -log10(FDR) >= 20 are indicated. d GO analysis of alternatively spliced genes between E9.0 and E8.0 mouse embryos, showing the top five to six ranked terms. e Validation by RT-PCR for genes with SE events identified by rMATS. PSI values are shown below the gel pictures. The information of target exons is shown on the right panel. Empty box, constitutive exon; black box, skipped exon. The experiment was repeated three independent times with similar results. Gapdh was used as a loading control.

Fig. 2 Genetic deletion of Strap causes mouse embryonic lethality. a Representative litters of E7.5-10.5 embryos from intercrosses of B6 Strap+/− mice. Compared with a wild-type E9.5, a Strap−/− embryo shows size and morphological defects. For E7.5-E8.5 embryos, the experiment was repeated five independent times with similar results. For E9.5-E10.5 embryos, fourteen independent experiments were repeated and had similar results. For E7.5-8.5 and E9.5 Strap−/− embryos, scale bar equals 300 μm. For E9.5 WT and E10.5 embryos, scale bar equals 1 mm. ec, ectoplacental cone; cd, chorionic dome; fn, posterior neural folds; fb, front brain; fl, front limb; hl, hind limb; t, tail. b Histograms showing the survival percentages of WT, wild type; Het, heterozygous; Homo, homozygous. Data represent one of three independent experiments. c, d qRT-PCR was used to quantify the relative mRNA levels in E8.5 (c) and E9.5 embryos (d). P-values based on unpaired two-tailed Student’s t-test. Error bars indicate the mean ± SD from n = 5 biological replicates. WT, wild type; Het, heterozygous; Homo, homozygous. Data represent one of three independent experiments. e Paraffin embedded tissue sections from mouse ESC teratomas were stained with three germ-layer markers as indicated. The experiment was repeated two independent times with similar results (n = 3 biological replicates per group). Scale bar, 50 μm. Magnification, ×40.
The interactions between STRAP and some of the above-mentioned partners were confirmed by probing same immunoprecipitated NEs using indicated antibodies (Fig. 2c). Treatment with RNase led to unchanged or enhanced the association of STRAP with U2 proteins (such as U2A’, SF3A/3B subunits, and SR140), suggesting protein–protein interactions. There was no interaction, however, between STRAP and spliceosome disassembly protein DDX15 (Supplementary Fig. 3d). Of note, there was pull-down of TFIIIC110 and STRAP in the absence of DNase (Supplementary Fig. 3e), indicating that STRAP acts as a bridge to link the complex of protein(s) and co-transcriptional RNA:DNA hybrid structures. We will explore this possibility in future work. Immunofluorescence staining assays confirmed that STRAP co-localized with various spliceosomal components in the nucleus (Fig. 3d). These results suggest a function of STRAP in the spliceosomal complex.

Fig. 3 Intracellular STRAP binds with subunits of the spliceosome complex. a List of STRAP-binding proteins for spliceosome complex subunits. Anti-STRAP antibody was used to immunoprecipitate STRAP and co-precipitate its interacting proteins from mouse ESC nuclear extracts (NE). The binding partners were identified by LC-MS/MS. b Ranked curve for STRAP binding-partners (see method). Enriched complexes (colored dots) are those that meet the P-value cutoff (P < 5.012e−2). Top enriched complexes are indicated. c Interactions between indicated proteins and STRAP were confirmed by co-immunoprecipitation and Western blotting analyses. Samples were prepared as in (a) and treated with or without ribonuclease A (RNase A). 1% of lysates were loaded as input control. The levels of co-precipitated proteins after RNase treatment (relative to untreated control) are shown below the gel pictures. The experiment was independently repeated twice with similar results. d The localization of STRAP (green) and indicated proteins (red) in mouse ESC cells were analyzed by immunofluorescence. Cell nuclei were labeled by DAPI (blue) staining. Scale bar, 10 µm. Images were captured by a Keyence microscope. The experiment was repeated three times independently with similar results.

STRAP-mediated AS events are involved in mouse early neuroectoderm lineage commitment. Since loss of STRAP affects expression patterns of early brain region-specific genes in vivo, we established N2B27-induced EB models to direct ESCs into neuroectodermal fate30,31 and, later, to neuronal cell identity32. The numbers, size, and morphology of EBs derived from WT and Strap KO ESCs appeared similar over time.
(Supplementary Fig. 4a). In addition, the expression of STRAP was steady during differentiation of WT EBs (Supplementary Fig. 4b). To define the molecular features of developing EBs, quantitative PCR (qPCR) assays were performed over the course of time. At day 6-9, expression of neuroectoderm-related genes was comparable between the two groups, (Supplementary Fig. 4c). Brain developmental markers appeared as early as day 9, but, at day 11, most of them were low in Strap KO EBs as compared with WT counterparts (Supplementary Fig. 4d), consistent with in vivo studies (Fig. 2c). Together, the results suggest that STRAP functions during early neural differentiation. We thus conducted RNA-sequencing using 9-day-old EBs, in which WT and KO cells commit to the similar lineage. Global comparisons using principle components analysis (PCA) and hierarchical clustering analysis revealed that EB replicates in the same group had high reproducibility, and that their identity was close to previously reported neuroectoderm organoids\(^3\) than to other developmental derivatives.\(^4,5\) (Supplementary Fig. 4e, f). Therefore, our N2B27-induced EB model molecularly resembles the mouse embryonic germ layer specification and provides us an ideal model to dissect the role of STRAP in mouse early embryogenesis.

We next characterized transcripts and splicing variants with genome-wide profiling. Only 154 genes were altered in transcriptome profiling of Strap-KO ESCs relative to that of ESC WT cells (Supplementary Data 4), suggesting that STRAP is dispensable for mouse ESCs viability and identity.\(^18\) Loss of STRAP in 9-day-old EBs resulted in 146 genes upregulated and 261 genes downregulated (Supplementary Data 4). As determined by use of the rMATS tool, 454 classical AS events (in 397 genes) were detected in Strap-KO EBs at day 9 (Fig. 4a and Supplementary Data 5). Of all AS, SE events were most frequent (49.1%, 223 of 454) (Fig. 4a, b). Retained introns were also overrepresented (18.1%, 82 of 454) as compared to other types (Fig. 4a). Thus, STRAP appears to function in promoting both exon inclusion and skipping (Fig. 4a), implicating its unbiased regulation for SE. GO function analysis revealed that STRAP-regulated AS events are involved in protein catabolism, molecular binding, and cell motility (Fig. 4c). Specifically, the events included neural development, ubiquitin-related genes, and transcription factors (Supplementary Fig. 4g--i). To assess the accuracy of our analysis of AS, we measured 28 SEs using 9-day-old EBs (Supplementary Fig. 4k), demonstrating a strong correlation with splicing efficiency (Supplementary Fig. 4l).

By comparing splice sites scores in responsive and unresponsive SEs, we found that both STRAP-enhanced and STRAP-repressed exons have weaker 3′ splice site (ss) than those found in unresponsive SE (unpaired Wilcoxon test, Fig. 4d, e). In contrast, we observed unbiased strength in either the 5′ss of SE or their flanking sites (Fig. 4d, e). To examine how STRAP correlates with other splicing factors in terms of the splicing level, we used a public database\(^36\) and compared the properties with AS genes between STRAP and several core reported AS factors. STRAP was well correlated with SF3 family members and PRPF3 (Fig. 4f), suggesting the potential cooperation of STRAP with these proteins. Only a few AS outcomes were affected by deletion of STRAP, either in mouse ESCs (Supplementary Data 6, \(n = 130\), EB vs. ESC, \(p < 0.0001\), Fisher’s exact test) or in mouse intestinal-specific tissues (Supplementary Data 6, \(n = 210\), EB vs. cKO, \(p < 0.0001\), Fisher’s exact test), suggesting that STRAP mediates AS in a cell-type-specific manner. To explore overall mRNA isoforms regulated by STRAP during neuroectoderm differentiation, we reanalyzed RNA-seq data using the MAJIQ (modeling alternative junction inclusion quantification) tool\(^37\) to quantify both classic binary AS events and complicated local splicing variants (LSVs). We obtained 1462 altered LSVs (of 1163 genes) upon deletion of STRAP in 9-day-old EBs (\(|dPSI| > 0.1, p < 0.05\), Supplementary Data 7), as exemplified by Dnmt3b and Zkscan1 (Fig. 4g). Moreover, these results had a significant concordance with data from rMATS (Fig. 4h). Thus, these analyses revealed a previously unknown role of STRAP involving the transcriptional diversity at exon levels during a certain germ layer differentiation.

**STRAP displays genome-wide mRNA binding patterns.** We next asked whether STRAP binds to its target RNAs in vivo. To this end, we utilized eCLIP-seq, which yields high complex libraries at single nucleotide resolution with enhanced technical and biological reproducibility\(^38\). Radiolabeled STRAP-bound RNA complexes from mouse embryos and EBs were visualized (Fig. 5a). Stringent RNase I treatment markedly eliminated STRAP-RNA bands (Fig. 5a), suggesting that the robustness and specificity of the STRAP-RNA complexes identified by eCLIP. As shown in Supplementary Fig. 5a, reads density was highly correlated across replicate samples, suggesting that these STRAP-binding clusters were due to intrinsic biological functions instead of technical variations. We obtained, in total, 9686 STRAP high-confidence peaks located at 2047 individual transcripts on a genome-wide scale (Supplementary Data 8). Most STRAP-peaks were defined by certain exons and introns, as well as their intersections (Fig. 5b and Supplementary Fig. 5b). We next leveraged STRAP reads along all exon/intron (5′ss) and intron/exon (3′ss) borders. In comparison to the input, STRAP signals in embryonic samples were enriched within ~50 nucleotides upstream of the 5′ss and ~25 nucleotides downstream of the 3′ss (Fig. 5c). Similar binding patterns were also evident in EBs (Fig. 5d). Thus, the binding of STRAP in close proximity to both 5′ and 3′ of splicing sites suggests its role in demarcating exonic regions. In addition, these featured occupancies were largely dependent on their transcript abundance (Supplementary Fig. 5c), in agreement with a previous report for CLIP-seq data\(^39\). Notably, among these exonic hits, 964 peaks appeared on annotated alternative exons and 2112 peaks were associated with constitutive exons (within the 300 nt region either upstream or downstream of exons, Supplementary Data 9), thus providing evidence for its involvement of STRAP in the AS process.

To identify the binding site consensus sequences for STRAP, we mapped STRAP binding peaks and identified the top-ranking enriched motifs (Fig. 5e). The fractions of targeted sequences were comparable (Fig. 5e, % of Targets), implying that STRAP cooperates with other RNA-processing regulators to recognize different types of sequences. To verify the direct binding of STRAP with RNAs, we conducted RNA electrophoretic mobility shift assays (REMSA) using a probe targeting the 3′ UTR of the Nnat transcript, which encompasses two putative STRAP consensus binding sequences (Fig. 5e). STRAP binds to the WT sequence but not to the mutant (MUT) probe (Fig. 5f). The addition of excess cold probe competed effectively for binding during the assay (Fig. 5f). These results demonstrate the RNA binding specificity of STRAP. To illustrate the biological functions of the loci related to STRAP peaks, we performed GO analysis, revealing genes for nervous system development, cytoskeleton, centrosome, and spliceosomal complex (Fig. 5g). Examples of STRAP binding peaks for neural transcripts are shown in Supplementary Fig. 5d--f. These data suggest that transcriptome-wide binding sites are differentially bound by STRAP during early development of mouse embryonic brain.

**STRAP dynamically regulates neuronal exon splicing during EB differentiation.** We next determined if STRAP binding is linked to its regulated genes at AS levels. Using the annotated
regions. Figure 6b shows a general increase in STRAP occupancy thus mapped STRAP binding peaks to alternative exons and their STRAP-RNA interactions in STRAP-dependent AS events. We binding. These results prompted us to assess the distribution of developmentally regulates certain AS events through direct

STRAP binding (Fig. 6a), supporting our hypothesis that STRAP unique genes with AS events in 9-day-old EBs was linked to

types and non-classic types) and that 22.4% (319 of 1423) of events was associated with its regulated AS genes (both classic

genes list derived from merged eCLIP peaks as a reference
dataset, we found that 15.6% (319 of 2047) of STRAP binding events was associated with its regulated AS genes (both classic types and non-classic types) and that 22.4% (319 of 1423) of unique genes with AS events in 9-day-old EBs was linked to STRAP binding (Fig. 6a), supporting our hypothesis that STRAP developmentally regulates certain AS events through direct binding. These results prompted us to assess the distribution of STRAP-RNA interactions in STRAP-dependent AS events. We thus mapped STRAP binding peaks to alternative exons and their flanking constitutive exons, as well as surrounding intronic regions. Figure 6b shows a general increase in STRAP occupancy surrounding the 3′ss of STRAP-enhanced exons and neighboring intronic regions, in addition to 5′ss of upstream flanking exons. Similar trends were also evident in STRAP-repressed exons (Fig. 6c), indicating that STRAP collaborates with other splicing regulator(s) to cause the final splicing outcomes. Furthermore, we performed UV crosslinking and immunoprecipitation (RIP) with anti-STRAP antibody to examine three eCLIP-hit genes, Tcf7l2, Ctnnd1, and Fat1. qPCR profiles further confirmed that STRAP was associated with these targeted transcripts (Supplementary Fig. 6a–c). Upon deletion of STRAP, there was no change in their transcripts (Supplementary Fig. 6d–f), ruling out the possibility that STRAP interacts to target RNAs due to their transcriptional abundance. Together, these results indicate that STRAP selectively binds AS regions for their processing.
Fig. 4 Extensive AS events occur in response to Strap loss in lineage-committed EB cells. 

The WT EB model underwent neuronal differentiation after long-term N2B27 culture with ascorbic acid, as evidenced by the high percentage of CD24+/CD56+ neuronal cells at day 14 (Supplementary Fig. 6g) and gradually increased neurogenesis markers over time (Supplementary Fig. 6h). Compared to a WT parallel, KO EBs had lower levels of the CD24+/CD56+ subpopulation and unaltered expression patterns for selected neuronal genes (except for Pax6) along the differentiation (Supplementary Fig. 6g, h), indicating that EBs failed to undergo neuronal lineage upon deletion of Strap. We further assessed the terminal differentiation of EBs by assessing their electrophysiological properties. In whole-cell patch clamp recordings, we observed a small percentage of KO cells (20%) capable of producing immature single action potentials (APs) upon depolarization and a prominent sag upon hyperpolarization at day 12 compared to 40% of WT cells (Fig. 6d, e). In both genotypes, the percentages of cells capable of producing APs and sags increased with time, reaching 85% in WT vs. 25% in KO at day 22 (Fig. 6d, e). None of the cells in this phase of maturation produced trains of APs in response to depolarizing steps. However, a few WT cells at day 22 showed single or multiple “rebound” APs and spontaneous AP firing at resting membrane potentials (RMP) upon termination of hyperpolarizing pulses (examples in Fig. 6d), middle and bottom right panel, respectively), indicating a faster and more pronounced course of neuronal differentiation in WT than in KO cells.

During maturation of neuronal cells, stage-specific-switch exons in genes have distinct functions. We intersected eCLIP genes with rMATS-derived classic SE genes (FDR < 0.1) and thus found 98 hits (Fig. 6f). Here, we investigated exons from two murine genes, Nnat and Mark3. AS of neuronatin (Nnat) generates two spliced isoforms, depending on the usage of exon2. The short mutant has a more potent role in neural patterning than the full-length one. In Nnat, extensive STRAP-binding sites located on exon1 and exon3 (including 5’ and 3’ UTRs, respectively) were detected by use of embryonic samples (Fig. 6g). RIP-qPCR validated that STRAP peaks on 14-day-old WT EBs are strongly present on the indicated boundaries as compared to 11-day-old parallels (Fig. 6h). Little enrichment was evident in other regions (Supplementary Fig. 6i). These interactions were not caused by RNA abundance (Supplementary Fig. 6k). We also observed a lower exon 2 inclusion in WT EBs compared to its earlier parallels and KO groups (Fig. 6i, j), suggesting that STRAP temporally regulates exon2 inclusion in Nnat.

Another example is Mark3, which encodes a kinase involved in the phosphorylation of MAP2 and MAP4. Two spliced variants of Mark3 are differently expressed in neural progenitors (exon16 inclusion) and neuronal cells (exon16 skipping). In embryos, accumulated STRAP-binding peaks are located on Mark3 exon15, as determined by eCLIP (Fig. 6k). The transcripts with target sites were confirmed by RIP in 14-day-old WT EBs (Fig. 6l and Supplementary Fig. 6j), which were also coupled with more exon16 skipping relative to other parallels (Fig. 6m, n). Again, RNA products had no effect on these binding signals, as there were no altered Mark3 transcripts on the indicated days (Supplementary Fig. 6l).

To eliminate the possibility that SE events were caused by differentiation arrest rather than STRAP-dependent binding regulation, we developed a Tet-On inducible short hair RNA (shRNA) system against STRAP in mouse E14 ESC cells. Three-day treatment of Doxycycline (Dox) resulted in an efficient reduction of STRAP expression in 14-day-old EBs derived from E14 cells (Fig. 6o, left bottom). Compared with control groups, induced knockdown of STRAP produced more included exons for Nnat and Mark3 transcripts (Fig. 6o), consistent with observations with the KO EBs. These data suggest that, during EB differentiation, STRAP is involved in the selection of neural exons in a local concentration-dependent manner.

STRAP is involved in the assembly of 17S U2 snRNP proteins. As the RNase treatment did not interfere with STRAP binding to U2 snRNP components, indicating that these interactions depended on protein–protein contacts, we focused on the downstream effects of Strap knockdown. First, we measured the abundance of U2 proteins in whole-cell lysates extracted from ESCs and EBs and found higher levels of SF3A subunits and SR140 in STRAP KO 14-day-old EBs relative to WT parallels (Fig. 7a, b), suggesting that STRAP is involved in balancing homeostasis of U2 proteins at a specific stage. Similar trends were observed for Tet-On inducible shSTRAP EBs at day 14 (Supplementary Fig. 7a), ruling out the indirect effects caused by STRAP-KO impaired differentiation.

We next addressed the question of whether STRAP deletion affects formation of the U2 complex. Among U2 interactors, the SF3B complex is initially recruited to 12S core particles (the U2 snRNA, the Sm core proteins, and the U2A' and U2B' complex) to generate pre-cmature 15S complex, which subsequently incorporates SF3A subunits to form the mature 17S snRNP. Lastly, the functional U2 snRNP is recruited to 3’ss of the intron, together with the U1 complex at the 5’ss to form pre-spleosomes. To determine whether STRAP has effects on the assembly of the 17S U2 complex, we accomplished co-immunoprecipitations (co-IP) using an SF3B1 antibody on NEs from the indicated cells. In ESCs, interactions between U2 components were independent of STRAP (Fig. 7c, right panel). However, during EBs differentiation
loss of STRAP led to a dissociation between SF3B1 and two SF3A subunits, whereas the interaction between SF3B1 and U2A' was independent of STRAP expression (Fig. 7c, right panel; 7d). Together, the results suggest that STRAP developmentally regulates assembly of the 17S complex. We also tested the interaction between SR140 and CHERP, a binding partners among 17S U2 components. Although there was no change in CHERP expression (Fig. 7e, left panel), SR140 was less co-precipitated with CHERP in KO EBs relative to the WT group (Fig. 7e, right panel; 7f).

To map the binding domain(s) of STRAP responsible for its function, we generated deletion mutants of STRAP with Flag epitope tag and performed co-IP assays with HA-tagged SR140. The results revealed that the third WD-40 domain is necessary for the binding of STRAP with SR140 (Fig. 7g). To test whether
certain splicing variants could be regulated by the WD-40 domain (s) of STRAP, we performed gain-of-function assays in MEF cells using a mini-gene splicing tool, which contains UPF3A exon 4 in a SR140-dependent manner49. Compared to WT MEFs, the usage of exon 4 was significantly reduced in STRAP KO MEFs and could be partially rescued by overexpression of STRAP-Flag, STRAP (1–4)-Flag, and STRAP (3–4)-Flag mutants (Fig. 7h) and of SR140 (Supplementary Fig. 7b). Similar assays using other mini-gene tools in two gene contexts (Nnat and Ppp2r2d) also revealed that the third WD-40 domain of STRAP is essential for its splicing activity (Supplementary Fig. 7c, d).

**Discussion**

In this study, we show that, for mice, the deletion of *Strap* leads to early embryonic lethality and, in EBs, to abnormal differentiation (Fig. 2 and Supplementary Fig. 4c, d), indicating an essential role of STRAP in mouse early embryo development and differentiation of ESCs. Since the AS program tightly controls the post-transcriptional genes necessary for cell differentiation and development, we reason that these defects are caused, in part, by AS events upon Strap KO. The following evidence supports this concept: (i) Substantial AS patterns, instead of limited transcriptional regulation, are affected upon *Strap* KO in lineage-committed murine ESCs; (ii) The involvement of STRAP for the assembly of the 17S U2 snRNP complex suggests its role in AS modification; (iii) STRAP contributes to the selection of important neuronal exons during EB differentiation, as exemplified by *Nnat* and *Mark3; and (iv) *Strap*-MO in Xenopus embryos influences neuronal patterning and impairs the body axis.

Our work reveals that deletion of STRAP has effects mainly on exon skipping during the lineage commitment, which is distinct from the recently elucidated AS programs during the neural differentiation of human iPSCs, in which increased intron retention is a predominant feature of AS at early stage55. Among STRAP-mediated AS events, the binding typically takes place upstream of alternative exons (Fig. 6b, c), in agreement with the position-dependent splicing rule39,56. We also show the dynamics of STRAP enrichment on its RNA targets, thereby raising a possibility that it bridges splicesome complexes mediating AS recognition and sequential catalysis during development.

Various RBPs preferentially bind low-complexity motifs comprised primarily of one or two base types57,58, which is reflected in the abundance of GA-rich and CU-rich motifs for STRAP (Fig. 5e). A possible explanation is that these motifs facilitate cooperative binding of STRAP with other RBPs and enhance splice site recognition during exon ligation. In addition, distal intronic regions are reservoirs of highly conserved RNA cis elements necessary for splicing regulation59. Our RNA-binding map also reveals that some of STRAP peaks are located far from target exons (>200 nt). To date, the most commonly used models for protein-RNA binding and CLIP-seq experiments are stable cell...
lines or regionally dissected tissues. These models, however, are unable to identify, temporally, RBP targets with stage-specific switch exons during cell differentiation. Here, we present an EB-based model, which facilitates understanding of the RNA processes during development.

In sum, these studies define a genome-wide AS network that STRAP regulates in a coordinated manner. We reveal the position-dependent and context-dependent effects of STRAP on splicing regulation during neuroectodermal lineage commitment, which adds another layer of complexity to the cross-
regulatory networks between RBPs and elucidates the role of STRAP in RNA biogenesis.

Methods

Generation of Strap +/- mice. Mouse mutants with deletions were created using the previously described zinc finger nucleases (ZFNs) protocol. In brief, ZFNs targeting the Strap gene were electrointroduced into V6.5 ESCs. Strap+/+ or Strap−/− ESCs were subsequently intercrossed for the indicated experiments. Postimplantation embryos were isolated from naturally mated timed-pregnant mice. To genotype embryos, yolk sacs were lysed at 55 °C overnight in lysis buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM MgCl2, 0.45% NP40, and 0.45% Tween-20) supplemented with 1 µg/ml proteinase K (Invitrogen). The primers used for genotyping are listed in Supplementary Data 10. All animal experiments were carried out according to the guidelines for the care and use of laboratory animals of the University of Alabama at Birmingham Institutional Animal Care and Use Committee (IACUC). We have complied with all relevant ethical regulation for mice testing and research. Animals were housed at a controlled temperature (23 °C) and humidity (53%) under a 12:12 light-dark cycle and received standard mouse chow and water.

Derivation of mouse ESC lines. E3.5 blastocysts were collected from Strap+/+ females mated with Strap−/− males. ESC lines were derived from E3.5 blastocysts in defined medium (Knockout-D-MEM medium, 20% Knockout Serum Replacement, penicillin/streptomycin, 2 mM L-glutamine, 1x MEM non-essential amino acids, 100 mM 8-mercaptoethanol, and 1000x recombinant mouse LIF) and cultured on MEF-feeder cells. We generated a total of 24 independent cell lines: 5 wild type, 14 heterozygous, and 5 Strap−/−.

ESCs and EBs culture. All established ESC lines were routinely cultured on MEF-feeder cells in standard ESC culture medium (Knockout D-MEM medium, 15% ES-conditioned DMEM/F-12 medium and Neurobasal Medium, which was supplemented with fresh 0.5 mM DTT, protease inhibitors (Roche), and 0.2 mM PMF) on ice. Nuclear pre-mRNA was subjected to centrifugation (1000×g, at 4 °C) and then suspended in Buffer C (20 mM Tris-HCl (pH 7.9), 1.5 mM MgCl2, 0.2 mM EDTA, 250 mM NaCl, 0.2 µM DTT, 0.5% SDS, 0.5 mM DTT and protease inhibitors). Soluble nuclear proteins were separated by centrifugation (25000×g for 30 mins, at 4 °C) and dialyzed against Buffer D (20 mM Tris-HCl (pH 7.9), 50 mM KCl, 0.2 mM EDTA, 10% glycerol, 100 mM 8-mercaptoethanol) for 7 h. Samples were centrifuged at 16000×g for 30 mins at 4 °C and the supernatant was kept at −80 °C.

Nuclear extracts (NEs) preparation. Mouse ESCs (1 × 106) were homogenized in Buffer A (10 mM Tris-HCl (pH 7.9), 10 mM KCl, 10% glycerol, 1.5 mM MgCl2 supplemented with fresh 0.5 mM DTT, protease inhibitors (Roche), and 0.2 mM PMF) on ice. Nuclear NEs were used for in vitro transcription and RNA labeling using the DIG RNA Labeling kit and T3 RNA polymerase (Roche). Primer sequences are listed in Supplementary Data 10.

Immunoprecipitation. Mouse ESC NEs were treated either with RNase A (100 µg/ml) for 15 min or with DNase I (10 U) for 30 min at 37 °C. Immunoprecipitations were then performed with samples (with or without treatment) resuspended in standard immunoprecipitation lysis buffer (150 mM NaCl, 10 mM EDTA, 0.02% NaN3, 50 mM NaF and 0.5% Nonidet P-40) and rotated overnight with beads cross-linked anti-STRAP antibody (4 µg/mg lysates, Bethyl, A304-735A). Rabbit IgG (3 µg/mg lysates, Cell signaling, #2729) was used as a negative control. For NE samples from mouse ESCs or EBs, immunoprecipitations were performed using either beads cross-linked with anti-psiSE antibody (4 µg/mg lysates, Bethyl, A300-996A) or anti-CHEP antibody (crosslinked to anti-IgG antibody (4 µg/mg lysates, Santa Cruz, sc-100650).

Proteomics analysis. We independently performed the proteomics experiments twice with similar results. On each time, we had an NE-specific sample pulldowned by an anti-STRAP antibody (n = 1) and a parallel control sample precipitated by an anti-IgG antibody (n = 1). Complex proteins from each sample were separated by SDS Bis-Tris gels (4–12%, Invitrogen). The gels were stained with Sypro Ruby. Each lane was cut into 6 MW fractions, and each gel plug was equilibrated in 100 mM ammonium bicarbonate (AmBc). Each gel plug was then digested with Trypsin Gold (Promega) following the manufacturer’s instruction, and peptide extracts were reconstituted in 0.1% formic acid/ dMnO4 at ~0.1 µg/µl. Peptide digests were injected onto a 1260 Infinity nPLLC Stark (Agilent Technologies), and separated using a 75 micron ID, 0.15 µm pulled tip C-18 column (Jupiter C-18 300 Å, 5 micron, Phenomenex). This system ran in-line with a Thermo Orbitrap Velos Pro hybrid mass spectrometer equipped with a nano-electrospray source (Thermo Fisher Scientific), and all data were collected in CID mode. The nPLLC was configured with binary mobile phases that included solvent A (0.1%FA in dH2O2), and solvent B (0.1%FA in 13% dH2O / 85% ACN), programmed as follows: 10 min; 1% B (2 µL/min, load), 90 min at 5%–40%B (linear: 0.5 µL/min, analyze), 50 min at 70%B (2 µL/min, wash), 10 min at 5%B (2 µL/min, equilibrate). Following each parent ion scan (300-1200 m/z @ 60k resolution), fragmentation data (MS2) were collected on the top most intense 15 ions. For data dependent scans, charge state screening and dynamic exclusion were enabled with a repeat count of 2, repeat duration of 30 s, and exclusion duration of 90 s.
The XCalibur RAW files were further collected in profile mode, centroided and converted to MzXML using ReAdW (version 3.5.1). The mgf files were then created using MzXML2Search (included in TPP v. 3.5) for all scans. The data was searched using SEQUEST (v.27 rec12. dta files), which was set for two maximum missed cleavages, a precursor mass window of 20 ppm, trypsin digestion, variable modification C @ 57.0293, and M @ 15.9949. Searches were performed with a species specific subset of the UniRef100 database. The list of peptide IDs generated based on SEQUEST search results were filtered using Scaffold (version 3.0). The filter cut-off values were set with a minimum peptide length of >5 AA’s, with no MH + 1 charge states, with peptide probabilities of >80% C.I., and with the number of peptides per protein ≥ 2. The protein probabilities were then set to a > 99.0% C.I., and an FDR < 1.0%. STRAP binding-partners were ranked by a COMPLEAT online tool (http://www.flyrnai.org/compleat)\textsuperscript{29}. The enrichment P-value was calculated based on the permutation: generate 1000 random lists of input data then
calculate the P-value based on the distribution of the scores with random lists. Sub-networks for STRAP interaction with spliceosomal partners were also computed by this tool.

**Immunofluorescence staining.** Cells were fixed in 4% paraformaldehyde for 10 min at room temperature, blocked, and then incubated at 4 °C overnight with the following antibodies: SR140 (at the dilution of 1:150, Santa Cruz, sc-398718); SF3A2 (at the dilution of 1:250, Santa Cruz, sc-390444); SF3B1 (at the dilution of 1:150, Santa Cruz, sc-514655); SYF1 (at the dilution of 1:150, Santa Cruz, sc-271037); or STRAP (at the dilution of 1:200, Bethyl, A304-735A). After washing, cells were incubated with goat anti-rabbit Alexa Fluor 488 antibody (at the dilution of 1:150, Life Technologies, A-11080) or goat anti-mouse Alexa Fluor 555 antibody (at the dilution of 1:150, Life Technologies, A-21422) and counter-stained with DAPI to detect nuclei.

**Western blotting.** Proteins were separated by 10% SDS-PAGE and probed with primary antibodies. Primary antibodies, at the dilution of 1:1000 for each, included: SNRPA (Santa Cruz, sc-376027); U2A’ (Santa Cruz, sc-393804); SR140 (Santa Cruz, sc-398718); HELIC2 (Santa Cruz, sc-393170); PRPF3 (Santa Cruz, sc-101330); SYF1 (Santa Cruz, sc-9713); SF3B1 (Santa Cruz, sc-9713) (with exons 3–5); DRB1 (Santa Cruz, sc-390444); SF3B1 (Santa Cruz, sc-514655); STRAP (BD Transduction Labs, #611346); DXD15 (Santa Cruz, sc-271686); TFIIC110 (Santa Cruz, sc-81406); GAPDH (Cell Signaling, #2118), and CHERP (Santa Cruz, sc-100650). β-Actin (Sigma, A5316) was used at the dilution of 1:10,000.

**Glutathione S-transferase (GST)-fused protein purification and RNA EMSA.** Full-length STRAP was PCR amplified from previously reported pcDNA3-STRAP-Flag vector and subcloned into pGEX-4T-1 GST expression vector using BamHI and XhoI restriction sites. The induction was performed by adding 0.4 mM isopropyl-β-D-thiogalactopyranoside at 30 °C for 3.5 h. Whole bacterial lysates were applied to a glutathione Sepharose 4B (GE Healthcare Life Science, 4500193) column, and GST-tagged proteins were purified according to the manufacturer’s instructions. To perform RNA-EMSA and supershift analyses, a Chemiluminescent Immunoprecipitation System (Invitrogen) was used. The signal intensities of the bands were quantified by the ImageJ program. Primers used are listed in Supplementary Data 10.

**Expression plasmds.** The truncated coding regions from the full-length STRAP-Flag vector were inserted into the pcDNA3.0 Flag vector using BamHI and XhoI restriction sites. The primers used for subcloning are listed in Supplementary Table 10. Plasmid SR140-HA was a gift from Dr. Ying Feng (Chinese Academy of Sciences, #2118). For co-IP assays, HepG2 cells were transfected with various plasmids using Lipofectamine3000 (Thermo Fisher Scientific), and lysates were subjected to co-immunoprecipitation with an anti-HA antibody (3 μg/ml lysates, Bethyl, A190-208A) and probed with either an antibody for anti-Flag (at the dilution of 1:15,000, Sigma, F3165) or anti-HA (at the dilution of 1:1,000, Bethyl, A190-208A).

**Mini-gene reporter assays.** Nnat and Pprr2r2d mini-genes were constructed by amplifying genomic sequences spanning exons 1–3 of Nnat and exons 3–5 of Pprr2r2d respectively, which were then cloned into pcDNA3.1 vectors (BamHI and XhoI restriction sites for Nnat, KpnI, and EcoRI restriction sites for Pprr2r2d). The primers used for subcloning are listed in Supplementary Table 10. The UPP3A gene was kindly provided by Dr. Ying Feng. Mini-gene vector and STRAP truncating variants were co-transfected into MEF cell lines using Lipofectamine3000 reagent and the cells were cultured for a further 48 h until RNA isolation.

**Inducible STRAP knockdown in mouse ESCs.** For RNA interference in mouse E14 ESCs, an shRNA against Strap was cloned into pLKO-Tet-On (Agel and EcoRI sites). The target sequences of shRNA was as follows: GCCAGGGATATTTACCATTATGTTTCCAGA. For the knockdown experiment, pLKO-Tet-On-based lentiviral vectors and packaging plasmids pMD2.G and psPAX2 were co-transfected into 293T cells using Lipofectamine 3000 reagent. The supernatant was collected after 48 h and passed through a 0.45 μm filters (Millipore). E14 cells were cultured in the viral supernatant in the presence of 5 μg/ml polybrene (Sigma) for 48 h. After infection of cells, selection was started by adding 4 μg/ml of puromycin to E14 ESCs. ESCs were then formed to EBs and underwent differentiation in N2B27 medium. shSTRAP expression was induced by the addition 1 μg/ml of Dox to the culture medium of 11-day-old EBs.

**Patch clamp recordings and flow cytometry.** Whole cell patch-clamp recordings in current-clamp mode were performed with a borosilicate microelectrode filled with a potassium gluconate-based internal solution. EBs were held at resting membrane potential (RMP) and 500 ms current steps from −200 pA to 700 pA were delivered. Single cells digested from EBs were collected at time points and stained with CD24-FTC (2 μl/test, MACS, #130-110-825) and CD56-APC (2 μl/test, R&D systems, FAB7820A) antibodies followed by flow cytometry. Becton Dickinson FACSDiva (version 8.0) analyzed the positive population.

**RNA-seq library preparation and sequencing.** For analysis of differential gene expression, samples were lysed with Trizol, treated with DNase (QIAGEN) and purified using RNeasy Minielute Cleanup Kit (QIAGEN). Ribosomal RNA was removed from each RNA extraction using Ribo-Zero Gold RNA Removal kit (Illumina). Indexed libraries were pooled and sequenced at a final concentration of 1.8 pmol/μl on Illumina NextSeq 500 using paired-end chemistry with a 75-bp read length. Cutadapt (version 2.2) was used to trimming primer adapters from raw FASTQ files. Sequencing reads were mapped to Gencode GRCh38 p4 Release M11 using STAR version 2.5.2b (options:--outUnmapped Fastx;--outSAMtype Coordinate;--outSAMAttribute All). Transcript abundances were calculated using Cufflinks version 2.2.1 with options:--library-type fr-firststrand; --G -l Cuffmerge was then used to merge the transcript files from Cufflinks into one file. Following Cuffmerge, Cufquant was used to quantify the
transcript abundances, followed by differential gene expression using Cuffdiff. Differentially expressed genes with a P value < 0.01, as well as a log2 fold change >1 were further analyzed further.

Splicing analysis from RNA-seq data. AS was inferred from the RNA-seq data using the Multivariate Analysis of Transcript Splicing (rMATS) version 3.2.5 software for replicates, which processed the above STAR alignments (in BAM format) to provide the AS events\(^\text{17}\). Parameter values for rMATS were set by optimizing the accuracy of the MATS estimation of splicing difference of known standards (genes previously analyzed by RT-PCR). The final rMATS analysis of STRAP-responsive alternative splicing was accomplished with options -b1, -b2, -gtf, -t paired, -len 74, and -novSS. The output from rMATS was then filtered using the criteria of: FDR < 0.05 and absolute value of ΔPSI ≥ 10% (unless otherwise noted). MAJIQ and Voda (https://biociphers.bitbucket.io/majiq/index.html) were used to detect\(^\text{37}\), quantify and visualize local splicing variations (LSVs) from the RNA-Seq data. Briefly, the MAJIQ build tool used the alignment BAM files from STAR along with the gene annotation file (GFF3) to de

In the presence of STRAP

Neuroectoderm lineage

ESC

\(\text{5' ss} \quad \text{3' ss} \quad \text{S} \quad \text{17S U2} \quad \text{Splicesome} \)

In the absence of STRAP

Defective differentiation

ESC

\(\text{5' ss} \quad \text{3' ss} \quad \text{S} \quad \text{17S U2} \quad \text{Splicesome} \)
Fig. 8 Loss of STRAP triggers aberrant AS across species. a, b Strap-MO (50 ng) was injected alone or with Myc-Strap (1 ng) into the marginal zone region of two dorsal blastomeres in 4-cell-stage Xenopus embryos. The embryos were cultured to the neurula stage; representative embryos are shown. The experiment was repeated three independent times with similar results. MO, morpholin ogonucleotides. Scale bar, 1 mm. c Evolutionary conservation of sequence (red) across vertebrates within the putative splicing regions in the Cep57 and CkI genes. Extracted splicing sequences are referred to by MaxEntScan definition. d RT-PCR analysis of splicing in the presence of Noggin RNA with or without Strap antisense MO in Xenopus early tailbud stage. Quantifications of PSI values are shown below the gel. e Fraction of mis-splicing calculated from one of three independent analyses is shown in the bar graph. Error bars indicate the mean ± SD from n = 3 biological replicates. P-values were determined by unpaired two-tailed t-tests. f CkI splice-MO (50 ng) was injected into the dorsal animal regions of 4-cell to 8-cell stage Xenopus embryos. RT-PCR analysis of splicing with CkI splice-MO in Xenopus at the early tailbud stage. Quantifications of the PSI values are shown below the gels. The experiment was repeated three independent times with similar results. g Embryos were cultured until the tailbud or tadpole stages; representative embryos are shown here. The experiment was repeated three independent times with similar results. h Model describing the role of STRAP in mediating AS patterns. Upper, under conditions of normal levels of STRAP, it cooperates with the 17S U2 snRNP complex to regulate certain splicing sites, resulting in a balance of altered splicing patterns during neuroectoderm lineage commitment. Bottom, loss of STRAP has a negative effect on assembly of the 17S U2 complex. The fidelity of pre-mRNA splicing is also disrupted, causing high or low alternative transcripts or retaining ones. Blue box: constitutive exon; Yellow box: alternative exon; s: STRAP; 175 U2: 175 U2 snRNP complex; 155 U2, 155 U2 snRNP complex; 5/3′ ss: 5/3′ splicing site. The schematic diagram was created by the authors.

**STRAP-RNA binding assay.** Nucleotide resolution UV crosslinking at 254 nm (400 mJ/cm²) and immunoprecipitation were performed. Two replicates from two independent WT EB or embryos were used for generating protein-RNA binding samples. Lysates generated from the crosslinked cells were treated with Turbo DNase (Ambion) and RNase I (1: 50 for high or 1:500 for low dilution, Ambion) for 5 min at 37°C to digest the genomic DNA and trim the RNA to short fragments of an optimal size range. RNA-protein complexes were immunoprecipitated with 100 µl of protein A Dynabeads (Life Technologies) and 10 µg of anti-STRAP antibody. Following stringent high salt washes (50 mM Tris-HCl; 1 M NaCl, 1 mM EDTA; 1% NP-40, 0.5% Tergitol 5, 0.5% sodium deoxycholate), the immunoprecipitated RNA was 5′-end labeled with radioactive [γ-32P]ATP62. The immunoprecipitated complexes were separated by SDS-PAGE and transferred to a nitrocellulose membrane followed by the exposure to a Fuji film at −80°C. eCLIP correlation analysis. To evaluate the biological reproducibility of eCLIP data, pairwise comparisons were performed to use 1.4–2.3 × 10⁵ uniquely mapped reads without PCR duplicates. All reads with read depths >3 were obtained using Samtools (version 1.3.1) and in-house Perl (version 5.26.3) scripts. The resulting log2 (Read depth) values were then plotted with the ggplot2 (version 3.2.1) package in R, followed with calculation of the Pearson correlation coefficient.

**Establishment of eCLIP-seq libraries.** eCLIP libraries were generated based on the standardized eCLIP experimental protocol63 with minor modifications. Briefly, UV-crosslinked WT (50 ng) or 400 ng; 400 ng; 400 ng; 0.3 ng) embryos were used for generating STRAP-RNA binding assay. Fig. 8 Loss of STRAP triggers aberrant AS across species. a, b Strap-MO (50 ng) was injected alone or with Myc-Strap (1 ng) into the marginal zone region of two dorsal blastomeres in 4-cell-stage Xenopus embryos. The embryos were cultured to the neurula stage; representative embryos are shown. The experiment was repeated three independent times with similar results. MO, morpholin ogonucleotides. Scale bar, 1 mm. c Evolutionary conservation of sequence (red) across vertebrates within the putative splicing regions in the Cep57 and CkI genes. Extracted splicing sequences are referred to by MaxEntScan definition. d RT-PCR analysis of splicing in the presence of Noggin RNA with or without Strap antisense MO in Xenopus early tailbud stage. Quantifications of PSI values are shown below the gel. e Fraction of mis-splicing calculated from one of three independent analyses is shown in the bar graph. Error bars indicate the mean ± SD from n = 3 biological replicates. P-values were determined by unpaired two-tailed t-tests. f CkI splice-MO (50 ng) was injected into the dorsal animal regions of 4-cell to 8-cell stage Xenopus embryos. RT-PCR analysis of splicing with CkI splice-MO in Xenopus at the early tailbud stage. Quantifications of the PSI values are shown below the gels. The experiment was repeated three independent times with similar results. g Embryos were cultured until the tailbud or tadpole stages; representative embryos are shown here. The experiment was repeated three independent times with similar results. h Model describing the role of STRAP in mediating AS patterns. Upper, under conditions of normal levels of STRAP, it cooperates with the 17S U2 snRNP complex to regulate certain splicing sites, resulting in a balance of altered splicing patterns during neuroectoderm lineage commitment. Bottom, loss of STRAP has a negative effect on assembly of the 17S U2 complex. The fidelity of pre-mRNA splicing is also disrupted, causing high or low alternative transcripts or retaining ones. Blue box: constitutive exon; Yellow box: alternative exon; s: STRAP; 175 U2: 175 U2 snRNP complex; 155 U2, 155 U2 snRNP complex; 5/3′ ss: 5/3′ splicing site. The schematic diagram was created by the authors.

**Strand-RNA binding assay.** Nucleotide resolution UV crosslinking at 254 nm (400 mJ/cm²) and immunoprecipitation were performed. Two replicates from two independent WT EB or embryos were used for generating protein-RNA binding samples. Lysates generated from the crosslinked cells were treated with Turbo DNase (Ambion) and RNase I (1: 50 for high or 1:500 for low dilution, Ambion) for 5 min at 37°C to digest the genomic DNA and trim the RNA to short fragments of an optimal size range. RNA-protein complexes were immunoprecipitated with 100 µl of protein A Dynabeads (Life Technologies) and 10 µg of anti-STRAP antibody. Following stringent high salt washes (50 mM Tris-HCl; 1 M NaCl, 1 mM EDTA; 1% NP-40, 0.5% Tergitol 5, 0.5% sodium deoxycholate), the immunoprecipitated RNA was 5′-end labeled with radioactive [γ-32P]ATP62. The immunoprecipitated complexes were separated by SDS-PAGE and transferred to a nitrocellulose membrane followed by the exposure to a Fuji film at −80°C. eCLIP correlation analysis. To evaluate the biological reproducibility of eCLIP data, pairwise comparisons were performed to use 1.4–2.3 × 10⁵ uniquely mapped reads without PCR duplicates. All reads with read depths >3 were obtained using Samtools (version 1.3.1) and in-house Perl (version 5.26.3) scripts. The resulting log2 (Read depth) values were then plotted with the ggplot2 (version 3.2.1) package in R, followed with calculation of the Pearson correlation coefficient.

**Merged eCLIP-seq peaks.** As the resulting eCLIP reads were highly correlated, we merged all eCLIP clusters to ensure both specificity and sensitivity across the reproducible read data63. The merged peaks were generated based on the following strategy: (1) all eCLIP clusters common to, at least two of four libraries, would be chosen; and (2) if the two peaks from various groups have intersected regions, both of them would be retained and merged. After merged peaks were generated, the self-comparison was performed, and the redundancy was eliminated. In-house Perl (version 5.26.3) scripts were used for comparison of peak regions and generation of the final list of merged peaks.

**Motif analysis.** Motif analysis on peak regions was performed using HOMER software (http://homer.ucsd.edu/homer/). eCLIP profiles were referred to by the peak identifier. For each position of a peak, the number of CLIP sequences covering each base for a distance of 100 bases in the exon and 300 bases in the intron from all exon boundaries. For each position of a boundary, the relative density was defined as the total number of reads in the sample/the total number of reads in the control. For composite maps for STRAP-binding peaks on induced exon-inclusion or skipping events, a window between two constitutive exons was taken and divided into 100-bp equally sized bins. Only normalized peak size by allowing a given position was considered as an occupation.

**RNA map analysis.** Based on genomic coordinates of each peak, the gene and transcript information was obtained through Mutalyzer (version 2.0.29 http://mutalyzer.nl) and the corresponding read depth information was extracted from the BAM files using Samtools (version 1.3.1). If one peak region covered multiple genes/transcripts, generally, the longest protein-coding gene/transcript would be chosen for following analysis. In exon-intron/exon-exon boundary analysis, two kinds of boundaries were included in the analysis: (1) the boundaries that were covered by at least one peak; and (2) the exon-intron or intron-exon boundaries that were closest to one of two terminals of a peak, but not covered by this peak. For example, if peak-1 covered a part of exon-3, exon-3, exon-4, and a part of exon-4, then the boundaries between (1) intron-2 and exon-3, (2) exon-3 and intron-3, (3) intron-3 and exon-4, (4) exon-4 and intron-4, (5) intron-4 and exon-5 would be included in the analysis. This resulted in 11456-19058 unique exon-intron boundaries and 11444-16993 intron-exon boundaries. In-house Perl (version 5.26.3) scripts were created for selection of genes/transcripts and to count the number of CLIP sequences covering each base for a distance of 100 bases in the exon (GI 300 bases in the intron) from all exon boundaries. For each position of a boundary, the relative density was defined as the total number of reads in the sample/the total number of reads in the control. For composite maps for STRAP-binding peaks on induced exon-inclusion or skipping events, a window between two constitutive exons was taken and divided into 100-bp equally sized bins. Only normalized peak size by allowing a given position was considered as an occupation.

**Gene ontology analysis.** Gene Ontology (GO) enrichment analysis was performed with in-house Perl scripts based on gene-GO association file (ftp://ftp.geneontology.org/go/go-associations/gene_association.mgi.gz), and P-values were determined by Chi-squared test followed by Bonferroni correction.

**Xenopus.** Xenopus laevis frogs (1–2 years old male and female, Nasco) were used following the institutional IACUC protocol 09658 at the University of Alabama at Birmingham. We have compiled with all relevant ethical regulations for animal testing and research. Female frogs were primed with 800 units/frog of human chorionic gonadotropin hormone (Sigma-Aldrich) the night before use65. Embryos were...
obtained by in vitro fertilization and dejellied with 2% cysteine solution before injection of indicated molecules. The translational-blocking antisenese morpholino oligonucleotide (MO) was designed using Gene Tools service as follows: 5′-CCAC TAGGGAGGCTCTAGCATC-3′ for Strap targeting the start translation site (ATG-MO); 5′-CTTGCGTTCGGTGTCACTCTGTTT-3′ for Ckt1 ATG-MO; and 5′-TTGATGACCTTCTTGAAAACAGGA-3′ for Ckt1 spliced isoform (splice-MO). We performed PCR-based cloning to add the Myc-tag at the N-terminus of TAGCGAGGGCTTCATGTCAAT-3

from Gene Expression Omnibus database under the accession codeGSE116153 and them, gene expression data for neuroectoderm-like EBs and mouse E14.5 brain were from European Nucleotide Archive under the accession codePRJEB4513. All the other

targeting the translation start site

and isoform regulation in mammalian tissues. Nature 419, 188–192 (2002).

Harb. Perspect. Biol. 6, 137–148 (2015).

Nature 373, 126–128 (1988).

by in vitro fertilization and dejellied with 2% cysteine solution before

yielding the change in the mouse.

320

and isoform regulation in mammalian tissues. Science 338, 539–543 (2012).

321

322

323

and isoform regulation in mammalian tissues. Science 338, 539–543 (2012).

324

Nature 373, 126–128 (1988).

References

1. Mitiku, N. & Baker, J. C. Genomic analysis of gastrulation and organogenesis in the mouse. Dev. Cell 13, 897–907 (2007).

2. Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature 500, 593–597 (2013).

3. Biase, F. H., Cao, X., Zhong, S. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. Genome Res. 24, 1787–1796 (2014).

4. Mohammed, H. et al. Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. Cell Rep. 20, 1215–1228 (2017).

5. Wu, L. et al. The landscape of accessible chromatin in mammalian preimplantation embryo. Nature 534, 652–657 (2016).

6. Inoue, A. et al. Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature 547, 419–424 (2017).

7. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. Science 338, 1593–1599 (2012).

8. Will, C. L. & Lührmann, R. Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3, a003707 (2011).

9. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701–718 (2009).

10. Wahl, M. C. & Lührmann, R. SnapShot: spliceosome dynamics I. Cell 161, 1471–e1 (2015).

11. Vazquez, A. P. & O’Reilly, D. Variant snRNPs: new players within the spliceosome system. RNA Biol. 15, 17–25 (2018).

12. Hunt, S. L., Hsuan, J. J., Totty, N. & Jackson, R. J. Unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. Genes Dev. 13, 437–448 (1999).

13. Halder, S. K. et al. Oncogenic function of a novel WD-domain protein, STRAP, in human carcinogenesis. Cancer Res. 66, 6156–6166 (2006).

14. Jin, L., Yu, T., Yuan, G. & Datta, P. K. STRAP promotes stemness of human colorectal cancer via epigenetic regulation of the NOTCH pathway. Cancer Res. 77, 5464–5478 (2017).

15. Grümmer, M. et al. Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization. Hum. Mol. Genet. 14, 3099–3111 (2005).

16. Cauchi, R. J. & Cauchi, R. J. SMN and Gemins: ‘We are family’ … or are we? Insights into the partnership between Gemins and the spinal muscular atrophy disease protein SMN. Bioessays 32, 1077–1089 (2010).

17. Shen, S. et al. mATMs: Robust and flexible detection of differential alternative splicing from replicate RNA-seq data. Proc. Natl Acad. Sci. USA 111, E5593–E5601 (2014).

18. Lin, S. Serine threonine kinase receptor-associated protein deficiency impairs mouse embryonic stem cell lines cell commitment through CYP26A1-mediated retinoic acid homeostasis. Stem Cell 13, 1385–1396 (2015).

19. Chen, W. V., Delrow, J., Corrin, P. D., Frazier, J. P. & Soriano, P. Identification and validation of PDGF transcriptional targets by microarray-coupled gene-trap mutagenesis. Nat. Genet. 36, 304–312 (2004).

20. Camus, A. The morphogenetic role of midline mesendoderm and ectoderm in the development of the forebrain and the midbrain of the mouse embryo. Development 127, 1799–1818 (2000).

21. Storm, E. M., Rubenstein, J. R. & Martin, G. R. Dosage of Fgf8 determines whether cell survival is positively or negatively regulated in the developing forebrain. Proc. Natl Acad. Sci. USA 100, 1757–1762 (2003).

22. Veenholt, J. V. et al. Specification of dopaminergic subsets involves interplay of En1 and Ptx3. Development 140, 3373–3384 (2013).

23. Gavialis, A., Ruhbarg, C., Livet, J., Henderson, C. E. & Krumlauf, R. Neuronal defects in the hindbrain of Hoxa1, Hoxb1 and Hoxb2 mutants reflect regulatory interactions among these Hox genes. Development 130, 5663–5679 (2003).

24. Lagutin, O. V. et al. Sis3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. Genes Dev. 17, 367–379 (2003).

25. Tanaka, M., Chen, Z., Bartunkova, S., Yamasaki, N. & Itohu, S. The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. Development 126, 1269–1280 (1999).

26. Morrissey, E. E. et al. GATA6 regulates HNF4 and is required for embryo development. Genes Dev. 12, 3579–3590 (1998).

27. Burgess, R., Rawls, A., Brown, D., Bradley, A. & Olson, E. N. Requirement of the paraxis gene for somite formation and musculoskeletal patterning. Nature 384, 570–573 (1996).

28. Makarov, E. M., Owen, N., Bottrill, A. & Makarova, O. V. Functional mammalian spicoposomal complex E contains SMN complex proteins in addition to U1 and U2 snRNPs. Nucleic Acids Res. 40, 2639–2652 (2012).

29. Vinayagam, A. et al. Protein complex-based analysis framework for high-throughput data sets. Sci. Signal. 6, 1–11 (2013).

30. Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat. Biotechnol. 21, 183–186 (2003).

31. Meinhardt, A. et al. 3D reconstitution of the patterned neural tube from embryonic stem cells. Stem Cell Rep. 3, 987–999 (2014).

32. Choi, H. W. et al. In vivo differentiation of induced pluripotent stem cells into neural stem cells by chimera formation. PLoS ONE 12, e0170735 (2017).

33. Li, Q. et al. p53 integrates temporal WDR5 inputs during neuroectoderm and mesoderm differentiation of mouse embryonic stem cells. Cell Rep. 30, 465–480 (2020).

34. Ayoub, A. E. et al. Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-dimensional mRNA sequencing. Proc. Natl Acad. Sci. USA 108, 14950–14955 (2011).

35. Weber, M. et al. The tissue-specific transcriptomic landscape of the mid-gestational mouse embryo. Development 141, 2325–2330 (2014).

36. Han, H. et al. Multilayered control of alternative splicing regulatory networks by transcription factors. Mol. Cell 65, 539–553 (2017).

37. Vaquero-Garcia, J. et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. eLife 5, 11752–11782 (2016).

38. Nestruev, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods 13, 508–514 (2016).
39. Yeo, G. W. et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat. Struct. Mol. Biol. 16, 130–137 (2009).
40. Weyn-Vanhentenryck, S. M. et al. Precise temporal regulation of alternative splicing during neural development. Nat. Commun. 9, 2188 (2018).
41. Joseph, R., Dou, D. & Tsang, W. Neuronatin mRNA: alternatively spliced forms of a novel brain-specific mammalian developmental gene. Brain Res. Mol. Brain Res. 295, 1–8 (2018).
42. Lin, H. H. et al. Neuronatin promotes neural lineage in ESCs via Ca signaling. Stem Cells 28, 2810–2818 (2010).
43. Will, C. L. et al. Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. EMBO J. 18, 4978–4988 (2002).
44. Sunset, C. The SF3b complex: splicing and beyond. Cell Mol. Life Sci. https://doi.org/10.1007/s00018-020-03493-z (2020).
45. Wang, Q. et al. U2-related proteins CHERP and SR140 contribute to colorectal tumorigenesis via alternative splicing regulation. Int. J. Caner 145, 2728–2739 (2019).
46. Fox-Walsh, K. L. et al. The architecture of pre-mRNAs affects mechanisms of splice site pairing. Proc. Natl Acad. Sci. USA 102, 16176–16181 (2005).
47. Gelfman, S. et al. Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. Genome Res. 22, 35–50 (2012).
48. Hollander, D., Naftelberg, S., Lev-Maor, G., Kornblihtt, A. R. & Ast, G. How are short exons flanked by long introns defined and committed to splicing. Trends Genet. 32, 596–606 (2016).
49. Emanuele, M. J. & Stukenberg, P. T. Xenopus Cep57 is a novel kinetochore component involved in microtubule attachment. Cell 130, 893–905 (2007).
50. Ninomiya, K., Kataoka, N. & Hagiwara, M. Stress-responsive maturation of RNA splicing. J. Cell Sci. 2005, 1–9 (2011).
51. Yi, Y. et al. Splicing activation by Rbfox requires self-aggregation through its RNA binding domains. Nat. Struct. Mol. Biol. 25, 1140–1146 (2018).
52. Fu, X. D. & Ares, M. Jr. Context-dependent control of alternative splicing by Rbfox proteins. Nat. Rev. Genet. 15, 689–701 (2014).
53. Michael, T. L. et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nat. Struct. Mol. Biol. 20, 1434–1442 (2013).
54. Datta, P. K. & Moses, H. L. STRAP and Smad7 synergize in the inhibition of transforming growth factor β signaling. Mol. Biol. Cell 11, 3157–3167 (2000).
55. Ahn, E. Y. et al. SON controls cell cycle progression by coordinated regulation of RNA splicing. Mol. Cell 42, 185–198 (2011).
56. Konig, J. et al. CLIP transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. J. Vis. Exp. 50, E6268 (2011).
57. Chakrabarti, A. M. et al. Data science issues in studying protein-RNA interactions with CLIP technologies. Annu. Rev. Biomed. Data Sci. 1, 235–261 (2018).
58. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cells identities. Mol. Cell 38, 576–589 (2010).
59. Popov, I. K. et al. The RhoGEF protein Plekhg5 regulates apical constriction of bottle cells during gastrulation. Development 145, 1–11 (2018).