

# Insertion of an Alu element in a lncRNA leads to primate-specific modulation of alternative splicing

Shanshan Hu<sup>1,2</sup>, Xiaolin Wang<sup>1,2</sup> & Ge Shan<sup>1</sup>

Noncoding RNAs, mobile elements, and alternative splicing are all critical for the regulation of gene expression. Here we show that a conserved noncoding RNA acquires a new function due to the insertion of a mobile element. We identified a noncoding RNA, termed 5S-OT, which is transcribed from 5S rDNA loci in eukaryotes including fission yeast and mammals. 5S-OT plays a *cis* role in regulating the transcription of 5S rRNA in mice and humans. In the anthropoidea suborder of primates, an antisense Alu element has been inserted at the 5S-OT locus. We found that in human cells, 5S-OT regulates alternative splicing of multiple genes *in trans* via Alu/anti-Alu pairing with target genes and by interacting with the splicing factor U2AF65. This *trans* effect of 5S-OT in splicing might be exploited in biotechnological applications.

Noncoding RNAs (ncRNAs) can play conserved or lineage-specific roles<sup>1–3</sup>. One of the oldest ncRNAs in cells is 5S rRNA, which is found in all domains of life<sup>4,5</sup>. In essentially all metazoans, 5S rRNA genes with intergenic sequences cluster into tandem repeats<sup>5,6</sup>. Although the organization and sequences of 5S rRNA genes are highly conserved, the intergenic sequences in the 5S rDNA cluster are not<sup>4–6</sup>. The transcription of 5S rRNA by RNA polymerase (pol) III from the internal promoter of the 5S rDNA gene is well studied in eukaryotes<sup>7,8</sup>. For many pol III-transcribed genes, including 5S rRNA, the chromatin immunoprecipitation (ChIP) signals of RNA pol II have been observed to be close to one another, although whether pol II transcribes at these loci is elusive<sup>9,10</sup>.

We set out to investigate the potential existence of pol II transcripts from the 5S rDNA loci, speculating that these RNAs might be involved in local coupling between pol II and pol III. We identified a previously undescribed long ncRNA (lncRNA) that has a *cis* role in regulating pol III transcription of 5S rRNA in mammalian cells. Furthermore, we found that in human cells this lncRNA gains functions in modulating alternative splicing *in trans* via RNA-RNA pairing and by interacting with a splicing factor.

## RESULTS

### Identification of mammalian 5S rRNA-overlapping transcripts

In a screen for possible new transcripts arising from 5S rDNA loci, we identified an RNA with sequence overlap (either sense or antisense) with 5S rRNA and with a high probability of possessing a poly(A) tail in both mice and humans (Fig. 1a). We characterized the transcript further in mice and humans with an array of assays. First, strand-specific primer-mediated reverse-transcription PCR (RT-PCR) showed that this RNA was a sense transcript that overlapped with 5S rRNA in both mice and humans (Supplementary Fig. 1a). 5' and 3' rapid

amplification of cDNA ends (RACE), full-length RT-PCR and northern blots demonstrated that this transcript was 847 nt and 354 nt in mice and humans, respectively (Fig. 1b,c). We termed this molecule 5S rRNA overlapped transcript (5S-OT).

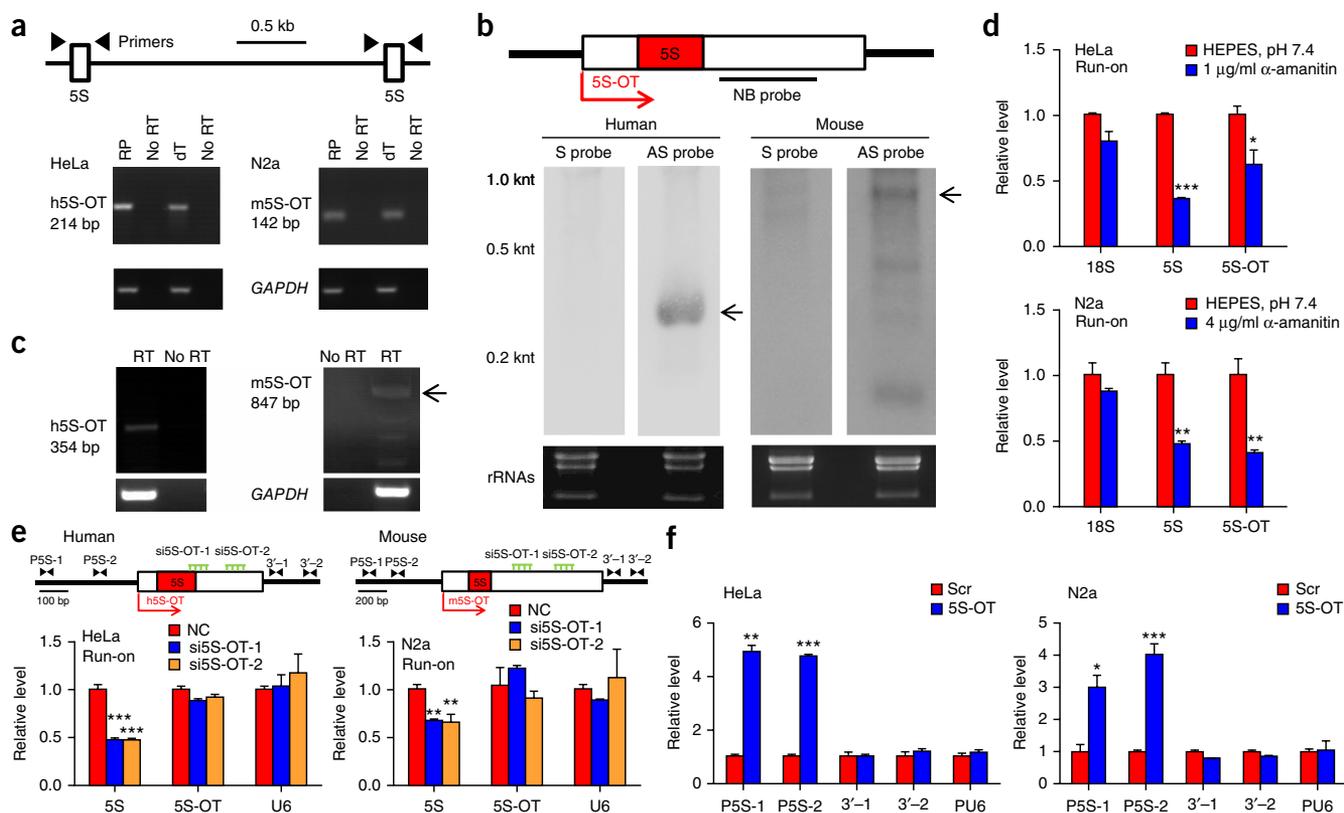
5S-OT was expressed in essentially all cell types and tissues examined in both mice and humans (Supplementary Fig. 1b,c). 5S-OT might be a pol II-synthesized transcript with a 3' poly(A), as inferred from the results of RT-PCR and 3' RACE with oligo(dT) as the RT primer (Fig. 1a,c). Specific inhibition of pol II with  $\alpha$ -amanitin<sup>11</sup> resulted in a decreased level of human (h5S-OT) and mouse (m5S-OT) transcripts (Supplementary Fig. 1d). ChIP with an anti-pol II antibody showed that pol II bound to the promoter and formed a peak around the first nucleosome (~200 bp downstream) of the transcription start site of 5S-OT in both human and mouse cells (Supplementary Fig. 1e). Coding possibility analysis<sup>12</sup> suggested that 5S-OT in mice and humans is noncoding (Supplementary Fig. 1f).

### *Cis* role of 5S-OT in regulating 5S rRNA transcription

Unexpectedly, specific inhibition of pol II with  $\alpha$ -amanitin resulted in decreased production of nascent 5S-OT as well as nascent 5S rRNA in both human (HeLa and 293T) and mouse (N2a and 3T3) cells, as detected by nuclear run-on assays (Fig. 1d and Supplementary Fig. 2a). These results suggest that 5S-OT has a *cis* regulatory role in the transcription of 5S rRNA, and we sought to further dissect this *cis* role by manipulating the expression level of 5S-OT. The nature of sequence repeats and the essentiality of 5S rRNA prevent the application of recombinant technologies such as CRISPR. Hence, we used RNA interference (RNAi) and antisense oligonucleotides (ASOs). Knockdown of h5S-OT or m5S-OT with two independent short interfering RNAs (siRNAs) resulted in decreased production of nascent 5S rRNA in both human and mouse cells (Fig. 1e and

<sup>1</sup>CAS Key Laboratory of Innate Immunity and Chronic Disease, CAS Center for Excellence in Molecular Cell Science, School of Life Sciences, University of Science and Technology of China, Hefei, China. <sup>2</sup>These authors contributed equally to this work. Correspondence should be addressed to G.S. (shange@ustc.edu.cn).

Received 16 June; accepted 7 September; published online 3 October 2016; doi:10.1038/nsmb.3302



**Figure 1** Identification, characterization, and *cis* effects of mammalian 5S-OT. (a) Identification of 5S-OT in humans and mice by RT-PCR. RT, reverse transcription; RP and dT, random primer and oligo(dT), respectively, used for RT; *GAPDH* mRNA, positive control. Primers used to amplify the transcripts across 5S rRNA sequences from the 5S rDNA repeats are indicated by black arrowheads. (b) Northern blots of human and mouse 5S-OT, using sense and antisense probes. The position of the northern blot (NB) probe is indicated above the blot images. rRNA bands are shown as an indication of equal loading. RNA samples were from HeLa (human) and N2a (mouse) cells. Uncropped images of northern blots are shown in **Supplementary Data Set 1**. (c) RT-PCR of the full-length h5S-OT and m5S-OT (indicated by an arrow). (d) Nuclear run-on assays showing the decrease in 5S and 5S-OT transcription after 24-h  $\alpha$ -amanitin treatment at a concentration of 1  $\mu$ g/ml and 4  $\mu$ g/ml in HeLa and N2a, respectively. 18S rRNA (a pol I transcript) is a negative control. (e) Nuclear run-on assays showing a decrease in 5S transcription after 5S-OT knockdown with two individual siRNAs in HeLa and N2a. The corresponding target site of the siRNA is shown (details in Online Methods). NC, siRNAs with scrambled sequences. U6 snRNA, pol III transcript used as a negative control for 5S rRNA. (f) Pull-down of 5S rDNA loci in ChIP with antisense probes against 5S-OT in HeLa or N2a. Scr, control oligonucleotide with scrambled sequence. Positions of 5S rDNA sites (P5S-1, P5S-2, 3'-1, and 3'-2) examined are indicated in e. The U6 promoter was used as a negative control. In **d-f**, error bars, s.e.m. from three independent experiments. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  by two-tailed Student's *t* test. Source data for **d-f** are available online.

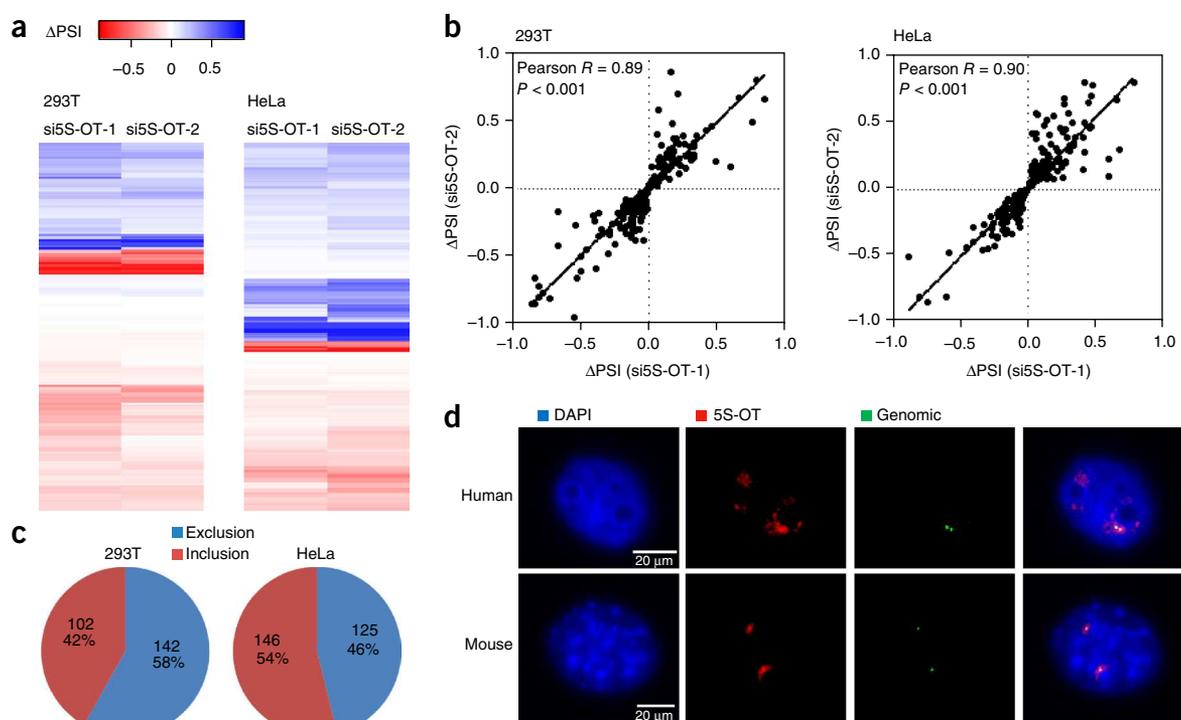
**Supplementary Fig. 2b,c**). We observed a similar phenomenon in human cells after knockdown of 5S-OT with ASOs (**Supplementary Fig. 2d**). Chromatin isolation by RNA purification (ChIRP) assays<sup>13</sup> showed that 5S-OT RNA bound to the 5S-OT promoter in both mouse and human cells (**Fig. 1f** and **Supplementary Fig. 2e**). These results together suggest that there is a pol II–pol III coupling mechanism in the transcription of 5S-OT and 5S rRNA. 5S-OT RNA associates with the chromatin of the 5S rDNA cluster and has a *cis* role in promoting the transcription of 5S rRNA in mammalian cells.

### Human 5S-OT modulates alternative splicing

To evaluate whether 5S-OT exerts functions other than its *cis* effect, we knocked down its expression with siRNAs and performed RNA sequencing. After the knockdown of h5S-OT with two independent siRNAs, we observed that the splicing of more than 200 exons was significantly altered in two human cell lines, HEK293T and HeLa (**Fig. 2a–c** and **Supplementary Fig. 3a**). We quantified cassette exons with significant changes ( $P < 0.01$ ) in splicing after h5S-OT knockdown and found 102 and 146 exons with increased inclusion and 142 and 125 exons with increased exclusion in

293T and HeLa cells, respectively (**Fig. 2c**). We called these exons with significant changes h5S-OT-sensitive exons. Thus, there were 244 h5S-OT-sensitive exons out of the 6,144 alternatively spliced exons detected in 293T cells and 271 h5S-OT-sensitive exons out of the 6,071 alternatively spliced exons detected in HeLa cells. A very small portion of h5S-OT-sensitive exons overlapped between 293T and HeLa cells (**Supplementary Fig. 3b**). We randomly selected and confirmed changes in alternative splicing for several cassette exons after knockdown of h5S-OT with siRNA or ASOs (**Supplementary Fig. 3c**). In the two examples examined with western blotting, the changes in the isoform ratio at the protein level were consistent with the alterations at the mRNA level, although the changes in protein levels were very subtle (**Supplementary Fig. 3c**). The full range of effects in protein levels resulting from the modulation of alternative splicing by h5S-OT remain to be further investigated.

Compared with changes in alternative splicing, changes in gene expression were less pronounced, and 226 (out of 16,978 expressed genes) and 34 (out of 17,372 expressed genes) genes in 293T and HeLa cells, respectively, demonstrated significantly altered expression levels



**Figure 2** Human 5S-OT has a *trans* role in modulating alternative splicing. **(a)** Heat maps of the changes in the percentage spliced in ( $\Delta$ PSI) compared with the negative control (bioinformatics details in Online Methods) after knockdown of h5S-OT in 293T and HeLa cells with two independent siRNAs. siRNAs with scrambled sequences were used as a negative control. Cassette exons with significant ( $P < 0.01$ ) changes in PSI are shown.  $P$  values were generated by two-tailed Mann–Whitney  $U$  test. **(b)** Correlation plot of  $\Delta$ PSI from two independent siRNAs in human 293T and HeLa cells. **(c)** Percentage inclusion or exclusion of h5S-OT-sensitive exons in HeLa and 293T cells. Numbers of exons significantly affected in the same direction of inclusion or exclusion for both siRNAs are shown. **(d)** Double FISH of human and mouse 5S-OT RNA (red) and 5S rDNA genomic loci (green) in human (HeLa,  $n = 42$ ) and mouse (N2a,  $n = 36$ ) cells. 5S-OT is localized in the nucleus (blue; DAPI nuclear stain); mouse 5S-OT is confined at the genomic loci of 5S rDNA, whereas human 5S-OT is not.

(two-fold change or greater,  $P < 0.01$ , with only changes consistent across two independent knockdowns counted, **Supplementary Fig. 4a,b**). Knockdown of m5S-OT with siRNA resulted in some changes in alternative splicing and gene expression (**Supplementary Fig. 4c,d**). 98 out of 6,970 cassette exons detected in mouse N2a cells showed significant changes ( $P < 0.01$ ) in alternative splicing, a much smaller effect compared with the changes observed after knockdown of h5S-OT in human cells (**Supplementary Fig. 4d**).

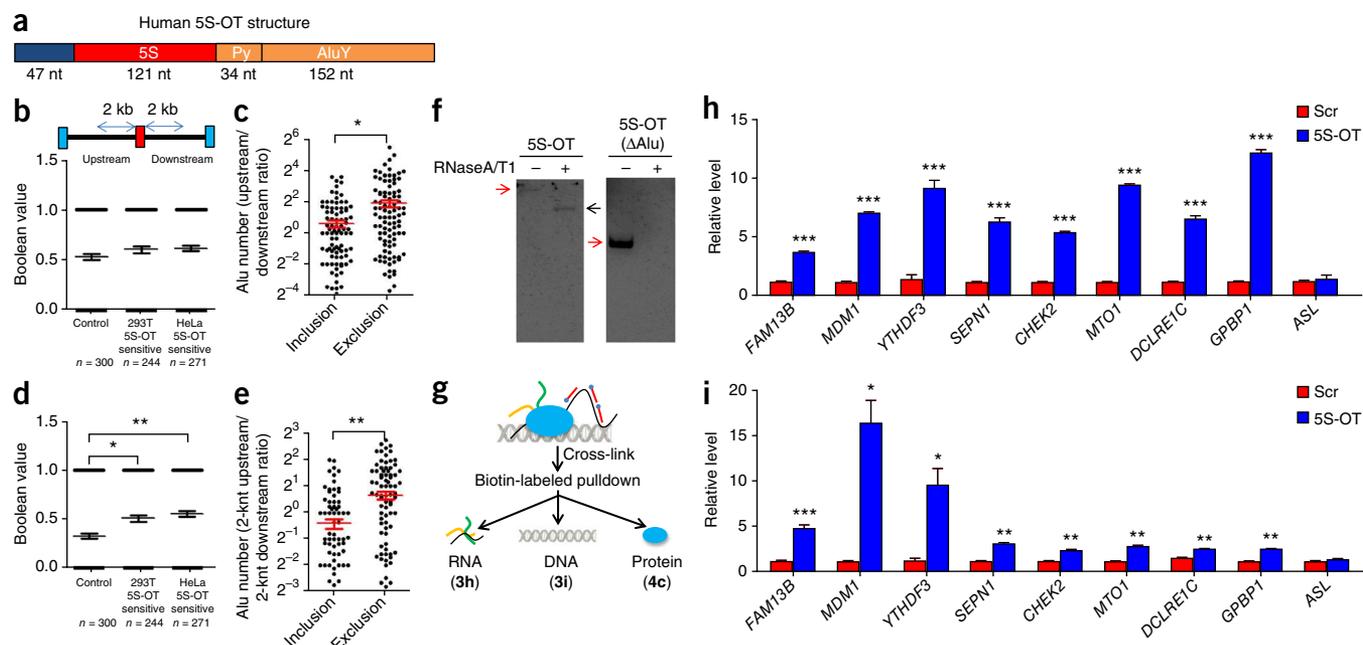
These results suggest that h5S-OT has a *trans* role in modulating the alternative splicing of multiple genes. If this is true, h5S-OT RNA should localize to regions other than the 5S rDNA loci. To confirm this, we used fluorescence *in situ* hybridization (FISH), which indicated that 5S-OT localized exclusively to the nuclei of human and mouse cells (**Fig. 2d**). This cellular localization further suggests that 5S-OT is noncoding. One salient feature from the FISH results was that h5S-OT was not confined to the genomic loci of 5S rDNA, whereas m5S-OT localization was more confined to the 5S rDNA gene cluster (**Fig. 2d**). We estimated that the copy numbers of 5S-OT in human HeLa and 293T cells were  $\sim 169$  and  $\sim 475$  copies per cell, respectively; in comparison, the copy numbers of m5S-OT in 3T3 and N2a cells were  $\sim 17$  and  $\sim 22$  copies per cell, respectively (**Supplementary Fig. 4e**).

### h5S-OT modulates alternative splicing via Alu pairing

How does 5S-OT modulate the alternative splicing of multiple exons in human cells? Inspection of h5S-OT sequences identified an Alu element antisense to the 3' region (**Fig. 3a**). Alu is a

primate-specific transposable element, and the Alu element in the h5S-OT belongs to the AluY subfamily<sup>14</sup>. AluY sequences are also present in the corresponding 5S-OT region of the chimpanzee, gorilla, orangutan, and green monkey transcripts but are absent in the owl monkey transcript; thus, the Alu sequence may have become part of the 5S-OT after the separation of Old World monkeys and New World monkeys<sup>14</sup>.

Could h5S-OT, with its antisense Alu sequences, mediate alternative splicing by targeting Alu elements in coding genes? Bioinformatic analyses showed that more than 90% of genes with h5S-OT-sensitive exons have sense Alu sequences in their precursor mRNAs (pre-mRNAs), although we observed essentially the same high percentage of sense Alu in all genes with cassette exons (**Supplementary Fig. 4f**). We then investigated introns immediately upstream or downstream of the h5S-OT-sensitive exons and found no significant differences in the presence of sense Alu in these exons compared with control cassette exons (**Fig. 3b**). Interestingly, exons with increased inclusion after h5S-OT knockdown tended to have more sense Alu in their downstream introns, whereas exons with increased exclusion in h5S-OT knockdown had more Alu in their upstream introns (**Fig. 3c**). Further analyses revealed that h5S-OT-sensitive exons had significantly higher chances of possessing sense Alu within 2 kilonucleotides (knt) of the 5' or 3' end (**Fig. 3d**). We examined sequences within 2 knt of either end and found that exons with increased inclusion after h5S-OT knockdown had more Alu in their downstream (3') sequences, whereas exons with increased exclusion after h5S-OT knockdown had more Alu in their upstream



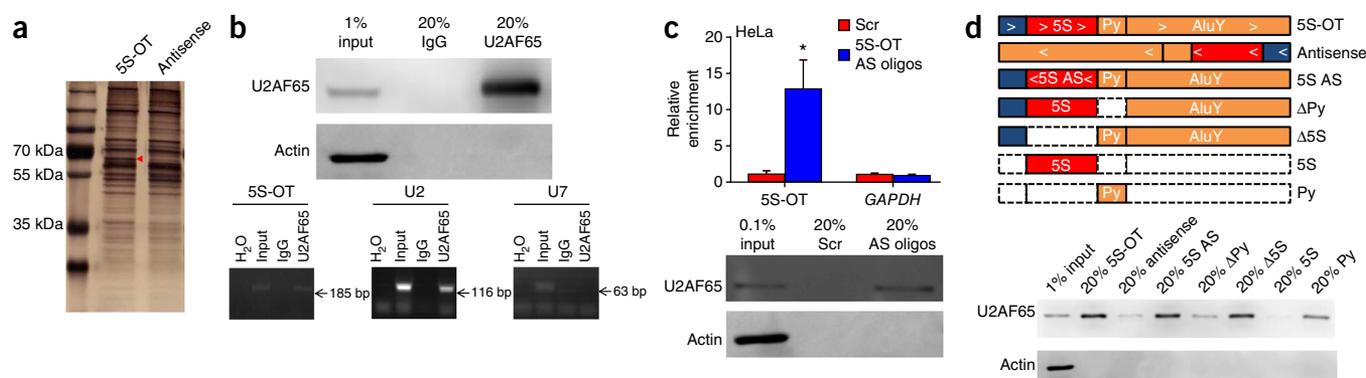
**Figure 3** h5S-OT has antisense AluY sequences and interacts with target pre-mRNAs. (a) Human 5S-OT structure. (b) Boolean distribution of sense Alu sequences within the upstream or downstream introns of h5S-OT-sensitive cassette exons. (c) Ratio of sense Alu numbers in the upstream/downstream introns of h5S-OT-sensitive cassette exons with inclusion or exclusion. (d) Boolean distribution of sense Alu sequences within 2 knt upstream or downstream of h5S-OT-sensitive cassette exons. In both **b** and **d**, controls were 300 randomly chosen cassette exons. 0, no sense Alu sequences present; 1, sense Alu sequences present. (e) Ratio of sense Alu numbers in the 2 knt upstream/downstream of h5S-OT-sensitive exons with inclusion or exclusion. (f) RNase-protection assays with h5S-OT probe (red arrow). The protected RNA band is indicated with a black arrow. Right, assays with a fragment of h5S-OT without the antisense Alu ( $\Delta$ Alu; red arrow), used for comparison. Uncropped images are shown in **Supplementary Data Set 1**. (g) Experimental process for RNA-RNA pulldown (shown in **h**), ChIRP (shown in **i**), and RNA-protein pulldown (shown in **Fig. 4c**) with biotinylated antisense oligonucleotides specific for h5S-OT. (h) RNA pulldown of h5S-OT showing coprecipitation of the corresponding pre-mRNAs of h5S-OT-sensitive genes in 293T cells. ASL, argininosuccinate lyase (an h5S-OT-insensitive gene with sense Alu sequences in its pre-mRNA) is a negative control. (i) Pull-down of the chromatin regions of multiple h5S-OT-sensitive genes in 293T cells by ChIRP. ASL, negative control. In **h** and **i**, Scr, biotinylated oligonucleotides with scrambled sequences. Data are from three independent experiments. In **b**, **d**, **h** and **i**, error bars, s.e.m.  $^*P < 0.05$ ;  $^{**}P < 0.01$ ;  $^{***}P < 0.001$  by two-tailed Student's *t* test. In **c** and **e**, error bars, s.e.m.  $^*P < 0.05$ ;  $^{***}P < 0.01$  by two-tailed Mann-Whitney *U* test. Source data for **h** and **i** are available online.

(5′) sequences (**Fig. 3e**). Together, these analyses demonstrated a positive correlation between sense Alu sequences in pre-mRNAs and sensitivity to h5S-OT, and a clear association between the distribution of sense Alu in the flanking sequences of exons and the h5S-OT-mediated effect of exon inclusion or exclusion. It is also possible that a small fraction of changes observed in alternative splicing after h5S-OT knockdown might be indirect, because the splicing patterns of several splicing factors were also altered after h5S-OT knockdown (**Supplementary Fig. 4g**).

Through RNase-protection assays (RPAs), we found that h5S-OT but not the control RNA (lacking the antisense Alu sequences) formed double-stranded RNA *in vitro* from isolated total RNA transcripts (**Fig. 3f**). RNA pulldown experiments with biotinylated antisense oligonucleotides specific to h5S-OT pulled down the corresponding transcript from eight h5S-OT-sensitive genes that we examined (**Fig. 3g,h**). Because the majority of introns are spliced cotranscriptionally on chromatin, h5S-OT may interact with the chromatin of the h5S-OT-sensitive genes containing h5S-OT-sensitive exons. To examine this possibility, we performed ChIRP assays in 293T cells and found that h5S-OT indeed interacted with the chromatin of the eight h5S-OT-sensitive genes examined (**Fig. 3i**). These results suggested that h5S-OT modulates alternative splicing by antisense-sense Alu pairing, although the specificity for which exons are targeted may not be based solely on Alu pairing and remains to be further investigated.

### h5S-OT interacts with U2AF65

The effects of h5S-OT in alternative splicing may also be mediated by proteins. To identify proteins interacting with h5S-OT, we performed pulldowns with biotin-labeled h5S-OT incubated with cell lysates and identified the associated proteins through mass spectrometry (**Fig. 4a** and **Supplementary Fig. 5a,b**). The splicing factor U2AF65 has been identified as an h5S-OT-binding protein<sup>15</sup>. We further confirmed the interactions through RNA immunoprecipitation (RIP) with antibody against U2AF65 and by RNA pulldown with antisense oligonucleotides against 5S-OT (**Fig. 4b,c**). There is a polypyrimidine tract (Py) in the sequence of h5S-OT (**Fig. 3a** and **Supplementary Fig. 5c**), and U2AF65 is known to bind to Py created by the insertion of antisense Alu<sup>16</sup>. Presumably, the Py site in h5S-OT has evolved from the poly(A) tail of the antisense Alu after its genomic insertion (**Supplementary Fig. 5c**). We verified that the Py in h5S-OT was an U2AF65-binding site by using a variety of constructs (**Fig. 4d**). Thus, it is not the antisense Alu sequence but the Py site that is created by the insertion of the Alu element, which is important for U2AF65 recruitment. For comparison, we examined another splicing factor, PTBP1 (ref. 17), which is also known to bind Py sites, and found no interaction between PTBP1 and h5S-OT (**Supplementary Fig. 5d**). m5S-OT did not interact with U2AF65, even when this RNA was overexpressed in mouse cells (**Supplementary Fig. 5e,f**), whereas h5S-OT interacted with mouse U2AF65 protein when h5S-OT was artificially expressed in mouse cells (**Supplementary Fig. 5g**).



**Figure 4** h5S-OT interacts with U2AF65. (a) Silver staining of proteins pulled down with biotinylated h5S-OT RNA and a negative control with antisense sequences. Red triangle denotes the band identified as U2AF65 by mass spectrometry. (b) Pull-down of h5S-OT in RIP with an antibody to U2AF65. Western blots showing efficient pull-down of U2AF65 with  $\beta$ -actin as a negative control. U2 snRNA, which is known to interact with U2AF65, is a positive control for h5S-OT in RT-PCR, and U7 snRNA is a negative control. (c) RNA pull-down of h5S-OT with 5S-OT antisense oligonucleotides (AS oligos), showing coprecipitation of U2AF65. Pull-down efficiency is shown in the bar graph, and proteins were examined through western blotting. *GAPDH* mRNA is a negative control for the RNA pull-down, and  $\beta$ -actin is a negative control for western blotting. Scr, oligonucleotide with scrambled sequences. Error bars, s.e.m. from three independent experiments. \* $P < 0.05$  by two-tailed Student's *t* test. (d) RNA pull-down assays showing that the Py site of h5S-OT RNA is the binding site of U2AF65.  $\beta$ -actin, negative control. Constructs for the *in vitro* transcription of the RNAs used in the RNA pull-down assays are shown above. For b–d, uncropped images of are shown in **Supplementary Data Set 1**. Source data for c are available online.

### U2AF65 and Alu pairing are required for h5S-OT's *trans* roles

To explore the potential involvement of U2AF65 in the *trans* roles of h5S-OT, we performed siRNA knockdown of U2AF65 and sequenced the RNA samples. The splicing of 835 and 750 exons was significantly altered ( $P < 0.01$ ) after knockdown of U2AF65 with two independent siRNAs in HeLa and 293T cells, respectively (Fig. 5a and Supplementary Fig. 6a,b). We termed these exons U2AF65-sensitive exons. U2AF65 has been shown to directly bind at least 88% of 3' splice sites (3' SSs), although a previous study has reported that global U2AF65 knockdown significantly altered the splicing of only 445 exons (102 increased exclusion and 343 increased inclusion out of the 6,915 alternatively spliced exons detected) in HeLa cells<sup>18</sup>. Approximately 87% of these 445 exons were also identified as U2AF65-sensitive exons in HeLa cells in our study (Supplementary Fig. 6c). Although U2AF65 is involved in the majority of alternative-splicing events; only a small portion of these events are sensitive to global decreases in U2AF65 levels<sup>18</sup>. Moreover, U2AF65 has a complex polar effect in alternative splicing<sup>18</sup> (Fig. 5a).

Out of the 271 h5S-OT-sensitive exons in HeLa cells, 158 (~58.3%) were also U2AF65-sensitive exons; for 293T cells, the overlap was 100 out of 244 exons (~41.0%) (Fig. 5b). Knockdown of either h5S-OT or U2AF65 did not change the expression level of the other interaction partner (Supplementary Fig. 6d,e), thus suggesting that these partners do not stabilize or destabilize each other. As a comparison, exons showing altered splicing after PTBP1 knockdown had little overlap with h5S-OT-sensitive exons (Supplementary Fig. 6f).

Bioinformatic analyses revealed that exons sensitive to both h5S-OT and U2AF65 were positively correlated in their changes in exclusion or inclusion (Fig. 5c). We randomly selected and confirmed some of the changes in alternative splicing by using RT-PCR (Supplementary Fig. 7a). Overexpression of U2AF65 could not compensate for the alternative-splicing effects of h5S-OT knockdown in the examples that we examined (Supplementary Fig. 7b), thus indicating that h5S-OT primarily affects the distribution rather than the global level of U2AF65 protein.

Previous research has shown that U2AF65 facilitates the exonization of Alu elements in human cells<sup>19</sup>. In agreement with the hypothesis that h5S-OT recruits U2AF65, exons generated from Alu

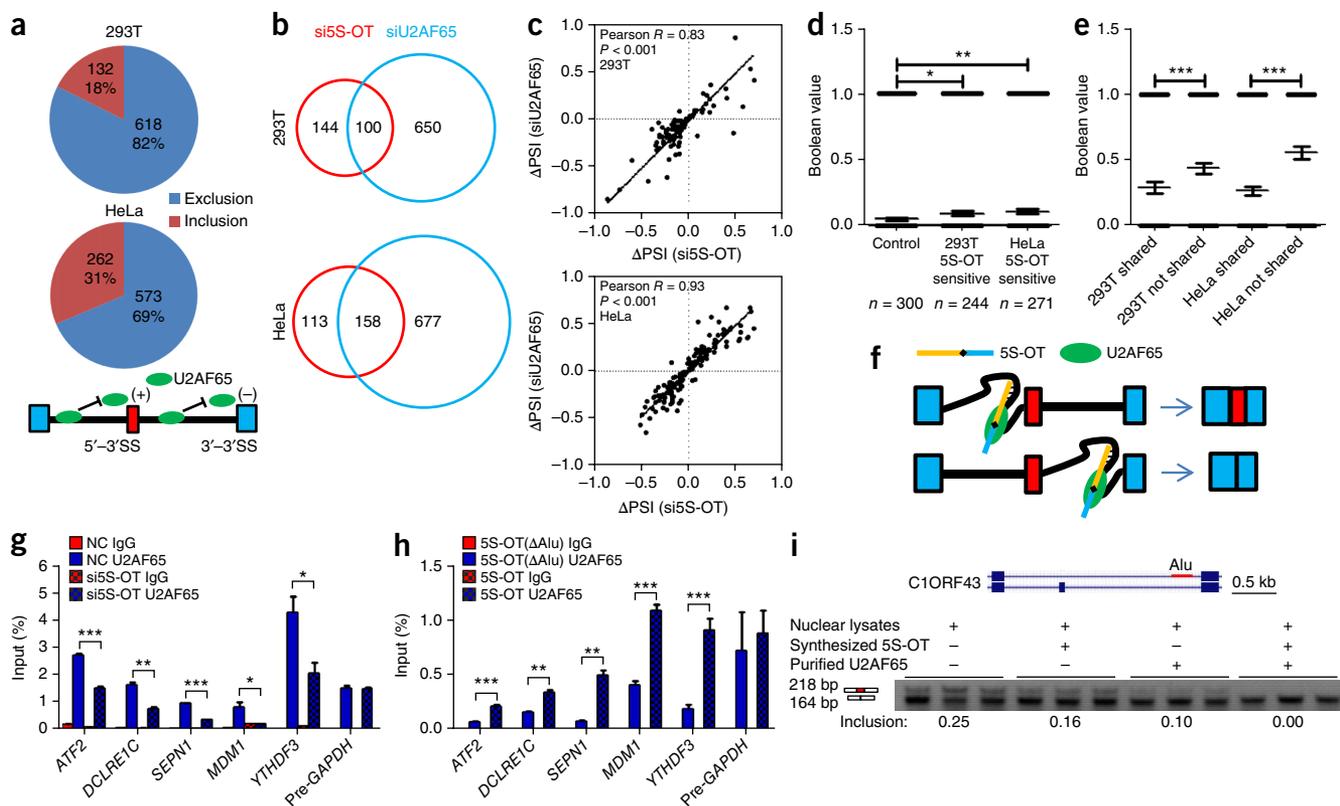
exonization were more sensitive to h5S-OT knockdown (Fig. 5d). Genes sensitive to h5S-OT but not to U2AF65 were more likely to have multiple (three or more) sense Alu sequences in their pre-mRNA than genes sensitive to both h5S-OT and U2AF65 in either 293T or HeLa cells (Fig. 5e). From these data along with results shown in Figure 3, we present a model in which h5S-OT recruits U2AF65 and consequently modulates alternative splicing, and inclusion or exclusion in the affected cassette exon is determined by the polar effect of U2AF65 (Fig. 5f). After h5S-OT knockdown, U2AF65 binding to pre-mRNAs of h5S-OT-sensitive genes was decreased (Fig. 5g and Supplementary Fig. 7c). However, overexpression of h5S-OT led to increased binding of U2AF65 to pre-mRNAs of h5S-OT-sensitive genes (Fig. 5h and Supplementary Fig. 7d). To further verify our hypothesis, we performed *in vitro* splicing assays with nuclear lysates, purified U2AF65 protein and synthesized h5S-OT RNA. Indeed, the examined cassette exon of *CIORF43* was more significantly affected when U2AF65 protein and h5S-OT RNA were added together than when either component was added alone (Fig. 5i).

### h5S-OT may have physiological functions

We further examined whether h5S-OT has regulatory functions in the differentiation of THP-1 cells, a more physiologically relevant context<sup>20</sup>. In the process of differentiation of THP-1 cells, a classic model of human macrophage differentiation<sup>20</sup>, h5S-OT expression was significantly increased (Fig. 6a). Knockdown of h5S-OT led to decreased differentiation efficiency of THP-1 cells (Fig. 6b) and to altered splicing of 174 and 173 genes ( $P < 0.01$ ) in undifferentiated and differentiating THP-1 cells, respectively (Fig. 6c).

### h5S-OT-based technology to manipulate alternative splicing

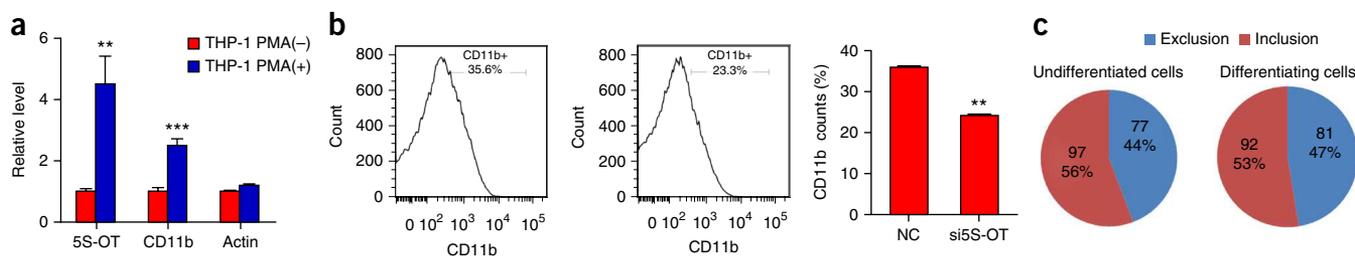
If our understanding of the *trans* role of h5S-OT in modulating alternative splicing is correct, then replacing the antisense Alu elements of h5S-OT with a gene-specific antisense sequence might convert a gene not regulated by the genuine h5S-OT into a gene regulated by the 'gene-specific' h5S-OT. Indeed, we were able to achieve interference in the splicing of targeted exons by overexpressing gene-specific h5S-OT, and the effect required the Py but not the 5S sequences in h5S-OT (Fig. 7a). This result indicates that the *trans* effect of h5S-OT



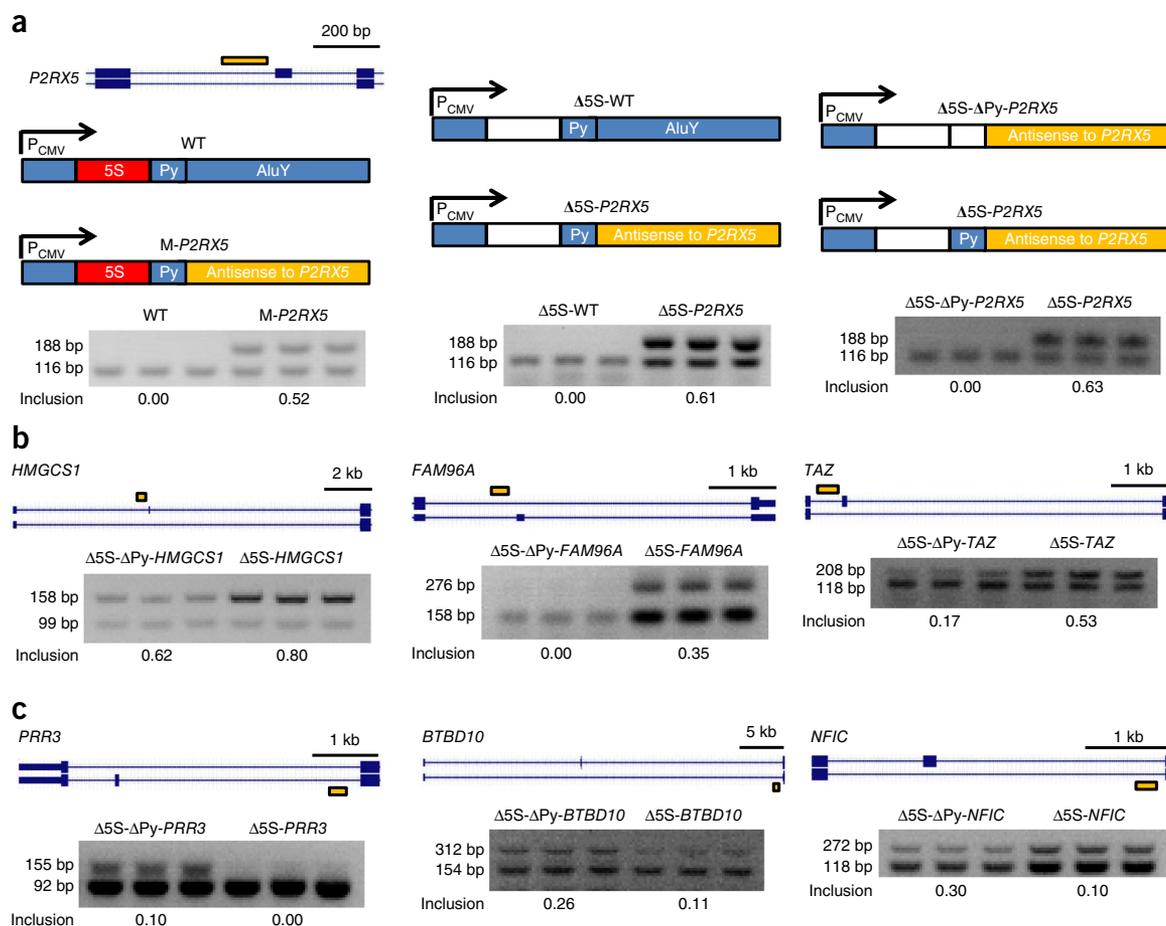
**Figure 5** h5S-OT modulates alternative splicing via RNA-RNA interaction and by recruiting U2AF65. **(a)** U2AF65-sensitive exons in HeLa and 293T cells. Numbers of exons significantly affected in the same direction of inclusion or exclusion for both siRNAs are shown. The known working model of U2AF65 polar effects is shown below; (+), promotes exon inclusion (red); (-), promotes exon exclusion (blue). **(b)** Numbers of h5S-OT- and U2AF65-sensitive exons in 293T and HeLa cells. **(c)** Correlation plot of  $\Delta$ PSI for exons sensitive to both h5S-OT and U2AF65 in 293T and HeLa cells. **(d)** Boolean distribution of exons generated from the Alu exonization of h5S-OT-sensitive exons. Control, 300 randomly chosen cassette exons. 0, no exonization of Alu; 1, exonization of Alu. **(e)** Boolean distribution of sense Alu sequences in the pre-mRNAs of h5S-OT-sensitive genes that were either sensitive to U2AF65 (shared) or insensitive to U2AF65 (not shared) in 293T and HeLa cells. 1, presence of multiple ( $\geq 3$ ) sense Alu sequences; 0, absence of multiple ( $< 3$ ) sense Alu sequences. **(f)** Model of h5S-OT in modulating alternative splicing by bringing U2AF65 to target the upstream or downstream 3' SSs of cassette exons through antisense-sense Alu pairing. **(g)** U2AF65 RIP of pre-mRNA of h5S-OT-sensitive genes with h5S-OT knockdown in 293T cells. NC, siRNAs with scrambled sequences. **(h)** U2AF65 RIP of pre-mRNA of h5S-OT-sensitive genes with h5S-OT overexpression in 293T cells. h5S-OT without the antisense Alu ( $\Delta$ Alu) was used as a comparison. In **g** and **h**, GAPDH pre-mRNA is a negative control, and data are from three independent experiments. **(i)** *In vitro* splicing assays of C10RF43 with or without the application of purified U2AF65 protein and h5S-OT RNA. The position of the sense Alu element in C10RF43 is indicated. In **d**, **e**, **g**, and **h**, error bars, s.e.m.; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  by two-tailed Student's *t* test. Source data for **g** and **h** are available online.

in splicing may be adopted as a biotechnology to manipulate alternative splicing of specific genes of interest. Targeting the upstream intron of a cassette exon with gene-specific 5S-OT increased inclusion (Fig. 7b), whereas targeting the downstream intron of a cassette

exon with gene-specific 5S-OT decreased inclusion (Fig. 7c). The effects of these gene-specific 5S-OTs matched the polar effect of U2AF65 (ref. 18). However, in some of the tested cases, the gene-specific 5S-OT did not work well (or might have required further



**Figure 6** h5S-OT regulates the differentiation of THP-1. **(a)** Real-time PCR showing increased expression of 5S-OT after 4-h phorbol 12-myristate 13-acetate (PMA) treatment in THP-1. CD11b, marker of THP-1 differentiation used as a positive control;  $\beta$ -actin, negative control. **(b)** Fluorescence-activated cell sorting (FACS) with CD11b staining. Bar graph to the right shows quantification. **(c)** h5S-OT-sensitive exons in undifferentiated and differentiating THP-1 cells. In **a** and **b**, error bars, s.e.m. from three independent experiments. \*\* $P < 0.01$ ; \*\*\* $P < 0.01$  by two-tailed Student's *t* test. Source data for **a** and **b** are available online.



**Figure 7** *Trans* effects of h5S-OT can be adapted as a biotechnology. (a) Modulation of alternative splicing of *P2RX5*, a gene not sensitive to h5S-OT, by tailored gene-specific h5S-OT. M-*P2RX5*, modulated *P2RX5*; WT, wild-type construct expressing h5S-OT as a negative control. The position of the gene-specific targeting site is indicated for the corresponding gene. The Py, but not the 5S sequences, is essential for the *trans* effect of gene-specific h5S-OT. (b) RT-PCR showing that targeting the upstream 3' SS with gene-specific h5S-OT increases inclusion of the corresponding cassette exon. (c) RT-PCR showing that targeting the downstream 3' SS with gene-specific h5S-OT increases exclusion of the corresponding cassette exon.

optimization) (Supplementary Fig. 8a). h5S-OT is one of various endogenous lncRNAs identified to play a *trans* role in modulating alternative splicing<sup>21,22</sup>, and *trans* delivery of a splicing factor to a pre-mRNA with artificial RNA to manipulate the splicing pattern has already been used as a biotechnology<sup>23</sup>.

### 5S-OT is an ancient lncRNA

We discovered RNA transcripts with sequences overlapping with the 5S rRNA, and with a high possibility of having a poly(A) tail, in *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Danio rerio*, but not in *Saccharomyces cerevisiae* (Fig. 8a). The novel transcript is sense with respect to 5S rRNA in *S. pombe*, *C. elegans*, *D. melanogaster*, and *D. rerio* (Fig. 8b). Thus, these RNAs along with h5S-OT and m5S-OT can collectively be referred to as 5S-OT. 5S-OT is a lncRNA that is relatively conserved in eukaryotes including fission yeast and humans, is transcribed from 5S rDNA loci and contains ultraconserved 5S rRNA sequences.

### DISCUSSION

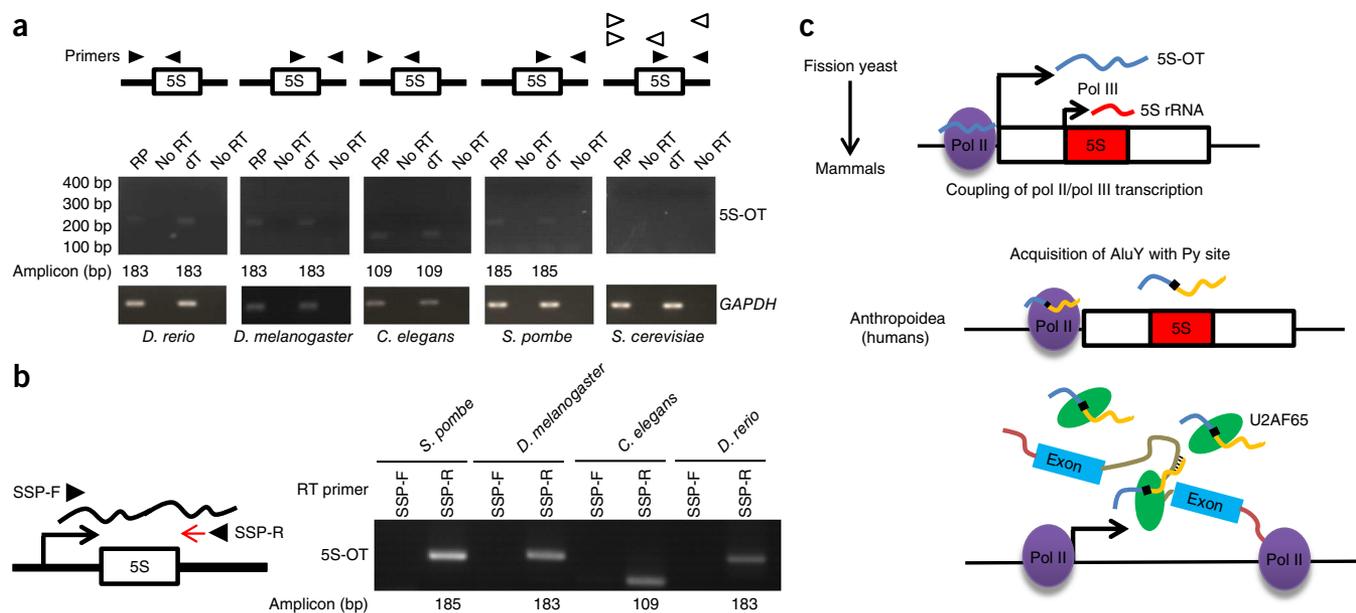
In mice and humans, 5S-OT plays a *cis* role in coupling transcription of 5S-OT by pol II with the transcription of 5S rRNA by pol III. An insertion of an antisense Alu element in the anthropoidea suborder of primates has added a Py to 5S-OT, thereby conferring the ability

to interact with U2AF65 and regulate alternative splicing of a subset of mRNAs via anti-Alu/Alu pairing (Fig. 8c).

It is still possible that 5S-OT may possess *trans* functions in other branches of eukaryotes, although these effects would be unlikely to depend on U2AF65 and RNA-RNA pairing. The 3' portion of m5S-OT is not a long repetitive sequence, and there is no Py site in m5S-OT or in the 5S-OTs of the other model organisms examined (Supplementary Fig. 8b). m5S-OT also showed no interaction with U2AF65 (Supplementary Fig. 5e,f).

Both h5S-OT and m5S-OT promoters may contain a CpG island (Supplementary Fig. 8c), thus indicating that the local chromatin structure and transcriptional activity may be subjected to CpG-island-dependent regulation<sup>24</sup>. m5S-OT did not have the high expression levels seen for h5S-OT (Fig. 2d and Supplementary Fig. 4e). Multiple Em for Motif Elicitation (MEME) analysis revealed that the human but not the mouse 5S-OT gene promoter has three binding motifs for ERG2 (Supplementary Fig. 8c), a transcriptional activator expressed in an array of human cells. Interestingly, the promoter of the ERG2 gene itself has also been involved in the evolution of mammalian lineages<sup>25</sup>.

We also cannot rule out the possibility that some of the *trans* effects of h5S-OT might be mediated by its *cis* effect on the transcription of 5S rRNA, although the total levels of 5S rRNA were not significantly



**Figure 8** Evolution and working model of 5S-OT. (a) Identification of 5S-OT in multiple model organisms by RT-PCR. RT, reverse transcription; RP and dT, random primer and oligo(dT) used for RT; GAPDH mRNA, positive control. Primers used to amplify the transcripts across 5S rRNA sequences are indicated above the gels. Three pairs of primers were used for *S. cerevisiae*, and no RT-PCR product was obtained (the gel image shows only results from the primer pair indicated with filled triangles). (b) Strand specific RT-PCR of 5S-OT in different species. SSP-F, RT primer in the sense direction relative to 5S rRNA. SSP-R, RT primer antisense to 5S rRNA. (c) Model for the evolution of 5S-OT in eukaryotes, its *cis* function in regulating the transcription of 5S rRNA in mammals, and its *trans* roles in modulating alternative splicing in humans.

changed despite a decreased transcriptional rate after 5S-OT knock-down in both human and mouse cells (Fig. 1e and Supplementary Fig. 2b–d). This phenomenon and observations from previous studies indicate that there may be a mechanism to maintain a homeostatic level of 5S rRNA in cells<sup>8,26</sup>. Our data together demonstrated that the *trans* roles of h5S-OT in modulating alternative splicing are primarily accomplished with RNA-RNA complementation mediated by the antisense Alu sequences at the 3' end, and by bringing U2AF65 via its Py site to the target pre-mRNAs.

Although our data showed that the target specificity of h5S-OT requires anti-Alu–Alu pairing, and the effect of h5S-OT in alternative splicing involves U2AF65, multiple factors may determine which exons are targeted for inclusion or exclusion. Because most splicing occurs cotranscriptionally, h5S-OT would presumably affect only alternative splicing of genes whose genomic loci colocalize with h5S-OT lncRNA. The complex effect of U2AF65 and interactions among splicing factors may also contribute to the effects of h5S-OT. We observed that both h5S-OT- and U2AF65-sensitive exons tended to have relatively stronger 3' SSs for their immediate downstream introns than the 3' SSs of their immediate upstream introns, independently of whether these exons showed changes in exclusion or inclusion after knockdown of h5S-OT or U2AF65 (Supplementary Fig. 8d); however, the reason for this phenomenon is unclear. The molecular mechanism of the target specificity of h5S-OT is more complex than the presence of sense Alu sequences and remains to be further investigated.

Inhibition of pol II with low doses of  $\alpha$ -amanitin also suppresses the expression of U6 small nuclear RNA (snRNA), a pol III transcript, although there is debate about whether this is an indication of a direct pol II–pol III coupling or an indirect association due to side effects of pol II inhibition<sup>27,28</sup>. The identification of 5S-OT RNA and its *cis* effect in humans and mice suggests a direct coupling in the transcriptional activity of pol II and pol III at the 5S rDNA loci via the

production of a lncRNA. The detailed mechanisms underlying how mammalian 5S-OT lncRNA associates with the 5S rDNA chromatin and further regulates the transcription of 5S rRNA require further investigation.

Alu elements are major players in shaping the genome and evolution of primates<sup>14,29</sup>. The observed *trans* effect of h5S-OT in modulating alternative splicing expands the functional mechanisms of lncRNAs and also describes a new role of Alu elements, which are already known to play multiple roles in processes including RNA editing, alternative splicing, and circular-RNA biogenesis<sup>29–31</sup>. Some other primate-specific mobile elements such as LINE-1 also take part in lineage-specific gene-expression regulation, pathology, and evolution<sup>32,33</sup>.

Recently, a mammalian-specific alternative-splicing event has been identified in the splicing factor PTBP1 and found to alter PTBP1's splicing-regulatory activities<sup>34</sup>. This observation may explain some of the differences in alternative splicing between birds and mammals<sup>34</sup>. This is an example in which the evolution of a protein-coding gene has created distinctions in alternative splicing between species. 5S-OT may be an example of a noncoding gene that has evolved to participate in the species-specific modulation of alternative splicing.

5S-OT is a relatively conserved lncRNA in eukaryotic cells. The conservation of lncRNAs can be related to sequences, synteny, functions, or structure<sup>35–37</sup>. Eukaryotic 5S-OTs share a common 5' portion of their 5S rRNA sequences, owing to their syntenic transcription from the 5S rDNA loci in eukaryotic organisms including fission yeast and humans, and they play a *cis* role in regulating the expression of 5S rRNA, at least in mice and humans. The molecular evolution of lncRNAs in specific animal lineages may have allowed certain lncRNAs to perform specific functions<sup>36,37</sup>. In the case of 5S-OT in Old World monkeys and apes including humans, insertion of a lineage-specific antisense AluY sequence enables this lncRNA to interact with the splicing factor U2AF65 and to further regulate alternative splicing of a subset of mRNAs *in trans*.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** The sequences of h5S-OT and m5S-OT have been deposited in GenBank under accession numbers [KX756089](#) and [KX756090](#). RNA-sequencing data have been deposited in the Gene Expression Omnibus database under accession number [GSE85709](#).

*Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).*

## ACKNOWLEDGMENTS

We thank X. Wang (HUST), J. Liu (HFUT), and Y. Zhou (Wuhan University) for reagents. We thank L. Chen and other members of the Shan laboratory for discussions and technical support. This work was supported by the National Basic Research Program of China (2015CB943000 to G.S.), the National Natural Science Foundation of China (91519333 and 31471225 to G.S.), the CAS Key Laboratory of Innate Immunity and Chronic Disease (G.S.), and the Fundamental Research Funds for the Central Universities (WK2070000034 to G.S.).

## AUTHOR CONTRIBUTIONS

G.S. conceived the project and supervised its execution. G.S., S.H., and X.W. designed experiments, analyzed data, and wrote the manuscript. S.H. and X.W. performed the experiments. X.W. performed bioinformatic analyses. All authors discussed the results and made comments on the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

- Quinn, J.J. & Chang, H.Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17**, 47–62 (2016).
- Cech, T.R. & Steitz, J.A. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
- Guttman, M. & Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
- Bogdanov, A.A., Dontsova, O.A., Dokudovskaya, S.S. & Lavrik, I.N. Structure and function of 5S rRNA in the ribosome. *Biochem. Cell Biol.* **73**, 869–876 (1995).
- Ciganda, M. & Williams, N. Eukaryotic 5S rRNA biogenesis. *Wiley Interdiscip. Rev. RNA* **2**, 523–533 (2011).
- Haeusler, R.A. & Engelke, D.R. Spatial organization of transcription by RNA polymerase III. *Nucleic Acids Res.* **34**, 4826–4836 (2006).
- Paule, M.R. & White, R.J. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.* **28**, 1283–1298 (2000).
- Hu, S., Wu, J., Chen, L. & Shan, G. Signals from noncoding RNAs: unconventional roles for conventional pol III transcripts. *Int. J. Biochem. Cell Biol.* **44**, 1847–1851 (2012).
- Oler, A.J. *et al.* Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nat. Struct. Mol. Biol.* **17**, 620–628 (2010).
- Barski, A. *et al.* Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. *Nat. Struct. Mol. Biol.* **17**, 629–634 (2010).
- Tata, J.R., Hamilton, M.J. & Shields, D. Effects of  $\alpha$ -amanitin *in vivo* on RNA polymerase and nuclear RNA synthesis. *Nat. New Biol.* **238**, 161–164 (1972).
- Wang, L. *et al.* CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74 (2013).
- Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
- Batzer, M.A. & Deininger, P.L. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3**, 370–379 (2002).
- Ruskin, B., Zamore, P.D. & Green, M.R. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**, 207–219 (1988).
- Deininger, P. Alu elements: know the SINES. *Genome Biol.* **12**, 236 (2011).
- Gooding, C., Kemp, P. & Smith, C.W. A novel polypyrimidine tract-binding protein paralog expressed in smooth muscle cells. *J. Biol. Chem.* **278**, 15201–15207 (2003).
- Shao, C. *et al.* Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nat. Struct. Mol. Biol.* **21**, 997–1005 (2014).
- Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453–466 (2013).
- Auwerx, J. The human leukemia cell line, THP-1: a multifaceted model for the study of monocyte-macrophage differentiation. *Experientia* **47**, 22–31 (1991).
- Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).
- Gonzalez, I. *et al.* A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat. Struct. Mol. Biol.* **22**, 370–376 (2015).
- Skordis, L.A., Dunckley, M.G., Yue, B., Eperon, I.C. & Muntoni, F. Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc. Natl. Acad. Sci. USA* **100**, 4114–4119 (2003).
- Deaton, A.M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
- Marcovitz, A., Jia, R. & Bejerano, G. “Reverse genomics” predicts function of human conserved non-coding elements. *Mol. Biol. Evol.* **33**, 1358–1369 (2016).
- Li, M. & Gu, W. A critical role for noncoding 5S rRNA in regulating Mdmx stability. *Mol. Cell* **43**, 1023–1032 (2011).
- Listerman, I., Bledau, A.S., Grishina, I. & Neugebauer, K.M. Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III. *PLoS Genet.* **3**, e212 (2007).
- White, R.J. Transcription by RNA polymerase III: more complex than we thought. *Nat. Rev. Genet.* **12**, 459–463 (2011).
- Ule, J. Alu elements: at the crossroads between disease and evolution. *Biochem. Soc. Trans.* **41**, 1532–1535 (2013).
- Jeck, W.R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).
- Li, Z. *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* **22**, 256–264 (2015).
- Konkel, M.K., Walker, J.A. & Batzer, M.A. LINEs and SINES of primate evolution. *Evol. Anthropol.* **19**, 236–249 (2010).
- Denli, A.M. *et al.* Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* **163**, 583–593 (2015).
- Guerousov, S. *et al.* An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **349**, 868–873 (2015).
- Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
- Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends Genet.* **30**, 121–123 (2014).
- Morris, K.V. & Mattick, J.S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–437 (2014).

## ONLINE METHODS

**Culture of cell lines and model organisms.** HeLa, HEK293T, A549, HepG2, HCT116, N2a, NIH3T3, and THP-1 cells all originated from the ATCC. Cells were tested for mycoplasma by a PCR-based method as well as DAPI staining, to ensure the absence of contamination. Among all cell lines used, only HEK293T was found in the database of commonly misidentified cell lines that is maintained by ICLAC and NCBI Biosample. HEK293T was used as one of the cell lines to confirm the findings from the other human cell lines, and it was authenticated by short-tandem-repeat profiling. The other cell lines used in this study were not authenticated. THP-1 cells were maintained with RPMI-1640; A549 cells were cultured with F-12K; HEK293T, HeLa, HepG2, HCT116, N2a, and NIH3T3 cells were cultured with DMEM. All cells were cultured under standard conditions including 10% FBS and 1% penicillin/streptomycin at 37 °C under 5% CO<sub>2</sub>. For THP-1 differentiation, cells were induced by PMA (Sigma) at a concentration of 50 ng/ml for 4 h, and then the PMA was washed off. After an overnight incubation, cells were harvested for use in FACS experiments or RNA extraction. All the model organisms including *S. cerevisiae* (BJ2168), *S. pombe* (PRI09), *C. elegans* (N2, male and hermaphrodite), *D. melanogaster* (CS, male and female), and *D. rerio* (AB, male and female) were cultured according to standard methods. Protocols involving *D. rerio* were approved by the Institutional Animal Care and Use Committee at The University of Science and Technology of China. For experiments using the model organisms, no statistical method was used to predetermine sample size. The experiments were not randomized and were not performed with blinding to the conditions of the experiments.

**Plasmid construction.** All plasmids were constructed with restriction-enzyme digestion and ligation or with recombinant methods (Vazyme c113-02). Oligonucleotide sequences for primers used in plasmid construction, probe preparation, siRNAs, and biotin-labeled nucleic acids are listed in **Supplementary Table 1**. The shRNA plasmid for knockdown of PTBP1 mRNA (shPTBP1-1, TRCN000001062) with negative-control shRNA (SHC002) was obtained from the MISSION shRNA Library (Sigma). For the overexpression of FLAG-tagged PTBP1 and G3BP2, the vector was p3×FLAG-Myc-CMV-24. For the overexpression of h5S-OT, M-gene and m5S-OT, the promoter in the vector was the CMV promoter. All plasmids were sequenced for confirmation. Further information about these plasmids is available upon request.

**Transfection of plasmids, siRNAs and ASOs.** Plasmid transfection was conducted with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Transfection of siRNAs was conducted with Lipofectamine 2000 or Oligofectamine (Invitrogen) according to the manufacturer's protocol. All siRNAs were subjected to BLAST searching to ensure the absence of hits with more than 17-nt matches in the corresponding genomes<sup>38</sup>. 2'-O-methyl RNA/DNA antisense oligonucleotides (ASOs), which were modified by changing the five nucleotides at the 5' and 3' ends into 2'-O-methyl ribonucleotides, were synthesized by RiboBio. All bases of ASOs were converted into phosphorothioate oligonucleotides<sup>39</sup>.

**PCR reactions.** RNA was extracted with TRIzol reagent (Invitrogen) according to the manufacturer's protocol. For RT-PCR, complementary DNA was synthesized from RNA with a GoScript Reverse Transcription System (Promega) according to the manufacturer's protocol, with the corresponding primers. For PCR with DNA template, DNA was isolated with phenol/chloroform extraction. Quantitative real-time PCR was performed with GoTaq SYBR Green qPCR Master Mix (Promega) on a PikoReal 96 real-time PCR system (Thermo Scientific) according to standard procedures. All PCR products were sequenced for confirmation. Owing to the sequence overlap between 5S rRNA and 5S-OT, primers specific for 5S-OT (avoiding the overlapped region) were used to amplify 5S-OT. Primers for 5S rRNA would inevitably amplify the 5S sequence also present in 5S-OT, although the total amount as well as the amount of nascent transcript (in the nuclear run-on assays) of 5S rRNA was far greater (generally ≥30 fold) than that of 5S-OT in both human and mouse cells. In the RT-PCR of 5S-OT from model organisms, total RNA was isolated from *S. cerevisiae* (mixed stage), *S. pombe* (mixed stage), *C. elegans* (mixed stage), *D. melanogaster* (adults), and *D. rerio* (adults). All primer information is included in **Supplementary Table 1**.

**3' RACE and 5' RACE.** 3' RACE and 5' RACE of h5S-OT and m5S-OT were performed with a SMARTer RACE cDNA Amplification kit (Clontech) with the

primers listed in **Supplementary Table 1**. 5' RACE of the SMARTer RACE cDNA Amplification kit relies on the terminal transferase activity of the RT enzyme, which adds 3–5 residues to the 3' end of the first-strand cDNA, thus ensuring amplification of the full 5' end of the RNA. 3' RACE with the SMARTer RACE cDNA Amplification kit relies on the 3' poly(A) tail of the RNA.

**Northern blotting.** Sense and antisense digoxigenin-labeled RNA probes were prepared with a DIG Northern Starter Kit (Roche) with the corresponding PCR products used as a template for T7 transcription, according to the manufacturer's protocol. 20 µg of total RNA and RiboRuler High Range RNA Ladder (Thermo Scientific) were loaded on a 2% agarose gel containing 1% formaldehyde and run for 1 h in MOPS buffer. RNA was transferred onto Hybond-N+ membranes (GE Healthcare) by capillary transfer. Hybridization was performed at 60 °C overnight. Detection was performed according to the manufacturer's protocol (Roche, DIG Northern Starter Kit). Images were taken with an ImageQuant LAS4000 Biomolecular Imager (GE Healthcare).

**Fluorescence in situ hybridization (FISH) and DNA/RNA double FISH.** RNA probes were generated with a Transcript Aid T7 High Yield Transcription Kit (Thermo Scientific), with the corresponding insertion in the T vector as a template, and then labeled with Alexa Fluor546, by using a ULYSIS Nucleic Acid Labeling Kit (Invitrogen), which added a fluor on every G in the probe to amplify the fluorescence intensity. Fixed cells and RNA probes were denatured at 80 °C for 10 min and then incubated at 42 °C for 15–17 h with 30 ng/µl human Cot-1 DNA (Life Technologies) and 500 ng/µl yeast total RNA (Ambion). Slides were washed with 2× SSC at 45 °C for 10 min. For DNA/RNA double FISH, DNA probes were amplified with FAM-labeled primers by using genomic DNA as a template; hybridization was then performed under the same conditions as those for the RNA FISH.

**Quantification of RNA copy number per cell.** DNA fragments corresponding to human 5S-OT and mouse 5S-OT were amplified with cDNA, and then purified fragments were used to plot standard curves through real-time PCR. HeLa, 293T, 3T3 and N2a total RNA was extracted from 1.0 × 10<sup>5</sup> cells, and cDNA was then synthesized. The copy numbers per cell in each cell line were calculated on the basis of cell numbers and the Ct value by using the standard curve.

**RNase-protection assay (RPA).** RPAs were carried out as previously described with modifications<sup>40</sup>. Biotin-labeled RNAs were prepared with a Biotin RNA labeling kit (Epicentre), then treated with RNase-free DNase I (Promega) and extracted with phenol/chloroform/isoamyl alcohol. 10 pmol labeled RNAs was mixed with 20 µg total RNA, precipitated with 3 M NaAc and absolute ethyl alcohol, then resuspended with hybridization buffer (1 mM EDTA, 80% formamide, 400 mM NaCl, and 40 mM PIPES, pH 6.4). After denaturation at 90 °C for 5 min, the RNA mixture was hybridized at 45 °C overnight, treated with RNase A/T1 (Ambion) for 15 min at 30 °C, extracted with phenol/chloroform/isoamyl alcohol, and separated on 5% urea page gels. RNA was transferred onto Hybond-N+ membranes (GE Healthcare). Detection was performed with a Chemiluminescent Biotin-labeled Nucleic Acid Detection Kit according to the manufacturer's protocol (Beyotime Biotechnology). Biotin-labeled h5S-OT RNA was the experimental probe, and a fragment of h5S-OT without the antisense Alu ( $\Delta$ Alu probe) was used for comparison (the membrane with the  $\Delta$ Alu probe was overexposed to better demonstrate the absence of a protected RNA band).

**Nuclear run-on assay.** Nuclear run-on assays were carried out as described previously<sup>31</sup>, with modifications. For nuclear isolation, cells were rinsed with PBS and harvested in ice-cold hypotonic solution (150 mM KCl, 4 mM MgOAc, and 10 mM Tris-HCl, pH 7.4) and were pelleted by centrifugation. Then pellets were resuspended in lysis buffer (150 mM KCl, 4 mM MgOAc, 10 mM Tris-HCl, pH 7.4, 0.5% NP-40, and 10% glycerol). The crude nuclei were then prepared by sucrose density gradient centrifugation. The nuclear run-on mixture (10 mM ATP, CTP, GTP, BrUTP, and the crude nuclei) was incubated at 30 °C for 5 min in the run-on buffer (10 mM Tris-HCl, pH 7.5, 5 mM MgCl<sub>2</sub>, 150 mM KCl, 1% sarbyl, and 2% DTT) in the presence of RNase inhibitor (Promega). The RNA was isolated by TRIzol reagent (Life Technologies), per the manufacturer's instructions, and DNA was removed by DNase I (Promega) treatment. Nascent transcripts were immunoprecipitated with anti-BrdU antibody (Abcam, ab1893;

validation provided on the manufacturer's website.) and converted to cDNA for use in real-time PCR assays.

**ChIRP.** The ChIRP protocol was modified from a previously described method<sup>13</sup>. Log-phase cells were cross-linked in a UV cross-linker (UVP) at 200-mJ strength. The cells were pelleted and resuspended in swelling buffer (0.1 M Tris, pH 7.0, 10 mM KOAc, and 15 mM MgOAc, with freshly added 1% NP-40, 1 mM DTT, Complete protease inhibitor, and 0.1 U/μl RNase inhibitor) for 10 min on ice. Cell suspensions were then homogenized and pelleted at 2,500g for 5 min. Nuclei were further lysed in nuclear lysis buffer (50 mM Tris, pH 7.0, 10 mM EDTA, and 1% SDS, with freshly added 1 mM DTT, Complete protease inhibitor, and 0.1 U/μl RNase inhibitor) on ice for 10 min and were sonicated with a Sonics Vibra-Cell until most chromatin had solubilized, and DNA was in the size range of 100–500 bp. Chromatin was diluted in two volumes of hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris, pH 7.0, 1 mM EDTA, 15% formamide, 1 mM DTT, protease inhibitor, and 0.1 U/μl RNase inhibitor). Biotin probes (100 pmol) were added to 3 ml of diluted chromatin, which was mixed by end-to-end rotation at 37 °C for 4 h. M-280 Streptavidin Dynabeads (Life Technologies) were washed three times in nuclear lysis buffer, blocked with 500 ng/μl yeast total RNA and 1 mg/ml BSA for 1 h at room temperature, then washed three times again in nuclear lysis buffer before being resuspended. 100 μl washed/blocked Dynabeads was added per 100 pmol of biotin-DNA oligonucleotides, and the mixture was then rotated for 2 h at 37 °C. Beads were captured with magnets (Life Technologies) and washed five times with a 40× volume of wash buffer (2× SSC, 0.5% SDS, 0.1 mM DTT, and fresh PMSF). Beads were then subjected to RNA elution and DNA elution.

**Chromatin immunoprecipitation (ChIP).** ChIP was carried out as previously described, with modifications<sup>31</sup>. Cells were cross-linked in a UV cross-linker (UVP) at 200-mJ strength. After being washed with PBS twice, cell pellets were lysed in 1 ml of SDS lysis buffer (1% (w/v) SDS, 10 mM EDTA, and 50 mM Tris-HCl, pH 8.1) containing Complete protease-inhibitor cocktail (Roche) and were incubated for 20 min on ice. Cell extracts were sonicated for 5 min with a Sonics Vibra-Cell to obtain up to 500-bp DNA fragments. A 100-μl sample of the supernatant was saved as input. The remaining sample was diluted 1:10 in ChIP dilution buffer (0.01% (w/v) SDS, 1.1% (v/v) Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, and 167 mM NaCl) containing protease inhibitors. The chromatin solution was precleared and immunoprecipitated with antibody to pol II (Santa Cruz Biotechnology, sc-9001); validation provided on the manufacturer's website.). The immunocomplexes were eluted in 1% (w/v) SDS and 50 mM NaHCO<sub>3</sub>, and cross-links were reversed for 6 h at 65 °C. Samples were digested with proteinase K for 1 h at 45 °C, and the DNA was extracted with phenol/chloroform/isoamyl alcohol. Eluted DNA was subjected to quantitative real-time PCR detecting enriched genomic DNA regions with the corresponding PCR primer pairs.

**Western blotting.** For western blots, samples were separated on SDS-PAGE gels and then transferred to PVDF membranes (Millipore). Membranes were processed according to the ECL western blotting protocol (GE Healthcare). The following antibodies were used in western blots: anti-U2AF65 (Sigma, U4758); anti-G3BP-2a (Santa Cruz Biotechnology, sc-161612); anti-pol II (Santa Cruz Biotechnology, sc-9001); anti-NOSIP (Santa Cruz Biotechnology, sc-137117); anti-KIAA1191 (Santa Cruz Biotechnology, sc-243180); anti-FLAG (Sigma, F1804); and anti-β-actin (Transgene, HC201). Antibody validation is provided on the manufacturers' websites.

**Pulldown of biotinylated RNA.** RNA pulldown was carried out as previously described, with modifications<sup>41</sup>. Biotin-labeled RNAs were prepared with a Biotin RNA labeling kit (Epicentre). Biotinylated RNAs were then treated with RNase-free DNase I (Promega) and extracted with phenol/chloroform/isoamyl alcohol. Cell lysates were prepared from 10<sup>7</sup> cells that were washed with cold 1× PBS and lysed in 4 mL of RNA immunoprecipitation (RIP) buffer (150 mM KCl, 25 mM Tris, pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP-40, and 1× protease-inhibitor cocktail (Roche)) for 30 min at 4 °C with gentle agitation, after which the lysates were sonicated for 10 min. Lysates were cleared of cell debris by centrifugation at 13,000 r.p.m. for 20 min. Protein concentrations were determined with a BCA protein assay kit (Pierce) with BSA as a standard. Ten picomoles of biotinylated RNA was heated for 10 min to 60 °C and slow-cooled over the course of 40 min

to 4 °C. RNA was mixed with 1 mg of cell lysate in RIP buffer supplemented with 0.1 mg/mL yeast total RNA, 5 mM MgCl<sub>2</sub>, and 1 U/mL RNase inhibitor (Promega) and incubated for 2 h at 4 °C with gentle rotation. Twenty-five microliters of washed M280 Streptavidin magnetic Dynabeads (Invitrogen) was added to each binding reaction and further incubated for 2 h at 4 °C. Beads were washed with supplemented RIP buffer once and supplemented RIP buffer with 500 mM NaCl twice (with each wash carried out for 5 min at 4 °C) and then boiled in SDS loading buffer and subjected to SDS-PAGE. Proteins on the gel were visualized with a silver-staining kit (Sangon) or western blotting.

**Mass spectrometry.** Specific silver-stained bands were cut, digested and extracted. The masses of the peptides in the extract were then measured by MS to obtain the peptide mass fingerprints. Next, peptides were selected to undergo fragmentation via tandem MS. Both the MS and tandem MS data were searched against protein sequence databases to determine the proteins present in the gel.

**RNA pulldown with biotinylated antisense oligonucleotides.** RNA pulldown with 5'-biotinylated AS oligos was modified from a previously described method<sup>11</sup>. Cells were cross-linked in a UV cross-linker (UVP) at 200-mJ strength. The cells were pelleted and resuspended in RIPA buffer (50 mM Tris-Cl, pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 0.1% SDS, 1 mM DTT, Complete protease inhibitor, and 0.1 U/μl RNase inhibitor) for 10 min on ice, then harvested and sonicated for 10 min. Lysate was cleared of cell debris by centrifugation at 13,000 r.p.m. for 20 min. Biotinylated AS oligos (100 pmol) were added to the supernatant at 4 °C for 2 h. M-280 Streptavidin Dynabeads (Life Technologies) were washed three times in RIPA buffer, blocked with 500 ng/μl yeast total RNA and 1 mg/ml BSA for 1 h at room temperature, then washed three times again in RIPA buffer before being resuspended. 50 μl washed/blocked Dynabeads was added per 100 pmol of biotin-DNA oligonucleotides, and the mixture was then rotated for 4 h at 4 °C. Beads were captured with magnets (Life Technologies) and washed five times with RIPA buffer supplemented with 500 mM NaCl. RNAs and proteins were eluted from beads for further analysis.

**Cross-linking immunoprecipitation (CLIP) assay.** CLIP was carried out as previously described with some modifications<sup>31</sup>. Briefly, the cultured cells were irradiated in a UV cross-linker (254 nm, 400 mJ/cm<sup>2</sup>, 1 min) and then harvested in ice-cold lysis buffer (10 mM HEPES, pH 7.4, 200 mM NaCl, 30 mM EDTA, and 0.5% Triton-X 100), 100 units/ml RNasin Plus RNase Inhibitor (Promega), 1.5 mM DTT, and 1× protease-inhibitor cocktail (Sangon). Cells were sonicated for 5 min with a Sonics Vibra-Cell, the cell suspension was centrifuged at 12,000g for 15 min at 4 °C, and the supernatant was collected. Antibody or IgG (as control) was added for antigen coupling and incubated 1 h at 4 °C, and then Protein G Dynabeads (Life Technology) suspension was then added and allowed to bind for at least 3 h at 4 °C. The antibody-Protein G bead complexes were washed five times with lysis buffer, and one-fifth of the bead volume after the last wash was saved for western blotting. The remaining antibody-Protein G bead complexes were resuspended in 50 μl elution buffer (100 mM Tris, pH 7.8, 10 mM EDTA, and 1% SDS) and were digested with 30 μg of proteinase K at 65 °C for 1 h; this was followed by extraction with phenol/chloroform, pH 4.2, to obtain RNA. The following antibodies were used: anti-U2AF65 (Sigma, U4758) and anti-FLAG (Sigma, F1804). Antibody validation is provided on the manufacturers' websites.

**Protein purification.** The recombinant human U2AF65 was expressed in *E. coli* BL21(DE3) cells with the pET-22b vector, as induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) at 16 °C for 20 h. The protein was purified with a Ni<sup>2+</sup>-chelating column and then subjected to size-exclusion chromatography on a Superdex 75 16/60 column.

**In vitro splicing assay.** *In vitro* splicing assays (25 μl) contained 1 mM ATP, CTP, UTP, and GTP, 50 mM MgCl<sub>2</sub>, and 8 U HeLa nuclear extract in the splicing buffer and were carried out with a HeLaScribe Nuclear Extract *in vitro* Transcription System (Promega). 100 ng of C1ORF43 linear plasmid was added to the nuclear extract, with or without the addition of 100 ng recombinant U2AF65 and 100 ng *in vitro*-transcribed h5S-OT RNA. Reactions were incubated for 1 h at 30 °C. RNA was recovered with TRIzol reagent and converted to cDNA for use in RT-PCR analyses.

**Fluorescence-activated cell sorting (FACS).** Antibodies used for flow cytometric analyses were APC-Cy7-conjugated lineage-marker monoclonal antibodies CD11b (BD Biosciences; validation provided on the manufacturer's website). Samples were treated with FcBlock (BD Biosciences) for 15 min before incubation with antibody cocktail for an additional 30 min. Data were acquired on an LSRII flow cytometer (BD Biosciences) and analyzed with FlowJo v7.6.2 (Tree Star).

**mRNA-library preparation and sequencing.** For mRNA-seq, 5 µg total RNA treated with siRNA or plasmid was iron-fragmented at 95 °C and then subjected to end repair and 5'-adaptor ligation. Then reverse transcription was performed with random primers containing 3' adaptor sequences and randomized hexamers. The cDNAs were purified and amplified, and PCR products of 200–500 bp were purified, quantified and stored at –80 °C until sequencing. For high-throughput sequencing, the libraries were prepared according to the manufacturer's instructions and subjected to 151-nt paired-end sequencing with an Illumina Nextseq 500 system. We sequenced each library to a depth of 10–50 million read pairs. To obtain clean reads, adapters were removed with cutadapt.

**Bioinformatic analyses of gene-expression levels, PSI of cassette exons, and repetitive-element annotation.** Gene-expression levels were calculated with TopHat2 (cutoff with  $P < 0.01$ , RPKM  $\geq 1$  and fold change  $\geq 2$ ) with the following genome releases: *Homo sapiens*, hg19; *Mus musculus*, mm9. Repetitive-element annotation was from the UCSC RepeatMasker database (rmsk.txt, updated 2009/4/27)<sup>42</sup>. The pipeline for PSI first aligned the reads to the genome by using Bowtie, allowing two mismatches, to filter out the continuous mapping reads. The unmapped reads were aligned to a custom library of exon-exon junctions (EEJs) with Bowtie (–v 2 –m 20–best) with at least a 4-nt overhang. The custom EEJ library was generated by using existing RNA-seq data, EST and cDNA evidence, gene annotations, and evolutionary conservation. For the case of single exon-skipping events, we generated EEJs for E1AS, ASE2 and E1E2 (AS represents the alternative exon, and E1 and E2 represent the neighboring constitutive exons).

$$\text{PSI} = \frac{\Sigma(\text{E1AS}) + \Sigma(\text{ASE2})}{\Sigma(\text{E1AS}) + \Sigma(\text{ASE2}) + 2\Sigma(\text{E1E2})}$$

$\Delta\text{PSI}$  was calculated by subtracting the PSI of the knockdown group from that of the control, and  $P$  values were generated with two-sided Mann–Whitney  $U$  tests.

**Analysis of splice-site strength.** MaxEntScan was used to calculate maximum entropy scores for 23-bp 3' splice sites<sup>43</sup>.

**CpG-island analysis.** We used the classic but arbitrary definition of a CpG island as a region of at least 200 bp in length<sup>44</sup> with a GC percentage  $>50\%$  and an observed-to-expected CpG ratio greater than 60%.

**Coding-potential analysis.** The Coding Potential Assessment Tool (CPAT)<sup>12</sup> was used to estimate the degree of evolutionary pressure on sequence substitutions acting to preserve an ORF in both human and mouse 5S–OT. Briefly, CPAT is an alignment-free method used to rapidly recognize coding and noncoding transcripts. It uses a logistic regression model built with four sequence features: open-reading-frame size, open-reading-frame coverage, Fickett TESTCODE statistic, and hexamer usage bias.

**Image processing and quantification.** Cell images were processed with ImageJ image-acquisition software, and color channels were also merged in ImageJ. Quantification of RT–PCR band intensities was also performed with ImageJ. An area above each band, of the same size as the corresponding band, was used for background subtraction.

**Confocal microscopy.** FISH images were taken with an AndorXonEM+ DV897K EM CCD camera mounted on an Andor Revolution XD laser confocal microscopy system (Andor Technology), with Andor IQ 10.1 software.

**Statistical analysis.** Either Student's  $t$  tests or Mann–Whitney  $U$  tests were used to calculate  $P$  values, as indicated in the figure legends. For Student's  $t$  tests, the values reported in the graphs represent averages of three independent experiments, with error bars showing s.e.m. After analysis of variance with  $F$  tests, the statistical significance and  $P$  values were evaluated with Student's  $t$  tests.

38. Shan, G. RNA interference as a gene knockdown technique. *Int. J. Biochem. Cell Biol.* **42**, 1243–1251 (2010).
39. Dubin, R.A., Kazmi, M.A. & Ostrer, H. Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene* **167**, 245–248 (1995).
40. Carey, M.F., Peterson, C.L. & Smale, S.T. The RNase Protection Assay. *Cold Spring Harb. Protoc.* **2013**, pdb.prot071910 (2013).
41. Maamar, H., Cabili, M.N., Rinn, J. & Raj, A. linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes Dev.* **27**, 1260–1271 (2013).
42. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
43. Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
44. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).