ARTICLE

doi:10.1038/nature09479

U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation

Daisuke Kaida[1], Michael G. Berg[1], Ihab Younis[1], Mumtaz Kasim[1], Larry N. Singh[1], Lili Wan[1] & Gideon Dreyfuss[1]

In eukaryotes, U1 small nuclear ribonucleoprotein (snRNP) forms spliceosomes in equal stoichiometry with U2, U4, U5 and U6 snRNPs; however, its abundance in human far exceeds that of the other snRNPs. Here we used antisense morpholino oligonucleotide to U1 snRNA to achieve functional U1 snRNP knockdown in HeLa cells, and identified accumulated unspliced pre-mRNAs by genomic tiling microarrays. In addition to inhibiting splicing, U1 snRNP knockdown caused premature cleavage and polyadenylation in numerous pre-mRNAs at cryptic polyadenylation signals, frequently in introns near (<5 kilobases) the start of the transcript. This did not occur when splicing was inhibited with U2 snRNA antisense morpholino oligonucleotide or the U2-snRNP-inactivating drug spliceostatin A unless U1 antisense morpholino oligonucleotide was also included. We further show that U1 snRNA–pre-mRNA base pairing was required to suppress premature cleavage and polyadenylation from nearby cryptic polyadenylation signals located in introns. These findings reveal a critical splicing-independent function for U1 snRNP in protecting the transcriptome, which we propose explains its overabundance.

Messenger RNAs in eukaryotic cells are produced from primary transcripts (pre-mRNAs) by extensive post-transcriptional processing, including 5′ end capping, removal of introns by splicing, and 3′ end cleavage and polyadenylation[1–4]. Each splicing reaction is carried out by a spliceosome, a large RNA–protein complex comprised predominantly of snRNPs[5–8]. The U1, U2, U4, U6 and U5 snRNPs are components of the major (U2-type) spliceosome, whereas a much less abundant (~1%) minor (U12-type) spliceosome is comprised of U11, U12, U4atac, U6atac and U5 snRNPs[5,9–11]. The snRNPs, aided by specific RNA-binding proteins, recognize by snRNA–pre-mRNA base pairing, canonical sequences within pre-mRNAs that define the major- and minor-class introns, including the intron–exon junctions at the 5′- and 3′-splice sites. U1 snRNP has an essential role in defining the 5′ splice site by RNA–RNA base pairing via the 5′ nine nucleotide sequence of U1 snRNA. To form the catalytic core of the spliceosome, the snRNPs come together in 1:1 stoichiometry as a modular machine[5]. However, the abundance of the various snRNPs in cells does not reflect their equimolarity in the spliceosomes. This is particularly striking for U1 snRNP which, at an estimated copy number of ~10^6 molecules per human cell (HeLa), is much more abundant than the other snRNPs in higher eukaryotes[12]. The potential role of the different amounts of the snRNPs is not known.

Our interest in exploring a potential function for cellular snRNP abundance arose from earlier observations that deficiency in the survival of motor neurons (SMN) protein, a key component in snRNP biogenesis[13–17], perturbs the normal abundance of snRNPs in cells (the snRNP repertoire)[18,19] and causes widespread splicing abnormalities[19]. The possible effect of snRNP abundance changes on splicing and the molecular consequences of SMN deficiency in general are of importance because SMN deficiency is the cause of spinal muscular atrophy, an often fatal motor neuron degenerative disease[20–22]. However, the snRNP repertoire changes that occur in an SMN-deficient spinal muscular atrophy mouse model vary in different tissues and are not uniform for all of the snRNPs[18,19], including both downregulation and upregulation in the levels of several snRNPs simultaneously, making them difficult to recapitulate. To circumvent this, we investigated the effect of functional reduction of individual snRNPs on the transcriptome using antisense morpholino oligonucleotide (AMO). Our experiments revealed an unexpected function for U1 snRNP in protecting pre-mRNAs from premature cleavage and polyadenylation (PCPA), distinct from its role in splicing.

Functional knockdown of U1 snRNP with AMO

To decrease the amount of functional U1 snRNP, we designed an AMO that covers the 5′ end of U1 snRNA (U1 AMO) to block its binding to 5′ splice sites. To confirm the binding of U1 AMO to U1 snRNP and determine the amount required to inhibit it in cells, we performed an RNase H protection assay. Extracts from cells transfected with a scrambled control AMO[23,24] or various concentrations of U1 AMO were incubated with RNase H and an antisense DNA oligonucleotide probe also complementary to U1 snRNA's 5′ end sequence (Fig. 1a). A dose-dependent decrease in the amount of cleaved U1 snRNA was observed as the amount of transfected U1 AMO was increased (Fig. 1a), indicating that the U1 AMO prevented the antisense DNA oligonucleotide probe from binding and eliciting RNase H digestion. Complete or near-complete interference with U1 snRNA 5′ base pairing in cells was observed with 7.5 μM of U1 AMO (Fig. 1a). In addition, we used *in situ* hybridization with a LNA probe complementary to U1 snRNA's 5′ sequence (nucleotides 1–25) to determine if the U1 AMO was bound to the same sequence in cells. The images (Fig. 1b) demonstrate that the U1 AMO indeed shields U1 snRNA's 5′ sequence in a dose-dependent manner and that this sequence is completely inaccessible at 7.5 μM U1 AMO, the concentration that was used in all subsequent U1 AMO transfection experiments. To confirm that the U1 AMO inhibited the activity of U1 snRNP directly, we tested its effect on the *in vitro* splicing. As shown in Fig. 1c, U1 AMO, but not control AMO, strongly decreased the amount of spliced product for Ad2ΔIVS. Thus, U1 AMO functionally inactivated U1 snRNP both *in vivo* and *in vitro*.

Unspliced pre-mRNA accumulation after splicing inhibition

To obtain a high-resolution global picture of the transcriptome changes that occurred upon U1 snRNP knockdown, including effects

[1]Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6148, USA.

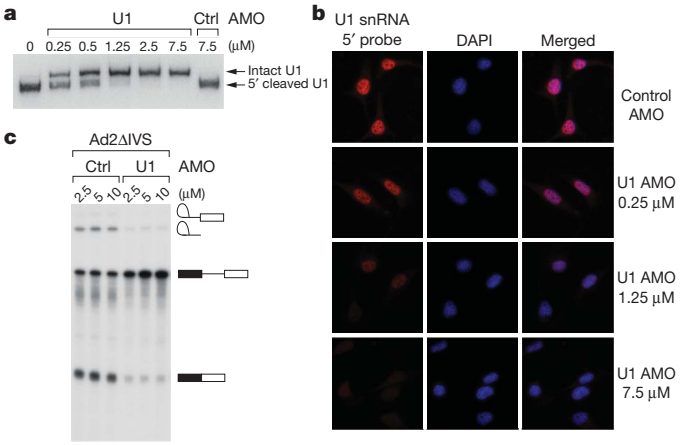

Figure 1 | U1 AMO binds to the 5′ sequence of U1 snRNA and inhibits its splicing activity. **a**, HeLa cells were transfected with the indicated concentrations of control and U1 AMO for 8 h. RNase H protection assay was performed using total cell extracts and U1 snRNA was detected by northern blotting. **b**, *In situ* hybridization was performed on HeLa cells transfected with varying concentrations of U1 AMO, as indicated, for 8 h using a biotin-labelled LNA probe to the U1 snRNA (left panels) followed by fluorescent Alexa Fluor 594 streptavidin conjugate detection. Nuclei were visualized with DAPI (middle panels) and merged images are shown (right panels). All images were taken at ×40 magnification. **c**, [α-^{32}P]UTP-labelled Ad2ΔIVS pre-mRNA was spliced *in vitro* in the presence of control or U1 AMOs at the indicated concentrations. Splicing product identities are depicted to the right of the gel.

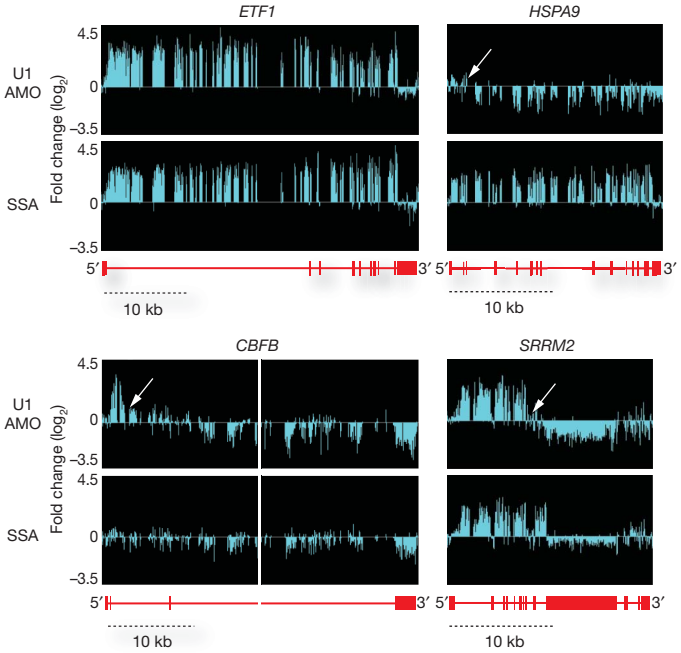

Figure 2 | Genomic tiling arrays identify unspliced pre-mRNAs after U1 AMO and SSA treatment. RNA samples prepared from control or U1-AMO-transfected cells (7.5 μM, 8 h) or SSA-treated cells (100 ng ml^{-1}, 8 h) were analysed using genomic tiling array. Fold changes ($\log_2$) of signal intensities of U1-AMO-transfected and SSA-treated cells compared to control cells are shown above the corresponding structure of each gene. With a scale shown below, gene structures are depicted in red, with horizontal lines indicating introns and boxes indicating exons. The middle part of the *CBFB* gene (~35 kb) was removed. White arrows indicate points showing an abrupt drop of the signal (inflection points).

on introns and exons, we analysed total RNA prepared from HeLa cells transfected with either U1 or control AMOs using Affymetrix GeneChip Human Tiling 2.0R E arrays. This high-density genomic tiling array includes tiled probes (25-mer oligonucleotides spaced at 10 nucleotides) covering the entire genomic sequence of chromosomes 5, 7 and 16, which are estimated to contain ~3,600 genes. All experiments were carried out as separate biological triplicates and treatments were for 8 h to allow transfected cells to recover and for sufficient signals to accumulate above background. As a reference, we treated cells in parallel with the potent and general splicing inhibitor spliceostatin A (SSA), which targets the splicing factor SF3b, a component of U2 snRNP[25]. As the amount of each pre-mRNA is typically very small and difficult to detect, large increases in intron signals provided the most definitive evidence for the accumulation of unspliced pre-mRNAs, ensuring that the corresponding sequences were actively transcribed and their unspliced transcripts are sufficiently stable to be scored. Statistical analysis was performed to identify significant changes that exceeded the following thresholds: fold-change ≥2, *P*-value <0.01 and length of affected region ≥100 nucleotides (corresponding to three or more consecutive probes). This identified 319 genes that showed accumulation typically of one or more introns in either U1-AMO-transfected (211 genes) or SSA-treated (216 genes) cells. From the outset we expected two patterns due to splicing inhibition. A general reduction in all signals from a transcript could reflect that unspliced pre-mRNA is less stable and rapidly degraded; however, we did not include them in our analysis as they could also potentially result from transcriptional downregulation. We also expected that U1 AMO and SSA would show similar patterns of intron accumulation throughout the transcript as a result of splicing inhibition in general, but this was observed for only 98 genes (30.7%), as exemplified by *ETF1* (Fig. 2). A small number of genes (41 genes; 12.9%) showed other profiles but did not exhibit any coherent pattern. Unexpectedly, however, most genes (180 genes; 56.4%) showed different patterns for U1 AMO and SSA (*HSPA9*, *SRRM2* and *CBFB* in Fig. 2).

U1 snRNP knockdown causes PCPA

Most notably, the majority of the affected genes in the U1-AMO-transfected cells showed a similar pattern consisting of strong intron signals that terminated prematurely relative to SSA, for example in *HSPA9*, *SRRM2* and *CBFB* as indicated by arrows in Fig. 2. This usually occurred in introns in the first quarter of the gene with a strong 5′ bias, as shown for several functionally diverse genes in Fig. 3. Notably, in U1-AMO-transfected cells there was an abrupt decrease in unspliced intron signals frequently within 3–5 kb from the start of the transcript. From the point at which a sharp drop in the signals in U1-AMO-treated cells was observed, all downstream signals, including exons, were consistently lower than those of control and the SSA-treated samples. To define the termination point, we characterized cDNA fragments from this region (indicated by arrows in Fig. 3), including 3′ rapid amplification of cDNA ends (RACE) products for transcripts of several genes that showed this pattern from U1-AMO-transfected cells, of which *NR3C1* and *STK17A* were sequenced (Fig. 4a). Surprisingly, this revealed that the transcripts had a poly(A) sequence at the 3′ end that is not found in the genomic sequence and which was added post-transcriptionally ~20 nucleotides downstream of a potential polyadenylation signal (PAS), typically AAUAAA (Fig. 4a). These findings indicate that pre-mRNAs are prematurely cleaved and polyadenylated within an intron when the levels of functional U1 snRNP are reduced.

To determine if these putative PASs are functionally relevant, we constructed a *NR3C1* mini-gene comprising exons 2 and 3 which flank a truncated intron 2 where the cleavage and polyadenylation occurred. An identical plasmid was also constructed with a mutation in the putative PAS in intron 2 inferred from the tiling array results (Fig. 4b). Cells expressing each of these plasmids were then transfected with either control or U1 AMOs. Wild-type plasmid-expressing cells showed PCPA at the cryptic PAS only after U1 AMO treatment (Fig. 4b), consistent with the result of the endogenous *NR3C1* gene. No cleavage and polyadenylation occurred in the transcript containing a mutant PAS in either control or U1-AMO-transfected

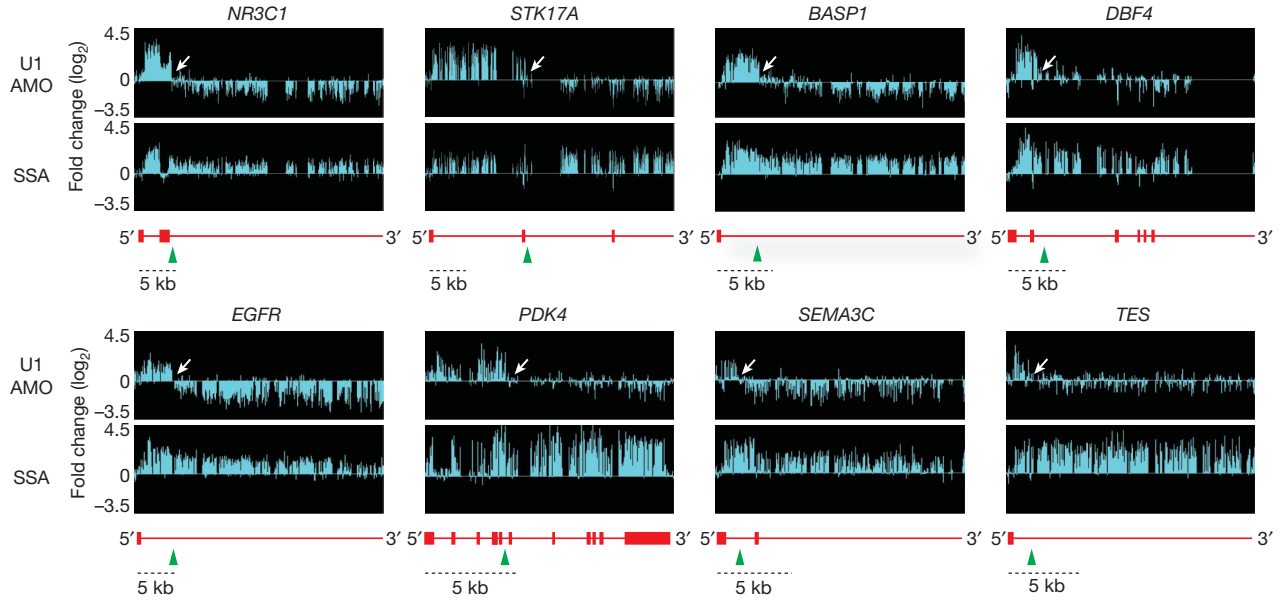

Figure 3 | Premature termination in introns of pre-mRNAs in U1-AMO-transfected cells. Representative examples of genes differentially affected by U1 AMO and SSA are shown. The sudden drop in signals in U1-AMO-transfected cells is indicated by green arrowheads and white arrows. With a scale shown below, gene structures are depicted in red with horizontal lines indicating introns and boxes indicating exons.

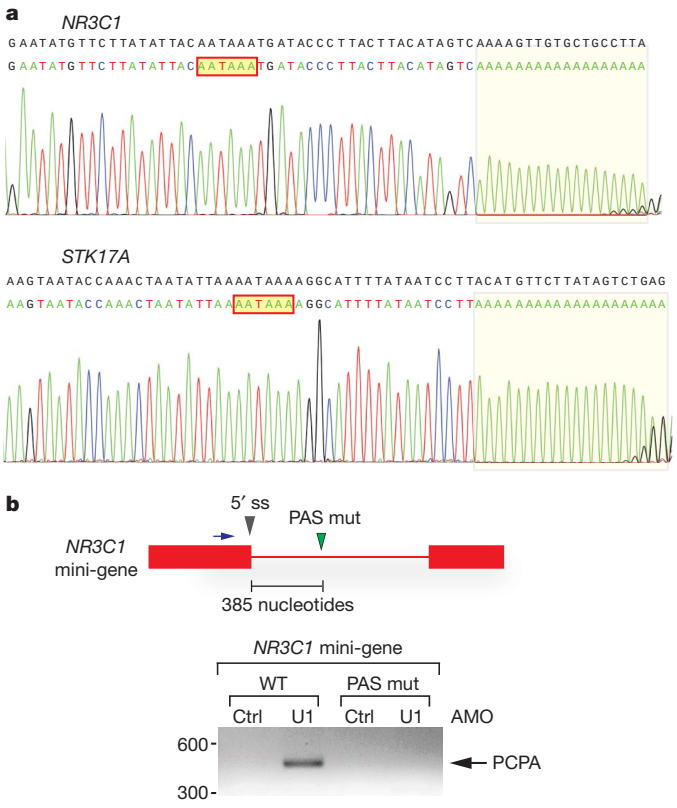

Figure 4 | Prematurely terminated pre-mRNAs are polyadenylated from cryptic PASs in introns. **a**, 3′ RACE using nested PCR was performed to detect polyadenylated mRNAs using total RNA from U1-AMO-transfected cells (7.5 μM, 8 h). Sequencing results of the 3′ RACE products for the *NR3C1* and *STK17A* genes are shown with the corresponding genomic sequence in black. The poly(A) tails are shaded and the putative PASs are indicated in red-bordered boxes. **b**, HeLa cells were transfected with the wild-type and PAS mutant (mutated from AAUAAA to GAAUUC) *NR3C1* mini-gene construct followed by either control or U1 AMOs. 3′ RACE was performed as described in **a**. The mini-gene structure is depicted above the gel, with a blue arrow indicating the forward primer for 3′ RACE. Grey arrowhead, PAS mutation; PCPA, premature cleavage and polyadenylation. Band sizes (bp) are indicated to the left of the gel.

cells (Fig. 4b), demonstrating that this cryptic PAS is functional and that this process is similar to that which occurs normally at the 3′ ends of mRNAs.

PCPA suppression is a non-splicing function of U1 snRNP

Unlike U1 AMO, tiling arrays with SSA or U2 AMO (which generally showed a tiling array pattern of splicing inhibition similar to that of SSA; Supplementary Fig. 1) did not show PCPA. To test this directly, we used an oligo(dT) reverse primer to amplify *NR3C1* and *STK17A* transcripts in control, U1 or U2-AMO-treated cells and SSA-treated cells (Fig. 5a). As expected, PCPA was observed for U1 AMO; however, none or very little was seen in U2-AMO- and SSA-treated cells, which could be due to destabilization of U1 snRNP binding upon U2 snRNP inhibition. Notably, PCPA still occurred in cells treated simultaneously with U1 AMO and SSA, indicating that the effect of U1 inhibition is dominant over splicing inhibition (Fig. 5b). We conclude that U1 suppresses cleavage and polyadenylation and that this is not a consequence of the splicing inhibition that it causes, as neither U2 AMO nor SSA showed this effect.

The functional PAS from which cleavage and polyadenylation occurred in the *NR3C1* mini-gene upon U1 snRNP depletion is less than 400 nucleotides downstream of the 5′ splice site and it thus seemed likely that U1 snRNP that is base paired to this 5′ splice site suppresses utilization of this cryptic PAS. To test this, we mutated the 5′ splice site, which inactivated splicing as evidenced by the lack of mRNA (Fig. 5c). Whereas the wild-type mini-gene showed PCPA only upon U1 AMO treatment, in 5′ splice site mutant transfected cells, PCPA was observed even with control morpholino treatment. This indicates that U1 snRNP that is base paired to the 5′ splice site is able to suppress the cryptic PAS (Fig. 5c). However, treatment of the 5′ splice site mutant with U1 AMO resulted in ~4-fold increase in PCPA from the same cryptic PAS and concomitantly ~8-fold decrease in polyadenylation from the normal PAS at the transcript's 3′ end (Fig. 5c, d). Therefore, U1 snRNP that is base paired to sequences other than the 5′ splice site also suppresses PCPA. The large number of potential U1 snRNP binding sites in introns precluded identification of other sites from which U1 snRNP might suppress premature utilization of this PAS (Supplementary Fig. 2). These findings indicate that the PCPA results from loss of the base pairing of the 5′ end of U1 snRNA with the pre-mRNA, indicating a

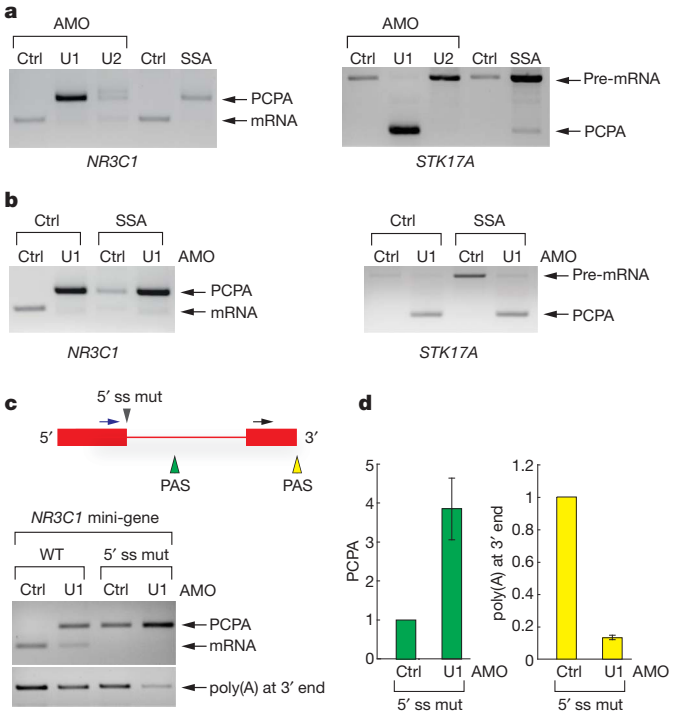

Figure 5 | U1 snRNP suppression of premature cleavage and polyadenylation from a nearby cryptic PAS is splicing independent and requires base pairing. a, 3′ RACE was carried out as described in Fig. 4a on the endogenous *NR3C1* and *STK17A* genes using RNA samples from HeLa cells transfected with control, U1 (7.5 μM, 8 h) or U2 AMO (25 μM, 8 h), or treated with SSA (100 ng ml^{-1}, 8 h). **b**, 3′ RACE was carried out as described in Fig. 4a using RNA samples from HeLa cells transfected with control or U1 AMOs (7.5 μM) with or without SSA (100 ng ml^{-1}) for 8 h. **c**, HeLa cells were transfected with wild-type and 5′ splice site mutant *NR3C1* (mutated from AAGGTAAGA to GTCCATTCA) mini-gene. 3′ RACE was performed as described in Fig. 4a. The mini-gene structure is depicted. Blue arrow, forward primer to detect PCPA; black arrow, forward primer to detect polyadenylation at 3′ end; grey arrowhead, 5′ splice site mutation; green and yellow arrowheads, polyadenylation signals in intron and at 3′ end, respectively. An unspliced and normally cleaved product was too large to detect. **d**, Quantification of PCPA and normal polyadenylation at the 3′ end represented as fold change in control, which was set to one, and U1-AMO-treated cells was performed using real-time PCR. Error bars indicate s.d. ($n = 3$).

U1 snRNP function other than and independent of its known function in splicing.

Discussion

Our experiments here revealed an unexpected function for U1 snRNP in protecting transcripts from PCPA in addition to and independent of its role in splicing. As a reference for U1 AMO, the general splicing inhibitor SSA[25], which inactivates the U2 snRNP component SF3b (refs 5, 26), allowed identification of introns that were stable enough and that accumulate to significantly detectable levels when their splicing was inhibited. As expected, the patterns observed for U1 AMO and SSA (which is similar to that of U2 AMO) showed that both efficiently inhibited splicing. However, U1 snRNP functional reduction had an additional and marked effect, resulting in the failure to produce full-length pre-mRNA from the majority of genes in our data set. We showed that this was due to premature cleavage and polyadenylation from a cryptic PAS, typically in an intron and frequently within the first few kilobases (<5 kb) from the start of RNA polymerase II transcripts. A non-splicing role for a snRNP has been previously shown for U2 snRNP in the 3′-end formation of histone mRNAs[27,28].

The mechanism by which U1 snRNP suppresses PCPA is not presently known. However, as it occurs from canonical PASs, it is reminiscent of previous observations on the capacity of tethered U1 snRNP to regulate normal 3′-end cleavage and polyadenylation from the natural PASs in the last exon[29], and may have features in common with it. For example, the U1 snRNP protein U1-70K can interact directly with the poly(A) polymerase (PAP)[30,31] and inhibit polyadenylation. Targeting 5′-mutated U1 snRNAs with complementarity to sequences in the vicinity (within <500 nucleotides) of the natural PAS at the 3′-terminal exon results in degradation of the transcript because cleavage occurs without addition of a poly(A) tail, leaving the transcript vulnerable to 3′ exonucleases[32]. A considerable number of genes that we surveyed, but that were not included in our analysis, showed a decrease in exon signals or in both introns and exons throughout the transcript in U1-AMO-treated cells. It is possible that in these cases cleavage occurred without subsequent polyadenylation, and the transcript was therefore rapidly degraded. Alternatively, cleavage and polyadenylation may have occurred very close to the transcription start site, making these transcripts difficult to detect. These scenarios are nevertheless consistent with a role for U1 snRNP in suppressing cleavage and polyadenylation throughout the entire pre-mRNA by a similar machinery that until now was thought to only process the 3′ end of mRNA, in an even larger number of genes than our data set presents.

Stochastically, canonical PASs (most frequently AAUAAA or AUUAAA) occur every 2,000 nucleotides, although in several of the genes we studied, including *NR3C1*, *STK17A* and *BASP1*, cryptic PASs are found every 500–800 nucleotides. The strong 5′ bias with which PCPA occurred in these genes upon U1 snRNP functional knockdown indicates that one of the first few cryptic PASs is used. Up to the point at which PCPA occurred, these transcripts also contained many cryptic 5′ splice sites (Supplementary Fig. 2). We propose the following model to explain our observations. Pre-mRNA processing factors, including splicing factors, hnRNP proteins, snRNPs and 3′-end cleavage and polyadenylation factors co-transcriptionally associate with nascent transcripts[33–37]. Direct association of cleavage/polyadenylation factors with the C-terminal domain of RNA polymerase II in the transcription elongation complex has been demonstrated[36]. U1 snRNP associates with nascent transcripts, by base pairing with cognate sequences on the nascent pre-mRNA, including 5′ splice sites and cryptic 5′ splice sites, and inhibits the cleavage/polyadenylation machinery from attacking the pre-mRNA at cryptic PASs. We envisage that when U1 snRNP's base pairing is prevented, as is the case in U1-AMO-transfected cells, cleavage and polyadenylation occur co-transcriptionally at the first actionable PAS that the transcription elongation complex encounters. By actionable PAS, we mean one that has the necessary hexanucleotide consensus and is in an RNP context that makes it accessible and susceptible to attack by the cleavage/polyadenylation machinery unless U1 snRNP base-paired in the vicinity is able to protect it. We suggest that under normal circumstances, this encounter happens after the last strong U1 binding site (5′ splice site or a cryptic 5′ splice site) in the 3′ untranslated region of the terminal exon because a sufficient density of U1 snRNP base paired throughout protects the transcript up to that point. The likelihood of normal or premature termination may be enhanced by the presence of pausing sites[38].

U1 snRNP bound to 5′ splice sites may thus serve a dual purpose—in splicing and suppression of PCPA. The perimeter of U1 snRNP's protective zone is not known, but its binding to 5′ splice site alone is unlikely to be able to protect the majority of introns, which in humans average ~3.4 kb in length[39]. Furthermore, if suppression of actionable PASs was provided only via U1 snRNP bound to 5′ splice sites, 5′ splice-site mutations would be expected to cause premature termination, as opposed, for example, to exon skipping, which would be extremely deleterious and, to our knowledge, has not been observed. Additional U1 snRNP binding sites, including cryptic 5′ splice sites, may function as tethering sites for its activity in suppression of cleavage and polyadenylation in introns. Viewed from this perspective, sequences referred to as cryptic 5′ splice sites may serve a non-splicing purpose to recruit U1 snRNP to protect introns. It is also reasonable to

consider that modulating U1 snRNP levels or its binding at sites that protect actionable PASs could be a mechanism for regulating gene expression, including downregulation of the mRNA or switching expression to a different mRNA produced from a prematurely terminated pre-mRNA. We suggest that the vulnerability to PCPA would be expected to increase with increasing intron size if U1 snRNP and cognate base-pairing sites are not available to protect it. We propose that the large excess of U1 snRNP over what is required for splicing in human cells serves an additional critical biological function, to suppress PCPA in introns and protect the integrity of the transcriptome.

METHODS SUMMARY

Antisense morpholino oligonucleotide (AMO) transfection was performed by electroporation. The sequences of the U1 and control AMOs (Gene Tools) are 5′-GGTATCTCCCCTGCCAGGTAAGTAT-3′ and 5′-CCTCTTACCTCAGTTACAATTTATA-3′, respectively[23,24]. RNase H protection assay was carried out using AMO-transfected cell extracts and antisense DNA oligonucleotide for U1 snRNA (5′-CAGGTAAGTAT-3′). After RNase H treatment, RNA samples were purified and analysed by northern blotting with a U1 snRNA probe (5′-CAAATTATGCAGTCGAGTTTCCCACATTTG-3′). *In situ* hybridization of U1 snRNA was performed with a biotin-labelled LNA probe (5′-GGTATCTCCCCTGCCAGGTAAGTAT-3′). Nuclei were stained by DAPI. For *in vitro* splicing, [α-^{32}P]UTP labelled Ad2ΔIVS pre-mRNA was prepared as previously described[40]. *In vitro* splicing reactions were carried out in 293T whole cell extracts prepared as previously described[41]. Splicing products were resolved on denaturing PAGE, and gels were autoradiographed. For tiling array, labelled cDNA targets were prepared and applied to Affymetrix GeneChip human tiling 2.0R E arrays. Arrays were scanned to produce .CEL files. The .CEL files were analysed using the Affymetrix Tiling Analysis Software (TAS) to produce .BED files of signal intensity and *P*-value. Overlapping regions of two data sets were chosen using Galaxy (http://galaxy.psu.edu/)[42]. We produced .BAR files from the .CEL files using TAS to visualize on the Integrated Genome Browser (Affymetrix). For 3′ RACE, cDNA was synthesized from total RNA using an oligo dT18-XbaKpnBam primer. The first and second (nested) PCR reactions were performed using gene-specific forward primers and the XbaKpnBam reverse primer. For 3′ RACE of the *NR3C1* mini-gene, pcDNA3.1-5′ primer was used as the first primer to distinguish mini-gene RNA from endogenous *NR3C1* RNA. To construct the *NR3C1* mini-gene, DNA fragments of *NR3C1* intron 1-exon 2-intron 2 and *NR3C1* intron 2-exon 3 were amplified and subcloned into pcDNA3.1 vector. The poly(A) site and 5′ splice site were mutated in this construct where indicated. Sequences of all primers are listed in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 February; accepted 9 September 2010.
Published online 29 September 2010.

1. Danckwardt, S., Hentze, M. W. & Kulozik, A. E. 3′ end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* **27,** 482–498 (2008).
2. Gu, M. & Lima, C. D. Processing the message: structural insights into capping and decapping mRNA. *Curr. Opin. Struct. Biol.* **15,** 99–106 (2005).
3. Mandel, C. R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3′-end processing. *Cell. Mol. Life Sci.* **65,** 1099–1122 (2008).
4. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136,** 688–700 (2009).
5. Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–718 (2009).
6. Kambach, C., Walke, S. & Nagai, K. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr. Opin. Struct. Biol.* **9,** 222–230 (1999).
7. Nilsen, T. W. The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* **25,** 1147–1149 (2003).
8. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92,** 315–326 (1998).
9. Hall, S. L. & Padgett, R. A. Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* **271,** 1716–1718 (1996).
10. Patel, A. A. & Steitz, J. A. Splicing double: insights from the second spliceosome. *Nature Rev. Mol. Cell Biol.* **4,** 960–970 (2003).
11. Tarn, W. Y. & Steitz, J. A. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell* **84,** 801–811 (1996).
12. Baserga, S. J. & Steitz, J. A. in *The RNA World* (eds Gesteland, R. F. & Atkins, J. F.) 359–381 (Cold Spring Harbor Laboratory Press, 1993).
13. Fischer, U., Liu, Q. & Dreyfuss, G. The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* **90,** 1023–1029 (1997).
14. Liu, Q., Fischer, U., Wang, F. & Dreyfuss, G. The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell* **90,** 1013–1021 (1997).
15. Meister, G., Buhler, D., Pillai, R., Lottspeich, F. & Fischer, U. A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nature Cell Biol.* **3,** 945–949 (2001).
16. Pellizzoni, L., Yong, J. & Dreyfuss, G. Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298,** 1775–1779 (2002).
17. Wan, L. *et al.* The survival of motor neurons protein determines the capacity for snRNP assembly: biochemical deficiency in spinal muscular atrophy. *Mol. Cell. Biol.* **25,** 5543–5551 (2005).
18. Gabanella, F. *et al.* Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. *PLoS ONE* **2,** e921 (2007).
19. Zhang, Z. *et al.* SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133,** 585–600 (2008).
20. Cifuentes-Diaz, C., Frugier, T. & Melki, J. Spinal muscular atrophy. *Semin. Pediatr. Neurol.* **9,** 145–150 (2002).
21. Iannaccone, S. T., Smith, S. A. & Simard, L. R. Spinal muscular atrophy. *Curr. Neurol. Neurosci. Rep.* **4,** 74–80 (2004).
22. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80,** 155–165 (1995).
23. König, H., Matter, N., Bader, R., Thiele, W. & Muller, F. Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. *Cell* **131,** 718–729 (2007).
24. Matter, N. & Konig, H. Targeted 'knockdown' of spliceosome function in mammalian cells. *Nucleic Acids Res.* **33,** e41 (2005).
25. Kaida, D. *et al.* Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nature Chem. Biol.* **3,** 576–583 (2007).
26. Krämer, A. *et al.* Structure-function analysis of the U2 snRNP-associated splicing factor SF3a. *Biochem. Soc. Trans.* **33,** 439–442 (2005).
27. Friend, K., Lovejoy, A. F. & Steitz, J. A. U2 snRNP binds intronless histone pre-mRNAs to facilitate U7-snRNP-dependent 3′ end formation. *Mol. Cell* **28,** 240–252 (2007).
28. Kyburz, A., Friedlein, A., Langen, H. & Keller, W. Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3′ end processing and splicing. *Mol. Cell* **23,** 195–205 (2006).
29. Millevoi, S. & Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. *Nucleic Acids Res.* **38,** 2757–2774 (2009).
30. Vagner, S., Ruegsegger, U., Gunderson, S. I., Keller, W. & Mattaj, I. W. Position-dependent inhibition of the cleavage step of pre-mRNA 3′-end processing by U1 snRNP. *RNA* **6,** 178–188 (2000).
31. Gunderson, S. I., Polycarpou-Schwarz, M. & Mattaj, I. W. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol. Cell* **1,** 255–264 (1998).
32. Fortes, P. *et al.* Inhibiting expression of specific genes in mammalian cells with 5′ end-mutated U1 small nuclear RNAs targeted to terminal exons of pre-mRNA. *Proc. Natl Acad. Sci. USA* **100,** 8264–8269 (2003).
33. Calvo, O. & Manley, J. L. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* **17,** 1321–1327 (2003).
34. Dreyfuss, G., Kim, V. N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nature Rev. Mol. Cell Biol.* **3,** 195–205 (2002).
35. Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J. & Nogues, G. Multiple links between transcription and splicing. *RNA* **10,** 1489–1498 (2004).
36. Perales, R. & Bentley, D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* **36,** 178–191 (2009).
37. Proudfoot, N. New perspectives on connecting messenger RNA 3′ end formation to transcription. *Curr. Opin. Cell Biol.* **16,** 272–278 (2004).
38. Gromak, N., West, S. & Proudfoot, N. L. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol. Cell. Biol.* **26,** 3986–3996 (2006).
39. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
40. Pellizzoni, L., Kataoka, N., Charroux, B. & Dreyfuss, G. A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* **95,** 615–624 (1998).
41. Kataoka, N. & Dreyfuss, G. A simple whole cell lysate system for *in vitro* splicing reveals a stepwise assembly of the exon-exon junction complex. *J. Biol. Chem.* **279,** 7009–7013 (2004).
42. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15,** 1451–1455 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the members of our laboratory, especially J. Yong and J. Bachorik, for discussions and comments on this manuscript. We thank M. Yoshida for providing spliceostatin A. We also thank D. A. Baldwin and H. Rodriguez at the Microarray Core Facility at the University of Pennsylvania School of Medicine for help with the tiling array. This work was supported by the Association Française Contre les Myopathies (AFM). G.D. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions D.K., M.G.B., I.Y., M.K., L.N.S. and L.W. designed and performed experiments and contributed to data analysis. G.D. is responsible for the project planning and experimental design. All authors contributed to writing the paper.

Author Information The tiling array data have been submitted to the GEO database under the accession number GSE24179. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.D. (gdreyfuss@hhmi.upenn.edu).

METHODS

Cell culture and antisense morpholino oligonucleotide transfection. HeLa cells were cultured as previously described. HeLa cells ($\sim 10^7$ cells per transfection) were trypsinized, washed twice with PBS and re-suspended in 400 µl of DMEM without serum. After mixing cells with morpholino oligonucleotide, they were transferred to a 0.4-cm gap cuvette (Bio-rad). Electroporation was performed using a Bio-Rad Gene pulser at 960 µF and 280 V. After electroporation, cells were cultured for 8 h in 6-well plates with 2 ml DMEM. The sequence of the 25-mer U1 AMO is 5′-GGTATCTCCCCTGCCAGGTAAGTAT-3′, which is complementary to nucleotides 1–25 in human U1 snRNA. The sequence of the 25-mer U2 AMO is 5′-TGATAAGAACAGATACTACACTTGA-3′, which is complementary to nucleotides 27–51 in human U2 snRNA. The 25-mer scrambled sequence control AMO is 5′-CCTCTTACCTCAGTTACAATTTA TA-3′, as previously described[23,24]. AMOs were obtained from Gene Tools, LLC.

RNA preparation and 3′ RACE. Total RNA was extracted from cultured cells using TRIzol (Invitrogen). cDNA was synthesized using SuperScript III reverse transcriptase (Invitrogen) using oligo dT18-XbaKpnBam primer for 3′ RACE according to the manufacturer's directions. 3′ RACE was carried out using the first and second (nested) forward primers and XbaKpnBam reverse primer.

For 3′ RACE of the *NR3C1* mini-gene, pcDNA3.1-5′ primer was used as a forward primer to distinguish mini-gene RNA from endogenous *NR3C1* RNA. PCR products were cut with HindIII to distinguish PCR products of prematurely polyadenylated RNA from mRNA spliced and polyadenylated at the canonical PAS at the 3′ end. Quantification of PCPA and polyadenylation at the canonical PAS at the 3′ end in control and U1-AMO-treated cells (Fig. 5d) was performed using real-time PCR. Primer sequences are listed in Supplementary Table 1.

Tiling array target preparation, hybridization and data analysis. Total RNAs were prepared from HeLa cells transfected with control or U1 AMOs (7.5 µM), or SSA (100 ng ml^{-1}) or methanol for 8 h. Labelled cDNA targets were prepared using the GeneChip WT amplified Double-Stranded cDNA Synthesis Kit and the GeneChip WT Double-Stranded DNA Terminal Labelling Kit (Affymetrix) according to the manufacturer's directions. The end-labelled cDNA targets were applied to GeneChip human tiling 2.0R E arrays (Affymetrix). Hybridization was performed using F450-001 fluidics wash and stain script on the Affymetrix GeneChip Fluidics Station 450. Arrays were scanned using the Affymetrix GCS 3000 7G GeneChip Operating Software (GCOS) to produce .CEL files.

For tiling array analysis, we used .CEL files and the Affymetrix Tiling Analysis Software (TAS) version 1.1 to produce .BED files of the following signal intensity and *P*-value (bandwidth = 50, Min. Run = 100, Max. Gap = 100, fold change ≥2-fold, *P*-value <0.01). Overlapping regions of two files of signal intensity and *P*-value were chosen using Galaxy (http://galaxy.psu.edu/)[42]. We produced .BAR files from the .CEL files using TAS to visualize on the Integrated Genome Browser (IGB) (Affymetrix).

***In vitro* splicing.** [α-^{32}P]UTP-labelled Ad2ΔIVS pre-mRNA was prepared as previously described[40]. *In vitro* splicing reactions were carried out in 293T whole cell splicing extracts prepared as previously described[41]. Increasing amounts of control and U1 AMOs were added to reactions that were incubated for 90 min at 30 °C. Splicing products were purified with TRIzol, resolved on denaturing PAGE, and gels were autoradiographed.

RNase H protection assay and northern blotting. HeLa cells were transfected with AMOs and incubated for 8 h. Cells were harvested and total cell extract was prepared using RSB-100 buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 2.5 mM $MgCl_2$). The cell extract was incubated with 1.5 U of RNase H (Promega) and 5 µM antisense DNA oligonucleotide in a 20-µl reaction for 25 min at 30 °C. Antisense DNA oligonucleotide for U1 snRNA is 5′-CAGGTAAGTAT-3′. After RNase H treatment, RNA samples were purified and analysed by northern blotting with [γ-^{32}P]ATP-labelled U1 snRNA probe, the sequence of which is 5′-CAAATTATGCAGTCGAGTTTCCCACATTTG-3′.

Plasmid construction. To construct the *NR3C1* mini-gene, DNA fragments of *NR3C1* intron 1-exon 2-intron 2 and *NR3C1* intron 2-exon 3 were amplified from genomic DNA from HEK293T cells using NR3C1 int1 for and NR3C1 int2 rev XhoI, and NR3C1 int2 for XhoI and NR3C1 Ex3 rev XhoI. The *NR3C1* intron 1-exon 2-intron 2 fragment was digested with XhoI and EcoRI, and the *NR3C1* intron 2-exon 3 fragment was digested with XhoI. These fragments were subcloned into pcDNA3.1 vector (Invitrogen). The intron 2 of *NR3C1* was truncated from 80 kb to 2 kb to facilitate cloning, mutagenesis and transfection. To introduce mutations, the QuikChange II Site-Directed Mutagenesis Kits (Stratagene) was used with *NR3C1* poly(A) site mutation forward and *NR3C1* poly(A) site mutation reverse, and *NR3C1* 5′ SS mutation forward and *NR3C1* 5′ SS mutation reverse primers. Sequences of all primers are listed in Supplementary Table 1.

***In situ* hybridization.** *In situ* hybridization of U1 snRNA in HeLa cells transfected with control or U1 AMO for 8 h was performed with a biotin-labelled LNA probe (5′-GGTATCTCCCCTGCCAGGTAAGTAT-3′) obtained from Exiqon. The protocol was essentially as described by the manufacturer (Exiqon). Hybridization was performed in 50% formamide, 2× SSC, 50 mM sodium phosphate (pH 7), and 10% dextran sulphate, containing 10 nM LNA probe, at ~ 20 °C below the melting temperature (T_m) for 1 h (50 °C) in a humidified chamber. After hybridization, cells were washed in 2× SSC + 0.1% Triton X-100, followed by detection with a fluorescent Alexa Fluor 594 streptavidin conjugate. Cells were washed 3 × 5 min at 37 °C in 4× SSC + 0.1% Triton X-100, followed by washes in 2× and 1× SSC, with a final wash in PBS at 25 °C. Nuclei were stained by DAPI.