Title: *INSIGHTS INTO THE U1 SMALL NUCLEAR RIBONUCLEOPROTEIN COMPLEX SUPERFAMILY*

Authors *Guiro, J[1] and O'Reilly D[2*]*

Affiliation *[1] Institute of Biosciences, University of Sao Paulo, Rua do Matao, 05508-090 Sao Paulo, SP, Brazil*

[2] Sir William Dunn School of Pathology, South Parks Road. Oxford, OX1 3RE

Correspondence ** O'Reilly D: dawn.oreilly@path.ox.ac.uk*

Key words *U1 snRNA, snRNP, splicing, polyadenylation, stem cells, development*

Word count *5,527 (excluding references and Figure Legends)*

The 164 bp U1 small nuclear (sn) RNA is one of the most abundant non-coding (nc) RNA in human cells, estimated to be in the region of 10^6 copies/cell. Although best known for its role in pre-mRNA splicing events, research over the past 20 years has revealed diverse functions of this ncRNA in mammalian cell types. Excellent reviews exist detailing the role of U1 snRNA in pre-mRNA splicing events. This review highlights what is currently known regarding the additional roles, snRNP composition, expression profiles and the genomic organization of this ncRNA.

THE HETEROGENEOUS U1 SNRNP COMPLEX

*The U1 snRNA primarily exerts its function in the form of a ribonucleoprotein complex consisting of three U1-specific proteins, U1-A, U1-70K and U1-C, and 8 Sm proteins Sm-B/B', -D1,- D2, -D3, -E, -F, which are shared with other U-rich containing snRNAs (U2, U4/46 and U5)(1) (**Figure 1**). Each Sm protein binds a single base of the highly conserved Sm motif (AUUUGUG in the case of the U1 snRNA) in the order Sm-E, G, D3, B, D1, D2, and Sm-F, forming a stable ring structure around the motif (2,3). Assembly of the Sm proteins occurs in the cytoplasm and is catalyzed by the Survival of Motor Neuron (SMN) complex (4-6). U1-70K and U1-A both contain RNA binding domains (RBDs) and bind the stem loop I (SLI) and stem loop II (SLII) structure of the U1 snRNA, respectively, in the nucleus (2). The zinc-finger-containing U1-C protein does not bind the U1 snRNA directly and is recruited through protein:protein interactions with U1-70K and Sm-D3 (2,7).*

Structural analysis of the native U1 snRNP, purified from HeLa cells, indicates that at least 4 major isoforms exist (8). The Sm-B and Sm-B'-containing U1 snRNPs are present in equal abundance. Sm-B' is a spliced variant of Sm-B, which differs by 9 residues, introducing an additional proline-rich glycine/arginine/glycine (GRG) repeat at the end of its C-terminal domain (9). These repeats are common to other Sm proteins, including Sm-D1 and –D3, and are typically dimethylated in the cytoplasm by methyltransferase enzymes, including PRMT4/CARM1 and PRMT5 (10-14). Both the native Sm-B- and -B'- containing U1 snRNP complexes are fully dimethylated in vivo and thought to play important roles in U1 snRNP assembly in the cytoplasm. In particular, dimethylation of SmB/B' enhances protein:protein interactions with the Tudor domain of SMN1 (15,16). Interestingly, Sm-B' is expressed in all tissues analyzed with the exception of brain. The functional consequence of this is currently not clear. Since many splicing regulators contain RG motifs which are important for protein:protein interactions, the lack of such a repeat in the Sm-B-containing U1 snRNP complex found in the brain likely plays an important role in alternative splicing events by abrogating/enhancing U1 snRNP activity with brain-specific splicing modulators.

Two different isoforms of the U1-70K are also known to associate with the U1 snRNA in human cells (8,17). Isoform 1 contains a nine amino acid insertion containing a serine residue (Ser 226), which is known to be phosphorylated in vivo. 70% of the native U1 snRNP complex consists of this isoform. U1-70K is the only reported U1-associated protein which is phosphorylated and multiple variants exist in human cells as a consequence of differential phosphorylation at 6 sites, 5 of which are located within the C-terminal arginine/serine (RS) domain and 1 in the N-terminal RNA binding domain (RBD) (18,19). Although little is known about the functional consequences of particular phosphorylated residues or the specific enzymes involved in phosphorylation/dephosphorylation, this post-translational modification is important for regulating the splicing activity of U1 snRNP specifically. Studies have shown that purified

thiophosphorylated U1 snRNP complexes, which contain a phosphatase-resistant hyperphosphorylated form of U1-70K, are unable to complement splicing activity of U1-depleted nuclear extracts, whereas purified U1 snRNP complexes can fully restore activity in vitro. Although the molecular basis of this is unclear, hyperphosphorylation of U1-70K does not seem to interfere with the association of the mature spliceosome but specifically blocks activation of the first step of splicing. Moreover, enhanced phosphorylation of a serine residue in the RBD (Ser 140), which occurs during apoptosis, results in U1-70K relocation from sites of active splicing (perichromatin fibrils) to sites of storage (Interchromatin granules) (20,21). Additionally, phosphorylation on Serine 226 (isoform 1-specific) enhances U1-70K interaction with U1-C, whereas Isoform 2 interacts more strongly with Sm-B/B'. Since U1-C participates in 5'ss selection by stabilizing the base pairing between the 5' end of the U1 snRNA and the 5' splice-site region, altering the affinity of U1-C for the U1 snRNA could affect 5' splice site choice (2,22-26). Consequently, the existence of U1-70K isoforms and the numerous variants associated with their differential phosphorylation could serve as a platform for integrating differential protein:protein interactions, which would influence both U1 snRNP activity and /or location underpinning alternative splicing events in different cell types and/or during development.

In addition to these four major isoforms of U1 snRNP complexes, several of the U1 containing proteins are subject to additional post-translational modifications, including methylation of the zinc-finger domain of U1-C by CARM1 and dimethylation of the unstructured C-terminal tails of Sm-D1 (9 sites) and –D3 (4 sites) by PRMT5 (11,16,27). Moreover, the U1 snRNA is also post-transcriptionally edited. Two uridines at position +5 and +6 (+1 being the first base of the U1 snRNA) are converted to pseudouridines (28). The additional imino group in pseudouridine not only confers rigidity to the local RNA structure by promoting local base stacking but also contributes to base-pairing interactions (29,30). In support of this, U2 snRNA lacking pseudouridine in the region required for interaction with the pre-mRNA branch site, is unable to reconstitute pre-mRNA splicing activity following injection into U2 depleted Xenopus oocytes, whereas control U2 isolated from untreated cells completely reconstituted the splicing (31,32). Furthermore, four adenosine bases are methylated at positions +1, +2, +65 and +70. Although the biological significance of these modifications has not been determined, they are thought to enhance association of U1-specific proteins to the U1 snRNA. In particular, methylation of the backbone ribose of adenosines A65 and A70 facilitates contact with the U1-A protein (33). This suggest that in some tissues at least the control of U1 snRNA function could be modulated by post-transcription editing events, which could influence specific protein:U1 RNP interactions.

U1 AND VARIANT (V)U1 SNRNA GENES.

The heterogeneity of the native U1 snRNP is not only restricted to variations in the U1 associated proteins but also extends to the associated U1 snRNAs. RNA species migrating in polyacrylamide gels with the U1 snRNA were first observed in RNA extracts purified from rat hematoma (34) and HeLa cells (35). However, the first vU1 gene, referred to as U1b, was cloned and sequenced from mouse cells and differs from U1 snRNA by seven base changes, including a single base insertion (36). All base changes occur within SLII and disrupt U1-A association (7). Unlike U1, which is constitutively expressed, U1b is restricted to mouse fetal tissues, adult tissues that contain a significant number of undifferentiated stem cells and transformed cells,

including teratocarcinoma, lymphoma, and Friend cells; accounting for almost 40% of total U1 snRNA in these cell types (37-39).

The mechanism by which U1 and U1b snRNAs differentially accumulate in various cell types is not well understood but is thought to be controlled at the level of transcription rather than RNA stability. In support of this, U1b snRNA is stable in fully differentiated cells, whereas the U1b promoter is largely inactive (40). Distal sequences (>2 kb from the transcription start site) and sequences within the immediate 5' flanking region, but not including the PTF/SNAPc (PSE-binding transcription factor, snRNA-activating protein complex) binding site, are important for developmental regulation (41). This region is not well conserved with the U1 snRNA gene, supporting a model whereby developmental control of U1b expression involves either sequence-specific recruitment of a repressor molecule(s) to down-regulate expression during differentiation or the association of specific enhancer protein(s) to promote expression in stem cells. However, it remains unclear what these factors might be or even, for that matter, the biological significance of a tissue-restricted vU1 snRNA. Since U1b snRNA contains the same 5'end as U1, it is impossible to predict potential pre-mRNA targets for regulation through nucleotide sequence complementarity. Moreover, as the numbers of base changes are so few, specific knockdown experiments to assess its global function(s) are also difficult. Renewed interest in these vU1 snRNAs was sparked again when developmentally-regulated vU1 snRNAs were purified and sequenced from RNP complexes across a number of different species, including frog (42), fly (43), moths (44) and sea urchin (45,46). Like mouse, these vU1 snRNA genes retain good conservation of known sequences elements required for basal (proximal sequence element (PSE) (PTF/SNAPc binding site)) and activated (distal sequence element (DSE) (Oct1 and Sp1 binding site) expression. Furthermore, sequences flanking these elements, although highly conserved amongst the vU1 snRNA genes of a particular species, are less well conserved compared to their respective U1 snRNA, reinforcing the model that tissue-specific factors are likely key regulatory components modulating the differential expression of vU1 snRNA genes in the different species (43,47,48). Unfortunately, the similarly high degree of conservation between these vU1 snRNAs and their corresponding species-specific U1 snRNA also limited the potential to investigate their biological function(s) further.

In contrast, the discovery of vU1 snRNAs expressed in human cancer and embryonic stem cells, which contain numerous base changes in their ncRNA regions, including the 5' ends, provided the first real opportunity to address the biological significance of this class of human vU1 snRNAs (49-52). The human genome encodes approximately 141 vU1 snRNA genes (Feb 2009 hg19 build) and many are dispersed throughout the human genome (http://www.genenames.org/rna/snRNA) (53-55). Several of the dispersed vU1 genes have no homology in their flanking regions to each other or to the U1 snRNA genes and are typically flanked by repeat elements or stretches of adenines or thymines (56-58). While the lack of any recognizable promoter sequences and 3' end processing elements would suggest that such U1 copies are transcriptionally inactive, vU1 snRNAs matching these genes copies have previously been detected in human cells (51,59). Although no known function(s) has been assigned to these vU1 copies, it does suggest, at least, that U1 snRNA expression can be controlled by mechanisms other than that which has been previously described for the RNU1.1-4 U1 genes (60). The remaining vU1 genes are encoded within a multigene locus on Chromosome 1q21-22 (54). These vU1 snRNA genes have good conservation

*of promoter, U1 snRNA-encoding and 3' flanking sequences and recruit snRNA-specific factors, including PTF/SNAPc and RNA Pol II phosphorylated on Serine 7 of the Carboxyl Terminal Domain (CTD) heptapeptide of the large subunit, for active transcription (52,61,62). The majority of these genes encode vU1 snRNAs that deviate in sequence from U1 snRNA by numerous bases changes and/or small deletions (52) (**Figure 1**). Like the vU1 snRNAs previously described in other species, this group of human vU1 snRNAs are primarily expressed in human stem and carcinoma cells, contain a trimethyl cap structure at the 5' end and are packaged into Sm-containing complexes. With regards the core proteins, including U1-C, U1-A and U1-70K, only vU1 snRNAs with the highest identity to U1 are capable of associating with all of these proteins in vivo in general (52,63). Gel shift assays using recombinant core proteins with in vitro transcribed U1/vU1 snRNA demonstrate that most of the vU1 snRNAs analyzed display a range of binding capabilities, with varying affinities, for the U1-specific core proteins when compared to the U1 snRNA. Since all vU1 snRNAs analyzed form high molecular complexes in vitro, this data suggests that several vU1 snRNAs are packaged into novel RNP complexes. These uncharacterized RNPs may have additional roles in regulating vU1 snRNA activity by directing precise recruitment to target pre-mRNAs via specific protein:protein interactions with RNA processing regulators. Moreover, these unique proteins may also impact a completely novel function on the vU1 snRNP. While future work will be required to determine the contribution of each vU1 snRNA, an intriguing possibility is that the modulation of vU1 snRNA expression profiles in different cell types and/or during development profoundly impacts on the proteome output of each cell type.*

DIFFERENT FLAVOURS TO U1 SNRNP ACTIVITY

U1 snRNA was first proposed to participate in mRNA biogenesis in the 1970s following the observation that sequences within the 5' end of the U1 snRNA were complementary to exon/intron junction sequences located at the 5' end of introns (64-66). Subsequently, using U1 snRNP-specific antibodies, it was shown that the U1 snRNP could specifically associate to the 5'ss in vitro (1,67) and in the mid-1980s it was firmly demonstrated that the U1 snRNA forms RNA:RNA base pair interactions with the 5'ss in vivo (68). While 99% of all splice sites contain a GU at the first two intronic positions, which are complementary to A7 and C8 in U1 (1 being the first base of the U1 snRNA), only 0.85% of splice sites are complementary to the first nine bases of the U1 snRNA (69). This increases to 5% considering that U1 snRNA binds 5' ss using alternative base pairing registers, including shifting base pairing interactions by one or more nucleotides or allowing nucleotides to bulge out of the 5' ss-U1 snRNA duplex to increase complementarity (70,71). These data reflect the tremendous flexibility and redundancy which is allowed between the 5' ss-U1 snRNA base-pairing to achieve efficient splicing at 5' ss. In addition, splicing regulators, including the serine arginine (SR)-containing and hnRNP proteins, play a major role in regulating U1 snRNP recruitment enabling seemingly weak 5' ss to be efficiently recognised and processed by the spliceosome machinery and much effort over the past 20 years has focused on understanding the mechanism of action of these SR and hnRNP enhancers/silencers in modulating splice site choice (72-76). Nonetheless, there is growing evidence that splicing can occur independently of U1 snRNP/snRNA, implicating other mechanisms regulating alternative splicing in mammalian cells.

Immunodepletion of HeLa nuclear extracts of U1 snRNP or blocking its activity with complementary oligonucleotides abolishes splicing activity in vitro. However this activity is restored with increasing concentrations of SR proteins (77,78). These were among the first reports raising questions regarding the requirement for U1 snRNP activity in 5' ss recognition and spliceosome assembly. Subsequent reports

provided more evidence of naturally occurring U1-independent splicing mechanisms. Neurofibromatosis type I is a multisystem disorder with complete penetrance by the age of 5. One of the most common mutations associated with the disease resides in the 5' ss of exon 29, which causes aberrant splicing at that exon. Attempts to rescue this mutation, however, using a minigene approach with suppressor U1 snRNAs proved unsuccessful (79). Another study showed that U1 snRNA was absent from the E complex formed on intron 9 of the human ATP synthase γ subunit (hIFγ) pre-mRNA in spliceosome assembly assays, whereas U2 and U2AF levels were normal (80). Moreover, exon 9-10 of hIFγ was efficiently spliced in U1 snRNA inactivated HeLa nuclear extracts and Xenopus oocytes, re-enforcing the notion that hIFγ exon 9 is a natural substrate for U1 independent splicing and that U1 independent splicing mechanisms are conserved in other species.

The mechanism(s) of naturally-occurring U1-independent splicing is currently unclear but some reports have highlighted the importance of additional factors in circumventing the requirement for U1 snRNP in early spliceosome assembly. The U1-C protein, for example, is known to contact the 5' ss directly in the absence of U1 snRNP base pairing (2,26,81). In support of this, an in vitro SELEX assay using normal or 5' end cleaved U1 snRNA-containing extracts demonstrated little differences in the functional 5'ss sequences selected, drawing into question the requirement for U1 snRNA base-pairing (82). Additional evidence is also available implicating a role for U6 snRNA, which replaces U1 prior to the first transesterification reaction, as the efficiency of splicing in SR enriched extracts is affected by the level of complementarity to U6 snRNA (83,84). However, there is growing evidence that alternative mechanisms of spliceosome assembly exist since U1 snRNA independent 5' ss are processed as efficiently as U1-dependent 5' ss, in the absence of U1 snRNA. Likely candidates are the human vU1 snRNAs, which have previously been shown to contribute to mRNA 3' end processing (52). Many vU1 snRNAs differ in their respective 5' ends and could participate in atypical 5' ss selection. In support of this, cryptic 5' ss, rather than canonical 5' ss, were often selected in the SR-enriched U1 snRNA-depleted splicing events and the efficiency of splicing varies with different mRNA substrates. Moreover, some vU1 snRNAs may have evolved to participate in splicing regulation in the absence of RNA:RNA base pairing. As discussed above, the majority of vU1 snRNAs are likely packaged into novel RNP complexes since they lack conservation to U1 snRNA in regions known to be required for core-protein binding. Consequently, splicing regulators could play important roles in specifically recruiting vU1 snRNPs to particular target 5' ss via interactions with novel vU1- associated factors. In agreement with this, sequences flanking the 5' ss and downstream of the branch site are known to contribute to U1-independent splicing mechanisms (85). Following vU1 snRNP recruitment, splice site choice is subsequently mediated through direct binding of the U1-C protein, for example, which is known to be associated with several vU1 snRNP complexes (52,63). In agreement with this, no psoralen cross-linking of RNA to the 5' ss(s) were detected in the SR-enriched splicing extracts pre-treated with 2' O-methyl oligonucleotides targeting the 5' end of the U1 snRNA, re-enforcing the notion that vU1 snRNPs may use alternative mechanisms in regulating mRNA processing. Furthermore, since many vU1 snRNAs are differentially expressed, they could contribute to tissue-specific, developmentally-regulated alternative splicing events. In support of this, a U1 independent isoform of hIFγ is known to exist in many different tissues, with the exception of brain, heart and skeleton muscle (80). Further research is needed to fully ascertain the requirements for U1 independence, the likely role for vU1 snRNAs in U1 bypass reactions and the prevalence of such mechanism in mammalian cells.

THE MULTIFACETED U1/vU1 SNRNP COMPLEX

a) Role in 3' end processing and directionality of transcription

*U1 snRNP is classically known for its role in pre-mRNA splicing events, playing an integral part in splice site selection and spliceosome assembly by base-pairing to the 5' ss (**Figure 3**). However, the finding that U1 snRNA levels far exceed other spliceosomal associated snRNA levels led to the notion that U1 snRNA may have additional roles in the cell apart from splicing regulation. In support of this, recent evidence has highlighted a global role for U1 snRNP in regulating transcript length and directionality of transcription via competition with components of the polyadenylation machinery (**Figure 2**)(86,87). Using morpholino oligonucleotides to block the base pairing activity of U1 snRNA followed by tiling array analysis to quantitate changes in transcript levels and composition, researchers observed a sharp reduction in mRNA 3' ends instead of the expected stabilization of intronic RNA, which would be indicative of a splicing defect (87). Subsequent cloning and sequencing of these short stable mRNA products revealed the presence of a poly A (pA) tail added downstream of a bona fida pA site. The requirement for both the U1 snRNA and the pA site was corroborated with mini-gene experiments demonstrating the need for a functional 5' ss and AAUAAA hexamer. Importantly, this effect was not observed when splicing was disrupted by pharmacologically abrogating U2 snRNP's activity, supporting a splicing-independent role for U1 snRNP in the suppression of internal poly A site usage. This additional function of U1 snRNP was further substantiated by other studies demonstrating a relationship between the amount of available U1 snRNA in the cell and the corresponding length of the mRNA transcripts generated. For example, complete inhibition of U1 snRNA activity resulted in the production of very short stable polyadenylated mRNAs whereas lowering the concentration of the U1-specific morpholino, and thereby increasing the amount of active U1 snRNA available, caused a progressive lengthening of the mRNAs synthesized (88). Furthermore, reactivation of internal pA site suppression was achieved using decoy RNA molecules designed to impair U1 snRNA base-pairing with 5' ss sequences upstream of the target pA site (89).*

*Interestingly, bioinformatics analysis of regions flanking transcription start sites indicated a bias in the positioning of U1 snRNA binding sites relative to transcription start sites. The number of U1 snRNA binding sites are typically reduced in the antisense direction with an asymmetric enrichment in the sense direction (86) (**Figure 2**). It has been known for some time that RNA Pol II transcription is bidirectional at promoters (90,91). Transcription in the antisense direction is quickly terminated giving rise to short products that are rapidly turned over by the exosome (92). In contrast, transcription in the sense direction proceeds productively. The authors proposed that specific recruitment of U1 snRNP in the sense direction ensures suppression of the pA sites enabling transcription to proceed productively. This work was further corroborated in another study demonstrating that the short antisense transcripts were indeed prematurely terminated at functional pA sites (93).*

b) Requirements for polyadenylation control

The involvement of a component(s) of the splicing machinery in the regulation of pA site usage was evident from an early study demonstrating that pA site usage was position-dependent. When a strong synthetic pA site was placed downstream of a terminal pA site it was exclusively used but when placed within an intron it was silenced completely. Importantly, reactivation of the synthetic pA site usage was observed following mutation of the proximal 5' ss (94). The direct involvement of U1 snRNA in polyadenylation

control, and specifically its suppression, was later demonstrated in studies investigating viral gene expression. Work on human and bovine papillamona virus late gene expression identified an inhibitory element within the 3' UTR, upstream of the pA site, which contained 5' ss–like sequence motifs complementary to the 5' end of U1 snRNA. Mutations of these motifs abrogated the inhibitory effect, which was restored upon co-expression of suppressor U1 snRNAs containing complimentary base changes in the 5' end (95,96). These data were further supported by other work investigating pA site selection in the long terminal repeat (LTR) regions that flank the HIV-1 proviral genome. Both LTRs contain identical pA sites but the use of the proximal pA site needs to be restricted to enable efficient viral gene expression. The critical determinant of this suppression also relied on a 5' ss sequence located downstream of the pA site. As expected, mutation of this motif relieved the suppression and co-expression of mutant U1 snRNAs engineered to bind the mutant 5' ss re-instated the inhibitory effect on the pA site (97,98). Interestingly, the recruitment of U1 snRNP downstream of the pA site also contributed to interference with mRNA cleavage, enabling transcription and proper processing of the HIV-1 downstream genes and transcripts, respectively (99). These data highlight the existence of distinct mechanisms of U1 snRNP-mediated cleavage and polyadenylation control with the position of the U1 snRNA binding site relative to the pA site being a major determinant.

The range of U1 snRNP's inhibitory affect seems to be limited to within 1 kb of the pA site and since 5' ss can be up to 100's of kb apart, it suggests that the abundant pseudo 5' ss are functionally important in mediating U1 snRNP's role in pA silencing (99). This finding has far-reaching implications for the interpretation of disease associated mutations as base changes at pseudo 5' ss are not normally predicted to cause disease. However, it is now evident that such mutations could lead to the activation of a downstream pA site with the consequent generation of a shortened protein with potentially pathogenic properties if expressed in the wrong cell or at an inappropriate stage of development.

pA suppression is thought to be orchestrated through direct binding of the U1-70K protein with the poly A polymerase (PAP) (100). U1-70K contains four PAP regulatory domains (PRDs) within its SR domain, which contact the PAP enzyme and physically interferes with its ability to add a pA tail to the 3' end of the message. In support of this, knockdown of U1-70K leads to a general activation of internal sites and tethering a MS2/U1-70K fusion protein to a modified U1 snRNP in which SLI has been replaced by an MS2-binding loop is sufficient to mediate pA suppression (89,101). Although the mechanism by which U1 snRNP contributes to mRNA cleavage inhibition is currently not known, it is possible that U1-70K is also involved as it has been reported to interact with components of the mammalian cleavage factor 1 complex (CF1); in particular, the 25 kDa subunit (102). Consequently, association of U1-70K with the CF1 complex may also impede its binding and prevent cleavage of the pre-mRNA at the upstream pA site.

Interesting, the developmentally regulated human vU1 snRNAs are also known to be involved in pA site suppression in spite of the fact that many are packaged into snRNP complexes that lack the U1-70K protein (52,63). Recent evidence has shown that splicing modulators, including SRP20, SRP75 and U2AF65, contain SR domains that closely match the U1-70K consensus and two of these factors, SRP20 and U2AF65, can interact with and modulate the polyadenylation machinery (103,104). This data suggests that pA site usage may be regulated by alternate mechanisms in vivo. Since vU1 snRNAs are thought to form novel RNP complexes they could interact directly or indirectly with such factors to facilitate their role in pA

suppression. Further analysis is required to determine the specific role of U2AF65 and the additional SR factors in pA suppression and how individual vU1 snRNPs target specific subsets of pre-mRNAs for pA control in vivo.

Although it is still unclear how suppression of pA site usage by U1 snRNP is alleviated at the terminal pA sites, it is apparent that U1 snRNP plays a major non-splicing role in regulating the output from our genes by safeguarding the integrity of all transcripts and ensuring the correct product is generated by RNA pol II transcribing in the correct direction. The physiological role of this property of U1 snRNP and its exploitation in the treatment of human diseases and cancers is discussed in detail in a recent review (105)

Role in transcription initiation

*In addition to U1 snRNP's contribution to pre-mRNA splicing and 3' end regulation, there is also evidence indicating a possible role in transcription initiation, which is independent of its RNA processing roles (**Figure 3**). The presence of intronic sequences, and in particular the requirement for the 5' ss, have long been known to have a stimulatory effect on the amount of mRNA produced from transgenes either transiently or stably expressed in different cell types (106,107). The role for U1 snRNA in enhancing this gene output was later confirmed in experiments using mutant HIV minigene constructs and suppressor U1 snRNAs (108). Moreover, mutations within the 5' ss or altering its distance from the transcription start site affect the association of Pol II and recruitment of general transcription factors, including TFIIB and TFIIH (109). The link between U1 snRNA and transcription was later supported by the finding that the U1 snRNA specifically co-purifies with the general transcription factor TFIIH (110). TFIIH contains a cyclin dependent kinase (CDK)-activating kinase (CAK) complex, which phosphorylates the RNA Pol II large subunit carboxyl-terminal domain (CTD) at the early stages of transcription to promote transcription initiation (111-115). This kinase activity of TFIIH consists of three subunits CDK7, cyclin H and an assembly factor MAT-1. SLII of the U1 snRNA interacts with the cyclin H component and the addition of cyclin H peptides, which specifically disrupts this interaction, dramatically impairs the kinase activity of an immunoaffinity purified CAK complex when assessed against a CTD substrate (116). Although the molecular basis for this interaction is unclear, this work highlights an important role for U1 snRNA in the early events of transcription initiation.*

In addition to TFIIH, U1 snRNA is also known to associate with other factors implicated in transcription initiation, including TAF-15 ($TAF_{II}68$). TAF15 is a putative RNA/ssDNA binding protein and seems to form two separate complexes with U1, a U1 snRNP-containing and U1-snRNA-containing complex, which are distinct from the TAF-15-containing TFIID complex (117-119). However, the functional significance to this interaction has currently not been determined.

Further analysis is necessary to establish the specific contribution(s) U1 snRNA makes to transcription initiation and how regulation of U1 snRNA association could affect the expression of different genes in different cell types, for example.

CONSERVATION OF VU1 SNRNA GENES.

Two large scale comparative studies analysing the evolution of spliceosomal proteins and splicing signals revealed a selective expansion of SR protein gene families in metazoans and hnRNP gene families in

vertebrates alongside an increased degeneracy at the 5'ss and 3'ss (120,121). Although the expansion of splicing regulatory factors correlates well with alternative splicing being widespread in multicellular organisms, there appears to be little correlation between degeneracy at the 5'ss and evolution of the U1 snRNA genes (121). The bases located at the 5' end of U1 snRNA are very well conserved across all species analysed, supporting a role for splicing regulators as key modulators of splicing activity. However, in both of these studies only the highly expressing U1 snRNA genes were considered as vU1 snRNA genes were thought to represent non-functional U1 snRNA pseudogenes. In light of what we now know regarding the differential expression of vU1 snRNA genes in different species, it may be worthwhile re-considering the potential contribution U1/vU1 snRNA gene products make to the evolution of alternative splicing in higher organisms. On the basis of RNA structure alone, secondary structure predictions indicate that despite the numerous base changes, all human vU1 snRNAs identified can adopt a typical clover-leaf structure (2,122). Wherever a base change has occurred, there appears to have been some evolutionary pressure to introduce a compensatory change at the corresponding position to maintain this structure (52). This is inconsistent with the idea that vU1 snRNA genes have evolved neutrally and supports the idea that evolutionary pressure maintains and diversifies vU1 snRNA genes in many different species. In fact, a simple comparative analysis of multiple genomes, using the UCSC web browser, indicates the existence of numerous vU1 snRNA gene sequences in different species including monkeys, mouse, fly, plants and protozoa. As expected, the Saccharomyces cerevisiae yeast genome does not contain any vU1 snRNA gene copies, which is consistent with the fact that 5' ss within the 3% of intron-containing genes are highly conserved and complementary to the 5' end of the yeast U1 snRNA. While there appears to be good conservation of the human vU1 snRNA genes between monkey and human, there is clear evidence of interspecies variations. Taken together, this analysis suggests that vU1 snRNA genes, like SR and hnRNP gene families, have recently appeared during evolution, which correlates with a similar increase in 5' ss degeneracy. This expansion of vU1 snRNA genes is consistent with the view that the concurrent increase in the abundance of alternatively spliced mRNA isoforms is particularly associated with multicellular organisms.

Considering the functional repertoire of the U1 snRNA, this non-coding RNA is in no doubt a truly fascinating molecule. Consequently, the existence of a diverse population of vU1 snRNAs, which are differentially-expressed, developmentally-regulated and evolutionary-conserved, shouldn't be ignored. Low abundant vU1 snRNAs expressed in different tissues and at different stages of development could contribute to the fidelity of mRNA processing and underpin regulated mRNA processing events in many different cells types and in response to different environment stimuli. De-regulated expression of variant U snRNAs has previously been reported to cause disease phenotypes in both mouse and humans (123-126). A five nucleotide deletion within mouse U2 snRNA genes causes ataxia and neurodegeneration and four mutations in the human $U4_{atac}$ gene is associated with microcephalic osteodysplastic primordial dwarfism type 1. Moreover, deletion of the chromosome region 1q12-21, encompassing all of the human vU1 snRNA genes, causes severe neurological dysfunction reinforcing the notion that vU1 snRNAs may exert tissue-specific effects and be important for normal development.

REFERENCES

1. Lerner, M.R. and Steitz, J.A. (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A*, **76**, 5495-5499.

2. Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J. and Nagai, K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature*, **458**, 475-480.
3. Weber, G., Trowitzsch, S., Kastner, B., Luhrmann, R. and Wahl, M.C. (2010) Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J*, **29**, 4172-4184.
4. Chari, A., Golas, M.M., Klingenhager, M., Neuenkirchen, N., Sander, B., Englbrecht, C., Sickmann, A., Stark, H. and Fischer, U. (2008) An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell*, **135**, 497-509.
5. Fischer, U., Liu, Q. and Dreyfuss, G. (1997) The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell*, **90**, 1023-1029.
6. Liu, Q., Fischer, U., Wang, F. and Dreyfuss, G. (1997) 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.
7. Bach, M., Krol, A. and Luhrmann, R. (1990) Structure-probing of U1 snRNPs gradually depleted of the U1-specific proteins A, C and 70k. Evidence that A interacts differentially with developmentally regulated mouse U1 snRNA variants. *Nucleic Acids Res*, **18**, 449-457.
8. Hernandez, H., Makarova, O.V., Makarov, E.M., Morgner, N., Muto, Y., Krummel, D.P. and Robinson, C.V. (2009) Isoforms of U1-70k control subunit dynamics in the human spliceosomal U1 snRNP. *PLoS One*, **4**, e7202.
9. van Dam, A., Winkel, I., Zijlstra-Baalbergen, J., Smeenk, R. and Cuypers, H.T. (1989) Cloned human snRNP proteins B and B' differ only in their carboxy-terminal part. *EMBO J*, **8**, 3853-3860.
10. Friesen, W.J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G.S., Van Duyne, G., Rappsilber, J., Mann, M. and Dreyfuss, G. (2001) The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol Cell Biol*, **21**, 8289-8300.
11. Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L. and Luhrmann, R. (2000) The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *J Biol Chem*, **275**, 17122-17129.
12. Friesen, W.J., Wyce, A., Paushkin, S., Abel, L., Rappsilber, J., Mann, M. and Dreyfuss, G. (2002) A novel WD repeat protein component of the methylosome binds Sm proteins. *J Biol Chem*, **277**, 8243-8247.
13. Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C. and Fischer, U. (2001) Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr Biol*, **11**, 1990-1994.
14. Gonsalvez, G.B., Tian, L., Ospina, J.K., Boisvert, F.M., Lamond, A.I. and Matera, A.G. (2007) Two distinct arginine methyltransferases are required for biogenesis of Sm-class ribonucleoproteins. *J Cell Biol*, **178**, 733-740.
15. Brahms, H., Meheus, L., de Brabandere, V., Fischer, U. and Luhrmann, R. (2001) Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA*, **7**, 1531-1542.
16. Cheng, D., Cote, J., Shaaban, S. and Bedford, M.T. (2007) The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol Cell*, **25**, 71-83.
17. Spritz, R.A., Strunk, K., Surowy, C.S., Hoch, S.O., Barton, D.E. and Francke, U. (1987) The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding. *Nucleic Acids Res*, **15**, 10373-10391.
18. Tazi, J., Kornstadt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C. and Luhrmann, R. (1993) Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature*, **363**, 283-286.

19. Woppmann, A., Patschinsky, T., Bringmann, P., Godt, F. and Luhrmann, R. (1990) Characterisation of human and murine snRNP proteins by two-dimensional gel electrophoresis and phosphopeptide analysis of U1-specific 70K protein variants. *Nucleic Acids Res*, **18**, 4427-4438.
20. Monneaux, F., Hoebeke, J., Sordet, C., Nonn, C., Briand, J.P., Maillere, B., Sibillia, J. and Muller, S. (2005) Selective modulation of CD4+ T cells from lupus patients by a promiscuous, protective peptide analog. *J Immunol*, **175**, 5839-5847.
21. Dieker, J., Cisterna, B., Monneaux, F., Decossas, M., van der Vlag, J., Biggiogera, M. and Muller, S. (2008) Apoptosis-linked changes in the phosphorylation status and subcellular localization of the spliceosomal autoantigen U1-70K. *Cell Death Differ*, **15**, 793-804.
22. Heinrichs, V., Bach, M., Winkelmann, G. and Luhrmann, R. (1990) U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science*, **247**, 69-72.
23. Will, C.L., Rumpler, S., Klein Gunnewiek, J., van Venrooij, W.J. and Luhrmann, R. (1996) In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res*, **24**, 4614-4623.
24. Chen, J.Y., Stands, L., Staley, J.P., Jackups, R.R., Jr., Latus, L.J. and Chang, T.H. (2001) Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol Cell*, **7**, 227-232.
25. Stark, H., Dube, P., Luhrmann, R. and Kastner, B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, **409**, 539-542.
26. Du, H. and Rosbash, M. (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature*, **419**, 86-90.
27. Miranda, T.B., Khusial, P., Cook, J.R., Lee, J.H., Gunderson, S.I., Pestka, S., Zieve, G.W. and Clarke, S. (2004) Spliceosome Sm proteins D1, D3, and B/B' are asymmetrically dimethylated at arginine residues in the nucleus. *Biochem Biophys Res Commun*, **323**, 382-387.
28. Reddy, R., Henning, D. and Busch, H. (1981) Pseudouridine residues in the 5'-terminus of uridine-rich nuclear RNA I (U1 RNA). *Biochem Biophys Res Commun*, **98**, 1076-1083.
29. Yu, A.T., Ge, J. and Yu, Y.T. (2011) Pseudouridines in spliceosomal snRNAs. *Protein Cell*, **2**, 712-725.
30. Wu, G., Yu, A.T., Kantartzis, A. and Yu, Y.T. (2011) Functions and mechanisms of spliceosomal small nuclear RNA pseudouridylation. *Wiley Interdiscip Rev RNA*, **2**, 571-581.
31. Zhao, X. and Yu, Y.T. (2007) Incorporation of 5-fluorouracil into U2 snRNA blocks pseudouridylation and pre-mRNA splicing in vivo. *Nucleic Acids Res*, **35**, 550-558.
32. Zhao, X. and Yu, Y.T. (2004) Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in Xenopus oocytes. *RNA*, **10**, 681-690.
33. McConnell, T.S., Lokken, R.P. and Steitz, J.A. (2003) Assembly of the U1 snRNP involves interactions with the backbone of the terminal stem of U1 snRNA. *RNA*, **9**, 193-201.
34. Busch, H. (1974) *The Cell Nucleus*. New York: Academic Press.
35. Zieve, G. and Penman, S. (1976) Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell*, **8**, 19-31.
36. Marzluff, W.F., Brown, D.T., Lobo, S. and Wang, S.S. (1983) Isolation and characterization of two linked mouse U1b small nuclear RNA genes. *Nucleic Acids Res*, **11**, 6255-6270.
37. Lund, E., Kahan, B. and Dahlberg, J.E. (1985) Differential control of U1 small nuclear RNA expression during mouse development. *Science*, **229**, 1271-1274.
38. Kato, N. and Harada, F. (1985) New U1 RNA species found in Friend SFFV (spleen focus forming virus)-transformed mouse cells. *J Biol Chem*, **260**, 7775-7782.

39. Matsuda, M., Nomura, T. and Kameyama, T. (1986) Isolation of U1-snRNP(s) from mouse teratocarcinoma cells using immunochemical and biochemical procedures: high proportion of U1a-snRNP to U1b-snRNP. *J Biochem*, **99**, 895-900.
40. Caceres, J.F., McKenzie, D., Thimmapaya, R., Lund, E. and Dahlberg, J.E. (1992) Control of mouse U1a and U1b snRNA gene expression by differential transcription. *Nucleic Acids Res*, **20**, 4247-4254.
41. Cheng, Y., Lund, E., Kahan, B.W. and Dahlberg, J.E. (1997) Control of mouse U1 snRNA gene expression during in vitro differentiation of mouse embryonic stem cells. *Nucleic Acids Res*, **25**, 2197-2204.
42. Forbes, D.J., Kirschner, M.W., Caput, D., Dahlberg, J.E. and Lund, E. (1984) Differential expression of multiple U1 small nuclear RNAs in oocytes and embryos of Xenopus laevis. *Cell*, **38**, 681-689.
43. Lo, P.C. and Mount, S.M. (1990) Drosophila melanogaster genes for U1 snRNA variants and their expression during development. *Nucleic Acids Res*, **18**, 6971-6979.
44. Sierra-Montes, J.M., Pereira-Simon, S., Smail, S.S. and Herrera, R.J. (2005) The silk moth Bombyx mori U1 and U2 snRNA variants are differentially expressed. *Gene*, **352**, 127-136.
45. Santiago, C. and Marzluff, W.F. (1989) Expression of the U1 RNA gene repeat during early sea urchin development: evidence for a switch in U1 RNA genes during development. *Proc Natl Acad Sci U S A*, **86**, 2572-2576.
46. Nash, M.A., Sakallah, S., Santiago, C., Yu, J.C. and Marzluff, W.F. (1989) A developmental switch in sea urchin U1 RNA. *Dev Biol*, **134**, 289-296.
47. Krol, A., Lund, E. and Dahlberg, J.E. (1985) The two embryonic U1 RNA genes of Xenopus laevis have both common and gene-specific transcription signals. *EMBO J*, **4**, 1529-1535.
48. Stevenson, K.A., Yu, J.C. and Marzluff, W.F. (1992) A conserved region in the sea urchin U1 snRNA promoter interacts with a developmentally regulated factor. *Nucleic Acids Res*, **20**, 351-357.
49. Lund, E. (1988) Heterogeneity of human U1 snRNAs. *Nucleic Acids Res*, **16**, 5813-5826.
50. Patton, J.G. and Wieben, E.D. (1987) U1 precursors: variant 3' flanking sequences are transcribed in human cells. *J Cell Biol*, **104**, 175-182.
51. Kyriakopoulou, C., Larsson, P., Liu, L., Schuster, J., Soderbom, F., Kirsebom, L.A. and Virtanen, A. (2006) U1-like snRNAs lacking complementarity to canonical 5' splice sites. *Rna*, **12**, 1603-1611.
52. O'Reilly, D., Dienstbier, M., Cowley, S.A., Vazquez, P., Drozdz, M., Taylor, S., James, W.S. and Murphy, S. (2013) Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Res*, **23**, 281-291.
53. Denison, R.A., Van Arsdell, S.W., Bernstein, L.B. and Weiner, A.M. (1981) Abundant pseudogenes for small nuclear RNAs are dispersed in the human genome. *Proc Natl Acad Sci U S A*, **78**, 810-814.
54. Lindgren, V., Bernstein, L.B., Weiner, A.M. and Francke, U. (1985) Human U1 small nuclear RNA pseudogenes do not map to the site of the U1 genes in 1p36 but are clustered in 1q12-q22. *Mol Cell Biol*, **5**, 2172-2180.
55. Manser, T. and Gesteland, R.F. (1982) Human U1 loci: genes for human U1 RNA have dramatically similar genomic environments. *Cell*, **29**, 257-264.
56. Bernstein, L.B., Manser, T. and Weiner, A.M. (1985) Human U1 small nuclear RNA genes: extensive conservation of flanking sequences suggests cycles of gene amplification and transposition. *Mol Cell Biol*, **5**, 2159-2171.
57. Denison, R.A. and Weiner, A.M. (1982) Human U1 RNA pseudogenes may be generated by both DNA- and RNA-mediated mechanisms. *Mol Cell Biol*, **2**, 815-828.
58. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T. and Gesteland, R.F. (1981) Direct repeats flank three small nuclear RNA pseudogenes in the human genome. *Cell*, **26**, 11-17.

59. Somarelli, J.A., Mesa, A., Rodriguez, C.E., Sharma, S. and Herrera, R.J. (2014) U1 small nuclear RNA variants differentially form ribonucleoprotein particles in vitro. *Gene*, **540**, 11-15.
60. Egloff, S., O'Reilly, D. and Murphy, S. (2008) Expression of human snRNA genes from beginning to end. *Biochem Soc Trans*, **36**, 590-594.
61. Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D. and Murphy, S. (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, **318**, 1777-1779.
62. James Faresse, N., Canella, D., Praz, V., Michaud, J., Romascano, D. and Hernandez, N. (2012) Genomic study of RNA polymerase II and III SNAPc-bound promoters reveals a gene transcribed by both enzymes and a broad use of common activators. *PLoS Genet*, **8**, e1003028.
63. Somarelli, J.A., Mesa, A., Rodriguez, C.E., Sharma, S. and Herrera, R.J. (2014) U1 small nuclear RNA variants differentially form ribonucleoprotein particles in vitro. *Gene*.
64. Catterall, J.F., O'Malley, B.W., Robertson, M.A., Staden, R., Tanaka, Y. and Brownlee, G.G. (1978) Nucleotide sequence homology at 12 intron--exon junctions in the chick ovalbumin gene. *Nature*, **275**, 510-513.
65. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) Are snRNPs involved in splicing? *Nature*, **283**, 220-224.
66. Reddy, R., Ro-Choi, T.S., Henning, D. and Busch, H. (1974) Primary sequence of U-1 nuclear ribonucleic acid of Novikoff hepatoma ascites cells. *J Biol Chem*, **249**, 6486-6494.
67. Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J.A. (1983) The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell*, **33**, 509-518.
68. Zhuang, Y. and Weiner, A.M. (1986) A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell*, **46**, 827-835.
69. Sheth, N., Roca, X., Hastings, M.L., Roeder, T., Krainer, A.R. and Sachidanandam, R. (2006) Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res*, **34**, 3955-3967.
70. Roca, X. and Krainer, A.R. (2009) Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. *Nat Struct Mol Biol*, **16**, 176-182.
71. Roca, X., Akerman, M., Gaus, H., Berdeja, A., Bennett, C.F. and Krainer, A.R. (2012) Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev*, **26**, 1098-1109.
72. Cao, W. and Garcia-Blanco, M.A. (1998) A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding. *J Biol Chem*, **273**, 20629-20635.
73. Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garcia-Blanco, M.A. and Manley, J.L. (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**, 119-124.
74. Xiao, X., Wang, Z., Jang, M., Nutiu, R., Wang, E.T. and Burge, C.B. (2009) Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol*, **16**, 1094-1100.
75. Long, J.C. and Caceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem J*, **417**, 15-27.
76. Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T. and Chabot, B. (2007) hnRNP proteins and splicing control. *Adv Exp Med Biol*, **623**, 123-147.
77. Crispino, J.D., Blencowe, B.J. and Sharp, P.A. (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science*, **265**, 1866-1869.
78. Tarn, W.Y. and Steitz, J.A. (1994) SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev*, **8**, 2704-2717.
79. Raponi, M., Buratti, E., Dassie, E., Upadhyaya, M. and Baralle, D. (2009) Low U1 snRNP dependence at the NF1 exon 29 donor splice site. *FEBS J*, **276**, 2060-2073.

80. Fukumura, K., Taniguchi, I., Sakamoto, H., Ohno, M. and Inoue, K. (2009) U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res*, **37**, 1907-1914.
81. Rossi, F., Forne, T., Antoine, E., Tazi, J., Brunel, C. and Cathala, G. (1996) Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5' splice site-U1 snRNP interaction. *J Biol Chem*, **271**, 23985-23991.
82. Lund, M. and Kjems, J. (2002) Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. *Rna*, **8**, 166-179.
83. Crispino, J.D. and Sharp, P.A. (1995) A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes Dev*, **9**, 2314-2323.
84. Hwang, D.Y. and Cohen, J.B. (1996) U1 snRNA promotes the selection of nearby 5' splice sites by U6 snRNA in mammalian cells. *Genes Dev*, **10**, 338-350.
85. Crispino, J.D., Mermoud, J.E., Lamond, A.I. and Sharp, P.A. (1996) Cis-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing. *RNA*, **2**, 664-673.
86. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. and Sharp, P.A. (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, **499**, 360-363.
87. Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L. and Dreyfuss, G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, **468**, 664-668.
88. Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L. *et al.* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell*, **150**, 53-64.
89. Vorlova, S., Rocco, G., Lefave, C.V., Jodelka, F.M., Hess, K., Hastings, M.L., Henke, E. and Cartegni, L. (2011) Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol Cell*, **43**, 927-939.
90. Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845-1848.
91. Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A. and Sharp, P.A. (2008) Divergent transcription from active promoters. *Science*, **322**, 1849-1851.
92. Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H. and Jensen, T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, **322**, 1851-1854.
93. Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R. *et al.* (2013) Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol*, **20**, 923-928.
94. Levitt, N., Briggs, D., Gil, A. and Proudfoot, N.J. (1989) Definition of an efficient synthetic poly(A) site. *Genes Dev*, **3**, 1019-1025.
95. Furth, P.A. and Baker, C.C. (1991) An element in the bovine papillomavirus late 3' untranslated region reduces polyadenylated cytoplasmic RNA levels. *J Virol*, **65**, 5806-5812.
96. Furth, P.A., Choe, W.T., Rex, J.H., Byrne, J.C. and Baker, C.C. (1994) Sequences homologous to 5' splice sites are required for the inhibitory activity of papillomavirus late 3' untranslated regions. *Mol Cell Biol*, **14**, 5278-5289.
97. Ashe, M.P., Griffin, P., James, W. and Proudfoot, N.J. (1995) Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev*, **9**, 3008-3025.
98. Ashe, M.P., Pearson, L.H. and Proudfoot, N.J. (1997) The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *EMBO J*, **16**, 5752-5763.
99. Vagner, S., Ruegsegger, U., Gunderson, S.I., Keller, W. and Mattaj, I.W. (2000) Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP. *RNA*, **6**, 178-188.

100. Gunderson, S.I., Polycarpou-Schwarz, M. and Mattaj, I.W. (1998) U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell*, **1**, 255-264.
101. Abad, X., Vera, M., Jung, S.P., Oswald, E., Romero, I., Amin, V., Fortes, P. and Gunderson, S.I. (2008) Requirements for gene silencing mediated by U1 snRNA binding to a target sequence. *Nucleic Acids Res*, **36**, 2338-2352.
102. Awasthi, S. and Alwine, J.C. (2003) Association of polyadenylation cleavage factor I with U1 snRNP. *RNA*, **9**, 1400-1409.
103. Ko, B. and Gunderson, S.I. (2002) Identification of new poly(A) polymerase-inhibitory proteins capable of regulating pre-mRNA polyadenylation. *J Mol Biol*, **318**, 1189-1206.
104. Lou, H., Neugebauer, K.M., Gagel, R.F. and Berget, S.M. (1998) Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol Cell Biol*, **18**, 4977-4985.
105. Spraggon, L. and Cartegni, L. (2013) U1 snRNP-Dependent Suppression of Polyadenylation: Physiological Role and Therapeutic Opportunities in Cancer. *Int J Cell Biol*, **2013**, 846510.
106. Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E. and Palmiter, R.D. (1988) Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci U S A*, **85**, 836-840.
107. Zago, P., Baralle, M., Ayala, Y.M., Skoko, N., Zacchigna, S., Buratti, E. and Tisminetzky, S. (2009) Improving human interferon-beta production in mammalian cell lines by insertion of an intronic sequence within its naturally uninterrupted gene. *Biotechnology and applied biochemistry*, **52**, 191-198.
108. Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A. and Proudfoot, N.J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev*, **16**, 2792-2799.
109. Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H. and Kjems, J. (2008) A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell*, **29**, 271-278.
110. Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N.J. and Akoulitchev, A. (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat Struct Biol*, **9**, 800-805.
111. Shiekhattar, R., Mermelstein, F., Fisher, R.P., Drapkin, R., Dynlacht, B., Wessling, H.C., Morgan, D.O. and Reinberg, D. (1995) Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature*, **374**, 283-287.
112. Akoulitchev, S., Makela, T.P., Weinberg, R.A. and Reinberg, D. (1995) Requirement for TFIIH kinase activity in transcription by RNA polymerase II. *Nature*, **377**, 557-560.
113. Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D. and Ansari, A.Z. (2009) TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell*, **34**, 387-393.
114. Serizawa, H., Makela, T.P., Conaway, J.W., Conaway, R.C., Weinberg, R.A. and Young, R.A. (1995) Association of Cdk-activating kinase subunits with transcription factor TFIIH. *Nature*, **374**, 280-282.
115. Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P. and Bentley, D.L. (2009) TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol Cell Biol*, **29**, 5455-5464.
116. O'Gorman, W., Thomas, B., Kwek, K.Y., Furger, A. and Akoulitchev, A. (2005) Analysis of U1 small nuclear RNA interaction with cyclin H. *J Biol Chem*, **280**, 36920-36925.
117. Jobert, L., Pinzon, N., Van Herreweghe, E., Jady, B.E., Guialis, A., Kiss, T. and Tora, L. (2009) Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep*, **10**, 494-500.

118. Bertolotti, A., Lutz, Y., Heard, D.J., Chambon, P. and Tora, L. (1996) hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J*, **15**, 5022-5031.
119. Leichter, M., Marko, M., Ganou, V., Patrinou-Georgoula, M., Tora, L. and Guialis, A. (2011) A fraction of the transcription factor TAF15 participates in interactions with a subset of the spliceosomal U1 snRNP complex. *Biochimica et biophysica acta*, **1814**, 1812-1824.
120. Barbosa-Morais, N.L., Carmo-Fonseca, M. and Aparicio, S. (2006) Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Res*, **16**, 66-77.
121. Schwartz, S.H., Silva, J., Burstein, D., Pupko, T., Eyras, E. and Ast, G. (2008) Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res*, **18**, 88-103.
122. Krol, A., Westhof, E., Bach, M., Luhrmann, R., Ebel, J.P. and Carbon, P. (1990) Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res*, **18**, 3803-3811.
123. Edery, P., Marcaillou, C., Sahbatou, M., Labalme, A., Chastang, J., Touraine, R., Tubacher, E., Senni, F., Bober, M.B., Nampoothiri, S. *et al.* (2011) Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science*, **332**, 240-243.
124. He, H., Liyanarachchi, S., Akagi, K., Nagy, R., Li, J., Dietrich, R.C., Li, W., Sebastian, N., Wen, B., Xin, B. *et al.* (2011) Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science*, **332**, 238-240.
125. Jia, Y., Mu, J.C. and Ackerman, S.L. (2012) Mutation of a U2 snRNA gene causes global disruption of alternative splicing and neurodegeneration. *Cell*, **148**, 296-308.
126. Mefford, H.C., Sharp, A.J., Baker, C., Itsara, A., Jiang, Z., Buysse, K., Huang, S., Maloney, V.K., Crolla, J.A., Baralle, D. *et al.* (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med*, **359**, 1685-1699.

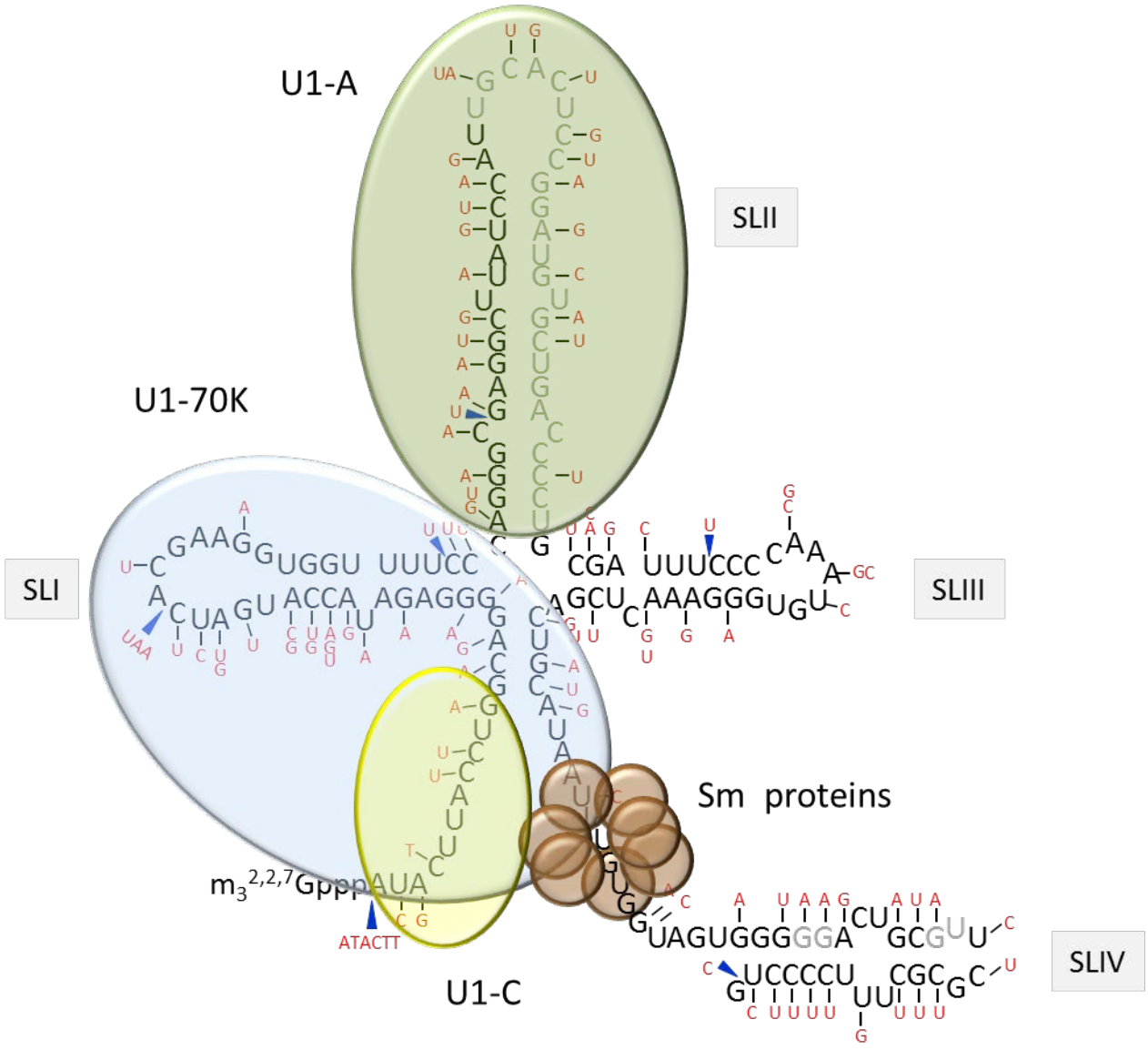

Figure 1: Composite figure illustrating the heterogenetity of the U1 and vU1 snRNAs

The U1 snRNA is illustrated in black in the form of a clover leaf structure. Base changes, which have been documented for human variant (v)U1 snRNAs located on chromosome 1q12-21, are illustrated in red. Insertion of additional bases is denoted by a blue triangle. Deleted regions, associated with some vU1 snRNAs are denoted in grey. $m_3^{2\ 2,7}$Gppp represents the snRNA-specific trimethyl cap structure. The positions of the U1-specific proteins U1-70K (blue), U1-A (green), U1-C (yellow) and Sm proteins (brown) are illustrated. U1-C does not associate with the U1 snRNA directly but interacts specifically with U1-70K and Sm-D3. U1-70k specifically interacts with stem loop II (SLI) of the U1 snRNA and makes additional contacts with U1-C and Sm-B/B' and Sm–D3.

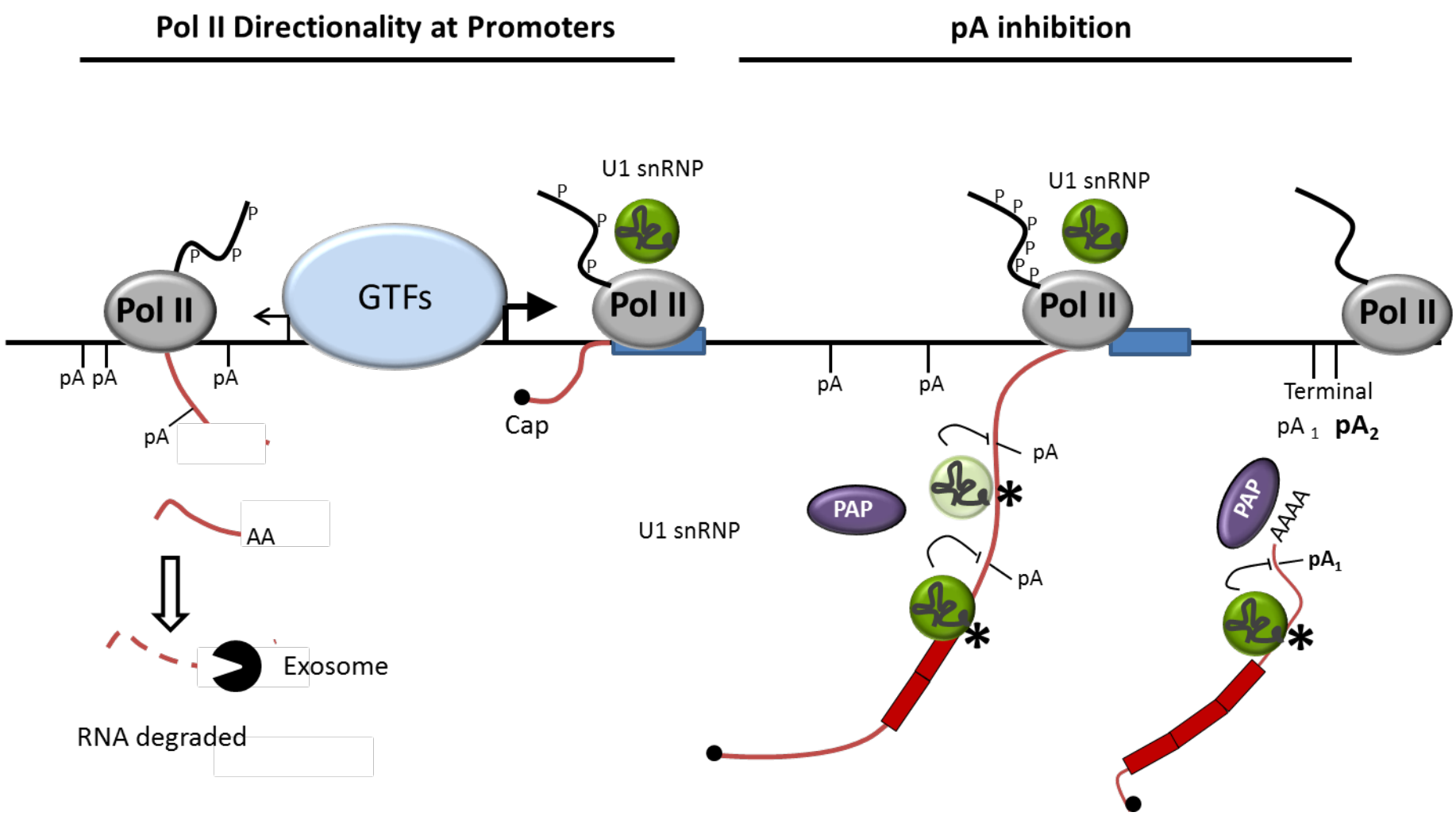
Pol II Directionality at Promoters
pA inhibition
U1 snRNP
Pol II
GTFs
pA pA
pA
Cap
AA
Exosome
RNA degraded
U1 snRNP
PAP
pA
Terminal
pA 1 pA2
AAAA
pA1

Figure 2: U1/vU1 snRNAs participates in pA site selection and Pol II directionality at promoters.

U1 snRNA binding sites (*) are located at exon/intron junctions and within intronic and UTR regions of pre-mRNAs. In addition, poly A (pA) hexamers are also found within intronic regions and the UTRs. If U1 snRNP associates with the pre-mRNA in the vicinity of a pA site, both cleavage and polyadenylation at the pA site can be inhibited. The U1-specific protein, U1-70K, contributes to this block in polyadenylation by interfering with the association of the poly A polymerase (PAP), which adds a poly adenosine tail to the end of the mRNA. Depending on where U1 snRNA binds, varying length mRNAs can be generated as a result of cleavage/polyadenylation at internal 'cryptic' pA sites or alternative pA sites within the 3'UTRs (terminal pA sites). vU1 snRNPs, denoted in light green, have also been shown to participate in polyadenylation control.

Motif analysis of regions flanking the promoters of human genes has confirmed specific enrichment of pA sites upstream and U1 snRNA binding sites downstream of the transcription start sites. Consequently, transcripts generated in the antisense direction are cleaved and polyadenylated early during transcription due to the lack of U1 snRNA binding sites to recruit U1 snRNP to block polyadenylation. These short polyadenylated RNAs are rapidly degraded by the exosome complex. Productive elongation by Pol II is favoured in the sense direction owning to the increased number of U1 snRNA binding sites which enable U1 snRNA to bind and block cleavage/polyadenylation at internal cryptic pA sites in favour of proper polyadenylation at the end.

The DNA and RNA are depicted as black and red lines, respectively. Exons are denoted as blue (DNA) or red boxes (RNA). The start of transcription is illustrated by an arrow. Differential phosphorylation of the C-terminal domain of the large subunit of Pol II (black wavy line) throughout the transcription cycle is noted by 'P'. Cap: m_7G trimethy cap structure.

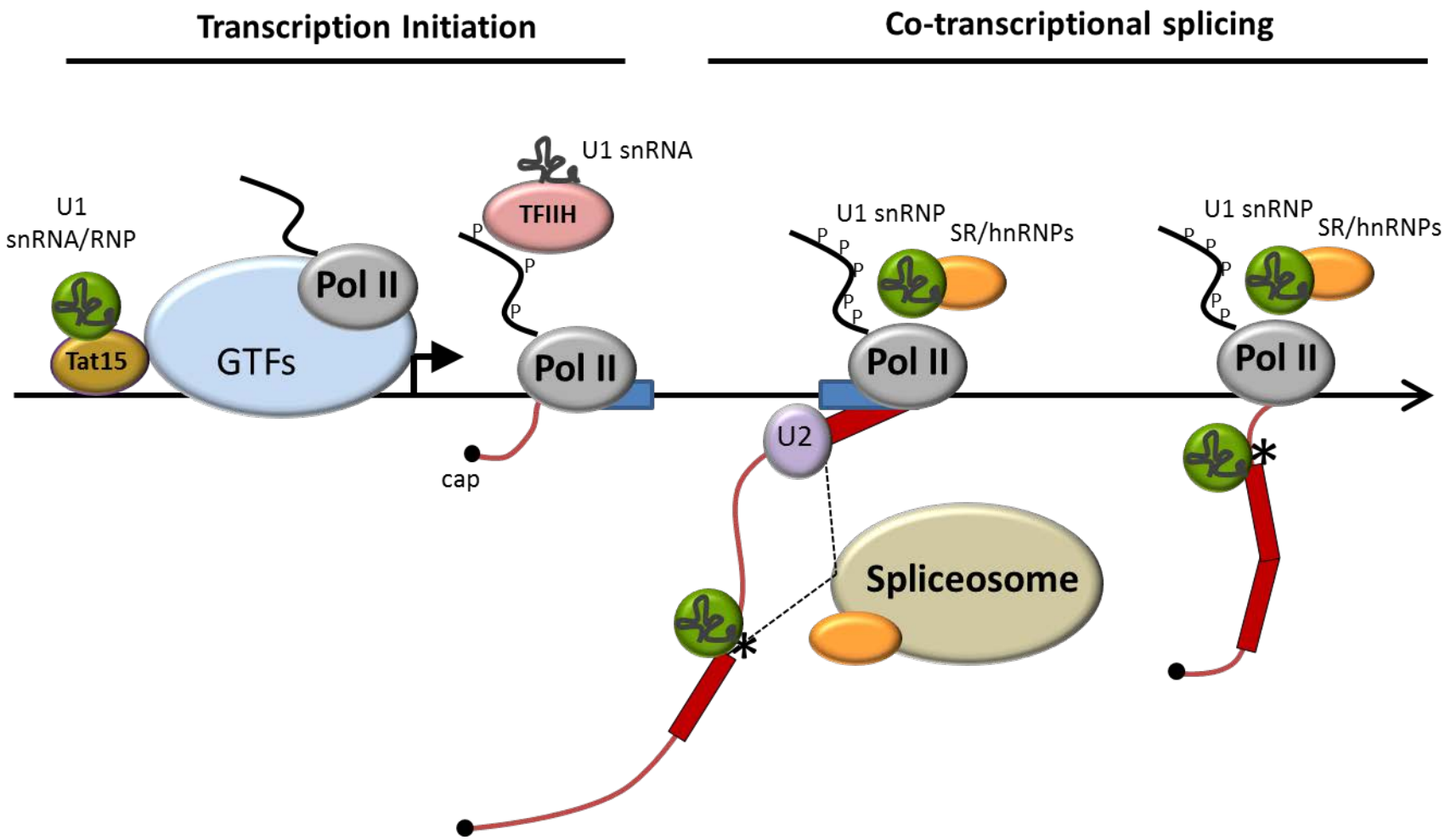
Transcription Initiation
Co-transcriptional splicing
U1
snRNA/RNP
Tat15
GTFs
Pol II
U1 snRNA
TFIIH
P
Pol II
cap
U1 snRNP
SR/hnRNPs
Pol II
U2
Spliceosome
U1 snRNP
SR/hnRNPs
Pol II

Figure 3: U1 snRNA participates in transcription initiation and pre-mRNA splicing.

U1 snRNA is recruited to the transcription initiation complex by Taf15, which associates with the pre-initiation complex either through protein:protein interactions with the general transcription factors (GTFs) or through recognition of specific promoter elements. Taf15 forms two complexes with the U1 snRNA: the U1 snRNA alone or as part of a ribonucleoprotein complex (U1 snRNA/RNP). Following initiation, TFIIH phosphorylates the C-terminal domain of the large subunit of Pol II (black wavy line) to promote productive elongation. U1 snRNA interacts with the TFIIH associated kinase complex and is thought to enhance the rate of transcription initiation/re-initiation.

U1 snRNP also associates with the transcribing polymerase during productive elongation, which is thought to position U1 snRNA in close proximity to the emerging pre-RNA to enhance base-pairing interactions with the 5' splice sites (*). Together with splicing factors (SR and hnRNP proteins), U2 snRNA and the Spliceosome, U1 snRNP facilitates excision of the intronic sequences and splices together the exonic sequences to generate a translatable mRNA. Figure 3 is labelled as in Figure 2 above with the addition of U2: U2 snRNP; SR: serine-arginine splicing factors; hnRNP: heterogeneous nuclear ribonucleoproteins.