Identification and transcriptional regulation of eukaryotic small nuclear RNA genes.

PhD in Biochemistry and Molecular Biology
XXI cycle: 2006-2008

Gloria Fiorino

Department of Biochemistry and Molecular Biology
University of Parma, Italy

PhD Coordinator: Prof. Gian Luigi Rossi

PhD Tutor: Prof. Giorgio Dieci

Contents

Small nuclear RNAs: a brief introduction

An abundance of RNA regulators: non coding RNAs

RNA molecules are active participants in the regulation, catalysis and control of many fundamental cellular processes, role that was thought to be restricted to proteins until a few years ago. These regulatory RNAs, that do not encode a protein, are referred to as non-protein-coding RNAs (ncRNAs). They are characterized by different size, tissue specific expression pattern and biological functions. Despite their differences, they elicit their biological responses through one of the three basic mechanism: catalysing biological reactions, binding to and modulating the activity of a protein, or base-pairing with a target nucleic acid.

The structure of a ncRNA is crucial for its functionality; as happens for proteins, ncRNAs fold into specific higher order structures that impart a function to the molecule. Often these RNAs require several partner proteins; that is, the functional unit is the non-coding ribonucleoprotein (ncRNP) (fig. 1). The steps of ncRNP assembly are often not well defined but can be regulated to control the activity of the complex.

ncRNA activities are important at many levels of gene expression such as splicing, mRNA turnover, gene silencing and translation (see table 1). Regulatory RNAs include small interfering and micro RNA, mainly involved in postranscriptional regulation, and other “structural molecules” including ribosomal, transfer, small nucleolar, small cytoplasmic (sc), and small nuclear (sn) RNAs that function in the processing, translation and degradation of other RNA molecules. Recent studies had drawn the attention on the regulatory roles of such “structural RNA class”. Pol III-transcribed Y RNAs are required for chromosomal DNA replication (Christov et al., 2006), while 7SK RNA and U1 snRNA influence messenger RNA transcription; the first represses transcript elongation by Pol II by binding the elongation factor P-TEFb (Nguyen et al., 2001) while the second co-purifies with the multisubunit transcription factor TFIIH and is involved both in transcription and initiation by pol II (Kwek et al., 2002). Human RNAse P RNA is required for transcription of tRNA genes, thus linking transcription with processing in regulation of tRNA gene expression (Jarrous et al., 2007). The line between regulatory and structural classes has became less evident since the recent discovery of snRNA-like transcription units that encode anti-sense transcripts capable of post-transcriptional gene regulation (Pagano et al., 2007).

Interestingly many of these ncRNA are expressed in a tissue-specific manner, suggesting specific and regulated functions of the RNAs, rather than fundamental housekeeping roles played in all tissues. The small cytoplasmic BC200 RNA (and its rodent functional counterpart BC1) are specifically detected in the central nervous system, where they could be involved in translation of dendritic mRNAs (Lin, Y. et al. (2001). Dittmar et al. (2006) demonstrated tissue-specific differences in the expression of individual tRNA species.

Finally the widespread functions of ncRNA makes it likely that alterations in their levels and activity will compromise diverse cellular processes; a putative role of BC1/BC200 in memory processes is in agreement with a recent study showing that neocortical expression of BC200 RNA is up-regulated in Alzheimer disease (AD) brains (Mus et al., 2007).

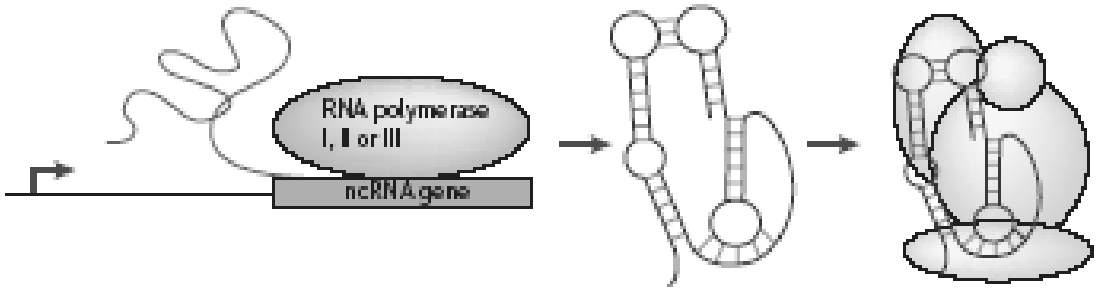

Fig. 1 The structure of a non-coding RNA is crucial for its function. ncRNA genes can be transcribed by either RNA pol I, II, III. ncRNAs fold into specific structures that impart a function to the molecule. Often these RNAs are incorporated into large complexes that contain proteins and sometimes other nucleic acids (Goodrich and Kugel, 2006)

ncRNA	Species	Functions
7SKsnRNA	Human	Inhibition of P-TEFb and RNAPII elongation
H1 RNA (RNase P RNA)	Human	tRNA maturation, RNAPIII transcription
U1 snRNA	Human/yeast	mRNAsplicing, stimulation of RNAPII transcription/ mRNA splicing
U2 snRNA	Human/yeast	mRNAsplicing, stimulation of RNAPII elongation/ mRNA splicing
U4 snRNA	Human/yeast	mRNAsplicing
U5 snRNA	Human/yeast	mRNAsplicing
U6 snRNA	Human/yeast	mRNAsplicing
MRP RNA	Human/yeast	rRNA processing, mithocondrial DNA replication/ cell cycle progression
U7 snRNA	Human/yeast	Histone pre-mRNA 3'end formation
C/D snoRNAs	Human/yeast	2' O-methylation of rRNA, snRNAs and tRNAs; rRNA processing
H/ACA snoRNAs	Human/yeast	Pseudouridylation of rRNAs, snRNAs and tRNAs; rRNA processing
tRNA	Human/yeast	mRNA translation
Y scRNA	Human	Nc RNA degradation;chromosomal replication
Vault RNA	Human	Nucleocytoplasmic trafficking, multidrug resistance
SRP RNA	Human	Protein translocation to the endoplasmic reticulum
miRNA	Human	Gene silencing; translational repression or cleavage of target mRNas
SINE-encoded RNAs	Human	RNa editing; alternative splicing; chromosomal recombination; gene-expression regulation; cell stress response; miRNA target
siRNA	Human	Gene silencing; cleavage of RNAs derived from viruses, retroelements and repeat sequences
BC200	Human	Possible role in translation of dendritic mRNA

Table 1. Functions of ncRNA in human and yeast species.

The snRNAs

snRNAs comprise a small group of highly abundant, non-polyadenylated, non-protein-coding transcripts that function in the nucleoplasm. They assemble with numerous protein factors to form small nuclear ribonucleoproteins (snRNPs). The snRNAs can be divided into two classes: Sm-class RNAs (U1, U2, U4, U5, U7) are characterized by a 5'-trimethylguanosine cap, a 3' stem loop and by their ability to associate to a group of seven Sm proteins, through the so-called Sm site. Lsm-class RNAs (U6) contain a monomethylphosphate cap and a 3' stem-loop, terminating in a stretch of uridines (fig 2). With the exception of the U7 snRNP, which function in histone pre-mRNA 3' processing, the other snRNPs form the core of the spliceosome and catalyse the removal of introns from pre-mRNA (Matera et al., 2007).

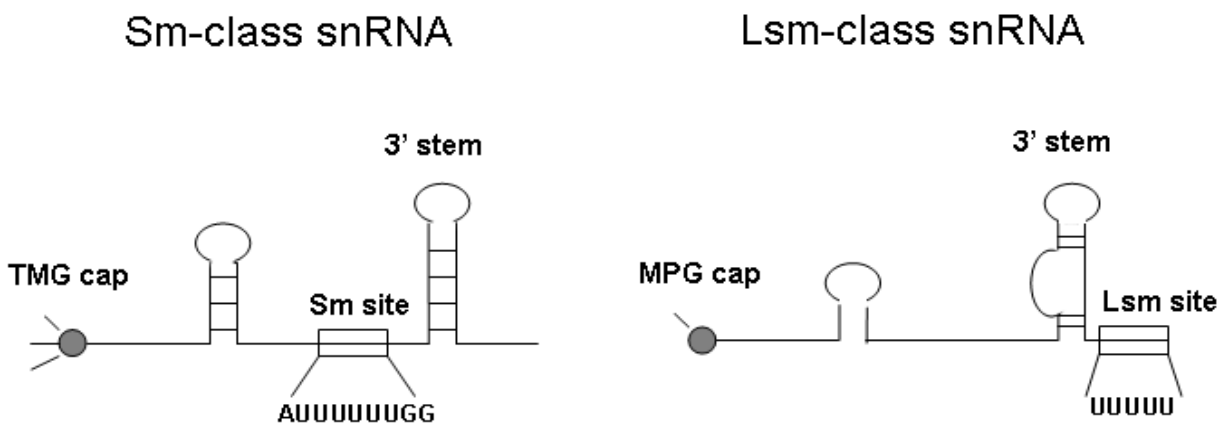

Fig. 2 Feautures of Sm- and Lsm-class small nuclear RNAs. Sm-class snRNA contain three important recognition elements: a 5'-trimethylguanosine (TMG) cap, an Sm-protein-binding site (Sm site) and a 3' stem-loop structure. Lsm-class snRNA contain a 5'-monomethylphosphate (MPG) cap, a 3' stem and terminate in a stretch of uridine residues (Lsm site).

snRNA gene promoter structure

Promoters of snRNA genes are characterized by typical features. First, snRNA genes usually contain transcriptional elements that are unique to this gene group. Secondly, the promoters of both RNA pol II and RNA pol III snRNA genes within a species are structurally related. Thirdly, although snRNA promoters are highly conserved within the same species, they vary greatly between different genera indicating that they have evolved rapidly. Fig. 3 shows the structure of snRNA promoters from *Homo sapiens* (Hs) *Arabidopsis thaliana* (At) *and Drosophila melanogaster* (Dm) and illustrates how snRNA promoters have diverged during evolution, maintaining close similarity between those recognized by pol II and those recognized by pol III (Hernandez, 2001).

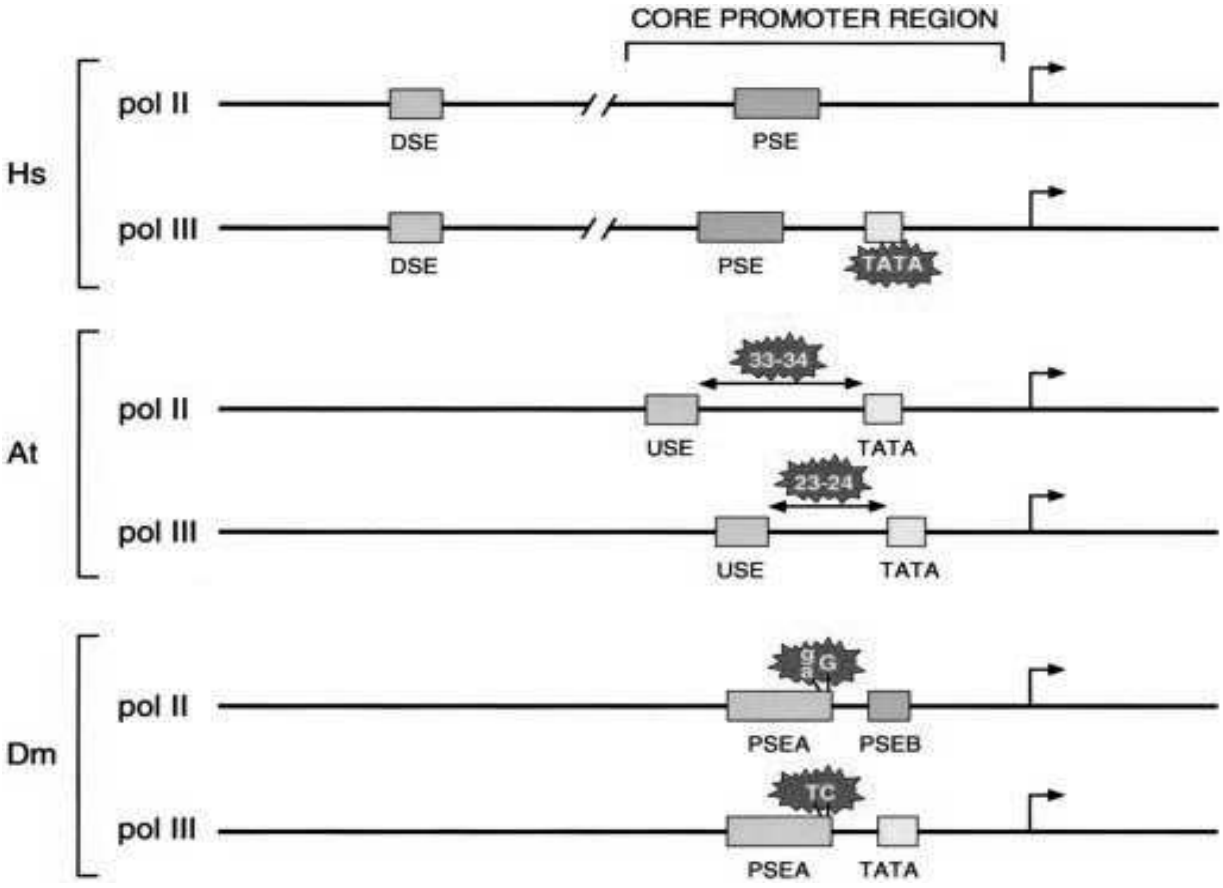

Fig. 3 Structure of the *H. sapiens* (Hs), *A. thaliana* (At) and *D. melanogaster* (Dm) snRNA promoters. (Hernandez, 2001)

Members of the human snRNA and some related scRNA gene families are characterized by a diagnostic arrangement of promoter elements, minimally including a distal sequence element (DSE) that serves as enhancer and a proximal sequence element (PSE) that is located in the core promoter region upstream from the start sit of transcription. Some genes contain a TATA box located adjacently to the PSE. The TATA element acts as a determinant for polymerase specificity: the combination of the extragenic PSE and TATA elements directs recruitment of the RNA polymerase III-specific transcriptional machinery whereas the absence of a TATA box specifies recruitment of the RNA pol II-specific transcriptional apparatus. U1 and U2 snRNA promoters and the U6 snRNA promoters serve respectively as prototypic pol II and pol III snRNA promoters. Both DSE and PSE can be interchanged between pol II and pol III promoters with no effect on polymerase specificity, while mutation in the TATA box induces RNA polymerase II transcription from the U6 promoter and insertion of the TATA box conversely causes RNA polymerase III transcription from the U2 promoter (Lobo and Hernandez ,1989; Mattaj et al.,1988). Other categories of RNA polymerase III-transcribed genes, such as vault RNA, contain a combination of extragenic PSE-like and TATA promoter elements along with canonical intragenic elements typically utilized for RNA polymerase III transcription of transfer RNA genes (Vilalta et al., 1994). The A. thaliana snRNA promoters consist of an upstream sequence element (USE) and a TATA box. The USE is a plant snRNA gene-specific enhancer and the spacing between the USE and the TATA box is the major determinant of RNA polymerase specificity (Waibel and Filipowicz, 1990).

In *D. melanogaster* both Pol II and III promoters contain a quite conserved 21 bp element called the PSEA located at about 42 bp upstream of the start site. A PSEB or a canonical TATA box are located downstream the PSEA in Pol II and Pol III snRNAs respectively (Zamrod et al., 1993). Polymerase specificity is determined by position 19 and 20 of the PSEA, g/aG in Pol II and TC in Pol III promoters (Jensen, 1998). Transcription of Schizosaccharomyces pombe U2 gene is directed by two essential promoter elements: spUSE centered at -55, which functions as an activator and a TATA box at -26 (Zhou and Lobo-Ruppert, 2001). In *S. cerevisiae* only the Pol III U6 snRNA transcripton has been studied. The promoter contains a TATA box and other two elements are required for correct efficient transcription: an A box located within the coding region, as in tRNA genes, and a B box located at an anomalous position in 3' flanking region (Eschenlauer , 1993).

The PSE binding factor

The PSE sequence (fig 4A) is recognized by the snRNA activating protein complex (SNAPc), also known as the PTF transcription factor (PTF) or PSE-binding protein (PBP). SNAPc is composed of at least five proteins SNAP190, SNAP 50, SNAP45, SNAP43 and SNAP19 (fig.4B). DNA binding by SNAPc requires both SNAP190 and SNAP50 which directly bind to DNA via their Myb and zinc-fingers domains respectively (Wong et al, 1998; Jawdekar et al, 2006). SNAP43 is required for assembly of SNAP190 and SNAP50 into a DNA binding competent complex (Hinkley et al., 2003). These subunits are widely conserved through evolution, with omologues in vertebrate species, Drosophila, Trypanosomes and plants. In contrast human SNAP19 and SNAP45 subunits are dispensable for transcription in vitro and are not so widely conserved, suggesting that these vertebrate-specific SNAPc subunits may have acquired specialized regulatory roles for snRNA transcription (Mittal et al., 1999). SNAPc plays a central role in snRNA transcription, being involved in pre-initiation complex assembly, including direct promoter recognition, and serving as target for numerous activators and repressor of transcription. For example SNAP19 can either interact with the Oct-1 activator (Ford et al., 1998) and p53 tumor suppressor (Gridasova and Henry, 2000).

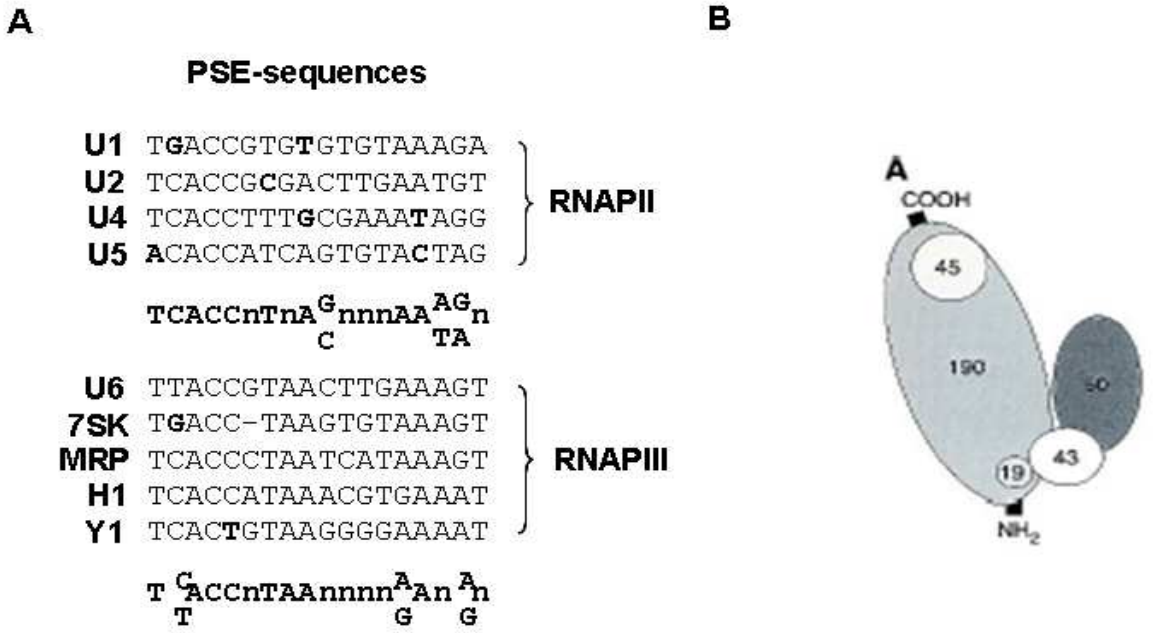

Fig. 4 (A) sequence comparison of snRNA genes core promoters. Consensus sequence is indicated below (Jawdekar and Henry, 2008) **(B) Schematic representation of SNAPc subunit organization.** (Hernandez, 2001)

Pre-initiation complex assembly at snRNA gene promoter

Pol II and Pol III transcribed genes share the initial promoter recognition by SNAPc complex.

For TATA-less snRNA genes, SNAPC binding is a prelude to the recruitment of traditional components of the general transcription machinery, including TBP, TFIIA,TFIIB,TFIIE, and TFIIF (Kuhulman et al., 1999), used also for mRNA transcription by RNA polymerase II. The role, if any of TFIIH in snRNA gene transcription is not clear. Interestingly U1 snRNA was found associated to TFIIH to stimulate RNA polymerase II transcription of certain mRNA genes (Kwek et al., 2002) suggesting that U1 snRNA may contribute to RNA polymerase II regulation. An intriguing possibility is that TFIIH complexes differing by the presence or absence of U1 snRNA participate in active transcription of mRNA and snRNA genes.

SNAPc associates with TBP (Sadowski et al., 1996); although no homologues for SNAPC have been described in yeast, it appears that a role for SNAPC in coordinating TBP activity at snRNA genes is evolutionarily conserved.

As with TATA-less RNA polymerase II-transcribed snRNA genes, SNAPC recruits TBP to RNA polymerase III-transcribed genes; the SNAPC/TBP juxtaposition results in the recruitment of a TFIIB-related factor called Brf2 (Schramm et al, 2000). Brf2 differs substantially from Brf1 used for transcription of other RNA polymerase III-transcribed genes in humans, and from the Brf used for U6 snRNA gene transcription in yeast. Promoter recruitment of TFIIB for RNA polymerase II transcription or the TFIIB-related Brf2 factor for RNA polymerase III transcription, as dictated by the absence or presence of the TATA box, serves as a critical determinant for utilization of different RNA polymerases in humans. The assembly of SNAPC, TBP, and Brf2 facilitates recruitment of the SANT- 187 domain containing protein Bdp1 (Jawdekar et al., 2006), a factor globally utilized for all RNA polymerase III transcription (Schramm and Hernandez,2002).

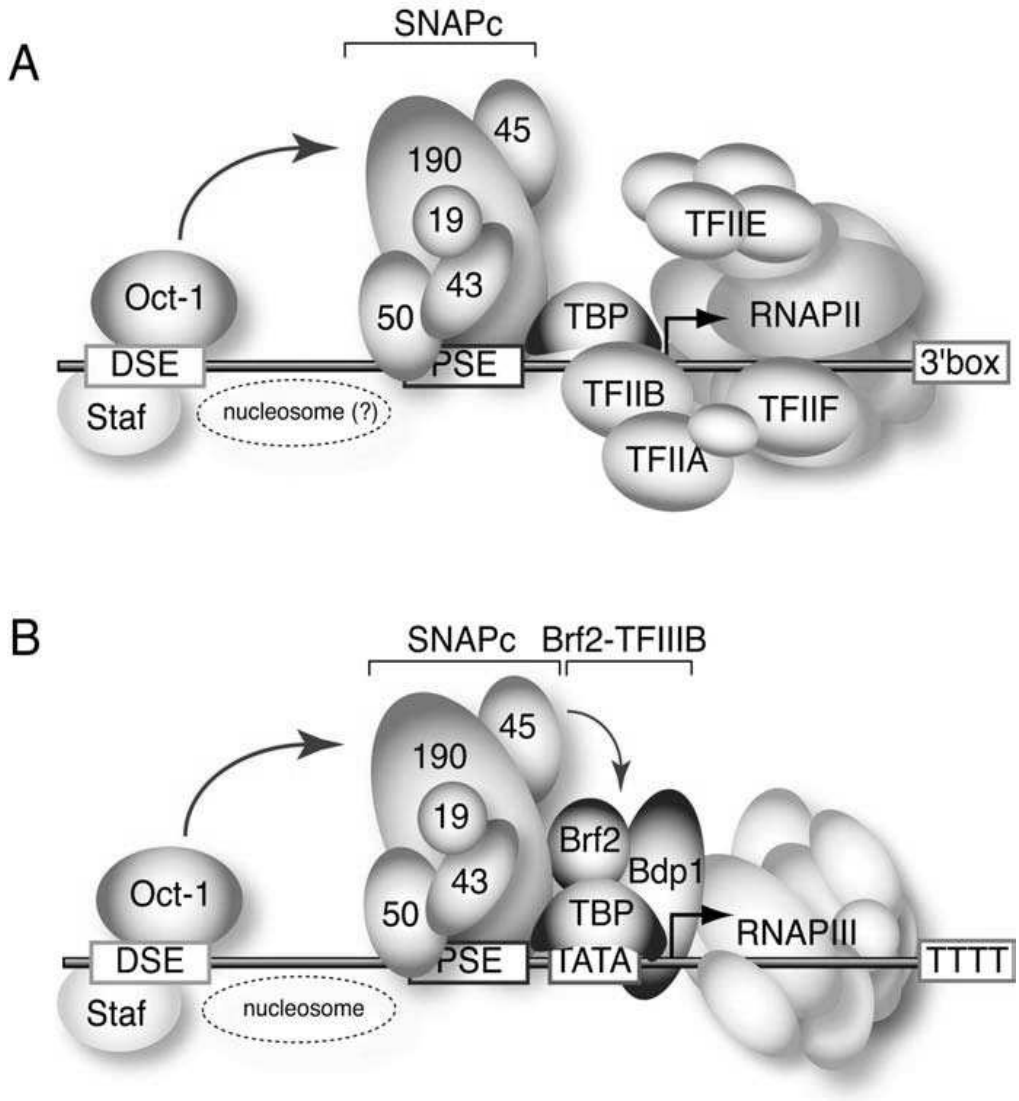

Fig. 5 Factors involved in human snRNA gene transcription. The transcription of human snRNA genes by RNA polymerases II (A) and III (B) involves a combination of shared factors including the Staf and Oct-1 activators and the general transcription factors SNAPC and TBP along with additional factors specialized for transcription by only one polymerase. RNA polymerase II and III termination is directed by the 3′ box or by the TTTT terminator, respectively. The numbers within SNAPC represent the apparent molecular weights of each subunit (Jawdekar and Henry, 2008).

Activation of snRNA gene transcription

The high-level expression of snRNA in cells depends upon the DSE, typically located at about 200 bp upstream the transcription start site. The DSE is composed by various protein binding site, but one of them is almost invariably the octamer sequence ATGCAAAT, recognized by the POU domain activator protein Oct-1 (Carbon et al., 1987). In addiction the DSE contains a Staf-responsive element recognized by the snRNA transcriptional activator Staf, also known as SPH binding factor (Myslinski et al., 1998). Some snRNA genes harbor Sp1 binding sites adjacent to the DSE. These proteins function as activators synergizing to recruit SNAPc. The mechanism is better understood for Oct-1, wherein direct contacts between the Oct-POU domain and the SNAP190 subunit of SNAPc contributes to increased SNAPc recruitment (Mittal, 1996). The strict spacing between the DSE and PSE is conserved in most, but not all, snRNA genes, and at least for U6 and 7SK snRNA gene expression, a nucleosome positioned between these promoter elements contributes to activated transcription (Boyd et al, 2000; Zhao et al, 2001) spatially juxtaposing the DSE and PSE to facilitate direct interactions between Oct-1 and SNAP190 (fig 6). This observation suggest that factors modulating chromatin structure are important for regulated transcription. Consistently, a higher proportion of histone H3 at U6 promoters is acetylated in cells that maintain higher levels of RNA polymerase III transcription (GWJ, unpublished observations). The opposing activity of histone deacetylase (HDAC) factors is also important for transcriptional repression by the Retinoblastoma (RB) and p53 tumor suppressor proteins.

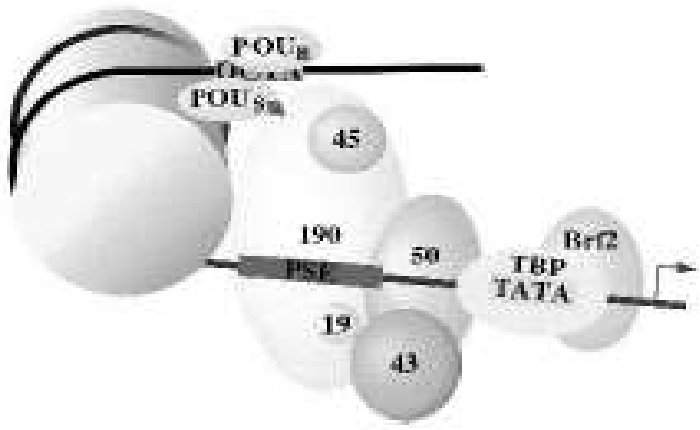

Fig. 6 Assembly of the human U6 snRNA initiation complex (Hernandez, 2001)

Transcription and 3'-end formation of snRNA genes.

In common with pre-mRNAs, newly pol II-transcribed pre-snRNAs are co-transcriptionally capped at the 5'-end by several enzymatic activities, with the attachment of a 7-methyl-guanosine (m^7G) residue to the γ-phosphate through a 5'-phosphoester linkage (Mattaj, 1986). Transcription and RNA processing reactions have been found to be closely linked in vivo; at the heart of this lies the repetitive CTD (C-terminal domain) of Pol II largest subunit, which serves to recruit and interact with many processing factors. Recruitment of processing factors by Pol II CTD is closely linked to its position along the gene and the phosphorylation state of specific serine residues. Recently phosphorylation of the serine in position 5 (Ser^5) and 2 (Ser^2) of the CTD has been related to activation of 5'-end capping and 3'-end formation of snRNA transcripts (Sylvain and Murphy, 2008). Pol II transcribed genes transcription in fact is linked with downstream RNA processing events. Indeed, appropriate snRNA termination depends upon the 3′ box, which is recognized by RNA polymerase II when recruited to PSE-containing genes (Hernandez and Lucito, 1988); if snRNA promoter is replaced by a mRNA promoter, 3'-end is not formed and the transcripts are polyadenilated. One candidate target in coupling U1 and U2 snRNA transcription to subsequent 3′ end processing is the Integrator complex, which is composed of at least twelve polypeptides and associates with the carboxy terminal domain (CTD) of the RNA polymerase II largest subunit. Intriguingly, serine 7 of the CTD is also specifically required for both snRNA gene transcription and downstream 3′ end formation but has no effects on mRNA transcription and is also critical for recruitment of the Integrator complex to U1 and U2 snRNA genes (Egloff et al., 2007).

In constrast to the vertebrate snRNAs, the 3'-ends of two snRNAs, U2 and U5L, in *S.cerevisiae* are formed by processing events unlinked to transcription (Chanfreau et al., 1997; Abuo Elela and Ares, 1998). Rnt1p, a homolog of bacterial RNase III, is strongly implicated in 3'-end formation; in fact U2 and U5L snRNA accumulation is impaired in cells that have a temperature-sensitive mutation in the RNT1 gene and purified, recombinant Rnt1p can cleave in vitro putative RNA stem-loop structures formed by U2 and U5 3'-flanking regions. However in the rnt1 mutant cells the steady-state levels of U1, U4, and U5S are increased at the restrictive temperature (Chanfreau et al., 1997), suggesting that their 3' ends may be generated by an

alternative mechanism. In the fission yeast *Schizosaccharomyces pombe* Pac1p, a RNAse III ortholog, is required for correct end-formation (Dewang et al., 1999). As in S. pombe and in S. cerevisiae, studies in plants (Connelly and Filipowicz, 1993) and in sea urchins (Wendelburg and Marzluff, 1992) indicate that the formation of the 3'ends of snRNAs can be dissociated from transcription.

Lsm-class snRNA genes are transcribed by Pol III; the run of uridines that forms the Lsm-binding site at the 3'-end also doubles as a Pol III transcription terminator. U6 snRNA is O-methylated on the 5' terminal (γ) phosphate in humans. Synthesis of the γ-methyl phosphate cap of U6 RNA in human cells is dependent upon structural determinants at the base of a conserved 5'-terminal stem (Singh et al,1990). However is not known whether yeast U6 RNA receives a γ-methyl phosphate cap.

snRNA nuclear export system.

In higher eukaryotes, Lsm class snRNAs never leave the nucleus, whereas the biogenesis of the Sm-class snRNPs requires export in the cytoplasmic for maturations events; the 5' cap structure and the length of the RNA are the key determinant in nuclear export. Following transcription and 3' processing in the nucleus, newly transcribed Sm-class snRNAs are transported to the cytoplasm by an export complex that contains the snRNA-specific export adaptor protein PHAX, the export receptor chromosome region maintainance CRM1, the cap-binding complex CBC and a Ran GTPase. These factors dissociate from the pre snRNA in the cytoplasm after binding with the survival of motor neuron (SMN) complex and desphorilation of PHAX (Segref et al., 2001). The SMN recognizes specific sequence elements (the Sm protein-binding site and the 3' stem-loop) and recruits a set of seven Sm proteins to form the core RNP. Following assembly on the Sm core, the 7-methylguanosine (m^7G) cap is hypermethylated to form 2,2,7-trimethylguanosine (TMG) cap structure and the 3' end is trimmed by an unidentified exonuclease (Exo). Methylguanosine caps of yeast snRNAs are also hypermethylated but the subcellular compartment in which hypermethylation occurs is not known. The formation of the TMG cap triggers the assembly of the import complex, which includes the import adaptor snurportin-1 (SPN) and the import receptor importin-β (Imp-β). After reimport into the nucleus, snRNPs together with numerous other splicing factors assemble into the functional spliceosome.

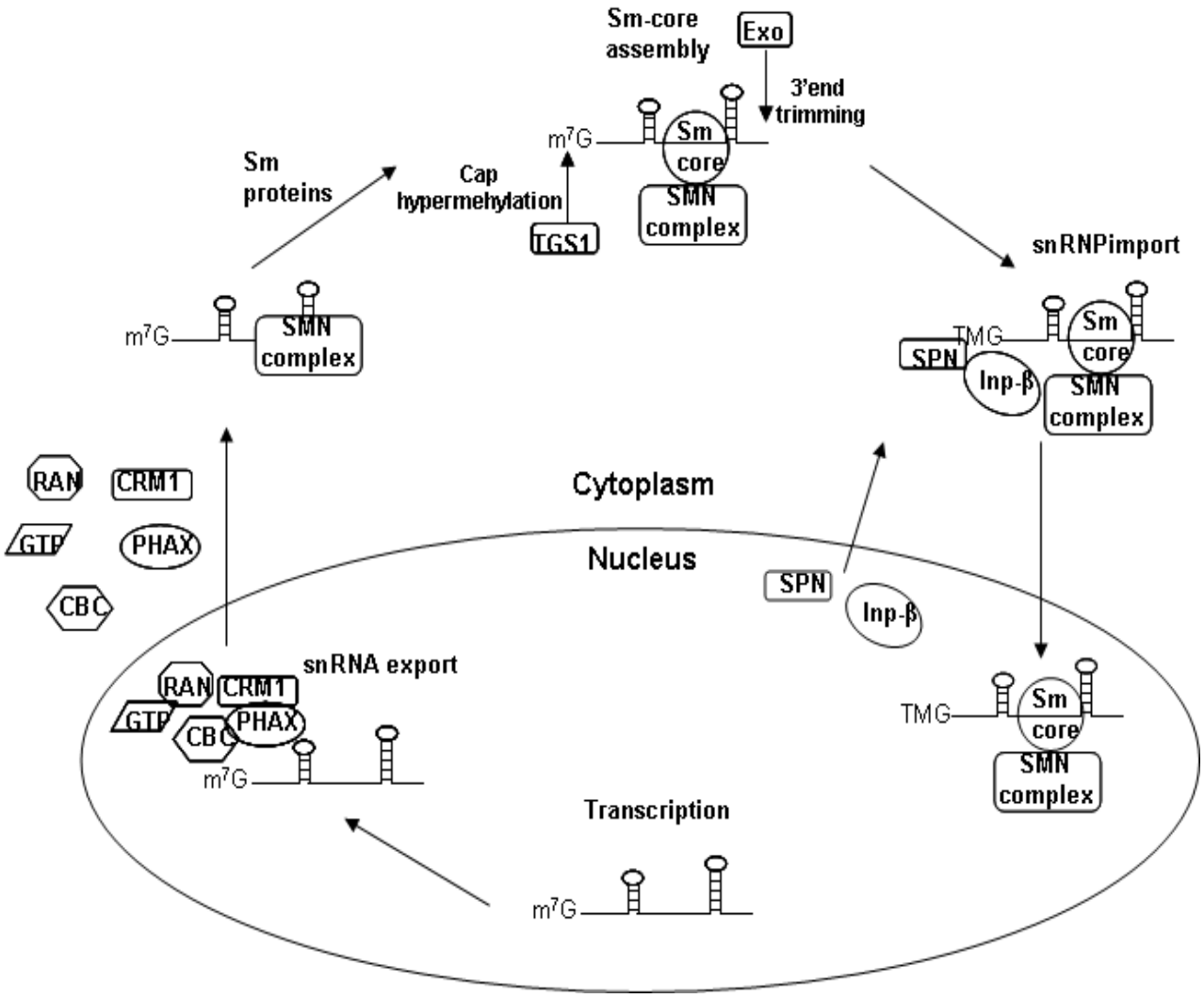

Fig. 7 SnRNA export system. Following transcription by Pol II, pre-snRNA are exported to the cytoplasm.. The snRNA-export complex consist of export adaptor protein (PHAX), the export receptor chromosome region maintainance (CRM1), the cap-binding complex (CBC) and the GTP-bound form of Ran GTPase. These factor dissociates in the cytoplasm. The survival of motor neuron (SMN) recruits the Sm-proteins to form the Sm-core RNP. Following assembly on the Sm core, the 7-methylguanosine (m^7G) cap is hypermethylated by trimethylguanosine syntase-1 (TGS1) to form 2,2,7-trimethylguanosine (TMG) cap structure and the 3' end is trimmed by an unidentified exonuclease (Exo). TMG and associated Sm-proteins provide a targeting signal for re-import into the nucleus by the import adaptor snurportin-1 (SPN) and the import receptor importin-β (Imp-β).

Cell cycle regulation of snRNA gene transcription.

The demand for snRNA and other non coding RNAs depends upon the metabolic state of the cell and therefore their transcription is regulated during changes in cells growth and cells cycle progression. Recent evidence suggests that transcription of snRNA genes is restricted at various points during the cell cycle by the protein kinase CK2 and the RB 300 tumor suppressor protein. Currently, the cell cycle regulation of RNA polymerase III transcription is better understood; RNA polymerase III transcription is most active during the late G1, S and G2 phases of the cell cycle and is repressed during the M and G0/early G1 phases. The protein kinase CK2 is suggested to play a major role in M-phase repression of snRNA transcription. CK2 associates with multiple components of the polymerase III transcriptional apparatus, including the polymerase itself, and SNAPC, and phosphorylates both the Bdp1 component of the Brf2–TFIIIB complex (Hu et al., 2004) and the SNAP190 component of SNAPC (Gu et al., 2005) reducing their promoter association. RB tumor suppressor protein silences RNA polymerase III transcription at the G0/early G1 stage the cell cycle. Interestingly, a role for RB family members for the regulation of snRNA gene transcription by RNA polymerase II has not been observed (Hirsch et al., 2004), suggesting that the distinct usage of SNAPC and Brf2–TFIIIB specifies RB targeting of RNA polymerase III-transcribed snRNA genes. As the products of these non-coding RNA genes contribute substantially to the biosynthetic capacity of the cell, the repression of snRNA gene transcription and other RNA polymerase III transcripts by RB likely plays an important role in growth control limiting tumor progression. During U6 repression, RB stably associates with the promoter via protein–protein interactions with components of SNAPC and Brf2–TFIIIB. Thus, during snRNA gene repression, RB likely inhibits steps subsequent to RNA polymerase III recruitment, potentially including promoter escape, open complex formation, elongation, and termination.

snRNAs function: pre-mRNA splicing.

Pre-messenger RNA splicing refers to a process catalyzed by the spliceosome, in which non coding intron sequences are excised and coding exons are ligated together through a two-step reaction to form mature mRNA. In the first of the two nucleophilic reactions, the 2'-OH of an intronic "branch point" adenosine attacks the phosphodiester backbone at the 5'-splice site, creating a branched lariat intermediate and a free 3'-OH on the 5' exon. This hydroxyl then attacks the 3'-splice site, ligating the exons and releasing the intron in the lariat form. The chemistry of these two trans-esterifications reactions is relatively simple; however pre-mRNA splicing requires accuracy and precise regulation and is therefore catalyzed by the spliceosome, one of the largest and most complex molecular machines in the cell.

The spliceosome is composed of over 200 different proteins and five RNA components (U1, U2, U4, U5, U6 snRNAs) that form a dynamic and elaborate network of RNA-RNA, RNA-protein and protein-protein interactions (Nilsen, 2003).

The snRNPs have at least three important functions: recognition of splicing sites and branchpoint sequences, folding of pre-mRNA into a reactive structure and alignment of splice sites and finally possibly direct contribution to catalysis.

In the early stages of spliceosome assembly the 5' splice site base pairs with U1 snRNA; U1 base pairing is well established in both yeast and mammalian splicing system. U2 snRNP recognizes the branch point adenosine, and a short intermolecular helix forms between U2 snRNA and the consensus sequence within the intron . This U2-intron duplex serves to position the branch point adenosine for its role as the nucleophile during the first catalytic step. The U4/U6.U5 tri-snRNP next joins the spliceosome and U1 at the 5' splice site is exchanged for U5 snRNA on the exon side of the splice site.

When tri-snRNP joins the spliceosome U6 snRNA, after U4 snRNA dissociation, is also exchanged for U1 but on the intro side of the 5' splice site and forms a base pairing interaction with the U2 snRNA that juxtaposes the 5' splice site and branch site, the reactant of the first of the two trans-esterification reactions. The highly conserved 5' end of the U2 snRNA is made up of several stem-loop and single stranded regions that interact with the U6 snRNA and the snRNP proteins. Finally the two exons are bound and kept in alignment partially via interactions with a conserved loop in U5 snRNA.

Promoter regulatory elements in *S. cerevisiae* Pol II-transcribed snRNA genes

Abstract

Small nuclear RNAs (U1 to U6) are abundant capped RNAs synthesized by RNA Polymerase (Pol) II, with the exception of the U6 snRNA, whose gene is transcribed by Pol III.

Even if the function of these RNAs has been studied in detail, little is known about their transcription. In particular, in Saccharomyces cerevisiae, only the Pol III U6 snRNA promoter has been characterized, while nothing is known about the promoter architecture of the other Pol II-transcribed snRNA genes.

To begin to characterize the promoter architecture of these RNA genes, we used comparative sequence analysis. The alignment of the orthologous sequences from four different species of Saccharomyces identified conserved elements upstream of the transcription start site.

Some evolutionarily conserved sequence blocks matched the consensus binding site of known general regulatory factors (Rap1, Abf1, Reb1) while others did not match any known motif. The analysis also revealed the conservation of a few regulatory elements known to be involved in different pathways (RRPE - ribosomal RNA processing element; PACE - Proteasome Associated Control Element).

To better understand the involvement of the identified putative promoter elements in transcription we prepared sequence-tagged version of two yeast snRNA genes, LRS1 (coding for U2 snRNA) and SNR7 (coding for U5 snRNA), to be used as reporter genes for in vivo expression analysis.

Mutational analysis of the upstream conserved motifs revealed their influence on snRNA gene transcription, possibly involving the generation or maintenance of nucleosome-free promoter regions.

Introduction

Small nuclear RNA genes (snRNAs) are a small group of highly abundant, essential, non-polyadenylated, non-protein-coding transcripts that function in the nucleoplasm. They assemble with numerous protein factors to form small nuclear ribonucleoproteins (snRNPs), component of the spliceosome, the complex that catalyzes the splicing of pre-mRNA.

All are synthesized by RNA Polymerase (Pol) II, with the exception of the U6 snRNA, whose gene is transcribed by Pol III. Pol II-transcribed snRNA genes U1, U2, U4, U5 (Sm class) snRNAs share common features: a 5′-trimethylguanosine cap, a 3′ stem–loop and a binding site (the Sm site) for a group of seven Sm proteins. (for a review see Matera et al., 2007).

As compared to the metazoan snRNAs, little is known about snRNA synthesis in unicellular organism, even if their function has been studied in detail.

Human snRNA genes contain compact promoters that are recognized by increasingly well-characterized transcription factors. Pol II and Pol III snRNA promoters share common elements: a loosely conserved proximal sequence element (PSE), located at about position -55 upstream the transcription start site, which recruit the snRNA activating protein complex (SNAPc or PTF) and define the transcription initiation site, and a distal sequence element (DSE) located around position -200 that function as enhancer. This motif usually contains an octamer sequence that recruit the POU domain transcription factor Oct-1, as well as an SPH element that recruits the zinc finger transcription factor Staf. Core Pol III snRNA promoters contain in addition a TATA box downstream of the PSE. Both PSE and DSE can be interchanged between Pol II and Pol III promoters with no effect on RNA polymerase specificity, that is determined only by the presence or absence of the TATA box. (Hernandez, 2001).

In *D. melanogaster* both Pol II and III promoters contain a quite conserved 21 bp element, called the PSEA, located at about 42 bp upstream the start site. A PSEB or a canonical TATA box are located downstream the PSEA in Pol II and Pol III snRNAs respectively (Zamrod et al., 1993). Polymerase specificity is determined by position 19 and 20 of the PSEA, g/aG in Pol II and TC in Pol III promoters (Jensen, 1998).

Plant snRNA gene promoters consist of an upstream sequence element (USE) and TATA box at -30. The major determinant of RNA polymerase specificity is the spacing between the USE and the TATA box (Filipowicz et al., 1990).

Transcription of Schizosaccharomyces pombe U2 gene is directed by two essential promoter elements: spUSE centered at -55, which functions as an activator element and a TATA box at -26 (Zhou and Lobo-Ruppert, 2001).

In *S. cerevisiae* only the Pol III U6 snRNA transcription has been studied in detail. The promoter contains a TATA box, and two other elements are required for correct efficient transcription: an A box located within the transcribed region and a B box located in the 3' flanking region, downstream of the termination signal (Eschenlauer , 1993).

The identification of regulatory elements within a promoter is a key step in understanding gene transcription. Even if these elements are short DNA sequences, often only 5 to 20 bp in length, they are critical for gene regulation. Cis-regulatory elements are commonly conserved across evolution. Comparative sequence analysis for regulatory element discovery is an high-throughput method to identify putative functional elements whose effective role in transcription can be verified by experimental procedures.

In this work we combined computational and experimental approaches to gain insight into transcriptional regulation of Pol II transcribed snRNA genes and their promoter architecture.

Materials and Methods

Comparative sequence analysis:

The coding sequences of *S. cerevisiae* snRNAs were recovered from SGD (Saccharomyces Genome Database). The homology search in the fungal genomes (*S. paradoxus* , *S. mikatae*, *S. bayanus* and *S. kudriazevii)* was carried out using BLASTN program (http://seq.yeastgenome.org/cgi-bin/blast-fungal.pl). As result of this search the orthologous snRNA coding sequences were identified. The snRNA gene 5'-flanking sequences of the different genomes were recovered by search in the NCBI databases (www.ncbi.nlm.nih.gov). The sequences (400 bp) upstream the transcription start site were then aligned with ClustalX program.

Amplification and cloning of DNA templates:

U5M-YEp352 template: We generated a tagged form of SNR7 (U5 snRNA gene) by inserting a 22-bp sequence tag into the transcribed region (position +130). We first generated two PCR products, from yeast genomic DNA (strain S288C), using the high fidelity Pfu DNA polymerase (Promega) and gene-specific pairs of oligonucleotide primers: U5_fw _SacI in combination with U5_rev_BamHI (carrying the oligo-tag underlined in table 1) to produce the upstream half of the construct, containing 130 bp of the transcribed sequence plus 248 bp of 5'-flanking region; U5_fw _BamHI in combination with U5_rev_SphI to produce the downstream half, containing 84 bp of the transcribed sequence plus 292 bp of 3'-flanking region. The upstream and downstream fragments were then digested with SacI/BamHI and BamHI/SphI respectively, and inserted in SacI/SphI-cut Yep352 vector through a single ligation reaction.

U5M 5'-mutant templates:

U5M 5'- mutants were obtained by PCR amplification using U5M-YEp352 as template and specific mutagenic forward primers. All the fragments were cloned in Yep352 vector (SmaI site) and sequence verified by dideoxy chain termination sequencing.

U2M-YEp352 template: A tagged form of LSR1 (U2 snRNA gene) was generated by inserting a 22-bp sequence tag into the transcribed region (position +387). We first generated two PCR products from yeast genomic DNA (strain S288C) using the high fidelity Pfu DNA polymerase (Promega) and gene-specific pairs of oligonucleotide primers: U2_SacI_fw in combination with U2_BamHI_rev (carrying the oligo-tag underlined in table 2) to produce the upstream half of the construct, containing 387 bp of the transcribed sequence plus 251 bp of 5'-flanking region; U2_BamHI_fw in combination with U2_SphI_rev to produce the downstream half containing 792 bp of the transcribed sequence plus 129 bp of 3'-flanking region. The fragments were digested with SacI/BamHI and BamHI/SphI respectively and inserted in SacI/SphI-cut Yep352 vector through a single ligation reaction.

U2M 5'- mutants templates:

U2M 5'- mutants were obtained by PCR amplification using U2M-YEp352 as template and specific mutagenic forward primers. All the fragments were cloned in Yep352 vector (SmaI site) and sequence verified by dideoxy chain termination sequencing.

All oligonucleotides sequences used for amplification and cloning of DNA templates are listed in tables 1and 2.

Primer name	Primer sequence
U5_fw _SacI	CCCGGAGCTCCTGCTTTACATCGATGACAAAGG
U5_rev_BamHI	GCGCGGATCCTCACTCGTAAGACTGCTCTATGGAGACAAC ACCCGGA
U5_fw _BamHI	GCGCGGATCCAACAGGTAAAGCTGTCCGTTAC
U5_rev _SphI	ACATGCATGCTTATGCAAATGCTCCTTTACGCG
U5M_+160_fw	AGTATTCTCATCACGATTAACG
U5M_Abf1_down_fw	AGTATTCTCAGATCGATTATTCAATATGAAAAAAAAAATTGA AAATTTTGTAG
U5M_poliA_down_fw	ATTGATCTTGACGTAGAAACGGAGTGCTCAGTA
U5M_RRPE_down_fw	ATTGAAAATTTTGTAGAAACGGAG
U5M_TATA_fw	TATAAAAAGCGCATAGTAAGACTT
U5M_ΔTATA_fw	CTTTTCTTTCTTTTTGTTTTAAAACC

Table 1 List of oligonucleotides used in amplification and cloning of U5M template and U5M 5'-mutant forms.

Primer name	Primer sequence
U2_SacI_fw	CCCGGAGCTCACACTTTCACTACGTGTATAACG
U2_BamHI_rev	GCGCGGATCCTCACTCGTAAGACTGCGACAGGGAAGAGTATGAAGC
U2_BamHI_fw	GCGCGGATCCTTTCCGAGCCGTTTATGTCC
U2_SphI_rev	ACATGCATGCCCAATTAGTGCACACACATAC
U2_Rap_fw	GGGGTGTATGGGTGTGGGTGG
U2_Rpn4_fw	GGTGGCAAAAAAAACCTAGCAAC
U2_poliT_fw	CCGGAGCTCGCAACGCTCTATGTTTCTTTTC
U2_+127_fw	CCGTTTCCGATGGGCCACTCG

Table 2 List of oligonucleotides used in amplification and cloning of U2M template and U2M 5'-mutant forms.

Yeast strains:

Strain	Genotype	Source
BY4742	Matα, his3-Δ1, leu2-Δ0, lys2-Δ0, ura3-Δ0	Yeast Knock Out Collection (Open Biosystem)
BY4741	Matα, his3-Δ1, leu2-Δ0, met15-Δ0, ura3-Δ0	Yeast TAP Fusion Collection (Open Biosystem)
ABF1-TAP	Matα, his3-Δ1, leu2-Δ0, met15-Δ0, ura3-Δ0, ABF1-TAP-HIS3MX6	Yeast TAP Fusion Collection (Open Biosystem)
STB3-TAP	Matα, his3-Δ1, leu2-Δ0, met15-Δ0, ura3-Δ0, STB3-TAP-HIS3MX6	Yeast TAP Fusion Collection (Open Biosystem)
RAP1-TAP	Matα, his3-Δ1, leu2-Δ0, met15-Δ0, ura3-Δ0, RAP1-TAP-HIS3MX6	Yeast TAP Fusion Collection (Open Biosystem)
RPN4-TAP	Matα, his3-Δ1, leu2-Δ0, met15-Δ0, ura3-Δ0, RPN4-TAP-HIS3MX6	Yeast TAP Fusion Collection (Open Biosystem)

Table 3 Yeast strains used in this work.

***In vivo* RNA analyses:**

U5M-YEp352 transformants: Yeast cells (BY4742 strain) were transformed with the different U5M-YEp352 templates and high-copy number plasmid YEp352 by the lithium acetate procedure (Ito et al., 1983), and resulting transformants were selected for uracil auxotrophy. RNA extraction was performed according to a previously described procedure (Schmitt et al., 1990). Total RNA samples (10 µg) were fractionated on a 6% polyacrylamide/7 M urea gel, then transferred to a positively charged nylon membrane (Gene Screen Plus, Perkin Elmer). The filter was hybridised with a 5'-labeled probe (5'-GGATCCTCACTCGTAAGACTGC) complementary to the oligonucleotide inserted as previously described in the coding region of SNR7 gene. Hybridization was carried out overnight at 28°C in 5X SSC, 5X Denhardt's solution , 0,1 mg/ml denatured salmon sperm DNA, 0,5 % (w/v) SDS, followed by one 10-minute washing with 2X SSC solution containing 0,1% SDS and a short washing in 1X SSC solution containing 0,1 % SDS. Hybridization products were visualized and quantified by phosphorimaging. A $tRNA^{Ala}$ specific primer (Ala_AGC_probe: 5'-GGAGACCTCTCCCATGCTAAGGGAGCGCGC) was used for normalization.

U2M-YEp352 transformants: Yeast cells (BY4742 strain) were transformed with the different U2M-YEp352 templates and high-copy number plasmid YEp352 by the lithium acetate procedure (Ito et al., 1983), and resulting transformants were selected for uracil auxotrophy. RNA extraction was performed according to a previously described procedure (Schmitt et al., 1990). Total RNA samples (10 µg) were fractionated on a 1,2% agarose, 1,9% formaldehyde gel, then transferred to a positively charged nylon membrane (Gene Screen Plus, Perkin Elmer). The filter was hybridised with a 5'-labeled probe (5'-GGATCCTCACTCGTAAGACTGC) complementary to the oligonucleotide inserted as previously described in the coding region of LSR1 gene. Hybridization was carried out overnight at 28°C in 5X SSC, 5X Denhardt's solution , 0,1 mg/ml denatured salmon sperm DNA, 0,5 % (w/v) SDS, followed by one 10-minute washing with 2X SSC solution containing 0,1% SDS and a short washing in 1X SSC solution containing 0,1 % SDS. Hybridization products were visualized and quantified by phosphorimaging. A $tRNA^{Ala}$ specific primer (Ala_AGC_probe: 5'-GGAGACCTCTCCCATGCTAAGGGAGCGCGC) was used for normalization.

Chromatin immunoprecipitation:

Cross-linked chromatin was prepared essentially as described (Kuras, 2004; Kuras and Struhl, 1999). Yeast strains expressing tandem affinity purification (TAP) protein-tagged version of Abf1, Stb3, Rap1, Rpn4 proteins were from the Yeast TAP Fusion Collection (Open Biosystem) (Ghaemmaghami et al., 2003). BY4741 (table 3) was used as nontagged control strain. Yeast cultures (200 ml) were grown exponentially in glucose-containing medium to OD_{600} = 0,5. Cultures were treated with formaldehyde (1% final concentration) for 10 minutes at room temperature, with occasional swirling, and then quenched with glycine (240 mM final concentration) for 5 minutes at room temperature. Cells were collected, washed once with cold TBS (20 mM Tris-HCl, pH 7.5; 150 mM NaCl), once with FA-lysis buffer (50mM HEPES-KOH, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0,1% sodium dioxycholate, 0,1% SDS, 1 mM PMSF) and resuspended in 1ml of FA-lysis buffer containing 0.5% SDS. An equal volume of glass beads (diameter 0,5 mm) was added and the cells were disrupted with vortexing (15 min, 4 °C). Glass beads were removed and samples were diluited into 8 ml of FA-lysis buffer. Chromatin was pelleted by centrifugation for 20 minutes at 20000g, washed twice with 1,5 ml of lysis buffer for 1 h at 4 °C and sonicated to reduce DNA fragments to an average size of 300bp. Sonicated samples were diluited to a volume of 4 ml, centrifugated for 20 min at 20000g and the supernatant was transferred to clean tubes in 800 µl aliquots. Before proceeding with immunoprecipitation , 400 µl of solution were put aside and marked as input samples. For immunoprecipitation of TAP-tagged proteins chromatin solution was incubated over night at 4°C with IgG-Sepharose (GE Healthcare). Immunoprecipitated DNA was purified as described (Kuras and Struhl, 1999). Beads were washed twice for 4 min in 1,4 ml FA-lysis buffer with 275 mM NaCl, twice in 1,4 ml FA-lysis buffer with 500 mM NaCl, once in 1,4 ml of 10 mM Tris-HCl, pH 8.0, 0,25M LiCl, 1mM EDTA, 0,5% N-P40, 0,5% sodium deoxycholate, and once in 1,4 ml TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). DNA was eluted by heating the beads for 10 minutes at 65°C in 400 µl of elution buffer (50 mM Tris–HCl, pH 7.5, 10 mM EDTA, 1% SDS). To reverse crosslinks, pronase (0,8 mg/ml) was added and samples were incubated for 1 h at 42 °C and for 5h at 65 °C. After extraction with phenol-chloroform-isoamyl alchol and chloroform, DNA was precipitated with ethanol-sodium acetate in the presence of 20 µg of glycogen, and resuspended in TE buffer.

DNA Amplification:

DNA samples were amplified by PCR using GoTaq polymerase (Promega) and specific primer pairs (Table 4). Typically 1/100 of the immunoprecipitated DNA and 1/30000 of the total DNA Input were used. The specific primers for the intergenic region ARS540 of Chr V (Intergenic_V1; Intergenic_V2bis) were used as internal control. Reactions were carried out in 10 µl and contained 0,25 µM each primer, 0,1 mM dNTPs, 0,06 mCi/ml of α^{32}P-dCTP (Perkin Elmer). PCR products were fractionated on a 6% polyacrylamide gel, visualized and quantified by phosphorimaging. The fraction of immunoprecipitated material for a specific fragment was calculated by ratio of immunoprecipitated DNA over total DNA. Control chromatin immunoprecipitation experiments were performed with the untagged BY4741 strain.

The primers used in DNA amplification are listed in table 4.

Primer name	Primer sequence
U2_rap_rpn4_fw	ACACTTTCACTACGTGTATAACG
U2_rap_rpn4_rev	CGAGTGGCCCATCGGAAACG
U5_Abf1_fw	TAACTTCCTATTTGAGTTCGTGG
U5_Abf1_rev	CATTTAACAAAAAGTCTTACTATGC
Intergenic_V1	GGCTGTCAGAATATGGGGCCGTA
Intergenic_V2bis	GACCCGAGGGTATGGTTTTCACAAG

Table 4 List of oligonucleotides used as primers for the PCR on immunoprecipitated chromatin. The amplicon length was set to 144 bp and 127bp for gene-specific primers U2 and U5, respectively, and to 112bp for intergenic control region.

Primer extension:

Total RNA was purified from yeast cells transformed with U5M_YEp352 according to a previously described procedure (Schmitt et al., 1990). Reverse transcription of the isolated RNA was carried out using Superscript III reverse transcriptase (Invitrogen), following the manufacturer's protocol by annealing with a specific labeled probe (5'-GGATCCTCACTCGTAAGACTGC) complementary to the oligonucleotide inserted as previously described in the coding region of SNR7 gene. The reactions were performed at 55 °C for 1h and contained 3 µg of total RNA, 2 pmol 5' labeled probe, 1X Superscript III buffer, 1 unit/µl of SUPERase-In (Ambion), 200 unit/µl of Superscript III and RNase-free water to a final volume of 50 µl. The enzyme was heat-inactivated for 15 minutes at 70°C. The extension products were precipitated with ethanol/ammonium acetate, separated by electrophoresis on 7% polyacrylamide/7 M urea gel and detected by phosphorimaging. The 5' end was mapped by direct comparison with dideoxy chain termination sequencing reactions (Kit Thermo-Sequenase, GE Healthcare) run on the same gel. Total RNA from yeast cells transformed with U5M_ΔTATA_YEp352 was used as a negative control.

Results and discussion

Phylogenetic Footprinting of snRNA genes promoters in *Saccharomyces*.

To begin to define the Pol II-transcribed snRNA promoters architecture in *Saccharomyces cerevisiae,* we used a comparative analysis approach to identify putative functional elements upstream of snRNA genes. Because functional sequences are maintained in evolution, these elements should stand out by the characteristic of being highly conserved across related species. This kind of approach, in fact, has been used recently with success (Cliften et al., 2003) (Kellis et al., 2003) (McCutcheon and Eddy, 2003). We selected for the analysis members of the *Saccharomyces sensu strictu* group, *S. paradoxus* , *S. mikatae*, *S. bayanus* and *S. kudriazevii*, because these species, even if closely related to *S. cerevisiae*, have sufficient sequence divergence to allow identification of motif with high degree of conservation by a simple sequence alignment.

Through an homology search in the fungal genomes in the Saccharomyces Genome Database (see methods and materials) we performed ClustalX-multiple alignment of the region (400 bp) upstream the start site of Pol II-transcribed snRNA genes. The phylogenetic footprints obtained are shown in fig. 1 (U1 snRNA), fig. 2 (U2 snRNA), fig. 3 (U4 snRNA) and fig. 4 (U5 snRNA).

All *S. cerevisiae* snRNA genes are characterized by the presence of a highly conserved TATA box ($TATAAA^{T}/_{A}$) in canonical position (-95;-90 upstream the start site) that matches with the consensus obtained by Basehoar et al., 2004.

Apart from the TATA box the analysis identified conserved sequence patterns in *Saccharomyces* snRNA promoters; some evolutionary conserved elements matched the consensus binding site of known transcriptional proteins while others did not match any known regulatory motif.

The most strikingly conserved region in the **U1 snRNA** promoter (figure 1) is located 157-149 base pair upstream of the transcriptional start site; the motif exactly match the Reb1p binding sequence (CGGGTAA or TTACCCG). Reb1p is an essential DNA binding protein that has been implicated in the activation of transcription by polymerase II, in the termination of transcription by Pol I and in the organization of nucleosomes.

A T-rich sequence is located 10 bp downstream of the Reb1p site; a combined effect of these two elements in nucleosome positioning has been observed previously in yeast promoters (Angermayr, Oechsner et al.,2003).

Sequence alignment also revealed a conserved motif AAAtCCTC, whose location just upstream the transcription start site could suggest an involvement in start site selection (Kuehner and Brow, 2006)

The alignment of **U2 snRNA** promoter regions revealed a conserved motif at position -206, -193 whose sequence conforms to the Rap1p binding site. The Repressor Activator Protein RAP1 is well known for its involvement in gene activation and repression, telomere structure, function and replication. In addition to these roles, this protein can also participate in the formation of boundary elements, stimulate meiotic recombination and transcriptional activation by opening chromatin (Morse, 2000).

Close to the Rap1p binding motif, the sequence GGTGGCAA stands out by his high conservation; the motif exactly matches the Rpn4p binding site, also known as PACE (Proteasome Associated Control Elements), a common motif in the promoters of proteasome genes and other several genes for factors involved in cell wall synthesis, protein folding, mRNA stability and processing (Mannhaupt et al., 1999).

The location of a poly(dT) sequence downstream of Rap1p site is conserved in Saccharomyces group. T-tracts are abundant genomic DNA elements that operate not by recruitment of specific transcription factors but rather by their intrinsic DNA rigid structure that might affect nucleosome stability and enhance accessibility to nearby sequences. A synergistic effect of Rap1p and T-rich elements in transcriptional activation been studied in rp (ribosomal protein) gene promoters (Goncalves et al., 1995).

An element located at position -127, showing an high degree of conservation, didn't match with any known transcriptional protein binding site.

The alignment identified also the conserved sequence TTAAATCCCC located just upstream the start site, whose location suggests involvement in transcription start site selection, as observed also in U1 snRNA promoter.

The alignment of **U4 snRNA** and **U5 snRNA** promoter regions revealed common motifs. An Abf1p binding motif is located at similar positions (-144 and -150 in U4 and U5 respectively) in the upstream regions of these genes and in both cases it is coupled with a poly(dT-dA) sequence (see fig.3 and 4).

The autonomously replicating sequence-binding factor 1 (ABF1) is known as a multifunctional DNA binding protein involved in transcriptional regulation, DNA-replication, chromatin remodelling and gene silencing. The promoter regions of numerous yeast genes contain ABF1 binding site and these genes are involved in diverse range of cellular functions, leading to the notion that Abf1p acts as a global transcriptional regulator (Miyake et al., 2004). Also, the large number of Abf1p consensus binding site in yeast genome argue for its global role in gene regulation. The conserved positions emerging from the alignments of U4 and U5 promoter regions match the Abf1p consensus TatCGTattgcaTGAT from Beinoraviciūte-Kellner (2005).

The combined presence of Abf1p binding site and poly(dT-dA) elements has been observed in many other yeast promoters; this protein, in fact, has a relatively weak transcriptional activation potential on its own but synergize strongly with other transcription factors. Direct observations show that ABF1 can remodel chromatin near its binding site, often requiring a downstream T-rich element to create a nucleosome-free region (Goncalves et al., 1995; Lascaris et al., 2000).

The comparative analysis in U5 snRNA promoters revealed a conserved motif at position -121 whose sequence exactly matches a motif known as RRPE (ribosomal RNA processing element), overrepresented in genes involved in ribosome biogenesis. Liko et al (2007) identified Stb3 as an RRPE binding protein that would mediate the inhibition of the transcriptional response to fresh glucose in RRPE containing genes.

The conserved element AAaACtCC upstream the start site in U4 snRNA has recently been identified as a functional initiator element by Kuehner and Brow (2006).

The conserved motifs identified by the comparative analysis of Pol II-transcribed snRNA promoters are summarized in **Table 5**.

snRNA gene	Conserved motif	Location	Putative binding factor
U1	ATTACCCG	-157	Reb1p
U1	AAAtCCTC	-8	
U2	GtgTaTGGGTGT	-206	Rap1p
U2	GGTGGCAAA	-192	Rpn4p
U2	TTTTTTTTTTTTT	-142	
U2	CGtTTCCGATGG	-127	
U2	TTAAATCCCC	-10	
U4	ATCGTGtaNAaTGA	-144	Abf1p
U4	TTTTTTt	-113	
U4	AAaACtCC	-8	
U5	ATCACNNTNaACGA	-150	Abf1p
U5	aaAAaAAA	- 131	
U5	TGAAAATTTT	-121	RRPE element

Table 5. Conserved motifs in snRNA gene promoter regions. Capital letters in sequences stand for highly conserved positions.

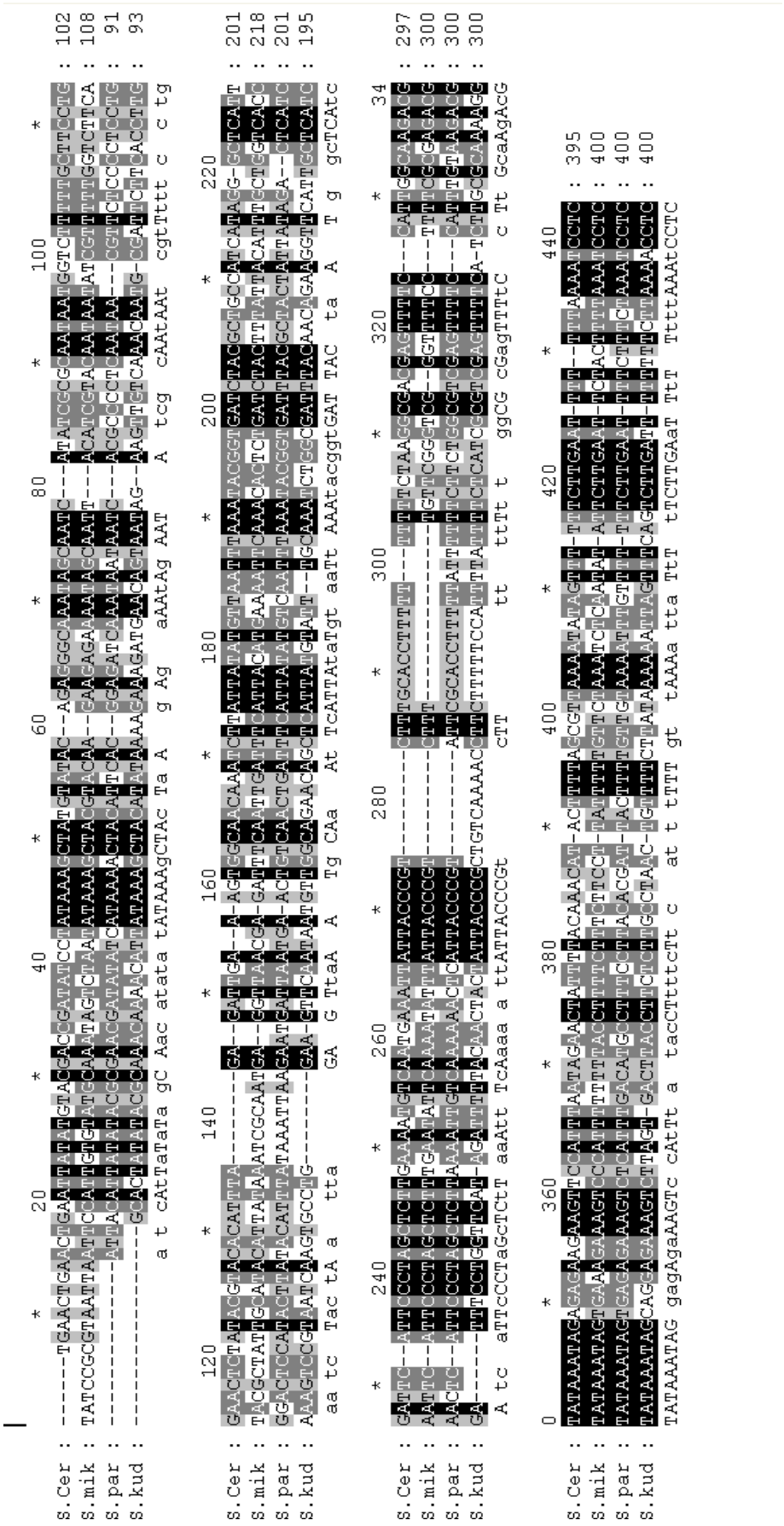

Figure.1 Four-way ClustalX alignment of 400 bp upstream of snR19 (U1 snRNA). Black boxes stand for sequence invariance across all four species: *S. cerevisiae* (S.cer), *S. mikatae* (S.mik), *S. paradoxus* (S.par), *S. kudriazevii* (S.kud).

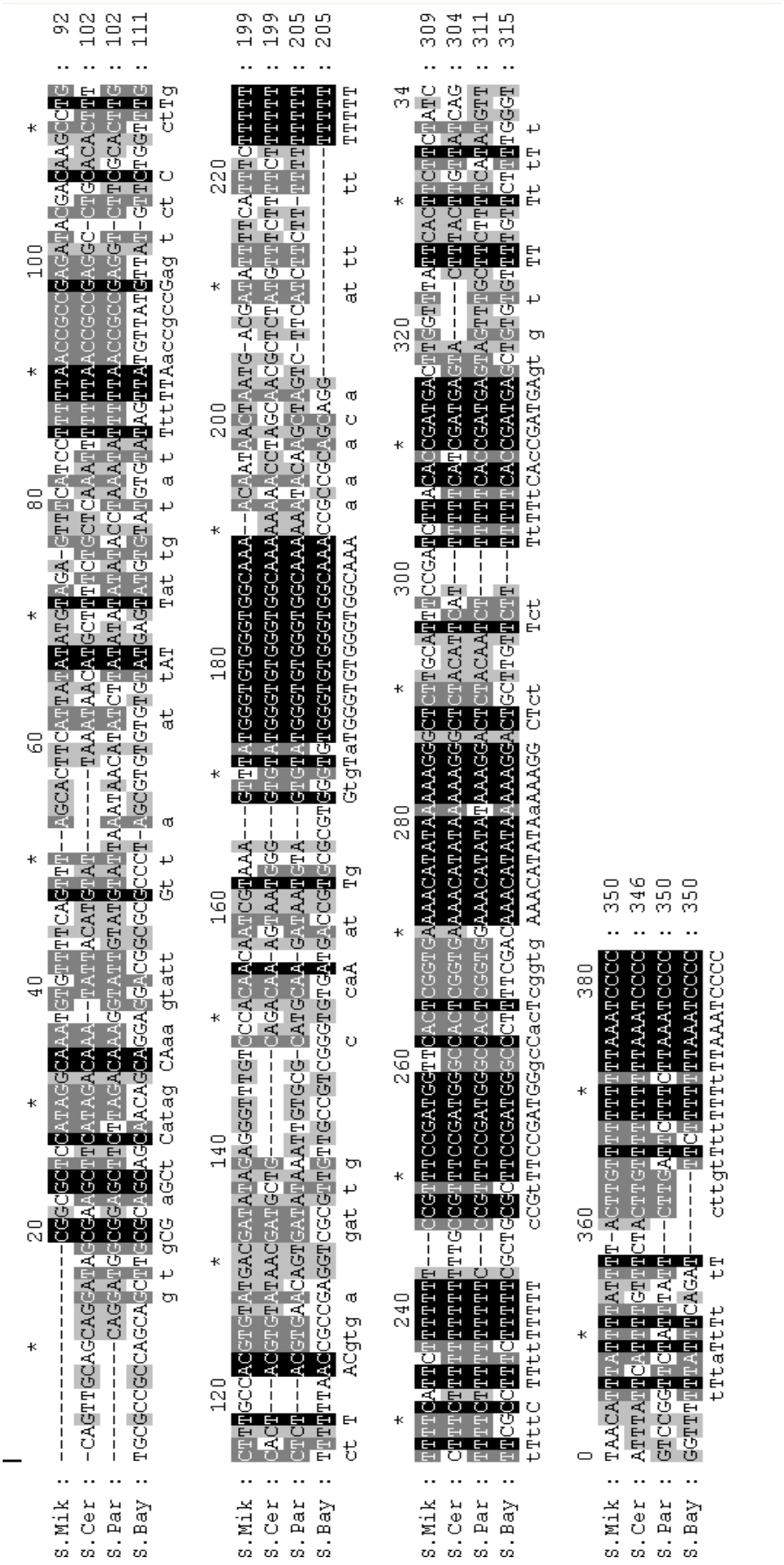

Figure.2 Four-way ClustalX alignment of 400 bp upstream of LSR1 (U2 snRNA). Black boxes stand for sequence invariance across all four species: *S. cerevisiae* (S.cer), *S. mikatae* (S.mik), *S. paradoxus* (S.par), *S. bayanus* (S.bay).

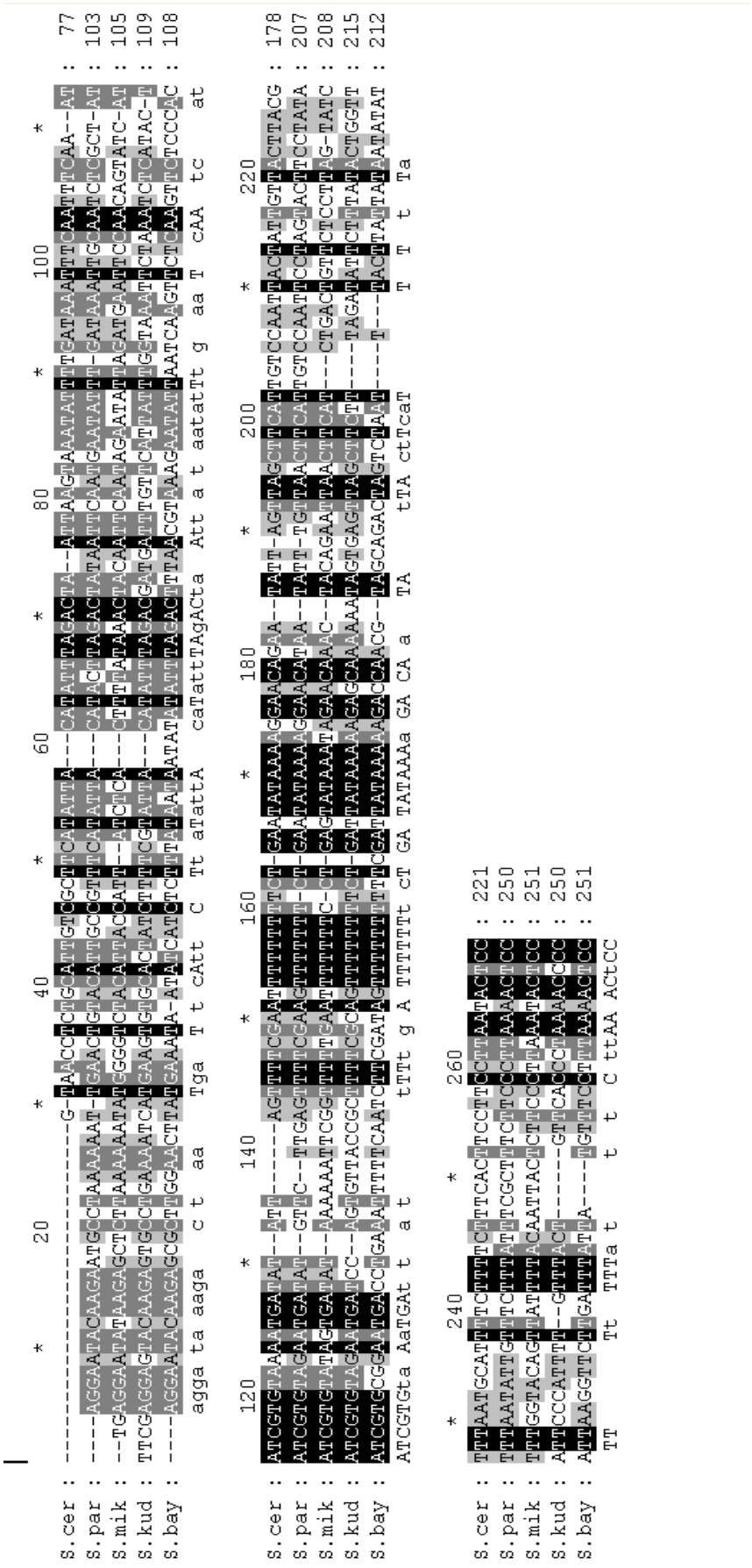

Fig. 3 Five-way ClustalX alignment of 400 bp upstream of SNR14 (U4 snRNA). Black boxes stand for sequence invariance across all four species: *S. cerevisiae* (S.cer), *S. mikatae* (S.mik), *S. paradoxus* (S.par), *S. bayanus* (S.bay), *S. kudriazevii* (S.kud).

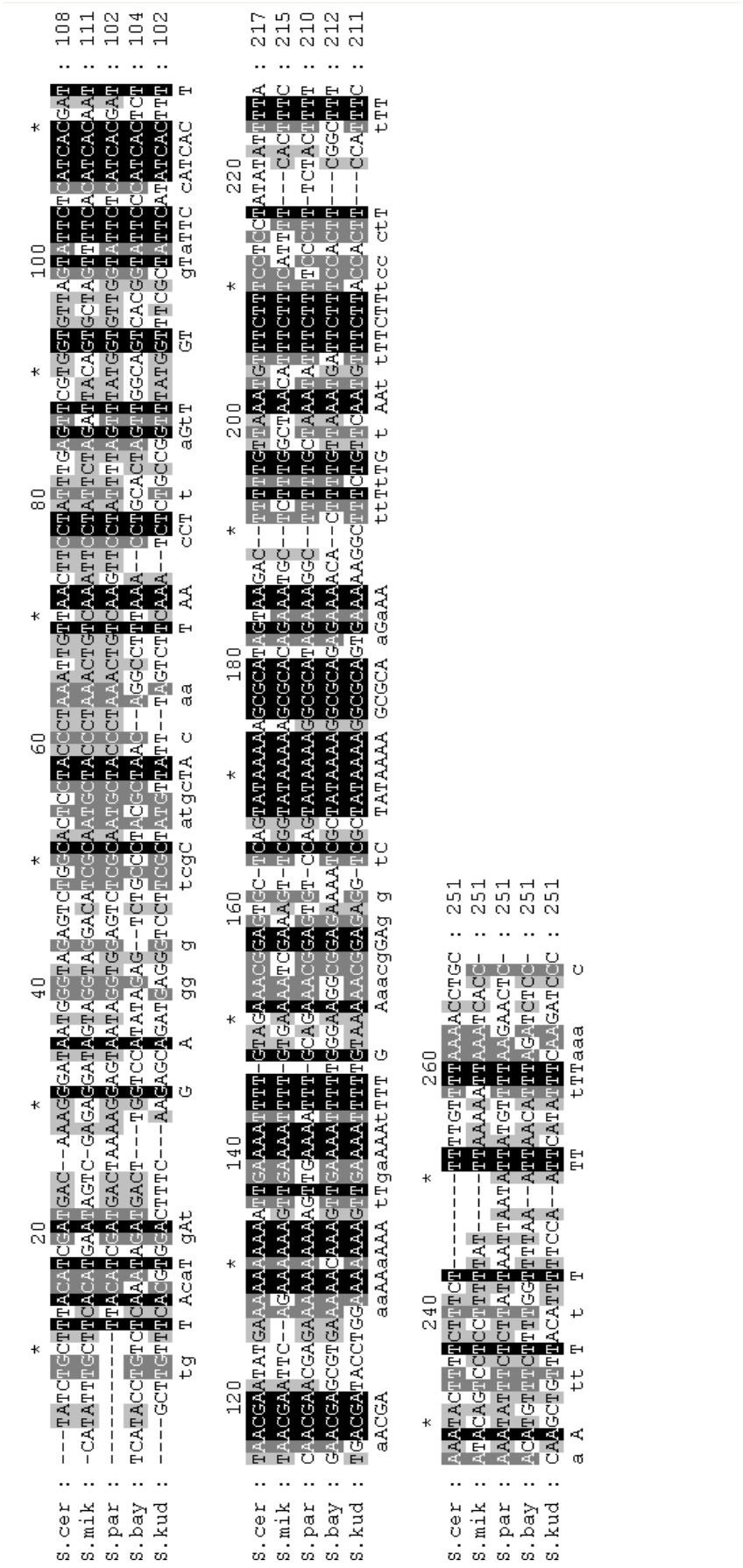

Fig. 4 Five-way ClustalX alignment of 400 bp upstream of SNR7 (U5 snRNA). Black boxes stand for sequence invariance across all four species: *S. cerevisiae* (S.cer), *S. mikatae* (S.mik), *S. paradoxus* (S.par), *S. bayanus* (S.bay), *S. kudriazevii* (S.kud).

Generation of LRS1 and SNR7 reporter genes for *in vivo* expression analysis.

To better understand the involvement of the identified putative promoter elements in transcription regulation we have chosen two snRNA genes LRS1 (coding for U2 snRNA) and SNR7 (coding for U5 snRNA), to be used as reporter genes for in vivo expression analysis. Because snRNA genes are essential, a strain deleted for one of these genes would be unable to survive.

In this work we prepared tagged versions of these genes, cloned into high-copy vector, to identify their expression and monitor the effects on transcription of changes in their promoter structures. We used as tag a 22 nucleotides sequence (5'-GCAGTCTTACGAGTGAGGATCC-3') whose BLAST analysis showed no significant similarities in the *S. cerevisiae* genome.

The oligonucleotide was inserted within the coding regions with the aim to obtain stable tagged-RNAs whose secondary structure were as similar as possible to RNAs transcribed from wild type genes.

The secondary structure of U5 snRNA is characterized by a highly conserved stem-loop (S1-L1) flanked by a moderately conserved internal loop (IL1); this 39-nucleotide domain contains all U5-specific sequences essential for splicing activity (Frank et al., 1994). We decided to insert the oligonucleotide at position +130 in the internal loop 2 (IL2) whose length is known to be quite variable in fungi and not required for U5 activity (Frank et al., 1994). The insertion of the tag within the U5 sequence is represented in fig 5.

U2 snRNA is 1179 nucleotides long, six times larger than its mammalian counterpart (188 nucleotides).; the 5' domain is highly conserved (120 nucleotides) and essential for viability while deletion of the central 945 nucleotides has no effect on growth rate (Shuster and Guthrie, 1988). Since only the secondary structure of the 5' end domain has been studied, we used the structure prediction programme *mfold* (http://mfold.bioinfo.rpi.edu/) to find the proper position for the tag. The secondary structure is mainly characterized by a long stem; we have decided to insert the tag in position +387 within a loop to reduce the disruption of base pairing (fig 5).

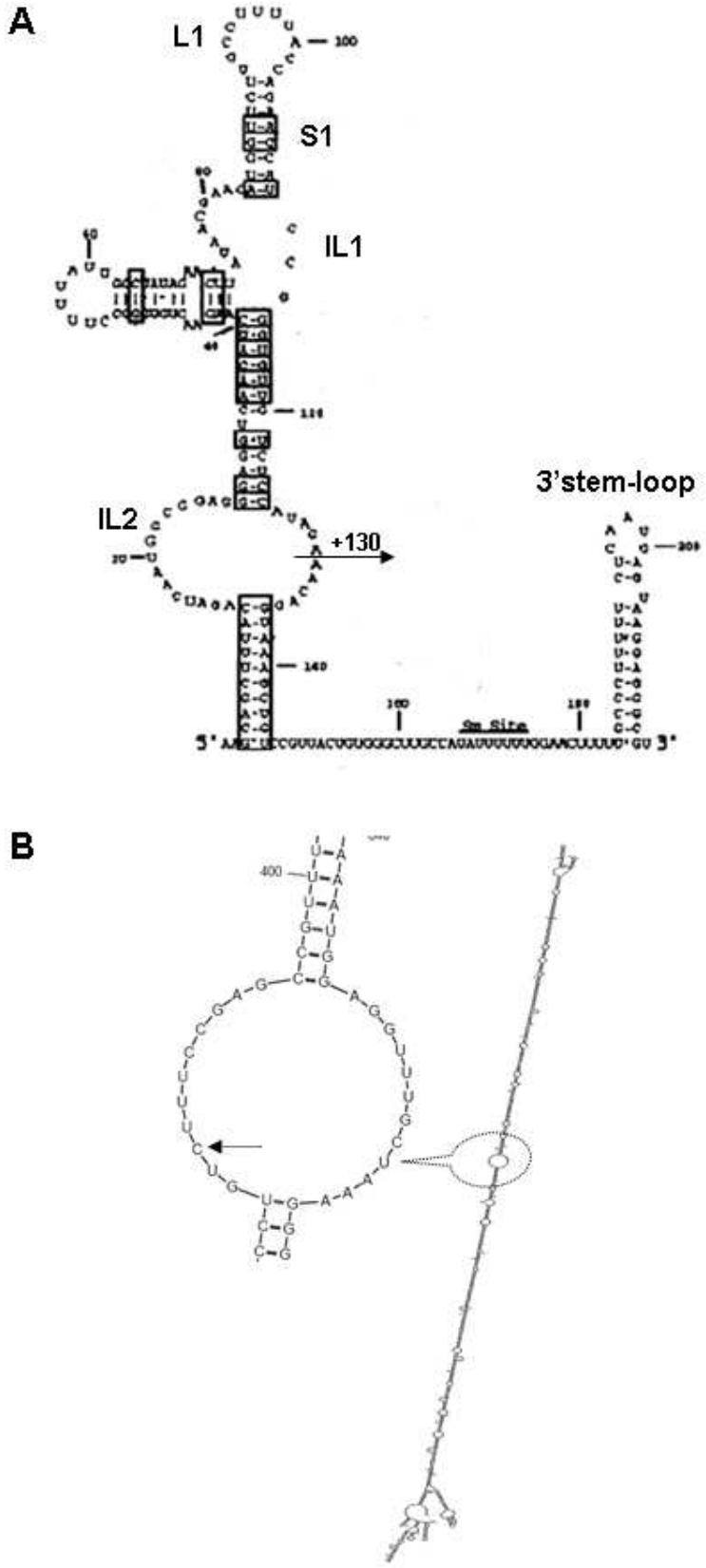

Fig. 5 (A) Current model of U5 snRNA secondary structure (Frank et al., 1994). An arrow specifies the position chosen for tag-insertion. **(B) Secondary structure of U2 snRNA from mfold.** The loop chosen for insertion is enlarged aside. An arrow specifies the tag-position.

Marked U5 snRNA and U2 snRNA genes produces stable RNAs.

The marked U5 snRNA gene (U5M) contains 248 bp of 5'-flanking region, a transcribed region of 236 bp rather than 214 bp as in the wild type and 314 bp of 3' flanking region. The marked U2 snRNA gene (U2M) contains 251 bp of 5'-flanking region, transcribed region of 1201 bp rather than 1179 bp as in the wild type and 129 bp of 3' flanking region(fig 6). Both templates were cloned into the Yep352 high-copy vector (see methods and materials for details).

To verify whether stable RNAs could be generated from such templates, we tested gene expression in vivo. Yeast cells (BY4741 strain) were transformed with U5M-YEp352, U2M-YEp352 and empty vector Yep352 as a control. Total RNA was extracted and probed with a radiolabeled oligonucleotide complementary to the inserted tag.

As shown in fig. 6, the probe detected correctly transcribed marked RNAs from cells transformed with U5M-YEp352 (lane 2,3) and U2M-YEp352 (lane 5,6) while no specific signal was observed in cells transformed with the empty vector (lane 1,4).

The analysis showed that the insertion of the tag didn't introduce significant alterations in RNAs secondary structure since templates generates stable RNAs of the expected length.

When cells were transformed with U5M templates, the specific probe detected two marked RNAs of different length. This result can be explained by the fact that U5 snRNAs of S. cerevisiae are expressed in two forms, U5 short (U5S) and U5 long (U5L), both products of the SNR7 gene. These forms differ for the presence of absence of a stem-loop at their 3' end due to alternative cleavage pathway of pre-U5RNA (see fig 5A) (Patterson and Guthrie, 1987; Chanfreau et al., 1997).

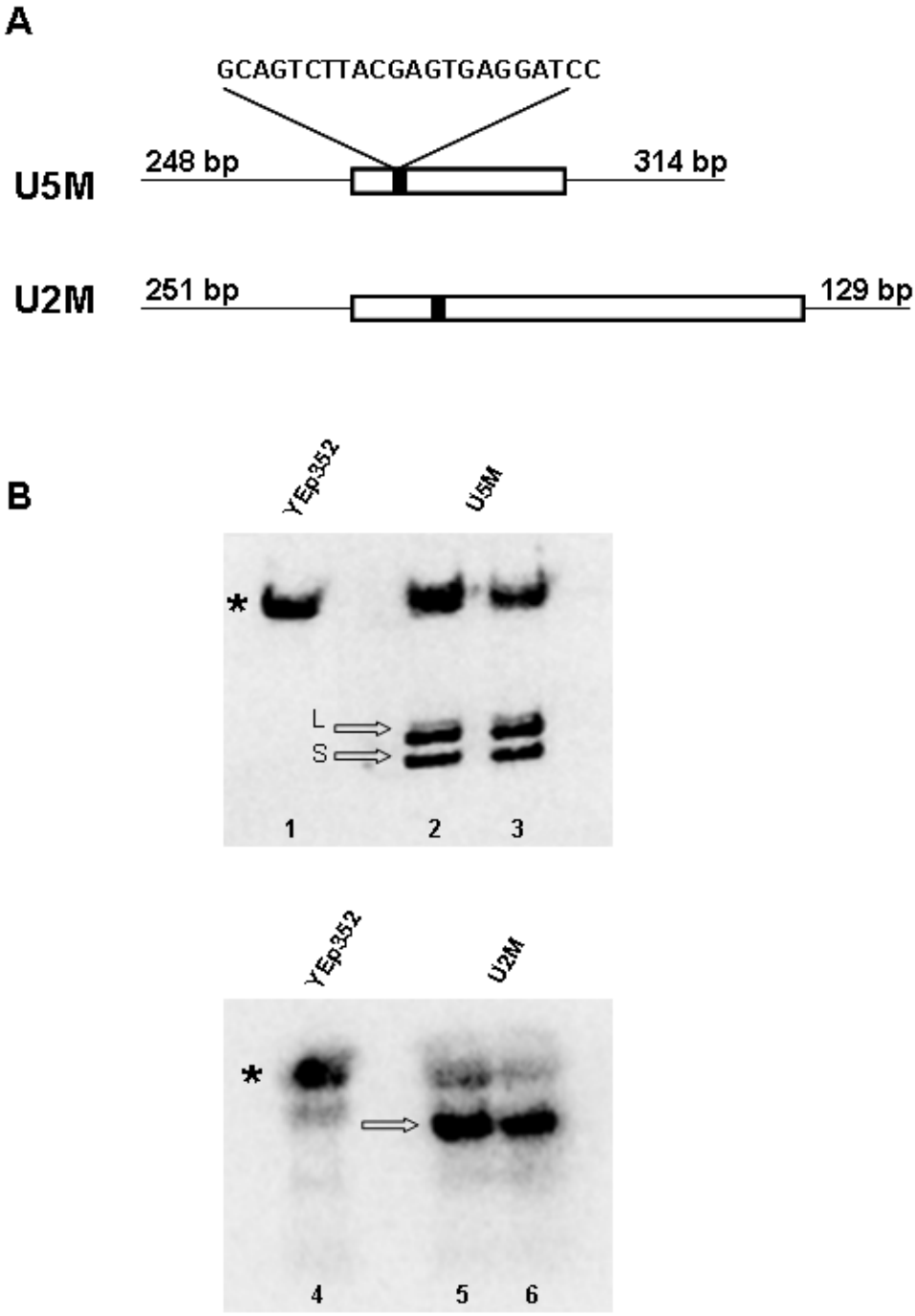

Fig. 6 (A) Schematic representation of U5M and U2M templates. Black boxes indicates the position of the tag within the coding region. **(B) In vivo expression of marked U5M and U2M templates.** Total RNA extracted from BY4741 transformed with the empty YEp352 vector (lane 1,4), U5M-YEp352 (lane 2,3) or U2M-YEp352 (lane 5,6) was gel fractionated and probed with a radiolabeled oligonucleotide complementary to the inserted tag. The migration position of the two form of U5M RNA, large and small (L and S) and of U2M RNA are indicated by an arrow. The asterisk indicates non-specific hybridization.

Marked U5 snRNA is correctly initiated in vivo.

Small nuclear gene transcription starts predominantly at a single position that corresponds to a unique mature 5' end. Since precise placement of the transcription start site may be required to ensure a proper expression, we questioned if marked RNAs were correctly initiated.

To determine the transcription start site of U5M RNA we used a primer extension assay. The reaction was performed on total RNA extracted from cells transformed with U5M-Yep352 template with a specific labeled probe complementary to the tag inserted as previously described in the coding region of SNR7 gene. As shown in fig 7B (lane 5) the analysis gave rise to one major signal corresponding to initiation position. The transcription initiation site was mapped by electrophoresing the primer extension products adjacent to dideoxy chain termination sequencing reactions performed with the same end-labeled oligonucleotide (Fig 7B, lane 1,2,3,4). Total RNA from yeast cells transformed with U5M_ΔTATA_YEp352, lacking all the upstream region upstream necessary to transcription (fig 7A), was used as a control(fig 7B, lane 6).

The result indicates that marked U5 RNA initiates transcription *in vivo* at the same A residue as wild type. Since we have found a single start site, the two specific signal found in the northern blot assay (fig 6B) have to differ in their 3' end as supposed.

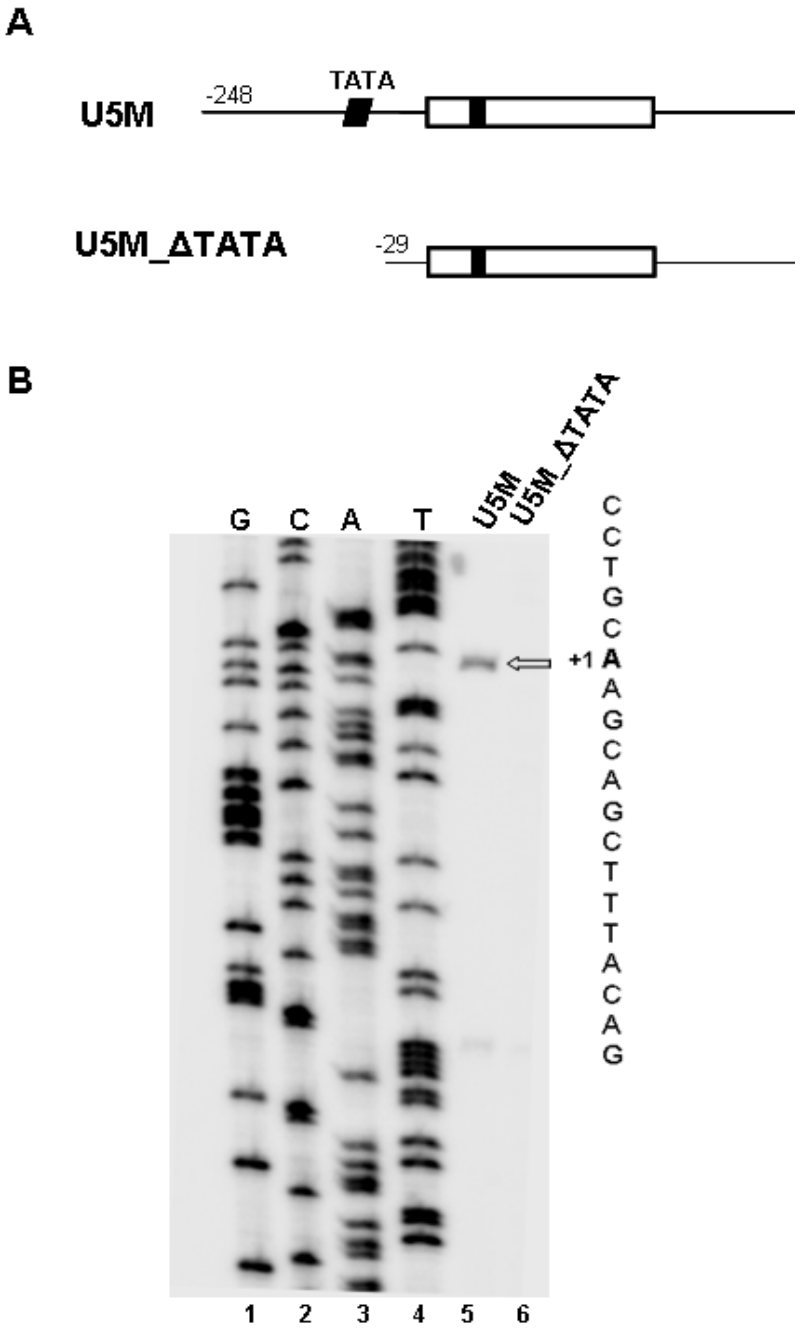

Fig. 7 (A) Schematic representation of U5M and U5M_ΔTATA templates. Black boxes indicates the position of the tag within the coding region. **(B) Primer extension.** Total RNA extracted from BY4741 transformed with U5M-Yep352 (lane 5) or U5M-ΔTATA-YEp352 (lane 6) was hybridized to [γ-^{32}P]5'-labelad probe complementary to the inserted tag. The adenosine residue at position +1(start site) is indicated by an arrow. On the right side, the sequence surrounding the U5 transcription start site is shown. For comparison, the U5M template was sequenced using the same end-labeled oligonucleotide (lane 1, 2, 3, 4).

SNR7 promoter region contains positive *cis*-acting elements.

To investigate in more detail the conserved elements found in U5 promoter region, we began to generate progressive deletions in 5' flanking region of U5M gene. A schematic representation of templates is shown in fig 8A.

Plasmids were transformed into BY4741 strain and the expression level was analyzed by northern blot (fig 8B).

Deletion of sequences upstream the Abf1 binding site caused a slight decrease in expression (fig 8B, lane 3), thus indicating that this region is required for optimal transcription, even if phylogenetic footprinting didn't reveal any conserved element. A significant drop in transcription was observed when the region containing the conserved elements ABF1 and poly(dA) were eliminated (compare lane 3 and 5).

To verify the role of ABF1 element, we introduced point mutations in conserved position of the Abf1p consensus motif (TCA→GAT; ACG→TTC); these mutations diminished transcription (compare lane 3 and 4) and deletion of poli(dA) element caused a further transcriptional decrease (compare lane 3, 4 and 5). Our results suggest that U5 promoter-driven expression is positively influenced by both these elements; their combined action has been already observed in many yeast promoter in relation to chromatin organization (Goncalves et al., 1995; Lascaris et al., 2000).

RRPE site directed mutagenesis (TGAAAATTTT→TGATCTTGAC) revealed a positive role of this element in transcription (compare lane 5 and 6); similar requirement of an RRPE for basal snoRNA gene transcription has been observed in our laboratory (Milena Preti, unpublished observations).

The presence of only a TATA box upstream of SNR7 ensures low, residual basal transcription (lane 7), that is abolished when the TATA box is also removed (lane 8).

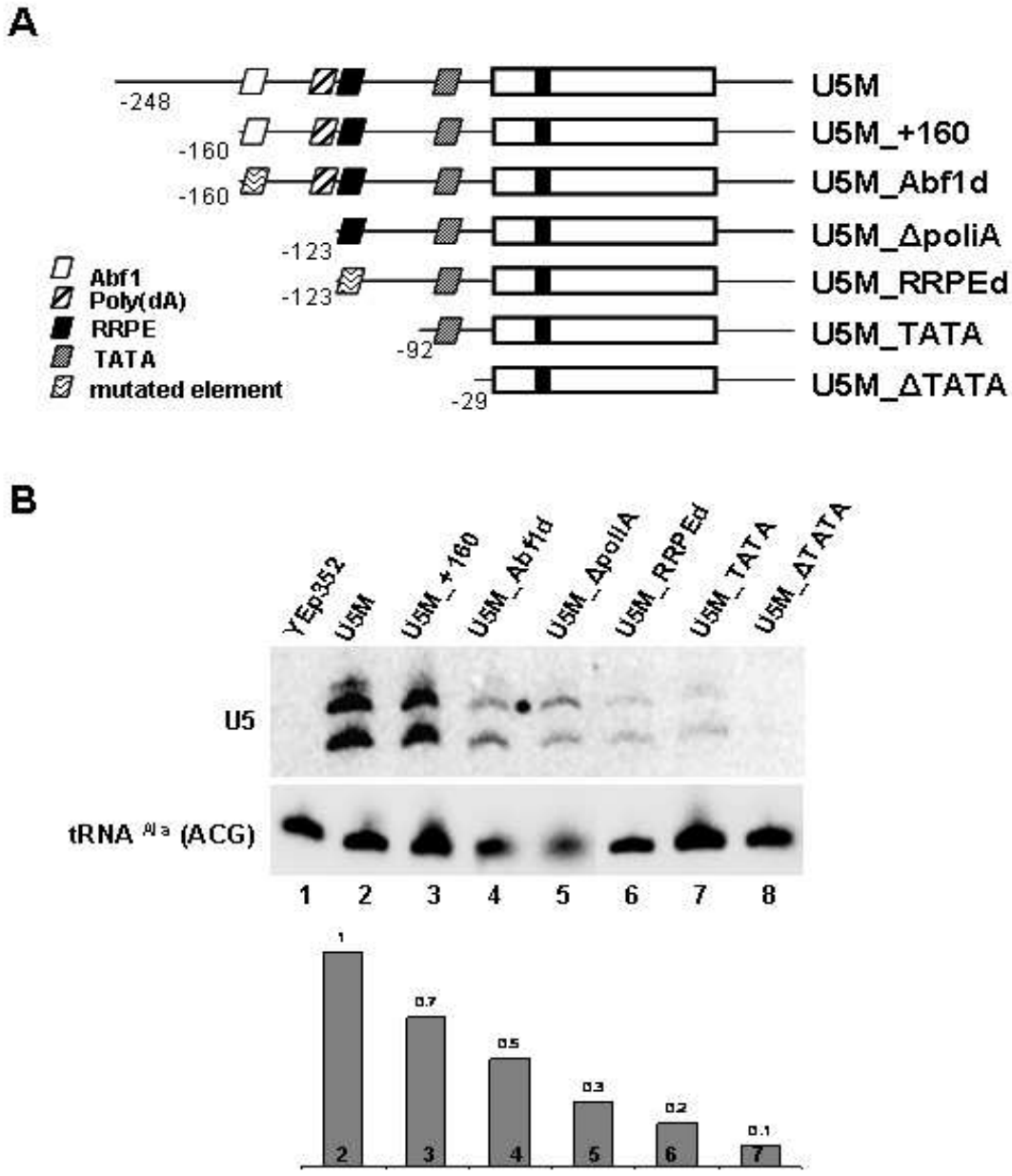

Fig. 8 (A) Schematic representation of U5M 5'-mutant templates. Black boxes indicates the position of the tag within the coding region. **(B) In vivo expression of U5M 5'-mutant templates.** Total RNA (10 µg) was extracted from BY4741 cells transformed with the empty YEp352 vector (lane 1) and U5M 5'-mutant templates.(lane 2-8) was gel fractionated and probed with a radiolabeled oligonucleotide complementary to the inserted tag. The same blot was also probed for $tRNA^{Ala}$ as an internal standard. The histogram represents RNA quantitative value after normalization with the standard.

Sequence upstream the TATA box are required for LRS1 transcription in vivo.

A preliminary series of U2M promoter deletion mutant was made to determine the minimal region required for transcription. A schematic representation of templates is shown in fig 8A.

Plasmids were transformed into the BY4741 strain and the expression level was analyzed by northern blot (fig 9B). Values were normalized with a probe specific for $tRNA^{Ala}$ used as an internal standard.

The deletion of 5'-flanking region upstream of the Rap1p binding site caused a reduction of transcription level (fig 9B; compare lanes 2,3); also observed for the U5 promoter, this region, even if does not contain any conserved motifs, seems to be important for optimal transcription.

When the 78 nucleotides containing RAP1, RPN4 and poly(dT) were removed, LRS1 transcription was abolished (lane 4); this results shows that one or more of these elements are necessary for transcription. At variance from what we observed in the case of the U5 snRNA gene, TATA box alone is not able to ensure basal transcription level. Similar results has been obtained in previous analysis of RP gene promoters, where Rap1p was found to be necessary and sufficient for TFIID recruitment (Mencia et al., 2002). The region upstream the TATA box contains strong *cis*-acting elements; RAP1 and T-tracks are in fact known for their positive role in transcription (Morse, 2000; Goncalves et al., 1995).

It would be interesting to investigate in more detail the role of each element, in particular the possible role of Rpn4, a protein involved in proteasome-ubyquitin system, to acquire information about the regulatory significance of this site.

A possible role in transcription of the conserved motif at position -127, that didn't match with any known protein, need also to be verified.

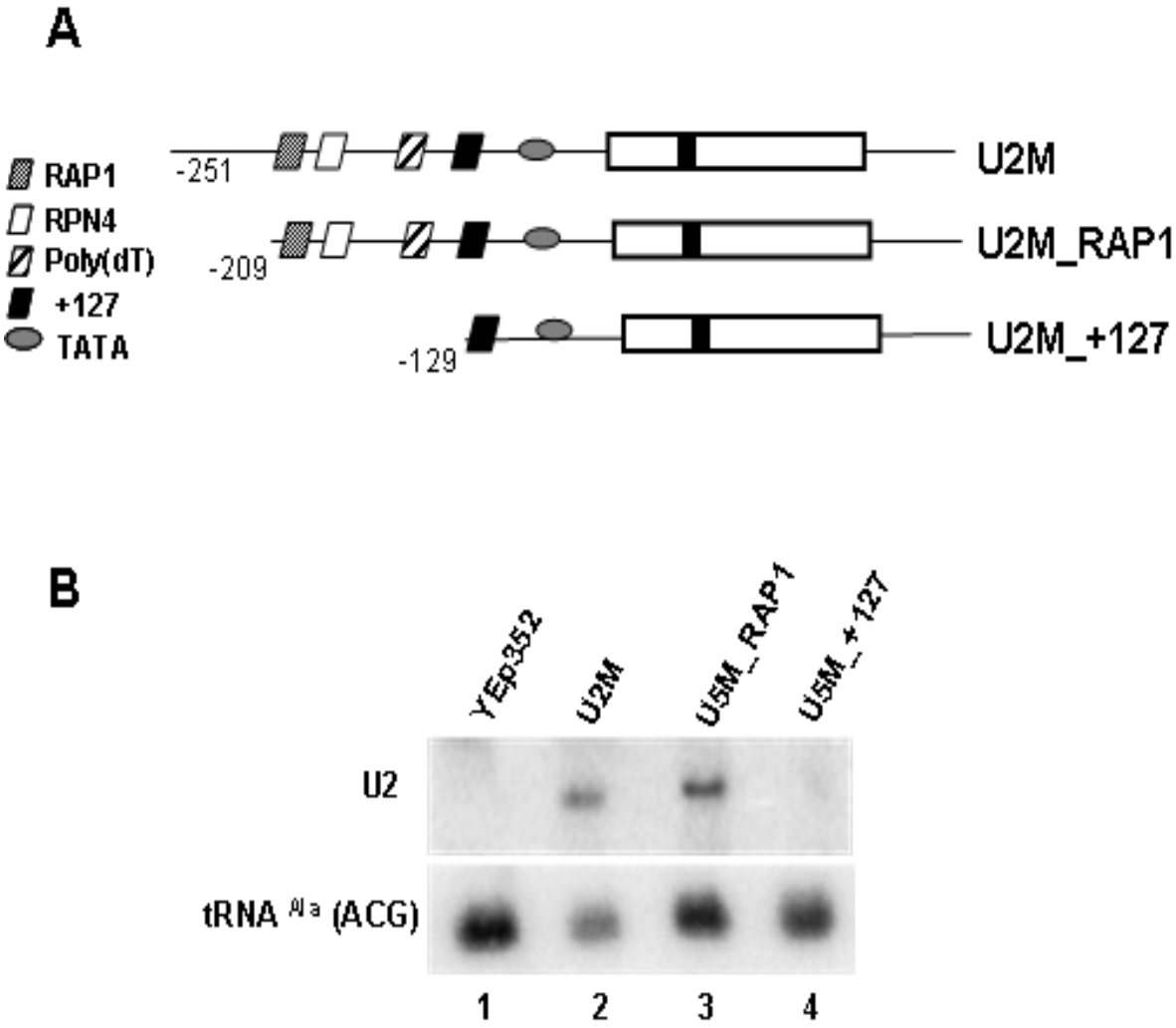

Fig. 9 (A) Schematic representation of U2M 5'-mutant templates. Black boxes indicates the position of the tag within the coding region. **(B) In vivo expression of U2M 5'-mutant templates.** Total RNA (10 µg) was extracted from BY4741 transformed with the empty YEp352 vector (lane 1) and U2M 5'-mutant templates.(lane 2-4) and gel fractionated. A radiolabeled oligonucleotide complementary to the inserted tag was used as probe. The same blot was hybridized with a probe specific for $tRNA^{Ala}$ as an internal standard.

Rap1p interacts with LSR1 promoter region *in vivo*.

To gain insight into the involvement and mechanism of actions of Ab1p, Rap1p and Rpn4p in snRNA gene transcriptional regulation, we verified physical association of these proteins to their target promoters. We also checked SNR7 promoter for occupancy by Stb3p, a recently discovered RRPE-binding protein (Liko et al.,2007).
We performed chromatin immunoprecipitation experiments from yeast strains expressing tandem affinity purification (TAP) tagged version of Abf1, Stb3, Rap1 and Rpn4 proteins. The untagged BY4741 strain was used as control.
Independent ChIP experiments were carried out, and immunoprecipitated DNA was analyzed by PCR amplification. The results of ChIP analyses are reported in Fig. 10. Occupancy values are relative to occupancy at intergenic region ARS540 Chr 5, that was used as a reference.
As shown in fig 10A, Abf1p is not found specifically associated to SNR7 promoter region; however we have demonstrated that promoter activity depends on Abf1 binding site (fig 10B). It is conceivable that regulations may not require stable binding of Abf1p to the promoter region. This mode of action, defined “hit and run”, has been previously suggested for Abf1p (Schroeder et al., 1998).
We tested if Stb3p could represent the protein binding to the SNR7 RRPE. ChIP analysis didn’t show any enrichment of Stb3 at the SNR7 promoter; anyway not all RRPE-containing genes are influenced by this protein, whose mode of action remain unclear (Liko et al., 2007).
When we analyzed LRS1 promoter for enrichment we found a strong association of Rap1p (fig 10B); since Rap1p binding is continuously required for enhancement, this result is a further confirmation of Rap1 involvement in transcription (see fig 10B).
In vivo binding of Rpn4p showed only a slight enrichment; we can’t anyway exclude that the low occupancy value obtained could be related to the extremely short half-life of this protein ($t_{1/2}$ <2 min).

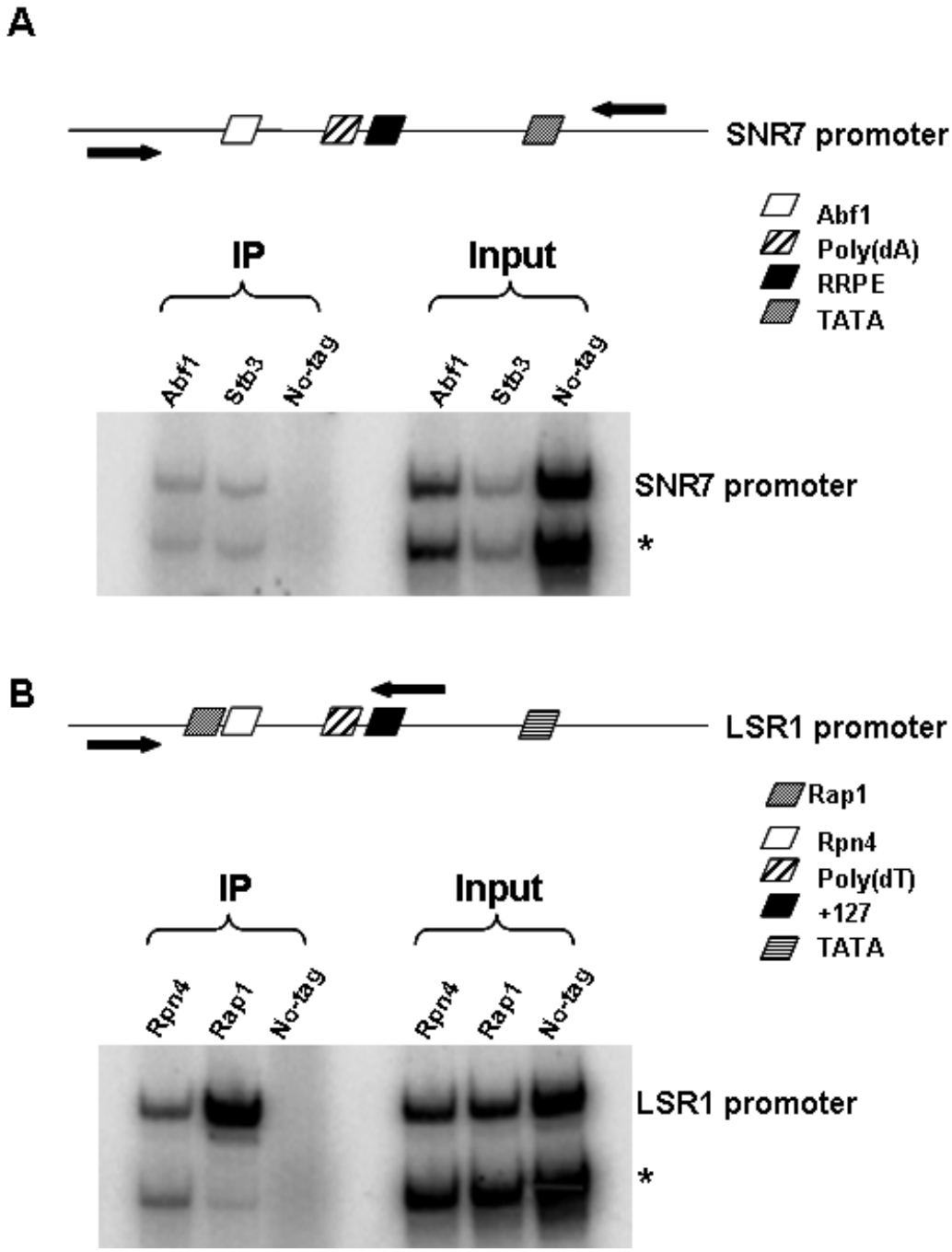

Fig. 10 In vivo occupancy at SNR7 (A) and LRS1(B) promoters. ChIP were performed from yeast strains expressing tandem affinity purification (TAP) protein-tagged version of Abf1, Stb3, Rap1and Rpn4 proteins. BY4741was used as non-tagged control strain. Immunoprecipitated DNA was analyzed for enrichment by PCR. The asterisk indicates the intergenic region ARS540 Chr 5, used as an internal standard.

Promoter are schematically represented; regions amplified by PCR in the ChIP experiment are identified by arrow pairs.

Conclusion

Defining co-occurrence and spatial relationship of individual binding sites is an important step in understanding the regulatory content of promoter regions.

The primary aim of this work was to identify putative regulatory elements in yeast Pol II-transcribed snRNA gene promoters, starting from phylogenetic footprinting. The comparative analysis showed elements conserved across *Saccharomyces sensu stricto* group, a closely related set of species in which most of the genes and regulatory elements are shared.

This approach gave a rich haul of information. The common pattern in all Pol II-transcribed gene is the presence in their promoter region of General Regulatory Factors (GRFs) Abf1p (U4 and U5 snRNAs), Reb1p (U1 snRNA) and Rap1p (U2 snRNA). These three abundant and essential transcription factors have many target sites in yeast genome acting as multifunctional proteins. They can enhance both activation and repression of transcription, but are also involved in silencing and regulation of DNA replication initiation. The GRFs seems to share a common mechanism of action; indeed the binding site for one GRF within a promoter can be exchanged with another (Fourel et al, 2002).

Their binding motif usually has little intrinsic regulatory activity and is often found in combination with poly(dT-dA) sequences, that operate not by recruitment of specific transcription factors but rather by their intrinsic DNA rigid structure. It has been hypothesized that these elements synergise in local opening of chromatin which then permits increased binding of other transcription factors. With this respect, it is remarkable that in snRNA gene promoters a T-rich or A-rich elements are found downstream the GRF binding site.

To better understand the involvement of the identified putative promoter elements in transcription regulation we have chosen two snRNA genes LRS1 (coding for U2 snRNA) and SNR7 (coding for U5 snRNA), to be used as reporter genes for in vivo expression analysis. We have prepared tagged versions of these genes by inserting an oligonucleotide within the transcribed region. Such marked templates generated stable and corrected initiated RNAs, thus allowing *in vivo* promoter analysis.

In vivo expression analysis of 5'-mutated versions of marked genes allowed us to identify the combined positive role of Abf1p-poly(dA) element and Rap1p-poly(dT)

element in SNR7 and LRS1 transcription respectively. The requirement for Rap1p-poly(dT) element was absolute, since deletion of this region caused complete loss of transcription; an analogous elimination of Abf1p-poly(dA) element caused a strong reduction in SNR7 expression. An important role in regulation of Rap1p is confirmed by its tight association to LRS1 promoter, revealed by ChIP analysis.

A high resolution atlas of nucleosome occupancy in yeast has been presented by Lee et al (2007). Both SNR7 and LSR1 promoter are characterized by a free nucleosome region that extends up to 100 bp upstream of the Abf1p and Rap1p binding site respectively. Future work will investigate if these elements function as chromatin-reorganizing factor preventing the deposition of nucleosomes in the region close to their binding site, by analysing if the loss of Abf1p and Rap1p binding or deletion of poly(dT-dA) elements are associated with changes in chromatin structure.

GRFs usually amplify the effect of neighbouring regulatory sites; we have identified the positive *cis*-acting element RRPE (ribosomal RNA processing element) in U5 snRNA promoter. It was interesting to find the same positive role of RRPE in snoRNA transcription (data not shown); this could suggest a possible common regulatory strategy of snRNA and snoRNA transcription. Further investigation will also be directed toward understanding the strong conservation of Rpn4p in U2 snRNA promoter and its possible relationship with the ubiquitin-proteasome pathway.

Characterization

of novel snRNA gene-like

transcriptional units

in the human genome

Abstract

The role of polymerase (Pol) III in eukaryotic transcription is commonly thought of as being restricted to a small set of housekeeping, highly expressed non-protein-coding (nc) RNA genes. Recent studies, however, have remarkably expanded the set of known Pol III synthesized ncRNA. By means of computer search for upstream promoter elements (proximal sequence element and distal sequence element) typical of small nuclear RNA genes, we have identified in the human genome some putative transcription units.

In this work we analyzed the *in vitro* transcription properties in HeLa and SKNBE neuroblastoma cells nuclear extracts of some of these putative units, showing that their promoter elements were actually able to support Pol III-dependent transcription.

In particular, we identified a novel 171 nt ncRNA, 17A, whose transcription was driven by the presence of a PSE located at -59, which ensure efficient transcription, and a TATA box located at -28, that directs start site selection. Indeed, in the absence of the TATA box, another A-T rich element located downstream, not affected by the presence of the PSE, could direct an efficient alternative initiation of a shorter transcript. We have also defined the promoter architectures of two other ncRNA genes, 38A and 29A, characterized by the presence of pol III-type3 elements, PSE and TATA box, both necessary to ensure an efficient level of transcription. Finally we identify a TATA less PSE-dependent promoter for 51A.

Chromatin immunoprecipitation experiments conducted in SKNBE and SHSY5Y neuroblastoma cells using antibodies directed against SNAPc and other components of the Pol III-type 3 transcription machinery, demonstrated its association with the 38A and 51A promoters, suggesting a Pol III-dependent *in vivo* expression of these genes.

Some of the characterized transcriptional units (17A, 38A, 51A) are internal to known protein-coding genes where alternative splicing events takes place, while another (29A) was identified as an Alu.

Further studies will be required to better characterize these transcription units *in vivo* and the function of their products, even though their location suggest a possible regulatory role and an increasing complexity of the by pol III transcriptome.

Introduction

RNA molecules are active participant in regulating, catalysis and controlling of many fundamental cellular processes, role that was reserved for proteins just few years ago; as happens for proteins the structure of these RNAs is crucial for their functionality. These regulatory RNAs, that do not encode a protein, are referred as non coding RNAs and fold into specific higher order structures that impart a function to the molecule. They are characterized by different size, tissue specific expression pattern and biological functions. Pol III is specialized in transcription of ncRNA, being its role commonly restricted to small set of housekeeping, highly expressed non-protein-coding (nc) RNA genes. Recent studies, however, have remarkably expanded the set of known Pol III synthesized ncRNAs, suggesting that gene-specific Pol III regulation is more common than previously appreciated. Newly identified Pol III transcripts include small nucleolar RNAs, microRNAs, short interspersed nuclear element-encoded or tRNA-derived RNAs. Recent advances in mammalian genome studies are bringing to light the occurrence of a widespread transcription of non coding regions devoted to the regulation of the protein coding genome expression (Mattick, 2004). In addiction, the involvement of Pol III transcripts in gene regulation and, more generally, in several important RNA dependent functions makes it likely that alterations in their levels and activity will compromise diverse cellular processes.

Comprehensive transcriptome analyses, genome wide location studies of transcription factors and computational searches for ncRNA sequences and DNA regulatory regions in whole genome are powerful tools in discovering new ncRNAs.

Pol III-transcribed genes are characterized by three promoter types called type 1-3. Type 1 consist of an internal control region, which can be subdivided into A box, intermediate element and C box. The type 2 promoter consist of an intragenic A and B boxes. Type 3 promoter is totally extragenic. The best characterized type 3 promoters belong to snRNA U6, 7SK RNA, H1 RNA and RNase MRP RNA genes. The sequences required for efficient basal transcription are a TATA element located between -30 and -25, which acts as the major determinant of the polymerase specificity, and a proximal sequence element (PSE) between -66 and -47. The PSE recruits a stable protein complex, known as SNAPc or PTF, containing five subunits.

Activated transcription is provided by the distal sequence element (DSE) located between -260 and -190. DSEs are composed of many functional submotifs that can be present either simultaneously or separately. Two of these are often the octamer and the Staf motifs that binds respectively Oct-1, a homodomain transcriptional activator, and Staf, a seven zinc finger protein (Hernandez, 2001).

Starting from the hypothesis the human genome might contain pol III transcription units each specifically regulating one or more specific pol II genes, our laboratory performed a screening in the human genome for pol III type-3 regulatory elements and identified a set of about 30 novel putative pol III-transcribed units. A detail investigation revealed that one of this novel non-coding genes, called 21A, played a role in control of the proliferation of some tumor cells (Pagano et al, 2007). In these work we selected four of these identified units (17A, 38A, 29A, 51A) that were internal to known protein genes, to be characterized by experimental determinations, showing that their promoter elements were actually able to support Pol III-dependent *in vitro* transcription.

Materials and Methods

Human genomic DNA isolation:

Extraction of genomic DNA was performed as detailed below. Four volumes of PBS (50mM potassium phosphate; 150 mM NaCl) were added to a sample of saliva (4 ml) vortexed and then centifuged at 1800xg for 5 min. The supernatant was carefully removed and the pellet rinsed using additional 2 ml of PBS, before centrifugation again at 1800xg for 5 min. After removal of the supernatant, the pellet was resuspended in 360 µl of PBS. An equal volume of lysis buffer (50 mM Tris-HCl, pH 8; 10% (w/v) SDS) was added and the mixture was incubated at 65 °C for 30 min. Pronase was added to a final concentration of 1 mg/ml followed by incubation at 45 °C for 1h and 30 min and treatment with RNase (final concentration 75 µg/ml; incubation at 37 °C for 45 min). After extraction with phenol-chloroform, DNA was precipitated at RT with isopropyl alcohol in the presence of 20 µg of glycogen and centrifugated for 10 min at 14000 rpm. The pellet was finally resuspended in 50 µl of AE buffer (10 mM Tris-HCl, pH 8). DNA concentration was measured by spectrophotometry at 260 nm. Purity of DNA was assessed using the ratio of $OD_{260/280}$ with a value of 1.8 -2.0 being of good purity.

Amplification and cloning of DNA templates:

Some of the putative transcriptional units obtained by a computer analysis were selected for experimental determinations. DNA fragments were PCR-amplified from human genomic DNA using Taq DNA polymerase (Fermentas) and cloned into pGEM –T easy vector (Promega). 12A, 14A, 17A, 22A, 24A, 29A, 31A, 38A, 41A, 42A, 43A and 51A templates were digested with SphI-SacI and subcloned into pNEB193 vector. All constructs were verified by sequencing and purified with the Qiagen Midi Kit according to manufacturer's instructions before to be used for *in vitro* transcription reactions. All oligonucleotides used in amplification and cloning of the selected units are listed in table 1.

Primer name	Primer sequence
12A_fw	GAAATAACTTATAAAACAGGTCATCC
12A_rev	GGAAAATTAATGTATTATAAATGCAAAGG
14A_fw	ACTGATGTATGATTATATCTTATTTTGG
14A_rev	TTTGTCTGTGGACAAATTGTCTCC
20A_fw	GGTAATGGATTAATTGATTCAGACC
20A_rev	ACAGGAAAGTAGACAGCGGAGCTGG
22A_fw	AAACAATAATCATCCTTTGTAAAGAGC
22A_rev	TATGCAGAAATCTAATACAGGAAGC
23A_fw	ATACCAGGTCTATAACACCTGAGC
23A_rev	TACTATAACTACAATCCATTACTACAC
24A_fw	GCAACATAATACACAAGAAGAAGAAC
24A_rev	CTTGCTATAAACATTTACTTGTGTGG
29A_fw	GGTATTTTGGTGTTTCAGCCTTTCC
29A_rev	ACATTGAATCACCTTATGAG
31A_fw	CTTTCTTGAAATATTCTCATTGTTACC
31A_rev	TAGTGACTTACTGGAGAGGATAGG
33A_fw	ATACGAAGGCATAGTAGAAGAAAGC
33A_rev	CAGACTAATGTCTTCCTGAGAGG
34A_fw	TTTCTGCTTGTGCCCATTACAACC
34A_rev	TACAGGCGTAGGCCACTACAGC
36A_fw	GCCCATTGCCAAGTCCTTCACG
36A_rev	AAATCAAAAGTTGTCTCACTCAGG
37A_fw	AAAAGAAAAGAAACCATTGTTTAATCC
37A_rev	TAACTTCTATCAAAATCTCTGCTGG
38A_fw	CTAGCAATAGCAATCAGACCAGAG
38A_rev	TTCAGGGTGTCCTGTTGGTACC
41A_fw	TAGACACACATATGTAAATTCACTCC
41A_rev	ATTGCTTACCATTTTGCATTATCACC
42A_fw	TCCAAGTTTACAAATAGGGTCTCC
42A_rev	ATAAATCAAAAGTTGTCTCACTCAGG
43A_fw	AAAAGCCCAATGTTTAACATATCC
43A_rev	AACCTGCGTGTTTAAAAAGAGTCC
44A_fw	GTTACATAAGCTTTCTATGCCTTGG
44A_rev	AGCTGCACCCTGGCACTGTCC
45A_fw	TAATTGAGAAGCAACTCCAGCTCTAGC
45A_rev	GGAACTCTGATAATTTCGTTTATAGG
48A_fw	GTCGGAGGGTCTTCTTGCATGG
48A_rev	AACCCATTGAATGAAGGCCTAGC
51A_fw	ACAAACTCCATCTGCAATTCCTCG
51A_rev	CAGGTATAGGAGGGGTGCAGC

Table 1 List of oligonucleotides used in amplification and cloning of putative transcriptional units selected for experimental determinations.

17A templates: Constructs were obtained by PCR amplification using 17A_pNEB as template (containing 90 bp of the 5'-flanking region and 34 bp of the 3'-flanking region) and specific mutagenic forward primers. All the fragments were cloned in pNEB193 vector (SmaI site), sequence verified and purified with the Qiagen Midi Kit.

38A templates:

Constructs were obtained by PCR amplification using 38A_pNEB as template (containing 347 bp of the 5'-flanking region) and specific mutagenic forward primers. 38A_poliT template was generate by amplification from 38A template with a specific reverse primer containing a sequence of 9T. All the fragments were cloned in pNEB193 vector (SmaI site), sequence verified and purified with the Qiagen Midi Kit. The fusion 38A_7SK DNA fragment was obtained by amplifying the 7SK coding region with a specific primer containing the 38A upstream sequence from -68 bp to -27 bp. The hybrid product was cloned into pNEB193 vector (SmaI site). The 5'-mutant of 38A_7SK were obtained by PCR amplification with specific primers forward. The fragments were cloned into the SmaI site of pNEB193 vector.

29 templates:

Constructs were obtained by PCR amplification using 29A_pNEB as template (containing 522 bp of the 5'-flanking region) and specific mutagenic forward primers. 29A_poliT template was generate by amplification from 38A template with a specific reverse primer containing a sequence of 9T. All the fragments were cloned in pNEB193 vector (SmaI site), sequence verified and purified with the Qiagen Midi Kit. The fusion 29A_7SK DNA fragment was realised generating two PCR products amplifying from 29A_pNEB template and 7SK template using the high fidelity Pfu DNA polymerase (Promega) and gene-specific pairs of oligonucleotide primers: 29A_SacI_fw in combination with 29A_BamHI_rev to realise the first insert containing 29A 5'-flanking region from positon – 513 to -25; 7SK_BamHI_fw in combination with 7SK_SphI_rev to realise the second insert containing 7SK the transcribed sequence plus 18 bp of 5'-flanking region and 124 bp of 3'-flanking region. The fragments were digested with SacI/BamHI and BamHI/SphI respectively and inserted in pNEB193 (sacI/ sphI sites). The 5'-mutant of 29A_7SK were obtained by PCR amplification with

specific primers forward. The fragments were cloned into the SmaI site of pNEB193 vector.

All oligonucleotides sequences used for amplification and cloning of 17A, 29A, 38A, 51A templates are listed in table 2.

Primer name	Primer sequence
17A_CC_fw	CCTCACCATAAAAGTGAAATAATG
17A_AA_fw	CCTCAAAATAAAAGTGAAATAATGTTGC
17A_TATA1_fw	AATAAATAGTGCAAAATATTAACAAAG
17A_TATA2_fw	*AAATATTAACAAAG*ACACAATTGAATA
38A_deltaDSE_fw	TAATAACAACATATCTGAAAAAGACGC
38A_deltaTATA_fw	GTTGAAGAAGACACATATAAATAGA
38A_DSE_fw	TATTTGCATATAAAAATAGTTAGAAATAAATTTAACC
38A_poliT_rev	AAAAAAAAACTATTTCTGTGATGCATATCCTTG
38A_7SK_fw	TAACCATAAAGGTGAAATATTTGTATACCGATAACTAT AAAGCTTGTGCGCCGCTTGG
7SK_Rev	CGGGAGGTGGAGGTTACAGTGAGC
38A_7SK_TATA_fw	ACTATAAAGCTTGTGCGCCGC
7SK_NP_fw	GCTTGTGCGCCGCTTGGGTACC
29A_deltaDSE _fw	GGAACTTTATGTCGCTACC
29A_ ΔTATA_fw	AGACACTGAATTCTAACTAGACGC
29A_ FW_SacI	CGCCGAGCTCGGTATTTTGGTGTTTCAGCCTTTCC
29A_ Rev_ BamHI	TACTGGATCCTATTTATTGTGAGTTCTAGTAATTTC
7sK_ fw_ BamHI	TATAGGATCCGCGCCGCTTGGGTACCTCGG
7sk _Rev_ SphI	ACATGCATGCGGAGGTGGAGGTTACAGTGAGC
29A_710-1270 fw	GGAACTTTATGTCGCTACC
29A_AA_7SK_ fw	GGAACTTTATGTCGCTAAAATAAATG
29A_TATA_7SK	*AATAAATAGGATCCGCGCCGC*
29A_deltaDSE _poliT_rev	AAAAAAAAACCTGAGCTCAAGCAACCC

Table 2 List of oligonucleotides used in amplification and cloning 17A, 29A, 38A, 51A templates.

Preparation of SKNBE extract:

We prepared the extract according to Dignam et al (1983). We prepared 10 plates (15 cm) of SKNBE cells grown to about 90% confluence. Each plate was washed once with 10 ml of PBS; cells were trypsinized and collected by centrifugation at 4°C for 3 min at 2000 rpm. The pellet was washed one with cold PBS (3 ml) and the packed cell volume was determined. Cells were resuspended in 2.5 volumes of hypotonic S100 buffer (20 mM Hepes/KOH, pH 7.9; 10 mM KCl; 1.5 mM $MgCl_2$), containing 0.2 % NP40, 3 mM DTT and 0.5 mM PMSF, and directly centrifuged at 12000 rpm for 30 sec at 4°C. The supernatant (cytoplasmic extract S100) was carefully removed and dialyzed against a 10% glycerol containing buffer (20 mM Hepes/KOH, pH 7.9; 60 mM KCl; 3 mM DTT; 0.5 mM PMSF). The pellet was analysed under the microscope to verify the integrity of the nuclei then resuspended in 1.5 volumes of nuclear extract buffer (20 mM Hepes/KOH, pH 7.9; 10 mM KCl; 1.5 mM $MgCl_2$; 20% glycerol; 3 mM DTT; 0.5 mM PMSF). 0.5 volumes of the same buffer containing 1.2 M KCL were added drop wise at 4°C to the nuclei-buffer slurry over a period of half an hour, moderately vortexing. This mixture turn over-head at 4°C for 30 min then was centrifugated at 40000 g for 30 min at 4°C. The resulting supernatant (nuclear extract) dialyzed against a 20% glycerol containing buffer (20 mM Hepes/KOH, pH 7.9; 60 mM KCl; 3 mM DTT; 0.5 mM PMSF; 0.2 mM EDTA).The extract was frozen as aliquots in liquid nitrogen and stored at -80°C. The protein concentration was 10 mg/ml.

***In vitro* transcription analyses:**

Transcription reactions were carried out in a final volume of 25 µl in the presence of 2 µg of template DNA and HeLa cell or SKNBE nuclear extract (100 µg) supplemented with 50 ng of recombinant human TBP. The standard transcription mix contained: 5 mM creatine phosphate, 70 mM KCl , 5 mM $MgCl_2$, 20 mM Tris/HCl pH 8, 1 mM DTT, 2 µg/ml α-amanitin, 0.5 mM CTP,ATP,GTP, 25 µM/ 10 µCi UTP /[α-^{32}P]UTP, SUPERase IN (Ambion, 10 U), glycerol 10 % (v/v). The reactions were incubated for 1h at 30 °C. The products were phenol extracted and precipitated with ammonium acetate. Radiolabeled transcripts were separated on 6% polyacrylamide/7 M urea

gel,quantified and visualized by phosphorimaging using a Personal Molecular Imager FX (Bio-Rad).
We used as internal standard an RNA of 118 nt in length synthesized in vitro by T7 RNA polymerase (Amersham), following the manufacturer's protocol. The RNA was phenol extracted and precipitated within each sample.

Primer extension:

A double scale transcription reaction was performed as described, without including radiolabeled UTP. A parallel reaction ,no-DNA containing, was conducted as a control.
Reverse transcription reactions were carried out using Superscript III reverse transcriptase (Invitrogen), following the manufacturer's protocol. The purified transcripts were resuspended in a final volume of 12 µl in the presence of 0.5 mM dNTPs and 1 pmole of specific 5'-end-radiolabeled probe (table 6) . The mixture was heated at at 65°C for 5 min and a mixture providing 50 mM Tris/HCl pH 8, 75 mM KCl, 3 mM $MgCl_2$,5 mM DTT, SUPERase IN (Ambion, 10 U) and 200 U Superscript III reverse transcriptase (Invitrogen) was added to a final volume of 20 µl. The reactions were incubated for 1h at 60 °C and subsequently for 15 min at 70°C to inactivate the enzyme. The products were precipitated with ammonium acetate and gel-fractionated. A parallel reaction was conducted as a control in absence of Superscript III reverse transcriptase.
The 5' end was mapped by direct comparison with dideoxy chain termination sequencing reactions (Kit Thermo-Sequenase, GE Healthcare) run on the same gel.
Oligonucleotides used as specific probe in primer extension reactions are listed in table 3.

Primer name	Primer sequence
17A_probe	GGACTTTCCAAGATTGCCCAGGG
38A_probe	GGTTAATACAGACATTTTAACATTGTTA
29A_probe	GGATTACACACAGGAAGCCACCG
51A_probe	CCCTCATGGCACTTGGAGATTTG

Table 3 List of oligonucleotides used as specific probe in primer extension analyses.

Chromatin immunoprecipitation:

Cross-linked chromatin was prepared essentially as describe (Kurdistani S, 2003). Human cells (SKNBE and SHSY5Y line) were grown to approximately 75% confluence (two plates) and then crosslinked with formaldeyde for 30 min. Cells were washed twice in cold PBS (each plate with 10 ml) complemented with protease inhibitors (Complete, Roche: 1 tablet). Cells were resuspended in 500 µl of PBS, scraped into microcentrifuge tubes and pelleted for 2 min at 10000 rpm at 4°C. The pellet was then suspended in 400 µl of lysis buffer (1% SDS; 50 mM Tris/HCl, pH 8; 20 mM EDTA; protease inhibitors). Cells were incubated 10 min on ice and sonicated to reduce DNA fragments to an average size of 300 bp. The lysate was clarify by centrifugation for 10 min at maximum speed at 4°C. The supernatant was transferred to clean tube. Before proceeding with immunoprecipitation, 100 µl of solution were put aside as input samples, diluted to 1 ml with buffer DB (1.67 mM Tris/HCl, pH 8; 0.001 % SDS; 0.1 % Triton X-100; 0.12 mM EDTA; 16.7 mM NaCl). Beads (30 µl, slurry 50%) were added, followed by incubation of the tubes on a nuotator at 4°C for 1h.

For immunoprecipitation chromatin solution was incubated over night at 4°C with 10 µl of antibody (α-SNAPc or α-Brf2 or α-pol III). Fifty microliters of a 50% (v/v) suspension of protein A –Sepharose beads was added and incubated for 2 h at 4°C. The protein A beads were pelleted at room temperature for 3 min at 8000 rpm. The supernatant was removed and the beads were resuspended in 200 µl of wash buffer (50 mM HEPES, pH 7.9, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, 140 mM NaCl) and transferred to a 0.45-µm filter unit (Millipore Ultrafree-MC). The tubes were rinsed with another 200 µl of the wash buffer, and the rmaining beads were added to the filter unit. The filter was rotated 15 min at 4°C and spun at 8000 rpm for 3 min. The flow-through fraction was discarded. The washing step was repeated once with the wash buffer containing 500 mM NaCl, and twice with LiCl Buffer (20 mM Tris-HCl, pH 8, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA, 250 mM LiCl) and TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). To elute chromatin from the beads 100 µl of elution buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 1% SDS) was added to the filter unit which was then incubated at 65°C for 10 min and spun at 13000 rpm for 3 min. Elution was repeated once more. The flow-through were combined and incubated at 65 °C overnight. The sample was treated with RNase

(conc 0.04 mg/ml, incubation at 37°C for 30 min) and proteinase K ((conc 0.08 mg/ml, incubation at 56°C for 2 h). After extraction with phenol-chloroform, DNA was precipitated overnight at -20 °C with ethanol-sodium acetate in the presence of 20 µg of glycogen, and resuspended in TE buffer (10 µl). For input samples 5 µl of sonicated crosslinked chromatin were added to 245 µl TE with 1% SDS and incubated overnight at 65 °C. These input samples were then treated with Proteinase K, phenol/chloroform extracted, ethanol precipitated and dissolved in 100 µl TE.

DNA Amplification:

DNA samples were amplified by PCR using GoTaq polymerase (Promega) and specific primer pairs (Table 4). Typically 1 µl of the immunoprecipitated DNA and 1µl of the total DNA Input were used. The specific primers for the GAPDH exon 2 were used as internal control. Reactions were carried out in 10 µl and contained 0,25 µM primers, 0,1 mM dNTPs, 0,06 mCi/ml di α^{32}P-dCTP (Perkin Elmer). PCR products were fractionated on a 6% polyacrylamide gel, visualized and quantified by phosphorimaging. The fraction of immunoprecipitated material for a specific fragment was calculated by ratio of immunoprecipitated DNA over total DNA.

The primers used in DNA amplification are listed in table 4.

Primer name	Primer sequence
38A_fw_deltaDSE	TAATAACAACATATCTGAAAAAGACGC
38A_rev_ChIP	GTGTCTTCTTCAACTTTTTTTATC
51A_fw	ACAAACTCCATCTGCAATTCCTCG
51A_rev_ChIP	GAGATTTGAAAGGACTGCAGG
U6_fw	GTACAAAATACGTGACGTAGAAAG
U6 _rev	GGTGTTTCGTCCTTTCCACAAG
GAPDH _fw	AGGTCATCCCTGAGCTGAAC
GAPDH_rev	CCACCTGGTGCTCAGTGTAG

Table 4 List of oligonucleotides used as primers for the PCR on imunoprecipitated chromatin. The amplicon length was set to 143 bp, 151 bp, 150 bp for gene-specific primers 38A, 51A, U6 respectively and to 189 bp for GAPDH control region.

Results and discussion

RNA polymerase III-dependent transcription of putative transcription units identified by computational analysis.

In a previously work from our laboratory, we identified a novel set of ncRNA screening the human genome for upstream promoter elements typical of type 3 -pol III promoters (proximal sequence element and distal sequence element). In particular we tested H1 PSE as query sequence for the search of similar elements in the human genome by using the BLAST algorithm (http://www.ncbi.nlm.nih.gov/BLAST) and selected among the sequences obtained those that contained a DSE sequence element within a distance of 1000 bp upstream of the PSE. We further investigated if those units had features compatible a pol III promoter structure: a TATA-like element downstream of the PSE and a termination signal (run of at least four T residues). To complete the definition of the transcriptional unit, we assumed the transcriptional start site about 30 bp downstream the TATA box. We finally selected 31 novel putative units with pol III transcribed snRNA gene features (Pagano et al.,2007). When these non coding sequences were used to challenge the human genome database, it was found that some of them were internal to known or predicted protein-coding genes.

We have already investigated 21A transcriptional unit (table 5) and its regulatory activity (Pagano et al.,2007).

To test if those units were actually transcribed, we selected for experimental determination those characterized by a PSE-TATA distance close to the canonical distance of 20 bp. These units were PCR amplified from genomic human DNA and inserted in pGEM-T easy vector. After a preliminary *in vitro* transcription assay, we decided to continue the analysis with the templates generating the most interesting transcription patterns (data not shown). The selected units (12A, 14A, 17A, 22A, 24A, 29A, 31A, 38A, 41A, 42A, 43A and 51A) were subcloned in pNEB193, a vector producing a lower non-specific transcription background, and *in vitro* transcription reactions were performed. The reactions were carried out in HeLa extract in the presence of α-amanitin (2 μg/ml), to ensure pol III specific transcription. We used transcription of the empty vector pNEB193 as a control.

Tr.Unit	PSE-TATA distance	Tr.Length(nt)	Hum.Gen.Map	BLAST Human Genome
12A	25	99	2p24.3	-
14A	19	148	3p12	-
17A	24	159	9q22-q31	GPR51(intron3) Sense
20A	21	547	14q22.1	-
21A	76	333	8q24.1	CENPF(intron7,14,18)Antisense
22A	23	235	6q16-q21	-
23A	24	200	Xq21.3	-
24A	28	307	12q21	-
27A	51	91	7q22	-
29A	15	360	11p15	ASCL3(intron1)Sense
30A	-	258	Xp11.4	-
31A	25	231	12q21	-
32A	-	140	17q21	-
33A	31	210	1q32.2	-
34A	17	33	5q15	-
35A	50	351	8p11.2	-
36A	16	122	3p12	-
37A	17	49	14q13	-
38A	17	354	4p15.31	KCNIP4(intron1)Antisense
39A	78	76	Xp11.3	FLJ22843(intron10)Antisense
40A	55	484	11p15	-
41A	28	79	2q31	-
42A	17	122	3p12.3	-
43A	13	65	4q34.3	-
44A	27	218	4q13.3	-
45A	21	78	4p14	APBB2(intron1)Antisense
47A	58	48	2q22.1	-
48A	20	405	11q24.2	-
50A	6	156	21q21	-
51A	17	273	11q23.2	SORL1(intron1)Sense
52A	50	142	8p11.2	-

Table 5. Novel putative snRNA gene-like transcriptional units. Units selected for experimental determinations are written in bold.

Comparison of the transcription profiles with the one obtained with the empty vector (fig 1B, lane 9-10) allowed us to identify pol III RNAs selectively transcribed from the inserted units.

In vitro transcription in HeLa extract of template **12A** and **22A** generated a group of specific strong signals between 55-80 nt for the former and 90 nt for the latter (lane 1,3). Since putative transcript length was predicted to be 99 nt and 235 nt respectively, this multiple-transcript pattern could be the result of either RNA processing or RNA multiple initiation/termination. Other termination signals beside the canonical oligo(T) sequences can in fact affect pol III termination (Thomann et al., 1989).

An interesting result was obtained from **17A** and **43A** template transcription (lane 2,8). The analysis revealed the presence of signals of the expected length (table 1), with the 17A transcript more strongly expressed than the 43A RNA. Both seemed to be the specific result of transcription from these novel units, as the inserts did not contain any other sequence beside the promoter and the putative coding regions (fig 1A). The same also for **51A** template whose promoter is characterized only by the presence of a PSE and a TATA box; the about 400 nt-sized RNA (lane 11), longer than expected, could be the result of a termination signal positioned downstream the one assigned in our previous analysis (Pagano et al., 2007).

Promoter region of the other units contains one or more DSE motifs upstream the PSE element. The identification of specific signals in transcription of a wide sequence could be complicated by the possible presence of non-specific products from the 5' flanking region. This was the case for **24A** template whose analysis revealed two main signals, about 55 and 200 nt in length (lane 4 and 15), **38A** template which generated differently sized RNAs (lane 12), **41A** template with two transcripts of 150 and 50 nt (lane 5) and **42A** template with a strong band of 60 nt (lane 7). None of the transcripts was of the expected length; were they specific signals or not?

The analysis on **31A** template identified a strong signal but the RNA size, more longer than expected (lane 13) suggested a read-through transcript.

We have already identified **29A** as a short interspersed element AluJb (see Pagano et al., 2007); in addition its 5' flanking region contains an AluSg. As expected, its transcription resulted in abundant differently sized RNAs (lane 5 and 14); again, it was not clear which DNA sequence directed transcription.

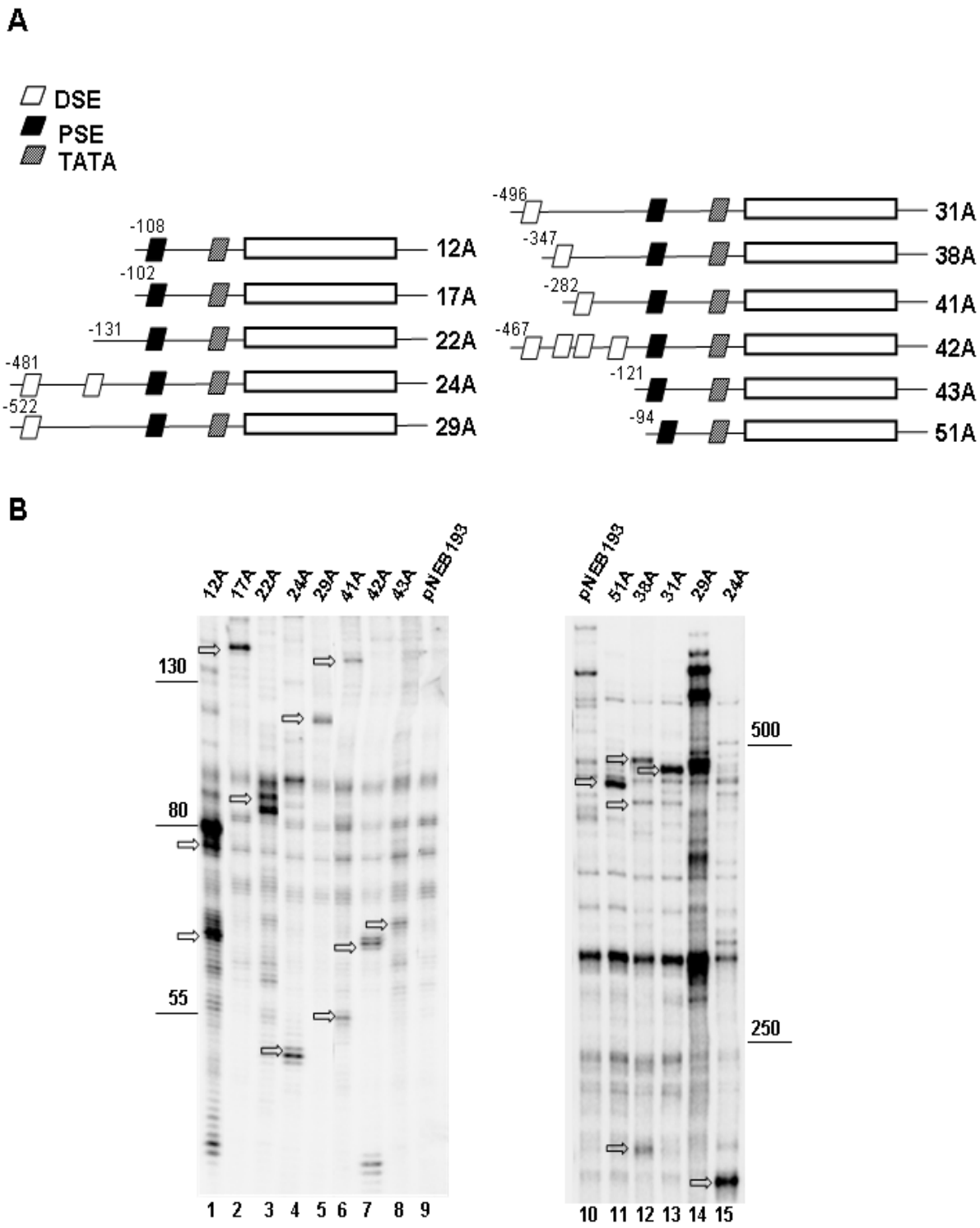

Fig. 1 In vitro transcription of putative transcription units. (A) Schematic representation of the transcription units selected for the transcription assay. (B) In vitro transcription reactions were performed in HeLa nuclear extract supplemented with 50 ng of recombinant human TBP, in the presence of α-amanitin (2 µg/ml). Transcription of the empty vector pNEB193 (lane 9,10) was used as a control. Transcription products were radiolabeled during synthesis, gel-fractionated and directly visualized. Radiolabeled RNA size markers were loaded on the same gel; their migration positions and length (in nt) are indicated on each side.

Starting from these preliminary results, showing that polymerase III could transcribe these transcription units in vitro, our next goal was to clarify if the identified type 3 promoter elements were actually able to support and regulate pol III-dependent accurate transcription and in which way.
In this work we focused on transcriptional features of those units that were internal to known protein genes (in sense or in antisense configuration) since this peculiar location could be related to the existence of regulatory RNAs (Pagano et al.,2007).

Promoter region of 17A contains a positive *cis*-acting PSE and two functional TATA boxes.

The putative transcription unit 17A maps to intron 3 of GPR51 gene, coding for the GABA B receptor subunit 2 (GABA B_2r)(fig 2). GPR51 is involved in inhibition of neuronal activity through G protein-coupled second-messenger system, which regulate the release of neurotransmitters and the activity of ion channels and adenylyl cyclase.
In addition to the PSE, promoter region contains two A-T rich elements: one quite close to the PSE (11bp downstream), the other one located at the canonical PSE-TATA distance of 24 bp (fig 2A). Before proceeding in investigation of element's function, we tried to predict the putative transcribed region. We predicted the transcription start site (the closest Py-Pu sequence about 30 bp downstream the second A-T motif) and the termination signal (run of 5 consecutive T) according to the general features of pol III transcription. Putative transcript length was set to 159 nt.
In vitro transcription analysis in HeLa extracts showed a pol III dependent production of a single transcript (fig. 2D) whose accurate initiation was confirmed by primer extension (fig. 2E, lane 4). Differently from what we predicted, the transcription started at the A residue located 21 bp downstream the first A-T rich motif (TAAATA), thus acts as a dominant TATA box. The ncRNA selectively transcribed from 17A unit was 171 nt long. RNA secondary structure from the structure prediction programme *mfold* (http://mfold.bioinfo.rpi.edu/) is shown in fig. 2C.
When we transcribed 17A in SKNBE neuroblastoma cells extract we also detected also a shorter specific signal (fig 2D); since the sequence of 5 T seemed to be a strong terminator, we postulated that these transcripts could differ in their 5' ends.

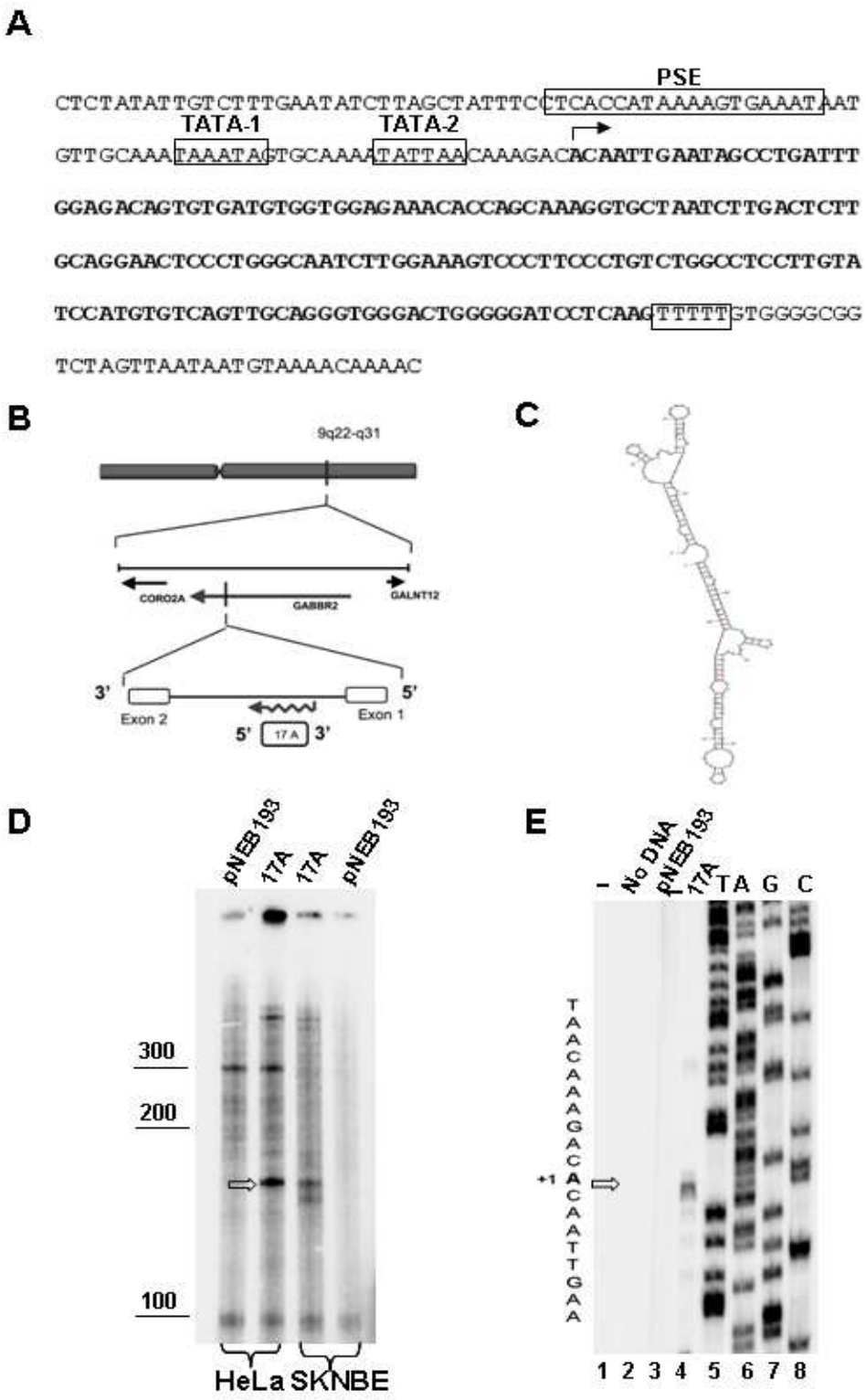

Figure.2 Transcription properties of the unit 17A. (A) nucleotide sequence of unit17A inserted in pNEB193 vector. Transcribed sequence is in bold. (B) Schematic view of GPR51locus. (C) Secondary structure of 17A RNA from *mfold*. (D) In vitro transcription of 17A in Hela and SKNBE extract. Empty vector pNEB193 was used as a control. RNA size markers were loaded on the same gel; their migration positions and length (in nt) are indicated on the left side. (E) Primer extension analysis: Lane 1, no reverse transcriptase during primer extension; lane 2 no DNA during in vitro transcription. Lane 4, primer extension product (indicated by an arrow). Lane 5-8, sequencing reactions.

To identify in more detail the role of the identified promoter elements, we generated progressive deletions in 5' flanking region of unit 17A (fig 3A).

In vitro transcription activity of the different constructs was tested in HeLa extract and the efficiencies were normalized to an internal standard synthesized using T7 RNA polymerase, phenol extracted and precipitated together with Pol III-synthesized RNAs (fig 3B).

The deletion of the region upstream the PSE didn't affect transcription (data not shown) while further deletion of the PSE element caused a transcriptional decrease of about 3.3 fold (fig.3B, lane 4). To gain insight into the mechanism of regulation, we introduced a double point mutation at the two C residues in position 4 and 5 in the PSE sequence (CC→AA). These residues are positionally invariant in human PSE sequences both in pol II and pol III PSE-dependent genes (for PSE consensus see Jawdekar and Henry, 2008). Also PSE consensus between human and mouse H1 RNA contains invariant CC (Carbon P, 2001). The constructs where PSE was mutated or not were indicated as 17A_AA and 17A_CC. Mutation reduced transcription of about 4 fold (fig 3B, lane 3). This result demonstrates that the two residues in the PSE play a crucial role in transcription.

It's interesting to observe that TATA-2 can replace TATA-1 in directing efficient transcription; the construct containing only TATA-2 generated a shorter transcript whose expression level was no affected by the loss of the PSE (fig 3C, lane 5). The result shows that TATA box is determinant for the start site selection.

If In HeLa extract TATA-1 directed transcription was prevailing, in SKNBE extract both TATA elements seemed functional (fig 2 D).

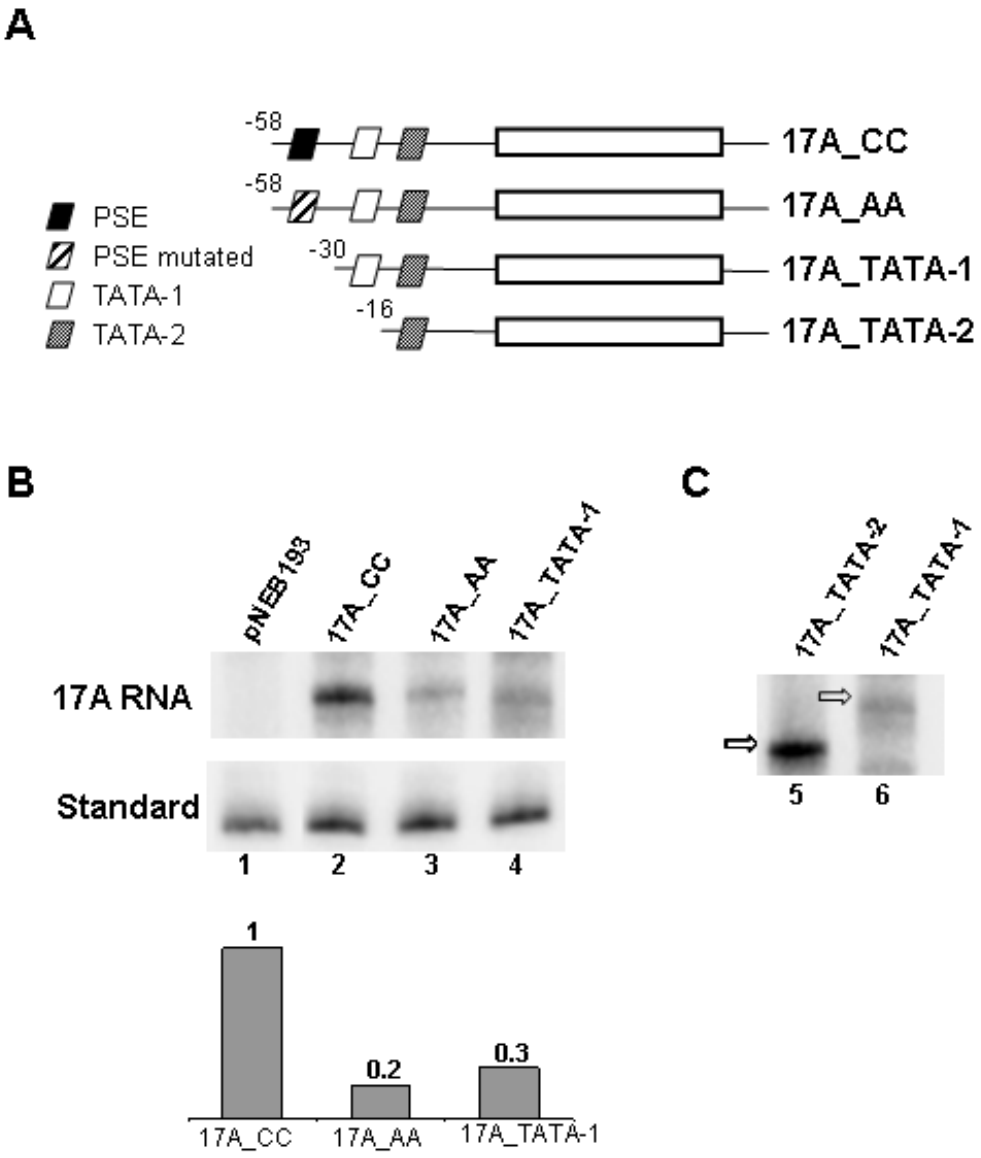

Fig. 3 Transcriptional analysis of 17A promoter region. (A) Schematic representation of 17A templates. (B) In vitro transcription analysis. Reactions were performed in HeLa extract supplemented with 50 ng of recombinant human TBP, in the presence of α-amanitin (2 µg/ml). Transcription of the empty vector pNEB193 (lane 1) was used as a control. The histogram represents RNA quantitative value relative to 17A_CC value after normalization with the RNA used as standard. (C) Alternative transcription supported by the two TATA boxes; an arrow indicates the transcribed RNAs.

PSE and TATA box support pol III dependent transcription of unit 38A.

The putative transcription unit 38A maps in intron 1 of KCNIP4 gene, encoding a member of the family of voltage-gated potassium channel-interacting proteins (KCNIPs). Members of the KCNIP family are small calcium binding proteins. They may regulate neuronal excitability in response to changes in intracellular calcium.
Promoter region of this unit contains a putative DSE motif 195 bp upstream of the PSE element (fig 4A). The preliminary *in vitro* transcription analyses identified three differently sized specific RNAs, but none of the expected length of 354 nt. Since the region cloned in the template was quite wide, the interpretation of the transcription pattern was complicated by a possible presence of non specific-products.
To clarify the result of the *in vitro* transcription, we decided to generate a series of deletion in the 5'-flanking region of 38A. We created also a template where the DSE was positioned 23 bp upstream the PSE as in H1 RNA (Carbon P et al, 2001), to test if this element could stimulate transcription. The templates shown in figure 4B were tested in *in vitro* transcription assay both in Hela and SKNBE nuclear extracts (fig 4C).
Comparing the transcription pattern of 38A_ΔDSE template and of 38A_ΔDSE template, not containing any of the promoter elements, allowed us to identify a unique specific signal, probably result of a read-through. It's known in fact that other termination signals beside the canonical oligo(T) sequences can affect pol III termination (Thomann et al., 1989). Anyway when we added a strong termination signal (9T) within the transcribed region, the cut RNA was well detected (fig 4E), showing that PSE and TATA element could actually support transcription. Accurate initiation was verified by primer extension analysis: the start site mapped at the A residue located 23 bp downstream the TATA element (fig 4D).
The octamer motif recruits the transcription factor Oct-1, well known for its stimulatory role (review: Hernadez N, 2001). At variance from what we expected, the insertion of the DSE next to the PSE produced a dramatic decrease in transcription (fig 4D, lane 10). Further investigation are required to study the possible role of the DSE element in transcriptional repression and the mechanism by which Oct-1 mediates its repression function. The role of Oct-1 as transcriptional repressor has been documented in other human gene promoters (Leung K et al., 2002).

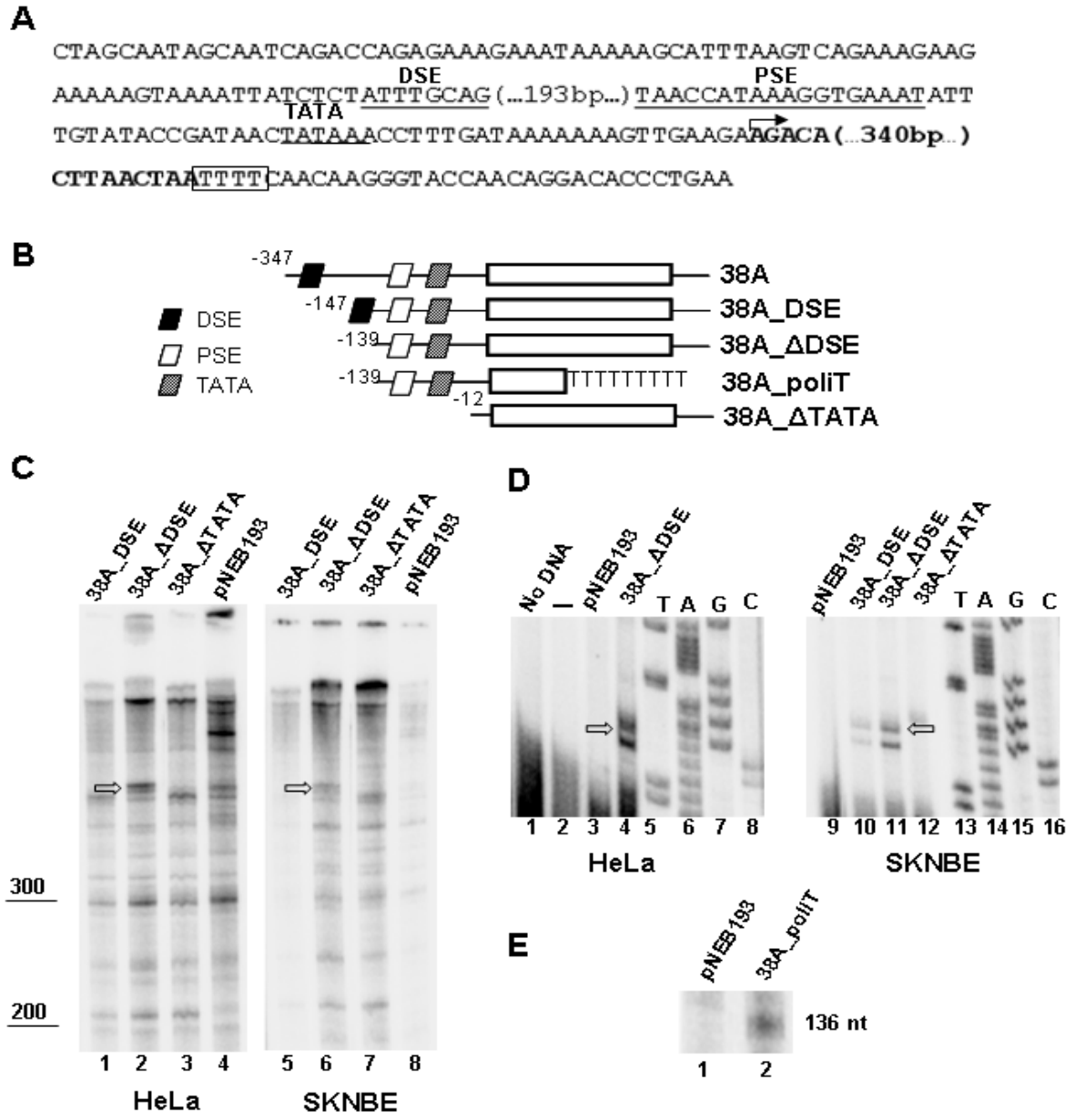

Figure.4 Transcription properties of the unit 38A. (A) nucleotide sequence of unit 38A inserted in pNEB193 vector. Transcribed sequence is in bold. (B) Schematic representation of 38A templates. (C) In vitro transcription of 38A in Hela and SKNBE extracts. Empty vector pNEB193 was used as a control. RNA size markers were loaded on the same gel; their migration positions and length (in nt) are indicated on the left side. (D) Primer extension analysis: Lane 1, no DNA during in vitro transcription ; lane 2 no reverse transcriptase during primer extension; lane 4,11 primer extension product (indicated by an arrow). Lane 5-8, 13-16 sequencing reactions. (E) In vitro transcription of 38A containing a poly T sequence as termination signal.

To better characterize the promoter activity, we decided to investigate the quantitative role of the PSE in transcription using 7SK as a reporter gene. 7SK RNA is an abundant 331 nt ncRNA that is transcribed by pol III. *In vitro* transcription of the 7SK template is easy to monitor and dependent only on the presence of external elements (Murphy S et al., 1987), ideal features for a reporter gene. We generated a fusion construct attaching the 5'-flanking sequence of 38A from -68 bp to -27 bp to the transcribed region of 7SK (fig 5A). This fusion template supported efficient transcription of an RNA of the expected length. Primer extension analysis showed that the initiation site was unchanged in the fusion template with respect to the natural 7SK gene (lanes 5,7). Removal of the sequence upstream the TATA box resulted in a fourfold drop in transcriptional efficiency (fig 5B, lane 3); anyway the TATA box alone was able to direct a basal transcription level. Values were normalized with an internal standard RNA synthesized using T7 RNA polymerase. This result showed that *in vitro* the PSE acts as a positive regulator of transcription, in agreement with studies demonstrating the requirement of this element for efficient transcription of other well known pol III- type 3 genes. The factor binding to the PSE element has been best characterized in the human system and is variously known as PTF or SNAPc (Jawdekar and Henry, 2008). To determine whether SNAPc complex directly occupies 38A promoter *in vivo,* ChIP experiments were performed with SKNBE and SHSY5Y neuroblastoma cells and antibodies directed against SNAPc. In addition, we tested the binding of other components of the Pol III-type 3 transcription machinery, with antibodies directed against Brf2, one of the multiple subunits of the pol III transcription initiator factor, or polymerase III itself. Recovered DNA segments were analyzed by PCR with specific primer to the 38A promoter (see Matherials and methods for details). A primer pairs specific to GAPDH exon 2 were used for internal standard control. The pol III-type 3 U6 snRNA promoter was analysed as positive control to verify the efficiency of the procedure. As shown in fig 5, 38A promoter region was specifically immunoprecipitated with anti-SNAPc and anti-Brf2 antibodies in both cell lines. In contrast promoter enrichment in the anti-Pol III immunoprecipitated samples was higher than that observed for the GAPDH negative control only in SKNBE cell line. Anyway also the enrichment for positive control U6 snRNA was lower than expected, suggesting an inefficient immnoprecipitation with the anti-pol III antibody. The presence of pol III, SNAPc and Brf2 in the promoter region of 38A is an important result to postulate active *in vivo* expression of this unit.

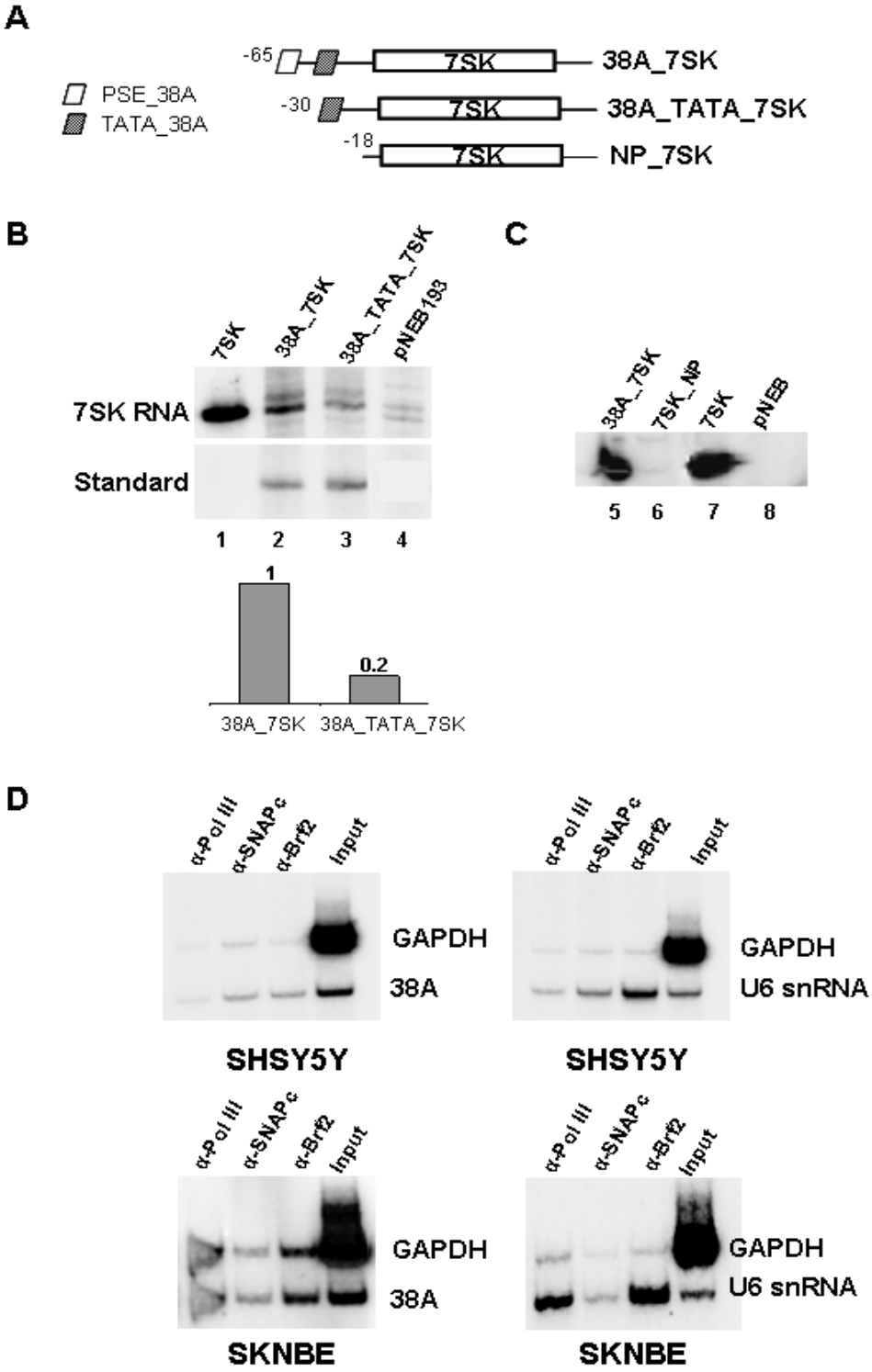

Fig. 5 Cis-trans factor in transcriptional activity of 38A (A) Schematic representation of 38A_7SK templates. (B) In vitro transcription in HeLa extract of 38A_7SK templates. Natural 7SK (lane 1) was used as positive control. The histogram represents RNA quantitative value relative to 38A_7SK value after normalization with the RNA used as standard. (C) Primer extension analysis of 38A_7SK transcript. Natural 7SK (lane 7) was used as positive control. (D) ChIP analysis of *in vivo* association of SNAPc, Brf2 and pol III to 38A promoter region. ChIP experiments were performed with SKNBE and SHSY5Y cell lines. Primer specific to GAPDH exon 2 were used as internal standard control. The pol III-type 3 U6 snRNA promoter was analyzed to verify the efficiency of the procedure.

The pol III transcription of *Alu* RNA 29A is affected by external promoter elements, PSE and TATA box.

The putative transcription unit 29A maps in the intron 1 of ASCL3 transcription factor (basic helix-loop-helix transcription factors family), essential for the determination of cells fate and development and differentiation of numerous tissues (Jonsson et al., 2004). The promoter of 29A unit was characterized by the presence of a DSE (-498), a PSE (-62) and a TATA box (-29) at canonical position for pol III-type 3 genes. Sequence inspection of the 3' end of the putative transcript revealed the absence of more than three consecutive thymidine residues; a run of 18 A residues was then considered as a Pol III terminator (Emerson and Roeder, 1984; Hess et al, 1985) (fig 6A). The analysis with Repeat Masker algorithm (http://repeatmasker.org) evidenced the presence of two *Alu* elements within the 29A sequence: an *AluSg* from the position -504 to -16 in 5'-flanking region and an *AluJb* in the putative transcribed region from position +60. *Alu* elements are the most abundant repetitive elements in the human genome and belong to the SINE (short interspersed elements) family. The pol III-transcribed *Alu* are usually characterized by the presence of internal A and B boxes, helped by an upstream enhancer. The preliminary analysis of *Alu* 29A identified only external putative regulatory elements; anyway length (about 300 bp) and the dimeric secondary structure composed of two similar but distinct monomers (left and right arm) are typical of *Alu* elements (Hasler and Strub, 2007).

In vitro transcription of 29A unit resulted in a complex pattern. The first step to clarify the transcription properties was to generate a series of 5'-end deleted construct (fig6B), thus removing the interference by the upstream AluSg. We performed *in vitro* transcription reactions in HeLa and SKNBE nuclear extracts programmed with the truncated constructs. Again we weren't able to precisely identify specific signals in the transcription pattern (fig 6C). To determine if the putative promoter elements identified could direct pol III transcription, we analyzed in vitro transcripts by primer extension; the analysis confirmed the presence of a specific transcript from 29A template initiated at the C residue positioned 29 bp downstream the TATA box (fig 6D). When the region upstream the PSE was removed (lane 3,12) we detected a unique strong signal from transcription in HeLa and SKNBE extract. When promoter region was entirely removed, transcription dropped completely (lane 13), showing that external elements, PSE and TATA, were necessary for efficient transcription.

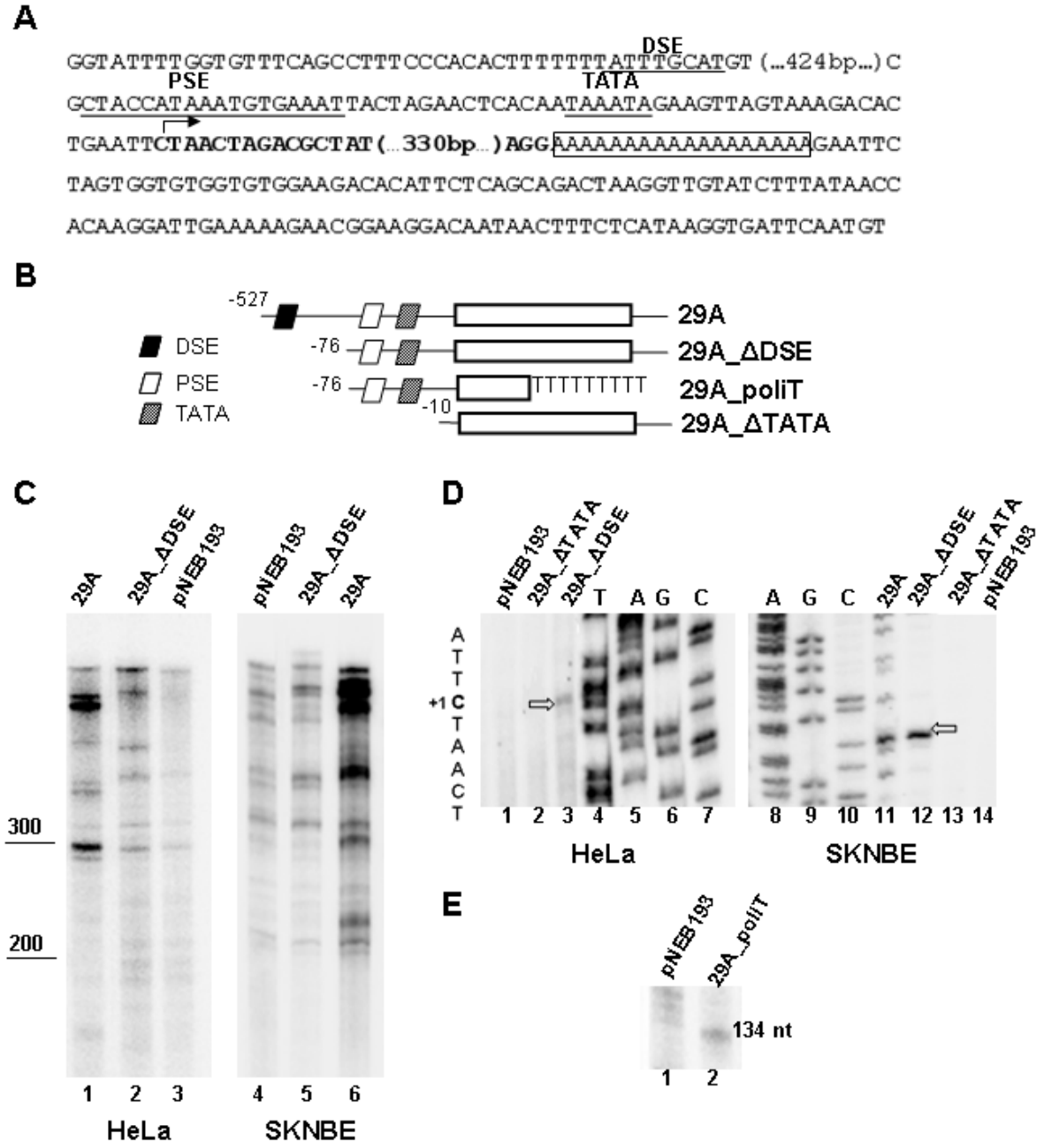

Fig.6 Transcription properties of the unit 29A. (A) nucleotide sequence of unit 29A inserted in pNEB193 vector. Transcribed sequence is in bold. (B) Schematic representation of truncated 29A templates. (C) In vitro transcription of 29A in Hela and SKNBE extracts. Empty vector pNEB193 was used as a control. RNA size markers were loaded on the same gel; their migration positions and length (in nt) are indicated on the left side. (D) Primer extension analysis: primer extension product was indicated by an arrow. Lane 4-7, 8-11 sequencing reactions. (E) In vitro transcription in HeLa extract of 29A containing a poly T sequence as termination signal.

When we added a strong termination signal (9T) within the transcribed region, the truncated RNA product was well detected (fig 6E), confirming that PSE and TATA element could actually support transcription.

The role of the PSE in 29A expression was further investigated using 7SK RNA as a reporter gene, as previously done for 38A. We generated a fusion construct attaching the 5'-flanking sequence of 29A from -535 bp to -23 bp to the coding sequence of 7SK (fig 7A).

This fusion template supported efficient transcription of an RNA of the expected length. Primer extension analysis verified that the initiation site was unchanged in the fusion template respect to natural 7SK gene (7C lane 8,9). Transcription values were normalized with an internal standard RNA synthesized using T7 RNA polymerase. Removal of the sequence upstream the PSE resulted in enhancement of specific transcription thus removing the interference by the upstream AluSg (compare lane 2,3). Further deletion of the PSE caused a fourfold drop in transcriptional efficiency (fig 7B, lane 5): TATA box alone was thus able to direct a basal transcription level. The involvement of the PSE was investigated more in detail introducing a double point mutation at the two C residues in position 4 and 5 in the PSE sequence (CC→AA) as already done in the analysis of 17A promoter. Transcription efficiency of the mutated template 29A_AA_7SK was half respect to the wild type. Even if transcription reduction caused by mutation and deletion of PSE wasn't exactly comparable, both these results suggest that *in vitro* the PSE acts as a positive signal for transcription as it happens for other well known pol III- type 3 genes.

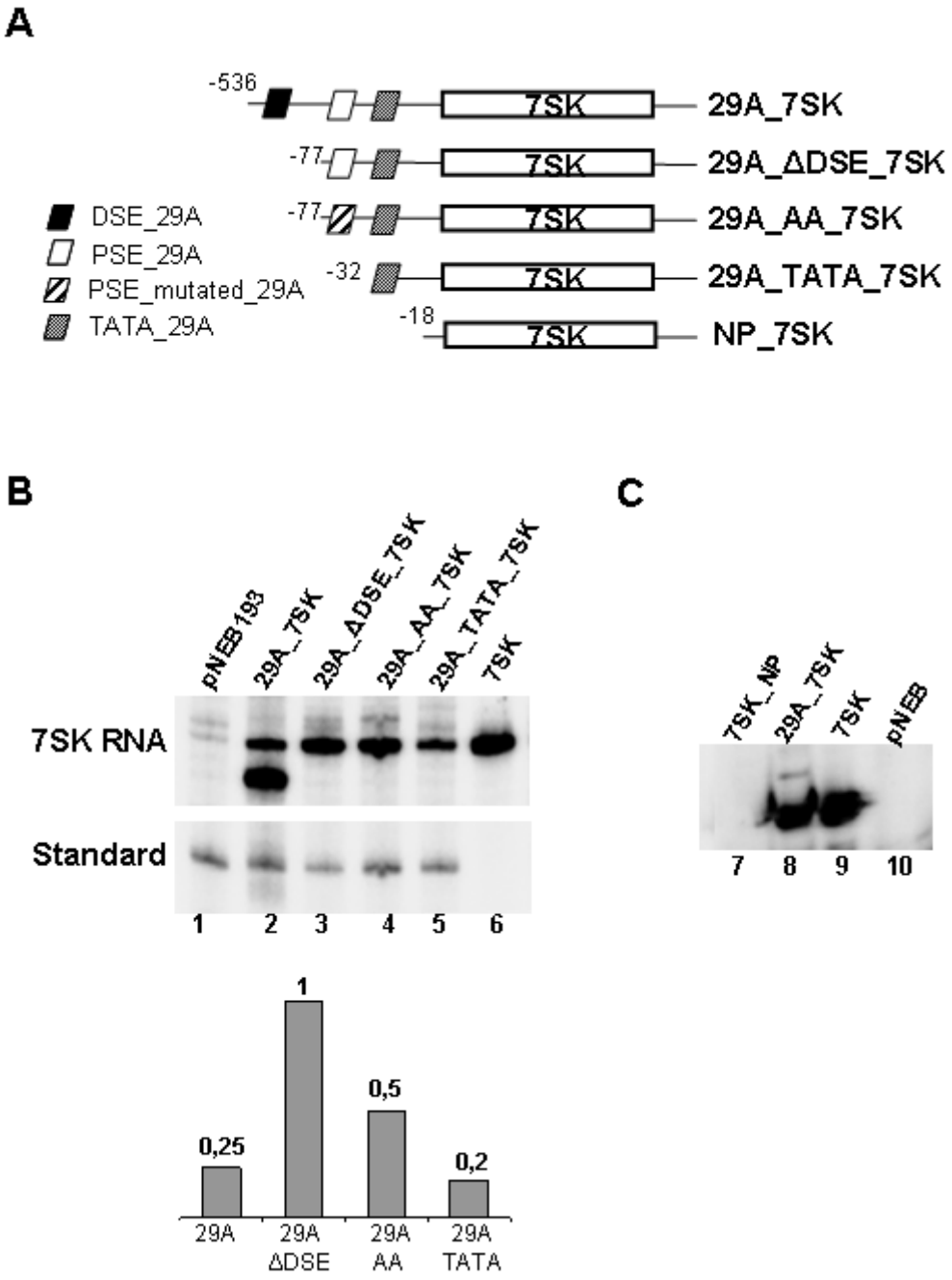

Fig. 7 Regulation of 29A transcription by PSE element. (A) Schematic representation of 29A_7SK templates. (B) In vitro transcription in HeLa extract of 29A_7SK templates. Natural 7SK (lane 6) was used as positive control. The histogram represents RNA quantitative value relative to 29A_ΔDSE_7SK value after normalization with the RNA used as standard. (C) Primer extension analysis of 29A_7SK transcript. Natural 7SK (lane 9) was used as positive control.

51A RNA transcription is supported by a TATA less PSE-dependent promoter.

Unit 51A maps in the intron 1 of SORL1, encoding a protein that belongs to the family of vacuolar protein sorting 10 (VPS10) domain-containing receptor proteins. This gene is strongly expressed in the central nervous system. The preliminary analysis of 51A promoter identified a TATA box located downstream the PSE (17 bp) (fig 8A). In vitro transcription of 51A template resulted in a specific signal (fig 8B), that was much stronger when transcription was carried out in nuclear extract from neuroblastoma cells (SKNBE). Transcript length could be explained by a the read-through of polymerase III across the first potential terminator, a T_4 run, to the next available downstream stop signal (T_6). A secondary structure model of 51A RNA from the structure prediction programme *mfold* is shown in fig 8D. Primer extension analysis confirmed the efficient transcription of 51A both in HeLa and SKNBE cells nuclear extract of a transcript initiated at the A residue upstream the TATA box (fig 8C, lane 4,12). This result suggested that in this particular case the PSE, located at position -32 with respect to the TSS (the canonical position for TATA box), might act as the main actor in selection of the start site. SNAPc binding to the PSE of snRNA gene promoters is a prelude to the recruitment of the general transcription machinery; in the case of pol III-transcribed snRNA genes SNAPc recruits TBP to the TATA box. The SNAPc/TBP juxtaposition results in a recruitment of a TFIIB-related factor called Brf2 (Jawdekar and Henry, 2008). To verify whether the pol III transcription machinery interacts with the 51A promoter in vivo, we performed ChIP experiments in SKNBE and SHSY5Y neuroblastoma cells using antibodies directed against SNAPc, Brf2 or a subunit of RNA polymerase III. Recovered DNA fragments were analyzed by PCR with primer specific for the 51A promoter region. A Primer pair amplifying a region of GAPDH exon 2 were used as a specificity control. The pol III-type 3 U6 snRNA promoter was analysed as a positive control for interaction with the Pol III machinery. As shown in fig. 9, pol III machinery components specifically immunoprecipitated with 51A promoter region in both cell lines, suggesting a possible PSE-dependent pol III active transcription *in vivo*. 51A transcribed region was finally analyzed with Pol3scan program (Pavesi et al., 1994) to verify the presence of putative internal regulatory elements; a B box-like motif (CTTTCAAATCT) was found at position +74. Further investigation are required to test the possible cooperation of extragenic and intragenic elements in transcriptional regulation of 51A RNA.

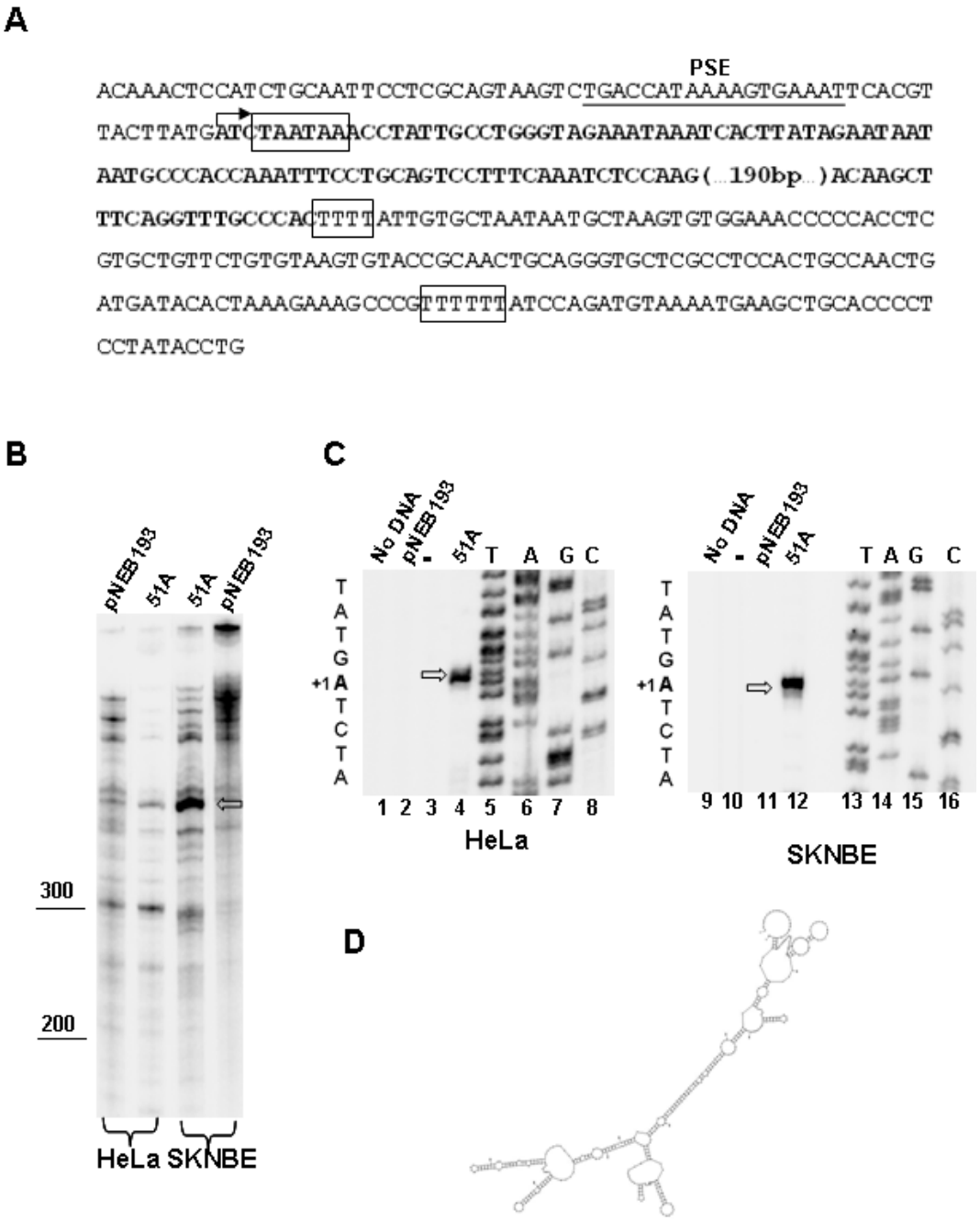

Fig. 8 Transcription properties of the unit 51A. (A) nucleotide sequence of unit 51A inserted in pNEB193 vector. Transcribed sequence is in bold. (B) In vitro transcription of 51A in Hela and SKNBE extract. Empty vector pNEB193 was used as a control. RNA size markers were loaded on the same gel; their migration positions and length (in nt) are indicated on the left side. (C) Primer extension analysis: Lane 3,10 no reverse transcriptase during primer extension; lane 1,9 no DNA during in vitro transcription. In lanes 4, 12 the primer extension product is indicated by an arrow. Lane 5-8, 13-16 sequencing reactions. (D) Secondary structure of 51A RNA from *mfold* (http://mfold.bioinfo.rpi.edu/).

Fig. 9 Occupancy at 51A promoter region. Cross-linked chromatin from SHSY5Y and SKNBE cells was immunoprecipitated with anti-SNAPc, anti-Brf2 or anti-Pol III antibodies and analyzed by PCR amplification for enrichment of 51A promoter. Primer specific to GAPDH exon 2 were used as internal standard control. The pol III-type 3 U6 snRNA promoter was also analyzed as a positive interaction control.

Conclusion

Pol III transcriptome appeared as a group of ncRNA, with a few well-known abundant members and an unexplored realm of non abundant transcripts. In past few year computational searches for ncRNA sequences and DNA regulatory regions in whole genome has expanded the knowledge concerning the ncRNAs and their cellular roles. Starting from these observations, our laboratory performed a screening in the human genome for elements typical of snRNA gene promoters (PSE and DSE) and identified a set of about 30 novel putative pol III-transcribed units (Pagano et al, 2007). In these work we selected four of these identified units (17A, 38A, 29A, 51A) to be characterized by experimental determinations.

We first tested their transcription properties to verify if core promoter elements were actually able to direct efficient pol III transcription. Though *in vitro* transcription and primer extension analyses, we defined the boundary of two novel ncRNAs, 17A and 51A, 171 nt and 423 nt in length respectively. Correct *in vitro* termination required at least a run of five T , as demonstrated in correctly-ended transcription of 17A . In 51A the sequence of four T was not sufficient for termination by Pol III, that read-through this sequence up to the next stop signal (6T). 38A and 29A RNAs were correctly initiated *in vitro* but their 3'-end weren't clearly identified; maybe alternative stop signals instead of canonical T sequences were required for termination.

Since these units were the results of an homology research for H1 RNA -like PSE elements within the human genome, all core promoters shared the presence of a PSE. This element was located at similar position in 17A, 38A and 29A promoters, upstream of TATA box elements located in canonical position for start site selection. PSE-TATA distance was instead slightly shorter than expected (Myslinski et al., 2001). As happens in other well known pol III type-3 genes, the PSE within promoter region of 17A, 38A and 29A, acted as a positive regulator of transcription; mutation or deletion of this element caused in fact a reduction of about four fold in expression of 17A natural transcript, or of an auxiliary reporter gene (7SK) attached at the promoter region of 38A and 29A. 51A RNA transcription was supported by a TATA-less PSE-dependent promoter, even if PSE position (-32 upstream the start site) seemed more canonical for a TATA box rather than a PSE element.

Core promoter elements and their position are summarized in table 1.

The factor binding to the PSE element has been best characterized in the human system and is variously known as PTF or SNAPc (see for a review Jawdekar and Henry, 2008). Chip experiments confirmed the binding of SNAPc at 38A and 51A promoters; this result, together with the observed occupancy of the same regions by other component of pol III machinery (Brf2 and pol III itself) suggested an active *in vivo* expression of these units.

Interestingly some of these novel units were internal to known protein-coding genes, one being in sense (17A) and two in antisense (38A, 51A), in regions where alternative splicing events takes place. A growing number of endogenous antisense RNA transcripts have been reported in the last years in eukaryotic organism where they exert control at various levels of gene expression, such as transcription, mRNA processing, splicing, stability transport and translation (Knee and Murphy, 1997; Hastings et al., 2000). In addiction recent evidence suggests that antisense intronic ncRNA may play a key role in a range of human diseases (Larvogna et al., 2004). The peculiar location of 17A (GPR51-intron3), 38A (KCNIP4-intron1) and 51A (SORL1-intron1) RNAs could suggest an involvement in splicing of these proteins , all directly or indirectly associated to amyloid precursor protein (APP) processing and therefore possibly to Alzheimer disease (AD).

As further intriguing observation the 29A RNA is located in a chromosomal region associated to a tumor suppressor activity, and was identified as an *Alu*. *Alu* RNAs are known as posttranscriptional regulator of gene expression, for example by affecting protein translation, alternative splicing and mRNA stability (Hasler and Strub, 2007).

Further studies will be required to better characterize these transcription units *in vivo* and the function of their products, that are suggestive of an increasing complexity of regulation by pol III transcriptome.

Transcriptional unit	PSE position	TATA position	PSE-TATA distance
17A	-56	-27	11
38A	-64	-23	15
29A	-62	-23	17
51A	-32	/	/

Table 1. Core promoter elements and their position within 17A, 38A, 29A, and 51A transcriptional units

References

Abou Elela S, Ares M (1998) Depletion of yeast RNase III blocks correct U2 39 end formation and results in polyadenylated but functional U2 snRNA. *EMBO J* 17,3738–3746.

Angermayr M, Oechsner U, Bandlow W. (2003) Reb1p-dependent DNA bending effects nucleosome positioning and constitutive transcription at the yeast profilin promoter. *J Biol Chem.* 278, 17918-26.

Basehoar AD, Zanton SJ, Pugh BF. (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell*. 116, 699-709.

Beinoraviciūte-Kellner R, Lipps G, Krauss G. (2005) In vitro selection of DNA binding sites for ABF1 protein from Saccharomyces cerevisiae. *FEBS Lett*. 579, 4535-4540.

Boyd D, Greger H, Murphy S (2002) In vivo footprinting studies suggest a role for chromatin in transcription of the human 7SK gene. *Gene* 247,33–44.

Chanfreau G, Elela SA, Ares M Jr, Guthrie C. (1997) Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev*. 11, 2741-2751.

Cheng C, Yeung C, Hoo R, Chow B, Leung (2002) Oct-1 is involved in the transcriptional repression of the gonadotropin-releasing hormone receptor gene. *Endocrinology*. 143, 4693-4701.

Christov CP, Gardiner TJ, Szuts D, Krude T (2006) functional requirement of non coding Y RNAs for human chromosomal DNA replication. *Mol Cell. Biol.* 26, 6993-7004.)

Connelly S, Filipowicz W (1993) Activity of chimeric U small nuclear RNA (snRNA)/mRNA genes in transfected protoplasts of Nicotiana plumbaginifolia: U snRNA 39-end formation and transcription initiation can occur independently in plants. *Mol Cell Biol* 13:6403–6415.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A. and Johnston, M. (2003) Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science*, 301, 71-76.

Dewang Z, Frendewey D, Lobo Rupper S (1999) Pac1p, an RNase homolog, is required for formation of the 3’ end of U2 snRNA in Schizosaccharomyces pombe. *RNA* 5, 1083-1098.

Dignam JD, Lebovitz RM, Roeder RG. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian. *Nucleic Acids Res*. 11, 1475-1489.

Dittmar, K.A. (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet*. 2, e221

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* 318,1777–1779.

Emerson B, Roeder R, (1984) DNA sequences and transcription factor interactions of active and inactive forms of mammalian 5 S RNA genes. *J. Biol. Chem.* 259, 7926-7935.

Eschenlauer JB, Kaiser MW, Gerlach VL, Brow DA. (1993) Architecture of a yeast U6 RNA gene promoter. *Mol Cell Biol.* 13, 3015-26.

Ford E, Strubin M, Herandez N (1998) The Oct-1 POU domain activates snRNA gene transcription by contacting a region in the SNAPc largest subunit that bears sequence similarities to the Oct-1 coactivator OBF-1. *Genes Dev.* 12, 3528-3540.

Fourel G, Miyake T, Defossez P, Li R, Gilson E. (2002) General regulatory factors as genome partitioners. *J. Biol. Chem.* 277, 41736–41743.

Frank DN, Roiha H, Guthrie C. (1994) Architecture of the U5 small nuclear RNA. *Mol Cell Biol*., 14, 2180-2190.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. (2003) Global analysis of protein expression in yeast. *Nature*, 425, 737-741.

Goncalves PM, Griffioen G, Minnee R, Bosma M, Kraakman LS, Mager H, Planta RJ. (1995) Transcriptional activation of yeast ribosomal protein genes requires additional elements apart from binding sitee for Abf1p or Rap1p. *Nucleic Acids Res*, 23, 1475-1480.

Goodrich JA, Kugel JE (2006) Non-coding-RNA regulators of polymerase II transcription. *Nature* 7, 612-616.

Gridasova A, Henry RW (2000) The retinoblastoma tumor suppressor protein targets dinstinc general transcription factors to regulate RNA polymerase III gene expression. *Mol. Cell. Biol.* 20, 9182-9191.

Gu L, Esselman WJ, Henry RW (2005) Cooperation between small nuclear RNA-activating protein complex (SNAPC) and TATA-box-binding protein antagonizes protein kinase CK2 inhibition of DNA binding by SNAPC. *J. Biol. Chem*. 280,27697–27704.

Häsler J, Samuelsson T, Strub K. (2007) Useful 'junk': Alu RNAs in the human transcriptome. Cell Mol Life 64, 1793-800.

Hastings M, Ingle H, Lazar M, Munroe S (2000) Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. *J. Biol. Chem.* 275, 11507-11513.

Hernandez N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem.* 276,26733-26736.

Hernandez N, Lucito R (1988) Elements required for transcription initiation of the human U2 snRNA gene coincide with elements required for snRNA 3′ end formation, *Embo. J.* 7, 3125–3134.

Hess J, Perez-Stable C, Wu GJ, Weir B, Tinoco I Jr, Shen CK (1985) End-to-end transcription of an Alu family repeat. A new type of polymerase-III-dependent terminator and its evolutionary implication. *J Mol Biol.* 184,7-21.

Hinkley CS, Hirsh H, Gu L, Lamere B, Henry RW (2003) The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition. *J. Biol. Chem*. 278, 18649-18657.

Hirsch HA, Jawdekar GW, Lee KA, Gu L, Henry RW (2004) Distinct mechanisms for repression of RNA polymerase III transcription by the retinoblastoma tumor suppressor protein*, Mol. Cell Biol*. 24,5989–5999

Hu P, Samudre K, Wu S, Sun Y, Hernandez N (2004) CK2 phosphorylation of Bdp1 executes cell cycle-specific RNA polymerase III transcription repression. *Mol. Cell* 16, 81–92.

Ito H, Fukuda Y, Murata K, Kimura A. (1983) Transformation of intact yeast cells treated with alkali cations. *J Bacteriol*.153,163-168.

Jarrous N, Reiner R (2007) Human RNAse P: a tRNA processing enzyme and transcription factor. *Nucleic Acids Res.* 35, 3519-3524

Jawdekar G, Henry RW (2008) Transcriptional regulation of human small nuclear RNA genes. *Biochimica et Biofisica acta.* 1779, 295-305.

Jawdekar G, Hanzlowsky A, Hovde S, Jelencic B, Feig M, Geiger JH, Henry RW (2006) The unorthodox SNAP50 zinc finger domain contributes to co-operative promoter recognition by human SNAPc. *J. Biol. Chem*. 281, 31050-31060.

Jensen RC, Wang Y, Hardin SB, Stumph WE. (1998) The proximal sequence element (PSE) plays a major role in establishing the RNA polymerase specificity of Drosophila U-snRNA genes. *Nucleic Acids Res*. 26, 616-22.

Jonsson M, Björntorp Mark E, Brantsing C, Brandner JM, Lindahl A, Asp J. (2004) Hash4, a novel human achaete-scute homologue found in fetal skin. *Genomics.* 84:859-66.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B. and Lander, E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, 423, 241-254.

Knee R, Murphy P (1997) Regulation of gene expression by natural antisense RNA transcripts. *Neurochem. Int*. 31, 379-392.

Kwek KY, Murphy S, Furger A, Thomas B, Kimura H, Proudfoot NJ, Akoulitchev A (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat. Struct. Biol.* 9, 800-805.

Kuehner, J.N. & Brow, D.A.(2006) Quantitative analysis of i*n vivo* initiator selection by yeast polymerase II supports a scanning model *J. Biol. Chem.* 281, 14119–14128.

Kuhlman TC, Cho H, Reinberg D, Hernandez N. (1999) The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter.*Mol. Cell. Biol*. 19, 2130–2141.

Kuras, L. (2004) Characterization of protein-DNA association in vivo by chromatin immunoprecipitation. *Methods Mol. Biol.*, 284, 147-162.

Kuras, L. and Struhl, K. (1999) Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature*, 399, 609-613.

Kurdistani S, Grunstein M (2003) In vivo protein-protein and protein-DNA crosslinking for genomewide binding microarray. *Methods.* 31, 90-95.

Lascaris RF, De Groot E, Mager WH, Planta RJ. (2000) Different roles for Abf1p and T-rich promoter element in nucleosome organization of the yeast RPS28A gene. *Nucleic Acids Res*. 28, 1390-1396.

Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson C M, Casari G, (2004) In search of antisense. *Trends Biochem. Sci*. 29, 88-94.

Lee W, Tillo D, Bray N, Morse RH, Davis R, Hughes T, Nislow C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics* 39, 1235-1244

Lin, Y. Brosius J, Tiedge H (2001) Neuronal BC1 RNA: co-expression with growth-associated protein-43 messenger RNA. *Neuroscience* 103, 465–479

Liko D, Slattery MG, Heideman W. (2007) Stb3 binds to ribosomal RNA processing element motifs that control transcriptional responses to growth in Saccharomyces cerevisiae. *J Biol Chem*. 282,26623-26628.

Lobo SM, Hernandez N (1989) A 7 bp mutation converts a human polymerase II snRNA promoter into an RNA polymerase III promoter. *Cell* 58, 55-67.

Mannhaupt G, Schnall R, Karpov V, Vetter I, Feldmann H. (1999) Rpn4 acts as a transcription factor by bindig PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. *FEBS Lett* 450, 27-34.

Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol*. 8, 209-220.

Mattaj IW, Dathan NA, Parry HD, Carbon P, Krol A (1988) Changing the RNA polymerase specificity of U snRNA genes promoters. *Cells* 55, 435-442.

Mattai IW (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* 46, 905-911.

Mattick JS (2004) RNA regulation: a new genetic? *Nat. Rev. Genet*. 5,316-323.

Mencía M, Moqtaderi Z, Geisberg JV, Kuras L, Struhl K. (2002) Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Mol Cell*. 9, 823-833.

McCutcheon, J.P. and Eddy, S.R. (2003) Computational identification of non-coding RNAs in Saccharomyces cerevisiae by comparative genomics. *Nucleic Acids Res*, 31, 4119-4128.

Mittal V, MA B, Hernandez N (1999) SNAPc: a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 Pou domain. *Genes Dev.* 13, 1807-1821.

Mittal V, Cleary A, Herr W, Hernandez N (1996) The Oct-1 POU-specific domain can stimulate small nuclear RNA gene transcription by stabilizing the basal transcription complex SNAPc, *Mol. Cell Biol*. 16, 1955–1965

Miyake T, Reese J, Loch CM, Auble DT, Li R. (2004) Genome-wide analysis of ARS (autonomously replicating sequence) binding factor 1(Abf1p)-mediated transcriptional regulation in Saccharomyces cerevisiae*. J Biol Chem.* 279,34865-34872.

Morse RH (2000) RAP, RAP, open up! New wrinkles for RAP1 in yeast. *Trends Genet* 16, 51-53.

Myslinski E, Amè J, Krol A, Carbon P. (2001) An unusually compact external promoter for RNA polymeras III transcription in the human H1RNA gene. *Nucleic Acids Res*. 29, 2502-2509.

Myslinski E, Krol A, Carbon P (1998) ZNF76 and ZNF143 are two human homologs of the transcriptional activator staf. *J. Biol. Chem.* 273, 21998–22006.

Murphy S, Liegro C, Melli M. (1987) The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* 51, 81-87.

Mus E, Hof P, Tiedge H (2007) Dendritic BC200 RNA in aging and in Alzheimer's disease. *Proc. Natl. Acad. Sci*. 104, 10679–10684

Nguyen VT, Kiss T, Michels A, Bensaude O (2001) 7SK small nuclear RNA binds and inhibits the activity of CDK9/cyclyn T complexes. *Nature* 414, 317-322.

Nilsen TW (2003) The spliceosome: the most complex macromolecular machine in the cell? *BioEssays* 25, 1147-1149.

Pagano A, Castelnuovo M, Tortelli F, Ferrari R, Dieci G, Cancedda R. (2007) New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. *PLoS Genet.* 3, 174-184.

Patterson B, Guthrie C. (1987) An essential yeast snRNA with a U5-like domain is required for splicing in vivo. *Cell* 49, 613-624.

Pavesi A, Conterio F, Bolchi A, Dieci G, Ottonello S (1994) Nucleic. Acids Res. 22, 1247-1256.

Sadowski CL, Henry RW, Kobayashi R, Hernandez N (1996) The SNAP45 subunit of the small nuclear RNA (snRNA) activating protein complex is required for RNA polymerase II and III snRNA gene transcription and interacts with the TATA box binding protein, *Proc. Natl. Acad. Sci*. 93,4289–4293.

Schmitt, M.E., Brown, T.A. and Trumpower, B.L. (1990) A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Res*. 18, 3091-3092.

Schroeder SC, Weil A. (1998) Genetic test of the role of Abf1p in driving transcription of the yeast TATA box binding protein-encoding gene, SPT15. *J Biol Chem.* 273,19884-19891.

Schramm L, Pendergrast PS, Sun Y, Hernandez N (2000) Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters. *Genes Dev.* 14,2650–2663.

Schramm L, Hernandez N (2002) Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* 16, 2593–2620.

Segref A, Mattaj I, Ohno W (2001) The evolutionary conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export. *RNA* 7, 351-360

Shuster EO, Guthrie C. (1988) Two conserved domains of yeast U2 snRNA are separated by 945 nonessential nucleotides. *Cell.* 55, 41-48.

SinghR, Gupta S, Reddy R. (1990) Capping of mammalian U6 snRNA in vitro is directed by a conserved stem-loop and AUAUAC sequence: conversion of a noncapped RNA into a capped RNA. *Mol. Cell. Biol.* 10, 939-946.

Sylvain E, Murphy S (2008) Role of the C-terminal domain of RNA polymerase II in expression of small nuclear RNA genes. *Biochem. Soc. Trans.* 36, 537-539.

Thomann H, Schmutzler C, Hudepohl U, Blow M, Gross HJ. (1989) Genes, variant genes and pseudogenes of the human $tRNA^{val}$ gene family expression and pre-tRNA maturation in vitro. *J Mol. Biol* 209,505-523.

Vilalta, A. et al. (1994) The rat vault RNA gene contains a unique RNA polymerase III promoter composed of both external and internal elements that function synergistically. J. Biol. Chem. 269, 29752–29759

Waibel F, Filipowicz W. (1990) RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. *Nature* 346, 199-202.

Wang KL, Warner JR. (1998) Positive and negative autoregulation of REB1 transcription in Saccharomyces cerevisiae. *Mol Cell Biol*., 18, 4368-76.

Wendelburg BJ, Marzluff WF (1992) Formation of the 3’ end of sea urchin U1 small nuclear RNA occurs independently of the conserved 3’ box and on transcripts initiated from a histone promoter. *Mol Cell Biol* 12, 4132–4141.

Wong MW, Henry RW, Ma B, Kobayashi R, Klages N, Matthias P, Strubin M, Hernandez N (1998) The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1, *Mol. Cell. Biol*.18, 368-377

Zamrod Z, Tyree CM, Song Y, Stumph WE. (1993) In vitro transcription of a Drosophila U1 small nuclear RNA gene requires TATA box-binding protein and two proximal cis-acting elements with stringent spacing requirements. *Mol Cell Biol.* 9,5918-27.

Zhao X, Pendergrast P, Hernandez N (2001) A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain. *Mol. Cell* 7, 539–549.

Zhou D, Lobo-Ruppert SM. (2001) Transcription of the Schizosaccharomyces pombe U2 gene in vivo and in vitro is directed by two essential promoter elements. *Nucleic Acids Res*. 29, 2003-11.