

UNIVERSITÀ DEGLI STUDI DI NAPOLI
“ FEDERICO II ”



DOTTORATO DI RICERCA IN
GENETICA
E
MEDICINA MOLECOLARE
XVII CICLO

“Un’analisi genomica per l’identificazione di nuovi geni che codificano piccoli RNA nucleolari in *Drosophila melanogaster* ”

Tutore

Ch.ma Prof.ssa
Maria Furia

Candidato

Dr.ssa
Maria Carmela Accardo

Coordinatore:

Ch.mo Prof.
Carmelo Bruno Bruni

Anno Accademico 2005/2006

Ai miei genitori

INDICE

1.INTRODUZIONE	4
1.1 Una categoria in espansione: la classe dei geni che non codificano per proteine (“non coding RNA genes”)	4
1.2 Classificazione degli snoRNA.....	20
1.3 Organizzazione genomica e biogenesi degli snoRNA.....	26
1.4 Localizzazione cellulare degli snoRNA	34
1.5 Composizione e formazione delle particelle snoRNP	36
1.6 Origine evolutiva degli snoRNA dagli archebatteri agli eucarioti	42
1.7 Funzioni degli snoRNA	45
2. RISULTATI.....	50
2.1 Introduzione al lavoro sperimentale	50
2.2 Impostazione del lavoro sperimentale	55
2.3 Identificazione, tramite analisi bioinformatica, di putativi geni codificanti per snoRNA nel genoma di <i>Drosophila melanogaster</i>	65
2.4 Convalida sperimentale dei putativi geni identificati	76
2.4.1 Parte I: Analisi dei putativi geni identificati in base alla presenza di siti di metilazione conservati fra le sequenze degli rRNA di <i>Saccharomyces cerevisiae</i> e di <i>Drosophila melanogaster</i>	76
2.4.1.2 Geni a localizzazione intergenica	88
2.4.1.3 Geni a localizzazione intronica.....	92
2.4.1.4 Geni a localizzazione esonica.....	99
2.4.1.5 Geni localizzati in regioni soggette a splicing alternativo.....	112

2.4.2 Parte II: Analisi dei putativi geni identificati in base alla presenza di siti di metilazione conservati fra le sequenze degli rRNA di <i>Drosophila melanogaster</i> e <i>Homo sapiens</i>	114
2.4.2.1 Geni a localizzazione intergenica.....	117
2.4.2.2 Geni a localizzazione intronica.....	125
2.4.2.3 Geni a localizzazione esonica.....	133
2.4.2.4 Geni localizzati a cavallo di una giunzione esone/introne del gene ospite.....	139
2.5 Nuovi RNA non-coding?.....	146
2.6 Ricerca di nuovi putativi target.....	161
3. DISCUSSIONE.....	172
4. MATERIALI E METODI.....	191
4.1 Ceppo di <i>Drosophila melanogaster</i> utilizzato.....	191
4.2 Terreno per la coltura in massa di <i>D.melanogaster</i>	191
4.3 Mantenimento dei ceppi.....	192
4.4 Ricerca in banca dati e analisi di sequenza.....	192
4.5 Estrazione di DNA a alto peso molecolare.....	195
4.6 Amplificazione del DNA con la tecnica della PCR (reazione a catena della polimerasi).....	196
4.7 Recupero dei frammenti amplificati per PCR mediante l'uso del kit Nucleospin extract (M-medical).....	197
4.8 Clonaggio di frammenti di PCR (pMOS Blue Blunt ended cloning kit)..	197
4.9 Protocollo per la minipreparazione del DNA plasmidico.....	199
4.10 Protocollo per la midipreparazione di DNA plasmidico.....	200
4.11 Protocollo per la reazione di sequenza del DNA plasmidico (kit USB). 201	

4.12 Elettroforesi su gel di poliacrilammide in condizioni denaturanti per il sequenziamento dei cloni ottenuti	202
4.13 Trasferimento del gel su carta, esposizione per autoradiografia e sviluppo	203
4.14 Protocollo per l'estrazione dello RNA da adulti ed embrioni di D. melanogaster	203
4.15 Protocollo per l'estrazione dello RNA da larve di D. melanogaster (di 1°, 2° e 3° stadio)	204
4.16 Elettroforesi dello RNA e trasferimento su filtro di nylon (Northern blot)	205
4.17 Preparazione in vitro di una molecola di DNA radioattiva copiata su di un templatato noto: NICK TRANSLATION	206
4.18 Protocollo di preibridazione ed ibridazione degli RNA trasferiti sul filtro di nylon	207
4.19 Primer extension	207
4.20. Elenco delle soluzioni utilizzate:	210
5.BIBLIOGRAFIA	212

1.INTRODUZIONE

1.1 Una categoria in espansione: la classe dei geni che non codificano per proteine (“non coding RNA genes”)

L'interesse verso l'attività trascrizionale del genoma è stata focalizzata negli ultimi vent'anni, almeno per quanto riguarda gli eucarioti superiori, quasi esclusivamente sulla scoperta dei geni che codificano per proteine, parallelamente l'RNA è stato considerato a lungo una molecola accessoria, coinvolta prevalentemente nella mediazione dei processi di trascrizione e traduzione. Quest'idea può oggi essere considerata approssimativamente corretta nei Procarioti, i cui genomi sono quasi interamente composti da sequenze codificanti compatte, con poche eccezioni di geni che codificano per RNA con funzioni regolative e che occupano in genere non più dell' 1% dell' intero genoma. Negli organismi superiori, al contrario, essa è stata messa in crisi dagli esiti del progetto di sequenziamento del genoma, in quanto è emerso che il numero di geni che codificano per proteine è notevolmente più basso di quello atteso. E' oggi evidente che, mentre i genomi dei procarioti posseggono meno del 25% di DNA non coding, quelli degli eucarioti semplici ne posseggono almeno il 50%, fino a raggiungere il 98.5% nell'uomo (il cui genoma ha però dimensioni tre volte maggiori rispetto ai procarioti; Levane and Tjian, 2003; Buchler *et al.*, 2003). Un'ampia frazione del genoma degli eucarioti superiori dà infatti origine a trascritti poliadenilati che non codificano per proteine, mentre studi di genomica comparata indicano che la quantità di DNA non codificante, rispetto a quella che codifica per proteine cresce in funzione della complessità di un organismo. Nei procarioti, il numero di molecole di natura proteica coinvolte nei processi regolativi appare crescere in

maniera logaritmica con l'aumentare delle dimensioni del genoma ed il punto in cui il sistema sembra arrivare a saturazione è vicino al limite massimo di grandezza osservato per un genoma batterico, il che conferma una limitazione della complessità genomica e funzionale di questi organismi, dovuta al mancato adattamento dei loro meccanismi di regolazione (Croft *et al.*, 2003; Gagen and Mattick, 2004). Negli organismi più complessi, la quantità crescente di trascritti non codificanti per proteine ha portato a considerare l'ipotesi che gli RNA non coding possano essere i componenti di un sistema di regolazione parallelo a quello guidato dalle proteine e forse molto più sofisticato. Da qui l'idea che negli organismi superiori si sia verificata una riorganizzazione dei meccanismi di regolazione, fondamentale per l'evoluzione e per lo sviluppo di una maggiore complessità del genoma, basata sull'ipotesi che gli RNA non coding rappresentino la frazione principale di molecole geneticamente attive presenti nel genoma degli organismi più complessi (Mattick, 2001). Originariamente, il termine "RNA non coding" è stato utilizzato per indicare esclusivamente molecole di RNA eucariotici poliadenilati e dotati di CAP al 5', cioè tipici trascritti della RNA polimerasi II che mancavano di uno schema di traduzione aperto (ORF). Al giorno d'oggi questa definizione è di fatto estesa ad indicare un'ampia e variegata categoria di trascritti che non codificano per proteine e la cui classificazione è basata su criteri (quali la loro localizzazione cellulare, la funzione che svolgono, la loro lunghezza o il loro coefficiente di sedimentazione) diversi tra loro e non ancora integrati, per cui può accadere che uno stesso RNA non coding venga indicato con nomi differenti ed appaia classificato in più di una categoria.

Nei Procarioti, le prime molecole di RNA non coding, comunemente indicate come sRNA (small RNA o small regulatory RNA) sono state identificate per caso durante esperimenti di marcatura metabolica o analisi diretta mediante diverse procedure di frazionamento (Ikemura and Dahlberg, 1978; Lee *et al.*, 1978; Andersen *et al.*, 1987; Altuvia *et al.*, 1997). La prima ricerca su scala genomica condotta nei microrganismi è partita dall'osservazione che i piccoli RNA identificati derivavano da regioni intergeniche ed erano generalmente conservati in specie strettamente correlate (Wassarman *et al.*, 2001; Rivas *et al.*, 2001). Studi successivi sono stati indirizzati verso l'individuazione, all'interno di regioni intergeniche, di promotori e di sequenze e strutture proprie dei terminatori della trascrizione rho-indipendenti (Argaman *et al.*, 2001; Chen *et al.*, 2002). In un altro approccio, invece, è stata sfruttata la possibilità di associazione degli sRNA batterici con proteine RNA-binding note, come nel caso degli sRNA che legano lo "chaperone" Hfq; questa ricerca ha portato all'identificazione di numerosi RNA non coding che erano sfuggiti alle precedenti analisi (Zhang *et al.*, 2003; Wilderman *et al.*, 2004; Lenz *et al.*, 2004). In *E.coli*, il batterio nel quale è stato condotto il maggior numero di analisi genetiche, il numero di geni codificanti per sRNA era pari a 10 fino al 1999 (Wassarman *et al.*, 1999), per poi salire a 62 nel 2003 (Hersherberg *et al.*, 2003; Vogel *et al.*, 2003) ed essere stimato tra 118 e 260 in un recente studio basato sull'analisi di sequenze conservate e di regioni intergeniche (Zhang *et al.*, 2004). Gli RNA non coding procariotici possono essere classificati in 3 categorie, sulla base delle loro caratteristiche funzionali: gli sRNA con funzione strutturale o enzimatica; i piccoli RNA coinvolti nel controllo della

qualità del processo di traduzione e la più ampia classe degli sRNA con funzioni regolative (Storz, 2005). Nella prima categoria rientrano 2 sRNA altamente conservati; la Ribonucleasi P, responsabile del processing all'estremità 5' dei precursori degli RNA transfer e di alcuni RNA ribosomali (Frank *et al.* 1998; Gopalan *et al.*, 2002), e l'RNA 4.5S coinvolto nel trasporto delle proteine attraverso le membrane (Herskovit *et al.*, 2000). Queste due molecole di piccoli RNA non coding sono l'unico esempio di sRNA essenziali conosciuti finora in *E.coli*. Alla seconda categoria appartiene il tmRNA di *E.coli*, trovato anche in numerosi altri batteri, così come nei genomi mitocondriali di alcuni eucarioti (Williams and Bartel, 1996; Williams, 2002). Il suo nome, tmRNA, riflette la sua capacità di funzionare sia come tRNA che come mRNA; questo RNA viene chiamato anche SsrA (small stable RNA), dal nome assegnatogli nella prima identificazione (Lee *et al.*, 1998). E' stato osservato che parte della sua struttura si organizza come un tRNA, acquisendo la capacità di legare un amminoacido (alanina; Kamine *et al.*, 1994). Quando il processo di traduzione si interrompe prematuramente, il tmRNA interviene, funzionando come un normale tRNA; la traduzione riprende però dal registro di lettura interno del tmRNA rilasciando un polipeptide con una estremità C-terminale di 11 aa chiamata SsrA-tag. La coda di 11 aa sintetizzata grazie all'intervento del tmRNA etichetta il polipeptide per la degradazione ad opera delle proteasi e libera il ribosoma dalla fase di stallo.

Nella classe più ampia di RNA non coding procariotici rientrano i piccoli RNA che hanno funzioni regolative. Tra questi è possibile operare un'ulteriore classificazione, cioè tra gli sRNA che interagiscono con proteine e gli sRNA

regolativi antisenso che agiscono attraverso interazioni RNA:RNA. In *E.coli*, attualmente sono 2 le proteine la cui attività è regolata dal legame con uno sRNA; la RNA polimerasi, che è bersaglio dello sRNA 6S (Wassarman *et al.*, 2000) e la proteina CsrA (Carbon storage regulatory protein) che regola l'accumulo del carbonio e la cui funzione è inibita dal legame di almeno 2 sRNA, CsrB e CsrC (Liu *et al.*, 1997; Weibacher *et al.*, 2002). Lo sRNA 6S è processato da un mRNA che codifica per un prodotto la cui funzione è sconosciuta (Hu *et al.*, 1985) ed è l'unico sRNA procariotico conosciuto che agisce sul processo trascrizionale piuttosto che a livello post-trascrizionale. Esso si lega ed inibisce la RNA polimerasi che contiene il fattore vegetativo sigma 70 e non ha effetto su quella che contiene il fattore stazionario sigma 8 o RpoS, per cui sembrerebbe andare ad impedire il riconoscimento del promotore durante la fase stazionaria. L'RNA 6S ha una struttura che ricorda quella di un promotore sigma 70, per cui è plausibile che l'interazione con la RNA polimerasi avvenga proprio nelle regioni in cui quest'ultima lega il promotore (Wassarman and Storz, 2000). La proteina CsrA, invece, agisce come inibitore della traduzione, legandosi alla 5'UTR del suo mRNA bersaglio. Gli sRNA CsrB e CsrC ostacolano l'azione regolativa di CsrA, in quanto si legano ad essa attraverso dei motivi di sequenza ripetuti capaci di legare CsrA, simili a quelli presenti sul suo mRNA target (Romeo *et al.*, 1998). Un numero significativo di sRNA agisce come regolatore antisenso e va ad alterare la stabilità o ad inibire la traduzione di molecole di mRNA mediante la formazione di un doppio filamento RNA:RNA. Una piccola parte di questi sRNA agisce in *cis*, in quanto è codificata sul filamento opposto a quello del messaggero bersaglio;

questo meccanismo d'azione è tipico delle molecole di RNA antisenso dei fagi e dei plasmidi (Brantl, 2002; Wagner and Brantl, 1998), anche se alcuni sRNA che agiscono in *cis* sono stati trovati anche nei batteri (SokA e RdID). Essi fanno parte di sistemi di stabilizzazione di plasmidi replicanti nelle cellule di *E.coli* e agiscono come antidoti contro l'azione di tossine stabili che uccidono le cellule "plasmid-free". SokA inibisce la traduzione del messaggero del gene *hok* (host killing) che codifica per una tossina associata alla membrana lunga 52 aa che causa danni irreversibili alla membrana cellulare ed è per questo letale per la cellula ospite (Gerdes *et al.*, 1997), mentre RdID agisce sul messaggero del gene LdR-D (Long direct Repeat-D) che codifica per un peptide di 35 aa la cui iper-espressione causa la condensazione del nucleotide e la conseguente morte della cellula ospite (Kawano, 2002). La maggior parte degli sRNA antisenso codificati dai batteri agisce in *trans*, in quanto mappa lontano dal proprio bersaglio e necessita dell'intervento di una molecola di RNA "chaperone" chiamata Hfq. Una recente caratterizzazione di questo RNA ha rivelato che esso è omologo alle proteine Sm e Sm-like che formano il "core" dei complessi di splicing e di degradazione degli mRNA nelle cellule eucariotiche e negli archeobatteri (per una review, vedi He and Parker, 2000; Pannone and Wolin, 2000). Come le proteine Sm e Sm-like, Hfq si lega a sequenze ricche in AU. Quattro sRNA, lunghi approssimativamente 100 nt, che legano Hfq sono stati studiati in dettaglio: DsrA, OxyS, Spot42 e RyhB. Tutti vengono codificati da geni con un singolo promotore e ognuno di essi finisce con un terminatore rho-indipendente; essi vengono solitamente espressi in risposta a segnali indotti da situazioni di stress. Oxy-R è legato allo stress

ossidativo, RyhB è regolato dal repressore Fur, un regolatore globale dei geni che modulano l'acquisizione del ferro; la sintesi di Spot42 è regolata dalla proteina cAMP e dal suo recettore ed infine, DsrA viene prodotto in risposta ad abbassamenti della temperatura (Altuvia *et al.*, 1997; Massè and Gottesman, 2002; Møller *et al.*, 2002). La funzione dello "chaperone" Hfq sembra non essere limitata al legame con gli sRNA. Infatti è stato visto che esso lega anche regioni ricche in AU sull'mRNA target favorendo, *in vitro*, il legame tra gli sRNA e i loro target (Møller *et al.*, 2002; Zhang *et al.*, 2002; Lease and Woodson, 2004). Non è ancora ben chiaro se l'appaiamento con il messaggero ne induca la degradazione o se vada ad inibire il processo di traduzione; è possibile che il processo degradativo sia successivo a quello di inibizione della traduzione.

Fino a questo momento, gli studi sulle molecole di RNA non coding sono stati condotti fondamentalmente in parallelo nei sistemi procariotici ed eucariotici e l'analisi dei meccanismi d'azione di molti ncRNA, sia nei batteri che negli organismi superiori, ha evidenziato una serie di sorprendenti similitudini in tutti gli organismi esaminati. Negli eucaroti, la combinazione di metodi sperimentali e computazionali fondati prevalentemente sull'analisi di sequenze e strutture conservate presenti nelle molecole di RNA non coding già identificate (Lagos-Quintana *et al.*, 2001; Lee and Ambros, 2001; Elbashir *et al.*, 2001; El-Mabrouk and Lisacek, 1996; Lowe and Eddy, 1997; 1999; Le *et al.*, 1988; Seffens *et al.*, 1999; Olivas *et al.*, 1997), ha portato all'identificazione di due classi di RNA funzionali, distinguibili sulla base delle funzioni svolte: gli RNA non coding costitutivi (housekeeping ncRNA), che sono espressi costitutivamente e sono

necessari per le normali funzioni e la sopravvivenza della cellula, e gli RNA non coding regolativi (regulatory ncRNA) che sono espressi solo in certi stadi dello sviluppo o del differenziamento cellulare o in risposta a stimoli esterni (vedi Szymanski and Barcizewski, 2002; Morey and Avner., 2004). La classe degli RNA non coding ‘housekeeping’ comprende ovviamente gli RNA transfer (tRNA), gli RNA ribosomali (rRNA) impegnati nel processo di traduzione dell’informazione genetica ed i piccoli RNA nucleari (snRNA) coinvolti nel processo di splicing. In questa categoria rientra inoltre l’RNA associato alla telomerasi (TERC) che funge da stampo per la sintesi delle estremità dei cromosomi. TERC contiene motivi tipici di un’altra classe di RNA non coding, gli snoRNA H/ACA (che saranno descritti nel paragrafo successivo) ed elementi di sequenza essenziali per l’attività stessa della telomerasi (Mitchell *et al.*, 1999; Chen *et al.*, 2000).

Le molecole di RNA non coding di tipo regolativo, invece, possono agire nella regolazione a livello trascrizionale o post-trascrizionale, o come modulatrici dell’attività delle proteine ed infine intervenire come determinanti nella distribuzione di RNA e proteine all’interno della cellula (Szymanski and Barcszeski, 2002; Storz, 2002). Sulla base delle dimensioni è possibile identificare due grandi categorie di RNA con funzioni regolative: i lunghi RNA non coding (large ncRNA), ossia molecole di dimensioni superiori anche ai 10000 nt coinvolti prevalentemente in processi di silenziamento genico e rimodellamento della cromatina negli eucarioti superiori (Sleutels *et al.*, 2002; Erdmann *et al.*, 2000; Avner and Heard, 2001) ed i piccoli RNA non coding (small ncRNA), che comprendono la famiglia dei piccoli RNA guida (miRNA

e siRNA), i cui membri, lunghi circa 20-25 nucleotidi, regolano lo sviluppo in *C. elegans*, in *Drosophila melanogaster* e nei mammiferi (Lagos-Quintana *et al.*, 2001; Lee and Ambros, 2001; Ruvkun, 2001; Grosshans and Slack, 2002) ed altri piccoli RNA le cui dimensioni possono variare tra i 60 e i 500 nt e che dirigono le modifiche post-trascrizionali di molecole di RNA target (Bachellerie *et al.*, 2002; Kiss, 2002; Kiss *et al.*, 2004).

I lunghi RNA non coding vengono trascritti dai promotori della Polimerasi II e sono spesso sottoposti a splicing alternativo e poliadenilati, per cui potrebbero essersi evoluti da geni che hanno perso la capacità di codificare per proteine. Alcuni di essi, identificati negli insetti ed in cellule di mammifero, sono implicati nel silenziamento genico e nei processi di rimodellamento della cromatina di ampie regioni cromosomiche (Sleutels *et al.*, 2002; Erdmann *et al.*, 2000; Avner *et al.*, 2001). Ne sono un esempio Xist (X inactive specific transcript), e roX1/roX2 (RNA on the X), coinvolti nel fenomeno di compensazione del dosaggio del cromosoma X nei mammiferi ed in *Drosophila*, rispettivamente. Nello specifico, Xist va ad inattivare uno dei due cromosomi X presenti nelle cellule sessuali femminili (Avner and Heard, 2001), mentre roX1 e roX2 inducono l'iper-attivazione trascrizionale dell'unico cromosoma X nelle cellule XY di *Drosophila* (Smith *et al.*, 2000). Altre lunghe molecole di ncRNA sono coinvolte nel fenomeno di imprinting. Il primo gene di questo tipo ad essere identificato è stato il gene umano H19. Questo gene è localizzato sul cromosoma 11p15.5 adiacente al gene IGF2 (Insulin Like Growth Factor 2) ed entrambi sono coinvolti in un fenomeno definito di "imprinting "reciproco". H19 infatti mostra una specifica espressione

dell'allele paterno mentre IGF2 una specifica espressione di quello materno. Durante lo sviluppo, H19 viene trascritto ad alti livelli nella placenta, nell'embrione e nella maggior parte dei tessuti fetali, ma dopo la nascita la sua espressione appare significativamente ridotta (Looijenga *et al.*, 1997). Sebbene l'RNA H19 mostri, nei mammiferi, una significativa omologia di sequenza, non sono stati identificati schemi di traduzione (ORF), conservati anche se è stata osservata una comune struttura secondaria (Juan *et al.*, 2000) che è fondamentale per la sua funzionalità. E' stato proposto che i pattern di espressione "reciproci" mostrati dai due geni H19 e IGF2 durante lo sviluppo riflettano differenti ruoli svolti dai loro prodotti. IGF2 è un fattore di crescita che stimola la proliferazione cellulare, mentre l'RNA H19 potrebbe essere responsabile del differenziamento (Looijenga *et al.*, 1997). L'estremità al 5' del gene H19 contiene una regione caratterizzata da metilazione allele-specifica regolata da fenomeni di imprinting. L'allele materno che viene trascritto attivamente non è metilato, mentre l'allele paterno, inattivo, è metilato. La regione soggetta a questo tipo di modificazione differenziale si estende nel locus H19 del topo per 2 kb a monte del promotore del gene (Kanduri *et al.*, 2000) ed un pattern di metilazione simile è stato osservato anche per il gene H19 umano. Il modello proposto per questo fenomeno suggerisce che un isolatore trascrizionale localizzato a monte del gene H19 isola *Ins2* e *Igf2* dal suo enhancer. Quando la regione di controllo dell'imprinting (ICR: Imprinting Control Region) sul cromosoma di origine materna non è metilata, le 2 regioni ipersensibili all'attacco di nucleasi, sovrapposte a regioni ricche in CG conservate in topo ed in uomo, sono legate dall'isolatore cromatinico CTFC.

Quando CTFC occupa la regione ICR, viene bloccata l'attività della copia materna di Igf2 e H19 viene trascritto. L'evento di metilazione della regione ICR, indotto per effetto paterno, previene il legame del fattore CTFC inattivando l'isolatore trascrizionale e permettendo all'enhancer di H19 di attivare i geni Igf2 e Ins 2 (Hark *et al.*, 2000; Bell and Felsenfeld, 2000; Wilkins *et al.*, 2000; Juan *et al.*, 2000). Modificazioni nello stato di metilazione delle regioni localizzate a monte del gene H19 determinano la perdita dell'imprinting di H19 o di IGF2 e l'espressione biallelica di uno dei due geni può causare una crescita cellulare anomala caratteristica delle cellule neoplastiche (Ulaner *et al.*, 2003). Il coinvolgimento di H19 nello sviluppo del cancro sembrerebbe essere duplice; esso può agire o come gene soppressore di tumori, in quanto la sua espressione riduce la tumorigenicità e la crescita di alcune linee cellulari maligne (Hao *et al.*, 1993; Isfort *et al.*, 1997), o avere un'azione oncogenica in quanto risulta espresso ad alti livelli in alcune forme di cancro del polmone, della mammella e della vescica (Ariel *et al.*, 2000; Lottin *et al.*, 2002). Questi risultati contraddittori potrebbero essere spiegati dal fatto che l'espressione del gene H19 dipende dallo stato di sviluppo delle cellule o potrebbe essere dovuto all'espressione di forme derivanti da fenomeni di splicing alternativo che sono cellula o tessuto-specifiche (Lottin *et al.*, 2002) o ancora dall'espressione di specifiche proteine di legame al trascritto di H19 (Ioannidis *et al.*, 2004). Un altro esempio di trascritto non coding implicato nel fenomeno di imprinting genomico è *Air* (*Anti-sense* IGF2r RNA), un ncRNA che si origina dal secondo introne del gene IGF2r la cui espressione è correlata con la repressione di 3 geni che codificano per proteine (Sleutels *et al.*, 2002).

Altre lunghe molecole di ncRNA sembrano essere importanti per la sopravvivenza cellulare: tra queste, particolarmente interessanti sono quelle coinvolte nella genesi di alcune malattie umane. Infatti, alcuni ncRNA sono stati correlati a malattie umane neurodegenerative, come lo SCA8 legato all'ataxia spinocerebrale di tipo 8 (Mutsuddi *et al.*, 2004), e i ncRNA associati alla schizofrenia (DISC2: "Disrupted in Schizophrenia 2"; Millar *et al.*, 2001;) o ancora il gene *MALAT-1* che codifica per un RNA non coding lungo 8 kb associato al cancro al polmone (Ji *et al.*, 2003).

La categoria degli "small regulatory RNA" comprende invece i piccoli RNA citoplasmatici (scRNA) e due principali classi ben definite, ossia quella dei piccoli RNA guida, che include microRNA (miRNA) e gli small interfering RNA (siRNA), e quella dei piccoli RNA nucleolari (snoRNA). Esempi dei piccoli RNA citoplasmatici sono l'RNA SRP, componente principale della particella di riconoscimento del segnale (vedi Koch *et al.*, 2003) ed i piccoli RNA non coding BC1 e BC200, la cui funzione sembra essere in qualche modo legata al cancro. L'RNA BC1 è infatti espresso in maniera specifica nelle cellule tumorali (Chen *et al.*, 1997a), ma non nei corrispondenti tessuti normali. Allo stesso modo è stato osservato che l'RNA BC200 è espresso nei carcinomi umani della mammella, della cervice, dell'esofago e di altri organi, ma non nelle cellule sane (Chen *et al.*, 1997b). Un'iper-espressione di BC200 è stata osservata nelle forme invasive di carcinoma della mammella e questo ncRNA è stato recentemente utilizzato come nuovo marker molecolare per questo tipo di tumore (Iacoangeli *et al.*, 2004). Una delle classi più ampie di piccoli RNA regolativi eucariotici è rappresentata dai minuscoli RNA non coding (lunghi

21-25 nt) chiamati microRNA e dagli short interfering RNA (siRNA). I microRNA (definiti anche RNA antisenso *trans*-encoded) derivano da trascritti endogeni più lunghi (pri-miRNA) contenenti strutture a forcina lunghe circa 70 nt, e sono coinvolti nel silenziamento genico post-trascrizionale. Essi infatti mediano la repressione della traduzione di specifici mRNA e nelle cellule agiscono mediante un meccanismo che prevede la formazione di un appaiamento imperfetto con la 3'UTR del messaggero bersaglio (Lee *et al.*, 1993; Bartel and Chen, 2004; Ambros *et al.*, 2004; He and Sontheimer, 2004). Gli siRNA sono prodotti da lunghe molecole di RNA a doppio filamento, spesso di origine esogena, ed agiscono su sequenze bersaglio attraverso appaiamento perfetto, inducendone il silenziamento genico tramite il meccanismo dell'RNA interference (Cullen, 2004). Sia i miRNA che gli siRNA vengono processati attraverso meccanismi simili, che prevedono il taglio dei rispettivi precursori ad opera di endonucleasi che agiscono in maniera specifica su regioni a doppio filamento; nel caso dei miRNA, la struttura a forcina interna viene tagliata dalla endonucleasi Drosha, che agisce in coppia con un'altra proteina chiamata DGCR8 nell'uomo e Pasha in *Drosophila*; successivamente il pri-miRNA viene processato per azione di un'altra endonucleasi, DICER, che libera molecole di RNA a doppio filamento lunghe 22 nt caratterizzate da estremità 3'-protruding di 2 nt (Lee *et al.*, 2003). Anche i precursori degli siRNA vengono processati dall'enzima DICER, ma senza l'intervento di Drosha (vedi Bartel *et al.*, 2004; Cullen *et al.*, 2004). I piccoli RNA a doppio filamento generati, sia nel caso dei miRNA che degli siRNA, vengono successivamente incorporati nel complesso ribonucleoproteico

“RISC” che comprende un membro della famiglia delle proteine Argonata. In generale solo uno dei filamenti del duplex di RNA è inglobato nel complesso; la scelta di quale dei due filamenti debba essere ritenuto all’interno del RISC sembra sia determinata dalla stabilità relativa dell’estremità 5’ del duplex e sembra quindi che venga inglobata quella meno strettamente appaiata (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003). A questo punto il RISC forma un complesso con la molecola di RNA bersaglio ed a seconda del tipo di appaiamento che si stabilisce con l’RNA target si può indurre o la repressione della traduzione, come avviene per i miRNA che si appaiano in maniera imperfetta, o indurre il taglio del messaggero, nel caso degli siRNA, che si appaiano in maniera perfetta con l’RNA bersaglio (Zeng *et al.*, 2003). Il pathway di formazione del complesso RISC attivo appare quindi comune anche se le proteine della famiglia Argonata che si associano al complesso possono essere diverse. I primi due miRNA identificati sono stati *lin-4* e *let-7*, isolati in un’analisi genetica sul controllo temporale dello sviluppo in *C.elegans* (Lee *et al.*, 1993; Reinhart *et al.*, 2000). *lin-4* e *let-7*, lunghi rispettivamente 22 nt e 21 nt, si appaiano con la sequenza complementare all’estremità 3’ del messaggero target e questa interazione determina l’arresto della traduzione. Nelle piante, e finora in un solo caso anche nei mammiferi, è stato visto che per molti miRNA l’appaiamento con il messaggero bersaglio è perfetto e ciò ne induce anche la degradazione, attivando il pathway dell’RNA interference (Llave *et al.*, 2002; Hutvagner and Zamore, 2002a). Per i miRNA è stato evidenziato un coinvolgimento in numerosi processi di sviluppo negli animali e nelle piante (Seitz *et al.*, 2003). Fra l’altro è stato suggerito che l’alterazione del pattern di

espressione di alcuni miRNA potrebbe essere responsabile di cambi nel programma genetico delle cellule e riflettersi in una crescita cellulare tumorale (McManus, 2003). E' stato infatti dimostrato che diverse linee cellulari neoplastiche mostrano differenti pattern di espressione di alcuni miRNA, i cui profili di espressione cambiano fra cellule normali e maligne ed in funzione dell'origine dei tumori. E' inoltre interessante notare che in tutti i casi in cui è stata osservata una correlazione tra cancro e miRNA, l'espressione di queste piccole molecole risultava essere ridotta. Sembra quindi evidente che i miRNA potrebbero giocare un ruolo chiave nel mantenimento dell'omeostasi cellulare, andando ad inibire l'espressione di geni coinvolti nel processo di carcinogenesi. Un esempio è rappresentato dai miRNA umani miR-143 e miR-145, per i quali è stata osservata una riduzione del livello basale di produzione di molecole di miRNA mature nelle fasi adenomatose e cancerose della neoplasia colonrettale (Michael *et al.*, 2003). L'espressione dei miRNA cambia anche in risposta all'induzione del differenziamento cellulare, come dimostrato nel caso delle cellule leucemiche umane (Kasashima *et al.*, 2004). Una delle regioni del cromosoma 13p14 nell'uomo che è frequentemente deleta nelle cellule B leucemiche (B-CCL: B-Cell Chronic Lymphocytic leukaemia) ospita 2 geni che codificano per miRNA: miR-15 e miR-16 (Cimmino *et al.*, 2005); non è ancora chiaro tuttavia se questi miRNA siano o meno coinvolti nel processo di differenziamento delle cellule B. Un' ulteriore evidenza del legame esistente tra i miRNA e cancro è sostenuta dalla scoperta che mutazioni nei geni necessari per la biosintesi di queste molecole causano difetti dello sviluppo ed insorgenza di tumori (Hutvagner and Zamore, 2002b; Carmell *et al.*, 2002).

Un'altra famiglia di piccoli ncRNA eucariotici è quella degli snoRNA (small nucleolar RNA), che sono implicati nella maturazione o modificazione di RNA target tramite appaiamento specifico di sequenza. In questi ultimi anni sono state individuate e caratterizzate molte molecole di snoRNA in diversi organismi, quali il lievito *S. cerevisiae* (Lowe and Eddy, 1999), le piante *A. thaliana* ed *O. sativa* (Barneche *et al.*, 2001; Brown *et al.*, 2001; Liang *et al.*, 2002), la *D. melanogaster* (Yuan *et al.*, 2003; Accardo *et al.*, 2004), il topo ed il ratto (Hüttenhofer *et al.*, 2001; Cavallé *et al.*, 2001) ed infine l'Uomo (Vitali *et al.*, 2003). Recentemente gli snoRNA sono stati identificati anche in organismi eucariotici primitivi, quali *E. gracilis* (Russell *et al.*, 2004), *T. brucei* (Uliel *et al.*, 2004), *D. discoideum* (Aspegren *et al.*, 2004) e *G. lamblia* (Yang *et al.*, 2005). Infine, molecole snoRNA-like sono state caratterizzate anche negli archebatteri (Omer *et al.*, 2000). Questi ultimi dati, come vedremo in seguito in dettaglio, dimostrano l'importanza degli snoRNA e delle loro funzioni che risultano conservate nel corso dell'evoluzione e che risalgono ai progenitori comuni degli eucarioti e degli archebatteri.

1.2 Classificazione degli snoRNA

Negli eucarioti, quasi tutti gli snoRNA fino ad oggi individuati rientrano in due grandi classi: la classe degli snoRNA di tipo C/D e la classe degli snoRNA di tipo H/ACA. Questa suddivisione è basata sulla presenza di distinti motivi di sequenza (chiamati box C, D, H e ACA) conservati nel corso dell'evoluzione in tutte le specie studiate (Kiss, 2001). Gli snoRNA della classe C/D sono responsabili della metilazione dei gruppi 2'-*O*-idrossilici dei ribosi, mentre gli snoRNA della classe H/ACA dirigono la conversione delle uridine in pseudouridine nelle molecole degli RNA bersaglio (Nicoloso *et al.*, 1996; Ni *et al.*, 1997; Weinstein and Steitz, 1999). La lunghezza delle molecole di snoRNA di entrambe le classi varia mediamente fra circa i 70–350 nt.

Gli snoRNA di classe C/D contengono due corti motivi di sequenza conservati: la box C (5'RUGAUGA3') e la box D (5'CUGA3'), localizzati a pochi nucleotidi dalle estremità 5' e 3', rispettivamente (Fig.1A). I due motivi, però, si trovano generalmente ravvicinati in una struttura secondaria terminale, che coinvolge 4-5 nt ad entrambe le estremità, e che, nota importante, è critica per la stessa biogenesi e la localizzazione nucleolare della molecola di snoRNA. Il motivo C/D inoltre, come vedremo, è essenziale per la maturazione, la stabilità, la localizzazione nucleolare e la funzione degli snoRNA di questa classe (Cavaillé *et al.*, 1996; Caffarelli *et al.*, 1996; Lange *et al.*, 1998; Samarsky *et al.*, 1998; Villa *et al.*, 2000). Gli snoRNA di tipo C/D contengono talvolta, nella loro porzione centrale, una copia interna della box C, chiamata box C' (Kiss-Laszlo *et al.*, 1998) e una copia addizionale della box D, chiamata similmente box D', localizzata internamente (Tycowski *et al.*, 1996).

snoRNA di classe C/D

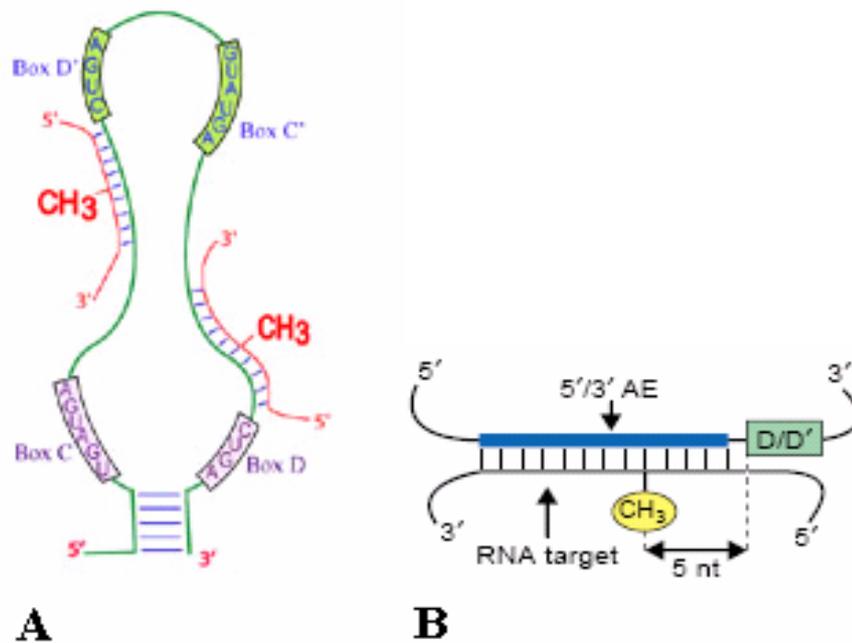


Fig.1. Struttura e funzione degli snoRNA di classe C/D.

A. Rappresentazione schematica di una canonica molecola di snoRNA di classe C/D. Il gruppo metilico rappresenta il nucleotide metilato sullo RNA target all'interno della doppia elica formata tra la sequenza guida dello snoRNA e lo RNA target. **B.** Ogni sequenza guida antisense (AE) media la 2'-O-metilazione al nucleotide posizionato 5 nt a monte delle box D/D'.

Immediatamente a monte della box D e/o D', si trovano elementi guida antisense, lunghi 10-21 nt, complementari in maniera perfetta alla molecola di RNA bersaglio. Tramite questo appaiamento si viene a formare una doppia elica nella quale il nucleotide da modificare è sempre posizionato esattamente 5 nt a monte delle box D o D' (Fig.1B; Kiss-Laszlo *et al.*, 1996). A questa classe,

infine, appartiene lo snoRNA U14, che presenta entrambe le funzioni di metilazione e di taglio dello rRNA (Liang *et al.*, 1997).

Gli snoRNA di classe H/ACA presentano una caratteristica struttura secondaria composta di due domini a “forcina” (hairpin), connessi da un motivo a singolo filamento (hinge) e seguiti da una corta coda terminale. Un motivo conservato chiamato box H (ANANNA, dove N sta per qualsiasi nucleotide), è localizzato nel filamento “hinge”, mentre un altro motivo conservato, la box ACA, si ritrova sempre a tre nt di distanza dall’estremità 3’ ed è quindi localizzato nella coda della molecola dello snoRNA (Fig.2A; Balakin *et al.*, 1997; Ganot *et al.*, 1997).

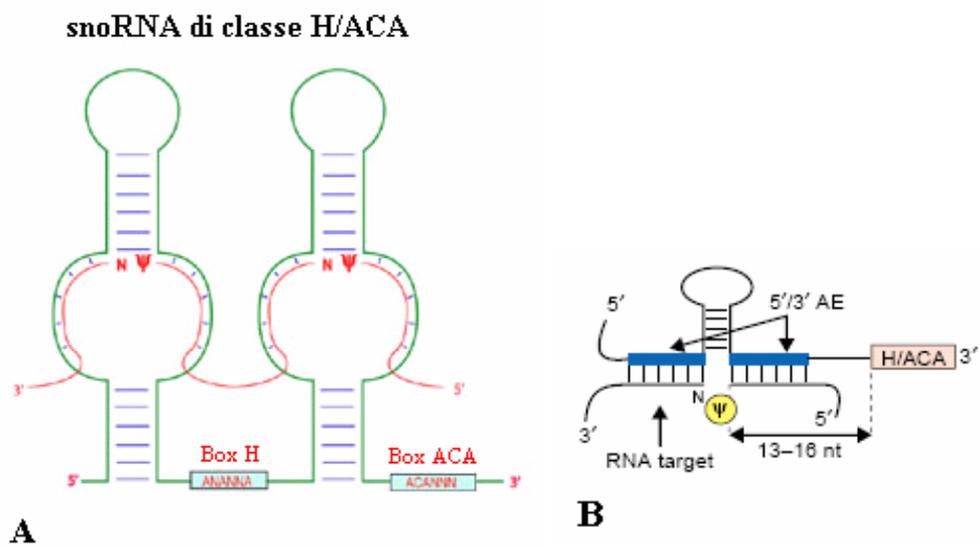


Fig.2. Struttura e funzione degli snoRNA di classe H/ACA.

A. Rappresentazione schematica di una canonica molecola di snoRNA di classe H/ACA. Nei loop interni delle strutture a forcina sono presenti le sequenze guida bipartite complementari all'RNA target che fiancheggiano i due nucleotidi non appaiati, di cui uno è l'uridina che verrà pseudouridilata. **B.** Il sito di pseudouridilazione è posizionato a 13-16 nt a monte delle box H/ACA.

Ciascuno snoRNA di classe H/ACA contiene una sequenza guida bipartita all'interno di uno o entrambi i loop dei due grandi domini a forcina. Questa sequenza si appaia perfettamente alla molecola di RNA bersaglio, formando con esso due corti duplex di 3-10 bp e lasciando al centro due soli nucleotidi non appaiati, tra cui uno corrisponde all'uridina che verrà pseudouridilata (Fig.2B). La distanza tra l'uridina bersaglio e le box H/ACA è sempre di 14-16 nt, così che la distanza da queste box determina quale sito subirà la pseudouridilazione (Ganot *et al.*, 1997). In lievito e nei mammiferi, inoltre, i due domini a forcina sono essenziali per la corretta attività di pseudouridilazione, anche quando lo snoRNA di tipo H/ACA contiene una sola sequenza guida (Bortolin *et al.*, 1999). Alcuni snoRNA di questa classe, gli snoRNA E2 e E3, anziché svolgere la funzione di pseudouridilazione hanno un ruolo importante nella maturazione dell'RNA ribosomale mediando il taglio nelle sequenze 5'ETS (Enright *et al.*, 1996). Sorprendentemente, in *T.brucei* e negli archebatteri, gli RNA che guidano la pseudouridilazione possiedono un singolo dominio a forcina seguito dalla box ACA e sembrano mancare completamente del motivo H (Liang *et al.*, 2001; Rozhdestvensky *et al.*, 2003). Ricordiamo poi che nel complesso ribonucleoproteico della telomerasi dei vertebrati, il componente ad RNA presenta i motivi tipici H/ACA, insieme agli elementi di sequenza essenziali per l'attività stessa della telomerasi (Mitchell *et al.*, 1999). I target preferenziali delle modificazioni operate dagli snoRNA sono gli RNA ribosomali, ma è stato dimostrato che essi possono andare a modificare anche molecole di RNA differenti, come gli snRNA (Tycowsky *et al.*, 1998; Kiss, 2001) o molecole di RNA messaggero (Vitali *et al.* 2005).

Recentemente sono state identificate anche molecole di snoRNA che si accumulano nei Cajal Bodies (organelli nucleari) e che per questo sono state chiamate “scaRNA” (small Cajal-body-specific RNA) (Darzacq *et al.*, 2002). Il Cajal Body (CB), descritto per la prima volta nel 1903, è un organello nucleare di struttura sferica presente in numero variabile nei nuclei (da 1 a 5). Il marcatore più usato per l’individuazione dei CB è la Coilina, una proteina che è stata indicata come il punto di legame tra questi organelli ed il nucleolo. Gli scaRNA possono essere strutturalmente e funzionalmente identici agli snoRNA che dirigono le 2’-O-metilazioni e le pseudouridilazioni. Gran parte delle molecole di scaRNA finora identificate presentano i motivi tipici delle molecole di snoRNA di classe H/ACA e solo due scaRNA di tipo C/D, U90 e U91, sono stati finora identificati (Darzacq *et al.*, 2002). E’ stato dimostrato che uno specifico elemento segnale assicura la ritenzione degli scaRNA di tipo H/ACA nei CB. Tale elemento di sequenza è chiamato CAB box ed è presente nei loop terminali di entrambe le forcine dei domini H/ACA (Richard *et al.*, 2003). La sequenza consenso per la CAB è 5’-UGAG-3’, dove le ultime due basi sono altamente conservate, mentre le prime due lo sono meno (Fig.3). Alterazioni della A o G nella CAB box aboliscono del tutto la sua funzione, dimostrando il suo ruolo critico nella maturazione e/o nell’accumulo degli scaRNA, anche se non è ancora noto in dettaglio il meccanismo d’azione. Tuttavia, alcuni scaRNA hanno strutture più complesse. Ad esempio alcuni contengono sia i motivi H/ACA che C/D e la loro lunghezza può essere compresa fra i 300-500 nt. E’ stato inoltre descritto uno scaRNA conservato nei mammiferi, chiamato U93, composto da due motivi H/ACA ripetuti (Kiss *et*

al., 2002). Gli scaRNA possono quindi dirigere sia le reazioni di 2'-*O*-metilazione che di pseudouridilazione sugli snRNA trascritti dalla RNA pol II, U1, U2, U4 e U5, che si localizzano transientemente nei CB prima di migrare nel nucleo. Mentre lo snRNA U6, trascritto dalla RNA pol III, è localizzato transientemente nel nucleolo per subire ulteriori modificazioni e non lascia il nucleo durante la sua maturazione (Lange and Gerbi, 2000), gli snRNA trascritti dalla RNA pol II (U1,U2,U4 e U5) sono invece esportati nel citoplasma, dove si associano con le proteine Sm, vengono processati alle loro estremità 3' e subiscono l'ipermetilazione del CAP. In seguito le snRNP assemblate ritornano nel nucleo, dove si associano ai CB per ulteriori modificazioni da parte dei suddetti scaRNA.

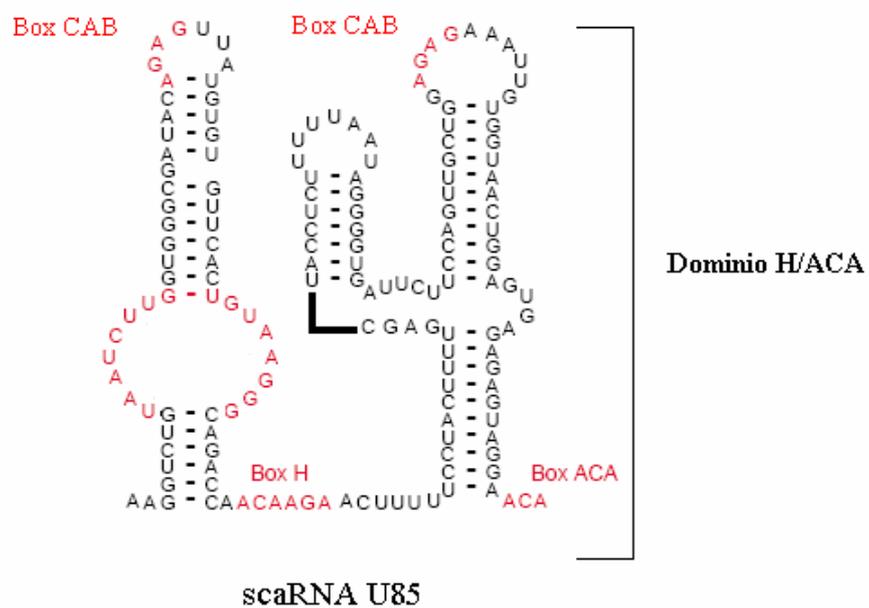


Fig.3. Posizione della box CAB che determina la localizzazione nei Cajal Bodies degli scaRNA con i motivi H/ACA.

La box CAB è localizzata nei loop terminali di una o di entrambe le strutture a forcina dei domini H/ACA. Nel loop interno della prima struttura a forcina è evidenziata anche la sequenza guida bipartita complementare ad una molecola di RNA target.

Recentemente è stato dimostrato che lo scaRNA umano U100, che è evolutivamente conservato, interagisce con lo snRNA U6 nel CB senza però modificarlo, ma agendo probabilmente da “RNA chaperone” e agevolando la formazione della particella snRNP U6 e la sua mobilità nel nucleoplasma (Vitali *et al.*, 2003).

1.3 Organizzazione genomica e biogenesi degli snoRNA

Entrambe le classi di snoRNA C/D ed H/ACA presentano caratteristiche simili per quanto riguarda la loro particolare organizzazione genomica ed i meccanismi di biosintesi. Nel lievito fino ad oggi sono stati caratterizzati solo otto snoRNA intronici (Lee *et al.*, 2003). La maggioranza degli snoRNA di lievito è infatti trascritta in maniera indipendente come unità mono- o policistroniche (Brown *et al.*, 2003). In più, sempre nel lievito, le sequenze dei trascritti indipendenti sono caratterizzate dalla presenza a monte di promotori TATA box-like e di siti di riconoscimento per il fattore di trascrizione “Rap1p” (Qu *et al.*, 1999) (Fig.4A). Nelle piante, sia l’organizzazione genica che la modalità di maturazione degli snoRNA appare un po’ differente da quella descritta in lievito e successivamente nei vertebrati. Infatti, oltre alla presenza di unità in prevalenza policistroniche di snoRNA trascritte indipendentemente, nelle piante sono stati individuati cluster di snoRNA policistronici all’interno di un singolo introne (Liang *et al.*, 2002; Fig.4B). Questa scoperta è in netto contrasto con la regola “uno snoRNA per introne” osservata nei vertebrati e nel lievito. Per quanto riguarda la maturazione, i cluster di snoRNA policistronici intronici richiedono due attività endonucleolitiche: la prima sul pre-mRNA e/o

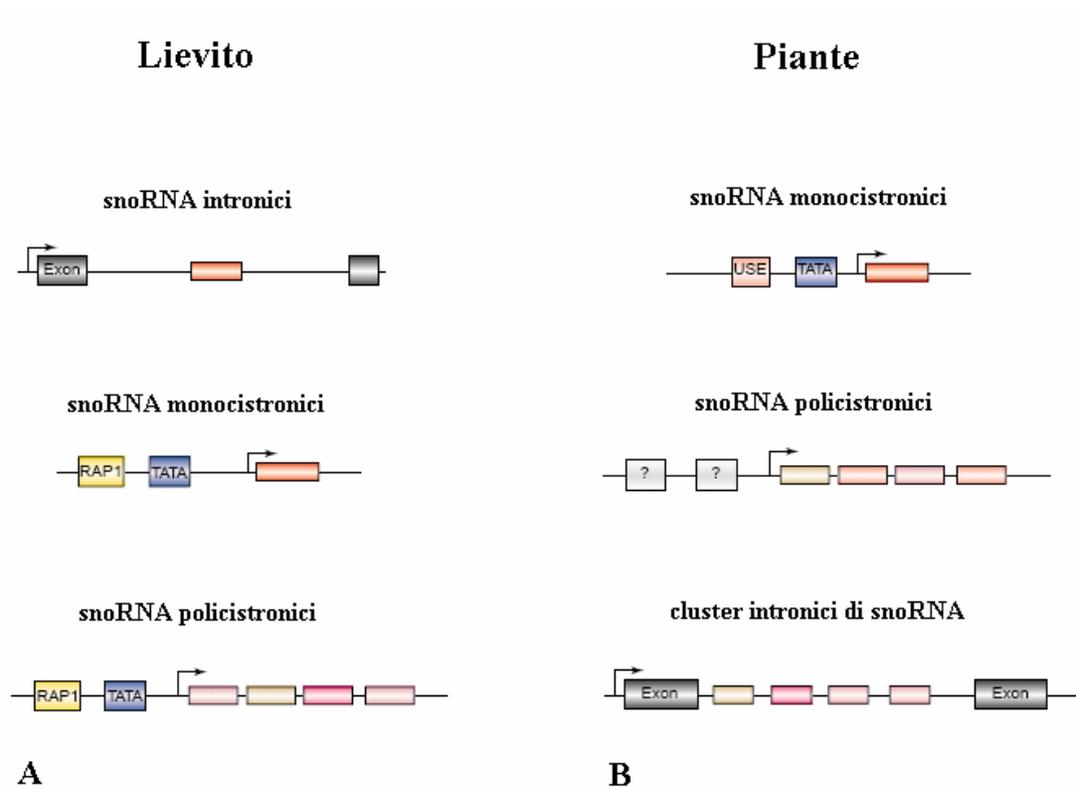


Fig.4 Organizzazione genomica degli snoRNA in lievito ed in piante.

A. In lievito gli snoRNA sono per lo più trascritti come unità mono- o policistroniche, ma possono essere presenti anche in introni di geni ospiti. **B.** Nelle piante, molti snoRNA sono trascritti come unità policistroniche, oppure organizzati in cluster intronici. Nel caso delle unità policistroniche le sequenze promotrici a monte non sono bene caratterizzate. Abbreviazione: USE, Upstream Sequence Element.

sull'introne eliminato, la seconda alle estremità delle singole unità di snoRNA, le quali alla fine sono maturate da digestioni esonucleolitiche (Fig.5). E' da ricordare che nelle piante i trascritti indipendenti non possiedono promotori TATA-like né siti di riconoscimento per fattori di trascrizione come nel lievito, e che le sequenze necessarie per l'espressione degli snoRNA non sono state ancora identificate.

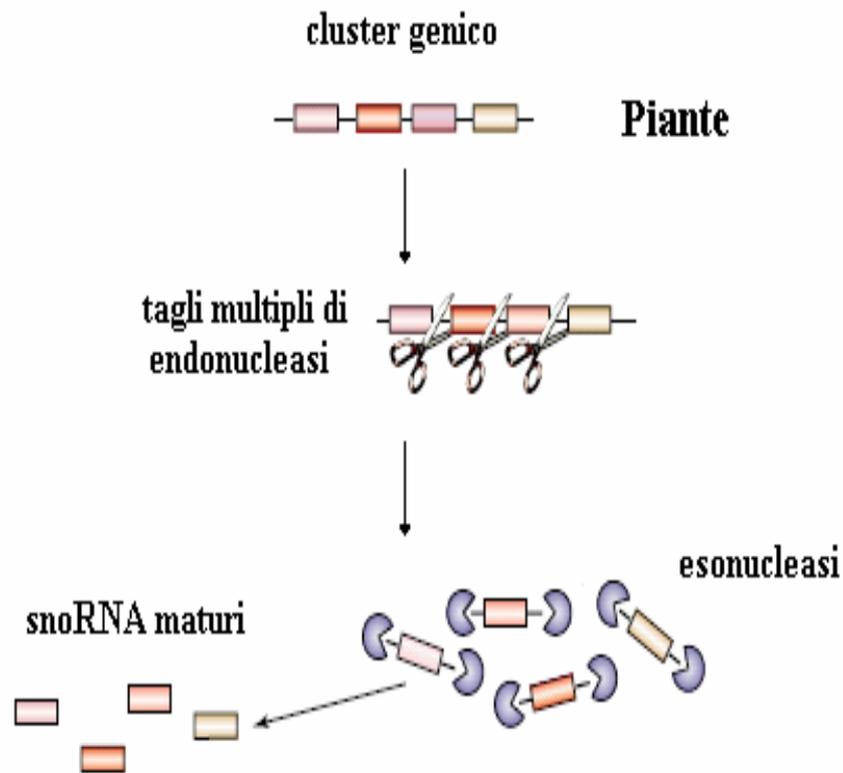


Fig.5 Maturazione dei cluster intronici nelle piante.

Nelle piante i cluster intronici di snoRNA richiedono due attività endonucleolitiche: la prima sul pre-mRNA e/o sull'introne eliminato, la seconda alle estremità delle singole unità di snoRNA, le quali alla fine sono maturate da digestioni esonucleolitiche.

Nei vertebrati, l'organizzazione genomica degli snoRNA è piuttosto differente rispetto a quanto descritto nel lievito e nelle piante. La maggioranza degli snoRNA dei vertebrati infatti è localizzata in regioni introniche di geni ospiti, spesso codificanti proteine coinvolte nella sintesi e nella funzione dell'apparato traduzionale (Maxwell and Fournier, 1995; Weinstein and Steitz, 1999; Kiss *et al.*, 2001). Altri snoRNA sono invece presenti in introni di geni i cui trascritti, anche se subiscono lo splicing e la poliadenilazione, sembrano non codificare alcuna proteina funzionale (Tycowski *et al.*, 1996; Bortolin *et al.*, 1998; Pelczar and Filipowicz, 1998; Smith and Steitz, 1998; Fig.6A). Questi geni sono quindi

ospiti in “ncRNA genes” e sembrano funzionare solamente come veicoli per la produzione degli snoRNA. I primi membri identificati di questa classe sono i geni di mammifero “*UHG*” e “*gas5*” che contengono, nei loro introni, differenti snoRNA di tipo C/D (8 snoRNA il primo gene: U22, U35-U41 e 10 snoRNA il secondo: U44, U47, U74-U76, U78-U81) ed i geni “U17HG” e “U19HG” che ospitano entrambi, nei loro introni, un solo tipo di snoRNA di classe H/ACA (due copie dello snoRNA U17 nel gene U17HG ed una singola molecola dello snoRNA U19 nel gene U19HG). Due caratteristiche accomunano i geni dei vertebrati che ospitano snoRNA. Come già detto, ospitano un singolo snoRNA per introne (regola del “uno-snoRNA-per-introne) e spesso presentano una sequenza al 5’terminale ricca in oligopirimidine, chiamata sequenza 5’-TOP. I promotori dei geni TOP sono peculiari, in quanto mancano della TATA box. E’ stato dimostrato che la sequenza TOP all’estremità 5’ di questi geni è implicata nel loro controllo traduzionale (Levy *et al.*, 1991) e probabilmente anche trascrizionale (de Turrís *et al.*, 2004). Gran parte degli snoRNA a localizzazione intronica presente nei vertebrati è prodotta mediante un meccanismo che coinvolge la linearizzazione del cappio intronico tramite un taglio endonucleolitico, seguito da digestioni ad entrambe le estremità da parte di esonucleasi specifiche che rilasciano le molecole di snoRNA mature (Tycowski *et al.*, 1993; Kiss and Filipowicz, 1993 e 1995; Fig.6B). In questo caso, quindi, la sintesi dello snoRNA dipende dall’avvenuta maturazione del trascritto ospite (pre-mRNA). La partecipazione attiva del complesso di splicing nel rilascio degli snoRNA maturi è confermata dall’identificazione di molti snoRNA umani di tipo C/D localizzati in una

posizione fissa all'interno degli introni. La distanza di 70-80 nt rispetto al sito di splicing è stata infatti dimostrata essere critica per un'efficiente espressione degli snoRNA di tipo C/D *in vivo* (Hirose *et al.*, 2001). Un meccanismo alternativo, meno comune, è costituito da tagli endonucleolitici del pre-mRNA nelle regioni introniche fiancheggiando la sequenza dello snoRNA, seguiti da digestioni esonucleolitiche maturative (Caffarelli *et al.*, 1996). Questi dati sono stati ottenuti in esperimenti di microiniezione in oociti di *X.laevis*, in cui l'attività di linearizzazione del coppia intronico è assente. Da questo meccanismo deriva che il rilascio dello snoRNA può anche essere indipendente dallo splicing del trascritto primario del gene ospite, il quale, una volta tagliato dalla endonucleasi, non produce alcun mRNA ed è destinato ad essere degradato. Nelle cellule somatiche, dove è presente l'attività di linearizzazione, entrambi i meccanismi di maturazione potrebbero essere attivi. Recentemente, infatti, è stato dimostrato che le sequenze 5'-TOP, tipiche dei geni che ospitano snoRNA nei loro introni, determinano uno specifico rapporto tra la produzione dello snoRNA maturo e le molecole di mRNA "spliced". Se il trascritto del gene ospite inizia con questi elementi di sequenza, lo snoRNA può infatti originarsi sia tramite meccanismo splicing-dipendente sia tramite taglio endonucleolitico del pre-mRNA. Questo indica che le sequenze 5' TOP sono capaci di reclutare fattori specifici che controllano il "pathway" di maturazione dei trascritti che le contengono (de Turris *et al.*, 2004). Queste unità di trascrizione indipendenti presentano tutte le caratteristiche di un classico gene: un promotore, una sequenza codificante lo snoRNA, e le sequenze di terminazione. Nei geni degli snoRNA U3, U8 e U13 nei vertebrati

sono trascritti dalla RNA polimerasi II e possiedono a monte entrambi gli elementi prossimali e distali a monte del sito d'inizio della trascrizione

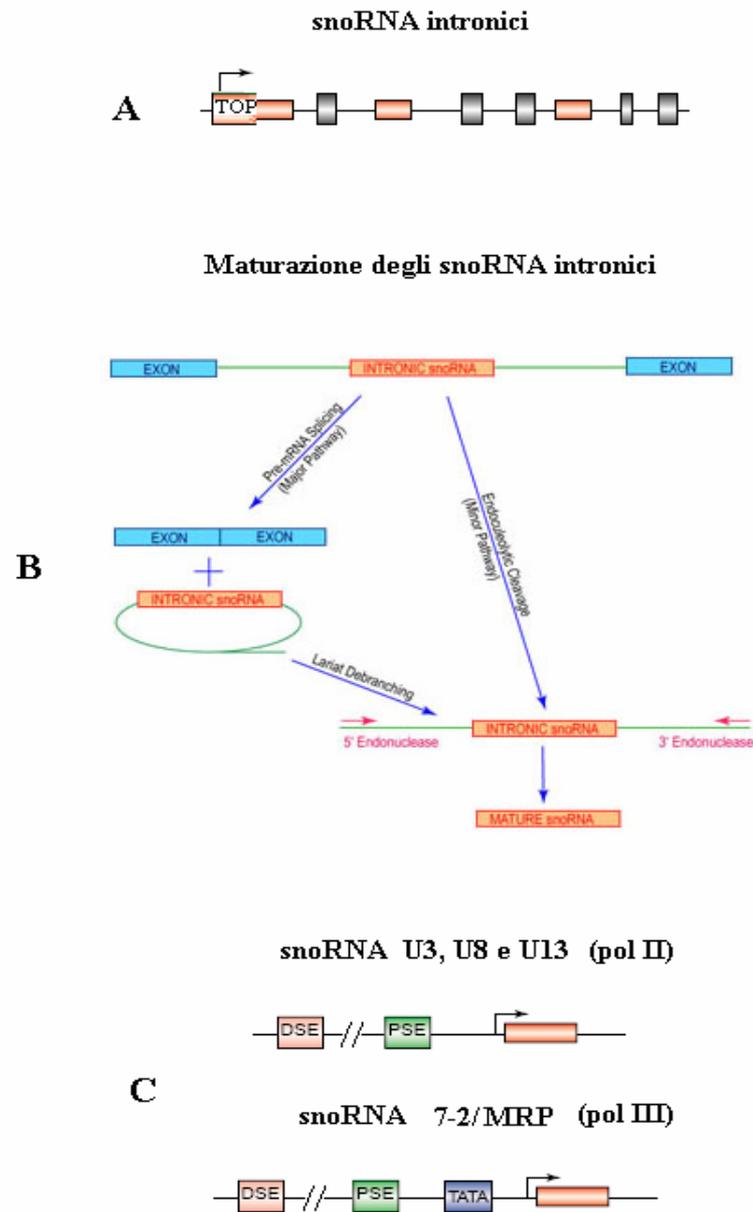


Fig.6 Organizzazione genica e maturazione degli snoRNA nei vertebrati.
A. SnoRNA intronici in geni ospiti in cui sono presenti le sequenze TOP al 5'. **B.** Maturazione degli snoRNA intronici tramite i due meccanismi splicing dipendente e indipendente. **C.** SnoRNA trascritti indipendentemente con a monte sequenze promotrici tipiche della RNA pol II e III.

e simili in sequenza a quelli per i geni degli snRNA (Savino *et al.*, 1992; Fig.6C). Il gene che produce lo snoRNA della RNasi MRP invece è trascritto dalla RNA polimerasi III ed ha a monte elementi promotori misti per le RNA polimerasi II e III: PSE e DSE accoppiati ad una sequenza TATA box (Kiss *et al.*, 1992). Recentemente sono stati identificati e caratterizzati due snoRNA che presentano, alle loro estremità 5', il cap di metilguanosa, caratteristica che conferma che i rispettivi geni sono trascritti indipendentemente. Ciascuno di questi snoRNA è formato da due domini C/D, uno dei quali si compone di box D e C la cui sequenza è difforme da quella consenso, ed entrambi sono privi di una sequenza antisense complementare allo rRNA o ad alcun snRNA; in più, i geni presentano, a monte, elementi promotori TATA l'uno, e TATA-less l'altro (Tycowski *et al.*, 2004). Questi dati suggeriscono che i vertebrati ancestrali possedessero snoRNA trascritti indipendentemente, e solo più tardi, nell'evoluzione, la maggior parte di questi geni si sia spostata negli introni di geni ospiti. Il lievito rappresenta il miglior organismo modello per la descrizione delle componenti proteiche che intervengono nei processi di maturazione degli snoRNA. Le unità monocistroniche e policistroniche di snoRNA di lievito sono inizialmente riconosciute dall'endonucleasi "Rnt1p", che taglia ai lati delle unità contenenti gli snoRNA, liberando questi ultimi dal trascritto primario (Chanfreau *et al.*, 1998). La proteina Rnt1p è omologa alla RNasi III batterica e riconosce il tetraloop con sequenza AGNN. Nel lievito, ma si pensa anche in altri organismi, il complesso dell' "esosoma" gioca un ruolo chiave nel generare le estremità 3' degli RNA e quindi non solo degli snoRNA (Allmang *et al.*, 1999). L'esosoma è composto di 11 subunità

(subcomplesso nucleare), che presentano omologia di sequenza con note ribonucleasi e per alcune delle quali è stata dimostrata *in vitro* l'attività 3'-5' esonucleasica. L'attività 5'-3' esonucleasica è svolta, invece, dalle RNasi nucleari "Rat1p" e "Xrn1p", conservate largamente negli eucarioti, che operano così il processo di maturazione delle estremità 5' degli snoRNA (Petfalski *et al.*, 1998). Nel caso delle unità monocistroniche, i motivi di sequenza C e D risultano essere importanti per almeno due meccanismi. Il primo riguarda la corretta formazione dell'estremità 3', in quanto le due box C/D sono riconosciute da una proteina che lega gli snoRNA di tipo C/D (la fibrillarina, vedi paragrafi successivi), che a sua volta interagisce con alcune componenti proteiche del complesso APT e precisamente con "Ref2p" e "Ssu72p" (Nedea *et al.*, 2003; Morlando *et al.*, 2004). Il complesso APT è coinvolto nella formazione delle estremità 3' degli snRNA e snoRNA nel lievito, tramite un meccanismo poliA-indipendente. Il secondo meccanismo si basa sulla necessità che le box C/D siano appaiate per la formazione del 5'-CAP ipermetilato (2,2,7-trimetilguanosina CAP) presente al 5' degli snoRNA trascritti in maniera indipendente dalla RNA pol II. E' stato dimostrato infatti, tramite esperimenti di microiniezione in oociti di *X. laevis*, che i motivi C/D sono necessari e sufficienti per la formazione del m₃G-CAP. In particolare, mutazioni nelle sequenze C o D portano all'assenza del CAP ipermetilato ed in più, sequenze minime di RNA composte essenzialmente dai motivi C/D, risultano ancora essere efficienti substrati per l'ipermetilazione del CAP, suggerendo l'esistenza di un comune meccanismo per questo tipo di modificazione, conservato in diversi organismi (Speckmann *et al.*, 2000; Jacobson and Pederson, 1998).

Infine, poco si sa in merito alle proteine coinvolte nel processo di rilascio degli snoRNA nei Vertebrati, fatta eccezione per l'endonucleasi "XendoU", isolata in *X. laevis*, che riconosce e taglia una specifica sequenza ricca di U nella struttura secondaria dello snoRNA U16, operando, quindi, attraverso il meccanismo splicing-indipendente (Laneve *et al.*, 2003; Gioia *et al.*, 2005).

1.4 Localizzazione cellulare degli snoRNA

Vari studi hanno indicato che il nucleolo è la sede di molte funzioni che finora non erano state associate a questa struttura subnucleare. E' stata infatti osservata la localizzazione nucleolare per molecole coinvolte in diversi processi, come la molecola di RNA della RNasi P (Jacobson *et al.*, 1997) e della telomerasi (Mitchell *et al.*, 1999), nonché il componente a RNA della particella che riconosce il segnale (Jacobson e Pederson, 1998). Inoltre, è stato osservato che nelle cellule di lievito la maturazione endonucleolitica dei pre-tRNA, catalizzata dalla RNasi P, ha luogo nel nucleolo (Bertrand *et al.*, 1998). Da ultimo, anche per gli mRNA di *myc* e *myoD* è stata osservata una localizzazione transiente nel nucleolo di linee cellulari tumorali umane (Bond and Wold, 1993), così come per molti dei trascritti cellulari e dell'HIV (Michienzi *et al.*, 2000). Per quanto riguarda gli snoRNA di classe C/D, è stato dimostrato che le box C e D sono entrambe essenziali per la localizzazione nucleolare degli snoRNA U3, U8 e U14 (Lange *et al.*, 1998). Le due box infatti, insieme alle proteine ad esse legate, oltre a dirigere la corretta maturazione di entrambe le estremità 5' e 3' ed a fornire la stabilità metabolica alla molecola, sono coinvolte nel trasporto degli snoRNA dal nucleoplasma, dove vengono rilasciati i pre-snoRNA, al nucleolo, dove gli snoRNA vengono

maturati e rilasciati per le loro funzioni. E' stato ipotizzato che gli snoRNA sintetizzati transitino dapprima nei CB, dove, con molta probabilità, subiscono ulteriori modificazioni prima di migrare nel nucleolo (Narayanan *et al.*, 1999). Come già accennato, alcuni snoRNA, però, anziché spostarsi nel nucleolo, rimangono nei CB, grazie al particolare motivo CAB box presente nella loro sequenza. Questi snoRNA (scaRNA) rappresentano un nuovo sottoinsieme degli snoRNA e la loro scoperta rende necessaria una revisione dell'iniziale nomenclatura, basata sulla localizzazione esclusiva di queste molecole nel nucleolo. E' poi interessante notare che studi in lievito hanno portato a descrivere particolari strutture nucleari chiamate *nucleolar bodies* (NB), che presentano funzioni simili a quelle dei CB (Verheggen *et al.*, 2001). Forti evidenze che gli snoRNA subiscano eventi di maturazione nei NB derivano dalla scoperta che l'enzima Tgs1p, implicato nella ipermetilazione di alcuni snoRNA (vedi sopra), è localizzato in queste strutture nucleari (Mouaikel *et al.*, 2002). Parallelamente, una proteina omologa a Tgs1p è risultata localizzata nei CB dei vertebrati (Verheggen *et al.*, 2002).

1.5 Composizione e formazione delle particelle snoRNP

Entrambe le classi di snoRNA di classe C/D e H/ACA svolgono le loro funzioni biologiche come componenti di complessi ribonucleoproteici a localizzazione nucleolare (snoRNP). Ciascun complesso è composto da uno specifico snoRNA associato ad un ristretto numero di proteine che sono caratteristiche di ciascuna delle due classi.

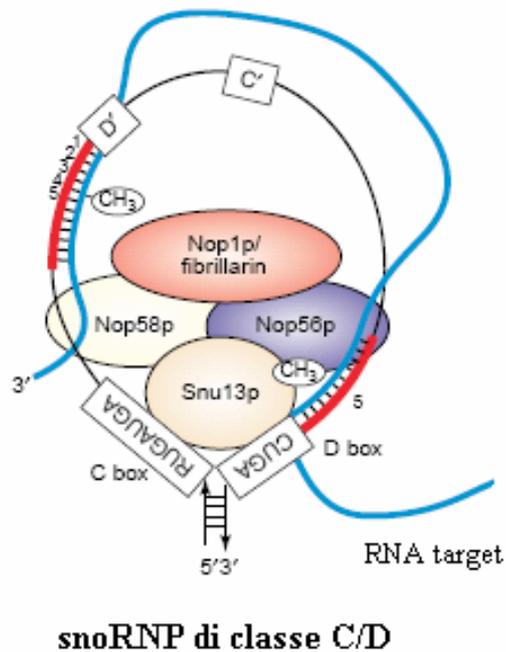


Fig.7 SnoRNP di classe C/D.

La proteina Snu13p interagisce con entrambe le box C e D degli snoRNA e riconosce il motivo di struttura K-turn. Nop1p (la fibrillarina) interagisce con entrambe le box D e D' (non mostrato) mentre le proteine Nop58p e Nop56p riconoscono la box D e la box D', rispettivamente. A monte delle box D e D' sono mostrate le sequenze guida antisenso e complementari alla molecola di RNA target.

In aggiunta a queste proteine, che costituiscono il “core” proteico, altre proteine sembrano associarsi transientemente alle snoRNP di ciascuna classe per assicurare il loro corretto assemblaggio e trasporto nel nucleolo. Le snoRNP di tipo C/D sono formate da un complesso minimo (“core”) composto di quattro proteine conservate ed essenziali (Fig.7), che nel lievito sono Snu13p (15.5 kDa nell’uomo), Nop1p (fibrillarina nell’uomo), Nop56p e Nop58p (hNop56p e hNop58p nell’uomo). Snu13p si lega in maniera specifica al motivo di struttura secondaria formato dall’appaiamento delle box C e D, chiamato K-turn, ritrovato anche nelle strutture secondarie dello snRNA U4 e negli rRNA ribosomali. Un canonico motivo K-turn contiene due corti stem: lo stem I contiene due appaiamenti G-C , mentre lo stem II contiene due appaiamenti G-A non canonici, con in mezzo un “bulge” di tre nucleotidi, formando un motivo asimmetrico decisivo per il riconoscimento da parte di Snu13p (15.5 kDa) (Fig.8A). L’interazione tra il K-turn e Snu13p , comunque, non riguarda le box interne C’ e D’, caratteristica questa osservata in tutti gli eucarioti (Szewczak *et al.*, 2002). Nop1p (fibrillarina) è riconosciuta essere l’enzima con attività di 2’-O-metiltransferasi. Questa proteina possiede, infatti, una sequenza aminoacidica tipica delle metiltransferasi SAM- dipendenti (Niewmierzycka e Clarke, 1999). Inoltre, mutazioni puntiformi nel gene *Nop1p* in posizioni interne al dominio proteico metiltransferasi-like bloccano la metilazione di tutti i ribosi nelle cellule di lievito (Tollervey *et al.*, 1993). Le altre proteine del “core” delle C/D snoRNP, Nop56p e Nop58p, presentano sequenze aminoacidiche molto simili tra di loro (Newman *et al.*, 2000). p50 e p55, due proteine nucleari correlate strutturalmente, si associano poi alle snoRNP di tipo

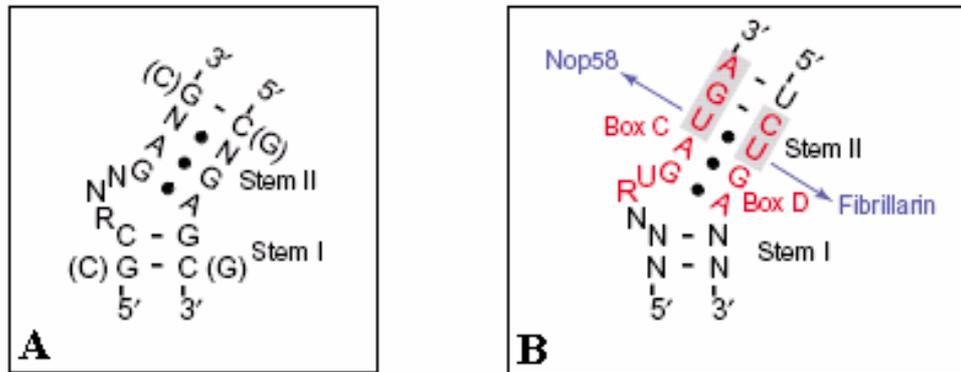


Fig.8 Confronto tra il motivo K-turn “consenso” e quello presente negli snoRNA di classe C/D.

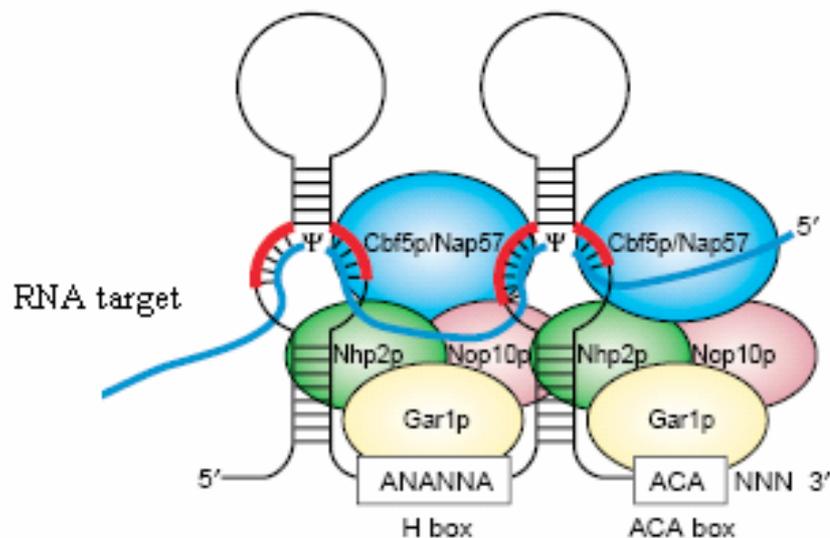
A. Motivo K-turn canonico, comprendente due corti stem: la stem I con due appaiamenti G-C e la stem II con due appaiamenti G•A non canonici, separate da un “bulge” di tre nucleotidi. **B.** Motivo K-turn presente negli snoRNA di classe C/D. Nop58 e la fibrillarina interagiscono con le uridine nella stem II.

C/D suggerendo una loro partecipazione alla biogenesi di questi complessi. Le proteine p50 e p55 presentano un’attività DNA elicastica che in precedenti studi è stata collegata ai meccanismi di rimodellamento della cromatina ed alla trascrizione (Newmann *et al.*, 2000). Un’altra proteina transitoriamente associata alle snoRNP di tipo C/D è la fosfoproteina Nop140, presente nel nucleolo e nei CB, che potrebbe facilitare la formazione e il trasporto delle snoRNP di questa classe (Yang *et al.*, 2000). Anche la proteina SMN (*Survival of Motor Neurons*) sembra essere implicata nel metabolismo delle snoRNP di tipo C/D tramite la sua interazione con il dominio proteico di Nop1p, ricco in glicine ed arginine (dominio GAR). Questa interazione è alterata nel caso in cui la proteina SMN è mutata, condizione patologica che si riscontra nel caso dei pazienti affetti da atrofia spino-muscolare (Jones *et al.*, 2001; Pellizzoni *et al.*, 2001).

Nei vertebrati è stato osservato che la proteina di 15.5 kDa, omologa alla proteina Snu13p del lievito, è la prima proteina che interagisce con la molecola dello snoRNA, e il suo legame al motivo K-turn è essenziale per il corretto assemblaggio delle restanti proteine che compongono la snoRNP di tipo C/D (Watkins *et al.*, 2002). Dopo l'avvenuta interazione tra il motivo K-turn e la proteina di 15.5 kDa, entrambe le proteine Nop58p e Nop1p (fibrillarina) si legano alla molecola dello snoRNA. Precisamente, le due proteine riconoscono il residuo di uridina dello stem II della box C e della box D, rispettivamente (Cahill *et al.*, 2002) (Fig.8B). Inoltre, Nop1p (fibrillarina) riconosce, insieme a Nop56b, anche la box interna D'. Sembra quindi che una comune caratteristica dei motivi C/D e C'/D' sia responsabile dell'interazione con la fibrillarina, la probabile metiltransferasi, e che alcuni tra i nucleotidi che mediano l'interazione siano proprio all'interno delle Box D e D' (vedi Henras *et al.*, 2004).

Ritornando al meccanismo di maturazione degli snoRNA nei vertebrati, e particolarmente nei mammiferi, aggiungiamo alcune particolari osservazioni riguardo gli snoRNA di tipo C/D, integrando le componenti proteiche descritte. Nel rilascio splicing-dipendente, operato dalle esonucleasi, la distanza di 70-80 nt tra lo snoRNA ed il sito 3' accettore, come già accennato, si è dimostrata essere una proprietà importante per la corretta maturazione della molecola di snoRNA (Hirose *et al.*, 2003). Nel caso degli snoRNA localizzati in una tale posizione all'interno dell'introne del gene ospite, la proteina di 15.5 kDa si lega per prima, durante lo splicing, sulla molecola di snoRNA ancora legata al trascritto primario, seguita poi dalla fibrillarina e quindi da Nop56p e

58p. Invece, per gli snoRNA localizzati a maggiore distanza dal sito 3' di splicing, è una lunga struttura a forcina stabile, esterna allo snoRNA, a determinare il rilascio della molecola. Inoltre, in questo caso la proteina 15.5 kDa e la fibrillarina si legano indipendentemente del processo di splicing, che comunque facilita la formazione della lunga forcina esterna. Le proteine comuni alle snoRNP di tipo H/ACA includono le proteine evolutivamente conservate di lievito Cbf5p (discherina nell'uomo, MFL in *Drosophila*, Gar1p, Nhp2p e Nop10p, tutte essenziali per la reazione di pseudouridilazione (Fig.9).



snoRNP di classe H/ACA

Fig.9 snoRNP di classe H/ACA.

Cbf5p (la discherina nell'uomo) interagisce probabilmente con entrambi i motivi H e ACA (non mostrato). Sono mostrate le sequenze bipartite nei loop interni delle strutture a forcina dei domini H/ACA che interagiscono con la molecola di RNA target.

Cbf5p corrisponde alla componente catalitica di questa classe in quanto sono presenti, nella sua sequenza, domini proteici tipici delle pseudouridine sintasi batteriche. Inoltre, mutazioni puntiformi nel suo gene (*Cbf5*) aboliscono in vivo le pseudouridilazioni delle molecole di rRNA in cellule di lievito (Zebarjadian *et al.*, 1999). In più si pensa, in base ad osservazioni sulla caratteristica distanza tra ciascuna di queste due box ed il sito da modificare che, con molta probabilità, Cbf5p si leghi ad entrambe le box H e ACA (Ganot *et al.*, 1997). Cbf5p, in base a dati di omologia delle sequenze aminoacidiche, è correlata alla proteina TruB di *E. coli* che catalizza la formazione di pseudouridine nei T loops di molti tRNA. La struttura cristallina di TruB legata all' RNA suggerisce che sia Cbf5p che TruB riconoscano la molecola di RNA nella stessa maniera. Nhp2p, Nop10p e Cbf5p costituiscono le proteine essenziali delle snoRNP di tipo H/ACA. Anche per questa classe di snoRNP ricordiamo le interazioni con le proteine descritte in precedenza, p50, p55, SNM e Nopp140, che si legano transitoriamente al complesso delle snoRNP facilitandone la biosintesi ed il trasporto nel nucleolo.

Mutazioni che alterano la sequenza della discherina, la proteina umana omologa a Cbf5p, sono causa della discheratosi congenita X-linked, una malattia ereditaria umana dovuta ad alterazione della biogenesi dei ribosomi e ad accorciamento della lunghezza dei telomeri. Questo ultimo aspetto è probabilmente dovuto ad un'alterata interazione tra la proteina mutata e la molecola di RNA di tipo H/ACA del complesso della telomerasi (Vulliamy *et al.*, 2001).

1.6 Origine evolutiva degli snoRNA dagli archebatteri agli eucarioti

Anche gli archebatteri, come già accennato, posseggono RNA snoRNA-like, chiamati “sRNA”, che guidano le 2'-*O*-metilazioni e le pseudouridilazioni dei nucleotidi target presenti sulle molecole di rRNA e tRNA. La presenza di snoRNA di classe C/D e H/ACA in questi organismi suggerisce fortemente che l'impiego di RNA guida per le modifiche nucleotidiche post-trascrizionali sia un meccanismo evolutivamente antico (Tran *et al.*, 2004). Benché i nucleotidi modificati sulle molecole di rRNA siano pochi in numero e differenti nelle posizioni rispetto a quelli negli eucarioti, gli elementi strutturali tipici degli RNA guida (i motivi C/D e H/ACA, K-turn, C'/D' Box e il dominio a forcina) sono ben conservati tra gli archebatteri e gli eucarioti. L'analisi delle sRNP di classe C/D e H/ACA ha inoltre rivelato una composizione proteica simile a quella delle rispettive snoRNP eucariotiche. Nelle sRNP di tipo C/D degli archebatteri sono state caratterizzate tre proteine che formano il “core” e che presentano omologia con le proteine del “core” eucariotiche: la proteina ribosomale L7 (omologa alla proteina di 15.5kD eucariotica), una singola proteina Nop56/58p e la fibrillarina (Fig.10A). Le sRNP di tipo H/ACA sono composte invece da quattro proteine che compongono il “core”, di cui tre omologhe a quelle eucariotiche (Cbf5p, Gar1p e Nop10p). La quarta proteina eucariotica, Nhp2p, presenta un'alta similarità con la proteina ribosomale L7 degli Archebatteri, che, come abbiamo visto, è omologa anche della proteina di 15.5kD eucariotica (Fig.10B).

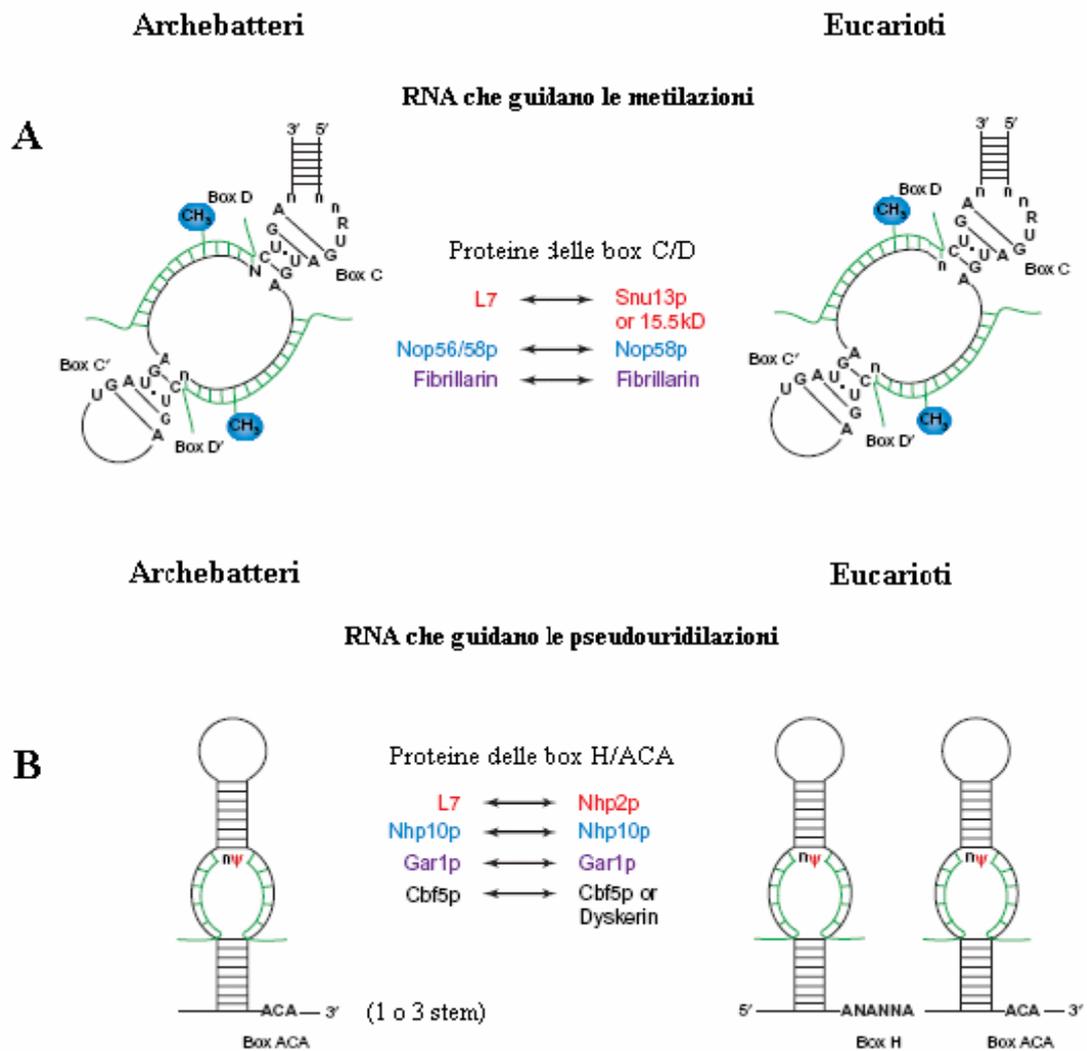


Fig.10 Composizione dei complessi sRNP di classe C/D e H/ACA degli archebatteri.

A. Composizione dei complessi sRNP di classe C/D e confronto con i complessi snoRNPs eucariotici. La proteina L7 (omologa alla proteina Snu13p) riconosce il motivo conservato K-turn. **B.** Proteine dei complessi sRNP di classe H/ACA. La proteina L7 è presente in entrambi i complessi sRNP. Negli archebatteri le strutture a forcina (stem e loop) dei domini H/ACA sono presenti in singola copia o in 3 copie per molecola di sRNA.

Questa interessante scoperta suggerisce che in origine le due classi di sRNA formassero un comune complesso RNP ancestrale (Ganot *et al.*, 1997). La proteina L7 riconosce il motivo conservato K-turn presente in entrambe le classi degli sRNA degli archebatteri (Rozhdestvensky *et al.*, 2003). Ciò potrebbe implicare che anche la proteina Nhp2p eucariotica possa riconoscere un ipotetico motivo K-turn presente negli snoRNA H/ACA eucariotici. Tuttavia, un motivo K-turn non è stato ancora identificato in questa classe di snoRNA, anche se è presto per escludere tale possibilità, vista la grande variabilità mostrata dal motivo K-turn nelle strutture ripiegate di RNA. L'apparente assenza nei batteri dei complessi di modificazione dei nucleotidi guidati da RNA indica che i primi complessi ad RNP di questo tipo derivino da un ramo evolutivo comune al regno degli archebatteri e degli eucarioti, posteriore alla divergenza dai batteri. L'uso di specifici enzimi per le modificazioni dello RNA nei batteri ha comportato nell'evoluzione un limitato numero di nucleotidi modificati; al contrario, gli rRNA degli eucarioti e degli archebatteri posseggono un numero largamente maggiore di modificazioni. È stato quindi suggerito che il grande numero di RNP di tipo C/D e H/ACA in questi due regni si sia sviluppato attraverso variazioni della sequenza guida (Lafontaine and Tollervey, 1998). Molecole di RNA di tipo C/D e di tipo H/ACA, che apportavano benefici alla sintesi o alla funzione dei ribosomi sarebbero stati selezionati positivamente e conservati. Questo modello spiega la mancanza di conservazione dei siti modificati tra i due regni e tra organismi dello stesso regno.

1.7 Funzioni degli snoRNA

La maggior parte degli snoRNA è implicata nella modificazione e maturazione del pre-rRNA nel nucleolo delle cellule eucariotiche. La biogenesi dello rRNA eucariotico, come è noto, richiede una numerosa serie di passaggi di maturazione che prevedono precisi tagli esoendonucleolitici del trascritto primario, con la conseguente rimozione di estese regioni “spacer” e la produzione delle molecole mature di rRNA presenti nelle subunità ribosomali (28S, 18S e 5.8S). Prima di questi tagli maturativi, il pre-rRNA appena trascritto dalla RNA polimerasi I subisce una complessa serie di modificazioni in posizioni specifiche e funzionalmente importanti per la corretta attività del ribosoma. Queste modificazioni sono essenzialmente di due tipi: la 2'-*O*-metilazione e la pseudouridilazione, entrambe dirette dagli snoRNA (Smith and Steitz, 1997). In aggiunta alle reazioni di modificazione, un piccolo numero di snoRNA media il processo di taglio e maturazione del pre-rRNA. Questi snoRNA sono U3, U8, U14 e U22 nei vertebrati appartenenti alla classe C/D, ed snR10 e snR30 in lievito appartenenti alla classe H/ACA (Tollervey, 1987; Tollervey, 1996; Sharma and Tollervey, 1999; Morrissey and Tollervey, 1993). Negli ultimi anni è stato dimostrato che anche le modificazioni dei piccoli RNA spliceosomali trascritti dalle RNA polimerasi II–III, gli snRNA U1, U2, U4, U5 (pol II) e U6 (pol III), sono dirette dagli snoRNA (Tycowski *et al.*, 1998; Kiss, 2001).

In tutti questi casi di modificazione, sia a carico dello rRNA che degli snRNA, un nucleotide presente nel trascritto precursore è riconosciuto da specifiche interazioni che si stabiliscono tra l'RNA bersaglio e lo snoRNA. Si viene così a

formare una corta elica che seleziona il sito bersaglio, mentre la funzione catalitica della specifica modificazione viene svolta da apposite proteine legate allo snoRNA (che formano i complessi snoRNP). Lo snoRNA agisce quindi nel posizionare correttamente queste componenti catalitiche (Filipowicz and Pogačic, 2002). Negli archebatteri è stato dimostrato che anche i tRNA possono essere bersagli delle modificazioni operate dagli snoRNA (Dennis *et al.*, 2001), mentre in *Trypanosoma brucei* si è osservato che l'RNA SL (spliced leader RNA) è pseudouridilato in una specifica posizione (a 12 basi dal sito 5' di splicing) e che questa modificazione è operata da uno snoRNA chiamato SLA1 (Liang *et al.*, 2001). Nei vertebrati un crescente numero di snoRNA, specialmente della classe C/D, è espresso esclusivamente, o quasi, nel sistema nervoso centrale. Almeno sette snoRNA di tipo C/D derivati dagli introni di un'unità trascrizionale complessa e sottoposta al fenomeno dell'imprinting nella linea germinale paterna, chiamata IC-SNURF-SNRPN hanno mostrato un'espressione specifica nel cervello (Runte *et al.*, 2001; Cavallè *et al.*, 2000 e 2001). Per due di essi, HBII-52 e HBII-85, è stata osservata l'assenza di espressione nei cervelli di pazienti affetti dalla sindrome di Prader-Willi (Cavallè *et al.*, 2000). Questa interessante scoperta ha suggerito che i due snoRNA potrebbero essere coinvolti nell'eziologia di questa malattia. La caratterizzazione, nell'uomo, di una nuova traslocazione t(4;15) (q27; q11.2) associata alla sindrome di Prader-Willi ha suggerito inoltre l'ipotesi che la comparsa della patologia sia legata, almeno in parte, all'assenza dello snoRNA HBII-85, anche se i putativi target di questo snoRNA restano per il momento sconosciuti (Schule *et al.*, 2005). Anche per gli altri snoRNA ad espressione

cervello-specifica non è stata trovata complementarietà di sequenza con rRNA, snRNA o altri RNA stabili. Tuttavia, lo snoRNA HBII-52 umano mostra una lunga sequenza di 18 nt complementare alla molecola di mRNA del recettore della serotonina 5-HT_{2C} umano. E' stato recentemente dimostrato che questo snoRNA di tipo C/D dirige la 2'-O-metilazione di un residuo adenosinico, nel trascritto del recettore della serotonina 5-HT_{2C}, noto per essere deamminato ad inosina mediante un processo di editing operato dall'enzima ADAR2 (Fig.11).

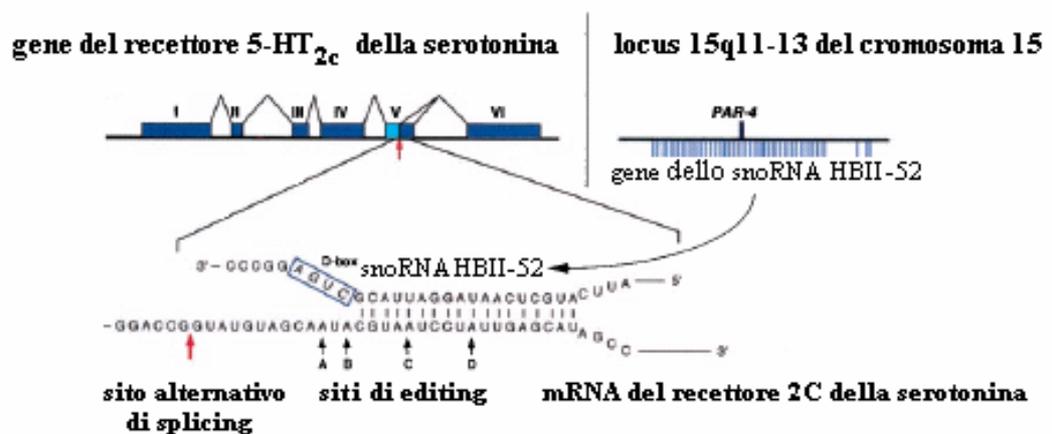


Fig.11 SnoRNA HBI-52 di classe C/D.

Tra lo mRNA del recettore 5-HT_{2C} della serotonina e lo snoRNA HBII-52 si instaura un'interazione che coinvolge una regione sottoposta ad editing. In particolare l'adenina metilata corrisponde ad uno dei quattro siti di conversione da A a I (il sito C) operata dall'enzima ADAR 2.

Poiché la 2'-O-metilazione inibisce la deamminazione dell'adenosina, lo snoRNA HBII-52 sembrerebbe svolgere un'importante funzione regolatrice, antagonizzando il meccanismo dell'editing operato dall'enzima ADAR2 (Vitali *et al.*, 2005). Questi ultimi dati dimostrano la grande varietà di possibili substrati modificati dagli snoRNA e le loro molteplici funzioni cellulari, che

vanno al di là della biogenesi dello rRNA. In generale, le possibili funzioni giustificano il gran numero finora identificato di snoRNA, per i quali non si è ancora trovato l'elemento antisense bersaglio (snoRNA "orfani"). In questi anni sono stati infatti identificati numerosi snoRNA cosiddetti "orfani", in quanto mancano di regioni di complementarietà con rRNA, snRNA, tRNA o altri noti RNA stabili (Hüttenhofer *et al.*, 2001; Kiss *et al.*, 2001). Quest'osservazione suggerisce che il repertorio di substrati della 2'-O-metilazione e della pseudouridilazione guidate dagli snoRNA sia significativamente maggiore di quello finora noto. Inoltre, è anche possibile che alcuni snoRNA orfani partecipino a funzioni cellulari diverse dalle modificazioni degli RNA, quali il taglio dei trascritti precursori dei rRNA, oppure operino come molecole "chaperone" per il corretto ripiegamento delle molecole di RNA bersaglio.

Poiché i dati recenti sullo snoRNA HBII-52 indicano che anche le molecole di mRNA possono essere bersaglio di modificazione, appare probabile che molti snoRNA orfani possano in effetti modificare le molecole di mRNA. Nonostante il loro coinvolgimento in processi cellulari fondamentali, l'identificazione delle molecole di snoRNA è risultata finora difficile a causa della mancanza, nella loro sequenza, di moduli di lettura aperti per la traduzione (ORF: Open Reading Frame) o di altre caratteristiche, come la coda di poli(A), che hanno semplificato, invece, l'individuazione dei trascritti dei geni codificanti sequenza, di moduli di lettura aperti per la traduzione (ORF: Open Reading Frame) o di altre caratteristiche, come la coda di poli(A), che hanno semplificato, invece, l'individuazione dei trascritti dei geni codificanti

proteine o delle lunghe molecole di ncRNA poliadenilate. In questi anni, l'identificazione di nuovi snoRNA e di molecole di snmRNA (small non-messenger RNA) è avvenuta principalmente mediante due tipi di analisi: un'analisi sperimentale, basata sulla costruzione di librerie di cDNA da frazioni di RNA di piccolo peso molecolare (circa 50-500 nt) (RNomica sperimentale) ed un'analisi computazionale dei genomi, basata sulla ricerca di sequenze e motivi conosciuti (inclusi elementi trascrizionali) e confronti filogenetici (RNomica computazionale).

2. RISULTATI

2.1 Introduzione al lavoro sperimentale

La genomica strutturale, ossia la determinazione sistematica di tutte le sequenze rappresentate in un genoma, è stata finora focalizzata quasi esclusivamente sulle sequenze che codificano per proteine. Sebbene sia un luogo comune parlare dei geni e dei loro prodotti proteici, migliaia di sequenze presenti nei genomi degli organismi eucariotici producono trascritti che esplicano la loro funzione senza produrre proteine (vedi per delle review Hüttenhofer *et al.*, 2005; Mattick, 2005).

Non è difficile quindi capire come la “RNomica”, cioè lo studio degli RNA funzionali e delle loro interazioni a livello genomico, rivesta un ruolo di estrema importanza, sia teorica che pratica, nella biologia moderna. Il primo passo verso il raggiungimento di una più ampia e soddisfacente comprensione di questo interessante campo di ricerca è rappresentato dallo sviluppo di metodologie versatili ed affidabili capaci di facilitare l’identificazione di quelle molecole di RNA che possono sfuggire alle analisi condotte con i metodi classici. Come già ampiamente discusso nel paragrafo introduttivo, la categoria degli RNA non-coding comprende sia RNA costitutivi (housekeeping) che RNA con funzioni regolative (regulatory; vedi Morey and Avner, 2004 per una review). In quest’ultima classe rientrano i piccoli RNA nucleolari (snoRNA) che intervengono nei processi di modificazione post-trascrizionale a carico di molecole di RNA target, fra cui soprattutto gli RNA ribosomali.

Gli approcci più utilizzati per l’identificazione di nuovi geni che producono snoRNA sono stati fondamentalmente due: uno sperimentale, l’altro

computazionale. L'approccio sperimentale prevede la costruzione di una libreria di cDNA a partire da frazioni di piccoli RNA (in genere delle dimensioni di circa 50-500 nt) i cui cloni vengono selezionati tramite ibridazione con sonde corrispondenti ad RNA già noti, in modo da identificare come possibili nuovi geni quelli che danno un segnale di ibridazione non significativo con le sonde utilizzate (vedi per esempio Hüttenofer *et al.*, 2001; Brown *et al.*, 2001; Yuan *et al.*, 2003). L'approccio computazionale più efficace per l'identificazione di molecole di snoRNA si basa sulla ricerca di elementi strutturali e motivi di sequenza che includono preferibilmente elementi di regolazione trascrizionale, in combinazione con analisi filogenetiche. La combinazione della "RNomica sperimentale" e di quella computazionale è la condizione ideale, ma tutti i risultati vanno definitivamente confermati da analisi del profilo di espressione e studi funzionali dei nuovi geni identificati. In questi ultimi anni numerose molecole di snoRNA sono state individuate e caratterizzate in diversi organismi. Uno degli organismi più utilizzati per l'analisi *in vivo* dei principali processi biologici è la *Drosophila melanogaster*, da tempo identificata come sistema modello grazie agli enormi vantaggi sperimentali legati alla facilità di allevamento, alla brevità del ciclo vitale ed alla facilità di manipolazione genetica. Infatti, all'enorme mole di dati genetici già accumulati, si aggiunge la possibilità di disporre di una serie di sofisticate tecniche di transgenesi e dei sistemi binari di manipolazione genetica GAL4-UAS e Flp-FRT, la cui efficacia è supportata dalla disponibilità di un genoma da tempo interamente sequenziato (Adams *et al.*; 2000).

In *D. melanogaster*, il primo snoRNA della classe C/D (U21), è stato identificato nel 1996 all'interno del quinto introne del gene ARF-1 ed ha rappresentato il primo esempio di snoRNA intronico negli invertebrati (Renalier *et al.*, 1996). Il primo snoRNA di tipo H/ACA identificato è stato invece lo snoRNA H1, presente nel quarto introne del gene *minifly* di *Drosophila* (Giordano *et al.*, 1999), un gene che codifica per la proteina MFL che è omologa alla proteina Cbf5p di lievito ed alla discherina umana. Le proteine della classe MFL/Cbf5p/discherina rappresentano una delle quattro componenti del “core” proteico delle snoRNP di classe H/ACA presenti negli eucarioti e svolgono l'attività catalitica di pseudouridilazione delle molecole di RNA target. Solo dopo alcuni anni (nel 2001), dalle analisi degli appaiamenti di sequenza di alcuni snoRNA umani con la sequenza genomica di *Drosophila* tramite l'algoritmo BLAST, sono stati individuati altri snoRNA di classe C/D che risultavano conservati tra Uomo e *Drosophila*. Questi snoRNA si presentavano in differenti copie e localizzati in introni di geni verosimilmente non-coding, chiamati “dUHG1” e “dUHG2”, per la loro correlazione evolutiva con i geni UHG (Unknown snoRNA Host Gene) dei mammiferi (Tycowski and Steitz, 2001). Più recentemente sono stati identificati 66 piccoli RNA stabili di *Drosophila*, tra cui erano rappresentate molecole di snoRNA di entrambe le classi (Yuan *et al.*, 2003). L'approccio sperimentale utilizzato per questo lavoro è stato basato sulla costruzione di una genoteca di cDNA ottenuta a partire da una frazione di RNA totale, di dimensioni comprese fra 50 e 500 nucleotidi. Nel caso di questa ricerca sono stati caratterizzati 35 snoRNA, di cui solo 15 appartenenti alla classe C/D e 20 alla classe H/ACA; sono state

inoltre identificate due specie di RNA definite H/ACA-like, cioè piccoli RNA che presentano entrambi i motivi H- e ACA-like all'interno di una struttura secondaria simile a quella degli snoRNA di classe H/ACA. In più, questo approccio molecolare ha permesso di identificare anche due snoRNA che presentano un bassissimo livello di conservazione con i corrispondenti U3 e U14 precedentemente isolati da altre specie, la cui identificazione non sarebbe stata possibile mediante analisi di omologia di sequenza. U3 è il più abbondante snoRNA nelle cellule; è essenziale per la vita in lievito ed è ubiquitario in tutti gli eucarioti. La sua funzione, insieme a U14, è quella di mediare alcuni tagli sulle molecole di pre-rRNA mediati da appaiamenti specifici che consentono a U3 e U14 di operare come molecole “chaperone” nel ripiegamento del pre-rRNA. Le interazioni con il pre-rRNA sono garantite da caratteristici motivi presenti nella sequenza di U3: le box GAC, A e A', presenti nella porzione 5' della molecola (Samarsky and Fournier, 1998). In *Drosophila* le due sequenze Dm-818 e Dm-830, pur mostrando i caratteristici motivi sopra elencati, presentano un bassissimo livello di conservazione nell'intera sequenza con lo snoRNA U3. Come U3, anche lo snoRNA U14 funziona come chaperone del pre-rRNA (Liang *et al.*, 1997). In lievito e nei vertebrati questo snoRNA funziona anche come guida nella 2'-*O*-metilazione dell'rRNA 18S (Bachelierie and Cavallé, 1997). L'identificazione in *D. melanogaster* dello snoRNA U14-like, che manca del potenziale per guidare la metilazione dell'rRNA 18S, ha mostrato che questa funzione addizionale non è conservata fra i metazoi.

All'inizio del mio progetto di Dottorato, e fino a due anni fa, le ricerche dei geni che codificano snoRNA condotte in *Drosophila melanogaster* (Tycowski and Steitz, 2001; Yuan *et al.*, 2003) avevano quindi portato all'identificazione di un numero di geni molto inferiore rispetto a quelli emersi dalle ricerche condotte in altri organismi (Lowe and Eddy, 1999; Barneche *et al.*, 2001; Brown *et al.*, 2001; Liang *et al.*, 2002; Hüttenhofer *et al.*, 2001; Cavallé *et al.*, 2001; Vitali *et al.*, 2003). Sebbene in *D. melanogaster* non fossero ancora stati mappati sperimentalmente i siti metilati sullo rRNA, era tuttavia plausibile ritenere che il loro numero fosse intermedio tra i valori caratterizzati in lievito (55 residui metilati) e quelli in Uomo (107 residui metilati) (Lowe and Eddy, 1999; Maden, 1990). Non si aveva poi alcuna informazione sul numero dei siti metilati presenti sulle molecole di snRNA, tRNA o presumibilmente sugli mRNA, elemento che avrebbe ovviamente portato ad incrementare ulteriormente il numero dei geni degli snoRNA presente nel genoma della *Drosophila*. Appariva di conseguenza probabile che molte molecole di snoRNA fossero sfuggite alle analisi molecolari di Yuan *et al.*, (2003) perché sottorappresentate nella genoteca di cDNA, a causa della loro scarsa abbondanza o per la presenza di inusuali strutture secondarie o di modificazioni nucleotidiche che avrebbero potuto impedire la progressione della trascrittasi inversa.

2.2 Impostazione del lavoro sperimentale

L'obiettivo principale del mio progetto di Dottorato è stato quello di identificare, mediante un approccio di "RNomica" computazionale, potenziali geni codificanti per snoRNA presenti nell'intero genoma di *Drosophila melanogaster*. Una prima fase del progetto è stata condotta in collaborazione con il Professor Calogero dell'Università di Torino e ha riguardato la ricerca bioinformatica dei putativi snoRNA. Tale ricerca è stata realizzata avvalendosi del programma SNOSCAN, un efficace algoritmo già utilizzato per l'identificazione di molecole di snoRNA C/D in lievito (Lowe and Eddy, 1999). In una seconda fase, i putativi geni candidati emersi dall'analisi bioinformatica sono stati analizzati in esperimenti di Northern blot, tesi a confermare la loro espressione. Lo SNOSCAN è in grado di identificare possibili snoRNA appartenenti alla classe C/D che guidano le modificazioni di putative molecole di RNA target, ed opera a partire dalla conoscenza delle regioni riconosciute dalle sequenze antisense degli snoRNA e quindi dei siti metilati. Il programma analizza in maniera sequenziale 6 caratteristiche proprie di questi geni; la presenza di **Box D** e **box C**, della **regione di complementarità** con l'eventuale molecola target, l'eventuale presenza di una **box D'** addizionale, la presenza del **sito di metilazione** all'interno della molecola target, e lo **stem terminale** di complementarità, se presente (vedi Fig.1 per una schematizzazione della struttura di un tipico snoRNA). Il programma, inoltre, tiene conto anche della distanza relativa tra le varie componenti esaminate all'interno della sequenza del putativo snoRNA, critica

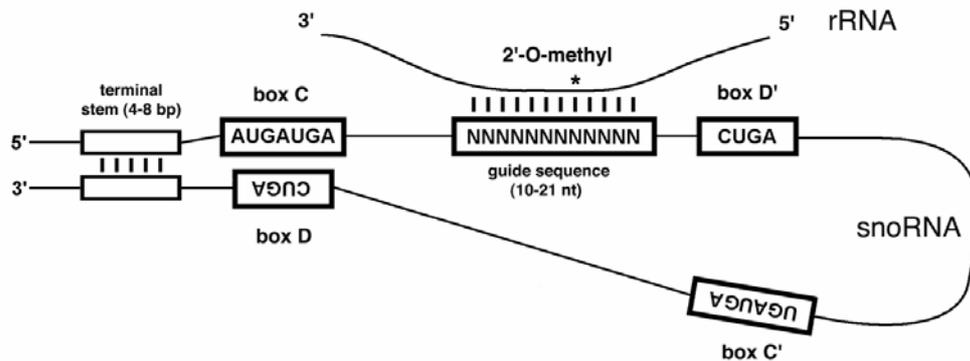
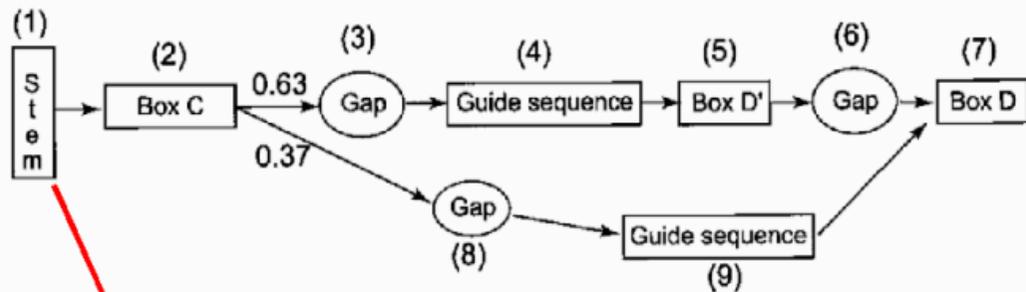


Figura 1: Rappresentazione schematica della struttura di un tipico snoRNA della classe C/D. Il sito di 2'-O-metilazione sullo rRNA è indicato (*) all'interno dell'elica formata dalla sequenza guida dello snoRNA con lo rRNA.

per ridurre i falsi positivi. A ciascun possibile candidato viene poi assegnato uno “score” rispetto ad un modello probabilistico (vedi Fig.2 per una schematizzazione del modello probabilistico). I putativi candidati vengono considerati validi sulla base di uno score logaritmico finale che tiene conto delle informazioni relative a ciascuna caratteristica analizzata.

Essendo gli RNA ribosomali i target preferenziali per la maggior parte degli snoRNA coinvolti nel processo di 2'-O-metilazione del ribosio la mia analisi è stata condotta partendo dalle sequenze degli RNA ribosomali 18S, 28S e 5.8S di *Drosophila melanogaster*. Tuttavia, siccome i siti di metilazione in *Drosophila* non erano ancora stati determinati sperimentalmente (e non lo sono tuttora, se non in minima parte), è stato necessario condurre prima un'analisi genomica comparata con la sequenza delle molecole di rRNA di organismi nei quali fossero già noti i siti di metilazione. E' stato quindi realizzato un allineamento strutturale tra gli rRNA di *D.melanogaster* e quelli di *S.cerevisiae* e *H. sapiens* (<http://www.rna.icmb.utexas.edu>), utilizzando i siti di metilazione

SnoRNA Search



State number	Feature	Model	Consensus	Feature score (bits)		
				Best	Average	Worst
1	Terminal stem	SCFG, 4 to 8 bp	6 bp (when present)	7.60	3.09	0.35
2	Box C	7-bp ungapped HMM	AUGAUGA	12.73	11.63	5.84
3	Gap	Duration model	Length: 6 to 10 bp	-1.59	-2.09	-4.76
4	Guide sequence	HMM	12-bp duplex	15.67	11.11	2.54
5	Box D'	4-bp ungapped HMM	CUGA	7.34	4.85	-3.74
6	Gap	Duration model	Length: 36 to 45 bp	-1.59	-2.43	-5.36
7	Box D	4-bp ungapped HMM	CUGA	8.05	7.92	5.43
8	Gap	Duration model	Length: 56 to 75 bp	-1.50	-2.10	-4.17
9	Guide sequence	HMM	14-bp duplex	18.96	13.98	9.95

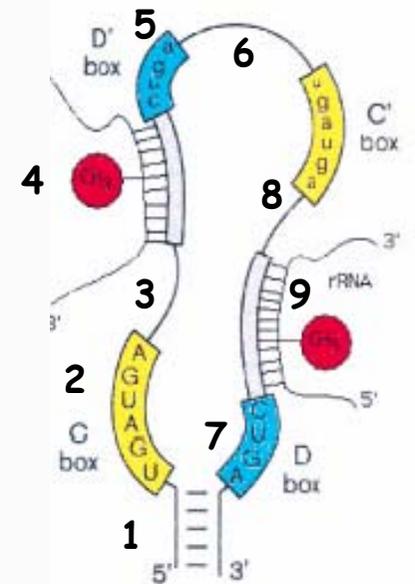


Figura 2: schematizzazione del modello probabilistico utilizzato dal programma SNOSCAN. Nel riquadro a destra sono evidenziate le caratteristiche strutturali delle molecole di snoRNA associate a ciascuno stato del modello.

già definiti sperimentalmente in questi due organismi (http://bioinf.scri.sari.ac.uk/cgi.bin/plant_snoRNA/conservation).

Quest'approccio ha consentito di individuare all'interno delle sequenze conservate tra gli rRNA di lievito e di Uomo dei putativi siti di metilazione presenti sugli rRNA di *Drosophila* (vedi Tabella 1). L'analisi ha portato a stabilire che, tra i putativi siti di metilazione conservati tra lievito e Uomo, il 91% di quelli presenti sull'rRNA 18S e sul 28S sono conservati anche in *Drosophila*; sorprendentemente 3 siti (Cm414/Cm462; Cm1638/Cm1705 sull'rRNA il 18S; 1124/1909 presente sull'rRNA 28S, vedi tabella 1) risultavano conservati tra lievito e Uomo ma non in *Drosophila*. L'assenza in *Drosophila* di questi siti di metilazione era un dato inaspettato ed alquanto inspiegabile, che abbiamo ipotizzato essere legato alla sensibilità dell'algoritmo utilizzato. Quest'ipotesi è stata recentemente avvalorata dai risultati riportati da Huang *et al.* (2005). Questi autori hanno infatti identificato, tramite analisi molecolari, 2 snoRNA C/D capaci di dirigere la metilazione dei siti Cm414/Cm462 e Cm168/Cm1705 presenti sull'rRNA 18S, mentre il sito 1124/1909 presente sull'rRNA 28S sembra essere bersaglio di pseudouridilazione da parte dello snoRNA H/ACA 28S- ψ 1192. E' poi interessante notare che 4 siti di metilazione presenti sull'rRNA 18S (21%) e 4 di quelli presenti sul 28S (9%) rappresentano dei putativi siti conservati solo tra *Drosophila* e Uomo, mentre inoltre 2 siti di metilazione sull'rRNA 18S e 6 sull'rRNA 28S sono specifici di *Drosophila* essendo stati predetti dal programma SNOSCAN.

Tabella 1: Putativi snoRNAs identificati nel genoma di Drosophila utilizzando il programma SNOSCAN e i siti di metilazione conservati tra Drosophila e lievito e Drosophila e uomo.

Putativo RNA target	SnoRNA identificati in Dm	Siti di metilazione predetti	SnoRNA umani	Siti di metilazione umani	SnoRNA di lievito	Siti di metilazione in lievito	Conservazione dei siti di metilazione
18S	DmSnR74/Z4	Am28	U27	Am27	SnR74	Am28	DHY
18S	Dm425	Am425	-	-	-	-	D [@]
18S	DmSnR54	Am1061	U59	Am1031	SnR54	Am975	DHY
18S	DmSnR55	Um1356	U33	Um1326	SnR55	Um1267	DHY
18S	DmSnR40	Gm1358	-	-	-	Gm1269	DY
18S	DmSnR56a	Gm1620	U25	Gm1490	SnR56	Gm1427	DHY
18S	Dm1643	Cm1643	-	-	-	-	D [@]
18S	-	-	U14	Cm462	U14	Cm414	HY*
18S	DmU16	Am441	U16	Am484	-	-	DH [°]
18S	-	Am627	U36a	Am668	SnR47	Am619	DHY
18S	-	-	U43	Cm1705	SnR70	Cm1638	HY*
18S	DmU54	Gm603	U54	Gm644	-	-	DH [°]
18S	-	Cm475	U56	Cm517	-	-	DH [°]
18S	-	-	U61	Um1442	-	-	H
18S	-	Am161	U44	Am166	-	-	DH [°]
18S	-	Am549	Nd	Am590	SnR41	Am541	DHY
18S	-	Um1086	Nd	Um1056	SnR31	Um998	DHY
18S	-	-	-	-	SnR53	Am796	Y
18S	-	Um586	Nd	Um627	SnR77	Um578	DHY
25S	DmU18(a, b)	Am773	U18	Am1313	U18	Am647	DHY
25S	DmU18adj	Cm774	-	-	-	Cm648	DY

Continuazione Tabella 1

25S	DmSnR39-59ab	Am981	U51/U32a	Am1511	SnR39-59	Am805	DHY
25S	DmSnR72/Z2	Am1050	-	-	SnR72	Am874	DY
25S	DmSnR60(a-d)	Gm1082	U80	Gm1612	SnR60	Gm906	DHY
25S	DmSnR61(a-c)	Am1321	U38ab	Am1858	SnR61	Am1131	DHY
25S	DmG1322	Gm1322	-	-	-	-	D [@]
25S	DmU1332	Um1332	-	-	-	-	D [@]
25S	DmU24	Cm1652	U24	Cm2338	U24	Cm1435	DHY
25S	DmSnR62(a, b)	Um2133	U34	Um2824	SnR62	Um1886	DHY
25S	Dm2466	Am2466	-	-	-	-	D [@]
25S	DmSnR47(a-c)	Am2527	U36	Am3703	SnR47	Am2218	DHY
25S	-	-	-	-	SnR75	Gm2286	Y
25S	DmSnR64(a-c)	Cm2644	U74	Cm3820	SnR64	Cm2335	DHY
25S	DmSnR66	Um2724	U52	Um3904	SnR78	Um2414	DHY
25S	Dm3049	Cm3049	-	-	-	-	D [@]
25S	DmSnR67	Gm3080	U31	Gm4166	SnR67	Gm2616	DHY
25S	Dm311 & Dm3112a	Gm3112	-	-	-	-	D [@]
25S	DmSnR48(a-c)	Gm3254	-	-	SnR48	Gm2788	DY
25S	Dm3403(a-d)	Cm3403	-	-	-	-	D [@]
25S	DmSnR71(a, b)	Cm3406	U29	Am4493	SnR71	Am2943	DHY
25S	DmSnR69(a, b)	Cm3408	-	-	SnR69	Cm2945	DY
25S	-	Gm2596	-	-	SnR75	Gm2286	DY
25S	DmU30	Am2628	U30	Am3804	-	-	DH [°]
25S	-	Cm3419	U35	Cm4506	SnR73	Cm2956	DHY
25S	DmSnR39B	Gm979	Nd	Gm1509	SnR39B	Gm803	DHY
25S	-	Am2588	U15	Am3764	SnR13	Am2279	DHY
25S	DmU37	Am2521	U37	Am3697	-	-	DH [°]
25S	-	-	U49	Cm2409	-	-	H
25S	DmU55/39	Cm2101	U55 / U39	Cm2791	-	-	DH [°]

Continuazione Tabella 1

25S	-	Um3287	U65	Um4373	SnR34	Um2823	DHY
25S	DmSnR58(a, b)	Cm787	Nd	Cm1327	SnR58	Cm661	DHY
25S	-	Um3341	U65	Um4427	SnR34	Um2877	DHY
25S	-	Um3185	U31	Gm4166	SnR67	Gm2616	DHY
25S	-	-	U53	Am3848	-	-	H
25S	-	Um2728	U52	Um3904	SnR78	Um2414	DHY
25S	-	Gm3276	-	-	SnR38	Gm2812	DY
25S	-	Gm979	Nd	Gm1509	SnR39B	Gm803	DHY
25S	DmU79	Am2633	U79	Am3809	-	-	DH°
25S	-	Am3101	-	-	SnR68	Am2637	DY
25S	-	-	N.d.	1909	SnR5	1124	HY*
25S	-	Um1177	N.d.	Um1717	SnR49	Um990	DHY
25S	-	Um1191	N.d.	Um1731	SnR5	Um1004	DHY

La numerazione degli rRNA di lievito e Uomo è basata sulle sigle di accesso alla banca dati Genbank: M21017 (D), U13369 (H) e della banca dati per gli snoRNA guida della metilazione per *S.cerevisiae*. H-Uomo. D-Drosophila, Y-lievito. (I dati relativi a lievito e all'Uomo sono stati estratti dal sito http://bioin.scri.sari.ac.uk/cgi-bin/plant_snorna=conservation)

*Siti di metilazione conservati solo tra lievito e Uomo.

°Siti di metilazione conservati solo tra Drosophila e Uomo

@Siti di metilazione specifici di Drosophila.

I putativi siti di metilazione individuati attraverso i due allineamenti condotti sono poi stati forniti come parametri iniziali al programma SNOSCAN e definiti attendibili se consentivano l'identificazione, rispettivamente nel genoma di lievito o in quello umano, del corrispondente snoRNA con uno "score" maggiore di 20 bits, che è il parametro di default definito da Lowe and Eddy, 1999. Le putative sequenze geniche identificate sono state poi allineate con lo scaffold genomico di *D. melanogaster* (disponibile presso <http://www.flybase.org>) mediante l'utilizzo dell'algoritmo BLAST, allo scopo di definire con precisione la localizzazione genomica dei putativi geni di snoRNA identificati dall'analisi computazionale.

La conferma sperimentale dell'espressione dei putativi geni identificati attraverso la ricerca bioinformatica è stata effettuata determinando il loro profilo di espressione attraverso esperimenti di Northern blot. L'analisi per Northern blot rappresenta uno dei metodi standard utilizzati per l'identificazione e l'analisi dei trascritti e fornisce inoltre informazioni sui pattern di espressione temporali, così come sull'identificazione di potenziali precursori o prodotti di degradazione. Nei miei esperimenti, le sequenze fiancheggianti i geni candidati, ricavate dalla banca dati (<http://www.flybase.org>) sono state analizzate per la sintesi e la progettazione di oligonucleotidi opportuni (di lunghezza compresa tra i 18 e i 30 nt) che hanno consentito di amplificare, mediante una reazione di PCR, un tratto di sequenza genomica lungo mediamente 500 bp, comprendente il putativo gene identificato dallo SNOSCAN. Il frammento è stato poi clonato e successivamente marcato con isotopi radioattivi (vedi Materiali e Metodi); la

sonda ottenuta è stata utilizzata per ibridare membrane sulle quali erano state precedentemente trasferite aliquote di RNA totale estratto dai vari stadi di sviluppo della *Drosophila*, allo scopo di evidenziare anche eventuali profili di espressione non costitutivi, come ipotizzabile per eventuali molecole di snoRNA che venissero a mancare della regione di complementarità con rRNA, tRNA, snRNA o altri RNA stabili (Hüttenhofer *et al.*, 2001; Kiss *et al.*, 2001). Considerando la tendenza dei geni codificanti per snoRNA ad essere organizzati in cluster, è stata poi condotta un'analisi delle sequenze fiancheggianti ciascun gene candidato per cercare eventuali isoforme localizzate nelle adiacenze e per verificare la presenza di altri snoRNA sia di tipo sia C/D che H/ACA, attraverso l'uso del programma BLAST o per analisi diretta delle sequenze. Il progetto segue schematizzato in figura 3.

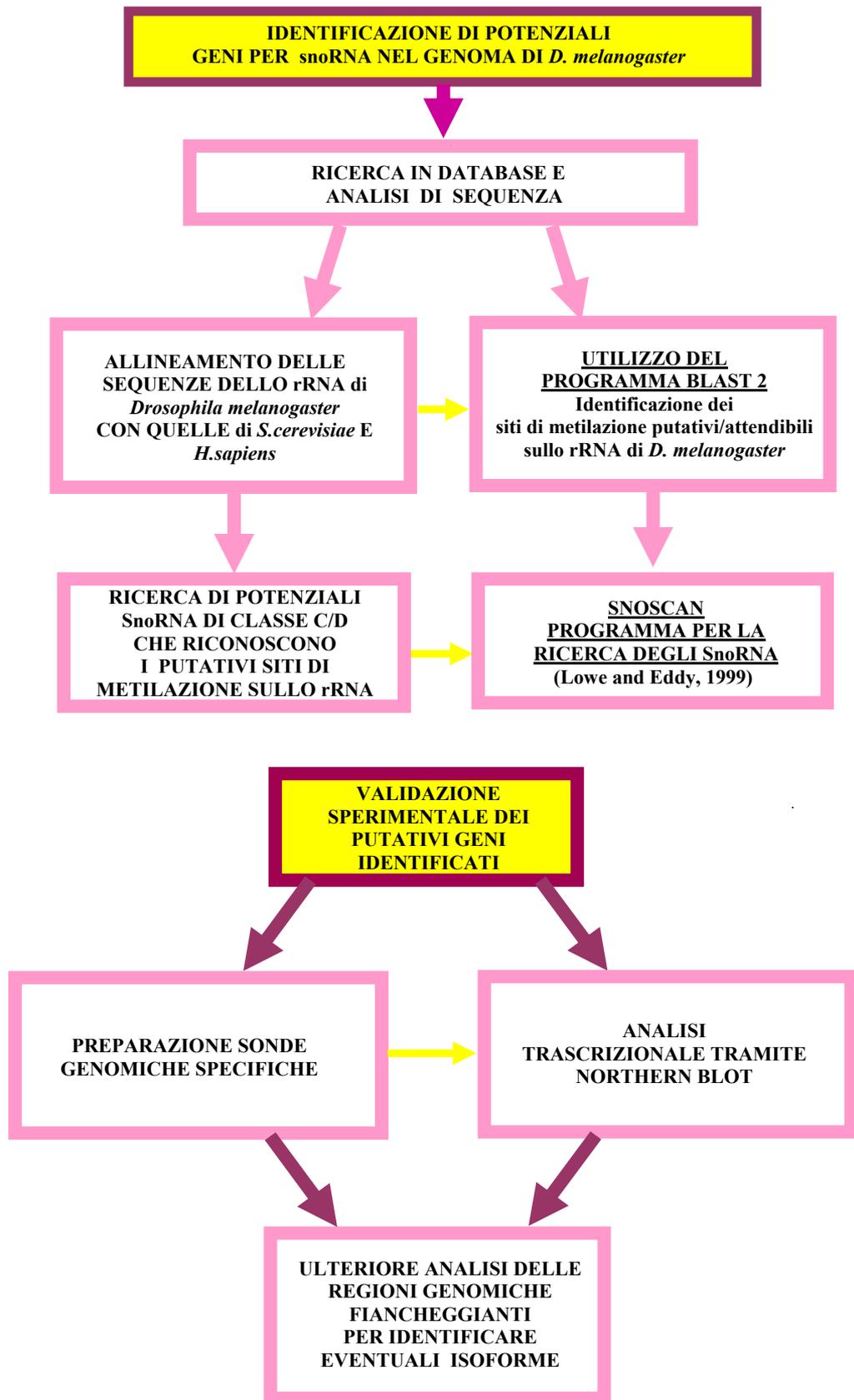


Figura 3. Schematizzazione delle fasi dell'approccio di "RNomica computazionale" utilizzato per l'identificazione di putativi geni codificanti per snoRNA nel genoma di *Drosophila melanogaster*.

2.3 Identificazione, tramite analisi bioinformatica, di putativi geni codificanti per snoRNA nel genoma di *Drosophila melanogaster*.

L'analisi del genoma di *Drosophila*, realizzata utilizzando come punto di partenza i putativi siti di metilazione conservati tra gli rRNA di *S.cerevisiae*, *D.melanogaster* e *H.sapiens* ha condotto all'identificazione di ben 100 (44 derivati dall'analisi condotta tra lievito e *Drosophila* e 56 dall'analisi condotta tra *Drosophila* e Uomo) putativi geni codificanti per snoRNA di tipo C/D presumibilmente coinvolti nella metilazione degli RNA ribosomali 18S e 28S, mentre non è stato identificato alcun gene presumibilmente coinvolto nella metilazione dello rRNA 5.8S.

Per quel che riguarda la localizzazione cromosomica di tutti i geni identificati, la maggioranza di questi è risultata prevalentemente distribuita sui cromosomi 2 e 3 di *Drosophila*, una minoranza mappa sul cromosoma X, ed infine solo 2 sul quarto cromosoma; nessun gene è stato identificato sul cromosoma Y, che insieme al cromosoma 4 è noto essere essenzialmente eterocromatico (Fig.4 e 5). Nelle tabelle 2 e 3 sono riportati i risultati dell'analisi delle caratteristiche strutturali di ciascun gene candidato e la rispettiva localizzazione genomica, nonché la complementarietà di sequenza con l'RNA ribosomale. Per i 44 candidati identificati dall'analisi dei siti conservati tra gli rRNA di lievito e *Drosophila* è emerso che l'assoluta maggioranza (40/44) presenta la box D' addizionale; quasi la metà mostra entrambe le box canoniche (19/44), mentre una frazione di poco inferiore (18/44) presenta la box D canonica e la box D' variante (in media in un solo nucleotide); solo in tre casi (3/44) entrambe le box D e D' risultano varianti, anche se in un solo nucleotide (Fig.6).

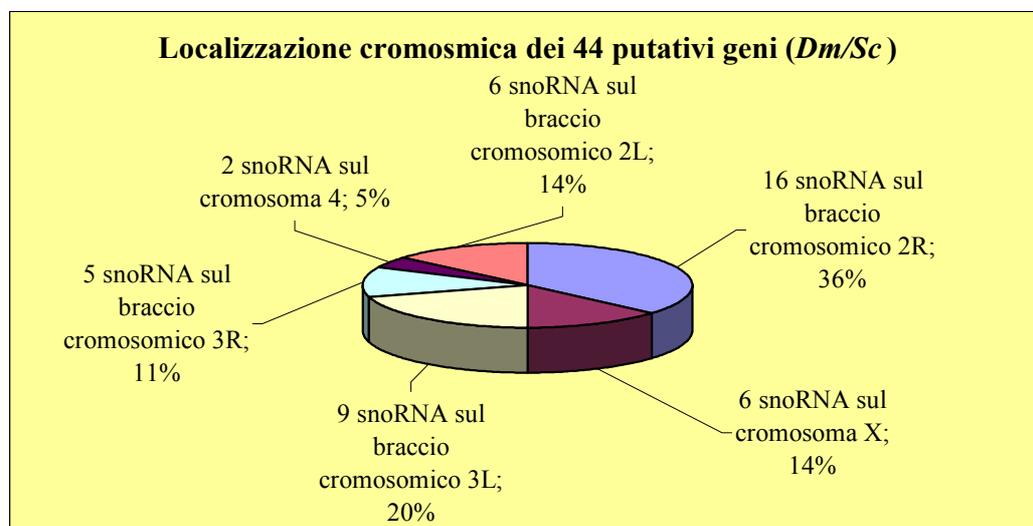


Figura 4

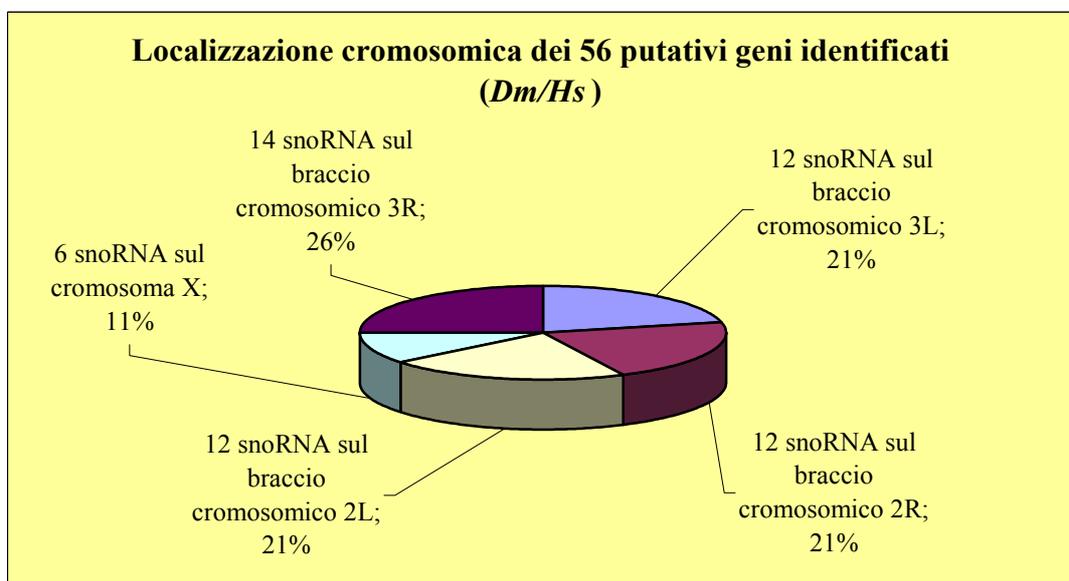


Figura 5

**Caratteristiche delle box D e D' dei putativi geni candidati
(*Dm/Sc*)**

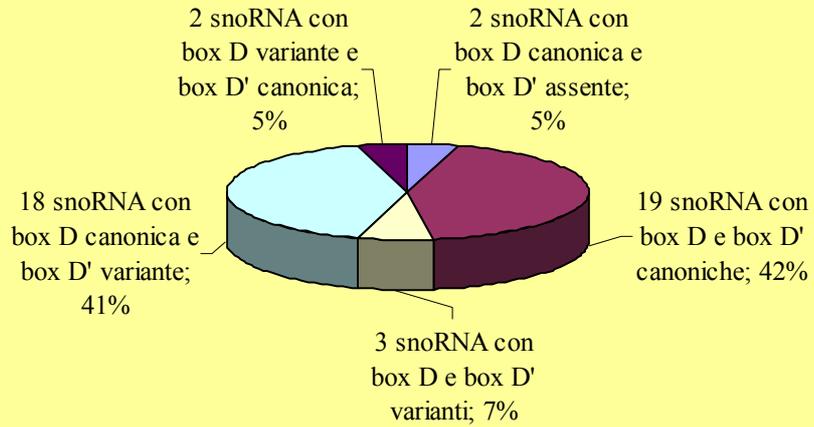


Figura 6

**Posizionamento della sequenza di riconoscimento per lo
rRNA per i putativi geni (*Dm/Sc*)**

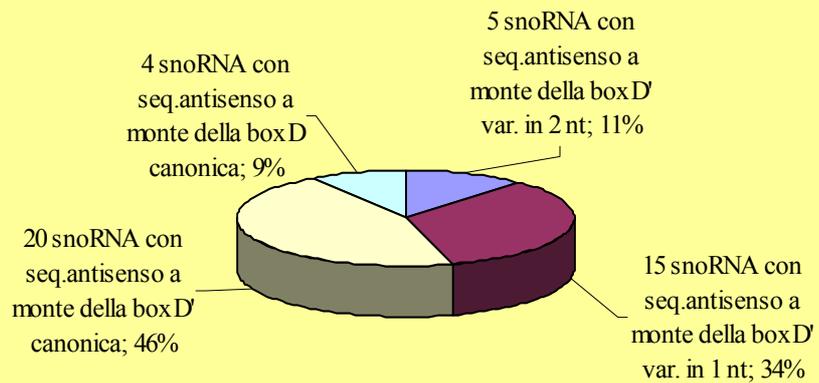


Figura 7

La sequenza antisenso che mostra complementarità con l'RNA ribosomiale è risultata localizzata per la maggior parte dei geni candidati (40/44) a monte della box D' (Fig.7); un dato questo abbastanza insolito. Anche per i 56 geni identificati partendo dai siti di metilazione conservati tra gli rRNA di *Drosophila* e Uomo l'assoluta maggioranza (49/56) presenta la box D' addizionale; per quanto riguarda, invece, le loro caratteristiche strutturali, è emerso che solo 13 putativi snoRNA presentano entrambe le box D e D' (*CUGA*) canoniche, mentre la maggioranza (27/56) mostra la box D canonica e la box D' variante (in media in un solo nucleotide, vedi Fig.8). Inoltre, analogamente a quanto osservato per i geni identificati partendo dall'allineamento con l'rRNA di lievito, la maggioranza dei putativi candidati (49/56) presenta la sequenza antisenso posizionata a monte della box D', mentre solo per 7 di essi tale sequenza è posizionata a monte della box D (Fig.9).

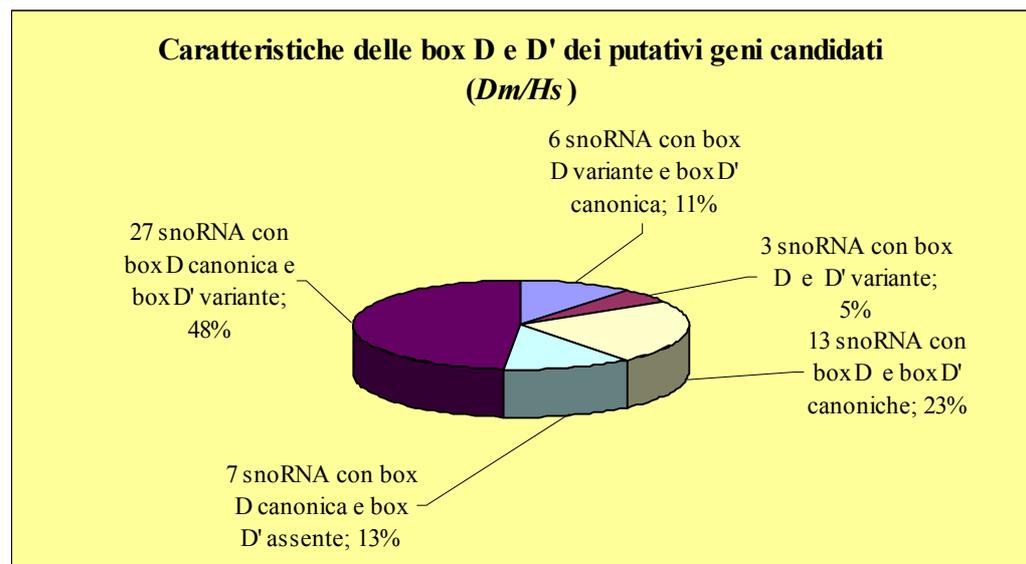


Figura 8

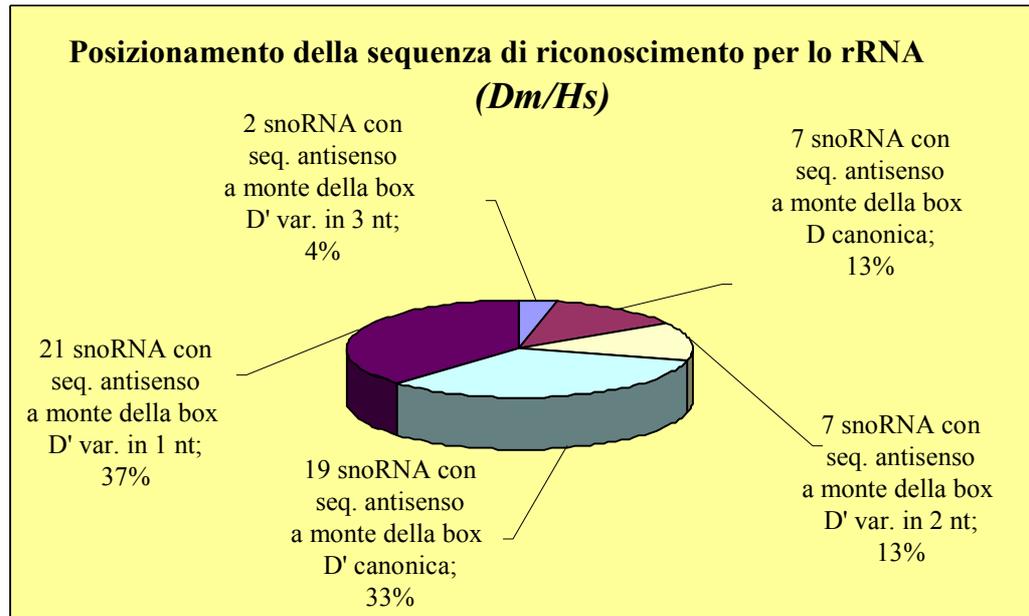


Figura 9

Come già notato, questa caratteristica è insolita perché la sequenza guida complementare alla molecola di RNA target è posizionata a monte della box D, e solo in pochi casi era stata finora riscontrata un'organizzazione di tipo diverso (Tycowski *et al.*, 1998; Darzacq *et al.*, 2002). Questo dato ha suggerito l'ipotesi che le molecole identificate potessero essere bifunzionali e cioè che la sequenza antisense a monte della box D potesse riconoscere sequenze target diverse dallo rRNA. Quest'ipotesi appare rafforzata dall'osservazione che dal nostro screening computazionale è emerso che per i 44 geni identificati dall'allineamento delle sequenze degli rRNA di lievito e *Drosophila*, solo il 65% (29/44) rispetta l'appaiamento previsto con lo rRNA, mentre per quelli identificati in base all'allineamento Uomo/*Drosophila*, solo 10/56 putativi snoRNA (18%) presentavano un appaiamento perfetto con lo RNA ribosomale. Anche questo risultato è apparso abbastanza inaspettato, essendo il nostro approccio basato in modo specifico sull'identificazione di snoRNA potenzialmente capaci di guidare la metilazione di rRNA; nello stesso tempo,

esso ha però dimostrato anche l'efficacia dell'approccio utilizzato. Infatti, lo SNOSCAN, pur partendo dalla complementarità con l'RNA ribosomale, determina lo score finale sulla base dell'insieme delle caratteristiche strutturali ed è quindi stato in grado di identificare come geni candidati anche quelli la cui sequenza antisenso non era perfettamente complementare ai potenziali siti metilati presenti sull'rRNA. L'insieme di queste osservazioni ha suggerito che alcuni degli snoRNA identificati che mancavano della complementarità con lo rRNA potessero appartenere alla categoria di snoRNA cosiddetti "orfani", per i quali restava da definire l'RNA bersaglio. Nei mammiferi, la maggior parte degli "orfani" è espressa in modo tessuto-specifico; inoltre l'identificazione di nuovi snoRNA espressi esclusivamente nel cervello e privi di complementarità per lo rRNA o per gli snRNA, ha per la prima volta suggerito nuovi possibili ruoli per gli snoRNA (Cavaillè *et al.*, 2001). Infatti, gli snoRNA possono essere coinvolti nella metilazione non solo degli rRNA 18S, 28S e 5.8S, ma anche di molecole di snRNA (Tycowsky *et al.*, 1998; Kiss, 2001) o di molecole di RNA messaggero (Vitali *et al.* 2005), regolando quindi l'espressione di geni che codificano per proteine. Le molecole di snoRNA capaci di modificare i messaggeri potrebbero agire a livello post-trascrizionale modificando gli mRNA che transitano nel nucleolo, ostacolando i processi di editing, influenzando lo splicing alternativo dei trascritti primari oppure ostacolando la traduzione od il riconoscimento degli mRNA da parte di proteine RNA-binding. Recentemente è stato dimostrato che lo snoRNA "orfano" di mammifero MBII-52, che presenta una lunga sequenza di 18 nucleotidi complementare alla molecola di mRNA del recettore 5-HT_{2C} della

serotonina (Vitali *et al.*, 2005), può regolare lo splicing alternativo dello mRNA bersaglio mediante la modificazione di un nucleotide adiacente al sito di splicing. La modificazione sembrerebbe anche svolgere un ruolo antagonista rispetto al processo di editing, andando a ridurre di circa 200 volte l'efficienza di deaminazione del residuo adenosinico bersaglio della 2'-O-metilazione.

Tabella 2: Caratteristiche strutturali e localizzazione dei 44 putativi snoRNA identificati in base all'analisi computazionale dei siti conservati tra gli rRNA di lievito e Drosophila.

Simbolo SNOSCAN	Localizzazione CROMOSOMICA	Localizzazione GENICA	MISMATCH Seq.Guida/rR NA	Caratteristiche D Box (CUGA)	Caratteristiche D'Box (CUGA)	Localizzazione Seq.ANTISENS O
Dm 18S.001	2R	intergenico	0/13	canonica	canonica	upstream D' Box
Dm 18S.002	3L	intergenico	0/10	canonica	canonica	upstream D' Box
Dm 18S.003	2L	intergenico	0/10	canonica	canonica	upstream D' Box
Dm 18S.004	3L	intronic	0/10	canonica	var.in 2nt(CAAA)	upstream D' Box
Dm 18S.005 ^b	3R	intronic	0/11	canonica	canonica	upstream D' Box
Dm 18S.005a ^{b,c}	3R	intronic	0/11	canonica	canonica	upstream D' Box
Dm 18S.007¹	2L	intronic¹	0/12	canonica	canonica	upstream D' Box
Dm 18S.008	2R	intertrasposonico	0/13	canonica	canonica	upstream D' Box
Dm 18S.009¹	2R	intronic¹	0/14	canonica	canonica	upstream D' Box
Dm 18S.010	X	Esonico	0/14	canonica	assente	upstream D Box
Dm 18S.011	3R	intr/esone	0/14	canonica	canonica	upstream D Box
Dm 28S.002	3L	intronic	0/13	canonica	var.in 1nt (AUGA)	upstream D' Box
Dm 28S.002a	3L	intronic	0/13	canonica	var.in 1nt (AUGA)	upstream D' Box
Dm 28S.004	3L	esonico	3/12	var.in 1nt (AUGA)	canonica	upstream D' Box
Dm 28S.005^b	3L	intronic	0/14	var.in 1nt (AUGA)	var.in 1nt (AUGA)	upstream D' Box
Dm 28S.005a^{b,c}	3L	intronic	0/14	var.in 1nt (AUGA)	var.in 1nt (AUGA)	upstream D' Box
Dm 28S.006	X	intergenico	2/13	canonica	var.in 2nt(CGAA)	upstream D' Box
Dm 28S.007	2L	esonico	2/9	canonica	canonica	upstream D' Box
Dm 28S.008^b	2R	intergenico	0/12	canonica	canonica	upstream D' Box
Dm 28S.008a^{b,c}	2R	intergenico	1/12	canonica	canonica	upstream D' Box
Dm 28S.010^b	2R	intronic	0/13	canonica	canonica	upstream D' Box
Dm 28S.011^b	2R	intronic	0/13	canonica	canonica	upstream D' Box
Dm 28S.010/011a^{b,c}	2R	intronic	1/13	canonica	canonica	upstream D' Box
Dm 28S.010/011b^{b,c}	2R	intronic	0/13	canonica	canonica	upstream D' Box
Dm 28S.012	2L	intergenico	1/13	canonica	canonica	upstream D' Box
Dm 28S.013	X	intronic	1/10	canonica	canonica	upstream D' Box
Dm 28S.014	3L	intergenico	1/12	canonica	var.in 1nt (CUGU)	upstream D' Box
Dm 28S.015	2R	intertrasposonico	2/15	canonica	var.in 2nt(CAAA)	upstream D' Box
Dm 28S.016	2R	intronic	0/10	canonica	canonica	upstream D' Box
Dm 28S.017	3L	eso e intro	1/9	canonica	var.in 1nt(CAGA)	upstream D' Box
Dm 28S.018	3R	eso e intro	0/15	canonica	var.in 2nt(UUGA)	upstream D' Box
Dm 28S.018a	3R	intronic	0/14	canonica	var.in 2nt(UUGA)	upstream D' Box
Dm 28S.019	X	intergenico	0/11	canonica	var.in 1nt(CGGA)	upstream D Box
Dm 28S.020	2L	esonico	0/10	canonica	canonica	upstream D' Box

Continuazione Tabella 2

Dm 28S.021	2R	intronic	0/14	canonica	var.in 1nt(CUGG)	upstream D' Box
Dm 28S.023	2L	intronic	0/12	canonica	canonica	upstream D' Box
Dm 28S.023a^b	2L	intronic	0/12	canonica	canonica	upstream D' Box
Dm 28S.023b^{b,c}	2L	intronic	0/12	canonica	canonica	upstream D' Box
Dm 28S.024	2R	esonico	1/13	var.in 1nt (CCGA)	canonica	upstream D' Box
Dm 28S.025	4	intertrasposonico	2/14	canonica	var.in 1nt (CCGA)	upstream D' Box
Dm 28S.026	2R	intronic	1/11	canonica	canonica	upstream D' Box
Dm 28S.027	2R	intronic	0/14	canonica	var.in 1nt (AUGA)	upstream D' Box
Dm 28S.028	2R	intergenico	0/16	canonica	var.in 1nt (AUGA)	upstream D' Box
Dm 28S.031¹	2R	intronic	0/11	canonica	var.in 1nt (UUGA)	upstream D' Box
Dm 28S.032¹	2R	Intronic	0/11	canonica	var.in 1nt (UUGA)	upstream D' Box
Dm 28S.036a^b	3L	esonico	1/11	canonica	assente	upstream D' Box
Dm 28S.036b^{b,c}	3L	eso/intro	Nessun appaiamento	canonica	assente	assente
Dm 28S.038	X	esonico	2/15	canonica	var.in 2nt(AUCA)	upstream D' Box
Dm 28S.039	X	esonico	2/12	canonica	canonica	upstream D' Box
Dm 28S.040	3R	esonico	3/21	canonica	var.in 1nt (CAGA)	upstream D' Box
Dm 28S.041¹	2R	intronic	0/14	canonica	canonica	upstream D' Box
Dm 28S.043	3L	intronic	0/13	canonica	var.in 1nt (AUGA)	upstream D' Box
Dm 28S.043a	3L	intronic	Nessun appaiamento	var.in 2nt(CCAA)	assente	assente
Dm 28S.044	3R	intergenico	1/18	var.in 1nt(AUGA)	var.in 1nt(CCGA)	upstream D' Box
Dm 28S.046a^b	4	esonico	0/11	canonica	canonica	upstream D' Box
Dm 28S.046b^{b,c}	4	esonico	0/11	canonica	canonica	upstream D' Box
Dm 28S.046c^{b,c}	4	esonico	0/11	canonica	canonica	upstream D' Box
Dm 28S.046d^{b,c}	4	esonico	0/11	canonica	canonica	upstream D' Box

¹SnoRNA identificati anche da Tycowski and Steitz, 2001.

^bTutti gli snoRNA indicati con lo stesso numero (ex:Dm_28S.018, Dm_28S.018a) sono localizzati nello stesso cluster

^cIsoforme identificate attraverso ricerche per omologia tramite BLAST delle sequenze circostanti il gene originariamente identificato dallo SNOSCAN, che presentano sequenze altamente correlate.

I geni confermati sperimentalmente sono riportati in grassetto.

Tabella 3: Caratteristiche strutturali e localizzazione dei 56 putativi snoRNA identificati in base all'analisi computazionale dei siti conservati tra gli rRNA di Uomo e Drosophila.

Simbolo SNOSCAN	Localizzazione Cromosomica	Localizzazione Genica	Mismatch Seq. Guida / rRNA	D Box (CUGA) Caratteristiche	D' Box (CUGA) Caratteristiche	Localizzazione della sequenza antisense
Dm 18S.1019	X	esonico	3 / 17 bp	D Box var. in 1 nt (AUGA)	D' Box canonica	upstream D' Box
Dm 18S.1018	X	intronic	2 / 14 bp	D Box var. in 1 nt (AUGA)	D' Box canonica	upstream D' Box
Dm 18S.1001	3L	intergenico	0 / 10 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 18S.1045	2L	intergenico	1 / 12 bp	D Box canonica	D' Box var. in 1 nt (CUGC)	upstream D' Box
Dm 18S.1031	3R	intronic	1 / 10 bp	D Box canonica	Assente	upstream D Box
Dm 28S.1014	X	esonico	1 / 14 bp	D Box canonica	Assente	upstream D Box
Dm 28S.1016	X	esonico	2 / 19 bp	D Box var. in 1 nt (AUGA)	D' Box var. in 1 nt (CUUA)	upstream D' Box
Dm 28S.1017	X	esonico	0 / 11 bp	D Box canonica	D' Box var. in 3 nt (AGGC)	upstream D' Box
Dm 28S.1053	2L	intertrasposonico	1 / 13 bp	D Box var. in 1 nt (AUGA)	D' Box var. in 1 nt (CAGA)	upstream D' Box
Dm 28S.1015	X	intronic	2 / 15 bp	D Box var. in 1 nt (AUGA)	D' Box canonica	upstream D' Box
Dm 28S.1050	2L	esonico	1 / 12 bp	D Box canonica	D' Box var. in 2 nt (AUGG)	upstream D' Box
Dm 28S.1049	2L	eso. / intr. ^^	1 / 11 bp	D Box canonica	Assente	upstream D Box
Dm 28S.1051	2L	intertrasposonico	1 / 10 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1052	2L	intergenico	2 / 13 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1037	2R	esonico	1 / 11 bp	D Box var. in 1 nt (AUGA)	D' Box canonica	upstream D' Box
Dm 28S.1036	2R	esonico	1 / 10 bp	D Box canonica	D' Box var. in 2 nt (CCGC)	upstream D' Box
Dm 28S.1035	2R	esonico	1 / 14 bp	D Box canonica	D' Box var. in 1 nt (CUGG)	upstream D' Box
Dm 28S.1032	2R	eso. / intr. ^^	1 / 13 bp	D Box canonica	Assente	upstream D Box
Dm 28S.1033	2R	esonico	1 / 11 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1005	3L	esonico	2 / 13 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1003	3L	intronic	2 / 16 bp	D Box var. in 1 nt (AUGA)	D' Box canonica	upstream D' Box
Dm 28S.1001	3L	intronic	2 / 17 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1002	3L	intronic	2 / 15 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1004	3L	eso. / intr. ^^	1 / 13 bp	D Box canonica	D' Box var. in 1 nt (CUGU)	upstream D' Box
Dm 28S.1030	3R	esonico	2 / 17 bp	D Box canonica	Assente	upstream D Box
Dm 28S.1027	3R	intergenico	0 / 10 bp	D Box canonica	D' Box var. in 2 nt (UUCA)	upstream D' Box
Dm 28S.1029	3R	esonico	1 / 12 bp	D Box canonica	D' Box var. in 1 nt (CCGA)	upstream D' Box
Dm 28S.1028	3R	esonico	1 / 12 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1056	2L	intergenico	0 / 10 bp	D Box var. in 1 nt (AUGA)	D' Box var. in 1 nt (CUUA)	upstream D' Box
Dm 28S.1043	2R	esonico e intronic ^	2 / 11 bp	D Box canonica	Assente	upstream D Box
Dm 28S.1010	3L	intertrasposonico	2 / 12 bp	D Box canonica	D' Box var. in 1 nt (CAGA)	upstream D' Box
Dm 28S.1026	3R	intergenico	1 / 15 bp	D Box canonica	D' Box var. in 1 nt (AUGA)	upstream D' Box

Continuazione Tabella 3

Dm 28S.1044	2R	intronico	2 / 11 bp	D Box canonica	D' Box var. in 1 nt (GUGA)	upstream D' Box
Dm 28S.1025	3R	eso. / intr. ^^	1 / 12 bp	D Box canonica	D' Box var. in 1 nt (GUGA)	upstream D' Box
Dm 28S.1006*	3L	intergenico	0 / 12 bp	D Box canonica	D' Box var. in 1 nt (UUGA)	upstream D' Box
Dm 28S.1007	3L	intergenico	0 / 12 bp	D Box canonica	D' Box var. in 1 nt (UUGA)	upstream D' Box
Dm 28S.1008**	3L	intergenico	0 / 12 bp	D Box canonica	D' Box var. in 1 nt (UUGA)	upstream D' Box
Dm 28S.1046	2L	intergenico	2 / 13 bp	D Box var. in 1 nt (AUGA)	D' Box canonica	upstream D' Box
Dm 28S.1047	2L	esonico	3 / 15 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1048	2L	eso. / intr. ^^	1 / 10 bp	D Box canonica	D' Box var. in 1 nt (CAGA)	upstream D' Box
Dm 28S.1042	2R	eso. / intr. ^^	0 / 12 bp	D Box canonica	D' Box var. in 1 nt (UUGA)	upstream D' Box
Dm 28S.1041	2R	esonico	1 / 10 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1011*	3L	intergenico	1 / 17 bp	D Box canonica	D' Box var. in 3 nt (CCUC)	upstream D' Box
Dm 28S.1012**	3L	intergenico	1 / 17 bp	D Box canonica	D' Box var. in 2 nt (AUAA)	upstream D' Box
Dm 28S.1013	3L	intronico	1 / 11 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1024	3R	intergenico	2 / 11 bp	D Box canonica	D' Box var. in 1 nt (CGGA)	upstream D' Box
Dm 28S.1055	2L	esonico	3 / 15 bp	D Box canonica	D' Box var. in 1 nt (CAGA)	upstream D' Box
Dm 28S.1054	2L	esonico	2 / 11 bp	D Box canonica	D' Box var. in 2 nt (AAGA)	upstream D' Box
Dm 28S.1038	2R	eso. / intr. ^^	0 / 14 bp	D Box canonica	D' Box var. in 1 nt (CUGG)	upstream D' Box
Dm 28S.1039	2R	eso. / intr. ^^	2 / 13 bp	D Box canonica	D' Box var. in 2 nt (CCGC)	upstream D' Box
Dm 28S.1040	2R	eso e intr^	0 / 10 bp	D Box canonica	D' Box var. in 1 nt (CAGA)	upstream D' Box
Dm 28S.1009	3L	eso. / intr. ^^	1 / 10 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1021	3R	esonico	1 / 11 bp	D Box canonica	D' Box canonica	upstream D' Box
Dm 28S.1020	3R	eso. / intr. ^^	1 / 18 bp	D Box canonica	Assente	upstream D' Box
Dm 28S.1022	3R	intergenico	1 / 14 bp	D Box canonica	D' Box var. in 1 nt (UUGA)	upstream D' Box
Dm 28S.1023	3R	intronico	1 / 16 bp	D Box canonica	D' Box var. in 2 nt (CGAA)	upstream D' Box

*1006 e 1011 presentano sequenze altamente correlate ma box D' diverse e diverse sequenze antisenso.

**1008 e 1012 presentano sequenze altamente correlate ma box D' diverse e diverse sequenze antisenso.

^^snoRNA localizzati in regioni sottoposte a splicing alternativo.

^snoRNA localizzati a cavallo di giunzioni esone/introne.

I geni confermati sperimentalmente sono riportati in grassetto.

2.4 Convalida sperimentale dei putativi geni identificati

2.4.1 Parte I: Analisi dei putativi geni identificati in base alla presenza di siti di metilazione conservati fra le sequenze degli rRNA di *Saccharomyces cerevisiae* e di *Drosophila melanogaster*.

Dai saggi tesi a confermare sperimentalmente i putativi geni identificati tramite l'approccio bioinformatico basato sull'allineamento dell'rRNA di lievito e *Drosophila* è emerso che 26 dei 44 geni candidati identificati sono effettivamente espressi in *Drosophila*. Dei geni identificati, 4 presentano sequenze antisense complementari all'rRNA 18S mentre 22 mostrano complementarità con il 28S. L'analisi delle sequenze fiancheggianti i nuovi snoRNA identificati ha portato all'identificazione di isoforme altamente correlate ad alcuni snoRNA inizialmente identificati dallo SNOSCAN (Dm_28S.002, Dm_28S.005, Dm_28S.008, Dm_28S.010/011, Dm_28S.018, Dm_28S.023, Dm_28S.036, Dm_28S.043, Dm_28S.046, vedi tabelle 4 e 5), portando così il numero totale di geni confermati sperimentalmente a 39 confermando che la maggior parte dei geni identificati è organizzata in cluster. Solo sei dei geni identificati erano stati precedentemente descritti da altri autori (Tycowski and Steitz, 2001; Yuan *et al.*, 2003). È interessante notare che di questi, solo uno (chiamato DmSnR61c nel nostro screening, indicato come Dm-797 da Yuan *et al.*, 2003 e depositato in banca dati come Z1), è stato identificato partendo dall'analisi di un libreria di cDNA di *Drosophila* (Yuan *et al.*, 2003), mentre gli altri 5 sono risultati essere membri dei cluster dUgh1 o dUhg2 precedentemente identificati attraverso analisi di omologia di sequenza con i geni UHG1 e UHG2 di mammifero (Tycowski *et al.*, 1996).

Tabella 4: Lista degli snoRNA di *Drosophila melanogaster* identificati in base all'allineamento fra l'rRNA 18 S di lievito e di *Drosophila* e confermati dall'analisi trascrizionale.

putativo sito di metilazione sull'rRNA 18S di <i>Dm</i> (sito di metilazione omologo in <i>S.c.</i>). snoRNA guida di <i>S.c.</i>	Sigla SNOSCAN Nome assegnato	SNOSCAN (score)	snoRNAs presenti nello stesso cluster	Localizzazione genomica Cromosoma, contig (inizio..fine)	Localizzazione genica	Lunghezza stimata dello snoRNA
A 28 (A 28) SnR74/Z4	Dm_18S.009 [@] (SnoRNA:U27:54Eb) DmSnR74/Z4	31.50	Dm_28S.031 [@] Dm_28S.032 [@] Dm_28S.041 [@]	2R, NT_033778 (12766358..12766481)	dUhg1: CG14486	71
A 425 (-)	Dm_18S.011 Dm425	26.92	-	3R, NT_033777 (7048578..7048666)	Introne 3 CG4863-RA; Introne 2 CG4863-RE; Esone 3 CG4863-RD (<i>Ribosomal protein L3</i>)	80
A 1061 (A 973) SnR54	Dm_18S.001 DmSnR54	30.15	-	2R, NT_033778 (9686765..9686892)	Intergenica CG12863/CG10131	72
G 1620 (G 1425) SnR56	Dm_18S.007 [@] (SnoRNA:U25:30E) DmSnR56 a	30.05	-	2L, NT_033779 (9886700..9886777)	dUhg2:CR32873	67

L'assenza del sito di metilazione omologo in *S. cerevisiae* indica che in *D. melanogaster* il putativo sito di metilazione è stato predetto solo in seguito all'identificazione del corrispondente snoRNA guida.

[@]snoRNA identificati anche da Tycowski and Steitz, 2001.

Tabella 5: Lista degli snoRNA di *Drosophila melanogaster* identificati in base all'allineamento fra l'rRNA 28 S di lievito e di *Drosophila* e confermati dall'analisi trascrizionale.

Putativo sito di metilazione sull'rRNA 28S di <i>Dm</i> (sito di metilazione omologo in <i>S.c.</i>). snoRNA guida di <i>S.c.</i>	Sigla SNOSCAN Nome assegnato	SNOSCAN (score)	snoRNAs presenti nello stesso cluster	Localizzazione genomica Cromosoma, contig (inizio..fine)	Localizzazione genica	Lunghezza Stimata dello snoRNA
A 773 (A 647) U18	Dm_28S.002* DmU18 a	31.46	-	3L, NT_037436 (12981769..12981850)	Introne 3 CG11271-RA, RB, RF; Introne 2 CG11271-RC (putative Ribosomal protein S12)	82
	Dm_28S.002a* DmU18 b	-	-	3L, NT_037436 (12981134..12981195)	Introne 2 CG11271-RA, RB, RC, RF; (putative Ribosomal protein S12)	84
C 774 (C 648) U18adj	°⊗Dm_28S.004 DmU18adj	26.05	-	3L,NT_037436 (21544792..21544926)	Esone 4 CG11306 (putative alpha-1,2-mannosyltransferase activity)	135
C 787 (C 661) SnR58	Dm_28S.005* DmSnR58 a	25.20	Dm_28S.043* Dm_28S.043a*	3L, NT_037436 (807898..808016)	Introne 4 CG13900-RB; introne 10 CG13900-RA (putative damaged_DNA binding activity)	119
	Dm_28S.005a* DmSnR58 b	-		3L, NT_037436 (807903..808008)	Introne 4 CG13900-RB; introne 10 CG13900-RA (putative damaged_DNA binding activity)	107
G 3080 (G 2616) SnR67	Dm_28S.041@ (snoRNA:U31:54Ea) DmSnR67	29.38	Dm_28S.031@ Dm_18S.009@ Dm_18S.032@	2R, NT_033778 (12762927..12763009)	dUhg1:CG14486	83

Continuazione Tabella 5

	Dm_28S.026 DmSnR48 a	23.19	-	2R, NT_033778 (19961174..19961258)	Introne 3 CG18506 (function unknown)	85
zG 3254 (G 2790) SnR48	Dm_28S.027 DmSnR48 b	27.93	Dm_28S.008* Dm_28S.008a* Dm_28S.028	2R, NT_033778 (4157505..4157597)	Intergenico CG13741/CG8078	93
	Dm_28S.028 DmSnR48 c	26.96	Dm_28S.008* Dm_28S.008a* Dm_28S.027	2R, NT_033778 (4156738..4156833)	GH14469, putative RNA non coding gene	96
	Dm_28S.046 a* Dm3403 a	20.62	-	4, NC_004353 (779886..780018)	Esone 33 gene <i>bt</i> CG32019	133
	Dm_28S.046 b* Dm3403 b	-	-	4, NC_004353 (772774..772883)	Esone 32 gene <i>bt</i> CG32019	109
C 3403 (-)	Dm_28S.046 c* Dm3403 c	-	-	4, NC_004353 (771067..771181)	Esone 30 gene <i>bt</i> CG32019	114
	Dm_28S.046 d* Dm3403 d	-	-	4, NC_004353 (767748..767852)	Esone 29 gene <i>bt</i> CG32019	104

Continuazione Tabella 5

A 3406 (A 2943) SnR71	Dm_28S.031 [@] (snoRNA:U29:54Eb) DmSnR71 a	29.25	Dm_18S.009 [@] Dm_28S.032 [@] Dm_28S.041 [@] Dm_18S.009 [@] Dm_28S.031 [@] Dm_28S.041 [@]	2R, NT_033778 (12763939..12764038)	dUhg1:CG14486	100
	Dm_28S.032 [@] (snoRNA:U29:54Ed) DmSnR71 b	28.84		2R, NT_033778 (12764695..12764795)	dUhg1:CG14486	101
G 3112 (-)	Dm_28S.043* Dm3112	23.09		3L, NT_037436 (808163..808253)	Introne 2 CG13900-RB ; Introne 8 CG13900-RA (putative damaged_DNA binding activity)	91
	⊗Dm_28S.043a* Dm3112 a	-	Dm_28S.005* Dm_28S.005a*	3L, NT_037436 (808410..808451)	Introne 1 CG13900-RB ; Introne 7 CG13900-RA (putative damaged_DNA binding activity)	91
C 3408 (C 2945) SnR69	°⊗Dm_28S.036 a* DmSnR69 a	27.55	-	3L, NT_037436 (17767021..17767175)	Esone 1 CG14586 (Glutamated-gated-ion channel activity)	155
	°⊗Dm_28S.036 b* DmSnR69 b	-	-	3L, NT_037436 (17764880..17765031)	°Esone 3/Introne 3 CG14586 (Glutamated-gated-ion channel activity)	151

Continuazione Tabella 5

U 2133 (U 1886) SnR62	Dm_28S.018* DmSnR62 a	34.24	-	3R, NT_033777 (1451090..1451158)	°Esone 3 CG 1475, RH27364;introne 2 CG 1475, per SD27659 (putative Ribosomal L13 protein)	69
	Dm_28S.018a* DmSnR62 b	-	-	3R, NT_033777 (1450906..1450971)	Introne 2 CG 1475 (putative Ribosomal L13 protein)	65
A 981 (A 805) SnR39-59	Dm_28S.008* DmSnR39-59 a	33.99	Dm_28S.027 Dm_28S.028	2R, NT_033778 (4157248..4157335)	Intergenico CG13741/CG8078	88
	Dm_28S.008a* DmSnR39-59 b	-	-	2R, NT_033778 (4156990..4157057)	Intergenico CG13741/CG8078	79
C 2644 (C 2335) SnR64	Dm_28S.023* DmSnR64 a	32.17	-	2L, NT_033779 (20617555..20617735)	GH14469, putative RNA non coding gene	181
	Dm_28S.023a* DmSnR64 b	-	-	2L, NT_033779 (20617800..20617970)	GH14469, putative RNA non coding gene	171
	Dm_28S.023b* DmSnR64 c	-	-	2L, NT_033779 (20618031..20618173)	GH14469, putative RNA non coding gene	143

Continuazione Tabella 5

A 1321 (A 797) SnR61	Dm_28S.016 [#] (Dm-797) DmSnR61 c	29.07	-	2R, NT_033778 (6957832..6957916)	Introne 1 <i>Eflalpha48D</i>	80
	Dm_28S.010* DmSnR60 a	38.37	Dm_28S.011* Dm_28S.010/011a* Dm_28S.010/011b*	2R, NT_033778 (19228887..19228978)	Introne 8 <i>mfl/Nop60B:CG3333</i>	92
	Dm_28S.011* DmSnR60 b	36.40	Dm_28S.010* Dm_28S.010/011a* Dm_28S.010/011b*	2R, NT_033778 (19229126..19229217)	Introne 9 <i>mfl/Nop60B:CG3333</i>	92
G 1082 (G 906) SnR60	Dm_28S.010/011a* DmSnR60 c	-	Dm_28S.010* Dm_28S.011 Dm_28S.010/011b*	2R, NT_033778 (19228662..19228736)	Introne 7 <i>mfl/Nop60B:CG3333</i>	92
	Dm_28S.010/011b* DmSnR60 d	-	Dm_28S.010* Dm_28S.011* Dm_28S.010/011b*	2R, NT_033778 (19228415..19228490)	Introne 6 <i>mfl/Nop60B:CG3333</i>	92
A 2527 (A 2218) SnR47	Dm_28S.019 DmSnR47 a	24.53	-	X, NC_004354 (19775793..19775877)	Intergenico <i>CG9570/CG9571</i>	85

Continuazione Tabella 5

G 1322 (-)	°⊗Dm_28S.040 DmG1322	22.81	-	3R, NT_033777 (12139343..12139450)	° Esone 3 gene <i>sulf1</i> CG6725 (N-acetylglucosamine -6-sulfatase-activity; required for normal pattern formation of embryonic cuticole)	108
U 1332 (-)	⊗Dm_28S.039 DmU1332	21.69	-	X, NC_004354 (17818091..17818221)	Esone 1 gene <i>dik</i> CG7098 (component of histone acetyltransferase complex)	131
C 1652 (C1435) U24	⊗Dm_28S.017 DmU24	23.15	-	3L, NT_037436 (3201959..3202060)	Introne 3 CG12740- RA,RB,RD; esone 3 CG12740-RC (putative Ribosomal L28e protein)	102

L'assenza del sito di metilazione omologo in *S. cerevisiae* indica che in *D. melanogaster* il putativo sito di metilazione è stato predetto solo in seguito all'identificazione del corrispondente snoRNA guida. *Tutti gli snoRNAs indicati con lo stesso numero (es: Dm_28S.018, Dm_28S.018a) sono localizzati nello stesso cluster; quando lo score dello SNO SCAN non è presente, le isoforme degli snoRNA sono state identificate tramite analisi di BLAST delle sequenze fiancheggianti il gene originariamente identificato da SNO SCAN ed hauna sequenza altamente correlata.

#snoRNA identificati anche da Yuan et al, 2003.

@snoRNA identificati anche da Tycowski e Steitz, 2001.

⊗snoRNA con sequenze antisense (≥13 nt), localizzate a monte delle box D o D', non perfettamente complementari alla sequenza dell'rRNA.

°Geni trascritti con polarità opposta rispetto al gene ospite.

La scoperta che 4 membri del cluster dUhg1 (U27:54Eb; U29:54Eb; U29:54Ed; U31:54Ea), insieme con un membro del cluster dUhg2 (U25:30E), sono potenzialmente capaci di metilare siti presenti sugli rRNA 18S/28S conservati tra lievito e *Drosophila* offre un chiaro segno della loro origine evolutiva, così come della loro elevata correlazione con la controparte presente nei mammiferi (Tycowski *et al.*, 1996). Pur essendo il nostro screening basato su un'analisi *in silico*, la corrispondenza dei nostri dati con quelli emersi dagli studi molecolari ha confermato la validità del metodo utilizzato per la ricerca ed ha evidenziato la parziale sovrapposibilità dell'approccio bioinformatico con quello molecolare. Al contrario, le precedenti analisi molecolari non hanno portato all'identificazione di alcuno dei geni identificati dal nostro screening, probabilmente a causa della loro scarsa abbondanza o per la presenza di inusuali strutture secondarie o variazioni nucleotidiche che possono aver impedito la progressione della trascrittasi inversa. Un dato interessante è che l'organizzazione dei geni identificati è risultata molto più variegata di quanto precedentemente immaginato (vedi tabella 6). Infatti, la maggior parte dei geni (22/39) (56%) mostra una localizzazione intronica e 6/39 (15%) una localizzazione intergenica. E' interessante notare poi che 8/39 (21%) hanno mostrato un'organizzazione genomica del tutto imprevista, essendo localizzati all'interno di esoni di geni ospiti che codificano per proteine, in 4 casi perfino trascritti con polarità opposta (DmG1322, dmU18adj e DmSnR69a-b). In altri 2/39 casi (5%) i geni sono risultati localizzati in una regione genomica soggetta a splicing alternativo, per cui la loro localizzazione può risultare intronica od esonica, in dipendenza del pattern di splicing considerato. Infine, è stato identificato un solo gene localizzato a cavallo di una giunzione esone/introne (Fig.10).

Tabella 6: Nuovi snoRNA identificati in base all'allineamento lievito/Drosophila la cui espressione è stata confermata dall'analisi trascrizionale.

GENI A LOCALIZZAZIONE INTRONICA			
Nome	Cromosoma	Localizzazione genica	Gene ospite
DmSnR48a	2R	Introne 3	CG18506
DmSnR74/Z4(18S)*§	2R	Introne 14	dUhg1 (CG14486)
DmSnR56(18S)a**§	2L	Introne 2	dUhg2 (CR32873)
DmSnR61c [§]	2R	Introne 1	<i>EflalphaD8</i>
DmSnR67*§	2R	Introne 1	dUhg1(CG14486)
DmSnR71a*§	2R	Introne 3	dUhg1(CG14486)
DmSnR71b*§	2R	Introne 5	dUhg1(CG14486)
DmU18a*	3L	Introne 2	<i>RpS12</i> (CG11271)
DmU18b*	3L	Introne 3	<i>RpS12</i> (CG11271)
DmSnR58a*	3L	Introne 10	CG13900
DmSnR58b*	3L	Introne 10	CG13900
Dm3112a*	3L	Introne 7	CG13900
Dm3112a*	3L	Introne 8	CG13900
DmSnR60a*	2R	Introne 8	<i>Nop60B</i> (CG3333)
DmSnR60b*	2R	Introne 9	<i>Nop60B</i> (CG3333)
DmSnR60c*	2R	Introne 7	<i>Nop60B</i> (CG3333)
DmSnR60d*	2R	Introne 6	<i>Nop60B</i> (CG3333)
DmSnR64a*	2L	Introne 1	GH14469
DmSnR64b*	2L	Introne 2	GH14469
DmSnR64c*	2L	Introne 3	GH14469
DmSnR62a*	3R	Introne 2	<i>pRpL13</i> (CG1475)
DmSnR62b*	3R	Introne 2	<i>pRpL13</i> (CG1475)
GENI A LOCALIZZAZIONE INTERGENICA			
Nome	Cromosoma	Localizzazione genica	
DmSnR54(18S)	2R	CG12863/CG10131	
DmSnR39-59a*	2R	CG13741/CG8078	
DmSnR39-59b*	2R	CG13741/CG8078	
DmSnR47	X	CG9570/CG9571	
DmSnR48b*	2R	CG13741/CG8078	
DmSnR48c*	2R	CG13741/CG8078	
GENI A LOCALIZZAZIONE ESONICA			
Nome	Cromosoma	Localizzazione genica	Gene ospite
DmG1322§	3R	Esone 3	<i>Sulfl</i> (CG6725)
DmU1332	X	Esone 1	<i>dik</i> (CG7098)
Dm3403a*	4	Esone 33	<i>bt</i> (CG32019)
Dm3403b*	4	Esone 32	<i>bt</i> (CG32019)
Dm3403c*	4	Esone 30	<i>bt</i> (CG32019)
Dm3403d*	4	Esone 29	<i>bt</i> (CG32019)
DmSnR69a*	3L	Esone 1	CG14586
DmU18adj	3L	Esone 4	CG11306
GENI LOCALIZZATI IN REGIONI DI SPLICING O A CAVALLO DI UNAGIUNZIONE ESONE/INTRONE			
Nome	Cromosoma	Localizzazione genica	Gene ospite
DmU24 [@]	3L	Esone1-Introne3	<i>Rp28</i> (CG12740)
Dm425 (18S) [@]	3R	Esone3-Introne3/2	<i>RpL3</i> (CG4863)
DmSnR69b [°]	3L	Esone3/Introne3	CG14586

*Geni organizzati in cluster; §Geni identificati anche da Tycowski and Steitz, 2001; §Geni identificati anche da Yuan et al., 2003; @Geni localizzati in una regione soggetta a splicing alternativo; °Gene localizzato a cavallo di una giunzione esone/introne

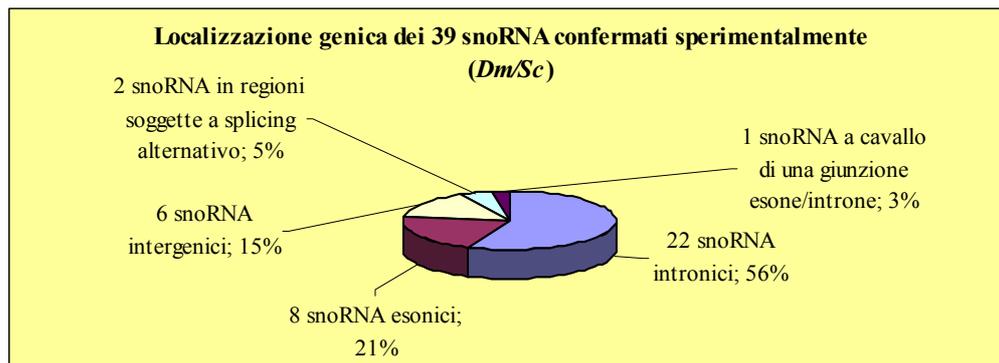


Figura 10

Il mio approccio sperimentale ha portato poi alla scoperta di 8 nuovi cluster (vedi Tabella 7) oltre a quelli identificati da Tycowski (2001); questi sono risultati composti da copie ripetute in tandem altamente correlate di uno stesso gene (i cluster DmSnR39-59/DmSnR48, DmU18, DmSnR60, DmSnR62, DmSnR64, DmSnR69 e Dm3403) o da un insieme di isoforme dello stesso gene e di geni eterologhi (il cluster DmSnR58/Dm3112). Questi dati implicano che la duplicazione dei geni degli snoRNA si è verificata frequentemente durante l'evoluzione del genoma di *Drosophila*.

Tabella 7: Nuovi cluster genici identificati in base all'allineamento lievito/*Drosophila* la cui espressione è stata confermata dall'analisi trascrizionale.

Nome	Chr	N° geni x Cluster	Localizzazione genomica	Gene ospite
DmsnR64*	2R	3	Intronico	GH14469
DmU18	3L	2	Intronico	CG11271
DmSnR58-G3112	3L	4	Intronico	CG13900
DmSnR60	2R	4	Intronico	Nop60B(CG3333)
DmSnR62	3R	2	Esonico/Intronico	CG1475
DmSnR69	3L	2	Esone/Introne	CG14586
DmSnR48-SnR39-59	2R	4	Intergenico	CG13741/CG8078
Dm3403	4	4	Esonico	<i>bt</i> (CG32019)

*Cluster identificato successivamente anche da Huang *et al.* (2005)

In 2 casi, inoltre, l'analisi delle sequenze fiancheggianti ha rivelato la presenza di geni codificanti per snoRNA della classe H/ACA (DmSnR60a-b-c-d; DmSnR64a-b), mentre in un solo caso (DmU18a) sono stati trovati motivi

addizionali caratteristici degli snoRNA che si localizzano all'interno dei Cajal bodies (scaRNA) (Darzacq *et al.*, 2002) suggerendo che il gene DmU18a possa appartenere alla categoria degli scaRNA. Tra gli 8 cluster identificati, 5 sono costituiti da snoRNA a localizzazione intronica, (DmSnR60a-b-c-d; DmSnR64a-b; DmSnR58a-b/DmG1332a-b; DmU18a-b; DmSnR62a-b). È interessante notare che l'esempio dei geni Dm3403 ha invece fornito il primo caso di cluster genico a localizzazione esonica: infatti, i 4 geni che compongono il cluster sono tutti localizzati nello stesso gene ospite in esoni successivi o molto vicini tra loro. Il cluster DmSnR69 comprende uno snoRNA esonico (DmSnR69a) ed uno localizzato a cavallo di una giunzione esone-introne del gene ospite (DmSnR69b), mentre il cluster DmSnR39-59/DmSnR48 ha una localizzazione intergenica. Tutti i geni identificati, tranne due (DmU18 e DmU1332), sono risultati espressi in maniera costitutiva durante lo sviluppo, come previsto per molecole di snoRNA che guidano eventi di modificazione sugli rRNA. Le dimensioni dei trascritti identificati non si sono discostate significativamente da quelle predette dallo SNOSCAN, tranne che per i geni DmU18a-b, per il quale è stata evidenziata una lunghezza di circa 300 nt, una dimensione compatibile con le strutture complesse tipiche di quegli snoRNA appartenenti alla categoria degli scaRNA (Darzacq *et al.*, 2002). Di seguito saranno riportate le descrizioni dettagliate dei geni confermati sperimentalmente, divisi in quattro categorie sulla base della loro localizzazione genica; discuterà preferenzialmente quelli che hanno mostrato caratteristiche più interessanti.

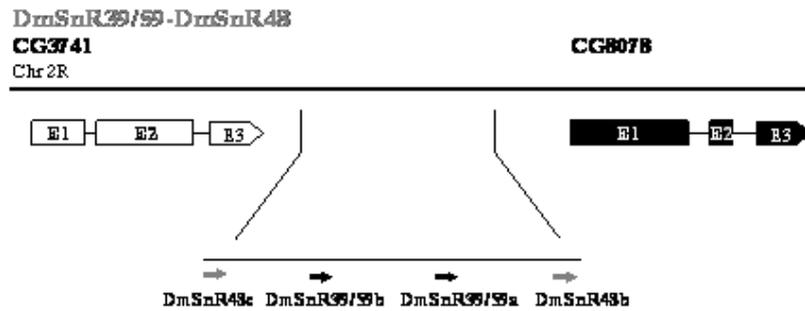
2.4.1.2 Geni a localizzazione intergenica

L'organizzazione genomica dei geni a localizzazione intergenica confermati è schematizzata nella figura 11. I geni a localizzazione intergenica confermati sperimentalmente (vedi tabella 4: DmSnR54; vedi tabella 5: DmSnR39-59a, DmSnR39-59b, DmSnR48c, DmSnR48b, DmSnR47a) hanno presentato tutti un profilo di espressione costitutivo e la maggior parte di essi è risultata organizzata in cluster. L'esistenza in banca dati di una EST (EN01530) che copre l'intera regione del cluster DmSnR39-59/DmSnR48, lascerebbe supporre che si tratti di un tipico cluster dUhg. Tuttavia l'organizzazione dei geni appare peculiare, in quanto due dei quattro snoRNA che lo costituiscono sono localizzati in introni non adiacenti.

Il cluster DmSnR39-59/DmSnR48

Il cluster DmSnR39-59/DmSnR48 è localizzato nella banda politenica 45A2 del braccio destro del cromosoma 2 di *D.melanogaster*, tra i geni CG13741 e CG8078 (Fig.12A), trascritti con polarità opposta. La regione intergenica nella quale è localizzato il cluster è lunga 967 bp e la distanza tra i geni codificanti gli snoRNA è di circa 250 bp. Il cluster è composto da due geni altamente correlati, DmSnR48c e DmSnR48b, separati da due copie dello snoRNA DmSnR39-59 (a-b). La peculiare localizzazione dei due geni DmSnR48c-b potrebbe essere il risultato di un evento di duplicazione seguito da trasposizione in regioni adiacenti. Nei mammiferi sono stati descritti esempi di "C/D snoRNA retrogenes" (Bachellerie *et al.*, 2002), caratterizzati da una coda di poli(A) al 3' della sequenza codificante lo snoRNA e fiancheggiati da una

A



B



C



Figura 11: Organizzazione genomica dei nuovi geni identificati in regioni intergeniche.

A. Il cluster DmSnR9-59/DmSnR48 localizzato nella regione compresa tra i geni CG3741 e CG8078: le frecce grigie indicano le due copie omologhe dei geni DmSnR48b-c, mentre le frecce nere indicano le due copie omologhe dei geni DmSnR58a-b. B. In rosso è indicato il gene DmSnR54, localizzato tra i geni CG12863 e CG10131 indicati rispettivamente in bianco e in nero. C. In arancio è indicato il gene DmSnR47, localizzato tra i geni CG6570 e CG9571. I rettangoli bianchi indicano gli esoni e le linee orizzontali gli introni: le dimensioni non sono in scala.

coppia di sequenze dirette ripetute, lunghe 10-15 nt e generate probabilmente dall'inserzione di una sequenza simile ad un cDNA. E' stato quindi ipotizzato che queste molecole si siano evolute da un gene ancestrale in seguito ad eventi di retrotrasposizione seguiti da una serie di duplicazioni in tandem nei loci di inserzione (Bachellerie *et al.*, 2002). Lo stesso meccanismo non appare ipotizzabile per gli snoRNA DmSnR48c-b, dato che essi non presentano code di poli(A) all'estremità 3' e non sono fiancheggiati da sequenze ripetute simili a quelle descritte per i "retrogenes snoRNA C/D" di mammifero.

In figura 12A è schematizzata l'organizzazione genomica del cluster intergenico DmSnR39-59/DmSnR48 e sono mostrati gli esperimenti di Northern blot che ne confermano l'effettiva espressione nel genoma di *D.melanogaster*. Per tutti e quattro i geni le sequenze guida antisenso si trovano a monte della box D', e l'appaiamento con lo rRNA 28S è di perfetta complementarità. Come si può osservare (Fig.12B), il risultato dell'ibridazione sottolinea il profilo di espressione costitutivo del gene, tipico delle molecole di snoRNA che agiscono nella modificazione dello rRNA. La lunghezza della molecola di snoRNA DmSnR39-59a stabilita attraverso l'approccio bionformatico (88 nt) è stata confermata attraverso elettroforesi su gel di poliacrilammide al 6% e l'estremità 5' del gene è stata determinata tramite un esperimento di Primer extension (Fig.12C).

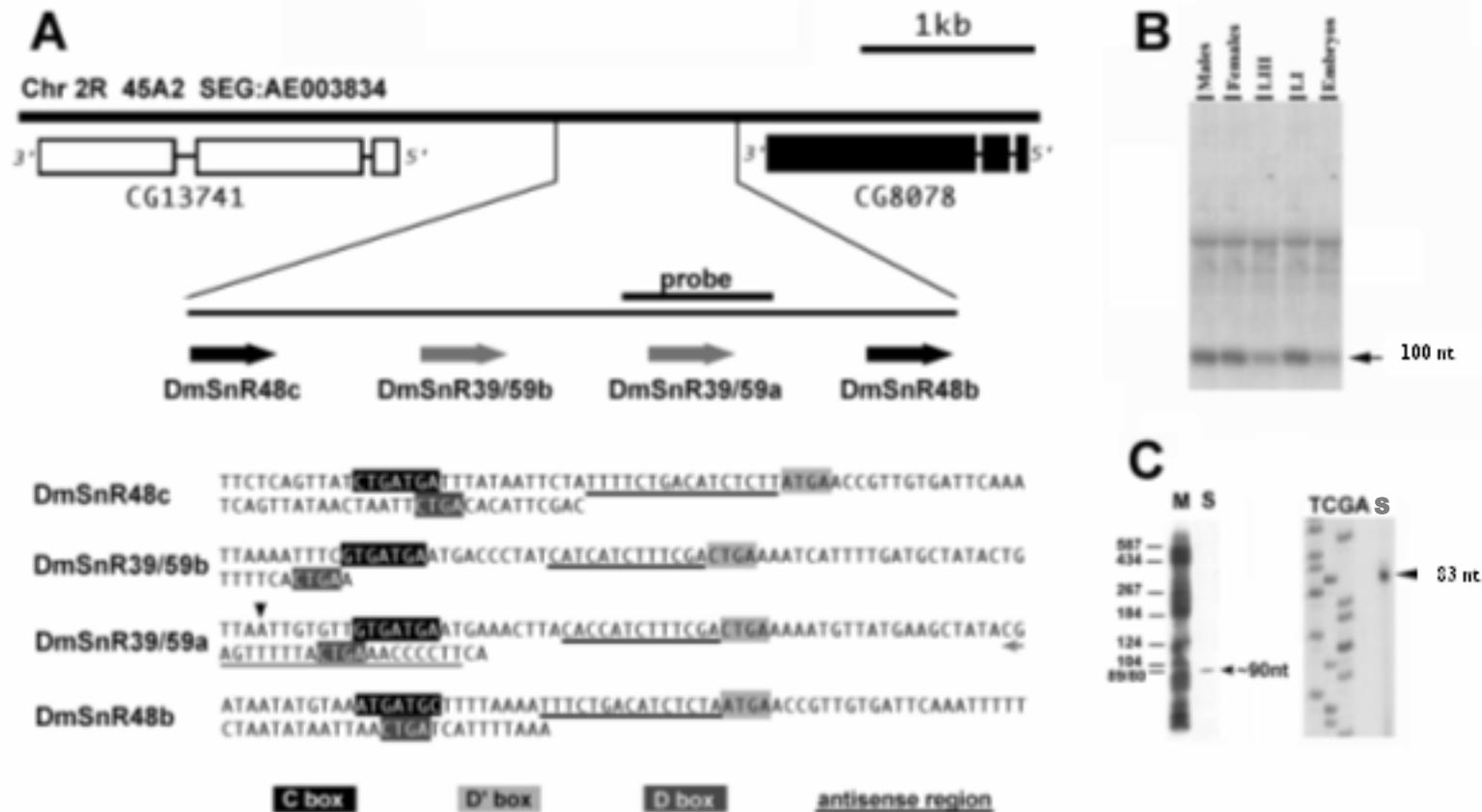


Figura 12: Organizzazione genomica del Cluster DmSnR39-59/DmSnR48

A. In alto, l'organizzazione genomica del cluster Dm SnR48-SnR39-59 nella regione 45A2 del braccio destro del cromosoma 2 di *D. melanogaster*. Il cluster è localizzato nella regione tra i geni CG13741 e CG8078 ed è composto dai geni altamente correlati DmSnR48c e DmSnR48b, separati da due copie degli snoRNA Dm SnR39-59. La sequenza degli snoRNA è riportata in basso; le box C, D e D' sono indicate rispettivamente dai rettangoli nero, grigio e grigio chiaro. La linea grigia all'estremità 3' dello snoRNA DmSnR39-59a rappresenta l'oligonucleotide utilizzato nell'analisi di primer extension e la freccia nera indica l'estremità 5' determinata sperimentalmente. La sequenza antisense complementare all'rRNA 28S è sottolineata. **B.** Northern blot di sviluppo osservato per DmSnR39-59a. 6 µg di RNA totale sono stati caricati in ciascun pozzetto (LI e LIII, primo e terzo instar larvale, rispettivamente). **C.** La dimensione dello snoRNA DmSnR39-59a è stata stimata intorno ai 90 nt mediante elettroforesi su gel di poliacrilammide al 6% (a sinistra) (M: marker di peso molecolare; S: 4µg di RNA totale) e da analisi per primer extension (a destra), utilizzando il primer descritto in A (TCGA: reazione di sequenza usata come marker di peso molecolare; S: campione).

2.4.1.3 Geni a localizzazione intronica

Come già accennato, la maggioranza dei geni ha localizzazione intronica (vedi tabelle 5 e 6). La maggior parte di questi è organizzata in cluster, tranne i due geni DmU24 e DmSnR48a. I geni codificanti per snoRNA che hanno localizzazione intronica sono presenti con una frequenza piuttosto alta negli introni di geni coinvolti nella biogenesi dei ribosomi, o più generalmente nel processo di traduzione genica. La mia ricerca ha infatti identificato 10 geni codificanti per snoRNA localizzati all'interno di introni di geni ospiti con tali caratteristiche (vedi tabella 4: *RpL3*; vedi tabella 5: *RpS12*, *RpL13*, *RpL28e*, *Nop60B*, CG12740, CG1475). Tuttavia, in 15 casi il nuovo gene identificato è risultato localizzato in introni od esoni di un gene ospite la cui funzione è attualmente sconosciuta o non appare affatto correlata con il processo di traduzione genica (vedi tabella 5: CG11306, CG13900, *Eflalpha48D*, CG14586, CG18506, *Sulf1*, *dik*, *bt*). Due cluster hanno mostrato caratteristiche interessanti: il cluster costituito dai geni DmU18a-b, la cui analisi trascrizionale ha rivelato la presenza di un trascritto le cui dimensioni (300 nt) suggerivano l'appartenenza alla classe degli scaRNA, ed il cluster costituito dai geni DmSnR62a-b, che ha mostrato un'organizzazione tipica dei cluster intronici finora identificati solo nelle piante (Qu *et al.* 2001; Liang *et al.* 2002; Chen *et al.* 2003) ed in *Drosophila* solo per gli snoRNA di classe H/ACA (Huang *et al.*, 2004). L'organizzazione genomica dei geni a localizzazione intronica la cui espressione è stata confermata dall'analisi trascrizionale è schematizzata nella figura 13.

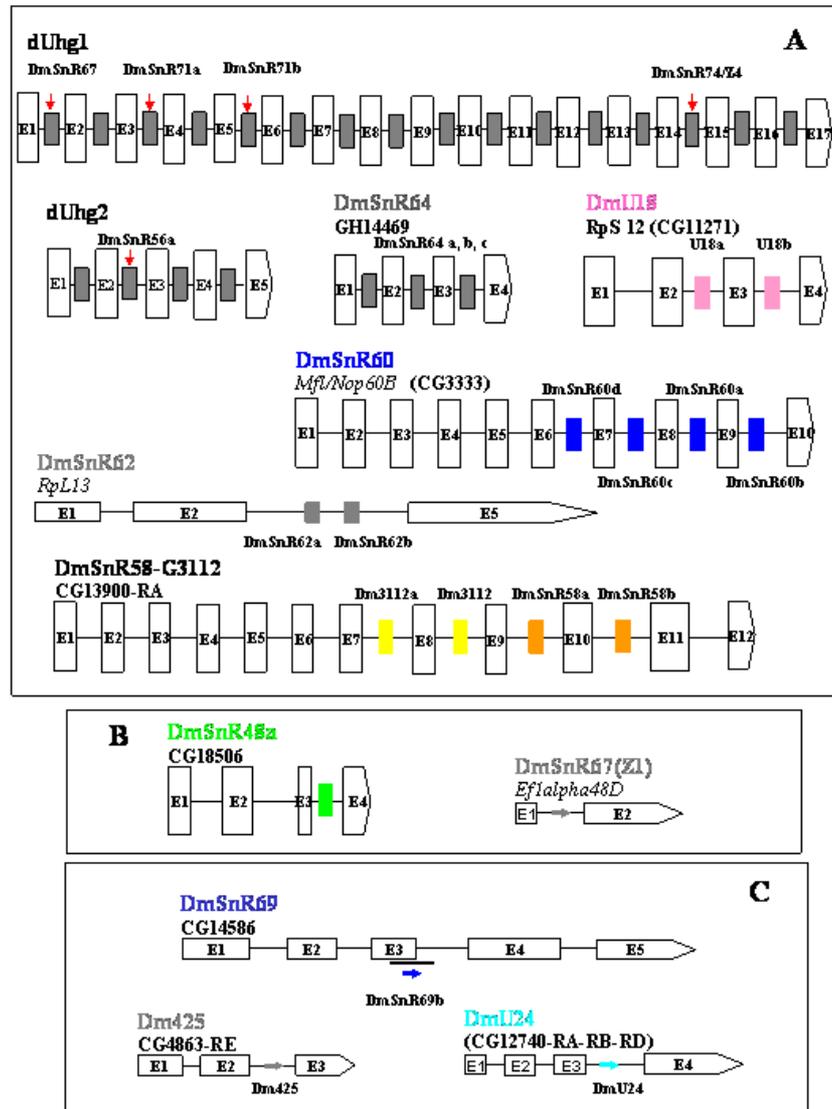
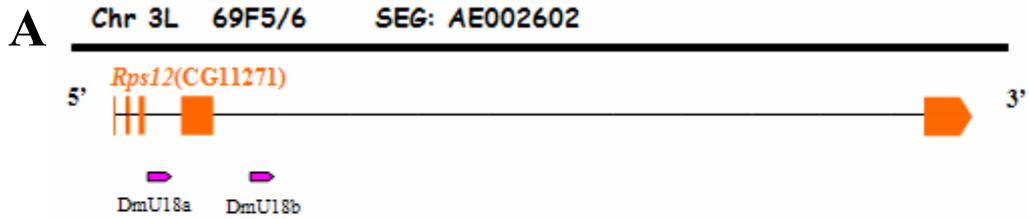


Figura 13: Organizzazione genomica dei nuovi geni identificati in regioni introniche.

A. Cluster dUhg1 e dUhg2: le frecce verticali indicano gli snoRNA già identificati da Tycowski and Steitz, 2001. Cluster DmSnR64: in grigio sono indicati i geni DmSnR64a-b-c, localizzati rispettivamente negli introni 1, 2 e 3 del gene GH14469; cluster DmU18: in rosa sono indicati i geni DmU18a e DmU18b localizzati rispettivamente nell'introne 2 e 3 del gene CG11271(*Rps12*); cluster DmSnR60, in blu sono indicati i geni DmSnR60a-b-c-d localizzati rispettivamente negli introni 6, 7, 8 e 9 del gene *Nop60B* (CG3333); cluster DmSnR62, in grigio sono rappresentati i geni DmSnR62a-b, localizzati nell'introne 2 del gene *RpL13*; cluster DmSnR58/DmG3112 costituito da copie omologhe del gene DmSnR58 indicate in arancio e dai geni DmG3112a-b indicati in giallo, localizzati rispettivamente negli introni 7, 8 e negli introni 9 e 10 del gene CG13900. **B.** Schematizzazione dei geni intronici non organizzati in cluster: in verde è rappresentato il gene DmSnR48a localizzato nell'introne 3 del gene CG18506; in grigio è rappresentato il gene DmSnR61c, identificato anche da Yuan *et al.* 2003. **C.** Organizzazione dei geni intronici localizzati in regioni sottoposte a splicing alternativo o cavallo di una giunzione esone/introne: in blu è indicato il gene DmSnR69b localizzato a cavallo della giunzione tra l'esone 3 e l'introne 3 del gene CG14586; in grigio è rappresentato il gene Dm425, localizzato nel secondo introne del trascritto alternativo RE del gene CG4863; in celeste il gene DmU24 localizzato nel terzo introne dei trascritti alternativi RA-RB-RD del gene CG12740. I rettangoli bianchi indicano gli esoni e le linee orizzontali gli introni, le dimensioni non sono in scala.

Il cluster DmU18

In figura 14A è schematizzata l'organizzazione genomica del cluster DmU18, i cui geni sono localizzati rispettivamente negli introni 3 e 2 del gene (*RpS12*) *Ribosomal protein S12*; (CG11271). Il gene CG11271 è localizzato nella banda poligenica 69F5-6 del braccio sinistro del cromosoma 3 di *D.melanogaster* ed è costituito da 4 esoni e 3 introni. Esso dà origine a due trascritti, CG11271-RA e CG11271-RB, lunghi rispettivamente 625 nt e 718 nt, che differiscono a livello del quarto esone e codificano per due proteine (S12E e L7AE) che hanno in comune la porzione N-terminale. Le sequenze antisense sono localizzate per entrambi gli snoRNA a monte della box D' ed entrambe presentano un appaiamento perfetto con lo rRNA 28S. Nell'esperimento di Northern blot mostrato in figura 14C l'ibridazione è stata effettuata con una sonda genomica di circa 450 bp comprendente la sequenza dell' isoforma DmU18a; i risultati non solo hanno confermato l'effettiva espressione del gene ma hanno rivelato la presenza di due tipi di trascritti, uno di circa 100 nt presente nelle larve di III instar sia precoci e che tardive, nelle pupe e negli adulti, ed un secondo trascritto, lungo circa 300 nt, presente solo nella larve di II stadio e di III stadio precoce (Fig. 14B). Il primo segnale corrisponde in effetti alla lunghezza dello snoRNA determinata dall'analisi bioinformatica, mentre la lunghezza del secondo trascritto è molto differente da quella stimata dal programma SNOSCAN. Un' analisi delle sequenze fiancheggianti, volta a mappare più in dettaglio il gene DmU18a, mi ha portato ad evidenziare ulteriori motivi di sequenza tipici degli scaRNA di classe H/ACA-C/D, cioè le molecole di RNA che si localizzano nei Cajal Bodies (CB; Kiss *et al.* 2002).



DmU18a: TGGTCTGACTGTGATGAAATTAATTGGTCCGTGTTTATGAATTTAT
ATTGATTA AAAACCTTCCAATTACTGAGCAAATGAC

DmU18b: TGATGAAATTAATTGGTCCGTGTTTATGAATTTCCATTGATTATAA
ATTCTCCAATTACTGACCAAGCAAGGCTAAGTCGAGTT

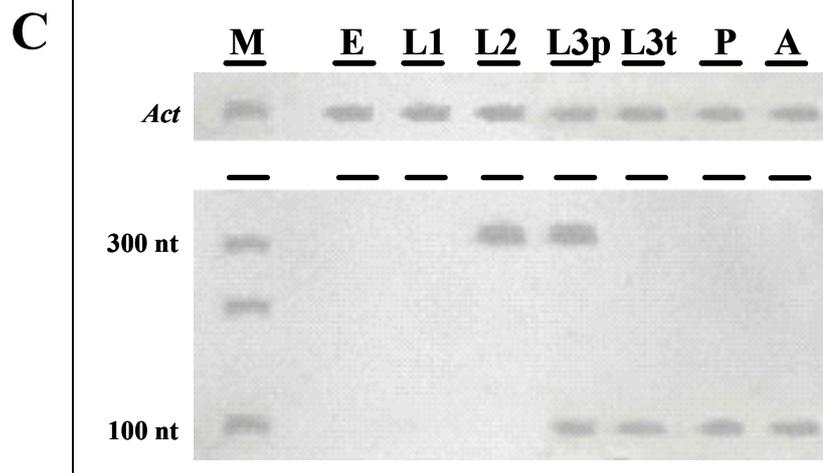
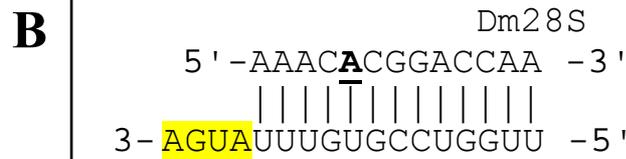
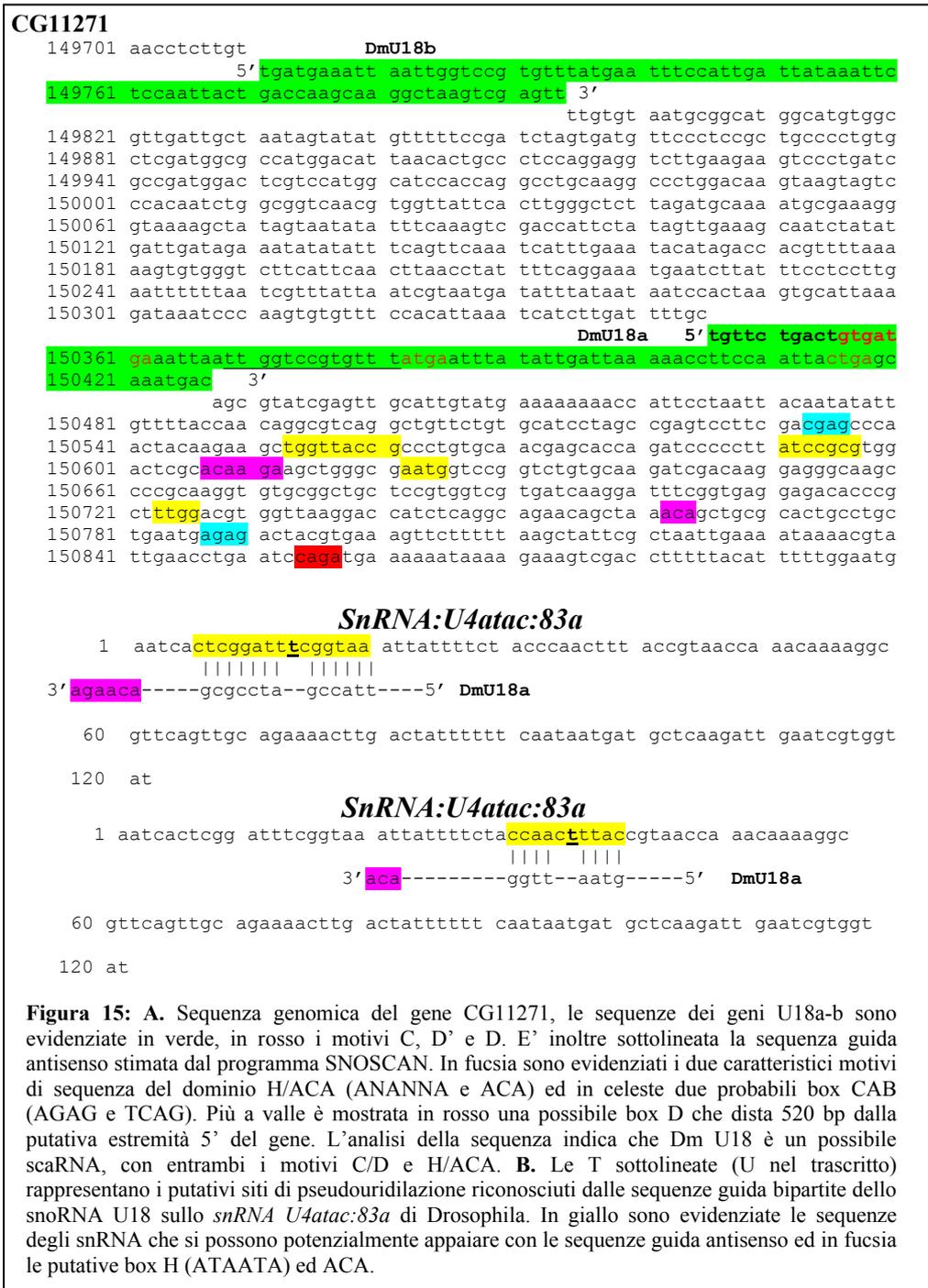


Figura 14: Identificazione del Cluster DmU18

A. I geni DmU18a e DmU18b indicati in rosa, sono localizzati nella regione 69F5/6 del braccio sinistro del cromosoma 3. Gli snoRNA si trovano rispettivamente negli introni 2 e 3 del gene CG11271. Sequenze dei geni DmU18a-b: in rosso sono indicate le box C, D e D'. La regione antisense complementare all' rRNA è sottolineata. **B.** E' mostrato l' appaiamento perfetto tra la sequenza guida antisense dello snoRNA U18a (uguale a quella dello snoRNA U18b) e lo rRNA 28S. L'adenina sottolineata e in grassetto rappresenta la putativa base metilata. **C.** Analisi trascrizionale dello snoRNA Dm U18a: in alto è riportato l' esperimento di normalizzazione con Actina; in basso il trascritto da 300 nt presente solo nelle larve di II e III stadio precoce (LII e LIIIp) quello da 100 nt corrispondente allo snoRNA è assente negli embrioni (E), nelle larve di I e II stadio (LI e LII). M: marker di dimensioni note (RNA ladder low range-Fermentas).

Come mostrato in figura 15, a valle della sequenza genica di DmU18a è presente un motivo H (ANANNA) e, più a valle, un motivo ACA. Sono presenti inoltre i due motivi CAB (AGAG) e (TCAG) necessari per la ritenzione degli scaRNA nei CB. Infine è presente una possibile box D non canonica (CAGA) a valle delle box H/ACA che potrebbe rappresentare l'estremità 3' della molecola di scaRNA (Fig.15A). Poichè è noto che gli scaRNA mediano le reazioni di modificazione sugli small nuclear RNA (U1, U2, U3, U5 e U12 i principali), per avvalorare l'ipotesi che il DmU18a appartenesse alla categoria degli scaRNA ho ricercato possibili target alternativi avvalendomi della banca dati Noncode (<http://noncode.bioinfo.org.cn>). Tale ricerca mi ha portato ad individuare due possibili siti di pseudouridilazione sullo snRNA:U4atac:83a, la cui complementarità con la sequenza antisenso bipartita del dominio H/ACA, lunga 13 nt include un solo mismatch (Fig.15B), caratteristica questa considerata compatibile con il riconoscimento della molecola di RNA bersaglio (Darzacq *et al.*, 2002). L'insieme di queste osservazioni indicherebbe dunque che DmU18a potrebbe rappresentare il secondo esempio di snoRNA di tipo C/D-H/ACA finora identificato in *Drosophila* (Richard *et al.*, 2003). Un altro dato significativo emerso dall'analisi trascrizionale dello snoRNA U18a è che esso ha mostrato un profilo di espressione regolato, essendo assente nei primi stadi di sviluppo (Embrioni, Larve di I e II instar; Fig. 14C). Questa situazione è abbastanza insolita, soprattutto se si considera che esso ha come bersaglio una molecola di rRNA che è ovviamente espressa in maniera costitutiva; in più il gene è localizzato all'interno di un introne di un ospite che codifica per una

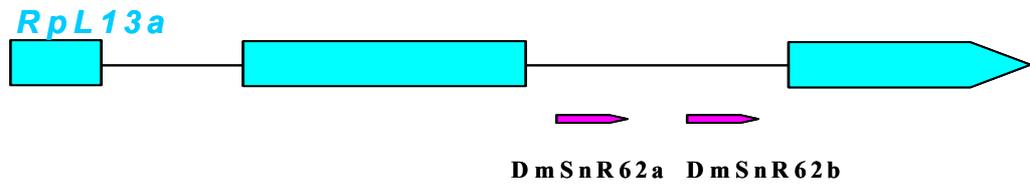
proteina ribosomale, anch'esso ovviamente espresso in maniera costitutiva. Un profilo regolato durante lo sviluppo è stato osservato per pochi altri snoRNA in *Drosophila* (Yuan *et al.*, 2003) e sarà quindi interessante definire più in dettaglio le funzioni e la regolazione dell'espressione del gene DmU18a.



Il cluster DmSnR62

Il cluster DmSnR62 è costituito da 2 copie ripetute in tandem del gene DmSnR62 (a-b) entrambe localizzate nell'introne 2 del gene *RpL13a* (Fig.16A). La presenza di più copie ripetute in tandem di uno stesso gene codificante per snoRNA all'interno di un singolo introne, caratteristica dei cluster intronici identificati nelle piante, era stata riscontrata finora in *Drosophila* solo per gli snoRNA appartenenti alla classe H/ACA (Huang *et al.*, 2004). Gli snoRNA che presentano questo tipo di organizzazione vengono processati attraverso un meccanismo che prevede tagli sia endonucleolitici che esonucleolitici, sostanzialmente differente da quello che agisce nella maturazione degli snoRNA intronici singoli. Il gene ospite *RpL13a* è situato nella banda politenica 83B6-B7 del braccio destro del cromosoma 3 di *Drosophila melanogaster* e produce un unico trascritto (CG1475-RB) lungo 742 nt che codifica per la proteina ribosomale *RpL13a*. I geni DmSnR62a-b presentano la stessa sequenza antisenso localizzata a monte della box D', che però è lunga 15 nt per DmSnR62a e 14 nt per DmSnR62b; l'appaiamento con l'rRNA 28S è in entrambi i casi di perfetta complementarietà (Fig.16B). L'esperimento di Northern blot condotto utilizzando una sonda genomica di 338 bp ha mostrato che lo snoRNA DmSnR62a è espresso costitutivamente durante lo sviluppo; il segnale di ibridazione evidenzia un trascritto di circa 80 nt che è in accordo con le dimensioni stimate dallo SNOSCAN.

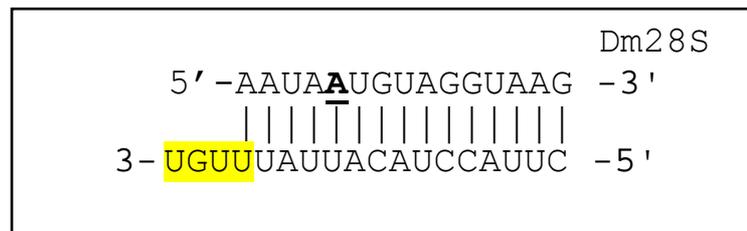
A Chr 3R 83B6/B7



DmSnR62a: ACATGATGAACATTACATTCTTACCTACATTATTTGA
 ACTAGAAACCTGTAACCAACTCTCAACTGATG

DmSnR62b: CATGATGTTTTCAAACCTCTATTACCTACATTATTTGAA
 CGAGAAACCGTGTAACAACAAACATAACTGA

B



C

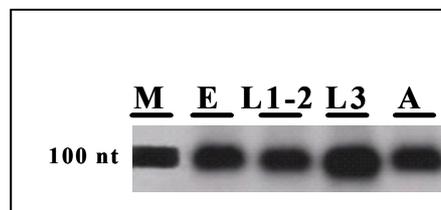
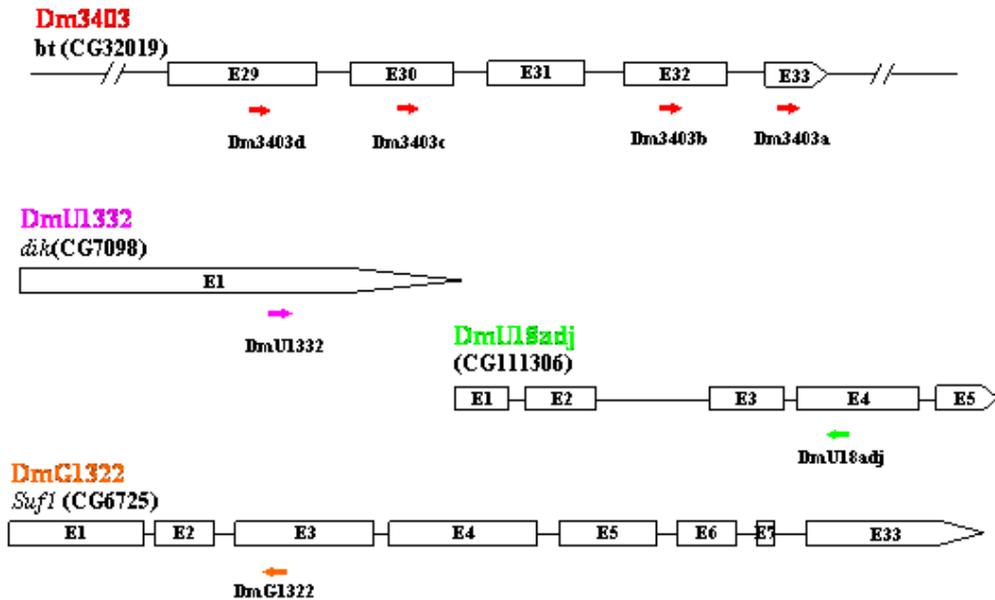
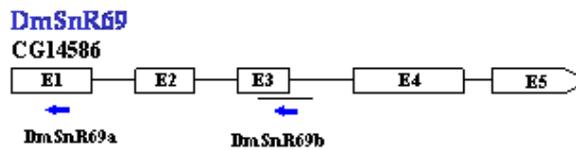
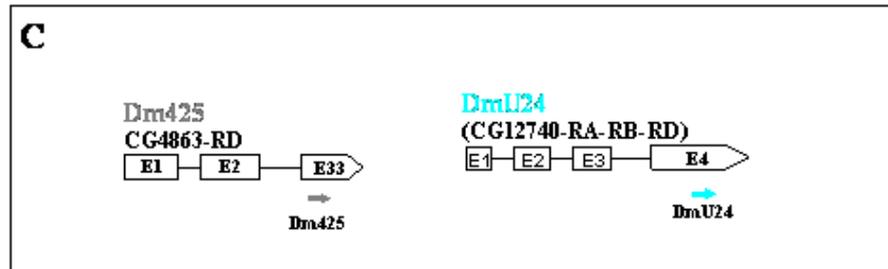


Figura 16: Identificazione del cluster DmSnR62

A. Il cluster DmSnR62 è localizzato nella regione 83B6-B7 del braccio destro del cromosoma 3 nel secondo introne del gene *RpL13a*; in figura è evidenziata la localizzazione dei due geni DmSnR62a e DmSnR62b all'interno dello stesso introne. Sequenza dei due geni costituenti il cluster Dm SnR62 a e DmSnR62 b: in rosso sono indicate le box C, D' e D e la regione antisense complementare all' rRNA è sottolineata. B. Appaiamento della sequenza antisense dello snoRNA, lunga 15 nt e l' rRNA 28S, da notare la presenza di complementarietà perfetta, la putativa A metilata è indicata in grassetto e sottolineata. C. Analisi trascrizionale del gene DmSnR62a: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 100 nt.

2.4.1.4 Geni a localizzazione esonica

Uno dei risultati più sorprendenti ottenuti nel corso del nostro screening è stata la scoperta di 8 geni a localizzazione esonica (vedi tabella 5: DmU18adj, DmG1322, DmU1332, Dm3403a-b-c-d e DmSnR69a), di cui 4 trascritti con polarità opposta rispetto al gene ospite (DmG1322, DmU18adj e DmSnR69a-b). 5 di questi geni sono organizzati in cluster (i geni Dm3403a-b-c-d ed il gene DmSnR69a-b). Il cluster DmSnR69 è costituito dal gene DmSnR69a e dalla sua isoforma DmSnR69b, che risulta essere localizzata a cavallo di una giunzione esone/introne del gene ospite. La localizzazione all'interno di esoni non era mai stata osservata finora in nessun organismo per gli snoRNA. E' stata evidenziata da studi condotti per alcuni snoRNA intronici di lievito e vertebrati (Caffarelli *et al.*, 1996; Lee *et al.*, 2003; de Turris *et al* 2004) l'esistenza di un meccanismo di biogenesi indipendente dallo splicing che rappresenta una via minoritaria di biosintesi degli snoRNA intronici. Per gli snoRNA di *Drosophila* localizzati all'interno di esoni, questo tipo di biogenesi dovrebbe invece rappresentare non una via minoritaria, bensì l'unico meccanismo di sintesi. L'organizzazione genomica dei geni a localizzazione esonica confermati è schematizzata nella figura 17.

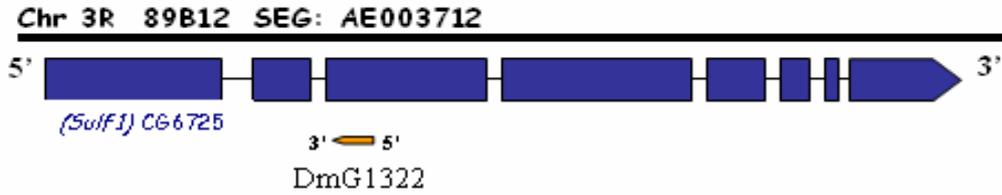
A**B****C****Figura 17: Organizzazione dei nuovi geni identificati in regioni esoniche.**

A. Il Cluster esonico Dm3403: in rosso sono rappresentati i geni Dm3403a-b-c-d localizzati rispettivamente negli esoni 33, 32, 29 e 30 del gene *bt*; in rosa è indicato il gene DmU1332 localizzato all'interno dell'unico esone del gene *dik*; in verde è indicato il gene DmU18adj, localizzato nel quarto esone del gene CG11306 e in arancio è rappresentato il gene DmG1322, localizzato nel terzo esone del gene CG6725 e orientato con polarità opposta al gene ospite. **B.** In blu è indicato il gene DmSnR69a, localizzato nel primo esone del gene CG14586 e in cluster col gene DmSnR69b. **C.** Geni esonici localizzati in regioni sottoposte a splicing alternativo o cavallo di una giunzione esone/introne: in grigio è rappresentato il gene Dm425 localizzato nel secondo introne del trascritto alternativo RE del gene CG4863; in celeste il gene DmU24 localizzato nel terzo introne dei trascritti alternativi RA-RB-RD del gene CG12740. I rettangoli bianchi indicano gli esoni e le linee orizzontali gli introni, le dimensioni non sono in scala con quelle reali.

Lo snoRNA DmG1322

Il gene DmG1322 ha mostrato un tipo di organizzazione particolare: esso è situato nel terzo esone del gene *Sulfated (Sulf1)* ed è trascritto con polarità opposta al gene ospite (Fig.18). La molecola dello snoRNA DmG1322 risulta quindi perfettamente complementare all' mRNA (o al trascritto primario) del gene ospite per tutta la sua lunghezza, e non soltanto per il breve tratto dell'elemento antisenso (10-21 nt), come avviene nella maggior parte dei casi. Questo dato suggerisce la possibilità che questo snoRNA ed il trascritto del gene ospite possano appaiarsi formando molecole di RNA a doppia elica, lunghe abbastanza da essere in grado di innescare il processo di degradazione noto come "RNA interference". Il gene *Sulf1* è localizzato nella banda politenica 89B12 sul braccio destro del cromosoma 3 ed è costituito da 8 esoni e 7 introni; è un gene indispensabile per la formazione del normale sviluppo della cuticola embrionale e codifica per una proteina, nota come N-acetilglucosamina-6-solfatasi (G6S) che appartiene alla famiglia delle solfatasi extracellulari. Per il gene omologo umano, *HSulf-1*, è stato recentemente osservato un coinvolgimento nella regolazione della crescita cellulare; l'assenza di *HSulf-1* nelle linee cellulari di carcinoma squamoso del collo e della testa potenzia l'azione del fattore di crescita "heparin-binding", determina un aumento della motilità e dell'invasività delle cellule tumorali e inibisce l'apoptosi indotta da stress con un conseguente aumento della tumorigenicità (Lai *et al.*, 2004). La sequenza guida antisenso dello snoRNA DmG1322, situata a monte della box D', presenta perfetta complementarietà per 9 nt con lo rRNA 28S, e continua in maniera imperfetta per un ulteriore tratto di 12 nt.

A



DmG1322: CGGA~~ACTCTCA~~ATGAGGAAGCTATCTGGCCAGTTGTCTCGCACAGCG
CGATTTCTGCTCAGAAGCAGTGGCAGTATGCTGCGCCCATCCATATGCTGAGGC
GTTGGCA

B



C

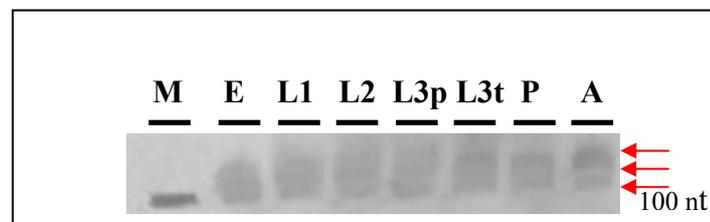


Figura 18: Identificazione del gene Dm G1322

A. Il gene DmG1322 è localizzato nella regione 89B12 del braccio destro del cromosoma 3, all'interno del terzo esone del gene CG6725 (*Sulf1*), ed è trascritto con polarità opposta al gene ospite. E' riportata la sequenza dello snoRNA Dm G1322 e in rosso sono indicate le box C, D' e D e la regione antisense complementare all' rRNA è sottolineata. B. Appaiamento della sequenza antisense dello snoRNA, lunga 18 nt e l' rRNA 28S, da notare la complementarietà perfetta per i primi 9 nt a valle della box D', la putativa G metilata è indicata in grassetto e sottolineata. C. Analisi trascrizionale del gene DmG1322: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M, marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 100 nt. Le frecce rosse evidenziano il processo di trimming esonucleolitico. M: marker di peso molecolare noto (RNA ladder, low-range-Fermentas).

L'analisi del profilo di espressione dello snoRNA DmG1322, ospitato dal gene *Sulfl*, ha mostrato che esso è espresso in tutti gli stadi di sviluppo; il segnale di ibridazione indica la presenza di un trascritto di 100 nt, in accordo con la previsione dello SNOSCAN sulle dimensioni del gene in esame. Tuttavia, come si può osservare nella figura 18C, l'esperimento di Northern blot evidenzia una serie di bande che probabilmente riflettono il processo di maturazione mediante "trimming" esonucleolitico alle estremità di questo snoRNA DmG1322.

Lo snoRNA DmU1332

La snoRNA DmU1332 è localizzato nel gene ospite *Diskette (Dik)*, situato nella banda politenica 16F6 del cromosoma X di *Drosophila melanogaster*. In figura 19 è schematizzata la struttura del gene ospite, la localizzazione dello snoRNA e la sua sequenza. Il gene *Dik*, indicato in banca dati anche con la denominazione *ADA3*, non presenta introni e codifica un prodotto proteico con un putativa attività di acetiltrasferasi degli istoni, che rappresenta un componente del complesso attivatore della trascrizione Ada2/Gcn5/Ada3 (Aguilera *et al.*, 1999; Muratoglu *et al.*, 2003; Kusch *et al.*, 2003). La lunghezza della molecola dello snoRNA DmU1332 determinata sperimentalmente è di circa 130 nt, in accordo con i dati dell'analisi computazionale. La sequenza guida antisenso, localizzata a monte della box D', presenta perfetta complementarietà con lo rRNA 28S per 8 nt, ed imperfetta per i successivi 6 nt. I risultati emersi dall'analisi trascrizionale sono abbastanza insoliti: infatti, dall'ibridazione con una sonda genomica di 507 bp comprendente il gene in esame è emerso che il gene DmU1332 non è espresso nello stadio embrionale e

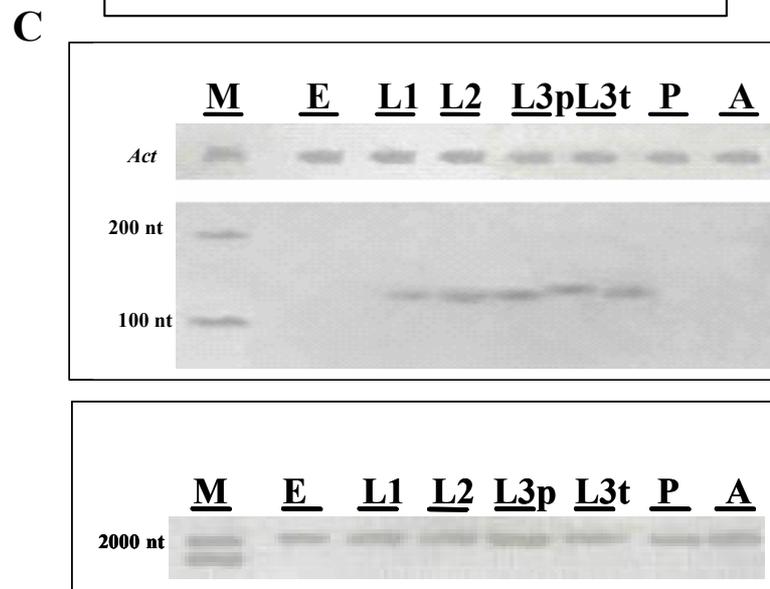
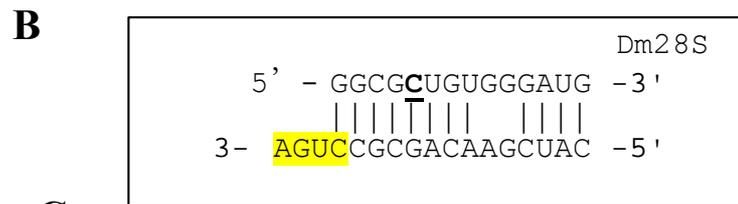
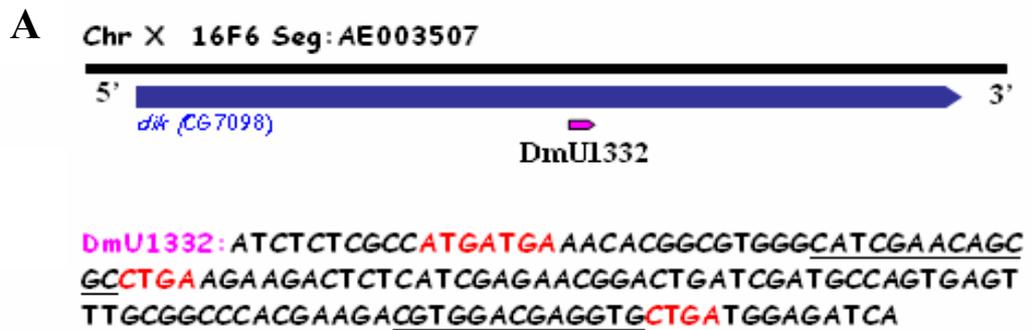


Figura 19: Identificazione del gene Dm U1332

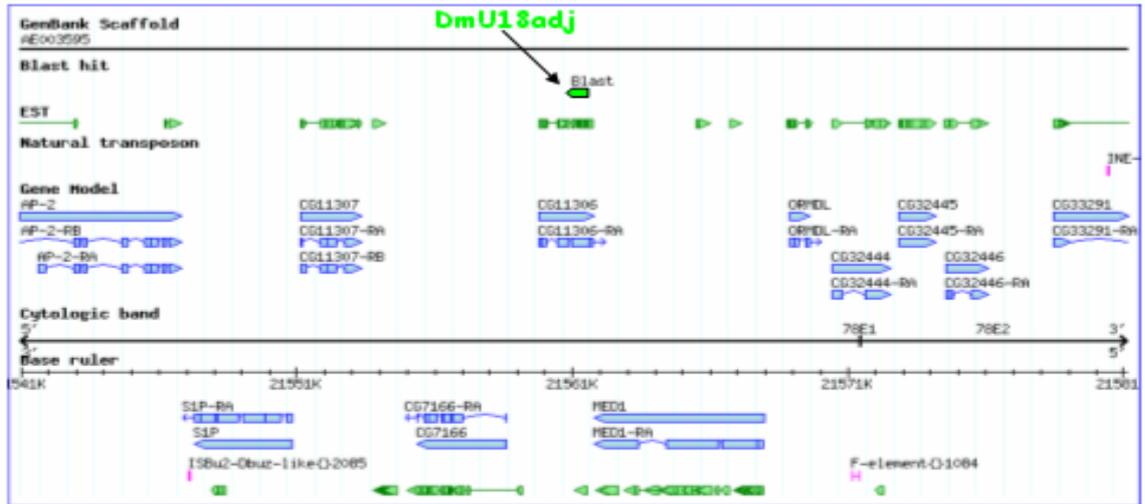
A. Il gene DmU1332 è localizzato nella regione 16F6 del braccio destro del cromosoma X all'interno del gene *dik*, che è ininterrotto. Sequenza del gene Dm U1332: in rosso sono indicate le box C, D' e D ed è sottolineata la regione antisense complementare all' rRNA. **B.** Appaiamento tra la sequenza guida antisense e l'rRNA, la putativa citosina metilata è indicata in grassetto e sottolineata. **C.** Analisi trascrizionale del gene DmU1332, nel pannello in alto è riportato l'esperimento di normalizzazione con Actina ed è mostrato il trascritto di circa 130 nt corrispondente allo snoRNA che è assente negli embrioni (E) e nelle larve di 1° stadio (I); nel pannello in basso è mostrato l'esperimento di Northern blot condotto sulla frazione di poly(A)⁺ che mostra l'espressione costitutiva del gene ospite *dik* (E: embrioni; L1: larve di 1° stadio; LIIIp: larve di ° stadio precoce; LIII: larve di 3° stadio tardivo; P: pupe; A: adulti), lungo circa 2000 nt. M: marker di peso molecolare (RNA ladder, low range-Fermentas).

nel primo stadio larvale, un profilo di espressione differente da quello del gene ospite che è invece espresso in maniera costitutiva. L' esperimento di Northern blot condotto sulla frazione di RNA poly(A)⁺ ha difatti rivelato un segnale di ibridazione di circa 2000 nt che corrisponde al trascritto del gene ospite, presente in tutti gli stadi di sviluppo, confermando così l'ipotesi dell'esistenza di un profilo di espressione differente tra il gene ospite, e lo snoRNA (Fig.18C). Questo dato ha avvalorato ulteriormente l'ipotesi di un meccanismo di biogenesi splicing indipendente dello snoRNA, già ipotizzato per i geni a localizzazione esonica, ma ancora più evidente in questo caso, essendo il gene ospite privo di introni.

Lo snoRNA DmU18adj

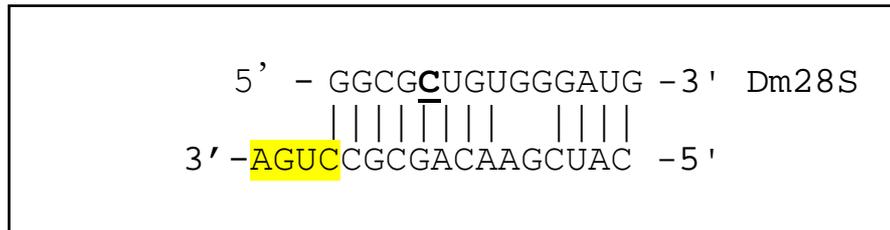
Un altro gene a localizzazione esonica è il gene DmU18adj, localizzato nel quarto esone del gene CG11306 (Fig.20A) e trascritto anch'esso con polarità opposta rispetto al gene ospite. Il gene CG11306 mappa nella banda politenica 78D6, sul braccio sinistro del cromosoma 3, e codifica per una proteina con una putativa attività alfa-1,2-mannosiltrasferasica (<http://www.flybase.org>). Anche per lo snoRNA DmU18adj la sequenza guida antisenso presenta complementarità perfetta con l'rRNA 28S per 8 nt e prosegue con un appaiamento imperfetto per altri 5 nt. La ricerca di un possibile target tra le sequenze degli snRNA, non ha avuto successo e quindi anche per questo gene è plausibile considerare l'rRNA target effettivo. Lo snoRNA presenta un profilo di espressione costitutivo, come confermato dall'esperimento di Northern blot mostrato nella figura 20B, nel quale è stato osservato un segnale di ibridazione in tutti gli stadi di sviluppo esaminati.

A



DmU18adj: CGACAGACAGTATGATGAATTCGTCGCCTTTTTCCGTGTGCTG
CTGAAGGCTCTTTAGGTGACTAACCTCGCACGGAGGGTAAACCCGGTG
 TGTTTTGAAGGGCACATCCACAGTTGCAGAAT**ATGA**ATTCGGTCC

B



C

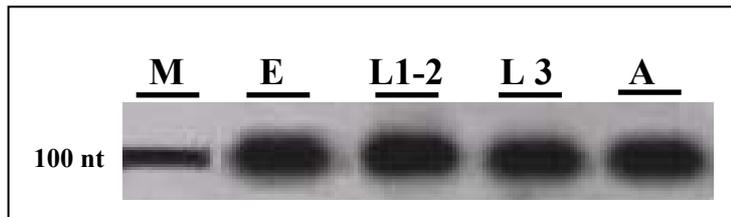


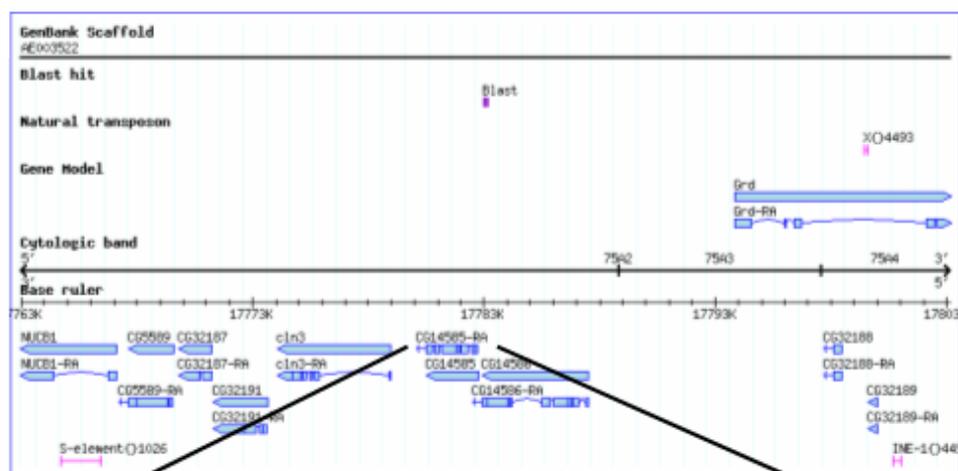
Figura 20: Identificazione del gene Dm U18adj

A. Il gene DmU18adj, indicato dalla freccia in alto, è localizzato nella regione 78D6 del braccio sinistro del cromosoma 3, nel quarto esone del gene CG11306 ed è orientato con polarità opposta. Sequenza del gene Dm U18adj: in rosso le box C, D' e D ed è sottolineata la sequenza antisense complementare all' rRNA. B. Appaiamento della sequenza antisense con l'rRNA 28S. C. Analisi trascrizionale del gene DmU18 adj: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 100 nt.

Il cluster DmSnR69

Nella figura 21A è rappresentata l'organizzazione genomica del gene DmSnR69a e della sua isoforma DmSnR69b, individuata mediante analisi diretta delle sequenze fiancheggianti. L'organizzazione genomica di queste due isoforme è alquanto peculiare, in quanto il gene DmSnR69a è localizzato all'interno del primo esone del gene CG14586, mentre la sua isoforma è localizzata a cavallo di una giunzione esone/introne, precisamente tra l'esone 3 e l'introne 3 del gene ospite, entrambi sono trascritti con polarità opposta rispetto al gene ospite. Il gene CG14586 è localizzato nella banda politenica 75A2 sul braccio sinistro del cromosoma 3 e codifica per un putativo recettore del canale ionico del glutammato. In figura 21 è indicata anche la sequenza genomica del gene DmSnR69a, così come è stata identificata dal programma SNOSCAN e quella della sua isoforma DmSnR69b. La lunghezza della molecola di snoRNA DmSnR69a determinata computazionalmente è di 155 nt, mentre quella dell'isoforma DmSnR69b è rispettivamente di 151 nt. Entrambi i geni presentano una sola box D; lo snoRNA DmSnR69a presenta una sequenza antisenso che è perfettamente complementare all' rRNA 28S per soli 8 nt, seguiti da un tratto di imperfetto appaiamento di 3 nt. La sequenza guida antisenso dello snoRNA DmSnR69b è invece variante (Fig.21C) e non è affatto complementare all'rRNA. La ricerca all'interno della banca dati Noncode (<http://noncode.bioinfo.org.cn>) non ha però identificato alcun possibile target tra gli snRNA per nessuna delle due isoforme, per cui, mentre per lo snoRNA DmSnR69a si può considerare l'ipotesi che l'rRNA sia effettivamente bersaglio della metilazione, lo snoRNA DmSnR69b è per il momento da considerarsi "orfano".

A

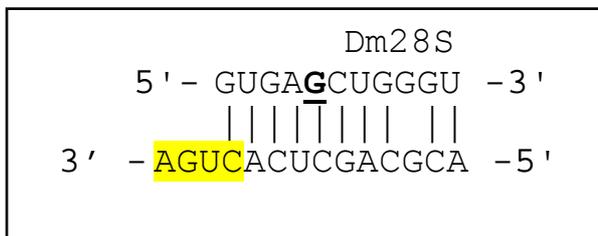


DmSnR69b
DmSnR69a

DmSnR69a:
TCCCAATAGGATGATGAACAGCGGAGCAGCGTACTCCAGACC CACGGCCAC CGTGTGA
TTATGCTGGTAGCAGTGGAGTTTGTCTCCATCCAAGACCGCTCGTCTGAAATGAAC TCCTGTCT
CCAGCATACCGAGCTCACTGAAATTGGGTT

DmSnR69b:
ACTTGATCACAAAACTGTTCCAGAACCGTGTGCAACTTGGTCTTGTCTAA
AGAACCGGCTCTGCTTCGATGTGATTCCGGTCTCCAGCATTCCACCTGTAGTTGGGTGAGTAGCT
GTCCGTAGGTGGAGTTCAGGTGAACCAATTCG

B



C

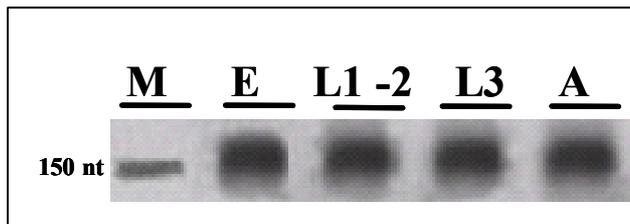
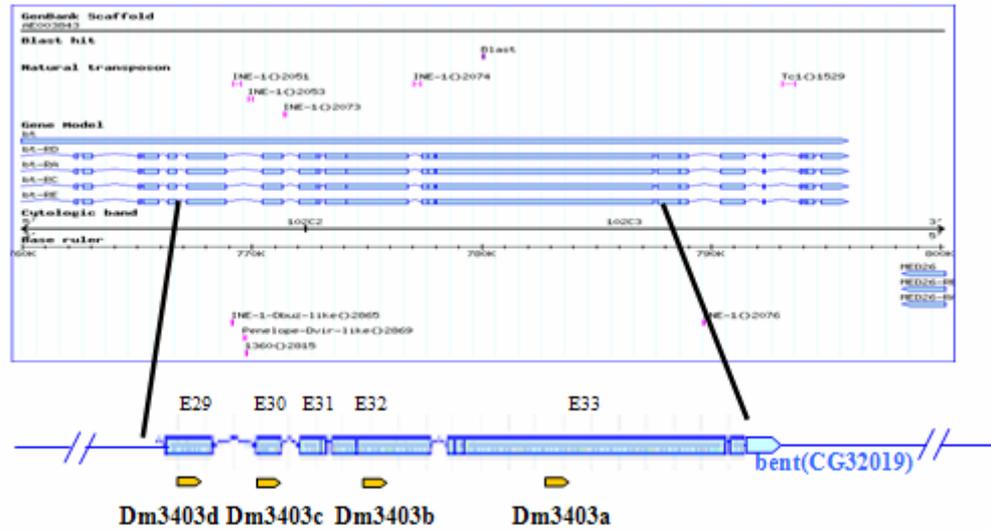


Figura 21: Identificazione del cluster DmSnR69
 A. I geni DmSnR69a e DmSnR69b sono localizzati nella regione genomica 75A2, sul braccio sinistro del cromosoma 3, rispettivamente nell'esone 1 e a cavallo della giunzione esone3/introne3 del gene CG14586. In basso sono riportate le sequenze genomiche dei geni DmSnR69a-b, in rosso sono indicate le box C e D, la regione antiseno complementare all' rRNA è sottolineata; da notare nel caso del gene DmSnR69b la sequenza guida antiseno è variante e non complementare all'rRNA. B. Appaiamento della sequenza antiseno dello snoRNA DmSnR69a con l'rRNA; la putativa guanina metilata è indicata in grassetto e sottolineato. C. Analisi trascrizionale del gene DmSnR69a: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 150 nt

Il cluster Dm3403

Il cluster Dm3403 è costituito dal gene Dm3403a e dalle sue tre isoforme Dm3403b, Dm3403c, Dm3403d, individuate mediante analisi delle sequenze fiancheggianti la sequenza inizialmente identificato da SNOSCAN. Questo cluster rappresenta il primo esempio di cluster esonico ed i geni che lo costituiscono (Dm3403a, Dm3403b, Dm3403c e Dm3403d) sono localizzati rispettivamente all'interno degli esoni 33, 32, 30 e 29 del gene *bent* (*bt*, CG32019). Il gene *bt* è situato nella banda politenica 102C2-3 sul cromosoma 4; esso mostra un'organizzazione abbastanza complessa, essendo costituito da ben 45 esoni e codifica per una proteina-chinasi citoplasmatica probabilmente coinvolta nel metabolismo dei carboidrati. La sequenza guida antisense è uguale per tutti e mostrato complementarità perfetta con l'rRNA 28S per un tratto di 10 nt (Fig.22B).

A



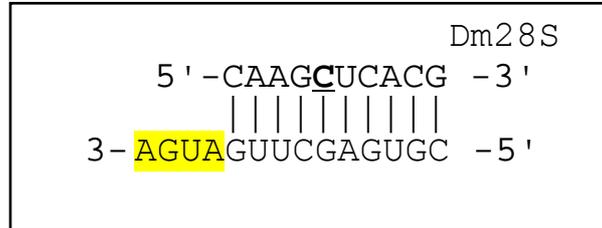
Dm3403a: CGCTGAAATCAGT**GATGAC**ATCA**CCAATGCTCACGTTCTGACTTGATT**
 GAAGGCCTTAAATATGAATTCCGTGTCAGAGCAGTAAACAAAGCCGGTCCAGGC
 TCACCAAGTGATGCA**CTGA**AACACATGTG

Dm3403b: CCAATGCTCACGTT**CTGACTTGATTGAAGGCCTTAAATATGAATTCCGT**
 GTCAGAGCAGTAAACAAAGCCGGTCCAGGCTCACCAAGTGATGCA**CTGA**AACACA
 TGTG

Dm3403c: CGCTGAAATCAGT**GATGAC**ATCA**CCAATGCTCACGTTCTGACTTGATT**
 GAAGGCCTTAAATATGAATTCCGTGTCAGAGCAGTAAACAAAGCCGGTCCAGGC
 TCACCAAGTGATGCA**CTG**

Dm3403d: AATGCTCACGTT**CTGACTTGATTGAAGGCCTTAAATATGAATTCCGTG**
 CAGAGCAGTAAACAAAGCCGGTCCAGGCTCACCAAGTGATGCA**CTGA**AACACAT

B



C

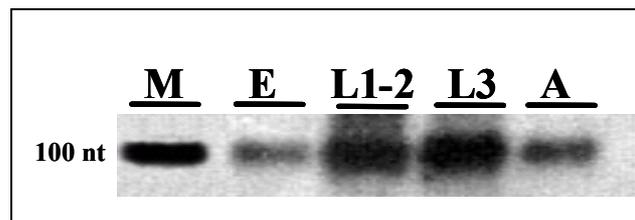
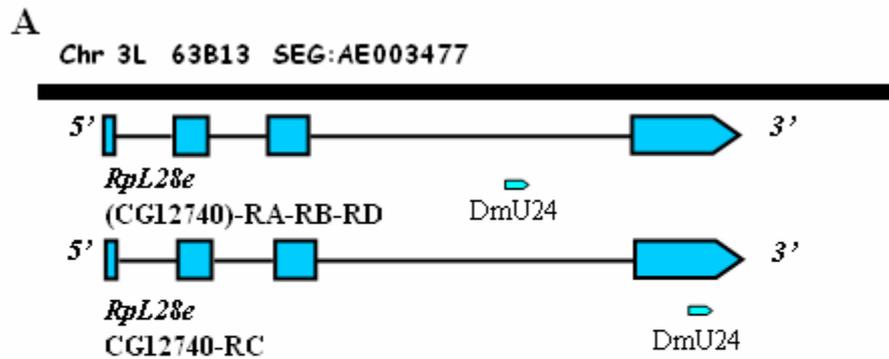


Figura 22: Identificazione del cluster Dm3403

A. I geni Dm3403a-b-c-d sono localizzati nella regione 102C2-3 del cromosoma 4, rispettivamente negli esoni 33, 32, 31 e 29 del gene bent (CG32019), sono riportate le sequenze dei geni Dm3403a-b-c-d e in rosso sono indicate le box C, D' e D ed è sottolineata la sequenza antisenso. B. Appaiamento tra la sequenza antisenso e l'rRNA 28S, notare la perfetta complementarità, la putativa citosina metilata è sottolineata e in grassetto. C. Analisi trascrizionale del gene Dm3403a: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 100 nt.

2.4.1.5 Geni localizzati in regioni soggette a splicing alternativo

Due geni, Dm425 e DmU24 (vedi tabella 5) sono risultati localizzati in una regione genomica sottoposta a splicing alternativo; la loro localizzazione può risultare intronica od esonica, in dipendenza del pattern di splicing considerato per il gene ospite. Ad esempio, il gene DmU24 mappa in una regione del gene *RpL28* (*Ribosomal protein L28*) ed è localizzato nell'introne 1 dei trascritti CG12740-RA-RB-RD del gene ospite, mentre per il trascritto CG12740-RC è presente nell'esone 1 (Fig. 23A). Il gene *RpL28* è situato nella banda politenica 63B13 sul braccio sinistro del cromosoma 3. La sequenza guida antisense dello snoRNA DmU24 è situata a monte della box D' e l'appaiamento parte 1 nt a valle della box D', condizione questa considerata compatibile con il riconoscimento dell'rRNA bersaglio (Kiss-Laszlo *et al.*, 1996; Cavallè *et al.* 1996; Fig. 23B). L'esperimento di Northern blot, condotto utilizzando una sonda genomica di 400 bp comprendente lo snoRNA DmU24 ha mostrato che DmU24 presenta un profilo di espressione (Fig.23C), nel quale è presente un segnale di ibridazione in tutti gli stadi di sviluppo esaminati.



DmU24: GGTTCCTTGTCATGATGATCTTTTTTTTAAATGCTACACA
 ACCAAGATCGCAGAAATGATCGTGTGATAAACTTATCT
 GAAGCTCTCAAAGCTGAAGCTTATCG

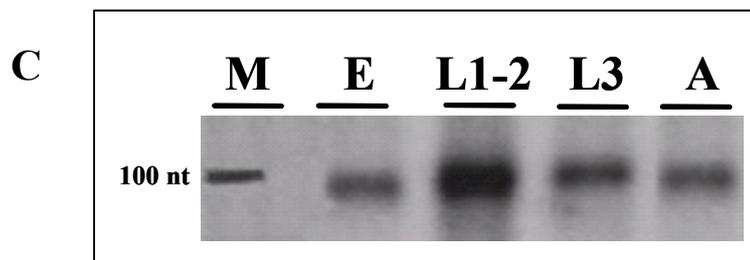
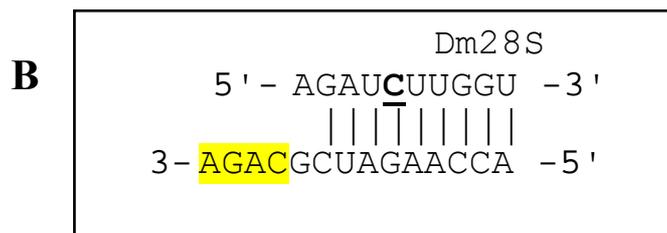


Figura 23: Identificazione del gene DmU24.

A. Il gene DmU24 è localizzato nel gene *Ribosomal protein L28e* (CG12740) situato nella regione 63B13 del braccio sinistro del cromosoma 3 in una regione sottoposta a splicing alternativo per cui può trovarsi o nell'introne 3 o nell'esone 3 del gene ospite a seconda del pattern di splicing. Sequenza del gene: in rosso sono indicate le box C, D' e D ed è sottolineata la sequenza antisense. **B.** Appaiamento della sequenza antisense con l'rRNA, notare la perfetta complementarità, la putativa base metilata è indicata in grassetto e sottolineata. **C.** Analisi trascrizionale del gene DmU24: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 100 nt.

2.4.2 **Parte II:** Analisi dei putativi geni identificati in base alla presenza di siti di metilazione conservati fra le sequenze degli rRNA di *Drosophila melanogaster* e *Homo sapiens*.

Dai saggi tesi a confermare sperimentalmente l'effettiva espressione dei putativi geni codificanti per snoRNA identificati tramite quest'approccio bioinformatico è emerso che solo 10 dei 56 geni candidati risultavano effettivamente espressi in *Drosophila melanogaster* (Tabella 8). E' plausibile che alcuni geni siano sfuggiti alla conferma sperimentale perché poco espressi, o espressi secondo un profilo di espressione temporale e spaziale molto ristretto e quindi al di sotto del limite di apprezzabilità della tecnica utilizzata. Per 2 geni (Dm1006 e Dm1007) lo SNOSCAN ha identificato anche due isoforme alternative, Dm1011 e Dm1012 shiftate di pochi nucleotidi e l'analisi delle sequenze fiancheggianti il gene Dm1008 ha portato all'identificazione di un'isoforma aggiuntiva denominata Dm1008a. Dall'analisi della localizzazione dei 10 nuovi geni identificati è emerso che solo 2 geni erano localizzati in introni di geni ospiti che codificano per proteine, 4 mostravano una localizzazione intergenica (di cui 3 organizzati in un unico cluster), ed altri 2 erano localizzati in esoni dei geni ospiti. Infine, 2 geni sono risultati localizzati a cavallo di giunzioni esone/introne (vedi Fig.24). L'analisi trascrizionale ha poi evidenziato, per ulteriori 6 sonde, la presenza di trascritti con lunghezza notevolmente superiore a quella predetta per gli snoRNA. Le sonde hanno infatti identificato 4 trascritti di lunghezza compresa tra 700 nt e 1000 nt, e due di circa 500 nt, per i quali non è stata notata nella sequenza genomica la

Tabella 8: Lista degli snoRNA di *Drosophila melanogaster* identificati in base all'allineamento fra l'rRNA 18/28 S di *Drosophila* e Uomo e confermati dall'analisi trascrizionale.

Simbolo SNOSCAN Nome assegnato	Sito putativo 28S in Dm (sito omologo in H.s.) SnoRNA di <i>H.sapiens</i>	Score (SNOSCAN)	Lunghezza (nt) (SNOSCAN)	Localizzazione all'interno del gene ospite	Box aggiuntivi	Putativo target
Dm_28S.1006 DmU79/62A12a		31.20	94		-	rRNA 28S
Dm_28S.1007 DmU79/62A12b	A (Am2633) U79	28.77	87		-	rRNA 28S
Dm_28S.1008 DmU79/62A12c		27.91	106	Intergenica tra CG7974/RpL23a	-	rRNA 28S
Dm_28S.1011 DmU79/62A12a	A (Am 2628) U30	22.57	94			rRNA 28S
Dm_28S.1012 DmU79/62A12b		21.83	110			rRNA 28S
Dm_28S.1046 DmU30/21E2	A (Am 2628) U30	22.08	107	Intergenica tra U1-21.1/gluRIII	-	snRNA:U1:21D
Dm_28S.1044 DmU79/46E3	A (Am2633) U79	21.10	138	Introne 1 <i>eiger</i>	-	snRNA:U1:82Eb snRNA:U12:83B
Dm_18S.1031 DmU54/88F5-6	G (Gm603) U54	20.16	100	Introne 1 CG14869	2 box D' 1 box C'	snRNA:U2:34ABa
Dm_28S.1047 DmU30/23F6	A (Am 2628) U30	21.67	119	Esone 3 <i>tim</i>	2 box D' 1 box C'	-
Dm_28S.1021 DmU55/39/83B2/4	C (Cm 2101) U55/39-67/D10	21.98	143	Esone 3 CG2926	1 motivo H/ACA 2 motivi CAB 1 box D'	snRNA:U4:25F snRNA:U5:23D snRU12
Dm_18S.1009 DmU55/39/67D/10	C (Cm 2101) U55/39	24.04	97	Esone2/Introne2 <i>Ect</i>	2 box D' 1 box C	-
Dm_28S.1040 DmU55/39/53F5-7	C (Cm 2101) U55/39	24.14	103	Esone7/Introne5 CG30460	1 motivo H/ACA 2 motivi CAB 1 box D'	snRNA:U1:82Eb snRNA:U1:82Eb

presenza di motivi addizionali di sequenza tali da ipotizzare la loro classificazione nell'ambito della classe degli scaRNA. Pertanto, tali trascritti al momento non possono essere considerati appartenenti né alla categoria degli snoRNA, né a quella degli scaRNA e saranno quindi descritti separatamente.

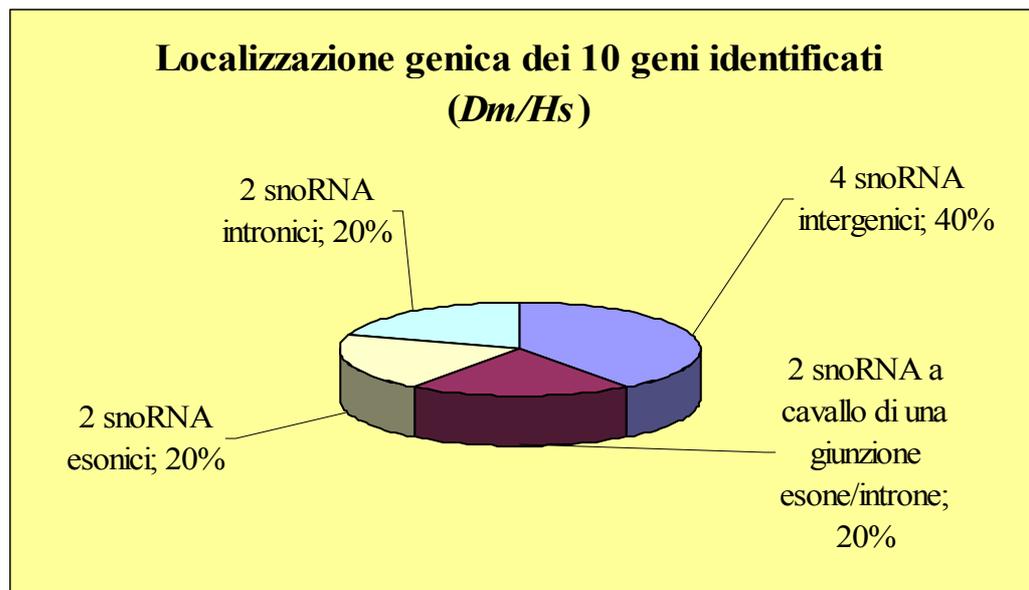


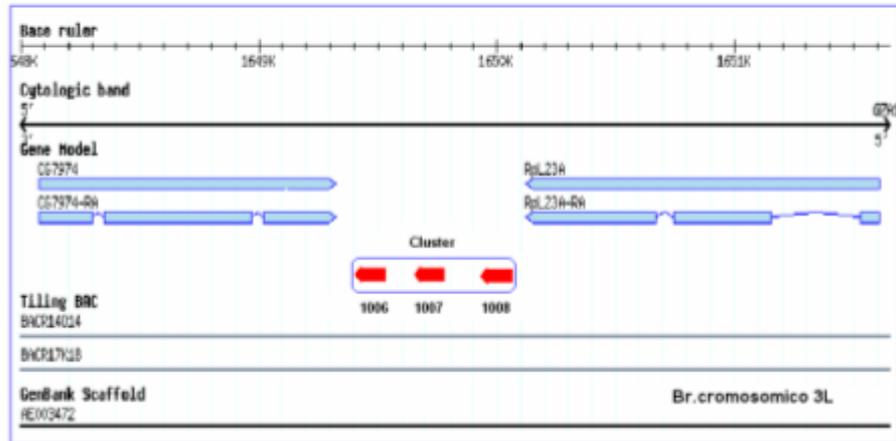
Figura 24

2.4.2.1 Geni a localizzazione intergenica

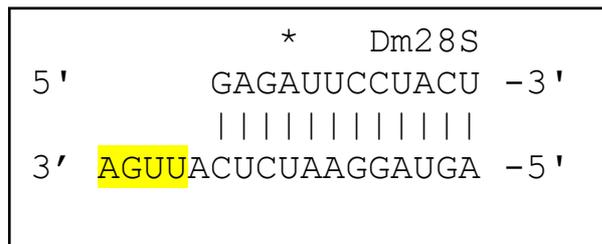
I nuovi geni a localizzazione intergenica (Dm1006, Dm1007, Dm1008, Dm1046, vedi Tabella 8) hanno presentato tutti un profilo di espressione costitutivo e dimensioni corrispondenti a quelle predette dal programma SNOSCAN; 3 di questi geni sono organizzati in cluster (Dm1006, Dm1007, Dm1008) ed hanno mostrato caratteristiche che inducono ad ipotizzare l'esistenza di una comune trascrizione policistronica, nonché di una sovrapposizione trascrizionale con il gene adiacente *RpL23a*.

Il cluster formato dai geni Dm1006, Dm1007 e Dm1008 è infatti risultato localizzato tra il gene CG7974, trascritto con polarità opposta, ed il gene *RpL23a*, trascritto con lo stesso orientamento e localizzato a sole 225 bp (Fig.25A) dall'estremità 5' del cluster. Le sequenze guida antisenso, lunghe 12 nt, si trovano in tutti e tre membri del cluster a monte della box D' e l'appaiamento previsto con lo rRNA 28S è di complementarità perfetta (Fig.25B). Un'attenta analisi delle sequenze a monte del gene 1008, ha poi rivelato la presenza di un'ulteriore putativa variante genica, da me denominata Dm1008a (Figura 26A). La sequenza del putativo snoRNA 1008a diverge significativamente da quella degli altri membri ad eccezione di un breve segmento di omologia con il gene 1008 (Fig.26B), questo gene non presenta a monte della box D', la sequenza antisenso complementare a all' rRNA e deve quindi, per ora, essere considerato "orfano".

A



B



C

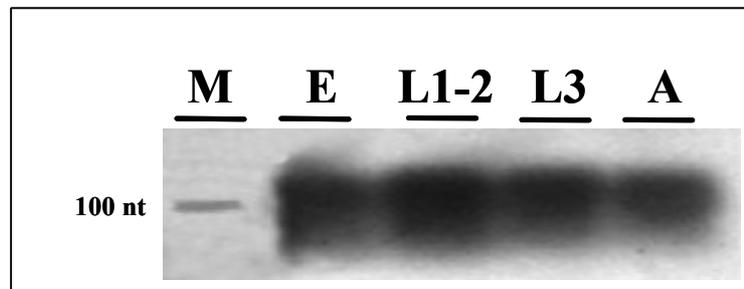
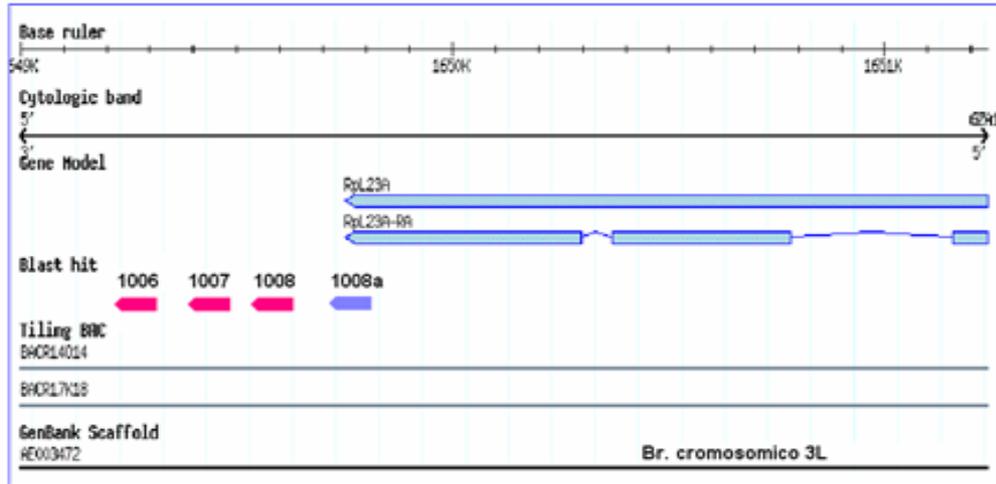


Figura 25: Identificazione dei geni Dm1006, Dm1007 e Dm1008

A. Gli snoRNA 1006, 1007 e 1008, rappresentati in rosso, sono localizzati sul braccio cromosomico 3L, in posizione intergenica tra il gene *CG7974* trascritto con orientamento opposto, ed il gene *RpL23a*, trascritto con la stessa polarità. Il cluster copre una regione genomica di 532 bp. B. Appaiamento tra la sequenza guida antisenso e lo rRNA 28S. L'adenina asteriscata rappresenta la putativa base metilata. Notare la perfetta doppia elica lunga 12 nt formata dalle due sequenze appaiate. C. Analisi trascrizionale dello snoRNA Dm 1006: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 100 nt.

A



B

Esone 3 Rpl23a

1141 AACACCCOTGG TCTTCCTCAC ACACCTGCGC GCCAACAAGA ACCACGTGCG TGCCCGCAGTG
 1201 CGCAAGCTCT ACGACATCAA GGTGGCCAAG GTGAACGTGC TCATCCGGCC CGATCGCCAG
 1261 AAGAAGGCCT ATGTGCGCCT GCGCGCGAC TACGACGCC TGGACATTGC CAACAAGATC
 1321 GGCATCATAT AAGCGCTGCA ATCTGCGGTA ACGTGTGGAA TGGGCATCGC GTTCTACATT
 1381 TTATATGTAG ATTCTTTTCT ATAAACAAG TAATGCTAAG CTGACACGGT

Dm 1008a 5' AGAGCGAGGA

1441 TGAACGTTTC AAAGTTGTCT GGGTTTAATA AAACCACTCA AAAACTAAGT GTGGGTTTTTC
 1501 AATGTGGGCT CAATGGGTTT TGAGTTT 3'

CTT TCAGTTTTCT CTTTCAGATT GGATACTTCA

1561 GGTTTTAGCG CTCTAAATTG GTGAGTTTTT AATAACAAGC AATCTTTCAA TTGCTAATA
 1621 CTACCTTCTA TCGCAGGCTG TCTTTAAATG GGAAGCTTAA CTTGCTAAAC GAAGAGGTAA
 1681 TTAAGTTTTT ATTTTGTAAT TTTAATGTAT

Dm28S 1008 AACACTTAGC TGATGATATT ACGAGTAGGA

1741 ATCTCATTGA AATAAOCATG AAATCAATTT TATCTCCACT GATAGCATTG TGCTCAAGGA
 1801 AGCTGAACTC AACCAACCA AGCTAATCAA GACTTTATTT CGTTTCAGTT TTCAAICTGG
 1861 CTCGTATCTG GATCCTTGA TATGAAATTC TTCAAGGTGT GTATATTTTCG GATGGAGTAG
 1921 TTTGAATTG

Dm28S 1007 T AATCTGCTCT GATGATATTT CGAGTAGGAA TCTCATGAG CTACCTGTAG

1981 CTATGAAATC AATTTTAATC TCCCACTGAT CAAATCATT CAAACAACA TGGCTTCTCT
 2041 TAAAGCGCTT TATTTAACAA TTCTTTTTTA ATTCATTTCG GTTGCOCAT TGTTCOCCGA
 2101 TTCTCTGGAC TAACCTCTTA TAAGGTGTGT ATCTGTTTTA GTTACTGACA AC

Dm28S 1006 TTCTTTAT

2161 TGTGATGATA TTTTGAGTAG GAATCTCATT GAGCCTCAGT CATTCGTTGC TTTGAAATTA
 2221 ATTTTGTTCT TCACTGATTG CAAGCATAAA AGCAAATATG TTCATCGTTG TGAAGCATAA

C

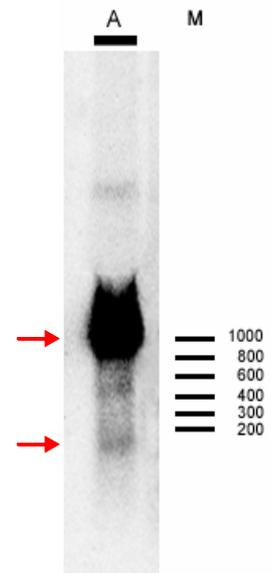


Figura 26: Identificazione dell' isoforma Dm1008a

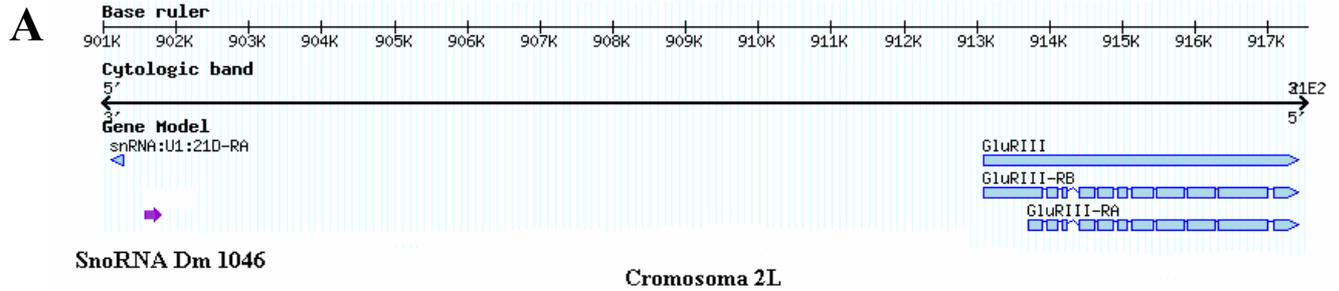
A. Lo snoRNA 1008a, rappresentato in viola, è localizzato a monte del gene Dm1008 e risulta sovrapposto all'ultimo esone del gene *Rpl23a*. B. Una parte dell'esone 3 del gene *Rpl23a* è rappresentata in blu; la porzione di sovrapposizione con Dm1008a, lunga circa 60 bp, è in grassetto; sono rappresentate anche le sequenze degli altri membri del cluster genico, mentre in rosso sono evidenziati i putativi motivi di sequenza della classe degli snoRNA C/D, varianti nelle posizioni nucleotidiche sottolineate. La putativa sequenza antisense capace di riconoscere lo rRNA 28S è sottolineata a monte della box D'. In verde sono evidenziati gli oligonucleotidi usati per l'amplificazione tramite PCR della regione genomica di 418 bp utilizzata come sonda nell'esperimento di Northern blot riportato a lato. C. L'analisi mediante Northern blot conferma l'espressione dello snoRNA 1008a in una preparazione di RNA totalI ottenuta da individui adulti. Il segnale di ibridazione è di circa 100 nt, come indicato dalla freccia in basso. Si osserva inoltre un segnale di circa 1000 nt corrispondente al trascritto del gene *Rpl23a* ed indicato dalla freccia in alto. M: marker di RNA di dimensioni note (RNA ladder low range -Fermentas).

Un dato particolarmente interessante è che questo gene risulta parzialmente sovrapposto (60 bp) all'ultimo esone (esone 3) del gene *RpL23a*, estendendosi poi per altre 40 bp nella breve regione intergenica che separa *RpL23a* da Dm1008, il primo gene del cluster (Fig.26B). Questi dati hanno suggerito la possibilità di una regolazione coordinata tra il gene che codifica per la proteina ribosomale L23, situato al 5', ed i geni degli snoRNA localizzati al 3'. Era quindi interessante verificare la possibilità che anche il gene 1008a venisse trascritto. L'esperimento di Northern blot effettuato con la sonda genomica opportuna ha in effetti confermato l'espressione dello snoRNA 1008a in preparazioni di RNA totali ottenute da individui adulti (Fig.26C); bisogna inoltre notare che la sonda riconosce, come atteso, anche il trascritto del gene *RpL23a*. L'effettiva espressione del gene 1008a ha suggerito la possibilità che la sovrapposizione del cluster dei geni 1006, 1007, 1008 e 1008a ed il gene *RpL23a* fornisce un esempio in *Drosophila* di un'organizzazione genica del tutto simile a quella recentemente descritta per una significativa percentuale dei geni umani presenti sui cromosomi 21 e 22. Infatti, la mappatura dei siti di legame per fattori di trascrizione (TFBS) Sp1, p53 e cMyc sui cromosomi umani 21 e 22 ha rivelato che la maggior parte di questi siti (36%) si trova nelle vicinanze o entro 5 kb dal 3' di geni protein-coding, e presumibilmente marcano il punto d'inizio della trascrizione di geni non-coding (Cawley *et al.*, 2004). E' anche interessante notare che il programma SNOSCAN ha identificato una box D' alternativa sia nella sequenza dello snoRNA 1006 che nella sequenza dello snoRNA 1008; queste box D' alternative presentano a monte una sequenza antisense complementare allo rRNA 28S capace di

Le molecole di snoRNA Dm 1006/Dm 1011 e Dm 1008/Dm 1012 potrebbero avvalersi dell'utilizzo alternativo delle due sequenze antisenso e guidare differenti eventi di metilazione sulla medesima molecola di rRNA target. E' altresì plausibile che le box D' presenti nelle isoforme Dm 1011 e Dm 1012, varianti in ben 3 e 2 posizioni nucleotidiche, possano non essere affatto funzionali.

Lo snoRNA Dm1046

Il gene Dm 1046 potrebbe essere trascritto in maniera indipendente e ciò rende questo gene particolarmente interessante, dato che gli snoRNA a localizzazione intergenica e non organizzati in cluster, sono una rarità nel sottoregno dei metazoi. Dm1046 è localizzato nella banda politenica 21E2 del braccio sinistro del cromosoma 2 (2L) (Fig.28A), fra il gene *gluRIII*, che mappa circa 11.4 kb a valle ed il gene *U1-21.1*, trascritto con polarità opposta e situato sole 320 bp a monte. Il gene *gluRIII* codifica per un prodotto proteico con attività di recettore cationico del glutammato ed è coinvolto nella trasmissione sinaptica nella giunzione neuromuscolare. Il gene adiacente in 5', *U1:21:1*, lungo circa 171 bp, codifica invece per lo snRNA U1 di tipo a di *Drosophila*, un gene coinvolto nell'attività di splicing della molecola di pre-mRNA attraverso la formazione dello "spliceosoma" di tipo-U2. La sequenza antisenso dello snoRNA Dm1046, localizzata a monte della box D', non presenta un tratto di perfetta complementarità con lo rRNA. Questo dato mi ha spinto a ricercare altri possibili target di modificazione all'interno della banca dati Noncode (<http://noncode.bioinfo.org.cn>).



```

901444 GGCTACATAT ATGAAAATAA TGATACTTTT AAAATTTTATT GCAGATGTCG GGACGGCGCG
901504 AACGCCATT CCGGCTACCA AAAATTACAC GCACGAGTTA TTCACATTAG GAATAATCGC
901564 AGAGGTCAAC TCAGCCGAGG TGCAATGGCC AAGCCTCACT CCGGAGGAAC CGCCTTCGTG
901624 ATCACGGTTA ACCTCTACGC CAGGTAAGTA TGCTTTCTC AGCTCAGGGA ATGGGTTTCC
901684 ATCAGCTGAT GCCAGAAGC AGTTGGGAAT TATTGCAGTG AAAGTCCGTG AAAGCTTAC
901744 CCCCAGAAAGT ATGTTAGCAA AATGTTACAT CTGCAGTTGG GTTGATTTT GTAAACTGTG
901804 CGTTGATATT GTGGTTCTTA TCTGCTTCCA CCCTTATACA TATTATATGT AACCTTTGGA
901864 TTTTACTGTT TCATTTAAAT GATTATATTA ACGAGCGAAT AATGATAAGA ATAATTTTAA
901924 TTAAACTAA ACAATCTGAG AGGCCACAA TAGTTTTACA AGCCACGAAA AT
                                     Dm28S 1046                                     5' ATCAGCAA
901984 TAAAATGGTG AAAATCGCC GGAATATGGT TAACCTGACA ACGAAACAGA GCATCTTCTA
902044 TACGGAATAT TTATGGGGTA GTAATAATG ATTTTGCAG 3'
                                     A ATTCCATAGG CCGCAAAAAA
902104 CTAATAATCG AAATAGTTTC GGTAECTCGA TAGAAAGTTC GATGGTGTGA ATGGATCAGC
902164 GGGAGACTGG GGCGCCGAAA GTGATTTAAT TTATGGCAAT TACAAATTGT TATTAGAAAT
902224 TTGCGGTCAA CGAGGGCGCA CACGACCGCA GGTGAATTG AAATCAAATG GAAAAATGCG
902284 AAATGAACCA GGCTGAAATA GCAAAAAGC AGGAGAGCAA AAATTATTGT TGACATAAGC

```

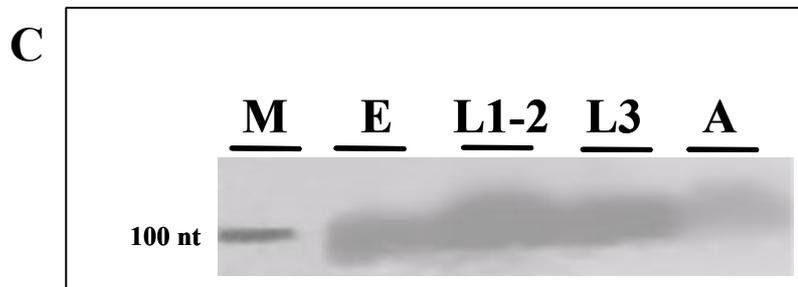
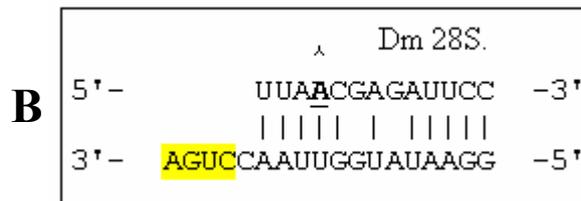
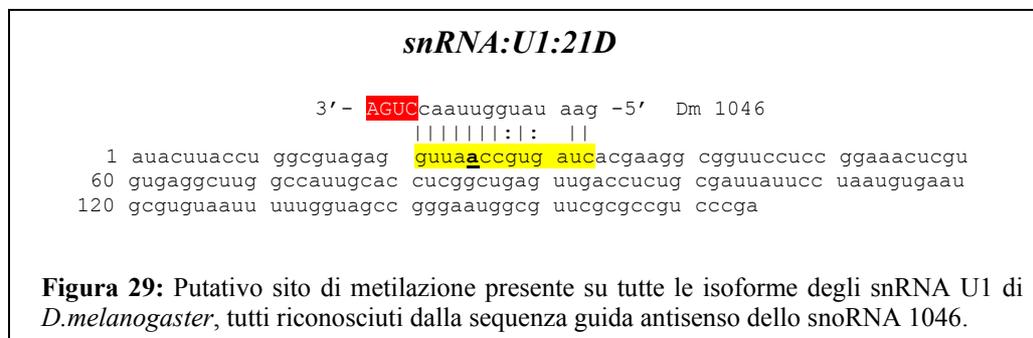


Figura 28: Identificazione del gene Dm28S 1046

A. Il gene Dm 1046 è localizzato sul braccio cromosomico 2L nella regione intergenica tra il gene per lo snRNA *U1:21* ed il gene *gluRIII*. In basso è riportata la sequenza dello snoRNA Dm1046 con i motivi C, D'e D indicati in rosso: la sequenza guida antisenso è sottolineata. In verde sono mostrati gli oligonucleotidi utilizzati per la PCR ed in azzurro è evidenziata la sequenza genomica a monte del gene *U1:21D*, con i due elementi promotori a monte PSEA (5'-TAATTCCCAACTGCTTCTGGC-3') e PSEB (5'-CATGGAAA-3') in viola **B**. Appaiamento tra la sequenza guida antisenso e lo rRNA 28S: notare la presenza di due mismatches. L'adenina marcata è la putativa base metilata. **C**. Analisi trascrizionale dello snoRNA Dm 1046: l'espressione è pressocchè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 100 nt.

Tale approccio è risultato fruttuoso ed ha consentito di individuare un possibile sito di metilazione sulla stessa molecola di snRNA U1, all'interno di una regione perfettamente conservata in tutte le isoforme di questo gene finora identificate (cioè 7) in *D. melanogaster*. La figura 29 mostra infatti l'appaiamento tra lo snoRNA e la sequenza del snRNA U1:21D, il cui gene, come già detto, è localizzato solo 320 bp a monte ed è trascritto in direzione opposta. E' quindi possibile ipotizzare che la trascrizione dello snoRNA Dm1046 e di uno dei suoi possibili target, lo snRNA U1:21D, sia soggetta ad una comune regolazione. Il gene snRNA U1:21D contiene nella breve regione intergenica che lo separa da Dm1046 le sequenze conservate PSEA e PSEB che sono gli elementi basilari per la trascrizione degli snRNA di *D. melanogaster* da parte RNA polimerasi II (Jensen *et al.*, 1998; McNamara-Schroeder *et al.*, 2001).



La vicinanza tra i due elementi di sequenza ed il gene Dm1046 suggerisce che gli stessi elementi possano promuovere anche la trascrizione dello snoRNA, in maniera coordinata a quella del suo RNA bersaglio, lo snRNA U1:21D. E' interessante notare che quest'organizzazione è del tutto insolita, e rappresenterebbe il primo esempio di regolazione coordinata della trascrizione fra uno snoRNA ed il suo target di modificazione.

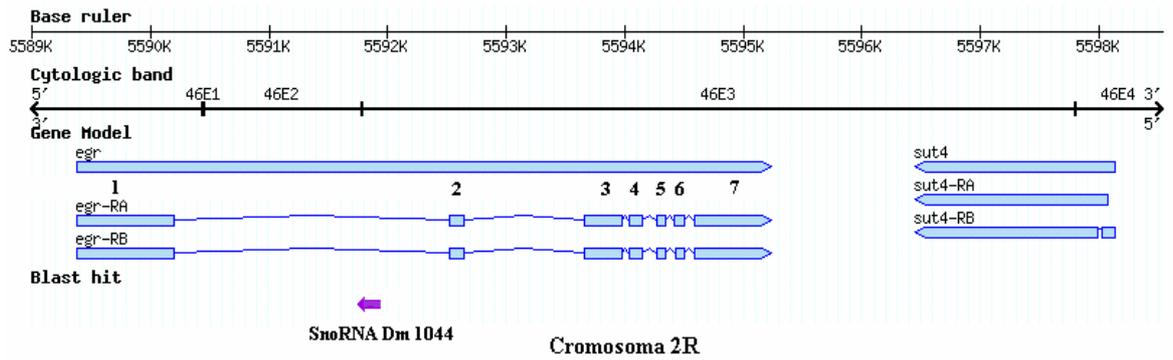
2.4.2.2 Geni a localizzazione intronica

La conferma sperimentale dei geni a localizzazione intronica (Dm1044 e Dm1031; vedi Tabella 8) ha evidenziato la presenza di trascritti di dimensioni atipiche (300-400 nt) per molecole appartenenti alla classe degli snoRNA C/D. Inoltre, per entrambi i geni non era presente una regione di perfetta complementarità con l'RNA ribosomiale. L'insieme di questi dati mi ha spinto a ricercare nelle sequenze fiancheggianti quelle identificate da SNOSCAN la presenza di motivi addizionali C/D, H/ACA o CAB per verificare la possibilità che queste molecole insolitamente lunghe potessero avere una struttura più complessa, di tipo composito.

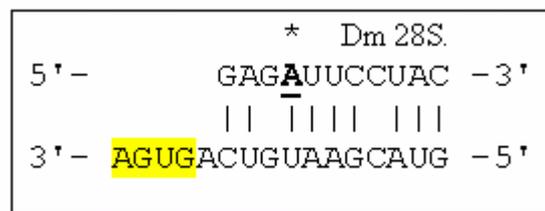
Lo snoRNA Dm1044

Il gene per lo snoRNA Dm 1044 è localizzato nella banda politenica 46E2 sul cromosoma 2R, all'interno del primo introne del gene ospite *eiger*, ed è trascritto con polarità opposta a quella del gene ospite (Fig.30A), un tipo di arrangemento genico che, come già notato, potrebbe riflettere un meccanismo di biogenesi basato sulla produzione alternativa degli mRNA/snoRNA sovrapposti. Il gene ospite, *eiger*, codifica per il primo ligando appartenente alla superfamiglia dei fattori di necrosi tumorale (TNF: Tumor Necrosis Factor) scoperto negli invertebrati. La proteina EIGER, lunga 409 aa, presenta un dominio idrofobico transmembrana ed un dominio C-terminale extracellulare che mostra omologia con il TNF. *Eiger* è espresso in tutti gli stadi di sviluppo della *Drosophila*, prevalentemente nel sistema nervoso, dove sembrerebbe

A



B



C

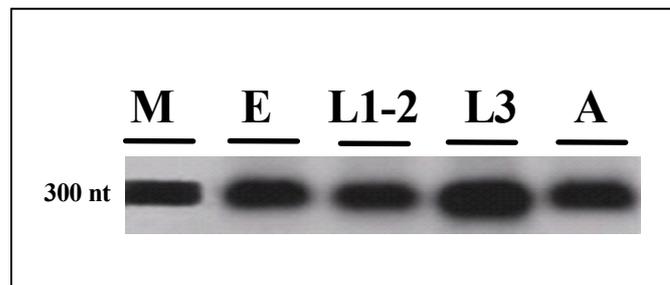


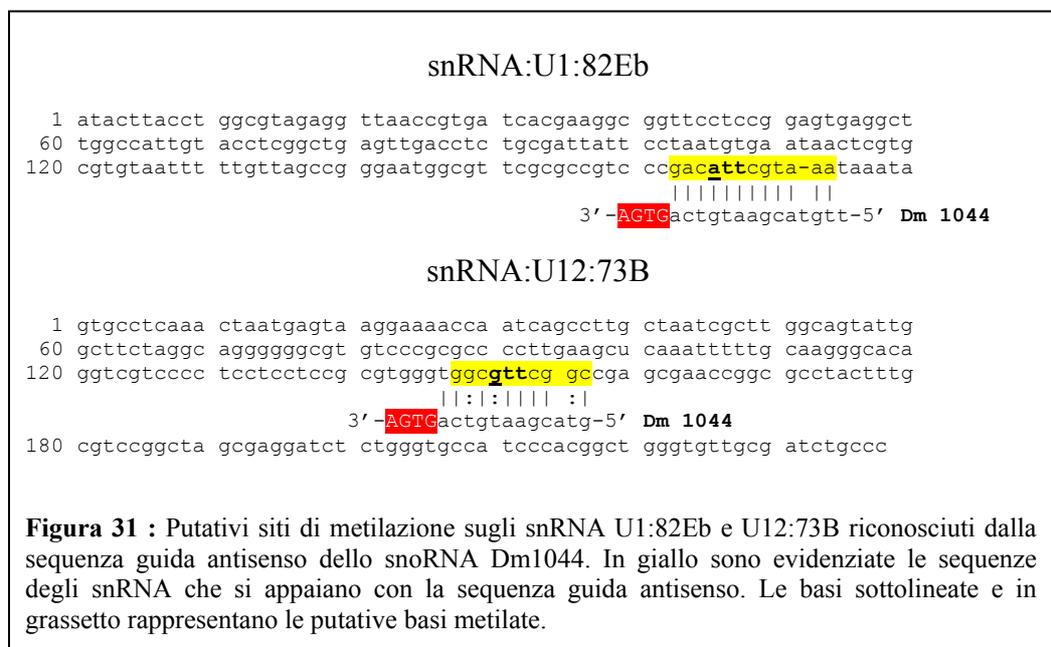
Figura 30: Identificazione del gene Dm1044.

A. Lo snoRNA 1044 indicato in viola, è localizzato sul braccio cromosomico 2R sul filamento antisense al gene *eiger*, all'interno del primo introne. In basso è indicata la sequenza genomica con i motivi C, D' e D in rosso e la sequenza antisense sottolineata.

B. Appaiamento tra la sequenza guida antisense dello snoRNA Dm1044 e lo rRNA 28S. L'adenina sottolineata e in grassetto rappresenta la putativa base metilata. Notare i due mismatches tra le due sequenze.

C. Analisi trascrizionale dello snoRNA Dm1044: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 300 nt.

regolare la proliferazione delle cellule progenitrici dei neuroni permettendo il normale sviluppo del sistema nervoso (Moreno and Basler, 2003). L'analisi trascrizionale ha evidenziato la presenza di un trascritto di 300 nt in tutti gli stadi di sviluppo, una lunghezza che non coincide con quella stimata dal programma SNOSCAN (Fig.30C). Tuttavia un'analisi preliminare delle sequenze fiancheggianti, volta a mappare più in dettaglio il gene Dm1044, ha escluso la presenza dei motivi H/ACA o di motivi addizionali C/D canonici. Poiché l'appaiamento tra sequenza guida antisense dello snoRNA Dm1044 ed rRNA non è risultato perfetto (2 mismatches; Fig.30B) ho intrapreso anche per questo gene la ricerca di un putativo target tra le sequenze degli snRNA contenuti nella banca dati Noncode (<http://noncode.bioinfo.org.cn>). Questa ricerca ha consentito di individuare due possibili siti di metilazione su due diversi snRNA, lo snRNA U1.82Eb e lo snRNA U12 (Fig.31).



Lo snRNA U1.82Eb è uno snRNA U1 di tipo b la cui sequenza è variante rispetto a quella canonica per la presenza di una coda addizionale al 3'

terminale. In *D. melanogaster* sono state isolate solo 7 copie di geni per lo snRNA U1, di cui 3 sono varianti (U1a U1b e U1c), mentre nell'Uomo ne sono state finora identificate 30, nel topo 40 e in *X. laevis* 1000 (Goodal *et al.*, 1991). E' stato dimostrato che la variante snRNA U1b è espressa specificamente durante gli stadi embrionali dello sviluppo nelle cellule di *Drosophila* capaci di differenziare, mentre lo snRNA U1a è abbondante in tutti gli altri stadi ed è espresso soprattutto nelle linee cellulari differenziate terminalmente (Patrick and Mount, 1990). Il sito riconosciuto dallo snoRNA Dm1044 è specifico per la regione 3' terminale dello snRNA U1 di tipo b che, come descritto, è assente nelle varianti geniche di tipo a e c. L'altro possibile target, lo snRNA U12, è invece presente in un'unica copia nel genoma della *Drosophila* e la sua lunghezza di 238 nt lo rende il più lungo snRNA U12 finora conosciuto. Lo snRNA U12 media il processo di splicing degli introni di classe ATAC, i cui siti donatori 5' e accettori 3' sono costituiti dai dinucleotidi AT e CA rispettivamente. In tale meccanismo lo snRNA U12 interagisce con gli snRNA U4-U6 atac. In conclusione, appare probabile che lo snoRNA Dm1044 vada in effetti a modificare gli snRNA U1 e U12, anziché lo rRNA.

Lo snoRNA Dm1031

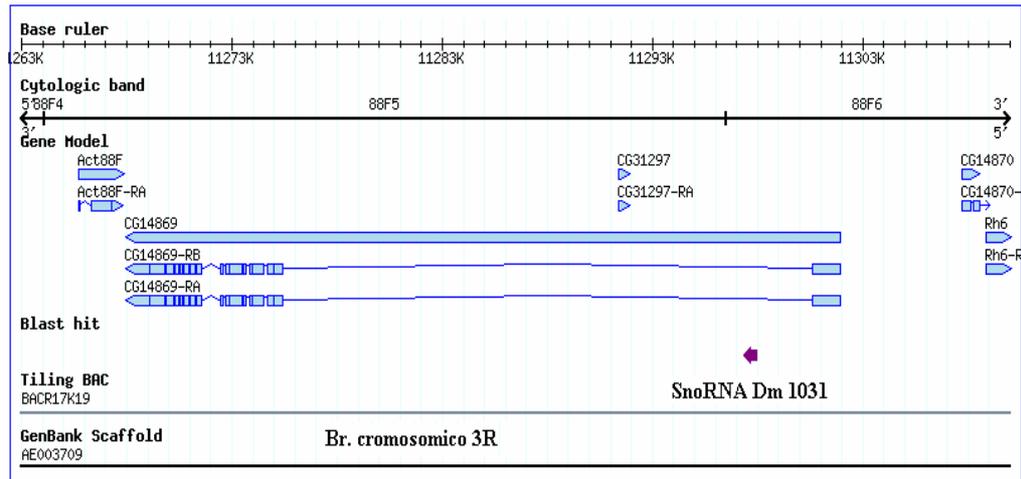
Il secondo gene a localizzazione intronica identificato nel corso di questo screening è stato il gene Dm1031, che mappa a cavallo delle bande citologiche 88F5 ed 88F6 sul braccio destro del cromosoma 3 (3R). Dm1031 è localizzato all'interno del 1° introne del gene CG14869, lungo 2.5 kb, e mostra il

medesimo orientamento trascrizionale del gene ospite (Fig. 32A). Il gene CG14869 è lungo 33 kb e codifica per due trascritti, entrambi composti da 17 esoni: CG14869-RA, di circa 6969 nt e CG14869-RB, di circa 6968 nt, i cui prodotti proteici, rispettivamente di 1537 e 1688 aa, possiedono un'attività metalloendopeptidasica e sono putativamente coinvolti nel meccanismo di comunicazione cellulare (<http://www.flybase.org>). Il pattern di espressione di questo gene ed il suo reale ruolo biologico sono tuttavia ancora da definire e l'unico allele finora disponibile è quello wild type (<http://www.flybase.org>).

L'analisi del profilo di espressione ha evidenziato la presenza un trascritto di circa 350 nt in tutti gli stadi di sviluppo; in larve di 3° stadio si osserva invece un segnale di ibridazione della lunghezza di circa 500 nt (Fig.32C). E' interessante notare che un trascritto di questa lunghezza non corrisponde a nessun trascritto del gene *eiger* attualmente riportato in FlyBase (<http://www.flybase.org>).

L'analisi delle sequenze fiancheggianti ha evidenziato la presenza di un ulteriore motivo D canonico (CTGA) e di un ulteriore motivo C, variante in un solo nucleotide (GTGATGC) rispetto alla sequenza consenso; questi nuovi motivi potrebbero definire la reale estremità 5' dello snoRNA 1031. A valle della sequenza è presente un' ulteriore box D canonica che potrebbe invece definirne l'estremità 3'(Fig.32B). Considerando le box aggiuntive identificate sia al 5' che al 3', la lunghezza del gene potrebbe diventare di 500 nt, una lunghezza corrispondente al segnale di ibridazione rilevato in larve del 3° stadio.

A

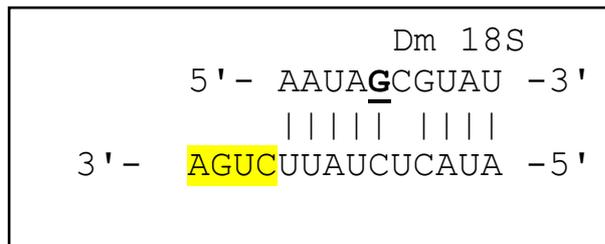


Introne 1 (*CG14869*)

```

3661 CGATTAATGA TCGTAAATA TTCAGCTCGT CTACCCGTGA TTAATGCCAC ACATAACCAT
3721 CATTCTGTAT ACCGATTCTC ATGCAAATGA GTAGGGCGTG TCCTTGCAGC CCTTTGCATG
3781 GTGGATGGGG TGACAAATAA TTGTCGGAA TATTAATGCC GTCAGAGAAC AAAATAGAAA
3841 TAAACCACAG GCAGTAAAAA GGCCAGAAAA CGAGCAAATA ATTTTAGGTC TGATGCCGGG
3901 GCGTGACTTT GGGGTGCTGT GCTTATGCTT TAAGCTGGTT ATTTTGAATT TAATACAATT
3961 TTTTTCGCCA AAAGATATAT ACGTGATCTG AATGCAAACA CTTCTGCGAA GGCCCGCATC
4021 CCCCACCCCT CCCCTCGGGA AACCACCCAC GTTCCCATA AAAGCAATTT TGCTCATTTT
4081 CTAGCAAACG AAAA Dm18S 1031
5' CTGTCA AAATGATGAA AAAAATACTT ATAAAACCTCT AGGGAATTAT
4141 TTCGAGCAAG AGAGCCGACA CTGACATACT CTATTCTGAT TTCTAAGCAT TTG 3'
AACACTT
4201 TTCAAAAATG TTGTAGATCC CTAAACATA AAAGTTAACG ACTACAAGCC ACCTTGCCTA
4261 TAAATTAATC ATTGACTTAT AAGTAGTAAT CATATTTGTA TTGTTTACCA CGCTAATTTA
4321 TTAATCTGA GCCATAGGAA GTGT TAGGA AAATATGATC AAAGATTCC GGGCCTGACA
4381 AGTGATTGCT CATCTTAATT AATTCCTTTT TAATTATCA GTCAAGATTA ATGTTAAACA
  
```

B



C

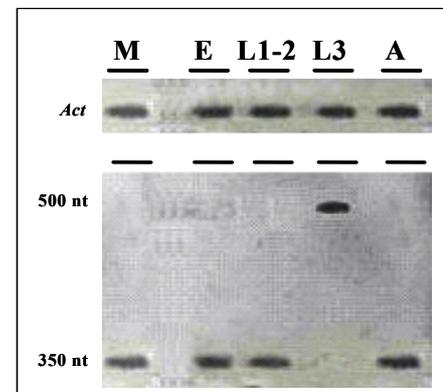
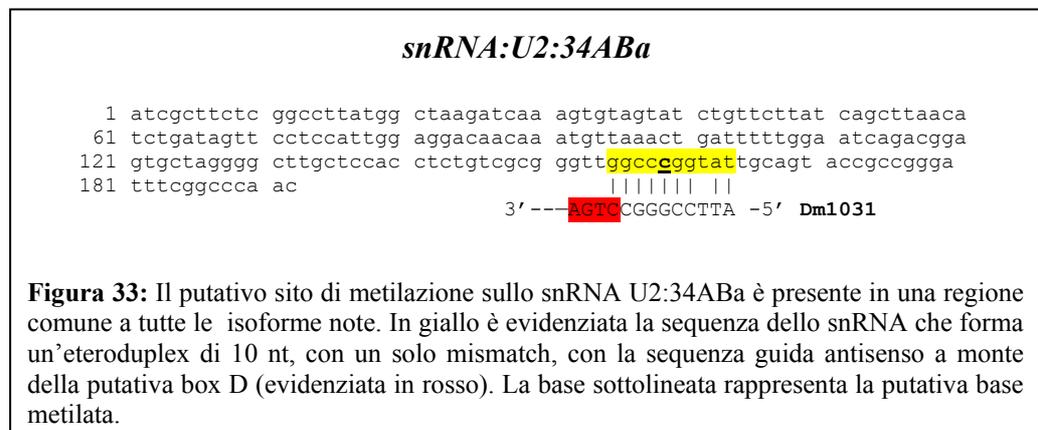


Figura 32: Identificazione del gene Dm1031.

A. Gene Dm1031, localizzato sul braccio cromosomico 3R, nel primo introne del gene CG14869. In basso è mostrata la sequenza genomica dello snoRNA con in rosso i motivi C e D. In fucsia sono evidenziati ulteriori motivi di sequenza: un' ulteriore box D e una box C a monte, ed una box D con la putativa sequenza guida antisenso (in azzurro), a valle della sequenza genomica. E' inoltre sottolineata la sequenza guida antisenso stimata dal programma SNOSCAN. **B.** Appaiamento tra la sequenza guida antisenso dello snoRNA Dm1044 e lo rRNA 28S. L'adenina sottolineata e in grassetto rappresenta la putativa base metilata. Notare i due mismatches tra le sequenze. **C.** Analisi trascrizionale del gene Dm1031: un segnale di 350 nt è presente in E (embrioni), L I-II (larve di 1°-2° stadio), ed A (adulti). In larve di 3° stadio (LIII) viene rilevato invece un segnale di ibridazione di circa 500 nt. In alto è riportata la normalizzazione con Actina. M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas).

Considerando invece solo la box aggiuntiva identificata al 3', la distanza di quest'ultima dall'estremità 5' predetta da SNOSCAN è di circa 300nt, cioè corrispondente a quella dei segnali di ibridazione rilevati nelle preparazioni di RNA estratti da embrioni, larve di 1°-2° stadio ed adulti.

L'identificazione delle box D aggiuntive mi ha spinto a verificare la presenza di possibili target riconosciuti dalle sequenze guida antisenso situate a monte fra le sequenze depositate nella banca dati Noncode (<http://noncode.bioinfo.org.cn>). Questa ricerca mi ha portato ad identificare un possibile snRNA target per la putativa sequenza antisenso della box D identificata a valle dell'estremità 3' predetta dallo SNOSCAN. Tale sequenza potrebbe appaiarsi con una sequenza comune a tutte le isoforme note dello snRNA U2, determinando la formazione di un eteroduplex snoRNA/ snRNA interrotto da un solo mismatch, come mostrato in figura 33:



Nel genoma di *D. melanogaster* sono state identificate 6 isoforme dello snRNA U2. Queste isoforme si assemblano in una snRNP, formando insieme alla snRNP U6 il "core" catalitico dello spliceosoma di tipo U2, che determina lo splicing della maggior parte degli introni. E' stato osservato che le modifiche guidate dagli snoRNA sulle molecole di snRNA possono determinare il legame di altre proteine specifiche alle snRNP. Lo snRNA U2

pare infatti capace di legare gli specifici fattori di splicing SF3a ed SF3b solo dopo la sua modificazione (Yan *et al.*, 1996).

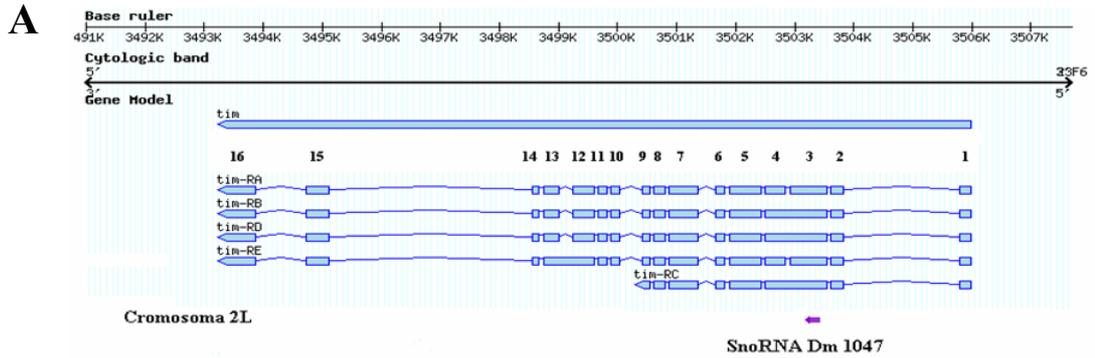
E' interessante notare che la struttura dello snoRNA Dm1031, così come risulta dall' identificazione dei nuovi motivi tipici della classe di uno snoRNA C/D sopra descritti, è simile a quella dello snoRNA umano di classe C/D codificato in un introne del gene KIAA1731. Questo snoRNA, denominato mgU2-19/30 (Tycowski *et al.*, 2004) guida la metilazione dei residui G19 ed A3 sullo snRNA U2 ed è costituito da un doublet (doppietta) di snoRNA di classe C/D separati da un tratto a singolo filamento di circa 170 nt. Lo snoRNA si accumula in cellule Hela sia sotto forma di "doublet" (mgU2-19/30) lunga circa 250 nt che sotto forma di "singlets" (mgU2-19 e mgU2-30) lunghe 80-90 nt; ma una delle "singlet", mgU2-19, appare molto meno abbondante dell'altra (mgU2-30). La molecola costituita dai due domini (mgU2-19/30) sembra localizzarsi nei Cajal Bodies, mentre le molecole costituite dai singoli domini sembrerebbero localizzarsi nel nucleolo. L'accumulo e la maturazione dello snoRNA Dm1031 potrebbero essere di tipo analogo e verranno approfondite in successive analisi sperimentali. Sulla base dei dati finora ottenuti appare infatti possibile ipotizzare che il segnale di circa 500 nt potrebbe corrispondere ad una molecola contenente entrambi i domini C/D, mentre il segnale di circa 300 nt dovrebbe invece corrispondere alla molecola di snoRNA "processata", contenente un solo dominio C/D. Si noti che le dimensioni delle bande relative alle due possibili forme di "singlets" sono di lunghezza molto simile e quindi difficilmente risolvibili su gel di agarosio.

2.4.2.3 Geni a localizzazione esonica

Come già commentato, uno dei risultati più interessanti emersi dal mio lavoro è stata l'identificazione di geni che codificano per snoRNA all'interno di regioni esoniche, un'organizzazione finora non ancora identificata per nessuna molecola di snoRNA nei metazoi. Tuttavia, anche per i geni Dm1047 e Dm1021 (vedi Tabella 8) è stata evidenziata la presenza di trascritti di dimensioni superiori a quelle tipiche degli snoRNA di classe C/D. Entrambi i geni hanno mostrato un profilo di espressione costitutivo e la presenza di motivi strutturali addizionali che per uno dei due (Dm1021) mi ha spinto ad ipotizzare l'appartenenza alla classe degli snoRNA localizzati all'interno dei Cajal bodies (scaRNA; Kiss *et al.* 2002).

Lo snoRNA Dm1047

Tra gli snoRNA a localizzazione esonica, il gene Dm 1047 è particolarmente interessante, essendo ospitato nel terzo esone di un gene, *timeless (tim)*, che codifica per una proteina importante per il mantenimento del ritmo circadiano della *Drosophila* (Fig.34A). Mutazioni amorfe del gene *tim* sono recessive e provocano negli adulti anomalie nel ritmo circadiano, nonché problemi locomotori, di accoppiamento e di sviluppo (Bae *et al.*, 1998). Il gene *tim* è localizzato nella banda politenica 23F6 sul braccio sinistro del cromosoma 2 (2L). *tim* è costituito di 16 esoni e codifica per 5 trascritti alternativi, tim RA (~ 4939 nt), RB (~ 5035 nt), RC (~ 3109 nt), RD (~ 5035 nt) e RE (~ 5176 nt). Lo snoRNA Dm1047 si origina dal terzo esone, in una regione che appare essere inclusa in tutti i 5 trascritti (Fig.34A). Tra la sequenza antisenso dello snoRNA Dm1047 e la regione dell'rRNA non è presente un perfetto

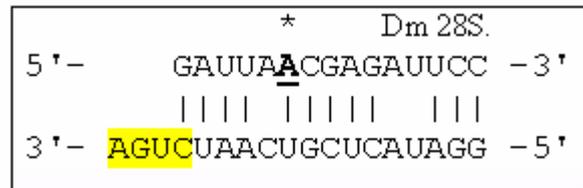


```

3504461 GGTATGGAC TGGTACTAG CAACTCCGCA GTTGTACAGC GGGTTCTOCT CCTGGGGTTG
3504401 CTGGAGGGC GATACCTATG TGGTCAACCC GAATGCATTG GGTGAGTAAT CAATATAGGA
3504341 CTACTACATT AATTAGGTAT TTAACCTATT TTCAATCCOC AGCCATTCTG GAGGAGATCA
3504281 ACTACAAGCT CACCTATGAG GACCAAACAC TGCGCACCTT TCGACGGGCC ATTGGATTTG
3504221 GCCAGAATGT GAGGTCAGAC CTGATACCGC TGCTGGAGA
                                Dm28S 1047      5' A TCCCAAGCAT GATCGGGTCC
3504161 TGGAGTCGGT CATCCGGATA CTGCTCAATC TGA CGGTGOC GGTGGAGTGC CTCTTCTCCG
3504101 TGGACGTGAT GTACCGCACG GATGTGGGTC GCCACACCAT CTTCGAGCTG AACAACTGC
3504041 TGTACACCAG 3'
                                CAAGGAAGCA TTTACCGAGG CCAGGAGCAC CAAGAGCGTG GTGGAGTACA
3503981 TGAACACAT ACTGGAGTCG GACCCTAAGC TGTOGCOGCA CAAATGCGAT CAAATCAACA
3503921 ACTGTCTGCT GCTGCTGAGA AATATCCTTC ACATTCCAGA GACGCATGCC CATTGCGTGA
3503861 TGCCATGAT GCAGTCGATG CCGCATGGCA TCTOCATGCA GAACAGATC TTGTGGAATC
3503801 TCTTCATCCA GAGCATCGAC AAGTTACTOC TGTATCTGAT GACCTGTCCG CAGAGAGCCT
3503741 TCTGGGGAGT GACCATGGTG CAACTGATTG CTTTGATCTA CAAGGATCAG CATGTGAGCA

```

B



C

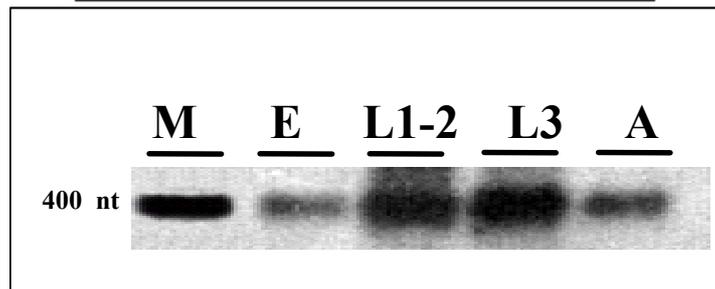


Figura 34: Identificazione del gene Dm1047

A. Il gene Dm1047 indicato in viola, è localizzato sul braccio cromosomico 2L nel terzo esone del gene *timeless*. In basso è mostrata la sequenza genomica dello snoRNA Dm1047 con indicate in rosso le box C, D' e D. In azzurro sono evidenziati ulteriori motivi di sequenza C e D a valle, presumibilmente inclusi nella molecola dello snoRNA. E' inoltre sottolineata la sequenza guida antisenso. **B.** Appaiamento tra la sequenza guida antisenso e lo rRNA 28S. L'adenina asteriscata rappresenta la putativa base metilata. Notare i tre mismatches tra le due sequenze appaiate. **C.** Analisi trascrizionale del gene Dm 1047: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 400 nt.

appaiamento (3 mismatches su 15 bp), un dato che suggerisce che l'effettivo RNA target modificato da questo snoRNA potrebbe essere diverso (Fig.34B). Si noti inoltre che la lunghezza del trascritto evidenziato per lo snoRNA Dm1047 è di circa 400 nt, sensibilmente maggiore della lunghezza stimata dal programma SNOSCAN. L'analisi delle sequenze genomiche fiancheggianti lo snoRNA 1047 ha mostrato la presenza di due motivi CTGA a valle e di una sequenza ATGATGC tra di esse. Il motivo CTGA più a valle dista dall'estremità 5' stimata da SNOSCAN circa 420 bp, una lunghezza abbastanza conforme a quella del trascritto identificato nell'esperimento di Northern blot. Ciò suggerisce che questo gene potrebbe contenere 2 box D' interne alternative, entrambe canoniche, ed una sequenza C' addizionale (Fig.34A). Un'ulteriore ricerca di possibili siti di modificazione presenti sulle sequenze degli snRNA di *D. melanogaster* non ha portato ad alcun risultato positivo, così che anche questo snoRNA deve al momento essere considerato "orfano".

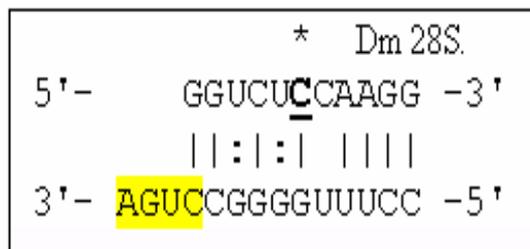
Lo snoRNA Dm1021

Il gene Dm 1021 risulta localizzato nel terzo esone del gene CG2926 e mostra il medesimo orientamento trascrizionale del gene ospite (Fig.35A). CG2926 è localizzato nella banda citologica 83B2-4, sul braccio destro del cromosoma 3. Questo gene non è ancora ben caratterizzato e codifica per un unico trascritto molto lungo, circa 7449 nt, composto da 6 esoni, il cui prodotto proteico (2296 aa) possiede un putativo dominio di "DNA binding" ed è probabilmente coinvolto nel meccanismo di ubiquitinizzazione delle proteine, facendo parte del complesso dell'ubiquitina ligasi.

A



B



C

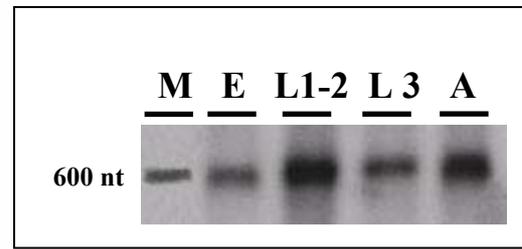
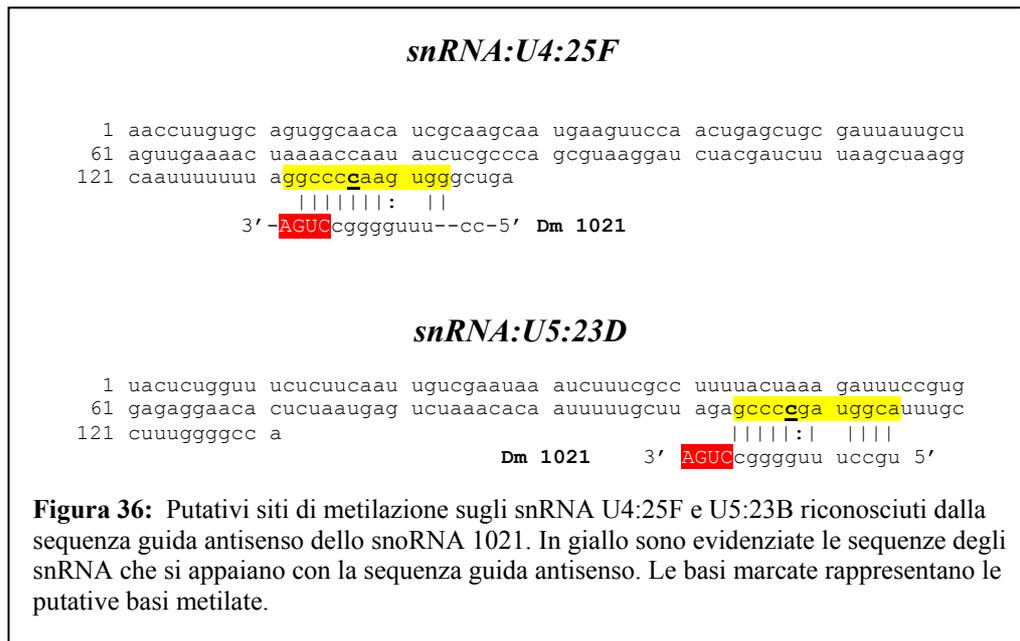


Figura 35: Identificazione del gene Dm1021.

A. Il gene Dm1021 indicato in viola, è localizzato sul braccio cromosomico 3R nel terzo esone del gene *CG2926*. La distanza dell'inizio della sequenza dello snoRNA dal sito accettore (3') del terzo esone è di circa 35 bp. I motivi C, D' e D sono segnati in rosso. In azzurro sono evidenziati i due caratteristici motivi di sequenza del dominio H/ACA (AnAnnA e ACA) e in viola due probabili box CAB (T/CGAG). Più a valle è evidenziata una possibile box D che dista dalla putativa estremità 5' del gene circa 580 bp. L'analisi della sequenza indica che Dm1021 è un possibile scaRNA, con entrambi i motivi C/D e H/ACA. E' inoltre sottolineata la sequenza guida antisense stimata dal programma SNOSCAN. B. Appaiamento tra la sequenza guida antisense e lo rRNA 28S. La citosina asteriscata rappresenta la putativa base metilata. Notare la presenza di 1 mismatch nella doppia elica formata dalle due sequenze appaiate e di due appaiamenti degenerati. C. Analisi trascrizionale dello snoRNA Dm1021: l'espressione è pressochè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 600 nt.

La lunghezza del trascritto osservato per lo snoRNA Dm1021 di circa 600 nt è molto difforme dalla lunghezza di 143 nt stimata dal programma SNOSCAN ed insolitamente lunga per una molecola di snoRNA. L'analisi delle sequenze genomiche fiancheggianti volta a mappare più in dettaglio lo snoRNA ha infatti mostrato i motivi di sequenza tipici degli scaRNA di classe H/ACA-C/D che mediano le reazioni di modificazione sugli snRNA nei Cajal Bodies. Infatti, come mostrato in figura 36A, a valle della sequenza genica di Dm 1021 è presente il motivo H (ANANNA) e, più a valle, il motivo ACA. Sono presenti inoltre i due motivi CAB (TGAG) necessari per la ritenzione degli scaRNA nei CB. Infine è mostrata una possibile box D non canonica (CTGG) che dista dalla box C presente al 5' circa 580 bp, che potrebbe determinare un'estremità 3' conforme alla lunghezza di circa 600 nt evidenziata nell'esperimento di Northern blot (Fig.35C). La ricerca per ulteriori possibili target riconosciuti da Dm 1021 ha portato all'identificazione di due putativi siti, localizzati rispettivamente sugli snRNA U4:25F e U5:23D (Fig.36), la cui complementarità con la sequenza guida antisenso include entrambi i casi include un singolo mismatch, una situazione considerata compatibile con il riconoscimento della molecola di RNA bersaglio (Darzacq *et al.*, 2002). Entrambi gli snRNA putativamente riconosciuti da Dm 1021 sono presenti nel genoma di *D. melanogaster* in varie copie: 3 copie per lo snRNA U4 e 7 copie per lo snRNA U5; tuttavia, solo una copia genica di ciascun tipo di snRNA è riconosciuta dalla sequenza antisenso di Dm 1021. Infatti, la sequenza genica di snU4:25F risulta abbastanza divergente rispetto a quella degli altri geni snU4 precedentemente caratterizzati ed esso sembra essere funzionale, vista la



presenza a monte delle sequenze promotrici altamente conservate di questa classe genica, PSEA e PSEB (Mount and Salz, 2000). Un'ulteriore analisi volta a ricercare un possibile target di pseudouridilazione sugli snRNA riconosciuto da una delle sequenze bipartite dei domini H/ACA di Dm1021 ha portato anche all'identificazione di un appaiamento significativo, lungo 12 bp, tra la sequenza bipartita a monte della box H e la regione 3' terminale dello snRNA U12. Queste osservazioni indicherebbero che Dm1021 potrebbe rappresentare il terzo esempio di snoRNA di tipo C/D-H/ACA finora identificato in *Drosophila*.

2.4.2.4 Geni localizzati a cavallo di una giunzione esone/introne del gene ospite

I due geni localizzati a cavallo di giunzioni esone/introne di geni ospiti che codificano per proteine (Dm1009 e Dm1040, vedi Tabella 8), hanno entrambi mostrato un profilo di espressione regolato durante lo sviluppo e la capacità di codificare per trascritti di dimensioni superiori a quelle normalmente attribuite agli snoRNA di classe C/D (cioè rispettivamente 400 nt e 500 nt). Anche per questi due geni sono stati ricercati nelle sequenze fiancheggianti motivi strutturali addizionali. Per uno dei due geni (Dm1040) infatti è stata osservata, come per lo snoRNA Dm1021 esonico, un tipo di organizzazione caratteristica delle molecole di scaRNA.

Lo snoRNA Dm1009

Per lo snoRNA Dm 1009 l'analisi computazionale posiziona la sequenza identificata da SNOSCAN a cavallo del secondo esone e il secondo introne del gene *ectodermal (ect)* di *D. melanogaster*. Più della metà della sequenza dello snoRNA (51 nt) cade, infatti, all'interno della sequenza codificante il gene *ect* (Fig.37A). Questo gene, localizzato sul braccio sinistro del cromosoma 3 (3L) nella banda politenica 67D10, ha una lunghezza di circa 11.6 kb e codifica per tre differenti trascritti: RA (~ 2101 nt), RB (~ 2136 nt) e RC (~ 2055 nt) che producono tre prodotti proteici, ECT –PA, PB (581 aa) e PC (283 aa). La funzione biologica delle proteine ECTODERMAL è tuttora in discussione, e nella “Flybase” (<http://www.flybase.org>) non sono ancora riportati ceppi

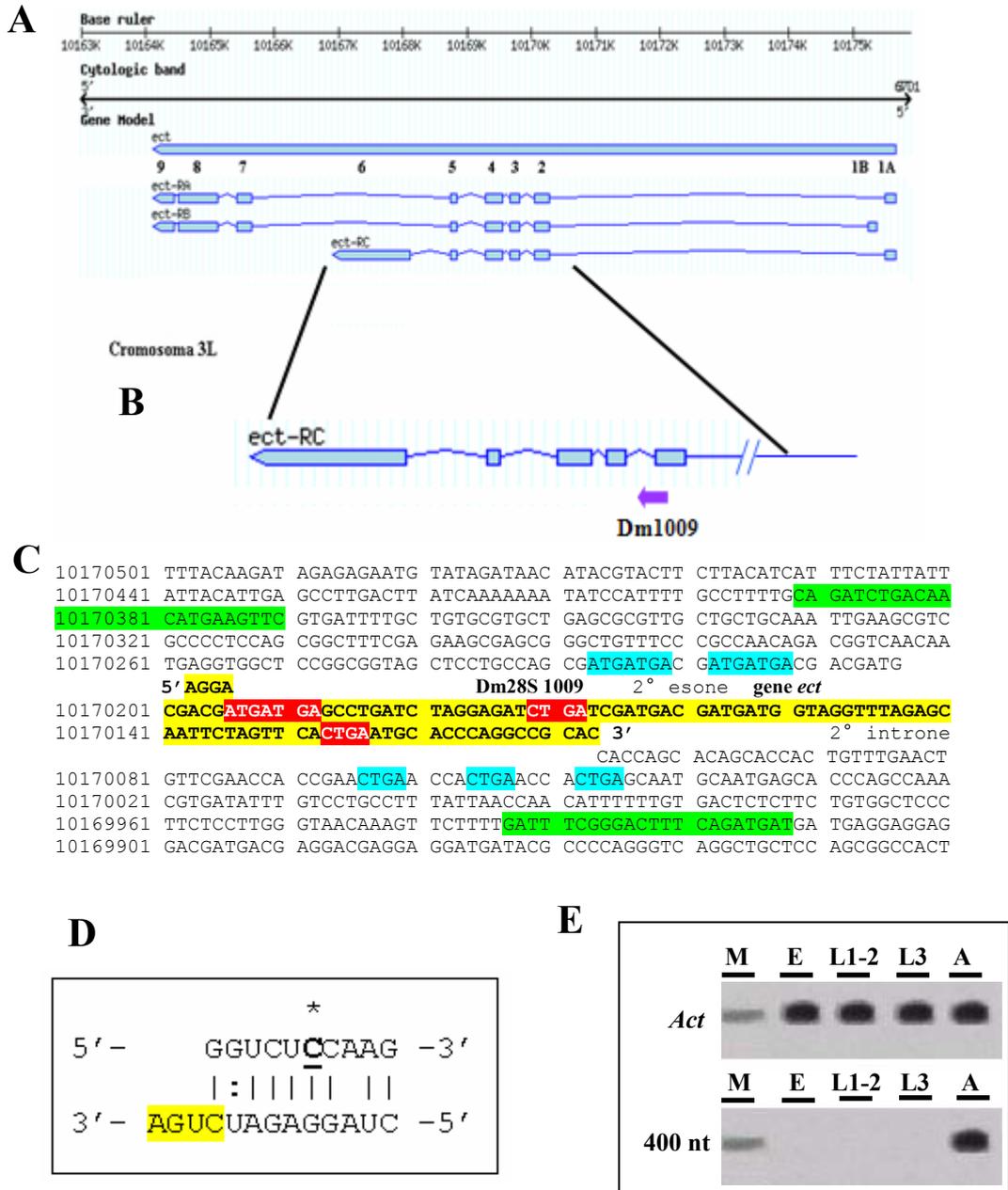


Figura 37: Identificazione del gene Dm1009.

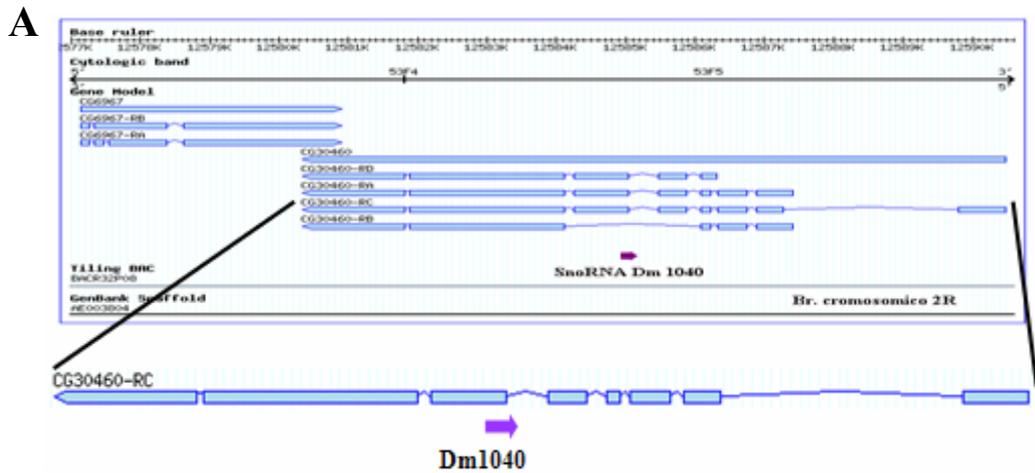
A. Il gene Dm1009, indicato in viola è localizzato sul braccio cromosomico 3L all'interno del gene *ect*, e più precisamente l'analisi computazionale posiziona la sua sequenza tra l'esone II e l'introne II del gene ectodermale (*ect*). B. Ingrandimento che evidenzia la localizzazione a cavallo della giunzione esone 2/introne 2 del gene *ect*. C. La sequenza dello snoRNA Dm 1009 è evidenziata in giallo con in rosso i motivi C, D' e D. In verde sono evidenziati gli oligonucleotidi usati per l'amplificazione tramite PCR. In azzurro sono evidenziati ulteriori motivi di sequenza, due alternative box C a monte e tre box D a valle della sequenza genomica. E' inoltre sottolineata la sequenza guida antisense stimata dal programma SNOSCAN. D. Appaiamento tra la sequenza guida antisense e lo rRNA 28S. La citosina asteriscata rappresenta la putativa base metilata. E. Analisi trascrizionale del gene Dm 1009: è presente esclusivamente negli adulti (A) e il segnale di ibridazione si localizza intorno i 400 nt, in alto è riportata la normalizzazione con Actina. M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas).

o dati diretti che possono determinare con certezza la possibile funzione molecolare o cellulare attribuibile a queste proteine. E' noto però che *ect* è espresso per un breve periodo durante l'embriogenesi in tutti i tessuti derivati dall'ectoderma, eccetto quelli del sistema nervoso (Raha *et al.*, 1990).

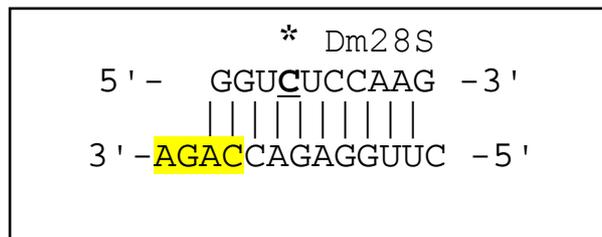
La lunghezza del trascritto del gene Dm1009 emersa dall'analisi di Northern blot è di circa 400 nt, difforme quindi dalla lunghezza di 97 nt stimata dal programma SNOSCAN. Nella Fig.37C sono evidenziati diversi motivi di sequenza ATGATGA e CTGA, presenti rispettivamente a monte e a valle della sequenza dello snoRNA, che ne potrebbero definire le reali estremità, in accordo con le dimensioni evidenziate nell'esperimento di Northern blot (Fig.37E). Un'ulteriore ricerca di possibili siti di modificazione sulle sequenze degli snRNA di *D. melanogaster* riconosciuti dalla sequenza guida antisenso di Dm1009 non ha però portato ad alcun risultato positivo.

Lo snoRNA Dm1040

Il gene Dm 1040 è localizzato sul cromosoma 2R, nella banda politenica 53F5 e viene trascritto con polarità opposta al gene ospite CG30460 (Fig.38A). CG30460 è un gene complesso, lungo 101.6 kb, che codifica per quattro trascritti alternativi: CG30460-RA (~ 5954 nt), CG30460-RB (~ 4790 nt), CG30460-RC (~ 6491 nt) e CG30460-RD (~ 5111 nt) per i cui rispettivi prodotti proteici, indicati come: CG30460-PA (1520 aa), CG30460-PB (1132 aa), CG30460-PC (1868 aa) e CG30460-PD (1402 aa) non è stata ipotizzata ancora alcuna funzione, in quanto non sono presenti domini a funzione nota. L'analisi computazionale ha identificato lo snoRNA 1040 sul filamento antisenso



B



C

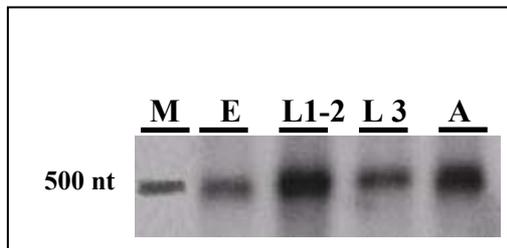
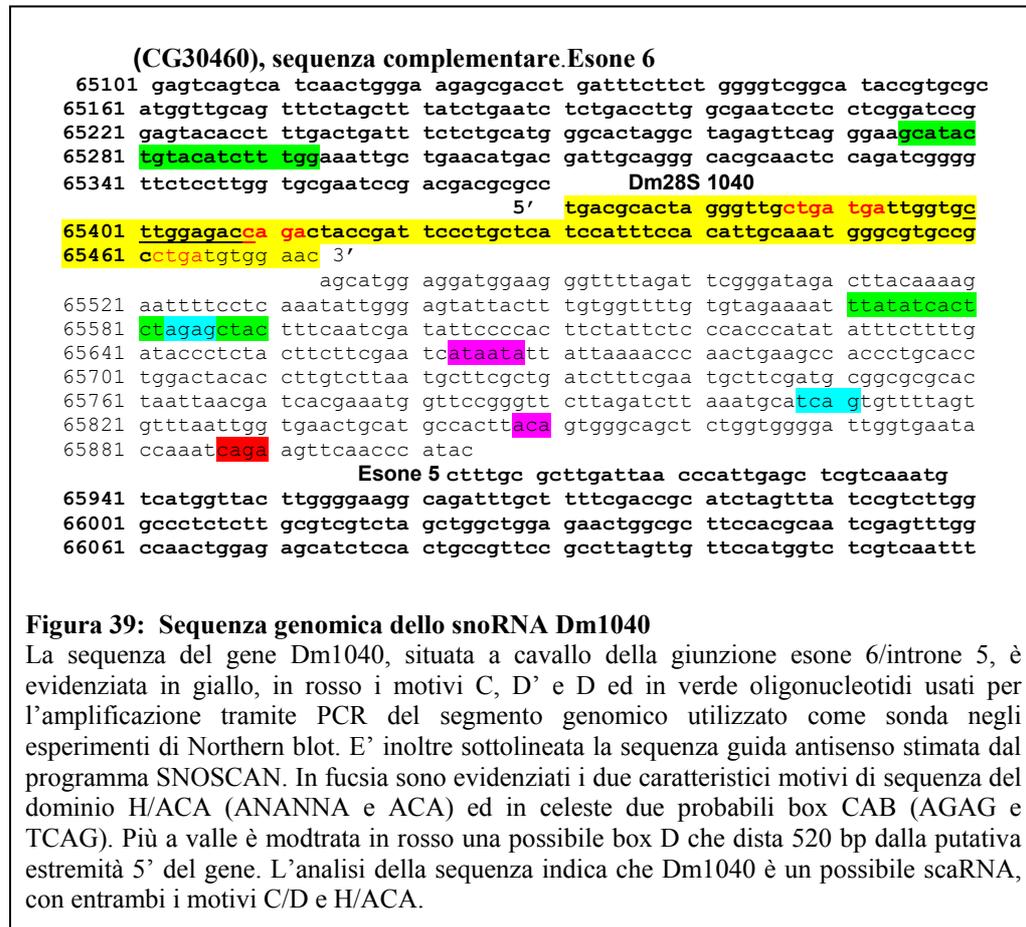


Figura 38: Identificazione del gene Dm1040

A. Il gene Dm1040, indicato in viola è localizzato sul braccio cromosomico 2R sul filamento antisense del gene *CG30460*, a cavallo della giunzione introne 5/esone 6 dei trascritti CG30460-RA, CG30460-RC e CG30460-RD, mentre si trova in corrispondenza del terzo introne del trascritto alternativo CG30460-RB. E' mostrato inoltre un ingrandimento che evidenzia la localizzazione a cavallo della giunzione esone/introne. **B.** Appaiamento tra la sequenza guida antisense dello snoRNA 1040 e lo rRNA 28S. La citosina asteriscata rappresenta la putativa base metilata. **C.** Analisi trascrizionale dello snoRNA Dm1040: l'espressione è pressocchè costante in tutti gli stadi di sviluppo (E: embrioni, L I-II: larve di 1° stadio, L III: larve di 3° stadio ed A: adulti). M: marker di RNA di dimensioni note (RNA ladder low range-Fermentas). Il segnale di ibridazione si localizza intorno ai 500 nt.

del gene ospite, in corrispondenza della giunzione introne 5/esone 6. E' interessante notare che l' esone 6 è incluso nei trascritti alternativi CG30460-RA, CG30460-RC e CG30460-RD, ma è escluso dal trascritto maturo alternativo CG30460-RB, così che il gene Dm1040 è situato a cavallo di una giunzione esone/introne sottoposta a splicing alternativo. La sequenza guida complementare allo rRNA 28S, lunga 10 nt, si trova a monte della box interna D' e l'appaiamento con la molecola di rRNA target risulta perfetto (Fig.38B). L' esperimento di Northern blot mostra che lo snoRNA Dm 1040 è espresso in tutti gli stadi di sviluppo di *D. melanogaster*, anche se il livello di espressione sembra essere maggiore negli stadi larvali, soprattutto in larve di 1°-2° instar (Fig.38C). Tuttavia, poichè la lunghezza del trascritto identificato è di circa 500 nt, molto difforme dalla lunghezza di 103 nt stimata dal programma SNOSCAN ed insolitamente lunga per una molecola di snoRNA, è stata effettuata un' analisi delle sequenze fiancheggianti volta a mappare più in dettaglio il gene. Quest' analisi ha portato ad evidenziare ulteriori motivi di sequenza tipici degli scaRNA di classe H/ACA-C/D, cioè le molecole di RNA che mediano le reazioni di modificazione sugli snRNA presenti nei Cajal Bodies (CB). Infatti, come mostrato in figura 39, a valle della sequenza genica di Dm 1040 è presente un motivo H (ANANNA) e, più a valle, un motivo ACA. Sono presenti inoltre i due motivi CAB (AGAG) e (TCAG), necessari per la ritenzione degli scaRNA nei CB. Infine, è presente una possibile box D non canonica (CAGA) a valle delle box H/ACA che potrebbe rappresentare l'estremità 3' della molecola (Fig.39). E' interessante notare che la lunghezza

di circa 520 nt del putativo scaRNA appare conforme a quella evidenziata dall'esperimento di Northern blot.



Poichè è noto che gli scaRNA mediano le reazioni di modificazione sugli small nuclear RNA (U1, U2, U3, U5 e U12 i principali), ho ricercato i possibili target avvalendomi della banca dati Noncode (<http://noncode.bioinfo.org.cn>). Tale approccio mi ha consentito di individuare due possibili siti di pseudouridilazione comuni a tutti gli snRNA di tipo U1 (Fig.40), la cui complementarità con la sequenza antisense bipartita del dominio H/ACA, lunga 13 nt, in entrambi i casi include una interruzione in una singola posizione, caratteristica considerata compatibile con il riconoscimento della molecola di RNA bersaglio (Darzacq *et al.*, 2002). In *D. melanogaster* sono

2.5 Nuovi RNA non-coding?

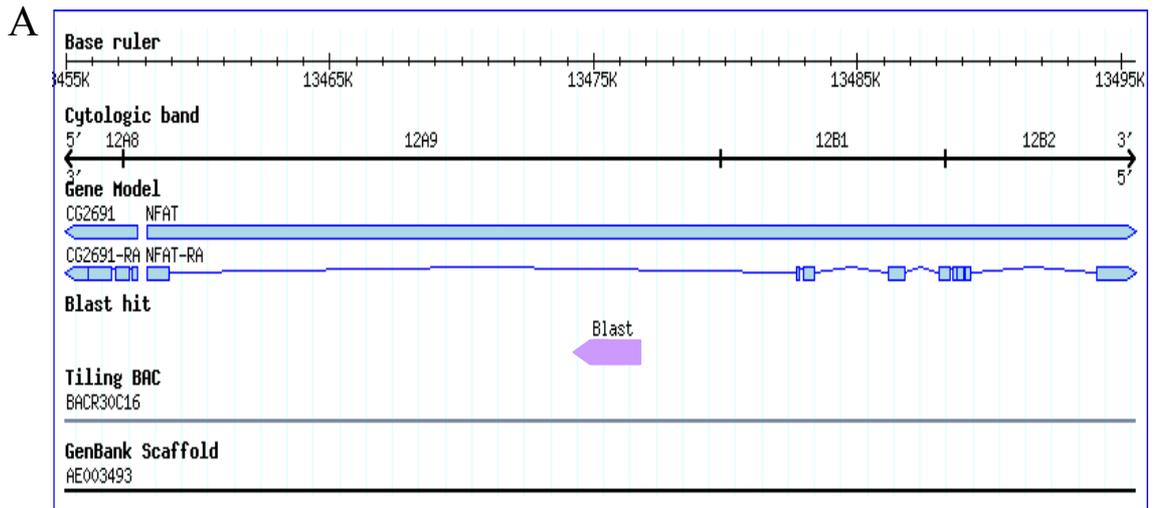
L'analisi trascrizionale per 4 delle sonde utilizzate negli esperimenti di Northern blot ha evidenziato la presenza di trascritti di dimensioni che differivano notevolmente da quelle predette per gli snoRNA. Per queste sonde la lunghezza dei trascritti era compresa infatti fra 700 nt e 1000 nt; per altre due sonde la lunghezza dei trascritti era comparabile a quella dei membri della classe degli scaRNA (cioè 500 nt), ma non è stata notata nelle sequenze genomiche adiacenti la presenza di ulteriori motivi di sequenza conservati o di eventuali sequenze antisenso capaci di riconoscere molecole di snRNA. Di conseguenza, la classificazione dei trascritti evidenziati da queste 6 sonde appare assai dubbia. Tuttavia, nessuno di questi trascritti appare coincidere con molecole di mRNA o EST annotate ad oggi per rispettivi geni ospiti nel sito <http://www.flybase.org>; inoltre, effettuando una ricerca con BLAST, non ho trovato alcun allineamento significativo con altri loci del genoma di *D. melanogaster*. I trascritti osservati potrebbero quindi rappresentare delle nuove putative molecole di RNA non-coding (pncRNA), che si originano dai geni ospiti e verranno perciò provvisoriamente definiti come pncRNA1015a, pncRNA1015b, pncRNA1016, pncRNA1019, pncRNA1020, pncRNA1028, e pncRNA1049 (pnc: putativo non coding) definizioni che derivano dai codici attribuiti ai putativi snoRNA corrispondenti, dai quali quest'analisi è inizialmente partita. Le principali caratteristiche dei 7 pncRNA identificati sono riassunte nella tabella 9; alcuni dettagli sulla loro possibile organizzazione genomica e sulle rispettive putative funzioni verranno brevemente discussi nel paragrafo successivo.

Tabella 9: Putativi non coding RNA identificati dall' analisi trascrizionale

Simbolo SNOSCAN Nome assegnato	Sito putativo 28S in Dm (sito omologo in H.s.) SnoRNA di <i>H.sapiens</i>	Score (SNOSCAN)	Lunghezza stimata (SNOSCAN)	Lunghezza reale	Localizzazione stimata (putativo gene ospite)
Dm_28S.1015a Dm_28S.1015b DmU49	<i>C</i> (Cm3340) U49	24.58	99	1000 nt 800 nt	Introne 1 (<i>NFAT</i>)
Dm_28S.1016 DmU49	<i>C</i> (Cm3340) U49	23.47	100	500 nt	Esone 2 (<i>dec-1</i>)
Dm_28S.1019 DmU55/39	<i>C</i> (Cm2101) U55/39	23.14	127	1000 nt	Esone 23 (<i>para</i>)
Dm_28S.1020 Dm55/39	<i>C</i> (Cm2101) U55/39	25.75	160	500 nt	Esone1/Introne 1/Esone2 (CG14701)
Dm_28S.1028 DmU49	<i>C</i> (Cm3340) U49	23.29	119	700 nt	Esone 5 (<i>Larp</i>)
Dm_28S.1049 DmU49	<i>C</i> (Cm3340) U49	21.69	143	1000 nt	Esone1/Introne1 (CG33123)

Putativi non coding RNA 1015a e 1015b

Questi due nuovi trascritti sono stati identificati mediante l' utilizzo di una sonda lunga 522 bp designata per convalidare l'espressione dello snoRNA Dm1015 (Fig.41). L'analisi mediante Northern blot ha rivelato un trascritto di lunghezza superiore ai 1000 nt (pnc1015a) presente in tutti gli stadi di sviluppo, ed un altro di circa 800 nt (pnc1015b) presente esclusivamente in larve del 1° instar ed in adulti (Fig.41C). La localizzazione del putativo ncRNA 1015 è particolarmente interessante. Esso si origina nel primo introne del gene Nuclear Factor of activated T cell (*NFAT*), situato sul cromosoma X (Fig.41 A), ed è trascritto con orientamento opposto al gene ospite. Il gene *NFAT*, lungo 38.9 kb, è descritto in banca dati (<http://www.flybase.org>) produrre un solo trascritto di 5858 nt, che codifica per una proteina di 1419 aa che possiede domini di legame al DNA ed altri domini tipici della famiglia dei fattori di trascrizione, come il dominio RHD (Rel Homology Domain), la cui fosforilazione regola l'attività del fattore, modulando di conseguenza l'espressione dei relativi geni target. Mutazioni che cadono nel gene *NFAT* determinano difetti nella neuroanatomia della *Drosophila* alterando lo sviluppo degli occhi e le sinapsi; alcune mutazioni sono poi letali (Abdelilah-Seyfried *et al.*, 2000; Tseng and Hariahan, 2002). In seguito ad un meccanismo di defosforilazione mediato da una fosfatasi calcio-dipendente, il fattore di trascrizione *NFAT* viene accumulato nel nucleo dove diventa trascrizionalmente attivo. Le dimensioni del trascritto del putativo gene Dm1015 e la sua localizzazione sul filamento opposto a quello del suo gene ospite lascerebbero ipotizzare la possibilità che il gene *NFAT* possa essere regolato dal pncRNA 1015.



B

```

5'
13475554 TGCATATAAG GTTCACCCTA ATATAGTGCG CATACGCAGC GCGGCCCGA TGCATAAAAT
13475614 AGAGCGAAAA GAGATGAAGA TGGAAATGTC AGGTTCCAGA TCGATCTTTT GATTTTAATA
13475674 TTGATATCGC GGCTCCATCT CTCTCCCTCT CGCTCTTTTG CCGTCCTCTC TCTTTGTGGC
13475734 TGCTTTATCA ATTTGCGTTT GCAATCAGTG CCGACAAGGA AAGCAACAAC AACAAACAGA
13475794 AACTAGGCGC GCGATCTGCT TGCTCACACA CACATCGAAA TCAATTAAAC CCGTTGTGCGT
13475854 CATTCGGCCA CCGGTACACT TGTGCTCCA TCTCAGTCGG CTCTTGCTCT CTCTCTGGG
13475914 CAGCTCTGCT CCTCATCCTC CTCTTAATA CATGCCCTC TGCTCGCTCT CTCGCTCACC
13475974 ACGTCTTCAT TCAGCATCAT TCGATTCGAT TCGTTCTTTT CGCCCGGACC GGACACTTGA
13476034 GAGAGAGAGA ATTCTCCGGC TCGGTTTCGG ATTCGGGTTT GAGTCTGCGC TAAAGTTATA
13476094 TAGCGCGCCC ACCTTTTTC GCCGCAGTCC ATCGATCTCA GCGTAAGTGT TCGCATCGCC
13476154 ACTCGTCGCC CTCTTCCTC TTCCGATTAC GGGTCACGGT TTGCGGGTTA CGGTCTTGCG 3

```

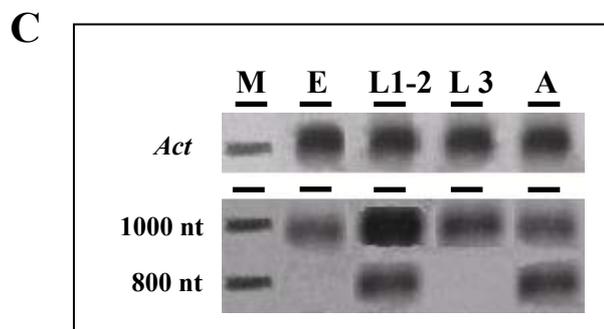


Figura 41: Organizzazione della regione genomica da cui si origina il pncRNA 1015

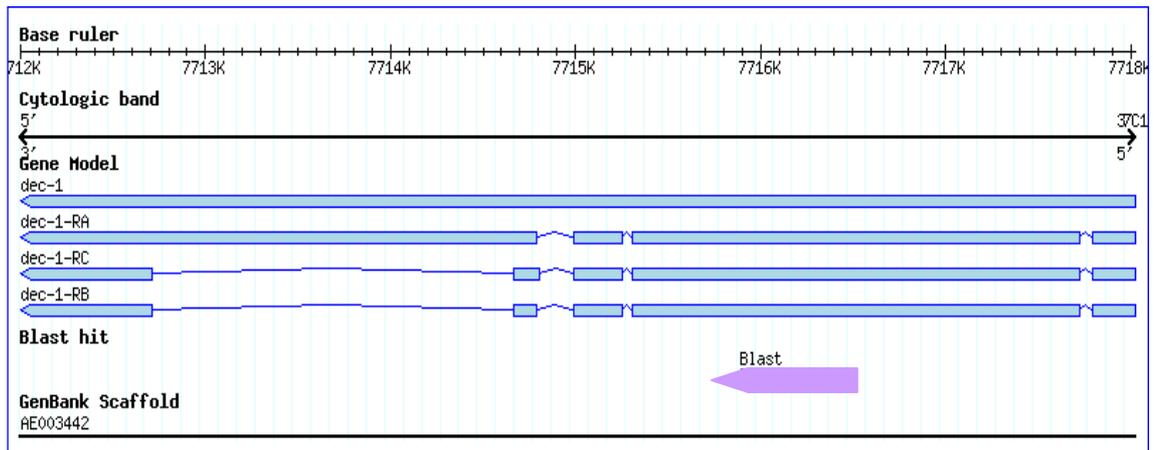
A. Regione genomica nella quale è localizzato il gene *NFAT* (*Nuclear factor of activated T cell*), in viola è mostrata la porzione corrispondente alla sonda di 522 bp amplificata mediante PCR e successivamente utilizzata nell'esperimento di Northern blot mostrato in **C**. **B.** La regione del gene *NFAT* comprendente circa 700 bp del primo introne. In giallo è evidenziata la sequenza complementare dello snRNA 1015, predetta da SNOSCAN e trascritto con polarità opposta, in verde sono evidenziati gli oligonucleotidi utilizzati per l'amplificazione tramite PCR della regione genomica successivamente utilizzata come sonda. **C.** Analisi trascrizionale del pncRNA 1015 (E:embrioni, LI-II:larve di 1° e 2° stadio; LIII: larve di 3° stadio; A: adulti; M: marker di peso molecolare, RNA ladder low range-Fermentas): in alto è riportata la normalizzazione con Actina; sotto il trascritto da 1000 nt espresso in tutti gli stadi ed in basso il trascritto da 800 nt presente solo nelle larve di 1° stadio e negli adulti.

E' interessante notare che l'omologo murino di NFAT è stato osservato essere regolato da una molecola di ncRNA la cui sequenza è conservata anche nell'Uomo. E' stata infatti identificata una molecola di RNA non coding, denominata NRON(Non coding Repressor Factor of NFAT) che dà origine a trascritti alternativi di dimensioni comprese tra gli 800 e i 3700 nt ed agisce in complesso con l'importina Beta ed altre proteine nella repressione dell'attività trascrizionale di NFAT. L'azione di questo complesso sembrerebbe regolare il traffico nucleare di NFAT impedendone l'accumulo nel nucleo e la conseguente attivazione trascrizionale, attraverso un meccanismo di interazione RNA-proteina (Willingham *et al.*, 2005). Diversamente da quanto osservato nel topo per l'RNA non coding NRON, la putativa molecola di ncRNA 1015, di *Drosophila* potrebbe agire attraverso un meccanismo antisenso sul messaggero del gene *NFAT*, formando un RNA a doppia elica lungo abbastanza da innescare il processo di silenziamento genico mediato dall'RNAi. Quest'ipotesi classificherebbe il putativo non coding RNA Dm1015 come uno dei potenziali RNA regolativi che agiscono attraverso meccanismi mediati da interazioni RNA-RNA.

Il putativo non coding RNA 1016

Il putativo non-coding RNA 1016 presenta una lunghezza di circa 500 nt ed è stato identificato mediante l'utilizzo di una sonda di 560 bp mirata inizialmente a convalidare l'espressione dello snoRNA Dm1016, identificato da SNOSCAN nel secondo esone del gene *Defective chorion-1(dec-1)* e trascritto con la stessa polarità del gene ospite (Fig. 42A).

A



B

ESONE2 gene *dec-1*

5'

```

7716958 GGATCAGACT CTGTCCAAGG AGGATATCGT CCAGATTATG GCGTATGCCT ATCGCATGGC
7716898 CAACGAACA ATGGAGAGCG AGAAGGGAA GCAGGACAAG GTTTACGCGG CCTACAGGAC
7716838 GGAACAGAAT CCAATGATGA TGCAGCAAAG ACAATGGTCG GAGGAGCAGG CCAAGATCCA
7716778 ACAGAATCAA CAGCAGATCC AGCAAAATCC AATGATGATG CAGCAACGAC AATGGTCGGA
7716718 GGAGCAGGCC AAGATCCAAC AGAATCAACA GCAGATCCAG CAAAATCCAA TGATGATGCA
7716658 GCAACGACAA TGGTCGGAGG AGCAGGCCAA GATCCAACAG AATCAACAGC AGATCCAGCA
7716598 AAATCCAATG ATGATGCAGC AAAGACAATG GTCGGAGGAG CAGGCTAAGA TCCAACAGAA
7716538 TCAACAGCAG ATCCAGCAA ATCCAATGAT GGTGCAGCAA AGACAATGGT CGGAGGAGCA
7716478 GGCCAAGATC CAACAGAATC AACAGCAGAT CCAGCAAAT CCAATGATGA TGCAGCAAAG
7716418 ACAATGGTCG GAGGAGCAGG CCAAGATTCA GCATGATCAA CAGATGGCAC AACAGATGGC
7716358 ACAGCAAGGT CTC ATGATGA CGGAGCAGAG GA AAAGGCAG TGGTCAGAAG ATCAGGCCAA
7716298 AATTCAGCAG GCTCAACAGA TGGCCCAACA GACACCCATG ATGATGCCAC AGATGCAACA
  
```

C

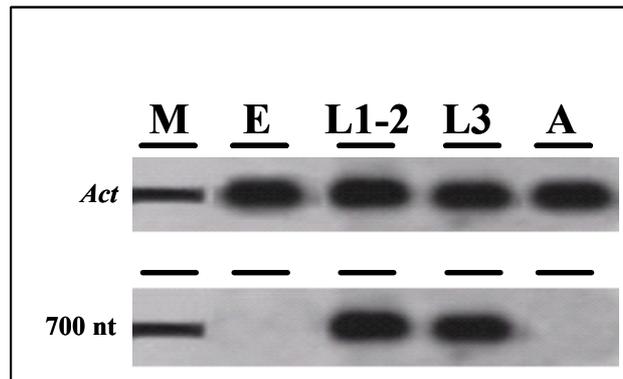
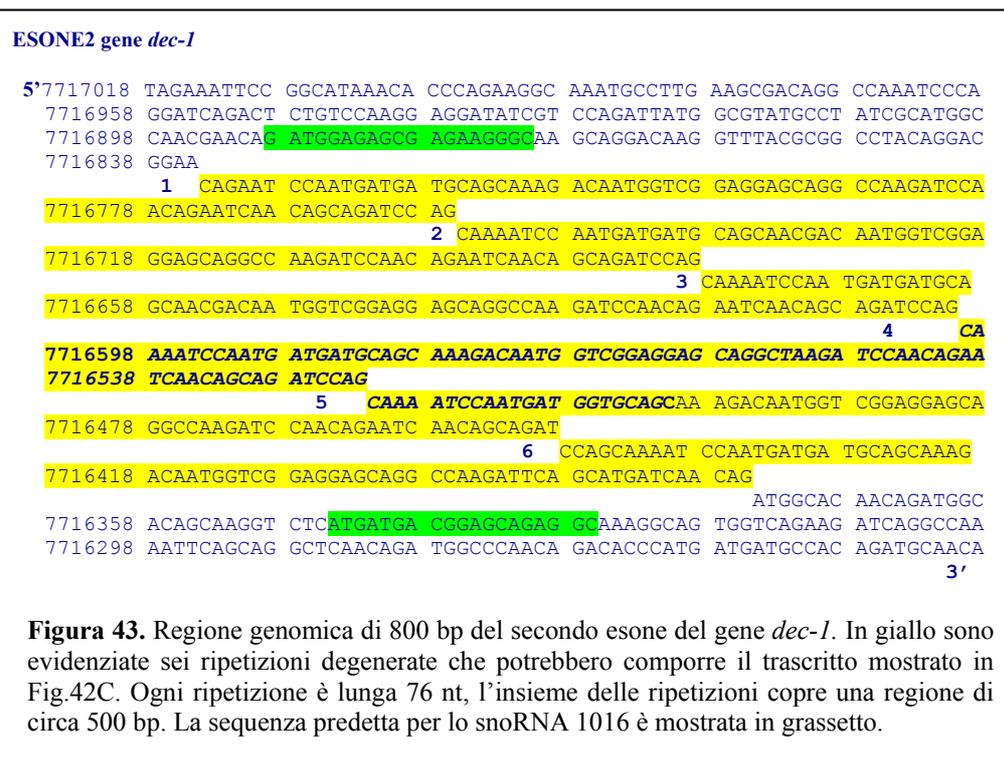


Figura 42: Organizzazione della regione genomica da cui si origina il pncRNA1016

A. Regione genomica nella quale è localizzato il gene *dec-1*: in viola è mostrata la sonda di circa 560 bp utilizzata nell'esperienza di Northern blot. **B.** Sequenza genomica di circa 700 bp comprendente il secondo esone del gene *dec-1*. In giallo è evidenziata la sequenza predetta da SNOSCAN per lo snoRNA 1016, in verde sono indicati gli oligonucleotidi utilizzati per l'amplificazione della sonda di 560 bp utilizzata nell'esperienza di Northern blot. **C.** Analisi trascrizionale del pncRNA 1016 (E:embrioni, LI-II:larve di 1° e 2° stadio; LIII: larve di 3° stadio; A: adulti; M: marker di peso molecolare, RNAladder low range-Fermentas): in alto è riportata la normalizzazione con Actina; sotto il trascritto da 700 nt espresso solo nelle larve di 1° stadio e nelle larve di 3°stadio.

Dec-1 codifica per prodotti proteici multipli generati sia da eventi di splicing alternativo che di processamento proteolitico. Le diverse isoforme proteiche si ritrovano nelle cellule follicolari, dove sono necessarie per la corretta formazione delle membrane dell'ovulo. Le mutazioni che cadono in questo gene alterano la formazione del chorion e della membrana vitellina, impedendo la corretta oogenesi (Badciong *et al.*, 1998). Il trascritto di circa 500 nt rilevato dall'esperienza di Northern blot (fig.42C) non corrisponde a nessuno dei trascritti noti del gene ospite, tutti notevolmente più lunghi (dec-1 RA 5699 nt; dec-1 RB 3740 nt; e dec-1 RC 3754 nt). Il pncRNA 1016 sembra poi essere regolato nello sviluppo, in quanto è assente in adulti ed embrioni ma è molto abbondante in larve del 1° e 3° instar. Analizzando le sequenze fiancheggianti del putativo snoRNA 1016 ho osservato che la sequenza predetta da SNOSCAN fa parte in realtà di una regione composta da 6 ripetizioni in tandem (458bp; Fig.43) che potrebbe comporre la struttura del pncRNA 1016.

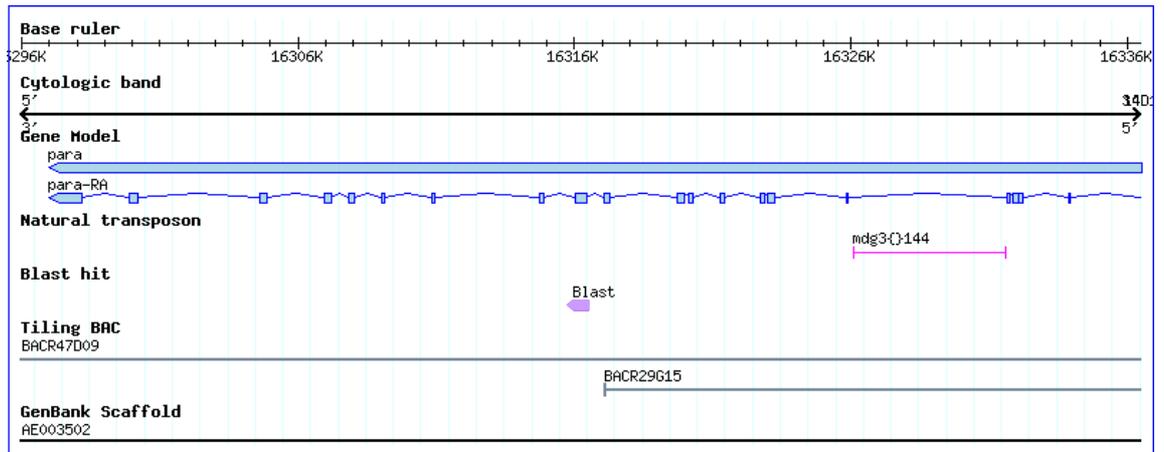


Si potrebbe inoltre ipotizzare per questo trascritto una funzione regolativa sul gene *dec-1*, dato che il pncRNA1016 pare espresso solo negli stadi in cui il gene ospite non si esprime, ovvero in embrioni, larve di 1° instar e larve di 3° instar, (stadi in cui le cellule follicolari non sono ancora differenziate), mentre non si esprime in adulti, dove invece *dec-1* è specificamente espresso (nelle femmine adulte durante l'oogenesi).

Il putativo non coding RNA 1019

Il putativo non-coding RNA 1019, identificato nell'esperimento di Northern blot volto a confermare l'espressione dello snoRNA 1019, deriva dall'esone 23, costitutivamente espresso, del gene *paralytic (para)* il quale è descritto produrre un unico trascritto di 6848 nt (Fig. 44). Il gene *para* codifica per una proteina integrale di membrana che costituisce un canale per gli ioni Na⁺ voltaggio dipendente ed è coinvolto in vari processi della trasmissione nervosa, come la trasmissione sinaptica e la contrazione muscolare, ma anche in processi comportamentali, come il corteggiamento dei maschi di *Drosophila* (Amichot *et al.*, 2003; Anholt *et al.*, 2001). Il putativo ncRNA 1019 ha una lunghezza di circa 1000 nt, e come si può vedere nel pannello di ibridazione mostrato in figura 44C, esso è presente in tutti gli stadi tranne che in adulti e risulta più abbondante nelle larve di 1° stadio.

A



B

gene *para*

```

16317072 TGTATAACTA TAACTATAAC TATACAGATG ATTGGCAACT CAATTAACCA CCAAGACAAT
16317012 AGACTGGAAC ATGAGCTAAA CATAGAGGT TTGTCCTTAC AGGACGACGA CACTGCCAGC
16316952 ATTATTCAT ATGGTAGCCA TAAGATCGA CCATTCRAGG ACGAGAGCCA CAAGGGCAGC
16316892 GCCGAGACGA TGGAGGGCGA GGGGAGCGC GACGCCAGCA AGGAGGATTT AGGTCTCGAC
16316832 GAGGAACTGG ACGAGGAGGG CGAATGCGAG GAGGGCCCGC TCGACGGTGA TATCATTTT
16316772 CATGCAACAG ACGAGGATAT ACTCGATGAA TATCCAGCTG ATTGCTGCC C GATTCGTAC
16316712 TATAAGAAAT TTCCGATCTT AGCC
                                     5' GGTGAC GATGACTCGC CGTTCTGGCA AGGATGGGGC
16316652 AATTTACGAC TGAAAC TTT TCAATTAATT GAAATTAAT ATTTTGAAC AGCTGTATC
16316592 ACTATGATTT TRATGAGTAG CTTAG 3'
                                     CTTTG GTAAGTCAAC TACAAAATTA AATTAAATTA
16316532 AATTTATCGA GAAAAAGAAA AAACGTTTAA AAAAAAATA ATATTAAATC AAAAGACTCT
16316472 AAAACAGACT CAAAGACTCT GTTTGCCATT TAGATTGTTG CAAATTTGTG TCATTTGCTT
16316412 TTTTAAAATG CTAAAACGC AAATGTAGC TTGCAACA GA TCACTAAATT TAAACCAAT
16316352 TAAATCAATT CAATCAATAA GCATTCAAA TAAACAAA C AATTGAATAT GTAGTTAATC

```

C

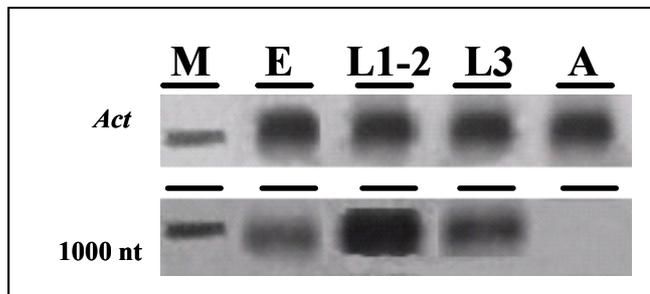


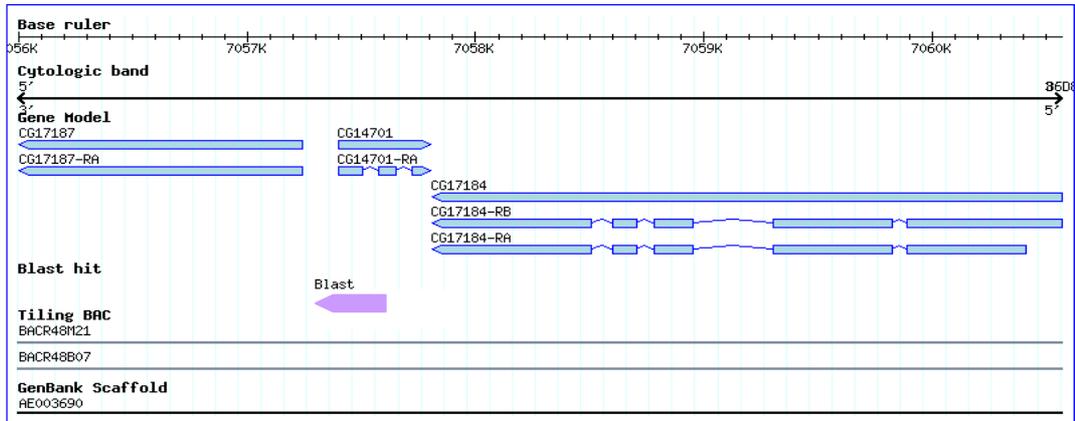
Figura 44: Organizzazione della regione genomica da cui si origina il pncRNA 1019

A. Regione genomica nella quale è localizzato il gene paralytic (*para*), in viola è mostrata la regione corrispondente alla sonda di circa 560 bp utilizzata nell'esperimento di Northern mostrato in C. La sonda è localizzata nell'esone 23, costitutivamente espresso del gene. B. Sequenza genomica comprendente parte dell'introne 22, tutto l'esone 23 e parte dell'introne 23 del gene *para*. In grassetto sono evidenziati gli esoni, in giallo è evidenziata la sequenza del putativo snoRNA 1019 predetta da SNOSCAN, in verde gli olinucleotidi utilizzati per l'amplificazione della sonda di 500 bp utilizzata nell'esperimento di Northern blot. C. Analisi trascrizionale del pnc 1019 (E: embrioni, LI-II: larve di 1° e 2° stadio; LIII: larve di 3° stadio; A: adulti; M: marker di peso molecolare, RNA ladder low range-Fermentas), in alto è riportata la normalizzazione con Actina ed in basso è indicato il trascritto da 1000 nt che è espresso in tutti gli stadii tranne che in adulti ed è più abbondante negli stadi larvali LI-II.

Il putativo non coding RNA 1020

Il putativo non-coding RNA Dm1020, di circa 500 nt, è espresso soprattutto in larve del 1°-2° instar ed in adulti (Fig. 45). Esso è stato identificato mediante ibridazione della sonda di 516 bp (Fig.45A) volta a confermare l'espressione dello snoRNA 1020. Il putativo snoRNA 1020, lungo 160 nt, è stato identificato da SNOSCAN in una regione che comprende parte del primo esone, il primo introne e parte del secondo esone del gene CG14701, quindi a cavallo delle giunzioni esone 1/introne 1 e introne1/esone2. Il pncRNA è trascritto con polarità opposta rispetto al gene CG14701, il quale ha funzione ignota ed è descritto produrre un solo trascritto lungo 241 nt (Fig.45C) che codifica una putativa proteina con un dominio zinc-finger di legame al DNA. Come si può osservare nel pannello di ibridazione in figura 45C, l'espressione di questo trascritto è modulata nello sviluppo.

A



B

gene *CG14701*

```

5' cgcaaca caa cgtgccatt tgtttggcgt ggattaatct aagtatcgc aaaaaaact
145681 caaatataat atgaccttta ggcgctgaat taagctgaag gtacgtcatg agcatctatc
145741 acgacgaggt ggagatcgag gacttcgagt acgacgagga ggaggagatg tactactatc
145801 cctgtccatg cggcgcacga ttccagatct ccaaggtaac gagtctgata cgcaaccttt
145861 ttttttctgg actcccata tcaacttgta aattccact ttaggaggag ctaatcgagg
145921 gcgaggaggt ggccacctgt cccagctgct ccctagtcac caaggtcata tacgatccgg
145981 taagtgactc attcagaaat acacctagaa ctcccagaa aactcaagca atctgatttt
146041 ccgaatttca ggagatgttc aaagctgagg aggatgaaga aagtgcgctg aacgagaagc
146101 tcggcgacctgaagctcgag agaactaac cagtttg gttaattaa 3'
  
```

C

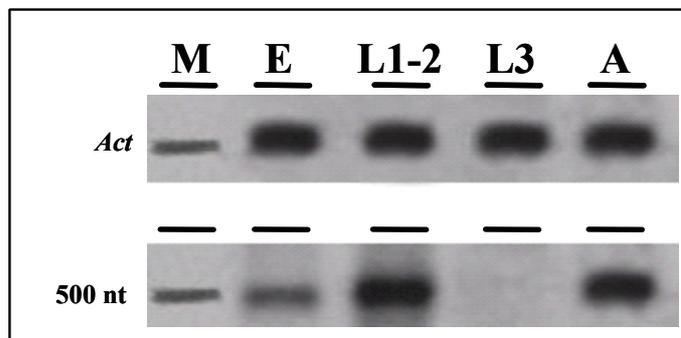


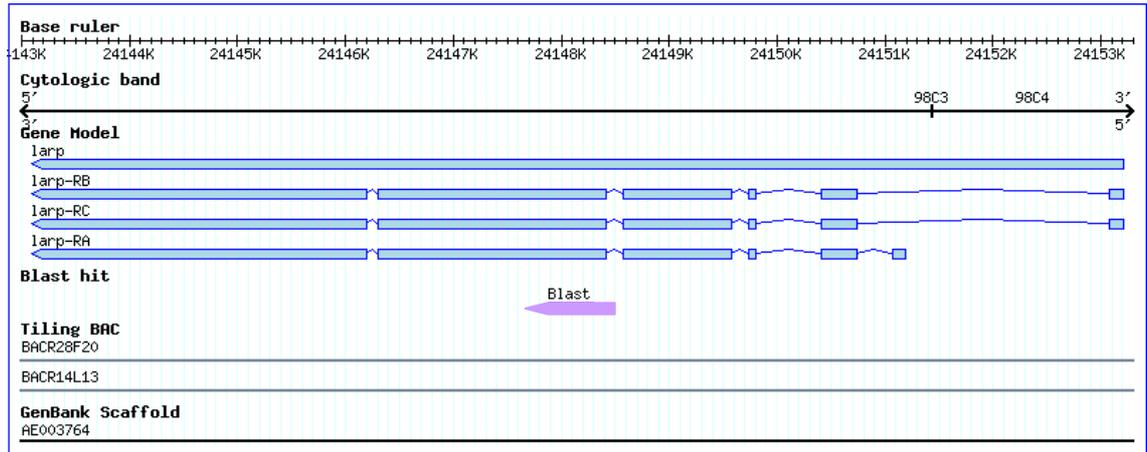
Figura 45: Organizzazione della regione genomica da cui si origina il pncRNA 1020

A. Regione genomica nella quale è localizzato il gene ospite CG14701, in viola è mostrata la sonda genomica di 516 bp utilizzata nell'esperimento di Northern blot mostrato in C. **B.** Sequenza genomica completa del gene CG14701, lungo 402 bp. In grassetto sono evidenziati gli esoni. In giallo è evidenziata la sequenza complementare del putativo snoRNA 1020 predetta da SNOSCAN. In verde gli oligonucleotidi utilizzati per l'amplificazione di una sonda di 500 bp utilizzata nell'esperimento di Northern blot. La sonda copre tutto il gene, una porzione della regione intergenica e poche basi al 3' del CG17184. **C.** Analisi trascrizionale del pncRNA 1020 sviluppo (E:embrioni, LI-II:larve di 1° e 2° stadio; LIII: larve di 3° stadio; A: adulti; M: marker di peso molecolare, RNAladder low range-Fermentas), in alto è riportata la normalizzazione con Actina ed in basso è indicato il trascritto da 500 nt che è espresso in tutti gli stadi tranne che nelle larve di 3° stadio.

Il putativo non coding RNA 1028

Il putativo non-coding RNA 1028 ha una lunghezza di circa 700 nt ed è prevalentemente espresso negli embrioni; il trascritto è stato identificato nell'esperimento di Northern blot volto a confermare l'espressione dello snoRNA Dm1028. Il pncRNA 1028 si origina dall'esone 5, costitutivamente espresso, del gene *Larp* (*La related protein*). Questo gene produce 3 trascritti: Larp-RA, Larp-RB e Larp-RC, rispettivamente di 6673 nt, 6749 nt e 6749 nt (Fig.46A). Le isoforme proteiche codificate dal *Larp* possiedono un dominio simile alla proteina La che agisce frequentemente da autoantigene nel Lupus eritematoso sistemico. Nel nucleo, La agisce come fattore di trascrizione specifico della RNA pol III, dove lega la regione 3' UTR dei trascritti nascenti assistendo il folding e la maturazione (Alfano *et al.*, 2004), mentre nel citoplasma La riconosce la classe degli mRNA 5'-TOP (Keene, 2003). Il gene *Larp*, ancora poco caratterizzato, sembra anche essere coinvolto nel meccanismo di morte cellulare programmata di tipo II (mediata da autofagia), un meccanismo di distruzione della cellula operato dal suo stesso contenuto lisosomiale, che si realizza attraverso la formazione di numerosi vacuoli autofagici intracellulari (Gorski *et al.*, 2003). In figura 46A è mostrata l'organizzazione genomica e la sequenza della sonda utilizzata nell'esperimento di Northern blot, nonché il profilo di espressione osservato per pncRNA 1028, che sembra essere espresso solo negli stadi embrionali (Fig.46C).

A



B

Esone 6 gene *Larp*

```

24149205 ATACTATTTT AGCGTCGATA ATCTTACCGG CGACTTCTTT TTGCGCCGCA AAATGGACCC

24149145 CGAGGGCTAT ATCCAGTTA CTCTGATCGC CTCGTTCCAT CGCGTTCTCG CACTGAC CAC
24149085 GATGTGGCT GTGATC TAA ACGCCATTAA GGAATCGGAT AAACCTGGAGC TCTTTGAGGG
24149025 CTATAAGGTG CGCACCAAGA CGACACCGAC CACCTGGCCC ATCACGGAAG TGCCGGAGGT
24148965 CAACGAAGGG GAACCGAAGG CTATTGGCAC ACTGGAGCAG GAGCAGT
                                         5' TGG AGCAGAATGA
24148905 TGGTCAGGAA AAGCTGGAGG AGCAGACTGA GGCGGATTCC CCGCCGCCCA TTCTCACATC
24148845 GGCGATGGCC ACAAAGCCGC TGAACAGCAT T 3'
                                         CCGCCGCCG CCAATGCCCG GAAACCCCCA
24148785 GAATCTGGTG CCCAAATGT TGCAGGATAA GCAACAGTCG AGATCCTCGA CCATTGCGGC
24148725 TCTAAACTCT GTGAATGCGA TGATGCCCT GACCCAGCAA GTGGAGGGTG GTGCCCGTGA
24148665 GCTGGCTGGA CATTGAGTG GACTGGCTGA GAGTGTGAAG CCCAAGTCGA CATCGACGCC
24148605 AGACAAGAGA AATGCTGCTT CAGCGGGCAA TGGAGCTGGG TCAGCAGCTG CTTTGGTGGC
  
```

C

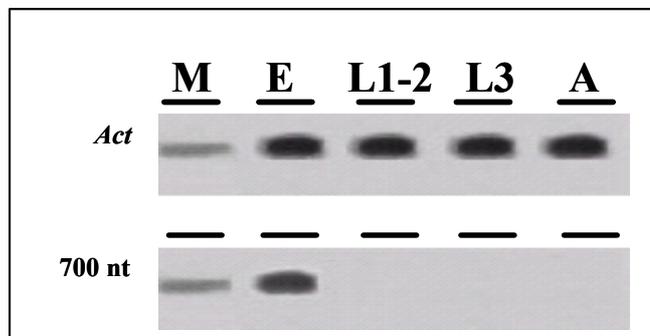


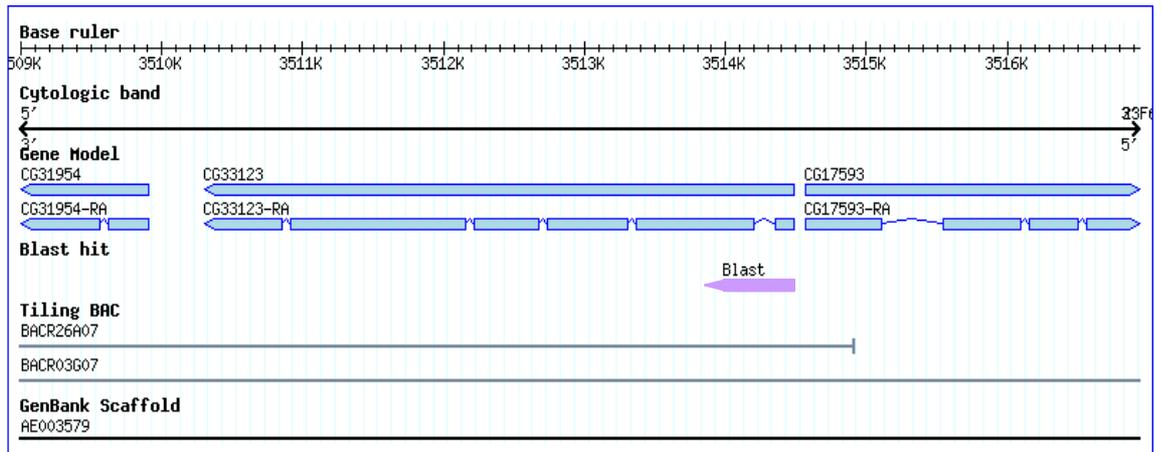
Figura 46: Organizzazione della regione genomica da cui si origina il pncRNA 1028

A. Il gene *La related protein (Larp)* è localizzato sul braccio destro del cromosoma 3(3R) e produce 3 trascritti: Larp-RA, Larp-RB e Larp-RC, rispettivamente di 6673 nt, 6749 nt e 6749 nt. In viola è mostrata la sequenza corrispondente alla sonda utilizzata nell'esperimento di Northern blot mostrato in C. La sonda è localizzata nel quinto esone costitutivo del gene *Larp*. B. Sequenza genomica di circa 700 bp del quinto esone del gene *Larp*. In giallo è evidenziata la sequenza del putativo snoRNA 1028 predetta da SNOSCAN, in verde gli olinucleotidi utilizzati per l'amplificazione della sonda di 403 nt utilizzata nell'esperimento di Northern blot. C. Analisi trascrizionale del pncRNA 1028 (E:embrioni, LI-II:larve di 1° e 2° stadio; LIII: larve di 3° stadio; A: adulti; M: marker di peso molecolare, RNA ladder low range-Fermentas), in alto è riportata la normalizzazione con Actina ed in basso è indicato il trascritto da 700 nt che è espresso solo negli stadi embrionali.

Il putativo non coding RNA 1049

Il putativo non-coding RNA 1049 presenta una lunghezza di 1000 nt ed è stato identificato mediante l'utilizzo della sonda, lunga 560 bp, mirata inizialmente alla validazione del putativo snoRNA 1049. Il putativo snoRNA 1049, lungo 100 nt è stato identificato da SNOSCAN a cavallo della giunzione esone 1/introne 1 del CG33123, trascritto con medesimo orientamento (Fig. 47). Il gene ospite è descritto produrre un unico trascritto di 3791 nt il cui prodotto proteico è previsto avere una localizzazione citoplasmatica ed una putativa attività perossidasi, nonché altre funzioni attribuite in base alla similarità di sequenza, quali l'attività amminoacil-tRNA sintetasica. La reale funzione genica della ORF CG33123 non è stata ancora determinata, tuttavia la relativa proteina sembra coinvolta nella risposta allo stress ossidativo. (<http://www.flybase.org>). Il putativo non coding RNA identificato non corrisponde a nessun trascritto del gene riportato ad oggi in banca dati (www.flybase.org) e risulta espresso esclusivamente in adulti, come si può osservare nel pannello di ibridazione mostrato in figura 47 C.

A



CG33123

1 CCGCCTATTG **TGCTGCCACA CTGGCCGCAA** CACGTTGCAC TGCTCGAAAC ATATTAAGTT
 61 TTTTGCTAGT GCCACGATTT **GAATGAAGAA** ATTGTTCTAA ACAATTACGA ATATCATTGA
 121 G
 5' **AACTAAGCT ATGGTGAGCA** TTGGCCCGGG AATGGGAATC **CCCATAAAT** CGATGAGTCA
 181 CAGGAACTGC CACGAGGACG **AGCCGGCTGA** TGCAATTGGC **CG** 3'
 CGGCATTT CTGCATTTGA
 241 AAGGCAATTA ATATGCTAAT AAGTTTACTT TTTTACCAT AACAGGCCAC **AACCGAGCGC**
 301 **AAGGGCACCT** TTAAGGTGGA **GTATCTACAG** AAGATCGAGC GGGAGGTGCA GCAGCGATGG
 361 GAGGCGGAGC GAGTGCACGA GTCGGACGCG CCGACGGCCC CAAAGAAGCG GCAGGCAGAG
 421 AAGTTCTTCG **TGACCTTTC** **GTTCCCTAC** ATGAACGGAC GCCTCCACCT GGGCCACACC
 481 **TTCTCGTTGT** **CAAAGCCGGA** **GTACTCGATG** CGCTACCATC GCCTGAAGGG ACGGCGCGTA
 541 CTCTGGCCCT TCGGATTCCA CTGCACCGGC ATGCCCATCA AGGCTTGTGC GGACAAGTTA
 601 ACGCGCGAAC TGGAGCAGTT CGGCTTTC A CACAGTTCC CCGAGACCGA GGAAGTGGTC

B

C

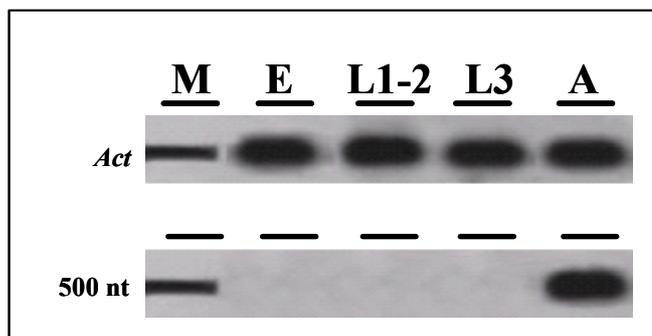


Figura 47: Organizzazione della regione genomica da cui si origina il pncRNA 1049

A. IL CG33123 è localizzato sul braccio sinistro del cromosoma 2 (2L) e codifica per un unico trascritto di 3791 nt. In viola è mostrata la sequenza corrispondente alla sonda utilizzata nell'esperimento di Northern blot che ha portato all'identificazione del putativo ncRNA 1049 mostrato in C. **B.** E' mostrata la sequenza genomica comprendente il primo esone, il primo introne e circa 400 bp del secondo esone del gene CG33123. Gli esoni sono evidenziati in grassetto. In giallo è evidenziata la sequenza del putativo snoRNA 1049 predetta da SNOSCAN, in verde gli olinucleotidi utilizzati per l'amplificazione della sonda di circa 500 bp. **C.** Analisi trascrizionale del pncRNA 1049 (E:embrioni, LI-II:larve di 1° e 2° stadio; LIII: larve di 3° stadio; A: adulti; M: marker di peso molecolare, RNA ladder low range-Fermentas), in alto è riportata la normalizzazione con Actina ed in basso è indicato il trascritto da 500 nt che è espresso solo negli adulti.

2.6 Ricerca di nuovi putativi target

Come ampiamente discusso nei paragrafi precedenti, la maggior parte dei 52 snoRNA identificati nel corso di questo lavoro presenta la sequenza guida antisense posizionata a monte della box D', ma in 9 casi (DmU18adj, DmG1322, DmSnR69b, Dm3112a, Dm1008a, Dm1011, Dm1012 e Dm1047) tale sequenza manca di significativa complementarità con le molecole di rRNA 18S, 28S o 5.8S. Poiché anche la box D manca di significativa complementarità con l'rRNA, questi nuovi snoRNA appartengono quindi alla categoria dei cosiddetti "orfani", per i quali restava ancora da definire una molecola bersaglio. A questo proposito è importante ricordare che recentemente è stato riportato in letteratura un esempio di snoRNA "orfano", lo snoRNA di mammifero MBII-52, che è in grado di modificare come molecola bersaglio una molecola di RNA messaggero, cioè l'mRNA del recettore della serotonina 5-HT_{2C} (Vitali *et al.*, 2005). E' quindi possibile ipotizzare che specifiche molecole di RNA messaggero possano rappresentare i bersagli degli snoRNA "orfani". Partendo da questi presupposti ho condotto, tramite BLAST, una ricerca sull'intero trascrittoma di *Drosophila* andando ad allineare le sequenze antisense degli snoRNA "orfani" (lunga 13 nt) con tutti i trascritti (EST, cDNA) presenti nella banca dati (<http://www.flybase.org/blast>; BLAST database: alla sezione Annotated genes: NT). Questa ricerca ha portato risultati interessanti, dato che ben 41 mRNA possono rappresentare putativi target per 7 (DmU18adj, DmG1322, Dm3112a, Dm1008a, Dm1011, Dm1012 e Dm1047) dei nuovi snoRNA da me identificati e classificabili come "orfani". Nella tabella 10 sono riportate le caratteristiche generali

Tabella 10 : SnoRNA la cui sequenza antisenso è complementare a trascritti di geni che codificano per proteine (mRNA)

SnoRNA	Putativo gene target	Localizzazione sequenza target	Appaiamento Seq.Antisenso/Seq.target
Dm_28S.002- DmU18adj	<i>Polychaetoid</i> (CG17782) (Function:JNK cascade)	Esone 1	D' box 3'- UCAG cugugccuuuuc -5' 5' - gacacggaaaag- 3'-mRNA
	CG11306	Esone 4	D' box 3'- UCAG cugugccuuuuc -5' 5' - gacacggaaaag- 3'-mRNA
	CG4203	Esone 1	D' box 3'- UCAG cugugccuuuuc -5' 5' - gacacggaaaag- 3'-mRNA
	CG31226	Esone 2	D' box 3'- UCAG cugugccuuuuc -5' 5' - gacacggaaaag- 3'-mRNA
	CG30096	Esone 3	D' box 3'- UCAG cugugccuuuuc -5' 5' - gacacggaaaag- 3'-mRNA
	CG8709	Esone 6	D' box 3'- UCAG cugugccuuuuc -5' 5' - gacacggaaaag- 3'-mRNA

Segue da Tabella 10

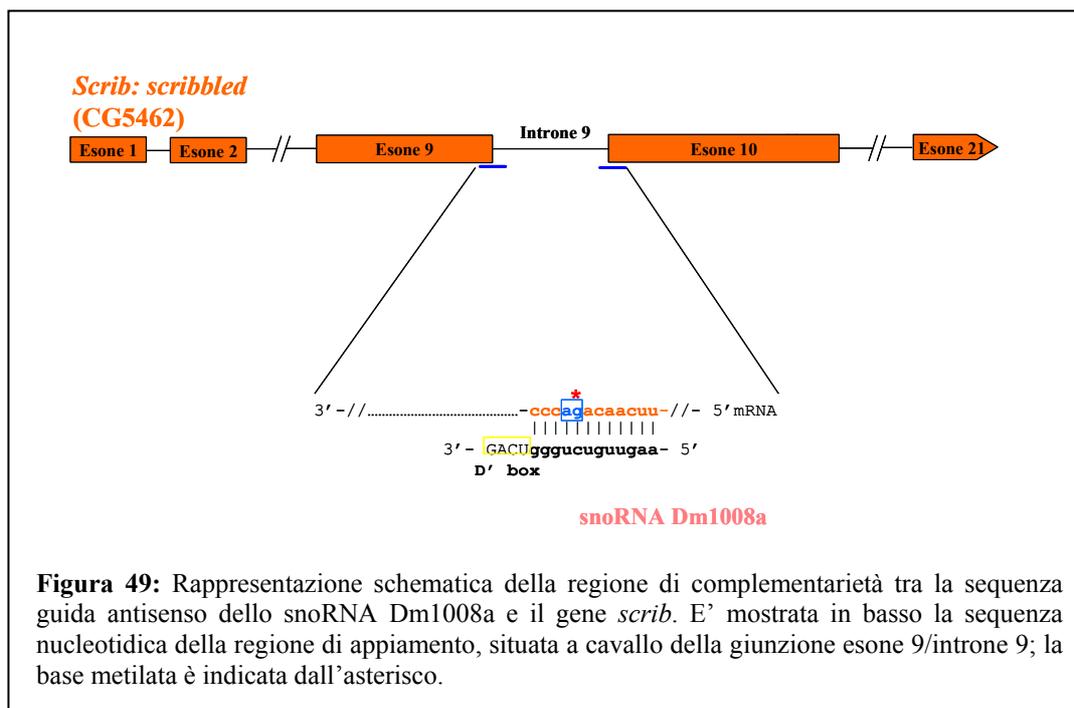
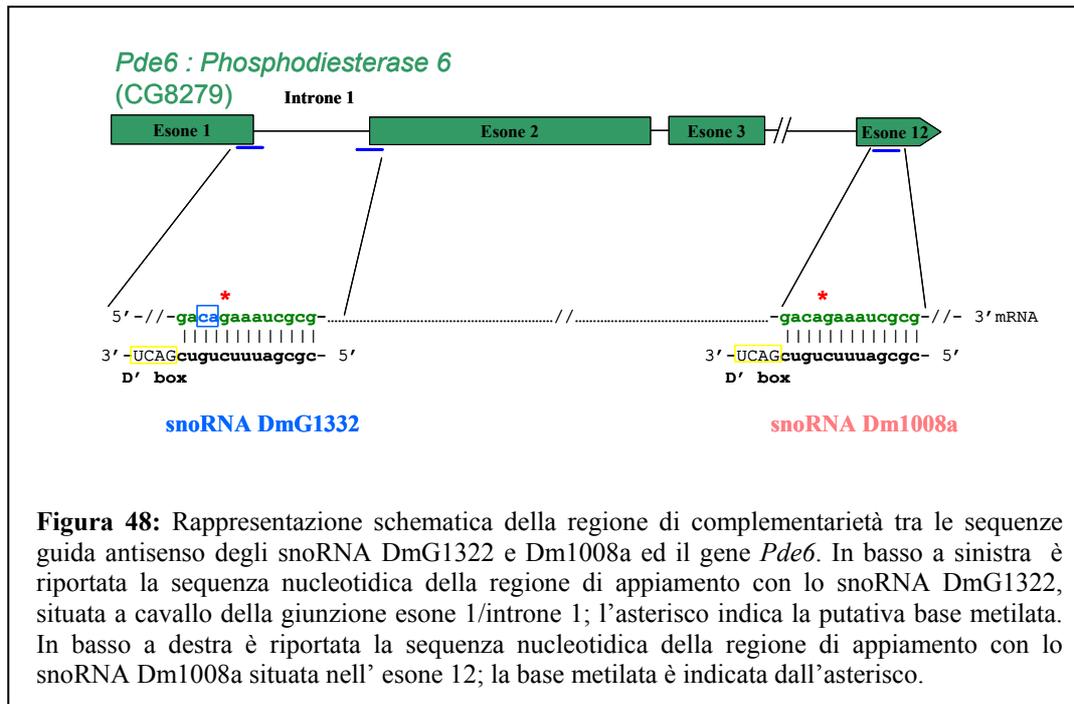
Dm_28S.043A- Dm3112a	CG7671	Estone 1	<p>D' box</p> <p>3'- UCAG cgcgauuucugcu -5'</p> <p> </p> <p>5' - gcgc<u>u</u>aaagacga- 3'mRNA</p>
	CG7309	Estone 3	<p>D' box</p> <p>3'- UCAG cgcgauuucugcu -5'</p> <p> </p> <p>5' - gcgc<u>u</u>aaagacga- 3'mRNA</p>
	CG13957	Estone 5	<p>D' box</p> <p>3'- UCAG cgcgauuucugcu -5'</p> <p> </p> <p>5' - gcgc<u>u</u>aaagacga- 3'mRNA</p>
	CG14116	Estone 6	<p>D' box</p> <p>3'- UCAG cgcgauuucugcu -5'</p> <p> </p> <p>5' - gcgc<u>u</u>aaagacga- 3'mRNA</p>
	CG31510	Estone 2	<p>D' box</p> <p>3'- UCAG cgcgauuucugcu -5'</p> <p> </p> <p>5' - gcgc<u>u</u>aaagacga- 3'mRNA</p>

Segue da Tabella 10

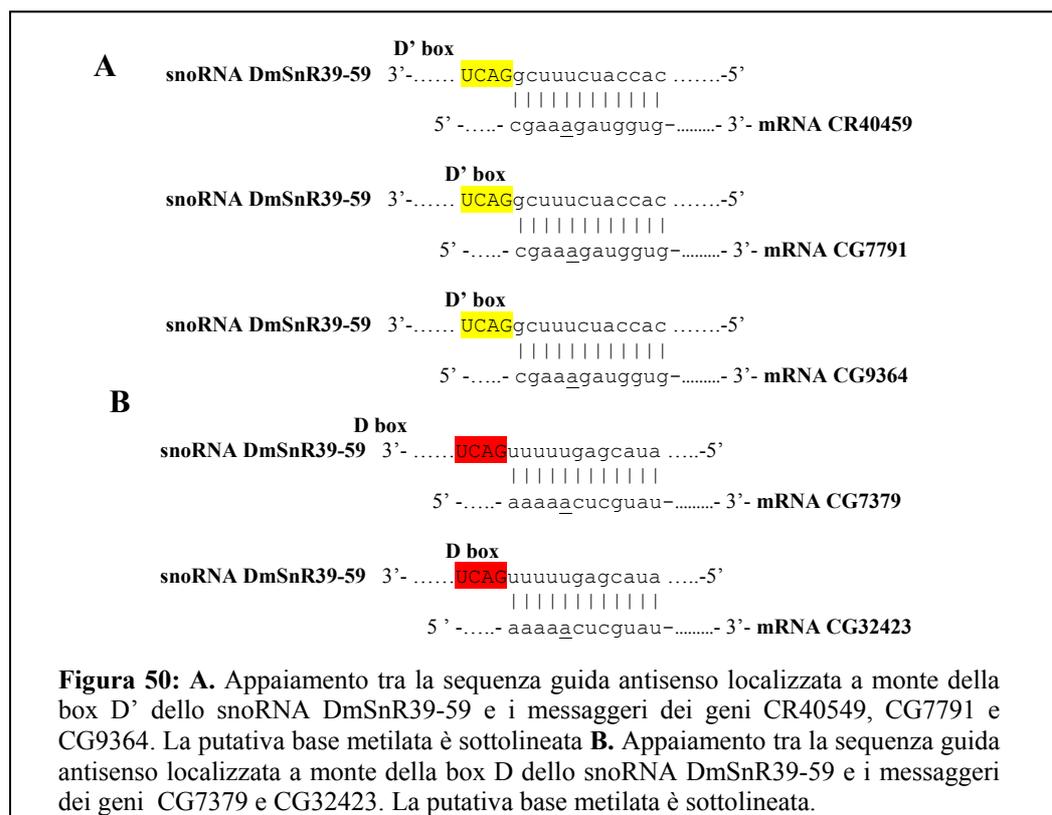
Dm_28S.1047	<i>Hexo1</i> (CG1318) (Hexosaminidase 1, function: polysacccharide metbolism)	Estone 4	D' box 3'- UCAG uaacugcuc <u>au</u> agg -5' 5' - auug <u>a</u> cgag <u>ua</u> ucc- 3'mRNA
	<i>Nep5</i> (CG6265) (Neprylisin 5, function: proteolysis and peptidolysis)	Estone 6	D' box 3'- UCAG uaacugcuc <u>au</u> agg -5' 5' - auug <u>a</u> cgag <u>ua</u> ucc- 3'mRNA
	<i>Kz</i> (CG3228) (Function: nuclear mRNA splicing)	Estone 3	D' box 3'- UCAG uaacugcuc <u>au</u> agg -5' 5' - auug <u>a</u> cgag <u>ua</u> ucc- 3'mRNA
	CG31852	Estone 2	D' box 3'- UCAG uaacugcuc <u>au</u> agg -5' 5' - auug <u>a</u> cgag <u>ua</u> ucc- 3'mRNA

dei geni i cui trascritti rappresentano putativi target, la localizzazione precisa della putativa sequenza bersaglio, nonché la regione di appaiamento tra la sequenza guida dello snoRNA e quella bersaglio. Come mostrato in tabella, la maggior parte dei trascritti identificati come putativi target codificano per proteine con attività enzimatica. Discuterò solo, come esempio, uno dei dati più interessanti, che riguarda gli snoRNA DmG1322 e Dm1008a, le cui sequenze guida sono risultate potenzialmente capaci di modificare nucleotidi immediatamente adiacenti a siti di giunzione esone/introne dei putativi geni target. Infatti, lo snoRNA DmG1322 riconosce come putativo target il trascritto del gene *Pde6* (*Phosphodiesterase 6*; CG8279) la base immediatamente adiacente al sito di splicing tra l'esone 1 e l'introne 1 dell'mRNA, mentre lo snoRNA Dm1008a ha come putativo target il trascritto del gene *scrib* (*scribbled*; CG5462), ed è potenzialmente capace di modificare una base coinvolta nel sito di splicing tra l'esone 9 e l'introne 9 (Fig.48 e 49). Inoltre, la sequenza antisenso dello snoRNA Dm1008a è risultata anch'essa complementare al trascritto del gene *Pde6*, anche se il tratto di complementarità è localizzato più a valle rispetto a quello che si appaia con lo snoRNA DmG1332, ed esattamente all'interno dell'esone 12 (Fig.48). Quindi, l'mRNA del gene *Pde6* presenta due potenziali siti di metilazione in due punti diversi della molecola che possono essere bersaglio di due differenti snoRNA "orfani". Di tutti gli snoRNA "orfani" ho anche saggiato la potenzialità di appaiamento della seconda sequenza antisenso posizionata a monte della box D e la ricerca ha avuto esito positivo per 6/9 snoRNA "orfani" (DmU18adj, DmG1322, Dm1011, Dm1047, Dm1008a e DmSnR69b),

portando ad identificare almeno 12 putativi mRNA target (dati non mostrati). I dati ottenuti suggeriscono quindi che gli snoRNA inizialmente definiti orfani possono in realtà essere delle molecole che svolgono la loro funzione regolativa sulle molecole di pre-mRNA piuttosto che quelle di rRNA.



Questo dato apre enormi prospettive allo studio del ruolo biologico svolto da queste molecole. Poiché è possibile che anche gli snoRNA che mostrano perfetto appaiamento con l'rRNA possano riconoscere come target addizionali altre molecole di RNA mediante appaiamento con la seconda sequenza antisenso, o addirittura con la stessa sequenza guida che si appaia all'rRNA, ho effettuato una ricerca analoga per tutti i nuovi geni identificati. Dati interessanti, benché preliminari, sono emersi da tale analisi: infatti, alcuni snoRNA capaci di appaiarsi con l'rRNA sono in grado di appaiarsi utilizzando la stessa sequenza guida anche con messaggeri che ne rappresentano quindi addizionali target; in più, ulteriori messaggeri vengono riconosciuti come putativi target anche dalla seconda sequenza antisenso “orfana”. Un esempio fra tutti è quello rappresentato dallo snoRNA DmSnR39-59, la cui sequenza antisenso, localizzata a monte della box D' non è perfettamente complementare all'rRNA ma è risultata invece complementare per un tratto di 12 nt ai trascritti di ben tre geni diversi: CR40459, CG7791 e CG9364. In aggiunta, anche la sequenza antisenso localizzata a monte della box D di questo snoRNA ha mostrato complementarità perfetta con i trascritti di 2 geni, CG7379 e CG32423 (Fig.50). Questi risultati sostengono con forza l'ipotesi di una multifunzionalità delle molecole di snoRNA. Inoltre l'osservazione che per alcuni putativi trascritti target la base selezionata per la metilazione è una A suggerisce l'ipotesi che in alcuni casi le modificazioni operate dagli snoRNA sui messaggeri possano modulare l'efficienza del processo di editing operato dagli enzimi ADAR, considerato che la metilazione del ribosio diminuisce di molto l'efficienza del processo di deamminazione operato da questi enzimi (Vitali *et*



al., 2005). Inoltre recenti dati dell'esistenza di una relazione tra l'efficienza del processo di editing e la scelta dei siti alternativi di splicing (Agrawal and Stormo, 2005) sostengono ulteriormente l'ipotesi di un significato funzionale molto più complesso di quelli finora immaginato per le molecole di snoRNA, che io presumo possa essere pienamente elucidato in un prossimo futuro.

3. DISCUSSIONE

Il numero di molecole di snoRNA identificate in diversi organismi eucariotici è cresciuto notevolmente negli ultimi anni grazie all'utilizzo, spesso combinato, di approcci sperimentali basati sulla costruzione e l'analisi di genoteche di cDNA ed approcci computazionali, basati sull'utilizzo di algoritmi di ricerca specifici per queste molecole. Tuttavia, l'eterogeneità delle caratteristiche mostrate finora da questi geni nei diversi organismi lascia supporre che il numero di quelle finora identificate possa rappresentare soltanto la punta di un iceberg la cui reale dimensione deve ancora essere determinata. Lo scopo del presente lavoro è stato quello di condurre un'analisi genomica tesa ad identificare nuovi geni che codificano per piccoli RNA nucleolari in *Drosophila melanogaster* attraverso un approccio di RNomica computazionale. Tale analisi è stata condotta utilizzando lo SNOSCAN, un efficiente algoritmo già sperimentato con successo in lievito (Lowe and Eddy, 1999) e specifico per l'identificazione degli snoRNA appartenenti alla classe C/D. La ricerca operata dallo SNOSCAN viene condotta a partire dalle sequenze delle putative molecole di RNA bersaglio in corrispondenza dei siti di metilazione presenti; essa si basa sull'analisi sequenziale delle caratteristiche di struttura specifiche degli snoRNA ed assegna a ciascun candidato uno "score" calcolato rispetto ad un modello probabilistico. I putativi siti di metilazione di *D.melanogaster*, non ancora determinati sperimentalmente, sono stati ottenuti mediante allineamento delle sequenze degli rRNA di lievito, *Drosophila* e Uomo e sono stati poi forniti, come parametri iniziali, al programma SNOSCAN per identificare nel genoma di *Drosophila* le sequenze degli snoRNA potenzialmente capaci di modificare tali siti. Quest' approccio ha portato all'identificazione di 100 putativi geni per

snoRNA della classe C/D che superavano il valore soglia di attendibilità assegnato dal programma SNOSCAN (score>20), tutti presumibilmente coinvolti nella metilazione degli RNA ribosomali 18S e 28S. Al contrario, l'analisi non ha portato ad identificare alcun gene presumibilmente coinvolto nella metilazione dello rRNA 5.8S. L'analisi trascrizionale ha successivamente confermato l'espressione di 36 geni codificanti per snoRNA, di cui 26 emersi dall'allineamento delle sequenze dei siti di metilazione di lievito con l'rRNA *Drosophila* e 10 dall'allineamento tra i siti di metilazione di Uomo e l'rRNA di *Drosophila*. La maggior parte dei nuovi geni ha mostrato un profilo di espressione costitutivo; tuttavia, per 5 snoRNA (DmU18, DmU1332, Dm1031, Dm1040 e Dm1044) è stata osservata un'espressione regolata durante lo sviluppo, mentre in altri 7 casi (Dm U18, Dm1009, Dm1021, Dm1031, Dm1040, Dm1044 e Dm1047) sono stati evidenziati trascritti le cui dimensioni erano comprese tra i 300 nt e i 500 nt, cioè superiori a quelle predette per le molecole di snoRNA della classe C/D. Una successiva analisi delle sequenze fiancheggianti i geni confermati sperimentalmente ha condotto all'identificazione di 16 isoforme altamente correlate ad alcuni snoRNA inizialmente trovati dallo SNOSCAN (13 per l'approccio lievito/*Drosophila* e 3 per quello Uomo/*Drosophila*), portando così a 52 il numero totale dei geni per snoRNA della classe C/D identificati nel genoma di *Drosophila melanogaster*. L'analisi ha in due casi anche portato ad identificare geni codificanti per snoRNA della classe H/ACA che mappavano nelle regioni immediatamente fiancheggianti i geni identificati (DmSnR60a-b e DmSnR64a-b-c). E' da notare che per l'assoluta maggioranza dei geni confermati sperimentalmente (25/36), lo

score assegnato dal programma SNOSCAN era significativamente superiore al valore soglia di attendibilità. Infatti, per quanto riguarda l'approccio lievito/Drosophila, la media degli scores assegnati dal programma SNOSCAN per i geni di cui è stata confermata l'espressione è pari a 32.8, contro una media di 24.17 osservata per i geni risultati negativi all'analisi trascrizionale. Analogamente, per l'approccio Uomo/Drosophila la media degli scores relativa ai geni confermati è pari a 30.4 contro quella di 23.46 per i geni la cui espressione non è stata confermata.

La percentuale di geni di cui è stata confermata la reale espressione è risultata essere notevolmente superiore per l'approccio basato sull'allineamento delle sequenze degli rRNA di lievito e Drosophila (26/44, 60%) rispetto a quella osservata nel caso dell'allineamento delle sequenze degli rRNA di Uomo e Drosophila (10/56, 18%). La maggiore percentuale di geni identificati nell'approccio basato sull'allineamento tra le sequenze degli rRNA di lievito e Drosophila può presumibilmente essere messa in relazione con le caratteristiche dei nuovi snoRNA identificati che presentavano tutte le caratteristiche strutturali canoniche e dimensioni tipiche degli snoRNA di classe C/D (ad eccezione del solo DmU18, che sembra appartenere alla categoria degli scaRNA di tipo C/D-H/ACA). Al contrario, la maggior parte degli snoRNA identificati mediante allineamento con l'rRNA umano (7/13) ha presentato caratteristiche strutturali (box D o D' varianti) e/o dimensioni differenti da quelle comunemente attribuite agli snoRNA C/D. Almeno 3 di queste molecole presentano infatti una struttura complessa tipica degli scaRNA di tipo C/D-H/ACA o scaRNA di tipo C/D (D1021, Dm1040 e Dm1031). L'identificazione di molecole più complesse nel

caso dell'approccio basato sull'allineamento tra le sequenze di rRNA di Uomo e *Drosophila* potrebbe essere legata all'acquisizione di nuove funzioni da parte degli snoRNA nel genoma dei metazoi. Infatti, mentre l'assoluta maggioranza degli snoRNA identificati nel primo approccio è presumibilmente coinvolta nella modificazione degli RNA ribosomali 18S o 28S, almeno 6 tra quelli identificati nel secondo approccio sono presumibilmente coinvolti nella metilazione degli snRNA. E' noto che le modificazioni delle molecole di snRNA possono alterare la loro capacità di associazione con le componenti del complesso dello splicesoma ed in alcuni casi influenzare specificamente l'efficienza del processo di splicing (Donmez *et al.*, 2004). Sebbene siano state identificate anche in lievito molecole di snoRNA capaci di operare la metilazione degli snRNA (come nel caso dello snoRNA mgU6-47 di *S.pombe*; Zhou *et al.*, 2002), il numero ridotto di geni dotati di introni (3-4%) presente nel genoma di questo organismo lascia ipotizzare che questa funzione abbia acquisito un'importanza molto maggiore nel corso dell'evoluzione del genoma dei metazoi, nei quali è noto che la maggior parte dei geni è interrotta da introni ed utilizza estesamente meccanismi di splicing alternativo.

Per quanto riguarda le caratteristiche strutturali dei nuovi geni identificati in entrambi gli approcci, 48/52 di essi presentano una box D' addizionale; quasi la metà (24/52) mostra entrambe le box canoniche, mentre una frazione di poco inferiore (20/52) presenta la box D canonica e la box D' variante (in media in un solo nucleotide); in 2/52 casi la box D è variante in 1 nt e la box D' canonica; 3/52 presentano la sola box D canonica e 1/52 la sola box D variante in 1 nt. Infine, solo in tre casi (3/52) entrambe le box D e D' risultano varianti,

anche se in un solo nucleotide. La sequenza antisenso che mostra complementarità con l'RNA ribosomale è risultata localizzata per la maggior parte dei geni candidati (48/52) a monte della box D', caratteristica questa abbastanza insolita perché la sequenza guida complementare alla molecola di RNA target è di solito posizionata a monte della box D, e solo in pochi casi era stata finora riscontrata un'organizzazione di tipo diverso (Tycowski *et al.*, 1998; Darzacq *et al.*, 2002). Questo dato ha suggerito l'ipotesi che le molecole identificate potessero essere bifunzionali e cioè che la sequenza antisenso a monte della box D potesse riconoscere sequenze target diverse dallo rRNA. Una caratteristica generale per una corretta funzionalità delle molecole di snoRNA è il perfetto appaiamento tra la sequenza guida ed il suo target; tuttavia, dalla nostra analisi è emerso che solo il 67% dei geni confermati sperimentalmente (35/52) rispettava l'appaiamento previsto con lo rRNA per una lunghezza superiore ai 10 nt. Per almeno 4 geni la regione di perfetta complementarità con la molecola di rRNA si estende per un tratto ancora più breve, di 8-9 nt. Questa condizione è stata arbitrariamente considerata compatibile con il riconoscimento della molecola di RNA target sulla base dei dati emergenti in letteratura (vedi per una review Bachellerie *et al.*, 2002) e porta ad avanzare l'ipotesi che la lunghezza minima del tratto di perfetta complementarità, finora stimata generalmente di almeno 10 nt (Kiss-Lazslo *et al.*, 1996; Nicoloso *et al.*, 1996) possa essere ulteriormente abbassata a tratti di lunghezza inferiore. Per 13 su 52 geni la sequenza guida antisenso non ha mostrato complementarità con l'rRNA, suggerendo la possibilità che questi nuovi snoRNA potessero appartenere alla categoria di snoRNA cosiddetti

“orfani”. La possibilità che questi snoRNA potessero riconoscere come target molecole di RNA di tipo diverso dagli rRNA è stata saggiata cercando nuovi possibili target nella banca dati Noncode (<http://noncode.bioinfo.org.cn>). Quest’analisi ha portato ad individuare per 5 snoRNA “orfani” putativi siti di metilazione sulle molecole di snRNA. In particolare, due di questi siti sono riconosciuti dallo snoRNA Dm1044 su due diversi snRNA (U1:82Eb e U12), mentre il gene Dm1021 riconosce altri due siti, rispettivamente sugli snRNA U4:25F e U5:23D. Quest’ultimo snoRNA, classificabile come putativo scaRNA di tipo C/D-H/ACA, è quindi implicato sia nelle reazioni di metilazione che di pseudouridilazione. Infine, un putativo sito target dello snoRNA Dm1046 è presente su una sequenza comune a tutte le isoforme dello snRNA U1. A sorpresa, una variante genica dello snRNA U1, U1:21D, è presente a sole 300 bp dal gene Dm1046 e viene trascritta con orientamento opposto, condividendo con lo snoRNA una corta regione regolativa presente in 5’. E’ suggestiva l’idea che la trascrizione dello snoRNA Dm1046 e del suo possibile target, lo snRNA U1:21D, possano essere soggette ad una comune regolazione, e quest’ipotesi verrà saggiata in futuri esperimenti.

In definitiva, il numero dei putativi snoRNA identificati per i quali resta da definire un putativo RNA bersaglio è stata quindi ridotta a 9. Dato che era molto stimolante l’ipotesi che queste molecole, finora identificate come “orfani”, potessero riconoscere molecole di mRNA, ho saggiato la possibilità che anche le molecole di snoRNA che mostravano perfetto appaiamento con l’rRNA 18S/28S potessero riconoscere mediante la stessa o diversa sequenza guida antisense, anche altre molecole di RNA bersaglio. Tale analisi è stata

basata sull'ipotesi che l'appaiamento della sequenza antisenso posizionata a monte della seconda box o addirittura della stessa sequenza antisenso che si appaia all'rRNA poteva portare la stessa molecola di snoRNA a modificare target multipli. La ricerca di nuovi possibili target è stata effettuata tramite BLAST sull'intero trascrittoma di *Drosophila* (<http://www.flybase.org>; BLAST database: alla sezione Annotated genes: NT) utilizzando tutte le sequenze antisenso (13nt) degli snoRNA “orfani” ed ha portato ad individuare 41 mRNA che potrebbero rappresentare putativi target per 7 di questi snoRNA. La maggior parte dei trascritti identificati come putativi target codificano per proteine con attività enzimatica. In due casi (DmG1332 e Dm1008a), le sequenze guida degli snoRNA sono risultate potenzialmente capaci di modificare nucleotidi immediatamente adiacenti a siti di giunzione esone/introne. Infatti, lo snoRNA DmG1322 riconosce come putativo target il trascritto del gene *Pde6* (*Phosphodiesterase 6*; CG8279) nelle adiacenze della giunzione tra esone 1/ introne1, mentre lo snoRNA Dm1008a ha come putativo target il trascritto del gene *scrib* (*scribbled*; CG5462), nel dinucleotide di giunzione tra esone 9/introne 9. E' poi interessante notare che l'mRNA del gene *Pde6* rappresenta un potenziale target anche per lo snoRNA orfano Dm1008a, in corrispondenza dell'esone 12. Per tutti gli snoRNA “orfani” ho anche saggiato la potenzialità di appaiamento della sequenza antisenso localizzata a monte della seconda box D ed in 6 casi su 9 questa è perfettamente complementare ad almeno 12 molecole di mRNA (dati non mostrati). L'analisi effettuata ha così identificato 5 snoRNA orfani (DmG1322, DmU18adj, Dm1008a, Dm1011 e Dm1047) per i quali entrambe le sequenze

antisense possono essere potenzialmente funzionali e riconoscere molecole di mRNA come putativi target. Allo scopo di confermare ulteriormente l'ipotesi di bifunzionalità o multifunzionalità delle molecole di snoRNA ho poi avviato una ricerca di nuovi putativi target per tutte le sequenze antisense delle molecole di snoRNA fin qui identificate. I dati ottenuti finora sembrano sostenere l'ipotesi di multifunzionalità; dal momento che alcuni snoRNA capaci di appaiarsi con l'rRNA sono in grado di appaiarsi anche con mRNA che potrebbero quindi rappresentarne ulteriori target. Inoltre, putativi target addizionali vengono riconosciuti anche dalla sequenza antisense presente a monte della seconda box D, quando essa è presente. Futuri esperimenti dovranno tuttavia confermare la reale presenza di metilazione dei siti identificati e chiarirne il ruolo funzionale. Tuttavia, i dati ottenuti, seppur preliminari, potrebbero aprire enormi prospettive allo studio del ruolo biologico svolto dalle molecole degli snoRNA. E' poi interessante ricordare che quando la base selezionata per la metilazione è una A, la sua modificazione può antagonizzare l'efficienza del processo di editing operato dagli enzimi ADAR (Vitali *et al.*, 2005).

L'insieme dei dati ottenuti ha quindi dimostrato l'efficacia dell'approccio utilizzato nell'identificazione degli snoRNA della classe C/D. Infatti, il programma SNOSCAN, pur partendo dalla complementarità con l'rRNA ribosomale, determina lo score finale sulla base dell'insieme delle caratteristiche strutturali ed è quindi stato in grado di identificare come geni candidati anche quelli la cui sequenza antisense non era perfettamente complementare ai potenziali siti di metilazione presenti sull'rRNA.

Per quanto riguarda l'organizzazione genomica dei geni che codificano per snoRNA, i dati ottenuti (Tycowski *et al.*, 1996; Tycowski and Steitz, 2001; Yuan *et al.*, 2003; Huang *et al.*, 2004, 2005) hanno evidenziato un'ampia ed efficiente utilizzazione delle regioni introniche. Come in altri organismi, i geni codificanti per gli snoRNA sono risultati localizzati frequentemente in introni di geni i cui prodotti proteici sono coinvolti nel processo di traduzione. E' interessante notare che recentemente in *Drosophila* sono state evidenziate due principali tipologie di organizzazione genica per gli snoRNA di tipo C/D e H/ACA. Un'organizzazione in cluster, di tipo dUhg, veicola soprattutto l'espressione degli snoRNA di classe C/D, mentre un'organizzazione in cluster intronici è stata descritta preferenzialmente per gli snoRNA H/ACA (Huang *et al.*, 2004, 2005). I geni all'interno dei cluster dUhg sono ripetuti in tandem ed ospitati in introni di geni non-coding secondo la regola "uno snoRNA per introne", come osservato nei vertebrati ed in lievito. Tali geni producono solitamente un trascritto poliadenilato il cui "processing" libera gli snoRNA ed un trascritto maturo privo di registro di traduzione (ORF) e con funzione sconosciuta. I cluster intronici, invece, sono costituiti da geni codificanti snoRNA che si ripetono in tandem nello stesso introne del gene ospite e la loro maturazione avviene attraverso un meccanismo che prevede tagli sia endonucleolitici che esonucleolitici, cioè un meccanismo sostanzialmente differente da quello che agisce nella maturazione degli altri snoRNA presenti in singola copia in un introne. Questo tipo di organizzazione è ampiamente rappresentata nel genoma delle piante ed è stata finora descritta in *Drosophila* solo per gli snoRNA di classe H/ACA (Huang *et al.*, 2004, 2005).

Dai risultati emersi dal mio lavoro l'organizzazione degli snoRNA di tipo C/D è apparsa però molto più variegata di quanto precedentemente immaginato. Per la prima volta infatti sono stati identificati geni per snoRNA localizzati all'interno di esoni (10/52) di geni ospiti codificanti per proteine. In due casi i geni identificati mappano in regioni sottoposte a splicing alternativo, per cui la loro localizzazione può essere esonica od intronica a seconda del pattern di splicing del gene ospite, mentre in 3 casi la localizzazione è a cavallo di una giunzione esone/introne. In 5 casi inoltre (DmG1322, DmU18adj, DmSnR69a, DmSnR69b e Dm1040) i geni hanno mostrato orientamento trascrizionale opposto a quello dell'ospite. L'orientamento trascrizionale opposto fa sì che gli snoRNA siano complementari ai trascritti primari del gene ospite per tutta la loro lunghezza. Ciò renderebbe questi snoRNA potenzialmente capaci di modulare l'espressione dei geni ospiti, dato che l'appaiamento potrebbe mascherare i siti riconosciuti dallo spliceosoma, regolando lo splicing del pre-mRNA o addirittura determinarne la degradazione, innescando il processo noto come "RNA interference". La localizzazione all'interno di ORF o putative ORF era stata finora osservata in *Drosophila* solo per 8 snmRNA, mentre un orientamento opposto rispetto al gene ospite era stato osservato solo per 2 snmRNA codificati all'interno di introni (Yuan *et al.*, 2003). Questo tipo di organizzazione genomica potrebbe riflettere un meccanismo di biogenesi basato sulla produzione alternativa delle molecole di snoRNA e di mRNA sovrapposte, come dimostrato negli esperimenti condotti su oociti dello *X.laevis* per gli snoRNA intronici U16 e U18 (Caffarelli *et al.*, 1996). Anche studi condotti in lievito e nei vertebrati (Lee *et al.*, 2003; de Turrís *et al.*, 2004)

hanno dimostrato l'esistenza di un meccanismo di biogenesi splicing-indipendente, indicato come via minoritaria di sintesi per la produzione di alcuni snoRNA intronici; tale meccanismo dovrebbe rappresentare per gli snoRNA esonici l'unica via di sintesi possibile. Per la maggior parte degli altri geni identificati (24/52) è stata osservata una localizzazione intronica; di questi, in un solo caso (Dm 1044) il gene risulta orientato con polarità opposta rispetto al gene ospite. Infine, una frazione non trascurabile (13/52) dei geni identificati risultata localizzata in regioni intergeniche. 14 geni sono risultati ospitati in introni od esoni di geni che codificano per proteine non coinvolte nel processo di traduzione. Tra questi, esempi interessanti sono rappresentati dal gene *Sulfated* (*Sulf1*) che ospita nel suo terzo esone il gene DmG1322 ed è trascritto con polarità opposta; dal gene *eiger*, che ospita invece nel suo primo introne il gene Dm1044, anch'esso trascritto con polarità opposta; dal gene *timeless* (*tim*) che ospita nel suo terzo esone Dm1047 ed infine dal gene *dik*, un gene ininterrotto che ospita al suo interno il gene DmU1332. Il gene *Sulf1* è indispensabile per il normale sviluppo della cuticola embrionale e codifica per una proteina (N-acetilglucosamina-6-solfatasi, G6S) che appartiene alla famiglia delle solfatasi extracellulari. Per il gene omologo umano, *HSulf-1*, è stato recentemente osservato un coinvolgimento nella regolazione della crescita cellulare; l'assenza di *HSulf-1* nelle linee cellulari di carcinoma squamoso del collo e della testa potenzia l'azione del fattore di crescita "heparin-binding", determina un aumento della motilità e dell'invasività delle cellule tumorali ed inibisce l'apoptosi indotta da stress con un conseguente aumento della tumorigenicità (Lai *et al.*, 2004). Il gene *eiger* codifica invece per il primo

membro della superfamiglia dei fattori di necrosi tumorale (TNF: Tumor Necrosis Factor) scoperto negli invertebrati. *Eiger* è espresso in tutti gli stadi di sviluppo della *Drosophila*, prevalentemente nel sistema nervoso, dove sembrerebbe regolare la proliferazione delle cellule progenitrici dei neuroni permettendo il normale sviluppo del sistema nervoso (Moreno and Basler, 2002). Il gene *timeless* codifica invece per una proteina coinvolta nel mantenimento del ritmo circadiano della *Drosophila*; mutazioni amorfe del gene *tim* sono recessive e provocano negli adulti di *Drosophila* anomalie nel ritmo, nonché problemi locomotori, di accoppiamento e di sviluppo (Bae *et al.*, 1998). Il gene *Dik*, indicato in banca dati anche con la denominazione *ADA3*, non presenta introni e codifica un prodotto proteico con un putativa attività di acetiltrasferasi degli istoni che è un componente del complesso attivatore della trascrizione Ada2/Gcn5/Ada3 (Muratoglu *et al.*, 2003; Kusch *et al.*, 2003). Per quanto riguarda il gene DmU1332, identificato all'interno del gene *Dik*, i risultati emersi dall'analisi trascrizionale sono stati abbastanza insoliti: infatti, è emerso che tale gene non è espresso nello stadio embrionale e nel primo stadio larvale; mostrando quindi un profilo di espressione differente da quello del gene ospite, che si esprime invece in maniera costitutiva. Questo dato ha avvalorato ulteriormente per questo snoRNA l'ipotesi di un meccanismo di biogenesi splicing indipendente già generalmente avanzato per tutti i geni a localizzazione esonica ma ancora più evidente in questo caso, essendo il gene ospite privo di introni e quindi il suo trascritto non sottoposto a splicing.

Come già affermato precedentemente, la maggior parte dei geni identificati è risultata essere organizzata in cluster. Sono stati identificati 9 nuovi cluster

costituiti da copie ripetute in tandem ed altamente correlate tra loro a livello di sequenza (rispettivamente i cluster DmSnR39-59/DmSnR48, DmU18, DmSnR60, DmSnR62, DmSnR64, DmSnR69, Dm3403 e Dm1006/1007/1008) e solo in un caso il cluster è risultato formato da geni sia omologhi che eterologhi (il cluster DmSnR58/Dm3112). Questi dati implicano che la duplicazione dei geni degli snoRNA si è verificata frequentemente durante l'evoluzione del genoma di *Drosophila*. È interessante notare che i geni Dm3403a-b-c-d hanno fornito il primo caso di cluster genico a localizzazione esonica; infatti i 4 geni che compongono il cluster sono tutti localizzati nello stesso gene ospite, in esoni successivi o molto vicini tra loro. Il cluster DmSnR69 comprende invece uno snoRNA esonico (DmSnR69a) ed uno localizzato a cavallo di una giunzione esone-introne del gene ospite (DmSnR69b), mentre i cluster DmSnR39-59/DmSnR48 e Dm1006/1007/1008 hanno una più comune localizzazione intergenica. Il cluster a localizzazione intergenica DmSnR39-59/DmSnR48 ha poi mostrato un'organizzazione particolare, in quanto è costituito da 2 copie del gene DmSnR48 (a-b) separate da 2 copie del gene DmSnR39-59, un tipo di localizzazione che potrebbe essere il risultato di un evento di duplicazione seguito da trasposizione in regioni adiacenti, secondo un meccanismo di mobilità ipotizzato per i "C/D snoRNA retrogenes" di mammifero (Bachellerie *et al.*, 2002). In più, i risultati recentemente riportati da Huang *et al.*, (2005) hanno confermato l'ipotesi, dettata dalla presenza in banca dati (<http://www.flybase.org>) di una EST (EN01530) che copre l'intera regione del cluster DmSnR39-59/DmSnR48, che questo cluster rappresenti un tipico esempio di organizzazione tipo dUhg

(dUhg4; Huang *et al.*, 2005). Il cluster costituito dai geni Dm1006, Dm1007, Dm1008 e dall'isoforma Dm1008a ha anche mostrato un'organizzazione peculiare. Infatti, l'isoforma Dm1008a si sovrappone in parte all'ultimo esone del gene *RpL23a*; e l'effettiva trascrizione di questa isoforma suggerisce un'organizzazione genica del tutto simile a quella recentemente descritta per una significativa percentuale dei geni umani presenti sui cromosomi 21 e 22 (Cawley *et al.*, 2004). Dalla mappatura dei siti di legame per i fattori di trascrizione Sp1, p53 e cMyc è stato infatti osservato che la maggior parte di questi siti (36%) sono situati al 3' di geni coding e presumibilmente marcano il punto di inizio della trascrizione di geni non coding parzialmente sovrapposti. Bisogna inoltre evidenziare che il programma SNOSCAN ha identificato due box D e D' alternative all'interno degli snoRNA Dm1006 e Dm1008, shiftata di pochi nucleotidi e per le quali naturalmente sono identificabili due sequenze antisense differenti che potrebbero consentire a queste molecole di guidare differenti eventi di metilazione sulla molecola di rRNA target. In conclusione, i cluster dUhg finora identificati in *D. melanogaster* sono 8 (Huang *et al.*, 2005) e rappresentano un'importante tipo di organizzazione genica per gli snoRNA di classe C/D nel genoma di questo organismo. Molto interesse suscita l'origine evolutiva di questi geni, dato che gli introni che codificano gli snoRNA e non gli esoni risultano conservati dagli Invertebrati all'Uomo (Tycowski and Steitz, 2001). Appare quindi probabile che in questo tipo di trascritti non coding solo gli introni producano molecole funzionali. Quest'osservazione porta a riconsiderare la ben nota definizione di "esoni" come le regioni contenenti le

principali informazioni genetiche. Sono tre le ipotesi avanzate per cercare di spiegare l'origine evolutiva dei dUhg:

- Gli esoni di questi geni codificavano originariamente per una proteina, ma hanno perso questa proprietà durante l'evoluzione;
- Gli esoni codificavano per un unico RNA non codificante, la cui funzione nel tempo è diventata obsoleta o non è stata ancora identificata;
- I geni UHG e UHG-like si sono originati da geni codificanti snoRNA policistronici, simili a quelli ritrovati in piante, in lievito e *D. melanogaster*, mediante la conversione delle regioni che separano i geni che codificano gli snoRNA in esoni. A questo proposito è interessante notare che recentemente è stato identificato in *C. elegans* uno pseudogene, *rpl-7aps*, correlato al gene *rpl-7a* che codifica per la proteina ribosomale L7a. L'origine di questo pseudogene è chiaramente legata ad un evento di duplicazione e lo pseudogene mostra il 97% di identità con il gene ancestrale. Tuttavia, lo pseudogene *rpl-7aps* ha perso la capacità di codificare la proteina ribosomale, in quanto entrambi i trascritti che produce contengono codoni di stop prematuri (PTC) e vengono rapidamente degradati dal pathway del nonsense-mediated decay (Mitrovich and Anderson, 2005). Pur avendo perso la capacità di codificare una proteina, lo pseudogene ha conservato del gene ancestrale la funzione di codificare uno snoRNA di classe H/ACA. Questi dati suggeriscono l'ipotesi che le sequenze degli pseudogeni espressi possano essere mantenute nel genoma grazie alla capacità di codificare per snoRNA funzionali, anche se le mutazioni accumulate non consentono più la produzione di una proteina e forniscono quindi un supporto sperimentale

all'ipotesi che i geni UHG si siano originati da un gene ancestrale, probabilmente di tipo protein-coding. Tale ipotesi inoltre suggerisce anche l'esistenza di un possibile tipo di vantaggio selettivo a favore del mantenimento delle sequenze degli pseudogeni nei genomi degli organismi eucariotici.

Ritornando ai dati emersi dalla mia ricerca, un'interessante scoperta ha riguardato il cluster intronico DmSnR62: infatti i geni di tale cluster, DmSnR62a e DmSnR62b, sono localizzati entrambi nell'introne 2 del gene ospite *RpL13a*. La presenza di più copie ripetute in tandem di uno stesso gene codificante per snoRNA all'interno di un singolo introne, caratteristica dei cluster intronici identificati nelle piante, era stata riscontrata finora in *Drosophila* solo per gli snoRNA appartenenti alla classe H/ACA (Huang *et al.*, 2004). Il cluster DmSnR62 rappresenta quindi un'ulteriore novità riguardo il tipo di organizzazione dei geni che codificano snoRNA di tipo C/D nel genoma di *Drosophila*.

L'analisi trascrizionale effettuata per i 100 putativi geni identificati dall'analisi computazionale, oltre a confermare l'espressione dei 52 geni già citati, ha rivelato anche la presenza di trascritti la cui lunghezza o le cui caratteristiche di sequenza sono risultate non attribuibili in maniera semplice alla categoria degli snoRNA. Poiché questi trascritti non corrispondono nemmeno ad mRNA derivati dai rispettivi geni ospiti, né è stato osservato alcun allineamento con altri loci del genoma di *D. melanogaster* mediante l'utilizzo dell'algoritmo BLAST, è plausibile che essi definiscano nuovi putativi non-coding RNA. I trascritti in questione, denominati rispettivamente

pncRNA1015a, pncRNA105b, pncRNA1016, pncRNA1019, pncRNA1020, pncRNA1028 e pncRNA1049, potrebbero quindi risultare particolarmente interessanti. Ad esempio, i pncRNA1015a e pncRNA1015b, rispettivamente di circa 1000 nt ed 800 nt, derivano dal primo introne del gene *NFAT* (*Nuclear factor of activated T cell*). Essendo trascritti con polarità opposta a quello del gene ospite, entrambi potrebbero rappresentare dei trascritti naturali antisenso (*cis*-NAT) e fungere da regolatori di *NFAT*. Il pncRNA1016, identificato mediante l'utilizzo di una sonda che mappa nel secondo esone del gene *defective chorion 1* (*dec-1*), risulta espresso esclusivamente negli stadi di sviluppo in cui il gene ospite non è espresso. Si potrebbe ipotizzare per questo trascritto una funzione regolativa sul gene *dec-1*, dato che il pncRNA1016 pare espresso solo negli stadi in cui il gene ospite non si esprime. Infatti, *dec-1* si esprime solo in mosche adulte durante l'oogenesi, mentre il putativo ncRNA1016 è evidenziabile solo in embrioni, larve di 1° e di 3° e sembra inoltre accumularsi in maggiori quantità in questi ultimi due stadi. Per gli altri putativi ncRNA non è possibile ad oggi ipotizzare alcuna funzione, anche se tutti presentano la comune caratteristica di essere regolati nello sviluppo. Tra questi, i più interessanti sembrano essere il putativo ncRNA 1019, di circa 1000 nt, che si accumula in tutti gli stadi tranne che in adulti e deriva dal gene *paralytic* (*para*), che codifica per una proteina integrale di membrana che costituisce un canale del Na⁺ voltaggio dipendente; infine, il putativo ncRNA 1028, di circa 700 nt, mappa nel gene *La-related protein* (*Larp*), coinvolto nella morte cellulare programmata.

Concludendo, nel loro complesso i risultati ottenuti dimostrano che la ricerca da me condotta è andata ben oltre l'obiettivo iniziale. In fase di realizzazione, infatti, la ricerca non solo si è rivelata fruttuosa per l'identificazione di nuovi geni codificanti per snoRNA nel genoma di *Drosophila melanogaster*, ma ha evidenziato una serie di informazioni significative riguardo l'organizzazione genomica e strutturale di questi geni, suggerendo anche l'esistenza di un meccanismo di biogenesi non ancora ben caratterizzato ma già ipotizzato in altri organismi. La ricerca ha altresì condotto all'identificazione di nuove putative molecole di ncRNA, strutturalmente diverse dagli snoRNA, la cui futura caratterizzazione potrà gettare le basi per delineare un quadro più completo dei rispettivi ruoli funzionali e della specifica organizzazione. A tale proposito uno degli obiettivi futuri sarà quello di procedere alla caratterizzazione funzionale dettagliata di molti dei geni identificati, selezionati in base a caratteristiche interessanti e particolari. Il numero sempre crescente di nuove classi di ncRNA funzionali identificate in vari organismi mostra infatti che il campo della "RNomica" è in forte espansione e c'è quindi da aspettarsi che nuove ed interessanti categorie di ncRNA possano essere identificate in un prossimo futuro. Da qui l'esigenza di progettare urgentemente nuovi strumenti, sia bioinformatici che sperimentali, volti all'identificazione e caratterizzazione di queste categorie di trascritti. Una nuova grande prova attende i ricercatori negli anni a venire: una piena comprensione dei processi biologici fondamentali degli organismi superiori richiederà una profonda conoscenza sia del Proteoma che del Trascrittoma e delle loro complesse interazioni.

4. MATERIALI E METODI

4.1 Ceppo di *Drosophila melanogaster* utilizzato

Oregon (wild-type)

Ciclo vitale di *Drosophila melanogaster* a 25°C

STADIO	ORE	GIORNI
Embrione	0-24	0-1
Larva 1° stadio	24-48	1-2
Larva 2° stadio	48-72	2-3
Larva 3° stadio	72-120	3-5
Prepupa	120-132	5-5.5
Pupa	132-216	5.5-9
Adulto	216	9-10

4.2 Terreno per la coltura in massa di *D.melanogaster*

Il terreno su cui vivono i moscerini nel nostro laboratorio viene preparato utilizzando un terreno per la coltura di insetti (Instant *Drosophila* medium) fornito dalla Sigma-aldrich. Per la preparazione di 1 lt di terreno di coltura vengono aggiunti ad un litro di acqua, 300 gr. di Instant *Drosophila* medium, 10 gr. di agar e il tutto viene portato ad ebollizione; si lascia bollire per 5' e si aggiungono successivamente 2.5 gr. di metil p-idrossibenzoato, sciolto in una

appropriata quantità di alcool etilico 95%, come antimuffa. Il terreno di coltura così preparato viene poi versato in appositi barattoli cilindrici utilizzati per il mantenimento dei ceppi di *Drosophila*. Si lascia raffreddare coprendo i barattoli con della garza per evitare che possano verificarsi contaminazioni del terreno e si aspetta fino a che il terreno sia completamente asciutto; infine si tappano i barattoli con del cotone da batteriologia.

4.3 Mantenimento dei ceppi

I ceppi vengono fatti crescere in camere termo-regolate ad una temperatura di 22°C in barattoli contenenti il terreno suddetto e vengono trasferiti su terreno fresco ogni 14/15 giorni. Il trasferimento viene operato previa analisi al microscopio del fenotipo dei moscerini per accertarsi che nessun individuo con caratteristiche differenti da quelle corrispondenti al ceppo in esame sia presente nel barattolo. Si selezionano poi una trentina di individui di ciascun sesso e si trasferiscono in un nuovo barattolo. Il numero di individui da trasferire non è fisso ma dipende dal genotipo dei moscerini e un numero maggiore è richiesto se essi sono caratterizzati da bassa fertilità.

4.4 Ricerca in banca dati e analisi di sequenza

Le sequenze genomiche disponibili all'indirizzo (ftp://ftp.ncbi.nih.gov/genomes/Drosophila_melonogaster) sono state analizzate per la ricerca di potenziali geni per snoRNA della classe C/D presumibilmente coinvolti nella metilazione di siti sugli rRNA (potenzialmente conservati tra lievito e *Drosophila* utilizzando il programma SNOSCAN; Lowe and Eddy,

1999, <http://rna.wustl.edu/snoRNAdb/code/>). Lo SNOSCAN sfrutta un efficiente algoritmo per identificare geni candidati che codificano per putativi snoRNA coinvolti nel processo di 2'-O-metilazione di molecole di RNA target ed opera a partire dalla conoscenza delle regioni riconosciute dalle sequenze antisense degli snoRNA e dei siti metilati. Il programma analizza in maniera sequenziale 6 caratteristiche proprie di questi geni; la presenza di Box D e box C, della regione di complementarità con l'eventuale molecola target, l'eventuale presenza della box D' addizionale se la regione di complementarità non è perfettamente adiacente alla box D, la presenza del sito di metilazione all'interno della molecola target, e lo stem terminale di complementarità, se presente. Il programma, inoltre, tiene conto anche della distanza relativa tra le varie componenti esaminate all'interno della sequenza del putativo snoRNA, informazione critica per ridurre i falsi positivi. Per identificare i geni codificanti per snoRNA, lo SNOSCAN necessita delle sequenze degli rRNA (18S, 28S e 5.8S) e di una lista dei siti di metilazione sugli rRNA. I siti di metilazione in *Drosophila* non ancora determinati sperimentalmente sono stati ottenuti realizzando un allineamento strutturale tra gli rRNA di *D.melanogaster* e quelli di *S.cerevisiae* e *H. sapiens* (<http://www.rna.icmb.utexas.edu>), utilizzando i siti di metilazione definiti sperimentalmente in questi due organismi (http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snoRNA/conservation). I risultati della ricerca operata dallo SNOSCAN consistono in una sintesi delle informazioni relative a ciascuna sequenza degli snoRNA predetti e dalla sequenza stessa riportata in formato FASTA. Le informazioni relative a ciascun candidato includono:

Il nome della sequenza ricercata ed il sito di inizio e fine dello snoRNA all'interno della sequenza;

Valore di attendibilità (score) più alto ritrovato;

Sequenza target e localizzazione del target;

Numero totale di basi appaiate e mismatches nella sequenza guida;

Lunghezza della sequenza dello snoRNA candidato.

Nella schermata finale dello SNOSCAN (vedi Fig.1) è inclusa anche una rappresentazione grafica della regione di appaiamento tra sequenza guida e target e la struttura secondaria dello stem:

```
(B) snoscan Output
>> snR24 26.40 (1-87) Cmpl: ySc-25S-Aml1447 (U24) 12/0 bp Gs-DpBox: 18 (18) Len: 87 TS
Meth site found: 1447 (U24) Guide Seq Sc: 11.88 (21.36 -1.12 -7.36 -1.00)
*
Db seq: 5'- AGUAGCAAUUAU -3' ySc-25S (1444-1456)
|||||
Qry seq: 3'- AGACUUCUUCGUUUAUA -5' snR24 (29-18)

Terminal stem: +- [C Box]-N- ACU - 5' Stem Sc: 0.84 (3 bp)
| |||
+---[D Box] - UGAA - 3' Stem Transit Sc: -1.11

>Summary [ C Box ] -- -- [ Cmpl/ Mism ] X [D'Bx] -- -- [D Bx] Length
>Meth Aml1447 [AUGAUGU] -- 6 bp -- [ 12 / 0 ] 1 [CAGA] -- 47 bp -- [CUGA] 87 bp
>Sc 26.40 [ 7.48 ] -- -1.59 -- [ 21.36 bits ] [3.94] -2.44 [8.05]

Candidate sequence:
>snR24 26.40 (1-87) Cmpl: ySc-25S-Aml1447 Len: 87
TCAAAATGATGTAATAACATATTTGCTACTTCAGATGGAACTTTGAGTTCCGAATGAGACA
TACCAATTATCACCAAGATCTCTGATG
```

Figura 1

4.5 Estrazione di DNA a alto peso molecolare

2-10 gr di moscerini vengono omogeneizzati in azoto liquido e poi risospesi in una soluzione 100 mM NaCl, 10 mM EDTA, 20 mM Tris-HCl, 0.5% Tryton X-100 (pH 8.5). Mediante filtrazione su garza di nylon esso viene dunque raccolto in tubi corex. Si centrifuga a 5000 rpm per 10' a 4 °C. Si elimina il sovrantante, si diluisce il sedimento, rappresentato in gran parte da nuclei, con una soluzione 20mM Tris-HCl pH 8, 10mM EDTA, 100mM NaCl; si centrifuga 5' a 5000 rpm, si elimina il sovrantante e si ripete per altre due volte l'operazione di lavaggio del sedimento. Si risospende quest'ultimo in 1.5 ml della stessa soluzione e si aggiunge 1 ml di SDS 4%; questo detergente rompe le membrane nucleari determinando un aumento di viscosità della soluzione. Dopo 5' si aggiunge 1 ml di Perclorato di Sodio 1M e si lascia in ghiaccio per 10'. Si estrae poi più volte con Fenolo/Cloroformio/Alcool Isoamilico; si precipita il DNA aggiungendo 0.6 volumi di isopropanolo e si lava 3 volte con ETOH 70%. Si essicca e si risospende in 500 µl di TE 1x.

Per eliminare l'RNA si aggiungono RNasi A in concentrazione finale 20 µg/ml e RNasi T1 in concentrazione finale pari a 1000u/ml. Dopo 1 hr a 37 °C si precipita aggiungendo 3/5 di volume di una soluzione PEG 20%, NaCl 2M e lasciando in ghiaccio per 30'. Si centrifuga a 10000 rpm per 10', si lava due volte con ETOH 70%, si essicca e si risospende in 500 µl di TE1x.

Infine il DNA è letto allo spettrofotometro per determinarne la concentrazione e la purezza. La lettura a 260 nm serve a determinare la concentrazione del DNA , sapendo che 1 mg/ml ha un'assorbanza di 22 O.D. La lettura a 280 nm serve invece a determinare la concentrazione delle proteine presenti. Se il

rapporto tra la lettura a 260 nm e quella a 280 nm è tra 1.8 e 2, il DNA è considerato privo di proteine; un rapporto inferiore è indice, invece, di un DNA più o meno ricco di proteine.

4.6 Amplificazione del DNA con la tecnica della PCR (reazione a catena della polimerasi)

La PCR é un metodo che consente di effettuare la sintesi di sequenze di DNA definite. la reazione sfrutta due primers oligonucleotidici, che fiancheggiano la sequenza del DNA da amplificare. Una serie ripetuta di cicli, che coinvolgono la denaturazione del DNA stampo, il legame dei primers allo stampo e l'estensione dei primers legati, ad opera della Taq Polimerasi, produce un accumulo esponenziale di un frammento specifico di DNA.

La miscela di reazione è così costituita:

buffer per Taq (Pharmacia) 1/10 del volume

200µM di dNTP

100ng di ciascun oligonucleotide

100ng di DNA genomico

1U di Taq(pharmacia)

H₂O fino a 50 µl.

Le condizioni per l' amplificazione sono:

1' a 94°C di denaturazione

45s a 55°C di annealing

45s a 70°C di extension

per 30 cicli.

4.7 Recupero dei frammenti amplificati per PCR mediante l'utilizzo del kit Nucleospin extract (M-medical)

I prodotti di amplificazione vengono caricati su un gel all' 1% di agarosio e dopo aver localizzato la banda di DNA di interesse con l'uso di una lampada ad UV, si taglia il pezzo di agarosio che la contiene e si ripone in una provetta Eppendorf da 1.5mL. Si aggiungono 300µl di NT1 (buffer di lisi) per 100 mg di gel tagliato e si incuba a 50°C per 5-10 min. Una volta sciolto il gel viene trasferito nella colonnina con membrana alla quale si legherà il DNA. Si centrifuga 1 min a 8000 rpm, si aggiungono 600µl di NT3 (soluzione di Etanolo) e si centrifuga 1 min a 11000 rpm. Si aggiungono 200 µl di NT3. Si centrifuga 2 min a 11000 rpm. Si trasferisce la colonnina in una provetta pulita e si eluisce il DNA con 20-25 µl di NE (Buffer di eluizione), tenuto 1 min a T ambiente, centrifugando 1 min a 11000rpm. Il frammento di DNA viene così eluito dalla membrana e raccolto nella provetta e viene conservato a- 20° C. Il frammento viene quantizzato agli UV dopo elettroforesi su gel di agarosio 1% in TAE 1X mediante confronto con un marker di peso molecolare.

4.8 Clonaggio di frammenti di PCR (pMOS Blue Blunt ended cloning kit)

Si calcola la quantità di inserto richiesto mediante l' equazione:

$Z \times 50 / 2887 \times 2.5/1 = X$ ng di inserto, in cui Z è la lunghezza in bp dell'inserto, 2887 è la lunghezza in bp del vettore, 50 sono i ng di vettore utilizzati e 2.5/1 il rapporto molare inserto/vettore utilizzato (può andare da 1/1 a 2.5/1). La quantità di inserto utilizzata può essere in eccesso se la probabilità

di avere multimeri di inserto è bassa (inserto molto lungo). Conoscendo la concentrazione dell'inserto si ricava il volume di frammento da aggiungere.

Reazione di fosforilazione dell'inserto:

H ₂ O	Y μ l
Inserto	X μ l
pK Buffer 10 X	1 μ l
DTT	0.5 μ l
pK Enzime Mix	1 μ l
Volume totale	10 μ l

Si centrifuga brevemente a 13200 rpm per raggruppare i volumi e si incuba la reazione a 22°C per 40 min. Si inattiva la proteina chinasi (pK) a 75°C per 10 min. Si tiene in ghiaccio per 2 min e si centrifuga a 13200 rpm per 1 minuto. Si prepara poi la miscela la reazione di ligasi: aggiungendo al campione 1 μ l di vettore pMosBlue; 1 μ l di DNA ligasi T4 per un volume totale di 12 μ l e si incuba a 22°C per 2h. Trasformazione: circa 30 min prima dalla fine dell'incubazione si prende una piastra con ampicillina tenuta a 4°C e la si lascia a temperatura ambiente. Si scongela in ghiaccio un'aliquota di cellule (E.coli) competenti conservate a - 80°C. Si aggiungono, 100 μ l di cellule competenti per reazione di ligasi e si tengono in ghiaccio per 10 min, dopodiché si sottopongono a shock termico incubando la miscela di reazione a 37°C per 5 min. Successivamente si piastrano le cellule con un'ansa di vetro in

condizioni di sterilità. Le piastre capovolte sono lasciate a 37°C overnight per favorire la crescita delle cellule trasformate:

Informazioni sul vettore plasmidico pMosBlue:

Linearizzazione al sito di taglio EcoR V

Lunghezza: 2887 bp

Gene per la resistenza all'antibiotico ampicillina

Promotore dell'Rna Polimerasi T7

Polilinker (a monte e a valle del sito EcoRV)

Efficienza di trasformazione circa del 75%.

4.9 Protocollo per la minipreparazione del DNA plasmidico

Versare in ogni tubo Eppendorf 1.5 ml di terreno LB con le cellule batteriche trasformate. [Riporre il resto delle colture (o inoculi) a 4°C]. Centrifugare per 1 min a 13200 rpm ed eliminare il surnatante. Aggiungere 200 µl di Sol.A e risospendere il pellet (al vortex). Aggiungere 200 µl di Sol. B; agitare bene per inversione. Aggiungere 200 µl di Sol. C ed agitare bene per inversione. Centrifugare 20 min a 13200 rpm a 4°C. Trasferire il surnatante (600 µl) in Eppendorf pulite. Aggiungere 1 ml di EtOH 95%. Agitare per inversione. Si lascia a -20°C per almeno 30 min. Centrifugare per 15 min a 13200 rpm a 4°C. Eliminare il surnatante ed aggiungere 200 µl di EtOH 70%. Agitare per inversione. Centrifugare per 5 min a 13200 rpm. Eliminare il surnatante ed asciugare bene il pellet. Risospendere in 30 µl di TE ed Aggiungere 2 µl di RNasi A. Incubare 20 min a 37°C. Conservare le miniprep. a -20°C (fino al sequenziamento del clone positivo).

4.10 Protocollo per la midipreparazione di DNA plasmidico

Versare 50 ml di terreno LB, contenente le cellule batteriche trasformate e tenute in coltura a 37°C o/n, in apposite falcon da centrifuga. Centrifugare per 6 min a 5000 rpm. Eliminare il surnatante e risospendere il pellet in 1 ml di GTE lasciando 5 min a T ambiente. Aggiungere 2 ml di Sol.B agitare e tenere 10 min in ghiaccio. Aggiungere 1.5 ml di Sol III agitare e tenere 5 min in ghiaccio. Centrifugare per 15 min a 10000 rpm e trasferire il surnatante in una nuova falcon, filtrandolo su garza per evitare di versare una piccola quota di pellet. Aggiungere 0.6 volumi di isopropanolo, agitare e tenere in ghiaccio da 20 min ad 1h. Centrifugare per 20 min a 10000 rpm. Versare il surnatante ed asciugare il pellet. Risospendere in 1 ml di TE, aggiungere 1 ml di LiCl 5M e tenere 20 min in ghiaccio. Centrifugare per 20 min a 10000 rpm e trasferire il surnatante in una nuova falcon. Precipitare di nuovo con 2.5 volumi di etanolo 95% e mettere a -20°C per almeno 30 min. Centrifugare per 15 min a 10000 rpm. Eliminare il surnatante e risospendere in 500µl di TE e aggiungere 20µl di RNasi A (10mg/µl). Incubare per 20 min a 37°C. Aggiungere 500µl di Polietilenglicol (PEG) e tenere in ghiaccio per almeno 1 ora. Centrifugare per 20 min a 13200 rpm ed eliminare il surnatante. Risospendere in 200µl di TE e lasciare a T ambiente per qualche minuto. Aggiungere 200µl di Fenolo, agitare e centrifugare per 15 min a 13200 rpm. Prelevare la fase acquosa e versarla in un nuovo tubo Eppendorf. Aggiungere 200µl di Fenolo/Cloroformio e ripetere la centrifugazione. Prelevare la fase acquosa e versarla in una nuova provetta. Aggiungere 200µl di Cloroformio/Alcool-isoamilico e centrifugare. Prelevare la fase acquosa e aggiungere 20µl di Sodio Acetato 3M e precipitare con 2.5

volumi di Etanolo 95%. Si lascia a -20°C overnight. Centrifugare per 15 min a 13200 rpm, aspirare il surnatante ed aggiungere 200µl di Etanolo 70%. Centrifugare per 5 min a 13200 rpm. Estrarre il surnatante ed asciugare bene il pellet di DNA. Risospendere in una adeguata quantità di TE. Conservare a 4°C. La concentrazione del DNA plasmidico estratto viene determinata dal valore dall'assorbanza a 260nm sapendo che il coefficiente di estinzione molare è 0.022.

4.11 Protocollo per la reazione di sequenza del DNA plasmidico (kit USB)

Si utilizza come fonte di DNA stampo il plasmide pMOSBlue, utilizzato per il clonaggio dei frammenti di snoRNA. Ai lati del sito di clonaggio (EcoRV), questo plasmide ingegnerizzato contiene le sequenze per i due primer, T7 e U19, che consentono la lettura di entrambi i filamenti delle sequenze di interesse. Si prepara la soluzione di denaturazione, composta da 4% idrossido di sodio 10N e acqua distillata. Si aggiungono 20µl di Sol.denaturante a 20µl di DNA (5µg e si incuba a temperatura ambiente per 10 min). Si precipita poi il DNA con 7µl di sodio acetato 3M ed 120µl di etanolo 95% e si incuba a -80°C per almeno 1 ora. Si centrifuga a 13200 rpm per 15 min a 4°C e si lascia asciugare bene il pellet. In seguito, si risospende, in 2µl di primer (T7 o U19), 2µl di buffer di annealing e 10µl di acqua distillata. La reazione di annealing viene incubata prima a 65°C per 5 min poi a 37°C per 10 min ed infine a temperatura ambiente per almeno 5 min. Si diluisce la RNA polimerasi T7 modificata per le sequenze nell'apposito buffer. Alla reazione di annealing si

aggiungono 3 μ l di labelling mix, 0,5 μ l di dATP α -35S e 2 μ l di T7 polimerasi. Si incuba la reazione a temperatura ambiente per 5 min e poi si aggiungono 4,5 μ l in ogni provetta contenente ciascuna i 4 differenti ddNTP/dNTP in rapporto 1/100 e si incuba a 37°C per 15-20 min. Infine si aggiungono 5 μ l di soluzione STOP, composta da formammide e colorante per la corsa, e si conserva a -20°C per un breve tempo. Tutti i reagenti sono contenute nel kit della USB.

4.12 Elettroforesi su gel di poliacrilammide in condizioni denaturanti per il sequenziamento dei cloni ottenuti

Mettere in agitazione 50 ml di soluzione contenente 6% acrilammide/bis-acrilammide (40% w/v), 8M urea, TBE 1X e acqua. Versare contemporaneamente 250 μ l di persolfato di ammonio 10% (APS) e 25 μ l di TEMED. Lasciare polimerizzare per circa 1 min, quindi versare con attenzione la soluzione tra le lastre (precedentemente pulite con alcool ed acetone per eliminare granelli di polvere e residui di poliacrilammide) evitando di formare bolle. In seguito porre un pettine all'estremità per creare il fronte di corsa ed applicare delle pinze per diminuire lo spessore del gel ed assicurare le due lastre tra di loro. Lasciare polimerizzare o/n. Creare i pozzetti di caricamento ponendo il pettine tra le due lastre e versare il tampone di corsa (TBE 1X) ai due elettrodi. Fare una precorsa per riscaldare il gel di 30 min. Caricare le reazioni di sequenza precedentemente denaturate a 90°C per 5 min sul gel di poliacrilammide. La durata della corsa è indicativa della lunghezza della sequenza campione da leggere. Generalmente nel nostro laboratorio, per la

lettura delle sequenze degli snoRNA clonati in frammenti lunghi 400-500 bp, si adoperano due combinazioni di corsa: una corsa di circa 5-7 ore a 1800V, che consente la lettura delle basi poste 300-500 nucleotidi a valle del primer utilizzato e una corsa di circa 3-4 ore, nelle stesse condizioni, che invece permette di leggere quelle basi poste 10-300 nucleotidi dal primer utilizzato.

4.13 Trasferimento del gel su carta, esposizione per autoradiografia e sviluppo

Terminata la corsa dei campioni, il gel viene trasferito su un foglio di carta Whatmann 3MM tramite semplice compressione della carta sul gel, il quale aderisce ad essa. In seguito, si lascia asciugare il gel per circa 30 min a 80°C al gel-dryer. Si controlla prima la bontà delle sequenze, tramite la misura delle conte per minuto (cpm) con il geiger, dopodichè il gel è esposto ad una lastra autoradiografica e lasciato nell'apposita cassetta da esposizione a temperatura ambiente per tutto il tempo necessario. Lo sviluppo delle sequenze consiste nell'immergere la lastra esposta, dapprima in una soluzione di sviluppo, che consente la visualizzazione delle bande tipiche di sequenza, quindi nella soluzione di fissaggio, dove le bande sono definitivamente fissate alla lastra.

4.14 Protocollo per l'estrazione dello RNA da adulti ed embrioni di *D. melanogaster*

Si raccoglie un numero adatto di individui (20-25) in tubi Eppendorf da 1.5 ml sterili, in cui vengono pestati, in 300µl di buffer di omogeneizzazione. Si aggiungono altri 200µl di buffer (vol.totale 500µl) e di proteinasi K alla

concentrazione finale di 100µg/µl e si incubare a 37°C per 1 ora. Si aggiungono 500µl di fenolo pH 4 e si centrifuga per 5 min a 13200 rpm a 4°C. Si preleva la fase superiore acquosa e si trasferisce in una nuova provetta sterile. Si ripete l'estrazione con fenolo. Si aggiungono 500µl di cloroformio/isoamilico e si centrifuga per 5 min a 13200 rpm a 4°C. Si preleva la fase superiore acquosa. Si aggiunge sodio acetato a concentrazione finale di 200mM. Si precipita con due volumi di etanolo assoluto (95%) e si pone il tutto a -80°C overnight. Si centrifuga per 15 min a 13200 rpm a 4°C, si elimina l'etanolo 95% e si aggiungono 200µl di etanolo 70%. Si centrifuga per 5 min a 13200 rpm e si elimina l'etanolo 70%, lasciando asciugare il campione. Si risospende in 30µl di H₂O DEPC e si conserva il campione di RNA a -80°C.

4.15 Protocollo per l'estrazione dello RNA da larve di *D. melanogaster* (di 1°, 2° e 3° stadio)

Si raccoglie un numero adatto di larve in tubi Eppendorf da 1.5ml sterili. Si aggiungono 300µl di buffer di omogeneizzazione per RNA da larve e si pestano velocemente. Si aggiungono altri 200µl di buffer (vol.totale 500µl) e proteinasi K (250µg/ml finale) e si incuba a 50°C per 1 ora. A fine incubazione si aggiungono 250µl di fenolo pH 4 e si tiene in agitazione 5 min, poi 250µl di cloroformio/isoamilico e si tiene in agitazione 5 min. Si centrifuga 5 min a 13200 a 4°C. Si preleva la fase acquosa superiore. Si aggiungono 500µl di cloroformio/isoamilico e si agita per 5 min. Si centrifuga per 5 min a 13200 rpm a 4°C. Si preleva la fase acquosa superiore Si precipita con due volumi di etanolo 95% e si precipita a -80°C overnight. Si centrifuga 15 min a 13200

rpm, si elimina l'etanolo 95% e si aggiungono 200µl di etanolo 70%. Si centrifuga per 5 min a 13200 rpm e si elimina l'etanolo 70%. Si lascia asciugare il campione. Si risospende in 30µl di H₂O DEPC e si conserva il campione di RNA a -80°C.

4.16 Elettroforesi dello RNA e trasferimento su filtro di nylon (Northern blot)

Si prepara un gel di agarosio denaturante all'1,2% sciogliendo 1,8gr. di agarosio in 110 ml di acqua distillata a cui si aggiungono 15ml di RB 10X . Si lascia bollire e quando la temperatura del gel scende a circa 65°C, si aggiungono, sotto cappa, 24,9 ml di formaldeide. Si caricano 6µl di RNA totale (almeno 6-7 µg di RNA) in 19µl di soluzione di caricamento, composta da formammide 50%, formaldeide 2.2M, RB 1X e 0,5% bromuro d'etidio in acqua DEPC. Si caricano anche 4µl marker (1µg/2µl; Fermentas) composto da bande di RNA a peso molecolare noto. Si denaturano i campioni per 15 min a 65°C. Infine ai campioni si aggiungono 2,5µl di una soluzione Ficoll 25%, 0,05% Bromofenolo, 0,05% Xilene Cianolo. Si caricano i campioni sul gel, precedentemente preparato, e si effettua la corsa elettroforetica per circa 2-3 ore a 120V (tampone di corsa RB 1X). A fine corsa, il gel viene adagiato su un supporto di vetro che poggia su una vaschetta contenente 300-400 ml di SSC 10X. La soluzione è in contatto con il gel mediante un ponte formato da due fogli imbevuti di carta Whatman 3MM. Un filtro di nylon (Hybond-N Amersham) delle stesse dimensioni del gel viene adagiato sul gel evitando di formare bolle d'aria. Successivamente, si pongono tre fogli di carta Whatmann

3MM, e su di essi una pila di circa 10 cm di comune carta da banco e alla sommità un peso adatto. Si lascia trasferire l'RNA overnight. Si rimuove delicatamente il filtro di nylon, lo si fa asciugare a temperatura ambiente 30 min e si fissa l'RNA al filtro a 80°C in stufa sotto vuoto circa 2 ore. La parte del filtro corrispondente al marker viene tagliata. Il marker viene colorato con blu di metilene e conservato.

4.17 Preparazione in vitro di una molecola di DNA radioattiva copiata su di un template noto: NICK TRANSLATION

Questa tecnica è basata sulla capacità della DNasi I di tagliare aspecificamente il DNA creando delle estremità 3' OH che fungono da innesco per la Polimerasi I. Quest'enzima rimuove nucleotidi del template e li sostituisce con quelli presenti nella miscela di reazione, includendo quelli radioattivi. La miscela di reazione contiene: 250mM Tris-HCl (pH 7.7), 50mM Solfato di Magnesio, 1mM DTT, 15 µM dTTP, 15 µM dGTP, 15 µM dATP, 15-20 µM[alfa 32P]dCTP (attività specifica 3000 Ci/mM). A questa miscela si aggiungono 250-500 ng di DNA template e 2µl di una miscela enzimatica contenente DNA Polimerasi I e DNasi I. La reazione avviene a 15 °C per 1 ora, dopodiché la reazione viene bloccata aggiungendo EDTA in concentrazione finale 0.2M/l Il DNA viene infine separato dai nucleotidi non incorporati mediante gel-filtrazione su colonna di Sephadex G-50.

4.18 Protocollo di preibridazione ed ibridazione degli RNA trasferiti sul filtro di nylon

Il filtro di nylon sul quale è stato trasferito l'RNA viene messo in incubazione, per 15 min a 55°, con una soluzione contenente 20ml di Sol di preibridazione/ibridazione 1X composta da buffer fosfato 0,5M, SDS 5% ed EDTA 10mM, precedentemente riscaldata a 55°C. Successivamente si sostituisce la soluzione di pre-ibridazione con 20 ml della medesima soluzione in cui viene aggiunta la sonda di DNA marcata con α -³²PdCTP (mediante nick translation) e precedentemente denaturata per 20 min a 100°C. Il filtro viene incubato in questa soluzione a 55°C over night. Il filtro viene poi sottoposto a dei lavaggi con la soluzione di ibridazione 0,1X preriscaldata a 55°C per ridurre le ibridazioni aspecifiche. Infine il filtro viene fatto asciugare a temperatura ambiente e viene poi controllata con un Geiger (rivelatore di radioattività) l'intensità del segnale di ibridazione. Il filtro viene poi messo in esposizione con una lastra autoradiografica in apposite cassette e lasciato ad una temperatura di -80°C per un tempo che varia in funzione dell'intensità del segnale rivelato.

4.19 Primer extension

Questa tecnica è utilizzata per mappare e quantizzare le estremità 5' dello RNA. Lo RNA da testare è ibridato con un eccesso di DNA primer a singolo filamento marcato al 5'. E' utilizzata poi la trascrittasi inversa per estendere questo primer e produrre il cDNA complementare all' RNA stampo. Gli oligonucleotidi sono stati marcati come descritto precedentemente.

Si pongono 100 ng di DNA primer con 50µg di RNA totale. Si aggiunge 0.1 volumi di sodio acetato (pH 5.2) e 2.5 volumi di etanolo e si pone 30 minuti a -20°C. Si centrifuga a 12.000g per 10 minuti a 4°C. Il pellet è sciacquato con etanolo 70% e lasciato asciugare. Si risospende il pellet in 30 µl di buffer di ibridazione, così costituito:

40 mM di piperazine-N,N'-bis(2-etanesulfonico acido) pH 6.4

1mM EDTA pH 8.0

0.4M NaCl

80% di formammide.

Dopo aver risospeso bene si incuba la soluzione a 85°C per 10 minuti per far denaturare gli acidi nucleici.

Si trasferisce a per l'annealing e si lascia 12 ore.

Si aggiunge 170µl di acqua e 400 µl di etanolo, si lascia 1 ora a -20°C.

Si centrifuga a 12.000g per 15 minuti a 4°C, e il pellet viene sciacquato con 500 µl di etanolo 70%. Si lascia asciugare.

L'ibrido RNA/primer viene risospeso in 20 µl di buffer reverse transcriptase.

Il buffer è costituito da:

50mM di TRIS-HCl pH 7.6

60mM di KCl

10mM di MgCl₂

1mM di ciascun dNTP

1mM di ditiotreitolo

1 unità/µl inibitori di RNase

50µg/ml di actinomicina D

Si aggiungono 50 unità di transcriptase inversa, pipettando bene. Si aggiunge 1 µl di EDTA 0.5M pH8, e 1 µl di DNAase-RNAase free 5 µg/µl. Si incuba per 30 minuti a 37°C. Si addiziona 150 µl di TE pH 7.6 contenente 0.1M di NaCl e 200 µl di fenolo: cloroformio. Si pone al vortex per 30 secondi, e si centrifuga a 12.000g per 5 minuti a temperatura ambiente. Si trasferisce la fase superiore acquosa in un tubo pulito, a cui si aggiunge 500 µl di etanolo. Si precipita 1 ora a -20°C. Si centrifuga 15 minuti a 4°C, si sciacqua il pellet con etanolo 70% e si lascia asciugare. Il pellet è risospeso in 4 µl di TE pH 7.4. Si aggiungono 6 µl di buffer formammide e si pipetta bene.

Il buffer è costituito da:

80% formammide

10mM di EDTA pH 8

1mg/ml di xilene cianolo

1mg/ml di blue di bromofenolo

Si incuba 5 minuti a 95°C per denaturare e si pone immediatamente in ghiaccio. Il DNA marcato è analizzato per elettroforesi su un gel di poliacrilammide/urea 7M. Per determinare la dimensione si utilizza il marker V della Boeringer Mannheim e si carica anche la sequenza di DNA preparata con gli stessi oligonucleotidi.

4.20. Elenco delle soluzioni utilizzate:

TAE 50X : 2M Tris-acetato; 0.05M EDTA pH 8.0

TBE 5X : 0.445M Tris; 0.445M Acido borico; 0.01M EDTA

TE 10X : 0.1M Tris-Cl pH 7.0; 0.01M EDTA pH 8.0

LB : 1% Cloruro di sodio; 1% Triptone; 0.5% Estratto di lievito

SSC 20X : 3M Cloruro di sodio; 0.3M Sodio citrato tribasico

RB 10X (pH 7.0) : 0.2M Mops; 0.05M Sodio acetato; 0.01M EDTA

PEG 13%-Sodio cloruro 1.6M

Sol A : 50mM Tris-Cl pH 7.5; 10mM EDTA

Sol B : 0.2M Idrossido di sodio ; 1% SDS

Sol C : 1.32M Potassio acetato pH 4.8

Sol III : 3M Potassio-acetato pH 4.2

H2O DEPC : H2O distillata, 0.1% Dietilpirocarbonato.

Buffer fosfato : 1M NaH₂PO₄-Na₂HPO₄ pH 7.0

Blu di metilene : 0.04% blu di metilene; 0.5M Sodio-acetato pH 5.2

GTE : 25mM Tris, 10mM EDTA, 50mM Glucosio.

Cloroformio isoamilico : cloroformio-alcool isoamilico 24:1

Buffer di omogeneizzazione per l'estrazione di RNA da adulti ed embrioni:

50mM Tris-Cl pH 7.5;

60mM Sodio cloruro;

20mM EDTA pH 8.0;

H2O DEPC a volume;

1% SDS

Buffer di omogeneizzazione per l'estrazione di RNA da larve:

20mM Tris-Cl pH 7.5;

0,2M Sodio cloruro;

20mM EDTA pH 8.0;

H2O DEPC a volume

2% SDS

Buffer di omogeneizzazione per l'estrazione del DNA genomico:

10mM Tris-Cl (pH 7.5);

60mM Sodio cloruro;

10mM EDTA;

0.15mM Spermina

0.15mM Spermidina

5.BIBLIOGRAFIA

- Accardo M.C., Giordano E., Riccardo S., Digilio F.A., Iazzetti G., Calogero R.A. and Furia M. (2004).** RNomics: a computational search for box C/D snoRNA genes in the *D. melanogaster* genome. *Bioinformatics*, 20: 3293-3301.
- Adams M.D., Celniker S.E. et al (2000).** The genome sequence of *Drosophila melanogaster*. *Science*, 287: 2185-2195.
- Allmang C., Kufel J. et al. (1999).** Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.*, 18: 5399-5410.
- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990).** Basic local alignment search tool. *J. Mol. Biol.*, 5: 403-410.
- Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G. (1997).** A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*. 1997. 90(1):43-53.
- Ambros V. (2004)** The functions of animal microRNAs. *Nature*. 16;431(7006):350-5. Review
- Andersen K, Wilke-Douglas M. (1987).** Genetic and physical mapping and expression in *Pseudomonas aeruginosa* of the chromosomally encoded ribulose biphosphate carboxylase genes of *Alcaligenes eutrophus*. *J Bacteriol.*;169(5):1997-2004.
- Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. (2001).** Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol*. 26;11(12):941-50.
- Ariel I, Sughayer M, Fellig Y, Pizov G, Ayesh S, Podeh D, Libdeh BA, Levy C, Birman T, Tykocinski ML, de Groot N, Hochberg A. (2000)** The imprinted H19 gene is a marker of early recurrence in human bladder carcinoma. *Mol Pathol.*;53(6):320-3.
- Aspegren A., Hinas A. et al. (2004).** Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. *Nucleic Acids Research*, 32: 4646-4656.
- Avner P, Heard E. (2001).** X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet.*; 2(1):59-67.
- Bachellerie J.P. and Cavallé J. (1997).** Guiding ribose methylation of rRNA. *Trends Biochem. Sci.*, 22: 257-261.
- Bachellerie J.P., Cavallé J. and Hüttenhofer A. (2002).** The expanding snoRNA world. *Biochimie*, 84: 775-790.

Bae K, Lee C, Sidote D, Chuang KY, Edery I. (1998). Circadian regulation of a *Drosophila* homolog of the mammalian Clock gene: PER and TIM function as positive regulators..Mol Cell Biol.18(10):6142-51.

Balakin A.G., Smith L. and Fournier M.J. (1996). The RNA world of nucleolus: two major families of small RNAs defined by different box elements with related functions. Cell, 86: 823-834.

Ballarino M., Morlando M., Pagano F., Fatica A., Bozzoni I. (2005). The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in *Saccharomyces cerevisiae*. Mol. Cell. Bio., 25 : 5396-5403.

Barneche F., Gaspin C., Guyot R., Echeverria M. (2001). Identification of 66 box C/D snoRNAs in *Arabidopsis thaliana*: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2'-O-methylation sites. J Mol Biol., 311: 57-73.

Bartel DP, Chen CZ. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet.5(5):396-400. Review

Bell AC, Felsenfeld G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. Nature; 405(6785):482-5.

Bertrand E., Houser-Scott F., Kendall A., Singer R.H. and Engelke D.R. (1998). Nucleolar localization of early tRNA processing. Genes Dev., 12: 2463-2468.

Bond V.C. and Wold B. (1993). Nucleolar localization of myc transcripts. Mol. Cell. Biol., 13: 3221-3230.

Bortolin M.L. and Kiss T. (1998). Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. RNA, 4: 445-454.

Bortolin M.L., Ganot P. and Kiss T. (1999). Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. EMBO J., 18: 457-469.

Brantl S. (2002). Antisense RNAs in plasmids: control of replication and maintenance. Plasmid. 2002 Nov;48(3):165-73.

Brown J.W., Clark G.P., Leader D.J., Simpson C.G., Lowe T. (2001). Multiple snoRNA gene clusters from *Arabidopsis*. RNA, 7: 1817-32.

Brown J.W., Echeverria M. and Liang-Hu Qu. (2003). Plant snoRNAs: functional evolution and new modes of gene expression. Trends Plant Sci., 8: 42-49.

Buchler NE, Gerland U, Hwa T. (2003). On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A.*; 00(9):5136-41.

Bussemakers M. J. (1999). DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.*, 59 : 5975-5979.

Caffarelli E., Fatica A., Prislei S., De Gregorio E., Fragapane P. and Bozzoni I. (1998). Processing of the iron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *EMBO J.*, 15: 1121-1131.

Cahill N.M., Friend K., Speckmann W. and Steitz J.A. (2002). Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J.*, 21: 3816-3828.

Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* 16(21):2733-42. Review.

Cavaillé J. and Bachellerie J.P. (1996). Processing of fibrillarin-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by common stem-box terminal structure. *Biochimie*, 78: 443-456.

Cavaillé J., Buiting K., Kiefmann M., Lalonde M., Brannan C.I., Horsthemke B., Bachellerie J.P., Brosius J. and Hüttenhofer A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *PNAS*, 97: 14311–14316.

Cavaillé J., Vitali P., Basyuki E., Huttenhofer A. and Bachellerie J.P. (2001). A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *The Journal of biological Chemistry*, 276: 26374-26383.

Cawley S., Bekiranov S., Huck H., Kapranov P., Sekinger E., Kampa D., Piccolboni A., Sememtchenko V., Cheng J., Williams A.J., Wheeler R., Wong B., Drenkow J., Yamanaka M., Patel S., Bubaker S., Tammana H., Helt G., Strhl and Gingeras T.R. (2004). Unbiased widespread mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNA. *Cell*, Vol 116: 499-509.

Chanfreau G., Legrain P., Jacquier A. (1998). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J Mol Biol.*, 284: 975-88.

Chen W, Heierhorst J, Brosius J, Tiedge H. (1997a). Expression of neural BC1 RNA: induction in murine tumours. *Eur J Cancer.* 33(2):288-92.

Chen W, Bocker W, Brosius J, Tiedge H. (1997b). Expression of neural BC200 RNA in human tumours. *J Pathol.* 183(3):345-51.

Chen JL, Blasco MA, Greider CW. (2000). Secondary structure of vertebrate telomerase RNA. *Cell.*; 100(5):503-14.

Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M, Rassenti L, Alder H, Volinia S, Liu CG, Kipps TJ, Negrini M, Croce CM. (2005) miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A.* 102(39):13944-9.

Cullen BR. (2004). Derivation and function of small interfering RNAs and microRNAs. *Virus Res.* 102(1):3-9. Review.

Darzacq X., Jàdy B.E., Verheggen C., Kiss A.M., Bertrand E. and Kiss T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *The Embo Journal*, **21**: 2746-2756.

Dennis P.P., Omer A., Lowe T. (2001). A guided tour: small RNA function in Archea. *Molecular Microbiology*, 40: 509-519.

De Turrís V., Di Leva G., Caldarola S., Loreni F., Amaldi F., Bozzoni I. (2004). TOP promoter elements control the relative ratio of intron-encoded snoRNA versus spliced mRNA biosynthesis. *J Mol Biol.*, **344**: 383-94.

Donmez G, Hartmuth K, Luhrmann R. (2004). Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA.* 10(12):1925-33.

Eddy S.R. (2001). Non-coding RNA genes and the modern RNA world. *Nat Rev Genet.*, 2: 919-29.

Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* 20(23):6877-88.

El-Mabrouk N, Lisacek F. (1996). Very fast identification of RNA motifs in genomic DNA. Application to tRNA search in the yeast genome. *J Mol Biol.*; 264(1):46-55.

Enright C.A., Maxwell E.S., Eliceiri G.L., Sollner-Webb B. (1996). 5'ETS rRNA processing facilitated by four small RNAs: U14, E3, U17, and U3. *RNA*, 2: 1094-9.

Erdmann VA, Szymanski M, Hochberg A, Groot N, Barciszewski J. (2000). Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res.* 28(1):197-200.

Filipowicz W. and Pogacic V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol.*, 14: 319-27.

Frank DN, Pace NR. (1998). Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu Rev Biochem.* 67:153-80.

Ganot P., Bortolin M.L., Kiss T. (1997). Site-specific pseudouridine formation in pre-ribosomal RNA is guided by small nucleolar RNAs. *Cell*, 89: 799-809.

Ganot P., Caizergues-Ferrer M., Kiss T. (1997). The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, 11: 941-56.

Geerlings T.H., Vos J.C., and Rauè H.A. (2000). The final step in the formation of 25S rRNA in *Saccharomyces cerevisiae* is performed by 59r 39 exonucleases. *RNA*, 6: 1698-1703.

Gerdes K, Gulyaev AP, Franch T, Pedersen K, Mikkelsen ND. (1997). Antisense RNA-regulated programmed cell death. *Annu Rev Genet.*31:1-31.

Gioia U., Laneve P., Dlakic M., Arceci M., Bozzoni I., Caffarelli E. (2005). Functional characterization of XendoU, the endoribonuclease involved in small nucleolar RNA biosynthesis. *J Biol Chem.*, 280: 18996-9002.

Giordano E., Peluso I., Senger S., Furia M. (1999). Minifly, a *Drosophila* gene required for ribosome biogenesis. *J Cell Biol.*, 144: 1123-33.

Gopalan V, Vioque A, Altman S. (2002). RNase P: variations and uses. *J Biol Chem.*; 277(9):6759-62.

Grosshans H, Slack FJ. (2002). Micro-RNAs: small is plentiful. *J Cell Biol.*; 156(1):17-21.

Hao Y, Crenshaw T, Moulton T, Newcomb E, Tycko B. (1993) Tumour-suppressor activity of H19 RNA. *Nature*;365(6448):764-7.

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*; 405(6785):486-9.

He W, Parker R. (2000). Functions of Lsm proteins in mRNA degradation and splicing. *Curr Opin Cell Biol.*; 12(3):346-50.

He Z, Sontheimer EJ. (2004) "siRNAs and miRNAs": a meeting report on RNA silencing. *RNA*. 10(8):1165-73.

Heiss N.S., Bachner D., Salowsky R., Kolb A., Kioschis P., Poustka A. (2000). Gene structure and expression of the mouse dyskeratosis congenita gene, *dkc1*. *Genomics*, 67: 153-63.

Henras A.K., Dez C. and Henry Y. (2004). RNA structure and function in C/D and H/ACA s(no)RNPs. *Current Opinion in Structural Biology*, 14: 335-343.

Hershberg R, Altuvia S, Margalit H. (2003). A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res.* 31(7):1813-20.

Herskovits AA, Bibi E. (2000). Association of *Escherichia coli* ribosomes with the inner membrane requires the signal recognition particle receptor but is independent of the signal recognition particle. *Proc Natl Acad Sci U S A*; 97(9):4621-6.

Hirose T. and Steitz J.A. (2001). Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *PNAS*, 98: 12914–12919.

Hirose T., Shu M.D., Steitz J.A. (2003). Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol Cell.*, 12: 113-23.

Hu JC, Gross CA. (1985). Mutations in the sigma subunit of *E. coli* RNA polymerase which affect positive control of transcription. *Mol Gen Genet.* 199(1):7-13.

Huang Z.P., Zhou H., Liang D. and Liang-Hu Qu. (2004). Different expression strategy: multiple intronic gene clusters of box H/ACA snoRNA in *Drosophila melanogaster*. *J. Mol. Biol.*, 341: 669-683.

Huang Z.P., Zhou H., He H.L., Chen C.L., Liang D. and Liang-Hu Qu. (2005). Genome – wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs. *RNA*. Vol. 11: 8

Hüttenhofer A, Brosius J, Bachellerie JP. (2002). RNomics: identification and function of small, non-messenger RNAs. *Curr Opin Chem Biol.* 6(6):835-43. Review.

Hüttenhofer A, Schattner P, Polacek N. (2005). Non-coding RNAs: hope or hype? *Trends Genet.* 21(5):289-97. Review.

Hutvagner G, Zamore PD. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. *Science*. 297(5589):2056-60. Epub 2002 Aug 1.

Hutvagner G, Zamore PD. (2002) RNAi: nature abhors a double-strand. *Curr Opin Genet Dev.* 12(2):225-32.

Hüttenhofer A., Brosius J. and Bachellerie J.P. (2002b). RNomics: identification and function of small, non-messenger RNAs. *Current Opinion in Chemical Biology*, 6: 835-843.

Iacoangeli A, Lin Y, Morley EJ, Muslimov IA, Bianchi R, Reilly J, Weedon J, Diallo R, Bocker W, Tiedge H. (2004). BC200 RNA in invasive and preinvasive breast cancer. *Carcinogenesis*. 25(11):2125-33.

Ikemura T, Dahlberg JE. (1973). Small ribonucleic acids of *Escherichia coli*. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis. *J Biol Chem*. 248(14):5024-32.

Ioannidis P, Kottaridi C, Dimitriadis E, Courtis N, Mahaira L, Talieri M, Giannopoulos A, Iliadis K, Papaioannou D, Nasioulas G, Tragas T. (2004) Expression of the RNA-binding protein CRD-BP in brain and non-small cell lung tumors. *Cancer Lett*. 209(2):245-50.

Isfort RJ, Cody DB, Stuard SB, Randall CJ, Miller C, Ridder GM, Doersen CJ, Richards WG, Yoder BK, Wilkinson JE, Woychik RP. (1997). The combination of epidermal growth factor and transforming growth factor-beta induces novel phenotypic changes in mouse liver stem cell lines. *J Cell Sci*. 110 (Pt 24):3117-29.

Jacobson M.R., Cao L.G., Taneja K., Singer R.H., Wang Y.L. and Pederson Y. (1997). Nuclear domains of the RNA subunit of RNase P. *Journal of Cell Science*, 110: 829-837.

Jacobson M.R. and Pederson T. (1998). Localization of signal recognition particle RNA in the nucleolus of mammalian cells. *Proc. Natl. Acad. Sci.*, 95: 7981–7986.

Jacobson M.R. and Pederson T. (1998). A 7-methylguanosine cap commits U3 and U8 small nuclear RNAs to the nucleolar localization pathway. *Nucleic Acid Research*, 26: 756-760.

Jensen R.C., Wang Y., Hardin S.B., Stumph W.E. (1998). The proximal sequence element (PSE) plays a major role in establishing the RNA polymerase specificity of *Drosophila* U-snrRNA genes. *Nucleic Acids Res.*, 26 :616-22.

Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Muller-Tidow C. (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*. 22(39):8031-41.

Jones K.W., Gorzynski K., Hales C.M., Fischer U., Badbanch F., Terns R.M. and Terns M.P. (2001). Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. *The Journal of biological Chemistry*, 276: 38645–38651.

Juan V, Crain C, Wilson C. (2000). Evidence for evolutionarily conserved secondary structure in the H19 tumor suppressor RNA. *Nucleic Acids Res.*; 28(5):1221-7.

Kamine J, Bakker A, Desrosiers RC. (1994). Mapping of RNA transcribed from a region of the Herpesvirus saimiri genome required for oncogenicity. *J Virol.*; 52(2):532-40.

Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanekov VV. (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol.*; 10(14):853-6.

Kasashima K, Nakamura Y, Kozu T. (2004) Altered expression profiles of microRNAs during TPA-induced differentiation of HL-60 cells. *Biochem Biophys Res Commun*. 322(2):403-10.

Kawano M, Oshima T, Kasai H, Mori H. (2002). Molecular characterization of long direct repeat (LDR) sequences expressing a stable mRNA encoding for a 35-amino-acid cell-killing peptide and a cis-encoded small antisense RNA in *Escherichia coli*. *Mol Microbiol.*;45(2):333-49.

Khvorova A, Reynolds A, Jayasena SD.(2003) . Functional siRNAs and miRNAs exhibit strand bias. *Cell*. 115(2):209-16. Erratum in: *Cell*. 2003 Nov 14;115(4):505.

Kiss T., Marshallsay C., Filipowicz W. (1992). 7-2/MRP RNAs in plant and mammalian cells: association with higher order structures in the nucleolus. *EMBO J.*, 11: 3737-46.

Kiss T., Filipowicz W. (1993). Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. *EMBO J.*, 12: 2913-20.

Kiss T, Filipowicz W. (1995). Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev.*, 9: 1411-24.

- Kiss T., Bortolin M.L., Filipowicz W. (1996).** Characterization of the intron-encoded U19 RNA, a new mammalian small nucleolar RNA that is not associated with fibrillarin. *Mol Cell Biol.*, 16: 1391-400.
- Kiss-László Z., Henry Y. and Kiss T. (1998).** Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *The EMBO Journal*, 17: 797–807.
- Kiss T. (2001).** Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *The EMBO Journal*, 20: 3617–3622.
- Kiss T. (2002).** Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell.*; 109(2):145-8.
- Kiss T, Jady BE. (2004).** Functional characterization of 2'-O-methylation and pseudouridylation guide RNAs. *Methods Mol Biol.*; 265:393-408.
- Koch HG, Moser M, Muller M. (2003).** Signal recognition particle-dependent protein targeting, universal to all kingdoms of life. *Rev Physiol Biochem Pharmacol.* 146:55-94.
- Kusch T, Guelman S, Abmayr SM, Workman JL. (2003).** Two *Drosophila* Ada2 homologues function in different multiprotein complexes. *Mol Cell Biol.* 23(9):3305-19.
- Lafontaine D.L., Tollervey D. (1998).** Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem Sci.*, 23: 383-8.
- Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. (2001).** Identification of novel genes coding for small expressed RNAs. *Science*; 294(5543):853-8.
- Lai DZ, Weng SJ, Qi LQ, Yu CM, Fu L, Yu T, Chen W. (2004) .** Construction of two robust CHO cell lines resistant to apoptosis and adapted to protein-free medium by over-expression of Igf-1/bcl-2 or bcl-2/cyclin E genes. 20(1):66-72. Chinese.
- Laneve P., Altieri F., Fiori M.E., Scaloni A., Bozzoni I. and Caffarelli E. (2003).** Purification, cloning, and characterization of XendoU, a novel endoribonuclease involved in processing of intron-encoded small nucleolar RNAs in *Xenopus laevis*. *The Journal of Biological Chemistry*, 278: 13026–13032.
- Lange T.S., Borovjagin et al. (1998).** Conserved boxes C and D are essential nucleolar localization elements of U14 and U8 snoRNAs. *EMBO J.*, 17: 3176-3187.

Lange T.S., Ezrokhi M. et al. (1998). Nucleolar localization elements of *Xenopus laevis* U3 small nucleolar RNA. *Mol. Biol. Cell*, 9: 2973-2985.

Lange T.S. and Gerbi S.A. (2000). Transient nucleolar localization of U6 small nuclear RNA in *Xenopus laevis* oocytes. *Mol. Biol. Cell*, 11: 2419-2428.

Le SV, Chen JH, Currey KM, Maizel JV Jr. (1988). A program for predicting significant RNA secondary structures. *Comput Appl Biosci.*; 4(1):153-9.

Lease RA, Woodson SA. (2004). Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J Mol Biol.*;344(5):1211-23.

Lee RC, Feinbaum RL, Ambros V. (1993) . The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell.*;75(5):843-54.

Lee C.Y., Kim YJ. (1998). Requirement for a functional interaction between mediator components Med6 and Srb4 in RNA polymerase II transcription. *Mol Cell Biol.*;18(9):5364-70

Lee RC, Ambros V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*; 294(5543):862-4.

Lee C.Y., Lee A. et al. (2003). The roles of endonucleolytic cleavage and exonucleolytic digestion in the 5'-end processing of *S. cerevisiae* box C/D snoRNAs. *RNA*, 9: 1362-70.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. (2003). The nuclear RNase III Droscha initiates microRNA processing. *Nature*.425(6956):415-9.

Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. (2004). The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell*. 2004 Jul 9;118(1):69-82.

Levine M., Tjian R. (2003). Transcription regulation and animal diversity. *Nature*. 2003 Jul 10;424(6945):147-51.

Levy S., Avni D. et al. (1991). Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc Natl Acad Sci U S A*, 88: 3319-23.

Liang W.Q., Jeffrey A.C. et al. (1997). The rRNA-processing function of the yeast U14 small nucleolar RNA can be rescued by a conserved RNA helicase-like protein. *Mol. And Cell. Biol.*, 17: 4124-4132.

Liang X.H., Liu L. et al. (2001). Identification of the first trypanosome H/ACA RNA that guides pseudouridine formation on rRNA. *J. Biol. Chem.*, 276: 40313-40318.

Liang D., Zhou H. et al. (2002). A novel gene organization: intronic snoRNA gene clusters from *Oryza sativa*. *Nucleic Acids Res.* 30: 3262-72.

Liu MY, Gui G, Wei B, Preston JF 3rd, Oakford L, Yuksel U, Giedroc DP, Romeo T. (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem.* 1997 Jul 11;272(28):17502-10.

Llave C, Kasschau KD, Rector MA, Carrington JC. (2002). Endogenous and silencing-associated small RNAs in plants. *Plant Cell.* 14(7):1605-19.

Lo P.C. and Mount S.M. (1990). *Drosophila melanogaster* genes for U1 snRNA variants and their expression during development. *Nucleic Acids Res.*, 18: 6971-9.

Looijenga LH, Gillis AJ, van Gurp RJ, Verkerk AJ, Oosterhuis JW. (1997). X inactivation in human testicular tumors. XIST expression and androgen receptor methylation status. *Am J Pathol.* 151(2):581-90.

Lottin S, Adriaenssens E, Dupressoir T, Berteaux N, Montpellier C, Coll J, Dugimont T, Cury JJ. (2002) Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. *Carcinogenesis.* 23(11):1885-95.

Lowe T.M., Eddy S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*; 25(5):955-64.

Lowe T.M. and Eddy S.R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science*, 283: 1168-1171.

Masse E, Gottesman S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A.*;99(7):4620-5.

Mattick JS. (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* 2001 Nov;2(11):986-9

Mattick JS. (2005). The functional genomics of noncoding RNA. *Science.* 309(5740):1527-8.

- Michienzi A., Cagnon L., Bahner I., Rossi J.J. (2000).** Ribozyme-mediated inhibition of HIV 1 suggests nucleolar trafficking of HIV-1 RNA. *PNAS*, 97: 8955–8960.
- Mitchell J.R., Wood E. and Collins C. (1999).** A telomerase component is defective in the human disease dyskeratosis congenita. *Nature*, 402.
- Mitrovich Q.M. and Anderson P. (2005).** mRNA surveillance of expressed pseudogenes in *C. elegans*. *Current biology*, Vol 15: 963-967.
- Møller T, Franch T, Udesen C, Gerdes K, Valentin-Hansen P. (2002).** Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev.*;16(13):1696-706.
- Møller T, Franch T, Hojrup P, Keene DR, Bachinger HP, Brennan RG, Valentin-Hansen P. (2002).** Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell.*;9(1):23-30.
- Moreno E, Yan M, Basler K. (2002).** Evolution of TNF signaling mechanisms: JNK-dependent apoptosis triggered by Eiger, the *Drosophila* homolog of the TNF superfamily. *Curr Biol.* 12(14):1263-8.
- Morey C, Avner P. (2004).** Employment opportunities for non-coding RNAs. *FEBS Lett.* 2004 Jun 1;567(1):27-34.
- Morlando M., Ballarino M., Greco P., Caffarelli E., Dichtand B. and Bozzoni I. (2004).** Coupling between snoRNP assembly and 30 processing controls box C/D snoRNA biosynthesis in yeast. *EMBO J.*, 23: 2392–2401.
- Morrissey J.P. AND Tollervey D. (1993).** Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Molecular and Cellular Biology*, 13: 2469-2477.
- Mouaikel J., Verheggen C., Bertrand E., Tazi J., and Bordonne R. (2002).** Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Molecular Cell.* 9: 891-901.
- Muratoglu S, Georgieva S, Papai G, Scheer E, Enunlu I, Komonyi O, Cserpan I, Lebedeva L, Nabirochkina E, Udvardy A, Tora L, Boros I. (2003).** Two different *Drosophila* ADA2 homologues are present in distinct GCN5 histone acetyltransferase-containing complexes. *Mol Cell Biol.* 23(1):306-21.
- Narayanan A., Speckmann W., Terns R. and Terns M.P. (1999).** Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Molecular Biology of the Cell*, 10: 2131–2147.

Nedea E., He X., Kim M., Pootoolal J., Zhong G., Canadien V., Hughes T., Buratowski S., Moore C.L. and Greenblatt J.(2003). Organization and function of APT, a subcomplex of the yeast Cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *The Journal of Biological Chemistry*, 278: 33000–33010.

MacNamara -Schroeder K.J., Hennessey R.F. et al. (2001). The *Drosophila* U1 and U6 gene proximal sequence elements act as important determinants of the RNA polymerase specificity of small nuclear RNA gene promoters in vitro and in vivo. *J Biol Chem*, 276: 31786-92.

Maden B.E. (1990). The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucl. Acid. Res. Mol. Biol.*, 39: 241-303.

Maxwell E.S. e Fournier M.J. (1995). The small nucleolar RNAs. *Annu. Rev. Biochem.*, 65: 897-934.

Mc Manus TE, Moore JE, Crowe M, Dunbar K, Elborn JS. (2003). A comparison of pulmonary exacerbations with single and multiple organisms in patients with cystic fibrosis and chronic *Burkholderia cepacia* infection. *J Infect.* 46(1):56-9.

Michael MZ, O' Connor SM, van Holst Pellekaan NG, Young GP, James RJ. (2003). Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol Cancer Res.* 1(12):882-91.

Millar JK, Christie S, Anderson S, Lawson D, Hsiao-Wei Loh D, Devon RS, Arveiler B, Muir WJ, Blackwood DH, Porteous DJ. (2001) Genomic structure and localisation within a linkage hotspot of Disrupted In Schizophrenia 1, a gene disrupted by a translocation segregating with schizophrenia. *Mol Psychiatry.* 6(2):173-8.

Mitrovich QM, Anderson P. (2005). mRNA surveillance of expressed pseudogenes in *C. elegans*. *Curr Biol.* 15(10):963-7.

Mutsuddi M, Marshall CM, Benzow KA, Koob MD, Rebay I. (2004) The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with stau1 in *Drosophila*. *Curr Biol.* 14(4):302-8.

Newman D.R, Kuhn J.F., Shanab G.M., Maxwell E.S. (2000). Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA*, 6: 861-79.

Ni J., Tien A.L, Fournier M.J. (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*, 89: 565-73.

Nicoloso M., Qu L.H., Michot B., Bachellerie J.P. (1996). Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species

points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. *J Mol Biol.*, 260: 178-95.

Niewmierzycka A. and Clarke S. (1999). S-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase. *J Biol Chem.*, 274: 814-24.

Ofengand J. and Bakin A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J Mol Biol.*, 266: 246-68.

Olivas WM, Muhrad D, Parker R. (1997). Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs. *Nucleic Acids Res.*;25(22):4619-25.

Omer A.D, Lowe T.M., Russell A.G., Ebhardt H., Eddy S.R., Dennis P.P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science*, 288: 517-22.

Pannone BK, Wolin SL. (2000). Sm-like proteins wRING the neck of mRNA. *Curr Biol.*; 10(13):R478-81.

Pelczar P. and Filipowicz W. (1998). The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol Cell Biol.* 18 :4509-18.

Pellizzoni L., Baccon J., Charroux B., Dreyfuss G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Curr Biol.*, 11: 1079-88.

Perkins D.O., Jeffries C., Sullivan P., (2004). Expanding the “ central dogma” : the regulatory role of non protein coding genes and implication for the genetic liability to schizophrenia. *Mol. Psychiatry*, 10 : 69-78.

Petfalski E., Dandekar T., Henry Y., Tollervey D. (1998). Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol Cell Biol.*, 18: 1181-9.

Qu L.H., Henras A., Lu Y.J., Zhou H., Zhou W.X., Zhu Y.Q., Zhao J., Henry Y., Caizergues-Ferrer M., Bachellerie J.P. (1999). Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol Cell Biol.*, 19: 1144-58.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *403(6772):901-6.*

- Renalier M.H., Nicoloso M., Qu L.H., Bachellerie J.P. (1996).** SnoRNA U21 is also intron-encoded in *Drosophila melanogaster* but in a different host-gene as compared to warm-blooded vertebrates. *FEBS Lett.*, 379: 212-6.
- Richard P., Darzacq X., Bertrand E., Jady B.E., Verheggen C., Kiss T. (2003).** A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J.*, 22: 4283-93.
- Rivas E, Eddy SR. (2001).** Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics.* 2:8.
- Romeo T. (1998).** Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol Microbiol.*;29(6):1321-30.
- Rozhdestvensky T.S., Tang T.H., Tchirkova I.V., Brosius J., Bachellerie J.P., Huttenhofer A. (2003).** Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, 31: 869-77.
- Runte M., Hüttenhofer A., Gross S., Kiefmann M., Horsthemke B., Buiting K. (2001).** The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum Mol Genet.*, 10: 2687-700.
- Runte M., Kroisel P.M., Gillessen-Kaesbach G., Varon R., Horn D., Cohen M.Y., Wagstaff J., Horsthemke B., Buiting K. (2004).** SNURF-SNRPN and UBE3A transcript levels in patients with Angelman syndrome. *Hum Genet.*, 114: 553-61.
- Russell A.G., Schnare M.N., Gray M.W. (2004).** Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. *RNA.*, 10: 1034-46.
- Ruvkun G. (2001).** Molecular biology. Glimpses of a tiny RNA world. *Science*; 294(5543):797-9.
- Samarsky D.A., Fournier M.J., Singer R.H., Bertrand E. (1998).** The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J.*, 17: 3747-57.
- Samarsky D.A. and Fournier M.J. (1998).** Functional mapping of the U3 small nucleolar RNA from the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol.*, 18: 3431-44.
- Savino R., Hitti Y., Gerbi S.A. (1992).** Genes for *Xenopus laevis* U3 small nuclear RNA. *Nucleic Acids Res.*, 20: 5435-42.

Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell*. 115(2):199-208.

Seffens W, Digby D. (1999). mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Res.*; 27(7):1578-84.

Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, Bachellerie JP, Ferguson-Smith AC, Cavaille J. (2003). Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat Genet*. 34(3):261-2.

Sharma K. and Tollervey D. (1999). Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. *Mol Cell Biol.*, 19 :6012-9.

Sleutels F, Zwart R, Barlow DP. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*; 415(6873):810-3.

Smith C.M., Steitz J.A. (1997). Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, 89: 669-72.

Smith C.M., Steitz J.A. (1998). Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol Cell Biol.*, 18: 6897-909.

Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC. (2000). The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol Cell Biol.*; 20(1):312-8.

Snyder M. and Gerstein M. (2003). Genomics. Defining genes in the genomics era. *Science*, **300: 258-60.**

Speckmann W.A., Terns R.M., Terns M.P. (2000). The box C/D motif directs snoRNA 5'-cap hypermethylation. *Nucleic Acids Res.*, 28: 4467-73.

Storz G. (2002). An expanding universe of noncoding RNAs. *Science*; 296(5571): 1260-3.

Storz G, Altuvia S, Wassarman KM. (2005). An abundance of RNA regulators. *Annu Rev Biochem*. 2005;74:199-217.

Szewczak L.B., DeGregorio S.J., Strobel S.A., Steitz J.A. (2002). Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem. Biol.*, 9: 1095-107.

Szymanski M, Barciszewski J. (2002). Beyond the proteome: non-coding regulatory RNAs. *Genome Biol.* 2002;3(5): reviews0005.

Tam W., Ben-Yehuda D., Hayward W.S. (1997). Bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. *Mol. Cell. Biol.*, 17 : 1490-1502.

Tollervey D., Lehtonen H., Jansen R., Kern H., Hurt E.C. (1993). Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell.*, 72: 443-57.

Tollervey D. (1987). A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J.*, 6: 4169-75.

Tollervey D. (1996). Small nucleolar RNAs guide ribosomal RNA methylation. *Science*, 273: 1056-7.

Tran E., Brown J., Maxwell E.S. (2004). Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus? *Trends Biochem. Sci.*, 29: 343-50.

Tycowski K.T., Shu M.D., Steitz J.A. (1993). A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes Dev.*, 7: 1176-90.

Tycowski K.T., Shu M.D., Steitz J.A. (1996). A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, 379: 464-6.

Tycowski KT, You ZH, Graham PJ, Steitz JA. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol Cell.*, 2: 629-38.

Tycowski K.T., Steitz J.A. (2001). Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *Eur J Cell Biol.*, 80: 119-25.

Tycowski K.T., Aab A., Steitz J.A. (2004). Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr Biol.*, 14: 1985-95.

Ulaner GA, Yang Y, Hu JF, Li T, Vu TH, Hoffman AR (2003). CTCF binding at the insulin-like growth factor-II (IGF2)/H19 imprinting control region is insufficient to regulate IGF2/H19 expression in human tissues. *Endocrinology.* 10:4420-6.

Uliel S., Liang X.H., Unger R., Michaeli S. (2004). Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions. *Int J Parasitol.*, 34: 445-54.

Verheggen C., Mouaikel J., Thiry M., Blanchard J.M., Tollervey D., Bordonne R., Lafontaine D.L., Bertrand E. (2001). Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain. *EMBO J.*, 20: 5480-90.

Verheggen C., Lafontaine D.L., Samarsky D., Mouaikel J., Blanchard J.M., Bordonne R., Bertrand E. (2002). Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J.*, 21: 2736-45.

Villa T., Ceradini F., Bozzoni I. (2000). Identification of a novel element required for processing of intron-encoded box C/D small nucleolar RNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol.*, 20: 1311-20.

Vitali P., Royo H., Seitz H., Bachellerie J.P., Hüttenhofer A., Cavaille J. (2003). Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.*, 31: 6543-51.

Vitali P., Basyuk E., Le Meur E., Bertrand E., Muscatelli F., Cavaille J., Hüttenhofer A. (2005). ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J Cell Biol.*, 169: 745-53.

Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Hüttenhofer A, Wagner EG. (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*; 31(22):6435-43.

Vulliamy T., Marrone A., Goldman F., Dearlove A., Bessler M., Mason P.J., Dokal I. (2001). The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature*, 413: 432-5.

Wagner H, Brantl S. (1998). Kissing and RNA stability in antisense control of plasmid replication. *Trends Biochem Sci.* ; 23(12):451-4.

Wassarman KM, Zhang A, Storz G. (1999). Small RNAs in *Escherichia coli*. *Trends Microbiol.* 1999 Jan;7(1):37-45.

Wassarman KM, Storz G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell.*; 101(6):613-23.

Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.*;15(13):1637-51.

Watkins N.J., Dickmanns A., Luhrmann R. (2002). Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol.*, 22: 8342-52.

Weinstein L.B. and Steitz J.A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol.*, 11: 378-84.

Wilderman PJ, Sowa NA, FitzGerald DJ, FitzGerald PC, Gottesman S, Ochsner UA, Vasil ML. (2004). Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc Natl Acad Sci USA.*;101(26):9792-7.

Williams KP, Bartel DP. (1996). Phylogenetic analysis of tmRNA secondary structure. *RNA.* 1996 Dec;2(12):1306-10.

Williams KP. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.*; 30(4):866-75.

Yang C.Y., Zhou H., Luo J., Qu L.H. (2005). Identification of 20 snoRNA-like RNAs from the primitive eukaryote, *Giardia lamblia*. *Biochem Biophys Res Commun.*, 328: 1224-31.

Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T. (2000). Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140. *Mol Biol Cell.*, 11: 567-77.

Yuan G., Klambt C., Bachellerie J.P., Brosius J., Huttenhofer A. (2003). RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.*, 31: 2495-507.

Zebarjadian Y., King T., Fournier M.J., Clarke L., Carbon J. (1999). Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol Cell Biol.*, 19: 7461-72.

Zeng Y, Yi R, Cullen BR. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A.* 100(17):9779-84. Epub 2003 Aug 5.

Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G. (2002). The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell.*;9(1):11-22.

Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. (2003). Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol.*;50(4):1111-24.

Zhang Y, Zhang Z, Ling L, Shi B, Chen R. (2004). Conservation analysis of small RNA genes in Escherichia coli. *Bioinformatics*. 2004 Mar 22;20(5):599-603.

Zhou H, Chen YQ, Du YP, Qu LH. (2002). The Schizosaccharomyces pombe mgU6-47 gene is required for 2'-O-methylation of U6 snRNA at A41. *Nucleic Acids Res*. 30(4):894-902.