UNIVERSIDADE DE SÃO PAULO INSTITUTO DE QUÍMICA

Programa de Pós Graduação em Ciências Biológicas (Bioquímica)

ANA CAROLINA AYUPE DE OLIVEIRA

Biogênese, estabilidade e localização sub-celular de RNAs não-codificadores longos expressos em regiões intrônicas do genoma humano

Versão corrigida da Tese conforme resolução CoPGr 5890 O original se encontra disponível na Secretaria de Pós-Graduação do IQ-USP

> SÃO PAULO Data do Depósito na SPG: 14/02/2012

ANA CAROLINA AYUPE DE OLIVEIRA

Biogênese, estabilidade e localização sub-celular de RNAs não-codificadores longos expressos em regiões intrônicas do genoma humano

Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Doutor em Ciências (Bioquímica)

Área de Concentração: Bioquímica

Orientador: Dr. Eduardo Moraes Rego Reis

São Paulo

Ana Carolina Ayupe de Oliveira

Biogênese, estabilidade e localização sub-celular de RNAs não-codificadores longos expressos em regiões intrônicas do genoma humano

Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Doutor em Ciências (Bioquímica)

Aprovado em:		
Banca Examinadora		
Prof. Dr.		
Instituição:		
Assinatura:		
Prof. Dr	 	
Instituição:		
Assinatura:		
Prof. Dr	 	
Instituição:		
Assinatura:		
Prof. Dr.		
Instituição:		
Assinatura:		
Prof. Dr	 	
Instituição:		
Assinatura:		

AGRADECIMENTOS

Gostaria de agradecer, primeiramente, ao meu querido Deus, que sempre esteve ao meu lado e nunca me desemparou.

Aos meus pais, Antonio e Angela, que não mediram esforços para me apoiar e incentivar, sempre que precisei.

Aos meus irmãos, Tatiane, Fabrícius e Juliana pela amizade constante.

Ao meu marido, Felipe, pelo companherismo, paciência e suporte em todos os momentos. Você foi essencial para eu ter chegado até aqui.

Ao Irzo e a Célia, Patrícia e Armando, Júnior e Adrina pelas orações e palavras de apoio.

Ao Dr. Eduardo Reis, pela orientação e ensinamentos durante meu período de doutoramento.

Ao Dr. Sergio Verjovski-Almeida, pela participação durante a execução do meu projeto de doutorado, com críticas, sugestões e a permissão para uso da estrutura de seu laboratório.

Aos amigos do laboratório, Lauren, Giulliana, Carlos, Murilo, Letícia, Santiago, Kleber, Dinar, Ângela, Kátia, Mariana, Otto, Bianca, Esther e Jefferson por estarem sempre dispostos a ajudar e por tornarem o ambiente de trabalho tão prazeroso.

À Ana Tahira, Yuri, Vinícius, Rodrigo por todo o empenho e ajuda em situações decisivas desta jornada.

À Ana Paula, que inúmeras vezes tornou minha vida muito mais fácil.

À Bruna, pela colaboração durante um período do desenvolvimento desta tese.

À Dra. Aline da Silva, pela apresentação ao programa de pós-graduação do Instituto de Química.

À Dra. Mari Cleide Sogayar e à sua aluna Maria Fernanda Forni por me ensinarem a manipulação de cultura de células.

À Dra. Bettina Malnic e à Jussara Michaloski pelos ensinamentos de hibridização in situ.

À Dra. Carla Rosenberg por permitir o uso do equipamento *scanner* Agilent e à Tatiane Rodrigues pela ajuda.

À Ana Cláudia Carreira e ao Marcos Demasi que me ajudaram na execução de muitos protocolos novos.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro.

A todos os funcionários do Instituto de Química pelos serviços prestados.

RESUMO

Oliveira, A. C. A. **Biogênese, estabilidade e localização sub-celular de RNAs não-codificadores longos expressos em regiões intrônicas do genoma humano.** 2012. 148p. Tese de Doutorado - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

Trabalhos recentes indicam que a maior parte do transcriptoma de células de mamíferos é composto por RNAs não-codificadores de proteínas (ncRNAs). Nosso grupo tem identificado e caracterizado ncRNAs longos (>200 nt), sem splicing, expressos em regiões intrônicas de genes codificadores de proteína. Contudo, a biogênese, processamento e localização subcelular desta classe de RNAs permanecem desconhecidos. Este trabalho teve como objetivos i) investigar a contribuição da RNA Polimerase II (RNAP II) na transcrição de ncRNAs intrônicos, ii) avaliar a meia-vida destes ncRNAs em relação a mRNAs, e iii) verificar a distribuição sub-celular de ncRNAs intrônicos. Os resultados obtidos indicaram que ncRNAs intrônicos são predominantemente transcritos pela RNAP II a partir de regiões promotoras funcionalmente semelhantes as que controlam a transcrição de mRNAs. Ensaios de estabilidade revelaram que, em média, ncRNAs intrônicos possuem meia-vida igual ou maior (3,4h a 4,2h) do que mRNAs (3,1h). A maior parte dos ncRNAs intrônicos possui estrutura cap 5', sugerindo que sejam estabilizados para desempenhar papéis na biologia da célula que não dependam de um rápido turnover. A maior parte dos ncRNAs intrônicos é exportada para o citoplasma, indicando que devam exercer alguma função biológica neste compartimento. Em conjunto, este trabalho fornece informações novas a respeito da biogênese, estabilidade e localização sub-celular ncRNAs intrônicos expressos em células humanas, contribuindo para avançar o conhecimento sobre esta classe de transcritos celulares.

Palavras-chaves: RNAs não-codificadores intrônicos, transcrição eucariótica, estabilidade de RNAs, localização sub-celular de RNAs, oligoarranjos de DNA, transcriptoma.

ABSTRACT

Oliveira, A. C. A. **Biogenesis, stability and sub-cellular localization of long non-coding RNAs expressed in intronic regions of the human genome**. 2012. 148p. Tese de Doutorado - PhD Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Recent studies have shown that most of the mammalian transcriptome is comprised of noncoding RNAs (lncRNAs). Our group has identified and characterized long (>200 nt), unspliced lncRNAs expressed in intronic regions of protein coding genes. However, the biogenesis, processing, stability and subcellular localization of members from this RNA class remain unknown. The aims of this work were i) to investigate the contribution of RNA Polymerase II (RNAP II) to the transcription of intronic, ii) to evaluate the half-life of these ncRNAs relative to mRNAs, and iii) determine their subcellular distribution. Our results indicate that intronic ncRNAs are predominantly transcribed by RNAP II from promoter regions functionally similar to those that control the transcription of mRNAs. Stability assays revealed that intronic ncRNAs have an average half-life equal or greater (3.4h to 4.2h) than mRNAs (3.1h). The majority of intronic ncRNAs have 5' cap modification suggesting that these transcripts are stabilized, possibly to exert roles in the biology of the cell that does not depend on a rapid turnover. Although intronic ncRNAs do not encode proteins, most of these transcripts are transported to the cytoplasm which indicates that they may perform some biological function in this compartment. Altogether, this study reveals with novel information regarding the biogenesis, stability and subcellular localization of intronic ncRNAs expressed in human cells, thus contributing to advance the knowledge on this class of cellular transcripts.

Key words: intronic non-coding RNAs, eukaryotic transcription, RNA stability, RNA subcellular localization, DNA oligoarrays, transcriptome.

Lista de abreviaturas e siglas

5'-Exo: Terminator 5'-Phosphate -Dependent Exonuclease

ARE: AU-Rich Element

Cap 5': 7-metilguanosina na extremidade 5' de RNAs

CAGE: Cap Analyses of Gene Expression

CBC: Cap Binding Complex CCL: Cell Cycle Length

cDNA: DNA complementar

ChIP-seq: Chromatin Imunoprecipitation followed by DNA sequencing

Ct: Cycle Threshold

Cy3: Cyanine -3-CTP

Cy5: Cyanine -5-CTP

DD: Diferencialmente Detectado

DNA: Deoxyribonucleic Acid

dsRNA: double strand RNA

EC: Enriquecido no Citoplasma

EJC: Exon Junction Complex

EN: Enriquecido no Núcleo

ENCODE: Encyclopedia of DNA Elements Consortium

EST: Expressed Sequence Tag

FANTOM: Functional Annotation of Mouse Consortium

FDR: False Discovery Rate

GO: Gene Ontology

H3K4me3: trimetilação da lisina 4 da histona 3

KS: teste Kolmogorov-Sminorv

lincRNAs: long intergenic noncoding RNAs

Lowess: Locally Weighted Scatterplot Smoothing

miRNA: microRNA

MPSS: Massively Parallel Signature Sequencing

mRNA: RNA mensageironcRNA: RNA não codificador de proteína

NDD: Não Diferencialmente Detectado

NE: Não Enriquecido

nt: nucleotídeos

ORF: Open Reading Frame

PCR: Polymerase Chain Reaction

PET: Paired-End Tags

PIN: ncRNA Parcialmente INtrônico

piRNAs: piwi-interacting RNAs

qPCR: PCR quantitativa

RACE: Rapid Amplification of cDNA Ends

RIN: RNA Integrity Number

RISC: RNA Induced Silencing Complex

RNA: Ribonucleic Acid

RNAP: RNA Polimerase

RNA-FISH: RNA Fluorescence In Situ Hybridization

RNP: Ribonucleoproteínas

rRNA: RNA ribossomal

RT: Reverse Transcription

SAGE: Serial Analysis of Gene Expression

SAM: Significance Analysis of Microarrays

siRNA: small interfering RNA

snRNA: small nuclear RNA

snoRNA: small nucleolar RNA

TAP: Tobacco Acid Pyrophosphatase

TIN: ncRNA $\underline{\mathbf{T}}$ otalmente $\underline{\mathbf{IN}}$ trônico

tRNA: RNA transportador

UTR: mRNA untranslated region

ÍNDICE

1	. INTRODUÇÃO	12
	1.1 A complexidade do transcriptoma eucariótico	12
	1.2 RNAs não-codificadores longos em eucariotos	16
	1.3 RNA Polimerases de eucariotos e seus produtos	18
	1.4 Processo de transcrição pela RNA Polimerase II	21
	1.5 Localização sub-celular de mRNAs e de ncRNAs em eucariotos	22
	1.6 ncRNAs longos e regulação da expressão gênica em eucariotos	28
	1.6. 1 Regulação transcricional de mRNAs	28
	1.6.2 Regulação pós-transcricional	33
	1.7 Vias de degradação de mRNAs em eucariotos	36
	1.8 Estabilidade de mRNAs e de ncRNAs em eucariotos	39
2	. OBJETIVOS	42
	2.1 Objetivo geral	42
	2.2 Objetivos específicos	42
3	. MATERIAL E MÉTODOS	43
	3.1 Linhagem celular humana e condições de cultivo	43
	3.2 Fracionamento celular	44
	3.3 Extração de RNA, purificação e tratamento com DNase	44
	3.4 Avaliação da qualidade do RNA	45
	3.5 Tratamento enzimático para análise de presença da modificação cap 5'	46
	3.6 Desenho de oligonucleotídeos iniciadores (<i>Primers</i>)	46
	3.7 Transcrição reversa e síntese de cDNA	47
	3.8 Reação da Polimerase em Cadeia (<i>Polymerase Chain Reaction -</i> PCR)	47
	3.8.1 PCR	47
	3.8.2 PCR quantitativa em tempo real	48
	3.9 Northern blot	49
	3.10 Análise da expressão gênica através de microarranjos de oligonucleotídeos (ou	
	oligoarranjos)	50
	3.10.1 <i>Design</i> do oligoarranjo	
	3.10.2 Síntese, amplificação e marcação do RNA	
	3.10.3 Hibridização	
	3.10.4 Processamento de Imagem	
	3.10.5 Tratamento dos dados	53

3.10.6 Métodos estatísticos
3.10.7 Métodos de enriquecimento de categorias gênicas54
3.10.8 Métodos de bioinformática5.
4. RESULTADOS55
4.1 RNAs não-codificadores intrônicos expressos em células HeLa59
4.2 Estimativa em larga-escala da estabilidade de ncRNAs intrônicos em células HeLa6
4.2.1 Estimativa da meia-vida de ncRNAs intrônicos e mRNAs
4.2.2 Análise de enriquecimento de categorias gênicas entre transcritos com meia-vida curta (< 2,5h) ou longa (> 3,5h)
4.3 Avaliação da contribuição da RNA Polimerase II na biogênese de ncRNAs intrônicos 79
4.4 Avaliação da presença de cap 5' em ncRNAs intrônicos
4.5 Análise da localização sub-celular de ncRNAs intrônicos
5. DISCUSSÃO
5.1 Meia-vida de ncRNAs intrônicos e de mRNAs codificadores de proteínas10-
5.2 RNA Polimerase II e a biogênese de ncRNAs intrônicos
5.3 ncRNAs intrônicos e a modificação cap 5'
5.4 Localização sub-celular de mRNAs e ncRNAs intrônicos
6. CONCLUSÕES12:
7. REFERÊNCIAS
8. ANEXOS

1. INTRODUÇÃO

1.1 A complexidade do transcriptoma eucariótico

Até recentemente, acreditava-se que o transcriptoma eucariótico era composto por RNAs mensageiros, elementos intermediários que transferem para o citoplasma a informação genética para a biossíntese de proteínas, e por RNAs transportadores (tRNAs), RNAs ribossomais (rRNAs), e pequenos RNAs nucleares (*small nuclear* RNAs, snRNA; *small nucleolar* RNAs, snoRNA), que exercem funções estruturais e catalíticas no processamento do mRNA ou na tradução protéica. Entretanto, um nível maior de complexidade do transcriptoma começou a ser evidenciado a partir da descoberta dos *small interfering* RNAs (siRNAs) e microRNAs (miRNAs).

O primeiro microRNA, lin-4, foi identificado em 1993 como um regulador endógeno de genes que controlam o desenvolvimento pós-embrionário em *Caenorhabditis elegans* (Lee et al. 1993). Em 1999, foi mostrado que o silenciamento de RNAs em plantas era acompanhado pelo aparecimento de uma sequência de 20-25 nucleotídeos complementares à sequência do alvo silenciado (Tomari and Zamore 2005). Cinco anos mais tarde, Mello e colaboradores reportaram que RNAs exógenos dupla fita (dsRNA) eram capazes de silenciar genes específicos através de um mecanismo chamado interferência de RNA (RNAi) (Mello and Conte 2004). Pouco tempo depois foi documentada a conversão de dsRNAs em siRNAs contendo 21-23 nucleotídeos. Em 2001, o termo miRNA foi introduzido (Lagos-Quintana et al. 2001) e passou a designar uma classe ampla de pequenos RNAs reguladores amplamente representada em animais e plantas. A partir deste ponto, duas categorias de pequenos RNAs passaram a ser caracterizadas em detalhe quanto ao seu papel na regulação pós-transcricional da expressão gênica: miRNAs, que atuam como reguladores de genes endógenos, e siRNAs,

que exercem uma função na defesa da integridade do genoma em resposta a ácidos nucléicos invasores como vírus, transposons e transgenes (Carthew and Sontheimer 2009).

A identificação dos siRNAs e miRNAs marcou apenas o início de uma série de descobertas que têm promovido revisões no conceito do "dogma central da biologia molecular" originalmente proposto por Francis Crick (Crick 1970). Segundo este conceito o fluxo da informação genética se dá primordialmente no sentido DNA→RNA→proteína. Os RNAs têm um papel de elemento intermediário, enquanto as proteínas seriam responsáveis por exercer a maior parte das funções estruturais, catalíticas e regulatórias das células.

Análises genômicas em larga escala têm revelado que a complexidade biológica de um organismo não é diretamente relacionada ao número de genes codificadores de proteínas de seu genoma (Consortium 1998; Adams et al. 2000; Lander et al. 2001; Waterston et al. 2002). Por exemplo, vertebrados recentes, como humanos e camundongos, têm aproximadamente 20.000-25.000 genes, enquanto eucariotos ancestrais, como Drosophila e Caenorhabditis elegans, têm um número comparável de genes codificadores de proteínas, 13.000 e 19.000, respectivamente (Consortium 1998; Adams et al. 2000; Lander et al. 2001; Waterston et al. 2002). Parte deste paradoxo pode ser explicada, por exemplo, pelo aumento das taxas de splicing alternativo, do uso de promotores e sítios de terminação alternativos e dos eventos de edição de RNAs em organismos mais complexos, permitindo assim a expressão de um conjunto maior de isoformas protéicas (Nagasaki et al. 2005; Taft et al. 2007; Farajollahi and Maas 2010). Além disso, observa-se que o aumento da complexidade biológica está positivamente associado ao aumento da proporção de sequências não-codificadoras presentes no genoma dos organismos, que pode variar de 0,25% do total no genoma de procariotos até > 98% no genoma humano (Taft et al. 2007). Estas observações sugerem que a aquisição de novos RNAs não codificadores (ncRNAs) possa fornecer um repertório maior de moléculas

com funções regulatórias que permitiram o desenvolvimento de eucariotos superiores (Taft et al. 2007).

Com a consolidação e popularização de abordagens para o estudo de transcriptomas em larga-escala, a transcrição generalizada do genoma eucariótico e a existência de transcritos relativamente estáveis mapeando em regiões intrônicas e intergênicas têm sido detectadas em diferentes espécies eucarióticas, incluindo fungos, plantas, nematóides, insetos e mamíferos (Amaral and Mattick 2008). Estes estudos incluem a clonagem e sequenciamento de bibliotecas de cDNAs (Okazaki et al. 2002; Yamada et al. 2003; Imanishi et al. 2004; Ota et al. 2004; Seki et al. 2004), técnicas como SAGE (*Serial Analysis of Gene Expression*), CAGE (*Cap Analysis of Gene Expression*) e MPSS (*Massively Parallel Signature Sequencing*) (Chen et al. 2002; Saha et al. 2002; Meyers et al. 2004; Carninci et al. 2005; Jongeneel et al. 2005), *tiling arrays* genômicos (Kapranov et al. 2002; Rinn et al. 2003; Yamada et al. 2003; Bertone et al. 2004; Schadt et al. 2004; Stolc et al. 2004; Cheng et al. 2005; David et al. 2006; He et al. 2007) e mais recentemente sequenciamento em larga escala (Guttman et al. 2010; Kapranov et al. 2010; Almeida et al. 2011; Mercer et al. 2011)

Em 2007, o projeto ENCODE (Encyclopedia of DNA Elements Consortium) analisou intensivamente 1% do genoma humano através de várias técnicas, incluindo tiling arrays genômicos de alta densidade e sequenciamento com alta profundidade de bibliotecas de cDNA, tendo demonstrado que ao menos 90% do DNA genômico é transcrito em diferentes células humanas, sendo que desta fração apenas 1,2% é capaz de codificar para proteínas (Birney et al. 2007; Kapranov et al. 2007). A análise do transcriptoma de organismos eucarióticos ancestrais usando tiling arrays genômicos mostrou que a transcrição generalizada do genoma não é restrita aos mamíferos, mas extende-se a moscas (>85% do genoma de Drosophila melanogaster é transcrito nas primeiras 24h do desenvolvimento embrionário), nematóides (>70% do genoma é transcrito em C. elegans em populações de diferentes

estágios de desenvolvimento) e leveduras (>85% do genoma é expresso meio rico em nutrientes) (Amaral and Mattick 2008). No genoma murino, o projeto FANTOM3 (*Functional Annotation of Mouse 3*) clonou e sequenciou mais de 100.000 sequências de cDNA completas de diversos tecidos e estágios de desenvolvimento, revelando que mais de 70% do genoma de camundongo é transcrito, tendo sido descritos mais de 30.000 novos ncRNAs, incluindo transcritos intrônicos, intergênicos e antissenso não-codificadores longos (Carninci et al. 2005).

As descobertas mencionadas acima têm alterado e expandido a compreensão da organização dos genomas eucarióticos. A observação do aumento da transcrição generalizada nos organismos eucariotos ao longo da escala evolutiva tem sido utilizada como argumento para apoiar a idéia de que os ncRNAs contribuam com funções regulatórias essenciais para o ganho de complexidade, de modo análogo aos mecanismos já bem descritos de *splicing* alternativo, utilização de promotores e sítios de terminação alternativos e modificações póstranscricionais (Taft et al. 2007). Vários estudos têm mostrado que ncRNAs são regulados durante o desenvolvimento (Dinger et al. 2008; Cesana et al. 2011), exibem expressão tecido específica (Ravasi et al. 2006; Nakaya et al. 2007; Mercer et al. 2008), localizam-se em compartimentos subcelulares (Hutchinson et al. 2007; Clemson et al. 2009; Sasaki et al. 2009; Sunwoo et al. 2009), são regulados durante o ciclo celular (Hung et al. 2011; Yang et al. 2012) e são associados com uma variedade de doenças humanas, incluindo o câncer (Reis et al. 2004; Qureshi et al. 2010; Taft et al. 2010; Gibb et al. 2011; Tahira et al. 2011). Em adição, evidências de seleção evolucionária têm sido observadas em alguns ncRNAs (Ponjavic et al. 2007; Guttman et al. 2009; Mestdagh et al. 2010; Braconi et al. 2011).

Contudo, existe um debate intenso na literatura sobre o real significado biológico da transcrição generalizada do genoma, motivado pelo fato de que a maioria dos ncRNAs são

expressos em baixos níveis na célula e a maior parte destes transcritos ainda não foi caracterizada funcionalmente (Struhl 2007; van Bakel et al. 2010).

1.2 RNAs não-codificadores longos em eucariotos

A variedade de estudos de identificação dos RNAs expressos em células de eucariotos, apresentados no item anterior, tem mostrado que o transcriptoma eucariótico é extraordinariamente complexo e consiste de milhares de transcritos sobrepostos, expressos em ambas as fitas, incluindo ncRNAs que localizam-se em regiões intergênicas ou intrônicas, que exibem diferentes sítios de iniciação, terminação, padrões de *splicing* e de expressão em diferentes tipos celulares (Gingeras 2007; Mattick 2011).

Têm sido convencional na literatura classificar ncRNAs em função do tamanho para efeito de sistematização: ncRNAs longos quando maiores que 200 nucleotídeos e ncRNAs curtos, quando menores que 200 nucleotídeos (Clark and Mattick 2011). Entre os ncRNAs curtos (< 200 nt) se incluem miRNAs, siRNAs, *piwi-interacting* RNAs (piRNAs), pequenos RNAs associados a regiões promotoras e de iniciação da transcrição, entre outros (Girard et al. 2006; Carthew and Sontheimer 2009; Fejes-Toth et al. 2009; Taft et al. 2009). Entre os ncRNAs longos incluem-se RNAs com 200 até vários milhares de nucleotídeos, com ou sem *splicing* (Qureshi et al. 2010). Estes transcritos mapeiam em regiões genômicas frequentemente associadas a ligação de fatores de transcrição, indicando que são potencialmente regulados de forma análoga aos mRNAs (Cawley et al. 2004; Birney et al. 2007). Adicionalmente, alguns ncRNAs longos servem de precursores para ncRNAs curtos, como miRNAs e snoRNAs (Mattick and Makunin 2005).

Recentemente, foram identificados ncRNAs longos (>1kb) intergênicos, lincRNAs (long intergenic noncoding RNAs) expressos em células e tecidos de camundongo e humano, a partir da análise de regiões de cromatina transcricionalmente ativa e sequenciamento de

RNA (Guttman et al. 2009; Khalil et al. 2009; Guttman et al. 2010; Cabili et al. 2011). LincRNAs tem expressão tecido-específica, são conservados evolutivamente e alguns têm sido descritos como responsáveis por recrutar e conferir especificidade de alvo a complexos remodeladores da cromatina (Guttman et al. 2009; Khalil et al. 2009; Guttman et al. 2010; Cabili et al. 2011).

Nosso grupo tem focado o estudo na caracterização de uma classe de ncRNAs definida por serem longos, sem *splicing* e localizados em regiões intrônicas do genoma (Reis et al. 2004; Louro et al. 2007). Uma análise utilizando bancos de dados públicos de ESTs (*Expressed Sequence Tags*) e mRNAs revelou a existência de cerca de 55.000 *contigs* mapeando em regiões totamente intrônicas (TIN) e mais de 12.000 *contigs* mapeando em regiões parcialmente intrônicas (PIN), ou seja, apresentando sobreposição com um exon (Nakaya et al. 2007). Foi observado também que, pelo menos 74% dos genes humanos anotados possuem regiões intrônicas transcricionalmente ativas e uma análise de potencial codificador indicou que 98% e 90% dos transcritos TIN e PIN, respectivamente, não possuem potencial para codificar proteínas, provavelmente constituindo-se em ncRNAs (Nakaya et al. 2007).

A partir da análise de ESTs realizada, foi construído um oligoarranjo intron-exon contendo 44 mil sondas, sintetizado sob encomenda para o estudo destes ncRNAs (Nakaya et al. 2007). Com o uso desta plataforma, foram identificadas assinaturas de transcrição intrônica tecido-específicas para fígado, próstata e rim humano (Nakaya et al. 2007). Um subconjunto de transcritos intrônicos, com orientação parcialmente intrônica, apresentou um padrão de expressão correlacionada ao gene codificador de proteína expresso no mesmo *locus*, indicando que os ncRNAs intrônicos podem ter um papel na regulação do padrão de *splicing* alternativo ou ainda, regular a abundância das mensagens codificadoras (Nakaya et al. 2007). De maneira interessante, os ncRNAs TIN antissenso que apresentaram maior expressão foram

transcritos a partir de introns de genes codificadores de proteínas enriquecidos significativamente na categoria de ontologia gênica "Regulação da Transcrição" (Nakaya et al. 2007).

Outro trabalho realizado em nosso grupo, demonstrou que um subconjunto de ncRNAs intrônicos têm sua transcrição diretamente regulada por andrógeno, através da ligação do hormônio à elementos regulatórios ARE (*Androgen Response Element*), de maneira semelhante ao que ocorre com mRNAs (Louro et al. 2007). Outros estudos realizados no laboratório, têm apontado também para um papel dos ncRNAs intrônicos no controle da expressão gênica em doenças como o câncer (Reis et al. 2004; Louro et al. 2007; Brito et al. 2008; Tahira et al. 2011).

1.3 RNA Polimerases de eucariotos e seus produtos

Os genes nucleares em eucariotos são transcritos por três RNAs polimerases (RNAPs), denominadas RNAP I, RNAP II e RNAP III (Raha et al. 2010). A RNAP I é responsável pela transcrição dos RNAs ribossomais, exceto o 5S rRNA (Russell and Zomerdijk 2006). A RNAP III transcreve ncRNAs curtos importantes para a capacidade traducional, como o 5S, rRNA, RNAse P, RNAse MRP e todos os RNAs transportadores (Oler et al. 2010). Em adição, a RNAP III transcreve uma lista crescente de ncRNAs com funções alternativas que conectam esta polimerase a biologia do *splicing* (U6 snRNA), além de RNAs virais (VA-I e VA-II), miRNAs, RNAs derivados de regiões repetitivas do DNA (*Mammalian-wide Interspersed Repeat* (MIR) e elementos Alu), reguladores da transcrição pela RNAP II (7SK e BC2) e RNAs associados a resistência à múltiplas drogas (Vault) (Oler et al. 2010). Os ncRNAs longos BC1, que é específico de roedores e BC200A, que é específico de primatas, também são transcritos pela RNAP III (Oler et al. 2010). Estes dois ncRNAs são localizados especificamente em compartimentos dendritícos pós-sinápticos, onde modulam a síntese local

de proteínas através da repressão da iniciação da tradução dependente de eIF4A (Qureshi et al. 2010). Dezenas de novos sítios de ligação da RNAP III foram identificados em regiões intergênicas e alguns sítios em regiões intrônicas, contudo sua significância ainda não foi determinada (Oler et al. 2010).

A RNAP II produz uma larga variedade de transcritos, sendo os RNAs mensageiros codificadores de proteínas o seu protótipo. Além dos mRNAs, alguns tipos de RNAs não codificadores de proteínas com papéis já bem definidos são transcritos pela RNAP II, como os snRNAs, snoRNAs, a maioria dos miRNAs e muitos ncRNAs ainda pouco conhecidos.

Os snRNAs U1, U2, U4 e U5 são sintetizados pela RNAP II e fazem parte do spliceossomo (Hopper 2006). Os snoRNAs são moléculas muito estáveis processados a partir de regiões intrônicas que tem papel bem estabelecido na modificação de rRNAs (Mattick 2011). Alguns têm mostrado ter expressão tecido específica, contexto-dependente ou, ainda regulada por *imprinting* genômico (Mattick 2011).

Os miRNAs são RNAs não codificadores curtos, com ~23 nucleotídeos, que exercem funções regulatórias através da repressão traducional, desestabilização de mRNAs ou uma combinação de ambos (Carthew and Sontheimer 2009; Inui et al. 2010). Novas descobertas indicam que miRNAs também estão envolvidos na regulação transcricional, juntamente com os fatores de transcrição (Martinez and Walhout 2009). Estes RNAs curtos são transcritos a partir de longos transcritos primários (pri-miRNAs) que podem estar localizados em regiões intrônicas, antissenso ou intergências, os quais são semelhantes a mRNAs, possuindo cap 5' e cauda poli-A (Lee et al. 2004). Cada pri-miRNAs contém uma ou mais regiões intramoleculares de dsRNA (hairpin) que são clivados por um complexo nuclear multiprotéico denominado Microprocessor, cujos componentes centrais são a Drosha, uma endonuclease RNase III, e aproteína ligadora de RNA dupla fita DGCR8/Pasha (Carthew and Sontheimer 2009; Inui et al. 2010). Este processamento gera precursores (pre-miRNAs) de 70

nucleotídeos que são exportados para o citoplasma pela exportina 5 (Bartel and Chen 2004). No citoplasma, o pre-miRNA é reconhecido pela *Dicer*, outra endonuclease RNase III, e por uma proteína ligadora de RNA, TRBP (*TAR* RNA *Binding Protein*) (Carthew and Sontheimer 2009; Inui et al. 2010). A *Dicer* cliva este precursor, gerando um miRNA maduro dupla fita de ~23 nucleotídeos. Os siRNAs também são processados pela *Dicer*, porém a partir de longos dsRNAs e a *Drosha* não está envolvida no processo (Carthew and Sontheimer 2009). Recentemente, uma via de maturação independente da *Dicer* também foi reportada (Cheloufi et al. 2010). Uma vez processados, os miRNA e siRNAs maduros passam a fazer parte do complexo RISC (*RNA Induced Silencing Complex*), tendo a função de direcionar o complexo aos seus respectivos alvos através de interações mediadas por pareamento de base (Carthew and Sontheimer 2009; Inui et al. 2010).

Embora os dados descritos nos itens anteriores demonstrem que a maior parte dos transcritos humanos são ncRNAs longos transcritos a partir de regiões intrônicas e intergênicas, até o momento nenhum trabalho estudou diretamente o mecanismo de síntese destes transcritos. Uma parcela considerável dos ncRNAs longos parece ser transcrita pela RNAP II, já que muitos destes transcritos compartilham características estruturais com mRNAs, como a poliadenilação na cauda 3', *splicing* e estrutura cap 5' (Erdmann et al. 2001; Numata et al. 2003; Willingham and Gingeras 2006; Qureshi et al. 2010). Além disso, experimentos de imunoprecipitação da cromatina acompanhado de *tiling arrays* ou sequenciamento de última geração demonstraram que a RNAP II e muitos sítios de ligação de seus fatores de transcrição, e ainda, modificações de histonas relacionadas à iniciação da transcrição, estão localizados em muitas regiões genômicas distantes de promotores canônicos de genes codificadores de proteínas (Cawley et al. 2004; Birney et al. 2007; Guttman et al. 2009; Rozowsky et al. 2009). Entretanto, estudos recentes apontam para a possibilidade de que uma fração considerável dos ncRNAs não possua estas características.

Cheng e colaboradores, através de experimentos utilizando *tiling arrays* demonstraram que quase metade dos transcritos encontrados em células humanas não são poliadenilados (Cheng et al. 2005) e uma possibilidade é que sejam transcritos pela RNAP III (Mattick 2011). Estudos mais recentes também têm relatado que muitos transcritos não poliadenilados correspondem a RNAs ainda não conhecidos que mapeiam em regiões intrônicas e intergênicas do genoma (Wu et al. 2008; Yang et al. 2011).

1.4 Processo de transcrição pela RNA Polimerase II

A transcrição é uma função central na regulação da expressão gênica que ocorre no núcleo de células eucarióticas em coordenação com outros processos nucleares. Em células eucarióticas, a produção de mRNA maduro que sai do núcleo para ser traduzido em proteínas requer precisa e extensiva modificação do transcrito nascente. Os mRNAs codificadores de proteínas recém-sintetizados (pré-mRNAs) pela RNAP II passam por vários passos de processamento (Luna et al. 2008).

O primeiro passo é a adição da estrutura 7-metilguanosina cap na extremidade 5' do transcrito (cap 5'). Quando o transcrito alcança 20-30 nucleotídeos de extensão, o cap 5' é adicionado, protegendo o transcrito nascente de ser degradado. Em seguida, o cap 5' é ligado pelo complexo CBC (*Cap Complex Binding*), que parece ser o primeiro complexo protéico a se ligar no pré-mRNA, sendo requerido para os passos sub-sequentes de *splicing*, transporte nuclear, tradução e decaimento do mRNA (Aguilera 2005; Carmody and Wente 2009). O próximo evento é o *splicing*, onde um grupo de proteínas é adicionado no sítio de fusão do exon com o intron (Carmody and Wente 2009). Estas proteínas são definidas como EJC (*Exon-Junction Complex*). A adição de cap 5' e o *splicing* são importantes para o recrutamento do complexo TREX, que é altamente conservado e é essencial para a exportação do mRNA (Cheng et al. 2006). O processamento final do pré-mRNA envolve os eventos de

clivagem 3' e poliadenilação. O sítio de poliadenilação é reconhecido na região 3' UTR (*Untranslated Region*), resultando na clivagem do pré-mRNA 20 a 30 bases depois deste sítio pelos fatores CstF, CPSF, CFIm e CFIIm em mamíferos e CF1A, CF1B e CPF, em leveduras. A cauda poli-A é adicionada na extremidade 3' do corte, pela poli-A Polimerase, e ligada pela proteína de ligação poli-A (Carmody and Wente 2009). O processo de poliadenilação facilita a liberação do mRNA do sítio de transcrição e sua exportação através do NPC (*Nuclear Pore Complex*), além de ser importante na sua estabilidade e iniciação da tradução (Moore and Proudfoot 2009).

Alguns pré-mRNAs podem, ainda, sofrer edições pós-transcricionais, como substituições de bases (Farajollahi and Maas 2010). Estas modificações podem gerar ou depletar sítios de *splicing*, regular *splicing* alternativo e influenciar na dinâmica de sítios de *splicing* constitutivo (Farajollahi and Maas 2010). As edições de bases podem também promover retenção nuclear de mRNAs e substituições de aminoácidos durante o processo de tradução, constituindo-se em um processamento muito importante na regulação da expressão gênica (Farajollahi and Maas 2010).

Durante todo o processamento, o pré-mRNA se associa com proteínas ligadoras de RNA formando um complexo ribonucleoproteína (RNP), que especifica seu processamento, transporte, localização sub-celular, tradução e estabilidade (Aguilera 2005; Luna et al. 2008). O RNP, muitas vezes, é parte de uma grande estrutura chamada grânulo de transporte de RNA, que é transportada, por proteínas motoras juntamente com elementos do citoesqueleto, para o destino final do mRNA na célula (Besse and Ephrussi 2008).

1.5 Localização sub-celular de mRNAs e de ncRNAs em eucariotos

Virtualmente todas as células são polarizadas, particionando de forma diferencial seus conteúdos moleculares para uma variedade de organelas, compartimentos e interfaces de

membrana que executam funções biológicas regulatórias especializadas (Lecuyer et al. 2007). Com a descoberta do peptídeo sinal, pensava-se que o direcionamento das proteínas para os seus vários destinos sub-celulares ocorria somente após a tradução (Blobel and Dobberstein 1975). Contudo, mais recentemente, têm sido mostrado que a localização da proteína também pode ser controlada através da localização do mRNA antes de sua tradução (Bashirullah et al. 1998; Czaplinski and Singer 2006).

O processo de localização do mRNA acoplado à sua tradução tem sido reconhecido como um mecanismo poderoso para restringir espacialmente e temporalmente a expressão gênica a sítios específicos dentro de células altamente polarizadas (revisado em (Besse and Ephrussi 2008; Martin and Ephrussi 2009)). A habilidade de restrição da síntese do mRNA ao local adequado é imprescindível, em particular, para proteínas que podem ser tóxicas ou deletérias às células quando expressas ectopicamente. Além disso, a localização do transcrito permite a imposição de muitos níveis de controle da expressão gênica.

Alguns estudos têm revelado que um vasto número de mRNAs ocupam localizações sub-celulares específicas, por exemplo, apical-basal, associado à membrana, centrossomos, microtúbulos, o que indica que a localização do mRNA é um mecanismo importante usado pelas células para estabelecer estruturas e compartimentos funcionalmente distintos (Lecuyer et al. 2007; Besse and Ephrussi 2008).

Uma vantagem de restringir a expressão gênica espacialmente é a possibilidade de se obter alta resolução temporal, dado que um estímulo local pode regular a tradução, ao invés de requerer um sinal para que o mRNA seja transcrito no núcleo e exportado para o citoplasma para iniciar a tradução (revisado em (Besse and Ephrussi 2008; Martin and Ephrussi 2009)). Assim, moléculas de mRNA localizadas podem ser submetidas a muitos ciclos de síntese de proteínas e, ainda, evitar o custo energético de mover cada molécula de

proteína individualmente para o local de sua atuação (revisado em (Besse and Ephrussi 2008; Martin and Ephrussi 2009)).

O direcionamento do mRNA para uma localização específica envolve muitos passos. O "endereço" celular dos transcritos é codificado por elementos regulatórios em *cis*, ou motivos, no RNA. Estes elementos, chamados "elementos de localização" ou "*zipcodes*" são mais frequentemente encontrados na região 3' UTR, todavia, em alguns casos podem ser encontrados na 5' UTR ou na sequência codificadora (Besse and Ephrussi 2008; Martin and Ephrussi 2009).

Os elementos de localização são reconhecidos por proteínas ligadoras de RNAs específicas que, frequentemente, têm funções tanto na localização dos transcritos quanto na regulação de sua tradução (Martin and Ephrussi 2009). Como já mencionado no item anterior, vários estudos indicam que o processamento de pré-mRNAs no núcleo é requerido para o recrutamento de proteínas de ligação de RNA que determinam sua eventual localização no citoplasma (Aguilera 2005; Giorgi and Moore 2007; Luna et al. 2008). O mecanismo de *splicing*, por exemplo, é um passo importante para o transporte dos mRNAs para o citoplasma (Valencia et al. 2008), sendo necessário para o recrutamento de proteínas essenciais no processo de exportação.

Diferente de mRNAs, muitos ncRNAs não são sofrem o processamento de *splicing* (Ravasi et al. 2006; Nakaya et al. 2007), enquanto, contraditoriamente, alguns ncRNAs sofrem *splicing* e permanecem no núcleo (Khalil et al. 2009), não sendo exportados para o citoplasma. Também existem mecanismos independentes de *splicing* para o recrutamento de fatores de transporte o para o RNA nascente, como é o caso de mRNAs *unspliced* de histonas, onde fatores *cis* na sua sequência promovem a ligação de fatores de exportação (Huang and Carmichael 1997; Erkmann et al. 2005). Contudo, ainda não se sabe se tais vias poderiam ser responsáveis pela exportação de ncRNAs (Clark and Mattick 2011).

Uma análise global da localização de mRNAs durante a embrigênese de *Droshophila*, usando RNA-FISH (RNA-*Fluorescence In Situ Hybridization*), revelou que 71% dos genes expressos avaliados apresentaram um padrão de localização sub-celular bem definido (Lecuyer et al. 2007). Os mRNAs foram agrupados em mais de 30 categorias, incluindo RNAs localizados em pólos embriônicos, no citoesqueleto, na maquinaria de divisão celular, no núcleo, em *foci* citoplasmático entre outros. Em vista da frequência e variedade de localização dos mRNAs, os autores sugeriram que as vias de tráfego dos RNAs podem ter um impacto sobre todos os aspectos da arquitetura e função celular (Lecuyer et al. 2007).

Lecuyer e colaboradores também observaram que uma parte dos transcritos expressos foi retida no núcleo durante alguns estágios do desenvolvimento (Lecuyer et al. 2007). Quando o mRNA era observado localizado no núcleo, pouca ou nenhuma expressão de sua proteína podia ser observada na célula, mas em estágios onde a proteína foi expressa abundantemente, pôde-se verificar um padrão difuso de localização do mRNA no citoplasma da célula (Lecuyer et al. 2007). Este exemplo sugere que a retenção nuclear é um mecanismo regulatório adicional na coordenação do momento preciso de expressão da proteína na célula. Relatos de retenção nuclear controlando a tradução do mRNA em proteína e retenção nuclear em fases específicas do desenvolvimento também já foram documentados em mamíferos (Prasanth et al. 2005; Chen and Carmichael 2009).

De maneira interessante, alguns mRNAs localizados parecem ter outros papéis funcionais na célula, além de codificar para proteínas. É o caso, por exemplo, do mRNA VegT, em oócitos de *Xenopus* (Kloc et al. 2005). Este mRNA localiza-se no córtex vegetal, onde interage com filamentos de citoqueratina, juntamente com o ncRNA Xlsirts, sendo essencial para a manutenção do citoesqueleto de citoqueratina e apropriada formação dos grânulos germinativos, propiciando subsequente desenvolvimento da linhagem germinativa (Kloc et al. 2005). De modo similar, um grande número de RNAs, particularmente RNAs

ribossomais, assim como muitos transcritos ainda não caracterizados, se associam com o fuso mitótico (Blower et al. 2005). Os autores observaram que o tratamento com RNAse A causa rompimento no fuso, todavia, tratamento com inibidores de tradução não tem nenhum efeito aparente, argumentando para a possibilidade destes RNAs desenvolverem papéis na organização estrutural do fuso mitótico (Blower et al. 2005).

Alguns ncRNAs já foram associados com estruturas sub-celulares específicas. O ncRNA Gomafu, por exemplo, é encontrado em domínios sub-nucleares em neurônios pósmitóticos e células progenitoras neurais em diferenciação (Sone et al. 2007). Gomafu tem funções associadas ao desenvolvimento da retina e variantes de sequências têm sido relacionadas à susceptibilidade ao infarto do miocárdio em humanos (Ishii et al. 2006; Rapicavoli et al. 2010).

O ncRNA MALATI(*Metastasis Associated Lung Adenocarcinoma Transcript 1*), que tem expressão aumentada em vários tipos de cânceres, é retido no núcleo e localiza-se em sub-estruturas denominadas *speckles* nucleares, que são complexos ribonucleoprotéicos enriquecidos em fatores de *splicing* de pré-mRNA, incluindo snRNPs (*small nuclear ribonucleoprotein particles*) e proteínas SR (*Serine/arginine-Rich proteins*) (Lamond and Spector 2003; Hutchinson et al. 2007; Lin et al. 2007). MALAT1 está envolvido na regulação de *splicing* alternativo, sequestrando vários fatores de *splicing* SR para os *speckles* (Tripathi et al. 2010). A depleção de MALAT1 altera a localização e atividade de fatores de *splicing*, alternado o padrão de *splicing* alternativo para um grupo de pré- mRNAs (Tripathi et al. 2010). MALAT1 também controla processos complexos como a invasão de trofoblastos na parede uterina durante o desenvolvimento e sinaptogênese (Tseng et al. 2009; Bernard et al. 2010).

O ncRNA NEAT1 (Nuclear Enriched Abundant Transcript 1) ou MEN-epsilon/beta (Multiple Endocrine Neoplasia epsilon/beta) localiza-se na sub-estrutura nuclear

paraspeckles, que é localizada adjacente aos speckles e está envolvida na retenção nuclear de RNAs com edições adenosina à inosina (A→I) e de RNAs que são induzidos em células diferenciadas (Bond and Fox 2009). Recentemente, foi observado que o ncRNA NEAT1 é um componente essencial para a formação e manutenção de paraspeckles (Clemson et al. 2009; Sasaki et al. 2009; Sunwoo et al. 2009). Chen e colaboradores observaram também que NEAT1 não é expresso em células tronco embrionárias humanas e que paraspeckles estão ausentes nestas células, só surgindo após a diferenciação, concomitante com a expressão de NEAT1 (Chen and Carmichael 2009). De forma semelhante, Sunwoo e colaboradores mostraram que a expressão de NEAT1/paraspeckles é induzida com a diferenciação de mioblastos em miotúbulos, ocorrendo um aumento de três vezes na expressão de NEAT1 e um aumento no tamanho e no número de paraspeckles (Sunwoo et al. 2009).

Um estudo de média-escala utilizando hibridização *in situ*, revelou que a maioria dos 88 ncRNAs expressos em células de Purkinje, no cerebelo de camundongo, apresentou evidência de localização sub-celular, contribuindo para a idéia que muitos destes RNAs não codificadores de proteínas podem ocupar locais específicos na célula (Mercer et al. 2008). Dos ncRNAs expressos, 29% mostrou um padrão de localização nuclear difuso, 61% apresentou um padrão de *foci* e 10% parece ser localizado ao longo do corpo celular (Mercer et al. 2008).

RNAs não-codificadores de proteínas parecem exibir uma variedade de localizações, o que, provavelmente, pode ser uma característica geral dos ncRNAs, nas quais podem desempenhar uma variedade de pápeis em diferentes domínios sub-celulares (Clark and Mattick 2011). Embora a maioria dos trabalhos, até o momento, tenha focado em descrever a localização de ncRNAs individualmente, um estudo em linhagens celulares humanas usando *tiling arrays* genômicos relatou que ~30% dos ncRNAs longos são encontrados exclusivamente no núcleo, ~15% são encontradas exclusivamente no citoplasma e ~50% são

encontrados em ambos os compartimentos (Kapranov et al. 2007). Esta distribuição subcelular implica que os ncRNAs podem controlar a regulação gênica à nível transcricional e pós-transcricional e o entendimento deste fenômeno pode ajudar na expansão do conhecimento a respeito da biologia celular e da função destes RNAs não codificadores de proteínas (Clark and Mattick 2011).

1.6 ncRNAs longos e regulação da expressão gênica em eucariotos

1.6. 1 Regulação transcricional de mRNAs

Ao nível transcricional, ncRNAs longos podem agir como estrutura de base, possibilitando a ligação, recrutamento, ou coordenação transcricional de complexos remodeladores da cromatina ativadores ou repressores à um *locus* genômico específico (Wang and Chang 2011). As mudanças na expressão gênica promovidas por ncRNAs podem ocorrer na vizinhança de seu *locus* (*cis*) ou em genes distantes de seu *locus* (*trans*).

Um dos mais bem estudados mecanismos de regulação por ncRNAs em *cis* é referente ao centro de inativação do cromossomo X, Xic (X *inactivation center*), um *locus* gênico que especifica vários ncRNAs, entre estes o Xist (Chow and Heard 2009). Xic controla o silenciamento de um dos cromossomos X em embriões de fêmeas de mamíferos, para promover a compensação de dose entre os sexos. Uma das primeiras mudanças que ocorrem durante a etapa de silenciamento de um dos cromossomos X é o recrutamento em *cis* do complexo de remodelamento da cromatina polycomb 2 (PRC2) pelo ncRNA RepA, que é originado da extremidade 5' de Xist (Zhao et al. 2008). RepA inicialmente recruta PRC2 para o futuro cromossomo X a ser inativado (Xi), contudo o ncRNA Tsix, que é antissenso ao Xist e tem papel bem estabelecido como antagonista de Xist, pode inibir esta interação competindo com RepA pela ligação ao PRC2 (Chow and Heard 2009; Senner and Brockdorff 2009).

Xi, de maneira que RepA pode recrutar o PRC2 e ativar a transcrição de Xist (Chow and Heard 2009; Senner and Brockdorff 2009). A indução de Xist é acompanhada pelo recrutamento de PRC2 que promove modificações de cromatina, trimetilação na lisina 27 da histona H3 (H3K27), no cromossomo Xi levando ao silenciamento de sua transcrição (Zhao et al. 2008). Mais recentemente, foi mostrado que uma proteína de matriz, hnRNP U (heterogeneous nuclear ribonucleoprotein U), pode ser necessária para o acúmulo de Xist no Xi (Hasegawa et al. 2010). Xist interage com esta proteína, de forma que a depleção de hnRNP U faz com que Xist deixe de se acumular ao longo do cromossomo Xi e passe a se localizar difusamente pelo nucleoplasma (Hasegawa et al. 2010).

Mecanismos de repressão da atividade transcricional similares têm sido reportados para outros ncRNAs longos. O ncRNA Air silencia a transcrição em *cis* de uma região de 400kb que inclui os genes Slc22a3, Slc22a2 e Igf2r no cromossomo paternal. Análogo à Xist, Air é retido no núcleo (Seidl et al. 2006) e parece cobrir o *locus imprinted* no cromossomo paterno (Nagano et al. 2008). Todavia, ao invés de se localizar uniformemente ao longo do domínio silenciado, Air acumula-se preferencialmente no promotor de Slc22a3, recrutando o complexo modificador de cromatina G9a metiltransferase que metila a lisina 9 da histona H3, causando silenciamento do promotor Slc22a3 no alelo paterno (Nagano et al. 2008).

De forma semelhante, o ncRNA Kcnq1ot1 está associado ao silenciamento bidirecional de cerca de 10 genes silenciados por *imprinting* no cromossomo paterno no *locus* Kcnq1 (Mancini-Dinardo et al. 2006). O ncRNA Kcnq1ot1 acumula-se não uniformemente ao longo do *locus* Kcnq1 e exerce sua função via interação com proteínas dos complexos G9a e PRC2 (Pandey et al. 2008). De forma interessante, certos genes no *locus* Kcnq1 *imprinted* são silenciados apenas na placenta, provavelmente, porque o ncRNA Kcnq1ot1 interage com G9a e PRC2 de maneira linhagem específica, resultando no estabelecimento de um estado

repressivo de modificações de histonas destes genes somente em certos tipos celulares (Pandey et al. 2008).

Um mecanismo parecido também ocorre em plantas, sugerindo que a relação ncRNA-PRC2 é um mecanismo de repressão gênica evolutivamente conservado (Wang and Chang 2011). Por exemplo, o ncRNA COLDAIR (*Cold Assisted Intronic Noncoding* RNA), em *Arabidopsis*, é necessário para o estabelecimento e manutenção de um estado repressivo da cromatina durante o período de frio (Heo and Sung 2011). COLDAIR é transcrito na direção senso de um intron de seu gene alvo, FLC, um forte repressor floral. Este ncRNA desempenha um papel crítico em direcionar o complexo PRC2 para a cromatina do gene FLC durante a vernalização, promovendo sua repressão gênica através da trimetilação de H3K27 (Heo and Sung 2011). Similarmente, em leveduras, vários ncRNAs antissenso, em numerosos *loci* gênicos, promovem o silenciamento da transcrição senso por afetar os estados de metilação e acetilação da cromatina (Camblong et al. 2007; van Dijk et al. 2011).

Em contraste ao grupo de ncRNAs longos que regulam a expressão gênica em cis, ainda se conhecem poucos exemplos bem caracterizados de ncRNAs longos que agem em trans. Um trabalho recente investigando lincRNAs regulados pelo fator de transcrição supressor tumoral p53, revelou um ncRNA, lincRNA-p21, que promove silenciamento de múltiplos genes localizados ao longo do genoma (Huarte et al. 2010). O promotor de lincRNA-p21 é ativado diretamente por p53 em resposta a dano de DNA e a sua expressão ectópica promove mudanças na expressão gênica e induz apoptose. Uma pesquisa dos fatores lincRNA-p21 identificou hnRNP que interagem com (heterogenous nuclear ribonucleoprotein K), um componente de um complexo repressor que age na via de p53. A interação de hnRNP K com o domínio 5' de lincRNA-p21 foi necessária, mas não suficiente para indução de apoptose, o que sugere que outras regiões do RNA são requeridas para recrutar outros fatores ou complexos para a cromatina (Huarte et al. 2010).

Centenas de ncRNAs longos são sequencialmente expressos ao longo dos eixos de desenvolvimento temporal e espacial dos *loci* humanos homeobox (Hox), onde definem domínios de cromatina com diferentes padrões de metilações de histona e acessibilidade à RNA polimerase (Rinn et al. 2007). Um destes ncRNAs, o HOTAIR (Hox *Transcript Antisense* RNA), origina-se do *locus* HOXC e silencia a transcrição ao longo de 40kb do *locus* HOXD, em *trans*, por induzir um estado de cromatina repressivo, que é proposto ocorrer através de recrutamento de PRC2 (Rinn et al. 2007). Adicionalmente, a expressão de HOTAIR tem sido recentemente associada à metástase (Gupta et al. 2010). Foi observado elevada expressão de HOTAIR em metástase de câncer de mama e sua depleção reduz a invasividade de células que expressam altos níveis de proteínas do complexo PRC2 (Gupta et al. 2010).

LincRNAs, como HOTAIR, são capazes de alterar e regular estados epigenéticos através do recrutamento de complexos modificadores da cromatina em *trans*. Em suporte a esta idéia, múltiplos ncRNAs expressos em vários tipos celulares mostraram-se associados a PRC2, de forma que a depleção mediada por siRNA de uma parte destes ncRNAs promoveu o enriquecimento de genes normalmente reprimidos por PRC2 (Khalil et al. 2009; Zhao et al. 2010).

O mecanismo de associação de ncRNAs longos aos complexos modificadores de cromatina pode resolver o paradoxo de como estes complexos que, frequentemente, possuem domínios de ligação a RNA, mas pouca especificidade de sequência para ligação em DNA, são capazes de se ligar a cromatina e promover a regulação da expressão gênica através de modificações na sua estrutura (Mercer et al. 2009).

Adicionalmente, têm sido sugerido que o processo de transcrição de ncRNAs por si só, ao invés do produto de transcrição, no caso o próprio ncRNA, pode ser funcional por facilitar uma estrutura de cromatina aberta em promotores de genes codificadores de proteínas, aumentando a acessibilidade aos fatores de transcrição e a RNA polimerase (Dinger et al. 2009; Atkinson et al. 2011). Um exemplo é a transcrição de ncRNAs longos a montante do *locus* fbp1⁺ de *Schizosaccharomyces pombe*, que induz remodelamento de cromatina que é crítico para ativação da transcrição de genes codificadores de proteínas a jusante (Hirota et al. 2008). De maneira interessante, foi observado que a transcrição de ncRNAs inicia-se em múltiplos sítios a montante do promotor fbp1⁺, causando uma região de cromatina aberta que procede progressivamente até o sítio de início da transcrição do mRNA. A inserção de um terminador transcricional na região a montante do mRNA abole a cascata de transcrição dos ncRNAs e promove alteração progressiva da cromatina, resultando em redução do recrutamento de fatores de transcrição ao promotor de fbp1⁺ e mínima indução da transcrição do mRNA (Hirota et al. 2008). Um mecanismo similar de remodelamento da cromatina promovido pela transcrição de ncRNAs foi mostrado no *locus* ade6-M26 de *S. Pombe* (Hirota and Ohta 2009).

Em alguns casos, ambos, a transcrição de um *locus* e o transcrito resultante podem ser funcionais, como foi observado no *locus* humano DHFR, onde a transcrição de um promotor a montante do mRNA e o ncRNA resultante contribuem para a repressão do gene codificador de proteínas, DHFR (Martianov et al. 2007). O ncRNA DHFR inibe a expressão do mRNA em *cis* e em *trans*, através da formação de uma estrutura tripla hélice RNA-DNA com o promotor DHFR, que interage diretamente com o fator de transcrição TFIIB, resultando no rompimento do complexo de pré-iniciação no promotor do mRNA (Martianov et al. 2007).

Dependendo do locus gênico, a transcrição do ncRNA pode ter profundos efeitos, influenciando negativamente ou positivamente na expressão de genes codificadores na vizinhança. Em alguns casos, a ação da transcrição é suficiente para ter consequências funcionais, mas é provável que muitos dos ncRNAs produzidos ainda possam desempenhar papéis adicionais, ainda não conhecidos (Wilusz et al. 2009).

Transcritos nascentes podem também representar alvos que recrutam fatores efetores, como é o caso, por exemplo, da transcrição de ncRNAs longos associados com a região promotora do gene *cyclin* D1 (Wang et al. 2008). Estes ncRNAs agem cooperativamente recrutando e modulando a atividade da proteína ligadora de RNA TLS (*Translocated in Liposarcoma*), em resposta a radiação ionizante. A ligação dos ncRNAs induz mudança conformacional de TLS, possibilitando a inibição da atividade de histona acetil transferase dos complexos CBP (CREB-*Binding Protein*) e p300 *histone acetyltransferase*, através de sua porção amino-terminal, silenciando a expressão de *cyclin* D1 (Wang et al. 2008). A habilidade de alguns ncRNAs, de se ligar e recrutar proteínas ligadoras de RNA para as regiões promotoras, representa mais uma expansão do repertório de regulação disponível para o programa transcricional (Wang et al. 2008).

NcRNAs longos podem também agir como co-fatores que modulam a atividade de fatores de transcrição. Em camundongo, o ncRNA Evf2 é transcrito de um *enhancer* ultraconservado e recruta o fator de transcrição DLX2 para o mesmo *enhancer* para induzir a expressão do gene adjacente, Dlx6 (Feng et al. 2006). Recentemente, estes ncRNAs transcritos a partir de regiões *enhancers* foram definidos como uma nova classe de ncRNAs, *enhancer* RNAs ou eRNAs (Kim et al. 2010). Em conformidade com o exemplo citado acima, o nível de expressão destes eRNAs foi correlacionado positivamente com o nível de síntese de mRNA nos genes na sua vizinhança, sugerindo que a síntese de eRNA ocorre especificamente em *enhancers* que estão ativamente engajados na síntese de mRNA (Kim et al. 2010; Wang et al. 2011).

1.6.2 Regulação pós-transcricional

A habilidade de ncRNAs longos em reconhecer sequências complementares propicia interações altamente específicas que são essenciais na regulação de vários passos do

processamento de mRNAs, incluindo *splicing* alternativo, edição, transporte, tradução e degradação (Mercer et al. 2009). A maioria dos genes de mamíferos expressam transcritos antissenso, que podem constituir uma classe de ncRNAs importante na regulação da dinâmica do mRNA (He et al. 2008).

NcRNAs antissenso podem mascarar elementos chaves na sequência do mRNA através da formação de RNA *duplexes*, como é o caso do Zeb2/Sip1 ncRNA, que é complementar a um sítio de *splicing* 5' de um intron do mRNA Zeb2/Sip1 (Beltran et al. 2008). O gene Zeb2/Sip1 é um repressor transcricional de *E-cadherin*, cuja expressão é altamente regulada durante a transição epitelial-mesenquimal (EMT). A expressão do ncRNA é induzida após EMT e previne o *splicing* de um intron que contém um *Internal Ribosome Entry Site* (IRES), requerido para a eficiente tradução e expressão da proteína Zeb2/Sip1 (Beltran et al. 2008). Em adição a este exemplo, existem muitos outros transcritos antissenso não codificadores de proteínas envolvidos na modulação de padrões de *splicing* alternativo do gene ao qual se se sobrepõem (Krystal et al. 1990; Munroe and Lazar 1991; Yan et al. 2005).

Certos ncRNAs podem parear com miRNAs, inibindo a sua habilidade de interagir com seus mRNAs alvos. Este mecanismo é usado pelo ncRNA IPS1 (*Induced by Phosphate Starvation* 1) em *Arabidopsis thaliana* (Franco-Zorrilla et al. 2007). O ncRNA IPS1 se liga e sequestra o miRNA miR-399 através de complementariedade de sequência quase perfeita, resultando na expressão aumentada dos genes alvos de miR-399 (Franco-Zorrilla et al. 2007).

O ncRNA rncs-1, em *C. elegans*, interfere no processamento de transcritos de dsRNA pela *Dicer* (Hellwig and Bass 2008). Apesar de conter uma dupla hélice quase perfeita de ~300 pares de bases, este transcrito não é substrato da *Dicer* por apresentar estruturas ramificadas flanqueando a porção central da dupla hélice que inibem seu processamento. Ao invés disso, o ncRNA rncs-1 atua em *trans*, inibindo a atividade da *Dicer*. Após superexpressão ou deleção da atividade de rncs-1 *in vivo*, os níveis de expressão de certos siRNAs

foram diminuídos ou aumentados, respectivamente, com as correspondentes mudanças nos seus níveis de mRNAs alvos (Hellwig and Bass 2008). Com isso, têm sido proposto que o ncRNA rcns1 se liga à *Dicer* ou assessora a ligação de proteínas ligadoras de dsRNA para competir com outros dsRNAs envolvidos em silenciamento gênico (Hellwig and Bass 2008).

Alternativamente, o anelamento do ncRNA pode promover a ligação de complexos protéicos efetores da degradação do mRNA, de uma maneira análoga ao exercido pelo complexo RISC quando associado a siRNAs e miRNAs (Mercer et al. 2009). RNA *duplexes* resultantes do anelamento de transcritos complementares, ou mesmo de ncRNAs longos com *hairpins* internos que podem ser processados em RNAs curtos, aumentam a possibilidade de que muitos ncRNAs longos possam participar em vias de silenciamento de RNA (Ogawa et al. 2008).

Um papel muito interessante é desempenhado pelo ncRNA longo NRON (*Noncoding Repressor of* NFAT), que participa do controle do tráfego nuclear do fator de transcrição NFAT (*Nuclear Factor of Activated T cells*). A depleção do ncRNA aumenta significativamente os níveis nucleares de NFAT, sugerindo que NRON funciona como um regulador específico do tráfego de NFAT para o núcleo (Willingham et al. 2005; Sharma et al. 2011). Provavelmente, ainda existem muitas outras funções exercidas pelos ncRNAs na biologia da célula que ainda não foram descobertas. A observação que muitos ncRNAs se localizam no citoplasma reforçam ainda mais esta idéia (Kapranov et al. 2007).

Dois estudos recentes empregaram sequenciamento de última geração para determinar se ncRNAs podem ser encontrados em associação com os ribossomos (Ingolia et al. 2011; Wilson and Masel 2011). Notavelmente, estes estudos revelaram que muitos RNAs previamente anotados como ncRNAs, em leveduras (Wilson and Masel 2011) e e em células tronco embrionárias de camundongos (Ingolia et al. 2011) estão no citoplasma, engajados na maquinaria de tradução. Este trabalho chama a atenção para o fato de que uma fração de

transcritos anotados como não-codificadores podem, na realidade, codificar para pequenos peptídeos.

Adicionalmente, o sequenciamento em larga escala do transcriptoma mitocondrial humano tem revelado a presença de ncRNAs nesta organela, onde podem estar envolvidos em funções específicas (Mercer et al. 2011; Mercer et al. 2011).

1.7 Vias de degradação de mRNAs em eucariotos

Mudanças rápidas e precisas na expressão gênica são essenciais para todos os aspectos da fisiologia celular, e células eucarióticas efetivamente regulam sua expressão gênica através de rápido ajuste do transcriptoma. A abundância de um transcrito é determinada pelo balanço entre a produção do mRNA e sua degradação, mas a velocidade com que as células ajustam seus níveis de mRNA é criticamente dependente da taxa de degradação (Munchel et al. 2011).

O processo de degradação do mRNA é um evento chave na determinação da estabilidade e dos níveis no estado estacionário de todas as espécies de mRNA. Em adição, as células continuamente requerem a eliminação de moléculas malformadas e a reciclagem dos produtos resultantes das reações de processamento (Schmid and Jensen 2008). Em células de eucariotos, a degradação dos produtos derivados da RNAP II é, na maioria das vezes, realizada por exonucleases. Pelo fato destas enzimas, geralmente, agirem em substratos de RNA com extremidade fita simples acessível, RNAs estáveis são equipados com características protetoras em suas extremidades (Schmid and Jensen 2008). No caso de mRNAs, a estrutura cap 5' e a cauda poli-A conferem proteção ao mRNA.

De acordo com estas considerações, a regulação da degradação do mRNA é promovida, principalmente, ao nível de remoção do cap 5' e da cauda poli-A (Schmid and Jensen 2008). Enquanto exonucleases 5'→3' degradam RNAs com a extremidade 5' desprotegida de uma maneira não específica, o decaimento da extremidade 3' dos RNAs é

realizado, na maioria dos casos, por um complexo de exonucleases altamente controlado, o exossomo (Schmid and Jensen 2008).

No citoplasma, o exossomo participa da degradação de mRNAs funcionais, auxiliando no controle da abundância do mRNA e nas taxas de síntese de proteínas, juntamente com as exonucleases 5'→3'. O exossomo citoplasmático também detecta e degrada RNAs com defeitos estruturais, através das vias de *nonsense-mediated decay* (NMD) e *non-stop decay* (NSD). A via de *nonsense-mediated decay* degrada mRNAs com códon de terminação prematuro, enquanto a via de *nonstop decay* degrada mRNAs que perderam o códon de terminação. Adicionalmente, também existe a via de *no-go decay* (NGD), onde o exossomo degrada RNAs cuja tradução é paralisada nos ribossomos.

A degradação de mRNAs no citoplasma é iniciada pela remoção da cauda poli-A por deadenialases específicas. Em alguns casos, poli-A polimerases citoplasmáticas adicionam uma cauda-poliA transiente na extremidade 3' de RNAs truncados que marca estes RNAs para degradação (Slomovic et al. 2010). Uma vez que a cauda poli-A é suficientemenete curta, o mRNA entra na via de degradação. Alguns mRNAs deadenilados são degradados pelo exossomo citoplasmático 3'→5', enquanto outros tem a estrutura cap 5' removida pelo complexo de *decapping* (Dcp1 e Dcp2) e são rapidamente degradados pela exonuclease 5'→3' Xrn1 (*exoribonuclease* 1) (Schmid and Jensen 2008). Ambas as vias podem agir redundantemente em paralelo, contudo a contribuição de cada uma destas vias exonucleotícas permanece em debate (Schmid and Jensen 2008).

Em adição, mRNAs de histonas, que não são poliadenilados, são rapidamente degradados acompanhando o processo de oligouridilação, que consiste na adição de uridinas na extremidade 3' do RNA, levando ao simultâneo *decapping* por Dcp1, acompanhado por degradação 5'→3' por Xrn1 e degradação 3'→5' pelo exossomo (Mullen and Marzluff 2008).

Ainda permanece a ser determinado se a oligouridilação está envolvida na degradação de outros RNAs.

Várias proteínas envolvidas nas vias de degradação 5'→3' e 3'→5', miRNAs e siRNAs têm sido localizadas em um domínio citoplasmático discreto denominado *P-bodies* (Balagopal and Parker 2009). Além disso, após inibição da degradação de RNA, vários mRNAs são detectados nos *P-bodies*, sugerindo a participação desta estrutura na degradação de RNAs (Balagopal and Parker 2009).

No núcleo, exonucleases 5'→3' (Rat1 em levedura, XRN2, em humanos, e Xrn1 em ambos) e a atividade 3'→5' do exossomo degradam RNAs que são produtos de maturação, no caso de mRNAs, os introns que são excisados pelo mecanismo de *splicing* (Houseley et al. 2006). O exossomo nuclear detecta especificamente RNAs precursores defeituosos e promove sua degradação, e, além disso, degrada mRNAs normais como parte do controle dos seus níveis de expressão (Kuai et al. 2005; Houseley et al. 2006).

Alguns elementos regulatórios em *cis*, presentes na sequência do mRNA, são importantes na determinação da meia-vida basal de um mRNA específico, se o mRNA será estabilizado ou desetabilizado em resposta a diferentes estímulos e, muitas vezes, determinam como será sua degração (Hollams et al. 2002). Estes elementos regulatórios são reconhecidos por fatores *trans*, que controlam a degradação dos RNAs, incluindo o exossomo, exonucleases 5'→3', siRNA, miRNA e proteínas de ligação a RNA (Hollams et al. 2002; Krol et al. 2010).

Um elemento bastante conhecido por promover instabilidade ao RNA é o elemento rico em AU (AU *Rich Element*- ARE). Elementos ARE estão presentes em muitos mRNAs que codificam proteínas nas quais sua expressão transiente é importante, como os fatores de crescimento e proto-oncogenes (Hollams et al. 2002; Houseley et al. 2006). O exossomo, juntamente com exonucleses 5'→3', promovem rápida degradação de mRNAs contendo o

elemento ARE, via interação direta com ARE ou através de seu recrutamento por proteínas de ligação a ARE (Moore 2002; Stoecklin et al. 2006; Schmid and Jensen 2008). Proteínas como a HuR, membro da família de proteínas ligadoras de RNA ELAV, podem se ligar a ARE, protegendo os mRNAs de serem degradados, adicionando assim mais um nível de regulação na estabilidade dos mRNAs. (Moore 2002; Lebedeva et al. 2011).

A junção de exons (*exon junction*) constitui outro elemento regulatório, que permite a distinção, pela maquinaria de NMD, entre códons de terminação normais e prematuros através da ligação de proteínas do complexo EJC (Le Hir et al. 2001). Outros elementos *cis* são conhecidos, como por exemplo, elementos responsivos a ferro, elementos ricos em citosina, *stem-loops* em histonas e introns retidos, que parecem conferir maior estabilidade ao mRNA (Hollams et al. 2002; Zhao and Hamilton 2007). A contribuição destes elementos para a degradação de ncRNAs ainda não é conhecida.

1.8 Estabilidade de mRNAs e de ncRNAs em eucariotos

Os mRNAs exibem uma variedade de meia-vidas, que muitas vezes estão correlacionadas com o papel funcional da proteína por eles codificada (Hollams et al. 2002). Proteínas que exercem funções que são constantemente necessárias na célula são caracterizadas por mRNAs com meia-vida mais longa, enquanto proteínas que são necessárias somente durante um certa fase do ciclo celular, etapa do desenvolvimento, crescimento ou diferenciação, ou em resposta a um estímulo externo, muitas vezes, têm meia-vida mais curta (Hollams et al. 2002).

A meia-vida dos mRNAs pode aumentar ou diminuir em resposta a diferentes estímulos, como fatores ambientais, fatores mitógenos e de crescimento, hormônios e mensageiros secundários que são liberados por cascatas de sinalização (Hollams et al. 2002). Um exemplo é o mRNA VEGF, que em condições de hipóxia tem sua meia-vida aumentada

de ~43 min para ~106 min (Levy et al. 1996). O fator de crescimento EGF prolonga a meiavida do mRNA do seu receptor EGFR (Jinno et al. 1988), e de forma semelhante, o metabólito dihidrotestosterona regula a estabilidade do mRNA do receptor de andrógeno (Yeap et al. 1999). A exposição a UV-B promove estabilização de muitos mRNAs com meiavida curta em células de mamíferos (Gowrishankar et al. 2005). Contrariamente, em leveduras, choque térmico moderado promove rápida degradação de mRNAs que codificam para proteínas ribossomais (Herruer et al. 1988), enquanto ausência de glicose, aminoácidos ou estresse osmótico induzido por açúcar causam estabilização de um grupo de mRNAs (Jona et al. 2000; Benard 2004; Greatrix and van Vuuren 2006).

Mudanças na estabilidade do mRNA também podem ser importantes durante diferenciação celular ou desenvolvimento embrionário. Durante a diferenciação muscular, a meia-vida de alguns mRNAs de genes músculo-específicos é prolongada, declinando quando a diferenciação está completa (Rudnicki et al. 1992). A expressão de uma variedade de transcritos é controlada por mudanças na estabilidade do mRNA durante o desenvolvimento neuronal (Ratti et al. 2006). Alterações na estabilidade de RNAs podem também estar associadas à várias doenças, incluindo câncer, nefropatias como diabetes, atrofia muscular e desordens neurológicas como doença de Alzheimer (Hollams et al. 2002; Champelovier et al. 2006; Benjamin and Moroni 2007; Heier et al. 2007).

Têm sido mostrado que muitos ncRNAs compartilham de vias de degradação utilizadas pelos mRNAs codificadores de proteínas e que alguns têm sua meia-vida alterada em resposta a estímulos externos ou fisiológicos. A depleção de dois componentes do exossomo (Rrp46 e Rrp44) em fibroblastos humanos, através de interferência de RNA acompanhada de *tilling array*, revelou uma classe de ncRNAs instáveis, que mapeiam a montante de promotores de genes codificadores de proteínas conhecidos. A depleção de outros fatores envolvidos em degradação como, Xrn1, Xrn2 e DCP2, não causou nenhum

efeito na estabilidade destes ncRNAs, sugerindo que estes ncRNAs são degradados especificamente pelo exossomo (Preker et al. 2008). De maneira semelhante, a análise do transcriptoma de leveduras mutantes para o componente Rrp6 do exossomo também revelou uma classe de transcritos instáveis que são rapidamente degradados pelo exossomo nuclear logo após sua síntese (Davis and Ares 2006; Neil et al. 2009; Xu et al. 2009).

Existem evidências que condições fisiológicas podem afetar a natureza destes transcritos instáveis, tornando-os estáveis. Por exemplo, a perda da proteína Rrp6 leva a estabilização de um transcrito antissenso ao gene PHO84, causando a repressão da transcrição do mRNA (Camblong et al. 2007). Intrigantemente, o mesmo fenótipo é observado durante o envelhecimento cronológico, quando a proteína Rrp6 mostra menor associação com o *locus* PHO84 (Camblong et al. 2007).

Em leveduras, uma classe de ncRNA meióticos é ativamente degradada durante o ciclo celular mitótico pelo exossomo nuclear, tornando-se estabilizada quando a célula entra em diferenciação meiótica (Lardenois et al. 2011). O sequenciamento fita-específico de última geração de leveduras mutantes para a exonuclease Xrn1, identificou 1.658 transcritos instáveis sensíveis a degradação pela maquinaria 5'→3' (van Dijk et al. 2011). Mais de 50% dos transcritos identificados têm orientação antissenso a genes codificadores de proteínas e acumulam-se em meio contendo lítio, indicando um possível papel para esta nova classe de ncRNAs na resposta adaptativa à mudanças de condições de crescimento (van Dijk et al. 2011).

Estes estudos reforçam a idéia que a estabilidade é um componente essencial na regulação da expressão gênica (Jacob and Monod 1961) e indicam que a modulação da maquinaria de degradação da célula parece estar intimamente relacionada com a regulação da expressão gênica envolvendo classes de ncRNAs instáveis, que sob determinadas condições tornam-se estabilizadas, podendo exercer sua função na célula.

2. OBJETIVOS

2.1 Objetivo geral

Investigar a biogênese, estabilidade e localização sub-celular de RNAs não-codificadores de proteínas intrônicos expressos em células humanas utilizando abordagens em larga-escala para avançar o conhecimento sobre a transcrição, processamento e endereçamento celular desta classe de transcritos.

2.2 Objetivos específicos

- 1. Estudar a biossíntese de ncRNAs intrônicos em células HeLa através de inibição específica da RNAP II com α-amanitina, seguido da extração do RNA das células tratadas, marcação e hibridização com oligoarranjos customizados contendo sondas que interrogam a expressão de mRNAs codificadores de proteínas e ncRNAs intrônicos gerados em uma fração de *loci* gênicos (oligoarranjos intron-exon).
- 2. Verificar a presença da modificação co-transcricional cap 5' em ncRNAs intrônicos, através de tratamento de RNA total com enzimas que removem o cap 5' e degradam RNAs com 5'-monofosfato, seguido de marcação e hibridização com oligoarranjos intronexon.
- 3. Estimar a estabilidade de ncRNAs intrônicos pelo tratamento de células HeLa com o inibidor transcricional actinomicina-D durante diferentes tempos (0, 1, 3, 6 e 8h), seguido de extração, marcação do RNA e hibridização em oligoarranjos intron-exon.
- 4. Avaliar a distribuição nuclear/citoplasmática de ncRNAs intrônicos e mRNAs através de fracionamento sub-celular de células HeLa, seguido de extração dos RNAs enriquecidos nas frações nuclear e citoplasmática, marcação e hibridização com oligoarranjos intronexon.

3. MATERIAL E MÉTODOS

3.1 Linhagem celular humana e condições de cultivo

Para a realização deste trabalho foi utilizada a linhagem celular humana de adenocarcinoma cervical, HeLa (CCL-2), que é amplamente utilizada na comunidade científica para diversos estudos. A linhagem foi adquirida da ATCC (*American Type Culture Collection*) e cultivada à 37 °C em atmosfera de 5% de CO₂ em meio de cultura apropriado. O meio utilizado, DMEM (*Dulbecco's Modified Eagle Medium*) (Cultilab), foi suplementados com SFB (Soro Fetal Bovino) (Cultilab), 10%, bicarbonato de sódio 1,2 g/L, glicose 4,2g/L e com os antibióticos penicilina (100U/mL) e estreptomicina (100μg/mL). Para a manutenção das células em cultura, ao atingirem aproximadamente 80% da densidade de saturação, as células foram lavadas com solução salina PBS (*Phosphate Buffered Saline*) (NaCl 140mM, KCl 2,7mM, Na₂HPO₄ 8mM e KH₂PO₄ 1,5mM, pH 7,2) e subcultivadas com o uso de solução de tripsina 0,1% (Cultilab). Os estoques celulares foram mantidos no meio de cultivo com 10% de DMSO (dimetilsulfóxido) (Sigma) e armazenado em freezer a -80 °C.

Para os experimentos de inibição de RNA Polimerase II, as células foram semeadas (~2 x 10⁶ células HeLa em placas de 56 cm²) e cultivadas por 24 horas, quando o meio de cultura foi trocado por meio fresco com α-amanitina (Sigma) 50µg/mL ou com o veículo (H₂O MilliQ estéril) e tratadas por 9 horas (Lee et al. 2004; Nakaya et al. 2007; Raha et al. 2010).

Para os experimentos de meia vida, as células foram semeadas (~2 x 10⁶ células HeLa em placas 56 cm²) e cultivadas por 24 horas, quando o meio de cultura foi trocado por meio fresco com actinomicina-D (Invitrogen) 10 μg/mL ou com o veículo (DMSO) e coletadas em diferentes tempos, 0,1, 3, 6 e 8 horas (Ebralidze et al. 2008).

3.2 Fracionamento celular

O fracionamento celular foi realizado conforme o protocolo descrito em Topisrovic et al. (Topisirovic et al. 2003). As células (~3 x 10⁷) foram coletadas e sedimentadas (conforme será descrito no item 3.3), em seguida foram lavadas em PBS pH 7,4 gelado, ressuspendidas gentilmente em tampão de lise (Tris pH 8,4 10mM, NaCl 140mM, MgCl2 1,5 mM, 0.5% NP-40, 1mM DTT (ditiotreitol) and RNasin (100 U/ml) (Promega)) e incubadas em gelo durante 5 minutos. A eficiência da lise, presença de células intactas *versus* núcleos livres, foi verificada através da análise de uma alíquota de 1uL do lisado por microscopia de luz, utilizando o microscópio Olympus CKX41. Em seguida, as suspensões celulares foram centrifugadas a 1.000g durante 5 min a 4 °C e o sobrenadante foi salvo como fração citoplasmática. Os sedimentos (frações nucleares) foram ressuspendidos em tampão de lise acrescido de 3,3% de deoxicolato de sódio e 6,6% de tween 40 e submetidos à nova centrifugação a 1.000g para remoção completa de resquícios da fração citoplasmática. O sobrenadante foi descartado e o sedimento foi coletado como fração nuclear. As frações nucleares e citoplasmáticas foram ressuspendidas 1mL e 4mL respectivamente de TRizol (Invitrogen) e armazenados a -80 °C.

3.3 Extração de RNA, purificação e tratamento com DNase

Células HeLa foram cultivadas em frascos apropriados até a confluência desejada, lavadas com PBS e descoladas da superfície do frasco por tratamento com 0,1% de tripsina seguido da adição de meio de cultura contendo SFB, ou apenas removidas do frasco raspando as células com *cell scraper* ("*rodo*" de borracha) em PBS (no caso dos tratamentos com α-amanitina e actinomicina-D, item 3.1). Em seguida, as células foram sedimentadas por centrifugação a 325 xg por 5 minutos e o sobrenadante foi removido. Os sedimentos de células foram utilizados imediatamente (no caso do fracionamento celular, item 3.2) ou

mantidos a -80 °C até extração do RNA. O isolamento do RNA foi realizado utilizando TRizol (Invitrogen), de acordo com o protocolo do fabricante, e eluído em um volume de 100 μL de água tratada com DEPC (Invitrogen). Em seguida, o RNA é submetido a uma etapa de purificação e tratamento com DNaseI utilizando o kit RNAspin (GE Healthcare), de acordo com o protocolo do fabricante, com alteração do tempo de incubação da DNaseI de 15 min para 1 hora. O RNA purificado é armazenado a -80 °C.

3.4 Avaliação da qualidade do RNA

As amostras foram quantificadas por sua densidade óptica a 260nm, e sua pureza atestada pela razão 260/280nm. As medidas foram realizadas no espectrofotômetro NanoDrop (Thermo Scientific).

O RNA foi analisado quanto à sua qualidade utilizando-se o equipamento Bioanalyzer 2100 (Agilent Technologies), que realiza uma eletroforese capilar de alta tensão. A integridade das amostras foi avaliada com o programa 2100 Expert (Agilent). Este programa atribui um valor de integridade de RNA (RNA Integrity Number, RIN), que permite uma estimativa da integridade através de todo traçado eletroforético da amostra, e não apenas da razão entre os RNA ribossomais. Apenas amostras que tinham uma boa integridade (RIN > 8) foram utilizadas em todos os experimentos.

Além disso, somente RNAs que não apresentavam contaminação com DNA genômico foram utilizados para os ensaios. Para certificar a ausência de DNA genômico nas amostras de RNA, foi realizada uma PCR (50μL, em 40 ciclos) do gene α-tubulina utilizando os *primers* α-Tubulina (tabela 1) e 1μg de RNA total de cada amostra (sem transcrição reversa) como molde para a reação de PCR (conforme será descrito no item 3.8.1).

3.5 Tratamento enzimático para análise de presença da modificação cap 5'

A enzima Tobacco Acid Pyrophospahatase (TAP; Epicentre Biotechnologies) cliva a ligação pirofosfato do nucleotídeo com a estrutura cap, guanosina metilada 5'-terminal, deixando um monofosfato na extremidade 5'. A enzima Terminator 5'-Phosphate-Dependent Exonuclease (5'-Exo; Epicentre Biotechnologies) é uma exonuclease processiva $5' \rightarrow 3'$ que digere RNAs 5'- monofosfato. Inicialmente, 5µg de RNA total de células HeLa foram tratados com 15U da enzima Tobacco, na presença de 40U do inibidor de RNAse RNAseOUT (Invitrogen) durante 2h a 37 °C. Em paralelo, foram feitos dois tratamentos sem enzima, um deles foi utilizado como controle negativo e o outro foi submetido à próxima etapa de clivagem enzimática pela 5'-Exo (TAP-/5'-Exo+). Após isso, as 3 amostras foram purificadas usando RNeasy Micro Kit (Qiagen), de acordo com o protocolo do fabricante. O RNA recuperado após o tratamento com TAP foi submetido ao tratamento com a enzima 5'-Exo (1U para 1ug de RNA), na presença de 40U de RNAseOUT (Invitrogen) (TAP⁺/5'-Exo⁺), 2h a 37 °C. A amostra TAP⁻/5'-Exo⁺ foi sujeita ao mesmo tratamento enzimático com 5'-Exo, enquanto o controle negativo foi novamente sujeito às mesmas condições de reação, exceto pelo fato de não receber adição de enzima (TAP-/5'-Exo-). Em seguida, as amostras foram purificadas usando RNeasy Micro Kit (Qiagen) de acordo com o protocolo do fabricante.

3.6 Desenho de oligonucleotídeos iniciadores (*Primers*)

Os *primers* utilizados neste trabalho (Tabela 1) foram desenhados utilizando-se o programa Primer Express versão 3.0 (Applied Biosystems), de acordo com os parâmetros padrões.

Tabela 1: Sequências dos primers utilizados. F (Forward) e R (Reverse).

Nome do primer	Sequência de nucleotídeos		
α-Tubulina_F	TCAACACCTTCTTCAGTGAAACG		
α-Tubulina_R	AGTGCCAGTGCGAACTTCATC		
c-Myc_F	TCAAGAGGTGCCACGTCTCC		
c-Myc_R	TCTTGGCAGCAGGATAGTCCTT		
pré-tRNA ^{Tyr} _F	AAAAAACCGCACTTGTCTCCTTCG		
pré-tRNA ^{Tyr} _R	CCTTCGATAGCTCAGCTGGTAGAG		
7SK_F	GACATCTGTCACCCCATTGA		
7SK_R	GCGCAGCTACTCGTATACCC		
18S rRNA_F	GCAGGCGCGGTAACC		
18S rRNA_R	AAGCTTATGACCCGCACTTACTG		
45S rRNA_F	GTACCGGCCGTGCGTACTTA		
45S rRNA_R	CTCGCCGCGCTCTACCTA		
5S rRNA_F	AGGCGCCTCCTTCAGCGTCT		
5S rRNA_R	CAGGCGGTCTCCCATCCAAG		
snRNA U15A_F	GAAGAGATGATGACGAGTCTGACTTG		
snRNA U15A_R	GAAATTACTTCAACCAGGGCTCTTT		

3.7 Transcrição reversa e síntese de cDNA

Para a transcrição reversa e síntese de cDNA fita simples foi utilizado 1μg de RNA total de cada condição experimental, de acordo com o que foi necessário. Cada amostra foi transcrita com o kit SuperScript III First Strand Synthesis Mix (Invitrogen), utilizando *primers* oligo-dT(20) ou *random hexamer*, segundo as recomendações do fabricante, com incubação à 50 °C. Uma pré-incubação de 3 min a 85 °C foi introduzida no protocolo, antes da fase de anelamento, para desnaturar completamente o RNA. Após o término da fase de anelamento, a reação foi mantida a 50 °C, para a adição da enzima em tampão pré-aquecido à 50 °C, eliminando-se a etapa de resfriamento no gelo.

3.8 Reação da Polimerase em Cadeia (*Polymerase Chain Reaction - PCR*)

3.8.1 PCR

Para amplificar os transcritos de interesse, foi realizada reação em cadeia da polimerase (PCR) utilizando-se 0,2 µM de *primers* específicos para cada transcrito e 1U de Taq Polimerase (GoTaq Green Master Mix; Promega), segundo às instruções do fabricante.

Utilizou-se 250ng de DNA genômico de células HeLa como controle positivo da reação e H₂O DEPC (Invitrogen) como controle negativo. O produto da reação foi submetido à eletroforese em gel de agarose 1,7% contendo brometo de etídeo 0,5 μg/mL em tampão TAE (Tris-Acetato 40mM, EDTA 1mM). O produto da reação foi visualizado em um fotodocumentador MiniBis Pro (DNR Bio-Imaging Systems) sob luz UV e as imagens foram adquiridas através do programa GelCapture (DNR Bio-Imaging Systems).

3.8.2 PCR quantitativa em tempo real

As reações da PCR quantitativa em tempo real (qPCR) foram feitas em triplicata em um volume de 20μL contendo 10μL de Sybr Green Master Mix (Applied Biosystems), 5μL de cDNA da transcrição reversa (diluído 1:6) e 3,2μM de *primers Foward* e *Reverse* específicos para cada gene. A reação foi realizada no aparelho 7500 Real–Time PCR System (Applied Biosystems) utilizando os parâmetros padrão do aparelho.

O ciclo da PCR no qual a intensidade de fluorescência do *Sybr Green*, ligado à dupla fita de DNA amplificado, é detectada determina o Ct do gene (*Cycle threshold* – ciclo referência). Este Ct é uma medida relativa da quantidade inicial de moléculas do transcrito alvo presente no cDNA. Quanto menor o Ct, mais expresso é o transcrito. Para tornar as medidas de expressão comparáveis entre as amostras usa-se como referência, o Ct de um gene constitutivo ou que não tenha sua expressão alterada nas diferentes condições de estudo. O 5S rRNA foi utilizado como normalizador no experimento de α-amanitina e o snRNA U15A foi utilizado como normalizador no experimento de tratamento enzimático com as enzimas TAP e 5'-Exo (Tabela 1). Calcula-se o delta-Ct (ΔCt) pela diferença entre a média das triplicatas (réplicas técnicas) do Ct do gene em análise e o Ct médio do normalizador. A partir dos valores de ΔCt calcula-se o valor de delta-delta-Ct (ΔΔCt) que consiste na subtração do ΔCt obtido nas diferentes condições testadas (tratamento e controle). A diferença de expressão

relativa do gene em estudo, *fold change*, é calculada como 2^{-ΔΔCt}. Para a análise de meia-vida, o *fold change* foi calculado a partir do delta-Ct obtido pela diferença entre a média do Ct detectado para o gene em análise em células tratadas com actinomicina D e o Ct médio do mesmo gene detectado em células controle não-tratadas.

A significância estatística da diferença de expressão dos transcritos testados por qPCR entre condições distintas (tratamento e controle) foi medida utilizando o teste t de *Student* (pareado, bicaudal, homocedástico), considerando-se como limiar de significância p-valor <0,05. Neste teste, foram considerados os valores médios de ΔCt de cada condição, referentes a pelo menos 3 réplicas biológicas.

3.9 Northern blot

A confirmação do enriquecimento das frações nucleares e citoplasmáticas foi realizada através de northern blot usando-se como sondas oligonucleotídeos biotinilados para o transcrito com localização predominantemente nuclear. U6 snRNA (5'-GAATTTGCGTGTCATCCTTGCGCAGGGGCCATGCTAA-3') e para o transcrito com tRNA^{Lys} localização predominantemente citoplasmática, (5'-7CTCATGCTCTACCGACTGAGCTAGCCGGGC-3') (Topisirovic et al. 2003); note que 7 indica biotina.

Um total de 10ug de RNA total de cada fração celular foi aplicado em gel de agaroseformaldeído 1%, ácido bórico 20mM, pH 8,3; EDTA 0,25mM; e 6,5% de formaldeído
(Merck). Antes de aplicadas no gel, as amostras foram desnaturadas a 60 °C por 10 minutos
em uma solução contendo formamida, formaldeído e tampão de corrida (20% glicerol; ácido
bórico 20mM, pH 8,3; EDTA 0,25mM; 50% formamida; 16% formaldeído; e azul de
bromofenol). Após 5 minutos no gelo, as amostras foram aplicadas no gel e submetidas à
eletroforese a 70 volts em tampão de corrida durante cerca de 4 horas.

O gel foi lavado com 4x SSC (NaCl 0,6M e citrato de sódio 60mM, pH 6,45) e 10x SSC (NaCl 1,5M, citrato de sódio 0,15M, pH 6,45) durante 20 minutos para eliminar o excesso de formaldeído. As amostras foram transferidas por capilaridade a uma membrana de *nylon* carregada positivamente (Ambion) com 10x SSC durante 16 horas, e os RNAs foram fixados na membrana através da aplicação de radiação ultravioleta utilizando o *UV Crosslinker* (Biorad). As membranas foram pré-hibridizadas em ULTRAhyb buffer (Ambion) por 1h a 45 °C e depois hibridizadas com 30pM de cada sonda separadamente durante 16h a 45 °C. Os sinais foram detectados usando CDP Star Chemiluminescence (Ambion) de acordo com as instruções do fabricante.

3.10 Análise da expressão gênica através de microarranjos de oligonucleotídeos (ou oligoarranjos)

3.10.1 Design do oligoarranjo

O oligoarranjo contém 44 mil sondas, incluindo sondas para as regiões de interesse e sondas controles, representadas por fragmentos de DNA fita-simples de 60 nucleotídeos. As sondas foram desenhadas de acordo com as especificações da Agilent, o que inclui conteúdo C-G entre 35 e 55%, TM entre 68 a 76 °C, ausência de bases repetidas consecutivamente mais do que 7 vezes ou 8 ou mais bases derivadas de regiões repetitivas, e um viés para a região 3' do transcrito (Hughes et al. 2001). O oligoarranjo foi desenhado em nosso laboratório e sintetizado pela Agilent (Nakaya et al. 2007). A anotação das sondas e suas sequências encontram-se disponíveis no GEO, número de acesso GPL4051.

3.10.2 Síntese, amplificação e marcação do RNA

Para os experimentos de fracionamento celular foram utilizados 500ng de RNA total para obtenção de cRNA amplificado e marcado com Cyanine 3-CTP (Cy3) ou Cyanine 5-CTP

(Cy5), utilizando o kit Quick Amp Labeling Kit, two-color, de acordo com as instruções do fabricante. Já para os experimentos de inibição da RNAP II, meia vida e presença de cap 5' foram utilizados 200ng de RNA total para obtenção de cRNA amplificado e marcado com Cy3 e Cy5, utilizando o kit Low Input Quick Amp Labeling Kit, two-color (Agilent Technologies) de acordo com as instruções do fabricante. Em ambos os kits, a quantidade inicial de RNA total é misturada a 2ul de um RNA sintético controle (RNAs *spikes*), na diluição especificada pelo fabricante (Agilent Technologies). Os RNAs *spikes* são constituídos por 10 RNAs sintéticos poli-adenilados presentes em diferentes quantidades conhecidas e são derivados do transcriptoma de adenovírus E1A. Estes transcritos não possuem complementaridade a sequências humanas e são otimizados para hibridização com sondas presentes nos oligoarranjos Agilent, possibilitando o controle da qualidade da amplificação e marcação das amostras hibridizadas.

O cRNA obtido na reação de amplificação e marcação foi purificado com RNeasy Mini Kit (Qiagen) de acordo com o protocolo do fabricante, mas sem a etapa de digestão com DNaseI. A incorporação do fluoróforo é confirmada através da medida de sua absorbância utilizando o espectrofotômetro NanoDrop (NanoDrop Technologies). Como recomendado pelo fabricante, cRNA amplificado com uma incorporação mínima de 8pmol fluoróforo/micrograma cRNA foi usado para hibridização. A Figura 1 esquematiza as etapas do processo de amplificação e marcação do RNA utilizando os kits citados acima.

RNA total RNA total Amostra A Amostra B mRNA AAAA 3' mRNA MMLV-RT MMLV-RT Oligo dT-Promoter Primer Oligo dT-Promoter Primer cDNA cDNA 2ª. Fita 2ª. Fita cDNA cDNA T7 RNA Polimerase T7 RNA Polimerase Cy5-CTP, NTPs Cy3-CTP, NTPs cRNA cRNA 2ª. Fita 2ª. Fita cDNA cDNA cRNA Purificado cRNA Purificado Ç~0000 cRNA cRNA Oligo Microarrays

cRNA Amplificado

Figura 1: Método de amplificação de cRNA e marcação com os fluoróforos Cy3 e Cy5 (Agilent Technologies).

3.10.3 Hibridização

Para cada lâmina de oligoarranjo, 825 ng de amostras de cRNA de duas condições distintas, marcadas independentemente com Cy3 e Cy5 foram usados para a hibridização, de acordo com o protocolo *Agilent's ISH Kit Plus* (Agilent Technologies) por 17 horas a 65 °C.

Posteriormente, as lâminas foram lavadas com o *Agilent SSPE wash protocol v. 2.1* (Agilent Technologies).

3.10.4 Processamento de Imagem

Após a hibridização com os alvos fluorescentes, os oligoarranjos foram digitalizados com uma resolução de 5μm usando o *scanner High-Resolution Microarray scanner* (Agilent Technologies). As intensidades de fluorescência foram extraídas usando o *Feature Extraction programa* (FE). O programa reconhece o local de cada sonda (*spot*), obtém a intensidade daquela região registrada pelo *scanner* e aplica a subtração da intensidade do fundo local (em torno do *spot*). O FE também calcula um test t para identificar se o valor de intensidade obtido no *spot* está significativamente (p < 0,05) acima dos valores de *background* local.

3.10.5 Tratamento dos dados

Os critérios de filtragem e processamento dos dados referentes a cada um dos experimentos utilizando oligoarranjos (actinomicina D, α-amanitina, cap 5' e localização subcelular) encontram-se detalhados nos resultados (vide itens 4.2, 4.3, 4.4 e 4.5). Para as análises de expressão descritas no item 4.1 foi utilizado o processo de normalização por quantil (Bolstad et al. 2003) para tornar as intensidades entre as réplicas comparáveis. Durante a normalização, os valores de intensidade de cada réplica são ordenados de maneira independente, segundo sua posição no *ranking*. Em seguida é calculado o valor médio para cada posição do *ranking* e este substitui o valor inicial de intensidade em todas as réplicas. A normalização foi feita utilizando a platforma R (Team 2011).

3.10.6 Métodos estatísticos

Para a análise dos transcritos diferencialmente expressos após o tratamento de células com o inibidor transcricional α-amanitina foi utilizado um programa específico para análise

de dados de microarranjo, SAM (*Significance Analysis of Microarrays*) (Tusher et al. 2001). O SAM atribui um *score* para cada transcrito de acordo com a alteração da expressão deste transcrito em relação ao desvio-padrão das medidas réplicas. Para os casos de transcritos cujo *score* atribuído é maior do que um limiar determinado pelo usuário, o programa emprega permutações para estimar a taxa de falsos positivos (*False Discovery Rate* - FDR), ou seja, a fração destes transcritos que poderiam ter sido identificados ao acaso. Ao final da análise o programa agrega um valor de *q* para cada transcrito, que mede a sua significância estatística. Este valor de *q* (q-valor) representa a menor taxa de FDR na qual um transcrito é considerado significativo (Tusher et al., 2001). Uma medida de q-valor <0,05% foi utilizada como critério para considerar um transcrito diferencialmente detectado entre as condições testadas.

Para as demais análises (cap 5', localização sub-celular) não se utilizou o teste estatístico realizado pelo SAM devido ao pequeno número de réplicas, optando-se pela utilização de um critério de razão de aumento (*fold-change*) para definição dos transcritos diferencialmente detectados entre as condições analisadas (vide itens 4.4 e 4.5).

Na análise de actinomicina D foi utilizado um pacote em *Perl* que tem funções para cálculo de regressão linear e vários parâmetros relacionados (http://search.cpan.org/~randerson/Statistics-LineFit-0.07/lib/Statistics/LineFit.pm). O pacote fornece o valor de uma estatística t, que é utilizada para determinação da significância do coeficiente angular obtido para cada regressão linear ajustada.

3.10.7 Métodos de enriquecimento de categorias gênicas

Foi realizada análise de enriquecimento nas categorias função molecular, processo biológico e componente celular, definidas por termos de GO (*Gene Ontology*), utilizando-se o programa DAVID (Huang da et al. 2009; Huang da et al. 2009). Como referência, na análise para os genes codificadores de proteínas foram usados todos os genes representados no

microarranjo e para os transcritos intrônicos foram usados todos os genes para os quais existem transcritos TIN ou PIN presentes no microarranjo. O teste estatístico Hipergeométrico *Benjamini & Hochberg's FDR multiple testing correction* foi empregado, considerando-se um nível de significância de ≤ 0.05 .

3.10.8 Métodos de bioinformática

Para realizar as análises bioinformáticas correspondentes aos transcritos TIN e PIN expressos em células HeLa foi utilizado o conjunto de programas de código aberto BEDTools (Quinlan and Hall 2010). Estes programas utilizam arquivos em formato BED, contendo coordenadas genômicas, facilitando a busca de elementos de interesse, como transcritos, genes, regiões promotores, sítios de ligação de fatores de transcrição, entre outros, obtidos no *UCSC Genome Browser* (http://genome.ucsc.edu/). O uso do programa BEDTools permite o cruzamento de coordenadas genômicas entre os transcritos analisados (TIN ou PIN) com os elementos de interesse.

Os transcritos TIN e PIN considerados como expressos em células HeLa, nos experimentos de oligoarranjo referentes ao inibidor α-amanitina, foram utilizados para as análises bioinformáticas. O critério para considerar um transcrito como expresso foi que este apresentasse expressão acima do *background* nas 8 réplicas controle hibridizadas. As análises realizadas incluíram a identificação dos seguintes elementos de interesse provenientes de experimentos realizados com HeLaS3, variante da linhagem HeLa:

- H3K4me3 (trimetilação da lisina 4 da histona 3);
- RNAP II;
- -CAGE tags (CAP analysis of gene expression).

Cada uma das bibliotecas utilizadas acima foi processada para ser utilizadas na análise. Duas bibliotecas de HeLaS3 H3K4me3 foram obtidas utilizando os *tracks* que se

encontravam no *UCSC* Genome Browser (http://genome.ucsc.edu/) fazendo parte do projeto ENCODE. Somente coordenadas genômicas que haviam sido detectadas nas duas bibliotecas foram utilizadas para a análise, totalizando em 36.706 tags. Após a identificação destes tags, foram montados clusters para juntar as coordenadas genômicas que se sobrepuseram, resultando em 35.328 cluster tags. Somente os cluster tags que não apresentaram sobreposição com 5' UTRs (RefSeq e UCSC genes) foram utilizados resultando em 26.425 (75%, 26.425/35.328) clusters tags. Esta última filtragem foi realizada para minimizar a associação de H3K4me3 referentes aos transcritos codificadores de proteínas com os transcritos estudados, pois esta marca de cromatina está associada a regiões promotoras de genes e se estendem para o início da transcrição (Mikkelsen et al. 2007; Guttman et al. 2009).

A biblioteca referente às coordenadas de RNA polimerase II foi obtida no trabalho de Rozowsky e colaboradores (Rozowsky et al. 2009). As coordenadas referentes à RNAP II tiveram que ser convertidas de hg18 para hg19 resultando em 24.680 tags. A mesma metodologia de retirar as coordenadas que se sobrepunham aos 5'UTR também foi utilizada como anteriormente, resultando em 14.071 tags no final. As bibliotecas de CAGE tags foram obtidas do projeto realizado em RIKEN OSC (RIKEN Omics Science Center) e estão disponíveis no UCSC Genome Browser (http://genome.ucsc.edu/). As técnicas utilizadas para a produção das bibliotecas estão descritas na própria página do Genome Browser (Carninci and Hayashizaki 1999; Kodzius et al. 2006; Valen et al. 2009). Somente as bibliotecas referentes à linhagem HeLaS3 foram utilizadas. Na primeira etapa foram montados clusters tags para juntar as coordenadas genômicas que se sobrepunham, resultando em 1.893.993 clusters tags. Em seguida, foram removidos clusters tags que se sobrepunham aos 5' UTRs e a outros exons. Neste caso, foram retirados clusters que se sobrepunham a exons, pelo fato destes poderem representar possíveis inícios alternativos de transcrição do próprio codificador de proteína, produzindo uma nova isoforma como discutido no trabalho de Valen e

colaboradores (Valen et al. 2009). Assim, a biblioteca de *clusters tags* ficou composta por 1.489.677 coordenadas genômicas, representando 77% da biblioteca inicial.

Para realizar a análise de identificação dos elementos de interesse, a coordenada genômica mais próxima da ponta 5' do transcrito (TIN ou PIN) foi selecionada e a distância anotada. Somente foram consideradas as coordenadas genômicas que não se sobrepunham, ou que se sobrepunham parcialmente a região do transcrito analisado para identificar uma possível região para o início de transcrição associado ao TIN ou PIN.

Como controle, foi utilizado um conjunto de 100 ou 10 grupos compostos por sequências randômicas referentes aos transcritos TIN e PIN, respectivamente, que apresentavam tamanho e conteúdo de CG similar aos das sequências não codificadoras analisadas. O desenho dessas sequências foi baseado nas características dos transcritos. No caso dos transcritos TIN senso foram desenhadas sequências totalmente intrônicas que possuíam orientação senso, o mesmo ocorreu para os ncRNAs TIN antissenso. O desenho dos grupos randômicos referentes aos transcritos PIN foi realizado utilizando coordenadas genômicas de regiões exônicas com acréscimo de nucleotídeos em 5kb em ambas as pontas para mimetizar os transcritos parcialmente intrônicos. Como a base de dados para a busca de sequências randômicas PIN estava limitada pelas regiões exônicas mais 5kb a montante e a jusante, a quantidade de grupos randômicos desenhados para essas sequências parcialmente intrônicas (PIN) foi reduzida para 10.

Os grupos randômicos foram analisados exatamente da mesma forma que as sequências dos transcritos TIN e PIN. Para inferir um valor estatístico nas análises, foi utilizada a análise Kolmogorov-Smirnov, que mede a diferença de distribuição entre dois grupos atribuindo um valor estatístico. Para isso foi utilizado um *script* montado em R com o pacote Deducer para comparar os grupos de RNAs não codificadores e codificadores de proteínas (mRNAs) com os 100 grupos randômicos. Somente foram consideradas com

significância estatística as análises que apresentaram todos os valores de p-value < 0.05 comparado aos 100 ou 10 grupos randômicos.

4. RESULTADOS

4.1 RNAs não-codificadores intrônicos expressos em células HeLa

Células de adenocarcinoma de cérvix (HeLa), uma linhagem amplamente utilizada na comunidade científica, foram utilizadas neste trabalho para os estudos de estabilidade, biogênese, processamento por adição de cap na extremidade 5' e localização sub-celular de RNAs não-codificadores intrônicos (ncRNAs intrônicos). Todos estes estudos envolveram a hibridização de RNAs amplificados linearmente e marcados fluorescentemente com oligoarranjos intron-exon e serão detalhados adiante (vide próximos itens dos resultados). Os oligoarranjos utilizados contém 44 mil sondas e foram desenhados em nosso laboratório, tendo sido sintetizados, sob encomenda, pela Agilent (figura 2) (Nakaya et al. 2007). Cada lâmina contém, além de controles positivos e negativos, sondas para 7.135 transcritos totalmente intrônicos (TIN) e 4.440 transcritos parcialmente intrônicos (PIN), selecionados a partir de uma análise de mapeamento no genoma humano do conjunto de sequências expressas (mRNAs e ESTs) presentes nos bancos públicos (Nakaya et al. 2007). Estão também presentes no microarranjo 8.786 sondas representando regiões exônicas dos mRNAs para os quais se incluíram sondas para transcritos TIN ou PIN. O esquema da figura 2 ilustra os tipos de sondas presentes no oligoarranjo.

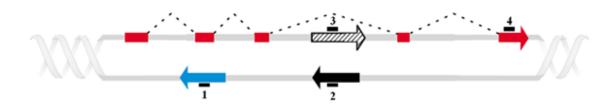


Figura 2: Desenho do oligoarranjo intron-exon 44k. Sonda 1: sonda para transcritos parcialmente intrônicos antissenso. Sondas 2 e 3: pares de sondas reverso-complementares para cada uma das duas possíveis fitas dos transcritos totalmente intrônicos desenhadas para detecção de transcritos TIN antissenso ou senso. Sonda 4: sondas disponíveis comercialmente pela empresa Agilent, representando um exon de um gene codificador de proteínas conhecido.

Inicialmente, foi feita uma análise para determinar os transcritos exônicos e ncRNAs intrônicos expressos em células HeLa (figura 3). Realizamos a extração de RNA de 4 réplicas biológicas de culturas sub-confluentes de células HeLa. Após tratamento com DNAse para eliminar possíveis contaminantes e verificação da integridade do RNA isolado, foram gerados alvos fluorescentes marcados com Cy3 ou Cy5 a partir de cada réplica, gerando 8 réplicas no total (4 réplicas biológicas x 2 réplicas técnicas de cada). Após hibridização com o oligoarranjo intron-exon e extração dos dados de intensidade (vide item 3.10) foram feitas análises para detectar os transcritos codificadores e não-codificadores de proteína expressos em HeLa. Para um transcrito ser considerado expresso, estipulamos como filtro a detecção da respectiva sonda acima do *background* em todas as 8 réplicas.

Foi observada a detecção positiva de 6.894 sondas para mRNAs codificadores de proteínas, o que corresponde a 78% do total de sondas presentes na lâmina (figura 3). Para os ncRNAs intrônicos observou-se a detecção de uma fração menor de sondas comparado aos mRNAs; para os transcritos TIN antissenso, PIN antissenso e TIN senso foram detectadas 1.514 (21% do total), 2.105 (48% do total) e 2.671 (37% do total de sondas presentes na lâmina), respectivamente, do total de transcritos presente no oligoarranjo (figura 3). As sondas que interrogam ncRNAs presentes no oligoarranjo foram desenhadas a partir de transcritos não-codificadores detectados em mais de 20 tipos de tecidos humanos (Nakaya et al. 2007). A detecção de uma fração menor de ncRNAs em HeLa quando comparado aos mRNAs, está de acordo com relatos da literatura que indicam que, em geral, RNAs não codificadores apresentam expressão tecido ou célula específica (Ravasi et al. 2006; Nakaya et al. 2007; Mercer et al. 2008).

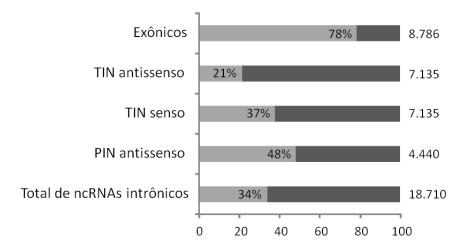


Figura 3: Expressão de transcritos exônicos e ncRNAs intrônicos em célula HeLa. Na extrema direita do gráfico encontram-se os totais de sondas presentes no oligoarranjo para cada categoria de transcrito, exônicos, TIN antissenso, TIN senso e PIN antissenso. Em cinza claro está mostrado a porcentagem de sondas que apresentaram expressão acima do *background* nas 8 réplicas hibridizadas e a fração em cinza escuro representa as sondas que não foram expressas acima do *background* nas 8 réplicas.

Com o intuito de comparar a abundância relativa de ncRNAs intrônicos e mRNAs codificadores de proteínas, foi realizada uma análise da distribuição das intensidades de sinal das sondas para transcritos exônicos, TIN antissenso, TIN senso e PIN antissenso expressos em células HeLa (figura 4, painel à esquerda). É possível observar que os transcritos codificadores de proteínas apresentam intensidades de expressão mais altas que os ncRNAs intrônicos (figura 4), sendo em média 13 vezes mais alta, com um valor mediano de intensidade de 190. A distribuição de intensidades para todas as classes de transcritos intrônicos é similar, e as medianas dos transcritos TIN antissenso, TIN senso e PIN antissenso são, respectivamente, 14, 14 e 15. Este resultado está de acordo com a observação de outros estudos que relataram que ncRNAs são, geralmente, menos abundantes que os mRNAs codificadores de proteínas (Kampa et al. 2004; Nakaya et al. 2007).

A análise do tamanho da sequencia dos *contigs* correspondentes aos ncRNAs intrônicos interrogados no oligoarranjo e detectados como expressos em células HeLa, revelou que estes transcritos são longos, com uma fração apresentando mais de 1.000

nucleotídeos (figura 4, painel à direita). A mediana do tamanho predito para os contigs que constituem os transcritos intrônicos TIN antissenso, TIN senso e PIN antissenso, foi respectivamente, 723, 757 e 774 nucleotídeos.

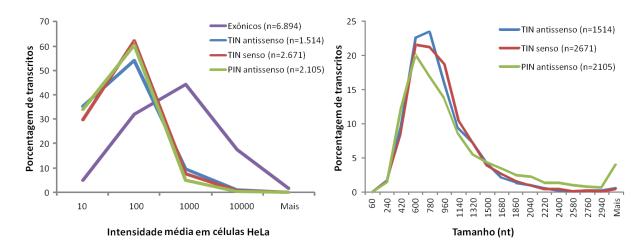


Figura 4: Distribuição das intensidades de sinal das sondas para transcritos exônicos e ncRNAs intrônicos obtida através da hibridização de RNA de células HeLa no oligoarranjo intron-exon 44k (painel à esquerda). Distribuição do tamanho (nt) predito para os ncRNAs intrônicos expressos em HeLa (painel à direita). Somente sondas com intensidade de sinal acima do background nas 8 réplicas hibridizadas (4 réplicas independentes e 4 réplicas técnicas) foram consideradas na análise.

Com o intuito de expandir o conhecimento sobre a biogênese dos ncRNAs intrônicos, fizemos análises utilizando dados públicos referentes a elementos associados a atividade transcricional. Inicialmente foi realizada uma análise para verificar o enriquecimento da enzima RNAP II na vizinhança dos transcritos intrônicos expressos em células HeLa (figura 5). Para isso foram utilizados dados de imunoprecipitação da cromatina com anticorpos para RNAP II, seguidos de sequenciamento em larga escala (ChIP-seq) (Rozowsky et al. 2009). Os dados são provenientes da linhagem HeLa S3, que é uma variante clonal originada de células HeLa. Após o cruzamento das coordenadas genômicas das extremidades 5' dos ncRNAs intrônicos expressos em Hela, verificamos enriquecimento significativo (teste Kolmogorov-Smirnov (KS), p <0,05) da enzima RNAP II na vizinhança (até 10 kb) do provável início de transcrição, em relação a um conjunto controle. Este conjunto é formado por sequências

genômicas, em igual número e com o mesmo tamanho e conteúdo de GC dos ncRNAs intrônicos, selecionadas aleatoriamente a partir de todas as regiões intrônicas do genoma (figura 5) (vide item 3.10.8). O resultado obtido sugeriu que ao menos uma fração dos ncRNAs intrônicos seja sintetizada pela RNAP II.

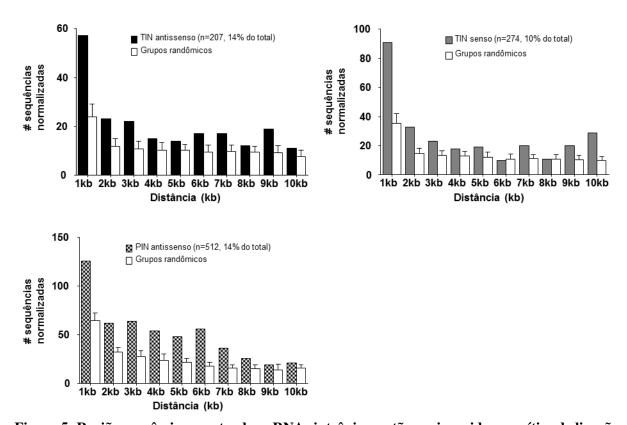


Figura 5: Regiões genômicas contendo ncRNAs intrônicos estão enriquecidas em sítios de ligação da RNAP II. A distribuição da distância (eixo x) relativa ao acúmulo da RNA Polimerase II na vizinhança dos ncRNAs intrônicos foi calculada. Para comparação, a distribuição das distâncias foi realizada com um número igual de sequências de grupos randômicos que continham sequências genômicas com características similares dos transcritos expressos em células HeLa (tamanho e conteúdo de CG). O número de transcritos ("n") utilizado para gerar as distribuições está indicado entre parênteses.

De forma semelhante à análise realizada para verificar enriquecimento de RNAP II na vizinhança dos ncRNAs intrônicos, cruzamos as coordenadas das extremidades 5' dos transcritos intrônicos expressos em Hela com as coordenadas de *tags* cap 5' expressas na linhagem HeLa S3, obtidas a partir de bibliotecas de CAGE *tags* (CAP *analysis of gene expression*) (Carninci and Hayashizaki 1999; Kodzius et al. 2006; Valen et al. 2009). Após

isso, comparamos a distribuição das distâncias das CAGE *tags* relativas aos prováveis sítios de início da transcrição dos ncRNAs intrônicos e de um grupo controle, como descrito anteriormente (figura 6) (vide item 3.10.8). A distribuição de distância das CAGE *tags* foi significativamente diferente da observada para o grupo controle (teste KS, p <0,05), mostrando enriquecimento de CAGE *tags* na vizinhança a montante dos ncRNAs intrônicos (figura 6). A alta porcentagem de ncRNAs intrônicos que possuem um CAGE *tag* a menos de 1kb da extremidade 5' (40 a 50%) indica que o início de transcrição real encontra-se próximo do início predito para muitos dos transcritos interrogados no oligoarranjo. Uma vez que as bibliotecas de CAGE são geradas a partir de frações enriquecidas em RNA poliadenilado e com cap 5', este resultado também corrobora a idéia de que os ncRNAs intrônicos podem constituir unidades transcricionais produzidas pela RNAP II.

A fim de acumular mais evidências quanto a possíveis elementos envolvidos na regulação da transcrição intrônica, também foi investigado o enriquecimento na marca trimetilação da lisina 4 da histona 3 (H3K4me3), uma modificação de cromatina associada à regiões promotoras ativas da RNAP II (Mikkelsen et al. 2007; Guttman et al. 2009), em relação ao início de transcrição conhecido dos ncRNAs intrônicos expressos em células HeLa (figura 7) (vide item 3.10.8). Para isso foram utilizados os dados de duas bibliotecas de HeLa S3, disponíveis no *UCSC Genome Browser* (http://genome.ucsc.edu/), construídas através de experimentos de imunoprecipitação da cromatina com anticorpos para H3K4me3, seguido de sequenciamento em larga escala (ChIP-seq). A distribuição de distância das marcas H3K4me3 para os ncRNAs TIN antissenso e TIN senso foi significativamente diferente (teste KS, p <0,05) da observada para o grupo controle (mesmo das análises anteriores), mostrando haver um enriquecimento de H3K4me3 na vizinhança a montante dos ncRNAs TIN (figura 7). Não foi observado enriquecimento de H3K4me3 na vizinhança de transcritos PIN antissenso.

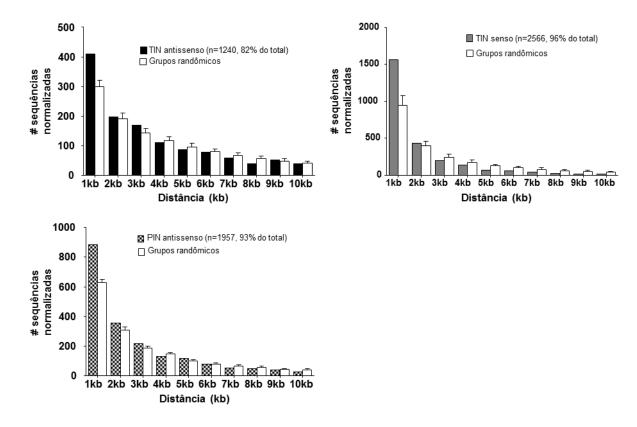


Figura 6: Regiões genômicas contendo ncRNAs intrônicos estão enriquecidas em sítios de iniciação da transcrição de RNAs com cap 5'. A distribuição da distância (eixo x) relativa à presença de CAGE tags (CAP analysis of gene expression) na vizinhança dos ncRNAs intrônicos foi calculada. Para comparação, a distribuição das distâncias foi realizada com um número igual de sequências de grupos randômicos que continham sequências genômicas com características similares dos transcritos expressos em células HeLa (tamanho e conteúdo de CG). O número de transcritos ("n") utilizado para gerar as distribuições está indicado entre parênteses.

Adicionalmente, pôde-se observar que nas três análises realizadas, RNAP II, CAGE *tags* e marcas H3K4me3, houve um enriquecimento significativo (teste qui-quadrado, p<0,05), na vizinhança até 1kb do sítio conhecido de início de transcrição dos ncRNAs intrônicos, exceto para o PIN antissenso na análise da marca H3K4me3. Este resultado sugere que o sítio de iniciação da transcrição dos ncRNAs intrônicos seja, de fato, próximo do predito através de análise de ESTs, e que a estrutura da região promotora destes ncRNAs seja semelhante a de genes codificadores de proteína transcritos pela RNAP II.

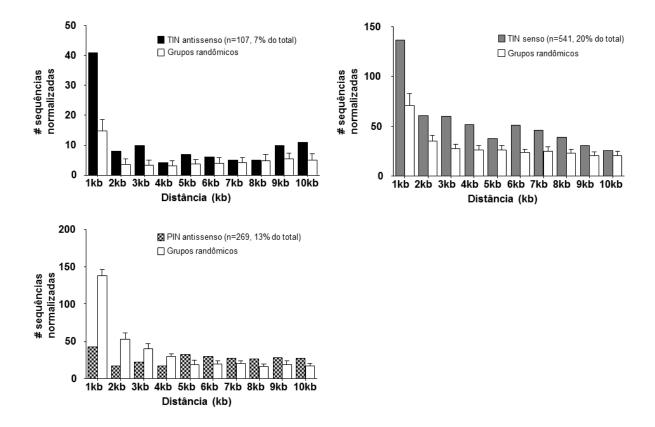


Figura 7: Regiões genômicas contendo ncRNAs intrônicos estão enriquecidas em marcas de cromatina associadas à regiões promotoras. A distribuição da distância (eixo x) associada às marcas de H3K4me3 relativa aos ncRNAs intrônicos foi calculada. Para comparação, a distribuição das distâncias foi realizada com um número igual de sequências de grupos randômicos que continham sequências genômicas com características similares dos transcritos expressos em células HeLa (tamanho e conteúdo de CG). O número de transcritos ("n") utilizado para gerar as distribuições está indicado entre parênteses.

Para investigar a relação entre a transcrição intrônica e a transcrição do mRNA no mesmo *locus*, fomos verificar se os transcritos intrônicos estão expressos de maneira concomitante a expressão do mRNA do mesmo *locus* gênico. A figura 8, painel à esquerda, mostra um diagrama de Venn com sobreposição da expressão dos transcritos intrônicos e exônicos correspondentes. Ao menos 57% dos *loci* gênicos interrogados no oligoarranjo e expressos em células HeLa (2.948/6.894) geram transcritos exônicos e ncRNAs intrônicos de pelo menos uma das classes (TIN antissenso, TIN senso ou PIN antissenso) (figura 8, painel à esquerda). Este resultado está de acordo com a observação prévia do nosso grupo de que a maior parte dos *loci* humanos (>74%) gera transcritos não-codificadores que mapeiam em

regiões intrônicas (Nakaya et al. 2007). Uma fração de 17%, 9% e 10% do total de transcritos intrônicos expressos, respectivamente, TIN antissenso, TIN senso e PIN antissenso, são gerados em *loci* gênicos em que não foi detectada a expressão do mRNA codificador correspondente, o que sugere que a expressão destes transcritos seja independente do mRNA codificador do mesmo *locus* (figura 8, painel à esquerda). Para 43% dos *loci* gênicos presentes no oligoarranjo detectou-se apenas a expressão de transcritos exônicos (figura 8, painel à esquerda).

Do total de ncRNAs intrônicos expressos em *loci* que também produzem mRNAs (figura 8, interseções no painel à esquerda), a maior parte apresentou correlação positiva com a abundância do mRNA, ≥0,5 (TIN antissenso 39%, TIN senso 41% e PIN antissenso 44%) e uma fração menor apresentou correlação negativa, ≤ -0,5 (TIN antissenso 23%, TIN senso 21% e PIN antissenso 19%) (figura 8, painel à direita). A observação de correlação predominantemente positiva entre a abundância de transcritos intrônicos e exônicos pode sugerir que estes ncRNAs tenham um papel em modular em *cis* a transcrição ou a estabilidade do mRNA codificador de proteína correspondente.

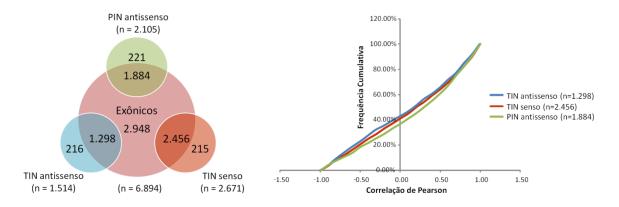


Figura 8: Expressão de mRNAs e ncRNAs intrônicos em Hela. Diagrama de Venn mostrando a sobreposição da expressão, em células HeLa, entre os *loci* gênicos de transcritos exônicos e de ncRNAs TIN antissenso, TIN senso e PIN antissenso (painel à esquerda) e correlação de pearson entre a expressão dos transcritos exônicos e cada uma das três classes de ncRNAs intrônicos expressos no mesmo *locus* gênico (painel à direita).

4.2 Estimativa em larga-escala da estabilidade de ncRNAs intrônicos em células HeLa

Para estudar a estabilidade dos RNAs não-codificadores intrônicos de forma ampla, medimos as mudanças na sua expressão após a inibição da transcrição com actinomicina D durante diferentes tempos de tratamento seguido da marcação de sondas fluorescentes e hibridização com o oligoarranjo intron-exon. A actinomicina D é um antibiótico contendo polipeptídeos cíclicos, que forma um complexo estável com o DNA fita simples no complexo de iniciação da transcrição, bloqueando a elongação da transcrição pelas RNA Polimerases (Sobell 1985).

Para este experimento, foi extraído RNA de duas réplicas biológicas independentes de células HeLa tratadas com actinomicina D ou com seu veículo (controle) em cinco tempos diferentes (0, 1, 3, 6 e 8h). Após esta etapa inicial, as amostras de RNA referentes a cada condição foram processadas para hibridização em oligoarranjos intron-exon para avaliação em larga-escala da meia-vida dos ncRNAs intrônicos. O desenho experimental utilizado foi baseado na comparação do cRNA das amostras tratadas nos diferentes tempos com uma mesma amostra de cRNA referência, proveniente de um *pool* (referência) preparado a partir de quantidades iguais de RNA de duas réplicas independentes de amostras controle (1, 3, 6 e 8h). Cada tempo de tratamento foi hibridizado em uma lâmina de oligoarranjo intron-exon 44k, de forma que foram usadas cinco lâminas. As amostras de cRNA de células tratadas com actinomicina D, coletadas nos diferentes tempos, foram marcadas com Cy3, e a amostra de cRNA referência foi marcada com Cy5 (figura 9). A estratégia utilizada elimina a necessidade de utilização de *dye swap*, visto que todas as amostras são hibridizadas contra uma mesma referência e, posteriormente, as razões normalizadas pela referência são comparadas entre si (Simon and Dobbin 2003).

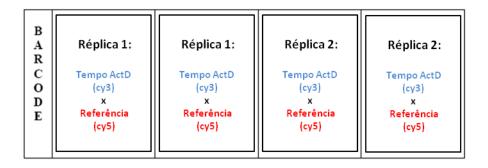


Figura 9: Esquema das hibridizações no oligoarranjo de 4x44k para análise de meia-vida. A lâmina contém 4 *arrays* independentes. cRNA marcado com Cy3, proveniente de células tratadas com actinomicina D, em diferentes tempos (0, 1, 3, 6 e 8h), foi hibridizado com cRNA marcado com Cy5, proveniente de células tratadas com o veículo (Referência). Utilizou-se 1 lâmina para cada tempo de tratamento. Os experimentos foram realizados com 2 réplicas independentes de cRNA de células tratadas com actinomicina D e um *pool* de amostras controles de duas réplicas independentes coletadas nos tempos 1, 3, 6 e 8h (Referência). Cada réplica biológica independente foi hibridizada duas vezes (réplica técnica). Tempo; refere-se a cada um dos 5 tempos hibridizados (0, 1, 3, 6 e 8h); ActD; actinomicina D.

A fim de se verificar a reprodutibilidade dos experimentos de hibridização no oligoarranjo, os valores das intensidades dos sinais de cada uma das réplicas foram comparados através da medida da correlação entre as réplicas dos experimentos (Quackenbush 2002). Foi possível observar uma alta correlação entre as réplicas técnicas da amostra referência (média da correlação entre as amostras 0,98 ± 0,01). Também foi obtida boa correlação entre as medidas de intensidade das réplicas técnicas e biológicas das amostras tratadas com actinomicina D por 0h, 1h, 3h, 6h e 8h, sendo a média das correlações, respectivamente, 0,89± 0,11; 0,99± 0,01; 0,97± 0,02; 0,98± 0,01; 0,97± 0,02. Assim, os valores de intensidade de todas as réplicas hibridizadas foram utilizados para as análises de expressão.

Inicialmente, foram consideradas na análise todas as sondas detectadas com intensidade acima do *background* nas 20 réplicas técnicas da amostra referência (tabela 2). Um filtro adicional foi usado, onde somente sondas com intensidade de sinal 2 vezes maior que o *background* nas 20 réplicas técnicas da amostra referência foram consideradas como

detectadas e mantidas na análise (tabela 2). Este critério foi estipulado para evitar estimações errôneas de meia-vida para transcritos expressos em níveis muito baixos, próximos ao background, onde pequenas mudanças de intensidade, após o tratamento com actinomicina D, poderiam ser interpretadas como evidência de alta estabilidade (Sharova et al. 2009). Foram calculadas razões de intensidades (actinomicina D/referência) em logaritmo na base 10 (log10) para cada sonda. Em seguida, a taxa de degradação ou decaimento ("d") dos transcritos foi estimada através da regressão linear dos valores de razão das intensidades em log10 referentes aos 5 tempos de tratamento com actinomicina D versus o tempo t, como mostra a equação: y = a + bt, onde "t" é o tempo, "b" é o coeficiente angular e "a" é o intercepto e "d"= -b*ln(10) (Sharova et al. 2009). Apenas as sondas cujos modelos de regressão linear ajustados apresentaram coeficiente angular "b" dentro do intervalo de confiança da regressão de 95%, através de teste-t (p <0,05) foram consideradas válidas para determinação da meia-vida (tabela 2). O teste t foi calculado para cada valor de coeficiente angular e testa a hipótese nula dos dados obtidos se adequarem a um modelo de regressão linear.

Tabela 2: Número de sondas de cada classe com expressão acima do *background* no experimento de hibridização com oligoarranjo intron-exon de RNA de células HeLa tratadas com actinomicina D (0, 1, 3, 6 e 8h); número de sondas detectadas (intensidade pelo menos 2 vezes maior que o *background*) e número de sondas válidas para o cálculo da meia-vida.

Classe do	#Sondas	#Sondas com	#Sondas detectadas	#Sondas válidas (%
transcrito	presentes no	expressão acima	(% do total das	das expressas acima
	oligoarranjo	0 \	expressas acima do	0 // \
		do total da classe	background)	das detectadas)

Exônicos	8.786	7.129 (81%)	5.551 (78%)	5.480 (77%; 99%)
TIN	7.135	1.992 (28%)	401 (20%)	331 (17%; 83%)
antissenso				
TIN senso	7.135	3.548 (50%)	713 (20%)	646 (18%; 91%)
PIN	4.440	2.480 (56%)	409 (17%)	371 (15%; 91%)
antissenso				

Como a mesma massa de RNA foi usada para as hibridizações em todos os tempos de tratamento com actinomicina D, a degradação de RNAs menos estáveis, após o bloqueio da transcrição, resulta no aumento da abundância relativa de RNAs mais estáveis. Para corrigir este efeito, normalizamos a taxa de decaimento "d" de todas as sondas válidas, pela média das taxas de decaimento dos transcritos considerados mais estáveis, ou seja, aqueles que tiveram sua abundância relativa aumentada ao longo do tratamento. Apenas um pequeno número de transcritos (23 transcritos, 0,3% do total) teve sua abundância relativa aumentada, o que foi explicitado pela taxa de decaimento com sinal negativo ou coeficiente angular com sinal positivo, enquanto a maior parte dos transcritos apresentou uma diminuição na abundância ao longo do tempo, como era esperado. Portanto, 23 transcritos foram utilizados para a normalização dos dados.

A meia-vida de todos os RNA foi estimada pela fórmula: $t_{1/2} = \ln(2)/d$, onde " $t_{1/2}$ " é a meia-vida e "d" é a taxa de decaimento, como exemplificado na figura 10. Foi estabelecido um limite máximo de 24h para o valor de meia-vida calculado, visto ser difícil estimar valores de meia-vida muito longos baseados em um experimento cujo último ponto de medida foi o tempo 8h.

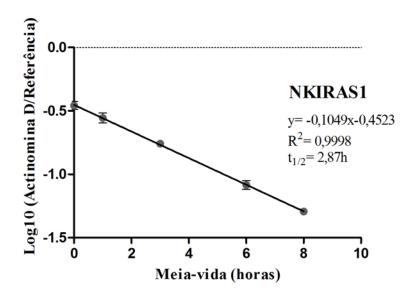


Figura 10: Determinação da meia-vida através da inibição da transcrição com actinomicina D. Células HeLa foram tratadas com actinomicina D durante 0, 1, 3, 6 e 8h e os níveis de expressão de

mRNA NKIRAS1 foram obtidos por hibridização no oligoarranjo intron-exon. Regressão linear (y=a + bx) foi usada para estimar a taxa de decaimento dos RNAs, d=-b*ln(10) e a meia-vida, $t_{1/2}=ln(2)/d$. Após a normalização da taxa de decaimento, pela média da taxa de decaimento dos transcritos mais estáveis, a meia-vida passou a ser $t_{1/2}=2.54h$. O coeficiente de determinação (R^2), a equação do modelo de regressão linear e a meia-vida ($t_{1/2}$) estimada antes da normalização pelos transcritos mais estáveis estão indicados. As razões de intensidades (actinomicina D/referência), em logaritmo na base 10, estão representadas como a média de duas réplicas independentes e duas réplicas técnicas \pm erro padrão da média.

4.2.1 Estimativa da meia-vida de ncRNAs intrônicos e mRNAs

Através das análises descritas anteriormente, foi possível estimar a meia-vida de 1.348 transcritos não-codificadores intrônicos, fração que corresponde a 17%, 15% e 18% do total de sondas para os ncRNAs TIN antissenso, TIN senso e PIN antissenso, respectivamente, expressas na lâmina considerando-se apenas o critério de detecção de intensidade acima do *background* nas réplicas da amostra referência ("sondas com expressão acima do *background*"; (vide item anterior; tabela 2). Como os transcritos intrônicos são expressos em baixos níveis na célula, e na análise realizada foi necessário assegurar que as sondas válidas tivessem intensidade pelo menos duas vezes acima do *background* ("sondas detectadas"; vide item anterior), uma grande fração dos ncRNAs intrônicos foi excluída da análise para garantir uma maior confiabilidade dos resultados. Já os transcritos exônicos, como são mais abundantes na célula, não foram muito afetados pelo limiar de intensidade estipulado, de modo que foi possível calcular a meia-vida para 77% do total de sondas exônicas expressas na lâmina com intensidade acima do *background* (vide item anterior; tabela 2).

A diminuição da quantidade de RNA medida pelo oligoarranjo ao longo dos tempos de tratamento se adequou bem ao modelo de regressão linear ajustado para o cálculo da meiavida, de modo que foram consideradas válidas (teste t, p<0,05; vide item anterior) 99% das sondas exônicas detectadas e 88% das sondas intrônicas detectadas (vide item anterior; tabela 2).

A figura 11 mostra a distribuição da frequência dos valores de meia-vida para as diferentes classes de transcritos. Em nossa análise, a mediana da meia-vida estimada para os transcritos exônicos foi de 3,1h (figura 11), tempo relativamente curto comparado com trabalhos da literatura (Yang et al. 2003; Sharova et al. 2009).

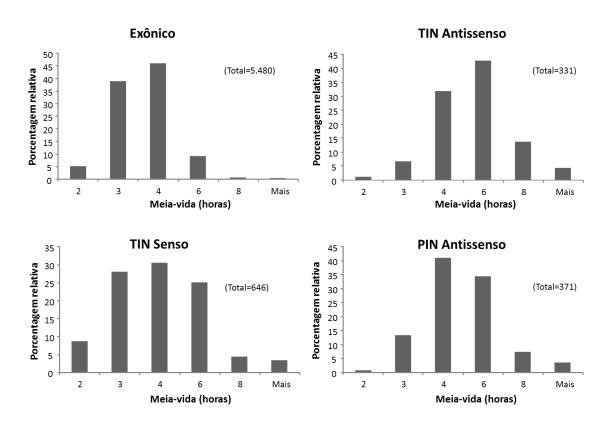


Figura 11: Distribuição de meia-vida dos transcritos exônicos e ncRNA intrônicos. O número de transcritos em cada classe está indicado entre parênteses.

Yang e colaboradores calcularam a meia-vida de 5.245 genes em células de hepatoma humano (HepG2) e a mediana estimada foi de 10h (Yang et al. 2003). Um trabalho mais recente avaliou a meia-vida de 19.977 genes em células tronco embrionárias de camundongo (MC1), estimando uma mediana de 7,1h (Sharova et al. 2009). Ao compararmos as taxas de decaimento de 1.601 transcritos exônicos que foram comuns entre as análises de células HeLa (nossa análise) e células HepG2 (dados de Yang e colaboradores) obtivemos uma correlação moderada mas estatisticamente significativa (r=0,53; p<0,001) (figura 12, painel à esquerda).

Usando a base de dados HOMOLOGENE (Wheeler et al. 2008), identificamos 3.440 transcritos exônicos ortólogos entre humano e camundongo, para os quais taxas de decaimento foram estimadas em células HeLa (nossa análise) e em células MC1 (dados de Sharova e colaboradores). Comparando-se as taxas de decaimento dos transcritos exônicos, identificados como ortólogos de camundongo, com os dados de Sharova, de forma semelhante ao resultado obtido com células HepG2, observamos novamente uma correlação moderada e estatisticamemente significativa (0,52; p<0,001) (figura 12, painel à direita). Apesar da mediana da meia-vida para RNAs codificadores de proteínas estimada em nossa análise ter sido bem menor que a dos trabalhos citados acima (Yang et al. 2003; Sharova et al. 2009), obtivemos uma correlação estatisticamente significante ao comparar as taxas de decaimento de nossa análise com os dados de hepatoma humano e de células tronco embrionárias de camundongo. Uma possível explicação, para a observação de valores de meia-vida mais curtos em nossa análise, pode ser atribuída ao menor tempo de duração do ciclo celular das células HeLa em relação ao das linhagens celulares utilizadas nos estudos mencionados; o tempo de geração das células HeLa é cerca de 14-16h enquanto o tempo de geração de HepG2 é de aproximadamente 50h (Whitfield et al. 2002; Yang et al. 2003).

A mediana dos valores de meia-vida estimados para os transcritos intrônicos foi de 4,2h para os TIN antissenso, 3,4h para os TIN senso e 3,9h para os PIN antissenso. Notavelmente, os ncRNAs intrônicos apresentaram em média uma estabilidade comparável ou superior a observada para transcritos exônicos (3,1h).

Fomos verificar se, ao compararmos a meia-vida de transcritos intrônicos e exônicos em uma mesma faixa de intensidade de expressão, ainda era possível observar a maior estabilidade dos ncRNAs intrônicos em relação aos exônicos (figura 13). Para isso, comparamos apenas as distribuições de meia-vida dos transcritos exônicos e intrônicos com intensidade menor ou igual 100. As frações de transcritos com intensidade até 100, do total de

transcritos de cada classe, exônicos, TIN antissenso, TIN senso e PIN antissenso, para os quais estimamos a meia-vida, foram, respectivamente, 32%, 72%, 78% e 81% (figura 13). Nota-se uma menor porcentagem de transcritos exônicos em relação a intrônicos com baixo sinal de intensidade (menor que 100), em conformidade o resultado mostrado no item 4.1. A figura 13 sugere que os ncRNAs intrônicos apresentam maior estabilidade que os transcritos exônicos, independente da faixa de intensidade analisada. Desta forma, elimina-se a possibilidade da maior estabilidade observada para os ncRNAs intrônicos em relação aos transcritos exônicos, por algum motivo, estar atrelada simplesmente a sua menor intensidade de expressão.

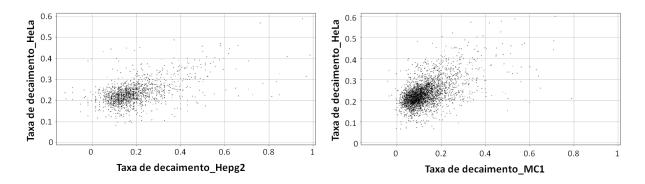


Figura 12: Comparação das taxas de decaimento obtidas para os transcritos exônicos em células HeLa com as observadas em trabalhos da literatura. No painel à esquerda, as taxas de decaimento de 1.601 transcritos exônicos comuns as análises de células HeLa e células HepG2 (Yang et al. 2003) foram comparadas. No painel à direita, as taxas de decaimento de 3.440 transcritos exônicos ortólogos, comuns as análises de células HeLa e células de camundongo, MC1(Sharova et al. 2009) foram comparadas.

Utilizamos o teste estatístico de Kolmogorov-Smirnov (KS) para determinar a significância estatistica da diferença entre as distribuições de meia-vida das diferentes classes de transcritos, apresentadas na figura 13. Este teste calcula a máxima diferença de distância entre duas distribuições (d-valor) e estima um p-valor para esta distância, que é indicativo da probabilidade de dois conjuntos de dados se distribuírem da mesma maneira.

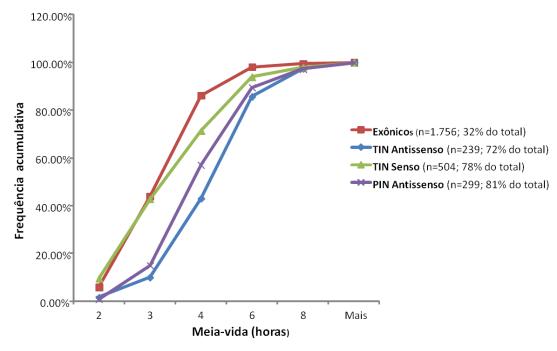


Figura 13: Comparação da meia-vida entre transcritos exônicos e transcritos intrônicos TIN antissenso, TIN senso e PIN antissenso, com intensidade de expressão até 100 (relativo ao tempo 0h de tratamento com actinomicina D). O número de transcritos em cada classe está indicado entre parênteses e o valor de porcentagem refere-se à fração de transcritos com sinal de intensidade até 100, do total de transcritos em cada classe para os quais se têm valores de meia-vida calculados.

O padrão de maior estabilidade em relação aos transcritos exônicos é mais evidente para os TIN antissenso (teste KS, d=0,43; p=0) e PIN antissenso (teste KS, d-valor=0,29; p=0) (figura 14). Visivelmente, ocorre uma sobreposição nas distribuições de meia-vida entre transcritos exônicos e TIN senso nos tempos mais curtos (< 3h), contudo, ainda pode-se dizer que a classe de trancritos TIN senso também apresenta uma maior proporção de transcritos com maior meia-vida em relação aos exônicos (teste KS, d=0,15; p=6,35E⁻⁸) (figura 13).

Dentre os transcritos intrônicos, os TIN senso apresentaram uma distribuição de meiavida bem distinta em relação aos TIN e PIN antissenso, representando a classe com menor estabilidade (figura 13). Através do teste KS, observamos que existe uma distância de 32% (d-valor=0,32; p=1,33E⁻¹⁵) e 28% (d-valor=0,28; p=4,98E⁻¹³), respectivamente, entre as distribuições de meia-vida dos TIN senso e TIN antissenso e TIN senso e PIN antissenso. Esta distância entre as distribuições cai para 14% ao compararmos os TIN e PIN antissenso

(d-valor=0,14; p=0,01). Portanto, os transcritos TIN senso parecem diferir dos demais transcritos intrônicos, possivelmente por incluir sub-populações de transcritos com natureza distinta, tais como pré-mRNAs e/ou trechos de introns resultantes do *splicing* de pré-mRNAs.

4.2.2 Análise de enriquecimento de categorias gênicas entre transcritos com meiavida curta (< 2,5h) ou longa (> 3,5h)

Estudos prévios em eucariotos observaram que proteínas envolvidas em processos altamente regulados decaem relativamente rápido, em contraste com aquelas envolvidas em funções metabólicas centrais (Ross 1995; Wang et al. 2002; Yang et al. 2003). Para verificar se nossas análises de meia-vida em células HeLa seguiram esta mesma relação geral documentada, fizemos uma análise de enriquecimento em categorias de termos de ontologia gênica, incluindo os domínios função molecular, processo biológico e componente celular. Avaliamos o enriquecimento em termos GO para 892 transcritos exônicos com meia-vida menor que 2,5h (menos estáveis) e para 1.389 com meia-vida maior que 3,5h (mais estáveis) (tabelas 3 e 4).

Os transcritos codificadores de proteínas menos estáveis apresentaram enriquecimento em categorias relacionadas à regulação da transcrição, ciclo celular, regulação de apoptose, dentre outros. Já os transcritos exônicos mais estáveis apresentaram enriquecimento em categorias relacionadas à elongação da tradução e constituinte estrutural do ribossomo. Portanto, a maioria destes resultados encontra-se em conformidade com a literatura (Narsai et al. 2007; Friedel et al. 2009; Sharova et al. 2009). Nenhuma categoria de GO foi encontrada significativamente enriquecida entre os *loci* gênicos que expressam ncRNAs intrônicos.

Tabela 3: Análise de enriquecimento em categorias de ontologia gênica (GO) de transcritos exônicos com meia-vida menor que 2,5h.

Categoria	Termo	# de genes	P-valor	P-valor
				corrigido*
GOTERM_BP_FAT	GO:0045449~regulation of transcription	234	<0,001	<0,001
GOTERM_MF_FAT	GO:0003677~DNA binding	208	<0,001	<0,001
GOTERM_BP_FAT	GO:0051252~regulation of RNA metabolic process	156	<0,001	<0,001
GOTERM_MF_FAT	GO:0030528~transcription regulator activity	142	<0,001	<0,001
GOTERM_CC_FAT	GO:0031981~nuclear lumen	145	<0,001	<0,001
GOTERM_CC_FAT	GO:0005730~nucleolus	61	<0,001	<0,001
GOTERM_MF_FAT	GO:0008270~zinc ion binding	200	<0,001	<0,001
GOTERM_BP_FAT	GO:0006357~regulation of transcription from RNA polymerase II promoter	80	<0,001	<0,03
GOTERM_CC_FAT	GO:0005694~chromosome	44	< 0,001	<0,001
GOTERM_BP_FAT	GO:0051172~negative regulation of nitrogen compound metabolic process	56	<0,001	<0,002
GOTERM_BP_FAT	GO:0051253~negative regulation of RNA metabolic process	40	<0,001	<0,002
GOTERM_BP_FAT	GO:0016568~chromatin modification	39	<0,001	<0,05
GOTERM_BP_FAT	GO:0006325~chromatin organization	42	<0,001	<0,002
GOTERM_MF_FAT	GO:0003682~chromatin binding	25	<0,001	<0,05
GOTERM_BP_FAT	GO:0019941~modification-dependent protein catabolic process	57	<0,001	<0,002
GOTERM_CC_FAT	GO:0015630~microtubule cytoskeleton	52	< 0,001	< 0,05
GOTERM_BP_FAT	GO:0016573~histone acetylation	11	<0,001	<0,003
GOTERM_BP_FAT	GO:0007049~cell cycle	79	<0,001	<0,03
GOTERM_CC_FAT	GO:0005819~spindle	21	<0,001	<0,002
GOTERM_BP_FAT	GO:0030521~androgen receptor signaling pathway	11	<0,001	<0,002
GOTERM_BP_FAT	GO:0006974~response to DNA damage stimulus	42	<0,001	<0,004
GOTERM_BP_FAT	GO:0006259~DNA metabolic process	49	<0,001	<0,04
GOTERM_BP_FAT	GO:0006281~DNA repair	34	<0,001	< 0,05
GOTERM_BP_FAT	GO:0042981~regulation of apoptosis	70	<0,001	< 0,05

^{*} P-valor corrigido para múltiplo teste (Benjamini-Hocheberg).

Tabela 4: Análise de enriquecimento em categorias de ontologia gênica (GO) de transcritos exônicos com meia-vida maior que 3,5h.

Categoria	Termo	# de genes	P-valor	P-valor
				corrigido*
GOTERM_MF_FAT	GO:0003735~structural constituent of ribosome	20	<0,001	<0,001
GOTERM_BP_FAT	GO:0006414~translational elongation	15	<0,001	<0,02
GOTERM_CC_FAT	GO:0005840~ribosome	23	<0,001	<0,004

^{*}P-valor corrigido para múltiplo teste (Benjamini-Hocheberg).

4.3 Avaliação da contribuição da RNA Polimerase II na biogênese de ncRNAs intrônicos

Com o objetivo de responder se os ncRNAs intrônicos são transcritos pela RNAP II, realizamos tratamento de células HeLa com α- amanitina. A α-amanitina é um peptídeo cíclico de oito aminoácidos que inibe a enzima RNAP II por bloquear sua translocação sobre o DNA durante a elongação da transcrição (Bushnell et al. 2002; Gong et al. 2004) Primeiramente, verificamos a eficiência do tratamento com α- amanitina utilizando como controle um transcrito sintetizado pela RNAP II, mRNA de c-Myc. Após esta validação inicial, réplicas de RNA total de células tratadas e não-tratadas com α-amanitina foram amplificadas, marcadas fluorescentemente e hibridizadas no oligoarranjo intron-exon. Esta abordagem possibilitou o conhecimento dos ncRNAs intrônicos presentes no oligoarranjo e expressos em células HeLa que tem como enzima transcricional a RNAP II.

Realizamos a extração de RNA de 4 réplicas biológicas de células HeLa tratadas e não-tratadas (controle) com α- amanitina (50µg/ml) durante 9h, seguindo essencialmente protocolos de trabalhos já publicados (Lee et al. 2004; Nakaya et al. 2007; Raha et al. 2010). A eficácia da inibição da RNAP II foi avaliada através da quantificação do mRNA de c-Myc, por PCR em tempo real, em amostras de RNA tratadas e não-tratadas com o inibidor (figura 14). Como controles negativos do experimento, utilizamos os transcritos pre-tRNA^{Tyr} e 7SK, que são sintetizados pela RNAP III, que não é sensível a baixas concentrações de α-amanitina, tais como às utilizadas nestes experimentos, e os RNAs 18S rRNA e 45S rRNA, transcritos pela RNAP I, que é muito resistente ao tratamento com α- amanitina (Jacob et al. 1970; Kedinger et al. 1970; Lindell et al. 1970). O transcrito 5S rRNA, sintetizado pela RNAP III, foi utilizado como normalizador do experimento. Através deste ensaio, foi possível verificar que o tratamento com α-amanitina foi eficiente, diminuindo em 46 vezes a

abundância relativa de c-Myc nas células tratadas com α- amanitina e não afetou de maneira estatisticamente significativa a abundância dos controles negativos (teste t, p<0,05; figura 14).

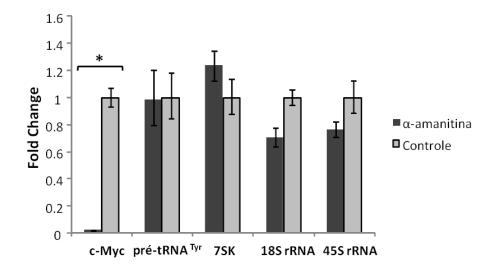


Figura 14: Inibição da expressão de c-Myc pelo tratamento de células HeLa com α-amanitina. Células HeLa foram incubadas na presença ou ausência de 50µg/mL de α-amanitina durante 9 h e os níveis de cada transcrito foram determinados por qPCR. mRNA de c-Myc foi utilizado como controle da eficiência da inibição da RNA Polimerase II. Os RNAs pre-tRNA^{Tyr} e 7SK, transcritos pela RNA Polimerase III, e os RNAs 18S rRNA e 45S rRNA, transcritos pela RNA Polimerase I, foram utilizados como controles negativos. O 5S rRNA, transcrito pela RNA Polimerase III, foi utilizado como normalizador do experimento. Os valores estão expressos em *fold change* relativo às células controle não- tratadas (barras cinza). As medidas foram realizadas em 4 réplicas independentes. O asterisco indica que a diferença entre as condições é estatisticamente significativa (teste t, p<0,05).

Após a validação inicial do tratamento, fizemos a amplificação e marcação com Cy3 ou Cy5 de quantidades iguais de RNA tratado com α-amanitina e RNA controle e hibridizamos estes alvos fluorescentes com duas lâminas de oligoarranjo intron-exon, conforme mostrado na figura abaixo (figura 15). Cada réplica biológica referente a uma condição foi marcada separadamente com Cy3 e Cy5, de modo que o fluoróforo da amostra tratada e controle foram trocados (*dye swap*), permitindo assim a hibridização em duplicata para cada réplica biológica (réplica técnica) (figura 15). Essa estratégia *dye swap* foi utilizada com o intuito de corrigir viéses resultantes do uso de dois fluoróforos distintos, eliminando assim possíveis diferenças na incorporação dos fluoróforos e diferenças na intensidade de

captação do sinal quando os mesmos são excitados com luz UV no momento do *scanning* (Quackenbush 2002).



Figura 15: Esquema das hibridizações no oligoarranjo de 4x44k. A lâmina contém 4 *arrays* independentes. RNA amplificado e marcado com cy3 ou cy5, proveniente de células tratadas e nãotratadas com α-amanitina, foi hibridizado em cada *array*. Para cada réplica biológica hibridizada, hibridizamos também uma réplica técnica (*dye swap*). Os experimentos foram realizados com 4 réplicas independentes de cada condição.

Comparando-se os valores das intensidades dos sinais obtidos da hibridização de réplicas técnicas e biológicas das amostras tratadas com α- amanitina, foram observados altos índices de correlação, sendo a média entre estes 0,98± 0,01. O mesmo padrão foi observado comparando-se as réplicas técnicas e biológicas das amostras controle, sendo a média das correlações obtidas 0,98± 0,01, o que indica uma boa reprodutibilidade das hibridizações realizadas. Com a finalidade de preservar possíveis transcritos cuja inibição de sua transcrição pela α-amanitina resultasse em sua completa depleção após 9h de tratamento, consideramos nesta análise todas as sondas detectadas acima do *background* nas 8 réplicas controle. Esse critério foi ecolhido pelo fato de ser esperado maior sinal nestas amostras, pelo menos para as sondas de transcritos exônicos, pois sabidamente são transcritos pela RNAP II.

Dados de expressão gênica gerados em experimentos de hibridização com oligoarranjos utilizando 2 cores, geralmente, necessitam de uma etapa de normalização para corrigir viéses resultantes da incorporação preferencial ou diferenças de eficiência de emissão de luz pelos diferentes fluoróforos, como citado anteriormente (Quackenbush 2002). Para a correção destes efeitos, o programa que faz a extração dos dados de intensidade das lâminas de oligoarranjo, Feature Extraction, aplica um método de normalização baseado em Lowess (Local Weighted Scaterplot Smoothing). Este método assume a existência de um mesmo número de genes diferencialmente expressos aumentados e diminuídos entre duas amostras em uma determinada faixa de intensidade (Yang et al. 2002). Contudo, no experimento de inibição da transcrição com α-amanitina espera-se que, ao menos no que diz respeito aos transcritos exônicos que são sabidamente transcritos pela RNAP II, a abundância relativa dos transcritos diminua ao longo do tempo de tratamento com o inibidor. Assim, a utilização da estratégia de normalização por *Lowess* poderia distorcer os dados de expressão, forçando que as razões se distribuíssem em torno de 1 (ou log2 = 0). Como alternativa para corrigir os efeitos de incorporação diferencial dos fluoróforos e de diferença de sensibilidade do scanner para os diferentes comprimentos de onda de cada fluoróforo, foi utilizada a hibridização réplica com troca de fluoróforo (dye swap), como mostra a equação: 0,5*log2[(controle Cy5 $/\alpha$ -amanitina Cy3)*(controle Cy3/ α -amanitina Cy5)]). Desta forma, ao final do dye swap ficamos com medidas da razão em log2 (controle/α-amanitina) de 4 réplicas biológicas independentes, cada uma proveniente da média de suas duas réplicas técnicas. Adicionalmente, foi incluído um filtro para selecionar apenas transcritos que apresentaram mudança de abundância, ou seja, diminuição ou aumento após tratamento com α- amanitina, concordantes em todas as réplicas (tabela 5).

Pelo fato de um dos controles negativos utilizados nos experimentos de PCR em tempo real, 18S rRNA, ter apresentado uma diminuição de cerca de 30% após o tratamento

com a droga, apesar desta proporção não ter sido considerada estatisticamente significante (teste t, p<0,05), optamos por considerar um critério de no mínimo 40% de diferença entre as medidas de intensidade de α-amanitina e controle para aumentar a confiabilidade de que a diferença observada entre as amostras é proveniente do efeito inibitório da droga. Utilizando este critério, os transcritos considerados como detectados e que apresentaram pelo menos 40% de diferença entre as intensidades de α-amanitina e controle (equivalente a um *fold-change* ≥ 1,66) em cada uma das 4 réplicas de razão resultantes da média do *dye swap*, foram submetidos a um teste estatístico específico para análise de oligoarranjos denominado SAM (*Significance Analysis of Microarrays*) (Tusher et al. 2001). Utilizando-se o SAM, obtivemos um conjunto de transcritos considerados diferencialmente detectados, com q-valor <0,05%, entre as amostras de RNA provenientes de células tratadas com α-amanitina e controle (tabela 5).

Tabela 5: Número de sondas de cada classe detectadas no oligoarranjo e número de sondas diferencialmente detectadas entre células HeLa tratadas e não-tratadas com α-amanitina durante 9h.

Classe de transcrito	#Sondas presentes no oligoarranjo	#Sondas detectadas (% do total da classe)	#Sondas diferencialmente detectadas (% das detectadas)
Exônicos	8.786	6.403 (73%)	3.868 (60%)
TIN antissenso	7.135	1.008 (14%)	344 (34%)
TIN senso	7.135	1.500 (21%)	853 (57%)
PIN antissenso	4.440	1.482 (33%)	556 (38%)

Através do gráfico MxA (Yang et al. 2002), que compara as razões, no caso, considerando-se o *dye swap* das réplicas, (0,5*log2[(controle Cy5/α-amanitina Cy3)*(controle Cy3/α-amanitina Cy5)]) em função das intensidades (0,25*log2[controle Cy5*α-amanitina Cy3*controle Cy3*α-amanitina Cy5]), pode-se perceber que todos os transcritos considerados diferencialmente expressos após o tratamento com α-amanitina estão aumentados nas

amostras controle em relação às amostras tratadas, o que significa que tiveram sua abundância diminuída após o tratamento (figura 16).

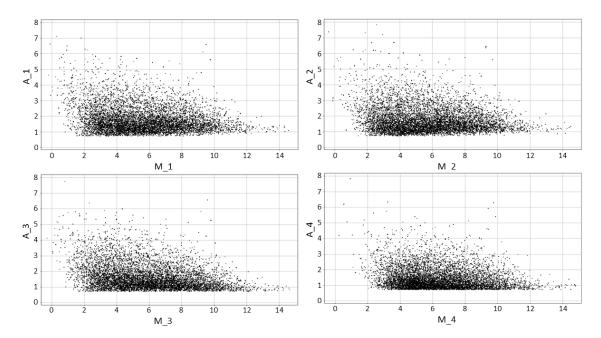


Figura 16: Gráficos M vs. A das hibridizações em oligoarranjos intron-exon de RNA de células HeLa tratadas e não-tratadas com α-amanitina. O logaritimo da razão (controle/α-amanitina) das medidas de intensidades (M) dos transcritos considerados diferencialmente expressos após o tratamento com α-amanitina (fold change \ge 1,66, q-valor <0,05%) está plotado em função do logaritmo do produto das intensidades (controle*α-amanitina) (A). Os números 1, 2, 3 e 4 referem-se a 4 hibridizações réplicas independentes.

Cerca de 60% dos transcritos exônicos detectados (3.868/6.403) apresentaram variação significativa com o tratamento com α-amanitina (tabela 5). Uma fração comparável de sondas que interrogam TIN senso apresentou variação significativa com o tratamento (57%). Em contraste, uma porcentagem menor de transcritos intrônicos não-codificadores antissenso TIN e PIN, 34% e 38%, respectivamente, foi afetada pelo tratamento com o inibidor de RNAP II. Portanto, parece haver uma resposta diferente à inibição da RNAP II entre transcritos intrônicos e exônicos e também entre intrônicos com orientação antissenso e senso.

Como verificamos que apenas 60% dos transcritos exônicos foram considerados como diferencialmente detectados em nossa análise e que, sabidamente, a totalidade dos mRNAs são transcritos pela RNAP II, fomos verificar se aqueles transcritos exônicos considerados como não-diferencialmente detectados possuem maior meia-vida em relação aos diferencialmente detectados, o que apoiaria o resultado observado. Para isto, utilizamos os dados provenientes das análises do tratamento com actinomicina D (vide item 4.2) (figura 17). Verificamos que, de fato, os transcritos exônicos considerados como não-diferencialmente detectados em nossa análise apresentaram maior meia-vida em relação aos diferencialmente detectados (teste KS, p<0,001). Fizemos a mesma análise para os transcritos não-codificadores intrônicos (figura 17) e observamos resultado semelhante para as três classes de ncRNAs intrônicos (teste KS, p<0,001), o que sugere que, pelo menos, uma fração dos transcritos intrônicos que consideramos como não-diferencialmente detectados em nossa análise, também possam ser produtos de síntese da RNAP II que possuem meia-vida mais longa.

Como foi visto que os transcritos exônicos possuem meia-vida mais curta que os intrônicos, sobretudo que os antissensos (vide item 4.2), a observação de que uma fração maior de exônicos foi afetada pelo tratamento com α-amanitina pode ser uma consequência da diferença de estabilidade entre estas classes de RNAs. Em virtude desta diferença, foi avaliado o efeito da inibição da RNAP II na abundância dos transcritos exônicos e intrônicos que apresentaram meia-vida igual ou superior a 3h, na tentativa de eliminar parte da discrepância de meia-vida entre estes transcritos (tabela 6). Uma proporção equivalente a 56% do total de sondas para transcritos exônicos apresentou meia-vida igual ou superior a 3h (tabela 6). Para os ncRNAs intrônicos, 94%, 71% e 86% do total de sondas para os transcritos TIN antissenso, TIN senso e PIN antissenso, respectivamente, apresentaram meia-vida igual ou superior a 3h (tabela 6).

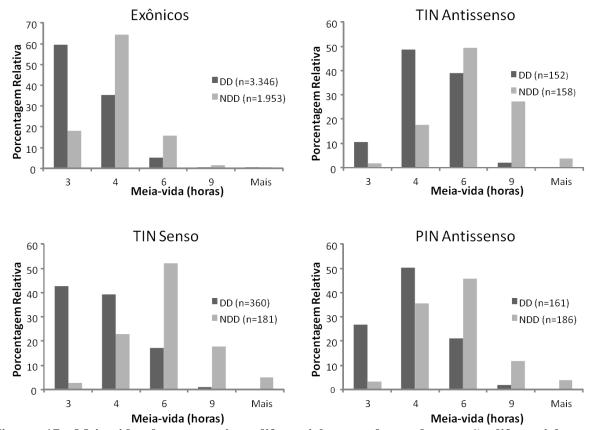


Figura 17: Meia-vida dos transcritos diferencialmente detectados e não-diferencialmente detectados após tratamento de células HeLa com 50μg/mL de α-amanitina por 9h. Para determinação da meia-vida destes transcritos foram utilizados os dados referentes ao experimento de células HeLa tratadas com actinomicina D, descritos no item 4.2. O número de transcritos está indicado entre parênteses. DD (Diferencialmente Detectados), NDD (Não Diferencialmente Detectados). As distribuições de meia-vida dos transcritos DD e NDD mostraram-se estatisticamente diferentes (teste KS, p<0,001) para todas as classes de transcritos analisadas (exônicos e intrônicos antissenso e senso).

Das sondas para mensagens codificadoras de proteínas detectadas com meia-vida igual ou superior a 3h, uma proporção de 46% mostrou-se significativamente (q-valor <0,05%) diminuída com o tratamento com α-amanitina (tabela 6). Proporções semelhantes foram observadas para transcritos intrônicos, de modo que 47% dos TIN antissenso, 54% dos TIN senso e 40% dos PIN senso detectados com meia-vida igual ou superior a 3h apresentaram diminuição significativa (q-valor <0,05%) com o tratamento (tabela 6). A observação de proporções semelhantes de transcritos das diferentes classes, exônicos e intrônicos senso e antissenso afetados pelo tratamento, quando analisados RNAs com meia-vida superior a 3h,

sugere que os ncRNAs intrônicos apresentam a mesma sensibilidade a α -amanitina que os mRNAs.

Tabela 6: Número de sondas de cada classe detectadas no oligoarranjo e número de sondas diferencialmente detectadas entre células HeLa tratadas e não-tratadas com α -amanitina durante 9h, considerando-se apenas os transcritos com meia-vida superior a 3h.

Classe do transcrito	#Sondas presentes no oligoarranjo	#Sondas detectadas (% do total da classe)	#Sondas detectadas com dados de meia-vida (% do total das detectadas)	#Sondas detectadas com meia- vida ≥ 3h (% do total das detectadas com dados de meia-vida)	#Sondas diferencialmente detectadas com meia-vida ≥ 3h (% do total das detectadas com meia-vida≥ 3h)
Exônicos	8.786	6.403 (73%)	5.299 (83%)	2.963 (56%)	1.358 (46%)
TIN	7.135	1.008 (14%)	310 (31%)	291 (94%)	136 (47%)
antissenso					
TIN senso	7.135	1.500 (21%)	541 (36%)	382 (71%)	206 (54%)
PIN	4.440	1.482 (33%)	347 (23%)	298 (86%)	118 (40%)
antissenso			•		· · · · · · · · · · · · · · · · · · ·

4.4 Avaliação da presença de cap 5' em ncRNAs intrônicos

Para investigar se os transcritos ncRNA intrônicos possuem a modificação 7-metil guanosina na extremidade 5' (cap 5'), característica de mRNAs transcritos pela RNAP II, foi proposto como estratégia experimental avaliar se o tratamento enzimático para remoção do cap 5' com a enzima *Tobacco Acid Pyrophosphatase* (TAP) facilita a posterior digestão dos ncRNAs intrônicos por uma enzima 5'-exonuclease (5'-Exo). Os RNAs com a extremidade 5' protegida pela presença de cap são resistentes à digestão pela 5'-Exo, enquanto RNAs sem cap ou parcialmente degradados são digeridos pela enzima.

Foram feitas três condições de tratamento: i) amostra tratada com TAP e 5'-Exo (TAP⁺/5'-Exo⁺), ii) amostra tratada apenas com 5'-Exo (TAP⁻/5'-Exo⁺) e iii) amostra controle, sem tratamento com as enzimas Tobacco e 5'-Exonuclease (TAP⁻/5'-Exo⁻). A amostra TAP⁻

/5'-Exo foi submetida à incubação com mesmo tampão e temperatura do tratamento com a enzima TAP e em seguida incubada com mesmo tampão e temperatura do tratamento com a enzima 5'-Exo, contudo sem adição de nenhuma das enzimas. De modo semelhante, fizemos incubação prévia da amostra a ser tratada exclusivamente com 5'-Exo, com o mesmo tampão e temperatura do tratamento com a enzima TAP (TAP-/5'-Exo+), o que assegura que as amostras em estudo (TAP+/5'-Exo+, TAP-/5'-Exo+ e TAP-/5'-Exo-) ficaram sujeitas as mesmas condições e portanto, susceptíveis ao mesmo nível de degradação inespecífica do RNA por RNAses contaminantes (figura 18). Foram feitas 4 réplicas independentes de cada uma das 3 condições de tratamento.

A eficácia do ensaio enzimático realizado para determinação de presença ou ausência de cap 5' foi verificada através de PCR quantitativa em tempo real. O snoRNA (*small nucleolar* RNA), snRNA U15A, que não possui cap 5' (Tycowski et al. 1993), foi usado como controle negativo para o tratamento com as enzimas (figura 19, painel à direita) e como normalizador do experimento de PCR quantitativa (figura 19, painel à esquerda).

As amostras tratadas com 5'-Exo, com e sem tratamento prévio com TAP (TAP+/5'-Exo+ e TAP-/5'-Exo+) foram amplificadas, marcadas com Cy3 ou Cy5 e hibridizadas com oligoarranjos intron-exon para verificação da presença de cap 5' entre os transcritos não-codificadores intrônicos (figura 20). A amostra TAP-/5'-Exo- foi utilizada apenas como controle da qualidade do RNA após os períodos de incubação, mas não foi hibridizada no oligoarranjo (figura 18).

Para avaliar a presença de cap 5' em ncRNAs intrônicos em larga-escala, as 4 réplicas independentes de RNA tratado com as enzimas TAP e 5'-Exo (TAP+/5'-Exo+ e TAP-/5'-Exo+), após amplificação e marcação com Cy3 ou Cy5, foram hibridizadas no oligoarranjo intronexon . Foi feita uma marcação invertida, *dye swap* dos fluoróforos, para cada uma das réplicas independentes, de forma que foram hibridizadas duas réplicas técnicas de cada réplica

biológica. Novamente, como descrevemos no item 4.3, utilizamos a hibridização com *dye swap* para corrigir possíveis efeitos provenientes da co-hibridização de duas amostras marcadas com fluoróforos distintos. Contudo, tivemos problemas na hibridização de uma das duas lâminas utilizadas, ficando com os dados resultantes de apenas uma das lâminas, referentes a 2 réplicas independentes (figura 20).

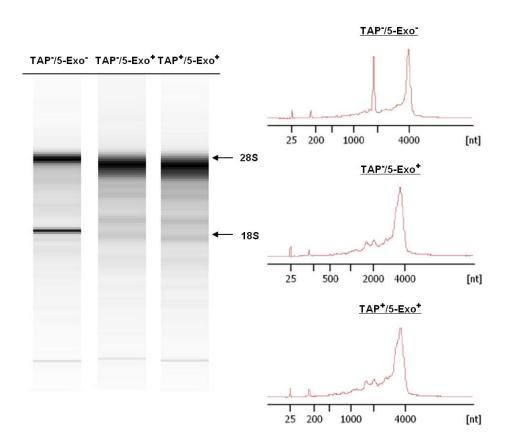


Figura 18: Eletroforese capilar das amostras de RNA total de células HeLa tratadas com as enzimas 5'-Exonuclease (5'-Exo) e *Tobacco Acid Pyrophosphatase* (TAP). Os painéis representam o perfil analítico de preparações de 200ng de RNA total após tratamento com as enzimas TAP'/5'-Exo⁺, TAP'/5'-Exo⁺ e TAP⁺/5'-Exo⁺, obtidos a partir de eletroforese capilar no Agilent 2100 Bioanalyzer (Agilent Technologies). O painel à esquerda mostra a imagem virtual da eletroforese reconstruída após análise no Bioanalyzer para as três condições em análise. O painel à direita apresenta o eletroferograma, com os dois picos maiores representando os rRNAs 18S e 28S (da esquerda para a direita, respectivamente). As bandas das subunidades 28S e 18S do rRNA, com uma razão próxima de 2, na amostra TAP'/5-Exo⁻ atestam a integridade da amostra de RNA utilizada como controle para a degradação por RNAses contaminantes. É possível verificar que houve uma degradação das bandas dos RNAs ribossomais 28S e 18S nas amostras TAP'/5'-Exo⁺ e TAP⁺/5'-Exo⁺, o que era desejável, visto que estas duas amostras foram tratadas com a enzima 5'-Exo.

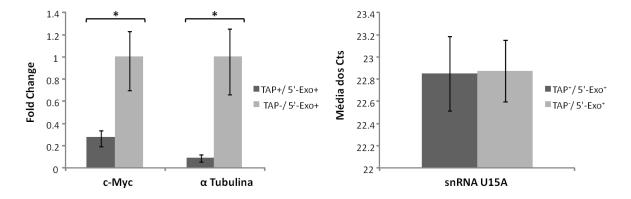


Figura 19: Avaliação da eficácia do ensaio utilizando as enzimas *Tobacco Acid Pyrophosphatase* (TAP) e 5'-Exonuclease (5'-Exo) para determinação da presença ou ausência de cap 5'. RNA total de células HeLa foi submetido a tratamento com as enzimas i) TAP, que remove o cap 5', e 5'-Exo, que degrada RNAs 5'-monofosfato (TAP⁺/5'-Exo⁺) , ou ii) apenas com 5'-Exo (TAP⁻/5'-Exo⁺). Em seguida, os níveis de cada transcrito foram determinados por qPCR. mRNA de α-Tubulina e c-Myc foram utilizados como controles positivos para a presença de cap 5' (painel à esquerda) e o snRNA U15A foi utilizado como controle negativo (painel à direita). Os valores estão expressos em *fold change* relativo às células controle (TAP⁻/5'-Exo⁺) e normalizados pelos níveis do snRNA U15A (barras cinzas). As medidas foram realizadas em 4 réplicas independentes. O asterisco indica que a diferença entre as condições é estatisticamente significativa (teste t, p<0,05).

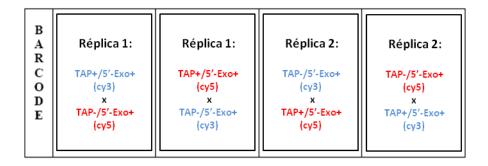


Figura 20: Esquema das hibridizações no oligoarranjo de 4x44k. A lâmina contém 4 *arrays* independentes. Amostras de RNA TAP⁺/5'-Exo⁺ e TAP⁻/5'-Exo⁺ amplificadas e marcadas com Cy3 ou Cy5 foram hibridizadas em cada *array*. Para cada réplica biológica hibridizada, hibridizamos também uma réplica técnica (*dye swap*). Os experimentos foram realizados com 2 réplicas independentes de cada condição.

Foram observados altos índices de correlação entre os valores de intensidades dos sinais obtidos na hibridização de réplicas técnicas e biológicas das amostras tratadas com TAP⁺/5'-Exo⁺, sendo a média entre estes 0,97± 0,01. O mesmo padrão foi observado comparando-se as réplicas técnicas e biológicas das amostras tratadas com TAP⁻/5'-Exo⁺,

sendo a média das correlações obtidas 0,97± 0,01, apontando para a boa reprodutibilidade dos experimentos.

Todas as sondas que apresentaram intensidade de sinal acima do *background* nas 4 réplicas TAP'/5'-Exo⁺ hibridizadas foram consideradas como detectadas na análise (tabela 7). Esta abordagem possibilitou que fossem mantidos na análise os transcritos com expressão ausente em TAP⁺/5'-Exo⁺, o que é desejável. Como no caso da análise de α-amanitina (vide item 4.3), optamos por não utilizar um método de normalização clássico baseado em *Lowess* (Yang et al. 2002). Também nos experimentos de cap 5' assumimos a existência predominante de transcritos com expressão diminuída ou inalterada (no caso do transcrito não ter cap 5') ao comparar-se RNA tratado com TAP'/5'-Exo⁺ em relação a RNA tratado com TAP'/5'-Exo⁺. Para cada uma das sondas, utilizamos os dados referentes à média da razão das intensidades do *dye swap* de duas réplicas técnicas hibridizadas (Yang et al. 2002): 0,5*log2{[TAP'/5'-Exo⁺(Cy5)/TAP⁺/5'-Exo⁺(Cy3)]*[TAP/5'Exo⁺(Cy3)/TAP⁺/5'-Exo⁺(Cy5)]. Desta forma, partimos de 4 réplicas (2 biológicas e 2 técnicas) de cada condição de tratamento (TAP'/5'-Exo⁺ e TAP⁺/5'-Exo⁺) e ao final do *dye swap* ficamos com 2 medidas de razão das intensidades em log2 (TAP'/5'-Exo⁺/ TAP⁺/5'-Exo⁺).

Para um transcrito ser considerado portador de cap 5', a sua razão TAP/5'-Exo⁺/TAP⁺/5'-Exo⁺ deveria ser igual ou maior que 2 (*fold-change* ≥ 2) nas duas réplicas, o que corresponde a pelo menos 50% de degradação do transcrito após remoção do cap 5' (tabela 7). Utilizando este critério, conseguimos confirmar a presença desta modificação cotranscricional em 59% dos transcritos exônicos detectados em nosso modelo (tabela 7). Os 41% dos transcritos codificadores de proteínas, cuja modificação cap 5' não foi detectada, apontam para uma deficiência de sensibilidade do método empregado ou da alta estringência da análise. Contudo, optamos por utilizar um critério estrito para considerar um transcrito com cap 5' pelo fato de termos apenas 2 réplicas de medidas de razão das intensidades, o que

impossibilitou a aplicação de métodos estatísticos robustos para identificação de transcritos diferencialmente detectados entre as amostras TAP⁻/5'-Exo⁺ e TAP ⁺/5'-Exo⁺.

Tabela 7: Número de sondas de cada classe detectadas no oligoarranjo e número de sondas com cap 5'.

Classe do transcrito	#Sondas presentes no oligoarranjo	#Sondas detectadas (% do total da classe)	#Sondas com cap 5' (% das detectadas)
Exônicos	8.786	7.216 (82%)	4.289 (59%)
TIN antissenso	7.135	1.948 (27%)	989 (51%)
TIN senso	7.135	2.861 (40%)	865 (30%)
PIN antissenso	4.440	2.796 (63%)	1.807 (65%)

A presença de cap 5' foi observada em 51% dos transcritos TIN antissenso, 30% dos TIN senso e 65% dos PIN antissenso detectados no oligoarranjo (tabela 7). Ao menos metade dos ncRNAs intrônicos TIN antissenso e mais da metade dos transcritos PIN antissenso, detectados em nossa análise, apresentaram pelo menos 50% de degradação após o tratamento com enzimas que removem o cap 5' e degradam RNAs 5'-monofosfato, indicando a presença de cap na extremidade 5' destes RNAs em proporção, em geral, comparável a observada para os mRNAs (tabela 7).

Dentre as três classes de ncRNAs intrônicos analisadas, é bem visível que os transcritos TIN senso foram os que apresentaram a resposta mais discrepante aos tratamentos enzimáticos para detecção da presença de cap 5', em relação aos transcritos exônicos, controles positivos para a presença de modificação cap 5'. Como é sabido que a modificação cap 5' é importante para a proteger os RNAs contra a degradação por exonucleases 5'→3' (Wilusz et al. 2001), fomos verificar a meia-vida dos transcritos TIN senso considerados com e sem cap 5' (figura 21). Para fins comparativos, avaliamos também a meia-vida dos transcritos considerados com e sem cap 5' das demais classes, TIN e PIN antissenso e exônicos (figura 21). A distribuição de meia-vida dos transcritos considerados com e sem cap 5' não foi estatisticamente diferente entre os transcritos exônicos (teste KS), como já era

esperado, visto que todos os transcritos codificadores de proteína apresentam cap 5', independente de sua meia-vida. Para os transcritos TIN e PIN antissenso a distribuição de meia-vida dos transcritos considerados com e sem cap 5' também não foi diferente (teste KS), demonstrando que, de modo semelhante ao que se observa para os transcritos exônicos, parece não existir diferença, pelo menos de tempo de meia-vida, entre os transcritos considerados com e sem cap 5' nas análises realizadas (figura 21).

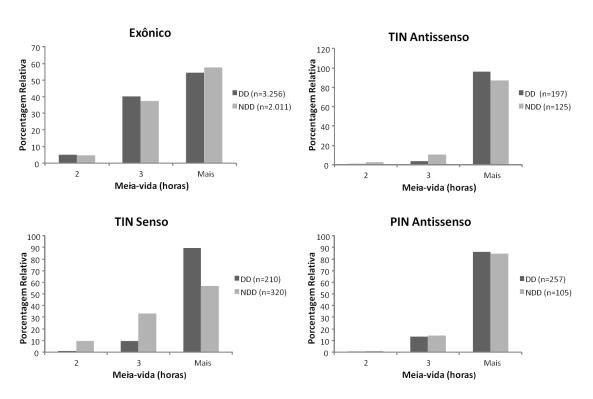


Figura 21: Meia-vida dos transcritos diferencialmente detectados após o tratamento com as enzimas TAP e/ou 5'-Exo, ou seja, considerados com e sem cap 5'. Alíquotas de RNA de células HeLa foram tratadas com as enzimas TAP /5'-Exo⁺ ou TAP +5'-Exo⁺ para a detecção de cap 5'. Para determinação da meia-vida foram utilizados os dados referentes ao experimento de células HeLa tratadas com actinomicina D, descritos no item 4.2. DD (Diferencialmente Detectados) e NDD (Não-Diferencialmente Detectados). O número de transcritos está indicado entre parênteses. As distribuições de meia-vida dos transcritos DD e NDD mostraram-se estatisticamente diferentes (teste KS, p<0,05) apenas para a classe de transcritos TIN senso.

Nota-se que, novamente, a classe TIN senso apresentou comportamento distinto das demais, sendo a diferença entre as distribuições de meia-vida dos transcritos considerados

com e sem cap 5' estatisticamente significante (teste KS, p <0,001) (figura 21). Uma proporção maior (43%) de transcritos TIN senso considerados como sem cap 5' apresentou meia-vida mais curta (<3h) em relação aos transcritos considerados com cap 5' (10%). Contudo, ainda assim é possível afirmar que pelo menos uma fração dos ncRNAs intrônicos da classe TIN senso sofre modificação co-transcricional por adição de cap 5', semelhante ao processamento que ocorre com os RNAs codificadores de proteínas e demais ncRNAs intrônicos (tabela 7).

4.5 Análise da localização sub-celular de ncRNAs intrônicos

A fim de documentar a distribuição dos ncRNAs intrônicos na célula, foram realizados experimentos de fracionamento celular para obtenção de frações nucleares e citoplasmáticas de células HeLa, através de lise em condições que rompem a membrana plasmática, mas mantém o envelope nuclear intacto, seguido da marcação e hibridização de sondas fluorescentes com oligoarranjos intron-exon.

A qualidade dos RNAs obtidos no fracionamento foi avaliada por eletroforese capilar. A observação de razões 28S/18S próximas de 2 nas amostras fracionadas confirmou a integridade do RNA disponível para os experimentos planejados (figura 22). Nota-se um sensível enriquecimento do precursor rRNA 32S nas fração nuclear em relação à fração citoplasmática. Estes resultados são semelhantes aos obtidos com o método original de fracionamento descrito na literatura (Gondran et al. 1999) e indicam que a contaminação nuclear nas frações citoplasmáticas obtidas é ausente ou muito limitada. Para análise da pureza das preparações de núcleo e citoplasma obtidas, foram também realizados ensaios de *northern blot* com sondas para RNAs com localização predominantemente nuclear (U6 snRNA) ou citoplasmática (tRNA^{Lys}) (figura 23). Os ensaios de *northern blot* confirmaram o enriquecimento dos controles nuclear e citoplasmático, respectivamente, nas frações

enriquecidas em RNA nuclear e citoplasmático, e demonstraram a existência de baixa contaminação cruzada entre as frações, atestando para a boa qualidade das preparações (figura 23).

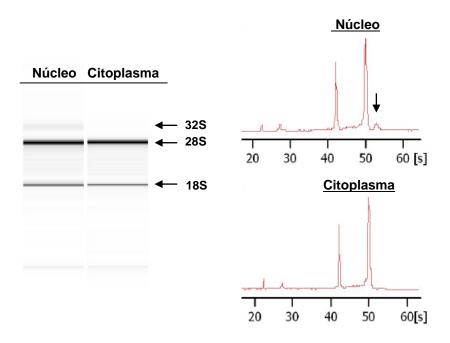


Figura 22: Integridade e qualidade do RNA total de células HeLa obtido por fracionamento celular. Os painéis representam o perfil analítico das preparações de 200ng de RNA total de citoplasma e núcleo após tratamento com DNAse, obtidos a partir de eletroforese capilar no Agilent 2100 Bioanalyzer (Agilent Technologies). O painel à esquerda mostra a imagem de eletroforese virtual reconstruída após análise no Bioanalyzer. O painel à direita apresenta o eletroferograma, com os dois picos maiores representando os rRNAs 18S e 28S (da esquerda para a direita, respectivamente). A seta indica a presença do precursor 32S rRNA, observado apenas na fração nuclear. O enriquecimento do precursor 32S na fração nuclear indica boa qualidade do fracionamento e limitada contaminação nuclear das frações citoplasmáticas. A razão próxima de 2 das bandas das subunidades 28 S e 18 S do rRNA indica boa integridade das amostras de RNA utilizadas.

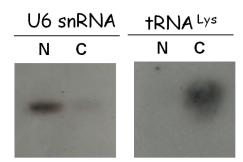


Figura 23: Obtenção de frações enriquecidas em RNA nuclear ou citoplasmático. Alíquotas de frações enriquecidas em RNA nuclear ou citoplasmático (10 μg de RNA total), isoladas de células HeLa, foram analisadas por *northern blot* utilizando sondas biotiniladas complementares a RNAs com localização predominantemente nuclear (U6 snRNA) ou citoplasmática (tRNA ^{Lys}).

Para o experimento de hibridização no oligoarranjo intron-exon, alíquotas de frações de RNA nuclear e citoplasmático preparadas em paralelo, a partir de duas culturas de células HeLa, foram amplificadas, marcadas fluorescentemente (Cy3 ou Cy5) e hibridizadas em quadriplicata. Foi feita uma marcação invertida, *dye swap* dos fluoróforos, para cada fração de RNA nuclear e citoplasmático independente, de forma que foram hibridizadas duas réplicas técnicas (marcação e hibridização independentes) de cada réplica biológica (fracionamento independente). O esquema das hibridizações está apresentado na figura 24.

A reprodutibilidade dos experimentos foi avaliada através da medida da correlação dos valores de intensidade entre as réplicas dos experimentos (biológicas e técnicas). Foi obtida uma boa correlação entre as réplicas de núcleo (r=0,92±0,04) e citoplasma (r=0,86±0,05) comparadas separadamente. Medidas de correlação muito inferiores foram obtidas ao compararmos as intensidades das amostras de núcleo e citoplasma (r=0,48±0,08), indicando a existência de grandes diferenças entre as duas frações.

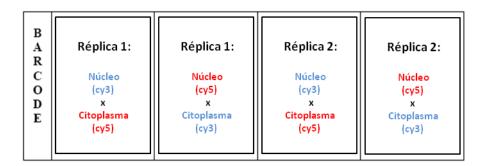


Figura 24: Esquema das hibridizações no oligoarranjo de 4x44k. A lâmina contém 4 *arrays* independentes. RNA de frações de núcleo e citoplasma de células HeLa foi amplificado, marcado com Cy3 ou Cy5 e hibridizado em cada *array*. Para cada réplica de fracionamento independente hibridizada, hibridizamos também uma réplica técnica (*dye swap*).

Para considerar uma sonda como detectada em nosso experimento, estipulamos como critério a detecção acima do *background* nas 4 réplicas de núcleo e citoplasma (tabela 8). A correlação relativamente baixa (r=0,48±0,08) observada entre as medidas de intensidade das

réplicas de núcleo e citoplasma foi um indicativo que as duas frações celulares apresentam perfis de expressão gênica bem distintos, o que já se encontra documentado na literatura (Schwanekamp et al. 2006; Trask et al. 2009). Com isso, novamente, como nas análises de α-amanitina e cap 5' (vide itens 4.3 e 4.4), não foi possível a utilização de métodos de normalização como *Lowess*, visto que teríamos que assumir que a maioria das sondas estivessem igualmente expressas em ambos os canais (Yang et al. 2002). Alternativamente, foi utilizada a média da razão das intensidades do *dye swap* das duas hibridizações réplicas (Yang et al. 2002), como mostra a equação: 0,5*log2[(Núcleo Cy5/Citoplasma Cy3)*(Núcleo Cy3/Citoplasma Cy5)]) para correção dos efeitos de incorporação diferencial dos fluoróforos e de diferença de sensibilidade do *scanner* para os diferentes comprimentos de onda de cada fluoróforo. Desta forma, ao final do *dye swap* ficamos com medidas de razão de intensidade em log2 (núcleo/citoplasma) de 2 réplicas biológicas independentes, cada uma proveniente da média de suas duas réplicas técnicas.

Como a maioria do RNA total da célula localiza-se no citoplasma (Meister et al. 2004; Barthelson et al. 2007) e foi utilizada a mesma massa de RNA nuclear e citoplasmático nas hibridizações no oligoarranjo intron-exon, existe a possibilidade de um viés experimental que refletisse em um aparente enriquecimento nuclear. Para atenuar este efeito, calculamos um fator de correção (2,1) baseado nas proporções de massa de RNA total nuclear e citoplasmático obtidas em oito fracionamentos independentes. Nos fracionamentos realizados, o RNA total nuclear representou, em média, 32% ±8,7 do RNA total da célula, de maneira que a massa de RNA total citoplasmática correspondeu a mais que o dobro da massa de RNA nuclear (68% do RNA total da célula). Fizemos também um ajuste adicional, para corrigir diferenças de amplificação entre as amostras de núcleo e citoplasma. Para isto, calculamos um fator de correção baseado nas diferenças entre as medidas de intensidades observadas para um

conjunto de RNAs sintéticos controle adicionado às alíquotas de RNA (*spikes*) e detectados por sondas controles presentes na lâmina (vide item 3.10.2).

Após a realização do dye swap das réplicas técnicas, ficamos com 2 medidas de razão das intensidades (núcleo/citoplasma). Com isso, não foi possível a utilização de métodos estatísticos para a identificação de sondas significativamente diferencialmente distribuídas entre as frações de núcleo e citoplasma. Alternativamente, foi estabelecido um critério de que a razão núcleo/citoplasma deveria ser maior ou igual a 3 (fold- $change \ge 3$) nas duas réplicas, para considerar uma sonda como enriquecida nas frações nuclear ou citoplasmática. Os dados obtidos estão apresentados na figura 25 e tabela 8.

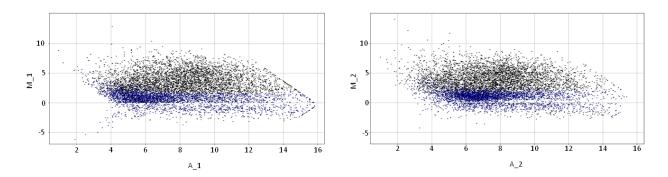


Figura 25: Gráficos M vs. A das hibridizações em oligoarranjo intron-exon de RNA das frações núcleo e citoplasma de células HeLa. O logaritimo da razão (núcleo/ citoplasma) das medidas de intensidades (M) está plotado em função do logaritmo de seus produtos (núcleo*citoplasma) (A). Os pontos pretos são os transcritos considerados enriquecidos em uma das frações, núcleo ou citoplasma, (fold change ≥3 nas duas réplicas) e os pontos azuis são os transcritos não considerados como enriquecidos em uma das frações. Os números 1 e 2 referem-se a 2 hibridizações réplicas independentes.

É interessante notar que os mRNAs codificadores de proteínas apresentaram maior fração de transcritos considerados acumulados em um dos compartimentos celulares (68% dos detectados) comparados aos ncRNAs intrônicos, TIN antissenso (17% dos detectados), TIN senso (46% dos detectados) e PIN antissenso (32% dos detectados) (tabela 8). Observamos que quase a totalidade dos mRNAs codificadores considerados enriquecidos em um dos compartimentos apresentou-se enriquecido no núcleo (98% do total dos mRNAs enriquecidos

em um dos compartimentos) (tabela 8). De forma semelhante, verificamos que para os ncRNAs intrônicos que apresentaram enriquecimento em um dos compartimentos, este enriquecimento ocorreu, quase que exclusivamente, no núcleo da célula (TIN antissenso 99%, TIN senso 99% e PIN antissenso 97%) (tabela 8). A classe PIN antissenso foi a que apresentou maior fração de transcritos enriquecidos no citoplasma (3% do total dos RNAs enriquecidos em um dos compartimentos).

Tabela 8: Distribuição de mRNAs codificadores de proteína e ncRNAs intrônicos nas frações nuclear/citoplasmática de células HeLa.

presentes no oligoarranjo	detectadas (% do total da classe)	enriquecidas em uma das frações (% das detectadas)	nuclear (% das enriquecidas)	citoplasmático (% das enriquecidas)
8.786	5.002 (57%)	3.422 (68%)	3.366 (98%)	56 (2%)
7.135	755 (11%)	129 (17%)	128 (99%)	1 (1%)
7.135	1.221 (17%)	566 (46%)	561 (99%)	5 (1%)
4.440	1.224 (26%)	390 (32%)	377 (97%)	13 (3%)
	8.786 7.135 7.135	8.786 5.002 (57%) 7.135 7.135 1.221 (17%)	(% do total da classe)	(% do total da classe)

É importante destacar que existe uma fração maior de ncRNA intrônicos que são detectados em ambos os compartimentos e não foram considerados enriquecidos em nenhum deles (*fold change* <3, tabela 9). Portanto, embora não se observe enriquecimento no citoplasma, estes transcritos parecem ser exportados para este compartimento, onde podem estar desempenhando alguma função.

Para excluir a possibilidade de que os ncRNAs intrônicos detectados como enriquecidos no citoplasma pudessem ter potencial para codificar proteínas, fizemos uma análise utilizando o programa CPC (*Coding Potential Calculator*) (Kong et al. 2007). Verificamos que a maior parte dos transcritos intrônicos acumulados no citoplasma (18 do total de 19) não apresentaram potencial codificador, indicando que a maior parte dos

transcritos intrônicos acumulados no citoplasma são, de fato, RNAs não codificadores de proteínas.

Tabela 9: Distribuição de mRNAs codificadores de proteína e ncRNAs intrônicos não enriquecidos em uma das frações, núcleo/citoplasma, de células HeLa (fold change <3 em pelo menos uma das duas réplicas). NE (Não Enriquecidos).

Classe do transcrito	#Sondas NE em uma das frações (% do total de detectadas- tabela 8)	#Sondas NE com maior intensidade no núcleo (% do total de NE)	#Sondas NE com maior intensidade no citoplasma (% do total de NE)
Exônicos	1.580 (32%)	1.149 (73%)	431 (27%)
TIN antissenso	626 (83%)	540 (86%)	86 (14%)
TIN senso	655 (54%)	588 (90%)	67 (10%)
PIN antissenso	834 (68%)	703 (84%)	131 (16%)

Com o intuito de entender um pouco mais sobre a dinâmica de distribuição dos RNAs dentro da célula, combinamos as análises de tratamento com actinomicina D (vide item 4.2) e fracionamento celular núcleo/citoplasma. Comparamos a distribuição das medidas de meiavida dos RNAs considerados enriquecidos no núcleo (EN), no citoplasma (EC) e a dos não enriquecidos em um dos compartimentos (NE) (fold change <3 em pelo menos uma das duas réplicas), tanto para os transcritos exônicos quanto para os intrônicos (figura 26). Não foi possível fazer essa análise para cada classe de ncRNAs intrônicos separadamente, devido ao pequeno número de transcritos enriquecidos no citoplasma em cada uma destas classes.

A comparação da meia-vida dos RNAs EN com a dos RNAs EC, tanto para mRNAs codificadores de proteínas quanto para os ncRNAs intrônicos, mostrou que os transcritos considerados como enriquecidos no núcleo possuem meia-vida mais curta (mediana mRNA= 3h; ncRNA intrônicos= 3,1h) que os enriquecidos no citoplasma (mediana mRNA= 4h; ncRNA intrônicos= 4,3h) (teste KS, p <0,001) (figura 26). Os RNAs NE parecem ter uma estabilidade intermediária entre os EN e EC. Verificamos que os RNAs NE apresentaram maior estabilidade que os RNAs EN, tanto para os mRNAs (mediana= 3,4h vs. mediana= 3h,

respectivamente) quanto para os ncRNAs intrônicos (mediana= 4,2h vs. mediana= 3,1h, respectivamente) (teste KS, p <0,001) (figura 27). Os mRNAs NE mostraram-se significativamente diferentes dos mRNAs EC (mediana= 3,4h vs. mediana= 4h, respectivamente) (teste KS, p <0,001), mas o mesmo resultado não foi observado ao comparar-se os ncRNAs intrônicos NE e EC, provavelmente devido ao pequeno conjunto amostral (n=12) (figura 26).

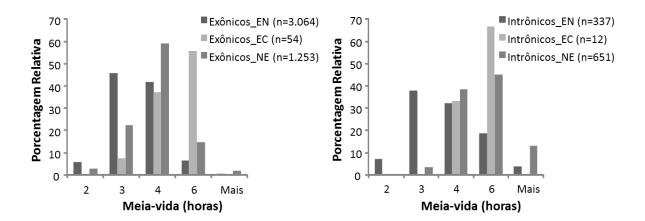


Figura 26: Distribuição das medidas de meia-vida dos transcritos considerados como enriquecidos no núcleo, no citoplasma e dos transcritos não considerados enriquecidos em um dos compartimentos, através de fracionamento de células HeLa. Para determinação da meia-vida destes transcritos, foram utilizados os dados referentes ao experimento de células HeLa tratadas com actinomicina D, descritos no item 4.2. EN (Enriquecidos no Núcleo), EC (Enriquecidos no Citoplasma) e NE (Não Enriquecidos em um dos compartimentos). O número de transcritos está indicado entre parênteses. As distribuições de meia-vida dos RNAs EN e EC mostraram-se estatisticamente diferentes (teste KS, p<0,05) para os transcritos exônicos e intrônicos e as distribuições NE e EC foram estatisticamente diferentes apenas para os transcritos exônicos (teste KS, p<0,05).

Para avaliar a presença de categorias gênicas com funções específicas entre os mRNAs enriquecidos nos sub-compartimentos celulares foi investigado o enriquecimento de termos de ontologia gênica (GO), incluindo os domínios função molecular, processo biológico e componente celular dos mRNAs acumulados nas fração nuclear (3.033 transcritos; *fold change* \geq 5) e citoplasmática (56 transcritos; *fold change* \geq 3) de células HeLa (tabelas 10 e 11). Vários termos de GO mostraram-se representados significativamente

(p<0,05) entre os transcritos enriquecidos no núcleo. Estes termos incluem processos biológicos como ciclo celular, mitose, processamento de mRNA, resposta a dano de DNA e resposta celular ao estresse. Os transcritos exônicos acumulados no citoplasma apresentaram enriquecimento na categoria função molecular, constituinte estrutural de ribossomo. Nenhuma categoria de GO foi encontrada significativamente enriquecida para os ncRNAs intrônicos enriquecidos no núcleo ou no citoplasma.

Tabela 10: Análise de enriquecimento em categorias de ontologia gênica (GO) de transcritos exônicos enriquecidos no núcleo.

Categoria	Termo	# de genes	P-valor	P-valor
				corrigido*
GOTERM_CC_FAT	GO:0005654~nucleoplasm	249	<0,001	<0,001
GOTERM_CC_FAT	GO:0005730~nucleolus	176	< 0,001	< 0,001
GOTERM_CC_FAT	GO:0005856~cytoskeleton	310	<0,001	<0,003
GOTERM_MF_FAT	GO:0003723~RNA binding	179	< 0,001	< 0,001
GOTERM_CC_FAT	GO:0005813~centrosome	77	< 0,001	< 0,001
GOTERM_CC_FAT	GO:0005874~microtubule	77	< 0,001	< 0,007
GOTERM_MF_FAT	GO:0004386~helicase activity	58	< 0,001	< 0,001
GOTERM_BP_FAT	GO:0000280~nuclear division	66	< 0,001	< 0,002
GOTERM_BP_FAT	GO:0007067~mitosis	66	< 0,001	< 0,002
GOTERM_CC_FAT	GO:0005819~spindle	60	< 0,001	< 0,001
GOTERM_BP_FAT	GO:0007049~cell cycle	205	< 0,001	< 0,002
GOTERM_BP_FAT	GO:0048285~organelle fission	67	< 0,001	< 0,002
GOTERM_BP_FAT	GO:0030163~protein catabolic process	170	< 0,001	<0,003
GOTERM_BP_FAT	GO:0006259~DNA metabolic process	132	< 0,001	< 0,005
GOTERM_BP_FAT	GO:0033554~cellular response to stress	147	< 0,001	< 0,03
	GO:0006974~response to DNA damage		< 0,001	< 0,03
GOTERM_BP_FAT	stimulus	107		
GOTERM_MF_FAT	GO:0000166~nucleotide binding	504	< 0,001	< 0,003
GOTERM_MF_FAT	GO:0005524~ATP binding	357	< 0,001	< 0,05
GOTERM_BP_FAT	GO:0016071~mRNA metabolic process	107	< 0,001	< 0,003
GOTERM_BP_FAT	GO:0006396~RNA processing	137	< 0,001	< 0,002
GOTERM_CC_FAT	GO:0005694~chromosome	123	<0,001	< 0,001
	GO:0010608~posttranscriptional regulation		< 0,001	
GOTERM_BP_FAT	of gene expression	72		< 0,002
GOTERM_BP_FAT	GO:0043039~tRNA aminoacylation	20	<0,001	<0,03
	GO:0000775~chromosome, centromeric		< 0,001	
GOTERM_CC_FAT	region	36		< 0,002
GOTERM_CC_FAT	GO:0000793~condensed chromosome	37	<0,001	<0,03
GOTERM_CC_FAT	GO:0000151~ubiquitin ligase complex	35	< 0,002	< 0,05

^{*}P-valor corrigido para múltiplo teste (Benjamini-Hocheberg).

Tabela 11: Análise de enriquecimento em categorias de ontologia gênica (GO) de transcritos exônicos enriquecidos no citoplasma.

Categoria	Termo	# de genes	P-valor	P-valor
				corrigido*
GOTERM_CC_FAT	GO:0044429~mitochondrial part	9	<0,001	<0,006
GOTERM_CC_FAT	GO:0070469~respiratory chain	4	<0,001	<0,009
GOTERM_CC_FAT	GO:0005739~mitochondrion	10	<0,004	< 0,03
GOTERM_MF_FAT	GO:0003735~structural constituent of		<0,001	
	ribosome	5		< 0,02
GOTERM_CC_FAT	GO:0005840~ribosome	5	< 0,001	< 0,005

^{*}P-valor corrigido para múltiplo teste (Benjamini-Hocheberg).

5. DISCUSSÃO

5.1 Meia-vida de ncRNAs intrônicos e de mRNAs codificadores de proteínas

A estabilidade de um transcrito é determinada pelo equilíbrio entre as suas taxas de síntese e de degradação (Yang et al. 2003). A degradação do RNA, por sua vez, é um componente importante na regulação gênica, que participa do controle da concentração de cada transcrito no estado estacionário (Sharova et al. 2009). Em humanos, estima-se que a estabilidade seja responsável pelo controle do nível de mRNA de 5-10% de todos os genes (Bolognani and Perrone-Bizzozero 2008). Alterações no controle da estabilidade de muitos mRNAs têm sido relacionadas a diversos estados patológicos como câncer, respostas inflamatórias crônicas e doenças coronarianas (Wilusz et al. 2001). Apesar da importância da taxa de degradação para determinação dos níveis funcionais de cada transcrito na célula, pouco se sabe sobre a estabilidade dos RNAs não-codificadores de proteínas. Alguns poucos estudos realizados até o momento estimaram a meia-vida de ncRNAs individualmente (Wilusz et al. 2008; Redrup et al. 2009; Sunwoo et al. 2009). Nenhum trabalho descreveu globalmente a estabilidade de ncRNAs, que possuem funções ainda pouco conhecidas na célula. Com o intuito de avançar o conhecimento neste âmbito ainda pouco explorado, avaliamos a meia-vida de RNAs não-codificadores intrônicos em células HeLa, utilizando um oligoarranjo customizado.

Como controle, inicialmente, caracterizamos a meia-vida dos mRNAs representados no oligoarranjo e para os quais já existem relatos na literatura. A mediana da meia-vida estimada para os mRNAs codificadores de proteínas foi de 3,1h, tempo relativamente curto comparado com a mediana de 10h, estimada por um trabalho descrevendo análise de meia-vida em larga escala em células de carcinoma de fígado humano, HepG2 (Yang et al. 2003). A sobreposição limitada entre os mRNAs medidos neste trabalho com os mRNAs, cuja meia-

vida foi estimada no estudo de Yan e colaboradores, pode explicar parcialmente as diferenças observadas. Uma explicação alternativa para estas diferenças de meia-vida é que, apesar de ambas as linhagens serem humanas, o tempo de duração do ciclo celular de cada uma é bem distinto, 50h em HepG2 (Yang et al. 2003) e 14-16h em células HeLa (Whitfield et al. 2002). Trabalhos na literatura já descreveram a existência de uma relação proporcional entre a meia-vida e o tempo de duração do ciclo celular (*Cell Cycle Length*- CCL), onde o tempo mediano de meia-vida dos mRNAs expressos foi cerca de 20-25% do tempo de extensão do ciclo celular. Isto foi visto para vários organismos, incluindo células humanas (HepG2; mediana da meia-vida= 10h, CCL= 50h) (Yang et al. 2003), *Escherichia coli* (mediana da meia-vida= 5 min, CCL= 20 min) (Bernstein et al. 2002), levedura (mediana da meia-vida= 3,8h, CCL= 19h) (Narsai et al. 2007). Esta mesma relação também foi observada em nossa análise, onde a mediana da meia-vida representou 19-22% do tempo de duração do ciclo celular de HeLa.

Comparando-se as taxas de decaimento medidas em nosso trabalho, em HeLa, com as medidas obtidas por Yang e colaboradores, em HepG2, foi observado uma correlação estatisticamente signicativa (r=0,52; p<0,001). Apesar dos experimentos terem sido realizados com tipos celulares distintos, em laboratórios diferentes, tempos de tratamento com actinomicina D diferentes e utilizando diferentes plataformas de microarranjo, observamos uma correlação moderada entre as taxas de decaimento, demonstrando haver consistência entre as medidas obtidas em ambas as linhagens celulares, apesar da discrepância entre as meias-vidas estimadas. Também foi observada uma correlação moderada e estatisticamente significativa entre as taxas de decaimento em HeLa e células embrionárias de camundongo, MC1, (r=0,53; p<0,001), indicando haver uma conservação entre as vias de degradação entre espécies de mamíferos, o que encontra-se em conformidade com o resultado observado por Sharova e colaboradores (Sharova et al. 2009). O fato de não termos observado uma maior

correlação ao compararmos as taxas de decaimento de HeLa com outra linhagem humana (HepG2), frente à correlação obtida entre HeLa e uma linhagem de camundongo, deve-se possivelmente a diferença entre o número de genes para os quais foi estimada uma taxa de decaimento entre os dois trabalhos (5.245 genes em HepG2 e 19.977 genes em MC1). Apesar de HeLa e HepG2 pertencerem a mesma espécie, a estimativa de meia-vida para ambos os trabalhos foi realizada em plataformas diferentes e que não incluíam todos os genes humanos, enquanto o trabalho usando MC1 utilizou uma plataforma contendo o genoma completo de camundongo, possibilitando uma maior sobreposição entre o número de genes com medida de meia-vida entre HeLa e MC1.

Os mRNAs detectados como menos estáveis na análise de meia-vida realizada, apresentaram enriquecimento em categorias gênicas relacionadas à regulação da transcrição, ciclo celular, regulação da apoptose, entre outros. Outros trabalhos também observaram enriquecimento de mRNAs menos estáveis em categorias semelhantes (Yang et al., 2003; Narsai et al., 2007; Sharova et al., 2009; Friedel et al., 2009). Uma possível interpretação para este resultado é que um alto turnover do RNA possibilita alterações rápidas nas suas concentrações no estado estacionário, através de mudanças na taxa de transcrição, o que é importante para mRNAs com funções regulatórias na célula (Yang et al. 2003; Friedel et al. 2009). Em contraste, mudanças nas taxas de transcrição de mRNAs mais estáveis demoram mais tempo para alterar os níveis totais do mRNA, e uma vez estabelecidas na célula, persitem por mais tempo, o que é desejável para mensagens que codificam proteínas que participam em processos constitutivos da célula. Esta característica poderia justificar a observação de enriquecimento de genes envolvidos com elongação da tradução e proteínas estruturais do ribosomo entre os mRNAs com meia-vida mais longa. Este resultado também foi amparado por descrições prévias na literatura (Narsai et al. 2007; Friedel et al. 2009; Sharova et al. 2009).

De maneira interessante, a estabilidade observada para os transcritos intrônicos, foi maior do que a observada para os RNAs codificadores de proteínas. A mediana da meia-vida calculada para os ncRNAs TIN antissenso, TIN senso e PIN antissenso foi, respectivamente, 4,2h, 3,4h e 3,9h, enquanto a mediana da meia-vida estimada para os RNAs codificadores de proteínas foi de 3,1h. Para assegurar que os resultados obtidos não estivessem atrelados as diferenças nas abundâncias entre os transcritos, principalmente pelo fato de os ncRNAs intrônicos serem menos abundantes que os mRNAs, adotamos um critério bastante rigoroso na análise. Somente sondas com intensidade de sinal pelo menos 2 vezes maior que o background nas 20 réplicas técnicas da amostra referência foram mantidas na análise, previnindo, deste modo, a superestimação da estabilidade de transcritos expressos em níveis próximos ao background (Sharova et al. 2009). Além disso, comparamos também a distribuição de meia-vida entre transcritos intrônicos e exônicos em uma mesma faixa de intensidade. Ao comparar as distribuições da meia-vida de ncRNAs intrônicos e mRNAs, com intensidade de sinal menor que 100 relativo ao tempo 0h de tratamento com actinomicina D, continuamos obtendo o mesmo resultado, demonstrando assim, que a maior estabilidade observada para os transcritos intrônicos não é um viés associado a sua menor abundância. Este resultado encontra-se de acordo com o trabalho de Sharova e colaboradores, que não observaram correlação entre a estabilidade de mRNAs e seus níveis de expressão em células embrionárias de camundongo (Sharova et al. 2009).

Dentre os transcritos intrônicos, os TIN senso apresentaram uma maior proporção de transcritos com meia-vida mais curta (<3h). A distribuição de meia-vida desta classe de ncRNAs se sobrepôs a das mensagens codificadoras de proteínas nos tempos menores que 3h, o que não ocorreu para os TIN e PIN antissenso. Entretanto, observamos também a existência, nesta classe, de uma maior proporção de transcritos com meia-vida mais longa (>6h) comparado aos mRNAs, o que contribuiu para que suas distribuições de meia-vida fossem

estatisticamente diferentes (teste KS, p<0,001). Estes resultados parecem indicar que a classe TIN senso é constituida por uma população heterogênea de RNAs, o que poderia explicar as diferenças observadas na distribuição de meia-vida em relação aos TIN e PIN antissenso. Como estes transcritos têm a mesma orientação dos transcritos codificadores de proteína, é possível que parte destes RNAs, especialmente os que possuem meia-vida mais curta, representem trechos de introns removidos pelo mecanismo de splicing ou ainda, fragmentos de pré-mRNA. Embora tenhamos selecionado frações de RNA poliadenilado, utilizando primers oligo dT para o preparo das amostras de cRNA que foram hibridizadas (vide item 3.10.2, figura 1), existe a possibilidade de que regiões ricas em adenina dentro de introns tenham servido como sítios internos de anelamento, possibilitando a transcrição reversa destas mensagens durante o preparo do cRNA (Furuno et al. 2006). Por outro lado, já foi documentado a existência de ncRNAs intrônicos com a mesma orientação do mRNA expresso no locus gênico e que exercem uma função na célula. Por exemplo, o ncRNA COLDAIR (Cold Assisted Intronic Noncoding RNA), em Arabidopsis, é um ncRNA intrônico longo com orientação senso, que promove a repressão gênica do mRNA codificador de proteínas FLC, que é transcrito no mesmo locus, através do direcionamento do complexo PRC2 para a cromatina do seu gene alvo (Heo and Sung 2011).

O conceito que introns recém-excisados acumulam-se em baixos níveis na célula, porque são rapidamente degradados, é largamente difundido (Clement et al. 1999). Entretanto, existem evidências da existência de RNAs intrônicos, excisados de transcritos primários de mRNA (pré-mRNA), com específica localização sub-celular e meia-vida relativamente significante, comparável a meia-vida dos mRNAs c-Myc e c-Fos que possuem meia-vida curta (Clement et al. 1999; Clement et al. 2001). Estas evidências corroboram a hipótese de que ao menos uma fração dos transcritos TIN senso com menor meia-vida possam representar introns removidos do pré-mRNA pelo mecanismo de *splicing* e que são estabilizados na

célula. Alternativamente, alguns dos transcritos TIN senso podem representar introns na vizinhança de exons específicos de isoformas alternativas de determinados pré-mRNAs que ainda não sofreram splicing. Uma publicação recente, em cérebro fetal humano, mostrou que isoformas gênicas provenientes de splicing alternativo são mais lentamente processadas do que isoformas gênicas constitutivas (Ameur et al., 2011). Através de sequenciamento em larga escala do transcriptoma, foi obtida uma cobertura de sequência significativamente maior em introns mapeados na vizinhança de exons específicos de isoformas alternativas, comparado a introns na vizinhança de exons de isoformas constitutivas (Ameur et al. 2011). De maneira muito interessante, outro trabalho observou que, enquanto o splicing constitutivo ocorre cotranscricionalmente no sítio de transcrição, o que já é um consenso, alguns processamentos de splicing alternativo ocorrem pós-transcricionalmente na sub-estrutura nuclear speckles, onde fatores de splicing e reguladores transcricionais estão concentrados (Vargas et al. 2011). Estas descobertas reforçam a possibilidade de que parte dos transcritos TIN senso possam representar mensagens imaturas de mRNA, que ainda serão destinadas ao processamento póstranscricional por splicing alternativo. Portanto, é necessário certa cautela ao considerar a classe TIN senso como constituída exclusivamente por RNAs não-codificadores gerados a partir de promotores independentes do mRNA no mesmo *locus*.

Cabe ainda ressaltar que, apesar das evidências obtidas de que os ncRNAs intrônicos apresentam certa estabilidade, não se pode desconsiderar que transcritos que são rapidamente degradados possam ser estabilizados, sob determinadas circunstâncias, para o desempenho de alguma função na célula. Como exemplo, um trabalho realizado com leveduras verificou que RNAs antissenso expressos no *locus* PHO84 acumulam-se em células com mutações no componente Rrp6 do exossomo ou durante o envelhecimento celular (Camblong et al. 2007). Em condições normais, estes RNAs antissenso são normalmente transcritos e prontamente degradados, mas mediante a uma mudança de condição fisiológica são estabilizados na célula

e exercem sua função de reprimir a expressão do gene PHO84, por direcionar a histona deacetilase Hda 1 para o seu *locus* (Camblong et al. 2007).

Por fim, alguns trabalhos têm mostrado ncRNAs longos com importantes funções biológicas que apresentam certo grau de estabilidade. MALAT1, um ncRNA longo, envolvido em vários tipos de cânceres, retido em *speckles* nucleares e regulador de *splicing* alternativo mostrou-se muito estável, com uma meia-vida maior que 12h (Hutchinson et al. 2007; Wilusz et al. 2008; Tripathi et al. 2010). RNAs não-codificadores longos como o Air, Xist e Kcnq1ot1, com meia-vida estimada de 2,1h, 4,6h e 3,4h, respectivamente, têm sido associados com regulação epigenética por promovem *imprinting* em sua vizinhança genômica através do recrutamento de complexos de silenciamento da cromatina (Seidl et al. 2006; Redrup et al. 2009). Os transcritos não-codificadores de proteína MEN epsilon e MEN beta são RNAs estáveis, com meia-vida de 4h e 8h, respectivamente, e são essenciais para a integridade estrutural da sub-estrutura nuclear *paraspeckles* (Sunwoo et al. 2009). A observação de que ncRNAs intrônicos possuem, em média, estabilidade igual ou superior a de mRNAs, contribui para o argumento de que estes transcritos podem desempenhar funções importantes na célula que não dependam de um rápido *turnover* (Mattick 2003; Szymanski et al. 2003; Reis et al. 2004; Reis et al. 2005).

5.2 RNA Polimerase II e a biogênese de ncRNAs intrônicos

Um trabalho prévio de nosso grupo, realizado em células de adenocarcinoma de próstata LNCaP e com a mesma plataforma de oligoarranjo intron-exon usada neste estudo, verificou que a maior parte dos transcritos intrônicos, que foi alterada após o tratamento das células com α-amanitina, apresentou uma diminuição significativa em sua abundância, indicando que sua transcrição é realizada pela RNAP II (Nakaya et al., 2007). Uma vez que a expressão de ncRNAs é geralmente descrita como sendo altamente tecido-específica

(Kapranov et al. 2002; Ravasi et al. 2006; Nakaya et al. 2007; Sasaki et al. 2007; Mercer et al. 2008), neste trabalho revisitamos na linhagem HeLa o papel da RNAP II na biogênese de ncRNAs intrônicos. Foi observado que a totalidade dos transcritos intrônicos interrogados no oligoarranjo teve sua abundância significativamente diminuida após tratamento das células HeLa com α-amanitina (SAM, q-valor <0,05%).

Cerca de 60% das mensagens codificadoras de proteínas mostraram-se significativamente diminuídas (SAM, q-valor <0,05%) após a inibição da transcrição mediada pela α-amanitina, enquanto a fração restante, presumivelmente composta por mRNAs mais estáveis, não foi afetada. Uma fração menor de ncRNAs intrônicos antissenso foi afetada pelo tratamento com α-amanitina (TIN antissenso 34% e PIN antissenso 38%) frente à proporção de mensagens codificadoras de proteínas afetada (60%). A observação de que transcritos TIN e PIN antissenso têm uma maior proporção de RNAs com meia vida-longa quando comparados aos mRNAs codificadores de proteínas (vide item 5.1) sugere que, de forma semelhante ao observado para mRNAs codificadores de proteína, a diferença de sensibilidade à α-amanitina seja explicada pela maior estabilidade destes RNAs.

Ao serem avaliadas apenas as proporções de transcritos intrônicos antissenso e exônicos com meia-vida mais longa (maior ou igual a 3h) que foram significativamente diminuídos (q-valor <0,05%) com o tratamento, a diferença inicial observada entre as frações de transcritos TIN e PIN antissenso e exônicos (TIN antissenso 34%, PIN antissenso 38%, exônicos 60%) afetadas pela inibição da transcrição torna-se bastante reduzida (TIN antissenso 47%, PIN antissenso 40% e exônicos 46%). Esta aproximação das proporções de transcritos intrônicos antissenso e de mRNAs afetados pela α-amanitina, ao comparar-se apenas RNAs com meia-vida mais longa, corrobora a hipótese de que os ncRNAs intrônicos teriam uma mesma resposta à inibição da RNAP II que os transcritos codificadores de proteínas. Em concordância com a premissa que os transcritos não afetados pelo tratamento

com α-amanitina são mais estáveis, observamos que os mRNAs e ncRNAs intrônicos antissenso não-diferencialmente detectados após o tratamento apresentaram meia-vida mais longa que os diferencialmente detectados (KS, p<0,001). Essa observação corrobora, mais uma vez, com a conjectura que os RNAs não afetados de maneira significativa pelo tratamento com α-amanitina devem também ser transcritos pela RNAP II.

Os transcritos intrônicos TIN senso mostraram uma resposta bem distinta ao tratamento com α-amanitina em relação aos demais ncRNAs intrônicos, o que foi evidenciado por uma fração de transcritos com detecção diminuída equiparável a dos mRNAs (57%). Este resultado encontra-se em conformidade com a maior proporção de transcritos com meia-vida curta observada nesta classe, frente às classes de ncRNAs intrônicos antissenso (vide item 5.1). De maneira semelhante ao padrão obtido para os mRNAs e transcritos intrônicos antissenso, verificamos que os ncRNAs intrônicos TIN senso não-diferencialmente detectados apresentaram meia-vida mais longa que os diferencialmente detectados (KS, p<0,001). Desta forma, apesar de ser necessário cautela ao considerar todos os transcritos desta classe como ncRNAs, visto que alguns destes RNAs podem constituir-se em pré-mRNAs ainda não totalmente processados pela maquinaria de *splicing* (Ameur et al. 2011) (vide item 5.1), podese dizer que estes transcritos com orientação senso foram afetados pela α-amanitina de maneira similar aos mRNAs (exônicos 60%, TIN senso 57%, considerando-se todos os detectados; exônicos 46% e TIN senso 54%, considerando-se detectados com meia-vida maior ou igual a 3h).

Os resultados obtidos através da inibição da RNAP II por α-amanitina demonstraram que as três classes de ncRNAs intrônicos, TIN antissenso, TIN senso e PIN antissenso foram afetadas pelo tratamento com o inibidor de forma forma semelhante às mensagens codificadoras de proteínas. Deste modo, sugerimos que, possivelmente, não apenas os transcritos considerados como significativamente afetados pelo tratamento, mas uma fração

maior, abrangendo boa parte dos ncRNAs intrônicos não significativamente afetados, teriam como enzima transcricional a RNAP II.

Além dos experimentos de inibição com α-amanitina, foram também realizadas análises informáticas para obter evidências adicionais do envolvimento da RNAP II na transcrição de ncRNAs intrônicos. O cruzamento das coordenadas genômicas dos ncRNAs com as coordenadas de sítios de ligação de RNAP II obtidos a partir de experimentos de ChIP-seq (Rozowsky et al. 2009) revelou o enriquecimento significativo da polimerase na vizinhança dos transcritos intrônicos que foram expressos em células HeLa (vide item 4.1). Como a RNAP II se acumula em regiões próximas aos promotores de genes ativos e inativos (Brodsky et al. 2005; Kim et al. 2005; Guenther et al. 2007; Muse et al. 2007; Rahl et al. 2010; Jishage et al. 2012), estes dados também reforçam a noção de que transcritos intrônicos, especialmente uma fração dos TIN com orientação senso, sejam originados a partir de unidades transcricionais independentes. Em concordância com estes resultados, duas análises informáticas adicionais usando dados disponíveis UCSC Genome no Browser (http://genome.ucsc.edu/), onde as coordenadas genômicas dos ncRNAs foram cruzadas com as coordenadas de tags de cap 5' (CAGE tags) e com coordenadas de regiões de cromatina apresentando H3K4me3, uma modificação de histona relacionada a promotores transcrionalmente ativos, também mostraram enriquecimento destes dois elementos na vizinhança do início de transcrição conhecido para os ncRNAs intrônicos expressos em HeLa. A única exceção foi a classe PIN antissenso, que não apresentou enriquecimento na marca H3K4me3 em relação ao grupo controle. Contudo, pode-se sugerir que em conjunto estas observações contribuem para o argumento de que os ncRNAs intrônicos antissenso e, ao menos uma parte dos que possuem orientação senso, sejam transcritos pela RNAP II, a partir de regiões promotoras semelhantes às que controlam a transcrição de mRNAs codificadores de proteínas.

Apesar destes resultados coletivamente apontarem para uma possível biogênese dos ncRNAs intrônicos pela RNAP II, não se pode excluir a possibilidade de que alguns destes RNAs sejam transcritos por outras RNA polimerases. As três RNA polimerases de eucariotos podem ser diferenciadas através de suas diferentes sensibilidades à α-amanitina. A droga não afeta a atividade da RNAP I, inibe fortemente a RNAP II e em altas concentrações inibe a RNAP III (Jacob et al. 1970; Kedinger et al. 1970; Lindell et al. 1970). Neste trabalho, utilizamos α-amanitina na concentração de 50μg/ml, considerando que nesta concentração a droga tem efeito inibitório preferencialmente sobre a RNAP II. No entanto, há trabalhos que descrevem que em concentrações de α-amanitina que inibem seletivamente a RNA Polimerase II, um grupo específico de genes da RNAP III tem sua transcrição diminuída (Listerman et al. 2007; Raha et al. 2010). Portanto, é possível especular que alguns dos ncRNAs intrônicos com abundância diminuída após o tratamento com α-amanitina sejam alvos transcricionais da RNAP III, regulados indiretamente pela RNAP II.

5.3 ncRNAs intrônicos e a modificação cap 5'

A presença da modificação cap 5' é um forte indício de um evento de iniciação específico, promovido pela RNAP II (Kapranov 2009). Na maioria dos eucariotos, os transcritos sintetizados pela RNAP II são modificados co-transcricionalmente em sua extremidade 5' pela adição da estrutura 7-metil guanosina (cap 5') (Cougot et al. 2004). A presença de cap na extremidade 5' de ncRNAs tem sido reportada em uma série de publicações (Carninci et al. 2005; Wyers et al. 2005; 2009; Guttman et al. 2009; Flynn et al. 2011), mas até o momento ainda não foi documentada sistematicamente em transcritos não-codificadores longos originados em regiões intrônicas do genoma.

Visto que, boa parte dos ncRNAs intrônicos expressos em células HeLa são transcritos pela RNAP II, realizamos experimentos para obter medidas experimentais diretas da presença

da modificação cap 5' nestes transcritos mediante ensaio de digestão de RNA total com duas enzimas, uma que remove o cap 5' (TAP) e outra que degrada RNAs com 5'-monofosfato (5'-Exo) seguido de hibridização com oligoarranjo intron-exon. Considerando-se uma diminuição de pelo menos 50% na abundância do RNA da amostra tratada com TAP+/5'-Exo+ em relação à amostra tratada com TAP+/5'-Exo+ (vide item 4.4), como indicativo da presença de cap 5', observamos a presença da modificação cap 5' em uma fração considerável dos transcritos intrônicos; pelo menos metade dos ncRNAs intrônicos antissenso detectados em nossa análise apresentaram a modificação cap 5' (TIN 51% e PIN 65%). Visto que a totalidade dos mRNAs são transcritos pela RNAP II e possuem cap 5', chama a atenção que tenhamos detectado a presença de cap 5' em apenas 59% dos mRNAs representados no oligoarranjo. Esse resultado indica um limite de detecção de cap 5' na abordagem experimental utilizada e sugere que a fração de ncRNAs intrônicos antissenso considerada como dotada de cap 5' esteja subestimada.

De maneira intrigante, novamente, os transcritos TIN senso apresentaram uma comportamento diferente dos demais ncRNAs intrônicos, TIN e PIN antissenso (vide itens 5.1 e 5.2). Uma fração aproximadamente 2 vezes menor dos transcritos TIN senso apresentou a modificação cap 5' (30%), comparada às classes de intrônicos antissenso (TIN 51% e PIN 65%) e também aos mRNAs (59%). Como evidência adicional de que a classe TIN senso deve ser composta por uma população mais heterogênea de RNAs, verificou-se que, ao contrário do que foi observado para os ncRNAs intrônicos antissenso e para os mRNAs, a distribuição de meia-vida dos transcritos TIN senso, considerados com e sem cap 5', foi significativamente diferente (KS, p<0,001). Os transcritos TIN senso com cap 5' mostraram ter uma meia-vida mais longa que os sem cap 5'. Como a estrutura cap 5' é um fator importante na estabilidade de RNAs transcritos pela RNAP II (Cougot et al. 2004), o padrão de menor meia-vida verificada para os transcritos TIN senso considerados sem cap 5' é

bastante plausível. Consistente com esta noção, uma publicação atual verificou que em células mutantes de levedura que perderam a capacidade de remoção da estrutura cap 5' pelas proteínas DCP1 e DCP2, observa-se o acúmulo de diversos ncRNAs longos (Geisler et al. 2012). Neste estudo, os autores verificaram que um destes ncRNAs, o GAL10, desempenha um papel crítico na manutenção da repressão gênica de genes indutíveis por galactose, no seu próprio *locus*. Durante a ausência deste sinal de ativação, o ncRNA GAL10 é estabilizado, reprimindo a expressão gênica através de um mecanismo envolvendo deacetilação de histonas (Geisler et al. 2012). Na presença de galactose, a meia-vida deste ncRNA é reduzida drasticamente, através de uma via dependente da remoção da estrutura cap 5', de modo a permitir uma rápida ativação dos genes envolvidos no metabolismo da galactose (Geisler et al. 2012). Portanto, a estrutura cap 5' é um fator importante na modulação da estabilidade deste ncRNA.

Em conjunto, os dados gerados sugerem fortemente que a classe TIN senso seja composta por RNAs com e sem cap 5'. Uma parte desta população, com cap 5' e estabilizada por perídos mais longos na célula, parece ser sintetizada pela RNAP II a partir de unidades transcricionais independentes. Alternativamente, não se pode excluir a possibilidade de que alguns transcritos TIN senso representem pré-mRNA destinados a *splicing* alternativo (vide item 5.1), visto que o cap 5' é adicionado logo no início da transcrição (Bentley 2005; Ameur et al. 2011). Outra parte, sem cap 5' e com meia-vida mais curta, parece ser proveniente do processo de *splicing* do pré-mRNA. Uma vez que esses transcritos TIN senso sem cap 5' e com meia-vida mais curta se acumulam nas células e apresentam relativa estabilidade (3,2h), é plausível que este conjunto seja constituído por *intron lariats* que não sejam simples intermediários de degradação, mas que tenham adquirido alguma função celular.

5.4 Localização sub-celular de mRNAs e ncRNAs intrônicos

Neste estudo, observou-se a existência de grandes diferenças no acúmulo de RNAs entre os compartimentos núcleo e citoplasma de células HeLa. Em concordância com nosso resultado, na literatura já se encontra documentado que os níveis de RNA nuclear e citoplasmático são fracamente correlacionados e que existe um perfil de expressão gênica bem distinto ao se comparar RNA nuclear e citoplasmático (Schwanekamp et al. 2006; Trask et al. 2009). A investigação da localização sub-celular das mensagens codificadoras de proteínas, para as quais existem sondas no oligoarranjo utilizado, revelou que, da fração de 68% considerada enriquecida em um dos compartimentos, a maior parte (98%) encontra-se acumulada no núcleo da célula. Uma vez que a tradução de mRNAs ocorre no citoplasma, não está totalmente claro a razão da existência de um "pool" nuclear de mRNAs. Uma possibilidade é que o sinal medido no oligoarranjo possa refletir mRNAs imaturos. Essa hipótese é pouco provável, uma vez que foi utilizado um método que favorece a marcação de mRNAs poliadenilados e esta modificação 3' ocorre pós-transcricionalmente (Bentley 2005). Vários estudos apontam para a existência de RNAs poliadenilados estáveis enriquecidos no núcleo de mamíferos (Herman et al. 1976; Carter et al. 1991; Visa et al. 1993; Huang et al. 1994). Um trabalho mais recente, utilizando tiling arrays para examinar as sequências transcritas de dez cromossomos, em oito linhagens diferentes, verificou que a quantidade total de sequências poliadeniladas transcritas, detectadas exclusivamente no núcleo versus citoplasma, foi respectivamente, 10% e 3%, e que 25% do RNA poliadenilado nuclear foi associado com regiões exônicas bem anotadas (Cheng et al. 2005).

A existência de mecanismos de regulação da expressão gênica envolvendo a retenção de RNAs poliadenilados no núcleo poderia, pelo menos em parte, explicar o resultado observado de que a grande maioria dos mRNAs diferencialmente acumulados entre os compartimentos núcleo e citoplasma encontra-se enriquecida no núcleo da célula. O gene

mCAT2 (Mouse Cationic Amino Acid Transporter 2), codifica dois RNAs diferentes através do uso de promotores e sítios de poliadenilação alternativos (Prasanth et al. 2005). Um mRNA de ~4,2kb que codifica para a proteína mCAT2, responsável pelo transporte de aminoácidos catiônicos, e um RNA de ~8kb, denominado CTN-RNA que é retido no núcleo da célula, nos paraspeckles (Prasanth et al. 2005). O mecanismo de retenção envolve a modificação do CTN-RNA através da edição de resíduos de adenosina à inosina (A→I) na sua região 3'-UTR (discutido adiante) (Prasanth et al. 2005). Sob condições de estresse, o CTN-RNA é clivado pós-transcricionalmente na região 3'-UTR, eliminando o sinal de retenção nuclear e gerando um transcrito maduro contendo a região codificadora do mRNA mCAT2 e uma região 5'-UTR exclusiva (Prasanth et al. 2005). Este RNA é transportado para o citoplasma, onde é traduzido na proteína mCAT2, que desempenha um importante papel em resposta ao estresse (Prasanth et al. 2005). Deste modo, o acúmulo de CTN-RNA no núcleo propicia uma resposta rápida da célula ao estresse. De forma similar, Kay e colaboradores observaram que o *locus* gênico do fator estimulante de migração (MSF), que é uma isoforma truncada da fibronectina oncofetal, produz dois diferentes transcritos através sítios alternativos de poliadenilação, semelhante ao locus do mCAT2 (Kay et al. 2005). Os dois transcritos têm a mesma região codificadora e diferem apenas no comprimento de sua região 3'-UTR. Uma isoforma tem 5,9kb e é enriquecida na fração nuclear, enquanto a outra tem 2,1kb e é enriquecida na fração citoplasmática, onde codifica para a proteína MSF. Fibroblastos fetais secretores de MSF têm níveis significantemente mais baixos do mRNA de 5,9kb no núcleo e níveis mais altos do mRNA de 2,1kb no citoplasma, que fibroblastos adultos não secretores de MSF. Fibroblastos adultos, induzidos a secretar MSF através de tratamento com o fator de crescimento transformante- β1, apresentam mudanças nos níveis dos dois diferentes transcritos de MSF, de forma a apresentar um padrão de expressão semelhante ao encontrado para os fibroblastos fetais secretores. Baseado nisto, foi sugerido que a expressão da proteína MSF é regulada pela clivagem da região 3'-UTR do precursor de 5,9kb retido no núcleo. Estes dois exemplos constituem um mecanismo celular sofisticado para a rápida produção pós-transcricional de mRNAs citoplasmáticos, no qual um RNA capaz de codificar para proteínas é estocado no núcleo até que haja indução de seu processamento e transporte para o citoplasma para subsequente tradução.

Neste contexto, é interessante destacar que foi observado um enriquecimento significativo (teste Benjamini-Hocheberg, p<0,05) dos mRNAs acumulados no núcleo em categorias de ontologia gênica associadas à resposta ao estresse celular e dano de DNA. Este dado nos possibilita sugerir que alguns dos mRNAs acumulados no núcleo em células HeLa, também podem ter sua liberação para o citoplasma regulada mediante um estímulo externo de estresse ou dano de DNA. Contudo, ainda são necessários estudos detalhados dos mRNAs vistos como acumulados no núcleo para confirmar se o mecanismo de retenção nuclear, de fato, é responsável por parte do enriquecimento nuclear observado para os transcritos codificadores de proteínas.

Há relatos na literatura que a hiper-edição de bases A→I de regiões UTRs leva à retenção de mRNAs dentro do núcleo, diminuindo assim a síntese da proteína por ele codificada, através da prevenção de seu transporte para o citoplasma, como é o caso do CTN-RNA, citado anteriormente (Zhang and Carmichael 2001; Prasanth et al. 2005). Esta edição é uma modificação pós-transcricional de RNAs, catalisada por enzimas da família ADAR (*Adenosine Deaminases Acting on RNA*) e ocorre principalmente em regiões não-codificadoras do RNA, tipicamente em repetições Alu (Levanon et al. 2004; Eisenberg et al. 2005). Resíduos múltiplos de inosina no RNA editado interagem com proteínas presentes em *paraspeckles*, p54^{nrb}, PSF, matrin 3, PSP1, resultando na retenção do complexo ribonucleoprotéico no núcleo (Zhang and Carmichael 2001; Prasanth et al. 2005). Edições A→I podem ser detectadas comparando-se múltiplas sequências transcritas de um mesmo

gene, de forma que as adenosinas editadas são representadas como guanosinas nas sequências de cDNA.

Em uma pesquisa de todos os transcritos humanos presentes no Genbank, Levanon e colaboradores identificaram 1.637 genes com sítios de edição A→I e 92% destes sítios editados foram associados com repetições Alu (Levanon et al. 2004). Comparando-se a lista de transcritos exônicos enriquecidos no núcleo de células HeLa com a lista de genes identificados no trabalho de Levanon, verificamos sobreposição de apenas 192 transcritos (6% dos 3.366 transcritos exônicos enriquecidos no núcleo). É possível que alguns mRNAs identificados neste trabalho como enriquecidos no núcleo possam representar isoformas alternativas que ficam retidas neste compartimento a espera de um sinal molecular que promova seu transporte para o citoplasma, onde codifica para tradução de proteínas. Também, não se pode excluir a possibilidade de que outros mRNAs presentes em nossa lista de transcritos enriquecidos no núcleo também possam apresentar modificações A→I ainda não descritas.

De forma semelhante ao padrão observado para os mRNAs, os ncRNAs intrônicos que apresentaram enriquecimento em um dos compartimentos da célula, mostraram estar enriquecidos no núcleo (TIN antissenso 99%, TIN senso 99% e PIN antissenso 97%) (vide lista de ncRNAs enriquecidos no núcleo, anexo 1). Já foram descritos ncRNAs longos que acumulam-se no núcleo e exercem importantes papéis na regulação transcricional, remodelamento da cromatina e na arquitetura nuclear (Wilusz et al. 2009; Chen and Carmichael 2010). A função destes ncRNAs, muitas vezes, está intimamente associada à sua localização específica dentro do compartimento nuclear. É o caso, por exemplo, dos ncRNAs MEN epsilon e MEN beta, que ao mesmo tempo que acumulam-se em *paraspeckles*, são responsáveis pelo estabelecimento e manutenção desta sub-estrutura, de modo que sua depleção resulta na desintegração de *paraspeckles* (Clemson et al. 2009; Sasaki et al. 2009;

Sunwoo et al. 2009). Da mesma forma, MALAT1, um ncRNA acumulado em speckles, que é uma sub-estrutura nuclear envolvida no armazenamento, modificação e montagem da maquinaria de processamento de mRNA, modula o splicing alternativo através de sua interação com fatores de splicing SR (Serine/Arginine) localizados em speckles (Tripathi et al. 2010). Também podemos citar os ncRNAs Xist e Kcnq1ot1, que estão localizados na região perinucleolar durante a fase S do ciclo celular, e promovem o silenciamento da cromatina (Zhang et al. 2007; Mohammad et al. 2008). Assim, conhecer a localização sub-celular de um ncRNA pode fornecer pistas importantes para avançar o conhecimento sobre sua possível função biológica dentro da célula. Com esta expectativa, o plano de trabalho proposto para o presente projeto de doutoramento tinha como objetivo inicial a identificação e caracterização de ncRNAs intrônicos localizados em sub-regiões específicas do núcleo, através de hibridização fluorescente in situ, RNA-FISH (RNA-Fluorescence In Situ Hybridization). Foram selecionados vários candidatos enriquecidos no núcleo a partir das análises de microarranjo e realizada investigação de sua localização através de RNA-FISH. As tentativas de localização sub-celular de ncRNAs intrônicos através de RNA-FISH resultaram, em geral, na obtenção de candidatos enriquecidos no núcleo, mas distribuídos de maneira homogênea pelo compartimento (vide anexo 2). Não obtivemos sucesso em detectar ncRNAs intrônicos localizados em foci discretos no núcleo, apesar de termos investido bastante tempo no estudo da localização destes ncRNAs.

De maneira interessante, uma fração dos ncRNAs intrônicos, sobretudo, da classe PIN antissenso (3% do total dos RNAs enriquecidos em um dos compartimentos) apresentou transcritos enriquecidos no citoplasma da célula (vide lista de ncRNAs intrônicos enriquecidos no citoplasma, anexo 1). Cheng e colaboradores também observaram a presença de transcritos intrônicos no citoplasma (Cheng et al. 2005). Apesar de ainda escassas, já existem algumas documentações na literatura sobre funções citoplasmáticas atribuídas aos

ncRNAs. NRON (*Noncoding Repressor of NFAT*) é um ncRNA longo, integrante de um complexo citoplasmático, juntamente com a proteína IQGAP (*IQ motif containing GTPase activating protein*) e proteínas quinases inibidoras de NFAT (*Nuclear Factor of Activated T cells*), que é responsável por manter NFAT em um estado latente (Sharma et al. 2011). NFAT é um fator de transcrição, que reside no citoplasma em uma forma altamente fosforilada. Sob estímulo de cálcio, é desfosforilado e transloca-se para o núcleo, onde promove a ativação transcricional de genes alvos específicos. A depleção de NRON aumenta a desfosforilação de NFAT e seu transporte para o núcleo, indicando que este ncRNA tem um papel importante na regulação do tráfego de NFAT (Sharma et al. 2011).

Outra função de ncRNAs no citoplasma está associada ao recrutamento de proteínas promotoras de decaimento de mRNA. Gong e colaboradores identificaram vários ncRNAs que se ligam na região 3'-UTR de certos mRNAs, através de pareamento imperfeito baseado em repetições Alu, e verificaram que esta ligação resulta na degradação destes mRNAs via Staufen-1 (STAU-1) mediated messenger RNA decay (SMD) (Gong and Maquat 2011). O mecanismo de degradação de mRNAs traducionalmente ativos por SMD envolve a ligação da região 3'-UTR do mRNA alvo pela proteína STAU1, que se liga à RNAs dupla fita. Desta forma, o ncRNA, ao se ligar à região 3' UTR do mRNA, facilita a formação de um sítio de ligação para a proteína STAU1, o que desencadeia o processo de degradação do mRNA (Gong and Maquat 2011). O grupo observou também que um ncRNA, individualmente, pode induzir a degradação de vários mRNAs alvos de SMD e, que distintos ncRNAs podem induzir a degradação de um mesmo alvo. À luz destas descobertas, a observação de ncRNAs intrônicos no citoplasma torna-se bastante atraente e abre a perspectiva de que estes transcritos possam exercer funções regulatórias fora do núcleo. Apesar de uma fração pequena destes RNAs (2% do total de transcritos intrônicos enriquecidos em um dos compartimentos) ter apresentado enriquecimento citoplasmático, menos de 50% destes transcritos (TIN antissenso 17%, TIN senso 46% e PIN antissenso 32%) apresentou enriquecimento em um dos compartimentos, indicando que foram detectados tanto no núcleo quanto no citoplasma, o que é intrigante e sugere que estes RNAs embora não codifiquem para proteína, sejam exportados para o citoplasma para exercer alguma função biológica.

O fenômeno de retenção nuclear está limitado a um número restrito de RNAs e o enriquecimento de um RNA em um compartimento celular específico, núcleo/citoplasma, é principalmente o resultado do balanço entre mecanismos regulatórios mais gerais. O RNA é transcrito no núcleo, onde se acumula no estado estacionário, que é uma função complexa resultante das taxas de síntese, processamento, degradação e/ou transporte para o citoplasma (Moore 2002; Orphanides and Reinberg 2002). Para os RNAs que são transportados para o citoplasma, o RNA acumula-se no estado estacionário, de acordo com as taxas de exportação do núcleo, de tradução nos polissomos, no caso de mRNAs, e de degradação (Orphanides and Reinberg 2002). Portanto, o enriquecimento de mRNAs e ncRNAs intrônicos no núcleo relativo ao citoplasma, sugere que a taxa de síntese destes transcritos seja relativamente alta e, reciprocamente, sua taxa de liberação para o citosol seja baixa e/ou sua taxa de degradação no citoplasma seja alta. Da mesma forma, o enriquecimento de transcritos no citoplasma relativo ao núcleo sugere que a estabilidade destes transcritos seja relativamente alta e que suas taxas de transcrição sejam baixas. Um trabalho antigo, utilizando experimentos de *pulse-chase*, em linfócitos, relatou que, pelo menos, 50% dos RNAs citoplasmáticos poliadenilados têm meiavida muito curta (menos de 17 minutos), enquanto o remanescente é altamente estável (meiavida maior que 24h) (Berger and Cooper 1975). Esse encontro pode justificar, pelo menos em parte, o fato de termos detectado poucos RNAs enriquecidos no citoplasma, sugerindo que a taxa de decaimento de muitos RNAs no citoplasma é alta, fazendo com que, no estado estacionário, estejam mais acumulados no núcleo. Em conformidade com esta hipótese, ao compararmos a meia-vida dos ncRNAs intrônicos e dos mRNAs considerados enriquecidos no núcleo ou no citoplasma, verificamos que os transcritos enriquecidos no citoplasma apresentaram meia-vida significativamente maior (teste KS, p<0,05) que os transcritos enriquecidos no núcleo. Reforçando esta idéia, os mRNAs acumulados no citoplasma apresentaram enriquecimento estatisticamente significativo (teste Benjamini-Hocheberg, p<0,05) na categoria de GO associada à função molecular de constituinte estrutural de ribossomo, que por sua vez, também mostrou-se enriquecida entre os transcritos com meia-vida mais longa (vide item 5.1). De acordo com estas observações, é possível sugerir que os transcritos detectados como enriquecidos no citoplasma representem a fração de RNAs mais estáveis observados por Berger e colaboradores.

6. CONCLUSÕES

Os resultados obtidos nas análises de meia-vida, tratamento com α-amanitina para verificação de presença de RNAP II, tratamento com as enzimas TAP e 5'-Exo para verificação da presença de cap 5' e distribuição sub-celular núcleo e citoplasma estão sumarizados na tabela 12. As inferências obtidas através destes estudos estão descritas nos itens abaixo, juntamente com as conclusões quanto aos dados de expressão dos transcritos intrônicos e exônicos em células HeLa e as análises bioinformáticas realizadas.

Tabela 12: Número de transcritos intrônicos e exônicos com meia-vida maior, igual ou menor que 3h; detectados como transcritos pela RNAP II; portadores de cap 5'; enriquecidos no núcleo; no citoplasma ou igualmente distribuídos no núcleo e no citoplasma. Entre parênteses está indicada a porcentagem do total de transcritos detectados.

Classe do transcrito	Meia-vida ≥3h	Meia-vida <3h	RNAP II	Cap 5'	Núcleo	Citoplasma	Igualmente distribuídos
Exônicos	3.069	2.411	1.358	4.289	3.366	56	1.580
	(56%)	(44%)	(46%)	(59%)	(67%)	(1%)	(32%)
TIN	305	26	136	989	128	1	626
antissenso	(92%)	(8%)	(47%)	(51%)	(17%)	(0,1%)	(83%)
TIN senso	409	237	206	865	561	5	655
	(63%)	(37%)	(54%)	(30%)	(46%)	(0,4%)	(54%)
PIN	319	52	118	1.807	377	13	834
antissenso	(86%)	(14%)	(40%)	(65%)	(31%)	(1%)	(68%)

- ncRNAs intrônicos são expressos em níveis menores na célula comparados aos mRNAs codificadores de proteínas.
- Do total de ncRNAs intrônicos expressos em *loci* que também produzem mRNAs, a maior parte (62%) apresentou correlação significativa, positiva ou negativa, com a abundância do mRNA, sugerindo que ao menos uma fração destes ncRNAs exerça uma função em *cis* na modulação da transcrição ou da estabilidade do mRNA codificador de proteína.
- Uma fração (10%) de ncRNAs intrônicos não apresentou expressão concomitante com
 o mRNA codificador de proteína do *locus* correspondente, o que sugere que a
 expressão destes transcritos seja independente do mRNA codificador.

- Os ncRNAs intrônicos, TIN antissenso, TIN senso e PIN antissenso foram afetados
 pelo tratamento com o inibidor transcricional α-amanitina na mesma proporção que
 mRNAs codificadores de proteínas, o que sugere que esta classe diversificada de
 RNAs não-codificadores seja predominantemente transcrita pela RNAP II.
- Uma proporção semelhante de ncRNAs intrônicos antissenso e mRNAs foi afetada pelo tratamento com enzimas que removem o cap 5' e degradam RNAs com extremidade 5' monofosfato, indicando que estes transcritos são predominantemente dotados da modificação co-transcricional 5' cap.
- Foi detectada a presença de cap 5' em uma proporção menor de transcritos intrônicos TIN senso (30%). Este resultado indica que, embora devam existir transcritos TIN senso gerados a partir de unidades transcricionais independentes, esta classe também deve incluir *intron lariats* removidos do pré-mRNA através do *splicing*, que se acumulam na célula durante tempo suficiente que permita sua detecção.
- Análises bioinformáticas utilizando dados públicos mostraram o enriquecimento significativo de i) sítios de ligação de RNAP II, ii) CAGE tags e de iii) marcas de cromatina associadas à regiões promotoras ativas (H3K4me3) na vizinhança a montante dos transcritos intrônicos expressos em células HeLa. Os resultados obtidos indicam que esta classe de RNAs é predominantemente transcrita pela RNAP II a partir de regiões promotoras semelhantes às que controlam a transcrição de mRNAs codificadores de proteínas.
- Esta é a primeira vez que é realizada uma análise de meia-vida em larga escala de um conjunto de RNAs longos sem potencial para codificar proteínas. Os ncRNAs intrônicos mostraram ser mais estáveis (TIN antissenso=4,2h; TIN senso=3,4h; PIN antissenso=3,9h) que os RNAs codificadores de proteínas (3,1h) representados no oligoarranjo intron-exon. O acúmulo de ncRNAs intrônicos no estado estacionário

- corrobora com a idéia de que estes transcritos possam desempenhar funções relevantes na biologia da célula.
- Os transcritos TIN senso apresentaram uma maior proporção de transcritos com meiavida mais curta em relação aos demais ncRNAs intrônicos analisados (TIN e PIN antissenso), semelhante a de mRNAs codificadores. Esta classe apresenta uma maior amplitude de valores de meia-vida, o que reforça a idéia de que compreenda transcritos com diferentes naturezas: *intron lariats*, pré-mRNAs e ncRNAs transcritos de forma independente do mRNA.
- Dentre os ncRNAs intrônicos interrogados nas análises de fracionamento sub-celular, uma fração considerável, maior que 50%, mostrou-se igualmente distribuída entre o núcleo e citoplasma. Estes dados indicam que estes RNAs, embora não codifiquem para proteína, são exportados para o citoplasma, podendo exercer alguma função biológica neste compartimento.
- Quase a totalidade dos mRNAs codificadores de proteínas e ncRNAs intrônicos que apresentou enriquecimento em um dos compartimentos, núcleo ou citoplasma, mostrou-se enriquecida no núcleo (98%). Este resultado pode refletir uma alta taxa de síntese destes transcritos relativamente à taxa de transporte para o citosol, e/ou uma alta taxa de degradação dos RNAs no citoplasma. Alternativamente, é plausível que um pequeno número de mRNAs e ncRNAs intrônicos identificados como enriquecidos no núcleo sejam ativamente retidos através de mecanismos póstranscricionais (ex. edição de RNA).

7. REFERÊNCIAS

- Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A. Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kennison, K. A. Ketchum, B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L. Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacleb, M. Palazzolo, G. S. Pittman, S. Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. Saunders, F. Scheeler, H. Shen, B. C. Shue, I. Siden-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z. Y. Wang, D. A. Wassarman, G. M. Weinstock, J. Weissenbach, S. M. Williams, WoodageT, K. C. Worley, D. Wu, S. Yang, Q. A. Yao, J. Ye, R. F. Yeh, J. S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X. H. Zheng, F. N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H. O. Smith, R. A. Gibbs, E. W. Myers, G. M. Rubin and J. C. Venter (2000). The genome sequence of Drosophila melanogaster. Science 287(5461): 2185-2195.
- Aguilera, A. (2005). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. Curr Opin Cell Biol **17**(3): 242-250.
- Almeida, G. T., M. S. Amaral, F. C. Beckedorff, J. P. Kitajima, R. Demarco and S. Verjovski-Almeida (2011). Exploring the Schistosoma mansoni adult male transcriptome using RNA-seq. Exp Parasitol.
- Amaral, P. P. and J. S. Mattick (2008). Noncoding RNA in development. Mamm Genome **19**(7-8): 454-492.
- Ameur, A., A. Zaghlool, J. Halvardson, A. Wetterbom, U. Gyllensten, L. Cavelier and L. Feuk (2011). Total RNA sequencing reveals nascent transcription and widespread cotranscriptional splicing in the human brain. Nat Struct Mol Biol **18**(12): 1435-1440.
- Atkinson, S. R., S. Marguerat and J. Bahler (2011). Exploring long non-coding RNAs through sequencing. Semin Cell Dev Biol.
- Balagopal, V. and R. Parker (2009). Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. Curr Opin Cell Biol **21**(3): 403-408.
- Bartel, D. P. and C. Z. Chen (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet **5**(5): 396-400.

- Barthelson, R. A., G. M. Lambert, C. Vanier, R. M. Lynch and D. W. Galbraith (2007). Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. BMC Genomics **8**: 340.
- Bashirullah, A., R. L. Cooperstock and H. D. Lipshitz (1998). RNA localization in development. Annu Rev Biochem **67**: 335-394.
- Beltran, M., I. Puig, C. Pena, J. M. Garcia, A. B. Alvarez, R. Pena, F. Bonilla and A. G. de Herreros (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev **22**(6): 756-769.
- Benard, L. (2004). Inhibition of 5' to 3' mRNA degradation under stress conditions in Saccharomyces cerevisiae: from GCN4 to MET16. RNA **10**(3): 458-468.
- Benjamin, D. and C. Moroni (2007). mRNA stability and cancer: an emerging link? Expert Opin Biol Ther **7**(10): 1515-1529.
- Bentley, D. L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol **17**(3): 251-256.
- Berger, S. L. and H. L. Cooper (1975). Very short-lived and stable mRNAs from resting human lymphocytes. Proc Natl Acad Sci U S A **72**(10): 3873-3877.
- Bernard, D., K. V. Prasanth, V. Tripathi, S. Colasse, T. Nakamura, Z. Xuan, M. Q. Zhang, F. Sedel, L. Jourdren, F. Coulpier, A. Triller, D. L. Spector and A. Bessis (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J **29**(18): 3082-3093.
- Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao and S. N. Cohen (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A **99**(15): 9697-9702.
- Bertone, P., V. Stolc, T. E. Royce, J. S. Rozowsky, A. E. Urban, X. Zhu, J. L. Rinn, W. Tongprasit, M. Samanta, S. Weissman, M. Gerstein and M. Snyder (2004). Global identification of human transcribed sequences with genome tiling arrays. Science **306**(5705): 2242-2246.
- Besse, F. and A. Ephrussi (2008). Translational control of localized mRNAs: restricting protein synthesis in space and time. Nat Rev Mol Cell Biol **9**(12): 971-980.
- Birney, E., J. A. Stamatoyannopoulos, A. Dutta, R. Guigo, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, M. S. Kuehn, C. M. Taylor, S. Neph, C. M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J. A. Greenbaum, R. M. Andrews, P. Flicek, P. J. Boyle, H. Cao, N. P. Carter, G. K. Clelland, S. Davis, N. Day, P. Dhami, S. C. Dillon, M. O. Dorschner, H. Fiegler, P. G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K. D. James, B. E. Johnson, E. M. Johnson, T. T. Frum, E. R. Rosenzweig, N. Karnani, K. Lee, G. C. Lefebvre, P. A. Navas, F. Neri, S. C. Parker, P. J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F. S. Collins, J. Dekker, J. D. Lieb, T. D. Tullius, G. E. Crawford, S. Sunyaev, W. S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I. L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H. A. Hirsch, E. A. Sekinger, J. Lagarde, J. F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J. S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M. C. Dickson, D. J. Thomas, M. T. Weirauch, J. Gilbert, J. Drenkow, I. Bell, X. Zhao, K. G. Srinivasan, W. K. Sung, H. S. Ooi, K. P. Chiu, S. Foissac, T. Alioto, M. Brent, L. Pachter, M. L. Tress, A. Valencia, S. W. Choo, C. Y. Choo, C. Ucla, C. Manzano, C. Wyss, E. Cheung, T. G. Clark, J. B. Brown, M. Ganesh, S. Patel, H. Tammana, J. Chrast, C. N. Henrichsen, C. Kai, J. Kawai, U. Nagalakshmi, J. Wu, Z. Lian, J. Lian, P. Newburger, X. Zhang, P.

- Bickel, J. S. Mattick, P. Carninci, Y. Hayashizaki, S. Weissman, T. Hubbard, R. M. Myers, J. Rogers, P. F. Stadler, T. M. Lowe, C. L. Wei, Y. Ruan, K. Struhl, M. Gerstein, S. E. Antonarakis, Y. Fu, E. D. Green, U. Karaoz, A. Siepel, J. Taylor, L. A. Liefer, K. A. Wetterstrand, P. J. Good, E. A. Feingold, M. S. Guyer, G. M. Cooper, G. Asimenos, C. N. Dewey, M. Hou, S. Nikolaev, J. I. Montoya-Burgos, A. Loytynoja, S. Whelan, F. Pardi, T. Massingham, H. Huang, N. R. Zhang, I. Holmes, J. C. Mullikin, A. Ureta-Vidal, B. Paten, M. Seringhaus, D. Church, K. Rosenbloom, W. J. Kent, E. A. Stone, S. Batzoglou, N. Goldman, R. C. Hardison, D. Haussler, W. Miller, A. Sidow, N. D. Trinklein, Z. D. Zhang, L. Barrera, R. Stuart, D. C. King, A. Ameur, S. Enroth, M. C. Bieda, J. Kim, A. A. Bhinge, N. Jiang, J. Liu, F. Yao, V. B. Vega, C. W. Lee, P. Ng, A. Yang, Z. Moqtaderi, Z. Zhu, X. Xu, S. Squazzo, M. J. Oberley, D. Inman, M. A. Singer, T. A. Richmond, K. J. Munn, A. Rada-Iglesias, O. Wallerman, J. Komorowski, J. C. Fowler, P. Couttet, A. W. Bruce, O. M. Dovey, P. D. Ellis, C. F. Langford, D. A. Nix, G. Euskirchen, S. Hartman, A. E. Urban, P. Kraus, S. Van Calcar, N. Heintzman, T. H. Kim, K. Wang, C. Qu, G. Hon, R. Luna, C. K. Glass, M. G. Rosenfeld, S. F. Aldred, S. J. Cooper, A. Halees, J. M. Lin, H. P. Shulha, M. Xu, J. N. Haidar, Y. Yu, V. R. Iyer, R. D. Green, C. Wadelius, P. J. Farnham, B. Ren, R. A. Harte, A. S. Hinrichs, H. Trumbower, H. Clawson, J. Hillman-Jackson, A. S. Zweig, K. Smith, A. Thakkapallayil, G. Barber, R. M. Kuhn, D. Karolchik, L. Armengol, C. P. Bird, P. I. de Bakker, A. D. Kern, N. Lopez-Bigas, J. D. Martin, B. E. Stranger, A. Woodroffe, E. Davydov, A. Dimas, E. Eyras, I. B. Hallgrimsdottir, J. Huppert, M. C. Zody, G. R. Abecasis, X. Estivill, G. G. Bouffard, X. Guan, N. F. Hansen, J. R. Idol, V. V. Maduro, B. Maskeri, J. C. McDowell, M. Park, P. J. Thomas, A. C. Young, R. W. Blakesley, D. M. Muzny, E. Sodergren, D. A. Wheeler, K. C. Worley, H. Jiang, G. M. Weinstock, R. A. Gibbs, T. Graves, R. Fulton, E. R. Mardis, R. K. Wilson, M. Clamp, J. Cuff, S. Gnerre, D. B. Jaffe, J. L. Chang, K. Lindblad-Toh, E. S. Lander, M. Koriabine, M. Nefedov, K. Osoegawa, Y. Yoshinaga, B. Zhu and P. J. de Jong (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447**(7146): 799-816.
- Blobel, G. and B. Dobberstein (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J Cell Biol **67**(3): 835-851.
- Blower, M. D., M. Nachury, R. Heald and K. Weis (2005). A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. Cell **121**(2): 223-234.
- Bolognani, F. and N. I. Perrone-Bizzozero (2008). RNA-protein interactions and control of mRNA stability in neurons. J Neurosci Res **86**(3): 481-489.
- Bolstad, B. M., R. A. Irizarry, M. Astrand and T. P. Speed (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics **19**(2): 185-193.
- Bond, C. S. and A. H. Fox (2009). Paraspeckles: nuclear bodies built on long noncoding RNA. J Cell Biol **186**(5): 637-644.
- Braconi, C., N. Valeri, T. Kogure, P. Gasparini, N. Huang, G. J. Nuovo, L. Terracciano, C. M. Croce and T. Patel (2011). Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. Proc Natl Acad Sci U S A **108**(2): 786-791.
- Brito, G. C., A. A. Fachel, A. L. Vettore, G. M. Vignal, E. R. Gimba, F. S. Campos, M. A. Barcinski, S. Verjovski-Almeida and E. M. Reis (2008). Identification of protein-coding and intronic noncoding RNAs down-regulated in clear cell renal carcinoma. Mol Carcinog 47(10): 757-767.

- Brodsky, A. S., C. A. Meyer, I. A. Swinburne, G. Hall, B. J. Keenan, X. S. Liu, E. A. Fox and P. A. Silver (2005). Genomic mapping of RNA polymerase II reveals sites of cotranscriptional regulation in human cells. Genome Biol **6**(8): R64.
- Bushnell, D. A., P. Cramer and R. D. Kornberg (2002). Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. Proc Natl Acad Sci U S A **99**(3): 1218-1222.
- Cabili, M. N., C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev and J. L. Rinn (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev **25**(18): 1915-1927.
- Camblong, J., N. Iglesias, C. Fickentscher, G. Dieppois and F. Stutz (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell **131**(4): 706-717.
- Carmody, S. R. and S. R. Wente (2009). mRNA nuclear export at a glance. J Cell Sci 122(Pt 12): 1933-1937.
- Carninci, P. and Y. Hayashizaki (1999). High-efficiency full-length cDNA cloning. Methods Enzymol **303**: 19-44.
- Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, R. Kodzius, K. Shimokawa, V. B. Bajic, S. E. Brenner, S. Batalov, A. R. Forrest, M. Zavolan, M. J. Davis, L. G. Wilming, V. Aidinis, J. E. Allen, A. Ambesi-Impiombato, R. Apweiler, R. N. Aturaliya, T. L. Bailey, M. Bansal, L. Baxter, K. W. Beisel, T. Bersano, H. Bono, A. M. Chalk, K. P. Chiu, V. Choudhary, A. Christoffels, D. R. Clutterbuck, M. L. Crowe, E. Dalla, B. P. Dalrymple, B. de Bono, G. Della Gatta, D. di Bernardo, T. Down, P. Engstrom, M. Fagiolini, G. Faulkner, C. F. Fletcher, T. Fukushima, M. Furuno, S. Futaki, M. Gariboldi, P. Georgii-Hemming, T. R. Gingeras, T. Gojobori, R. E. Green, S. Gustincich, M. Harbers, Y. Hayashi, T. K. Hensch, N. Hirokawa, D. Hill, L. Huminiecki, M. Iacono, K. Ikeo, A. Iwama, T. Ishikawa, M. Jakt, A. Kanapin, M. Katoh, Y. Kawasawa, J. Kelso, H. Kitamura, H. Kitano, G. Kollias, S. P. Krishnan, A. Kruger, S. K. Kummerfeld, I. V. Kurochkin, L. F. Lareau, D. Lazarevic, L. Lipovich, J. Liu, S. Liuni, S. McWilliam, M. Madan Babu, M. Madera, L. Marchionni, H. Matsuda, S. Matsuzawa, H. Miki, F. Mignone, S. Miyake, K. Morris, S. Mottagui-Tabar, N. Mulder, N. Nakano, H. Nakauchi, P. Ng, R. Nilsson, S. Nishiguchi, S. Nishikawa, F. Nori, O. Ohara, Y. Okazaki, V. Orlando, K. C. Pang, W. J. Pavan, G. Pavesi, G. Pesole, N. Petrovsky, S. Piazza, J. Reed, J. F. Reid, B. Z. Ring, M. Ringwald, B. Rost, Y. Ruan, S. L. Salzberg, A. Sandelin, C. Schneider, C. Schonbach, K. Sekiguchi, C. A. Semple, S. Seno, L. Sessa, Y. Sheng, Y. Shibata, H. Shimada, K. Shimada, D. Silva, B. Sinclair, S. Sperling, E. Stupka, K. Sugiura, R. Sultana, Y. Takenaka, K. Taki, K. Tammoja, S. L. Tan, S. Tang, M. S. Taylor, J. Tegner, S. A. Teichmann, H. R. Ueda, E. van Nimwegen, R. Verardo, C. L. Wei, K. Yagi, H. Yamanishi, E. Zabarovsky, S. Zhu, A. Zimmer, W. Hide, C. Bult, S. M. Grimmond, R. D. Teasdale, E. T. Liu, V. Brusic, J. Quackenbush, C. Wahlestedt, J. S. Mattick, D. A. Hume, C. Kai, D. Sasaki, Y. Tomaru, S. Fukuda, M. Kanamori-Katayama, M. Suzuki, J. Aoki, T. Arakawa, J. Iida, K. Imamura, M. Itoh, T. Kato, H. Kawaji, N. Kawagashira, T. Kawashima, M. Kojima, S. Kondo, H. Konno, K. Nakano, N. Ninomiya, T. Nishio, M. Okada, C. Plessy, K. Shibata, T. Shiraki, S. Suzuki, M. Tagami, K. Waki, A. Watahiki, Y. Okamura-Oho, H. Suzuki, J. Kawai and Y. Hayashizaki (2005). The transcriptional landscape of the mammalian genome. Science **309**(5740): 1559-1563.

- Carter, K. C., K. L. Taneja and J. B. Lawrence (1991). Discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus. J Cell Biol **115**(5): 1191-1202.
- Carthew, R. W. and E. J. Sontheimer (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell **136**(4): 642-655.
- Cawley, S., S. Bekiranov, H. H. Ng, P. Kapranov, E. A. Sekinger, D. Kampa, A. Piccolboni, V. Sementchenko, J. Cheng, A. J. Williams, R. Wheeler, B. Wong, J. Drenkow, M. Yamanaka, S. Patel, S. Brubaker, H. Tammana, G. Helt, K. Struhl and T. R. Gingeras (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell **116**(4): 499-509.
- Cesana, M., D. Cacchiarelli, I. Legnini, T. Santini, O. Sthandier, M. Chinappi, A. Tramontano and I. Bozzoni (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell **147**(2): 358-369.
- Champelovier, P., V. Pautre, M. Elatifi, I. Dupre, B. Rostaing, A. Michoud, F. Berger and D. Seigneurin (2006). Resistance to phorbol ester-induced differentiation in human myeloid leukemia cells: a hypothetic role for the mRNA stabilization process. Leuk Res **30**(11): 1407-1416.
- Cheloufi, S., C. O. Dos Santos, M. M. Chong and G. J. Hannon (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature **465**(7298): 584-589.
- Chen, J., M. Sun, S. Lee, G. Zhou, J. D. Rowley and S. M. Wang (2002). Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. Proc Natl Acad Sci U S A **99**(19): 12257-12262.
- Chen, L. L. and G. G. Carmichael (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell **35**(4): 467-478.
- Chen, L. L. and G. G. Carmichael (2010). Decoding the function of nuclear long non-coding RNAs. Curr Opin Cell Biol **22**(3): 357-364.
- Cheng, H., K. Dufu, C. S. Lee, J. L. Hsu, A. Dias and R. Reed (2006). Human mRNA export machinery recruited to the 5' end of mRNA. Cell **127**(7): 1389-1400.
- Cheng, J., P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, J. Long, D. Stern, H. Tammana, G. Helt, V. Sementchenko, A. Piccolboni, S. Bekiranov, D. K. Bailey, M. Ganesh, S. Ghosh, I. Bell, D. S. Gerhard and T. R. Gingeras (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308(5725): 1149-1154.
- Chow, J. and E. Heard (2009). X inactivation and the complexities of silencing a sex chromosome. Curr Opin Cell Biol **21**(3): 359-366.
- Clark, M. B. and J. S. Mattick (2011). Long noncoding RNAs in cell biology. Semin Cell Dev Biol **22**(4): 366-376.
- Clement, J. Q., S. Maiti and M. F. Wilkinson (2001). Localization and stability of introns spliced from the Pem homeobox gene. J Biol Chem **276**(20): 16919-16930.
- Clement, J. Q., L. Qian, N. Kaplinsky and M. F. Wilkinson (1999). The stability and fate of a spliced intron from vertebrate cells. RNA 5(2): 206-220.
- Clemson, C. M., J. N. Hutchinson, S. A. Sara, A. W. Ensminger, A. H. Fox, A. Chess and J. B. Lawrence (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell **33**(6): 717-726.
- Consortium, I. H. G. S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science **282**(5396): 2012-2018.
- Cougot, N., E. van Dijk, S. Babajko and B. Seraphin (2004). 'Cap-tabolism'. Trends Biochem Sci **29**(8): 436-444.

- Crick, F. (1970). Central dogma of molecular biology. Nature 227(5258): 561-563.
- Czaplinski, K. and R. H. Singer (2006). Pathways for mRNA localization in the cytoplasm. Trends Biochem Sci **31**(12): 687-693.
- David, L., W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis and L. M. Steinmetz (2006). A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci U S A **103**(14): 5320-5325.
- Davis, C. A. and M. Ares, Jr. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A **103**(9): 3262-3267.
- Dinger, M. E., P. P. Amaral, T. R. Mercer and J. S. Mattick (2009). Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. Brief Funct Genomic Proteomic 8(6): 407-423.
- Dinger, M. E., P. P. Amaral, T. R. Mercer, K. C. Pang, S. J. Bruce, B. B. Gardiner, M. E. Askarian-Amiri, K. Ru, G. Solda, C. Simons, S. M. Sunkin, M. L. Crowe, S. M. Grimmond, A. C. Perkins and J. S. Mattick (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res **18**(9): 1433-1445.
- Ebralidze, A. K., F. C. Guibal, U. Steidl, P. Zhang, S. Lee, B. Bartholdy, M. A. Jorda, V. Petkova, F. Rosenbauer, G. Huang, T. Dayaram, J. Klupp, K. B. O'Brien, B. Will, M. Hoogenkamp, K. L. Borden, C. Bonifer and D. G. Tenen (2008). PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. Genes Dev 22(15): 2085-2092.
- Eisenberg, E., K. Adamsky, L. Cohen, N. Amariglio, A. Hirshberg, G. Rechavi and E. Y. Levanon (2005). Identification of RNA editing sites in the SNP database. Nucleic Acids Res **33**(14): 4612-4617.
- Erdmann, V. A., M. Z. Barciszewska, A. Hochberg, N. de Groot and J. Barciszewski (2001). Regulatory RNAs. Cell Mol Life Sci **58**(7): 960-977.
- Erkmann, J. A., R. Sanchez, N. Treichel, W. F. Marzluff and U. Kutay (2005). Nuclear export of metazoan replication-dependent histone mRNAs is dependent on RNA length and is mediated by TAP. RNA **11**(1): 45-58.
- Farajollahi, S. and S. Maas (2010). Molecular diversity through RNA editing: a balancing act. Trends Genet **26**(5): 221-230.
- Fejes-Toth, K., V. Sotirova, R. Sachidanandam, G. Assaf and J. Hannon G (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature **457**(7232): 1028-1032.
- Feng, J., C. Bi, B. S. Clark, R. Mady, P. Shah and J. D. Kohtz (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes Dev **20**(11): 1470-1484.
- Flynn, R. A., A. E. Almada, J. R. Zamudio and P. A. Sharp (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc Natl Acad Sci U S A **108**(26): 10460-10465.
- Franco-Zorrilla, J. M., A. Valli, M. Todesco, I. Mateos, M. I. Puga, I. Rubio-Somoza, A. Leyva, D. Weigel, J. A. Garcia and J. Paz-Ares (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet **39**(8): 1033-1037.
- Friedel, C. C., L. Dolken, Z. Ruzsics, U. H. Koszinowski and R. Zimmer (2009). Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. Nucleic Acids Res **37**(17): e115.
- Furuno, M., K. C. Pang, N. Ninomiya, S. Fukuda, M. C. Frith, C. Bult, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, J. S. Mattick and H. Suzuki (2006). Clusters of internally primed transcripts reveal novel long noncoding RNAs. PLoS Genet **2**(4): e37.

- Geisler, S., L. Lojek, A. M. Khalil, K. E. Baker and J. Coller (2012). Decapping of Long Noncoding RNAs Regulates Inducible Genes. Mol Cell.
- Gibb, E. A., C. J. Brown and W. L. Lam (2011). The functional role of long non-coding RNA in human carcinomas. Mol Cancer **10**: 38.
- Gingeras, T. R. (2007). Origin of phenotypes: genes and transcripts. Genome Res **17**(6): 682-690.
- Giorgi, C. and M. J. Moore (2007). The nuclear nurture and cytoplasmic nature of localized mRNPs. Semin Cell Dev Biol **18**(2): 186-193.
- Girard, A., R. Sachidanandam, G. J. Hannon and M. A. Carmell (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature **442**(7099): 199-202.
- Gondran, P., F. Amiot, D. Weil and F. Dautry (1999). Accumulation of mature mRNA in the nuclear fraction of mammalian cells. FEBS Lett **458**(3): 324-328.
- Gong, C. and L. E. Maquat (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature **470**(7333): 284-288.
- Gong, X. Q., Y. A. Nedialkov and Z. F. Burton (2004). Alpha-amanitin blocks translocation by human RNA polymerase II. J Biol Chem **279**(26): 27422-27427.
- Gowrishankar, G., R. Winzen, F. Bollig, B. Ghebremedhin, N. Redich, B. Ritter, K. Resch, M. Kracht and H. Holtmann (2005). Inhibition of mRNA deadenylation and degradation by ultraviolet light. Biol Chem **386**(12): 1287-1293.
- Greatrix, B. W. and H. J. van Vuuren (2006). Expression of the HXT13, HXT15 and HXT17 genes in Saccharomyces cerevisiae and stabilization of the HXT1 gene transcript by sugar-induced osmotic stress. Curr Genet **49**(4): 205-217.
- Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch and R. A. Young (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell **130**(1): 77-88.
- Gupta, R. A., N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M. C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar and H. Y. Chang (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464(7291): 1071-1076.
- Guttman, M., I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn and E. S. Lander (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458(7235): 223-227.
- Guttman, M., M. Garber, J. Z. Levin, J. Donaghey, J. Robinson, X. Adiconis, L. Fan, M. J. Koziol, A. Gnirke, C. Nusbaum, J. L. Rinn, E. S. Lander and A. Regev (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol **28**(5): 503-510.
- Hasegawa, Y., N. Brockdorff, S. Kawano, K. Tsutui and S. Nakagawa (2010). The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. Dev Cell **19**(3): 469-476.
- He, H., J. Wang, T. Liu, X. S. Liu, T. Li, Y. Wang, Z. Qian, H. Zheng, X. Zhu, T. Wu, B. Shi, W. Deng, W. Zhou, G. Skogerbo and R. Chen (2007). Mapping the C. elegans noncoding transcriptome with a whole-genome tiling microarray. Genome Res **17**(10): 1471-1477.
- He, Y., B. Vogelstein, V. E. Velculescu, N. Papadopoulos and K. W. Kinzler (2008). The antisense transcriptomes of human cells. Science **322**(5909): 1855-1857.

- Heier, C. R., R. G. Gogliotti and C. J. DiDonato (2007). SMN transcript stability: could modulation of messenger RNA degradation provide a novel therapy for spinal muscular atrophy? J Child Neurol **22**(8): 1013-1018.
- Hellwig, S. and B. L. Bass (2008). A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes. Proc Natl Acad Sci U S A **105**(35): 12897-12902.
- Heo, J. B. and S. Sung (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science **331**(6013): 76-79.
- Herman, R. C., J. G. Williams and S. Penman (1976). Message and non-message sequences adjacent to poly(A) in steady state heterogeneous nuclear RNA of HeLa cells. Cell **7**(3): 429-437.
- Herruer, M. H., W. H. Mager, H. A. Raue, P. Vreken, E. Wilms and R. J. Planta (1988). Mild temperature shock affects transcription of yeast ribosomal protein genes as well as the stability of their mRNAs. Nucleic Acids Res **16**(16): 7917-7929.
- Hirota, K., T. Miyoshi, K. Kugou, C. S. Hoffman, T. Shibata and K. Ohta (2008). Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature **456**(7218): 130-134.
- Hirota, K. and K. Ohta (2009). Cascade transcription of mRNA-type long non-coding RNAs (mlonRNAs) and local chromatin remodeling. Epigenetics **4**(1): 5-7.
- Hollams, E. M., K. M. Giles, A. M. Thomson and P. J. Leedman (2002). MRNA stability and the control of gene expression: implications for human disease. Neurochem Res **27**(10): 957-980.
- Hopper, A. K. (2006). Cellular dynamics of small RNAs. Crit Rev Biochem Mol Biol **41**(1): 3-19.
- Houseley, J., J. LaCava and D. Tollervey (2006). RNA-quality control by the exosome. Nat Rev Mol Cell Biol **7**(7): 529-539.
- Huang da, W., B. T. Sherman and R. A. Lempicki (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res **37**(1): 1-13.
- Huang da, W., B. T. Sherman and R. A. Lempicki (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **4**(1): 44-57.
- Huang, S., T. J. Deerinck, M. H. Ellisman and D. L. Spector (1994). In vivo analysis of the stability and transport of nuclear poly(A)+ RNA. J Cell Biol **126**(4): 877-899.
- Huang, Y. and G. G. Carmichael (1997). The mouse histone H2a gene contains a small element that facilitates cytoplasmic accumulation of intronless gene transcripts and of unspliced HIV-1-related mRNAs. Proc Natl Acad Sci U S A **94**(19): 10104-10109.
- Huarte, M., M. Guttman, D. Feldser, M. Garber, M. J. Koziol, D. Kenzelmann-Broz, A. M. Khalil, O. Zuk, I. Amit, M. Rabani, L. D. Attardi, A. Regev, E. S. Lander, T. Jacks and J. L. Rinn (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell **142**(3): 409-419.
- Hughes, T. R., M. Mao, A. R. Jones, J. Burchard, M. J. Marton, K. W. Shannon, S. M. Lefkowitz, M. Ziman, J. M. Schelter, M. R. Meyer, S. Kobayashi, C. Davis, H. Dai, Y. D. He, S. B. Stephaniants, G. Cavet, W. L. Walker, A. West, E. Coffey, D. D. Shoemaker, R. Stoughton, A. P. Blanchard, S. H. Friend and P. S. Linsley (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19(4): 342-347.
- Hung, T., Y. Wang, M. F. Lin, A. K. Koegel, Y. Kotake, G. D. Grant, H. M. Horlings, N. Shah, C. Umbricht, P. Wang, B. Kong, A. Langerod, A. L. Borresen-Dale, S. K. Kim, M. van de Vijver, S. Sukumar, M. L. Whitfield, M. Kellis, Y. Xiong, D. J. Wong and

- H. Y. Chang (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet **43**(7): 621-629.
- Hutchinson, J. N., A. W. Ensminger, C. M. Clemson, C. R. Lynch, J. B. Lawrence and A. Chess (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics 8: 39.
- Imanishi, T., T. Itoh, Y. Suzuki, C. O'Donovan, S. Fukuchi, K. O. Koyanagi, R. A. Barrero, T. Tamura, Y. Yamaguchi-Kabata, M. Tanino, K. Yura, S. Miyazaki, K. Ikeo, K. Homma, A. Kasprzyk, T. Nishikawa, M. Hirakawa, J. Thierry-Mieg, D. Thierry-Mieg, J. Ashurst, L. Jia, M. Nakao, M. A. Thomas, N. Mulder, Y. Karavidopoulou, L. Jin, S. Kim, T. Yasuda, B. Lenhard, E. Eveno, C. Yamasaki, J. Takeda, C. Gough, P. Hilton, Y. Fujii, H. Sakai, S. Tanaka, C. Amid, M. Bellgard, F. Bonaldo Mde, H. Bono, S. K. Bromberg, A. J. Brookes, E. Bruford, P. Carninci, C. Chelala, C. Couillault, S. J. de Souza, M. A. Debily, M. D. Devignes, I. Dubchak, T. Endo, A. Estreicher, E. Eyras, K. Fukami-Kobayashi, G. R. Gopinath, E. Graudens, Y. Hahn, M. Han, Z. G. Han, K. Hanada, H. Hanaoka, E. Harada, K. Hashimoto, U. Hinz, M. Hirai, T. Hishiki, I. Hopkinson, S. Imbeaud, H. Inoko, A. Kanapin, Y. Kaneko, T. Kasukawa, J. Kelso, P. Kersey, R. Kikuno, K. Kimura, B. Korn, V. Kuryshev, I. Makalowska, T. Makino, S. Mano, R. Mariage-Samson, J. Mashima, H. Matsuda, H. W. Mewes, S. Minoshima, K. Nagai, H. Nagasaki, N. Nagata, R. Nigam, O. Ogasawara, O. Ohara, M. Ohtsubo, N. Okada, T. Okido, S. Oota, M. Ota, T. Ota, T. Otsuki, D. Piatier-Tonneau, A. Poustka, S. X. Ren, N. Saitou, K. Sakai, S. Sakamoto, R. Sakate, I. Schupp, F. Servant, S. Sherry, R. Shiba, N. Shimizu, M. Shimoyama, A. J. Simpson, B. Soares, C. Steward, M. Suwa, M. Suzuki, A. Takahashi, G. Tamiya, H. Tanaka, T. Taylor, J. D. Terwilliger, P. Unneberg, V. Veeramachaneni, S. Watanabe, L. Wilming, N. Yasuda, H. S. Yoo, M. Stodolsky, W. Makalowski, M. Go, K. Nakai, T. Takagi, M. Kanehisa, Y. Sakaki, J. Quackenbush, Y. Okazaki, Y. Hayashizaki, W. Hide, R. Chakraborty, K. Nishikawa, H. Sugawara, Y. Tateno, Z. Chen, M. Oishi, P. Tonellato, R. Apweiler, K. Okubo, L. Wagner, S. Wiemann, R. L. Strausberg, T. Isogai, C. Auffray, N. Nomura, T. Gojobori and S. Sugano (2004). Integrative annotation of 21,037 human genes validated by full-length cDNA clones. PLoS Biol 2(6): e162.
- Ingolia, N. T., L. F. Lareau and J. S. Weissman (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell **147**(4): 789-802.
- Inui, M., G. Martello and S. Piccolo (2010). MicroRNA control of signal transduction. Nat Rev Mol Cell Biol **11**(4): 252-263.
- Ishii, N., K. Ozaki, H. Sato, H. Mizuno, S. Saito, A. Takahashi, Y. Miyamoto, S. Ikegawa, N. Kamatani, M. Hori, Y. Nakamura and T. Tanaka (2006). Identification of a novel noncoding RNA, MIAT, that confers risk of myocardial infarction. J Hum Genet **51**(12): 1087-1099.
- Jacob, F. and J. Monod (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3: 318-356.
- Jacob, S. T., E. M. Sajdel and H. N. Munro (1970). Specific action of alpha-amanitin on mammalian RNA polymerase protein. Nature **225**(5227): 60-62.
- Jinno, Y., G. T. Merlino and I. Pastan (1988). A novel effect of EGF on mRNA stability. Nucleic Acids Res **16**(11): 4957-4966.
- Jishage, M., S. Malik, U. Wagner, B. Uberheide, Y. Ishihama, X. Hu, B. T. Chait, A. Gnatt, B. Ren and R. G. Roeder (2012). Transcriptional Regulation by Pol II(G) Involving Mediator and Competitive Interactions of Gdown1 and TFIIF with Pol II. Mol Cell **45**(1): 51-63.

- Jona, G., M. Choder and O. Gileadi (2000). Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast. Biochim Biophys Acta **1491**(1-3): 37-48.
- Jongeneel, C. V., M. Delorenzi, C. Iseli, D. Zhou, C. D. Haudenschild, I. Khrebtukova, D. Kuznetsov, B. J. Stevenson, R. L. Strausberg, A. J. Simpson and T. J. Vasicek (2005). An atlas of human gene expression from massively parallel signature sequencing (MPSS). Genome Res **15**(7): 1007-1014.
- Kampa, D., J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, J. Drenkow, A. Piccolboni, S. Bekiranov, G. Helt, H. Tammana and T. R. Gingeras (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res 14(3): 331-342.
- Kapranov, P. (2009). From transcription start site to cell biology. Genome Biol 10(4): 217.
- Kapranov, P., S. E. Cawley, J. Drenkow, S. Bekiranov, R. L. Strausberg, S. P. Fodor and T. R. Gingeras (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science 296(5569): 916-919.
- Kapranov, P., J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermuller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana and T. R. Gingeras (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316(5830): 1484-1488.
- Kapranov, P., G. St Laurent, T. Raz, F. Ozsolak, C. P. Reynolds, P. H. Sorensen, G. Reaman, P. Milos, R. J. Arceci, J. F. Thompson and T. J. Triche (2010). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. BMC Biol 8: 149.
- Kapranov, P., A. T. Willingham and T. R. Gingeras (2007). Genome-wide transcription and the implications for genomic organization. Nat Rev Genet **8**(6): 413-423.
- Kay, R. A., I. R. Ellis, S. J. Jones, S. Perrier, M. M. Florence, A. M. Schor and S. L. Schor (2005). The expression of migration stimulating factor, a potent oncofetal cytokine, is uniquely controlled by 3'-untranslated region-dependent nuclear sequestration of its precursor messenger RNA. Cancer Res 65(23): 10742-10749.
- Kedinger, C., M. Gniazdowski, J. L. Mandel, Jr., F. Gissinger and P. Chambon (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. Biochem Biophys Res Commun **38**(1): 165-171.
- Khalil, A. M., M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B. E. Bernstein, A. van Oudenaarden, A. Regev, E. S. Lander and J. L. Rinn (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106(28): 11667-11672.
- Kim, T. H., L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green and B. Ren (2005). A high-resolution map of active promoters in the human genome. Nature **436**(7052): 876-880.
- Kim, T. K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman and M. E. Greenberg (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature 465(7295): 182-187.
- Kloc, M., K. Wilk, D. Vargas, Y. Shirato, S. Bilinski and L. D. Etkin (2005). Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes. Development **132**(15): 3445-3457.

- Kodzius, R., M. Kojima, H. Nishiyori, M. Nakamura, S. Fukuda, M. Tagami, D. Sasaki, K. Imamura, C. Kai, M. Harbers, Y. Hayashizaki and P. Carninci (2006). CAGE: cap analysis of gene expression. Nat Methods **3**(3): 211-222.
- Kong, L., Y. Zhang, Z. Q. Ye, X. Q. Liu, S. Q. Zhao, L. Wei and G. Gao (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 35(Web Server issue): W345-349.
- Krol, J., I. Loedige and W. Filipowicz (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet **11**(9): 597-610.
- Krystal, G. W., B. C. Armstrong and J. F. Battey (1990). N-myc mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. Mol Cell Biol **10**(8): 4180-4191.
- Kuai, L., B. Das and F. Sherman (2005). A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 102(39): 13962-13967.
- Lagos-Quintana, M., R. Rauhut, W. Lendeckel and T. Tuschl (2001). Identification of novel genes coding for small expressed RNAs. Science **294**(5543): 853-858.
- Lamond, A. I. and D. L. Spector (2003). Nuclear speckles: a model for nuclear organelles. Nat Rev Mol Cell Biol 4(8): 605-612.
- Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M.

- Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen (2001). Initial sequencing and analysis of the human genome. Nature **409**(6822): 860-921.
- Lardenois, A., Y. Liu, T. Walther, F. Chalmel, B. Evrard, M. Granovskaia, A. Chu, R. W. Davis, L. M. Steinmetz and M. Primig (2011). Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc Natl Acad Sci U S A **108**(3): 1058-1063.
- Le Hir, H., D. Gatfield, E. Izaurralde and M. J. Moore (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J **20**(17): 4987-4997.
- Lebedeva, S., M. Jens, K. Theil, B. Schwanhausser, M. Selbach, M. Landthaler and N. Rajewsky (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. Mol Cell **43**(3): 340-352.
- Lecuyer, E., H. Yoshida, N. Parthasarathy, C. Alm, T. Babak, T. Cerovina, T. R. Hughes, P. Tomancak and H. M. Krause (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell **131**(1): 174-187.
- Lee, R. C., R. L. Feinbaum and V. Ambros (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75**(5): 843-854.
- Lee, Y., M. Kim, J. Han, K. H. Yeom, S. Lee, S. H. Baek and V. N. Kim (2004). MicroRNA genes are transcribed by RNA polymerase II. EMBO J **23**(20): 4051-4060.
- Levanon, E. Y., E. Eisenberg, R. Yelin, S. Nemzer, M. Hallegger, R. Shemesh, Z. Y. Fligelman, A. Shoshan, S. R. Pollock, D. Sztybel, M. Olshansky, G. Rechavi and M. F. Jantsch (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol **22**(8): 1001-1005.
- Levy, A. P., N. S. Levy and M. A. Goldberg (1996). Post-transcriptional regulation of vascular endothelial growth factor by hypoxia. J Biol Chem **271**(5): 2746-2753.
- Lin, R., S. Maeda, C. Liu, M. Karin and T. S. Edgington (2007). A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. Oncogene **26**(6): 851-858.
- Lindell, T. J., F. Weinberg, P. W. Morris, R. G. Roeder and W. J. Rutter (1970). Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. Science 170(3956): 447-449
- Listerman, I., A. S. Bledau, I. Grishina and K. M. Neugebauer (2007). Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III. PLoS Genet **3**(11): e212.
- Louro, R., H. I. Nakaya, P. P. Amaral, F. Festa, M. C. Sogayar, A. M. da Silva, S. Verjovski-Almeida and E. M. Reis (2007). Androgen responsive intronic non-coding RNAs. BMC Biol 5: 4.
- Luna, R., H. Gaillard, C. Gonzalez-Aguilera and A. Aguilera (2008). Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. Chromosoma **117**(4): 319-331.
- Mancini-Dinardo, D., S. J. Steele, J. M. Levorse, R. S. Ingram and S. M. Tilghman (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev **20**(10): 1268-1282.

- Martianov, I., A. Ramadass, A. Serra Barros, N. Chow and A. Akoulitchev (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature **445**(7128): 666-670.
- Martin, K. C. and A. Ephrussi (2009). mRNA localization: gene expression in the spatial dimension. Cell **136**(4): 719-730.
- Martinez, N. J. and A. J. Walhout (2009). The interplay between transcription factors and microRNAs in genome-scale regulatory networks. Bioessays **31**(4): 435-445.
- Mattick, J. S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays **25**(10): 930-939.
- Mattick, J. S. (2011). The central role of RNA in human development and cognition. FEBS Lett **585**(11): 1600-1616.
- Mattick, J. S. and I. V. Makunin (2005). Small regulatory RNAs in mammals. Hum Mol Genet **14 Spec No 1**: R121-132.
- Meister, G., M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng and T. Tuschl (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell **15**(2): 185-197.
- Mello, C. C. and D. Conte, Jr. (2004). Revealing the world of RNA interference. Nature **431**(7006): 338-342.
- Mercer, T. R., M. E. Dinger and J. S. Mattick (2009). Long non-coding RNAs: insights into functions. Nat Rev Genet **10**(3): 155-159.
- Mercer, T. R., M. E. Dinger, S. M. Sunkin, M. F. Mehler and J. S. Mattick (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A 105(2): 716-721.
- Mercer, T. R., D. J. Gerhardt, M. E. Dinger, J. Crawford, C. Trapnell, J. A. Jeddeloh, J. S. Mattick and J. L. Rinn (2011). Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol **30**(1): 99-104.
- Mercer, T. R., S. Neph, M. E. Dinger, J. Crawford, M. A. Smith, A. M. Shearwood, E. Haugen, C. P. Bracken, O. Rackham, J. A. Stamatoyannopoulos, A. Filipovska and J. S. Mattick (2011). The human mitochondrial transcriptome. Cell **146**(4): 645-658.
- Mestdagh, P., E. Fredlund, F. Pattyn, A. Rihani, T. Van Maerken, J. Vermeulen, C. Kumps, B. Menten, K. De Preter, A. Schramm, J. Schulte, R. Noguera, G. Schleiermacher, I. Janoueix-Lerosey, G. Laureys, R. Powel, D. Nittner, J. C. Marine, M. Ringner, F. Speleman and J. Vandesompele (2010). An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours. Oncogene 29(24): 3583-3592.
- Meyers, B. C., T. H. Vu, S. S. Tej, H. Ghazal, M. Matvienko, V. Agrawal, J. Ning and C. D. Haudenschild (2004). Analysis of the transcriptional complexity of Arabidopsis thaliana by massively parallel signature sequencing. Nat Biotechnol **22**(8): 1006-1011.
- Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander and B. E. Bernstein (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature **448**(7153): 553-560.
- Mohammad, F., R. R. Pandey, T. Nagano, L. Chakalova, T. Mondal, P. Fraser and C. Kanduri (2008). Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. Mol Cell Biol **28**(11): 3713-3728.
- Moore, M. J. (2002). Nuclear RNA turnover. Cell **108**(4): 431-434.
- Moore, M. J. and N. J. Proudfoot (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell **136**(4): 688-700.

- Mullen, T. E. and W. F. Marzluff (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. Genes Dev 22(1): 50-65.
- Munchel, S. E., R. K. Shultzaberger, N. Takizawa and K. Weis (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol Biol Cell 22(15): 2787-2795.
- Munroe, S. H. and M. A. Lazar (1991). Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. J Biol Chem **266**(33): 22083-22086.
- Muse, G. W., D. A. Gilchrist, S. Nechaev, R. Shah, J. S. Parker, S. F. Grissom, J. Zeitlinger and K. Adelman (2007). RNA polymerase is poised for activation across the genome. Nat Genet 39(12): 1507-1511.
- Nagano, T., J. A. Mitchell, L. A. Sanz, F. M. Pauler, A. C. Ferguson-Smith, R. Feil and P. Fraser (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science **322**(5908): 1717-1720.
- Nagasaki, H., M. Arita, T. Nishizawa, M. Suwa and O. Gotoh (2005). Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes. Gene **364**: 53-62.
- Nakaya, H. I., P. P. Amaral, R. Louro, A. Lopes, A. A. Fachel, Y. B. Moreira, T. A. El-Jundi, A. M. da Silva, E. M. Reis and S. Verjovski-Almeida (2007). Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. Genome Biol 8(3): R43.
- Narsai, R., K. A. Howell, A. H. Millar, N. O'Toole, I. Small and J. Whelan (2007). Genomewide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. Plant Cell **19**(11): 3418-3436.
- Neil, H., C. Malabat, Y. d'Aubenton-Carafa, Z. Xu, L. M. Steinmetz and A. Jacquier (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature **457**(7232): 1038-1042.
- Numata, K., A. Kanai, R. Saito, S. Kondo, J. Adachi, L. G. Wilming, D. A. Hume, Y. Hayashizaki and M. Tomita (2003). Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. Genome Res **13**(6B): 1301-1306.
- Ogawa, Y., B. K. Sun and J. T. Lee (2008). Intersection of the RNA interference and X-inactivation pathways. Science **320**(5881): 1336-1341.
- Okazaki, Y., M. Furuno, T. Kasukawa, J. Adachi, H. Bono, S. Kondo, I. Nikaido, N. Osato, R. Saito, H. Suzuki, I. Yamanaka, H. Kiyosawa, K. Yagi, Y. Tomaru, Y. Hasegawa, A. Nogami, C. Schonbach, T. Gojobori, R. Baldarelli, D. P. Hill, C. Bult, D. A. Hume, J. Quackenbush, L. M. Schriml, A. Kanapin, H. Matsuda, S. Batalov, K. W. Beisel, J. A. Blake, D. Bradt, V. Brusic, C. Chothia, L. E. Corbani, S. Cousins, E. Dalla, T. A. Dragani, C. F. Fletcher, A. Forrest, K. S. Frazer, T. Gaasterland, M. Gariboldi, C. Gissi, A. Godzik, J. Gough, S. Grimmond, S. Gustincich, N. Hirokawa, I. J. Jackson, E. D. Jarvis, A. Kanai, H. Kawaji, Y. Kawasawa, R. M. Kedzierski, B. L. King, A. Konagaya, I. V. Kurochkin, Y. Lee, B. Lenhard, P. A. Lyons, D. R. Maglott, L. Maltais, L. Marchionni, L. McKenzie, H. Miki, T. Nagashima, K. Numata, T. Okido, W. J. Pavan, G. Pertea, G. Pesole, N. Petrovsky, R. Pillai, J. U. Pontius, D. Qi, S. Ramachandran, T. Ravasi, J. C. Reed, D. J. Reed, J. Reid, B. Z. Ring, M. Ringwald, A. Sandelin, C. Schneider, C. A. Semple, M. Setou, K. Shimada, R. Sultana, Y. Takenaka, M. S. Taylor, R. D. Teasdale, M. Tomita, R. Verardo, L. Wagner, C. Wahlestedt, Y. Wang, Y. Watanabe, C. Wells, L. G. Wilming, A. Wynshaw-Boris, M. Yanagisawa, I. Yang, L. Yang, Z. Yuan, M. Zavolan, Y. Zhu, A. Zimmer, P. Carninci, N. Hayatsu, T. Hirozane-Kishikawa, H. Konno, M. Nakamura, N. Sakazume, K. Sato, T. Shiraki, K. Waki, J. Kawai, K. Aizawa, T. Arakawa, S. Fukuda, A. Hara, W.

- Hashizume, K. Imotani, Y. Ishii, M. Itoh, I. Kagawa, A. Miyazaki, K. Sakai, D. Sasaki, K. Shibata, A. Shinagawa, A. Yasunishi, M. Yoshino, R. Waterston, E. S. Lander, J. Rogers, E. Birney and Y. Hayashizaki (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature **420**(6915): 563-573.
- Oler, A. J., R. K. Alla, D. N. Roberts, A. Wong, P. C. Hollenhorst, K. J. Chandler, P. A. Cassiday, C. A. Nelson, C. H. Hagedorn, B. J. Graves and B. R. Cairns (2010). Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. Nat Struct Mol Biol 17(5): 620-628.
- Orphanides, G. and D. Reinberg (2002). A unified theory of gene expression. Cell **108**(4): 439-451.
- Ota, T., Y. Suzuki, T. Nishikawa, T. Otsuki, T. Sugiyama, R. Irie, A. Wakamatsu, K. Hayashi, H. Sato, K. Nagai, K. Kimura, H. Makita, M. Sekine, M. Obayashi, T. Nishi, T. Shibahara, T. Tanaka, S. Ishii, J. Yamamoto, K. Saito, Y. Kawai, Y. Isono, Y. Nakamura, K. Nagahari, K. Murakami, T. Yasuda, T. Iwayanagi, M. Wagatsuma, A. Shiratori, H. Sudo, T. Hosoiri, Y. Kaku, H. Kodaira, H. Kondo, M. Sugawara, M. Takahashi, K. Kanda, T. Yokoi, T. Furuya, E. Kikkawa, Y. Omura, K. Abe, K. Kamihara, N. Katsuta, K. Sato, M. Tanikawa, M. Yamazaki, K. Ninomiya, T. Ishibashi, H. Yamashita, K. Murakawa, K. Fujimori, H. Tanai, M. Kimata, M. Watanabe, S. Hiraoka, Y. Chiba, S. Ishida, Y. Ono, S. Takiguchi, S. Watanabe, M. Yosida, T. Hotuta, J. Kusano, K. Kanehori, A. Takahashi-Fujii, H. Hara, T. O. Tanase, Y. Nomura, S. Togiya, F. Komai, R. Hara, K. Takeuchi, M. Arita, N. Imose, K. Musashino, H. Yuuki, A. Oshima, N. Sasaki, S. Aotsuka, Y. Yoshikawa, H. Matsunawa, T. Ichihara, N. Shiohata, S. Sano, S. Moriya, H. Momiyama, N. Satoh, S. Takami, Y. Terashima, O. Suzuki, S. Nakagawa, A. Senoh, H. Mizoguchi, Y. Goto, F. Shimizu, H. Wakebe, H. Hishigaki, T. Watanabe, A. Sugiyama, M. Takemoto, B. Kawakami, K. Watanabe, A. Kumagai, S. Itakura, Y. Fukuzumi, Y. Fujimori, M. Komiyama, H. Tashiro, A. Tanigami, T. Fujiwara, T. Ono, K. Yamada, Y. Fujii, K. Ozaki, M. Hirao, Y. Ohmori, A. Kawabata, T. Hikiji, N. Kobatake, H. Inagaki, Y. Ikema, S. Okamoto, R. Okitani, T. Kawakami, S. Noguchi, T. Itoh, K. Shigeta, T. Senba, K. Matsumura, Y. Nakajima, T. Mizuno, M. Morinaga, M. Sasaki, T. Togashi, M. Oyama, H. Hata, T. Komatsu, J. Mizushima-Sugano, T. Satoh, Y. Shirai, Y. Takahashi, K. Nakagawa, K. Okumura, T. Nagase, N. Nomura, H. Kikuchi, Y. Masuho, R. Yamashita, K. Nakai, T. Yada, O. Ohara, T. Isogai and S. Sugano (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet **36**(1): 40-45.
- Pandey, R. R., T. Mondal, F. Mohammad, S. Enroth, L. Redrup, J. Komorowski, T. Nagano, D. Mancini-Dinardo and C. Kanduri (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32(2): 232-246.
- Ponjavic, J., C. P. Ponting and G. Lunter (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res **17**(5): 556-565.
- Prasanth, K. V., S. G. Prasanth, Z. Xuan, S. Hearn, S. M. Freier, C. F. Bennett, M. Q. Zhang and D. L. Spector (2005). Regulating gene expression through RNA nuclear retention. Cell **123**(2): 249-263.
- Preker, P., J. Nielsen, S. Kammler, S. Lykke-Andersen, M. S. Christensen, C. K. Mapendano, M. H. Schierup and T. H. Jensen (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science **322**(5909): 1851-1854.
- Quackenbush, J. (2002). Microarray data normalization and transformation. Nat Genet **32 Suppl**: 496-501.

- Quinlan, A. R. and I. M. Hall (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**(6): 841-842.
- Qureshi, I. A., J. S. Mattick and M. F. Mehler (2010). Long non-coding RNAs in nervous system function and disease. Brain Res **1338**: 20-35.
- Raha, D., Z. Wang, Z. Moqtaderi, L. Wu, G. Zhong, M. Gerstein, K. Struhl and M. Snyder (2010). Close association of RNA polymerase II and many transcription factors with Pol III genes. Proc Natl Acad Sci U S A 107(8): 3639-3644.
- Rahl, P. B., C. Y. Lin, A. C. Seila, R. A. Flynn, S. McCuine, C. B. Burge, P. A. Sharp and R. A. Young (2010). c-Myc regulates transcriptional pause release. Cell **141**(3): 432-445.
- Rapicavoli, N. A., E. M. Poth and S. Blackshaw (2010). The long noncoding RNA RNCR2 directs mouse retinal cell specification. BMC Dev Biol **10**: 49.
- Ratti, A., C. Fallini, L. Cova, R. Fantozzi, C. Calzarossa, E. Zennaro, A. Pascale, A. Quattrone and V. Silani (2006). A role for the ELAV RNA-binding proteins in neural stem cells: stabilization of Msi1 mRNA. J Cell Sci 119(Pt 7): 1442-1452.
- Ravasi, T., H. Suzuki, K. C. Pang, S. Katayama, M. Furuno, R. Okunishi, S. Fukuda, K. Ru, M. C. Frith, M. M. Gongora, S. M. Grimmond, D. A. Hume, Y. Hayashizaki and J. S. Mattick (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res 16(1): 11-19.
- Redrup, L., M. R. Branco, E. R. Perdeaux, C. Krueger, A. Lewis, F. Santos, T. Nagano, B. S. Cobb, P. Fraser and W. Reik (2009). The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. Development **136**(4): 525-530.
- Reis, E. M., R. Louro, H. I. Nakaya and S. Verjovski-Almeida (2005). As antisense RNA gets intronic. OMICS **9**(1): 2-12.
- Reis, E. M., H. I. Nakaya, R. Louro, F. C. Canavez, A. V. Flatschart, G. T. Almeida, C. M. Egidio, A. C. Paquola, A. A. Machado, F. Festa, D. Yamamoto, R. Alvarenga, C. C. da Silva, G. C. Brito, S. D. Simon, C. A. Moreira-Filho, K. R. Leite, L. H. Camara-Lopes, F. S. Campos, E. Gimba, G. M. Vignal, H. El-Dorry, M. C. Sogayar, M. A. Barcinski, A. M. da Silva and S. Verjovski-Almeida (2004). Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. Oncogene 23(39): 6684-6692.
- Rinn, J. L., G. Euskirchen, P. Bertone, R. Martone, N. M. Luscombe, S. Hartman, P. M. Harrison, F. K. Nelson, P. Miller, M. Gerstein, S. Weissman and M. Snyder (2003). The transcriptional activity of human Chromosome 22. Genes Dev **17**(4): 529-540.
- Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal and H. Y. Chang (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell **129**(7): 1311-1323.
- Ross, J. (1995). mRNA stability in mammalian cells. Microbiol Rev **59**(3): 423-450.
- Rozowsky, J., G. Euskirchen, R. K. Auerbach, Z. D. Zhang, T. Gibson, R. Bjornson, N. Carriero, M. Snyder and M. B. Gerstein (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol **27**(1): 66-75.
- Rudnicki, M. A., T. Braun, S. Hinuma and R. Jaenisch (1992). Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. Cell **71**(3): 383-390.
- Russell, J. and J. C. Zomerdijk (2006). The RNA polymerase I transcription machinery. Biochem Soc Symp(73): 203-216.
- Saha, S., A. B. Sparks, C. Rago, V. Akmaev, C. J. Wang, B. Vogelstein, K. W. Kinzler and V. E. Velculescu (2002). Using the transcriptome to annotate the genome. Nat Biotechnol 20(5): 508-512.

- Sasaki, Y. T., T. Ideue, M. Sano, T. Mituyama and T. Hirose (2009). MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci U S A **106**(8): 2525-2530.
- Sasaki, Y. T., M. Sano, T. Ideue, T. Kin, K. Asai and T. Hirose (2007). Identification and characterization of human non-coding RNAs with tissue-specific expression. Biochem Biophys Res Commun **357**(4): 991-996.
- Schadt, E. E., S. W. Edwards, D. GuhaThakurta, D. Holder, L. Ying, V. Svetnik, A. Leonardson, K. W. Hart, A. Russell, G. Li, G. Cavet, J. Castle, P. McDonagh, Z. Kan, R. Chen, A. Kasarskis, M. Margarint, R. M. Caceres, J. M. Johnson, C. D. Armour, P. W. Garrett-Engele, N. F. Tsinoremas and D. D. Shoemaker (2004). A comprehensive transcript index of the human genome generated using microarrays and computational approaches. Genome Biol 5(10): R73.
- Schmid, M. and T. H. Jensen (2008). The exosome: a multipurpose RNA-decay machine. Trends Biochem Sci **33**(10): 501-510.
- Schwanekamp, J. A., M. A. Sartor, S. Karyala, D. Halbleib, M. Medvedovic and C. R. Tomlinson (2006). Genome-wide analyses show that nuclear and cytoplasmic RNA levels are differentially affected by dioxin. Biochim Biophys Acta **1759**(8-9): 388-402.
- Seidl, C. I., S. H. Stricker and D. P. Barlow (2006). The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. EMBO J **25**(15): 3565-3575.
- Seki, M., M. Satou, T. Sakurai, K. Akiyama, K. Iida, J. Ishida, M. Nakajima, A. Enju, M. Narusaka, M. Fujita, Y. Oono, A. Kamei, K. Yamaguchi-Shinozaki and K. Shinozaki (2004). RIKEN Arabidopsis full-length (RAFL) cDNA and its applications for expression profiling under abiotic stress conditions. J Exp Bot 55(395): 213-223.
- Senner, C. E. and N. Brockdorff (2009). Xist gene regulation at the onset of X inactivation. Curr Opin Genet Dev **19**(2): 122-126.
- Sharma, S., G. M. Findlay, H. S. Bandukwala, S. Oberdoerffer, B. Baust, Z. Li, V. Schmidt, P. G. Hogan, D. B. Sacks and A. Rao (2011). Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex. Proc Natl Acad Sci U S A **108**(28): 11381-11386.
- Sharova, L. V., A. A. Sharov, T. Nedorezov, Y. Piao, N. Shaik and M. S. Ko (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res **16**(1): 45-58.
- Simon, R. M. and K. Dobbin (2003). Experimental design of DNA microarray experiments. Biotechniques **Suppl**: 16-21.
- Slomovic, S., E. Fremder, R. H. Staals, G. J. Pruijn and G. Schuster (2010). Addition of poly(A) and poly(A)-rich tails during RNA degradation in the cytoplasm of human cells. Proc Natl Acad Sci U S A 107(16): 7407-7412.
- Sobell, H. M. (1985). Actinomycin and DNA transcription. Proc Natl Acad Sci U S A **82**(16): 5328-5331.
- Sone, M., T. Hayashi, H. Tarui, K. Agata, M. Takeichi and S. Nakagawa (2007). The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J Cell Sci 120(Pt 15): 2498-2506.
- Stoecklin, G., T. Mayo and P. Anderson (2006). ARE-mRNA degradation requires the 5'-3' decay pathway. EMBO Rep **7**(1): 72-77.
- Stolc, V., Z. Gauhar, C. Mason, G. Halasz, M. F. van Batenburg, S. A. Rifkin, S. Hua, T. Herreman, W. Tongprasit, P. E. Barbano, H. J. Bussemaker and K. P. White (2004). A gene expression map for the euchromatic genome of Drosophila melanogaster. Science **306**(5696): 655-660.

- Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol **14**(2): 103-105.
- Sunwoo, H., M. E. Dinger, J. E. Wilusz, P. P. Amaral, J. S. Mattick and D. L. Spector (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res **19**(3): 347-359.
- Szymanski, M., V. A. Erdmann and J. Barciszewski (2003). Noncoding regulatory RNAs database. Nucleic Acids Res **31**(1): 429-431.
- Taft, R. J., E. A. Glazov, N. Cloonan, C. Simons, S. Stephen, G. J. Faulkner, T. Lassmann, A. R. Forrest, S. M. Grimmond, K. Schroder, K. Irvine, T. Arakawa, M. Nakamura, A. Kubosaki, K. Hayashida, C. Kawazu, M. Murata, H. Nishiyori, S. Fukuda, J. Kawai, C. O. Daub, D. A. Hume, H. Suzuki, V. Orlando, P. Carninci, Y. Hayashizaki and J. S. Mattick (2009). Tiny RNAs associated with transcription start sites in animals. Nat Genet 41(5): 572-578.
- Taft, R. J., K. C. Pang, T. R. Mercer, M. Dinger and J. S. Mattick (2010). Non-coding RNAs: regulators of disease. J Pathol **220**(2): 126-139.
- Taft, R. J., M. Pheasant and J. S. Mattick (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays **29**(3): 288-299.
- Tahira, A. C., M. S. Kubrusly, M. F. Faria, B. Dazzani, R. S. Fonseca, V. Maracaja-Coutinho, S. Verjovski-Almeida, M. C. Machado and E. M. Reis (2011). Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. Mol Cancer 10: 141.
- Team, R. D. C. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing. from http://www.R-project.org/.
- Tomari, Y. and P. D. Zamore (2005). Perspective: machines for RNAi. Genes Dev 19(5): 517-529.
- Topisirovic, I., B. Culjkovic, N. Cohen, J. M. Perez, L. Skrabanek and K. L. Borden (2003). The proline-rich homeodomain protein, PRH, is a tissue-specific inhibitor of eIF4E-dependent cyclin D1 mRNA transport and growth. EMBO J **22**(3): 689-703.
- Trask, H. W., R. Cowper-Sal-lari, M. A. Sartor, J. Gui, C. V. Heath, J. Renuka, A. J. Higgins, P. Andrews, M. Korc, J. H. Moore and C. R. Tomlinson (2009). Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs. RNA 15(10): 1917-1928.
- Tripathi, V., J. D. Ellis, Z. Shen, D. Y. Song, Q. Pan, A. T. Watt, S. M. Freier, C. F. Bennett, A. Sharma, P. A. Bubulya, B. J. Blencowe, S. G. Prasanth and K. V. Prasanth (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell **39**(6): 925-938.
- Tseng, J. J., Y. T. Hsieh, S. L. Hsu and M. M. Chou (2009). Metastasis associated lung adenocarcinoma transcript 1 is up-regulated in placenta previa increta/percreta and strongly associated with trophoblast-like cell invasion in vitro. Mol Hum Reprod **15**(11): 725-731.
- Tusher, V. G., R. Tibshirani and G. Chu (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A **98**(9): 5116-5121.
- Tycowski, K. T., M. D. Shu and J. A. Steitz (1993). A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. Genes Dev **7**(7A): 1176-1190.
- Valen, E., G. Pascarella, A. Chalk, N. Maeda, M. Kojima, C. Kawazu, M. Murata, H. Nishiyori, D. Lazarevic, D. Motti, T. T. Marstrand, M. H. Tang, X. Zhao, A. Krogh, O. Winther, T. Arakawa, J. Kawai, C. Wells, C. Daub, M. Harbers, Y. Hayashizaki, S.

- Gustincich, A. Sandelin and P. Carninci (2009). Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. Genome Res **19**(2): 255-265.
- Valencia, P., A. P. Dias and R. Reed (2008). Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc Natl Acad Sci U S A **105**(9): 3386-3391.
- van Bakel, H., C. Nislow, B. J. Blencowe and T. R. Hughes (2010). Most "dark matter" transcripts are associated with known genes. PLoS Biol 8(5): e1000371.
- van Dijk, E. L., C. L. Chen, Y. d'Aubenton-Carafa, S. Gourvennec, M. Kwapisz, V. Roche, C. Bertrand, M. Silvain, P. Legoix-Ne, S. Loeillet, A. Nicolas, C. Thermes and A. Morillon (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature **475**(7354): 114-117.
- Vargas, D. Y., K. Shah, M. Batish, M. Levandoski, S. Sinha, S. A. Marras, P. Schedl and S. Tyagi (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell 147(5): 1054-1065.
- Visa, N., F. Puvion-Dutilleul, F. Harper, J. P. Bachellerie and E. Puvion (1993). Intranuclear distribution of poly(A) RNA determined by electron microscope in situ hybridization. Exp Cell Res **208**(1): 19-34.
- Wang, D., I. Garcia-Bassets, C. Benner, W. Li, X. Su, Y. Zhou, J. Qiu, W. Liu, M. U. Kaikkonen, K. A. Ohgi, C. K. Glass, M. G. Rosenfeld and X. D. Fu (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature 474(7351): 390-394.
- Wang, K. C. and H. Y. Chang (2011). Molecular mechanisms of long noncoding RNAs. Mol Cell **43**(6): 904-914.
- Wang, X., S. Arai, X. Song, D. Reichart, K. Du, G. Pascual, P. Tempst, M. G. Rosenfeld, C. K. Glass and R. Kurokawa (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature 454(7200): 126-130.
- Wang, Y., C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag and P. O. Brown (2002). Precision and functional specificity in mRNA decay. Proc Natl Acad Sci U S A 99(9): 5860-5865.
- Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C.

- Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody and E. S. Lander (2002). Initial sequencing and comparative analysis of the mouse genome. Nature **420**(6915): 520-562.
- Wheeler, D. L., T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. Dicuccio, R. Edgar, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, O. Khovayko, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, V. Miller, J. Ostell, K. D. Pruitt, G. D. Schuler, M. Shumway, E. Sequeira, S. T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, R. L. Tatusov, T. A. Tatusova, L. Wagner and E. Yaschenko (2008). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 36(Database issue): D13-21.
- Whitfield, M. L., G. Sherlock, A. J. Saldanha, J. I. Murray, C. A. Ball, K. E. Alexander, J. C. Matese, C. M. Perou, M. M. Hurt, P. O. Brown and D. Botstein (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13(6): 1977-2000.
- Willingham, A. T. and T. R. Gingeras (2006). TUF love for "junk" DNA. Cell **125**(7): 1215-1220.
- Willingham, A. T., A. P. Orth, S. Batalov, E. C. Peters, B. G. Wen, P. Aza-Blanc, J. B. Hogenesch and P. G. Schultz (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science **309**(5740): 1570-1573.
- Wilson, B. A. and J. Masel (2011). Putatively noncoding transcripts show extensive association with ribosomes. Genome Biol Evol 3: 1245-1252.
- Wilusz, C. J., M. Wormington and S. W. Peltz (2001). The cap-to-tail guide to mRNA turnover. Nat Rev Mol Cell Biol **2**(4): 237-246.
- Wilusz, J. E., S. M. Freier and D. L. Spector (2008). 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell **135**(5): 919-932.
- Wilusz, J. E., H. Sunwoo and D. L. Spector (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev **23**(13): 1494-1504.
- Wu, Q., Y. C. Kim, J. Lu, Z. Xuan, J. Chen, Y. Zheng, T. Zhou, M. Q. Zhang, C. I. Wu and S. M. Wang (2008). Poly A- transcripts expressed in HeLa cells. PLoS One **3**(7): e2803.
- Wyers, F., M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour, J. Boulay, B. Regnault, F. Devaux, A. Namane, B. Seraphin, D. Libri and A. Jacquier (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell **121**(5): 725-737.
- Xu, Z., W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Munster, J. Camblong, E. Guffanti, F. Stutz, W. Huber and L. M. Steinmetz (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature 457(7232): 1033-1037.
- Yamada, K., J. Lim, J. M. Dale, H. Chen, P. Shinn, C. J. Palm, A. M. Southwick, H. C. Wu, C. Kim, M. Nguyen, P. Pham, R. Cheuk, G. Karlin-Newmann, S. X. Liu, B. Lam, H. Sakano, T. Wu, G. Yu, M. Miranda, H. L. Quach, M. Tripp, C. H. Chang, J. M. Lee, M. Toriumi, M. M. Chan, C. C. Tang, C. S. Onodera, J. M. Deng, K. Akiyama, Y.

- Ansari, T. Arakawa, J. Banh, F. Banno, L. Bowser, S. Brooks, P. Carninci, Q. Chao, N. Choy, A. Enju, A. D. Goldsmith, M. Gurjal, N. F. Hansen, Y. Hayashizaki, C. Johnson-Hopson, V. W. Hsuan, K. Iida, M. Karnes, S. Khan, E. Koesema, J. Ishida, P. X. Jiang, T. Jones, J. Kawai, A. Kamiya, C. Meyers, M. Nakajima, M. Narusaka, M. Seki, T. Sakurai, M. Satou, R. Tamse, M. Vaysberg, E. K. Wallender, C. Wong, Y. Yamamura, S. Yuan, K. Shinozaki, R. W. Davis, A. Theologis and J. R. Ecker (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome. Science 302(5646): 842-846.
- Yan, M. D., C. C. Hong, G. M. Lai, A. L. Cheng, Y. W. Lin and S. E. Chuang (2005). Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas. Hum Mol Genet **14**(11): 1465-1474.
- Yang, E., E. van Nimwegen, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnasco and J. E. Darnell, Jr. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res **13**(8): 1863-1872.
- Yang, F., F. Yi, Z. Zheng, Z. Ling, J. Ding, J. Guo, W. Mao, X. Wang, X. Ding, Z. Liang and Q. Du (2012). Characterization of a carcinogenesis-associated long non-coding RNA. RNA Biol 9(1).
- Yang, L., M. O. Duff, B. R. Graveley, G. G. Carmichael and L. L. Chen (2011). Genomewide characterization of non-polyadenylated RNAs. Genome Biol **12**(2): R16.
- Yang, Y. H., S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai and T. P. Speed (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res **30**(4): e15.
- Yeap, B. B., R. G. Krueger and P. J. Leedman (1999). Differential posttranscriptional regulation of androgen receptor gene expression by androgen in prostate and breast cancer cells. Endocrinology **140**(7): 3282-3291.
- Zhang, L. F., K. D. Huynh and J. T. Lee (2007). Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell **129**(4): 693-706.
- Zhang, Z. and G. G. Carmichael (2001). The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell **106**(4): 465-475.
- Zhao, C. and T. Hamilton (2007). Introns regulate the rate of unstable mRNA decay. J Biol Chem **282**(28): 20230-20237.
- Zhao, J., T. K. Ohsumi, J. T. Kung, Y. Ogawa, D. J. Grau, K. Sarma, J. J. Song, R. E. Kingston, M. Borowsky and J. T. Lee (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40(6): 939-953.
- Zhao, J., B. K. Sun, J. A. Erwin, J. J. Song and J. T. Lee (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science **322**(5902): 750-756.

8. ANEXOS

Anexo 1 - ncRNAs intrônicos enriquecidos no núcleo e no citoplasma de células HeLa.

Anexo 2 - RNA-FISH (Fluorescent In Situ Hybridization)

Clonagem das sondas de RNA-FISH

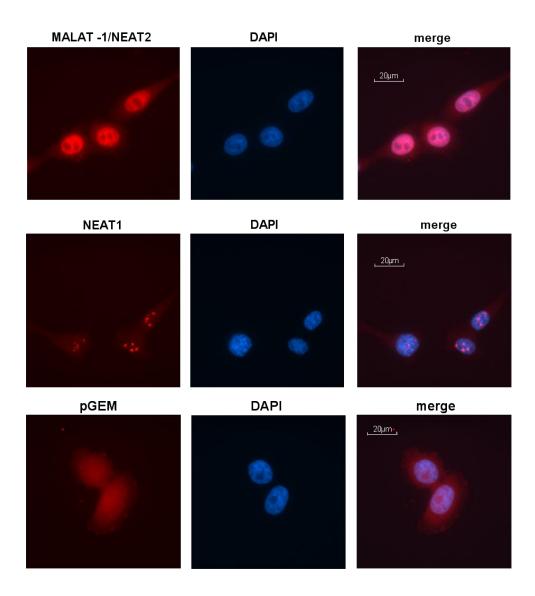
Após amplificação dos transcritos de interesse por PCR, o produto da reação foi submetido à eletroforese em gel de agarose 1,7% contendo brometo de etídeo 0,5 µg/mL em tampão TAE (Tris-Acetato 40mM, EDTA 1mM). O produto da reação foi visualizado em um fotodocumentador MiniBis Pro (DNR Bio-Imaging Systems) sob luz UV. As bandas correspondentes aos produtos de PCR obtidos foram cortadas do gel e purificadas com o kit Quiack Gel Extraction (Qiagen), de acordo com o protocolo do fabricante. Os produtos da PCR purificados foram clonados em pGEM-T Easy Vector System (Promega), de acordo com o protocolo do fabricante. Em seguida, foi realizada diálise por 2 horas utilizando-se uma membrana VS de 0,025um (Millipore) para a remoção dos sais da reação. A reação dialisada foi utilizada para a transformação de células competentes de Escherichia coli, cepa DH10B (Invitrogen), através de eletroporação. Imediatamente após a eletroporação, 400µL de meio SOC (bacto-triptona 20 mg/ml, extrato de levedura 5mg/ml, NaCl 0,584mg/ml, KCl 0,186mg/ml, MgCl2 10mM, MgSO4 10mM, glicose 20mM, pH 7,0) foram adicionados às células, que foram incubadas a 37 °C por 1h. Posteriormente, as bactérias foram semeadas em meio LB-ágar (peptona de caseína 1%, extrato de levedura 0,5%, NaCl 1%, ágar bacteriológico 1,4%, pH 7,0) contendo ampicilina (10 mg/mL), X-Gal 20mg/ml (5-bromo-4chloro-3-indolyl-b-D-galactopyranoside) **IPTG** 20 mg/ml(Isopropyl β-D-1thiogalactopyranoside) e mantidas a 37 °C, *overnight*. As colônias selecionadas foram coletadas e inoculadas em 5mL de meio LB líquido com ampicilina (10 mg/mL) a 37 °C, *overnight*, com agitação constante. A extração do DNA plasmidial contendo o inserto foi realizada com o Plasmid Mini Kit (Qiagen), de acordo com o protocolo do fabricante. O DNA obtido foi quantificado por sua densidade óptica a 260 nm e sua pureza foi atestada pela razão 260/280nm. As medidas foram realizadas no espectrofotômetro NanoDrop (Thermo Scientific). As construções foram utilizadas para a preparação de sondas do RNA-FISH através da reação de *nick-translation*, como descrito a seguir.

Preparo das sondas e detecção de ncRNAs através de RNA-FISH

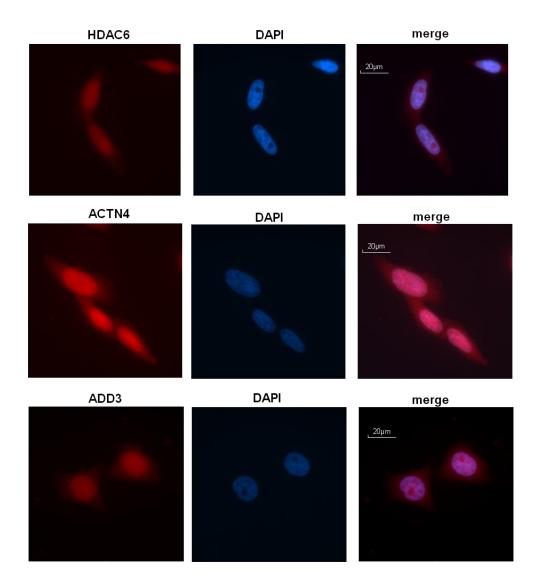
Os experimentos de hibridização *in situ* fluorescente (RNA-Fluorescent In Situ Hybridization) foram baseados nos protocolos descritos e estabelecidos no laboratório do Dr. David L. Spector (Lamond and Spector 2003; Prasanth et al. 2005). As sondas de cDNA foram produzidas por reação de *nick-translation*, segundo protocolo do fabricante (Nick Translation Kit; Abbot Molecular), na presença de 0,01mM do análogo fluorescente de dUTP (dUTP-Texas Red; Abbot Molecular), utilizando os transcritos clonados em vetor pGEM, como descrito no item anterior.

Células HeLa com ~70% de confluência foram crescidas em lamínulas e fixadas em solução de PBS pH 7.4 contendo 4% formaldeído por 15 min à temperatura ambiente. Após 3 lavagens em PBS (10 minutos cada) com agitação, as células foram permeabilizadas em gelo através de incubação por 10 minutos em PBS contendo 0.5% Triton-X100 (Sigma) e 10mM inibidor de RNAses VRC (Vanadyl Ribonucleoside Complexes; New England Biolabs). As células permeabilizadas foram lavadas 3 vezes em PBS (10 minutos cada) sob agitação, e saturadas em 2X SSC (NaCl 0,3M e citrato de sódio 0,03 M) por 10 minutos. Em seguida, as lamínulas foram incubadas com 2μg de sonda de cDNA fluorescente em 20 μl de tampão de

hibridização (1 μg/ml tRNA de *E. coli*, 2X SSC, 10% dextran sulfato) por 16h a 37 °C, em uma câmara úmida. Após a hibridização, as lamínulas foram submetidas à lavagens sucessivas com 2X SSC/50% formamida a 37 °C (5 minutos), 2X SSC (5 minutos) e 1X SSC (5 min) e, posteriormente, incubadas em solução de 4X SSC contendo o corante de DNA DAPI (Diamino-Phenol-Indol, Invitrogen). As lamínulas foram montadas em lâminas para microscopia com meio de montagem para fluorescência (Vectashield). As imagens foram obtidas através de microscopia de fluorescência em um microscópio invertido (TE 300 NIKON) e estão apresentadas nas figuras abaixo.



Controles utilizados nos experimentos de localização sub-celular de ncRNAs nucleares através de RNA-FISH. Imagem de fluorescência com aumento de 630 X. Células HeLa foram hibridizadas com sondas de cDNA fluorescentes marcadas com Texas Red (vermelho) complementares aos transcritos MALAT-1/NEAT2 (painel superior) e NEAT1 (painel central). Nota-se a distribuição desses RNAs em *foci* nucleares discretos, que já foi mostrado na literaura corresponder a *speckles* (MALAT-1/NEAT2) ou *paraspeckles* (NEAT1) nucleares. Como controle negativo foi utilizada uma sonda fluorescente complementar ao vetor de clonagem vazio (pGEM) (painel inferior). O DNA nuclear foi corado com DAPI (azul), permitindo a sobreposição das imagens nos painéis à direita ("merge").



NcRNAs intrônicos com localização predominantemente nuclear observados através de RNA-FISH. Imagem de fluorescência com aumento de 630 X. Células HeLa foram hibridizadas com sondas de cDNA fluorescentes marcadas com Texas Red (vermelho) complementares aos transcritos intrônicos HDAC6 (painel superior), ACTN4 (painel central) e ADD3 (painel inferior). Nota-se a localização predominante dos transcritos no núcleo da célula, dispersos de forma homogênea pelo compartimento nuclear. O DNA nuclear foi corado com DAPI (azul), permitindo a sobreposição das imagens nos painéis à direita ("merge").

Anexo 3 - Súmula Curricular