UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)

MURILO SENA AMARAL

Identificação de RNAs longos não-codificadores de proteínas regulados por micro-RNAs

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:

30/10/2013

MURILO SENA AMARAL

Identificação de RNAs longos não-codificadores de proteínas regulados por micro-RNAs

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(a): Prof. Dr. Sergio Verjovski-Almeida

São Paulo

2013

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Amaral, Murilo Sena
A485i Identificação de RNAs longos não-codificadores de proteínas regulados por micro-RNAs / Murilo Sena Amaral. -- São Paulo, 2013.
147p.

Tese (doutorado) - Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica.
Orientador: Verjovski-Almeida, Sergio

1. Regulação gênica 2. Expressão gênica 3. Controle gênico 4. RNA I. T. II Verjovski-Almeida, Sergio, orientador.

574.88 CDD

À minha mãe pelo amor incondicional.

Ao meu pai por sua dedicação.

Aos meus irmãos, por serem mais do que somente irmãos.

Todo santo dia, força e alegria.

AGRADECIMENTOS

Ao Prof. Dr. Sergio Verjovski-Almeida, competente e comprometido pesquisador e orientador que me ajudou a construir uma visão crítica essencial para o estudo científico. Agradeço também a confiança depositada em mim ao me atribuir projetos paralelos a este, em colaboração com outros pesquisadores. Essas experiências foram cruciais para a minha formação.

À minha família, suporte essencial para a minha vida. À minha mãe Edivê, ao meu pai Nilton (Foka Senna), que não mediram esforços para me proporcionar sempre o melhor. Aos meus irmãos Soraia e Diogo, aos novos irmãos Wagno e Jeane, à Vó Linda e aos meus sobrinhos Sara, Ryan e Sofia por encherem a minha vida de alegria.

Ao Luan Perusso, por todos os anos juntos que valeram totalmente a pena. Obrigado por estar ao meu lado nos momentos bons e ruins, pela paciência e por me ensinar a ser menos rígido com a vida.

À minha nova família recebida: Ronize, Samantha e Dorival.

A todos os meus familiares e amigos, que mesmo distantes, sempre me apoiaram e mostraram admiração pelo meu trabalho.

Ao Prof. Dr. Eduardo Reis, pelas críticas e sugestões que com certeza me ensinaram muito e ajudaram a melhorar os trabalhos desenvolvidos.

À Profa. Dra. Juliana Lopes Rangel Fietto que me iniciou na carreira científica e me mostrou um lado extremamente cativante do ser humano. Aos Professores Dr. Rogelio Lopes Brandão, Dr. Ieso de Miranda Castro, Dr. Luis Carlos Crocco Afonso e Dr. Luciano Gomes Fietto pelas oportunidades no começo da iniciação científica.

À Dra. Layla Farage Martins e à Cláudia Siqueira Campos pelo suporte nos sequenciamentos e pela incansável disposição em ajudar.

Ao Rodrigo Borges, bioinformata muito competente pela imensa ajuda com as análises de bioinformática.

À Ana Paula, secretária muito eficiente, por tornar a minha vida e de todos no laboratório muito mais simples.

Ao Jefferson dos Santos, Renato Alvarenga e Denise Yamamoto pela ajuda diária no laboratório.

Aos amigos Carlos de Ocesano, Lauren Camargo, Santiago Vilella, Maria Emilia Zenteno, Ângela Fachel e Ana Carolina Tahira pela partilha de momentos inesquecíveis dentro e fora do laboratório.

Ao amigo Felipe Beckedorff, por compartilhar padronizações, implementações e, principalmente, pelas discussões científicas, sugestões e críticas nos projetos.

Aos amigos do laboratório Ana Ayupe, Giulliana, Letícia, Kátia, Mariana, Bianca, Kleber, Dinar, Otto, Ester, Lucas, Gabriel e Yuri pelo suporte em vários momentos.

Aos amigos Vinícius Maracajá-Coutinho, Valdir Blasios, Renato Grempel e Cícero Alves pelos bons momentos compartilhados.

À República Cruz Vermelha, que após mais de 40 anos, continua sendo a casa de muitos dos seus ex-alunos e amigos.

Aos amigos do vôlei pelo refúgio, amizade, pelas viagens compartilhadas e pelas vitórias e derrotas: Carlos, PT, Zozó, Banzai, Rafa, João, Simões, Guga, Black, Gui, Thiago, Chu, Felipe, Osama, Lele, Marani, Marino, Anderson, Thenner, Marcus, Diogo, Gabriel, Eduardo Rocha, Marcelo Mendes, Felipe Mochiute e Rodrigo Nicoletti.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e à Financiadora de Estudos e Projetos (FINEP) pelo suporte financeiro.

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

RESUMO

Amaral, M.S. **Identificação de RNAs longos não-codificadores de proteínas regulados por micro-RNAs.** 2013. 147p. Tese de Doutorado - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

Estudos recentes têm revelado que a maior parte dos transcritos gerados em células humanas é composta por RNAs não-codificadores de proteínas (ncRNAs). Uma parte desses ncRNAs compreende a classe de RNAs curtos, que possuem menos que 200 nucleotídeos. Os micro-RNAs (miRNAs) fazem parte dessa classe e têm sido alvo de grande interesse, pois são preditos como possíveis reguladores de mais de 60% dos RNAs mensageiros (mRNAs) humanos. Outra classe dos ncRNAs é composta por ncRNAs longos (lncRNAs, com mais de 200 nucleotídeos), que são transcritos a partir de regiões intergênicas e intrônicas do genoma humano e possuem várias funções, muitas delas relacionadas ao controle da expressão de mRNAs. Recentemente, os lncRNAs têm sido caracterizados quanto à sua estrutura e função. No entanto, muito pouco se sabe sobre os mecanismos pelos quais os lncRNAs são regulados. Este trabalho teve como objetivo avaliar se lncRNAs são regulados por miRNAs em células humanas. Para tanto, identificamos lncRNAs ligados ao complexo de silenciamento induzido por RNA (RISC) em células da linhagem HeLa, utilizando um método aqui desenvolvido de geração de bibliotecas de cDNA direcionadas para sequenciamento em larga escala na plataforma 454/Roche. Em paralelo, sequenciamos os miRNAs ligados ao RISC nestas mesmas células. Os resultados obtidos mostram que centenas de lncRNAs de diversas classes se ligam ao RISC em células HeLa, juntamente com milhares de mRNAs e várias centenas de miRNAs. Entre os miRNAs, encontramos 37 que são preditos como alvejando os lncRNAs detectados. Estes miRNAs constituem possíveis reguladores dos lncRNAs e, portanto, nosso trabalho estabelece um mapa experimental de interações diretas entre lncRNAs e miRNAs. Dentre os lncRNAs identificados ligados ao RISC neste trabalho, destaca-se o *TUG1*, lincRNA sabidamente envolvido na regulação de genes relacionados à apoptose e ao ciclo celular. Mostramos por ensaio de super-expressão de miRNAs e qPCR que *TUG1* é regulado pelo miRNA-148b, um dos miRNAs por nós detectados que possui um sítio alvo altamente conservado em mamíferos localizado na extremidade 3’ de *TUG1*. Em conjunto, este

trabalho contribui para o entendimento da regulação dos níveis de expressão de lncRNAs em células humanas e abre perspectivas para a modulação de miRNAs como estratégia de regulação dos níveis e das funções de lncRNAs.

Palavras-chave: RNAs longos não-codificadores, micro-RNAs, regulação da expressão gênica, RNA-Seq fita-específico.

ABSTRACT

Amaral, M.S. **Identification of long non-protein coding RNAs regulated by micro-RNAs.** 2013. 147p. PhD Thesis - Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Recent studies have revealed that the largest fraction of the transcripts generated in human cells is composed of non-protein coding RNAs (ncRNAs). A portion of these RNAs encompasses the class of short RNAs, which are less than 200 nucleotides in length. Micro-RNAs (miRNAs) are part of this class and are of great interest, as they are predicted to target over 60% of the human messenger RNAs (mRNAs). Another class of ncRNAs is composed of long ncRNAs (lncRNAs, longer than 200 nucleotides), which are transcribed from intergenic and intronic regions of the human genome and have several functions, many of them related to the control of the mRNA expression. Recently, the structure and function of lncRNAs have been characterized. However, little is known about the mechanisms involved in lncRNA regulation. This work aimed to evaluate whether lncRNAs are regulated by miRNAs in human cells. For this purpose, we identified lncRNAs bound to the RNA-induced silencing complex (RISC) in HeLa cells using a method developed here for the generation of strand-specific cDNA libraries for large scale RNA-sequencing in the 454/Roche plataform. In parallel, we sequenced the miRNAs bound to RISC in these cells. Our results show that hundreds of lncRNAs from diverse classes are bound to RISC in HeLa cells, along with thousands of mRNAs and several hundred miRNAs. Among the miRNAs we identified 37 that are predicted to target the detected lncRNAs. These miRNAs are possible regulators of the lncRNAs, and therefore our work establishes an experimental map of direct interactions between lncRNAs and miRNAs. The lncRNA *TUG1*, a lincRNA involved in the regulation of genes related to apoptosis and cell cycle, was identified among the lncRNAs bound to RISC. We showed by miRNA over-expression and qPCR that *TUG-1* is regulated by the miRNA-148b, which is one of the miRNAs detected in our sequencings and has a binding site highly conserved in mammals located at the *TUG1* 3` end. Taken together, our results contribute to the understanding of the regulation of the lncRNA expression levels in human cells and open perspectives for the modulation of miRNAs as a strategy to regulate the levels and functions of lncRNAs.

Keywords: Long non-coding RNAs, micro-RNAs, regulation of gene expression, strand-specific RNA-Seq.

LISTA DE ABREVIATURAS E SIGLAS

5-AZA: 5-*Aza-2'-deoxicitidina*

Ago2: Argonauta2

ARE: *AU-Rich Element*

aRNAs: RNAs amplificados

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

cDNA: DNA complementar

Ct: *Cycle Threshold*

DMEM: *Dulbecco's Modified Eagle Medium*

DMSO: Dimetilsulfóxido

DNA: *Deoxyribonucleic Acid*

DNMT: DNA-metiltransferase

dscDNA: *double-stranded cDNA*

emPCR: PCR em emulsão

ENCODE: *Encyclopedia of DNA Elements*

ESTs: *Expressed Sequence Tags*

FANTOM: *Functional Annotation of Mouse*

H3K4me3: trimetilação da lisina 4 da histona 3

H3K36me3: trimetilação da lisina 36 da histona 3

HOTAIR: Hox Transcript Antisense RNA

IgG: Imunoglobulina G

Kb: *kilobases*

lincRNA: ncRNA longo intergênico

lncRNA: RNA longo não-codificador de proteína

MALAT-1: *Metastasis-Associated Lung Adenocarcinoma Transcript 1*

miRNA: micro-RNA

mRNA: RNA mensageiro

ncRNA: RNA não-codificador de proteína

nt: Nucleotídeos

PCR: *Polymerase Chain Reaction*

piRNAs: *piwi-interacting RNA*

PRC2: *Polycomb Repressive Complex 2*

qPCR: PCR quantitativo

RIN: *RNA Integrity Number*

RIP: *RNA-binding protein immunoprecipitation*

RISC: *RNA-Induced Silencing Complex*

RNAP II: RNA polimerase II

RNA-Seq: Sequenciamento em larga escala de RNA

RNP: Ribonucleoproteína

RPKM: *Reads per kilobase of exon model per million mapped reads*

rRNA: RNA ribossomal

RT: *Reverse Transcription*

snoRNA: *small nucleolar RNA*

snRNA: *small nuclear RNA*

sscDNA: single-stranded cDNA

tRNA: RNA transportador

TUG1: *Taurine upregulated gene 1*

UTR: *untranslated region*

SUMÁRIO

1. INTRODUÇÃO 14
1.1 A complexidade do transcritoma de eucariotos 14
1.2 RNAs longos não-codificadores de proteínas em eucariotos 18
1.3 micro-RNAs 22
1.4 Regulação da expressão de lncRNAs 25
1.4.1 Regulação transcricional de lncRNAs 26
1.4.2 Regulação pós-transcricional de lncRNAs 29
2. OBJETIVOS 36
2.1 Objetivo Geral 36
2.2 Objetivos Específicos 36
3. MATERIAIS E MÉTODOS 37
3.1 Linhagem celular e condições de cultivo 37
3.2 Imunoprecipitação nativa de complexos ribonucleoprotéicos – RIP 37
3.3 Extração de RNA, purificação e tratamento com DNase 39
3.4 Avaliação da qualidade do RNA 41
3.5 Western Blot 42
3.6 Desenvolvimento de método de geração de bibliotecas fita-específicas para sequenciamento de RNAs longos no 454/Roche 43
3.7 Sequenciamento em larga escala de RNAs longos 49
3.8 Sequenciamento em larga-escala de RNAs curtos 50
3.9 Análises de Bioinformática 50
3.9.1 Análise Bioinformática de RNAs longos 50
3.9.2 Análise Bioinformática de RNAs curtos 52
3.9.3 Identificação de lncRNAs como possíveis alvos de miRNAs 53
3.10 *Crosslink and immunoprecipitation* – CLIP 53
3.11 Desenho de oligonucleotídeos iniciadores (*primers*) 54
3.12 Transcrição Reversa e Síntese de cDNA 56
3.13 Reação da Polimerase em Cadeia (*Polymerase Chain Reaction* - PCR) 56
3.13.1 PCR 56

3.13.2 PCR quantitativo em tempo real (qPCR) 56
3.14 Superexpressão de miRNAs 58
4. RESULTADOS 59
4.1 Imunoprecipitação do complexo de silenciamento induzido por RNA contendo a proteína Argonauta2 59
4.2 Desenvolvimento de método de geração de bibliotecas fita-específicas para sequenciamento de RNAs longos no 454/Roche 61
4.3 Identificação de lncRNAs ligados ao complexo RISC 69
4.4 Validação da interação entre lncRNAs e o complexo RISC 74
4.5 Identificação de miRNAs ligados ao complexo RISC 78
4.5.1 Identificação dos miRNAs já anotados em bancos de dados 81
4.5.2 Identificação de possíveis novos miRNAs humanos 82
4.6 Identificação de lncRNAs alvejados por miRNAs 83
4.7 Regulação do lincRNA *TUG1* pelo miRNA-148b 86
5. DISCUSSÃO 90
5.1 Desenvolvimento de método de geração de bibliotecas fita-específicas para sequenciamento de RNAs longos no 454/Roche 90
5.2 lncRNAs regulados por miRNAs 95
6 CONCLUSÕES 102
7 REFERÊNCIAS BIBLIOGRÁFICAS 103
APÊNDICE 126
ANEXOS ELETRÔNICOS 148
ANEXO 149

1. INTRODUÇÃO

1.1 A complexidade do transcritoma de eucariotos

O desenvolvimento de novas tecnologias nos campos da genômica e transcritômica nas últimas décadas tem permitido o aprofundamento do estudo do perfil transcricional de diversos organismos. O estudo do transcritoma de vários eucariotos, por exemplo, tem revelado que a maior parte do genoma desses organismos é transcrita, mas não possui evidência de ser traduzida em proteínas (Carninci *et al.*, 2005; David *et al.*, 2006; Birney *et al.*, 2007; Gerstein *et al.*, 2010; Roy *et al.*, 2010; Banfai *et al.*, 2012; Derrien *et al.*, 2012; Djebali *et al.*, 2012).

Esse entendimento de que a maior parte dos genomas de eucariotos é transcrita refinou o conceito de dogma central da biologia molecular, inicialmente proposto por Francis Crick em 1958 (Crick, 1958). De acordo com o dogma inicial, o fluxo da informação genética se daria essencialmente do DNA para o RNA e do RNA para proteínas (Crick, 1958). Nesse contexto, pelo fato das proteínas desempenharem papéis cruciais em praticamente todos os processos celulares, o RNA era visto principalmente como uma molécula intermediária (o RNA mensageiro, ou mRNA) que levaria a informação para a síntese de proteínas, enquanto as funções estruturais, catalíticas e regulatórias seriam desempenhadas primordialmente por proteínas.

Com o desenvolvimento de técnicas de biologia molecular nas décadas de 1960 e 1970, passou-se a considerar que outras formas de transferência de informação entre moléculas biológicas eram possíveis (Crick, 1970), tais como de RNA para DNA e mesmo entre proteínas (Gibbons e Hunter, 1967). Além disso, os papéis de outros RNAs, além do mRNA, passaram a ser mais bem definidos, tais como RNAs transportadores (tRNAs), RNAs ribossomais (rRNAs) e pequenos RNAs nucleares e nucleolares (*small nuclear RNAs*, snRNAs e *small nucleolar RNAs*, snoRNAs), que exercem funções estruturais e catalíticas no processamento do mRNA ou na tradução proteica

(Holley *et al.*, 1965; Lerner e Steitz, 1979). Apesar da evolução na compreensão das funções desempenhadas pelos RNAs na célula, o papel funcional dos RNAs como produto dos genes permaneceu por muito tempo focado na produção de proteínas, com algumas poucas exceções.

Assim, foi principalmente nas últimas duas décadas que o estudo dos transcritomas de eucariotos começou a revelar que existe uma grande quantidade e diversidade de RNAs sendo transcrita nas células, o que aumenta consideravelmente a complexidade da transcrição eucariótica. As análises dos transcritomas de *Drosophila melanogaster*, *Caenorhabditis elegans*, levedura e camundongo, por exemplo, mostraram que pelo menos 70% do genoma desses organismos são transcritos (Carninci *et al.*, 2005; David *et al.*, 2006; Amaral e Mattick, 2008; Gerstein *et al.*, 2010; Roy *et al.*, 2010). A maioria dos RNAs que são transcritos não são traduzidos em proteínas e são denominados RNAs não-codificadores de proteínas (ncRNAs). Esses ncRNAs têm sido classificados, para efeito de sistematização, em função do seu tamanho: ncRNAs longos (lncRNAs), quando maiores que 200 nucleotídeos (nt), ou ncRNAs curtos, quando menores que 200 nt (Clark e Mattick, 2011).

Entre os lncRNAs, incluem-se RNAs que possuem de 200 até vários milhares de nucleotídeos, com ou sem *splicing* (Rinn e Chang, 2012; Kung *et al.*, 2013; Ma *et al.*, 2013). Entre os ncRNAs curtos (< 200 nt), estão incluídos os micro-RNAs (miRNAs), snRNAs, snoRNAs, *piwi-interacting* RNAs (piRNAs), pequenos RNAs associados à regiões promotoras, entre outros (Girard *et al.*, 2006; Carthew e Sontheimer, 2009; Fejes-Toth *et al.*, 2009; Taft *et al.*, 2009). Os lncRNAs e miRNAs serão abordados em detalhes nas próximas seções. Acredita-se que esses ncRNAs desempenhem papéis importantes na regulação da expressão gênica transcricional e pós-transcricional, contribuindo para o desenvolvimento, para a diversificação das funções de organismos superiores e até mesmo para a sofisticação da regulação dos processos cognitivos (Mattick, 2011; Sabin *et al.*, 2013).

Nesse contexto, análises de genomas de diversos organismos revelaram que a diversidade biológica de organismos mais complexos não está relacionada diretamente à quantidade de genes codificadores de proteínas (Taft *et al.*, 2007). Estima-se, por exemplo, que o número de genes codificadores de proteínas para eucariotos recentes mais complexos, como humanos e camundongo, é de cerca de 20.000-25.000 (Waterston *et al.*, 2002; Collins *et al.*, 2004). Essa quantidade de genes codificadores de proteínas é bem próxima do número de genes de eucariotos ancestrais menos complexos como *Drosophila* e *C. elegans*, que possuem cerca de 15.000 e 20.000 genes codificadores de proteínas, respectivamente (Consortium, 1998; Adams *et al.*, 2000).

Certamente, a maior complexidade do transcritoma eucariótico responde em parte por esse aumento na sofisticação e na complexidade biológicas de organismos eucariotos, possivelmente determinadas pela diversidade de ncRNAs, capazes de fornecer um repertório maior de moléculas com funções regulatórias que teriam permitido o desenvolvimento dos eucariotos (Mattick, 2007; Taft *et al.*, 2007). De fato, observa-se que o aumento da complexidade biológica está positivamente associado ao aumento da proporção de sequências não-codificadoras de proteínas 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).

Outros fatores contribuem para o aumento da complexidade biológica de organismos eucariotos. Entre esses fatores estão, por exemplo, o 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 proteicas (Nagasaki *et al.*, 2005; Birney *et al.*, 2007; Taft *et al.*, 2007; Farajollahi e Maas, 2010).

A fim de se explorar a complexidade do transcritoma de humanos, teve início na década passada o projeto ENCODE (*Encyclopedia of DNA Elements*). Em sua primeira fase ou projeto-piloto, o ENCODE analisou 1% do genoma humano por meio de diversas técnicas, que incluíram

arrays genômicos de alta densidade e sequenciamento em larga escala de bibliotecas de cDNA, e foi estimado que 90% do genoma humano tinha potencial de ser transcrito (Birney *et al.*, 2007). Ao témino do projeto ENCODE, observou-se que essa taxa era de 74,7% (Consortium *et al.*, 2012; Derrien *et al.*, 2012; Djebali *et al.*, 2012). Além disso, concluiu-se que os transcritos não-codificadores de proteínas estão predominantemente localizados no citoplasma de células humanas e que aproximadamente 6% de todos os transcritos codificadores e não-codificadores de proteínas anotados são prováveis precursores de RNAs curtos (Djebali *et al.*, 2012).

Apesar de todas essas evidências, existe um debate intenso na comunidade científica sobre o significado biológico da transcrição generalizada dos genomas de eucariotos, que é motivado pelo fato de que a maioria dos ncRNAs são expressos em baixos níveis na célula e de que a maior parte desses transcritos ainda não foi caracterizada funcionalmente (Struhl, 2007; van Bakel *et al.*, 2010). A caracterização detalhada dos mecanismos de ação e das funções dos ncRNAs irá revelar a sua real contribuição para o incremento da complexidade do transcritoma de eucariotos. Uma iniciativa que é parte do ENCODE, denominada GENCODE (Harrow *et al.*, 2012), deu início a este processo por meio da anotação atualizada e completa dos genes codificadores de proteínas e ncRNAs de humanos (Derrien *et al.*, 2012).

1.2 RNAs longos não-codificadores de proteínas em eucariotos

Embora a descoberta de lncRNAs específicos seja anterior à descoberta dos miRNAs (Brannan *et al.*, 1990; Brown *et al.*, 1991; Lee *et al.*, 1993)[1], somente recentemente um maior entendimento a respeito dos lncRNAs foi iniciado (Rinn e Chang, 2012; Kung *et al.*, 2013). Os lncRNAs podem ser transcritos a partir de ambas as fitas do genoma humano, incluindo regiões intergênicas e principalmente intrônicas (Derrien *et al.*, 2012; St Laurent *et al.*, 2012), e exibem diferentes sítios de iniciação, terminação, padrões de *splicing* e de expressão em diferentes tipos celulares (Birney *et al.*, 2007; Mattick, 2011). A definição e a nomeação dos lncRNAs ainda estão evoluindo na literatura em função da sua recente identificação e caracterização, e diferentes classes ou categorias funcionais de lncRNAs têm sido descritas (Wright e Bruford, 2011).

Algumas dessas classes são mais estudadas, como a de ncRNAs longos intergênicos (lincRNAs) (Guttman *et al.*, 2009; Khalil *et al.*, 2009; Guttman *et al.*, 2010; Ulitsky e Bartel, 2013), que são relativamente mais fáceis de se caracterizar experimentalmente pois não possuem sobreposição com genes codificadores de proteínas.

Outra classe de lncRNAs, composta por ncRNAs longos intrônicos, possuem sobreposição na mesma orientação (lncRNAs senso) ou na orientação oposta (lncRNAs antissenso) a de introns de genes codificadores de proteínas (Nakaya *et al.*, 2007; Rearick *et al.*, 2011; Chorev e Carmel, 2012; Derrien *et al.*, 2012). Nosso grupo de pesquisa tem focado seu estudo na caracterização dessa classe de lncRNAs (Reis *et al.*, 2004; Louro *et al.*, 2007; Nakaya *et al.*, 2007; Beckedorff *et al.*, 2013). Assim, uma análise utilizando bancos de dados públicos de ESTs (*Expressed Sequence Tags*) e mRNAs revelou que pelo menos 74% dos genes humanos anotados possuem regiões

[1] Os lncRNAs *H19* e *XIST* foram primeiramente identificados no início dos anos 1990, entretanto suas características e funções só foram elucidadas após os anos 2000.

intrônicas capazes de gerar transcritos estáveis e uma análise do potencial codificador indicou que mais de 90% dos possíveis transcritos originados dessas regiões 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 e, com o uso desta plataforma, foram identificadas assinaturas de transcrição intrônica tecido-específicas para amostras de fígado, próstata e rim humanos (Nakaya *et al.*, 2007).

A variedade de classes de lncRNAs novos encontrados tem crescido muito rapidamente. Recentemente, Zhang e colaboradores descreveram a identificação de lncRNAs intrônicos circulares que se acumulam em células humanas (Zhang *et al.*, 2013). Esses lncRNAs intrônicos circulares são abudantes no núcleo das células e não possuem sítios conservados como alvos de miRNAs (Zhang *et al.*, 2013). Os autores mostram que um dos lncRNAs intrônicos circulares se associa com a maquinaria da RNAP II (RNA Polimerase II) e facilita a elongação da transcrição do gene codificador de proteína parental, o que sugere que esse pode ser um mecanismo geral de regulação positiva de mRNAs exercida por essa classe de lncRNAs (Zhang *et al.*, 2013).

Os lncRNAs estão envolvidos na regulação de uma grande variedade de processos biológicos (Nagano e Fraser, 2011; Batista e Chang, 2013), tais como: o desenvolvimento e diferenciação de células T (Hu *et al.*, 2013), a pluripotência (Guttman *et al.*, 2011), o comprimento do telômero (Azzalin *et al.*, 2007), a adipogênese (Sun *et al.*, 2013), a invasividade e metástase (Beltran *et al.*, 2008) e a resposta ao dano do DNA (Huarte *et al.*, 2010; Wan *et al.*, 2013), entre outros.

Um dos processos recentemente identificados, por exemplo, envolve o lncRNA denominado *NeST* (*nettoie Salmonella pas Theiler's [cleanup Salmonella not Theiler's]*), que foi mostrado conferir resistência a infecções virais e diminuir a patogênese induzida pela bactéria

Salmonella enterica em camundongos (Gomez *et al.*, 2013). *NeST* se liga ao compontente do complexo de metilação da lisina 4 da histona H3 denominado WDR5 (*WD repeat-containing protein 5*) e leva à ativação epigenética do gene *IFN-gama* (*interferon-gama*) (Gomez *et al.*, 2013).

Já se sabe que os lncRNAs podem atuar por meio de diversos outros mecanismos moleculares (Clark e Mattick, 2011; Wang e Chang, 2011), tanto por meio da regulação da transcrição quanto através de mecanismos pós-transcricionais. Em relação à regulação da transcrição, é sugerido que o processo de transcrição de lncRNAs por si só, em vez do produto de transcrição, no caso o próprio lncRNA, 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 à enzima RNA polimerase (Dinger *et al.*, 2009; Kornienko *et al.*, 2013). Por outro lado, o processo de transcrição do lncRNA pode prejudicar a transcrição do gene codificador de proteína, como no caso do lncRNA *Airn (antisense Igf2r RNA noncoding)*, cuja sobreposição transcricional com o promotor do gene *Igf2r* (*Insulin-like growth factor 2 receptor*) interfere com o recrutamento da RNAP II na ausência de cromatina repressiva (Latos *et al.*, 2012).

Outra forma de regulação da transcrição de genes codificadores de proteínas por lncRNAs envolve o controle epigenético (Lee, 2012; Mattick, 2012), seja por meio do recrutamento de complexos modificadores da cromatina, seja através da regulação dos níveis de metilação do DNA.

Um dos lncRNAs mais bem caracterizados que age por meio de mecanismo epigenético é o *HOTAIR* (*Hox Transcript Antisense RNA*), que se origina do *locus HOXC* e silencia a transcrição ao longo de 40kb do *locus HOXD*, em *trans*, ao induzir um estado de cromatina repressivo que é proposto ocorrer através do recrutamento do complexo PRC2 (*Polycomb Repressive Complex 2*) (Rinn *et al.*, 2007). O aumento da expressão de *HOTAIR* tem sido associado à metástase (Gupta *et al.*, 2010) e a tumores colorretais, intestinais, pancreáticos e de mama e fígado (Beckedorff *et al.*, 2013).

Recentemente, nosso grupo de pesquisa identificou um lncRNA antissenso *unspliced* denominado *ANRASSF1* (*Antisense Intronic Noncoding RASSF1 RNA*) (Beckedorff *et al.*, 2013). *ANRASSF1* forma um híbrido lncRNA/DNA em células humanas, liga-se ao PRC2 e direciona esse complexo para o promotor do gene codificador de proteína *RASSF1A (RAS-association domain family member 1A*), levando ao silenciamento e à diminuição da expressão somente dessa isoforma do gene *RASSF1* (Beckedorff *et al.*, 2013). Esses achados demonstraram um novo mecanismo de repressão epigenética, no qual um lncRNA antissenso reprime seletivamente a expressão de uma das isoformas do gene *RASSF1*.

Di Ruscio e colaboradores recentemente identificaram um lncRNA não-poliadenilado enriquecido no núcleo de células humanas denominado *ecCEBPA* (*extracoding CEBPA*) (Di Ruscio *et al.*, 2013). *ecCEBPA* se liga à DNMT1 (*DNA (cytosine-5)-methyltransferase 1*) e previne a metilação do locus do gene codificador de proteína *CEBPA*, sugerindo que lncRNAs possam ser alvos de estratégias terapêuticas objetivando uma demetilação de genes específicos (Di Ruscio *et al.*, 2013).

Por outro lado, a habilidade de ncRNAs longos em reconhecer sequências complementares propicia interações altamente específicas que são essenciais em vários passos da regulação pós transcricional de mRNAs, incluindo *splicing* alternativo, edição, transporte, tradução e degradação (Mercer *et al.*, 2009).

Evidências crescentes têm apontado para o envolvimento de lncRNAs em diversas doenças humanas (Taft *et al.*, 2010; Esteller, 2011; Wapinski e Chang, 2011; Lee e Bartolomei, 2013), especialmente o câncer (Prensner e Chinnaiyan, 2011; Nie *et al.*, 2012; Reis e Verjovski-Almeida, 2012; Spizzo *et al.*, 2012; Beckedorff *et al.*, 2013). Por isso, lncRNAs são considerados possíveis alvos em estratégias terapêuticas (Lee e Bartolomei, 2013; Wahlestedt, 2013).

Nesse sentido, Du e colaboradores avaliaram os perfis de expressão de 10.207 lncRNAs em aproximadamente 1.300 tumores humanos de diversos tipos (Du *et al.*, 2013). Foram identificados lncRNAs associados com subtipos de câncer e com prognóstico clínico, e dois desses lncRNAs foram validados como necessários para o crescimento de células tumorais de próstata (Du *et al.*, 2013). Esse estudo gerou, portanto, um catálogo de lncRNAs que são possíveis alvos a serem explorados como biomarcadores ou em terapias anti-tumorais (Du *et al.*, 2013).

O desenvolvimento de estratégias terapêuticas que utilizem lncRNAs como alvos, no entanto, ainda depende de um melhor entendimento de como a expressão e a manutenção dos níveis dos lncRNAs são regulados em processos fisiológicos e patológicos na célula.

1.3 micro-RNAs

Os micro-RNAs (miRNAs) possuem cerca de 21 nt de tamanho e compreendem uma grande família de ncRNAs curtos que atuam como importantes reguladores pós-transcricionais da expressão gênica em metazoários e plantas (Bartel, 2009; Carthew e Sontheimer, 2009). Atualmente estão descritos mais de 2.500 miRNAs maduros em humanos (versão 20 do banco de dados miRBase, junho de 2013) (Kozomara e Griffiths-Jones, 2011) e estudos indicam que mais de 60% dos genes codificadores de proteínas humanos podem ter os seus níveis regulados por miRNAs (Friedman *et al.*, 2009; Krol *et al.*, 2010).

A maioria dos miRNAs apresenta sequências evolutivamente conservadas (Lagos-Quintana *et al.*, 2003) e muitos têm a expressão modulada espacial ou temporalmente (Krichevsky *et al.*, 2003; Kasashima *et al.*, 2004; Liu *et al.*, 2004; Wienholds *et al.*, 2005), participando da regulação de uma grande variedade de processos fisiológicos (Ambros, 2004).

Em animais, os miRNAs são gerados pela via canônica a partir de RNAs precursores que possuem cap-5' e cauda poli-A (pri-miRNAs), gerados pela RNAP II em unidades transcricionais independentes, ou em regiões intrônicas ou exônicas de genes (Lee *et al.*, 2004). Após sua transcrição, o pri-miRNA forma uma estrutura secundária característica em formato de grampo (*hairpin*) que é clivada pelo complexo Microprocessador, composto pelas enzimas nucleares da família RNase III Drosha e a proteína de ligação a RNA dupla-fita DGCR8/Pasha (Han *et al.*, 2004). Após clivagem do pri-miRNA, é gerado um pre-miRNA com cerca de 70 nucleotídeos, que é então exportado para o citoplasma pela Exportina-5 (XPO5) (Yi *et al.*, 2003; Lund *et al.*, 2004). No citoplasma, o pre-miRNA é processado pela enzima Dicer, gerando a forma madura do miRNA, com aproximadamente 21 a 23 nt (Hutvagner *et al.*, 2001; Ketting *et al.*, 2001).

Alternativamente, miRNAs podem ser gerados independentemente da via canônica. Vários eventos em que miRNAs são gerados através desse tipo de processamento têm sido relatados e, na maioria deles, o intermediário precursor é clivado pela enzima Dicer (Babiarz *et al.*, 2008; Bogerd *et al.*, 2010). Em outra forma de geração de miRNAs independente da via canônica, o precursor do miRNA não é clivado pela Dicer, mas sim pela proteína Argonauta2 (Ago2) (Cheloufi *et al.*, 2010; Cifuentes *et al.*, 2010). Os miRNAs também podem ser gerados independentemente da *Drosha*, a partir de regiões intrônicas curtas (*mirtrons*), como um subproduto da maquinaria de *splicing* (Sibley *et al.*, 2012). Recentemente foram descritos *simtrons (splicing-independent mirtron-like miRNAs)*, precursores de miRNAs similares aos *mirtrons* mas que são gerados a partir de introns de maneira dependente de Drosha e independente de DGCR8 ou Dicer (Havens *et al.*, 2012).

De uma maneira geral, em animais, os miRNAs regulam o nível do estado estacionário de seu alvo por meio do pareamento de bases com sequências parcialmente complementares no mRNA alvo, em geral na porção 3'UTR deste, levando à redução da expressão da proteína através de uma variedade de mecanismos que envolvem a desestabilização e a degradação do mRNA

(Huntzinger e Izaurralde, 2011); esse pareamento parcial também pode levar à redução da expressão da proteína pela repressão da tradução (Huntzinger e Izaurralde, 2011). Em plantas, geralmente, o pareamento perfeito dos miRNAs com os seus alvos leva à clivagem e degradação do mRNA (Llave *et al.*, 2002).

Atualmente, no entanto, estudos apontam para um cenário de regulação mais complexo e flexível do que o modelo clássico no qual miRNAs animais reprimem o mRNA via pareamento com a porção 3'UTR do seu alvo (Pasquinelli, 2012). Os dados atuais mostram que miRNAs podem se ligar a uma ampla variedade de alvos (Poliseno *et al.*, 2010), que não existe um padrão único de pareamento do miRNA com o mRNA (Shin *et al.*, 2010; Chi *et al.*, 2012) e que várias regiões da sequência do mRNA podem ser alvejadas, incluindo a 5' UTR e a região codificadora (Grey *et al.*, 2010; Reczko *et al.*, 2012). Em plantas, estudos recentes indicam que miRNAs também podem atuar através do controle da tradução (Yu e Wang, 2010). De maneira não convencional, os miRNAs têm sido, inclusive, relacionados ao aumento da tradução sob determinadas condições específicas em células de mamíferos (Orom *et al.*, 2008).

Os miRNAs medeiam a diminuição dos níveis do estado estacionário do mRNA alvo na célula por meio do recrutamento do *RNA-induced silencing complex* (RISC), um complexo ribonucleoprotéico que contém uma proteína argonauta (Ago) e uma proteína GW182 (*glycine-tryptophan repeat containing protein*) (Fabian e Sonenberg, 2012). Existem quatro membros da subfamília das proteínas argonautas (Ago1-4) que, juntamente com GW182, são responsáveis pelo recrutamento de proteínas que irão propiciar a desestabilização dos RNAs-alvo, tais como deadenilases e exonucleases (Huntzinger e Izaurralde, 2011). Em humanos, somente a proteína Ago2 foi mostrada ter capacidade de clivagem do alvo (Liu *et al.*, 2004; Meister *et al.*, 2004).

Muitos miRNAs desempenham funções importantes e conferem homeostase e robustez a diversos processos biológicos, tais como o controle do ciclo celular, o crescimento e a diferenciação

celulares, a apoptose e o desenvolvimento embrionário (Ebert e Sharp, 2012). Como consequência, os miRNAs estão envolvidos em uma grande variedade de doenças, especialmente distúrbios neurológicos, doenças cardiovasculares e câncer (Esteller, 2011; Mendell e Olson, 2012). Além disso, o uso de miRNAs extracelulares como biomarcadores de doenças tem sido alvo de interesse (Boeri *et al.*, 2011; Di Stefano *et al.*, 2011), dada a sua presença em fluidos corporais tais como soro, plasma, saliva e líquido amniótico (Cortez *et al.*, 2011).

1.4 Regulação da expressão de lncRNAs

Apesar dos recentes avanços na elucidação da estrutura, dos mecanismos de ação e das funções dos lncRNAs, muito pouco se sabe sobre os mecanismos pelos quais os lncRNAs são regulados ou se os níveis dos lncRNAs são controlados da mesma maneira que os níveis dos mRNAs (Geisler *et al.*, 2012). Apenas para alguns lncRNAs específicos, como o *MALAT-1* (*metastasis-associated lung adenocarcinoma transcript 1*), têm-se acumulado informações consideráveis a respeito da sua regulação (Gutschner *et al.*, 2013). O entendimento dos mecanismos de regulação dos lncRNAs será de grande importância, uma vez que os lncRNAs são considerados como possíveis marcadores moleculares ou alvos terapêuticos em diversas doenças humanas (Lee e Bartolomei, 2013).

1.4.1 Regulação transcricional de lncRNAs

Tem sido documentado que a transcrição de alguns lncRNAs ocorre de maneira semelhante à de mRNAs codificadores de proteínas, com a presença de fatores de transcrição na região genômica em que os lncRNAs mapeiam (Birney *et al.*, 2007). Foi mostrado, por exemplo, que lncRNAs conservados são regulados pelo fatores de transcrição Oct4 (*octamer-binding transcription factor 4*) e Nanog (*nanog homeobox*) em mESCs (*mouse embryonic stem cells*) e que essa regulação está envolvida na manutenção do estado de pluripotência dessas células (Sheik Mohamed *et al.*, 2010). Além disso, experimentos de imunoprecipitação da cromatina com anticorpo anti-RNAP II, acompanhados de *tiling arrays* ou sequenciamento de segunda geração, demonstraram que os sítios de ligação da RNAP II e de seus fatores de transcrição estão localizados em regiões de lncRNAs ou em regiões genômicas distantes de promotores canônicos de genes codificadores de proteínas (Cawley *et al.*, 2004; Guttman *et al.*, 2009; Conley e Jordan, 2012). No entanto, nem todos os lncRNAs são transcritos pela RNAP II e uma possibilidade é que eles sejam transcritos pela RNAP III (Dieci et al., 2007; Mattick, 2011).

Também há indícios na literatura de que os lncRNAs possam ser regulados por mecanismos epigenéticos, como modificações de histonas e metilação do DNA em regiões próximas a seus sítios de transcrição. Domínios compostos de marcas de histonas relacionadas a regiões de promotores ativos (H3K4me3, trimetilação da lisina 4 da histona 3) juntamente com marcas de transcrição ativa ao longo do corpo do transcrito (H3K36me3, trimetilação da lisina 36 da histona 3) são utilizados para a identificação e caracterização de diferentes classes de lncRNAs (Guttman *et al.*, 2009; Conley e Jordan, 2012). Além disso, alterações nestas marcas têm sido relacionadas à regulação dos níveis e consequentemente das funções dos lncRNAs durante o processo de diferenciação celular (Wu *et al.*, 2010).

Um estudo recente do grupo do Dr. Eduardo M. Reis, em nosso Departamento, avaliou a contribuição da metilação do DNA na regulação da expressão de lncRNAs intrônicos senso e antissenso por meio do tratamento de linhagens celulares tumorais humanas com o agente desmetilante do DNA 5-AZA (5-Aza-2'-deoxicitidina), seguido de extração do RNA e hibridização em *oligoarray* introns-exons[2]. Os resultados encontrados mostraram que o tratamento com 5-AZA alterou a expressão de um subconjunto de lncRNAs intrônicos. Embora tenha sido observado que a densidade de CGIs (ilhas de dinucleotídeos CpG) na vizinhança de lncRNAs intrônicos seja inferior à encontrada para mRNAs, a análise de reexpressão gênica após tratamento com 5-AZA nas linhagens revelou um subconjunto de 555 lncRNAs intrônicos antissenso e 404 lncRNAs intrônicos senso que têm a sua expressão regulada por metilação do DNA. Esses dados reforçam a noção de que a expressão de lncRNAs está sujeita a regulações transcricionais em células humanas.

Tem aumentado o número de relatos na literatura de lncRNAs regulados por estímulos extracelulares e muitos lncRNAs estão sujeitos à regulação por hormônios. Nosso grupo de pesquisa identificou 39 lncRNAs intrônicos cujos níveis foram significativamente alterados pela exposição de células humanas da linhagem de adenocarcinoma de próstata LNCaP ao hormônio andrógeno. Foi mostrado que este hormônio se liga a elementos regulatórios ARE (*Androgen Response Element*) na região genômica dos lncRNAs, de maneira semelhante à que ocorre com mRNAs (Louro *et al.*, 2007). Além disso, outros grupos também têm relatado a regulação de lncRNAs pelo hormônio andrógeno (Wang *et al.*, 2011; Takayama *et al.*, 2013; Yang *et al.*, 2013).

[2] Camargo, L. Expressão de RNAs não codificadores intrônicos longos em linhagens celulares humanas e o seu controle epigenético por metilação do DNA. (2012). Tese (Doutorado em Bioquímica) – Instituto de Química, Universidade de São Paulo, São Paulo, 2012.

Utilizando a técnica de GRO-Seq (*global nuclear run-on and sequencing*), Hah e colaboradores mostraram que lncRNAs de diversas classes são regulados pelo hormônio estrógeno em células tumorais de mama (Hah *et al.*, 2011), incluindo 686 RNAs intergênicos e 429 RNAs intrônicos antissenso. Apesar das funções desses lncRNAs regulados por estrógeno não terem sido descritas, especula-se que esses lncRNAs devam desempenhar funções antagônicas às funções do estrógeno, uma vez que a maioria desses transcritos teve a sua expressão diminuída após o tratamento das células com esse hormônio (Hah *et al.*, 2011).

Recentemente, centenas de lncRNAs, incluindo 54 pseudogenes, também foram mostrados como regulados pela citocina TNF-alfa (*tumor necrosis fator-alfa*) em células de camundongo expostas a estímulos inflamatórios (Rapicavoli *et al.*, 2013). Foi encontrado um pseudogene denominado *Lethe*, que é induzido por citocinas pró-inflamatórias e que regula negativamente a sinalização por NF-kappaB (*nuclear factor kappa-light-chain-enhancer of activated B cells*) por meio da ligação a este fator e prevenção de sua ligação ao DNA (Rapicavoli *et al.*, 2013).

Na busca por transcritos com possíveis papéis nos processos de aterosclerose e hipertensão, outro trabalho identificou lncRNAs conhecidos e novos regulados pelo peptídeo angiotensina II em células vasculares de músculo liso (Leung *et al.*, 2013). De uma maneira geral, esses trabalhos indicam que os lncRNAs podem ser regulados de modo a atuar como moléculas versáteis que agem por diferentes mecanismos na resposta a estímulos extracelulares.

1.4.2 Regulação pós-transcricional de lncRNAs

Muitos lncRNAs sofrem modificações estruturais após a sua transcrição que são comuns com os mRNAs, o que sugere que mecanismos pós-transcricionais de regulação da expressão de mRNAs também podem ser compartilhados por lncRNAs. Entre essas características, estão a presença da estrutura *cap* na ponta 5' do lncRNA (Geisler *et al.*, 2012) e a possibilidade dos lncRNAs serem poliadenilados (Kapranov *et al.*, 2010) e de sofrerem *splicing* (Cabili *et al.*, 2011). A presença de tais características em lncRNAs indicam que proteínas que reconhecem essas estruturas em mRNAs para regulação do seu metabolismo poderiam também interagir com lncRNAs.

Neste sentido, um estudo recente mapeou em levedura o padrão de interação entre lncRNAs e 13 proteínas classicamente relacionadas ao processamento, exportação e reciclagem de mRNAs (Tuck e Tollervey, 2013). Foram encontrados muitos lncRNAs associados a proteínas envolvidas no controle de qualidade de RNAs, indicando que esses representam lncRNAs instáveis sujeitos à degradação no núcleo (Tuck e Tollervey, 2013). Entretanto, foi encontrado um grupo de lncRNAs que são alvejados pelas proteínas Hrp1 (*nuclear polyadenylated RNA-binding protein 4*) e Nab2 (*nuclear polyadenylated RNA-binding protein 2*), que estão relacionadas à clivagem da extremidade 3', poliadenilação e consequente exportação dos lncRNAs para o citoplasma (Tuck e Tollervey, 2013), indicando o papel dessas proteínas na regulação dos níveis e na localização dos lncRNAs. A proteína HuR (*Hu Antigen R*), reguladora da estabilidade de mRNAs, é capaz de se ligar a lncRNAs, sugerindo que muitos lncRNAs estão sujeitos a mecanismos de regulação por meio da sua estabilização ou desestabilização em células humanas (Lebedeva *et al.*, 2011).

Apesar dessas evidências, outros trabalhos têm relatado que muitos lncRNAs não possuem as mesmas estruturas encontradas em mRNAs. Estudos iniciados na década passada utilizando

tiling arrays para investigar a presença de poliadenilação em transcritos mostraram que 19,4%, 43,7% e 36,9% dos transcritos identificados em células humanas foram considerados poliadenilados, não-poliadenilados e bimórficos, respectivamente (Cheng *et al.*, 2005). Mais recentemente, estudos utilizando sequenciamento em larga escala também relataram que muitos transcritos não-poliadenilados correspondem a RNAs ainda não conhecidos que mapeiam em regiões intrônicas e intergênicas do genoma humano (Wu *et al.*, 2008; Yang *et al.*, 2011). Para esses lncRNAs, outros motivos ou estruturas secundárias presentes em suas sequências devem contribuir para a regulação dos seus níveis de expressão (Tuck e Tollervey, 2013).

Modificações na estrutura de tRNAs e rRNAs, tais como metilações em citosinas ou guanosinas, são conhecidas como reguladoras da expressão e função destes RNAs (Motorin *et al.*, 2010; Cantara *et al.*, 2011). Recentemente, Amort e colaboradores mostraram que os lncRNAs *HOTAIR* e *XIST* (*X-inactive specific transcript*) também são metilados em suas citosinas em regiões funcionalmente importantes para estes lncRNAs, necessárias para a sua associação com complexos modificadoras da cromatina (Amort *et al.*, 2013). Os autores observaram que a metilação da citosina na região A do lncRNA *XIST* afeta a sua ligação ao complexo modificador da cromatina PRC2, sugerindo que a metilação de citosinas em lncRNAs pode representar um mecanismo comum de regulação da sua função (Amort *et al.*, 2013).

A localização sub-celular de lncRNAs também pode representar um mecanismo pós-transcricional de controle dos seus níveis de expressão. Por um lado, a necessidade de localização dos lncRNAs em regiões definidas das células tem sido associada ao desempenho de suas funções em diferentes domínios sub-celulares (Clark e Mattick, 2011) . Entretanto, é possível que esta localização específica também seja um mecanismo celular de regulação da expressão dos lncRNAs (Djebali *et al.*, 2012), como já bem documentado para RNAs codificadores de proteínas (Lecuyer *et al.*, 2007; Besse e Ephrussi, 2008; Martin e Ephrussi, 2009).

No caso dos mRNAs, essa localização específica está relacionada à expressão temporal destes RNAs na célula e já foram documentados motivos nos mRNAs (fatores em *cis*) que são reconhecidos por proteínas de ligação a RNA (fatores em *trans*) (Martin e Ephrussi, 2009). Um estudo revelou que durante o desenvolvimento embrionário de *Drosophila*, 71% dos genes expressos avaliados apresentaram um padrão de localização sub-celular bem definido (Lecuyer *et al.*, 2007). Os mRNAs foram agrupados em cerca de 35 categorias de acordo com a sua localização, as quais incluem RNAs localizados na maquinaria de divisão celular, nos pólos embriônicos, no núcleo e no citoesqueleto (Lecuyer *et al.*, 2007).

No caso de lncRNAs, ainda são pouco conhecidos os fatores que determinam a sua localização sub-celular e qual a influência destes fatores na regulação dos seus níveis de expressão, especialmente porque a maioria do estudos tem focado na descrição da localização de lncRNAs individuais (Clark e Mattick, 2011).

O lncRNA *MALAT-1*, por exemplo, 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*) (Hutchinson *et al.*, 2007). Esse lncRNA foi originalmente descoberto como um fator prognóstico de sobrevivência de pacientes com adenocarcinoma de grau I ou com carcinoma de células escamosas (Ji *et al.*, 2003; Hutchinson *et al.*, 2007). *MALAT-1* tem expressão aumentada em diversos tipos de câncer (Gutschner *et al.*, 2013) e é essencial para a modulação da interação dos *speckles* nucleares com fatores de *splicing* (Hutchinson *et al.*, 2007). Além disso, esse lncRNA está envolvido na regulação do *splicing* alternativo de um subconjunto de transcritos (Tripathi *et al.*, 2010) e do ciclo celular, por meio do controle da expressão do fator de transcrição B-MYB (Tripathi *et al.*, 2013).

Outro lncRNA, denominado *NEAT1* (*Nuclear Enriched Abundant Transcript 1*) ou MEN-epsilon/beta (*Multiple Endocrine Neoplasia epsilon/beta*), é um componente essencial para a formação e manutenção das estruturas sub-celulares denominadas *paraspeckles* (Clemson *et al.*, 2009; Sasaki *et al.*, 2009; Sunwoo *et al.*, 2009). Os *paraspeckles* estão envolvidos na retenção nuclear de RNAs com edições de adenosina à inosina (A→I) (Bond e Fox, 2009).

Um maior entendimento dos fatores que determinam a localização sub-celular e consequentemente a regulação dos níveis dos lncRNAs depende de estudos que avaliem o padrão de distribuição de classes distintas de lncRNAs em diferentes tecidos e localizações sub-celulares bem definidas. Neste sentido, Mercer e colaboradores mostraram por meio de hibridização *in situ* que a maioria dos 88 ncRNAs expressos em células de Purkinje, no cerebelo de camundongo, apresentou evidência de uma localização sub-celular bem definida (Mercer *et al.*, 2008). Dos 88 ncRNAs expressos nas células de Purkinje, 25 (29%) apresentaram um padrão de localização nuclear difuso, 54 (61%) apresentaram um padrão semelhante a *speckles* e 9 (10%) se mostraram localizados ao longo do corpo celular (Mercer *et al.*, 2008), sugerindo expressão regulada dos ncRNAs.

O grupo de pesquisa do Dr. Eduardo M. Reis, em nosso Departamento, estudou a localização sub-celular de lncRNAs intrônicos por meio do fracionamento sub-celular de lisado de células HeLa seguido de extração do RNA de diferentes compartimentos e hibridização em *oligoarray* introns-exons[3]. Os resultados encontrados demonstraram que uma fração considerável dos lncRNAs intrônicos interrogados nas análises (maior que 50%), mostrou-se igualmente

[3]Ayupe, A. C. "Biogênese, estabilidade e localização sub-celular de RNAs não-codificadores longos expressos em regiões intrônicas do genoma humano." (2012). Tese (Doutorado em Bioquímica) – Instituto de Química, Universidade de São Paulo, São Paulo, 2012.

distribuída entre o núcleo e citoplasma. Esses dados indicam que lncRNAs, embora não codifiquem para proteínas, sofrem algum tipo de regulação para que sejam exportados para o citoplasma e exerçam alguma função biológica nesse compartimento.

Outra possível forma de regulação pós-transcricional é aquela em que lncRNAs são alvejados e silenciados por miRNAs. Nesse sentido, duas ferramentas de predição de alvos de miRNAs publicadas recentemente passaram a incorporar algoritmos que buscam em lncRNAs possíveis sítios de ligação de miRNAs (Jeggari *et al.*, 2012; Paraskevopoulou *et al.*, 2013). No entanto, o número de falso-positivos em predições computacionais de alvos de miRNAs é alto, e a regulação exercida pelo miRNA deve ser avaliada e validada experimentalmente (Chi *et al.*, 2009). Essas validações podem ser feitas para um grupo pequeno de miRNAs e alvos selecionados, mas como um miRNA pode regular vários alvos e um gene pode ser regulado por vários miRNAs (Pasquinelli, 2012), estudos de identificação de alvos de miRNAs em larga escala, especialmente aqueles capazes de identificar interações diretas entre miRNAs e alvos, têm sido propostos (Arvey *et al.*, 2010).

Ainda não se sabe se lncRNAs são alvejados por miRNAs de maneira disseminada para controle de suas funções na célula, pois apenas um pequeno número de estudos validaram interações entre miRNAs e lncRNAs. No mecanismo mais caracterizado de interação miRNA-lncRNA, os lncRNAs atuam como esponjas naturais (Ebert e Sharp, 2010). Neste mecanismo, um lncRNA compete com um mRNA por miRNAs que o alvejam. Ao competirem pelos mesmos miRNAs, os níveis de transcritos codificadores de proteínas aumentam, porque o lncRNA possui sítios de ligação para miRNAs semelhantes ou idênticos aos sítios de ligação do mRNA em questão (Franco-Zorrilla *et al.*, 2007; Cazalla *et al.*, 2010; Poliseno *et al.*, 2010; Cesana *et al.*, 2011; Hansen *et al.*, 2013; Memczak *et al.*, 2013).

Recentemente, dois grupos identificaram RNAs circulares que agem como esponjas para miRNAs (Hansen *et al.*, 2013; Memczak *et al.*, 2013). Hansen e colaboradores (Hansen *et al.*, 2013) mostraram que um RNA circular denominado *CDR1 NAT (Natural Antisense Transcript),* previamente identificado pelo mesmo grupo (Hansen *et al.*, 2011), age como uma esponja para o miR-7, suprimindo a atividade deste miRNA, e que o RNA circular *Sry* (*sex-determining region Y*) também age como uma esponja para o miR-138. Além de identificar o *CDR1 NAT* e seu papel regulador do miR-7, (Memczak *et al.*, 2013) identificaram milhares de outros RNAs circulares em humanos, em camundongos e em *C. elegans.* Estes RNAs têm potencial de atuarem como esponjas, uma vez que suas sequências são enriquecidas em sítios complementares a miRNAs, sugerindo que o efeito esponja realizado por RNAs circulares pode representar um fenômeno geral de regulação dos níveis de miRNAs.

Em outro mecanismo menos caracterizado, a regulação do lncRNA pelo miRNA ocorre de forma indireta: o miR-29 regula o nível dos mRNAs de proteínas DNA metiltransferases, que por sua vez regulam o lincRNA *MEG3* em câncer hepatocelular (Braconi *et al.*, 2011).

Permanecem pouco caracterizados os mecanismos de regulação direta por miRNAs dos níveis de lncRNAs que exercem funções já descritas, diferentes da função esponja. Hansen e colaboradores já haviam descrito que o *CDR1 NAT* é regulado pelo miR-671, levando à clivagem deste *NAT* de uma maneira dependente da proteína Ago2 (Hansen *et al.*, 2011), com concomitante diminuição nos níveis do mRNA de *CDR1*. Entretanto, não foi mostrada a implicação funcional desta regulação nem o mecanismo de controle do mRNA de *CDR1* pelo *CDR1 NAT*. Como não foram encontrados sítios complementares ao miR-671 na região 3’ UTR do mRNA de *CDR1*, foi sugerido que *CDR1 NAT* não atua por meio de um mecanismo de esponja para este miRNA (Hansen *et al.*, 2011), mas que o *CDR1 NAT* poderia conferir estabilidade ao mRNA de *CDR1* através do pareamento de bases direto, conforme já observado (Faghihi *et al.*, 2008).

Será importante identificar lncRNAs que exercem funções já identificadas, tais como o recrutamento de proteínas modificadoras da cromatina ou o recrutamento de proteínas envolvidas em *splicing* (Wang e Chang, 2011; Beckedorff *et al.*, 2013) e que sejam alvejados por miRNAs. Além disso, dada a participação de lncRNAs em processos fisiológicos e patológicos e o desenvolvimento de estratégias terapêuticas baseadas em miRNAs (Kasinski and Slack 2011), será importante caracterizar como os miRNAs eventualmente regulam os níveis de lncRNAs.

2. OBJETIVOS

2.1 Objetivo Geral

Investigar a possível regulação de RNAs longos não-codificadores de proteínas (lncRNAs) por micro-RNAs (miRNAs) na linhagem celular humana HeLa, a fim de avançar o conhecimento sobre a regulação do metabolismo de lncRNAs.

2.2 Objetivos Específicos

1. Identificar lncRNAs que interajam com o complexo de silenciamento induzido por RNA (RISC) por meio de imunoprecipitação de RISC na linhagem celular HeLa, seguida de isolamento dos RNAs co-imunoprecipitados e sequenciamento destes em larga escala.
2. Desenvolver um método de geração de bibliotecas de cDNA direcionadas que permita o sequenciamento fita-específico na plataforma 454/Roche e a identificação de lncRNAs ligados ao complexo RISC.
3. Identificar miRNAs ligados ao complexo RISC na linhagem celular HeLa e que possuam sítios alvo preditos computacionalmente em lncRNAs, sendo portanto possíveis reguladores dos respectivos lncRNAs.
4. Avaliar a possível regulação de lncRNAs por miRNAs por meio da superexpressão de miRNAs selecionados seguida da medida dos níveis dos lncRNAs candidatos a alvo.

3. MATERIAIS E MÉTODOS

3.1 Linhagem celular e condições de cultivo

Para a realização deste trabalho, foi utilizada a linhagem celular humana de adenocarcinoma cervical HeLa (código ATCC - *American Type Culture Collection* - CCL-2). Esta linhagem é amplamente utilizada na comunidade científica e tem sido usada como modelo no estudo das relações entre miRNAs, proteínas argonauta e RNAs-alvo (Janas *et al.*, 2012). A linhagem foi adquirida da ATCC e cultivada a 37ºC em atmosfera de 5% de CO_2 em meio de cultura apropriado, conforme recomendações da ATCC. O meio recomendado e utilizado foi DMEM (*Dulbecco's Modified Eagle's Medium*) (Cultilab), suplementado com SFB (Soro Fetal Bovino) (Cultilab) 10%, bicarbonato de sódio 1,2 g/L, glicose 4,2 g/L e com os antibióticos penicilina (100 U/mL) e estreptomicina (100 µg/mL). Ao atingirem 60-80% da densidade de saturação, as células foram subcultivadas ou utilizadas para realização de manipulações específicas de cada experimento. Para a manutenção em cultura, as células foram lavadas com solução salina PBS (*Phosphate Buffered Saline*) (NaCl 140 mM, KCl 2,7 mM, Na_2HPO_4 8 mM e KH_2PO_4 1,5 mM, pH 7,2) ao atingirem aproximadamente 80% da densidade de saturação e subcultivadas após o uso de solução de tripsina 0,1% (código número 000012, VitroCell). Os estoques celulares foram armazenados em *freezer* a -80ºC, nos quais as células foram mantidas no meio de cultivo com 10% de DMSO (dimetilsulfóxido) (Sigma).

3.2 Imunoprecipitação nativa de complexos ribonucleoprotéicos – RIP

Para cada experimento de imunoprecipitação de complexos ribonucleoproteicos (RIP-*RNA-binding protein immunoprecipitation*), células HeLa foram subcultivadas em duas garrafas

de 175 cm^2 (Sarstedt) cada, até a densidade de saturação de 80%, quando então foram lavadas com PBS e descoladas da superfície da garrafa por tratamento por 5 minutos com 0,1% de tripsina, seguido da adição de meio de cultura DMEM contendo SFB. As células foram transferidas para tubos *falcon* (TPP), uma alíquota foi retirada de cada tubo *falcon* e as células foram contadas em uma câmara de Neubauer. Em seguida, uma alíquota de volume de meio contendo $2x10^7$ células HeLa de cada *falcon* foi centrifugada separadamente a 300 x g por 7 minutos e o sobrenadante foi removido. Essas alíquotas foram usadas posteriormente nas imunoprecipitações com o anticorpo específico ou com anticorpo controle. Após descarte do sobrenadante, os sedimentos das células foram lavados com PBS, as células foram novamente centrifugadas a 300 x g por 7 minutos e o sobrenadante foi descartado. A partir desse passo, foi utilizado o *Magna RIP RNA-Binding Protein Immunoprecipitation Kit* (17-700; Millipore), com algumas modificações, conforme descrito a seguir. Os sedimentos das células foram suspendidos em 200 µL de tampão de lise contendo 1 µL de coquetel de inibição de proteases e 0,5 µL de inibidor de RNase, mantidos por 5 minutos no gelo e então congelados em *freezer* a -80ºC por pelo menos 72 horas, mas por não mais que 30 dias. Em seguida, foram seguidas as recomendações do fabricante: 5 µg de anticorpo anti-Ago2 (03-110; Millipore) ou 5 µg de anticorpo IgG de camundongo não imunizado (12-371; Millipore) foram adicionados a 50 µL de esferas magnéticas que contêm proteínas A/G; lavou-se o excesso de anticorpo duas vezes com 500 µL de *RIP Wash Buffer* e, em seguida, o lisado celular previamente preparado foi adicionado ao complexo esfera-anticorpo. Antes da incubação, foi retirada uma alíquota do lisado, que foi denominada "Proteína *Input*" e foi usada posteriormente no experimento de *western blot*, conforme descrito no item 3.5. Após incubação por 14 horas a 4ºC, foram realizadas seis lavagens do complexo esfera-anticorpo-lisado celular com o *RIP Wash Buffer*. Após as lavagens, foi retirada uma alíquota de 10% (50 µL) do volume final de 500 µL da amostra para confirmação da imunoprecipitação da proteína Ago2 por *western blot* (conforme

descrito no item 3.5) e em seguida as proteínas foram degradadas com proteinase K por 30 minutos a 55ºC. O RNA foi extraído e purificado conforme descrito no item 3.3. Para otimização deste protocolo, foram testados a quantidade de células utilizadas, a concentração de anticorpo primário ligado às esferas e o tempo de incubação do lisado celular com o complexo esfera-anticorpo. Após os testes, as condições descritas acima levaram a uma maior eficiência de imunoprecipitação da proteína Ago2 e a uma maior recuperação de RNAs co-imunoprecipitados.

Os RNAs obtidos foram utilizados para a geração de bibliotecas de cDNA para sequenciamento na plataforma 454/Roche, conforme descrito no item 3.6, para validação dos transcritos detectados nestes sequenciamentos por PCR quantitativo em tempo real (qPCR), conforme descrito no item 3.13.2, ou para sequenciamento de RNAs curtos, conforme descrito no item 3.8.

3.3 Extração de RNA, purificação e tratamento com DNase

Os RNAs co-imunoprecipitados no item 3.2 foram extraídos por meio da utilização de fenol:clorofórmio:álcool isoamílico seguida de clorofórmio, e precipitados com etanol 100% a -80°C por 16 horas, de acordo com recomendações do *Magna RIP RNA-Binding Protein Immunoprecipitation Kit* (17-700; Millipore).

Para garantir a ausência de DNA genômico nuclear ou mitocondrial contaminantes entre os RNAs co-imunoprecipitados, estes RNAs foram tratados com 2 unidades/µg de RNA de TURBO™ DNase (AM2238; Ambion) por 45 minutos a 37ºC. Essa enzima tem uma maior afinidade por DNA do que a DNase I, comumente usada para remoção de DNA de amostras de RNA. A remoção completa de DNA é especialmente importante, uma vez que os RNAs co-imunoprecipitados foram em seguida usados para geração de bibliotecas de cDNA para

sequenciamento em larga escala. A avaliação da ausência de DNA genômico contaminante entre os RNAs foi feita conforme descrito no item 3.4. Após tratamento com TURBO™ DNase, os RNAs co-imunoprecipitados foram purificados utilizando o *miRNeasy Mini Kit* (217004; Qiagen). Foram seguidas as recomendações do fabricante, de modo a preservar os RNAs curtos (entre 18 e 200 nt) juntamente com RNAs longos (maiores que 200 nt).

Para a extração dos RNAs totais nos experimentos de super-expressão de miRNAs (conforme descrito no item 3.14), células HeLa foram transfectadas por 24 horas com diferentes *miRNA mimics* ou controles, lavadas com PBS e descoladas da superfície da garrafa por tratamento por 5 minutos com 0,1% de tripsina seguido da adição de meio de cultura contendo SFB. Em seguida, as células foram sedimentadas por centrifugação a 300 x g por 5 minutos e o sobrenadante foi removido. Os sedimentos das células foram lavados com PBS, as células foram novamente centrifugadas a 300 x g por 5 minutos e o sobrenadante foi descartado. Os sedimentos de células foram mantidos a -80ºC até a extração do RNA. O isolamento do RNA foi realizado utilizando TRIzol (15596-026; Life Technologies), 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 foi submetido a uma etapa de tratamento com TURBO™ DNase (AM2238;Ambion), conforme descrito acima para os RNAs co-imunoprecipitados. Os RNAs foram purificados utilizando o *miRNeasy Mini Kit* (217004, Qiagen), de acordo com o protocolo do fabricante e armazenados a -80ºC.

3.4 Avaliação da qualidade do RNA

Para certificar a ausência de DNA genômico nas amostras de RNAs co-imunoprecipitados ou de RNA total, foi realizada uma PCR (em 50 µL de reação, com 40 ciclos) utilizando-se *primers* que anelam no retrogene de múltiplas cópias da histona H3. Na ausência de DNA contaminante na amostra de RNA, esta PCR nas condições descritas abaixo não gera produto, pois não é feita transcrição reversa. A PCR gera *amplicons* com cerca de 680 pares de bases quando está presente DNA genômico contaminante no RNA. Foram utilizados 5 ng de cada RNA co-imunoprecipitado ou 1 µg de RNA total de cada amostra, sem transcrição reversa, como molde para a reação de PCR (conforme será descrito no item 3.13.1). Somente os RNAs que não apresentavam contaminação com DNA genômico foram utilizados para os ensaios.

Os RNAs co-imunoprecipitados foram quantificados utilizando o reagente RiboGreen (R11490; Invitrogen) conforme protocolo do fabricante. O perfil eletroforético dos RNAs co-imunoprecipitados foi avaliado utilizando-se o equipamento Bioanalyzer 2100 (Agilent Technologies), que realiza uma eletroforese capilar de alta tensão. Foram utilizados os *kits RNA 6000 Pico Kit* (5067-1513; Agilent Technologies) e *Small RNA Kit* (5067-1548; Agilent Technologies).

As amostras de RNA total 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 integridade no equipamento Bioanalyzer 2100 (Agilent Technologies) utilizando o *RNA 6000 Nano Kit* (5067-1511; Agilent Technologies). A integridade das amostras foi avaliada com o programa 2100 Expert (Agilent Technologies). Esse programa atribui um valor de integridade para o RNA total *(RNA Integrity Number*, RIN), que permite uma estimativa da integridade por meio da análise das intensidades de

todo o traçado eletroforético da amostra, ou seja, dos RNAs de todos os tamanhos, e não apenas da razão entre as intensidades nos picos das bandas dos RNAs ribossomais 18S e 28S. Apenas amostras que tinham uma boa integridade (RIN > 8) foram utilizadas em todos os experimentos.

3.5 Western Blot

A alíquota denominada "Proteína *Input*" e as alíquotas retiradas após imunoprecipitação com anticorpo Anti-Ago2 ou com anticorpo IgG normal de camundongo não-imunizado (item 3.2) foram separadas por SDS-PAGE com 10% de poliacrilamida, ao lado do padrão de massa molecular de proteínas *Prestained Protein Marker* (SM#1811; Fermentas) e posteriormente transferidas para uma membrana de nitrocelulose Hybond-C Extra (Amersham Biosciences) utilizando o transferidor semi-seco TE 77 PWR (Amersham Biosciences).

As membranas foram bloqueadas em 10 mL de solução de bloqueio (400 mg BSA /10 mL PBS; Sigma) durante 1 hora. Em seguida, foram incubadas com o anticorpo primário anti-Ago2 (MABE253, Millipore; na diluição 1:500) em uma nova solução de bloqueio por 16 horas a 4ºC, sendo, a seguir, lavadas com PBS Tween 0,1%, duas vezes por 15 minutos e duas vezes por 5 minutos. Após as lavagens, as membranas foram incubadas com o anticorpo secundário anti-IgG de rato IRDye 800CW (926-32219; LI-COR Biosciences) por 1 hora e, em seguida, as lavagens descritas acima foram repetidas. O sinal de fluorescência foi analisado com o *Odyssey Infrared Imaging System* (LI-COR Biosciences) e as intensidades foram quantificadas com o *Odyssey Software Aplicativo v3* (LI-COR Biosciences).

3.6 Desenvolvimento de método de geração de bibliotecas fita-específicas para sequenciamento de RNAs longos no 454/Roche

Os RNAs obtidos após imunoprecipitação de Ago2 ou imunoprecipitação controle (item 3.2) foram usados para geração de bibliotecas de cDNA para sequenciamento de RNAs longos na plataforma 454/Roche. Não existe um *kit* comercial nem metodologia estabelecida para geração de bibliotecas direcionadas ou fita-específicas a serem sequenciadas na plataforma 454/Roche, ou seja, bibliotecas que sejam marcadas assimetricamente em suas extremidades durante a sua geração, de modo a permitirem saber qual fita do cDNA será sequenciada. Bibliotecas direcionadas permitem identificar, após o sequenciamento, a partir de qual fita do DNA genômico o RNA tenha sido transcrito. Essa informação é essencial para determinação da orientação dos RNAs, incluindo lncRNAs que são transcritos dentro de *loci* de genes codificadores de proteínas (lncRNAs que mapeiam na orientação senso ou antissenso em relação ao mRNA de um determinado *locus*). Portanto, para geração de bibliotecas de cDNA direcionadas para sequenciamento no 454/Roche, foi desenvolvido um método, cujas etapas estão representadas na figura 1, a seguir.

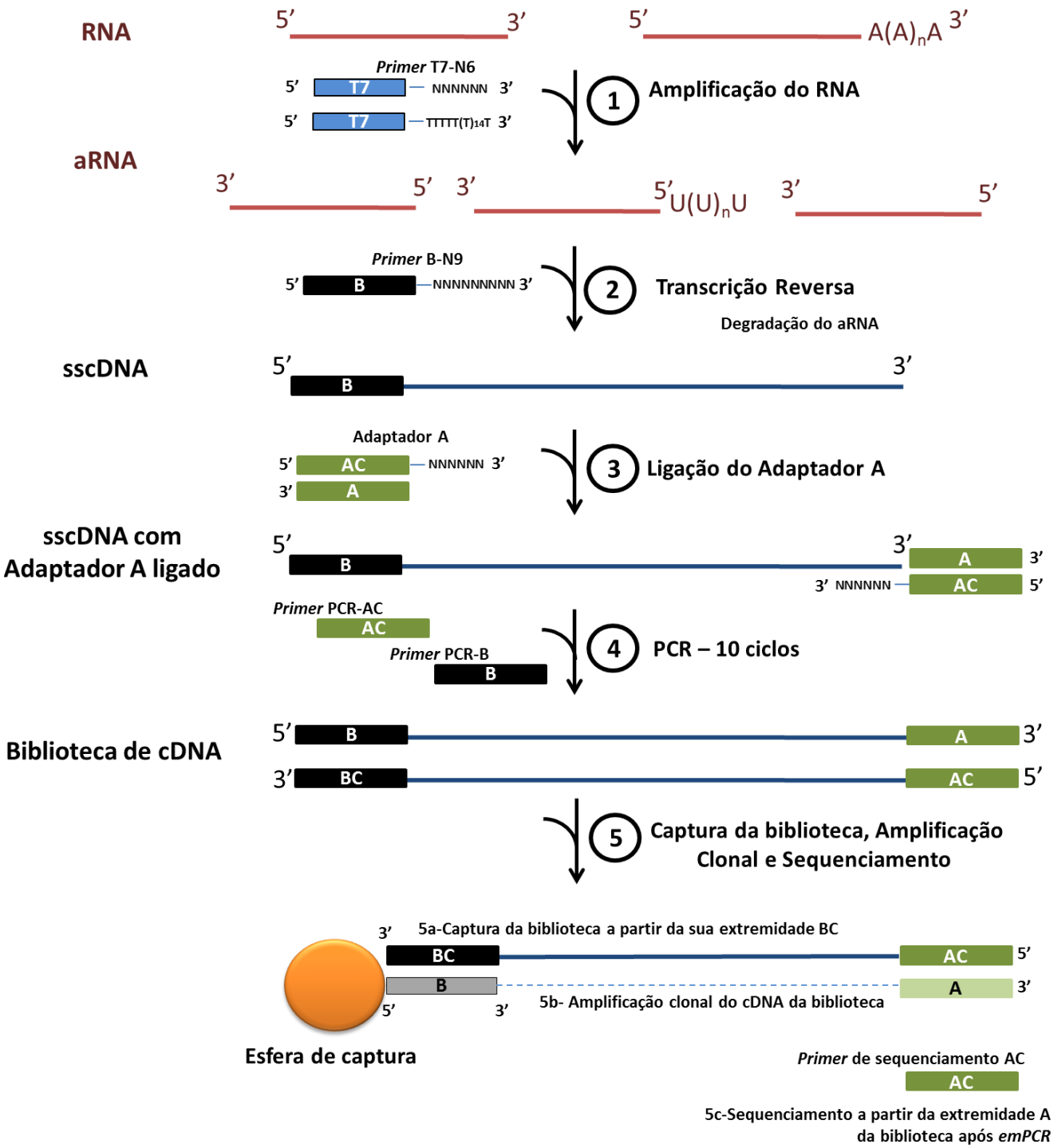

Figura 1: Desenho esquemático representando o método desenvolvido para geração de bibliotecas de cDNA direcionadas para sequenciamento na plataforma 454/Roche. A construção da biblioteca de cDNA começa essencialmente na Etapa 2, na qual são necessários 2 µg de RNA. Caso a quantidade de RNA disponível seja menor do que 2 µg, faz-se a amplificação do RNA utilizando a enzima T7-RNA polimerase (Etapa 1). Este é o caso, por exemplo, do RNA obtido por imunoprecipitação do RISC;

os RNAs longos, contendo ou não caudas poli(A) de diferentes tamanhos, foram extraídos e purificados a partir do imunoprecipitado, e levados para a Etapa 1. **Etapa 1.** Os RNAs (linhas vermelhas) são inicialmente amplificados com *primers* que contém a sequência T7 para reconhecimento pela T7 RNA polimerase. São usados *primers* randômicos, além do *primer* oligo(dT), na reação de transcrição reversa para que RNAs que já tenham sido deadenilados também sejam usados como molde. Ao final da amplificação, os RNAs amplificados (aRNAs) têm a sua orientação invertida em relação ao RNA original, e neste caso os cDNAs serão sequenciados a partir da extremidade correspondente ao 3' do RNA original, como se vê nesta figura. Caso haja quantidade suficiente de RNA (2 ug) ele é levado diretamente para a Etapa 2, e neste caso os cDNAs serão sequenciados a partir da extremidade correspondente ao 5' do RNA original (não mostrado na figura). **Etapa 2.** Os aRNAs são utilizados em uma reação de transcrição reversa realizada com um *primer* randômico de nove bases acoplado à sequência correspondente ao adaptador B da tecnologia 454/Roche para geração de cDNAs simples fita (*single-stranded cDNAs*, ou *sscDNAs*) com o adaptador B em sua extremidade 5'. Esta sequência do adaptador B é utilizada mais adiante, para pareamento das moléculas de cDNA ao oligunocleotídeo presente nas esferas durante o processo de captura dos cDNAs (etapa 5a), e subsequente etapa de amplificação clonal do protocolo 454/Roche (etapa 5b). **Etapa 3.** Às extremidades 3' dos sscDNAs é ligado o Adaptador A, que contém a sequência de nucleotídeos que será utilizada para hibridização do *primer* da PCR do protocolo do 454/Roche durante o processo de amplificação clonal por PCR em emulsão, e durante o sequenciamento (como mostrado na etapa 5c). **Etapa 4.** Os sscDNAs com adaptador A ligado são amplificados em uma PCR de dez ciclos com *primers* B e AC (A-Complementar), para garantir que todas as moléculas de cDNA contenham os dois adaptadores nas duas extremidades. **Etapa 5**: Segundo o protocolo do 454/Roche, a biblioteca de cDNA gerada na etapa 4 é pareada às esferas de captura a partir da sua extremidade BC (B-Complementar) (etapa 5a). A proporção de esferas e cDNA é tal que apenas uma molécula de cDNA é ligada por esfera, e sofre amplificação clonal em uma PCR em emulsão (*emPCR*). A geração da primeira fita de DNA (linha azul tracejada) que permanecerá ligada covalentemente às esferas, que é o primeiro passo da *emPCR*, está representada esquematicamente na etapa 5b. No total, são feitos 50 ciclos de amplificação na *emPCR*, o que resulta no recobrimento de cada esfera com aproximadamente um trilhão de cópias do cDNA original. As fitas de cDNA (linha azul tracejada) que permanecem ligadas covalentemente às esferas após a *emPCR* possuem a sequência A na sua extremidade 3'. Estas fitas são posteriormente sequenciadas a partir das suas extremidades A (etapa 5c), com o *primer* do sequenciamento cuja sequência corresponde à do *primer* AC. Cada esfera fica contida em um poço da placa de sequenciamento, e o instrumento coleta a imagem simultânea de aproximadamente 1 milhão de poços, medindo a quimioluminescência gerada cada vez que um nucleotídeo é incorporado.

Antes de serem usados na preparação de bibliotecas de cDNA, 40 ng de cada RNA co-imunoprecipitado foram amplificados utilizando o *Message Amp II aRNA Kit* (AM1751; Ambion) – passo 1 da figura 1. Foram seguidas as recomendações do fabricante com as seguintes adaptações: durante a síntese da primeira fita de cDNA, além do *primer* T7-dT foi adicionado à reação um *primer* que consiste em 6 bases randômicas na extremidade 3', contendo na extremidade 5' a sequência correspondente ao promotor reconhecido pela RNA polimerase T7 (*primer* T7-N6, Tabela 1, concentração final na reação de 2,5 ng/µL). A realização da transcrição reversa com *primer* randômico é essencial, pois além de muitos lncRNAs não serem poliadenilados (Yang *et al.*, 2011), no momento da captura dos RNAs ligados a Ago2, que são o objeto de estudo deste trabalho, muitos destes RNAs já podem ter sido deadenilados. Foi utilizado um tempo de incubação durante a transcrição *in vitro* de 14 horas e somente uma etapa de amplificação.

Tabela 1: Sequências dos *primers* utilizados na geração de bibliotecas de cDNA fita-específicas para sequenciamento no 454/Roche. * representam ligações fosforotioato, adicionadas para aumentar a estabilidade do *primer*. Sublinhada está destacada a sequência da *key* da tecnologia do 454/Roche.

Nome do oligonucleotídeo iniciador (*primer*)	**Sequência do oligonucleotídeo iniciador (*primer*)**
Amplificação do RNA co-imunoprecipitado	
T7-N6	TAATACGACT CACTATAGGGNNNNNN
Geração de bibliotecas de cDNA fita-específicas	
B-N9	C*C*T*A*TCCCCTGTGTGCCTTGGCAGTCTCAGNNNNN*N*N*N*N
sscDNA Oligo A-Prime	C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGNN*N*N*N*N
sscDNA Oligo A	C*T*G*A*GTCGGAGACACGCAGGGATGAG*A*T*G*G
PCR-AC	C*C*A*T*CTCATCCCTGCGTGTCTC*C*G*A*C
PCR-B	C*C*T*A*TCCCCTGTGTGCCTTGGC*A*G*T*C

Dois microgramas dos aRNAs (RNAs amplificados) obtidos foram submetidos à transcrição reversa - passo 2 da figura 1- utilizado um *primer* (*primer* B-N9, Tabela 1, concentração final na reação de 2,5ng/μL) que consiste em 9 bases randômicas na extremidade 3' mais uma sequência de 30 bases na extremidade 5' que será usada para anelamento à sequência do oligunocleotídeo ligado covalentemente às esferas durante o processo de amplificação clonal (PCR em emulsão feita para aumentar a quantidade de sequências na superfície de esferas e permitir o sequenciamento no equipamento 454/Roche). Para transcrição reversa, foi utilizado o kit *SuperScript III First Strand Synthesis Mix* (18080-400; Invitrogen) em um volume final de 20 μL, segundo as recomendações do fabricante para transcrição reversa com *primer* randômico. Uma pré-incubação de 3 minutos a 85ºC foi introduzida no protocolo, antes da fase de anelamento, para desnaturar completamente o RNA.

A reação de transcrição reversa foi tratada com 4 unidades de RNase H (18021-071; Invitrogen) e 1 μg/mL de RNase A (R4875; Sigma) por 20 minutos a 37ºC e por 5 minutos a 85ºC (etapa não mostrada na figura 1). Em seguida, a primeira fita de cDNA gerada foi purificada utilizando o reagente RNAClean XP (A63987; Agencourt) com um volume correspondente a 1,8 vezes o volume de reação de transcrição reversa e seguindo as recomendações do fabricante. A primeira fita dos cDNAs foi quantificada utilizando o reagente RiboGreen (R11490; Invitrogen) conforme protocolo do fabricante e o seu perfil eletroforético foi avaliado utilizando *RNA 6000 Pico Kit* (5067-1513, Agilent Technologies) no equipamento Bioanalyzer 2100 (Agilent Technologies).

Em seguida, os sscDNAs (*single-stranded cDNAs*) tiveram um Adaptador A de dupla fita ligado às suas extremidades 3' - passo 3 da figura 1. O Adaptador A contém em uma das fitas - a que será ligada ao sscDNA- a sequência de nucleotídeos que será utilizada para hibridização do *primer* do protocolo do 454/Roche em dois momentos: primeiro, durante o processo de

amplificação clonal (PCR em emulsão feita para aumentar a quantidade de sequências na superfície de esferas e permitir o sequenciamento no equipamento 454/Roche) e segundo, durante o sequenciamento propriamente dito, que ocorre na placa denominada *Pico Titer Plate* (PTP), dentro do instrumento 454/Roche. O Adaptador A foi gerado por meio do anelamento do sscDNA Oligo A (Tabela 1) com o sscDNA Prime A (Tabela 1), utilizando as seguintes temperaturas: 5 minutos a 80°C, 7 minutos a 65°C, 7 minutos a 60°C, 7 minutos a 55°C, 7 minutos a 50°C, 7 minutos a 45°C, 7 minutos a 40°C, 7 minutos a 35°C, 7 minutos a 30°C e 7 minutos a 25°C.

A ligação do Adaptador A aos sscDNAs foi feita em 33 µL de reação contendo 3,3 µL do 10X *T4 DNA Ligase Buffer* (#B0202S; New England Biolabs), 6,67 µM do Adaptador A e 2.500 U da T4 DNA Ligase (#M0202S; New England Biolabs). A reação foi incubada por 2 horas a 16°C e, em seguida, foram adicionados 67 µL de 1x TE (pH 8,0) para parar a reação. Os sscDNAs com o Adaptador A ligado foram purificados utilizando o reagente RNAClean XP (A63987; Agencourt) com um volume correspondente a 1,8 vezes o volume de reação de ligação e seguindo as recomendações do fabricante. Os sscDNAs com o Adaptador A ligado foram quantificados utilizando o reagente RiboGreen (R11490; Invitrogen) conforme protocolo do fabricante e o seu perfil eletroforético foi avaliado utilizando *RNA 6000 Pico Kit* (5067-1513; Agilent Technologies) no equipamento Bioanalyzer 2100 (Agilent Technologies).

Em seguida, os cDNAs foram utilizados em uma PCR com 10 ciclos para garantir a presença do Adaptador A e do *primer* B em todos os cDNAs - passo 4 da figura 1. Esta reação de PCR foi feita em um volume final de 50 µL contendo: 1x Advantage 2 buffer (S1799; Clontech), 1x Advantage 2 polymerase mix (S1798; Clontech), 400 µM de dNTPs (28-4065-61; GE Healthcare), 2 µM do Primer AC (Tabela 1) e 2 µM do Primer B (Tabela 1). A reação de PCR foi feita segundo as seguintes condições: 96°C por 4 minutos, 10 ciclos de 94°C por 30 segundos e 64°C por 1 minuto, e finalmente 68°C por 3 minutos. Para purificação dos dscDNAs, estes foram

submetidos a dois ciclos de tratamento com o reagente AMPure XP PCR Purification (A63881; Agencourt) utilizando a proporção de 0,60 μL do reagente para 1 μL de reação de PCR (0,60:1). O reagente contem esferas de magnetita em um tampão adequado para precipitação do DNA em sua superfície, e a proporção entre reagente e cDNA determina uma seleção de tamanho dos cDNAs purificados. A proporção 0,6:1 usada foi ajustada para recuperar dscDNAs purificados com tamanhos entre 300 e 1000 pares de bases. Os dscDNAs foram quantificados utilizando o reagente PicoGreen (P7589; Invitrogen) conforme protocolo do fabricante e o seu perfil eletroforético foi avaliado utilizando *High Sensitivity DNA Kit* (5067-4626; Agilent Technologies) no equipamento Bioanalyzer 2100 (Agilent Technologies).

3.7 Sequenciamento em larga escala de RNAs longos

As bibliotecas de cDNA geradas (item 3.6) foram usadas para amplificação clonal das suas sequências em microrreatores, de acordo com o protocolo para PCR em emulsão em larga escala denominado *GS-FLX-Titanium-emPCR-Lib-L-LV Method*. As emulsões foram quebradas e as esferas recuperadas. A contagem das esferas foi realizada no contador de partículas Coulter Counter Multisizer 3, modelo Z1 (Beckman Coulter) e 2 milhões de esferas foram depositadas em cada uma das regiões de uma *PicoTiter Plate* (PTP). O sequenciamento foi feito na plataforma 454/Roche, seguindo as recomendações do protocolo padrão denominado *GS-FLX-Titanium-Sequencing-Method* no CATG - Centro Avançado de Tecnologias em Genômica - no Instituto de Química da Universidade de São Paulo.

3.8 Sequenciamento em larga-escala de RNAs curtos

O sequenciamento dos RNAs curtos (< 200 nt) obtidos após imunoprecipitação de Ago2 ou imunoprecipitação controle (item 3.2) foi realizado no *The Scripps Research Institute*, em San Diego, California, Estados Unidos. Foram usados 15 ng dos RNAs de cada amostra como material inicial para geração de bibliotecas de cDNA utilizando o protocolo denominado TruSeq *Small RNA Sample Prep Kit* (código de catálogo RS-930-1012, parte número 15004197, Illumina). As bibliotecas de cDNA foram separadas por eletroforese e foi recuperada a fração com tamanho menor que 150 pb, que foi usada para sequenciamento. Note que depois de ligados os adaptadores do protocolo Illumina, os miRNAs passam a ter 147 pb. As 6 bibliotecas, consistindo de 3 réplicas biológicas de cada condição (imunoprecipitação de Ago2 ou imunoprecipitação controle), foram geradas com *barcodes* diferentes e sequenciadas em conjunto em uma canaleta do equipamento HiSeq2000/Illumina. A análise das imagens obtidas e o *base calling* foram realizados pelo *software* do próprio instrumento, e a separação das sequências referentes a cada biblioteca foi feita de acordo com o *barcoding*.

3.9 Análises de Bioinformática

3.9.1 Análise Bioinformática de RNAs longos

Para a identificação dos transcritos longos ligados ao RISC ou transcritos obtidos nas imunoprecipitações controle, os *reads* obtidos após o sequenciamento das bibliotecas de cDNA na plataforma 454/Roche foram mapeados contra o genoma humano (versão GRCh37/hg19, baixado do UCSC Genome Browser, http://hgdownload.soe.ucsc.edu/downloads.html). Foi utilizada a ferramenta BLAT (Kent, 2002) e como ponto de corte sequências com no mínimo 70 % de cobertura e 90% de identidade. Após o mapeamento, as sequências foram separadas pela fita do genoma à qual elas alinhavam utilizando *scripts* em Perl desenvolvidos pelo nosso grupo. Em

seguida, para a geração de sequências consenso (*isotigs*), foi realizado um *assembly* com as sequências de cada fita separadamente, utilizando o *software* Newbler (versão 2.6, Roche). Para detecção dos transcritos sequenciados, os *isotigs* obtidos após *assembly* foram mapeados nos transcritos GENCODE (versão 14) (Harrow *et al.*, 2012) utilizando a ferramenta BLAT (Kent, 2002). As sequências obtidas foram visualizadas no site http://genome.ucsc.edu/ após geração de *tracks* em formato *bigBed*.

Para a quantificação da abundância dos transcritos longos ligados ao complexo de silenciamento induzido por RNA, ou transcritos obtidos nas imunoprecipitações controle, *scripts* em Perl foram utilizados para cálculo do valor de RPKM (*reads per kilobase of exon model per million mapped reads*) para cada transcrito (Mortazavi *et al.*, 2008), de acordo com a seguinte fórmula:

$$RPKM = \frac{10^9\, C}{N\, L}$$

na qual C é o número de sequências mapeadas que se alinham a exons dos transcritos, N é o número total de sequências mapeadas no experimento e L é a soma dos exons em pares de bases.

Essa normalização permite a comparação entre transcritos de diferentes comprimentos e entre amostras com profundidades de sequenciamento distintas. Os transcritos GENCODE foram mapeados em um gráfico de dispersão utilizando a função *smoothScatter* do programa estatístico R (Ihaka, 1996), de acordo com a média dos seus valores de RPKM obtidos nos sequenciamentos derivados das imunoprecipitações de Ago2 ou das imunoprecipitações controle.

3.9.2 Análise Bioinformática de RNAs curtos

As sequências de RNAs curtos obtidas foram inicialmente filtradas utilizando o FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Foram retiradas as sequências correspondentes aos adaptadores da tecnologia Illumina, assim como sequências menores que 18 nucleotídeos ou com qualidade inferior a Q30.

Para a identificação do conjunto de transcritos curtos ligados ao RISC, as sequências filtradas foram alinhadas com o banco de dados de RNAs curtos presente em (Berninger *et al.*, 2008), que contém um amplo repertório de RNAs curtos já descritos. Para tanto, foi utilizado o algoritmo Oligomap (Berninger *et al.*, 2008) e todos os alinhamentos perfeitos ou com até 1 erro (*mismatch*, inserção ou deleção) foram obtidos.

Para o mapeamento, identificação e quantificação de miRNAs conhecidos foi utilizado o software miRDeep2 (Friedlander *et al.*, 2012). Como arquivos de entrada foram usados o genoma humano (versão GRCh37/hg19, baixado do UCSC Genome Browser http://hgdownload.soe.ucsc.edu/downloads.html) e os miRNAs maduros e precursores de miRNAs [baixados do miRBase (Kozomara e Griffiths-Jones, 2011), versão 19 ftp://mirbase.org/pub/mirbase/19/].

O miRDeep2 (Friedlander *et al.*, 2012) também foi usado para a identificação de possíveis novos miRNAs entre as sequências utilizando o *miRDeep2 Module*. Esse algoritmo extrai as coordenadas genômicas que flanqueiam as sequências de possíveis novos miRNAs e calcula as probabilidades de que essa região genômica dê origem a um precursor de miRNA baseado na sua estrutura secundária e nas sequências usadas como entrada no mapeamento. O algoritmo então calcula um *score*, que representa a probabilidade da sequência genômica ser um precursor de

miRNA, assim como os números de possíveis miRNAs novos, falsos-positivos e falsos-negativos identificados em cada *score*.

3.9.3 Identificação de lncRNAs como possíveis alvos de miRNAs

A fim de identificar os miRNAs possíveis reguladores dos lncRNAs encontrados como ligados ao RISC nos sequenciamentos, foi feita uma busca no banco de dados miRCode (Jeggari *et al.*, 2012) pelos miRNAs que possuem sítios complementares aos lncRNAs identificados. Para isso, foram utilizados scripts em Perl desenvolvidos pelo grupo. O miRCode (Jeggari *et al.*, 2012) busca computacionalmente e descreve um conjunto de possíveis sítios alvos de miRNAs no transcritoma de RNAs não codificadores de proteínas longos, utilizando o GENCODE (Harrow *et al.*, 2012) para busca.

3.10 *Crosslink and immunoprecipitation* – CLIP

A técnica de CLIP (*crosslink and immunoprecipitation*) foi realizada segundo (Ule *et al.*, 2005) com algumas modificações, conforme descrito a seguir. Para cada experimento de CLIP, células HeLa foram subcultivadas em duas placas de 147,8 cm^2 (TPP) até a densidade de saturação de 80%, quando então o meio foi descartado e as células foram lavadas duas vezes com PBS. As células de cada uma das duas placas foram irradiadas no gelo com 300 mJ/cm^2 utilizando uma incubadora Stratalinker (Catálogo 400071; Stratagene) e em seguida removidas da placa através de raspagem com *cell scraper* ("*rodo*" de borracha) em PBS e transferidas para um tubo *falcon* (TPP). Uma alíquota foi retirada de cada tubo *falcon* e as células foram contadas utilizando-se uma câmara de Neubauer. Em seguida, uma alíquota de volume de PBS contendo $2x10^7$ células HeLa de cada *falcon* foi centrifugada separadamente a 300 x g por 7 minutos e o sobrenadante foi removido.

Essas alíquotas foram usadas posteriormente nas imunoprecipitações com o anticorpo específico ou com anticorpo controle. A partir desse passo foi utilizado o *Magna RIP RNA-Binding Protein Immunoprecipitation Kit* (17-700; Millipore), conforme descrito no item 3.2. A diferença é que, durante a etapa de lavagem, o *RIP Wash Buffer* foi substituído pelos tampões A (1x PBS, 0,1% SDS, 0,5% desoxicolato e 0,5% NP-40) e B (5x PBS, 0,1% SDS, 0,5% desoxicolato e 0,5% NP-40). Foram feitas duas lavagens com 500 µL cada do tampão A e duas lavagens com 500 µL cada do tampão B. Os RNAs não foram digeridos com RNase antes de serem isolados, para que os RNAs completos fossem recuperados para quantificação por PCR quantitativo em tempo real, conforme descrito no item 3.13.2.

3.11 Desenho de oligonucleotídeos iniciadores (*primers*)

Os *primers* utilizados para validação dos transcritos ligados ao RISC (RNAs obtidos nos itens 3.2 ou 3.10) ou para quantificação dos RNAs após superexpressão dos miRNAs (RNAs obtidos no item 3.14) foram desenhados no programa Primer Express versão 3.0 (Applied Biosystems), de acordo com os parâmetros padrão, e estão descritos na Tabela 2.

Tabela 2: Sequências dos *primers* utilizados na avaliação da qualidade do RNA, na validação da interação entre RNAs e complexo de silenciamento induzido por RNA e na avaliação do efeito da superexpressão de miRNAs. F:*Forward*; R:*Reverse*

Nome do oligonucleotídeo iniciador (*primer*)	Sequência do oligonucleotídeo iniciador (*primer*)
Avaliação da qualidade do RNA	
HistoneH3_F	ACGGCTCGTACAAAGCAGAC
HistoneH3_R	CCGCTGAAACTTGTTCACTG
Validação da interação entre RNAs e complexo de silenciamento induzido por RNA	
Linc553_F	TTCTGAGTCAGGCCAAAGGT

Linc553_R	TCGTCACCCACAACAAACTG
Linc488_F	GGCATTTGCCTGCTTCATAC
Linc488_R	TTGTGTTTACTGTGTGAGACAAATGA
NPM1P12_F	TTGCTGCTGATGAAGATGATG
NPM1P12_R	GTGCATTTTTGGCTTGGAGT
PIPSL_F	CCGCTTCTTTATGTCGGACT
PIPSL_R	AGCCGCTGATTGAACTCTGT
SPATA2_F	AGCTCCAAGGGCTACATCAA
SPATA2_R	TCTGCCACCTACATGGTCAA
YOD1_F	ATTTTGCTGGGACTTCATGG
YOD1_R	AGTGGTACCCTTGACGCTTG
TBCB_F	ACGGGTTGGACGAGATATGA
TBCB_R	AATGGGGTCAGGAGAAAAGG
Avaliação do efeito da superexpressão de miRNAs	
XPO6_F	GCCAACCAGAAAAGTGTGCT
XPO6_R	TGTAGTAGCGCAGGTCGTTG
Validação da interação entre RNAs e complexo de silenciamento induzido por RNA e avaliação do efeito da superexpressão de miRNAs	
TUG1_F	TAGCAGTTCCCCAATCCTTG
TUG1_F	CACAAATTCCCATCATTCCC
U1_F	GGGAGATACCATGATCACGAAGGT
U1_R	CCACAAATTATGCAGTCGAGTTTCCC
GAPDH_F	GCTCTCTGCTCCTCCTGTTC
GAPDH_R	ACGACCAAATCCGTTGACTC

3.12 Transcrição Reversa e Síntese de cDNA

Para a transcrição reversa e síntese de cDNA fita simples, foram utilizados 5 µL da suspensão de RNAs co-imunoprecipitados obtidos nos itens 3.2 ou 3.10 ou 1 µg de RNA total de cada condição experimental obtidos no item 3.14. Cada amostra foi transcrita com o *kit SuperScript III First Strand Synthesis Mix* (18080-400; Invitrogen), utilizando *primers* hexâmeros randômicos em um volume final de 20 µL, segundo as recomendações do fabricante. Uma pré-incubação de 3 min a 85ºC foi introduzida no protocolo, antes da fase de anelamento, para desnaturar completamente o RNA.

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

3.13.1 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 1 U de Taq Polimerase (GoTaq Green Master Mix; Promega), segundo as instruções do fabricante. Foram utilizados 250 ng de DNA genômico de células HeLa como controle positivo da reação e água DEPC (Invitrogen) como controle negativo. O produto da reação foi submetido a eletroforese em gel de agarose 1,7% contendo 0,5 µg/mL de brometo de etídio em tampão TAE (40mM Tris-Acetato, 1mM EDTA). 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).

3.13.2 PCR quantitativo em tempo real (qPCR)

As reações da PCR quantitativa em tempo real (qPCR) foram feitas em triplicata em um volume de 20 µL contendo 10 µL de Sybr Green Master Mix (Applied Biosystems), 5 µL de reação

contendo o cDNA da transcrição reversa (diluído 1:2 para transcrição reversa dos RNAs co-imunoprecpitados e diluído 1:6 para transcrição reversa de RNAs totais) e 3,2 µM de *primers Foward* e *Reverse* específicos para cada gene (Tabela 2). A reação foi realizada no equipamento 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 ou DNA genômico. 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 transcrito (normalizador) que não tenha sua expressão alterada nas diferentes condições de estudo. Para os experimentos de RIP e CLIP, utilizou-se o transcrito *U1* como normalizador, um RNA que faz parte do *spliceossomo* e que não é esperado se ligar a Ago2. Para os experimentos de superexpressão de miRNAs, foi utilizado o *GAPDH* como normalizador.

Calculou-se o delta-Ct (ΔCt) pela diferença entre a média das triplicatas (réplicas técnicas) do Ct do transcrito 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 (RIP-Anti-Ago2 e RIP-Controle, por exemplo). A diferença de expressão relativa do gene em estudo, *fold change*, é calculada como $2^{-\Delta\Delta Ct}$.

A significância estatística da diferença de expressão dos transcritos testados por qPCR entre condições distintas (imunoprecipitação com anticorpo anti-Ago2 *versus* imunoprecipitação controle ou tratamento com miRNA teste *versus* tratamento com miRNA 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.

3.14 Superexpressão de miRNAs

Células HeLa foram crescidas em garrafas de 25 cm^2 (Sarstedt) em meio DMEM (Cultilab) suplementado com SFB (Cultilab) 10%, 1,2 g/L de bicarbonato de sódio, 4,2 g/L de glicose e com os antibióticos penicilina (100 U/mL) e estreptomicina (100 µg/mL). Quando atingiram a confluência de 30%, as células foram lavadas com PBS e meio DMEM sem antibiótico foi adicionado. Em seguida, as células foram transfectadas com 25 nM de cada um dos *miRNAs mimics*: mirVana™ miRNA *mimic* hsa-miR-148b (MC10264, #4464066, Life Technologies) ou mirVana™ miRNA *mimic* Negative Control #1 (#4464058; Life Technologies) separadamente, utilizando o reagente Lipofectamine™ RNAi Max (13778150; Life Technologies).

Após 24 horas, as células foram lavadas com PBS e descoladas da superfície da garrafa por tratamento com 0,1% de tripsina seguido da adição de meio de cultura contendo SFB. Em seguida, as células foram sedimentadas por centrifugação a 300 x g por 5 minutos e o sobrenadante foi removido. Os sedimentos das células foram lavados com PBS, as células foram novamente centrifugadas a 300 x g por 5 minutos e o sobrenadante foi descartado. Os sedimentos de células foram mantidos a -80ºC até extração do RNA. O RNA foi extraído conforme descrito no item 3.3, sua integridade foi atestada de acordo com a descrição do item 3.4 e a quantificação dos transcritos de interesse foi feita de acordo com o item 3.13.2.

4. RESULTADOS

4.1 Imunoprecipitação do complexo de silenciamento induzido por RNA contendo a proteína Argonauta2

A fim de buscarmos lncRNAs possíveis alvos de miRNAs de maneira direta, recuperamos os RNAs longos obtidos pela imunoprecipitação nativa (RIP, *RNA-binding protein immunoprecipitation*) do complexo de silenciamento induzido por RNA (RISC, *RNA-induced silencing complex*), a partir do lisado de células de adenocarcinoma de cérvix HeLa em cultura. Para alvo da imunoprecipitação, escolhemos como componente do RISC a proteína Ago2, pois apesar dos relatos de que as quatro proteínas argonautas de humanos compartilham os mesmos miRNAs (Wang *et al.*, 2012), Ago2 é a única argonauta com capacidade de clivagem do alvo em mamíferos (Liu *et al.*, 2004; Meister *et al.*, 2004), e é capaz de se associar com a maioria dos miRNAs em alguns tecidos (Wang *et al.*, 2012). A linhagem HeLa teve seu genoma, epigenoma e transcritoma recentemente analisados (Adey *et al.*, 2013; Landry *et al.*, 2013), e tem sido utilizada na comunidade científica como modelo no estudo das relações entre miRNAs, proteínas argonauta e RNAs-alvo (Janas *et al.*, 2012).

Utilizamos anticorpo anti-Ago2 nas imunoprecipitações e, em paralelo, foram feitas imunoprecipitações controle, nas quais foi utilizado anticorpo IgG normal de camundongo não-imunizado, a fim de se avaliar a contribuição de interações inespecíficas para a recuperação de RNAs durante o ensaio. A figura 2 mostra a imunoprecipitação de Ago2, assim como o perfil eletroforético dos RNAs longos extraídos e purificados dos imunoprecipitados, tanto no ensaio controle quanto no ensaio com anticorpo anti-Ago2.

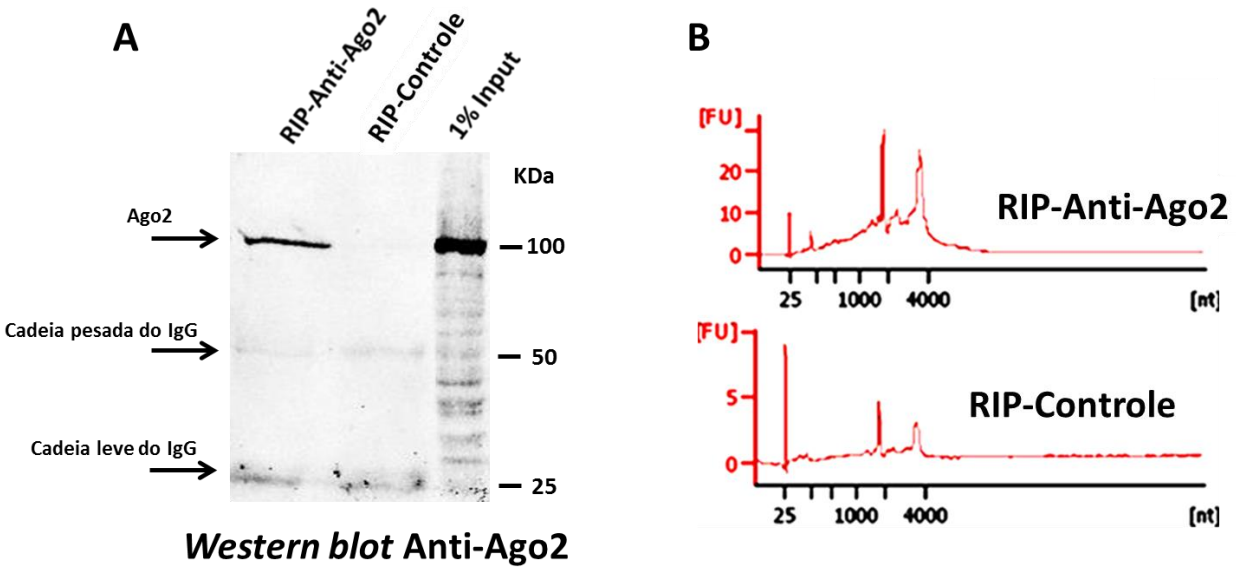

Figura 2. Imunoprecipitação em células HeLa da proteína Argonauta2, que é parte do complexo de silenciamento induzido por RNA (RISC), e caracterização dos RNAs longos co-imunoprecpitados. **A**. *Western blot* do conteúdo de proteínas imunoprecipitadas no ensaio RIP. As frações retiradas após imunoprecipitação com anticorpo anti-Ago2 (RIP-Anti-Ago2) ou com IgG normal de camundongo não-imunizado (RIP-Controle), ou 1% da fração inicial, denominada "Proteína *Input*" foram corridos em gel de SDS-PAGE e as proteínas foram transferidas para uma membrana de nitrocelulose que foi incubada com o anticorpo anti-Ago2, e revelada com anticorpo secundário fluorescente. Uma banda de aproximadamente 100 kDa foi detectada nas amostras RIP-Anti-Ago2 e 1% *Input*, mas não na fração imunoprecipitada com IgG normal de camundongo não-imunizado. **B**. Os painéis representam o perfil analítico dos RNAs purificados a partir da fração imunoprecipitada com anticorpo anti-Ago2 ou com IgG normal de camundongo não-imunizado; foram analisados 3 ng de cada RNA em eletroforese capilar no equipamento Agilent 2100 Bioanalyzer (Agilent Technologies) usando-se o *chip RNA 6000 Pico Kit*, adequado para detectar RNAs com tamanhos entre 200 e 6000 nt. O pico em 25 nt é um marcador adicionado a cada amostra, que é usado pelo instrumento para definir o tamanho mínimo detectado na corrida.

Após as imunoprecipitações, os RNAs longos co-imunoprecipitados obtidos foram identificados por meio de sequenciamento em larga escala na plataforma 454/Roche (RIP-Seq), conforme realizado para outros complexos ribonucleoprotéicos (Zhao *et al.*, 2010). Como a plataforma 454/Roche não possui protocolo para geração de bibliotecas de cDNA direcionadas, foi desenvolvido um método, conforme descrito na próxima seção.

4.2 Desenvolvimento de método de geração de bibliotecas fita-específicas para sequenciamento de RNAs longos no 454/Roche

A plataforma 454/Roche não possui protocolo para geração de bibliotecas de cDNA fita-específicas. As primeiras indicações de preparação de bibliotecas de cDNA, segundo o protocolo padrão da 454/Roche denominado *GS-FLX Titanium General Library Preparation Method* sugeriam que a primeira fita do cDNA fosse gerada com oligonucleotídeo oligod(T) e que, após a geração da segunda fita do cDNA, o cDNA dupla-fita fosse nebulizado e os fragmentos tivessem as suas extremidades polidas. Em seguida, adaptadores sem direcionalidade eram ligados às extremidades 5' e 3' dos cDNAs. Além de não fornecer a direcionalidade dos cDNAs sequenciados, este método foi reportado na literatura como sendo ineficiente ao gerar perdas excessivas (Maricic e Paabo, 2009). Além disso, este método também requer a titulação em pequeno volume de emPCR, para que seja encontrada a proporção ideal entre o número de cópias de cDNA e o número de esferas no protocolo da 454/Roche. Esta proporção ideal é usada no passo de emPCR em larga escala para geração das esferas contendo, em cada esfera, vários bilhões de cópias da mesma molécula de cDNA; as esferas serão usadas posteriormente no sequenciamento. Foi mostrado que esta etapa de titulação apresenta comportamento irregular e pouco confiável quando é utilizado cDNA (Sandberg *et al.*, 2009).

Com o aprimoramento da tecnologia 454/Roche, foi lançado um novo protocolo denominado *GS-FLX Titanium-cDNA Rapid Library Preparation Method*, que reduziu as perdas durante o processo de geração das bibliotecas de cDNA. No entanto, este protocolo também não permite a geração de bibliotecas de cDNA direcionadas, e ainda requer a titulação da biblioteca de cDNA em pequeno volume de emPCR.

Como a direcionalidade dos RNAs é uma informação importante na identificação e caracterização de lncRNAs, as limitações citadas dos protocolos 454/Roche para construção de

bibliotecas de cDNA fita-específicas nos impulsionaram a desenvolver um método de geração de bibliotecas de cDNA direcionadas.

Para o desenvolvimento do método próprio, inicialmente nós nos inspiramos em duas publicações que relatavam a geração de bibliotecas de cDNA direcionadas para sequenciamento na plataforma 454/Roche. Uma destas publicações realizava o sequenciamento do genoma completo de RNA do vírus H1N1 (Simons e Hutchison, 2007), enquanto a outra tratava do sequenciamento do transcritoma de células não tumorais e tumorais derivadas de tecido prostático humano para detecção de genes de fusão (Maher *et al.*, 2009). Neste método, os RNAs poliA+ são fragmentados por hidrólise alcalina e em seguida é realizada transcrição reversa com oligonucleotídeos randômicos para geração de cDNAs fita-simples. A estes cDNAs fita-simples são ligados adaptadores A e B dupla-fita com orientação específica: o adaptador A possui uma extremidade 5' não pareada composta de cinco nucleotídeos, permitindo a sua ligação às extremidades 3' de cDNAs fita-simples, enquanto o adaptador B possui uma extremidade 3' não pareada composta de cinco nucleotídeos randômicos, permitindo a sua ligação às extremidades 5' de cDNAs fita-simples (Maher *et al.*, 2009). Por fim, o cDNA fita-simples é amplificado em uma reação de PCR com 20 ciclos, usando *primers* complementares aos adaptadores A e B para garantir que os cDNAs obtidos tenham os dois adaptadores, um em cada extremidade. É importante notar que, devido à baixa eficiência da reação da enzima DNA ligase nestas condições, uma fração considerável dos cDNAs contêm apenas um adaptador em uma das extremidades, ou mesmo nenhum adaptador (Maricic e Paabo, 2009), e esta fração é variável entre os experimentos, daí sendo necessária a titulação de cada biblioteca por PCR de emulsão em pequena escala.

Este método alternativo de geração de bibliotecas de cDNA para sequenciamento no 454/Roche traz as seguintes vantagens:

a- as sequências são geradas de forma direcionada, ou seja, elas contêm a informação de qual fita do DNA genômico (senso ou antisenso) gerou o RNA, já que os adaptadores A e B foram ligados ao cDNA com orientação específica;

b- obtém-se uma cobertura mais uniforme do gene por ser utilizado oligonucleotideo randômico durante a transcrição reversa (Wang *et al.*, 2009);

c- com a amplificação por PCR, em uma etapa anterior à emPCR, todos os cDNAs gerados e aplicados às esferas da emPCR terão os dois adaptadores em suas extremidades. A garantia da presença dos dois adaptadores nas extremidades de todas as moléculas de cDNA descarta a necessidade da etapa de titulação em pequeno volume de PCR de emulsão, pois com um adaptador em cada extremidade da molécula de cDNA dupla-fita, necessita-se de apenas 2 cópias de DNA por esfera no passo de PCR de emulsão em larga escala. Essa proporção está relacionada ao fato de que apenas uma das duas fitas é a fita a ser sequenciada (pois apenas uma das fitas contem o *primer* A na extremidade 5'; a outra fita contem o *primer* B na extremidade 5' e não pode ser sequenciada).

Este método foi usado com sucesso na descrição do perfil transcricional de vermes adultos machos de *Schistosoma mansoni* pelo nosso grupo (Almeida *et al.*, 2012). As análises realizadas nesse trabalho (Almeida *et al.*, 2012) confirmaram a transcrição em 990 regiões nas quais os genes eram somente preditos, corrigiram predições gênicas errôneas e catalogaram 2.185 novas unidades transcricionais em regiões intrônicas ou intergênicas do genoma desse parasita. Estas novas unidades transcricionais podem representar novos lncRNAs, UTRs de genes codificadores de proteínas ou novos genes codificadores de proteínas.

Entretanto, esse método também apresenta as seguintes desvantagens:

a- a forma de ligação dos adaptadores com orientação específica não é ótima para as condições de sequenciamento, e pode levar a um enriquecimento na região 3' dos transcritos . O

adaptador A é ligado na extremidade 5´ do cDNA fita-simples (equivalente à extremidade 3’ do RNA). Como o sequenciamento 454/Roche tem início a partir do *primer* A, muitas vezes a leitura inicia-se pelo equivalente à cauda poli-A dos RNAs. Assim, as sequências geradas nesse método tendem a apresentar um enriquecimento na região 3’ dos transcritos. Outra desvantagem é que o tratamento dos dados feito pelo sequenciador 454/Roche, nos primeiros ciclos de sequenciamento, envolve uma normalização do sinal utilizando a leitura da luminescência dos primeiros 10 ciclos. Deste modo, ocorre uma interferência prejudicial na normalização do sinal durante os primeiros ciclos de sequenciamento quando existe uma cauda poli-A muito longa em muitas das sequências, pois este trecho de homopolímero gera um sinal muito intenso com viés para a base A (adenosina);

b- o uso de um alto número de ciclos (20 ciclos) durante a amplificação da biblioteca por PCR pode levar uma distorção da abundância relativa dos transcritos conforme descrito para bibliotecas sequenciadas na plataforma Illumina em (Mamanova *et al.*, 2010);

c- por fim, sabe-se que a enzima transcriptase reversa é capaz de gerar uma segunda fita espúria de cDNA através da sua atividade de DNA polimerase dependente de DNA (Spiegelman *et al.*, 1970). O início da polimerização da segunda fita de cDNA a partir da primeira fita do cDNA pode ocorrer tanto por uma volta na extremidade 3’ da primeira fita do cDNA quanto pelo uso de fragmentos de RNA ou do próprio *primer* usado na reação de transcrição reversa (Gubler, 1987). Portanto, neste método, caso a transcriptase reversa gere a segunda fita do cDNA a partir da primeira fita do cDNA, os adaptadores A e B poderão se ligar nas extremidades 3’ e 5’, respectivamente, das moléculas da segunda fita do cDNA. Isso levará ao sequenciamento de moléculas de forma invertida.

Visando à resolução das desvantagens citadas acima, desenvolvemos um novo método de geração de bibliotecas de cDNA direcionadas para sequenciamento no 454/Roche (figura 1 da seção 3.6), com as modificações descritas a seguir:

a- para evitar os problemas gerados pela baixa eficiência de ligação dos adaptadores e para controlar a extremidade na qual o adaptador A é adicionado, elaboramos uma nova forma de incorporação destes adaptadores aos cDNAs. Assim, a transcrição reversa passou a ser feita com oligonucleotídeos randômicos que já contêm na sua extremidade 5' a sequência do *primer* B do protocolo 454/Roche (*primer* B-N9, tabela 1, item 3.6). Com essa alteração, dois problemas foram contornados: um deles é evitar uma etapa limitante do processo que é a baixa eficiência da T4 DNA Ligase que acrescentava os adaptadores A e B (Maricic e Paabo, 2009). Neste novo método (figura 1 da seção 3.6), só ocorre a ligação do adaptador A, visto que o adaptador B já está incorporado ao oligonucleotídeo randômico usado durante a transcrição reversa. Além disso, o forte sinal da cauda poli-A no início dos ciclos de sequenciamento passa a ser evitado, uma vez que a leitura inicia-se na extremidade em que está o *primer* A e termina na extremidade em que está o *primer* B e, neste novo método, o *primer* B encontra-se incorporado ao oligonucleotídeo randômico, localizando-se portanto na extremidade correspondente ao 3' do RNA (veja figura 1, no item 3.6). Dessa forma, quando houver cauda poli-A em um fragmento de RNA, esta será a última região a ser sequenciada e não irá interferir na normalização do sinal; no máximo será considerada pelo *software* do sequenciador como uma região homopolimérica de baixa qualidade na sequência;

b- para evitar a distorção na abundância relativa dos transcritos, reduziu-se o número de ciclos da etapa final de PCR. Após a síntese da primeira fita do cDNA, o adaptador A dupla fita é ligado à extremidade 3' do cDNA, que corresponde à extremidade 5' do RNA. Em

seguida, os cDNAs são submetidos a uma PCR de 10 ciclos, tanto para garantir a presença dos dois adaptadores nas duas extremidades dos cDNAs, quanto para gerar a segunda fita do cDNA. Essa etapa de geração da segunda fita do cDNA se faz necessária, pois é esta fita que irá parear nas esferas, visto que o oligonucleotídeo ancorado às esferas possui a mesma sequência do *primer* B incorporado ao oligonucleotídeo randômico usado na geração da primeira fita do cDNA. O uso de 10 ciclos na reação de PCR possivelmente diminui a distorção na abundância dos transcritos quando comparado a bibliotecas geradas com 20 de ciclos de PCR;

c- para evitar a perda de fragmentos curtos gerados durante a lise alcalina do RNA e consequentemente a perda de amostragem de algumas regiões dos transcritos, a etapa de lise foi eliminada do protocolo.

Este novo método foi usado com sucesso na análise da expressão gênica diferencial entre vermes adultos machos, vermes adultos fêmeas e ovos de *Schistosoma mansoni* pelo nosso grupo (DeMarco *et al.*, 2013). Os RNAs poliA+ destes diferentes estágios foram extraídos e utilizados para a geração de bibliotecas de cDNA de acordo com o novo método acima descrito, o que permitiu a identificação de transcritos e vias diferencialmente ativados nos distintos estágios do parasita (DeMarco *et al.*, 2013).

Além de introduzir as vantagens descritas acima em relação ao método de (Maher *et al.*, 2009), o novo método desenvolvido ainda manteve as propriedades de maior cobertura dos transcritos (em função da utilização de oligonucleotídeos randômicos durante a transcrição reversa) e de eliminação da etapa de titulação (em decorrência da PCR final de 10 ciclos que garante adaptadores nas duas extremidades dos cDNAs e permite a utilização de 2 cópias de cDNA por esfera).

Para confirmar que a proporção de 2 cópias de cDNA por esfera no novo método era efetivamente a ideal, foi realizada a titulação da biblioteca gerada com os RNAs poliA+ de vermes adultos machos de *Schistosoma mansoni* em pequeno volume de emPCR (tabela 3). Na titulação, foram utilizadas diferentes proporções de moléculas de cDNA para esferas, e foi medido o enriquecimento de esferas positivas de acordo com o protocolo *GS-FLX-Titanium-emPCR-Lib-L-SV*. O método de titulação indicado no protocolo da 454/Roche descreve que foi determinado experimentalmente que um enriquecimento em torno de 8% de esferas positivas é o enriquecimento ideal para que se alcance 1 cópia "efetiva" de DNA por esfera; este enriquecimento de 8% foi obtido com 2 cópias de cDNA por esfera (Tabela 3), como esperado.

Tabela 3. Titulação da proporção de moléculas de cDNA que devem ser ligadas às esferas para obtenção de complexos cDNA:esferas eficientes para sequenciamento. Foram utilizados os cDNAs gerados pelo método desenvolvido neste trabalho, a partir de RNAs poliA+ extraídos do verme adulto macho de *Schistosoma mansoni*. Para cada proporção de cDNA:esferas, foi quantificada a porcentagem de esferas enriquecidas, segundo o protocolo 454/Roche.

Proporção de cDNA:esferas (*Copies per bead, cpb*)	**Porcentagem obtida de esferas enriquecidas**
1	6,15
2	8,15
4	10,4
6	23,4

Para avaliar o desempenho deste novo método frente aos eventos de geração de segunda fita espúria de cDNA pela transcriptase reversa, foram comparadas as proporções de transcritos antissenso invertidos entre as bibliotecas sequenciadas em (Almeida *et al.*, 2012) que foram geradas segundo o método descrito em (Maher *et al.*, 2009), com as bibliotecas analisadas em (DeMarco *et al.*, 2013), que foram geradas com o novo método aqui desenvolvido. Para isso, foi

calculada a proporção de transcritos supostamente antissenso que mapeava em transcritos contendo múltiplos exons, e que apresentavam *splicing* com as bordas intron/exon exatamente nos mesmos locais da sequência, porém na fita oposta. Sabe-se que o par aceptor/doador canônico de *splicing* (GT/AG) não ocorre de maneira invertida (CT/AC) e que, portanto, os supostos RNAs antissenso que apresentavam tal padrão de *splicing* representam um artefato de geração da biblioteca.

Enquanto as bibliotecas geradas com o novo método desenvolvido (DeMarco *et al.*, 2013) apresentaram uma proporção média de transcritos antissenso invertidos de 0,5% dos transcritos antissenso *spliced*, as bibliotecas geradas com o método anterior (Almeida *et al.*, 2012) tiveram a proporção de 5% de transcritos antissenso invertidos em relação a todos os transcritos antissenso *spliced* daquelas bibliotecas. Esse dado demonstra que a modificação na forma de adição dos adaptadores ao cDNA, mais precisamente a incorporação do *primer* B durante a etapa de transcrição reversa e o estabelecimento de um único evento de ligação do adaptador A, levou a um significativo aumento na acurácia da determinação de direcionalidade dos transcritos obtidos e no desempenho do novo método desenvolvido.

4.3 Identificação de lncRNAs ligados ao complexo RISC

Para a identificação de lncRNAs ligados ao complexo RISC e regulados por miRNAs, os RNAs longos obtidos a partir das frações imunoprecipitadas com anticorpo anti-Ago2 (RIP-Anti-Ago2) ou com IgG normal de camundongo não-imunizado (RIP-Controle) (item 4.1) foram usados na geração de bibliotecas de cDNA de acordo com o novo método descrito no item anterior.

O método desenvolvido para geração de bibliotecas de cDNA direcionadas para sequenciamento no 454/Roche foi novamente validado por meio da titulação destas bibliotecas de cDNA de acordo com o protocolo denominado *GS-FLX-Titanium-emPCR-Lib-L-SV*. Após a titulação e enriquecimento, confirmou-se que a proporção de 2 cópias de cDNA por esfera levou a um enriquecimento de aproximadamente 8% de esferas, e que portanto esta proporção gera uma suspensão de esferas prontas para sequenciamento no 454/Roche. Deste ponto em diante, as titulações foram dispensadas.

Para sequenciamento, foram geradas duas bibliotecas de cada condição (2 réplicas biológicas de RIP-Anti-Ago2 e 2 réplicas biológicas de RIP-Controle). Estas bibliotecas foram utilizadas na PCR de emulsão em larga escala de acordo com o protocolo denominado *GS-FLX-Titanium-emPCR-Lib-L-LV* e em seguida foram sequenciadas no instrumento 454-GS-FLX. A tabela 4 mostra a quantidade de sequências somadas das duas corridas para cada condição e o tamanho médio dessas sequências.

Tabela 4: Número e tamanho médio das sequências obtidas a partir das amostras de RNAs longos co-imunopreciptados com anticorpo anti-Ago2 (RIP-Anti-Ago2) ou no ensaio controle (RIP-Controle).

Amostra	Sequências com *Phred Quality Score* maior que Q20	Tamanho médio das sequências (pares de bases)
RIP-Anti-Ago2	1.539.938	343
RIP-Controle	738.982	347

Após o sequenciamento das amostras de RNAs longos (item 3.7), foram feitas análises (item 3.9.1) para detectar os transcritos longos codificadores e não-codificadores de proteína ligados ao complexo RISC na linhagem celular HeLa. Para determinação de um limiar a partir do qual um transcrito poderia ser considerado como ligado ao RISC, foram estabelecidos: 1- um número mínimo de sequências por transcrito para que cada transcrito fosse considerado detectado acima do ruído e 2- um enriquecimento de sequências do transcrito nos sequenciamentos RIP-Anti-Ago2 em relação aos sequenciamentos RIP-Controle.

Para ser considerado como detectado acima do ruído, cada transcrito deveria ser detectado em pelo menos um dos dois sequenciamentos de cada tipo (RIP-Anti-Ago2 ou RIP-Controle) com mais de 3 sequências, ou ser detectado nos dois sequenciamentos do mesmo tipo (RIP-Anti-Ago2 ou RIP-Controle) com mais de uma sequência em cada. A figura 3A mostra o número de transcritos GENCODE (Harrow *et al.*, 2012) não detectados em nenhum sequenciamento, assim como o número de transcritos detectados somente na RIP-Anti-Ago2, somente na RIP-Controle ou em ambos os sequenciamentos. A figura 3B mostra as classes dos transcritos detectados exclusivamente nos sequenciamentos RIP-Anti-Ago2.

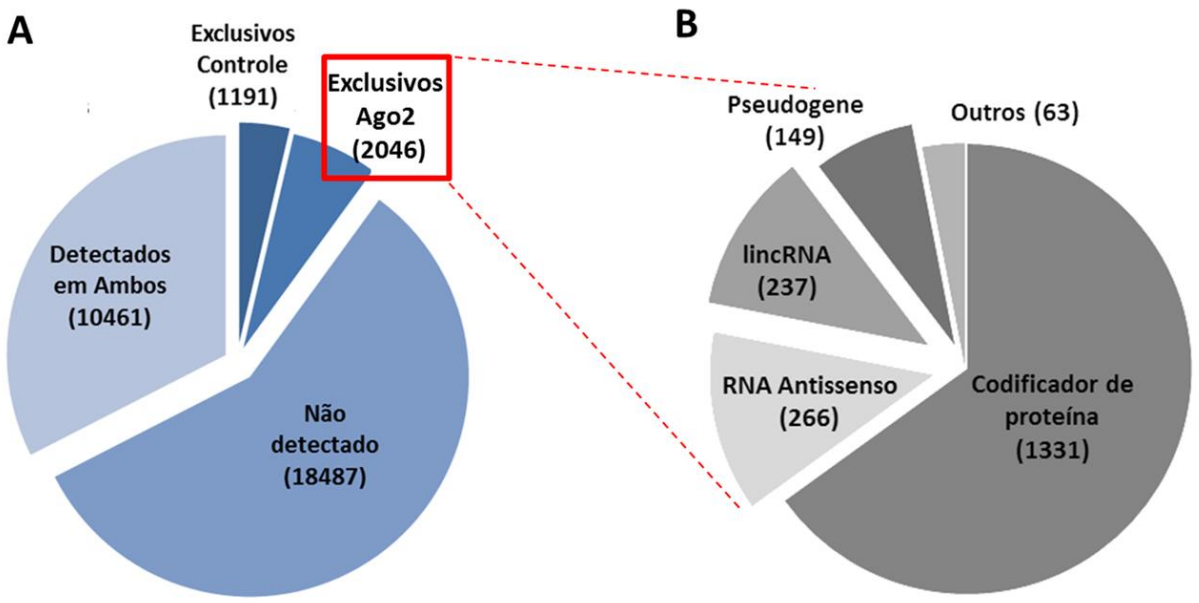

Figura 3. Detecção dos transcritos anotados no GENCODE v.14 (Harrow *et al.*, 2012) nos sequenciamentos RIP-Anti-Ago2 e RIP-Controle. A. São apresentados o número de transcritos não detectados em nenhum sequenciamento, assim como o número de transcritos detectados somente na RIP-Anti-Ago2, somente na RIP-Controle ou em ambos os sequenciamentos. **B.** Os números de transcritos detectados somente nos sequenciamentos RIP-Anti-Ago2 são mostrados de acordo com as suas classes.

Para determinação do enriquecimento dos transcritos nos sequenciamentos RIP-Anti-Ago2 em relação aos sequenciamentos RIP-Controle, um parâmetro útil para aqueles detectados em ambas as condições, foi calculado o valor de RPKM (*reads per kilobase of exon model per million mapped reads*) (Mortazavi *et al.*, 2008) para cada transcrito. O cálculo do RPKM permite a comparação dos valores de enriquecimento entre diferentes transcritos de diferentes sequenciamentos, pois faz a normalização pelo tamanho de cada transcrito e pela profundidade do sequenciamento (número de sequências) de cada biblioteca.

A figura 4 mostra a distribuição dos 32.185 transcritos anotados no GENCODE (Harrow *et al.*, 2012) em um gráfico de dispersão, de acordo com as médias dos seus valores de RPKM nos sequenciamentos RIP-Controle (eixo x) e RIP-Anti-Ago2 (eixo y). Transcritos não detectados em

nenhum sequenciamento, com valores de RPKM de zero ou muito próximo de zero representam o maior grupo ("nuvem" azul próxima de [0,0]). Transcritos com valores de x iguais a zero e com valores de y diferentes de zero indicam a população de transcritos detectados como ligados exclusivamente ao RISC (pontos que formam a linha vertical que aponta para [0,0]).

Assim, foi encontrado um conjunto de 2.940 transcritos com um enriquecimento maior ou igual a quatro vezes em relação ao RIP-Controle (à esquerda e acima da linha vermelha "Linha de Enriquecimento (4:1)", figura 4) que foi considerado como ligado ao RISC. Destes 2.940 transcritos, 2.046 foram detectados exclusivamente no ensaio RIP-Anti-Ago2 e 894 transcritos foram encontrados com enriquecimento maior ou igual a quatro vezes na fração RIP-Anti-Ago2 em relação ao RIP-Controle.

Foi encontrado outro conjunto, que corresponde a 2.438 transcritos com enriquecimento entre 1 e 4 em vezes na fração RIP-Anti-Ago2 em relação ao RIP-Controle (região entre a linha vermelha e a linha preta no gráfico, figura 4); decidimos não analisar mais este conjunto, considerado de baixo enriquecimento. O terceiro conjunto compreende 26.807 transcritos não enriquecidos na RIP-Anti-Ago2 (região abaixo da linha preta no gráfico, figura 4), dos quais 18.487 transcritos não foram detectados em nenhum sequenciamento (nuvem azul próxima a [0,0]), 1.191 transcritos foram detectados somente na RIP-Controle e 7.129 transcritos foram detectados na RIP-Controle e também na RIP-Anti-Ago2 com enriquecimento menor que 1.

No total, foram detectados 12.507 transcritos em pelo menos um sequenciamento RIP-Anti-Ago2, incluindo mRNAs codificadores de proteínas e lncRNAs, sendo 2.940 transcritos detectados exclusivamente na RIP-Anti-Ago2 ou com um enriquecimento maior ou igual a quatro vezes em relação ao RIP-Controle, que são os de interesse em nosso trabalho. Além disso, foram detectados outros 2.438 transcritos com baixo enriquecimento, entre 1 e 4 em vezes em relação ao RIP-

Controle, e 7.129 transcritos foram detectados na RIP-Anti-Ago2 com enriquecimento menor que 1.

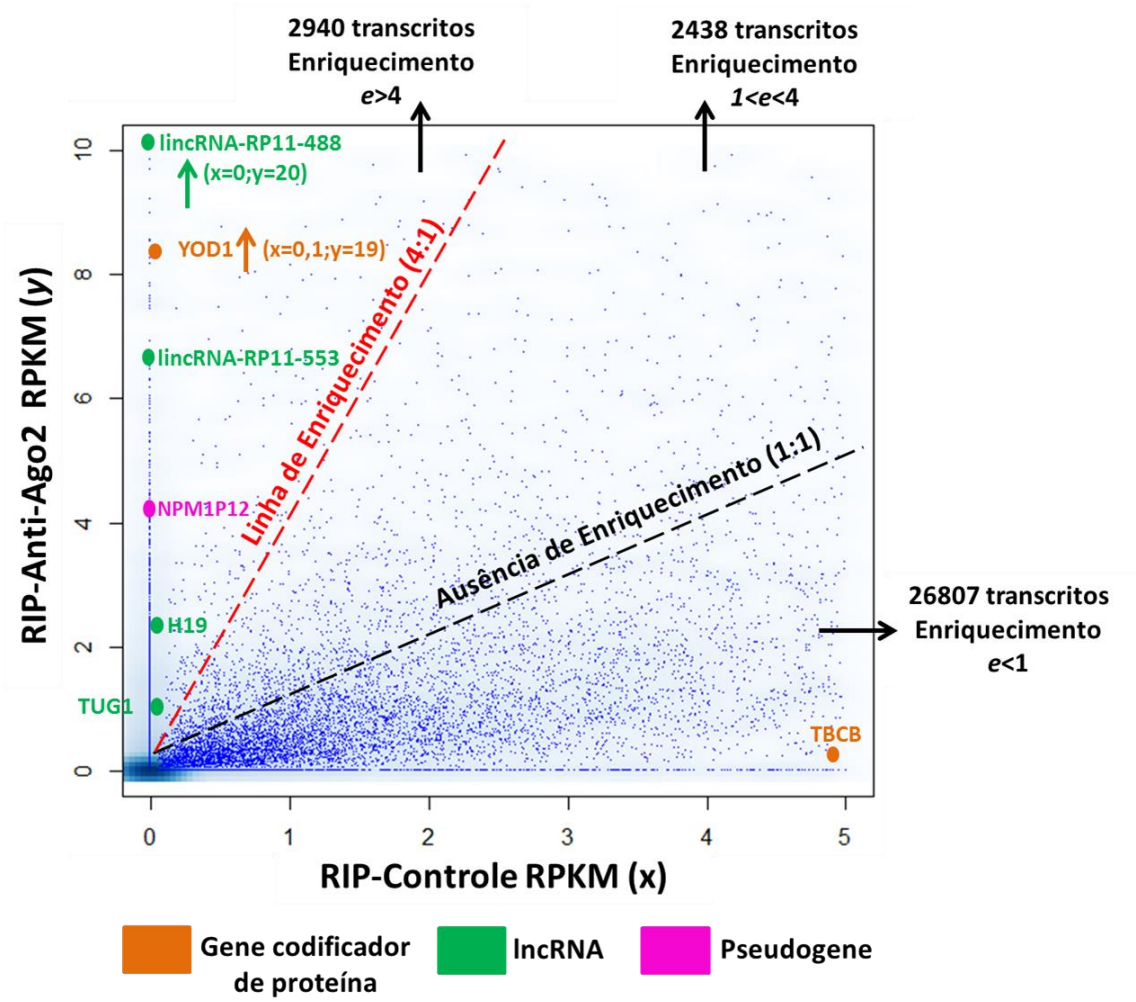

Figura 4. Transcritos presentes na versão 14 do banco de dados GENCODE (Harrow *et al.*, 2012) e detectados nos sequenciamentos RIP-Anti-Ago2 e/ou RIP-Controle. Estão mostrados todos os transcritos detectados nos sequenciamentos e que estão anotados no GENCODE, incluindo mRNAs codificadores de proteínas e lncRNAs. A abundância dos transcritos detectados nos sequenciamentos RIP-Anti-Ago2 e/ou RIP-Controle foi avaliada por meio do cálculo do valor de RPKM (*reads per kilobase of exon model per million mapped reads*) para cada transcrito. O gráfico de dispersão mostra cada transcrito (pontos azuis) de acordo com seus valores de RPKM no sequenciamento RIP-Anti-Ago2 (eixo y) e no sequenciamento RIP-Controle (eixo x). A linha pontilhada preta mostra a relação de abundância 1:1 (Ago2:Controle), ou seja, ausência de enriquecimento. A linha vermelha mostra a relação 4:1

(Ago2:Controle), tomada como o limite para ser considerada como evidência de enriquecimento - segundo (Zhao *et al.*, 2010). Os quatro lncRNAs selecionados como de interesse estão coloridos em verde, e possuem uma relação Ago2:Controle bem acima do limite arbitrário 4:1. Um pseudogene de interesse (lilás) e dois genes codificadores de proteína selecionados como controle (laranja) também estão indicados, um controle positivo e outro controle negativo.

Com relação aos lncRNAs, foram encontrados 686 transcritos correspondendo a lncRNAs entre os 2.940 transcritos detectados com um enriquecimento maior ou igual a quatro vezes em relação ao RIP-Controle, dos quais 664 lncRNAs foram encontrados exclusivamente nos sequenciamentos RIP-Anti-Ago2 e 22 lncRNAs foram encontrados com um enriquecimento maior ou igual a quatro vezes em relação ao RIP-Controle.

Esse conjunto de 686 lncRNAs foi considerado como ligado ao RISC e compreende 271 RNAs antissenso, 244 lincRNAs, 155 pseudogenes, 4 RNAs anotados como ncRNAs e 12 transcritos que geram miRNAs. A lista dos 686 lncRNAs considerados ligados ao RISC e a abundância relativa de transcritos (RPKM) detectados na fração ligada ao RISC para cada lncRNA está apresentada no Apêndice 1.

4.4 Validação da interação entre lncRNAs e o complexo RISC

A fim de validarmos a interação entre os RNAs longos detectados nos sequenciamentos RIP-Anti-Ago2 e o complexo RISC, foram realizadas novas imunoprecipitações nativas e os níveis de alguns RNAs selecionados foram medidos por qPCR. Foram escolhidos e validados dois lncRNAs intergênicos enriquecidos nos sequenciamentos RIP-Anti-Ago2 que possuem funções parcialmente descritas na literatura: *TUG1* e *H19* (figura 5A). *TUG1* é um lincRNA induzido por p53 em resposta ao dano do DNA (Guttman *et al.*, 2009), que se liga a PRC2 (*Polycomb Repressive Complex 2*) e está envolvido na regulação de genes que controlam o ciclo celular e apoptose (Khalil

et al., 2009). *H19* é um lincRNA com efeito de supressão tumoral tanto *in vitro* quanto *in vivo*, cujo mecanismo de ação está envolvido com a geração do miR-675 (Keniry *et al.*, 2012) e que recentemente foi mostrado agir como esponja dos miRNAs da família let-7 (Kallen *et al.*, 2013). Outros dois lincRNAs com funções ainda desconhecidas (*lincRNA-RP11-553L6.5* e *lincRNA-RP11-488L18.10*), detectados exclusivamente nos sequenciamentos RIP-Anti-Ago2, também foram validados (figura 5A).

Nos últimos anos, pseudogenes têm sido implicados na regulação de transcritos codificadores de proteínas ao competirem com um mRNA pelos miRNAs que o alvejam (Ebert e Sharp, 2010). O pseudogene *NPM1P1,2* detectado nos sequenciamentos RIP-Anti-Ago2 foi validado por RIP-qPCR (figura 5B). Também foram escolhidos e validados um mRNA detectado exclusivamente nos sequenciamentos RIP-Anti-Ago2 (*SPATA2*) e um mRNA (*YOD1*) enriquecido nos sequenciamentos RIP-Anti-Ago2 (figura 5C). Como controle negativo, um mRNA não enriquecido no sequenciamento RIP-Anti-Ago2 (*TBCB*) também foi testado e não mostrou enriquecimento significativo na nova imunoprecipitação com anticorpo anti-Ago2 (figura 5C).

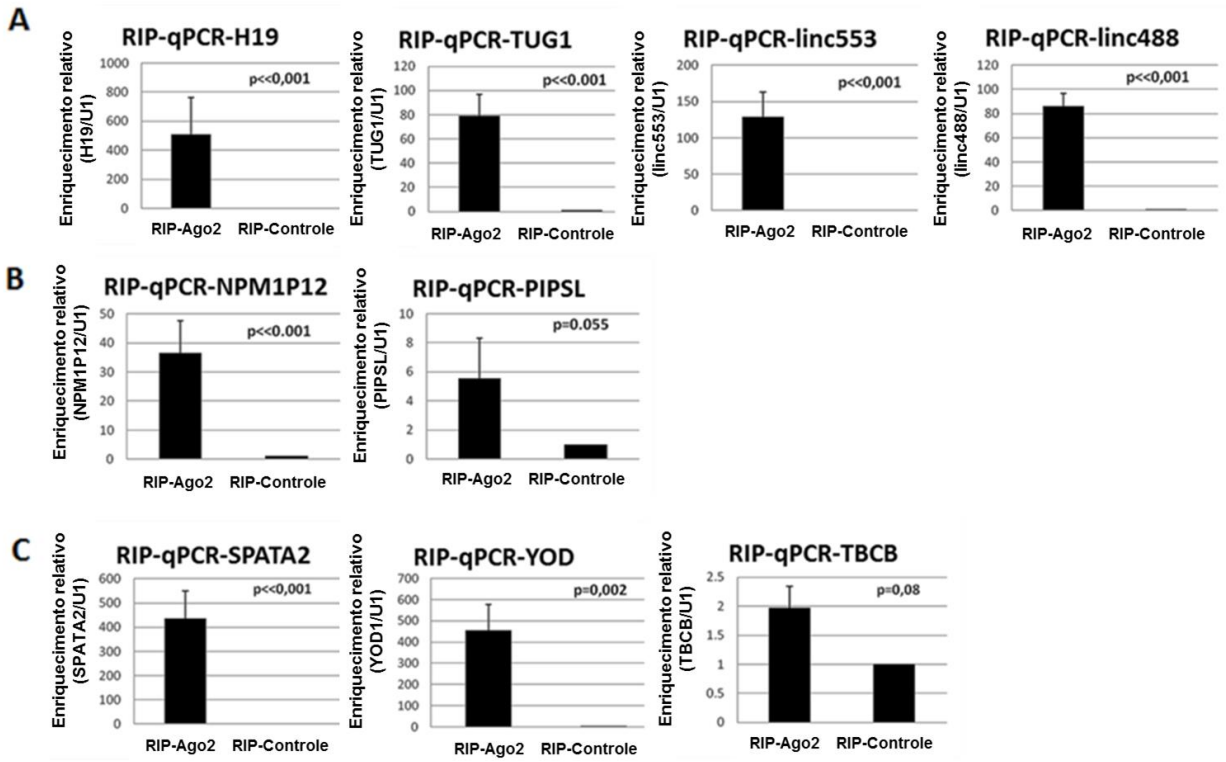

Figura 5. Validação por *RNA-binding protein immunoprecipitation* (RIP) seguido por PCR quantitativo em tempo real (RIP-qPCR) da interação entre RISC e RNAs longos selecionados. (**A**) lincRNAs, (**B**) pseudogenes e (**C**) RNAs codificadores de proteínas enriquecidos nos sequenciamentos RIP-Anti-Ago2 foram isolados por RIP e sua abundância na fração RIP-Ago2 relativa ao RIP-Controle foi medida por qPCR. Em (C) está mostrado também o gene *TBCB*, que foi escolhido como controle negativo, e que o método de RIP-qPCR confirmou como não significativamente enriquecido. Os dados mostram a média ± D.P. de três a quatro réplicas biológicas.

A validação da interação entre RISC e os RNAs longos encontrados como enriquecidos no sequenciamento RIP-Anti-Ago2 também foi realizada por meio de outro método de imunoprecipitação denominado CLIP (*crosslink and immunoprecipitation*), seguido de qPCR. O método CLIP (Ule *et al.*, 2005) é usado como um ensaio alternativo mais estringente para testar as interações RNA-proteína *in vivo*, e está baseado na habilidade do UV-254 nm ligar covalentemente o RNA à proteína. Como o *crosslink* ocorre somente a curtas distâncias, e como neste método os

complexos são lavados com altas concentrações de sal, é possível se detectar com mais precisão as interações diretas entre RNA e proteína, podendo-se evitar eventuais artefatos de reassociação de RNAs a Ago2 durante o isolamento do RNA (Riley *et al.*, 2012).

Os mesmos transcritos validados por RIP-qPCR foram testados no ensaio de CLIP-qPCR e se mostraram enriquecidos nas frações CLIP imunoprecipitadas com anticorpo anti-Ago2 em relação à imunoprecipitação CLIP controle (figura 6).

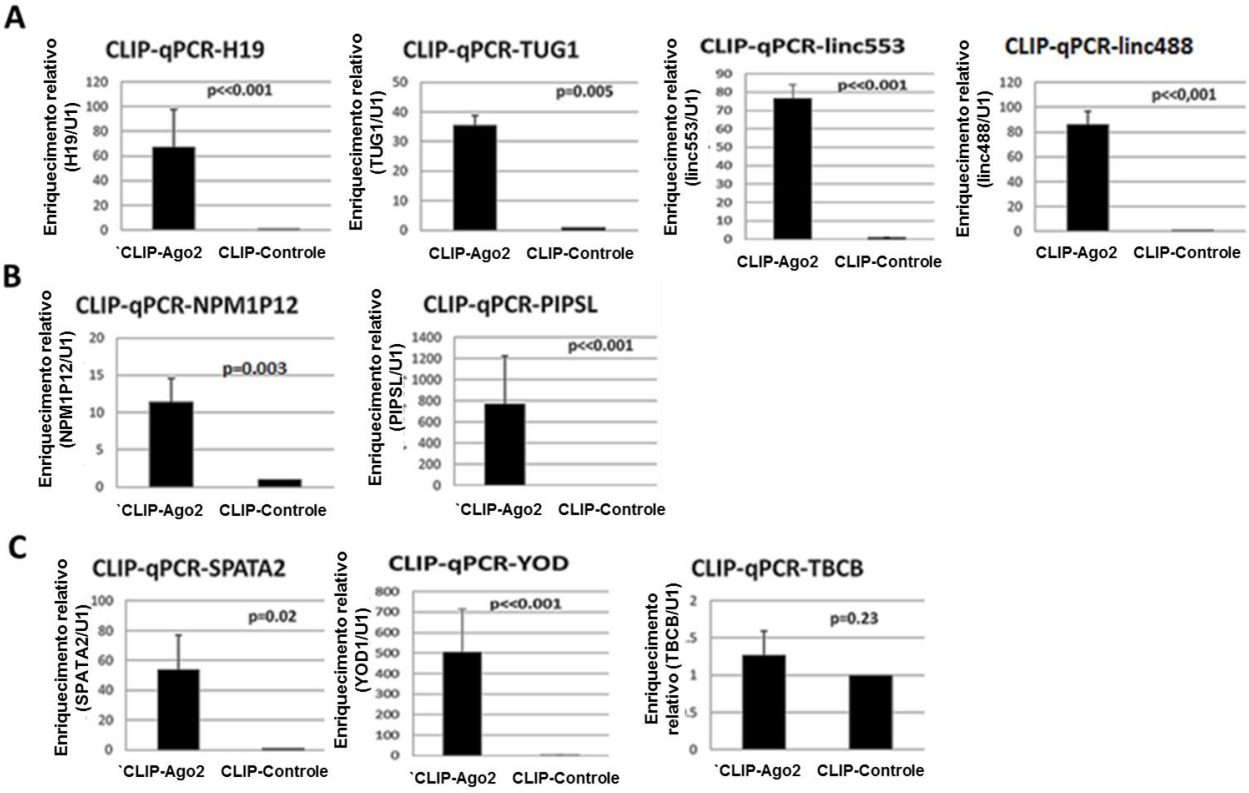

Figura 6. Validação por *crosslink and immunoprecipitation* (CLIP), seguido por PCR quantitativo em tempo real (CLIP-qPCR), da interação entre RISC e RNAs longos selecionados. (**A**) lincRNAs, (**B**) pseudogenes e (**C**) RNAs codificadores de proteínas enriquecidos nos sequenciamentos RIP-Anti-Ago2 foram isolados por CLIP e sua abundância no CLIP-Ago2 relativa ao CLIP-Controle foi medida por qPCR. Em (C) está mostrado também o gene *TBCB*, que foi escolhido como controle negativo, e que o

método de CLIP-qPCR confirmou como não significativamente enriquecido. Os dados mostram a média ± D.P. de três a quatro réplicas biológicas.

4.5 Identificação de miRNAs ligados ao complexo RISC

A fim de identificar miRNAs que possam alvejar os lncRNAs previamente encontrados ligados ao complexo RISC, parte dos RNAs co-imunoprecipitados nos dois ensaios RIP-Anti-Ago2 ou dois ensaios RIP-Controle em HeLa foi usada para o sequenciamento em larga escala de RNAs curtos. Foi feita uma terceira réplica biológica das imunoprecipitações e, portanto, foram sequenciados RNAs curtos de 3 réplicas biológicas do ensaio RIP-Anti-Ago2 e 3 réplicas biológicas do ensaio RIP-Controle. O sequenciamento foi realizado em uma canaleta do equipamento HiSeq2000, na plataforma Illumina.

Após a eliminação dos adaptadores das sequências obtidas, estas foram filtradas para retirada das sequências menores que 18 nucleotídeos ou com qualidade menor que Q30. A tabela 5 sumariza os dados relativos às sequências obtidas de RNAs curtos.

Tabela 5: Sumário dos resultados dos sequenciamentos das bibliotecas geradas com as amostras de RNAs curtos co-imunoprecipitados com anticorpo anti-Ago2 (RIP-Anti-Ago2) ou no ensaio controle (RIP-Controle).

Amostra	Sequências Totais	Número de sequências após remoção de adaptadores	Porcentagem de sequências após remoção de adaptadores	Número de sequências após filtragem	Porcentagem de sequências após filtragem
RIP-Anti-Ago2	77.453.338	76.653.505	99,0%	71.617.172	92,5%
RIP-Controle	52.563.959	50.852.949	96,7%	42.721.997	81,3%

Após a remoção dos adaptadores e filtragem das sequências por qualidade, foram obtidos RNAs curtos principalmente na faixa de 18 a 40 nucleotídeos de tamanho nas sequências de ambos os ensaios. Os RNAs curtos imunoprecipitados na RIP-Anti-Ago2 mostraram enriquecimento na faixa 21-24 nucleotídeos (Figure 7A), já sugerindo que miRNAs maduros são as principais espécies de RNAs curtos encontrados ligados ao complexo Ago2/RISC.

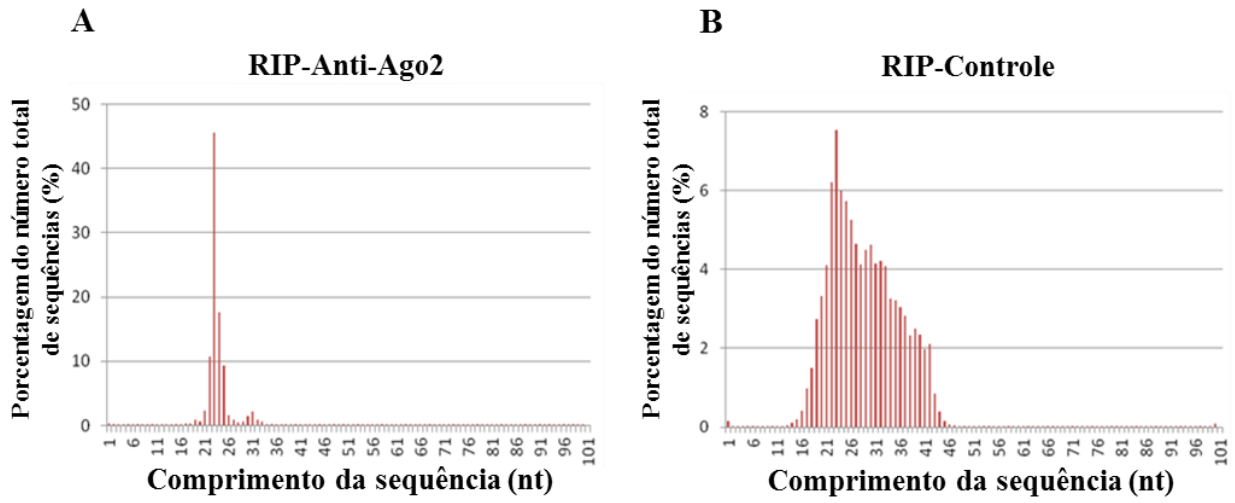

Figura 7. Distribuição de tamanhos dos RNAs curtos sequenciados obtidos a partir da imunoprecipitação com anticorpo Anti-Ago2 (RIP-Anti-Ago2) **(A)** ou imunoprecipitação controle, feita com IgG normal de camundongo (RIP-Controle) **(B)**.

Por outro lado, os RNAs imunoprecipitados com anticorpo IgG de camundongo (RIP-Controle) mostraram distribuição de tamanhos mais ampla, com sequências variando principalmente entre 15 e 45 nucleotídeos (figura 7B). Esse padrão de tamanhos já sugere que outras espécies de RNAs curtos ou produtos de degradação de RNAs mais longos (snoRNAs, tRNAs, rRNAs e mRNAs) podem ter sido obtidos após precipitação inespecífica que pode ocorrer durante o ensaio RIP-Controle.

Para averiguar essa possibilidade, nós buscamos, inicialmente, identificar o repertório de RNAs curtos sequenciados nas duas condições, a fim de comparar o perfil de RNAs curtos obtidos

a partir da RIP-Anti-Ago2 com o perfil da RIP-Controle. Para tanto, as sequências de RNAs curtos foram analisadas conforme descrito em (Hafner *et al.*, 2010). Para identificação dos RNAs curtos já conhecidos, as sequências filtradas foram alinhadas com as sequências do banco de dados de RNAs curtos descrito em (Berninger *et al.*, 2008), que contém um amplo repertório atualizado de RNAs curtos já catalogados. Para tanto, foi utilizado o algoritmo Oligomap (Berninger *et al.*, 2008).

A figura 8 mostra a distribuição das sequências em diferentes classes de RNAs curtos. A grande maioria dos RNAs curtos imunoprecipitados no ensaio RIP-Anti-Ago2 (cerca de 50%) corresponde a miRNAs maduros (figura 8A). Uma classe de miRNAs que se mostrou bastante abundante é a da família de miRNAs let-7, que contém membros bastante conservados entre espécies, muitos dos quais envolvidos com desenvolvimento e câncer (Roush e Slack, 2008). Outros RNAs contaminantes, tais como rRNA, tRNA e snRNAs e snoRNAs representam apenas uma pequena parcela (menos de 5%) dos RNAs sequenciados co-imunoprecipitados com Ago2/RISC. Um percentual de contaminação similar também foi observado em outro trabalho de imunoprecipitação de componentes do complexo RISC (Hafner *et al.*, 2010).

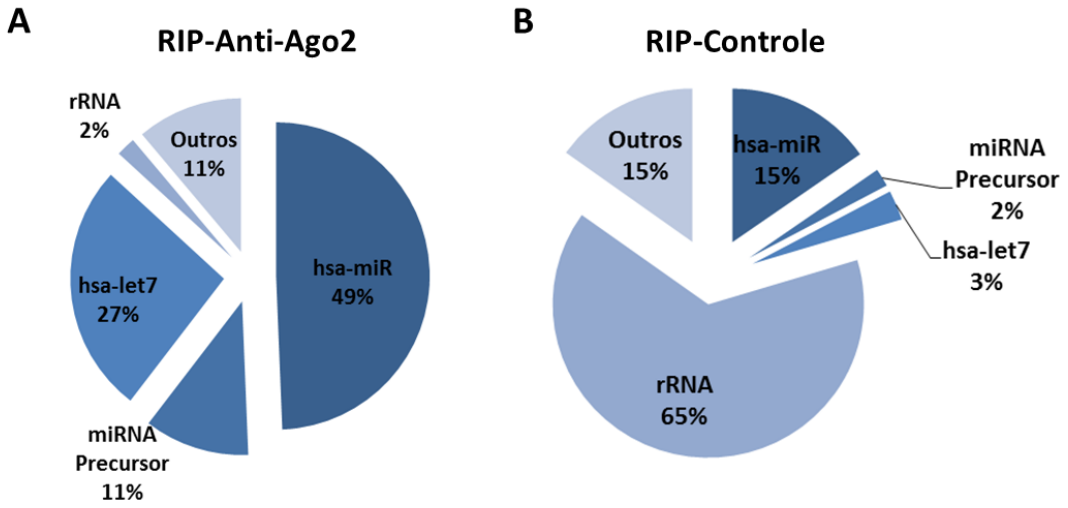

Figura 8. Diversidade dos RNAs curtos sequenciados obtidos a partir imunoprecipitação com anticorpo Anti-Ago2 (RIP-Anti-Ago2) **(A)** ou imunoprecipitação controle, feita com IgG normal de camundongo (RIP-Controle) **(B)**. As sequências de cada condição foram alinhadas com as sequências de um banco de dados de RNAs curtos conhecidos (Berninger *et al.*, 2008) e separadas de acordo com as classes às quais mapeavam.

Por outro lado, a maior parte dos RNAs encontrados na imunoprecipitação RIP-Contole correspondem a produtos de degradação de rRNAs (cerca de 65%), refletindo a abundância celular dos rRNAs (figura 8B). A quantidade de sequências alinhadas com outros RNAs, tais como tRNA, sn/snoRNA e mRNAs é maior que na imunoprecipitação RIP-Anti-Ago2, o que demonstra uma maior precipitação inespecífica destes RNAs no ensaio RIP-Controle.

4.5.1 Identificação dos miRNAs já anotados em bancos de dados

Em seguida, nós partimos para a identificação e quantificação dos miRNAs já anotados que se encontravam entre os RNAs curtos sequenciados. Para tanto, nós cruzamos as sequências de RNAs curtos obtidas com os miRNAs presentes na versão 19 do miRBase (Kozomara e Griffiths-Jones, 2011) utilizando o algoritmo miRDeep2 (Friedlander *et al.*, 2012). Foi encontrado um

conjunto de 812 miRNAs humanos conhecidos que foram aqui detectados como ligados ao RISC em células HeLa. Destes, 492 miRNAs foram encontrados exclusivamente nos sequenciamentos RIP-Anti-Ago2, enquanto 320 miRNAs foram detectados com um enriquecimento nos sequenciamentos RIP-Anti-Ago2 igual ou superior a quatro vezes em relação aos sequenciamentos RIP-Controle. A lista dos 812 miRNAs humanos detectados a partir de imunoprecipitações do complexo RISC em células HeLa e já presentes no banco de dados miRBase v.19, que foram anotados pelo algoritmo miRDeep2, está disponível no Arquivo 1 do Anexo, juntamente com o número de sequências detectadas para cada um desses miRNAs.

4.5.2 Identificação de possíveis novos miRNAs humanos

A fim de identificar possíveis novos miRNAs a partir das sequências de RNAs curtos ligados a Ago2/RISC obtidas em nosso sequenciamento, nós utilizamos miRDeep2 (Friedlander *et al.*, 2012). Para esta função, o miRDeep2 extrai as coordenadas genômicas que flanqueiam a sequência do possível novo miRNA e calcula a probabilidade de que essa região genômica dê origem a um precursor de miRNA baseado na sua estrutura secundária e na sequência usada como entrada no mapeamento. O algoritmo então calcula um *score*, que representa a probabilidade da sequência genômica ser um precursor daquele miRNA.

Utilizando o miRDeep2 (Friedlander *et al.*, 2012), nós identificamos 336 possíveis novos miRNAs com *score* maior ou igual a 5, um valor com nível de sinal-ruído seguro que corresponde a uma porcentagem de predição verdadeira maior que 90%, e que é recomendado pelos desenvolvedores do algoritmo (Friedlander *et al.*, 2012). Esses novos miRNAs podem representar novos reguladores de lncRNAs, de mRNAs e de importantes processos biológicos em células

humanas. A lista dos 336 novos miRNAs humanos encontrados, contendo as coordenadas genômicas dos possíveis precursores, está disponível no Arquivo 2 do Anexo.

4.6 Identificação de lncRNAs alvejados por miRNAs

A fim de identificarmos lncRNAs que sejam alvejados por miRNAs que também foram sequenciados em nosso experimentos RIP-Anti-Ago2, nós buscamos no banco de dados miRCode (Jeggari *et al.*, 2012) os lncRNAs por nós identificados e que possuem miRNAs preditos computacionalmente como possíveis reguladores. Em seguida, nós buscamos entre os miRNAs encontrados como possíveis reguladores de lncRNAs, aqueles que foram detectados em nossos sequenciamentos de RNAs curtos ligados ao complexo RISC. Nesta análise, optamos por realizar uma busca mais conservadora e utilizamos apenas os 207 miRNAs membros de famílias altamente conservadas de acordo com (Friedman *et al.*, 2009).

Considerando todos os 12.507 transcritos longos detectados em pelo menos um sequenciamento RIP-Anti-Ago2, nós identificamos 90 lncRNAs que possuem pelo menos um miRNA detectado em nossos sequenciamentos de RNAs curtos e que estava predito no miRCode como possível regulador do lncRNA correspondente. Entre estes 90 lncRNAs, encontram-se 15 lincRNAs, 1 RNA antissenso, 70 pseudogenes, 2 RNAs anotados como ncRNAs e 2 transcritos que geram miRNAs. A lista dos 90 lncRNAs que foram detectados em pelo menos um sequenciamento RIP-Anti-Ago2 e que possuem pelo menos um miRNA predito como regulador sequenciado em nossos ensaios está apresentada no Apêndice 2.

Já quando consideramos apenas os 686 lncRNAs que foram detectados exclusivamente no ensaios RIP-Anti-Ago2 ou com enriquecimento igual ou superior a quatro vezes em relação ao RIP-Controle, ou seja, os lncRNAs considerados como ligados ao complexo RISC, encontramos

30 lncRNAs que possuem pelo menos um dos 37 miRNAs detectados em nossos sequenciamentos de RNAs curtos e previsto no miRCode como possível regulador. Entre estes 30 lncRNAs, encontram-se 6 lincRNAs, 1 RNA antissenso, 20 pseudogenes, 1 RNA anotado como ncRNA e 2 genes que geram miRNAs. A Tabela 6 mostra a lista dos 30 lncRNAs ligados ao RISC que possuem pelo menos um miRNA como possível regulador também detectado ligado ao RISC em nossos sequenciamentos, assim como os 37 miRNAs que possivelmente alvejam cada um destes 30 lncRNAs.

Tabela 6: Relação dos 30 RNAs longos não-codificadores de proteínas detectados exclusivamente nos sequenciamentos RIP-Anti-Ago2 em HeLa e dos miRNAs sequenciados neste trabalho que os alvejam

Símbolo do lncRNA	**Classe (anotação do GENCODE)**	**Micro-RNAs detectados nos sequenciamentos e com predição de alvejar o lncRNA**
HCG14	antisense	miR-551[a]
LINC00476	lincRNA	miR-1244;miR-20b-5p;miR-590-5p;miR-219-5p;miR-876-3p; miR-372;miR-140-5p;miR-193a-3p;miR-152;miR-455-5p; miR-107;miR-301b
AC074391.1	lincRNA	miR-363-3p;miR-153
PART1	lincRNA	miR-4770;miR-363-3p;miR-142-3p;miR-449c-5p;miR-590-5p; miR-22-3p;miR-219-5p;miR-152;miR-301b
CROCCP2	lincRNA	miR-363-3p;miR-449c-5p;miR-20b-5p;miR-22-3p;miR-217; miR-551a; miR-372;miR-193a-3p;miR-455-5p;miR-107
TMEM191C	lincRNA	miR-363-3p;miR-1244;miR-22-3p;miR-876-3p;miR-217; miR-140-5p;miR-107
TUG1	lincRNA	miR-132;miR-137;miR-138;miR-140;miR-141;miR-142-3p; miR-143;miR-148b; miR-153;miR-190;miR-193;miR-216a; miR-219;miR-22;miR-23b-3p;miR-31;miR-103a;miR-33a-3p; miR-152
MIR100	miRNA	miR-210
AL441988.1	miRNA	miR-153;miR-363-3p

LINC00205	non_coding	miR-153
SCAMP1	Processed-transcript	miR-4770;miR-363-3p;miR-137;miR-142-3p;miR-153; miR-20b-5p; miR-210;miR-22-3p;miR-217;miR-551a;miR-215; miR-372;miR-33a-3p;miR-107
UBXN8	Processed-transcript	miR-217;miR-193a-3p;miR-590-5p
PCBP1-AS1	Processed-transcript	miR-363-3p;miR-142-3p;miR-1244;miR-20b-5p;miR-590-5p; miR-22-3p; miR-876-3p;miR-217;miR-215;miR-375; miR-193a-3p;miR-455-5p; miR-4770;miR-449c-5p; miR-219-5p; miR-372;miR-140-5p;miR-33a-3p;miR-107
BNIP3P1	Pseudogene	miR-137;miR-20b-5p;miR-210;miR-107;miR-301b
CIDECP	Pseudogene	miR-363-3p;miR-137;miR-20b-5p;miR-590-5p;miR-22-3p; miR-372;miR-152
OR7E14P	pseudogene	miR-4770
SDHAP3	pseudogene	miR-152;miR-107
GBAP1	pseudogene	miR-219-5p;miR-217;miR-153;miR-33a-3p;miR-152; miR-22-3p
HSD17B7P2	pseudogene	miR-4770;miR-20b-5p;miR-219-5p;miR-372
TMED11P	pseudogene	miR-375
HSP90AA4P	pseudogene	miR-215;miR-375;miR-193a-3p;miR-455-5p
PIP5K1P1	pseudogene	miR-142-3p;miR-372;miR-33a-3p;miR-107
CYP4F30P	pseudogene	miR-363-3p;miR-372
FAM86FP	pseudogene	miR-4770;miR-449c-5p;miR-372;miR-107
ALMS1P	pseudogene	miR-4770;miR-193a-3p;miR-210
KIAA0664L3	pseudogene	miR-4770;miR-142-3p;miR-449c-5p;miR-210;miR-22-3p; miR-215;miR-372
SLC6A10P	pseudogene	miR-363-3p;miR-137;miR-142-3p;miR-449c-5p; miR-1244; miR-22-3p; miR-876-3p;miR-375;miR-372;miR-140-5p; miR-193a-3p; miR-152
GGT3P	pseudogene	miR-219-5p;miR-363-3p;miR-137;miR-217;miR-33a-3p; miR-152; miR-22-3p
AMZ2P1	pseudogene	miR-4770;miR-363-3p;miR-142-3p;miR-449c-5p;miR-153; miR-215; miR-455-5p;miR-301b
RPS3AP47	pseudogene	miR-219-5p

4.7 Regulação do lincRNA *TUG1* pelo miRNA-148b

Como pode ser observado na Tabela 6, o lincRNA *TUG1* foi um dos lncRNAs detectados como ligados ao complexo RISC. *TUG1* é um lincRNA induzido por p53 em resposta ao dano do DNA (Guttman *et al.*, 2009) e em células humanas se liga a PRC2 (*Polycomb Repressive Complex 2*) e está envolvido na regulação de genes que controlam o ciclo celular e apoptose (Khalil *et al.*, 2009).

Entre os 812 miRNAs detectados nos nossos sequenciamentos de RNAs curtos (Arquivo 2 do Anexo), 19 miRNAs possuem *seeds* com sítios complementares em *TUG1* (Tabela 6). A fim de validar se o lincRNA *TUG1* é regulado por algum destes miRNAs, nós inicialmente selecionamos dois miRNAs que foram detectados nos nossos sequenciamentos cuja *seed* 5'-CAGUGC-3' é complementar ao sítio mais conservado para ligação de miRNAs de *TUG1*: o miR-148b e o miR-152. Estes miRNAs são parte da família mir-148/152, que tem sido envolvida no início e desenvolvimento de processos tumorais (Chen *et al.*, 2013). Esses miRNAs foram selecionados pois o sítio presente no lincRNA *TUG1* complementar à *seed* 5'-CAGUGC-3' encontrada nos miRNAs possui 100% de conservação entre primatas e 74% de conservação entre mamíferos. Essa alta conservação do sítio alvo é indicativo de funcionalidade da regulação miRNA-lncRNA.

A figura 9 mostra o *locus* genômico do lincRNA *TUG1*, assim como *tracks* correspondentes aos dois sequenciamentos de RNAs ligados a Ago2/RISC (RIP-Anti-Ago2-1 e RIP-Anti-Ago2-2) e aos dois sequenciamentos controle (RIP-Controle-1 e RIP-Controle-2). No topo da figura está mostrada a localização do *locus* no cromossomo humano 22 (traço vermelho). *Reads* mapeados no lincRNA *TUG1* detectados nos diferentes sequenciamentos estão mostrados como barras pretas. Logo abaixo dos *reads* das corridas RIP-Anti-Ago2 estão os miRNAs preditos como possíveis

reguladores de *TUG1* de acordo com o algoritmo miRCode (Jeggari *et al.*, 2012). Circulados em vermelho estão aqueles miRNAs que foram detectados no nosso sequenciamento de RNAs curtos. A seta vermelha indica os miRNAs da família mir-148/152.

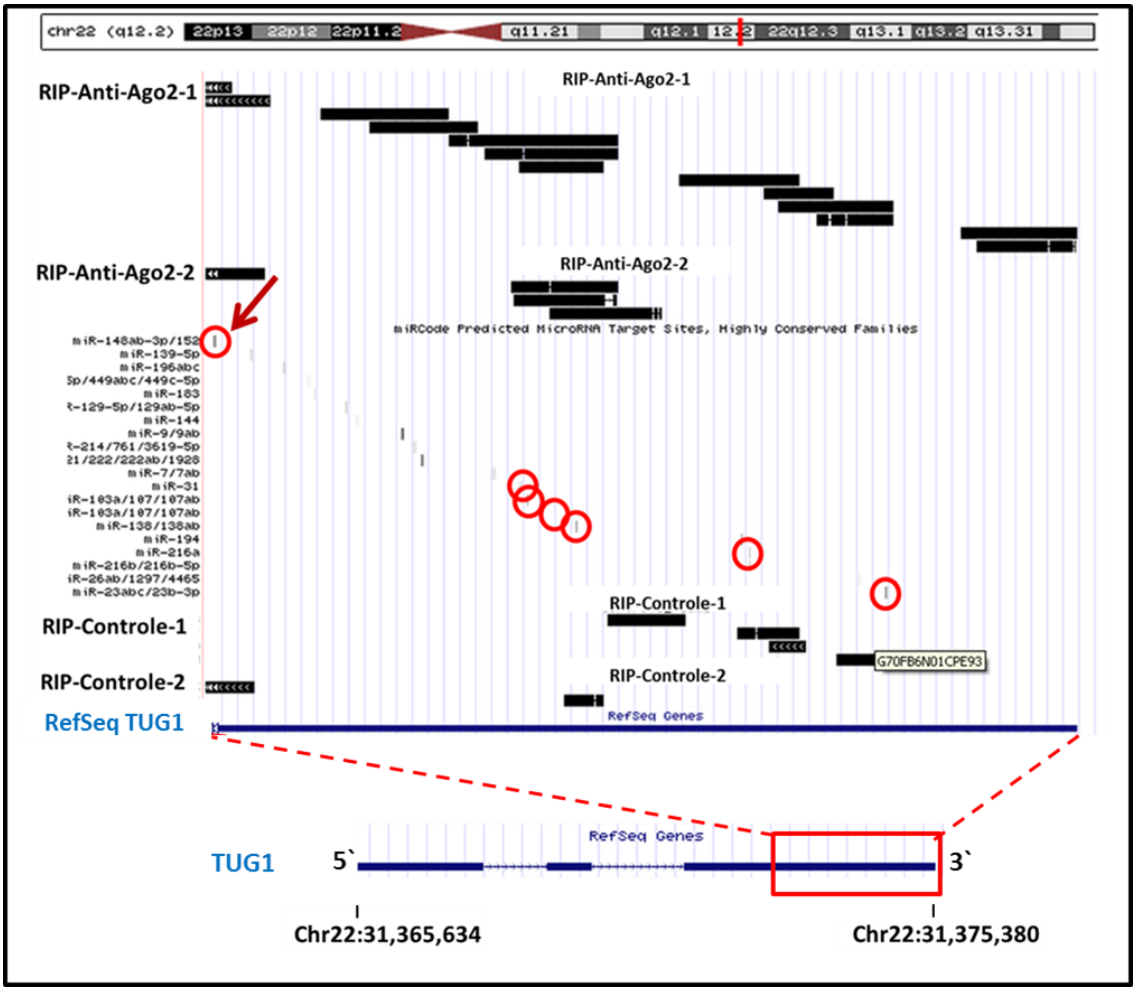

Figura 9. *Locus* genômico do lincRNA *TUG1* com seus possíveis miRNAs reguladores. No topo da figura está mostrada a localização do *locus* no cromossomo humano 22 (traço vermelho). Na base da figura, em azul, está mostrado o RefSeq do transcrito *TUG1*, e logo acima o trecho ampliado da porção 3'deste transcrito, onde está presente o sítio complementar ao miR-148b, além dos sítios complementares aos outros miRNAs indicados. Nas barras pretas no meio da figura estão mostrados os *reads* correspondentes aos dois sequenciamentos de RNAs ligados a Ago2/RISC (RIP-Anti-Ago2-1 e RIP-Anti-Ago2-2) e aos dois sequenciamentos controle (RIP-Controle-1 e RIP-Controle-2). Abaixo dos *reads* das corridas RIP-Anti-Ago2 estão os miRNAs preditos como possíveis reguladores de *TUG1* de acordo com o

algoritmo miRCode (Jeggari *et al.*, 2012). Circulados em vermelho estão aqueles miRNAs que foram detectados nos nossos sequenciamentos de RNAs curtos. A seta vermelha indica o sítio de interação dos miRNAs da família mir-148/152.

O miR-148b foi mostrado atuar como um supressor tumoral que inibe a proliferação celular e cuja expressão está diminuída em câncer gástrico (Song *et al.*, 2011). Além disso, o miR-148b é capaz de sensibilizar células tumorais à droga pró-apoptótica navitoclax (Lam *et al.*, 2010). Como o lincRNA *TUG1* tem um papel na regulação de genes que controlam o ciclo celular e a apoptose (Khalil *et al.*, 2009), nós investigamos se o miR-148b poderia regular os níveis de *TUG1* (figura 10A). Para tanto, células HeLa foram transfectadas em triplicata com 25nM de hsa-miR-148b ou com 25 nM de miRNA controle e, após 24h, o RNA total foi extraído e qPCR realizada para medida dos níveis dos RNAs. Observamos que o nível do RNA *TUG1* sofreu uma forte redução de 78% na transfecção feita com o miR-148b, quando comparada à transfecção controle (figura 10B). Por outro lado, o RNA estrutural *U1*, que não é esperado ser regulado pelo miR-148b, não sofreu alteração em seus níveis (figura 10B). Esses resultados mostram que o lincRNA *TUG1* está sendo regulado pelo miR-148b.

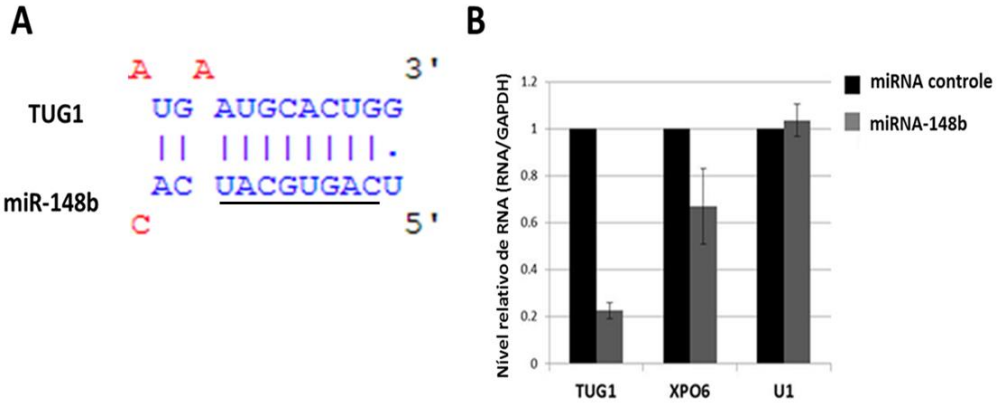

Figura 10. O lincRNA *TUG1* é regulado pelo miR-148b. A. Pareamento de bases entre o miR-148b e o lincRNA *TUG1* apontado pela predição *in silico*. No topo está mostrado o trecho da porção 3' do lincRNA *TUG1* contendo o sítio do tipo 7mer-m8 que é complementar ao trecho do miR-148b que contem a *seed*, indicada pelas bases sublinhadas na parte de baixo. **B.** Células HeLa foram transfectadas com hsa-miR-148b ou com miRNA controle (25nM, *Applied Biosystems*). Após 24h, o RNA total foi extraído e qPCR realizada para medida dos níveis dos RNAs. *XPO6* é um mRNA controle, predito *in silico* como regulado pelo miR-148b (Maragkakis *et al.*, 2009). O lncRNA *U1* é um RNA controle negativo, não esperado como sendo regulado por miR-148b. Os ensaios foram feitos em triplicata biológica; barras de erro referem-se ao desvio-padrão.

5. DISCUSSÃO

5.1 Desenvolvimento de método de geração de bibliotecas fita-específicas para sequenciamento de RNAs longos no 454/Roche

O desenvolvimento das plataformas de sequenciamento em larga escala de segunda geração iniciado na década passada (Mardis, 2013) tem permitido o aprofundamento do conhecimento em diversas áreas da biologia. Essas áreas incluem o estudo de organismos não-modelo (Rowan *et al.*, 2011), da metagenômica (Gevers *et al.*, 2012; Martins *et al.*, 2013) e principalmente da biologia humana, seja por meio do estudo da genômica (Koboldt *et al.*, 2013) e epigenômica (Rivera e Ren, 2013), seja através da utilização do sequenciamento de segunda geração na medicina personalizada (Wheeler *et al.*, 2008) ou na prática clínica, ainda que a real aplicabilidade neste contexto permaneça um alvo de intensa discussão (Biesecker *et al.*, 2012).

A fim de se estudar o conjunto de mensagens transcritas em determinada condição em certo organismo, as plataformas de sequenciamento de segunda geração têm sido utilizadas para o sequenciamento de cDNAs gerados a partir de RNAs (RNA-Seq) (Ozsolak e Milos, 2011). Essa técnica é usada para a identificação e quantificação da expressão diferencial de RNAs longos e curtos, para a caracterização do *splicing* alternativo, para a descoberta de transcritos novos, ainda não anotados, e para a detecção de genes que sejam produtos de fusão. Recentemente, a técnica de RNA-Seq também tem sido empregada no estudo das interações entre RNAs e proteínas (Konig *et al.*, 2011; Tuck e Tollervey, 2013), para caracterização do conteúdo de RNAs presente em complexos ribonucleoprotéicos.

Por um lado, muitas das limitações iniciais enfrentadas pelos sequenciamentos de segunda geração foram superadas, tais como o aumento do tamanho e da qualidade das sequências, o desenvolvimento de algoritmos para mapeamento rápido de sequências curtas e o aumento da reprodutibilidade e da acessibilidade às tecnologias e aos dados gerados (Nekrutenko e Taylor,

2012). No entanto, muitos desafios ainda permanecem. Entre esses desafios estão a obtenção de amostras com qualidade adequada que tenham sido coletadas sob aspectos legais e éticos bem estabelecidos, o barateamento dos custos e a simplificação da manipulação das amostras, o desenvolvimento de métodos de processamento das amostras que atendam às necessidades de aplicações distintas em diferentes plataformas de sequenciamento, a capacidade de armazenamento, de interpretação e integração dos dados gerados e a transposição dos achados para a prática clínica (Metzker, 2010). Nesse contexto, com o objetivo de permitir a aplicação de RNA-Seq de maneira fita-específica na plataforma 454/Roche (uma vez que não há *kits* comerciais disponíveis nem metodologia estabelecida para tal), nós desenvolvemos, neste trabalho, um método de geração de bibliotecas de cDNA direcionadas (DeMarco *et al.*, 2013).

Além da incorporação de direcionalidade nas bibliotecas de cDNA geradas, outra motivação para o desenvolvimento deste método de geração de bibliotecas de cDNA fita-específicas foi o barateamento dos custos de preparação dessas bibliotecas, uma tendência que vem sendo observada na literatura para outras plataformas de sequenciamento em larga escala (Dunham e Friesen, 2013). Apesar do aumento da disponibilidade de *kits* para geração de bibliotecas de cDNA direcionadas em algumas plataformas de sequenciamento de segunda geração, na maioria das vezes os custos destes *kits* ainda são elevados.

As estratégias de geração de bibliotecas fita-específicas para sequenciamento em larga escala se baseiam em três princípios: sequenciamento direto da primeira fita do cDNA (Lipson *et al.*, 2009; Ozsolak *et al.*, 2010), marcação química seletiva do RNA ou da segunda fita do cDNA (He *et al.*, 2008; Parkhomchuk *et al.*, 2009) e ligação de adaptadores de maneira assimétrica às extremidades de cDNAs ou de RNAs, metodologia esta utilizada na plataforma Illumina (Lister *et al.*, 2008). No que se refere à acurácia na determinação da direcionalidade dos transcritos, cada um desses métodos apresenta limitações particulares.

Como a enzima transcriptase reversa é capaz de gerar uma segunda fita espúria de cDNA através da sua atividade de DNA polimerase dependente de DNA (Spiegelman *et al.*, 1970; Gubler, 1987), os métodos de sequenciamento direto da primeira fita de cDNA podem fornecer informação incorreta a respeito da direcionalidade de um cDNA caso a segunda fita espúria seja gerada e sequenciada (Ozsolak e Milos, 2011). Os métodos de marcação química do RNA ou da segunda fita do cDNA dependem da eficiência desta marcação, como no caso em que o RNA é submetido a tratamento com bissulfito de sódio para conversão de citosinas em uracilas (He *et al.*, 2008). Após a conversão, a sequência do RNA tratado com bissulfito de sódio só pode representar uma das duas possíveis fitas de DNA, no entanto os RNAs que não sofrem conversão não podem ter a sua direcionalidade determinada (He *et al.*, 2008). Além disso, a análise computacional das sequências obtidas por esse método é mais trabalhosa, uma vez que as sequências devem ser alinhadas a dois genomas referência, um deles tendo todas as citosinas convertidas em timinas (He *et al.*, 2008).

O método por nós desenvolvido de geração de bibliotecas de cDNA direcionadas para sequenciamento no 454/Roche (DeMarco *et al.*, 2013) baseia-se na incorporação aos cDNAs de sequências de adaptadores com direcionalidade conhecida. Em geral, os métodos de geração de bibliotecas de cDNA direcionadas que se baseiam na ligação de adaptadores às extremidades de cDNAs ou de RNAs realizam a ligação de dois adaptadores, um em cada extremidade do RNA ou cDNA (Maher *et al.*, 2009). A principal alteração implementada no método aqui desenvolvido foi a incorporação da sequência do *primer* B da tecnologia 454/Roche no *primer* usado para transcrição reversa (DeMarco *et al.*, 2013). Com isso, o protocolo passou a ter somente um evento de ligação (do adaptador A à extremidade 3’ da primeira fita do cDNA), o que melhorou consideravelmente a acurácia do método na determinação da orientação dos transcritos obtidos. Nesse sentido, encontramos no método aqui desenvolvido uma taxa de erro de apenas 0,5% (DeMarco *et al.*, 2013)

para a proporção média de transcritos representando eventos artefatuais, identificados pela presença de *reads* que mapeiam na fita antissenso de genes codificadores e mostram sítios de *splicing* invertidos, ao passo que esta proporção foi de 5% em um método prévio de geração de bibliotecas direcionadas para sequenciamento na plataforma 454/Roche (Almeida *et al.*, 2012) que incluía o uso de ligase em dois eventos de adição de adaptadores.

Apesar das vantagens descritas para o método aqui desenvolvido (DeMarco *et al.*, 2013), algumas limitações ainda permanecem, como a quantidade inicial de RNA necessária para geração das bibliotecas, que pode ser limitante para alguns tipos de aplicações. Outro fator que deve ser observado pelos pesquisadores ao utilizar este método é a tendência da enzima DNA ligase em unir preferencialmente alguns tipos de sequências (Housby e Southern, 1998; Faulhammer *et al.*, 2000), o que pode gerar um viés caso o adaptador A se ligue prioritariamente a algum tipo específico de sequência de cDNA. Este viés da DNA ligase é pouco explorado na literatura, enquanto que os vieses sequência-específicos adicionados pelas RNA ligases nos métodos de geração de bibliotecas de sequenciamento na plataforma Illumina têm sido mais apontados (Jayaprakash *et al.*, 2011). Por fim, sabe-se que a etapa de amplificação das sequências presente no método desenvolvido neste trabalho pode introduzir artefatos como, por exemplo, vieses relacionados ao conteúdo GC de diferentes amostras (Dohm *et al.*, 2008; Kozarewa *et al.*, 2009).

Na realidade, os problemas acima descritos estão presentes em todas as tecnologias de sequenciamento em larga escala de segunda geração. Assim, para comparar sete diferentes métodos existentes de RNA-Seq fita-específicos, Levin e colaboradores avaliaram um total de treze bibliotecas de cDNA direcionadas utilizando o genoma de *Saccharomyces cerevisiae* como referência (Levin *et al.*, 2010). Foram avaliadas apenas bibliotecas direcionadas geradas para sequenciamento na plataforma Illumina (Levin *et al.*, 2010), dada a indisponibilidade de métodos estabelecidos ou *kits* para geração de bibliotecas direcionadas na plataforma 454/Roche. Das treze

bibliotecas analisadas, onze foram geradas e sequenciadas na plataforma Illumina GAIIx naquele trabalho (Levin *et al.*, 2010) e incluíam duas variações de quatro dos métodos, e as outras duas bibliotecas usadas eram de trabalhos prévios da literatura. Após o sequenciamento e a avaliação das performances das bibliotecas em termos de complexidade e fita-especificidade, além da regularidade da cobertura dos genes, concluiu-se que dois dos métodos apresentaram melhor desempenho: o método que se baseia na incorporação de dUTP à segunda fita do cDNA, com subsequente degradação desta fita com a enzima uracil-N-glicosilase e sequenciamento da primeira fita do cDNA (Parkhomchuk *et al.*, 2009) foi o que apresentou melhor desempenho; o segundo melhor método foi o protocolo de ligação de adaptadores ao RNA da tecnologia Illumina. Apesar da comparação do desempenho dos métodos de geração de bibliotecas fita-específicas realizada naquele trabalho (Levin *et al.*, 2010), estudos comparativos que permitam a caracterização dos artefatos e vieses específicos de cada técnica, como aqueles relacionados à amplificação das sequências ou à ligação dos adaptadores anteriormente citados, ainda precisam ser realizados.

Nos últimos anos surgiram novas tecnologias de sequenciamento em larga escala com grande potencial para superação das limitações apresentadas pelas atuais tecnologias. Entre essas novas tecnologias estão o sequenciamento direto do RNA (Ozsolak *et al.*, 2009) e o uso de nanoporos para detecção das bases por meio de alterações na corrente elétrica conduzida por um canal à medida que cada uma das bases passa através deste poro (Schneider e Dekker, 2012). A implantação definitiva destas tecnologias ainda depende atualmente da superação de limitações técnicas e da disponibilização comercial de equipamentos para testes na rotina laboratorial (Metzker, 2010).

5.2 lncRNAs regulados por miRNAs

Trabalhos e projetos multicêntricos que se iniciaram na última década revelaram que cerca de 90% do genoma humano pode ser transcrito, mas apenas 2% é transcrito e traduzido em proteínas (Birney *et al.*, 2007; Banfai *et al.*, 2012; Djebali *et al.*, 2012). Esses transcritos não traduzidos representam RNAs não-codificadores de proteínas e compõem um vasto e heterogêneo conjunto de RNAs curtos e longos, cujas funções na célula têm sido relacionadas principalmente à regulação da expressão de genes codificadores de proteínas (Sabin *et al.*, 2013).

Embora a descoberta de lncRNAs específicos seja anterior à descoberta dos miRNAs (Brannan *et al.*, 1990; Brown *et al.*, 1991; Lee *et al.*, 1993), somente recentemente um maior entendimento a respeito dos lncRNAs tem sido alcançado (Rinn e Chang, 2012). Se por um lado tem-se avançado na elucidação da estrutura, mecanismos de ação e funções dos lncRNAs (Ponting *et al.*, 2009), por outro lado muito pouco se sabe sobre como os lncRNAs são regulados (Geisler *et al.*, 2012). Uma possibilidade é que os lncRNAs sejam alvejados e silenciados por miRNAs como forma de regulação pós-transcricional. A fim de avaliar esta hipótese, nós encontramos e estabelecemos, no nosso trabalho, um catálogo de 686 lncRNAs ligados ao complexo RISC, compreendo RNAs antissenso, precursores de miRNAs e pseudogenes, entre os 2.940 transcritos encontrados ligados a este complexo.

Recentemente, duas ferramentas de predição de alvos de miRNAs publicadas passaram a incorporar algoritmos que visam analisar lncRNAs como possíveis alvos de miRNAs: o miRCode (Jeggari *et al.*, 2012) buscou sequências complementares a *seeds* de miRNAs em 10.419 lncRNAs de humanos e gerou um mapa de possíveis alvos de miRNAs neste conjunto. Para estender as possíveis interações entre miRNAs e lncRNAs, o DIANA-LncBase (Paraskevopoulou *et al.*, 2013) analisou, além de lncRNAs de humanos, lncRNAs de camundongo como possíveis alvos de

miRNAs. Apesar das predições computacionais serem úteis em uma etapa exploratória de busca por miRNAs com possibilidade de regularem RNAs longos, diferentes algoritmos geram resultados divergentes em função de critérios distintos utilizados e o número de falso-positivos nestas predições ainda é muito alto (Rajewsky, 2006; Witkos *et al.*, 2011). Por isso, a regulação exercida pelo miRNA deve ser avaliada e validada experimentalmente (Chi *et al.*, 2009). Nesse sentido, torna-se importante o uso de técnicas que sejam capazes de detectar a interação entre RNAs longos, miRNAs e o complexo RISC, como a imunoprecipitação do complexo RISC usada no nosso trabalho.

Relatos da literatura já observavam que uma parte dos RNAs ligados às proteínas argonauta eram transcritos a partir de regiões não codificadoras do genoma de mamíferos (Chi *et al.*, 2009; Hafner *et al.*, 2010), no entanto nenhum destes trabalhos analisou os lncRNAs entre os RNAs detectados. Chi e colaboradores realizaram imunoprecipitação da proteína Ago2 após *crosslink* do tecido cerebral de camundongos e sequenciaram os RNAs co-imunoprecipitados em larga escala (Chi *et al.*, 2009). Foi observado que 6% das sequências obtidas correspondiam a RNAs provenientes de regiões intergênicas do genoma do camundongo, 5% correspondiam a lincRNAs anotados em (Guttman *et al.*, 2009) ou outros RNAs não-codificadores de proteínas anotados no projeto FANTOM3 (Maeda *et al.*, 2006) e 12% mapeavam em regiões intrônicas de genes codificadores de proteínas e poderiam, portanto, representar RNAs não-codificadores de proteínas longos transcritos dentro dos *loci* de genes codificadores de proteínas (Chi *et al.*, 2009). No nosso trabalho, encontramos proporções comparáveis de lncRNAs ligados ao RISC: entre os 2.940 transcritos encontrados ligados ao RISC, 8% representam lincRNAs, 9% representam RNAs antissenso e 5% representam pseudogenes, totalizando 22% de todos os transcritos detectados ligados ao RISC (686 lncRNAs no total).

Outro trabalho (Hafner *et al.*, 2010) realizou imunoprecipitação das proteínas Ago1, Ago2, Ago3 e Ago4 utilizando a técnica de PAR-CLIP (*Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation*) (Hafner *et al.*, 2012) em células embrionárias de rim HEK293 e sequenciaram em larga escala os RNAs co-imunoprecipitados. Novamente, os lncRNAs não foram analisados neste trabalho (Hafner *et al.*, 2010). No entanto, os autores observaram que 14% das sequências obtidas mapeavam em regiões intrônicas de genes codificadores de proteínas RefSeq (Pruitt *et al.*, 2007) e poderiam, portanto, representar lncRNAs intrônicos ou antissenso. Além disso, entre as sequências mapeadas em regiões exônicas, 4% correspondiam à 5' UTR, 50% à CDS e 46% à 3'UTR (Hafner *et al.*, 2010). A técnica de PAR-CLIP permite a identificação da sequência que está diretamente em contato com a proteína imunoprecipitada, uma vez que é feito *crosslink* entre o RNA e a proteína e ocorre digestão com RNases do RNA não ligado (Hafner *et al.*, 2012). A identificação da sequência específica do RNA ligada à proteína não é possível no caso da imunoprecipitação nativa (RIP), realizada em nosso trabalho, pois não é feito *crosslink* para que sejam capturadas interações mais estáveis entre os RNAs e as proteínas, e não é feita digestão com RNases para se recuperar apenas o RNA que não é digerido por estar ligado à proteína (Zhao *et al.*, 2010). Apesar disso, o mapeamento das sequências obtidas no nosso trabalho nos genes RefSeq (Pruitt *et al.*, 2007) mostrou proporções de mapeamento similares às encontradas em (Hafner *et al.*, 2010): cerca de 2% das sequências mapearam na 5'UTR, 52% mapearam em CDS, 41% mapearam em regiões 3'UTR, e 5% mapearam em regiões intrônicas (dados não mostrados). Esses achados confirmam observações da literatura que sugerem sobreposição nos RNAs detectados nos ensaios de PAR-CLIP e RIP (Mukherjee *et al.*, 2011).

Recentemente, Jalali e colaboradores (Jalali *et al.*, 2013) analisaram as sequências geradas no trabalho de (Hafner *et al.*, 2010) na busca de lncRNAs ligados às proteínas Ago1, Ago2, Ago3 e Ago4 em células embrionárias de rim HEK293. Foram encontrados 113 lncRNAs ligados a pelo

menos uma destas proteínas e destes, 25 lncRNAs possuíam algum miRNA como possível regulador. No nosso trabalho, foram detectados 686 lncRNAs nos sequenciamentos de RNAs co-imunoprecipitados com o complexo RISC e, destes, 30 lncRNAs possuem algum miRNA como possível regulador. A baixa proporção de lncRNAs que possuem miRNAs como possíveis reguladores [22% no trabalho de (Jalali *et al.*, 2013) e 4,3% no nosso trabalho] pode estar relacionada ao fato de que outros RNAs curtos são capazes de se ligar a proteínas argonauta e talvez possam contribuir para a regulação dos lncRNAs (Czech e Hannon, 2011). Além disso, no nosso trabalho, optamos por fazer uma busca mais conservadora por interações miRNA-lncRNA. Para isso, buscamos apenas por miRNAs de famílias com alta conservação de acordo com (Friedman *et al.*, 2009), além de termos buscado miRNAs possíveis reguladores dos lncRNAs apenas entre aqueles que também foram sequenciados em HeLa nos nossos próprios ensaios.

A fim de identificar possíveis interações entre lincRNAs e miRNAs, um recente estudo buscou padrões de expressão correlacionados entre miRNAs e lincRNAs em tecidos não-tumorais e tumorais de mama (Juan *et al.*, 2013). Foram encontrados nestes tecidos 90 lincRNAs com correlação de expressão invertida com miRNAs, os quais possuíam papéis conhecidos no câncer de mama e pelo menos um sítio alvo predito presente no lincRNA (Juan *et al.*, 2013). A busca pelos padrões de expressão apresentados pelos pares lncRNA-miRNA encontrados no nosso trabalho em dados públicos contendo amostras tumorais e não-tumorais, assim como a avaliação experimental da possível funcionalidade dessas interações poderão indicar possíveis papéis destes pares no estabelecimento e/ou desenvolvimento de tumores.

Os papéis funcionais da regulação dos lncRNAs pelos miRNAs ainda são pouco conhecidos. Até o momento, o mecanismo mais caracterizado de interação miRNA-lncRNA é aquele em que os lncRNAs atuam como esponjas naturais (Ebert e Sharp, 2010). Neste mecanismo, um lncRNA compete com um mRNA por miRNAs que o alvejam (Franco-Zorrilla *et al.*, 2007;

Cazalla *et al.*, 2010; Poliseno *et al.*, 2010; Cesana *et al.*, 2011; Hansen *et al.*, 2013; Kallen *et al.*, 2013; Memczak *et al.*, 2013). Nós não descartamos a possibilidade de que alguns dos lncRNAs encontrados no nosso trabalho, além dos pseudogenes, atuem como esponjas. De fato, foi mostrado recentemente que o lncRNA *H19*, encontrado nos nossos sequenciamentos ligado ao RISC, age como uma esponja molecular modulando a disponibilidade de miRNAs da família let-7 em células humanas de carcinoma embrionárias PA.1 e impedindo a diferenciação de mioblastos C1C12 de camundongo induzida por let-7 (Kallen *et al.*, 2013).

Por outro lado, é provável que muitos lncRNAs sejam regulados a fim de que a sua própria função seja modulada, dado que muitos lncRNAs possuem diversas funções na célula, tais como sinais, guias e blocos de sustentação para proteínas (Wang e Chang, 2011). Neste sentido, escolhemos o lincRNA *TUG1* para avaliação da sua possível regulação pelo miR-148b. *TUG1* tem expressão ativada pela proteína p53 em resposta ao dano do DNA em células de camundongo (Guttman *et al.*, 2009), e em células humanas se liga ao PRC2 e direciona este complexo proteico para a repressão de genes relacionados ao ciclo celular ou apoptose (Khalil *et al.*, 2009). O miR-148b foi mostrado como sendo um supressor tumoral que inibe a proliferação celular e cuja expressão está diminuída em câncer gástrico (Song *et al.*, 2011), além de sensibilizar células tumorais à droga pró-apoptótica navitoclax (Lam *et al.*, 2010). Mostramos que a superexpressão do miR-148b reduz significativamente os níveis do lincRNA *TUG1*. Será importante confirmar que o miR-148b regula o nível do lncRNA *TUG1* por meio de uma interação direta usando-se ensaio de gene repórter, além de avaliar a implicação funcional desta regulação na célula. Será também interessante estudar a regulação exercida por miRNAs sobre outros lncRNAs aqui identificados como candidatos, e que já estão descritos na literatura como exercendo diversas funções, tais como o recrutamento de proteínas modificadoras da cromatina ou o recrutamento de proteínas envolvidas em *splicing* (Wang e Chang, 2011; Beckedorff *et al.*, 2013).

Neste sentindo, dois trabalhos iniciaram recentemente a caracterização da regulação exercida por miRNAs sobre lncRNAs (Leucci *et al.*, 2013; Zhang *et al.*, 2013). Leucci e colaboradores mostraram, por meio de superexpressão e silenciamento do miR-9, que este miRNA regula os níveis do lncRNA *MALAT-1* na linhagem celular de linfoma de Hodgkin L428, apesar da implicação dessa regulação na função do *MALAT-1* não ter sido avaliada (Leucci *et al.*, 2013). Em outro trabalho, foi mostrado que o lncRNAs *GAS5* é negativamente regulado pelo miR-21 (Zhang *et al.*, 2013). Foi mostrado que a superexpressão e o silenciamento do miR-21 diminuem e aumentam, respectivamente, os níveis do *GAS5* em células tumorais de mama MCF-7, e que esses ncRNAs apresentam padrões de expressão invertidos em tumores de mama e em linhagens celulares tumorais e não-tumorais de mama (Zhang *et al.*, 2013). Apesar disso, os autores não descartam a possibilidade de que o *GAS5* haja como uma esponja para o miR-21, dado que são conhecidos outros RNAs longos como reguladores do miR-21 (Kim *et al.*, 2011).

Um dos critérios utilizados por nós para seleção dos miRNAs possíveis reguladores do lincRNA *TUG1* foi a alta conservação do sítio complementar ao miRNA no lncRNA alvo. O sítio complementar ao miR-148b no lincRNA *TUG1* possui 100% de conservação entre primatas e 74% de conservação entre mamíferos. Na contramão das indicações atuais de regulação de lncRNAs por miRNAs, um recente trabalho mostrou existir uma evidência bastante limitada de alvos evolutivamente conservados de miRNAs em lincRNAs (Alaei-Mahabadi e Larsson, 2013). Neste trabalho, os autores analisaram a presença de sítios complementares a *seeds* de miRNAs em mRNAs e lincRNAs. No caso do mRNAs, foram encontrados enriquecimentos variando de 1,6 a 6,1 vezes de sítios alvo conservados de miRNAs em relação a sítios randômicos, principalmente na região 3'UTR do mRNA, conforme descrito na literatura (Friedman *et al.*, 2009). No caso dos lincRNAs, foi analisado um conjunto de 2.121 lincRNAs e não foi observado enriquecimento de sítios alvo conservados de miRNAs nem ao longo de todo o transcrito, nem nas extremidades 3'

ou 5’ desses lincRNAs (Alaei-Mahabadi e Larsson, 2013). Foi observado um discreto enriquecimento de 1,8 vezes apenas em um grupo de 169 lncRNAs predominantemente citoplasmáticos, dos quais 8 lincRNAs possuíam 12 sítios complementares a miRNAs conservados (Alaei-Mahabadi e Larsson, 2013). Existem evidências na literatura de que lncRNAs são regulados por miRNAs (Hansen *et al.*, 2011; Leucci *et al.*, 2013; Zhang *et al.*, 2013), e os dados encontrados no nosso trabalho aumentam essas evidências. Dada a conhecida diversidade de classes de lncRNAs (Rinn e Chang, 2012), os resultados de Alaei-Mahabad e colaboradores podem ser interpretados como uma sugestão de que existam classes ou subfamílias específicas de lncRNAs que possam ser reguladas por miRNAs. Estudos que avaliem a contribuição de miRNAs para a regulação de lncRNAs com diferentes graus de conservação, incluindo lncRNAs não conservados e que tenham evoluído recentemente, tais como *HOTAIR* (Schorderet e Duboule, 2011) e *XIST* (Nesterova *et al.*, 2001), irão contribuir para o entendimento da diversidade de interações entre miRNAs e lncRNAs.

6 CONCLUSÕES

Neste trabalho, desenvolvemos um método de geração de bibliotecas de cDNA direcionadas para sequenciamento em larga escala na plataforma 454/Roche que se mostrou mais acurado na determinação da orientação dos transcritos sequenciados que um método previamente existente na literatura. O método aqui desenvolvido está disponível para ser usado pela comunidade científica em experimentos que envolvam o sequenciamento em larga escala de transcritos na plataforma 454/Roche com a determinação da direcionalidade dos RNAs sequenciados, e foi útil na identificação de RNAs longos não codificadores de proteínas (lncRNAs).

O método de geração de bibliotecas de cDNA direcionadas aqui desenvolvido foi usado na identificação de lncRNAs ligados ao complexo RISC na linhagem celular HeLa. Dada a crescente participação de lncRNAs em processos normais e patológicos e o recente desenvolvimento de estratégias terapêuticas baseadas em miRNAs, torna-se importante avaliar a regulação de lncRNAs por miRNAs. Nesse sentido, catalogamos neste trabalho 686 lncRNAs que se ligam ao complexo RISC em células HeLa. Destes, 30 lncRNAs têm potencial de serem alvejados por miRNAs que também foram detectados como ligados ao RISC em nossos sequenciamentos de RNAs curtos. Esse catálogo abre amplas possibilidades de estudo de pares lncRNA-miRNA que podem participar de processos fisiológicos em células humanas.

Por fim, mostramos por meio de superexpressão do miR-148b em células HeLa que o lncRNA *TUG1* é regulado por este miRNA. Abordagens experimentais, tais como a avaliação do impacto da modulação dos níveis do miR-148b e do lncRNA *TUG1* sobre a apoptose e o ciclo celular, devem ser realizadas a fim de que se determine como essa regulação tem implicações funcionais no contexto celular.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Adams, M. D.; Celniker, S. E.; Holt, R. A.; Evans, C. A.; Gocayne, J. D.; Amanatides, P. G.; Scherer, S. E.; Li, P. W.; Hoskins, R. A.; Galle, R. F.; George, R. A.; Lewis, S. E.; Richards, S.; Ashburner, M.; Henderson, S. N.; Sutton, G. G.; Wortman, J. R.; Yandell, M. D.; Zhang, Q.; Chen, L. X.; Brandon, R. C.; Rogers, Y. H.; Blazej, R. G.; Champe, M.; Pfeiffer, B. D.; Wan, K. H.; Doyle, C.; Baxter, E. G.; Helt, G.; Nelson, C. R.; Gabor, G. L.; Abril, J. F.; Agbayani, A.; An, H. J.; Andrews-Pfannkoch, C.; Baldwin, D.; Ballew, R. M.; Basu, A.; Baxendale, J.; Bayraktaroglu, L.; Beasley, E. M.; Beeson, K. Y.; Benos, P. V.; Berman, B. P.; Bhandari, D.; Bolshakov, S.; Borkova, D.; Botchan, M. R.; Bouck, J.; Brokstein, P.; Brottier, P.; Burtis, K. C.; Busam, D. A.; Butler, H.; Cadieu, E.; Center, A.; Chandra, I.; Cherry, J. M.; Cawley, S.; Dahlke, C.; Davenport, L. B.; Davies, P.; de Pablos, B.; Delcher, A.; Deng, Z.; Mays, A. D.; Dew, I.; Dietz, S. M.; Dodson, K.; Doup, L. E.; Downes, M.; Dugan-Rocha, S.; Dunkov, B. C.; Dunn, P.; Durbin, K. J.; Evangelista, C. C.; Ferraz, C.; Ferriera, S.; Fleischmann, W.; Fosler, C.; Gabrielian, A. E.; Garg, N. S.; Gelbart, W. M.; Glasser, K.; Glodek, A.; Gong, F.; Gorrell, J. H.; Gu, Z.; Guan, P.; Harris, M.; Harris, N. L.; Harvey, D.; Heiman, T. J.; Hernandez, J. R.; Houck, J.; Hostin, D.; Houston, K. A.; Howland, T. J.; Wei, M. H.; Ibegwam, C.; Jalali, M.; Kalush, F.; Karpen, G. H.; Ke, Z.; Kennison, J. A.; Ketchum, K. A.; Kimmel, B. E.; Kodira, C. D.; Kraft, C.; Kravitz, S.; Kulp, D.; Lai, Z.; Lasko, P.; Lei, Y.; Levitsky, A. A.; Li, J.; Li, Z.; Liang, Y.; Lin, X.; Liu, X.; Mattei, B.; McIntosh, T. C.; McLeod, M. P.; McPherson, D.; Merkulov, G.; Milshina, N. V.; Mobarry, C.; Morris, J.; Moshrefi, A.; Mount, S. M.; Moy, M.; Murphy, B.; Murphy, L.; Muzny, D. M.; Nelson, D. L.; Nelson, D. R.; Nelson, K. A.; Nixon, K.; Nusskern, D. R.; Pacleb, J. M.; Palazzolo, M.; Pittman, G. S.; Pan, S.; Pollard, J.; Puri, V.; Reese, M. G.; Reinert, K.; Remington, K.; Saunders, R. D.; Scheeler, F.; Shen, H.; Shue, B. C.; Siden-Kiamos, I.; Simpson, M.; Skupski, M. P.; Smith, T.; Spier, E.; Spradling, A. C.; Stapleton, M.; Strong, R.; Sun, E.; Svirskas, R.; Tector, C.; Turner, R.; Venter, E.; Wang, A. H.; Wang, X.; Wang, Z. Y.; Wassarman, D. A.; Weinstock, G. M.; Weissenbach, J.; Williams, S. M.; WoodageT; Worley, K. C.; Wu, D.; Yang, S.; Yao, Q. A.; Ye, J.; Yeh, R. F.; Zaveri, J. S.; Zhan, M.; Zhang, G.; Zhao, Q.; Zheng, L.; Zheng, X. H.; Zhong, F. N.; Zhong, W.; Zhou, X.; Zhu, S.; Zhu, X.; Smith, H. O.; Gibbs, R. A.; Myers, E. W.; Rubin, G. M. ; Venter, J. C. (2000). "The genome sequence of Drosophila melanogaster." Science 287(5461): 2185-2195.

Adey, A.; Burton, J. N.; Kitzman, J. O.; Hiatt, J. B.; Lewis, A. P.; Martin, B. K.; Qiu, R.; Lee, C. ; Shendure, J. (2013). "The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line." Nature 500(7461): 207-211.

Alaei-Mahabadi, B. ; Larsson, E. (2013). "Limited evidence for evolutionarily conserved targeting of long non-coding RNAs by microRNAs." Silence 4(1): 4.

Almeida, G. T.; Amaral, M. S.; Beckedorff, F. C.; Kitajima, J. P.; DeMarco, R. ; Verjovski-Almeida, S. (2012). "Exploring the Schistosoma mansoni adult male transcriptome using RNA-seq." Exp Parasitol 132(1): 22-31.

Amaral, P. P. ; Mattick, J. S. (2008). "Noncoding RNA in development." Mamm Genome 19(7-8): 454-492.

Ambros, V. (2004). "The functions of animal microRNAs." Nature 431(7006): 350-355.

Amort, T.; Souliere, M. F.; Wille, A.; Jia, X. Y.; Fiegl, H.; Worle, H.; Micura, R. ; Lusser, A. (2013). "Long non-coding RNAs as targets for cytosine methylation." RNA Biol 10(6): 1003-1008.

Arvey, A.; Larsson, E.; Sander, C.; Leslie, C. S. ; Marks, D. S. (2010). "Target mRNA abundance dilutes microRNA and siRNA activity." Mol Syst Biol 6: 363.

Azzalin, C. M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E. ; Lingner, J. (2007). "Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends." Science 318(5851): 798-801.

Babiarz, J. E.; Ruby, J. G.; Wang, Y.; Bartel, D. P. ; Blelloch, R. (2008). "Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs." Genes Dev 22(20): 2773-2785.

Banfai, B.; Jia, H.; Khatun, J.; Wood, E.; Risk, B.; Gundling, W. E., Jr.; Kundaje, A.; Gunawardena, H. P.; Yu, Y.; Xie, L.; Krajewski, K.; Strahl, B. D.; Chen, X.; Bickel, P.; Giddings, M. C.; Brown, J. B. ; Lipovich, L. (2012). "Long noncoding RNAs are rarely translated in two human cell lines." Genome Res 22(9): 1646-1657.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell 136(2): 215-233.

Batista, P. J. ; Chang, H. Y. (2013). "Long noncoding RNAs: cellular address codes in development and disease." Cell 152(6): 1298-1307.

Beckedorff, F. C.; Amaral, M. S.; Deocesano-Pereira, C. ; Verjovski-Almeida, S. (2013). "Long non-coding RNAs and their implications in cancer epigenetics." Biosci Rep 33(4).

Beckedorff, F. C.; Ayupe, A. C.; Crocci-Souza, R.; Amaral, M. S.; Nakaya, H. I.; Soltys, D. T.; Menck, C. F.; Reis, E. M. ; Verjovski-Almeida, S. (2013). "The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation." PLoS Genet 9(8): e1003705.

Beltran, M.; Puig, I.; Pena, C.; Garcia, J. M.; Alvarez, A. B.; Pena, R.; Bonilla, F. ; de Herreros, A. G. (2008). "A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition." Genes Dev 22(6): 756-769.

Berninger, P.; Gaidatzis, D.; van Nimwegen, E. ; Zavolan, M. (2008). "Computational analysis of small RNA cloning data." Methods 44(1): 13-21.

Besse, F. ; Ephrussi, A. (2008). "Translational control of localized mRNAs: restricting protein synthesis in space and time." Nat Rev Mol Cell Biol 9(12): 971-980.

Biesecker, L. G.; Burke, W.; Kohane, I.; Plon, S. E. ; Zimmern, R. (2012). "Next-generation sequencing in the clinic: are we ready?" Nat Rev Genet 13(11): 818-824.

Birney, E.; Stamatoyannopoulos, J. A.; Dutta, A.; Guigo, R.; Gingeras, T. R.; Margulies, E. H.; Weng, Z.; Snyder, M.; Dermitzakis, E. T.; Thurman, R. E.; Kuehn, M. S.; Taylor, C. M.; Neph, S.; Koch, C. M.; Asthana, S.; Malhotra, A.; Adzhubei, I.; Greenbaum, J. A.; Andrews, R. M.; Flicek, P.; Boyle, P. J.; Cao, H.; Carter, N. P.; Clelland, G. K.; Davis, S.;

Day, N.; Dhami, P.; Dillon, S. C.; Dorschner, M. O.; Fiegler, H.; Giresi, P. G.; Goldy, J.; Hawrylycz, M.; Haydock, A.; Humbert, R.; James, K. D.; Johnson, B. E.; Johnson, E. M.; Frum, T. T.; Rosenzweig, E. R.; Karnani, N.; Lee, K.; Lefebvre, G. C.; Navas, P. A.; Neri, F.; Parker, S. C.; Sabo, P. J.; Sandstrom, R.; Shafer, A.; Vetrie, D.; Weaver, M.; Wilcox, S.; Yu, M.; Collins, F. S.; Dekker, J.; Lieb, J. D.; Tullius, T. D.; Crawford, G. E.; Sunyaev, S.; Noble, W. S.; Dunham, I.; Denoeud, F.; Reymond, A.; Kapranov, P.; Rozowsky, J.; Zheng, D.; Castelo, R.; Frankish, A.; Harrow, J.; Ghosh, S.; Sandelin, A.; Hofacker, I. L.; Baertsch, R.; Keefe, D.; Dike, S.; Cheng, J.; Hirsch, H. A.; Sekinger, E. A.; Lagarde, J.; Abril, J. F.; Shahab, A.; Flamm, C.; Fried, C.; Hackermuller, J.; Hertel, J.; Lindemeyer, M.; Missal, K.; Tanzer, A.; Washietl, S.; Korbel, J.; Emanuelsson, O.; Pedersen, J. S.; Holroyd, N.; Taylor, R.; Swarbreck, D.; Matthews, N.; Dickson, M. C.; Thomas, D. J.; Weirauch, M. T.; Gilbert, J.; Drenkow, J.; Bell, I.; Zhao, X.; Srinivasan, K. G.; Sung, W. K.; Ooi, H. S.; Chiu, K. P.; Foissac, S.; Alioto, T.; Brent, M.; Pachter, L.; Tress, M. L.; Valencia, A.; Choo, S. W.; Choo, C. Y.; Ucla, C.; Manzano, C.; Wyss, C.; Cheung, E.; Clark, T. G.; Brown, J. B.; Ganesh, M.; Patel, S.; Tammana, H.; Chrast, J.; Henrichsen, C. N.; Kai, C.; Kawai, J.; Nagalakshmi, U.; Wu, J.; Lian, Z.; Lian, J.; Newburger, P.; Zhang, X.; Bickel, P.; Mattick, J. S.; Carninci, P.; Hayashizaki, Y.; Weissman, S.; Hubbard, T.; Myers, R. M.; Rogers, J.; Stadler, P. F.; Lowe, T. M.; Wei, C. L.; Ruan, Y.; Struhl, K.; Gerstein, M.; Antonarakis, S. E.; Fu, Y.; Green, E. D.; Karaoz, U.; Siepel, A.; Taylor, J.; Liefer, L. A.; Wetterstrand, K. A.; Good, P. J.; Feingold, E. A.; Guyer, M. S.; Cooper, G. M.; Asimenos, G.; Dewey, C. N.; Hou, M.; Nikolaev, S.; Montoya-Burgos, J. I.; Loytynoja, A.; Whelan, S.; Pardi, F.; Massingham, T.; Huang, H.; Zhang, N. R.; Holmes, I.; Mullikin, J. C.; Ureta-Vidal, A.; Paten, B.; Seringhaus, M.; Church, D.; Rosenbloom, K.; Kent, W. J.; Stone, E. A.; Batzoglou, S.; Goldman, N.; Hardison, R. C.; Haussler, D.; Miller, W.; Sidow, A.; Trinklein, N. D.; Zhang, Z. D.; Barrera, L.; Stuart, R.; King, D. C.; Ameur, A.; Enroth, S.; Bieda, M. C.; Kim, J.; Bhinge, A. A.; Jiang, N.; Liu, J.; Yao, F.; Vega, V. B.; Lee, C. W.; Ng, P.; Yang, A.; Moqtaderi, Z.; Zhu, Z.; Xu, X.; Squazzo, S.; Oberley, M. J.; Inman, D.; Singer, M. A.; Richmond, T. A.; Munn, K. J.; Rada-Iglesias, A.; Wallerman, O.; Komorowski, J.; Fowler, J. C.; Couttet, P.; Bruce, A. W.; Dovey, O. M.; Ellis, P. D.; Langford, C. F.; Nix, D. A.; Euskirchen, G.; Hartman, S.; Urban, A. E.; Kraus, P.; Van Calcar, S.; Heintzman, N.; Kim, T. H.; Wang, K.; Qu, C.; Hon, G.; Luna, R.; Glass, C. K.; Rosenfeld, M. G.; Aldred, S. F.; Cooper, S. J.; Halees, A.; Lin, J. M.; Shulha, H. P.; Xu, M.; Haidar, J. N.; Yu, Y.; Iyer, V. R.; Green, R. D.; Wadelius, C.; Farnham, P. J.; Ren, B.; Harte, R. A.; Hinrichs, A. S.; Trumbower, H.; Clawson, H.; Hillman-Jackson, J.; Zweig, A. S.; Smith, K.; Thakkapallayil, A.; Barber, G.; Kuhn, R. M.; Karolchik, D.; Armengol, L.; Bird, C. P.; de Bakker, P. I.; Kern, A. D.; Lopez-Bigas, N.; Martin, J. D.; Stranger, B. E.; Woodroffe, A.; Davydov, E.; Dimas, A.; Eyras, E.; Hallgrimsdottir, I. B.; Huppert, J.; Zody, M. C.; Abecasis, G. R.; Estivill, X.; Bouffard, G. G.; Guan, X.; Hansen, N. F.; Idol, J. R.; Maduro, V. V.; Maskeri, B.; McDowell, J. C.; Park, M.; Thomas, P. J.; Young, A. C.; Blakesley, R. W.; Muzny, D. M.; Sodergren, E.; Wheeler, D. A.; Worley, K. C.; Jiang, H.; Weinstock, G. M.; Gibbs, R. A.; Graves, T.; Fulton, R.; Mardis, E. R.; Wilson, R. K.;

Clamp, M.; Cuff, J.; Gnerre, S.; Jaffe, D. B.; Chang, J. L.; Lindblad-Toh, K.; Lander, E. S.; Koriabine, M.; Nefedov, M.; Osoegawa, K.; Yoshinaga, Y.; Zhu, B. ; de Jong, P. J. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature 447(7146): 799-816.

Boeri, M.; Verri, C.; Conte, D.; Roz, L.; Modena, P.; Facchinetti, F.; Calabro, E.; Croce, C. M.; Pastorino, U. ; Sozzi, G. (2011). "MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer." Proc Natl Acad Sci U S A 108(9): 3713-3718.

Bogerd, H. P.; Karnowski, H. W.; Cai, X.; Shin, J.; Pohlers, M. ; Cullen, B. R. (2010). "A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs." Mol Cell 37(1): 135-142.

Bond, C. S. ; Fox, A. H. (2009). "Paraspeckles: nuclear bodies built on long noncoding RNA." J Cell Biol 186(5): 637-644.

Braconi, C.; Kogure, T.; Valeri, N.; Huang, N.; Nuovo, G.; Costinean, S.; Negrini, M.; Miotto, E.; Croce, C. M. ; Patel, T. (2011). "microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer." Oncogene 30(47): 4750-4756.

Brannan, C. I.; Dees, E. C.; Ingram, R. S. ; Tilghman, S. M. (1990). "The product of the H19 gene may function as an RNA." Mol Cell Biol 10(1): 28-36.

Brown, C. J.; Ballabio, A.; Rupert, J. L.; Lafreniere, R. G.; Grompe, M.; Tonlorenzi, R. ; Willard, H. F. (1991). "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." Nature 349(6304): 38-44.

Cabili, M. N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A. ; Rinn, J. L. (2011). "Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses." Genes Dev 25(18): 1915-1927.

Cantara, W. A.; Crain, P. F.; Rozenski, J.; McCloskey, J. A.; Harris, K. A.; Zhang, X.; Vendeix, F. A.; Fabris, D. ; Agris, P. F. (2011). "The RNA Modification Database, RNAMDB: 2011 update." Nucleic Acids Res 39(Database issue): D195-201.

Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M. C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C.; Kodzius, R.; Shimokawa, K.; Bajic, V. B.; Brenner, S. E.; Batalov, S.; Forrest, A. R.; Zavolan, M.; Davis, M. J.; Wilming, L. G.; Aidinis, V.; Allen, J. E.; Ambesi-Impiombato, A.; Apweiler, R.; Aturaliya, R. N.; Bailey, T. L.; Bansal, M.; Baxter, L.; Beisel, K. W.; Bersano, T.; Bono, H.; Chalk, A. M.; Chiu, K. P.; Choudhary, V.; Christoffels, A.; Clutterbuck, D. R.; Crowe, M. L.; Dalla, E.; Dalrymple, B. P.; de Bono, B.; Della Gatta, G.; di Bernardo, D.; Down, T.; Engstrom, P.; Fagiolini, M.; Faulkner, G.; Fletcher, C. F.; Fukushima, T.; Furuno, M.; Futaki, S.; Gariboldi, M.; Georgii-Hemming, P.; Gingeras, T. R.; Gojobori, T.; Green, R. E.; Gustincich, S.; Harbers, M.; Hayashi, Y.; Hensch, T. K.; Hirokawa, N.; Hill, D.; Huminiecki, L.; Iacono, M.; Ikeo, K.; Iwama, A.; Ishikawa, T.; Jakt, M.; Kanapin, A.; Katoh, M.; Kawasawa, Y.; Kelso, J.; Kitamura, H.; Kitano, H.; Kollias, G.; Krishnan, S. P.; Kruger, A.; Kummerfeld, S. K.; Kurochkin, I. V.; Lareau, L. F.; Lazarevic, D.; Lipovich, L.; Liu, J.; Liuni, S.; McWilliam, S.; Madan Babu, M.; Madera, M.; Marchionni, L.; Matsuda, H.; Matsuzawa, S.; Miki, H.; Mignone, F.;

Miyake, S.; Morris, K.; Mottagui-Tabar, S.; Mulder, N.; Nakano, N.; Nakauchi, H.; Ng, P.; Nilsson, R.; Nishiguchi, S.; Nishikawa, S.; Nori, F.; Ohara, O.; Okazaki, Y.; Orlando, V.; Pang, K. C.; Pavan, W. J.; Pavesi, G.; Pesole, G.; Petrovsky, N.; Piazza, S.; Reed, J.; Reid, J. F.; Ring, B. Z.; Ringwald, M.; Rost, B.; Ruan, Y.; Salzberg, S. L.; Sandelin, A.; Schneider, C.; Schonbach, C.; Sekiguchi, K.; Semple, C. A.; Seno, S.; Sessa, L.; Sheng, Y.; Shibata, Y.; Shimada, H.; Shimada, K.; Silva, D.; Sinclair, B.; Sperling, S.; Stupka, E.; Sugiura, K.; Sultana, R.; Takenaka, Y.; Taki, K.; Tammoja, K.; Tan, S. L.; Tang, S.; Taylor, M. S.; Tegner, J.; Teichmann, S. A.; Ueda, H. R.; van Nimwegen, E.; Verardo, R.; Wei, C. L.; Yagi, K.; Yamanishi, H.; Zabarovsky, E.; Zhu, S.; Zimmer, A.; Hide, W.; Bult, C.; Grimmond, S. M.; Teasdale, R. D.; Liu, E. T.; Brusic, V.; Quackenbush, J.; Wahlestedt, C.; Mattick, J. S.; Hume, D. A.; Kai, C.; Sasaki, D.; Tomaru, Y.; Fukuda, S.; Kanamori-Katayama, M.; Suzuki, M.; Aoki, J.; Arakawa, T.; Iida, J.; Imamura, K.; Itoh, M.; Kato, T.; Kawaji, H.; Kawagashira, N.; Kawashima, T.; Kojima, M.; Kondo, S.; Konno, H.; Nakano, K.; Ninomiya, N.; Nishio, T.; Okada, M.; Plessy, C.; Shibata, K.; Shiraki, T.; Suzuki, S.; Tagami, M.; Waki, K.; Watahiki, A.; Okamura-Oho, Y.; Suzuki, H.; Kawai, J. ; Hayashizaki, Y. (2005). "The transcriptional landscape of the mammalian genome." Science 309(5740): 1559-1563.

Carthew, R. W. ; Sontheimer, E. J. (2009). "Origins and Mechanisms of miRNAs and siRNAs." Cell 136(4): 642-655.

Cawley, S.; Bekiranov, S.; Ng, H. H.; Kapranov, P.; Sekinger, E. A.; Kampa, D.; Piccolboni, A.; Sementchenko, V.; Cheng, J.; Williams, A. J.; Wheeler, R.; Wong, B.; Drenkow, J.; Yamanaka, M.; Patel, S.; Brubaker, S.; Tammana, H.; Helt, G.; Struhl, K. ; Gingeras, T. R. (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.

Cazalla, D.; Yario, T. ; Steitz, J. A. (2010). "Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA." Science 328(5985): 1563-1566.

Cesana, M.; Cacchiarelli, D.; Legnini, I.; Santini, T.; Sthandier, O.; Chinappi, M.; Tramontano, A. ; Bozzoni, I. (2011). "A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA." Cell 147(2): 358-369.

Cheloufi, S.; Dos Santos, C. O.; Chong, M. M. ; Hannon, G. J. (2010). "A dicer-independent miRNA biogenesis pathway that requires Ago catalysis." Nature 465(7298): 584-589.

Chen, Y.; Song, Y. X. ; Wang, Z. N. (2013). "The microRNA-148/152 family: multi-faceted players." Mol Cancer 12: 43.

Cheng, J.; Kapranov, P.; Drenkow, J.; Dike, S.; Brubaker, S.; Patel, S.; Long, J.; Stern, D.; Tammana, H.; Helt, G.; Sementchenko, V.; Piccolboni, A.; Bekiranov, S.; Bailey, D. K.; Ganesh, M.; Ghosh, S.; Bell, I.; Gerhard, D. S. ; Gingeras, T. R. (2005). "Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution." Science 308(5725): 1149-1154.

Chi, S. W.; Hannon, G. J. ; Darnell, R. B. (2012). "An alternative mode of microRNA target recognition." Nat Struct Mol Biol 19(3): 321-327.

Chi, S. W.; Zang, J. B.; Mele, A. ; Darnell, R. B. (2009). "Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps." Nature 460(7254): 479-486.

Chorev, M. ; Carmel, L. (2012). "The function of introns." Front Genet 3: 55.

Cifuentes, D.; Xue, H.; Taylor, D. W.; Patnode, H.; Mishima, Y.; Cheloufi, S.; Ma, E.; Mane, S.; Hannon, G. J.; Lawson, N. D.; Wolfe, S. A. ; Giraldez, A. J. (2010). "A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity." Science 328(5986): 1694-1698.

Clark, M. B. ; Mattick, J. S. (2011). "Long noncoding RNAs in cell biology." Semin Cell Dev Biol 22(4): 366-376.

Clemson, C. M.; Hutchinson, J. N.; Sara, S. A.; Ensminger, A. W.; Fox, A. H.; Chess, A. ; Lawrence, J. B. (2009). "An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles." Mol Cell 33(6): 717-726.

Collins, F. S.; Lander, E. S.; Rogers, J.; Waterston, R. H. ; Conso, I. H. G. S. (2004). "Finishing the euchromatic sequence of the human genome." Nature 431(7011): 931-945.

Conley, A. B. ; Jordan, I. K. (2012). "Epigenetic regulation of human cis-natural antisense transcripts." Nucleic Acids Res 40(4): 1438-1445.

Consortium, E. P.; Dunham, I.; Kundaje, A.; Aldred, S. F.; Collins, P. J.; Davis, C. A.; Doyle, F.; Epstein, C. B.; Frietze, S.; Harrow, J.; Kaul, R.; Khatun, J.; Lajoie, B. R.; Landt, S. G.; Lee, B. K.; Pauli, F.; Rosenbloom, K. R.; Sabo, P.; Safi, A.; Sanyal, A.; Shoresh, N.; Simon, J. M.; Song, L.; Trinklein, N. D.; Altshuler, R. C.; Birney, E.; Brown, J. B.; Cheng, C.; Djebali, S.; Dong, X.; Dunham, I.; Ernst, J.; Furey, T. S.; Gerstein, M.; Giardine, B.; Greven, M.; Hardison, R. C.; Harris, R. S.; Herrero, J.; Hoffman, M. M.; Iyer, S.; Kelllis, M.; Khatun, J.; Kheradpour, P.; Kundaje, A.; Lassman, T.; Li, Q.; Lin, X.; Marinov, G. K.; Merkel, A.; Mortazavi, A.; Parker, S. C.; Reddy, T. E.; Rozowsky, J.; Schlesinger, F.; Thurman, R. E.; Wang, J.; Ward, L. D.; Whitfield, T. W.; Wilder, S. P.; Wu, W.; Xi, H. S.; Yip, K. Y.; Zhuang, J.; Bernstein, B. E.; Birney, E.; Dunham, I.; Green, E. D.; Gunter, C.; Snyder, M.; Pazin, M. J.; Lowdon, R. F.; Dillon, L. A.; Adams, L. B.; Kelly, C. J.; Zhang, J.; Wexler, J. R.; Green, E. D.; Good, P. J.; Feingold, E. A.; Bernstein, B. E.; Birney, E.; Crawford, G. E.; Dekker, J.; Elinitski, L.; Farnham, P. J.; Gerstein, M.; Giddings, M. C.; Gingeras, T. R.; Green, E. D.; Guigo, R.; Hardison, R. C.; Hubbard, T. J.; Kellis, M.; Kent, W. J.; Lieb, J. D.; Margulies, E. H.; Myers, R. M.; Snyder, M.; Starnatoyannopoulos, J. A.; Tennebaum, S. A.; Weng, Z.; White, K. P.; Wold, B.; Khatun, J.; Yu, Y.; Wrobel, J.; Risk, B. A.; Gunawardena, H. P.; Kuiper, H. C.; Maier, C. W.; Xie, L.; Chen, X.; Giddings, M. C.; Bernstein, B. E.; Epstein, C. B.; Shoresh, N.; Ernst, J.; Kheradpour, P.; Mikkelsen, T. S.; Gillespie, S.; Goren, A.; Ram, O.; Zhang, X.; Wang, L.; Issner, R.; Coyne, M. J.; Durham, T.; Ku, M.; Truong, T.; Ward, L. D.; Altshuler, R. C.; Eaton, M. L.; Kellis, M.; Djebali, S.; Davis, C. A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; Xue, C.; Marinov, G. K.; Khatun, J.; Williams, B. A.; Zaleski, C.; Rozowsky, J.; Roder, M.; Kokocinski, F.; Abdelhamid, R. F.; Alioto, T.; Antoshechkin, I.; Baer, M. T.; Batut, P.; Bell, I.; Bell, K.; Chakrabortty, S.; Chen, X.; Chrast, J.; Curado, J.; Derrien, T.; Drenkow, J.; Dumais, E.; Dumais, J.; Duttagupta, R.;

Fastuca, M.; Fejes-Toth, K.; Ferreira, P.; Foissac, S.; Fullwood, M. J.; Gao, H.; Gonzalez, D.; Gordon, A.; Gunawardena, H. P.; Howald, C.; Jha, S.; Johnson, R.; Kapranov, P.; King, B.; Kingswood, C.; Li, G.; Luo, O. J.; Park, E.; Preall, J. B.; Presaud, K.; Ribeca, P.; Risk, B. A.; Robyr, D.; Ruan, X.; Sammeth, M.; Sandu, K. S.; Schaeffer, L.; See, L. H.; Shahab, A.; Skancke, J.; Suzuki, A. M.; Takahashi, H.; Tilgner, H.; Trout, D.; Walters, N.; Wang, H.; Wrobel, J.; Yu, Y.; Hayashizaki, Y.; Harrow, J.; Gerstein, M.; Hubbard, T. J.; Reymond, A.; Antonarakis, S. E.; Hannon, G. J.; Giddings, M. C.; Ruan, Y.; Wold, B.; Carninci, P.; Guigo, R.; Gingeras, T. R.; Rosenbloom, K. R.; Sloan, C. A.; Learned, K.; Malladi, V. S.; Wong, M. C.; Barber, G. P.; Cline, M. S.; Dreszer, T. R.; Heitner, S. G.; Karolchik, D.; Kent, W. J.; Kirkup, V. M.; Meyer, L. R.; Long, J. C.; Maddren, M.; Raney, B. J.; Furey, T. S.; Song, L.; Grasfeder, L. L.; Giresi, P. G.; Lee, B. K.; Battenhouse, A.; Sheffield, N. C.; Simon, J. M.; Showers, K. A.; Safi, A.; London, D.; Bhinge, A. A.; Shestak, C.; Schaner, M. R.; Kim, S. K.; Zhang, Z. Z.; Mieczkowski, P. A.; Mieczkowska, J. O.; Liu, Z.; McDaniell, R. M.; Ni, Y.; Rashid, N. U.; Kim, M. J.; Adar, S.; Zhang, Z.; Wang, T.; Winter, D.; Keefe, D.; Birney, E.; Iyer, V. R.; Lieb, J. D.; Crawford, G. E.; Li, G.; Sandhu, K. S.; Zheng, M.; Wang, P.; Luo, O. J.; Shahab, A.; Fullwood, M. J.; Ruan, X.; Ruan, Y.; Myers, R. M.; Pauli, F.; Williams, B. A.; Gertz, J.; Marinov, G. K.; Reddy, T. E.; Vielmetter, J.; Partridge, E. C.; Trout, D.; Varley, K. E.; Gasper, C.; Bansal, A.; Pepke, S.; Jain, P.; Amrhein, H.; Bowling, K. M.; Anaya, M.; Cross, M. K.; King, B.; Muratet, M. A.; Antoshechkin, I.; Newberry, K. M.; McCue, K.; Nesmith, A. S.; Fisher-Aylor, K. I.; Pusey, B.; DeSalvo, G.; Parker, S. L.; Balasubramanian, S.; Davis, N. S.; Meadows, S. K.; Eggleston, T.; Gunter, C.; Newberry, J. S.; Levy, S. E.; Absher, D. M.; Mortazavi, A.; Wong, W. H.; Wold, B.; Blow, M. J.; Visel, A.; Pennachio, L. A.; Elnitski, L.; Margulies, E. H.; Parker, S. C.; Petrykowska, H. M.; Abyzov, A.; Aken, B.; Barrell, D.; Barson, G.; Berry, A.; Bignell, A.; Boychenko, V.; Bussotti, G.; Chrast, J.; Davidson, C.; Derrien, T.; Despacio-Reyes, G.; Diekhans, M.; Ezkurdia, I.; Frankish, A.; Gilbert, J.; Gonzalez, J. M.; Griffiths, E.; Harte, R.; Hendrix, D. A.; Howald, C.; Hunt, T.; Jungreis, I.; Kay, M.; Khurana, E.; Kokocinski, F.; Leng, J.; Lin, M. F.; Loveland, J.; Lu, Z.; Manthravadi, D.; Mariotti, M.; Mudge, J.; Mukherjee, G.; Notredame, C.; Pei, B.; Rodriguez, J. M.; Saunders, G.; Sboner, A.; Searle, S.; Sisu, C.; Snow, C.; Steward, C.; Tanzer, A.; Tapanan, E.; Tress, M. L.; van Baren, M. J.; Walters, N.; Washieti, S.; Wilming, L.; Zadissa, A.; Zhengdong, Z.; Brent, M.; Haussler, D.; Kellis, M.; Valencia, A.; Gerstein, M.; Raymond, A.; Guigo, R.; Harrow, J.; Hubbard, T. J.; Landt, S. G.; Frietze, S.; Abyzov, A.; Addleman, N.; Alexander, R. P.; Auerbach, R. K.; Balasubramanian, S.; Bettinger, K.; Bhardwaj, N.; Boyle, A. P.; Cao, A. R.; Cayting, P.; Charos, A.; Cheng, Y.; Cheng, C.; Eastman, C.; Euskirchen, G.; Fleming, J. D.; Grubert, F.; Habegger, L.; Hariharan, M.; Harmanci, A.; Iyenger, S.; Jin, V. X.; Karczewski, K. J.; Kasowski, M.; Lacroute, P.; Lam, H.; Larnarre-Vincent, N.; Leng, J.; Lian, J.; Lindahl-Allen, M.; Min, R.; Miotto, B.; Monahan, H.; Moqtaderi, Z.; Mu, X. J.; O'Geen, H.; Ouyang, Z.; Patacsil, D.; Pei, B.; Raha, D.; Ramirez, L.; Reed, B.; Rozowsky, J.; Sboner, A.; Shi, M.; Sisu, C.; Slifer, T.; Witt, H.; Wu, L.; Xu, X.; Yan, K. K.; Yang, X.; Yip, K. Y.; Zhang, Z.; Struhl, K.; Weissman, S. M.; Gerstein,

M.; Farnham, P. J.; Snyder, M.; Tenebaum, S. A.; Penalva, L. O.; Doyle, F.; Karmakar, S.; Landt, S. G.; Bhanvadia, R. R.; Choudhury, A.; Domanus, M.; Ma, L.; Moran, J.; Patacsil, D.; Slifer, T.; Victorsen, A.; Yang, X.; Snyder, M.; White, K. P.; Auer, T.; Centarin, L.; Eichenlaub, M.; Gruhl, F.; Heerman, S.; Hoeckendorf, B.; Inoue, D.; Kellner, T.; Kirchmaier, S.; Mueller, C.; Reinhardt, R.; Schertel, L.; Schneider, S.; Sinn, R.; Wittbrodt, B.; Wittbrodt, J.; Weng, Z.; Whitfield, T. W.; Wang, J.; Collins, P. J.; Aldred, S. F.; Trinklein, N. D.; Partridge, E. C.; Myers, R. M.; Dekker, J.; Jain, G.; Lajoie, B. R.; Sanyal, A.; Balasundaram, G.; Bates, D. L.; Byron, R.; Canfield, T. K.; Diegel, M. J.; Dunn, D.; Ebersol, A. K.; Ebersol, A. K.; Frum, T.; Garg, K.; Gist, E.; Hansen, R. S.; Boatman, L.; Haugen, E.; Humbert, R.; Jain, G.; Johnson, A. K.; Johnson, E. M.; Kutyavin, T. M.; Lajoie, B. R.; Lee, K.; Lotakis, D.; Maurano, M. T.; Neph, S. J.; Neri, F. V.; Nguyen, E. D.; Qu, H.; Reynolds, A. P.; Roach, V.; Rynes, E.; Sabo, P.; Sanchez, M. E.; Sandstrom, R. S.; Sanyal, A.; Shafer, A. O.; Stergachis, A. B.; Thomas, S.; Thurman, R. E.; Vernot, B.; Vierstra, J.; Vong, S.; Wang, H.; Weaver, M. A.; Yan, Y.; Zhang, M.; Akey, J. A.; Bender, M.; Dorschner, M. O.; Groudine, M.; MacCoss, M. J.; Navas, P.; Stamatoyannopoulos, G.; Kaul, R.; Dekker, J.; Stamatoyannopoulos, J. A.; Dunham, I.; Beal, K.; Brazma, A.; Flicek, P.; Herrero, J.; Johnson, N.; Keefe, D.; Lukk, M.; Luscombe, N. M.; Sobral, D.; Vaquerizas, J. M.; Wilder, S. P.; Batzoglou, S.; Sidow, A.; Hussami, N.; Kyriazopoulou-Panagiotopoulou, S.; Libbrecht, M. W.; Schaub, M. A.; Kundaje, A.; Hardison, R. C.; Miller, W.; Giardine, B.; Harris, R. S.; Wu, W.; Bickel, P. J.; Banfai, B.; Boley, N. P.; Brown, J. B.; Huang, H.; Li, Q.; Li, J. J.; Noble, W. S.; Bilmes, J. A.; Buske, O. J.; Hoffman, M. M.; Sahu, A. O.; Kharchenko, P. V.; Park, P. J.; Baker, D.; Taylor, J.; Weng, Z.; Iyer, S.; Dong, X.; Greven, M.; Lin, X.; Wang, J.; Xi, H. S.; Zhuang, J.; Gerstein, M.; Alexander, R. P.; Balasubramanian, S.; Cheng, C.; Harmanci, A.; Lochovsky, L.; Min, R.; Mu, X. J.; Rozowsky, J.; Yan, K. K.; Yip, K. Y. ; Birney, E. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature 489(7414): 57-74.

Consortium, I. H. G. S. (1998). "Genome sequence of the nematode C. elegans: a platform for investigating biology." Science 282(5396): 2012-2018.

Cortez, M. A.; Bueso-Ramos, C.; Ferdin, J.; Lopez-Berestein, G.; Sood, A. K. ; Calin, G. A. (2011). "MicroRNAs in body fluids--the mix of hormones and biomarkers." Nat Rev Clin Oncol 8(8): 467-477.

Crick, F. (1970). "Central dogma of molecular biology." Nature 227(5258): 561-563.

Crick, F. H. (1958). "On protein synthesis." Symp Soc Exp Biol 12: 138-163.

Czech, B. ; Hannon, G. J. (2011). "Small RNA sorting: matchmaking for Argonautes." Nat Rev Genet 12(1): 19-31.

David, L.; Huber, W.; Granovskaia, M.; Toedling, J.; Palm, C. J.; Bofkin, L.; Jones, T.; Davis, R. W. ; Steinmetz, L. M. (2006). "A high-resolution map of transcription in the yeast genome." Proc Natl Acad Sci U S A 103(14): 5320-5325.

DeMarco, R.; Oliveira, K. C. P.; Almeida, G. T.; Amaral, M. S.; Beckedorff, F. C. F. ; Verjovski-Almeida, S. (2013). "Comparative transcriptomic profiles of Schistosoma mansoni adult male, adult female and egg stages " Submitted.

Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D. G.; Lagarde, J.; Veeravalli, L.; Ruan, X.; Ruan, Y.; Lassmann, T.; Carninci, P.; Brown, J. B.; Lipovich, L.; Gonzalez, J. M.; Thomas, M.; Davis, C. A.; Shiekhattar, R.; Gingeras, T. R.; Hubbard, T. J.; Notredame, C.; Harrow, J. ; Guigo, R. (2012). "The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression." Genome Res 22(9): 1775-1789.

Di Ruscio, A.; Ebralidze, A. K.; Benoukraf, T.; Amabile, G.; Goff, L. A.; Terragni, J.; Figueroa, M. E.; De Figueiredo Pontes, L. L.; Alberich-Jorda, M.; Zhang, P.; Wu, M.; D'Alo, F.; Melnick, A.; Leone, G.; Ebralidze, K. K.; Pradhan, S.; Rinn, J. L. ; Tenen, D. G. (2013). "DNMT1-interacting RNAs block gene-specific DNA methylation." Nature.

Di Stefano, V.; Zaccagnini, G.; Capogrossi, M. C. ; Martelli, F. (2011). "microRNAs as peripheral blood biomarkers of cardiovascular disease." Vascul Pharmacol 55(4): 111-118.

Dieci, G.; Fiorino, G.; Castelnuovo, M.; Teichmann, M. ; Pagano, A. (2007). "The expanding RNA polymerase III transcriptome." Trends Genet 23(12): 614-622.

Dinger, M. E.; Amaral, P. P.; Mercer, T. R. ; Mattick, J. S. (2009). "Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications." Brief Funct Genomic Proteomic 8(6): 407-423.

Djebali, S.; Davis, C. A. ; et_al. (2012). "Landscape of transcription in human cells." Nature 489(7414): 101-108.

Dohm, J. C.; Lottaz, C.; Borodina, T. ; Himmelbauer, H. (2008). "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing." Nucleic Acids Res 36(16): e105.

Du, Z.; Fei, T.; Verhaak, R. G.; Su, Z.; Zhang, Y.; Brown, M.; Chen, Y. ; Liu, X. S. (2013). "Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer." Nat Struct Mol Biol 20(7): 908-913.

Dunham, J. P. ; Friesen, M. L. (2013). "A cost-effective method for high-throughput construction of illumina sequencing libraries." Cold Spring Harb Protoc 2013(9).

Ebert, M. S. ; Sharp, P. A. (2010). "Emerging roles for natural microRNA sponges." Curr Biol 20(19): R858-861.

Ebert, M. S. ; Sharp, P. A. (2012). "Roles for microRNAs in conferring robustness to biological processes." Cell 149(3): 515-524.

Esteller, M. (2011). "Non-coding RNAs in human disease." Nat Rev Genet 12(12): 861-874.

Fabian, M. R. ; Sonenberg, N. (2012). "The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC." Nat Struct Mol Biol 19(6): 586-593.

Faghihi, M. A.; Modarresi, F.; Khalil, A. M.; Wood, D. E.; Sahagan, B. G.; Morgan, T. E.; Finch, C. E.; St Laurent, G., 3rd; Kenny, P. J. ; Wahlestedt, C. (2008). "Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase." Nat Med 14(7): 723-730.

Farajollahi, S. ; Maas, S. (2010). "Molecular diversity through RNA editing: a balancing act." Trends Genet 26(5): 221-230.

Faulhammer, D.; Lipton, R. J. ; Landweber, L. F. (2000). "Fidelity of enzymatic ligation for DNA computing." J Comput Biol 7(6): 839-848.

Fejes-Toth, K.; Sotirova, V.; Sachidanandam, R.; Assaf, G. ; Hannon G, J. (2009). "Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs." Nature 457(7232): 1028-1032.

Franco-Zorrilla, J. M.; Valli, A.; Todesco, M.; Mateos, I.; Puga, M. I.; Rubio-Somoza, I.; Leyva, A.; Weigel, D.; Garcia, J. A. ; Paz-Ares, J. (2007). "Target mimicry provides a new mechanism for regulation of microRNA activity." Nat Genet 39(8): 1033-1037.

Friedlander, M. R.; Mackowiak, S. D.; Li, N.; Chen, W. ; Rajewsky, N. (2012). "miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades." Nucleic Acids Res 40(1): 37-52.

Friedman, R. C.; Farh, K. K.; Burge, C. B. ; Bartel, D. P. (2009). "Most mammalian mRNAs are conserved targets of microRNAs." Genome Res 19(1): 92-105.

Geisler, S.; Lojek, L.; Khalil, A. M.; Baker, K. E. ; Coller, J. (2012). "Decapping of long noncoding RNAs regulates inducible genes." Mol Cell 45(3): 279-291.

Gerstein, M. B.; Lu, Z. J.; Van Nostrand, E. L.; Cheng, C.; Arshinoff, B. I.; Liu, T.; Yip, K. Y.; Robilotto, R.; Rechtsteiner, A.; Ikegami, K.; Alves, P.; Chateigner, A.; Perry, M.; Morris, M.; Auerbach, R. K.; Feng, X.; Leng, J.; Vielle, A.; Niu, W.; Rhrissorrakrai, K.; Agarwal, A.; Alexander, R. P.; Barber, G.; Brdlik, C. M.; Brennan, J.; Brouillet, J. J.; Carr, A.; Cheung, M. S.; Clawson, H.; Contrino, S.; Dannenberg, L. O.; Dernburg, A. F.; Desai, A.; Dick, L.; Dose, A. C.; Du, J.; Egelhofer, T.; Ercan, S.; Euskirchen, G.; Ewing, B.; Feingold, E. A.; Gassmann, R.; Good, P. J.; Green, P.; Gullier, F.; Gutwein, M.; Guyer, M. S.; Habegger, L.; Han, T.; Henikoff, J. G.; Henz, S. R.; Hinrichs, A.; Holster, H.; Hyman, T.; Iniguez, A. L.; Janette, J.; Jensen, M.; Kato, M.; Kent, W. J.; Kephart, E.; Khivansara, V.; Khurana, E.; Kim, J. K.; Kolasinska-Zwierz, P.; Lai, E. C.; Latorre, I.; Leahey, A.; Lewis, S.; Lloyd, P.; Lochovsky, L.; Lowdon, R. F.; Lubling, Y.; Lyne, R.; MacCoss, M.; Mackowiak, S. D.; Mangone, M.; McKay, S.; Mecenas, D.; Merrihew, G.; Miller, D. M., 3rd; Muroyama, A.; Murray, J. I.; Ooi, S. L.; Pham, H.; Phippen, T.; Preston, E. A.; Rajewsky, N.; Ratsch, G.; Rosenbaum, H.; Rozowsky, J.; Rutherford, K.; Ruzanov, P.; Sarov, M.; Sasidharan, R.; Sboner, A.; Scheid, P.; Segal, E.; Shin, H.; Shou, C.; Slack, F. J.; Slightam, C.; Smith, R.; Spencer, W. C.; Stinson, E. O.; Taing, S.; Takasaki, T.; Vafeados, D.; Voronina, K.; Wang, G.; Washington, N. L.; Whittle, C. M.; Wu, B.; Yan, K. K.; Zeller, G.; Zha, Z.; Zhong, M.; Zhou, X.; Ahringer, J.; Strome, S.; Gunsalus, K. C.; Micklem, G.; Liu, X. S.; Reinke, V.; Kim, S. K.; Hillier, L. W.; Henikoff, S.; Piano, F.; Snyder, M.; Stein, L.; Lieb, J. D. ; Waterston, R. H. (2010). "Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project." Science 330(6012): 1775-1787.

Gevers, D.; Knight, R.; Petrosino, J. F.; Huang, K.; McGuire, A. L.; Birren, B. W.; Nelson, K. E.; White, O.; Methe, B. A. ; Huttenhower, C. (2012). "The Human Microbiome Project: a community resource for the healthy human microbiome." PLoS Biol 10(8): e1001377.

Gibbons, R. A. ; Hunter, G. D. (1967). "Nature of the scrapie agent." Nature 215(5105): 1041-1043.

Girard, A.; Sachidanandam, R.; Hannon, G. J. ; Carmell, M. A. (2006). "A germline-specific class of small RNAs binds mammalian Piwi proteins." Nature 442(7099): 199-202.

Gomez, J. A.; Wapinski, O. L.; Yang, Y. W.; Bureau, J. F.; Gopinath, S.; Monack, D. M.; Chang, H. Y.; Brahic, M. ; Kirkegaard, K. (2013). "The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus." Cell 152(4): 743-754.

Grey, F.; Tirabassi, R.; Meyers, H.; Wu, G.; McWeeney, S.; Hook, L. ; Nelson, J. A. (2010). "A viral microRNA down-regulates multiple cell cycle genes through mRNA 5'UTRs." PLoS Pathog 6(6): e1000967.

Gubler, U. (1987). "Second-strand cDNA synthesis: mRNA fragments as primers." Methods Enzymol 152: 330-335.

Gupta, R. A.; Shah, N.; Wang, K. C.; Kim, J.; Horlings, H. M.; Wong, D. J.; Tsai, M. C.; Hung, T.; Argani, P.; Rinn, J. L.; Wang, Y.; Brzoska, P.; Kong, B.; Li, R.; West, R. B.; van de Vijver, M. J.; Sukumar, S. ; Chang, H. Y. (2010). "Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis." Nature 464(7291): 1071-1076.

Gutschner, T.; Hammerle, M. ; Diederichs, S. (2013). "MALAT1 -- a paradigm for long noncoding RNA function in cancer." J Mol Med (Berl) 91(7): 791-801.

Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M. F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B. W.; Cassady, J. P.; Cabili, M. N.; Jaenisch, R.; Mikkelsen, T. S.; Jacks, T.; Hacohen, N.; Bernstein, B. E.; Kellis, M.; Regev, A.; Rinn, J. L. ; Lander, E. S. (2009). "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals." Nature 458(7235): 223-227.

Guttman, M.; Donaghey, J.; Carey, B. W.; Garber, M.; Grenier, J. K.; Munson, G.; Young, G.; Lucas, A. B.; Ach, R.; Bruhn, L.; Yang, X.; Amit, I.; Meissner, A.; Regev, A.; Rinn, J. L.; Root, D. E. ; Lander, E. S. (2011). "lincRNAs act in the circuitry controlling pluripotency and differentiation." Nature 477(7364): 295-300.

Guttman, M.; Garber, M.; Levin, J. Z.; Donaghey, J.; Robinson, J.; Adiconis, X.; Fan, L.; Koziol, M. J.; Gnirke, A.; Nusbaum, C.; Rinn, J. L.; Lander, E. S. ; Regev, A. (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.

Hafner, M.; Landthaler, M.; Burger, L.; Khorshid, M.; Hausser, J.; Berninger, P.; Rothballer, A.; Ascano, M., Jr.; Jungkamp, A. C.; Munschauer, M.; Ulrich, A.; Wardle, G. S.; Dewell, S.; Zavolan, M. ; Tuschl, T. (2010). "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP." Cell 141(1): 129-141.

Hafner, M.; Lianoglou, S.; Tuschl, T. ; Betel, D. (2012). "Genome-wide identification of miRNA targets by PAR-CLIP." Methods 58(2): 94-105.

Hah, N.; Danko, C. G.; Core, L.; Waterfall, J. J.; Siepel, A.; Lis, J. T. ; Kraus, W. L. (2011). "A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells." Cell 145(4): 622-634.

Han, J.; Lee, Y.; Yeom, K. H.; Kim, Y. K.; Jin, H. ; Kim, V. N. (2004). "The Drosha-DGCR8 complex in primary microRNA processing." Genes Dev 18(24): 3016-3027.

Hansen, T. B.; Jensen, T. I.; Clausen, B. H.; Bramsen, J. B.; Finsen, B.; Damgaard, C. K. ; Kjems, J. (2013). "Natural RNA circles function as efficient microRNA sponges." Nature 495(7441): 384-388.

Hansen, T. B.; Wiklund, E. D.; Bramsen, J. B.; Villadsen, S. B.; Statham, A. L.; Clark, S. J. ; Kjems, J. (2011). "miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA." EMBO J 30(21): 4414-4422.

Harrow, J.; Frankish, A.; Gonzalez, J. M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B. L.; Barrell, D.; Zadissa, A.; Searle, S.; Barnes, I.; Bignell, A.; Boychenko, V.; Hunt, T.; Kay, M.; Mukherjee, G.; Rajan, J.; Despacio-Reyes, G.; Saunders, G.; Steward, C.; Harte, R.; Lin, M.; Howald, C.; Tanzer, A.; Derrien, T.; Chrast, J.; Walters, N.; Balasubramanian, S.; Pei, B.; Tress, M.; Rodriguez, J. M.; Ezkurdia, I.; van Baren, J.; Brent, M.; Haussler, D.; Kellis, M.; Valencia, A.; Reymond, A.; Gerstein, M.; Guigo, R. ; Hubbard, T. J. (2012). "GENCODE: the reference human genome annotation for The ENCODE Project." Genome Res 22(9): 1760-1774.

Havens, M. A.; Reich, A. A.; Duelli, D. M. ; Hastings, M. L. (2012). "Biogenesis of mammalian microRNAs by a non-canonical processing pathway." Nucleic Acids Res 40(10): 4626-4640.

He, Y.; Vogelstein, B.; Velculescu, V. E.; Papadopoulos, N. ; Kinzler, K. W. (2008). "The antisense transcriptomes of human cells." Science 322(5909): 1855-1857.

Holley, R. W.; Apgar, J.; Everett, G. A.; Madison, J. T.; Marquisee, M.; Merrill, S. H.; Penswick, J. R. ; Zamir, A. (1965). "Structure of a Ribonucleic Acid." Science 147(3664): 1462-1465.

Housby, J. N. ; Southern, E. M. (1998). "Fidelity of DNA ligation: a novel experimental approach based on the polymerisation of libraries of oligonucleotides." Nucleic Acids Res 26(18): 4259-4266.

Hu, G.; Tang, Q.; Sharma, S.; Yu, F.; Escobar, T. M.; Muljo, S. A.; Zhu, J. ; Zhao, K. (2013). "Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation." Nat Immunol 14(11): 1190-1198.

Huarte, M.; Guttman, M.; Feldser, D.; Garber, M.; Koziol, M. J.; Kenzelmann-Broz, D.; Khalil, A. M.; Zuk, O.; Amit, I.; Rabani, M.; Attardi, L. D.; Regev, A.; Lander, E. S.; Jacks, T. ; Rinn, J. L. (2010). "A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response." Cell 142(3): 409-419.

Huntzinger, E. ; Izaurralde, E. (2011). "Gene silencing by microRNAs: contributions of translational repression and mRNA decay." Nat Rev Genet 12(2): 99-110.

Hutchinson, J. N.; Ensminger, A. W.; Clemson, C. M.; Lynch, C. R.; Lawrence, J. B. ; Chess, A. (2007). "A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains." BMC Genomics 8: 39.

Hutvagner, G.; McLachlan, J.; Pasquinelli, A. E.; Balint, E.; Tuschl, T. ; Zamore, P. D. (2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." Science 293(5531): 834-838.

Ihaka, R. a. G., R. (1996). "R: A language for data analysis and graphics." Journal of Computational and Graphical Statistics 5(3): 299-314.

Jalali, S.; Bhartiya, D.; Lalwani, M. K.; Sivasubbu, S. ; Scaria, V. (2013). "Systematic transcriptome wide analysis of lncRNA-miRNA interactions." PLoS One 8(2): e53823.

Janas, M. M.; Wang, B.; Harris, A. S.; Aguiar, M.; Shaffer, J. M.; Subrahmanyam, Y. V.; Behlke, M. A.; Wucherpfennig, K. W.; Gygi, S. P.; Gagnon, E. ; Novina, C. D. (2012). "Alternative RISC assembly: binding and repression of microRNA-mRNA duplexes by human Ago proteins." RNA 18(11): 2041-2055.

Jayaprakash, A. D.; Jabado, O.; Brown, B. D. ; Sachidanandam, R. (2011). "Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing." Nucleic Acids Res 39(21): e141.

Jeggari, A.; Marks, D. S. ; Larsson, E. (2012). "miRcode: a map of putative microRNA target sites in the long non-coding transcriptome." Bioinformatics 28(15): 2062-2063.

Ji, P.; Diederichs, S.; Wang, W.; Boing, S.; Metzger, R.; Schneider, P. M.; Tidow, N.; Brandt, B.; Buerger, H.; Bulk, E.; Thomas, M.; Berdel, W. E.; Serve, H. ; Muller-Tidow, C. (2003). "MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer." Oncogene 22(39): 8031-8041.

Juan, L.; Wang, G.; Radovich, M.; Schneider, B. P.; Clare, S. E.; Wang, Y. ; Liu, Y. (2013). "Potential roles of microRNAs in regulating long intergenic noncoding RNAs." BMC Med Genomics 6 Suppl 1: S7.

Kallen, A. N.; Zhou, X. B.; Xu, J.; Qiao, C.; Ma, J.; Yan, L.; Lu, L.; Liu, C.; Yi, J. S.; Zhang, H.; Min, W.; Bennett, A. M.; Gregory, R. I.; Ding, Y. ; Huang, Y. (2013). "The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs." Mol Cell.

Kapranov, P.; St Laurent, G.; Raz, T.; Ozsolak, F.; Reynolds, C. P.; Sorensen, P. H.; Reaman, G.; Milos, P.; Arceci, R. J.; Thompson, J. F. ; Triche, T. J. (2010). "The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA." BMC Biol 8: 149.

Kasashima, K.; Nakamura, Y. ; Kozu, T. (2004). "Altered expression profiles of microRNAs during TPA-induced differentiation of HL-60 cells." Biochem Biophys Res Commun 322(2): 403-410.

Keniry, A.; Oxley, D.; Monnier, P.; Kyba, M.; Dandolo, L.; Smits, G. ; Reik, W. (2012). "The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r." Nat Cell Biol 14(7): 659-665.

Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." Genome Res 12(4): 656-664.

Ketting, R. F.; Fischer, S. E.; Bernstein, E.; Sijen, T.; Hannon, G. J. ; Plasterk, R. H. (2001). "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans." Genes Dev 15(20): 2654-2659.

Khalil, A. M.; Guttman, M.; Huarte, M.; Garber, M.; Raj, A.; Rivea Morales, D.; Thomas, K.; Presser, A.; Bernstein, B. E.; van Oudenaarden, A.; Regev, A.; Lander, E. S. ; Rinn, J. L. (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, Y. J.; Park, S. J.; Choi, E. Y.; Kim, S.; Kwak, H. J.; Yoo, B. C.; Yoo, H.; Lee, S. H.; Kim, D.; Park, J. B. ; Kim, J. H. (2011). "PTEN modulates miR-21 processing via RNA-regulatory protein RNH1." PLoS One 6(12): e28308.

Koboldt, D. C.; Steinberg, K. M.; Larson, D. E.; Wilson, R. K. ; Mardis, E. R. (2013). "The next-generation sequencing revolution and its impact on genomics." Cell 155(1): 27-38.

Konig, J.; Zarnack, K.; Luscombe, N. M. ; Ule, J. (2011). "Protein-RNA interactions: new genomic technologies and perspectives." Nat Rev Genet 13(2): 77-83.

Kornienko, A. E.; Guenzl, P. M.; Barlow, D. P. ; Pauler, F. M. (2013). "Gene regulation by the act of long non-coding RNA transcription." BMC Biol 11: 59.

Kozarewa, I.; Ning, Z.; Quail, M. A.; Sanders, M. J.; Berriman, M. ; Turner, D. J. (2009). "Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes." Nat Methods 6(4): 291-295.

Kozomara, A. ; Griffiths-Jones, S. (2011). "miRBase: integrating microRNA annotation and deep-sequencing data." Nucleic Acids Res 39(Database issue): D152-157.

Krichevsky, A. M.; King, K. S.; Donahue, C. P.; Khrapko, K. ; Kosik, K. S. (2003). "A microRNA array reveals extensive regulation of microRNAs during brain development." Rna 9(10): 1274-1281.

Krol, J.; Loedige, I. ; Filipowicz, W. (2010). "The widespread regulation of microRNA biogenesis, function and decay." Nat Rev Genet 11(9): 597-610.

Kung, J. T.; Colognori, D. ; Lee, J. T. (2013). "Long noncoding RNAs: past, present, and future." Genetics 193(3): 651-669.

Lagos-Quintana, M.; Rauhut, R.; Meyer, J.; Borkhardt, A. ; Tuschl, T. (2003). "New microRNAs from mouse and human." Rna 9(2): 175-179.

Lam, L. T.; Lu, X.; Zhang, H.; Lesniewski, R.; Rosenberg, S. ; Semizarov, D. (2010). "A microRNA screen to identify modulators of sensitivity to BCL2 inhibitor ABT-263 (navitoclax)." Mol Cancer Ther 9(11): 2943-2950.

Landry, J. J.; Pyl, P. T.; Rausch, T.; Zichner, T.; Tekkedil, M. M.; Stutz, A. M.; Jauch, A.; Aiyar, R. S.; Pau, G.; Delhomme, N.; Gagneur, J.; Korbel, J. O.; Huber, W. ; Steinmetz, L. M. (2013). "The genomic and transcriptomic landscape of a HeLa cell line." G3 (Bethesda) 3(8): 1213-1224.

Latos, P. A.; Pauler, F. M.; Koerner, M. V.; Senergin, H. B.; Hudson, Q. J.; Stocsits, R. R.; Allhoff, W.; Stricker, S. H.; Klement, R. M.; Warczok, K. E.; Aumayr, K.; Pasierbek, P. ; Barlow, D. P. (2012). "Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing." Science 338(6113): 1469-1472.

Lebedeva, S.; Jens, M.; Theil, K.; Schwanhausser, B.; Selbach, M.; Landthaler, M. ; Rajewsky, N. (2011). "Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR." Mol Cell 43(3): 340-352.

Lecuyer, E.; Yoshida, H.; Parthasarathy, N.; Alm, C.; Babak, T.; Cerovina, T.; Hughes, T. R.; Tomancak, P. ; Krause, H. M. (2007). "Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function." Cell 131(1): 174-187.

Lee, J. T. (2012). "Epigenetic regulation by long noncoding RNAs." Science 338(6113): 1435-1439.

Lee, J. T. ; Bartolomei, M. S. (2013). "X-inactivation, imprinting, and long noncoding RNAs in health and disease." Cell 152(6): 1308-1323.

Lee, R. C.; Feinbaum, R. L. ; Ambros, V. (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell 75(5): 843-854.

Lee, Y.; Kim, M.; Han, J.; Yeom, K. H.; Lee, S.; Baek, S. H. ; Kim, V. N. (2004). "MicroRNA genes are transcribed by RNA polymerase II." EMBO J 23(20): 4051-4060.

Lerner, M. R. ; Steitz, J. A. (1979). "Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus." Proc Natl Acad Sci U S A 76(11): 5495-5499.

Leucci, E.; Patella, F.; Waage, J.; Holmstrom, K.; Lindow, M.; Porse, B.; Kauppinen, S. ; Lund, A. H. (2013). "microRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus." Sci Rep 3: 2535.

Leung, A.; Trac, C.; Jin, W.; Lanting, L.; Akbany, A.; Saetrom, P.; Schones, D. E. ; Natarajan, R. (2013). "Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells." Circ Res 113(3): 266-278.

Levin, J. Z.; Yassour, M.; Adiconis, X.; Nusbaum, C.; Thompson, D. A.; Friedman, N.; Gnirke, A. ; Regev, A. (2010). "Comprehensive comparative analysis of strand-specific RNA sequencing methods." Nat Methods 7(9): 709-715.

Lipson, D.; Raz, T.; Kieu, A.; Jones, D. R.; Giladi, E.; Thayer, E.; Thompson, J. F.; Letovsky, S.; Milos, P. ; Causey, M. (2009). "Quantification of the yeast transcriptome by single-molecule sequencing." Nat Biotechnol 27(7): 652-658.

Lister, R.; O'Malley, R. C.; Tonti-Filippini, J.; Gregory, B. D.; Berry, C. C.; Millar, A. H. ; Ecker, J. R. (2008). "Highly integrated single-base resolution maps of the epigenome in Arabidopsis." Cell 133(3): 523-536.

Liu, C. G.; Calin, G. A.; Meloon, B.; Gamliel, N.; Sevignani, C.; Ferracin, M.; Dumitru, C. D.; Shimizu, M.; Zupo, S.; Dono, M.; Alder, H.; Bullrich, F.; Negrini, M. ; Croce, C. M. (2004). "An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues." Proc Natl Acad Sci U S A 101(26): 9740-9744.

Liu, J.; Carmell, M. A.; Rivas, F. V.; Marsden, C. G.; Thomson, J. M.; Song, J. J.; Hammond, S. M.; Joshua-Tor, L. ; Hannon, G. J. (2004). "Argonaute2 is the catalytic engine of mammalian RNAi." Science 305(5689): 1437-1441.

Llave, C.; Xie, Z.; Kasschau, K. D. ; Carrington, J. C. (2002). "Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA." Science 297(5589): 2053-2056.

Louro, R.; Nakaya, H. I.; Amaral, P. P.; Festa, F.; Sogayar, M. C.; da Silva, A. M.; Verjovski-Almeida, S. ; Reis, E. M. (2007). "Androgen responsive intronic non-coding RNAs." Bmc Biology 5: -.

Lund, E.; Guttinger, S.; Calado, A.; Dahlberg, J. E. ; Kutay, U. (2004). "Nuclear export of microRNA precursors." Science 303(5654): 95-98.

Ma, L.; Bajic, V. B. ; Zhang, Z. (2013). "On the classification of long non-coding RNAs." RNA Biol 10(6): 925-933.

Maeda, N.; Kasukawa, T.; Oyama, R.; Gough, J.; Frith, M.; Engstrom, P. G.; Lenhard, B.; Aturaliya, R. N.; Batalov, S.; Beisel, K. W.; Bult, C. J.; Fletcher, C. F.; Forrest, A. R.; Furuno, M.; Hill, D.; Itoh, M.; Kanamori-Katayama, M.; Katayama, S.; Katoh, M.; Kawashima, T.; Quackenbush, J.; Ravasi, T.; Ring, B. Z.; Shibata, K.; Sugiura, K.; Takenaka, Y.; Teasdale, R. D.; Wells, C. A.; Zhu, Y.; Kai, C.; Kawai, J.; Hume, D. A.; Carninci, P. ; Hayashizaki, Y. (2006). "Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs." PLoS Genet 2(4): e62.

Maher, C. A.; Kumar-Sinha, C.; Cao, X. H.; Kalyana-Sundaram, S.; Han, B.; Jing, X. J.; Sam, L.; Barrette, T.; Palanisamy, N. ; Chinnaiyan, A. M. (2009). "Transcriptome sequencing to detect gene fusions in cancer." Nature 458(7234): 97-U99.

Mamanova, L.; Andrews, R. M.; James, K. D.; Sheridan, E. M.; Ellis, P. D.; Langford, C. F.; Ost, T. W. B.; Collins, J. E. ; Turner, D. J. (2010). "FRT-seq: amplification-free, strand-specific transcriptome sequencing." Nature Methods 7(2): 130-U163.

Maragkakis, M.; Reczko, M.; Simossis, V. A.; Alexiou, P.; Papadopoulos, G. L.; Dalamagas, T.; Giannopoulos, G.; Goumas, G.; Koukis, E.; Kourtis, K.; Vergoulis, T.; Koziris, N.; Sellis, T.; Tsanakas, P. ; Hatzigeorgiou, A. G. (2009). "DIANA-microT web server: elucidating microRNA functions through target prediction." Nucleic Acids Res 37(Web Server issue): W273-276.

Mardis, E. R. (2013). "Next-generation sequencing platforms." Annu Rev Anal Chem (Palo Alto Calif) 6: 287-303.

Maricic, T. ; Paabo, S. (2009). "Optimization of 454 sequencing library preparation from small amounts of DNA permits sequence determination of both DNA strands." Biotechniques 46(1): 51-+.

Martin, K. C. ; Ephrussi, A. (2009). "mRNA localization: gene expression in the spatial dimension." Cell 136(4): 719-730.

Martins, L. F.; Antunes, L. P.; Pascon, R. C.; de Oliveira, J. C.; Digiampietri, L. A.; Barbosa, D.; Peixoto, B. M.; Vallim, M. A.; Viana-Niero, C.; Ostroski, E. H.; Telles, G. P.; Dias, Z.; da Cruz, J. B.; Juliano, L.; Verjovski-Almeida, S.; da Silva, A. M. ; Setubal, J. C. (2013). "Metagenomic analysis of a tropical composting operation at the sao paulo zoo park reveals diversity of biomass degradation functions and organisms." PLoS One 8(4): e61928.

Mattick, J. S. (2007). "A new paradigm for developmental biology." J Exp Biol 210(Pt 9): 1526-1547.

Mattick, J. S. (2011). "The central role of RNA in human development and cognition." FEBS Lett 585(11): 1600-1616.

Mattick, J. S. (2012). "RNA driving the epigenetic bus." EMBO J 31(3): 515-516.

Meister, G.; Landthaler, M.; Patkaniowska, A.; Dorsett, Y.; Teng, G. ; Tuschl, T. (2004). "Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs." Mol Cell 15(2): 185-197.

Memczak, S.; Jens, M.; Elefsinioti, A.; Torti, F.; Krueger, J.; Rybak, A.; Maier, L.; Mackowiak, S. D.; Gregersen, L. H.; Munschauer, M.; Loewer, A.; Ziebold, U.; Landthaler, M.; Kocks, C.; le Noble, F. ; Rajewsky, N. (2013). "Circular RNAs are a large class of animal RNAs with regulatory potency." Nature 495(7441): 333-338.

Mendell, J. T. ; Olson, E. N. (2012). "MicroRNAs in stress signaling and human disease." Cell 148(6): 1172-1187.

Mercer, T. R.; Dinger, M. E. ; Mattick, J. S. (2009). "Long non-coding RNAs: insights into functions." Nat Rev Genet 10(3): 155-159.

Mercer, T. R.; Dinger, M. E.; Sunkin, S. M.; Mehler, M. F. ; Mattick, J. S. (2008). "Specific expression of long noncoding RNAs in the mouse brain." Proc Natl Acad Sci U S A 105(2): 716-721.

Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nat Rev Genet 11(1): 31-46.

Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L. ; Wold, B. (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods 5(7): 621-628.

Motorin, Y.; Lyko, F. ; Helm, M. (2010). "5-methylcytosine in RNA: detection, enzymatic formation and biological functions." Nucleic Acids Res 38(5): 1415-1430.

Mukherjee, N.; Corcoran, D. L.; Nusbaum, J. D.; Reid, D. W.; Georgiev, S.; Hafner, M.; Ascano, M., Jr.; Tuschl, T.; Ohler, U. ; Keene, J. D. (2011). "Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability." Mol Cell 43(3): 327-339.

Nagano, T. ; Fraser, P. (2011). "No-nonsense functions for long noncoding RNAs." Cell 145(2): 178-181.

Nagasaki, H.; Arita, M.; Nishizawa, T.; Suwa, M. ; Gotoh, O. (2005). "Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes." Gene 364: 53-62.

Nakaya, H. I.; Amaral, P. P.; Louro, R.; Lopes, A.; Fachel, A. A.; Moreira, Y. B.; El-Jundi, T. A.; da Silva, A. M.; Reis, E. M. ; Verjovski-Almeida, S. (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 Biology 8(3): -.

Nekrutenko, A. ; Taylor, J. (2012). "Next-generation sequencing data interpretation: enhancing reproducibility and accessibility." Nat Rev Genet 13(9): 667-672.

Nesterova, T. B.; Slobodyanyuk, S. Y.; Elisaphenko, E. A.; Shevchenko, A. I.; Johnston, C.; Pavlova, M. E.; Rogozin, I. B.; Kolesnikov, N. N.; Brockdorff, N. ; Zakian, S. M. (2001). "Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence." Genome Res 11(5): 833-849.

Nie, L.; Wu, H. J.; Hsu, J. M.; Chang, S. S.; Labaff, A. M.; Li, C. W.; Wang, Y.; Hsu, J. L. ; Hung, M. C. (2012). "Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer." Am J Transl Res 4(2): 127-150.

Orom, U. A.; Nielsen, F. C. ; Lund, A. H. (2008). "MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation." Mol Cell 30(4): 460-471.

Ozsolak, F.; Goren, A.; Gymrek, M.; Guttman, M.; Regev, A.; Bernstein, B. E. ; Milos, P. M. (2010). "Digital transcriptome profiling from attomole-level RNA samples." Genome Res 20(4): 519-525.

Ozsolak, F. ; Milos, P. M. (2011). "RNA sequencing: advances, challenges and opportunities." Nat Rev Genet 12(2): 87-98.

Ozsolak, F.; Platt, A. R.; Jones, D. R.; Reifenberger, J. G.; Sass, L. E.; McInerney, P.; Thompson, J. F.; Bowers, J.; Jarosz, M. ; Milos, P. M. (2009). "Direct RNA sequencing." Nature 461(7265): 814-818.

Paraskevopoulou, M. D.; Georgakilas, G.; Kostoulas, N.; Reczko, M.; Maragkakis, M.; Dalamagas, T. M. ; Hatzigeorgiou, A. G. (2013). "DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs." Nucleic Acids Res 41(Database issue): D239-245.

Parkhomchuk, D.; Borodina, T.; Amstislavskiy, V.; Banaru, M.; Hallen, L.; Krobitsch, S.; Lehrach, H. ; Soldatov, A. (2009). "Transcriptome analysis by strand-specific sequencing of complementary DNA." Nucleic Acids Res 37(18): e123.

Pasquinelli, A. E. (2012). "MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship." Nat Rev Genet 13(4): 271-282.

Poliseno, L.; Salmena, L.; Zhang, J.; Carver, B.; Haveman, W. J. ; Pandolfi, P. P. (2010). "A coding-independent function of gene and pseudogene mRNAs regulates tumour biology." Nature 465(7301): 1033-1038.

Ponting, C. P.; Oliver, P. L. ; Reik, W. (2009). "Evolution and functions of long noncoding RNAs." Cell 136(4): 629-641.

Prensner, J. R. ; Chinnaiyan, A. M. (2011). "The emergence of lncRNAs in cancer biology." Cancer Discov 1(5): 391-407.

Pruitt, K. D.; Tatusova, T. ; Maglott, D. R. (2007). "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins." Nucleic Acids Res 35(Database issue): D61-65.

Rajewsky, N. (2006). "microRNA target predictions in animals." Nat Genet 38 Suppl: S8-13.

Rapicavoli, N. A.; Qu, K.; Zhang, J.; Mikhail, M.; Laberge, R. M. ; Chang, H. Y. (2013). "A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics." Elife 2: e00762.

Rearick, D.; Prakash, A.; McSweeny, A.; Shepard, S. S.; Fedorova, L. ; Fedorov, A. (2011). "Critical association of ncRNA with introns." Nucleic Acids Res 39(6): 2357-2366.

Reczko, M.; Maragkakis, M.; Alexiou, P.; Grosse, I. ; Hatzigeorgiou, A. G. (2012). "Functional microRNA targets in protein coding sequences." Bioinformatics 28(6): 771-776.

Reis, E. M.; Nakaya, H. I.; Louro, R.; Canavez, F. C.; Flatschart, A. V. F.; Almeida, G. T.; Egidio, C. M.; Paquola, A. C.; Machado, A. A.; Festa, F.; Yamamoto, D.; Alvarenga, R.; da Silva, C. C.; Brito, G. C.; Simon, S. D.; Moreira, C. A.; Leite, K. R.; Camara-Lopes, L. H.; Campos, F. S.; Gimba, E.; Vignal, G. M.; El-Dorry, H.; Sogayar, M. C.; Barcinski, M. A.; da Silva, A. M. ; Verjovski-Almeida, S. (2004). "Antisense intronic non-coding RNA levels

correlate to the degree of tumor differentiation in prostate cancer." Oncogene 23(39): 6684-6692.

Reis, E. M. ; Verjovski-Almeida, S. (2012). "Perspectives of Long Non-Coding RNAs in Cancer Diagnostics." Front Genet 3: 32.

Riley, K. J.; Yario, T. A. ; Steitz, J. A. (2012). "Association of Argonaute proteins and microRNAs can occur after cell lysis." RNA 18(9): 1581-1585.

Rinn, J. L. ; Chang, H. Y. (2012). "Genome regulation by long noncoding RNAs." Annu Rev Biochem 81: 145-166.

Rinn, J. L.; Kertesz, M.; Wang, J. K.; Squazzo, S. L.; Xu, X.; Brugmann, S. A.; Goodnough, L. H.; Helms, J. A.; Farnham, P. J.; Segal, E. ; Chang, H. Y. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by Noncoding RNAs." Cell 129(7): 1311-1323.

Rivera, C. M. ; Ren, B. (2013). "Mapping human epigenomes." Cell 155(1): 39-55.

Roush, S. ; Slack, F. J. (2008). "The let-7 family of microRNAs." Trends Cell Biol 18(10): 505-516.

Rowan, B. A.; Weigel, D. ; Koenig, D. (2011). "Developmental genetics and new sequencing technologies: the rise of nonmodel organisms." Dev Cell 21(1): 65-76.

Roy, S.; Ernst, J.; Kharchenko, P. V.; Kheradpour, P.; Negre, N.; Eaton, M. L.; Landolin, J. M.; Bristow, C. A.; Ma, L.; Lin, M. F.; Washietl, S.; Arshinoff, B. I.; Ay, F.; Meyer, P. E.; Robine, N.; Washington, N. L.; Di Stefano, L.; Berezikov, E.; Brown, C. D.; Candeias, R.; Carlson, J. W.; Carr, A.; Jungreis, I.; Marbach, D.; Sealfon, R.; Tolstorukov, M. Y.; Will, S.; Alekseyenko, A. A.; Artieri, C.; Booth, B. W.; Brooks, A. N.; Dai, Q.; Davis, C. A.; Duff, M. O.; Feng, X.; Gorchakov, A. A.; Gu, T.; Henikoff, J. G.; Kapranov, P.; Li, R.; MacAlpine, H. K.; Malone, J.; Minoda, A.; Nordman, J.; Okamura, K.; Perry, M.; Powell, S. K.; Riddle, N. C.; Sakai, A.; Samsonova, A.; Sandler, J. E.; Schwartz, Y. B.; Sher, N.; Spokony, R.; Sturgill, D.; van Baren, M.; Wan, K. H.; Yang, L.; Yu, C.; Feingold, E.; Good, P.; Guyer, M.; Lowdon, R.; Ahmad, K.; Andrews, J.; Berger, B.; Brenner, S. E.; Brent, M. R.; Cherbas, L.; Elgin, S. C.; Gingeras, T. R.; Grossman, R.; Hoskins, R. A.; Kaufman, T. C.; Kent, W.; Kuroda, M. I.; Orr-Weaver, T.; Perrimon, N.; Pirrotta, V.; Posakony, J. W.; Ren, B.; Russell, S.; Cherbas, P.; Graveley, B. R.; Lewis, S.; Micklem, G.; Oliver, B.; Park, P. J.; Celniker, S. E.; Henikoff, S.; Karpen, G. H.; Lai, E. C.; MacAlpine, D. M.; Stein, L. D.; White, K. P. ; Kellis, M. (2010). "Identification of functional elements and regulatory circuits by Drosophila modENCODE." Science 330(6012): 1787-1797.

Sabin, L. R.; Delas, M. J. ; Hannon, G. J. (2013). "Dogma derailed: the many influences of RNA on the genome." Mol Cell 49(5): 783-794.

Sandberg, J.; Stahl, P. L.; Ahmadian, A.; Bjursell, M. K. ; Lundeberg, J. (2009). "Flow cytometry for enrichment and titration in massively parallel DNA sequencing." Nucleic Acids Research 37(8): -.

Sasaki, Y. T.; Ideue, T.; Sano, M.; Mituyama, T. ; Hirose, T. (2009). "MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles." Proc Natl Acad Sci U S A 106(8): 2525-2530.

Schneider, G. F. ; Dekker, C. (2012). "DNA sequencing with nanopores." Nat Biotechnol 30(4): 326-328.

Schorderet, P. ; Duboule, D. (2011). "Structural and functional differences in the long non-coding RNA hotair in mouse and human." PLoS Genet 7(5): e1002071.

Sheik Mohamed, J.; Gaughwin, P. M.; Lim, B.; Robson, P. ; Lipovich, L. (2010). "Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells." RNA 16(2): 324-337.

Shin, C.; Nam, J. W.; Farh, K. K.; Chiang, H. R.; Shkumatava, A. ; Bartel, D. P. (2010). "Expanding the microRNA targeting code: functional sites with centered pairing." Mol Cell 38(6): 789-802.

Sibley, C. R.; Seow, Y.; Saayman, S.; Dijkstra, K. K.; El Andaloussi, S.; Weinberg, M. S. ; Wood, M. J. (2012). "The biogenesis and characterization of mammalian microRNAs of mirtron origin." Nucleic Acids Res 40(1): 438-448.

Simons, J. F. ; Hutchison, S. K. (2007) "RNA Virus Sequencing Using the Genome Sequencer FLX System." Application Note No. 8 / October 2007.

Song, Y. X.; Yue, Z. Y.; Wang, Z. N.; Xu, Y. Y.; Luo, Y.; Xu, H. M.; Zhang, X.; Jiang, L.; Xing, C. Z. ; Zhang, Y. (2011). "MicroRNA-148b is frequently down-regulated in gastric cancer and acts as a tumor suppressor by inhibiting cell proliferation." Mol Cancer 10: 1.

Spiegelman, S.; Burny, A.; Das, M. R.; Keydar, J.; Schlom, J.; Travnicek, M. ; Watson, K. (1970). "DNA-directed DNA polymerase activity in oncogenic RNA viruses." Nature 227(5262): 1029-1031.

Spizzo, R.; Almeida, M. I.; Colombatti, A. ; Calin, G. A. (2012). "Long non-coding RNAs and cancer: a new frontier of translational research?" Oncogene 31(43): 4577-4587.

St Laurent, G.; Shtokalo, D.; Tackett, M. R.; Yang, Z.; Eremina, T.; Wahlestedt, C.; Urcuqui-Inchima, S.; Seilheimer, B.; McCaffrey, T. A. ; Kapranov, P. (2012). "Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells." BMC Genomics 13: 504.

Struhl, K. (2007). "Transcriptional noise and the fidelity of initiation by RNA polymerase II." Nat Struct Mol Biol 14(2): 103-105.

Sun, L.; Goff, L. A.; Trapnell, C.; Alexander, R.; Lo, K. A.; Hacisuleyman, E.; Sauvageau, M.; Tazon-Vega, B.; Kelley, D. R.; Hendrickson, D. G.; Yuan, B.; Kellis, M.; Lodish, H. F. ; Rinn, J. L. (2013). "Long noncoding RNAs regulate adipogenesis." Proc Natl Acad Sci U S A 110(9): 3387-3392.

Sunwoo, H.; Dinger, M. E.; Wilusz, J. E.; Amaral, P. P.; Mattick, J. S. ; Spector, D. L. (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.

Taft, R. J.; Glazov, E. A.; Cloonan, N.; Simons, C.; Stephen, S.; Faulkner, G. J.; Lassmann, T.; Forrest, A. R.; Grimmond, S. M.; Schroder, K.; Irvine, K.; Arakawa, T.; Nakamura, M.; Kubosaki, A.; Hayashida, K.; Kawazu, C.; Murata, M.; Nishiyori, H.; Fukuda, S.; Kawai, J.; Daub, C. O.; Hume, D. A.; Suzuki, H.; Orlando, V.; Carninci, P.; Hayashizaki, Y. ;

Mattick, J. S. (2009). "Tiny RNAs associated with transcription start sites in animals." Nat Genet 41(5): 572-578.

Taft, R. J.; Pang, K. C.; Mercer, T. R.; Dinger, M. ; Mattick, J. S. (2010). "Non-coding RNAs: regulators of disease." J Pathol 220(2): 126-139.

Taft, R. J.; Pheasant, M. ; Mattick, J. S. (2007). "The relationship between non-protein-coding DNA and eukaryotic complexity." Bioessays 29(3): 288-299.

Takayama, K.; Horie-Inoue, K.; Katayama, S.; Suzuki, T.; Tsutsumi, S.; Ikeda, K.; Urano, T.; Fujimura, T.; Takagi, K.; Takahashi, S.; Homma, Y.; Ouchi, Y.; Aburatani, H.; Hayashizaki, Y. ; Inoue, S. (2013). "Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer." EMBO J 32(12): 1665-1680.

Tripathi, V.; Ellis, J. D.; Shen, Z.; Song, D. Y.; Pan, Q.; Watt, A. T.; Freier, S. M.; Bennett, C. F.; Sharma, A.; Bubulya, P. A.; Blencowe, B. J.; Prasanth, S. G. ; Prasanth, K. V. (2010). "The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation." Mol Cell 39(6): 925-938.

Tripathi, V.; Shen, Z.; Chakraborty, A.; Giri, S.; Freier, S. M.; Wu, X.; Zhang, Y.; Gorospe, M.; Prasanth, S. G.; Lal, A. ; Prasanth, K. V. (2013). "Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB." PLoS Genet 9(3): e1003368.

Tuck, A. C. ; Tollervey, D. (2013). "A Transcriptome-wide Atlas of RNP Composition Reveals Diverse Classes of mRNAs and lncRNAs." Cell 154(5): 996-1009.

Ule, J.; Jensen, K.; Mele, A. ; Darnell, R. B. (2005). "CLIP: a method for identifying protein-RNA interaction sites in living cells." Methods 37(4): 376-386.

Ulitsky, I. ; Bartel, D. P. (2013). "lincRNAs: genomics, evolution, and mechanisms." Cell 154(1): 26-46.

van Bakel, H.; Nislow, C.; Blencowe, B. J. ; Hughes, T. R. (2010). "Most "dark matter" transcripts are associated with known genes." PLoS Biol 8(5): e1000371.

Wahlestedt, C. (2013). "Targeting long non-coding RNA to therapeutically upregulate gene expression." Nat Rev Drug Discov 12(6): 433-446.

Wan, G.; Hu, X.; Liu, Y.; Han, C.; Sood, A. K.; Calin, G. A.; Zhang, X. ; Lu, X. (2013). "A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation." EMBO J.

Wang, D.; Garcia-Bassets, I.; Benner, C.; Li, W.; Su, X.; Zhou, Y.; Qiu, J.; Liu, W.; Kaikkonen, M. U.; Ohgi, K. A.; Glass, C. K.; Rosenfeld, M. G. ; Fu, X. D. (2011). "Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA." Nature 474(7351): 390-394.

Wang, D.; Zhang, Z.; O'Loughlin, E.; Lee, T.; Houel, S.; O'Carroll, D.; Tarakhovsky, A.; Ahn, N. G. ; Yi, R. (2012). "Quantitative functions of Argonaute proteins in mammalian development." Genes Dev 26(7): 693-704.

Wang, K. C. ; Chang, H. Y. (2011). "Molecular mechanisms of long noncoding RNAs." Mol Cell 43(6): 904-914.

Wang, Z.; Gerstein, M. ; Snyder, M. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nature Reviews Genetics 10(1): 57-63.

Wapinski, O. ; Chang, H. Y. (2011). "Long noncoding RNAs and human disease." Trends Cell Biol 21(6): 354-361.

Waterston, R. H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J. F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P.; Antonarakis, S. E.; Attwood, J.; Baertsch, R.; Bailey, J.; Barlow, K.; Beck, S.; Berry, E.; Birren, B.; Bloom, T.; Bork, P.; Botcherby, M.; Bray, N.; Brent, M. R.; Brown, D. G.; Brown, S. D.; Bult, C.; Burton, J.; Butler, J.; Campbell, R. D.; Carninci, P.; Cawley, S.; Chiaromonte, F.; Chinwalla, A. T.; Church, D. M.; Clamp, M.; Clee, C.; Collins, F. S.; Cook, L. L.; Copley, R. R.; Coulson, A.; Couronne, O.; Cuff, J.; Curwen, V.; Cutts, T.; Daly, M.; David, R.; Davies, J.; Delehaunty, K. D.; Deri, J.; Dermitzakis, E. T.; Dewey, C.; Dickens, N. J.; Diekhans, M.; Dodge, S.; Dubchak, I.; Dunn, D. M.; Eddy, S. R.; Elnitski, L.; Emes, R. D.; Eswara, P.; Eyras, E.; Felsenfeld, A.; Fewell, G. A.; Flicek, P.; Foley, K.; Frankel, W. N.; Fulton, L. A.; Fulton, R. S.; Furey, T. S.; Gage, D.; Gibbs, R. A.; Glusman, G.; Gnerre, S.; Goldman, N.; Goodstadt, L.; Grafham, D.; Graves, T. A.; Green, E. D.; Gregory, S.; Guigo, R.; Guyer, M.; Hardison, R. C.; Haussler, D.; Hayashizaki, Y.; Hillier, L. W.; Hinrichs, A.; Hlavina, W.; Holzer, T.; Hsu, F.; Hua, A.; Hubbard, T.; Hunt, A.; Jackson, I.; Jaffe, D. B.; Johnson, L. S.; Jones, M.; Jones, T. A.; Joy, A.; Kamal, M.; Karlsson, E. K.; Karolchik, D.; Kasprzyk, A.; Kawai, J.; Keibler, E.; Kells, C.; Kent, W. J.; Kirby, A.; Kolbe, D. L.; Korf, I.; Kucherlapati, R. S.; Kulbokas, E. J.; Kulp, D.; Landers, T.; Leger, J. P.; Leonard, S.; Letunic, I.; Levine, R.; Li, J.; Li, M.; Lloyd, C.; Lucas, S.; Ma, B.; Maglott, D. R.; Mardis, E. R.; Matthews, L.; Mauceli, E.; Mayer, J. H.; McCarthy, M.; McCombie, W. R.; McLaren, S.; McLay, K.; McPherson, J. D.; Meldrim, J.; Meredith, B.; Mesirov, J. P.; Miller, W.; Miner, T. L.; Mongin, E.; Montgomery, K. T.; Morgan, M.; Mott, R.; Mullikin, J. C.; Muzny, D. M.; Nash, W. E.; Nelson, J. O.; Nhan, M. N.; Nicol, R.; Ning, Z.; Nusbaum, C.; O'Connor, M. J.; Okazaki, Y.; Oliver, K.; Overton-Larty, E.; Pachter, L.; Parra, G.; Pepin, K. H.; Peterson, J.; Pevzner, P.; Plumb, R.; Pohl, C. S.; Poliakov, A.; Ponce, T. C.; Ponting, C. P.; Potter, S.; Quail, M.; Reymond, A.; Roe, B. A.; Roskin, K. M.; Rubin, E. M.; Rust, A. G.; Santos, R.; Sapojnikov, V.; Schultz, B.; Schultz, J.; Schwartz, M. S.; Schwartz, S.; Scott, C.; Seaman, S.; Searle, S.; Sharpe, T.; Sheridan, A.; Shownkeen, R.; Sims, S.; Singer, J. B.; Slater, G.; Smit, A.; Smith, D. R.; Spencer, B.; Stabenau, A.; Stange-Thomann, N.; Sugnet, C.; Suyama, M.; Tesler, G.; Thompson, J.; Torrents, D.; Trevaskis, E.; Tromp, J.; Ucla, C.; Ureta-Vidal, A.; Vinson, J. P.; Von Niederhausern, A. C.; Wade, C. M.; Wall, M.; Weber, R. J.; Weiss, R. B.; Wendl, M. C.; West, A. P.; Wetterstrand, K.; Wheeler, R.; Whelan, S.; Wierzbowski, J.; Willey, D.; Williams, S.; Wilson, R. K.; Winter, E.; Worley, K. C.; Wyman, D.; Yang, S.; Yang, S. P.; Zdobnov, E. M.; Zody, M. C. ; Lander, E. S. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature 420(6915): 520-562.

Wheeler, D. A.; Srinivasan, M.; Egholm, M.; Shen, Y.; Chen, L.; McGuire, A.; He, W.; Chen, Y. J.; Makhijani, V.; Roth, G. T.; Gomes, X.; Tartaro, K.; Niazi, F.; Turcotte, C. L.; Irzyk, G.

P.; Lupski, J. R.; Chinault, C.; Song, X. Z.; Liu, Y.; Yuan, Y.; Nazareth, L.; Qin, X.; Muzny, D. M.; Margulies, M.; Weinstock, G. M.; Gibbs, R. A. ; Rothberg, J. M. (2008). "The complete genome of an individual by massively parallel DNA sequencing." Nature 452(7189): 872-876.

Wienholds, E.; Kloosterman, W. P.; Miska, E.; Alvarez-Saavedra, E.; Berezikov, E.; de Bruijn, E.; Horvitz, H. R.; Kauppinen, S. ; Plasterk, R. H. (2005). "MicroRNA expression in zebrafish embryonic development." Science 309(5732): 310-311.

Witkos, T. M.; Koscianska, E. ; Krzyzosiak, W. J. (2011). "Practical Aspects of microRNA Target Prediction." Curr Mol Med 11(2): 93-109.

Wright, M. W. ; Bruford, E. A. (2011). "Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature." Hum Genomics 5(2): 90-98.

Wu, Q.; Kim, Y. C.; Lu, J.; Xuan, Z.; Chen, J.; Zheng, Y.; Zhou, T.; Zhang, M. Q.; Wu, C. I. ; Wang, S. M. (2008). "Poly A- transcripts expressed in HeLa cells." PLoS One 3(7): e2803.

Wu, S. C.; Kallin, E. M. ; Zhang, Y. (2010). "Role of H3K27 methylation in the regulation of lncRNA expression." Cell Res 20(10): 1109-1116.

Yang, L.; Duff, M. O.; Graveley, B. R.; Carmichael, G. G. ; Chen, L. L. (2011). "Genomewide characterization of non-polyadenylated RNAs." Genome Biol 12(2): R16.

Yang, L.; Lin, C.; Jin, C.; Yang, J. C.; Tanasa, B.; Li, W.; Merkurjev, D.; Ohgi, K. A.; Meng, D.; Zhang, J.; Evans, C. P. ; Rosenfeld, M. G. (2013). "lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs." Nature 500(7464): 598-602.

Yi, R.; Qin, Y.; Macara, I. G. ; Cullen, B. R. (2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." Genes Dev 17(24): 3011-3016.

Yu, B. ; Wang, H. (2010). "Translational inhibition by microRNAs in plants." Prog Mol Subcell Biol 50: 41-57.

Zhang, Y.; Zhang, X. O.; Chen, T.; Xiang, J. F.; Yin, Q. F.; Xing, Y. H.; Zhu, S.; Yang, L. ; Chen, L. L. (2013). "Circular Intronic Long Noncoding RNAs." Mol Cell 51(6): 792-806.

Zhang, Z.; Zhu, Z.; Watabe, K.; Zhang, X.; Bai, C.; Xu, M.; Wu, F. ; Mo, Y. Y. (2013). "Negative regulation of lncRNA GAS5 by miR-21." Cell Death Differ 20(11): 1558-1568.

Zhao, J.; Ohsumi, T. K.; Kung, J. T.; Ogawa, Y.; Grau, D. J.; Sarma, K.; Song, J. J.; Kingston, R. E.; Borowsky, M. ; Lee, J. T. (2010). "Genome-wide identification of polycomb-associated RNAs by RIP-seq." Mol Cell 40(6): 939-953.

APÊNDICE

Apêndice 1 – Lista dos 686 RNAs longos não codificadores de proteínas detectados ligados ao complexo RISC nos sequenciamentos RIP-Anti-Ago2. Os lncRNAs estão ordenados de acordo com a anotação do GENCODE, e estão mostrados na seguinte sequência: antisense, lincRNAs, miRNAs (ou seja, transcritos que dão origem a miRNAs), non-coding RNAs, polymorphic-pseudogenes, processed-transcripts (ou seja, que não possuem ORFs de acordo com a definição do GENCODE) e pseudogenes. Dentro de cada classe de anotação estão mostrados primeiro os transcritos exclusivos (lncRNAs detectados exclusivamente nos ensaios RIP-Anti-Ago2) ordenados pelo RPKM, e depois os lncRNAs detectados no ensaio RIP-Anti-Ago2 com razão de enriquecimento (RPKM RIP-Anti-Ago2/ RPKM RIP-Controle) igual ou superior a quatro vezes em relação aos ensaios RIP-Controle.

Símbolo do lncRNA	Classe (anotação do GENCODE)	Razão RPKM RIP-Anti-Ago2/ RPKM RIP-Controle	RPKM RIP-Anti-Ago2
RP11-950C14.7	antisense	Exclusivo	17,62
RP11-798M19.3	antisense	Exclusivo	14,82
AP001631.9	antisense	Exclusivo	11,92
RP4-803J11.2	antisense	Exclusivo	11,23
RP13-39P12.3	antisense	Exclusivo	9,68
RP11-144G6.4	antisense	Exclusivo	9,21
AP001601.2	antisense	Exclusivo	7,45
EIF1AX-AS1	antisense	Exclusivo	6,63
RP11-123B3.2	antisense	Exclusivo	6,61
RP11-656E20.5	antisense	Exclusivo	5,99
RP11-857B24.5	antisense	Exclusivo	5,99
CTD-2555O16.2	antisense	Exclusivo	5,84
RP11-881M11.4	antisense	Exclusivo	5,39
RP11-613M10.6	antisense	Exclusivo	4,97
RP11-16P20.3	antisense	Exclusivo	4,92
RP11-626G11.3	antisense	Exclusivo	4,89
AC091736.11	antisense	Exclusivo	4,76
AC005037.3	antisense	Exclusivo	4,68
RP11-335O4.3	antisense	Exclusivo	4,62
IGBP1-AS1	antisense	Exclusivo	4,62
RP11-8J9.4	antisense	Exclusivo	4,47
AC005104.3	antisense	Exclusivo	4,46
AC004237.1	antisense	Exclusivo	4,29
RP11-632F7.3	antisense	Exclusivo	4,13
RP11-126K1.6	antisense	Exclusivo	4,05
CTD-2203K17.1	antisense	Exclusivo	4,01
RP4-800F24.1	antisense	Exclusivo	3,91
RP11-342M3.5	antisense	Exclusivo	3,90
CTD-2013N24.2	antisense	Exclusivo	3,88
ALG13-AS1	antisense	Exclusivo	3,69
AC079145.4	antisense	Exclusivo	3,64
AC131971.3	antisense	Exclusivo	3,60
TBL1XR1-AS1	antisense	Exclusivo	3,47
RP11-456I15.2	antisense	Exclusivo	3,42
DIPAH3-AS2	antisense	Exclusivo	3,41
CTD-2515A14.1	antisense	Exclusivo	3,40
RP11-69D4.5	antisense	Exclusivo	3,38
CTD-3065B20.3	antisense	Exclusivo	3,33
AP000695.6	antisense	Exclusivo	3,22

RP11-677I18.3	antisense	Exclusivo	3,20
RP11-715J22.3	antisense	Exclusivo	3,15
RP11-473O4.5	antisense	Exclusivo	3,13
AJ003147.9	antisense	Exclusivo	3,13
RP5-1085F17.4	antisense	Exclusivo	3,13
RP11-432I5.6	antisense	Exclusivo	3,08
RP11-398C13.4	antisense	Exclusivo	3,08
RP1-292L20.1	antisense	Exclusivo	3,07
AC008074.3	antisense	Exclusivo	2,98
RP11-656E20.4	antisense	Exclusivo	2,98
AC091133.1	antisense	Exclusivo	2,98
RP11-266N13.2	antisense	Exclusivo	2,96
HM13-AS1	antisense	Exclusivo	2,94
RP5-875H18.4	antisense	Exclusivo	2,91
RP11-66D17.5	antisense	Exclusivo	2,89
RP4-781K5.2	antisense	Exclusivo	2,84
RP11-124N14.3	antisense	Exclusivo	2,79
RP11-165P7.1	antisense	Exclusivo	2,68
CTD-3065B20.2	antisense	Exclusivo	2,64
AF001548.5	antisense	Exclusivo	2,58
RP11-649G15.2	antisense	Exclusivo	2,48
AC007163.6	antisense	Exclusivo	2,46
IGBP1-AS2	antisense	Exclusivo	2,44
AC016738.4	antisense	Exclusivo	2,37
RP11-354M20.3	antisense	Exclusivo	2,33
RP11-872D17.4	antisense	Exclusivo	2,31
RP1-142L7.5	antisense	Exclusivo	2,30
RP13-638C3.3	antisense	Exclusivo	2,26
RP11-327F22.1	antisense	Exclusivo	2,25
RP11-415I12.5	antisense	Exclusivo	2,22
AP006216.12	antisense	Exclusivo	2,22
RP11-38L15.3	antisense	Exclusivo	2,21
RP5-864K19.4	antisense	Exclusivo	2,20
RP11-166N17.1	antisense	Exclusivo	2,16
RP1-67K17.3	antisense	Exclusivo	2,06
RP11-864G5.3	antisense	Exclusivo	1,99
LATS2-AS1	antisense	Exclusivo	1,98
AC005154.5	antisense	Exclusivo	1,91
RP11-376O6.2	antisense	Exclusivo	1,91
NAV2-AS1	antisense	Exclusivo	1,90
AC007271.3	antisense	Exclusivo	1,88
RP5-833A20.1	antisense	Exclusivo	1,87
AC012358.8	antisense	Exclusivo	1,86

CTD-2201E9.4	antisense	Exclusivo	1,84
RP11-537A6.9	antisense	Exclusivo	1,79
CTC-508F8.1	antisense	Exclusivo	1,77
HCG14	antisense	Exclusivo	1,74
RP11-540O11.7	antisense	Exclusivo	1,74
P4HA2-AS1	antisense	Exclusivo	1,71
RP11-542C16.1	antisense	Exclusivo	1,71
RP11-21M24.3	antisense	Exclusivo	1,70
RP11-265N6.1	antisense	Exclusivo	1,70
RP11-182N22.8	antisense	Exclusivo	1,69
RP11-707G14.1	antisense	Exclusivo	1,68
RP11-162D16.2	antisense	Exclusivo	1,63
RP11-310I9.1	antisense	Exclusivo	1,61
JARID2-AS1	antisense	Exclusivo	1,58
RP11-961A15.1	antisense	Exclusivo	1,57
RP11-20I23.6	antisense	Exclusivo	1,57
ZMYM4-AS1	antisense	Exclusivo	1,57
CTC-467M3.1	antisense	Exclusivo	1,55
RP4-781B1.2	antisense	Exclusivo	1,55
RP5-1050D4.3	antisense	Exclusivo	1,52
RP1-225E12.2	antisense	Exclusivo	1,51
RP11-109P14.10	antisense	Exclusivo	1,49
RP11-686G8.2	antisense	Exclusivo	1,48
RP5-956O18.2	antisense	Exclusivo	1,45
AC073934.6	antisense	Exclusivo	1,45
RP11-367H5.7	antisense	Exclusivo	1,45
RP3-511B24.5	antisense	Exclusivo	1,44
RP11-540O11.6	antisense	Exclusivo	1,44
RP3-325F22.3	antisense	Exclusivo	1,42
RP1-43E13.2	antisense	Exclusivo	1,42
AC016745.3	antisense	Exclusivo	1,41
RP11-490D19.6	antisense	Exclusivo	1,40
RP5-956O18.3	antisense	Exclusivo	1,40
MED14-AS1	antisense	Exclusivo	1,40
RP11-428C19.4	antisense	Exclusivo	1,39
AC093171.1	antisense	Exclusivo	1,39
RP11-346K17.4	antisense	Exclusivo	1,38
RP11-292F9.2	antisense	Exclusivo	1,36
AC021016.8	antisense	Exclusivo	1,34
RP11-80H5.4	antisense	Exclusivo	1,34
COL4A2-AS1	antisense	Exclusivo	1,33
CTD-2113L7.1	antisense	Exclusivo	1,33
AC100748.2	antisense	Exclusivo	1,33

RP11-656E20.3	antisense	Exclusivo	1,30
RP11-6E9.4	antisense	Exclusivo	1,30
RP11-359P5.1	antisense	Exclusivo	1,28
AF127577.10	antisense	Exclusivo	1,28
RP11-867O8.5	antisense	Exclusivo	1,27
CTC-506B8.1	antisense	Exclusivo	1,27
RP11-159F24.6	antisense	Exclusivo	1,26
IPO9-AS1	antisense	Exclusivo	1,26
RP11-180O5.2	antisense	Exclusivo	1,25
RP11-73M7.1	antisense	Exclusivo	1,22
RP1-59D14.1	antisense	Exclusivo	1,18
PRRT3-AS1	antisense	Exclusivo	1,17
RP11-361H10.2	antisense	Exclusivo	1,16
RP11-288I21.1	antisense	Exclusivo	1,11
RP5-884G6.2	antisense	Exclusivo	1,10
RP11-23N2.4	antisense	Exclusivo	1,09
RP11-441O15.3	antisense	Exclusivo	1,08
RP11-458D21.2	antisense	Exclusivo	1,08
RP11-659E9.4	antisense	Exclusivo	1,08
RP11-499O7.7	antisense	Exclusivo	1,08
CTC-338M12.5	antisense	Exclusivo	1,05
AC073072.5	antisense	Exclusivo	1,05
AC009473.1	antisense	Exclusivo	1,04
RP11-18C24.6	antisense	Exclusivo	1,04
CTD-3037G24.3	antisense	Exclusivo	1,03
MEIS1-AS3	antisense	Exclusivo	1,02
RP11-506O24.1	antisense	Exclusivo	1,00
RP11-466F5.8	antisense	Exclusivo	1,00
AC013269.4	antisense	Exclusivo	0,99
RP11-473I1.5	antisense	Exclusivo	0,98
RP11-506H20.1	antisense	Exclusivo	0,93
RP4-647C14.3	antisense	Exclusivo	0,92
AC007679.3	antisense	Exclusivo	0,92
RP11-133L14.5	antisense	Exclusivo	0,92
RP11-181K12.2	antisense	Exclusivo	0,91
RP11-80G7.1	antisense	Exclusivo	0,90
AC006460.2	antisense	Exclusivo	0,89
RP3-327A19.5	antisense	Exclusivo	0,89
RP11-307C12.11	antisense	Exclusivo	0,87
RP11-556O5.5	antisense	Exclusivo	0,86
RP11-412D9.4	antisense	Exclusivo	0,85
RP11-804N13.1	antisense	Exclusivo	0,84
RP11-540O11.4	antisense	Exclusivo	0,83

RP11-91K8.1	antisense	Exclusivo	0,80
SMARCA5-AS1	antisense	Exclusivo	0,76
CTC-498J12.3	antisense	Exclusivo	0,76
RP11-177A2.4	antisense	Exclusivo	0,74
RP11-596D21.1	antisense	Exclusivo	0,74
RP3-400N23.6	antisense	Exclusivo	0,74
AC009299.2	antisense	Exclusivo	0,71
HCG15	antisense	Exclusivo	0,70
RP11-492A10.1	antisense	Exclusivo	0,69
RP11-153F1.1	antisense	Exclusivo	0,69
RP11-431J24.2	antisense	Exclusivo	0,69
RP11-797H7.5	antisense	Exclusivo	0,68
RP11-251A15.4	antisense	Exclusivo	0,66
CTC-575N7.1	antisense	Exclusivo	0,65
AC005358.3	antisense	Exclusivo	0,63
U52111.14	antisense	Exclusivo	0,61
RP11-783K16.14	antisense	Exclusivo	0,59
RP11-736E3.1	antisense	Exclusivo	0,59
RP11-20I20.4	antisense	Exclusivo	0,59
CTD-2213F21.2	antisense	Exclusivo	0,57
RP11-465N4.4	antisense	Exclusivo	0,55
RP11-511B23.2	antisense	Exclusivo	0,55
RP11-146F11.1	antisense	Exclusivo	0,55
RP11-283G6.3	antisense	Exclusivo	0,54
AC074138.3	antisense	Exclusivo	0,53
CTD-2154I11.2	antisense	Exclusivo	0,52
CTD-3107M8.4	antisense	Exclusivo	0,51
RP11-560I19.4	antisense	Exclusivo	0,50
RP11-225L12.2	antisense	Exclusivo	0,50
RP11-537H15.3	antisense	Exclusivo	0,50
AC005971.3	antisense	Exclusivo	0,50
AC025335.1	antisense	Exclusivo	0,50
RP11-490O6.2	antisense	Exclusivo	0,48
AC012487.2	antisense	Exclusivo	0,48
RP11-417F21.1	antisense	Exclusivo	0,47
RP11-37C7.3	antisense	Exclusivo	0,47
RP1-39G22.7	antisense	Exclusivo	0,47
RP11-216B9.6	antisense	Exclusivo	0,46
AP000640.2	antisense	Exclusivo	0,46
CTD-2540C19.1	antisense	Exclusivo	0,45
RP11-386G11.10	antisense	Exclusivo	0,45
RP11-98I9.4	antisense	Exclusivo	0,44
RP11-101E7.2	antisense	Exclusivo	0,44

RP11-157P1.5	antisense	Exclusivo	0,43
RP11-526I2.1	antisense	Exclusivo	0,41
CTD-2574D22.2	antisense	Exclusivo	0,41
CTC-338M12.6	antisense	Exclusivo	0,40
RP11-63M22.2	antisense	Exclusivo	0,40
RP11-445H22.3	antisense	Exclusivo	0,40
RP11-877E17.2	antisense	Exclusivo	0,39
RP11-186N15.3	antisense	Exclusivo	0,38
AC004041.2	antisense	Exclusivo	0,37
TMEM106A-AS1	antisense	Exclusivo	0,37
RP11-522I20.3	antisense	Exclusivo	0,36
RP11-736K20.5	antisense	Exclusivo	0,34
AC007038.7	antisense	Exclusivo	0,33
DIAPH2-AS1	antisense	Exclusivo	0,33
RP11-575L7.2	antisense	Exclusivo	0,32
RP11-434I12.2	antisense	Exclusivo	0,31
RP4-639F20.1	antisense	Exclusivo	0,31
RP11-158K1.3	antisense	Exclusivo	0,31
AC009227.2	antisense	Exclusivo	0,30
OSGEPL1-AS1	antisense	Exclusivo	0,30
RP11-742B18.1	antisense	Exclusivo	0,30
RP11-382A18.1	antisense	Exclusivo	0,29
RP11-299G20.2	antisense	Exclusivo	0,28
RP13-638C3.4	antisense	Exclusivo	0,28
AC091729.9	antisense	Exclusivo	0,26
NALCN-AS1	antisense	Exclusivo	0,25
RP11-57H12.3	antisense	Exclusivo	0,24
CTA-445C9.14	antisense	Exclusivo	0,24
RP11-15B17.1	antisense	Exclusivo	0,23
RP11-439E19.10	antisense	Exclusivo	0,21
HOXA11-AS	antisense	Exclusivo	0,20
RP11-50D9.3	antisense	Exclusivo	0,19
RP11-110G21.1	antisense	Exclusivo	0,18
RP11-560G2.1	antisense	Exclusivo	0,17
RP11-849H4.4	antisense	Exclusivo	0,17
RP11-21M24.2	antisense	Exclusivo	0,17
AC009133.14	antisense	Exclusivo	0,16
RP11-196E1.3	antisense	Exclusivo	0,16
RP11-34P13.7	antisense	Exclusivo	0,16
AC009501.4	antisense	Exclusivo	0,16
RP11-378J18.3	antisense	Exclusivo	0,15
FAM13A-AS1	antisense	Exclusivo	0,14
AC011747.7	antisense	Exclusivo	0,13

RP11-643M14.1	antisense	Exclusivo	0,12
RP4-694B14.5	antisense	Exclusivo	0,12
RP4-545C24.1	antisense	Exclusivo	0,11
RP11-265M18.2	antisense	Exclusivo	0,09
CTD-2047H16.4	antisense	Exclusivo	0,09
RP11-438E5.1	antisense	Exclusivo	0,08
FLNB-AS1	antisense	Exclusivo	0,07
RP11-327J17.3	antisense	Exclusivo	0,05
RP11-473M20.14	antisense	Exclusivo	0,05
RP11-164P12.4	antisense	12,01	101,51
CTD-2636A23.2	antisense	11,63	12,93
CTA-204B4.6	antisense	8,35	42,81
RP11-665G4.1	antisense	4,87	31,50
RP11-488L18.10	lincRNA	Exclusivo	20,17
CTD-2023M8.1	lincRNA	Exclusivo	15,10
RP11-549K20.1	lincRNA	Exclusivo	13,76
RP11-543D5.2	lincRNA	Exclusivo	8,37
RP11-530C5.1	lincRNA	Exclusivo	7,60
RP11-553L6.5	lincRNA	Exclusivo	6,52
RP4-724E16.2	lincRNA	Exclusivo	5,43
AC073834.3	lincRNA	Exclusivo	4,25
LINC00340	lincRNA	Exclusivo	3,83
RP11-196I18.4	lincRNA	Exclusivo	3,81
RP11-760D2.1	lincRNA	Exclusivo	3,72
RP4-594I10.2	lincRNA	Exclusivo	3,42
AL592494.4	lincRNA	Exclusivo	3,39
RP11-618G20.4	lincRNA	Exclusivo	3,31
RP11-773D16.1	lincRNA	Exclusivo	3,24
RP11-7F17.4	lincRNA	Exclusivo	3,06
AC116609.3	lincRNA	Exclusivo	2,81
RP11-254F7.2	lincRNA	Exclusivo	2,74
AC090421.1	lincRNA	Exclusivo	2,74
XXbac-BPG248L24.1	lincRNA	Exclusivo	2,68
RP11-344N17.8	lincRNA	Exclusivo	2,66
RP11-139E24.1	lincRNA	Exclusivo	2,60
RP11-661C8.2	lincRNA	Exclusivo	2,57
RP11-469M7.1	lincRNA	Exclusivo	2,43
CTA-313A17.5	lincRNA	Exclusivo	2,41
RP11-774O3.3	lincRNA	Exclusivo	2,26
AC093642.6	lincRNA	Exclusivo	2,23
RP11-277P12.10	lincRNA	Exclusivo	2,21
RP1-180E22.3	lincRNA	Exclusivo	2,16
CTD-2004A9.1	lincRNA	Exclusivo	2,15

LINC00265	lincRNA	Exclusivo	2,11
RP11-317N8.3	lincRNA	Exclusivo	2,08
RP11-342C23.4	lincRNA	Exclusivo	2,00
RP11-666A1.5	lincRNA	Exclusivo	1,97
RP11-420G6.4	lincRNA	Exclusivo	1,96
RP11-638I2.8	lincRNA	Exclusivo	1,87
RP11-315D16.4	lincRNA	Exclusivo	1,81
RP11-195E2.4	lincRNA	Exclusivo	1,77
RP11-396C23.2	lincRNA	Exclusivo	1,77
RP11-314D7.4	lincRNA	Exclusivo	1,76
RP11-202I11.2	lincRNA	Exclusivo	1,68
AC007966.1	lincRNA	Exclusivo	1,67
AC012322.1	lincRNA	Exclusivo	1,67
RP13-942N8.1	lincRNA	Exclusivo	1,67
CTC-552D5.1	lincRNA	Exclusivo	1,66
RP11-83M16.6	lincRNA	Exclusivo	1,65
AC012668.2	lincRNA	Exclusivo	1,61
RP11-766N7.4	lincRNA	Exclusivo	1,61
RP11-461A8.5	lincRNA	Exclusivo	1,58
AC093627.9	lincRNA	Exclusivo	1,57
RP11-701H24.2	lincRNA	Exclusivo	1,56
RP11-30K9.6	lincRNA	Exclusivo	1,54
RP3-471C18.2	lincRNA	Exclusivo	1,54
CTA-342B11.1	lincRNA	Exclusivo	1,52
RP11-457P14.5	lincRNA	Exclusivo	1,52
RP4-735C1.6	lincRNA	Exclusivo	1,51
LINC00159	lincRNA	Exclusivo	1,45
RP11-761N21.1	lincRNA	Exclusivo	1,44
RP11-357D18.1	lincRNA	Exclusivo	1,43
RP11-305L7.3	lincRNA	Exclusivo	1,42
RP4-633O19__A.1	lincRNA	Exclusivo	1,41
RP11-403P17.3	lincRNA	Exclusivo	1,41
AC002310.12	lincRNA	Exclusivo	1,41
CTD-3195I5.5	lincRNA	Exclusivo	1,39
RP11-575H3.1	lincRNA	Exclusivo	1,36
AC073115.7	lincRNA	Exclusivo	1,35
RP11-798K23.4	lincRNA	Exclusivo	1,34
RP4-710M16.2	lincRNA	Exclusivo	1,32
RP11-332J15.1	lincRNA	Exclusivo	1,30
AC073115.6	lincRNA	Exclusivo	1,29
RP11-23E10.3	lincRNA	Exclusivo	1,28
AC017074.2	lincRNA	Exclusivo	1,28
CTD-2016O11.1	lincRNA	Exclusivo	1,28

CTB-1I21.1	lincRNA	Exclusivo	1,27
LINC00476	lincRNA	Exclusivo	1,27
RP11-486O13.4	lincRNA	Exclusivo	1,27
RP11-370I10.2	lincRNA	Exclusivo	1,27
RP11-611L7.1	lincRNA	Exclusivo	1,26
RP11-498P14.5	lincRNA	Exclusivo	1,24
RP11-789C2.1	lincRNA	Exclusivo	1,23
AC062020.1	lincRNA	Exclusivo	1,23
RP11-120D5.1	lincRNA	Exclusivo	1,18
CTD-2091N23.1	lincRNA	Exclusivo	1,15
RP11-65J3.1	lincRNA	Exclusivo	1,14
RP11-114F10.2	lincRNA	Exclusivo	1,14
AC139099.4	lincRNA	Exclusivo	1,12
RP11-66B24.5	lincRNA	Exclusivo	1,12
LA16c-83F12.6	lincRNA	Exclusivo	1,10
RP11-202K23.1	lincRNA	Exclusivo	1,09
RP11-480G7.1	lincRNA	Exclusivo	1,09
RP13-608F4.8	lincRNA	Exclusivo	1,08
LA16c-325D7.2	lincRNA	Exclusivo	1,08
AC005682.8	lincRNA	Exclusivo	1,06
RP11-318M2.2	lincRNA	Exclusivo	1,06
RP5-1024G6.5	lincRNA	Exclusivo	1,06
AC010894.3	lincRNA	Exclusivo	1,02
RP11-624M8.1	lincRNA	Exclusivo	1,01
AC018359.1	lincRNA	Exclusivo	1,01
RP11-347E10.1	lincRNA	Exclusivo	1,00
RP11-445O3.2	lincRNA	Exclusivo	1,00
RP1-102E24.8	lincRNA	Exclusivo	0,99
RP11-1000B6.5	lincRNA	Exclusivo	0,98
RP11-685F15.1	lincRNA	Exclusivo	0,98
RP11-202G18.1	lincRNA	Exclusivo	0,97
RP11-19N8.4	lincRNA	Exclusivo	0,95
RP11-178A10.1	lincRNA	Exclusivo	0,95
RP11-38G5.4	lincRNA	Exclusivo	0,95
RP11-452J13.1	lincRNA	Exclusivo	0,92
RP4-704D21.2	lincRNA	Exclusivo	0,92
RP11-654D12.2	lincRNA	Exclusivo	0,92
RP11-760D2.5	lincRNA	Exclusivo	0,91
RP11-13N12.1	lincRNA	Exclusivo	0,88
RP11-87C12.5	lincRNA	Exclusivo	0,88
AP001505.9	lincRNA	Exclusivo	0,87
AC079305.11	lincRNA	Exclusivo	0,87
AC019117.1	lincRNA	Exclusivo	0,86

AL163953.3	lincRNA	Exclusivo	0,86
RP11-626G11.4	lincRNA	Exclusivo	0,85
RP11-151N17.2	lincRNA	Exclusivo	0,83
RP11-276H19.1	lincRNA	Exclusivo	0,82
RBM26-AS1	lincRNA	Exclusivo	0,82
RP11-115C21.2	lincRNA	Exclusivo	0,82
AC007255.8	lincRNA	Exclusivo	0,81
AC092614.2	lincRNA	Exclusivo	0,79
RP11-819C21.1	lincRNA	Exclusivo	0,78
RP5-936J12.1	lincRNA	Exclusivo	0,77
RP11-348I14.4	lincRNA	Exclusivo	0,76
RP11-538P18.2	lincRNA	Exclusivo	0,76
RP11-31E23.1	lincRNA	Exclusivo	0,76
AC195454.1	lincRNA	Exclusivo	0,75
RP11-368L12.1	lincRNA	Exclusivo	0,74
AC091814.2	lincRNA	Exclusivo	0,73
LINC00116	lincRNA	Exclusivo	0,72
LINC00493	lincRNA	Exclusivo	0,71
RP11-87H9.3	lincRNA	Exclusivo	0,71
CTD-2291D10.2	lincRNA	Exclusivo	0,70
RP11-363G15.2	lincRNA	Exclusivo	0,69
RP11-350F4.2	lincRNA	Exclusivo	0,68
RP11-645N11.2	lincRNA	Exclusivo	0,67
RP11-589B3.6	lincRNA	Exclusivo	0,67
ZNF718	lincRNA	Exclusivo	0,66
RP11-252P19.1	lincRNA	Exclusivo	0,64
AP000696.2	lincRNA	Exclusivo	0,63
RP4-794H19.2	lincRNA	Exclusivo	0,63
AC092835.2	lincRNA	Exclusivo	0,63
AC073218.2	lincRNA	Exclusivo	0,62
CTD-2354A18.1	lincRNA	Exclusivo	0,62
CTC-504A5.1	lincRNA	Exclusivo	0,61
AC108488.3	lincRNA	Exclusivo	0,60
RP11-279N8.3	lincRNA	Exclusivo	0,58
RP11-707G14.6	lincRNA	Exclusivo	0,58
RP5-1172A22.1	lincRNA	Exclusivo	0,57
AC074391.1	lincRNA	Exclusivo	0,57
LIFR-AS1	lincRNA	Exclusivo	0,57
RP11-111F5.4	lincRNA	Exclusivo	0,57
RP11-83B20.1	lincRNA	Exclusivo	0,56
RP11-508N12.2	lincRNA	Exclusivo	0,55
RP11-122C21.1	lincRNA	Exclusivo	0,55
PART1	lincRNA	Exclusivo	0,54

RP11-175K6.1	lincRNA	Exclusivo	0,54
RP11-442N24__B.1	lincRNA	Exclusivo	0,53
RP11-29H23.1	lincRNA	Exclusivo	0,52
AC087590.3	lincRNA	Exclusivo	0,52
RP11-181D18.2	lincRNA	Exclusivo	0,51
RP11-305P14.1	lincRNA	Exclusivo	0,50
CTD-2126E3.1	lincRNA	Exclusivo	0,47
RP11-304M2.2	lincRNA	Exclusivo	0,47
RP11-371I1.2	lincRNA	Exclusivo	0,47
U82695.10	lincRNA	Exclusivo	0,46
CTD-2001C12.1	lincRNA	Exclusivo	0,46
RP11-513I15.6	lincRNA	Exclusivo	0,46
CROCCP2	lincRNA	Exclusivo	0,45
RP11-327L3.1	lincRNA	Exclusivo	0,45
RP11-108O10.2	lincRNA	Exclusivo	0,44
RP1-274L7.1	lincRNA	Exclusivo	0,44
RP11-202P11.1	lincRNA	Exclusivo	0,44
AC092937.2	lincRNA	Exclusivo	0,42
AC007392.3	lincRNA	Exclusivo	0,42
LINC00256A	lincRNA	Exclusivo	0,41
LINC00571	lincRNA	Exclusivo	0,41
CTD-2291D10.1	lincRNA	Exclusivo	0,39
RP11-568N6.1	lincRNA	Exclusivo	0,38
RP11-557H15.3	lincRNA	Exclusivo	0,37
RP11-509E16.1	lincRNA	Exclusivo	0,37
RP11-138I17.1	lincRNA	Exclusivo	0,36
RP11-420A23.1	lincRNA	Exclusivo	0,36
LINC00230A	lincRNA	Exclusivo	0,35
LINC00230B	lincRNA	Exclusivo	0,35
RP11-413E6.8	lincRNA	Exclusivo	0,34
RP11-706O15.3	lincRNA	Exclusivo	0,33
RP3-326I13.1	lincRNA	Exclusivo	0,32
AP000295.7	lincRNA	Exclusivo	0,32
RP5-1070A16.1	lincRNA	Exclusivo	0,32
CTC-428G20.3	lincRNA	Exclusivo	0,32
RP11-213G2.2	lincRNA	Exclusivo	0,32
AC105760.2	lincRNA	Exclusivo	0,31
AC018816.4	lincRNA	Exclusivo	0,31
RP11-66D17.3	lincRNA	Exclusivo	0,30
RP5-1030M6.2	lincRNA	Exclusivo	0,30
AC002480.3	lincRNA	Exclusivo	0,30
RP11-551L14.4	lincRNA	Exclusivo	0,29
RP11-204J18.2	lincRNA	Exclusivo	0,28

RP11-774D14.1	lincRNA	Exclusivo	0,28
RP11-398F12.1	lincRNA	Exclusivo	0,28
TMEM191C	lincRNA	Exclusivo	0,28
RP11-213H15.3	lincRNA	Exclusivo	0,27
AC046143.7	lincRNA	Exclusivo	0,27
RP11-250B2.3	lincRNA	Exclusivo	0,26
RP11-867G23.3	lincRNA	Exclusivo	0,26
RP1-78O14.1	lincRNA	Exclusivo	0,25
RP5-1186N24.3	lincRNA	Exclusivo	0,25
RP11-179A16.1	lincRNA	Exclusivo	0,25
RP11-497D6.4	lincRNA	Exclusivo	0,25
RP5-821D11.7	lincRNA	Exclusivo	0,24
CTA-445C9.15	lincRNA	Exclusivo	0,24
RP11-317N12.1	lincRNA	Exclusivo	0,23
CTD-2001E22.2	lincRNA	Exclusivo	0,23
RP3-341D10.4	lincRNA	Exclusivo	0,22
RP4-781K5.8	lincRNA	Exclusivo	0,20
LINC00339	lincRNA	Exclusivo	0,20
CECR7	lincRNA	Exclusivo	0,18
AC091814.3	lincRNA	Exclusivo	0,18
RP11-989E6.10	lincRNA	Exclusivo	0,18
RP11-421L10.1	lincRNA	Exclusivo	0,17
AC007359.6	lincRNA	Exclusivo	0,17
RP11-206L10.3	lincRNA	Exclusivo	0,15
AC078941.1	lincRNA	Exclusivo	0,14
RP11-887P2.1	lincRNA	Exclusivo	0,14
RP11-430B1.2	lincRNA	Exclusivo	0,14
CTB-43E15.3	lincRNA	Exclusivo	0,13
DICER1-AS1	lincRNA	Exclusivo	0,13
RP11-557H15.2	lincRNA	Exclusivo	0,12
RP11-620J15.2	lincRNA	Exclusivo	0,11
RP1-80B9.2	lincRNA	Exclusivo	0,09
LINC00277	lincRNA	Exclusivo	0,08
AF064858.6	lincRNA	Exclusivo	0,06
TUG1	lincRNA	13,34	1,50
H19	lincRNA	9,95	2,28
MCM3AP-AS1	lincRNA	5,86	5,94
OIP5-AS1	lincRNA	5,12	24,00
RP4-550H1.6	lincRNA	4,16	94,48
RP11-197P3.4	lincRNA	4,04	14,45
MIR568	miRNA	Exclusivo	22,69
MIR1204	miRNA	Exclusivo	18,69
AC098650.1	miRNA	Exclusivo	14,73

MIRLET7A2	miRNA	Exclusivo	9,84
MIR296	miRNA	Exclusivo	8,98
MIR100	miRNA	Exclusivo	8,98
AC097455.1	miRNA	Exclusivo	8,26
MIR298	miRNA	Exclusivo	8,17
AL023096.1	miRNA	Exclusivo	7,64
AC068704.2	miRNA	Exclusivo	7,56
AP005264.1	miRNA	Exclusivo	6,47
AL441988.1	miRNA	Exclusivo	5,17
RP11-1029J19.4	Non-coding	Exclusivo	1,33
RP11-530C5.2	Non-coding	Exclusivo	1,08
LINC00205	Non-coding	Exclusivo	1,03
MIR17HG	Non-coding	Exclusivo	0,77
GBA3	polymorphic-pseudogene	Exclusivo	0,61
SEMA3B	polymorphic-pseudogene	Exclusivo	0,09
RP11-143P4.2	processed-transcript	Exclusivo	7,85
AL591893.1	processed-transcript	Exclusivo	5,33
CTD-2503O16.4	processed-transcript	Exclusivo	4,88
SCAMP1	processed-transcript	Exclusivo	4,52
STEAP2-AS1	processed-transcript	Exclusivo	4,00
RP11-589M4.3	processed-transcript	Exclusivo	3,67
RP11-290L1.5	processed-transcript	Exclusivo	3,50
RP11-494M8.4	processed-transcript	Exclusivo	2,86
RP11-91P24.7	processed-transcript	Exclusivo	2,46
AC012501.2	processed-transcript	Exclusivo	2,45
AC008592.5	processed-transcript	Exclusivo	2,16
RP11-499E18.1	processed-transcript	Exclusivo	2,14
CTD-2066L21.3	processed-transcript	Exclusivo	2,02
RP11-79P5.6	processed-transcript	Exclusivo	1,77
RP11-956E11.1	processed-transcript	Exclusivo	1,70
RP11-256L6.3	processed-transcript	Exclusivo	1,67
HIST2H2BA	processed-transcript	Exclusivo	1,67
RP11-563D10.1	processed-transcript	Exclusivo	1,51
RP11-57C13.6	processed-transcript	Exclusivo	1,45
RP1-90L14.1	processed-transcript	Exclusivo	1,43
RP11-611E13.2	processed-transcript	Exclusivo	1,42
CTD-2507G9.1	processed-transcript	Exclusivo	1,41
RP5-905H7.9	processed-transcript	Exclusivo	1,39
RP11-753H16.5	processed-transcript	Exclusivo	1,32
RP11-167H9.4	processed-transcript	Exclusivo	1,29
RP11-309L24.2	processed-transcript	Exclusivo	1,29
RP11-423H2.2	processed-transcript	Exclusivo	1,27
RP11-508M1.5	processed-transcript	Exclusivo	1,26

UBXN8	processed-transcript	Exclusivo	1,21
RP11-101P17.6	processed-transcript	Exclusivo	1,16
RP11-101C11.1	processed-transcript	Exclusivo	1,13
RP4-798P15.2	processed-transcript	Exclusivo	1,09
AC124789.1	processed-transcript	Exclusivo	1,08
LINC00443	processed-transcript	Exclusivo	1,07
AC022007.5	processed-transcript	Exclusivo	1,04
LINC00322	processed-transcript	Exclusivo	1,03
RP4-669L17.11	processed-transcript	Exclusivo	1,02
RP11-845M18.3	processed-transcript	Exclusivo	1,00
AC073043.1	processed-transcript	Exclusivo	0,95
RP11-323I15.5	processed-transcript	Exclusivo	0,93
RP11-3L8.3	processed-transcript	Exclusivo	0,90
RP4-735C1.4	processed-transcript	Exclusivo	0,89
RP11-104H15.7	processed-transcript	Exclusivo	0,80
RP11-85K15.2	processed-transcript	Exclusivo	0,76
AC005281.1	processed-transcript	Exclusivo	0,66
XX-CR54.3	processed-transcript	Exclusivo	0,65
PCBP1-AS1	processed-transcript	Exclusivo	0,59
RP1-240B8.3	processed-transcript	Exclusivo	0,54
RP11-457M11.2	processed-transcript	Exclusivo	0,52
RP11-529K1.2	processed-transcript	Exclusivo	0,49
RP11-108L7.7	processed-transcript	Exclusivo	0,47
AC018642.1	processed-transcript	Exclusivo	0,47
SNHG10	processed-transcript	Exclusivo	0,47
RP11-383C5.4	processed-transcript	Exclusivo	0,45
RP1-117O3.2	processed-transcript	Exclusivo	0,43
RP3-416H24.1	processed-transcript	Exclusivo	0,42
AC103702.1	processed-transcript	Exclusivo	0,35
RP11-100L22.2	processed-transcript	Exclusivo	0,33
AC137934.1	processed-transcript	Exclusivo	0,32
RP11-793H13.3	processed-transcript	Exclusivo	0,31
RP11-399D6.2	processed-transcript	Exclusivo	0,31
RP11-600K15.1	processed-transcript	Exclusivo	0,30
RP11-403I13.4	processed-transcript	Exclusivo	0,28
RP1-90J20.7	processed-transcript	Exclusivo	0,26
RP11-758M4.4	processed-transcript	Exclusivo	0,26
AC017099.3	processed-transcript	Exclusivo	0,19
RP11-974F13.6	processed-transcript	Exclusivo	0,19
RP11-477J21.6	processed-transcript	Exclusivo	0,13
MAGI2-AS3	processed-transcript	Exclusivo	0,09
RP11-739N20.2	processed-transcript	18,31	96,78
RP5-894A10.2	processed-transcript	13,48	5,09

CTB-104F4.2	processed-transcript	9,44	32,94
RP5-1061H20.4	processed-transcript	8,74	30,63
RP11-384K6.6	processed-transcript	6,39	24,82
RP11-659G9.3	processed-transcript	6,09	164,53
DCP1A	processed-transcript	4,70	26,35
RP11-611O2.5	processed-transcript	4,40	14,41
AC091304.3	pseudogene	Exclusivo	12,32
RP11-152C15.1	pseudogene	Exclusivo	11,14
CHCHD2P6	pseudogene	Exclusivo	10,89
RP11-2F9.3	pseudogene	Exclusivo	7,49
PTP4A1P7	pseudogene	Exclusivo	6,18
BNIP3P1	pseudogene	Exclusivo	4,99
RP1-96H9.5	pseudogene	Exclusivo	4,69
NPM1P12	pseudogene	Exclusivo	4,32
C10orf102	pseudogene	Exclusivo	3,60
AC140481.7	pseudogene	Exclusivo	3,44
AC093642.5	pseudogene	Exclusivo	3,34
PIPSL	pseudogene	Exclusivo	3,14
CIDECP	pseudogene	Exclusivo	3,02
AC016734.2	pseudogene	Exclusivo	3,00
FAM86EP	pseudogene	Exclusivo	2,89
OR7E14P	pseudogene	Exclusivo	2,70
SDHAP3	pseudogene	Exclusivo	2,68
CEP57L1P1	pseudogene	Exclusivo	2,61
AC126544.2	pseudogene	Exclusivo	2,47
MYL8P	pseudogene	Exclusivo	2,44
RP11-445P17.4	pseudogene	Exclusivo	2,42
RP13-98N21.2	pseudogene	Exclusivo	2,19
RP13-608F4.1	pseudogene	Exclusivo	2,14
AC097523.2	pseudogene	Exclusivo	2,06
RP11-863K10.4	pseudogene	Exclusivo	1,87
RP11-497D6.2	pseudogene	Exclusivo	1,83
FABP5P1	pseudogene	Exclusivo	1,76
RP11-152F13.7	pseudogene	Exclusivo	1,59
AC113554.1	pseudogene	Exclusivo	1,46
KRT8P12	pseudogene	Exclusivo	1,44
GBAP1	pseudogene	Exclusivo	1,43
HSD17B7P2	pseudogene	Exclusivo	1,41
TMED11P	pseudogene	Exclusivo	1,35
AC002519.6	pseudogene	Exclusivo	1,30
AP000356.1	pseudogene	Exclusivo	1,30
RP11-440K22.1	pseudogene	Exclusivo	1,28
SDHDP1	pseudogene	Exclusivo	1,26

PPP1R2P3	pseudogene	Exclusivo	1,26
AC140481.1	pseudogene	Exclusivo	1,21
CYP4F43P	pseudogene	Exclusivo	1,17
CTD-3092A11.1	pseudogene	Exclusivo	1,16
INTS4L2	pseudogene	Exclusivo	1,03
HSP90AA4P	pseudogene	Exclusivo	1,00
RP11-297L17.6	pseudogene	Exclusivo	0,96
PIP5K1P1	pseudogene	Exclusivo	0,91
CYP4F30P	pseudogene	Exclusivo	0,88
FAM86FP	pseudogene	Exclusivo	0,87
TAB3P1	pseudogene	Exclusivo	0,84
RP11-543B16.2	pseudogene	Exclusivo	0,80
ABCC6P1	pseudogene	Exclusivo	0,79
ANKRD19P	pseudogene	Exclusivo	0,78
ALMS1P	pseudogene	Exclusivo	0,78
RP11-3N2.13	pseudogene	Exclusivo	0,75
BMPR1APS2	pseudogene	Exclusivo	0,71
RP11-480I12.5	pseudogene	Exclusivo	0,68
AC027612.1	pseudogene	Exclusivo	0,68
KIAA0664L3	pseudogene	Exclusivo	0,66
LRRC37A6P	pseudogene	Exclusivo	0,60
CNOT6LP1	pseudogene	Exclusivo	0,50
GUSBP2	pseudogene	Exclusivo	0,49
RP1-34B21.6	pseudogene	Exclusivo	0,47
ZNF192P1	pseudogene	Exclusivo	0,44
OR2A9P	pseudogene	Exclusivo	0,43
RP11-152F13.2	pseudogene	Exclusivo	0,37
EPS15P1	pseudogene	Exclusivo	0,35
SLC6A10P	pseudogene	Exclusivo	0,33
HTR7P1	pseudogene	Exclusivo	0,33
GGT3P	pseudogene	Exclusivo	0,31
GUSBP11	pseudogene	Exclusivo	0,29
ZNF876P	pseudogene	Exclusivo	0,28
AACSP1	pseudogene	Exclusivo	0,25
WBP11P1	pseudogene	Exclusivo	0,24
RP11-642P15.1	pseudogene	Exclusivo	0,24
AMZ2P1	pseudogene	Exclusivo	0,21
RPS3AP47	pseudogene	9,92	135,34
RPLP0P2	pseudogene	5,02	18,96

Apêndice 2 – Lista dos 90 RNAs longos não codificadores de proteínas e dos micro-RNAs detectados nos sequenciamentos RIP-Anti-Ago2 em HeLa e que são preditos como pares lncRNA-miRNA.

Símbolo do lncRNA	Classe (anotação do GENCODE)	Micro-RNAs detectados nos sequenciamentos e com predição de alvejar o lncRNA
HCG14	Antisense	miR-551a
AC074391.1	lincRNA	miR-363-3p;miR-153
LINC00476	lincRNA	miR-1244;miR-20b-5p;miR-590-5p;miR-219-5p;miR-876-3p; miR-372; miR-140-5p;miR-193a-3p;miR-152;miR-455-5p; miR-107;miR-301b
PART1	lincRNA	miR-4770;miR-363-3p;miR-142-3p;miR-449c-5p;miR-590-5p; miR-22-3p;miR-219-5p;miR-152;miR-301b
CROCCP2	lincRNA	miR-363-3p;miR-449c-5p;miR-20b-5p;miR-22-3p;miR-217; miR-551a;miR-372;miR-193a-3p;miR-455-5p;miR-107
TMEM191C	lincRNA	miR-363-3p;miR-1244;miR-22-3p;miR-876-3p;miR-217; miR-140-5p;miR-107
COX10-AS1	lincRNA	miR-363-3p;miR-142-3p;miR-1244;miR-20b-5p;miR-22-3p; miR-219-5p;miR-876-3p;miR-217;miR-372;miR-140-5p
PLK1S1	lincRNA	miR-142-3p;miR-449c-5p;miR-1244;miR-22-3p;miR-219-5p; miR-876-3p;miR-217;miR-215;miR-375;miR-140-5p
ENTPD3-AS1	lincRNA	miR-142-3p;miR-20b-5p;miR-590-5p;miR-372
SEC22B	lincRNA	miR-363-3p;miR-142-3p;miR-375;miR-193a-3p;miR-33a-3p; miR-301b
MAL2	lincRNA	miR-363-3p;miR-137;miR-153;miR-1244;miR-20b-5p;miR-22-3p; miR-876-3p;miR-375;miR-140-5p;miR-33a-3p;miR-152;miR-301b
LINC00467	lincRNA	miR-107
ZNF518A	lincRNA	miR-137;miR-142-3p;miR-153;miR-20b-5p;miR-590-5p;miR-22-3p; miR-217;miR-375;miR-455-5p;miR-301b;miR-4770;miR-219-5p; miR-372;miR-107
NBPF1	lincRNA	miR-4770;miR-137;miR-142-3p;miR-1244;miR-22-3p;miR-876-3p; miR-375;miR-140-5p;miR-193a-3p;miR-152;miR-455-5p;miR-107
LINC00152	lincRNA	miR-4770;miR-153;miR-193a-3p;miR-107
TUG1	lincRNA	miR-132;miR-137;miR-138;miR-140;miR-141;miR-142-3p;miR-143; miR-148b;miR-153;miR-190;miR-193;miR-216a;miR-219;miR-22; miR-23b-3p;miR-31;miR-103a;miR-33a-3p
MIR100	miRNA	miR-210
AL441988.1	miRNA	miR-153;miR-363-3p
LINC00205	non-coding	miR-153
CASC2	non-coding	miR-363-3p;miR-153;miR-1244;miR-20b-5p;miR-590-5p; miR-876-3p; miR-217;miR-215;miR-193a-3p;miR-152;miR-301b; miR-4770;miR-449c-5p;miR-372;miR-140-5p;miR-107
SCAMP1	processed-transcript	miR-4770;miR-363-3p;miR-137;miR-142-3p;miR-153;miR-20b-5p; miR-210;miR-22-3p;miR-217;miR-551a;miR-215;miR-372; miR-33a-3p; miR-107
UBXN8	processed-transcript	miR-217;miR-193a-3p;miR-590-5p
PCBP1-AS1	processed-transcript	miR-363-3p;miR-142-3p;miR-1244;miR-20b-5p;miR-590-5p; miR-22-3p; miR-876-3p;miR-217;miR-215;miR-375;miR-193a-3p; miR-455-5p;miR-4770;miR-449c-5p;miR-219-5p;miR-372; miR-140-5p; miR-33a-3p;miR-107
POLDIP2	processed-transcript	miR-137;miR-20b-5p;miR-193a-3p;miR-152;miR-455-5p;miR-107; miR-22-3p
VPS11	processed-transcript	miR-137;miR-153;miR-590-5p;miR-22-3p;miR-219-5p;miR-217; miR-215; miR-375;miR-372;miR-193a-3p;miR-455-5p;miR-301b

C14orf167	processed-transcript	miR-4770;miR-449c-5p;miR-153;miR-1244;miR-22-3p;miR-876-3p; miR-217;miR-140-5p;miR-193a-3p;miR-152;miR-455-5p;miR-301b
LRRC37BP1	processed-transcript	miR-4770;miR-153;miR-20b-5p;miR-590-5p;miR-22-3p;miR-219-5p; miR-217;miR-375;miR-372;miR-193a-3p;miR-33a-3p;miR-152; miR-107
CWC15	processed-transcript	miR-4770;miR-363-3p;miR-449c-5p;miR-20b-5p;miR-22-3p; miR-215; miR-375;miR-193a-3p;miR-107;miR-301b
ALDH3B1	processed-transcript	miR-4770;miR-449c-5p;miR-153;miR-20b-5p;miR-22-3p;miR-372; miR-33a-3p;miR-152;miR-107
SLC37A4	processed-transcript	miR-4770;miR-142-3p;miR-449c-5p;miR-153;miR-1244;miR-22-3p; miR-876-3p;miR-140-5p;miR-193a-3p
RECQL4	processed-transcript	miR-142-3p;miR-449c-5p;miR-193a-3p;miR-210;miR-152;miR-107; miR-22-3p
PRIM2	processed-transcript	miR-20b-5p;miR-217;miR-375;miR-372;miR-33a-3p;miR-152; miR-107; miR-301b
PPP1R9B	processed-transcript	miR-4770;miR-449c-5p;miR-20b-5p;miR-215;miR-372;miR-152
BNIP3P1	pseudogene	miR-137;miR-20b-5p;miR-210;miR-107;miR-301b
CIDECP	pseudogene	miR-363-3p;miR-137;miR-20b-5p;miR-590-5p;miR-22-3p; miR-372; miR-152
OR7E14P	pseudogene	miR-4770
SDHAP3	pseudogene	miR-152;miR-107
GBAP1	pseudogene	miR-219-5p;miR-217;miR-153;miR-33a-3p;miR-152;miR-22-3p
HSD17B7P2	pseudogene	miR-4770;miR-20b-5p;miR-219-5p;miR-372
TMED11P	pseudogene	miR-375
HSP90AA4P	pseudogene	miR-215;miR-375;miR-193a-3p;miR-455-5p
PIP5K1P1	pseudogene	miR-142-3p;miR-372;miR-33a-3p;miR-107
CYP4F30P	pseudogene	miR-363-3p;miR-372
FAM86FP	pseudogene	miR-4770;miR-449c-5p;miR-372;miR-107
ALMS1P	pseudogene	miR-4770;miR-193a-3p;miR-210
KIAA0664L3	pseudogene	miR-4770;miR-142-3p;miR-449c-5p;miR-210;miR-22-3p;miR-215; miR-372
SLC6A10P	pseudogene	miR-363-3p;miR-137;miR-142-3p;miR-449c-5p;miR-1244; miR-22-3p; miR-876-3p;miR-375;miR-372;miR-140-5p;miR-193a-3p; miR-152
GGT3P	pseudogene	miR-219-5p;miR-363-3p;miR-137;miR-217;miR-33a-3p;miR-152; miR-22-3p
AMZ2P1	pseudogene	miR-4770;miR-363-3p;miR-142-3p;miR-449c-5p;miR-153;miR-215; miR-455-5p;miR-301b
RPS3AP47	pseudogene	miR-219-5p
CA5BP1	pseudogene	miR-876-3p;miR-137;miR-551a;miR-1244;miR-372;miR-140-5p; miR-193a-3p;miR-152;miR-107
PIPSL	pseudogene	miR-142-3p;miR-153;miR-1244;miR-22-3p;miR-219-5p;miR-876-3p; miR-140-5p;miR-33a-3p;miR-107
TMEM198B	pseudogene	miR-4770;miR-449c-5p;miR-22-3p;miR-219-5p;miR-372;miR-33a-3p
ZNF833P	pseudogene	miR-4770;miR-363-3p;miR-153;miR-219-5p;miR-217;miR-33a-3p; miR-107
DPY19L2P2	pseudogene	miR-363-3p;miR-137;miR-153;miR-1244;miR-590-5p;miR-22-3p; miR-219-5p;miR-876-3p;miR-217;miR-375;miR-140-5p; miR-193a-3p; miR-152;miR-107;miR-301b
MSL3P1	pseudogene	miR-372;miR-20b-5p;miR-22-3p
BTN2A3P	pseudogene	miR-4770;miR-363-3p;miR-142-3p;miR-449c-5p;miR-20b-5p;

		miR-372; miR-33a-3p;miR-152;miR-455-5p;miR-107;miR-301b
ATP8B5P	pseudogene	miR-363-3p;miR-153;miR-1244;miR-20b-5p;miR-876-3p;miR-217; miR-375;miR-4770;miR-449c-5p;miR-372;miR-140-5p;miR-107
PMS2CL	pseudogene	miR-449c-5p;miR-1244;miR-20b-5p;miR-590-5p;miR-876-3p; miR-372; miR-140-5p;miR-33a-3p;miR-107
SRSF8	pseudogene	miR-4770;miR-142-3p;miR-20b-5p;miR-22-3p;miR-217;miR-375; miR-372;miR-301b
AFG3L1P	pseudogene	miR-137;miR-1244;miR-20b-5p;miR-590-5p;miR-876-3p;miR-215; miR-375;miR-193a-3p;miR-152;miR-301b;miR-449c-5p;miR-210; miR-372;miR-140-5p;miR-33a-3p;miR-107
RP9P	pseudogene	miR-217;miR-372;miR-20b-5p
STAG3L4	pseudogene	miR-4770;miR-1244;miR-22-3p;miR-876-3p;miR-140-5p; miR-193a-3p; miR-455-5p;miR-107
NSUN5P2	pseudogene	miR-137;miR-449c-5p;miR-551a;miR-210;miR-455-5p;miR-107
TTC3P1	pseudogene	miR-137;miR-142-3p;miR-590-5p;miR-22-3p;miR-217;miR-375; miR-372; miR-193a-3p;miR-152;miR-107;miR-301b
EGFEM1P	pseudogene	miR-4770;miR-153;miR-20b-5p;miR-217;miR-372;miR-193a-3p; miR-33a-3p;miR-301b
TUBBP1	pseudogene	miR-142-3p;miR-449c-5p;miR-455-5p;miR-22-3p
CRYBB2P1	pseudogene	miR-153;miR-20b-5p;miR-590-5p;miR-22-3p;miR-217;miR-215; miR-372;miR-33a-3p;miR-455-5p;miR-107;miR-301b
RCN1P2	pseudogene	miR-449c-5p;miR-107
HLA-P	pseudogene	miR-137;miR-20b-5p;miR-590-5p;miR-215;miR-301b;miR-4770; miR-219-5p;miR-372;miR-33a-3p
SRGAP2P2	pseudogene	miR-215
PLEKHA8P1	pseudogene	miR-4770;miR-215;miR-372;miR-107
RPL31P11	pseudogene	miR-219-5p;miR-590-5p;miR-33a-3p;miR-107
PI4KAP2	pseudogene	miR-142-3p;miR-153;miR-20b-5p;miR-22-3p;miR-217;miR-215; miR-193a-3p; miR-152;miR-455-5p;miR-301b;miR-4770; miR-449c-5p; miR-210;miR-219-5p;miR-372;miR-107
FKBP9L	pseudogene	miR-4770;miR-142-3p;miR-449c-5p;miR-1244;miR-210;miR-876-3p; miR-372;miR-140-5p;miR-193a-3p;miR-152;miR-107
HNRNPA3P3	pseudogene	miR-217
COX5BP2	pseudogene	miR-22-3p
SDHAP1	pseudogene	miR-449c-5p;miR-20b-5p;miR-217;miR-375;miR-152; miR-455-5p;miR-107
HERC2P3	pseudogene	miR-142-3p;miR-449c-5p;miR-1244;miR-876-3p;miR-215;miR-375; miR-140-5p;miR-193a-3p;miR-33a-3p;miR-152;miR-107
SCAND2	pseudogene	miR-4770;miR-137;miR-449c-5p;miR-20b-5p;miR-590-5p;miR-217; miR-215;miR-372;miR-33a-3p;miR-152;miR-107
HERC2P9	pseudogene	miR-142-3p;miR-449c-5p;miR-153;miR-1244;miR-20b-5p; miR-219-5p; miR-876-3p;miR-217;miR-215;miR-375;miR-140-5p; miR-193a-3p;miR-33a-3p;miR-152;miR-107
UBE2Q2P1	pseudogene	miR-876-3p;miR-1244;miR-140-5p;miR-455-5p;miR-107;miR-22-3p
SEPT7P2	pseudogene	miR-4770;miR-22-3p;miR-219-5p;miR-217;miR-215;miR-375; miR-152; miR-455-5p
ZNF767	pseudogene	miR-4770;miR-449c-5p;miR-153;miR-1244;miR-210;miR-876-3p; miR-215;miR-372;miR-140-5p;miR-455-5p;miR-107
SRP72P2	pseudogene	miR-219-5p;miR-137;miR-375;miR-33a-3p;miR-152
ANKRD36B	pseudogene	miR-137;miR-153;miR-1244;miR-20b-5p;miR-590-5p;miR-876-3p; miR-215;miR-375;miR-193a-3p;miR-301b;miR-4770;miR-449c-5p; miR-219-5p;miR-372;miR-140-5p;miR-107

WASH3P	pseudogene	miR-137;miR-217;miR-193a-3p;miR-33a-3p;miR-455-5p;miR-22-3p
HMGB1P5	pseudogene	miR-142-3p;miR-107
HERC2P2	pseudogene	miR-142-3p;miR-1244;miR-20b-5p;miR-876-3p;miR-215;miR-375; miR-193a-3p;miR-152;miR-4770;miR-449c-5p;miR-210;miR-219-5p; miR-372;miR-140-5p;miR-33a-3p;miR-107
FTH1P10	pseudogene	miR-215

ANEXOS ELETRÔNICOS

1. **Artigo Publicado 1:** BECKEDORFF, FELIPE C. ; AYUPE, ANA C. ; CROCCI-SOUZA, RENAN ; **AMARAL, MURILO S**. ; NAKAYA, HELDER I. ; SOLTYS, DANIELA T. ; MENCK, CARLOS F. M. ; REIS, EDUARDO M. ; VERJOVSKI-ALMEIDA, SERGIO . The Intronic Long Noncoding RNA *ANRASSF1* Recruits PRC2 to the *RASSF1A* Promoter, Reducing the Expression of RASSF1A and Increasing Cell Proliferation. PLOS Genetics (Online), v. 9, p. e1003705, 2013.

2. **Artigo Publicado 2:** BECKEDORFF, F. C.*, **AMARAL, M. S.***, DEOCESANO‐PEREIRA, C. AND VERJOVSKI‐ALMEIDA, S. (2013) Long non-coding RNAs and their implications in cancer epigenetics. Biosci. Rep. 33(4):art:e00061.

3. **Artigo Publicado 3:** ALMEIDA, GIULLIANA TESSARIN ; **AMARAL, MURILO SENA** ; BECKEDORFF, FELIPE CESAR FERRAREZI ; KITAJIMA, JOÃO PAULO ; DEMARCO, RICARDO ; VERJOVSKI-ALMEIDA, SERGIO . Exploring the *Schistosoma mansoni* adult male transcriptome using RNA-seq. Experimental Parasitology, v. 1, p. 01, 2011.

4. **Arquivo 1:** Lista dos 812 miRNAs humanos conhecidos detectados pelo algoritmo miRDeep2 a partir de imunoprecipitações do complexo RISC em células HeLa.

5. **Arquivo 2:** Lista dos possíveis novos 336 miRNAs humanos preditos pelo algoritmo miRDeep2.

OBSERVAÇÃO: Todos os anexos encontram-se em arquivos no CD-ROM anexo a esta tese.

ANEXO

SÚMULA CURRICULAR

Dados Pessoais:

Nome: **Murilo Sena Amaral**
Nascimento: 28 de julho de 1985; Naturalidade: Almenara- MG; Nacionalidade: Brasileiro
Endereço: Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes, 748 sala 1200 CEP: 05508-000 - São Paulo, SP.
Telefone: trabalho (011) 3091-2173 R209 E-mail: sena@iq.usp.br

1) Formação:

1. Farmacêutico-Bioquímico pela Universidade Federal de Ouro Preto – 2003-2007
2. Estudande de Doutorado em Ciências Biológicas- Bioquímica, Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo

Orientação Dr. Sergio Verjovski-Almeida – 2008-atual.

2) Histórico profissional:

1. Aluno de Iniciação Cientifica

Local: Departamento de Ciências Biológicas - Instituto de Ciências Exatas e Biológicas - Universidade Federal de Ouro Preto, Ouro Preto, MG
Projeto: Clonagem e expressão heteróloga das apirases de *Leishmania sp.*
Orientadora: Dra. Juliana Lopes Rangel Fietto.
Bolsista PET/CAPES - Vigência da Bolsa: 01/01/04 a 31/04/07

2. Analista de Laboratório de Controle de Qualidade de Vinhos

Local: Hartford Family Winery, Forestville, California, EUA
Principais atividades: Análises laboratoriais de amostras (uvas, sucos, vinhos) durante o processo de produção suco-vinho, incluindo porcentagem de sólidos e perfil enzimático, entre outras.
Supervisor: Damon Harvey
Período: 01/08/07 a 31/12/07

3. Aluno de Doutorado Direto

Local: Departamento de Bioquímica - Instituto de Quimíca – USP, São Paulo –SP.
Projeto: Caracterização da interação entre RNAs não codificadores de proteínas (ncRNAs) e proteínas com domínios de ligação a RNA
Orientador: Prof. Dr. Sergio Verjovski Almeida.
Bolsista CNPq: Vigência da Bolsa: 01/05/08 a 31/04/13
Provável término em 12/2013.

3) Resultados de pesquisa mais relevantes

Artigos completos publicados em periódicos

Beckedorff, Felipe C. ; Ayupe, Ana C. ; Crocci-Souza, Renan ; **Amaral, Murilo S**. ; Nakaya, Helder I. ; Soltys, Daniela T. ; Menck, Carlos F. M. ; Reis, Eduardo M. ; Verjovski-Almeida, Sergio . The Intronic Long Noncoding RNA *ANRASSF1* Recruits PRC2 to the *RASSF1A* Promoter, Reducing the Expression of RASSF1A and Increasing Cell Proliferation. PLOS genetics (online), v. 9, p. E1003705, 2013.

Beckedorff, F. C.*, **Amaral, M. S**.*, Deocesano-Pereira, C. And Verjovski-Almeida, S. (2013) Long Non-Coding RNAs and their implications in cancer epigenetics. Biosci. Rep. 33(4):Art:E00061
***These authors equally contributed to this work.*

Almeida, G.T., **Amaral, M.S.**, Beckedorff, F.C.F., Kitajima, J.P., DeMarco, R., Verjovski-Almeida, S. Exploring the *Schistosoma mansoni* adult male transcriptome using RNA-seq. Experimental Parasitology. 2011. http://dx.doi.org/10.1016/j.exppara.2011.06.010.

DeMarco, R.; Oliveira, K. C. P.; Almeida, G. T.; **Amaral, M. S.**; Beckedorff, F. C. F. ; Verjovski-Almeida, S. Comparative transcriptomic profiles of *Schistosoma mansoni* adult male, adult female and egg stages. 2013. ***Submitted***.

Santos R.F., **Amaral, M.S**., Maciel, T.E.F., Vidigal P.M.P., Bastos, M.S, Beckedorff, F. C., Almeida, M.R., Cruz, A.K., Kitajima, J.P., Afonso, L.C.C., Verjovski-Almeida, S. , Fietto,J.L.R. RNA-Seq in *Leishmania* Reveals Distinct Patterns of Internal Polyadenylation in Truncated RNA Molecules. 2013. ***Submitted***

DeOcesano-Pereira,C., Parreira K.S., **Amaral, M.S**, Ayupe, A.C., Jacysyn J.F. , Amarante-Mendes, G.P., Reis, E.M.,Verjovski-Almeida, S..Long Noncoding RNA *INXS* Mediates *BCL-XS* Induced Apoptosis. 2013. ***Submitted***

Resumos publicados em anais de congressos

GOMES, M. P. B. ; RANGEL, L. P. ; **AMARAL, M. S**. ; VERJOVSKI-ALMEIDA, S. ; CORDEIRO, Y. ; SILVA, J. L. . Isolation and characterization of RNA molecules bound to recombinant mouse prion protein. In: XLII Reunião Anual da SBBq, 2013. Anais da XLII Reunião Anual da SBBq, 2013.

AMARAL, M. S. ; BECKEDORFF, F. C. F. ; REIS, E. M. R. ; VERJOVSKI-ALMEIDA, S.. Transcriptome-wide identification of Argonaute2-bound RNAs reveals targeting of long non-coding RNAs. In: Regulatory and Non-coding RNAs, 2012, Cold Spring Harbor. Regulatory and Non-coding RNAs, 2012.

REIS, E. M. R.; BECKEDORFF, F. C. F.;. **AMARAL, M. S**.; VERJOVSKI-ALMEIDA, S. Transcriptional Map and the Effect of Androgen on a Human Cancer Cell Line Using RNA-Seq. In: Cell Symposium 2011- Regulatory RNAs, 2011, Chicago. Regulatory RNAs, 2011

FIETTO, J. L. R. ; SANTOS,R.F. ; MACIEL, T. E. ; VIDIGAL, P. M. P. ; **AMARAL, M. S**. ; BECKEDORFF, F. C. F. ; LEITE, P. M. ; ALMEIDA, M. ; CRUZ, A. K. ; AFONSO, L. C. C. ; KITAJIMA, J.P. ; VERJOVSKI-ALMEIDA, S. . *Leishmania (Viannia) Braziliensis* transcriptome by RNA-SEQ: Looking for genes related to virulence in two distinct isolates. In: XXVII Annual Meeting of the Brazilian Society of protzoology/XXXVIII Annual Meeting on basic research in Chagas' Disease, 2011, 2011, Foz do Iguaçu. XXVII Annual Meeting of the Brazilian Society of protzoology/XXXVIII Annual Meeting on basic research in Chagas' Disease, 2011, 2011.

AMARAL, M. S. ; REIS, E. M. R. ; VERJOVSKI-ALMEIDA, S. . Transcriptome-wide identification of Argonaute2-bound RNAs reveals transcripts originated from intronic genomic regions. In: Functional Genomics and Systems Biology, 2011, Hinxton - Cambridge. Functional Genomics and Systems Biology, 2011.

PIERRY, P. M. ; DASILVA, P. I. P. ; SANTANA, W. O. ; BECKEDORFF, F.C.F. ; **AMARAL, M. S.** ; ALMEIDA, L. G. P. ; VASCONCELOS, A. T. R. ; VERJOVSKI-ALMEIDA, S. . Sequenciamento e comparação do genoma de uma cepa nao-virulenta de *Xylella fastidiosa*. In: 56º Congresso Brasileiro de Genética, 2010, Guarujá. 56º Congresso Brasileiro de Genética, 2010.

SANTANA, W. O. ; **AMARAL, M. S**. ; BECKEDORFF, F.C.F. ; COLETTA-FILHO, H. D. ; ALMEIDA, L. G. P. ; VASCONCELOS, A. T. R. ; DASILVA, A. M. . Pirossequenciamento, genômica comparativa e evolução de *Xylella fastidiosa*. In: 56º Congresso Brasileiro de Genética, 2010, Guarujá. 56º Congresso Brasileiro de Genética, 2010.

AMARAL, M. S. ; VERJOVSKI-ALMEIDA, S. . Caracterização da interação entre RNAs não codificadores (ncRNAs) intrônicos e complexos protéicos modificadores da cromatina. In: 56º Congresso Brasileiro de Genética, 2010, Guarujá. 56º Congresso Brasileiro de Genética, 2010.

4) Atividades Extras

2007- Bolsista DAAD (Deutscher Akademischer Austauschdienst) – Estágio Iniciação Científica- Universitat Konstanz, Alemanha.

2009- Treinamento Sequenciador 454/Roche.

2009 - Ministrante do IV Curso de Verão – "Expressão gênica diferencial entre tecidos não-tumorais e tumorais". Coordenador: Prof. Dr. Bayardo Baptista Torres - 12 a 23 de Janeiro de 2009

2011- Ministrante do VI Curso de Inverno – "Análise da Expressão Gênica por Sequenciamento em Larga Escala (RNA-Seq)". Coordenador: Prof. Dr. Eduardo Moraes Rego Reis – 04 a 15 de Julho de 2011

2012-Atual – Revisor Periódico PLOS One

2013- Treinamento Sequenciador MiSeq/Illumina