

**UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE QUÍMICA**

**Programa de Pós Graduação em Ciências Biológicas (Bioquímica)**

**LAUREN CAMARGO**

**Expressão de RNAs não codificadores intrônicos longos  
em linhagens celulares humanas e o seu controle  
epigenético por metilação do DNA**

**Versão original da tese defendida**

**SÃO PAULO**

**Data do Depósito na SPG:**

**20/07/2012**

LAUREN CAMARGO

Expressão de RNAs não codificadores intrônicos longos em  
linhagens celulares humanas e o seu controle epigenético por  
metilação do DNA

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

Área de Concentração: Bioquímica

Orientador: Dr. Eduardo Moraes Rego Reis

São Paulo

2012

**Lauren Camargo**

**Expressão de RNAs não codificadores intrônicos longos em linhagens celulares humanas e o seu controle epigenético por metilação do DNA**

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

Aprovado em: \_\_\_\_\_

**Banca Examinadora**

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

## DEDICATÓRIA

*Ao meu pai, que plantou em mim a semente da curiosidade científica.*

*À minha mãe, pelo exemplo de força e determinação.*

*Devo tudo a vocês.*

## **AGRADECIMENTOS**

Apesar da autoria individual, essa tese de doutorado é fruto do trabalho, dedicação e incentivo de muitas pessoas, assim agradeço:

Ao Prof. Dr. Eduardo Reis, pela oportunidade de desenvolver essa tese de doutorado e pela sua orientação durante esses anos.

Ao Prof. Dr. Sergio Verjovski-Almeida, por suas contribuições críticas e sugestões essenciais, e pelo compartilhamento da estrutura de seu laboratório.

À minha família, Dani, Baleia e Cissa, pelo apoio incondicional, sempre me incentivando nos momentos em que mais precisei.

Aos amigos:

Ângela Fachel e Carlos De Ocesano, pelo apoio psicológico, e amizade durante esse período de desafios.

Ana Tahira, principalmente pela amizade, mas também pela ajuda com as análises informáticas.

Santiago Villela, Maria Emília Zenteno, Murilo Sena e Katia Oliveira, pelos inúmeros ótimos momentos compartilhados dentro e fora do laboratório.

Desejo que sigam comigo durante toda a vida!

Aos demais amigos de laboratório, Ana Ayupe, Felipe, Kleber, Dinar, Mariana, Letícia, Otto, Vinícius, Rodrigo, Esther, Bianca, Giulliana e Jefferson por estarem sempre dispostos a ajudar e por tornarem o ambiente de trabalho mais prazeroso.

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

Ao Breno Pimentel, pela colaboração durante um período do desenvolvimento desta tese.

Aos ex-membros do laboratório, Camila Egídio, Yuri Moreira, Renato Alvarenga e Helder Nakaya, pelas trocas de ideias e amizade.

Aos demais amigos conquistados durante esse período em que estive no Instituto de Química: Pererê, Matheus, Luciana Gomes, Joana, Maria Fernanda, Eliezer, entre outros. Muito obrigada pelos momentos de descontração e cafezinhos nos corredores!

À Prof. Dra. Mari Cleide Sogayar por ter mantido seu laboratório sempre de portas abertas e por ter me ensinado a manipular células em cultura.

À Prof. Dra. Anamaria A. Camargo e Dra. Paula Asprino, que gentilmente disponibilizaram apoio técnico para os ensaios de conversão do DNA com bissulfito de sódio.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro.

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

*“(...) Sou só uma rede vazia diante dos olhos humanos na escuridão e de dedos habituados à longitude do tímido globo de uma laranja. Caminho como tu, investigando as estrelas sem fim e em minha rede, durante a noite, acordo nu. A única coisa capturada é um peixe preso dentro do vento.”*

*Pablo Neruda (1904 - 1973)*

## RESUMO

Camargo, L. **Expressão de RNA não codificadores intrônicos longos em linhagens celulares humanas e seu controle epigenético por metilação do DNA**. 2012. 227p. Tese de Doutorado - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

Estudos recentes têm revelado que uma fração significativa do transcriptoma de eucariotos é composta por RNAs não codificadores longos (lncRNAs). Este trabalho investigou o padrão de expressão de um conjunto de lncRNAs originados a partir de regiões intrônicas de genes codificadores de proteínas em três linhagens celulares tumorais humanas utilizando microarranjos de DNA customizados. Realizamos uma série de análises *in silico* com a perspectiva de identificar propriedades globais desses transcritos, tais como a abundância relativa em diferentes tecidos, características evolutivas, estruturais e regulatórias, além de possíveis funções celulares. Avaliamos também a contribuição da metilação do DNA, um mecanismo de silenciamento epigenético da expressão de genes codificadores de proteínas, na regulação da expressão de lncRNAs intrônicos.

Observamos que uma fração dos lncRNAs intrônicos detectados nas linhagens estudadas são conservados evolutivamente, tem padrão de expressão tecido específico, e está enriquecida em elementos regulatórios na sua extremidade 5'. Foram identificados subconjuntos de lncRNAs intrônicos possivelmente atuando sobre genes associados a vias regulatórias importantes para o controle do desenvolvimento de organismos e ciclo celular. Comparativamente a mRNAs, uma menor proporção de lncRNAs intrônicos possui ilhas CpGs (CGIs) na vizinhança de seu início de transcrição. Apesar disso, observamos que um subconjunto desses transcritos teve sua expressão sensível ao tratamento com o agente desmetilante de DNA 5-AZA, demonstrando que lncRNAs intrônicos transcritos podem estar sujeitos a regulação transcricional mediada por metilação do DNA.

Dentre os lncRNAs intrônicos regulados por metilação do DNA, destaca-se o lncRNA AS-APP, cuja expressão aumentou em 25 a 80 vezes nas linhagens celulares DU-145 e HEK293, respectivamente, após tratamento com 5-AZA. Este lncRNA possui uma CGI metilada e um promotor ativo a cerca de 4 kb de distância do seu início de transcrição conhecido. O aumento da transcrição do lncRNA AS-APP após desmetilação do DNA correlacionou-se a uma diminuição significativa dos níveis de expressão do mRNA do gene *APP*. Este resultado sugere uma possível ação regulatória em *cis* do lncRNA AS-APP no *locus APP*, um importante gene envolvido na doença de Alzheimer e com expressão associada ao prognóstico de alguns tipos de câncer. Os resultados obtidos neste trabalho reforçam a ideia de que lncRNAs intrônicos constituem unidades transcricionais independentes que se encontram sobre controle regulatório nos diferentes tipos celulares. Foi gerado também um catálogo de lncRNAs intrônicos regulados por metilação que permitirá a seleção de candidatos com maior potencial de relevância funcional para caracterização detalhada.

**Palavras-chave:** RNAs não codificadores intrônicos, metilação do DNA, epigenética, expressão gênica.

## ABSTRACT

Camargo, L. **Long intronic noncoding RNA expression in human cell lines and its DNA methylation epigenetic control.** 2012. 227p. PhD Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Recent studies have revealed that a significant fraction of the eukaryotic transcriptome is composed of long noncoding RNAs (lncRNAs). This work investigated the expression pattern in three human tumor cell lines of a set of lncRNAs originated from intronic regions of protein coding RNAs, using custom DNA oligoarrays. *In silico* analyses were performed to identify global properties of these transcripts such as relative abundance in different human tissues, regulatory, evolutionary and structural aspects, as well as their possible cellular functions. In addition, we evaluated the contribution of DNA methylation, an important epigenetic mechanism that control the expression of protein coding genes, in the regulation of intronic lncRNAs expression. We found that a fraction of the intronic lncRNAs detected in the cell lines are evolutionarily conserved, show a tissue specific expression pattern, and is enriched in regulatory elements at their 5' end region. Subsets of intronic lncRNAs possibly acting on genes associated to important regulatory pathways controlling organism development and cell cycle were identified.

A smaller proportion of intronic lncRNAs relative to mRNAs displayed CpG islands (CGI) in the vicinity of the transcription start site. Notwithstanding, we observed that a subset of these transcripts responded to treatment with the DNA demethylation agent 5-AZA, demonstrating that intronic lncRNAs may be under transcriptional regulation mediated by DNA methylation.

Among intronic lncRNAs regulated by DNA demethylation, stands out AS-APP lncRNA, which was up regulated 25 to 80 times in DU-145 and HEK293 cell lines following 5-AZA treatment, respectively,. This lncRNAs has a methylated CGI and an active promoter at 4-kb upstream from

its known transcription start *site*. Increased AS-APP lncRNA transcription following DNA demethylation correlated with a significant decrease of *APP* gene messenger RNA levels. This finding suggests a possible *cis*-regulatory action of the lncRNA AS-APP in the *APP* locus, an important gene involved in Alzheimer disease and whose expression is associated with prognosis of different cancer types. The results obtained in this study reinforce the idea that intronic lncRNAs constitute independent transcriptional units under regulatory control in the different cell types. It was generated a catalog of intronic lncRNAs regulated by DNA methylation that will allow the selection of candidates with higher potential of functional relevance for detailed characterization.

**Keywords:** intronic noncoding RNAs, DNA methylation, epigenetics, gene expression.

## LISTA DE ABREVIATURAS E SIGLAS

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

5mC: 5-metilcitosina

BRE: elemento de reconhecimento do fator de transcrição TFIIB

CAGE: *Cap Analysis of Gene Expression*

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

cDNA: DNA complementar

CGI: ilha de dinucleotídeos CpG

ChIP: *Chromatin Immunoprecipitation*

ChIP-seq: *Chromatin Immunoprecipitation followed by DNA sequencing*

CPC: *Coding Potential Calculator*

cRNA: RNA complementar

Ct: *Cycle Threshold*

Cy3: *Cyanine -3-CTP*

Cy5: *Cyanine -5-CTP*

DAVID: *Database for Annotation, Visualization and Integrated Discovery*

DMEM: *Dulbecco's Modified Eagle Medium*

DMSO: Dimetilsulfóxido

DNMTs: DNA-metiltransferases

DPE: *Downstream Promoter Element*

dsRNA: *double strand RNA*

ENCODE: *Encyclopedia of DNA Elements Consortium*

EST: *Expressed Sequence Tag*

F.E.T: Fração de Expressão do Total

FANTOM: *Functional Annotation of Mouse Consortium*

FDR: *False Discovery Rate*

GAGE 2A: nome do gene *G antigen 2A*

GAPDH: nome do gene *glyceraldehyde-3-phosphate dehydrogenase*

gDNA : DNA genômico

GO: *Gene Ontology*

GSTP1: nome do gene *glutathione S-transferase pi 1*

H3K27me3: trimetilação da lisina 27 da histona 3

H3K4me3: trimetilação da lisina 4 da histona 3

H3K9me3: trimetilação da lisina 9 da histona 3

HDACs: Histona deacetilases

HMTs: Histonas metilases

Inr: elemento iniciador

KS: teste Kolmogorov-Sminorv

lincRNA: *long intergenic noncoding RNA*

lncRNA: *long noncoding RNA*

Lowess: *Locally Weighted Scatterplot Smoothing*

MBP: *methyl binding protein*

miRNA: *microRNAs*

mRNA: *RNA mensageiro*

NCBI: *National Center for Biotechnology Information*

ncRNA: *noncoding RNA*

NPTX2: *nome do gene neuronal pentraxin II*

ORF: *Open Reading Frame*

PCR: *Polymerase Chain Reaction*

piRNAs: *piwi-interacting RNA*

poli(A): *repetições de adenina encontrada na extremidade 3' de RNAs mensageiros*

PRC1: *Polycomb Repressor Complex 1*

PRC2: *Polycomb Repressor Complex 2*

RACE-PCR: *Rapid Amplification of cDNA Ends*

RISC: *RNA Induced Silencing Complex*

RNAP II: *RNA Polimerase II*

RNA-seq: *sequenciamento de RNA em larga escala*

RPKM: *Reads per Kilobase of exon model per Million mapped reads*

rRNA: RNA ribossomal

RT: *Reverse Transcription*

RT-PCR: transcrição reversa seguida de PCR

SAM: S-adenosil-L-metionina

SAM: *Significance Analysis of Microarrays*

snoRNAs: *small nucleolar RNA*

snRNAs: *small nuclear RNA*

TAE: tampão Tris-Acetato-EDTA

TFIIB: Fator de Transcrição IIB

T<sub>m</sub>: temperatura de *melting* (desnaturação)

tRNA: RNA transportador

TSS: *Transcription Start Site*

UTR: *Untranslated Region*

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
1.1	A complexidade do transcriptoma eucariótico	19
1.2	RNAs não codificadores (ncRNAs) em eucariotos	22
1.3	Controle epigenético da expressão gênica em eucariotos	28
1.4	RNAs não codificadores longos (lncRNAs) e regulação da expressão gênica em eucariotos	44
<b>2</b>	<b>OBJETIVOS</b>	<b>50</b>
2.1	Objetivos gerais	50
2.2	Objetivos específicos	50
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>51</b>
3.1	Linhagens celulares humanas e condições gerais de cultivo	51
3.2	Tratamento das linhagens celulares com o agente desmetilante 5-Aza-2'-deoxicitidina (5-AZA)	52
3.3	Extração e purificação de RNA total	55
3.4	Avaliação do rendimento, pureza e integridade do RNA obtido	56
3.5	Análise da expressão gênica em larga escala	58
3.5.1	Lâmina de oligonucleotídeos com 44 mil sondas (oligoarray) enriquecida em transcritos não codificadores	58
3.5.2	Obtenção dos alvos fluorescentes	59
3.5.3	Desenho experimental e hibridização	61
3.5.4	Avaliação do potencial codificador de lncRNAs intrônicos detectados nas linhagens DU-145, MCF-7 e Mia PaCa 2	63
3.5.5	Identificação de lncRNAs intrônicos em bibliotecas de RNA-seq	64

3.5.6	Comparação de intensidade de expressão entre lncRNAs intrônicos e mRNAs codificadores de proteínas -----	67
3.5.7	Correlação de expressão entre lncRNAs intrônicos e mRNAs detectados -----	68
3.5.8	Identificação de elementos regulatórios na extremidade 5' conhecida dos lncRNAs intrônicos -----	70
3.5.9	TransMaps – Análise de conservação de expressão -----	73
3.5.10	PhastConsElements 46way – Análise de conservação de sequência -----	74
3.5.11	RNAz – Predição de estrutura secundária -----	75
3.5.12	Definição de lncRNAs intrônicos e mRNAs codificadores de proteínas regulados pelo tratamento com 5-AZA - Análise estatística -----	75
3.5.13	Anotação funcional dos genes codificadores de proteínas regulados pelo tratamento com 5-AZA -----	76
<b>3.6</b>	<b>Abordagens para validação da expressão fita específica de lncRNAs intrônicos regulados pelo tratamento com 5-AZA -----</b>	<b>77</b>
3.6.1	Transcrição reversa fita específica -----	77
3.6.2	RACE-PCR -----	80
<b>3.7</b>	<b>Clonagem e sequenciamento dos produtos obtidos -----</b>	<b>82</b>
3.7.1	Quantificação de RNAs por PCR em tempo real -----	83
<b>3.8</b>	<b>Quantificação absoluta de RNAs por PCR em tempo real -----</b>	<b>85</b>
<b>3.9</b>	<b>Identificação de uma possível região promotora para o lncRNA AS-APP e seu estudo funcional -----</b>	<b>87</b>
3.9.1	Identificação da predição informática do possível promotor do lncRNA AS-APP ----	87
3.9.2	Amplificação e clonagem da região promotora -----	89
3.9.3	Lipofecção de células HEK293 para ensaios de atividade promotora com gene repórter- -----	90
3.9.4	Ensaio de atividade promotora -----	91
<b>3.10</b>	<b>Extração, quantificação e fragmentação de DNA genômico (gDNA) -----</b>	<b>92</b>
3.10.1	Extração do gDNA -----	92
3.10.2	Avaliação do rendimento, pureza e integridade do gDNA -----	92

3.10.3	Fragmentação do gDNA .....	93
<b>3.11</b>	<b>Análise de metilação do DNA .....</b>	<b>93</b>
3.11.1	Digestão do DNA metilado utilizando a endonuclease McrBC.....	93
3.11.2	Sequenciamento do DNA tratado com bissulfito de sódio.....	95
<b>4</b>	<b>RESULTADOS .....</b>	<b>103</b>
<b>4.1</b>	<b>Expressão em larga escala de ncRNAs longos (lncRNAs) intrônicos e mRNAs codificadores de proteínas em linhagens tumorais humanas.....</b>	<b>103</b>
<b>4.2</b>	<b>LncRNAs intrônicos são expressos em níveis mais baixos do que mRNAs codificadores de proteínas .....</b>	<b>109</b>
<b>4.3</b>	<b>LncRNAs intrônicos apresentam padrão de expressão mais linhagem/tecido específico do que mRNAs codificadores de proteínas.....</b>	<b>110</b>
<b>4.4</b>	<b>Correlação de expressão entre lncRNAs intrônicos e mRNAs codificadores de proteínas .....</b>	<b>115</b>
<b>4.5</b>	<b>Identificação de elementos regulatórios localizados à 5' dos lncRNAs intrônicos .....</b>	<b>125</b>
<b>4.6</b>	<b>Conservação evolutiva dos lncRNAs intrônicos.....</b>	<b>128</b>
<b>4.7</b>	<b>Regulação da expressão de lncRNAs intrônicos e mRNAs por metilação do DNA ---</b>	<b>134</b>
<b>4.8</b>	<b>Seleção de lncRNAs intrônicos regulados por metilação do DNA para validação ----</b>	<b>141</b>
<b>4.9</b>	<b>Expressão fita específica de lncRNAs intrônicos .....</b>	<b>142</b>
<b>4.10</b>	<b>Determinação da estrutura primária de lncRNAs intrônicos por RACE 5' e 3' .....</b>	<b>144</b>
<b>4.11</b>	<b>Quantificação absoluta da expressão de lncRNAs intrônicos regulados por metilação --- .....</b>	<b>147</b>
<b>4.12</b>	<b>Expressão do mRNA codificador de proteínas correlaciona-se negativamente com o padrão de expressão do lncRNA intrônicos antisense produzido no locus APP .....</b>	<b>148</b>
<b>4.13</b>	<b>Identificação de possível região promotora do lncRNA AS-APP .....</b>	<b>152</b>

<b>5</b>	<b>DISCUSSÃO</b>	<b>162</b>
<b>5.1</b>	<b>Expressão de lncRNAs intrônicos e mRNAs nas linhagens tumorais DU-145, MCF-7 e Mia PaCa 2.</b>	<b>162</b>
<b>5.2</b>	<b>Correlação de expressão entre lncRNAs intrônicos e mRNAs codificadores de proteínas</b>	<b>163</b>
<b>5.3</b>	<b>Identificação de elementos regulatórios localizados à 5' de lncRNAs intrônicos</b>	<b>168</b>
<b>5.4</b>	<b>Conservação evolutiva de lncRNAs intrônicos</b>	<b>170</b>
<b>5.5</b>	<b>Regulação da expressão de lncRNAs intrônicos por metilação do DNA</b>	<b>173</b>
<b>6</b>	<b>CONCLUSÕES E CONSIDERAÇÕES FINAIS</b>	<b>181</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>183</b>
	<b>APÊNDICE</b>	<b>209</b>
	<b>ANEXO</b>	<b>228</b>

# 1 INTRODUÇÃO

## 1.1 A complexidade do transcriptoma eucariótico

Nos últimos anos, análises da expressão gênica em larga escala têm demonstrado que a atividade transcricional do genoma humano, e de outros mamíferos e eucariotos superiores é muito maior do que havia sido previsto, sendo que a maioria desses novos transcritos mapeia fora de regiões anotadas do genoma (Willingham and Gingeras 2006). Uma análise pioneira da atividade transcricional dos cromossomos 21 e 22 com *microarrays* de alta densidade (*tiling arrays*) indicou que a diversidade de transcritos em células humanas é aproximadamente 10 vezes maior do que o predito pela anotação gênica existente (Kapranov et al. 2002). Kampa e colaboradores, em uma análise da atividade dos mesmos cromossomos em 11 linhagens celulares, detectaram que a menor parte dos transcritos detectados representavam genes codificadores de proteínas (Kampa et al. 2004). Consistentemente, esta abundância de transcritos tem sido confirmada por outros experimentos de *tiling array* cobrindo grandes regiões genômicas em humanos e murinos (Bertone et al. 2004; Cheng et al. 2005; Kapranov et al. 2005) e também por diferentes abordagens experimentais: sequenciamento de bibliotecas de cDNA (Okazaki et al. 2002; Numata et al. 2003; Carninci et al. 2005), MPSS (*Massively Parallel Signature Sequencing*) (Jongeneel et al. 2005), SAGE (*Serial Analysis of Gene Expression*) (Chen et al. 2002; Saha et al. 2002), ou predições computacionais (Washietl et al. 2005; Glusman et al. 2006).

Com o emprego da tecnologia de *tiling arrays*, foi possível avaliar a atividade transcricional da porção não repetitiva do genoma humano (Bertone et al. 2004). Utilizando apenas a fração poli(A)+ de RNA isolado de fígado de vários indivíduos, um total de 13.889 unidades transcricionais longas foram detectadas, das quais 65% mapeiam em regiões não anotadas do genoma. Deste grupo, cerca de 35% são transcritos que mapeiam inteiramente dentro

de íntrons, sendo que metade são transcritos na mesma orientação do gene codificador e a outra metade é transcrita na fita oposta, gerando RNAs antissenso.

Todas estas observações têm alterado a compreensão da organização e funcionamento dos genomas eucarióticos. Ao invés de concentrar toda a atividade transcricional em uma fração minoritária que contem genes codificadores de proteína, o genoma parece ser formado por redes intrincadas de regiões transcritas, incluindo porções codificadoras e não codificadoras, gerando moléculas de RNA que se sobrepõe total ou parcialmente, na mesma orientação ou em orientações opostas (Kapranov et al. 2005). Em uma análise do transcriptoma de camundongos e humanos foi observado que grande parte do genoma é transcrito em ambas as fitas, chamadas de “florestas transcricionais”, intercaladas por regiões de desertos transcricionais (Carninci et al. 2005). Em muitos casos, em um mesmo *locus* é observada a expressão de cadeias de pelo menos três transcritos independentes com sobreposição senso/antissenso ou com promotores bidirecionais (Engstrom et al. 2006). Mais ainda, dados experimentais indicam que mais de 70% dos transcritos em mamíferos possuem sobreposição com um RNA antissenso (Katayama et al. 2005). Estas observações somam-se ao fato de que a maioria dos transcritos possui sítios de iniciação da transcrição, de terminação e padrões de *splicing* alternativos (Carninci et al. 2005), aumentando ainda mais a complexidade do transcriptoma. Mais recentemente, o projeto ENCODE (*Encyclopedia of DNA Elements Consortium*) analisou intensivamente 1% do genoma humano utilizando várias abordagens, incluindo *tiling arrays* genômicos de alta densidade e sequenciamento “*deep sequencing*” de bibliotecas de cDNA (RNA-seq), e demonstrou que ao menos 90% do DNA genômico é transcrito em diferentes células humanas, sendo que desta fração, apenas 1,2% é capaz de codificar proteínas (Birney et al. 2007; Kapranov et al. 2007).

A análise do transcriptoma de organismos eucarióticos inferiores usando *tiling arrays* genômicos mostrou que a transcrição pervasiva do genoma não é restrita aos mamíferos, mas

estende-se a moscas (>85% do genoma de *Drosophila melanogaster* é transcrito nas primeiras 24h do desenvolvimento embrionário), nematóides (>70% do genoma é transcrito em *C. elegans* em populações de diferentes estágios de desenvolvimento) e leveduras (>85% do genoma é expresso em meio rico em nutrientes) (Amaral and Mattick 2008).

Somada às análises do transcriptoma em larga escala em diversos organismos, análises genômicas têm revelado que a complexidade biológica de um organismo não é diretamente relacionada ao número de genes codificadores de proteínas de seu genoma, mas sim com o aumento da proporção não codificadora do transcriptoma dessas espécies. Por exemplo, vertebrados superiores, como humanos e camundongos, têm aproximadamente 20.000 a 25.000 genes. Comparativamente, apesar da menor complexidade biológica, eucariotos inferiores como, *Drosophila melanogaster* e *Caenorhabditis elegans*, têm um número similar de genes codificadores de proteínas, 13.000 e 19.000, respectivamente (Consortium 1998; Adams et al. 2000; Lander et al. 2001; Waterston et al. 2002). A partir desse aparente paradoxo, foi observado que enquanto a complexidade biológica de um organismo aumenta, observa-se o aumento proporcional de sequências não codificadoras em seu genoma, que pode variar de 0,25% do total no genoma de procariotos até mais de 98% no genoma humano (Taft et al. 2007). Estas observações sugerem que o ganho de complexidade funcional que ocorre ao longo da escala evolutiva não depende apenas do maior número de genes codificadores de proteínas, e que a aquisição de novos RNAs não codificadores (ncRNAs) possa fornecer um repertório maior de moléculas com funções regulatórias, permitindo o desenvolvimento complexo de eucariotos superiores (Taft et al. 2007).

## 1.2 RNAs não codificadores (ncRNAs) em eucariotos

Atualmente existem fortes indícios de que a expressão de ncRNAs provenientes de regiões intrônicas com orientação transcricional senso ou antissenso em relação ao mRNA do mesmo *locus* gênico e de regiões intergênicas ativas seja um fenômeno bastante geral, observado em múltiplos modelos biológicos e medido através de diversas técnicas (Mattick and Makunin 2006). A funcionalidade destes transcritos é um tema em discussão pela comunidade científica. Há a hipótese de que ncRNAs possam representar meramente um ruído transcricional resultante do reconhecimento ilegítimo de promotores pela RNA polimerase II (RNAP II) (Kapranov et al. 2005; Struhl 2007; van Bakel et al. 2010). Por outro lado, tem sido proposto que os ncRNAs ainda não caracterizados possam desempenhar funções regulatórias (Mattick and Makunin 2006). Especula-se que estes RNAs poderiam compor redes de sinalização que controlariam e integrariam processos celulares complexos, com papel determinante na ontogenia e evolução dos eucariotos (Mattick 1994; Mattick 2001; Amaral and Mattick 2008).

Argumentos a favor da expressão funcional de ncRNAs são apoiados por estudos que combinam ensaio de imunoprecipitação da cromatina (*Chromatin Immunoprecipitation Assays - ChIP*) e *tiling arrays*. Esta estratégia permitiu o mapeamento de sítios regulatórios, em alta resolução, ao longo do genoma humano e mostrou que existe ligação de fatores de transcrição não só na região promotora de genes codificadores de proteínas, mas também em regiões intrônicas e intergênicas do genoma, o que apoia a existência disseminada da transcrição (Martone et al. 2003; Cawley et al. 2004; Euskirchen et al. 2004; Kim et al. 2005). De fato, nos últimos anos, vários ncRNAs têm sido caracterizados, com funções associadas à regulação de uma grande variedade de processos celulares tais como *splicing* alternativo, edição de RNA (conversão de adenosinas em inosinas), silenciamento gênico, compensação de dose, *imprinting*

genômico, metilação de DNA e formação de heterocromatina (Chen and Carmichael 2010; Moran et al. 2012; Rinn and Chang 2012).

Tem sido convencional na literatura, para efeito de sistematização, classificar ncRNAs em função do seu tamanho: ncRNAs longos (lncRNAs) quando maiores que 200 nucleotídeos e ncRNAs curtos, menores que 200 nucleotídeos (Clark and Mattick 2011). Entre os ncRNAs curtos (< 200 nt) estão inclusos os microRNAs (miRNAs), *small nuclear RNAs* (snRNAs), *small nucleolar RNAs* (snoRNAs), *piwi-interacting RNAs* (piRNAs), pequenos RNAs associados à regiões promotoras, entre outros (Girard et al. 2006; Carthew and Sontheimer 2009; Fejes-Toth et al. 2009; Taft et al. 2009). Entre os ncRNAs longos incluem-se RNAs com 200 até vários milhares de nucleotídeos, com ou sem *splicing* (Qureshi et al. 2010). Alguns lncRNAs servem de precursores para ncRNAs curtos como miRNAs e snoRNAs (Mattick and Makunin 2005).

Dentre os ncRNAs curtos, os mais estudados e caracterizados funcionalmente são os microRNAs (miRNAs). Os miRNAs compreendem uma grande família de pequenos RNAs que atuam como reguladores pós-transcricionais chave da expressão gênica em metazoários e plantas (Bartel 2009; Carthew and Sontheimer 2009). Os miRNAs são originados a partir de RNAs precursores (pri-miRNAs), tipicamente gerados pela RNAP II em unidades transcricionais independentes, ou em regiões intrônicas ou exônicas de genes. Após sua transcrição, o pri-miRNA forma uma estrutura secundária característica em forma de grampo (*hairpin*) que é clivada sequencialmente por enzimas da família RNase III. A enzima nuclear *Drosha* cliva o pri-miRNA dando origem ao precursor pré-miRNA, com cerca de 70 nucleotídeos, que é em seguida, exportado para o citoplasma e processado pela enzima Dicer gerando a forma madura do miRNA (Kim et al. 2009). Alternativamente, microRNAs podem ser gerados independentemente de *Drosha*, a partir de regiões intrônicas curtas (*mirtrons*), como um subproduto da maquinaria de *splicing* (Sibley et al. 2012).

As moléculas de miRNA atuam como RNAs antissenso, direcionando especificamente os RNAs alvos ou à degradação ou à inibição da tradução, dependendo de se sua complementaridade for total ou parcial, respectivamente (Bartel 2004; Petersen et al. 2006). O processo de repressão da expressão gênica envolve a via de “interferência de RNA”, ou RNAi (RNA interference), no qual os miRNAs se associam a complexos contendo endonucleases, denominados RISC (*RNA induced silencing complex*), onde direcionam o silenciamento pós-transcricional de genes específicos (Bartel 2004).

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 enorme gama de processos fisiológicos (Ambros 2004), inclusive em tumorigênese (Esquela-Kerscher and Slack 2006; Ventura and Jacks 2009). Um miRNA pode ter muitos alvos e um mRNA pode ser alvo de vários miRNAs, ampliando, assim, sua influência em redes de regulação da expressão gênica (Farh et al. 2005; Lim et al. 2005). Alguns estudos indicam que aproximadamente 50% dos genes humanos têm a expressão modulada por miRNAs (Krol et al. 2010), sendo que já foram descritos cerca de 1900 miRNAs em humanos (Kozomara and Griffiths-Jones 2011).

Além dos miRNAs, existem outras classes já conhecidas de pequenos ncRNAs funcionalmente relevantes (Liu et al. 2005; Mattick and Makunin 2005). Os snoRNAs possuem de 60 a 300 nt e são originados a partir do processamento de íntrons excisados ou da clivagem endonucleolítica de transcritos primários. Sua atuação se baseia no pareamento de pequenos trechos de sua sequência com RNAs alvo, sendo tipicamente envolvidos no processo de modificação de RNAs ribossomais, no qual direcionam a atuação de proteínas que promovem a pseudouridilação ou metilação de bases específicas (Kiss 2002).

Os piRNAs (*Piwi-interacting RNAs*), encontrados associados com proteínas Argonauta da subfamília PIWI, são pequenos RNAs regulatórios (de 26 a 30 nt) extremamente abundantes em testículos de ratos e camundongos, que estão envolvidos no processo de espermatogênese em mamíferos, onde atuam dirigindo o silenciamento da expressão de retrotransposons (Girard et al. 2006; Grivna et al. 2006; Siomi et al. 2011).

A classe de lncRNAs, arbitrariamente definida apenas com base no tamanho de suas sequências (> 200 nt), é uma classe de transcritos eucarióticos sem potencial codificador para proteínas que têm recentemente emergido a partir de estudos do transcriptoma de diversos organismos (Okazaki et al. 2002; Bertone et al. 2004; Birney et al. 2007; Kapranov et al. 2007; Amaral et al. 2011).

Esta classe compreende ncRNAs intergênicos, bem como ncRNAs intragênicos que se sobrepõem a ambas as fitas de *loci* gênicos codificadores de proteínas (Carninci et al. 2005; Katayama et al. 2005; Kapranov et al. 2007), incluindo lncRNAs associados a promotores (Kapranov et al. 2007), lncRNAs intrônicos (Louro et al. 2009), transcritos bidirecionais (Katayama et al. 2005), isoformas alternativas não codificadoras de genes codificadores de proteínas (Hashimoto et al. 2009) e mRNAs que possuem papel regulatório como RNAs não traduzidos (Candeias et al. 2008; Chooniedass-Kothari et al. 2010).

Muitos lncRNAs detectados compartilham características estruturais comuns a mRNAs, tais como presença de cauda poli(A), cap 5' e *splicing*, sendo chamados de mRNA-like ncRNAs (Erdmann et al. 2001; Numata et al. 2003). Porém, estudos recentes demonstram que uma fração majoritária dos RNAs expressos em células de mamíferos não possui estas características. Foi observado que cerca de metade de todos os transcritos gerados em células humanas são retidos no núcleo, e que cerca de 44% dos RNAs nunca são poliadenilados e aproximadamente 37% são bimórficos, encontrados nas formas poli(A)+ e poli(A)- (Cheng et al. 2005). Mais da metade dos

RNAs poli(A)- nunca exportada para o citoplasma são transcritos intrônicos. Uma parcela dos lncRNAs encontrados com as metodologias baseadas em síntese de cDNA com oligo(dT) não correspondem necessariamente a RNAs pós-transcricionalmente poliadenilados, mas a transcritos ainda maiores detectados devido à existência de trechos ricos em adeninas em sua sequência que servem de sítios internos de anelamento (*internal priming*) (Furuno et al. 2006).

Estudos de genômica comparativa têm indicado a relevância funcional das regiões não codificadoras de proteína, ao revelar que em torno de 5% do genoma humano se encontra sob pressão seletiva, apesar de apenas 1,5% ser composto por sequências codificadoras (Waterston et al. 2002; Siepel et al. 2005). Ainda, a maior parte das regiões ultraconservadas, sequências de pelo menos 200 pb com 100% de identidade entre o genoma humano e murino (Bejerano et al. 2004), e também das sequências preservadas da inserção de transposons (Simons et al. 2006) são localizadas em regiões intrônicas e intergênicas.

Recentemente, foram identificados lncRNAs (>1kb) intergênicos, coletivamente denominados como lincRNAs (*long intergenic noncoding RNAs*). Estes transcritos têm expressão tecido específica, são conservados evolutivamente e alguns têm sido descritos como responsáveis por recrutar e conferir especificidade de alvo à complexos remodeladores da cromatina (Guttman et al. 2009; Khalil et al. 2009; Guttman et al. 2010; Cabili et al. 2011).

Suportando a ideia de que muitos lncRNAs possam ter relevância fisiológica, tem sido observado que a expressão de um grande número dessas moléculas em humanos e camundongos ocorre de forma tecido específica e é regulada em resposta a sinais externos e mudanças ambientais (Cawley et al. 2004; Kampa et al. 2004; Reis et al. 2004; Ravasi et al. 2006; Nakaya et al. 2007). Por exemplo, o nível da expressão de lncRNAs é regulado em células humanas em

diferenciação induzida por ácido retinóico, com uma resposta similar à observada para a população de mRNAs (Cawley et al. 2004).

Finalmente, uma informação importante vem da constatação de que lncRNAs estão envolvidos no desenvolvimento de doenças genéticas em humanos (Costa 2005; Szymanski et al. 2005), como por exemplo, o câncer (Gibb et al. 2011; Prensner and Chinnaiyan 2011; Reis and Verjovski-Almeida 2012).

Nosso grupo tem focado o estudo na caracterização de uma classe de lncRNAs sem *splicing* e originados de regiões intrônicas do genoma, com orientação senso ou antissenso em relação aos genes em que mapeiam (Reis et al. 2004; Louro et al. 2007). Um trabalho realizado em nosso laboratório demonstrou que um subconjunto dos lncRNAs intrônicos têm sua transcrição diretamente regulada por andrógeno, de maneira semelhante ao que ocorre com mRNAs (Louro et al. 2007). Além disso, foi demonstrado que subconjuntos desses lncRNAs intrônicos são expressos em tecidos de rim, próstata e fígado humanos (Nakaya et al. 2007), em células tronco mesenquimais (Secco et al. 2009), em células CD34<sup>+</sup> de pacientes com síndrome mielodisplásica (Baratti et al. 2010), bem como em tumores de pâncreas (Tahira et al. 2011), próstata (Reis et al. 2004), rim (Brito et al. 2008) e em leucemia mielóide crônica (Silveira, *et al.*, 2012, submetido)<sup>1</sup>. Estes resultados apontam para um papel de lncRNAs intrônicos no controle da expressão gênica em tecidos humanos, e em especial, em doenças como no câncer.

---

<sup>1</sup> Silveira, R. A.; Fachel, A. A.; Moreira, Y. B.; De Souza, C. A.; Costa, F. F.; Verjovski-Almeida, S.; Pagnano, K. B. (2012). "Protein-coding genes and noncoding RNAs are differentially expressed in dasatinib-treated chronic myeloid leukemia patients with resistance to imatinib." Submetido.

### 1.3 Controle epigenético da expressão gênica em eucariotos

A história da epigenética está ligada ao estudo da evolução e desenvolvimento. As alterações fenotípicas das células durante o desenvolvimento de organismos multicelulares ocorrem a despeito de elas compartilharem o mesmo genótipo. Desde que são células tronco embrionárias até serem células completamente diferenciadas, todas as células de um organismo compartilham a mesma sequência de DNA, no entanto, diferem muito em relação ao perfil de genes que expressam. Como definição primordial, epigenética é o estudo das mudanças na função gênica mitoticamente e meioticamente herdáveis que não podem ser explicadas pelas alterações na sequência do DNA. Nos últimos anos, no entanto, o termo epigenética tem sofrido uma grande evolução em seu significado, principalmente devido ao aumento do conhecimento sobre os mecanismos moleculares que regulam a expressão gênica em eucariotos. Atualmente, epigenética pode ser definida molecularmente como a soma de alterações da cromatina que coletivamente estabelecem e propagam padrões diferentes de expressão gênica e silenciamento a partir de um mesmo genoma.

Os diferentes fenômenos epigenéticos estão ligados pelo fato de que o DNA não se encontra no interior do núcleo das células. Este encontra-se intimamente complexado à proteínas especializadas (histonas) formando o que chamamos de cromatina (Kornberg 1974). Inicialmente a função da cromatina foi estabelecida como sendo uma forma passiva de empacotamento e organização do DNA, no entanto, com o passar dos anos, vários mecanismos de modificações covalentes e não covalentes da cromatina foram descobertos e demonstrados como influenciadores da regulação da expressão gênica. Estes mecanismos incluem modificações pós-traducionais das histonas, remodelamento da cromatina dependente de energia que mobilizam ou alteram a estrutura dos nucleossomos, troca de variantes de histonas dentro e fora

dos nucleossomos, a ligação de RNAs não codificadores à cromatina e metilação do DNA em resíduos de citosinas, usualmente, em dinucleotídeos CpGs. Estas modificações têm sido implicadas na manutenção da estabilidade do genoma, regulação da transcrição e como contribuintes para a memória epigenética. Muitas, mas nem todas essas modificações da cromatina são reversíveis. As marcas transitórias são relevantes, pois elas impõem mudanças na cromatina em resposta a diferentes estímulos internos e externos à célula (Jaenisch and Bird 2003), e dessa maneira, regulam o acesso e a processividade da maquinaria transcricional (Sims et al. 2004). No entanto, algumas das modificações das histonas, a metilação do DNA e as alterações da estrutura dos nucleossomos podem ser estáveis e herdáveis durante as divisões celulares, estabelecendo uma “memória celular”. Por essas perspectivas, as assinaturas da cromatina podem ser vistas como um sistema altamente organizado de estocagem de informação, podendo compartimentalizar regiões do genoma e acomodar respostas a sinais ambientais que ditam diferentes programas de expressão gênica.

Embora as histonas possam ser pós-transcricionalmente modificadas em diferentes regiões de suas sequências peptídicas, as regiões N-terminais ou caudas N-terminais são preferencialmente modificadas. A razão para isso é que na estrutura dos nucleossomos, as caudas N-terminais estão expostas, possibilitando o acesso de complexos enzimáticos modificadores da cromatina (Baneres et al. 1997). As modificações nas caudas das histonas que regulam a estrutura da cromatina são, em geral, dinâmicas, altamente reguladas e transientes, de forma a se adaptarem a mudanças fisiológicas nas células. Inicialmente, foi proposto que as modificações nas caudas das histonas poderiam ser vistas como um novo código biológico (Jenuwein and Allis 2001), que determinassem o reconhecimento e a ligação de proteínas específicas a essas caudas, culminando na expressão ou repressão gênica. Ensaio de imunoprecipitação da cromatina em larga escala em diversos genomas têm encontrado fortes correlações entre a trimetilação da lisina

4 da histona H3 (H3K4me3) e regiões promotoras ativadas (Strahl and Allis 2000; Santos-Rosa et al. 2002; Bernstein et al. 2005), bem como da trimetilação da lisina 9 (Hall et al. 2002; Lippman et al. 2004; Martens et al. 2005) e da lisina 27 (Litt et al. 2001; Ringrose et al. 2004) da histona H3 (H3K9me3 e H3K27me3) e o silenciamento da heterocromatina. No entanto, foi observado que diferentes combinações de modificações podem resultar em diferentes efeitos na expressão gênica e esses efeitos podem ser diferentes em espécies diferentes, tornando o código de histonas não universal. Um exemplo é a hiperacetilação da H4K12, que pode ter efeito de repressão da expressão gênica nos genomas de levedura e drosófila (Turner et al. 1992), mas esse efeito, não é observado no genoma humano. Para adicionar outro nível de sofisticação e precisão, cada sucessiva modificação produz um novo contexto onde outras modificações podem ser feitas. Assim, a ordem e a posição das modificações, além do tipo, determinam o contexto no qual essas alterações existem e sinergicamente produzem um efeito regulatório. Esse, talvez seja a maior limitação do código de histonas, o fato de que uma modificação não traduz apenas um efeito biológico, e sim a acumulação de diferentes modificações em combinação parece definir e contribuir para os papéis biológicos (Henikoff 2005).

Nas últimas décadas, foi identificado um grande número de enzimas modificadoras da cromatina. Essas enzimas frequentemente fazem parte de grandes complexos que podem catalisar a incorporação ou remoção das modificações covalentes para resíduos de aminoácidos específicos nas histonas. Muitos tipos de modificações covalentes das histonas têm sido identificados e caracterizados; estas incluem a acetilação, metilação, fosforilação, ubiquitinação, sumoilação, ADP-ribosilação, biotinilação das histonas e isomerização de prolinas, entre outras.

A acetilação e a metilação das histonas, principalmente H3 e H4, foram as primeiras modificações a serem descritas e foram propostas correlacionarem com mudanças positivas e negativas na atividade transcricional, respectivamente (Allfrey et al. 1964). O efeito primário

produzido pela acetilação da histona é a parcial neutralização das cargas positivas das histonas, reduzindo sua afinidade pelo DNA (Hong et al. 1993). Além disso, as caudas de histonas acetiladas podem diretamente recrutar fatores de transcrição, por meio de certos domínios proteicos, tais como bromodomínios e cromodomínios (Zeng and Zhou 2002). A importância funcional da acetilação das histonas depende completamente de uma deacetilação eficiente e acurada. Embora esteja claro que, em geral, a hipoacetilação de histonas está relacionada com a repressão da transcrição (Wade 2001), alguns estudos genéticos em leveduras sugerem um papel na ativação da transcrição para as histonas deacetilases (HDACs) (Wang et al. 2002).

A metilação de histonas foi reportada há muitos anos atrás (Murray 1964). Diferentemente de algumas modificações de histonas que podem afetar as funções da cromatina por alterar a carga e interações estruturais dentro do nucleossomo, a metilação de histonas é uma modificação relativamente inerte que parece funcionar como um sítio de nucleação para proteínas efetoras. Essas proteínas efetoras frequentemente recrutam outras proteínas com as quais interage causando seu impacto sobre a cromatina. Histonas metiltransferases (HMTs) catalisam a transferência de grupos metil a partir da S-adenosil-L-metionina (SAM) para resíduos de argininas ou lisinas nas histonas. As argininas podem ser mono- ou di-metiladas enquanto que lisinas podem ser mono-, di- ou tri-metiladas pelas HMTs (Zhang and Reinberg 2001). Em 2004, foi identificada uma enzima, a LSD1, que é capaz de catalisar a mono- e di- demetilação específica da lisina 4 da histona 3 por meio de uma reação de amino oxidase (Shi et al. 2004). A função da metilação é dependente não somente de qual resíduo de aminoácido esteja sendo afetado, mas também da interação estabelecida entre a metilação e as outras modificações de histona (Turner 2002). As mais amplamente estudadas marcas de metilações em lisinas foram encontrada sobre H3K4, K9, K27, K36, K79 e H4K20. Em geral, a metilação das lisinas H3K4, K36 e K79 são encontradas próximas a unidades transcricionais ativas, enquanto que a metilação

das lisinas H3K9 e H4K20 são marcos de regiões silenciadas ou heterocromáticas. A metilação de argininas em histonas ocorre sobre as argininas H3R2, R8, R17 e R26 e H4R3. Essas marcas possuem papel de definir ambos os *status* ativo e reprimido da cromatina.

Outro importante mecanismo pelo qual são induzidas transições na cromatina é pelo recrutamento de complexos remodeladores da cromatina que utilizam energia gerada pela hidrólise do ATP para mudar a composição dos nucleossomos de uma maneira não covalente. Os complexos remodeladores da cromatina frequentemente trabalham em conjunto com as enzimas modificadoras da cromatina ativadoras (HDACs, HMTs, etc.). As proteínas que compõem os complexos remodeladores de cromatina são, em geral, categorizados em duas famílias: os SNF2H/ISWI e os SWI/SNF. A família SNF2H/ISWI mobiliza os nucleossomos ao longo do DNA (Tsukiyama et al. 1995; Varga-Weisz et al. 1997), enquanto que a família SWI/SNF transientemente altera a estrutura do nucleossomo, expondo contatos entre histonas e DNA.

As diferenças na composição da cromatina que podem ocorrer devido à presença de variantes de histonas contribuem para o estabelecimento de regiões cromossômicas com funções especializadas. Por exemplo, a troca da histona *core* H2A pela sua variante H2A.Z correlaciona-se com a troca das atividades de repressão para a promoção transcricional (Kamakaka and Biggins 2005). A CENP-A, uma variante centromérica específica da H3, é essencial para a função dos centrômeros durante a segregação cromossômica. A macroH2A é uma variante da histona H2 que está especificamente associada com o cromossomo X inativo em mamíferos (Chadwick et al. 2001). Muitos trabalhos na literatura têm documentado que variantes de histona possuem seu próprio padrão de susceptibilidade a modificações, especificado pelo pequeno número de aminoácidos diferentes que ocorrem numa variante e não na outra (Henikoff and Ahmad 2005; Sarma and Reinberg 2005). Além disso, algumas variantes de histona possuem distintos aminoácidos em seus domínios amino- e carboxi-terminais com atividade reguladora da

cromatina própria e com afinidade específicas por fatores de transcrição. Muitos dos complexos remodeladores da cromatina são específicos para a troca de histonas específicas por suas variantes. Por exemplo, a troca da H3 por sua variante H3.3 ocorre via a ação do complexo remodelador HIRA (*histone regulator A*) (Tagami et al. 2004) e a troca de H2A por H2A.Z é realizada pela atividade do complexo remodelador SWR1 (Swi2/Snf2-related ATPase 1) (Mizuguchi et al. 2004).

A metilação do DNA é o mecanismo epigenético correlacionado com a repressão gênica há mais tempo conhecido (Holliday and Pugh 1975; Riggs 1975; Razin and Riggs 1980) e está presente em grau variado em quase todos os eucariotos. No genoma de mamíferos a metilação do DNA ocorre pela adição do grupo metil ao carbono 5 de resíduos de citosinas, principalmente em dinucleotídeos CpGs. Sabe-se que os dinucleotídeos CpG estão pouco representados no genoma de mamíferos. No entanto, comuns em determinadas regiões, denominadas ilhas CpG (CGIs). Essa menor frequência de CpGs no genoma é principalmente explicada pelo fato de que as 5-metil-citosinas (5mC) são espontaneamente convertidas a timinas ao longo do tempo evolutivo. Somente CpGs que estão sobre alguma pressão seletiva, ou nunca são metilados, ou ainda, são metilados apenas transitoriamente permanecem no genoma. Em células somáticas humanas, cerca de 1% do total de bases do DNA são citosinas metiladas, afetando assim, de 70% a 80% do total de dinucleotídeos CpG do genoma (Ehrlich et al. 1982). Classicamente, as CGIs foram definidas como regiões genômicas do DNA de pelo menos 200 pb com alta frequência de dinucleotídeos CpGs (mais de 50% de GC e uma razão maior que 0,6 do número de dinucleotídeos CpGs observados em relação ao número esperado com base na frequência de Cs e Gs dentro do segmento) (Gardiner-Garden and Frommer 1987), que encontram-se em promotores de genes e também em regiões intragênicas (Jones 1999). No entanto, há muita discussão na literatura sobre a exata definição de CGIs regulatórias (Illingworth and Bird 2009). A análise em larga escala dos

cromossomos 21 e 22 mostrou que regiões do DNA maiores do que 500 pb e com conteúdo de CG igual ou maior a 55% e uma razão de CG observada/esperada maior ou igual a 65% foram mais associadas com regiões promotoras de genes (Takai and Jones 2002). A maior parte das CGIs não se encontram metiladas durante todos os estágios do desenvolvimento ou em todos os tipos teciduais (Antequera and Bird 1993; Wang and Leung 2004). A distribuição de DNA metilado no genoma mostra seu enriquecimento em regiões não codificadoras, por exemplo, heterocromatina centromérica e em elementos repetitivos (transposons), mas não nas CGIs de genes ativos (Bird 1986).

Estudos iniciais estimaram que aproximadamente 60% dos genes humanos possuem CGIs localizadas em suas regiões promotoras. Os promotores de mamíferos podem ser separados em duas classes: os promotores enriquecidos em elementos *TATA-box*, que são mais conservados e mais definidos e os promotores ricos em CpGs, que são mais plásticos (Carninci et al. 2006). Existem outros elementos na sequência do DNA que estão associados ao *core* de promotores em mamíferos, como o elemento iniciador (*Inr*), o elemento promotor *downstream* (*DPE*) e o elemento de reconhecimento de TFIIB (*BRE*). Todos esses elementos podem ocorrer simultaneamente em um mesmo promotor, embora determinadas combinações sejam mais comuns que outras. Os elementos *TATA-box* estão associados à promotores fortes, à genes de expressão tecido-específica e frequentemente ocorrem em conjunto com o elemento *Inr* ou sequências *Inr-like* (Sandelin et al. 2007). Promotores que contém CGIs são mais frequentemente associados a genes de expressão ubíqua (Schug et al. 2005), embora haja muitas exceções, incluindo genes expressos especificamente no cérebro (Gustincich et al. 2006). Somente uma fração dos promotores associados a CGIs contém elementos *TATA-like* (Sandelin et al. 2007).

O padrão da metilação do DNA de células somáticas normais é um estado que pode ser herdado ao longo das sucessivas divisões celulares, o que sugere que a metilação possa ser

considerada como um fenômeno de memória celular (Bird 2002). A metilação de manutenção nomeia o processo que reproduz o padrão de metilação do DNA entre as gerações celulares.

Uma vasta alteração no perfil de metilação do DNA pode ser observado durante a embriogênese (Santos et al. 2002). Nos estágios iniciais do desenvolvimento, os níveis de metilação do DNA das células germinativas declinam para aproximadamente 30% do nível típico observado nas células somáticas. O reestabelecimento do padrão básico de metilação é então atingido no estágio de implantação do embrião (Monk et al. 1987), por meio da metilação *de novo* do DNA. A metilação *de novo* de CGIs está envolvida no silenciamento estável da expressão gênica durante o desenvolvimento embrionário, sendo este fato essencial para o desenvolvimento completo do embrião (Li et al. 1992). Esse fenômeno tem sido diretamente implicado em *imprinting* genômico (Reik and Walter 2001) e inativação do cromossomo X (Avner and Heard 2001). O evento da metilação *de novo* em células germinativas ou nos estágios iniciais do desenvolvimento sugere que esta seja particularmente ativa nestes estágios. Há evidências, no entanto, de que a metilação *de novo* possa também ocorrer em células somáticas adultas. Uma fração significativa de todas as CGIs humanas é propensa a sofrer metilação em certos tecidos durante o envelhecimento, ou em estados patológicos, como o câncer (Baylin and Herman 2000).

As DNA-metiltransferases (DNMTs) são as enzimas responsáveis pela transferência do radical metil do SAM para o carbono 5 da citosina. De acordo com o seu substrato genômico preferido as DNMTs de mamíferos são agrupadas em duas classes gerais: as *de novo* e a de manutenção (Jaenisch and Bird 2003). As primeiras, DNMT3a e DNMT3b, são as responsáveis pela metilação de citosinas do DNA previamente não metiladas (Okano et al. 1999). Já a última, DNMT1, é responsável pela manutenção do padrão de metilação durante a replicação do DNA (Bestor 2000). Há ainda duas outras DNMTs; a DNMT2 que possui baixa atividade de DNA-

metiltransferase *in vitro* (Hermann et al. 2003) e *in vivo* (Okano et al. 1998). E a DNMT3L, que é uma proteína relacionada às DNMTs, mas não possui atividade intrínseca de DNAmetiltransferase. No entanto, se faz fisicamente necessária para a modulação da atividade das DNMT3a e DNMT3b (Suetake et al. 2004).

O entendimento dos mecanismos pelo qual as DNMTs envolvidas na metilação *de novo* são recrutadas para as sequências de DNA alvo nos estágios da embriogênese ainda é incompleto, no entanto, sabe-se que em alguns casos o substrato para essas enzimas é o DNA nucleossomal e que as modificações das histonas dentro dos nucleossomos influenciam profundamente a habilidade das DNMTs em induzir a metilação (Ooi et al. 2007). Estudos genômicos em larga escala têm demonstrado que genes contendo CGI em seus promotores e que estão silenciados pelo complexo modificador da cromatina *Polycomb* tornam-se metilados mais frequentemente que outros genes associados à CGIs (Ohm et al. 2007; Schlesinger et al. 2007; Widschwendter et al. 2007; Gal-Yam et al. 2008). Esta evidência sugere que o estado silenciado do gene precede a metilação do DNA. Os conhecimentos mais atuais a cerca do direcionamento da metilação *de novo* do DNA provém de estudos em cultura celular de sistemas modelo. Estes estudos têm sugerido que há pelo menos três principais mecanismos possíveis pelo qual a metilação *de novo* do DNA é direcionada para a sua sequência alvo no DNA. O primeiro mecanismo sugere que as próprias enzimas DNMT3a e DNMT3b possuam sítios específicos de reconhecimento e ligação ao DNA ou à cromatina (Ge et al. 2004). Um segundo mecanismo para direcionamento das DNMTs envolveria o recrutamento de repressores transcricionais ou outros fatores que a direcionaria à região promotora dos genes alvo (Di Croce et al. 2002; Brenner et al. 2005). Por último, um mecanismo de direcionamento da metilação *de novo* envolveria a resposta desencadeada por RNA interferência (RNAi). O silenciamento transcricional mediado por RNAi em plantas frequentemente resulta em metilação *de novo* do gene silenciado (Mette et al. 2000).

Dois estudos independentes têm reportado um similar mecanismo da metilação *de novo* induzida por RNAi em mamíferos que acarretaram em silenciamento gênico (Kawasaki and Taira 2004; Morris et al. 2004). No entanto, há algumas controvérsias nas generalidades destas observações, visto que outros estudos observaram o silenciamento transcricional mediado por RNAi na ausência de metilação do DNA (Svoboda et al. 2004; Ting et al. 2005).

Quando genes associados à CGIs estão ativos, seus promotores são usualmente caracterizados por regiões depletadas de nucleossomos, assim, os níveis de expressão destes genes são controlados por fatores de transcrição (Gal-Yam et al. 2006). Essas regiões promotoras são flanqueadas por nucleossomos que contém a variante de histona H2A.Z (Zilberman et al. 2008) e são marcados pela trimetilação da lisina 4 na histona H3 (H3K4me3). Foi mostrado em camundongos, que a ocorrência da marca H3K4me3 é mantida pela presença da proteína CFP1, também conhecida por *CXXC finger protein 1*. Esta é uma proteína que reconhece regiões ricas em CpGs e recruta a H3K4 metiltransferase (Thomson et al. 2010). O estado não metilado destas CGIs é presumivelmente assegurado pela presença da proteína TET1, as quais são frequentemente encontradas em TSS ativos de promotores com alto conteúdo de CpGs. Já foi mostrado que TET1 converte as 5mC dessas regiões em 5-hidroximetilcitosinas (Williams et al. 2011), mantendo o TSS ativo livre de citosinas metiladas.

Os promotores que contém CGIs podem ser reprimidos por vários mecanismos, podendo ou não ocorrer a metilação do DNA dessas regiões. Por exemplo, genes que codificam proteínas reguladoras do desenvolvimento embrionário, tais como, *MYOD1* ou *PAX6*, são reprimidos pelo complexo modificador da cromatina *Polycomb* tanto em células tronco embrionárias como em células diferenciadas que não expressam esses genes. As CGIs nos promotores desses genes não são metilados, provavelmente devido a presença de nucleossomos com as marcas H3K4me3 e H2A.Z, antagonistas da metilação do DNA.

No entanto, alguns genes reprimidos têm seus promotores associados à CGIs metiladas. Estes promotores são usualmente restritos a genes que possuem seu estado reprimido estável ao longo do tempo e possuem nucleossomos marcados por H3K27me3 (Taberlay et al. 2011). Exemplos incluem genes que sofrem *imprinting*, genes localizados sobre o cromossomo X inativo, genes que são exclusivamente expressos em células da linhagem germinativa e genes envolvidos com a carcinogênese.

Embora muitas das CGIs encontradas no genoma estejam localizadas em promotores de genes, elas também podem existir em regiões intragênicas e intergênicas (Lorincz et al. 2004; Eckhardt et al. 2006; Illingworth et al. 2010). Inicialmente acreditava-se que as CGIs situadas em regiões intragênicas, salvo algumas exceções, permaneciam desmetiladas. No entanto, estudos recentes têm demonstrado que muitas das CGIs intragênicas são metiladas em vários tecidos (Eckhardt *et al.*, 2006), por exemplo, cerca de 34% de todas as CGIs intragênicas são metiladas no cérebro humano (Maunakea et al. 2010). Mais que isso, Maunakea e colaboradores, mostraram que a maioria das CGIs metiladas encontra-se em regiões intrônicas e intergênicas, enquanto que menos de 3% de CGIs metiladas estão nos promotores a 5' de genes (Maunakea et al. 2010). Diversos estudos têm demonstrado uma correlação positiva entre transcrição ativa e metilação do DNA no corpo de genes no genoma de animais e plantas (Hellman and Chess 2007; Cokus et al. 2008; Lister et al. 2009; Feng et al. 2010), no entanto, o papel dessa metilação, que é tecido específica, ainda não está bem claro.

A metilação que ocorre no corpo dos genes tem sido associada ao silenciamento de elementos repetitivos no DNA (Yoder et al. 1997). Foi proposto que a metilação do DNA nestas regiões é responsável pelo bloqueio da iniciação da transcrição de elementos repetitivos, enquanto que ao mesmo tempo, permite a transcrição do gene hospedeiro. Recentemente, estudos em larga escala, têm hipotetizado outros papéis para a metilação do DNA no corpo de genes

(Laurent et al. 2010). Tem sido mostrado que a metilação intragênica do DNA pode controlar a expressão de isoformas alternativas de genes, marcando inícios da transcrição alternativos (Maunakea et al. 2010), e o *splicing* (Shukla et al. 2011). Estudos mostraram que éxons são mais metilados que íntrons, e que há uma transição do grau de metilação do DNA nas fronteiras entre éxons e íntrons, sugerindo um possível papel da metilação intragênica na regulação do *splicing* (Shukla et al. 2011). Reforçando essa ideia, foi mostrado que os nucleossomos, que são sítios onde a metilação do DNA preferencialmente ocorre (Chodavarapu et al. 2010), são mais frequentemente encontrados em éxons quando comparado às regiões intrônicas (Schwartz et al. 2009). Outra função proposta para a metilação do DNA encontrada no corpo de genes, é que ela poderia controlar a expressão de transcritos não codificadores antissensos que, por sua vez, regulariam a expressão do mRNA do gene hospedeiro através do pareamento senso/antissenso (Shenker and Flanagan 2012). Esta ideia é suportada pelo fato de que mais de 90% de regiões intragênicas desmetiladas que sofrem metilação do DNA *de novo* de uma maneira tecido específica durante o desenvolvimento são associadas à marca de histona computacionalmente preditora de TSS, H3K4me3 (Straussman et al. 2009). De fato, já se conhece lncRNAs intrônicos (Seidl et al. 2006; Redrup et al. 2009) e miRNAs oriundos de íntrons (Lopez-Serra and Esteller 2012) que têm sua expressão regulada por metilação do DNA. Outro dado importante a respeito de CGIs intragênicas metiladas é a observação de que estão presentes em genes preferencialmente envolvidos em desenvolvimento, morfogênese e organogênese, sugerindo que estas regiões regulatórias internas representariam um ajuste fino de processos complexos, como o desenvolvimento embrionário (Straussman et al. 2009). Outro aspecto interessante é o fato de que as CGIs intragênicas são alvos preferenciais para a metilação *de novo* durante a oncogênese (Nguyen et al.).

A metilação do DNA está intimamente ligada ao silenciamento transcricional de alguns genes (Boyes, 1991; Siegfried, 1999), e muitos esforços têm sido feitos para desvendar os mecanismos através dos quais estes eventos ocorrem. Estudos iniciais realizados por Lock e colegas mostraram que a metilação do gene *HPRT* sobre o cromossomo X inativo ocorre após o cromossomo ter sido completamente inativado, o que sugere que a metilação do DNA é um evento final nas etapas do silenciamento gênico, funcionando como uma “chave” para reforçar o estado silenciado dos genes ligados ao cromossomo X (Lock et al. 1987). No entanto, há evidências na literatura de que a metilação do DNA também pode ter um papel importante na diferenciação de células tronco hematopoiéticas nas etapas iniciais do silenciamento gênico (Challen et al. 2012). Como já mencionado acima, o processo da metilação *de novo* requer nucleossomos como substrato para a DNMT3A, no entanto, os TSSs ativos são depletados dessas moléculas (Ooi et al. 2007). Além disso, para que a metilação *de novo* ocorra, esses nucleossomos não devem conter a marca H3K4me3, que está associada a genes ativos. Em 2011, You e colaboradores examinaram a cinética do silenciamento dos genes *OCT4* e *NANOG* em células de carcinoma embrionárias durante a diferenciação por ácido retinóico e mostraram que primeiramente um nucleossomo deve estar presente no promotor para então acontecer o recrutamento da DNMT3A e subsequentemente ocorrer a metilação *de novo* do DNA (You et al. 2011).

Uma das propostas para explicar a repressão da expressão gênica mediada por metilação do DNA seria a de que a metilação pudesse diretamente reprimir a transcrição. Segundo esta proposta, ela bloquearia a ligação de ativadores transcricionais à região promotora dos genes evitando a consequente montagem da maquinaria de transcrição (Prendergast and Ziff 1991). Outro modelo incluiria a participação de proteínas ligadoras de metil-CpG (MBPs). Estas proteínas reconheceriam o DNA metilado e recrutariam corepressores para silenciar diretamente

a expressão gênica (Boyes, 1991; Nan, 1998). Um terceiro mecanismo envolveria a repressão da expressão gênica com a alteração da estrutura da cromatina (Buschhausen *et. al.*, 1987). Assim, as DNMTs além de sua atividade catalítica, poderiam exercer uma função não enzimática no silenciamento transcricional. Elas, quando ligadas ao DNA interagiriam bioquimicamente e recrutariam consigo uma série de outras enzimas envolvidas com modificações e remodelamento da cromatina, como as HMTs (Fuks *et. al.*, 2003; Geiman, *et. al.*, 2004) e as HDACs (Bai *et. al.*, 2005).

Consistente com a ideia de que o câncer é uma doença que tem sua origem devido a falhas na expressão de genes relacionados ao programa de desenvolvimento embrionário, os processos epigenéticos têm sido estudados com grande interesse, a fim de se entender o desenvolvimento e a progressão das doenças neoplásicas.

A perda de metilação do DNA em dinucleotídeos CpGs foi a primeira anormalidade epigenética a ser identificada em células tumorais (Feinberg and Vogelstein 1983), sendo mostrado que ela ocorre extensivamente nos genomas destas células e em variados tipo de sequências, que inclui elementos repetitivos, retrotransposons, promotores pobres em CpGs, íntrons e desertos gênicos (Rodriguez et al. 2006). A hipometilação do DNA pode contribuir para a carcinogênese de diferentes maneiras. Primeiramente, a hipometilação do DNA pode levar à instabilidade genômica. A diminuição de DNA metilado em sequências repetitivas pode favorecer a recombinação mitótica dessas regiões, causando deleções, translocações e rearranjos cromossômicos (Eden et al. 2003; Esteller 2008). Por exemplo, muitos cânceres humanos que frequentemente contêm translocações cromossômicas não balanceadas são caracterizados por uma severa hipometilação em sequências pericentroméricas (Yeh et al. 2002). Além disso, a extensiva desmetilação do DNA de sequências centroméricas em células tumorais pode ter um papel importante para a aneuploidia (Xu et al. 1999; Karpf and Matsui 2005). Outro fato é que a

hipometilação de retrotransposons pode resultar em sua ativação e translocação para outras regiões do genoma, provocando rearranjos cromossômicos, e quebra de genes. Por exemplo, hipometilação de retrotransposons L1 (*long interspersed nuclear elements*) é característico de câncer colorretal (Howard et al. 2008). Outra maneira da hipometilação do DNA contribuir para a carcinogênese, é ativando proto-oncogenes. Exemplo de genes que são afetados pela hipometilação inclui *R-Ras*, *cyclin D2* e *maspin* em cânceres gástricos, *MAGE* em melanoma, *HPV16* em câncer cervical, *S100A4* em câncer de cólon e outros vários genes em câncer pancreático (Feinberg and Tycko 2004; Wilson et al. 2007). Além disso, o mecanismo de hipometilação tem sido proposto para a ativação do miRNA let-7a-3, implicado em câncer de cólon e endométrio (Brueckner et al. 2007). Finalmente, a hipometilação pode levar a perda de *imprinting* genômico. Por exemplo, a perda de *imprinting* do gene *IGF2*, um importante fator de crescimento, resulta em aumento do risco de neoplasia colorretal (Cui et al. 2003) e contribui para o desenvolvimento de tumor de Wilms (Feinberg 1999).

Uma série de evidências tem indicado que CGIs podem ser submetidas a uma metilação anormal *de novo* em estados patológicos como o câncer (Keshet et al. 2006). Os genomas de células malignas são caracterizados por hipermetilação em CGIs localizados em regiões promotoras de genes supressores de tumor, mesmo no contexto da massiva hipometilação que ocorre nestas células. A importância desse fenômeno pode ser compreendida considerando-se a frequência com que ela ocorre, a natureza dos genes envolvidos e a similar vantagem seletiva concedida por esta alteração à célula tumoral em relação às mutações gênicas (Baylin and Herman 2000). Muitos genes supressores de tumor envolvidos em cânceres familiares que estão mutados na linhagem germinativa podem ser inativados por hipermetilação de seus promotores em cânceres esporádicos.

A significância da perda da função gênica mediada epigeneticamente em câncer tem sido esclarecida pelo reconhecimento de que este processo precede e parece ser essencial para diversos eventos genéticos que dirigem a progressão tumoral (Baylin and Herman 2000). Esta ligação pode ser de uma forma indireta, estando associada com hipermetilação de promotores de genes chave em estágios iniciais da tumorigênese. Por exemplo, a hipermetilação de *p16INK4A* ocorre em estágios pré-malignos da progressão tumoral (Belinsky et al. 1998; Kiyono et al. 1998; Nuovo et al. 1999). A observação de alterações no padrão de metilação de CGIs localizadas em genes relacionados com tumorigênese, e que ocorrem nos estágios iniciais da transformação maligna, sugere que a caracterização do *status* de metilação do DNA em amostras clínicas possa ser utilizada como ferramenta para diagnóstico precoce do câncer.

O estudo do perfil de metilação do DNA em CGIs de genes supressores de tumores levou a observação de que cada tipo de câncer pode ter seu próprio perfil específico de hipermetilação (Esteller 2008). Esta informação, em combinação com marcas genéticas e citogenéticas pode ser muito útil para a classificação de subtipos tumorais quanto à sua agressividade ou sensibilidade a quimioterápicos.

Além da hipermetilação do DNA levar a uma direta inativação de genes supressores de tumor, ela também pode desencadear uma inativação indireta de outros genes, através do silenciamento de fatores de transcrição e genes de reparo de DNA. A hipermetilação do DNA em promotor de fatores de transcrição como *RUNX3* em câncer esofágico (Long et al. 2007) e *GATA-4* e *GATA-5* em cânceres colorretal e gástricos (Akiyama et al. 2003), levou a inativação de seus genes alvos. O silenciamento de genes de reparo, como *BRCA1* bloqueia o reparo de erros genéticos e capacita as células a acumular outras lesões, culminando na transformação neoplásica (Dobrovic and Simpfendorfer 1997).

#### 1.4 RNAs não codificadores longos (lncRNAs) e regulação da expressão gênica em eucariotos

Os lncRNAs podem regular a expressão gênica de eucariotos, funcionando como “*scaffolds*” e guias para o recrutamento ou coordenação transcricional de complexos remodeladores da cromatina ativadores ou repressores (Wang and Chang 2011). As mudanças na expressão gênica promovidas por lncRNAs podem ocorrer na vizinhança (*cis*) ou em genes distantes de seu *locus* (*trans*). Um dos mais bem estudados mecanismos de regulação por lncRNAs em *cis* é aquele responsável pela inativação do cromossomo X. A inativação do cromossomo X envolve o lncRNA *Xist*, que tem por função recrutar os complexos repressivos *polycomb* PRC1 e PRC2 ao centro de inativação do X (XIC). O complexo PRC2 inclui enzimas modificadoras de histonas, como EZH2, que catalisa a trimetilação da H3K27. O complexo PRC1 se liga ao XIC promovido pela H3K27me3 causando a ubiquitinação de H2A. As modificações da cromatina desencadeadas pelas proteínas dos complexos repressivos *polycomb* culminam com a metilação do DNA ao longo de todo o cromossomo X inativo (Avner and Heard 2001; Heard 2005).

Mecanismos de repressão da atividade transcricional similares têm sido reportados para outros lncRNAs. Por exemplo, o lncRNA intrônico *Air* silencia a transcrição em *cis* de uma região de 400 kb que inclui os genes *Slc22a3*, *Slc22a2* e *Igf2r* no cromossomo paterno. Quando expresso, *Air* acumula-se no promotor de *Slc22a3*, recrutando o complexo modificador de cromatina G9a metiltransferase que metila a lisina 9 da histona H3, causando silenciamento do promotor *Slc22a3* no alelo paterno (Seidl et al. 2006; Nagano et al. 2008). De forma semelhante, o lncRNA *Kcnq1ot1* está associado ao silenciamento bidirecional de cerca de 10 genes silenciados por imprinting no cromossomo paterno (Mancini-Dinardo et al. 2006). O lncRNA *Kcnq1ot1* acumula-se não uniformemente ao longo do locus *Kcnq1* e exerce sua função via

interação com proteínas dos complexos G9a e PRC2 (Pandey et al. 2008; Redrup et al. 2009). De forma interessante, certos genes no locus *Kcnq1* são silenciados por imprinting apenas na placenta, provavelmente, porque o lncRNA *Kcnq1ot1* interage com G9a e PRC2 de maneira linhagem específica, resultando no estabelecimento do estado repressivo destes genes somente em certos tipos celulares (Pandey et al. 2008). Um mecanismo semelhante também ocorre em plantas, sugerindo que a relação entre ncRNA e complexos repressivos da cromatina é um mecanismo de repressão gênica evolutivamente conservado (Wang and Chang 2011). Por exemplo, o lncRNA COOLDAIR (Cold Assisted Intronic Noncoding RNA), em *Arabidopsis thaliana*, é necessário para o estabelecimento e manutenção de um estado repressivo da cromatina durante o período de frio (Heo and Sung 2011). COOLDAIR é transcrito na direção senso de um íntron de seu gene alvo, *FLC*, um forte repressor floral. Este lncRNA desempenha um papel crítico em direcionar o complexo PRC2 para a cromatina na região do gene *FLC* durante a vernalização<sup>2</sup>, promovendo sua repressão gênica através da trimetilação da H3K27 (Heo and Sung 2011). Similarmente, em leveduras, vários ncRNAs antissenso, em numerosos loci gênicos promovem o silenciamento da transcrição de mRNAs afetando os estados de metilação e acetilação da cromatina (Camblong et al. 2007; van Dijk et al. 2011).

Além dos lncRNAs que regulam a expressão gênica em cis, existem também lncRNAs bem caracterizados que agem em trans na regulação da expressão gênica. Vários lincRNAs, por exemplo, são capazes de alterar e regular estados epigenéticos através do recrutamento de complexos modificadores da cromatina, como PRC1 e PRC2, em trans (Khalil et al. 2009; Zhao et al. 2010).

---

<sup>2</sup> Vernalização é o processo pelo qual as plantas são induzidas a florescer através da exposição a temperaturas baixas não congelantes.

Centenas de lncRNAs são expressos sequencialmente ao longo dos eixos de desenvolvimento temporal e espacial dos loci humanos homeobox (Hox), onde definem domínios de cromatina com diferentes padrões de metilações de histona e acessibilidade à RNA polimerase (Rinn et al. 2007). Um destes lncRNAs, o lincRNA HOTAIR (Hox Transcript Antisense RNA), origina-se do locus HOXC e silencia a transcrição ao longo de 40 kb do locus HOXD, em trans, por induzir um estado de cromatina repressivo, que é proposto ocorrer através de recrutamento de PRC2 (Rinn et al. 2007). Adicionalmente, a expressão de HOTAIR tem sido recentemente associada à metástase de mama (Gupta et al. 2010) e ao baixo prognóstico de câncer de pâncreas (Kim et al. 2012). O mecanismo de associação de lncRNAs aos complexos modificadores de cromatina pode resolver o paradoxo de como estes complexos que, frequentemente, possuem domínios de ligação a RNA, mas pouca especificidade de sequência para ligação em DNA, são capazes de se ligar especificamente à cromatina e promover a regulação da expressão gênica através de modificações na sua estrutura (Mercer et al. 2009).

Tem sido sugerido, na literatura, que o processo de transcrição de ncRNAs por si só, pode ser funcional para o controle epigenético de genes relacionados. Esta transcrição induziria uma estrutura de cromatina aberta em promotores de genes codificadores de proteínas, facilitando a acessibilidade de fatores de transcrição e da RNAP II (Dinger et al. 2009; Atkinson et al. 2011). A transcrição de lncRNAs upstream ao locus *fbp1+* de *Schizosaccharomyces pombe* induz remodelamento de cromatina que é crítico para ativação da transcrição de genes codificadores de proteínas localizados downstream (Hirota et al. 2008). Foi observado que a transcrição de ncRNAs inicia-se em múltiplos sítios upstream do promotor de *fbp1+*, induzindo uma região de cromatina aberta que progride progressivamente até o sítio de início da transcrição do mRNA. A inserção de um terminador transcricional na região upstream do mRNA abole a cascata de transcrição dos ncRNAs e promove alteração progressiva da cromatina, resultando em redução do

recrutamento de fatores de transcrição ao promotor de *fbp1+* e mínima indução da transcrição do mRNA (Hirota et al. 2008).

Em alguns casos, a ação do lncRNA pode depender da transcrição do mRNA do locus onde ele atua, para promover a regulação da expressão gênica. O lncRNA DHFR, transcrito a partir de uma região localizada upstream ao locus do gene DHFR, através da formação de uma estrutura tripla RNA-DNA com o promotor desse gene, interage diretamente com o fator de transcrição TFIIB, resultando no rompimento do complexo de pré-iniciação no promotor do mRNA (Martianov et al. 2007).

Transcritos nascentes também podem representar alvos que recrutam fatores proteicos efetores, regulando-os alostericamente. Um exemplo são os lncRNAs associados à região promotora do gene ciclina D1 (Wang et al. 2008). Estes lncRNAs agem cooperativamente recrutando e modulando a atividade da proteína ligadora de RNA TLS (Translocated in Liposarcoma), em resposta à radiação ionizante. A ligação dos lncRNAs induz mudança conformacional na proteína, que por sua vez, inibe a atividade de histona acetiltransferase dos complexos CBP (CREB-Binding Protein) e p300 histone acetyltransferase, silenciando a expressão de cyclin D1 (Wang et al. 2008).

LncRNAs também podem agir como cofatores que modulam a atividade de fatores de transcrição. Em camundongo, o lncRNA *Evf2* é transcrito de um enhancer ultraconservado. Ele recruta o fator de transcrição DLX2 para o seu locus induzindo a expressão do gene adjacente, *Dlx6* (Feng et al. 2006). Recentemente, lncRNAs transcritos a partir de regiões enhancers foram definidos como uma nova classe de ncRNAs, enhancer RNAs ou eRNAs (Kim et al. 2010). Trabalhos mostraram que o nível de expressão destes eRNAs está correlacionado positivamente com o nível de síntese de mRNA dos genes na sua vizinhança, sugerindo que a síntese de eRNA ocorre especificamente em enhancers que estão ativamente engajados na síntese de mRNA (Kim

et al. 2010; Wang et al. 2011). A habilidade de alguns lncRNAs em se ligar e recrutar proteínas ligadoras de RNA para regiões promotoras, representa mais uma expansão do repertório de regulação disponível para o programa transcricional (Wang et al. 2008).

LncRNAs podem reconhecer sequências complementares, propiciando interações altamente específicas que são essenciais na regulação de processos que controlam a expressão gênica, incluindo *splicing* alternativo, edição, transporte, tradução e degradação de mRNAs (Mercer et al. 2009).

Os lncRNAs antissenso podem mascarar elementos chaves para a regulação do *splicing* na sequência do mRNA através da formação de RNA duplexes, produzindo transcritos alternativos. Um exemplo é o Zeb2/Sip1 ncRNA, que é complementar a um sítio de *splicing* 5' de um íntron do mRNA Zeb2/Sip1 (Beltran et al. 2008). O gene Zeb2/Sip1 é um repressor transcricional de E-caderina, cuja expressão é altamente regulada durante a transição epitelial-mesenquimal (EMT). A expressão do ncRNA é induzida após EMT e previne o *splicing* de um íntron que contém um sítio interno de entrada no ribossomo (Birney et al.) requerido para a eficiente tradução e expressão da proteína Zeb2/Sip1 (Beltran et al. 2008). Em adição a este exemplo, existem muitos outros transcritos antissenso não codificadores de proteínas envolvidos na modulação de padrões de *splicing* alternativo do gene ao qual se sobrepõem (Krystal et al. 1990; Munroe and Lazar 1991; Yan et al. 2005).

Outro mecanismo de regulação transcricional exercido por lncRNAs envolve o pareamento específico com miRNAs, inibindo a habilidade do miRNA em interagir com seus mRNAs alvos. Este mecanismo é usado pelo ncRNA IPS1 (Induced by Phosphate Starvation 1) em *Arabidopsis thaliana* (Franco-Zorrilla et al. 2007). O ncRNA IPS1 se liga e sequestra o miRNA miR-399 através de complementaridade de sequência quase perfeita, resultando na expressão aumentada dos genes alvos de miR-399 (Franco-Zorrilla et al. 2007).

Em *C. elegans*, foi documentado outro mecanismo de ação de lncRNAs no controle da expressão gênica. Transcritos não codificadores, como o lncRNA *rncs-1* podem interagir com a enzima Dicer, acarretando na desregulação do processamento de dsRNA envolvidos no mecanismo de RNAi, causando superexpressão de seus genes alvos (Hellwig and Bass 2008). Alternativamente, o anelamento de lncRNAs pode promover ligação de complexos proteicos efetores da degradação do mRNA, de uma maneira análoga ao exercido pelo complexo RISC quando associado à siRNAs e miRNAs (Mercer et al. 2009). RNA duplexes resultantes do anelamento de transcritos complementares, ou mesmo de lncRNAs com hairpins internos podem ser processados em RNAs curtos, aumentando a possibilidade de muitos lncRNAs participarem de vias de silenciamento de RNA (Ogawa et al. 2008).

A expressão de lncRNAs tem profundos efeitos na biologia das células eucarióticas, influenciando negativamente ou positivamente a expressão de genes codificadores de proteínas. Há muitos exemplos, onde sua ação regula transcricionalmente ou pós-transcricionalmente a expressão de mRNAs codificadores de proteínas, no entanto, é muito provável que lncRNAs possam desempenhar papéis adicionais (Wilusz et al. 2009).

## 2 OBJETIVOS

### 2.1 Objetivos gerais

Investigar a expressão, regulação e possíveis papéis biológicos de RNAs não codificadores longos (lncRNAs) gerados em regiões intrônicas de *loci* gênicos codificadores de proteínas em linhagens celulares humanas.

Investigar a contribuição da metilação do DNA no controle na biogênese de lncRNAs intrônicos.

### 2.2 Objetivos específicos

1. Gerar perfis de expressão de lncRNAs intrônicos senso e antissenso nas linhagens celulares tumorais humanas de próstata (DU-145), mama (MCF-7) e pâncreas (Mia PaCa 2) utilizando uma plataforma de *oligoarray* customizada.
2. Avaliar propriedades globais dos lncRNAs intrônicos, tais como a abundância relativa em tecidos humanos, conservação evolutiva e estrutural, e presença de elementos regulatórios característicos de genes transcritos pela RNA Polimerase II através de meta-análises com dados disponíveis publicamente.
3. Investigar associações *in cis* ou *in trans* entre a expressão de lncRNAs intrônicos e mRNAs que evidenciem possíveis papéis funcionais.
4. Avaliar o efeito do agente desmetilante do DNA 5-Aza-2'-deoxicitidina (5-AZA) sobre a expressão de lncRNAs intrônicos.
5. Caracterizar o *status* de metilação e a atividade promotora de regiões do DNA que controlam a transcrição de lncRNA intrônicos.

### 3 MATERIAIS E MÉTODOS

#### 3.1 Linhagens celulares humanas e condições gerais de cultivo

As células de linhagens humanas utilizadas neste trabalho foram adquiridas da ATCC (*American Type Culture Collection*) e cultivadas a 37°C em atmosfera de 5% de CO<sub>2</sub> em meio de cultura apropriado (Tabela 1). Os meios utilizados foram DMEM (*Dulbecco's Modified Eagle Medium*) (Cultilab Materiais para Cultura de Células) e RPMI 1640 (*Roswell Park Memorial Institute*) (Cultilab Materiais para Cultura de Células) suplementados com 10% (vol/vol) de soro fetal bovino (SFB) (Cultilab Materiais para Cultura de Células), 4,5 g/L de glicose, 1 mM de piruvato de sódio, 1,5 g/L de bicarbonato de sódio e antibióticos estreptomicina (100 ug/mL) e penicilina (100 U/mL). Para a propagação das linhagens em cultura, as células ao atingirem cerca de 90% de confluência, foram lavadas 2 vezes com solução salina PBS estéril (*Phosphate Buffered Saline*) (NaCl 140 mM; KCl 2,7 mM; Na<sub>2</sub>HPO<sub>4</sub> 8 mM; KH<sub>2</sub>PO<sub>4</sub> 1,5 mM; pH 7,2) e subcultivadas com solução de tripsina/EDTA (2,5 g/L) (Vitrocell). Os estoques celulares foram mantidos nos respectivos meios de cultivo acrescidos com DMSO 10% estéril em reservatório contendo nitrogênio líquido.

**Tabela 1: Linhagens celulares e meios de cultura utilizados neste trabalho.**

Linhagem Celular	Origem	Meio de Cultivo
DU 145 - ATCC HTB-81™	Carcinoma de próstata metastático	DMEM
MCF 7 - ATCC HTB-22™	Adenocarcinoma de mama metastático	DMEM
Mia PaCa 2 - ATCC CRL-1420™	Carcinoma de pâncreas	DMEM
HEK 293 - ATCC CRL-1573™	Rim embrionário	DMEM
LNCaP - ATCC CRL-1740™	Carcinoma de próstata metastático	RPMI 1640 + 1,2 g/L de HEPES

### 3.2 Tratamento das linhagens celulares com o agente desmetilante 5-Aza-2'-deoxicitidina (5-AZA)

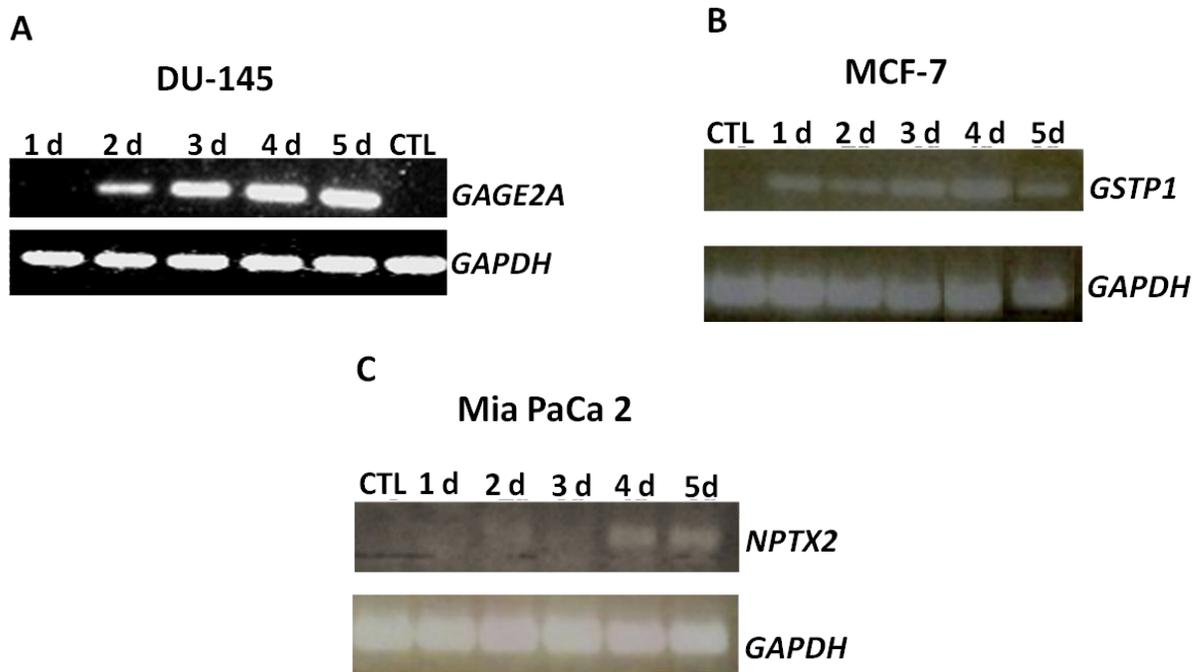
Para a análise da reexpressão de genes silenciados por metilação do DNA, cerca de  $2 \times 10^6$  células foram tratadas com 20 mL de meio DMEM acrescido de 5-Aza-2'-deoxicitidina (Sigma) na concentração de  $5 \mu\text{M}$  em placas de  $147,8 \text{ cm}^2$ . Esta droga é um análogo de citidina que se incorpora ao DNA durante a sua replicação, inibindo de forma irreversível as DNA-metiltransferases e conseqüentemente a metilação do DNA das células-filhas. Como controle do tratamento, células foram incubadas sob as mesmas condições e pelo mesmo período de tempo com o meio de cultura sem a adição de 5-AZA.

A condição experimental tempo de exposição (1 a 5 dias) com o agente desmetilante 5-AZA foi padronizada analisando-se o perfil de reexpressão de genes descritos na literatura como silenciados devido à hipermetilação do DNA de suas regiões promotoras nas linhagens DU-145, MCF-7 e Mia PaCa 2 (Jhaveri e Morrow, 1998; De Backer *et al.*, 1999; Sato *et al.*, 2003). Os genes escolhidos para padronização foram *GAGE 2A* para a linhagem de próstata, *GSTP1* para a linhagem de mama e o gene *NPTX2* para a linhagem de pâncreas. Nas três linhagens, o perfil de reexpressão desses genes foi avaliado realizando a síntese de cDNAs com o kit *SuperScript® III First-Strand Synthesis SuperMix* (Invitrogen) seguido de PCR semi-quantitativo (25 ciclos). O

cDNA foi produzido a partir de 1 µg de RNA total extraído das células tratadas com o agente desmetilante e das células controle e 5 µM do *primer* oligo-dT. Após 3 minutos a 85 °C para desnaturação das estruturas secundárias dos RNAs, é adicionado à reação 2 µL do mix de enzimas *SuperScript III/RNaseOUT* em tampão *First-Strand Reaction Mix*, que contém 10 mM de MgCl<sub>2</sub> e 1 mM de cada dNTP. A transcrição reversa foi realizada a 50°C por 50 minutos. Em seguida, a reação é finalizada, incubando-a por 5 minutos a 85 °C para inativação das enzimas.

A reação de PCR foi realizada com quantidades equivalentes a 8 ng de RNA reversamente transcrito (utilizando uma estimativa de 2% de RNA reversamente transcrito na população de RNA total), 5 unidades da enzima *Go Taq DNA Polymerase* (Promega), 0,2 mM de dNTPs e 0,2 µM de cada *primer* específico em tampão recomendado pelo fabricante, que contém 1,5 mM de MgCl<sub>2</sub>. Em seguida, os produtos da PCR foram analisados por eletroforese em gel de agarose 1% em TAE com 0,5 µg/mL de brometo de etídio (Figura 1). Como controle da reexpressão após tratamento com 5-AZA foi utilizado o gene *housekeeping GAPDH*. Os *primers* utilizados para as reações de PCR (

Tabela 2) foram desenhados com os parâmetros padrões do programa *Primer3* ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)).



**Figura 1: Padronização do tempo de tratamento das linhagens celulares DU-145, MCF-7 e Mia PaCa 2 com o agente desmetilante 5-AZA.** Células das três linhagens celulares foram mantidas em cultura com 5  $\mu$ M de 5-AZA por 1 a 5 dias (1d a 5d). Os seus RNAs totais foram extraídos e usados como molde para a síntese de cDNA. O cDNA foi amplificado por 25 ciclos com pares de *primers* específicos para os transcritos (A) *GAGE2A*, (B) *GSTP1* e (C) *NPTX2* nas linhagens DU-145, MCF-7 e Mia PaCa 2, respectivamente. O transcrito do gene de expressão constitutiva *GAPDH* foi usado como controle.

**Tabela 2: Sequências dos *primers* desenhados para avaliação do efeito do tratamento das células das linhagens DU-145, MCF-7 e Mia PaCa 2 com o agente desmetilante 5-AZA.**

Nome do <i>primer</i>	Sequência de oligonucleotídeos (5' -> 3')*
GAGE 2	F: GACCAAGACGCTACGTAG
	R: CCATCAGGACCATCTTCA
GSTP1	F: ATGACTATGTGAAGGCACTG
	R: AGGTTACGTACTIONCAGGGGA
NPTX2	F: CATCGAGCTGCTCATCAAC
	R: CTGCTCTTGCCAAGGATC
GAPDH	F: TGCACCACCAACTGCTTAGC
	R: GGCACTGGACTGTGGTCATGAG

\* F = Forward; R = Reverse

Após esta etapa de padronização, ficou estabelecido o tempo de 3 dias de tratamento para a linhagem DU-145 e 4 dias de tratamento para as linhagens MCF-7 e Mia PaCa 2.

Após cultivo e tratamento pelo tempo padronizado, as células foram lavadas com PBS e descoladas da superfície da placa com o auxílio de um *cell scraper* em PBS. Em seguida, as células foram sedimentadas por centrifugação a 2.000 rpm por 10 min e o sobrenadante foi removido. Os sedimentos foram utilizados imediatamente ou mantidos a -80 °C até as extrações de RNA e/ou DNA e/ou proteínas.

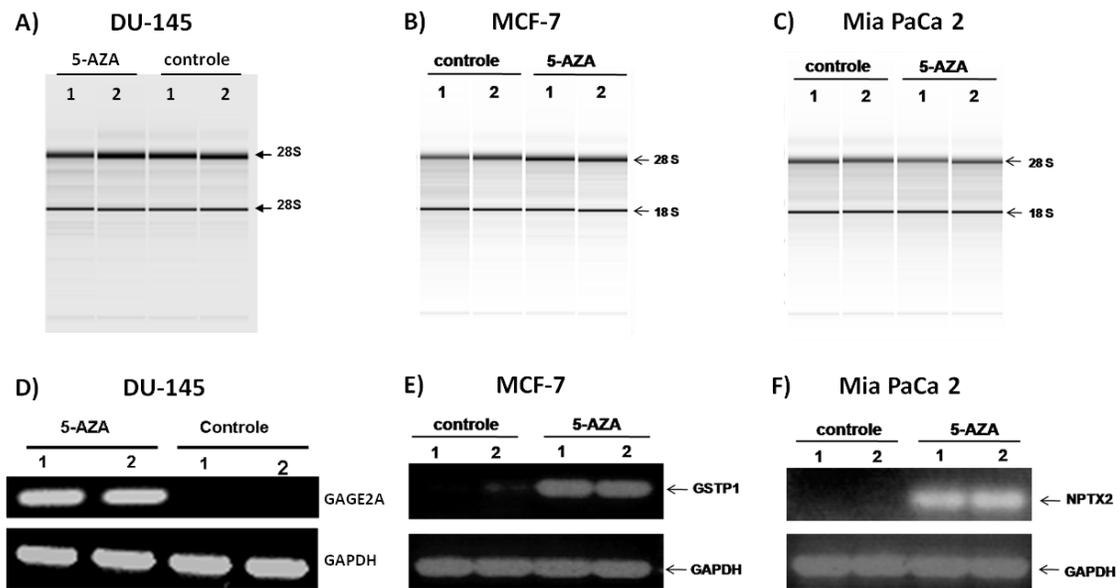
### 3.3 Extração e purificação de RNA total

A extração do RNA foi realizada utilizando-se o reagente *Trizol* (Invitrogen) na proporção de  $10^7$  células para cada 1 mL. Todo o procedimento seguiu o protocolo padrão do fabricante. O *Trizol* é uma solução ácida, monofásica de fenol e isotiocianato de guanidina que possibilita a extração de RNAs através dos mesmos princípios do método originalmente desenvolvido por Chomczynski e Sacchi (Chomczynski e Sacchi, 1987). Após extração, as amostras de RNAs foram purificadas utilizando-se o kit *Illustra RNAspin Mini RNA Isolation* (GE Healthcare). Esta é uma importante etapa do protocolo de extração de RNAs de boa qualidade e livres de contaminação com DNA genômico, pois inclui uma incubação por 1 hora a temperatura ambiente com DNase I (GE Healthcare) enquanto as amostras estão ligadas à membrana de sílica nas colunas. Após a purificação, as amostras de RNAs extraídas foram eluídas com 60 µL de água tratada com DEPC (Ambion) e armazenadas a -80°C. Para certificar a ausência de DNA genômico nas amostras de RNA, foi realizada uma reação de PCR (40 ciclos) com 5 unidades da enzima *Go Taq DNA Polymerase* (Promega), 0,2 mM de dNTPs, 0,2 µM de *primers* para o gene *GAPDH* (Tabela 2) e 1µg de RNA total de cada amostra (sem transcrição reversa) como molde para a reação. Em seguida, os produtos da PCR foram analisados por eletroforese em gel de agarose 1% em TAE com 0,5 µg/mL de brometo de etídio. A ausência de banda indica um RNA extraído sem contaminação com DNA genômico.

### 3.4 Avaliação do rendimento, pureza e integridade do RNA obtido

O RNA total isolado foi quantificado pela medida óptica de absorbância a 260 nm no espectrofotômetro *NanoDrop-1000* (Thermo Scientific). A presença de proteínas e compostos fenólicos contaminantes foi avaliada no mesmo equipamento através da determinação das razões de absorbância nos comprimentos de onda de 260/280 nm e 260/230 nm, respectivamente, sendo consideradas amostras puras as que obtiveram razões 260/280 nm próximas de 2,0 e 260/230 entre 1,8 e 2,2.

A qualidade dos RNAs extraídos foi avaliada utilizando-se o equipamento *BioAnalyser 2100* (Agilent Technologies). Esse equipamento realiza uma eletroforese capilar de alta tensão com as amostras de RNA, o que resulta na obtenção de perfis eletroforéticos que quando comparados ao perfil do padrão (*RNA 6000 ladder*) torna possível uma comparação das intensidades das bandas correspondentes aos RNAs ribossomais 28S e 18S, os mais abundantes nas células eucarióticas. A observação de razões 28S/18S próximas de 2 para as amostras confirmou a boa qualidade do RNA disponível para os experimentos planejados (Figura 2 A, B e C)



**Figura 2: Qualidade dos RNAs e validação do tratamento de células com agente desmetilante 5-AZA.** Qualidade das extrações e purificações dos RNAs totais de (A) DU-145, (B) MCF-7 e (C) Mia PaCa 2. Aliquotas de aproximadamente 200 ng foram examinadas por eletroforese capilar no equipamento *BioAnalyzer* (Agilent Technologies). As bandas nas alturas de 4.000 nt e 2.000 nt indicam, respectivamente, a integridade das subunidades 28S e 18S do RNA ribossomal. A reexpressão dos genes (D) *GAGE2A* em DU-145, (E) *GSTP1* em MCF-7 e (F) *NPTX2* em Mia PaCa 2, após o tratamento com 5-AZA indicam a eficiência do tratamento realizado nas células. 1 e 2: representam réplicas biológicas.

Como podemos observar na figura acima, as amostras de RNA total providas de células tratadas com 5-AZA apresentaram boa qualidade. Assim, essas amostras foram amplificadas, marcadas com os fluoróforos Cy5 e Cy3 e hibridizadas nas lâminas de *oligoarray* como descrito na subseção 3.5.3 da seção Material e Métodos.

### 3.5 Análise da expressão gênica em larga escala

#### 3.5.1 Lâmina de oligonucleotídeos com 44 mil sondas (oligoarray) enriquecida em transcritos não codificadores

Para avaliação do efeito da desmetilação do DNA induzido pelo tratamento com o agente desmetilante 5-AZA na expressão de transcritos não codificadores de proteínas e de mRNAs codificadores de proteínas em larga escala foi utilizada uma lâmina de *oligoarray* construída com a tecnologia de síntese de oligonucleotídeos *in situ* (tecnologia *SurePrint*) (Agilent Technologies). A lâmina de oligonucleotídeos (*oligoarray*) utilizada é composta por 31.063 sondas de 60-mer, desenhadas por nosso grupo, seguindo os critérios recomendados pelo fabricante (Agilent Technologies) e descritos por Hughes e colaboradores (2001) (Hughes *et al.*, 2001). Além destas sondas, o *oligoarray* contém controles positivos e negativos desenhados pelo fabricante para avaliação da qualidade das hibridizações. Os transcritos representados no *oligoarray* foram definidos após extensiva análise de mapeamento de sequências expressas no genoma humano (Nakaya *et al.*, 2007). Resumidamente, todas as sequências de ESTs, mRNAs e RefSeqs humanas depositadas no GenBank e mapeadas na montagem de maio de 2004 do genoma humano (hg17), disponível no *Genome Browser* (<http://genome.ucsc.edu>), foram filtradas e as sequências expressas com sobreposição foram agrupadas em *clusters* cujas coordenadas foram mapeadas no genoma. Os *clusters* foram definidos como “transcritos” e classificados quanto à localização e orientação relativa a genes anotados. A lista com as sequências de todos os oligos presentes no *oligoarray* que utilizamos estão depositadas no banco público do NCBI (<http://www.ncbi.nlm.nih.gov/geo/>) sob o número de acesso GPL9193. Os transcritos não codificadores sondados com a lâmina são longos (>200nt) (Nakaya *et al.*, 2007) e divididos em duas classes: os lncRNAs intrônico antissenso, cuja expressão ocorre a partir da fita

do DNA oposta à fita que transcreve o mRNA codificador de proteínas de seu *locus* hospedeiro, podendo ou não sobrepor parcialmente a um éxon do mRNA e os lncRNAs senso, cuja expressão ocorre a partir de regiões totalmente intrônicas, na mesma fita do DNA dos quais são transcritos os mRNAs do *locus* hospedeiro.

A lâmina é composta por 10.901 sondas que medem a expressão de lncRNAs antissenso e 6.831 sondas que medem a expressão de lncRNAs senso; além de 13.274 sondas que medem mRNAs de genes codificadores de proteínas. Todas estas sondas foram reanotadas em abril de 2012 usando a montagem mais recente do genoma humano (hg19).

### **3.5.2 Obtenção dos alvos fluorescentes**

A amplificação e marcação dos RNAs com fluoróforos foi realizada utilizando o kit de amplificação *Low RNA Input Fluorescent Linear Amplification Kit* (Agilent Technologies), de acordo com as instruções do fabricante (Figura 3). Este protocolo garante a manutenção da representação relativa das mensagens do transcriptoma mesmo após amplificação do RNA. De modo geral, cada amostra de 500 ng de RNA total das células DU-145, MCF-7 e Mia PaCa 2 tratadas ou não (controle) com 5-AZA foi amplificada e marcada separadamente em duplicata com Cy3 ou Cy5. O método usa um *primer* oligo dT acoplado ao sítio de reconhecimento da enzima T7 RNA polimerase para a produção de cDNAs a partir de RNAs poliadenilados presentes nas amostras, assim como de RNAs que têm na sua sequência regiões ricas em adenosina. Após a produção do cDNA, é adicionada à reação a enzima T7 RNA polimerase, que a partir de seu sítio de reconhecimento presente nas moléculas de cDNAs transcreve o material alvo, gerando o cRNA (RNA complementar), que possui em sua sequência CTPs marcados com cianina 3 (Cy3) ou cianina 5 (Cy5). Como controle da amplificação e marcação, é adicionado à reação uma mistura de 10 diferentes RNAs poliadenilados de adenovírus sintetizados *in vitro*

(*Spike mix*), com diferentes proporções conhecidas. O *oligoarray* construído pela Agilent possui sondas para estes RNAs controle.

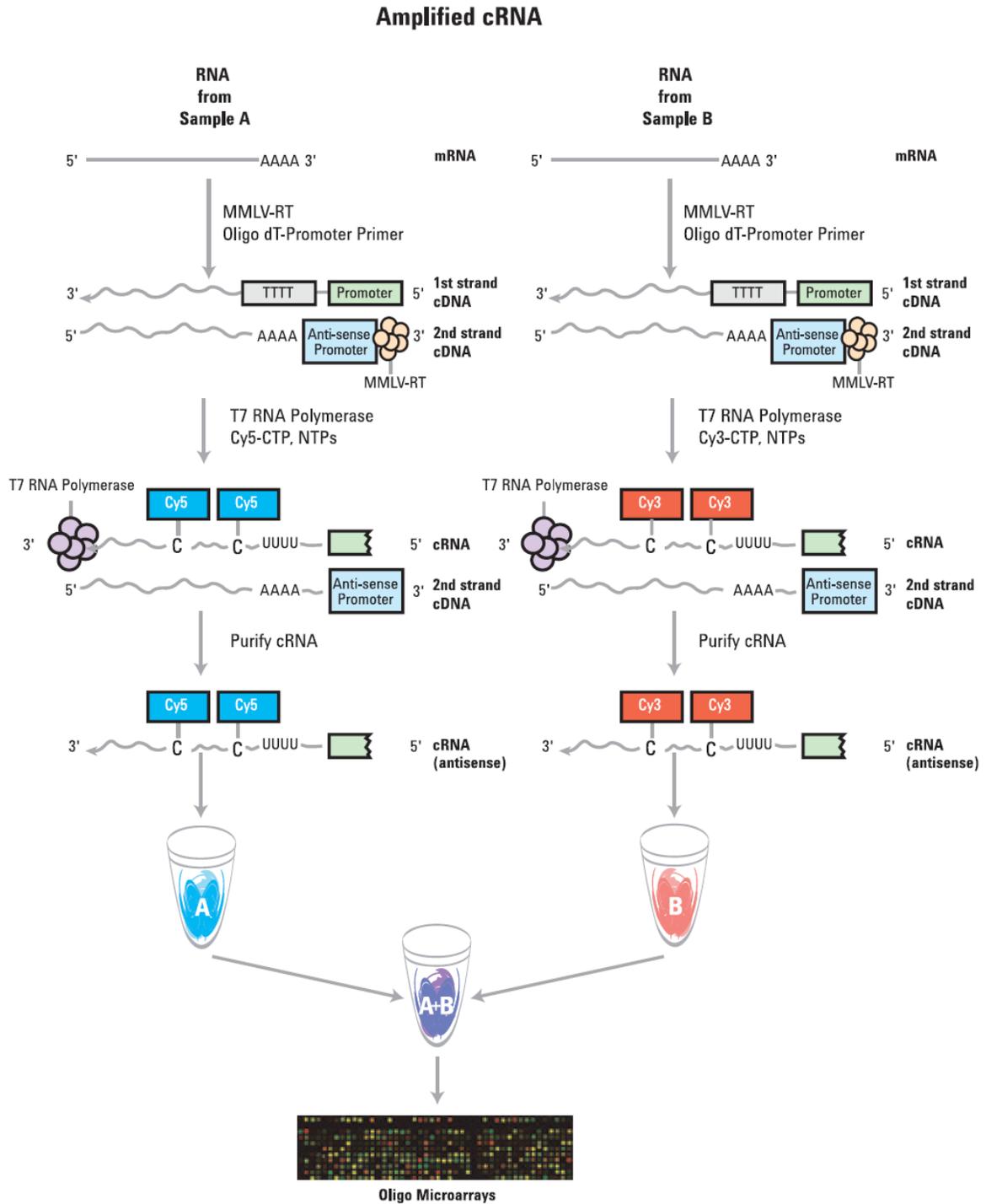


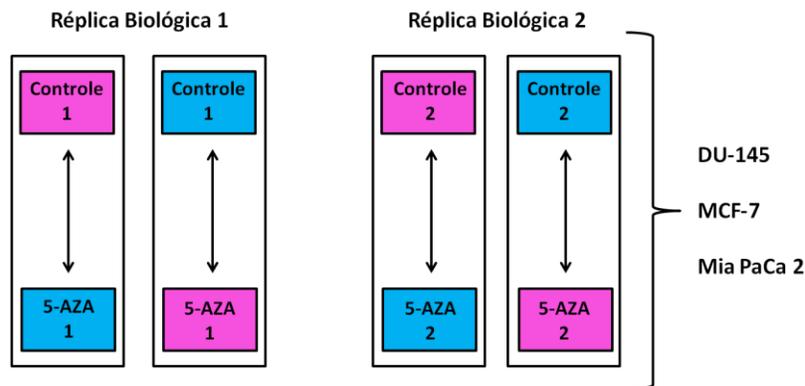
Figura 3: Método de amplificação de cRNA e marcação com fluoróforos (Agilent Technologies).

O cRNA obtido na reação de amplificação e marcação foi purificado com o *RNeasy Mini Kit* (Qiagen) de acordo com o protocolo do fabricante. A incorporação do fluoróforo é avaliada medindo-se sua absorvância no espectrofotômetro NanoDrop-1000 (NanoDrop Technologies) a 570 nm (Cy3) e 670 nm (Cy5). Como recomendado pelo fabricante, somente cRNAs amplificados com uma incorporação mínima de 8 pmol de fluoróforo por micrograma de cRNA foram usados nas hibridizações.

### 3.5.3 Desenho experimental e hibridização

Como estratégia de hibridização, escolhemos a comparação direta entre os perfis de expressão de células tratadas com o agente desmetilante e seus respectivos controles.

Para cada linhagem foram realizadas quatro hibridizações, sendo duas réplicas biológicas (tratamentos com 5-AZA independentes) e duas réplicas técnicas (marcações independentes) (Figura 4).



**Figura 4: Desenho experimental para as hibridizações.** Amostras de RNAs obtidas de duas réplicas biológicas (1 e 2) de cada linhagem estudada (DU-145, MCF-7 e Mia PaCa 2) tratadas independentemente com o agente desmetilante (5-AZA) e não tratadas (Controle) foram marcadas com Cy5 (azul) e Cy3 (rosa). Setas indicam o par de amostras hibridizado em cada *oligoarray*. Foi realizada uma réplica técnica para cada réplica biológica invertendo-se os fluoróforos, totalizando quatro medidas de intensidade de expressão para cada sonda em cada linhagem.

Quantidades iguais (825 ng) de cRNA de cada amostra tratada e controle marcadas com Cy5 ou Cy3 foram misturadas e hibridizadas nos *oligoarrays*. Para compor réplicas técnicas, os pares de cRNAs tratado e controle das réplicas biológicas foram marcados com os fluoróforos invertidos.

Os alvos fluorescentes foram hibridizados nos *oligoarrays* seguindo o protocolo de hibridização *Agilent Gene Expression Hybridization Kit* (Agilent Technologies) por 17 horas a 65 °C. A seguir, as lâminas foram lavadas com os tampões recomendados pelo fabricante e escaneadas com o *scanner GenePix 4000B* (Molecular Devices Corporation, Sunnyvale, CA) afim de se obter os valores de intensidade de cada alvo hibridizado no *oligoarray*. As intensidades de fluorescência foram extraídas usando o programa *Agilent Feature Extraction* (Agilent Technologies). O programa calcula a intensidade de sinal de cada sonda subtraindo a intensidade de fundo local (*background*). Em seguida, normaliza a intensidade do conjunto de sondas da lâmina, utilizando o método LOWESS (*Local Weighted Scatterplot Smoothing*), que procura compensar a diferença de sensibilidade inerente aos fluoróforos Cy5 e Cy3 (Quackenbush, 2001).

A primeira etapa no processamento dos dados de expressão foi a filtragem das sondas detectadas nos experimentos para cada linhagem. Para tanto, utilizou-se como critério de inclusão na análise, sondas que medem seus alvos com valores de intensidade estatisticamente acima do *background* (Teste T; *p-value* < 0,05) em três das quatro réplicas no grupo tratado ou no grupo controle.

### 3.5.4 Avaliação do potencial codificador de lncRNAs intrônicos detectados nas linhagens DU-145, MCF-7 e Mia PaCa 2

O programa CPC (*Coding Potential Calculator*) (Kong *et al.*, 2007) calcula o potencial codificador de transcritos analisando características intrínsecas de suas sequências. Primeiramente, procura por predição de ORFs (*Open Reading Frames*) nas três fases de leitura possíveis, utilizando o *framefinder software*. Havendo identificação de alguma ORF, o programa analisa a sua qualidade e calcula um *score* (LOG-ODDS SCORE) levando em consideração o tamanho, a qualidade e a integridade<sup>3</sup> da ORF. O programa realiza também um BLASTX para verificar se a ORF predita possui alguma similaridade com proteínas depositadas no banco de dados do *UniRef 90 (UniProt Reference Clusters)*. Se a ORF predita possuir pelo menos 90% de identidade com alguma proteína conhecida calcula-se um HIT SCORE, considerando o número de *hits* encontrados pelo BLASTX e a qualidade desses *hits* (*E-value*). Uma vez que é possível encontrar ORFs preditas em transcritos não codificadores ao acaso, e que quando isto ocorre os *hits* tendem a estar aleatoriamente nas três fases de leitura, o programa calcula um *score* denominado FRAME SCORE que avalia a distribuição dos *hits* encontrados pelo BLASTX nas três fases de leitura possíveis. O programa CPC calcula uma pontuação denominada CODING POTENTIAL SCORE que leva em consideração todos os cálculos acima. Desta forma, transcritos preditos como não codificadores possuem uma pontuação menor que zero e os codificadores de proteína possuem uma pontuação acima de zero. A pontuação que fica no intervalo de [-1; 0] e [0; 1] é predita como potencial não codificador fraco e potencial codificador fraco, respectivamente. Logo, quanto mais longe o CODING POTENTIAL SCORE calculado estiver do zero, maior é a confiabilidade na predição do potencial codificador/não codificador.

---

<sup>3</sup> Uma ORF predita íntegra significa que ela começa com um *start codon* e termina com um *stop codon* na mesma fase de leitura.

### 3.5.5 Identificação de lncRNAs intrônicos em bibliotecas de RNA-seq

Foram utilizados neste trabalho dados de RNA-seq disponíveis publicamente no UCSC *Genome Browser* (<http://genome.ucsc.edu>): Burge RNA-seq (Wang *et al.*, 2008) e ENCODE *Caltech RNA-seq Libraries* (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=268692585&c=chr1&g=wgEncodeCaltechRnaSeq>).

RNA-seq é um método de análise do transcriptoma em larga escala que usa tecnologias de “*deep-sequencing*” para o sequenciamento dos RNAs expressos em uma determinada célula ou tecido. Esta técnica é especialmente útil para a obtenção de informação a respeito de padrões de *splicing* de RNAs e, dependendo de sua cobertura, para a determinação da presença ou ausência de classes de RNAs de baixa abundância, como os RNAs não codificadores de proteínas (Mortazavi *et al.*, 2008). De um modo geral, o método consiste na produção de uma biblioteca de cDNAs contendo adaptadores ligados a uma ou a ambas as suas pontas, a partir de uma população de RNAs (total ou fracionada, tal como poli(A)+, por exemplo). Em seguida, essas moléculas são sequenciadas em larga escala, obtendo-se sequências curtas, entre 30 – 400 pb, dependendo da tecnologia usada para o sequenciamento. As sequências resultantes são então alinhadas ao genoma ou transcriptoma de referência ou montadas “*de novo*” para a produção de um mapa do transcriptoma, que consiste da estrutura transcricional (presença de íntrons e éxons em genes, regiões de *splicing*, isoformas gênicas, variações de sequência em regiões transcritas, por exemplo) e/ou dos níveis de expressão gênica.

As bibliotecas de RNA-seq do projeto ENCODE *Caltech RNA-seq Libraries* utilizadas foram geradas a partir de RNA poli(A)+ obtidos de células das linhagens: GM12878, linhagem celular linfoblásticoide; H1-hESC, célula-tronco embrionária; HeLa-S3, carcinoma cervical; HepG2, carcinoma de fígado, HUVEC, células endoteliais da veia umbilical humana; K562,

leucemia mielóide crônica e NHEK, queratinócitos epidermais. Elas foram geradas preservando a informação da orientação original dos transcritos, sequenciadas com a tecnologia de sequenciamento em larga escala *Illumina Genome Analyser* (Illumina) e mapeadas na montagem hg19 do genoma humano.

Os arquivos disponibilizados no UCSC *Genome Browser* apresentavam os dados em formato bigWig, que foram transformados em formato wiggle para as análises. O formato de arquivo wiggle exhibe os *reads* sequenciados de forma gráfica, como picos de densidade de *reads* ao longo das coordenadas do genoma. A partir desses dados não é possível estimar o nível de expressão dos transcritos, uma vez que a informação do número de *reads* detectados que compõem os picos sobre a coordenada não está disponível. Assim, com essas bibliotecas foram realizadas apenas análises de identificação dos lncRNAs detectados em nossas linhagens em estudo. Isto foi feito a partir dos cruzamentos das coordenadas genômicas de nossos lncRNAs com as coordenadas das regiões expressas nas sete bibliotecas de RNA-seq.

Avaliamos a cobertura de sequenciamento para regiões intrônicas e exônicas das bibliotecas do projeto ENCODE *Caltech RNA-seq Libraries* a fim de se verificar se estas possuem algum viés de cobertura. A análise da cobertura de sequenciamento foi realizada computando-se o nível de saturação das regiões intrônicas e exônicas ao se amostrar aleatoriamente *tags* sequenciadas a partir das sete bibliotecas. As *tags* amostradas foram mapeadas no genoma humano em relação aos Gene IDs de genes, levando-se em consideração a sua localização exônica ou intrônica. Após cada rodada de sorteio o número de novos Gene IDs mapeados e sua localização (região intrônica ou exônica) foram computados para a construção da curva de saturação.

As bibliotecas de RNA-seq Burge (Wang *et al.*, 2008) foram geradas a partir de RNA poli(A)+ de nove tecidos de doadores humanos saudáveis. O sequenciamento em larga escala foi

realizado com o sequenciador *Illumina Genome Analyzer* (Illumina). A Tabela 3 apresenta o número de *reads* mapeados na montagem hg19 do genoma humano.

**Tabela 3: Número de *reads* detectados nas bibliotecas do projeto RNA-seq Burge mapeados na montagem hg19 do genoma humano.**

Tecido	Número de <i>reads</i> mapeados na montagem hg19 do genoma humano
Adiposo	17.635.339
Cérebro	10.962.620
Mama	10.557.297
Cólon	17.817.798
Coração	10.630.949
Fígado	11.571.688
Linfonodo	16.221.077
Músculo esquelético	14.212.800
Testículo	18.856.562

A coordenadas genômicas dos *reads* obtidos de cada uma das nove bibliotecas do projeto RNA-seq Burge foram cruzadas com as coordenadas genômicas dos lncRNAs detectados em pelo menos uma das três linhagens, DU-145, MCF-7 e Mia PaCa 2. A seguir, o número de *reads* por quilobase por milhão (RPKM) foi calculado como definido por Mortazavi e colaboradores (Mortazavi *et al.*, 2008):

$$RPKM = \frac{N^{\circ} \text{ reads}}{N^{\circ} \text{ total de reads da biblioteca} \times \text{tamanho do transcrito}} \times 10^9$$

A expressão dos transcritos detectados em pelo menos uma das linhagens em estudo (DU-145, MCF-7 e Mia PaCa 2) e identificados nas bibliotecas do projeto RNA-seq Burge foi

apresentado como uma fração da expressão total do respectivo transcrito entre as nove bibliotecas estudadas (F.E.T.). A abordagem desta análise foi realizada de acordo com os métodos publicados por Cabili e colegas (Cabili *et al.*, 2011):

$$V' = \frac{\log_2 (V + 1)}{\sum_{i=1}^n \log_2 (v_i + 1)}$$

$V'$  corresponde à fração de expressão e  $V = (v_1, \dots, v_n)$  corresponde a estimativa da abundância do transcrito calculado anteriormente e expresso pelo valor de RPKM.

### **3.5.6 Comparação de intensidade de expressão entre lncRNAs intrônicos e mRNAs codificadores de proteínas**

Para a comparação da intensidade de expressão entre os lncRNAs e mRNAs codificadores de proteínas detectados em pelo menos uma das linhagens DU-145, MCF-7 e Mia PaCa 2 foi necessário normalizar os valores de intensidade medidos por cada sonda válida nas análises individuais de cada experimento entre todos experimentos. O método de normalização realizado foi normalização por quantil (Bolstad *et al.*, 2003). Esta normalização ordena os valores de intensidade de expressão medidos pelas sondas válidas de cada experimento independentemente. Para cada posição do ranque, é calculada uma média da intensidade de expressão entre os experimentos. O valor de intensidade médio calculado para o ranque substitui cada valor original de intensidade de expressão. Esta normalização foi feita utilizando a plataforma R (Team, 2011).

Os agrupamentos dos transcritos detectados entre as linhagens foi realizado utilizando o programa *SpotFire* (SpotFire Inc.).

### 3.5.7 Correlação de expressão entre lncRNAs intrônicos e mRNAs detectados

As análises de correlação de expressão entre os lncRNAs intrônicos e os mRNAs codificadores de proteínas expressos a partir dos mesmos *loci* gênicos (correlação em *cis*) e de correlação de expressão entre lncRNAs intrônicos e mRNAs codificadores de proteínas expressos a partir de *loci* gênicos diferentes (correlação em *trans*) foi realizada para o conjunto de transcritos cuja expressão pôde ser detectada nas três linhagens. Esse critério foi utilizado com a finalidade de se obter uma melhor confiabilidade na interpretação dos resultados, uma vez que as correlações observadas ou não devem ocorrer nas três linhagens. Os coeficientes de correlação de *Pearson* foram calculados a partir dos dados normalizados por quantil, como descrito na subseção 3.5.4.3. da seção Materiais e Métodos.

As análises de enriquecimento em termos de *Gene Ontology (GO)* (Ashburner *et al.*, 2000) para os mRNAs cuja expressão se correlaciona com a expressão de lncRNAs oriundos dos mesmos *loci* gênicos foram realizadas utilizando o programa DAVID (*Database for Annotation, Visualization and Integrated Discovery*) (Huang *da et al.*, 2009a; Huang *da et al.*, 2009b). Como referência para o cálculo do enriquecimento foram utilizados os GENE IDs de todos os mRNAs detectados nas linhagens DU-145, MCF-7 e Mia PaCa 2 e que possuíam lncRNAs originados a partir de seus *loci* gênicos também detectados. Somente termos com *EASE score* < 0,05 (Hosack *et al.*, 2003) foram considerados enriquecidos nas análises.

Para tornar a análise de correlação em *trans*, também chamada de “*guilt by association*” computacionalmente viável, ela foi realizada para os 20% mais expressos lncRNAs intrônicos, senso e antissenso, e todos os mRNAs codificadores de proteínas que estivessem expressos nas três linhagens em estudo. Além disso, foram retirados da análise lncRNAs intrônicos senso que

possuíam correlação de expressão positiva ( $r \geq 0,5$ ) com o mRNA expresso a partir do mesmo *locus* gênico, uma vez que esses têm uma maior probabilidade de serem pré-mRNAs.

A análise de enriquecimento em termos *GO* (Ashburner *et al.*, 2000) para os mRNAs com expressão gênica correlacionadas com a expressão de lncRNAs em *trans* foi realizada utilizando o programa GENOMICA (<http://genomica.weizmann.ac.il>) (Segal *et al.*, 2004). O programa foi desenvolvido para identificar grupos de genes (módulos) que são induzidos ou reprimidos coordenadamente em uma ampla variedade de condições biológicas. Os mapas de módulos de associação são criados entre dados de expressão gênica derivados de diferentes condições experimentais e grupos de genes com algum significado biológico em particular, por exemplo, pertencerem a uma mesma categoria de *GO*. Em nossa análise, o programa foi utilizado para a construção de módulos de associação entre lncRNAs intrônicos e os termos de *GO* para os transcritos coexpressos (Hung *et al.*, 2011). Ou seja, para cada lncRNA foram definidos grupos de genes coexpressos (grupo de mRNAs com expressão positivamente ou negativamente correlacionada com um determinado lncRNA) nas três linhagens em estudo. Para identificar lncRNAs intrônicos que estão associados com cada um dos termos de *GO* foram construídos mapas de módulos gênico das associações entre cada grupo de coexpressão versus os termos de *Gene Ontology*. O cálculo de enriquecimento para os termos atribuídos foi realizado em relação a todos os mRNAs expressos que possuíam lncRNAs intrônicos detectados originados a partir de seus *loci* gênicos nas três linhagens ( $p\text{-value} \leq 0,05$  - com correção para múltiplos testes de Bonferroni).

### 3.5.8 Identificação de elementos regulatórios na extremidade 5' conhecida dos lncRNAs intrônicos

As análises para identificação de elementos regulatórios da expressão gênica nas regiões genômicas localizadas a 5' dos 5.625 e 3.504 lncRNAs intrônicos antissenso e senso, respectivamente, foi realizada utilizando o conjunto de programas de código aberto *BEDTools* (Quinlan e Hall, 2010). Estes programas utilizam arquivos em formato BED obtidos no UCSC *Genome Browser* (<http://genome.ucsc.edu/>) que contém coordenadas genômicas, facilitando a busca de elementos de interesse, como transcritos, genes, regiões promotoras, sítios de ligação de fatores de transcrição, entre outros. O uso dos programas *BEDTools* permitiu o cruzamento de coordenadas genômicas entre os lncRNAs intrônicos analisados e os elementos de interesse.

O enriquecimento da presença da enzima RNAP II e da modificação H3K4me3 foi avaliado utilizando dados de imunoprecipitação da cromatina com anticorpos específicos para estes elementos seguidos de sequenciamento em larga escala (ChIP-seq). Estes dados estão disponíveis publicamente no UCSC *Genome Browser* (*Broad/MGH ENCODE*) e foram gerados a partir de quatro linhagens celulares diferentes (ChIP-seq de RNAP II)<sup>4</sup> e 13 linhagens celulares diferentes (ChIP-seq da modificação H3K4me3)<sup>5</sup>. Para evitar que os elementos regulatórios encontrados estivessem associados a genes, as coordenadas genômicas dos *reads* de ChIP-seq de RNAP II e da marca H3K4me3 que se sobrepunham à região 5'UTRs de RefSeqs e UCSC genes foram filtradas, de modo que todos os elementos que fossem encontrados em até 1 kb *upstream*

---

<sup>4</sup> HeLa-S3, carcinoma cervical; HUVEC, linhagem de células endoteliais da veia umbilical; K562, leucemia mieloide crônica e NHEK, queratinócitos epidermais.

<sup>5</sup> Dnd41, células T leucêmicas; GM128878, linfócitos B; H1-hESC, células tronco embrionárias; HeLa-S3, carcinoma cervical; HepG2, carcinoma hepatocelular; HMEC, células epiteliais mamárias; HSMM, mioblastos de músculo esquelético; HUVEC, linhagem de células endoteliais da veia umbilical; K562, leucemia mieloide crônica; NH-A, astrócitos; NHDF-Ad, fibroblastos dermais adultos; NHLF, fibroblastos de pulmão e NHEK, queratinócitos epidermais.

dessa região ou na própria região 5'UTR saíram da análise. Após a filtragem, restaram 238.243 (77%) coordenadas genômicas nas bibliotecas de RNAP II e 510.278 (70%) coordenadas genômicas nas bibliotecas de H3K4me3. Para a análise, as coordenadas genômicas dos *reads* de ChIP-seq de RNAP II foram convertidas da montagem de 2006 do genoma humano (hg18) para a montagem de 2009 (hg19) com a ferramenta *Batch Coordinate Conversion (liftOver)* (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>).

As coordenadas referentes às CGIs foram determinadas através do cálculo do *score* epigenético como definido por Bock e colegas (Bock *et al.*, 2007). Esse *score* epigenético é uma combinação de valores que levam em consideração o (i) status de metilação, (ii) atividade promotora e (iii) estado da cromatina. Para a análise foram utilizadas 46.155 predições de CGIs. Assim como para as bibliotecas de RNAP II e H3K4me3, as CGIs também foram filtradas em relação às regiões 5'UTRs de RefSeqs e UCSC genes, reduzindo o número de CGIs na análise para 30.395 (66%).

As bibliotecas de CAGE *tags* (*Cap Analysis of Gene Expression*) utilizadas foram produzidas a partir RNA poli(A)+ extraído de nove linhagens celulares diferentes. Elas foram obtidas do projeto realizado em RIKEN OSC (RIKEN Omics Science Center) (Carninci e Hayashizaki, 1999; Kodzius *et al.*, 2006; Valen *et al.*, 2009) e estão disponíveis publicamente no UCSC *Genome Browser* (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=279293583&c=chr7&g=wgEncodeRikenCag>).

A primeira etapa da análise foi a montagem de *clusters tags* que juntassem as coordenadas genômicas de diferentes CAGE *tags*, que se sobrepusessem. Além disso, foram filtrados os *clusters tags* que se sobrepunham a regiões 5'UTRs de RefSeqs e UCSC genes. Também foram retiradas da análise *clusters* de CAGE *tags* que se sobrepunham a éxons, uma vez que estas

podariam representar possíveis inícios alternativos de transcrição de genes codificadores de proteínas (Valen *et al.*, 2009).

As coordenadas genômicas de cada um dos elementos regulatórios foram cruzadas com as coordenadas genômicas das extremidades 5' preditas pela montagem de *contigs* dos lncRNAs intrônicos antissenso e senso realizada em 2007 por Nakaya e colegas (Nakaya *et al.*, 2007). Para as análises de enriquecimento foi computada a coordenada genômica do elemento regulatório encontrado mais próximo do início predito de cada um dos transcritos não codificadores em análise em até 10 kb de distância do mesmo. Foram consideradas as coordenadas que não se sobrepunham, ou que se sobrepunham parcialmente, à coordenada do transcrito intrônico analisado.

Para fins de comparação, as análises de enriquecimento de elementos regulatórios na extremidade 5' foi realizada também para os 6.611 genes codificadores de proteínas que foram detectados nas análises de expressão.

Como controle, foi utilizado um conjunto de sequências genômicas randômicas que apresentavam tamanhos e distribuição cromossômica similar a das sequências não codificadoras analisadas. As sequências dos grupos randômicos foram analisadas exatamente da mesma forma que as sequências dos lncRNAs intrônicos.

Para agregar um valor estatístico nas análises, foi utilizado o teste estatístico Kolmogorov-Smirnov, que testa a diferença de distribuição entre dois grupos. Para isso foi utilizado um *script* escrito em R com o pacote *Deducer* para comparar os grupos de RNAs não codificadores e codificadores de proteínas (mRNAs) com os grupos randômicos. Somente foram consideradas com significância estatística as análises que apresentaram todos os valores de *p-value* < 0,05 comparado aos grupos randômicos.

### 3.5.9 TransMaps – Análise de conservação de expressão

A análise de conservação de expressão para os 5.625 lncRNAs intrônicos antissenso e 3.504 lncRNAs intrônicos senso detectados nas linhagens celulares humanas DU-145, MCF-7 e Mia PaCa 2 foi realizada identificando-se ortólogos expressos em espécies de vertebrados. Essa identificação de foi obtida cruzando-se as coordenadas genômicas dos lncRNAs com as coordenadas genômicas de transcritos de vertebrados que foram sintenticamente mapeados no genoma humano pelo método de alinhamento de mRNAs entre espécies denominado TransMap (Zhu *et al.*, 2007). TransMap mapeia todos os transcritos conhecidos (ex., cDNAs full-length e outros presentes em bancos de dados como RefSeq ou UCSC) e ESTs entre espécies de vertebrados usando alinhamentos sintênicos. Primeiramente os transcritos são alinhados nos genomas nativos usando BLAT. Em seguida, o TransMap projeta os alinhamentos dos cDNAs entre as espécies no genoma alvo utilizando o algoritmo de alinhamento BLASTZ (Schwartz *et al.*, 2003). BLASTZ é um algoritmo de alinhamento altamente sensível e otimizado para alinhar sequências genômicas ortólogas divergentes (Stanke *et al.*, 2008). Das 26 espécies disponíveis para a análise, oito<sup>6</sup> foram excluídas, pois apresentam menos de 10.000 regiões sintenticamente mapeadas na montagem hg19 do genoma humano. Os números de lncRNAs antissenso e senso que possuem ortólogos expressos (TransMaps) foram comparados com o número de ortólogos expressos (TransMaps) encontrados para um conjunto de sequências de mesmo tamanho e com a mesma distribuição cromossômica geradas randomicamente. O teste estatístico Qui-Quadrado foi realizado.

---

<sup>6</sup> Elefante (loxAfr3), babuíno (papHam1), lagarto (anoCar1), canguru (macEug1), sagui (calJac1), gato (felCat3), *Monodelphis domestica* (monDom5) e ornitorrinco (ornAna1).

### 3.5.10 PhastConsElements 46way – Análise de conservação de sequência

Os elementos evolutivamente conservados *PhastConsElements 46way* foram obtidos no UCSC *Genome Browser* (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=279496285&c=chr21&g=cons46way>). As 46 espécies contidas no alinhamento múltiplo, bem como os métodos utilizados para se medir a conservação evolutiva estão descritas na página do UCSC *Genome Browser*. O número de elementos conservados em cada grupo de espécies está distribuído da seguinte forma:

- Vertebrados (46 espécies; n = 5.163.775)
- Placentários (32 espécies; n = 3.743.478)
- Primatas (9 espécies; n = 725.627)

As coordenadas lncRNAs detectados em pelo menos uma das três linhagens foram cruzadas com as coordenadas genômicas dos elementos conservados presentes em cada grupo de espécies usando a ferramenta *BEDTools* (Quinlan e Hall, 2010). O número de transcritos identificados como conservados em cada grupo de espécies foi dividido pelo número total de elementos presentes em cada grupo para torná-los comparáveis entre si. Semelhantemente ao que foi realizado para a análise de TransMaps, coordenadas genômicas de um conjunto de sequências de mesmo tamanho e com a mesma distribuição cromossômica geradas randomicamente foram cruzadas com as coordenadas genômicas dos elementos conservados em cada grupo de espécies. O teste estatístico Qui-Quadrado foi realizado comparando-se as frequências de elementos conservados obtidas para os lncRNAs com as frequências de elementos conservados obtidas para as sequências randômicas.

### 3.5.11 RNaz – Predição de estrutura secundária

O programa RNaz (Gruber *et al.*, 2007) determina a predição de estrutura secundária baseando-se em dois parâmetros importantes:

- conservação estrutural
- estabilidade termodinâmica

A partir dos seis melhores alinhamentos das sequências dos lncRNAs intrônicos entre 46 espécies de vertebrados diferentes usando o algoritmo *RNAalifold* o programa constrói uma sequência consenso predita (Hofacker *et al.*, 2002). A seguir, o algoritmo utiliza informações termodinâmicas e filogenéticas para determinar a estrutura desta predição. RNaz determina a conservação estrutural calculando a taxa de energia livre da conformação consenso em relação à taxa de energia livre da sequência de fita simples em cada espécie (SCI – *Structure Conservation Index*). Além disso, o programa determina um *score* para estabilidade da sequência. Mutações compensatórias que preservam a estrutura secundária são consideradas como um adicional na pontuação no modelo de energia livre. Assim, esses dois parâmetros são levados em consideração para classificar o alinhamento da sequência como sendo estruturalmente conservada ou não em uma janela determinada. Para a predição da estrutura secundária foram utilizadas janelas de 120 bp. Apenas foram consideradas estruturas que apresentaram RNaz *score*  $\geq 0,5$ .

### 3.5.12 Definição de lncRNAs intrônicos e mRNAs codificadores de proteínas regulados pelo tratamento com 5-AZA - Análise estatística

Os mRNAs de genes codificadores de proteínas e os lncRNA intrônicos significativamente diferencialmente expressos após tratamento com 5  $\mu$ M de 5-AZA foram definidos aplicando-se o teste estatístico para análise de *microarrays* SAM (*Significance Analysis*

of *Microarrays*) (Tusher *et al.*, 2001) sobre os valores dos logaritmos de base 2 das razões das intensidades medidas em cada sonda, entre as amostras tratadas e controles. O SAM confere um *score* para cada gene na análise baseado na mudança de expressão deste gene em relação ao desvio-padrão de diversas medições repetidas. Para os genes com um *score* maior do que um valor mínimo ajustável (*threshold*), o programa usa repetidas permutações para estimar a probabilidade, ou FDR (False Discovery Rate), de estes genes serem identificados ao acaso. Cada gene encontrado também está associado a um *q-value*, que representa a menor taxa de FDR na qual aquele gene é considerado significativo (Tusher *et al.*, 2001). *Q-value* < 10% foi utilizado como critério para considerar um transcrito diferencialmente expresso entre as condições testadas.

### **3.5.13 Anotação funcional dos genes codificadores de proteínas regulados pelo tratamento com 5-AZA**

As anotações funcionais dos mRNAs de genes codificadores de proteínas regulados pela desmetilação do DNA gerada pelo tratamento com o agente 5-AZA foram feitas utilizando a ferramenta *Ingenuity Pathways Analysis* (Ingenuity Systems). Os números de acesso dos genes codificadores de proteínas expressos (Gene IDs), bem como os dados de diferença de expressão entre tratado e controle (*fold change*) e a significância estatística (*q-value*), gerada a partir da análise de SAM foram importados no *software Ingenuity Pathway Analysis 7.1*. A análise foi feita com *q-value* < 10% como critério de corte para considerar genes diferencialmente expressos. O programa cria redes para genes com funções similares e interações já reportadas. Um *score* é computado para cada categoria funcional de acordo com o número de genes considerados diferencialmente expressos que estão presentes nela. Este *score* reflete o logaritmo

negativo do *p-value* em base 10, que indica a chance dos genes diferencialmente expressos que pertencem a uma categoria estarem associados a ela devido ao acaso.

### **3.6 Abordagens para validação da expressão fita específica de lncRNAs intrônicos regulados pelo tratamento com 5-AZA**

#### **3.6.1 Transcrição reversa fita específica**

A produção do cDNA fita específica foi realizada utilizando o RNA total extraído como descrito anteriormente (subseção 3.3.) e o *kit ThermoScript™ RT-PCR System* (Invitrogen). Inicialmente, amostras de 4 µg de RNA total misturados a 2 µL de dNTP mix (10 mM) e 1 µL de *primer* específico para cada transcrito (10 µM) são incubados por 3 minutos a 85 °C para desnaturação do RNA total. A seguir, a temperatura da reação é abaixada para a temperatura de anelamento ótima para cada *primer* (Tabela 4) por 5 minutos. Após esse intervalo de tempo, é adicionada aos tubos a mistura contendo: 15 unidades da enzima *ThermoScript™ RT*, 40 unidades da enzima *RNase OUT™*, 1 µL de DTT 0,1 M em tampão *cDNA Synthesis Buffer* (Tris acetato 250 mM, pH 8,4; acetato de potássio 375 mM e acetato de magnésio 40 mM). A reação é então realizada, incubando as amostras por 1 hora na temperatura de anelamento ideal do *primer* de cada reação (ver Tabela 4). Em seguida, a reação é finalizada, incubando-a por 5 minutos a 85 °C para inativação das enzimas.

**Tabela 4: Sequências dos *primers* utilizados para as reações de transcrição reversa fita específica.**

Nome do <i>primer</i>	Sequência de nucleotídeos (5' -> 3')	Temperatura de anelamento (°C)
lncRNA AS-SMG6-SENDO	GTATACAGGCGAAAGTCTAAGAGGACAG	65
lncRNA AS-SMG6-ANTISENDO	TGTGAAAAATGTCAAAGTATCCACAGAAAC	64
lncRNA AS-APP-SENDO	GTCCTAAAACAGGAGCAAGGTGGTTAATTC	65
lncRNA AS-APP-ANTISENDO	GGGATTGCGTATCTAAACAACTAGAAAAGG	64
lncRNA AS-CELF2-SENDO	GTAACCCCTAAAACCTCCAGGCGAGTAAGT	66
lncRNA AS-CELF2-ANTISENDO	ACTCTAGAGCAGTTTAGGGATGAAGTCTGG	66

Para cada transcrito não codificador a ser validado, foram preparadas quatro reações em paralelo sob as mesmas condições de incubação: (1) Reação Senso, reação na qual foi utilizado como *primer* para a transcriptase reversa um oligonucleotídeo específico complementar à mesma fita da qual o gene codificador de proteína é transcrito; (2) Reação Antissenso, na qual foi utilizado como *primer* para a transcriptase reversa um oligonucleotídeo específico complementar à fita cuja orientação é antissenso em relação ao gene codificador de proteína; (3) Reação SP, reação controle na qual não é adicionado oligonucleotídeo iniciador. Essa reação é importante para termos a certeza de que o cDNA produzido nas reações (1) e (2) foi especificamente produzido a partir dos iniciadores específicos para cada uma das fitas adicionados em cada reação e não por um artefato de auto anelamento ou pelo anelamento de uma outra molécula de RNA presente no pool total de RNA; (4) Reação SE, reação controle na qual a enzima transcriptase reversa não é adicionada. Esta reação é importante para descartarmos a possibilidade de amplificação de alguma contaminação por DNA genômico nas amostras de RNA ou em algum reagente da reação de transcrição reversa, na reação de PCR subsequente à produção do cDNA.

Após a transcrição reversa, a presença ou ausência das mensagens não codificadoras senso e antissenso no pool de RNA total das células foi avaliada por PCR. Para isso, um par de

*primers* para cada um dos transcritos foi desenhado com o programa *Primer3* ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)), utilizando os parâmetros padrões, numa região interna ao *primer* utilizado na reação de transcrição reversa (*primer nested*). As sequências desses *primers* estão mostradas na Tabela 5. A reação de PCR foi realizada com quantidades equivalentes a 8 ng de RNA reversamente transcrito (utilizando uma estimativa de 2% de RNA reversamente transcrito na população de RNA total), 5 unidades da enzima *Go Taq DNA Polymerase* (Promega), 0,2 mM de dNTPs e 0,2 µM de cada *primer* específico em tampão recomendado pelo fabricante, que contém 1,5 mM de MgCl<sub>2</sub>. Em seguida, os produtos da PCR foram analisados por eletroforese em gel de agarose 1% em TAE com 0,5 µg/mL de brometo de etídio.

**Tabela 5: Sequência dos *primers nested* utilizados nas reações de RT-PCR fita específica.**

Nome do <i>primer</i>	Sequência de nucleotídeos (5' -> 3')*
lncRNA AS-SMG6-NT	F: TTTAAGCCAGGGACAAGCAC
	R: CTACAGACGCAGCCAGTCAG
lncRNA AS-APP-NT	F: CTGAGATTCCAGCCACCAG
	R: GCATTTGGAAATGAGCACTTG
lncRNA AS-CELF2-NT	F: GCAGTTTAGGGATGAAGTCTGG
	R: AACTCCAGGCGAGTAAGTTGAG

\* F = Forward; R = Reverse

### 3.6.2 RACE-PCR

A metodologia para amplificação de extremidades desconhecidas de fragmentos de cDNA, RACE-PCR (*Rapid Amplification of cDNA Ends*) foi realizada para obtenção da sequência completa de alguns transcritos intrônicos identificados como regulados pelo tratamento com 5-AZA. Amplificações por RACE das extremidades 5' e 3' de moléculas de cDNA foram realizadas utilizando bibliotecas comerciais (Clontech) geradas a partir de RNA poli(A)+ de tecido de próstata (*Human Prostate Marathon-Ready cDNA*).

#### **Planejamento dos *primers***

Como sugerido pelo manual do usuário das bibliotecas, para cada transcrito foram desenhadas quatro sequências de *primers* específicos e complementares às sequências das ESTs que mapeiam na região interna dos *contigs* de lncRNAs intrônicos. Em cada uma das extremidades dos transcritos foram desenhados dois *primers* orientados para as pontas da mensagem, sendo que um deles ancora-se a aproximadamente 150 nucleotídeos e o outro (*primer nested*) a aproximadamente 100 nucleotídeos das extremidades dos *contigs*. O segundo *primer* (*primer nested*) possibilita a realização de uma reação de PCR sequencial (*Nested*), a partir do produto da reação com o primeiro *primer* diluído, o que aumenta a especificidade dos produtos gerados. Os *primers* utilizados (Tabela 6 e Tabela 7) foram desenhados com o programa *Primer3* ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)).

**Tabela 6: Sequência dos *primers* utilizados para as reações de RACE 5'.**

Nome do <i>primer</i> *	Sequência de oligonucleotídeos (5'→3')
lncRNA AS-SMG6_RC5	CTGCCAAGACCCTTACCCTCTCCTCTCC
lncRNA AS-SMG6_RC5_NT	TTAACCCCTCTCCCACATCTGGAAAACC
lncRNA AS-APP_RC5	GCCACCATCATATGTTTGGTCTGTCTGTG
lncRNA AS-APP_RC5_NT	CCACGCTTTCCAAGGAGGTAAGTGTG
lncRNA AS-APP_2RC5	GGTCAGGTTGAATGGAAGCCACGA
lncRNA AS-APP_2RC5_NT	GGGGGTGCAGAACCTAGGAGTGGCTTAC
lncRNA AS-CELF2_RC5	AGTTCCTCCACATGGTGGTCACCTCTCC
lncRNA AS-CELF2_RC5_NT	CTGAGCCCACTGCCCTCAACTACTCG

\* RC5 = RACE 5'; NT = *Nested***Tabela 7: Sequência dos *primers* utilizados para as reações de RACE 3'.**

Nome do <i>primer</i> *	Sequência de oligonucleotídeos (5'→3')
lncRNA AS-SMG6_RC3	GATAAGTGTGGCTGGGGTGGTGGTATGC
lncRNA AS-SMG6_RC3_NT	CTGCGTCTGTAGCAGCCCCACTATTCC
lncRNA AS-APP_RC3	GCTATTCTGACATGGAGGGCACATTGG
lncRNA AS-APP_RC3_NT	TGTAGGTGTGTAGCCAGGTGTGTAGCA
lncRNA AS-CELF2_RC3	CCAGGCGAGTAAGTTGAGGGCAGTGG
lncRNA AS-CELF2_RC3_NT	CAGGGCCAGACTTCATCCCTAAACTGC

\* RC3 = RACE 3'; NT = *Nested*

## Reação de RACE-PCR

As reações de RACE-PCR com as bibliotecas de cDNA comerciais foram realizadas em um volume de 50 µL, com: 5 µL de cDNA molde; 0,2 mM de dNTP mix; 1X *Advantage 2 PCR Buffer* (Clontech); 1X *Advantage 2 Polymerase Mix* (Clontech); 0,2 µM de *primers* específicos para o RACE 5' ou RACE 3', e 0,2 µM do *primer* universal (*Adaptor Primer 1*) complementar ao adaptador existente nas porções 5' e 3' de cada molécula de cDNA presente na biblioteca. As reações de RACE-PCR *Nested* são essencialmente iguais às reações de RACE-PCR já descritas, com a diferença de que o molde para a reação é o produto de PCR da primeira reação diluído 200

vezes e de que os *primers* específicos à sequência (*primer nested*) e o *primer* universal (*Adaptor Primer 2*) são complementares a uma região interna ao anelamento dos *primers* anteriores.

### **Touchdown PCR**

Todas as reações foram incubadas no termociclador GeneAmp PCR System (Applied Biosystems). Foi realizado o *Touchdown PCR* para amplificação dos produtos, com o objetivo de aumentar a especificidade da reação de RACE-PCR. Este tipo de reação utiliza, nos ciclos iniciais da PCR, temperaturas de anelamento e extensão alguns graus acima da  $T_m$  (temperatura de *melting*) dos *primers*. Embora o anelamento dos *primers* seja menos eficiente nessas condições, ele é mais específico, o que permite a acumulação de produto específico no início da reação, que será utilizado como molde nos ciclos subsequentes, realizados em temperaturas mais baixas ( $T_m$  dos *primers*). As condições utilizadas para ciclagem foram: 94 °C por 1 minuto para desnaturação das fitas duplas de cDNA, seguido por 5 repetições de 94 °C por 30 segundos e 72 °C por 4 minutos para anelamento e extensão; mais 5 repetições de 94 °C por 30 segundos e 70 °C por 4 minutos e mais 25 repetições de 94 °C por 30 segundos e 68 °C por 4 minutos.

### **3.7 Clonagem e sequenciamento dos produtos obtidos**

Os produtos obtidos foram separados por eletroforese em gel de agarose 1%. As bandas que possivelmente corresponderiam às mensagens específicas foram cortadas do gel e purificadas com o kit *Wisard SV Gel and PCR Clean-Up System* (Promega) conforme orientações do fabricante. Os fragmentos purificados foram então clonados no plasmídeo *pGEM-T Easy* (Promega) com a enzima *T4 DNA ligase* (Promega) segundo as orientações do fabricante. A seguir, as ligações foram submetidas à diálise por 2 horas e utilizadas para eletroporação em bactérias *E. coli* DH10B eletrocompetentes. Após o crescimento das bactérias em meio LB sólido

contendo ampicilina por 16 horas a 37 °C, uma colônia de cada clonagem foi crescida em 5 mL de meio LB líquido contendo ampicilina por mais 16 horas sob agitação a 37 °C. Esse inóculo foi utilizado para a extração do DNA plasmidial com o kit *Pure Yield™ Plasmid Miniprep System* (Promega) seguindo o protocolo do fabricante. Os plasmídios purificados foram então sequenciados com o método dideoxi no sequenciador *MegaBace* (Armeshan) utilizando para a reação de sequenciamento o kit *BigDye Terminator v3.1 Cycle Sequencing* (Applied Biosystems) e 2,5 µM de um *primer* complementar à região presente no vetor, que flanqueia o inserto. Confirmadas as identidades das porções 5' e 3' obtidas, através de mapeamento no genoma humano, o fragmento completo foi amplificado a partir das bibliotecas de cDNA com um novo par de *primers* ancorados nas novas extremidades conhecidas do lncRNA. Procedimentos similares de clonagem e sequenciamento foram repetidos com as mensagens completas.

### **3.7.1 Quantificação de RNAs por PCR em tempo real**

#### **Planejamento e otimização dos *primers***

Os oligonucleotídeos (*primers*) utilizados para quantificação da expressão gênica neste trabalho (Tabela 8) foram planejados utilizando o programa *Primer Express* versão 2.0 (Applied Biosystems). Os principais parâmetros definidos para o desenho foram: tamanho dos fragmentos amplificados entre 100 e 150 pb, conteúdo CG entre 30% e 80 %, ausência de capacidade de formação de dímeros ou de estrutura secundária e temperatura de anelamento entre 58 °C e 60 °C, de acordo com o algoritmo *Nearest Neighbor* (Breslauer *et al.*, 1986).

**Tabela 8: Sequência de *primers* utilizados para quantificação da expressão gênica.**

Nome do <i>primer</i>	Sequência de nucleotídeos (5' -> 3')*
lncRNA AS-CELF2	F: GCAGTTTAGGGATGAAGTCTGG
	R: AACTCCAGGCGAGTAAGTTGAG
lncRNA AS-APP	F: TTCTGACATGGAGGGCACATT
	R: CGCTTTCCAAGGAGGTAACT
APP-codificador	F: TCTCTGAAGTGAAGATGGATGC
	R: GATTGCACCTTTGTTGAACC

\* F = Forward; R = Reverse

Otimizações das condições ideais para utilização dos pares de *primers* desenhados foram realizadas para a garantia de eficiências comparáveis na amplificação das amostras experimentais. Para os pares de *primers* utilizados na quantificação absoluta, as eficiências foram calculadas com base na regressão linear ajustada às curvas de calibração específicas construídas para cada um deles (ver subseção 3.6.3.2.). A concentração ideal dos pares de *primers* na reação de PCR em tempo real foi determinada realizando reações idênticas de amplificação, porém com quantidades crescentes desses, de 200 nM – 1.000 nM, até que o ciclo de referência (Ct – número do ciclo em que a quantidade de alvo amplificada alcança um limiar fixo na faixa exponencial de amplificação) da reação se estabilizasse em um valor mínimo. Para cada concentração de par de *primer* testada, foi avaliada a ausência de formação de dímeros dos *primers* através da análise de curvas de dissociação de cada reação. Após estes testes foi definida a concentração de 800 nM para os pares de *primers* utilizados.

### 3.8 Quantificação absoluta de RNAs por PCR em tempo real

Para a quantificação da expressão fita específica de RNAs não codificadores intrônicos utilizamos a técnica de PCR em tempo real absoluto. O primeiro passo para a quantificação das amostras com essa técnica é a construção de uma curva de calibração específica para cada RNA alvo. A curva, utilizada como padrão externo para o cálculo do número de moléculas de RNA por massa de RNA total usado na reação de transcrição reversa, é construída a partir da diluição seriada de um DNA padrão de concentração conhecida. Esse DNA padrão deve ser obtido com os mesmos *primers* específicos que serão utilizados na quantificação por PCR em tempo real, ou seja, deve ter a mesma composição de deoxinucleotídeos que os *amplicons* da reação.

O DNA padrão para a construção da curva de calibração específica foi obtido pela amplificação de 25 ng de DNA genômico por PCR, realizada com 5 unidades da enzima *Go Taq DNA Polymerase* (Promega), 0,2 mM de dNTPs e 0,2  $\mu$ M de cada *primer* específico em tampão recomendado pelo fabricante, que contém 1,5 mM de  $MgCl_2$ . Em seguida, todo o produto da PCR foi analisado por eletroforese em gel de agarose 1% em TAE com 0,5  $\mu$ g/mL de brometo de etídio. Confirmada a presença de banda única no gel, estas foram cortadas e purificadas com o *Kit Wizard® SV Gel and PCR Clean-Up System* (Promega), segundo recomendações do fabricante. Os produtos purificados foram quantificados por espectrofotometria (absorbância a 260 nm) no equipamento NanoDrop-1000 (Thermo Scientific). As concentrações medidas foram convertidas em número de cópias por  $\mu$ L usando a constante de Avogrado ( $6,023 \times 10^{23}$ ) e a massa molecular dos fragmentos amplificados (número de bases do produto de PCR multiplicado pela massa molecular média de um par de deoxinucleotídeos, que é 660 Da). Uma solução estoque de concentração inicial  $1,0 \times 10^{-1}$  ng/ $\mu$ L foi preparada. A partir dessa solução estoque, foram realizadas até 9 diluições seriadas de 10 vezes que foram mantidas aliquotadas e

congeladas a  $-80\text{ }^{\circ}\text{C}$ . Para maximizar a acurácia da curva de calibração, os pontos de diluição máximo e mínimo deveriam incluir a quantidade esperada de RNA alvo nas amostras experimentais.

A reação de PCR em tempo real foi realizada no equipamento *7500 Real Time PCR Systems* (Applied Biosystems) com o reagente *SYBR® Green* (Applied Biosystems). A técnica se baseia na detecção da fluorescência emitida por este reagente quando o mesmo intercala-se entre as fitas de DNA resultantes dos ciclos de amplificação realizados na reação de PCR. Os valores de fluorescência, que representam a quantidade de produto amplificada são medidos durante cada ciclo. Quanto maior a quantidade de *template* presente no início da reação, menor será o número de ciclos necessário para alcançar um ponto no qual o sinal de fluorescência fica significativamente acima de um nível de *threshold* definido e comum para todas as reações a serem diretamente comparadas. Esse ponto é conhecido como Ct (threshold cycle), e sempre ocorrerá durante a fase exponencial da amplificação. Assim, a quantificação não será afetada por qualquer limitação de algum componente da reação em sua fase platô. A curva de calibração é então criada, plotando-se os valores de Ct obtidos para cada uma das diluições do padrão *versus* os valores de logaritmo das concentrações correspondentes (expressas em números de cópias /  $\mu\text{L}$ ).

A partir da curva de calibração plotada, pode-se ajustar uma regressão linear cuja equação da reta fornece parâmetros importantes para avaliação da qualidade do ensaio de PCR em tempo real. O coeficiente de determinação da curva ( $R^2$ ) é uma medida do desempenho do ensaio, ou seja, de quanto os pontos da curva de calibração estão ajustados dentro da regressão linear. Um valor de 1 indica um perfeito encaixe entre a regressão linear e os pontos da curva. Somente valores de  $R^2$  maiores que 0,99 foram considerados aceitáveis nos experimentos apresentados nesse trabalho. Já o coeficiente de regressão linear indica a eficiência de amplificação do ensaio.

Coefficientes de regressão de -3,32 indicam uma eficiência de amplificação de 100%. Foram considerados aceitáveis nos experimentos apresentados nesse trabalho, coeficientes de regressão entre -3,1 e -3,6 (90% a 110% de eficiência, respectivamente). Além de se determinar os parâmetros acima, a partir da equação da reta pode-se obter a quantificação dos RNAs experimentais:

$$\text{Quantidade de moléculas de RNA} = 10^{(Ct-b)/m}$$

Na equação,  $b$  é o valor do intercepto do eixo Y (Ct quando a quantidade é igual a 1) e  $m$  é o valor do coeficiente de regressão.

As reações de PCR foram realizadas como recomendado pelo protocolo da Applied Biosystems. Resumidamente, foram adicionados à reação 5  $\mu$ L do cDNA (*template*) diluído uma vez, 5  $\mu$ L da mistura do par de *primers* na concentração de 800 nM e 10  $\mu$ L do reagente *SYBR® Green PCR Master Mix* (Applied Biosystems). Todas as quantificações por PCR em tempo real foram realizadas com três réplicas biológicas independentes e as significâncias estatísticas foram testadas com o teste estatístico “*Student’s t-test*”.

### **3.9 Identificação de uma possível região promotora para o lncRNA AS-APP e seu estudo funcional**

#### **3.9.1 Identificação da predição informática do possível promotor do lncRNA AS-APP**

A identificação do promotor foi realizada a partir de uma busca informática por predições na região de 5 kb de sequência genômica localizada *upstream* do início (extremidade 5’) conhecido do lncRNA AS-APP utilizando os parâmetros padrões do programa FPRM (<http://linux1.softberry.com/berry.phtml?topic=fprom&group=programs&subgroup=promoter>).

### Planejamento dos *primers* para clonagem

Para avaliarmos a atividade promotora da região regulatória predita para o lncRNA intrônico AS-APP pelo programa FPRM utilizamos o ensaio de gene repórter. Para isso, a região promotora predita foi clonada de forma orientada a montante do gene da enzima luciferase de *Photinus pyralis* no vetor *pGL3-Basic Vector* (Promega). Os pares de *primers* utilizados (Tabela 9) foram planejados para amplificar a região promotora em estudo e adicionar especificamente o sítio de reconhecimento da enzima de restrição NheI (sequências em vermelho nos *primers* da Tabela 9) em uma das pontas da sequência amplificada e o sítio de restrição da enzima HindIII (sequências em azul nos *primers* da Tabela 9) na outra ponta da sequência. Paralelamente, outro par de *primers* foi planejado para amplificar essa mesma região, no entanto, com os sítios de restrição das enzimas trocados em relação às pontas da sequência, a fim de se poder também amplificar e clonar a sequência promotora invertida.

**Tabela 9: Sequência dos *primers* utilizados para clonagem da predição de promotor para o lncRNA AS-APP.**

Nome do primer	Sequência de oligonucleotídeos (5'→3')*
lncRNA AS-APP-Promotor	F: AAGAGCTAGCACAGCACAGCTGTCAAAGG
	R: ACATAAGCTTGCTTCTCCATGTTACCCTTA
lncRNA AS-APP-Promotor Invertido	F: AAGAAAGCTTACAGCACAGCTGTCAAAGG
	R: ACATGCTAGCGCTTCTCCATGTTACCCTTA

\* F = Forward; R = Reverse

### 3.9.2 Amplificação e clonagem da região promotora

A região promotora predita foi amplificada por PCR com os dois pares de *primers* de clonagem sítio dirigida a partir do DNA genômico extraído de células da linhagem DU-145. As reações foram realizadas em termociclador *GeneAmp PCR System* (Applied Biosystems) com 25 ng do DNA, 5 unidades da enzima *Go Taq DNA Polymerase* (Promega), 0,2 mM de dNTPs e 0,2 µM de cada *primer* específico em tampão recomendado pelo fabricante, que contém 1,5 mM de MgCl<sub>2</sub>. Em seguida, todo o produto das PCRs foi analisado por eletroforese em gel de agarose 1% em TAE com 0,5 µg/mL de brometo de etídio. As bandas correspondentes aos *amplicons* específicos foram cortadas e purificadas com o *Kit Wizard® SV Gel and PCR Clean-Up System* (Promega), segundo recomendações do fabricante. Os produtos purificados foram quantificados por espectrofotometria (absorbância a 260 nm) no equipamento NanoDrop-1000 (Thermo Scientific). Alíquotas de 1.500 ng de cada um dos fragmentos purificados, bem como 1.000 ng do vetor *pGL3-Basic Vector* (Promega) foram digeridas, em paralelo, com 10 unidades das enzimas de restrição NheI e HindIII (New England BioLabs) em tampão *NEBuffer 2*, adicionado de 0,1 µg/µL de BSA a 37°C por 3 horas, seguido por 15 minutos a 65°C para inativação das enzimas. Após as digestões, os fragmentos foram novamente purificados com o *Kit Wizard® SV Gel and PCR Clean-Up System* (Promega), quantificados por espectrometria e ligados ao vetor com a enzima *T4 DNA ligase* (Promega) respeitando a proporção de 1 massa de DNA inserto para 3 massas de vetor, *overnight*, segundo as orientações do fabricante. A seguir, as ligações foram submetidas à diálise por 2 horas e utilizadas para eletroporação em bactérias *E. coli* DH10B eletrocompetentes. Após o crescimento das bactérias em meio LB sólido contendo ampicilina por 16 horas a 37 °C, colônias de cada clonagem foram transferidas para tubos de ensaio contendo 5 mL de meio LB líquido contendo ampicilina e crescidas por 16 horas a 37 °C. Esse inóculo foi

utilizado para a extração do DNA plasmidial com o kit *Pure Yield<sup>TM</sup> Plasmid Miniprep System* (Promega) seguindo o protocolo do fabricante. Para a confirmação da clonagem sítio dirigida, os plasmídios purificados foram sequenciados com o método dideoxi no sequenciador *MegaBace* (Armehsan) utilizando para a reação de sequenciamento o kit *BigDye Terminator v3.1 Cycle Sequencing* (Applied Biosystems) e 2,5  $\mu\text{M}$  de um *primer* complementar à região presente no vetor, que flanqueia o inserto.

### **3.9.3 Lipofecção de células HEK293 para ensaios de atividade promotora com gene repórter**

Cerca de  $10^5$  células da linhagem HEK293 foram crescidas em meio DMEM contendo soro fetal bovino por 24 horas em placas de cultura de 24 poços (Corning). Após esse período, os meios de cultura foram trocados por meios frescos e 50  $\mu\text{L}$  da mistura de 1.000 ng do vetor pGL3-Basic contendo a sequência promotora predita para o lncRNA AS-APP clonada, 125 ng do vetor pRL-SV40 (Promega) e 3,4  $\mu\text{L}$  do reagente de transfecção *FuGENE HD* (Promega) foram adicionados às células. Paralelamente, foram transfectados também com o reagente FuGENE HD, os vetores pGL3-Basic contendo a sequência promotora predita para o lncRNA AS-APP invertida clonada, pGL3-Basic vazio (controle negativo) e pGL3-Promoter/SV40 (Promega). O vetor pGL3-Promoter/SV40 é essencialmente o vetor pGL3-Basic contendo a sequência do promotor viral SV40 clonado *upstream* do gene da luciferase (controle positivo). O vetor pRL-SV40 foi transfectado nas células juntamente com todos os vetores pGL3 com a finalidade de normalizar os ensaios de atividade que se quer comparar, reduzindo as variações observadas decorrentes de diferenças nas eficiências de transfecção. Esse vetor é um plasmídeo que contém o promotor viral SV40 clonado *upstream* do gene da luciferase de *Renilla reniformis*. Suas medidas de luminescência são utilizadas para normalizar as medidas de luminescência obtidas a partir da

reação catalisada pela enzima luciferase de *Photinus pyralis*. Para compor réplicas biológicas, três diferentes transfecções de cada um dos vetores pGL3-Basic (pGL3-promotor, pGL3-promotor invertido, pGL3-vazio e pGL3-Promoter/SV40) foram realizadas. Após 48 horas, as células transfectadas foram coletas e lisadas com 100 µL do reagente *Passive Lysis Buffer* (Promega) e congeladas a -80 °C até o momento do ensaio de atividade promotora.

### 3.9.4 Ensaio de atividade promotora

A atividade promotora foi medida com o ensaio *Dual-Luciferase Reporter Assay System* (Promega), seguindo as recomendações do fabricante. Com este método as medidas da atividade da enzima luciferase de *Photinus pyralis* é obtida pela a adição de seu substrato específico (luciferina). Em seguida, sua atividade é inativada e é obtida a medida da atividade da luciferase de *Renilla reniformis* pela adição de seu substrato (coelenterazina). Resumidamente, 100 µL do reagente *Luciferase Assay Reagent II* (LAR II) e 20 µL do lisado de células foram transferidos para placas *OptiPlate-96 White Opaque* de 96 poços (Thermo Fisher Scientific) e misturados com a pipeta. A primeira medida de luminescência, gerada pela reação da enzima luciferase de *Photinus pyralis*, foi obtida pelo luminômetro *MicroLumat Plus* (Microplate Luminometer LB – Berthold Technologies) com o programa *WinGlow* (Berthold Technologies). Em seguida, 100 µL do reagente *Stop & Glo Reagent* foi adicionado ao mesmo poço, misturado com a pipeta e a segunda medida de luminescência, gerada pela reação da enzima luciferase *Renilla reniformis* foi obtida. Razões entre as medidas de luminescência obtidas com as enzimas luciferase de *Photinus pyralis* e de *Renilla reniformis* foram calculadas (Unidade Relativa de luz, LRU). Três medidas de luminescência para cada lisado de células foram realizadas (réplicas técnicas).

### **3.10 Extração, quantificação e fragmentação de DNA genômico (gDNA)**

#### **3.10.1 Extração do gDNA**

Os sedimentos de células coletados foram ressuspensos em 500 µL de TE (Tris 10mM, EDTA 1mM) e lisados com o tampão de lise (20 mM Tris pH 8.0; 4 mM EDTA pH 8,0; 20 mM NaCl; 1 % SDS) contendo 20 µL de proteinase K (10 mg/mL). As amostras foram incubadas a 55 °C por 16-18 horas para digestão das proteínas celulares. A seguir, às amostras, foi adicionado 2 µL de RNase A (20 mg/mL), e as mesmas, foram incubadas a 37 °C por 1 hora. A extração propriamente dita do DNA genômico foi realizada com 600 µL de fenol-clorofórmio-álcool isoamílico (25:24:1) pH 8,0, agitando-se o tubo por 5 minutos e centrifugando a 1.800xg por 15 minutos a 4 °C. Após a centrifugação a fase aquosa foi coletada e transferida para um tubo novo. O ciclo de extração foi repetido 2 vezes. A seguir o gDNA foi precipitado por 16-18 horas a -20 °C com 1.400 µL de acetato de sódio (75 mM) em etanol absoluto. Após essa etapa de precipitação as amostras foram centrifugadas a 1.800xg por 15 minutos a 4 °C. O sobrenadante foi descartado e o gDNA sedimentado foi lavado com 1 mL de etanol 70% e centrifugado a 1.800xg por 10 minutos. Novamente, o sobrenadante foi descartado e o sedimento foi seco à temperatura ambiente. A seguir o gDNA foi ressuspensado suavemente em 300 µL de TE (Tris 10mM, EDTA 1mM). Para acelerar a solubilização e inativar DNases as amostras foram incubadas por 30 minutos a 65 °C. Após completa solubilização, as amostras de gDNA extraídas foram estocadas a 4 °C.

#### **3.10.2 Avaliação do rendimento, pureza e integridade do gDNA**

O DNA genômico isolado foi quantificado pela medida óptica de absorvância a 260 nm no espectrofotômetro NanoDrop-1000 (Thermo Scientific). A presença de proteínas e compostos

fenólicos contaminantes foi avaliada no mesmo equipamento através da determinação das razões de absorbância nos comprimentos de onda de 260/280 nm e 260/230 nm, respectivamente, sendo consideradas amostras puras as que obtiveram razões 260/280 nm próximas de 1,8 e 260/230 entre 1,8 e 2,2.

A integridade do DNA genômico extraído foi avaliada por eletroforese em gel de agarose 0,8%.

### **3.10.3 Fragmentação do gDNA**

O DNA genômico (100 µg) de cada amostra foi fragmentado aleatoriamente utilizando-se o sonicador *Sonic Dismembrator (Model 500; Fischer Scientific)*. As amostras foram submetidas a quatro pulsos de 10 segundos com amplitude de 11% e intervalos de 30 segundos. A eficiência de sonicação foi avaliada por eletroforese em gel de agarose 2%.

## **3.11 Análise de metilação do DNA**

Com a finalidade de se avaliar o *status* de metilação do DNA de possíveis regiões promotoras de lncRNAs antissenso e intrônicos senso, realizamos principalmente duas técnicas:

### **3.11.1 Digestão do DNA metilado utilizando a endonuclease McrBC**

A endonuclease McrBC é capaz de reconhecer sequências genômicas contendo metilcitosinas em uma ou ambas as fitas do DNA e clivá-las. Essa endonuclease possui uma característica especial em relação às outras enzimas de restrição, pois apresenta uma menor especificidade no reconhecimento de sítios de ligação ao DNA ( $G/AC_m(N)40-3000(N)G/AC_m$ ), o que torna a frequência de corte dessa enzima maior. A enzima tem atividade dependente da

presença de dois sítios contendo citosinas metiladas que podem estar separados por até 3 kb de distância (separação ótima se dá entre 55 – 103 pb).

A digestão de 4 µg de amostras de DNA metilado e de DNA desmetilado (tratado com 5-AZA) foi realizada com 32 unidades da endonuclease McrBC (NEB) em tampão fornecido pela empresa (10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl<sub>2</sub> e 1 mM Dithiothreitol, pH 7,9) acrescido de 0,1 µg/ µl de BSA e 2 mM de GTP por 16 horas a 37°C. Em paralelo, uma reação sem a enzima foi realizada para controle. Após a digestão, o DNA remanescente da digestão foi quantificado por PCR em tempo real. Os valores de delta Ct foram calculados normalizando-se as medidas de Ct das amostras tratadas com a enzima com as medidas obtidas para as mesmas amostras submetidas ao protocolo sem a adição da enzima. As sequências dos *primers* utilizados na reação de PCR em tempo real estão listadas na Tabela 10.

**Tabela 10: Sequência dos *primers* utilizados para avaliação da digestão de CGIs com a enzima McrBC.**

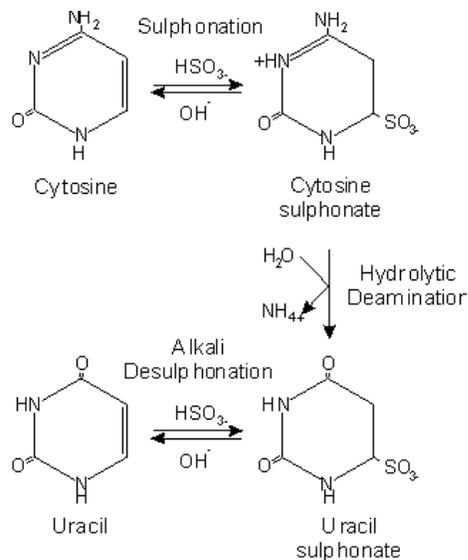
Nome do primer	Sequência de oligonucleotídeos (5' -> 3')*
CGI_IncRNA AS-APP	F: ATGGTTTTGCTGTCCAATTCA
	R: GCAGCAGAACGGCTACGAA
CGI_IncRNA AS-CELF2	F: TGTGGGCAACACCATATTGG
	R: ATGGATGTGTTGGTGCTGATGA
CGI_ACTB	F: CGCGCCGAGTCCTTAGG
	R: TCCGACCAGTGTTCCTTT

\* F = Forward; R = Reverse

### 3.11.2 Sequenciamento do DNA tratado com bissulfito de sódio

#### Conversão do gDNA por bissulfito de sódio

O tratamento com bissulfito de sódio desamina os resíduos de citosinas presentes no DNA, modificando-os a uracilas, enquanto que as 5-metilcitosinas, em geral, permanecem intactas, o que permite a distinção entre elas. A conversão de citosinas a uracilas é realizada em três reações: (1) Sulfonação; (2) Desaminação hidrolítica e (3) Dessulfonação álcali (Figura 5). Após a conversão, o DNA deve ser precipitado e pode ser clonado e sequenciado. As sequências são então alinhadas e comparadas com a sequência do DNA genômico não convertido.



**Figura 5: Reações químicas que convertem citosinas em uracilas durante o tratamento do DNA com bissulfito de sódio.**

A primeira etapa para a conversão do DNA por bissulfito de sódio consiste na desnaturação das amostras. Para isso, 2  $\mu\text{g}$  de DNA genômico, em um volume de 17  $\mu\text{L}$  é incubado com 2  $\mu\text{L}$  de NaOH 3 M e 1  $\mu\text{L}$  da solução carreadora (20 mg/mL de glicogênio) por 15 minutos a 37  $^\circ\text{C}$ , seguido de 2 minutos a 90  $^\circ\text{C}$  e gelo. A seguir, é adicionado aos tubos 208

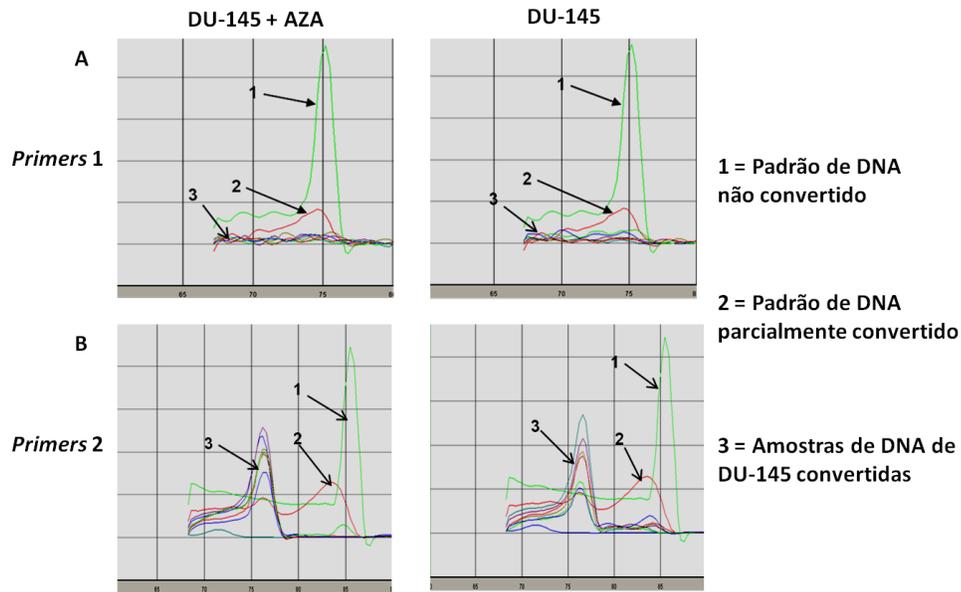
$\mu\text{L}$  da solução saturada de bissulfito de sódio fresca (5,33 M; pH 5) e 12  $\mu\text{L}$  de hidroquinona 10 mM, para prevenir oxidação do bissulfito de sódio. As amostras são incubadas por 3 horas a 70 °C em banho seco, cobertos com papel alumínio para evitar a exposição da hidroquinona à luz. Após a conversão das citosinas em uracilas sulfonadas pelo bissulfito de sódio, o DNA deve ser purificado. Para isso, 950  $\mu\text{L}$  de *Wizard Minipreps DNA Purification Resin* (Promega) foram adicionados às amostras. Os tubos foram homogeneizados em agitador mecânico por 5 a 10 minutos. Seringas de 5 mL sem o êmbolo foram conectadas às *Wizards Minicolumns* (Promega). As misturas de amostra e resina foram adicionadas às seringas com o auxílio de um pipetador automático. Os êmbolos foram recolocados nas seringas e pressionados até a completa passagem das amostras pelas colunas. Para lavagem do DNA, foram passados pelas colunas, através das seringas, 4 mL de isopropanol 80%. A seguir, as colunas foram colocadas em tubos *ependorfs* de 1,5 mL e centrifugadas a 10.000 RPM por 3 minutos para secagem. A eluição do DNA foi realizada centrifugando as colunas a 14.000 RPM por 3 minutos em novos tubos *ependorf* de 1,5 mL, com 50  $\mu\text{L}$  de água MilliQ autoclavada pré-aquecida a 80 °C. A reação de dessulfonação álcali das uracilas sulfonadas foi realizada incubando-se as amostras com 5,5  $\mu\text{L}$  de NaOH 3 M a 37 °C por 15 minutos. A precipitação do DNA convertido foi realizada incubando-se as amostras com 75  $\mu\text{L}$  de acetato de amônio 5 M; pH 7,0 por 5 minutos. A seguir, 1  $\mu\text{L}$  de glicogênio (20 mg/mL) e 350  $\mu\text{L}$  de etanol 100% gelado foram adicionados às amostras e as mesmas foram incubadas a -20 °C por 16-18 horas. A seguir, as amostras foram centrifugadas a 14.000 RPM por 30 minutos a 4 °C e lavadas com 500  $\mu\text{L}$  de etanol 70%, centrifugando-se mais uma vez os tubos a 14.000 RPM por 15 minutos a 4 °C. Os sedimentos de DNA convertido foram secos em temperatura ambiente e ressuspensos em 22  $\mu\text{L}$  água MilliQ autoclavada.

### **Eficiência de conversão do DNA por bissulfito de sódio**

A verificação da eficiência do tratamento com bissulfito de sódio, foi realizada quantificando-se por PCR em tempo real as amostras tratadas. Para isso foram utilizados dois pares de *primers* especialmente desenhados para essa finalidade.

Um dos pares de *primers* (*primer 1*) é complementar a uma sequência de DNA não convertido por bissulfito e que é rica em dinucleotídeos CpGs não metilados. Na ausência de conversão por bissulfito, ocorre amplificação de uma banda com *temperatura de melting* ( $T_m$ ) definida. Após conversão por bissulfito, não há mais complementaridade e anelamento dos *primers* e, portanto não ocorre amplificação da região.

O segundo par de *primers* (*primer 2*) flanqueia uma região do DNA rica em CpGs desmetilados. A conversão não afeta o anelamento dos *primers*, no entanto, as citosinas contidas na região do *amplicon*, quando as amostras são eficientemente tratadas com bissulfito de sódio, convertem-se a uracilas. Isto torna a *temperatura de melting* ( $T_m$ ) do *amplicon* convertido mais baixa que a temperatura do *amplicon* não convertido. Após a amplificação por PCR as *temperaturas de melting* podem ser determinadas pela curva de dissociação dos produtos e a eficiência de conversão avaliada pela comparação da área sob cada um dos picos de dissociação.



**Figura 6: Eficiência do tratamento das amostras de DNA extraídos de DU-145 com bissulfito de sódio.** Amostras de gDNA de células de DU-145 tratadas com 5-AZA e controle foram convertidas com o tratamento por bissulfito de sódio. A seguir, as amostras de gDNA de DU-145 convertidas (3), os padrões de DNA não convertido (1) e parcialmente convertido (2) foram amplificados por PCR em tempo real utilizando dois pares de *primers*. O par de *primer 1* (A) amplifica apenas gDNA não convertido. O par de *primer 2* (B) é capaz de amplificar alvos convertidos ( $T_m \approx 76^\circ\text{C}$ ) ou não ( $T_m \approx 81^\circ\text{C}$ ) por bissulfito. Notar a ausência de amplificação pelo par de *primer 1* das amostras de gDNA de DU-145 convertido (painéis superiores – seta 3) e a presença de amplificação pelo par de *primer 2* de um *amplicon* com  $T_m$  menor (painéis inferiores – seta 3) em relação a  $T_m$  do *amplicon* amplificado a partir de um padrão de DNA não convertido (painéis inferiores - seta 1).

Na Figura 6A (PCR realizada com o par de *primers 1*) não se observa amplificação de produtos para as amostras de gDNA de DU-145 convertido. Na Figura 6B, nota-se a amplificação exclusiva de bandas com  $T_m$  compatível ao produto amplificado a partir de um DNA convertido ( $T_m \approx 76^\circ\text{C}$ ). Não se observa a amplificação de produtos com  $T_m \approx 81^\circ\text{C}$  nas amostras de gDNA de DU-145 tratadas com bissulfito, indicando portanto uma alta eficiência de conversão das amostras.

### Planejamento dos *primers* para clonagem das CGIs

Após a conversão do DNA genômico por bissulfito de sódio, as CGIs testadas quanto ao seu *status* de metilação devem ser amplificadas e clonadas para posterior sequenciamento e análise. O desenho dos *primers* para amplificação do DNA modificado é um desafio, uma vez que o tratamento converte citosinas não metiladas em uracilas, as quais serão posteriormente copiadas como timinas na amplificação por PCR. Essas modificações reduzem a complexidade das sequências, o que aumenta a possibilidade de formação de auto anelamento entre os *primers* e ao mesmo tempo aumenta a possibilidade de os *primers* encontrarem múltiplos sítios de anelamento no DNA, o que levaria à produção de *amplicons* secundários. Além disso, o tratamento tornam as duas fitas do DNA não mais complementares entre si. Assim, o par de *primers* a ser desenhado deve amplificar apenas uma das fitas do DNA. Neste trabalho, para a amplificação das CGIs, os *primers* específicos (Tabela 11) foram desenhados com o auxílio do programa *Methyl Primer Express* (Applied Biosystems). Em geral, esse programa leva em consideração os critérios comuns ao desenho de *primers* para PCRs convencionais, além de desenhar *primers* que evitem a presença de sítios CpG em sua sequência, possibilitando a amplificação do DNA caso ele esteja metilado ou não. Caso não seja possível evitar sítios CpGs nas sequências dos *primers*, devido a limitações da sequência *input*, o programa atribui à posição da citosina do sítio CpG as duas possibilidades de bases. Além disso, as sequências de cada *primer* devem conter no mínimo 3 citosinas não-CpGs, para maximizar a amplificação seletiva das fitas de DNA modificadas pelo tratamento com bissulfito.

**Tabela 11: Sequência dos *primers* desenhados para clonagem das CGIs a partir do DNA convertido com bissulfato de sódio.**

Nome do primer	Sequência de oligonucleotídeos (5'→3')*
CGI_IncRNA AS-APP_BSP	F: TTTGTGTTATAGTATAGTTGTTAAAAGG
	R: CTCCTCTTATTTTTCAAATTAAC
CGI_IncRNA AS-CELF2_BSP	F: GGTTGTGGGTAATATTATATTGGG
	R: AAACACCCACAAAAACAAAAT
CGI_ACTB_BSP	F: GGAGGTTTTGTGTAGAGAAAAG
	R: CTCCTTTATCCCAATCTAA

\* F = Forward; R = Reverse

### Clonagem e sequenciamento do DNA genômico convertido por bissulfato de sódio

As CGIs analisadas foram amplificadas por PCR a partir de 50 ng de DNA genômico convertido com bissulfato de sódio. A reação de PCR foi realizada com 5 unidades da enzima *Go Taq DNA Polymerase* (Promega), 0,2 mM de dNTPs e 0,2 µM de cada *primer* em tampão recomendado pelo fabricante, que contém 1,5 mM de MgCl<sub>2</sub>.

Os produtos obtidos foram separados por eletroforese em gel de agarose 1% em TAE com 0,5 µg/mL de brometo de etídio. As bandas correspondentes aos *amplicons* específicos foram cortadas do gel e purificadas com o kit *Wisard SV Gel and PCR Clean-Up System* (Promega) conforme orientações do fabricante. Os fragmentos purificados foram clonados no plasmídeo *pGEM-T Easy* (Promega) com a enzima *T4 DNA ligase* (Promega), segundo as orientações do fabricante. A seguir, as ligações foram submetidas à diálise por 2 horas e utilizadas para eletroporação em bactérias *E. coli* DH10B eletrocompetentes. Após o crescimento das bactérias em meio LB sólido contendo ampicilina por 16 horas a 37 °C, cerca de 30 a 40 colônias de cada clonagem foram transferidas para placas de cultura de 96 poços (Corning) contendo 100 µL de meio TB líquido acrescido de ampicilina. As colônias foram crescidas por 16 horas a 37 °C.

Para o sequenciamento, cada clone de bactéria foi submetido à reação de PCR (PCR de colônia). A reação de PCR de colônia foi realizada com 5 unidades da enzima *Taq DNA Polymerase* (Bioline), 1,8 nM de dNTPs, 2,6 pmol de cada *primer* (*primers* que são complementares à região que flanqueia o sítio múltiplo de clonagem do vetor *pGEM-T Easy*), 22 µM de MgCl<sub>2</sub> em tampão recomendado pelo fabricante. O volume de 15 µL de reação foi transferido para microplacas para PCR de 96 poços (ABgene), seguida da transferência de 1-2 µL das culturas crescidas em meio TB. A reação foi realizada em termociclador GeneAmp PCR System (Applied Biosystems) nas seguintes condições: 4 minutos a 95 °C, 40 ciclos de 45 segundos a 95 °C, 45 segundos a 57 °C e 1 minuto a 72 °C, seguido por 5 minutos a 72 °C. Para a verificação da correta amplificação dos produtos inseridos no vetor, 1 µL de cada reação foi submetido a eletroforese em gel de agarose 1% em TAE com 0,5 µg/mL de brometo de etídio. Os produtos do PCR de colônia foram sequenciados com o método dideoxi no sequenciador *MegaBace* (Armeshan) utilizando para a reação de sequenciamento o kit *BigDye Terminator v3.1 Cycle Sequencing* (Applied Biosystems) e 2,5 µM do *primer* que anela no vetor.

### **Análise informática das sequências dos clones**

As sequências geradas foram analisadas com a ferramenta disponível *on line* QUMA (QUantification tool for Methylation Analysis - <http://quma.cdb.riken.jp/>) (Kumaki *et al.*, 2008). Essa ferramenta inclui muitas das funções necessárias para o processamento de dados de sequenciamento bissulfito, além de prover uma análise de qualidade das sequências analisadas. Devido à baixa complexidade das sequências após a conversão por bissulfito, erros de sequenciamento são frequentes durante a análise. Para controlar a qualidade das sequências analisadas, a ferramenta adota como parâmetros o limite inferior de 90% de identidade e o número máximo de 10 *mismatches* entre as sequências analisadas e a sequência genômica.

A conversão de citosinas por bissulfito incompleta é outro problema frequentemente encontrado nas análises de sequenciamento. Devido ao baixo nível de metilação do DNA em citosinas não CpGs no genoma de mamíferos, a presença de um certo número de citosinas não CpGs não convertidas em um único clone de sequenciamento bissulfito tem sido um indicador prático de conversão por bissulfito incompleta. Assim, para o controle de qualidade da conversão por bissulfito, o *software* adota uma eficiência mínima de 95% de conversão de citosinas (razão entre o número de citosinas não CpGs convertidas pelo número total de citosinas não CpGs presentes na sequência) e o número máximo de 5 citosinas não convertidas em sítios não CpGs como parâmetros. Outra vantagem desse *software* é a possibilidade de realizar uma análise estatística entre os perfis de metilação de dois grupos de sequências convertidas por bissulfito que se deseja comparar. A significância estatística da diferença entre os dois grupos de sequencia em cada sítio CpG é avaliado com o teste exato de Fisher, enquanto que o grupo inteiro dos sítios CpGs é avaliado com o teste Mann-Whitney U-test.

## 4 RESULTADOS

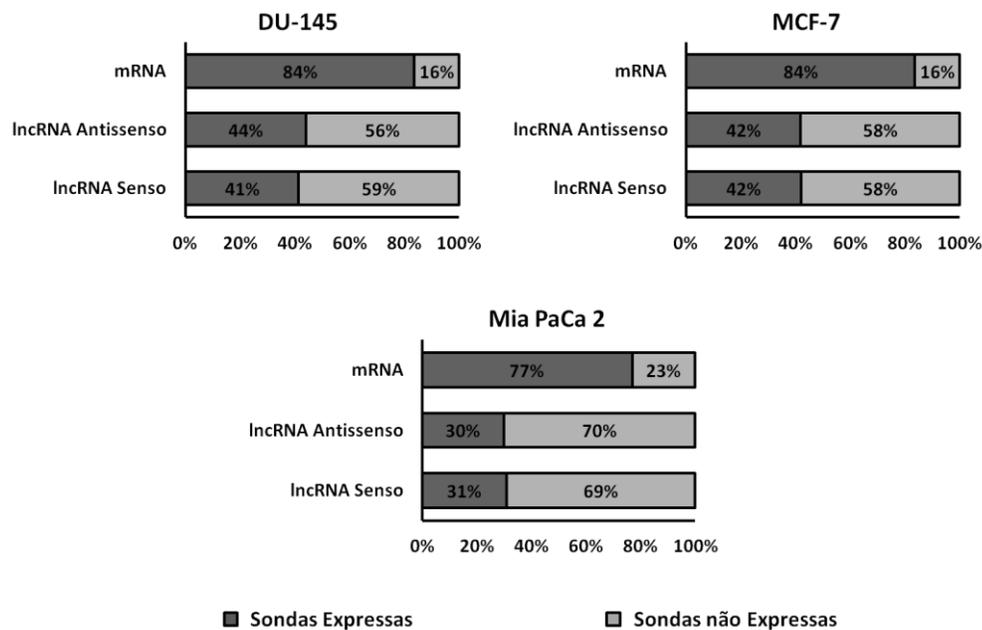
### 4.1 Expressão em larga escala de ncRNAs longos (lncRNAs) intrônicos e mRNAs codificadores de proteínas em linhagens tumorais humanas

As linhagens tumorais derivadas de tumores próstata (DU-145), mama (MCF-7) e pâncreas (Mia PaCa 2) tiveram seu padrão de expressão gênica global analisado através da hibridização de seus conteúdos de RNA com *oligoarrays* customizados em nosso laboratório e sintetizados pela Agilent. Estas lâminas utilizadas contém 13.274 sondas que medem a expressão de mRNAs codificadores de proteínas, 6.831 sondas que medem a expressão de lncRNAs com evidências de serem transcritos a partir de regiões intrônicas (lncRNAs senso) de genes codificadores de proteína com sequência curada (conjunto RefSeq) e 10.901 sondas que medem a expressão de lncRNAs com evidências de serem transcritos a partir da fita oposta de genes codificadores de proteínas (lncRNAs antissenso).

A estratégia de hibridização está mostrada na figura 4 da subseção 3.5.3 da seção de Materiais e Métodos. Após a extração dos dados, os valores de intensidade foram filtrados. Para o RNA ser considerado expresso na linhagem em análise, o sinal de intensidade da sonda deveria estar significativamente acima do *background* da lâmina em três das quatro réplicas de medidas de expressão. A Tabela 12 lista o número de sondas de cada tipo consideradas expressas nas linhagens celulares. A Figura 7 mostra a proporção de sondas expressas e não expressas, nas células de cada uma das linhagens estudadas, em relação ao total de sondas na lâmina de *oligoarray*, por classe de transcrito.

Tabela 12: Número de sondas expressas nas linhagens celulares estudadas.

Tipos de Sondas	Lâmina 44K	DU-145		MCF-7		Mia PaCa 2	
		Expressos	Não Expressos	Expressos	Não Expressos	Expressos	Não Expressos
mRNA	13274	11093	2181	11099	2175	10208	3066
ncRNA Antissenso	10901	4815	6086	4565	6336	3262	7639
ncRNA Senso	6831	2825	4006	2875	3956	2113	4718
Outros	57	37	20	41	16	35	22
<b>Total</b>	<b>31063</b>	<b>18770</b>	<b>12293</b>	<b>18580</b>	<b>12483</b>	<b>15618</b>	<b>15445</b>



**Figura 7: Proporção de sondas expressas e não expressas, nas três linhagens em estudo, em relação ao total de sondas na lâmina de oligonucleotídeos, por classe de transcrito.** Foram considerados RNAs expressos aqueles que possuíam sinal de intensidade significativamente acima do *background* da lâmina em três das quatro réplicas medidas. Em cinza escuro está representada a proporção de sondas consideradas expressas na análise e em cinza claro, a proporção de sondas consideradas não expressas na análise.

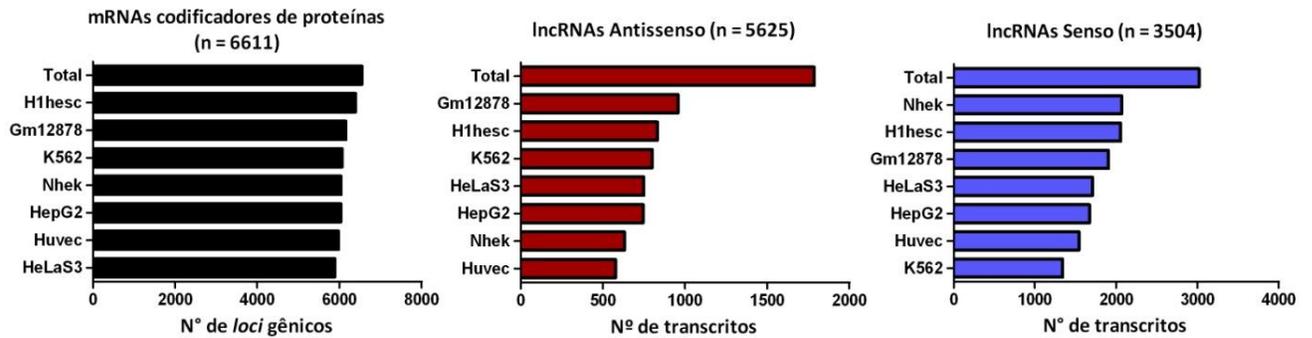
Foi detectada a expressão de 10 mil a 11 mil sondas que medem mRNAs codificadores de proteínas em cada uma das linhagens. Esses números representam a expressão de cerca de 77% a 84% das sondas que medem mRNAs presentes na lâmina de oligonucleotídeos (Figura 7). As sondas detectadas representam 6.278 (DU-145), 6.255 (MCF-7) e 6.191 (Mia PaCa 2) diferentes *loci* gênicos expressos em cada uma das linhagens. Em relação aos lncRNAs com orientação antissenso ao *locus* codificador de proteínas hospedeiro, foram detectadas 4.815 (DU-145, 44% do total de sondas presentes na lâmina), 4.565 (MCF-7, 42% do total de sondas presentes na lâmina) e 3.262 (Mia PaCa 2, 30% do total de sondas presentes na lâmina) sondas positivas em cada uma das linhagens (Tabela 12 e Figura 7). Além disso, foram detectados lncRNAs cuja expressão ocorre a partir de uma região intrônica, na mesma fita que o mRNA pertencente ao *locus* do gene codificador de proteínas hospedeiro (lncRNAs senso): 2.825 (41% do total presente na lâmina) em DU-145, 2.875 (42% do total de sondas presentes na lâmina) em MCF-7 e 2.113 (31% do total de sondas presentes na lâmina) em Mia PaCa 2 (Tabela 12 e Figura 7).

Essa análise indicou uma menor proporção de lncRNAs intrônicos expressos em cada uma das linhagens em relação a proporção de mRNAs. Está descrito na literatura que RNAs não codificadores apresentam em geral perfis de expressão tecido e/ou célula específico (Ravasi *et al.*, 2006; Mercer *et al.*, 2008; Cabili *et al.*, 2011). O resultado descrito acima corrobora a ideia de que lncRNAs originados em regiões intrônicas possuem características de expressão celular/tecidual comparáveis às observadas para lncRNAs intergênicos, além de corroborar também resultados prévios de nosso grupo utilizando amostras de tecido e a mesma plataforma de oligonucleotídeos (Nakaya *et al.*, 2007).

Os lncRNAs intrônicos antissenso e senso expressos em pelo menos uma das três linhagens tiveram seu potencial codificador de proteínas avaliado pelo programa CPC – *Coding Potential Calculator* (Kong *et al.*, 2007). Para ambas as classes de lncRNAs em estudo, mais de

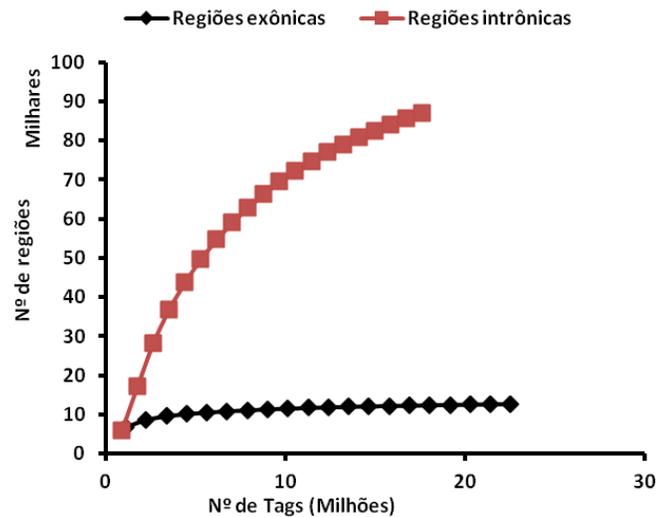
99% dos transcritos expressos em nossas análises possuem pouco ou nenhum potencial codificador: 5.610 dos 5.625 lncRNAs antissenso e 3.498 dos 3.504 lncRNAs senso.

Com o intuito de obter evidências adicionais de que os lncRNAs antissenso e senso considerados detectados em nossas análises são transcritos reais e não resultantes da hibridização inespecífica com as sondas da lâmina de oligonucleotídeos, as coordenadas genômicas dos 9.129 lncRNAs considerados expressos em pelo menos uma das três linhagens estudadas foram cruzadas com as coordenadas genômicas disponíveis publicamente de RNAs detectados pela metodologia de RNA-seq em 7 bibliotecas de diferentes linhagens celulares (*ENCODE Caltech RNA-seq Libraries*, <http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=268692585&c=chr1&g=wgEncodeCaltechRnaSeq>). As 7 bibliotecas utilizadas para comparação com os dados detectados por nossa lâmina de *oligoarray* foram geradas a partir de RNA poli(A)+ obtidos de 7 linhagens celulares, preservando a informação da orientação original dos transcritos. A Figura 8 mostra o resultado da análise de sobreposição de coordenadas genômicas dos 6.611 mRNAs, dos 5.625 lncRNAs antissenso e dos 3.504 lncRNAs senso expressos em pelo menos uma das três linhagens estudadas com as coordenadas genômicas dos RNAs detectados nas 7 bibliotecas analisadas pelo projeto *ENCODE Caltech RNA-seq Libraries*.



**Figura 8: Número de lncRNAs intrônicos antissenso e senso e mRNAs detectados em pelo menos uma das três linhagens estudadas (DU-145, MCF-7 e Mia PaCa 2) e identificados em outras linhagens celulares e por outra metodologia.** As coordenadas genômicas dos 6.611 mRNAs codificadores de proteínas, 5.625 lncRNAs antissenso e 3.504 lncRNAs senso foram cruzadas com as coordenadas genômicas de RNAs detectados em sete linhagens celulares diferentes por RNA-seq (ENCODE Caltech RNA-seq Libraries). Na figura estão plotados os números de transcritos detectados em nossos experimentos com *oligoarrays* e que também foram detectados em pelo menos uma das sete linhagens do projeto ENCODE Caltech RNA-seq Libraries, por classe de RNA.

Na figura acima podemos notar que a quase totalidade (99%) dos mRNAs expressos em pelo menos uma das três linhagens também apresentam expressão em pelo menos uma das 7 linhagens analisadas nas bibliotecas de RNA-seq Caltech. Para os lncRNAs antissenso, essa sobreposição foi de 32% e para os lncRNAs senso, 86%. A menor sobreposição de coordenadas encontrada para os lncRNAs em relação aos mRNAs pode ser explicada por algumas razões. Como já demonstrado na literatura (Cabili *et al.*, 2011), e como será discutido mais adiante em nossas análises (Figura 10), os ncRNAs possuem abundância de expressão mais baixa que os mRNAs. É sabido, no entanto, que transcritos com menor abundância de expressão são mais difíceis de serem cobertos por experimentos de RNA-seq que transcritos com abundância de expressão maiores. Assim, grandes esforços de sequenciamento devem ser realizados para amostrar a totalidade dessa classe de RNAs. Realizamos uma análise para determinar a cobertura do transcriptoma exônico e intrônico nas 7 bibliotecas de RNA-seq utilizadas (Figura 9).



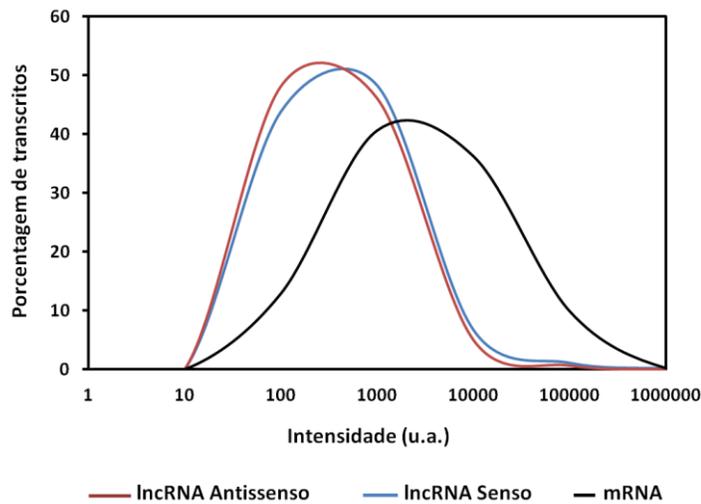
**Figura 9: Curva de saturação da cobertura de sequenciamento das bibliotecas de RNA-seq do projeto ENCODE *Caltech RNA-seq Libraries*.** O gráfico representa a curva de saturação das *tags* sequenciadas por RNA-seq a partir das 7 bibliotecas de cDNA do projeto ENCODE *Caltech RNA-seq Libraries*. O número de regiões exônicas (curva preta) representa a contagem de GENE IDs e o número de regiões intrônicas representa a contagem de íntrons pertencentes a GENE IDs (curva vermelha). No eixo X estão representados os números de *tags* referentes às 7 bibliotecas do projeto.

Observamos a saturação do número de regiões codificadoras detectadas nestas bibliotecas, que correspondem a cerca de 12.600 genes (Figura 9, curva preta). Por outro lado, observamos que o número de regiões intrônicas não se encontra saturada na profundidade de cobertura atingida com o sequenciamento destas bibliotecas. Mesmo após sequenciar cerca de 40 milhões de *tags*, a curva cumulativa indica que o número de regiões intrônicas detectadas encontra-se longe da saturação (Figura 9, linha vermelha). A saturação alcançada nos experimentos de RNA-seq foi insuficiente para detectar o conjunto de transcritos intrônicos expressos nas 7 linhagens. Este resultado explica, ao menos em parte, a menor porcentagem de sobreposição observada entre os lncRNAs detectados em nossos experimentos e transcritos intrônicos detectados nos experimentos de RNA-seq em comparação a sobreposição observada para os mRNAs. Outra possibilidade de causa para a baixa sobreposição de coordenadas entre os lncRNAs intrônicos e RNAs sequenciados a partir das 7 bibliotecas do projeto ENCODE *Caltech*

*RNA-seq Libraries*, é o fato de que os ncRNAs possuem um padrão de expressão com maior especificidade em relação ao tipo de tecido/tipo celular que os mRNAs em geral (Cabili *et al.*, 2011). Este fator também deve contribuir para a menor proporção de lncRNAs em relação aos mRNAs expressos nas 3 linhagens testadas que são confirmados nas 7 bibliotecas de RNA-seq analisadas.

#### **4.2 LncRNAs intrônicos são expressos em níveis mais baixos do que mRNAs codificadores de proteínas**

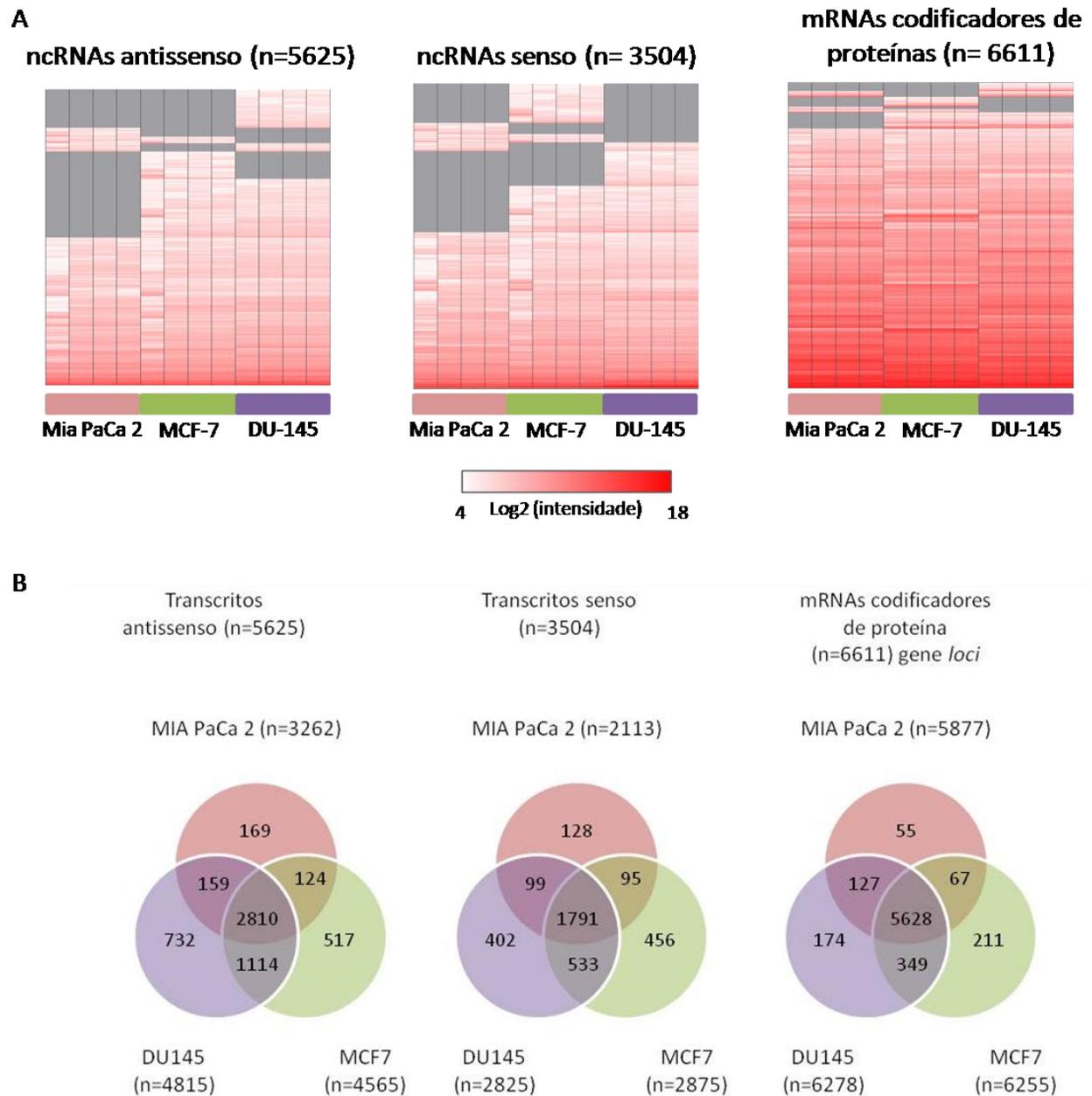
Comparando a distribuição das intensidades de sinal das sondas consideradas expressas em pelo menos uma das três linhagens em estudo, pudemos observar que os lncRNAs apresentam intensidades de expressão menores que os mRNAs, sendo as medianas de intensidade de expressão dos lncRNAs antissenso de 90 u.a. e a dos lncRNAs senso de 98 u.a. A mediana de intensidade de expressão encontrada para os mRNAs codificadores de proteínas foi de 785 u.a. (Figura 10).



**Figura 10: Distribuição das intensidades de sinal das sondas complementares a lncRNAs intrônicos antissenso ou senso e mRNAs detectadas nas linhagens DU-145, MCF-7 e MIA PaCa 2.** Após extração dos dados, todas as sondas consideradas detectadas em pelo menos uma das três linhagens (com sinal de expressão significativamente acima do *background* da lâmina em três das quatro réplicas de cada linhagem) tiveram suas intensidades normalizadas por quantil. Para os dados normalizados foram calculadas as frequências de cada faixa de intensidade. As frequências por intensidade dos 3.504 lncRNAs senso, 5.625 lncRNAs antissenso e 11.879 sondas de mRNAs detectados em pelo menos uma das três linhagens estão mostradas pelas linhas azul, vermelha e preta, respectivamente.

#### 4.3 lncRNAs intrônicos apresentam padrão de expressão mais linhagem/tecido específico do que mRNAs codificadores de proteínas

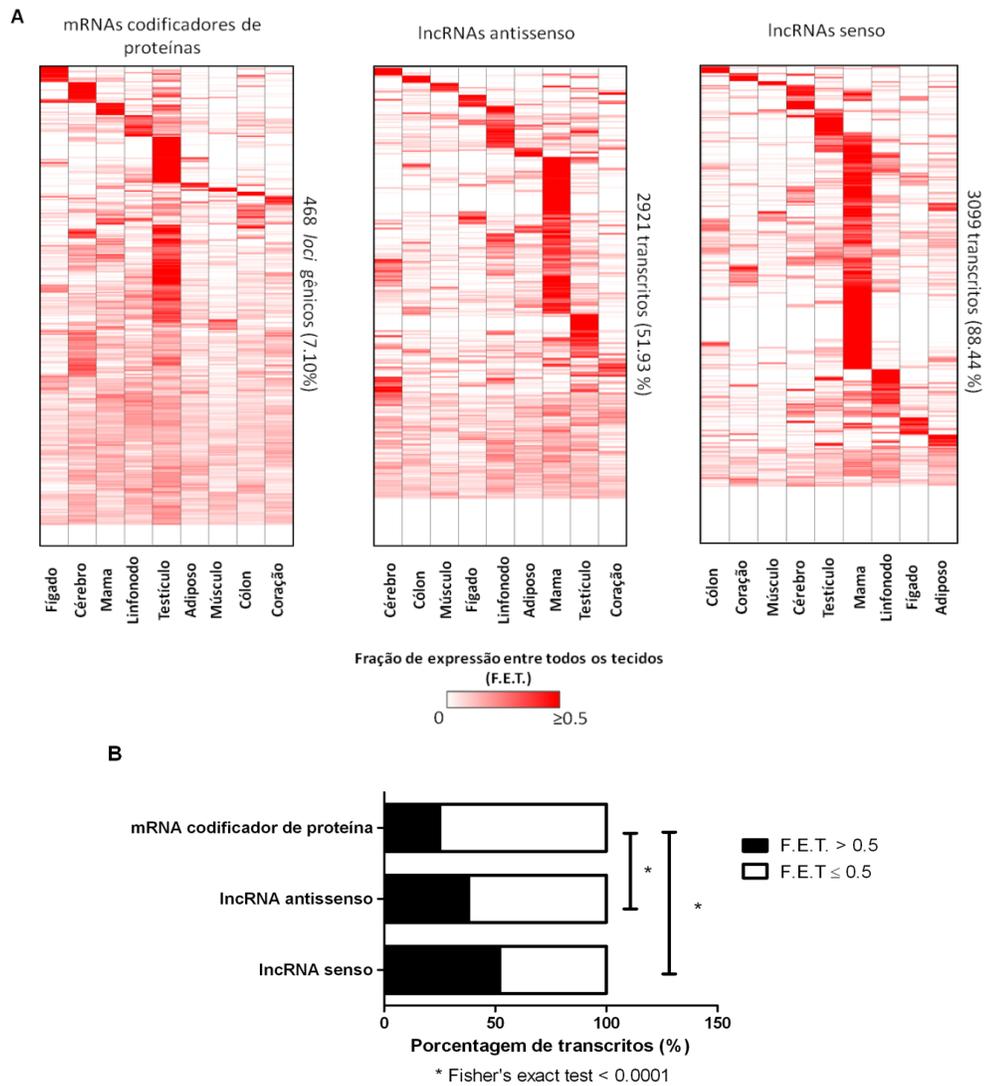
Realizamos o agrupamento dos dados de intensidade dos lncRNAs intrônicos para investigar o seu padrão de expressão entre as linhagens analisadas (Figura 11). Podemos observar uma menor sobreposição entre os lncRNAs expressos em pelo menos uma das três linhagens em comparação aos mRNAs codificadores de proteínas, o que demonstra que os lncRNAs intrônicos são expressos de uma maneira mais linhagem-específica do que os mRNAs codificadores de proteínas. Do total de transcritos de cada classe de RNA detectados em pelo menos uma das três linhagens, apenas 72% dos lncRNAs senso (2.518/3.504) e 75% dos lncRNAs antissenso (4.207/5.625) são expressos em mais do que uma linhagem. Por outro lado, a maior parte dos mRNAs (6.171/6.611, 93%) foi detectada em mais do que um tipo celular.



**Figura 11: Especificidade de expressão de lncRNAs intrônicos antissenso e senso e mRNAs codificadores de proteínas entre as linhagens DU-145, MCF-7 e Mia PaCa 2.** No painel **A** estão agrupadas as intensidades de expressão detectadas para 5.625 lncRNAs antissenso, 3.504 lncRNAs senso e 6.611 mRNAs expressos em pelo menos uma das três linhagens analisadas. Nas matrizes de expressão, cada linha representa um transcrito individual. A escala de cor representa o logaritmo na base 2 dos valores de intensidade de expressão normalizados por quantil entre as três linhagens. A cor cinza representa a ausência de expressão daquele transcrito naquela linhagem. No painel **B** estão os diagramas de venn mostrando a sobreposição de transcritos expressos entre as linhagens DU-145, MCF-7 e Mia PaCa 2, por classe de transcrito analisada.

Com o objetivo de explorar melhor o padrão de expressão tecidual dos lncRNAs intrônicos, foi realizada uma análise de abundância de expressão relativa de 6.488 RNAs expressos em nosso estudo, com um *dataset* de RNA-seq disponível publicamente (RNA-seq Burge) (Wang *et al.*, 2008) composto por nove bibliotecas de diferentes tecidos. Nesta análise, as coordenadas genômicas referentes aos RNAs não codificadores intrônicos longos antissenso, senso e codificadores de proteínas do nosso conjunto de expressos foi cruzada com as coordenadas genômicas dos *reads* de RNA-seq das nove bibliotecas de cDNA originadas de diferentes tecidos. Para cada transcrito representado em nossa lâmina de oligonucleotídeos foi determinada a sua expressão em cada uma das bibliotecas de RNA-seq, representado pelo número de *reads* por quilobase por milhão (RPKM) (Mortazavi *et al.*, 2008). Em seguida, a expressão relativa de cada transcrito entre as 9 bibliotecas de RNA-seq foi determinado, e representado como uma fração da expressão total (F.E.T), como proposto por (Cabili *et al.*, 2011) e descrito na seção de Materiais e Métodos (ver subseção 3.5.5). Nesta análise, foram usados apenas RNAs detectados em pelo menos uma de nossas três linhagens e que possuíam expressão menor ou igual a 1,5 RPKM nas nove bibliotecas do *dataset* de RNA-seq. Utilizamos esse limiar de intensidade para incluir na análise apenas mRNAs com nível de expressão semelhante a dos lncRNAs. Isso foi feito para assegurarmos que as diferenças de expressão relativa de lncRNAs e mRNAs observadas entre os nove tecidos presentes no *dataset* de RNA-seq não eram devido à menor expressão relativa dos lncRNAs em relação aos mRNAs. Na Figura 12A estão representadas as frações de expressão de 6.488 RNAs (2.921 lncRNAs antissenso, 3.099 lncRNAs senso e 468 mRNAs de genes codificadores de proteínas) detectados em pelo menos uma das três linhagens e que possuíam expressão menor ou igual a 1,5 RPKM nas nove bibliotecas de RNA-seq. Baseado nesse agrupamento pode-se observar que a abundância relativa

de transcritos não codificadores de proteínas exibem um perfil mais tecido-específico que os transcritos codificadores de proteínas.



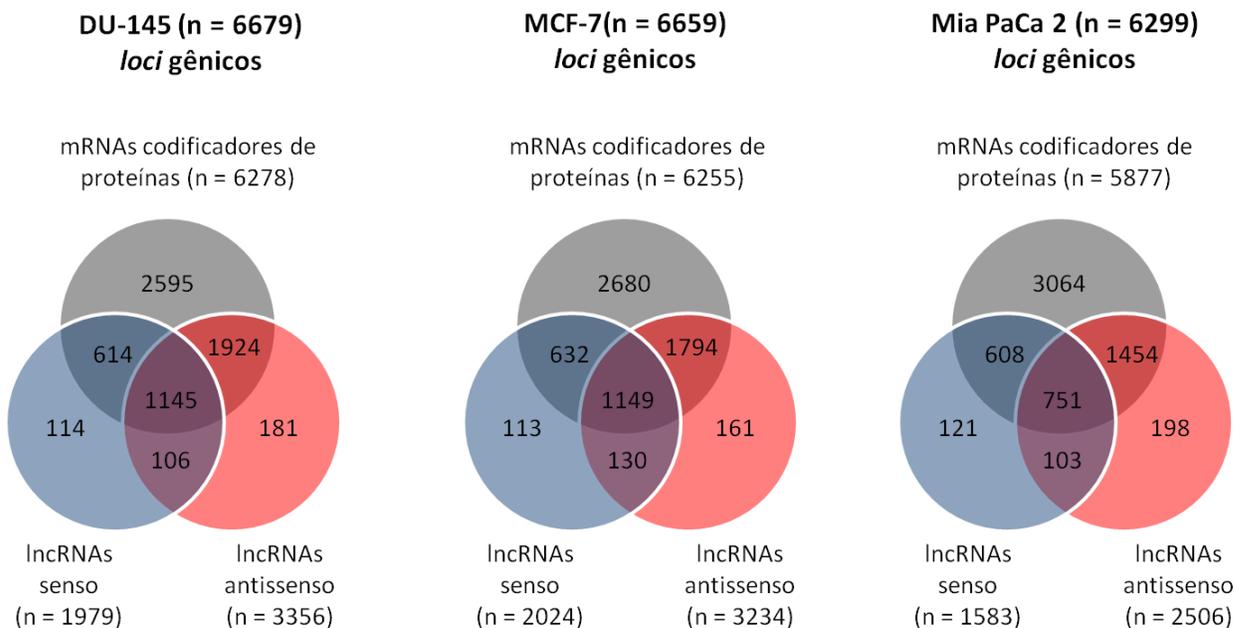
**Figura 12: Abundância relativa da expressão de lncRNAs intronicos antissenso, senso e de mRNAs codificadores de proteínas entre nove tecidos diferentes.** No painel **A** estão agrupados os 2.921 lncRNAs antissenso, 3.099 lncRNAs senso e 468 mRNAs codificadores de proteínas expressos em pelo menos uma das três linhagens em nosso estudo e que possuem intensidade de expressão menor que 1,5 RKPM em um *dataset* composto por RNAs detectados por RNA-seq em nove tecidos saudáveis diferentes. As porcentagens se referem ao número de transcritos cuja expressão é menor ou igual a 1,5 RKPM do total de transcritos expressos em pelo menos uma das linhagens DU-145, MCF-7 e Mia PaCa 2. Nas matrizes, cada linha representa um transcrito individual e cada coluna um tecido presente no *dataset* de RNA-seq. A escala de cor representa a fração de expressão entre os nove tecidos (F.E.T.), onde branco significa fração zero, ou seja, naquele tecido o transcrito não é expresso e vermelho forte significa fração maior que 0,5, ou seja, aquele tecido contribui para mais da metade da expressão total do transcrito nos nove tecidos. No painel **B** estão plotadas as porcentagens de transcritos com F.E.T. > 0,5 por classe de RNA. As proporções observadas para os lncRNAs antissenso e senso são significativamente diferentes em relação à proporção observada para os mRNAs codificadores de proteínas (Fisher's exact test;  $p$ -value < 0,0001).

Na Figura 12B estão plotadas as proporções de transcritos em cada classe com abundância de expressão tecido-específica, definido por um valor de F.E.T  $> 0,5$  em um tecido (Cabili *et al.*, 2011). Entre os lncRNAs, 52% dos transcritos intrônicos senso e 38% dos intrônicos antissenso apresentam abundância de expressão tecido específica, enquanto que entre os mRNAs codificadores de proteínas, apenas 25% apresentam essa característica (Figura 12B). Essas diferenças nas abundâncias relativas de expressão entre os nove tecidos analisados no *dataset* de RNA-seq entre lncRNAs e mRNAs é estatisticamente significativa (*Fisher's exact test*; *p-value*  $< 0,0001$ ). Além disso, pode-se afirmar que essas diferenças não são resultantes dos baixos níveis de expressão dos lncRNAs uma vez que foram utilizados na análise apenas transcritos codificadores e não codificadores de proteínas presentes no *dataset* de RNA-seq que estavam dentro de uma mesma faixa de expressão (RPKM).

#### **4.4 Correlação de expressão entre lncRNAs intrônicos e mRNAs codificadores de proteínas**

Um questionamento a respeito dos lncRNAs senso expressos de regiões intrônicas do genoma é se estes são realmente transcritos independentes dos mRNAs codificadores de proteínas, possuindo funções próprias dentro da célula, ou se representam a medida de pré-mRNAs ou de produtos de *splicing* desses pré-mRNAs que em um momento posterior serão degradados pela célula. Embora tenhamos selecionado frações de RNA poli(A)<sup>+</sup>, utilizando *primers* oligo-dT, para o preparo das amostras de cRNA que foram hibridizadas (subseção 3.5.2. da seção de Materiais e Métodos), existe a possibilidade de que trechos ricos em adenina, dentro de íntrons, tenham servido como sítios internos de anelamento (*internal priming*), possibilitando a transcrição reversa destas mensagens (Furuno *et al.*, 2006). Por outro lado, já foi documentada a existência de ncRNAs intrônicos com a mesma orientação do mRNA expresso no locus gênico e

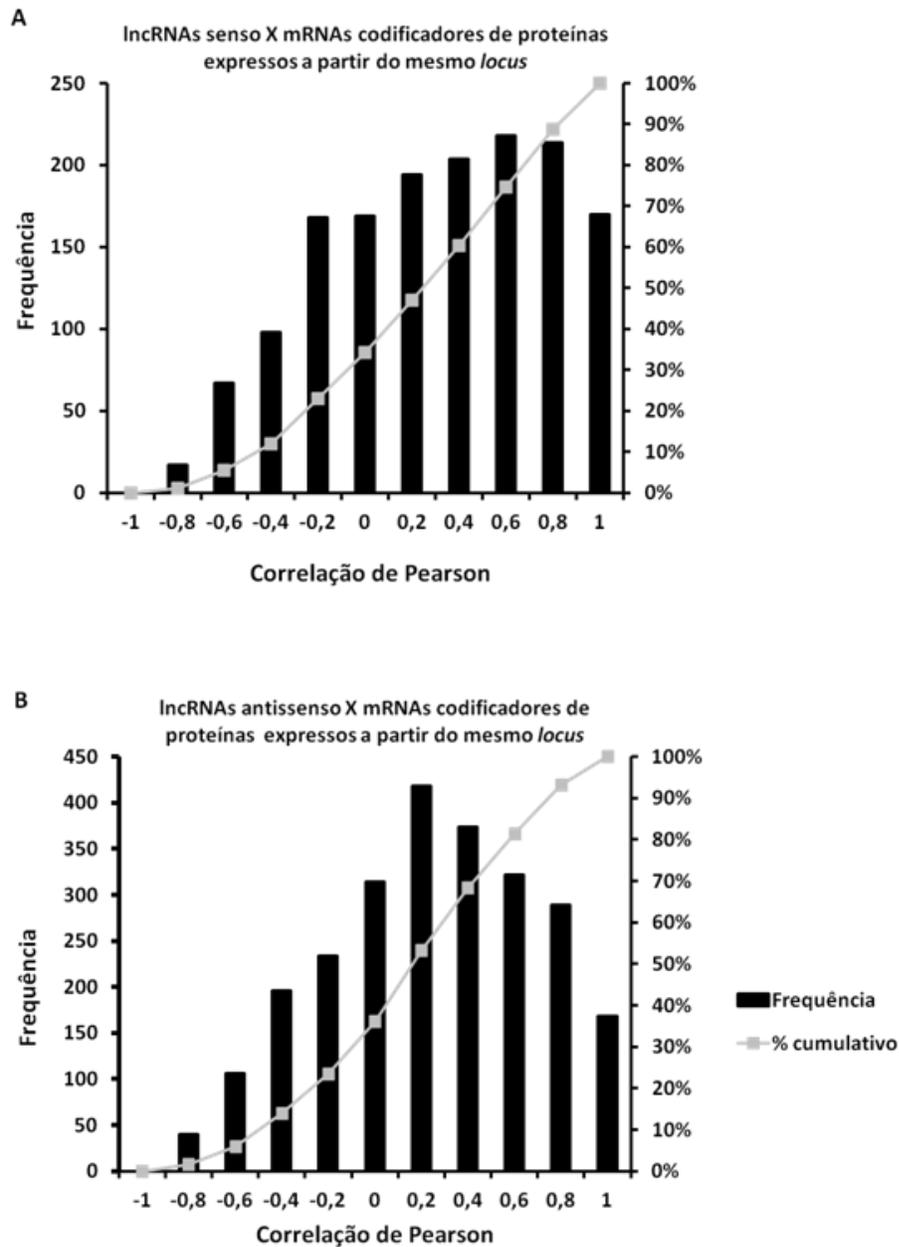
que exercem função na célula (Hill *et al.*, 2006; Heo e Sung, 2011). Outro questionamento geral sobre os lncRNAs intrônicos, tanto com orientação senso quanto antissenso, é se eles são produtos de transcrição específica e regulada ou, alternativamente, são produtos espúrios de uma transcrição inespecífica em regiões de cromatina aberta para a transcrição dos mRNAs codificadores de proteínas daqueles *loci*. Para abordar e investigar essas possibilidades de uma maneira global foi feita uma análise para identificar quantos e quais lncRNAs intrônicos são transcritos de forma independente do mRNA de seu *locus* hospedeiro. Para isso, foi analisada a associação entre a expressão de mRNAs e lncRNAs em 6.679 *loci* gênicos em DU-145, 6.659 em MCF-7 e 6.299 em Mia PaCa 2 (Figura 13).



**Figura 13: Transcrição de lncRNAs intrônicos antissenso e senso concomitante com a expressão de mRNAs no mesmo locus gênico.** Os diagramas de venn apresentam os números de *loci* gênicos com sobreposição de expressão entre lncRNAs e mRNAs codificadores de proteínas em cada linhagem estudada. Foram analisados 6.679, 6.659 e 6.299 *loci* gênicos em DU-145, MCF-7 e Mia PaCa 2, respectivamente.

Como pode-se observar na Figura 13 acima, para uma parte dos *loci* gênicos interrogados (40% a 49% do total de *loci* detectados) detectou-se exclusivamente o mRNA codificador de proteína. Para uma pequena fração dos *loci* gênicos interrogados nas três linhagens, detectou-se apenas a expressão de lncRNAs intrônicos senso ou antisenso (6% a 7 % do total de *loci* detectados). A ausência de detecção do mRNA nesses *loci* é um forte indício de que ao menos nesses casos os lncRNAs intrônicos sejam originados a partir de transcrição independente.

Para 45% a 55% do total de *loci* em que se detectou atividade transcricional, detectou-se simultaneamente a expressão de ao menos um lncRNA intrônico e do mRNA codificador de proteína. A fim de investigar a existência de associação entre a expressão do mRNA e lncRNA intrônico no mesmo *locus*, testou-se a correlação entre a expressão destes transcritos analisando os dados das 3 linhagens (Figura 14). Foram analisados apenas lncRNAs cuja expressão do par mRNA/lncRNA foi detectada nas três linhagens, resultando em um total de 1.519 lncRNAs senso e 2.461 lncRNAs antisenso. Esse critério foi utilizado com a finalidade de se obter uma maior confiabilidade na interpretação dos resultados, uma vez que as correlações calculadas, ou a ausência das mesmas, foram determinadas em linhagens distintas e não entre réplicas da mesma linhagem. Com essa análise foi possível observar que a maioria, tanto dos lncRNAs senso (69%) como dos lncRNAs antisenso (76%), possui uma baixa correlação de expressão ( $-0,6 < r < 0,6$ ) com os respectivos mRNAs expressos a partir do mesmo *locus*. Este resultado sugere que a maior parte dos lncRNAs intrônicos deva ser processada e acumulada de maneira distinta do mRNA codificador de proteínas expresso no mesmo *locus*. Essa informação é relevante, principalmente para os lncRNAs intrônicos senso, uma vez que a baixa correlação observada para a maioria dos transcritos dessa classe sugere que esse conjunto de RNAs não represente apenas pré-mRNAs ou produtos do processamento de mRNAs (*splicing*).



**Figura 14: Correlação de expressão entre IncRNAs intrônicos (senso e antissenso) e mRNAs codificadores de proteínas expressos a partir do mesmo locus gênico.** Os valores de intensidade de expressão medidos nos experimentos com *oligoarrays* para os IncRNAs e mRNAs foram normalizados por quantil. Os coeficientes de correlação de *Pearson* entre os valores de intensidade normalizados dos IncRNAs e dos mRNAs expressos a partir do mesmo locus gênico foram calculados para aqueles transcritos expressos nas três linhagens celulares em estudo (DU-145, MCF-7 e Mia PaCa 2). Nos gráficos estão plotadas as frequências de IncRNAs por intervalo de coeficiente de correlação de *Pearson* ( $r$ ) calculado e as proporções cumulativas de transcritos em cada intervalo. (A) 1.519 IncRNAs senso e (B) 2.461 IncRNAs antissenso.

Uma pequena fração dos lncRNAs intrônicos apresentou correlação de expressão com o mRNA codificador de proteína do mesmo *locus*. A fração de lncRNAs que apresentaram correlação positiva ( $r \geq 0,6$ ), 25% para os senso e 19% para os antissensos, ou negativa ( $r \leq -0,6$ ), 6% para os senso e antissensos pode sugerir que um subconjunto dessas classes de lncRNAs tenha a transcrição corregulada com a transcrição do mRNA expresso a partir do mesmo *locus* gênico. Esta correlação de expressão levanta a hipótese de que este subconjunto de lncRNAs possa exercer algum tipo de papel na modulação *in cis* da transcrição e/ou na estabilidade dos mRNAs. Embora não tenha sido observado significância estatística (teste  $\chi^2$ ,  $p\text{-value} < 0,3$ ), a maior proporção de pares mRNAs/lncRNAs intrônicos senso correlacionados positivamente ( $r \geq 0,6$ ) do que pares mRNAs/lncRNAs antissenso (25% versus 19%,) sugere que os primeiros incluam pré-mRNAs ou *intron lariats* resultantes do *splicing*, cuja abundância reflete o nível de expressão do mRNA codificador.

O conjunto de mRNAs que possuem lncRNAs correlacionadamente expressos a partir de seu *locus* foi usado para uma análise de enriquecimento de categorias de *Gene Ontology* (GO) (Ashburner *et al.*, 2000) usando o *software* disponível *on line* DAVID (*Database for Annotation, Visualization and Integrated Discovery*) (Huang *et al.*, 2009) . Entre os 468 mRNAs codificadores de proteína cuja expressão está correlacionada ( $r \leq -0,6$  e  $r \geq 0,6$ ) com a expressão de lncRNAs intrônicos senso do mesmo *locus* gênico observa-se o enriquecimento de genes pertencentes aos termos GO “*muscle organ development*”, “*muscle cell differentiation*”, “*anatomical structure morphogenesis*”, “*developmental process*” e “*protein binding*” (Tabela 13).

**Tabela 13: Categorias de GO enriquecidas entre os mRNAs codificadores de proteína expressos nas três linhagens e que possuem alta correlação ( $r \leq -0,6$  e  $r \geq 0,6$ ) de expressão com lncRNAs intrônicos senso transcritos a partir de seus loci gênicos.**

<i>Categoria</i>	<i>GO ID</i>	<i>GO Name</i>	<i>Nº de genes</i>	<i>Enriquecimento</i>	<i>p-value</i>
GOTERM_BP	GO:0007517	muscle organ development	12	2,2	0,003
GOTERM_BP	GO:0009653	anatomical structure morphogenesis	46	1,4	0,004
GOTERM_BP	GO:0042692	muscle cell differentiation	10	2,3	0,006
GOTERM_BP	GO:0032502	developmental process	95	1,2	0,009
GOTERM_MF	GO:0005515	protein binding	263	1,1	0,008

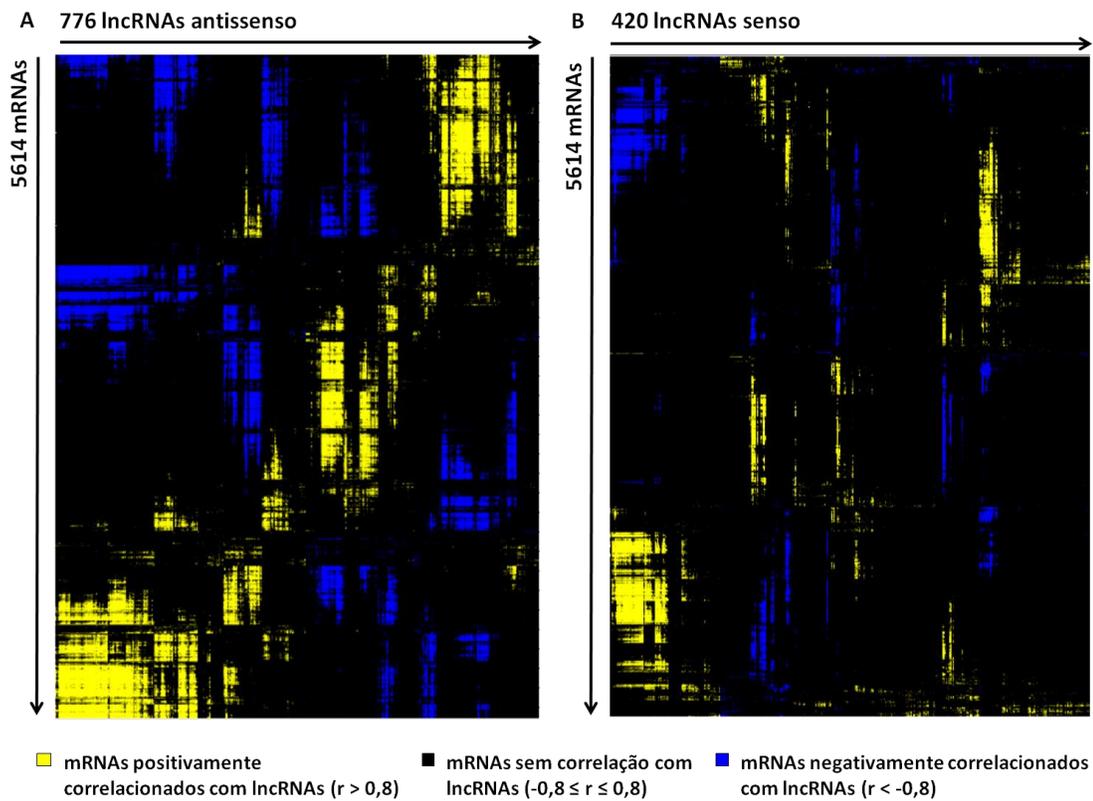
Para os 603 mRNAs codificadores de proteína que possuem correlação de expressão com lncRNAs antissenso expressos a partir de seu locus gênicos, a mesma análise indicou o enriquecimento de genes referentes aos termos GO “*actin filament-based process*”, “*regulation of cellular metabolic process*”, “*regulation of cellular process*”, “*metallopeptidase activity*” e “*nucleus*” (Tabela 14).

**Tabela 14: Categorias de GO enriquecidas entre os mRNAs codificadores de proteína expressos nas três linhagens e que possuem alta correlação ( $r \leq -0,6$  e  $r \geq 0,6$ ) de expressão com lncRNAs antissenso transcritos a partir da fita oposta ao seus loci gênicos.**

<i>Categoria</i>	<i>GO ID</i>	<i>GO Name</i>	<i>Nº de genes</i>	<i>Enriquecimento</i>	<i>p-value</i>
GOTERM_BP	GO:0065007	biological regulation	266	1,1	0,008
GOTERM_BP	GO:0030029	actin filament-based process	18	1,8	0,008
GOTERM_BP	GO:0031323	regulation of cellular metabolic process	147	1,2	0,008
GOTERM_BP	GO:0080090	regulation of primary metabolic process	139	1,2	0,009
GOTERM_BP	GO:0050794	regulation of cellular process	246	1,1	0,009
GOTERM_CC	GO:0005634	nucleus	226	1,1	0,006
GOTERM_MF	GO:0008237	metallopeptidase activity	15	2,0	0,008

Uma vez que a maioria dos lncRNAs intrônicos senso e antissenso não possuem expressão correlacionada com a expressão do mRNA transcrito a partir do mesmo locus nós fomos explorar a possibilidade de os lncRNAs apresentarem correlação de expressão com mRNAs de outros loci gênicos (análise de correlação em *trans* ou análise “*guilt by association*”)

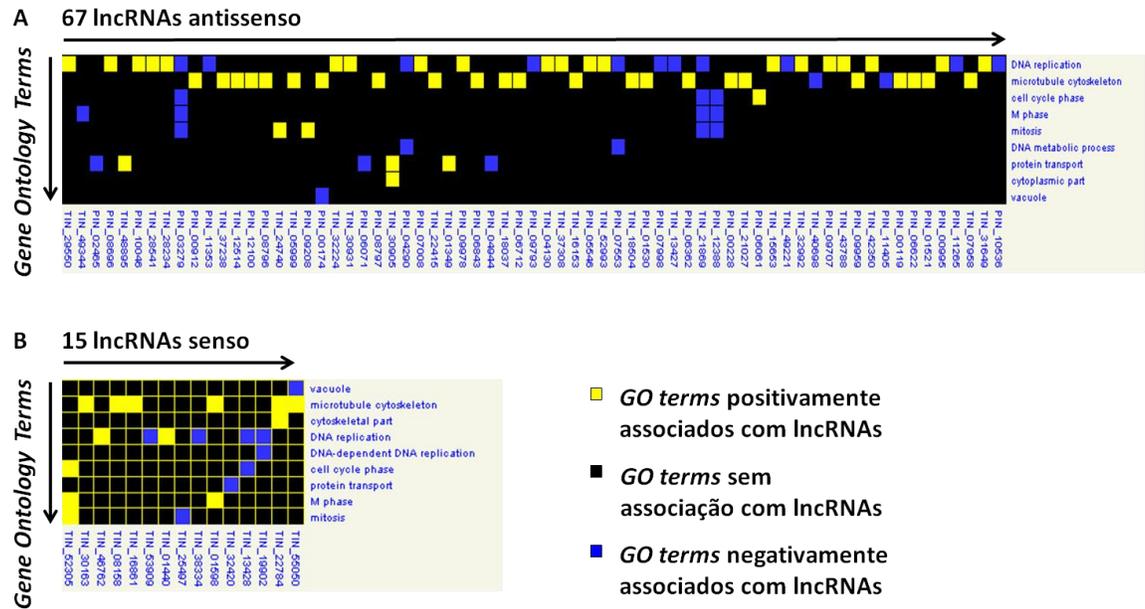
(Guttman *et al.*, 2009; Hung *et al.*, 2011). Além disso, fomos avaliar se esses mRNAs de expressão correlacionada com a expressão de lncRNAs fazem parte de alguma categoria de *Gene Ontology* enriquecida em nossa análise. Para esta análise foram utilizados os valores de intensidade de expressão normalizados por quantil de todos os 5.614 mRNAs codificadores de proteínas expressos nas três linhagens em estudo e dos 20% mais expressos lncRNAs em cada linhagem, e que estavam expressos nas três linhagens. Além disso, para a análise dos lncRNAs intrônicos senso, foram retirados os transcritos que apresentavam correlação positiva ( $r \geq 0,5$ ) com os mRNAs expressos a partir do mesmo *locus*, uma vez que esses transcritos possuem uma maior chance de serem pré-mRNAs. Assim, na análise foram computados 776 lncRNAs antissenso e 420 lncRNAs senso. Os coeficientes de correlação de *Pearson* entre os valores de expressão desses lncRNAs e todos os mRNAs expressos nas três linhagens foram calculados e expressos em uma matriz de correlação (Figura 15).



**Figura 15: Correlação de expressão entre lncRNAs e mRNAs expressos nas 3 linhagens.** Os valores de intensidade de expressão dos mRNAs e lncRNAs expressos nas linhagens DU-145, MCF-7 e Mia PaCa 2 foram normalizados por quantil. Em seguida, os valores de expressão dos 20% lncRNAs antisense mais expressos em cada linhagem, e expressos nas três linhagens (A) e os 20% lncRNAs senso mais expressos em cada linhagem, expressos nas três linhagens e que não possuem alta correlação de expressão com o mRNA transcrito a partir de seu *locus* hospedeiro (B) foram correlacionados com os valores de expressão dos 5.614 mRNAs expressos nas três linhagens em estudo. Os coeficientes de correlação de *Pearson* foram calculados e plotados na matriz de correlação. Nas linhas estão representados cada um dos 5.614 mRNAs e nas colunas cada um dos 776 lncRNAs antisense (A) e 420 lncRNAs senso (B). Em amarelo estão representadas correlações altamente positivas ( $r > 0,8$ ) e em azul, correlações altamente negativas ( $r < -0,8$ ). Em preto estão representadas as ausências de correlação ( $-0,8 \leq r \leq 0,8$ ).

Com o objetivo de explorar melhor as coexpressões observadas, foi realizada uma análise de enriquecimento em categorias funcionais de *Gene Ontology* (GO) para os mRNAs com expressão positivamente ( $r > 0,8$ ) e negativamente ( $r < -0,8$ ) correlacionadas com a expressão de lncRNAs. Para essa análise foi utilizado o programa GENOMICA, disponível *on line* (<http://genomica.weizmann.ac.il/>) (Segal *et al.*, 2004). Para cada lncRNA foi definido um grupo

de genes coexpressos (grupo de mRNAs com expressão positivamente ou negativamente correlacionada com um determinado lncRNA) nas três linhagens em estudo. Para identificar lncRNAs associados com cada um dos termos das categorias de *GO* foi construído um mapa de módulos gênico, que estabelece associações entre cada grupo de coexpressão e os termos no *GO*. O cálculo de enriquecimento para os termos *GO* atribuídos foi realizado em relação a todos os mRNAs expressos nas três linhagens ( $p\text{-value} \leq 0,05$  - com correção para múltiplos testes de Bonferroni) (Figura 16).



**Figura 16: Enriquecimento em categorias funcionais (*Gene Ontology*) de mRNAs que possuem perfis de coexpressão com lncRNAs em *trans* nas linhagens DU-145, MCF-7 e Mía PaCa 2.** As matrizes de correlação construídas entre os 776 lncRNAs antissenso e os 5.614 mRNAs expressos nas três linhagens e os 420 lncRNAs senso e os 5.614 mRNAs expressos nas três linhagens foram usadas como *input* para o *software* GENOMICA (Segal *et al.*, 2004). O programa foi utilizado para a construção de mapas de módulos de associação entre cada grupo de mRNAs coexpressos com um determinado lncRNA e os termos *Gene Ontology*. Para a definição dos grupos de coexpressão foi utilizado o critério de alto coeficiente de correlação de Pearson ( $r > 0,8$  e  $r < -0,8$ ) entre os valores de expressão de cada lncRNA e os valores de expressão de cada um dos 5.614 mRNAs. O enriquecimento em termos de *Gene Ontology* dos genes coexpressos com lncRNAs foi calculado ( $p\text{-value} \leq 0,05$  – com correção de Bonferroni para múltiplos testes). Em (A) estão apresentados os 67 lncRNAs antissenso e em (B) os 15 lncRNAs senso cujos perfis de coexpressão com mRNAs nas linhagens DU-145, MCF-7 e Mía PaCa 2 estão associados com pelo menos um dos termos de *GO* enriquecidos. Em amarelo estão representadas associações positivas entre o lncRNA e os termos de *GO* e em azul estão representadas associações negativas entre o lncRNA e os termos de *GO*. Em preto estão representadas ausência de associações significativas.

Essa análise identificou 67 lncRNAs antissenso cujos perfis de expressão estão correlacionados com os perfis de expressão de mRNAs expressos nas três linhagens e que estão associados com os seguintes termos de *Gene Ontology*: "*DNA replication*", "*microtubule cytoskeleton*", "*cell cycle phase*", "*M phase*", "*mitosis*", "*DNA metabolic process*", "*protein transport*", "*cytoplasmic part*" e "*vacuole*". Também foram identificados 15 lncRNAs senso

cujos perfis de expressão correlacionam-se com os perfis de expressão de mRNAs nas três linhagens. Curiosamente este conjunto está associado com os mesmos termos de *Gene Ontology*: "DNA replication", "microtubule cytoskeleton", "cell cycle phase", "M phase", "mitosis", "DNA-dependent DNA replication", "protein transport", "cytoskeletal part" e "vacuole".

#### 4.5 Identificação de elementos regulatórios localizados à 5' dos lncRNAs intrônicos

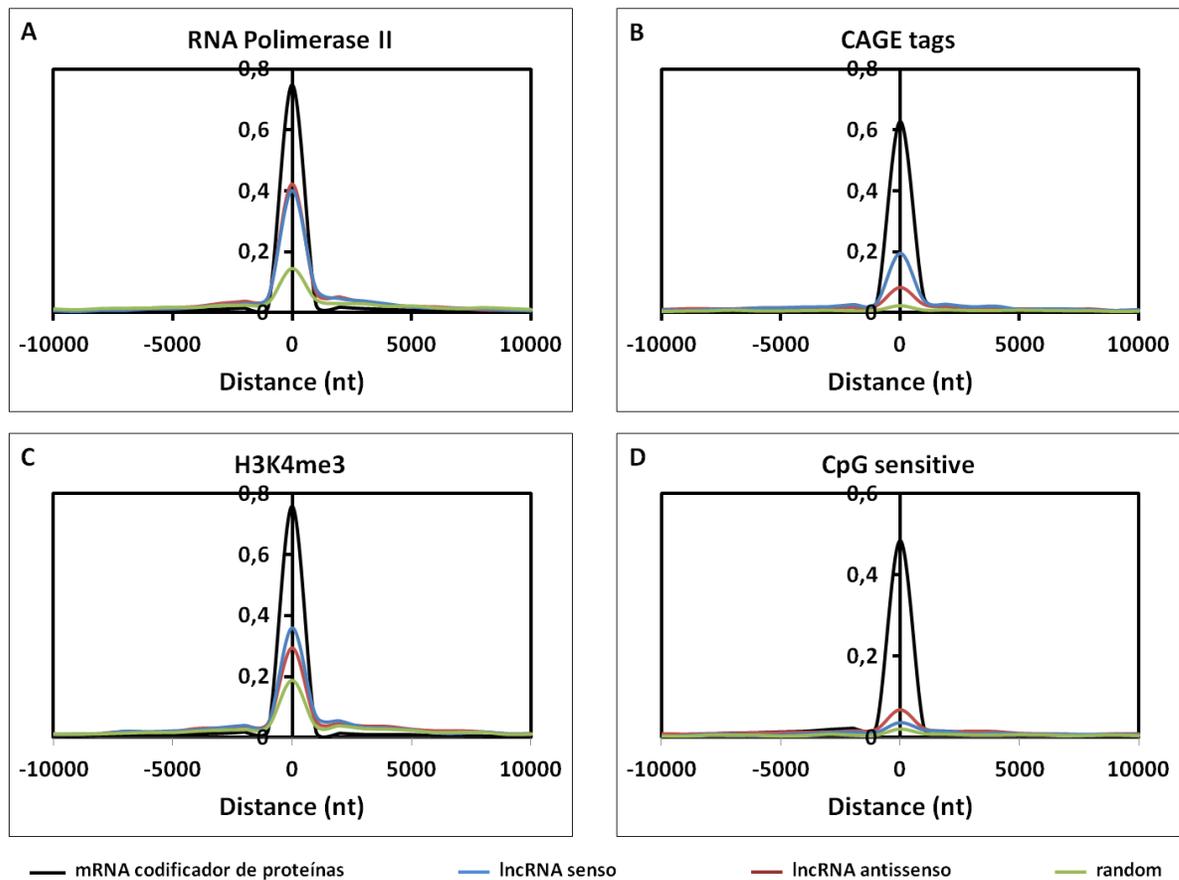
Uma vez que a maior parte dos lncRNAs intrônicos senso ou antissenso não apresentou correlação de expressão com o mRNA codificador de proteínas expresso a partir do mesmo *locus*, levantamos a hipótese de que uma parte desses lncRNAs devem ser transcritos e regulados independentemente do seu gene hospedeiro. Para investigar essa hipótese fizemos buscas por elementos na cromatina associadas à atividade transcricional disponíveis em bancos de dados públicos nas regiões genômicas situadas nas extremidades 5' preditas pela montagem de *contigs* dos lncRNAs intrônicos antissenso e senso realizada em nosso grupo (Nakaya *et al.*, 2007). É importante chamar a atenção para o fato de que o tamanho total dos transcritos intrônicos não foi ainda confirmado, de forma que não é possível definir com precisão o seu início de transcrição.

O enriquecimento da presença da enzima RNAP II na vizinhança dos 5.625 lncRNAs antissenso e 3.504 lncRNAs intrônicos senso expressos em pelo menos uma das três linhagens foi avaliado utilizando dados de imunoprecipitação da cromatina seguido de sequenciamento em larga escala (ChIP-seq) (Rozowsky *et al.*, 2009). As coordenadas genômicas dos picos de ChIP-seq de Pol II foram cruzadas com as coordenadas genômicas das extremidades 5' conhecidas dos lncRNA intrônicos detectados nas linhagens em estudo. Para a análise de enriquecimento foram computados os picos de ChIP-seq de RNAP II encontrados em até 10 kb de distância do início predito dos transcritos. Após o cruzamento das coordenadas genômicas, verificamos enriquecimento significativo (teste Kolmogorov-Smirnov (KS), *p-value* < 0,05) da enzima RNAP

II na vizinhança (até 10 kb) do provável início de transcrição, em relação a um conjunto controle formado por sequências genômicas selecionadas aleatoriamente a partir de regiões intrônicas do genoma (Figura 17A). Este resultado sugeriu que ao menos uma fração dos lncRNAs intrônicos sejam sintetizados pela RNAP II.

De forma semelhante, cruzamos as coordenadas das extremidades 5' dos lncRNAs intrônicos expressos em pelo menos uma das três linhagens em estudo, com as coordenadas de sequências de cDNAs que possuem cap 5' (*CAGE tags*) (*Cap Analysis of Gene Expression*) expressas em nove bibliotecas diferentes geradas a partir de RNAPoliA+ (Carninci e Hayashizaki, 1999; Kodzius *et al.*, 2006; Valen *et al.*, 2009). Após o cruzamento das coordenadas genômicas, verificamos um enriquecimento entre 20% a 8% de *CAGE tags* próximo à posição 0 em relação aos TSSs preditos para os lncRNAs senso e antissenso, respectivamente (Figura 17B).

Já foi descrito que trimetilações na lisina 4 da histona 3 (H3K4me3) e trimetilações na lisina 36 da histona 3 (H3K4me36) são capazes de apontar para regiões transcricionalmente ativas, sendo que a marca H3K4me3 está mais associada às regiões promotoras (Guttman *et al.*, 2009), enquanto que a marca H3K36me3 está mais associada às regiões transcritas em si (Mikkelsen *et al.*, 2007). Além disso, sabe-se CGIs também são elementos regulatórios característicos de certos tipos de promotores de mamíferos (Carninci *et al.*, 2006). Essencialmente, a mesma abordagem de cruzamento de coordenadas realizada para os dados públicos de CHIP-seq de RNAP II foram realizadas para verificar enriquecimento de marcas de H3K4me3 e presença de CGIs (Figura 17, C e D).



**Figura 17: TSSs preditos dos lncRNAs intrônicos antissenso e senso expressos em DU-145, MCF-7 e Mia PaCa 2 estão enriquecidos em marcas associadas a regiões promotoras e *CAGE tags*.** Os gráficos acima mostram a distribuição das distâncias calculadas (eixo X) para cada um dos elementos em cada uma das classes de transcritos analisadas. (A) RNAP II, (B) *CAGE tags*, (C) H3K4me3 e (D) CGI. A fração de transcritos está representada no eixo Y e cada linha representa uma das classes de transcritos descritas: mRNAs codificadores de proteínas (linha preta), lncRNAs antissenso (linha vermelha), lncRNAs senso (linha azul) e sequências randômicas (linha verde).

Como esperado a maior parte dos mRNAs codificadores de proteínas apresentam marcas de regiões promotoras como RNAP II, H3K4me3 e ilhas CpG (Figura 17). Cerca de 40% dos lncRNAs antissenso e senso apresentaram enriquecimento da presença de RNAP II próximo à posição 0 em relação aos seus TSSs preditos (Figura 17A). Cerca de 30% e 35% dos lncRNAs antissenso e senso, respectivamente, também apresentaram enriquecimento da marca H3K4me3 próximo à posição predita para os seus TSSs (Figura 17C).

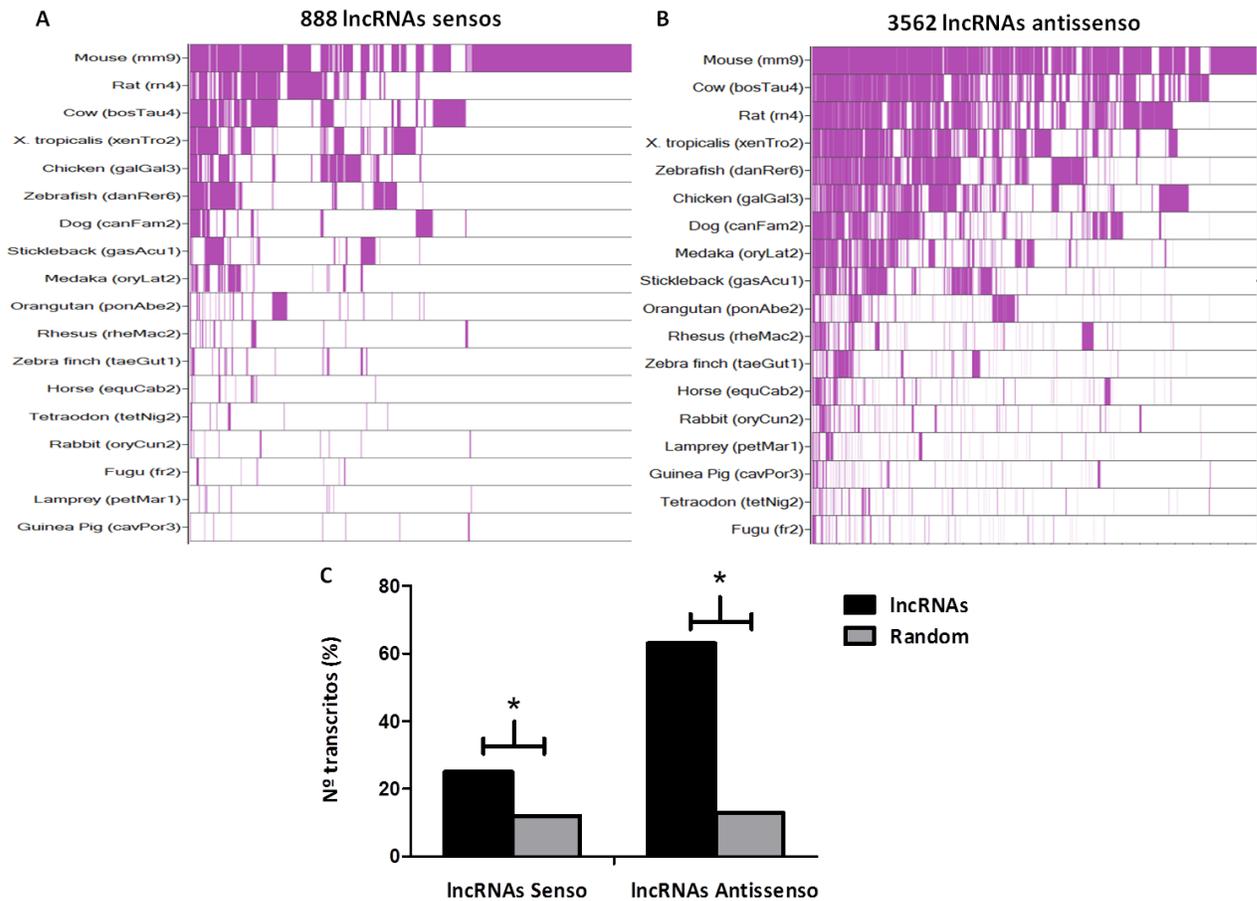
Enquanto que cerca de 50% dos mRNAs apresentaram enriquecimento de CGIs próximo aos TSSs preditos, menos de 10% dos lncRNAs intrônicos antissenso e senso apresentaram enriquecimento deste elemento regulatório, o que sugere que apenas uma pequena parte desses transcritos deva ser regulada por ilhas CpGs.

Para as três marcas de elementos regulatórios da transcrição analisados (RNAP II, H3K4me3 e CGIs), as frações de lncRNAs que apresentaram-se associados a essas marcas foram menores que as frações de mRNAs. Entretanto, essas frações foram maiores do que as encontradas para o grupo randômico testado (teste KS,  $p\text{-value} < 0,05$ ). Esses achados sugerem que ao menos uma fração desses lncRNAs apresentam regiões promotoras próprias e que sofram algum tipo de regulação com características semelhantes as observadas para mRNAs.

#### **4.6 Conservação evolutiva dos lncRNAs intrônicos**

Para identificar, dentre os lncRNAs intrônicos antissenso e senso detectados nas linhagens celulares humanas DU-145, MCF-7 e Mia PaCa 2, aqueles que possuem ortólogos expressos em outras espécies, nós cruzamos as coordenadas genômicas dos 5.625 lncRNAs antissenso e 3.504 lncRNAs senso com as coordenadas genômicas de transcritos de vertebrados mamíferos e não mamíferos que foram sintenticamente mapeados no genoma humano pelo método de alinhamento de mRNAs entre espécies denominado TransMap (Zhu *et al.*, 2007).

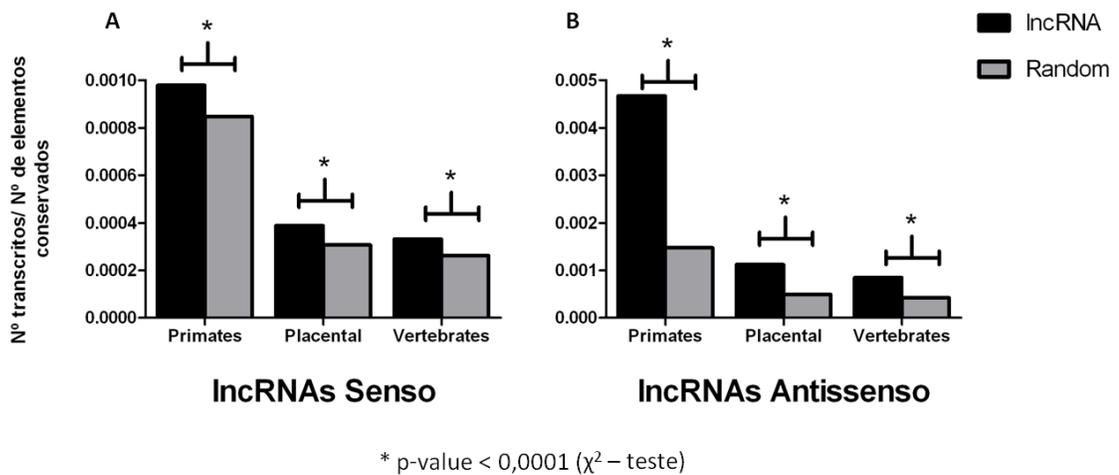
Dentre os lncRNAs senso e antissenso expressos em pelo menos uma das três linhagens analisadas, 888 (25% do total) e 3.562 (63% do total), respectivamente apresentaram conservação de expressão (Figura 18A e B). Esses conjuntos de lncRNAs que possuem ortólogos expressos em outras espécies são maiores que os conjuntos esperados quando se cruzam coordenadas genômicas de sequências humanas geradas randomicamente com as coordenadas do catálogo de ortólogos gerados por TransMap, ( $p\text{-value} < 0,0001$ ; Teste Qui-quadrado) (Figura 18C).



**Figura 18: lncRNAs intrônicos possuem transcritos ortólogos em outros vertebrados.** Distribuição de 888 lncRNAs senso (A) e 3.562 lncRNAs antissenso (B) humanos que possuem ortólogos expressos em pelo menos uma das 18 espécies de vertebrados mamíferos e não mamíferos analisadas usando o algoritmo TransMap (Zhu *et al.*, 2007). As coordenadas genômicas de todos os 3.504 lncRNAs senso e 5.625 lncRNAs antissenso identificados nas três linhagens de estudo foram cruzadas com as coordenadas genômicas do catálogo de ortólogos humanos de transcritos expressos em 18 espécies de vertebrados alinhados no genoma humano pelo algoritmo TransMap. Em cada linha está representada a distribuição dos ortólogos mapeados em cada uma das 18 espécies de vertebrados analisadas. Nas colunas estão representados cada um dos lncRNAs senso (A) e antissenso (B) que possuem ortólogos identificados. Em (C) estão plotadas as porcentagens de ortólogos identificados nas 18 espécies de vertebrados analisadas para cada classe de transcrito não codificador em relação às porcentagens de ortólogos identificados para sequências randômicas (Teste Qui-Quadrado,  $p$ -value < 0,0001).

Para cerca de 25% dos lncRNAs senso e 63% dos lncRNAs antissenso foram identificados ortólogos em outras espécies de vertebrados, desde peixes até primatas. Apesar de estarem presentes nas análises espécies mais próximas ao homem, como a do orangotango e a do macaco *rhesus*, a espécie com mais ortólogos identificados em ambas as análises foi a do camundongo. Isto pode ser explicado pelo fato de que o transcriptoma do camundongo possui alta cobertura, devido ao grande esforço de anotação funcional do genoma dessa espécie realizado pelo projeto *Functional Annotation of the Mouse Genome* - FANTOM (Okazaki *et al.*, 2002). A baixa representatividade de ortólogos observada nas espécies de primatas pode ser explicada pela menor cobertura disponível do transcriptoma dessas espécies.

Outra evidência de conservação evolutiva para os lncRNAs intrônicos senso e antissenso expressos nas linhagens utilizadas neste trabalho foi obtida a partir do cruzamento das coordenadas genômicas de todos os lncRNAs expressos em pelo menos uma das três linhagens com as coordenadas genômicas de elementos conservados entre 46 espécies de vertebrados. Além disso, foi avaliada a presença de elementos conservados subagrupando as espécies em primatas (9 espécies) e mamíferos placentários (32 espécies) (Figura 19). Para essa análise foi utilizado o *dataset* de elementos conservados disponível no UCSC *Genome Browser*.

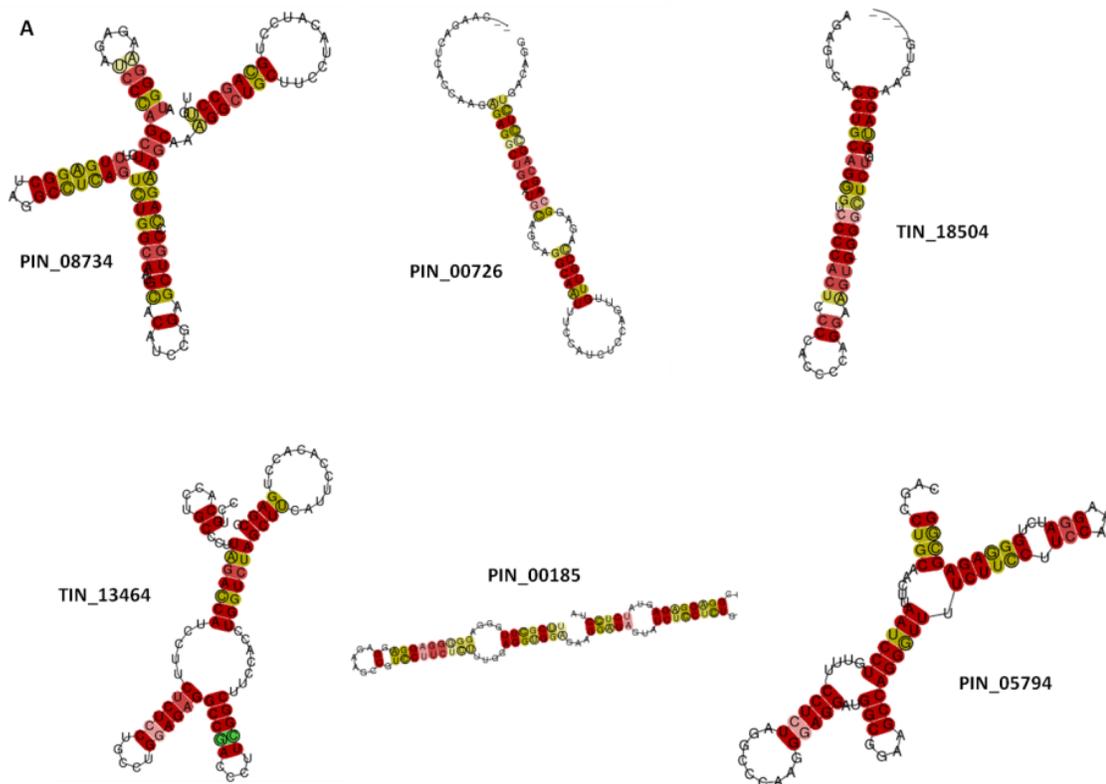


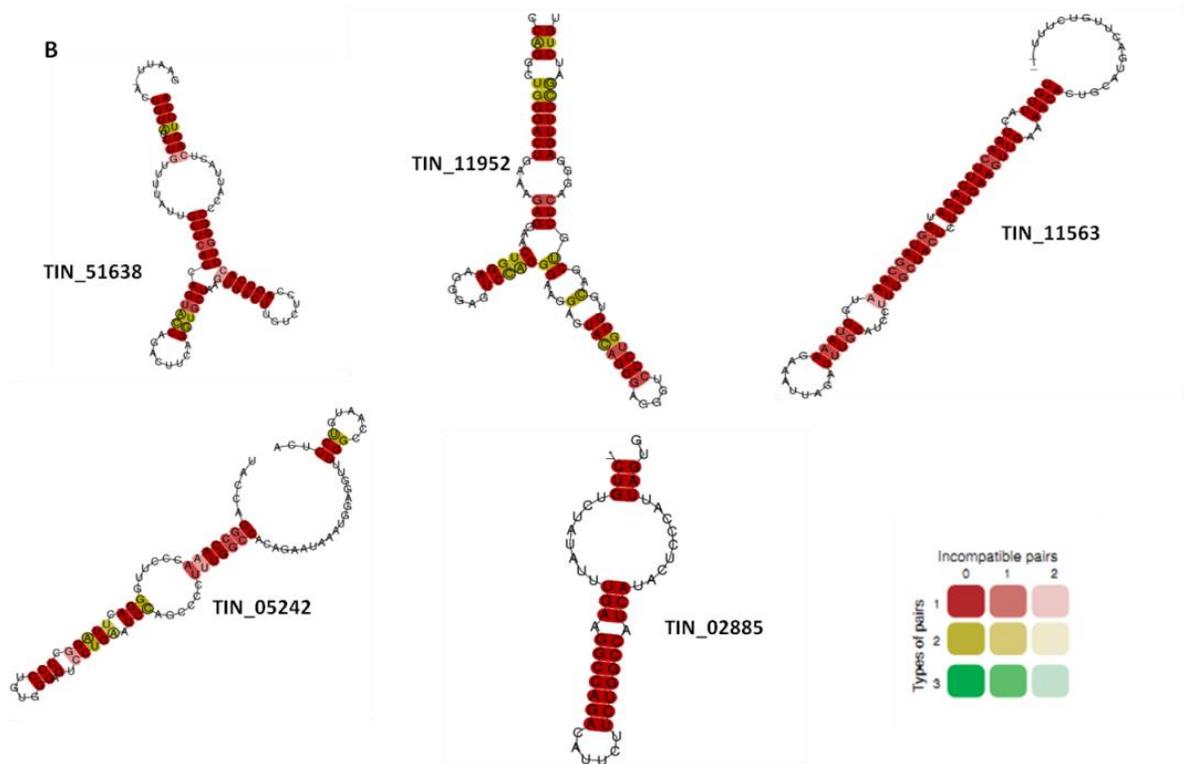
**Figura 19: Frequência de elementos evolutivamente conservados entre primatas, mamíferos placentários e vertebrados encontrados nos lncRNAs intrônicos.** As coordenadas genômicas dos 3.504 lncRNAs senso (A) e dos 5.625 lncRNAs antissenso (B) foram cruzadas com as coordenadas genômicas de elementos conservados evolutivamente entre 46 espécies de vertebrados e entre os subgrupos de primatas (9 espécies) e mamíferos placentários (32 espécies). As frequências de transcritos não codificadores conservados evolutivamente foram comparadas com as frequências de sequências geradas aleatoriamente conservadas evolutivamente (Teste Qui-Quadrado, *p-value* < 0,0001).

Foram identificados 4.377 (78% do total) e 1.711 (49% do total) lncRNAs intrônicos antissenso e senso, respectivamente, que comportam em sua sequência elementos conservados evolutivamente entre os vertebrados. Comparando-se as taxas de transcritos intrônicos que apresentaram sobreposição com coordenadas genômicas de elementos conservados entre organismos pertencentes a cada grupo evolutivo com as taxas de coordenadas genômicas geradas aleatoriamente podemos observar que tanto os lncRNAs senso como os lncRNAs antissenso possuem enriquecimento nesses elementos (*p-value* < 0,0001, Teste Qui-Quadrado). Além disso, pode-se observar, como esperado, uma maior frequência desses elementos entre os primatas, que entre placentários e vertebrados.

Uma grande parte dos ncRNAs bem caracterizados possuem estruturas características que são funcionais e bem conservadas ao longo do tempo evolutivo, como por exemplo, os rRNAs, os

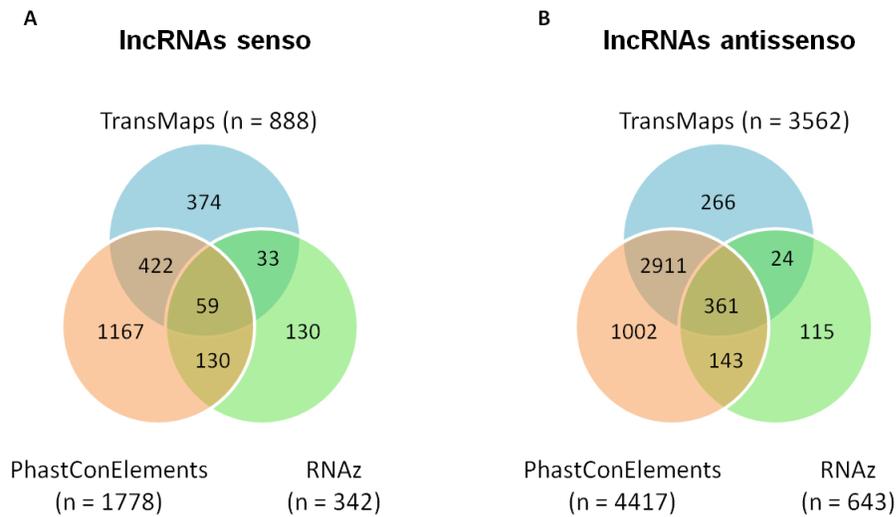
tRNAs, bem como os RNAs catalíticos, como por exemplo a RNase P e outros. A seleção natural agindo sobre a estrutura secundária desses RNAs não codificadores funcionais causa padrões de substituição de nucleotídeos característicos na sequência primária. Mutações consistentes e compensatórias troca um tipo de par de bases por outro nas regiões pareadas da molécula. Desse modo, há diferenças na variação de sequências entre regiões de *loop* e regiões de *stem*. Esses padrões podem ser explorados para discriminar ncRNAs funcionais expressos nas células (Washietl *et al.*, 2005). Utilizando o programa RNAz (Washietl *et al.*, 2005) identificamos estrutura secundária evolutivamente conservada para 623 (11% do total de 5.625) lncRNAs antissenso e 342 (10% do total de 3.504 lncRNAs senso). A Figura 20 mostra alguns exemplos de estruturas secundárias consenso obtidas para lncRNAs antissenso (Figura 20A) e para lncRNAs senso (Figura 20B).





**Figura 20: lncRNAs intrônicos antissenso e senso expressos nas linhagens DU-145, MCF-7 e Mia PaCa 2 possuem estrutura secundária conservada evolutivamente.** Na Figura estão representadas alguns exemplos de estruturas secundárias consenso previstas para lncRNAs antissenso (**A**) e senso (**B**). Os círculos indicam variações de pareamento de nucleotídeo nos *stems* e as cores indicam o número de diferentes tipos de pares de bases que suportam a seleção estabilizante sobre a estrutura e a escala de tom da cor indica número de pareamentos incompatíveis com a estrutura consenso no alinhamento múltiplo.

A Figura 21 mostra a sobreposição do número de transcritos não codificadores intrônicos senso e antissenso aos quais foram atribuídas predições de conservação evolutiva de expressão, de conservação evolutiva de sequência e predição de estrutura secundária conservada evolutivamente.



**Figura 21: Sobreposição do número de lncRNAs antissenso e senso que apresentaram conservação de expressão (TransMaps), conservação de sequência (PhastConElements) e predição de estrutura secundária evolutivamente conservada. (A) lncRNAs senso; (B) lncRNAs antissenso.**

Cinquenta e nove lncRNAs intrônicos senso (2% do total expresso) e 361 lncRNAs intrônicos antissenso (6% do total expresso) possuem predição de estrutura secundária evolutivamente conservada, além de conservação de expressão e de sequência (Figura 21). A pequena sobreposição pode ser explicada pelos diferentes critérios utilizados para o cálculo de estrutura e conservação por cada um dos algoritmos.

#### **4.7 Regulação da expressão de lncRNAs intrônicos e mRNAs por metilação do DNA**

Já está bem estabelecido na literatura (Esteller, 2008) que muitos genes importantes para o desenvolvimento do câncer têm sua expressão regulada pela hipermetilação do DNA. Com o objetivo de se avaliar a possibilidade desse mecanismo regular também a transcrição de lncRNAs intrônicos, células das linhagens tumorais DU-145, MCF-7 e Mia PaCa 2 foram tratadas com o agente desmetilante 5-aza-2'-deoxicitidina (5-AZA) (Figura 2, subseção 3.4 da seção de Materiais

e Métodos). As condições de tratamento para obter a reexpressão de genes silenciados por metilação foram padronizadas como descrito na subseção 3.2 da seção de Materiais e Métodos. Amostras de RNA provenientes de duas réplicas biológicas independentes de cada linhagem celular tratada ou não com 5-AZA foram amplificadas e marcadas com os fluoróforos Cy5 e Cy3. Os alvos fluorescentes foram em seguida hibridizados em duplicata com as lâminas de oligonucleotídeos customizadas já descritas, trocando-se os fluoróforos em cada réplica técnica (Figura 4 da subseção 3.5.3 da seção de Material e Métodos). Na Tabela 15 estão discriminados os números de sondas expressas nas linhagens tratadas com 5-AZA por classe de transcrito.

**Tabela 15: Número de sondas expressas em cada linhagem celular após tratamento com 5-AZA.**

Tipos de Sondas	Lâmina 44K	DU-145		MCF-7		Mia PaCa 2	
		Expressos	Não Expressos	Expressos	Não Expressos	Expressos	Não Expressos
mRNA	13274	11365	1909	11207	2067	10792	2482
ncRNA Antisenso	10901	5112	5789	4506	6395	3542	7359
ncRNA Senso	6831	3045	3786	2917	3914	2238	4593
Outros	57	41	16	42	15	35	22
<b>Total</b>	<b>31063</b>	<b>19563</b>	<b>11500</b>	<b>18672</b>	<b>12391</b>	<b>16607</b>	<b>14456</b>

A Tabela 16 mostra a proporção de transcritos com expressão estatisticamente significativamente diferente após o tratamento das células com 5-AZA em relação às células controle (SAM –  $q$ -value < 10%). A desmetilação do DNA pelo tratamento com o agente desmetilante 5-AZA alterou a expressão de apenas uma pequena fração de genes codificadores de proteínas e de lncRNAs.

**Tabela 16: Número de genes codificadores de proteínas e lncRNAs intrônicos antissenso e senso diferencialmente expressos após o tratamento de células das linhagens DU-145, MCF-7 e Mia PaCa 2 com 5-AZA.**

Classes de RNAs	DU-145 (SAM; 10% q-value)			MCF-7 (SAM; 10% q-value)			Mia PaCa 2 (SAM; 10% q-value)		
	EXPRESSOS	UP	DOWN	EXPRESSOS	UP	DOWN	EXPRESSOS	UP	DOWN
genes codificadores de proteínas	6432	392 (6%)	0	6328	639 (10%)	12 (0,2%)	6191	1600 (26%)	12 (0,2%)
lncRNA antissenso	5112	81 (2%)	0	4506	125 (3%)	2	3542	380 (11%)	6 (0,2%)
lncRNA senso	3045	70 (2%)	0	2917	163 (6%)	0	2238	212 (10%)	5 (0,2%)
Outros	41	6 (15%)	0	42	8 (19%)	0	35	8 (23%)	0
<b>Total</b>	<b>14630</b>	<b>549 (4%)</b>	<b>0</b>	<b>13793</b>	<b>935 (7%)</b>	<b>14 (0,1%)</b>	<b>12006</b>	<b>2200 (18%)</b>	<b>23 (0,2%)</b>

Nas células de DU-145, apenas 6% dos mRNAs codificadores de proteínas expressos foram considerados como diferencialmente expressos após tratamento das células com 5-AZA em relação às células controle. Nas células da linhagem MCF-7, a porcentagem de mRNAs diferencialmente expressos após tratamento com 5-AZA foi um pouco maior, 10% dos expressos. Em Mia PaCa 2 foi observado o maior efeito do agente desmetilante, 26% dos mRNAs codificadores de proteínas expressos foram significativamente afetados pela adição da droga no meio de cultura. Dentre os afetados pelo tratamento com 5-AZA, de 68% a 78% deles apresentaram CGIs em até 2 kb do seu 5'UTR.

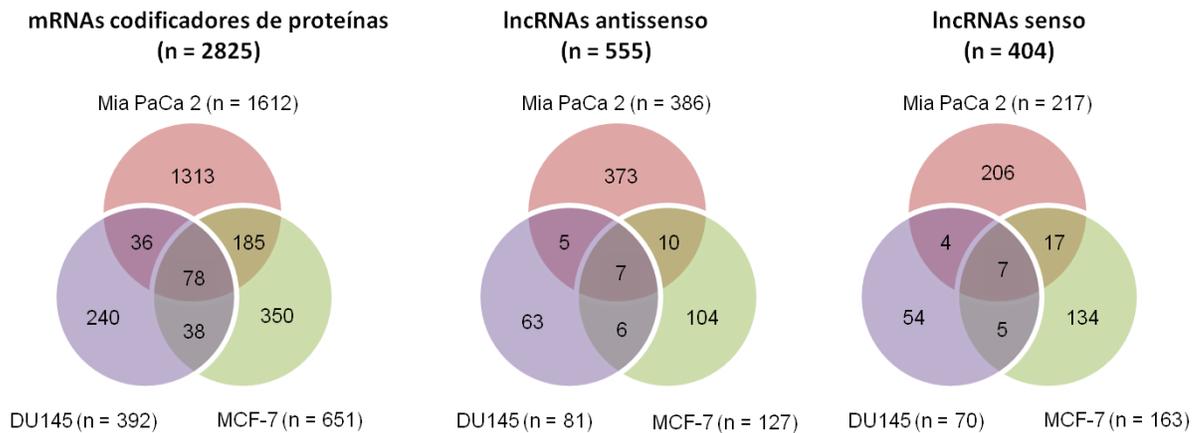
Em relação aos lncRNAs, o efeito do tratamento com o agente desmetilante 5-AZA foi ainda menor. Em DU-145, a proporção transcritos com expressão alterada após o tratamento foi de aproximadamente 2% do total de expressos, tanto para lncRNAs antissenso, como para os lncRNAs intrônicos senso. Em MCF-7, o tratamento com 5-AZA alterou a expressão de aproximadamente 3% dos lncRNAs antissenso e 6% dos lncRNAs senso. E em Mia PaCa 2, foi observada alteração de expressão em aproximadamente 11% dos lncRNAs antissenso e 10% dos

lncRNAs senso. Dentre estes lncRNAs intrônicos antissenso e senso afetados, cerca de 17% a 22% e 15% a 20% deles apresentaram CGIs em até 5kb de sua extremidade 5' conhecida, respectivamente.

As listas dos 50 mRNAs, 50 lncRNAs antissenso e 50 lncRNA senso com maior diferença de expressão em cada linhagem estão apresentadas nas tabelas em anexo (Apêndices de 1 a 9).

Interessantemente, em DU-145 não encontramos nenhum gene ou lncRNA com expressão diminuída com significância estatística ( $q\text{-value} < 10\%$ ). Em MCF-7 e Mia PaCa 2, o número de transcritos com expressão diminuída após o tratamento foi muito pequeno. Apenas 12 genes codificadores de proteínas foram encontrados diminuídos em cada uma das linhagens, sendo nenhum deles comuns a ambas. Para lncRNAs, encontramos apenas dois transcritos diminuídos em MCF-7 e 11 em Mia PaCa 2, não sendo nenhum deles comuns a ambas as linhagens também. Muito provavelmente, essa diminuição na expressão após tratamento ocorreu devido a um efeito indireto do tratamento com 5-AZA.

Tanto os mRNAs codificadores de proteínas diferencialmente expressos quanto os lncRNAs antissenso e senso diferencialmente expressos mostram-se afetados de forma linhagem específica. A Figura 22 apresenta os diagramas de venn que mostram a baixa sobreposição de transcritos identificados como diferencialmente expressos entre as linhagens DU-145, MCF-7 e Mia PaCa 2.



**Figura 22: mRNAs codificadores de proteínas e lncRNAs intrônicos diferencialmente expressos após tratamento com o agente desmetilante 5-AZA nas 3 linhagens.** Os diagramas de venn mostram os números de transcritos de cada classe que foram identificados como significativamente diferencialmente expressos (SAM  $q$ -value < 10%) após tratamento das linhagens DU-145, MCF-7 e Mia PaCa 2 com 5-AZA.

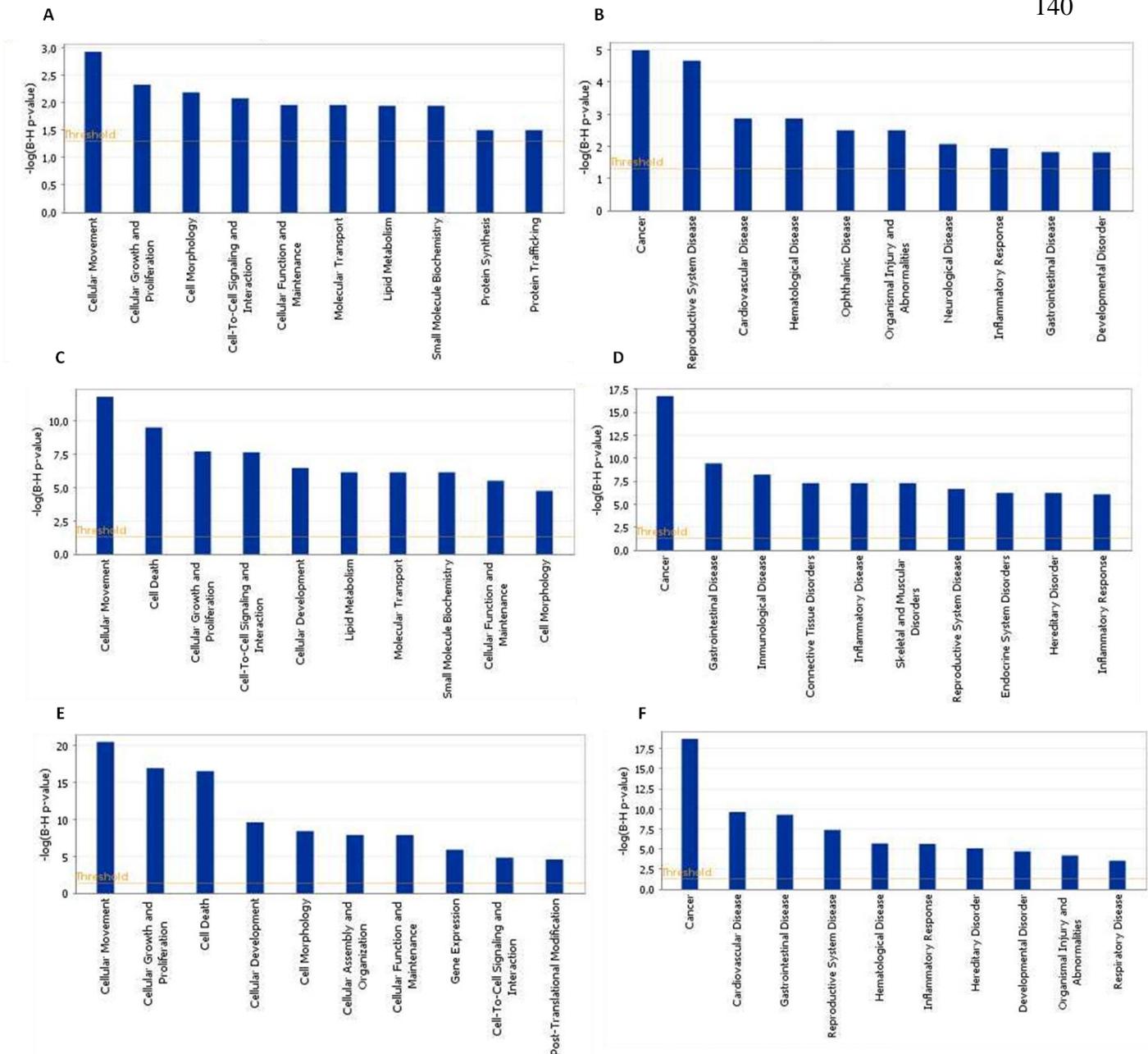
Entre os 2.825 mRNAs codificadores de proteínas diferencialmente expressos em pelo menos uma das três linhagens, somente 78 (3%) foram considerados como diferencialmente expresso nas três análises. Para os lncRNAs intrônicos a sobreposição foi de apenas 1% (7 de 555) e 2% (7 de 404) dos transcritos antissenso e senso, respectivamente.

Sobrepondo nossas listas de mRNAs diferencialmente expressos após o tratamento com o agente desmetilante do DNA (5-AZA) com as listas de genes já identificados em outros trabalhos como regulados por metilação do DNA e disponíveis publicamente no banco de dados *PubMeth* (Ongenaert *et al.*, 2008), encontramos uma pequena sobreposição. Do total de 362 genes diferencialmente expressos na linhagem de carcinoma de próstata DU-145, 6 já foram descritos como regulados por metilação do DNA em tumores de próstata. Oito dos 651 genes diferencialmente expressos na linhagem de adenocarcinoma de mama MCF-7, e 10 dos 1612 genes diferencialmente expressos na linhagem de carcinoma de pâncreas Mia PaCa 2, já foram descritos como regulados por metilação do DNA em tumores. A pequena sobreposição pode ser

explicada pela característica linhagem/tecido específica da metilação do DNA, e pela quantidade limitada de dados disponível no banco *PubMeth*.

Com o objetivo de investigar a presença de categorias biológicas específicas associadas com os genes regulados por metilação do DNA presentes em nossas análises usamos o *software Ingenuity Pathway Analysis* (<http://www.ingenuity.com>). Esta ferramenta foi utilizada para identificar redes de interações gênicas associadas a categorias biológicas funcionais contendo genes diferencialmente expressos após o tratamento com o agente desmetilante 5-AZA.

As 10 categorias biológicas funcionais e as 10 categorias de associação com doença/desordem mais relacionadas com os 392, 651 e 1.612 genes codificadores de proteínas cuja expressão foi regulada pelo tratamento com 5-AZA em DU-145, MCF-7 e Mia PaCa 2, respectivamente estão mostradas na Figura 23. A categoria biológica funcional e a doença mais associada com os genes diferencialmente expressos após tratamento com 5-AZA nas três linhagens foram “*Cellular Movement*” e “*Cancer*”, respectivamente.



**Figura 23: Top 10 categorias biológicas funcionais e doenças/desordens associadas aos genes codificadores de proteínas significativamente diferencialmente expressos após tratamento com o agente desmetilante 5-AZA. (A)** Top 10 categorias biológicas funcionais e **(B)** top 10 doenças/desordens associadas com os 392 genes afetados pela desmetilação do DNA em DU-145. **(C)** Top 10 categorias biológicas funcionais e **(D)** top 10 doenças/desordens associadas com os 651 genes afetados pela desmetilação do DNA em MCF-7. **(E)** Top 10 categorias biológicas funcionais e **(F)** top 10 doenças/desordens associadas com os 1.612 genes afetados pela desmetilação do DNA em Mia PaCa 2. A ferramenta *Ingenuity Pathways Analysis* calcula um  $p$ -value baseado na probabilidade de se obter o número de moléculas associadas às categorias de um *dataset* ao acaso. Nesta análise, o  $p$ -value foi corrigido utilizando a correção para múltiplos testes B-H. A linha *threshold* corresponde à  $-\log(\text{B-H } p\text{-value}) = 0,05$ .

#### **4.8 Seleção de lncRNAs intrônicos regulados por metilação do DNA para validação**

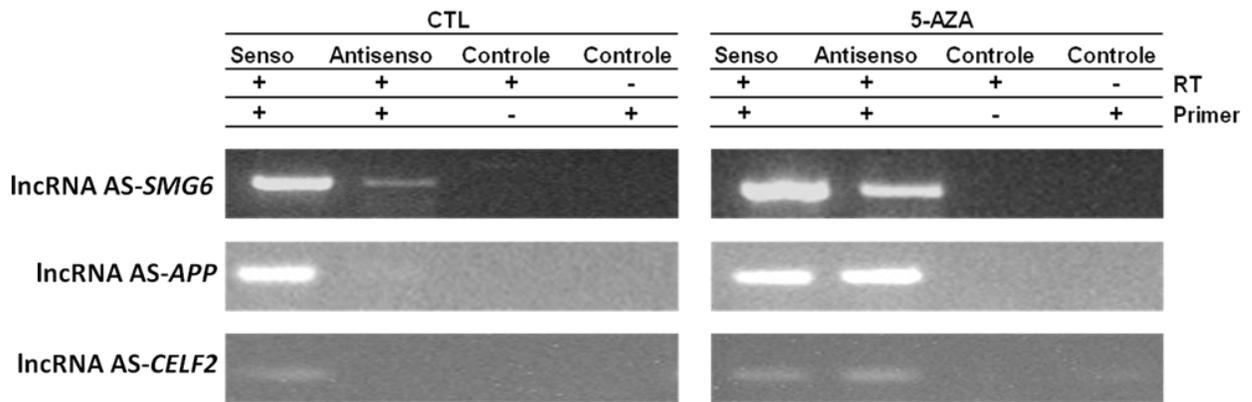
Nossos experimentos de expressão gênica em larga escala em células de três linhagens tumorais humanas após tratamento com o agente desmetilante 5-AZA identificaram 555 lncRNAs intrônicos antissenso e 404 lncRNAs intrônicos senso significativamente diferencialmente expressos em pelo menos uma das linhagens em estudo. Dentre esses, escolhemos três transcritos para validação do aumento de expressão após tratamento com 5-AZA observada nas análises em larga escala por PCR em tempo real e para validação de uma possível região promotora em sua extremidade 5' regulada por metilação do DNA (Tabela 17: ). Os critérios de escolha foram: (i) ser um transcrito de orientação antissenso à transcrição do mRNA do gene hospedeiro; (ii) ter sido medido, nos experimentos em larga escala, um aumento de pelo menos 50% em sua expressão (1,5 x) após o tratamento com 5-AZA em pelo menos duas das três linhagens estudadas e (iii) possuir uma predição de ilha CpG na região genômica em até 5 kb de distancia da extremidade 5' do transcrito. É importante chamar a atenção para o fato de que o tamanho total do transcrito intrônico e, portanto o seu início de transcrição preciso, é ainda desconhecido. As coordenadas dos contigs apresentadas na Tabela 17: são resultados da montagem de ESTs disponíveis no GenBank (Nakaya *et al.*, 2007). Como esses transcritos são em geral expressos em baixo número de cópias, é plausível que os transcritos completos se estendam em uma ou em ambas as direções.

**Tabela 17: LncRNAs intrônicos antissenso com expressão regulada pelo tratamento com 5-AZA selecionados para validação.**

<i>Antisense ncRNA Contig Name</i>	<i>Antisense ncRNA Contig Coordinate</i>	<i>Strand</i>	<i>Host Gene ID</i>	<i>Host Gene Symbol</i>	<i>Fold Change (AZA/CTL)</i>		
					<i>DU-145</i>	<i>MCF-7</i>	<i>Mia PaCa 2</i>
TIN_20703 lncRNA AS-SMG6	chr17:1965309-1966536	+	23293	SMG6	5,5	3,6	1,8
TIN_31722 lncRNA AS-APP	chr21:27259582-27260205	+	351	APP	4,1	8,2	5,8
TIN_05238 lncRNA AS-CELF2	chr10:11364897-11365170	-	10659	CELF2	2,6	-	1,6

#### 4.9 Expressão fita específica de lncRNAs intrônicos

Para confirmar a indução da expressão dos transcritos antissensos selecionados após tratamento com 5-AZA, primeiramente realizamos um experimento de transcrição reversa fita específica e verificamos a presença de expressão de cada uma das fitas antes (CTL) e após tratamento de células da linhagem DU-145 com 5  $\mu$ M de 5-aza-2'-deoxicitidina por 3 dias (5-AZA) (Figura 24).



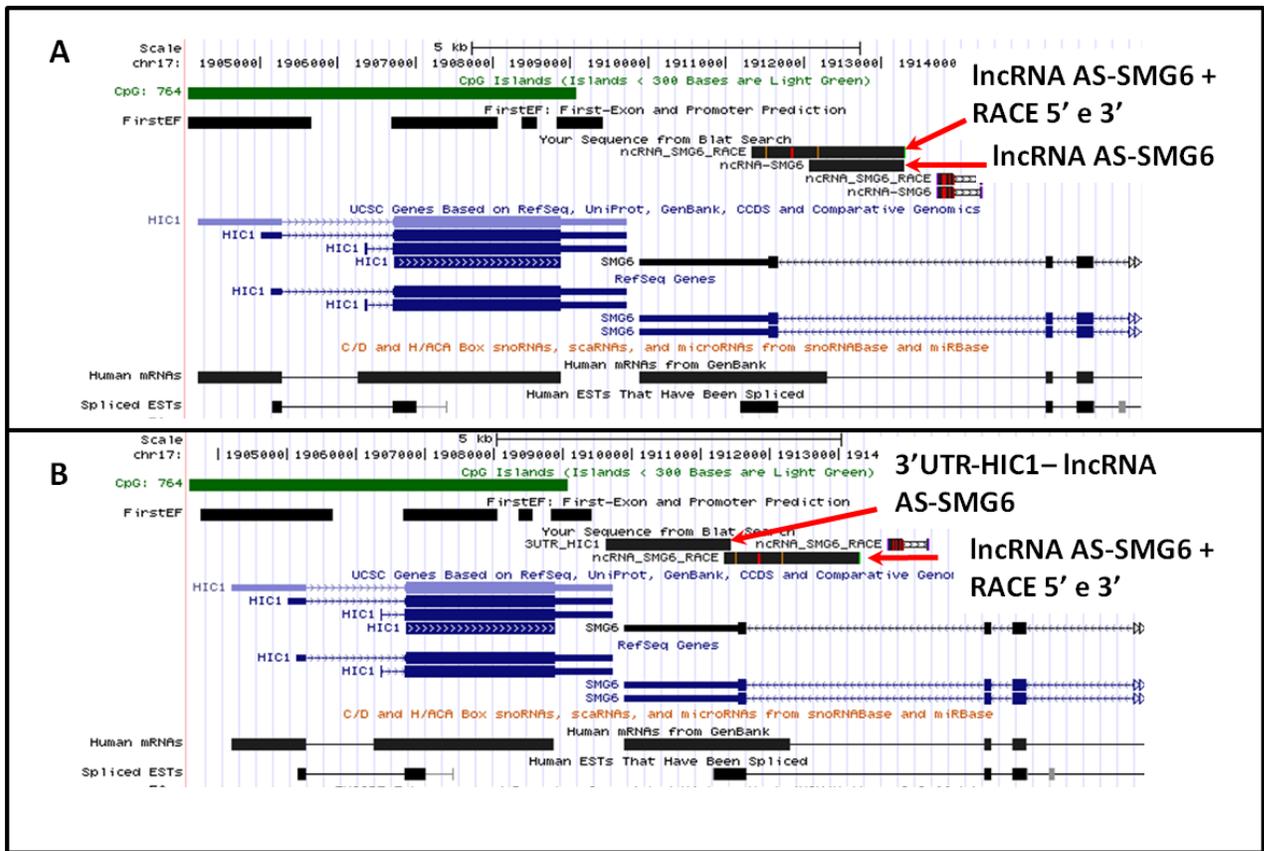
**Figura 24: Detecção de expressão fita específica dos lncRNAs antissensos escolhidos para validação em DU-145.** 4  $\mu$ g de RNA total extraído de células da linhagem DU-145 antes (CTL) e após tratamento com 5  $\mu$ M de 5-AZA (5-AZA) por 3 dias foram utilizados como molde para a síntese de cDNA utilizando *primers* específicos para as fitas senso e antisense dos lncRNAs AS-SMG6, AS-APP, AS-CELF2 utilizando o kit para transcrição reversa *TermoScript*<sup>®</sup> (Invitrogen). Como controles, foram realizadas uma transcrição reversa sem *primer*, para mostrar que o cDNA gerado não é um artefato causado por *self-annealing* de algum RNA da amostra, e uma transcrição reversa sem a enzima transcriptase reversa para mostrar ausência de contaminação das amostras com DNA genômico. A seguir, foram realizados 40 ciclos de amplificação por PCR de uma região interna aos *primers* utilizados na transcrição reversa de cada transcrito.

Foi detectada antes e após o tratamento com 5-AZA a expressão da fita senso das regiões genômicas que corresponde aos contigs dos lncRNAs antissenso AS-SMG6, AS-APP e AS-CELF2. Para o transcrito AS-SMG6 podemos observar que antes do tratamento com 5-AZA já podia ser detectada expressão da fita antisense, que se tornou, aparentemente maior após a desmetilação causada pelo tratamento com 5-AZA como sugerido pelos dados de análise em larga escala. Já para os transcritos AS-APP e AS-CELF2, a fita antisense mostrou-se expressa apenas após o tratamento com 5-AZA.

#### 4.10 Determinação da estrutura primária de lncRNAs intrônicos por RACE 5' e 3'

Para avançarmos na determinação de suas estruturas primárias, foram realizados experimentos de RACE 5' e 3' para dois dos três lncRNAs antissenso regulados por 5-AZA selecionados: AS-SMG6 e AS-APP.

A Figura 25A ilustra o resultado das reações de RACE 5' e 3' para o lncRNA AS-SMG6. A predição informática para o tamanho deste transcrito não codificador era inicialmente de 1.228 nucleotídeos. Após a extensão por RACE obtivemos um transcrito de 2.066 nucleotídeos com uma cauda poli-A contendo 112 nucleotídeos. Além disso, utilizando a ferramenta bioinformática *Polyadq* (Tabaska e Zhang, 1999) foi encontrado uma predição para um sinal de poliadenilação a 16 nucleotídeos *upstream* da cauda poli-A. Como se pode observar na Figura 25A, a extensão de cerca de 800 nucleotídeos no tamanho inicial da predição do transcrito ocorreu principalmente na extremidade 5' da predição do *contig*.

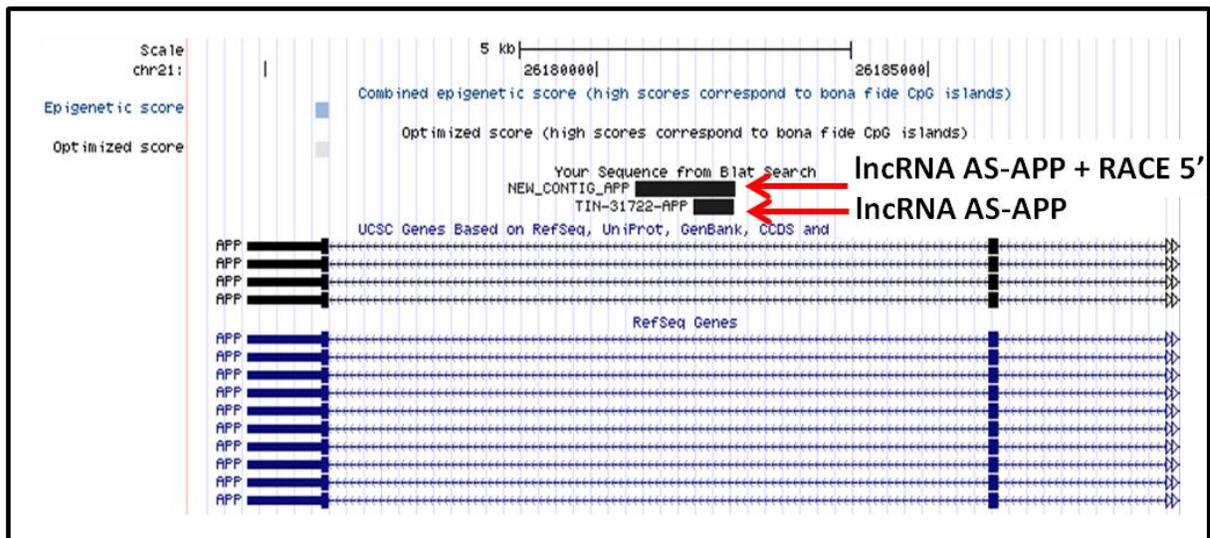


**Figura 25: Mapeamento genômico do *contig* do transcrito AS-SMG6 e dos resultados da sua extensão por RACE 5' e 3'.** Em (A) estão representados os mapeamentos da sequência predita originalmente pela montagem de ESTs realizada *in silico* para o transcrito **lncRNA AS-SMG6** e da sequência obtida após a extensão por RACE 5' e 3' (**AS-SMG6 + RACE 5' e 3'**). Em (B) estão representados os mapeamentos da sequência estendida do lncRNA AS-SMG6 e do resultado do sequenciamento do fragmento amplificado por RT-PCR com *primers* que anelam nas extremidades 3' UTR do gene *HIC1* e 5' do transcrito não codificador AS-SMG6 (**3' UTR-HIC1-AS-SMG6**), mostrando que ambos fazem parte de uma mesma unidade transcricional.

A extensão da extremidade 5' do transcrito não codificador juntamente com a localização próxima (~1.500 nucleotídeos) do 3' UTR do gene *HIC1* nos levou a pensar que ambas as mensagens pudessem fazer parte de uma única unidade transcricional. Para testar essa hipótese, foi realizada uma reação de RT-PCR com *primers* que anelam nas extremidades de ambos os transcritos. A reação gerou um fragmento do tamanho próximo ao esperado (~1800 pb), indicando que os transcritos fazem parte de uma mesma mensagem. O resultado do sequenciamento do

fragmento obtido está mostrado na Figura 25B. Desse modo ficou demonstrado que o transcrito não codificador AS-SMG6 cuja expressão é induzida por desmetilação do DNA faz parte da unidade transcricional do gene *HIC1*. Esse gene é um supressor de tumor e já foi demonstrado ser hipermetilado em câncer (Wales *et al.*, 1995).

O lncRNA transcrito a partir da fita oposta à fita que transcreve o mRNA do gene *APP* (lncRNA AS-APP) foi estendido em 867 nucleotídeos na sua extremidade 5' por RACE (Figura 26). Não foram obtidas sequências específicas para o lncRNA AS-APP nos experimentos de RACE 3'. Uma possibilidade para explicar esse resultado é que o 3' conhecido desse transcrito já seja o final na mensagem completa. A predição informática para o transcrito AS-APP inicial era de 624 nucleotídeos, após a extensão por RACE, o transcrito não codificador teve seu tamanho aumentado para 1.491 nucleotídeos. A localização genômica do transcrito encontra-se totalmente incluída na fita oposta ao último íntron do gene *APP* (Figura 26).

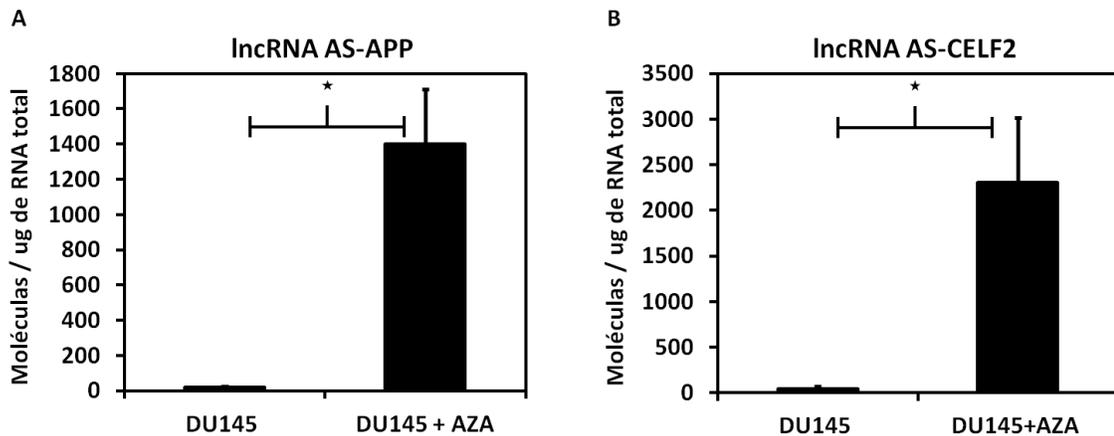


**Figura 26: Mapeamento genômico do contig do lncRNA AS-APP e do resultado de sua extensão por RACE 5'.** Na Figura estão representados os mapeamentos da sequência predita originalmente pela montagem de ESTs realizada *in silico* para o lncRNA AS-APP e da sequência obtida após a extensão por RACE 5' (lncRNA AS-APP + RACE 5').

#### 4.11 Quantificação absoluta da expressão de lncRNAs intrônicos regulados por metilação

Os resultados de transcrição reversa fita específica mostraram a existência de transcrição a partir de ambas as fitas nas regiões correspondentes aos *loci* dos lncRNAs antissensos selecionados (Figura 24). A expressão da fita senso medida nos experimentos de transcrição reversa fita específica pode corresponder ao pré-mRNA do gene codificador de proteína dos *loci* gênicos analisados, ou *intron lariats* resultantes do *splicing* que se acumulam na célula. Podem ser ainda transcritos com funções e biogênese independentes, gerados em regiões intrônicas com a mesma orientação do mRNA. Para analisar a expressão dos lncRNAs antissensos foi necessário utilizar um método de quantificação fita específica capaz de distinguir o transcrito intrônico antissenso de transcritos que se sobrepõe com orientação inversa (pré-mRNA, *intron lariats*). Para isso, utilizou-se cDNAs obtidos a partir de reações de transcrição reversa iniciadas por *primers* específicos complementares a fita antissenso. A abundância do transcrito antissenso foi a seguir determinada utilizando a metodologia de quantificação absoluta por PCR em tempo real.

A Figura 27 mostra o número de moléculas de cada lncRNA antissenso expressas na linhagem DU-145 antes e após tratamento com o agente desmetilante do DNA (5-AZA), por  $\mu\text{g}$  de RNA total usados nas reações de transcrição reversa fita específicas. Uma vez que mostramos que o lncRNA antissenso AS-SMG6 faz parte da região 3' UTR do gene codificador de proteínas *HIC1*, cuja transcrição já está caracterizada como sendo regulada por metilação do DNA, este transcrito deixou de ser um lncRNA candidato para a validação da indução da expressão após tratamento com o agente desmetilante 5-AZA.



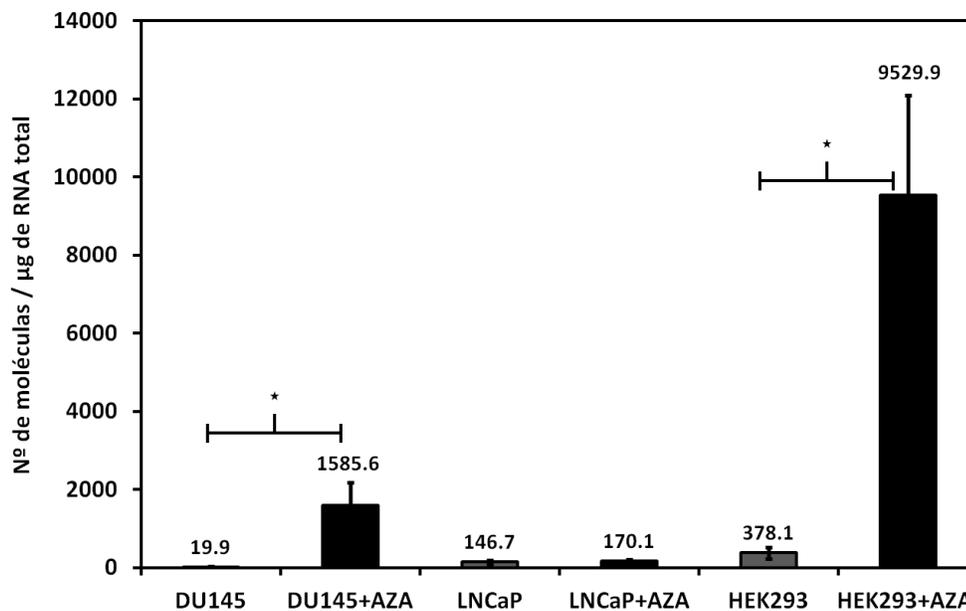
**Figura 27: Quantificação absoluta da expressão endógena e após o tratamento com 5-AZA dos lncRNAs antissenso AS-APP e AS-CELF2 em DU-145.** 4 ug de RNA total foram submetidos à transcrição reversa fita específica utilizando o kit *TermoScript®* (Invitrogen). A seguir os cDNAs obtidos foram quantificados por PCR em tempo real absoluto. Em **(A)** e **(B)** temos os números de moléculas por micrograma de RNA total usado na reação de transcrição reversa dos lncRNAs AS-APP e AS-CELF2, respectivamente, expressos em DU-145 antes e após tratamento com 5µM de 5-AZA por três dias. As barras de erro representam o desvio padrão em relação à média das 3 réplicas biológicas representada nos gráficos. O teste estatístico Teste-t para diferença entre as médias foi realizado ( $p$ -value < 0,05).

A determinação do número de moléculas dos lncRNAs antissenso AS-APP e AS-CELF2 antes e após o tratamento com o agente desmetilante 5-AZA mostraram que ambos os transcritos são regulados pela desmetilação do DNA induzida pelo tratamento, confirmando os resultados obtidos com os experimentos em larga escala. A expressão endógena desses lncRNAs nas células da linhagem DU-145 é muito baixa, cerca de 20 e 45 moléculas de cada lncRNA por ug de RNA total, respectivamente, sendo que o tratamento com 5-AZA aumentou a expressão desses lncRNAs em aproximadamente 70 e 50 vezes, respectivamente.

#### **4.12 Expressão do mRNA codificador de proteínas correlaciona-se negativamente com o padrão de expressão do lncRNA intrônicos antissenso produzido no *locus APP***

Tendo em vista uma possível ação do lncRNA antissenso AS-APP sobre a expressão do gene codificador de proteínas expresso a partir do mesmo *locus* gênico, medimos os níveis de

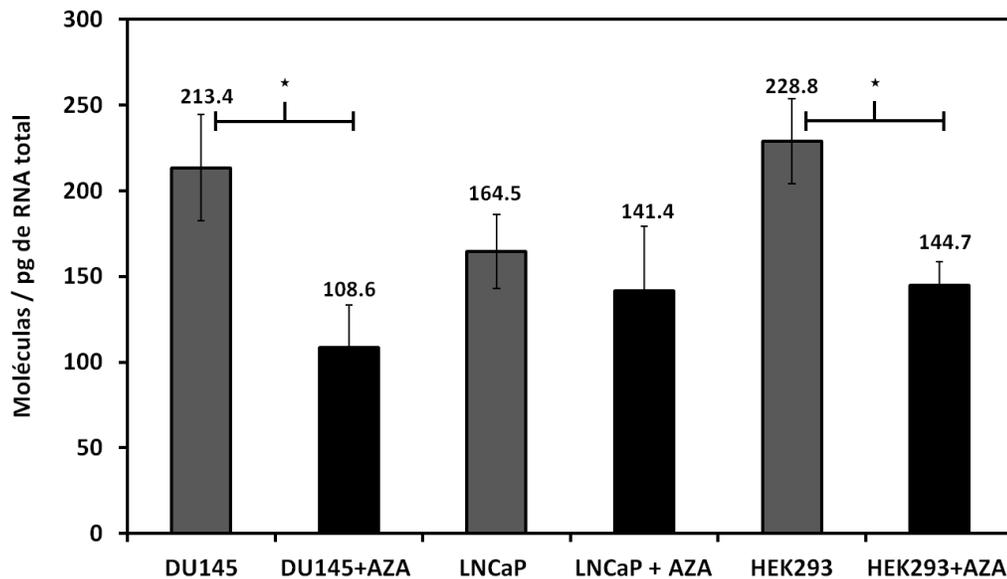
expressão de ambos os transcritos em células das linhagens DU-145 (linhagem de carcinoma de próstata), LNCaP (linhagem de carcinoma de próstata, responsiva a andrógeno) e HEK293 (linhagem embrionária de rim) tratadas e não com o agente desmetilante 5-AZA. Como podemos observar na Figura 28, o tratamento das células das linhagens DU-145 e HEK293 com 5  $\mu$ M do agente desmetilante 5-AZA por três dias aumentou a expressão do lncRNA AS-APP em cerca de 80 e 25 vezes, respectivamente. No entanto, esse efeito não foi observado na linhagem LNCaP.



**Figura 28: Quantificação absoluta da expressão endógena e após o tratamento com 5-AZA do lncRNA antissenso AS-APP em DU-145, LNCaP e HEK293.** 4  $\mu$ g de RNA total extraídos das linhagens DU-145, LNCaP e HEK293 controle e após tratamento com 5  $\mu$ M de 5-AZA por três dias foram submetidos à transcrição reversa fita específica utilizando o kit *TermoScript*<sup>®</sup> (Invitrogen). A seguir os cDNAs obtidos foram quantificados por PCR em tempo real absoluto. Nos gráficos estão plotados os números de cópias do transcrito lncRNA AS-APP por micrograma de RNA total usado na reação de transcrição reversa, expressos nas linhagens celulares com e sem a adição de 5-AZA no meio de cultura. As barras de erro representam o desvio padrão em relação à média das 3 réplicas biológicas representada nos gráficos. O teste estatístico Teste-t para diferença entre as médias foi realizado (*p-value* < 0,05).

Ao analisarmos a expressão do mRNA codificador da proteína APP, pudemos observar que ele está expresso nas células de ambas as linhagens tumorais de próstata (DU-145 e LNCaP) e também na linhagem embrionária de rim (HEK293). Curiosamente, o tratamento com o agente

desmetilante 5-AZA, por algum efeito indireto da desmetilação provocada pela droga, reduz os níveis de expressão do mRNA do gene APP em aproximadamente 50% e 40% nas linhagens DU-145 e HEK293, respectivamente. Na linhagem tumoral de próstata LNCaP, no entanto, o tratamento com 5-AZA não produziu o efeito observado para as outras linhagens (Figura 29).

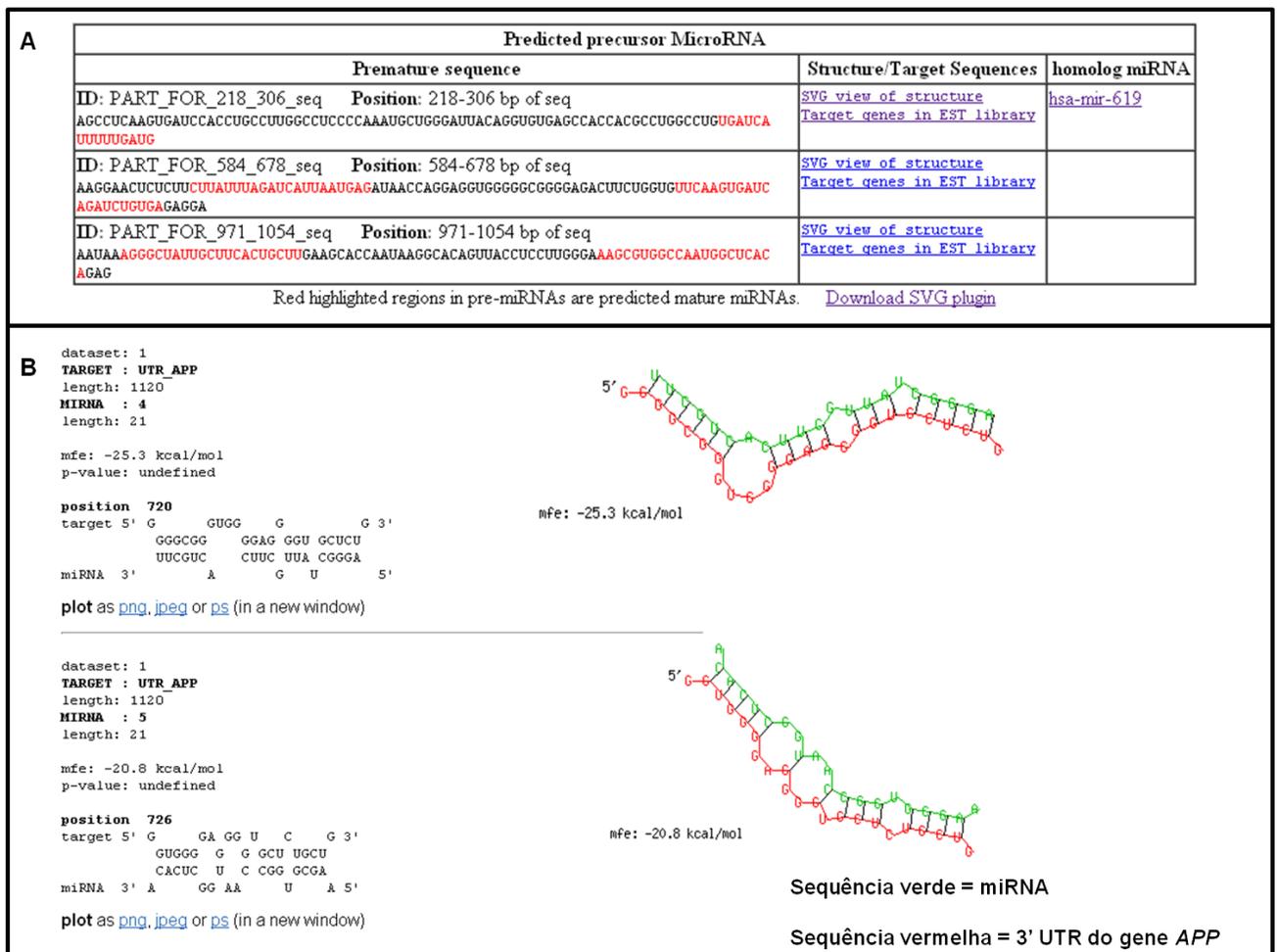


**Figura 29: Quantificação absoluta da expressão endógena e após o tratamento com 5-AZA do mRNA codificador da proteína APP em DU-145, LNCaP e HEK293.** Para a quantificação absoluta do mRNA do gene *APP*, 1 µg de RNA total extraído das células das linhagens DU-145, LNCaP e HEK293, tratadas com 5 µM de 5-AZA por três dias e controles foi submetido à transcrição reversa com oligo-dT utilizando o kit *SuperScript III* (Invitrogen). Nos gráficos estão representados os números de cópias do mRNA do gene *APP* por picograma de RNA total usado na reação de transcrição reversa, expressos nas linhagens celulares com e sem a adição de 5-AZA no meio de cultura. As barras de erro representam o desvio padrão em relação à média das 3 réplicas biológicas representada nos gráficos. O teste estatístico Teste-t para diferença entre as médias foi realizado ( $p$ -value < 0,05).

Comparando o efeito do tratamento com 5 µM do agente desmetilante 5-AZA por três dias sobre a expressão do mRNA do gene *APP* e a expressão do lncRNA AS-APP nas células das linhagens DU-145 e HEK293, pudemos observar um padrão oposto de expressão entre eles

(Figura 28 e Figura 29). Esse resultado pode indicar para uma possível ação em *cis* do lncRNA intrônico sobre a expressão do RNA codificador de proteínas do mesmo *locus*.

Um possível mecanismo de ação em *cis* do lncRNA AS-APP sobre a expressão do gene *APP* poderia ser o silenciamento de sua expressão via miRNA. Utilizando a ferramenta informática miRNAFinder (<https://sites.google.com/site/mirnatools/mirna-prediction>) encontramos possíveis precursores de pré-miRNA na estrutura primária do lncRNA AS-APP (Figura 30A). Alguns destes pré-miRNAs preditos, dão origem a miRNAs maduros que são capazes de alvejar e hibridizar na região 3' UTR do mRNA do gene *APP* (Figura 30B).

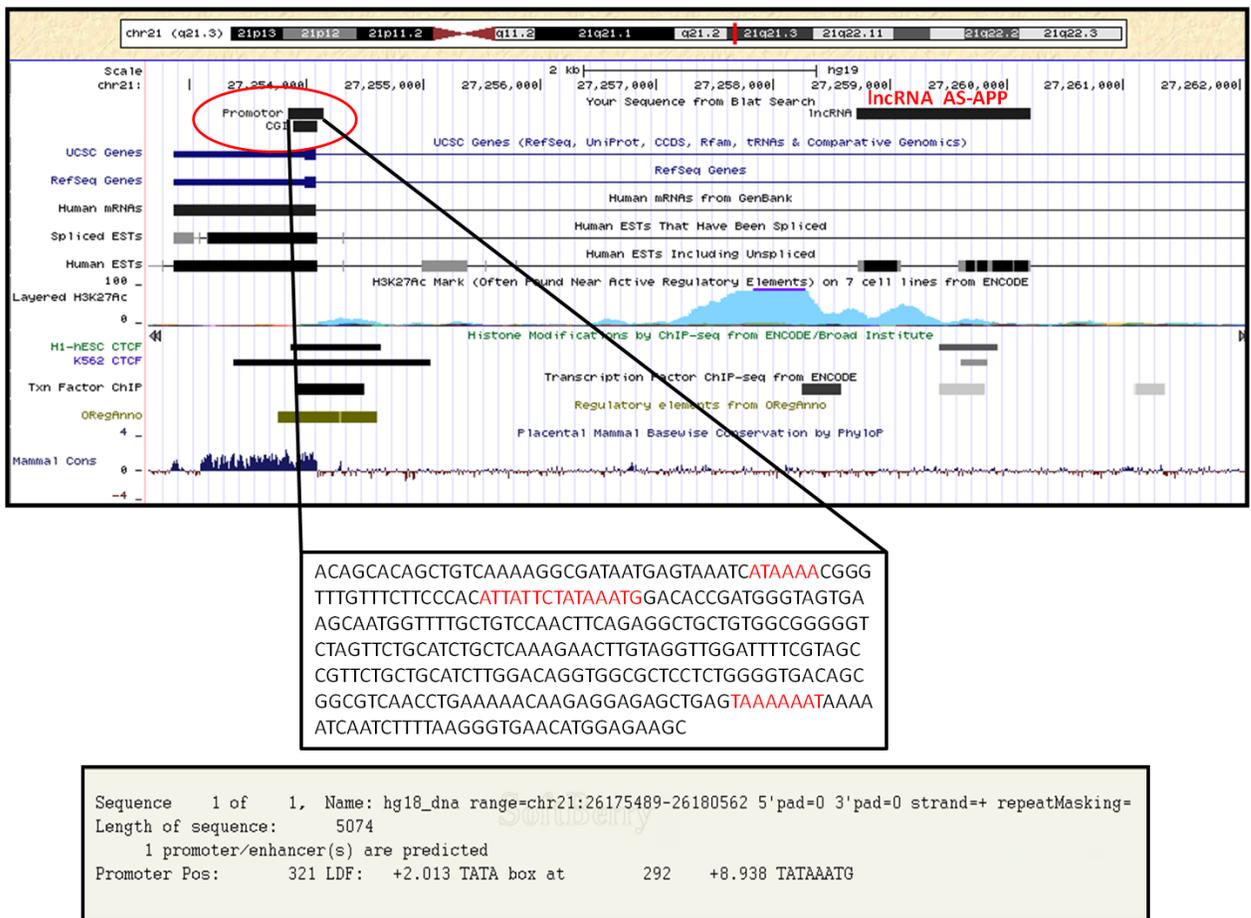


**Figura 30: Predições de precursores de miRNAs na estrutura primária do lncRNA AS-APP que alvejam o 3' UTR do gene APP.** A sequência nucleotídica do lncRNA AS-APP foi utilizada como *input* para o programa miRNAFinder. As predições de pré-miRNA identificadas geram miRNAs maduros que podem se hibridizar *in silico* na sequência 3' UTR do gene APP. Na figura estão mostrados exemplos de sequências de pré-miRNA identificados (A) e exemplos de hibridizações entre os miRNAs maduros preditos a partir do precursor pré-miRNA e o 3' UTR do gene APP (B).

#### 4.13 Identificação de possível região promotora do lncRNA AS-APP

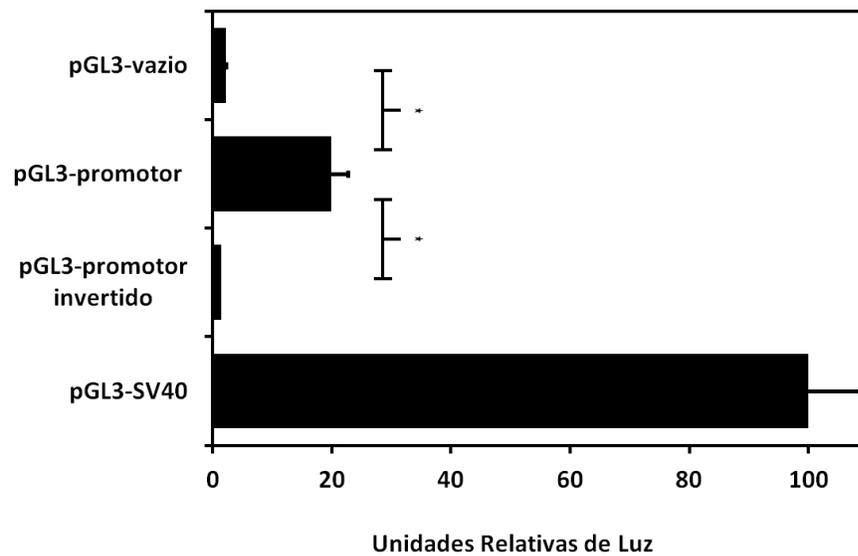
Uma análise *in silico* utilizando a ferramenta de análise bioinformática FPROM (<http://linux1.softberry.com/berry.phtml?topic=fprom&group=programs&subgroup=promoter>) (Adams *et al.*) identificou sequências características (*TATA box*) de um promotor em uma região a aproximadamente 4 kb de distância da extremidade 5' conhecida do lncRNA AS-APP e que se sobrepõem a uma CGI (Figura 31). No entanto, esta predição está localizada a um distância

relativamente grande do início do transcrito lncRNA AS-APP. Várias tentativas de se estender o transcrito em sua extremidade 5' foram realizadas. Como mostrado na Figura 26 o experimento de RACE 5' primeiramente realizado sobre a predição informática inicial resultante da montagem de ESTs disponíveis no GenBank, estendeu o transcrito em 867 nucleotídeos. Outros experimentos de RACE 5' baseando-se no transcrito estendido foram realizados sem sucesso. Tentamos também amplificar por RT-PCR a região localizada entre a predição da região promotora e o início do transcrito sem sucesso.



**Figura 31: Mapeamento genômico do lncRNA AS-APP e da predição informática de seu possível promotor.** A coordenada genômica do lncRNA AS-APP foi cruzada com o genoma humano utilizando a ferramenta BLAT disponível *site* do *Genome Browser* (<http://genome.ucsc.edu/index.html>). A predição de promotor identificada pela ferramenta FPROM (marcado com círculo vermelho) na região *upstream* ao transcrito lncRNA AS-APP está localizada a aproximadamente 4 kb de distância do 5' conhecido do mesmo.

Utilizando os dados disponíveis no *Genome Browser* podemos notar que a região identificada pela ferramenta FPROM como um possível promotor contém elementos regulatórios e sítios de ligação de fatores de transcrição. Para avaliar a sua atividade promotora, células da linhagem HEK293 foram transfectadas com um vetor contendo esse promotor clonado a montante do gene repórter que codifica a enzima luciferase. Após a transfecção a atividade promotora da predição, representada pela atividade da enzima luciferase foi medida em luminômetro utilizando o sistema *Dual-Luciferase Reporter Assay System* (Promega) (Figura 32).



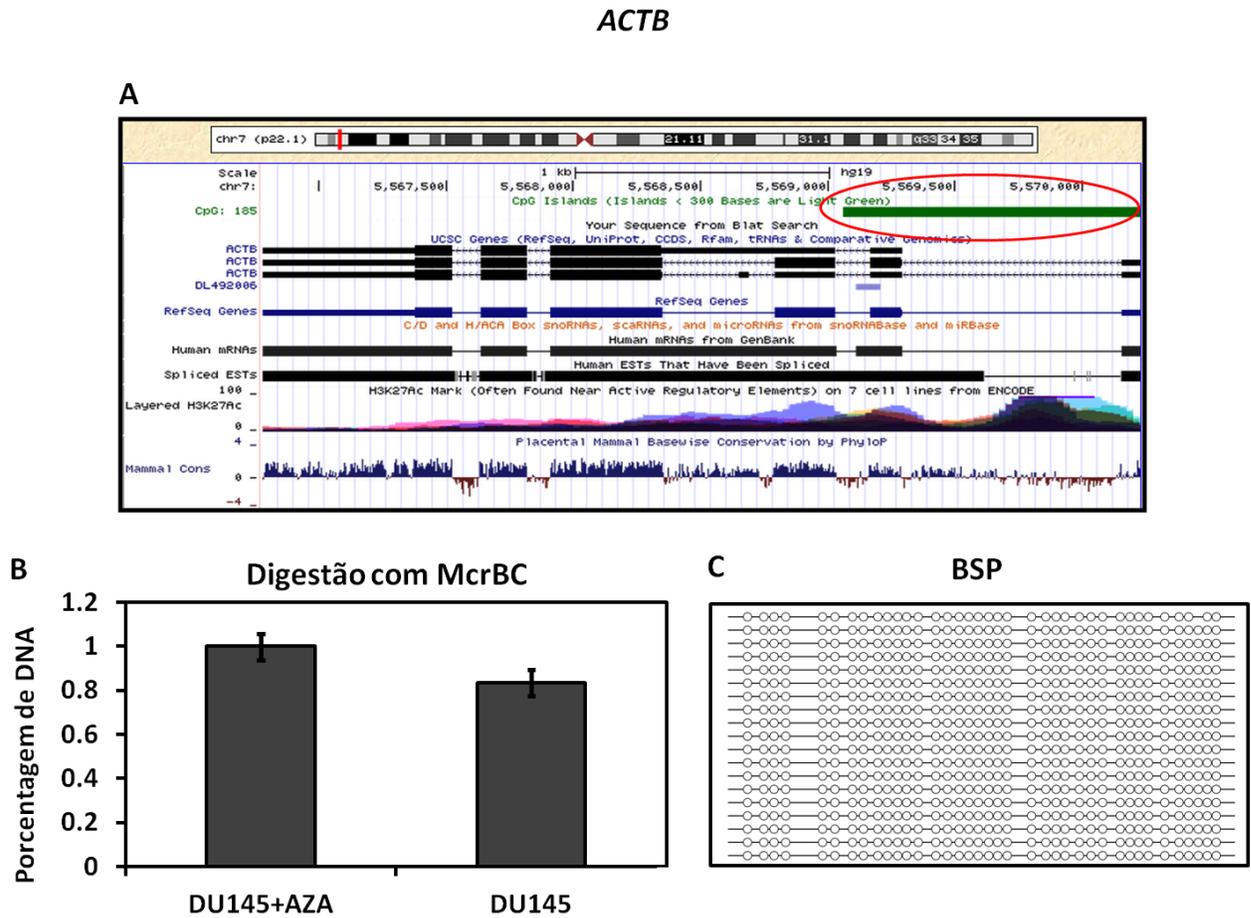
**Figura 32: Atividade do promotor do lncRNA AS-APP.** A predição do promotor obtida *in silico* pela ferramenta FPRM (pGL3-promotor) e seu complementar reverso (pGL3-promotor invertido) foram clonados *upstream* do gene repórter que codifica a enzima luciferase (de *Photinus pyralis*). Estes vetores, bem como, os controles positivo (pGL3-SV40) e negativo (pGL3-vazio) foram transfectados em células da linhagem HEK293. Após 48 horas de cultura, o ensaio de atividade foi realizado utilizando os reagentes do kit *Dual-Luciferase Reporter Assay System* (Promega). As medidas de luz emitida pela reação da luciferina com ATP catalisada pela enzima luciferase foram feitas utilizando um luminômetro. No gráfico estão plotadas as médias das medidas obtidas a partir das três réplicas biológicas realizadas e normalizadas pelo controle de transfecção (vetor contendo o promotor SV40 *upstream* do gene da luciferase de *Renilla reniformis* - pRL-SV40). As barras de erro representam o desvio padrão em relação às médias plotadas. O teste estatístico Teste-t para diferença entre as médias foi realizado ( $p$ -value < 0,05).

O experimento realizado mostrou que a predição do promotor localizado a 4 kb de distância da extremidade 5' conhecida do transcrito lncRNA AS-APP possui atividade promotora e que esta é unidirecional.

Uma vez que os lncRNAs AS-APP e AS-CELF2 foram identificados em nossas análises de expressão gênica em larga escala como candidatos a serem lncRNAs cuja transcrição é regulada por metilação do DNA, que ambos tiveram sua expressão induzida pelo tratamento com 5-AZA confirmada pelo método de quantificação absoluta por PCR em tempo real, e que ambos possuem

em sua extremidade 5' uma CGI, foram realizados experimentos para confirmar o *status* de metilação do DNA dessas ilhas.

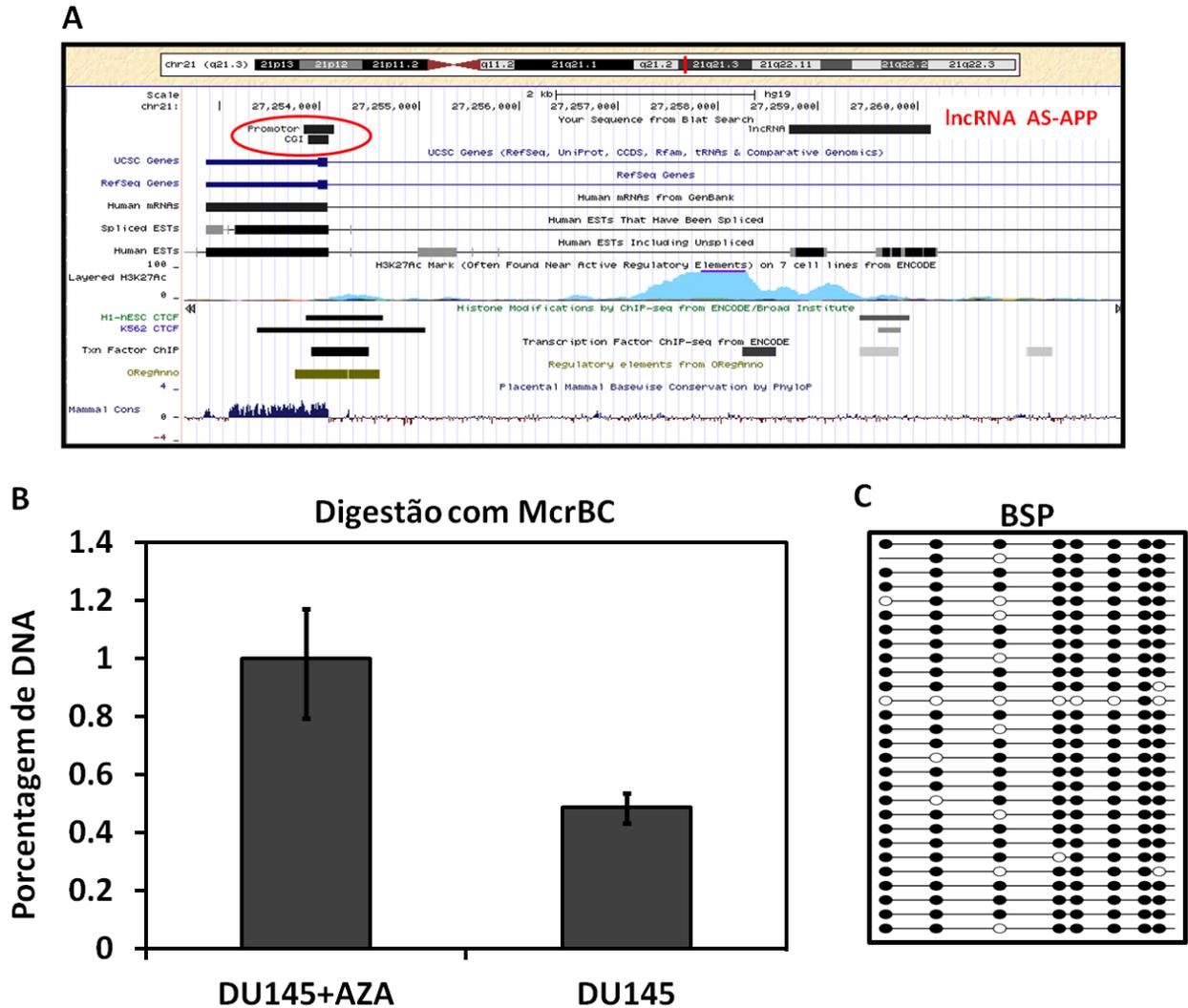
Para investigarmos o padrão de metilação das CGIs mais proximamente localizadas a 5' dos inícios conhecidos dos lncRNAs AS-APP e AS-CELF2 (Figura 34A e Figura 35A), utilizamos duas abordagens: digestão do DNA metilado com a endonuclease McrBC e sequenciamento do DNA convertido com bissulfito de sódio. A digestão do DNA metilado com a endonuclease McrBC é uma técnica relativamente simples e sensível de detecção de metilação do DNA em regiões regulatórias. Esta endonuclease é utilizada principalmente para a detecção da presença de metilação do DNA em regiões com baixa densidade de dinucleotídeos CpG (Irizarry *et al.*, 2008), característica encontrada por nós para a maior parte dos lncRNAs. A quantidade de DNA remanescente após a digestão realizada pela endonuclease pode ser medida por PCR quantitativo. Já a técnica de sequenciamento do DNA tratado com bissulfito de sódio é uma abordagem mais detalhada a respeito do *status* de metilação de cada dinucleotídeo CpG que se quer investigar dentro da CGI. Com esse método, citosinas não metiladas do DNA são convertidas a uracilas, enquanto que 5-metilcitosinas, que sofrem essa modificação muito mais lentamente, permanecem de maneira geral, como citosinas. A digestão do DNA com a endonuclease McrBC e o tratamento com bissulfito de sódio foram realizados em amostras de gDNA extraídos de células da linhagem DU-145. Como controle da digestão do DNA metilado com McrBC utilizamos a CGI localizada na região promotora do gene *ACTB* (Figura 33). Este gene é de expressão constitutiva em células da linhagem DU-145, ou seja, não se apresenta metilado nesta linhagem.



**Figura 33: Digestão com McrBC e sequenciamento bissulfito da ilha CpG localizada na região promotora do gene *ACTB*.** Em (A) está mostrado o contexto genômico da CGI (retângulo verde, circulado em vermelho) do gene *ACTB*. Em (B) Digestão do DNA com McrBC. Amostras de 4 µg de gDNA extraído de células da linhagem DU-145 tratadas com 5 µM do agente desmetilante 5-AZA por três dias (DU145+AZA) e controle foram submetidas à digestão com a endonuclease McrBC. A seguir, a quantidade de DNA remanescente após a digestão com McrBC, correspondente à CGI do gene *ACTB* foi quantificada por PCR em tempo real. Os resultados mostrados no gráfico foram normalizados em relação às medidas obtidas com as mesmas amostras submetidas ao protocolo sem a adição de enzima. As barras de erro referem-se ao desvio-padrão das médias das três réplicas biológicas medidas. Em (C) Sequenciamento do DNA convertido com bissulfito de sódio (BSP). As amostras de gDNA de células de DU-145 foram convertidas por bissulfito de sódio. A seguir, a CGI do gene *ACTB* foi amplificada a partir dessas amostras de gDNA convertido, clonadas e sequenciadas. No painel está mostrado o sequenciamento de 19 clones para essa ilha. Cada círculo representa uma citosina CpG sequenciada, sendo que círculos pretos indicam presença de metilação e círculos brancos indicam ausência de metilação.

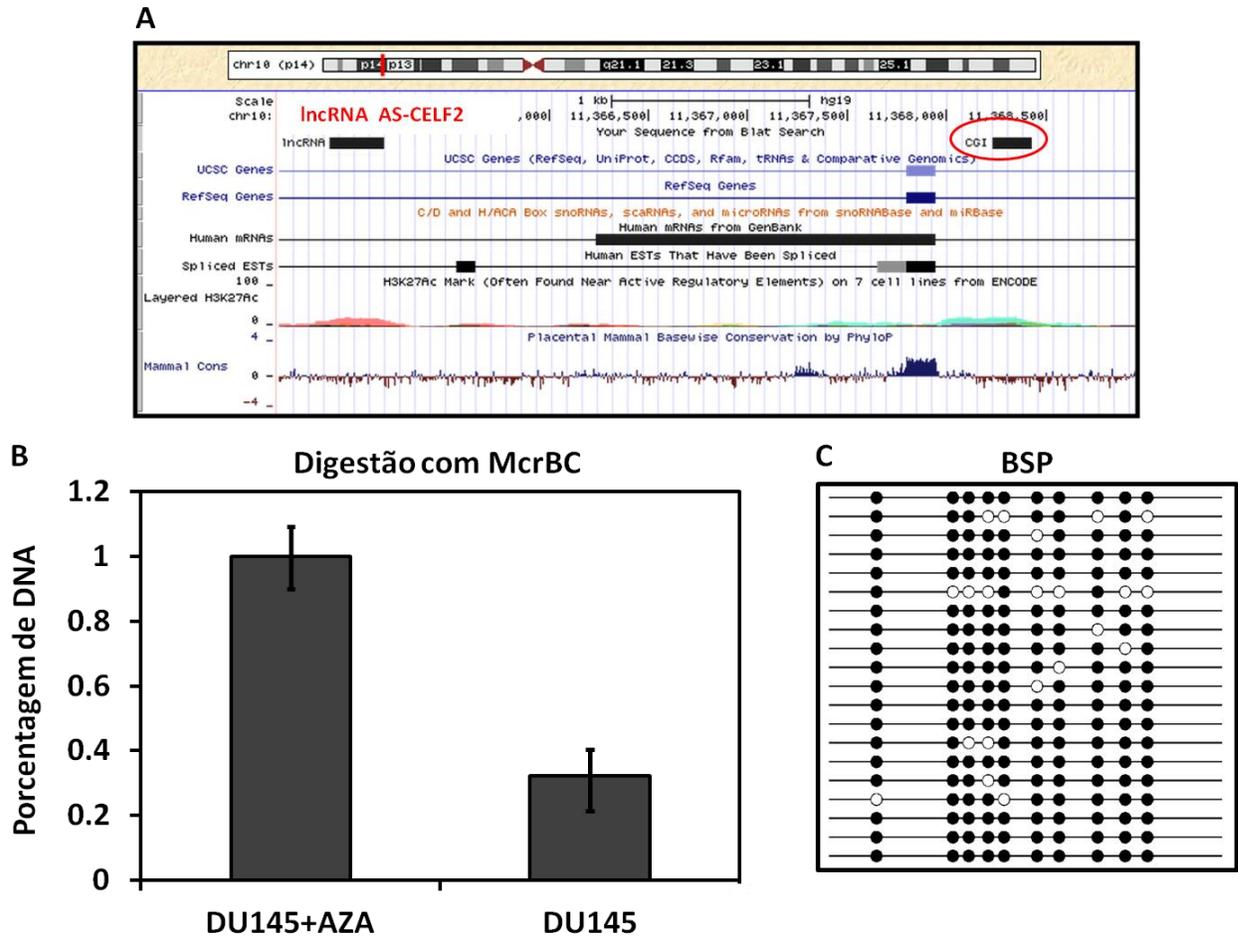
Como mostrado na Figura 33, painel C, o sequenciamento bissulfito confirmou que como esperado, a ilha CpG do gene *ACTB* não está metilada. Após digestão do DNA pela endonuclease McrBC, observamos uma diferença de cerca de 17% na abundância da região que compreende a CGI no promotor de *ACTB*, comparando células tratadas ou não com 5-AZA (Figura 33, painel B). Desse modo, foi definido para os nossos experimentos com digestão do DNA metilado pela endonuclease McrBC o valor de 17% de diferença de abundância de DNA remanescente entre DNA de células tratadas com 5-AZA e controle como produto de uma digestão inespecífica. Sendo assim, para que as CGIs dos lncRNAs AS-APP e AS-CELF2 fossem consideradas metiladas uma digestão maior que 17% deveria ser observada. Usando esse critério, pudemos confirmar a presença de metilação nas CGIs dos lncRNAs AS-APP e AS-CELF2 (Figura 34 e Figura 35, painéis B). A presença de metilação nas CGIs foi posteriormente confirmada por sequenciamento do DNA convertido com bissulfito de sódio (Figura 34 e Figura 35, painéis C).

## lncRNA AS-APP



**Figura 34: Caracterização do *status* de metilação do DNA em CGI localizada na região promotora putativa do lncRNA AS-APP.** Em (A) está mostrado o contexto genômico da CGI (retângulo preto, circulado em vermelho) mais proximamente localizada na extremidade 5' do lncRNA AS-APP. Em (B) **Digestão do DNA metilado pela endonuclease McrBC.** Amostras de 4  $\mu$ g de gDNA extraído de células da linhagem DU-145 tratadas com 5  $\mu$ M do agente desmetilante 5-AZA por três dias (**DU145+AZA**) e controle foram submetidas à digestão com a endonuclease McrBC. A seguir, a quantidade de DNA remanescente após a digestão, correspondente à região genômica da CGI foi quantificada por PCR em tempo real. Os resultados mostrados nos gráficos foram normalizados em relação às medidas obtidas com as mesmas amostras submetidas ao protocolo sem a adição de enzima. As barras de erro referem-se ao desvio-padrão das médias das três réplicas biológicas medidas. Em (C) **Sequenciamento do DNA convertido com bissulfito de sódio.** Amostras de gDNA de células de DU-145 foram convertidas por bissulfito de sódio. A seguir, a CGI foi amplificada, clonada e sequenciada. No painel estão mostrados 28 clones sequenciados. Cada círculo representa uma citosina CpG sequenciada, sendo que círculos pretos indicam presença de metilação e círculos brancos indicam ausência de metilação.

## IncRNA AS-CEL2



**Figura 35: Caracterização do *status* de metilação do DNA da CGI localizada na região promotora putativa do IncRNA AS-CEL2.** Em (A) está mostrado o contexto genômico da CGI (retângulo preto, circulado em vermelho) mais proximamente localizada na extremidade 5' do IncRNA AS-CEL2. Em (B) **Digestão do DNA metilado pela endonuclease McrBC.** Amostras de 4  $\mu$ g de gDNA extraído de células da linhagem DU-145 tratadas com 5  $\mu$ M do agente desmetilante 5-AZA por três dias (**DU145+AZA**) e controle foram submetidas à digestão com a endonuclease McrBC. A seguir, a quantidade de DNA remanescente após a digestão, correspondente à região genômica da CGI foi quantificada por PCR em tempo real. Os resultados mostrados nos gráficos foram normalizados em relação às medidas obtidas com as mesmas amostras submetidas ao protocolo sem a adição de enzima. As barras de erro referem-se ao desvio-padrão das médias das três réplicas biológicas medidas. Em (C) **Sequenciamento do DNA convertido com bissulfito de sódio.** Amostras de gDNA de células de DU-145 foram convertidas por bissulfito de sódio. A seguir, a CGI foi amplificada, clonada e sequenciada. No painel estão mostrados 20 clones sequenciados. Cada círculo representa uma citosina CpG sequenciada, sendo que círculos pretos indicam presença de metilação e círculos brancos indicam ausência de metilação.

A partir de ambas as abordagens para a análise do padrão de metilação do DNA pode-se constatar que as CGIs localizadas nas regiões a 5' dos lncRNAs AS-APP e AS-CELF2 estão metiladas. Com a metodologia de digestão do DNA metilado pela endonuclease McrBC, pudemos também comparar o efeito do tratamento com 5-AZA sobre a desmetilação daquelas CGIs analisadas. Para a CGI localizada a 5' do início conhecido do transcrito lncRNA AS-APP, foi medida uma porcentagem de digestão do DNA extraído de células da linhagem DU-145 pela endonuclease McrBC de 50% em relação ao DNA extraído de células da linhagem DU-145 tratadas com 5-AZA. Para a CGI localizada a 5' do início conhecido do transcrito lncRNA AS-CELF2, a porcentagem de digestão com McrBC medida foi de 70% do DNA metilado (células da linhagem DU-145) em relação ao DNA desmetilado (células da linhagem DU-145 tratadas com 5-AZA).

Em conjunto, esses resultados indicam que os lncRNAs AS-APP e AS-CELF2, cujos níveis de expressão aumentaram em cerca de 50 e 70 vezes após tratamento com o agente desmetilante 5-AZA têm sua transcrição controlada por metilação do DNA. Esses resultados explicam os baixos níveis de expressão desses lncRNAs na linhagem DU-145 (aproximadamente 20 e 45 moléculas dos transcritos AS-APP e AS-CELF2, respectivamente, por  $\mu\text{g}$  de RNA total utilizado para transcrição reversa (Figura 27)).

## 5 DISCUSSÃO

### 5.1 Expressão de lncRNAs intrônicos e mRNAs nas linhagens tumorais DU-145, MCF-7 e Mia PaCa 2.

Neste trabalho foram analisados o perfil de expressão de lncRNAs intrônicos senso e antissenso em três linhagens de células tumorais humanas com diferentes origens histológicas. Utilizando a metodologia de hibridização de RNAs marcados com sondas de oligonucleotídeos imobilizadas em lâminas customizadas (*oligoarrays*) identificamos 5.625 lncRNAs antissenso e 3.504 lncRNAs senso expressos em pelo menos uma das linhagens, sendo que uma fração significativa desses conjuntos de lncRNAs intrônicos também foi detectado, de forma independente, por outro método de análise do transcriptoma em outras sete linhagens celulares humanas (projeto *ENCODE Caltech RNA-seq Libraries*) e em nove tecidos humanos diferentes (RNA-seq Burge) (Wang *et al.*, 2008).

Foi observado que lncRNAs intrônicos são detectados em uma faixa de intensidade menor do que mRNAs codificadores de proteínas. Esta observação corrobora outros trabalhos na literatura que descreveram uma menor abundância relativa de ncRNAs em comparação a transcritos codificadores de proteínas (Kampa *et al.*, 2004; Ravasi *et al.*, 2006; Nakaya *et al.*, 2007; Guttman *et al.*, 2010; Cabili *et al.*, 2011).

Outra característica observada foi um padrão de expressão tecido-específico para a maior parte dos ncRNAs intrônicos representados na lâmina. O fato de lncRNAs intrônicos possuírem expressão preferencialmente restrita a tecidos específicos sugere que sua transcrição seja regulada de modo a exercer papéis importantes na especialização/diferenciação celular.

A observação de que a intensidade de expressão dos ncRNAs é menor em comparação aos mRNAs, tem levado alguns autores a sugerirem que estas mensagens não codificadoras

representam apenas ruído transcricional produzido pelo reconhecimento espúrio de promotores pela RNA polimerase II em regiões da cromatina que se encontram acessíveis para permitir a expressão de genes codificadores de proteínas (Cawley *et al.*, 2004; Kapranov *et al.*, 2005; Struhl, 2007; van Bakel *et al.*, 2010). Por outro lado, a baixa correlação observada entre a expressão do mRNA codificador de proteínas e do lncRNA expresso no mesmo *locus*, somada ao fato de que esses lncRNAs são expressos de maneira tecido específica podem ser interpretadas como uma indicação de que ao menos uma fração desses transcritos intrônicos sem potencial codificador de proteínas possam desempenhar funções biológicas importantes, possivelmente no controle fino de processos celulares complexos (Mattick e Makunin, 2006), tais como o controle epigenético da expressão gênica (Rinn e Chang, 2012) e o desenvolvimento (Amaral e Mattick, 2008).

## **5.2 Correlação de expressão entre lncRNAs intrônicos e mRNAs codificadores de proteínas**

Embora não haja na literatura evidências funcionais sobre os lncRNAs como um grupo, para diversos transcritos individuais têm sido estabelecidos papéis regulatórios na transcrição gênica, sendo que, esse papel pode ser exercido tanto em *cis*, quanto em *trans*. Muitos dos lncRNAs bem estudados, tais como aqueles envolvidos na compensação de dose e *imprinting* (*Air*, *Xist* e *Kcnq1ot1*), regulam a expressão gênica em *cis* (Peters e Robson, 2008; Lee, 2009), enquanto que outros, como *HOTAIR* e *linc-p21* podem regular a expressão de genes distantemente localizados, em *trans* (Rinn *et al.*, 2007; Gupta *et al.*, 2010; Huarte *et al.*, 2010). Entre todas as unidades transcricionais encontradas nos diversos organismos eucariotos, cerca de 5% a 70% delas possuem pelo menos um transcrito antissenso endogenamente expresso complementar agindo em *cis* (*cis-NATs*), sendo que a proporção exata é fortemente dependente

da cobertura do transcriptoma considerado (Lapidot e Pilpel, 2006). Por exemplo, utilizando abordagens em larga escala, foi identificado que cerca de 70% dos transcritos expressos em camundongos possuem parceiros antissenso (Katayama *et al.*, 2005), e que a perturbação do RNA antissenso pode alterar a expressão do gene senso (Rosok e Sioud, 2004). Em humanos, cerca de 20% dos transcritos codificadores formam pares senso-antissenso (Chen *et al.*, 2004), e quando analisado o *dataset* de CAGE *tags* humanas, foi identificado que para cada *locus* gênico codificador de proteínas está associado, em média, 6 *cis-NATs* (Conley *et al.*, 2008). *cis-NATs* têm suas funções dependentes da formação do pareamento com o mRNA complementar (Lapidot e Pilpel, 2006). Esse mecanismo de ação, implica que os pares *cis-NATs/mRNA* devam ser coexpressos ou inversamente expressos (Chen *et al.*, 2005). Recentemente, nosso grupo de pesquisa identificou e caracterizou dois lncRNAs intrônicos antissenso que possuem a função de regular a expressão dos genes codificadores de proteínas de seus respectivos *loci*. Um destes é antissenso ao *locus* do gene *RASSF1A*. Este lncRNA, *ANRASSF1*, é responsável por guiar o complexo repressivo da transcrição *polycomb* (PRC2) para a cromatina da região promotora do gene codificador de proteínas transcrito no mesmo *locus*, modulando a sua expressão e a proliferação celular (Beckedorff, *et al.*, 2012)<sup>7</sup>. O outro, é antissenso ao gene *BCL-X*. Este transcrito não codificador intrônico, atua sobre o processamento do mRNA do gene *BCL-X*, altera a abundância relativa das isoformas longa e curta desse gene, favorecendo a isoforma curta, e

---

<sup>7</sup> Beckedorff, F. C.; Crocci-Souza, R.; Amaral, M. S.; Nakaya, H. I.; Soltys, D. T.; Ayupe, A. C.; Menck, C. F. M.; Reis, E. M.; Verjovski-Almeida, S. (2012). "Intronic unspliced long noncoding RNA ANRASSF1 recruits PRC2 and modulates RASSF1A expression and cell proliferation." Submetido

contribuindo para a apoptose de células tumorais *in vitro* e a regressão de tumores *in vivo* (DeOcesano-Pereira, *et al.*, 2012)<sup>8</sup>.

Os lncRNAs intrônicos senso coexpressos com os mRNAs codificadores de proteínas do mesmo *locus*, cuja abundância de expressão reflete o nível de expressão do mRNA codificador, podem representar pré-mRNAs ou ainda, *intron lariats* resultantes do *splicing* dos pré-mRNAs. Outra hipótese para explicar a coexpressão desses lncRNAs senso com genes codificadores de proteínas expressos a partir do mesmo *locus* é que estes possam representar UTRs alternativos raramente expressos de transcritos codificadores de proteínas ainda não documentados. De fato, um dos candidatos selecionados para caracterização mostrou-se ser uma extensão da 3' UTR do mRNA codificador do gene *HIC1*.

Também já foi documentada a existência de lncRNAs intrônicos senso funcionais, que exercem seu papel sobre o transcrito expresso a partir do mesmo *locus*. Um exemplo é o lncRNA *COOLDAIR* (*Cold Assisted Intronic Noncoding RNA*), que em *Arabidopsis thaliana*, promove a repressão gênica do mRNA codificador de proteínas *FLC*, transcrito a partir do mesmo *locus*, através do direcionamento do complexo PRC2 para a cromatina de sua região promotora (Heo e Sung, 2011).

Cerca de 31% dos pares lncRNA senso/mRNA (468 do total de 1.519) e 25% dos lncRNA antissenso/mRNA (603 do total de 2.461) possuem correlação de expressão positiva ou negativa nas linhagens DU-145, MCF-7 e Mia PaCa 2. Entre os mRNAs codificadores de proteínas positivamente coexpressos ou inversamente expressos com os lncRNAs intrônicos senso e antissenso de nossas análises, estão significativamente enriquecidos genes associados a regulação

---

<sup>8</sup> DeOcesano-Pereira, C.; Parreira, K. S.; Amaral, M. S.; Ayupe, A. C.; Reis, E. M.; Verjovski-Almeida, S. (2012). "Intronic long noncoding RNA APOLLON modulates BCL-X alternative splicing, induces apoptosis in cell lines and promotes tumour regression *in vivo*." Em preparo.

de processos celulares (GO: 0050794), sub-localizados no núcleo das células (GO: 0005634), com atividades que envolvem a interação com proteínas (GO: 0005515), e envolvidos em processos desenvolvimentais (GO: 0032502), diferenciação celular (GO: 0042692) e morfogênese de estruturas anatômicas (GO: 0009653). Estes resultados corroboram a ideia de que lncRNAs possam regular finamente a expressão de genes que codificam para proteínas com funções regulatórias que se localizam no núcleo das células, possivelmente envolvidos no controle da transcrição de genes envolvidos na diferenciação celular e no desenvolvimento (Amaral e Mattick, 2008).

Um método bioinformático denominado “*guilt by association*” tem sido utilizado para identificar possíveis associações funcionais entre a expressão de lncRNAs e genes codificadores de proteínas (Guttman *et al.*, 2009). Este método permite identificar genes codificadores de proteínas e vias significativamente correlacionadas com um dado lncRNA a partir da análise de expressão gênica. Com base nas funções conhecidas dos genes codificadores de proteínas coexpressos, hipóteses são geradas para atribuir funções e potenciais regulatórios aos lncRNAs. Além disso, esta análise revela famílias de lncRNAs com base nas vias com as quais eles se associam ou não. Esta ferramenta tem predito diversos papéis para os lncRNAs, como envolvimento com diferenciação celular, pluripotência, adipogênese e câncer (Guttman *et al.*, 2009; Loewer *et al.*, 2010; Hung *et al.*, 2011). No trabalho de Huarte e colaboradores, foi observado que vários lncRNAs identificados como altamente correlacionados com a via do gene *p53* pela análise de “*guilt by association*” tem sua expressão modulada de uma maneira *p53*-dependente (Huarte *et al.*, 2010). Este mesmo trabalho demonstrou que um dos lncRNAs preditos a estarem associados a via de *p53*, chamado *lincRNA-p21*, é diretamente regulado por *p53*. Ele se associa à hnRNP-K, uma proteína nuclear que age como um repressor transcricional global e facilita a apoptose mediada por *p53* (Huarte *et al.*, 2010).

A análise de “*guilt by association*” para os lncRNAs intrônicos expressos nas linhagens estudadas revelou 67 lncRNAs antissenso e 15 lncRNAs senso cuja expressão está associada com processos biológicos que incluem a replicação do DNA e o ciclo celular. Este resultado é interessante uma vez que já foram identificados pela análise de “*guilt by association*” outros lncRNAs associados com estas mesmas funções biológicas (Hung *et al.*, 2011), sendo que alguns deles já foram caracterizados funcionalmente. Um exemplo é o lncRNA *CDKN2B-ASI* (também conhecido como *p15AS* ou *ANRIL*) que é transcrito a partir do *locus* do gene *CDKN2B*, um inibidor de quinase dependente de ciclina. Este lncRNA, age silenciando transcricionalmente o *locus INK4b-ARF-INK4a* (que inclui os genes *CDKN2B* e *CDKN2A*, ambos envolvidos na regulação do ciclo celular), através do direcionamento das subunidades EZH2 e CBX7 dos complexos repressores PRC2 e PRC1, respectivamente, induzindo a formação de heterocromatina no *locus* (Yu *et al.*, 2008; Yap *et al.*, 2010). Outro exemplo é o lncRNA *TERRA*. Este lncRNA controla o acesso da telomerase aos telômeros durante a fase S do ciclo celular (Flynn *et al.*, 2011). Desse modo, as funções hipotetizadas para os lncRNAs intrônicos expressos nas linhagens celulares DU-145, MCF-7 e Mia PaCa 2 pelas análises informáticas de correlação de expressão em *cis* e “*guilt by association*” corroboram as funções experimentalmente determinadas para outros lncRNAs já bem caracterizados na literatura. Ensaios experimentais de superexpressão e depleção desses lncRNAs intrônicos em modelos celulares devem ser realizados a fim de se validar essas possibilidades funcionais.

### 5.3 Identificação de elementos regulatórios localizados à 5' de lncRNAs intrônicos

Já foi demonstrado por nosso grupo de pesquisa que uma grande parte dos lncRNAs intrônicos são sintetizados pela RNAP II (Nakaya *et al.*, 2007) e (Ayupe, 2012)<sup>9</sup>. Na maioria dos eucariotos, os transcritos sintetizados por essa polimerase são modificados co-transcricionalmente em sua extremidade 5' pela adição da estrutura 7-metil guanosina (cap 5') (Cougot *et al.*, 2004). Assim, a presença dessa modificação é um forte indício de um evento de transcrição específico promovido pela RNAP II (Kapranov, 2009). A transcrição de lncRNAs catalisada pela RNAP II e a presença de cap na extremidade 5' de desses transcritos, têm sido reportada em uma série de publicações (Carninci *et al.*, 2005; Wyers *et al.*, 2005; Seidl *et al.*, 2006; 2009; Guttman *et al.*, 2009; Redrup *et al.*, 2009; Flynn *et al.*, 2011; Moran *et al.*, 2012). Não foi ainda estudada sistematicamente a presença da modificação 5' cap em lncRNAs intrônicos.

O mapeamento de sequências curtas derivadas de cDNAs que possuem cap 5' (*CAGE tags*) no genoma tem predito um grande número de promotores (Carninci *et al.*, 2005). Sabe-se também, que a RNAP II se acumula em regiões próximas à promotores de genes ativos e inativos (Brodsky *et al.*, 2005; Kim *et al.*, 2005; Guenther *et al.*, 2007; Muse *et al.*, 2007; Rahl *et al.*, 2010; Jishage *et al.*, 2012). Além disso, a presença de marcas ativadoras da cromatina estão geralmente associadas com regiões transcricionalmente ativas e transcritas pela RNAP II. Tem sido demonstrado que essas marcas são poderosas ferramentas para detectar a existência de novos lncRNAs, como no caso de lincRNAs identificados a partir de regiões de cromatina

---

<sup>9</sup> 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.

transcricionalmente ativa denominados “domínios K4-K36” (Guttman *et al.*, 2009)<sup>10</sup>. A presença de ilhas CpG (CGIs) é outro marco regulatório de regiões promotoras. Foi estimado que cerca de 60% a 70% dos promotores humanos estão associados com CGIs (Antequera e Bird, 1993; Saxonov *et al.*, 2006), sendo que a maioria deles não possuem elementos *TATA-box* (Sandelin *et al.*, 2007). Estes promotores são mais frequentemente associados com genes constitutivamente expressos nas células (*housekeeping genes*) (Schug *et al.*, 2005), embora existam muitas exceções, incluindo genes especificamente expressos no cérebro (Gustincich *et al.*, 2006).

Em nossas análises, observamos o enriquecimento significativo (teste KS;  $p$ -value < 0,05) para a presença de *CAGE tags*, RNAP II, CGIs e para a marca de histona H3K4me3 junto a extremidade 5' conhecida dos lncRNAs intrônicos expressos nas linhagens quando comparados à sequências genômicas randomicamente selecionadas.

A presença de *CAGE tags* foi associada a 20% e 8% das regiões próximas aos TSSs preditos a partir dos *contigs* de lncRNAs senso e antissenso, respectivamente, ou seja, foi observada a presença de *CAGE tags* em até no máximo 1 kb da extremidade conhecida do lncRNA. Os dados de *CAGE tags* disponíveis não compreendiam bibliotecas geradas a partir das linhagens analisadas. Considerando o padrão de expressão tecido-específico de lncRNAs intrônicos, a baixa cobertura ou a ausência da expressão dos lncRNAs interrogados na lâmina de oligonucleotídeos nas bibliotecas de *CAGE tags* podem explicar a baixa sobreposição observada acima.

Nós observamos a presença de ligação de RNAP II e da marca de transcrição ativa (H3K4me3) próximo ao TSS predito de cerca de 30% a 40% dos *contigs* de lncRNAs intrônicos (senso e antissenso) expressos nas linhagens DU-145, MCF-7 e Mia PaCa 2, sugerindo que uma

---

<sup>10</sup> Neste artigo, Guttman e colaboradores, identificaram aproximadamente 1.600 *large intervening non-coding RNAs* (lincRNAs) em quatro tipos celulares de camundongos associados às marcas de histona H3K4me3 e H3K36me3 utilizando uma abordagem informática.

porção dos lncRNAs seja transcrito pela RNAP II. Essas proporções de lncRNAs intrônicos informaticamente associados à transcrição realizada pela RNAP II corroboram as proporções observadas experimentalmente para um conjunto de lncRNAs intrônicos expressos em HeLa<sup>11</sup>. Apenas uma pequena fração, menos de 10% dos *contigs* de lncRNAs intrônicos senso e antissenso expressos nas linhagens possuem CGIs associadas à regiões próximas aos seus TSSs preditos. Por outro lado, observou-se a presença de CGIs em cerca de 50% dos mRNAs expressos. Este resultado está de acordo com o que se observa na literatura sobre a presença de CGIs em promotores de genes codificadores de proteínas (Antequera e Bird, 1993), e sugere que a regulação da expressão de lncRNAs intrônicos seja menos dependente da presença de CGIs nas regiões promotoras. Além disso, é sabido que muitos promotores que contém CGIs estão associados a genes com padrões de expressão não tecido-específico (Schug *et al.*, 2005), característica observada para a maioria dos mRNAs expressos em nossas análises e não para os lncRNAs intrônicos.

Estes dados, em conjunto com o fato de que a maioria do lncRNAs intrônicos não tem expressão correlacionada com os mRNAs codificadores de proteínas expressos a partir do mesmo *locus*, reforçam a ideia de que lncRNAs intrônicos são originados à partir de unidades transcricionais independentes, apresentando regiões regulatória próprias.

#### **5.4 Conservação evolutiva de lncRNAs intrônicos**

Em nossa análise, do total de lncRNAs intrônicos expressos em pelo menos uma das linhagens DU-145, MCF-7 e Mia PaCa 2, encontramos evidências de conservação evolutiva ao

---

<sup>11</sup> 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.

nível do DNA entre vertebrados para 4.417 (78% do total de expressos) e 1.778 (51% do total de expressos) lncRNAs antissenso e senso, respectivamente. Em uma análise semelhante, Chodroff e colaboradores encontraram que 59% de um conjunto de 2.055 lncRNAs intergênicos de camundongos se sobrepunham a elementos conservados entre vertebrados (Chodroff *et al.*, 2010). Foi atribuída a evidência de conservação de expressão para 3.562 (63% do total de expressos) lncRNAs antissenso e 888 lncRNAs senso (25% do total de expressos), uma vez que estes são sintenticamente pareados com pelo menos um transcrito ortólogo desde peixes até primatas. Além disso, para 52% (2.911 do total de expressos) e 13% (422 do total de expressos) dos lncRNAs intrônicos antissenso e senso, respectivamente, pudemos encontrar evidências de conservação de sequência e de conservação de expressão. Para ambas as análises foram testadas as probabilidades de se encontrar conservação ao acaso. Em ambas, foram encontrados um número estatisticamente significativo maior de lncRNAs associados à conservação que o esperado ao acaso. Um grupo menor de lncRNAs intrônicos, 623 (11% do total de expressos) antissenso e 342 (10% do total de expressos) senso, parecem possuir estrutura secundária estável evolutivamente conservada. Ao sobrepormos as três análises, encontramos os grupos de 361 (6% do total de expressos) e 59 (2% do total de expressos) lncRNAs intrônicos antissenso e senso, respectivamente, que possuem suas sequências conservadas e expressas entre vertebrados, além de possuírem estrutura secundária conservada evolutivamente.

Apesar de termos encontrado evidências de conservação, tanto estrutural, como de expressão, o que sugere uma possível funcionalidade para uma parte significativa dos lncRNAs intrônicos, vale ressaltar que a ausência ou baixa conservação não significa necessariamente que ncRNAs não possam ser funcionais. De fato, alguns lncRNAs bem caracterizados funcionalmente em mamíferos possuem pouca conservação de suas sequências em outras espécies, tais como o

*Xist*<sup>12</sup> e *HOTAIR*<sup>13</sup>. Sequências codificadoras de proteínas devem ser altamente conservadas uma vez que elas necessitam manter uma ORF que produza um produto proteico termodinamicamente estável. Similarmente, a alta conservação observada para os miRNAs e snoRNAs conhecidos (Cavaille *et al.*, 2000; Huttenhofer *et al.*, 2001; Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee e Ambros, 2001; Lagos-Quintana *et al.*, 2003) pode ser explicada pelo fato de que estas moléculas necessitam parear por complementaridade com outras sequências de ácidos nucleicos para exercerem suas funções. Principalmente os miRNAs, que são moléculas curtas (tipicamente entre 21 a 25 nt), apenas uma pequena mudança em suas sequências pode alterar fundamentalmente sua função. Além disso, muitos dos miRNAs parecem possuir múltiplos alvos na célula (Lewis *et al.*, 2005; Mattick e Makunin, 2005), o que poderia limitar severamente suas possibilidades de co-variação de sequência e deriva genética. Assim, o grau de pressão seletiva sob a qual uma determinada sequência está submetida é uma função do número e dos tipos de interações que devem ser mantidas para a manutenção da sua funcionalidade, mais do que a sequência nucleotídica precisa em si. Aquelas que possuem apenas um ou poucos parceiros podem estar submetidas a uma deriva genética mais rápida e também podem explorar novas interações em redes regulatórias. Assim, é provável que muitos dos lncRNAs ainda não caracterizados funcionalmente possam ter poucos alvos e exibirem pouca conservação. LncRNAs funcionais pobremente conservados poderiam evoluir rapidamente, uma vez que eles interagiriam com um número limitado de sequências alvo, um cenário que permitiria o aumento da co-variação ao longo do tempo evolutivo. Entre esses, poderiam estar incluídos os transcritos *cis*-

---

<sup>12</sup> *Xist* está presente em todos os mamíferos placentários (Eutheria) examinados até o momento, mas está ausente em marsupiais (Marsupialia) e monotremados (Monotremata) (Duret *et al.*, 2006; Davidow *et al.*, 2007; Hore *et al.*, 2007; Shevchenko *et al.*, 2007)

<sup>13</sup> *HOTAIR* existe em mamíferos, no entanto sua sequência é pouco conservada (Rinn *et al.*, 2007; He *et al.*, 2011). Um trecho de 239 nt no éxon 6 do *HOTAIR* é especialmente conservada em mamíferos (He *et al.*, 2011).

*NATs*, para os quais tem sido sugerido que sua função seja dependente do pareamento de bases do *cis-NAT* com o transcrito da fita oposta. Outra possibilidade é que somente curtos domínios dos lncRNAs interajam ativamente (e estão sobre pressão evolutiva) com seus alvos. Esta possibilidade é suportada pela observação de que muitos lncRNAs possuem curtos trechos de sequências que são altamente conservadas (Pang *et al.*, 2006). Embora alguns possam dar origem a produtos funcionais menores (por exemplo, miRNAs e snoRNAs), outros tais como o *Xist* são funcionais como moléculas maiores, e, em ambos os casos, não está claro qual a proporção do transcrito global está sob seleção e como os diferentes tipos de imposições funcionais afetam essa força de seleção. Além disso tudo, lncRNAs regulatórios podem não somente estar sob pressão de seleção negativa para exercerem funções em espécies relacionadas, mas também podem estar sujeitos à seleção positiva associada com a radiação fenotípica (Mattick e Gagen, 2001).

## **5.5 Regulação da expressão de lncRNAs intrônicos por metilação do DNA**

Um caminho para melhor compreendermos a relevância biológica dos lncRNAs intrônicos é o estudo dos mecanismos envolvidos em sua biossíntese e modulação em resposta a sinais intracelulares e ambientais. Neste trabalho, investigamos o papel da metilação do DNA, um mecanismo de regulação epigenética associada com o silenciamento da expressão gênica, no controle da expressão de lncRNAs intrônicos reprimidos em linhagens celulares tumorais humanas.

A droga 5-Aza-2'-deoxicitidina (5-AZA) é amplamente usada como um inibidor de metilação do DNA para induzir expressão gênica e diferenciação celular (Christman, 2002). Para exercer seus efeitos biológicos, a droga, que é um análogo de citosinas, deve ser incorporada ao DNA. O DNA contendo 5-AZA forma adutos covalentes com as DNA metiltransferases, depletando as células da atividade dessa enzima e causando uma desmetilação do DNA genômico

como consequência (Christman *et al.*, 1985). O tratamento com 5-AZA também causa várias outras mudanças nas células, incluindo ativação de genes silenciados (Jones *et al.*, 1982), descondensação da cromatina (Haaf e Schmid, 1989) e alteração no tempo de replicação do DNA (Jablonka *et al.*, 1985).

O tratamento das células das linhagens DU-145, MCF-7 e Mia PaCa 2 com o agente desmetilante do DNA 5-AZA alterou a expressão de uma pequena fração de genes codificadores de proteínas e lncRNAs intrônicos. Foram encontrados 392 (6% do total de expressos), 651 (10% do total de expressos) e 1.612 (26% do total) genes codificadores de proteínas regulados pela desmetilação do DNA provocada pelo tratamento com 5-AZA em DU-145, MCF-7 e Mia PaCa 2, respectivamente. Estas proporções de genes afetados pelo tratamento com a droga são dependentes da sensibilidade das células, da concentração de 5-AZA utilizada e do tempo de tratamento. Nossos resultados estão de acordo com o que está descrito na literatura em experimentos em larga escala de reexpressão de mRNAs induzida pelo tratamento com 5-AZA (Sato *et al.*, 2003; Lodygin *et al.*, 2005; Kim *et al.*, 2012). Entre os genes codificadores de proteínas reexpressos após o tratamento, apenas alguns já foram mostrados terem sua expressão regulada por metilação do DNA em amostras de tecidos tumorais humanos. Além disso, uma análise de enriquecimento em categorias funcionais mostrou que esses genes estão principalmente envolvidos com câncer. Assim, nosso estudo aponta para o fato de que há muitos genes cuja expressão é controlada por metilação do DNA em linhagens celulares derivadas de tumores humanos. O conjunto de genes reexpressos após tratamento com 5-AZA contém candidatos interessantes para serem testados como supressores de tumor silenciados em neoplasias em futuros estudos.

Entre os lncRNAs intrônicos a proporção de afetados foi menor do que a observada para mRNAs; apenas cerca de 2% a 11% dos lncRNAs expressos mostraram-se regulados nas

linhagens pelo tratamento com 5-AZA. Para cerca de 19% dos lncRNAs intrônicos antissenso e 17% dos lncRNAs intrônicos senso afetados pelo tratamento foi detectada pelo menos uma CGI localizada em até 5 kb de distância dos TSSs preditos. Embora as mudanças no padrão de metilação do DNA, que altera a expressão de genes codificadores de proteínas estejam classicamente ligadas a promotores ricos em CpGs<sup>14</sup> (Futscher *et al.*, 2002; Song *et al.*, 2005), estudos recentes têm mostrado que promotores com baixo conteúdo de CpGs também podem estar metilados e podem silenciar a expressão de genes de maneira tecido específica (Weber *et al.*, 2007). Além disso, muitas alterações no padrão de metilação do DNA importantes em câncer de cólon e na diferenciação celular não ocorrem em promotores, nem em CGIs, mas em trechos com baixa densidade de dinucleotídeos CpG (*CpG island shores*) localizados a até 2 kb de distância do promotor (Doi *et al.*, 2009; Irizarry *et al.*, 2009). Em 2007, Weber e colaboradores, observaram que a metilação do DNA que ocorre na linhagem somática, mas não ocorre na linhagem germinativa, está mais frequentemente associada a regiões promotoras com baixos/intermediários conteúdos de CpGs<sup>15</sup>, indicando que promotores com CGIs fracas são alvos preferenciais para a metilação *de novo* que ocorre durante o desenvolvimento e que o conteúdo de CpGs dos promotores é determinante para a dinâmica dessa metilação (Weber *et al.*, 2007). Estes promotores com baixo conteúdo de CpG seriam menos protegidos dos mecanismos da metilação *de novo* que ocorre nas células nos estágios do desenvolvimento e durante a tumorigênese. Desse modo, apesar de não termos encontrado CGIs localizadas em até 5 kb da extremidade 5' conhecida da maioria dos lncRNAs intrônicos regulados pela desmetilação do

---

<sup>14</sup> Promotores que contém uma CGI classicamente definida por Gardiner-Garden (Gardiner-Garden e Frommer, 1987), ou seja, regiões genômicas do DNA de pelo menos 200 pb com mais de 50% de GC e uma razão maior que 0,6 do número de dinucleotídeos CpGs observados em relação ao número esperado com base na frequência de Cs e Gs dentro do segmento.

<sup>15</sup> Promotores que não contém uma região de pelo menos 500 pb com uma razão do número de dinucleotídeos CpGs observados em relação ao número esperado com base na frequência de Cs e Gs dentro do segmento acima de 0,48.

DNA, não se descarta a possibilidade de que vários deles possam ser regulados por metilação do DNA, ainda mais considerando que para muitos deles a extremidade 5' conhecida não representa um verdadeiro TSS.

O paradigma de que a metilação do DNA controla aspectos da diferenciação celular necessita que ela atue sobre genes cujo padrão de expressão varie em diferentes tipos celulares. A observação de que após o tratamento com o agente desmetilante, conjuntos distintos de lncRNAs intrônicos apresentam alteração na expressão em cada uma das linhagens tumorais está de acordo com o que se observa na literatura para genes codificadores de proteínas expressos em tecidos tumorais (Costello *et al.*, 2000; Eckhardt *et al.*, 2006) e linhagens celulares (Lister *et al.*, 2009; Straussman *et al.*, 2009).

Entre os 555 lncRNAs intrônicos antissenso e 404 lncRNAs intrônicos senso potencialmente regulados pela desmetilação do DNA provocada pelo tratamento com 5-AZA, foram escolhidos três para validação por PCR em tempo real. Para a validação, privilegamos lncRNAs que tiveram sua expressão aumentada em pelo menos 50% em pelo menos duas das três linhagens estudadas. Para evitar a escolha de algum lncRNA intrônico que pudesse ser pré-mRNA ou produto de *splicing*, optamos por transcritos de orientação antissenso à transcrição do mRNA do gene hospedeiro.

O lncRNA intrônico AS-SMG6, localizado totalmente dentro do último íntron do gene SMG6 teve sua expressão antissenso regulada pelo tratamento com 5-AZA validada por transcrição reversa fita específica seguida de PCR. No entanto, um estudo posterior para se determinar o tamanho real do transcrito, realizado com a metodologia de RACE 5' e 3' identificou que este transcrito faz parte da unidade transcricional do gene *HIC1*, um gene supressor de tumor hipermetilado em câncer (Wales *et al.*, 1995).

A reexpressão fita específica dos lncRNAs intrônicos AS-CELF2 e AS-APP após o desmetilação do DNA foi confirmada por PCR em tempo real. O tratamento de células da linhagem DU-145 com o agente desmetilante 5-AZA aumentou a expressão desses transcritos em 70 e 50 vezes, respectivamente. Além disso, a metilação do DNA de uma CGI localizada em até 5 kb das extremidades 5' conhecidas desses transcritos foi confirmada por digestão do DNA metilado com McrBC e sequenciamento do DNA tratado com bissulfato de sódio.

Foi identificada informaticamente uma possível região promotora localizada a aproximadamente 5 kb de distância do início de transcrição predito pela montagem de *contigs* do transcrito lncRNA AS-APP realizada por Nakaya e colaboradores (Nakaya *et al.*, 2007) que coincide com a localização da CGI metilada. A atividade promotora dessa predição foi confirmada por ensaio de gene repórter. Várias tentativas de se estender a extremidade 5' do lncRNA AS-APP a fim de colocalizar seu início de transcrição com o promotor identificado foram realizadas. Com a metodologia de RACE 5' conseguimos estender o transcrito em 867 nt. Novas tentativas de RACE 5', baseando-se no transcrito estendido foram realizadas sem sucesso. Tentamos amplificar por RT-PCR a região localizada entre a predição da região promotora e o início do transcrito, no entanto, também não obtivemos sucesso. Uma abordagem definitiva para a determinação do tamanho real do transcrito poderia ser realizada utilizando a técnica de *Northern Blot*, no entanto, até o momento esta abordagem ainda não foi utilizada. Assim, ainda fica em aberto se a região promotora identificada com atividade promotora é responsável por promover a transcrição do lncRNA AS-APP.

A desmetilação do DNA causada pelo tratamento com a droga 5-AZA ativou a expressão do lncRNA AS-APP na linhagem tumoral de próstata DU-145. Esta ativação foi também observada na linhagem embrionária de rim HEK293, mas não na linhagem tumoral de próstata, responsiva a andrógeno (LNCaP). Interessantemente, ao se medir os níveis de mRNA do gene

codificador de proteínas *APP* nas linhagens DU-145 e HEK293, após tratamento com 5-AZA, observamos um padrão de expressão inverso ao obtido para o lncRNA antissenso. Na linhagem LNCaP, o mRNA do gene *APP*, que está expresso em níveis relativamente mais baixos que o observado nas outras linhagens, não variou significativamente com tratamento realizado. Estes resultados apontam para uma possível regulação em *cis* do gene *APP* pelo lncRNA antissenso expresso a partir de seu *locus*, nas linhagens DU-145 e HEK293. Outros experimentos para confirmar ou não essa hipótese e talvez desvendar o mecanismo de ação devem ser realizados. Primeiramente, *Western Blot* para a quantificação da proteína APP em ambas as linhagens antes e após o tratamento com o agente desmetilante deve ser realizado a fim de se confirmar no nível proteico a diminuição da expressão do gene após o tratamento com a droga. Além disso, para confirmar a hipótese de ação em *cis* do lncRNA AS-APP, esse resultado precisa ainda ser reproduzido em um modelo mais controlado, no qual o efeito secundário da desmetilação global que ocorre nestas células quando tratadas com a droga 5-AZA possa ser eliminado, o que possibilitaria a observação direta do efeito do lncRNA sobre a expressão do RNA mensageiro e da proteína APP. Uma abordagem adequada seria medir os níveis do RNA mensageiro e da proteína do gene *APP* em células DU-145 e HEK293 superexpressando o lncRNA AS-APP.

O gene *APP* (*amyloid precursor protein*) codifica uma proteína que tem papel importante na doença de Alzheimer (Seubert *et al.*, 1993). A proteína APP tem localização transmembrana, e possui um grande segmento extracelular que é constitutivamente clivado por uma série de proteases, coletivamente chamadas de  $\alpha$ -secretases (Lammich *et al.*, 1999). Um processamento alternativo amiloidogênico realizado por uma  $\beta$ -secretase nos neurônios leva a produção do  $\alpha\beta$ -peptídeo, que constitui o principal componente das placas amilóides no córtex cerebral de pacientes com a doença de Alzheimer (Annaert e De Strooper, 2002). A expressão dessa proteína em altos níveis também já foi correlacionada com o baixo prognóstico de câncer de próstata

(Takayama *et al.*, 2009) e com melanomas em estado avançado (Botelho *et al.*, 2010). Além disso, foi demonstrado que a expressão do gene é induzida por andrógeno em células LNCaP, e que a superexpressão desse gene nestas células, promove a proliferação celular. Em contrapartida, o *knockdown* desse gene por RNAi diminui esse fenótipo *in vitro* e *in vivo* (Takayama *et al.*, 2009). De forma semelhante, em células de melanoma metastáticas, o bloqueio da expressão do gene *APP* por RNAi reduz a proliferação celular, e leva essas células a se diferenciarem irreversivelmente, tornando-se mais suscetíveis a diversos agentes quimioterápicos (Botelho *et al.*, 2010).

Já foi demonstrado que o gene *APP* sofre regulação de sua expressão por metilação do DNA. A alta expressão desse gene em cérebro de pacientes com doença de Alzheimer está associada com a ausência de metilação do DNA no promotor desse gene. Em pacientes saudáveis, esta mesma região está metilada (West *et al.*, 1995).

A presença de metilação do DNA em regiões intragênicas em muitos casos está associada com ativação gênica (Hellman e Chess, 2007; Cokus *et al.*, 2008; Lister *et al.*, 2009; Feng *et al.*, 2010). Embora os mecanismos dessa regulação ainda não estejam bem claros, alguns autores têm proposto modelos para essa ação. Uma possibilidade é a de que regiões metiladas intragênicas possam demarcar promotores de transcritos antissenso (Straussman *et al.*, 2009). De fato, há exemplos de lncRNAs antissenso que agem em *cis* e/ou em *trans*, direcionando a formação de heterocromatina e metilação do DNA, como é o caso do lncRNA *CDKN2B-AS1* (Yu *et al.*, 2008), já discutido anteriormente e dos lncRNAs regulados por *imprinting* genômico, como o *Xist* (Heard, 2004), *Kcnq1ot1* (Redrup *et al.*, 2009) e *Air* (Seidl *et al.*, 2006). Todos esses lncRNAs são transcritos pela RNAP II a partir de um promotor sensível a metilação do DNA. Com base nesses dados da literatura podemos levantar a hipótese de que o lncRNA AS-APP atue em *cis*, recrutando complexos proteicos envolvidos no silenciamento epigenético do promotor do gene

*APP*. Para explorar essa hipótese, o *status* de metilação do DNA e a presença de marcas repressivas na região promotora do gene *APP*, antes e durante a superexpressão do lncRNA AS-APP deverão ser investigados.

Outra possibilidade para a ação do lncRNA AS-APP seria por um mecanismo pós-transcricional, via miRNA. Em uma análise preliminar utilizando o programa miRNAFinder identificamos algumas regiões dentro da sequência do lncRNA AS-APP que podem ser precursoras de pré-miRNAs. Algumas das predições de pré-miRNAs dão origem a miRNAs maduros que tem alta complementaridade, principalmente, em sua extremidade 5'<sup>16</sup> à região 3'UTR do mRNA do gene *APP*. O mecanismo de regulação pós-transcricional por miRNAs, via repressão da tradução já foi demonstrado para o gene *APP* (Patel *et al.*, 2008; Hebert *et al.*, 2009). No entanto, isto não exclui a possibilidade de que outros miRNAs, como por exemplo, algum originado a partir do lncRNA AS-APP, também possam alvejar o 3'UTR do mRNA desse gene, impedindo a tradução da proteína APP. Para verificar esta hipótese outros experimentos devem ser realizados. Inicialmente, experimentos de co-transfecção das predições de miRNA originados a partir do lncRNA AS-APP com o 3'UTR do gene *APP* clonado na região 3' de um gene repórter devem ser realizados para se verificar o potencial dessa regulação.

---

<sup>16</sup> A hibridização da extremidade 5' é um importante aspecto do mecanismo de regulação por miRNAs. Essa região é conhecida como “*seed sequence*” e corresponde às posições de 2 a 8 nt a partir da extremidade 5' na sequência do miRNA. Em geral, ela é perfeitamente pareada com o mRNA alvo, sendo importante para a definição do repertório de alvos de um miRNA.

## 6 CONCLUSÕES E CONSIDERAÇÕES FINAIS

- Milhares de lncRNAs intrônicos foram detectados em pelo menos uma das três linhagens tumorais humanas DU-145, MCF-7 e Mia PaCa 2. A detecção destes transcritos em linhagens celulares ou tecidos com origens histológicas distintas através de meta-análise com dados públicos de RNA-seq indica que esses transcritos são expressos e se acumulam nas células.
- A abundância de expressão dos lncRNAs intrônicos detectados é tecido específica, o que sugere que eles devam ter sua transcrição regulada nas células onde são expressos, não configurando ruídos transcricionais originados do reconhecimento inespecífico da RNAP II.
- Observamos o enriquecimento de elementos regulatórios na extremidade 5' de uma fração dos lncRNAs intrônicos analisados, configurando mais um indício de que estes transcritos constituem unidades transcricionais independentes, com regiões regulatórias próprias.
- O tratamento com o agente desmetilante do DNA 5-AZA, alterou a expressão de um subconjunto de lncRNAs intrônicos, o que reforça a noção de que a expressão destes transcritos esteja sujeita a regulações transcricionais nas células.
- O tratamento com o agente desmetilante 5-AZA, afetou a expressão de uma menor fração de lncRNAs em comparação com mRNAs codificadores de proteínas.
- Análises de correlação de expressão entre os lncRNAs intrônicos e mRNAs expressos a partir do mesmo *locus* e análises de “*guilt by association*” demonstraram que muitos desses lncRNAs podem estar agindo sobre vias regulatórias importantes para o controle do desenvolvimento de organismos e ciclo celular, apontando funções importantes para lncRNAs no surgimento e manutenção de cânceres.

- Análises de conservação de sequência nucleotídica e de conservação de expressão entre vertebrados, além de análises de formação de estrutura secundária estável mostraram que muitos dos lncRNAs intrônicos expressos nas linhagens tumorais DU-145, MCF-7 e Mia PaCa 2 são evolutivamente conservados. Esta observação aponta para um conjunto de lncRNAs candidatos a desempenharem funções biológicas que contribuem para um aumento do *fitness* adaptativo e sujeitos a seleção purificadora.
- Embora a densidade de CGIs na vizinhança de lncRNAs intrônicos seja inferior a encontrada para mRNAs, a análise de reexpressão gênica após tratamento com 5-AZA de células de linhagens tumorais humanas revelou um subconjunto de lncRNAs intrônicos que tem sua expressão regulada por metilação do DNA.
- O lncRNA AS-APP possui uma CGI metilada e um promotor ativo a cerca de 4 kb de distância do seu início de transcrição conhecido. AS-APP é reexpresso nas linhagens DU-145 e HEK293 após tratamento com 5-AZA. O aumento na expressão do lncRNA AS-APP se correlaciona com a diminuição significativa dos níveis de expressão do mRNA do gene *APP*, sugerindo uma possível ação em *cis* deste transcrito não codificador. Abordagens experimentais devem ser realizadas para confirmar a hipótese de ação em *cis* do lncRNA AS-APP sobre a expressão do gene codificador da proteína APP.

## 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.; Ferreira, 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.; Woodage, T.; 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.
- Akiyama, Y.; Watkins, N.; Suzuki, H.; Jair, K. W.; van Engeland, M.; Esteller, M.; Sakai, H.; Ren, C. Y.; Yuasa, Y.; Herman, J. G.; Baylin, S. B. (2003). "GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer." *Mol Cell Biol* 23(23): 8429-8439.
- Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. (1964). "Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis." *Proc Natl Acad Sci U S A* 51: 786-794.
- Amaral, P. P.; Clark, M. B.; Gascoigne, D. K.; Dinger, M. E.; Mattick, J. S. (2011). "lncRNADB: a reference database for long noncoding RNAs." *Nucleic Acids Res* 39(Database issue): D146-151.
- 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.
- Annaert, W. ; De Strooper, B. (2002). "A cell biological perspective on Alzheimer's disease." Annu Rev Cell Dev Biol 18: 25-51.
- Antequera, F. ; Bird, A. (1993). "Number of CpG islands and genes in human and mouse." Proc Natl Acad Sci U S A 90(24): 11995-11999.
- Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M. ; Sherlock, G. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet 25(1): 25-29.
- Atkinson, S. R.; Marguerat, S. ; Bahler, J. (2011). "Exploring long non-coding RNAs through sequencing." Semin Cell Dev Biol.
- Avner, P. ; Heard, E. (2001). "X-chromosome inactivation: counting, choice and initiation." Nat Rev Genet 2(1): 59-67.
- Avner, P. ; Heard, E. (2001). "X-chromosome inactivation: counting, choice and initiation." Nature reviews. Genetics 2(1): 59-67.
- Baneres, J. L.; Martin, A. ; Parello, J. (1997). "The N tails of histones H3 and H4 adopt a highly structured conformation in the nucleosome." J Mol Biol 273(3): 503-508.
- Baratti, M. O.; Moreira, Y. B.; Traina, F.; Costa, F. F.; Verjovski-Almeida, S. ; Olalla-Saad, S. T. (2010). "Identification of protein-coding and non-coding RNA expression profiles in CD34+ and in stromal cells in refractory anemia with ringed sideroblasts." BMC Med Genomics 3: 30.
- Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell 116(2): 281-297.
- Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell 136(2): 215-233.
- Baylin, S. B. ; Herman, J. G. (2000). "DNA hypermethylation in tumorigenesis: epigenetics joins genetics." Trends Genet 16(4): 168-174.
- Bejerano, G.; Pheasant, M.; Makunin, I.; Stephen, S.; Kent, W. J.; Mattick, J. S. ; Haussler, D. (2004). "Ultraconserved elements in the human genome." Science 304(5675): 1321-1325.
- Belinsky, S. A.; Nikula, K. J.; Palmisano, W. A.; Michels, R.; Saccomanno, G.; Gabrielson, E.; Baylin, S. B. ; Herman, J. G. (1998). "Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis." Proc Natl Acad Sci U S A 95(20): 11891-11896.
- 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.
- Bernstein, B. E.; Kamal, M.; Lindblad-Toh, K.; Bekiranov, S.; Bailey, D. K.; Huebert, D. J.; McMahon, S.; Karlsson, E. K.; Kulbokas, E. J., 3rd; Gingeras, T. R.; Schreiber, S. L. ; Lander, E. S. (2005). "Genomic maps and comparative analysis of histone modifications in human and mouse." Cell 120(2): 169-181.
- Bertone, P.; Stolc, V.; Royce, T. E.; Rozowsky, J. S.; Urban, A. E.; Zhu, X.; Rinn, J. L.; Tongprasit, W.; Samanta, M.; Weissman, S.; Gerstein, M. ; Snyder, M. (2004). "Global identification of human transcribed sequences with genome tiling arrays." Science 306(5705): 2242-2246.
- Bestor, T. H. (2000). "The DNA methyltransferases of mammals." Human molecular genetics 9(16): 2395-2402.

- Bird, A. (2002). "DNA methylation patterns and epigenetic memory." *Genes Dev* 16(1): 6-21.
- Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." *Nature* 321(6067): 209-213.
- 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.; Srinivasan, 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.; Bhide, 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.; Thakapallayil, 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.; Eyas, 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.
- Bock, C.; Walter, J.; Paulsen, M. ; Lengauer, T. (2007). "CpG island mapping by epigenome prediction." *PLoS Comput Biol* 3(6): e110.
- Bolstad, B. M.; Irizarry, R. A.; Astrand, M. ; Speed, T. P. (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." *Bioinformatics* 19(2): 185-193.
- Botelho, M. G.; Wang, X.; Arndt-Jovin, D. J.; Becker, D. ; Jovin, T. M. (2010). "Induction of terminal differentiation in melanoma cells on downregulation of beta-amyloid precursor protein." *J Invest Dermatol* 130(5): 1400-1410.
- Brenner, C.; Deplus, R.; Didelot, C.; Loriot, A.; Vire, E.; De Smet, C.; Gutierrez, A.; Danovi, D.; Bernard, D.; Boon, T.; Pelicci, P. G.; Amati, B.; Kouzarides, T.; de Launoit, Y.; Di Croce, L. ; Fuks, F. (2005). "Myc represses transcription through recruitment of DNA methyltransferase corepressor." *The EMBO journal* 24(2): 336-346.
- Breslauer, K. J.; Frank, R.; Blocker, H. ; Marky, L. A. (1986). "Predicting DNA duplex stability from the base sequence." *Proc Natl Acad Sci U S A* 83(11): 3746-3750.
- Brodsky, A. S.; Meyer, C. A.; Swinburne, I. A.; Hall, G.; Keenan, B. J.; Liu, X. S.; Fox, E. A. ; Silver, P. A. (2005). "Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells." *Genome Biol* 6(8): R64.
- Brueckner, B.; Stresemann, C.; Kuner, R.; Mund, C.; Musch, T.; Meister, M.; Sultmann, H. ; Lyko, F. (2007). "The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function." *Cancer Res* 67(4): 1419-1423.
- 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.
- Camblong, J.; Iglesias, N.; Fickentscher, C.; Dieppois, G. ; Stutz, F. (2007). "Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*." *Cell* 131(4): 706-717.
- Candeias, M. M.; Malbert-Colas, L.; Powell, D. J.; Daskalogianni, C.; Maslon, M. M.; Naski, N.; Bourougaa, K.; Calvo, F. ; Fahraeus, R. (2008). "P53 mRNA controls p53 activity by managing Mdm2 functions." *Nat Cell Biol* 10(9): 1098-1105.
- Carninci, P. ; Hayashizaki, Y. (1999). "High-efficiency full-length cDNA cloning." *Methods Enzymol* 303: 19-44.
- Carninci, P. ; Hayashizaki, Y. (1999). "High-efficiency full-length cDNA cloning." *Methods in enzymology* 303: 19-44.
- 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.; Kawasaki, 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.
- Carninci, P.; Sandelin, A.; Lenhard, B.; Katayama, S.; Shimokawa, K.; Ponjavic, J.; Semple, C. A.; Taylor, M. S.; Engstrom, P. G.; Frith, M. C.; Forrest, A. R.; Alkema, W. B.; Tan, S. L.; Plessy, C.; Kodzius, R.; Ravasi, T.; Kasukawa, T.; Fukuda, S.; Kanamori-Katayama, M.; Kitazume, Y.; Kawaji, H.; Kai, C.; Nakamura, M.; Konno, H.; Nakano, K.; Mottagui-Tabar, S.; Arner, P.; Chesi, A.; Gustincich, S.; Persichetti, F.; Suzuki, H.; Grimmond, S. M.; Wells, C. A.; Orlando, V.; Wahlestedt, C.; Liu, E. T.; Harbers, M.; Kawai, J.; Bajic, V. B.; Hume, D. A.; Hayashizaki, Y. (2006). "Genome-wide analysis of mammalian promoter architecture and evolution." *Nat Genet* 38(6): 626-635.
- Carthew, R. W.; Sontheimer, E. J. (2009). "Origins and Mechanisms of miRNAs and siRNAs." *Cell* 136(4): 642-655.
- Cavaille, J.; Buiting, K.; Kiefmann, M.; Lalande, M.; Brannan, C. I.; Horsthemke, B.; Bachellerie, J. P.; Brosius, J.; Huttenhofer, A. (2000). "Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization." *Proc Natl Acad Sci U S A* 97(26): 14311-14316.
- 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.
- Chadwick, B. P.; Valley, C. M.; Willard, H. F. (2001). "Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome." *Nucleic Acids Res* 29(13): 2699-2705.

- Challen, G. A.; Sun, D.; Jeong, M.; Luo, M.; Jelinek, J.; Berg, J. S.; Bock, C.; Vasanthakumar, A.; Gu, H.; Xi, Y.; Liang, S.; Lu, Y.; Darlington, G. J.; Meissner, A.; Issa, J. P.; Godley, L. A.; Li, W.; Goodell, M. A. (2012). "Dnmt3a is essential for hematopoietic stem cell differentiation." *Nat Genet* 44(1): 23-31.
- Chen, J.; Sun, M.; Hurst, L. D.; Carmichael, G. G.; Rowley, J. D. (2005). "Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts." *Trends Genet* 21(6): 326-329.
- Chen, J.; Sun, M.; Kent, W. J.; Huang, X.; Xie, H.; Wang, W.; Zhou, G.; Shi, R. Z.; Rowley, J. D. (2004). "Over 20% of human transcripts might form sense-antisense pairs." *Nucleic Acids Res* 32(16): 4812-4820.
- Chen, J.; Sun, M.; Lee, S.; Zhou, G.; Rowley, J. D.; Wang, S. M. (2002). "Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags." *Proc Natl Acad Sci U S A* 99(19): 12257-12262.
- Chen, L. L.; Carmichael, G. G. (2010). "Long noncoding RNAs in mammalian cells: what, where, and why?" *Wiley Interdiscip Rev RNA* 1(1): 2-21.
- 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.
- Chodavarapu, R. K.; Feng, S.; Bernatavichute, Y. V.; Chen, P. Y.; Stroud, H.; Yu, Y.; Hetzel, J. A.; Kuo, F.; Kim, J.; Cokus, S. J.; Casero, D.; Bernal, M.; Huijser, P.; Clark, A. T.; Kramer, U.; Merchant, S. S.; Zhang, X.; Jacobsen, S. E.; Pellegrini, M. (2010). "Relationship between nucleosome positioning and DNA methylation." *Nature* 466(7304): 388-392.
- Chodroff, R. A.; Goodstadt, L.; Sirey, T. M.; Oliver, P. L.; Davies, K. E.; Green, E. D.; Molnar, Z.; Ponting, C. P. (2010). "Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes." *Genome Biol* 11(7): R72.
- Chomczynski, P.; Sacchi, N. (1987). "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction." *Anal Biochem* 162(1): 156-159.
- Chooniedass-Kothari, S.; Hamedani, M. K.; Auge, C.; Wang, X.; Carascossa, S.; Yan, Y.; Cooper, C.; Vincett, D.; Myal, Y.; Jalaguier, S.; Cavailles, V.; Leygue, E. (2010). "The steroid receptor RNA activator protein is recruited to promoter regions and acts as a transcriptional repressor." *FEBS Lett* 584(11): 2218-2224.
- Christman, J. K. (2002). "5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy." *Oncogene* 21(35): 5483-5495.
- Christman, J. K.; Schneiderman, N.; Acs, G. (1985). "Formation of highly stable complexes between 5-azacytosine-substituted DNA and specific non-histone nuclear proteins. Implications for 5-azacytidine-mediated effects on DNA methylation and gene expression." *J Biol Chem* 260(7): 4059-4068.
- Clark, M. B.; Mattick, J. S. (2011). "Long noncoding RNAs in cell biology." *Semin Cell Dev Biol* 22(4): 366-376.
- Cokus, S. J.; Feng, S.; Zhang, X.; Chen, Z.; Merriman, B.; Haudenschild, C. D.; Pradhan, S.; Nelson, S. F.; Pellegrini, M.; Jacobsen, S. E. (2008). "Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning." *Nature* 452(7184): 215-219.

- Conley, A. B.; Miller, W. J. ; Jordan, I. K. (2008). "Human cis natural antisense transcripts initiated by transposable elements." Trends Genet 24(2): 53-56.
- Consortium, I. H. G. S. (1998). "Genome sequence of the nematode *C. elegans*: a platform for investigating biology." Science 282(5396): 2012-2018.
- Costa, F. F. (2005). "Non-coding RNAs: new players in eukaryotic biology." Gene 357(2): 83-94.
- Costello, J. F.; Fruhwald, M. C.; Smiraglia, D. J.; Rush, L. J.; Robertson, G. P.; Gao, X.; Wright, F. A.; Feramisco, J. D.; Peltomaki, P.; Lang, J. C.; Schuller, D. E.; Yu, L.; Bloomfield, C. D.; Caligiuri, M. A.; Yates, A.; Nishikawa, R.; Su Huang, H.; Petrelli, N. J.; Zhang, X.; O'Dorisio, M. S.; Held, W. A.; Cavenee, W. K. ; Plass, C. (2000). "Aberrant CpG-island methylation has non-random and tumour-type-specific patterns." Nat Genet 24(2): 132-138.
- Cougot, N.; van Dijk, E.; Babajko, S. ; Seraphin, B. (2004). "'Cap-tabolism'." Trends Biochem Sci 29(8): 436-444.
- Cui, H.; Cruz-Correa, M.; Giardiello, F. M.; Hutcheon, D. F.; Kafonek, D. R.; Brandenburg, S.; Wu, Y.; He, X.; Powe, N. R. ; Feinberg, A. P. (2003). "Loss of IGF2 imprinting: a potential marker of colorectal cancer risk." Science 299(5613): 1753-1755.
- Davidow, L. S.; Breen, M.; Duke, S. E.; Samollow, P. B.; McCarrey, J. R. ; Lee, J. T. (2007). "The search for a marsupial XIC reveals a break with vertebrate synteny." Chromosome Res 15(2): 137-146.
- De Backer, O.; Arden, K. C.; Boretti, M.; Vantomme, V.; De Smet, C.; Czekay, S.; Viars, C. S.; De Plaen, E.; Brasseur, F.; Chomez, P.; Van den Eynde, B.; Boon, T. ; van der Bruggen, P. (1999). "Characterization of the GAGE genes that are expressed in various human cancers and in normal testis." Cancer Res 59(13): 3157-3165.
- Di Croce, L.; Raker, V. A.; Corsaro, M.; Fazi, F.; Fanelli, M.; Faretta, M.; Fuks, F.; Lo Coco, F.; Kouzarides, T.; Nervi, C.; Minucci, S. ; Pelicci, P. G. (2002). "Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor." Science 295(5557): 1079-1082.
- 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.
- Dobrovic, A. ; Simpfendorfer, D. (1997). "Methylation of the BRCA1 gene in sporadic breast cancer." Cancer Res 57(16): 3347-3350.
- Doi, A.; Park, I. H.; Wen, B.; Murakami, P.; Aryee, M. J.; Irizarry, R.; Herb, B.; Ladd-Acosta, C.; Rho, J.; Loewer, S.; Miller, J.; Schlaeger, T.; Daley, G. Q. ; Feinberg, A. P. (2009). "Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts." Nat Genet 41(12): 1350-1353.
- Duret, L.; Chureau, C.; Samain, S.; Weissenbach, J. ; Avner, P. (2006). "The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene." Science 312(5780): 1653-1655.
- Eckhardt, F.; Lewin, J.; Cortese, R.; Rakyan, V. K.; Attwood, J.; Burger, M.; Burton, J.; Cox, T. V.; Davies, R.; Down, T. A.; Haefliger, C.; Horton, R.; Howe, K.; Jackson, D. K.; Kunde, J.; Koenig, C.; Liddle, J.; Niblett, D.; Otto, T.; Pettett, R.; Seemann, S.; Thompson, C.; West, T.; Rogers, J.; Olek, A.; Berlin, K. ; Beck, S. (2006). "DNA methylation profiling of human chromosomes 6, 20 and 22." Nat Genet 38(12): 1378-1385.
- Eden, A.; Gaudet, F.; Waghmare, A. ; Jaenisch, R. (2003). "Chromosomal instability and tumors promoted by DNA hypomethylation." Science 300(5618): 455.

- Ehrlich, M.; Gama-Sosa, M. A.; Huang, L. H.; Midgett, R. M.; Kuo, K. C.; McCune, R. A. ; Gehrke, C. (1982). "Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells." Nucleic Acids Res 10(8): 2709-2721.
- Engstrom, P. G.; Suzuki, H.; Ninomiya, N.; Akalin, A.; Sessa, L.; Lavorgna, G.; Brozzi, A.; Luzi, L.; Tan, S. L.; Yang, L.; Kunarso, G.; Ng, E. L.; Batalov, S.; Wahlestedt, C.; Kai, C.; Kawai, J.; Carninci, P.; Hayashizaki, Y.; Wells, C.; Bajic, V. B.; Orlando, V.; Reid, J. F.; Lenhard, B. ; Lipovich, L. (2006). "Complex Loci in human and mouse genomes." PLoS Genet 2(4): e47.
- Erdmann, V. A.; Barciszewska, M. Z.; Hochberg, A.; de Groot, N. ; Barciszewski, J. (2001). "Regulatory RNAs." Cell Mol Life Sci 58(7): 960-977.
- Esquela-Kerscher, A. ; Slack, F. J. (2006). "Oncomirs - microRNAs with a role in cancer." Nat Rev Cancer 6(4): 259-269.
- Esteller, M. (2008). "Epigenetics in cancer." N Engl J Med 358(11): 1148-1159.
- Euskirchen, G.; Royce, T. E.; Bertone, P.; Martone, R.; Rinn, J. L.; Nelson, F. K.; Sayward, F.; Luscombe, N. M.; Miller, P.; Gerstein, M.; Weissman, S. ; Snyder, M. (2004). "CREB binds to multiple loci on human chromosome 22." Mol Cell Biol 24(9): 3804-3814.
- Farh, K. K.; Grimson, A.; Jan, C.; Lewis, B. P.; Johnston, W. K.; Lim, L. P.; Burge, C. B. ; Bartel, D. P. (2005). "The widespread impact of mammalian MicroRNAs on mRNA repression and evolution." Science 310(5755): 1817-1821.
- Feinberg, A. P. (1999). "Mendel stayed home. Genomic imprinting and environmental disease susceptibility, National Institute of Environmental Health Sciences and Duke University Medical Center, Durham, NC, USA, 8-10 October 1998." Trends Genet 15(2): 46.
- Feinberg, A. P. ; Tycko, B. (2004). "The history of cancer epigenetics." Nat Rev Cancer 4(2): 143-153.
- Feinberg, A. P. ; Vogelstein, B. (1983). "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." Nature 301(5895): 89-92.
- 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.
- Feng, J.; Bi, C.; Clark, B. S.; Mady, R.; Shah, P. ; Kohtz, J. D. (2006). "The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator." Genes Dev 20(11): 1470-1484.
- Feng, S.; Cokus, S. J.; Zhang, X.; Chen, P. Y.; Bostick, M.; Goll, M. G.; Hetzel, J.; Jain, J.; Strauss, S. H.; Halpern, M. E.; Ukomadu, C.; Sadler, K. C.; Pradhan, S.; Pellegrini, M. ; Jacobsen, S. E. (2010). "Conservation and divergence of methylation patterning in plants and animals." Proc Natl Acad Sci U S A 107(19): 8689-8694.
- Flynn, R. A.; Almada, A. E.; Zamudio, J. R. ; Sharp, P. A. (2011). "Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome." Proc Natl Acad Sci U S A 108(26): 10460-10465.
- Flynn, R. L.; Centore, R. C.; O'Sullivan, R. J.; Rai, R.; Tse, A.; Songyang, Z.; Chang, S.; Karlseder, J. ; Zou, L. (2011). "TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA." Nature 471(7339): 532-536.
- 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.

- Furuno, M.; Pang, K. C.; Ninomiya, N.; Fukuda, S.; Frith, M. C.; Bult, C.; Kai, C.; Kawai, J.; Carninci, P.; Hayashizaki, Y.; Mattick, J. S.; Suzuki, H. (2006). "Clusters of internally primed transcripts reveal novel long noncoding RNAs." PLoS Genet 2(4): e37.
- Futscher, B. W.; Oshiro, M. M.; Wozniak, R. J.; Holtan, N.; Hanigan, C. L.; Duan, H.; Domann, F. E. (2002). "Role for DNA methylation in the control of cell type specific maspin expression." Nat Genet 31(2): 175-179.
- Gal-Yam, E. N.; Egger, G.; Iniguez, L.; Holster, H.; Einarsson, S.; Zhang, X.; Lin, J. C.; Liang, G.; Jones, P. A.; Tanay, A. (2008). "Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line." Proc Natl Acad Sci U S A 105(35): 12979-12984.
- Gal-Yam, E. N.; Jeong, S.; Tanay, A.; Egger, G.; Lee, A. S.; Jones, P. A. (2006). "Constitutive nucleosome depletion and ordered factor assembly at the GRP78 promoter revealed by single molecule footprinting." PLoS Genet 2(9): e160.
- Gardiner-Garden, M.; Frommer, M. (1987). "CpG islands in vertebrate genomes." J Mol Biol 196(2): 261-282.
- Ge, Y. Z.; Pu, M. T.; Gowher, H.; Wu, H. P.; Ding, J. P.; Jeltsch, A.; Xu, G. L. (2004). "Chromatin targeting of de novo DNA methyltransferases by the PWWP domain." J Biol Chem 279(24): 25447-25454.
- Gibb, E. A.; Brown, C. J.; Lam, W. L. (2011). "The functional role of long non-coding RNA in human carcinomas." Mol Cancer 10: 38.
- 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.
- Glusman, G.; Qin, S.; El-Gewely, M. R.; Siegel, A. F.; Roach, J. C.; Hood, L.; Smit, A. F. (2006). "A third approach to gene prediction suggests thousands of additional human transcribed regions." PLoS Comput Biol 2(3): e18.
- Grivna, S. T.; Beyret, E.; Wang, Z.; Lin, H. (2006). "A novel class of small RNAs in mouse spermatogenic cells." Genes Dev 20(13): 1709-1714.
- Gruber, A. R.; Neubock, R.; Hofacker, I. L.; Washietl, S. (2007). "The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures." Nucleic Acids Res 35(Web Server issue): W335-338.
- Guenther, M. G.; Levine, S. S.; Boyer, L. A.; Jaenisch, R.; Young, R. A. (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell 130(1): 77-88.
- 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.
- Gustincich, S.; Sandelin, A.; Plessy, C.; Katayama, S.; Simone, R.; Lazarevic, D.; Hayashizaki, Y.; Carninci, P. (2006). "The complexity of the mammalian transcriptome." J Physiol 575(Pt 2): 321-332.
- 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.; 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.
- Haaf, T. ; Schmid, M. (1989). "5-Azadeoxycytidine induced undercondensation in the giant X chromosomes of *Microtus agrestis*." Chromosoma 98(2): 93-98.
- Hall, I. M.; Shankaranarayana, G. D.; Noma, K.; Ayoub, N.; Cohen, A. ; Grewal, S. I. (2002). "Establishment and maintenance of a heterochromatin domain." Science 297(5590): 2232-2237.
- Hashimoto, K.; Ishida, E.; Matsumoto, S.; Shibusawa, N.; Okada, S.; Monden, T.; Satoh, T.; Yamada, M. ; Mori, M. (2009). "A liver X receptor (LXR)-beta alternative splicing variant (LXRBSV) acts as an RNA co-activator of LXR-beta." Biochem Biophys Res Commun 390(4): 1260-1265.
- He, S.; Liu, S. ; Zhu, H. (2011). "The sequence, structure and evolutionary features of HOTAIR in mammals." BMC Evol Biol 11: 102.
- Heard, E. (2004). "Recent advances in X-chromosome inactivation." Curr Opin Cell Biol 16(3): 247-255.
- Heard, E. (2005). "Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome." Curr Opin Genet Dev 15(5): 482-489.
- Hebert, S. S.; Horre, K.; Nicolai, L.; Bergmans, B.; Papadopoulou, A. S.; Delacourte, A. ; De Strooper, B. (2009). "MicroRNA regulation of Alzheimer's Amyloid precursor protein expression." Neurobiol Dis 33(3): 422-428.
- Hellman, A. ; Chess, A. (2007). "Gene body-specific methylation on the active X chromosome." Science 315(5815): 1141-1143.
- Hellwig, S. ; Bass, B. L. (2008). "A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes." Proc Natl Acad Sci U S A 105(35): 12897-12902.
- Henikoff, S. (2005). "Histone modifications: combinatorial complexity or cumulative simplicity?" Proc Natl Acad Sci U S A 102(15): 5308-5309.
- Henikoff, S. ; Ahmad, K. (2005). "Assembly of variant histones into chromatin." Annu Rev Cell Dev Biol 21: 133-153.
- Heo, J. B. ; Sung, S. (2011). "Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA." Science 331(6013): 76-79.
- Hermann, A.; Schmitt, S. ; Jeltsch, A. (2003). "The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity." J Biol Chem 278(34): 31717-31721.
- Hill, A. E.; Hong, J. S.; Wen, H.; Teng, L.; McPherson, D. T.; McPherson, S. A.; Levasseur, D. N. ; Sorscher, E. J. (2006). "Micro-RNA-like effects of complete intronic sequences." Front Biosci 11: 1998-2006.
- Hirota, K.; Miyoshi, T.; Kugou, K.; Hoffman, C. S.; Shibata, T. ; Ohta, K. (2008). "Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs." Nature 456(7218): 130-134.
- Hofacker, I. L.; Fekete, M. ; Stadler, P. F. (2002). "Secondary structure prediction for aligned RNA sequences." J Mol Biol 319(5): 1059-1066.
- Holliday, R. ; Pugh, J. E. (1975). "DNA modification mechanisms and gene activity during development." Science 187(4173): 226-232.
- Hong, L.; Schroth, G. P.; Matthews, H. R.; Yau, P. ; Bradbury, E. M. (1993). "Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA." J Biol Chem 268(1): 305-314.

- Hore, T. A.; Koina, E.; Wakefield, M. J. ; Marshall Graves, J. A. (2007). "The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals." Chromosome Res 15(2): 147-161.
- Hosack, D. A.; Dennis, G., Jr.; Sherman, B. T.; Lane, H. C. ; Lempicki, R. A. (2003). "Identifying biological themes within lists of genes with EASE." Genome Biol 4(10): R70.
- Howard, G.; Eiges, R.; Gaudet, F.; Jaenisch, R. ; Eden, A. (2008). "Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice." Oncogene 27(3): 404-408.
- Huang da, W.; Sherman, B. T. ; Lempicki, R. A. (2009). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." Nucleic Acids Res 37(1): 1-13.
- Huang da, W.; Sherman, B. T. ; Lempicki, R. A. (2009a). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc 4(1): 44-57.
- Huang da, W.; Sherman, B. T. ; Lempicki, R. A. (2009b). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." Nucleic Acids Res 37(1): 1-13.
- 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.
- Hughes, T. R.; Mao, M.; Jones, A. R.; Burchard, J.; Marton, M. J.; Shannon, K. W.; Lefkowitz, S. M.; Ziman, M.; Schelter, J. M.; Meyer, M. R.; Kobayashi, S.; Davis, C.; Dai, H.; He, Y. D.; Stephaniants, S. B.; Cavet, G.; Walker, W. L.; West, A.; Coffey, E.; Shoemaker, D. D.; Stoughton, R.; Blanchard, A. P.; Friend, S. H. ; Linsley, P. S. (2001). "Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer." Nat Biotechnol 19(4): 342-347.
- Hung, T.; Wang, Y.; Lin, M. F.; Koegel, A. K.; Kotake, Y.; Grant, G. D.; Horlings, H. M.; Shah, N.; Umbricht, C.; Wang, P.; Kong, B.; Langerod, A.; Borresen-Dale, A. L.; Kim, S. K.; van de Vijver, M.; Sukumar, S.; Whitfield, M. L.; Kellis, M.; Xiong, Y.; Wong, D. J. ; Chang, H. Y. (2011). "Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters." Nat Genet 43(7): 621-629.
- Huttenhofer, A.; Kiefmann, M.; Meier-Ewert, S.; O'Brien, J.; Lehrach, H.; Bachellerie, J. P. ; Brosius, J. (2001). "RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse." Embo J 20(11): 2943-2953.
- Illingworth, R. S. ; Bird, A. P. (2009). "CpG islands--'a rough guide'." FEBS Lett 583(11): 1713-1720.
- Illingworth, R. S.; Gruenewald-Schneider, U.; Webb, S.; Kerr, A. R.; James, K. D.; Turner, D. J.; Smith, C.; Harrison, D. J.; Andrews, R. ; Bird, A. P. (2010). "Orphan CpG islands identify numerous conserved promoters in the mammalian genome." PLoS Genet 6(9).
- Irizarry, R. A.; Ladd-Acosta, C.; Carvalho, B.; Wu, H.; Brandenburg, S. A.; Jeddelloh, J. A.; Wen, B. ; Feinberg, A. P. (2008). "Comprehensive high-throughput arrays for relative methylation (CHARM)." Genome Res 18(5): 780-790.
- Irizarry, R. A.; Ladd-Acosta, C.; Wen, B.; Wu, Z.; Montano, C.; Onyango, P.; Cui, H.; Gabo, K.; Rongione, M.; Webster, M.; Ji, H.; Potash, J. B.; Sabunciyan, S. ; Feinberg, A. P. (2009). "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores." Nat Genet 41(2): 178-186.

- Jablonka, E.; Goitein, R.; Marcus, M. ; Cedar, H. (1985). "DNA hypomethylation causes an increase in DNase-I sensitivity and an advance in the time of replication of the entire inactive X chromosome." Chromosoma 93(2): 152-156.
- Jaenisch, R. ; Bird, A. (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nat Genet 33 Suppl: 245-254.
- Jenuwein, T. ; Allis, C. D. (2001). "Translating the histone code." Science 293(5532): 1074-1080.
- Jhaveri, M. S. ; Morrow, C. S. (1998). "Methylation-mediated regulation of the glutathione S-transferase P1 gene in human breast cancer cells." Gene 210(1): 1-7.
- Jishage, M.; Malik, S.; Wagner, U.; Uberheide, B.; Ishihama, Y.; Hu, X.; Chait, B. T.; Gnatt, A.; Ren, B. ; Roeder, R. G. (2012). "Transcriptional Regulation by Pol II(G) Involving Mediator and Competitive Interactions of Gdown1 and TFIIF with Pol II." Mol Cell 45(1): 51-63.
- Jones, P. A. (1999). "The DNA methylation paradox." Trends in genetics : TIG 15(1): 34-37.
- Jones, P. A.; Taylor, S. M.; Mohandas, T. ; Shapiro, L. J. (1982). "Cell cycle-specific reactivation of an inactive X-chromosome locus by 5-azadeoxycytidine." Proc Natl Acad Sci U S A 79(4): 1215-1219.
- Jongeneel, C. V.; Delorenzi, M.; Iseli, C.; Zhou, D.; Haudenschild, C. D.; Khrebtukova, I.; Kuznetsov, D.; Stevenson, B. J.; Strausberg, R. L.; Simpson, A. J. ; Vasicek, T. J. (2005). "An atlas of human gene expression from massively parallel signature sequencing (MPSS)." Genome Res 15(7): 1007-1014.
- Kamakaka, R. T. ; Biggins, S. (2005). "Histone variants: deviants?" Genes Dev 19(3): 295-310.
- Kampa, D.; Cheng, J.; Kapranov, P.; Yamanaka, M.; Brubaker, S.; Cawley, S.; Drenkow, J.; Piccolboni, A.; Bekiranov, S.; Helt, G.; Tammana, H. ; Gingeras, T. R. (2004). "Novel RNAs Identified From an In-Depth Analysis of the Transcriptome of Human Chromosomes 21 and 22." Genome Res. 14(3): 331-342.
- Kampa, D.; Cheng, J.; Kapranov, P.; Yamanaka, M.; Brubaker, S.; Cawley, S.; Drenkow, J.; Piccolboni, A.; Bekiranov, S.; Helt, G.; Tammana, H. ; Gingeras, T. R. (2004). "Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22." Genome Res 14(3): 331-342.
- Kapranov, P. (2009). "From transcription start site to cell biology." Genome Biol 10(4): 217.
- Kapranov, P.; Cawley, S. E.; Drenkow, J.; Bekiranov, S.; Strausberg, R. L.; Fodor, S. P. ; Gingeras, T. R. (2002). "Large-scale transcriptional activity in chromosomes 21 and 22." Science 296(5569): 916-919.
- Kapranov, P.; Cheng, J.; Dike, S.; Nix, D. A.; Dutttagupta, R.; Willingham, A. T.; Stadler, P. F.; Hertel, J.; Hackermuller, J.; Hofacker, I. L.; Bell, I.; Cheung, E.; Drenkow, J.; Dumais, E.; Patel, S.; Helt, G.; Ganesh, M.; Ghosh, S.; Piccolboni, A.; Sementchenko, V.; Tammana, H. ; Gingeras, T. R. (2007). "RNA maps reveal new RNA classes and a possible function for pervasive transcription." Science 316(5830): 1484-1488.
- Kapranov, P.; Drenkow, J.; Cheng, J.; Long, J.; Helt, G.; Dike, S. ; Gingeras, T. R. (2005). "Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays." Genome Res 15(7): 987-997.
- Kapranov, P.; Willingham, A. T. ; Gingeras, T. R. (2007). "Genome-wide transcription and the implications for genomic organization." Nat Rev Genet 8(6): 413-423.
- Karpf, A. R. ; Matsui, S. (2005). "Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells." Cancer Res 65(19): 8635-8639.

- 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.
- Katayama, S.; Tomaru, Y.; Kasukawa, T.; Waki, K.; Nakanishi, M.; Nakamura, M.; Nishida, H.; Yap, C. C.; Suzuki, M.; Kawai, J.; Suzuki, H.; Carninci, P.; Hayashizaki, Y.; Wells, C.; Frith, M.; Ravasi, T.; Pang, K. C.; Hallinan, J.; Mattick, J.; Hume, D. A.; Lipovich, L.; Batalov, S.; Engstrom, P. G.; Mizuno, Y.; Faghihi, M. A.; Sandelin, A.; Chalk, A. M.; Mottagui-Tabar, S.; Liang, Z.; Lenhard, B. ; Wahlestedt, C. (2005). "Antisense transcription in the mammalian transcriptome." Science 309(5740): 1564-1566.
- Kawasaki, H. ; Taira, K. (2004). "Induction of DNA methylation and gene silencing by short interfering RNAs in human cells." Nature 431(7005): 211-217.
- Keshet, I.; Schlesinger, Y.; Farkash, S.; Rand, E.; Hecht, M.; Segal, E.; Pikarski, E.; Young, R. A.; Niveleau, A.; Cedar, H. ; Simon, I. (2006). "Evidence for an instructive mechanism of de novo methylation in cancer cells." Nat Genet 38(2): 149-153.
- 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, J. H.; Kang, S.; Kim, T. W.; Yin, L.; Liu, R. ; Kim, S. J. (2012). "Expression profiling after induction of demethylation in MCF-7 breast cancer cells identifies involvement of TNF-alpha mediated cancer pathways." Mol Cells 33(2): 127-133.
- Kim, K.; Jutooru, I.; Chadalapaka, G.; Johnson, G.; Frank, J.; Burghardt, R.; Kim, S. ; Safe, S. (2012). "HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer." Oncogene.
- Kim, T. H.; Barrera, L. O.; Qu, C.; Van Calcar, S.; Trinklein, N. D.; Cooper, S. J.; Luna, R. M.; Glass, C. K.; Rosenfeld, M. G.; Myers, R. M. ; Ren, B. (2005). "Direct isolation and identification of promoters in the human genome." Genome Res 15(6): 830-839.
- Kim, T. H.; Barrera, L. O.; Zheng, M.; Qu, C.; Singer, M. A.; Richmond, T. A.; Wu, Y.; Green, R. D. ; Ren, B. (2005). "A high-resolution map of active promoters in the human genome." Nature 436(7052): 876-880.
- Kim, T. K.; Hemberg, M.; Gray, J. M.; Costa, A. M.; Bear, D. M.; Wu, J.; Harmin, D. A.; Laptewicz, M.; Barbara-Haley, K.; Kuersten, S.; Markenscoff-Papadimitriou, E.; Kuhl, D.; Bito, H.; Worley, P. F.; Kreiman, G. ; Greenberg, M. E. (2010). "Widespread transcription at neuronal activity-regulated enhancers." Nature 465(7295): 182-187.
- Kiss, T. (2002). "Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions." Cell 109(2): 145-148.
- Kiyono, T.; Foster, S. A.; Koop, J. I.; McDougall, J. K.; Galloway, D. A. ; Klingelutz, A. J. (1998). "Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells." Nature 396(6706): 84-88.
- Kodzius, R.; Kojima, M.; Nishiyori, H.; Nakamura, M.; Fukuda, S.; Tagami, M.; Sasaki, D.; Imamura, K.; Kai, C.; Harbers, M.; Hayashizaki, Y. ; Carninci, P. (2006). "CAGE: cap analysis of gene expression." Nat Methods 3(3): 211-222.
- Kong, L.; Zhang, Y.; Ye, Z. Q.; Liu, X. Q.; Zhao, S. Q.; Wei, L. ; Gao, G. (2007). "CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine." Nucleic Acids Res 35(Web Server issue): W345-349.

- Kornberg, R. D. (1974). "Chromatin structure: a repeating unit of histones and DNA." Science 184(4139): 868-871.
- 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.
- Krystal, G. W.; Armstrong, B. C. ; Battey, J. F. (1990). "N-myc mRNA forms an RNA-RNA duplex with endogenous antisense transcripts." Mol Cell Biol 10(8): 4180-4191.
- Kumaki, Y.; Oda, M. ; Okano, M. (2008). "QUMA: quantification tool for methylation analysis." Nucleic Acids Res 36(Web Server issue): W170-175.
- Lagos-Quintana, M.; Rauhut, R.; Lendeckel, W. ; Tuschl, T. (2001). "Identification of novel genes coding for small expressed RNAs." Science 294(5543): 853-858.
- Lagos-Quintana, M.; Rauhut, R.; Meyer, J.; Borkhardt, A. ; Tuschl, T. (2003). "New microRNAs from mouse and human." Rna 9(2): 175-179.
- Lammich, S.; Kojro, E.; Postina, R.; Gilbert, S.; Pfeiffer, R.; Jasionowski, M.; Haass, C. ; Fahrenholz, F. (1999). "Constitutive and regulated alpha-secretase cleavage of Alzheimer's amyloid precursor protein by a disintegrin metalloprotease." Proc Natl Acad Sci U S A 96(7): 3922-3927.
- Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D. R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Myers, R. M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Schultz,

- R.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W. R.; de la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D. G.; Burge, C. B.; Cerutti, L.; Chen, H. C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; Furey, T. S.; Galagan, J.; Gilbert, J. G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L. S.; Jones, T. A.; Kasif, S.; Kasprzyk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A. F.; Stupka, E.; Szustakowski, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Yang, S. P.; Yeh, R. F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Patrinos, A.; Morgan, M. J.; Szustakowski, J.; de Jong, P.; Catanese, J. J.; Osoegawa, K.; Shizuya, H.; Choi, S.; Chen, Y. J. (2001). "Initial sequencing and analysis of the human genome." Nature 409(6822): 860-921.
- Lapidot, M.; Pilpel, Y. (2006). "Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms." EMBO Rep 7(12): 1216-1222.
- Lau, N. C.; Lim, L. P.; Weinstein, E. G.; Bartel, D. P. (2001). "An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*." Science 294(5543): 858-862.
- Laurent, L.; Wong, E.; Li, G.; Huynh, T.; Tsiganos, A.; Ong, C. T.; Low, H. M.; Kin Sung, K. W.; Rigoutsos, I.; Loring, J.; Wei, C. L. (2010). "Dynamic changes in the human methylome during differentiation." Genome Res 20(3): 320-331.
- Lee, J. T. (2009). "Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome." Genes Dev 23(16): 1831-1842.
- Lee, R. C.; Ambros, V. (2001). "An extensive class of small RNAs in *Caenorhabditis elegans*." Science 294(5543): 862-864.
- Lewis, B. P.; Burge, C. B.; Bartel, D. P. (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." Cell 120(1): 15-20.
- Li, E.; Bestor, T. H.; Jaenisch, R. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell 69(6): 915-926.
- Lim, L. P.; Lau, N. C.; Garrett-Engele, P.; Grimson, A.; Schelter, J. M.; Castle, J.; Bartel, D. P.; Linsley, P. S.; Johnson, J. M. (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs." Nature 433(7027): 769-773.
- Lippman, Z.; Gendrel, A. V.; Black, M.; Vaughn, M. W.; Dedhia, N.; McCombie, W. R.; Lavine, K.; Mittal, V.; May, B.; Kasschau, K. D.; Carrington, J. C.; Doerge, R. W.; Colot, V.; Martienssen, R. (2004). "Role of transposable elements in heterochromatin and epigenetic control." Nature 430(6998): 471-476.
- Lister, R.; Pelizzola, M.; Dowen, R. H.; Hawkins, R. D.; Hon, G.; Tonti-Filippini, J.; Nery, J. R.; Lee, L.; Ye, Z.; Ngo, Q. M.; Edsall, L.; Antosiewicz-Bourget, J.; Stewart, R.; Ruotti, V.; Millar, A. H.; Thomson, J. A.; Ren, B.; Ecker, J. R. (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature 462(7271): 315-322.
- Litt, M. D.; Simpson, M.; Gaszner, M.; Allis, C. D.; Felsenfeld, G. (2001). "Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus." Science 293(5539): 2453-2455.

- Liu, C.; Bai, B.; Skogerbo, G.; Cai, L.; Deng, W.; Zhang, Y.; Bu, D.; Zhao, Y.; Chen, R. (2005). "NONCODE: an integrated knowledge database of non-coding RNAs." Nucleic Acids Res 33(Database issue): D112-115.
- 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.
- Lock, L. F.; Takagi, N.; Martin, G. R. (1987). "Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation." Cell 48(1): 39-46.
- Lodygin, D.; Epanchintsev, A.; Menssen, A.; Diebold, J.; Hermeking, H. (2005). "Functional epigenomics identifies genes frequently silenced in prostate cancer." Cancer Res 65(10): 4218-4227.
- Loewer, S.; Cabili, M. N.; Guttman, M.; Loh, Y. H.; Thomas, K.; Park, I. H.; Garber, M.; Curran, M.; Onder, T.; Agarwal, S.; Manos, P. D.; Datta, S.; Lander, E. S.; Schlaeger, T. M.; Daley, G. Q.; Rinn, J. L. (2010). "Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells." Nat Genet 42(12): 1113-1117.
- Long, C.; Yin, B.; Lu, Q.; Zhou, X.; Hu, J.; Yang, Y.; Yu, F.; Yuan, Y. (2007). "Promoter hypermethylation of the RUNX3 gene in esophageal squamous cell carcinoma." Cancer Invest 25(8): 685-690.
- Lopez-Serra, P.; Esteller, M. (2012). "DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer." Oncogene 31(13): 1609-1622.
- Lorincz, M. C.; Dickerson, D. R.; Schmitt, M.; Groudine, M. (2004). "Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells." Nat Struct Mol Biol 11(11): 1068-1075.
- 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 Biol 5: 4.
- Louro, R.; Smirnova, A. S.; Verjovski-Almeida, S. (2009). "Long intronic noncoding RNA transcription: expression noise or expression choice?" Genomics 93(4): 291-298.
- Mancini-Dinardo, D.; Steele, S. J.; Levorse, J. M.; Ingram, R. S.; Tilghman, S. M. (2006). "Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes." Genes Dev 20(10): 1268-1282.
- Martens, J. H.; O'Sullivan, R. J.; Braunschweig, U.; Opravil, S.; Radolf, M.; Steinlein, P.; Jenuwein, T. (2005). "The profile of repeat-associated histone lysine methylation states in the mouse epigenome." EMBO J 24(4): 800-812.
- Martianov, I.; Ramadass, A.; Serra Barros, A.; Chow, N.; Akoulitchev, A. (2007). "Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript." Nature 445(7128): 666-670.
- Martone, R.; Euskirchen, G.; Bertone, P.; Hartman, S.; Royce, T. E.; Luscombe, N. M.; Rinn, J. L.; Nelson, F. K.; Miller, P.; Gerstein, M.; Weissman, S.; Snyder, M. (2003). "Distribution of NF-kappaB-binding sites across human chromosome 22." Proc Natl Acad Sci U S A 100(21): 12247-12252.
- Mattick, J. S. (1994). "Introns: evolution and function." Curr Opin Genet Dev 4(6): 823-831.

- Mattick, J. S. (2001). "Non-coding RNAs: the architects of eukaryotic complexity." EMBO Rep 2(11): 986-991.
- Mattick, J. S. ; Gagen, M. J. (2001). "The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms." Mol Biol Evol 18(9): 1611-1630.
- Mattick, J. S. ; Makunin, I. V. (2005). "Small regulatory RNAs in mammals." Hum Mol Genet 14 Spec No 1: R121-132.
- Mattick, J. S. ; Makunin, I. V. (2006). "Non-coding RNA." Hum Mol Genet 15 Suppl 1: R17-29.
- Maunakea, A. K.; Nagarajan, R. P.; Bilenky, M.; Ballinger, T. J.; D'Souza, C.; Fouse, S. D.; Johnson, B. E.; Hong, C.; Nielsen, C.; Zhao, Y.; Turecki, G.; Delaney, A.; Varhol, R.; Thiessen, N.; Shchors, K.; Heine, V. M.; Rowitch, D. H.; Xing, X.; Fiore, C.; Schillebeeckx, M.; Jones, S. J.; Haussler, D.; Marra, M. A.; Hirst, M.; Wang, T. ; Costello, J. F. (2010). "Conserved role of intragenic DNA methylation in regulating alternative promoters." Nature 466(7303): 253-257.
- 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.
- Mette, M. F.; Aufsatz, W.; van der Winden, J.; Matzke, M. A. ; Matzke, A. J. (2000). "Transcriptional silencing and promoter methylation triggered by double-stranded RNA." The EMBO journal 19(19): 5194-5201.
- Mikkelsen, T. S.; Ku, M.; Jaffe, D. B.; Issac, B.; Lieberman, E.; Giannoukos, G.; Alvarez, P.; Brockman, W.; Kim, T. K.; Koche, R. P.; Lee, W.; Mendenhall, E.; O'Donovan, A.; Presser, A.; Russ, C.; Xie, X.; Meissner, A.; Wernig, M.; Jaenisch, R.; Nusbaum, C.; Lander, E. S. ; Bernstein, B. E. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature 448(7153): 553-560.
- Mizuguchi, G.; Shen, X.; Landry, J.; Wu, W. H.; Sen, S. ; Wu, C. (2004). "ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex." Science 303(5656): 343-348.
- Monk, M.; Boubelik, M. ; Lehnert, S. (1987). "Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development." Development 99(3): 371-382.
- Moran, V. A.; Perera, R. J. ; Khalil, A. M. (2012). "Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs." Nucleic Acids Res.
- Morris, K. V.; Chan, S. W.; Jacobsen, S. E. ; Looney, D. J. (2004). "Small interfering RNA-induced transcriptional gene silencing in human cells." Science 305(5688): 1289-1292.
- 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.
- Munroe, S. H. ; Lazar, M. A. (1991). "Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA." J Biol Chem 266(33): 22083-22086.
- Murray, K. (1964). "The Occurrence of Epsilon-N-Methyl Lysine in Histones." Biochemistry 3: 10-15.
- Muse, G. W.; Gilchrist, D. A.; Nechaev, S.; Shah, R.; Parker, J. S.; Grissom, S. F.; Zeitlinger, J. ; Adelman, K. (2007). "RNA polymerase is poised for activation across the genome." Nat Genet 39(12): 1507-1511.

- Nagano, T.; Mitchell, J. A.; Sanz, L. A.; Pauler, F. M.; Ferguson-Smith, A. C.; Feil, R. ; Fraser, P. (2008). "The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin." *Science* 322(5908): 1717-1720.
- 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 Biol* 8(3): R43.
- Nguyen, C.; Liang, G.; Nguyen, T. T.; Tsao-Wei, D.; Groshen, S.; Lubbert, M.; Zhou, J. H.; Benedict, W. F. ; Jones, P. A. (2001). "Susceptibility of nonpromoter CpG islands to de novo methylation in normal and neoplastic cells." *J Natl Cancer Inst* 93(19): 1465-1472.
- Numata, K.; Kanai, A.; Saito, R.; Kondo, S.; Adachi, J.; Wilming, L. G.; Hume, D. A.; Hayashizaki, Y. ; Tomita, M. (2003). "Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection." *Genome Res* 13(6B): 1301-1306.
- Nuovo, G. J.; Plaia, T. W.; Belinsky, S. A.; Baylin, S. B. ; Herman, J. G. (1999). "In situ detection of the hypermethylation-induced inactivation of the p16 gene as an early event in oncogenesis." *Proc Natl Acad Sci U S A* 96(22): 12754-12759.
- Ogawa, Y.; Sun, B. K. ; Lee, J. T. (2008). "Intersection of the RNA interference and X-inactivation pathways." *Science* 320(5881): 1336-1341.
- Ohm, J. E.; McGarvey, K. M.; Yu, X.; Cheng, L.; Schuebel, K. E.; Cope, L.; Mohammad, H. P.; Chen, W.; Daniel, V. C.; Yu, W.; Berman, D. M.; Jenuwein, T.; Pruitt, K.; Sharkis, S. J.; Watkins, D. N.; Herman, J. G. ; Baylin, S. B. (2007). "A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing." *Nat Genet* 39(2): 237-242.
- Okano, M.; Bell, D. W.; Haber, D. A. ; Li, E. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." *Cell* 99(3): 247-257.
- Okano, M.; Xie, S. ; Li, E. (1998). "Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells." *Nucleic acids research* 26(11): 2536-2540.
- Okazaki, Y.; Furuno, M.; Kasukawa, T.; Adachi, J.; Bono, H.; Kondo, S.; Nikaido, I.; Osato, N.; Saito, R.; Suzuki, H.; Yamanaka, I.; Kiyosawa, H.; Yagi, K.; Tomaru, Y.; Hasegawa, Y.; Nogami, A.; Schonbach, C.; Gojobori, T.; Baldarelli, R.; Hill, D. P.; Bult, C.; Hume, D. A.; Quackenbush, J.; Schriml, L. M.; Kanapin, A.; Matsuda, H.; Batalov, S.; Beisel, K. W.; Blake, J. A.; Bradt, D.; Brusic, V.; Chothia, C.; Corbani, L. E.; Cousins, S.; Dalla, E.; Dragani, T. A.; Fletcher, C. F.; Forrest, A.; Frazer, K. S.; Gaasterland, T.; Gariboldi, M.; Gissi, C.; Godzik, A.; Gough, J.; Grimmond, S.; Gustincich, S.; Hirokawa, N.; Jackson, I. J.; Jarvis, E. D.; Kanai, A.; Kawaji, H.; Kawasawa, Y.; Kedzierski, R. M.; King, B. L.; Konagaya, A.; Kurochkin, I. V.; Lee, Y.; Lenhard, B.; Lyons, P. A.; Maglott, D. R.; Maltais, L.; Marchionni, L.; McKenzie, L.; Miki, H.; Nagashima, T.; Numata, K.; Okido, T.; Pavan, W. J.; Pertea, G.; Pesole, G.; Petrovsky, N.; Pillai, R.; Pontius, J. U.; Qi, D.; Ramachandran, S.; Ravasi, T.; Reed, J. C.; Reed, D. J.; Reid, J.; Ring, B. Z.; Ringwald, M.; Sandelin, A.; Schneider, C.; Semple, C. A.; Setou, M.; Shimada, K.; Sultana, R.; Takenaka, Y.; Taylor, M. S.; Teasdale, R. D.; Tomita, M.; Verardo, R.; Wagner, L.; Wahlestedt, C.; Wang, Y.; Watanabe, Y.; Wells, C.; Wilming, L. G.; Wynshaw-Boris, A.; Yanagisawa, M.; Yang, I.; Yang, L.; Yuan, Z.; Zavolan, M.; Zhu, Y.; Zimmer, A.; Carninci, P.; Hayatsu, N.; Hirozane-Kishikawa, T.; Konno, H.; Nakamura, M.; Sakazume, N.; Sato, K.; Shiraki, T.; Waki, K.; Kawai, J.; Aizawa, K.; Arakawa, T.; Fukuda, S.; Hara,

- A.; Hashizume, W.; Imotani, K.; Ishii, Y.; Itoh, M.; Kagawa, I.; Miyazaki, A.; Sakai, K.; Sasaki, D.; Shibata, K.; Shinagawa, A.; Yasunishi, A.; Yoshino, M.; Waterston, R.; Lander, E. S.; Rogers, J.; Birney, E.; Hayashizaki, Y. (2002). "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs." Nature 420(6915): 563-573.
- Ongenaert, M.; Van Neste, L.; De Meyer, T.; Menschaert, G.; Bekaert, S.; Van Criekinge, W. (2008). "PubMeth: a cancer methylation database combining text-mining and expert annotation." Nucleic Acids Res 36(Database issue): D842-846.
- Ooi, S. K.; Qiu, C.; Bernstein, E.; Li, K.; Jia, D.; Yang, Z.; Erdjument-Bromage, H.; Tempst, P.; Lin, S. P.; Allis, C. D.; Cheng, X.; Bestor, T. H. (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature 448(7154): 714-717.
- Pandey, R. R.; Mondal, T.; Mohammad, F.; Enroth, S.; Redrup, L.; Komorowski, J.; Nagano, T.; Mancini-Dinardo, D.; Kanduri, C. (2008). "Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation." Mol Cell 32(2): 232-246.
- Pang, K. C.; Frith, M. C.; Mattick, J. S. (2006). "Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function." Trends Genet 22(1): 1-5.
- Patel, N.; Hoang, D.; Miller, N.; Ansaloni, S.; Huang, Q.; Rogers, J. T.; Lee, J. C.; Saunders, A. J. (2008). "MicroRNAs can regulate human APP levels." Mol Neurodegener 3: 10.
- Peters, J.; Robson, J. E. (2008). "Imprinted noncoding RNAs." Mamm Genome 19(7-8): 493-502.
- Petersen, C. P.; Bordeleau, M. E.; Pelletier, J.; Sharp, P. A. (2006). "Short RNAs repress translation after initiation in mammalian cells." Mol Cell 21(4): 533-542.
- Prendergast, G. C.; Ziff, E. B. (1991). "Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region." Science 251(4990): 186-189.
- Prensner, J. R.; Chinnaiyan, A. M. (2011). "The emergence of lncRNAs in cancer biology." Cancer Discov 1(5): 391-407.
- Quackenbush, J. (2001). "Computational analysis of microarray data." Nat Rev Genet 2(6): 418-427.
- Quinlan, A. R.; Hall, I. M. (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics 26(6): 841-842.
- Qureshi, I. A.; Mattick, J. S.; Mehler, M. F. (2010). "Long non-coding RNAs in nervous system function and disease." Brain Res 1338: 20-35.
- Rahl, P. B.; Lin, C. Y.; Seila, A. C.; Flynn, R. A.; McCuine, S.; Burge, C. B.; Sharp, P. A.; Young, R. A. (2010). "c-Myc regulates transcriptional pause release." Cell 141(3): 432-445.
- Ravasi, T.; Suzuki, H.; Pang, K. C.; Katayama, S.; Furuno, M.; Okunishi, R.; Fukuda, S.; Ru, K.; Frith, M. C.; Gongora, M. M.; Grimmond, S. M.; Hume, D. A.; Hayashizaki, Y.; Mattick, J. S. (2006). "Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome." Genome Res 16(1): 11-19.
- Razin, A.; Riggs, A. D. (1980). "DNA methylation and gene function." Science 210(4470): 604-610.
- Redrup, L.; Branco, M. R.; Perdeaux, E. R.; Krueger, C.; Lewis, A.; Santos, F.; Nagano, T.; Cobb, B. S.; Fraser, P.; Reik, W. (2009). "The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing." Development 136(4): 525-530.

- Reik, W. ; Walter, J. (2001). "Genomic imprinting: parental influence on the genome." Nat Rev Genet 2(1): 21-32.
- Reis, E. M.; Nakaya, H. I.; Louro, R.; Canavez, F. C.; Flatschart, A. V.; 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-Filho, 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.
- Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation." Cytogenet Cell Genet 14(1): 9-25.
- Ringrose, L.; Ehret, H. ; Paro, R. (2004). "Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes." Mol Cell 16(4): 641-653.
- 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.; Bruggmann, 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.
- Rodriguez, J.; Frigola, J.; Vendrell, E.; Risques, R. A.; Fraga, M. F.; Morales, C.; Moreno, V.; Esteller, M.; Capella, G.; Ribas, M. ; Peinado, M. A. (2006). "Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers." Cancer Res 66(17): 8462-9468.
- Rosok, O. ; Sioud, M. (2004). "Systematic identification of sense-antisense transcripts in mammalian cells." Nat Biotechnol 22(1): 104-108.
- Rozowsky, J.; Euskirchen, G.; Auerbach, R. K.; Zhang, Z. D.; Gibson, T.; Bjornson, R.; Carriero, N.; Snyder, M. ; Gerstein, M. B. (2009). "PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls." Nat Biotechnol 27(1): 66-75.
- Saha, S.; Sparks, A. B.; Rago, C.; Akmaev, V.; Wang, C. J.; Vogelstein, B.; Kinzler, K. W. ; Velculescu, V. E. (2002). "Using the transcriptome to annotate the genome." Nat Biotechnol 20(5): 508-512.
- Sandelin, A.; Carninci, P.; Lenhard, B.; Ponjavic, J.; Hayashizaki, Y. ; Hume, D. A. (2007). "Mammalian RNA polymerase II core promoters: insights from genome-wide studies." Nat Rev Genet 8(6): 424-436.
- Santos-Rosa, H.; Schneider, R.; Bannister, A. J.; Sherriff, J.; Bernstein, B. E.; Emre, N. C.; Schreiber, S. L.; Mellor, J. ; Kouzarides, T. (2002). "Active genes are tri-methylated at K4 of histone H3." Nature 419(6905): 407-411.
- Santos, F.; Hendrich, B.; Reik, W. ; Dean, W. (2002). "Dynamic reprogramming of DNA methylation in the early mouse embryo." Dev Biol 241(1): 172-182.
- Sarma, K. ; Reinberg, D. (2005). "Histone variants meet their match." Nat Rev Mol Cell Biol 6(2): 139-149.
- Sato, N.; Fukushima, N.; Maitra, A.; Matsubayashi, H.; Yeo, C. J.; Cameron, J. L.; Hruban, R. H. ; Goggins, M. (2003). "Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays." Cancer Res 63(13): 3735-3742.

- Saxonov, S.; Berg, P. ; Brutlag, D. L. (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." Proc Natl Acad Sci U S A 103(5): 1412-1417.
- Schlesinger, Y.; Straussman, R.; Keshet, I.; Farkash, S.; Hecht, M.; Zimmerman, J.; Eden, E.; Yakhini, Z.; Ben-Shushan, E.; Reubinoff, B. E.; Bergman, Y.; Simon, I. ; Cedar, H. (2007). "Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer." Nat Genet 39(2): 232-236.
- Schug, J.; Schuller, W. P.; Kappen, C.; Salbaum, J. M.; Bucan, M. ; Stoeckert, C. J., Jr. (2005). "Promoter features related to tissue specificity as measured by Shannon entropy." Genome Biol 6(4): R33.
- Schwartz, S.; Kent, W. J.; Smit, A.; Zhang, Z.; Baertsch, R.; Hardison, R. C.; Haussler, D. ; Miller, W. (2003). "Human-mouse alignments with BLASTZ." Genome Res 13(1): 103-107.
- Schwartz, S.; Meshorer, E. ; Ast, G. (2009). "Chromatin organization marks exon-intron structure." Nat Struct Mol Biol 16(9): 990-995.
- Secco, M.; Moreira, Y. B.; Zucconi, E.; Vieira, N. M.; Jazedje, T.; Muotri, A. R.; Okamoto, O. K.; Verjovski-Almeida, S. ; Zatz, M. (2009). "Gene expression profile of mesenchymal stem cells from paired umbilical cord units: cord is different from blood." Stem Cell Rev 5(4): 387-401.
- Segal, E.; Friedman, N.; Koller, D. ; Regev, A. (2004). "A module map showing conditional activity of expression modules in cancer." Nat Genet 36(10): 1090-1098.
- Seidl, C. I.; Stricker, S. H. ; Barlow, D. P. (2006). "The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export." Embo J 25(15): 3565-3575.
- Seubert, P.; Oltersdorf, T.; Lee, M. G.; Barbour, R.; Blomquist, C.; Davis, D. L.; Bryant, K.; Fritz, L. C.; Galasko, D.; Thal, L. J. ; et al. (1993). "Secretion of beta-amyloid precursor protein cleaved at the amino terminus of the beta-amyloid peptide." Nature 361(6409): 260-263.
- Shenker, N. ; Flanagan, J. M. (2012). "Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research." Br J Cancer 106(2): 248-253.
- Shevchenko, A. I.; Zakharova, I. S.; Elisaphenko, E. A.; Kolesnikov, N. N.; Whitehead, S.; Bird, C.; Ross, M.; Weidman, J. R.; Jirtle, R. L.; Karamysheva, T. V.; Rubtsov, N. B.; VandeBerg, J. L.; Mazurok, N. A.; Nesterova, T. B.; Brockdorff, N. ; Zakian, S. M. (2007). "Genes flanking Xist in mouse and human are separated on the X chromosome in American marsupials." Chromosome Res 15(2): 127-136.
- Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A. ; Casero, R. A. (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." Cell 119(7): 941-953.
- Shukla, S.; Kavak, E.; Gregory, M.; Imashimizu, M.; Shutinoski, B.; Kashlev, M.; Oberdoerffer, P.; Sandberg, R. ; Oberdoerffer, S. (2011). "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." Nature 479(7371): 74-79.
- 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.
- Siepel, A.; Bejerano, G.; Pedersen, J. S.; Hinrichs, A. S.; Hou, M.; Rosenbloom, K.; Clawson, H.; Spieth, J.; Hillier, L. W.; Richards, S.; Weinstock, G. M.; Wilson, R. K.; Gibbs, R. A.;

- Kent, W. J.; Miller, W. ; Haussler, D. (2005). "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes." Genome Res 15(8): 1034-1050.
- Simons, C.; Pheasant, M.; Makunin, I. V. ; Mattick, J. S. (2006). "Transposon-free regions in mammalian genomes." Genome Res 16(2): 164-172.
- Sims, R. J., 3rd; Belotserkovskaya, R. ; Reinberg, D. (2004). "Elongation by RNA polymerase II: the short and long of it." Genes Dev 18(20): 2437-2468.
- Siomi, M. C.; Sato, K.; Pezic, D. ; Aravin, A. A. (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nat Rev Mol Cell Biol 12(4): 246-258.
- Song, F.; Smith, J. F.; Kimura, M. T.; Morrow, A. D.; Matsuyama, T.; Nagase, H. ; Held, W. A. (2005). "Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression." Proc Natl Acad Sci U S A 102(9): 3336-3341.
- Stanke, M.; Diekhans, M.; Baertsch, R. ; Haussler, D. (2008). "Using native and syntenically mapped cDNA alignments to improve de novo gene finding." Bioinformatics 24(5): 637-644.
- Strahl, B. D. ; Allis, C. D. (2000). "The language of covalent histone modifications." Nature 403(6765): 41-45.
- Straussman, R.; Nejman, D.; Roberts, D.; Steinfeld, I.; Blum, B.; Benvenisty, N.; Simon, I.; Yakhini, Z. ; Cedar, H. (2009). "Developmental programming of CpG island methylation profiles in the human genome." Nat Struct Mol Biol 16(5): 564-571.
- Struhl, K. (2007). "Transcriptional noise and the fidelity of initiation by RNA polymerase II." Nat Struct Mol Biol 14(2): 103-105.
- Suetake, I.; Shinozaki, F.; Miyagawa, J.; Takeshima, H. ; Tajima, S. (2004). "DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction." J Biol Chem 279(26): 27816-27823.
- Svoboda, P.; Stein, P.; Filipowicz, W. ; Schultz, R. M. (2004). "Lack of homologous sequence-specific DNA methylation in response to stable dsRNA expression in mouse oocytes." Nucleic acids research 32(12): 3601-3606.
- Szymanski, M.; Barciszewska, M. Z.; Erdmann, V. A. ; Barciszewski, J. (2005). "A new frontier for molecular medicine: noncoding RNAs." Biochim Biophys Acta 1756(1): 65-75.
- Tabaska, J. E. ; Zhang, M. Q. (1999). "Detection of polyadenylation signals in human DNA sequences." Gene 231(1-2): 77-86.
- Taberlay, P. C.; Kelly, T. K.; Liu, C. C.; You, J. S.; De Carvalho, D. D.; Miranda, T. B.; Zhou, X. J.; Liang, G. ; Jones, P. A. (2011). "Polycomb-repressed genes have permissive enhancers that initiate reprogramming." Cell 147(6): 1283-1294.
- 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.; Pheasant, M. ; Mattick, J. S. (2007). "The relationship between non-protein-coding DNA and eukaryotic complexity." Bioessays 29(3): 288-299.
- Tagami, H.; Ray-Gallet, D.; Almouzni, G. ; Nakatani, Y. (2004). "Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis." Cell 116(1): 51-61.
- Tahira, A. C.; Kubrusly, M. S.; Faria, M. F.; Dazzani, B.; Fonseca, R. S.; Maracaja-Coutinho, V.; Verjovski-Almeida, S.; Machado, M. C. ; Reis, E. M. (2011). "Long noncoding intronic

- RNAs are differentially expressed in primary and metastatic pancreatic cancer." Mol Cancer 10(1): 141.
- Takai, D. ; Jones, P. A. (2002). "Comprehensive analysis of CpG islands in human chromosomes 21 and 22." Proc Natl Acad Sci U S A 99(6): 3740-3745.
- Takayama, K.; Tsutsumi, S.; Suzuki, T.; Horie-Inoue, K.; Ikeda, K.; Kaneshiro, K.; Fujimura, T.; Kumagai, J.; Urano, T.; Sakaki, Y.; Shirahige, K.; Sasano, H.; Takahashi, S.; Kitamura, T.; Ouchi, Y.; Aburatani, H. ; Inoue, S. (2009). "Amyloid precursor protein is a primary androgen target gene that promotes prostate cancer growth." Cancer Res 69(1): 137-142.
- Team, R. D. C. (2011). "R: A language and environment for statistical computing. R Foundation for Statistical Computing." from <http://www.R-project.org/>.
- Thomson, J. P.; Skene, P. J.; Selfridge, J.; Clouaire, T.; Guy, J.; Webb, S.; Kerr, A. R.; Deaton, A.; Andrews, R.; James, K. D.; Turner, D. J.; Illingworth, R. ; Bird, A. (2010). "CpG islands influence chromatin structure via the CpG-binding protein Cfp1." Nature 464(7291): 1082-1086.
- Ting, A. H.; Schuebel, K. E.; Herman, J. G. ; Baylin, S. B. (2005). "Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation." Nat Genet 37(8): 906-910.
- Tsukiyama, T.; Daniel, C.; Tamkun, J. ; Wu, C. (1995). "ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor." Cell 83(6): 1021-1026.
- Turner, B. M. (2002). "Cellular memory and the histone code." Cell 111(3): 285-291.
- Turner, B. M.; Birley, A. J. ; Lavender, J. (1992). "Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei." Cell 69(2): 375-384.
- Tusher, V. G.; Tibshirani, R. ; Chu, G. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A 98(9): 5116-5121.
- Valen, E.; Pascarella, G.; Chalk, A.; Maeda, N.; Kojima, M.; Kawazu, C.; Murata, M.; Nishiyori, H.; Lazarevic, D.; Motti, D.; Marstrand, T. T.; Tang, M. H.; Zhao, X.; Krogh, A.; Winther, O.; Arakawa, T.; Kawai, J.; Wells, C.; Daub, C.; Harbers, M.; Hayashizaki, Y.; Gustincich, S.; Sandelin, A. ; Carninci, P. (2009). "Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE." Genome Res 19(2): 255-265.
- 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.
- van Dijk, E. L.; Chen, C. L.; d'Aubenton-Carafa, Y.; Gourvennec, S.; Kwapisz, M.; Roche, V.; Bertrand, C.; Silvain, M.; Legoix-Ne, P.; Loeillet, S.; Nicolas, A.; Thermes, C. ; Morillon, A. (2011). "XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast." Nature 475(7354): 114-117.
- Varga-Weisz, P. D.; Wilm, M.; Bonte, E.; Dumas, K.; Mann, M. ; Becker, P. B. (1997). "Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II." Nature 388(6642): 598-602.
- Ventura, A. ; Jacks, T. (2009). "MicroRNAs and cancer: short RNAs go a long way." Cell 136(4): 586-591.
- Wade, P. A. (2001). "Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin." Hum Mol Genet 10(7): 693-698.
- Wales, M. M.; Biel, M. A.; el Deiry, W.; Nelkin, B. D.; Issa, J. P.; Cavenee, W. K.; Kuerbitz, S. J. ; Baylin, S. B. (1995). "p53 activates expression of HIC-1, a new candidate tumour suppressor gene on 17p13.3." Nature medicine 1(6): 570-577.

- Wang, A.; Kurdistani, S. K. ; Grunstein, M. (2002). "Requirement of Hos2 histone deacetylase for gene activity in yeast." Science 298(5597): 1412-1414.
- 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, E. T.; Sandberg, R.; Luo, S.; Khrebtkova, I.; Zhang, L.; Mayr, C.; Kingsmore, S. F.; Schroth, G. P. ; Burge, C. B. (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature 456(7221): 470-476.
- Wang, K. C. ; Chang, H. Y. (2011). "Molecular mechanisms of long noncoding RNAs." Mol Cell 43(6): 904-914.
- Wang, X.; Arai, S.; Song, X.; Reichart, D.; Du, K.; Pascual, G.; Tempst, P.; Rosenfeld, M. G.; Glass, C. K. ; Kurokawa, R. (2008). "Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription." Nature 454(7200): 126-130.
- Wang, Y. ; Leung, F. C. (2004). "An evaluation of new criteria for CpG islands in the human genome as gene markers." Bioinformatics 20(7): 1170-1177.
- Washietl, S.; Hofacker, I. L.; Lukasser, M.; Huttenhofer, A. ; Stadler, P. F. (2005). "Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome." Nat Biotechnol 23(11): 1383-1390.
- Washietl, S.; Hofacker, I. L. ; Stadler, P. F. (2005). "Fast and reliable prediction of noncoding RNAs." Proc Natl Acad Sci U S A 102(7): 2454-2459.
- 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.; Esvara, P.; Eyraas, 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.
- Weber, M.; Hellmann, I.; Stadler, M. B.; Ramos, L.; Paabo, S.; Rebhan, M.; Schubeler, D. (2007). "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." *Nat Genet* 39(4): 457-466.
- West, R. L.; Lee, J. M.; Maroun, L. E. (1995). "Hypomethylation of the amyloid precursor protein gene in the brain of an Alzheimer's disease patient." *J Mol Neurosci* 6(2): 141-146.
- Widschwendter, M.; Fiegl, H.; Egle, D.; Mueller-Holzner, E.; Spizzo, G.; Marth, C.; Weisenberger, D. J.; Campan, M.; Young, J.; Jacobs, I.; Laird, P. W. (2007). "Epigenetic stem cell signature in cancer." *Nat Genet* 39(2): 157-158.
- 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.
- Williams, K.; Christensen, J.; Pedersen, M. T.; Johansen, J. V.; Cloos, P. A.; Rappaport, J.; Helin, K. (2011). "TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity." *Nature* 473(7347): 343-348.
- Willingham, A. T.; Gingeras, T. R. (2006). "TUF love for 'junk' DNA." *Cell* 125(7): 1215-1220.
- Wilson, A. S.; Power, B. E.; Molloy, P. L. (2007). "DNA hypomethylation and human diseases." *Biochim Biophys Acta* 1775(1): 138-162.
- Wilusz, J. E.; Sunwoo, H.; Spector, D. L. (2009). "Long noncoding RNAs: functional surprises from the RNA world." *Genes Dev* 23(13): 1494-1504.
- Wyers, F.; Rougemaille, M.; Badis, G.; Rousselle, J. C.; Dufour, M. E.; Boulay, J.; Regnault, B.; Devaux, F.; Namane, A.; Seraphin, B.; Libri, D.; Jacquier, A. (2005). "Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase." *Cell* 121(5): 725-737.
- Xu, G. L.; Bestor, T. H.; Bourc'his, D.; Hsieh, C. L.; Tommerup, N.; Bugge, M.; Hulten, M.; Qu, X.; Russo, J. J.; Viegas-Pequignot, E. (1999). "Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene." *Nature* 402(6758): 187-191.
- Yan, M. D.; Hong, C. C.; Lai, G. M.; Cheng, A. L.; Lin, Y. W.; Chuang, S. E. (2005). "Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas." *Hum Mol Genet* 14(11): 1465-1474.
- Yap, K. L.; Li, S.; Munoz-Cabello, A. M.; Raguz, S.; Zeng, L.; Mujtaba, S.; Gil, J.; Walsh, M. J.; Zhou, M. M. (2010). "Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a." *Mol Cell* 38(5): 662-674.
- Yeh, A.; Wei, M.; Golub, S. B.; Yamashiro, D. J.; Murty, V. V.; Tycko, B. (2002). "Chromosome arm 16q in Wilms tumors: unbalanced chromosomal translocations, loss of

- heterozygosity, and assessment of the CTCF gene." Genes Chromosomes Cancer 35(2): 156-163.
- Yoder, J. A.; Walsh, C. P. ; Bestor, T. H. (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet 13(8): 335-340.
- You, J. S.; Kelly, T. K.; De Carvalho, D. D.; Taberlay, P. C.; Liang, G. ; Jones, P. A. (2011). "OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes." Proc Natl Acad Sci U S A 108(35): 14497-14502.
- Yu, W.; Gius, D.; Onyango, P.; Muldoon-Jacobs, K.; Karp, J.; Feinberg, A. P. ; Cui, H. (2008). "Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA." Nature 451(7175): 202-206.
- Zeng, L. ; Zhou, M. M. (2002). "Bromodomain: an acetyl-lysine binding domain." FEBS Lett 513(1): 124-128.
- Zhang, Y. ; Reinberg, D. (2001). "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails." Genes Dev 15(18): 2343-2360.
- 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.
- Zhu, J.; Sanborn, J. Z.; Diekhans, M.; Lowe, C. B.; Pringle, T. H. ; Haussler, D. (2007). "Comparative genomics search for losses of long-established genes on the human lineage." PLoS Comput Biol 3(12): e247.
- Zilberman, D.; Coleman-Derr, D.; Ballinger, T. ; Henikoff, S. (2008). "Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks." Nature 456(7218): 125-129.

## APÊNDICE

### Apêndice 1 - Genes codificadores de proteínas diferencialmente expressos após tratamento da linhagem celular DU-145 com o agente desmetilante 5-AZA.

<i>Gene ID</i>	<i>Gene Symbol</i>	<i>Strand</i>	<i>Probe Coordinate</i>	<i>Fold Change (Log2(AZA/CTL))</i>
85453	TSPYL5	-	chr8:98285785-98285844	6,9
50511	SYCP3	-	chr12:102128796-102130847	6,9
93953	ACRC	+	chrX:70833291-70833350	5,5
285	ANGPT2	-	chr8:6360290-6360349	5,4
100128098	LOC100128098	+	chr10:17450221-17450280	5,3
56159	TEX11	-	chrX:69748974-69749033	5,2
143903	LAYN	+	chr11:111431460-111431519	5,0
4744	NEFH	+	chr22:29887194-29887253	4,9
8277	TKTL1	+	chrX:153558624-153558683	4,7
885	CCK	-	chr3:42299317-42299376	4,5
55384	MEG3	+	chr14:101327186-101327244	4,4
7802	DNALI1	+	chr1:38030666-38030725	4,3
54456	MOV10L1	+	chr22:50530553-50530612	4,1
4105	MAGEA6	+	chrX:151870169-151870228	4,1
6622	SNCA	-	chr4:90647685-90647744	4,1
54	ACP5	-	chr19:11685525-11685584	4,1
26289	AK5	+	chr1:78024290-78024349	4,0
9084	VCY	-	chrY:16097713-16097772	4,0
1264	CNN1	+	chr19:11661005-11661064	4,0
6364	CCL20	+	chr2:228681777-228681836	3,9
2242	FES	+	chr15:91436923-91436982	3,7
653140	C2orf84	+	chr2:24406202-24406382	3,6
8277	TKTL1	+	chrX:153551583-153551642	3,6
83733	SLC25A18	+	chr22:18072938-18072997	3,5
6364	CCL20	+	chr2:228680170-228680229	3,4
3635	INPP5D	+	chr2:234116417-234116476	3,4
93953	ACRC	+	chrX:70817800-70817859	3,3
126123	IZUMO2	-	chr19:50657856-50657886	3,3
79092	CARD14	+	chr17:78182633-78182692	3,2
5579	PRKCB	+	chr16:24227210-24227269	3,2
8626	TP63	+	chr3:189614963-189615022	3,2
3934	LCN2	+	chr9:130914244-130914303	3,2
84690	SPATA22	-	chr17:3346496-3346555	3,2
133584	EGFLAM	+	chr5:38464510-38464569	3,1
88	ACTN2	+	chr1:236925943-236926002	3,1
56159	TEX11	-	chrX:69898680-69898739	3,0

154197	PNLDC1	+	chr6:160240056-160240115	3,0
10942	PRSS21	+	chr16:2871094-2871409	2,9
7169	TPM2	-	chr9:35684492-35684549	2,9
3859	KRT12	-	chr17:39017641-39017698	2,9
6943	TCF21	+	chr6:134210941-134210985	2,9
667	DST	-	chr6:56479979-56480038	2,9
9148	NEURL	+	chr10:105352070-105352129	2,8
219731	LOC219731	-	chr10:11893319-11893378	2,8
4892	NRAP	-	chr10:115348649-115348708	2,8
5961	PRPH2	-	chr6:42664496-42664555	2,8
196951	C15orf33	-	chr15:49659750-49659809	2,7
5678	PSG9	-	chr19:43762430-43762489	2,7
23209	MLC1	-	chr22:50497948-50498007	2,7
8789	FBP2	-	chr9:97321090-97321149	2,7

**Apêndice 2 – LncRNAs intrônicos antissenso diferencialmente expressos após tratamento da linhagem celular DU-145 com o agente desmetilante 5-AZA.**

**50 lncRNAs Antissenso com maior diferença de expressão após tratamento com 5 µM de 5-AZA (SAM; q-valor < 10%)**

<i>Antisense ncRNA Contig Name</i>	<i>Antisense ncRNA Contig Coordinate</i>	<i>Strand</i>	<i>Host Gene ID</i>	<i>Host Gene Symbol</i>	<i>Fold Change (Log2(AZA/CTL))</i>
PIN_11039	chr7:74113224-74113702	-	2969	GTF2I	2,7
TIN_20703	chr17:1965309-1966536	+	23293	SMG6	2,5
TIN_38315	chr4:73970575-73971290	+	26057	ANKRD17	2,5
TIN_38318	chr4:73975995-73976457	+	26057	ANKRD17	2,4
TIN_31722	chr21:27259582-27260205	+	351	APP	2,0
PIN_00053	chr1:7870082-7871274	-	8863	PER3	2,0
TIN_40609	chr5:54750290-54751056	+	8611	PPAP2A	1,9
			20063		
PIN_06847	chr2:27666400-27667052	-	4	KRTCAP3	1,7
PIN_09759	chr5:43297768-43298287	+	3157	HMGCS1	1,6
PIN_00096	chr1:15792053-15793880	-	63036	CELA2A	1,5
TIN_03373	chr1:167789419-167791187	+	55811	ADCY10	1,5
	chr12:133707870-				
TIN_13499	133709104	-	7556	ZNF10	1,4
PIN_01484	chr10:75413651-75414087	+	79933	SYNPO2L	1,4
TIN_05238	chr10:11364897-11365170	-	10659	CELF2	1,4
PIN_03896	chr14:93299420-93300482	-	9950	GOLGA5	1,4
TIN_18020	chr15:71581150-71581677	-	79875	THSD4	1,4
PIN_07632	chr2:228680078-228680392	-	6364	CCL20	1,3
TIN_54326	chrX:39959017-39959361	+	54880	BCOR	1,3
PIN_09760	chr5:43298303-43299526	+	3157	HMGCS1	1,3
TIN_31450	chr20:54946800-54947591	+	6790	AURKA	1,3
PIN_06434	chr19:39329239-39330362	+	3191	HNRNPL	1,3
PIN_06742	chr2:10103631-10104363	-	29841	GRHL1	1,2
PIN_07962	chr20:47255339-47256700	+	57580	PREX1	1,2
PIN_03342	chr13:31711662-31713319	+	10808	HSPH1	1,2
PIN_06565	chr19:47763680-47764361	-	26093	CCDC9	1,2
PIN_08922	chr3:121447218-121448256	+	2804	GOLGB1	1,2
PIN_02657	chr12:15806740-15807363	+	2059	EPS8	1,2
PIN_06937	chr2:47605824-47606314	-	4072	EPCAM	1,2
TIN_05026	chr1:247071523-247072476	+	25909	AHCTF1	1,1
TIN_00727	chr1:26065627-26066146	-	57134	MAN1C1	1,1
PIN_05945	chr18:33846463-33847148	-	55034	MOCOS	1,1
PIN_03343	chr13:31723449-31724601	+	10808	HSPH1	1,1
TIN_45727	chr6:170671785-170672896	-	84498	FAM120B	1,0

PIN_01563	chr10:96099829-96101188	+	64318 19922	NOC3L	1,0
PIN_08994	chr3:137783248-137783833	+	1	DZIP1L	1,0
PIN_04398	chr15:90165420-90169127	-	90381	C15orf42	1,0
PIN_11737	chr8:131414067-131414510	+	50807	ASAP1	1,0
TIN_23254	chr18:11749382-11750327	-	2774	GNAL	1,0
PIN_01540	chr10:92678533-92679082	+	27063	ANKRD1	1,0
PIN_03080	chr12:101679904- 101680576	-	27340	UTP20	0,9
PIN_01053	chr1:200816196-200816712	-	23271	CAMSAP2	0,9
PIN_10140	chr5:167921413-167922008	-	5917	RARS	0,9
PIN_04028	chr15:39876085-39876654	-	7057	THBS1	0,9
TIN_42527	chr5:167935627-167936461	-	5917 28448	RARS	0,9
PIN_00785	chr1:151823237-151824027	+	6	THEM5	0,9
PIN_04125	chr15:45490715-45491663	+	90525	SHF	0,9
PIN_04417	chr15:91353979-91354700	-	641 19992	BLM	0,9
TIN_01681	chr1:57189745-57190853	+	0	C1orf168	0,9
PIN_08444	chr22:46724407-46725067	-	51512	GTSE1	0,9
PIN_00660	chr1:114945506-114948783	+	51592 13769	TRIM33	0,9
PIN_11566	chr8:56668725-56669168	+	5	TMEM68	0,9

**Apêndice 3 – LncRNAs intrônicos senso diferencialmente expressos após tratamento da linhagem celular DU-145 com o agente desmetilante 5-AZA.**

**50 lncRNAs senso com maior diferença de expressão após tratamento com 5 µM de 5-AZA (SAM; q-valor < 10%)**

<i>Sense ncRNA Contig Name</i>	<i>Sense ncRNA Contig Coordinate</i>	<i>Strand</i>	<i>Host Gene ID</i>	<i>Host Gene Symbol</i>	<i>Fold Change (Log2(AZA/CTL))</i>
TIN_45726	chr6:170667533-170668084	+	84498	FAM120B	1,8
TIN_13748	chr13:29249476-29250591	+	51371	POMP	1,8
TIN_00094	chr1:6516255-6516743	+	83715	ESPN	1,7
TIN_14995	chr13:114159537-114160378	+	55002	TMCO3	1,7
TIN_45798	chr7:2149012-2150325	-	8379	MAD1L1	1,7
TIN_45729	chr6:170680760-170681208	+	84498	FAM120B	1,4
TIN_31505	chr20:57464807-57465327	+	2778	GNAS	1,4
TIN_40399	chr5:37520029-37520535	+	55100	WDR70	1,3
TIN_13422	chr12:125622311-125623428	+	65985	AACS	1,3
TIN_53539	chr9:127494111-127494986	-	2649	NR6A1	1,3
TIN_21939	chr17:45010652-45011368	+	9570	GOSR2	1,3
TIN_02568	chr1:111500546-111501524	+	55791	LRIF1	1,3
TIN_15556	chr14:55748500-55750066	+	55030	FBXO34	1,2
TIN_11658	chr12:40207409-40207855	-	114134	SLC2A13	1,2
TIN_19650	chr16:50115713-50116262	+	55027	HEATR3	1,2
TIN_14435	chr13:76109743-76110448	-	170622	COMMD6	1,2
TIN_21756	chr17:40405061-40405851	-	6777	STAT5B	1,2
TIN_28113	chr2:122023744-122024277	-	29842	TFCP2L1	1,2
TIN_38646	chr4:88117145-88117813	-	57563	KLHL8	1,2
TIN_38319	chr4:73976998-73978621	-	26057	ANKRD17	1,2
TIN_11192	chr12:13048897-13050167	+	9052	GPRC5A	1,2
TIN_16740	chr14:105910579-105911244	+	9112	MTA1	1,2
TIN_53760	chr9:134487208-134487693	-	2889	RAPGEF1	1,1
TIN_39985	chr5:463341-463968	+	11336	EXOC3	1,1
TIN_17284	chr15:51213844-51214235	+	23431	AP4E1	1,1
TIN_30596	chr20:18540005-18540984	+	10483	SEC23B	1,1
TIN_28554	chr2:157185090-157185716	-	4929	NR4A2	1,1
TIN_13499	chr12:133707870-133709104	+	7556	ZNF10	1,0
TIN_39256	chr4:121734427-121734937	-	11107	PRDM5	0,9
TIN_26370	chr2:31329407-31330049	-	79623	GALNT14	0,9
TIN_40156	chr5:14682040-14682725	+	90268	FAM105B	0,9
TIN_00963	chr1:33726813-33727281	+	149076	ZNF362	0,9
TIN_33547	chr3:15362873-15363400	-	9467	SH3BP5	0,9
TIN_04314	chr1:212922712-212923657	-	25936	NSL1	0,9
TIN_03222	chr1:161861496-161861977	+	22926	ATF6	0,9
TIN_19395	chr16:23867017-23867674	+	5579	PRKCB	0,9

TIN_40653	chr5:56539722-56540996	+	65056	GPBP1	0,9
TIN_18122	chr15:72854318-72855246	+	25820	ARIH1	0,9
TIN_14689	chr13:99054660-99055293	+	10160	FARP1	0,9
TIN_28270	chr2:135159221-135159714	+	4249	MGAT5	0,9
TIN_00951	chr1:33483244-33484105	-	204	AK2	0,9
TIN_54559	chrX:70586485-70587178	+	618	BCYRN1	0,9
TIN_41547	chr5:112885149-112885714	+	64848	YTHDC2	0,8
TIN_20551	chr16:89599777-89600199	+	6687	SPG7	0,8
TIN_42330	chr5:150495528-150496404	-	309	ANXA6	0,8
TIN_20631	chr17:879210-880004	-	64359	NXN	0,8
TIN_17548	chr15:59222914-59224245	-	79811	SLTM	0,8
TIN_00460	chr1:19674629-19675951	-	832	CAPZB	0,8
TIN_28518	chr2:153290321-153290966	+	114793	FMNL2	0,8
TIN_27511	chr2:80171111-80171610	+	1496	CTNNA2	0,8
TIN_04731	chr1:235494086-235496388	+	9453	GGPS1	0,8

**Apêndice 4 – Genes codificadores de proteínas diferencialmente expressos após tratamento da linhagem celular MCF-7 com o agente desmetilante 5-AZA.**

**50 mRNAs com maior diferença de expressão após tratamento com 5 µM de 5-AZA (SAM; q-valor < 10%)**

<i>Gene ID</i>	<i>Gene Symbol</i>	<i>Strand</i>	<i>Probe Coordinate</i>	<i>Fold Change (Log2(AZA/CTL))</i>
7429	VIL1	+	chr2:219314095-219314154	5,3
93953	ACRC	+	chrX:70833291-70833350	5,2
7802	DNALI1	+	chr1:38030666-38030725	5,2
100128098	LOC100128098	+	chr10:17450221-17450280	5,1
4744	NEFH	+	chr22:29887194-29887253	5,0
50511	SYCP3	-	chr12:102128796-102130847	5,0
4680	CEACAM6	+	chr19:42275435-42275493	4,8
5918	RARRES1	-	chr3:158422630-158422689	4,6
3059	HCLS1	-	chr3:121350279-121350336	4,4
8789	FBP2	-	chr9:97321090-97321149	4,3
93953	ACRC	+	chrX:70817800-70817859	4,3
9547	CXCL14	-	chr5:134907529-134910291	4,2
9084	VCY	-	chrY:16097713-16097772	4,2
2950	GSTP1	+	chr11:67354050-67354109	4,1
83871	RAB34	-	chr17:27041349-27041408	4,1
54742	LY6K	+	chr8:143784821-143784880	4,1
8277	TKTL1	+	chrX:153558624-153558683	4,0
54	ACP5	-	chr19:11685525-11685584	3,9
8764	TNFRSF14	-	chr1:2495143-2495193	3,9
4680	CEACAM6	+	chr19:42275770-42275824	3,8
2192	FBLN1	+	chr22:45996933-45996992	3,8
83871	RAB34	-	chr17:27041859-27041918	3,7
4599	MX1	+	chr21:42831014-42831073	3,6
79132	DHX58	-	chr17:40253459-40253518	3,6
440335	LOC440335	+	chr16:4845788-4845847	3,6
54742	LY6K	+	chr8:143784626-143784685	3,5
6778	STAT6	-	chr12:57489651-57489710	3,5
6678	SPARC	-	chr5:151042185-151042244	3,5
4057	LTF	-	chr3:46479431-46479489	3,4
7570	ZNF22	+	chr10:45500488-45500547	3,4
2139	EYA2	+	chr20:45817096-45817155	3,2
7018	TF	+	chr3:133497549-133498039	3,2
55655	NLRP2	+	chr19:55505674-55505733	3,2
4599	MX1	+	chr21:42813643-42813702	3,1
133584	EGFLAM	+	chr5:38464510-38464569	3,1
144406	WDR66	+	chr12:122441655-122441714	3,1

51700	CYB5R2	-	chr11:7686352-7686411	3,1
9121	SLC16A5	+	chr17:73102150-73102208	3,1
6286	S100P	+	chr4:6698819-6698878	3,1
9823	ARMCX2	-	chrX:100910335-100910394	3,1
1396	CRIP1	+	chr14:105954700-105954841	3,1
9473	C1orf38	+	chr1:28213027-28213087	3,0
126123	IZUMO2	-	chr19:50657856-50657886	3,0
4281	MID1	-	chrX:10416703-10416762	2,9
55384	MEG3	+	chr14:101327186-101327244	2,8
11167	FSTL1	-	chr3:120113432-120113491	2,8
4162	MCAM	-	chr11:119180131-119180190	2,7
6277	S100A6	-	chr1:153507255-153507685	2,7
84894	LINGO1	-	chr15:77906026-77906085	2,7
3855	KRT7	+	chr12:52631297-52631354	2,7
63948	DMRTB1	+	chr1:53930465-53930524	2,7

**Apêndice 5 – LncRNAs intrônicos antissenso diferencialmente expressos após tratamento da linhagem celular MCF-7 com o agente desmetilante 5-AZA.**

**50 lncRNAs antissenso com maior diferença de expressão após tratamento com 5 µM de 5-AZA (SAM; q-valor < 10%)**

<i>Antisense</i>					
<i>lncRNA Contig Name</i>	<i>Antisense lncRNA Contig Coordinate</i>	<i>Strand</i>	<i>Host Gene ID</i>	<i>Host Gene Symbol</i>	<i>Fold Change (Log2(AZA/CTL))</i>
PIN_04900	chr16:68021709-68022136	+	64174	DPEP2	3,2
TIN_31722	chr21:27259582-27260205	+	351	APP	3,0
PIN_08129	chr21:42813321-42813875	-	4599	MX1	2,3
TIN_31891	chr21:36218040-36219259	+	861	RUNX1	2,2
TIN_11174	chr12:12636159-12636900	+	80824	DUSP16	2,1
TIN_20703	chr17:1965309-1966536	+	23293	SMG6	1,8
PIN_02641	chr12:12631807-12635533	+	80824	DUSP16	1,7
PIN_06434	chr19:39329239-39330362	+	3191	HNRNPL	1,7
PIN_11454	chr8:19276835-19278334	+	55790	CSGALNACT1	1,6
PIN_00785	chr1:151823237-151824027	+	284486	THEM5	1,6
TIN_40895	chr5:68391322-68391898	-	64924	SLC30A5	1,6
TIN_25908	chr2:8882277-8882965	+	57498	KIDINS220	1,6
TIN_14959	chr13:113663985-113664633	-	23263	MCF2L	1,5
PIN_06111	chr19:6745006-6746462	-	9322	TRIP10	1,4
TIN_13518	chr13:20286004-20287024	+	55269	PSPC1	1,4
PIN_03342	chr13:31711662-31713319	+	10808	HSPH1	1,4
PIN_00096	chr1:15792053-15793880	-	63036	CELA2A	1,3
TIN_42527	chr5:167935627-167936461	-	5917	RARS	1,3
PIN_10385	chr6:32809513-32812055	+	5696	PSMB8	1,2
PIN_01691	chr10:119026627-119028179	-	6571	SLC18A2	1,2
PIN_09916	chr5:102472345-102472561	-	23262	PPIP5K2	1,2
PIN_00636	chr1:111690113-111690997	-	10390	CEPT1	1,2
PIN_00287	chr1:36474755-36475555	-	192669	EIF2C3	1,2
PIN_06110	chr19:6742812-6743504	-	9322	TRIP10	1,2
PIN_10095	chr5:149785901-149787437	+	972	CD74	1,1
PIN_10405	chr6:33255833-33256390	+	9277	WDR46	1,1
PIN_05417	chr17:40262661-40263028	+	79132	DHX58	1,1
PIN_04311	chr15:74326582-74329243	-	5371	PML	1,0
PIN_05632	chr17:61790577-61792306	+	92335	STRADA	1,0
PIN_04504	chr16:1990660-1991535	+	51734	SEPX1	1,0
TIN_25012	chr19:34680468-34681442	-	26065	LSM14A	1,0
PIN_09270	chr4:8456308-8456669	-	152992	METTL19	1,0
PIN_03343	chr13:31723449-31724601	+	10808	HSPH1	1,0
PIN_02321	chr11:77102999-77104070	+	5058	PAK1	1,0
TIN_17512	chr15:58432414-58433762	-	366	AQP9	1,0

PIN_01231	chr1:237788592-237789135	-	6262	RYR2	1,0
PIN_04991	chr16:88712090-88713508	+	1535	CYBA	1,0
PIN_10799	chr6:160100331-160103790	+	6648	SOD2	0,9
PIN_06868	chr2:32664267-32664963	-	57448	BIRC6	0,9
TIN_38318	chr4:73975995-73976457	+	26057	ANKRD17	0,9
TIN_13876	chr13:35724421-35725254	-	26960	NBEA	0,9
TIN_08454	chr11:14327635-14328200	+	22800	RRAS2	0,9
TIN_44060	chr6:56405938-56406784	+	667	DST	0,9
PIN_07093	chr2:86348306-86348933	-	55037	PTCD3	0,9
TIN_44887	chr6:128637921-128638586	+	5796	PTPRK	0,9
TIN_09928	chr11:85772453-85774144	+	8301	PICALM	0,9
TIN_23996	chr18:60874703-60875401	+	596	BCL2	0,9
TIN_31220	chr20:45902716-45903040	+	23613	ZMYND8	0,9
TIN_19390	chr16:23595804-23596621	+	4706	NDUFAB1	0,9
TIN_14326	chr13:66925898-66926494	+	5101	PCDH9	0,9
PIN_05535	chr17:48154297-48154742	-	3675	ITGA3	0,9

**Apêndice 6 – LncRNAs intrônicos senso diferencialmente expressos após tratamento da linhagem celular MCF-7 com o agente desmetilante 5-AZA.**

**50 lncRNAs senso com maior diferença de expressão após tratamento com 5 µM de 5-AZA (SAM; q-valor < 10%)**

<i>Sense lncRNA Contig Name</i>	<i>Sense lncRNA Contig Coordinate</i>	<i>Strand</i>	<i>Host Gene ID</i>	<i>Host Gene Symbol</i>	<i>Fold Change (Log2(AZA/CTL))</i>
TIN_39401	chr4:141020240-141021142	-	55534	MAML3	2,4
TIN_11917	chr12:51553837-51554932	-	7024	TFCP2	2,2
TIN_53891	chr9:139330015-139330696	-	56623	INPP5E	1,6
TIN_40244	chr5:17245954-17246426	+	10409	BASP1	1,5
TIN_21189	chr17:20105973-20106593	+	92521	SPECC1	1,5
TIN_19964	chr16:69606014-69606434	+	10725	NFAT5	1,5
TIN_20631	chr17:879210-880004	-	64359	NXN	1,4
TIN_24135	chr18:77741564-77741973	-	10907	TXNL4A	1,4
TIN_30155	chr2:239243192-239243810	+	26146	TRAF3IP1	1,3
TIN_10365	chr11:115211408-115212013	-	23705	CADM1	1,3
TIN_22018	chr17:47405635-47406175	-	22834	ZNF652	1,2
TIN_26771	chr2:46529256-46529873	+	2034	EPAS1	1,2
TIN_27586	chr2:86396442-86397217	-	10989	IMMT	1,2
TIN_18498	chr15:90644180-90644773	-	3418	IDH2	1,2
TIN_00286	chr1:10418035-10420421	+	23095	KIF1B	1,2
TIN_53397	chr9:124081903-124082288	+	2934	GSN	1,1
TIN_41625	chr5:118619418-118620374	+	25816	TNFAIP8	1,1
TIN_05679	chr10:24782053-24782562	+	56243	KIAA1217	1,1
TIN_40399	chr5:37520029-37520535	+	55100	WDR70	1,1
TIN_11192	chr12:13048897-13050167	+	9052	GPRC5A	1,1
TIN_00394	chr1:16288343-16288563	+	7709	ZBTB17	1,1
TIN_31870	chr21:35891297-35892459	-	1827	RCAN1	1,1
TIN_01392	chr1:46057726-46058208	+	4678	NASP	1,1
TIN_22752	chr17:73734494-73735052	+	3691	ITGB4	1,1
TIN_41116	chr5:80395916-80397009	+	5924	RASGRF2	1,1
TIN_14625	chr13:97878066-97879841	+	10150	MBNL2	1,1
TIN_12593	chr12:96601022-96601590	+	2004	ELK3	1,1
TIN_17790	chr15:64494219-64495243	-	53944	CSNK1G1	1,1
TIN_54760	chrX:107921690-107923038	+	1287	COL4A5	1,0
TIN_23382	chr18:19278909-19279529	-	171586	ABHD3	1,0
TIN_26418	chr2:32606357-32607877	+	57448	BIRC6	1,0
TIN_41731	chr5:127463999-127464506	+	6558	SLC12A2	1,0
TIN_16957	chr15:40388378-40388609	-	90427	BMF	1,0
TIN_27159	chr2:64105633-64106144	+	7360	UGP2	1,0
TIN_24518	chr19:10837767-10838649	+	1785	DNM2	1,0
TIN_11243	chr12:14992316-14992847	-	420	ART4	1,0

TIN_02346	chr1:94134280-94134671	-	8412	BCAR3	1,0
TIN_25205	chr19:39833949-39834639	+	55095	SAMD4B	1,0
TIN_37096	chr3:192530952-192531611	-	151963	MB21D2	1,0
TIN_45726	chr6:170667533-170668084	+	84498	FAM120B	1,0
TIN_26770	chr2:46528369-46528966	+	2034	EPAS1	1,0
TIN_04054	chr1:202539777-202541961	+	4660	PPP1R12B	1,0
TIN_32177	chr21:44267619-44268550	-	10785	WDR4	1,0
TIN_11048	chr12:7303836-7304659	+	9746	CLSTN3	1,0
TIN_41618	chr5:118514190-118515001	+	1657	DMXL1	1,0
TIN_54789	chrX:109559435-109560135	-	9949	AMMECR1	1,0
TIN_20829	chr17:4743290-4743989	+	50488	MINK1	0,9
TIN_14047	chr13:43563187-43563719	-	94240	EPST11	0,9
TIN_52305	chr9:37826865-37827383	+	79269	DCAF10	0,9
TIN_45199	chr6:147656445-147657645	+	134957	STXBP5	0,9
TIN_41140	chr5:80999623-81000513	-	23635	SSBP2	0,9

**Apêndice 7 – Genes codificadores de proteínas diferencialmente expressos após tratamento da linhagem celular Mia PaCa 2 com o agente desmetilante 5-AZA.**

<i>Gene ID</i>	<i>Gene Symbol</i>	<i>Strand</i>	<i>Probe Coordinate</i>	<i>Fold Change (Log2(AZA/CTL))</i>
3945	LDHB	-	chr12:21790069-21790128	7,30
7980	TFPI2	-	chr7:93516575-93516634	7,17
664	BNIP3	-	chr10:133781312-133781370	7,13
3945	LDHB	-	chr12:21788449-21788508	6,96
50511	SYCP3	-	chr12:102128796-102130847	6,80
3945	LDHB	-	chr12:21795002-21795061	6,46
8277	TKTL1	+	chrX:153558624-153558683	6,16
3383	ICAM1	+	chr19:10396298-10396358	5,82
7980	TFPI2	-	chr7:93518473-93518532	5,58
4072	EPCAM	+	chr2:47613894-47613953	5,56
100128098	LOC100128098	+	chr10:17450221-17450280	5,37
4105	MAGEA6	+	chrX:151870169-151870228	5,35
85453	TSPYL5	-	chr8:98285785-98285844	5,35
6364	CCL20	+	chr2:228681777-228681836	5,28
79623	GALNT14	-	chr2:31133370-31133429	5,14
3486	IGFBP3	-	chr7:45952608-45952667	5,12
4744	NEFH	+	chr22:29887194-29887253	5,12
93953	ACRC	+	chrX:70833291-70833350	5,07
79746	ECHDC3	+	chr10:11805963-11806022	5,05
56300	IL36G	+	chr2:113743007-113743066	4,99
115908	CTHRC1	+	chr8:104390268-104390327	4,95
55930	MYO5C	-	chr15:52484644-52484703	4,92
57580	PREX1	-	chr20:47241619-47241677	4,90
55384	MEG3	+	chr14:101327186-101327244	4,90
6696	SPP1	+	chr4:88903920-88903979	4,88
6696	SPP1	+	chr4:88903920-88903979	4,88
6696	SPP1	+	chr4:88903920-88903979	4,84
6696	SPP1	+	chr4:88903920-88903979	4,84
6696	SPP1	+	chr4:88903920-88903979	4,82
6696	SPP1	+	chr4:88903920-88903979	4,78
6696	SPP1	+	chr4:88903920-88903979	4,76
9084	VCY	-	chrY:16097713-16097772	4,76
100131187	TSTD1	-	chr1:161007806-161007865	4,76
221303	FAM162B	-	chr6:117073387-117073445	4,76
6696	SPP1	+	chr4:88903920-88903979	4,74
115908	CTHRC1	+	chr8:104394743-104394802	4,71
144347	FAM101A	+	chr12:124800459-124800518	4,69
3914	LAMB3	-	chr1:209788278-209788337	4,63

100131187	TSTD1	-	chr1:161007764-161007823	4,54
93953	ACRC	+	chrX:70817800-70817859	4,51
3918	LAMC2	+	chr1:183213611-183213669	4,47
26136	TES	+	chr7:115897456-115897515	4,38
6591	SNAI2	-	chr8:49830710-49830769	4,35
55655	NLRP2	+	chr19:55505674-55505733	4,34
54742	LY6K	+	chr8:143784626-143784685	4,30
81831	NETO2	-	chr16:47143520-47143579	4,30
3624	INHBA	-	chr7:41729246-41729305	4,29
54742	LY6K	+	chr8:143784821-143784880	4,28
23209	MLC1	-	chr22:50497948-50498007	4,26
6364	CCL20	+	chr2:228680170-228680229	4,23
59345	GNB4	-	chr3:179118744-179118803	4,22

**Apêndice 8 – LncRNAs intrônicos antissenso diferencialmente expressos após tratamento da linhagem celular Mia PaCa 2 com o agente desmetilante 5-AZA.**

**50 lncRNAs antissenso com maior diferença de expressão após tratamento com 5 µM de 5-AZA (SAM; q-valor < 10%)**

<i>Antisense</i>					
<i>lncRNA Contig Name</i>	<i>Antisense lncRNA Contig Coordinate</i>	<i>Strand</i>	<i>Host Gene ID</i>	<i>Host Gene Symbol</i>	<i>Fold Change (Log2(AZA/CTL))</i>
TIN_18020	chr15:71581150-71581677	-	79875	THSD4	4,0
PIN_00785	chr1:151823237-151824027	+	284486	THEM5	4,0
PIN_10799	chr6:160100331-160103790	+	6648	SOD2	3,4
TIN_31722	chr21:27259582-27260205	+	351	APP	2,5
TIN_41328	chr5:95271087-95273124	+	22936	ELL2	2,1
PIN_07632	chr2:228680078-228680392	-	6364	CCL20	1,8
TIN_42527	chr5:167935627-167936461	-	5917	RARS	1,7
PIN_00053	chr1:7870082-7871274	-	8863	PER3	1,7
PIN_06554	chr19:47110760-47111729	-	808	CALM3	1,7
TIN_41924	chr5:134909106-134909411	+	9547	CXCL14	1,6
PIN_02990	chr12:69217974-69218538	-	4193	MDM2	1,6
PIN_01112	chr1:209795716-209796292	+	3914	LAMB3	1,6
PIN_06780	chr2:24303616-24303902	+	9540	TP53I3	1,5
PIN_03907	chr14:94848728-94850103	+	5265	SERPINA1	1,4
PIN_11366	chr7:150814865-150816037	-	116988	AGAP3	1,4
PIN_10480	chr6:42976986-42977385	-	5528	PPP2R5D	1,4
PIN_02929	chr12:57693659-57694043	+	22864	R3HDM2	1,3
PIN_02118	chr11:63403563-63403889	+	25923	ATL3	1,3
TIN_47306	chr7:80511635-80512110	+	10512	SEMA3C	1,2
PIN_09078	chr3:167405113-167405719	+	11235	PDCD10	1,2
PIN_00910	chr1:160260007-160260614	+	1314	COPA	1,2
PIN_01113	chr1:209807434-209808025	+	3914	LAMB3	1,2
PIN_11713	chr8:124037120-124037626	+	79139	DERL1	1,2
TIN_41515	chr5:112229030-112229698	+	7905	REEP5	1,2
PIN_10140	chr5:167921413-167922008	-	5917	RARS	1,2
PIN_09901	chr5:95241465-95242937	+	22936	ELL2	1,2
TIN_40609	chr5:54750290-54751056	+	8611	PPAP2A	1,2
TIN_45015	chr6:135713217-135714357	+	54806	AHI1	1,2
PIN_01142	chr1:223976260-223976917	+	7159	TP53BP2	1,2
PIN_09633	chr4:177248107-177248532	-	60559	SPCS3	1,1
TIN_12598	chr12:96651104-96651827	-	2004	ELK3	1,1
PIN_02561	chr12:6344234-6344619	-	928	CD9	1,1
PIN_05891	chr18:10547966-10548637	-	8774	NAPG	1,1
PIN_11579	chr8:63929671-63930526	+	8836	GGH	1,1
PIN_08177	chr21:46702114-46702544	+	23275	POFUT2	1,1

TIN_17512	chr15:58432414-58433762	-	366	AQP9	1,1
PIN_07027	chr2:70130184-70130539	-	11017	SNRNP27	1,1
PIN_03866	chr14:88935507-88936694	+	11099	PTPN21	1,1
TIN_28281	chr2:135623905-135624671	-	130013	ACMSD	1,1
PIN_09634	chr4:178272418-178273014	-	55247	NEIL3	1,1
TIN_06469	chr10:71170006-71170753	+	6865	TACR2	1,1
PIN_03893	chr14:93170707-93171352	+	5641	LGMN	1,1
PIN_03752	chr14:65216419-65217246	+	6710	SPTB	1,1
PIN_00096	chr1:15792053-15793880	-	63036	CELA2A	1,1
PIN_09521	chr4:110603490-110603958	-	55013	CCDC109B	1,0
TIN_45929	chr7:6209199-6210147	+	9265	CYTH3	1,0
PIN_10950	chr7:39609605-39610601	-	57002	YAE1D1	1,0
TIN_33678	chr3:19993246-19994751	-	5868	RAB5A	1,0
PIN_02291	chr11:71946304-71946889	-	3636	INPPL1	1,0
PIN_06434	chr19:39329239-39330362	+	3191	HNRNPL	1,0
PIN_04647	chr16:16225338-16226068	-	4363	ABCC1	1,0

**Apêndice 9 – LncRNAs intrônicos senso diferencialmente expressos após tratamento da linhagem celular Mia PaCa 2 com o agente desmetilante 5-AZA.**

**50 lncRNAs senso com maior diferença de expressão após tratamento com 5 µM de 5-AZA (SAM; q-valor < 10%)**

<i>Sense lncRNA Contig Name</i>	<i>Sense lncRNA Contig Coordinate</i>	<i>Strand</i>	<i>Host Gene ID</i>	<i>Host Gene Symbol</i>	<i>Fold Change (Log2(AZA/CTL))</i>
TIN_03775	chr1:183210410-183211966	+	3918	LAMC2	3,6
TIN_00094	chr1:6516255-6516743	+	83715	ESPN	3,0
TIN_48936	chr7:151443522-151444019	-	51422	PRKAG2	2,6
TIN_41327	chr5:95269206-95270787	-	22936	ELL2	2,5
TIN_27745	chr2:100436049-100436619	-	3899	AFF3	2,3
TIN_16740	chr14:105910579-105911244	+	9112	MTA1	2,2
TIN_05240	chr10:11799409-11799901	+	79746	ECHDC3	2,1
TIN_36803	chr3:182973029-182974266	+	23101	MCF2L2	1,9
TIN_41326	chr5:95268325-95269198	-	22936	ELL2	1,9
TIN_05679	chr10:24782053-24782562	+	56243	KIAA1217	1,9
TIN_46812	chr7:55147478-55148063	+	1956	EGFR	1,7
TIN_24241	chr19:3663989-3664802	-	23396	PIP5K1C	1,7
TIN_41316	chr5:95240456-95241112	-	22936	ELL2	1,5
TIN_00951	chr1:33483244-33484105	-	204	AK2	1,5
TIN_41116	chr5:80395916-80397009	+	5924	RASGRF2	1,4
TIN_33402	chr3:11512317-11513055	+	10533	ATG7	1,4
TIN_43738	chr6:38102940-38103877	+	60685	ZFAND3	1,3
TIN_41730	chr5:127451569-127452294	+	6558	SLC12A2	1,3
TIN_37839	chr4:38664162-38665044	-	79667	FLJ13197	1,3
TIN_13422	chr12:125622311-	+	65985	AACS	1,2

	125623428				
TIN_19546	chr16:30720545-30720818	+	10847	SRCAP	1,2
TIN_45512	chr6:160107899-160108623	-	6648	SOD2	1,2
TIN_46338	chr7:30507857-30508363	-	10392	NOD1	1,2
TIN_52305	chr9:37826865-37827383	+	79269	DCAF10	1,1
TIN_49559	chr8:23387442-23388827	+	51312	SLC25A37	1,1
TIN_17260	chr15:50750120-50750721	+	9101	USP8	1,1
TIN_19866	chr16:67802142-67802887	-	57610	RANBP10	1,1
TIN_48026	chr7:107241982-107243030	+	55973	BCAP29	1,1
TIN_00136	chr1:7770093-7770663	+	23261	CAMTA1	1,1
TIN_04299	chr1:212190442-212190945	-	25896	INTS7	1,1
TIN_43811	chr6:41770230-41770795	-	25862	USP49	1,1
TIN_33393	chr3:11471022-11471769	+	10533	ATG7	1,0
TIN_33399	chr3:11500077-11500766	+	10533	ATG7	1,0
TIN_33403	chr3:11519374-11520601	+	10533	ATG7	1,0
TIN_05634	chr10:23272919-23273422	+	219681	ARMC3	1,0
TIN_19344	chr16:22089792-22090351	+	146174	C16orf52	1,0
TIN_49946	chr8:48765436-48765779	-	5591	PRKDC	1,0
TIN_19062	chr16:11658022-11658874	-	9516	LITAF	1,0
TIN_07400	chr10:106031138-106032014	+	119391	GSTO2	1,0
TIN_03666	chr1:178799558-178800033	+	55103	RALGPS2	1,0
TIN_33388	chr3:11393567-11394842	+	10533	ATG7	1,0
TIN_00491	chr1:20080217-20081008	-	255104	TMCO4	1,0
TIN_07324	chr10:104696512-104697584	+	54805	CNNM2	1,0
TIN_00340	chr1:12370152-12370709	+	55187	VPS13D	1,0

TIN_41625	chr5:118619418- 118620374	+	25816	TNFAIP8	1,0
TIN_21443	chr17:29674979- 29675752	+	4763	NF1	1,0
TIN_41047	chr5:78696493- 78697222	-	9456	HOMER1	1,0
TIN_01641	chr1:54832469- 54833965	-	23648	SSBP3	1,0
TIN_12799	chr12:105625771- 105626539	-	55198	APPL2	0,9
TIN_50505	chr8:95269166- 95269905	-	2669	GEM	0,9
TIN_12744	chr12:104851533- 104852135	+	50515	CHST11	0,9

## ANEXO

### SÚMULA CURRICULAR

#### Dados pessoais

Nome: Lauren Camargo

Nascimento: 05 de janeiro de 1982

Naturalidade: Santos/SP

Nacionalidade: Brasileira

Endereço: Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo

Av. Profº. Lineu Prestes, 748, sala 1200, 05508-000 São Paulo, SP. Fone (11) 3091-2173 ramal:  
209

e-mail: [lauren@iq.usp.br](mailto:lauren@iq.usp.br)

#### Formação Acadêmica:

- Bióloga - Bacharelado e Licenciatura em Ciências Biológicas no Instituto de Biociências da Universidade de São Paulo, sito à Rua do Matão – Travessa 14, nº 321- Cidade Universitária – Butantã – CEP: 05508-900 – São Paulo/SP. Período: 2001 a 2006.
- Doutorado em Ciências Biológicas (Bioquímica). Instituto de Química da Universidade de São Paulo, sito à Av. Professor Lineu Prestes, 748, Cidade Universitária – Butantã – CEP: 05508-000 – São Paulo/SP - Período: 2006 até o presente. Previsão de término: maio de 2012.

#### Histórico Profissional

05/2005-06/2006 Aluna de Iniciação Científica, Departamento de Bioquímica, Instituto de Química – Universidade de São Paulo.

Projeto: Análise da expressão gênica em tumores de pâncreas através de microarrays de DNA

Orientador: Prof<sup>o</sup> Dr. Eduardo Moraes Rego Reis

07/2006 – Atual Aluna de doutorado Direto, Departamento de Bioquímica, Instituto de Química  
– Universidade de São Paulo

Projeto: Estudo do efeito da metilação do DNA sobre a regulação da transcrição de uma nova classe de RNAs antissenso intrônicos não codificantes.

Orientador: Prof<sup>o</sup> Dr. Eduardo Moraes Rego Reis

Bolsista FAPESP: Vigência da Bolsa: 01/01/07 a 31/08/11

## **Publicações**

### **Artigos completos publicados em periódicos**

IGREJA, C.; FRAGOSO, R.; CAIADO, F.; CLODE, N.; HENRIQUES, A.; CAMARGO, L.; REIS, E.M.; DIAS, S. (2007). “Detailed molecular characterization of cord blood-derived endothelial progenitors.” Experimental Hematology **36**: 193.e1-193.e15.

### **Trabalhos apresentados na foram oral em congressos nacionais e internacionais**

OLIVEIRA, Ana Carolina Ayupe de; CAMARGO, Lauren; VERJOVSKI-ALMEIDA, Sergio; Reis, Eduardo Moraes, 2008, Cold Spring Harbor, NY. Abstract book 2008 meeting on Dynamic Organization of Nuclear Function, 2008. p. 7.

### **Resumos publicados em anais de congressos**

CAMARGO, L.; VERJOVSKI-ALMEIDA, S.; REIS, E. Abstract # 5195: Effect of DNA methylation on the transcription of intronic noncoding RNAs in cancer cell lines AACR Meeting Abstracts, Apr 2009; 2009: 5195.

CAMARGO, L.; MARACAJÁ-COUTINHO, V.; VERJOVSKI-ALMEIDA, S.; REIS, E. M. Effect of DNA methylation on the transcription of intronic noncoding RNAs in cancer cell lines, Cold Spring Harbor, NY. Abstract book 2010 meeting on Systems Biology: Global Regulation of Gene Expression, 2010. p. 29.

DE OCESANO, C.; CAMARGO, L.; OLIVEIRA, A.C.A.; PARREIRA, K.S.; FACHEL, A.A.; VERJOVSKI-ALMEIDA, S. A new ncRNA tumor suppressor candidate regulator of *Bcl-X* gene. In XLI Reunião Anual da SBBq, 2011, Foz do Iguaçu – PR.

### **Atividades extras**

Orientação da aluna de graduação Ana Carolina Buratto durante seu período de Iniciação Científica no laboratório do Prof<sup>o</sup> Dr. Eduardo Reis e sob sua supervisão, agosto de 2008 – agosto de 2009.

Orientação do aluno de graduação Breno Pimentel Sampaio durante seu período de Iniciação Científica no laboratório do Prof<sup>o</sup> Dr. Eduardo Reis e sob sua supervisão, maio de 2010 – até o presente.

Monitoria no II Curso de Inverno “Temas Avançados de Bioquímica e Biologia Molecular” – Departamento de Bioquímica do Instituto de Química da USP, 16-27 de julho de 2007.

Monitoria no III Curso de Inverno “Temas Avançados de Bioquímica e Biologia Molecular” – Departamento de Bioquímica do Instituto de Química da USP, 21 de julho a 01 de agosto de 2008.