

**UNIVERSIDADE DE SÃO PAULO**

**INSTITUTO DE QUÍMICA**

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

**GUSTAVO STARVAGGI FRANÇA**

**Estudos sobre a organização genômica, evolução e  
expressão de microRNAs**

Versão Corrigida da Tese  
Conforme Resolução CoPGr 5890

A versão original encontra-se disponível na  
Secretaria de Pós-Graduação do IQ - USP

**São Paulo**

**Data de Depósito na SPG:**

**09/09/2015**

**GUSTAVO STARVAGGI FRANÇA**

**Estudos sobre a organização genômica, evolução e  
expressão de microRNAs**

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

Orientadora: Dra. Anamaria Aranha Camargo

Co-orientador: Dr. Pedro A. F. Galante

**São Paulo**

**2015**

**Ficha Catalográfica**

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

França, Gustavo Starvaggi

F814e Estudos sobre a organização genômica, evolução e expressão  
de microRNAs / Gustavo Starvaggi França -- São Paulo, 2015.  
145p.

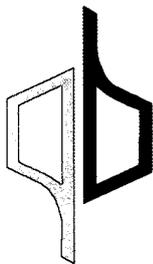
Tese (doutorado) - Instituto de Química da Universidade de  
São Paulo. Departamento de Bioquímica

Orientador : Camargo, Anamaria Aranha

Co-orientador : Galante, Pedro Alexandre Favoretto

1. Expressão gênica 2. Regulação gênica I. T. II. Camargo,  
Anamaria Aranha, orientador. III. Galante, Pedro Alexandre Favoretto

574.88 CDD



UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE QUÍMICA

**“Estudos sobre a organização genômica, evolução e expressão de microRNAs”**

*GUSTAVO STARVAGGI FRANÇA*

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências no Programa de Ciências Biológicas (Bioquímica) - Área de Concentração: Bioquímica.

Aprovado (a) por:

Prof. Dr. Anamaria Aranha Camargo  
(Orientadora e Presidente)

Prof. Dr. João Carlos Setubal  
IQ - USP

Prof. Dr. Eduardo Moraes Rego Reis  
IQ - USP

Prof. Dr. Diogo Meyer  
IB - USP

Prof. Dr. Helder Takashi Imoto Nakaya  
FCF - USP

SÃO PAULO

13 de novembro de 2015

## AGRADECIMENTOS

Agradeço sinceramente aos meus orientadores, Dra. Anamaria A. Camargo e Dr. Pedro A. F. Galante, por me acolherem em seus grupos de pesquisa, oferecendo os meios necessários para a realização deste trabalho.

Dentre inúmeros motivos, sou grato ao Pedro pela orientação, confiança depositada, incentivo, pelos ensinamentos científicos e pessoais, e por me expor sempre que possível aos bastidores da atividade científica, visando meu amadurecimento como pesquisador. Sobretudo, agradeço o apoio nos momentos decisivos e a liberdade concedida, permitindo que eu desenvolvesse minhas próprias idéias. Tudo isso contribuiu de maneira imensurável na minha formação.

Agradeço à Anamaria pela orientação, apoio, liberdade e pela forma acolhedora com que sempre me recebeu. A sua admirável busca pela excelência, disponibilidade em ajudar, e a maneira harmoniosa de lidar com questões científicas e pessoais me serviram de enorme aprendizado.

Agradeço à Prof. Dra. Maria Vibranovski pela oportunidade de colaborar em um projeto que, em princípio foi concebido como um estudo paralelo ao tema proposto para o meu Doutorado, mas que acabou transformando-se em uma das principais contribuições desta Tese. Agradeço a sua orientação precisa, incansável disponibilidade, rigor científico, sinceridade e por sempre oferecer pontos de vista positivos diante das dificuldades, transmitindo-os com entusiasmo e fazendo com que eu acreditasse mais no meu potencial.

Ao Prof. Dr. Diogo Meyer, pela oportunidade de trabalhar em seu laboratório no período antecedente ao Doutorado, por possibilitar a participação em sua disciplina e pelo exemplo de como abordar analiticamente problemas complexos em Evolução.

Ao Prof. Dr. Luiz Jorge Fagundes, pelo estímulo crucial num momento de descrença, sendo para mim um exemplo inspirador de dedicação ao trabalho, competência e conduta ética.

Ao Dr. Ludwig Christian Hinske, pela colaboração que resultou e um dos trabalhos apresentados nesta Tese.

À Prof. Dra. Gabriela Placoná Diniz e Elaine Guedes, pela oportunidade de colaboração.

Aos Drs. Érico T. Costa, Lilian T. Inoue e Luiza F. Andrade pela realização dos experimentos de validação funcional incluídos neste trabalho.

Aos Drs. Raphael Parmigiani e Camila Lopes, que conduziram o sequenciamento das linhagens celulares utilizadas neste trabalho.

Ao Prof. Dr. Manyuan Long (University of Chicago) e Dr. Yong E. Zhang (Institute of Zoology - Chinese Academy of Sciences) pelas sugestões e correções feitas no trabalho de evolução de microRNAs.

Aos Profs. Drs. Eduardo Reis, Carlos Hotta e Fábio Forti pelas contribuições no exame de qualificação.

Aos Profs. Drs. Pio Colepicolo e Daniela Bassères pela orientação durante o período de monitoria.

Ao Andrei Rozanski, pela amizade verdadeira com a qual pude sempre contar; pela sabedoria compartilhada, conversas permeadas de bom humor, incentivo e ótimos momentos vividos.

Ao Fábio Navarro, pela amizade, discussões intermináveis sobre o mundo acadêmico, pizzas e idéias empolgantes.

Ao Daniel Ohara, pelo excelente suporte técnico, paciência e disposição.

Ao Hugo A. M. Torres, pela constante ajuda “técnico-filosófica” na construção do miRIAD.

Ao Luis Felipe Campesato, pela amizade, churrascos e boas conversas.

Aos antigos e atuais membros do grupo de Bioinformática: Andrei, Bruna, Cibele, Daniel, Elisa, Fábio, Hugo e Rubens, pelas discussões científicas, ajudas e agradável convivência.

Aos ex- e atuais membros do Laboratório de Genômica Evolutiva: Bárbara, Rodrigo Francisco, Kelly, Márcia, Rodrigo Ramalho, Fábio Mendes e Débora.

Ao Instituto Ludwig de Pesquisa sobre o Câncer e seus funcionários, em especial à Renata e Viviani, competentes e prestativas nos assuntos administrativos.

Ao Instituto de Ensino e Pesquisa (IEP) do Hospital Sírio-Libanês pela infra-estrutura disponibilizada.

À Universidade de São Paulo, aos Profs. e funcionários do Instituto de Química e à secretaria de Pós-Graduação (Cibele, Milton, Marcelo e Emiliano).

À CAPES pelo apoio financeiro inicial.

À FAPESP pelas bolsas concedidas, financiamento e excelentes oportunidades oferecidas.

Por fim, sou imensamente grato à minha família. Ao meu tio Roberto, por seu altruísmo, mostrando-me caminhos alternativos. Aos meus tios e primos queridos. E, minha eterna gratidão aos meus pais Luiz e Anna Gina e ao meu irmão Guilherme, pelo apoio incondicional, compreensão e lições de vida. Obrigado por estarem ao meu lado!

*“A tarefa não é tanto ver aquilo que ninguém viu,  
mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.”*

Arthur Schopenhauer

## RESUMO

FRANÇA, G.S. **Estudos sobre a organização genômica, evolução e expressão de microRNAs**. 2015. 145p. Tese – Programa de Pós-Graduação em Ciências Biológicas (Bioquímica). Instituto de Química, Universidade de São Paulo, São Paulo.

Os microRNAs (miRNAs) são pequenos RNAs não codificadores de proteínas presentes na maioria dos eucariotos. Esses RNAs regulam a expressão gênica em nível pós-transcricional através do silenciamento de mRNAs-alvo que possuem sítios complementares às suas sequências, atuando em praticamente todos os processos celulares. Embora a estrutura e função dos miRNAs estejam bem caracterizadas, aspectos relacionados à sua organização genômica, evolução e atuação em doenças são tópicos que apresentam enormes lacunas. Nesta tese, utilizamos abordagens computacionais para investigar estes temas em três trabalhos. No primeiro, processamos e integramos um vasto volume de dados publicamente disponíveis referentes aos miRNAs e genes codificadores de proteínas para cinco espécies de vertebrados. Com isso, construímos uma ferramenta *web* que permite a fácil inspeção da organização genômica dos miRNAs em regiões inter e intragênicas, o acesso a dados de expressão de miRNAs e de genes codificadores de proteínas (classificados em genes hospedeiros e não hospedeiros de miRNAs), além de outras informações pertinentes. Verificamos que a ferramenta tem sido amplamente utilizada pela comunidade científica e acreditamos que ela possa facilitar a geração de hipóteses associadas à regulação dos miRNAs, principalmente quando estão inseridos em genes hospedeiros. No segundo estudo, buscamos compreender como o contexto genômico e a origem evolutiva dos genes hospedeiros influenciam a expressão e evolução dos miRNAs humanos. Nossos achados mostraram que os miRNAs intragênicos surgem preferencialmente em genes antigos (origem anterior à divergência de vertebrados). Observamos que os miRNAs inseridos em genes antigos têm maior abrangência de expressão do que os inseridos em genes novos. Surpreendentemente, miRNAs jovens localizados em genes antigos são expressos em um maior número de tecidos do que os intergênicos de mesma idade, sugerindo uma vantagem adaptativa inicial que pode estar relacionada com o controle da expressão dos genes hospedeiros, e como consequência, expondo-os a contextos celulares e conjuntos de alvos diversos. Na evolução a longo prazo, vimos que genes antigos conferem maior restrição nos padrões de expressão (menor divergência de expressão) para miRNAs intragênicos, quando comparados aos intergênicos. Também mostramos possíveis associações funcionais relacionadas ao contexto genômico, tais como o enriquecimento da expressão de miRNAs intergênicos em testículo e dos intragênicos em tecidos neurais. Propomos que o contexto genômico e a idade dos genes hospedeiros são fatores-chave para a evolução e expressão dos miRNAs. Por fim, buscamos estabelecer associações entre a expressão diferencial de miRNAs e a quimioresistência em câncer colorretal utilizando linhagens celulares sensíveis e resistentes às drogas 5-Fluoruracil e Oxaliplatina. Dentre os miRNAs identificados, o miR-342 apresentou níveis elevados de expressão nas linhagens sensíveis à Oxaliplatina. Com base na análise dos alvos preditos, detectamos uma significativa associação de miR-342 com a apoptose. A superexpressão de miR-342 na linhagem resistente SW620 evidenciou alterações na expressão de genes da via apoptótica, notavelmente a diminuição da expressão do fator de crescimento PDGFB, um alvo predito possivelmente sujeito à regulação direta pelo miR-342.

**Palavras-chave:** microRNAs, bioinformática, evolução molecular, expressão gênica, câncer colorretal

## ABSTRACT

FRANÇA, G.S. **Studies on the genomic organization, evolution and expression of microRNAs.** 2015. 145p. Tese – Programa de Pós-Graduação em Ciências Biológicas (Bioquímica). Instituto de Química, Universidade de São Paulo, São Paulo.

MicroRNAs (miRNAs) are short non-coding RNAs found in most eukaryotic species. These RNAs regulate gene expression at post-transcriptional level by silencing target mRNAs through base-pairing of complementary sequences, thus acting on virtually all cellular processes. Although the structure and function of miRNAs are well understood, several aspects related to their genomic organization, evolution and involvement with diseases are largely underexplored. In this thesis, we employed computational methods to investigate such issues in three different studies. In the first one, we have processed and integrated a large amount of public data related to miRNAs and coding genes for five vertebrate species. Then, we developed a webtool to allow the analysis of the miRNA genomic context in inter and intragenic regions, the access of miRNA and gene expression data (classified as host and non-host genes), as well as other relevant information. We noticed that the webtool has been largely used by the scientific community, and we believe that it can facilitate hypothesis generation related to miRNA regulation, especially when they are within host genes. In the following study, we sought to understand how the genomic context and the evolutionary origin of host genes can affect the expression and evolution of human miRNAs. Our findings showed that intragenic miRNAs are preferentially embedded within old host genes (originated before the split of vertebrates). We observed that miRNAs within old genes are more broadly expressed than those within young genes. Surprisingly, young miRNAs within old genes were expressed in more tissues than their intergenic counterparts, suggesting an initial adaptive advantage which might be related to their hosts' expression control, and as a consequence, exposing them to a more diverse cellular contexts and target genes. In the long run, we found that old host genes lead to expression constraints (lower expression divergence) between species for intragenic miRNAs, in respect to intergenic ones. We also showed possible functional associations related to miRNA genomic context, such as the enrichment of young intergenic miRNAs in testis, while young intragenic miRNAs were enriched in neural tissues. Thus, we propose that the genomic context and the age of the host genes are key factors in shaping the expression and evolution of miRNAs. Finally, we sought to establish associations between differential expression of miRNAs and chemoresistance in colorectal cancer using resistant and sensitive cell lines to 5-Fluoruracil and Oxaliplatin. Among differentially expressed miRNAs, miR-342 was highly expressed in sensitive cell lines to Oxaliplatin. Based on target prediction analysis, miR-342 is likely associated with apoptosis. The induced overexpression of miR-342 in SW620, a cell line resistant to Oxaliplatin, changed the expression levels of genes linked to the apoptosis pathway, notably the downregulation of PDGFB growth factor, which is a predicted target possibly subjected to direct regulation by miR-342.

**Keywords:** microRNAs, bioinformatics, molecular evolution, gene expression, colorectal cancer

## LISTA DE FIGURAS

<b>Figura 1.</b> Biogênese dos miRNAs. ....	18
<b>Figura 2.</b> Estrutura do pre-miRNA. ....	19
<b>Figura 3.</b> Organização genômica dos miRNAs. ....	21
<b>Figura 4.</b> Identificação de miRNAs. ....	25
<b>Figura 5.</b> Exemplos de interação miRNA-alvo e mecanismo de silenciamento. ....	28
<b>Figura 6.</b> Principais mecanismos de surgimento de novos miRNAs. ....	31
<b>Figura 7.</b> Esquema representativo do banco de dados que constitui a ferramenta miRIAD. ...	51
<b>Figura 8.</b> Estrutura da ferramenta miRIAD e como as camadas de informação estão organizadas. ....	65
<b>Figura 9.</b> Página inicial da aplicação web. ....	66
<b>Figura 10.</b> Sumário das principais informações apresentadas para o gene ERBB2 e seu miRNA intragênico mir-4728. ....	69
<b>Figura 11.</b> Sumário das principais informações apresentadas na ferramenta para o mir-483 e seu gene hospedeiro IGF2. ....	71
<b>Figura 12.</b> Representação do gene CLCN5 humano e seus miRNAs intragênicos. ....	73
<b>Figura 13.</b> Representação do gene Clcn5 em camundongo e seus miRNAs intragênicos. ....	74
<b>Figura 14.</b> Representação do gene HTR2C humano e seus miRNAs intragênicos. ....	75
<b>Figura 15.</b> Estatísticas de acesso da ferramenta miRIAD. ....	78
<b>Figura 16.</b> Idade dos miRNAs e genes hospedeiros humanos. ....	82
<b>Figura 17.</b> Abrangência de expressão de miRNAs de acordo com a idade e o contexto genômico. ....	84
<b>Figura 18.</b> Controles relacionados à análise de abrangência de expressão de miRNAs. ....	87
<b>Figura 19.</b> Nível de expressão dos miRNAs de acordo com a idade e o contexto genômico. ....	87
<b>Figura 20.</b> Co-expressão de miRNAs intragênicos e genes hospedeiros. ....	88
<b>Figura 21.</b> Expressão de miRNAs intergênicos em relação aos genes mais próximos. ....	90
<b>Figura 22.</b> Abrangência de expressão de miRNAs intergênicos e genes vizinhos. ....	91
<b>Figura 23.</b> Análise da expressão interespecífica de miRNAs. ....	93
<b>Figura 24.</b> Conservação de sequência dos miRNAs humanos. ....	94
<b>Figura 25.</b> Associações funcionais preditas in silico para miRNAs inter e intragênicos. ....	97
<b>Figura 26.</b> Heatmap dos valores de expressão normalizados dos miRNAs diferencialmente expressos. ....	100
<b>Figura 27.</b> Expressão de miR-342 e EVL por qRT-PCR. ....	101
<b>Figura 28.</b> Análise funcional dos alvos dos miRNAs diferencialmente expressos. ....	102
<b>Figura 29.</b> Ensaio de viabilidade celular por exclusão de iodeto de propídeo (PI) seguida de análise por citometria de fluxo. ....	104
<b>Figura 30.</b> Ensaio clonogênico bidimensional. ....	104
<b>Figura 31.</b> Análise da progressão do ciclo celular por marcação com PI. ....	105
<b>Figura 32.</b> Ensaios de ciclo celular e clonogênico bidimensional em células HCT116 transfectadas com miR-342. ....	107
<b>Figura 33.</b> Ensaios de ciclo celular e clonogênico bidimensional em células SW620 transfectadas com miR-342. ....	108
<b>Figura 34.</b> Genes da via apoptótica com expressão alterada após transfecção com miR-342 na linhagem SW620. ....	109

## LISTA DE TABELAS

<b>Tabela 1.</b> Sumário dos dados apresentados na ferramenta miRIAD. ....	66
<b>Tabela 2.</b> Estatísticas gerais do mapeamento dos RNAs pequenos no genoma humano e quantidade de sequências mapeadas em miRNAs maduros. ....	98

## SUMÁRIO

<b>1 Introdução</b> .....	<b>15</b>
1.1 MicroRNAs: Biogênese e estrutura .....	16
1.2 Contexto genômico dos miRNAs .....	19
1.3 Identificação de miRNAs pelo método de RNA-Seq .....	22
1.4 Reconhecimento de alvos e mecanismos de silenciamento .....	25
1.5 Evolução de miRNAs .....	30
1.6 MiRNAs e câncer.....	32
<b>2 Objetivos</b> .....	<b>34</b>
2.1 Objetivos gerais .....	35
2.2 Objetivos específicos .....	35
<b>3 Materiais e Métodos</b> .....	<b>36</b>
3.1 Dados públicos utilizados .....	37
3.1.1 Genomas de referência .....	37
3.1.2 Anotações de miRNAs .....	37
3.1.3 Anotações de genes codificadores de proteínas .....	37
3.1.4 Dados de expressão de miRNAs .....	38
3.1.5 Dados de expressão de genes codificantes de proteínas .....	38
3.2 Expressão de miRNAs e de genes codificadores de proteínas .....	39
3.2.1 Quantificação da expressão de miRNAs a partir de dados de RNA-Seq.....	39
3.2.2 Expressão de genes codificadores de proteínas a partir de dados de RNA-Seq ...	41
3.3 Métodos utilizados para análises evolutivas dos miRNAs humanos.....	41
3.3.1 Anotação dos miRNAs humanos de acordo com a posição genômica e anotação de genes hospedeiros.....	41
3.3.2 Definição das idades dos miRNAs.....	42
3.3.3 Definição das idades dos genes codificantes de proteínas .....	43
3.3.4 Análise estatística das idades dos genes hospedeiros.....	44
3.3.5 Análise de Ka/Ks.....	44
3.3.6 Abrangência de expressão de miRNAs e genes codificantes.....	45
3.3.7 Correlação da expressão entre miRNAs intragênicos e seus genes hospedeiros ..	45
3.3.8 Divergência de expressão de miRNAs entre espécies .....	46
3.3.9 Análise de conservação de sequências de miRNAs .....	47
3.3.10 Predição de alvos dos miRNAs.....	48
3.3.11 MiRNAs associados com doenças .....	48
3.3.12 Análise de enriquecimento funcional dos genes hospedeiros .....	49

3.4	Métodos utilizados para construção da ferramenta <i>web</i> miRIAD .....	49
3.4.1	Dados brutos.....	49
3.4.2	Arquitetura do banco de dados e interface web .....	50
3.4.3	Informações sobre genes hospedeiros e miRNAs .....	52
3.4.4	Alvos preditos .....	52
3.5	Métodos utilizados para análise de expressão diferencial de miRNAs em linhagens celulares de câncer colorretal.....	53
3.5.1	Sequenciamento dos RNAs pequenos em linhagens celulares de tumor colorretal .....	53
3.5.2	Seleção de linhagens celulares de câncer colorretal resistentes e sensíveis ao tratamento com drogas quimioterápicas.....	54
3.5.3	Expressão diferencial de miRNAs .....	54
3.5.4	Identificação de genes-alvo preditos para os miRNAs diferencialmente expressos .....	55
3.5.5	Análises dos alvos dos miRNAs diferencialmente expressos.....	56
3.6	Validação experimental da expressão de miR-342 e efeitos fenotípicos em linhagens resistentes e sensíveis à Oxaliplatina .....	56
3.6.1	Cultivo celular.....	57
3.6.2	Ensaio de viabilidade celular por exclusão de iodeto de propídeo .....	57
3.6.3	Ensaio de viabilidade celular por formação de colônias .....	57
3.6.4	Análise da progressão no ciclo celular por citometria de fluxo .....	58
3.6.5	Transfecções de oligonucleotídeos .....	59
3.6.6	Extração de RNA, miRNA e síntese de cDNA.....	59
3.6.7	Análise de expressão gênica pela metodologia de PCR em tempo real.....	60
3.6.8	Detecção da expressão de genes relacionados com vias apoptóticas.....	61
3.7	Infraestrutura do Laboratório de Bioinformática e linguagens de programação utilizadas .....	62
<b>4</b>	<b>Resultados .....</b>	<b>63</b>
4.1	miRIAD: uma ferramenta para a análise de miRNAs intragênicos.....	64
4.1.1	Características gerais da ferramenta.....	64
4.1.2	Sistema de busca da ferramenta miRIAD .....	67
4.1.3	Genes codificadores de proteínas .....	67
3.1.4	miRNAs intragênicos .....	69
3.1.5	Uso da ferramenta miRIAD para explorar genes e miRNAs .....	71
3.1.6	Estatísticas de acesso.....	75

4.2	Análise da influência dos genes hospedeiros e do contexto genômico na expressão e evolução de miRNAs .....	78
4.2.1	A maioria dos miRNAs intragênicos humanos são específicos de primatas e estão preferencialmente contidos em genes antigos.....	78
4.2.2	A idade dos genes hospedeiros e o contexto genômico afetam a expressão dos miRNAs.....	83
4.2.3	Evidências de que os miRNAs intragênicos antigos possuem expressão conservada.....	92
4.2.4	Associações funcionais entre os miRNAs e sua localização genômica .....	95
4.3	Análise da expressão diferencial de miRNAs e associações com a quimioresistência em câncer colorretal .....	98
4.3.1	Sequenciamento e identificação de microRNAs em linhagens celulares de câncer colorretal .....	98
4.3.2	Expressão diferencial de microRNAs em linhagens resistentes e sensíveis à drogas quimioterápicas .....	99
4.3.3	Enriquecimento de categorias funcionais para os alvos dos miRNAs diferencialmente expressos .....	102
4.3.4	Análise de viabilidade celular por exclusão de iodeto de propídeo e formação de colônias .....	103
4.3.5	Análise da progressão do ciclo celular .....	104
4.3.6	Efeitos da superexpressão de miR-342 na resposta ao tratamento com Oxaliplatina .....	105
4.3.7	Análise da expressão de genes da via apoptótica mediante superexpressão de miR-342.....	108
<b>5</b>	<b>Discussão.....</b>	<b>110</b>
5.1	Contribuição da ferramenta miRIAD para o estudo dos miRNAs intragênicos.....	111
5.2	A influência do contexto genômico e idade dos genes hospedeiros na expressão e evolução dos miRNAs humanos .....	113
5.3	MiRNAs envolvidos com a quimioresistência em câncer colorretal.....	118
<b>6</b>	<b>Conclusões .....</b>	<b>122</b>
	<b>Referências.....</b>	<b>125</b>
	<b>Apêndice .....</b>	<b>138</b>
	<b>Lista de Anexos.....</b>	<b>145</b>

## *1 Introdução*

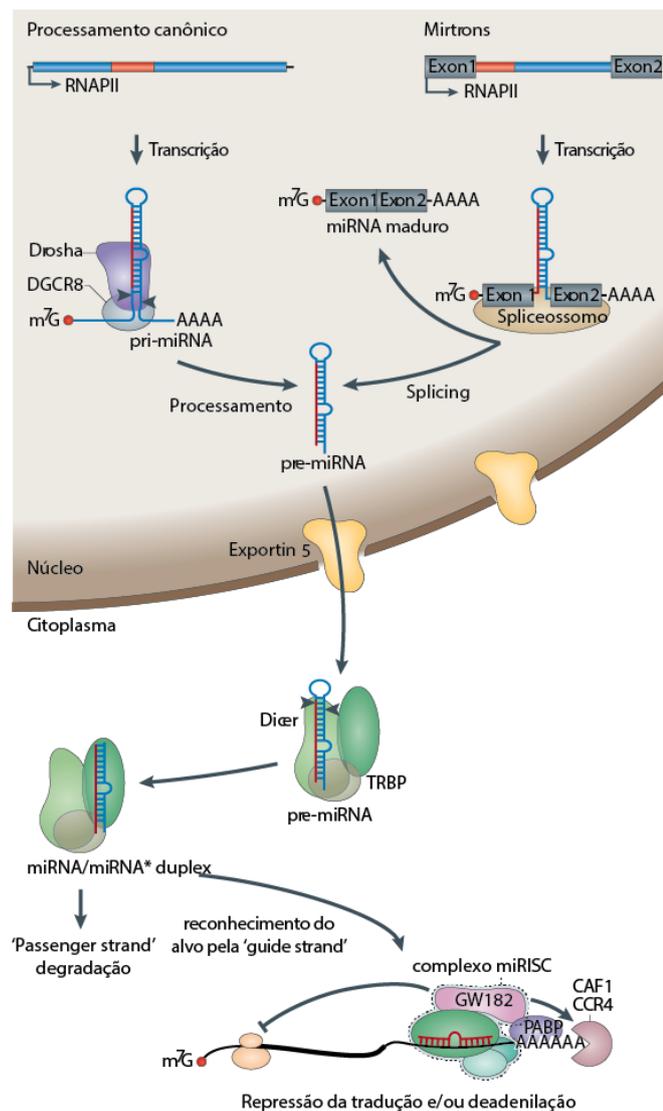
## 1.1 MicroRNAs: Biogênese e estrutura

O surgimento dos organismos multicelulares foi acompanhado de um notável aumento da complexidade dos mecanismos de regulação da expressão gênica, incluindo novos mecanismos de regulação pós-transcricional (Levine e Tjian 2003). A descoberta dos microRNAs (miRNAs) teve um profundo impacto na compreensão desse tema. Esta classe de RNAs não codificadores foi descoberta quando Lee *et al.* (1993) observaram que o gene *lin-4*, cuja perda de função resulta em um descompassamento da diferenciação celular durante o desenvolvimento larval de *C. elegans*, não codificava para uma proteína, mas para um RNA com tamanho aproximado de 61 nt (nucleotídeos) e outro de 22 nt. O RNA mais longo era capaz de formar uma estrutura predita em forma de *hairpin*, proposta como precursor do RNA mais curto. Estudos anteriores (Ambros e Horvitz 1987) já indicavam que o *lin-4* atua como regulador negativo de *lin-14*, um gene codificador de proteína também envolvido com a regulação temporal do desenvolvimento de *C. elegans*. Tanto Lee *et al.* (1993) quanto Wightman *et al.* (1993) verificaram que o mRNA de *lin-14* possuía sete sítios complementares aos transcritos de *lin-4* na sua região 3'UTR, mostrando que a diminuição dos níveis da proteína LIN-14 dependia da regulação por *lin-4*, especificamente quando os sítios complementares à *lin-4* estavam presentes na região 3'UTR em construções repórteres. Em conjunto, ambos os estudos foram pioneiros em sugerir um novo modelo de regulação pós-transcricional em que a expressão de *lin-14* é inibida pelo pareamento antisenso de *lin-4*. Esses estudos marcaram a descoberta de uma nova classe de RNAs curtos, posteriormente denominados de microRNAs (Lagos-Quintana *et al.* 2001), e que hoje constitui uma das maiores classes gênicas em plantas e animais (Pasquinelli 2012).

Os miRNAs são transcritos pela RNA polimerase II em longos RNAs dupla-fita primários (pri-miRNAs) cujo tamanho varia de centenas à milhares de bases (kb), contendo *cap* 5' 7-metilguanossina e cauda poly(A) (Cai *et al.* 2004; Lee *et al.* 2004). Na via de biogênese canônica,

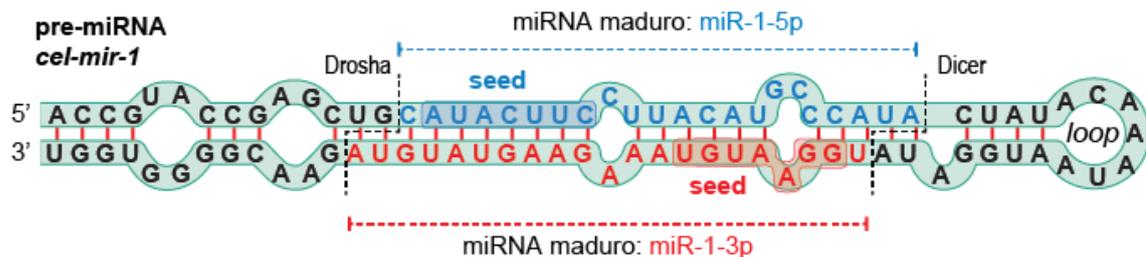
o pri-miRNA sofre duas etapas principais de processamento. A primeira é a clivagem no núcleo celular pela enzima RNase III Droscha, que converte o pri-miRNA em uma molécula precursora (pre-miRNA) de ~70 nt e esta forma uma estrutura secundária em *hairpin* (ou *stem-loop*), podendo excisar um ou mais pre-miRNAs (quando estes encontram-se agrupados em *clusters*). A Droscha faz parte de um complexo que contém uma proteína de ligação RNA dupla-fita (dsRNA) chamada DGCR8 em mamíferos, cujo papel está relacionado com o aumento da afinidade da RNase III pelo substrato, assegurando a precisão da clivagem (Denli *et al.* 2004). A clivagem pela Droscha deixa uma marca característica no RNA, que são dois nucleotídeos excedentes na extremidade 3'. O transportador nuclear Exportin 5 reconhece essa extremidade característica do pre-miRNA, facilitando o transporte para o citoplasma (Bohnsack *et al.* 2004). No citoplasma ocorre uma segunda etapa de processamento, dessa vez o *loop* terminal do pre-miRNA é clivado por outra RNase III, Dicer, produzindo o miRNA maduro dupla-fita de 18-24 nt (Figura 1). Da mesma maneira que a Droscha, a enzima Dicer interage com outra proteína de ligação dsRNA, chamada TRBP em mamíferos, que auxilia na estabilização e no posicionamento correto da clivagem (Hutvágner *et al.* 2001). O miRNA maduro é então incorporado no complexo RISC (*RNA-induced silencing complex*) contendo a proteína AGO2 (em mamíferos). A AGO2 é responsável pela interação direta com o miRNA, promovendo a acomodação estrutural da molécula no complexo de silenciamento. Quando o miRNA é incorporado ao RISC, uma de suas fitas é selecionada como miRNA maduro (*guide strand*) e mantida no complexo, enquanto a outra fita (*passenger strand*) é posteriormente degradada. Em geral, a fita selecionada é aquela que apresenta menor estabilidade termodinâmica na porção terminal 5', associada a um viés para presença de Us e purinas (Khvorova *et al.* 2003). A fita selecionada é responsável por guiar o complexo RISC para mRNAs-alvo contendo sítios complementares à sequência do miRNA maduro, cumprindo assim sua função no silenciamento da expressão gênica por promover a degradação do mRNA-alvo e/ou a inibição da tradução (ver seção 1.4). As etapas principais do

processamento dos miRNAs mencionadas acima estão ilustradas na Figura 1, que ressalta ainda a via biogênica não canônica dos “mirtrons”, em que o pre-miRNA está contido no próprio intron de um gene (geralmente codificador) e passa a ser excisado pela maquinaria de *splicing*, uma etapa que independe do processamento da Droscha.



**Figura 1. Biogênese dos miRNAs.** As etapas da via canônica envolvem a transcrição de um miRNA primário pela RNA polimerase II (RNAPII) e que é processado pela Drosha-DGCR8, gerando o pre-miRNA. O pre-miRNA é exportado para o citoplasma onde sofre um novo processamento pela Dicer-TRBP gerando o miRNA maduro que guia o complexo RISC para RNAs alvo. O complexo induz a repressão da tradução ou deadenilação (pela CAF1-CCR4) e posterior degradação. A proteína GW182 é um elemento central do complexo, funcionando como scaffold para outras proteínas. Na via não canônica, miRNAs contidos em introns podem ser excisados pela maquinaria de *splicing* sem sofrer a primeira etapa de processamento. Figura modificada de Krol *et al.* (2010).

O pre-miRNA possui uma estrutura típica em formato de *hairpin* com tamanho aproximado de 60 a 80 nt. A porção central é mantida pelo pareamento complementar, muitas vezes imperfeito, do braço 5' com o braço 3'. A porção central do RNA contém a região madura com cerca de 22 nt, chamada miRNA-miRNA\* duplex, enquanto a porção terminal é constituída por um *loop* que é excisado pela Dicer. O miRNA-miRNA\* duplex possui 2 nt livres nas extremidades 3', marcas típicas do processamento pelas endonucleases Drosha e Dicer (Ambros *et al.* 2003). Os miRNAs maduros possuem uma região definida como *seed*, que se estende do segundo ao oitavo nucleotídeo da sequência, sendo determinante para o reconhecimento dos alvos (ver mais detalhes na seção 1.4) (Figura 2).



**Figura 2. Região do pri-miRNA contendo o pre-miRNA.** A figura ilustra a estrutura secundária em formato de *hairpin* do pre-miRNA, onde estão indicadas as regiões maduras e *seed*, e sítios de processamento da Drosha e Dicer. Modificado de Watson *et al.* (2013).

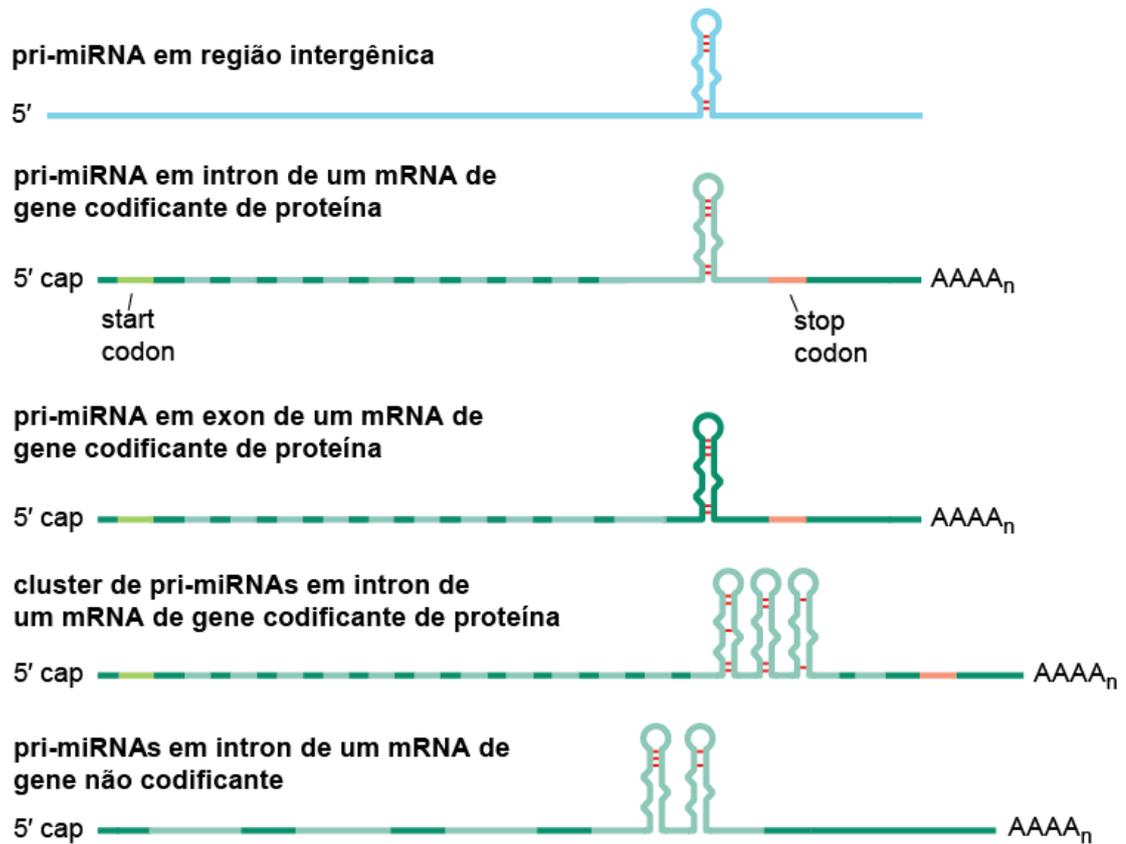
## 1.2 Contexto genômico dos miRNAs

Os miRNAs podem ser transcritos a partir de regiões intergênicas ou intragênicas, neste último caso, sobrepostos à regiões intrônicas ou exônicas de genes codificadores ou não-codificadores de proteínas (Figura 3). Ainda, os miRNAs apresentam-se como elementos únicos ou agrupados em *clusters* (Figura 3). Os membros de um *cluster* usualmente provém de uma origem evolutiva comum (seção 1.5) e em alguns casos podem atuar de forma coordenada na regulação de um mesmo processo biológico, como o cluster mir-183-96-192 em camundongo, cujos alvos estão envolvidos na via de sinalização da insulina (Xu e Wong 2008). Estima-se que

~35% dos miRNAs humanos estejam organizados em *clusters* (Altuvia *et al.* 2005), embora uma reanálise utilizando a anotação atual poderá indicar diferenças nesta estimativa.

Diversos trabalhos têm demonstrado um excesso significativo de miRNAs sobrepostos a genes codificadores de proteínas (Rodriguez *et al.* 2004; Campo-Paysaa *et al.* 2011; Hinske *et al.* 2010; Meunier *et al.* 2013). Este excesso é notadamente observado em espécies de vertebrados, em geral nas regiões intrônicas e com viés para a mesma orientação transcricional (orientação *sense*) dos seus genes hospedeiros (*host genes*) (Campo-Paysaa *et al.* 2011; Hinske *et al.* 2010). Cerca de 60% dos miRNAs anotados na espécie humana (miRBase v.20, Junho 2013) apresentam sobreposição com genes codificadores de proteínas, dos quais ~84% encontram-se na mesma orientação dos seus genes hospedeiros (Hinske *et al.* 2014).

O primeiro estudo a investigar a organização de miRNAs em regiões gênicas de modo sistemático revelou que dos 232 miRNAs anotados na época em humanos e camundongo, 161 (69%) estavam sobrepostos a regiões gênicas, sendo 40% (90) contidos em introns de genes codificadores de proteínas (Rodriguez *et al.* 2004). A presença de miRNAs predominantemente na mesma orientação transcricional dos genes hospedeiros indicam a possibilidade de co-transcrição. No trabalho de Rodriguez *et al.* (2004) essa hipótese foi testada para cinco casos, dos quais quatro apresentaram nítido padrão de co-expressão. Posteriormente, Baskerville e Bartel (2005) identificaram 16 casos adicionais de co-expressão ( $r > 0.4$ ) pela técnica de *MicroArray*. Essa questão foi abordada em maior escala por Ozsolak *et al.* (2008) que identificaram promotores associados à transcrição dos miRNAs e determinaram quais são potencialmente compartilhados com os genes hospedeiros. O estudo revelou que ~30% dos miRNAs intrônicos analisados são transcritos independentemente dos genes hospedeiros por promotores exclusivos, reforçando a idéia de que a co-regulação dos genes hospedeiros e miRNAs intragênicos é um evento comumente observado.



**Figura 3. Organização genômica dos miRNAs.** Representação das possíveis regiões genômicas a partir das quais os miRNAs podem ser transcritos. As linhas em verde claro representam regiões intrônicas, verde escuro exônicas e azul intergênicas. Modificado de Watson *et al.* (2013).

Associações funcionais e *feedbacks* regulatórios envolvendo miRNAs e genes hospedeiros ilustram como essa organização genômica pode ter um papel funcional importante em determinados fenótipos. O miR-26b (anotação atual: miR-26a-2), por exemplo, está contido no gene CTDSP2 e ambos são co-transcritos (Dill *et al.* 2012). Em um modelo de desenvolvimento de *zebrafish*, demonstrou-se que o processamento do mir-26b ocorre especificamente durante a diferenciação das células neuronais. Curiosamente, o miR-26b inibe diretamente a expressão do seu próprio gene hospedeiro. Como consequência, os níveis reduzidos da fosfatase CTDSP2 permitem que a RNA polimerase II transcreva genes de diferenciação neuronal em células-tronco neurais (Dill *et al.* 2012). Um outro exemplo é o papel cooperativo de miR-26a-1/2/b e seus respectivos genes hospedeiros CTDSPL, CTDSP2 e CTDSP1 na regulação do ciclo celular (Zhu *et al.* 2012). No caso do mir-483, localizado no segundo intron

do gene IGF2 (*Insulin-like growth factor 2*), um *feedback* regulatório positivo ocorre pelo pareamento incomum do miR-483-5p na região 5' UTR do mRNA de IGF2, o que favorece a associação da helicase DHX9, promovendo a transcrição do gene hospedeiro. O aumento da expressão de miR-483-5p e o consequente aumento de IGF2 está associado com a tumorigênese em modelos animais (Liu *et al.* 2013). Vários outros exemplos de interações envolvendo miRNAs e seus genes hospedeiros foram reportados (Barik 2008; Hinske *et al.* 2010; Kos *et al.* 2012), evidenciando a importância funcional desse tipo de organização genômica.

### 1.3 Identificação de miRNAs pelo método de RNA-Seq

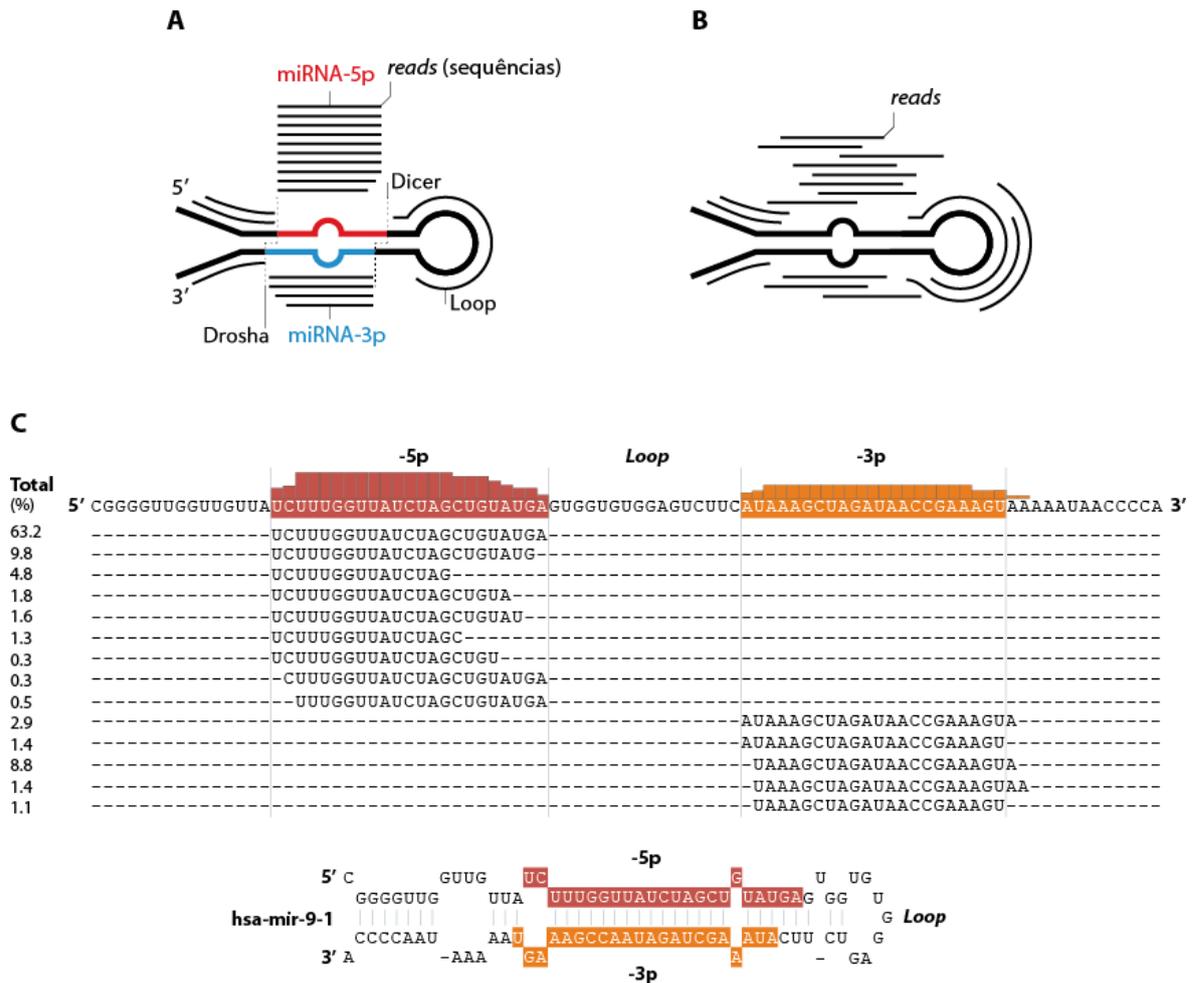
Inicialmente, a caracterização dos primeiros miRNAs ficou limitada às técnicas de genética direta (*forward genetics*) e clonagem, que impunham enormes dificuldades por causa do seu pequeno tamanho, efeitos fenotípicos muitas vezes discretos e expressão em condições ou tipos celulares específicos (Mendes *et al.* 2009). Este cenário mudou drasticamente com o desenvolvimento da tecnologia de RNA-Sequencing (RNA-Seq), que tem permitido a identificação e caracterização de novos transcritos a um baixo custo e alta eficiência. Graças a essa tecnologia, o repositório oficial de miRNAs, o miRBase (Kozomara e Griffiths-Jones 2011), tem sofrido um incremento notável nos últimos anos, contando atualmente (miRBase versão 21, Junho de 2014) com cerca de ~28.000 precursores anotados em um vasto conjunto de espécies. Na espécie humana por exemplo, há 1.881 precursores anotados e 2.588 miRNAs maduros, o que hoje constitui uma das maiores classes de genes do genoma humano.

A correta identificação dos miRNAs pelo método de RNA-Seq depende essencialmente das características intrínsecas que os distinguem de outros RNAs curtos. Um experimento de RNA-Seq usualmente gera dezenas ou centenas de milhões de sequências,

cuja localização genômica e posição na molécula precursora devem ser determinadas. Se as sequências geradas (*reads*) forem provenientes de miRNAs maduros, após o mapeamento contra o genoma de referência, as sequências mais abundantes devem conter entre 18-24 nt e apresentarem um padrão de mapeamento consistente com a estrutura da molécula precursora e com o processamento pela Drosha e Dicer, sendo predominante em um dos braços (5' ou 3') de miRNAs precursores (Figuras 4A e 4C). Além disso, *reads* provenientes da região madura geralmente apresentam distribuição uniforme em relação às extremidades correspondentes aos sítios de clivagem da Drosha e Dicer, em especial na extremidade 5' que define a região *seed*. Porém, em alguns casos pode haver uma pequena variação, resultado da clivagem em sítios distintos ou pela atividade de exonucleases e modificações pós-transcricionais (Pritchard *et al.* 2012) (Figuras 4A e 4C). Essas variantes são conhecidas como isomiRs. Por outro lado, se os *reads* não apresentarem este padrão característico, é improvável que sejam produto da clivagem precisa da Drosha e Dicer. Na prática, a distribuição dos *reads* ao longo do precursor seguirá um padrão próximo ao aleatório, por exemplo, e que possivelmente representam RNAs não processados completamente ou RNAs degradados (Figura 4B).

Após a identificação ou caracterização de novos miRNAs, o passo seguinte é a quantificação da expressão gênica, dada basicamente pelo número de *reads* que mapeiam nas respectivas regiões genômicas anotadas como miRNAs. Uma série de métodos de normalização que permitem a comparação da expressão de genes dentro e entre amostras foram desenvolvidos ao longo dos últimos anos, sendo que praticamente todos os métodos levam em conta como fator de normalização a quantidade de *reads* mapeados. Para miRNAs, dois dos métodos mais utilizados foram implementados nos pacotes DESeq (Anders e Huber 2010) e edgeR (Robinson *et al.* 2010). Ambos os métodos assumem que a maioria dos genes não são diferencialmente expressos. No primeiro método, o fator de normalização para cada

amostra (ex. tecido ou linhagem celular) é dado pela mediana da razão do número de *reads* correspondentes a cada gene sobre a média geométrica da contagem de reads do respectivo gene em todas as amostras (Seyednasrollah *et al.* 2013). No segundo, o fator de normalização chamado TMM (*Trimmed Mean of M-values*) é calculado para cada amostra tomando-se uma delas como referência e as outras como teste. Para cada amostra teste, o TMM é calculado pela média ponderada das razões do número de *reads* entre a amostra teste e a referência em escala log, após a exclusão dos genes mais e menos expressos. A normalização da expressão gênica é então obtida dividindo-se o número de *reads* mapeados no miRNA por este fator de normalização. A idéia é que para genes não diferencialmente expressos, o fator de normalização deve ser próximo de 1 (Seyednasrollah *et al.* 2013). Em resumo, esses métodos preocupam-se em considerar as diferenças na quantidade de *reads* mapeados para cada amostra e as diferenças extremas de expressão observadas entre os genes, tornando comparável a expressão gênica dentro e entre amostras.



**Figura 4. Identificação de miRNAs.** **A)** As seqüências geradas por um experimento de RNA-Seq provenientes de miRNAs maduros devem apresentar um padrão característico de mapeamento, precisa e abundantemente em uma das regiões maduras do precursor. **B)** Quando as seqüências apresentam um padrão aleatório de mapeamento, não se pode assumir os reads correspondam a miRNAs maduros. **C)** Exemplo do padrão de mapeamento do mir-9-1 humano, mostrando que a maioria dos reads correspondem à uma seqüência predominante, que em geral provém de uma das regiões maduras (-5p ou -3p). Figura adaptada de Berezikov (2011) e Pritchard *et al.* (2012).

#### 1.4 Reconhecimento de alvos e mecanismos de silenciamento

Os miRNAs podem inibir a expressão dos seus alvos através de dois mecanismos, promovendo a degradação e/ou a repressão da tradução do mRNA alvo (Bartel 2009). A interação dos miRNAs com seus alvos se dá pelo pareamento de bases complementares, usualmente na

região 3'UTR de RNAs mensageiros (Figura 5A). Ao longo dos últimos anos, uma série de parâmetros biológicos que definem o reconhecimento dos alvos foram elucidados (Grimson *et al.* 2007; Bartel 2009; Friedman *et al.* 2009). O mais importante é que o reconhecimento da sequência alvo em animais depende do pareamento completo da região *seed* (Figura 5A, exemplo 1), que compreende o trecho do segundo ao oitavo nucleotídeo da sequência madura (Figura 2). A presença de uma adenina na posição 1 e uma adenina ou uracila na posição 9 aumentam a eficiência da interação miRNA-mRNA (Grimson *et al.* 2007). Em animais, o pareamento na porção central e final do miRNA quase sempre é imperfeito. Em alguns casos, este pareamento pode exercer um efeito compensatório quando a complementaridade da *seed* não é perfeita (Figura 5A, exemplo 2) (Bartel 2009; Pasquinelli 2012). Além disso, regiões ricas em AU na sequência do mRNA conferem maior acessibilidade estrutural ao complexo RISC. Alvos localizados nas extremidades da região 3' UTR também estão associados a uma maior eficiência na atuação do miRNA (Grimson *et al.* 2007; Friedman *et al.* 2009; Bartel 2009). Múltiplos sítios-alvo para o mesmo ou diferentes miRNAs também contribuem para uma maior eficiência da repressão. Quando esses sítios estão próximos, um efeito cooperativo tende a ocorrer. Os níveis de expressão dos miRNAs e de seus alvos também influenciam a eficiência da regulação (Chang e Mendell 2007; Ameres e Zamore 2013).

Embora esses parâmetros sejam suficientes para definir boa parte das interações miRNA-mRNA, a complexidade dessas interações ainda está longe de ser bem compreendida. Há casos em que o miRNA liga-se à região codificante ou 5'UTR de mRNAs sem que necessariamente ocorra um pareamento perfeito da *seed* (Figura 5A, exemplo 3). Em outros casos o pareamento pode ocorrer na região central do miRNA maduro (ex. miR-124 e o gene RPTOR), e em plantas o pareamento em geral é praticamente perfeito entre toda a extensão do miRNA maduro e a região alvo (Figura 5A, exemplo 4) (Pasquinelli 2012). Além disso, a regulação por miRNAs não é

restrita a genes codificantes, podendo também regular outras classes de genes como lncRNAs e pseudogenes (Poliseno *et al.* 2010; Chiyomaru *et al.* 2014).

Para que a tradução ocorra, é necessário que os mRNAs contenham a estrutura 5' *cap* e uma cauda 3'-poly(A). Os fatores que se associam com o 5' *cap* e a cauda poly(A) interagem entre si: a proteína estabilizadora da cauda poly(A) PABP interage com o fator de iniciação eIF4G, que associa-se ao 5' *cap* (intermediado pelo eIF4E), configurando uma estrutura circular do mRNA que é eficientemente traduzido e protegido de degradação (Ameres e Zamore 2013). Acredita-se que a inibição mediada pelo complexo miRISC ocorre principalmente no início da tradução, interferindo no reconhecimento do 5' *cap* pelo complexo que liga-se a esta estrutura, que, intermediado pelo fator eIF4F, impede que a subunidade 40S do ribossomo se posicione para o início da tradução (Figura 5B) (Fabian *et al.* 2010; Huntzinger e Izaurralde 2011). Há evidências de que a tradução pode ser inibida após a iniciação, em que o complexo miRISC impede o avanço do ribossomo (modelo conhecido como *ribosome drop-off*) (Figura 5B) (Stefani e Slack 2008).

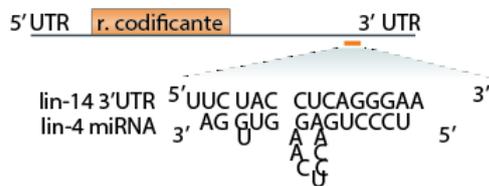
Além da inibição da tradução, os miRNAs podem induzir a degradação dos mRNAs alvos. Acredita-se que o complexo miRISC interage com o complexo deadenilase, que por intermédio das proteínas CCR4 e NOT, promove a desestabilização e degradação da cauda poly(A) do RNA mensageiro. Em seguida o *cap* é removido pela enzima DCP1-DCP2, tornando possível a degradação do mRNA por exonucleases 5'-3' (ex., XRN1) (Huntzinger e Izaurralde 2011) (Figura 5B).

Ainda é discutível o mecanismo majoritário (inibição da tradução ou degradação do mRNA) pelo qual ocorre o silenciamento por miRNAs. Estudos em mamíferos sugerem que a degradação do mRNA explica grande parte (~84%) da redução dos níveis protéicos (Guo *et al.* 2010). Entretanto, análises em *zebrafish* e *Drosophila* indicam que a repressão da tradução precede a degradação do mRNA (Bazzini *et al.* 2012; Djuranovic *et al.* 2012). Essas diferenças

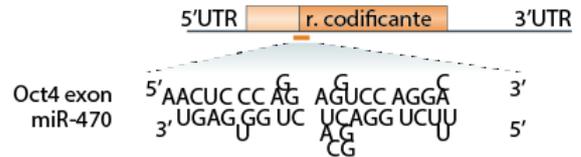
podem estar relacionadas com os tempos de medição dos níveis protéicos e dos mRNAs nos diferentes estudos após a transfecção do miRNA testado, ou com a eficiência da degradação após a desestabilização do mRNA (Hu e Coller 2012), fazendo com que a detecção seja favorável a um ou outro mecanismo. Independente da ordem dos eventos, via de regra, a consequência final é a diminuição dos níveis protéicos e de mRNAs (Selbach *et al.* 2008).

## A

### 1 *Caenorhabditis elegans* lin-14



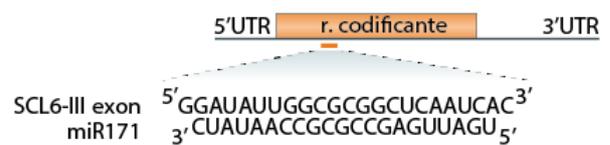
### 3 *Mus musculus* Oct4



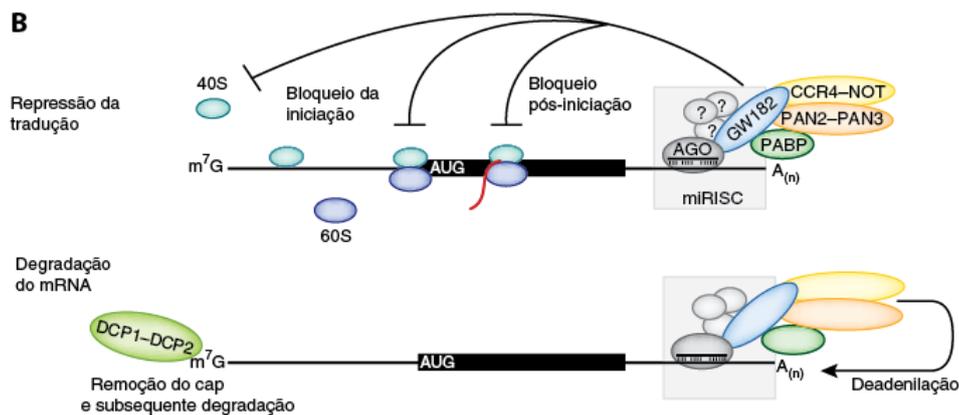
### 2 *Caenorhabditis elegans* lin-41



### 4 *Arabidopsis thaliana* SCL6-III



## B



**Figura 5. Exemplos de interação miRNA-alvo e mecanismo de silenciamento.** A) Exemplo 1: lin-4 e região alvo na 3'UTR do gene lin-14, mostrando o modelo clássico de pareamento perfeito da região seed. Exemplo 2: pareamento imperfeito da seed e pareamento compensatório na região central do miRNA. Exemplo 3: Alvo na região codificante do gene Oct4 de camundongo. Exemplo 4: Silenciamento em plantas, em que é comum o pareamento completo do miRNA, também em região codificante. Modificado de Pasquinelli (2012) B) Ilustração simplificada dos efeitos do complexo miRISC, que impe a tradução do mRNA desfavorecendo o reconhecimento do 5' cap pelo ribossomo ou a progressão do ribossomo, ou ainda pelo recrutamento de deadenilases (CCR4-NOT e PAN2-PAN3) e posterior degradação. Modificado de Fabian e Sonenberg (2012).

Os miRNAs têm o potencial de regular centenas de mRNAs, enquanto um mRNA pode ser regulado por vários miRNAs. Devido a enorme complexidade das interações miRNA-alvo, a caracterização de tais interações por validações experimentais é um processo extremamente laborioso (Bushati e Cohen 2007). Para lidar com essa dificuldade, inúmeras ferramentas computacionais de predição de alvos foram desenvolvidas (ex. TargetScan, miRanda, PicTar, miRmap, entre outros) e são extensivamente utilizadas neste tipo de análise. Estima-se que os miRNAs controlam a atividade de pelo menos 30% a 60% dos genes codificadores de proteínas em mamíferos (Friedman *et al.* 2009; Guo *et al.* 2010). Recentemente, o desenvolvimento de abordagens experimentais em larga escala baseadas na co-imunoprecipitação de proteínas ligadoras de RNA (RBPs – *RNA binding proteins*) e RNAs alvo, tais como PAR-CLIP, HITS-CLIP e iCLIP, têm se mostrado eficientes na detecção das interações miRNA-mRNA (Hausser e Zavolan 2014).

Os efeitos globais da regulação apresentam-se de maneiras distintas. Determinados miRNAs provocam uma repressão substancial dos seus alvos, a ponto de funcionarem como *switches* genéticos (ex. let-7 e o alvo lin-4). Entretanto, o que tipicamente ocorre é uma repressão parcial, cujos efeitos fenotípicos podem ou não ser claramente evidentes (Chen e Rajewsky 2007). A maioria dos miRNAs é responsável pelo ajuste fino da expressão gênica, mantendo seus alvos em níveis estáveis de acordo com o programa celular estabelecido. Neste sentido, os miRNAs podem tanto ajustar os níveis médios de expressão de seus alvos (*expression tuning*), quanto ajustar a variância da expressão (*expression buffering*). Neste último caso, a maior contribuição é na manutenção da estabilidade e robustez das redes regulatórias (Wu *et al.* 2009; Li e Zhang 2013).

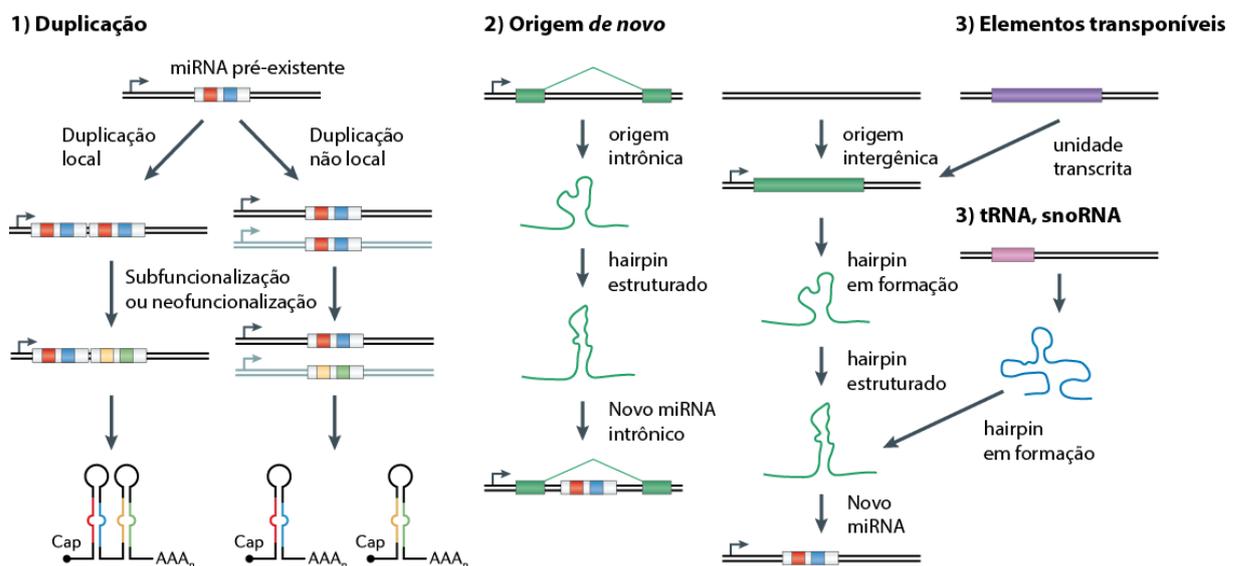
## 1.5 Evolução de miRNAs

Em animais, a distribuição filogenética dos miRNAs estende-se por praticamente todos os grupos de organismos. O mir-100, por exemplo, é compartilhado entre todos os bilaterados e está presente em cnidários (Grimson *et al.* 2008). Na base dos principais eventos de irradiação adaptativa observam-se concomitantes expansões do repertório de miRNAs. A presença de um conjunto de 34 miRNAs compartilhado entre protostômios e deuterostômios sugere uma expansão na base dos bilateria (Hertel *et al.* 2006). Outras expansões são observadas na origem dos vertebrados e após a separação dos mamíferos placentários (Hertel *et al.* 2006; Heimberg *et al.* 2008). Uma expansão ainda mais recente e numerosa ocorreu na base dos primatas, sendo que mais de 50% dos miRNAs humanos surgiu neste grupo (Iwama *et al.* 2013). Ao contrário dos genes codificadores de proteínas, a evolução dos miRNAs é um processo altamente dinâmico, caracterizado por altas taxas de ganhos e perdas, e expansões linhagem-específicas (Lu *et al.* 2008; De Wit *et al.* 2009; Berezikov 2011). Tais expansões têm sido relacionadas com o estabelecimento da complexidade morfológica dos vertebrados (Heimberg *et al.* 2008) e no surgimento de características espécie-específicas, como a participação do miR-941 na sinalização de neurotransmissores em humanos (Hu *et al.* 2012) e do miR-2786 no padrão de desenvolvimento das asas em lepidoptera (ex. borboletas) (Quah *et al.* 2015).

Diferentes mecanismos de origem de novos miRNAs têm sido propostos. Duplicações em *tandem* costumam gerar cópias locais de um miRNA pré-existente, dando origem à *clusters* de miRNAs cujos membros possuem uma origem em comum, enquanto duplicações não-locais dão origem a miRNAs em localizações genômicas distintas (Figura 6, exemplo 1). Com o tempo, os processos de subfuncionalização e neofuncionalização podem definir o destino funcional das novas cópias. Estima-se que cerca de um terço dos miRNAs humanos guardam homologia de sequência entre si, podendo ser agrupados em famílias gênicas (Berezikov 2011). Ao contrário dos genes codificantes, a origem *de novo* a partir de sequências intergênicas ou intrônicas é

frequente. Nestes casos, o acúmulo de mutações ao longo do tempo pode produzir uma estrutura em *hairpin*, que passa a ser transcrita e reconhecida pela maquinaria de processamento, dando origem a um novo miRNA (Berezikov 2011) (Figura 6, exemplo 2). A duplicação e origem *de novo* são os dois principais mecanismos de origem de novos miRNAs, ambos contribuindo de maneira equivalente para o repertório de miRNAs em vertebrados (Meunier *et al.* 2013). A origem de miRNAs a partir de regiões repetitivas (ex. transposons e retrotransposons), pseudogenes, snoRNAs e tRNAs também já foram reportadas (Devor 2006; Piriyaopongsa *et al.* 2007; Ender *et al.* 2008).

Em relação à conservação de sequência, a extensa região que compreende o pri-miRNA em geral é pouco conservada entre espécies. A conservação normalmente está confinada à sequência precursora, sendo que as regiões madura e *seed* apresentam maior grau de conservação (Mohammed *et al.* 2013).



**Figura 6. Principais mecanismos de surgimento de novos miRNAs.** Exemplo 1: Duplicações de um miRNA pré-existente, sejam em cópias locais (*tandem*) ou em diferentes regiões genômicas, seguida de sub- ou neofuncionalização. Exemplo 2: Origem a partir de sequências intrônicas ou intergênicas, que com o tempo acumulam mutações, favorecendo a formação de uma estrutura secundária que passa a ser reconhecida pela maquinaria de processamento de miRNAs. Exemplo 3: Sequências derivadas de elementos transponíveis e outros RNAs curtos podem dar origem a novos miRNAs. Figura adaptada de Berezikov (2011).

Os miRNAs antigos (ex. origem na base de Bilateria) tendem a apresentar expressão elevada e ampla (diversos tecidos), enquanto os miRNAs novos (ex. origem em primatas) comumente apresentam expressão baixa e restrita a um único ou poucos tecidos (tecido-específicos) (Chen e Rajewsky 2007; Meunier *et al.* 2013), podendo ser fixados ou perdidos em um curto período de tempo (Meunier *et al.* 2013; Lyu *et al.* 2014). Após uma fase adaptativa inicial, os miRNAs que foram preservados sofrem transição para uma fase conservativa e passam a ser gradualmente expressos em níveis mais elevados e em um maior número de tecidos, e passam então a ser efetivamente integrados nas redes regulatórias e sujeitos à seleção purificadora mais intensa (Lyu *et al.* 2014). Embora o impacto funcional dos miRNAs novos para determinação de um fenótipo específico possa ser discreto e discutível por causa do baixo nível de expressão, sugere-se que o principal papel desses miRNAs esteja associado com o processo de canalização<sup>1</sup>, contribuindo para a manutenção da estabilidade do transcriptoma (Hornstein e Shomron 2006; Wu *et al.* 2009).

## 1.6 MiRNAs e câncer

Pelo fato dos miRNAs atuarem como reguladores globais da expressão gênica, alterações nas etapas de biogênese, expressão, ou eventos mutacionais podem impactar processos celulares básicos como proliferação, apoptose, migração, etc. que, quando desregulados, contribuem para o surgimento do câncer. A expressão diminuída ou a deleção do *cluster* miR-15a e miR-16 em cerca de 68% dos pacientes com leucemia linfóide crônica constitui uma das primeiras e mais bem caracterizadas associações entre alterações da expressão de miRNAs e o câncer (Calin *et al.* 2004). Em tumores sólidos, a expressão aumentada de miR-17, miR-20, miR-21 e miR-92 são exemplos bem estabelecidos dessa desregulação (Volinia *et al.* 2006). Em câncer colorretal, foco

---

<sup>1</sup> Capacidade dos sistemas biológicos resistirem à perturbações genéticas e ambientais através de mecanismos que mantêm uma característica/fenótipo constante, diminuindo sua variabilidade (Hornstein e Shomron 2006).

de um dos trabalhos desenvolvidos nesta tese, diversos miRNAs (ex. miR-21, miR-126, miR-143, etc.) já foram identificados como diferencialmente expressos em comparação com o tecido normal, tendo como alvos reguladores-chave da proliferação (ex. KRAS), apoptose (ex. TP53), migração (ex. RECK3 e TIMP3), dentre outros processos (ver revisões em Slaby *et al.* (2009) e Mazez *et al.* (2013)).

A resistência aos compostos quimioterápicos representa uma das maiores barreiras para o tratamento de diversos tipos de câncer. Duas das drogas mais utilizadas no tratamento do câncer colorretal são a 5-Fluoruracil e Oxaliplatina (André *et al.* 2004). A 5-Fluoruracil é geralmente utilizada como primeira opção de tratamento, resultando em considerável melhora na sobrevida dos pacientes (Longley *et al.* 2003). Esta droga é um análogo da uracila, que, após entrar na célula, é convertida em diferentes metabólitos (FdUMP, FdUTP e FUTP), os quais inibem a síntese de RNA e a ação da timidilato sintase (TS). Essa enzima é responsável pela síntese de timidina (nucleosídeo associado à timina) a partir de dUMP, cuja atividade é inibida competitivamente por análogos da uracila, prejudicando a síntese de DNA e RNA (Longley *et al.* 2003). O miR-21 têm um papel crucial na resistência à 5-FU, cuja expressão aumentada em câncer colorretal está correlacionada com a inibição de MSH2, um importante componente do sistema de reparo (Valeri *et al.* 2010). A Oxaliplatina é utilizada (geralmente em combinação com 5-FU) com significativa eficácia no tratamento do câncer colorretal avançado (André *et al.* 2004). Ela exerce seu efeito citotóxico como um agente intercalante de DNA, em que o átomo da platina liga-se covalentemente ao N<sup>7</sup> de purinas, causando danos no DNA e que por consequência ativam mecanismos apoptóticos (Wang e Lippard 2005). Qian *et al.* (2013) demonstraram que a superexpressão de miR-143 está associada com a maior sensibilidade à essa droga na linhagem celular de câncer colorretal SW1116. Dessa forma, estes dois exemplos ilustram a importância dos miRNAs nos mecanismos de resistência à drogas e seu potencial como alvos terapêuticos e por isso é crucial que novos esforços melhor estabeleçam essas associações (ver revisão em Allen e Weiss (2010)).

## *2 Objetivos*

## 2.1 Objetivos gerais

Buscamos caracterizar o contexto genômico dos miRNAs em múltiplas espécies, integrando dados de anotação e expressão de miRNAs e genes codificadores de proteínas em uma ferramenta *web*. Além disso, propomos investigar como este contexto genômico influencia os padrões de expressão e evolução dos miRNAs humanos. Também como objetivo desta Tese, propomos identificar miRNAs que possam estar relacionados com mecanismos de resistência ao tratamento com drogas quimioterápicas em linhagens de câncer colorretal.

## 2.2 Objetivos específicos

- Construir um banco de dados para integração de dados de anotação e expressão de miRNAs e genes codificadores de proteínas para cinco espécies de vertebrados, disponibilizando essas informações em uma ferramenta *web* de fácil uso, com um mecanismo de busca eficiente e publicamente disponível. Este estudo é apresentado como artigo publicado no Anexo B.
- Determinar a origem evolutiva de miRNAs e genes codificadores de proteínas humanos, bem como os padrões de expressão em diferentes tecidos, relacionando-os com o contexto genômico, idade evolutiva dos genes hospedeiros e aspectos funcionais dessas duas classes de genes.
- Identificar miRNAs diferencialmente expressos em linhagens celulares de câncer colorretal resistentes e sensíveis à drogas quimioterápicas, buscando possíveis mecanismos associados através de validação experimental *in vitro* e análise de alvos potencialmente regulados.

### *3 Materiais e Métodos*

### 3.1 Dados públicos utilizados

#### 3.1.1 Genomas de referência

Os genomas de referência e suas respectivas versões foram obtidos do UCSC *Genome Browser* (<http://hgdownload.cse.ucsc.edu/downloads.html>) para as seguintes espécies: *Homo sapiens* (Humanos): hg19; *Pan troglodytes* (Chimpanzé): panTro4; *Gorilla gorilla* (Gorila): gorGor3; *Pongo abelii* (Orangotango): ponAbe2; *Macaca mulatta* (Macaco Rhesus): rheMac3; *Callithrix jacchus* (Saguí): calJac3; *Mus musculus* (Camundongo): mm10; *Rattus norvegicus* (Rato): rn5; *Bos taurus* (Vaca): bosTau7; *Canis familiaris* (Cachorro): canFam3; *Monodelphis domestica* (Opossum - gambá): monDom5; *Ornithorhynchus anatinus* (Ornitorrinco): ornAna1; *Gallus gallus* (Galinha): galGal4 e *Danio rerio* (Zebrafish): danRer7.

#### 3.1.2 Anotações de miRNAs

As sequências em formato FASTA das regiões precursoras e maduras dos miRNAs anotados para as espécies mencionadas acima foram obtidas do repositório oficial de miRNAs, o miRBase (<http://www.mirbase.org>; versão 20, Junho de 2013), bem como as coordenadas genômicas dessas regiões extraídas de arquivos em formato GFF3 (<http://gmod.org/wiki/GFF3>), disponíveis em: <ftp://mirbase.org/pub/mirbase/20/>

#### 3.1.3 Anotações de genes codificadores de proteínas

Os arquivos de anotação de genes e transcritos em formato GTF (<http://www.encodegenes.org/encodeformat.html>) utilizados para a quantificação da expressão gênica foram obtidos para *Homo sapiens*, *Macaca mulatta*, *Mus musculus*,

*Monodelphis domestica*, *Ornithorhynchus anatinus* e *Gallus gallus* através do banco Ensembl (<http://www.ensembl.org>) versão 71.

#### 3.1.4 Dados de expressão de miRNAs

Dados brutos de RNA-Seq de pequenos RNAs foram obtidos em formato .sra, disponíveis no NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>; GEO). Dados provenientes de tecidos normais para humanos foram obtidos dos seguintes *datasets*: GSE46622 (colorretal), GSE33858 (pulmão), GSE47720 (célula pancreática beta), GSE37686 (célula epitelial de retina), GSE32493 (célula B, linfócitos), GSE31617 (bexiga) e GSE19812 (células mononucleares de sangue periférico). Também foram utilizados os dados gerados por Meunier *et al.* (2013) (acesso: GSE40499), que contém arquivos de sequenciamento de RNAs pequenos para cinco tecidos (cérebro, cerebelo, coração, rim e testículo) de seis espécies (*Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Monodelphis domestica*, *Ornithorhynchus anatinus* e *Gallus gallus*). Para a espécie humana, os dados brutos de sequenciamento foram processados para obtenção dos níveis de expressão de miRNAs, totalizando 12 tecidos/tipos celulares para a espécie humana e 5 tecidos para as demais espécies.

#### 3.1.5 Dados de expressão de genes codificantes de proteínas

Os arquivos brutos de RNA-Seq em formato .sra foram obtidos do *Illumina Human Body Map 2.0 dataset*, disponível no EBI ArrayExpress (acesso: E-MTAB-513). Este conjunto de dados consiste no sequenciamento de RNAs de 16 tecidos humanos (tecido adiposo, glândula adrenal, ovário, mama, colon, coração, rim, leucócito, fígado, pulmão, linfonodo, prostata, músculo esquelético, testículo e glândula tireóide), que foram processados para obtenção dos

níveis de expressão de genes codificantes de proteínas. Também foram obtidos os dados gerados por Brawand *et al.* (2011) a partir do NCBI GEO (acesso: GSE30352). A partir deste *dataset*, foram utilizados os dados de amostras de cérebro, cerebelo, coração, rim, fígado e testículo de cinco espécies (*Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Monodelphis domestica* e *Gallus gallus*). Ambos os datasets obtidos foram gerados a partir de bibliotecas do tipo *paired-end*.

## 3.2 Expressão de miRNAs e de genes codificadores de proteínas

### 3.2.1 Quantificação da expressão de miRNAs a partir de dados de RNA-Seq

Para quantificar a expressão dos miRNAs, um *pipeline* (sequência de comandos e scripts) de bioinformática foi desenvolvido para este propósito. Para as sequências obtidas de bancos públicos em formato .sra, estas foram convertidas para o formato FASTq utilizando o programa fastq-dump do pacote SRA Toolkit (<http://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/2.3.3/>). Os *reads* (sequências ou leituras) foram então submetidos ao processo de remoção de sequências adaptadoras remanescentes utilizando o pacote FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Para as sequências geradas pela plataforma SOLiD 4 (seção 3.5.1), utilizamos a estratégia de trimagem sequencial dos adaptadores localizados na extremidade 3' implementada no programa SeqTrimMap 0.9 (Marco e Griffiths-Jones 2012). Após a remoção dos adaptadores, os *reads* foram mapeados nos genomas das respectivas espécies (seção 3.1.1) com o programa Bowtie 1.0.0 (Langmead *et al.* 2009) exigindo-se alinhamentos perfeitos (parâmetro -v 0) em no máximo 10 loci diferentes (-m 10), selecionando-se apenas os melhores alinhamentos (--best --strata). *Reads* menores que 16 nt foram descartados. As regiões mapeadas foram cruzadas com as coordenadas genômicas dos

miRNAs maduros anotados com o programa BEDTools (Quinlan e Hall 2010), de modo que houvesse total sobreposição com a região anotada. A expressão dos miRNAs maduros foi determinada pela soma do número de *reads* mapeados nas regiões correspondentes dos miRNAs anotados. Para contemplar pequenas diferenças decorrentes do processamento alternativo do miRNA maduro (isomiRs), admitimos a extensão de 3 nt na borda 3' em relação à anotação oficial. Como existem miRNAs maduros idênticos derivados de diferentes precursores, o número de *reads* que mapearam em múltiplas regiões no genoma com igual qualidade foi dividido pelo número correspondente de *loci* diferentes. Essa é uma estratégia comumente utilizada para evitar que a expressão de um determinado miRNA maduro com múltiplas cópias seja superestimada (Pritchard *et al.* 2012). A expressão de cada miRNA maduro foi normalizada levando em conta o total de *reads* mapeados em miRNAs com o pacote EdgeR 3.4.2 (Robinson *et al.* 2010).

Para as análises evolutivas de miRNAs (seção 3.3), utilizamos o valor de expressão do miRNA maduro mais expresso (5p ou 3p) como referência para expressão do precursor. Nos casos em que há miRNAs maduros idênticos para diferentes precursores, consideramos apenas o precursor de origem mais antiga (ver seção 3.3.2). Para todas as análises envolvendo expressão de miRNAs, foram considerados apenas aqueles cuja expressão normalizada foi  $> 1$  cpm<sup>2</sup> em pelo menos um tecido/tipo celular.

---

<sup>2</sup> CPM (*counts per million*) é uma unidade de expressão gênica, mais comumente utilizada quando o tamanho entre os genes comparados são aproximadamente equivalentes, e que corresponde ao número de *reads* reportados para um determinado gene normalizado pelo número de *reads* mapeados, neste caso, o número total de *reads* mapeados em miRNAs.

### 3.2.2 *Expressão de genes codificadores de proteínas a partir de dados de RNA-Seq*

Os arquivos de sequenciamento obtidos de dados públicos em formato .sra foram convertidos para o formato FASTq com o programa fastq-dump do pacote SRA Toolkit. Os *reads* foram mapeados nos genomas das respectivas espécies, usando os modelos de transcritos contidos nos arquivos GTF (seção 3.1.3) com o programa TopHat v.2.0.8 (Kim *et al.* 2013) (parâmetros *default*). Os alinhamentos foram filtrados com o pacote SAMTools (Li *et al.* 2009) exigindo-se qualidade mínima de alinhamento de 20 (-q 20) em escala Phred, o que corresponde a uma acurácia de 99%, e portanto virtualmente apenas alinhamentos únicos são considerados. A expressão normalizada dos genes foi calculada com o programa Cufflinks v.2.2.1 (Trapnell *et al.* 2012), cujo resultado é a expressão em valores de FPKM<sup>3</sup>. Apenas genes que tiveram expressão maior que 1 FPKM em pelo menos um tecido foram considerados.

## 3.3 Métodos utilizados para análises evolutivas dos miRNAs humanos

### 3.3.1 *Anotação dos miRNAs humanos de acordo com a posição genômica e anotação de genes hospedeiros*

Os miRNAs humanos obtidos do miRBase (versão 20) foram classificados como intragênicos caso as coordenadas das sequências precursoras estivessem sobrepostas por pelo menos uma base aos genes codificantes de proteínas anotados pelo Ensembl (versão 71), e em

---

<sup>3</sup> FPKM (*Fragments per kilobase of exon per million of mapped reads*) é uma unidade arbitrária de expressão gênica, mais comumente utilizada quando o tamanho dos genes a serem comparados diferem consideravelmente entre si, e que corresponde ao número de fragmentos (válido para bibliotecas do tipo paired-end) para cada mil bases (normalização por tamanho do transcrito) para cada milhão de reads mapeados (normalização pelo número de *reads* mapeados).

intergênicos para aqueles não sobrepostos. Os miRNAs intragênicos foram em seguida classificados na orientação senso ou antisenso em relação aos genes sobrepostos (genes hospedeiros), e posteriormente categorizados em intrônicos ou exônicos, dependendo da região gênica sobreposta. O transcrito mais logo do gene hospedeiro foi usado como referência. Para análises subsequentes, utilizamos uma definição estrita de miRNAs intragênicos, considerando apenas aqueles na mesma orientação dos genes hospedeiros (fita senso). Como controle para determinadas análises, miRNAs distantes entre si até 10 kb foram agrupados, com o objetivo de minimizar a potencial influência de grupos (*clusters*) de miRNAs que comumente são transcritos em um único RNA primário. Nestes casos, selecionamos aleatoriamente um dos miRNAs como representante do *cluster*. O critério de 10 kb foi escolhido com base em estudos prévios, demonstrando que este é um *cutoff* adequado para agrupar miRNAs em *clusters* (Altuvia *et al.* 2005). Agrupando-se miRNAs distantes à 50 kb não resultou em mudanças significativas nos principais resultados apresentados.

### 3.3.2 Definição das idades dos miRNAs

Para definir o período de origem dos miRNAs humanos na linhagem dos vertebrados, suas sequências precursoras foram interrogadas nos genomas de outras 13 espécies obtidas do UCSC Genome Browser (ver seção 3.1.1) com a finalidade de recuperar as sequências homólogas em cada espécie. A estratégia utilizada foi similar à de (Hu *et al.* 2011) e (Hu *et al.* 2012), com algumas modificações. Inicialmente, os melhores alinhamentos recíprocos são recuperados, usando BLAT ((Kent 2002), parâmetros: `stepSize = 5 -repMatch = 2253 minScore = 0, -minIdentity = 0`), BLASTN ((Altschul *et al.* 1997), parâmetros: `-word_size 8 -evaluate 1e-05`) e LiftOver ((Fujita *et al.* 2011), parâmetros *default*), exigindo-se que a região alinhada tenha pelo menos 70% e no máximo 130% do tamanho da sequência *query*. Em seguida, o miRNA

ortólogo foi definido quando a região alinhada teve suporte de pelo menos dois dos três métodos. As sequências maduras foram identificadas pelo alinhamento global das regiões ortólogas precursoras usando o programa CLUSTALW ((Larkin *et al.* 2007), parâmetros *default*). Para aproveitar as informações disponíveis no miRBase, as regiões ortólogas que possuíram sobreposição com anotações já conhecidas na mesma orientação transcricional e em pelo menos 50% do tamanho da região, foram usadas as anotações de regiões precursora e madura. Por fim, a idade dos miRNAs foi designada por números em ordem crescente ao longo da árvore filogenética das espécies analisadas, refletindo o grupo mais primitivo em que o miRNA ortólogo foi encontrado.

### 3.3.3 Definição das idades dos genes codificantes de proteínas

As idades dos genes codificantes de proteínas foram cedidas pelo Dr. Yong E. Zhang (<http://zhanglab.ioz.ac.cn/>), que ajustou o método de datação de genes empregado em Zhang *et al.* (2010) e Zhang *et al.* (2011) para uma versão mais recente de anotações gênicas (Ensembl v. 71). O método consiste em encontrar o *locus* referente a espécie humana com o melhor alinhamento recíproco sintênico nos arquivos de alinhamento genômicos disponíveis no UCSC Genome Browser, levando em conta a conservação de genes vizinhos. Descrições mais detalhadas dos métodos estão nos artigos originais e citados acima. Após a obtenção das idades dos genes, nós as reatribuímos levando em conta a árvore filogenética das espécies, permitindo comparar numa mesma escala filogenética as idades dos genes e dos miRNAs. Também utilizamos idades de genes obtidas por outros dois métodos (Ensembl e dados obtidos de Chen *et al.* (2012)) como parâmetro de comparação.

### 3.3.4 *Análise estatística das idades dos genes hospedeiros*

Para verificar se miRNAs intragênicos surgem preferencialmente em genes novos ou antigos, comparamos a proporção observada de genes antigos (idade = 1, ver Fig. 17) com a proporção esperada obtida de uma distribuição nula gerada por 10.000 amostras aleatórias formadas pelo mesmo número de genes de idade igual ou mais antiga que o número de miRNAs de uma determinada idade. Assumimos que a origem do gene hospedeiro precede à do seu miRNA intragênico, por isso utilizamos este procedimento de reamostragem. O p-valor empírico foi calculado como sendo a proporção de genes antigos maior ou igual à proporção observada. Alternativamente, a média da distribuição nula foi considerada como a proporção esperada de genes antigos e a diferença estatística para a proporção de genes hospedeiros antigos foi determinada pelo teste de Chi-quadrado. Casos duvidosos, em que a idade do gene hospedeiro foi definida como mais recente que à do miRNA foram removidos (N = 13).

### 3.3.5 *Análise de Ka/Ks*

Para analisar as distribuições das taxas de substituição não-sinônimas por sítios não sinônimos (Ka) em razão das taxas de substituições sinônimas por sítios sinônimos (Ks), utilizamos os valores de Ka e Ks computados para alinhamentos entre proteínas ortólogas da espécie humana e de camundongo. Estes valores foram diretamente obtidos do Ensembl v.71. Diferenças significativas entre as distribuições de Ka/Ks para genes hospedeiros e o restante dos genes codificantes foram calculadas com o teste de Kolmogorov-Smirnov (KS-test).

### 3.3.6 Abrangência de expressão de miRNAs e genes codificantes

Para determinar a abrangência de expressão de miRNAs e genes codificantes, utilizamos o índice de tecido-especificidade ( $\tau$ ) definido por Yanai *et al.* (2005), que leva em conta os níveis de expressão da amostra  $i$  normalizado pelo valor máximo de expressão encontrados nas amostras ( $x_i$ ) e o número de tecidos ( $N$ ) e que um gene é expresso (Equação 1). Para este cálculo, adotamos uma transformação logarítmica da expressão normalizada (adicionando 1 para lidar com valores de expressão menores que 1). Este índice varia de 0 a 1, onde valores próximos a 0 indicam ampla expressão e valores próximos a 1 indicam expressão tecido-específica. Por exemplo, expressão tecido-específica ( $\tau \geq 0.7$ ) em nossos dados indicam que miRNAs e genes codificantes são expressos em uma mediana de 5 (de 12) e 3 (de 16) tecidos, respectivamente. Expressão ampla ( $\tau \leq 0.3$ ) corresponde a miRNAs e genes expressos em uma mediana de 12 e 16 tecidos, respectivamente. Diferenças significativas entre as diferentes distribuições de abrangência de expressão foram calculadas por testes de Mann-Whitney. Correlações entre a abrangência de expressão de miRNAs e seus genes hospedeiros foram calculadas com testes de correlação de Spearman.

$$\tau = \frac{\sum_{i=1}^N (1 - x_i)}{N - 1} \quad (1)$$

### 3.3.7 Correlação da expressão entre miRNAs intragênicos e seus genes hospedeiros

Para verificar se as correlações entre a abrangência de expressão de miRNAs e genes hospedeiros estão associadas com a co-expressão das duas classes de genes, os dados de expressão de miRNAs e genes hospedeiros disponíveis para os mesmos tecidos (cérebro, cerebelo, coração, rim e testículo) provenientes de Meunier *et al.* (2013) e Brawand *et al.* (2011) foram usados para um teste de randomização. Este teste consiste em verificar se a expressão de

miRNAs e genes hospedeiros está significativamente associada em relação ao esperado ao acaso. Para um dado par miRNA-gene hospedeiro, os tecidos em que os pares miRNA-gene são expressos foram ordenados pelos níveis de expressão e computada a proporção de tecidos em que o miRNA é expresso na mesma ordem de tecidos em que seu gene hospedeiro. A média das proporções foi comparada com uma distribuição nula gerada pela randomização da ordem de expressão nos tecidos num total de 10.000 vezes. Adotamos esse procedimento, ao invés de um método de correlação, devido às discrepâncias inerentes aos níveis e amplitude de expressão entre miRNAs mais antigos e jovens, e também por causa do pequeno número de amostras (5 tecidos). Como os miRNAs antigos são mais expressos e em maior número de tecidos, sua expressão tende a correlacionar melhor com a expressão dos genes hospedeiros (os quais demonstramos ser em sua maioria antigos). Usando um teste de randomização em que apenas a ordem em dos tecidos em que os miRNAs são mais expressos é levada em conta, tais disproporções exercem um efeito menos acentuado nas comparações entre miRNAs de idades diferentes.

### 3.3.8 *Divergência de expressão de miRNAs entre espécies*

A divergência de expressão entre miRNAs humanos e seus respectivos ortólogos em cinco espécies (*Macaca mulatta*, *Mus musculus*, *Monodelphis domestica*, *Ornithorhynchus anatinus* e *Gallus gallus*) foi calculada pelas distâncias Euclidianas conforme sugerido e utilizado por Jordan *et al.* (2005) e Glazko e Mushegian (2010). Os valores de expressão normalizados (transformados em escala log<sub>2</sub>) em uma matriz de cinco tecidos (cérebro, cerebelo, coração, rim e testículo) foram usados para determinar as distâncias dos níveis de expressão em cinco tecidos (cérebro, cerebelo, coração, rim e testículo) para cada par miRNA humano-miRNA ortólogo.

### 3.3.9 Análise de conservação de sequências de miRNAs

A grau de conservação das sequências precursoras dos miRNAs foi investigada utilizando as pontuações (*scores*) PhyloP (Pollard *et al.* 2010) obtidos do UCSC Genome Browser. O programa PhyloP reporta *scores* de conservação ou evolução acelerada para cada base do genoma em relação ao esperado sob neutralidade, fazendo uso de alinhamentos múltiplos entre uma série de espécies (vertebrados, mamíferos e primatas). *Scores* positivos são proporcionais ao grau de conservação entre espécies, e *scores* negativos indicam sítios sob evolução acelerada (Pollard *et al.* 2010). Os *scores* previamente computados para alinhamentos entre espécies de vertebrados foram utilizados para miRNAs conservados além de mamíferos (classes de idade 1 e 2-4). Para miRNAs conservados em mamíferos placentários (classe de idade 5-6), utilizamos os *scores* compilados para alinhamentos de mamíferos, e para miRNAs conservados em primatas, utilizamos os *scores* compilados a partir de alinhamentos entre primatas (classe de idade 7-12). O PhyloP *score* de cada precursor foi determinado pela média dos *scores* das bases individuais. Para determinar as distribuições nulas dos *scores* nas regiões adjacentes aos miRNAs intragênicos, obtivemos as médias dos *scores* para 100 regiões intrônicas aleatoriamente selecionadas (com tamanhos iguais aos dos miRNAs) pertencendo aos genes hospedeiros. Para miRNAs intragênicos, os *scores* adjacentes foram calculados pela média de 100 regiões intergênicas num intervalo de 10 kb à justante ou à montante do miRNA analisado. Também foram calculados os PhyloP *scores* das regiões flanqueantes adjacentes aos miRNAs. Para determinar as distribuições de *scores* das regiões precursora e *seed* dos miRNAs, utilizamos o conjunto de miRNAs expressos, que foi controlado pela presença de sequências maduras idênticas e nível de expressão.

### 3.3.10 Predição de alvos dos miRNAs

Os alvos dos miRNAs humanos foram preditos com o algoritmo TargetScan 6.2, utilizando as sequências das regiões 3' UTR de genes codificantes de proteínas (considerando a isoforma mais longa por gene) obtidas do *website* do TargetScan (<http://www.targetscan.org>). Para obtermos um conjunto de genes com maior probabilidade de serem alvos de miRNAs, restringimos as predições para sítios do tipo 7mer-m8 e 8mer e com scores de contexto fornecidos pela ferramenta sendo  $< -0.25$ . Para que o número de genes-alvo fosse comparável entre os diferentes miRNAs, consideramos apenas aqueles presentes no conjunto de miRNAs expressos.

### 3.3.11 MiRNAs associados com doenças

Para testar se miRNAs inter- ou intragênicos estão sobre ou subrepresentados em conjuntos de dados que reportam associações de miRNAs com doenças, utilizamos os dados fornecidos pelas bases HMDD v.2.0 (Li *et al.* 2014) e PhenomiR 2.0 (Ruepp *et al.* 2010), que compilam uma série de informações de artigos publicados que evidenciam a participação de miRNAs cuja expressão foi alterada em doenças. Para avaliar o excesso ou depleção de miRNAs inter- ou intragênicos nestes conjuntos de dados, utilizamos testes de Fisher. Levando em conta que essas bases reportam dados para miRNAs anotados em diferentes períodos, e portanto não contemplam a anotação do repertório atual de miRNAs, utilizamos diferentes versões do miRBase (versão 19 para HMDD e 17 para PhenomiR) para obtenção das proporções de miRNAs inter e intragênicos anotados nos respectivos períodos em que essas bases foram disponibilizadas.

### 3.3.12 Análise de enriquecimento funcional dos genes hospedeiros

Para analisar o enriquecimento de termos referentes à categoria “processo biológico” do Gene Ontology (GO) (Ashburner *et al.* 2000) relacionados aos genes hospedeiros, a ferramenta AmiGO 2 (<http://amigo.geneontology.org/amigo>) foi utilizada (adotando a correção de Bonferroni para múltiplos testes assumindo  $P < 0.05$ ). Em seguida, utilizamos a ferramenta REViGO (<http://revigo.irb.hr/>) para sumarizar termos redundantes de uma mesma hierarquia. Análises utilizando a ferramenta DAVID 6.7 (Huang *et al.* 2009) produziram resultados bastante parecidos. Esta ferramenta também foi utilizada para avaliar se os genes hospedeiros estão enriquecidamente expressos em tecidos. Os p-valores foram considerados significativos para  $P < 0.05$ , adotando a correção para múltiplos testes de Benjamini-Hochberg.

## 3.4 Métodos utilizados para construção da ferramenta *web* miRIAD

### 3.4.1 Dados brutos

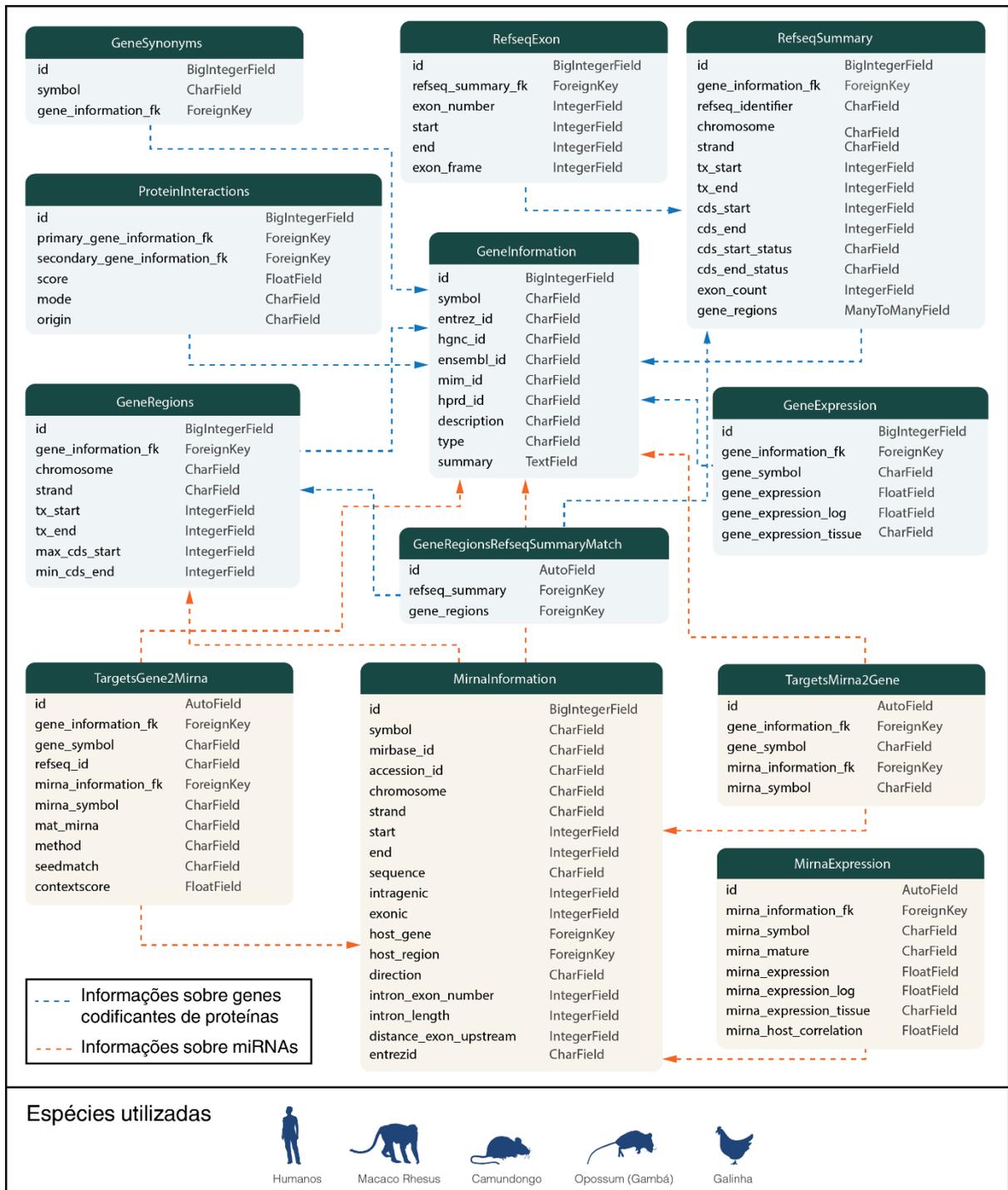
A construção da ferramenta por nós denominada miRIAD (Hinske *et al.* 2014) envolveu a integração de dados públicos de diferentes fontes. As coordenadas genômicas dos genes codificantes de proteínas e da estrutura exon-intron das isoformas de cada gene, assim como a descrição sumária dos genes foram obtidas do NCBI RefSeq para *Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Monodelphis domestica* e *Gallus gallus*. A anotação dos miRNAs foi obtida do miRBase, versão 20. Os alvos preditos e conservados dos miRNAs foram obtidos do TargetScan (<http://www.targetscan.org>, versão 6.2). Dados de interações proteína-proteína foram obtidas do *Human Protein Reference Database* (HPRD) e da base STRING (string-

db.org). Os dados de expressão gênica e de miRNAs foram obtidos e processados conforme as seções 3.2.1 e 3.2.2.

### 3.4.2 Arquitetura do banco de dados e interface web

Os dados brutos referentes a genes codificantes de proteínas e miRNAs foram processados em servidores locais utilizando *scripts* em linguagem de programação Python (<http://www.python.org>) para filtrá-los e organizá-los de forma adequada em um banco de dados relacional MySQL. Ao todo, o banco é composto por 12 tabelas para cada uma das 5 espécies (*Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Monodelphis domestica* e *Gallus gallus*), contendo aproximadamente 10 milhões de registros de informações relacionadas. As tabelas, seus respectivos campos e relações estão representadas na Figura 7.

A ferramenta *web* foi implementada utilizando o *framework* Django 1.5 (<https://www.djangoproject.com/>), escrito em linguagem Python. Basicamente, o *framework* se encarrega de fazer a conexão entre a aplicação web e o banco de dados, manejando as consultas e permitindo realizar operações, filtrar e customizar a apresentação dos dados. O sistema de busca foi implementado utilizando regras de expressões regulares que seguem uma hierarquia de prioridades, permitindo buscas flexíveis por diferentes identificadores de genes e miRNAs. Os *templates* das páginas foram escritos em HTML5 (<http://www.w3.org/TR/html5/>) e JavaScript (<https://en.wikipedia.org/wiki/JavaScript>) e os estilos definidos em CSS3 ([https://en.wikipedia.org/wiki/Cascading\\_Style\\_Sheets](https://en.wikipedia.org/wiki/Cascading_Style_Sheets)). Para gerar os diagramas das estruturas gênicas, a biblioteca GenomeTools foi utilizada.



**Figura 7.** Esquema representativo do banco de dados que constitui a ferramenta miRIAD. As tabelas são constituídas de seus respectivos campos, que são definidos pelo tipo de informação armazenada (CharField, FloatField, ForeignKey, etc.). As tabelas e flechas em azul representam informações referentes a genes codificantes e suas relações com outras tabelas. Tabelas e flechas em laranja referem-se à informações sobre miRNAs e suas relações com outras tabelas. Para cada espécie, um esquema de tabelas equivalente foi implementado.

### 3.4.3 Informações sobre genes hospedeiros e miRNAs

Os genes foram classificados em hospedeiros (*host*) e não hospedeiros (*non-host*) com base na presença de miRNAs sobrepostos. A classificação dos genes e informações adicionais foram armazenadas em três tabelas ('GeneInformation', 'GeneRegions' e 'GeneSynonyms'), como mostrado na Figura 7. Da mesma forma, os miRNAs foram classificados em intra- e intergênicos, com base na localização genômica. A tabela 'MirnaInformation' contém o nome oficial dos miRNAs, coordenadas genômicas do sequência precursora e seu gene hospedeiro, no casos dos intragênicos, assim como o intron ou exon sobreposto e a distância do miRNA ao exon mais próximo à 5' (Figura 7).

### 3.4.4 Alvos preditos

Os alvos conservados nas regiões 3' UTR para humanos e camundongo foram obtidos diretamente dos arquivos fornecidos pelo TargetScan (arquivo: Conserved\_Site\_Context\_Scores-.txt). Para as demais espécies utilizamos o algoritmo publicamente disponível do TargetScan para predição de alvos nas sequências 3' UTR dos respectivos organismos. O TargetScan define alvos de miRNAs procurando por segmentos complementares de 7 ou 8 bases (7mer e 8mer) nas sequências das regiões 3' UTR anotadas. Informações sobre conservação entre espécies são levadas em conta para definir o conjunto de alvos conservados.

### 3.5 Métodos utilizados para análise de expressão diferencial de miRNAs em linhagens celulares de câncer colorretal

#### 3.5.1 Sequenciamento dos RNAs pequenos em linhagens celulares de tumor colorretal

As 12 linhagens utilizadas foram cedidas pelo laboratório do Dr. John Mariadason (Instituto Ludwig de Pesquisa sobre o Câncer em Melbourne, Austrália), das quais 7 (Colo205, HCC2998, HCT116, HCT15, HT29, KM12 e SW620) provém do painel NCI60 e 5 (Caco-2, SW480, RKO, LIM1215 e LIM2405) foram estudadas no laboratório do Dr. Mariadason (Quadro 1, Apêndice). O RNA total de cada linhagem foi extraído, quantificado e seu grau de pureza e integridade determinados. A qualidade de todas as amostras, baseada na integridade do RNA (RIN), foi avaliada no Bioanalyzer. Todas as amostras apresentaram qualidade maior do que o recomendado (RIN maior do que 6). As bibliotecas de miRNA foram construídas a partir do RNA total das linhagens celulares utilizando-se o Total RNA-Seq *Kit*, seguindo as recomendações do fabricante (Life Technologies). As moléculas de RNAs pequenos foram submetidas à reação de transcrição reversa. O produto da reação foi separado em moléculas de 60 nt à 80 nt e amplificadas por PCR em emulsão. Após a amplificação, as *beads* recobertas por milhares de cópias de uma mesma molécula de DNA foram recuperadas, depositadas em uma lâmina específica e submetidas ao sequenciamento no SOLiD4 (Life Technologies). As etapas de extração, construção das bibliotecas de RNAs pequenos e sequenciamento foram executadas pela Dra. Camila Miranda Lopes (processo FAPESP 09/17785-5) sob a supervisão do Dr. Raphael Bessa Parmigiani.

### 3.5.2 Seleção de linhagens celulares de câncer colorretal resistentes e sensíveis ao tratamento com drogas quimioterápicas

A classificação das linhagens utilizadas neste estudo em resistentes e sensíveis à 5-FU e Oxaliplatina foram baseadas em resultados prévios obtidos a partir de um painel inicial de 30 linhagens de tumor colorretal, cuja resposta ao tratamento com drogas quimioterápicas foi avaliada através de ensaios de apoptose, inibição de crescimento e ensaio clonogênico. Os experimentos realizados para avaliar a resposta ao tratamento com 5-Fluorouracil (5-FU) estão descritos em Mariadason *et al.* (2003) e os experimentos realizados para avaliar a resposta à Oxaliplatina estão descritos em Arango *et al.* (2004). Do painel de 30 linhagens testadas, 11 foram gentilmente cedidas pelo laboratório do Dr. John Mariadason, a partir das quais foram construídas bibliotecas de RNAs pequenos e sequenciadas na plataforma SOLiD4. As 11 linhagens foram classificadas em resistentes e sensíveis, de acordo com a discrepância entre as porcentagens de apoptose perante ao tratamento com as drogas citadas. As linhagens consideradas resistentes à 5-FU foram HCT15, COLO205, SW620 e SW480, enquanto as sensíveis foram HCT116, KM12, LIM2405, HCC2998 (Figura S1, Apêndice). As linhagens consideradas resistentes à Oxaliplatina foram COLO205, SW620, HT29, KM12, LIM2405 e sensíveis foram SW480, RKO, CACO2, HCT116 e HCC2998 (Figura S2, Apêndice).

### 3.5.3 Expressão diferencial de miRNAs

A expressão diferencial dos miRNAs entre os grupos de linhagens sensíveis (S) e resistentes (R) ao tratamento com 5-FU (S: n = 4; R: n = 4) e Oxaliplatina (S: n = 5; R: n = 5) foi determinada com o uso do pacote estatístico edgeR (v. 2.6.12), implementado na linguagem

R, e que tem sido amplamente utilizado em análises de expressão diferencial em experimentos de RNA-seq. Os valores de expressão (soma do número de *reads* para cada miRNA) são normalizados entre as amostras e modelados de acordo com uma distribuição binomial negativa e um modelo linear generalizado (GLM).

As linhagens celulares que constituem os grupos S e R foram consideradas como réplicas biológicas. Foram incluídos apenas os miRNAs que apresentaram expressão maior que 1 cpm em no mínimo 3 amostras. Consideramos os miRNAs diferencialmente expressos aqueles que apresentaram diferenças estatisticamente significativas entre os grupos ( $FDR \leq 0.05$ ) e com valores de *fold-change*  $> 2$  em pelo menos 70% das comparações entre cada amostra pertencente aos grupos S e R.

#### 3.5.4 Identificação de genes-alvo preditos para os miRNAs diferencialmente expressos

Com o objetivo de investigar as possíveis funções associadas aos miRNAs que apresentaram expressão significativamente alterada entre grupos de linhagens resistentes e sensíveis às drogas, utilizamos o algoritmo de predição de alvos TargetScan 6.2 para identificar os potenciais genes alvo dos miRNAs. O TargetScan baseia-se em propriedades biológicas da interação miRNA-mRNA para atribuir *scores* que refletem uma maior probabilidade de que a interação seja de fato verificada experimentalmente. *Scores* mais negativos representam maiores probabilidades de que o sítio ou um conjunto de sítios-alvo sejam alvos para o respectivo miRNA.

Para selecionar os genes candidatos, consideramos os alvos conservados em vertebrados com *score*  $\leq -0.10$  e os alvos não conservados com *score*  $\leq -0.40$ . A escolha desses critérios é justificada por validações experimentais já obtidas para diversos alvos que apresentam *scores*

neste intervalo e o uso de critérios semelhantes em outros estudos (Dar *et al.* 2011; Jansen *et al.* 2011; Lu e Clark 2012).

### 3.5.5 Análises dos alvos dos miRNAs diferencialmente expressos

Os genes-alvo dos miRNAs diferencialmente expressos foram submetidos à análise de enriquecimento de categorias funcionais do Gene Ontology (GO) e vias de sinalização do KEGG (Kanehisa *et al.* 2012) através da ferramenta DAVID. A ferramenta agrupa as categorias funcionais enriquecidas em uma lista de genes de interesse em comparação com o conjunto de genes controle, neste caso, todos os genes humanos com anotações funcionais. A significância é dada pelo teste exato de Fisher, corrigida para múltiplas comparações. Nesta análise, foram selecionados os termos enriquecidos com  $P < 0.01$  e FDR  $< 5\%$  para as categorias pertencentes ao ramo “processo biológico” do GO e as vias de sinalização do KEGG com  $P < 0.01$  e FDR  $< 10\%$ .

## 3.6 Validação experimental da expressão de miR-342 e efeitos fenotípicos em linhagens resistentes e sensíveis à Oxaliplatina

As etapas experimentais apresentadas a seguir foram conduzidas pela Dra. Lilian Tiemi Inoue, sob a supervisão do Dr. Érico Tosoni Costa, pesquisadores do Instituto de Ensino e Pesquisa (IEP) do Hospital Sírio-Libanês e Instituto Ludwig de Pesquisa sobre o Câncer, respectivamente.

### 3.6.1 *Cultivo celular*

A linhagem celular HCT116 foi cultivada em meio McCoy's 5A e MEM (Minimum Essential Media) suplementado com 10% soro fetal bovino, em incubadora a 37°C e 5% CO<sub>2</sub>, enquanto a linhagem SW620 foi cultivada em meio L-15 (Leibovitz) suplementado com 10% soro fetal bovino, em incubadora a 37°C e 100% ar.

### 3.6.2 *Ensaio de viabilidade celular por exclusão de iodeto de propídeo*

Foram plaqueadas 2-3 x 10<sup>5</sup> células da HCT116 ou SW620 em placas p35 (*corning*). Após 24h, os meios foram substituídos por meios de cultura frescos na ausência ou presença de 10µM de oxaliplatina. Após 72h de tratamento, os sobrenadantes de cada placa foram coletados em um tubo, onde foram adicionadas as células tripsinizadas. As amostras foram centrifugadas e lavadas com PBS pH 7,3 (137mM NaCl; 2,7mM KCl; 10mM Na<sub>2</sub>HPO<sub>4</sub>; 1,8mM KH<sub>2</sub>PO<sub>4</sub>). As células foram ressuspensas em 1ml de PBS e foi adicionado 1µl de solução de 2mg/ml de iodeto de propídeo (PI) a cada amostra. Após incubação no gelo e no escuro por 5min, foram realizadas as análises em citômetro de fluxo FACSCalibur, utilizando o programa Cell Quest (*Becton Dickinson*). As células negativas para marcação com PI são consideradas viáveis.

### 3.6.3 *Ensaio de viabilidade celular por formação de colônias*

Os tratamentos com Oxaliplatina foram efetuados conforme descrito no item anterior. Após 72h de tratamento, 500 células em cada condição (controle e tratadas) foram semeadas em placas p35 (*Corning* – 35mm). Após 7-10 dias, as células foram lavadas com PBS e fixadas

com 1ml de 3,7% formaldeído diluído em PBS, por 15min à temperatura ambiente. Em seguida, as placas foram lavadas em água corrente e as células incubadas com 1ml de solução de 0,1% cristal violeta e 2% etanol, por 10min à temperatura ambiente. A contagem do número de colônias foi feita visualmente. Foram realizados três ensaios independentes, em duplicata e a análise estatística foi feita pelo teste t de Student.

#### 3.6.4 *Análise da progressão no ciclo celular por citometria de fluxo*

Os tratamentos com Oxaliplatina foram efetuados conforme descrito no item 3.6.2. Após 72h de tratamento, os sobrenadantes de cada condição foram coletados em um tubo, onde foram adicionadas as células tripsinizadas. As células foram lavadas com PBS e, em seguida, fixadas em 0,5ml de PBS e 4,5ml de 70% (v/v) de etanol gelado. Após incubação em gelo por no mínimo 2h, as amostras foram lavadas com 5ml de PBS e centrifugadas a 400g por 5min. Os *pellets* celulares foram ressuspensos em 0,3ml de solução de coloração de DNA preparada a fresco (20µg/ml iodeto de propídeo e 0,2mg de RNase diluídos em PBS). Depois de 30min de incubação à temperatura ambiente e no escuro, as amostras foram analisadas no citômetro de fluxo FACSCalibur, utilizando o programa computacional Cell Quest (*Becton Dickinson*). O conteúdo de DNA foi avaliado utilizando o detector FL2 em uma escala linear. Para eliminar os agregados celulares, a população de células a ser analisada foi selecionada a partir de um histograma bivariado demonstrando a área (FL2A) versus a largura (FL2W) do sinal FL2.

### 3.6.5 Transfecções de oligonucleotídeos

Para as transfecções celulares, foi utilizado o *kit* HiPerfect Transfection Reagent (*Qiagen*), de acordo com as instruções do fabricante. Basicamente, cerca de  $1,5 \times 10^5$  células HCT116 ou SW620 foram semeadas em placa p60 e cultivadas por 24h. Em seguida, foram incubados 420nM de miScript Inhibitor Negative Control (*Qiagen*) ou miScript miRNA inibitor anti-hsa-miR-342-3p (*Qiagen*) ou miScript miRNA *mimic* Syn-hsa-miR-342-3p (*Qiagen*) com 100µl de meio de cultura sem soro fetal bovino e 12µl do reagente de transfecção. Após incubação de 5min à temperatura ambiente, os complexos de transfecção foram adicionados gota-a-gota às células em cultura. Após cultivo por mais 24h, foram adicionados 10µM de oxaliplatina às células, que foram cultivadas por mais 48h antes do ensaio de análise de ciclo celular (descrito no item 3.6.4).

### 3.6.6 Extração de RNA, miRNA e síntese de cDNA

As extrações de RNA e miRNA foram realizadas utilizando o kit miRNeasy mini kit (*Qiagen*) de acordo com as instruções do fabricante. A qualidade e a integridade dos RNA extraídos foram analisadas aplicando-se cerca de 300ng de RNA total em gel de 1% agarose. Foram considerados íntegros os RNAs que apresentavam as bandas correspondentes aos RNA ribossômicos 28S e 18S evidentes e na razão 2:1. Para eliminar a contaminação com DNA genômico dos RNAs extraídos, estes foram tratados com DNase (*Ambion*) de acordo com as instruções do fabricante. Os produtos das reações foram analisados em gel de 8% acrilamida corado com prata e a ausência do fragmento de tamanho esperado (250pb) indicava a eficiência na eliminação do DNA genômico contaminante. Para a síntese de cDNA, um a dois microgramas de RNA total de cada amostra foram incubados a 65°C durante 5min na presença

de 250ng de oligo(dT), 50ng de iniciadores randômicos e 0,77mM dNTP, em um volume final de 13µl. Em seguida, as amostras foram incubadas em gelo e foram adicionados às mesmas 1x First Strand Buffer, 0,01M DTT, 40U RNase OUT Recombinant RNase Inhibitor (*Invitrogen*) e 200U de SuperScript III RT (*Invitrogen*). As reações foram incubadas a 50°C por 1h e depois a 70°C por 15min. Para a síntese de cDNA a partir dos miRNA, foi utilizado o *kit* miScript II RT (*Qiagen*), de acordo com as instruções do fabricante. Basicamente, a um micrograma de RNA foi adicionados 1x miScript HiSpec buffer, 1x miScript nucleics mix, miScript *reverse transcriptase* e água DEPC para um volume final de 20µl. Após a incubação a 37°C por 60min, a enzima foi inativada por incubação das amostras a 95°C por 5min.

### 3.6.7 *Análise de expressão gênica pela metodologia de PCR em tempo real*

Para avaliar os níveis de expressão de miR-342 nas amostras, foi utilizado o *kit* miScript SYBR Green PCR (*Qiagen*). Aproximadamente três nanogramas de cDNA foram incubados com 1x QuantiTect SYBR Green PCR Master Mix, 1x miScript universal primer, 1x miScript primer assay e água DEPC em volume suficiente para completar 25µl. Em seguida, a reação foi realizada em termociclador ABI Prism 7300 Sequence Detection System (*Applied Biosystems*).

Para a análise dos níveis de expressão do gene *Dnmt1*, utilizado como controle para verificar a eficiência da inibição de miR-342, foi desenhado um par de iniciadores específicos: *Dnmt1* F (5' GCTACCTGGCTAAAGTCAAA) e *Dnmt1* R (5' CCATTCCCACTCTACGG 3'). O gene *HMBS* foi utilizado como normalizador da reação, sendo utilizados os iniciadores *HMBS* F (5' GGCAATGCGGCTGCAA 3') e *HMBS* R (5' GGGTACCCACGCGAATCAC 3'). As concentrações ideais para a reação de PCR em tempo real de cada um desses pares de iniciadores foram previamente padronizadas. Em cada reação de amplificação, foram utilizados 10µl de SYBR Green (*Applied Biosystems*); um mix de iniciadores *Dnmt1* F e R a 400nM ou

um mix de iniciadores HMBS F e R a 300nM; e 100ng de cada cDNA. Todas as reações foram feitas em duplicata, incluindo a do controle negativo, onde foi utilizada água ao invés de cDNA (controle sem molde).

### 3.6.8 Detecção da expressão de genes relacionados com vias apoptóticas

Para avaliar o efeito na via apoptótica da superexpressão de miR-342 na linhagem SW620, utilizamos um sistema de detecção que baseia-se na tecnologia de SYBR green por PCR em tempo real e disponibiliza placas de 96-wells contendo 84 pares de primers relacionados com genes apoptóticos, cinco relacionados com genes controle de expressão constitutiva (do inglês ‘housekeeping genes’) e três com controle de qualidade. Os genes analisados estão apresentados na Quadro 2 (Apêndice). O protocolo utilizado acompanha o kit e é fornecido pelo fabricante (Superarray Bioscience) disponível em <http://www.sabiosciences.com/>. Basicamente, converte-se o RNA total em amostras de cDNA através do *kit* comercial RT<sup>2</sup> First Strand Kit (Superarray Bioscience). Mistura-se as amostras de cDNA com a solução comercial “RT<sup>2</sup> SYBR Green qPCR Master Mix” e distribui-se alíquotas iguais de 25ul em cada *well* da placa de “Apoptosis PCR Array System” (Superarray Bioscience). A análise dos resultados foi feita após o exame das curvas de dissociação e através do cálculo dos valores de Ct fornecidos para cada curva de amplificação. Finalmente a comparação par-a-par de cada gene expresso foi avaliada pelo método de  $\Delta\Delta Ct$ .

### 3.7 Infraestrutura do Laboratório de Bioinformática e linguagens de programação utilizadas

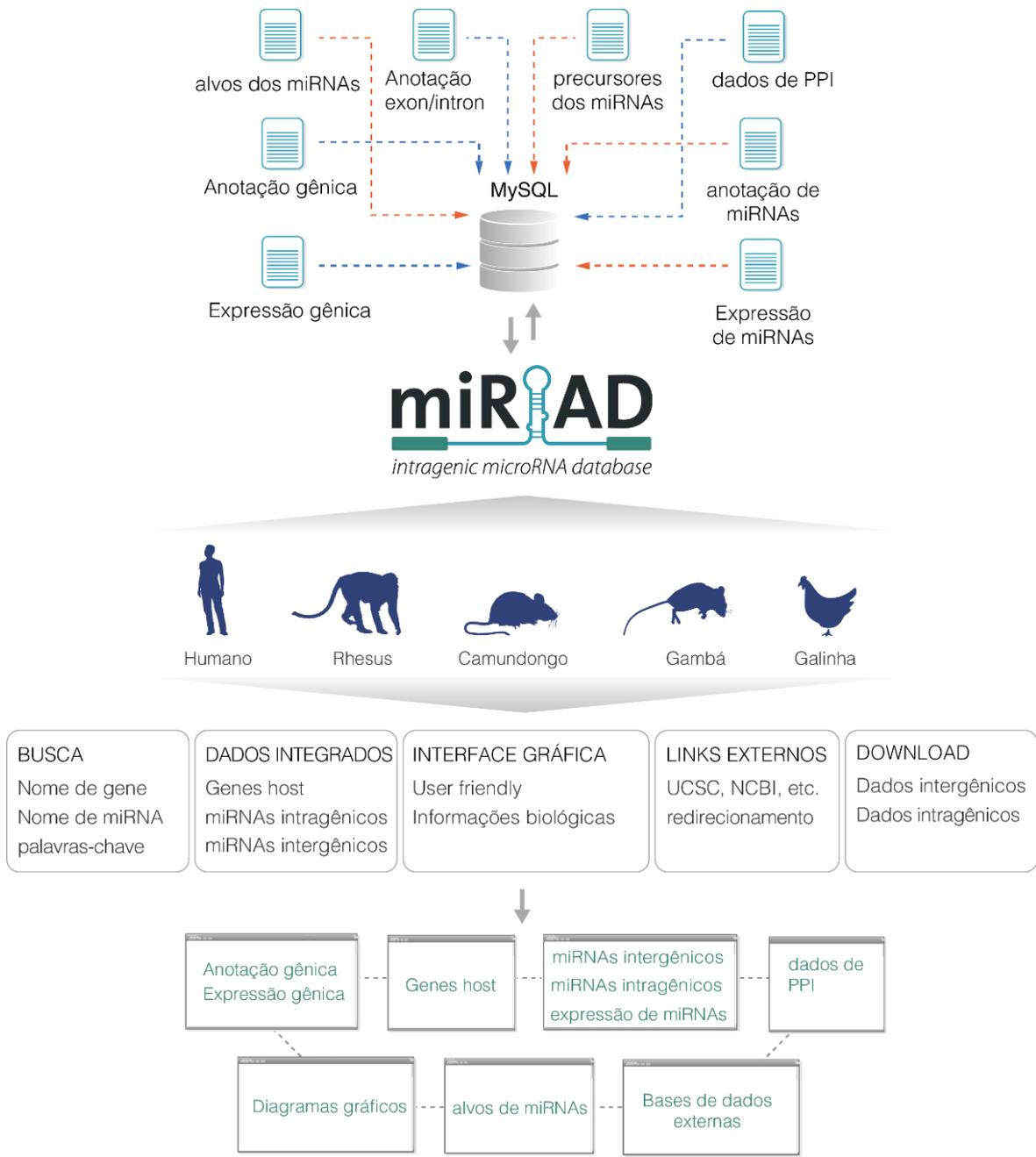
Todas as análises computacionais e processamento de dados biológicos que compõem este trabalho foram realizados utilizando a estrutura do Laboratório de Bioinformática do Centro de Oncologia Molecular do IEP – Instituto de Ensino e Pesquisa do Hospital Sírio-Libanês, chefiado pelo Dr. Pedro A. F. Galante. Atualmente, a estrutura conta com ~300 CPUs, ~150TB de capacidade de armazenamento de dados (*storage*) e mais de 1TB de memória RAM, se somados todos os servidores. Toda a estrutura está conectada em rede, funcionando como um *cluster*, e todos os computadores são baseados no Sistema Operacional (SO) Linux (<https://en.wikipedia.org/wiki/Linux>) e softwares instalados e gerados pelos membros do grupo seguem licenças de software livre GNU (<http://www.gnu.org/licenses/gpl-3.0.en.html>). Exceto quando foi necessário o uso de ferramentas publicamente disponíveis (ex. SAMtools, BEDTools, BLAT, etc.) para execução de tarefas gerais, a manipulação de dados e construção de *pipelines* (encadeamento de comandos e/ou programas para uma finalidade específica) envolveu a confecção de programas em linguagem Python, R e Shell *Scripts*.

## *4 Resultados*

## 4.1 miRIAD: uma ferramenta para a análise de miRNAs intragênicos

### 4.1.1 Características gerais da ferramenta

A estrutura da ferramenta está organizada de forma esquemática na Figura 8, evidenciando as camadas de informação que a compõe. A Figura 9 ilustra a página inicial de acesso, a partir da qual o usuário pode realizar buscas e navegar em outras seções. O foco principal dessa ferramenta é disponibilizar informações integradas de miRNAs intragênicos e seus genes hospedeiros (*host genes*), sumarizadas na Tabela 1. De forma geral, disponibilizamos dados de genes codificantes de proteínas (~22 mil genes em média), miRNAs anotados (~900 em média), alvos preditos de miRNAs, interações proteína-proteína e dados de expressão de miRNAs e genes codificantes para cinco e seis tecidos, respectivamente, para cinco espécies de vertebrados. Do total de miRNAs anotados, 57% (1072) são intragênicos na espécie humana, 29% (167) em macaco rhesus, 63% (745) em camundongo, 40% (179) em gambá e 52% (299) em galinha; destes 84, 54, 87, 88 e 84%, respectivamente, estão na mesma orientação transcricional dos genes hospedeiros. É importante ressaltar que muitas das discrepâncias nessas porcentagens devem refletir vieses de anotação, e não necessariamente um padrão evolutivo. Como é possível notar, as porcentagens de miRNAs intragênicos são similares entre humanos e camundongo, uma vez que são as duas espécies cuja anotação para miRNAs e genes codificantes é mais completa em relação às demais. Na página de cada gene hospedeiro ou miRNA intragênico, é possível visualizar a estrutura gênica e a posição do miRNA em cada isoforma, além da correlação da expressão entre gene hospedeiro e miRNA, miRNAs intragênicos que possuem como alvo seu próprio *host* e/ou parceiros de interação com seu gene hospedeiro. Mais detalhes da ferramenta podem ser obtidos diretamente em: <http://www.miriad-database.org/>



**Figura 8. Estrutura da ferramenta miRIAD e como as camadas de informação estão organizadas.** Setas em azul indicam dados relacionados a genes codificantes de proteínas e setas em laranja referem-se a dados de miRNAs. PPI: *protein-protein interactions*.

**Tabela 1. Sumário dos dados apresentados na ferramenta miRIAD.**

Dados	Humano	Rhesus	Camundongo	Gambá	Galinha
Genes codificadores de proteínas	20.53	22.553	29.664	20.55	16.953
Precursos de miRNAs	1.871	582	1.181	443	573
miRNAs intragênicos	1.072	167	745	179	299
miRNAs intergênicos	799	415	435	264	272
Genes hospedeiros	930	141	613	143	273
miRNAs na orientação sentido	902	90	645	145	90
miRNAs na orientação antisenso	170	77	95	12	28
Genes codificadores expressos	18.442	8.112	19.029	12.079	11.278
miRNAs expressos	1.111	475	784	405	465

**miRIAD**  
intragenic microRNA database

Home Download Statistics FAQ/Help Credits

Enter a **Gene** or a **microRNA** name:

Human SEARCH

examples

**About miRIAD**

miRIAD is a web search tool designed to help users to access integrated information concerning intragenic microRNAs and their host genes. For further details, please check out the [manuscript](#) published in *Database - The Journal of Biological Databases and Curation*.

**Browse data**

- » [Intragenic microRNAs](#)
- » [Intergenic microRNAs](#)
- » [Host genes](#)

Copyright © 2014 - miRIAD: intragenic microRNA database  
Supported browsers: Chrome - Firefox - Safari - IE8+ - Opera

home | download | statistics | faq & help | credits

**Figura 9. Página inicial da aplicação web.**

#### 4.1.2 Sistema de busca da ferramenta miRIAD

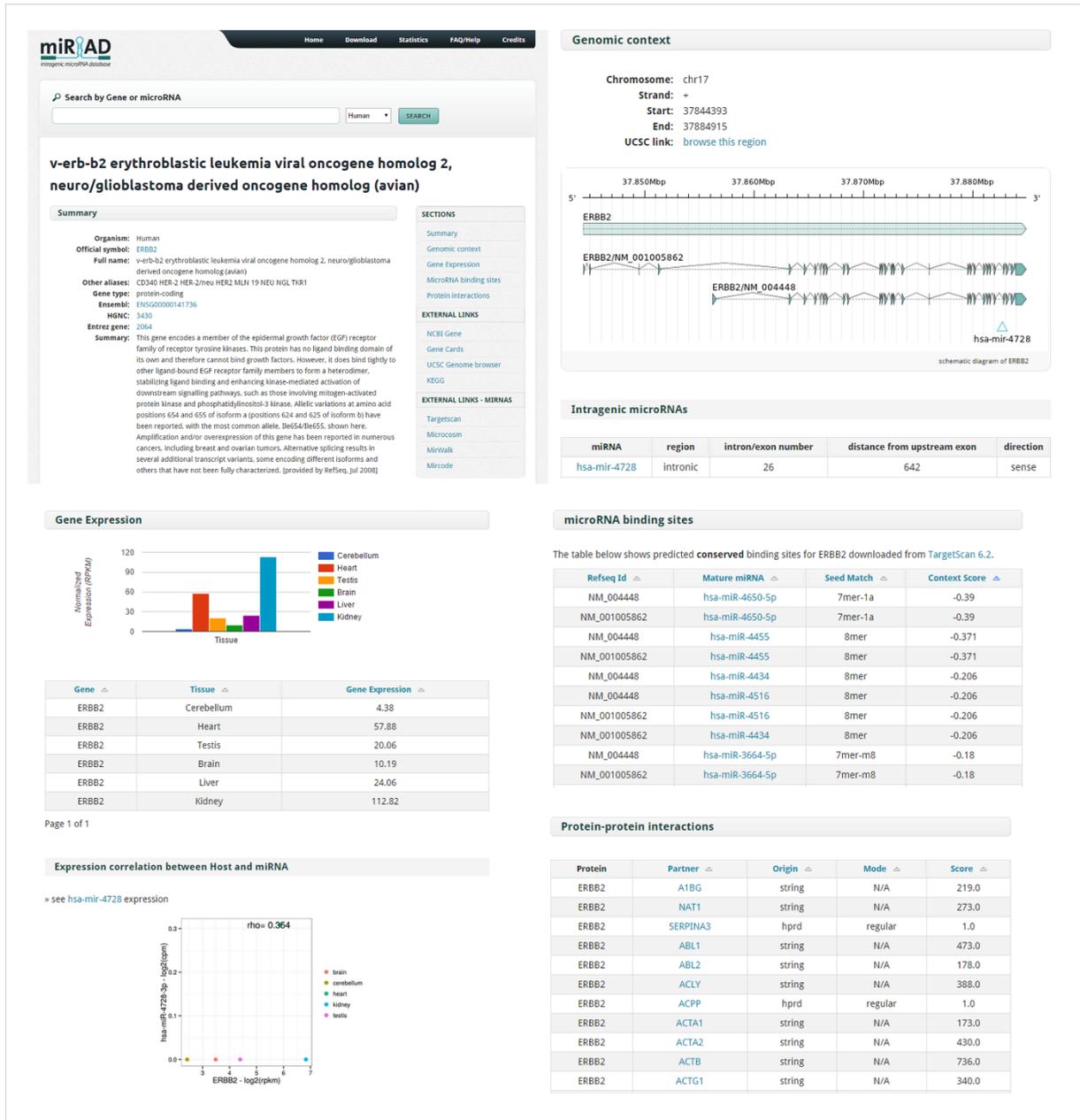
O sistema de busca foi desenvolvido e otimizado para ser eficiente, intuitivo, rápido e flexível, permitindo ao usuário buscar por diferentes termos, tais como nome oficial do miRNA e identificadores de genes (*Official symbol*, Ensembl ID, Entrez ID, HGNC ID ou *Gene Synonyms*) e palavras-chave (ex: ‘oncogene’, ‘kinase’, ‘transcription factor’, etc.). A flexibilidade da busca permite que termos não-exatos sejam procurados (*substrings*), retornando todos os resultados compatíveis com o termo desejado. Buscas por múltiplos termos ou listas de genes também são permitidas. Os resultados das buscas são listas de genes ou miRNAs, organizados por ordem de relevância compatível com o termo buscado (*input*). Os miRNAs intragênicos e genes hospedeiros são identificados por um símbolo específico, enquanto os demais genes e miRNAs também são mostrados. Detalhes das regras e de como proceder as buscas podem ser encontrados em:

<http://www.miriad-database.org/miriad/help#search>

#### 4.1.3 Genes codificadores de proteínas

Evidências recentes têm demonstrado que o contexto genômico dos genes hospedeiros pode revelar aspectos funcionais importantes sobre os miRNAs intragênicos (Baskerville e Bartel 2005; Hinske *et al.* 2010; Dill *et al.* 2012). Para cada gene, uma seção ‘Summary’ é exibida, contendo os identificadores mais comuns (*Official Symbol*, *Full name*, *Aliases*, Ensembl ID, HGNC e Entrez gene), bem como a descrição sumária das funções desempenhadas pelo gene (Figura 10). Além disso, informações sobre o contexto genômico, tais como a posição genômica de início e término de transcrição e orientação transcricional são evidenciados por representações gráficas das estruturas exon-intron (Figura 10). No caso

dos genes hospedeiros, são mostrados os miRNAs contidos em regiões intrônicas ou exônicas, suas posições relativas, a distância do miRNA ao exon a montante mais próximo e orientação (senso ou anti-senso). Também são apresentados dados de expressão gênica baseados em RNA-Seq para seis tecidos (cérebro, cerebelo, coração, rim, fígado e testículo), além da correlação de expressão com miRNAs intragênicos. Por fim, são listados os miRNAs que potencialmente ligam-se aos mRNAs alvo, além de interações proteína-proteína preditas. Casos em que os parceiros de interação de um gene hospedeiro são alvos preditos de um miRNA intragênico correspondente são destacados. Este tipo de informação é potencialmente útil porque pode indicar possíveis interações regulatórias envolvendo a participação destes três componentes (miRNA, gene hospedeiro e parceiros de interação do gene hospedeiro). Na seção destinada aos genes codificantes, há *links* contexto-específicos os quais redirecionam o usuário para as respectivas páginas nas bases de dados externas como NCBI Gene (<http://www.ncbi.nlm.nih.gov/gene>), UCSC Genome Browser (<https://genome.ucsc.edu/>), KEGG (<http://www.genome.jp/kegg/>) e TargetScan (<http://www.targetscan.org/>). As informações disponibilizadas pelo miRIAD para o exemplo do gene ERBB2 e seu miRNA intragênico mir-4728 estão representadas na Figura 10.

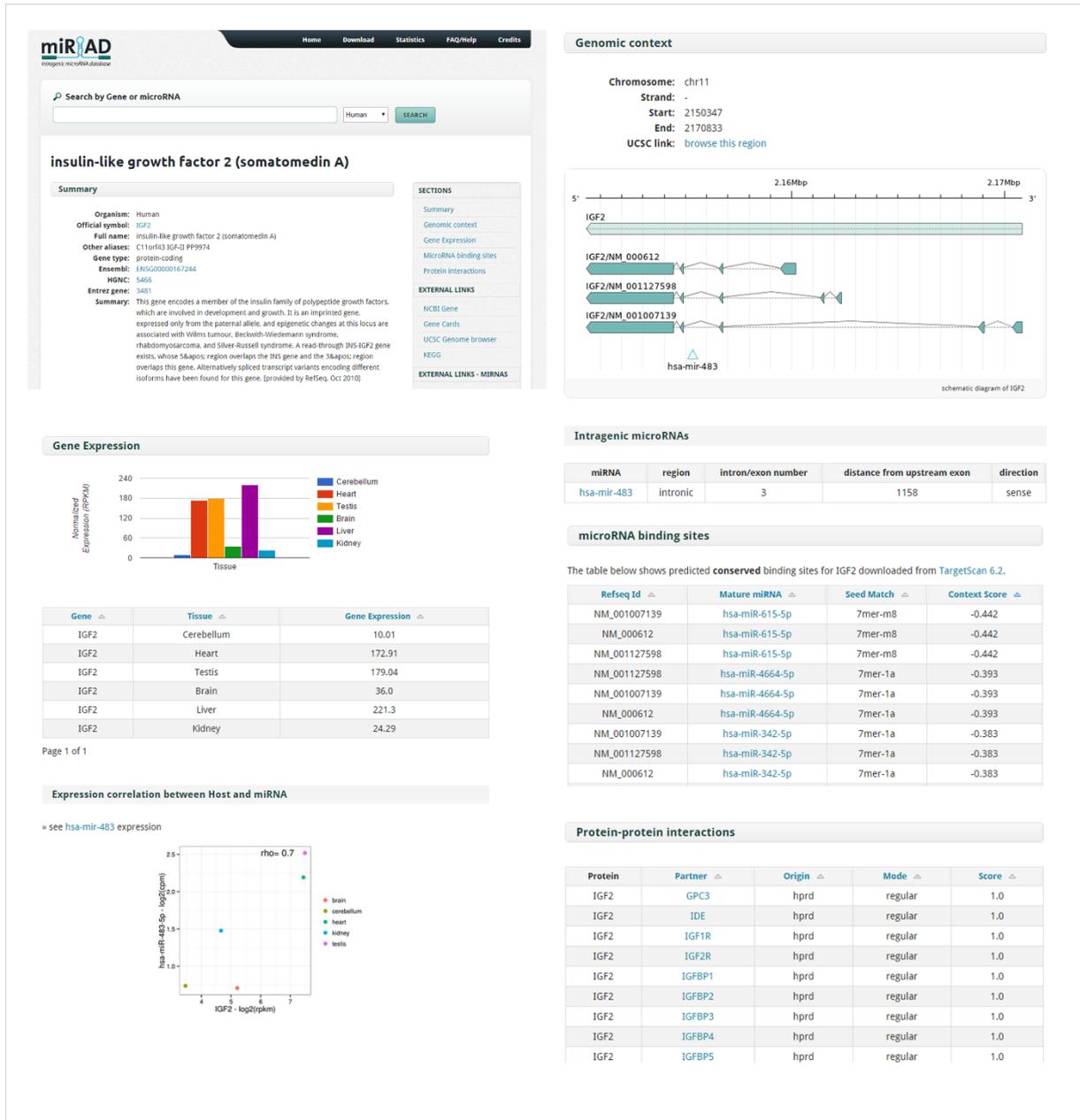


**Figura 10. Sumário das principais informações apresentadas para o gene ERBB2 e seu miRNA intragênico mir-4728.**

### 3.1.4 miRNAs intragênicos

De maneira equivalente aos genes codificantes, para cada pre-miRNA são apresentadas informações como o símbolo oficial, miRBase ID, contexto genômico e genes-alvo preditos. No caso dos miRNAs intragênicos, informações sobre o gene hospedeiro e sua posição relativa

são mostradas. A página dos miRNAs também contém dados de expressão de RNA-Seq para cinco tecidos (cérebro, cerebelo, coração, rim e testículo) e as correlações de expressão dos tecidos em comum com os genes hospedeiros. Além disso, uma série de *links* dinâmicos são disponibilizados para outras ferramentas e bases externas, tais como miRBase, miRDB, TargetScan, miRDIP, miRWalk, Tarbase, etc. (para a lista completa, ver o *website* da ferramenta). Essas informações estão reunidas e exemplificadas para o mir-483 e seu gene hospedeiro IGF2 (Figura 11). O gene IGF2 é um exemplo interessante captado pela nossa ferramenta. Este gene codifica para um fator de crescimento essencial para o crescimento e desenvolvimento do feto, e em geral encontra-se altamente expresso em diversos tumores malignos (Huff 2011). De acordo com os dados de expressão apresentados, há evidência de co-expressão entre o IGF2 e o mir-483 (coeficiente de correlação  $\rho = 0.7$ ). De fato, um trabalho recente demonstrou a existência de um *feedback* positivo entre o IGF2 e o mir-483, mostrando que ambos são co-expressos e que o miRNA tem como alvo a região 5' UTR do RNA mensageiro do próprio IGF2. Essa ligação do mir-483 à 5'UTR de IGF2 facilita a associação da helicase DHX9, promovendo a transcrição do seu próprio gene hospedeiro (Liu *et al.* 2013).

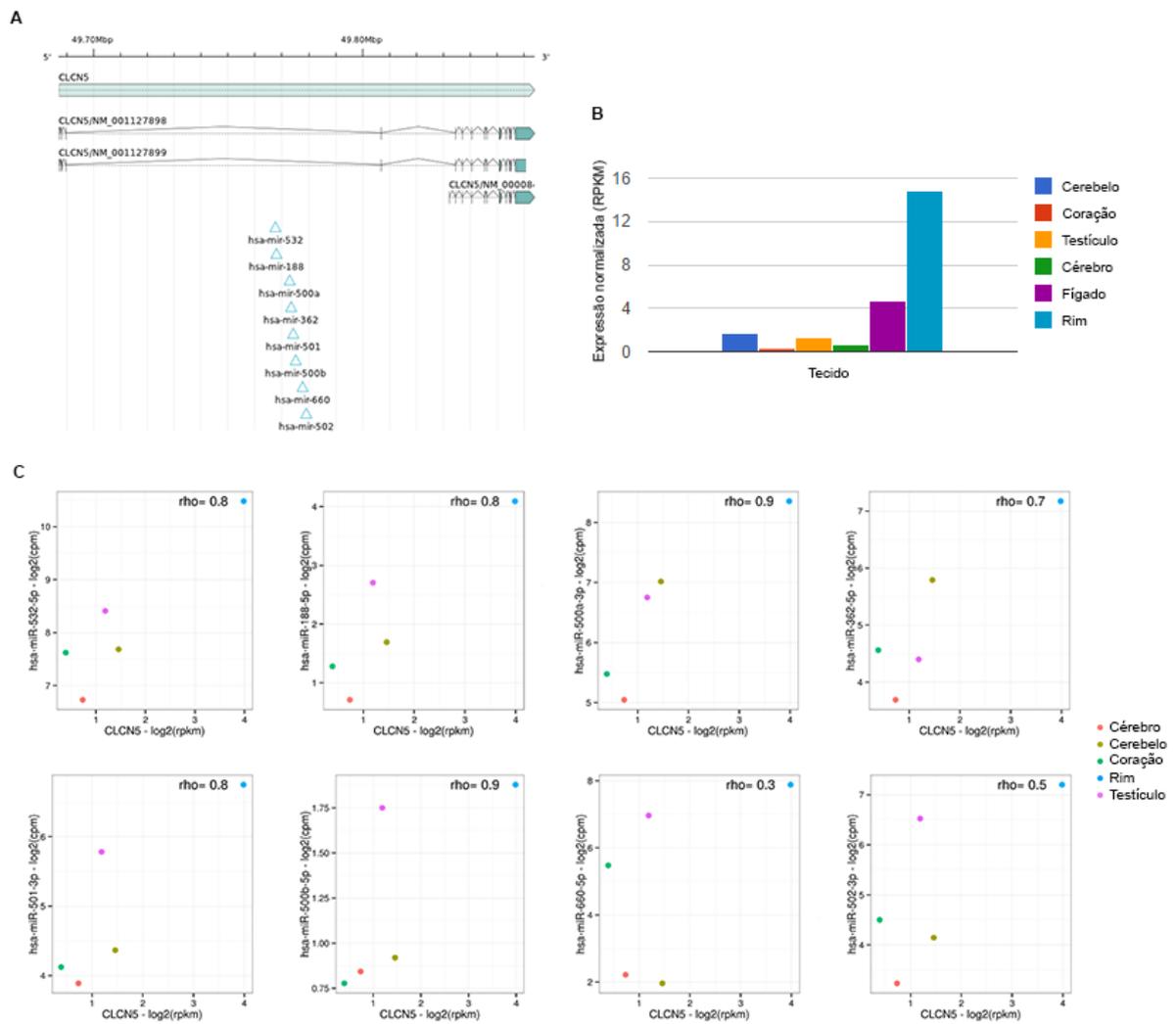


**Figura 11. Sumário das principais informações apresentadas na ferramenta para o mir-483 e seu gene hospedeiro IGF2.**

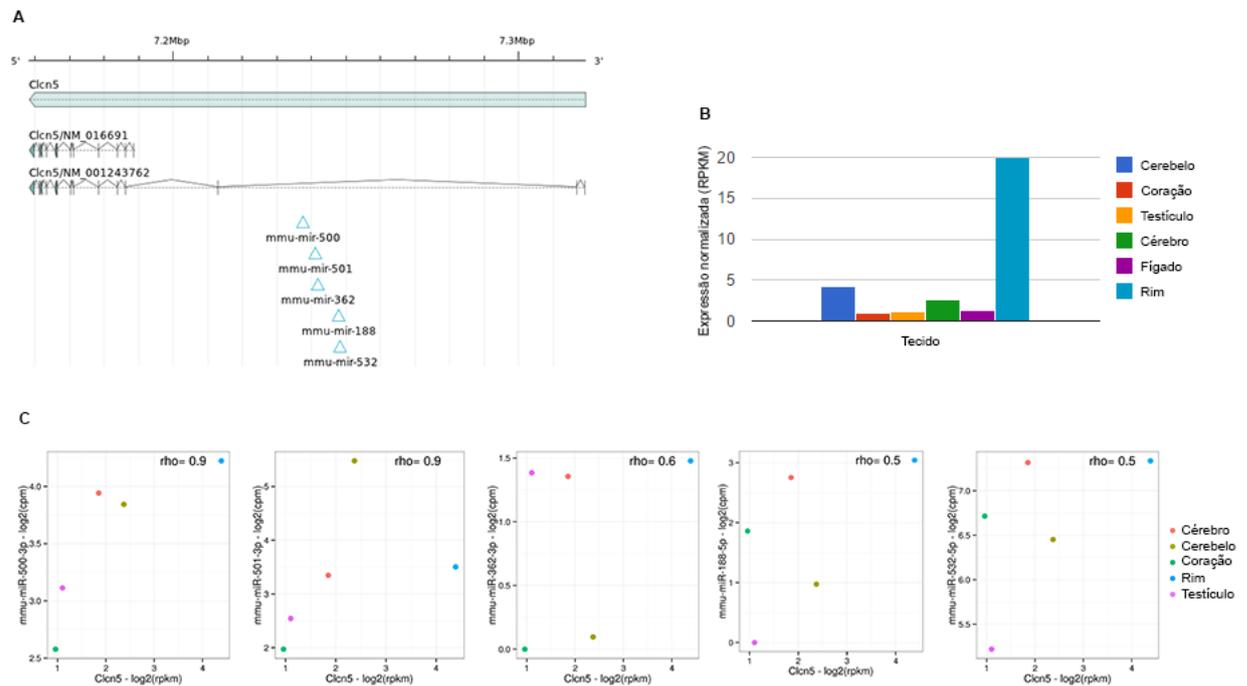
### 3.1.5 Uso da ferramenta miRIAD para explorar genes e miRNAs

Para demonstrar a utilidade prática do miRIAD, usamos um conjunto de genes que codificam para proteínas de superfície celular, catalogados por da Cunha *et al.* (2009). Estes genes são potencialmente interessantes como alvos para diagnóstico e terapias, uma vez que a

acessibilidade à ação de drogas pode ser facilitada pela exposição das proteínas codificadas na membrana celular (Rettig 1992). Do total de 3.702 genes identificados como codificantes para proteínas de superfície, 119 foram classificados como genes hospedeiros para 150 miRNAs, dos quais 87,3% são transcritos na mesma orientação. Em seguida, examinamos em mais detalhes os dois genes que contém o maior número de miRNAs intragênicos anotados, o CLCN5 e HTR2C. O gene CLCN5 codifica para um canal de cloro voltagem-dependente, assumindo importante papel na função dos túbulos renais. Mutações neste gene foram associadas com a doença renal conhecida como *Dent disease* (Grand *et al.* 2011). Este gene possui oito miRNAs intrônicos (Figura 12A), e alguns de seus transcritos possuem alvos preditos para um de seus miRNAs intragênicos, o miR-502. É interessante notar que este gene possui uma isoforma que tem seu início de transcrição a jusante dos miRNAs intragênicos (Figura 12A), o que poderia estar relacionado com uma possível ausência de expressão desses miRNAs se este transcrito for preferencialmente expresso em algum tecido ou em patologias; uma hipótese que pode ser investigada em estudos subsequentes. A análise da expressão do CLCN5 (Figura 12B) e dos seus miRNAs intragênicos sugerem co-expressão de pelo menos 6 dos 8 miRNAs intragênicos ( $\rho > 0.7$ ), sendo que tanto o gene hospedeiro quanto todos os seus miRNAs intragênicos têm os maiores níveis de expressão no tecido renal (Figura 12C). Enfatizando a importância da inclusão de outras espécies, é possível notar que o gene ortólogo *Clcn5* em camundongo possui cinco miRNAs intragênicos anotados, dos quais três possuem alta correlação de expressão com o gene hospedeiro ( $\rho > 0.6$ ) (Figura 13), sugerindo que a regulação desse gene e de alguns dos miRNAs é conservada, o que pode indicar uma relevante implicação funcional.

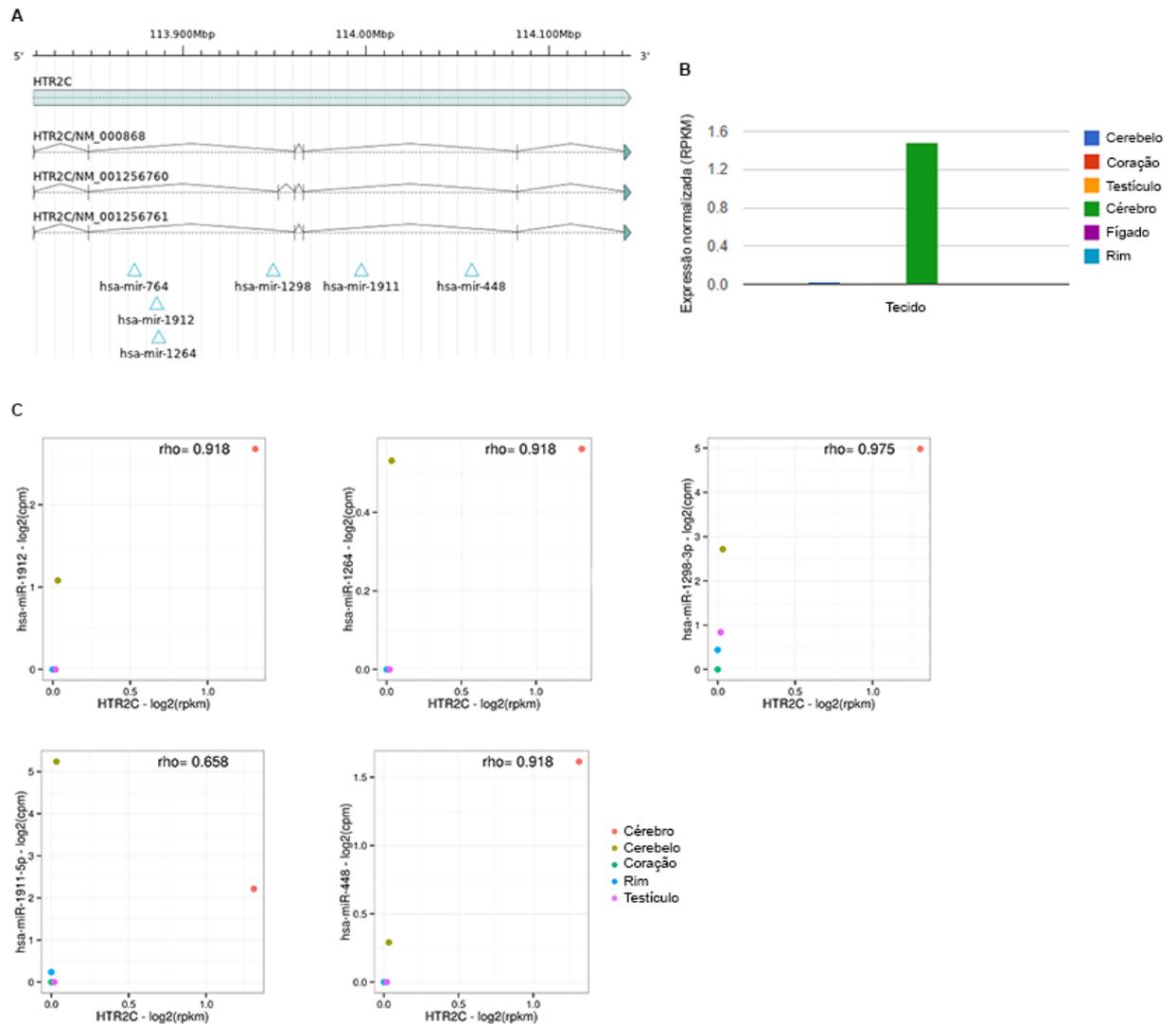


**Figura 12. Representação do gene CLCN5 humano e seus miRNAs intragênicos. A)** Diagrama das isoformas do gene e as posições dos miRNAs intragênicos nos transcritos. **B)** Expressão de CLCN5 em 6 tecidos. **C)** Correlações de expressão do gene CLCN5 e dos seus miRNAs intragênicos.



**Figura 13. Representação do gene *Clcn5* em camundongo e seus miRNAs intragênicos. A)** Diagrama das isoformas do gene e as posições dos miRNAs intragênicos nos transcritos. **B)** Expressão de *CLCN5* em 6 tecidos. **C)** Correlações de expressão do gene *Clcn5* e dos seus miRNAs intragênicos.

Outro gene investigado, o *HTR2C*, codifica para um receptor de serotonina e contém seis miRNAs intrônicos na mesma orientação (Figura 14A). Uma associação funcional entre o *HTR2C* e o miR-448, um dos miRNAs contidos neste gene, foi descrita por Kinoshita *et al.* (2010). O estudo mostrou que a expressão aumentada de *HTR2C* leva ao concomitante aumento da expressão de miR-448, que por sua vez inibe a expressão de *KLF5*, um fator de transcrição importante para diferenciação de adipócitos. Os dados apresentados na ferramenta reforçam a idéia de co-expressão do *HTR2C* e de seus miRNAs intragênicos ( $\rho > 0.65$ ), sendo ambos expressos preferencialmente no cérebro (Figura 14B e C). Os padrões de co-expressão também foram observados para camundongo e gambá (opossum) (ver em <http://miriad-database.org>). Também é interessante notar que o *HTR2C* está envolvido com importantes desordens neuropsiquiátricas (Muiños-Gimeno *et al.* 2011), e portanto as consequências funcionais da expressão conjunta deste gene e seus miRNAs intragênicos merecem ser investigadas.



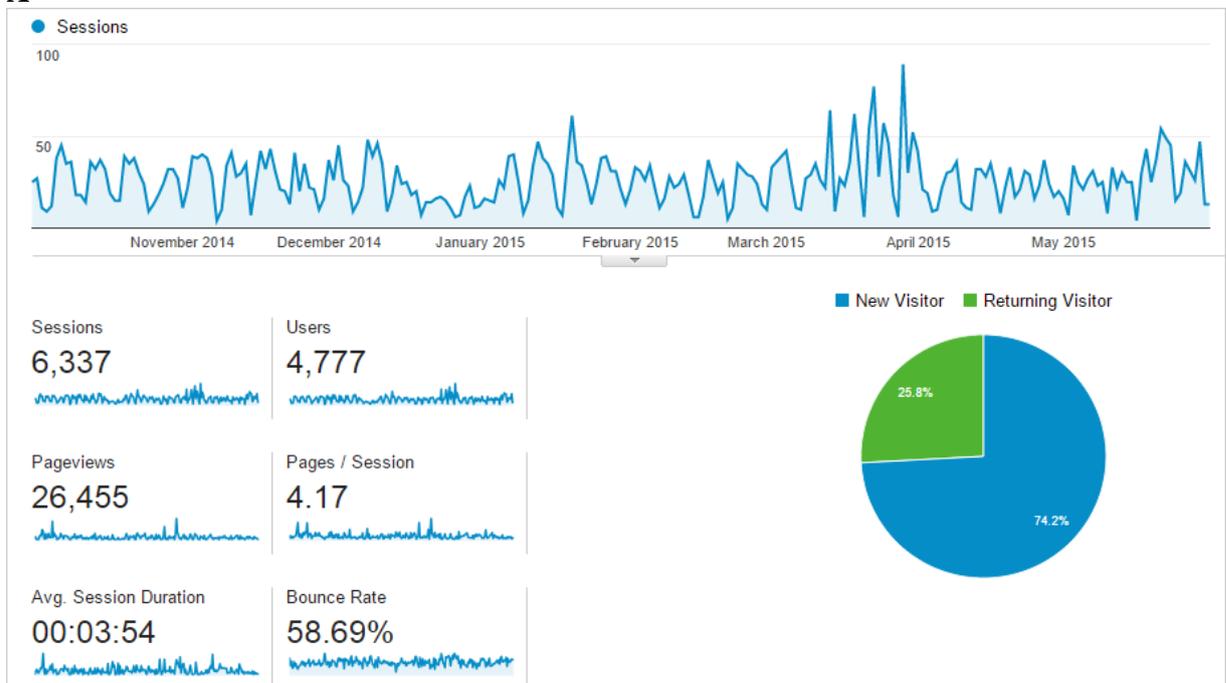
**Figura 14. Representação do gene HTR2C humano e seus miRNAs intragênicos.** A) Diagrama das isoformas do gene e as posições dos miRNAs intragênicos nos transcritos. B) Expressão de HTR2C em 6 tecidos. C) Correlações de expressão do gene HTR2C e dos seus miRNAs intragênicos. Como a expressão do miR-764 não foi detectada, não há correlação para este miRNA.

### 3.1.6 Estatísticas de acesso

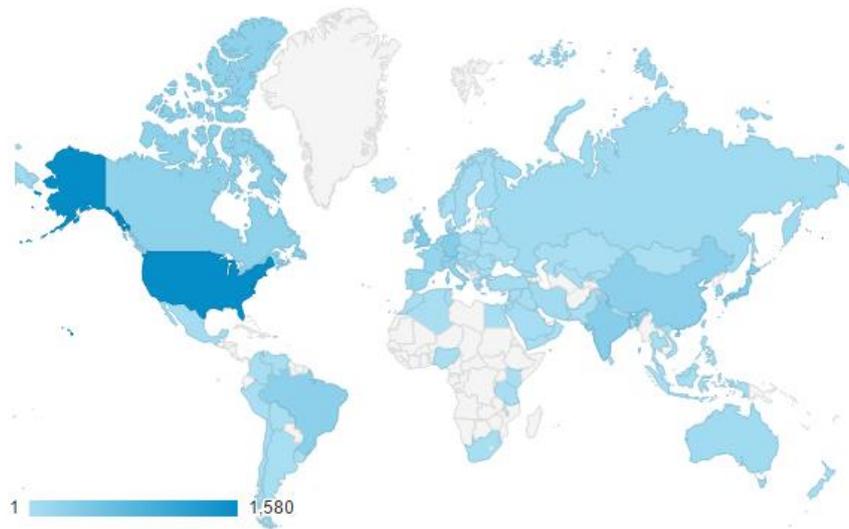
O acesso ao miRIAD é monitorado desde sua implementação através da ferramenta *Google Analytics* (<http://www.google.com/analytics/>), auxiliando na geração de estatísticas que indicam como a ferramenta tem sido utilizada. A Figura 15 mostra estatísticas geradas para o

período de Outubro de 2014 (mês em que o manuscrito foi publicado) até Maio de 2015. Durante este período, percebemos que a ferramenta tem sido bastante acessada, contando com 26.455 visualizações de páginas (média de 110 visualizações por dia) em 6.337 sessões (acessos únicos), sendo que cada sessão tem em média 4,17 visualizações de páginas que duram um tempo médio de ~4 minutos (Figura 15A). Isso indica um importante engajamento dos usuários. Em relação às estatísticas por região geográfica, houve acessos provenientes de praticamente todas as regiões (Figura 15B), sendo que a maior quantidade provém dos Estados Unidos (25% de todas as sessões), seguida do Reino Unido, Japão, Alemanha, Índia e China (Figura 15B). A Figura 15C mostra o número de acessos por página (as 10 mais visitadas no período). É interessante notar que, com exceção da página principal (*home*), a página de *downloads* tem sido a mais frequentemente acessada, indicando possivelmente o uso extensivo dos dados disponibilizados para processamento e análises locais.

A



B



Country ?	Acquisition			Behavior		
	Sessions ? ↓	% New Sessions ?	New Users ?	Bounce Rate ?	Pages / Session ?	Avg. Session Duration ?
	<b>6,337</b> % of Total: 100.00% (6,337)	<b>74.15%</b> Avg for View: 74.04% (0.15%)	<b>4,699</b> % of Total: 100.15% (4,692)	<b>58.69%</b> Avg for View: 58.69% (0.00%)	<b>4.17</b> Avg for View: 4.17 (0.00%)	<b>00:03:54</b> Avg for View: 00:03:54 (0.00%)
1.  United States	<b>1,580</b> (24.93%)	<b>83.80%</b>	<b>1,324</b> (28.18%)	<b>65.25%</b>	<b>4.03</b>	<b>00:02:59</b>
2.  United Kingdom	<b>375</b> (5.92%)	<b>69.33%</b>	<b>260</b> (5.53%)	<b>52.00%</b>	<b>4.54</b>	<b>00:06:24</b>
3.  Japan	<b>369</b> (5.82%)	<b>63.96%</b>	<b>236</b> (5.02%)	<b>56.10%</b>	<b>4.26</b>	<b>00:04:45</b>
4.  Germany	<b>342</b> (5.40%)	<b>81.58%</b>	<b>279</b> (5.94%)	<b>61.11%</b>	<b>3.44</b>	<b>00:02:42</b>
5.  India	<b>329</b> (5.19%)	<b>61.70%</b>	<b>203</b> (4.32%)	<b>50.46%</b>	<b>5.60</b>	<b>00:07:35</b>
6.  China	<b>285</b> (4.50%)	<b>66.67%</b>	<b>190</b> (4.04%)	<b>55.44%</b>	<b>3.08</b>	<b>00:05:07</b>
7.  Brazil	<b>282</b> (4.45%)	<b>59.57%</b>	<b>168</b> (3.58%)	<b>42.20%</b>	<b>5.30</b>	<b>00:03:47</b>
8.  Canada	<b>258</b> (4.07%)	<b>63.95%</b>	<b>165</b> (3.51%)	<b>51.55%</b>	<b>5.70</b>	<b>00:03:53</b>
9.  Italy	<b>245</b> (3.87%)	<b>73.88%</b>	<b>181</b> (3.85%)	<b>52.65%</b>	<b>5.49</b>	<b>00:04:35</b>
10.  South Korea	<b>207</b> (3.27%)	<b>60.87%</b>	<b>126</b> (2.68%)	<b>51.69%</b>	<b>4.54</b>	<b>00:05:10</b>

C

Page ?	Pageviews ?	Unique Pageviews ?	Avg. Time on Page ?	Entrances ?	Bounce Rate ?
	26,455 % of Total: 100.00% (26,455)	21,873 % of Total: 100.00% (21,873)	00:01:14 Avg for View: 00:01:14 (0.00%)	6,312 % of Total: 100.00% (6,312)	58.69% Avg for View: 58.69% (0.00%)
1. /	647 (2.45%)	642 (2.94%)	00:03:44	601 (9.52%)	88.18%
2. /miriad/	570 (2.15%)	414 (1.89%)	00:01:14	212 (3.36%)	29.25%
3. /miriad	526 (1.99%)	414 (1.89%)	00:00:50	388 (6.15%)	20.88%
4. /miriad/downloads	207 (0.78%)	178 (0.81%)	00:02:00	27 (0.43%)	66.67%
5. /miriad/help	132 (0.50%)	109 (0.50%)	00:02:19	23 (0.36%)	43.48%
6. /miriad/statistics	101 (0.38%)	77 (0.35%)	00:02:03	8 (0.13%)	62.50%
7. /miriad/intragenic	79 (0.30%)	60 (0.27%)	00:00:13	2 (0.03%)	0.00%
8. /miriad/credits	57 (0.22%)	49 (0.22%)	00:01:11	9 (0.14%)	55.56%
9. /miriad/miRNA/hsa-mir-6723/	53 (0.20%)	42 (0.19%)	00:03:40	42 (0.67%)	73.81%
10. /miriad-mouse/	48 (0.18%)	35 (0.16%)	00:01:16	19 (0.30%)	57.89%

**Figura 15. Estatísticas de acesso da ferramenta miRIAD.** A) Estatísticas gerais para o número de sessões e páginas visualizadas, porcentagem de usuários novos e os que retornaram ao website durante o período de Outubro de 2014 à Maio de 2015. B) Número sessões abertas de acordo com a localização geográfica e os 10 países de onde provém o maior número de acessos. C) As 10 páginas mais visitadas durante o período.

## 4.2 Análise da influência dos genes hospedeiros e do contexto genômico na expressão e evolução de miRNAs

### 4.2.1 A maioria dos miRNAs intragênicos humanos são específicos de primatas e estão preferencialmente contidos em genes antigos

Com o objetivo de investigar a influência dos genes hospedeiros e do contexto genômico na expressão e evolução dos miRNAs, estes foram classificados em inter- ou intragênicos, dependendo da sobreposição em nível genômico com genes codificantes. Observamos que o

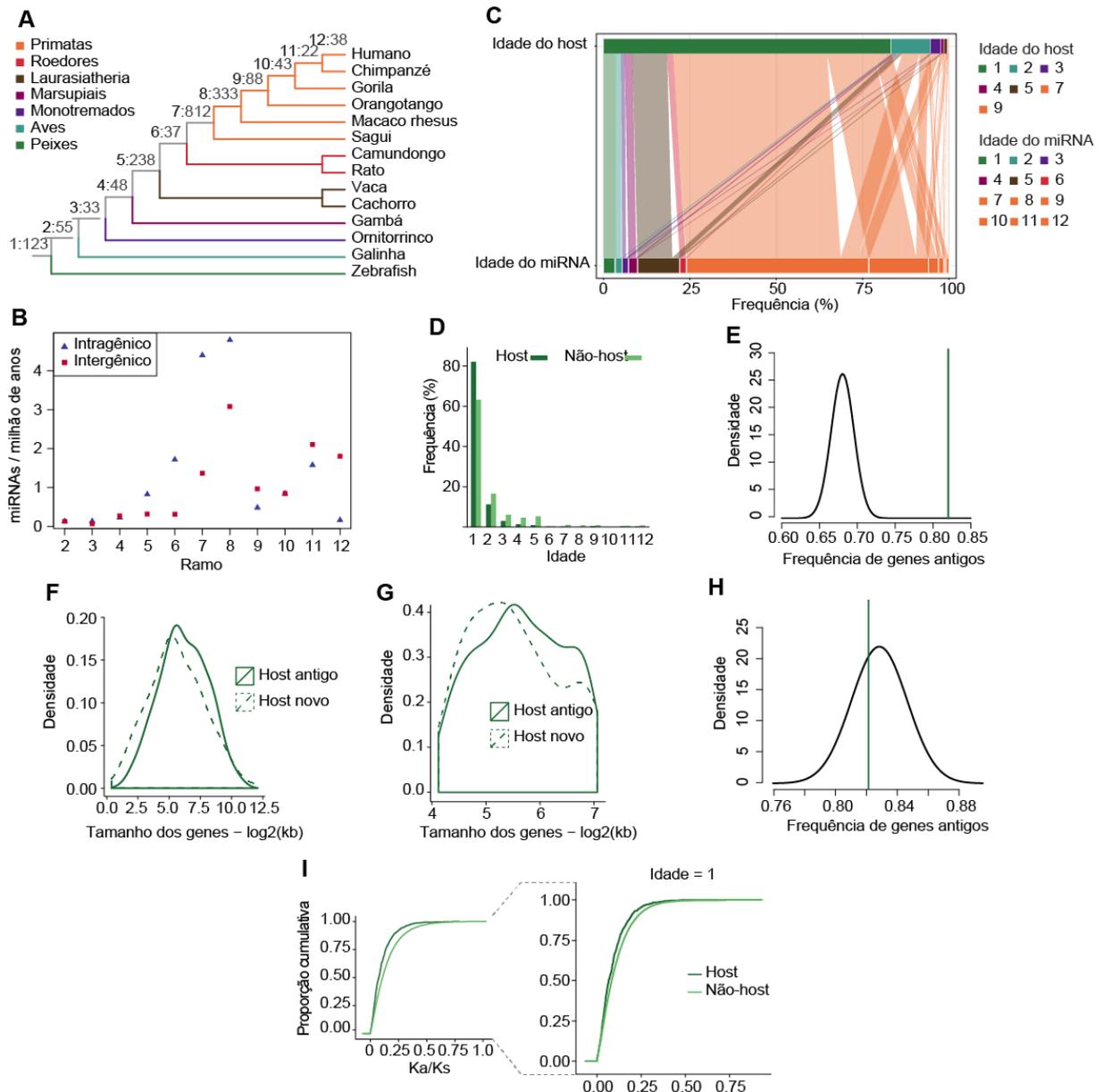
conjunto de miRNAs humanos (miRBase, v.20, N = 1870) é composto por ~40% de intergênicos e ~60% de intragênicos (contidos em 986 genes hospedeiros codificantes), sendo que 84% estão localizados na mesma orientação de transcrição dos genes hospedeiros. Os miRNAs intragênicos estão predominantemente em introns (90%), enquanto os que sobrepõem-se a exons estão, em sua maioria (~80%), em regiões não codificantes (5' UTR e 3' UTR). Com o intuito de analisar os padrões de origem de miRNAs inter- e intragênicos, as idades dos miRNAs humanos foram inferidas utilizando comparações de sequência com os genomas de outras 13 espécies de vertebrados (ver Materiais e Métodos). Com base nos períodos de surgimento, os miRNAs foram classificados em classes de idade (1: vertebrados, 2-4 amniotos, 5-6: mamíferos placentários e 7-12: primatas), usadas ao longo do trabalho em análises comparativas. Em concordância com estudos anteriores (Iwama *et al.* 2013), 85% dos miRNAs humanos emergiram após a divergência de mamíferos placentários (ramos 5-12), enquanto a maioria dos miRNAs humanos anotados (~70%) surgiu em primatas (prevalentemente nos ramos 7 e 8) (Figura 16A).

Considerando que os miRNAs intragênicos tornaram-se mais frequentes em vertebrados, especificamente em mamíferos (Meunier *et al.* 2013), decidimos caracterizar a dinâmica de surgimento de miRNAs inter- e intragênicos comparando a quantidade de miRNAs de cada categoria ao longo da evolução dos vertebrados (Figura 16A). Os resultados mostraram que há taxas distintas de surgimento de miRNAs inter- e intragênicos (considerando igual probabilidade (*i.e.*,  $p = 0,5$ )), especialmente na linhagem dos primatas. A maior taxa de surgimento de miRNAs intragênicos é vista no ramo 7 ( $P < 2,2 \times 10^{-16}$ , teste binomial), enquanto um excesso de miRNAs intergênicos existe nos ramos 9 ( $P = 0,01$ , teste binomial), 11 ( $P = 0,02$ , teste binomial) e 12 ( $P = 0,001$ , teste binomial). Contudo, possíveis cenários como evolução por duplicação de miRNAs e a inclusão de miRNAs sem evidência de expressão (o que poderiam ser considerados “pseudo-miRNAs”), podem afetar diretamente essas

interpretações. Por isso, optamos por considerar apenas os miRNAs com evidência de expressão e agrupar aqueles distantes entre si a menos de 10 kb (ver Materiais e Métodos). Os miRNAs intergênicos, em comparação com os intragênicos, foram mais frequentemente agrupados em *clusters* (odds ratio = 1,4,  $P = 6 \times 10^{-4}$ , teste exato de Fisher). No entanto, a taxa de surgimento de miRNAs intragênicos permaneceu significativamente maior no ramo 7 ( $P < 2,2 \times 10^{-16}$ , teste binomial), e os ramos 5 e 8 também apresentaram excesso de miRNAs intragênicos ( $P < 0,02$ ) (Figura 16B). Esses resultados revelam que miRNAs intragênicos humanos começaram a prevalecer (em relação aos intergênicos) em mamíferos placentários, sendo que a vasta maioria desses miRNAs foi adquirida em primatas, através de um substancial incremento de miRNAs contidos em genes codificantes.

Para testar a hipótese de que a idade dos genes hospedeiros afeta a expressão e evolução dos miRNAs, consideramos a localização genômica dos miRNAs como: intergênicos, intragênicos contidos em genes antigos e intragênicos contidos em genes novos (idade = 1 ou idade  $\geq 2$ , respectivamente; de acordo com Zhang *et al.* (2010) e Zhang *et al.* (2011) (Figura 16A). Observamos uma elevada fração de genes hospedeiros antigos (~83%) (Figura 16C), mesmo considerando apenas os miRNAs expressos e após o procedimento de agrupamento em *clusters*. Isso sugere que miRNAs intragênicos surgem ou são fixados preferencialmente em genes antigos (Figura 16D e Figura 16E;  $P < 0,0001$ , teste de randomização; ou  $P = 3,93 \times 10^{-11}$ , teste de Chi-quadrado). Essa diferença significativa também foi observada quando utilizamos idades de genes definidas por outros dois métodos alternativos (Ensembl e idades obtidas de Chen *et al.* (2012)). Entretanto, sabe-se que genes antigos são maiores que genes novos (Carmel *et al.* 2007; Wolf *et al.* 2009), o que de fato foi confirmado ( $P = 0,002$ , teste de Mann-Whitney; Figura 16F). A tendência dos miRNAs surgirem em genes antigos poderia então ser explicada simplesmente pelo fato desses genes compreenderem regiões genômicas maiores, e por isso encontraríamos um maior número de miRNAs em genes antigos. Para testar

essa hipótese, comparamos a proporção de miRNAs em genes antigos e novos utilizando subconjuntos que não diferem significativamente no tamanho ( $P = 0.27$ , teste de Mann-Whitney, Figura 16G). Mesmo utilizando este controle, observamos uma proporção equivalente de genes hospedeiros antigos em relação ao conjunto total de genes ( $P = 0,63$ , teste de randomização; ou  $P = 0,97$ , teste de Chi-quadrado, Figura 16H), sugerindo que o tamanho do gene não explica a diferença observada. Uma outra abordagem que comprova a idéia de que os genes hospedeiros são mais antigos é através da análise de Ka/Ks. Taxas de evolução são proporcionais aos valores de Ka/Ks, ou seja, quanto menor o valor de Ka/Ks, menor a taxa de evolução (evolução lenta), o que é comumente observado em genes antigos (Albà e Castresana 2005). Curiosamente, observamos que genes hospedeiros tendem a evoluir mais lentamente que os genes não hospedeiros, conforme denotado pela distribuição tendendo a valores de Ka/Ks menores ( $P < 2,2 \times 10^{-16}$ , teste de Kolmogorov-Smirnov (KS); Figura 16I). Este padrão se mantém inclusive quando comparamos somente as distribuições de genes antigos (idade = 1) (Figura 16I). Assim, os resultados apresentados até aqui apontam para um surgimento preferencial de miRNAs intragênicos em genes antigos, independentemente do tamanho e do método de definição de idade de gene, o que nos leva a sugerir que um fator seletivo pode estar associado a este padrão.



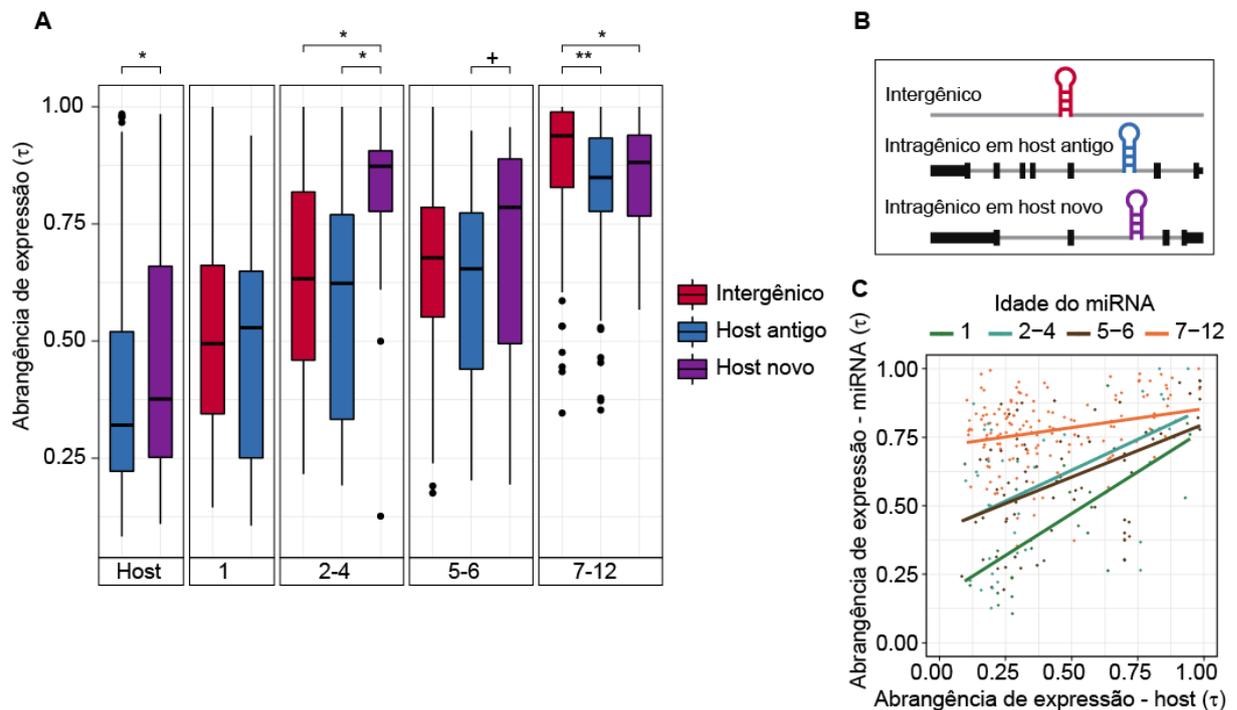
**Figura 16. Idade dos miRNAs e genes hospedeiros humanos.** **A)** Distribuição das idades dos miRNAs humanos em vertebrados, evidenciando o número absoluto de miRNAs que surgiu em cada ramo. **B)** Taxas de ganho de miRNAs inter- e intragênicos. O número de miRNAs por milhão de anos foi calculado com base no tempo de divergência entre os ramos. Os ramos 7 e 8 indicam o maior pico de ganho de miRNAs intragênicos ( $P < 0.01$ , teste Binomial). **C)** Relações de idade dos genes hospedeiros e dos miRNAs intragênicos. O comprimento das linhas horizontais é proporcional à frequência de genes e miRNAs de cada idade. **D)** Frequência relativa referente a cada idade dos genes *host* e *não-host*. Há um excesso significativo de genes *host* antigos (idade = 1) em relação aos *não-host* ( $P < 10^{-7}$ , teste de Chi-quadrado). **E)** Distribuição nula da frequência de genes antigos. Conjuntos de tamanhos iguais de genes tão ou mais antigos que os miRNAs de uma determinada idade foram amostrados randomicamente 10.000 vezes. A proporção de genes antigos observada está representada pela linha vertical verde. **F)** Distribuição do tamanho dos genes hospedeiros. O tamanho do gene foi tomado pela distância entre as coordenadas de início e fim de transcrição representadas em escala log<sub>2</sub>. O transcrito mais longo foi utilizado como referência. **G)** Distribuição do tamanho dos genes hospedeiros controlando pelo tamanho dos genes hospedeiros novos. O IQR (interquartile range) do tamanho dos genes novos foi usado para amostrar genes com tamanhos similares. **H)** Distribuição nula da frequência de genes hospedeiros antigos cujo tamanho não difere significativamente dos genes hospedeiros novos. A linha vertical

representa a frequência observada de genes hospedeiros antigos. **I**) Distribuição de valores de Ka/Ks (taxa de substituição não sinônima pela taxa de substituição sinônima), mostrando que os genes *host* possuem valores de Ka/Ks significativamente menores ( $P < 2.2 \times 10^{-16}$ , Kolgomorov-Smirnov test). Em detalhe, são mostradas as distribuições apenas para genes *host* e não-*host* antigos (idade = 1).

#### 4.2.2 A idade dos genes hospedeiros e o contexto genômico afetam a expressão dos miRNAs

Para investigar a influência do contexto genômico e da idade dos genes hospedeiros na expressão dos miRNAs, nós consideramos as classes de idade previamente definidas (1, 2-4, 5-6 e 7-12) e a posição genômica (inter- e intragênicos). Em seguida, determinamos a abrangência de expressão de miRNAs e de genes hospedeiros utilizando o índice de tecido-especificidade  $\tau$  (ver Materiais e Métodos) em 12 e 16 tecidos/tipos celulares, respectivamente. Os níveis de expressão, e particularmente a abrangência de expressão, são sabidamente correlacionadas com as taxas de evolução dos genes, o que significa que genes conservados tendem a ser altamente expressos e em diferentes tecidos (elevada abrangência de expressão) (Wolf *et al.* 2009; Park *et al.* 2012; Zhang *et al.* 2012). De fato, os genes hospedeiros antigos são mais amplamente expressos (*i.e.*,  $\tau$  menor) do que hospedeiros novos ( $P = 0,03$ , teste de Mann-Whitney, Figura 17A). Além disso, os miRNAs antigos também são mais amplamente expressos que os recentes (idade 1 vs 2-4, 5-6 e 7-12:  $P < 0,002$ ; idade 2-4 vs 7-12:  $P < 2,2 \times 10^{-16}$ ; idade 5-6 vs 7-12:  $P < 2,2 \times 10^{-16}$ , teste de Mann-Whitney; Figura 17A). Comparando a abrangência de expressão dentro da mesma classe de idade, vimos que, apesar do pequeno número de casos ( $N = 13$  para idades 2-4 e 5-6), miRNAs com origem em amniotos (idade  $\geq 2$ ) ou em mamíferos placentários (idade  $\geq 5$ ) contidos em genes novos tendem a apresentar maior tecido-especificidade do que aqueles localizados em genes antigos ou intergênicos (idade 2-4: miRNAs em genes hospedeiros novos vs antigos:  $P = 0,01$  e vs intergênicos:  $P = 0,01$ ; idade 5-6: miRNAs em genes hospedeiros novos vs antigos:  $P = 0,07$ ; Figura 17A, ver Figura 17B para uma ilustração esquemática). Curiosamente, miRNAs intragênicos primata-específicos (idade 7-12) têm maior

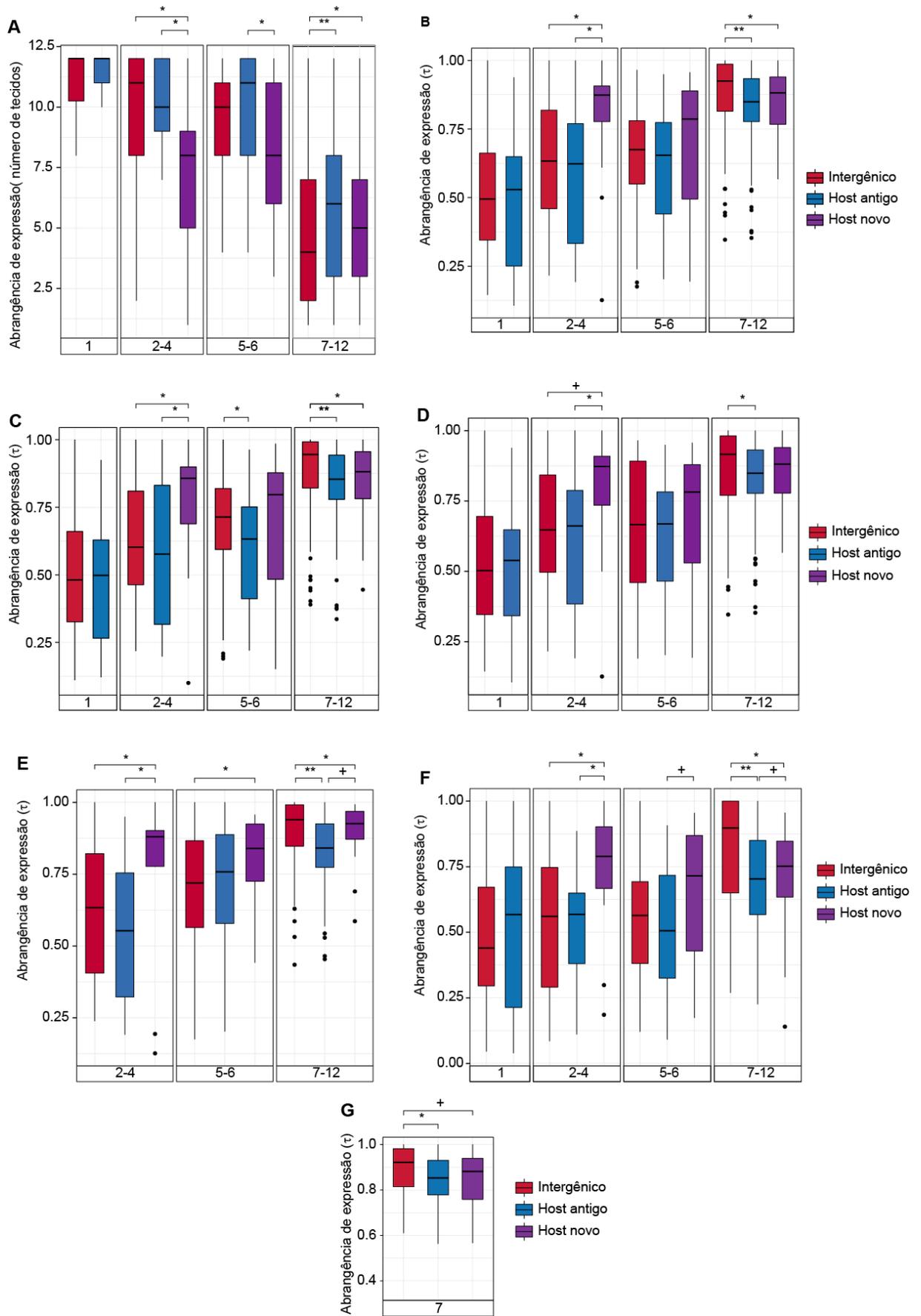
abrangência de expressão que os intergênicos de mesma idade, enquanto a maior diferença é observada para os miRNAs contidos em genes antigos (miRNAs em *hosts* antigos vs intergênicos:  $P = 1,6 \times 10^{-7}$ ; miRNAs em *hosts* novos vs intergênicos:  $P = 0,004$ ; Figura 17A). Esta mesma tendência é observada quando consideramos somente o número de tecidos em que os miRNAs são expressos como medida de abrangência de expressão (Figura 18A).



**Figura 17. Abrangência de expressão de miRNAs de acordo com a idade e o contexto genômico.** **A)** O primeiro painel mostra que genes *host* antigos são mais amplamente expressos que genes novos. Os demais painéis mostram a abrangência de expressão de miRNAs intergênicos (vermelho), intragênicos em *hosts* antigos (azul) e intragênicos em *hosts* novos (roxo) de acordo com a categoria de idade (1: vertebrados, 2-4: amniota, 5-6: mamíferos, 7-12: primatas). MiRNAs mais antigos são mais amplamente expressos (*i.e.*, quanto menor a idade, menor o  $\tau$ ). MiRNAs intragênicos em *hosts* novos são mais tecido-específicos em relação aos intergênicos e intragênicos em *hosts* antigos de mesma idade. MiRNAs novos (7-12) em *hosts* antigos são mais amplamente expressos que miRNAs intergênicos. (+ $P = 0,07$ ; \* $P < 0,05$ ; \*\* $P < 0,001$ , Mann-Whitney U test). **B)** Representação esquemática do contexto genômico dos miRNAs considerado nas análises. **C)** Correlações da abrangência de expressão de miRNAs intragênicos e seus genes hospedeiros (todas as correlações:  $\rho > 0,2$ ,  $P < 0,05$ , correlações de Spearman).

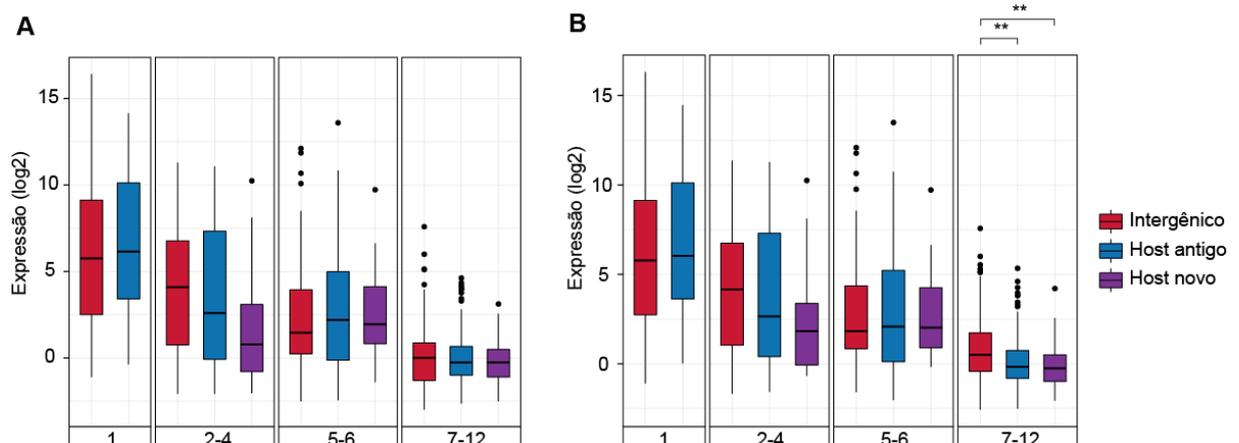
Para verificar a robustez dos resultados observados, recalculamos a abrangência de expressão dos miRNAs utilizando uma série de controles: (1) removendo dados de expressão

de testículo (Figura 18B), já que genes codificantes e não codificantes jovens usualmente apresentam expressão preferencial em testículo (Necsulea e Kaessmann 2014) e muito do que se conhece por transcrição “promíscua” é atribuída à permissividade da conformação da cromatina durante a espermatogênese (Soumillon *et al.* 2013). Dessa forma, excluimos a hipótese de “out of testis”, que prevê que genes novos são expressos primeiramente no testículo e evoluem para padrões de expressão amplos (Kaessmann 2010), o que poderia explicar parte das diferenças observadas na abrangência de expressão dos miRNAs. (2) Removendo dados de expressão de cérebro e cerebelo, evitando assim um viés inerente à complexidade e diversidade do transcriptoma de tecidos neurais (Ramsköld *et al.* 2009) (Figura 18C); (3) agrupando miRNAs cuja distância entre eles é menor do que 10 kb (Figura 18D); (4) utilizando idades de miRNAs definidas por outro método (Iwama *et al.* 2013) (Figura 18E); (5) usando dados de expressão unicamente gerados pelo mesmo estudo (Figura 18F), de modo a reduzir o potencial viés causado pela variabilidade técnica e/ou biológica, e finalmente (6) comparando miRNAs inter- e intragênicos somente de idade 7 (Figura 18G), já que miRNAs intergênicos mostraram-se o tipo mais frequente nos de idade 9-12 (Figura 16B). De modo consistente, todas essas análises dão suporte robusto a ideia de que a abrangência de expressão dos miRNAs é influenciada pela idade do gene hospedeiro.



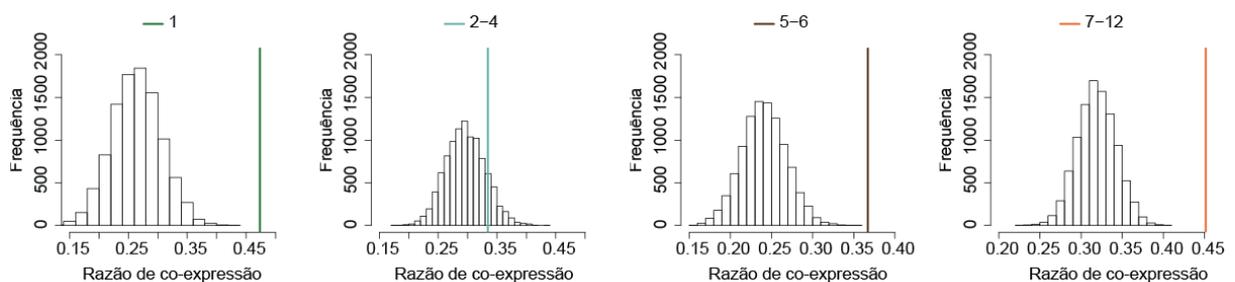
**Figura 18. Controles relacionados à análise de abrangência de expressão de miRNAs.** **A)** Abrangência determinada pela número de tecidos em que os miRNAs são expressos. **B)** Abrangência excluindo-se a expressão proveniente do testículo. **C)** Abrangência excluindo-se a expressão proveniente de cérebro e cerebelo. **D)** Abrangência de expressão com dados de miRNAs agrupados quando distantes até 10 kb. **E)** Abrangência de expressão utilizando idades de miRNAs definidas por Iwama *et al.* (2012). **F)** Abrangência de expressão utilizando dados de RNA-Seq provenientes do mesmo estudo (Meunier *et al.* 2013) para 5 tecidos. **G)** Abrangência de expressão para miRNAs jovens somente de idade = 7. (+)  $0,05 < P < 0,09$ ; (\*)  $P < 0,05$ ; (\*\*)  $P < 0,001$ .

Consistente com os achados acima, observamos correlações positivas e significativas entre a abrangência de expressão de miRNAs e dos seus genes hospedeiros (idade 1:  $\rho = 0,67$ ,  $P = 2,0 \times 10^{-4}$ ; idade 2-4:  $\rho = 0,36$ ,  $P = 0,03$ ; idade 5-6:  $\rho = 0,47$ ,  $P = 2,0 \times 10^{-5}$ , age 7-12:  $\rho = 0,20$ ,  $P = 0,01$ , correlações de Spearman; Figura 17C). Notadamente, a correlação dos miRNAs novos (idade 7-12) é basicamente explicada pela abrangência de expressão dos genes antigos (dado não mostrado). Ao contrário da análise de abrangência de expressão, não foram observadas diferenças significativas nos níveis de expressão entre miRNAs inter- e intragênicos pertencentes a uma mesma classe de idade (Figura 19A). A diferença observada entre os níveis de expressão mais elevados para miRNAs intergênicos novos devem-se essencialmente à expressão tecido-específica em testículo (Figura 19B).



**Figura 19. Nível de expressão dos miRNAs de acordo com a idade e o contexto genômico.** **A)** Boxplots da mediana de expressão dos miRNAs considerando os tecidos em que os miRNAs são expressos ( $> 1\text{cpm}$ ). Nota-se que os níveis de expressão são mais altos para os miRNAs mais antigos. Excluindo-se a expressão em testículo, não há diferenças significativas entre miRNAs pertencentes à uma mesma classe de idade. **B)** Quando consideramos todos os 12 tecidos (i.e. incluindo testículo), observamos que os miRNAs intergênicos jovens são mais expressos que os intragênicos (\*\*  $P < 0,001$ , teste de Mann-Whitney). Essa diferença deve-se, portanto, à expressão testículo-específica.

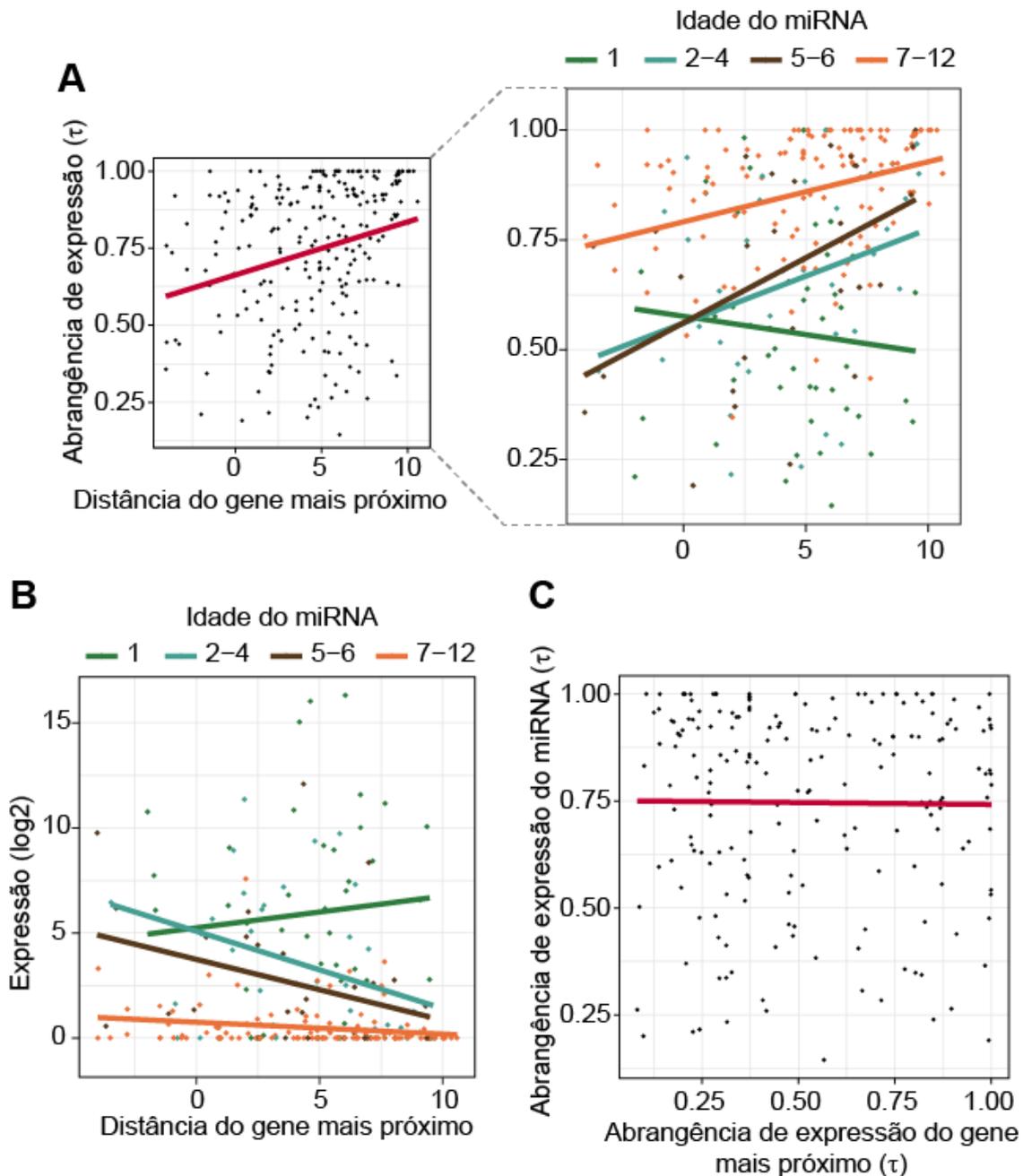
Levando em conta que os miRNAs intragênicos e seus genes hospedeiros são frequentemente co-transcritos (Baskerville e Bartel 2005; Marsico *et al.* 2013), espera-se que as correlações entre a abrangência de expressão dos miRNAs e dos genes nos quais estão inseridos estejam relacionadas com a co-expressão dos dois transcritos. Para testar essa hipótese, realizamos um teste de randomização (ver Materiais e Métodos), o que confirmou uma co-expressão significativa dos miRNAs intragênicos e seus genes hospedeiros em relação ao esperado ao acaso (todas as idades, exceto 2-4:  $P < 1,4 \times 10^{-4}$ ; Figura 20), sugerindo que a influência exercida pelos genes hospedeiros na abrangência de expressão dos miRNAs está de fato relacionada com pares miRNA-gene co-expressos.



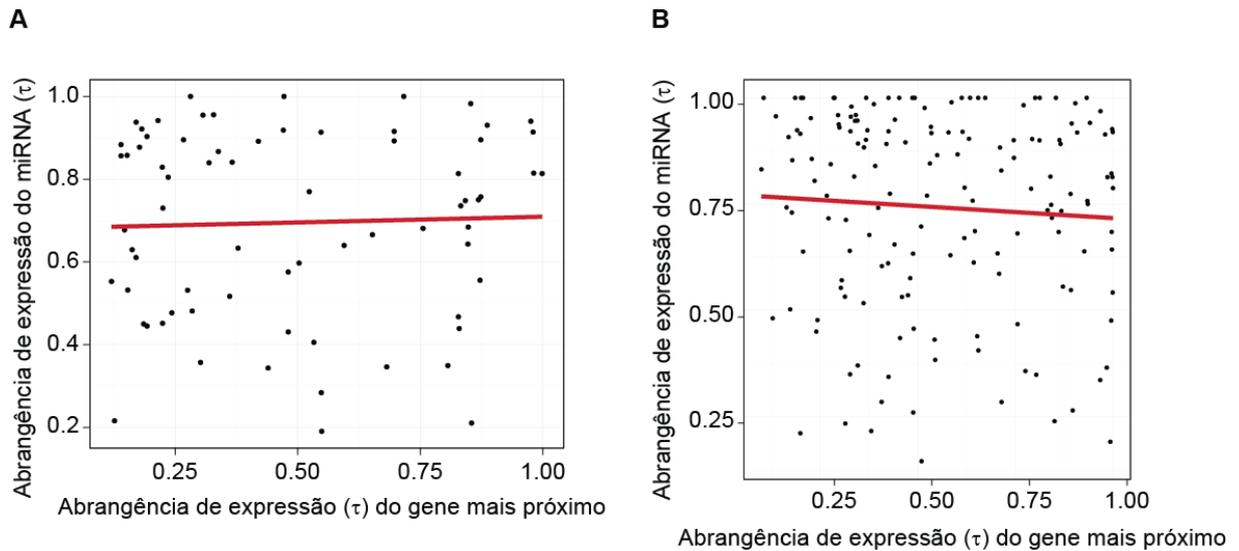
**Figura 20. Co-expressão de miRNAs intragênicos e genes hospedeiros.** Os histogramas representam as distribuições nulas das razões de co-expressão entre miRNAs e seus genes hospedeiros. Para cada par miRNA-gene, a ordem dos tecidos foram ranqueadas de acordo com o nível de expressão. Posteriormente a razão do número de tecidos em que o miRNA e o gene hospedeiro são expressos na mesma ordem foi contabilizada. A média das razões de co-expressão observadas (representadas por linhas verticais) foram comparadas com as médias das razões calculadas por randomização da ordem dos tecidos 10.000 vezes (idade 1:  $P < 0,0001$ ; idade 2-4:  $P = 0,12$ , idade 5-6:  $P < 0,0001$ ; idade 7-12:  $P < 0,0001$ ).

A influência exercida pelo contexto genômico na expressão dos miRNAs intergênicos também foi investigada. Para isso, o contexto genômico foi definido como a distância dos miRNAs para os genes codificantes mais próximos. Observamos que os miRNAs intergênicos distam dos genes mais próximos em dezenas de bases até mais de 1,5 Mb (megabases) (mediana = 34 kb), enquanto ~33% estão distantes até 10 kb. Resultados similares também foram observados para lincRNAs e os genes codificantes mais próximos (Cabali *et al.* 2011). Ao

investigarmos a relação entre a abrangência de expressão dos miRNAs intergênicos com a distância dos genes mais próximos, vimos que a abrangência tende a ser maior à medida que os miRNAs estão mais próximos dos genes ( $\rho = 0,31$ ,  $P = 1,84 \times 10^{-6}$ ; correlação de Spearman; Figura 21A). Correlações positivas e significativas foram observadas particularmente para miRNAs que surgiram após a divergência de aves (idade  $\geq 2-4$ ) (idade 1:  $\rho = -0,13$ ,  $P = 0,30$ ; idade 2-4:  $\rho = 0,36$ ,  $P = 0,03$ ; idade 5-6:  $\rho = 0,48$ ,  $P = 0,006$ ; idade 7-12:  $\rho = 0,36$ ,  $P = 6,37 \times 10^{-5}$ ; correlações de Spearman; Figura 21A). Da mesma forma, os níveis de expressão também estão correlacionados com a proximidade dos genes (idade 1:  $\rho = 0,16$ ,  $P = 0,30$ ; idade 2-4:  $\rho = -0,36$ ,  $P = 0,03$ ; idade 5-6:  $\rho = -0,36$ ,  $P = 0,04$ ; idade 7-12:  $\rho = -0,25$ ,  $P = 0,004$ ; correlações de Spearman; Figura 21B). Em contraste com os resultados encontrados para miRNAs intragênicos e seus genes hospedeiros, a abrangência de expressão dos miRNAs intergênicos e dos genes vizinhos não está correlacionada (Figura 21C), mesmo quando consideramos somente os miRNAs distantes a menos de 10 kb ou quando consideramos os genes mais próximos aqueles a jusante e na mesma orientação dos miRNAs (Figura 22A e B). Diferentemente dos genes que contém miRNAs, não encontramos nenhuma relação entre a idade dos genes e dos miRNAs com a proximidade dos genes vizinhos (dados não mostrados). Portanto, a proximidade dos genes vizinhos parece afetar a expressão dos miRNAs intergênicos (para aqueles de idade  $\geq 2$ ), mas não de maneira coordenada como ocorre com os genes hospedeiros e seus miRNAs intragênicos. Uma explicação alternativa para as correlações observadas pode estar relacionada com a atividade transcricional favorecida pela cromatina aberta ao redor dos genes vizinhos (Ebisuya *et al.* 2008; Cabili *et al.* 2011), ao invés da co-regulação.



**Figura 21. Expressão de miRNAs intergênicos em relação aos genes mais próximos.** **A)** Correlação entre a abrangência de expressão e distância dos genes mais próximos ( $\rho = 0.31$ ,  $P < 0,001$ , correlação de Spearman), mostrando que miRNAs intergênicos tendem a ter maior abrangência de expressão quando estão próximos a genes codificantes. Correlações para cada classe de idade são mostradas em detalhe (idade 2-4, 5-6 e 7-12:  $0,5 < \rho < 0,5$ ;  $P < 0,05$ ). **B)** Correlações entre os níveis de expressão dos miRNAs e a distância dos genes mais próximos para cada classe de idade. O nível de expressão é dado pela mediana da expressão dos miRNAs nos 12 tecidos analisados. Correlações significativas foram observadas para todas as idades (exceto idade 1) ( $-0,36 < \rho < -0,25$ ;  $P < 0,05$ ). **C)** Correlação entre a abrangência de expressão de miRNAs e dos genes mais próximos na mesma orientação transcricional. A correlação não é significativa ( $\rho < -0,03$ ;  $P = 0,6$ ). Para essas análises, miRNAs distantes até 10 kb foram agrupados.

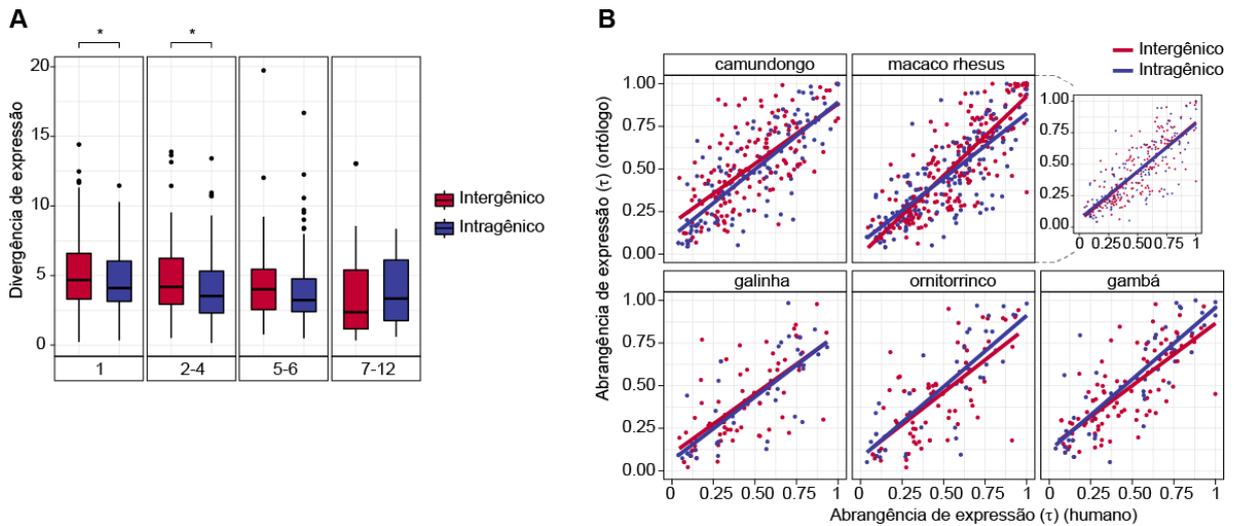


**Figura 22. Abrangência de expressão de miRNAs intergênicos e genes vizinhos. A)** Ausência de correlação entre a abrangência de expressão de miRNAs e dos genes mais próximos, considerando apenas os miRNAs distantes até 10 kb a montante dos genes. **B)** Ausência de correlação entre miRNAs intergênicos e os genes mais próximos considerando aqueles na mesma orientação e a jusante dos miRNAs.

Em resumo, os resultados apresentados até aqui mostram que a idade dos genes hospedeiros afeta a abrangência de expressão dos miRNAs intragênicos. Isso é especialmente relevante para miRNAs jovens (idade 7-12), que são expressos em um maior número de tecidos quando comparados com os intergênicos de mesma idade, sugerindo uma contribuição direta dos genes hospedeiros antigos para esse fenômeno. Aparentemente, os níveis de expressão dos miRNAs intragênicos não são afetados da mesma maneira. Em contrapartida, tanto a abrangência quanto os níveis de expressão dos miRNAs intergênicos estão associados com a proximidade de genes codificantes, o que possivelmente tem relação com o efeito do estado da cromatina nas regiões adjacentes.

#### 4.2.3 Evidências de que os miRNAs intragênicos antigos possuem expressão conservada

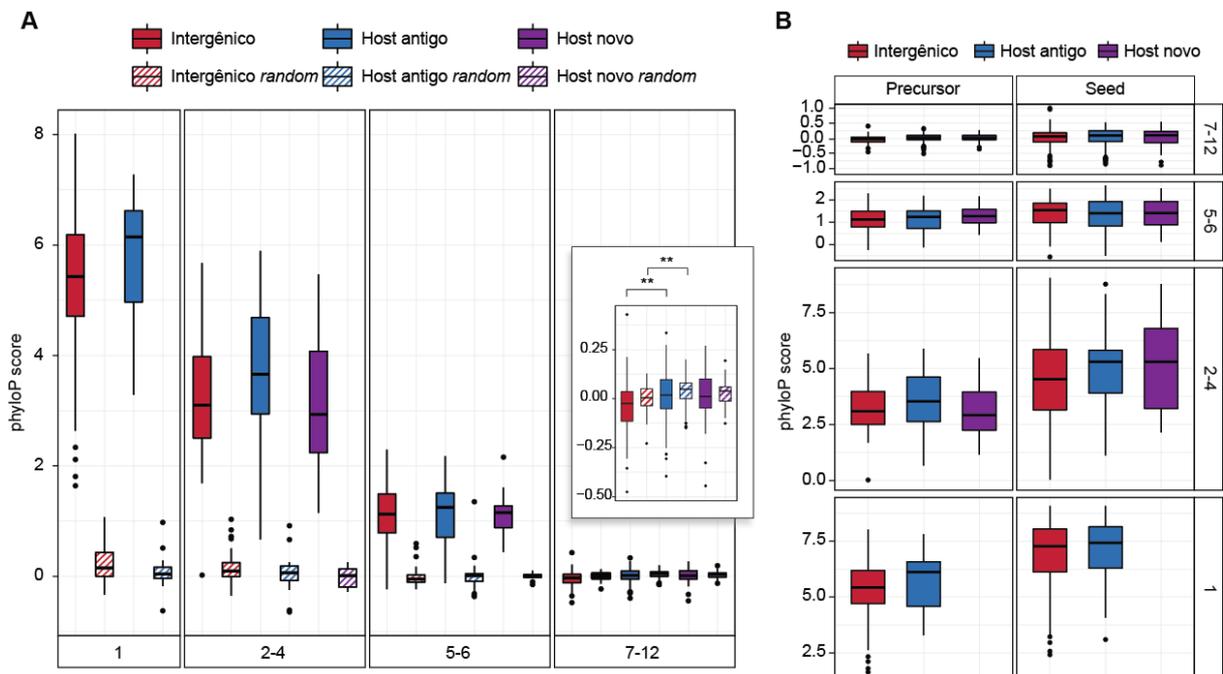
Diferenças na expressão gênica é reconhecidamente um dos fatores que determinam a diversidade fenotípica entre espécies (Carroll 2000). Embora a real contribuição da seleção natural para o estabelecimento dessas diferenças seja discutível (Khaitovich *et al.* 2004), atualmente aceita-se que a regulação gênica evolui sob seleção estabilizadora para um vasto conjunto de genes, o que implica em menor variação de expressão intra e interespecífica (Lemos *et al.* 2005). Neste sentido, questionamos se diferenças na expressão de miRNAs entre espécies são equivalentes para miRNAs inter- e intragênicos. Para responder essa questão, utilizamos os dados de expressão gerados por Meunier *et al.* (2013) para comparar os níveis de expressão dos miRNAs humanos em cinco tecidos (cérebro, cerebelo, coração, rim e testículo) com outras cinco espécies (macaco rhesus, camundongo, gambá, ornitorrinco e galinha). A quantificação da divergência de expressão foi determinada pelas distâncias Euclidianas dos níveis de expressão dos miRNAs humanos e dos seus ortólogos nas demais espécies (ver Materiais e Métodos). Curiosamente, observamos que miRNAs intragênicos antigos apresentam divergência de expressão significativamente menor que os intergênicos (idade 1:  $P = 0,019$ ; idade 2-4:  $P = 0,017$ , teste de Mann-Whitney, Figura 23A). Além disso, ao compararmos a abrangência de expressão dos miRNAs humanos e dos seus ortólogos nas respectivas espécies, observamos correlações com coeficientes maiores para miRNAs intragênicos antigos e de idade intermediária (*i.e.*, excluindo aqueles de idade 7) em detrimento dos intergênicos (diferenças significativas entre as correlações foram obtidas pelo método de transformação Z de Fisher e foram observadas para ornitorrinco ( $P = 0,02$ , teste unicaudal), gambá ( $P = 0,05$ , teste unicaudal) e camundongo ( $P = 0,01$ , teste unicaudal), Figura 23B). Esses resultados sugerem uma maior restrição da expressão (*constraint*) para miRNAs intragênicos quando consideramos períodos evolutivos mais distantes (anterior à divergência dos mamíferos placentários).



**Figura 23. Análise da expressão interespecífica de miRNAs.** **A)** Divergência de expressão entre miRNAs humanos e seus ortólogos em cinco espécies. A divergência de expressão foi calculada pelas distâncias Euclidianas dos níveis de expressão em cinco tecidos (cérebro, cerebelo, coração, rim e testículo) para cada par de ortólogos. Os boxplots representam as distribuições das distâncias conjuntas para cada par miRNA-ortólogo, de acordo com a classe de idade. Diferenças significativas foram avaliadas com testes de Mann-Whitney ( $*P < 0.02$ ). **B)** Correlações de Spearman entre a abrangência de expressão de miRNAs humanos e seus ortólogos em relação ao contexto genômico (inter- e intragênico). As correlações são: Galinha: intergênico:  $\rho = 0,70$ , intragênico:  $\rho = 0,78$ ; Ornitorrinco: intergênico:  $\rho = 0,68$ , intragênico:  $\rho = 0,83$ ; Opossum: intergênico:  $\rho = 0,76$ , intragênico:  $\rho = 0,85$ ; Camundongo: intergênico:  $\rho = 0,69$ , intragênico:  $\rho = 0,81$ ; Macaco Rhesus: intergênico:  $\rho = 0,86$ , intragênico:  $\rho = 0,79$ ; Macaco Rhesus excluindo miRNAs de idade 7 (mostrado em detalhe): intergênico:  $\rho = 0,79$ , intragênico  $\rho = 0,81$  (todas as correlações  $P < 1.0 \times 10^{-9}$ ). Para testar diferenças significativas entre os coeficientes de correlação, utilizamos a transformação Z de Fisher. Diferenças significativas entre correlações de miRNAs inter- e intragênicos foram observadas para Ornitorrinco ( $P = 0,02$ , unicaudal), Gambá ( $P = 0,05$ , unicaudal) e Camundongo ( $P = 0,01$ , unicaudal). Um mínimo de 1 cpm (contagem por milhão) em cada tecido foi adotado como *cutoff* de expressão para ambas as análises (A e B).

Em seguida, investigamos se a restrição de expressão observada para miRNAs intragênicos antigos é acompanhada de uma conservação diferencial em termos de sequência. Para isso, analisamos as distribuições dos PhyloP *scores* definidos para a sequência precursora dos miRNAs, o que revelou forte concordância com nossa definição de classes de idade. Assim, quanto mais antigos são os miRNAs, maiores são os PhyloP *scores* (Figura 24A). De forma geral, não foram observadas diferenças significativas na conservação de sequência entre miRNAs inter- e intragênicos de mesma idade, exceto no caso dos miRNAs intragênicos jovens (idade 7-12), que apresentaram *scores* maiores que os intergênicos (Figura 24A). Entretanto,

essa diferença parece estar relacionada com as regiões genômicas adjacentes, e não com as sequências dos miRNAs em si, conforme indicado pela análise das regiões intragênicas randômicas (Figura 24A) e para regiões imediatamente adjacentes aos miRNAs (dados não mostrados), que também possuem *scores* maiores que regiões intergênicas. Em relação à região *seed* – normalmente a região mais conservada dos miRNAs (Lewis *et al.* 2005; Mohammed *et al.* 2013) –, também observamos diferenças entre as duas categorias (dados não mostrados). Além disso, foi possível verificar que a sequência *seed* é de fato mais conservada que a região precursora, inclusive para os miRNAs jovens (idade 7-12) (Figura 24B). Esses resultados indicam que, a longo prazo, os miRNAs intragênicos podem estar sujeitos a uma menor variação de expressão (maior *constraint*) do que os intergênicos, o que a princípio não está refletido na conservação de sequência. Uma possível explicação para as diferenças de *constraint* de expressão pode estar relacionada com a influência do controle regulatório dos genes hospedeiros antigos (ver Discussão).



**Figura 24. Conservação de sequência dos miRNAs humanos.** A) A conservação é representada pela média dos phyloP scores da sequência precursora. Nota-se que os miRNAs antigos possuem *scores* mais elevados, refletindo a definição das classes de idade. Uma diferença significativa foi detectada entre

miRNAs inter- e intragênicos jovens (idade 7-12) (\*\* $P < 0,001$ , teste de Mann-Whitney). Entretanto, essa diferença também surge quando a comparação é feita com controles randomizados. **B)** Comparação da conservação das sequências ‘seed’ com as precursoras. Todas as comparações (e também com o controle randomizado) revelaram maior conservação das sequências ‘seed’ ( $P < 0,01$ , teste de Mann-Whitney), exceto para miRNAs intragênicos de idade 5-6 e 7-12 em *hosts* novos.

#### 4.2.4 Associações funcionais entre os miRNAs e sua localização genômica

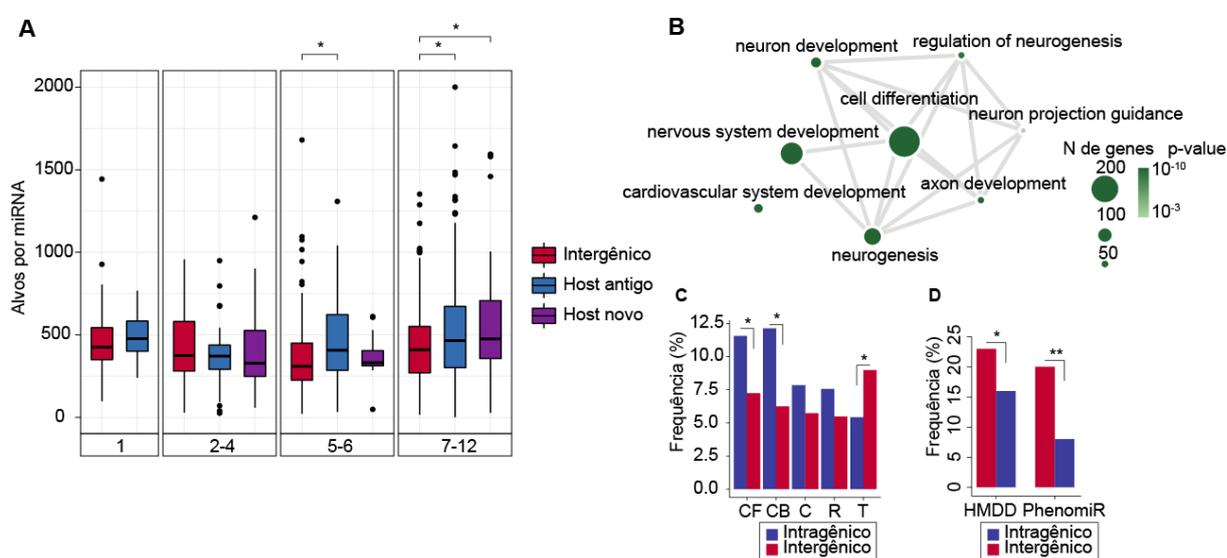
As diferenças observadas entre miRNAs inter e intragênicos motivaram a exploração de possíveis aspectos funcionais relacionados com essas duas categorias. Mostramos, por exemplo, que miRNAs intragênicos jovens possuem abrangência de expressão maior que os intergênicos (Figura 17A). Portanto, é razoável supor que esses miRNAs atinjam um repertório mais diverso de genes-alvo, uma vez que estão expressos em maior número de tecidos. Nós utilizamos o algoritmo TargetScan para predição de alvos de miRNAs, que indicou um conjunto de alvos mais diverso para os miRNAs intragênicos jovens quando comparados com os intergênicos (idade 7-12: intragênicos em *hosts* antigos vs intergênicos:  $P = 0,003$ , teste de Mann-Whitney; Figura 25A). Com base nesses resultados, especulamos que os miRNAs jovens, especialmente aqueles que surgem em genes antigos, por apresentarem maior abrangência de expressão, poderiam regular um repertório mais diverso de alvos, embora abordagens experimentais serão necessárias para corroborar esta idéia.

Considerando que as funções dos miRNAs intragênicos também podem estar relacionadas com as funções desempenhadas pelos genes hospedeiros, investigamos essa possibilidade realizando uma análise de enriquecimento funcional de categorias do Gene Ontology com os genes hospedeiros. Curiosamente, tais genes estão fortemente associados com funções neuronais (Figura 25B). Além disso, uma análise de enriquecimento de expressão com a ferramenta DAVID também revelou uma sobre-representatividade desses genes em cérebro ( $P = 1,1 \times 10^{-11}$ , p-valor corrigido para múltiplos testes com o método de Benjamini-Hochberg). Em seguida, testamos se a proporção de miRNAs inter e intragênicos também encontra-se

enriquecida em algum tecido. Encontramos que os miRNAs intragênicos estão sobre-representados (em relação aos intergênicos) em tecidos neurais (cérebro:  $P = 0,01$  e cerebelo:  $P = 0,002$ , teste exato de Fisher; Figura 25C), um resultado concordante com estudos anteriores que demonstraram expressão cérebro-específica de miRNAs jovens (Meunier *et al.* 2013). Portanto, a conexão entre o papel desempenhado pelos genes hospedeiros em tecidos neurais, com a sobre-representação de miRNAs intragênicos primata-específicos em cérebro e cerebelo, sugere uma associação funcional e uma contribuição conjunta desses genes codificantes e não-codificantes na evolução de interações regulatórias no tecido nervoso. Por outro lado, miRNAs intergênicos jovens (comparados aos intragênicos), estão marginalmente sobre-representados em testículo ( $P = 0,06$ , Figura 25C). Isto talvez esteja ligado a duas vertentes interpretativas sobre expressão testículo-específica, que invocam a expressão facilitada pelo remodelamento da cromatina durante a espermatogênese (Soumillon *et al.* 2013) ou a ação de pressões seletivas que favoreçam este padrão (Necsulea e Kaessmann 2014).

Considerando que a variação da expressão de miRNAs pode estar relacionada com a atividade regulatória dos genes hospedeiros (Baskerville e Bartel 2005), e que diversas doenças têm associações diretas ou indiretas com a expressão alterada de miRNAs, nós vislumbramos a possibilidade dos miRNAs inter- e intragênicos estarem diferentemente representados em doenças humanas. Para verificar essa hipótese, utilizamos informações fornecidas pelo HMDD v.2.0, um banco curado que reporta a participação de miRNAs em doenças. Observamos uma maior proporção de miRNAs intergênicos jovens (em relação aos intragênicos) associados à doenças (idade 7-12:  $P = 0,02$ , teste exato de Fisher; Figura 25D). Resultados similares também foram obtidos com dados provenientes de outro banco curado (PhenomiR) que compila informações de miRNAs com expressão alterada em doenças (idade 7-12:  $P = 1,5 \times 10^{-6}$ , teste exato de Fisher; Figura 25D). MiRNAs intergênicos de idades 1, 2-4 e 5-6, quando agrupados, também estão sobre-representados ( $P < 0,03$ ). Uma explicação para essas diferenças pode estar

relacionada com o modo como miRNAs respondem à perturbações na sua atividade transcricional. Como as regiões regulatórias de genes hospedeiros e miRNAs intragênicos são comumente compartilhadas (Marsico *et al.* 2013), é possível que a expressão de miRNAs intragênicos, especialmente os jovens, seja menos impactada por variações extremas por conta do controle regulatório bem estabelecido dos genes hospedeiros antigos, o que resultaria numa menor variação na expressão desses miRNAs (ver Discussão).



**Figura 25. Associações funcionais previstas in silico para miRNAs inter e intragênicos. (A)** Número de alvos previstos por miRNA de acordo com a idade e localização genômica ( $*P < 0,05$ , teste de Mann-Whitney). **(B)** Análise de enriquecimento de processos biológicos do Gene Ontology associados aos genes hospedeiros, evidenciando funções neuronais. **(C)** Proporção de miRNAs jovens (idade 7-12) expressos no cortex frontal (CF), cerebelo (CB), coração (C), rim (R) e testículo (T). MiRNAs intragênicos estão sobre-representados em tecidos neurais ( $*P < 0,05$ , teste exato de Fisher). **(D)** Proporção de miRNAs jovens (idade 7-12) inter- e intragênicos presentes em bancos de dados associados à doenças (HMDD e PhenomiR) em relação ao total de miRNAs jovens em cada categoria. Os miRNAs intergênicos estão sobre-representados em relação aos intragênicos ( $*P < 0,05$ ,  $**P < 0,01$ , teste exato de Fisher).

### 4.3 Análise da expressão diferencial de miRNAs e associações com a quimioresistência em câncer colorretal

#### 4.3.1 Sequenciamento e identificação de microRNAs em linhagens celulares de câncer colorretal

O sequenciamento gerou um total de 384.056.150 de *reads*, com variação entre 25 e 42 milhões de sequências por linhagem celular (Tabela 2). Em média, 82% das sequências alinharam perfeitamente no genoma humano. Do total de sequências geradas, em média 23% mapearam em miRNAs maduros anotados pelo miRBase (Tabela 2), uma proporção coerente com estudos que utilizaram metodologias similares (Schulte *et al.* 2010; Dhabhi *et al.* 2011).

**Tabela 2. Estatísticas gerais do mapeamento dos RNAs pequenos no genoma humano e quantidade de sequências mapeadas em miRNAs maduros.**

Linhagem celular	Total de <i>reads</i>	% <i>reads</i> mapeados	% <i>reads</i> mapeados em miRNAs
CACO2	25.178.281	84,05	20,99
COLO205	35.838.967	75,11	25,45
HCC2998	34.013.530	75,31	16,46
HCT116	27.107.238	88,09	33,58
HCT15	34.862.986	89,27	25,72
HT29	28.406.493	78,26	25,67
KM12	27.010.631	71,20	17,12
LIM2405	31.415.642	88,86	30,69
RKO	42.738.358	89,66	29,23
SW480	31.800.125	87,77	15,83
SW620	40.168.148	83,76	19,06
Total/Média	358.540.399	82,84	23,61

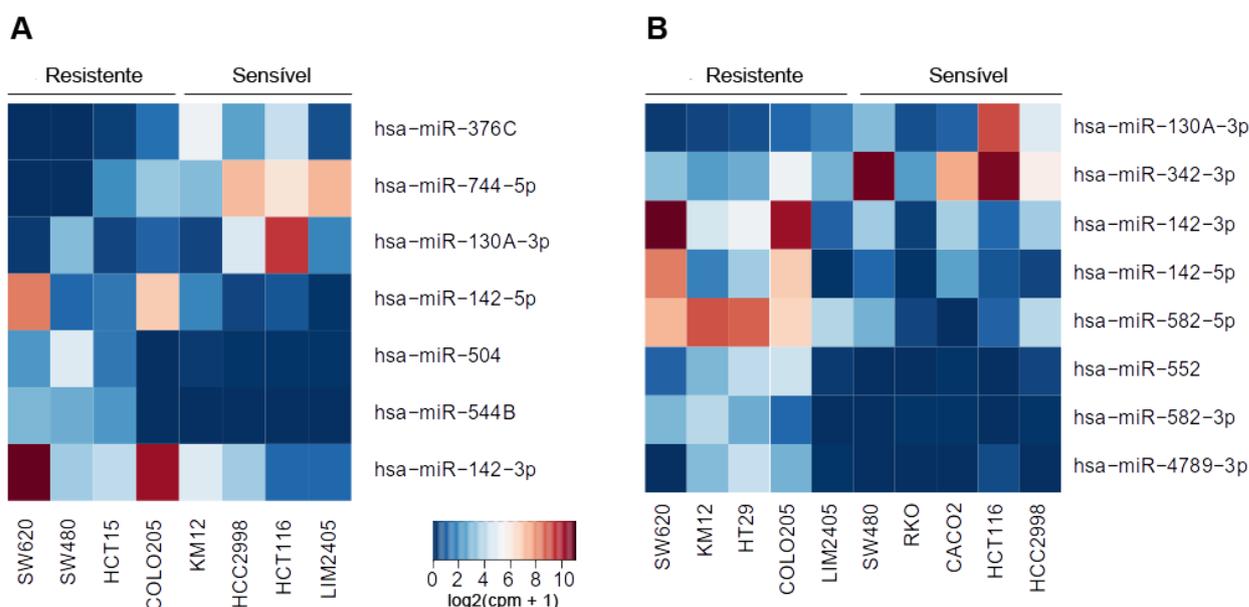
A maioria das sequências mapeadas contém entre 19 nt e 23 nt, um intervalo esperado que reflete o tamanho dos miRNAs maduros (Figura S3, Apêndice). Esses dados indicam que a construção das bibliotecas, o sequenciamento e o mapeamento dos *reads* foram conduzidos de forma adequada para as análises propostas. Em seguida, a quantidade de sequências

totalmente sobrepostas aos miRNAs conhecidos foram somadas para obtenção dos valores de expressão, e que posteriormente foram utilizados na análise da expressão diferencial entre os grupos de linhagens resistentes e sensíveis às drogas testadas.

#### *4.3.2 Expressão diferencial de microRNAs em linhagens resistentes e sensíveis à drogas quimioterápicas*

Os grupos de linhagens resistentes e sensíveis às drogas 5-FU e Oxaliplatina foram testados quanto à expressão diferencial de miRNAs. Os miRNAs muito pouco expressos foram eliminados das análises, excluindo-se aqueles com expressão < 1 cpm em menos de 3 linhagens. Do total de 1107 miRNAs maduros identificados (57,6% de todos os miRNAs maduros conhecidos – miRBase, versão 19), após a aplicação do filtro restaram 454 miRNAs com expressão em níveis consideráveis para o grupo de linhagens testadas para 5-FU e 495 miRNAs para o grupo testado para Oxaliplatina. O número de miRNAs maduros identificados, após a utilização do filtro de expressão também foi equivalente ao encontrado em estudos similares (Dhahbi *et al.* 2011; Hamfjord *et al.* 2012).

A Figura 26A mostra a expressão dos miRNAs diferencialmente expressos entre os grupos de linhagens resistentes e sensíveis à 5-FU. O agrupamento hierárquico dos miRNAs diferencialmente expressos, realizado pelo método de distâncias euclidianas, não foi capaz de separar completamente as linhagens sensíveis das resistentes (dados não mostrados). Entretanto, foi possível identificar miRNAs que apresentam papel importante na morte celular. Observamos que miR-504 é mais expresso nas linhagens resistentes (Figura 26A). Este miRNA foi descrito como regulador negativo do gene TP53 (Hu *et al.* 2010), sugerindo que a sensibilidade à 5-FU pode estar relacionada com níveis mais baixos desse miRNA.

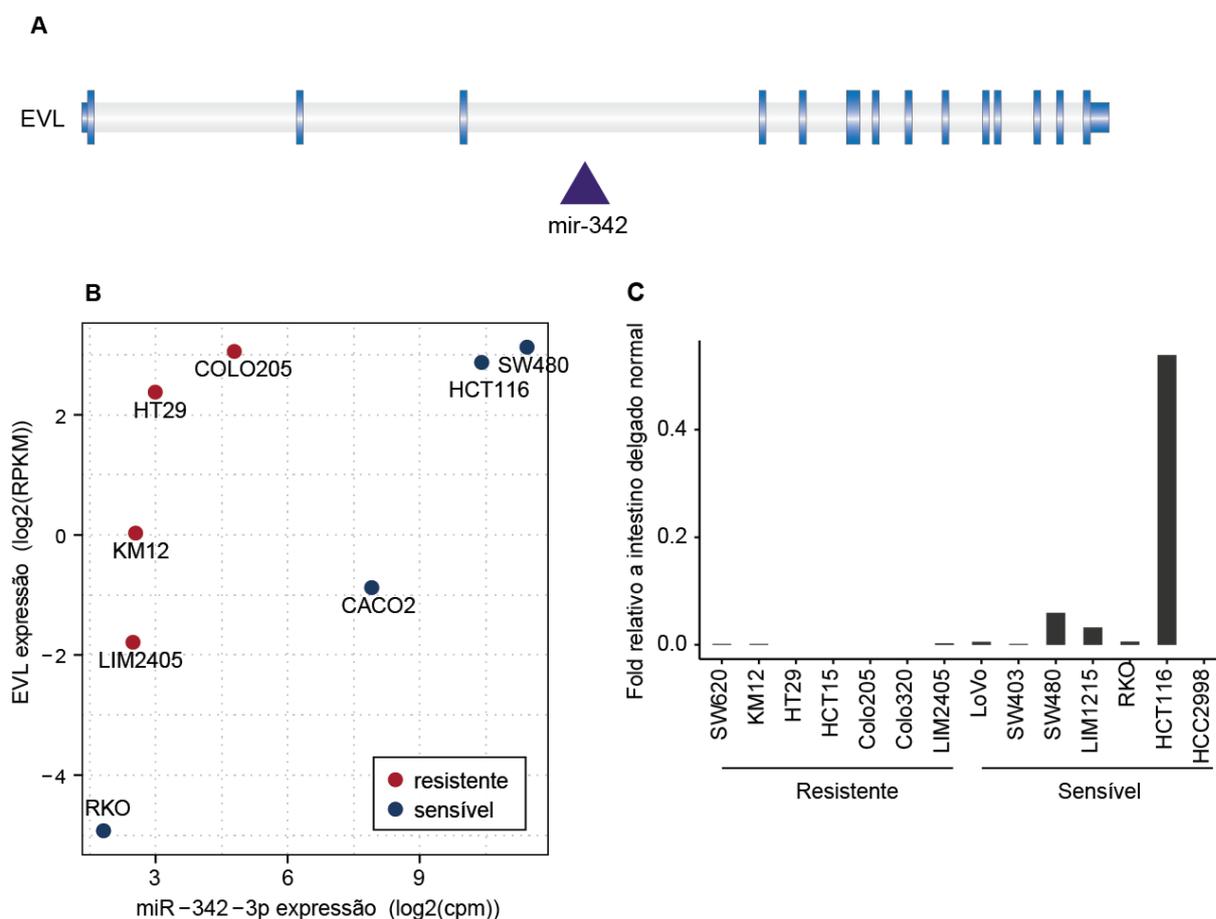


**Figura 26.** *Heatmap* dos valores de expressão normalizados dos miRNAs diferencialmente expressos (FDR < 0.01) entre linhagens resistentes e sensíveis ao tratamento com a droga 5-FU (A) e Oxaliplatina (B).

Outro miRNA potencialmente interessante é o miR-376C, que encontra-se menos expresso nas linhagens resistentes à 5-FU (Figura 26A). Células derivadas de melanoma transfectadas com miR-376C apresentaram proliferação e migração reduzida. Um dos seus alvos validados experimentalmente é o gene IGF1R, um receptor tirosina quinase conhecidamente envolvido com a progressão tumoral em melanoma (Zehavi *et al.* 2012).

Os miRNAs diferencialmente expressos entre linhagens resistentes e sensíveis à Oxaliplatina estão indicados na Figura 26B. Dentre os miRNAs identificados, o miR-342 foi o que mais nos chamou atenção por apresentar um padrão de expressão consistente com a literatura e em níveis testáveis experimentalmente, sendo mais expresso nas linhagens sensíveis à droga. Relacionado com esse resultado, estudos anteriores demonstraram que o miR-342 inibe a capacidade invasiva e proliferação de células SW480, caracterizada pela parada do ciclo celular na fase G(0)/G(1) e que a expressão aumentada de miR-342 foi capaz de inibir o crescimento tumoral e metástase pulmonar em camundongos (Wang *et al.* 2011).

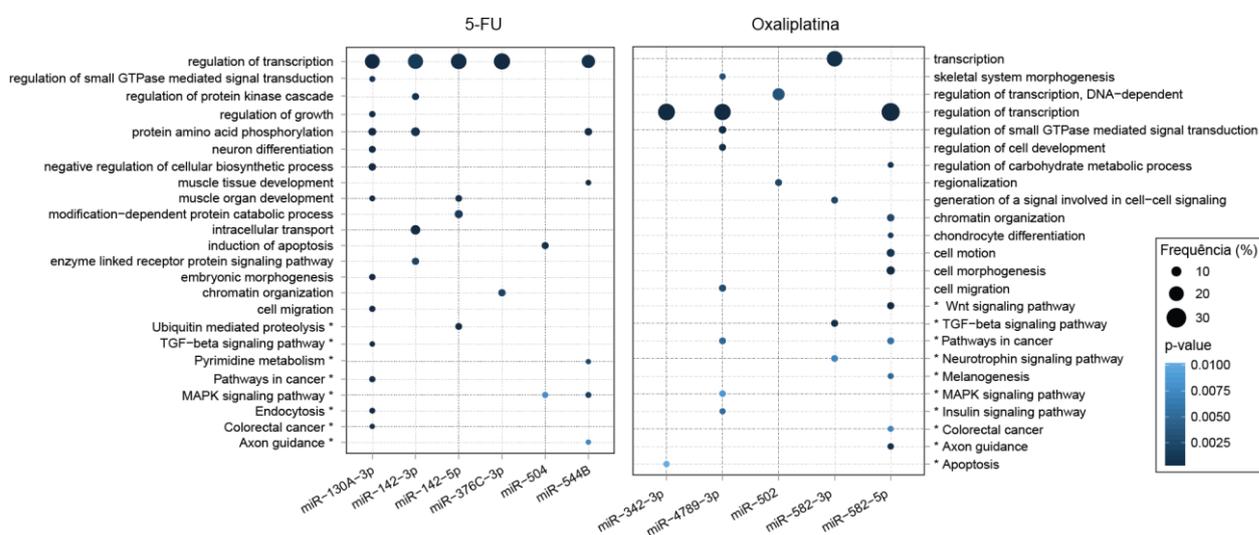
Adicionalmente, o miR-342 está localizado no terceiro intron do gene EVL (Figura 27A), um gene que participa da via de reparo de dupla-quebra de DNA (Takaku *et al.* 2009a). Esses dois genes são co-expressos e encontram-se silenciados em câncer colorretal devido a hipermetilação de ilhas CpG na região promotora do gene EVL (Grady *et al.* 2008). A co-expressão desses dois genes foi corroborada em nossas análises, evidenciada pela correlação da expressão de miR-342 e de EVL (utilizando dados de RNA-Seq de mRNAs disponíveis no laboratório) ( $\rho = 0,7$ ,  $P = 0,03$ ; Figura 27B). Também verificamos a maior expressão de miR-342 num painel que engloba linhagens adicionais sensíveis à Oxaliplatina por qRT-PCR ( $P = 0.05$ , teste t de Student; Figura 27C).



**Figura 27. Expressão de miR-342 e EVL por qRT-PCR.** A) Posição do miR-342 no terceiro intron do gene *host* EVL. B) Correlação dos valores de expressão de miR-342 e EVL ( $\rho = 0,7$ ,  $P = 0,03$ , correlação de Spearman). A expressão de EVL está indicada em RPKM. C) Expressão de miR-342 em linhagens resistentes e sensíveis Oxaliplatina determinadas por qRT-PCR. As medidas representam os valores de  $\Delta\Delta Ct$  relativos à IDN (intestino delgado normal).

### 4.3.3 Enriquecimento de categorias funcionais para os alvos dos miRNAs diferencialmente expressos

Com o objetivo de explorar possíveis funções desempenhadas pelos miRNAs diferencialmente expressos, selecionamos um conjunto de potenciais alvos preditos pelo algoritmo TargetScan 6.2. A Figura 28 mostra as principais categorias funcionais do Gene Ontology e vias do KEGG enriquecidas para os alvos dos miRNAs diferencialmente expressos nas linhagens tratadas com 5-FU e Oxaliplatina. Vimos que a maioria dos miRNAs tem seus alvos envolvidos com a regulação da transcrição. Os alvos preditos de miR-130A-3p e miR-582 são funcionalmente diversos e, notadamente, estão enriquecidos para *Pathways in cancer* e *Colorectal Cancer*, respectivamente. Consistente com estudos anteriores, observamos que os alvos do miR-504 e, em especial do miR-342-3p, estão significativamente associados com apoptose (Figura 28) (Grady *et al.* 2008; Hu *et al.* 2010).



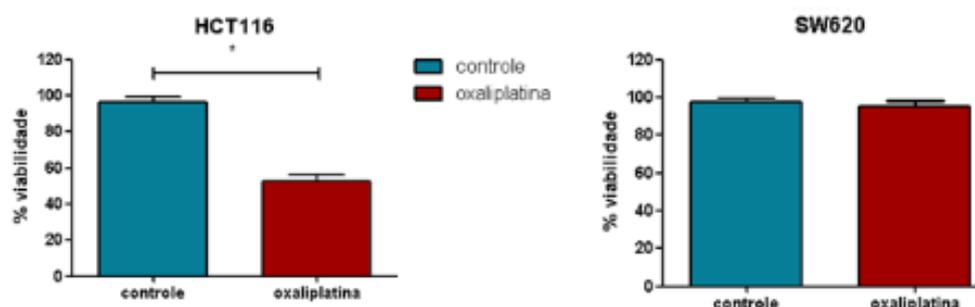
**Figura 28. Análise funcional dos alvos dos miRNAs diferencialmente expressos.** Foram incluídos na análise termos pertencentes à categoria “*biological process*” do Gene Ontology ( $P < 0.01$  e  $FDR < 5\%$ ) e vias de sinalização do KEGG (denotados com \*) ( $P < 0.01$  e  $FDR < 10\%$ ). O gráfico mostra 1 termo por *cluster* (score  $> 1.3$ ) definido pelo algoritmo da ferramenta DAVID. Os termos enriquecidos estão representados para o conjunto de alvos dos miRNAs diferencialmente expressos entre linhagens resistentes e sensíveis à 5-FU e Oxaliplatina.

Portanto, a análise de alvos e o fato do miR-342 apresentar maior expressão nas linhagens sensíveis à Oxaliplatina sugerem seu envolvimento com a via apoptótica. Além disso, Grady *et al.* (2008) mostraram que a expressão aumentada de miR-342 foi capaz de induzir a apoptose na linhagem HT-29. Outra evidência a favor desse papel está relacionada com a sensibilização de linhagens de tumor de mama à droga Tamoxifen pelo aumento da expressão de miR-342 (Cittelly *et al.* 2010). Em virtude desse conjunto de evidências, decidimos investigar o papel de miR-342 como promotor da apoptose em linhagens tratadas com Oxaliplatina utilizando abordagens experimentais, que foram conduzidas pela Dra. Lilian T. Inoue e Dr. Érico T. Costa, e descritas a seguir.

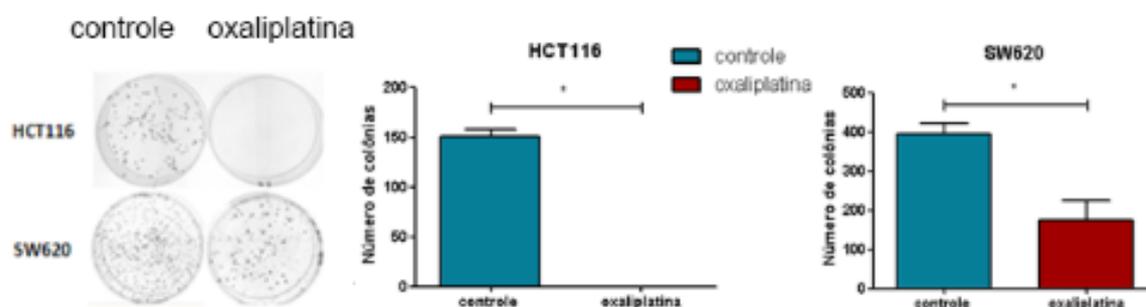
#### *4.3.4 Análise de viabilidade celular por exclusão de iodeto de propídeo e formação de colônias*

Para dar prosseguimento ao estudo do papel de miR-342 na morte celular por apoptose, optamos por realizar os experimentos em duas linhagens celulares que apresentaram fenótipos opostos frente ao tratamento com Oxaliplatina. Para confirmar a capacidade de resistência/sensibilidade determinada por Arango *et al.* (2004), selecionamos as linhagens HCT116 e SW620 classificadas como sensível e resistente à Oxaliplatina, respectivamente. Ensaio de viabilidade por MTT foram realizados para determinar o índice de citotoxicidade 50% (IC50) frente a doses crescentes da droga (dados não mostrados). Da mesma forma que o estudo anterior, estabelecemos a dose de 10  $\mu$ M e o tempo de 72h para os tratamentos com a droga. Em seguida, ensaios de exclusão por iodeto de propídeo (PI) foram realizados, evidenciando um significativo aumento de morte celular na linhagem HCT116, enquanto o mesmo não foi observado para SW620 (Figura 29). Complementariamente, a linhagem HCT116 apresentou redução de 100% do potencial clonogênico, enquanto a SW620 apresentou

redução de 50% deste potencial quando tratadas com Oxaliplatina (Figura 30). Com isso, confirmamos para estas duas linhagens o fenótipo sensível para HCT116 (IC50:  $3,2 \pm 3 \mu\text{M}$ ) e resistente para SW620 (IC50:  $51,7 \mu\text{M}$ ).



**Figura 29. Ensaio de viabilidade celular por exclusão de iodeto de propídeo (PI) seguida de análise por citometria de fluxo.** Células das linhagens tumorais de cólon HCT116 e SW620 foram cultivadas na presença de  $10\mu\text{M}$  de oxaliplatina por 72h. Após a incubação com PI, as células viáveis (PI-negativas) e inviáveis (PI-positivas) foram quantificadas por citometria de fluxo. Os experimentos foram realizados em triplicatas independentes. (\* $P < 0,05$ , teste t).

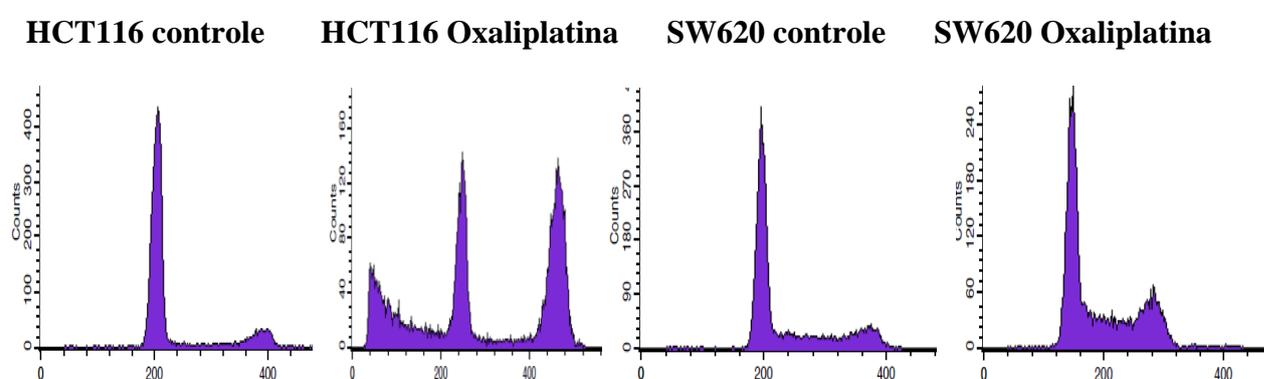


**Figura 30. Ensaio clonogênico bidimensional.** Células das linhagens tumorais de cólon HCT116 e SW620 foram cultivadas na presença de  $10\mu\text{M}$  de oxaliplatina por 72h. Posteriormente, quinhentas células foram mantidas em cultura por 10 dias em meio de cultura sem oxaliplatina. As colônias foram coradas com cristal violeta e contadas visualmente. Os experimentos foram realizados em duplicatas amostrais e triplicatas experimentais independentes. (\*  $p < 0,05$ , teste t).

#### 4.3.5 Análise da progressão do ciclo celular

Para melhor avaliação dos efeitos do tratamento com Oxaliplatina, quantificamos as células em cada fase do ciclo celular (G1, S, G2 e Mitose) e condição de subdiploidia

(denominada sub-G1), determinada pelo conteúdo de DNA por marcação com PI seguida de análise por citometria de fluxo. Para a linhagem HCT116, observamos intenso enriquecimento da fração subdiploide (sub-G1), ou seja, um acúmulo de células com conteúdo de DNA nuclear menor do que  $2n$ , indicativo de morte celular, seguida de degradação e perda de material genético (Figura 31), consistente com a definição do fenótipo “sensível”. A linhagem SW620 não apresentou alteração significativa na etapa sub-G1 do ciclo celular após exposição à Oxaliplatina, confirmando o fenótipo “resistente” (Figura 31).



**Figura 31. Análise da progressão do ciclo celular por marcação com PI.**

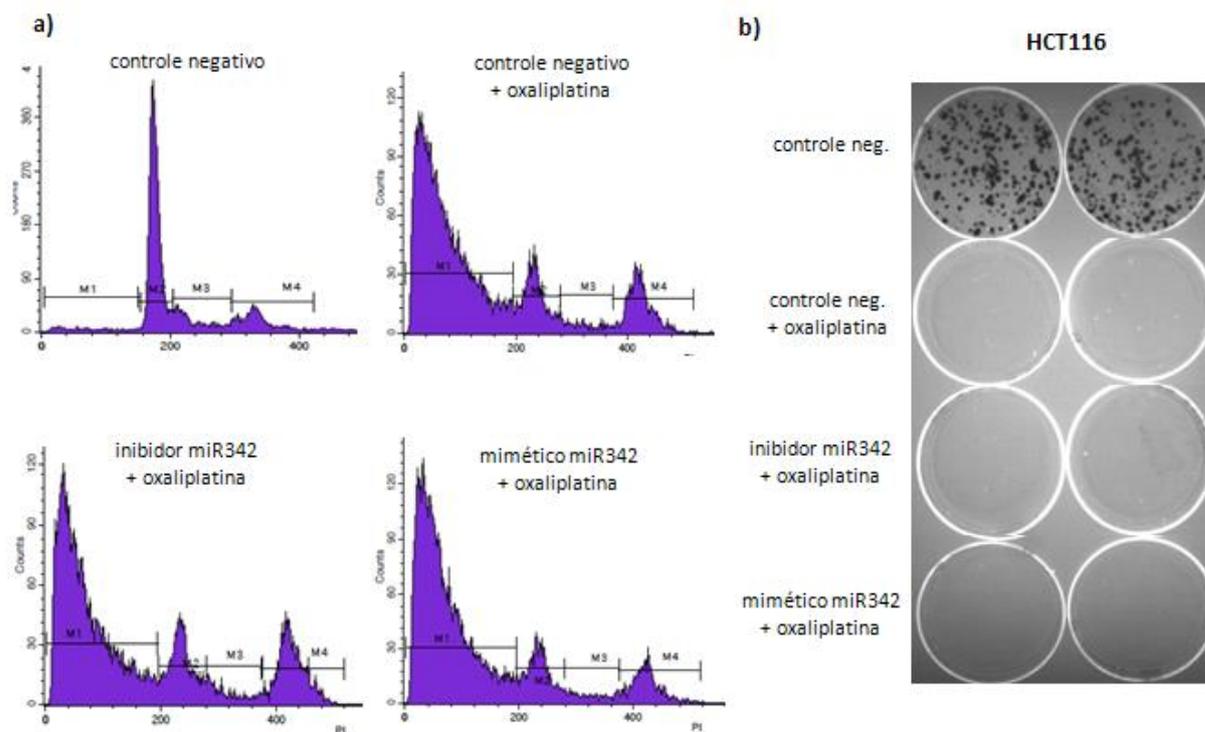
Células das linhagens tumorais de cólon HCT116, HT29, SW480 e SW620 foram cultivadas na presença de  $10\mu\text{M}$  de oxaliplatina por 72h. Os núcleos celulares foram extraídos e o DNA, marcado com iodeto de propídeo. A intensidade de fluorescência emitida pelo iodeto de propídeo está representada no eixo das abscissas. As análises foram feitas por citometria de fluxo e os resultados apresentados são representativos de triplicatas experimentais independentes.

#### 4.3.6 Efeitos da superexpressão de miR-342 na resposta ao tratamento com Oxaliplatina

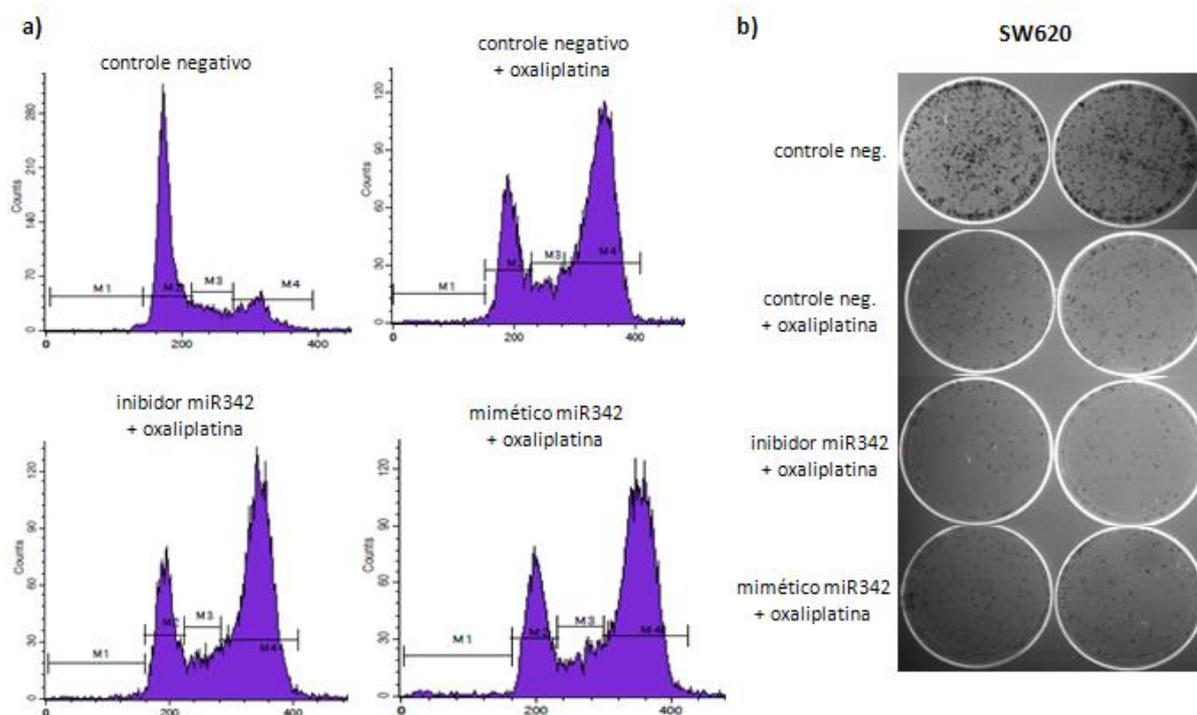
Para avaliar o impacto funcional da expressão de miR-342 no fenótipo de resistência/sensibilidade à Oxaliplatina, utilizamos como modelo para transfecções as linhagens HCT116 e SW620. Como a HCT116 apresentou os níveis mais elevados de miR-342, inicialmente consideramos analisar o efeito da inibição deste miRNA pela transfecção com o inibidor anti-miR-342. Entretanto, a verificação da eficiência de inibição esbarra em

dificuldades técnicas. Uma das alternativas é medir o efeito indireto da inibição pelos níveis de um de seus alvos. O DNMT1 é um alvo experimentalmente validado (Wang *et al.* 2011) e por isso foi escolhido como indicador. Se a inibição foi suficientemente eficaz, espera-se que os níveis de DNMT1 encontrem-se elevados em relação ao controle. Porém, não foi possível observar tal diferença (dados não mostrados). Por essa razão, focamos em analisar o efeito da superexpressão de miR-342 após 24, 55 e 72 horas de transfecção com a molécula mimética tanto em HCT116 quanto em SW620. A análise da expressão por qRT-PCR confirmou o significativo aumento de miR-342, cuja variação foi de 1.641 à 8.629 vezes maior em relação ao oligonucleotídeo controle negativo.

Em seguida, analisamos o impacto da inibição (mesmo que não tenha sido possível quantificar a eficiência da transfecção) e da superexpressão de miR-342 sobre a progressão do ciclo celular. Para isso, as células foram transfectadas com o miR-Ctrl (controle negativo), o inibidor ou o mimético de miR-342. Após 24 horas das transfecções as células foram incubadas com 10 uM de Oxaliplatina por mais 48 horas. Em seguida, foi feita a análise da porcentagem de cada fase do ciclo celular, conforme mencionado na seção 4.3.6. De acordo com a Figura 32A, não observamos diferenças nos perfis de ciclo celular em fase sub-G1 entre as células transfectadas com o inibidor ou mimético de miR-342 em relação ao controle, quando tratadas com Oxaliplatina. O ensaio de formação de colônias também não revelou diferenças entre os grupos Figura 32B. Os mesmos experimentos realizados para SW620 também não revelaram diferenças (Figura 33A e B). Estes resultados sugerem que a superexpressão ou a inibição de miR-342 não foram suficientes para alterar o padrão de ciclo celular no modelo testado.



**Figura 32. Ensaio de ciclo celular e clonogênico bidimensional em células HCT116 transfetadas com oligonucleotídeo controle negativo, inibidor de miR342 ou mimético de miR342 e expostas à 10 $\mu$ M de oxaliplatina por 48h. A) Perfis dos ciclos celulares após marcação dos DNA nucleares com iodeto de propídeo e análise por citometria de fluxo. B) Ensaio clonogênico bidimensional, onde mil células transfetadas e expostas à oxaliplatina foram cultivadas em meio sem o quimioterápico por 7 dias e, em seguida, coradas com solução de cristal violeta.**

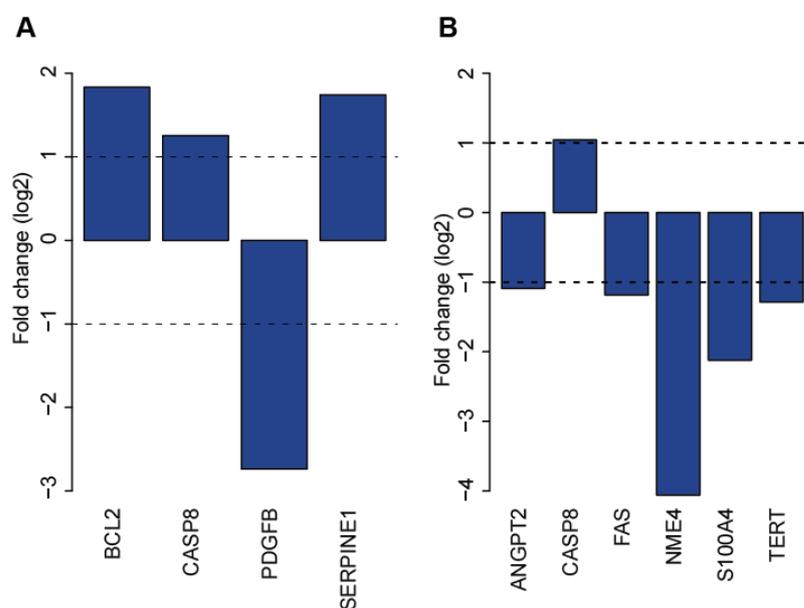


**Figura 33.** Ensaios de ciclo celular e clonogênico bidimensional em células SW620 transfectadas com oligonucleotídeo controle negativo, inibidor de miR342 ou mimético de miR342 e expostas à 10 $\mu$ M de oxaliplatina por 48h. **A)** Perfis dos ciclos celulares após marcação dos DNA nucleares com iodeto de propídeo e análise por citometria de fluxo. **B)** Ensaio clonogênico bidimensional, onde mil células transfectadas e expostas à oxaliplatina foram cultivadas em meio sem o quimioterápico por 7 dias e, em seguida, coradas com solução de cristal violeta.

#### 4.3.7 Análise da expressão de genes da via apoptótica mediante superexpressão de miR-342

Com o objetivo de melhor compreender a possível associação da expressão de miR-342 com a morte celular, especificamente por apoptose, avaliamos os padrões de expressão de 84 genes envolvidos na via de apoptose (definidos pelo *kit* comercial SuperArray; ver lista completa no Quadro 2 - Apêndice) 16 e 41 horas após a transfecção da linhagem resistente à Oxaliplatina SW620 com miR-342. Dentre os genes que apresentaram as maiores variações de expressão em relação ao controle, destacam-se ANGPT2, BCL2, CASP8, FAS, NME4, PDGFB, S100A4 e SERPINE1 e TERT (Figura 34 e Quadro 2 - Apêndice). Os genes ANGPT2, FAS, NME4, S100A4 e TERT apresentaram diminuição da expressão após a transfecção com

miR-342 em 41h (Figura 34B) mas não em 16h, o que sugere um efeito tardio, constituindo provavelmente uma consequência regulatória indireta da expressão do miR-342. Grosso modo, há evidências de um papel pró-apoptótico para ANGPT2 (envolvido com apoptose de células endoteliais (Bhandari *et al.* 2006)), além de FAS (Zhang *et al.* 1998) e NME4 (<http://www.genecards.org/>) e de um papel anti-apoptótico para TERT (Haendeler *et al.* 2003) e S100A4 (Orre *et al.* 2013). Em contrapartida, após 16h, observamos um aumento na expressão de BCL2, CASP8, SERPINE1 e diminuição de PDGFB (Figura 34A, Quadro 2 - Apêndice). O gene BCL2 é um importante anti-apoptótico (Chao e Korsmeyer 1998), enquanto CASP8 faz parte da cascata de caspases, que quando ativada promove a apoptose (McIlwain *et al.* 2013). A redução de PDGFB foi a mais expressiva em 16h, sendo este um fator de crescimento que age como um potente agente mitogênico, promovendo a proliferação celular (Andrae *et al.* 2008). É interessante notar que este gene possui dois sítios-alvo para o miR-342 preditos pela ferramenta microT (Paraskevopoulou *et al.* 2013) e um sítio predito pela ferramenta TargetScan, sugerindo a regulação direta por miR-342.



**Figura 34. Genes da via apoptótica com expressão alterada após transfecção com miR-342 na linhagem SW620.** Apenas genes com *fold change* > 2 (linha tracejada) em relação ao miR controle (oligonucleotídeo) foram considerados. A avaliação da expressão por qRT-PCR foi realizada 16 horas (A) e 41 horas (B) após a transfecção. Valores abaixo de 0 representam expressão diminuída e acima de 0 expressão aumentada em relação ao controle.

## *5 Discussão*

## 5.1 Contribuição da ferramenta miRIAD para o estudo dos miRNAs intragênicos

À medida que novos miRNAs são constantemente identificados, havendo um enorme aumento no número de novos transcritos anotados nos últimos anos, a quantidade de miRNAs intragênicos tem acompanhado essa tendência. Por exemplo, a versão 11 do miRBase (Abril de 2008) continha 678 precursores anotados para a espécie humana, dos quais 47% são intragênicos. Essa proporção aumentou para 53% (de um total de 1600 precursores) na versão 19 (Agosto de 2012), chegando a 57% (1066/1871) de miRNAs intragênicos (tomando como referência a sobreposição com genes codificantes do RefSeq) na versão 20 (Junho de 2013). Neste cenário, consideramos que o desenvolvimento da ferramenta miRIAD tem o potencial de contribuir para o acesso a informações integradas sobre miRNAs, seu contexto genômico e seus genes hospedeiros.

Os dados foram basicamente organizados em cinco camadas de informação. A primeira delas contém as informações de anotação dos genes codificantes de proteínas e miRNAs, incluindo os nomes oficiais, *aliases*, nomenclaturas referentes a outras bases de dados (Entrez gene, Ensembl, HGNC) e sumário da função desempenhada. A segunda camada provê informações sobre o contexto genômico dos miRNAs, genes codificantes (classificados em hospedeiros e não-hospedeiros), com a representação esquemática dos transcritos referentes a cada gene. A terceira camada é designada a dados de expressão por RNA-Seq de genes codificantes, miRNAs e à correlação da expressão dos miRNAs intragênicos e dos seus genes hospedeiros. Os alvos preditos dos miRNAs compõem a quarta camada, e a quinta está associada com informações adicionais como interações proteína-proteína, parceiros de interação dos genes hospedeiros que são alvos preditos dos miRNAs intragênicos. Adicionalmente, *links* dinâmicos para bases externas são disponibilizados. Todas essas informações são organizadas numa interface gráfica de fácil uso e integradas num banco

relacional MySQL. Para os usuários que desejem manipular as informações contidas no miRIAD localmente, os dados são disponibilizados para *download*. Informações específicas também podem ser requeridas diretamente com os administradores da ferramenta.

Acreditamos que as informações contidas no miRIAD proporcionam um ponto de partida para a geração de novas hipóteses acerca da regulação por miRNAs de um ou mais genes, principalmente quando estão inseridos em um contexto gênico. Esta possibilidade pôde ser evidenciada para os genes CLCN5 e HTR2C, os quais foram selecionados por uma rápida inspeção a partir de uma lista de genes de interesse. Para a maioria dos miRNAs contidos nesses genes, há fortes indícios de co-expressão com seus genes hospedeiros. Uma das consequências funcionais da co-expressão de miR-488 e HTR2C, por exemplo, está relacionada com a diferenciação de adipócitos (Kinoshita *et al.* 2010). De maneira similar, a co-expressão do IGF2 e miR-483 tem como resultado um *feedback* regulatório positivo em que o miR-483 promove a transcrição do seu próprio gene hospedeiro (Liu *et al.* 2013). Dessa forma, diversas outras consequências funcionais decorrentes da organização genômica de miRNAs contidos em genes codificantes de proteínas poderão ser reveladas em estudos futuros, e que em princípio, poderiam ter o miRIAD como ponto de partida para geração de hipóteses. Neste sentido, observamos um grande número de acessos provenientes de diversas regiões do mundo, o que ressalta o potencial dessa ferramenta para os propósitos mencionados. Por fim, enfatizamos que versões atualizadas estão sendo programadas, nas quais serão incluídas espécies adicionais, dados de expressão referentes a uma ampla gama de amostras e repopular o banco de dados com novas versões dos dados brutos (RefSeq, HPRD, miRBase, etc.). Para isso, os *scripts* e *pipelines* necessários foram desenvolvidos para que as atualizações sejam implementadas sob demanda.

## 5.2 A influência do contexto genômico e idade dos genes hospedeiros na expressão e evolução dos miRNAs humanos

O surgimento de miRNAs sobrepostos a outros genes tem sido recorrentemente documentado, no entanto, associações funcionais e as consequências evolutivas dessa organização genômica ainda são bastante desconhecidas. Nós abordamos essa questão investigando os padrões evolutivos aos quais os miRNAs humanos estão sujeitos sob a perspectiva da influência exercida pelo contexto genômico e genes hospedeiros. Observamos um excesso de miRNAs humanos intragênicos a partir da emergência dos mamíferos placentários. Além disso, vimos que a aquisição de grande parte do repertório de miRNAs intragênicos humanos acompanhou o surgimento massivo de miRNAs em primatas, mais precisamente nos ramos 7 e 8 (Figura 16) (que correspondem à divergência dos primatas do novo e velho mundo, respectivamente).

Nossas análises revelaram que os miRNAs intragênicos surgem, ou são fixados, preferencialmente em genes codificantes antigos, sugerindo importantes consequências funcionais e evolutivas. Vimos que os genes hospedeiros antigos têm maior abrangência de expressão que os hospedeiros novos, além de possuírem taxas de evolução mais lentas (menor  $Ka/Ks$ ), mesmo quando comparados com genes antigos que não abrigam miRNAs, o que sugere uma forte seleção purificadora atuando nos genes hospedeiros.

Neste trabalho mostramos que a idade dos genes hospedeiros afeta diretamente a abrangência de expressão dos miRNAs intragênicos. Especificamente, observamos que os miRNAs contidos em genes novos tendem a apresentar expressão tecido-específica quando comparados com miRNAs de mesma idade em genes antigos ou intergênicos. Surpreendentemente, os miRNAs jovens (primata-específicos) contidos em genes antigos são mais amplamente expressos que os intergênicos de mesma idade. Este efeito é correlacionado

com a abrangência de expressão dos genes hospedeiros, o que por sua vez está associado com a co-expressão dos pares miRNA-*host*. Embora discrepâncias entre a expressão do gene hospedeiro e seus miRNAs intragênicos tenham sido observadas, principalmente devido ao uso de promotores intrônicos independentes (Ozsolak *et al.* 2008; Marsico *et al.* 2013), os resultados encontrados são favoráveis ao cenário de co-regulação, o que deve ser válido particularmente para os miRNAs jovens, para os quais sugere-se que a regulação por promotores compartilhados com os genes hospedeiros seja significativamente frequente (Marsico *et al.* 2013). Além disso, o estado transcricional da cromatina envolvendo o gene hospedeiro e os miRNAs contidos na mesma região também pode exercer uma importante influência espaço-temporal na expressão dos miRNAs. De maneira análoga, mostramos que a expressão dos miRNAs intergênicos (exceto para os antigos (idade = 1)), parece ser afetada pela proximidade com genes vizinhos. Em contraste com os miRNAs intragênicos e os genes hospedeiros, não encontramos evidência de co-regulação entre os miRNAs intergênicos e genes vizinhos, o que nos leva a especular que a cromatina aberta em torno de genes transcionalmente ativos facilita a transcrição de miRNAs próximos, de maneira similar ao que ocorre com lincRNAs próximos a genes codificantes (Cabali *et al.* 2011).

Dado que a maioria dos miRNAs humanos reside em genes antigos, como tal contexto genômico poderia impactar a expressão e evolução dos miRNAs intragênicos? Recentemente, Popadin *et al.* (2014) analisaram as propriedades essenciais ligadas à idade dos genes. Os autores mostraram que cis-eQTLs (*cis- expressed quantitative trait loci*) associados à genes antigos possuem menor efeito na variância da expressão (*effect size*), são mais distantes do sítio de início de transcrição e afetam menor quantidade de tecidos que cis-eQTLs de genes novos. Em conjunto, essas observações indicam que há um aumento da restrição da variação de expressão (*expression constraints*) associada a genes antigos. Consistente com esse cenário evolutivo, nós encontramos que miRNAs intragênicos antigos (idade = 1) e de idade

intermediária (idade = 2-4), comparados aos intergênicos de mesma idade, apresentam menor divergência de expressão e maiores correlações de abrangência de expressão entre espécies. Além disso, observamos que miRNAs intragênicos jovens estão subrepresentados em dados de associações com doenças. Como essas associações são inferidas principalmente por causa da expressão alterada de miRNAs e assumindo que muitos desses miRNAs são co-regulados com seus genes hospedeiros, podemos supor que cis-eQTLs associadas aos genes hospedeiros podem afetar a expressão dos miRNAs contidos nesses genes. Portanto, sugerimos que o controle regulatório exercido pelos genes antigos pode conferir maior robustez na variação da expressão de miRNAs intragênicos antigos entre espécies, enquanto contribuiria para limitar a variação da expressão de miRNAs jovens em uma situação instável como em doenças. Uma investigação aprofundada, combinada com o mapeamento de eQTLs de miRNAs poderia trazer novas interpretações para essa questão.

Um importante achado deste trabalho é que a maior abrangência de expressão dos miRNAs intragênicos jovens, comparada com os intergênicos, está correlacionada com a abrangência de expressão dos genes hospedeiros antigos. De acordo com o modelo proposto por Lyu *et al.* (2014), a fixação de miRNAs de origem recente depende da sua integração nas redes regulatórias e interação com outros transcritos. Uma possível vantagem adquirida pelos miRNAs jovens inseridos em genes antigos é a sua incorporação em redes regulatórias impulsionada pelo controle espaço-temporal da expressão dos genes hospedeiros. Uma vez que esses miRNAs são expressos em uma maior quantidade de tecidos, eles poderiam regular uma maior variedade de alvos, conforme a análise de alvos preditos sugeriu, possivelmente acelerando o estabelecimento das relações entre os miRNAs e seus alvos, que com o tempo seriam moldadas pela seleção natural. Um argumento contrário a essa idéia é que a regulação de um vasto repertório de alvos por miRNAs que surgiram recentemente poderia resultar em efeitos deletérios, já que as relações miRNA-alvo não teriam ainda sido consolidadas.

Entretanto, numa fase evolutiva inicial, é razoável assumir que estes miRNAs raramente são expressos em níveis elevados a ponto de causar fortes mudanças de *fitness* (Lu *et al.* 2008; Wu *et al.* 2009). Mesmo para miRNAs altamente expressos e conservados, perturbações nos seus níveis de expressão podem produzir consequências sutis (Miska *et al.* 2007; Alvarez-Saavedra e Horvitz 2010). Além disso, análises experimentais sugerem funcionalidade mesmo para interações miRNA-alvo não conservadas, indicando que pelo menos durante a fase evolutiva inicial dos miRNAs, essas interações podem ser funcionais mas não deletérias. Neste sentido, o argumento aqui exposto encaixa-se com o papel dos miRNAs na hipótese de canalização (Hornstein e Shomron 2006; Lu e Clark 2012). Com base neste princípio, novos miRNAs serviriam principalmente como atenuadores (*buffering*) da expressão gênica, reduzindo a variação da expressão (*expression noise*) de seus alvos. Portanto, miRNAs jovens funcionariam como estabilizadores das redes regulatórias, em que as restrições evolutivas impostas a interações miRNA-alvo não são demasiadamente comprometedoras, pelo menos num período inicial (Wu *et al.* 2009).

Diversos miRNAs específicos de primatas (> 200) foram identificados por sua expressão em cérebro de humanos e chimpanzés, sugerindo um papel no estabelecimento da diversidade celular e programas regulatórios neste tecido (Berezikov *et al.* 2006). Nós adicionamos que os miRNAs intragênicos jovens, em conjunto com os seus genes hospedeiros, podem ter contribuído para este processo, já que observamos uma sobre-representação de miRNAs intragênicos primata-específicos em cérebro e cerebelo, além do enriquecimento dos genes hospedeiros expressos em cérebro e seu envolvimento com funções neuronais. Notadamente, há evidências de que miRNAs expressos em tecidos neurais possuem maior repertório de alvos que miRNAs não expressos nestes tecidos, e que os genes-alvo de miRNAs específicos de humanos estão associados com funções neuronais (Barbash *et al.* 2014). Um exemplo emblemático dessa rápida evolução é o caso do miRNA intragênico humano-

específico miR-941 e seu gene hospedeiro DNAJC5. Análises experimentais mostraram que este miRNA tem como alvo seu próprio *host*, e também outros dois genes cujos produtos interagem com a proteína codificada pelo DNAJC5 (Hu *et al.* 2012). Baseado na função do gene hospedeiro e de seus parceiros de interação, o miR-941 parece ter um papel em vias de sinalização de neurotransmissores (Hu *et al.* 2012).

Recapitulando as propriedades essenciais dos genes antigos, estes são caracterizados por estarem enriquecidos em redes regulatórias complexas e de alta conectividade (Popadin *et al.* 2014). Dessa forma, é plausível supor que miRNAs que surgiram em genes antigos poderiam atuar como reguladores dessas redes em tecidos neurais, interagindo com conjuntos de alvo em comum de maneira sinérgica ou antagônica (Hinske *et al.* 2010; Lutter *et al.* 2010; Gennarino *et al.* 2012). Em nota, um estudo recente assinalou a existência de ~3.500 novos miRNAs que não estavam anotados no miRBase, sendo que 95% e 57% foram determinados como primata e humano-específicos, respectivamente (Londin *et al.* 2015). Nós inspecionamos os dados publicados e observamos uma forte concordância com os padrões identificados aqui, o que significa que a maioria desses novos miRNAs são intragênicos (74%), com viés para orientação transcricional (76% orientação senso) contidos em genes antigos (80%). Estes genes hospedeiros também estão sobre-representados em cérebro ( $P = 1.0 \times 10^{-15}$ ). Por outro lado, os miRNAs intergênicos jovens podem ter contribuído para a regulação de RNAs mensageiros expressos preferencialmente no testículo, conforme sugerido pela sobre-representação desses miRNAs e expressão específica neste tecido. De fato, *clusters* de miRNAs primata-específicos ligados ao cromossomo X, que nós observamos serem intergênicos, apresentam expressão predominante em testículo (Zhang *et al.* 2007; Li *et al.* 2010; Meunier *et al.* 2013), tendo como alvos genes relacionados com a maturação do gameta masculino e morfologia do epidídimo (Li *et al.* 2010).

Em suma, neste trabalho apresentamos evidências de que os genes hospedeiros e o contexto genômico exercem forte influência na expressão e evolução dos miRNAs. Propomos que a idade do gene hospedeiro é uma propriedade chave na determinação dos padrões de expressão dos miRNAs intragênicos e também está relacionada com a variação de expressão interespecífica a longo prazo, enquanto promove maior abrangência de expressão aos miRNAs jovens. Uma vez que a expressão de um miRNA é afetada, seu destino evolutivo, em termos de interações com seus alvos, também é impactado. Nós propomos que o ambiente genômico transcricionalmente favorável e estritamente regulado dos genes hospedeiros antigos poderia acelerar a funcionalização dos miRNAs como agentes canalizadores, pelo menos durante sua fase adaptativa inicial (Lyu *et al.* 2014). Com isso, dependendo do contexto genômico em que os miRNAs surgem, este ambiente pode propiciar a ação da seleção adaptativa em novos miRNAs, favorecendo sua evolução (Kosik 2009; Ebert e Sharp 2012).

### **5.3 MiRNAs envolvidos com a quimioresistência em câncer colorretal**

O câncer colorretal é a terceira maior causa de morte por câncer no mundo (Parkin *et al.* 2002). Um dos maiores obstáculos na redução da mortalidade é a resistência de tumores metastáticos ao tratamento com drogas quimioterápicas (Bose *et al.* 2011). Com o objetivo de melhor compreender o papel dos miRNAs neste processo, abordamos esta questão comparando o perfil de expressão de miRNAs em linhagens sensíveis e resistentes às drogas 5-FU e Oxaliplatina previamente testadas e categorizadas nestes dois grupos de resposta (Mariadason *et al.* 2003; Arango *et al.* 2004). A análise das linhagens testadas para 5-FU revelou a expressão reduzida do miR-504 em linhagens sensíveis, cujos alvos preditos estão significativamente associados com a apoptose. Este miRNA foi descrito como regulador negativo do gene TP53,

um supressor tumoral que atua na regulação do ciclo celular induzindo a apoptose em resposta ao *stress* e danos no DNA. A superexpressão de miR-504 em células HCT116 foi responsável pela diminuição significativa dos níveis da proteína p53, promovendo o crescimento tumoral *in vivo* (Hu *et al.* 2010). Isso sugere que o miR-504 atua na via de apoptose mediada por TP53 e pode estar relacionado com a sensibilidade à 5-FU. Entretanto, a baixa expressão deste miRNA no modelo utilizado impediu que verificássemos a mesma tendência pela técnica de qRT-PCR. Os demais miRNAs identificados nesta análise também não apresentaram níveis de expressão suficientemente elevados para considerá-los nas validações experimentais (dados não mostrados).

Para as linhagens testadas com Oxaliplatina, identificamos o miR-342 como o miRNA com maior diferença de expressão entre os grupos resistente e sensível, sendo menos expresso nas linhagens resistentes. A expressão determinada por qRT-PCR confirmou essa tendência. É importante ressaltar que este miRNA tem sua expressão reduzida especificamente em câncer colorretal, tanto em linhagens celulares como em amostras de tecidos (Grady *et al.* 2008; Wang *et al.* 2011). Este miRNA está localizado na região intrônica do gene EVL e ambos são co-expressos. Em câncer colorretal a região promotora do EVL encontra-se frequentemente silenciada por metilação de ilhas CpG e por consequência a expressão de miR-342 também é silenciada (Grady *et al.* 2008). Curiosamente o EVL está envolvido no reparo por dupla-quebra de DNA através da interação com RAD51 (Takaku *et al.* 2009a, 2009b), um tipo de dano causado pela Oxaliplatina (Faivre *et al.* 2003), e por isso é plausível que o próprio gene hospedeiro do miR-342 também possa estar relacionado com mecanismos de resistência. Além disso, a análise de alvos preditos do miR-342 revelou um enriquecimento para genes relacionados com apoptose. De fato, a redução na quimiosensibilidade através de mudanças no balanço de proteínas que controlam a apoptose constitui um importante mecanismo de resistência (Allen e Weiss 2010), em especial para Oxaliplatina (Faivre *et al.* 2003).

Com base nestas evidências, prosseguimos o estudo a fim de verificar o papel de miR-342 como promotor da apoptose e sua possível relação com a sensibilização à Oxaliplatina através da modulação da sua expressão em uma linhagem sensível (HCT116) e outra resistente (SW620). Os ensaios de exclusão de iodeto de propídeo e clonogênico corroboraram esta classificação. Entretanto, a inibição ou superexpressão do miR-342 não foi capaz de reverter os fenótipos de sensibilidade e resistência para HCT116 e SW620, respectivamente, já que não observamos alterações significativas no padrão de ciclo celular quando comparadas com o controle negativo. É possível que a não detecção de efeitos fenotípicos evidentes após a superexpressão de miR-342 em HCT116 e SW620 frente ao tratamento com a Oxaliplatina esteja relacionada com o modelo experimental escolhido e com a intensidade do efeito de miR-342. As células HCT116 são consideravelmente sensíveis à Oxaliplatina (IC<sub>50</sub>: ~3µM), e portanto é provável que uma taxa ainda maior de apoptose/alteração no ciclo celular não seja detectável após a transfecção com miR-342 (assumindo que miR-342 tem um papel pró-apoptótico). Por outro lado, a SW620 tolera altos níveis de toxicidade (IC<sub>50</sub>: ~50µM), e por isso o efeito de miR-342 talvez não seja suficientemente forte para sensibilizá-la à Oxaliplatina. Ainda, é possível que a concentração de droga utilizada não tenha sido a ideal para avaliar o efeito da transfecção. Em condições normais (*i.e.*, na ausência do desafio com a droga), a expressão aumentada de miR-342 foi capaz de inibir a proliferação e invasão de SW480 e metástase *in vivo* (Wang *et al.* 2011) e induzir apoptose em HT-29 (Grady *et al.* 2008), ambas linhagens de câncer colorretal. Curiosamente, o efeito que esperávamos observar com a Oxaliplatina foi verificado por Cittelly *et al.* (2010) em modelo de câncer de mama, no qual células derivadas da linhagem MCF-7 foram sensibilizadas à Tamoxifen após transfecção com miR-342, o que evidencia sua associação com a quimioresistência.

Interpretamos que no modelo utilizado, este miRNA exerce seu papel na regulação da via apoptótica através do ajuste fino da expressão de seus alvos. A análise da expressão de

genes desta via mostrou que a superexpressão de miR-342 em SW620 alterou a expressão de nove genes com funções pró ou anti-apoptóticas, em 16 ou 41h após a transfecção do miRNA. De modo geral, observamos a inibição da expressão de genes pró e anti-apoptóticos em 41h (ANGPT2, FAS, NME4, TERT, S100A4), sugerindo alterações indiretas decorrentes do efeito de miR-342. Dentre os genes alterados após 16h de transfecção, destaca-se o PDGFB, que apresentou expressão diminuída de 6 vezes em relação ao controle. Consistente com a inibição de PDGFB, observamos que este gene possui dois sítios-alvo preditos para miR-342, indicando uma regulação direta. O PDGFB é um dos quatro ligantes de receptores PDGFR e atua como fator de crescimento, promovendo a proliferação celular de células de origem mesenquimal, além contribuir para a migração e diferenciação dessas células (Andrae *et al.* 2008). Em câncer colorretal a superexpressão de PDGFB está associada com maior proliferação de células tumorais em modelo animal e pior prognóstico (Nakamura *et al.* 2008). Por isso, o estudo dos mecanismos envolvidos com essa regulação poderá contribuir para a elucidação do possível papel supressor de miR-342.

Em resumo, os resultados deste trabalho corroboram a idéia de que o miR-342 promove a apoptose. Embora o efeito fenotípico da superexpressão do miR-342 não tenha sido detectável, os resultados indicam que o papel de miR-342 pode estar relacionado com o ajuste fino da expressão de genes da via apoptótica (*e.g.*, PDGFB), além de influenciar indiretamente a expressão de outros genes dessa via. Portanto, determinar quais são os principais alvos desse miRNA em tal contexto e como são regulados contribuirá para o entendimento dos mecanismos envolvidos com a apoptose e resistência à drogas mediado por miR-342.

## *6 Conclusões*

Nesta Tese, estudamos os miRNAs sob perspectivas da sua organização genômica, evolução e expressão diferencial em câncer. Tendo como foco a organização de miRNAs em regiões gênicas, construímos um banco de dados que integra informações de anotação, contexto genômico e expressão de miRNAs e de genes codificadores de proteínas para cinco espécies de vertebrados. Este banco serviu de base para a implementação de uma ferramenta *web* (miRIAD) cujo principal objetivo é o de facilitar a análise de miRNAs e genes codificadores, especialmente para os miRNAs intragênicos e genes hospedeiros. Acreditamos que a disponibilização desses dados pode contribuir para a geração de hipóteses relacionadas aos processos regulatórios relacionados com esta particular organização genômica, conforme exemplificamos para os genes *CLCN5* e *HTR2C* e seus miRNAs intragênicos. As estatísticas de acesso têm revelado um uso intenso e contínuo do miRIAD, o que vai ao encontro do propósito para o qual a ferramenta foi concebida.

Demonstramos evidências consistentes de que o contexto genômico e as características dos genes hospedeiros exercem forte influência na expressão e evolução dos miRNAs. Propomos que a idade dos genes hospedeiros é essencial na determinação dos padrões de expressão dos miRNAs intragênicos, relacionando-se com a divergência de expressão a longo prazo e com maior abrangência de expressão para os miRNAs intragênicos jovens. O fato dos miRNAs jovens residirem preferencialmente em genes antigos, e estes últimos conferirem maior abrangência de expressão aos primeiros, nos permitiu propor que o ambiente transcricionalmente ativo dos genes hospedeiros antigos pode acelerar o processo de funcionalização de novos miRNAs, facilitando a sua expressão em um maior número de tecidos e conseqüentemente expondo-os a um repertório diverso de alvos. Além disso, é possível que a evolução dos novos miRNAs intragênicos tenha tido um papel importante no processo de canalização, particularmente em tecidos neurais. Em contrapartida, nossos achados mostraram que miRNAs intergênicos jovens são mais tecido-específicos que os intragênicos e sofrem a

influência dos genes vizinhos, em termos de expressão. Dessa forma, sugerimos que a evolução de novos miRNAs está intimamente relacionada com o contexto genômico no qual se originam.

Na análise de expressão diferencial de miRNAs em linhagens sensíveis e resistentes às drogas 5-FU e Oxaliplatina, identificamos miRNAs associados com a apoptose. Em especial, o miR-342 mostrou-se o mais relevante para o modelo estudado, suportado por um importante conjunto de evidências. Observamos que este miRNA tem maior expressão nas linhagens sensíveis à Oxaliplatina e seus alvos preditos estão significativamente associados com a apoptose. Embora a superexpressão de miR-342 nas linhagens HCT116 e SW620 não tenha evidenciado alterações no ciclo celular e taxas de apoptose, observamos alterações na expressão de genes da via apoptótica, com destaque para uma redução significativa da expressão de um de seus alvos preditos, o PDGFB. Por fim, enfatizamos que a validação da interação entre miR-342 e PDGFB, e a escolha de um modelo experimental que favoreça a detecção dos efeitos fenotípicos dependentes de PDGFB poderá trazer importantes avanços para este estudo.

## *Referências*

- Albà MM, Castresana J. 2005. Inverse relationship between evolutionary rate and age of mammalian genes. *Mol Biol Evol* **22**: 598–606.
- Allen KE, Weiss GJ. 2010. Resistance may not be futile: microRNA biomarkers for chemoresistance and potential therapeutics. *Mol Cancer Ther* **9**: 3126–36.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H. 2005. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res* **33**: 2697–706.
- Alvarez-Saavedra E, Horvitz HR. 2010. Many Families of *C. elegans* MicroRNAs Are Not Essential for Development or Viability. *Curr Biol* **20**: 367–373.
- Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-jones SAM, Marshall M, *et al.* 2003. A uniform system for microRNA annotation. *RNA* **9**: 277–279.
- Ambros V, Horvitz HR. 1987. The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events. *Genes Dev* **1**: 398–414.
- Ameres SL, Zamore PD. 2013. Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* **14**: 475–88.
- Anders S, Huber W. 2010. DESeq: Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.
- Andrae J, Gallini R, Betsholtz C. 2008. Role of platelet-derived growth factors in physiology and medicine. *Genes Dev* **22**: 1276–1312.
- André T, Boni C, Mounedji-Boudiaf L, Navarro M, Tabernero J, Hickish T, Topham C, Zaninelli M, Clingan P, Bridgewater J, *et al.* 2004. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med* **350**: 2343–51.
- Arango D, Wilson a J, Shi Q, Corner G a, Arañes MJ, Nicholas C, Lesser M, Mariadason JM, Augenlicht LH. 2004. Molecular mechanisms of action and prediction of response to oxaliplatin in colorectal cancer cells. *Br J Cancer* **91**: 1931–46.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.
- Barbash S, Shifman S, Soreq H. 2014. Global coevolution of human micrnas and their target genes. *Mol Biol Evol* **31**: 1237–1247.
- Barik S. 2008. An intronic microRNA silences genes that are functionally antagonistic to its host gene. *Nucleic Acids Res* **36**: 5232–41.

- Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–33.
- Baskerville S, Bartel DP. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**: 241–7.
- Bazzini A a, Lee MT, Giraldez AJ. 2012. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**: 233–7.
- Berezikov E. 2011. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* **12**: 846–60.
- Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RH a. 2006. Diversity of microRNAs in human and chimpanzee brain. *Nat Genet* **38**: 1375–7.
- Bhandari V, Choo-Wing R, Lee CG, Zhu Z, NedreLOW JH, Chupp GL, Zhang X, Matthay MA, Ware LB, Homer RJ, *et al.* 2006. Hyperoxia causes angiopoietin 2-mediated acute lung injury and necrotic cell death. *Nat Med* **12**: 1286–1293.
- Bohnsack MT, Czaplinski K, Gorlich D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**: 185–191.
- Bose D, Zimmerman LJ, Pierobon M, Petricoin E, Tozzi F, Parikh a, Fan F, Dallas N, Xia L, Gaur P, *et al.* 2011. Chemoresistant colorectal cancer cells and cancer stem cells mediate growth and survival of bystander cells. *Br J Cancer* **105**: 1759–67.
- Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, *et al.* 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478**: 343–8.
- Bushati N, Cohen SM. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23**: 175–205.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**: 1915–27.
- Cai X, Hagedorn CH, Cullen BR. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**: 1957–1966.
- Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, *et al.* 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* **101**: 2999–3004.
- Campo-Paysaa F, Sémon M, Cameron RA, Peterson KJ, Schubert M. 2011. microRNA complements in deuterostomes: origin and evolution of microRNAs. *Evol Dev* **13**: 15–27.
- Carmel L, Rogozin IB, Wolf YI, Koonin E V. 2007. Evolutionarily conserved genes preferentially accumulate introns. *Genome Res* **17**: 1045–1050.

- Carroll SB. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**: 577–580.
- Chang T-C, Mendell JT. 2007. microRNAs in vertebrate physiology and human disease. *Annu Rev Genomics Hum Genet* **8**: 215–39.
- Chao DT, Korsmeyer SJ. 1998. BCL-2 family: regulators of cell death. *Annu Rev Immunol* **16**: 395–419.
- Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* **8**: 93–103.
- Chen W-H, Trachana K, Lercher MJ, Bork P. 2012. Younger genes are less likely to be essential than older genes, and duplicates are less likely to be essential than singletons of the same age. *Mol Biol Evol* **29**: 1703–6.
- Chiyoumaru T, Fukuhara S, Saini S, Majid S, Deng G, Shahryari V, Chang I, Tanaka Y, Enokida H, Nakagawa M, *et al.* 2014. Long non-coding RNA hotair is targeted and regulated by MIR-141 in human cancer cells. *J Biol Chem* **289**: 12550–12565.
- Cittelly DM, Das PM, Spoelstra NS, Edgerton SM, Richer JK, Thor AD, Jones FE. 2010. Downregulation of miR-342 is associated with tamoxifen resistant breast tumors. *Mol Cancer* **9**: 317.
- Da Cunha JPC, Galante PAF, de Souza JE, de Souza RF, Carvalho PM, Ohara DT, Moura RP, Oba-Shinja SM, Marie SKN, Silva WA, *et al.* 2009. Bioinformatics construction of the human cell surfaceome. *Proc Natl Acad Sci U S A* **106**: 16752–16757.
- Dar A a, Majid S, de Semir D, Nosrati M, Bezrookove V, Kashani-Sabet M. 2011. miRNA-205 suppresses melanoma cell proliferation and induces senescence via regulation of E2F1 protein. *J Biol Chem* **286**: 16606–14.
- De Wit E, Linsen SE V, Cuppen E, Berezikov E. 2009. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res* **19**: 2064–2074.
- Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231–235.
- Devor EJ. 2006. Primate microRNAs miR-220 and miR-492 lie within processed pseudogenes. *J Hered* **97**: 186–190.
- Dhahbi JM, Atamna H, Boffelli D, Magis W, Spindler SR, Martin DIK. 2011. Deep sequencing reveals novel microRNAs and regulation of microRNA expression during cell senescence. *PLoS One* **6**: e20509.
- Dill H, Linder B, Fehr A, Fischer U. 2012. Intronic miR-26b controls neuronal differentiation by repressing its host transcript, *ctdsp2*. *Genes Dev* **26**: 25–30.
- Djuranovic S, Nahvi A, Green R. 2012. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336**: 237–40.

- Ebert MS, Sharp P a. 2012. Roles for microRNAs in conferring robustness to biological processes. *Cell* **149**: 515–24.
- Ebisuya M, Yamamoto T, Nakajima M, Nishida E. 2008. Ripples from neighbouring transcription. *Nat Cell Biol* **10**: 1106–13.
- Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G. 2008. A Human snoRNA with MicroRNA-Like Functions. *Mol Cell* **32**: 519–528.
- Fabian MR, Sonenberg N. 2012. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat Struct Mol Biol* **19**: 586–93.
- Fabian MR, Sonenberg N, Filipowicz W. 2010. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79**: 351–379.
- Faivre S, Chan D, Salinas R, Woynarowska B, Woynarowski JM. 2003. DNA strand breaks and apoptosis induced by oxaliplatin in cancer cells. *Biochem Pharmacol* **66**: 225–237.
- Friedman RC, Farh KK-H, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**: 92–105.
- Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, *et al.* 2011. The UCSC genome browser database: Update 2011. *Nucleic Acids Res* **39**.
- Gennarino VA, D'Angelo G, Dharmalingam G, Fernandez S, Russolillo G, Sanges R, Mutarelli M, Belcastro V, Ballabio A, Verde P, *et al.* 2012. Identification of microRNA-regulated gene networks by expression analysis of target genes. *Genome Res* **22**: 1163–72.
- Glazko G, Mushegian A. 2010. Measuring gene expression divergence: the distance to keep. *Biol Direct* **5**: 51.
- Grady WM, Parkin RK, Mitchell PS, Lee JH, Kim Y-H, Tsuchiya KD, Washington MK, Paraskeva C, Willson JK V, Kaz a M, *et al.* 2008. Epigenetic silencing of the intronic microRNA hsa-miR-342 and its host gene EVL in colorectal cancer. *Oncogene* **27**: 3880–8.
- Grand T, L'Hoste S, Mordasini D, Defontaine N, Keck M, Pennaforte T, Genete M, Laghmani K, Teulon J, Lourdel S. 2011. Heterogeneity in the processing of CLCN5 mutants related to Dent disease. *Hum Mutat* **32**: 476–483.
- Grimson A, Farh KK-H, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**: 91–105.
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**: 1193–7.

- Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**: 835–40.
- Haendeler J, Hoffmann J, Rahman S, Zeiher AM, Dimmeler S. 2003. Regulation of telomerase activity and anti-apoptotic function by protein-protein interaction and phosphorylation. *FEBS Lett* **536**: 180–186.
- Hamfjord J, Stangeland AM, Hughes T, Skrede ML, Tveit KM, Ikdahl T, Kure EH. 2012. Differential expression of miRNAs in colorectal cancer: comparison of paired tumor tissue and adjacent normal mucosa using high-throughput sequencing. *PLoS One* **7**: e34150.
- Hausser J, Zavolan M. 2014. Identification and consequences of miRNA-target interactions - beyond repression of gene expression. *Nat Rev Genet*.
- Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ. 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci U S A* **105**: 2946–50.
- Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF. 2006. The expansion of the metazoan microRNA repertoire. *BMC Genomics* **7**: 25.
- Hinske LC, França GS, Torres H a. M, Ohara DT, Lopes-Ramos CM, Heyn J, Reis LFL, Ohno-Machado L, Kreth S, Galante P a. F. 2014. miRIAD-integrating microRNA inter- and intragenic data. *Database (Oxford)* **2014**: 1–9.
- Hinske LCG, Galante P a F, Kuo WP, Ohno-Machado L. 2010. A potential role for intragenic miRNAs on their hosts' interactome. *BMC Genomics* **11**: 533.
- Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. *Nat Genet* **38** **Suppl**: S20–4.
- Hu HY, Guo S, Xi J, Yan Z, Fu N, Zhang X, Menzel C, Liang H, Yang H, Zhao M, *et al.* 2011. MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genet* **7**: e1002327.
- Hu HY, He L, Fominykh K, Yan Z, Guo S, Zhang X, Taylor MS, Tang L, Li J, Liu J, *et al.* 2012. Evolution of the human-specific microRNA miR-941. *Nat Commun* **3**: 1145.
- Hu W, Chan CS, Wu R, Zhang C, Sun Y, Song JS, Tang LH, Levine AJ, Feng Z. 2010. Negative regulation of tumor suppressor p53 by microRNA miR-504. *Mol Cell* **38**: 689–99.
- Hu W, Collier J. 2012. What comes first: translational repression or mRNA degradation? The deepening mystery of microRNA function. *Cell Res* **22**: 1322–1324.
- Huang DW, Sherman BT, Lempicki R a. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.

- Huff V. 2011. Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene. *Nat Rev Cancer* **11**: 111–121.
- Huntzinger E, Izaurralde E. 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* **12**: 99–110.
- Hutvagner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, Zamore PD. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834–838.
- Iwama H, Kato K, Imachi H, Murao K, Masaki T. 2013. Human microRNAs originated from two periods at accelerated rates in mammalian evolution. *Mol Biol Evol* **30**: 613–26.
- Jansen BJH, Sama IE, Eleveld-Trancikova D, van Hout-Kuijjer M a, Jansen JH, Huynen M a, Adema GJ. 2011. MicroRNA genes preferentially expressed in dendritic cells contain sites for conserved transcription factor binding motifs in their promoters. *BMC Genomics* **12**: 330.
- Jordan IK, Mariño-Ramírez L, Koonin E V. 2005. Evolutionary significance of gene expression divergence. *Gene* **345**: 119–26.
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res* **20**: 1313–1326.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**.
- Kent WJ. 2002. BLAT - The BLAST-like alignment tool. *Genome Res* **12**: 656–664.
- Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Pääbo S. 2004. A neutral model of transcriptome evolution. *PLoS Biol* **2**: E132.
- Khvorova A, Reynolds A, Jayasena SD. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209–216.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.
- Kinoshita M, Ono K, Horie T, Nagao K, Nishi H, Kuwabara Y, Takanabe-Mori R, Hasegawa K, Kita T, Kimura T. 2010. Regulation of adipocyte differentiation by activation of serotonin (5-HT) receptors 5-HT2AR and 5-HT2CR and involvement of microRNA-448-mediated repression of KLF5. *Mol Endocrinol* **24**: 1978–1987.
- Kos A, Olde Loohuis NFM, Wiczorek ML, Glennon JC, Martens GJM, Kolk SM, Aschrafi A. 2012. A potential regulatory role for intronic microRNA-338-3p for its host gene encoding apoptosis-associated Tyrosine Kinase. *PLoS One* **7**: e31022.
- Kosik KS. 2009. MicroRNAs tell an evo-devo story. *Nat Rev Neurosci* **10**: 754–9.

- Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152–7.
- Krol J, Loedige I, Filipowicz W. 2010. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* **11**: 597–610.
- Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853–858.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace I, Wilm A, Lopez R, *et al.* 2007. ClustalW and ClustalX version 2. *Bioinformatics* **23**: 2947–2948.
- Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843–54.
- Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, Kim VN. 2004. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**: 4051–4060.
- Lemos B, Meiklejohn CD, Cáceres M, Hartl DL. 2005. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59**: 126–37.
- Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature* **424**: 147–51.
- Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15–20.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Li J, Liu Y, Dong D, Zhang Z. 2010. Evolution of an X-Linked Primate-Specific Micro RNA Cluster. *Mol Biol Evol* **27**: 671–683.
- Li J, Zhang Z. 2013. miRNA regulatory variation in human evolution. *Trends Genet* **29**: 116–24.
- Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q. 2014. HMDD v2.0: A database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* **42**.
- Liu M, Roth A, Yu M, Morris R, Bersani F, Rivera MN, Lu J, Shioda T, Vasudevan S, Ramaswamy S, *et al.* 2013. The IGF2 intronic miR-483 selectively enhances transcription from IGF2 fetal promoters and enhances tumorigenesis. *Genes Dev* **27**: 2543–8.

- Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, Hatzimichael E, Kirino Y, Honda S, Lally M, *et al.* 2015. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc Natl Acad Sci U S A* 201420955.
- Longley DB, Harkin DP, Johnston PG. 2003. 5-Fluorouracil: Mechanisms of Action and Clinical Strategies. *Nat Rev Cancer* 3: 330–8.
- Lu J, Clark AG. 2012. Impact of microRNA regulation on variation in human gene expression. *Genome Res* 22: 1243–54.
- Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu C-I. 2008. The birth and death of microRNA genes in *Drosophila*. *Nat Genet* 40: 351–5.
- Lutter D, Marr C, Krumsiek J, Lang EW, Theis FJ. 2010. Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. *BMC Genomics* 11: 224.
- Lyu Y, Shen Y, Li H, Chen Y, Guo L, Zhao Y, Hungate E, Shi S, Wu C-I, Tang T. 2014. New microRNAs in *Drosophila*--birth, death and cycles of adaptive evolution. *PLoS Genet* 10: e1004096.
- Marco A, Griffiths-Jones S. 2012. Detection of microRNAs in color space. *Bioinformatics* 28: 318–23.
- Mariadason JM, Arango D, Shi Q, Wilson AJ, Corner G a, Nicholas C, Aranes MJ, Lesser M, Schwartz EL, Augenlicht LH. 2003. Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. *Cancer Res* 63: 8791–812.
- Marsico A, Huska MR, Lasserre J, Hu H, Vucicevic D, Musahl A, Orom UA, Vingron M. 2013. PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol* 14: R84.
- Mazeh H, Mizrahi I, Ilyayev N, Halle D, Brücher B, Bilchik A, Protic M, Daumer M, Stojadinovic A, Itzhak A, *et al.* 2013. The Diagnostic and Prognostic Role of microRNA in Colorectal Cancer - a Comprehensive review. *J Cancer* 4: 281–95.
- McIlwain DR, Berger T, Mak TW. 2013. Caspase functions in cell death and disease. *Cold Spring Harb Perspect Biol* 5: 1–28.
- Mendes ND, Freitas AT, Sagot MF. 2009. Survey and Summary: Current tools for the identification of miRNA genes and their targets. *Nucleic Acids Res* 37: 2419–2433.
- Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H. 2013. Birth and expression evolution of mammalian microRNA genes. *Genome Res* 23: 34–45.

- Miska EA, Alvarez-Saavedra E, Abbott AL, Lau NC, Hellman AB, McGonagle SM, Bartel DP, Ambros VR, Horvitz HR. 2007. Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet* **3**: 2395–2403.
- Mohammed J, Flynt AS, Siepel A, Lai EC. 2013. The impact of age, biogenesis, and genomic clustering on *Drosophila* microRNA evolution. *RNA* **19**: 1295–308.
- Muiños-Gimeno M, Espinosa-Parrilla Y, Guidi M, Kagerbauer B, Sipilä T, Maron E, Pettai K, Kananen L, Navinés R, Martín-Santos R, *et al.* 2011. Human microRNAs miR-22, miR-138-2, miR-148a, and miR-488 are associated with panic disorder and regulate several anxiety candidate genes and related pathways. *Biol Psychiatry* **69**: 526–533.
- Nakamura Y, Tanaka F, Yoshikawa Y, Mimori K, Inoue H, Yanaga K, Mori M. 2008. PDGF-BB is a novel prognostic factor in colorectal cancer. *Ann Surg Oncol* **15**: 2129–2136.
- Necsulea A, Kaessmann H. 2014. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet*.
- Orre LM, Panizza E, Kaminsky VO, Vernet E, Gräslund T, Zhivotovsky B, Lehtiö J. 2013. S100A4 interacts with p53 in the nucleus and promotes p53 degradation. *Oncogene* **32**: 5531–40.
- Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, Zhang X, Song JS, Fisher DE. 2008. Chromatin structure analyses identify miRNA promoters. *Genes Dev* **22**: 3172–3183.
- Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG. 2013. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res* **41**.
- Park J, Xu K, Park T, Yi S V. 2012. What are the determinants of gene expression levels and breadths in the human genome? *Hum Mol Genet* **21**: 46–56.
- Parkin DM, Bray F, Ferlay J, Pisani P. 2002. Global cancer statistics, 2002. *CA Cancer J Clin* **55**: 74–108.
- Pasquinelli AE. 2012. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet* **13**: 271–82.
- Piriyapongsa J, Mariño-Ramírez L, Jordan IK. 2007. Origin and evolution of human microRNAs from transposable elements. *Genetics* **176**: 1323–1337.
- Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**: 1033–8.
- Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**: 110–21.

- Popadin KY, Gutierrez-arcelus M, Lappalainen T, Buil A, Steinberg J, Nikolaev SI, Lukowski SW, Bazykin GA, Seplyarskiy VB, Ioannidis P, *et al.* 2014. Gene age predicts the strength of purifying selection acting on gene expression variation in humans. *Am J Hum Genet* **95**:660–674.
- Pritchard CC, Cheng HH, Tewari M. 2012. MicroRNA profiling: approaches and considerations. *Nat Rev Genet* **13**: 358–69.
- Qian X, Yu J, Yin Y, He J, Wang L, Li Q, Zhang L-Q, Li C-Y, Shi Z-M, Xu Q, *et al.* 2013. MicroRNA-143 inhibits tumor growth and angiogenesis and sensitizes chemosensitivity to oxaliplatin in colorectal cancers. *Cell Cycle* **12**: 1385–94.
- Quah S, Hui JHL, Holland PWH. 2015. A Burst of miRNA Innovation in the Early Evolution of Butterflies and Moths. *Mol Biol Evol.*
- Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- Ramsköld D, Wang ET, Burge CB, Sandberg R. 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol* **5**.
- Rettig WJ. 1992. Immunogenetics of cell surface antigens of human cancer. *Curr Opin Immunol* **4**: 630–640.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**: 1902–10.
- Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ. 2010. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol* **11**: R6.
- Schulte JH, Marschall T, Martin M, Rosenstiel P, Mestdagh P, Schlierf S, Thor T, Vandesompele J, Eggert A, Schreiber S, *et al.* 2010. Deep sequencing reveals differential expression of microRNAs in favorable versus unfavorable neuroblastoma. *Nucleic Acids Res* **38**: 5919–28.
- Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**: 58–63.
- Seyednasrollah F, Laiho A, Elo LL. 2013. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform* 1–12.
- Slaby O, Svoboda M, Michalek J, Vyzula R. 2009. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. *Mol Cancer* **8**: 102.

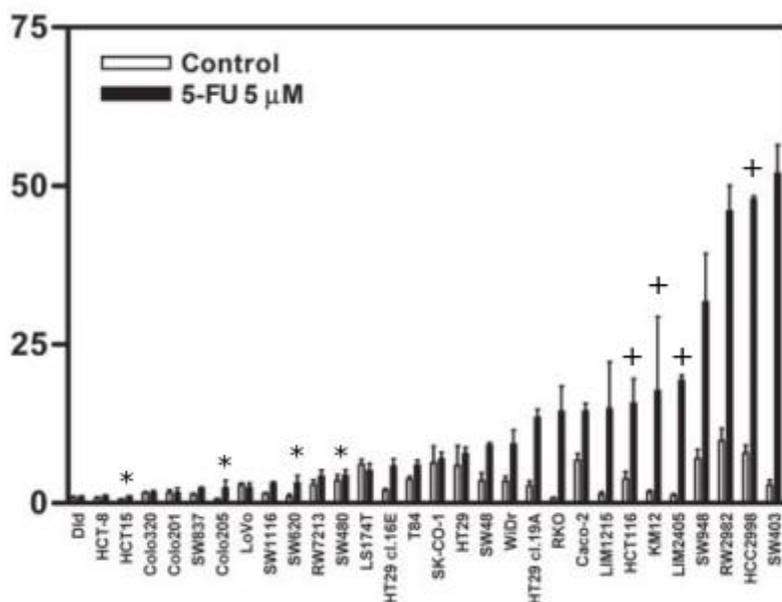
- Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthès P, Kokkinaki M, Nef S, Gnirke A, *et al.* 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep* **3**: 2179–90.
- Stefani G, Slack FJ. 2008. Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* **9**: 219–30.
- Takaku M, Machida S, Hosoya N, Nakayama S, Takizawa Y, Sakane I, Shibata T, Miyagawa K, Kurumizaka H. 2009a. Recombination activator function of the novel RAD51- and RAD51B-binding protein, human EVL. *J Biol Chem* **284**: 14326–36.
- Takaku M, Machida S, Nakayama S, Takahashi D, Kurumizaka H. 2009b. Biochemical analysis of the human EVL domains in homologous recombination. *FEBS J* **276**: 5841–8.
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562–78.
- Valeri N, Gasparini P, Braconi C, Paone A, Lovat F, Fabbri M, Sumani KM, Alder H, Amadori D, Patel T, *et al.* 2010. MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2). *Proc Natl Acad Sci U S A* **107**: 21098–103.
- Volinia S, Calin G a, Liu C-G, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, *et al.* 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* **103**: 2257–61.
- Wang D, Lippard SJ. 2005. Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov* **4**: 307–20.
- Wang H, Wu J, Meng X, Ying X, Zuo Y, Liu R, Pan Z, Kang T, Huang W. 2011. MicroRNA-342 inhibits colorectal cancer cell proliferation and invasion by directly targeting DNA methyltransferase 1. *Carcinogenesis* **32**: 1033–42.
- Watson, JD, Baker TA, Bell SP, Gann A, Levine M, Losick R, Harrison, SC. *Molecular Biology of the Gene*. 7ed. Pearson Education, Inc. 2014.
- Wightman B, Ha I, Ruvkun G, Hospital MG. 1993. Posttranscriptional Regulation of the Heterochronic Gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855–862.
- Wolf YI, Novichkov PS, Karev GP, Koonin E V, Lipman DJ. 2009. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. *Proc Natl Acad Sci U S A* **106**: 7273–80.
- Wu C-I, Shen Y, Tang T. 2009. Evolution under canalization and the dual roles of microRNAs: a hypothesis. *Genome Res* **19**: 734–43.

- Xu J, Wong C. 2008. A computational screen for mouse signaling pathways targeted by microRNA clusters. *RNA* **14**: 1276–1283.
- Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, *et al.* 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**: 650–9.
- Zehavi L, Avraham R, Barzilai A, Bar-Ilan D, Navon R, Sidi Y, Avni D, Leibowitz-Amit R. 2012. Silencing of a large microRNA cluster on human chromosome 14q32 in melanoma: biological effects of mir-376a and mir-376c on insulin growth factor 1 receptor. *Mol Cancer* **11**: 44.
- Zhang J, Cado D, Chen A, Kabra NH, Winoto A. 1998. Fas-mediated apoptosis and activation-induced T-cell proliferation are defective in mice lacking FADD/Mort1. *Nature* **392**: 296–300.
- Zhang R, Peng Y, Wang W, Su B. 2007. Rapid evolution of an X-linked microRNA cluster in primates. *Genome Res* **17**: 612–617.
- Zhang YE, Landback P, Vibranovski M, Long M. 2012. New genes expressed in human brains: implications for annotating evolving genomes. *Bioessays* **34**: 982–91.
- Zhang YE, Landback P, Vibranovski MD, Long M. 2011. Accelerated recruitment of new brain development genes into the human genome. *PLoS Biol* **9**: e1001179.
- Zhang YE, Vibranovski MD, Landback P, Marais G a B, Long M. 2010. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol* **8**.
- Zhu Y, Lu Y, Zhang Q, Liu J-J, Li T-J, Yang J-R, Zeng C, Zhuang S-M. 2012. MicroRNA-26a/b and their host genes cooperate to inhibit the G1/S transition by activating the pRb protein. *Nucleic Acids Res* **40**: 4615–25.

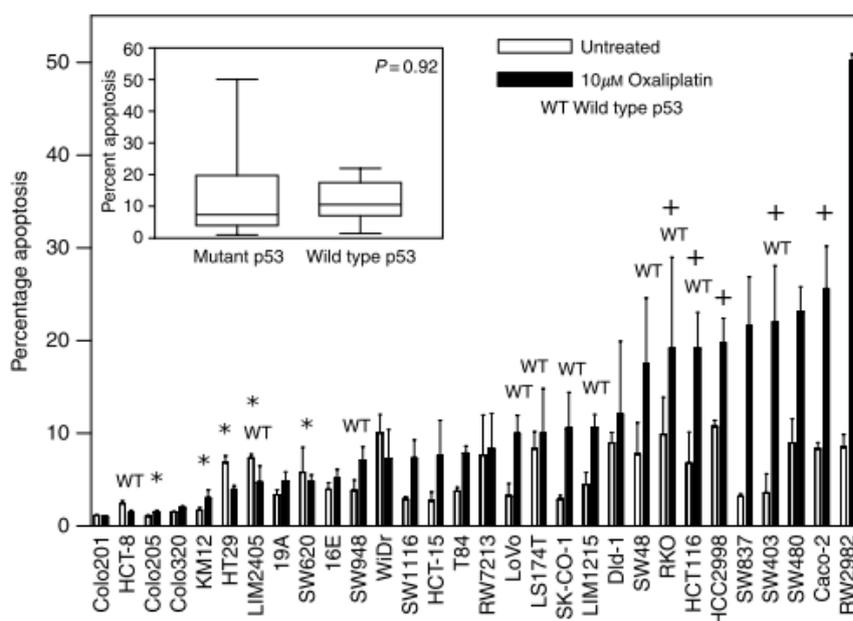
*Apêndice*

**Quadro 1. Linhagens celulares utilizadas no estudo de expressão diferencial de miRNAs.** Características e genotipagem das linhagens celulares. \*CIMP: CpG Island Methylator Phenotype. NA = não analisado.

Linhagem celular	Origem	Ras status	B-Raf status	p53 status	Mismatch repair Status	Expressão E-Caderina	CIMP status*
Caco-2	Carcinoma de cólon	Selvagem	Selvagem	Mutado	Funcional	Positiva	Negativo
Colo205	Adenocarcinoma coloretal derivado de sítio metastático: peritônio	Selvagem	Mutado V600E	Mutado	Funcional	NA	Positivo
HCC2998	Carcinoma de cólon	Mutado A146T	Selvagem	Mutado	Funcional	Positiva	Negativo
HCT116	Carcinoma de cólon	Mutado G13D	Selvagem	Selvagem	Deficiente	Positiva	Positivo
HCT-15	Adenocarcinoma de cólon	Mutado G13D	Selvagem	Mutado	Deficiente	NA	Positivo
HT29	Adenocarcinoma de cólon	Selvagem	Mutado V600E	Mutado	Funcional	Positiva	Positivo
KM12	Carcinoma de cólon	Selvagem	Selvagem	Mutado	Deficiente	Positiva	Positivo
LIM1215	Carcinoma de cólon	Selvagem	Selvagem	Selvagem	Deficiente	Positiva	Negativo
LIM2405	Carcinoma de cólon	Selvagem	Mutado V600E	Selvagem	Deficiente	Positiva	Positivo
RKO	Carcinoma de cólon	Selvagem	Mutado V600E	Selvagem	Deficiente	Negativa	Positivo
SW480	Adenocarcinoma de cólon	Mutado G12V	Selvagem	Mutado	Funcional	Positiva (fraco)	Negativo
SW620	Adenocarcinoma coloretal derivado de sítio metastático: linfonodo	Mutado G12V	Selvagem	Mutado	Funcional	NA	Negativo

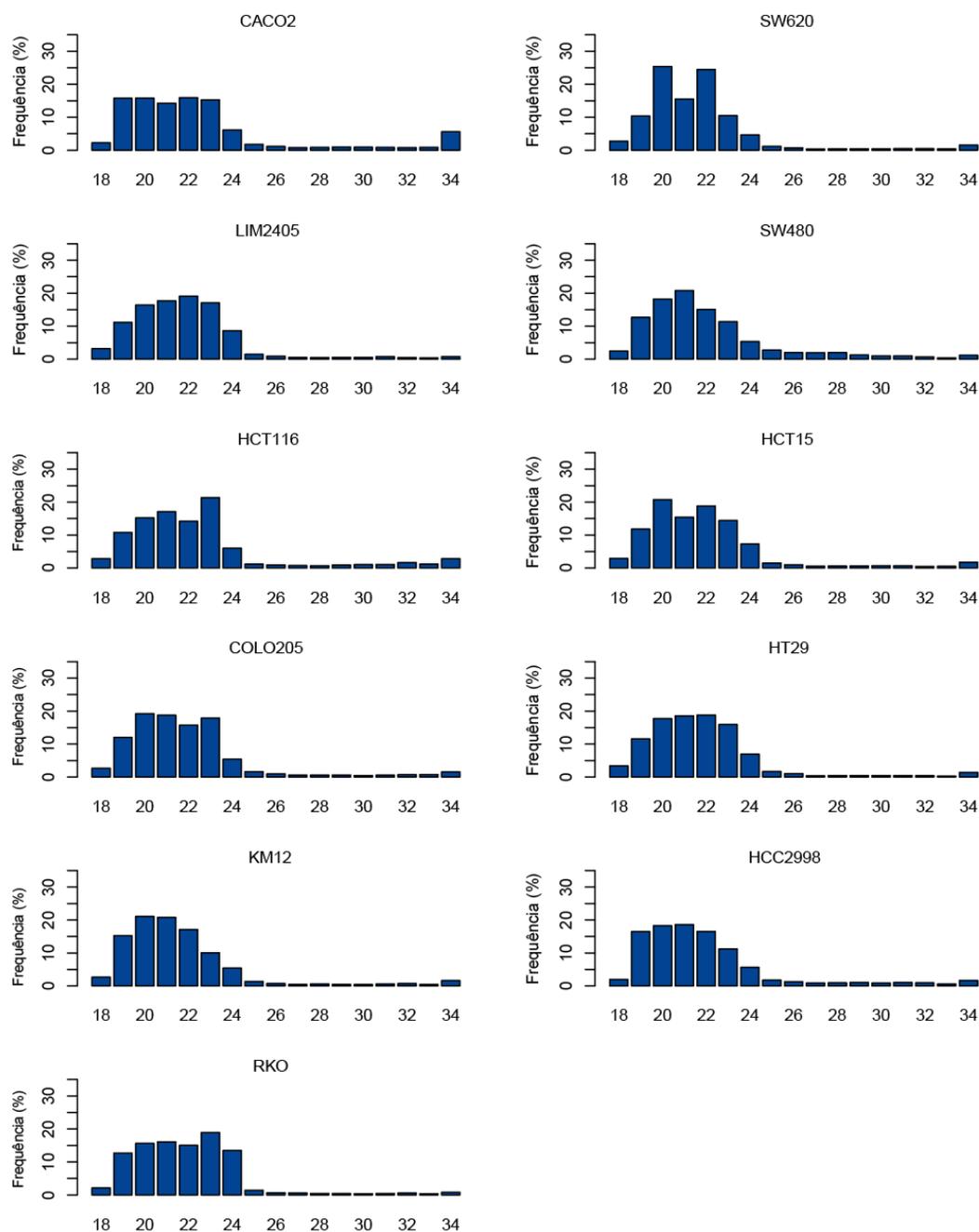


**Figura S1.** Taxas de apoptose em resposta ao tratamento com 5-FU em um painel de 30 linhagens de tumor coloretal. As células foram tratadas com 5  $\mu$ M por 72 h. Os experimentos foram feitos em triplicata. Os valores correspondem à média  $\pm$  desvio padrão. As linhagens marcadas com \* foram incluídas no grupo resistente e as linhagens marcadas com + incluídas no grupo sensível. Figura modificada de (Mariadason *et al.*, 2003). Detalhes metodológicos estão descritos no artigo original.



**Figura S2.** Taxas de apoptose em resposta ao tratamento com Oxaliplatina em um painel de 30 linhagens de tumor coloretal. As células foram expostas à 10  $\mu$ M por 72 h. Os experimentos foram feitos em

triplicata. Os valores correspondem à média  $\pm$  desvio padrão. As linhagens marcadas com \* foram incluídas no grupo resistente e as linhagens marcadas com + incluídas no grupo sensível. Figura modificada de (Arango *et al.*, 2004). Detalhes metodológicos estão descritos no artigo original.



**Figura S3. Frequências do número de *reads* mapeados no genoma humano de acordo com o tamanho da sequência em nucleotídeos.**

**Quadro 2. Experimento de qRT-PCR utilizando o kit SuperArray para genes da via apoptótica.** Os genes assinalados em vermelho apresentaram diminuição da expressão em relação ao controle (fold change > 2) e genes assinalados em verde apresentaram aumento da expressão em relação ao controle (fold change > 2).

Gene	Condições experimentais (valores CT)			$2^{-\Delta\Delta Ct}$	
	mirCtrl	mir342 16h	mir342 41h	mir342 16h	mir342 41h
AKT1	32.1	31.95	32.11	1.27	1.04
ANGPT1	22.49	22.39	23.25	1.07	0.59
ANGPT2	31.43	31.19	32.52	1.18	0.47
APAF1	31.72	31.02	31.95	1.62	0.85
ATM	<b>BAD</b>	<b>BAD</b>	<b>BAD</b>	0.00	0.00
BAD	31.36	30.99	31.87	1.29	0.70
BAX	25.89	25.42	26.12	1.39	0.85
BCL2	33.2	31.37	32.37	3.56	1.78
BCL2L1	22.92	22.87	22.57	1.04	1.27
BRCA1	27.78	27.16	28.56	1.54	0.58
CASP8	34.08	32.83	33.04	2.38	2.06
CCNE1	35	35	35	0.00	0.00
CDC25A	26.88	26.59	27.22	1.22	0.79
CDK2	<b>BAD</b>	<b>BAD</b>	<b>BAD</b>	0.00	0.00
CDK4	29.24	29.13	29.28	1.08	0.97
CDKN1A	27.81	28.05	28.1	0.85	0.82
CDKN2A	22.58	22.29	23.58	1.22	0.50
CFLAR	<b>BAD</b>	<b>BAD</b>	<b>BAD</b>	0.00	0.00
CHEK2	23.95	24.13	24.01	0.88	0.96
COL18A1	<b>BAD</b>	<b>BAD</b>	<b>BAD</b>	0.00	0.00
E2F1	34.28	33.66	33.79	1.54	1.40
ERBB2	35	35	35	0.00	0.00
ETS2	30.09	30.63	30.49	0.69	0.76
FAS	28.9	28.51	30.07	1.31	0.44
FGFR2	30.72	30.07	31.22	1.57	0.71
FOS	35	35	35	1.00	1.00
GZMA	<b>BAD</b>	<b>BAD</b>	<b>BAD</b>	0.00	0.00
HTATIP2	27.72	27.77	28.33	0.97	0.66
IFNA1	35	35	34.13	1.00	1.83
IFNB1	24.1	25.02	24.58	0.53	0.72
IGF1	27.6	27.86	28.09	0.84	0.71
IL8	31.21	31.24	31.98	0.98	0.59
ITGA1	35	35	35	0.00	0.00
ITGA2	28.47	28.55	29.28	0.95	0.57
ITGA3	35	35	35	1.00	1.00
ITGA4	35	35	35	1.00	1.00
ITGAV	35	35	35	1.00	1.00
ITGB1	32.27	31.39	32.2	1.84	1.05
ITGB3	27.03	26.47	27.93	1.47	0.54

ITGB5	BAD	BAD	BAD	0.00	0.00
JUN	35	35	35	1.00	1.00
MAP2K1	25.25	25.33	25.46	0.95	0.86
MCAM	30.82	30.32	31.77	1.41	0.52
MDM2	BAD	BAD	BAD	0.00	0.00
MET	BAD	BAD	BAD	0.00	0.00
MMP1	BAD	BAD	BAD	0.00	0.00
MMP2	35	35	35	1.00	1.00
MMP9	25.38	25.6	25.49	0.86	0.93
MTA1	35	35	35	0.00	0.00
MTA2	29.63	30.27	29.98	0.64	0.78
MTSS1	24.84	25.03	25.09	0.88	0.84
MYC	BAD	BAD	BAD	0.00	0.00
NFKB1	23.97	24.06	24.08	0.94	0.93
NFKBIA	25	25.06	25.71	0.96	0.61
NME1	BAD	BAD	BAD	0.00	0.00
NME4	33.56	33.13	37.59	1.35	0.06
PDGFA	23.81	24.24	23.81	0.74	1.00
PDGFB	31.37	34.06	32.11	0.15	0.60
PIK3R1	29.32	29.57	30.07	0.84	0.59
PLAU	35	35	35	0.00	0.00
PLAUR	28.22	28.21	28.39	1.01	0.89
PNN	29.39	29.54	29.76	0.90	0.77
RAF1	31.75	31.11	31.61	1.56	1.10
RB1	27.24	27.35	27.23	0.93	1.01
S100A4	27.83	27.23	29.94	1.52	0.23
SERPINB5	35	35	35	0.00	0.00
SERPINE1	34.57	32.83	34.58	3.34	0.99
SNCG	29.13	29.17	29.68	0.97	0.68
SYK	29.82	30.17	30.36	0.78	0.69
TEK	29.29	29.25	29.86	1.03	0.67
TERT	27.63	26.84	28.91	1.73	0.41
TGFB1	28.58	28.22	29.04	1.28	0.73
TGFBR1	33.43	32.96	32.55	1.39	1.84
THBS1	35	35	35	0.00	0.00
TIMP1	26.35	25.67	26.97	1.60	0.65
TIMP3	35	35	35	0.00	0.00
TNF	35	35	35	0.00	0.00
TNFRSF10B	27.16	26.95	27.21	1.16	0.97
TNFRSF1A	29.28	28.76	29.38	1.43	0.93
TNFRSF25	32.27	32.83	32.95	0.68	0.62
TP53	22.47	22.49	22.45	0.99	1.01

TWIST1	31.18	31.07	31.19	1.08	0.99
UCC1	27.7	27.46	28.32	1.18	0.65
VEGF	31.06	30.41	31.11	1.57	0.97
*18SrRNA	<b>BAD</b>	<b>BAD</b>	<b>BAD</b>	0.00	0.00
*HPRT1	21.9	22.2	21.8	0.81	1.04
*RPL13A	19.7	20.0	19.6	0.79	1.09
*GAPDH	22.8	22.4	23.4	1.27	0.66
*ACTB	18.1	18.0	18.5	1.05	0.76

\*controles endógenos

 *Downregulation*  
 *Upregulation*

**LISTA DE ANEXOS**

**Anexo A.** Súmula Curricular.

**Anexo B.** Artigo publicado: Database - The Journal of Biological Databases and Curation

**Anexo C.** Resumo do artigo submetido: *Host gene constraints and genomic context impact the expression and evolution of human microRNAs*

## **Anexo A – Súmula Curricular**

### **Dados Pessoais**

---

Nome: Gustavo Starvaggi França

Local e Data de Nascimento: São Paulo – SP, 17/04/1984

### **Educação**

---

#### **Doutorado em ciências – Bioquímica**

Universidade de São Paulo – USP

Instituto de Química, Departamento de Bioquímica

Orientadora: Dra. Anamaria A. Camargo

Co-orientador: Dr. Pedro A. F. Galante

2011-2015

#### **Mestrado em ciências – Bioquímica**

Universidade de São Paulo – USP

Instituto de Química, Departamento de Bioquímica

Orientador: Dr. Sandro José de Souza

2007-2010

#### **Bacharelado em Ciências Biológicas**

Universidade Federal do Paraná

2002-2006

### **Atividades Complementares**

---

#### **Estágio Supervisionado em Docência (PAE)**

Monitoria na Disciplina de Biologia Molecular (QBQ0212)

Departamento de Bioquímica, Instituto de Química – USP

2013-2013

#### **Bolsista de Treinamento Técnico TT4 (FAPESP)**

Departamento de Genética e Biologia Evolutiva – IB USP

2011-2011

#### **Bolsista de Iniciação Científica (CNPq)**

Departamento de Bioquímica – UFPR

2005-2006

### **Ocupação**

---

Bolsista de Doutorado, FAPESP.

2011-2015

Bolsista de Mestrado, FAPESP

2007-2009

## Publicações

---

- G.S. França**, M.D. Vibranovski and P.A.F. Galante. 2015. Host gene constraints and genomic context impact the expression and evolution of human microRNAs [submetido].
- E.C. Guedes, **G.S. França**, F.C. Koyama, L.N. Moreira, J. G. Alexandre, D. Figueroa, C.C. Irigoyen, P.A.F. Galante and G. P. Diniz. 2015. MicroRNA expression signature is altered in the cardiac remodeling induced by high fat diets. *Journal of Cellular Physiology*. Dec 7. doi: 10.1002/jcp.25280.
- R.S. Francisco, S. Buhler, J.M. Nunes, B. Bitarello, **G.S. França**, D. Meyer, A. Sanchez-Mazas. 2015. HLA supertype variation in human populations: new insights about the role of natural selection on the evolution of HLA-A and HLA-B polymorphisms. *Immunogenetics*, 67: 651-663
- C. Lopes-Ramos, F. C. Koyama, A. Habr-Gama, A.C.M. Salim, F. Bettoni, P.F. Asprino, **G. S. França**, J. Gama-Rodrigues, R.B. Parmigiani, R.O. Perez, P.A.F. Galante and A.A. Camargo. 2015. Comprehensive evaluation of the effectiveness of gene expression signatures to predict complete response to nCRT and guide surgical intervention in rectal cancer. *Cancer Genetics*, 208: 319-326.
- L.C. Hinske\*, **G. S. França\***, H.A.M. Torres, D.T. Ohara, C.M. Lopes-Ramos, J. Heyn, L.F.L. Reis, L. Ohno-Machado, S. Kreth and P.A.F. Galante. 2014. miRIAD – integrating microRNA inter- and intragenic data. *DATABASE-Oxford*, v. 2014, p. bau099. (\*equally contributed)
- G. S. França**, D.V. Cancherini and S.J. De Souza. 2012. Evolutionary history of exon shuffling. *Genetica*, 140: 249-257.
- D.V. Cancherini, **G. S. França** and S.J. De Souza. 2010. The role of exon shuffling in shaping protein-protein interaction networks. *BMC Genomics*, 11, S11.

## Participação em eventos

---

- 2015** The Biology of Genomes. Cold Spring Harbor Laboratory: Investigating the influence of the genomic context on the expression and evolution of human microRNAs. Cold Spring Harbor NY. Poster.
- 2014** SMBE 2014. Annual Meeting of the Society for Molecular Biology & Evolution: Host gene constraints impact the evolution of intragenic microRNAs. San Juan – Puerto Rico. Poster.
- 2014** Challenges and Solutions in Cancer Research and Treatment: microRNAs involved with chemoresistance in colorectal cancer. Rio de Janeiro – Brazil. Poster.
- 2013** X-meeting: 9<sup>th</sup> International Conference of the Brazilian Association for Bioinformatics and Computational Biology: Differentially regulated microRNAs and their associations with chemoresistance in colorectal cancer. Recife – Brazil. Poster.
- 2012** Intersections – 1<sup>st</sup> International Cooperative Cancer Symposium. Sao Paulo – Brazil. Poster

**2012** X-meeting: 8<sup>th</sup> International Conference of the Brazilian Association for Bioinformatics and Computational Biology: miRIAD: Intragenic microRNA database. Campinas – Brazil. Poster.

**2008** X-meeting: 4<sup>th</sup> International Conference of the Brazilian Association for Bioinformatics and Computational Biology – Evolutionary history of exon shuffling in eukaryotes. Salvador – Brazil. Poster.

**2007** X-meeting: 3<sup>th</sup> International Conference of the Brazilian Association for Bioinformatics and Computational Biology. Sao Paulo – Brazil. Poster.

**Anexo B - Artigo publicado:** Database – The Journal of Biological Databases and Curation



## Database tool

# miRIAD—integrating microRNA inter- and intragenic data

Ludwig Christian Hinske<sup>1,\*†</sup>, Gustavo S. França<sup>2,3,†</sup>, Hugo A. M. Torres<sup>2</sup>, Daniel T. Ohara<sup>2</sup>, Camila M. Lopes-Ramos<sup>2</sup>, Jens Heyn<sup>1</sup>, Luiz F. L. Reis<sup>2</sup>, Lucila Ohno-Machado<sup>4</sup>, Simone Kreth<sup>1</sup> and Pedro A. F. Galante<sup>2,\*</sup>

<sup>1</sup>Clinic of Anaesthesiology, Clinic of the University of Munich, Munich, Germany

<sup>2</sup>Centro de Oncologia Molecular, Hospital Sírio-Libanês, São Paulo, SP 01308-060, Brazil

<sup>3</sup>Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

<sup>4</sup>Division of Medial Informatics, University of California San Diego, La Jolla, CA 93093-0505, USA

\*Corresponding author: Tel: +55 11 3155 3704; Fax: +55 11 3155 4220; Email: pgalante@mochsl.org.br

Correspondence may also be addressed to Ludwig Christian Hinske. Tel: +49 89 4400 73410; Fax: +49 89 4400 78886; Email: christian.hinske@med.uni-muenchen.de

<sup>†</sup>These authors contributed equally to this work.

Citation details: Hinske,L.C., França,G.S., Torres,H.A.M., *et al.* miRIAD—integrating microRNA inter- and intragenic data. *Database* (2014) Vol. 2014: article ID bau099; doi:10.1093/database/bau099

Received 9 January 2014; Revised 4 September 2014; Accepted 9 September 2014

## Abstract

MicroRNAs (miRNAs) are a class of small (~22 nucleotides) non-coding RNAs that post-transcriptionally regulate gene expression by interacting with target mRNAs. A majority of miRNAs is located within intronic or exonic regions of protein-coding genes (host genes), and increasing evidence suggests a functional relationship between these miRNAs and their host genes. Here, we introduce miRIAD, a web-service to facilitate the analysis of genomic and structural features of intragenic miRNAs and their host genes for five species (human, rhesus monkey, mouse, chicken and opossum). miRIAD contains the genomic classification of all miRNAs (inter- and intragenic), as well as classification of all protein-coding genes into host or non-host genes (depending on whether they contain an intragenic miRNA or not). We collected and processed public data from several sources to provide a clear visualization of relevant knowledge related to intragenic miRNAs, such as host gene function, genomic context, names of and references to intragenic miRNAs, miRNA binding sites, clusters of intragenic miRNAs, miRNA and host gene expression across different tissues and expression correlation for intragenic miRNAs and their host genes. Protein–protein interaction data are also presented for functional network analysis of host genes. In summary, miRIAD was designed to help the research community to explore, in a user-friendly environment, intragenic miRNAs, their host genes and functional annotations with minimal effort, facilitating hypothesis generation and *in-silico* validations.

**Database URL:** <http://www.miriad-database.org>

## Introduction

Amongst regulatory mechanisms of gene expression in eukaryotes, microRNAs (miRNAs) have established a central role in the past two decades (1). These 22-nt short single-stranded RNA molecules guide the RNA-induced silencing complex to modulate the expression of target mRNAs (2). MicroRNA binding sites are most likely recognized by nucleotide sequences in the 3'-untranslated regions (3'-UTR) of target mRNAs. Binding of the miRNA-protein complexes to their targets results in either degradation or translational inhibition of the mRNA transcripts (2).

For humans, ~1900 miRNA genes have been identified (3), and more than half are located within genomic regions containing protein-coding genes (4–6). Hence, miRNA genes can be classified as either inter- or intragenic, and the latter sub-classified as intronic or exonic (4, 5). A substantial number of these intragenic miRNAs are co-transcribed, and consequently co-regulated with their host genes (4, 5, 7). Recent evidence suggests a functional linkage between intragenic miRNAs and their hosts on multiple levels, including direct and indirect interaction (8–10).

Despite the importance of these intragenic miRNAs, their exploration can be daunting, as much of the necessary information is not readily available and requires manual integration from multiple data sources (6, 11, 12). Although other databases exist that provide information related to intra- and intergenic miRNAs (12–15), some tools don't appear to be frequently updated (14), contain only an elementary set of information related to intragenic miRNAs and their host genes (13, 15) and/or their usage is complex and requires in-depth bioinformatics skills (12).

In the current manuscript, we present miRIAD, a web-service designed to examine intragenic miRNAs, their host genes and their functional annotations with a streamlined graphical data representation and an efficient information query system. miRIAD provides information regarding genomic context, gene function, gene interaction, miRNA targets and gene expression for five species, including human and mouse. miRIAD is publicly available at <http://www.miriad-database.org>.

## Materials and Methods

### Database architecture and raw data

Because miRIAD integrates a large set of data, processed information is stored in a MySQL relational database.

Supplementary Figure S1 provides an overview of the miRIAD database schema, its tables and their relations. To date, miRIAD consists of 60 tables in total, comprising 12 tables for each of the five species (human, rhesus monkey, mouse, opossum and chicken), containing ~10 million records of integrated information.

To construct miRIAD, we used several sets of publicly available data. The reference genomes (human genome sequence—GRCh37/hg19; rhesus genome—rheMac3; mouse genome—mm10/GRCm38; opossum genome—MonDom5; chicken genome—galGal4) were downloaded from UCSC Genome Browser (<http://genome.ucsc.edu>). The transcriptome sets were downloaded from the RefSeq project (<http://www.ncbi.nih.gov/refseq>) for all species. MicroRNA genomic coordinates, seed sequences and family information were retrieved from miRBase (<http://www.mirbase.org/>, release #20). Protein-protein interaction data were acquired from the Human Protein Reference Database (HPRD, downloaded from NCBI <http://www.ncbi.nih.gov>) and from EMBL's STRING database (<http://string-db.org/>). Gene expression data were obtained from Brawand *et al.* (16) (coding genes) and Meunier *et al.* (17) (miRNAs).

### Host gene and miRNA information

All known genes were classified either as host or non-host based on the presence of overlapping miRNAs for each species. This classification and additional information regarding known genes were stored in three tables (GeneInformation, GeneRegions and GeneSynonyms), as shown in Supplementary Figure S1.

All miRNA genes were classified either as intra- or intergenic, based on their genomic localization. The 'MirnaInformation' table contains the official name, genomic coordinates of the stem loop sequence and, if applicable, the host gene to which the miRNA is related. In case of multiple genes, the host gene assigned was the one on the same strand as the miRNA. If intronic, the intron number and the region length between the miRNA coordinates and the next exon upstream were calculated and stored.

### miRNA target prediction

miRIAD contains all conserved target sites within 3'UTRs from TargetScan (<http://www.targetscan.org/>, release #6.2) for human and mouse. In brief, TargetScan defines

miRNA targets by searching, within 3' UTR regions, for 8mer (exact match) and 7mer sites that match the seed region (position 2–7) of mature miRNAs. Information regarding interspecies conservation and match/mismatch profile are also used to define the final set of conserved targets (for further information, see <http://www.targetscan.org/>). miRIAD contains a total of 1141 miRNAs binding to 466569 mRNA targets from 14867 known protein coding genes for human. Target prediction information for human and mouse were directly downloaded from the TargetScan homepage (file `Conserved_Site_Context_Scores.txt`, release #62) and calculated for rhesus monkey, opossum and chicken miRNAs using the TargetScan tool kit, applied to all miRNAs and the 3'UTRs from these organisms.

### Gene and miRNA expression

To obtain expression for protein-coding genes, data from Brawand *et al.* (16) were downloaded from GEO (GSE30352) and aligned to the genome of each species using TopHat (version 2.0.8b) with default parameters (18). Normalized gene expression values for six tissues (brain, cerebellum, heart, liver, kidney and testis) from all species were computed by means of FPKM (19) with Cufflinks [version 2.2.1; (20)] using transcript annotations from Ensembl (version 71). To determine miRNA expression available for five tissues (brain, cerebellum, heart, kidney and testis) from all species, data from Meunier *et al.* (17) were downloaded from GEO (GSE40499) and reads were aligned to each genome with Bowtie version 1.0.0 using the following parameters: `-m 5 -v 0 -a -best -strata`. Only exact matches were considered, and reads aligned to >5 different loci were discarded. The 3' adaptors were removed using a sequential trimming strategy (21). Reads totally overlapping to mature miRNA coordinates annotated from miRBase (release 20) were counted and normalized for each species with EdgeR package version 2.6.12 (22). Host gene and intragenic miRNA expression correlations were calculated by Spearman's rank correlation using the normalized values (FPKM and CPM (counts per million) for coding genes and miRNAs, respectively).

## Results

### Database overview

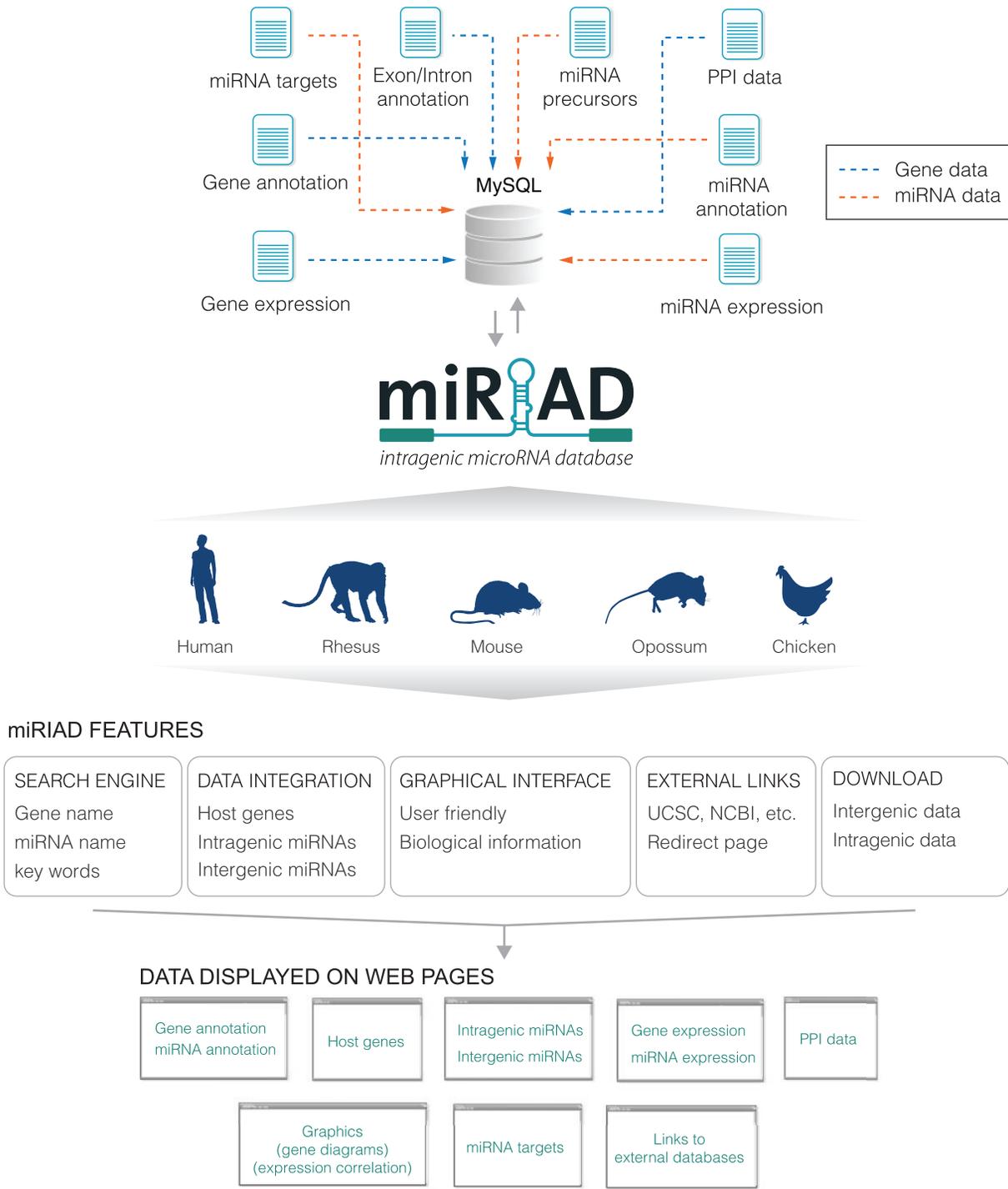
Figure 1 summarizes the main features, data sets and how information is presented in the miRIAD web tool. Most of miRIAD data related to intragenic miRNAs and their host genes is summarized in Table 1. To provide a useful platform, miRIAD integrates all known protein-coding genes

(~22k genes on average, for all five species), all known miRNAs (~900 on average, for all five species), miRNAs targets, validated and predicted protein–protein interactions and expression data for miRNAs and coding genes across five and six tissues, respectively. miRIAD classifies all miRNAs as intragenic or intergenic. It contains a total of 1072 (57%) for human; 167 (29%) for rhesus; 745 (63%) for mouse; 179 (40%) for opossum; 299 (52%) for chicken, additionally specifying whether or not they are transcribed in the same orientation as that of their host genes (84, 54, 87, 92 and 76% of intragenic miRNAs for human, rhesus, mouse, opossum and chicken, respectively). It is worth mentioning that some of the discrepancies between these percentages are likely due to the incompleteness of miRNA and gene annotation for individual species. As we can observe for human and mouse, which have the most complete annotated sets of coding genes and miRNAs, the values are quite similar. Additional complex information is also provided, such as the visualization of intragenic miRNAs within their host genes and positioning along the isoforms, expression correlation between intragenic miRNAs and their host genes, intragenic miRNAs binding to their own host genes and intragenic miRNAs binding to genes that are directly interacting with their host genes. These data are necessary in the identification and evaluation of putative negative or positive feedback mechanisms between miRNAs and host genes, (5, 23–25) and can offer a starting point for future analyses to reveal novel regulatory pathways.

### miRIAD query system

The miRIAD query system was developed and optimized to be fast, intuitive and functional. It lets the user search for several terms, such as miRNA symbol, gene name (Official Symbol, Ensembl ID, Entrez ID, HGNC ID or Gene Synonyms) and gene annotation keywords (e.g. 'oncogene', 'kinase', etc.). Searching for miRNAs follows the same principles as those used for coding genes, allowing for non-exact inputs (according to miRNA official nomenclature). It is also possible to query for multiple genes or miRNAs at once. The query system works in the same way for all five species.

The output for each searched term is a list of query matches organized by relevance, containing basic gene information for rapid inspection and selection. Names of host genes and intragenic miRNAs are readily identified by a particular tag (see web page for details). Moreover, non-host genes and intergenic miRNAs are also shown, because they may have indirect associations to intragenic miRNAs or host genes and are therefore also important. By clicking on a gene name, the user can access more



**Figure 1.** Overview of the miRIAD platform. Schematic representation of the main data presented in the web tool and how they are integrated and displayed. Blue arrows denote data related to protein-coding genes and orange arrows indicate data related to miRNAs. PPI: protein–protein interaction.

detailed information about any known coding or miRNA gene.

### Exploring host genes

In the recent past, it has become clear that functional aspects of intragenic miRNAs have to be viewed in the

context of their host genes (5, 7, 23, 24, 26). Therefore, information about all known protein-coding genes has been integrated into miRIAD to allow contextual search. For each protein-coding gene, miRIAD provides a ‘Summary’ section showing annotation data, such as official gene symbol, full gene name and gene name aliases, gene type and a gene function summary when publicly available.

**Table 1.** Summary of main miRIAD data

Data class	Human	Rhesus	Mouse	Opossum	Chicken
Known protein-coding genes	20 530	22 553	29 664	20 550	16 953
Known miRNA precursors	1871	582	1181	443	573
Intragenic miRNAs	1072	167	745	179	299
Intergenic miRNAs	799	415	435	264	272
Host genes	930	141	613	143	273
Sense miRNAs in respect to host orientation	902	90	645	145	90
Antisense miRNAs in respect to host orientation	170	77	95	12	28
Expressed coding genes	18 442	8 112	19 029	12 079	11 278
Expressed miRNAs	1111	475	784	405	465

Moreover, information regarding the genomic context, including the genomic position, transcription ‘start’ and ‘end’ and transcription orientation, is provided, as well as a graphical representation of the exon–intron structure of transcripts (Figure 2). If applicable, miRIAD presents miRNA name, genomic region (intronic/exonic), the intron/exon number where they are inserted, the distance to the closest upstream exon and transcriptional orientation, sense (miRNA and host in the same transcriptional orientation), or antisense (in opposite orientation). To facilitate the generation and evaluation of research hypotheses, expression data (based on RNA-Seq) of mRNAs across six tissues (brain, cerebellum, heart, kidney, liver and testis) as well as expression correlation between host genes and their intragenic miRNAs were included. All miRNAs potentially binding to a target gene are displayed under ‘miRNA binding sites’. Finally, the last section shows all known protein–protein interaction data for each gene. Cases in which interaction partners of a given host gene are targeted by its intragenic miRNA are explicitly shown. This kind of information is noteworthy because it can reveal unusual regulatory loops and may support findings or suggest future investigations. All these information are exemplified for the oncogene ERBB2 containing mir-4728 (Figure 2).

The gene section also provides links to external databases, such as NCBI Gene (<http://www.ncbi.nlm.nih.gov/gene/>), UCSC Genome Browser (<http://genome.ucsc.edu/>), Ensembl (<http://www.ensembl.org/>), KEGG (<http://www.genome.jp/kegg/>) and Targetscan (<http://www.targetscan.org/>). Most of these links are context-sensitive, easily redirecting the user to the gene of interest on the web page containing complementary data.

### Intragenic miRNAs

Intragenic miRNAs are the main focus of our web tool, even though we present information for all known miRNAs and protein-coding genes. For each pre-miRNA,

miRIAD provides a ‘Summary’ section with the official miRNA symbol, its full name, miRBase ID, target genes and the genomic context where each miRNA is mapped (Figure 3). For intragenic miRNAs, information about their intragenic position and location along the host genes are depicted by a graphical representation (Figure 3). Cases where an intragenic miRNA potentially targets its own host are highlighted for fast identification. Similar to the presentation of information about protein-coding genes, there are also expression data (based on RNAseq) for six tissues (brain, cerebellum, heart, kidney, liver and testis) and an expression correlation between intragenic miRNAs and their host genes. A set of context-sensitive links to external databases in the top right corner to access complementary information (miRBase, miRDB, Targetscan, mirgen, Magia, miRWalk and miRò) are also presented.

Figure 3 exemplifies the use of this information for mir-483 and its host IGF2. IGF2 produces the insulin-like growth factor 2, an essential protein for growth and development of the fetus and it is upregulated in several malignancies (27). According to our data, the expression of IGF2 and miR-483-5p are highly correlated ( $\rho = 0.7$ ). Accordingly, a recent report has uncovered a positive feedback between IGF2 and its intragenic mir-483, where the mature miR-483-5p molecule binds to the 5’UTR of IGF2 mRNA, promoting IGF2 transcription by facilitating the association of the helicase DHX9 (24).

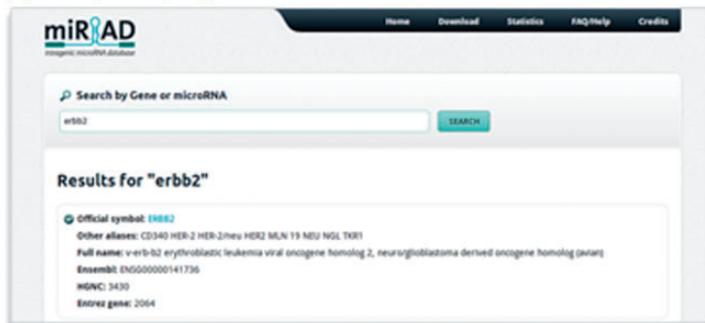
### Using miRIAD to explore a set of genes

In the following paragraph, we briefly illustrate how miRIAD can be used to explore a gene or a set of genes. Recently, da Cunha *et al.* (28) defined the set of all human genes coding for cell surface proteins (called surfaceome genes). These genes can be considered as potential targets for diagnostic and therapeutic interventions (28, 29).

The set of 3702 human surfaceome genes was retrieved from (28, 29) and submitted to miRIAD to initially be classified as host or non-host genes. In total, 119 surfaceome

## miRIAD Snapshots of Gene view

## (a) Example of query search results for "ERBB2"



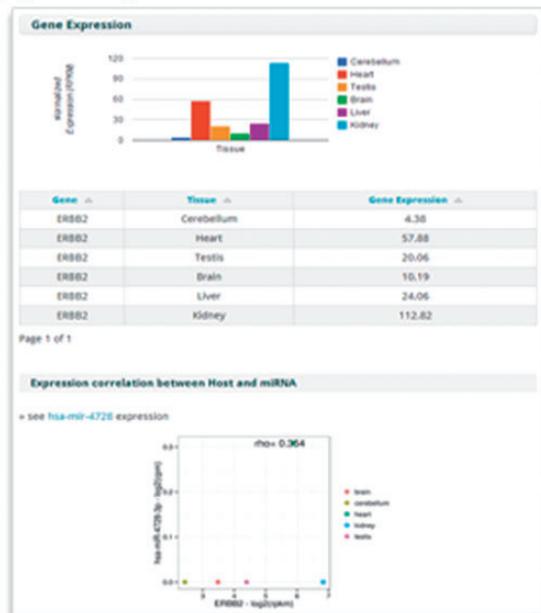
## (b) Detailed view of ERBB2 page: Summary and external links



## (c) Genomic context and graphical diagram



## (d) Gene expression data for ERBB2 and miR-4728



## (e) microRNA binding sites table

RefSeq ID	miRNA	Mature miRNA	Seed Match	Contact Score
NM_004448	hsa-miR-4555	hsa-miR-4555	Exact	-0.371
NM_001005862	hsa-miR-4555	hsa-miR-4555	Exact	-0.371
NM_004448	hsa-miR-4554	hsa-miR-4554	Exact	-0.206
NM_001005862	hsa-miR-4554	hsa-miR-4554	Exact	-0.206
NM_004448	hsa-miR-4516	hsa-miR-4516	Exact	-0.206
NM_001005862	hsa-miR-4516	hsa-miR-4516	Exact	-0.206

## (f) Protein-protein interactions table

Protein	Partner	Origin	Mode
ERBB2	SH2ONA3	hgnd	regular
ERBB2	ACFP	hgnd	regular
ERBB2	RTG	hgnd	regular
ERBB2	CAV1	hgnd	regular
ERBB2	CD44	hgnd	complex

## (g) Partners of ERBB2 targeted by its intragenic miR-4728

Name Symbol	RefSeq ID	miRNA	Mature miRNA	Seed Match	Contact Score
CAV1	NM_001172895	hsa-miR-4728	hsa-miR-4728-3p	Partial-mb	-0.047
CAV1	NM_001172896	hsa-miR-4728	hsa-miR-4728-3p	Partial-mb	-0.047
CAV1	NM_001172897	hsa-miR-4728	hsa-miR-4728-3p	Partial-mb	-0.047
CAV1	NM_001172898	hsa-miR-4728	hsa-miR-4728-3p	Partial-mb	-0.047
SH2	NM_001130040	hsa-miR-4728	hsa-miR-4728-3p	Exact	-0.268
SH2	NM_001130041	hsa-miR-4728	hsa-miR-4728-3p	Exact	-0.268
SH2	NM_001202859	hsa-miR-4728	hsa-miR-4728-3p	Exact	-0.268

Figure 2. A summary of the main information presented in miRIAD for the coding gene ERBB2 and its intragenic miR-4728.

genes are host genes for 150 intragenic miRNAs. Interestingly, most of these miRNAs (87.3%) are transcribed on the same orientation of their host genes, suggesting possible co-transcription (5). 140 of these intragenic miRNAs are actually inserted within intronic regions of surfaceome genes.

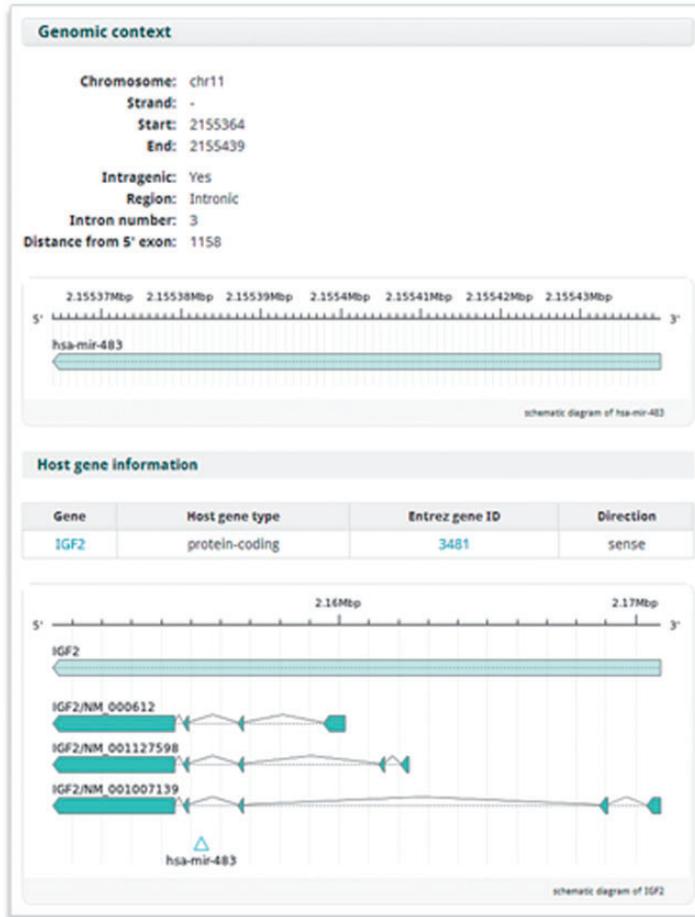
Next, we examined two genes in more detail. We selected the genes containing the largest number of intronic miRNAs, CLCN5 and HTR2C. In respect to CLCN5, mutations in its sequence have been proven to be associated

with diseases of renal tubules, resulting in chronic renal failure (30). This gene has eight intronic miRNAs, and surprisingly, some of their transcripts may be targeted by their intronic miR-502 (see miRIAD).

It is also striking that this host gene has isoforms starting transcription upstream of the miRNAs, which possibly could prevent co-expression between a CLCN5 transcript and those intronic miRNAs in some tissues or pathologies. Analysis of the expression data suggests co-expression or at least co-regulation between CLCN5 and its intronic

miRIAD Snapshots of microRNA view

(a) Example of the intragenic hsa-mir-483 and its host gene IGF2



(b) Expression of hsa-mir-483 and IGF2 correlation



(c) Potential targets of miR-483

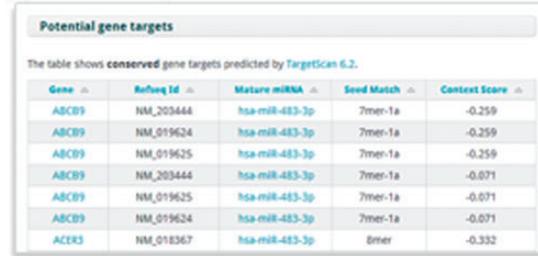


Figure 3. A summary of the main information presented in miRIAD for the intragenic mir-483 and its host gene IGF2.

miRNAs. CLCN5, as well as its intronic miRNAs are highly expressed in kidney. The expression correlations are high ( $\rho > 0.7$ , Spearman's rank correlation) for most of the intragenic miRNAs. The functional relationships between CLCN5 and its intronic miRNAs have not been explored yet, though, and deserve further exploration. Suggesting a conserved regulation, a similar pattern is found for Clcn5 gene in mouse, which has five annotated intragenic miRNAs and also a high expression correlation between miRNAs and the host gene.

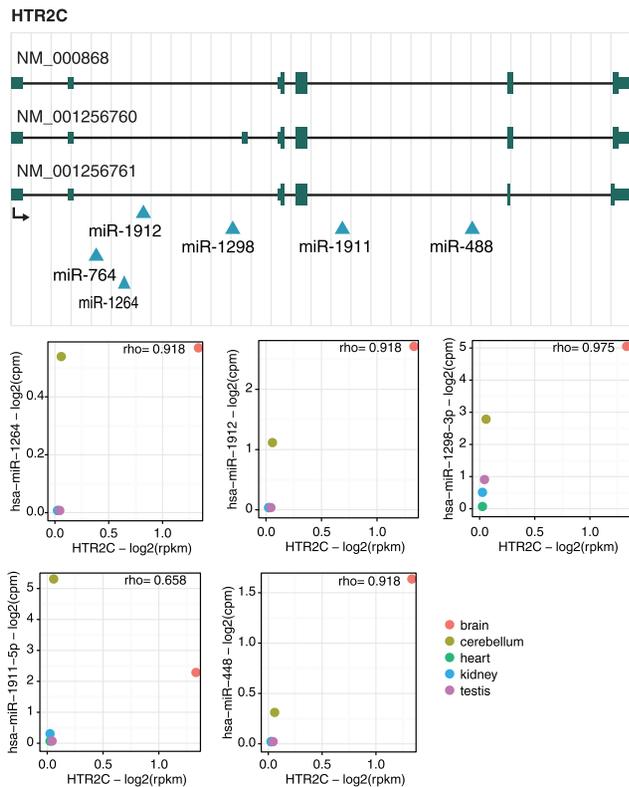
The second gene, HTR2C, encodes the 2C subtype of serotonin receptor and contains six intronic miRNAs (Figure 4). Similar to CLCN5, host and miRNAs have the same transcriptional orientation (see miRIAD web page for details). As reported by (10), up-regulation of HTR2C is involved in adipocyte differentiation by repressing the KLF5 gene through the expression of miR-448, a miRNA located in the fourth intron of HTR2C. Interestingly, our expression data show a highly positive

( $\rho > 8.5$ , Spearman's rank correlation) correlation between miRNAs and host gene, being expressed specifically in cerebellum and brain (Figure 4). The patterns of co-expression are also conserved in opossum and mouse. Moreover, HTR2C is tightly involved in important neuropsychiatric disorders (31); thus, the functional consequences of the concomitant expression of HTR2C and its intragenic miRNAs is tempting to investigate.

miRIAD helped us to identify two interesting gene loci involved in complex human diseases with this quick and unpretentious gene survey. We speculate that many other crucial host/miRNA regulatory mechanisms could be revealed by taking advantage of using miRIAD for initial and/or advanced exploration.

**Discussion and conclusion**

As the number of newly discovered miRNAs is constantly increasing, our understanding of the importance and the



**Figure 4.** HTR2C gene locus. Genomic mapping of HTR2C transcripts (NM\_000868, NM\_001256760 and NM\_001256761) and their six intragenic miRNAs (miR-1912, miR-764, miR-1264, miR-1298, miR-1911 and miR-488) as well as the expression correlation between HTR2C and these miRNAs. The diagram represents the gene structure according to UCSC genome browser. Because the expression of miR-764 could not be detected, expression correlation for this miRNA and its host gene is not shown.

frequency of intragenic miRNAs has also been expanding (5, 10, 13, 15). For example, the past miRBase release 11 (April 2008) had around 47% of intragenic miRNAs (3), and this proportion increased to 53% in the miRBase 19 (August 2012) and to 57% in the miRBase (20). miRIAD was created to help dealing with the challenges of unraveling the functional relationships between miRNAs and their host genes.

miRIAD data are organized in five layers of information. The first layer contains annotation for protein-coding and miRNA genes, including the official gene name, gene aliases and annotation. The second layer provides genomic information for host and miRNAs. The third layer contains gene expression for miRNAs and coding genes and expression correlation between intragenic miRNAs and their host genes. The fourth layer includes miRNA target prediction information (providing binding sites as well). The fifth layer contains additional information, which extends to protein–protein interaction data for host genes as well as interaction partners that are targeted by host’s intragenic miRNA. Additionally, a set of useful external links to other databases

is given. All these information are organized in a streamlined graphical web tool and full integrated into a MySQL relational database. For users who want to manipulate miRIAD information in a local environment, we provide links to download raw data and python code. Specific information not found in those files can be obtained upon request. Therefore, miRIAD can be used to investigate miRNAs in a very integrative context, with special attention to functional features, such as protein–protein interaction, miRNAs targeting host mRNAs or their partners in a functional network. We believe that our web tool can be used as a starting point for developing and testing new hypotheses related to miRNA gene regulation, for one gene or for large-scale data. Importantly, scripts have been developed and pipelined to deal with forthcoming updates.

miRIAD improvements, updates and further development will be ongoing. For example, we envision including additional species and other useful data, such as expression from unhealthy samples. Information on the last and upcoming updates can be found on the miRIAD website.

In conclusion, miRIAD provides a systematic, integrative, user-friendly, and easy-to-use platform to investigate inter- and intragenic miRNAs, host genes and their relationships for five species, including human and mouse. Users can query for and clearly retrieve miRNA and host gene information. Therefore, we believe that miRIAD can substantially improve the way in which we investigate intragenic miRNA and host genes.

## Supplementary Data

Supplementary data are available at *Database Online*.

## Acknowledgements

We thank all members of the Bioinformatics lab for suggestions.

## Funding

This work and some of the authors (P.A.F.G. and L.O.-M.) were partially supported by grants from National Institutes of Health (grant numbers D43TW007015, UH2HL108785 and U54HL108460). G.S.F. was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, process number 2011/19645-6). Funding for open access charge: National Institutes of Health [D43TW007015].

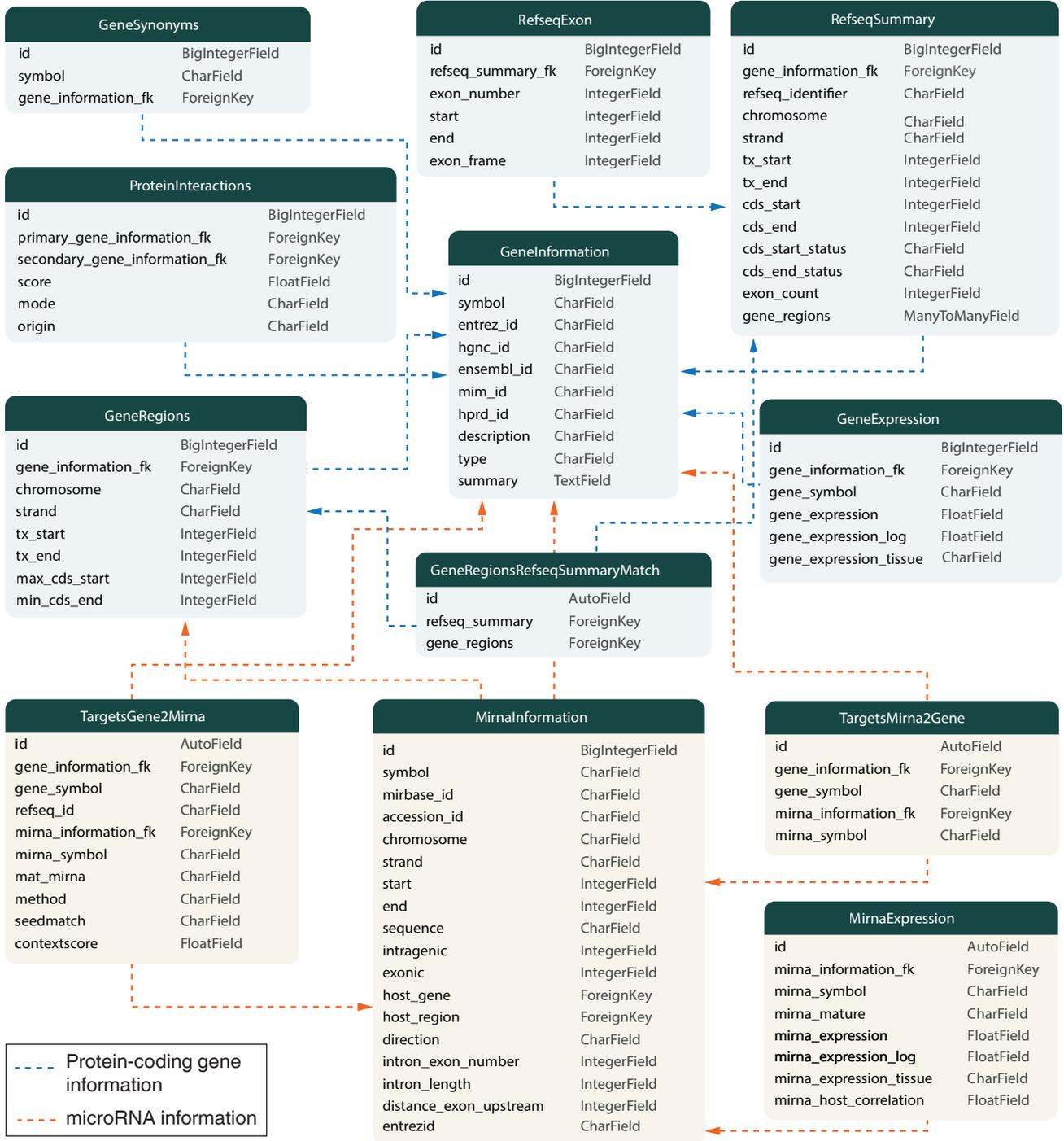
*Conflict of interest.* None declared.

## References

1. Krol,J., Loedige,I. and Filipowicz,W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, **11**, 597–610.
2. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

3. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
4. Rodriguez,A., Griffiths-Jones,S., Ashurst,J.L. *et al.* (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, **14**, 1902–1910.
5. Hinske,L.C.G., Galante,P.A.F., Kuo,W.P. *et al.* (2010) A potential role for intragenic miRNAs on their hosts' interactome. *BMC Genomics*, **11**, 533.
6. Hinske,L.C., Heyn,J., Galante,P.A.F. *et al.* (2013) Setting up an intronic miRNA database. *Methods Mol. Biol.*, **936**, 69–76.
7. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
8. Monteys,A.M., Spengler,R.M., Wan,J. *et al.* (2010) Structure and activity of putative intronic miRNA promoters. *RNA*, **16**, 495–505.
9. Yan,L., Hao,H., Elton,T.S. *et al.* (2011) Intronic microRNA suppresses endothelial nitric oxide synthase expression and endothelial cell proliferation via inhibition of STAT3 signaling. *Mol. Cell. Biochem.*, **357**, 9–19.
10. Kinoshita,M., Ono,K., Horie,T. *et al.* (2010) Regulation of adipocyte differentiation by activation of serotonin (5-HT) receptors 5-HT2AR and 5-HT2CR and involvement of microRNA-448-mediated repression of KLF5. *Mol. Endocrinol.*, **24**, 1978–1987.
11. Cho,S., Jang,I., Jun,Y. *et al.* (2013) MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res.*, **41**, D252–D257.
12. Meyer,L.R., Zweig,A.S., Hinrichs,A.S. *et al.* (2013) The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.*, **41**, D64–D69.
13. Godnic,I., Zorc,M., Jevsinek Skok,D. *et al.* (2013) Genome-wide and species-wide in silico screening for intragenic MicroRNAs in human, mouse and chicken. *PLoS One*, **8**, e65165.
14. Maselli,V., Di Bernardo,D. and Banfi,S. (2008) CoGemIR: a comparative genomics microRNA database. *BMC Genomics*, **9**, 457.
15. He,C., Li,Z., Chen,P. *et al.* (2012) Young intragenic miRNAs are less coexpressed with host genes than old ones: implications of miRNA-host gene coevolution. *Nucleic Acids Res.*, **40**, 4002–4012.
16. Brawand,D., Soumillon,M., Necsulea,A. *et al.* (2011) The evolution of gene expression levels in mammalian organs. *Nature*, **478**, 343–348.
17. Meunier,J., Lemoine,F., Soumillon,M. *et al.* (2013) Birth and expression evolution of mammalian microRNA genes. *Genome Res.*, **23**, 34–45.
18. Kim,D., Pertea,G., Trapnell,C. *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
19. Trapnell,C. and Salzberg,S.L. (2009) How to map billions of short reads onto genomes. *Nat. Biotechnol.*, **27**, 455–457.
20. Trapnell,C., Williams,B.A., Pertea,G. *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
21. Marco,A. and Griffiths-Jones,S. (2012) Detection of microRNAs in color space. *Bioinformatics*, **28**, 318–323.
22. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
23. Dill,H., Linder,B., Fehr,A. *et al.* (2012) Intronic miR-26b controls neuronal differentiation by repressing its host transcript, *ctdsp2*. *Genes Dev.*, **26**, 25–30.
24. Liu,M., Roth,A., Yu,M. *et al.* (2013) The IGF2 intronic miR-483 selectively enhances transcription from IGF2 fetal promoters and enhances tumorigenesis. *Genes Dev.*, **27**, 2543–2548.
25. Zhu,Y., Lu,Y., Zhang,Q. *et al.* (2012) MicroRNA-26a/b and their host genes cooperate to inhibit the G1/S transition by activating the pRb protein. *Nucleic Acids Res.*, **40**, 4615–4625.
26. Radfar,M.H., Wong,W. and Morris,Q. (2011) Computational prediction of intronic microRNA targets using host gene expression reveals novel regulatory mechanisms. *PLoS One*, **6**, e19312.
27. Pollak,M. (2008) Insulin and insulin-like growth factor signaling in neoplasia. *Nat. Rev. Cancer*, **8**, 915–928.
28. da Cunha,J.P.C., Galante,P.A.F., de Souza,J.E. *et al.* (2009) Bioinformatics construction of the human cell surfaceome. *Proc. Natl Acad. Sci. USA*, **106**, 16752–16757.
29. de Souza,J.E.S., Galante,P.A.F., de Almeida,R.V.B. *et al.* (2012) SurfaceomeDB: a cancer-orientated database for genes encoding cell surface proteins. *Cancer Immun.*, **12**, 15.
30. Gorvin,C.M., Wilmer,M.J., Piret,S.E. *et al.* (2013) Receptor-mediated endocytosis and endosomal acidification is impaired in proximal tubule epithelial cells of Dent disease patients. *Proc. Natl Acad. Sci. USA*, **110**, 7014–7019.
31. Mickey,B.J., Sanford,B.J., Love,T.M. *et al.* (2012) Striatal dopamine release and genetic variation of the serotonin 2C receptor in humans. *J. Neurosci.*, **32**, 9344–9350.

# Database schema



**Figure S1.** The tables, fields and relationships shown above were implemented for all species in miRIAD. For the sake of simplicity, the database schema is exemplified for human species, which comprises 12 tables. Considering all species, miRIAD is composed by 60 tables.



Human



Rhesus



Mouse



Opossum



Chicken

## Current species

**Anexo C – Resumo do artigo submetido:** *Host gene constraints and genomic context impact the expression and evolution of human microRNAs*

# **Host gene constraints and genomic context impact the expression and evolution of human microRNAs**

Gustavo S. França<sup>1,2</sup>, Maria D. Vibranovski<sup>3,\*</sup> & Pedro A. F. Galante<sup>1,\*</sup>

<sup>1</sup>Centro de Oncologia Molecular, Hospital Sírio-Libanês, 01308-060, São Paulo, SP, Brazil.

<sup>2</sup>Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, 05508-900, São Paulo, SP, Brazil.

<sup>3</sup>Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, 05608-900, São Paulo, SP, Brazil.

## **\*Corresponding authors**

Maria D. Vibranovski (mdv@ib.usp.br)

Department of Genetics and Evolutionary Biology

University of Sao Paulo

Rua do Matao 277, Sao Paulo, SP; Brazil - 05508-090

phone: +55 11 3091-0952

Pedro A. F. Galante (pgalante@mochsl.org.br)

Hospital Sírio-Libanês

Rua Daher Cutait, 69. Sao Paulo, SP; Brazil

Phone: +55 11 3155-3704

## **Abstract**

Increasing evidence have shown that recent miRNAs tend to emerge within coding genes. Here, we conjectured that human miRNA evolution is tightly influenced by the genomic context, especially by host genes. Our findings show a preferential emergence of intragenic miRNAs within old genes. We found that miRNAs within old host genes are significantly more broadly expressed than those within young ones. Surprisingly, young miRNAs within old genes are more broadly expressed than their intergenic counterparts, suggesting that young miRNAs had an initial advantage by residing in old genes, benefited by their hosts' expression control and through an exposure to diverse cellular contexts and target genes. Our results show that host genes may provide stronger expression constraints to intragenic miRNAs in the long run. We also show associated functional implications, highlighting the genomic context and host genes as driving factors for the expression and evolution of human miRNAs.