

RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo deste trabalho será disponibilizado somente a partir de 21/02/2019.

Avaliação meta-classificatória de ferramentas de predição de alvos de
microRNAs e análise de enriquecimento funcional de alvos utilizando
Homo sapiens como modelo biológico

Arthur Casulli de Oliveira

Botucatu, SP

2017

UNIVERSIDADE ESTADUAL PAULISTA

“Julio de Mesquita Filho”

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

Avaliação meta-classificatória de ferramentas de predição de alvos de microRNAs e análise de enriquecimento funcional de alvos utilizando *Homo sapiens* como modelo biológico

Candidato: Arthur Casulli de Oliveira

Orientador: Danilo Pinhal

Dissertação apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção do título de Mestre pelo Programa de Pós-Graduação em Ciências Biológicas (Genética).

Botucatu, SP

2017

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Oliveira, Arthur Casulli de.

Avaliação meta-classificatória de ferramentas de predição de alvos de microRNAs e análise de enriquecimento funcional de alvos utilizando Homo sapiens como modelo / Arthur Casulli de Oliveira. - Botucatu, 2017

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Danillo Pinhal

Capes: 20205007

1. MicroRNAs. 2. Genética humana. 3. Bioinformática.
4. Regulação da expressão gênica.

Palavras-chave: Bioinformática; Genética humana; RNAs não-codificadores; Regulação gênica.

Dedico este trabalho aos meus pais e avós, que me forneceram todo carinho e apoio necessário no decorrer do meu mestrado e à minha namorada, que sempre esteve ao meu lado me apoiando durante todos estes anos.

Agradeço:

Ao meu orientador, Prof. Dr. Danilo Pinhal, pelo suporte, orientação científica e amizade durante todos estes anos em que fui seu aluno e por sempre me incentivar a encarar novos desafios nesta jornada acadêmica.

À minha mãe, Fernanda Bressanelli Casulli, por todo amor e carinho e por sempre estar ao meu lado me apoiando e educando, nos momentos felizes e tristes da minha vida. Sem ela, não estaria perto de onde estou agora.

Aos meus avós, Ida Maria e Pilade, por todo amor, carinho e suporte que sempre me deram durante todos os anos da minha vida e, principalmente, durante anos de graduação e pós-graduação.

Ao meu irmão, Victor Casulli de Oliveira, por todos estes anos de amizade, brincadeiras e brigas e por sempre estar ao meu lado, nas risadas e nas tristezas.

Ao amor da minha vida, minha namorada Karina Gabriele Alves Dias (Nuros), por compartilhar comigo todos os momentos, bons e ruins, dos últimos quatro anos, pelos momentos divertidos que passamos juntos: jantares, parques, sessões de *Netflix*, cinema, etc e por tudo que o futuro ainda nos reserva. Te amo.

A todos os colegas do Laboratório Genômica e Evolução Molecular pelas conversas e discussões sobre os mais variados temas relacionados a pesquisa e ciência.

Aos amigos de trabalho Pedro (Batata), Marcos, e Luiz (Chokito) por todas as brincadeiras, risadas e colaborações realizadas durante todo o período do meu mestrado e por sempre me auxiliarem durante o desenvolvimento deste projeto.

Ao Prof. Dr. Ney Lemke, por todo suporte fornecido na área da computação e bioinformática, essenciais durante a realização deste projeto.

Ao Dr. Simon Moxon, pelo estágio realizado em Norwich, Inglaterra e por todo suporte e conhecimento transferido durante minha estadia no TGAC.

Aos meus grandes amigos e parceiros de vida, Julinho, Dox e Marsal, por todos os anos de grande amizade, jogatina e conversas e por estarem sempre comigo nos principais momentos da minha vida

Aos funcionários da seção de Pós-graduação por todas as dúvidas esclarecidas.

Ao Laboratório de Genômica e Evolução Animal, ao Departamento de Genética, à Pós-graduação em Genética, ao Instituto de Biociências de Botucatu e à Universidade Estadual Paulista pela estrutura cedida para a realização deste trabalho.

À CAPES pela bolsa de mestrado concedida no período de estudo.

A todas as pessoas que de alguma forma, direta ou indiretamente, auxiliaram na realização e finalização deste trabalho.

*Se, a princípio, a ideia não é absurda,
então não há esperança para ela.*

- Albert Einstein

Resumo

MicroRNAs (miRNAs) são pequenos RNAs não codificadores que regulam uma ampla gama de vias biológicas. Esta regulação ocorre através do pareamento complementar entre o miRNA e seu RNA mensageiro (mRNA) alvo, geralmente na região 3'UTR, inibindo a síntese proteica. Diversos trabalhos têm buscado determinar as funções biológicas desempenhadas pelos miRNAs por meio da identificação de seus alvos e posterior análise de enriquecimento funcional. Entretanto, as ferramentas de predição de alvos *in silico* disponíveis atualmente apresentam resultados pouco robustos e não há um consenso sobre a melhor ferramenta e estratégia para análise dos dados. Adicionalmente, a metodologia de enriquecimento funcional atual não leva em conta diversos fatores fundamentais atuantes na regulação dos alvos dos miRNAs, retornando resultados inconsistentes que culminam em experimentos de validação desnecessários e pouco específicos, com conseqüente desperdício de tempo e recursos. Desta maneira, o presente trabalho tem como objetivos (i) elaborar metodologia de predição de alvos com alta eficiência utilizando as ferramentas de bioinformática disponíveis e (ii) avaliar a regulação dos processos biológicos controlados pelos miRNAs através da análise de enriquecimento funcional, considerando o *fold-change* de seus mRNA alvo. Para tal, comparou-se as performances das três ferramentas de predição de alvos atualmente mais utilizadas (TargetScan, miRanda-mirSVR, e Pita), assim como testou-se todas as possibilidades de combinação dos dados gerados por cada ferramenta (uniões e/ou intersecções). A metodologia de união das ferramentas TargetScan + miRanda-mirSVR resultou na melhor performance, com o melhor balanço entre sensibilidade e especificidade. Posteriormente, dados de expressão de genes alvos obtidos por *microarray* após a superexpressão de onze miRNAs foram utilizados para as análises de enriquecimento funcional. Os alvos dos miRNAs foram agrupados manualmente em cinco *clusters* de acordo com seu *fold-change*. Os clusters foram então submetidos à análise de enriquecimento funcional. Os processos biológicos enriquecidos por esta análise foram distintos em cada cluster, sugerindo que os miRNAs regulam com intensidade semelhante genes associados a uma mesma função biológica, mas funções biológicas distintas são reguladas com intensidades diferentes. Os resultados obtidos neste projeto aprimoram significativamente a qualidade das análises *in silico* de predição de alvos, o que permitirá aos pesquisadores obterem resultados mais robustos durante a identificação de alvos dos miRNAs. Adicionalmente, a análise de enriquecimento funcional realizada sugere uma nova complexidade dos miRNAs, podendo justificar o fato de um único miRNA ser capaz de regular processos biológicos distintos com a especificidade demandada para cada processo dentro de um contexto celular.

Abstract

MicroRNAs (miRNAs) are short non-coding RNAs that regulates a wide range biological pathways. This regulation occurs by the complementary binding between miRNA and its target Messenger RNA (mRNA), mainly at 3'UTR region, blocking the protein synthesis. Several works try to identify the biological functions that miRNAs are assigned by detecting its mRNA targets and performing functional enrichment analysis using bioinformatic tools. However, *in silico* target prediction tools available nowadays often return little robust results and there is no consensus about a tool that highlights from the others or if combining the results from more than one tool improves the quality of the analysis. Moreover, the functional enrichment methodology used nowadays do not take in account several important aspects of the regulation of the miRNA targets, thus generates inconsistent results. This way, the objectives of this project are (i) to elaborate a target prediction method with high efficiency using the available tools and (ii) to evaluate the regulation of the biological process controlled by the miRNAs by functional analysis considering the fold-change levels of the target mRNAs. To do this, we compared the performances of the three most used target prediction tools (TargetScan, miRanda-mirSVR and Pita), as well as all combinatorial possibilities of these tools (unions and intersections). The union of TargetScan + miRanda-mirSVR returned the greatest performance, with the best balance between sensitivity e specificity. After, microarray data from gene expression after super-expression of eleven miRNAs were used for the functional enrichment analysis. The miRNA targets were grouped in five clusters according to their fold-change levels after the superexpression of the miRNAs. The clusters were individually submitted to functional enrichment analysis. The enriched biological processes were distinct in each cluster, suggesting that miRNAs control genes assigned with one function with similar intensity, but distinct biological processes are controlled with distinct intensities. The results obtained in this project improved the quality of *in silico* target prediction analysis, which can help researchers obtaining results with more quality when performing miRNA target prediction. Moreover, the functional enrichment analysis suggests a new complexity of miRNAs, and could justify the fact of a unique miRNA be capable of control several biological processes with the specificity required for each one within the cellular context.

Sumário

1. Introdução geral.....	11
1.1. A predição de alvos e seus principais atributos.....	14
1.1.1. Ferramentas de predição de alvos.....	16
1.2. O enriquecimento funcional dos alvos de miRNAs	18
1.2.1. Os problemas do enriquecimento funcional	19
2. Objetivos.....	21
3. Capítulo I: Análise de Predição de Alvos	22
3.1. Material e métodos	22
3.1.1. Resumo do workflow.....	22
3.1.2. Obtenção dos dados de predição de alvos	22
3.1.3. União e intersecção dos resultados das ferramentas.....	24
3.1.4. Cálculo da sensibilidade, especificidade, precisão e performance	24
3.1.5. Análise estatística	27
3.2. Resultados e discussão – Artigo Científico	28
4. Capítulo II: Análise de enriquecimento funcional.....	49
4.1. Material e Métodos.....	50
4.1.1. Resumo do Workflow.....	50
4.1.2. Obtenção dos dados de microarray.....	50
4.1.3. Agrupamento dos alvos em clusters de mRNA fold-change	50
4.1.4. Análise de enriquecimento funcional	51
4.1.5. Análise de conservação evolutiva.....	51
4.2. Resultados e discussão – Artigo Científico	53
5. Considerações finais	69
6. Referências bibliográficas	71

1. Introdução geral

MicroRNAs (miRNAs) são pequenos RNAs não-codificadores (~22 nucleotídeos) presentes no genoma de animais, plantas e, inclusive, vírus (Lee et al., 1993; Sunkar et al., 2005; Jia et al., 2008). Descobertos pioneiramente há mais de duas décadas em *Caenorhabditis elegans* (Lee et al., 1993), possuem atualmente a reconhecida importância de participar da regulação de uma vasta gama de processos biológicos, tais como diferenciação e proliferação celular, carcinogênese, resposta imune, morte celular, dentre outros (Ambros, 2004; Flynt et al., 2007, 2009; Liu e Olson, 2010; Shkumatava et al., 2009; Christodoulou et al., 2010; Takacs e Giraldez, 2011).

A via canônica da biogênese de um miRNA (Figura 1) tem início com a transcrição de uma longa molécula (~110 nucleotídeos) conhecida como miRNA primário (pri-miRNA) (Borchert et al., 2006). Essa molécula dobra-se em uma estrutura secundária em forma de grampo de cabelo (estrutura em *hairpin*), passando então a ser reconhecida pela enzima *Drosha*. A *Drosha* cliva o pri-miRNA na região caudal do *hairpin*, formando o miRNA precursor (pré-miRNA; ~70 nucleotídeos; Lee et al., 2006). O pré-miRNA é, então, exportado para o citoplasma através da proteína *exportina-5* (Lund et al., 2004) e processado pela enzima *Dicer*. Essa enzima cliva o pré-miRNA na região *loop*, formando uma molécula de RNA fita dupla (dsRNA) de aproximadamente 22 pares de bases. Proteínas da família argonauta associam-se com uma das fitas do dsRNA para formar o complexo de silenciamento induzido por RNA (RISC), dando origem ao miRNA maduro, ou canônico (Krutzfeldt et al., 2006), enquanto que a outra fita (miRNA*) pode ser degradada ou também se associar a um complexo RISC (Rand et al., 2005; Bang et al., 2014).

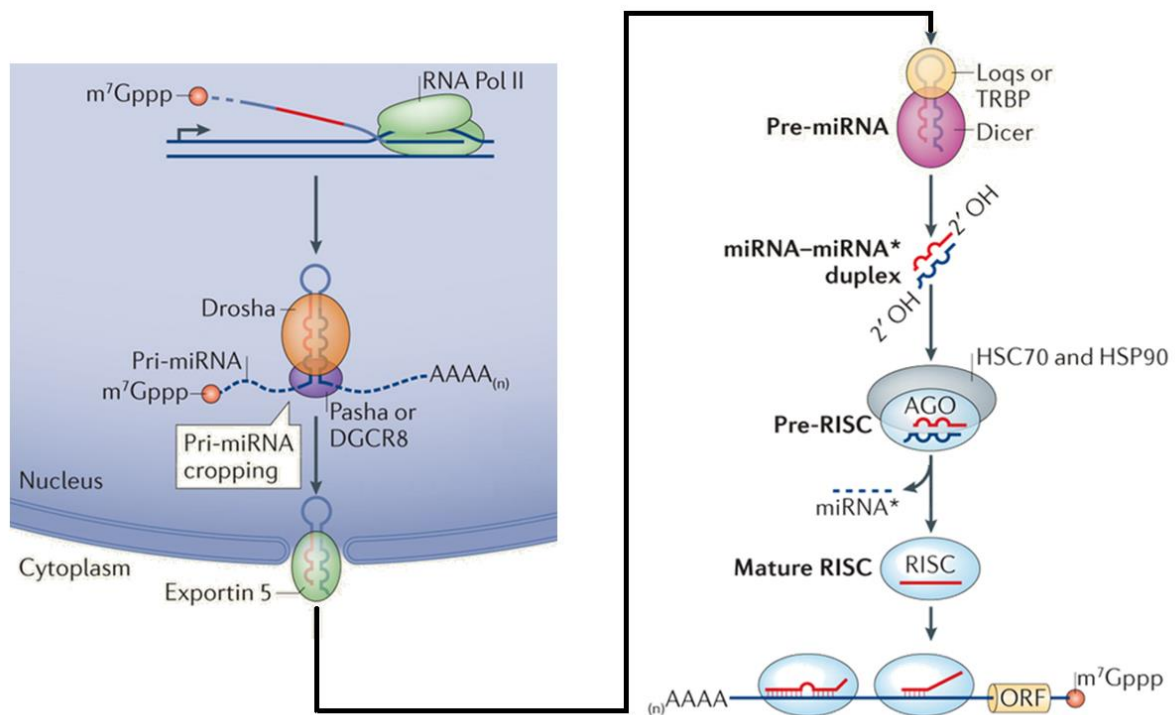


Figura 1: Via canônica da biogênese de miRNAs. Figura editada de Ameres e Zamore (2013).

Funcionalmente, os miRNAs atuam pelo pareamento simples de "Watson e Crick" com a sequência complementar presente na molécula do RNA mensageiro (mRNA) alvo. O miRNA maduro interage preferencialmente com a região 3'UTR do RNA mensageiro (mRNA) alvo por complementaridade total (nas plantas) ou parcial (nos animais) levando à inibição de sua síntese proteica (Lee e Dutta, 2009). Entretanto, adicionalmente à região 3'UTR, diversos trabalhos também detectaram sítios de interação em éxons (Tay et al., 2008; Reckzo et al., 2012; Schnall-Levin et al., 2010; Hausser et al., 2013) e na região 5'UTR (Lytle et al., 2007; Orom et al., 2008; Devlin et al., 2010; Zhou e Rigoutsos, 2014) da molécula de mRNA. Nas plantas, a alta complementaridade dos miRNAs tende a desencadear a clivagem do mRNA pelas proteínas da família argonauta (Tang et al., 2003; Lanet et al., 2009), enquanto que, a baixa complementaridade dos miRNAs com seu mRNA alvo, nos animais, geralmente não permite a clivagem do mRNA, sugerindo que neste grupo, a regulação da expressão gênica seja feita de maneira alternativa à clivagem (Karginov et al., 2010; Shin et al., 2010), embora haja exceções em ambos os grupos. Estudos realizados no *Danio rerio* (zebrafish) e em *Drosophila melanogaster* (drosófila) apontam que nos animais, os miRNAs

tendem a primeiramente inibir a tradução impedindo o acoplamento do ribossomo ao mRNA e posteriormente levar à degradação prematura deste mRNA (Bazzini et al., 2012; Mathonnet et al., 2007; Zdanowicz et al., 2009).

A interação de um miRNA com seu alvo é guiada principalmente por uma sequência de 7 nucleotídeos na região 5' do miRNA (nucleotídeos 2 a 8), chamada de sequência *seed*, embora importância cada vez maior esteja sendo atribuída ao pareamento 3' complementar nesse processo (Broughton et al., 2016). Cinco tipos de pareamentos da sequência *seed* são descritos atualmente: 8mer, 7mer-m8, 7mer-A1, 6mer e *offset*-6mer, apresentados na ordem do mais para o menos efetivo (Agarwal et al., 2015). O pareamento 8mer, se caracteriza pelo pareamento de sete nucleotídeos (2-8) com um "A" na posição "1" do 3'UTR. Estudos mostram que há uma preferência no reconhecimento de MREs que apresentam este nucleotídeo na posição "1" do 3'UTR, devido ao fato de este "A" ser uma região de assentamento das proteínas argonautas (Baek et al., 2008; Schirle et al., 2014). O pareamento 7mer-m8 representa um pareamento de sete nucleotídeos (2-8), porém sem a presença da "A" na posição "1". O pareamento 7mer-A1, é um pareamento de seis nucleotídeos (2-7), que contém um "A" na posição "1" do 3'UTR. O pareamento 6mer representa um pareamento de 6 nucleotídeos (2-7), enquanto que a *seed offset*-6mer caracteriza-se por um pareamento deslocado de 6 nucleotídeos (3-8), ambos sem a presença do "A" na posição "1". Entretanto, estes dois últimos tipos de pareamento apresentam baixa eficiência de regulação e são pouco conservados (Agarwal et al., 2015). Adicionalmente a estas interações baseadas na sequência *seed*, diversas outras interações não baseadas na *seed* foram detectadas (Clark et al., 2012; Clark et al., 2014; Chi et al., 2012), assim como foi demonstrada que a porção 3' dos miRNAs pode ser tão relevante quanto a região *seed* na detecção de seus alvos (Broughton et al., 2016).

Análises computacionais de predição de alvos de miRNAs indicam que um único miRNA pode ligar-se a centenas de mRNAs. Assim, cada miRNA regula uma gama extensa de processos biológicos distintos. Portanto, os miRNAs devem ser capazes de regular de forma específica cada processo biológico.

Desta maneira, a predição de genes alvo e a caracterização dos processos biológicos regulados são etapas fundamentais em diversas pesquisas envolvendo miRNAs e seus papéis biológicos. A predição de alvos de miRNAs atualmente é realizada através de ferramentas computacionais que avaliam diversos parâmetros envolvendo a interação miRNA-alvo. Já a caracterização dos processos biológicos regulados pelos miRNAs geralmente é feita através

de análises de enriquecimento funcional dos alvos preditos computacionalmente ou dos genes cuja expressão foi alterada através de técnicas de manipulação do miRNA estudado.

Entretanto, a predição computacional de alvos ainda está distante do ideal, pois as ferramentas exibem uma alta quantidade de interações falso-positivas ou são falhas em detectar interações genuínas. Adicionalmente, as análises de enriquecimento funcional dos alvos de miRNAs são realizadas sem que sejam considerados diversos fatores relevantes para que haja a interação miRNA-alvo (ex., pareamento da *seed*, acessibilidade do sítio de ligação, tamanho do 3'UTR). Desta maneira, diversas funções biológicas desempenhadas pelos miRNAs podem estar sendo equivocadamente propostas ou descartadas a priori.

1.1. A predição de alvos e seus principais atributos

Nos últimos anos, diversas propriedades importantes para o reconhecimento de um mRNA como alvo de um miRNA foram identificadas em animais, melhorando a compreensão destas interações e, conseqüentemente, aprimorando a predição computacional de alvos. Dentre todos os parâmetros descritos atualmente, alguns recebem maior destaque devido sua grande influência tanto na regulação quanto no reconhecimento de um mRNA como pleno alvo de um miRNA (Figura 2).

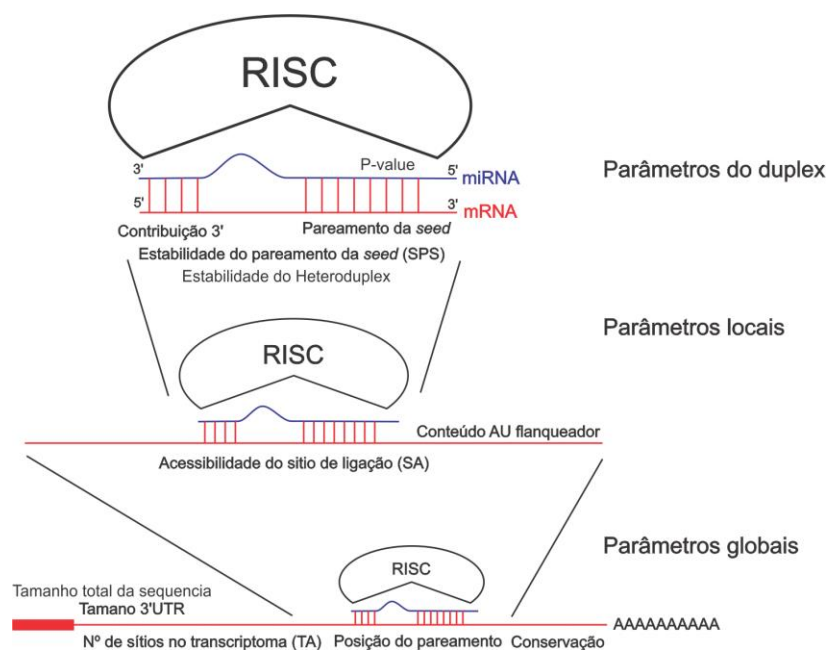


Figura 2: Principais parâmetros utilizados durante a predição de alvos. Figura editada de Betel et al. (2010).

Estes parâmetros podem ser divididas em três grandes grupos: os parâmetros do duplex, os parâmetros locais e os parâmetros globais (Betel et al., 2010). Os parâmetros do duplex contêm os parâmetros pareamento da *seed*, contribuição 3', estabilidade do pareamento da *seed* (*seed pairing stability* – SPS; Betel et al., 2010,), energia livre do heteroduplex e P-value (Miranda et al., 2010). Estes parâmetros aferem a interação direta entre o miRNA e seus alvos. O pareamento da *seed* avalia quantos nucleotídeos da região *seed* do miRNA estão pareado com o mRNA alvo. A contribuição 3' avalia se além do pareamento da *seed* ocorre também um pareamento da porção 3' do miRNA e o quão ele auxilia na regulação (Witkos et al., 2011). O SPS avalia quais os nucleotídeos que compõe a sequencia *seed* (Garcia et al., 2008). A energia livre do heteroduplex avalia se a mínima energia livre formada entre o miRNA e o mRNA alvo é suficiente para estabelecer uma hibridização. Finalmente, o P-value avalia a probabilidade da interação miRNA-alvo ter sido predita de maneira aleatória.

Os parâmetros locais aferem propriedades da 3'UTR do mRNA com influência direta no reconhecimento deste como um alvo de um miRNA qualquer. Estes parâmetros perfazem a acessibilidade do sítio de ligação (*site accessibility* – SA) e o conteúdo AU flanqueador. A SA avalia a capacidade do miRNA em desdobrar a potencial estrutura secundária formada na região de interação do miRNA (Kertsz et al., 2007), conhecida como elemento de reconhecimento de miRNAs (*miRNA recognition element* – MRE). O conteúdo AU flanqueador avalia o número de nucleotídeos “A” e “U” que flanqueiam os MREs, uma vez que altas concentrações destes nucleotídeos nestas regiões aumentam a eficiência da regulação (Grimsom et al., 2007).

Os parâmetros globais aferem propriedades da 3'UTR do mRNA alvo com influência indireta no reconhecimento deste como um alvo. Estes parâmetros são o tamanho total da sequencia e do 3'UTR, o número de sítios de ligação no transcriptoma (*transcriptome abundance* – TA), a posição do pareamento e o grau de conservação do 3'UTR. O tamanho do 3'UTR é importante de ser avaliado, uma vez que 3'UTRs mais longos sofrem maior grau de regulação (Sandberg et al., 2008), já o tamanho total da sequencia é relevante pois há maior probabilidade de predições falsas em sequencias maiores (Miranda et al., 2006). O TA avalia a quantidade de sítios de ligação de um mesmo miRNA em todo o transcriptoma, uma vez que quanto mais alvos esse miRNA regular, maior vai ser a diluição de seu efeito. A posição

do pareamento avalia a posição do MRE no 3'UTR, uma vez que MREs localizados nas porções terminais do 3'UTR apresentam um maior potencial regulatório (Grimsom et al., 2007). A conservação avalia a conservação dos MREs entre as espécies, uma vez que miRNAs mais conservados tendem a apresentar maior potencial regulatório (Grimsom et al., 2007).

1.1.1. Ferramentas de predição de alvos

A partir dos avanços da bioinformática nos últimos anos, diversas ferramentas de predição foram elaboradas na tentativa de se otimizar a busca por interações de genes alvos relacionados às vias regulatórias diversas nas quais os miRNAs estejam atuando. Atualmente, dezenas de ferramentas encontram-se disponíveis. Dentre elas, quatro têm sido amplamente utilizadas pela comunidade científica: TargetScan, miRanda-mirSVR, Pita e RNA22. As ferramentas Targetscan, miRanda-mirSVR e Pita consideram predições baseadas na sequência *seed* e nas regiões 3'UTR, enquanto a ferramenta RNA22 também considera interações não baseadas na *seed* e em todo o transcrito.

Apesar de todas estas ferramentas terem como objetivo a identificação de uma ampla gama de interações miRNA-alvo genuínas, elas são constituídas de parâmetros distintos (Tabela 1), proporcionando resultados divergentes entre elas.

Contudo, mesmo essas ferramentas mais avançadas, ainda geram predições de alvo pouco robustas, pois retornam (i) uma alta quantidade de interações falso-positivas ou (ii) são falhas em detectar interações genuínas. Adicionalmente, como demonstrado, essas ferramentas utilizam uma série de parâmetros divergentes, o que produz resultados inconsistentes quando comparadas entre si. Este fato acentua-se principalmente nas predições de alvos de miRNAs de animais, devido à possibilidade de pareamento incompleto entre o miRNA e seu alvo.

Desta maneira, apesar da disponibilidade de uma série de ferramentas de predição de alvos não há um consenso sobre a melhor maneira de utilizá-las. De fato, diversos experimentos de validação revelaram muitos resultados falso positivos e falso negativos, demonstrando que ainda há necessidade de futuras melhoras nas ferramentas. Na tentativa de minimizar estes resultados pesquisadores tem usado diversas estratégias para selecionar os alvos preditos, incluindo a utilização da intersecção ou união dos resultados de diversas

ferramentas. Entretanto, estas metodologias vêm sendo usadas de maneira indiscriminada, sem um critério bem definido e um teste comparativo para determinar a qualidade de tais estratégias. Desta maneira, ainda não se sabe a intersecção ou união dos resultados de mais de uma ferramenta de fato melhora a qualidade das análises de predição de alvo.

Tabela 1: Principais parâmetros utilizados pelas ferramentas TargetScan, miRanda-mirSVR, PiTa e RNA22.

Grupos	Atributos	TargetScan	miRanda-miRSVR	PiTa	RNA22
Parâmetros do duplex	Pareamento <i>seed</i>	X	X	X	X
	Contribuição 3' SPS	X	X	X	X
	Energia livro do heteroduplex				X
	P-value				X
	SA	X	X	X	
Parâmetros Locais	Conteúdo AU flanqueador	X	X	X	
Parâmetros Globais	TA	X			
	Posição do pareamento	X	X		
	Tamanho do 3' UTR	X	X		
	Tamanho total da sequencia				X
	Conservação	X	X	X	
-	Outros	X			X

1.2. O enriquecimento funcional dos alvos de miRNAs

Após a obtenção dos resultados provenientes da predição de alvos dos miRNAs, uma das principais análises realizadas com estes dados visa à busca e identificação dos papéis e

vias biológicas que cada miRNA participa (Bleazard et al., 2015). O método mais utilizado nesta análise é o enriquecimento funcional *in silico* dos alvos regulados pelos miRNAs. Este método consiste em três etapas: (i) identificar os genes regulados pelos miRNAs analisados, (ii) associar estes alvos com suas funções biológicas e (iii) calcular a super-representação estatística dos processos biológicos dos alvos dos miRNAs (Gusev et al., 2007).

A primeira etapa é geralmente realizada com a utilização de ferramentas de predição de alvo ou de *datasets* experimentais que avaliaram interação mRNA-alvo. Uma vez que tais experimentos em larga escala da interação mRNA-alvo, como por exemplo *chips* de *microarray* e *CLIP-seqs*, ainda são escassos a abordagem de predição de alvos permanece como a mais utilizada. Durante esta etapa os pesquisadores podem optar por analisar os alvos provenientes de um único miRNA ou de uma lista de miRNAs que possuem uma característica em comum, como por exemplo serem enriquecidos em determinado tecido ou estarem sub/super-expressos em pacientes com determinada doença.

A segunda etapa é geralmente realizada utilizando-se as anotações do *Gene Ontology* (GO; Ashburner et al., 2000), ou as vias biológicas do *Kyoto Encyclopedia of Genes and Genome* (KEGG; Kanehisa and Goto, 2000). O GO é um consórcio que agrega termos funcionais dos genes de diversas espécies de animais, plantas e microorganismos, dividindo-os em três categorias: processos biológicos, componentes celulares e funções moleculares. Os termos associados aos processos biológicos se referem às vias nas quais o gene contribui, como por exemplo *generation of neurons* (GO:0048699) e *response to stress* (GO:0006950). Os termos associados aos componentes celulares referem-se às partes da célula ou ambiente extracelular em que os genes atuam, como por exemplo *cytoplasmic vesicle* (GO:0031410) e *synapse part* (GO:0044456). Os termos associados às funções moleculares referem-se às atividades bioquímicas do gene, como por exemplo *hydrolase activity* (GO:0016787) e *protein complex binding* (GO:0032403) (Ashburner et al., 2000).

O KEGG, assim como o GO, é um banco de dados que agrega termos funcionais de diversas espécies de animais, plantas e microorganismos. Entretanto, diferentemente do GO, o KEGG fornece um mapa de vias biológicas, agrupando os genes segundo as grandes vias biológicas das quais participam, por exemplo, *fatty acid metabolism* e *lipid metabolism* (Kanehisa and Goto, 2000), ao invés de agrupá-los por processos biológicos relacionados a eventos específicos.

A terceira etapa consiste no teste de distribuição hipergeométrica, ou o teste de Fisher, utilizado para o enriquecimento dos dados. Neste contexto, a distribuição hipergeométrica calcula a probabilidade de um miRNA regular n genes num determinado processo biológico dado um total de x genes presentes na amostra. Por essa estratégia é possível testar se os genes alvos de miRNAs são controlados aleatoriamente ou se estão preferencialmente associados a determinados processos biológicos (Bleazard et al., 2015).

1.2.1. Os problemas do enriquecimento funcional

Apesar de este ser o método mais utilizado para determinação dos processos e vias biológicas controladas pelos miRNAs, ele não leva em consideração diversos fatores relevantes durante a interação miRNA-alvo, como os demonstrados na Tabela 1.

Bleazard et al. (2015) questiona a eficiência do modo pelo qual o enriquecimento funcional dos alvos de miRNAs é atualmente realizado, alertando que esta abordagem possui diversos problemas e enviesamento metodológico. Estes autores discutem que esta abordagem gera diversos resultados inespecíficos, como por exemplo retorna processos biológicos enriquecidos semelhantes independentemente da lista de alvos utilizada. Isso ocorre, pois, uma vez que os genes podem estar associados a mais de uma função biológica, isto pode tendenciar o aparecimento de determinados processos biológicos gerais, como por exemplo *regulation of biological process* (GO:0050789), *single-organism process* (GO:0044699), *multicellular organismal process* (GO:0032501), dentre outros, em diversas listas independentes.

Na tentativa de atenuar estes problemas, estes e outros pesquisadores têm buscado identificar abordagens alternativas à metodologia padrão de enriquecimento funcional. Bleazard et al. (2015), por exemplo, elaborou um cálculo empírico que também leva em consideração o número de MREs presentes em cada 3'UTR. Entretanto, além desta, diversas outras características influenciam a regulação final dos alvos dos miRNAs. Ignorar tais características ou o potencial regulatório de um miRNA sobre seus alvos como um todo durante a etapa de enriquecimento pode ocultar diversos padrões regulatórios e propriedades biológicas recorrentes dos miRNAs. Em outras palavras, uma vez que um único miRNA é capaz de regular diversos processos biológicos dentro de um mesmo contexto celular, é provável que cada processo seja regulado de modo diferencial e particular e que os genes alvo

atuantes num mesmo processo biológico sejam regulados com intensidades semelhantes, diferentemente de genes atuantes em outros processos biológicos cuja intensidade de regulação não estaria correlacionada (aqui, nos definimos “intensidade de regulação” como o grau de fold-change apresentado pelos mRNA após a super-expressão dos miRNAs através de mimetizadores). Tal categorização regulatória hipotética não pode ser avaliada com o uso das metodologias atuais de enriquecimento funcional, porém caso verdadeira, ajudaria a justificar a capacidade dos miRNAs em regular vias biológicas distintas com alta especificidade, atendendo às demandas próprias de cada via, e assim promovendo a homeostasia celular.

5. Considerações finais

A quantidade de trabalhos envolvendo o estudo do papel biológico dos miRNAs vem crescendo nos últimos anos. Nestes trabalhos, a predição *in silico* de alvos e o respectivo enriquecimento funcional dos mesmos estão quase sempre presentes. Entretanto, muito ainda se discute sobre a real eficiência de tais métodos, uma vez que em diversos casos trazem resultados pouco eluzivos ou de baixa qualidade. Neste contexto, os resultados apresentados neste trabalho poderão auxiliar os pesquisadores na realização destas duas etapas fundamentais durante a caracterização dos papéis biológicos desempenhados pelos miRNAs.

No capítulo I demonstramos que as últimas atualizações das ferramentas de predição de alvos forneceram valores de alta qualidade de especificidade e precisão, gerando predições quase totalmente livres de resultados falso-positivos. Estes dados, mostram que a técnica de intersecção, apesar de ainda ser muito adotada, atualmente não se faz necessária, trazendo prejuízos devido à perda da sensibilidade. Entretanto, os *cutoffs* utilizados pelas ferramentas excluem diversos alvos verdadeiros. Desta maneira, os desenvolvedores das ferramentas de predição de alvos devem agora investir na melhora da sensibilidade.

A união dos resultados das ferramentas TargetScan e miRanda-mirSVR obteve a elevada performance de 0.7 (em uma escala de 1 a -1), por agregar uma alta sensibilidade sem prejuízo de especificidade e precisão. Desta maneira, os resultados provenientes desta análise foram capazes de aprimorar a qualidade das análises *in silico* de predição de alvos, o que permitirá aos pesquisadores obterem resultados mais robustos durante a identificação de alvos dos miRNAs, economizando tempo e dinheiro em experimentos funcionais subsequentes.

No capítulo II, as análises de enriquecimento funcional, levando em consideração a intensidade da regulação fornecida pelos miRNAs, sugerem um novo nível de complexidade da atuação destas moléculas. Regular diversos genes associados a um mesmo processo biológico com intensidades semelhantes, enquanto processos biológicos distintos são regulados com intensidades diferentes é uma habilidade dos miRNAs que os torna capaz de regular de maneira precisa uma grande gama de funções biológicas distintas dentro de um mesmo contexto celular. Adicionalmente, nossas análises apontam que este é um fenômeno conservado entre as espécies de vertebrados analisadas, sugerindo que tal segregação regulatória vem sendo positivamente selecionada durante a evolução dos vertebrados.

O estudo aprofundado das funções biológicas reguladas pelos miRNAs miR-1, miR-7 e miR-9 trouxe também contribuições para a compreensão do comportamento dos miRNAs dentro da célula ou mesmo do organismo. Assim, a partir dos direcionamentos trazidos, novos estudos podem tentar identificar os padrões que causam a segregação da intensidade da regulação, o que possibilitaria a manipulação de uma via de ação dos miRNAs sem afetar as demais, minimizando os efeitos *off-target* comuns nos experimentos atuais.

6. Referências bibliográficas

- Agarwal V, Bell GW, Nam J, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. *eLife*. 4: e05005.
- Ambros, V (2004) The functions of animal microRNAs. *Nature*. 431: 350-5.
- Ameres & Zamore (2013) Ameres SL, Zamore PD. Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology*. 2013;14:475–488.
- Ashburner M., et al.. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, 25, 25–29.
- Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP 2008 The impact of microRNAs on protein output. *Nature*.;455(7209):64–71.
- Bandyopadhyay, S., Mitra, R. (2009). TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. *Bioinformatics*. **25(20)**:2625-31.
- Bang C, Batkai S, Dangwal S, Gupta SK, Foinquinos A, Holzmann A, Just A, Remke J, Zimmer K, Zeug A, Ponimaskin E, Schmiedl A, Yin X, Mayr M, Halder R, Fischer F, Engelhardt S, Wei Y, Schober A, Fiedler J, Thum T (2014) Cardiac fibroblast–derived microRNA passenger strand-enriched exosomes mediate cardiomyocyte hypertrophy. *J Clin Invest*;124(5):2136–46.
- Bazzini AA, Lee MT, Giraldez AJ (2012) Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* 336: 233–237.
- Betel D, Wilson M, Gabow A, Marks DS, Sander C (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res*. 36: D149-53.
- Betel D, Koppal A, Agius P, Sander C, Leslie C (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Gen Biol*. 11:R90.
- Bleazard T, Lamb JA, Griffiths-Jones S (2015) Bias in microRNA functional enrichment analysis. *Bioinformatics*, 31, 1592–1598.
- Borchert GM, Lanier W, Davidson BL (2006) RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol*. 13(12): 1097-1101.
- Broughton JP, Lovci MT, Huang JL, Yeo GW, Pasquinelli AE (2016) Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *64(2)*:320-333.
- Clarke C, Henry M, Doolan P, Kelly S, Aherne S, Sanchez N, Kelly P, Kinsella P, Breen L, Madden SF, Zhang L, Leonard M, Clynes M, Meleady P, Barron, N (2012) Integrated

- miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics*. 13:656.
- Chi SW, Hannon GJ, Darnell RB (2012) An alternative mode of microRNA target recognition. *Nat Struct Mol Biol*. 19(3):321-7.
- Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, Tsai TR, Ho SY, Jian TY, Wu HY, Chen PR, Lin NC, Huang HT, Yang TL, Pai CY, Tai CS, Chen WL, Huang CY, Liu CC, Weng SL, Liao KW, Hsu WL, Huang HD (2016) miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res*. 44(D1):D239-47
- Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ, Bork P, Arendt D 2010. Ancient animal microRNAs and the evolution of tissue identity. *Nature*. 463: 1084–88.
- Devlin AH, Thompson P, Robson T, McKeown SR (2010) Cytochrome P450 1B1 mRNA untranslated regions interact to inhibit protein translation. *Mol Carcinog*. 49(2):190-9.
- Enright AJ, John B, Gaul U, Tuschl T, Sander C and Marks DS (2003) MicroRNA targets in *Drosophila*. *Genome Biology*. 5;R1.
- Fan, X., Kurgan, L. (2015) Comprehensive overview and assessment of computational prediction of microRNA targets in animals. *Brief Bioinform*. 16(5):780-94.
- Flynt, AS N; Li, EJ; Thatcher, L; Solnica-Krezel, JG; Patton, JG (2007) Zebrafish miR-214 modulates Hedgehog signaling to specify muscle cell fate. *Nat Gen*. 39: 259-63.
- Flynt, AS; Thatcher, EJ; Burkewitz, K; Li, N; Liu, Y; Patton, JG (2009) *miR-8* microRNAs regulate the response to osmotic stress in zebrafish embryos *J Cell Biol*. 185(1): 115-27.
- Gaidatzis D, Nimwegen E, Hausser J, Zavolan M (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*. 8:69.
- Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lsc-6* and other microRNAs. *Nature Structural & Molecular Biology*. 18:1139–1146.
- Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*. 27(1):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*. 455:1193-1197.

- Gusev Y, Schmittgen TD, Lerner M, Postier R, Brackett D (2007) Computational analysis of biological functions and pathways collectively targeted by co-expressed microRNAs in cancer. *BMC Bioinformatics*. 8:S16.
- Hausser J, Syed AP, Bilén B, Zavolan M. (2013) Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res*. 23(4):604-15.
- Jia W, Li Z, Lun Z (2008) Discoveries and functions of virus-encoded MicroRNAs. *Chinese Science Bulletin*. 53:169–177.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS (2004). Human MicroRNA targets. *PLoS Biol*. 2:e363.
- Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol*. 57:19-53.
- Kanehisa M and Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28: 27–30.
- Karginov FV, Cheloufi S, Chong MM, Stark A, Smith AD, Hannon GJ (2010) Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. *Mol. Cell* 38, 781–788.
- Kertész M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition. *Nature Genetics*;39:1278–1284.
- Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. (2005) Combinatorial microRNA target predictions. *Nat Genet*.;37(5):495–500.
- Krutzfeldt J, Poy MN, Stoffel M (2006) Strategies to determine the biological function of microRNAs. *Nat Genet, Suppl*:S14-9.
- Lanet E, Delannoy E, Sormani R, Floris M, Brodersen P, Crete P, Voinnet O, Robaglia C (2009) Biochemical evidence for translational repression by Arabidopsis microRNAs. *Plant Cell*: 21, 1762–8.
- 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, Han J, Yeom KH, Jin H, Kim VN (2006) Drosha in primary microRNA processing. *CSH Symp Quant Biol*; 71:51-7.
- Lee YS, Dutta A (2009) MicroRNAs in cancer. *Annu Rev Pathol*, 4:199-227.
- 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.

- Liu N, Olson EN (2010) MicroRNA Regulatory Networks in Cardiovascular Development. *Dev Cell.* 18(4):510-25.
- Loher P and Rigoutsos I (2012) Interactive exploration of RNA22 microRNA target predictions. *Bioinformatics* 28, 3322–23.
- Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. *Science*, 303(5654):95-98.
- Lytle JR, Yario TA, Steitz JA (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A.* 104(23):9667-72.
- Mathonnet G, Fabian MR, Svitkin YV, Parsyan A, Huck L, Murata T, Biffo S, Merrick WC, Darzynkiewicz E, Pillai RS, Filipowicz W, Duchaine TF, Sonenberg N. (2007) MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* 317, 1764–1767.
- Molnár A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC (2007) miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature.* 447(7148):1126-9.
- Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA.* 13:1894–1910.
- Orom UA, Nielsen FC, Lund AH (2008) MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell.* 30(4):460-71.
- Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol.* **56(1)**:45–50.
- Powers, D.M.W. (2007). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies.* 2(1): 37-63.
- Rand TA, Petersen S, Du F, Wang X (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, 123(4):621-9.
- Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004). Fast and effective prediction of microRNA/target duplexes. *RNA* 10, 1507–1517.
- Reczko M, Maragkakis M, Alexiou P, Grosse I, Hatzigeorgiou AG (2012) Functional microRNA targets in protein coding sequences. *Bioinformatics.* 28:771–776.
- Reddy, K. B. (2015). MicroRNA (miRNA) in cancer. *Cancer. Cell. Int.* **15**:38.

- Schirle NT, Sheu-Gruttadauria J, MacRae IJ (2014) Structural basis for microRNA targeting. *Science*. 346:608–613.
- Shenoy, A., Blelloch, R.H. (2014). Regulation of microRNA function in somatic stem cell proliferation and differentiation. *Nat. Rev. Mol. Cell. Biol.* **15**(9):565-576.
- Shin C, Nam J, Farh KK, Chiang HR, Shkumatava A, Bartel DP (2010) Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell*; 38, 789–802.
- Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*. 320(5883):1643-7.
- Schnall-Levin M, Zhao Y, Perrimon N, Berger B (2010) Conserved microRNA targeting in *Drosophila* is as widespread in coding regions as in 3'UTRs. *Proc Natl Acad Sci U S A*. 107(36):15751-6.
- Shkumatava A, Stark A, Sive H, Bartel DP (2009) Coherent but overlapping expression of microRNAs and their targets during vertebrate development. *Genes Dev* 23: 466–481.
- Sturm M, Hackenberg M, Langenberger D, Frishman D (2010) TargetSpy: a supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics*. 11.
- Sunkar R, Girke T, Jain PK, Zhu JK (2005) Cloning and Characterization of MicroRNAs from Rice. *Plant Cell*. 17:1397-1411.
- Takacs CM and Giraldez AJ (2011) miR-430 regulates oriented cell division during neural tube development in zebrafish. *Dev Biol*. 409(2):442-450.
- Tang G, Reinhart BJ, Bartel DP, Zamore PD (2003) A biochemical framework for RNA silencing in plants. *Genes Dev*; 17, 49–63.
- Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos, I (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*. 455(7216):1124-8.
- Witkos TM, Koscianska E, Krzyzosiak WJ (2011) Practical Aspects of microRNA Target Prediction. *Curr Mol Med*. 11(2): 93–109.
- Zdanowicz A, Thermann R, Kowalska J, Jemielity J, Duncan K, Preiss T, Darzynkiewicz E, Hentze MW (2009) *Drosophila* miR2 primarily targets the m⁷ GpppN cap structure for translational repression. *Mol. Cell* 35, 881–888.
- Zhou H, Rigoutsos I (2014) MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. *RNA*. 20(9):1431-9. doi: 10.1261/rna.045757.114.