Biologia Molecular Computacional
IBI5035/QBQ2507 - 2023

Predição de alvos de microRNAs

Eduardo Moraes Rego Reis
Instituto de Química - USP

microRNAs

- Pequenos RNAs (18-22nt) codificados no genoma de eucariotos

- Complementares a mRNAs. Geralmente pareiam com na região 3'UTR

- Em geral tem mais de um mRNA-alvo.

- Regulam pós-transcricionalmente a expressão gênica: afetam a estabilidade do mRNA (aumentam a degradação) ou inibem a tradução do mRNA.

- Mais de 1000 microRNAs expressos no genoma humano, cada um potencialmente regulando a expressão de milhares de genes-alvo.
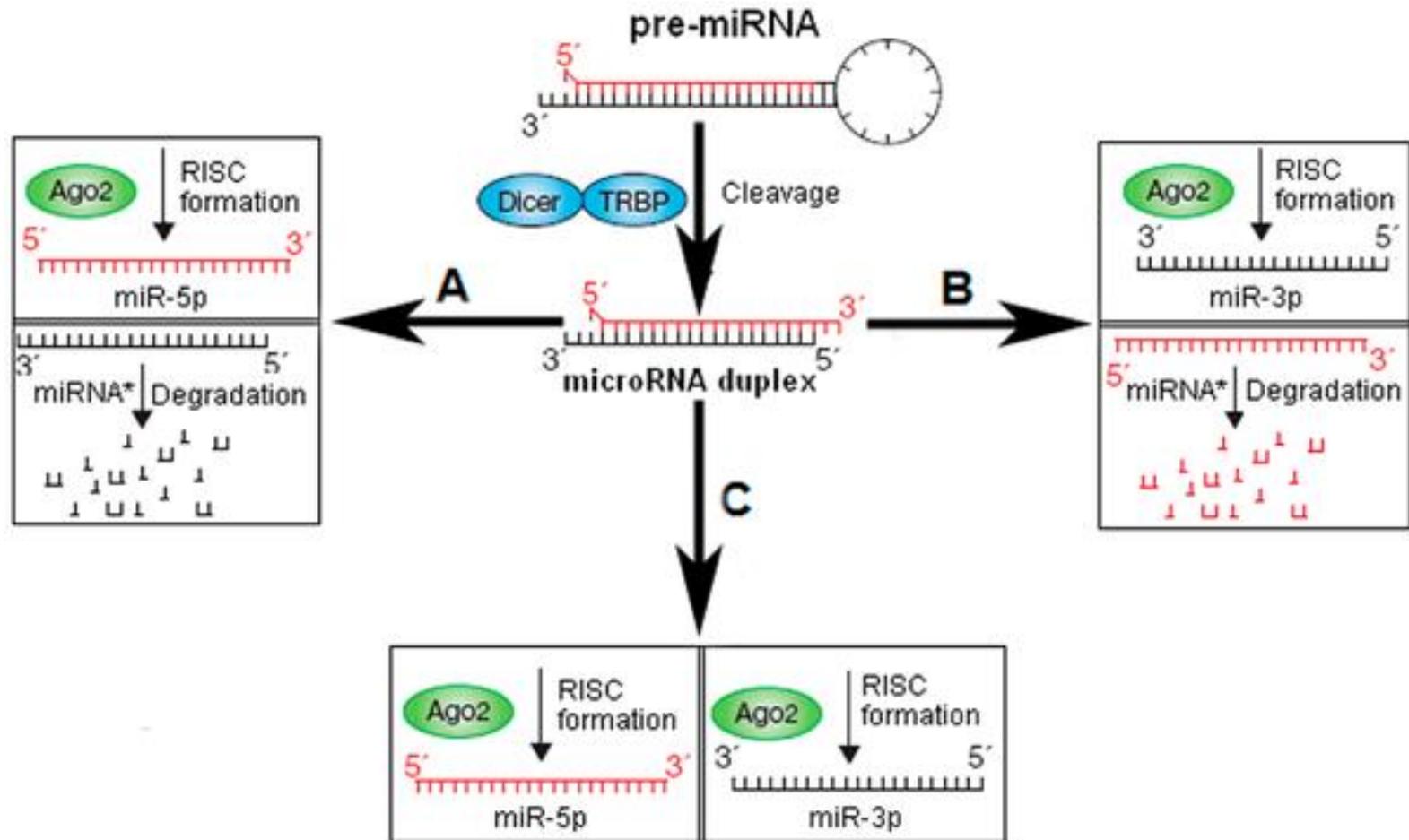
Biogênese de microRNAs
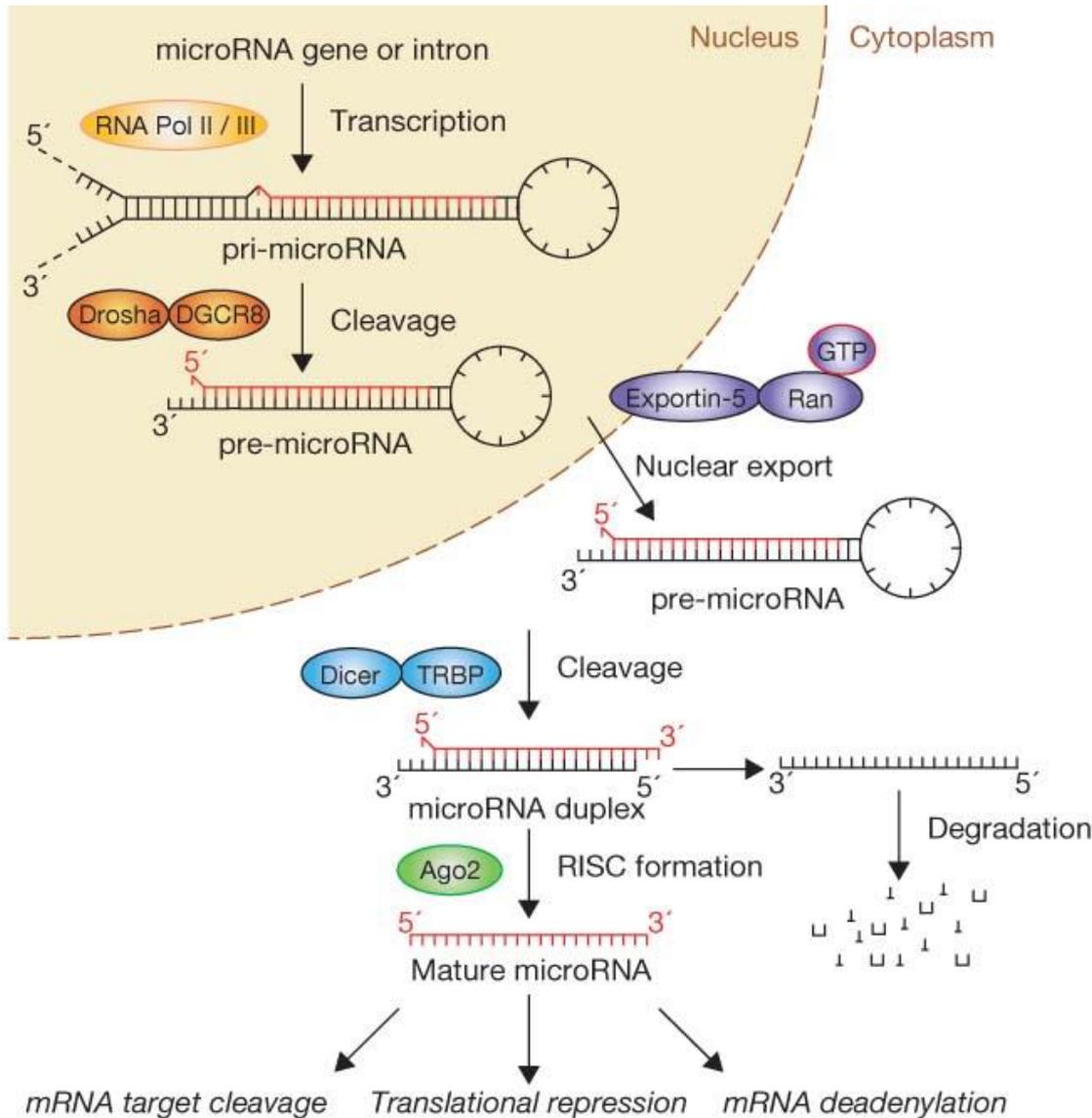
- No núcleo, Drosha-DCGR8 corta o pri-miRNA liberando um pre-miRNA em forma de grampo com ~70 nt

- O pre-miRNA é exportado para o citoplasma

- No citoplasma, a enzima Dicer corta o pre-miRNA e libera um duplex de RNA com ~22 nt que corresponde ao microRNA maduro

- Uma das fitas de RNA, a **fita guia** ("guide strand") é incorporada no complexo de silenciamento RISC ("RNA Induced Silencing Complex") e a **fita complementar** ("passenger strand") é degradada

Funções de microRNAs *in vivo*

Process	microRNA	Targets	Function	Evidence	Key citations
C. elegans					
Developmental timing	*lin-4* microRNA	*lin-14, lin-28*	Stem cell differentaition	LOF	Lee et al., 1993; Wightman et al., 1993
Developmental timing	*let-7* microRNA	*lin-41, hbl-1, daf-12, pha-4*	Stem cell differentaition	LOF	Reinhart et al., 2000; Abrahante et al., 2003; Lin et al., 2003; Slack et al., 2000; Grosshans et al., 2005
Developmental timing	miR-48, miR-84, miR-241	*hbl-1*	Stem cell differentaition	LOF	Abbott et al., 2005; Lin et al., 2005
Developmental timing	miR-48, miR-84	Unknown	Cessation of molting	LOF	Abbott et al., 2005
Organogenesis	miR-84	*let-60*	Differentiation/proliferation	GOF	Johnson et al., 2005
Differentiation	*lsy-6* microRNA	*cog-1*	Left-right asymmetry	LOF	Johnston and Hobert, 2003
Differentiation	miR-273	*die-1*	Left-right asymmetry	GOF	Chang et al., 2004
D. melanogaster					
Growth control and programmed cell death	*bantam* microRNA	*hid*	Proliferation/programmed cell death	LOF	Brennecke et al., 2003
Programmed cell death	miR-14	Unknown	Programmed cell death	LOF	Xu et al., 2003
Patterning and embryogenesis	miR-2a, -2b, -6, -7	*E(spl)/bHLH, bearded* families	Notch signalling	GOF	Brennecke et al., 2005; Lai, 2002; Lai et al., 2005; Stark et al., 2003
Embryogenesis and programmed cell death	miR-2, -6, -11, -13, -308	Unknown	Programmed cell death	2-O-Me-RNA	Leaman et al., 2005
D. rerio					
Differentiation and organogenesis	miR-430	Unknown	Neurogenesis	Dicer rescue	Giraldez et al., 2005
M. musculus					
Differentiation and organogenesis	miR-1	Hand2	Angiogenesis	GOF	Zhao et al., 2005
Differentiation and organogenesis	miR-181	Unknown	Hematopoiesis	GOF	Chen et al., 2004
Insulin secretion	miR-375	*Myotrophin (Mtpn)*	Exocytosis	2-O-Me-RNA	Poy et al., 2004
Human disease	miR-17, -18, -19a, -20, -19b-1, -92-1	Unknown	Tumorigenesis	GOF	He et al., 2005; O'Donnell et al., 2005
H. sapiens					
Human disease	miR-32	PFV-1	Viral defense	LNA	Lecellier et al., 2005

This table includes all microRNAs that have been analyzed in vivo using loss-of-function studies, and microRNAs for which a likely function has been demonstrated by using an indirect approach, e.g. mis-expression experiments. It table does not contain microRNAs for which a target mRNA has been predicted and validated using overexpression experiments, but for which no further functional characterization at the cellular or organismal level has been carried out.
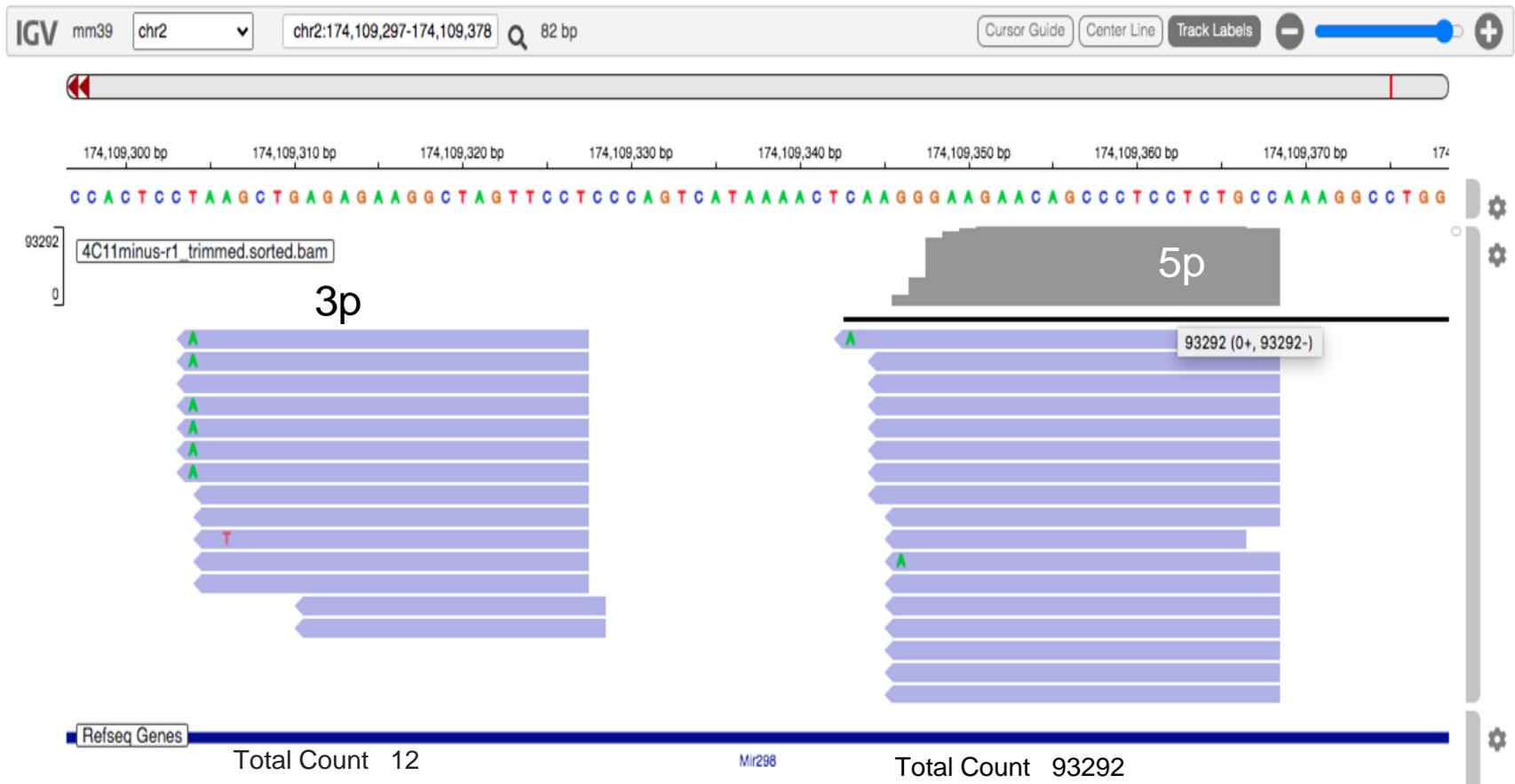
Alvarez-Garcia e Miska, 2005 Development 132:4653-4662

Em geral a fita com **menor estabilidade termodinâmica (maior ΔG) na extremidade 5'** do duplex de miRNA é carregada no complexo RISC

Nomenclatura 5p – 3p permite distinguir as 2 fitas do duplex

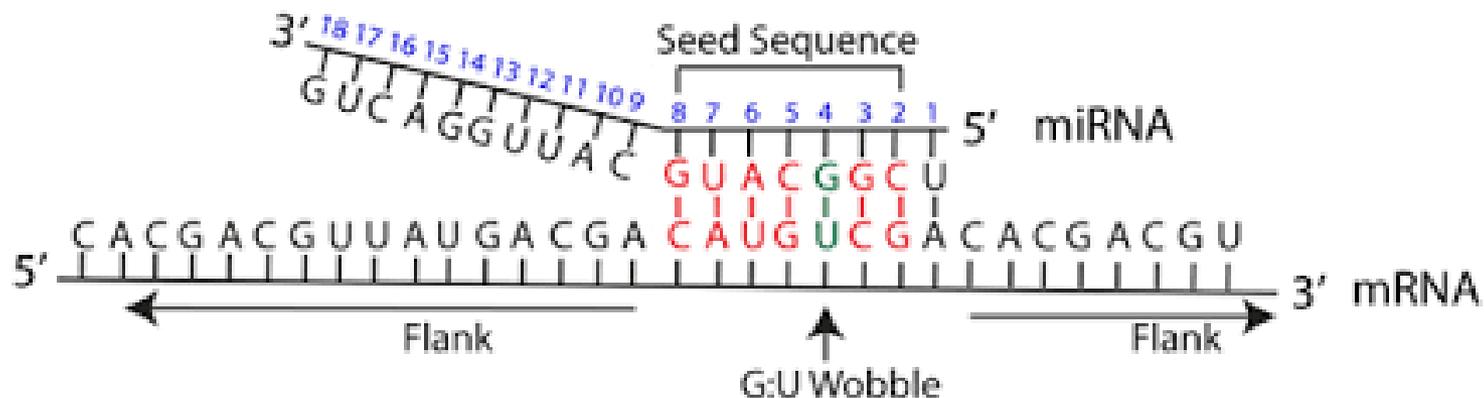# O sequenciamento de pequenos RNAs expressos indica a orientação 3p ou 5p dos microRNAs que se acumulam na célula/tecido

https://mirbase.org/hairpin/MI0000398

Literature search

27 open access papers mention mmu-mir-298
(136 sentences)

Sequence

350557 reads, 1278 reads per million, 103 experiments   Show Histogram

ccaggccuuuGGCAGAGGAGGGCUGUUCUUCCCuugaguuuuaugacuggGAGGAACUAGCCUUCUCUCAGCUUaggagugg
(((..(((..(((.(((((((((((((((((((((...(((((....)))))))))))))))))).))))))))).))).))).)))..)))

Structure

```
    gg   uu   A   -         -          uug    u
cca  ccu  GGC GAG GAGGGCU GUUCUUCCC   aguu u
|||  |||  ||| ||| ||||||| |||||||||   ||||
ggu  gga  UCG CUC CUUCCGA CAAGGAGgg   ucag a
  ga  -U   A   U          U          ---    u
```

Apenas o miR-298-5p é expresso nesse caso

A interação do microRNA e o mRNA–alvo depende de pequenas regiões de complementaridade (6 a 8 bases) na extremidade 5' do microRNA

Sequencias "seed"

Sítios canônicos e sítios atípicos envolvidos no pareamento entre microRNAs e RNAs alvo

Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009 Jan 23;136(2):215-33.

Existem famílias de microRNAs evolutivamente relacionados

- Podem possuir sequencias "seed" conservadas

- Em geral alvejam genes relacionados

Ex.:

hsa-miR-30 family

miR-30d	UGUAAACAUCCCCGACUGGAAG--
miR-30a	UGUAAACAUCCUCGACUGGAAG--
miR-30e	UGUAAACAUCCUUGACUGGAAG--
miR-30c	UGUAAACAUCCUACACUCUCAGC-
miR-30b	UGUAAACAUCCUACACUC--AGCU

miR-30 FI   3'-CATTTGTAGG-5'

Devido ao pequeno tamanho, existe uma alta probabilidade de microRNAs parearem ao acaso com outros RNAs

A **predição de alvos de microRNAs biologicamente relevantes** aumenta quando se considera:

i) a evidência de conservação evolutiva nos sítios de ligação e

ii) a presença de múltiplos sítios de ligação no mesmo mRNA (regulação cooperativa)

Whole genome comparison of real versus randomized miRNAs against the complete genomes of *D. melanogaster* and *D. pseudoobscura*

	Total hits	Total conserved hits	1 site	≥ 2 sites
		Apenas uma fração dos sítios são conservados	**Alguns miRNAs alvejam o mesmo mRNA em mais de um sítio**	
73 *D. melanogaster* miRNAs (A)	6,864	589	556	33
73 Random miRNAs (B)	5,152	204	201	3
Standard deviation (100 experiments)	± 132	± 43	± 40	± 3
Ratio (A/B)		2.9	2.8	11.0
Estimated false positives (%)		35%	36%	9%

Detected conserved hits (especially those with multiple detected sites in the 3' UTR) are significantly over-represented (2.8× and 11× as many cases, on average, respectively) in analyses with actual miRNAs compared to randomly shuffled miRNAs. The thresholds used for this analysis were S: 100; ΔG: -19 kcal/mol; ID: 70%.
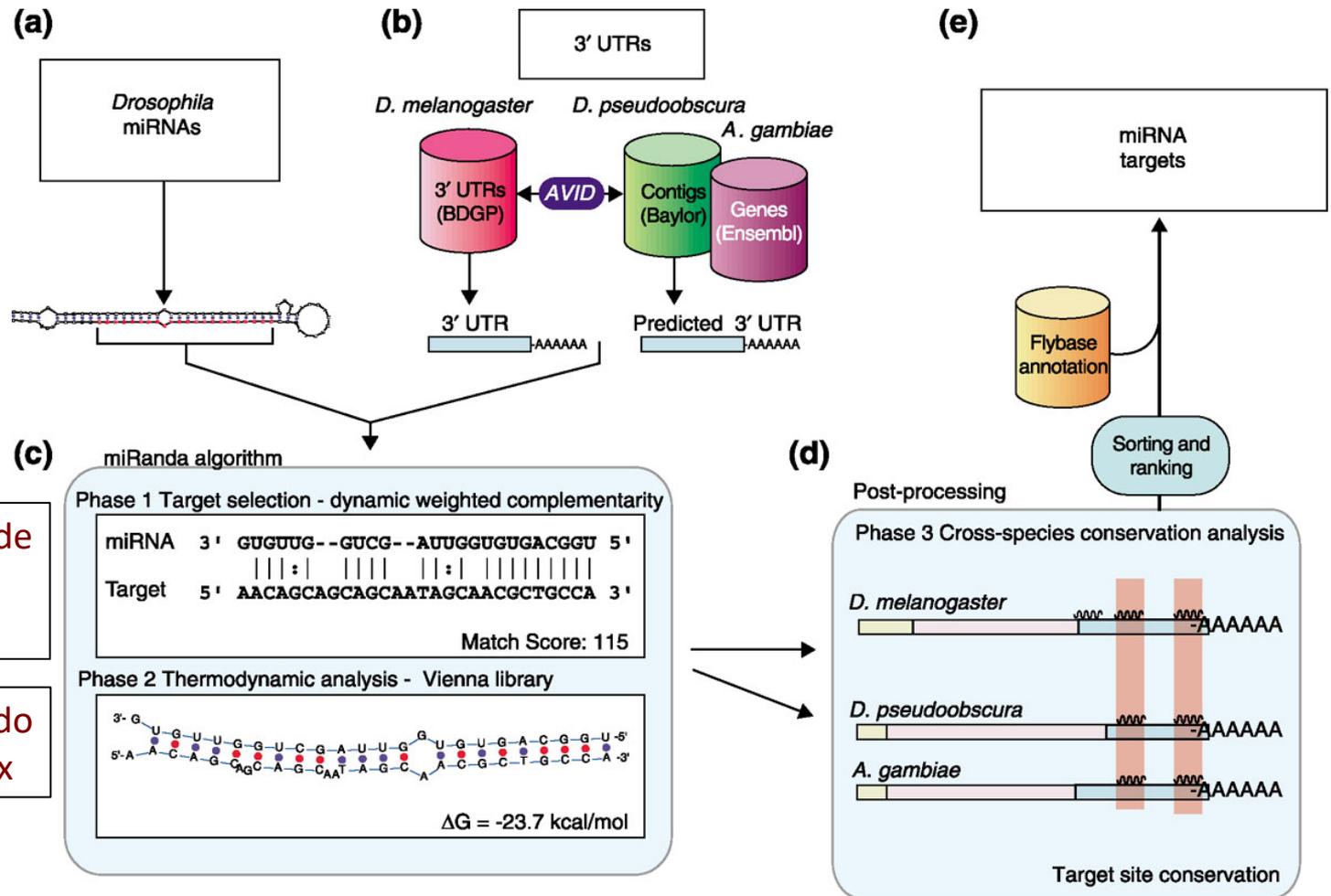
Caracteristicas de redes regulatórias mediadas por microRNAs

Multiplicidade: um microRNA alveja diferentes genes
Cooperatividade: multiplos sítios de ligação de um ou mais microRNAs em um mesmo mRNA alvo

Critérios utilizados para a predição de alvos de microRNAs

- Complementaridade entre sequencia "seed" do microRNA e mRNA-alvo.

- Existência de conservação evolutiva. Em geral há uma conservação maior na região "seed" do alvo.

- Propriedades termodinâmicas. Formação de pareamento estável entre microRNA-mRNA alvo a partir de critério (minimização de energia livre).

- Acessibilidade do sítio de ligação no mRNA-alvo (estruturas secundárias podem mascarar e impedir a ligação do microRNA)

- Presença de múltiplos sítios de ligação de microRNAs

- Modelos preditivos baseados em aprendizado de maquina que utilizam parametros teóricos (mencionados acima) e interações entre microRNA-mRNA alvo validadas experimentalmente.

- Uma combinação dos critérios acima.
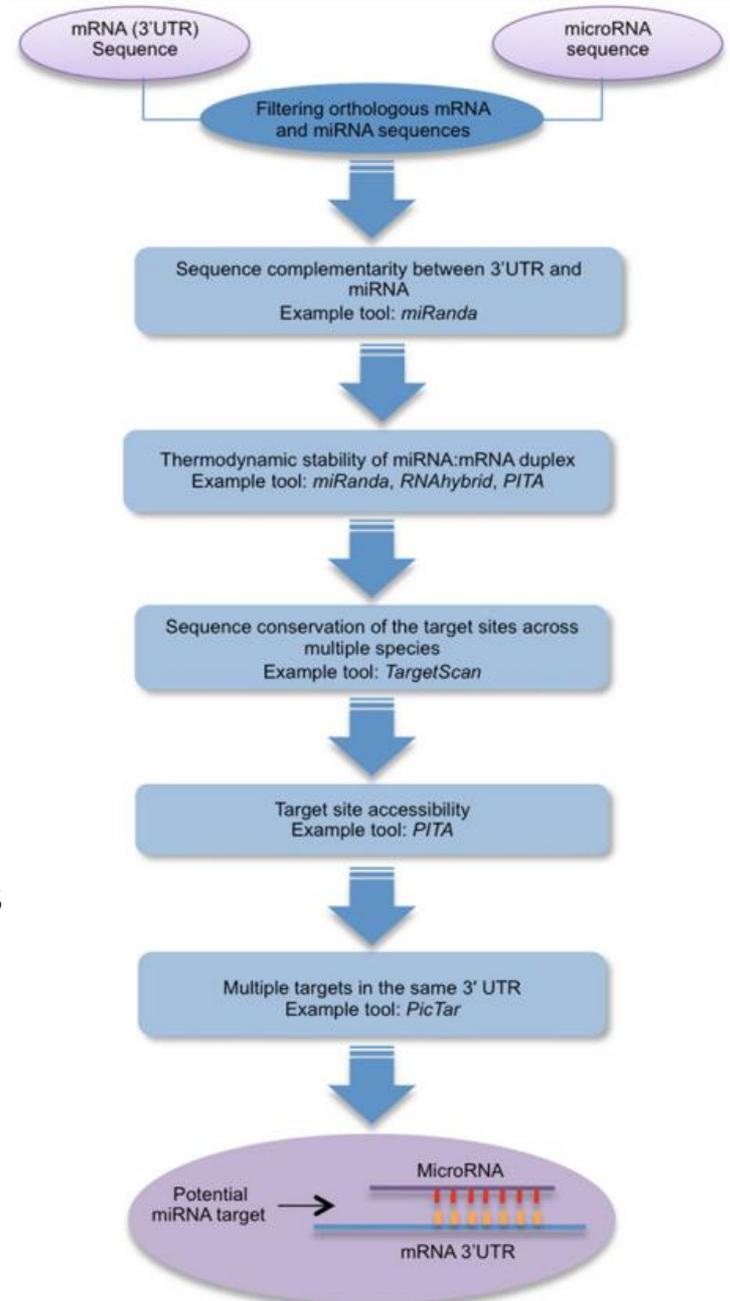
Fluxograma para identificação de genes-alvo de microRNAs (miRanda)

(a) Drosophila miRNAs

(b) 3' UTRs — D. melanogaster, D. pseudoobscura, A. gambiae — 3' UTRs (BDGP) ← AVID → Contigs (Baylor), Genes (Ensembl) — 3' UTR ┤AAAAAA — Predicted 3' UTR ┤AAAAAA

(c) miRanda algorithm

Phase 1 Target selection - dynamic weighted complementarity

```
miRNA   3' GUGUUG--GUCG--AUUGGUGUGACGGU 5'
           |||:| |||||  ||:| ||||||||||
Target  5' AACAGCAGCAGCAATAGCAACGCTGCCA 3'
```

Match Score: 115

Phase 2 Thermodynamic analysis - Vienna library

ΔG = -23.7 kcal/mol

(d) Post-processing

Phase 3 Cross-species conservation analysis

D. melanogaster, D. pseudoobscura, A. gambiae — ┤AAAAAA

Target site conservation

(e) miRNA targets — Flybase annotation — Sorting and ranking

1) Complementaridade de sequencia entre miRNA - RNA alvo

2) Energia livre do RNA-RNA duplex

3) Conservação dos sítios-alvo em outros genomas

Bancos de dados de microRNA e genes alvo

https://en.wikipedia.org/wiki/MicroRNA_and_microRNA_target_database

microRNA target gene databases [edit]

Name	Description	type	Link
StarBase	starBase is designed for decoding **miRNA-lncRNA**, miRNA-mRNA, miRNA-**circRNA**, miRNA-**pseudogene**, miRNA-sncRNA, **protein-lncRNA**, protein-sncRNA, protein-mRNA and protein-pseudogene interactions and **ceRNA** networks from 108 CLIP-Seq (HITS-CLIP, PAR-CLIP, iCLIP, CLASH) datasets. It also provides **Pan-Cancer Analysis** for microRNAs, lncRNAs, circRNAs and protein-coding genes from 6000 tumor samples.	database	website
StarScan	StarScan is developed for scanning small RNA (miRNA, piRNA, siRNA) mediated RNA cleavage events in lncRNA, circRNA, mRNA and pseudo genes from degradome sequencing data.	web-based software	website
Cupid	Cupid is a method for **simultaneous prediction of miRNA-target interactions and their mediated competing endogenous RNA (ceRNA) interactions**. It is an integrative approach significantly improves on miRNA-target prediction accuracy as assessed by both mRNA and protein level measurements in breast cancer cell lines. Cupid is implemented in 3 steps: Step 1: re-evaluate candidate miRNA binding sites in 3' UTRs. Step2: interactions are predicted by integrating information about selected sites and the statistical dependency between the expression profiles of miRNA and putative targets. Step 3: Cupid assesses whether inferred targets compete for predicted miRNA regulators. * Only the source code for step 3 is provided.	software (MATLAB)	website
TargetScan	Predicts biological targets of miRNAs by searching for the presence of sites that match the seed region of each miRNA. In flies and nematodes, predictions are ranked based on the probability of their evolutionary conservation. In zebrafish, predictions are ranked based on site number, site type, and site context, which includes factors that influence target-site accessibility. In mammals, the user can choose whether the predictions should be ranked based on the probability of their conservation or on site number, type, and context. In mammals and nematodes, the user can choose to extend the predictions beyond conserved sites and consider all sites.	database, webserver	website

microRNA databases [edit]

Name	Description	type	Link
deepBase	deepBase is a database for annotating and discovering small and long ncRNAs (microRNAs, siRNAs, piRNAs...) from high-throughput deep sequencing data.	database	website
miRBase	miRBase database is a searchable database of published miRNA sequences and annotation.	database	website
microRNA.org	microRNA.org is a database for Experimentally observed microRNA expression patterns and predicted microRNA targets & target downregulation scores.	database	website
miRGen 2.0	miRGen 2.0: a database of microRNA genomic information and regulation	database	website
miRNAMap	miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes	database	website
PMRD	PMRD: plant microRNA database	database	website
TargetScan	TargetScan7.0 classifies microRNAs according to their level of conservation (i.e., species-specific, conserved among mammals, or broadly conserved among vertebrates) and aggregates them into families based upon their seed sequence. It also annotates conserved isomiRs using small RNA sequencing datasets.[10]	database	website
VIRmiRNA	**VIRmiRNA** is the first dedicated resource on experimental **viral miRNA** and their **targets**. This resource also provides inclusive knowledge about anti-viral miRNAs known to play role in antiviral immunity of host.	Database	website

Banco de dados de microRNAs ("hairpin" e sequencia processada)

http://www.mirbase.org/

Genes-alvo de microRNAs validados experimentalmente

http://mirtarbase.mbc.nctu.edu.tw/php/index.php

miRTarBase

miRTarBase: the experimentally validated microRNA-target interactions database

As a database, miRTarBase has accumulated more than three hundred and sixty thousand miRNA-target interactions (MTIs), which are collected by manually surveying pertinent literature after NLP of the text systematically to filter research articles related to functional studies of miRNAs. Generally, the collected MTIs are validated experimentally by reporter assay, western blot, microarray and next-generation sequencing experiments. While containing the largest amount of validated MTIs, the miRTarBase provides the most updated collection by comparing with other similar, previously developed databases.

Major improvments

Features	miRTarBase 6.0	miRTarBase 7.0
Release date	2015/09/15	2017/09/15
Known miRNA entry	miRBase v20	miRBase v21
Known Gene entry	Entrez 2015	Entrez 2017
Species	18	23
Curated articles	4,966	8,510
miRNAs	3,786	4,076
Target genes	22,563	23,054
CLIP-seq datasets	138	231
Curated miRNA-target interactions	366,181	422,517
Text-mining technique to prescreen literature	NLP	Enhanced NLP
Download by validated miRNA-target sites	None	Yes
Browse by miRNA, gene, and disease	None	Yes

Current curation

Release 7.0: Sept. 15, 2017

Number of articles: 8,510

Number of species: 23

Number of target genes: 23,054

Number of miRNAs: 4,076

Number of miRNA-target interactions: 422,517
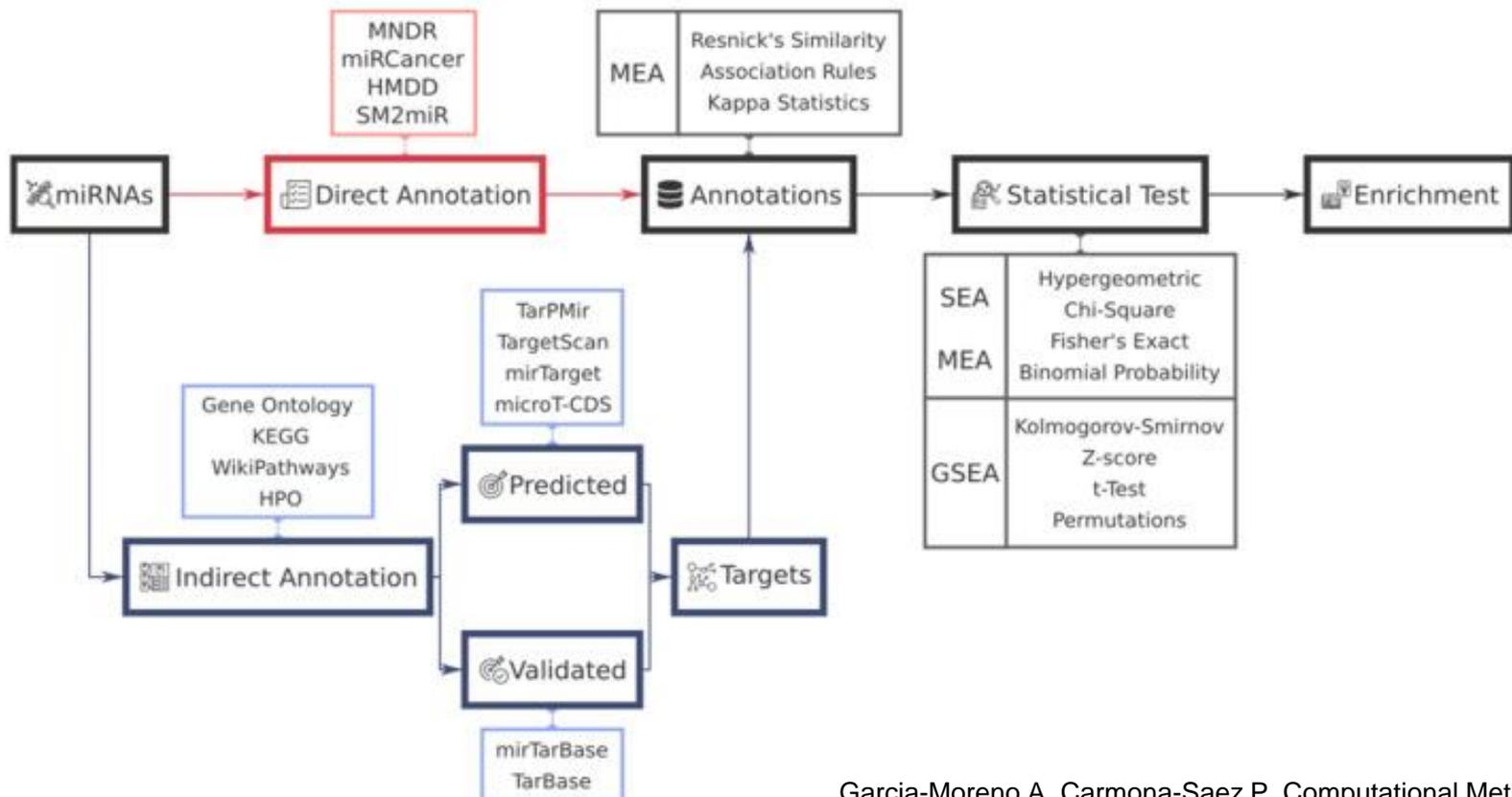
MicroRNA resources from ISBLAB

miRTar - An integrated web server for identifying miRNA-target interactions

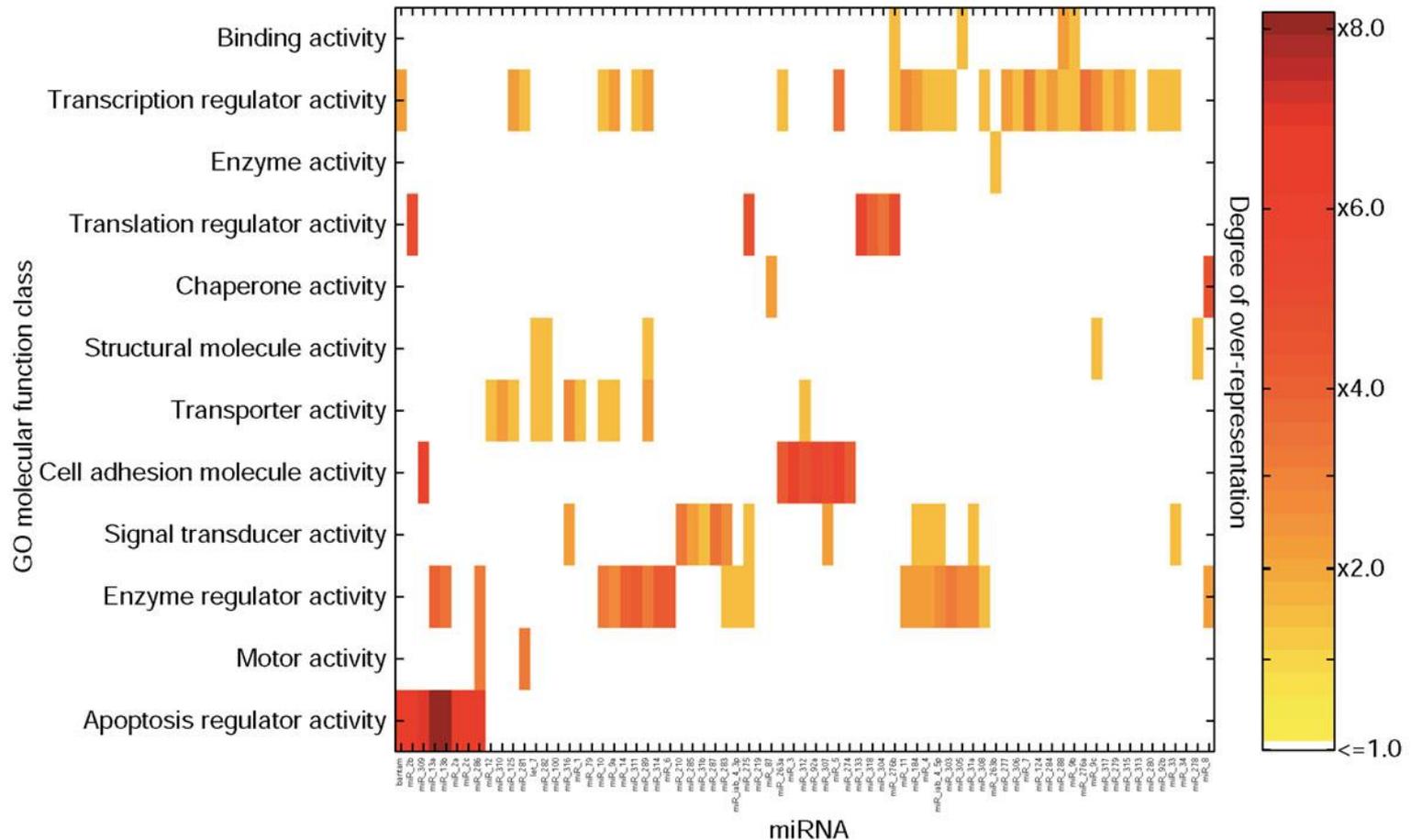# Anotação funcional de microRNAs e genes-alvo de regulação

Experimentos de RNAseq de pequenos RNA (miRNAseq) retornam listas de microRNAs com expressão diferencial em células ou tecidos.

É possível fazer uma inferência de vias moleculares reguladas por miRNAs de forma direta (pouco miRNAs tem esse grau de anotação) ou indireta, através da análise dos genes alvos.

A análise de enriquecimento de categorias funcionais entre os genes alvejados por um conjunto de miRNAs pode ser feita usando abordagens apresentadas em aula anterior (gProfiler, GSEA)

Garcia-Moreno A, Carmona-Saez P. Computational Methods and Software Tools for Functional Analysis of miRNA Data. Biomolecules. 2020 Aug 28;10(9):1252.
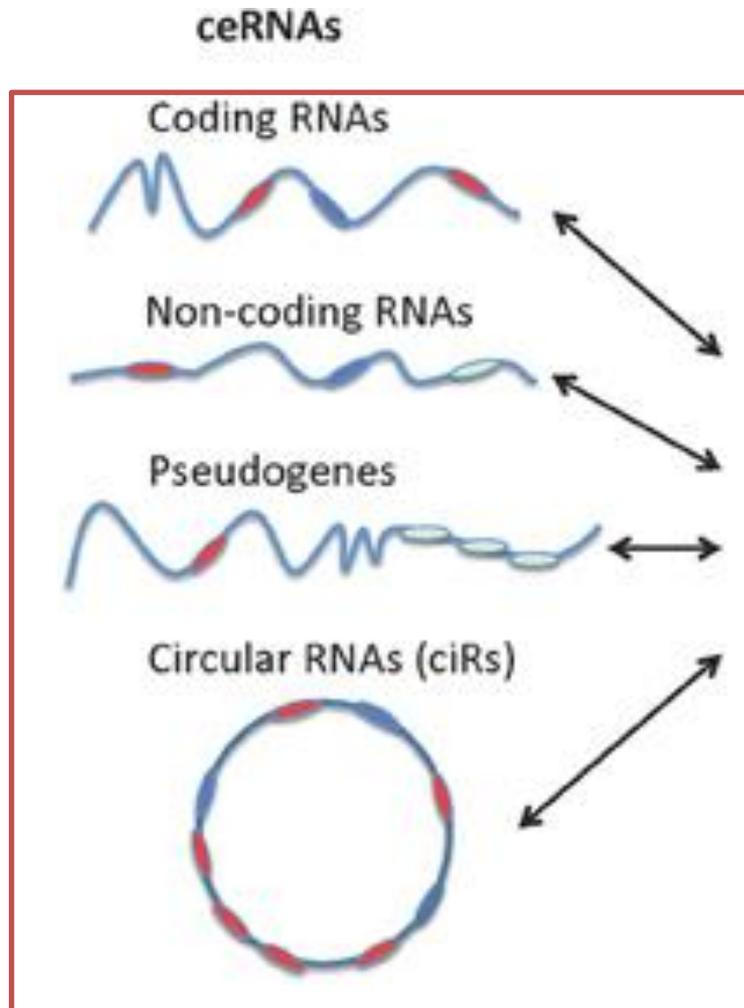
Diferentes microRNAs alvejam conjuntos de genes enriquecidos em diferentes categorias funcionais
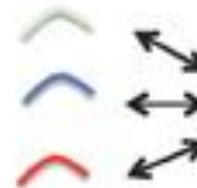
Diferentes classes de RNAs podem atuar como competidores endógenos de microRNAs (ceRNAs) e cooperar para o controle do nível de expressão de genes alvo

ceRNAs

Coding RNAs

Non-coding RNAs

Pseudogenes

Circular RNAs (ciRs)

miRNAs

mRNA Y

Possíveis ceRNAs

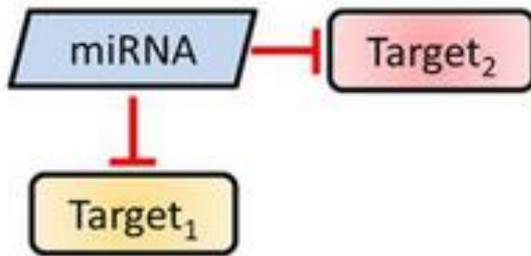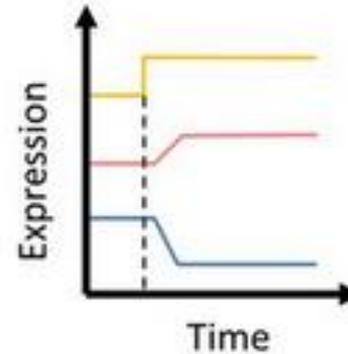ceRNAs possuem sítios de ligação de microRNAs reduzindo a disponibilidade destes para se ligar no mRNA-alvo

Gene-alvo de microRNA

Redes de co-expressão de miRNAs-ceRNAs envolvidas no ajuste fino da expressão gênica (controle pós-transcricional)
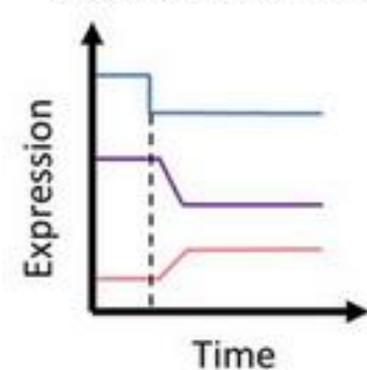
Bancos de dados de interações microRNA-ceRNAs
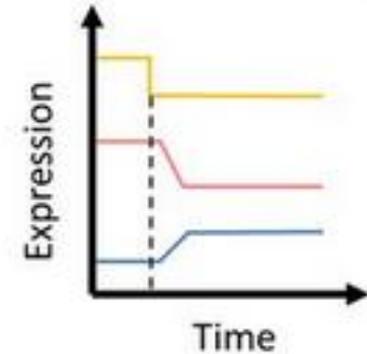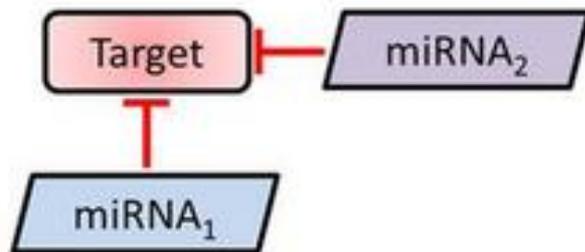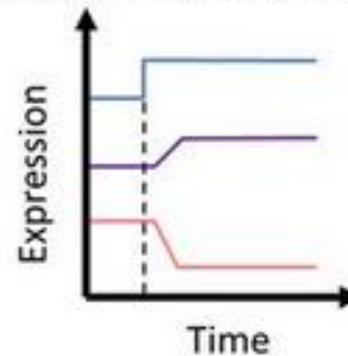
http://cupidtool.sourceforge.net/

Cupid: simultaneous reconstruction of miRNA-target and ceRNA networks

Cupid is a method for simultaneous prediction of miRNA-target interactions and their mediated competitive endogenous RNA (ceRNA) interactions. We showed that our integrative approach significantly improves on miRNA-target prediction accuracy as assessed by both mRNA and protein level measurements in breast cancer cell lines.
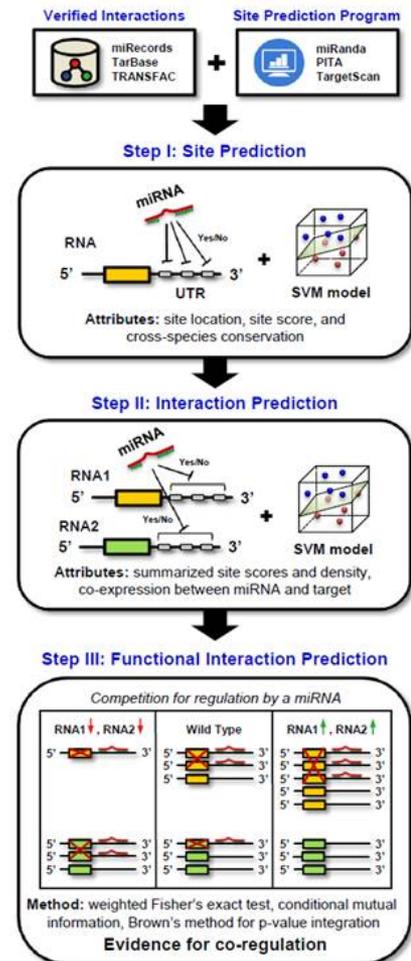
Cupid is implemented in 3 steps:

- o Step 1: re-evaluate candidate miRNA binding sites in 3' UTRs, as inferred by TargetScan, miRanda and PITA by integrating their scores, location in the 3' UTR, and cross-species conservation.
- o Step2: interactions are predicted by integrating information about selected sites, their multiplicity, and the statistical dependency between the expression profiles of miRNA and putative targets. Likelihoods for each predictive feature are computed based on a positive gold standard set.
- o Step 3: Cupid assesses whether inferred targets compete for predicted miRNA regulators.

- Ø Download Cupid v1.0 and its user guide
- Ø Download Cupid poster presented in Symposia on Cancer Research 2014 Illuminating Genomic Dark Matter "NcRNA in Disease and Cancer""

References:

- o Hua-Sheng Chiu*, María Rodríguez Martínez* *et al*. 2018. The number of titrated microRNA species dictates ceRNA regulation. *Nucleic Acids Research* (accepted).
- o Hua-Sheng Chiu *et al*. 2017. High-throughput validation of ceRNA regulatory networks. *BMC Genomics* 18: 418.
- o Hua-Sheng Chiu*, David Llobet-Navas*, *et al*., Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks. *Genome Research*. 2015 Feb;25(2):257-67
- o Pavel Sumazin*, Xuerui Yang*, Hua-Sheng Chiu*, *et al*., An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*. 2011 Oct 14;147(2):370-81

Verified Interactions — miRecords, TarBase, TRANSFAC
Site Prediction Program — miRanda, PITA, TargetScan

Step I: Site Prediction
RNA / miRNA / 5' / UTR / 3' / SVM model
Attributes: site location, site score, and cross-species conservation

Step II: Interaction Prediction
miRNA / RNA1 / RNA2 / SVM model
Attributes: summarized site scores and density, co-expression between miRNA and target

Step III: Functional Interaction Prediction
Competition for regulation by a miRNA
RNA1↓, RNA2↓ / Wild Type / RNA1↑, RNA2↑
Method: weighted Fisher's exact test, conditional mutual information, Brown's method for p-value integration
Evidence for co-regulation

Tutoriais - microRNAs

- **Tutorial na pasta de atividade do moodle**