Aprimoramento de uma ferramenta de predição de ncRNAs baseada em estruturas secundárias

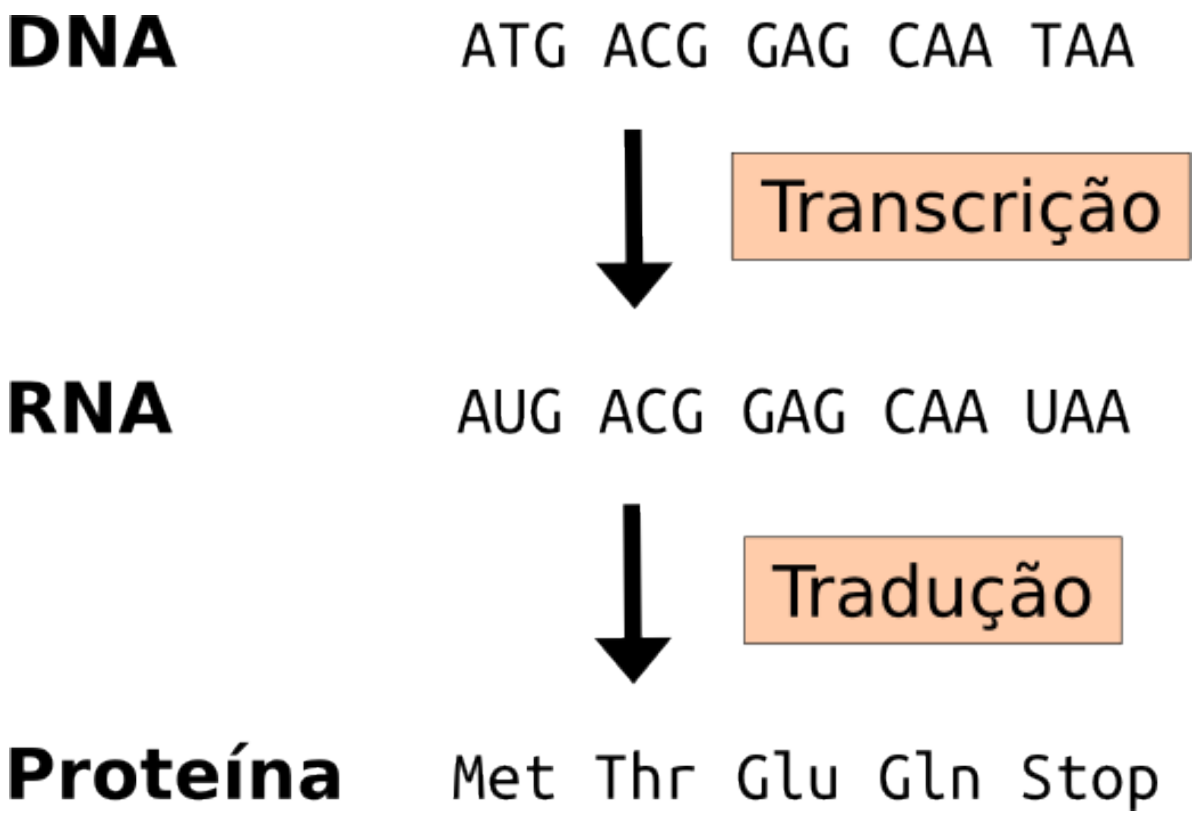
Trabalho de Formatura Supervisionado

Felipe Lamberti Amado
Orientador: Prof. Dr. Alan Mitchell Durham

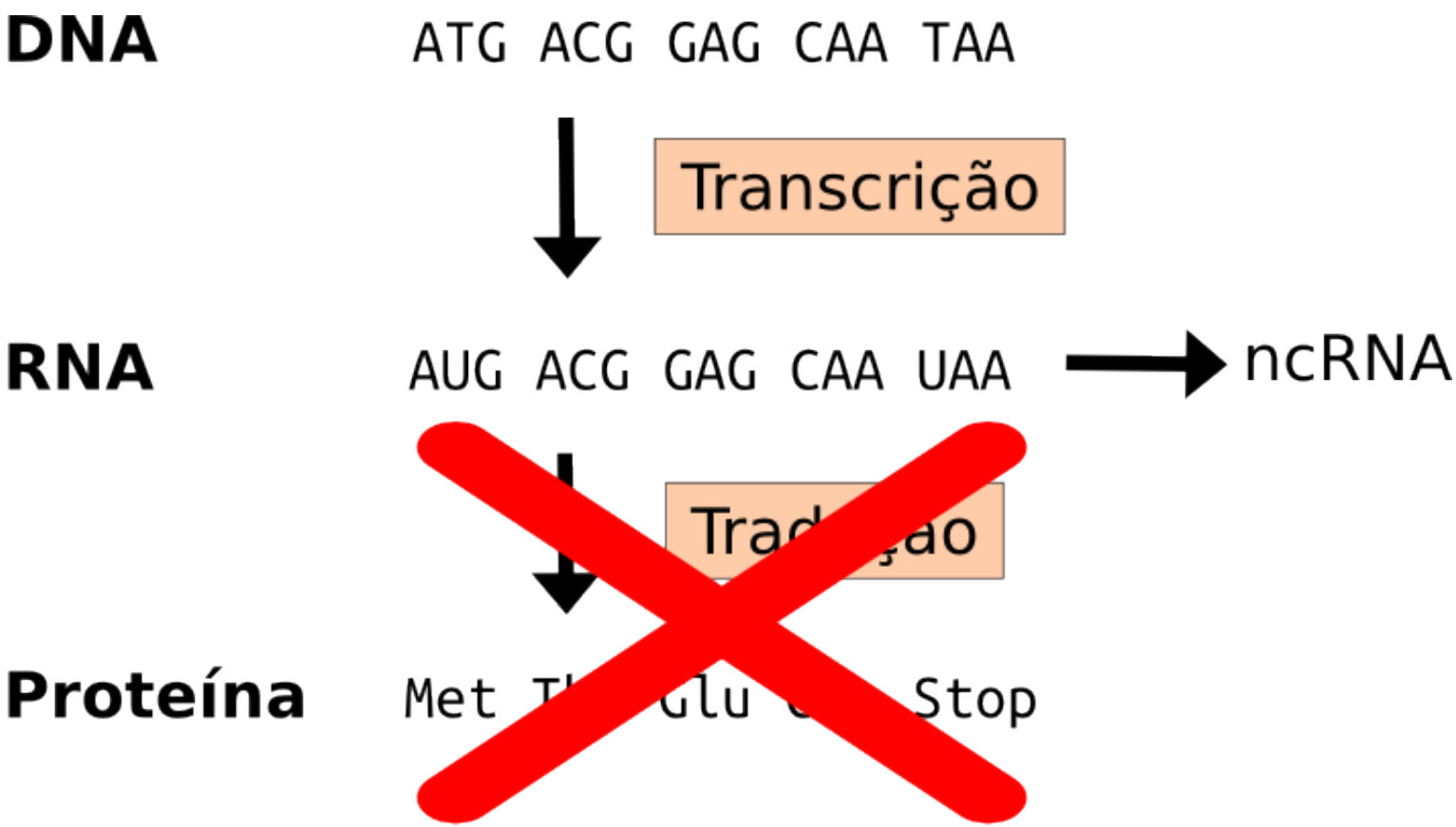
Instituto de Matemática e Estatística - USP

São Paulo, 11 de novembro de 2013

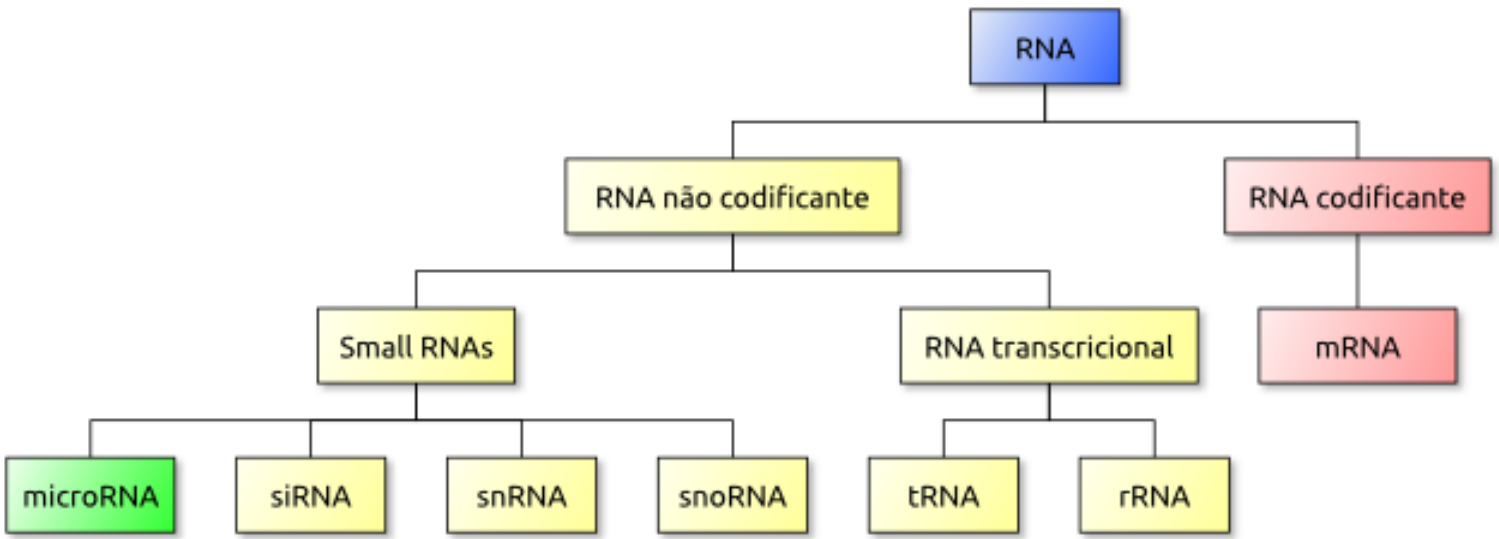
Dogma central da Biologia



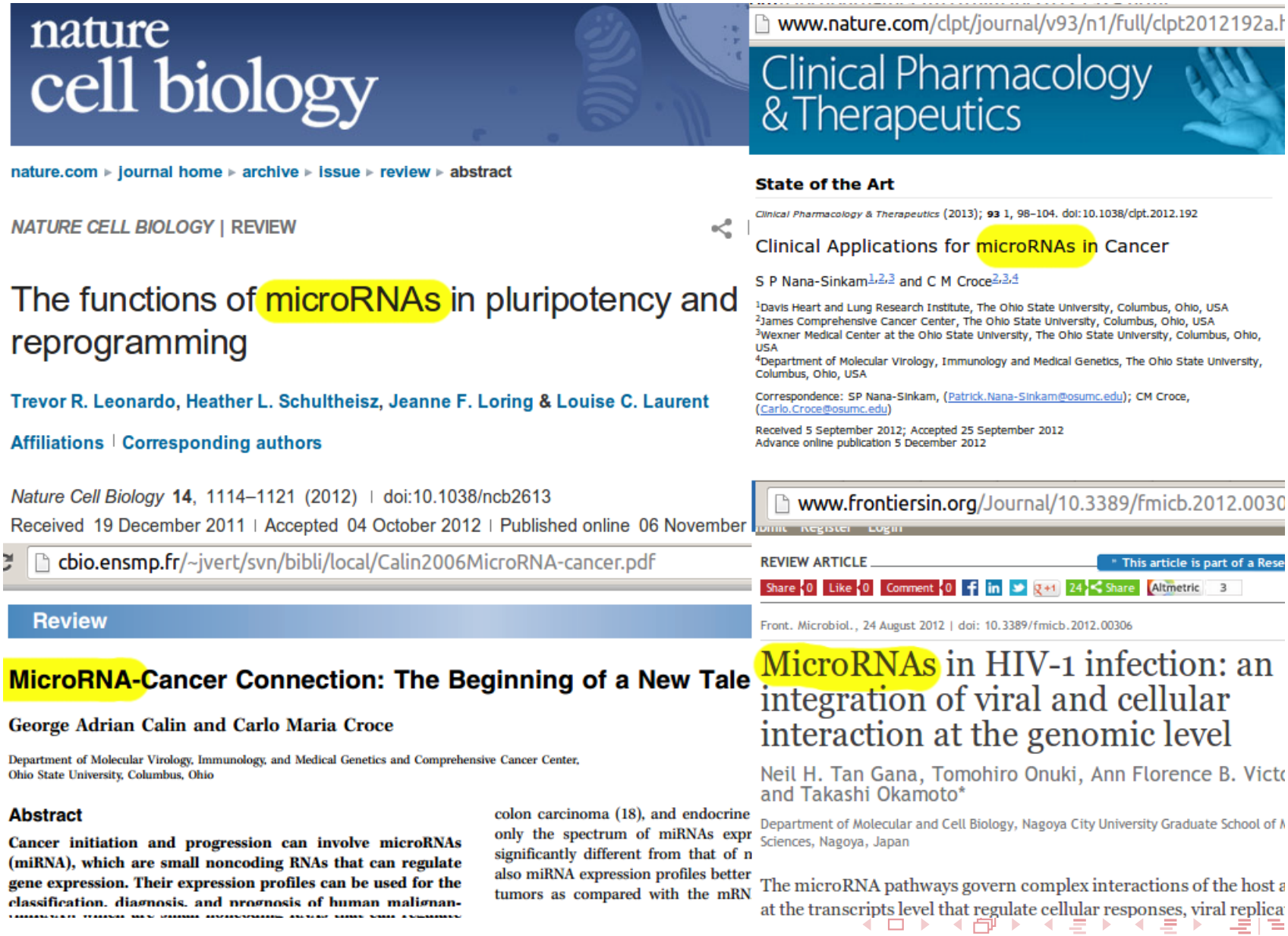
Dogma central - não previa ncRNAs



Tipos de RNA



http://www.nature.com/horizon/rna/background/pdf/microrna s.pdf



nature cell biology

nature.com ▸ journal home ▸ archive ▸ issue ▸ review ▸ abstract

NATURE CELL BIOLOGY | REVIEW

The functions of microRNAs in pluripotency and reprogramming

Trevor R. Leonardo, Heather L. Schultheisz, Jeanne F. Loring & Louise C. Laurent

Affiliations | Corresponding authors

Nature Cell Biology 14, 1114–1121 (2012) | doi:10.1038/ncb2613

Received 19 December 2011 | Accepted 04 October 2012 | Published online 06 November

www.nature.com/clpt/journal/v93/n1/full/clpt2012192a.h

Clinical Pharmacology & Therapeutics

State of the Art

Clinical Pharmacology & Therapeutics (2013); 93 1, 98–104. doi:10.1038/clpt.2012.192

Clinical Applications for microRNAs in Cancer

S P Nana-Sinkam[1,2,3] and C M Croce[2,3,4]

[1]Davis Heart and Lung Research Institute, The Ohio State University, Columbus, Ohio, USA
[2]James Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA
[3]Wexner Medical Center at the Ohio State University, The Ohio State University, Columbus, Ohio, USA
[4]Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, Ohio, USA

Correspondence: SP Nana-Sinkam, (Patrick.Nana-Sinkam@osumc.edu); CM Croce, (Carlo.Croce@osumc.edu)

Received 5 September 2012; Accepted 25 September 2012
Advance online publication 5 December 2012

cbio.ensmp.fr/~jvert/svn/bibli/local/Calin2006MicroRNA-cancer.pdf

Review

MicroRNA-Cancer Connection: The Beginning of a New Tale

George Adrian Calin and Carlo Maria Croce

Department of Molecular Virology, Immunology, and Medical Genetics and Comprehensive Cancer Center, Ohio State University, Columbus, Ohio

Abstract

Cancer initiation and progression can involve microRNAs (miRNA), which are small noncoding RNAs that can regulate gene expression. Their expression profiles can be used for the classification, diagnosis, and prognosis of human malignan-

colon carcinoma (18), and endocrine
only the spectrum of miRNAs expr
significantly different from that of n
also miRNA expression profiles better
tumors as compared with the mRN

www.frontiersin.org/Journal/10.3389/fmicb.2012.0030

REVIEW ARTICLE — This article is part of a Rese

Share 0 | Like 0 | Comment 0 | 24 Share | Altmetric 3

Front. Microbiol., 24 August 2012 | doi: 10.3389/fmicb.2012.00306

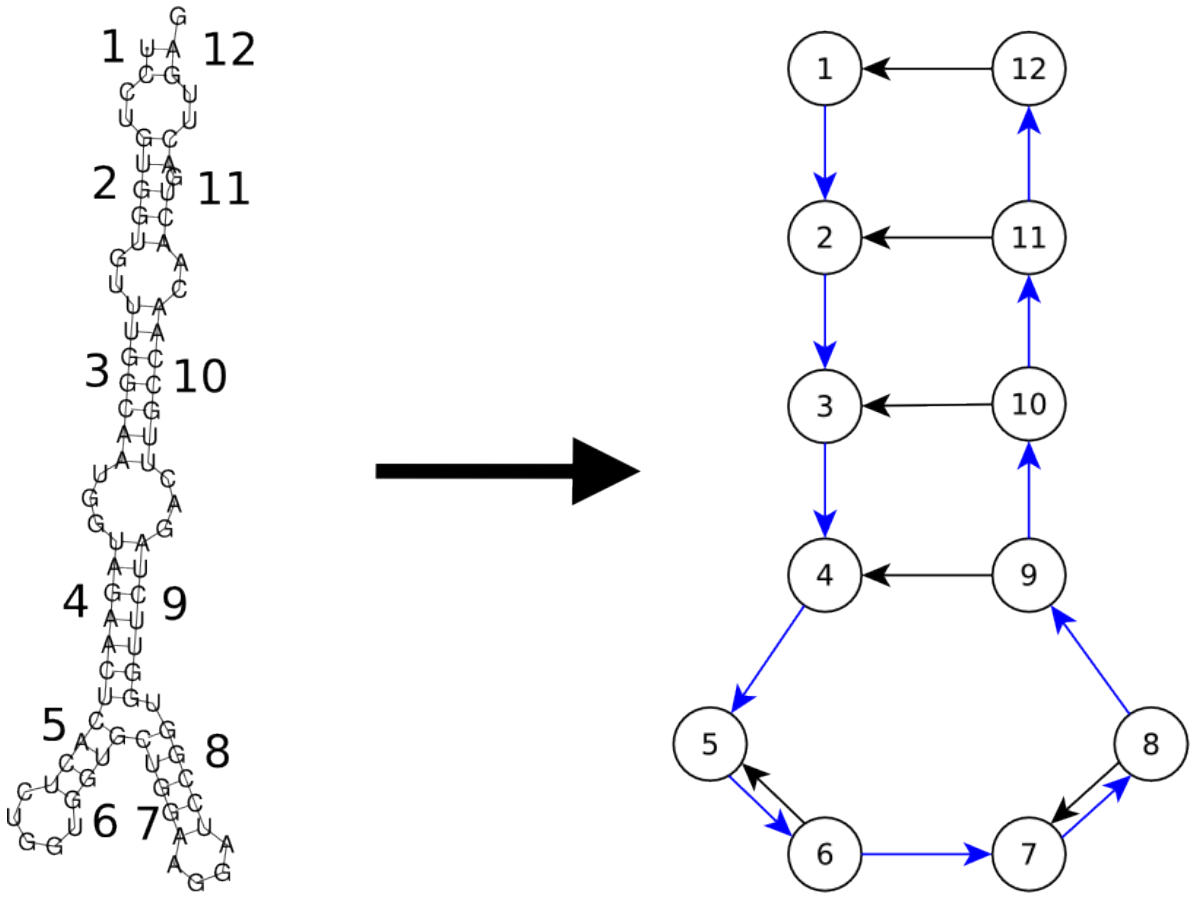
MicroRNAs in HIV-1 infection: an integration of viral and cellular interaction at the genomic level

Neil H. Tan Gana, Tomohiro Onuki, Ann Florence B. Victo and Takashi Okamoto*

Department of Molecular and Cell Biology, Nagoya City University Graduate School of M Sciences, Nagoya, Japan

The microRNA pathways govern complex interactions of the host a at the transcripts level that regulate cellular responses, viral replica

Como funciona o preditor?

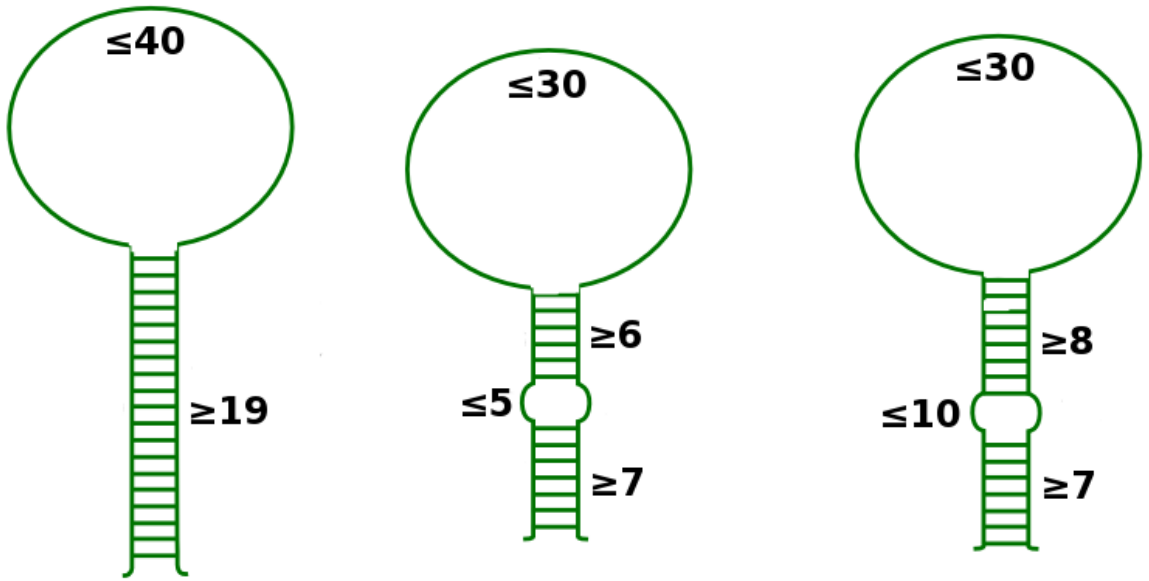


Arestas de i a j onde i $<$ j - 1 ocultadas por motivos estéticos.

Objetivos deste trabalho

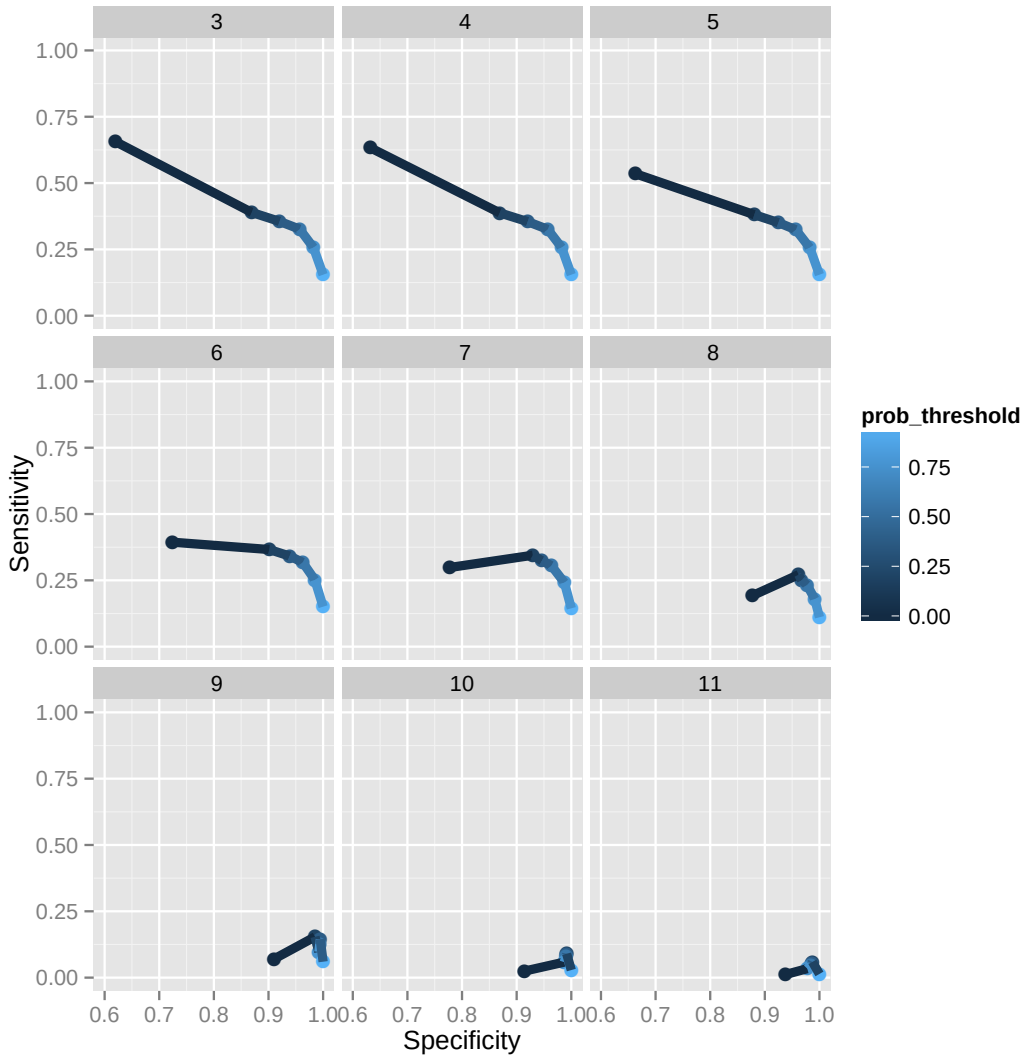
- Otimizar os parâmetros utilizados pelo preditor
- Fazer uma comparação justa com algum outro preditor de microRNAs
- Implementar uma extensão do preditor para realizar o treinamento automático

Parâmetros a serem otimizados



▸ Limiar de probabilidade	0,0	0,25	0,05
▸ Tamanho mínimo de hélice	2	4	6

Curvas ROC

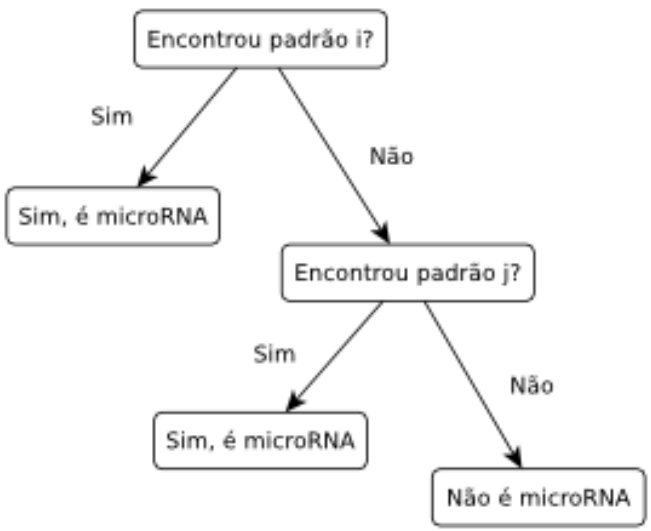

3
4
5
6
7
8
9
10
11
1.00
0.75
0.50
0.25
0.00
Sensitivity
0.6
0.7
0.8
0.9
1.0
Specificity
prob_threshold
0.75
0.50
0.25
0.00

Árvore de decisão

Sequência	Padrão 1	Padrão 2	...	Padrão N	microRNA real?
seqA	1	0	...	0	1
seqB	1	1	...	1	1
seqC	0	1	...	0	0
:	:	:	:	:	:

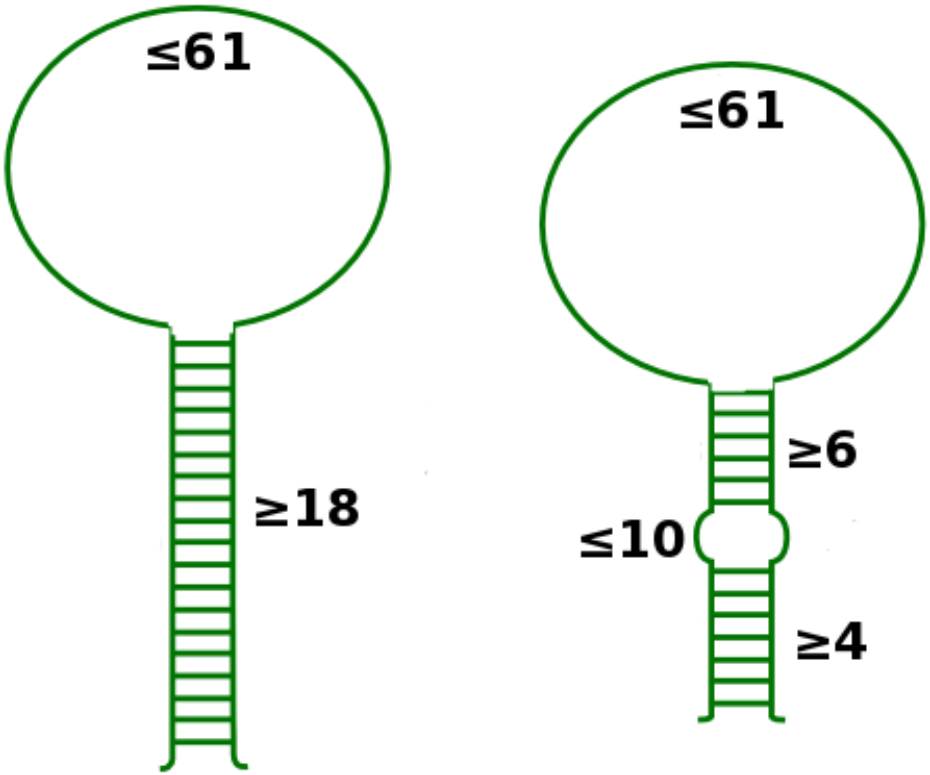
Árvore de decisão

Sequência	Padrão 1	Padrão 2	...	Padrão N	microRNA real?
seqA	1	0	...	0	1
seqB	1	1	...	1	1
seqC	0	1	...	0	0
:	:	:	:	:	:



$\equiv$ Padrão i v Padrão j

Padrões otimizados



Padrão I V **Padrão II**

Limiar de probabilidade: 0 Limiar de probabilidade: 0,57

Tamanho mínimo de hélice: 7 Tamanho mínimo de hélice: 5

Antes x Depois da otimização

Medida	Antes	Depois
Especificidade	87,538%	84,154%
Sensibilidade	71,692%	82,692%
Acurácia	79,615%	83,423%

$$\text{Especificidade} = TN/(TN + FN)$$
$$\text{Sensibilidade} = TP/(TP + FP)$$
$$\text{Acurácia} = (TP + TN)/(TP + FP + TN + FN)$$

Comparação entre os preditores

Medida	MiPred	Nosso preditor
Especificidade	88,076%	84,154%
Sensibilidade	80,615%	82,692%
Acurácia	84,346%	83,423%

Treinamento automático do preditor

- Feito encontrando subestruturas típicas em um conjunto de microRNAs
- Que foram encontradas utilizando um algoritmo de subgrafo maximal entre dois grafos
- Que, por sua vez, foi implementado utilizando uma redução ao problema do clique maximal

Redução ao problema do clique maximal

- **Entrada:** Grafos $G_1 =< V_1, A_1 > G_2 =< V_2, A_2 >$
- Grafo produto de arestas $H_e = G_1 \circ G_2$ possui como vértices o conjunto $V_H = E_1 \times E_2$
- Há uma aresta do nó $v = (v_1, v_2)$ ao $w = (w_1, w_2)$ em H_e sse $v_1 \neq w_1$ e $v_2 \neq w_2$ e (o nó que conecta as arestas v_1 e w_1 é compatível com o nó que conecta v_2 e w_2 OU v_1 e w_1 não são arestas adjacentes em G_1 e v_2 e w_2 também não são adjacentes em G_2)
- **Saída:** Clique maximal de H_e que corresponde a um subgrafo comum

Redução do problema ao problema do clique maximal

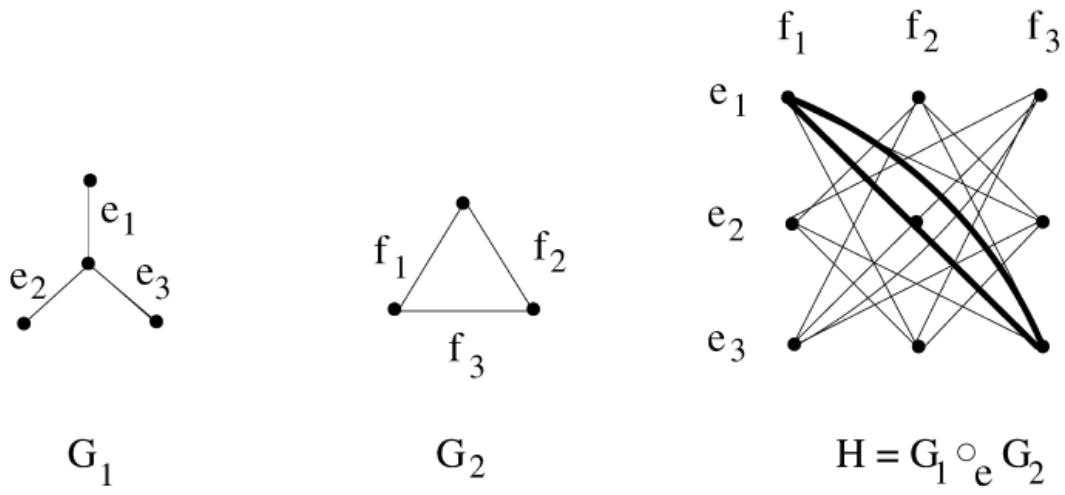
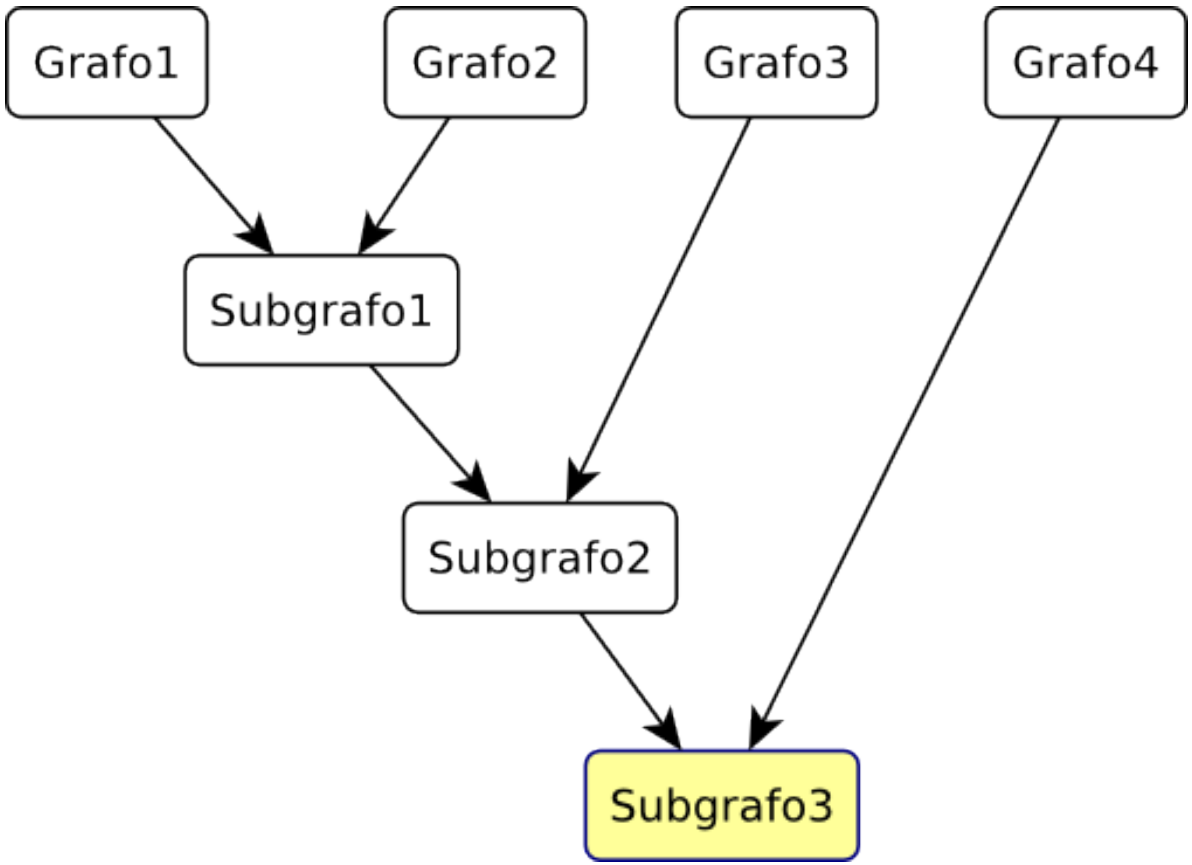


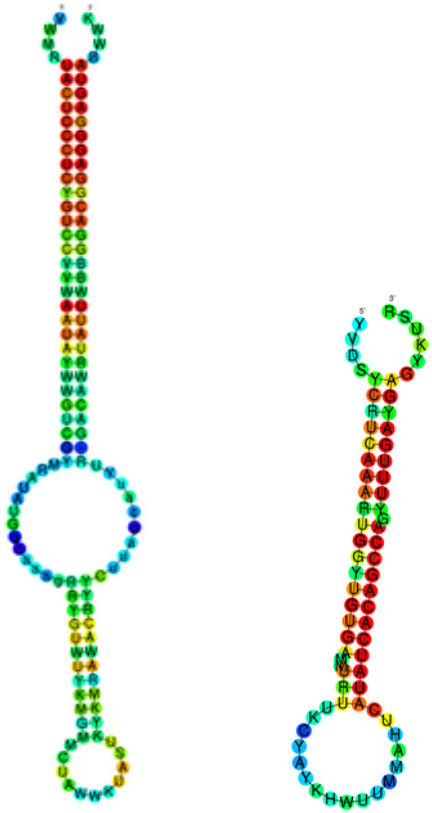
Fig. 2. An example in which cliques with three vertices do not necessarily define isomorphic subgraphs. Though the edge product graph H_e exibits a clique (bold edges) with the edge set $V_H = \{(e_1, f_1), (e_2, f_2), (e_3, f_3)\}$, G_1 (the Y-shape or triod) is not isomorphic to G_2 (the triangle).

Retirado do artigo de Koch de 2001.

Aplicando iterativamente



Família A x Família B



(a) (b)

http://rfam.sanger.ac.uk/

- O treinamento foi realizado em 21 sequências da família A, encontrando 4 subestruturas recorrentes
- 53,7% das sequências da família A foram identificadas com sucesso
- Apenas 2,7% das sequências da família B foram identificadas como sendo de A

Conclusões

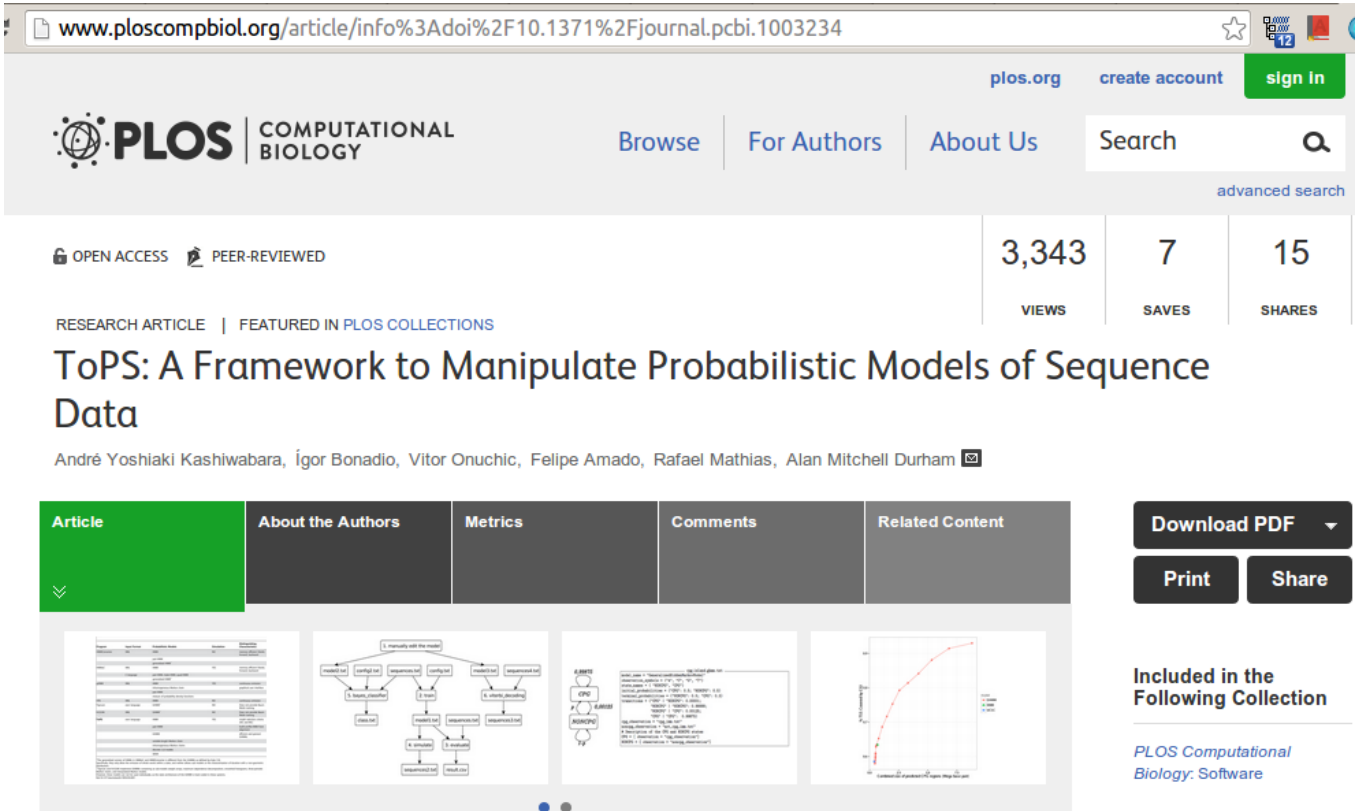
- Com a otimização dos parâmetros, obtivemos uma discreta melhora, mas suficiente para alcançar um nível de acurácia comparável a um dos melhores preditores de microRNA atualmente

Conclusões

- Com a otimização dos parâmetros, obtivemos uma discreta melhora, mas suficiente para alcançar um nível de acurácia comparável a um dos melhores preditores de microRNA atualmente
- Apesar de ter apresentado um bom resultado em um dos testes realizados, o algoritmo de treinamento automático ainda precisa de melhorias

O que mais me ajudou neste trabalho

- Disciplina de Algoritmo em grafos
- Disciplina de Biologia Computacional
- Disciplina de Data Mining (cursada durante intercâmbio)
- Iniciação científica



Dúvidas?

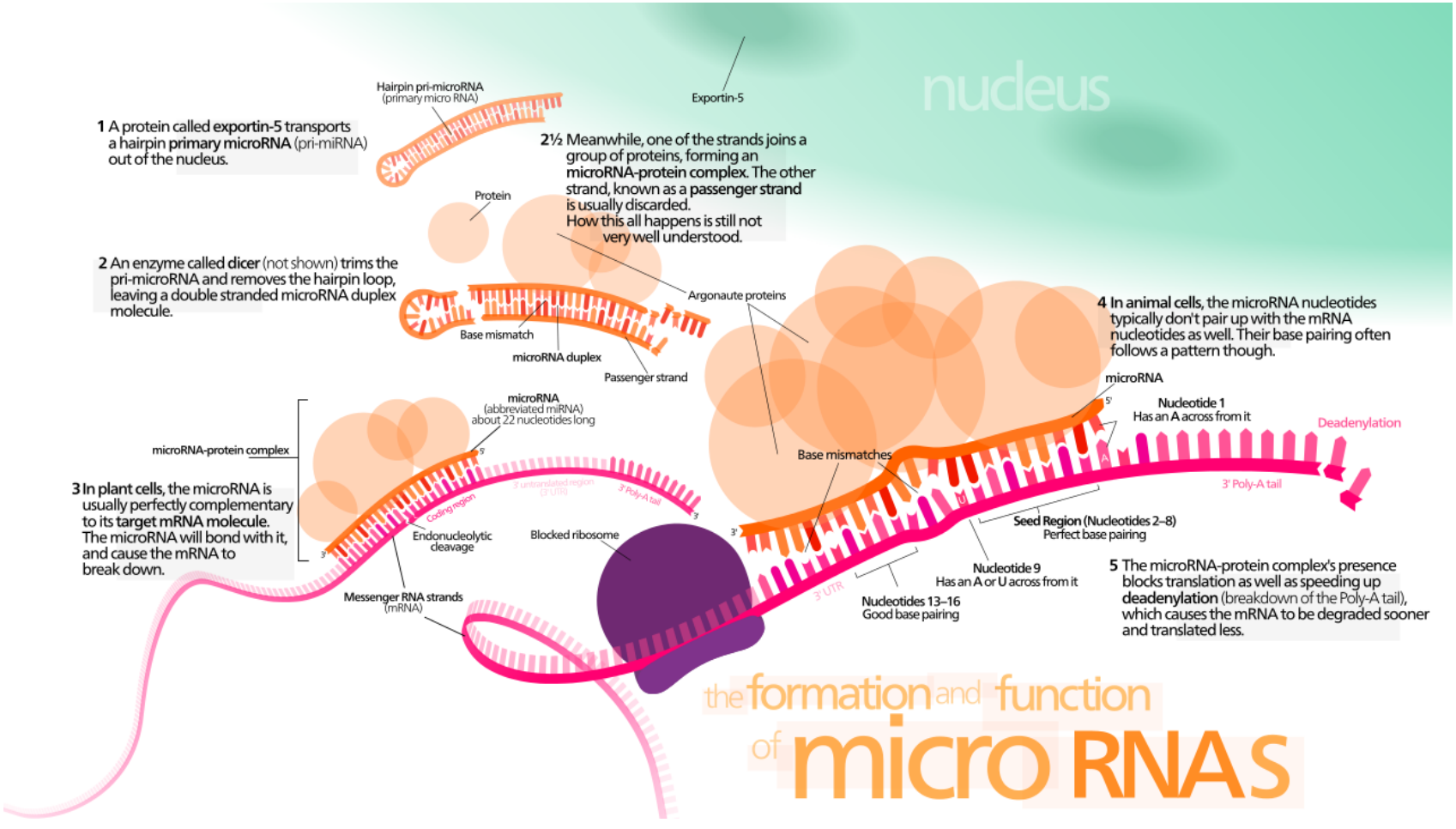
Bibliografia I

- Hu LL et al (2012). *Benchmark comparison of ab initio microRNA identification methods and software*. GMR.
- A Machado-Lima, HA del Portillo, AM Durham (2008). *Computational methods in noncoding RNA research.*. J Math Biol. 56(1-2):15-49
- Ina Koch (2001). *Enumerating all connected maximal common subgraphs in two graphs.* Elsevier.
- Sarah W. Burge et al (2013). *Rfam 11.0: 10 years of RNA families*. Nucleic Acids Res.
- Jiang P et al (2007). *MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features.* Nucleic Acids Res.

Bibliografia II

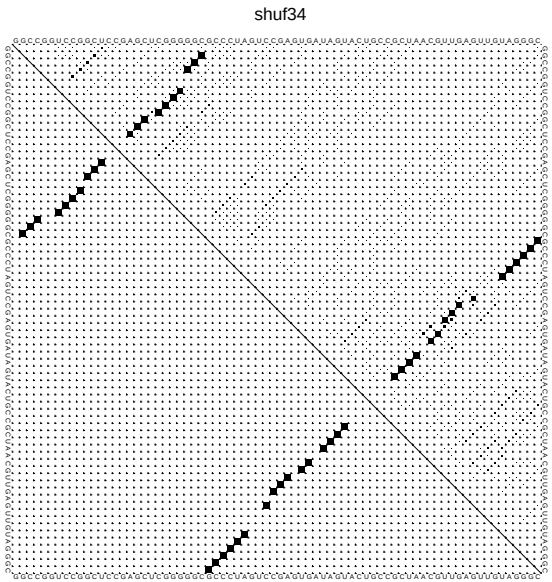
- Wei-Yin Loh and Yu-Shan Shih (1997). *Split Selection Methods For Classification Trees*. Statistica Sinica 7, 815-840
- Marilena V. Iorio et al (2005). *MicroRNA Gene Expression Deregulation in Human Breast Cancer*. AACR.

MicroRNA

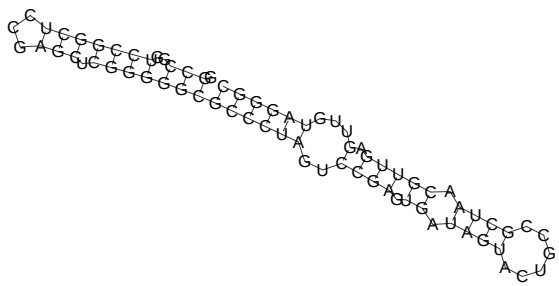


http://en.wikipedia.org/wiki/MicroRNA

RNAfold

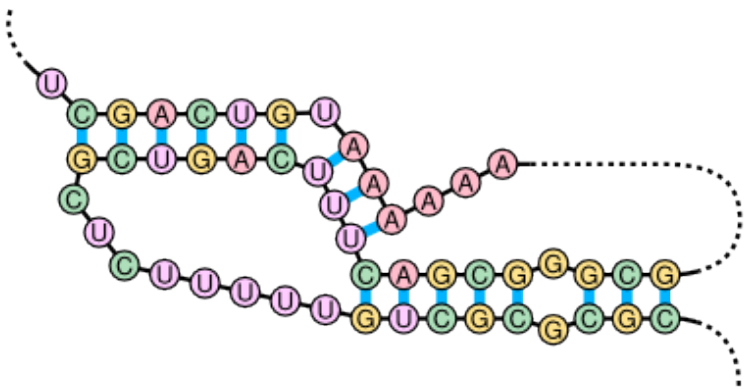


(c)



(d)

Pseudo-nó



(e)

http://en.wikipedia.org/wiki/File:Pseudoknot.svg