바이오 정보처리 연구 동향: 미세 RNA 분석을 중심으로

민혜영*, 윤성로**
*스탠퍼드대학교 통계학과
**고려대학교 전기전자전파공학부
e-mail : sryoon@korea.ac.kr

Bio Information Processing Trend: Deciphering microRNA Targets

Hyeyoung Min*, Sungroh Yoon**
*Department of Statistics, Stanford University, California, USA
**School of Electrical Engineering, Korea University, Seoul, Korea (Corresponding author)

요 약

기존의 실험을 통한 전통적인 생물학의 연구와는 달리, 미세 RNA (microRNA)의 연구에 있어 컴퓨터를 통한 프로그램 개발과 정보기술의 이용은 필수 불가결한 요소가 되었다. 컴퓨터를 바탕으로 한 대부분의 연구는 미세 RNA 를 발현하는 유전자와 미세 RNA 의 타겟 (target) 을 예측하는 두 가지 분야로 나누어 이루어지고 있다. 본 연구에서는 미세 RNA 의 타겟을 예측하는 프로그램 개발 시 이용되는 몇 가지 원칙들과 그 원칙들의 문제점을 서술하며, 현재 인터넷상에서 이용 가능한 프로그램들을 소개한다. 또한 컴퓨터를 통해 예측된 미세 RNA 의 타겟을 실험을 통해 검증하는 방법에 대해 논한다.

1. Introduction

MicroRNAs (miRNAs) are a class of small, non-coding regulatory RNAs that are important in post-transcriptional gene silencing (Bartel et al, 2004). They regulate gene expression by binding to 3' untranslated region (UTR) of their target mRNAs for cleavage or translational repression and play important roles in many biological processes.

Since the discovery of the very first miRNAs, computational approaches have been invaluable tools in understanding the biology of miRNAs (Bentwich et al., 2005; Rajewsky et al., 2006). Web-based-miRNA databases have been constructed and provided not only thousands of published miRNA sequences and annotation (miRBase Sequences) but also potential miRNA target genes (miRBase Targets). However, most computational approaches associated with miRNA research are miRNA gene detection and miRNA target prediction.

Researchers initially determined miRNA targets through experiments. However, due to the laborious nature of experiments and the absence of high-throughput experimental methods, it is inevitable to develop computational techniques to determine miRNA targets. In this paper, we summarize the principles to predict miRNAs and their targets, and introduce the currently available computational methods that have been developed for miRNA targets prediction.

2. Principles and concerns of miRNA target recognition

Target prediction and its biological validation have been major obstacles to miRNA researcher. Because miRNAs are short, and animal miRNAs have limited sequence complementarity to their targets, it is a challenging task to predict animal miRNA targets with high specificity.

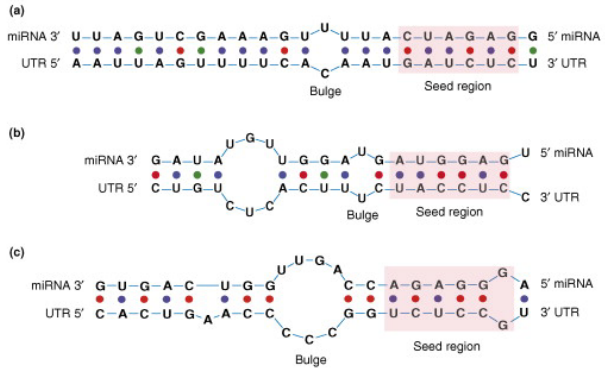
In order to develop computational algorithms identifying miRNA target genes, principles of miRNA target recognition are often established based on empirical evidences. For example, the importance of base pairing between miRNAs and their targets has been suspected according to the observation that the 'target site' of the lin-14 UTR is complementary to the 5' region of the lin-4 miRNA (Lee et al., 1993). Some features used by the mammalian target prediction programs are described below.

1) Base pairing pattern

Most target prediction programs identify potential binding sites according to specific pairing patterns. The binding sites can be classified into 3 categories (Maziere et al, 2007): (i) 5'-dominant canonical, (ii) 5'-dominant seed only and (iii) 3'-compensatory (Figure 1). MiRNA seed is defined as the consecutive 7 to 8 nt sequence starting from either the first or second base at the 5' end of an miRNA (Lewis et al, 2003). The 5' -dominant canonical sites have perfect base paring to the 5' end seed region and extensive base pairing to the 3' end of the miRNA with a bulge in the middle. The 5'-dominant seed only sites have perfect base pairing to the seed region and limited base pairing to the 3' end of the miRNA. The 3'-compensatory sites have a mismatch or wobble in the seed region of the miRNA, but have extensive base pairing to the 3' end of the miRNA to compensate for the weak binding at the 5' seed (Brennecke et al., 2005). However, the problem of using 5' dominant site is that 3' compensatory site having a mismatch or wobble in the seed region cannot be detected by most target prediction methods. Accordingly, it is of necessity to develop more computational algorithm to identify those 3' compensatory target sites with accuracy.

2) Thermodynamic stability of miRNA–mRNA duplexes

The thermodynamic properties of miRNA–mRNA duplex formation are analyzed by calculation of free energy (ΔG). The estimate free energy and secondary structure of the miRNA-mRNA duplex can be calculated by RNA folding program such as Vienna package (Wuchty et al., 1999). The free energy threshold is then calculated based on specificity and sensitivity. However, the thermodynamics can be removed without lowering the specificity of the algorithm by incorporating evolutionary conservation as an informational filter (Lewis et al., 2005).



(Figure 1) Approximate secondary structures of the three main types of target site duplex. (a) Canonical sites (b) Dominant seed sites (c) Compensatory sites (Adapted from Maziere et al., 2007)

3) Comparative sequence analysis for conservation

Cross-species sequence comparison is used to ask whether the target sequence has been evolutionarily conserved between related species. In order to reduce the number of false positives, many of the target prediction algorithms identify orthologous 3' UTR sequences and then perform conservation analysis across species. However, the use of conservation filter has a risk of increasing false negatives. Farh et al. (2005) demonstrated that many of the nonconserved target sites, which outnumber the conserved sites 10 to 1, are also functional and mediate repression. Thus, the presence of those nonconserved target sites should not be overlooked when designing an algorithm for target prediction.

Different methods have been developed for computational target prediction. Some of the currently available target prediction programs are listed in Tables 1

<Table 1> Computational methods for miRNA target prediction

Name	URL	Reference(s)
DIANA-microT	http://diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi	Kiriakidou *et al.*, 2004
GUUGle	http://bibiserv.techfak.uni-bielefeld.de/guugle	Gerlach *et al.*, 2006
miRanda	http://www.microrna.org/	Enright *et al.*, 2003
miTarget	http://cbit.snu.ac.kr/miTarget	Kim *et al.*, 2006
PicTar	http://pictar.bio.nyu.edu	Grun *et al.*, 2005
rna22	http://cbcsrv.watson.ibm.com/rna22.html	Miranda *et al.*, 2006
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid	Rehmsmeier *et al.*, 2004
TargetScan	http://genes.mit.edu/targetscan/	Lewis *et al.*, 2003
TargetScanS	http://genes.mit.edu/targetscan/	Lewis *et al.*, 2005

3. Experimental validation

Once miRNA targets are predicted with a fair degree of accuracy, the next step is to experimentally validate the miRNA – target interaction. Since computational methods are not perfect, and there is a risk of false-positive prediction, target validation in biological system is inevitable to complete the study of target prediction. Reporter assay is the most common method to check the interaction between miRNA and its target mRNA. Then, Northern blot analysis, quantitative real-time PCR (qPCR), or *in situ* hybridization is often performed to examine the co-expression of predicted miRNA and mRNA target gene. For thorough study, biological function can be examined through 'gain of function' or 'loss of function' experiment under *in vitro* or *in vivo* condition. However, those biological or biochemical experiments (even the reporter assay) are laborious, time-consuming, and expensive to deal with many pairs of miRNAs and their targets. Therefore, high-throughput experimental strategies should be developed for large-scale analysis of miRNA targets and their biological function.

References

[1] Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.
[2] Bentwich, I. (2005). Prediction and validation of microRNAs and their targets. FEBS Lett. 579, 5904–5910,
[3] Brennecke, J. et al., "Principles of microRNA- target recognition," *PLoS Biol.* 3, e85. 2005.
[4] Enright, A. J. et al., (2003). MicroRNA targets in Drosophila. Genome Biol. 5, R1.
[5] Farh, K. K., Grimson A., Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science 310, 1817 - 1821
[6] Gerlach, W., and Giegerich, R. (2006). GUUGle: a utility for fast exact matching under RNA complementary rules including G-U base pairing. Bioinformatics 22, 762–764.
[7] Kim, S. K., et al., (2006). miTarget: microRNA target-gene prediction using a support vector machine. BMC Bioinform. 7, 411.
[8] Lee, R. C. et al., (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843-854.
[9] Lewis, B. P. et al., (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120,15–20.
[10] Lewis, B. P. et al., (2003). Prediction of mammalian microRNA targets. Cell 115, 787–798.
[11] Maziere, P. et al., (2007). Prediction of microRNA targets. Drug Discov. Today. 12, 452 – 458.
[12] Miranda, K. C. et al., (2006). A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. Cell 126, 1203–1217.
[13] Rajewsky, N. (2006). microRNA target predictions in animals. Nat. Genet. 38(Suppl.), S8–S13.
[14] Rehmsmeier, M. et al., (2004). Fast and effective prediction of microRNA/target duplexes. RNA 10, 1507–1517.
[15] Wuchty, S et al., (1999). Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers 49, 145–165.