Computational Analysis of microRNA Recognition Site

Keishin Nishida¹
keishin@bio.c.u-tokyo.ac.jp

Riu Yamashita²
ryamasi@hgc.jp

Kengo Kinoshita²
kino@ims.u-tokyo.ac.jp

Kenta Nakai²
knakai@ims.u-tokyo.ac.jp

¹ Graduate School of Arts and Science, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
² Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

Keywords: microRNA, microarray

1 Introduction

MicroRNAs (miRNAs) are noncoding RNAs of about 22 nucleotides that suppress translation of target genes by binding to their mRNA, and thus have a role in gene regulation [1-3]. Recently, Lim et al. [1] transfected miRNAs into human cells and used microarrays to examine changes in the messenger RNA profile. These microarray profiles indicate that the 3’-untranslated regions of down-regulated messages have a significant propensity to pair to the 5’-region of the miRNAs, as expected if many of these messages are the direct targets of the miRNAs. This research, however, did not mention exact position of indispensable nucleotides for mRNA degradation in miRNAs. Here we present a visualization method to discover out the miRNAs recognition site by computational analysis, and report the location of miRNAs recognition site and its required length for mRNA degradation.

2 Material and Method

The sequence data of miRNAs were obtained from Lim et al. [1], and the mRNA sequence data were obtained from the Ensembl database [9]. The complementation at each position of mRNA sequence by a miRNA was considered. The length and start position within the miRNA sequence of any nucleotide stretch longer than 2 bases were recorded. We thereby obtained length vs position count matrices for each miRNA-mRNA pair. Matrices obtained with the same miRNA were added together, yielding general miRNA matrices.

The microarray data were obtained from NCBI GEO (GSE2075) [10]. miRNA specific gene expression profiles were linked to Ensembl transcripts based on the mapping of microarray oligo nucleotides on Ensembl mRNAs. The linked expression profiles that failed the assay quality control were removed, and the linked profile with the highest $p$-value was finally assigned to the corresponding gene. For each miRNAs, the length vs position count matrices of mRNA corresponding to downregulated genes with $p$-values smaller than 0.001 were summed, yielding downregulation miRNA matrices. The ratio of the downregulation matrix to the general matrix was calculated for each miRNA. Figure 1 illustrates the obtained position and length specific ratios, the dot color being set based on the calculated ratio for a fine position-length pair.

3 Results and Discussion

Using the 5’-untranslated region and the coding region, no significant results were obtained (data not shown). On the other hand, as shown for example in figure 1, the 3’-untranslated region showed significant contrasted dots for lengths 6 to 9 at the 5’ terminal position. This indicates that the first 6 to 9 nucleotides from the 5’ end of the miRNA are important for the mRNA recognition and are likely to be part of the miRNA recognition site. Our visualization of length vs position count matrices is fruitful to find out the required length and the important position in miRNA, and allowed to identify the 3’-untranslated region of miRNA as the region recognized by miRNAs.
Figure 1: An example of length and position count matrix plot for a 3'-untranslated region. The vertical axis indicates the length of the complementing stretch. Upper left diagonal axis indicates the start position of the complement stretch from the 5’ end of miRNA. This plot includes two matrices, miRNA-1 and miRNA-124. The length and position matrix of miRNA-1 uses red intensities, that of miRNA-124 uses green intensities. Therefore yellow indicates that both values are represented.

References