

Identification of microRNAs with regulatory potential using a matched microRNA-mRNA timecourse data

Vivek Jayaswal¹
vivek@maths.usyd.edu.au
Mark Lutherborrow²
m.lutherborrow@garvan.org.au

Yee Hwa Yang¹
jeany@maths.usyd.edu.au
David Ma²
d.ma@garvan.org.au

- ¹ School of Mathematics and Statistics, University of Sydney, Sydney, NSW, Australia
² Blood Stem Cell and Cancer Research Unit, St Vincent's Hospital Applied Biomedical Research, Darlinghurst, NSW, Australia

Abstract

We present a novel method for identifying regulatory microRNAs using matched microRNA-mRNA timecourse data. Our method also predicts target mRNAs and identifies the time lag between changes in microRNA and its target-mRNA's expression. We applied our method to a cancer data set and identified a few regulatory microRNAs, some of which have been experimentally validated.

Keywords: microRNA, microarray, timecourse, odds-ratio, multiple myeloma

1 Introduction

MicroRNAs (miRNAs) are small (~ 22 nucleotides) RNA molecules that play an important role in gene regulation by (i) repressing the translation of mRNA or (ii) cleaving the mRNA transcript. Over the past few years, some miRNAs have been shown to regulate the expression of cancer genes in humans.

We propose an odds-ratio (OR) statistic for the identification of miRNAs that potentially regulate gene expression. This statistic can also be used for – (i) identification of time lag between a change in miRNA expression and that of its target mRNAs, and (ii) ranking of regulatory miRNAs after combining results obtained from multiple miRNA target-prediction algorithms. These features are useful as currently there is no golden rule for the time lag between changes in miRNA and target-mRNA expression. Also, the miRNA target-prediction accuracy varies from one algorithm to another with no one algorithm being distinctly better. We illustrate our method using a longitudinal time course multiple myeloma data set.

2 Methods

We obtain t-statistics for the null hypothesis $H_0: \mu_{tg} = \mu_{0g}$ vs. the alternate hypothesis $H_1: \mu_{tg} \neq \mu_{0g}$, where μ_{tg} is the average expression of mRNA g at time point t and μ_{0g} is average expression at time point 0. Let $\mathbf{M}_g = [m_1, \dots, m_k]^T$ denote the classification of t-statistics as up-regulated (+1), down-regulated (-1) or not significant (0) for the g^{th} mRNA, where k is the number of time points excluding time point 0. Henceforth, we refer to \mathbf{M}_g as the discretized expression profile for mRNA g . Similarly, we obtain the discretized expression profiles for miRNAs.

We measure the association between a miRNA and its predicted target mRNAs using an OR-statistic. The OR-statistic is calculated using matched discretized expression profiles of miRNA and mRNA (explained later in Section 3). We test the null hypothesis that $OR = 1$ (i.e., a change in the expression of the target mRNAs is independent of a change in the miRNA's expression) using a chi-squared test with one degree of freedom. Since different miRNA-target-prediction algorithms return different targets for the same miRNA, the OR-statistics are algorithm-dependent. We propose combining the p-values returned by different algorithms using Fisher's combined test [1] to obtain the overall ranking of miRNAs; the higher ranked miRNAs being more likely to have regulatory potential. Finally, for the miRNAs with regulatory potential, we determine the odds of an mRNA being negatively

correlated to the relevant miRNA.

3 Results

We used our model to analyze miRNA and mRNA data generated from microarray assays performed on an in-vitro drug study for multiple myeloma. It had six time points - 0, 2 hrs, 4 hrs, 8 hrs, 24 hrs, and 48 hrs, with two replicates per time point for both miRNA and mRNA. The data was pre-processed and normalized [2] before discretization. The discretized data was used to (i) identify miRNAs with regulatory potential using the OR-statistic, (ii) integrate results obtained using different miRNA target-prediction algorithms, and (iii) identify mRNAs that are negatively correlated to the relevant miRNAs.

Identification of time lag: Since we were interested in miRNAs that regulate mRNA expression, we assumed that a change in target-mRNA's expression may occur with a time lag. Therefore, for Step (i) we considered five different time lags (Table 1) and obtained the OR-statistic for each miRNA-time lag combination. For example, the calculation of OR-statistic for Time lag 1 was based on whether a change in miRNA expression at time points 2, 4, 8, and 24 hrs produced a change in target mRNA's expression at time points 4, 8, 24, and 48 hrs, respectively.

Table 1. Different time lags for changes in miRNA and mRNA expressions. The numbers correspond to the matched miRNA-mRNA time points for various time lags – (i) 0 -> Time Lag 0, (ii) 1 -> Time Lag 1, (iii) 2 -> Time Lag 2, (iv) 3 -> Time Lag 3, and (v) 4 -> Time Lag 4

		miRNA time points				
		2 hrs	4 hrs	8 hrs	24 hrs	48 hrs
mRNA time points	2 hrs	0				
	4 hrs	1	0			
	8 hrs	2	1	0		
	24 hrs	3	2	1	0	
	48 hrs	4	3	2	1	0

We obtained 20 miRNAs with regulatory potential and these corresponded to 33 miRNA-time lag combinations. We observed that some miRNAs had regulatory potential for several time lags.

4 Discussion

A literature search revealed that some of the miRNAs identified by our method are oncogenic. This indicates that the proposed analytical method can be helpful in dissecting complex biological information generated from microarrays.

Acknowledgment

We thank Tiffany Khong and Andrew Spencer for providing us with the samples for microarray analysis.

References

- [1] Fisher, R. *Statistical Methods for Research Workers*. Oliver and Boyd, London, 1932.
- [2] Smyth, G.K. and Speed, T. Normalization of cDNA microarray data. *Methods*, 31: 265-273, 2003.