Inference of Gene Expression Regulation via microRNA Transfection

Y-h. Taguchi[1] and Jun Yasuda[2]

[1] Department of Physics, Chuo University, Tokyo 112-8551, Japan
tag@granular.com
http://www.granular.com/tag/
[2] COE Fellow, Graduate School of Medicine, Tohoku University,
Sendai 980-8575, Japan
jun.yasuda@jfcr.or.jp

Abstract. How each microRNA regulates gene expression is unknown problem. Especially, which gene is targeted by each microRNA is mainly depicted via computational method, typically without biological/ experimental validations. In this paper, we propose a computational method to detect gene expression regulation via miRNAs by the use of expression profile data and miRNA target prediction. This method is tested to miRNA transfection experiments to tumor cells and succeeded in inference of transfected miRNA.

Keywords: microRNA, target genes, tumor, computational inference.

1 Introduction

MicroRNAs (miRNAs) are post-transcriptional regulators of gene expression. It binds to target messenger RNAs (mRNAs) through complementary sequences in the three prime untranslated regions (3 UTRs) of the mRNA, and consequently suppresses the expression of the mRNAs. miRNAs are short RNA molecules, on average only 22 nucleotides long, and abundant in many human cell types. The human genome may encode over 1,000 miRNAs, and the coverage by all possible miRNAs may be about 60 % of mammalian genes.

On the other hand, how a miRNA regulates its target genes and which genes are regulated by a miRNA is unclear. Especially, the later is mainly depicted via computational prediction[1], without any biological/experimental validations. There are some direct ways to investigate the bindings of mRNAs to the miRNA-protein complexes, e.g., HITS-CLIP[2] but capability of these methods for identification of miRNA-mRNA relationship is limited since it is unlikely that all the potential target mRNAs for a miRNA simultaneously express in a cell.

Another experimental way to detect miRNA target genes is to analyze the difference of gene expression profiles with or without the transfection of the miRNA to a cell line. However, it is unrealistic to test all the miRNAs with this method because it is time and money consuming.

D.-S. Huang et al. (Eds.): ICIC 2010, LNCS 6215, pp. 672–679, 2010.

In this paper, we describe a computational method to detect miRNAs which regulate the transcriptomes in a cell in response to extracellular stimuli by analyzing the difference of gene expression profiles and computational miRNA target predictions. The validation of our methods were obtained through the analyses of the gene expression profiles with or without the transfection of single miRNAs and our algorithm can quite frequently predict the transfected miRNA.

2 Materials and Methods

2.1 Gene Expression Data for Transfection Experiment

We have downloaded transfection experiment[3] data set, CBX79, which is deposited at CIBEX data base[4] at Center for Information Biology and DNA Data Bank of Japan (DDBJ), National Institute of Genetics (Mishima, Japan). It includes two biological replicates of negative, mir-107, 185, and let-7a trasfection experiments, one day and three days after the transfection. Expression of 45015 genes (probes) are listed. Since our method is robust for the random noise of gene expression variance and the overall distribution of gene expression between technical replicates should be within the acceptable range, we did not apply any normalization procedure.

2.2 Inference of miRNA Which Regulates Target Genes Significantly

The way to detect miRNA whose target genes are significantly differently expressed between negative control and treated one is as follows. First, we have downloaded a list of conserved seed match in 3' UTRs of genes to each miRNA[1][5]. This includes 162 miRNA families. The reason why we do not use major target gene list, e.g., targetScan[6], PITA[7], pictar[8], miranda[9], and others, but use seed match is because Alexiou *et al*[10] recently reported that simple seed match often outperforms more complicated estimations of target genes. Then we have picked up genes which has at least one seed match for any miRNAs in those 3' UTRs. Then, 13270 genes remain.

Hereafter, we denote a set of these genes as G. Next, for each miRNA, m, we have listed genes which has at least one seed match in 3' UTRs. We denote this set of genes as G_m, where m denotes one of miRNA families. Also we define a set of genes, $G'_m \equiv G \setminus G_m$, which is a set of genes included into G, but not into G_m. After denoting expression of gene g under transfection of miRNA m_0, m_0 is one of mir-107, 185, let-7a, and Negative Control (NC), as $x_g^{m_0}$, we compute gene expression difference between post-miRNA transfection and NC,

$$\Delta x_g^{m_0} \equiv \log x_g^{m_0} - \log x_g^{NC}.$$

Then we apply two way t-test beween $\{\Delta x_g^{m_0} \mid g \in G_m\}$ and $\{\Delta x_g^{m_0} \mid g \in G'_m\}$. P-value, P_m, is computed for each miRNA, m. After applying FDR correction

[1] http://hollywood.mit.edu/targetrank/hsa_conserved_miR_family_ranked_targets.txt

(BH method[11]) to 162 P-values, we have selected ms whose FDR corrected P-value is less than 0.05 as miRNA which regulates target genes significantly. For t-test, we have used t.test module in base package in R[12].

2.3 Coincidence between Biological Replicates

We have also checked if two biological replicates satisfy reproducibility in three ways. Firstly, we employed Pearson correlation coefficients between log transformed P_ms and secondly, Spearman correlation coefficients between them. P-values for these are computed as well as 95 percentile significant interval for the form. Thirdly, we analyzed coincidence between significant miRNAs, ms between two biological replicates. If the first(second) replicates have $m_1(m_2)$ significant miRNAs and m_{12} miRNAs are selected for both replicates, P-value computed by binomial distribution $P(m_1,\mathrm{N},m_2/N)$ or $P(m_2,\mathrm{N},m_1/N)$, where $P(x,N,p)$ is the probability that x among N is selected when the probability of selection is p. N is the number of genes in G.

We have used cor.test module in base package of R for P-values of correlation coefficients and pbinom module for binomial distribution.

2.4 Significant Overlap between Target Genes

To compute P-values of accidental agreement between target genes, $P^O_{m,m'}$ of miRNAs m and m', we have employed binomial distribution $P(n_{mm'}, n_m, n_{m'}/N)$, where $n_{mm'}$ is the number of co-target genes, $n_m(n_{m'})$ is the number of target genes of m(m').

3 Results

Independent of conditions, i.e., date and transfected miRNA, our method almost always gets non-empty set of significant miRNAs, ms (see Table 1). Thus, in principle, our method can detect miRNA regulation of gene expression. Table 2 shows which miRNA significantly regulates target genes (full list is available as a supplementary material[13]). Most remarkably, P_m has the strong tendency to become smallest when $m = m_0$, especially for one day after transfection. Thus, our method has not only ability to detect miRNA regulation of genes, but also that to infer transfected miRNA correctly, as appearing the transfected one in the highest rank (first or second) for most of the present analyses (Table 1).

In Figs. 1, we have shown comparison of gene expression between $x_g^{m_0}$ and x_g^{NC}. It is generally clear that $x_g^{m_0} < x_g^{NC}$ only for $g \in G_{m_0}$, but not for $g \in G'_{m_0}$. Only exception is mir-185 transfection, replicate 2 where we have failed transfected miRNA correctly (see Table 2). Clearly, for this case, $x_g^{m_0} < x_g^{NC}$ stands for both $g \in G_{m_0}$ and $g \in G'_{m_0}$ (lower row in Fig. 1(b)).

Table 3 shows the results of several statistical tests for the coincidence between biological replicates. For five out of six cases, at least one of tests give the significant P-values < 0.05. Thus, biological replicates are good enough for inference of miRNA transfection.

Table 1. Numbers of significant miRNAs. The ranks of transfected micriRNAs are shown in square brackets.

	Transfected miRNA					
	mir-107		mir-185		let-7a	
time	replicate 1	replicate 2	replicate 1	replicate 2	replicate 1	replicate 2
day 1	25[1st]	36[2nd]	12[1st]	12[—]	2[1st]	2[1st]
day 3	60[17th]	98[—]	0[1st]	24[—]	1[1st]	33[8th]

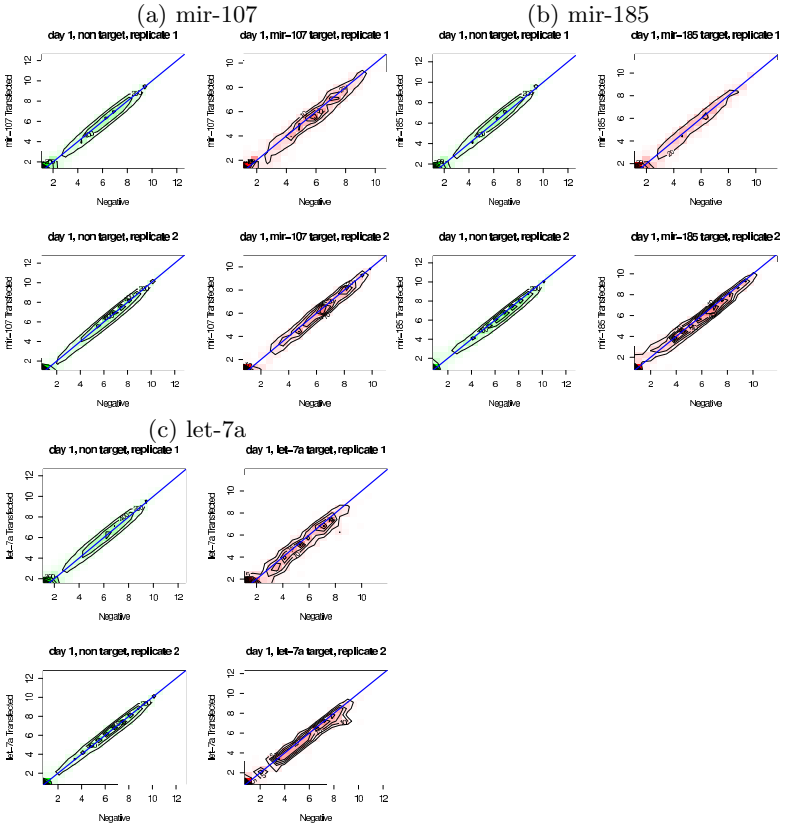

Fig. 1. Comparison between gene expression at one day after transfection (vertical axes) and negative control (horizontal axes). (a) mir-107, (b) mir-185, (c) let-7a . Right column (red) in each block indicates target genes ($g \in G_{m_0}$) and left column (green) in each block is others ($g \in G'_{m_0}$). The upper(lower) row in each block is replicate 1(2). Blue lines are diagonal, i.e., there are no difference between transfeted one and negative control.

Table 2. Top most significant miRNAs, one day or three days after trasnfection. Bold characters are those transfected.

mir-107 transfection, day 1			
replicate 1		replicate 2	
miRNA	P_m	miRNA	P_m
miR-103/107	2.64×10^{-8}	miR-29	1.34×10^{-10}
miR-30-5p	6.43×10^{-7}	**miR-103/107**	9.92×10^{-9}
miR-17-5p/20/93.mr/106/519.d	3.42×10^{-5}	miR-30-5p	2.71×10^{-8}
miR-452	3.73×10^{-5}	miR-17-5p/20/93.mr/106/519.d	3.45×10^{-7}
miR-25/32/92/363/367	2.49×10^{-4}	miR-129-5p	5.66×10^{-7}
mir-107 transfection, day 3			
replicate 1		replicate 2	
miRNA	P_m	miRNA	P_m
miR-17-5p/20/93.mr/106/519.d	1.31×10^{-11}	miR-17-5p/20/93.mr/106/519.d	4.39×10^{-15}
miR-29	8.42×10^{-9}	miR-186	3.27×10^{-13}
miR-25/32/92/363/367	1.07×10^{-7}	miR-203.1	6.56×10^{-13}
miR-181	1.44×10^{-7}	miR-374	2.46×10^{-12}
miR-448	2.30×10^{-7}	miR-369-3p	3.60×10^{-12}
mir-185 transfection, day 1			
replicate 1		replicate 2	
miRNA	P_m	miRNA	P_m
miR-185	8.64×10^{-12}	miR-496	1.87×10^{-7}
miR-326	3.15×10^{-6}	miR-495	1.96×10^{-5}
miR-15/16/195/424/497	1.60×10^{-4}	miR-320	1.98×10^{-5}
miR-410	2.01×10^{-4}	miR-181	2.52×10^{-5}
miR-491	3.98×10^{-4}	miR-381	3.93×10^{-5}
mir-185 transfection, day 3			
replicate 1		replicate 2	
miRNA	P_m	miRNA	P_m
—	—	miR-17-5p/20/93.mr/106/519.d	4.39×10^{-15}
—	—	miR-186	3.27×10^{-13}
—	—	miR-203.1	6.56×10^{-13}
—	—	miR-374	2.46×10^{-12}
—	—	miR-369-3p	3.60×10^{-12}
let-7a transfection, day 1			
replicate 1		replicate 2	
miRNA	P_m	miRNA	P_m
let-7/98	4.37×10^{-14}	**let-7/98**	9.55×10^{-16}
miR-196	3.94×10^{-4}	miR-196	2.81×10^{-4}
let-7a transfection, day 3			
replicate 1		replicate 2	
miRNA	P_m	miRNA	P_m
let-7/98	5.32×10^{-8}	miR-374	1.97×10^{-8}
—	—	miR-130/301	2.67×10^{-6}
—	—	miR-9	2.18×10^{-5}
—	—	miR-223	4.09×10^{-5}
—	—	miR-369-3p	4.96×10^{-5}

Table 3. Comparison of two biological replicates. Bold numbers indicate signifcant P-values (< 0.05). Bold asterisks (*****) indicate $P < 2.2 \times 10^{-16}$.

Transfection	mir-107		mir-185		let-7a	
Time	day 1	day 3	day 1	day 3	day 1	day 3
Pearson	0.67	0.67	0.03	0.16	0.86	0.28
	95 % confidence interval					
lower	0.57	0.58	-0.12	0.00	0.81	0.13
upper	0.74	0.75	0.18	0.30	0.89	0.42
P-value	*	*	0.7	**0.046**	*	**0.00029**
Spearman	0.52	0.70	0.05	0.09	0.28	0.13
P-value	*	*	0.53	0.27	**0.0003**	0.1
	# of significant miRNAs					
common	17	56	1	0	2	1
replicate 1	25	60	12	0	2	1
replicate 2	36	98	12	24	2	33
P-value	$\mathbf{1.6 \times 10^{-7}}$	$\mathbf{8.3 \times 10^{-10}}$	2.2×10^{-1}	—	**0**	**0**

4 Discussion

Although P_{m_0} is mostly the smallest, P_m with $m \neq m_0$ also can take the value as small as P_{m_0} (see Table 2). The reason why it occurs is not because their seed sequence is similar to those of trasnfected miRNA. Actually, seed sequence of miRNA whose P_m is as small as P_{m_0} is very different from that of transfected miRNA.

Thus, the reason why miRNAs not transfected often have P_ms as small as P_{m_0} may be because of the secondary effect. First of all, while transfected miRNA regulates target genes, these genes are targeted by other miRNAs, too. If these genes have significant overlaps with other miRNA's target genes, P_ms of those miRNAs often take significantly small P-value. It is impossible to exclude this possibility, since we do not consider miRNA expression at all. In some sense, it is remarkable that P_{m_0} often has the smallest P-value. Next, genes regulated by tnasfected miRNA induce expression of other genes which are not direct targets of transfected gene. If some of not transfected miRNAs' target genes have siginificant overlap with those secondary induced genes, those miRNA has inevitably significant small P-value. Actually, P_{m_0} for three days after transfection usually does not have small enough P-value. It possibly means that secondary induced genes are apparently more *regulated* than genes directly targeted by transfected miRNA. Thus, it results in the lack of detection of gene regulation by transfected miRNA.

The datasets of day 1for miR-185 transfection is only one case which failed showing significant coincidence between biological replicates. This fact is very interesting because this is also the only the case in which one of biological replicates (replicate 1) cannot infer transfected miRNA correctly within samples at day 1. This possibly means that this replicate may have failed to produce suitable gene expression profile in some sense. For other five cases (one day and three days after mir-107 transfection, three days after mir-185 transfection, one day and three days after let-7a transfection), at least one statistical test can provide

us significant P-value. Therefore, it is unlikely that the detection of regulation of target genes by not transfected miRNA is due to the analytical errors. It is important to clarify the biological reason of this unexpected result.

It is also interesting that miRNA target genes are generally more expressed than the other genes in the present datasets.(in Figs. 1, there are peaks around $\log x_g \simeq 7$, which is far from origin. This tendency cannot be seen in genes not targeted by any miRNA (not shown here). This fact may also be important to understand how each miRNA regulate genes in cancer formation/suppression.

One may wonder that accidental overlap of target genes results in significant regulation of target genes of non-transfected miRNAs. In order to check this point, we have computed P-values of significant overlap of target genes by $P(n_{mm_0}, n_m, n_{m_0}/N)$. As a result, even after correction considering multiple comparison, 156, 157 and 156 miRNAs among in total 162 miRNAs have significant overlap ($P < 0.05$) with transfected miRNA of mir-107, mir-185 and let-7a respectively. This means, almost all of non-transfected miRNAs have significant large number of common target genes with those transfected miRNA.

One may think that this is the evidence that our analysis is erroneous, but it is not the case. Actually, if we compute correlation coefficient between P_m and P^O_{m,m_0}, for most of cases, there are no significant correlations (see Table 4). This means, significant regulation of target genes of non-transfected miRNA cannot be explained by the accidental target gene overlap with those of transfected miRNA, m_0.

Only two exception among total 12 cases (i.e., two biological replicates $\times$ two time points (day 1 or day 3) $\times$ three transfection) is for day 3. Since day 3 is late date when the effects of transfected miRNA become weaker, apparent significant regulation of target genes of non-transfected miRNAs is caused not by accidental overlaps of target genes, but by secondary effect. In other words, these genes are regulated indirectly through direct regulation of target genes by transfection.

Table 4. Significance of correlation between P_m and P^O_{m,m_0}. Bold numbers are significant ($P < 0.05$)

Transfection	mir-107				mir-185				let-7a			
Date	day 1		day 3		day 1		day 3		day 1		day 3	
Replicates	1	2	1	2	1	2	1	2	1	2	1	2
correlation	0.031	0.050	0.15	0.22	0.052	-0.077	-0.045	0.23	0.080	-0.042	-0.019	0.11
P-value	0.70	0.53	0.058	**0.0059**	0.51	0.33	0.57	**0.0035**	0.31	0.6	0.82	0.15

5 Conclusion

In this paper, we have shown that gene expression profile combined with miRNA target genes predicted computationally can often correctly infer transfeted miRNA. This suggests that we may be able to infer miRNA regulation of genes solely from gene expressions without considering any other information than computationally predicted target genes.

References

1. Brennecke, J., Stark, A., Russell, R.B., Cohen, S.M.: Principles of microRNA-target recognition. PLoS Biol. 3, e85 (2005)
2. Chi, S.W., Zang, J.B., Mele, A., Darnell, R.B.: Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460, 479–486 (2009)
3. Takahashi, Y., Forrest, A.A.R., Maeno, E., Hashimoto, T., Daub, C.O., Yasuda, J.: MiR-107 and MiR-185 Can induce cell cycle arrest in human non small cell lung cancer cell lines? PLoS One 4, e6677 (2009)
4. Ikeo, K., Ishii, J., Tamura, T., Gojobori, T., Tateno, Y.: CIBEX: center for information biology gene expression database. C. R. Biol. 326, 1079–1082 (2003)
5. Nielsen, C.B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J., Burge, C.B.: Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. RNA 13, 1894–1910 (2007)
6. Friedman, R.C., Farh, K.K., Burge, C.B., Bartel, D.P.: Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105 (2009)
7. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., Segal, E.: The role of site accessibility in microRNA target recognition. Nat. Genet. 39, 1278–1284 (2007)
8. Krek, A., Grün, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., Rajewsky, N.: Combinatorial microRNA target predictions. Nat. Genet. 37, 495–500 (2005)
9. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S.: Human MicroRNA targets. PLoS Biol. 2, e363 (2004)
10. Alexiou, P., Maragkakis, M., Papadopoulos, G.L., Reczko, M., Hatzigeorgiou, A.G.: Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics 25, 3049–3055 (2009)
11. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300 (1995)
12. R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2009) ISBN 3-900051-07-0 `http://www.R-project.org`
13. `http://www.granular.com/ICIC2010/SPM.txt`