Computational Modeling of Post-Transcriptional Gene Regulation by MicroRNAs

RAYA KHANIN\(^1\) and VERONICA VINCIOTTI\(^2\)

ABSTRACT

MicroRNAs (miRNAs) have recently emerged as a new complex layer of gene regulation. MiRNAs act post-transcriptionally, influencing the stability, compartmentalization, and translation of their target mRNAs. Computational efforts to understand the post-transcriptional gene regulation by miRNAs have been focused on the target prediction tools, while quantitative kinetic models of gene regulation by miRNAs have so far largely been overlooked. We here develop a kinetic model of post-transcriptional gene regulation by miRNAs, focusing on the miRNAs' effect on increasing the target mRNAs' degradation rates. The model is fitted to a temporal microarray dataset where human mRNAs are measured upon transfection with a specific miRNA (miRNA124a). The proposed model exhibits good fit with many target mRNA profiles, indicating that such type of models can be used for studying post-transcriptional gene regulation by miRNA. In particular, the proposed kinetic model can be used for quantifying the miRNA-mediated effects on its targets in the miRNA mis-expression experiments. The model makes an experimentally verifiable prediction of the miRNA124a decay rate, quantifies the miRNA-mediated effect on the target mRNAs degradation, and yields a good correspondence between the inferred and experimentally measured decay rates of human target mRNAs.

Key words: maximum likelihood, Michaelis-Menten kinetics, microarray data, miRNA regulation, mRNA.

1. INTRODUCTION

Over the past few years, it has emerged that gene expression in plants and animals is post-transcriptionally regulated by microRNAs (miRNAs). The level of complexity of this stage of regulation rivals the transcriptional one (Hobert, 2004). Mature miRNAs are small (21–25 nucleotide) noncoding RNA molecules that influence the messenger RNAs (mRNAs). They are estimated to comprise 1–5% of animal genes, making them one of the most abundant classes of regulators. Their widespread and important role in animals is highlighted by recent estimates that up to 30% of an organism’s protein-coding genes are subject to miRNA-mediated control (Krek et al., 2005; Lewis et al., 2005; Stark et al., 2005; Xie et al.,...
and is evidenced by their evolutionary conservation (Rajewsky, 2006). MiRNAs play a central role in many biological processes, including developmental timing, cell proliferation, apoptosis, metabolism, cell differentiation, and morphogenesis (Alvarez-Garcia and Miska, 2006; Ambros, 2004). They might also help to maintain and define cell types by suppressing the expression of unwanted transcripts (Farh et al., 2005; Stark et al., 2005; Sood et al., 2006).

The mechanism by which miRNAs regulate gene expression is post-transcriptional, possibly influencing the stability, compartmentalization, and translation of mRNAs (Bagga et al., 2005; Carthew, 2006). The details of the microRNA-mediated post-transcriptional gene silencing are not known. In plants, miRNAs inhibit a target mRNA by near-perfect complementarity and direct mRNA target degradation (Standart and Jackson, 2007). In contrast to that, an animal miRNA typically makes imperfect partial base pair contacts with its target transcript. It has been thought that only if an animal miRNA encounters a target with complete complementarity, it can enter the RNAi pathway and guide target degradation. However, it has been reported that in C. elegans (Bagga et al., 2005) and mammalian cells (Wu et al., 2006) miRNAs accelerate the decay of mRNAs despite the fact that their 3'UTR regulatory sequences can only partially base-pair with the miRNA.

It is conceivable that miRNAs act through multiple cooperative mechanisms to repress their targets (Carthew, 2006). One model is that the interaction between a miRNA and its target results in the transport of mRNAs to P-/GW-bodies, where they are unavailable to the protein synthetic machinery but are subject to decapping and degradation by resident nucleases. An additional or alternative miRNA-mediated pathway has been demonstrated in mammalian cells where two different miRNAs—miR125b and let-7—increased mRNA decay rates upon association with their target mRNAs by promoting rapid deadenylation (Wu et al., 2006). According to Wu et al. (2006), this increased rate of deadenylation does not result from the diminished frequency of translation caused by miRNA binding. To summarize, it is now clear that, with some target mRNA/miRNA pairs, there is a significant increase in mRNA degradation rates and, consequently, a reduction in mRNA abundance (Standart and Jackson, 2007).

The current explosion of interest in miRNAs and their role in post-transcriptional gene regulation has created demand for experimental techniques to study them (Neely et al., 2006), as well as computational tools to identify their mRNA targets. Recent computational estimates indicate that an average miRNA may regulate hundreds of genes (Krek et al., 2005). This has also been confirmed experimentally in a number of microarray studies (Krützfeldt et al., 2005; Lim et al., 2005). Because experimental identification of miRNA targets is difficult, several groups have developed computational target prediction tools (Rajewsky, 2006).

Most computational efforts to understand the post-transcriptional gene regulation by miRNAs have been focused on the target prediction tools. Recently, kinetic models of post-transcriptional gene regulation by small RNAs (Levine et al., 2007b; Shimoni et al., 2007) small interfering RNAs (Malphettes and Fussenegger, 2006), and miRNAs (Levine et al., 2007; Xie et al., 2007) have been proposed. A quantitative model of gene regulation in E. coli by two classes of small RNAs (Levine et al., 2007b) demonstrated that small RNAs provide a novel mode of gene regulation with a threshold-linear response, a robust noise resistance characteristic, and a built-in capability for hierarchical cross talk. It has also been shown quantitatively that regulation by small RNAs is advantageous when fast responses to external signals are needed and that regulation by sRNA may provide fine-tuning of gene expression (Shimoni et al., 2007). In addition, small RNAs have been suggested to participate in sharpening a gene expression profile that was crudely established by a morphogen (Levine et al., 2007a).

MiRNAs function very much like small interfering RNAs, but these two types of small RNAs can be distinguished by their distinct pathways for maturation and by the logic by which they regulate gene expression (Du and Zamore, 2005). The coupled degradation of the target mRNA and its regulator (Levine et al., 2007b) is specific to the gene regulation by small RNAs, while miRNAs, incorporated into the RISC complex, do not degrade with their targets but return to the cytosol to begin a new round of target mRNA
repression. It is plausible, however, that due to increased endonucleolytic activity, miRNA is sometimes degraded in the process after a few cycles of target mRNA binding (Levine et al., 2007a).

A kinetic model of miRNA-mediated post-transcriptional gene silencing was recently introduced (Levine et al., 2007). These authors considered a two-step model where the binding of miRNA to the mRNA promotes a secondary process (e.g., ribosome run-off or deadenylation) that ultimately leads to mRNA accumulation in its processed state, perhaps in P-bodies. (Levine et al., 2007) show that the target mRNA and protein levels may be tuned by target-specific parameters while global effectors may alter this behavior for some—but not all—miRNA targets in the cell. Another recent paper considers the role of miRNA in the delayed negative feedback regulation of gene expression (Xie et al., 2007). It was shown that, by enhancing the decay of mRNAs, miRNA may be a destabilizing or stabilizing factor in the dynamics of gene expression with delayed negative feedback, depending on the severity of its effect on mRNA degradation. Both models (Levine et al., 2007; Xie et al., 2007) are theoretical and do not relate the models to quantitative experimental data.

In this paper, we introduce a minimal, yet plausible, model of post-transcriptional gene regulation by miRNAs, focusing on miRNA-mediated target mRNA degradation. The goal here is to apply the model to temporal gene expression data from a miRNA transfection experiment (Wang and Wang, 2006) and to quantify the overall kinetic parameters of the miRNA-mediated target downregulation.

2. MODEL

2.1. Model of microRNA-mediated mRNA degradation

One can start modeling the post-transcriptional effect of miRNAs on the expression level $\mu$ of its target mRNA for a gene, by noticing that these levels change due to production and degradation (Bolouri and Davidson, 2002). In the models of transcriptional gene regulation, it is commonly assumed that the production (transcription rate), $p$, depends on the state of the system and the availability of a single or multiple transcription factors. The mRNA degradation is commonly assumed to be the first-order process that occurs with a constant rate, $\delta$:

$$\dot{\mu}(t) = p(t; TF) - \delta \mu(t).$$

(1)

It has been shown experimentally that miRNAs directly affect the levels of their target transcripts (Bagga et al., 2005; Krützfeldt et al., 2005; Lim et al., 2005), presumably by accelerating their degradation rates, and thus lowering their expression levels. At least with some target mRNA/miRNA pairs, the increased degradation is via the normal pathway of deadenylation (Wu et al., 2006) followed by decapping and then degradation of the body of the mRNA, rather than via an initial endonucleolytic cleavage such as it occurs in RNA interference with small interfering RNAs that have perfect complementarity to the target mRNA (Standart and Jackson, 2007). Therefore, when a transcript is a target of a specific miRNA, its degradation rate, $\delta$, can no longer be considered a constant. In this case, $\delta$ becomes dependent on the level of the miRNA, $m$: $\delta = \delta_0 + \delta(m)$, where $\delta_0$ is the basal degradation rate of the target mRNA. The basal degradation here refers to the degradation process that is independent of the miRNA under consideration but it may reflect the regulation by other miRNAs that are present and target the mRNA transcript.

The exact functional form of the miRNA-mediated degradation of its target mRNA, $\delta(m)$, is not known except that the presence of miRNA enhances the degradation (Standart and Jackson, 2007). A plausible assumption is that the degradation rate of a target mRNA depends linearly on the miRNA level:

$$\delta(m) = \delta_0 + \tilde{d} \cdot m = \delta_0(1 + \tilde{d} \cdot m),$$

(2)

where $d \geq 0$ is the miRNA-mediated fold-change in the transcript mRNA degradation rate relative to the basal degradation rate, $\delta_0$. This mass-action equation has been used for modeling gene regulation by small RNAs (Levine et al., 2007b). A more general model describes the miRNA-mediated degradation rate by a saturative, Michaelis-Menten type, function:

$$\delta(m) = \delta_0 + \tilde{d} \cdot \frac{m}{k + m} = \delta_0 \left(1 + d \cdot \frac{m}{k + m}\right).$$

(3)
The above equation implies that for $m \ll k$, the increase in the miRNA mediated degradation is linear with $m$, while it saturates at higher levels of miRNA. Equations (2) and (3) include the case of the mRNA not being a target of the miRNA ($d = 0, \delta = \delta_0$). The kinetic parameters $d$ and $k$ in Equation (3)—$d$ in Equation (2)—are likely to depend on the strength of miRNA binding to its sites on the target mRNA 3'UTRs. To account for multiple sites for the same miRNA (Rajewsky, 2006), models (2) and (3) can be extended to include the cooperativity (or Hill coefficient), $h \geq 1$:

$$\delta(m) = \delta_0(1 + d \cdot m^h),$$

$$\delta(m) = \delta_0 \left(1 + \frac{d \cdot m^h}{k + m^h}\right).$$

An alternative or additional mechanism of miRNA-mediated transcript downregulation is modeled by considering three different target mRNA concentrations: free mRNA, bound miRNA-mRNA, and processed mRNA (e.g., localized in P-bodies) as well as free miRNA levels (Levine et al., 2007). The level of mRNA is affected if a major part of miRNA-mediated target regulation occurs in P-bodies. Alternatively, the level of mRNA is not affected if microRNA exert its effect on repressing translation. It is impossible to distinguish between the two different pathways of miRNA-mediated changes in target mRNA levels, whether they occur by increased degradation or by localization in the P-bodies (and therefore mRNA removal from the cytosol).

One miRNA may have several (up to hundreds of) target genes, indicative of target multiplicity (Krek et al., 2005; Lim et al., 2005). The expression level of each target mRNA transcript $i$ can be described by

$$\dot{\mu}_i(t) = p_i - \delta_{0i} \left(1 + d_i \frac{m}{k_i + m}\right) \mu_i(t)$$

with gene-specific kinetic parameters for transcription $p_i$, basal degradation rate $\delta_{0i}$, and parameters of the miRNA-mediated degradation, $d_i$ and $k_i$. Generally speaking, the transcription term $p_i$ depends on one or more transcription factor(s). In this work it is assumed to be constant and not dependent on other target genes or miRNAs. The steady state level of a mRNA transcript $i$ ($\dot{\mu}_i = 0$), given by

$$\mu_i = \frac{p_i}{\delta_{0i} \left(1 + d_i m/(k_i + m)\right)}.$$ 

can be used for analyzing microarray datasets of different tissues or cell-lines.

The dynamic model (6) is applicable to study miRNA-mediated effect on target mRNAs in the experiments where miRNA levels are modulated by transfections (Lim et al., 2005) or antagonomers (Krützfeldt et al., 2005). In the present paper, the model is fitted to a time-course microarray dataset where human mRNAs are measured at several time-points post miRNA124a transfection (Wang and Wang, 2006).

### 2.2. Model of microRNA decay

**MicroRNA levels are constant.** If the miRNA level in a cell-line or tissue does not change much with time, the value of $m$ in Equation (6) can be fixed at some level and a closed-form solution for each target mRNA can readily be written:

$$\mu_i(t) = \frac{p_i}{\delta_{0i} \lambda_i} + C_i e^{-\delta_{0i} \lambda_i t}, \quad \lambda_i = 1 + d_i \frac{m}{k_i + m}, \quad C_i = \frac{p_i}{\delta_{0i} \lambda_i} \left(1 - \frac{1}{\lambda_i}\right).$$

In modeling the transfection experiment of Wang and Wang (2006), it is assumed that at time $t = 0$ (transfection time), miRNA levels are at its maximum, $m = 1$, while the expression levels of mRNA targets are not affected yet by the miRNA and their levels are at the control values. Under the assumption that the miRNA level is constant, with this level fixed to $m = 1$ without loss of generality, and that there is no dependence between the gene targets, the 4 gene-specific kinetic parameters $\theta_i = \{p_i, \delta_{0i}, d_i, k_i\}$ can be fitted independently for each target gene using Equation (7).
MicroRNA levels are time dependent. It is quite plausible that miRNA levels are influenced by other miRNAs (Tuccoli et al., 2006) or/and transcription factors (Shalgi et al., 2007). In addition, miRNAs can decay before they are incorporated into the stable RISC complexes or degrade in the process of a few cycles of mRNA binding (Levine et al., 2007a). The latter can be accounted for, for example, by introducing a global parameter that represents the probability for a miRNA to be co-degraded with the mRNA in the processed state (Levine et al., 2007). Additionally, miRNA molecules, incorporated into the RISC complexes, are sequestered into the target mRNA/miRNA complexes, and are subsequently translocated to the P-bodies (Bagga et al., 2005). Some of the mRNA/miRNA complexes might be stored in the P-bodies to be released in response to stress (Bhattacharyya et al., 2006). In the miR124a transfection experiment considered in this work (Wang and Wang, 2006), most of the downregulated predicted targets show upregulation in the last time-point(s) (see Fig. 2 below). If the downregulation in the expression of the target mRNAs is solely due to the miRNA124a effect, then it is plausible that the subsequent upregulation of the target mRNAs is due to a decrease in the level of the regulator miRNA. This is likely to be due to a finite half-life of the miRNA124a in the cytoplasm. Incorporating miRNA decay due to cell growth/division, natural miRNA decay and the yet unknown effects of mRNA/miRNA storage mechanism on the dynamics of the miRNA levels, we here assume that the level of the miRNA decreases with a rate, $\delta_m$.

When the miRNA levels depend on time, $m = m(t)$, Equation (6) cannot be solved analytically even in the case of a simple linear decay model for the miRNA. As in our previous work on the reconstruction of transcription factor activity from gene expression data of the targets (Khanin et al., 2006), one analytical approach to solve the integral would be to approximate the unknown $m(t)$ profile by a piece-wise constant function with the values $m_i$ on each time-interval where the measurements were taken. In this work, we adopt a parametric approach. The simplest model assumes that the miRNA level changes due to degradation and production, $s(t)$,

$$\frac{dm(t)}{dt} = s(t) - \delta_m m(t).$$  \hspace{1cm} (8)

We will further refer to $\delta_m$ as the miRNA decay rate. In the miRNA transfection experiments (Wang and Wang, 2006), the miRNA level in the cell-line is at its maximum level at the initial (transfection) time $m(0) = 1$. Assuming no production of the miRNA in the cell-line where it is normally not present, $s(t) = 0$, the miRNA temporal profile is given by $m(t) = m_0 e^{-\delta_m t}$. This is consistent with the conducted miRNA overexpression experiment, where miRNA124a is transfected in the cell-line where it is normally not present (i.e., $s = 0$). In the cell-lines and tissues where a specific miRNA is present, its level can be approximated by $s/\delta_m$.

More complex models describing the dynamics of $m(t)$ can be considered. Future models should take into account various time-delays and dynamic storage of mRNA/miRNA complexes in P-bodies. In the light of limited experimental data available, we here ignore time-delays and use the assumption of miRNA decay with a constant rate, $\delta_m$.

3. METHODS

3.1. MicroRNA-regulated single input motif

The network structure wherein one miRNA post-transcriptionally regulates the levels of many target transcripts by enhancing their degradation rates is similar to the so-called Single Input Motif (SIM) that is common in transcription networks (Shen-Orr et al., 2002). In the transcription SIM, one transcription factor regulates (activates or represses) the transcription rate of many target genes. The miRNA-regulated SIM is condition-specific because many target mRNAs are presumably subject to control by other miRNAs whose levels do not change in the course of the experiment. Their effects are implicitly incorporated in the basal degradation rates of mRNAs.

The parameters of the SIM—that is the miRNA decay rate $\delta_m$, which is common to all of the targets in the SIM, and all the targets-specific parameters ($\delta_{oi}, d_i, k_i$)—are estimated using a maximum likelihood method, as described in the next section.
3.2. Maximum likelihood optimization

The maximum likelihood inference framework was developed in our previous work on reconstructing the transcription factor activity in a transcriptional SIM (Khanin et al., 2006). By analogy with transcriptional control, we refer to the regulatory module with one miRNA and many targets as the post-transcriptional SIM. The reconstruction procedure is based on finding the maximum likelihood for the whole miRNA-mediated SIM using the kinetic model (6) and (8) and gene expression data of the downregulated predicted target mRNAs. The observed gene expression of a target gene, taken on its original scale, is assumed to be log-normally distributed:

\[ g_i(t) \sim \text{lognorm}(a_i(t), \sigma_i^2), \quad E(g_i(t)) = \mu_i(t). \]  

(9)

where the location parameter, \( a_i(t) \), depends, via the true gene expression \( \mu_i \), on the miRNA profile, \( m(t) \), gene-specific kinetic parameters \( \theta_i = \{ \delta_i, d_i, k_i, p_i \} \), and the variances \( \sigma_i^2 \): \( a_i(t) = \log[\mu_i(t)] - \frac{1}{2} \sigma_i^2 \). In the absence of miRNA, the steady-state value of a gene \( i \) is given by \( \mu_i(m = 0) = p_i/\delta_{0i} = \tilde{p}_i \). This is used to reduce the number of kinetic parameters to 3: \( \{ \theta_i = d_i, \delta_{0i}, k_i \} \), while \( \tilde{p}_i \) is estimated from the initial and control time-points in the transfection experiment.

The kinetic parameters of post-transcriptional regulation are sought by maximizing the overall likelihood of the miRNA-mediated SIM:

\[ L_{\text{overall}}(\Theta, \Sigma^2; m) = \prod_{i=1}^{N} L_i(g_i(t); \theta_i, \sigma_i^2; m). \]  

(10)

Here \( \Theta \) represents the gene-specific kinetic parameters, \( \theta_i \), for all target mRNAs, \( i = 1, \ldots, N \), in the regulatory module; \( \Sigma^2 \) stands for the variances of the log-normal distribution, \( \sigma_i^2 \); and \( L_i(g_i(t); \theta_i, \sigma_i^2; m) \) is the gene-specific likelihood of a gene \( i \) given the observed data \( g_i(t) \) and the temporal profile of miRNA, \( m = m(t) \):

\[ L_i(g_i(t); \theta_i, \sigma_i^2, m) = \frac{1}{g_i(t) \sqrt{2\pi \sigma_i}} \exp \left\{ -\frac{1}{2} \left[ \frac{\log g_i(t) - a_i(t)}{\sigma_i} \right]^2 \right\}, \]  

(11)

where \( a_i(t) \) depends on the true gene expression \( \mu_i \), which is defined by the solution of the kinetic model (6). For the miRNA-mediated SIM, there are three kinetic parameters per gene to estimate for each value of \( \delta_m \) that is searched on a grid.

All computations were performed in the R statistical language. The computational method that maximizes the log-likelihood with respect to \( \Theta \) and \( \Sigma \) for all genes in the SIM is based on the conjugate gradient method and is implemented using the \texttt{optim} function in R. The search procedure uses multiple starts for locating the maximum with respect to all unknown parameters of interest. The R code is available from the authors.

3.3. Microarray data

The Affymetrix microarray dataset of the miRNA124a transfection experiment from Wang and Wang (2006) was used for this study. In this experiment, miRNA124 RNA duplex was transfected into HepG2 cell line and temporal measurements of 54,675 probe sets were taken at 4, 8, 16, 24, 32, 72, and 120 h post-transfection. A negative control sample was also transfected as reference for each of these time points. The RMA normalized data for this study can be downloaded from the Gene Expression Omnibus webpage (www.ncbi.nlm.nih.gov/geo) under the accession number GSE6207.

3.4. Predicted and downregulated targets

To determine the target genes for miRNA124a we used predictions from the Pictar algorithm (Krek et al., 2005). These targets can be downloaded from http://pictar.bio.nyu.edu. Amongst the pictar predicted targets, we determine the genes that are downregulated. A target is defined as being downregulated if its expression reduces by at least 40% at one (or more) time-points compared with the negative control.
The predicted targets that are found to be downregulated are used as the real miRNA targets for further analysis.

The pictar targets are given in RefSeq names. There are 787 unique RefSeq predicted targets corresponding to 510 unique UniGene gene symbols that are present on microarrays. Amongst the pictar genes present on the microarrays, 339 unique genes were found to be downregulated (corresponding to 655 transcripts) by at least 1.4 for at least one time-point. This data-file is available from the authors.

Since target mRNA profiles are highly correlated, we performed clustering of all data based on a correlation metric using the pamsam function from the smida R-library (www.stats.gla.ac.uk/~microarray/book/smida.html). This resulted in two automatically chosen clusters. Medoids of these two clusters have minima at the fourth and the seventh time-points correspondingly. The dataset has also been clustered into 10 clusters (pamsam with number of clusters fixed). This enables the identification of a greater variety of behaviors of gene profiles.

4. RESULTS

4.1. Predicted microRNA half-life

To estimate the miRNA decay rate, $\delta_m$, two sets of clusters were used for the ML optimization, consisting of two and 10, clusters respectively. Models for each $\delta_m$ are fitted independently to each target profile (medoid of each cluster). The miRNA decay rate is estimated as the value that gives the maximum likelihood for all genes and all $\delta_m$. Figure 1 shows the profile of the miRNA124a computed from Equation (8) with the MLE of the decay rate, $\delta_m = 0.024h^{-1}$ (solid line), and its 95% confidence bounds. The latter are calculated via a classical Wilks method: the MLE $\delta_m$ is perturbed, while keeping all other parameters fixed, such that the marginal confidence decreases by 95%, which corresponds approximately to a decrease in the likelihood by $2\chi^2_{1,0.95}$.

The decay rates estimated by our method correspond to a miRNA124a half-life of $29h$ and 95% confidence bounds $(26h, 50h)$. The miRNA124a profile is expressed in arbitrary units starting from the

![miRNA profile and confidence bounds](image)

**FIG. 1.** MicroRNA profile. The microRNA profile, given in arbitrary units, is computed from Equation (8) with the maximum likelihood estimate (MLE) for $\delta_m = 0.024h^{-1}$ (solid line) and its 95% confidence bounds (dashed lines) computed via a classical Wilks procedure. This corresponds to half-life of miRNA124a of $29h$ with the 95% confidence bounds $(26h, 50h)$. The above MLE for $\delta_m$ is computed by the ML inference procedure based on 10 clusters.
maximal value $m = 1$ at $t = 0$ (transfection time). This MLE for $\delta_m$ is computed by optimization on 10 clusters. Optimization on two clusters yields similar values (MLE $\delta_m = 0.02h^{-1}$ that corresponds to miRNA124a half-life of $35h$ with 95% confidence bounds being $(22h, 49.5h)$). We will further present results obtained on 10 clusters.

4.2. Model fit to medoids of 10 clusters

Figure 2 shows the reconstructed kinetic profiles of the 10 target medoids computed with the MLE for $\delta_m = 0.024h^{-1}$.

The cluster medoids are grouped according to their value ranges, and not by their functions, as we did not find any significant enrichment of GO categories in the clusters. Most of the medoids exhibit downregulation followed by upregulation at the last time-point. The model captures well the behavior of most profiles (clusters 1, 2, 4, 5, 10). The model fit to those medoids is good, with the goodness-of-fit (GOF) $p$-values computed using the Kolmogorov-Smirnov test being statistically significant at 5% level (Fig. 2). These $p$-values are computed using the R function `ks.test()`: if the model fits well to the data, then the distribution of the gene expression should be approximately lognormal (from Equation (9)). Given that the variance $\sigma^2$ is unknown, the gene expression can be approximated by a t-test distribution with 3 degrees of freedom (the number of observations minus the number of estimated parameters).

FIG. 2. Michaelis-Menten model fit to the 10 cluster medoids. The model fit for 10 medoid profiles computed by the ML inference procedure with the maximum likelihood estimate (MLE) for $\delta_m = 0.024h^{-1}$ (half-life 29h). Numbers stand for the data-points relative to the corresponding cluster medoid, and the lines represent the profiles fitted by model (3). Simulations were run using R function `optim` for 500 multiple starts per medoid. The GOF $p$-values for the clusters 1 to 10 are 0.75, 0.99, 0.73, 0.54, 0.49, 0.74, 0.37, 0.45, 0.6, and 0.61, respectively.
Clusters 6 and 9 show upregulation at some initial time-points followed by downregulation. This behavior can probably be attributed to high noise in the system. Medoids 3 and 8, which are on a much lower scale than the other medoids, also exhibit high noise and the medoids do not follow a well-defined pattern. Given just one replicate per time-point, the fit of the model to these clusters is still reasonably good (with large GOF p-values).

It has been demonstrated that multiple sites for the same miRNA on the target mRNA 3’UTR boost target repression (Rajewsky, 2006). We therefore explored whether the model with cooperativity \( h = 2 \) (Equation (5)) fits some of the clusters better. Indeed, we found that the model that incorporates the two binding sites for the miRNA achieved higher likelihoods for some clusters (6, 7, 8, and 10) (results not shown). This could be an indication that these clusters are enriched with miRNAs that have two (or more) binding sites for miRNA124a on their 3’UTRs. A higher Pictar score of a target can indicate the existence of multiple site for the miRNA. However, we did not find an enrichment of these clusters with the target miRNAs that have higher Pictar scores. One of the future directions for this work is to explore whether there is a dependence between the Pictar score of a target, and the fact that the kinetic model with multiple binding sites for the miRNA gives a better fit to the data.

4.3. Kinetics of microRNA-mediated target down-regulation

The kinetics of the miRNA-mediated downregulation of target miRNAs have not yet been measured. We here demonstrate that one of the ways of quantifying this regulation is by fitting a kinetic model to gene expression data measured at different time-points following the change in the levels of the specific miRNA by transfection or silencing. Indeed, from the ML reconstruction procedure, it is possible to estimate the numerical values for the kinetic parameters of the model. The MLE for the transcript-specific (or cluster-specific) kinetic parameters \( \delta_i, d_i, k_i \) are given in Table 1.

The rate-constants \( d_i \) and \( k_i \) determine the extent of the miRNA-mediated downregulation of the target miRNAs. In particular, \( d_i / (m + k_i) \) determines the fold-change in the target mRNA degradation rates relative to the basal rate, \( \delta_{0i} i = 1, ..., 10 \). For all but one cluster \( (i = 9) \), we found that the half-saturation constant is much bigger than \( m \sim 1 \) \( k_i \gg m \) indicating that the miRNA-mediated degradation does not saturate at these levels of miRNA and can therefore be adequately modeled by a linear dependence (Equation (2)). Following this, we have performed optimization using the linear model (2) and found overall very similar results (not shown). Indeed, a \( \chi^2 \)-test with one degree of freedom (the difference in the number of parameters between the two models) found no significant difference between the non-linear (Equation (3)) and linear (Equation (2)) models. Noting this fact, we here proceed with the non-linear model as it is a more general one and can be applied to the experimental data wherein the levels of miRNA change dramatically.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>( \delta_0 ) ( h^{-1} )</th>
<th>( d_i / (1 + k) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.009</td>
<td>6.4</td>
</tr>
<tr>
<td>2</td>
<td>0.029</td>
<td>1.54</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>2.94</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>8.82</td>
</tr>
<tr>
<td>5</td>
<td>0.012</td>
<td>2.0</td>
</tr>
<tr>
<td>6</td>
<td>2.63</td>
<td>0.20</td>
</tr>
<tr>
<td>7</td>
<td>0.007</td>
<td>3.25</td>
</tr>
<tr>
<td>8</td>
<td>0.0008</td>
<td>5.99</td>
</tr>
<tr>
<td>9</td>
<td>0.003</td>
<td>1.93</td>
</tr>
<tr>
<td>10</td>
<td>0.13</td>
<td>1.28</td>
</tr>
</tbody>
</table>

Maximum likelihood estimates for the cluster-specific parameters for the kinetic model of miRNA regulation (Equation (3)). The maximum log-likelihood for the SIM with one regulator and 10 medoids is 138.6, obtained at \( \delta_m = 0.024 h^{-1} \) (half-life 29h).
The effective miRNA-mediated fold-change increase in the target mRNAs degradation rates, given by $d_i/(1 + k_i)$, is well over 1 (with the exception of cluster 6) (Table 1). These values range from about 1.3 to 6.4, and reflect the maximum extent of the miRNA-mediated downregulation of each target medoid expression. Experiments wherein the mRNA expression is measured at different levels of miRNA would be particularly helpful in determining the miRNA dosage-dependent effect on the target downregulation.

Unlike the kinetics of the miRNA-mediated regulation, the degradation rates of human mRNAs have been measured in two cell-types (Yang et al., 2003). In particular, the human mRNAs decay rates have been obtained in the HepG2 cell-type, the same cell-type that is used in the microarray studies discussed in this paper (Wang and Wang, 2006). It is therefore interesting to directly compare the decay rates obtained in Yang et al. (2003) with the ones estimated by our ML inference procedure. To this end, we fixed the miRNA decay rate at its MLE obtained on 10 medoids ($\delta_m = 0.024h^{-1}$) and repeated the optimization for all downregulated picTar predicted target mRNAs. We mapped the genes from the Table 9 of Yang et al. (2003) with the gene-targets used in our optimization (188 unique gene names, corresponding to 388 transcripts). We found that about 60% of the mapped transcripts have their inferred basal degradation rates, $\delta_0$, within the interval mean ± SD (standard deviation) reported in Yang et al. (2003), while over 80% of the mapped transcripts have their MLE of $\delta_0$ within the interval mean ± 2·SD of Yang et al. (2003). This is a very good correspondence of the decay rates obtained from two independent studies. (The file with the gene names, measured and reconstructed decay rates can be obtained from the authors.) In our work, the target mRNAs degradation rates are inferred using a model-based approach from a rather small microarray dataset (seven time-points and controls, one replicate per time-point). The transfection experiment of Wang and Wang (2006) has not been specifically designed for measuring the mRNA decay rates, and the MLE of the mRNA decay rates are by-products of the reconstruction procedure that is inevitably based on various modeling assumptions and simplifications. We believe that such a good correspondence gives a strong support for the current modeling approach and can therefore be applied to other datasets.

The model shows good fit to many gene profiles, while the fit to the noisy profiles that do not follow specific patterns is relatively poor. This is similar to the results obtained for the cluster medoids (Fig. 2). Overall, GOF p-values obtained by the Kolmogorov-Smirnov test are statistically significant at the 5% level, indicating a good fit of the model to the data (not shown). The effective miRNA-mediated fold-change in the mRNA degradation rates for all targets has the mean 1.6, indicating that on average the target mRNA expression levels can be decreased 2.6-fold by the regulator miRNA compared to the state wherein this miRNA is present.

5. DISCUSSION

The motivation behind this paper stems from the need to develop comprehensive models of gene regulation on both transcriptional and post-transcriptional levels with the goal of further integration of such models in the overall complex regulatory network of genes. In this paper, we propose a simple model of post-transcriptional gene regulation by miRNA, focusing on miRNA-mediated effect on the target mRNA degradation rates. The model is based on the experimental findings that miRNAs can reduce the levels of their target mRNAs by enhancing their degradation rates (Bagga et al., 2005; Lim et al., 2005; Standart and Jackson, 2007; Wu et al., 2006). It is assumed that the downregulation of target mRNAs in the transfection experiment (Wang and Wang, 2006) is directly mediated by miRNA. As a consequence, experimentally observed up-regulation of targets at later time-points is primarily due to the decay of free miRNA. By fitting the model to the time-course microarray dataset (Wang and Wang, 2006), kinetic parameters of miRNA-mediated regulation as well as the miRNA-target half-life ($29h$ with 95% confidence bounds [26h, 50h]) are estimated. This estimate incorporates the effects of free miRNA decay, its sequestering into P-bodies as well as cell growth and division (and thereby miRNA dilution). MiRNA half-life times are difficult to measure experimentally. Our model applied to the temporal microarray datasets makes a prediction for the miRNA124a half-life that can be experimentally verified. This is the first attempt to the best of our knowledge to quantify the kinetics of the free miRNA in the cytoplasm as well as the miRNA-mediated downregulation of its target mRNAs.

We have found that the miRNA downregulating effect on the target mRNA degradation rates can adequately be described by either a non-linear or a linear model, $\delta = \delta_0 + d \cdot m$, where the parameter $d$...
COMPUTATIONAL MODELING BY MicroRNAs

reflects the fold-change in the mRNA degradation rate due to miRNA effect. In order to distinguish between the two models, miRNA mis-expression (overexpression and silencing) experiments are required where the miRNA levels can be measured. We have also found that the model that takes into account multiple sites for the same miRNA on the 3'UTR of the target mRNAs, gives a better fit to some mRNA profiles. However, we did not manage to confirm that those targets indeed have two binding sites on their 3'UTRs.

Due to current lack of experimental quantitative data we do not aim here to distinguish between complex mechanisms of miRNA-mediated gene silencing that still remain elusive and controversial (Standart and Jackson, 2007). Naturally, more complex models for miRNA-mediated target degradation should be developed in the future. Such models will have to include multiple-steps (Levine et al., 2007), be complemented by the models of miRNA-mediated inhibition of target translation, models of miRNA biogenesis and dynamics as well as modified models of RNA silencing pathways (Groenenboom et al., 2002). Clearly, further computational studies require more experimental data. With the little experimental data that is currently available, we believe that our approach gives a good illustration for the quantitative reconstruction of the kinetics of post-transcriptional gene regulation by microRNAs from microarray time-course in miRNA overexpression experiment.

ACKNOWLEDGMENTS

R.K. thanks Nikolaus Rajewsky for his encouragement of this work and for many helpful discussions on microRNAs and modeling in systems biology. We also thank X. Wang for helpful comments about the microarray dataset used in this work. This research was supported in part by the National Science Foundation (grant PHY05-51164) during R.K.’s visit to KITP, Santa Barbara, CA; Evolution of Molecular Networks program.

DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES


Address reprint requests to:

*Dr. Raya Khanin*

*Department of Statistics*

*University of Glasgow*

*Glasgow G12 8QQ, United Kingdom*

*E-mail: raya@stats.gla.ac.uk*