

## MOLECULAR BIOLOGY &amp; GENETICS

# Gene regulatory network stabilized by pervasive weak repressions: microRNA functions revealed by the May–Wigner theory

Yuxin Chen <sup>1</sup>, Yang Shen<sup>1,2</sup>, Pei Lin<sup>1</sup>, Ding Tong<sup>1,3</sup>, Yixin Zhao <sup>1</sup>, Stefano Allesina<sup>4</sup>, Xu Shen<sup>1</sup> and Chung-I Wu<sup>1,4,\*</sup>

## ABSTRACT

Food web and gene regulatory networks (GRNs) are large biological networks, both of which can be analyzed using the May–Wigner theory. According to the theory, networks as large as mammalian GRNs would require dedicated gene products for stabilization. We propose that microRNAs (miRNAs) are those products. More than 30% of genes are repressed by miRNAs, but most repressions are too weak to have a phenotypic consequence. The theory shows that (i) weak repressions cumulatively enhance the stability of GRNs, and (ii) broad and weak repressions confer greater stability than a few strong ones. Hence, the diffuse actions of miRNAs in mammalian cells appear to function mainly in stabilizing GRNs. The postulated link between mRNA repression and GRN stability can be seen in a different light in yeast, which do not have miRNAs. Yeast cells rely on non-specific RNA nucleases to strongly degrade mRNAs for GRN stability. The strategy is suited to GRNs of small and rapidly dividing yeast cells, but not the larger mammalian cells. In conclusion, the May–Wigner theory, supplanting the analysis of small motifs, provides a mathematical solution to GRN stability, thus linking miRNAs explicitly to ‘developmental canalization’.

**Keywords:** microRNAs, network stability, canalization, May–Wigner theory, systems biology, RNA crosstalk

## INTRODUCTION

Large networks characterize many biological systems. Of particular interest to ecologists and evolutionists are food webs and gene regulatory networks (GRNs). These large and highly connected networks are intrinsically unstable according to the May–Wigner theory [1]. Stability is defined as the speed at which every node would return to the equilibrium after perturbation. Hence, in large networks, the probability that all nodes would return to the equilibrium is diminished. The issue of stability has been extensively analyzed for food webs [1–3] but the underlying mathematics of stabilization should be common among networks.

Stability of GRNs is an especially challenging problem for several reasons. First, GRNs are typically large with thousands of nodes, each representing the abundance of an mRNA. Size alone may demand that GRNs evolve mechanisms for

stabilization. Second, there are as many types of GRNs as there are tissues in multicellular organisms. The stabilizing mechanism thus needs to be general. Third, unlike food webs that may tolerate substantial fluctuations in node values (i.e. species abundance), GRNs can only function within a small range of transcript abundance as transcriptome data show. Fourth, and perhaps most importantly, GRN stability may underlie the ‘developmental canalization’ proposed by Waddington [4].

Canalization is a metaphor for water traveling along canals [4]. Ever perturbed constantly, water always returns to the canal and flows along a pre-determined path. There are two types of motion: (i) flowing along the path and (ii) returning to the canal quickly after perturbation. These two motions occur in very different timescales and canalization (in the narrow sense) refers to the second motion. GRN canalization means the quick return to the

<sup>1</sup>School of Life Science, Sun Yat-Sen University, Guangzhou 510275, China;

<sup>2</sup>Target Discovery Research, Boehringer Ingelheim Pharma GmbH & Co KG, 88397 Biberach an der Riß, Germany;

<sup>3</sup>Department of Biostatistics, School of Public Health, Yale University, New Haven, CT 06520, USA and <sup>4</sup>Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

\*Corresponding author. E-mail: ciwu@uchicago.edu

Received 30 March 2019; Revised 7 June 2019; Accepted 10 June 2019

developmental path after the network is perturbed, for example, by cell divisions.

Although ‘canalized development’ originally referred to phenotypic stability, every level of network subsumed under the phenotype—including the metabolome, proteome and transcriptome—should be stable as well. A stable GRN, measured by the transcriptome, is hence necessary for developmental canalization and could be the first step in the process. For GRN stability, microRNAs (miRNAs) have been postulated to be the canalizing molecules [4–9]. These are a very large class of small regulatory RNAs that degrade mRNAs and repress translation, the former being particularly important [10].

While substantial literature have implicated a role for miRNAs in canalization [11–14], a rigorous theory that connects known miRNA actions to GRN stability is still absent. Many analyses have been carried out on small motifs of 2–5 nodes [5,9], but motifs cannot be easily expanded into a network. For example, adding only one extra node to a coherent motif can make it incoherent, and vice versa [7]. An alternative approach is the large RNA:RNA networks [15,16], which at present cannot address the stability issue. In this study, we provide a mathematical solution to GRN stability by applying the May–Wigner theory to the empirical data on miRNA activities.

## RESULTS

### Diffuse actions of miRNAs

For regulatory genes, miRNAs seem paradoxical for two reasons: (i) exclusive downregulation of their direct targets, and (ii) broad and weak repression of hundreds of target genes. In the conventional view, miRNAs repress targets in order to effect phenotypic changes, but that view is contradicted by their peculiar properties presented above [17]. By analyzing multiple miRNA targets and phenotypes concurrently, Liufu *et al.* [18] recently concluded that the role of miRNAs is indeed in minimizing phenotypic fluctuation (i.e. canalization) rather than effecting directional changes, as has been heatedly debated [17,19,20].

We first present the defining characteristics of miRNAs, i.e. weak and broad repression of mRNAs. Unlike previous analyses [21–29], this study pays special attention to weak interactions, which will later be subjected to mathematical interpretation.

### Number of targets

We examined 178 conserved miRNAs in human cells for their target sites following the common

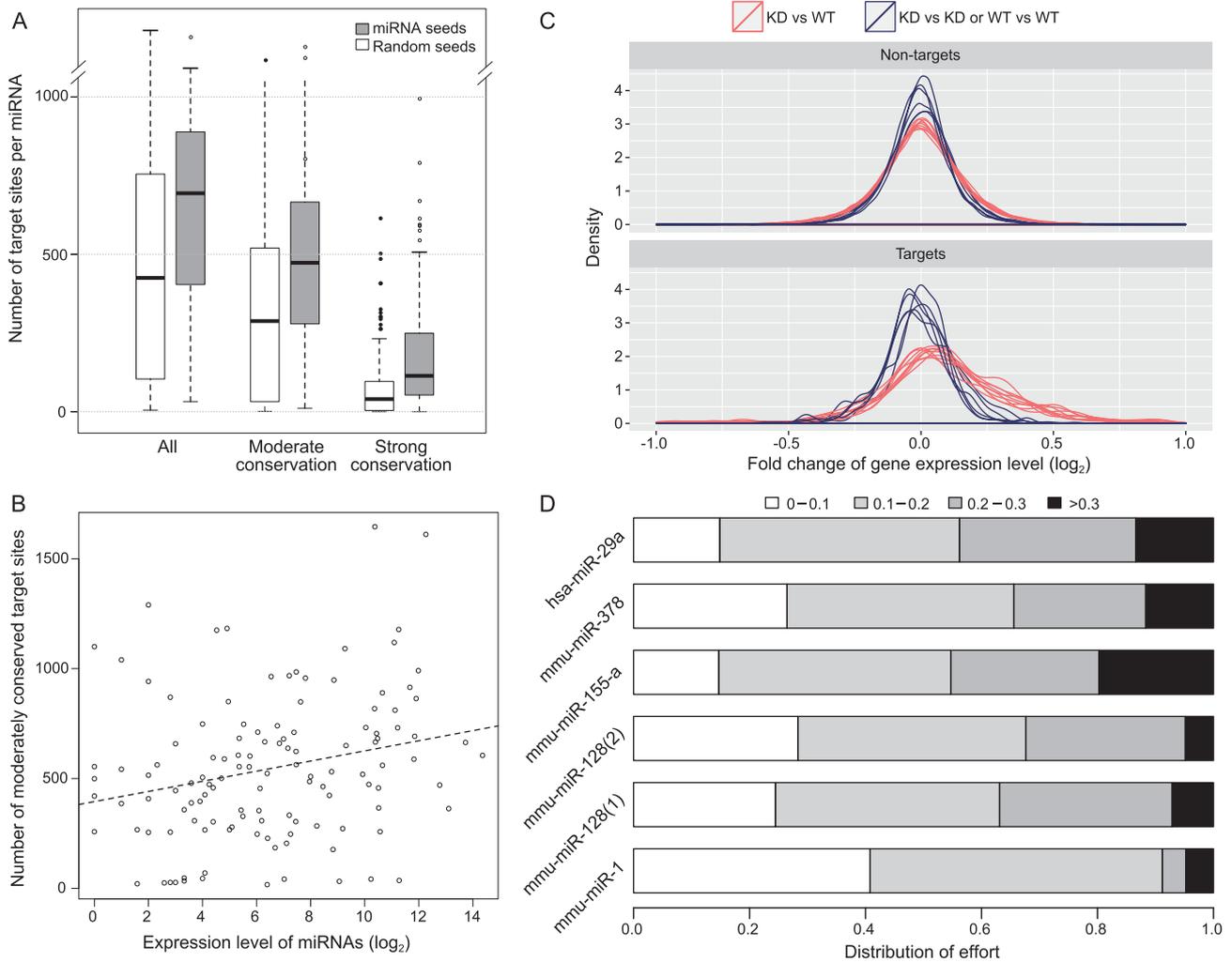
protocol (Fig. 1A; see *Methods*). Random seeds with the same CG content served as the control. If all potential targets are counted, the median number of target genes would be 694, >60% higher than the control. The numbers for the moderately and highly conserved targets were 473 and 114 (64 and 185% higher than the control), respectively. While highly conserved target sites are generally considered more reliable, Xu *et al.* [30] have shown that weakly conserved targets are also evolutionarily significant. Hence, the number of targets per human miRNA is likely to be between 100 and 500 [31–33] (Fig. S1-A, B).

The large number of target sites is even more puzzling for lowly expressed miRNAs. Given their limited repression capacity, these miRNAs might be expected to have far fewer targets. Figure 1B shows the prediction to be qualitatively true. However, the slope of the regression is extremely mild with a decrease of one-third of the target number when the expression decreases by >1000-fold (Fig. S1-C, D). Hence, if only strong repressions are functional, then more than one-half of the miRNAs expressed in any tissue would be non-functional.

### Strength of repression

With hundreds of targets, each miRNA is expected to exert weak effects on most targets. A typical example is given in Fig. 1C, which is based on six transcriptome data sets from the knockout line of hsa-29a miRNA (Fig. S2). The fold changes of target genes are symmetrically distributed around a peak that corresponds to ~3% repression. Note that the peak is not at 0%, as is the case for non-targets. Even though hsa-29a is moderately to highly expressed, the degradation of its targets is no more than 5%, on average.

Weak repression can still result in noise as long as the weak targeting collectively does not take up much of the total capacities of mRNAs. Therefore, we measured the fraction of each miRNA’s capacity that was used in weak repression. The distribution of effort (DOE) sums up all repressions of a certain strength, weighted by the expression level of the target gene. Figure 1D shows that miRNAs use most of their repression capacity to exert small influences on a large number of target genes. Indeed, only ~10% of the total repression capacity is used for the stronger repression (black bar, Fig. 1D). If we consider miRNAs that are themselves lowly expressed, DOE across all miRNAs would be even more biased toward weak repressions. We next analyzed weak repressions in the context of the GRN.



**Figure 1.** Predicted target number in relation to the observed derepression by miRNA knockout. (A) Number of miRNA target genes predicted by TargetScan (grey bars) vs. control (white bars) based on the shuffled seeds of the same miRNAs. The comparison is done at three levels of evolutionary conservation (see also Figure S1-A, B). (B) Correlation between the expression level of 109 miRNA seeds and the predicted number of moderately conserved targets. The correlation is positive but the slope is very small (see text and also Figure S1C, D). (C) Distribution of fold change in the expression of target genes in miRNA (hsp-29a) knockout lines between experiments and controls (red lines) and between controls (blue lines). The median increase upon miRNA knockout is <10%. (See also Figure S2). (D) DOE on target repression by each of 6 miRNAs. These efforts are categorized into four levels depending on the effect of repression, ranging from <10% to >30%. DOE sums up the repressions across all target genes, weighted by their expression level. Strong repression of >30% generally takes up ~10% of a miRNA's repression capacity.

### GRN stability in relation to expression repression

May pointed out that large interacting systems are difficult to stabilize, contradicting the belief that large systems are inherently stable [1]. The theory may be particularly suited to GRNs because cell functions depend on transcriptome stability [34], and losing even a small number of genes can have severe consequences [35,36]. Furthermore, GRNs are periodically perturbed by cell divisions and a speedy return to equilibrium seems vital to the cells. Because the stability in transcript abundance

has been analyzed mainly at the level of small motifs [5,9], we expanded the analysis to the network level.

In a GRN with  $N$  genes, let  $x_i(t)$  denote the mRNA concentration of gene  $i$  at time  $t$ . When the system is at an equilibrium,  $\frac{dx_i}{dt} = 0$  for all  $i$ 's. Here, we approximated small perturbations near the equilibrium by a linear system (although the system could be non-linear globally):

$$\frac{dx_i}{dt} = B_i - D_i x_i \quad (1 \leq i \leq N) \quad (1)$$

where  $B_i = b_i + S_i$ , with  $b_i$  being the hypothetical transcription rate if influences of all other genes on gene  $i$  cancel out. Also:

$$S_i = \sum_{j=1, j \neq i}^N a_{ij} x_j \tag{2}$$

where  $S_i$  is the aggregate effects of other genes on gene  $i$ , with  $a_{ij}$  being the strength of transcriptional regulation by specific protein–DNA binding of gene  $j$  on gene  $i$ .  $D_i$  is the decay rate of the mRNA of gene  $i$ , which would include possible autoregulation.

Following the approach of May [1] and Allesina *et al.* [2] for studying species interaction network (SIN) stability, we designated the interactions among genes by a matrix,  $M$ . The diagonal element,  $M_{ii}$ , represents the effect of  $x_i$  on itself and the off-diagonal element,  $M_{ij}$ , is the regulation strength of gene  $j$  on gene  $i$ .  $M$  is the Jacobian matrix:

$$M = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(X) & \frac{\partial F_1}{\partial x_2}(X) & \dots & \frac{\partial F_1}{\partial x_n}(X) \\ \frac{\partial F_2}{\partial x_1}(X) & \frac{\partial F_2}{\partial x_2}(X) & \dots & \frac{\partial F_2}{\partial x_n}(X) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial x_1}(X) & \frac{\partial F_n}{\partial x_2}(X) & \dots & \frac{\partial F_n}{\partial x_n}(X) \end{pmatrix} \tag{3}$$

where:

$$F_i(x_1, x_2, \dots, x_n) = F_i(X) = \frac{dx_i}{dt} = b_i + S_i(X) - D_i x_i \quad (1 \leq i \leq N) \tag{4}$$

Given (3) and (4), the elements of the matrix are:

$$M_{ii} = -D_i \text{ and } M_{ij} = a_{ij} \quad (i \neq j). \tag{5}$$

We note that the dynamics of change in  $x_i$  are linearized near equilibrium. Therefore, equations (3–7) are ‘local analysis’ as they approximate the true dynamics only in the small vicinity near equilibrium. The actual equilibrium is usually determined by empirical means, such as RNA sequencing. We first considered a network with only one gene ( $N = 1$ ) where the stability condition is:

$$\frac{\partial F_1}{\partial x_1}(X) = -D_1 < 0. \tag{6}$$

In other words, the slope of  $F_1$  at equilibrium is negative. In this system of  $N = 1$ , the local analysis approximates the equilibrium as:

$$x_1 = \frac{b_1}{D_1}. \tag{7}$$

By increasing  $D_1$  and  $b_1$  in proportion, this system could gain stability without changing the equilibrium and, indeed, the transcription and degradation have been shown to coevolve [37,38].

When  $N > 1$ , the stability of the system is measured in  $N$  orthogonal directions. The equivalent of  $N$  negative slopes pertaining to the stability is expressed as  $N$  negative eigenvalues, which is satisfied if and only if the leading eigenvalue of the matrix  $M$  is negative (‘eigenvalue’ in this paper only refer to the real part of eigenvalue, ‘leading eigenvalue’ is the eigenvalue have largest real part). The leading eigenvalue can be approximated as  $R - D$  [2,39].  $R$ , a function of the interaction strength (i.e. the off-diagonal elements), is the leading eigenvalue of the matrix  $M_0$ , which has the same off-diagonal elements as  $M$  but all diagonal elements are 0.  $D = \frac{\sum_{i=1}^N D_i}{N}$  ( $= \frac{\sum_{i=1}^N M_{ii}}{N}$ ) is the average degradation rate. Therefore, the stability condition is:

$$R - D < 0 \tag{8}$$

While  $R$  and  $D$  are usually obtained numerically, an analytical approximation can be derived from equation 1 of Tang *et al.* [3] when applied to actual transcription data of yeast and mammals (see later sections). Let the connectivity  $r$  be the proportion of non-zero  $M_{ij}$ s ( $i \neq j$ ), and let  $\mu$  and  $\sigma^2$  be the mean and variance of the non-zero off-diagonal elements. When  $\mu \approx 0$  or  $< 0$ , and  $E(M_{ij} \times M_{ji}) \approx 0$ ,

$$R \approx \sigma \sqrt{rN}. \tag{9}$$

Therefore, the stability condition is approximated by:

$$\sigma \sqrt{rN} - D < 0. \tag{10}$$

Equation (10) is suggestive of the roles of miRNAs, which increase  $D$  by catalytically degrading mRNAs [10]. The degradation can be expressed in two parts:

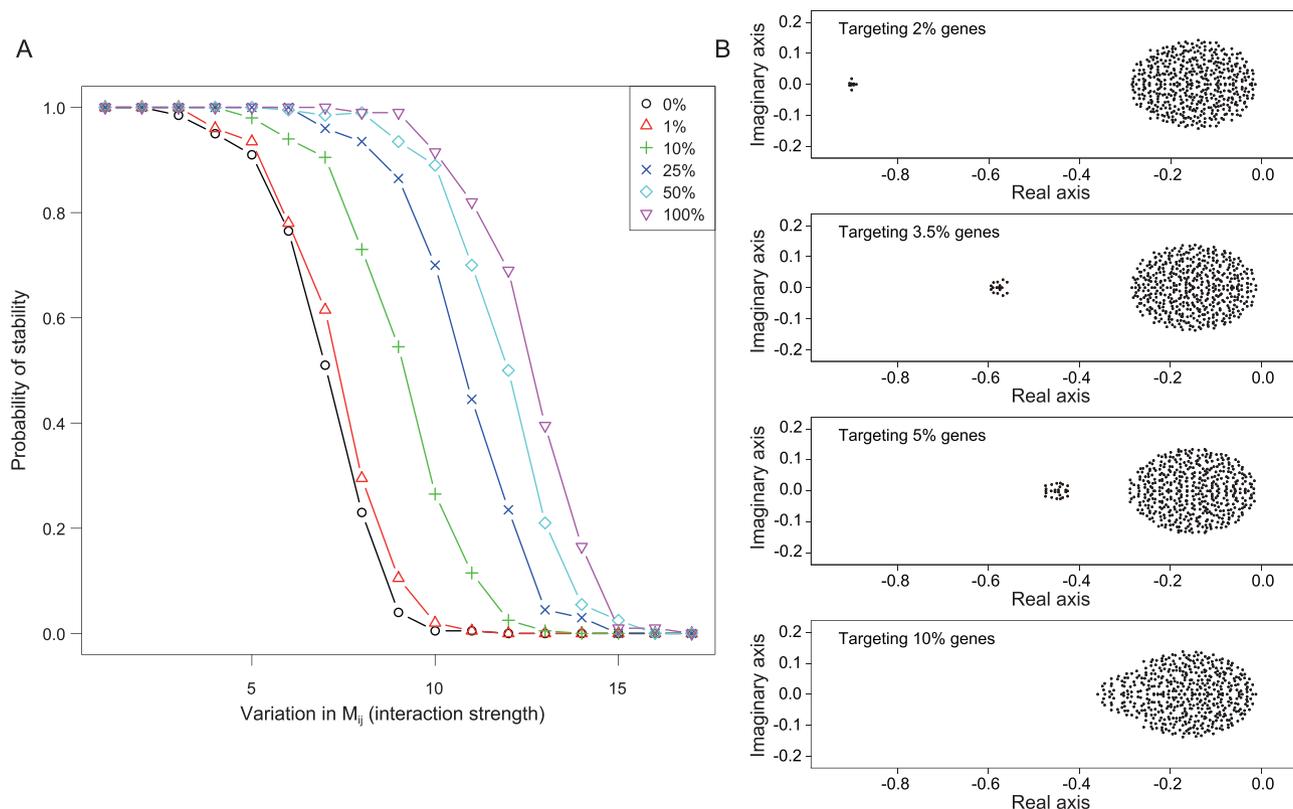
$$-D_i = -(d_i + m_i) \tag{11}$$

where  $d_i$  is the basal decay constant and  $m_i$  is the total effect of all miRNAs on the decay of gene  $i$ . Clearly, a larger  $D_i$  would make the system more stable.

### Properties and predictions of the theory germane to miRNA functions

#### Properties

We now briefly emphasize some key features of the model. First, stability in this study means that the entire system would return to exactly the same



**Figure 2.** GRN stability in relation to the spread of total miRNA repression in the GRN. (A) The  $y$ -axis is the probability of GRN stability, determined by the proportion of cases yielding a negative leading eigenvalue in 200 simulations. The  $x$ -axis is the variation in interaction strength presented as the (relative) standard deviation of  $M_{ij}$ . The repression is distributed over 1–100% of the entire GRN. While the total repression is constant, the probability of stability increases when the effect is spread more broadly over the network. The increase is most rapid from 1–25% and slows down gradually. (B) Distributions of eigenvalues as miRNA targeting becomes more diffuse. If the repression is concentrated on a few genes, only a small fraction of eigenvalues is affected, shown by the outliers on the left. Neither the bulk of the distribution nor the leading eigenvalue is noticeably changed. Only when the targeting is sufficiently broad would the entire distribution shift to the left, thus dragging along the leading eigenvalue.

equilibrium after small perturbations. Hence, the timescale is small for GRNs, in at most tens of minutes. At a larger timescale, the stability does not mean stasis as the equilibrium,  $x$  ( $x = b/D$ ) may gradually change. For example, through stages of development,  $b$  may change while  $D$  keeps the GRN close to the equilibrium. Note that stability and change (either evolution or development) are not antithetical. Second, a stable GRN is necessary but not sufficient for biological stability. Both the proteome and metabolome downstream may have to be stabilized as well. Third, in the theory, the diagonal elements are affected by miRNAs but miRNAs themselves are not in the network, as explained below.

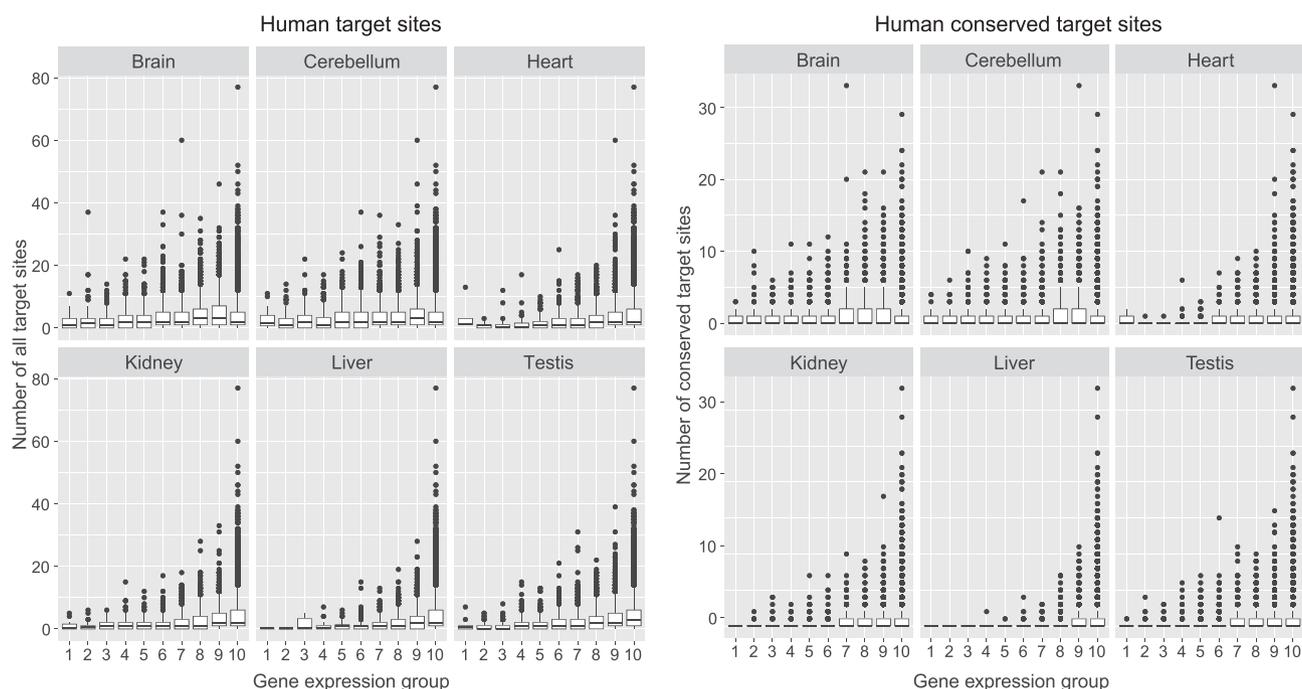
In mammalian cells, the total number of miRNA molecules has been reported to be in the same order of magnitude as the number of mRNAs [40,41]. Therefore, the abundance of miRNA per locus is > 100-fold greater than that of an average mRNA gene. Furthermore, the turnover of miRNAs occurs much more slowly than that of mRNAs. The half-life of

miRNAs in mammalian cells averages about 120 h in comparison with that of mRNAs at 6–8 h [42,43]. The estimates on miRNA half-lives vary partly because miRNA processing usually yields two products [44] and the minor product, so-called miR\*, may be quickly degraded [45,46]. The relevant population of miRs in this study is the major product of the highly expressed miRNAs. The abundance and slow turnover of miRNAs make them nearly unchanged in the time frame of local perturbation. These features also make them uniquely suited to be the canalizing molecules.

### Predictions

The theory makes several predictions, which provide a unified perspective on a suite of miRNA properties that have only been explained individually. One key property in fact has never been explained.

*Broad distribution of the degradation effect of miRNAs: how broad? The theory shows that the average*



**Figure 3.** Number of miRNA target sites on genes with different levels of expression, ranging from high to low from left to right in 10 different groups, each containing 10% of all genes. The left sets of panels are analyses of all targets of 109 miRNA seeds, and the right sets of panels are those of conserved targets. Analyses of two different levels of evolutionary conservation are shown but the pattern is observable in all (see Supplementary Data). For each level, six tissues are analyzed. Note that very highly expressed genes appear to avoid having a very large number of target sites. (See also Figure S3).

degradation,  $D = \frac{\sum_{i=1}^N D_i}{N}$ , or the total degradation,  $D_T = D \times N$ , is a main determinant of GRN stability. This property does not suggest how  $D_T$  should be distributed. In metazoan cells, the total repression effect is distributed among many miRNAs and the effect of each miRNA is further distributed over hundreds of target genes. Qualitatively, the total repression of miRNAs is diffusive over the entire GRN.

Here, we provide a quantitative evaluation of various distributions of miRNA targeting while keeping their aggregate effect constant, at 10% of the total (i.e.  $\sum_{i=1}^N m_i = 10\% \sum_{i=1}^N D_i$ ). Figure 2A shows that, given the same network complexity, the GRN is more likely to be stable when miRNA targeting becomes more diffuse (i.e. targeting more genes with less intensity; see *Methods*). When only 1% of the mRNAs are targeted for repression by all miRNAs, the probability of GRN stability is only slightly higher than a GRN without any repression. On the other hand, when the repression already covers 25% of all transcripts, further spread would have only incremental benefits on the stability.

An intuitive explanation is illustrated in Fig. 2B. When miRNAs target a small percentage of genes, only a few eigenvalues are affected and shifted very far to the negative side. The leading eigenvalue is hardly affected, hence resulting in only marginal improvement in GRN stability. The more diffuse the

targeting, the more eigenvalues are shifted to the left, eventually dragging the leading value down. Estimates of miRNA targeting fall in the range of 25–60% of all mRNAs in human cells [31,32], in reasonable accord with the theoretical prediction of >25%. The next section will explore whether targeting is randomly distributed among all mRNAs.

*Avoidance of very highly expressed mRNAs.* Another aspect is the expression levels of miRNA targets. In this theory, miRNAs are expected to avoid targeting very highly expressed genes. Given a fixed repression capacity, targeting the most highly expressed genes is a wasteful strategy for two reasons. First, highly expressed genes may act as ‘sponges’ [47], soaking up many miRNAs and leaving few available for other less highly expressed targets. Second, high abundance transcripts should be less affected by stochastic fluctuations; for example, after cell divisions. This prediction is supported by empirical data. Figure 3 is a typical example in which relatively few highly expressed genes have a high number of target sites (Fig. S3).

*Preference for transcription factor targeting.* Given the collective targeting by miRNAs, almost all classes of genes are their targets. In the entire GRN, transcription factors (TFs) are the well-known class of targets preferred by miRNAs and there are many

explanations [48–51]. Here, the GRN theory offers a new perspective based on the hierarchical structure within the network. Since TFs constitute a higher level and highly connected sub-network, the theory would predict TFs to be preferentially targeted by miRNAs. A detailed presentation is given in the Supplemental text where various explanations are compared.

### Comparisons of GRNs with and without miRNAs

An alternative approach to the function of miRNAs is to compare GRNs with and without miRNAs. We used human GRNs for the former and yeast GRNs for the latter. While the theory of network stability (Eqs. 8–10) requires sufficient repressions of mRNAs, there are other solutions besides the actions of miRNAs. In fact, repressing mRNAs non-specifically via RNases (as yeast cells do [52]) could be a more powerful means to confer GRN stability. The comparison may reveal different strategies of GRN stabilization and shed further light on the functionality of miRNAs. In comparing human and yeast GRNs, we estimated the diagonal and off-diagonal elements of the interaction matrix  $M$  separately, and then put them together to compare their stability.

### Degradation (the diagonal elements)

The degradation rates of transcripts in many GRNs have been measured, usually by turning off the transcription and monitoring the decay of mRNAs [53,54]. It is known that the mean half-life for yeast mRNAs is  $\sim 15$  min [55], whereas it is 4–8 h for human mRNAs [43]. Figure 4A and B show that the median decay constant (measured in molecules/hour) for yeast mRNAs is 17.4 times larger than that for human mRNAs. Interestingly, when calibrated against their respective cell doubling times of 1.5 and 24 h, the degradation rates are roughly equal (6.87 vs. 6.33) for yeast and human transcription factors (TFs, red shades of Fig. 4A and B).

One might expect the variation in degradation to be driven by natural selection to fine-tune the expression level,  $x_i$ . The regression of  $x_i$  over  $D_i$  is indeed significantly negative for yeast and human genes, but the correlation coefficient is small ( $R^2 = 0.364$  and  $0.101$  for yeast and human, respectively; Fig. 4C and D). Importantly, this is not a simple inverse relationship, as  $x_i$  spans three orders of magnitude and  $D_i$  varies by only one order (Fig. 4C and D). Even when we exclude the tails of the distributions (5% on either end),  $x_i$  still varies 10-fold more than  $D_i$ .

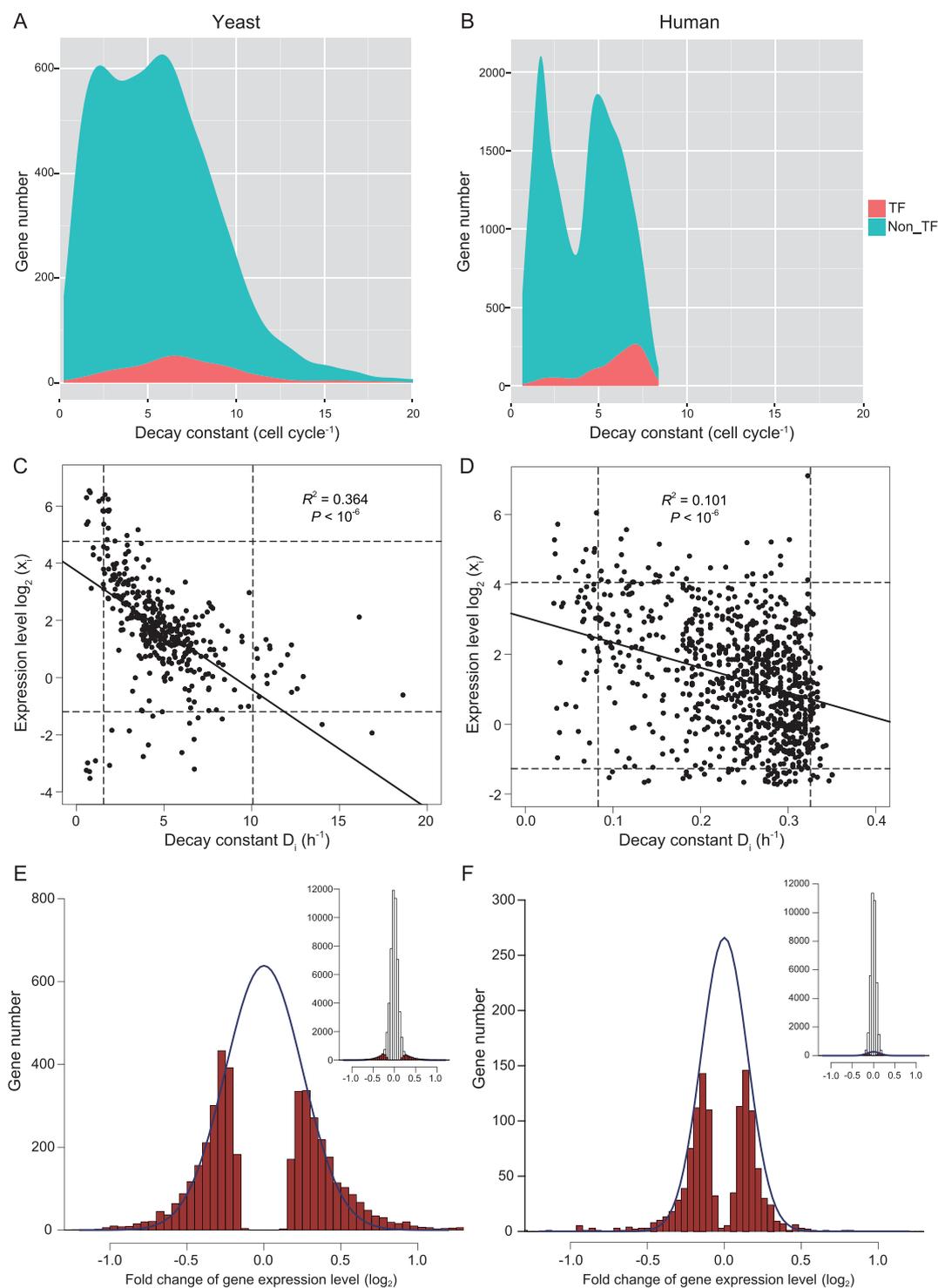
Therefore, the variation in expression is largely due to the variation in synthesis rather than degradation (see Supplement on  $D_i$  variation). If the many cellular components, including miRNAs and RNA-binding proteins, that function in mRNA degradation do not set the level of gene expression, the question is then ‘what roles may mRNA degradation play in the GRN?’.

### Strength of gene interaction (the off-diagonal elements)

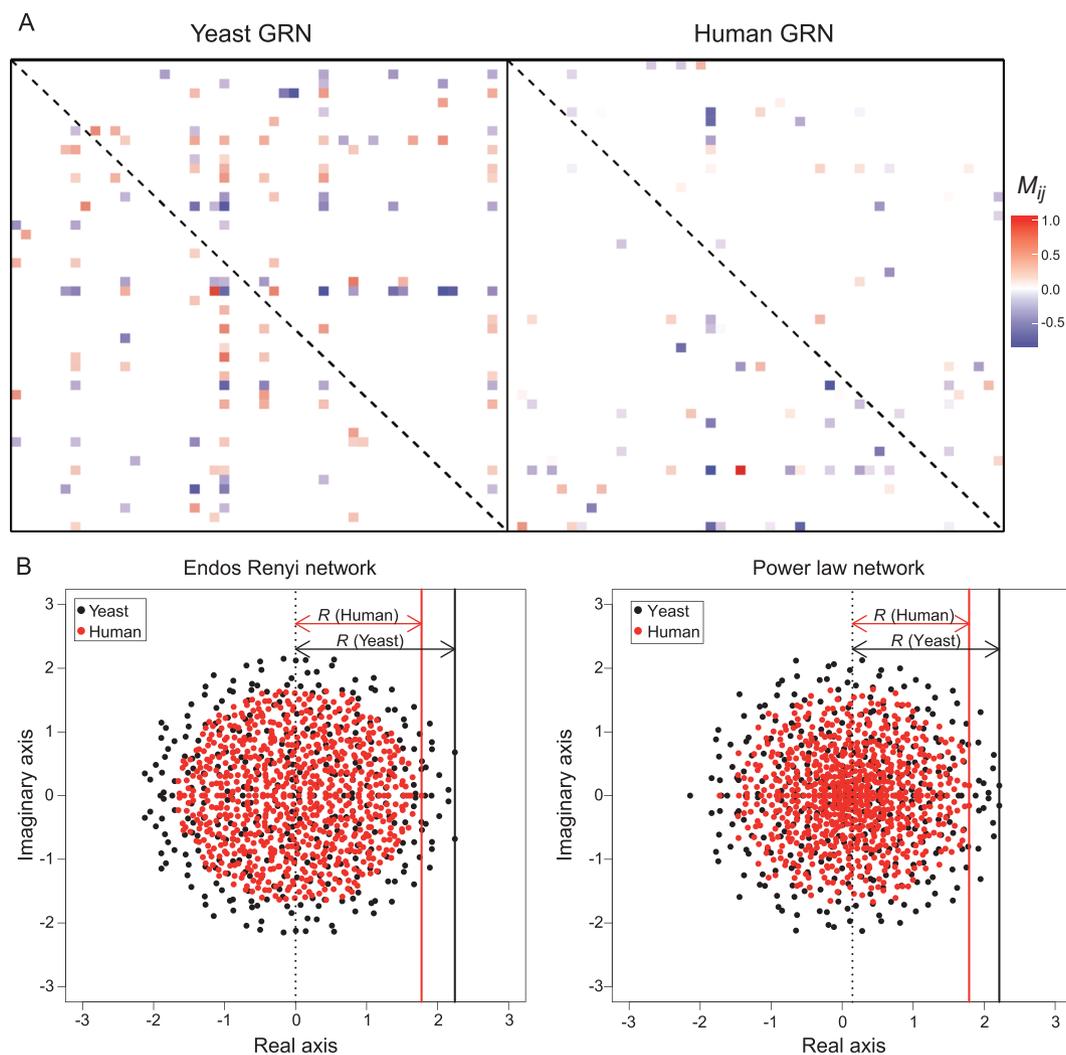
Figure 4A and B show that  $D$  is much larger in the yeast GRN than in human. Since the smaller  $R - D$  is, the greater the stability becomes (Eq. 8), yeast GRN could be much more stable than human GRN. Alternatively, yeast GRN might have a correspondingly larger  $R$  (Eq. 9) and the two GRNs would be comparably stable. We hence analyzed the measurements of  $M_{ij}$  based on experiments that delete or suppress the expression of one TF at a time [56,57]. These two databases are used for both biological and technical reasons, and they are also the largest publicly accessible databases. Nevertheless, the conclusion does not depend on the actual databases used. The TF sub-network is most responsible for the stability of the entire GRN, given its higher position within the hierarchy (see below). The effects of TF deletion/suppression are assayed by transcript analysis (see *Methods* and the Supplemental text for further explanations).

We now describe the construction of the yeast GRN. In order to determine the proper size of the network, we ranked genes by their expression in descending order. The set of the most highly expressed TFs with  $N = 356$  collectively accounted for 99% of total mRNAs.  $N = 356$  is the size of the yeast GRN. The procedures for estimating the regulation strength have been widely reported. Several were used here [56,57] (see *Methods*).

The distribution of the estimated regulation strength is given in Fig. 4E and its inset. Among all interactions, 4234 are significant with  $P < 0.001$ , yielding a connectivity of  $r = 0.076$ . Figure 4E shows the significant regulations with red bars, which are approximated by a normal distribution. The non-significant regulations are set to 0. The normal distribution containing all significant interactions is shown relative to the entire set in the inset. In summary, positive:negative regulation is evenly split with a 0.504:0.496 ratio. The mean ( $\mu$ ) and standard deviation ( $\sigma$ ) are 0.0144 and 0.432 (see legends). The mean and standard deviation of the absolute value of the interaction strength ( $|M_{ij}|$ ) are, respectively, 0.379 and 0.207. We note that, in order to construct the GRN, we estimate the mean, variance and



**Figure 4.** Measurements of degradation and interaction in GRNs. (A, B) Density plot of  $D_i$  (mRNA decay constant). TFs (shaded in red) have higher degradation rates than the rest of the transcriptome (blue). The rates shown are calibrated by the respective cell cycle time, by which the two systems are comparable in degradation. In actual time, yeast mRNAs are degraded 17 times faster than human mRNAs. (C, D) The relationship between the expression level of mRNAs and the decay constant. The y-axis spans three orders of magnitude while the x-axis spans only one order. Hence, the expression level of genes is only marginally affected by the degradation constant. The dotted lines mark 5% and 95% of the distribution. In this restricted range, Y also varies more than X by 10 fold. (E, F) Distribution of the interaction strength between genes in yeast and human GRN. The strength is the change in the abundance of mRNA of gene  $i$  upon the knockout (yeast) or knockdown (human) of gene  $j$ . Significant changes ( $P < 0.001$ ) are marked in red and approximated by a normal distribution marked by the blue line. The inset displays this portion of significantly changed genes relative to all genes.



**Figure 5.** The interaction matrix ( $M$ ) and its eigenvalues. (A) A random  $50 \times 50$  block of the interaction matrix,  $M$ , is shown for the yeast and human GRNs. Off-diagonal elements that are significantly positive or negative (see Fig. 4E and F) are indicated by a color of the heatmap. Diagonal elements are not shown (marked by the dashed line). Note that the yeast GRN is slightly denser than the human GRN. (B) Distribution of the eigenvalues of GRNs, which are complex numbers with real and imaginary parts. The GRN is constructed with the parameters obtained from the measurements of Fig. 4, with the off-diagonal elements following a normal distribution (Fig. 4E and F) and the diagonals set to zero. Two different network structures (random network and power-law) are modeled but the outputs are similar. The leading eigenvalue, marked by a solid vertical line, has a value of  $R$  (see equations (8) and (10)). The GRN stability requires that  $R - D < 0$ .

distribution of  $M_{ij}$ . The identities of specific nodes that are connected are not crucial in determining the identify. This aspect of GRN in relation to miRNA function will be discussed.

In the human GRN, the corresponding  $M_{ij}$  distribution is shown in Fig. 4F where  $N = 746$  and  $r = 0.031$ . The positive:negative split is 0.53:0.47,  $\mu = -0.0322$  and  $\sigma = 0.244$ . The mean and standard deviation of  $|M_{ij}|$  are 0.195 and 0.151 (see *Methods*). A presentation of a small portion ( $50 \times 50$ ) of each of the two GRNs is shown in Fig. 5A. The comparison visually portrays the difference in connectivity between human and yeast GRNs ( $r = 0.031$  vs.  $r = 0.076$ ) with the latter being denser. Therefore, while the human GRN is larger than that of

yeast ( $N = 746$  vs.  $N = 356$ ), the number of connections per node,  $\frac{N^2 r}{N}$ , is very similar with  $Nr = 23.1$  vs. 27.1. The effective sizes [58] are hence similar between the two GRNs. The interaction strength in the human network appears weaker, but only mildly so.

In both networks, significant regulations are not randomly distributed among nodes as a small fraction of nodes are disproportionately more connected than the rest [59,60]. Figure S4 presents the distribution of in-degree connections, or the number of significant regulations going toward a node as required in (2). The observed distribution is closer to the power-law than to random distribution, corroborating previous analyses [59,60].

### GRN stability in yeast vs. human: joint effects of diagonal and off-diagonal elements

With the off-diagonal  $M_{ij}$  values, the eigenvalues of the matrix  $M_0$  can be determined as shown in Fig. 5B for yeast (black dots) and human (red dots) GRNs. Note that the diagonal elements of  $M_0$  are set to 0, in comparison with  $M$  of (3). Marked by a vertical line in Fig. 5B,  $R$  roughly corresponds to the ‘radius’ of the eigenvalue distribution. The two panels of Fig. 5B also show that the eigenvalue distributions are not noticeably changed by the network structure (random vs. power-law interactions).

The  $R$  values in yeast and human GRNs differ very slightly (2.2 vs. 1.8 in Fig. 5B) over a wide range of cut-offs used in the estimation. Given that  $R$  is similar and  $D$  is 15 times larger in yeast,  $R - D$  is much more negative in yeast than in human. In other words, the yeast GRN is much more stable than the human GRN; thus, when perturbed, the yeast GRN should be able to return to equilibrium much more rapidly. Interestingly, yeast cells can divide 15 times faster than human cells (1.5 h vs. 24 h) and, hence, would be perturbed more frequently.

## DISCUSSION

### Comparative strategies of canalization

We applied the May–Wigner theory to two very different GRNs, one with and the other without miRNAs. The yeast GRN may be able to use a simpler strategy for GRN stabilization because unicellular organisms do not have different tissues with different cellular properties. Furthermore, because a typical haploid yeast cell is only 1% as large as an average-sized human cell [61,62], the transcription rate per unit volume can be much higher in yeast than in human cells. For these reasons, the yeast GRN may be able to have non-specific degradation of transcripts that is as rapid as transcription can keep up. This simple strategy would be neither feasible nor necessary for human GRNs. In metazoans, GRNs may need to adjust the strategy of stabilization in different cell types. Their larger cell volume also demands far more transcripts; thus, a high rate of mRNA degradation may stress the supply to a much larger degree. A suitable strategy for mammals would be to degrade mRNAs more selectively and modestly, and miRNAs might have evolved for these reasons.

### Canalization by miRNAs in metazoans

The pervasive weak action of miRNAs has been a contentious issue, giving rise to the view that most targets are biologically irrelevant [17,18,44,63–67]. Since the sum of weak repressions accounts for

>90% of the total activities of miRNAs, it is difficult to reconcile this view with a simple calculation. Instead, the May–Wigner theory suggests that weak repressions can cumulatively contribute to GRN stability [17]. Furthermore, the more diffuse the repression effect, the more stable the network.

In animals, miRNAs may be the true system-level regulators. It is their wiring pattern, rather than specific links between genes, that is germane to their function. Importantly, by stabilizing GRNs, miRNAs would stabilize the downstream phenotype, albeit indirectly. These molecules are hence the likely agents of developmental canalization proposed by Waddington >60 years ago [7,8,68–70]. A recent study [18] found that miRNAs often control the same phenotypes incoherently through multiple target genes. Incoherent control loops are usually associated with stasis rather than change [11–14]. Finally, the contrast between the diffuse actions of miRNAs in animals and the more concentrated repressions in plants [71,72] raises interesting questions about the divergence between plants and animals in relation to the ancestral functions of miRNAs.

*Caveats* - The theory presented here is a sharp break from the conventional views on miRNA functionality. Naturally, caveats need to be heeded. In the first section of the Supplement, most common objections, as well as our rebuttals, are given.

## METHODS

### Number of targets

Target sites of 178 conserved miRNA’s from eight closely related species were predicted in the human genome by seed matching using TargetScan (targetscan.org). Seed matches better than 7mer were considered target sites. The associated random seeds with the same CG content serve as the control. Target sites are stratified by conservation into three levels: all target sites, moderately conserved target sites and strongly conserved target sites (Table S2).

Similar analyses were applied to 94 conserved miRNAs in fly and 167 conserved miRNAs in mouse. Conserved miRNA family lists and untranslated regions (UTRs) were downloaded from TargetScan.org. The numbers of conserved miRNA families and 3’-UTRs are shown in Table S1. Data of miRNA expression come from Landgraf *et al.* and Lyu *et al.* [73,74].

### Strength of repression

To measure the regulation strength of single miRNAs, we examined the change of transcriptome after single miRNA knockout (or knockdown).

The transcriptome data of knockout or knockdown lines were downloaded from the Gene Expression Omnibus database. We collected ~50 data sets to examine their repeatability and selected 5 high-quality (correlation between replication > 0.99) data sets for further analysis (Table S3). The microarray data were processed by the Bioconductor package. To measure the strength of miRNA repression, we calculated the pairwise fold change of each sample, and separated the fold changes into two categories: (i) between an miRNA-deficient transcriptome and a wild-type one, and (ii) between two miRNA-deficient transcriptomes or two wild-type ones. These two categories of fold changes were directly compared by density plot.

DOE sums up all repressions of a certain strength, weighted by the expression level of the target gene. These efforts were categorized into four levels depending on the effect of repression, ranging from <10% to >30%. DOE sums up the repressions across all target genes, weighted by their expression level. Strong repression of >30% generally takes up ~10% of a miRNA's repression capacity.

### Simulations of the degradation effects of miRNAs

Random matrices were generated with the parameters  $N = 500$ ,  $r = 0.1$ ,  $M_{ij} (i \neq j) \sim N(0, a)$  and  $d = 0.1$ . The value  $a$  of  $N(0, a)$  increases from 0.1275 to 0.1675. The probabilities of stability were calculated from 200 simulations. We plotted four cases to illustrate the influence of target numbers on the eigenvalues (Fig. 2A and B).

### Decay rate

The mRNA half-life data were downloaded from public data [43,55], which contained the half-life measurements for 5656 yeast mRNAs and 10 290 human mRNAs. We transformed the half-life measurements into the decay rate using the exponential model:

$$d = \frac{\ln 2}{t}$$

where  $d$  is decay constant in units of  $\text{hours}^{-1}$  and  $t$  is the half-life in units of hours.

For further comparisons, the decay constants of yeast and human mRNAs were shown as the value normalized by the cell cycle time (1.5 h for yeast and 24 h for human), in units of  $\text{cell cycle}^{-1}$ .

### Strengths of gene interactions

To measure the strength of gene–gene transcriptional regulation, we examined the change of transcriptome after single TF knockout or knock-

down. The normalized transcriptome data were downloaded from public data [56,57]. The fold changes between the control and TF-deficient transcriptome were calculated with corresponding  $P$  values. Significant transcription regulations were defined by  $P < 0.001$ . We represented gene–gene interaction strength ( $M_{ij}$ ) with fold changes for significantly differential expressions.

### Construction of matrix $M$

The size of the matrix,  $N$ , is determined by the number of TFs that in aggregate account for 99% of the total transcripts. We generated two kinds of network: an Endos Renyi network and a power-law network, using R package(igraph) in order to determine the non-zero elements in the matrix. The values of non-zero elements in the network were sampled from the gene–gene regulations estimated above. The eigenvalues were calculated and plotted in the complex plane (Fig. 5B).

### Considerations of selective pressure

The arguments presented in this study are mechanistic in nature. For biological systems, these arguments have to incorporate selective forces as is done explicitly in other studies [75–77]. These considerations also underlie the overall approach of this current study.

### SUPPLEMENTARY DATA

Supplementary data are available at [NSR](#) online.

### ACKNOWLEDGEMENTS

We thank Jian Lu, Jacopo Grilli, Si Tang and Gyuri Barabas for help and comments at various stages of manuscript preparation.

### FUNDING

This work was supported by the National Natural Science Foundation of China (31730046 and 91731301), the National Basic Research Program (973 Program) of China (2014CB542006), the 985 Project (33000–18841204) and National Institute of Health, USA (S R01 GM058686-09 to WU, Chung-I).

### AUTHOR CONTRIBUTIONS

Conceptualization: C.W., Y.C. and Y.S.; methodology: Y.C., C.W. and S.A.; formal analysis: Y.C.; writing (original draft): Y.C., C.W. and Y.S.; writing (review and editing): C.W., Y.C., Y.S. and S.A.; supervision: C.W. and Y.C.

*Conflict of interest statement.* None declared.

## REFERENCES

- May RM. Will a large complex system be stable? *Nature* 1972; **238**: 15–7.
- Allesina S and Tang S. Stability criteria for complex ecosystems. *Nature* 2012; **483**: 205–8.
- Tang S, Pawar S and Allesina S. Correlation between interaction strengths drives stability in large ecological networks. *Ecol Lett* 2014; **17**: 1094–100.
- Waddington CH. Canalization of development and the inheritance of acquired characters. *Nature* 1942; **150**: 563–5.
- Hornstein E and Shomron N. Canalization of development by microRNAs. *Nat Genet* 2006; **38**: S20–4.
- Peterson KJ, Dietrich MR and McPeck MA. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays* 2009; **31**: 736–47.
- Wu CI, Shen Y and Tang T. Evolution under canalization and the dual roles of microRNAs: a hypothesis. *Genome Res* 2009; **19**: 734–43.
- Posadas DM and Carthew RW. MicroRNAs and their roles in developmental canalization. *Curr Opin Genet Dev* 2014; **27**: 1–6.
- Shen Y, Lv Y and Huang L *et al.* Testing hypotheses on the rate of molecular evolution in relation to gene expression using microRNAs. *Proc Natl Acad Sci USA* 2011; **108**: 15942–7.
- Eichhorn SW, Guo H and McGeary SE *et al.* mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell* 2014; **56**: 104–15.
- Tsang J, Zhu J and van Oudenaarden A. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol Cell* 2007; **26**: 753–67.
- Herranz H and Cohen SM. MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes Dev* 2010; **24**: 1339–44.
- Hausser J and Zavolan M. Identification and consequences of miRNA-target interactions—beyond repression of gene expression. *Nat Rev Genet* 2014; **15**: 599–612.
- Blevins R, Bruno L and Carroll T *et al.* MicroRNAs regulate cell-to-cell variability of endogenous target gene expression in developing mouse thymocytes. *PLoS Genet* 2015; **11**: e1005020.
- Salmena L, Poliseno L and Tay Y *et al.* A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011; **146**: 353–8.
- Thomson DW and Dinger M E. Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet* 2016; **17**: 272–83.
- Zhao Y, Shen X and Tang T *et al.* Weak regulation of many targets is cumulatively powerful—an evolutionary perspective on microRNA functionality. *Mol Biol Evol* 2017; **34**: 3041–6.
- Liufu Z, Zhao Y and Guo L *et al.* Redundant and incoherent regulations of multiple phenotypes suggest microRNAs' role in stability control. *Genome Res* 2017; **27**: 1665–73.
- Hunter SE, Finnegan EF and Zisoulis DG *et al.* Functional genomic analysis of the let-7 regulatory network in *Caenorhabditis elegans*. *PLoS Genet* 2013; **9**: e1003353.
- Ecsedi M, Rausch M and Grosshans H. The let-7 microRNA directs vulval development through a single target. *Dev Cell* 2015; **32**: 335–44.
- Bagga S, Bracht J and Hunter S *et al.* Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* 2005; **122**: 553–63.
- Johnson SM, Grosshans H and Shingara J *et al.* RAS is regulated by the let-7 microRNA family. *Cell* 2005; **120**: 635–47.
- Li Y, Wang F and Lee JA *et al.* MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes Dev* 2006; **20**: 2793–805.
- Karres JS, Hilgers V and Carrera I *et al.* The conserved microRNA miR-8 tunes atrophin levels to prevent neurodegeneration in *Drosophila*. *Cell* 2007; **131**: 136–45.
- Varghese J and Cohen SM. MicroRNA miR-14 acts to modulate a positive autoregulatory loop controlling steroid hormone signaling in *Drosophila*. *Genes Dev* 2007; **21**: 2277–82.
- Hyun S, Lee JH and Jin H *et al.* Conserved MicroRNA miR-8/miR-200 and its target USH/FOG2 control growth by regulating PI3K. *Cell* 2009; **139**: 1096–108.
- Iovino N, Pane A and Gaul U. miR-184 has multiple roles in *Drosophila* female germline development. *Dev Cell* 2009; **17**: 123–33.
- Bejarano F, Smibert P and Lai EC. miR-9a prevents apoptosis during wing development by repressing *Drosophila* LIM-only. *Dev Biol* 2010; **338**: 63–73.
- Luo W and Sehgal A. Regulation of circadian behavioral output via a MicroRNA-JAK/STAT circuit. *Cell* 2012; **148**: 765–79.
- Xu J, Zhang R and Shen Y *et al.* The evolution of evolvability in microRNA target sites in vertebrates. *Genome Res* 2013; **23**: 1810–6.
- Friedman RC, Farh KK and Burge CB *et al.* Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009; **19**: 92–105.
- Lewis BP, Burge CB and Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; **120**: 15–20.
- Hsu SD, Lin FM and Wu WY *et al.* miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res* 2011; **39**: D163–9.
- Jeong H, Mason SP and Barabasi AL *et al.* Lethality and centrality in protein networks. *Nature* 2001; **411**: 41–2.
- Greenberg F, Guzzetta V and Montes de Oca-Luna R *et al.* Molecular analysis of the Smith-Magenis syndrome: a possible contiguous-gene syndrome associated with del(17)(p11.2). *Am J Hum Genet* 1991; **49**: 1207–18.
- Brook-Carter PT, Peral B and Ward CJ *et al.* Deletion of the TSC2 and PKD1 genes associated with severe infantile polycystic kidney disease—a contiguous gene syndrome. *Nat Genet* 1994; **8**: 328–32.
- Dori-Bachash M, Shema E and Tirosh I. Coupled evolution of transcription and mRNA degradation. *PLoS Biol* 2011; **9**: e1001106.
- Haimovich G, Peral B and Ward CJ *et al.* Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* 2013; **153**: 1000–11.
- O'Rourke S and Renfrew D. Low rank perturbations of large elliptic random matrices. *Electron J Probab* 2014; **19**: 1–65.
- Liang Y, Ridzon D and Wong L *et al.* Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics* 2007; **8**: 166.
- Subkhankulova T, Gilchrist MJ and Livesey FJ. Modelling and measuring single cell RNA expression levels find considerable transcriptional differences among phenotypically identical cells. *BMC Genomics* 2008; **9**: 268.
- Gantier MP, McCoy CE and Rusinova I *et al.* Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. *Nucleic Acids Res* 2011; **39**: 5692–703.
- Tani H, Mizutani R and Salam KA *et al.* Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res* 2012; **22**: 947–56.
- Zhao Y, Lin P and Liufu Z *et al.* Regulation of large number of weak targets—new insights from twin-microRNAs. *Genome Biol Evol* 2018; **10**: 1255–64.
- Guo Y, Liu J and Elfenbein SJ *et al.* Characterization of the mammalian miRNA turnover landscape. *Nucleic Acids Res* 2015; **43**: 2326–41.
- Marzi MJ, Ghini F and Cerruti B *et al.* Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome Res* 2016; **26**: 554–65.

47. Ebert MS, Neilson JR and Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* 2007; **4**: 721–6.
48. Cui QH, Yu ZB and Purisima EO *et al.* Principles of microRNA regulation of a human cellular signaling network. *Mol Syst Biol* 2006; **2**: 46.
49. Chen CY, Chen ST and Fuh CS *et al.* Coregulation of transcription factors and microRNAs in human transcriptional regulatory network. *BMC Bioinformatics* 2011; **12**: S41.
50. Croft L, Szklarczyk D and Jensen LJ *et al.* Multiple independent analyses reveal only transcription factors as an enriched functional class associated with microRNAs. *BMC Syst Biol* 2012; **6**: 90.
51. Dannemann M, Prüfer K and Lizano E *et al.* Transcription factors are targeted by differentially expressed miRNAs in primates. *Genome Biol Evol* 2012; **4**: 552–64.
52. Wada T and Becskei A. Impact of methods on the measurement of mRNA turnover. *Int J Mol Sci* 2017; **18**: E2723.
53. Friedel CC, Dolken L and Ruzsics Z *et al.* Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res* 2009; **37**: e115.
54. Yang E, van Nimwegen E and Zavolan M *et al.* Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res* 2003; **13**: 1863–72.
55. Sun M, Schwalb B and Schulz D *et al.* Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* 2012; **22**: 1350–9.
56. Cusanovich DA, Pavlovic B and Pritchard JK *et al.* The functional consequences of variation in transcription factor binding. *PLoS Genet* 2014; **10**: e1004226.
57. Kemmeren P, Sameith K and van de Pasch LA *et al.* Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* 2014; **157**: 740–52.
58. Davis MP, Abreu-Goodger C and van Dongen S *et al.* Large-scale identification of microRNA targets in murine Dgcr8-deficient embryonic stem cell lines. *PLoS One* 2012; **7**: e41762.
59. Wagner A and Fell DA. The small world inside large metabolic networks. *Proc Biol Sci* 2001; **268**: 1803–10.
60. Guelzim N, Bottani S and Bourguin P *et al.* Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet* 2002; **31**: 60–3.
61. Jorgensen P, Nishikawa JL and Breitkreutz BJ *et al.* Systematic identification of pathways that couple cell growth and division in yeast. *Science* 2002; **297**: 395–400.
62. Puck TT, Marcus PI and Cieciora SJ. Clonal growth of mammalian cells in vitro; growth characteristics of colonies from single HeLa cells with and without a feeder layer. *J Exp Med* 1956; **103**: 273–83.
63. Seitz H. Redefining microRNA targets. *Curr Biol* 2009; **19**: 870–3.
64. Pinzon N, Li B and Martinez L *et al.* MicroRNA target prediction programs predict many false positives. *Genome Res* 2017; **27**: 234–45.
65. Ma F, Lin P and Chen Q *et al.* Direct measurement of pervasive weak repression by microRNAs and their role at the network level. *BMC Genomics* 2018; **19**: 362.
66. Seitz H. On the number of functional microRNA targets. *Mol Biol Evol* 2019; **36**: 1596–7.
67. Zhao Y, Shen X and Tang T *et al.* Weak regulation of many targets is cumulatively powerful- a reply to Seitz on microRNA functionality. *Mol Biol Evol* 2019; **36**: 1598–9.
68. Stark A, Brennecke J and Bushati N *et al.* Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* 2005; **123**: 1133–46.
69. Li X, Cassidy JJ and Reinke CA *et al.* A microRNA imparts robustness against environmental fluctuation during development. *Cell* 2009; **137**: 273–82.
70. Schmiedel JM, Klemm SL and Zheng Y *et al.* MicroRNA control of protein expression noise. *Science* 2015; **348**: 128–32.
71. Jones-Rhoades MW and Bartel DP. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 2004; **14**: 787–99.
72. Jones-Rhoades MW, Bartel DP and Bartel B. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol* 2006; **57**: 19–53.
73. Landgraf P, Rusu M and Sheridan R *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 2007; **129**: 1401–14.
74. Lyu Y, Shen Y and Li H *et al.* New microRNAs in *Drosophila*—birth, death and cycles of adaptive evolution. *PLoS Genet* 2014; **10**: e1004096.
75. Xu S, He Z and Zhang Z *et al.* The origin, diversification and adaptation of a major mangrove clade (Rhizophoraceae) revealed by whole-genome sequencing. *Natl Sci Rev* 2017; **4**: 721–34.
76. Wen H, Wang H-Y and He X *et al.* On the low reproducibility of cancer studies. *Natl Sci Rev* 2018; **5**: 619–24.
77. Wang H-Y, Chen Y and Tong D *et al.* Is the evolution in tumors Darwinian or non-Darwinian? *Natl Sci Rev* 2017; **5**: 15–7.