Translating ceRNA susceptibilities into correlation functions

Araks Martirosyan,[1,2] Matteo Marsili,[3] and Andrea De Martino[4,5,*]

[1]*Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy*
[2]*VIB-KU Leuven Center for Brain & Disease Research, Leuven, Belgium*
[3]*The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy*
[4]*Soft and Living Matter Lab, Institute of Nanotechnology (CNR-NANOTEC), Consiglio Nazionale delle Ricerche, Rome, Italy*
[5]*Human Genetics Foundation, Turin, Italy*

Competition to bind microRNAs induces an effective positive crosstalk between their targets, therefore known as 'competing endogenous RNAs' or ceRNAs. While such an effect is known to play a significant role in specific situations, estimating its strength from data and, experimentally, in physiological conditions appears to be far from simple. Here we show that the susceptibility of ceRNAs to different types of perturbations affecting their competitors (and hence their tendency to crosstalk) can be encoded in quantities as intuitive and as simple to measure as correlation functions. We confirm this scenario by extensive numerical simulations and validate it by re-analyzing PTEN's crosstalk pattern from TCGA breast cancer database. These results clarify the links between different quantities used to estimate the intensity of ceRNA crosstalk and provide new keys to analyze transcriptional datasets and effectively probe ceRNA networks *in silico*.

INTRODUCTION

MicroRNAs (miRNAs) are small non coding RNA (ncRNA) molecules that post-transcriptionally regulate a significant portion of the eukaryotic transcriptome via sequence-specific, protein-mediated binding in the cytoplasm [1]. Their primary effects on coding transcripts consist in inhibiting translation and fostering degradation [2]. Long ncRNAs, instead, can transiently sequester miRNAs, thereby altering their availability and overall repressive potential [3]. Following early observations concerning small regulatory RNAs in plants and bacteria [4, 5], competition to bind miRNAs has been hypothesized to cause an effective positive interaction ('crosstalk') between their coding and/or non-coding targets that may directly affect protein levels [6] (see Fig. 1A,B). Several experimental and modeling studies have clarified the conditions under which such a scenario may become biologically relevant, highlighting specifically how molecular levels and kinetic heterogeneities may control it [7–13]. So far, such a 'ceRNA effect' (whereby ceRNA stands for 'competing endogenous RNA') has been quantitatively validated in cases of differentiation [14], disease [15] or in presence of unphysiologically large transcriptional inputs [16]. Its significance in standard physiological conditions is therefore subject to scrutiny [17].

A major difficulty in detecting the ceRNA effect unambiguously in experiments or data lies in the fact that it should be disentangled from other mechanisms that may bear a similar impact, i.e. an effective positive coupling, on transcripts. Imagine a network of N ceRNA species interacting with M miRNA species. ceRNA levels m_i ($i = 1, \ldots, N$) fluctuate stochastically in time due to random synthesis and degradation events and to interactions with miRNAs, whose levels are also subject to random fluctuations. Denoting by $\langle \cdot \rangle$ the time-average in the steady state, an effective ceRNA-ceRNA dependence can be signaled by a statistical correlation coefficient such as Pearson's [8], i.e.

$$\rho_{ij} = \frac{\langle m_i m_j \rangle - \langle m_i \rangle \langle m_j \rangle}{[(\langle m_i^2 \rangle - \langle m_i \rangle^2)(\langle m_j^2 \rangle - \langle m_j \rangle^2)]^{1/2}}, \quad (1)$$

with the idea that, if ρ_{ij} is large enough, a perturbation altering the level of ceRNA j will cause part of the miRNA population to move from one target to the other, effectively broadcasting the perturbation from ceRNA j to ceRNA i through miRNA-mediated interactions. A more direct description of this mechanism is attained instead via *susceptibilities* like [7]

$$\chi_{ij} = \frac{\partial \langle m_i \rangle}{\partial b_j} \geq 0 \ , \quad (2)$$

where b_j stands for the transcription rate of ceRNA j. χ_{ij} quantifies the shift in the mean level of ceRNA i caused by a (small) variation in b_j, and a large χ_{ij} (assuming no direct control of ceRNA i by ceRNA j) points to miRNA-mediated crosstalk between ceRNAs i and j (see Fig. 1C).

While both χ_{ij} and ρ_{ij} capture aspects of ceRNA crosstalk seen in experiments, their underlying physical meaning is *a priori* different. Fluctuating miRNA levels naturally correlate co-regulated targets, so that a large ρ_{ij} is obtained when both ceRNAs respond to the stochastic dynamics of their regulator. This however does not necessarily imply a large χ_{ij}. In fact, χ_{ij} can be large even in absence of fluctuations in miRNA levels, i.e. as a consequence of competition alone. In such conditions, ρ_{ij} vanishes. χ_{ij} has indeed been found to be asymmetric under exchange of its indices (i.e. $\chi_{ij} \neq \chi_{ji}$ in general) [7], at odds with ρ_{ij} which is necessarily symmetric. It would therefore be important to clarify how quantities like (1) and (2) are related in miRNA-ceRNA networks, especially to understand whether responses to perturbations (a central quantity of interest for many potential applications of the ceRNA effect) can be encoded in quantities as intuitive and as simple to measure experimentally or from data as a Pearson correlation coefficient.

Here we show that the information conveyed by χ_{ij} is indeed captured by a correlation function similar to ρ_{ij}. On the other hand, ρ_{ij} is linked to a susceptibility, i.e. to the response

---
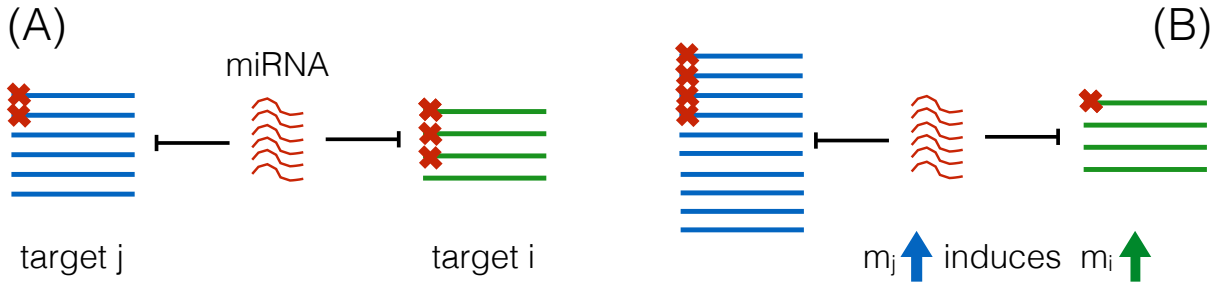
* andrea.demartino@roma1.infn.it

FIG. 1. (A) A miRNA species interacts with two different target RNAs who are competing to bind it, partially repressing their pools (repression being represented by red crosses). (B) In specific conditions, an increase in the level of one of the targets can induce a de-repression of the competitor, thereby establishing an effective positive coupling between the targets. Note that a similar effect can in principle be obtained between ceRNAs that are not co-regulated by the same miRNA species through chains of miRNA-mediated interactions.

of a target to a perturbation altering the level of its competitor, but the perturbation concerns the intrinsic *decay* rate of the competitor rather than its transcription (as is the case for χ_{ij}). In the following, we will derive these results and validate them by computer simulations and gene expression data analysis, and explore their consequences.

MATERIALS AND METHODS

Numerical simulations were performed using the Gillespie algorithm [18], an implementation of which, for a miRNA-ceRNA network, is available from https://github.com/araksm/ceRNA/. Gene expression data analysis was performed starting from 1098 breast cancer samples obtained from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/; project ID: TCGA-BRCA).

RESULTS

Theory

We start from the dynamics of molecular populations in a miRNA-ceRNA network, denoting by m_i the level of ceRNA species i (ranging from 1 to N), by μ_a the level of miRNA species a (ranging from 1 to M), and by c_{ia} the levels of miRNA-ceRNA complexes. In the deterministic limit where stochastic fluctuations are neglected, the time evolution of concentration variables is described by

$$
\dot{\mu}_a = \beta_a - \delta_a \mu_a - \sum_i k_{ia}^+ m_i \mu_a + \sum_i (k_{ia}^- + \kappa_{ia}) c_{ia}
$$

$$
\dot{m}_i = b_i - d_i m_i - \sum_a k_{ia}^+ m_i \mu_a + \sum_a k_{ia}^- c_{ia} \qquad (3)
$$

$$
\dot{c}_{ia} = k_{ia}^+ m_i \mu_a - \frac{c_{ia}}{\tau_{ia}} \quad ,
$$

with the different parameters denoting intrinsic synthesis (b_i, β_a) and degradation rates (d_i, δ_a), complex association/dissociation rates ($k_{ia}^\pm$) and complex processing rates (σ_{ia}

and κ_{ia} for stoichiometric and catalytic processing, respectively), while $\tau_{ia} = (\sigma_{ia} + \kappa_{ia} + k_{ia}^-)^{-1}$ represents the mean lifetime of the complex formed by miRNA a and ceRNA i. We note that if the mean lifetime of complexes is much shorter than that of free molecular species, i.e. if $\tau_{ia} \ll 1/d_i$ and $\tau_{ia} \ll 1/\delta_a$ for each i and a, miRNA-ceRNA complexes achieve a steady state much faster than miRNA and ceRNA levels. In such conditions, $\dot{c}_{ia} \simeq 0$ and one can eliminate complexes from (3) by replacing c_{ia} with its steady state value

$$
\langle c_{ia} \rangle = k_{ia}^+ \tau_{ia} m_i \mu_a \quad . \qquad (4)
$$

For $\kappa_{ia} + k_{ia}^- \ll \sigma_{ia}$ (i.e. when stoichiometric degradation without miRNA recycling is the dominant channel of complex processing), this allows to re-cast (3) in the form (see Supplementary Text 1)

$$
\dot{m}_i \simeq -m_i \frac{\partial L}{\partial m_i} \quad ,
$$
$$
\dot{\mu}_a \simeq -\mu_a \frac{\partial L}{\partial \mu_a} \quad , \qquad (5)
$$

where L is a function of all miRNA levels $\boldsymbol{\mu} = \{\mu_a\}$ and all ceRNA levels $\mathbf{m} = \{m_i\}$ given by

$$
L = -\sum_i (b_i \log m_i - d_i m_i)
$$
$$
- \sum_a (\beta_a \log \mu_a - \delta_a \mu_a) + \sum_{i,a} k_{ia}^+ m_i \mu_a \quad . \quad (6)
$$

One easily sees (see Supplementary Text 2) that L decreases along trajectories of (5), implying that its minimum describes the physically relevant steady state of (3) with $\mathbf{m} \neq \mathbf{0}$ and $\boldsymbol{\mu} \neq \mathbf{0}$.

If intrinsic molecular noise (arising from stochastic transcription and degradation events and from titration due to miRNA-ceRNA interactions) is added to (3), after a transient molecular levels will eventually stabilize and fluctuate over time around the steady state described by the minimum of L. We are interested in finding a compact and intuitive mathematical form for the correlations arising between the different

components in such conditions. Molecular noise is Poissonian, namely the strength of fluctuations affecting each variable is proportional to the square root of mean molecular levels (see e.g. [13] for an explicit representation in the context of a miRNA-ceRNA network), which makes our goal especially challenging. However we will see that the effects of molecular noise can be remarkably well approximated by a uniform "effective temperature" T representing the strength of fluctuations affecting all molecular species involved. In this case, one can describe fluctuations around the steady state as thermal fluctuations around a Boltzmann-Gibbs equilibrium state. This allows to compute averages of generic functions of $\mathbf{m}$ and $\boldsymbol{\mu}$ as "thermal averages", i.e.

$$\langle f \rangle = \frac{1}{Z(T)} \sum_{\mathbf{m},\boldsymbol{\mu}} f(\mathbf{m}, \boldsymbol{\mu}) e^{-L/T} \ , \qquad (7)$$

where

$$Z(T) = \sum_{\mathbf{m},\boldsymbol{\mu}} e^{-L/T} \qquad (8)$$

is a normalization factor, the deterministic limit being obtained for $T \to 0$. In particular, defining

$$\langle fg \rangle_c \equiv \langle fg \rangle - \langle f \rangle \langle g \rangle \ , \qquad (9)$$

by straightforward calculations one finds

$$\langle m_i \rangle = -T \frac{\partial}{\partial d_i} \log Z(T) \ , \qquad (10)$$

$$\langle m_i m_j \rangle_c = -T \frac{\partial \langle m_i \rangle}{\partial d_j} \equiv -T \omega_{ij} \ , \qquad (11)$$

$$\langle m_i \log m_j \rangle_c = T \chi_{ij} \ . \qquad (12)$$

Therefore, in this approximation, the susceptibility χ_{ij} [Eq. (2)] is linked to the correlation function [Eq. (12)]

$$X_{ij} = \langle m_i \log m_j \rangle_c \qquad (13)$$

which, as χ_{ij}, is not symmetric under the exchange of i and j, while the ceRNA-ceRNA covariance

$$C_{ij} = \langle m_i m_j \rangle_c \qquad (14)$$

is tied to the susceptibility ω_{ij} quantifying the change in $\langle m_i \rangle$ induced by a (small) change of the *intrinsic degradation rate* d_j of ceRNA j [Eq. (11)]. (Note that $\omega_{ij} \leq 0$.) The constant linking these quantities is the temperature T quantifying the strength of the uniform "effective noise".

Somewhat unexpectedly, the above results suggest that the susceptibility ω_{ij} must be symmetric under exchange of i and j, i.e., for instance, if the level of ceRNA i is altered by changing the intrinsic degradation rate of ceRNA j, then the reverse is also true. To check this property, one can calculate ω_{ij} explicitly for a system formed by N ceRNA species interacting with a single miRNA species at steady state by following a different route, specifically along the lines of [7]. Considering the repression strength to which ceRNAs i and j are subject at a given (mean) level $\langle \mu_1 \rangle$ of miRNA species 1 ($M = 1$

in this case), one finds, for each ceRNA, a soft "threshold" value of $\langle \mu_1 \rangle$, denoted by $\mu_{0,i1} \simeq d_i / k_{i1}^+$, such that i is unrepressed (resp. repressed or susceptible to changes in $\langle \mu_1 \rangle$) if $\langle \mu_1 \rangle \ll \mu_{0,i1}$ (resp. $\gg \mu_{0,i1}$ or $\simeq \mu_{0,i1}$). A direct calculation (see Supplementary Text 3) shows that ω_{ij} can attain large values only if both ceRNAs are susceptible to $\langle \mu_1 \rangle$, in which case one has ($i \neq j$)

$$\omega_{ij} \simeq -\frac{\langle \mu_1 \rangle}{\mu_{0,i1} \mu_{0,j1}} \frac{b_i b_j}{d_i d_j} \left(\delta_1 + \sum_\ell a_\ell \frac{b_\ell k_{\ell 1}^+}{d_\ell} \right)^{-1} \ , \qquad (15)$$

where $a_\ell = 0, 1, 1/4$ if ceRNA ℓ is repressed, unrepressed or susceptible, respectively. Eq. (15) confirms that ω_{ij} is indeed symmetric under exchange of i and j.

Concerning the approximations under which the the above results were obtained, we remark that we started by considering (3) in the limit of (i) fast complex equilibration, and (ii) miRNA-ceRNA complex processing dominated by the stoichiometric channel, with the former playing the key role in deriving the function L (see Supplementary Text 1). We note however that the overall scenario just described also holds for when complexes evolve over time scales much longer than those of free molecular levels, i.e. for $\tau_{ia} \gg 1/d_i$ and $\tau_{ia} \gg 1/\delta_a$. In particular, (3) can again be re-cast in the form of (5) with L given by (6), albeit with re-scaled transcription rates (see Supplementary Text 4 for details).

Therefore we conclude that, as long as molecular noise can be approximated by a uniform effective temperature,

(i) the ceRNA-ceRNA covariance $C_{ij} = \langle m_i m_j \rangle_c$ is a proxy for the susceptibility ω_{ij}, and

(ii) the correlation function $X_{ij} = \langle m_i \log m_j \rangle_c$ is a proxy for the susceptibility χ_{ij}.

Validation

We have validated the above scenario by simulating a small network involving 2 ceRNA and a single miRNA species via the Gillespie algorithm [18], where molecular noise is accounted for explicitly (see Supplementary Text 5). Results are summarized in Fig. 2, where we compare ω_{12}, ω_{21} and $C_{12} \equiv C_{21}$ on on hand, and χ_{12}, χ_{21}, X_{12} and X_{21} on the other, as computed from simulations (i.e. with the actual molecular noise), against the theoretical predictions. We considered three scenarios for the mean lifetime of miRNA-ceRNA complexes, namely those covered by the theory (i.e. complex equilibration much faster and much slower than the equilibration of miRNA and ceRNA levels) as well as the intermediate case where characteristic timescales are comparable for all variables.

One sees that theoretical predictions obtained in the "thermal noise" approximation agree remarkably well with simulations including the actual molecular noise. In particular, the full correspondence between the susceptibilities ω_{ij} and χ_{ij} and the (re-scaled) correlation functions C_{ij} and, respectively,
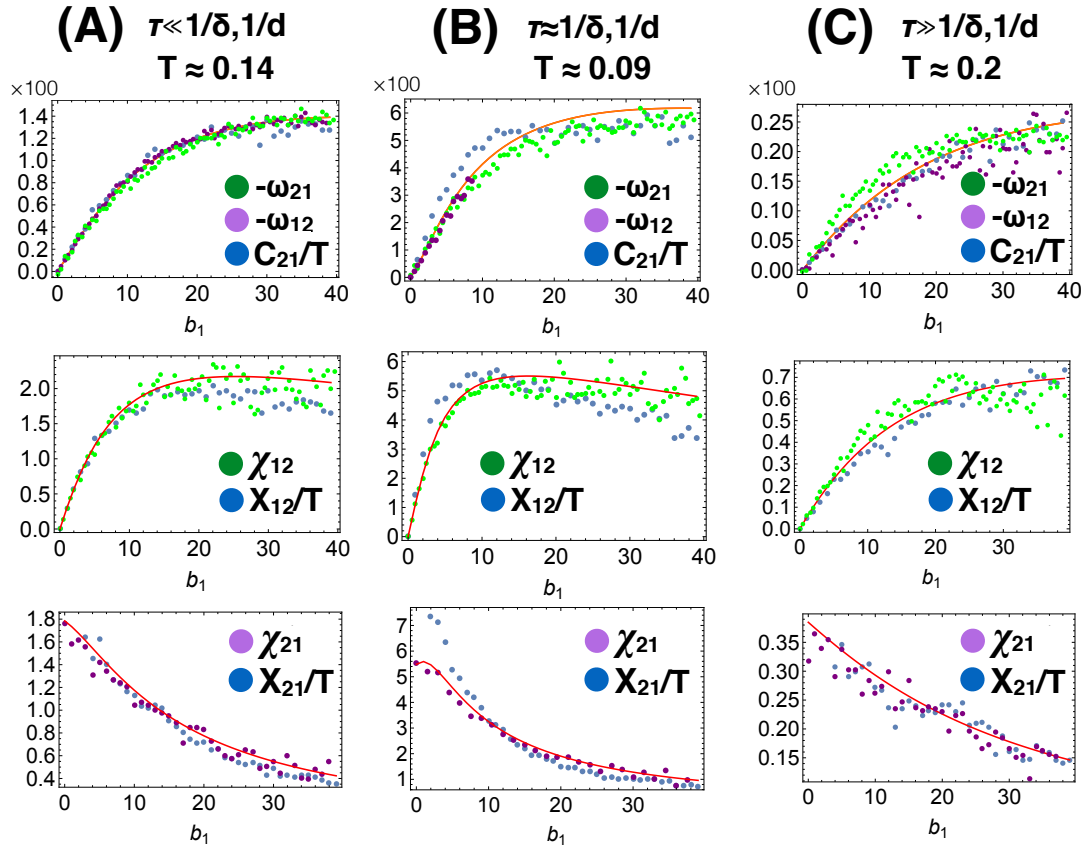
FIG. 2. Results from Gillespie simulations (markers) versus analytical predictions for susceptibilities (lines) for a miRNA-ceRNA network with $N = 2$ ceRNA and $M = 1$ miRNA species and different values of the transcription rate of ceRNA species 1, b_1. Numerical susceptibilities have been obtained by perturbing the system at steady state and recording the new steady state at which the system settles after a transient. Analytical lines correspond to the numerical derivatives of the steady state levels obtained from (3). Sets (A)–(C) are for different values of the mean lifetime of miRNA-ceRNA complexes τ, and refer, respectively, to the cases in which complex processing is much faster, comparable or much slower, than the degradation dynamics of free miRNAs and ceRNAs. Also reported is the value of T that provides the best fit in each case. See Table I for parameter values. Results obtained for stronger miRNA-ceRNA couplings are shown as Supplementary Figure 1.

X_{ij} is evident. Notice that a single global parameter $T \geq 0$ has been used to fit all data in each of the conditions. This shows how accurately the assumption of a uniform effective temperature can mimic the effects of intrinsic stochasticity. On the other hand, its limits might be reflected, at least in part, in the discrepancies that occur at high transcription rates.

These results confirm that (12) and (14) are indeed good predictors of the response of a ceRNA to a perturbation af-

fecting one of its competitors within a miRNA-ceRNA network. Notably, such correlation functions are easy to estimate from transcription data sets. Our framework therefore has the potential to offer new insight into post-transcriptional regulation, its system-level organization and its impact on cellular functions.

In order to test this idea, we analyzed the ceRNA scenario emerging from 1098 breast cancer samples obtained from TCGA, focusing on the widely studied oncosuppressor PTEN and its immediate competitors (i.e. the ceRNAs sharing at least one miRNA regulator with PTEN). In particular, we computed

$$C_{\text{PTEN,ceRNA}} = \langle m_{\text{PTEN}} m_{\text{ceRNA}} \rangle_c \quad (16)$$
$$X_{\text{ceRNA,PTEN}} = \langle m_{\text{ceRNA}} \log m_{\text{PTEN}} \rangle_c \quad (17)$$
$$X_{\text{PTEN,ceRNA}} = \langle m_{\text{PTEN}} \log m_{\text{ceRNA}} \rangle_c \quad (18)$$

for a set of candidate PTEN ceRNAs found in [15] by means of Mutually Targeted miRNA-Response Element Enrichment Analysis. Notice that the average appearing in Eqs. (16–18) is over samples and not over time. We expect however

Parameters [units]	Figs 2A	Figs 2B	Figs 2C
b_2, β_1 [molec. min^{-1}]	10	10	10
k_{11}^+ [molec.$^{-1}$ min^{-1}]	e^{-5}	e^{-5}	e^{-5}
k_{21}^+ [molec.$^{-1}$ min^{-1}]	e^{-6}	e^{-6}	e^{-6}
k_{11}^-, k_{21}^- [min^{-1}]	0.001	0.001	0.001
d_1, d_2, δ_1 [min^{-1}]	0.1	0.05	0.2
σ_{11}, σ_{21} [min^{-1}]	1	0.05	0.05
κ_{11}, κ_{21} [min^{-1}]	0.001	0.001	0.001

TABLE I. Parameters used in Fig 2. Note that $i \in \{1, 2\}$ while $a = 1$.

that, if the interaction network is conserved across samples, averages over samples should reproduce statistical averages such as (13), as different samples effectively represent different snapshots of the state of the network. Fig. 3A shows that when (16) (whose value is encoded in the color of markers) is large, both (17) and (18) tend to be large. According to (15), a large C_{ij} (or ω_{ij}) signals that both PTEN and its competitor are susceptible to changes in the level of at least one of their shared regulators. For such pairs, in addition, it has previously been shown that both χ_{ij} and χ_{ji} are expected to be large [7]. This implies a fully bi-directional crosstalk, i.e. any perturbation affecting the level of one species should affect the level of the other via miRNA-mediated regulation. Remarkably, this was experimentally shown to be the case in [15] for some of the ceRNAs we tested (e.g. SERINC1, VAPA), all of which are in this regime according to our analysis. Adding to this, we are also able to point to a number of other PTEN competitors, a perturbation of which should trigger a response by PTEN.

On the other hand, smaller values of (16) (orange markers in Fig. 3A) are associated to strongly asymmetric PTEN-ceRNA pairs for which (18) is much larger than (17). This suggests that PTEN will respond to an increase of its competitor's bare transcription rate (and not vice-versa), while no response of PTEN should be expected upon perturbing the bare decay rate of the same ceRNA as C_{ij} is small. Within the steady state theory of [7], ceRNA pairs with strongly different values of χ_{ij} and χ_{ji} pertain to cases where the responding ceRNA (PTEN here) is susceptible to variations in the miRNA levels while the perturbed one (PTEN's competitor) is fully repressed. Our data analysis fully confirms both this scenario and the theory presented here in linking such cases to low values of the bare covariance (14).

Finally, note (Fig. 3B) that the above information can not be retrieved if C_{ij} is replaced by the Pearson coefficient ρ_{ij}, Eq. (1), which just amounts to normalizing C_{ij} by the product of the standard deviations σ_{m_i} and σ_{m_j} of m_i and m_j. Indeed, using the value of ρ_{ij} to color-code PTEN's ceRNAs, one sees that the Pearson coefficient can mislead into expecting (or not expecting) a response to a perturbation when the actual susceptibilities are small (resp. large).

For instance, ρ_{ij} is rather small for the pair formed by PTEN and SLC1A2, which seems to suggest absence of mutual cross-talk between these two transcripts. However, while both $C_{\text{PTEN,SLC1A2}}$ and $X_{\text{SLC1A2,PTEN}}$ are small, $X_{\text{PTEN,SLC1A2}}$ is significant. This suggests that (i) SLC1A2 will not respond to a perturbation affecting the transcription rate of PTEN, and (ii) the pair will be insensitive to changes in each other's bare decay rate; however, (iii) PTEN *will be affected* by a change in the bare transcription rate of its competitor despite the small statistical correlation that exists between their levels. Likewise, the large value of the Pearson coefficient between PTEN and DTWD2 can mislead into generically expecting a response when instead the susceptibility is strongly perturbation-dependent. In particular, the level of DTWD2 should not be significantly modified by a change in the level of PTEN (as $X_{\text{DTWD2,PTEN}}$ is rather small) in spite of the large Pearson coefficient. Notice that, remarkably, for this pair, C_{ij} and ρ_{ij} take on very different values.

DISCUSSION

To sum up, we have identified [Eq.s (11) and (12)] a set of correlation functions that can serve as proxies for ceRNA susceptibilities to perturbations. Specifically, $C_{ij} = \langle m_i m_j \rangle_c$ is related to the susceptibility ω_{ij} quantifying ceRNA i's response to a change of the bare decay rate of ceRNA j, while $X_{ij} = \langle m_i \log m_j \rangle_c$ is related to the susceptibility χ_{ij} quantifying ceRNA i's response to a change of the bare transcription rate of ceRNA j. These relations are valid at steady state and within the approximations discussed, are fully confirmed by numerical simulations.

Most importantly, quantities like C_{ij} and X_{ij} can be easily estimated from data and possibly measured in experiments. An analysis of PTEN's emergent crosstalk pattern from TCGA breast cancer dataset using these functions has indeed shown that a map of ceRNA responses to perturbations affecting competitors can be constructed by combining the information provided by each, while the Pearson coefficient ρ_{ij} can be inaccurate in this respect. This opens the way to probing the structure and function of ceRNA networks *in silico* by straightforwardly analyzing transcriptional data, and provides a key to obtain testable transcriptome-scale predictions about ceRNA crosstalk.

Notice that our results apply without any modification to ceRNA pairs that don't share miRNA regulators, i.e. it is capable of identifying long-range crosstalk (i.e. interactions between ceRNAs that are separated by multiple miRNAs along the miRNA-ceRNA network) of the kind discussed in [19]. In this sense, they can provide insight into ceRNA crosstalk both at the local scale and at an extended, network-level scale.

From the viewpoint of physics, results like (11) and (12) are akin to the "fluctuation-response relations" that constitute a cornerstone of statistical mechanics [20]. Their derivation in our context has relied on an equilibrium framework that presupposes stationarity of molecular levels. Since ceRNA crosstalk can be substantially more complex away from the steady state [21], a more refined mathematical study will be required to extend the theory developed here to off-equilibrium dynamical regimes. On the other hand, our results open the way for the application of recently developed inference techniques [22] to the estimation of miRNA levels or kinetic parameters from ceRNA levels. Overall, the approach presented here provides new means to extract information on post-transcription regulation from sequencing and/or gene expression data, thereby potentially enhancing our ability to exploit the ceRNA effect for therapeutic purposes by allowing for the identification of better (i.e. more responsive) targets for intervention.
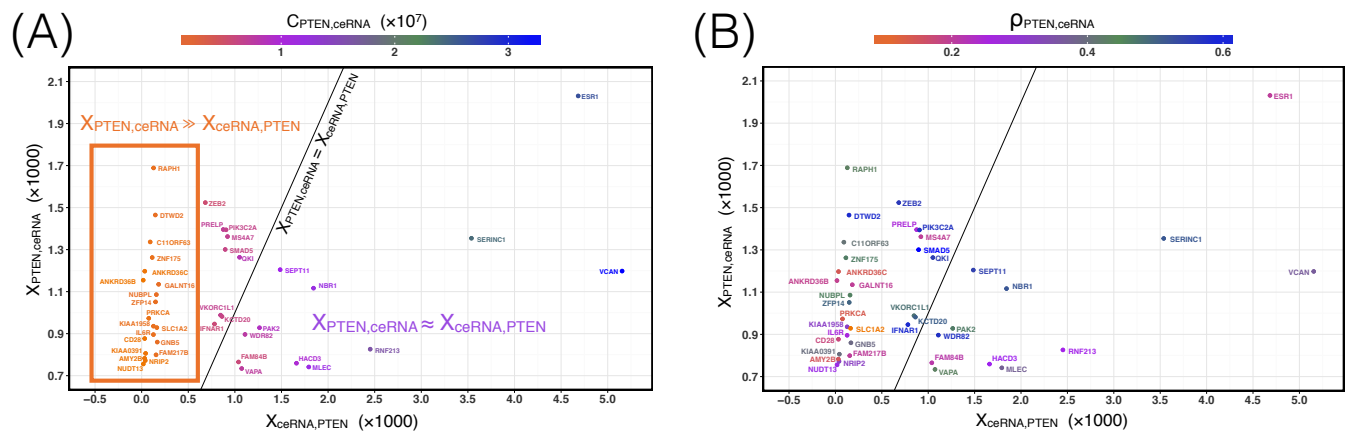
FIG. 3. Values of different correlation functions of PTEN with its ceRNAs computed from breast cancer samples from TCGA. (A) $X_{ceRNA,PTEN}$ and $X_{PTEN,ceRNA}$ are reported on the x and y axes respectively, while the color code gives the value of $C_{PTEN,ceRNA}$. (B) Same as (A), except that the color code now corresponds to the Pearson coefficient $\rho_{PTEN,ceRNA} = C_{ij}/(\sigma_{m_i}\sigma_{m_j})$. Notice that in (B) PTEN-ceRNA pairs occupy the same positions in the plane as in (A). Higher-resolution versions of both panels are provided as Supplementary Figure 2 and 3.

[1] Flynt AS, Lai EC. Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat Rev Genet (2008) 9:831-42

[2] Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet (2015) 16:421-33

[3] Guil S, Esteller M. RNA-RNA interactions in gene regulation: the coding and noncoding players. TIBS (2015) 40:248-56

[4] Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J. Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet (2007) 39:1033-1037

[5] Levine E, Zhang Z, Kuhlman T, Hwa T Quantitative Characteristics of Gene Regulation by Small RNA. PLoS Biol (2007) 5(9): e229

[6] Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell (2011) 146:353-8

[7] Figliuzzi M, Marinari E, De Martino A. MicroRNAs as a selective channel of communication between competing RNAs: a steady-state theory. Biophys J (2013) 104:1203-13

[8] Bosia C, Pagnani A, Zecchina R. Modelling competing endogenous RNA networks. PLoS One (2013) 8:e66609

[9] Noorbakhsh J, Lang AH, Mehta P. Intrinsic noise of microRNA-regulated genes and the ceRNA hypothesis. PLoS One (2013) 8:e72676

[10] Jens M, Rajewsky N. Competition between target sites of regulators shapes post-transcriptional gene regulation. Nat Rev Genet (2015) 16:113-26

[11] Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. Mol Cell (2014) 56:347-59

[12] Lai X, Wolkenhauer O, Vera J. Understanding microRNA-mediated gene regulatory networks through mathematical modelling. Nucleic Acids Res (2016) gkw550

[13] Martirosyan A, Figliuzzi M, Marinari E, De Martino A. Probing the limits to microRNA-mediated control of gene expression.

PLoS Comput Biol (2016) 12:e1004715

[14] Legnini I, Morlando M, Mangiavacchi A, Fatica A, Bozzoni I. A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis. Mol Cell (2014) 53:506-14

[15] Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, Karreth F, Poliseno L, Provero P, Di Cunto F, Lieberman J, Rigoutsos I, Pandolfi PP. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell (2011) 147:344-57

[16] Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. Mol Cell (2014) 54:766-76

[17] Denzler R, McGeary SE, Title AC, Agarwal V, Bartel DP, Stoffel M. Impact of microRNA levels, target-site complementarity, and cooperativity on competing endogenous RNA-regulated gene expression. Mol Cell (2016) 64:565-579

[18] Gillespie DT. Stochastic simulation of chemical kinetics. Annu Rev Phys Chem (2007) 58:35-55

[19] Nitzan M, Steiman-Shimony A, Altuvia Y, Biham O, Margalit H. Interactions between distant ceRNAs in regulatory networks. Biophys J (2014) 106:2254-66

[20] Marconi UM, Puglisi A, Rondoni L, Vulpiani A. Fluctuation-dissipation: response theory in statistical physics. Phys Rep (2008) 461:111-95

[21] Figliuzzi M, De Martino A, Marinari E. RNA-based regulation: dynamics and response to perturbations of competing RNAs. Biophys J (2014) 107:1011-22

[22] Roudi Y, Aurell E, Hertz J. Statistical physics of pairwise probability models. Front Comput Neurosci (2009) 3:22

Translating ceRNA susceptibilities into correlation functions

SUPPLEMENTARY MATERIAL

Araks Martirosyan,[1,2] Matteo Marsili,[3] and Andrea De Martino[4,5]

[1]*Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy*

[2]*VIB-KU Leuven Center for Brain & Disease Research, Leuven, Belgium*

[3]*The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy*

[4]*Soft and Living Matter Lab, Institute of Nanotechnology (CNR-NANOTEC),*

Consiglio Nazionale delle Ricerche, Rome, Italy

[5]*Human Genetics Foundation, Turin, Italy*

(Contains a Supplementary Text and 3 Supplementary Figures)

SUPPLEMENTARY TEXT

CONTENTS

1.   DERIVATION OF THE FUNCTION L FOR FAST COMPLEX PROCESSING

Starting from Eq. (3) of the Main Text, namely (with $\tau_{ia} = (\sigma_{ia} + \kappa_{ia} + k_{ia}^-)^{-1}$)

$$
\begin{aligned}
\dot{\mu}_a &= \beta_a - \delta_a \mu_a - \sum_i k_{ia}^+ m_i \mu_a + \sum_i (k_{ia}^- + \kappa_{ia}) c_{ia} \ , \\
\dot{m}_i &= b_i - d_i m_i - \sum_a k_{ia}^+ m_i \mu_a + \sum_a k_{ia}^- c_{ia} \ , \\
\dot{c}_{ia} &= k_{ia}^+ m_i \mu_a - \frac{c_{ia}}{\tau_{ia}} \ ,
\end{aligned}
\tag{1}
$$

we assume that complexes equilibrate much faster than miRNA and ceRNA levels and substitute c_{ia} with its steady state value $\langle c_{ia} \rangle = k_{ia}^+ \tau_{ia} m_i \mu_a$. One finds

$$
\begin{aligned}
\dot{\mu}_a &\simeq \beta_a - \delta_a \mu_a - \sum_i \frac{\sigma_{ia}}{\sigma_{ia} + \kappa_{ia} + k_{ia}^-} \, k_{ia}^+ m_i \mu_a \ , \\
\dot{m}_i &\simeq b_i - d_i m_i - \sum_a \frac{\sigma_{ia} + \kappa_{ia}}{\sigma_{ia} + \kappa_{ia} + k_{ia}^-} \, k_{ia}^+ m_i \mu_a \ .
\end{aligned}
\tag{2}
$$

For $\kappa_{ia} + k_{ia}^- \ll \sigma_{ia}$, this reduces to

$$
\begin{aligned}
\dot{\mu}_a &\simeq \beta_a - \delta_a \mu_a - \sum_i k_{ia}^+ m_i \mu_a \ , \\
\dot{m}_i &\simeq b_i - d_i m_i - \sum_a k_{ia}^+ m_i \mu_a \ ,
\end{aligned}
\tag{3}
$$

which is easily seen to be equivalent to Eq. (5) in the Main Text, with

$$L = -\sum_i (b_i \log m_i - d_i m_i) - \sum_a (\beta_a \log \mu_a - \delta_a \mu_a) + \sum_{i,a} k_{ia}^+ m_i \mu_a \ . \tag{4}$$

2. L DECREASES ALONG THE DYNAMICS

By direct differentiation and using the fact that (see Eq. (5) in the Main Text)

$$\frac{\dot{m}_i}{m_i} \simeq -\frac{\partial L}{\partial m_i} \ ,$$
$$\frac{\dot{\mu}_a}{\mu_a} \simeq -\frac{\partial L}{\partial \mu_a} \ , \tag{5}$$

one finds

$$\dot{L} = \sum_i \frac{\partial L}{\partial m_i} \dot{m}_i + \sum_a \frac{\partial L}{\partial \mu_a} \dot{\mu}_a = -\sum_i \frac{\dot{m}_i^2}{m_i} - \sum_a \frac{\dot{\mu}_a^2}{\mu_a} \leq 0 \ . \tag{6}$$

In other words, under the approximations discussed above, L decreases along the dynamics of the miRNA-ceRNA network. Therefore the minimum of L (which is unique by virtue of the concavity of L) describes a steady state of the dynamics (1).

3.   APPROXIMATE CALCULATION OF ω_{ij} FOR A SYSTEM WITH ONE MIRNA AND N CERNA SPECIES

Starting from Eq. (3) taken for N ceRNA species and a single miRNA species (we suppress its index for simplicity) in the limit $\sigma_i \gg k_i^- + \kappa_i$, the steady-state level of m_i reads

$$m_i = \frac{b_i}{d_i} F_i(\mu) \quad , \quad F_i(\mu) = \frac{\mu_{0,i}}{\mu_{0,i} + \mu} \quad , \quad \mu_{0,i} = \frac{d_i}{k_i^+} \ . \tag{7}$$

Now noting that

$$F_i(\mu) \simeq \begin{cases} 1 - \frac{\mu}{\mu_{0,i}} & \text{for } \mu \ll \mu_{0,i} \ \text{(ceRNA } i \text{ expressed)} \\ \frac{1}{2} - \frac{\mu - \mu_{0,i}}{4\mu_{0,i}} & \text{for } \mu \simeq \mu_{0,i} \ \text{(ceRNA } i \text{ susceptible)} \\ \frac{\mu_{0,i}}{\mu} & \text{for } \mu \gg \mu_{0,i} \ \text{(ceRNA } i \text{ repressed)} \end{cases} \tag{8}$$

$$\frac{dF_i}{d\mu} \equiv F_i' \simeq \begin{cases} -\frac{1}{\mu_{0,i}} & \text{for } \mu \ll \mu_{0,i} \ \text{(ceRNA } i \text{ expressed)} \\ -\frac{1}{4\mu_{0,i}} & \text{for } \mu \simeq \mu_{0,i} \ \text{(ceRNA } i \text{ susceptible)} \\ -\frac{\mu_{0,i}}{\mu^2} & \text{for } \mu \gg \mu_{0,i} \ \text{(ceRNA } i \text{ repressed)} \end{cases} \tag{9}$$

and that the steady state miRNA level can be approximated by [1]

$$\mu \simeq \frac{\beta - \sum_{i \in \text{Repr}} b_i - \frac{1}{4} \sum_{i \in \text{Susc}} b_i}{\delta + \sum_{i \in \text{Expr}} \frac{b_i k_i^+}{d_i} + \frac{1}{4} \sum_{i \in \text{Susc}} \frac{b_i k_i^+}{d_i}} \quad , \tag{10}$$

so that

$$\frac{\partial \mu}{\partial d_j} \simeq \chi_{\mu\mu} \frac{b_j}{d_j} \frac{\mu}{\mu_{0,j}} \times \begin{cases} 1 & \text{for } \mu \gg \mu_{0,i} \ (\text{ceRNA } j \text{ expressed}) \\ \frac{1}{4} & \text{for } \mu \simeq \mu_{0,i} \ (\text{ceRNA } j \text{ susceptible}) \\ 0 & \text{for } \mu \gg \mu_{0,i} \ (\text{ceRNA } j \text{ repressed}) \end{cases} \tag{11}$$

$$\chi_{\mu\mu} = \left(\delta + \sum_{i \in \text{Expr}} \frac{b_i k_i^+}{d_i} + \frac{1}{4} \sum_{i \in \text{Susc}} \frac{b_i k_i^+}{d_i} \right)^{-1} \quad , \tag{12}$$

we can compute the susceptibility ω_{ij} as

$$\omega_{ij} \equiv \frac{\partial m_i}{\partial d_j} = \frac{b_i}{d_i} F_i' \frac{\partial \mu}{\partial d_j} \qquad (i \neq j) \ . \tag{13}$$

One finds

$$\omega_{ij} \simeq -\chi_{\mu\mu} \frac{b_i b_j}{d_i d_j} W_{R(i),R(j)} \qquad (i \neq j) \ , \tag{14}$$

where $\widehat{W}$ is a 3×3 matrix that only depends on the regime $R(i)$ (repressed, susceptible or expressed) to which ceRNA i belongs. By considering the definitions of the different regimes in terms of the value of μ, all elements of $\widehat{W}$ are found to be $\ll 1$ (for instance, $W_{\text{Expr},\text{Expr}} = \mu/(\mu_{0,i}\mu_{0,j}) \ll 1$ as $\mu \ll \mu_{0,i}$ and $\mu \ll \mu_{0,j}$ if ceRNAs i and j are both expressed) except for $W_{\text{Susc},\text{Susc}}$, which is given by

$$W_{\text{Susc},\text{Susc}} = \frac{1}{16} \frac{\mu}{\mu_{0,i}\mu_{0,j}} \quad , \tag{15}$$

leading immediately to Eq. (15) of the Main Text.

4. CASE OF SLOW COMPLEX PROCESSING

Assuming complex levels c_{ia} are roughly stationary over time scales for which m_i and μ_a evolve (i.e. $\tau_{ia} \gg 1/d_i$ and $\tau_{ia} \gg 1/\delta_a$ for each i and a), then all terms in (1) that involve the variables c_{ia} can be taken to be roughly constant for short enough characteristic times. In such conditions, miRNAs are effectively transcribed at rates

$$\beta_a^{\text{eff}} \simeq \beta_a + \sum_i (k_{ia}^- + \kappa_{ia}) c_{ia} \ , \tag{16}$$

while ceRNAs are effectively transcribed at rates

$$b_i^{\text{eff}} \simeq b_i + \sum_a k_{ia}^- c_{ia} \ . \tag{17}$$

In this limit, (1) can again be cast as

$$\dot{m}_i \simeq -m_i \frac{\partial L}{\partial m_i} \ , $$
$$\dot{\mu}_a \simeq -\mu_a \frac{\partial L}{\partial \mu_a} \ , \tag{18}$$

with

$$L = -\sum_i (b_i^{\text{eff}} \log m_i - d_i m_i) - \sum_a (\beta_a^{\text{eff}} \log \mu_a - \delta_a \mu_a) + \sum_{i,a} k_{ia}^+ m_i \mu_a \ . \tag{19}$$

The main difference from the previous case lies in the fact that the minimum of L should now be computed self-consistently from the asymptotic value of c_{ia}: after the (fast) equilibration of m_i's and μ_a's following (18), a new steady state value for complexes is computed as $c_{ia} = k_{ia}^+ \tau_{ia} m_i \mu_a$, leading in turn to new values for the effective transcription rates β_a^{eff} and b_i^{eff} and hence to new values for m_i's and μ_a's from (18), and so on until convergence.

5. STOCHASTIC DYNAMICS OF A MIRNA-CERNA NETWORK

The time evolution of our miRNA-ceRNA network with N ceRNA species (labeled i), M miRNA species (labeled a) and intrinsic (molecular) noise is described by the system

$$\dot{\mu}_a = \beta_a - \delta_a \mu_a - \sum_i k_{ia}^+ m_i \mu_a + \sum_i (k_{ia}^- + \kappa_{ia}) c_{ia} + \eta_a \ , $$
$$\dot{m}_i = b_i - d_i m_i - \sum_a k_{ia}^+ m_i \mu_a + \sum_a k_{ia}^- c_{ia} + \xi_i \ , \tag{20}$$
$$\dot{c}_{ia} = k_{ia}^+ m_i \mu_a - \frac{c_{ia}}{\tau_{ia}} + \zeta_{ia} \ , $$

where $\tau_{ia} = (\sigma_{ia} + \kappa_{ia} + k_{ia}^-)^{-1}$ while η_a, ξ_a and ζ_{ia} represent stochastic variables. As each noise source contributes independently to the overall noise level, one has

$$\eta_a = \eta_{\mu_a} - \sum_i \zeta_{ia}^+ + \sum_i \zeta_{ia}^- + \sum_i \zeta_{ia}^\kappa \ , \tag{21}$$

$$\xi_i = \xi_{m_i} - \sum_a \zeta_{ia}^+ + \sum_a \zeta_{ia}^- \ , \tag{22}$$

$$\zeta_{ia} = \zeta_{ia}^\sigma + \zeta_{ia}^+ - \zeta_{ia}^- - \zeta_{ia}^\kappa \tag{23}$$

where ξ_{m_i}, η_{μ_a}, $\zeta_{ia}^{\pm}$, ζ_{ia}^{σ}, and ζ_{ia}^{κ} and are mutually independent zero-average random variables representing, respectively, the intrinsic noise in ceRNA levels, in miRNA levels, in the binding/unbinding dynamics of complexes, in the stoichiometric complex degradation channel and in the catalytic complex degradation channel. Correlations are, for each component, described by

$$
\begin{aligned}
\langle \xi_{m_i}(t)\xi_{m_i}(t')\rangle &= (d_i\overline{m}_i + b_i)\,\delta(t-t')\ , \\
\langle \xi_{\mu_a}(t)\xi_{\mu_a}(t')\rangle &= (\delta_a\overline{\mu}_a + \beta_a)\,\delta(t-t')\ , \\
\langle \zeta_{ia}^{+}(t)\zeta_{ia}^{+}(t')\rangle &= k_{ia}^{+}\overline{m}_i\overline{\mu}_a\,\delta(t-t')\ , \\
\langle \zeta_{ia}^{-}(t)\zeta_{ia}^{-}(t')\rangle &= k_{ia}^{-}\overline{c}_{ia}\,\delta(t-t')\ , \\
\langle \zeta_{ia}^{\sigma}(t)\zeta_{ia}^{\sigma}(t')\rangle &= \sigma_{ia}\overline{c}_{ia}\,\delta(t-t')\ , \\
\langle \zeta_{ia}^{\kappa}(t)\zeta_{ia}^{\kappa}(t')\rangle &= \kappa_{ia}\overline{c}_{ia}\,\delta(t-t')\ ,
\end{aligned}
\tag{24}
$$

where

$$
\overline{m}_i = \frac{b_i + \sum_a k_{ia}^{-}\overline{c}_{ia}}{d_i + \sum_a k_{ia}^{+}\overline{\mu}_a}\ ,
\tag{25}
$$

$$
\overline{\mu}_a = \frac{\beta_a + \sum_i (k_{ia}^{-} + \kappa_{ia})\overline{c}_{ia}}{\delta_a + \sum_i k_{ia}^{+}\overline{m}_i}\ ,
\tag{26}
$$

$$
\overline{c}_{ia} = \frac{k_{ia}^{+}\overline{\mu}_a\,\overline{m}_i}{\sigma_{ia} + k_{ia}^{-} + \kappa_{ia}}
\tag{27}
$$

denote the mean steady-state molecular levels. To obtain Fig. 2 of the Main Text, we have simulated the above system with $M = 1$, $N = 2$ using the Gillespie algorithm [2].

---

[1] Figliuzzi M, Marinari E, De Martino A. MicroRNAs as a selective channel of communication between competing RNAs: a steady-state theory. Biophys J (2013) 104:1203-13

[2] Gillespie DT. Stochastic simulation of chemical kinetics. Annu Rev Phys Chem (2007) 58:35-55
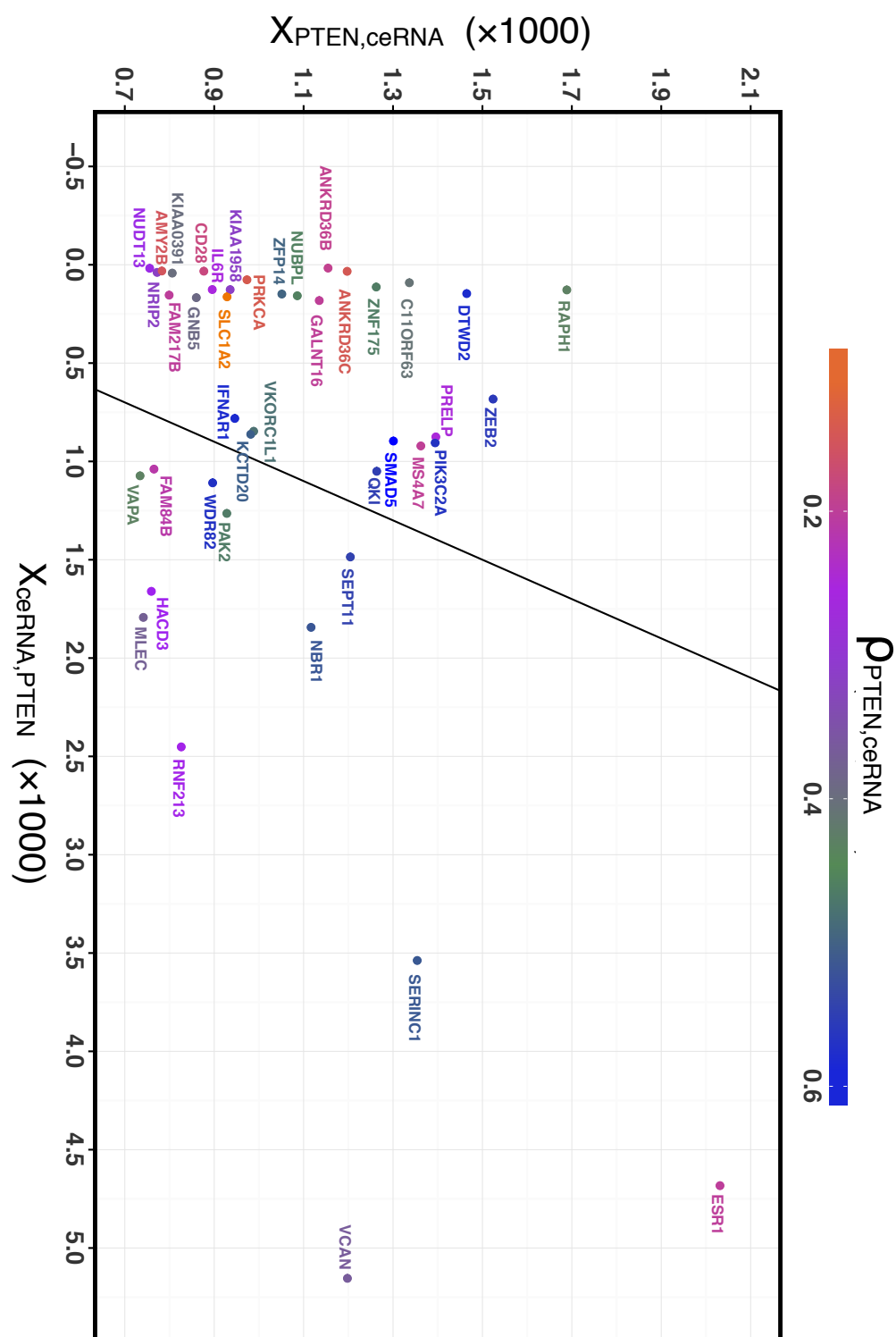
SUPPLEMENTARY FIGURES
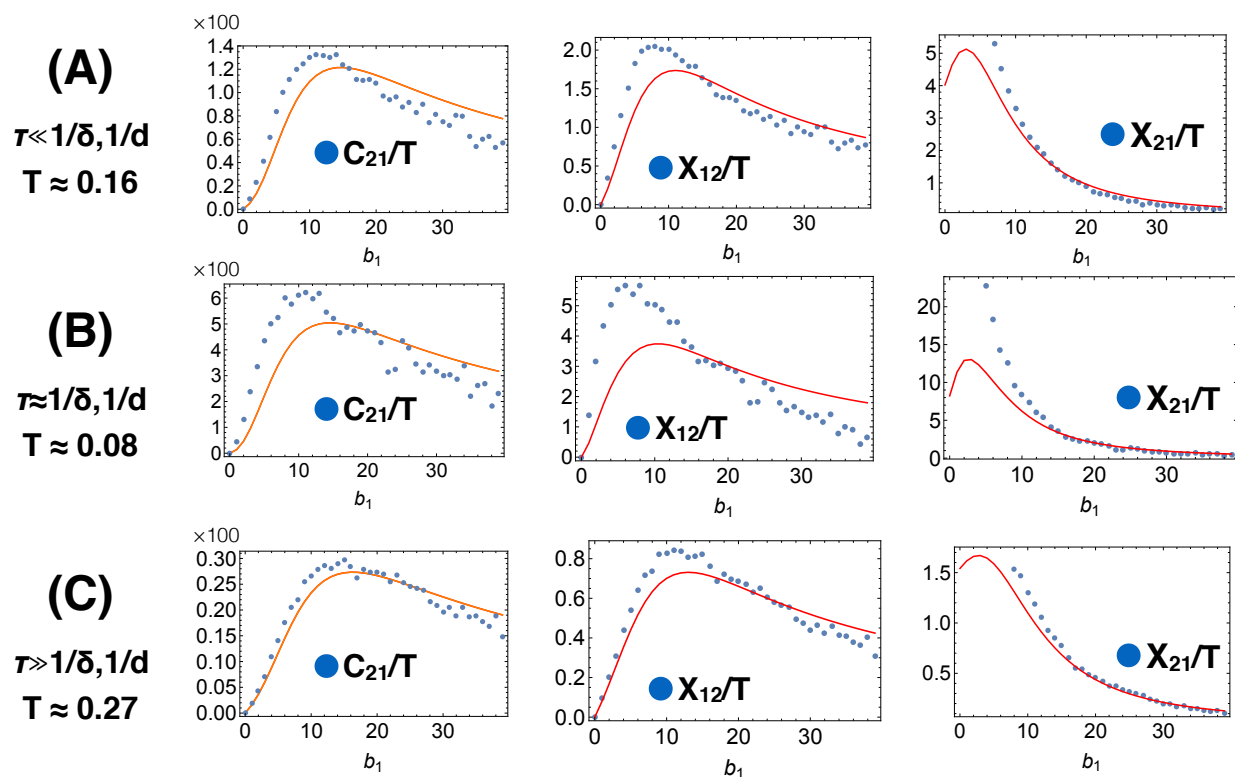
Supplementary Fig. 1. Same as Figure 2 in the Main Text but for stronger miRNA repression. Parameter values are as in Table I of the Main Text except for $k_{11}^+ = e^{-3}$ and $k_{21}^+ = e^{-4}$. Note that, for small values of b_1, m_1 gets too small to accurately estimate X_{ij}s.
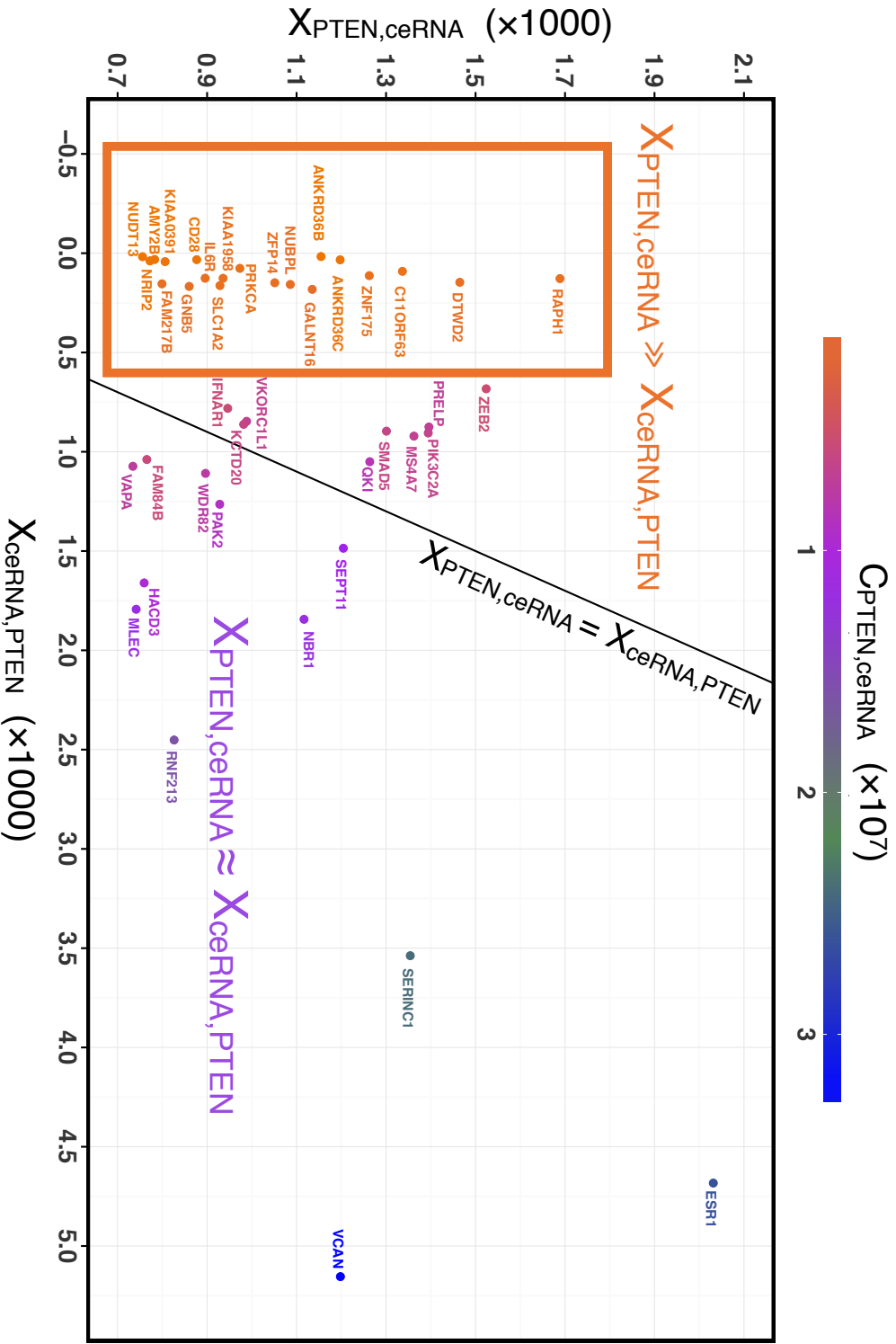
Supplementary Fig. 2. Large-scale version of Fig. 3A in the Main Text.

Supplementary Fig. 3. Large-scale version of Fig. 3B in the Main Text.