

Protocol

In silico and in vitro protocols for quantifying gene expression noise modulated by microRNAs



Characterizing the noise modulation pattern of microRNA is valuable for both microRNA function analysis and synthetic biology applications. Here we propose a coarse-grained model to simulate how the properties of microRNAs, competing RNAs, and microRNA response elements affect gene expression noise. We also detail an experimental protocol based on synthetic gene circuits and flow cytometry to quantify the noise. This framework is easy-to-use for the study and application of both microRNA and gene expression noise.

weilei92@tsinghua.edu.cn xwwang@tsinghua.edu.cn

microRNA modulates

and microRNA target composition to noise

Wei et al., STAR Protocols 3, 101205 March 18, 2022 © 2022 The Author(s). https://doi.org/10.1016/ j.xpro.2022.101205



Protocol

In silico and *in vitro* protocols for quantifying gene expression noise modulated by microRNAs

Lei Wei,^{1,3,4,*} Shuailin Li,^{2,3} and Xiaowo Wang^{1,5,*}

¹Ministry of Education Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, Bioinformatics Division, Beijing National Research Center for Information Science and Technology, Department of Automation, Tsinghua University, Beijing 100084, China

²School of Life Sciences, Tsinghua University, Beijing 100084, China

³These authors contributed equally

⁴Technical contact

⁵Lead contact

*Correspondence: weilei92@tsinghua.edu.cn (L.W.), xwwang@tsinghua.edu.cn (X.W.) https://doi.org/10.1016/j.xpro.2022.101205

SUMMARY

Characterizing the noise modulation pattern of microRNA is valuable for both microRNA function analysis and synthetic biology applications. Here we propose a coarse-grained model to simulate how the properties of microRNAs, competing RNAs, and microRNA response elements affect gene expression noise. We also detail an experimental protocol based on synthetic gene circuits and flow cytometry to quantify the noise. This framework is easy-to-use for the study and application of both microRNA and gene expression noise.

For complete details on the use and execution of this protocol, please refer to Wei et al. (2021).

BEFORE YOU BEGIN

Here we proposed a framework to simulate and observe gene expression noise modulated by micro-RNA (miRNA). The simulation approach includes a coarse-grained mathematical model and codes to solve the model. Here we adopted ordinary differential equations (ODEs) to describe and solve the whole system as ODEs can elaborately simulate the steady states as well as the dynamic behaviors of biochemical reactions in a numerical way when the molecule number in the reaction is not too rare to be regarded as a continuous value. The ODE model was based on previous studies that guantitatively analyzed the properties of endogenous competing RNAs (Ala et al., 2013; Miotto et al., 2019; Wei et al., 2019; Yuan et al., 2015, 2016) and was extended here to describe the composition of miRNA response elements (MREs). Biochemical reactions are stochastic even at steady states due to the widespread fluctuation, which leads to noise of gene expression. Here, we used the fluctuation-dissipation theorem to calculate the steady-state expression noise inspired by a previous study (Schmiedel et al., 2015). The fluctuation-dissipation theorem describes the correlations of observables' fluctuation under stochastic perturbations at thermodynamic equilibrium, which can be used to analyze gene expression noise with linear noise approximation at the equilibrium state (Elf and Ehrenberg, 2003). As the fluctuation-dissipation theorem is given by a Lyapunov equation (Van Kampen, 1992), here we solved the theorem by the continuous Lyapunov equation solution function lyap in MATLAB. We also proposed a method to evaluate the contribution of a certain biochemical reaction to noise. The protocol below describes a system with two species of competing RNAs and one species of miRNA, or one species of RNA with an MRE composed of two repetitive or composite miRNA targets and one species of miRNA. However, the number of species or miRNA targets can easily be extended according to the protocol.







Figure 1. The bidirectional promoter system

The experiment approach includes a bidirectional promoter system driving the expression of two fluorescent proteins, mKate2 and EYFP (Figure 1), which represent the transcription and translation level of the system, respectively. The system is widely used in miRNA and gene expression noise studies (Bosia et al., 2017; Mukherji et al., 2011; Schmiedel et al., 2015). MREs can be inserted to 3' untranslated region (3' UTR) of EYFP to modulate the expression (Figure 1). The intensity of mKate2 and EYFP can be quantified by flow cytometry. We also proposed a method to analyze the flow cytometry data to estimate the mean value and noise (coefficient of variation, CV) of EYFP expression. The protocol below was developed on HeLa cells. Key parameters and agents that can be modified when applying the protocol to other cell types are shown in the problem 1 of the troubleshooting part.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Doxycycline (Dox)	Clontech	#631311
Lipofectamine™ LTX Reagent with PLUS™ Reagent	Thermo Fisher Scientific	#A12621
Dulbecco's Modified Eagle Medium (DMEM), high glucose	Thermo Fisher Scientific	#11965092
Fetal Bovine Serum (FBS)	Thermo Fisher Scientific	#10437028
MEM Non-Essential Amino Acids Solution (NEAA) (100 \times)	Thermo Fisher Scientific	#11140050
Penicillin-Streptomycin	Thermo Fisher Scientific	#15140122
Trypsin-EDTA (0.05%), phenol red	Thermo Fisher Scientific	#25300120
Phosphate Buffered Saline (PBS)	Thermo Fisher Scientific	#10010023
Experimental models: Cell lines		
HeLa	ATCC	CCL-2
Oligonucleotides		
microRNA response elements	Wei et al. (2021)	N/A
Recombinant DNA		
Dual-fluorescent reporter systems	Wei et al. (2021)	GenBank: MZ542768
pDT7004	Li et al. (2015)	N/A

(Continued on next page)

Protocol



Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Executable model code and scripts	This paper	https://github.com/XWangLabTHU/ microRNA_noise
MATLAB (R2020b)	MathWorks	RRID: SCR_001622 https://www.mathworks.com/ products/matlab.html
R (v 3.6.1)	R Core Team	RRID: SCR_001905 https://www.r-project.org
Floreada.io	Floreada Devs	https://floreada.io
Other		
BD LSRFortessa™ cell analyzer	BD Biosciences	https://www.bdbiosciences.com/ en-ca/products/instruments/ flow-cytometers/research-cell- analyzers/bd-lsrfortessa

STEP-BY-STEP METHOD DETAILS

Download codes

© Timing: 10 min

1. The MATLAB codes for simulating gene expression noise and the R codes for analyzing flow cytometry data are available at https://github.com/XWangLabTHU/microRNA_noise. Please notice that all MATLAB codes should be downloaded into one same folder as the execution of *solve_noise.m* relies on the function *steady_state.m*. See troubleshooting 1 for details.

Simulating gene expression noise modulated by miRNA

© Timing: 20 min

In this section, we will describe how to set up a mathematical model to simulate the influence of miRNA properties on gene expression noise in MATLAB, using the code we provided.

The parameters of the mathematical model can be modified to fit the purpose of simulation. There are four main types of parameters that should be set before simulation: the kind of model, the fixed parameters, the altered parameters, and the simulation step size and range. After, the model could be simulated and visualized.

2. Set parameters for simulation.

- a. Open MATLAB. Open the folder where the MATLAB codes are deposited.
- b. Open the code file *solve_noise.m.* Please notice that in MATLAB, lines beginning with "%" denote comment lines in the code that will not be read as code in the software.
- c. Set the type of model on Line 7. There are three types of preset miRNA regulation models: i. for competing RNAs:

> type = 1;

ii. for repetitive targets of same miRNAs:

> type = 2;





iii. for multiple targets of different miRNAs:

> type = 3;

- d. Set the fixed and altered parameters. These parameters describe the production and degradation rates of all components in the biochemical reactions, which are shown in Figure 2. All parameters involved in the model should be set before simulation in the "set fixed parameters" section. All parameters that should be altered for simulating different conditions should be set in the "set altered parameters" section. We provide samples of altered parameters for all three model types, of which parameters for type 2 and 3 are commented on by default. If users are interested in type 2 or 3, they should uncomment (remove the "%" from the beginning of the line) the corresponding codes of the type they are interested and comment (add a "%" to the beginning of the line) the unrelated codes of other types. We provided default parameter settings in the code according to previous studies that describes perfect complementary miRNA target sites (Wei et al., 2019, 2021; Yuan et al., 2015). Users can modify these parameters by comparing their interested miRNA targets and the perfect complementary miRNA targets in this study.
- e. The simulation step size and range should be assigned in the "set simulation parameters" section. The simulation range determines the range that the gene expression is simulated in by setting the production rate of the target gene's mRNA (*kT*). We recommend users to set a wide range with a large step size to quickly determine the lower and upper limit of the range, and then narrow the step size to gain a refined simulation result.
- 3. Run solve_noise.m to simulate the model with the set parameters by clicking Run in the Contextual Tab Editor. The results will be written into the file exp_mean.mat and exp_CV.mat. exp_mean.mat is a matrix that stored the expression level of the target gene with different kT under different parameter settings that are set in step 2d. exp_CV.mat is a three-dimensional array that stored the expression noise (CV) of the target gene with different kT under different parameter settings contributed by different reactions. The first component of the third dimension of the array represents the total noise contributed by all reactions. All the other components of the third dimension of the order shown in Figure 2.
- 4. Open and run *plot_noise.m* to visualize the results stored in the file *exp_mean.mat* and *exp_CV.mat*.

Transient transfection of HeLa cells

© Timing: 2 days

Here we will introduce how to transfect Hela cells with three plasmids. The first plasmid can express two fluorescent proteins, mKate2 and EYFP, driven by a bidirectional promoter, which can be activated by binding of reverse tetracycline-controlled transactivator (rtTA) in the existence of doxycycline (Figure 1). The second plasmid can express rtTA constantly. The third plasmid does not encode any proteins and is used as a blank plasmid to increase transfection efficiency and reduce variation of plasmid transfection. miRNAs that can bind to 3'UTR of EYFP are endogenous.

Plasmids should be transfected into cells for further noise observation and quantification. Here we transfect HeLa cells via Lipofectamine LTX with PLUS Reagent according to the manufacturer's manual protocol with the cell density, amount of DNA and the reagents specified specifically for Hela cells transfected in 12-well plates compared to the manual (see next steps for details).

STAR Protocols Protocol



miRNA regulation model



Figure 2. The definition of parameters for simulation

The number represent the order of reactions for calculating their contributions in step 3. Reactions employed in each miRNA regulation model are shown. Types of all components and reactions are shown with different colors, shapes and line types.

5. Culture the cell lines into 12-well plates that you are interested in on the day before transfection with a proper density. The density of HeLa cells should be around 1.6 × 10⁵ cells/mL. The culture medium of HeLa cells is Dulbecco's Modified Eagle Medium (DMEM) with 10% Fetal Bovine





Serum (FBS), 1% MEM Non-Essential Amino Acids Solution (NEAA) and 1% Penicillin-Streptomycin. The total volume of the medium per well is 1 mL.

- 6. About 24 h later, make sure that the density of cells is around 70%.
- 7. For each well, Dilute 2 μ L Lipofectamine® LTX Reagent in 50 μ L serum-free DMEM.
- For each well, add 40 ng of the plasmids carrying the dual-fluorescent reporter, 40 ng of the plasmids carrying the rtTA gene and 420 ng of blank plasmids with no protein-coding sequences (for example, pDT7004 from (Li et al., 2015)) into 50 μL serum-free DMEM, and then add 1 μL PLUSTM Reagent.
- 9. Mixed the reagents prepared in step 7 and 8.
- 10. Incubate for 5 min at 20°C–25°C.
- 11. Changed the medium of cells in the 12-well plates with 900 μ L media, and then add 100 μ L mixture prepared in step 9 to cells.
- 12. Add doxycycline to the medium with a final concentration of 1 μ g/mL after adding the mixed reagents immediately.
- 13. Change the medium one day after transfection. Do not forget to add doxycycline into the medium again with the same concentration described in step 12.

Note: Make sure that the order of adding the reagents is strictly the same as the manufacturer's manual, which means that do not add the PLUSTM Reagent first to the medium and then add plasmids sequentially, because adding the PLUSTM Reagent first will decrease the co-transfection efficiency of the plasmids.

Note: It is important to add the blank plasmid with no protein-coding sequence. This procedure could supplement the total amount of the plasmid to enhance the transfection efficiency and reduce the variation of transfection. The amount of the plasmids could be adjusted to ensure a high transfection efficiency. For example, when the transfection efficiency is low, try to improve the relatively amount of the plasmids that express the reporter and rtTA in the total 500 ng plasmids (e.g., 80 ng of the plasmid that express the reporter + 80 ng of the plasmid that express rtTA + 340 ng of the blank plasmid per well).

Flow cytometry

© Timing: 4 h

The fluorescent intensity of cells can be observed by flow cytometry two days after transfection.

- 14. Digest cells from the 12-well plates using trypsin-EDTA (200 μ L for each well, 3 min), and then stop the reaction using 300 μ L culture medium described in step 5.
- 15. Centrifuge cells at 300 g for 5 min, wash cells using 500 μL phosphate buffered saline (PBS).
- 16. Repeat step 15.
- 17. Resuspend cells with PBS and transfer the cells into separate FACS tubes.
- 18. Detect fluorescent intensities of cells using flow cytometry machines and record no fewer than 10^5 singlets.
- 19. Export all records with the fcs format.

Note: The type of flow cytometer is flexible and the settings may need to be altered according to the equipment. We do not recommend using the high-throughput screening (HTS) way of flow cytometry to perform these procedures. When using HTS, the cells can hardly be mixed well and may cause poor data quality or damage the flow cytometry machine.

Note: Make sure cells have been washed twice before running in the flow cytometry machine. If the medium is not washed thoroughly, cells may clump together such that the machine may

STAR Protocols Protocol



be blocked by the clumped cells and bubbles can be generated, making the measurement of fluorescent intensity unreliable.

Note: Vortex cells to avoid clumped cells right before running samples in the flow cytometry machine.

Note: If the recording speed showing in the flow cytometry machine is too high (> 3,000 total events per second), dilute cells properly to make sure around 2,000 total events are recorded per second at the high-speed mode of the machine. Otherwise, too many cells might block the machine and generate bubbles in the machine.

Data analysis of flow cytometry

© Timing: 1 h

- 20. Gate living cells according to the forward scatter (FSC) and side scatter (SSC) value from the flow cytometry data via data analysis software. Save the fluorescent intensity values of these cells in a csv file. For instance, when analyzing the data by Floreada.io, press File/Open File(s) to import fcs files, gate living cells with FSC-A/SSC-A, FSC-H/FSC-W, SSC-H/SSC-W sequentially, and then store the data by pressing File/Save CSV Event File. Arrange all csv files into a folder named data in the same folder with analysis.r.
- 21. Run *analysis.r* to analyze all files in the folder *data*. For instance, when using Terminal, first direct the working directory to the working folder where the file *analysis.r* and the folder *data* are, and then execute

> Rscript analysis.r

When using RStudio, set the working folder as the working path and then execute *analysis.r*. The results will be stored in a folder named *result*.

EXPECTED OUTCOMES

The expected outcomes of the simulation with the default parameter settings in the code are shown in Figure 3. The line named condition 1, 2, 3 and 4 represent the simulation results under different altered parameters set in step 2d. *Expression* represents the expression level of the target gene. *Fold change* represents the expression level between condition 1 which is set as gene expression without the regulation of miRNA under the default parameter settings. CV represents the noise level of the target gene, which is calculated by dividing the standard deviation by mean.

The expected outcomes of flow cytometry experiments are shown in Figure 4. Cells recorded by flow cytometer are binned with the intensity of mKate2 in a logarithmic scale. EYFP_mean and mKate2_mean represent the mean value of EYFP and mKate2 of cells in a certain bin. EYFP_CV represents the noise level of EYFP in a certain bin.

LIMITATIONS

We only tested this protocol on HeLa cells. To make the protocol adaptable for other cell lines, the transfection condition may need to be modified to obtain high transfection efficiency. Besides, this experimental protocol is only suitable for studying regulatory elements in the post-transcriptional process. To study the noise modulation patterns in other gene regulatory stages, people may refer to Quarton et al. to develop proper experimental systems (Quarton et al., 2020).







Figure 3. An example of the visualized simulation results

Expression and CV represent the average gene expression level and noise level (CV) of the gene regulated by miRNA. Fold change represents the difference of gene expression level between the condition with and without the regulation of miRNA. Expression and fold change are shown in a logarithmic scale. This figure shows the simulation result of miRNA regulation model 1 (with competing RNAs). Condition 1 represents the result without miRNAs (kR = 0); condition 2 represents the result with miRNAs but without competing RNAs (kR > 0, kT2 = 0); condition 3 represents the result with miRNAs and strong competing RNAs (kR > 0, kT2 > 0, small koff2); condition 4 represents the results with miRNAs and weak competing RNAs (kR > 0, kT2 > 0, large koff2).

TROUBLESHOOTING

Problem 1

When running *solve_noise.m* in step 3, MATLAB reports the error: "Undefined function or variable 'steady_state'."

Potential solution

The code *solve_noise.m* relies on a function that calculate the steady state of ODEs which is provided in the file *steady_state.m*. Thus, it is necessary to store *steady_state.m* and *solve_noise.m* in the same folder. We recommend to download all the codes on GitHub together and perform simulation with the document structure the same as that on Github.

Problem 2

In step 4, the simulation results under different altered parameters don't share the same range of gene expression (such as Figure 3, the x-axis range of condition 1 is much shorter than other conditions).

Potential solution

Actually it is a right outcome of the simulation. As miRNA can repress the expression of the target gene, conditions with the same setting of kT (the production rate of the target gene's mRNA) may lead to different expression levels of the target genes' protein with or without miRNA. Thus, the range of the expression level of the target gene may be different under different altered parameter settings. If users want to observe the results of different conditions within a same expression level range, we recommend to set a large extent of kT to let the results of all conditions cover the whole range users are interested in, and then choose a proper x-axis limit to plot the results.

Problem 3

After performing step 5–13, The transfection efficiency of the reporter system into HeLa cells is low (e.g., lower than 10%).

Potential solution

Make sure that the blank plasmid which does not express any proteins is added in the transfection. Besides, make sure that plasmids are mixed first and then add the PLUS reagent to increase co-transfection

Protocol





Figure 4. An example of the visualized flow cytometry experiment results

EYFP_mean and mKate2_mean represent the average gene expression level of EYFP and mKate2 in each bin. All the mean values are shown in a logarithmic scale. EYFP_CV represents the noise level (CV) of EYFP in each bin.

efficiency. In addition, cell density, the amount of plasmids, the Lipofectamine LTX and PLUS reagent can be optimized according to the instructions of the manufacturer to maximize transfection efficiency.

Problem 4

The recording speed of the flow cytometer is too high (> 3,000 total events per second) or too low (< 500 total events per second).

Potential solution

If the recording speed is too high, please dilute cells with PBS properly to make sure around 2,000 total events are recorded per second. If the recording speed is too low, please check whether the machine has been set at the high-speed mode. Besides, clumped cells may exist at the bottom of tubes if cells are not mixed before recording. Low densities of cells at the seeding procedure (step 5) and loss of cells during washing (step 15–16) may also cause the low speed. If so, please notice that there may not be enough cells for downstream analysis.

Problem 5

Cells shown in the FSC-SSC plot in flow cytometry are not in the pre-specified region and have lower SSC than that in the region (step 18).

Potential solution

Make sure that the medium is washed away when the cells are to be measured by flow cytometry. Mix cells thoroughly to avoid clumped cells before running samples in the flow cytometry machine. Prime the machine multiple times or run water for around 5 min at high speed to get rid of the bubbles. Avoid using the HTS mode of the flow cytometry machine. Dilute cells properly to make sure around 2,000 cells are recorded each second at the high-speed mode of the machine.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xiaowo Wang (xwwang@tsinghua.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The code generated during this study is available at Github [https://github.com/XWangLabTHU/microRNA_noise, https://doi.org/10.5281/zenodo.5953915].

ACKNOWLEDGMENTS

This work has been supported by the State Key Research Development Program of China (2020YFA0906900) and the National Natural Science Foundation of China (62050152, 61773230, 61721003, and 62103227).

AUTHOR CONTRIBUTIONS

X.W., L.W., and S.L. designed and tested the protocol; L.W. and S.L. wrote the manuscript; X.W. supervised this study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Ala, U., Karreth, F.A., Bosia, C., Pagnani, A., Taulli, R., Leopold, V., Tay, Y., Provero, P., Zecchina, R., and Pandolfi, P.P. (2013). Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. Proc. Natl. Acad. Sci. U S A *110*, 7154–7159.

Bosia, C., Sgrò, F., Conti, L., Baldassi, C., Brusa, D., Cavallo, F., Cunto, F.D., Turco, E., Pagnani, A., and Zecchina, R. (2017). RNAs competing for microRNAs mutually influence their fluctuations in a highly non-linear microRNA-dependent manner in single cells. Genome Biol. *18*, 37.

Elf, J., and Ehrenberg, M. (2003). Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Res. 13, 2475–2484.

Li, Y., Jiang, Y., Chen, H., Liao, W., Li, Z., Weiss, R., and Xie, Z. (2015). Modular construction of mammalian gene circuits using TALE transcriptional repressors. Nat. Chem. Biol. 11, 207–213. Miotto, M., Marinari, E., and De Martino, A. (2019). Competing endogenous RNA crosstalk at system level. PLoS Comput. Biol. 15, e1007474.

Mukherji, S., Ebert, M.S., Zheng, G.X.Y., Tsang, J.S., Sharp, P.A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. Nat. Genet. 43, 854–859.

Quarton, T., Kang, T., Papakis, V., Nguyen, K., Nowak, C., Li, Y., and Bleris, L. (2020). Uncoupling gene expression noise along the central dogma using genome engineered human cell lines. Nucleic Acids Res. *48*, 9406–9413.

Schmiedel, J.M., Klemm, S.L., Zheng, Y., Sahay, A., Bluthgen, N., Marks, D.S., and van Oudenaarden, A. (2015). MicroRNA control of protein expression noise. Science 348, 128–132.

Van Kampen, N.G. (1992). Stochastic Processes in Physics and Chemistry (Elsevier).

Wei, L., Yuan, Y., Hu, T., Li, S., Cheng, T., Lei, J., Xie, Z., Zhang, M.Q., and Wang, X. (2019). Regulation by competition: a hidden layer of gene regulatory network. Quant. Biol. 7, 110–121.

Wei, L., Li, S., Zhang, P., Hu, T., Zhang, M.Q., Xie, Z., and Wang, X. (2021). Characterizing microRNAmediated modulation of gene expression noise and its effect on synthetic gene circuits. Cell Rep. 36, 109573.

Yuan, Y., Liu, B., Xie, P., Zhang, M.Q., Li, Y., Xie, Z., and Wang, X. (2015). Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. Proc. Natl. Acad. Sci. U S A *112*, 3158– 3163.

Yuan, Y., Ren, X., Xie, Z., and Wang, X. (2016). A quantitative understanding of microRNAmediated competing endogenous RNA regulation. Quant. Biol. *4*, 47–57.



