**Abstract**

A stochastic, exploratory model which simulated the molecular dynamic in each cell was adapted from a former work to investigate the feasibility of our project design. The expression of EGFP, which was a representative of a viral early gene expression product during the infection, and L7Ae, which was an inhibitor of EGFP expression, were simulated under the existence of micro-RNA(s) and the given cell condition. Our model results suggest that the different micro-RNA profiles between cancer cells and normal cells can distinguish the expression of EGFP. Future improvements include increasing the copy number of micro-RNA-L by selecting other species of micro-RNA which may improve the performance of the micro-RNA-L inhibition system.

---------------------------------------------------------------------------------------------------------------------------------

A stochastic, exploratory model which simulated the molecular dynamic in each cell was adapted from the model used by L. Wroblewska et al[1] to investigate whether our project design would be feasible.

In our design of the biological experiment, EGFP was used to validate our project design, which was considered as a representative of a viral early gene expression product during the infection. Therefore, this idea was also applied in the model; L7Ae was the inhibitor of EGFP expression by “binding” to the mRNAs of the EGFP and “inhibit” their “translation”, and was also simulated in our model.

**Part I: Assumption**

The different micro-RNA (miRNA, miR) profiles in control cells (which “expressed” neither miR-H nor miR-L), cancer cells (which only “expressed” miR-H) and normal cells (which only “expressed” miR-L) could distinguish the “expression” of EGFP.

**Part II: Model Description**

A detailed demonstration of the model is as shown in Fig. 1.

(Fig. 1)

In general, the model simulates the molecular dynamic in each cell by circularly calculating the probability of the events, including transcription, translation, substrates binding/unbinding, degradation and the inhibition of translation (caused by binding of L7Ae), and then executing one of the events according to the probability. The probability of an event is calculated by multiplying the amount of the reactant(s) involved and the factor(s) of the event (for example, the translation rate for the translation event), and then this value is divided by the sum of the product values of all events calculated as has been noted, and finally gets the probability of the event. All simulations were run in batches of 60 cells and the duration of 24 hours (1440 minutes).

1. A constant amount (as in the same batch of simulations, the same below), **N0**, of plasmid(s) is assumed in a single cell.

2. At some point, the cell “divides” and marks the start of the simulation. **nC**, **nL** are the numbers of the plasmid DNA (pDNA) of EGFP and L7Ae that can be “expressed” in the cell, respectively. To better simulate the transfection behavior, both nC and nL follow a Poisson distribution, which has a mean of N0 and a variance described by **SDfactor** (i.e. variance-to-mean rate, VMR).

3. Transcription: The gene of EGFP and L7Ae can be “transcribed” to message RNAs (mRNAs) of EGFP and L7Ae (**mC** and **mL**, respectively).

4. Translation: The mRNAs of EFGP and L7Ae can be “translated” to EGFP and L7Ae proteins (**C** and **L**, respectively).

5. Degradation of mRNAs and proteins: the degradation of mRNAs and proteins in the normal situation.

6. Substrates binding/unbinding: the binding/unbinding behavior of substrates, which include miR-H “bind to/unbind from” mRNA of L7Ae (mL7Ae), miR-L “bind to/unbind from” mEGFP, L7Ae “(doubly) bind to/unbind from” mEGFP.

7. The translation inhibition of mEGFP caused by L7Ae “(doubly) bind/unbind” to mEGFP: it is considered that when L7Ae doubly bind to mEGFP, the translation is inhibited as twice as much as when L7Ae singly binds to mEGFP.

8. The degradation of mRNAs caused by miR: After the miR “binds” to the target mRNA, the RISC complex can “degrade” the mRNA[2]. To simplify the simulation, it is considered that the total amount of miRs in a cell is constant; also, it is considered that the miRs will not be “degraded” during the event (the miRs are released/unbound from the “degraded” mRNAs).

**Part III: Model Results**

**The different miRNA profiles in control cells, cancer cells and normal cells could distinguish the “expression” of EGFP**

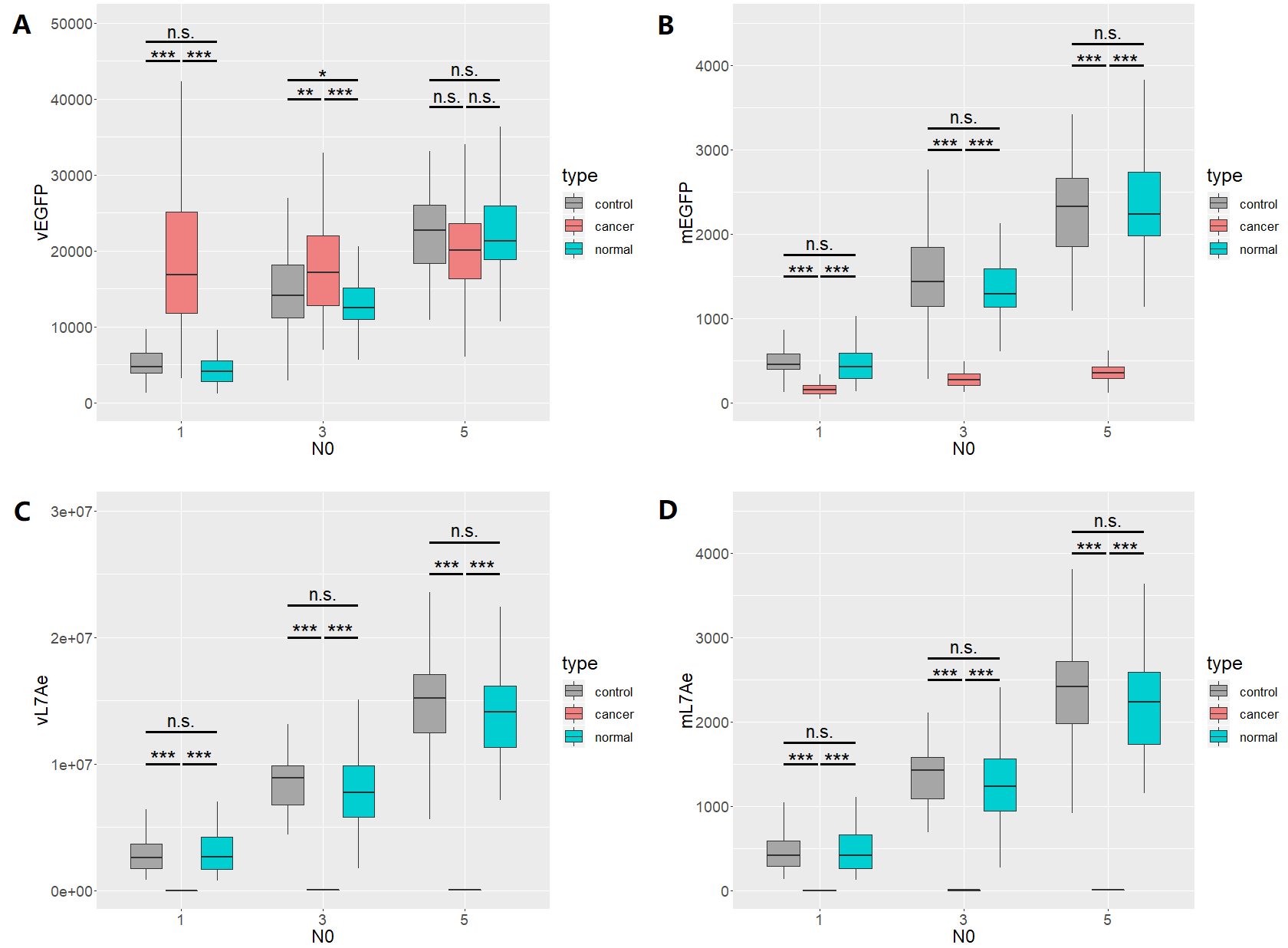
Some of our results show that the different miRNA profiles in control cells, cancer cells and normal cells can distinguish the “expression” of EGFP (N0=3, Fig. 2A), while others do not completely support this assumption (N0=1 or 5, Fig. 2A).

Further investigation was done by analyzing the amount of mEGFP (Fig. 2B), the “expression” of L7Ae (Fig. 2C), and the amount of mL7Ae (Fig. 2D). There is no significant difference of vL7Ae (variable forms of L7Ae proteins) or mL7Ae between normal cells and control cells, and there is a very significant difference of vL7Ae and mL7Ae between cancer cells and control cells (Fig. 2C-D). These results show that both the L7Ae inhibition system and the miR-H inhibition system work well.

There is no significant difference of mEGFP between control cells and normal cells (Fig. 2B), which suggests that the miR-L system does not work so well, probably because the total copy number of miR-L is too low.

Also, significant difference of mEGFP is observed between control cells and cancer cells, which is not expected, and it is even odd that the mEGFP in cancer cells is lower than in control cells (Fig. 2B). After excluding possible bugs, repeating the simulations with the exact same parameters and still getting the same result, we infer that a technical problem may occur during the simulation, which is that because the number of simulation repeats was limited, and only one event was executed in every simulation repeat, the operation of miR-H system may occupy the simulation repeat which may be used by mEGFP “synthesis”, and the copy number of mEGFP was unexpectedly affected.

In conclusion, these results show that the L7Ae inhibition system and the miR-H inhibition system work well while the miR-L inhibition system does not.



**Figure 2** Copy numbers of vEGFP (**A**), mEGFP (**B**), vL7Ae (**C**) and mL7Ae (**D**). The P-values are generated by Welch Two Sample t-test. Significance Symbols: \*, P-value<0.05; \*\*, P-value<0.01; \*\*\*, P-value<0.001; n.s., not significant, P-value≥0.05.

**Part IV: Parameters Configuration and Relevant Data**

**Table 1** Parameters configuration of the model

|  |  |  |
| --- | --- | --- |
| **Object** | **Description** | **Value** |
| N0 | starting copy number of plasmid(s) | various: 1,3,5 |
| SDfactor | the variance-to-mean rate of the Poisson distribution followed by nC and nL | 0.3 |
| kTS | transcription rate (1/min) | 1[1] |
| kTL | translation rate (1/min) | 8[1] |
| degR | mRNA degradation rate (1/min) | 0.002[1] |
| degP | protein degradation rate (1/min) | 5e-4[1] |
| sigma | L7Ae translation inhibition factor | 3e-3[1] |
| sigma2 | L7Ae translation inhibition factor (doubly bound) | 1.5e-3[1] |
| kON\_L | L7Ae binding rate (1/molec/min) | 0.00024[1] |
| kOFF\_L | L7Ae unbinding rate (1/molec/min) | 0.01[1] |
| kcat\_miR | mRNA degradation rate in RISC complex (1/min) | 4.26e-1[2] |
| kON\_miR | miR binding rate (1/molec/min) | 1e-1[2]\* |
| kOFF\_miR | miR unbinding rate (1/molec/min) | 4.14e-1[2] \* |
| miR-H | miR-H copy number in a cancer cell | 55[3,4,5] # |
| miR-L | miR-L copy number in a normal cell | 10[3,4,5] # |

\* kON\_miR and kOFF\_miR were calculated according to the formula Km=(kcat\_miR+kOFF\_miR)/kON\_miR. Km=8.4(nM)[2].

# These are estimated copy numbers of miR-H and miR-L according to the ACN (absolute copy number in a cell) - RPM (reads per million) relationship. It is estimated that in the MCF-7 cell line, the value of miR ACN is about 2-4 folds of RPM[3,4]. In our project design, we used hsa-miR-592 as miR-H and hsa-miR-885-5p as miR-L. The RPM of hsa-miR-592 and hsa-miR-885-5p in colon cancer (COAD) are 20.74 and 3.71, respectively, according to the starbase2 database[5]. Therefore, the estimated copy numbers of miR-H and miR-L are 55 and 10 in a cell, respectively.

**Part V: Conclusion and Future Improvements**

Our modeling suggests that our project design is feasible that the different micro-RNA profiles in cancer cells and normal cells can distinguish the expression of EGFP. The L7Ae inhibition system and the miR-H inhibition system work well. Future improvements include increasing the copy number of miR-L by selecting other species of miR which may improve the performance of the miR-L inhibition system.

**References**

[1] L. Wroblewska, T. Kitada, K. Endo, et al. Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery (2015). Nature Biotechnology 33(8), 839-841.

[2] B. Haley and P. D Zamore. Kinetic analysis of the RNAi enzyme complex (2004). Nature Structural and Molecular Biology 11(7), 599-606.

[3] Y. Song, D. Kilburn, J. H. Song, et al. Determination of absolute expression profiles using multiplexed miRNA analysis (2017). PLoS One 12(7), e0180988.

[4] B. Panwar, G. S. Omenn, Y. Guan. miRmine: A Database of Human miRNA Expression Profiles (2017). Bioinformatics 33(10), 1554-1560.

[5] JH. Li, S. Liu, H. Zhou, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data (2014). Nucleic Acids Research 42(Database issue): D92-7.