Vol.2013-BIO-35 No.3
2013/9/19

Evaluating Effectiveness of Accessibility to Infer RNA–RNA Interactions

Yuki Kato*,1,a) Tomoya Mori*,2 Kengo Sato[3] Shingo Maegawa[4] Hiroshi Hosokawa[4] Tatsuya Akutsu[2]

Abstract: RNA–RNA interaction is involved in post-transcriptional regulation of gene expression, and a good deal of effort has been made on computational prediction of RNA–RNA interactions from sequence data. Recently, a few studies have incorporated interaction site accessibility into their prediction methods, only indicating enhancement of predictive performance on limited interaction data. In this short report, we show the effectiveness of incorporating accessibility into our prediction model in both in silico and in vitro experiments. Our results reveal that elaborate incorporation of accessibility into our prediction method suggests the possibility of discerning real interacting RNAs.

1. Introduction

Regulatory non-coding RNAs often interact with other RNAs to perform their functions. Hence, identification of their interacting partners not only clarifies their functional roles but also exploits the possibility of completing interaction networks. Several computational methods have been developed to predict RNA–RNA interactions from sequence data, ranging from the approaches that look for the simplest type of interactions with almost perfect complementarity [1], [2], [3] to those aiming to predict more complex joint structures [4], [5], [6], [7], [8], [9]. Instead of taking each secondary structures into account explicitly, considering accessibility of interaction sites is also a good approach to improve their prediction accuracy [10], [11], [12], [13].

In general, there is a trade-off between the efficiency of a prediction method and the class of predictable structures. To clear this trade-off, we have recently proposed `RactIP`, a computational method for predicting RNA–RNA interACTions with complex joint structures using Integer Programming (IP) [14]. Although `RactIP` achieved better predictive performance in both run-time and accuracy as compared with other earlier methods for predicting RNA–RNA interactions, it was validated on a limited set of RNA pairs with known interaction sites available at that time, posing the question of how well it can work on RNA pairs whose interactions remain to be unknown.

Recently, Richter and Backofen [15] have shown that accessibility of experimentally verified interaction sites is significantly higher than that of predictive ones. Considering this observation, we present an improved version of `RactIP` that incorporates accessibility by modifying the IP formulation. In silico experiments on a newly compiled set of RNA pairs with known interaction sites showed that the new version of `RactIP` is superior in accuracy to the earlier `RactIP`. Moreover, for seven RNA pairs whose interactions were unknown, we performed in vitro experiments to determine whether they interact with the partners as well as inferring it by the new `RactIP`, yielding good results of their agreement.

In this short report, we overview our proposed method and preliminary results. Details of the method and results will be reported elsewhere.

2. Methods

RNA–RNA interaction can be interpreted as a set of external base pairs, which is also called an interaction site. Moreover, when two sets of internal base pairs, i.e., respective secondary structures in two RNAs are considered, the interacting structure is called a joint structure. Our model `RactIP` aims to predict RNA joint structures from a pair of RNA sequences by using IP. IP is one of the major techniques in optimization and has great flexibility to model a wide variety of combinatorial problems.

The improved version of `RactIP` presented here incorporates precomputed accessibility information into the earlier model. Accessibility of an interaction site is considered to be the probability of being unpaired in that sequence interval. This probability can be calculated by a partition function-based method such as `RNAplfold` [16] and `RNAup` [10] in the ViennaRNA package [17]. Accessibility information can be mainly incorporated into the linear constraints in our IP formulation.

[1] Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), 8916–5 Takayama, Ikoma, Nara 630–0192, Japan
[2] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611–0011, Japan
[3] Faculty of Science and Technology, Keio University, 3–14–1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223–8522, Japan
[4] Graduate School of Informatics, Kyoto University, 36–1 Yoshida-Honmachi, Sakyo-ku, Kyoto 606–8501, Japan
a) ykato@is.naist.jp
*These authors contributed equally to this work.

3. Results

3.1 Prediction of known RNA–RNA interactions

We compared the new version of RactIP that considers accessibility, which we call RactIP-a, with the old version of RactIP [14], RNAplex [2], its improved version RNAplex-a [13] and IntaRNA [11] by predicting interacting structures on a dataset of 80 RNA pairs with experimentally verified interaction sites. It should be noted that RactIP-a, RNAplex-a and IntaRNA are prediction models that take accessibility into consideration, although the way to incorporate accessibility is different from each other.

The results show that RactIP-a is better than RactIP and at least comparable to IntaRNA in prediction accuracy measured by Matthews correlation coefficient (MCC). In contrast, RNAplex-a, an accessibility-based extension of RNAplex, is significantly better in MCC than not only RNAplex but also the other competitive methods.

3.2 Inference of unknown RNA–RNA interactions

To investigate how well each prediction method stated above can work on RNA pairs whose interactions remain to be unknown, we carried out in silico and in vitro experiments on another set of seven RNA pairs. More specifically, each program was tested on the dataset to judge whether given two RNAs interact or not. At the same time, we performed in vitro experiments on the same seven RNA pairs to verify their actual interactions.

In computational experiments, all methods other than RactIP-a output interaction sites of all the given RNAs, whereas RactIP-a predicted external base pairs for only two out of seven RNA pairs. As for gel mobility shift assays, we observed that two out of seven pairs interacted, which are the same RNA pairs as predicted by RactIP-a.

4. Conclusion

We have extended our prediction method RactIP to take accessibility of interaction site into consideration. In silico experiments on the dataset with experimentally verified interaction sites have shown that the improved version of RactIP is superior in accuracy to the earlier one. The other experimental results on RNA pairs with unknown interactions have indicated that our proposed method opens up the possibilities of discriminating real interacting RNAs from putative ones.

RactIP-a outperforms the other methods in discriminating real interacting RNAs from putative ones. This evaluation is, however, qualitative one in terms of whether there is at least one external base pair between given two RNAs. To have more reliable evaluation, we should use quantitative measure such as melting temperature, which is also left as future work.

The program RactIP-a is available at http://code.google.com/p/ractip/.

Acknowledgments The authors would like to thank Dr. Andreas Richter for providing them with part of sequence data. This work was supported in part by JSPS KAKENHI [#24700296 to Y.K.].

References

[1] Rehmsmeier, M., Steffen, P., Höchsmann, M. and Giegerich, R.: Fast and effective prediction of microRNA/target duplexes, *RNA*, vol. 10, pp. 1507–1517 (2004).
[2] Tafer, H. and Hofacker, I.L.: RNAplex: a fast tool for RNA–RNA interaction search, *Bioinformatics*, vol. 24, pp. 2657–2663 (2008).
[3] Wenzel, A., Akbasli, E. and Gorodkin, J.: RIsearch: fast RNA–RNA interaction search using a simplified nearest-neighbor energy model, *Bioinformatics*, vol. 28, pp. 2738–2746 (2012).
[4] Pervouchine, D.D: IRIS: intermolecular RNA interaction search, *Genome Inform.*, vol. 15, pp. 92–101 (2004).
[5] Alkan, C., Karakoc, E., Nadeau, J.H., Sahinalp, S.C. and Zhang, K.: RNA–RNA interaction prediction and antisense RNA target search, *J. Comput. Biol.*, vol. 13, pp. 267–282 (2006).
[6] Chitsaz, H., Salari, R., Sahinalp, S.C. and Backofen, R.: A partition function algorithm for interacting nucleic acid strands, *Bioinformatics*, vol. 25, pp. i365–i373 (2009).
[7] Kato, Y., Akutsu, T. and Seki, H.: A grammatical approach to RNA–RNA interaction prediction, *Pattern Recognit.*, vol. 42, pp. 531–538 (2009).
[8] Huang, F.W.D., Qin, J., Reidys, C.M. and Stadler, P.F.: Target prediction and a statistical sampling algorithm for RNA–RNA interaction, *Bioinformatics*, vol. 26, pp. 175–181 (2010).
[9] Salari, R., Möhl, M., Will, S., Sahinalp, S.C. and Backofen, R.: Time and space efficient RNA–RNA interaction prediction via sparse folding, *Lect. Notes Bioinform.*, vol. 6044, pp. 473–490 (2010).
[10] Mückstein, U. *et al.*: Thermodynamics of RNA-RNA binding, *Bioinformatics*, vol. 22, pp. 1177–1182 (2006).
[11] Busch, A., Richter, A.S. and Backofen, R.: IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions, *Bioinformatics*, vol. 24, pp. 2849–2856 (2008).
[12] Salari, R., Backofen, R. and Sahinalp, S.C.: Fast prediction of RNA–RNA interaction, *Algorithms Mol. Biol.*, vol. 5, no. 5 (2010).
[13] Tafer, H., Amman, F., Eggenhofer, F., Stadler, P.F. and Hofacker, I.L.: Fast accessibility-based prediction of RNA–RNA interactions, *Bioinformatics*, vol. 27, pp. 1934–1940 (2011).
[14] Kato, Y. *et al.*: RactIP: fast and accurate prediction of RNA–RNA interaction using integer programming, *Bioinformatics*, vol. 26, pp. i460–i466 (2010).
[15] Richter, A.S. and Backofen, R.: Accessibility and conservation: general features of bacterial small RNA–mRNA interactions?, *RNA Biol.*, vol. 9, pp. 954–965 (2012).
[16] Bernhart, S.H., Hofacker, I.L. and Stadler, P.F.: Local RNA base pairing probabilities in large sequences, *Bioinformatics*, vol. 22, pp. 614–615 (2006).
[17] Lorenz, R. *et al.*: ViennaRNA package 2.0, *Algorithms Mol. Biol.*, vol. 6, no. 26 (2011).