

COMPUTATIONAL CHALLENGES IN THE STUDY OF SMALL REGULATORY RNAS

DORON BETEL

*Computational and Systems Biology Center, Memorial Sloan-Kettering Cancer Center
New York, NY 10065, U.S.A.*

CHRISTINA LESLIE

*Computational and Systems Biology Center, Memorial Sloan-Kettering Cancer Center
New York, NY 10065, U.S.A.*

NIKOLAUS RAJEWSKY

*Max Delbrück Centrum for Molecular Medicine
Berlin, Germany*

1. Introduction

Small regulatory RNAs are a class of non-coding RNAs that function primarily as negative regulators of other RNA transcripts. The principal members of this class are microRNAs and siRNAs, which are involved in post-transcriptional gene silencing. These small RNAs, which in their functional form are single-stranded and ~22 nucleotides in length, guide a gene silencing complex to an mRNA by complementary base pairing, mostly at the 3' untranslated region (3'UTR)^{1,2}. The association of the silencing complex to the conjugate mRNA results in silencing the gene either by translational repression or by degradation of the mRNA. The discovery of microRNAs and their regulatory mechanism has been at the center of a dogmatic shift of our view of non-coding RNAs and their biological role.

In recent years, microRNAs have emerged as a major class of regulatory genes central to a wide range of cellular activities, including stem cell maintenance, developmental timing, metabolism, host-viral interaction, apoptosis and neuronal gene expression and muscle proliferation³. Consequently, changes in the expression, sequence or target sites of microRNA are associated with a number of human genetic diseases⁴. Indeed, microRNAs are known to act both as tumor suppressors and oncogenes, and aberrant expression of microRNAs is associated with progression of cancer⁵. The importance of genetic regulation by microRNAs is reflected in their ubiquitous expression in almost all cell types as well as their conservation in most of the metazoan and plant species.

The molecular pathway of gene silencing by microRNAs is also the basis for RNA interference (RNAi), a powerful experimental technique that is used to selectively silence genes in living cells. This technique has gained wide use and is currently employed in a high throughput manner to investigate the effects of large scale gene repression⁶.

In addition to microRNAs and siRNAs, new types of regulatory small RNAs have been identified, including rasiRNAs⁷ in *Drosophila* and zebrafish, PIWI-interacting RNAs (piRNAs) in mammals⁸ and 21U-RNAs in *C. elegans*⁹. Collectively, the discovery of these sequences and their regulatory role has had a profound impact on our understanding on the post-transcriptional regulation of genes, suppression of transposable elements, heterochromatin formation and programmed gene rearrangement.

2. Session papers

The accelerated pace of biochemical and functional characterization of microRNAs and other small regulatory RNAs has been facilitated by computational efforts, such as microRNA target predictions, conservation and phylogenetic analysis, microRNA gene predictions and microRNA expression profiling. The papers in this session exemplify some of the primary challenges in this field and the novel approaches used to address them.

With the advent of pyrosequencing technology investigators can now identify many of the sparse and short genomic transcripts that have previously eluded detection. Not surprisingly, pyrosequencing has become the primary method for the detection and characterization of new microRNAs¹⁰ as well as the discovery of new regulatory RNAs such as piRNAs. One difficulty with this technology is the high rate of sequencing errors, which can be corrected to some degree by the assembly of partially overlapping fragments. The first paper in this session, by Vacic et al., addresses the problem of correcting sequencing errors in short reads that are typical in small RNA discovery where there is no fragment assembly step. They present a probabilistic framework to evaluate short reads by matching them against the genome from which the sequences are derived.

A central and still unresolved problem in the field of small regulatory RNAs is the prediction of the mRNA targets of a microRNA. Typical computational approaches search for a (near) perfect base-pairing between the 5' end of the microRNA and a complementary site in the 3' UTR of the potential target gene. Some algorithms also incorporate binding at the 3' end of the microRNA to the target or make use of conservation of target sites across species¹¹. So far, these sequence-based approaches result in a large number of predictions, suggesting that more refined rules governing microRNA-mRNA interactions remain to be discovered. In the second paper in the session, Long et al. provide new results in support of their recent energy-based model for microRNA target prediction.

They model the interaction between microRNA and target as a two-step hybridization reaction: (1) nucleation at an accessible target site, followed by (2) hybrid elongation to disrupt local target secondary structure and form the complete microRNA-target duplex. The authors present analysis of a set of microRNA-mRNA interactions that have been experimentally tested in mammalian systems.

Tissue-specific microRNA expression data can be also be exploited for target prediction and integrative models of microRNA gene silencing. The final paper in the session, from Huang et al., adopts such an approach in a development of their GenMiR model. Here, they integrate paired microRNA and mRNA expression data, predicted microRNA target sites, and mRNA sequence features associated with the predicted sites in a probabilistic approach for scoring candidate microRNA-mRNA target sites.

Acknowledgments

We thank all the authors who submitted papers to the session, and we gratefully acknowledge the reviewers who contributed their time and expertise to the peer review process.

References

1. D. P. Bartel, *Cell* **116** (2), 281 (2004).
2. P. D. Zamore and B. Haley, *Science (New York, N.Y)* **309** (5740), 1519 (2005).
3. C. Xu, Y. Lu, Z. Pan et al., *Journal of cell science* **120** (Pt 17), 3045 (2007); M. Kapsimali, W. P. Kloosterman, E. de Bruijn et al., *Genome Biol* **8** (8), R173 (2007).
4. J. S. Mattick and I. V. Makunin, *Human molecular genetics* **15 Spec No 1**, R17 (2006).
5. G. A. Calin and C. M. Croce, *Nature reviews* **6** (11), 857 (2006).
6. Y. Pei and T. Tuschl, *Nature methods* **3** (9), 670 (2006).
7. V. V. Vagin, A. Sigova, C. Li et al., *Science (New York, N.Y)* **313** (5785), 320 (2006).
8. V. N. Kim, *Genes & development* **20** (15), 1993 (2006).
9. J. G. Ruby, C. Jan, C. Player et al., *Cell* **127** (6), 1193 (2006).
10. K. Okamura, J. W. Hagen, H. Duan et al., *Cell* **130** (1), 89 (2007).
11. N. Rajewsky, *Nature genetics* **38 Suppl**, S8 (2006).