REGULATORY RNA

DRENA DOBBS

Department of Genetics, Development & Cell Biology, Iowa State University
Ames, IA 50011 USA
Email: ddobbs@iastate.edu

STEVEN E. BRENNER

Department of Plant & Microbial Biology, University of California
Berkeley, CA 94720 USA
Email: brenner@compbio.berkeley.edu

VASANT G. HONAVAR

Departments of Genomics & Bioinformatics and Neuroscience, Pennsylvania State University
University Park, PA 16802 USA
Email: vhonavar@ist.psu.edu

ROBERT L. JERNIGAN

L. H. Baker Center for Bioinformatics & Biological Statistics,
Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University
Ames, Iowa, 50011, USA
Email: jernigan@iastate.edu

ALAIN LAEDERACH

Department of Biology, University of North Carolina
Chapel Hill, North Carolina 27514, USA
Email: alain@unc.edu

QUAID MORRIS

Departments of Molecular Genetics and Computer Science, University of Toronto
Toronto, ON M5S 3E1, Canada
Email: quaid.morris@utoronto.ca

Advances in both experimental and computational approaches to genome-wide analysis of RNA transcripts have dramatically expanded our understanding of the ubiquitous and diverse roles of regulatory non-coding RNAs. This conference session includes presentations exploring computational approaches for detecting regulatory RNAs in RNA-Seq data, for analyzing i*n vivo* CLIP data on RNA-protein interactions, and for predicting interfacial residues involved in RNA-protein recognition in RNA–protein complexes and interaction networks.

Discoveries over the past decade have revealed the previously unsuspected diversity of non-coding RNAs, which are ubiquitous in living organisms and play key roles in regulating gene expression and in organizing genomes [1-2]. The last time PSB had an RNA-focused session was in 2010. Since then the field of RNA regulation has been transformed by new data on RNA-based regulation (e.g., ENCODE and ENCORE); new experimental methods for genome-wide probing of RNA structure [e.g., 3] and translation [e.g., 4]; new knowledge about RNA-protein interactions [e.g., 5-8]; and new computational techniques to predict RNA regulation [e.g., 9].

Even though the human genome encodes nearly as many "non-coding" RNAs as it does protein-encoding mRNAs, the cellular functions of most ncRNAs remain unknown [10]. Also, despite impressive recent progress in annotating RNA-binding proteins [6, 11-12] and characterizing their binding sites [7, 13], our understanding of the roles of RNA-binding proteins and the determinants of RNA-protein recognition lag far behind our understanding of the mechanisms and regulatory roles of transcription factors. What we know at present is that ncRNAs have important functions in both *pre-* (e.g., epigenetic) and *post*-transcriptional regulation of gene expression in eukaryotes, as well as in bacteria [14] and viruses [15]. The expanding roles of ncRNAs in normal development and disease [16-17], the discovery that the oldest known alternative RNA splicing is for gene regulation, rather than producing alternate protein isoforms [18], the unanticipated abundance of circular RNAs in the brain [19], and their emerging roles in human disease [20], and novel potential applications of small RNAs in personalized medicine [21] all present extraordinary opportunities for productive interactions and collaborations among biologists and computer scientists.

The goal of this session is to bring together scientists who investigate structures, functions and dynamics of RNA, RNA-protein complexes, and RNA-protein interaction networks, with a focus on identifying and functionally annotating regulatory RNAs and RNPs. This is a data-rich field within molecular biology, where many terabytes of transcriptomics data have already been collected but remain largely unanalyzed. The technical advances mentioned above have resulted in an enormous expansion of available RNA sequences, structures and expression data – and highlighted an increasing gap in our functional understanding. This *Regulatory RNA* discussion session will provide a timely opportunity to discuss recent successes in computational, biochemical, biophysical and genetic approaches for studying the roles of non-coding RNAs, RNPs and RNA-protein interaction networks. This is a first step toward addressing both an urgent need and outstanding prospects for integrating computational and experimental methods in the quest for a better understanding of the fascinating structures and functions of regulatory RNAs and RNA-protein complexes.

Session Contributions

Jennifer Doudna, a Howard Hughes Medical Institute Investigator at the University of California, Berkeley, will provide the invited lecture. Doudna's laboratory has provided key insights into RNA-mediated gene regulation, including landmark discoveries on the molecular mechanisms of RNA interference and translational control in eukaryotes and on CRISPR-Cas immunity in bacteria.

One of the major challenges in studying regulatory RNAs is accurately detecting and quantifying

them on a genome-wide scale. Many individual groups have grappled with both the biological (RNA decay and fragmentation) and quantitative (substring detection) aspects of the problem. The first paper in our session, from **Pena-Castillo** et al., focuses on the computational detection of small non-coding RNAs (sRNAs). sRNAs are bacterial regulatory RNAs that play important roles in stress response, virulence and a variety of other cellular processes. The authors present a systematic comparative assessment of available computational methods for detecting sRNAs in RNA-Seq data. They provide a valuable and balanced summary of the relative strengths and weaknesses of several approaches and propose a novel approach, ***sRNA-Detect***, which provides a higher retrieval rate than other methods.

Many regulatory RNAs carry out their functions by interacting with proteins, other RNAs or DNA. Recent technical advances in the detection of RNA-protein interactions *in vivo* include the development of protocols such as PAR-CLIP, HITS-CLIP and i-CLIP. The second paper in our session, from **Kassuhn** et al., provides a comparative evaluation of data processing tools for experimental CLIP-data. The paper describes a novel tool, ***Cseq-Simulator***, which can generate simulated datasets that can serve as valuable surrogates for an experimental gold standard CLIP dataset. Their comparisons point out which software tools are most useful for different purposes and identify some of the common pitfalls in genome-wide analysis of RNA-protein interactions.

Despite the tremendous advances in experimental approaches for detecting RNA-protein interactions, the sequence and structural determinants of specific recognition in RNA-protein complexes are still not well understood. In the final paper in our session, **Muppirala** et al. describe a computational method for predicting interfacial residues in RNA-protein complexes. They propose a machine learning approach that exploits sequence motifs found in the binding interfaces of known RNA-protein complexes. They propose a "partner-specific" method, ***PS-PRIP***, which simultaneously predicts interfacial residues for both the RNA and the protein components of complexes and is capable of predicting different interfaces for different binding partners. This method should be useful for identifying potential interfacial residues in RNA-protein complexes where structural information is not available.

The papers in this session provide an introduction to critical challenges facing researchers in the rapidly expanding field of regulatory RNA, and emphasize that this subject will undoubtedly continue to increase in importance for both basic and translational biology.

Acknowledgements

We thank all of the authors who submitted papers for this session and all of the reviewers who contributed their time and expertise. We acknowledge the Center for RNA Systems Biology P50 GM102706 to SEB and NIH R01 GM066387 to VGH, RLJ and DD for support. We are grateful to the PSB organizers for their support and especially Tiffany Murray for meeting coordination.

References

1. K. V. Morris and J. S. Mattick, *Nat Rev Genet.* **15**, 423 (2014).
2. J. W. Pek and K. Okamura, *Wiley Interdiscip Rev RNA*. **Oct 1.** doi: 10.1002/wrna.1309 (2015).

3. Y. Wan, K. Qu, Q. C. Zhang, R. A. Flynn, O. Manor, Z. Ouyang, J. Zhang, R. C. Spitale, M. P. Snyder, E. Segal and H. Y. Chang., *Nature.* **505**, 706 (2014).
4. N. T. Ingolia, L. F, Lareau and J. S. Weissman, *Cell* **147**, 789 (2011).
5. A. G. Baltz, M. Munschauer, B. Schwanhäusser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, and M. Landthaler, *Mol Cell.* **46**, 674 (2012).
6. A. Castello, B. Fischer, K. Eichelbaum, R. Horos, B. M. Beckmann, C. Strein, N. E. Davey, D. T. Humphreys, T. Preiss, L. M. Steinmetz, J. Krijgsveld and M. W. Hentze, *Cell* **149**, 1393 (2012).
7. D. Ray, H. Kazan, K. B. Cook, M. T. Weirauch, H. S. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L. H. Matzat, R. K. Dale, S. A. Smith, C. A. Yarosh, S. M. Kelly, B. Nabet, D. Mecenas, W. Li, R. S. Laishram, M. Qiao, H. D. Lipshitz, F. Piano, A. H. Corbett, R. P. Carstens, B. J. Frey, R. A. Anderson, K. W. Lynch, L. O. Penalva, E. P. Lei, A. G. Fraser, B. J. Blencowe, Q.D. Morris and T. R. Hughes, *Nature* **499**, 172, (2013).
8. K. B. Cook, T. R. Hughes and Q. D. Morris, *Brief Funct Genomics* **14**, 74 (2015).
9. H. Y. Xiong, B. Alipanahi, L. J. Lee, H. Bretschneider, D. Merico, R. K. Yuen, Y. Hua, S. Gueroussov, H.S. Najafabadi, T. R. Hughes, Q. Morris, Y. Barash, A. R. Krainer, N Jojic, S. W. Scherer, B. J. Blencowe, B. J. Frey, *Science* **347,** 1254806 (2015).
10. J. S. Mattick and J. L. Rinn, *Nat Struct Mol Biol.* **22**, 5 (2015).
11. A. N. Brooks, M. O. Duff, G. May, L. Yang , M. Bolisetty, J. Landolin, K. Wan, J. Sandler, S. E. Celniker, B. R. Graveley and S. E. Brenner, *Genome Res.* **Aug 20.** pii: gr.192518.115. (2015).
12. A. Re, T. Joshi, E. Kulberkyte, Q Morris and C. T. Workman, *Methods Mol Biol.* **1097,** 491, (2014).
13. X. Li, H. Kazan, H. D. Lipshitz and Q. D. Morris, *Wiley Interdiscip Rev RNA*. **5,** 111 (2014).
14. A. K. Chaudhary, D. Na and E. Y. Lee, *Biotechnol Adv*. **33**, 914 (2015).
15. K. T. Tycowski, Y. E. Guo, N. Lee, W. N. Moss, T. K. Vallery, M. Xie and J. A. Steitz, *Genes Dev.* **29**, 567 (2015).
16. P. J. Batista and H. Y. Chang, *Cell*, **153**, 1298 (2013).
17. H. Ling, K. Vincent, M. Pichler, R. Fodde, I. Berindan-Neagoe, F. J. Slack and G. A. Calin, *Oncogene* **34,** 5003, (2015).
18. L. F. Lareau and S. E. Brenner, *Mol Biol Evol*. **32**, 1072 (2015).
19. A. Rybak-Wolf, C. Stottmeister, P. Glažar, M. Jens, N. Pino, S. Giusti, M. Hanan, M. Behm, O. Bartok, R. Ashwal-Fluss, M. Herzog, L. Schreyer, P. Papavasileiou, A. Ivanov, M. Öhman, D. Refojo, S. Kadener, N. Rajewsky. *Mol Cell.* **58**, 870 (2015).
20. S. Qu, X. Yang, X. Li, J. Wang, Y. Gao, R. Shang, W. Sun, K. Dou and H. Li, *Cancer Lett.* **365,** 141 (2015).
21. A.C. Solem, M. Halvorsen, S. B. Ramos and A. Laederach, *Wiley Interdiscip Rev RNA*. **6,** 517 (2015).