Charting the Unexplored RNA-binding Protein Atlas of the Human Genome

Huiying Zhao[1], Yuedong Yang[1], Sarath Chandra Janga[2], Jason Chen[3], Heng Zhu[4], Cheng Kao[5], and Yaoqi Zhou[1]

[1]School Of Informatics, IUPUI/Center for Computational Biology & Bioinformatics, IU School of Medicine, Indiana [2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Illinois
[3]Graduate Institute of Systems Biology and Bioinformatics, National Central University, Taiwan,
[4]Department of Pharmacology and Molecular Sciences, Johns Hopkins School of Medicine, Maryland
[5]Department of Molecular & Cellular Biochemistry, Indiana University, Bloomington, Indiana

Abstract

Detecting protein-RNA interactions is challenging–both experimentally and computationally– because RNAs are large in number, diverse in cellular location and function, and flexible in structure. As a result, many RNA-binding proteins (RBPs) remain to be identified and characterized. Recently, we developed a bioinformatics tool called SPOT-Seq that integrates template-based structure prediction with RNA-binding affinity prediction to predict RBPs. Application of SPOT-Seq to human genome leads to doubling of RBPs from 2115 to 4296. Half of novel (>2000) RBPs are poorly or not annotated. The other half possesses Gene Ontology leaf IDs that are associated with known RBPs. In particular, we identified 36 novel RBPs in cancer, cardiovascular, diabetes and neurodegenerative pathways and 26 novel RBPs associated with disease-causing SNPs. Half of these disease-associating, predicted novel RBPs are annotated to interact with known RBPs. Accuracy of predicted novel RBPs is further validated by same confirmation rate of novel and annotated RBPs in human proteome microarrays experiments. The large number of predicted novel RBPs and their abundance in disease pathways and disease-causing SNPs are useful for hypothesis generation. These predicted novel human RBPs (>2000) with confidence level and their predicted complex structures with RNA can be downloaded from http://sparks.informatics.iupui.edu (yqzhou@iupui.edu)