Dissecting the function, evolution and disease associations of an atlas of human RNA binding proteins

Yaseswini Neelamraju¹, Seyedsasan Hashemikhabir¹, Sarath Chandra Janga¹, 2, 3,*

¹Department of Biohealth Informatics, School of Informatics and Computing, Indiana University Purdue University, 719 Indiana Ave Ste 319, Walker Plaza Building, Indianapolis, Indiana 46202
²Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 5021 Health Information and Translational Sciences (HITS), 410 West 10th Street, Indianapolis, Indiana, 46202
³Department of Medical and Molecular Genetics, Indiana University School of Medicine, Medical Research and Library Building, 975 West Walnut Street, Indianapolis, Indiana, 46202

RNA Binding Proteins (RBPs) play a central role in mediating post transcriptional regulation of genes. However less is understood about them and their regulatory mechanisms. In this study, we construct a repertoire of 1344 RBPs identified from several experimental studies and present a comprehensive analysis to understand their characteristics at a global scale. The domain architecture of RBPs enabled us to classify them into three groups - Classical (29%), Non-classical (19%) and Unclassified (52%). A higher percentage of proteins with unclassified domains reveals the presence of various uncharacterised motifs that can potentially bind RNA. In addition, enrichment of various unconventional superfamilies suggest that RBPs could form an integral part of the cellular architecture. Further, RBPs were found to be highly disordered compared to non-RBPs (p<2.2e-16, Fisher’s exact test), indicating a dynamic regulatory role of RBPs in cellular functioning. Evolutionary analysis in 62 different species showed that RBPs are highly conserved compared to non-RBPs (p<2.2e-16, Wilcoxon-test), reflecting the conservation of various biological processes like mRNA splicing, ribosome biogenesis. The expression patterns of RBPs from Human Body Map 2.0 revealed that ~60% of them are ubiquitously expressed while ~40% are tissue-specific. Additionally, non-classical proteins were found to be higher expressed than the classical proteins across all the tissues (p<0.001, Wilcoxon test). RBPs were also seen to be highly associated with several neurological disorders, cancer and inflammatory diseases. These analyses are made accessible to researchers in the form of a database called RNA Binding protein expression and disease dynamics database (READ DB).

Mentor : Sarath Chandra Janga, School of Informatics, Indiana University Purdue University, 719 Indiana Ave Ste 319, Walker Plaza Building, Indianapolis, Indiana 46202
Email: scjanga@iupui.edu