A Framework For Inferring Multivalent RNA Motifs And Their Candidate Regulators Of Splicing

Mariachiara Grieco 1,2 , Tommaso Becchi 1,3 , Gabriele Boscagli 3 , Francesca Priante 2 , Livia Caizzi 2,4 , Uberto Pozzoli 5 , Matteo Cereda 1,2,3 1. Department of Oncology and Hemato-Oncology, University of Milan, 20142 Milan, Italy 2. Italian Institute for Genomic Medicine, c/o IRCCS, Str. Prov.le 142, km 3.95, 10060 Candiolo (TO), Italy 3. IFOM ETS - The AIRC Institute of Molecular Oncology, Milan, Italy 4. Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy 5. Bioinformatics Lab, Scientific Institute IRCCS E. MEDEA, Bosisio Parini, Italy

Background Alternative splicing diversifies the proteome and its dysregulation is increasingly recognized as a hallmark of cancer. Beyond core spliceosomal catalysis, exon inclusion is finely tuned by auxiliary RNA-binding proteins (RBPs) that recognize clusters of short sequence motifs, known as multivalent RNA motifs (MRMs). However, inferring which RBP binds a given cis-acting motif and whether the interaction drives cancer-relevant splicing alterations remains difficult from sequence alone, limiting our understanding of the splicing code and the discovery of novel therapeutic intervention points. Methods We developed RNAmotifs 2.0, whose Motifs and Regulators of splicing (MaRs) module integrates in vivo binding (eCLIP) with functional splicing (RNA-seq) evidence from ENCODE to learn RBP-specific positional binding expectations. From user-supplied differentially spliced exons, the module scores candidate MRM-RBP associations and weights them by binding-signal robustness in an interpretable statistical model. Results Across 41 ENCODE knockdown datasets spanning 21 RBPs in HepG2 and K562, RNAmotifs 2.0 was run blind to the perturbed regulator and robustly prioritized it within the top five candidates in 80% of cases, with particularly strong recovery of repressive, exon-silencing regulators and prioritization confidence tracking the magnitude of the underlying splicing effect. Among the most robustly recovered regulators was HNRNPK, an RBP implicated in prostate cancer splicing dysregulation. We therefore challenged RNAmotifs 2.0 independently on RNA-seq from HNRNPK-depleted PC3 prostate cancer cells. Despite minimal exon overlap with the HepG2 eCLIP reference, RNAmotifs 2.0 correctly prioritized HNRNPK for both enhanced and silenced exons via CC-rich motifs matching its known binding specificity, demonstrating that the framework captures transferable, disease-relevant binding architecture rather than dataset-specific signatures, and extending its applicability to cancer transcriptomes beyond ENCODE. Conclusion RNAmotifs 2.0 offers an interpretable, generalizable framework for identifying candidate trans-acting splicing regulators directly from RNA-seq data, providing a route to expose actionable RBP-driven regulatory nodes for therapeutic targeting in cancer.