

Context-specific splicing regulatory network inference from large-scale alternative splicing data

Guangyuan Li, BS^{1,2}, Matthew Weirauch, PhD^{1,4}, Emily Miraldi, PhD^{1,3}, Nathan Salomonis, PhD^{1,2,5}

¹Department of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

²Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

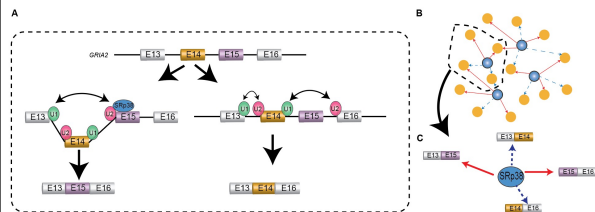
³Division of Immunology, University of Cincinnati College of Medicine, Cincinnati, OH, USA

⁴Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

⁵Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

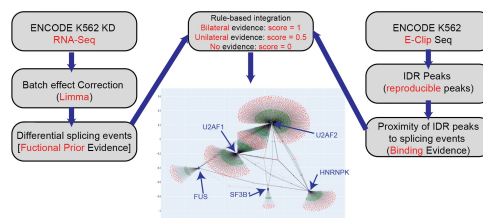


Splicing Regulatory Network (SRN)



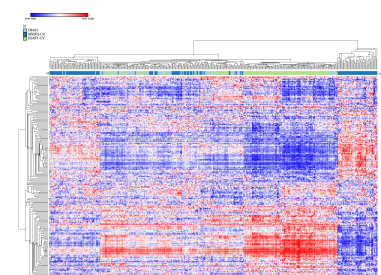
1. Alternative Splicing (AS) alters the composition of encoded protein and is tightly regulated by Splicing Factor (SF) via splicing site selection.
2. **Figure (A)** depicts binding of SRp48 lead to different resultant isoforms
3. Splicing Regulatory Network (SRN), shown in **Figure (B,C)** characterize the regulatory strength between splicing Factor and regulated alternative splicing events.

Building prior network using K562 cell



1. Splicing events were quantified from ENCODE K562 knock-down RNA-Seq data, followed by rigorous batch effect correction. Splicing that are present in the differentially spliced events in certain SF knockdown experiments were deemed functional evidence.
2. E-clip Seq reproducible peaks that fall into the proximal region of splicing events suggest binding evidence for potential regulation
3. We integrate the evidence from both sides via a rule-based schema, part of the prior network (U2AF1, U2AF2, FUS, SF3B1, HNRNPK) are shown above

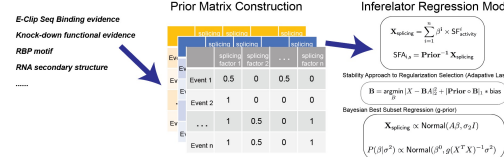
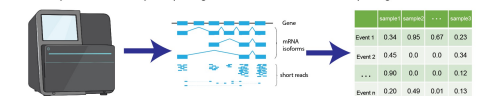
Inferred SF activity recapitulate the splicing architectures in AML



1. U2AF1-CV and SRSF1-CV are two clinically relevant cohorts that we previously identified, corresponding to poor and good prognosis respectively.
2. Splice-Inferelator predicted AML specific SRN, and then inferred the splicing factor activity in each sample.

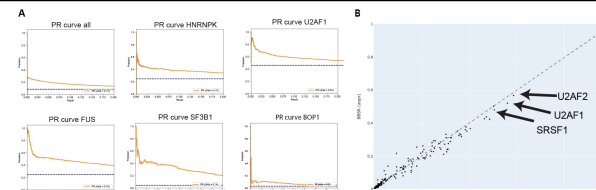
Splice-Inferelator: in-silico construction of SRN leveraging RNA-Seq and prior knowledge

Context-specific RNA Seq Splicing Event Identification Splicing Event Quantification



1. Splice-Inferelator is a multivariate regression model to define partial dependency between splicing event and splicing factor.
2. Splice-Inferelator utilizes either adaptive Lasso or G-prior to incorporate prior information into regression objective function.
3. Splice_inferelator derive splicing factor activity through context-specific transcriptomic data and prior known interactions

Predicting and validating Splicing Regulatory Network in HepG2 cell line



1. We apply splice-Inferelator algorithm on HepG2 cell line using the high-confident prior network derived from K562 cell line. We chose HepG2 for validation because the existence of gold standard deriving from E-clip and large-scale knock-down data.
2. Figure (A) shows the Area Under the Precision-Recall (AUPR) curve for all splicing factors (all) and five representative splicing factors (HNRNPK, U2AF1, FUS, SF3B1, BOP1).
3. Benchmark Two regression algorithms (G-prior and adaptive lasso) shows similar results while STARS-lasso performs marginally better for best-performing splicing factors (U2AF1, U2AF2, SRSF1)

Conclusion

- Splice-Inferelator is capable of building context-specific splicing regulatory network from transcriptomic data alone.
- Splice-Inferelator allows flexible incorporation of various sources of prior interactions
- Splice-Inferelator yields high-confident prediction measured by AUPR
- Splice-Inferelator can infer sample level splicing factor activity to nominate disease associated SF

Download

- Code: <https://github.com/frankligy/splice-inferelator>
- Manuscript in preparation