

Information-theoretic approach for detection of differential splicing from RNA-seq data

Axel Rasche, Ralf Herwig 

Max-Planck-Institute for Molecular Genetics, Dep. Computational Molecular Biology, Berlin, Germany

Rasche A and Herwig R (2015) *EMBnet.journal* **21**(Suppl A), e828. <http://dx.doi.org/10.14806/ej.21.A.828>

The computational prediction of alternative splicing from high-throughput sequencing data is inherently difficult and necessitates robust statistical measures because the differential splicing signal is overlaid by influencing factors such as gene expression differences and simultaneous expression of multiple isoforms, among others. In this work we describe ARH-seq (Rasche *et al.*, 2014), a discovery tool for differential splicing in case-control studies, that is based on the information-theoretic concept of entropy. ARH-seq works on high-throughput sequencing data and is an extension of the ARH method that was originally developed for exon microarrays (Rasche and Herwig, 2010). We show that the method has inherent features, such as independence of transcript exon number and independence of differential expression, what makes it particularly suited for detecting alternative splicing events from sequencing data. In order to test and vali-

date our workflow we challenged it with publicly available sequencing data derived from human tissues, and conducted a comparison with eight alternative computational methods. In order to judge the performance of the different methods we constructed a benchmark data set of true positive splicing events across different tissues, agglomerated from public databases, and show that ARH-seq is an accurate, computationally fast and high-performing method for detecting differential splicing events.

References

- Rasche A, Herwig R (2010) ARH: predicting splice variants from genome-wide data with modified entropy. *Bioinformatics* **26**(1), 84-90. <http://dx.doi.org/10.1093/bioinformatics/btp626>
- Rasche A, Lienhard M, Yaspo ML, Lehrach H, Herwig R (2014) ARH-seq: identification of differential splicing in RNA-seq data. *Nucleic Acids Res* **42**(14), e110. <http://dx.doi.org/10.1093/nar/gku495>