Computational Analysis of Alternative Splicing in Cancer using RNA-Seq

Moritz Aschoff[1], Agnes Hotz-Wagenblatt[1], Karl-Heinz Glatting[1], Matthias Fischer[4], Roland Eils[2,3] and Rainer König[2,3]

[1]Bioinformatics "HUSAR", Genomics Proteomics Core Facility, German Cancer Research Center
[2]Department of Theoretical Bioinformatics, German Cancer Research Center
[3]Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, and Bioquant, University of Heidelberg
[4]Department of Pediatric Oncology and Hematology and Center of Molecular Medicine Cologne, University Children's Hospital

m.aschoff@dkfz-heidelberg.de

Alternative splicing is a biological process that produces different mRNA isoforms from a single gene. Quantifying isoform abundances and detecting disease-specific isoforms are important tasks since aberrant splicing is known to be responsible for many diseases [1].
Isoform abundances can be inferred from short sequence reads produced via next-generation RNA sequencing (RNA-Seq). In comparison to microarray based approaches, RNA-Seq provides an even more precise measurement of levels of transcripts and their isoforms. Advantages of RNA-Seq include the ability to detect previously unknown splicing events, high sensitivity, low background signal as well as good reproducibility for replicates (cmp. [2]). However, there is no "box standard" software available for transcriptome analyses with RNA-Seq [3]. Therefore, method development efforts are required to enable the detection of disease- or cancer-specific splice variants that might serve as powerful biomarkers and are promising targets for treatment purposes.

In order to detect alternative splicing events that are significantly altered between biological conditions, we combined existing algorithms for short read mapping [4] with statistical approaches to analyse read count data. Our first approach focuses on the detection of single exons that undergo differential skipping or exon-modification events. A complementary approach is based on elementary linear algebra and tries to capture more complex differences in splicing that may affect several exons and originate from multiple isoforms.
We applied our analysis pipeline to high-throughput sequences of the transcriptomes from human brain and liver samples as well as neuroblastoma patients with low grade and high grade mortality (i.e. favourable (stage IVS according to the International Neuroblastoma Staging System [5]) and unfavourable (stage IV) outcome). In addition, we used a supervised machine learning approach to classify samples based on splicing information.

Our method allowed us to detect tumour stage-specific alternative splicing in neuroblastoma and tissue-specific alternative splicing in brain vs. liver samples. Classification performance reached a 85% success rate when classifying neuroblastoma patients and a 100% success rate when classifying brain vs. liver samples. Performance was estimated in a leave-one-out cross validation scheme. For many of the genes that were predicted with our method, we found additional evidence of splicing diversity. This includes reads that span exon-exon junctions, paired-end information that supports the presence of predicted isoforms and information derived from existing transcript annotations. Genes predicted to be differentially spliced between stage IV and stage IVS patients are involved in neuronal processes and apoptosis. Some of the genes are known to have tumour suppressor activity. Predictions are now being confirmed experimentally in the lab by RT-PCR.

[1] Kim E, Goren A and Ast G. Alternative splicing and disease. RNA Biology 2008 Mar 22;5(1).
[2] Wang Z, Gerstein M and Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics. 2009;10(1):57-63.

[3] Costa V, Angelini C, De Feis I and Ciccodicola A. Uncovering the Complexity of Transcriptomes with RNA-Seq. Journal of biomedicine & biotechnology 2010;2010:853916.
[4] Trapnell C, Pachter L and Salzberg S L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England). 2009;25(9):1105-11.
[5] Schwab M, Westermann F, Hero B and Berthold F. Neuroblastoma: biology and molecular and chromosomal pathology. The lancet oncology. 2003;4(8):472–80.