A pipeline for the discovery of alternative splicing events with Affymetrix Exon Arrays

Marc Johannes\textsuperscript{1}, Jan C. Brase\textsuperscript{1}, Holger Fröhlich\textsuperscript{2}, Stephan Gade\textsuperscript{1}, Ruprecht Kuner\textsuperscript{3}, Holger Sültmann\textsuperscript{3}, Tim Beilbarth\textsuperscript{3}

\textsuperscript{1} German Cancer Research Center, Im Neuenheimerfeld 580, 69120 Heidelberg, Germany
\textsuperscript{2} Cellzome AG, Meyerhofstrasse 1, 69117 Heidelberg, Germany
\textsuperscript{3} University of Göttingen, Humboldtallee 32, 37073 Göttingen, Germany

Introduction

Alternative splicing is one of many processes responsible for the diversity of the proteome. This diversity is achieved by including or excluding exons during the post-transcriptional processing. This process might be impaired in cancer-cells, thus, it is of particular interest to look for cancer-specific splice variants.

Method

Our approach to identify those alternatively spliced genes consists of 2 steps. First, fitting of the following linear model to the data

\[ x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \]

and in a second step applying an ANOVA to determine the significant effects. In the model,

- \( x_{ijk} \) denotes the background corrected probe intensities for the ith (\( i = 1, \ldots, n \)) exon and the kth (\( k = 1, \ldots, n_{ij} \)) probe in healthy (\( j = 1 \)) and cancer (\( j = 2 \)) tissue.
- \( \alpha_i \) accounts for splice variation common to both, healthy and diseased tissue.
- \( \beta_j \) represents general differences in the mean expression levels between the two tissue states.
- \( \gamma_{ij} \) captures combinations of m exons and j tissues as shown in figure 5.

Results

After applying the above described methods to the the exon array data, plots like the one shown in figure 6 are produced. With these plots and the corresponding p-values, the biologists are able to decide whether or not they want to further investigate a particular gene and validate the splicing event via RT-PCR.

The method has already been validated on a publicly available colon-cancer data set. Several cancerspecific splicing events that have already been validated by qRT-PCR were identified with low false positive rate.

References


Figures adapted from [1] and [2].