International Research Journal of Biotechnology

Extended Abstract
Vol. 11, Iss. 3
2020

Alternative Splice Variants in Gene Expression Values in Patients with Marfan's Syndrome

Wouter Ouwerkerk* and Aeilko H Zwinderman

Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Academic Medical Center, University of Amsterdam, The Netherlands

ABSTRACT

Alternative splicing of messenger RNAs provides cells with the opportunity to create protein isoforms of a multitude of functions from a single gene by excluding or including exons during post-transcriptional processing. Reconstructing the contribution of these splice variants on the total amount of gene expression remains difficult. We introduced a probabilistic formulation of the alternative splicing reconstruction problem using a finite mixture model, and provide a solution based on the maximum likelihood principle. Our model is based on the assumption that the expected expression level of exons in a particular splice variant is the same for all exons in that variant but allows for measurement error. In this algorithm the expression in a patient can be written as a weighted sum of the number of splice variant mixture multivariate Gaussian densities. We estimated the model parameter by maximizing the total likelihood using a Nelder and Mead optimization algorithm in R. To evaluate our algorithm we compared the AIC/BIC values of six models: Established optimal normal mixture modeling method, all exons are equally transcripted, the currently known splice variants, all possible splice variants, the known variants aided with the high prevalent variants of the all possible variants model, and manually selected splice variants. We applied the models to three genes (SLC2A10, TGFβR2 and FBN1), with 25, 29 and 265 possible splice variants, associated with Marfan's syndrome in gene/exon expression data of 63 patients with Marfan's syndrome. The models with the known splice variants aided with the high prevalent splice variants from the all possible splice variants had the best AI C/BBI C values for all three genes. In SLC2A10 and FBN1 there was one, in TGFβR2 two predominant splice variants. There are five basic modes of alternative splicing (depicted in Figure 1), of which exon skipping is most common in humans Predicting the contribution of these modes of splicing variation on gene expression data is difficult, especially in microarray data which returns highly fragmentary information from probes targeting specific exons or exon-exon junctions In reconstructing splice variants, formulating a splice graph traversal problem can be helpful especially when considering multiple traversals.

Established optimal normal mixture modeling method estimated by Mclust in R This is our null model, and positive control, because it has no constraints. Mclust has the disadvantage that it estimates clusters which almost surely are not corresponding to recognizable splice variants. The Mclust algorithm does not use constraints to simulate the biological process of alternative splicing. Mclust maximizes the likelihood using different parameter values for each exon in each splicing variant The primary objective of the simulation study was to validate our model and algorithms. The simulation was conducted using predefined splice variants and model parameters. We simulated variation of a set of two known splice variants for each of three genes existing of 5, 9 and 65 exons, and added one unknown variant.

Keywords: Alternative splicing; Marfan syndrome; Gene expression