Poster A-22

Comparison of Alternative Splicing Structures in Eukaryotes

Authors:
Sammeth, M (*GRIB - IMIM*)
Foissac, S (*GRIB - CRG*)
Guigó, R (*GRIB - IMIM*)

Short Abstract: In order to investigate the evolution of alternative splicing (AS), we compared the structural changes in AS patterns of orthologous genes in 10 species. Amongst other, we show species-specific biases in the distribution of the patterns and novel metrices to quantify the distance between different AS patterns.

Long Abstract:
Background: Recent analyzes of the complete human genome sequence have revealed the contrast between the molecular complexity of the cell and the apparently much lower number of identified genes. Several investigations have identified the phenomenon of alternative splicing (AS) as th main source of this molecular complexity. Moreover, AS has been shown to enable the genes to adapt the expression levels of variants to the corresponding cellular context, developmental status or state of disease.

Motivation: A still unsolved task is the investigation of evolutive mechanisms affecting AS in a gene. Preliminary work has focussed mainly on the investigation of variation in binding motifs for transcription factors to explain changes in a respective AS pattern. The strucutural comparison of AS patterns in orthologous genes across different genomes has still been unexplored.

Method: We present novel ways to describe and to compare AS patterns, including (1) a non-redundant notation system for AS patterns, (2) descriptive measures for AS structures, and (3) a novel metrics to quantify the evolutionary distance between the structures of AS patterns. Concerning (1), we describe a graph based on splice sites to flexibly model AS patterns without biasing for exons or introns. For a convenient analysis, the graph is projected into a non-redundant matrix representation that is highlighting splice sites shared by more than one variant. The notation is capable of describing arbitrarily long AS events with 2 or more variants. In the progress of (2), we firstly show the lack of definition for the general use of many terms established to describe AS. For instance, what is an AS event when regarding more than two variants at a time? Using the notation system of (1), we give formal definition for these terms to form a consistent language. Afterwards, novel measures have been defined to capture attributes of AS patterns, e.g., the degree (a complexity measure). To finally describe the difference between 2 AS patterns of corresponding genes (orthologs/paralogs), we set up a distance metrics (3).

Result: The described methods are applied to the genomic data of 10 species (human, mouse, rat, dog, cow, chicken, frog, fish, fruitfly, mosquito) as extracted from the Ensembl database. We show (1) the limitations of traditional terms to describe AS events in contrast

to the non-redundant matrix representation. Pooling the variants of a genome in groups of structurally distinct AS events, we created an atlas of AS patterns and their distribution in each of the genomes under investigation. By this, we show significant species-specific biases in the distribution of AS structures. In a further analysis, we compare more abstract attributes in the AS events. Using the novel descriptive measures from (2), we infer numerical values from the structures found in (1) which then are plotted along the genome and analyzed graphically across the 10 selected species. To evaluate the developed distance metrics (3), we retrieve orthologous and paralogous gene pairs from the EnsMart project. In the end, phylogenetic trees are inferred from the calculated distances and we present the graphs of some representative genes.