

LIFE SCIENCE INTERFACE/ DOCTORAL TRAINING CENTRE
First Year Project Proposal - 2005

Supervisor: Jotun Hein

Summer or Autumn Project

Title of Project: How many transcripts does it take to reconstruct the Splice Graph?

Description of project:

Motivation and Background: In recent years the phenomena of alternative splicing has emerged as functionally very important and occurring with high frequency. Alternative splicing is typically detected by sampling a series of transcripts that have been made by copying different configurations of exons. The alternative splicing graph (ASG) is then constructed so it minimally can explain these transcripts. The question here naturally arises, how small this minimal alternative splicing graph is relative to the true alternative splicing graph. This questions could be addressed both through analysing an increasing set of real transcripts from a data base and see how the reconstructed ancestral recombination graph grows or through a model of ASG with probability assigned to different paths. This could give an evaluation of how many transcripts to know the alternative splicing graph to a satisfactory degree.

Workplan: Write algorithm and small program that can construct minimal ASG from a series of transcripts. Test that this works on annotated human genome at NCBI/Ensembl. Use this program to test reliability from i) real data and from ii) simulated data. A measure of the size of the reconstructed ASG is needed here and could for instance be the number of nodes/edges or paths. An interesting question would here be if the ASG for a subset of transcript always is a sub-graph of the full ASG.

i. for a set or real genes with associated transcripts, randomize the order of the transcripts and reconstruct an increasing series of ASGs.

ii. Stochastic ASGs. This can be done at different levels of realism, starting with a segment of the real line and a series of "introns" that can be skipped by loops exiting and entering the real line to much more realistic models where the gene was modelled as DNA with as many features of a AS gene as possible. There is a limit to how realistic a model can be, since AS is not well understood. For these model (naïve or complex), probabilities to different paths (transcripts) must be chosen. The ASG will now generate transcripts with well defined probabilities.

Generate transcripts and reconstruct the ASGs. A question here will be what the minimal number of transcripts is that allows reconstruction of the full graph.

Comment: In continuation of this there are a long series of questions that could be addressed: How do you compare AS-graphs from different species? How do you determine by comparison, which alternative transcripts exists since they are functional and which are noise? If the project proceeds quickly, then it could be a possibility to make a stochastic model of how such a graph evolves and see if would be possible to find selected transcripts.

Literature

MECHANISMS OF ALTERNATIVE PRE-MESSENGER RNA SPLICING Douglas L. Black Annual Review of Biochemistry. Volume 72, Page 291-336, Jul 2003

Caceres JF, Kornblihtt AR. Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet. 2002 Apr;18(4):186-93.

Roca, X. et al.(2003) "Intrinsic differences between authentic and cryptic 5' splice sites" Nuc.Ac.Res. 31.21.6321-33.

Maniatis,T. and Tasic,B (2002) "Alternative pre-mRNA splicing and proteome expansion in metazoans" Nature
Fairbrother,WG et al.(2002) "Predictive Identification of Exonic Splicing Enhancers in Human Genes" Science 297.1007-

Location: *Oxford Centre for Gene Function*

Consumables required: