

Bioinformatics Day on "Alternative Splicing"

April 24th 2008 - 1pm - 5pm in the Oxford Centre for Gene Function, Oxford

On April 28th a Bioinformatics Day with focus on alternate splicing will take place. If you wish to attend this event, please send an email to Madeline Mitchell (mitchell@stats.ox.ac.uk) with your name and affiliation. There is a limit of 100 people in the seminar room in the OCGF, if early indications are that this is exceeded other venues can be found. Refreshments will be provided and are generously sponsored by Affymetrix.

1.00 Jotun Hein: **Short Welcome**

1.10 Jennifer Taylor (Wellcome Trust Centre for Human Genetics - WTCHG): **High-throughput observation of alternate splicing : Progress to date**

The observation of transcript diversity arising through alternate splicing on genome-wide is now a reality, made possible by emerging high-throughput approaches such as the Affymetrix Human Exon array, high-resolution tiling arrays and resequencing approaches. As expression profiling datasets emerge using these platforms, there is a substantial research effort underway to develop relevant analytical approaches that robustly normalize and summarise the data and infer transcript splice forms. This introduction will discuss the major high-throughput platforms sensitive to alternate splicing and the challenges and approaches identified to date with respect to mining these rich datasets.

1.35 Deborah Keenan (Affymetrix UK Ltd): **Genome-wide Alternative Splicing Analysis on a Single Array – The Affymetrix GeneChip® Exon 1.0 ST Array**

Alternative splicing is a major source of protein diversity for higher eukaryotic organisms and is frequently regulated in a developmental stage-specific or tissue-specific manner. Current estimates suggest that 50-75% (or more) of human genes have multiple isoforms. Exon-level expression profiling can now be conducted on a single microarray. The Affymetrix GeneChip® Exon Arrays, combined with the whole transcript amplification assay (WTA), enable researchers to understand at a global level how alternative splicing impacts crucial regulatory mechanisms in development, differentiation and disease pathogenesis. GeneChip Exon® arrays represent almost all protein-coding transcripts in the genome, including both empirically supported and predicted transcribed sequences for novel isoform discovery. Affymetrix Expression Console™ Software provides a platform for primary exon-level data analysis data quality control and normalisation. The NetAffx™ Analysis Centre provides array design information, probe sequence and functional annotation for faster translation of array data to biological stories. Further statistical evaluation and biological interpretation can be achieved through a number of GeneChip® compatible third-party software providers. Affymetrix continues to support academic and third-party software developers to further evolve ways for exploring Exon array data.

2.00 George Nicholson (Dept Statistics): **A robust statistical method for the detection of alternatively spliced mRNAs from Affymetrix exon arrays**

It is estimated that as many as 75% of genes in the human genome engage in production of multiple alternative splice variants (*mRNA isoforms*) from the same base sequence. Knowledge of which isoforms are created on specific physiological, genetic and environmental backgrounds will expedite the discovery of novel associations - as well as the characterisation of known ones - between genetic polymorphism and disease, and is also an important step in the integration of transcriptomics and proteomics.

The development of microarray technology over the past decade has permitted the study of mRNA transcript abundance on a genome-wide scale. It is now feasible to assess simultaneously the level of expression at all known exons in the human genome, using the Affymetrix Exon 1.0 ST Array platform. Statistical methodology for extracting informative signal from such a wealth of data is required. Here we describe a statistical method for the robust discovery of novel splice variants; the method is designed to attain a low rate of false discoveries.

2.25 Claire Vandiedonck (WTCHG): **Mapping haplotype-specific gene expression and alternative splicing in the human Major Histocompatibility Complex**

The human Major Histocompatibility Complex (MHC) has been a paradigm for genomics and disease association mapping. This is the most gene dense and polymorphic region in the genome with nearly half of known genes involved in the immune response. The region is associated with many diseases, notably autoimmune disorders, and recently the MHC was entirely re-sequenced for frequent haplotypes associated with common autoimmune diseases. Little is known about splicing in the MHC, but alternative splicing is predicted to play a major role in the immune system. Here, using an original hybrid array, including both a tiling path for the discovery of new transcripts and splicing probes to monitor every known or predicted splicing events, we draw the first transcriptional map of the MHC in the context of immune related haplotypes in lymphoblastoid cell lines. This study is the initial step of our attempt to map MHC expression and splicing QTLs which should facilitate fine mapping and functional analysis of genetic polymorphisms for future population-based association studies.

2.50 -3.30 BREAK

3.30 Rune Lyngsø (Dept. Statistics): **How many transcripts does it take to reconstruct the splice graph?**

Alternative splicing has emerged as an important biological process which increases the number of transcripts obtainable from a gene. Given a sample of transcripts, the alternative splicing graph (ASG) can be constructed-a mathematical object minimally explaining these transcripts. Most research has so far been devoted to the reconstruction of ASGs from a sample of transcripts, but little has been done on the confidence we can have in these ASGs providing the full picture of alternative splicing. There is evidence of a stochastic element to splicing events, so we introduce probabilistic models of transcript generation, under which growth of the inferred ASG is investigated. These models are used in novel methods to determine the nature of the collection of real transcripts from which the ASG was derived.

3.55 Gromak (Dunn School of Pathology): **Co-transcriptional cleavage of intronic RNAs as a mechanism of eukaryotic gene regulation**

Most human genes contain relatively short exons separated by long tracts of non-coding intronic sequences. Transcription of these long introns can result in the accumulation of vast amounts of pre-mRNA near the site of transcription which might interfere with the ability of the splicing machinery to find the correct splice sites. Therefore, removal of long introns from the nascent pre-mRNA transcript is likely to be a critical process requiring careful regulation. Our previous studies have demonstrated that exons of the nascent pre-mRNA are tethered to the elongating RNA polymerase II transcription complex, and that the continuity of the nascent intronic sequences is not essential for efficient constitutive splicing. Here we show that intronic cleavage can modulate alternative splicing if it is positioned between the exon and its intronic regulatory sequences. We speculate that many human introns contain naturally occurring target sequences for endonucleolytic co-transcriptional cleavage activity.

4.20 Eleanor Whitfield (EMBL-EBI, UK): **The Alternative Splicing and Transcript Diversity Database**

The Alternative Splicing and Transcript Diversity (ASTD) database is a new database available at the European Bioinformatics Institute. This database is a result of the merging, continuation and improvement of the Alternative Splicing Database (ASD) and the Alternative Transcript Diversity database (ATD), previously maintained at the EBI.

The ASTD database has genome wide alternative splicing data for human, mouse and rat. The aim of the database is to predict full-length transcripts for all three species using all publically available EST and mRNA data. Pairwise comparison of transcripts for each gene, allows us to display alternative splice events within the transcriptome. Transcripts are annotated for alternative transcription start sites, alternative polyadenylation sites, splice events, SNPs and splice site conservation information between orthologous species. eVOC annotations for each transcript also allows us to display potential expression patterns. Translations are derived for each transcript where possible, and, if relevant, are tagged as potential candidates for nonsense-mediated RNA decay.