## Poster L-22

**Splicing Arrays: Profiling transcriptome by statistical analysis of mRNA isoform expression**


**Authors:**
Murlidharan Nair (*IUSB*)
Hai-Ri, Li (*UCSD*)
Jessica A Whitaker (*IUSB*)
Xiang-Dong Fu (*UCSD*)

**Short Abstract:** Alternative splicing is a major source of proteome diversity. mRNA-isoform levels reflect transcriptional and post-transcriptional processing. Ability to detect the levels of each isoform in the cell significantly increases our understanding of the state of the cell. Using statistical approaches of multiple comparisons we have profiled transcriptome from isofom data.

**Long Abstract:**

Alternative splicing involves a combinatorial editing wherein two or more exons are joined together. Alternative splicing is a major source of proteome diversity. Querying the uniquely edited forms (mRNA isoforms) is reflective of events that occur during transcription as well as those that occur during the processing of the transcript. This processing of the primary transcript which is responsible for removing the introns is carried out by the splicing machinery. The resulting processed transcript is either translated into a protein or may serve as a functional RNA transcript. Splicing depends on the proper recognition of exons and needs to carried-out with extreme fidelity. Mutations annotated on the human genome reveals that about 10% affect canonical splice site sequence. Thus being able to detect the levels of each isoform in the cell significantly increases our understanding of the state of the cell.

Conventional microarrys that have been in use for sometime now are capable of detecting only the primary transcript and do not convey the complete picture of the state of the transcript that a cell harbors as a result of splicing. Since each transcript that result from splicing is capable for being translated into a protein product, the proteome picture that conventional arrays portray is not accurate.

In our recent study (Li, et al., 2006) we constructed splicing arrays to examine ~1500 mRNA isoforms from a set of genes that had been implicated in prostate cancer. The goal of the work was to identify signature mRNA isoforms that are characteristic of prostate cancer. In the study we also used the DASL assay (cDNA-mediated annealing, selection, extension and ligation) (Fan, et al., 2004) to circumvent the problem when dealing with partially degraded biological samples. The data so obtained was reflective of the levels of isoforms that were targeted for each gene.

The analysis of the data in itself presented a challenge. Being able to extract different views of the data that would be reflective of different underlying biological processes would be useful in understanding the state of the cell more thoroughly. One of the main questions we were interested in addressing was whether the data could reveal correlations or links between the coupled processes of transcriptional and posttranscriptional regulation.

Comparing multiple isoforms of the same gene and across different cell and tissue types can be elegantly performed using statistical multiple comparisons, which form the essence of hypothesis testing and model simplification in ANOVA. Using this approach we have compared the group means of expression of isoforms between cell and tissue types with a view to profile transcriptome. The approach also helps in identifying orthogonal sets of isoforms that can then be used for classification. In this poster we present our recent results of using multiple comparisons and new data analysis methods to reveal views of the data that provided significant biological insight into the underlying processes of transcriptional and post-transcriptional regulation.

Li, H.R., Wang-Rodriguez,J.,Nair, T.M, et al. Two-Dimensional Transcriptome Profiling: Identification of Messenger RNA Isoform Signatures in Prostate Cancer from Archived Paraffin-Embedded Cancer Specimens. Cancer Research 2006:66; 4079-88

Fan, J.B., Yeakley, J. M., Bibikova, M. et al. A versatile assay for high-throughput gene expression profiling on universal array matrices. Genome Res 2004:14; 878-85