Abstract

We developed a genome-based analysis method that uses data from expression microarray experiments (Affymetrix platform) to investigate alternative mRNA processing. This approach revealed numerous examples of genes whose expression patterns suggest strain-dependent differential mRNA processing in mouse.

Introduction

Current Affymetrix expression microarray designs typically include numerous redundant probe sets that interrogate different regions of the same gene. This is a side effect of a design process which creates probe sets matching alternative splicing or other types of mRNA variants. Thus, when redundant probe sets produce discordant results, such as differential expression in opposite directions, the most logical explanation is that the condition under investigation affects mRNA processing pathways, such as alternative splicing and alternative polyadenylation site choice.

Both types of alternative mRNA processing can have profound impacts on gene function when important functional motifs, such as conserved protein motifs or mRNA stability determinants, are affected. Discordant behavior of redundant probe sets therefore may represent an opportunity to study how alternative mRNA processing pathways operate. Here we investigate this idea using expression data collected from two mouse strains.

Methods

Data collection and pre-processing.

Labeled mRNA samples were prepared from the livers of six mice, three of strain AJ and three of strain B6. Labeled samples were hybridized to two Affymetrix 430_2 microarrays per mouse. The array data were processed using RMA and normalized using quantile-quantile normalization.

Redundant probe sets.

Affymetrix provides annotation data files that map probe set ids to Entrez Gene ids as well as "pat" files that map probe set design sequences onto the May 2004 mouse genome. We used these files to create and then screen a provisional list of redundant probe sets for use in our study.

Statistical Methodology

We used mixed effect ANOVA model to identify genes in which the redundant probe sets produced discordant differential expression readings across strains. For each gene, we fitted a mixed effects linear model:

\[ y_{ijk} = \mu + S_i + P_k + S_iP_k + M_{ijkl} + \epsilon_{ijk} \]

where \( y_{ijk} \) is a measure of gene expression produced by an individual probe set; \( \mu \) is the grand mean of all probe set readings for the gene being considered.

Results

The analysis identified 1,400 genes with adjusted p-values less than or equal to 0.05. We examined a number of these using the Integrated Genome Browser - see below [1]. In some cases, alternative mRNA processing alters the coding region and modifies conserved functional motifs, a common occurrence among alternatively spliced genes [2,3].

**Dypsi2 dihydroxypyrimidinase-like 2**

Conclusion

Based on these preliminary results, we conclude that analyzing the relative expression of redundant probe sets from Affymetrix expression microarray designs has the potential to reveal how diverse experimental conditions affect differential mRNA processing and degradation pathways.

References

[1] The Integrated Genome Browser is open source software available from [www.genome.org](http://www.genome.org)
