Poster G-2

Conservative sequences and proteins

Authors:
Yehoshua Sobolevsky (*Genome Diversity Center, Institute of Evolution, Haifa University*)
Edward N. Trifonov (*Genome diversity Center, Institute of Evolution, Haifa University*)

Short Abstract: Decamers which are present in at least 8 of total 15 (selected from different families) prokaryotic proteomes are listed. They are clustered into 201 groups of closely related decamers. The type(s) of proteins which contain the decamers from each group is presented. Several frequent group combinations and corresponding protein types are given.

Long Abstract:
15 bacterial genomes from different families are chosen. They are: Streptomyces coelicolor, Bradyrhizobium japonicum, Rhodopirellula baltica, Bacteroides thetaiotaomicron, Bacillus cereus, Methanosarcina acetivorans, Gloeobacter violaceus, Treponema denticola, Sulfolobus solfataricus, Thermus thermophilus, Fusobacterium nucleatum, Thermotoga maritima, Aquifex aeolicus, Chlamydophila pneumoniae, Nanoarchaeum equitans List of conservative decamers which are present in at least 8 of total 15 proteomes is obtained with help of method similar to Suffix Tree. This decamers are clustered into 201 groups of closely related decamers. It turns out that there is only one or sometimes few type of proteins which contain decamers from arbitrary definite group. The list of pairs (group and corresponding protein type) is presented. Protein may contains several conserved decamers in its sequence structure. These decamers belong to definite groups and so this protein gives us the definite combination of groups with definite distances between the groups (each group has definite main point or position with great frequency of occurrence). Several combinations of high occurrence are presented here.

Poster G-2

Comparison of Gene Expression in Murine Retina by SAGE and 2D-PAGE

Authors:
Tim Alcon (*Bioinformatics and Computational Biology, Iowa State University*)
Heather West Greenlee (*Biomedical Sciences, Iowa State University*)
Vasant Honavar (*Computer Science, Iowa State University*)

Short Abstract: Knowledge of gene expression can give insight into the networks of interactions active during development. We compare mRNA expression data from SAGE with protein abundance data from 2D-PAGE for five timepoints during the development of the murine retina. We find low overall correlation between mRNA abundance and protein abundance.

Long Abstract:

Knowing what genes are being expressed at various times during a developmental process can give insight into the pathways and networks of interactions that participate in the developmental program. Two methods that have been widely used to measure gene expression are SAGE and 2D-PAGE. SAGE (serial analysis of gene expression) is a method of identifying the population of mRNA transcripts present in cells. 2D-PAGE (two-dimensional polyacrylamide gel electrophoresis) together with tandem mass spectrometry is a method for identifying the protein population of cells. By comparing the data obtained by these two different methods, we can assess how well mRNA expression correlates with protein expression, and whether using both kinds of data might tell us more about the system being studied than either kind by itself. We compared expression data from these two different methods for the developing murine retina. The initial overlap between the two data sets consisted of 73 genes, but filtering out those spots with multiple protein IDs narrowed the list down to 55 genes. We combined the expression for distinct spots having the same protein ID, and also combined the expression for different tags that mapped to the same gene. We found a low level of concordance between the two methods (Spearman rank correlations ranging between .24 and .47 over the five different ages of development used in this study). These results indicate that such phenomena as post-translational modifications of proteins and differing lifetimes of proteins before they are degraded can have a significant effect on the levels of proteins detected. This suggests that transcript data and proteomics data can be used as complemetary sources of information regarding biological pathways. Where there are close matches, we can have increased confidence in the reliability of that data. Where there are discrepancies, we can investigate why they exist and perhaps be lead to more subtle relationships between genes.

Very few studies have examined development by comparing time-series data for mRNA expression and protein expression. To the best of our knowledge, those that have done so have used only two timepoints, whereas this study used five, permitting more informative comparisons of the dynamics of expression of individual genes during development.