Finding Time-Patterns in Temporal Gene Expression Data
Raveena Pendyam, Shahrazad Eslamian, Swati Gupta, Christopher Theisen, Guenter Tuscher (PhD)
Medical and Bioinformatics Graduate Program, School of Computing and Information Systems
Grand Valley State University, Grand Rapids, MI, USA

Introduction
Developments in microarray technology have led to similar or related biological questions. The statistical methodology of meta-analysis aims to combine results from independent but related studies. For example, a meta-analysis of five circadian microarray studies of Drosophila helped researchers to identify a novel set of rhythmically expressed genes. We advocate here a related approach to potentially extend confirmed results to other species or organs. In translational medicine or biology research, it is often based on measurements that have been obtained at different points in time. The biologist looks at these values not as individual points, but as a progression over time.

Our program (SPOT) helps the researcher find these patterns in large sets of microarray data. A researcher proceeds through three subsequent steps: first, selection of microarray data of interesting genes from NCBI GEO, second, translating the temporal measurements into time intervals, and third, defining temporal concepts like “peaks” based on those intervals. Then he/she can search for genes that exhibit that particular pattern within the previously selected data pool. With the continued growth in volume and complexity of gene expression data that are available.

Methods
One major challenge for comparing similar time course gene expression studies across different platforms is to probe mappings. The MicroArray Quality Consortium (FDDA supported) evaluated the concordance of measurements across seven platforms for humans on a validated dataset. One finding is that data from different platforms can be compared under certain precautions by mapping proofs to the RefSeq or AceView databases. We believe that in time-oriented studies a direct comparison at the probe level can be avoided by a transformation using a combination of a statistical and a knowledge-based abstraction (KSTA).

We use KSTA for temporal modeling that allows for the conversion of quantitative data into an interval-based qualitative representation. The resulting intervals are classified as "increasing", "decreasing", or "constant" (more levels possible) by use of the moderated t-statistic. A peak would be described by "increasing" interval followed immediately by a "decreasing" interval. The moderated t-statistic allows borrowing information across a group of genes to smooth variances and uses a single posterior variance estimate in a test.

The group of genes can be chosen from the same KEGG pathway as the selected genes. Our model assumes biological significance is measured by statistical significance, which is the implicit assumption in most gene expression time series publications. We expanded our platform SPOT that connects the user through a web interface hosted on an Apache server and utilizes R and a MySQL database, PHP, and JavaScript. The user can search the repository across selected platforms for time pattern like “peaks”. Our use case for testing the system was "Injuries". We report the results.

Results
For example, via NCBI GEO or EMBL ArrayExpress, it is important that researchers have efficient, flexible, powerful tools to query those data. NCBI GEO or GEOnetDB provide those tools for most types of searches. However, it is still difficult to adequately search for temporal patterns in time course gene expression experiments especially across different platforms. Those searches are necessary to explore biological processes. Temporal gene expression profiles allow for a more adequate characterization of gene function because biological systems are both developmental and dynamic. Data from time-oriented studies allow to study gene expression changes over time and thus to detect differential genes.

For example, via NCBI GEO or EMBL ArrayExpress, it is important that researchers have efficient, flexible, powerful tools to query those data. NCBI GEO or GEOnetDB provide those tools for most types of searches. However, it is still difficult to adequately search for temporal patterns in time course gene expression experiments especially across different platforms. Those searches are necessary to explore biological processes. Temporal gene expression profiles allow for a more adequate characterization of gene function because biological systems are both developmental and dynamic. Data from time-oriented studies allow to study gene expression changes over time and thus to detect differential genes.

Acknowledgments:
We would like to thank the following individuals without whose support this study would not have been possible: Dr. Anuška Dodečnik, Martin O'Connor (Stanford U) Dr. Craig Heads, Dr. Jeremy Miller (viz) Dr. Timothy Redmond, Dr. Mark Wilson (Stanford U), Ramya Gunala, Jayashree Goppaneni, Vincent K. Sinn, Lakshmi Mammal, Yuka Kitaba (GVSU).

References:

Figure 1: Selection of data of interests

Figure 2: Select time intervals for training

Conclusion
We report the results.