Fourier Analysis of Time Course Microarray Data and Its Relevance to Gene Expression Dynamics

Jerry Chen, F. Valafar, P. Paolini
Computational Science Research Center, San Diego State University, San Diego, CA 92182

Abstract

The overall aim of our biological research is to supplement traditional experimental techniques with computational and engineering methodologies for gaining more insight into gene and protein interaction networks and gene expression dynamics. The following study demonstrates the effectiveness of using a signal processing technique called the Fast Fourier Transform (FFT) on time course microarray (TCM) data for finding genes whose expressions oscillate over time.

Using FFT on previously published yeast TCM data, we find that 313 genes show periodic expression. Interestingly, there are four dominant periodicities, one of which matches the known yeast cell cycle periodicity. Gene annotation and GO functional analysis verify the presence of periodic cell cycle genes within the set of 313 genes found. Thus, Fourier analysis is a valuable tool for understanding gene expression dynamics.

Introduction

One of the exciting, ongoing research areas within the fields of bioinformatics and systems biology is the elucidation of gene and protein networks. While there is a large and important effort towards identifying the specific interactions among genes and proteins, there is also a need to understand the dynamics of gene and protein expression over time.

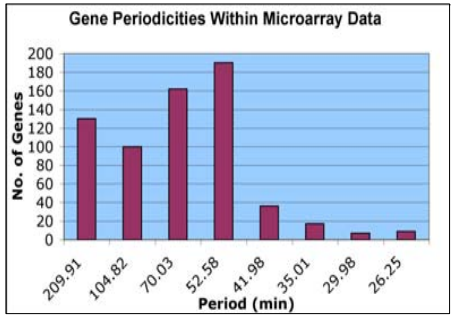
The goal of this study is to use methods from the field of signal processing to understand the dynamics of gene expression. For example, many biological processes such as the cell cycle, cardiac excitation-contraction and the circadian clock are periodic in nature and have underlying genes that are periodically expressed (Bar-Joseph (2004), Ahdesmaki, et.al. (2005)). Many research laboratories have focused on discovering such genes. Our current focus is on using Fourier analysis to study time-course microarrays.

References

Ahdesmaki, M., et al. (2005). BMC Bioinformatics 6: 117.
Bar-Joseph, Z. (2004). Bioinformatics 20(16): 2493-503.
Dennis, G., Jr., et al. (2003). Genome Biol. 4(5): p. P3.
Spellman, P. T., et al. (1998). Mol Biol Cell 9(12): 3273-97.

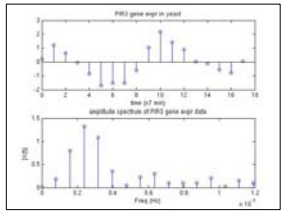
Data and Results

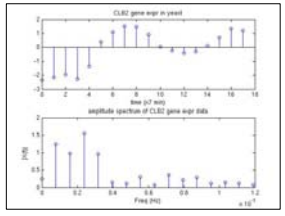
Finding Periodic Genes using Fast Fourier Transform



Left: periodic genes within the time course microarray data

Right: time and frequency domain of two representative genes





Annotation of Genes Periodic with the Cell Cycle

2D Heat Map for gene functional groups for cell cycle regulation (left) and mitosis (bottom).

Figure info. Each row is a gene, and each column is a biological process. Green indicates a gene-process association is present, while black indicates absence.

Bottom: significant GO terms found within genes of periodicity 70.03 min.

Category	Term	Count	%	PValue
GOTERM_BP_ALL	**cell division**	121	12.88%	**9.08E-16**
GOTERM_BP_ALL	establishment of cell polarity	56	4.55%	9.37E-13
GOTERM_CC_ALL	chromosome	48	12.88%	3.43E-12
GOTERM_BP_ALL	**cell cycle**	47	**23.48%**	6.29E-12
GOTERM_BP_ALL	cell communication	35	6.06%	2.95E-11
GOTERM_BP_ALL	cell budding	34	6.06%	7.62E-11
GOTERM_CC_ALL	chromatin	34	6.06%	2.16E-10
GOTERM_BP_ALL	establishment and/or maintenance of chromatin architecture	34	6.82%	2.30E-10
GOTERM_BP_ALL	establishment and/or maintenance of cell polarity	29	4.55%	1.11E-09
GOTERM_BP_ALL	cytokinesis, site selection	23	4.55%	1.89E-09
GOTERM_BP_ALL	**M phase of mitotic cell cycle**	22	8.33%	2.11E-09
GOTERM_BP_ALL	M phase	22	9.85%	2.26E-09
GOTERM_BP_ALL	cytokinesis, completion of separation	22	3.79%	2.36E-08
GOTERM_BP_ALL	**cytokinesis**	21	12.12%	5.27E-08
GOTERM_BP_ALL	interphase of mitotic cell cycle	20	7.58%	1.00E-07
GOTERM_CC_ALL	cell wall	20	19.70%	1.03E-07
GOTERM_BP_ALL	interphase	20	7.58%	1.99E-07
GOTERM_BP_ALL	**G1/S transition of mitotic cell cycle**	12	3.79%	1.11E-05
GOTERM_BP_ALL	regulation of cyclin dependent protein kinase activity	12	4.55%	1.15E-05
GOTERM_BP_ALL	DNA-dependent DNA replication	10	6.82%	6.86E-05
GOTERM_BP_ALL	**DNA replication**	10	9.09%	2.19E-04
GOTERM_BP_ALL	DNA repair	10	8.33%	2.43E-04
GOTERM_BP_ALL	**regulation of cell cycle**	9	13.64%	4.75E-04
GOTERM_BP_ALL	chromatin assembly or disassembly	8	6.82%	0.00163023
GOTERM_BP_ALL	mitotic cell cycle	8	16.67%	0.003679
GOTERM_BP_ALL	regulation of progression through cell cycle	7	13.64%	0.00427045
GOTERM_BP_ALL	**mitosis**	7	8.33%	0.00444373

Methods

Microarray Data
Time course microarray data is taken from published material for the Yeast Cell Cycle Analysis Project at Stanford (http://cellcycle-www.stanford.edu), analyzed in Spellman, et.al. Briefly, gene expression data were obtained every 7 min for 119 min from yeast cells synchronized by α-factor arrest. Expression fold change was measured by comparing synchronized cells to a non-synchronized control group.

Fast Fourier Transform (FFT)
FFT is a method used to compute the discrete Fourier transform of an evenly-spaced finite length signal. It converts a signal in the time domain into the frequency domain, showing the magnitude of each frequency component present within the signal. The formula is given by

$$X[k] = \frac{1}{N}\sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)nk}$$

where N is the signal length and k is the frequency. The FFT function in the MATLAB Signal Processing toolbox is used. Missing data values are filled in using linear interpolation.

Gene Annotation and Analysis
Gene annotation is done using the *S.Cerevisiae* genome database (http://www.yeastgenome.org). Heat maps and significant GO terms were obtained by using analysis tools provided online at Database for Annotation, Visualization and Integrated Discovery (DAVID; http://david.abcc.ncifcrf.gov).

Conclusion

Signal processing techniques such as Fourier analysis are a valuable tool for studying gene expression data from time-course microarrays and discovering genes involved in periodic biological processes. Future directions include (1) developing algorithms that can accurately handle missing, unevenly spaced and short time course data; and (2) annotating and analyzing the genes within this data set underlying other periodicities.

Acknowledgments

This work was supported by NIH NIDDK grant 5K90DK071512-03 "Training for a New Interdisciplinary Workforce" (J.C. is a predoctoral fellow), and by a grant from the California Metabolic Research Foundation (to P.P.).