**NET GEMM: Network Embedded analysis of Temporal Gene Expression using Mixed Models**

[Authors]

[Affiliation]

**ABSTRACT**

**Motivation**

Microarrays have become a routine tool in the biological enquiry geared to studying the impact of genetic or environmental perturbations. The outcome of a typical microarray experiment is the expression of the genes in the condition under which the experiment was performed. This results in a vast amount of data that is used to identify genes that are sensitive to the perturbation, the transcription factors regulating them, etc. These conventional methods (log-fold changes, clustering algorithms, etc) successfully identified condition-dependent transcriptional regulation. Applying this to temporal gene expression data overlooks the effect of time by inherently assuming that the conditions (time points) are independent. Furthermore, present methods to analyze gene expression data ignore the effect of the underlying interaction network of genes or proteins. The data is merely overlaid on the network to identify interactions among significantly changed genes. Substantial amount of additional information on the interaction dynamics could be obtained if time were to be treated appropriately in the context of the interaction network.

**Results**

In this manuscript, we please summarize the method…[describe briefly]. [Present statistics of the method, efficiency, characteristics, etc. How it performed on a toy model.] We used this method to analyze the dynamics of interaction between proteins based on the expression of the corresponding genes in *Saccharomyces cerevisiae*. We used two publicly available time-series datasets for the analysis. The first one measures the changes in the expression of genes during the transition from carbon-limitation to nitrogen limitation under aerobic or anaerobic conditions. We demonstrate the utility of the model in capturing environmental perturbations. The second dataset describes the transcriptional changes in *sfp1*Δ strain of *S. cerevisiae* and its reference strain following sudden exposure to glucose. The two experiments involve substantial transcriptional programming to adapt to environmental conditions or in the ability to assimilate environmental signals to coordinate metabolism with cellular processes. The method provided interesting insights into [need to be filled up after looking at the results].

**Availability**

The source code for NET GEMM is available from <http://www.website> [will fill in before submission. There should also be a readme.txt that should have info on input files needed, their formatting, choosing parameters, output file description and their interpretation].

1. **INTRODUCTION**

Microarrays have replaced the conventional method of determining the expression of a few genes with the ability to measure the true transcriptome rapidly [REFS]. They present a snapshot of the expression of the genes at the time of measurement. Given that there are thousands of genes in a genome, microarrays result in a vast amount of data. The conventional methods to analyze data have been (needs very brief elaboration of the methods here)

1. Simple log-fold changes in the expression of genes
2. Cluster analysis (PCA, hierarchical clustering, k-means, etc

These methods have been able to identify the condition-specific gene expression and map the transcription regulatory network by identifying binding sites of the promoters [REFS]. The interest in measuring temporal gene expression is increasing again, since dynamic data can potentially reveal causal relationships between the genes [This should be one of the deliverables from our analysis]. In order to extract such hidden information from time-series data, dedicated methods have to be developed. It is only recently that such methods to analyze temporal gene expression data have begun to be developed [REFS] - …..(a) EDGE, (b) STEM, (c) KELLER , (d) others that I missed. The fundamental feature of these methods is …(i) not take into account the regulatory interactions, (ii) not a true time-series, (iii)..

In this article, we present network-embedded analysis of temporal gene expression using mixture models (NET GEMM) that addresses these limitations by …….

The performance of this method was tested on a “toy model”. [Describe the toy model, how the model behaved, etc].

We applied this method to publicly available time-series gene expression data in *Saccharomyces cerevisiae*. Extensive gene expression datasets, physiological information and high-resolution regulatory interaction networks are available this yeast. We selected two time-series datasets in which the nutritional environment changed with time, one without any genetic perturbations and one with a deletion in the *SFP1* transcription factor. The first dataset consists of expression of genes during the gradual transition from carbon starvation to nitrogen starvation in a D-stat under aerobic or anaerobic conditions (Farzadfard et al., 2010). Almost a fourth of the genome underwent transcriptional changes in response to the transition. The dominant transcription factor that brought about these changes was Sfp1, which is known to assimilate signals from the environment and coordinates growth with metabolism (Marion et al., 2004). The second dataset measures the temporal changes in gene expression upon sudden exposure of a strain of *S. cerevisiae* in which Sfp1 was deleted to glucose (Cipollina et al., 2009).

Our method was able to identify [what new things…. – need to be filled up after the results become available].

1. **METHODS**

2.1 Description of the datasets

Temporal gene expression datasets were downloaded from Gene Expression Omnibus using accession numbers XXXXX and XXXXX. The two datasets were obtained using Affymetrix® platform. The first dataset contained the expression profiles of the genes in *S. cerevisiae* during the transition from carbon limitation to nitrogen limitation under aerobic or anaerobic conditions. The transition was achieved by gradual increment of glucose availability in the feed to the cells, while keeping the nitrogen concentration constant in a D-stat (Farzadfard et al., 2010). Beyond a certain concentration of glucose, nitrogen became the limiting nutrient. The cells underwent changes related to growth rate as well as metabolism. Analysis of genes whose expression significantly changed indicated that Sfp1 transcription factor played a dominant role in the bringing out the response to transition. In the interest of coherence, we chose a dataset that contains the temporal gene expression profiles in *sfp1* deletion mutant and its isogenic reference at different time points after pulsing steadily growing cells with glucose. The data was measured at six time points after the pulse. These data were analyzed using conventional methods, assuming that all time points are independent.

2.2 Construction of the interaction network

The yeast interaction network was constructed using data from previously published datasets. Interactions between proteins that occurred in at least two independent datasets were considered. These interactions were downloaded from BIND (Bader et al., 2003), MIPS (Mewes et al., 2002), MINT (Zanzoni et al., 2002), DIP (Xenarios et al., 2000) and BioGRID (Reguly et al., 2006) and literature data (5-6 references). The construction of this high-confidence network was described in detail previously (Musigkain et al., 2010). The transcriptional regulatory network (interactions between transcription factors and genes) was downloaded directly from YEASTRACT (Pedro et al., 2008). The two networks were combined and the nature of interactions was not distinguished for the analysis.

2.3 Construction of the toy network

2.4 Method description and algorithms

1. RESULTS

3.1 What came out of the toy model?

3.2 Any characteristics of the network properties from the two datasets together?

3.3 Dynamics of the interactions – coordination of metabolism with developmental processes

What are the interactions? How do they change with time? Think of showing the dynamics as a movie in the supplementary information.

3.4 Dynamics of the response to lack of Sfp1

Same thing here, another movie.

3.5 Anything else that I missed?

1. DISCUSSION
2. References